BREITBANDIGE SIGNALAUFBEREITUNG IN EIN- UND MEHRKANAL-MIKROFONANWENDUNGEN

Dissertation

zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften

vorgelegt von
Dipl.-Ing. Markus Noisternig

am
Institut für Elektronische Musik und Akustik
der Universität für Musik und darstellende Kunst Graz

1. Begutachter: O. Univ. Prof. Mag. art. DI Dr. techn. Robert Höldrich
2. Begutachter: Univ. Prof. Dr. phil. Gerhard Eckel

Graz, 31. Oktober 2017

Für Esther, Oskar und meinen Vater.

Danksagung

An erster Stelle gilt mein Dank Prof. Dr. Robert Höldrich für seine wissenschaftliche und methodische Unterstützung während der gesamten Bearbeitungsphase meiner Dissertation.

Außerdem möchte ich mich bei Doz. Dr. Peter Balazs, Dr. Wolfgang Kreuzer, Doz.^{in} Dr.^{in} Diana Stoeva (Österreichische Akademie der Wissenschaften Wien), Dr. Gilles Chardon (Centrale Supélec), Prof. Dr. Boaz Rafaely, Dr. Hai Morgenstern (Ben-Gurion University of the Negev), Dr. Filippo Maria Fazi (University of Southampton) und Dr. Martin Opitz (AKG Acoustics Wien) für die wertvollen Diskussionen und die produktive wissenschaftliche Zusammenarbeit im Rahmen der verschiedenen nationalen und internationalen Forschungskooperationen bedanken, die ganz wesentlich zum Entstehen dieser Arbeit beigetragen haben.

Dr. Franz Graf und DI Christoph Reitbauer (Joanneum Research Graz) danke ich für die angenehme und produktive Zusammenarbeit bei der Entwicklung der Hardware für das Mikrofonarray.

Daneben möchte ich mich bei allen Mitarbeiterinnen und Mitarbeitern am Institut für elektronische Musik und Akustik für die stets kollegiale Atmosphäre bedanken. Insbesondere danke ich Dr. Franz Zotter und Thomas Musil für die hilfreichen wissenschaftlichen Diskussionen, die angenehme Atmosphäre und die daraus erwachsene Freundschaft.

Meiner Frau Esther danke ich von ganzem Herzen für ihre unermüdliche Unterstützung, ihre Liebe und Motivation.

Kurzfassung

Das primäre Ziel der breitbandigen Signalaufbereitung liegt in der weitgehenden Unterdrückung von Störgeräuschen und Interferenzsignalen, bei möglichst geringer Verzerrung des Nutzsignals. Hierbei kann prinzipiell zwischen einkanaligen und mehrkanaligen Geräuschreduktionsverfahren unterschieden werden. Eine optimale Lösung des Problems lässt sich dadurch erreichen, dass die residualen Störgeräusche am Ausgang eines Mikrofonarrays durch Nachschalten eines einkanaligen Geräuschreduktionsfilters unterdrückt werden. In echtzeitfähigen Systemen darf dabei die Systemlatenz einen gewissen Grenzwert nicht überschreiten.

Die spektrale Subtraktion gehört zu den am weitesten verbreiteten Methoden in einkanaligen Mikrofonanwendungen. Die in dieser Arbeit vorgeschlagene Variante erlaubt eine flexible Handhabung der Unterdrückung des Musical Noise und ein schnelles Ansprechverhalten zur Reduktion der Verzerrungen transienter Klänge. Eine weitere Möglichkeit den Höreindruck zu verbessern besteht darin, die zeitlichen und spektralen Maskierungseffekte des menschlichen Gehörs zu nutzen, um Störungen nur dort zu unterdrücken, wo sie auch tatsächlich wahrnehmbar sind. Aus diesem Grund wird hier die Spektraltransformation mit einer eigens für diesen Zweck entwickelten recheneffizienten Gammaton-Filterbank mit kurzer Systemlatenz durchgeführt.

Sind Nutz- und Störquellen räumlich voneinander getrennt, kann mit mehrkanaligen Geräuschreduktionsverfahren das Signal-Stör-Verhältnis wesentlich verbessert werden. Mit dem in dieser Arbeit entworfenen robusten adaptiven Beamformer ist auch bei kleinen Mikrofonarrays und einer geringen Anzahl an Sensoren eine hohe und breitbandige richtungsabhängige Verstärkung möglich. Die adaptiven Prozesse gewährleisten eine hohe Robustheit gegenüber Bauteiltoleranzen, Fehlpositionierungen und reflexionsbehaftete Schallwellenausbreitung.

Modale Beamformer ermöglichen eine Signalaufbereitung mit möglichst hoher zeitlicher und räumlicher Bandbreite. Die räumlichen Abtastgitter haben dabei einen wesentlichen Einfluss auf die Fehlertoleranz des Systems. Es wird gezeigt, wie sich mit einigen wenigen Abtastpunkten im Inneren eines Arrays, die Interpolation des Schallfeldes stabilisieren lässt. Der Ansatz zur Berechnung der optimalen Abtastpunkte wird in Folge dahingehend erweitert, dass Rahmenbedingungen (wie z. B. vorgegebene Bauformen) berücksichtigt werden können. Dadurch wird eine einfache Implementierung der optimalen Abtastgitter gewährleistet. Simulationen unterschiedlicher Arraygeometrien zeigen, dass mit der vorgestellten Methode ein minimaler Interpolationsfehler über einen großen Wellenzahlbereich erzielt werden kann.

Abstract

This work studies broadband signal enhancement methods with the aim to suppress noise and interference signals and, at the same time, minimise target signal distortions. In general, they can be subdivided into single-channel and multichannel solutions. It can be shown, that combining a microphone array beamformer with a single-channel noise suppression post-filter can significantly improve the signal-to-noise ratio at the output of the system. To achieve almost real-time capability the overall latency must not exceed a certain threshold.

Spectral subtraction is one of the most popular single-channel approaches to noise suppression and signal enhancement. The proposed approach improves the spectral weighting rules to eliminate musical noise. The fast response time leads to a significant reduction of the distortion of transient sounds. For the time-frequency transform a computationally efficient analysis-synthesis filter bank is developed based on masking properties of the human auditory system. Hence less emphasis is given to frequencies where noise is inaudible. This reduces the target signal distortions and improves the auditory impression.

It is further shown that beamforming with microphone arrays can be used to significantly reduce the effects of noise and interference. A robust adaptive beamformer for small arrays with a low number of microphones is developed, which is based on the Generalised Sidelobe Canceler principle. It provides a high array gain over a broad frequency range and is robust against microphone mismatch, position errors, and multipath propagation.

Modal microphone array beamformers provide signal enhancement within both a large temporal and spatial frequency band. The spatial sampling grids have a strong influence on the error tolerance of the systems. It is shown that only a few microphones are needed in the interior of the considered microphone array to

ensure a low interpolation error in the frequency band of interest, and that most of the microphones have to be located on the boundary of the domain, with a non-uniform density depending on the shape of the domain. It is demonstrated that practical constraints can be included in the optimization process in order to ease the implementation of an array. Comparisons for some particular array geometries with design methods known from the literature are given, showing that the proposed approach results in significantly lower errors over an extended frequency range.

Erklärung

Hiermit bestätige ich, dass mir der *Leitfaden für schriftliche Arbeiten an der KUG* bekannt ist und ich diese Richtlinien eingehalten habe.

Graz, den 31/10/2017

.....
Unterschrift der Verfasserin/des Verfassers

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1

Einleitung

Aufgrund der rasanten technologischen Entwicklung in der Telekommunikation sind akustische Kommunikationssysteme heute nahezu allgegenwärtig (Mobiltelefone, Freisprechanlagen im Auto, Konferenzsysteme, etc.). Diese werden in teils sehr unterschiedlichen akustischen Umgebungen betrieben, die mitunter zu einer erheblichen Beeinträchtigung des Signals führen. Ziel ist die informationstragenden Signale (meist Sprache) möglichst ungestört von allen möglichen Interferenzen (andere Sprecher, Hintergrundgeräusche) zu übertragen. Ein typischer Funktionsblock in diesen Kommunikationssystemen ist daher die Signalaufbereitung. In dieser Arbeit werden die verschiedenen Methoden der Signalaufbereitung ausführlich diskutiert und zu mehreren bis dato offenen Problemen Lösungen vorgeschlagen.

Das primäre Ziel der breitbandigen Signalaufbereitung liegt in der weitgehenden Unterdrückung von Störgeräuschen und Interferenzsignalen, bei möglichst geringer Verzerrung des Nutzsignals und gleichzeitiger Minimierung der tonalen Artefakte. Das Gütekriterium ist dabei der Höreindruck, d. h. die subjektiv empfundene Qualität des Audiosignals. In dieser Arbeit wird nicht nur die spektrale, sondern auch die „räumliche Breitbandigkeit“ betrachtet.

Prinzipiell kann zwischen einkanaligen und mehrkanaligen Geräuschreduktionsverfahren unterschieden werden. Einkanalige Geräuschreduktionsverfahren beruhen auf der Schätzung der unterschiedlichen statistischen und spektralen Eigenschaften von Nutzsignal- und Störsignalkomponenten aus dem gestörten Gesamtsignal. Dieser Ansatz führt auf die klassischen Verfahren der Optimalfilterung, wie zum Beispiel das Wiener-Filter (vgl. Wiener, 1942, 1949; Levinson, 1946),

die spektrale Subtraktion (vgl. Boll, 1979) und das Kalman-Filter (vgl. Kalman, 1960; Kalman und Bucy, 1961). Mehrkanalige Geräuschreduktionsverfahren nutzen, neben den statistischen und spektralen Eigenschaften der Signale, auch die räumliche Trennung von Nutz- und Störquellen. Ganz allgemein, lassen sich mehrkanalige Verfahren in folgende Kategorien einteilen: (i) Geräuschkompensation unter Verwendung eines vom Nutzsignal akustisch entkoppelten Sensors in unmittelbarer Nähe der Störgeräuschquelle, wie z. B. die adaptive Geräuschkompensation (vgl. Widrow et al., 1975). (ii) Geräuschreduktion unter Verwendung einer Gruppe räumlich verteilter Sensoren. Typische Verfahren sind das Mikrofonarray-Beamforming (vgl. Van Veen und Buckley, 1988; Brandstein und Ward, 2001; Van Trees, 2002; Benesty et al., 2008a), das mehrkanalige Wiener-Filter (vgl. Chen et al., 2006; Cornelis et al., 2011; Yong et al., 2013) und das MIMO[1] (vgl. Huang et al., 2006, Kap. 3) Wiener-Filter.

Einkanalige Geräuschreduktionsverfahren

Das klassische Wiener-Filter (vgl. Wiener, 1942, 1949; Levinson, 1946) schätzt ein mit additivem Rauschen überlagertes Nutzsignal aus dem gestörten Gesamtsignal. Das Wiener-Filter ist in dem Sinne optimal, dass es den Mittelwert des quadratischen Fehlers (MMSE)[2] der Schätzung minimiert. Bei der Herleitung des Wiener-Filters wird die Kenntnis der Statistik der Signale vorausgesetzt. Nutzsignal und Störung werden dabei als stationäre, mittelwertfreie Prozesse angenommen, die nicht miteinander korreliert sind. Es ergibt sich ein lineares, zeitinvariantes Optimalfilter (siehe Kap. 4.1). Beim Wiener-Filter kommt es zu erheblichen Verzerrungen des Nutzsignals, die sich teils sehr störend auf den Höreindruck auswirken (vgl. Chen et al., 2006). In Umgebungen, in denen sich die Statistik der Signale zeitlich ändert, ist das Wiener-Filter nicht mehr optimal. Ist jedoch die Voraussetzung der Kurzzeitstationarität erfüllt, lässt sich durch blockweise Verarbeitung der Signale eine optimale Lösung finden. Bei kurzzeitstationären Signalen kommen dabei meist adaptive Filter zum Einsatz. Diese erlernen die Statistik der beteiligten Signale selbständig und approximieren das Optimalfilter (vgl. Haykin, 2002a). Kalman-Filter (Kalman, 1960; Kalman und Bucy, 1961) führen auf eine

[1] *Multiple-input multiple-output* (MIMO).
[2] *Minimum mean square error* (MMSE).

weitere Lösung des Schätzproblems in nichtstationären Umgebungen, werden aber in dieser Arbeit nicht weiter behandelt.

Die spektrale Subtraktion (vgl. Boll, 1979) ist eng mit dem Wiener-Filter verwandt und gehört zu den am weitesten verbreiteten Methoden zur Unterdrückung von Störgeräuschen in einkanaligen Mikrofonanwendungen. Bei der spektralen Subtraktion wird das Kurzzeit-Leistungsdichtespektrum des Störsignals geschätzt und vom gestörten Gesamtsignal abgezogen (vgl. Vary et al., 1998, Kap. 12.4). Dabei wird die Unkorreliertheit von Nutzsignal und Störung vorausgesetzt (siehe Kap. 4.2). In der Literatur finden sich zahlreiche Varianten der spektralen Subtraktion. Diese haben meist eine möglichst weitgehende Reduktion der Verzerrung des Nutzsignals bei gleichzeitiger Minimierung der Reststörungen zum Ziel (vgl. Benesty et al., 2008b, Kap. 44). In praxisnahen Anwendungen sind die Subtraktionsregeln von Ephraim und Malah (EMSR)[3] und die daraus abgeleiteten Verfahren weit verbreitet, da diese die meist als sehr störend empfundenen tonalen Artefakte der spektralen Subtraktion weitestgehend unterdrücken. Das Ephraim-Malah-Filter (vgl. Ephraim und Malah, 1983) kann als Wiener-Filter mit Korrekturterm zur Anpassung der spektralen Gewichte an die lokalen Signal-Stör-Verhältnisse (SNR)[4] interpretiert werden (vgl. Cappé, 1994). Das EMSR schätzt die Amplitude des ungestörten Signals aus dem Kurzzeitspektrum des gestörten Gesamtsignals. Der Kurzzeit-Amplituden-Schätzer (MMSE-STSA;[5] vgl. Ephraim und Malah 1983, 1984) ist für gaußverteilte, mittelwertfreie und statistisch voneinander unabhängige Nutz- und Störsignale optimal. Zur Berücksichtigung des logarithmischen Lautstärkeempfindens des menschlichen Gehörs kann auch die logarithmierte Amplitude des Nutzsignals geschätzt werden (MMSE-LSA;[6] vgl. Ephraim und Malah 1985). Beim MMSE-LSA ist die Verzerrung des Nutzsignals wesentlich geringer als beim MMSE-STSA (vgl. Cohen, 2006) und es wird ein wesentlich höherer PESQ-Score[7] (vgl. ITU-T-R P.862, 2001) erreicht, als mit anderen Methoden der spektralen Subtraktion. Weiterentwicklungen des MMSE-LSA berücksichtigen

[3] *Ephraim and Malah spectral suppression rule* (EMSR).
[4] *Signal to noise ratio* (SNR).
[5] *MMSE short-time spectral amplitude estimator* (MMSE-STSA).
[6] *MMSE short-time log-spectral amplitude estimator* (MMSE-LSA).
[7] *Perceptual evaluation of speech quality* (PESQ).

zudem die Wahrscheinlichkeit für das Auftreten des Nutzsignals (vgl. Malah et al., 1999; Martin et al., 2000) und glätten die spektralen Gewichte über eine rekursive Mittelung (vgl. Doblinger, 1995; Martin, 2001; Cohen, 2001, 2002, 2003, 2004b,c; Cohen und Berdugo, 2001a,b). Die Reststörungen klingen natürlicher, der Höreindruck verbessert sich. Letztere Ansätze sind sehr rechenaufwändig und zum Teil akausal (was eine zusätzliche Signalverzögerung zur Folge hat). Für den Entwurf der in dieser Arbeit betrachteten Zielanwendung – eines recheneffizienten Verfahrens mit möglichst kurzer Signallatenz – ist der von Wolfe und Godsill (2001) vorgeschlagene MMSE-SP[8] Schätzer von besonderem Interesse. Dieser lässt sich sehr recheneffizient implementieren und weist eine den MMSE-STSA/LSA Schätzern vergleichbare Störgeräuschunterdrückung auf. In Kap. 4.3 wird der MMSE-SP dahingehend modifiziert, dass die durch Fehler bei der Schätzung des zeitvarianten Störgeräuschspektrums auftretenden nichtlinearen Verzerrungen möglichst stark reduziert werden (vgl. Zotter, 2004; Noisternig et al., 2009).

Eine weitere Möglichkeit den Höreindruck zu verbessern besteht darin, die zeitlichen und spektralen Maskierungseffekte des menschlichen Gehörs zu nutzen, um Störungen nur dort zu unterdrücken, wo sie auch tatsächlich wahrnehmbar sind (vgl. McAulay und Malpass, 1980; Virag, 1995, 1999; Tsoukalas et al., 1997; Thiemann, 2001; Thiemann und Kabal, 2002; Hu und Loizou, 2003, 2004; Loizou, 2005; Jo und Yoo, 2009, 2010). Wird die Spektraltransformation als Filterbank realisiert, deren Kanäle an die Frequenzgruppen des menschlichen Gehörs angepasst sind, lassen sich die Gewichte der spektralen Subtraktion auf die Mithörschwelle begrenzen (vgl. Lorber und Höldrich, 1997; Gustafsson et al., 1998, 2002; Irino, 1999; Tuffy, 1999; Virag, 1999; Wolfe und Godsill, 2000; Hui, 2000; Lin und Ambikairajah, 2002; Lin et al., 2003). Dadurch werden maskierte Störungen weniger stark reduziert, wodurch sich die Verzerrung des Nutzsignals verringert. Die vorliegende Arbeit greift diesen Ansatz auf, wobei die Spektraltransformation mit der in Kap. 3 vorgestellten parallelen Bank sich überlappender Gammaton-Filter durchgeführt wird (vgl. Zotter, 2004; Noisternig et al., 2009). In der Literatur finden sich zahlreiche Studien zu linearen Gammaton-Filterbänken (vgl. Patterson et al., 1987, 1988; Patterson und Rice, 1987; Holdsworth et al.,

[8] *MMSE spectral power suppression rule*, (MMSE-SP).

1988), nichtlinearen und asymmetrischen Gammaton-Filterbänken (vgl. Carney und Yin, 1988; Carney, 1993; Pflüger et al., 1997; Lin et al., 2001a, 2002), Gammachirp-Filterbänken (vgl. Irino und Unoki, 1997, 1998, 1999, 2001) und komprimierenden Gammachirp-Filterbänken (vgl. Irino und Patterson, 2006a,b; Unoki et al., 2006). Eine umfassende Übersicht findet sich beispielsweise in Lopez-Poveda und Meddis (2001) und Duifhuis (2012). Nichtlineare Gammaton-Filter benötigen einen vergleichsweise hohen Rechenaufwand. Aus diesem Grund werden in dieser Arbeit ausschließlich lineare All-Pole, One-Zero und Three-Zero Gammaton-Filter (s. Kap. 2.3.1) als Teilfilter der auditiven Analyse-Synthese-Filterbank verwendet. Diese lassen sich sehr recheneffizient implementieren (vgl. Slaney, 1993; Lyon, 1996; Lin et al., 2001a).

Eine der wichtigsten Eigenschaften einer Analyse-Synthese-Filterbank ist die (nahezu) perfekte Rekonstruktion des Eingangssignals am Ausgang der Filterbank. Im einfachsten Fall wird das Originalsignal durch Summation der Teilbandsignale rekonstruiert. Allerdings weisen Gammaton-Filterbänke an den Übergängen benachbarter Teilbänder Phasendifferenzen auf, die bei der Summation zu erheblichen Signalverzerrungen führen. Zur Reduktion der Welligkeit am Systemausgang wird die Phase vor der Summation entzerrt (vgl. Hohmann, 2002; Zotter, 2004; Herzke und Hohmann, 2007; Noisternig et al., 2009). Eine weitere Möglichkeit eine nahezu perfekte Rekonstruktion zu erreichen besteht darin, die Synthese mit zeitgespiegelten Gammaton-Filtern (vgl. Kubin und Kleijn, 1999a,b; Irino und Unoki, 1998, 2001) durchzuführen. Um die Kausalität der Synthesefilter zu gewährleisten, ist hierbei eine zusätzliche zeitliche Verzögerung des Signals notwendig. Diese entspricht oft nicht den Anforderungen von Echtzeitsystemen. Über die Ansätze der linearen prädiktiven Entfaltung (vgl. Lin et al., 2001b) und der konvexen Projektion (vgl. Slaney et al., 1994; Sezan und Stark, 1982a,b; Youla und Webb, 1982; Mallat, 1989) lassen sich Sets von optimalen Synthesefiltern mit vergleichsweise geringer Systemlatenz herleiten. Allerdings ist auch hier der Rechenleistungsbedarf relativ hoch. Auch die Frame-Theorie (vgl. Christensen, 2003, 2008; Mallat, 2009; Vetterli et al., 2011) führt auf Synthesefilter mit einer nahezu perfekten Rekonstruktion (vgl. Feldbauer et al., 2005; Strahl und Mertins, 2009). Hierbei wird jedoch eine relativ hohe Anzahl an sich überlappenden Teilbändern benötigt. Aufgrund der hohen Güte der Teilfilter weichen diese stark von den psychophysikali-

schen Abstimmkurven ab. Die Maskierungseigenschaften des menschlichen Gehörs werden nicht mit ausreichender Genauigkeit nachgebildet. Zudem ist auch hier der Rechenaufwand vergleichsweise hoch, was die Implementierung auf digitalen Signalprozessoren erschwert.

Ziel dieser Arbeit ist eine recheneffiziente Implementierung von Filterbänken zur auditiven Signalverarbeitung mit möglichst kurzer Systemlatenz. Zur Signalanalyse wird, wie bereits angemerkt, eine parallele Bank sich überlappender Gammaton-Filter verwendet. Das Synthesefilter rekonstruiert das Originalsignal durch Summation der Teilbandsignale (vgl. Kap. 3). Zur Reduktion der Welligkeit am Ausgang des Systems wurde ein Kriterium abgeleitet, welches die Notwendigkeit angibt, vor der Summation das Vorzeichen zu wechseln (vgl. Zotter, 2004; Noisternig et al., 2009). Im Gegensatz zur Methode von Herzke und Hohmann (2007) erfordert dieser Ansatz keine Drehung der Phase. Anstatt der komplexwertigen Gammaton-Filter, können wesentlich recheneffizientere reellwertige Gammaton-Filter verwendet werden. Darüber hinaus wird das Transferverhalten des menschlichen Außen- und Mittelohres über die in Kap. 2.2.1 vorgestellten Außen-Mittelohr-Filterung nachgebildet.

Mehrkanalige Geräuschreduktionsverfahren

Eine Verbesserung des SNR führt bei einkanaligen Geräuschreduktionsverfahren zu einer Verbesserung des Höreindrucks und einer Verminderung der Höranstrengung (vgl. Weiss et al., 1975; Trine und Van Tasell, 2002; Bitzer et al., 2005; Hu und Loizou, 2008), jedoch nicht unweigerlich zu einer Verbesserung der Verständlichkeit (vgl. Hamacher et al., 2005; Hu und Loizou, 2007). Sind Nutz- und Störquellen räumlich getrennt, lässt sich mit mehrkanaligen Geräuschreduktionsverfahren neben dem Höreindruck auch die Verständlichkeit nachweislich verbessern (vgl. Hamacher et al., 2005; Chen et al., 2006). Dabei wird die Vorzugsrichtung (d. h. das Maximum der Empfindlichkeit) eines Mikrofonarray-Beamformers auf eine bestimmte Schallquelle ausgerichtet. Störungen und Interferenzen aus anderen Raumrichtungen werden, in Abhängigkeit von ihrer Einfallsrichtung und räumlich-zeitlichen Korrelation, abgeschwächt. Die einfachste Form ist der Delay-and-Sum-Beamformer (DAS-BF; vgl. Van Veen und Buckley, 1988; Schelkunoff,

1943; Dudgeon, 1977; Flanagan et al., 1985, 1991; Brandstein und Ward, 2001; Van Trees, 2002, Kap. 2), bei dem die Mikrofonsignale für eine gegebene Vorzugsrichtung durch zeitliche Entzerrung kohärent überlagert werden. Bei der Berechnung der Filtergewichte wird meist davon ausgegangen, dass sich die Schallquelle im Fernfeld des Arrays befindet. Die von Abhayapala et al. (2000) und Doclo und Moonen (2003b) vorgestellten Ansätze sind sowohl auf das Nahfeld als auch auf das Fernfeld anwendbar. Die Richtwirkung des DAS-BF ist stark frequenzabhängig. Beim Filter-and-Sum-Beamformer (FAS-BF) werden die Mikrofonsignale vor der Summation gefiltert. Dies erlaubt einen breitbandigen Entwurf der Richtcharakteristik. Bei der Herleitung der Filterkoeffizienten des FAS-BF wird ganz allgemein zwischen signalunabhängigen und signalabhängigen Ansätzen unterschieden. Bei den signalunabhängigen Ansätzen werden bestimmte Annahmen betreffend der Eigenschaften des Schallfeldes getroffen (vgl. Abschnitt 5.1.3). Entspricht das Schallfeld nicht den Annahmen, sind die Filter nicht mehr optimal. Diese Ansätze umfassen breitbandige Beamformer (vgl. Sydow, 1994; Doclo und Moonen, 2003a; Chen und Ser, 2009), superdirektive Beamformer (vgl. Cox et al., 1986; Bitzer und Simmer, 2001; Bitzer et al., 2001; Doclo und Moonen, 2007), Least-Squares Ansätze (vgl. Algazi und Suk, 1975; Doclo und Moonen, 2003b) und Maximum-SNR-Beamformer (Araki et al., 2007; Warsitz und Haeb-Umbach, 2007; Kolossa et al., 2008; Tanaka und Shiono, 2014). Signalabhängige Beamformer schätzen die statistischen Eigenschaften des Schallfeldes aus den Mikrofonsignalen und adaptieren die Filter, bis diese zur optimalen Lösung konvergieren. In der Praxis sind der Frost-Beamformer (vgl. Frost, 1972) und der Generalised Sidelobe Canceller (GSC; vgl. Griffiths, 1977; Griffiths und Jim, 1982; Buckley und Griffiths, 1986, Kap. 6.7) am weitesten verbreitet. Diese minimieren die Varianz des Signals am Ausgang des Systems, unter der Nebenbedingung unverzerrter Wiedergabe des Signals aus Vorzugsrichtung (MVDR/LCMV)[9]. Der MVDR-Ansatz wird beim GSC in zueinander orthogonalen Unterräumen formuliert. Dies führt auf eine Lösung ohne Zwangsbedingung, die sich in der Praxis relativ einfach implementieren lässt. Dazu wird einem Preprozessor, bestehend aus einem fixen Beamformer und einer Blockiermatrix, ein aktiver Störgeräuschunterdrücker nachgeschaltet. Theoretisch

[9] *Minimum variance distortionless response* (MVDR); *Linearly constrained minimum variance* (LCMV).

wird mit dem GSC eine signifikante Reduktion der Störgeräusche, bei sehr geringer Verzerrung des Nutzsignals erreicht (vgl. Jablon, 1986b; Bitzer et al., 1999b; Benesty et al., 2007, 2008b, Kap. 47). Wird das Nutzsignal in der Blockiermatrix jedoch nicht vollständig unterdrückt (vgl. Jablon, 1986a, 1987), kommt es zu Fehlanpassungen des aktiven Störgeräuschunterdrückers, die wiederum eine Verzerrung des entstörten Nutzsignals zur Folge haben. Zur Erhöhung der Robustheit wird die Blockiermatrix meist adaptiv ausgeführt (vgl. Nordebo et al., 1994; Hoshuyama et al., 1997, 1998, 1999; Hoshuyama und Sugiyama, 1996, 1999, 2001) und nur in Perioden adaptiert, in denen ein Nutzsignal vorhanden ist (vgl. Van Compernolle, 1990). Demzufolge hat die Steuerung der Adaption einen wesentlichen Einfluss auf die Robustheit der GSC-Implementierung. In Abschnitt 5.2.2 wird ein robuster adaptiver Beamformer mit verbesserter Adaptionskontrolle vorgestellt.

Der GSC ist für breitbandige Signale nur suboptimal (vgl. Monzingo und Miller, 1980; Simmer et al., 2001). In diesem Fall lässt sich mit einem mehrkanaligen Wiener-Filter (MWF) eine optimale Lösung erreichen (vgl. Cornelis et al., 2011). Das MWF kann als MVDR-Beamformer mit einem in Serie geschalteten einkanaligen Wiener-Filter interpretiert werden (vgl. Edelblute et al., 1967; Brooks und Reed, 1972; Monzingo und Miller, 1980; Simmer et al., 2001; Van Trees, 2002, Kap. 6; Herbordt, 2005). Dies motiviert den in dieser Arbeit verfolgten Ansatz, die residualen Störgeräusche am Ausgang eines Mikrofonarrays durch Nachschalten eines einkanaligen Geräuschreduktionsfilters zu reduzieren (vgl. Zelinski, 1988; Fischer und Simmer, 1996; Marro et al., 1998; Bitzer et al., 1999a, 2001; Simmer et al., 2001; McCowan und Bourlard, 2002, 2003; Hendriks et al., 2009). Um die Verzerrung der Transienten bei der Störgeräuschunterdrückung möglichst gering zu halten, wird das modifizierte Ephraim-Malah-Filter (vgl. Kap. 4.3) mit der auditiven Analyse-Synthese-Filterbank (vgl. Kap. 3) als Nachfilter verwendet.

Mit modalen Beamformern lässt sich die Richtwirkung eines Mikrofonarrays über alle Raumrichtungen steuern. Hierzu werden meist sphärische Mikrofonarrays verwendet. Diese tasten das Schallfeld an räumlich diskreten Ortspunkten auf der Kugeloberfläche ab und stellen dieses in Folge als eine Linearkombination von Kugelflächenfunktionen dar. Die Koeffizienten der Darstellung werden auch als Wellenspektrum bezeichnet. Die Kugelflächenfunktionen bilden eine orthonormale

Basis. Durch die raumdiskrete Abtastung kommt es bei Moden höherer harmonischer Ordnung zu räumlichem Aliasing, da diese durch die begrenzte Anzahl an Abtastpunkten nicht fehlerfrei abgebildet werden können (vgl. Li und Duraiswami, 2007; Rafaely et al., 2007; Meyer und Elko, 2008; Zotkin et al., 2008). Die verwendeten Abtastgitter haben dabei einen wesentlichen Einfluss auf die Fehlertoleranz der Schätzung der Koeffizienten des Wellenspektrums. Es gibt eine Reihe unterschiedlicher mathematischer Ansätze, das Quadraturproblem (d. h. eine numerisch bestmögliche Annäherung an das Integral über die Kugeloberfläche, bei einer möglichst geringen Anzahl an diskreten Stützstellen, zu erreichen) zu lösen (vgl. Atkinson und Han, 2012, Kap. 5; Fornberg und Martel, 2014; Reeger und Fornberg, 2015). Kap. 6.4 untersucht die Eigenschaften verschiedener Abtastgitter auf der Kugeloberfläche anhand der Konditionszahl und der Orthonormalität der Kugelflächenfunktionen. Durch die Nullstellen der sphärischen Bessel-Funktionen (als Lösung der Wellengleichung in radialer Richtung) werden sphärische Mikrofonarrays bei einigen Frequenzen instabil (vgl. Abhayapala und Ward, 2002; Gover et al., 2004; Rafaely, 2011). Meyer und Elko (2002) verwenden eine geschlossene, schallharte Kugel, um das Problem zu umgehen. Dieser Ansatz ist jedoch auf relativ kleine Arrayradien beschränkt. Dies führte zur Entwicklung zahlreicher alternativer Arraygeometrien. So können anstatt der ungerichteten Mikrofone, in Richtung der Flächennormale ausgerichtete Nierenmikrofone verwendet werden (vgl. Rahim und Davies, 1982; Meyer, 2001; Balmages und Rafaely, 2007). Aufgrund des starken Eigenrauschens von Nierenmikrofonen bei tiefen Frequenzen und Fehler bei der Ausrichtung und Positionierung der Mikrofone, ist dieser Ansatz in der Praxis nicht robust. Das Problem der schlechten Konditionierung lässt sich über ein aus zwei konzentrischen Kugelschalen mit unterschiedlichen Radien bestehendes Array lösen (vgl. Balmages und Rafaely, 2007; Parthy et al., 2009; Jin et al., 2014). Auf den beiden Kugelschalen werden meist dieselben Abtastgitter einer bestimmten Ordnung verwendet. Der wesentliche Nachteil dieses Ansatzes besteht darin, dass im Vergleich zu einem einfachen Array, die doppelte Anzahl an Mikrofonen benötigt wird.

Bei nicht-sphärischen Arraygeometrien wird meist eine geringere Anzahl an Abtastpunkten benötigt, um die Stabilität in der Umgebung der Nullstellen der Bessel-Funktionen zu verbessern. So lässt sich auch mit einigen wenigen Abtast-

punkten im Inneren der Kugel, die Robustheit über einen großen Wellenzahlbereich erhöhen (vgl. Rafaely, 2008). Die Positionen der inneren Abtastpunkte werden dabei über ein nichtlineares Optimierungsverfahren bestimmt, welches die Konditionszahl minimiert. Allerdings lässt sich mit dieser Methode nicht bestimmen, wie viele innere Abtastpunkte mindestens benötigt werden, um die Stabilität und Konvergenz des Optimierungsverfahrens zu gewährleisten. Abhayapala und Gupta (2009) tasten das Schallfeld mit mehreren kreisförmigen Arrays mit jeweils unterschiedlicher Anzahl an Abtastpunkten ab. Unter Ausnutzung bestimmter Eigenschaften der assoziierten Legendre- und sphärischen Bessel-Funktionen, lässt sich das Array über einen erweiterten Frequenzbereich stabilisieren. Das doppelseitige Konus-Array (vgl. Gupta und Abhayapala, 2010) ist ein weiterer Vorschlag zur Lösung des Nullstellenproblems. Dabei wird die radiale Orthogonalität von auf der Oberfläche eines doppelseitigen Konus ausgewerteten Bessel-Funktionen verwendet, um die Koeffizienten des Wellenspektrums über einen möglichst großen Frequenzbereich stabil zu schätzen. Da diese Schätzung nicht bei allen Frequenzen stabil ist, wird das Schallfeld über zwei oder mehrere Konusse abgetastet. Dies erfordert eine relativ hohe Anzahl an Abtastpunkten und ist für die meisten praktischen Anwendungen nicht geeignet. Alon und Rafaely (2012) tasten das Schallfeld auf der Oberfläche eines Rotationstorus ab. Dies erhöht die Robustheit gegenüber Rauschen, lässt sich in der Praxis jedoch schwer implementieren. In dieser Arbeit (vgl. Kap. 6.5) wurde eine Methode zum Entwurf optimaler Abtastgitter entwickelt, um das Schallfeld innerhalb des Volumens eines Arrays mit möglichst hoher Genauigkeit und über einen möglichst breiten Frequenzbereich zu interpolieren (vgl. Chardon et al., 2014b, 2015). Es wird gezeigt, wie sich die Wahl einer Basis auf die Stabilität der Schätzung der Koeffizienten des Wellenspektrums, welche die Grundvoraussetzung für eine stabile Interpolation ist, auswirkt. Die numerischen Instabilitäten in der Nähe der Nullstellen der sphärischen Bessel-Funktionen stellen ein besonderes Problem dar. Numerische Simulationen unterschiedlicher Abtastgitter zeigen, dass sich die Interpolation mit einigen wenigen Abtastpunkten im Inneren des Volumens über einen breiten Frequenzbereich stabilisieren lässt. Dies kann neben der Kugel auch ganz allgemein für konvexe und sternförmige Gebiete (wie z. B. ellipsoidische und kubische Arrays) gezeigt werden. Die Simulationen zeigen auch die Effizienz des vorgestellten Optimierungsansatzes, der über den ge-

samten Frequenzbereich den kleinsten Interpolationsfehler ergibt. Allerdings sind dabei die Abtastpunkte über das gesamte Volumen verteilt, was sich in der Praxis durch mechanische Einschränkungen meist nur schwer implementieren lässt. Aus diesem Grund wird der Ansatz in Folge dahingehend modifiziert, dass zusätzliche Randbedingungen (wie z. B. Einschränkungen in der Arraygeometrie) berücksichtigt werden können. Numerische Simulationen unterschiedlicher Arraygeometrien zeigen, dass sich mit dem modifizierten Ansatz nicht nur die Robustheit der Interpolation und der Schätzung der Koeffizienten des Wellenspektrums, sondern auch die praktische Realisierbarkeit eines Abtastgitters erhöhen lässt.

Kapitelübersicht

In Kapitel 2 werden die grundlegenden Eigenschaften der menschlichen Hörwahrnehmung im Hinblick auf die digitale Signalverarbeitung in der auditiven Domäne zusammenfassend beschrieben. Die aus den funktionalen Modellen des peripheren Gehörs (Kap. 2.1) abgeleiteten Außen-Mittelohr-Filter (Kap. 2.2) und Gammaton-Filter (Kap. 2.3) finden bei dem im Rahmen dieser Arbeit entwickelten und patentierten Verfahren zur einkanaligen Störgeräuschreduktion mit möglichst geringer Signallatenz (Patentschrift WO 2009/043066 A1, Noisternig et al. 2009) als Analyse-Synthese-Filterbank Anwendung (siehe auch Kap. 3 und 4 und Anhang D). Darüber hinaus wird das von Pflüger (1997) vorgeschlagene parametrische Außen-Mittelohr-Filter an die ISO-Norm 226:2003 (ISO, 2003) angepasst.

Kapitel 3 befasst sich mit der recheneffizienten Implementierung einer auditiven Analyse-Synthese-Filterbank mit möglichst kurzer Systemlatenz. Als Teilfilter werden die in Kap. 2.3.1 entworfenen linearen Gammaton-Filter verwendet. Eine der wichtigsten Eigenschaften einer Analyse-Synthese-Filterbank ist die (nahezu) perfekte Rekonstruktion des Eingangssignals am Ausgang des Systems. Idealerweise wird diese durch eine einfache Summation der Teilbandsignale erreicht. Aus der Betrachtung der Phase an den Übergängen benachbarter Teilbänder der Gammaton-Filterbank wird ein sehr einfaches Kriterium hergeleitet, welches die Notwendigkeit angibt, vor der Summation das Vorzeichen in einem Teilband zu wechseln (vgl. Zotter, 2004; Noisternig et al., 2009). Dadurch lässt sich die Welligkeit am Ausgang des Systems auch ohne rechenaufwändige Synthesefilter stark

reduzieren. Zudem kann die Filterbank anstatt aus komplexwertigen mit reellwertigen Gammaton-Filtern aufgebaut werden, was auf eine sehr recheneffiziente Implementierung der auditiven Analyse-Synthese-Filterbank führt.

Kapitel 4 diskutiert die Anwendung der spektralen Subtraktion auf die breitbandige Reduktion von Störgeräuschen in einkanaligen Mikrofonanwendungen. Dabei liegt das Hauptaugenmerk auf den Gewichtungsregeln von Ephraim und Malah, mit denen die Amplitude des ungestörten Signals aus dem Kurzzeitspektrum des gestörten Gesamtsignals geschätzt wird. Dabei kommen verschiedene Kurzzeit-Amplitudenschätzer und ein entscheidungsgesteuerter Ansatz (DDA)[10] zum Schätzen des *a-priori*-SNR zum Einsatz. Durch Darstellung der spektralen Gewichte auf einer Kennfläche über der vom *a-priori* und *a-posteriori*-SNR aufgespannten Ebene, lassen sich die Eigenschaften der unterschiedlichen Amplitudenschätzer sowie des entscheidungsgesteuerten Ansatzes sehr anschaulich darstellen (s. Kap. 4.3). Dies ermöglicht neue Einsichten in die grundlegende Funktionsweise des Ephraim-Malah-Filters zur einkanaligen Signalaufbereitung. Bei den aus der Literatur bekannten Ansätzen der spektralen Subtraktion muss stets ein Kompromiss zwischen der Unterdrückung des Musical Noise und den transienten Verzerrungen des Nutzsignals getroffen werden. Der in dieser Arbeit vorgeschlagene modifizierte DDA (s. Kap. 4.4) erlaubt eine wesentlich flexiblere Handhabung der Unterdrückung des Musical Noise, ein schnelles Ansprechverhalten zur Reduktion der Verzerrungen transienter Klänge sowie eine Verringerung des „Konstant-ξ-Effekts“ (s. a. Zotter, 2004, Noisternig et al., 2009 und Anhang D). Zur Spektraltransformation wird die in Kap. 3 vorgestellte auditive Analyse-Synthese-Filterbank verwendet. Die Ergebnisse eines informellen Hörtests zeigen eine tendenziell bessere Beurteilung des Höreindrucks bei Verwendung des modifizierten DDA gegenüber dem DDA von Ephraim und Malah.

Kapitel 5 befasst sich mit der Signalaufbereitung in mehrkanaligen Mikrofonanwendungen. Dazu wird in Kap. 5.1 zuerst die mathematische Beschreibung von Schallfeldern kurz zusammengefasst und das dem Mikrofonarray-Beamforming zugrunde liegende Signalmodell diskutiert. Zudem werden etablierte und weit ver-

[10] *Decision directed approach* (DDA).

breitete Maße zur Beurteilung der Leistungsfähigkeit von Mikrofonarrays vorgestellt, die dann später beim Entwurf modaler Beamformer aufgegriffen werden. In Kap. 5.2 wird ein robuster adaptiver Beamformer vorgestellt, der auf dem Prinzip des Generalised Sidelobe Cancellers beruht. Das Hauptaugenmerk liegt auf der Optimierung der Adaptionskontrolle, die einen großen Einfluss auf das Konvergenzverhalten der adaptiven Prozesse hat. Dazu werden die richtungsabhängigen Schwankungen der Schätzung des Signal-Interferenz-Verhältnisses minimiert und ein räumliches Kriterium eingeführt, welches die Signalleistungen aus unterschiedlichen Raumrichtungen ins Verhältnis setzt. Am Beispiel eines Mikrofonarrays kleiner Bauform mit einer geringen Anzahl an Sensoren wird gezeigt, dass mit der verbesserten Adaptionskontrolle eine hohe und breitbandige richtungsabhängige Verstärkung möglich ist. Aufgrund der adaptiven Ausführung ist der Beamformer zudem robust gegenüber Bauteiltoleranzen, Fehlpositionierungen und reflexionsbehaftete Schallwellenausbreitung.

In Kapitel 6 werden zuerst die Lösungen der Wellengleichung in Kugelkoordinaten (s. Kap. 6.1) diskutiert. Dem folgt der Entwurf (s. Kap. 6.2) und die Simulation (s. Kap. 6.3) unterschiedlicher modaler Beamformer. Dabei wird angenommen, dass das Schallfeld auf der Kugel strikt bandbegrenzt ist, sodass kein Aliasing entsteht. Die verwendeten Abtastgitter auf der Kugel haben einen wesentlichen Einfluss auf die Fehlertoleranz der Schätzung der Koeffizienten des Wellenspektrums. Kap. 6.4 zeigt das Verhalten verschiedener Abtastgitter anhand der Konditionszahl und der Orthonormalität der Kugelflächenfunktionen. Die Simulationen zeigen, dass sphärische Mikrofonarrays bei einigen Frequenzen instabil werden. Dies erklärt sich durch die Nullstellen der sphärischen Bessel-Funktionen (als radiale Lösungen der Wellengleichung), die bei der Inversion zu numerischen Instabilitäten führen. In Kap. 6.5 werden zwei Methoden zum Entwurf robuster modaler Mikrofonarrays vorgestellt. Als Maß für die Robustheit wird dabei der relative Fehler der Interpolation des Schallfeldes innerhalb des von dem jeweils betrachteten Array umschlossenen Volumen verwendet. Diese bestimmen einige wenige optimale Abtastpunkte im Inneren des Arrays, mit denen sich die Interpolation des Schallfeldes nahe der Nullstellen der Bessel-Funktionen stabilisieren lässt. Um die Implementierung zu erleichtern, wird der Ansatz zur Berechnung op-

timaler Abtastpunkte dahingehend erweitert, dass Rahmenbedingungen (wie z. B. mechanische Beschränkungen) mit berücksichtigt werden können. Mehrere Beispiele unterschiedlicher Arraygeometrien zeigen, dass mit der vorgestellten Methode der kleinste Interpolationsfehler über den größten Wellenzahlbereich erzielt werden kann.

Kapitel 7 fasst die gewonnen Erkenntnisse zusammen und gibt einen kurzen Ausblick auf aktuelle und zukünftige Arbeiten zu diesem Thema.

2

Grundlagen der menschlichen Hörwahrnehmung im Hinblick auf den Entwurf auditiver Filter

In diesem Kapitel werden grundlegende Eigenschaften der menschlichen Hörwahrnehmung im Hinblick auf die digitale Signalverarbeitung in der auditiven Domäne zusammenfassend beschrieben. Die aus den funktionalen Modellen des peripheren Gehörs (s. Kap. 2.1) abgeleiteten auditiven Filter (s. Kap. 2.3) finden bei dem im Rahmen dieser Arbeit entwickelten und patentierten Verfahren zur einkanaligen Störgeräuschreduktion mit möglichst geringer Signallatenz (Patentschrift WO 2009/043066 A1, Noisternig et al. 2009) als Analyse-Synthese-Filterbank Anwendung (s. a. Kapitel 3 und 4, sowie Anhang D).

Das Verständnis des komplexen Informationsverarbeitungssystems Gehör kann nach Hawkins (1995, Kap. 1) nur über eine multidisziplinäre Betrachtungsweise erreicht werden, die üblicherweise im übergeordneten Begriff der Psychoakustik zusammengefasst wird (vgl. dazu auch Pflüger, 1997, Kap. 2). Die Psychoakustik untersucht und beschreibt demnach den Zusammenhang zwischen quantitativ erfassbaren (d. h. physikalisch messbaren) Schallereignissen und dem subjektiven Höreindruck.

Die Modellbildung aus signaltheoretischer Sicht besteht in der möglichst exakten und vor allem auch recheneffizienten Nachbildung der psychologischen Reak-

tionen auf die akustischen Reize. Diese werden typischerweise in psychoakustischen Studien ermittelt. Es besteht demnach eine enge Bindung der Modellbildung an die experimentell ermittelten Daten, deren Varianz den Gültigkeitsbereich der Modelle begrenzt (vgl. Pflüger, 1997, Kap. 2.8). Die aus psychoakustischen Studien abgeleiteten Theorien bilden meist nur Teilaspekte der Hörwahrnehmung ab und stehen teils sogar im Widerspruch zu den für andere Teilaspekte experimentell ermittelten Daten. Nach Hudde (2005, S. 50) kann die menschliche Hörwahrnehmung durch keines der in der gegenwärtigen Literatur vorhandenen Modelle in ihrer Gesamtheit abgebildet werden.

Funktionale Modelle dienen der Nachbildung einzelner, den jeweiligen technischen Anwendungen angepassten Eigenschaften des peripheren Gehörs (vgl. dazu Hudde, 2005; Meddis und Lopez-Poveda, 2010). Die in psychoakustischen Versuchen ermittelten Daten werden dabei ohne Berücksichtigung der diesen Daten zugrunde liegenden physikalischen bzw. physiologischen Vorgänge modelliert (vgl. Pflüger, 1997, Kap. 4). Das Leistungsdichtespektrum (siehe z. B. Moore, 1995, Kap. 5) ist ein typisches Beispiel für ein funktionales Modell. Dieses aus dem Verfahren zur Ermittlung der Frequenzgruppenbreite (s. Kap. 2.2.4) abgeleitete Modell bildet die Grundlage funktional motivierter auditiver Filterbänke. Diese können zum Beispiel mit Hilfe von Gammaton-Filtern (s. Kap. 2.3) recheneffizient implementiert werden. Die für diese Arbeit entworfene Analyse-Synthese-Filterbank zur auditiven Signalverarbeitung (siehe Zotter, 2004; Noisternig et al., 2009) wird in Kapitel 3 ausführlich diskutiert.

2.1 Das periphere Gehör

Das periphere Gehör setzt sich aus dem Außen-, Mittel- und Innenohr zusammen (s. Abb. 2.1), die sowohl anatomisch wie auch funktionell eindeutig voneinander unterscheidbar sind (s. a. Møller, 2006, Kap. 1–4; Fastl und Zwicker, 2007, Kap. 3; Meddis und Lopez-Poveda, 2010; Kramme et al., 2011, Kap. 12; Pickles, 2012, Kap. 2–3; Moore, 2013, Kap. 1.6). Die Hauptaufgabe des peripheren Gehörs besteht in der Weiterleitung des am Ohr eintreffenden Schalls an das Innenohr, wo die Umwandlung in neuronale Signale stattfindet. Diese werden in weiterer Folge über den Hörnerv an die auditiven Verarbeitungszentren des Gehirns übertragen.

Zudem wird im peripheren Gehör die akustische Impedanz der Wellenausbreitung in Luft an die Impedanz der Wellenausbreitung im Innenohr[11] angepasst und die akustische Information durch Vorfilterung für die nachfolgende neuronale Verarbeitung aufbereitet.

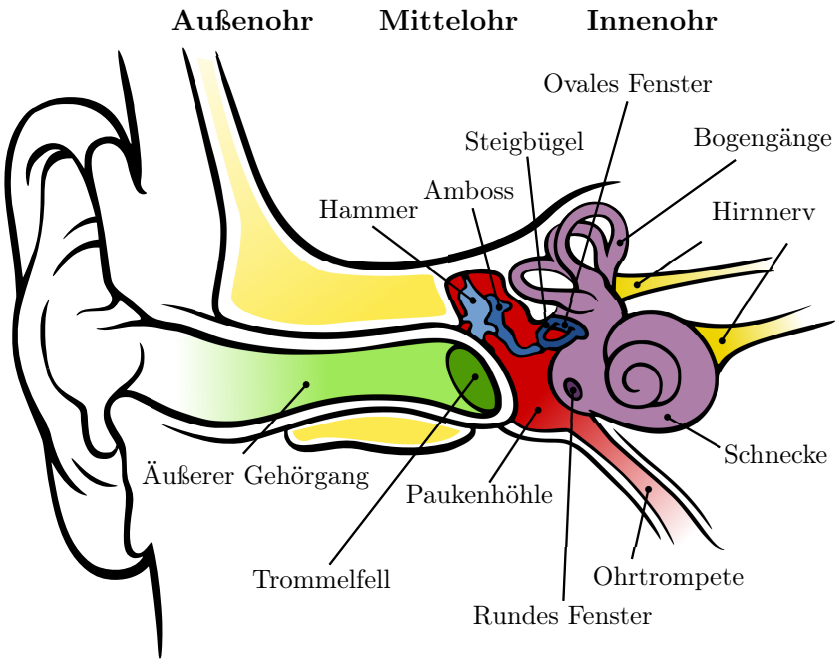


Abbildung 2.1: Aufbau des peripheren Gehörs[12].

Die Abschnitte 2.1.1 bis 2.1.3 geben einen Überblick über den anatomischen Aufbau des peripheren Gehörs und fassen die für den Entwurf einer auditiven Analyse-Synthese-Filterbank relevanten physiologischen Verarbeitungsschritte zusammen. Die daraus abgeleiteten perzeptiven Modelle werden zur praktischen Implementierung in die signaltheoretische Domäne übergeführt (s. Kap. 2.2). Dabei wird in dieser Arbeit vor allem auf eine recheneffiziente Umsetzung mit möglichst kurzer Systemlatenz geachtet, wobei die auditive Analyse-Synthese-Filterbank darüber hinaus eine nahezu perfekte Rekonstruktion des Originalsignals ermöglichen soll.

[11] Die Impedanz der Schallausbreitung in der Lymphflüssigkeit der Cochlea ist in etwa 4000-mal höher als die der Schallausbreitung in Luft (vgl. Meddis und Lopez-Poveda, 2010, S. 13).

[12] Bildnachweis: Chittka L. Brockmann (Creative Commons Licence, Version 2.5).

2.1.1 Das Außenohr

Das Außenohr umfasst die Ohrmuschel (Pinna) und den äußeren Gehörgang (siehe z. B. Møller, 2006, Kap. 1.3; Pickles, 2012, Kap. 2.1). Die einfallende Schallenergie wird über die Ohrmuschel, welche akustisch einem Trichter entspricht, aufgenommen und über die Luft im Gehörgang an das Trommelfell weitergeleitet. Der Körperschall (d. h. der Schalltransport mittels Knochenleitung) erzeugt im Innenohr einen um mehr als 40 dB geringeren Schallpegel als der durch die Luftausbreitung transportierte Schall (vgl. dazu Arnold und Burkard, 2000; Hudde, 2005; Klump, 2005). Bei Normalhörenden ist der akustische Reiz aufgrund von Knochenleitung somit perzeptiv irrelevant und kann in der Modellbildung vernachlässigt werden.

Ohrmuschel und Gehörgang erzeugen gemeinsam starke Resonanzen und damit verbundene Klangfärbungen (siehe z. B. Shaw, 1974a). Diese Resonanzen können in einer ersten Näherung durch das akustische Verhalten eines einseitig geschlossenen Rohrs, d. h. eines $\lambda/4$-Resonators, beschrieben werden (s. a. Kap. 2.2.1-A). Die Modellierung der Übertragungsfunktion des Außen- und Mittelohres wird in Kap. 2.2.1-B ausführlich beschrieben. Die daraus abgeleiteten Filter finden in weiterer Folge als Vorverarbeitungsstufe der auditiven Filterbank Anwendung.

Durch Beugung, Absorption und Reflexion einer einfallenden Schallwelle an Torso, Kopf und Ohrmuschel wird diese richtungsabhängig gefiltert (siehe z. B. Blauert, 1996, S. 63 ff). Die dabei zwischen dem linken und rechten Ohrsignal auftretenden Laufzeitunterschiede (*interaural time difference, ITD*) und Pegelunterschiede (*interaural level difference, ILD*) sind für die Lokalisation in der Transversalebene[13] von wesentlicher Bedeutung (vgl. Lord Rayleigh, 1907; Wightman und Kistler, 1989a,b; Abbagnaro et al., 1975; Middlebrooks und Green, 1991; Macpherson und Middlebrooks, 2002). Durch Reflexionen an der Pinna treten bei Frequenzen größer 5 kHz starke Kammfiltereffekte auf. Diese spielen vor allem bei der Lokalisation in der Medianebene[14] eine wesentliche Rolle (s. a. Blauert, 1996, S. 63 ff; Moore 1995, S. 321 ff; Hebrank und Wright, 1974; Mehrgardt

[13] Die Transversalebene ist eine auf den Körper bezogene Ebene, die diesen in einen oberen und einen unteren Abschnitt unterteilt. Es gibt unendlich viele Transversalebenen durch den Körper, wobei in der Akustik meist die Ebene die durch die Ohrachse verläuft als Bezugsystem verwendet wird.

[14] Als Medianebene wird die durch die Körpermitte verlaufenden Vertikalebene (auch Sagittalebene) bezeichnet.

und Mellert, 1977; Middlebrooks et al., 1989; Langendijk und Bronkhorst, 2002). Zudem findet im Fall fehlender Pinna-Filterung keine Außer-Kopf-Lokalisierung der wahrgenommenen Schallquelle statt (Wightman und Kistler, 1989a,b).

Bei der in Kapitel 3 entworfenen Filterbank steht vor allem die Frequenzgruppeneigenschaft der menschlichen Hörwahrnehmung (s. Kap. 2.2.4) im Vordergrund. Die räumliche Trennung der unterschiedlichen Schallquellen wird durch ein der einkanaligen Störgeräuschreduktion vorgeschaltetes Array räumlich verteilter Sensoren erreicht (s. Kapitel 5 und 6). Aus diesem Grund wird in dieser Arbeit bei der Modellbildung auditiver Filter nicht näher auf die Eigenschaften der Lokalisationswahrnehmung eingegangen. Eine ausführliche Zusammenfassung der unterschiedlichen Aspekte des Richtungshörens ist zum Beispiel in Blauert (1996), Hawkins (1995), Moore (1995), Terhardt (1998), Zwicker und Fastl (1999), Suzuki et al. (2011) und Xie (2013) zu finden.

2.1.2 Das Mittelohr

Das Mittelohr liegt in der luftgefüllten Paukenhöhle, die über die Eustachische Röhre mit dem Nasen-Rachenraum verbunden ist (vgl. Hudde, 2005; Møller, 2006, Kap. 1.4; Pickles, 2012, Kap. 2.2). Die vom Trommelfell aufgenommenen Luftschwingungen werden über die Gehörknöchelchen Hammer, Amboss und Steigbügel auf das ovale Fenster des Innenohres übertragen. Die Hauptfunktion des Mittelohres liegt in der Impedanzanpassung von der Schallwellenausbreitung in Luft zur Schallwellenausbreitung in Flüssigkeit. Dabei stellen die Hebelübersetzung der Gehörknöchelchen, sowie die Flächentransformation von Trommelfell zu ovalem Fenster die primären Einflussfaktoren dar. Als Übertragungsfunktion stellt sich Bandpasscharakteristik ein (vgl. Hudde und Engel, 1998a,b,c). Durch den Stapediusreflex[15] wird bei lauten Schallereignissen der Arbeitspunkt der Übertragungsstrecke aktiv verändert, um das Innenohr vor Schädigung zu schützen, wobei die Reaktionszeit im Mittel in etwa 100 ms beträgt (Metz, 1951; Møller, 1961). Dieser Schutzmechanismus stellt eine wesentliche Nichtlinearität des peripheren Gehörs dar.

[15] Der Stapediusreflex wird vom Hirnstamm ausgelöst und wirkt auf die an den Gehörknöchelchen sitzenden Muskeln *musculus stapedius* und *musculus tensor tympani*.

In Sprachkommunikationsanlagen werden Signale mit relativ geringen Schallpegeln am Hörerort wiedergegeben. Diese liegen im Allgemeinen weit unter den Grenzwerten zur Vermeidung lärmbedingter Hörschäden und somit auch unter dem Schwellwert des Eintretens des Stapediusreflexes. Aus diesem Grund wird diese Nichtlinearität des Mittelohres im folgenden Applikationsentwurf nicht nachgebildet. Das signaltheoretische Modell kann nach Pascal et al. (1998) jedoch dahingehend erweitert werden.

2.1.3 Das Innenohr

Das Innenohr liegt im Felsenbein[16] und bildet mit dem Gleichgewichtsorgan eine anatomische Einheit (vgl. Møller, 2006, Kap. 1.5; Pickles, 2012, Kap. 3; Duifhuis, 2012). Es besteht aus einem schneckenförmigen Gehäuse (*cochlea*), das in Längsrichtung in drei mit zwei unterschiedlichen Lymphflüssigkeiten befüllte Kammern aufgeteilt ist. Das eine Abdichtung gegenüber der Steigbügelfußplatte bildende ovale Fenster mündet in die Vorhoftreppe (*scala vestibuli*). Die Paukentreppe (*scala tympani*) ist zum Mittelohr hin über das runde Fenster abgeschlossen, welches dem Druckausgleich dient. Zwischen den parallel laufenden Treppen liegt als dritte Kammer der Schneckengang (*scala media* bzw. *ductus cochlearis*). Die Vorhof- und Paukentreppe sind mit Perilymphe, der Schneckengang hingegen mit Endolymphe gefüllt. Durch die unterschiedliche chemische Zusammensetzung beider Flüssigkeiten entsteht eine Spannungsdifferenz, die eine Umwandlung mechanischer Erregung in elektrische Impulse ermöglicht (Nin et al., 2008).

Auf den die Flüssigkeiten trennenden Membranen, Reissnersche Membran und Basilarmembran, bildet sich bei Anregung durch das Mittelohr eine Wanderwelle aus. Die Wanderwellentheorie wurde durch Békésy begründet (Békésy, 1928; Békésy und Wever, 1960; Wever, 1962; Békésy, 1964) und lässt sich nach Zwislocki (1946, 1948) durch den ortsvariablen Wellenwiderstand als Folge der unterschiedlichen Steifigkeit, Nachgiebigkeit und Dämpfung der Basilarmembran erklären (vgl. dazu auch Terhardt, 1998, S. 232–239; Zwicker und Fastl, 1999, S. 28–31). Die Hüllkurve der Wanderwelle steigt bei einer sinusförmigen Erregung

[16] Das Felsenbein (*pars petrosa, petrosum)* ist der härteste Knochen des Menschenschädels und ein Abschnitt des Schläfenbeins.

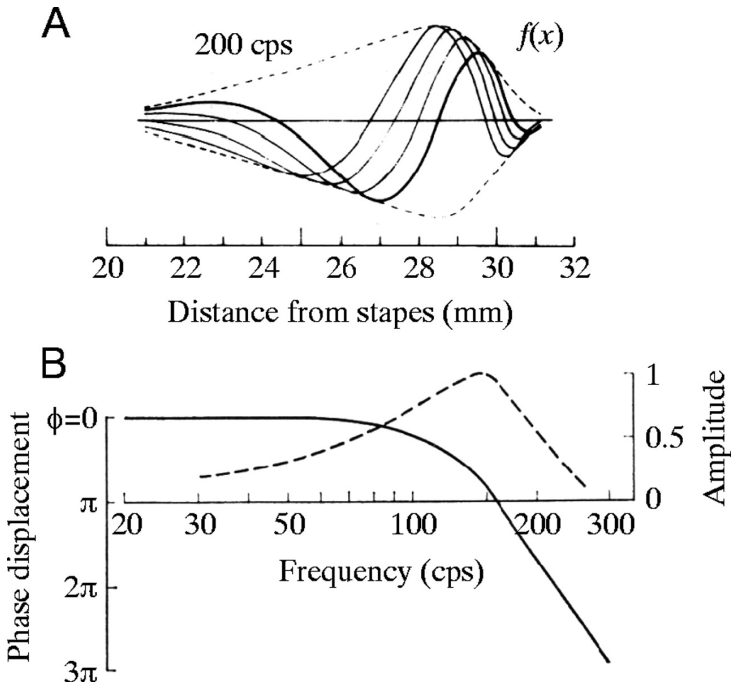


Abbildung 2.2: (A) Wanderwelle entlang der Basilarmembran bei einer Frequenz von 200 Hz. Die Einhüllende der maximalen Auslenkung während einer Schwingungsperiode ist als gestrichelte Linie dargestellt. Die durchgezogenen Linien zeigen vier Wellen mit unterschiedlichen Phasenwinkeln. (B) Phase der Auslenkung auf der Basilarmembran (durchgezogene Linie) und zughöriger Amplitudenverlauf (gestrichelte Linie). Illustration aus Békésy und Wever (1960).

langsam an und verebbt nach Erreichen des Maximums verhältnismäßig schnell (vgl. Abb. 2.2). Je höher die Frequenz, desto kürzer ist die Wanderwelle. Aufgrund dieser Frequenz-Orts-Transformation ergibt sich eine Spektralanalyse (vgl. dazu Fletcher, 1940; Greenwood, 1961, 1990). Das Helikotrema (*apex*) am oberen Ende der Cochlea stellt eine Verbindung zwischen Vorhoftreppe und Paukentreppe her. Über diese Öffnung wird bei sehr tiefen Frequenzen bzw. bei Gleichdruckänderungen eine Verschiebung der Flüssigkeitssäulen erreicht, sodass keine Sinnesreizung stattfindet. Auf der Basilarmembran befinden sich die inneren Haarzellen, die bei einer Auslenkung der Membran an der jeweiligen Stelle zum Feuern elektrischer Impulse angeregt werden. Dieser sensorische Transduktionsprozess wandelt die Reize in den Sinneszellen in Information um,

die anschließend vom Nervensystem verarbeitet werden kann (Møller, 2006, S. 48–56).

Wie mehrere Studien zeigen (vgl. Rhode, 1971; Johnstone et al., 1986; Dallos, 1992; Ruggero et al., 1997), kann mittels der äußeren Haarzellen das Schwingungsverhalten der Basilarmembran aktiv beeinflusst werden. Dabei werden diese über die efferenten (d. h. die vom Hirnstamm kommenden) Fasern des Hörnervs gesteuert und wirken als mechanische Verstärker. Diese aktive Verstärkung führt zu einer Steigerung der Empfindlichkeit und zu einer Erhöhung der Trennschärfe und sorgt demnach für eine feinere Frequenzabstimmung. Dieser Vorgang stellt die zweite wesentliche Nichtlinearität des peripheren Gehörs dar. Da die passiven Eigenschaften die Funktionalität des Innenohres dominieren wird, zur Vereinfachung der Modellbildung, diese Nichtlinearität hier nicht weiter berücksichtigt.

2.2 Modelle des peripheren Gehörs

Den meisten auditiven Modellen ist die strukturelle Anlehnung an den anatomischen Aufbau des peripheren Gehörs gemein (Hawkins, 1995, S. 1–14). Aufgrund der Komplexität der menschlichen Hörwahrnehmung werden bei der Modellbildung meist nur einzelne, an die technische Zielsetzung angepasste Eigenschaften des Gehörs nachgebildet. Außen- und Mittelohr werden dabei meist gemeinsam betrachtet, die Spektralanalyse wird der Basilarmembran zugeordnet. Der Transduktionsprozess sowie die Schallkodierung in den Fasern des Hörnervs werden bei den im Rahmen dieser Arbeit angewendeten Methoden nicht in die Modellbildung mit einbezogen.

In der auditiven Signalverarbeitung erfolgt die Einteilung meist in funktional motivierte und physiologisch motivierte Modelle (vgl. Pflüger 1997, Kap. 4; Hudde 2005). Bei den physiologischen Modellen steht die möglichst exakte Nachbildung der physiologischen und physikalischen Eigenschaften des Gehörs im Vordergrund. Im Gegensatz dazu bilden funktional motivierte Modelle lediglich die Eingangs-Ausgangs-Beziehungen der jeweiligen Verarbeitungsschritte nach. Zwischen diesen beiden Unterteilungsformen herrscht demnach eine gewisse Unschärfe.

2.2.1 Modellbildung eines Außen-Mittelohr-Filters

Obwohl das Mittelohr aktive Komponenten besitzt, wird dessen Funktionalität von den passiven Eigenschaften dominiert. Aus diesem Grund wird das Mittelohr, ebenso wie das Außenohr, zumeist als passives lineares Modell nachgebildet (Hawkins, 1995, S. 15–61). Die funktionale Nachbildung wird mit einem breitbandigen Bandpassfilter durchgeführt. Die Übertragungsfunktion (bzw. der Amplitudengang) dieses Filters kann nach folgenden Methoden abgeleitet werden.

A Physiologisch motivierte Modelle

Die der auditiven Signalverarbeitung zugrunde liegenden physiologischen Prozesse können über elektromechanische Analogien abgebildet beziehungsweise direkt über die akustische Zweitortheorie modelliert werden.

Modellbildung mittels elektromechanischer Analogien. Ausgehend von einer Kaskade unabhängiger Prozesse werden – in einem ersten Abstraktionsschritt – die physiologischen Modelle als mechanische Vorgänge formuliert. Dabei müssen Vereinfachungen in Kauf genommen werden und die Allgemeingültigkeit dieser Modelle geht dadurch weitgehend verloren (Hawkins, 1995, S. 15–61). Es ist oft vorteilhaft die meist komplexen mechanischen Schwingungsstrukturen in einem mathematisch einfacher zu beschreibenden elektrischen Schaltkreis zu analysieren. Unter Verwendung elektromechanischer Analogien werden – in einem zweiten Abstraktionsschritt – die mechanischen Modelle in elektrische Netzwerke übergeführt (Zollner und Zwicker, 1993, S. 123–148). Dabei muss einschränkend vorausgesetzt werden, dass die geometrische Ausdehnung der Elemente klein gegenüber der Wellenlänge des Schalls sein sollte. Nach Lerch et al. (2009, Kap. 11) sollte eine maximale Ausdehnung von $\lambda/10$ nicht überschritten werden. Die elektrischen Netzwerke lassen sich zur praktischen Implementierung – in einem dritten Abstraktionsschritt – meist relativ einfach in die digitale Domäne überführen.

Modellbildung mittels akustischer Zweitorsysteme. Bei dem von Terhardt (1998, S. 229–231) vorgeschlagenen Verfahren werden die physiologisch motivier-

ten Modelle bzw. die durch diese Modelle beschriebene Kaskade unabhängiger Prozesse, anhand der akustischen Zweitortheorie (vgl. Terhardt, 1998, S. 75–89) dargestellt. Dabei wird der Einfluss von Außen- und Mittelohr, wie folgend dargestellt, durch entsprechende Übertragungsfunktionen beschrieben.

Der **äußere Gehörgang** kann, unter der Annahme ebener Wellensausbreitung[17] (vgl. dazu Shaw und Teranishi, 1968; Stinson et al., 1982; Stinson, 1985; Rabbitt und Holmes, 1988), näherungsweise als einseitig abgeschlossenes Rohr betrachtet werden. Daraus ergibt sich für den Hörfrequenzbereich ein akustisches Zweitor

$$\begin{bmatrix} P_{\mathrm{E}}(s) \\ Q_{\mathrm{E}}(s) \end{bmatrix} = \mathbf{A} \begin{bmatrix} P_{\mathrm{T}}(s) \\ Q_{\mathrm{T}}(s) \end{bmatrix}, \tag{2.1}$$

mit den Frequenzfunktionen des Schalldrucks $P_{\mathrm{E}}(s)$, $P_{\mathrm{T}}(s)$ und der Schallschnelle $Q_{\mathrm{E}}(s)$, $Q_{\mathrm{T}}(s)$ am Eingang des Ohrkanals bzw. am Trommelfell, wobei s die komplexe Frequenzvariable der Laplace-Transformation bezeichnet. Die Ausbreitung einer fortschreitenden ebenen Welle in Rohren wird durch die akustische Flussimpedanz[18]

$$Z_{\mathrm{A}} = \frac{p}{q} = \frac{Z_0}{S} = \frac{\rho_0 c}{S} \tag{2.2}$$

beschrieben. Diese ergibt sich aus dem Quotienten von Schalldruck p und Schallfluss q und hängt über die Querschnittsfläche S mit der Feldimpedanz $Z_0 = \rho_0 c$ zusammen, wobei ρ_0 die Dichte und c die Schallausbreitungsgeschwindigkeit in Luft beschreiben. Die Kettenmatrix $\mathbf{A}$ berechnet sich nach Terhardt (vgl. dazu auch Møller, 1961; Hudde, 1983; Hammershøi und Møller, 1996; Hudde et al., 1999; Hiipakka et al., 2012) zu

$$\mathbf{A} = \begin{bmatrix} \cosh(\gamma l) & Z_{\mathrm{A}} \sinh(\gamma l) \\ \frac{1}{Z_{\mathrm{A}}} \sinh(\gamma l) & \cosh(\gamma l) \end{bmatrix}, \tag{2.3}$$

[17] Nach Farmer-Fedor und Rabbitt (2002) kann bei numerischen Simulationen, unter Berücksichtigung nichtebener Wellenausbreitung, eine genauere Annäherung an die im Ohrkanal gemessene Schalldruckverteilung erreicht werden. Zur Vereinfachung der Modellbildung wird der Einfluss nichtebener Wellenausbreitung hier nicht weiter berücksichtigt.

[18] Diese wird in der Literatur oft auch vereinfachend als "akustische Impedanz" bezeichnet.

mit einer angenommenen frequenzunabhängigen Dämpfungskonstante δ und mit $\gamma = \delta + \frac{s}{c}$. Zur einfachen Modellbildung kann die Länge l des Gehörgangs mit $l \approx 25\,\mathrm{mm}$ und dessen mittlere Querschnittsfläche mit $S \approx 0{,}5 \ldots 0{,}7\,\mathrm{cm}^2$ angenommen werden.

Die Übertragungsfunktion $G(s)$ wird durch das Verhältnis der Frequenzfunktionen des Schalldrucks am Trommelfell $P_\mathrm{T}(s)$ und des Schalldrucks am Eingang des Gehörgangs $P_\mathrm{E}(s)$ bestimmt. Dabei wird $G(s)$, wie aus den Gleichungen (2.1) und (2.3) leicht ersichtlich ist, hauptsächlich von der Trommelfellimpedanz $Z_\mathrm{T}(s) = P_\mathrm{T}(s)/Q_\mathrm{T}(s)$ bestimmt. Die Übertragungsfunktion ergibt sich zu

$$G(s) = \frac{P_\mathrm{T}(s)}{P_\mathrm{E}(s)} = \frac{Z_\mathrm{T}(s)}{Z_\mathrm{T}(s)\cosh(\gamma l) + Z_\mathrm{A}\sinh(\gamma l)}\,. \tag{2.4}$$

Nach Terhardt (1998, S. 229) kann für Frequenzen kleiner 7 kHz eine unendlich große Trommelfellimpedanz Z_T als Abschluss des Gehörgangs angenommen werden. Somit vereinfacht sich Gleichung (2.4) zu

$$G(s) \approx \frac{1}{\cosh(\gamma l)} = \frac{1}{\cosh\left((\delta + \frac{s}{c})l\right)}\,. \tag{2.5}$$

Der Betrag des Frequenzgangs (d. h. bei Betrachtung von $s = j2\pi f$) nimmt somit bei folgenden Frequenzen Maximalwerte an

$$f_n = (2n-1)\frac{c}{4l} \qquad \text{mit } n = 1, 2, 3, \ldots\,. \tag{2.6}$$

Nach Gleichung (2.6) stellt der $\lambda/4$-Resonator somit eine gute Näherung zur Berechnung der Resonanzfrequenzen des äußeren Gehörgangs dar.

Um die Übertragungsfunktion des **Mittelohres** zu modellieren, wird die Frequenzfunktion der Schnelle der Steigbügelfußplatte zur Frequenzfunktion des Schalldrucks am Trommelfell ins Verhältnis gesetzt. Der Literatur folgend (vgl. dazu Møller, 1961; Onchi, 1961; Zwislocki, 1962; Mehrgardt und Mellert, 1977; Kringlebotn und Gundersen, 1985) weist die Übertragunsfunktion

Tiefpassverhalten auf. Dieses kann näherungsweise durch die Gleichung

$$M(s) = C \cdot \frac{s}{(s+\omega_g)\left[(s+\omega_g)^2+\omega_g^2\right]} \tag{2.7}$$

beschrieben werden, wobei die Grenzfrequenz mit $f_g \approx 1500\,\mathrm{Hz}$ angenommen wird. Ein umfassender Modellansatz der Schwingungsübertragung im Mittelohr wird in Hudde und Weistenhofer (1997) beschrieben, wobei hier auch Nichtlinearitäten mit berücksichtigt werden.

B Funktional motivierte Modelle

Zur funktional motivierten Modellbildung der Übertragungsfunktion des Außen- und Mittelohres erfolgt eine qualitative Abschätzung des Amplitudenverlaufs aus den Kurven gleicher Lautheit (*equal loudness contours, ELC*), welche auch als Isophone bezeichnet werden. Diese stellen die psychoakustisch ermittelte Empfindlichkeit des Ohres bei Anregung mit harmonischen Schwingungen unterschiedlicher Frequenz dar. Der erste vollständige Satz Kurven gleicher Lautheit – bei Anregung im Freifeld und Schalleinfall von vorne – wurde von Fletcher und Munson (Fletcher und Munson, 1933; Fletcher, 1940) bestimmt. Die im ISO 226:1987 Standard (ISO, 1987) im Jahr 1987 festgelegten Isophone beruhen auf den von Robinson und Dadson (1956) ermittelten Kurven. Die Ergebnisse neuerer Studien (Brinkmann et al., 1994; Suzuki und Takeshima, 2004) zeigen jedoch deutliche Abweichungen von den Robinson-Dadson Kurven. Vor allem für Frequenzen kleiner 800 Hz sind diese Kurven wesentlich höher als die in ISO 226:1987 publizierten Daten. Aus diesem Grund wurde der ISO 226:1987 Standard im Jahr 2003 dahingehend revidiert und unter der Bezeichnung ISO 226:2003 (ISO, 2003) veröffentlicht. Bei der im folgenden Abschnitt diskutierten digitalen Nachbildung der Filterfunktion des Außen- und Mittelohres werden ausschließlich die Bezugshörschwellen der revidierten ISO-Norm angewendet. Diese sind in Abb. 2.3 dargestellt.

Die ISO-Kurven gleicher Lautheit weisen für Frequenzen über 1 kHz bei unterschiedlicher Lautheit einen sehr ähnlichen Amplitudenverlauf auf, der in etwa der Form der Ruhehörschwelle entspricht. Im Bereich unter 1 kHz werden die Kurvenverläufe mit steigender Lautstärke jedoch deutlich flacher. Daraus resultiert die Annahme, dass das interne Rauschen der Cochlea bei tiefen Frequenzen stärker

ausgeprägt ist als bei hohen Frequenzen (vgl. dazu Pflüger, 1997, Kap. 6). Dies wird dadurch begründet, dass der Pegel des internen Rauschens relativ gering ist und sich daher vorwiegend auf die Ruhehörschwelle auswirkt. Diese ist zu tiefen Frequenzen hin deutlich angehoben. Die Kurven gleicher Lautheit für hohe Pegel werden durch das Cochlea-Rauschen hingegen nicht beeinflusst (vgl. auch Nedzelnitsky, 1980).

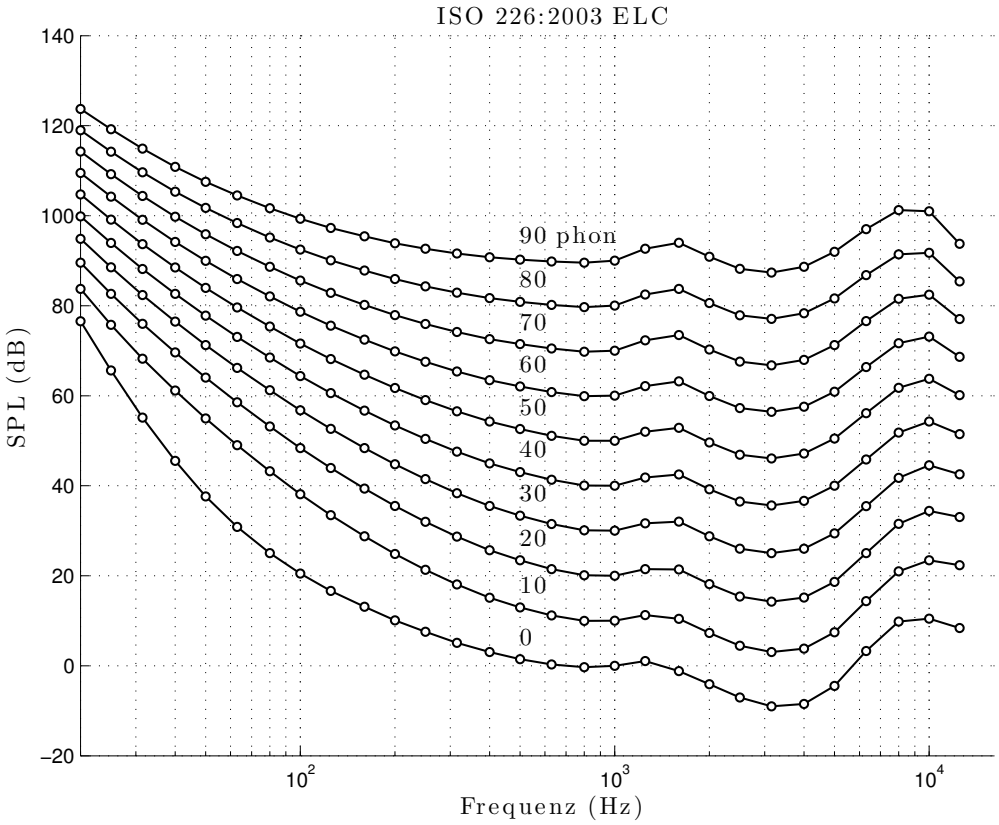


Abbildung 2.3: Kurven gleicher Lautheit (*equal loudness contours, ELC*) nach ISO 226:2003. Die Lautstärkepegel (*sound pressure levels, SPL*) wurden mit den Gleichungen nach ISO (2003, Abschnitt 4.1) berechnet. Die Kreise kennzeichnen die in der Normschrift publizierten Werte.

Der Amplitudenverlauf eines Außen-Mittelohr-Filters (AMF) kann nach Glasberg und Moore (1990) für Frequenzen unterhalb von 1 kHz durch die inverse 100-phon ISO-Kurve, sowie für darüber liegende Frequenzen über den Kurvenverlauf der inversen Ruhehörschwelle nachgebildet werden. Zwicker und Fastl (1999) modellieren den Amplitudenverlauf für Frequenzen größer 2 kHz über die Kurve

der inversen Ruhehörschwelle, verwenden für den Bereich kleiner 2 kHz jedoch einen konstanten Amplitudenverlauf mit 0 dB Verstärkung. Agerkvist (1994) schlägt vor über den gesamten Frequenzbereich die inverse 90-phon ISO-Kurve zu verwenden. Dies wird vor allem dadurch begründet, dass für die 100-phon ISO-Kurven bei 25 Hz, 10 kHz und 12,5 kHz keine experimentell ermittelten Daten vorliegen und bei Frequenzen über 1 kHz die Abweichung des Amplitudenverlaufs der 90-phon ISO-Kurve von dem der Ruhehörschwelle vernachlässigbar klein ist (vgl. dazu auch Pflüger, 1997; Terhardt, 1998; Zwicker und Fastl, 1999).

Glasberg und Moore (2002) verwenden zur digitalen Nachbildung der Filterfunktion des Außen- und Mittelohres ein FIR Filter[19] mit 4097 Koeffizienten. Der Amplitudengang entspricht dabei den in Moore et al. (1997) diskutierten Kurven. Pflüger (1997, S. 91–94) verwendet in einem alternativen Vorschlag zwei in Kaskade geschaltete IIR Filter (s. Abb. 2.4). Dabei bildet ein rekursives Tiefpassfilter 8ter Ordnung ($H_{\mathrm{AMF,RES\text{-}TP}}$) die Resonanzen der inversen Ruhehörschwelle bei 4 kHz und 12 kHz nach. Der Amplitudengang dieses Filters ist unter 1 kHz konstant 0 dB. Die Nachbildung der Filterflanke zu tiefen Frequenzen hin erfolgt durch ein in Serie geschaltetes Hochpassfilter 2ter Ordnung ($H_{\mathrm{AMF,HP}}$). Ein wesentlicher Vorteil der Kaskadierung besteht darin, dass sich die Flanke des Amplitudengangs unter 1 kHz mit einem einzigen Parameter steuern lässt. Dadurch lassen sich sowohl die Ruhehörschwelle als auch die inverse 100-phon ISO-Kurve nachbilden (s. Abb. 2.5). In Anwendungen der Telekommunikation werden Sprachsignale in der Regel auf eine Bandbreite $\leq$8 kHz beschränkt und mit Abtastfrequenzen $\leq$16 kHz digitalisiert. In diesem Fall entfällt die Nachbildung der Resonanz der inversen Ruhehörschwelle bei 12 kHz. Nach Zotter (2004) kann, wie in Abb. 2.6 dargestellt, zur Erhöhung der Recheneffizienz bei niedrigen Abtastraten das Tiefpassfilter 8ter Ordnung ($H_{\mathrm{AMF,RES\text{-}TP}}$) durch ein Resonanzfilter ($H_{\mathrm{AMF,RES}}$) mit nachgeschaltetem Tiefpassfilter ($H_{\mathrm{AMF,HP}}$) ersetzt werden. Eine effiziente und numerisch robuste Implementierung wird durch die Kaskadierung von IIR Filtern 2ter Ordnung (*second order section, SOS*) in kanonischer Direktform-2 (vgl. dazu Kammeyer und Kroschel, 1992, S. 60–62) erreicht (Abb. 2.4a).

[19] Nichtrekursive Filter (*finite impulse response, FIR*) sind gekennzeichnet durch eine Impulsantwort mit garantiert endlicher Länge. Rekursive Filter (*infinite impulse response, IIR*) führen

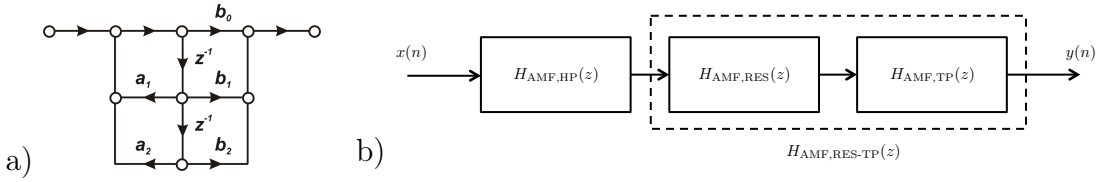


Abbildung 2.4: a) Biquadratische Filterstruktur, b) Struktur eines Außen-Mittelohr Filters (AMF), bestehend aus Hochpass-, Resonanz- und Tiefpassfilter.

Eine weitere Möglichkeit der funktionalen Nachbildung der Außen-Mittelohr-Übertragungsfunktion besteht in der Approximation des Amplitudenverlaufs mittels modifiziertem Yule-Walker Algorithmus (vgl. Friedlander und Porat, 1984; Friedlander und Sharman, 1985; Porat und Friedlander, 1985). Die Ableitung der Filterkoeffizienten beruht hierbei auf der Theorie autoregressiver Modelle und wird im letzten Teilkapitel dieses Abschnitts ausführlicher beschrieben. Das hier vorgestellte rekursive Filter 8ter Ordnung bildet, ähnlich dem Ansatz von Pflüger, die Resonanzen bei 4 kHz und 12 kHz nach. Die Nachbildung des Amplitudengangs zu tiefen Frequenzen hin erfolgt auch hier durch ein Hochpassfilter 2ter Ordnung ($H_{\mathrm{AMF,HP}}$). Der resultierende Amplitudenverlauf ist in Abb. 2.7 dargestellt.

Hochpassfilter für Frequenzen <1 kHz. Die benötigte Übertragungsfunktion lässt sich nach Pflüger (1997, S. 91–94) wie folgt realisieren

$$H_{\mathrm{AMF,HP}}(z) = \frac{(1 - r_{1,\mathrm{HP}} z^{-1})^2}{(1 - r_{2,\mathrm{HP}} z^{-1})^2} \tag{2.8}$$

wobei der Nullstellenradius mit $r_{1,\mathrm{HP}} = 1$ vorgeschlagen wird. In dieser Arbeit wird das AMF als Vorfilter der einkanaligen Störgeräuschreduktion (s. Kap. 4) verwendet. Die durch das AMF hervorgerufenen Signalverzerrungen lassen sich durch inverse Filterung rückgängig machen, sodass sich mit einem Nachfilter das Anregungssignal zurückgewinnen lässt. Um eine stabile Inversion des Filters zu gewährleisten, kann $r_{1,\mathrm{HP}} = 0{,}995$ angenommen werden, ohne dabei wesentliche Klangfärbungen des Signals zu erzeugen (s. a. Zotter, 2004, Kap. 3.1.1). Zur Nachbildung der inversen 100-phon ISO-Kurve wird in dieser Arbeit für breitbandige Signale und ei-

das Ausgangssignal an den Eingang zurück und haben eine unendlich lange Impulsantwort (vgl. Kammeyer und Kroschel, 1992; Oppenheim et al., 1998).

ne Abtastrate von $f_s = 44{,}1\,\text{kHz}$ der Polstellenradius mit $r_{2,\text{HP}} = 0{,}984$ festgelegt. Die Abweichung zu dem von Pflüger vorgeschlagenen Wert, $r_{2,\text{HP}} = 0{,}989$, erklärt sich durch die Verwendung unterschiedlicher ISO-Normen. Die hier verwendeten Kurven nach ISO 226:2003 verlaufen zu tiefen Frequenzen hin etwas steiler als die von Pflüger verwendeten ISO 226:1987 Kurven. Nach Pflüger (1997, S. 92) ist es von wesentlichem Vorteil, dass sich ohne Veränderung der globalen Filterstruktur ebenfalls der Verlauf der inversen Ruhehörschwelle nachbilden lässt. Dadurch lässt sich die Differenz von Außen-Mittelohr-Amplitudengang und Ruhehörschwelle, und somit das innere Rauschen der Cochlea, in geschlossener Form berechnen (s. a. Pflüger, 1997, S. 90). Wie aus Abb. 2.5 ersichtlich, wird die inverse Ruhehörschwelle bei dem von Pflüger vorgeschlagenen Polstellenradius von $r_{2,\text{HP}} = 0{,}957$ nur für Frequenzen größer 70 Hz mit ausreichender Genauigkeit nachgebildet. Der Amplitudengang des parametrisierbaren Hochpassfilters ist für Frequenzen größer 1 kHz konstant 0 dB.

In den Abb. 2.5 bis 2.7 ist die Änderung der unteren Flanke des Amplitudengangs als Funktion des Polstellenradius dargestellt. In Abb. 2.5 und 2.7 entspricht die graue Fläche einem Wertebereich von $r_{2,\text{HP}} = [0{,}957; 0{,}984]$, bei breitbandiger Modellierung und einer Abtastrate von $f_s = 44{,}1\,\text{kHz}$ (vgl. dazu auch Pflüger 1997, Abb. 6.4). In Abb. 2.6 wird, für bandbegrenzte Modelle und eine Abtastratenreduktion auf $f_s = 22{,}05\,\text{kHz}$, ein Wertebreich von $r_{2,\text{HP}} = [0{,}916; 0{,}962]$ dargestellt. Aus den Abbildungen ist leicht ersichtlich, dass bei den betrachteten Kaskaden-Filter-Modellen im Frequenzbereich unter 1 kHz die Differenz der Übertragungsfunktion des AMF zur Ruhehörschwelle ausschließlich vom Hochpassfilter $H_{\text{AMF,HP}}$ bestimmt wird.

Tiefpass- und Resonanzfilter für Frequenzen >1 kHz. Der Frequenzgang der modellierten Außen-Mittelohr-Übertragungsfunktion sollte für mittlere und hohe Frequenzen möglichst gut mit dem Kurvenverlauf der inversen Ruhehörschwelle übereinstimmen. Pflüger (1997, S. 91) schlägt zur breitbandigen Modellierung ($f_s = 44{,}1\,\text{kHz}$) ein rekursives Tiefpassfilter 8ter Ordnung vor

$$H_{\text{AMF,RES-TP}}(z) = 0{,}109\left(z^8 + z^7\right)\left(z^8 - 2{,}5359z^7 + 3{,}9295z^6 - 4{,}7532z^5 + \right.$$
$$\left. + 4{,}7251z^4 - 3{,}5548z^3 + 2{,}1396z^2 - 0{,}9879z + 0{,}2836\right)^{-1}\,, \quad (2.9)$$

welches die Resonanzen der inversen Ruhehörschwelle bei 4 kHz und 12 kHz nachbildet. Der Amplitudengang unter 1 kHz ist konstant 0 dB. Dadurch wird eine Entkopplung des parametrisierbaren Hochpassfilters ($H_{\mathrm{AMF,HP}}$) und des Tiefpassfilters ($H_{\mathrm{AMF,RES\text{-}TP}}$) erreicht. In Abb. 2.5 ist der Amplitudengang des Modellfilters im Vergleich zur inversen 100-phon ISO-Kurve und der inversen Ruhehörschwelle dargestellt.

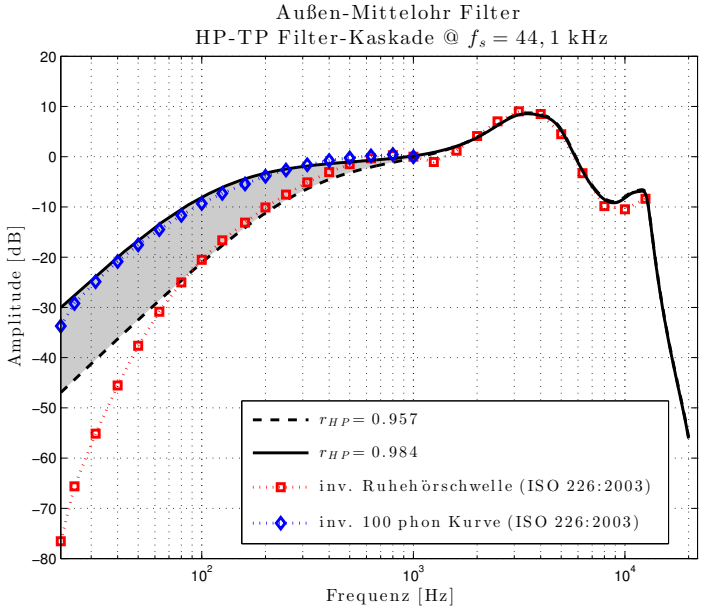


Abbildung 2.5: Amplitudengang eines Außen-Mittelohr-Filters (Hochpass-Tiefpass Filter-Kaskade) mit Polstellenradii $r_{2,\mathrm{HP}} = 0{,}957$ (gestrichelte Linie) und $r_{2,\mathrm{HP}} = 0{,}984$ (durchgezogene Linie) zu inverser Ruhehörschwelle (Quadrate) und inverser 100-phon ISO-Kurve bis 1 kHz (Rauten), vgl. dazu (Pflüger, 1997, S. 92), bei einer Abtastfrequenz von $f_s = 44100$ Hz.

Bei bandbegrenzten Modellen mit $f_s = 16$ kHz oder 22,05 kHz, die in der digitalen Sprachsignalverarbeitung Anwendung finden, kann das Tiefpassfilter 8ter Ordnung ($H_{\mathrm{AMF,RES\text{-}TP}}$) durch ein Resonanzfilter ($H_{\mathrm{AMF,RES}}$) mit nachgeschaltetem Tiefpassfilter ($H_{\mathrm{AMF,HP}}$) ersetzt werden (vgl. dazu Zotter, 2004). Das parametrische Resonanzfilter wird dabei in der von Zölzer et al. (2002, S. 52–54) vorgeschlagenen Regalia-Struktur mit Allpass-Filtern implementiert (vgl. dazu auch Regalia et al., 1988). Die Übertragungsfunktion des IIR-Resonanzfilters 2ter Ordnung kann wie folgt beschrieben werden (s. a. Zotter, 2004, Kap. 3.1.3):

$$H_{\mathrm{AMF,RES}}(z) = \frac{1}{1 - 2\,r_{\mathrm{RES}}\,\cos(\Theta_{\mathrm{RES}})z^{-1} + r_{\mathrm{RES}}z^{-2}} \cdot \left(\left[\frac{(1 + r_{\mathrm{RES}}^2)(1 - g)}{2} + g \right] - 2\,r_{\mathrm{RES}}\,\cos(\Theta_{\mathrm{RES}})z^{-1} + \left[\frac{(1 + r_{\mathrm{RES}}^2)(1 + g)}{2} - g \right] z^{-2} \right) \tag{2.10}$$

Mit den Parametern r_{RES}, Θ_{RES} und g können die Bandbreite, die Mittenfrequenz und die Überhöhung bzw. Absenkung gegenüber 0 dB Verstärkung unabhängig voneinander eingestellt werden. Die Inversion dieses Filters stellt im Regelfall kein Problem dar.

Zur besseren Modellierung des hohen Frequenzbereichs wird das Resonanzfilter ($H_{\mathrm{AMF,RES}}$) mit einem Tiefpassfilter in Serie geschaltet, wobei hier aufgrund der Bandbegrenztheit die Resonanz bei 12 kHz nicht nachgebildet wird. Die Übertragungsfunktion des Tiefpassfilters ergibt sich zu

$$H_{\mathrm{AMF,TP}}(z) = \frac{(1 + r_{1,\mathrm{TP}}z^{-1})^2}{(1 + r_{2,\mathrm{TP}}z^{-1})^2}\ . \tag{2.11}$$

Bei der Wahl der Null- und Polstellenradien ($r_{1,\mathrm{TP}}$, $r_{2,\mathrm{TP}}$) muss, um bei bandbegrenzten Signalen eine möglichst gute Approximation des Amplitudenverlaufs zu erhalten, die Abhängigkeit von der jeweiligen Abtastrate mit berücksichtigt werden (vgl. dazu auch Tabelle 2.1). Die Implementierung ergibt im Regelfall keine für die Inversion kritischen Nullstellenradien. Bei einer Bandbegrenzung kleiner 10 kHz spielt das Tiefpassfilter nur eine untergeordnete Rolle und kann zur Erhöhung der Recheneffizienz weggelassen werden.

In Abb. 2.6 ist der Amplitudengang des Modellfilters für $f_s = 22{,}05\,\mathrm{kHz}$ im Vergleich zur inversen 100-phon ISO-Kurve und der inversen Ruhehörschwelle dargestellt. Wie zu erkennen ist, wird die Außen-Mittelohr-Übertragungsfunktion bei 1 - 2 kHz nicht so exakt nachgebildet wie bei der Modellierung nach Pflüger. Der hauptsächliche Vorteil der Kaskadierung von IIR-Filtern 2ter Ordnung besteht somit in der höheren Recheneffizienz.

Tabelle 2.1 fasst die Parameter der Modellierung eines Außen-Mittelohrfilters mit einer Kaskade rekursiver Filter 2ter Ordnung zusammen. Wie in der jeweils

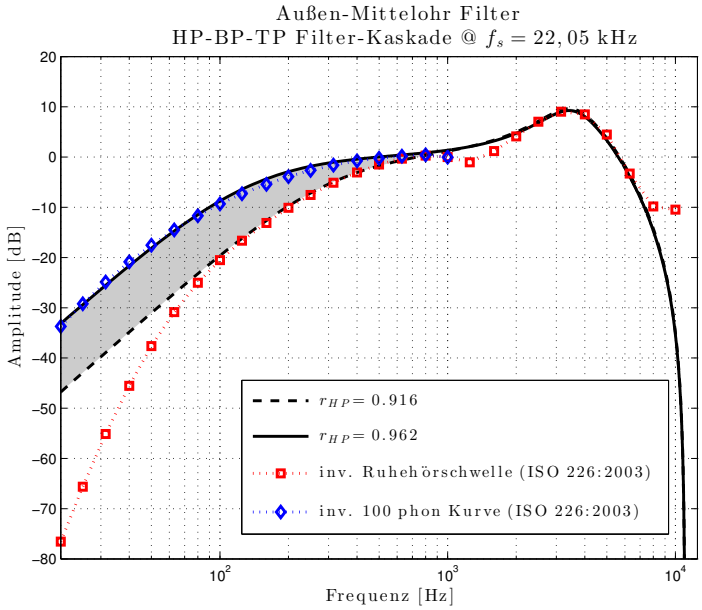


Abbildung 2.6: Amplitudengang des Außen-Mittelohr-Filters (Hochpass-Bandpass-Tiefpass Filter-Kaskade) mit Polstellenradii $r_{2,\mathrm{HP}} = 0{,}916$ (gestrichelte Linie) und $r_{2,\mathrm{HP}} = 0{,}962$ (durchgezogene Linie) zu inverser Ruhehörschwelle (Quadrate) und inverser 100-phon ISO-Kurve bis 1 kHz (Rauten) bei einer Abtastfrequenz von $f_s = 22050$ Hz.

zweiten Zeile ersichtlich, müssen bei Weglassen des Tiefpassfilters die Parameter des Resonanzfilters angepasst werden.

Tabelle 2.1: Zusammenstellung der AMF Parameter bei Abtastraten $f_s = 22{,}05$ kHz und $f_s = 16$ kHz (vgl. dazu Zotter, 2004, Kap. 3.1.4).

$\mathbf{f_s}$ [Hz]	$\mathbf{r}_{1,\mathrm{HP}}$	$\mathbf{r}_{2,\mathrm{HP}}$	$\mathbf{r}_{\mathrm{RES}}$	$\mathbf{\Theta}_{\mathrm{RES}}$	$\mathbf{g}_{\mathrm{RES}}$	$\mathbf{r}_{1,\mathrm{TP}}$	$\mathbf{r}_{1,\mathrm{TP}}$
22050	0,995	0,976	0,660	1,044	2,720	0,510	0,160
	0,955	0,976	0,660	0.980	2,720	-	-
16000	0,955	0,970	0,650	1,577	3,720	0,550	0,350
	0,955	0,970	0,650	1,420	3,360	-	-

Modifizierter Yule-Walker Algorithmus. Eine weitere Möglichkeit ein geeignetes Außen-Mittelohr-Filter zu berechnen, besteht in der Modellierung des gewünschten Amplitudengangs mit Hilfe des Modifizierten Yule-Walker (MYW)

Algorithmus[20] (vgl. dazu Friedlander und Porat, 1984; Friedlander und Sharman, 1985; Porat und Friedlander, 1985). Die Modellierung des Frequenzgangs beruht auf einem autoregressiven (*autoregressive moving average, ARMA*) Signalmodell, welches zum Beispiel in Haykin (2002a, S. 45–93) ausführlich diskutiert wird. Der MYW Algorithmus berechnet das Pol- und Nullstellenpolynom eines rekursiven Filters N-ter Ordnung, wobei die Optimierung der Modellbildung im Zeitbereich im Sinne der kleinsten Fehlerquadrate (*least squares approximation, LSA*) durchgeführt wird. Dabei kann der Algorithmus zur AMF Modellierung in folgende Arbeitsschritte unterteilt werden:

- Die Koeffizienten des Polstellenpolynoms werden, unter Verwendung der durch inverse Fourier Transformation (IFFT) aus dem gewünschten Amplitudengang erhaltenen Korrelationskoeffizienten, über die MYW Gleichungen bestimmt.

- Danach wird das Nullstellenpolynom in vier Schritten berechnet:

    1. Berechnung des Nennerpolynoms aus dem Leistungsspektrum der zu modellierenden Außen-Mittelohr Kurve.
    2. Die Übertragungsfunktion wird aus Pol- und Nullstellenpolynom zusammengesetzt und daraus die Frequenzantwort ermittelt.
    3. Die Impulsantwort des Filters wird mittels spektraler Faktorisierung (*spectral factorization technique*) berechnet (vgl. dazu Marple, 1987, S. 184–186).
    4. Optimierung des Nullstellenpolynoms durch LS-Approximierung der Impulsantwort.

Abb. 2.7 zeigt die Modellierung einer inversen Außen-Mittelohr Kurve mit Hilfe eines MYW Algorithmus 8ter Ordnung. Wie aus der Abbildung leicht ersichtlich ist, wird der gewünschte Amplitudenverlauf auch im hohen Frequenzbereich sehr gut angenähert. Dies ist vor allem dann wesentlich wenn die Bandbreite des zu verarbeitenden Audiosignals den gesamten Hörbereich umfasst. Um bei

[20] Der MYW Algorithmus ist in MATLAB® in der Funktion yulewalk.m (Signal Processing Toolbox) implementiert und wurde in dieser Arbeit zur Filtermodellierung verwendet.

möglichst geringer Modellordnung auch im tiefen Frequenzbereich eine ausreichend gute Approximation zu erreichen, wird ein Hochpass 2ter Ordnung in Kaskade geschaltet. Somit ergibt sich der in Abb. 2.7 als durchgezogene Linie dargestellte Amplitudenverlauf des modellierten AMF.

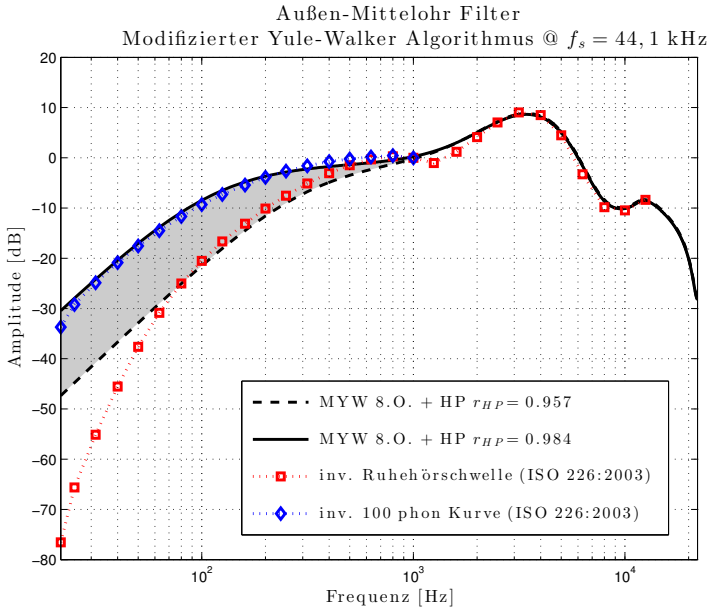


Abbildung 2.7: Amplitudengang eines Außen-Mittelohr-Filters (modifizierter Yule-Walker Algorithmus 8ter Ordnung mit in Kaskade geschaltetem Hochpass-Filter) mit Polstellenradii $r_{2,\mathrm{HP}} = 0{,}957$ (gestrichelte Linie) und $r_{2,\mathrm{HP}} = 0{,}984$ (durchgezogene Linie) zu inverser Ruhehörschwelle (Quadrate) und inverser 100-phon ISO-Kurve bis 1 kHz (Rauten) bei einer Abtastfrequenz von $f_s = 44100\,\mathrm{Hz}$.

C Richtungsbezogene Filterung

Bei physiologisch motivierten Außen-Mittelohr-Modellen wird angenommen, dass die räumliche Schallfeldfeldinformation bereits vollständig in dem am Eingang des äußeren Gehörgangs anliegenden Zeitsignal kodiert ist. Für die räumliche Kodierung ist die akustische Wirkung von Torso, Kopf und Pinna von grundlegender Bedeutung (s. a. Kap. 2.1.1). Die Modellierung der anatomisch bedingten Verzerrungen des Amplituden- und Phasenverlaufs der aus einer bestimmten Richtung einfallenden Schallwelle stellt somit das erste Glied einer physiologisch motivierten Signalverarbeitungskaskade dar. Diese Modelle lassen sich im Allgemeinen nur

durch sehr aufwendige numerische Verfahren, wie zum Beispiel der Randelementmethode (*boundary element method, BEM*), berechnen.

Bei funktional motivierten Modellen wird die räumliche Information hingegen mit linearen Filtern, den sogenannten Außenohrübertragungsfunktionen (*head related transfer functions, HRTF*), modelliert. Diese bilden die richtungsabhängigen Klangfärbungen als Funktion der Schalleinfallsrichtung, bzw. im Nahfeld als Funktion der Schallquellenposition, nach (siehe z. B. Blauert, 1974; Shaw, 1974b; Shaw und Vaillancourt, 1985; Møller et al., 1995; Brungart und Rabinowitz, 1999; Suzuki et al., 2011; Xie, 2013). Es ist oft von Vorteil die zeitlichen und spektralen Eigenschaften der Lokalisationswahrnehmung getrennt voneinander zu modellieren. HRTF sind kausal und lassen sich in einen minimalphasigen Anteil und eine Allpass-Übertragungsfunktion aufspalten (siehe z. B. Oppenheim et al., 1998, Kap. 5.6). Mehrgardt und Mellert (1977) haben gezeigt, dass die Phase der HRTF-Allpasskomponente in einem Frequenzbereich bis 10 kHz annähernd linear verläuft. Der Phasengang kann somit durch eine einfache, von der Schalleinfallsrichtung abhängige, Zeitverzögerung angenähert werden. Diese entspricht dem interauralen Laufzeitunterschied (ITD, s. Kap. 2.1.1). Zur Modellierung der ITD wird üblicherweise die frequenzunabhängige mittlere Gruppenlaufzeit aus den gemessenen HRTF geschätzt. Ein ausführlicher Vergleich der unterschiedlichen Schätzverfahren ist in Minnaar et al. (2000) sowie Katz und Noisternig (2014) zu finden. Die spektralen Komponenten der Außenohrübertragungsfunktionen werden über den minimalphasigen Anteil modelliert. Eine perzeptive Evaluierung dieses Modells ist zum Beispiel in Wightman und Kistler (1992), Kistler und Wightman (1992) und Kulkarni et al. (1999) zu finden. Zur weiteren Vereinfachung der Modellbildung wird in der Literatur zudem oft nur zwischen Freifeldkurve (d. h. bei Schalleinfall aus Blickrichtung) und Diffusfeldkurve (d. h. bei Schalleinfall aus allen Richtungen) unterschieden (vgl. dazu Vorländer, 1991; Terhardt, 1998; Zwicker und Fastl, 1999).

Außenohrübertragungsfunktionen können

- gemessen (s. a. Shaw und Teranishi, 1968; Abbagnaro et al., 1975; Mehrgardt und Mellert, 1977; Wightman und Kistler, 1989a; Pralong und Carlile, 1994; Brungart und Rabinowitz, 1999; Zotkin et al., 2006, u. v. a.),

- anhand von anthropometrischen Merkmalen modelliert (s. a. Middlebrooks, 1999a; Algazi et al., 2001, 2002; Zotkin et al., 2004; Nishino et al., 2007, u. v. a.), oder auch
- numerisch simuliert (s. a. Katz, 2001a,b; Otani und Ise, 2006; Kreuzer et al., 2009; Otani et al., 2009; Gumerov et al., 2010; Pollow et al., 2012, u. v. a.)

werden und weisen naturgemäß große interindividuelle Unterschiede auf (vgl. Wightman und Kistler, 1989a; Wenzel et al., 1993; Pralong und Carlile, 1996). Erfolgt die Wiedergabe virtueller Schallquellen mit Hilfe von nichtindividualisierten HRTF, verschlechtert sich sowohl die Lokalisationsgenauigkeit wie auch die Außer-Kopf-Lokalisation (siehe z. B. Wenzel et al., 1993; Middlebrooks, 1999b,a). Den Studien von Hofman et al. (1998), Shinn-Cunningham et al. (1998), Parseihian und Katz (2012) und Majdak et al. (2013) folgend, kann sich die menschliche Hörwahrnehmung in bestimmten Grenzen an veränderte spektrale Pinna-Cues anpassen. Das Richtungshören mit nichtindividualisierten HRTF kann demzufolge mit begrenzter Genauigkeit erlernt werden. Dies ist vor allem für praktische Anwendungen von Bedeutung, da die genannten Verfahren zur Ermittlung individualisierter HRTF sehr aufwendig sind.

Bei der Störgeräuschreduktion mittels spektraler Subtraktion können die Verzerrungen des Nutzsignals dadurch verringert werden, dass nicht wahrnehmbare Störgeräusche weniger stark unterdrückt werden (s. a. Kapitel 4). Dies wird typischerweise durch die Verwendung auditiver Filterbänke, welche die zeitliche und spektrale Maskierung des menschlichen Hörorgans nachbilden, erreicht. Bei räumlich verteilten Schallquellen muss, neben den für monophone Signale ermittelten Maskierungsfunktionen, auch die räumliche Demaskierung (*spatial release from masking, SRM*) der aus unterschiedlichen Raumrichtungen einfallenden Signalkomponenten berücksichtigt werden. Die SRM lässt sich jedoch durch keines der in der Literatur vorgestellten Modelle vollständig beschreiben (siehe z. B. Freyman et al., 2001; Arbogast et al., 2002; Brungart und Simpson, 2002; Noble und Perrett, 2002; Litovsky, 2005; Ebata, 2003; Marrone et al., 2008; Allen et al., 2011; Best et al., 2013; Glyde et al., 2013).

In dieser Arbeit werden die aus unterschiedlichen Raumrichtungen einfallenden Nutz- und Störsignalkomponenten mit einem Mikrofonarray-Beamformer (siehe Kapitel 5 und 6) weitgehend voneinander getrennt, bevor die Signale mit einer auditiven Filterbank (s. Kapitel 3) in den Spektralbereich transformiert werden. Durch die räumliche, zeitliche und spektrale Vorverarbeitung des mit dem Mikrofonarray aufgenommenen Schallfeldes kann neben der Simultanmaskierung und zeitlichen Verdeckung auch die räumliche Demaskierung in der spektralen Subtraktion berücksichtigt werden. Die Modellierung des räumlichen Hörens, insbesondere der SRM, ist bei diesem Ansatz nicht notwendig.

2.2.2 Modellbildung des Innenohres

Die im menschlichen Hörorgan durchgeführte Spektralanalyse wird hauptsächlich der Cochlea zugeordnet. Die sich auf der Basilarmembran ausbildende Wanderwelle transformiert die Frequenz eines Schallsignals in einen Ort maximaler Anregung. Aufgrund der Frequenz-Orts-Transformation kommt es zur Aufteilung der Signale in Teilbänder, bevor diese an den Hörnerv weitergeleitet werden (s. a. Kap. 2.1.3, Wanderwellentheorie nach Békésy). Pflüger (1997, S. 70) beschreibt drei Ebenen der Betrachtungsweise der auditiven Spektralanalyse: (i) die Basilarmembran, (ii) den Hörnerv und (iii) das Gesamtsystem der peripheren und zentralen Verarbeitung. Die Modellierung des Transduktionsprozesses[21] der Haarzellen wird aufgrund der Komplexität meist nicht im Detail durchgeführt. Somit beziehen sich die meisten Simulationen der auditiven Spektralanalyse auf die Ebene der Basilarmembran. Die unterschiedlichen Methoden können auch hier in physiologisch und funktional motivierte Modellbildung unterteilt werden. Dabei weisen physiologisch motivierte Modelle meist Kaskadenstruktur, funktional motivierte Modelle hingegen meist Parallelstruktur auf.

[21] Transduktion (von lat. *trans ducere* = Hindurchführung) beschreibt in der Physiologie die Umwandlung körperfremder chemischer, thermischer, mechanischer, o. ä. Energie in eine spezifische neuronale Signalform.

A Physiologisch motivierte Modelle

Physiologisch motivierte Modelle des Innenohres bilden die biophysikalische Funktionsweise der Cochlea nach. In der Literatur wird dabei meist zwischen makro- und mikromechanischer Modellbildung unterschieden (vgl. Zwicker, 1986; Lyon und Mead, 1988a,b; Kates, 1993; Giguere und Woodland, 1994a,b; Kates, 1995; Hawkins, 1995; Lyon, 1996). Bei der mikromechanischen Betrachtungsweise werden auch Details des Transduktionsprozesses im Cortischen Organ nachgebildet. Einfachere Modelle bilden hingegen meist nur die hydrodynamischen Eigenschaften der Basilarmembran nach (vgl. zum Beispiel Zwislocki, 1948; Zweig et al., 1976; Lyon und Mead, 1988a,b; de Boer, 1996; Givelberg und Bunn, 2003). Zur Vereinfachung werden physiologisch motivierte Modelle meist mit Hilfe von mathematisch einfacher beschreibbaren elektrischen Schaltkreisen dargestellt. Ein einfaches Modell zur Nachbildung der passiv angeregten Basilarmembran mit Hilfe von elektromechanischen Analogien wurde zum Beispiel von Lyon und Mead (1988b) vorgestellt. Andere Modelle versuchen neben der Nichtlinearität auch das aktive Verhalten der Cochlea (s. Kap. 2.1.3) zu berücksichtigen (vgl. van Netten und Duifhuis, 1983; Diependaal und Viergever, 1983), lassen sich meist jedoch nur numerisch simulieren (vgl. Diependaal et al., 1987). In dieser Arbeit werden ausschließlich funktional motivierte Modelle des Innenohres angewendet, die in den folgenden Teilkapiteln ausführlicher diskutiert werden.

B Funktional motivierte Modelle

Funktional motivierte Modelle des Innenohres bilden in erster Linie die experimentalphysiologischen Messdaten nach. Die Interpretation dieser Daten erfolgt anhand der den jeweiligen Prozessen zugrundeliegenden physikalischen (mechanischen) bzw. physiologischen Vorgängen. Genau genommen stellen funktional motivierte Modelle somit keine Näherung der mechanischen Vorgänge in der Cochlea, sondern der im auditiven System durchgeführten Signalanalyse dar (vgl. Pflüger, 1997, S. 71). Diese Modelle lassen sich vergleichsweise recheneffizient implementieren und finden aus diesem Grund bevorzugt Anwendung in der Audiosignalverarbeitung.

Bei funktional motivierten Modellen werden die Signale üblicherweise mit Hilfe von Filterbänken in die auditive Domäne übergeführt, wobei die Mittenfrequenzabstände und Bandbreiten an die spektrale Auflösung des Gehörs angepasst werden (s. a. Kap. 2.2.4). Der Literatur sind zahlreiche lineare, wie auch nichtlineare Implementierungen auditiver Filterbänke zu entnehmen (z. B. Patterson, 1987; Slaney, 1988, 1993; Cooke, 1991; Slaney et al., 1994; Hawkins, 1995; Colomes et al., 1995). Komplexere Modelle (z. B. Lopez-Poveda und Meddis, 2001; Meddis et al., 2010, 2013) bilden zudem den Transduktionsprozess der Haarzellen ab und sind für ein umfassendes Verständnis der menschlichen Hörwahrnehmung von zentraler Bedeutung. Diese Modelle sind meist sehr rechenintensiv und für Echtzeitanwendungen nicht geeignet. Zudem lassen sich die resultierenden Filter in den meisten Fällen nicht einfach invertieren, was deren Anwendung in Analyse-Synthese Systemen erschwert (vgl. dazu Necciari et al., 2013).

Bei der in dieser Arbeit vorgestellten auditiven Analyse-Synthese Filterbank (s. Kap. 3) wird bei der Modellbildung auf eine recheneffiziente Umsetzung mit möglichst geringer Systemlatenz geachtet, welche die Anwendung in Echtzeitsystemen (z. B. Freisprechanlagen in Telekommunikationsanlagen) erlaubt.

2.2.3 Leistungsdichtespektrum-Modell

Das Leistungsdichtespektrum-Modell (*power spectrum model, PSM*) bildet die Grundlage funktional motivierter auditiver Filterbänke (vgl. Fletcher, 1940) und wird in Moore (1993) ausführlich diskutiert. Zur Modellbildung werden folgende Annahmen getroffen (vgl. Moore, 1993; Pflüger, 1997, S. 12–13):

- Nachbildung des peripheren Gehörs mit einer parallelen Bank überlappender, linearer Bandpassfilter (auditive Filter).

- Bei der Wahrnehmung eines schmalbandigen Testsignals in einem Maskierer sind nur diejenigen Spektralkomponenten des Maskierers von Bedeutung, die um das Testsignal zentriert sind. Die Gewichtung der einzelnen Spektralkomponenten wird durch die Form der auditiven Filter bestimmt.

- Testsignal und Maskierer werden durch Leistungsdichtespektren repräsentiert, wobei die Länge des Zeitfensters das menschliche Zeitauflösungsvermögen übersteigt.

Das PSM berücksichtigt somit weder die Nichtlinearitäten des peripheren Gehörs, noch den Effekt des *off-frequency listenings*[22]. Durch die Zeitfensterung zur Bildung des Leistungsdichtespektrums gehen zudem Phaseninformationen und Kurzzeitänderungen der Signale nicht in die Modellbildung mit ein. In den folgenden Abschnitten werden die funktionalen Verarbeitungsblöcke des PSM und deren Implementierungsformen ausführlicher diskutiert.

2.2.4 Konzept der kritischen Bandbreite

Das Konzept der kritischen Bandbreite (*critical bandwidth, CB*) lässt sich auf Experimente von Fletcher (1940) zurückführen. Aus diesen Experimenten wird ersichtlich, dass zur Maskierung eines schmalbandigen Schallereignisses nur jene Rauschleistungen relevant sind, die in einem begrenzten Frequenzbereich um das Testsignal zentriert sind. Die Ermittlung der kritischen Bandbreiten, die gemeinhin auch als Frequenzgruppen bezeichnet werden, kann durch die Bestimmung der Mithörschwelle eines durch schmalbandiges Rauschen variabler Bandbreite maskierten Sinustons erfolgen. Andere Verfahren messen die Mithörschwelle von mehreren durch Maskierungsrauschen verdeckten Sinustönen variabler Anzahl bzw. von durch zwei Sinustönen maskierten schmalbandigen Rauschens variabler Bandbreite und Mittenfrequenz (s. a. Zwicker et al., 1957; Greenwood, 1961; Fastl und Zwicker, 2007, S. 149–173). Im Hörfrequenzbereich ergeben sich 24 Frequenzgruppen (vgl. Zwicker, 1961; Zwicker und Terhardt, 1980). Die Frequenzgruppenbreite lässt sich nach Zwicker und Terhardt (1980, Gl. 4) wie folgt als Funktion der Mittenfrequenz f_c berechnen (vgl. auch Zwicker und Fastl, 1999, S. 164)

$$CB(f_c) = 25 + 75 \left[1 + 1{,}4\ 10^{-6} f_c^2\right]^{0{,}69} . \tag{2.12}$$

[22] Bei der Detektion schmalbandiger Signale kann wesentliche Information auch von Filterkanälen stammen, die nicht um das Signal zentriert sind, bzw. kann die Information mehrerer Filter kombiniert werden (vgl. Patterson und Nimmo-Smith, 1980).

Mit Gl. (2.12) lassen sich die von Zwicker et al. (1957; 1961) ermittelten Daten mit einer Genauigkeit von ±10% annähern.

Aus den von Rosen und Baker (1994) mit Hilfe der „*notched noise*"-Methode ermittelten auditiven Filterformen (s. Kap. 2.2.5) lässt sich eine den kritischen Bandbreiten qualitativ entsprechende Einteilung der Frequenzgruppen in energieäquivalente rechteckige Bandbreiten (*equivalent rectangular bandwidth, ERB*) ableiten. Nach Moore und Glasberg (1983) können die ERB-Breiten als Funktion der Mittenfrequenz wie folgt bestimmt werden (s. a. Glasberg und Moore, 1990, Gl. 3):

$$ERB(f_c) = 0{,}1079 f_c + 24{,}7\,. \tag{2.13}$$

Diese Gleichung ist auch für unsymmetrische auditive Filter gültig. Aus den Gleichungen (2.12) und (2.13) ist ersichtlich, dass die ERB-Breiten über den gesamten Frequenzbereich schmäler als die kritischen Bandbreiten sind. Um den gesamten Hörfrequenzbereich abzudecken werden 39 ERB-Bänder benötigt. In Abb. 2.8 werden die Bark- und ERB-Breiten vergleichend gegenübergestellt.

Aus physiologischer Betrachtung entspricht die Frequenzgruppeneigenschaft des Gehörs der Frequenz-Orts-Transformation auf der Basilarmembran und der damit verbundenen Spektralanalyse. Diese Interpretation wird sowohl durch die Tonheitsskala wie auch die ERB-Skala repräsentiert. Üblicherweise werden folgende Hin- und Rücktransformationen der linearen Frequenzskala in die hörspezifischen Frequenzgruppenskalen verwendet.

Tonheitsskala z. In der Tonheitsskala entspricht $z = 1\,\text{Bark} = 100\,\text{mel}$ genau einer kritischen Bandbreite. Zwicker und Terhardt (1980) schlagen folgende Transformation in den Hörfrequenzbereich (f in Hz) vor:

$$z = 13 \arctan\left(0{,}00076 f\right) + 3{,}5 \arctan\left(\frac{f}{7500}\right)^2 . \tag{2.14}$$

Diese besitzt jedoch keine einfach zu berechnende Inverse und ist nach Traunmüller (1990) nicht mit den kritischen Bandbreiten in Gl. (2.12) kompatibel. Eine

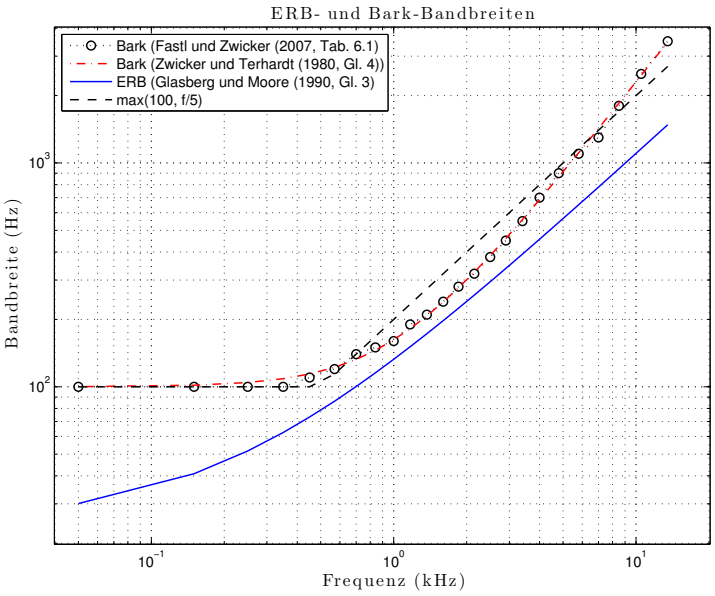


Abbildung 2.8: Vergleich der Bark- und ERB-Bandbreiten als Funktion der Frequenz: (punktierte Linie) Bark-Bandbreiten nach Fastl und Zwicker (2007, Tab. 6.1); (strichpunktierte Linie) approximierte Bark-Bandbreiten nach Zwicker und Terhardt (1980, Gl. 4), siehe auch Gl. (2.12); (durchgezogene Linie) ERB-Bandbreiten nach Glasberg und Moore (1990, Gl. 3), siehe auch Gl. (2.13); (gestrichelte Linie) Faustregel zur Abschätzung der Bark-Bandbreiten (Smith III und Abel, 1999, Abb. 11).

alternative Transformation ist in Traunmüller (1990, Gl. 6) zu finden:

$$\begin{aligned} z &= 26{,}81\, f/(1960 + f) - 0{,}53 \\ f &= 1966(z + 0{,}53)\ (26{,}28 - z) \end{aligned} \tag{2.15}$$

Mit den zugehörigen Bandbreiten (vgl. Traunmüller, 1990, Gl. 10):

$$CB(z) = 52548/(z^2 - 5256z + 690{,}39)\,. \tag{2.16}$$

Mit Gl. (2.15) lassen sich die experimentell ermittelten Daten in einem Frequenzbereich von $0{,}2 - 6{,}7\,\mathrm{kHz}$ mit einer Genauigkeit von $\pm 0{,}05$ Bark annähern. Die Messungenauigkeit der experimentell ermittelten Frequenzgruppenbreiten liegt in der Größenordnung von 0,1 Bark (siehe Smith III und Abel, 1999).

ERB-Skala ERBS. Nach Glasberg und Moore (1990, Gl. 4) lässt sich die ERB-Skala (ERB-Raten, *ERB rate scale*) durch Integration der reziproken ERB-Funktion, vgl. Gl. (2.13), berechnen.

$$\begin{aligned} ERBS_{\text{num}}(f) &= \int \frac{df}{ERB(f)} + \text{const} \\ &= 21{,}4 \log_{10}(1 + 0{,}00437\, f) \qquad (2.17) \\ &= 9{,}265 \ln(1 + 0{,}00437\, f)\ . \qquad (2.18) \end{aligned}$$

Dabei wird die Integrationskonstante so gewählt, dass sich $ERBS_{\text{num}}(0) = 0$ ergibt. Die Inverse obiger Gleichung lässt einfach berechnen und ergibt sich zu $f = 228{,}833 \left(e^{(ERBS_{\text{num}}/9{,}265)} - 1\right)$.

2.2.5 Auditive Filterformen

Die Form auditiver Filter wird – unter Annahme des Leistungsdichtespektrum-Modells – durch Messung der Hörschwelle eines maskierten Testtons bestimmt (vgl. Moore, 1995, S. 165–175). Als Maskierungssignal wird typischerweise breitbandiges Rauschen verwendet. Dadurch lassen sich Interferenzen innerhalb des Maskierers, sowie zwischen Maskierer und Testsignal vermeiden und die Mithörschwelle ausschließlich auf die Maskierung zurückführen. Es existieren zahlreiche Methoden der Formbestimmung, die sich durch die Wahl des Maskierungssignals unterscheiden. Bei der ‚*notched noise*'-Methode (vgl. Patterson, 1976; Patterson und Nimmo-Smith, 1980; Patterson und Henning, 1977; Shailer et al., 1989; Moore und Glasberg, 1983; Moore et al., 1990; Glasberg und Moore, 1990) wird die Mithörschwelle eines Sinustons, dessen Frequenz in der Mitte einer schmalbandigen Einkerbung gleichmäßig verdeckenden Breitbandrauschens liegt, als Funktion der Einkerbungsbreite bestimmt. Eine Erweiterung der Methode auf unsymmetrische Einkerbungen erlaubt die Messung unsymmetrischer auditiver Filterformen[23] (vgl. Patterson und Nimmo-Smith, 1980; Glasberg et al., 1984; Glasberg und Moore, 1990). Die auditiven Filterformen werden mit Hilfe der sogenannten roex-Kurven (siehe folgenden Abschnitt) mathematisch beschrieben. Eine weitere Methode zur

[23] Unsymmetrische auditive Filterformen treten bei hohen Pegeln wie auch bei gewissen Hörschädigungen auf.

Formbestimmung auditiver Filter stellt die „*rippled noise*“ Methode dar, bei der kammgefiltertes weisses Rauschen als Maskierer verwendet wird (vgl. Pick, 1980; Houtgast, 1977; Glasberg und Moore, 1990). Die Form der auditiven Filter wird entweder durch eine Fourier-Reihe angenähert oder, wie bei der *notched noise* Methode, über die roex-Funktionen beschrieben. Die mit der *rippled noise* Methode gewonnenen Filterformen weisen im Vergleich zur *notched noise* Methode einen breiteren Verlauf mit flacherer Spitze auf. Zudem kann mit dieser Methode nur ein geringer Dynamikbereich abgebildet werden (vgl. Moore, 1995, S. 172).

A roex - Funktionen

Patterson et al. (1982) nähern die auditiven Filterkurven mit Hilfe von roex-Funktionen (*rounded exponentials*) an, die in Abhängigkeit von der auf die Mittenfrequenz f_c normierten Frequenzvariable $g = |f - f_c|/f_c$ angegeben werden. Bei der **roex(p,r)-Funktion**

$$W(g) = (1 - r)(1 + pg)e^{-pg} + r \tag{2.19}$$

wird die auditive Filterform $W(g)$ mittels der Parameter p und r angenähert. Dabei steuert der Parameter p die Bandbreite und Flankenform und r die Abflachung der äußeren Flankenteile und somit die Dynamik des Filters (vgl. Abb. 2.9). Zur Beschreibung asymmetrischer Filterformen wird p typischerweise in p_l (für die tieffrequente Filterflanke) und p_u (für die hochfrequente Filterflanke) aufgeteilt. Glasberg und Moore (1986) zeigen zudem eine hohe Korrelation (0,92 bei $p < 0{,}01$) zwischen dem Parameter r, d. h. der Dynamikeinschränkung, und der Ruhehörschwelle (vgl. auch Nielsen, 1993). $W(g)$ stellt unter Berücksichtigung des Leistungsdichtespektrum-Modells eine Leistungsgröße dar. Aus den Messungen von Rosen und Baker (1994) ergeben sich folgende Parameter in Abhängigkeit vom Signalpegel P_S

$$p = \begin{cases} p_l = 39 - 0{,}42 \cdot P_S & \text{für } f < f_c \\ p_u = 27{,}1 & \text{für } f \geq f_c \end{cases} \tag{2.20}$$

$$r = \begin{cases} r_l = 10^{-4{,}67 + 0{,}042 \cdot P_S} & \text{für } f < f_c \\ r_u = 0 & \text{für } f \geq f_c \end{cases} \tag{2.21}$$

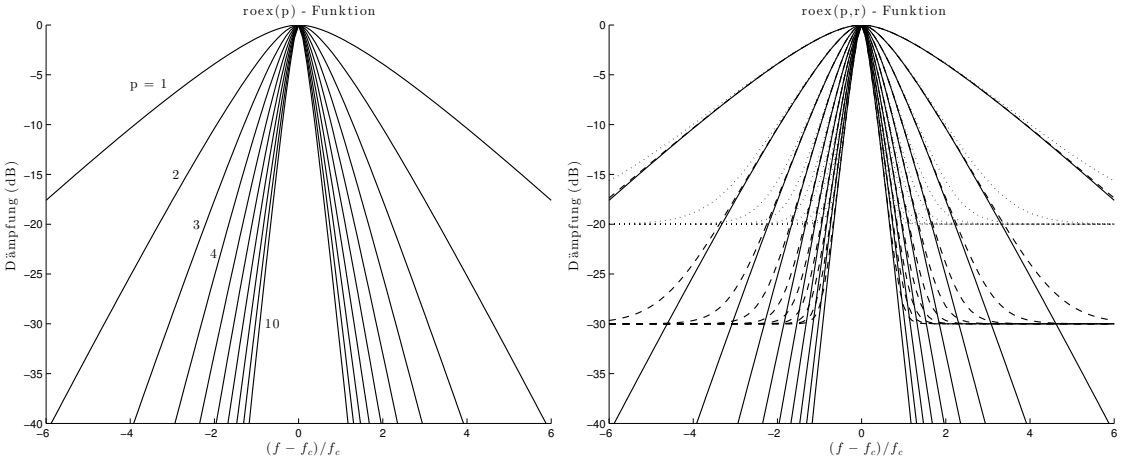


Abbildung 2.9: roex(p,r)-Funktion in Abhängigkeit von der normierten Frequenzvariable $(f - f_c)/f_c$ für unterschiedliche Parameter $p = [1, 8]$ (vgl. Pflüger, 1997, Abb. 2.12 und 2.13). Linkes Bild: $r = 0$. Rechtes Bild: $r = 0$ (durchgezogene Linien), $r = 0{,}001$ (gestrichelte Linien), $r = 0{,}01$ (punktierte Linien).

Bei der **roex(p,w,t)-Funktion** wird ein zusätzlicher Summationsterm zur genaueren Beschreibung des Filterverhaltens außerhalb des Durchlassbereichs eingeführt. Die auditiven Filter lassen sich wie folgt beschreiben (vgl. Abb. 2.10):

$$W(g) = (1 - w)(1 + pg)e^{-pg} + w(1 + tg)e^{-tg}\,, \tag{2.22}$$

wobei auch hier p zur besseren Modellierung der Asymmetrie in p_u und p_l aufgeteilt werden kann. Nach Rosen und Baker (1994) ist diese Beschreibungsform jedoch zu komplex für praktische Anwendungen. Auditive Filter werden meist über die wesentlich einfacheren roex(p,r)-Funktionen beschrieben.

B Implementierung auditiver Filter

Die Implementierung auditiver Filter kann mittels kaskadierter Tiefpassfilter (Kaskadenmodell), einer Bank paralleler Bandpassfilter, sowie nach dem Prinzip sogenannter Frequenzbereichsmodelle implementiert werden (vgl. Hawkins, 1995, S. 107–109).

Kaskadenmodelle. Die physiologisch motivierten Kaskadenmodelle simulieren die Wanderwellenausbreitung entlang der Cochlea mit Hilfe kaskadierter Tiefpass-

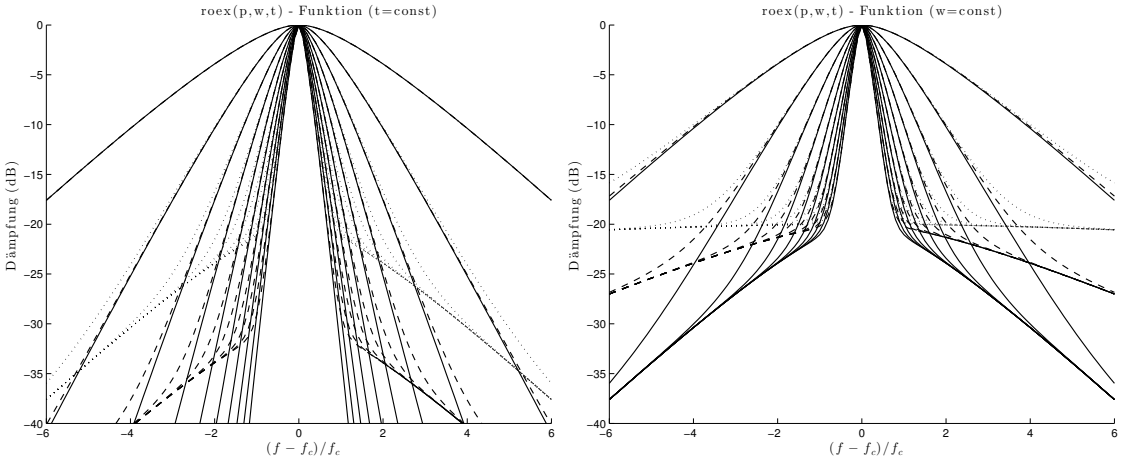


Abbildung 2.10: roex(p,w,t)-Funktion in Abhängigkeit von der normierten Frequenzvariable $(f - f_c)/f_c$ für unterschiedliche Parameter $p = [1, 8]$ (vgl. Pflüger, 1997, Abb. 2.14 und 2.15). Linkes Bild: $t = \text{const} = 1$; $w = 0$ (durchgezogene Linien), $w = 0{,}001$ (gestrichelte Linien), $w = 0{,}01$ (punktierte Linien). Rechtes Bild: $w = \text{const} = 0{,}01$; $t = 1$ (durchgezogene Linien), $t = 0{,}5$ (gestrichelte Linien) und $t = 0{,}1$ (punktierte Linien).

filter fallender Grenzfrequenz (vgl. Lyon und Mead, 1988b; Kates, 1993). Durch Parametrisierung jedes einzelnen Tiefpasses kann sowohl das nichtlineare wie auch das aktive Verhalten der Cochlea simuliert werden. Das *passive long-way model (PLWM)* nach Lyon (1982) und das *active short-wave model (ASWM)* nach Cooke et al. (1993, S. 94–116) ergeben sich aus der ein- bzw. zweidimensionalen Modellierung der hydrodynamischen Eigenschaften der Cochlea, wobei auch die Feuerungsraten der Nervenfasern simuliert werden. Lyon und Mead (1988a) zeigen anhand der VLSI-Implementierung[24] dieses Modells, dass die Gruppenlaufzeiten gegenüber den psychophysikalischen Messungen generell zu groß sind. Das Tiefpass-Kaskadenmodell nach Kates (1993, 1995) ist ähnlich aufgebaut wie das Modell von Lyon und Mead, nähert jedoch die Gruppenlaufzeiten der Cochlea wesentlich genauer an.

Parallel-Filterbankmodelle. Bei der Modellbildung wird das Verhalten der Cochlea durch eine Bank paralleler Bandpassfilter angenähert. Im Gegensatz zu

[24]VLSI (*very large scale integration*) bezeichnet den Integrationsgrad elektronischer Schaltkreise gemessen in Logikgattern (*gate equivalents*) und ist ein Maß für die Komplexität von Halbleiterstrukturen.

den Kaskadenmodellen erfolgt hier keine exakte Nachbildung des Wanderwellenverhaltens der Cochlea. Die einfachste dieser Modellbildung zuzuordnende Realisierungsform sind parallele Bandpassfilter konstanter Güte (*constant-Q filter bank*), deren Bandbreiten in etwa den Frequenzgruppenbreiten entsprechen. Dieser Ansatz beruht auf den Arbeiten von Youngberg und Boll (1978), Brown (1991) und Brown und Puckette (1992). Das gleichbleibende Verhältnis von Mittenfrequenz zu Bandbreite ergibt bei tiefen Frequenzen eine hohe Frequenzauflösung und zu hohen Frequenzen hin eine zunehmend genauere Zeitauflösung. Zur praktischen Implementierung werden die Analysefilter oft durch Mittelung der Amplituden in den jeweiligen Frequenzbändern angenähert (vgl. dazu Langendijk und Bronkhorst, 2002). Eine weitere typische Realisierungsvariante von constant-Q Filterbänken sind Oktav- und Terzbandfilter nach ANSI S1.11 (ANSI, 2004), siehe dazu auch Kuo et al. 2010 und Wei et al. 2010. Nach Smith III und Abel (1999) kann die ungleichmäßige spektrale Auflösung auch über Allpass-Strukturen erster Ordnung erreicht werden. Hierbei werden die Teilbandfilter über eine konforme Bilineartransformation an die Frequenzgruppen angepasst. O'Donovan und Furlong (2005) verwenden ein bilineares Zeit-Frequenz-Modell (auch *EarWig distribution, EWD*), eine Erweiterung des constant-Q Modells (*constant-Q modal distribution*) nach Pielemeier und Wakefield (1996), um die cochleären Zeit-Frequenz-Maskierungsfunktionen besser anzunähern.

Die constant-Q Filter nach Brown (1991) sind nicht invertierbar und aufgrund der hohen Frequenzauflösung (d. h. der hohen Anzahl von Kanälen pro Oktave) nur mit sehr hohem Rechenaufwand implementierbar. Bei der auditiven Wavelet-Transformation (AWT)[25], siehe auch Yang et al. (1992), Irino und Kawahara (1993), Chi et al. (2004) und Venkitaraman et al. (2014), wird die Impulsantwort der Basilarmembran über das Mutter-Wavelet angenähert. Dadurch ergibt sich ein an die cochleäre Verarbeitung angepasstes, jedoch einfach invertierbares Analyse-Synthese Filterbanksystem. Eine für möglichst kurze Signallatenzen optimierte AWT-Filterbank für Audio-Codecs wurde von Philippe et al. (1999) vorgestellt. Feldbauer et al. (2005) und Pichevar et al. (2011) zeigen recheneffiziente und näherungsweise invertierbare auditive Modelle, die ebenfalls für die

[25]Eine umfassende Einführung in die Wavelet- und Frame-Theorie ist in Christensen (2003, 2008), Mallat (2009) und Vetterli et al. (2011) zu finden.

Anwendung in Audio-Codecs optimiert wurden. Holighaus et al. (2013) verallgemeinern die mathematische Beschreibung von constant-Q Filterbänken mit Hilfe der Frame-Theorie. Dabei wird ein zum Analyse-Frame dualer Synthese-Frame (bzw. eine duale Basis) berechnet, woraus sich eine invertierbare und effiziente Implementierung der Teilbandfilter ableiten lässt. Dieser Ansatz wird bei der „*ERBlet*"-Transformation[26] von Necciari et al. (2013) dahingehend erweitert, dass die Bandbreiten der Teilbandfilter – genauer gesagt, die Bandbreiten der Zeit-Frequenz-Atome einer nichtstationären Gabor-Transformation (vgl. Balazs et al., 2011) – der ERB-Skala angepasst werden. Die Anpassung der Atome an die Maskierungsfunktionen der cochleären Verarbeitung erfolgt, wie in Swets et al. (1962), mit Gaußfunktionen. Dies ermöglicht die Konstruktion einer perfekt invertierbaren auditiven Zeit-Frequenz-Darstellung von Audiosignalen.

Patterson (1976) modelliert die in der Cochlea durchgeführte Signalanalyse über eine Bank paralleler Bandpassfilter, wobei die Übertragungsfunktion der Teilbandfilter mit Gaußfunktionen an die in psychoakustischen Experimenten ermittelten Maskierungsfunktionen angepasst wird. Ein erweitertes modulares Modell dieses Ansatzes ist in Patterson (1995) zu finden. Echtzeitimplementierungen auditiver Filter leiten sich vielfach vom Patterson-Holdsworth Modell (Patterson et al., 1987) ab. Dabei wird das Impulsverhalten der auditiven Nervenfasern mit einer Gammaton-Funktion angenähert (s. Kap. 2.3). Diese beschreibt die Impulsantwort der einzelnen Teilbandfilter, welche somit die Eigenschaften der Spektralanalyse der menschlichen Hörwahrnehmung modellieren. Gammaton-Filterbänke lassen sich gut parametrisieren. Für diese Filterform existieren Algorithmen zur recheneffizienten Implementierung, bestehend aus kaskadierten rekursiven Filtern niedriger Ordnung, die kurze Latenzzeiten ermöglichen (siehe z. B. Patterson et al., 1987, 1992; Slaney, 1993; Lyon, 1996; Hohmann, 2002). Darüber hinaus lässt sich das nichtlineare Verhalten der Cochlea durch Steuerung der Filterparameter in die Modellbildung mit einbeziehen (siehe z. B. Carney, 1993; Pflüger et al., 1997; Irino und Patterson, 1997, 2001). Meddis et al. (2001) und Lopez-Poveda und Meddis (2001) kombinieren mehrere Gammaton-Filter unterschiedlicher Bandbrei-

[26] Der Name ERBlet-Transformation leitet sich aus dem Verfahren selbst ab: Eine nichtstationäre Gabor/Wavelet-Transformation mit einer in ERB-Raten (vgl. Kap. 2.2.4) unterteilten Frequenzachse.

te um den Amplitudengang der Basilarmembran besser anzunähern. Mit Hilfe dieser DRNL (*dual-resonance nonlinear*) Filter lässt sich nicht nur die Frequenzabhängigkeit, sondern auch auch die Pegelabhängigkeit der experimentell ermittelten auditiven Filterformen nachbilden. Gammaton-Filterbänke lassen sich nur schwer invertieren. Nach Strahl und Mertins (2009) können mit Hilfe der Frame-Theorie Gammaton-Filterbänke mit nahezu perfekter Rekonstruktion abgeleitet werden. Um eine hohe Genauigkeit der Signalrekonstruktion zu erreichen wird hierbei jedoch eine relativ hohe Anzahl von sich überlappenden Teilbandfiltern benötigt. Eine ausführliche Diskussion der Analyse und Resynthese von Signalen mit Gammaton-Filterbänken wird in Kapitel 3 gegeben.

Frequenzbereichsmodelle. Nach Pflüger (1997, S. 77) lassen sich unter dem Überbegriff Frequenzbereichsmodelle diejenigen auditiven Modelle zusammenfassen, die in einer ersten Verarbeitungsstufe eine Zeit-Frequenz-Analyse beinhalten, welche nicht dem auditiven Modell selbst zugeordnet wird. Diese Modelle sind somit ebenfalls rein funktionaler Natur. Bei dieser Modellbildung kann, da keine direkte Verbindung zum physiologischen Hörmodell besteht, jede beliebige Zeit-Frequenz-Analyse verwendet werden und somit eine exakte Nachbildung der roex-Funktionen und des nichtlinearen Verhaltens des Gehörs erfolgen. Nachteilig wirkt sich hingegen die meist sehr rechenaufwendige Implementierung und die fehlende Echtzeitfähigkeit[27] der Algorithmen aus (vgl. bspw. Agerkvist, 1994).

2.3 Gammaton-Filter

Die Gammaton-Funktion wurde erstmals von Flanagan (1960a,b)[28] zur Modellierung der Schwingung der Basilarmembran im menschlichen Ohr eingeführt. Johannesma (1972), de Boer (1975) und de Boer und de Jongh (1978) haben

[27] Bei der klassischen Spektraltransformation wird meist ein begrenzter Signalausschnitt bzw. Signalblock mit Hilfe der schnellen Fourier Transformation (*fast Fourier transform*, FFT) und, im Fall von Resynthese, anschließender Fourier-Rücktransformation (IFFT) auf einmal transformiert. Dadurch können je nach Blocklänge wesentliche Systemlatenzen entstehen.

[28] Flanagan verwendete eine mit einer Gammafunktion modulierte Sinusschwingung. Diese wurde jedoch erst wesentlich später von Aertsen und Johannesma (1980) als „Gammaton-Funktion“ bezeichnet.

gezeigt, dass mit der revcor-Methode[29] an Katzen gemessene Impulsantworten auditiver Nervenzellen über die Gammaton-Funktion beschrieben werden können. Zudem lassen sich, den Studien von Schofield (1985) und Patterson et al. (1987) folgend, die roex-Funktionen (s. a. Kapitel 2.2.5) über die Gammaton-Funktion annähern und somit die von Patterson (1976) in experimentalpsychologischen Versuchen ermittelten auditiven Filter mit ausreichender Genauigkeit beschreiben (s. a. Carney und Yin, 1988; Patterson et al., 1992).

Recio und Rhode (2000) haben in elektrophysiologischen Experimenten an anästhesierten Chinchillas (*in vivo*) die Gruppenlaufzeit der Basilarmembran als Funktion der Anregungsfrequenz bestimmt (s. a. Rhode und Recio, 2000). Als Anregungssignal wurden dabei Klicks mit 20 μs Dauer verwendet. Die Ergebnisse dieser Studien zeigen, dass die Gammaton-Funktion den Phasengang der Basilarmembran nur mit unzureichender Genauigkeit abbildet. Bei phasenkritischen Anwendungen muss daher die Gruppenlaufzeit in Abhängigkeit von der Filtermittenfrequenz kompensiert werden (s. a. Kap. 3).

2.3.1 Lineare Gammaton-Filter

Die Impulsantwort $g_m(t)$ eines Gammaton-Filters (GTF) setzt sich aus einer Gammafunktion[30] multipliziert mit einer Trägerschwingung zusammen (vgl. Aertsen und Johannesma, 1980):

$$g_m(t) = a\, t^{m-1} e^{-bt} \cos(2\pi f_c t + \phi) \quad \forall\, t \geqslant 0, m \geqslant 1\,. \tag{2.23}$$

Die Gammafunktion beschreibt die Einhüllende der Impulsantwort, der Verstärkungsfaktor a dient der Skalierung. Die Filterordnung m bestimmt zum überwiegenden Teil die Flankensteilheit des annähernd symmetrischen Amplitudengangs. Der Parameter b bestimmt die Länge der Impulsantwort $g_m(t)$ und somit

[29] Die „*reverse correlation*“ (revcor) bzw. „*spike-triggered averaging*“ (STA) Methode ermittelt die korrelierte Aktivität der Neuronen, d. h. der Spikefolgen (Aktionspotenziale), bei Anregung mit zeitvarianten Stimuli (vgl. de Boer und Kuyper, 1968; Simoncelli et al., 2004). Die STA entspricht dem ersten Term einer Wiener-Volterra Reihenentwicklung der Systemantwort und bildet somit nur den linearen Teil des Systemverhaltens ab (vgl. Gazzaniga, 2004, S. 328).

[30] Genau genommen handelt es sich bei einer auf ganzzahlige Ordnungen $m \in \mathbb{N}$ beschränkten Gamma-Verteilung um eine sog. Erlang-Verteilung (siehe z. B. Forbes et al., 2011, Kap. 15).

die Bandbreite des Filters. Aus Gl. (2.23) ist ersichtlich, dass die Mittenfrequenz des GTF der Frequenz f_c der Trägerschwingung entspricht. Mit dem Parameter ϕ wird die Anfangsphase, d. h. die relative Position der Feinstruktur der Trägerschwingung zur Einhüllenden, festgelegt. Da die menschliche Hörwahrnehmung relativ unempfindlich gegenüber Phasenverzerrungen ist (siehe z. B. Patterson, 1987; Gockel et al., 2002; Carlyon und Shamma, 2003), wird die Anfangsphase zur Vereinfachung der Implementierung meist zu Null gesetzt. Die Impulsantwort und der Amplitudengang eines Gammaton-Filters 4ter Ordnung, mit einer Mittenfrequenz von 1000 Hz, sind in Abb. 2.11 dargestellt.

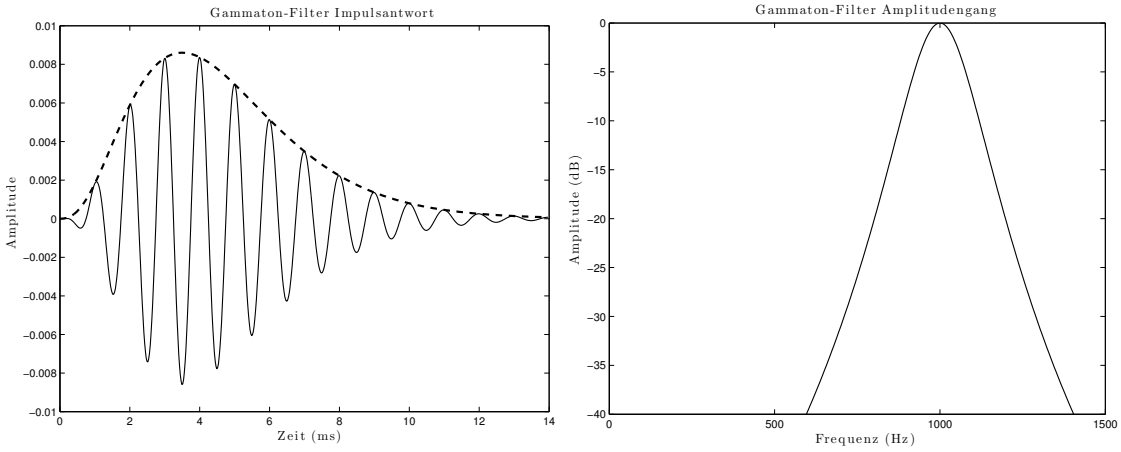


Abbildung 2.11: Impulsantwort und Amplitudengang eines GTF 4ter Ordnung mit einer Mittenfrequenz von $f_c = 1000\,\text{Hz}$. Im linken Bild sind der Realteil (durchgezogene Linie) und die Einhüllende (gestrichelte Linie) der Impulsantwort abgebildet. Das rechte Bild bildet den symmetrischen Amplitudengang des GTF ab.

Die Laplace-Transformierte der Gammaton-Funktion, Gl. (2.23), lässt sich für beliebige Ordnungen $\{m \mid m \in \mathbb{Z}, m > 0\}$ in geschlossen analytischer Form darstellen (s. Anhang A):

$$G_m(s) = (-1)^{m-1} \frac{(m-1)!}{2} \frac{(s+b-j\omega_c)^m + (s+b+j\omega_c)^m}{\left((s+b)^2+\omega_c^2\right)^m}\,. \tag{2.24}$$

$G_m(s)$ besitzt m konjugiert komplexe Polpaare bei $s_p = -b \pm j\omega_c$, sowie m reelle Nullstellen. Die Übertragungsfunktion lässt sich, unter Anwendung der Partialbruchzerlegung, in reelle Polynome 2ten Grades faktorisieren. Die Lage der Nullstellen wird dabei meist numerisch bestimmt.

Gammaton-Filter nach Gl. (2.23) sind linear, d. h. die resultierende Filterform ist pegelunabhängig. Nach Patterson et al. (1992) kann mit einer parallelen Bank linearer GTF der Ordnung $m = 4$ und $b = 2\pi \times 1{,}019 \times ERB(f_c)$ die Frequenzauflösung des menschlichen Gehörs ausreichend genau modelliert werden. Die $-3\,\mathrm{dB}$ Bandbreite entspricht dabei 0,887 × der ERB-Bandbreite (vgl. Patterson, 1994). Nach Holdsworth et al. (1988) lässt sich der Bandbreitenparameter b analoger GTF wie folgt für unterschiedliche Filterordnungen m bestimmen:

$$b = \alpha_m^{-1} ERB(f_c), \tag{2.25}$$

$$\alpha_m = \frac{\pi (2m-2)!\; 2^{-(2m-2)}}{(m-1)!^2}\,.$$

Daraus berechnet sich die $-3\,\mathrm{dB}$ Bandbreite zu

$$\Delta f_{b,3\mathrm{dB}} = c_m\, b = \hat{c}_m\, \alpha_m^{-1}\, ERB(f_c), \tag{2.26}$$

$$c_m = 2\sqrt{\sqrt[m]{2} - 1}.$$

Für den in Abb. 2.11 dargestellten GTF 4ter Ordnung ergeben sich folgende Werte für $\alpha_m^{-1} = 1{,}019$ und $c_m = 0{,}870$. Daraus berechnet sich der Bandbreitenparameter $b = \alpha_m^{-1} ERB(f_c) = 1{,}019 \times 128{,}14 = 130{,}57\,\mathrm{Hz}$ und die $-3\,\mathrm{dB}$ Bandbreite $\Delta f_{b,3\mathrm{dB}} = c_m\, b = 0{,}870 \times 130{,}57 = 113{,}59\,\mathrm{Hz}$. In Van Compernolle (1991, Gl. 2.6c) ist eine ähnliche mathematische Beziehung zur Berechnung der $-3\,\mathrm{dB}$ Bandbreite zeitkontinuierlicher Gammaton-Filter zu finden. Bei der Herleitung wird davon ausgegangen, dass die Amplitude eines GTF in der Nähe der Resonanzfrequenz nur von der Distanz zum nächstgelegenen m-fachen Pol und einem Verstärkungsfaktor abhängt. Mit dieser Annahme lässt sich, wie in Anhang A.5.1 gezeigt wird, ebenso die $-3\,\mathrm{dB}$ Bandbreite zeitdiskreter GTF berechnen.

Die Gammaton-Funktion in Gl. (2.23) kann relativ einfach vom zeitkontinuierlichen (analogen) in den zeitdiskreten (digitalen) Bereich[31] transformiert werden. In

[31] Ein zeitkontinuierliches (oder auch analoges) Signal ist eine Funktion $f : \mathbb{R} \to \mathbb{C}$. Ein zeitdiskretes Signal ist eine Folge $f : \mathbb{Z} \to \mathbb{C}$. Ein digitales Signal ist ein zeitdiskretes Signal dessen Bildbereich eine endliche Teilmenge von $\mathbb{C}$ ist.

der Literatur sind zahlreiche Implementierungen als FIR Filter (siehe z. B. Evans, 1986; Shackleton et al., 2000) und IIR Filter (siehe z. B. Patterson et al., 1987; Cooke, 1991; Van Compernolle, 1991; Slaney, 1993; Lyon, 1996; Pflüger, 1997; Hohmann, 2002; Zotter, 2004; Katsiamis et al., 2006) zu finden. Van Immerseel und Peeters (2003) und Lyon et al. (2010) geben eine umfassende Übersicht über die wichtigsten Implementierungsformen.

Cooke (1991) und Slaney (1993) kommen in ihren Studien zum Schluss, dass die Impulsinvarianz-Transformation die beste Übereinstimmung mit der Impulsantwort sowie dem Amplituden- und Phasengang analoger GTF ergibt. Dieses Ergebnis wurde in einer von Van Immerseel und Peeters (2003) durchgeführten vergleichenden Studie[32] unterschiedlicher Transformationsverfahren zur Herleitung digitaler GTF bestätigt. Dabei wird gezeigt, dass die von Cooke vorgeschlagene Impulsinvarianz-Transformation eines Tiefpass-Prototypfilters im Basisband (*baseband impulse invariant transformation*, BBIIT)[33] die beste Übereinstimmung mit der analogen Impulsantwort eines GTF ergibt. Ein vergleichbares Ergebnis kann mit der direkten Impulsinvarianz-Transformation (*impulse invariant transformation*, IIT) nach Slaney erzielt werden, wobei hier die Modulation und Demodulation vor und nach der Filterung entfällt. Dadurch verringert sich der Rechenaufwand gegenüber der BBIIT.

Anhang A fasst die Berechnung der Parameter analoger und digitaler GTF zusammen. In dieser Arbeit liegt das Hauptaugenmerk auf der recheneffizienten digitalen Implementierung von GTF als Teilbandfilter einer Analyse-Synthese-Filterbank zur auditiven Signalverarbeitung. Aus diesem Grund werden hier ausschließlich GTF bis zur 4ten Ordnung betrachtet, da diese die auditiven Maskierungskurven mit ausreichender Genauigkeit annähern und sehr recheneffizient mit rekursiven Filterstrukturen implementiert werden können (s. a. Patterson et al., 1992). Die zur digitalen Implementierung linearer GTF benötigte z-Transformierte wird in Anhang A.3 hergeleitet. Die Herleitung folgt dabei vorwiegend den Ausfüh-

[32] Van Immerseel und Peeters vergleichen den mittleren absoluten Fehler der Impulsantworten bzw. der Amplituden- und Phasengänge unterschiedlicher zeitdiskreter Realisierungen analoger Gammaton-Filter.

[33] Bei der BBIIT wird das Signal vor der Filterung durch Modulation mit der Mittenfrequenz ω_c in das Basisband ($\omega_c = 0$) verschoben. Dadurch reduziert sich das ursprüngliche Bandpassfilter auf ein Tiefpassfilter, welches sich wesentlich einfacher in die z-Ebene transformieren lässt. Das gefilterte Signal wird danach wieder in den Bandpassbereich zurückmoduliert.

rungen von Slaney (1993), Lyon (1996), Pflüger (1997), Zotter (2004) und Noisternig et al. (2009). Die allgemeine Form der z-Transformierten eines GTF der Ordnung m lautet (Slaney, 1993):

$$G_m(z) = \prod_{i=1}^{m} G_{m,i}(z) = \prod_{i=1}^{m} \frac{b_{0,i} + b_{1,i}z^{-1} + b_{2,i}z^{-2}}{1 + a_{1,i}z^{-1} + a_{2,i}z^{-2}} \tag{2.27}$$

Die Teilfilter $G_{m,i}(z)$ unterscheiden sich lediglich durch die Koeffizienten b_i. Die Berechnung der Koeffizienten a_i und b_i der Teilfilter wird in Anhang A.3 gezeigt. Vorsicht ist bei der Zusammenfassung der Teilfilter zu einem Gesamtfilter höherer Ordnung gegeben. Die Implementierung auf digitalen Systemen erfordert in diesem Fall eine sehr hohe numerische Genauigkeit der Koeffizienten. Aufgrund der endlichen Genauigkeit der Zahlendarstellung werden die Filter bei tiefen Mittenfrequenzen numerisch sehr instabil (s. a. Pflüger, 1997, S. 84). Eine Kaskade von m Zweipol-Filtern, Gl. (2.27), führt in den meisten Fällen zu einer numerisch stabileren Implementierung. Die Amplitudengänge zeitdiskreter linearer GTF unterschiedlicher Ordnung im Vergleich zu den auditiven Filterkurven nach Rosen und Baker (1994) sind in Abb. 2.12 dargestellt.

Reellwertige GTF m-ter Ordnung mit Anfangsphase $\phi = 0$ besitzen m identische konjugiert komplexe Polpaare bei $z_p = re^{\pm j\theta}$ und m reelle Nullstellen. Die Lage der Nullstellen auf der reellen Achse wird dabei durch die normierte Kreisfrequenz $\theta = \omega_c T$ und $r = e^{-bT}$ bestimmt. Wird die Anfangsphase der Trägerschwingung der Gammaton-Funktion, Gl. (2.23), zu $\phi = \pi/2$ gesetzt, verschiebt sich eine Nullstelle der Laplace-Transformierten ins Unendliche (siehe z. B. Slaney, 1993). Der sogenannte Sinus-Phasen GTF (Flanagan, 1960b; Lyon, 1996) besitzt somit nur $m - 1$ Nullstellen.

2.3.2 Lineare *All-Pole* Gammaton-Filter

Aufgrund ihres symmetrischen Amplitudengangs können lineare GTF die experimentalpsychologisch ermittelten auditiven Filterkurven vor allem zu hohen Frequenzen hin nicht mit ausreichender Genauigkeit abbilden (s. Abb. 2.12). Dies führte zur Entwicklung modifizierter Gammaton-Filter mit asymmetrischen Amplitudengängen. Slaney (1993) und Lyon (1996) schlagen eine All-Pol Implemen-

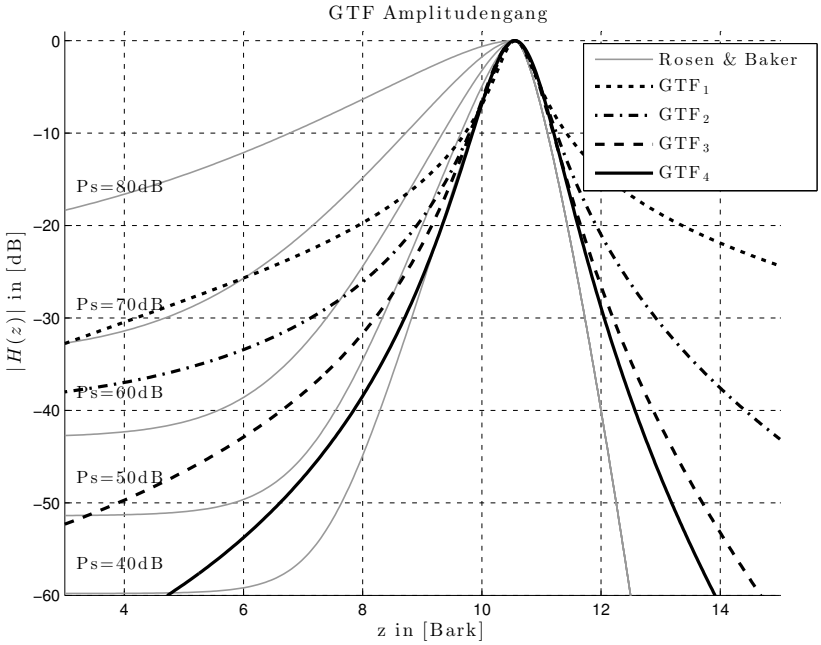


Abbildung 2.12: Amplitudengänge linearer Gammaton-Filter unterschiedlicher Ordnung im Vergleich zu den auditiven Filterkurven nach Rosen und Baker (1994) als Funktion des Signalpegels P_S (graue durchgezogene Linien) und einer Abtastfrequenz von 48 kHz.

tierung[34] durch Verschiebung der Nullstellen auf den Koordinatenursprung der z-Ebene vor. Nach Hohmann (2002) bleiben dabei sowohl die Einhüllende als auch die zeitliche Feinstruktur der GTF Impulsantwort erhalten. *All-Pole* Gammaton-Filter (APGF) der Ordnung m lassen sich als Kaskade von m Zweipol-Filtern realisieren (s. Anhang A.4):

$$G_{\mathrm{APGF},m}(z) = \frac{\hat{a}_{\mathrm{APGF},m}}{(1 - 2r\cos\theta\, z^{-1} + r^2 z^{-2})^m}\,, \tag{2.28}$$

wobei der Verstärkungsfaktor $\hat{a}_{\mathrm{APGF},m}$ wiederum der Skalierung der Amplitude bei der Mittenfrequenz auf 0 dB dient. Da die z-Transformierte lediglich m konjugiert komplexe Polpaare besitzt, lassen sich APGF sehr einfach und recheneffizient implementieren. Lyon (1996) zeigt, dass mit dem asymmetrischen APGF Ampli-

[34] Van Compernolle (1991, Kap. 2.4 und 2.5) erwähnte bereits vor Slaney die Möglichkeit Gammaton-Filter recheneffizient über All-Pol Filter anzunähern. Eine detaillierte mathematische Beschreibung der All-Pol Filterstrukturen wurde allerdings erst später von Slaney (1993, Kap. 3.5) veröffentlicht.

tudengang die experimentell ermittelten auditiven Filterformen besser angenähert werden als mit regulären GTF. Beim Filterentwurf ist jedoch auf die Abhängigkeit des Amplitudengangs von der Abtastfrequenz zu achten (s. a. Irino und Unoki, 1997). Abb. 2.13 zeigt die Amplitudengänge zeitdiskreter linearer APGF unterschiedlicher Ordnung im Vergleich zu den auditiven Filterkurven nach Rosen und Baker (1994).

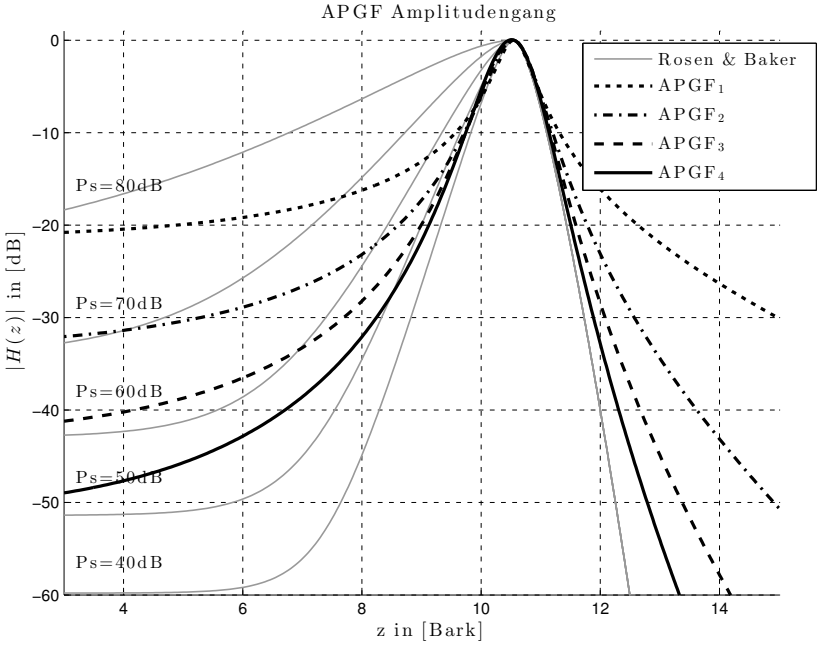


Abbildung 2.13: Amplitudengänge von APGF unterschiedlicher Ordnung im Vergleich zu den auditiven Filterkurven nach Rosen und Baker (1994) als Funktion des Signalpegels P_S (graue durchgezogene Linien) und einer Abtastfrequenz von 48 kHz.

APGF lassen sich sehr einfach parametrieren. Durch eine pegelabhängige Verschiebung der Polstellen kann somit das nichtlineare Verhalten der Cochlea angenähert werden (s. Slaney, 1993; Lyon, 1997; Pflüger et al., 1997). Die APGF Filterstruktur stellt zudem eine Verbindung zu den physiologisch motivierten Kaskadenmodellen der Cochlea, d. h. zur passiven Ausbreitung der Wanderwelle auf der Basilarmembran, her (s. Lyon und Mead, 1988a,b; Slaney und Lyon, 1993; Kates, 1993, 1995; Lyon, 1997).

Nach Lyon und Mead (1988b, Anhang 6) ergibt die Rücktransformation der APGF Übertragungsfunktion in den Zeitbereich eine Bessel-Funktion erster Art

mit halbzahligem Index[35] als Trägerschwingung, wie sie in frequenzmodulierten Signalen vorkommt (siehe z. B. Daniel, 1997, S. 247). Møller und Nilsson (1979)[36] zeigen ein frequenzmodulierendes Verhalten der Impulsantwort der Basilarmembran und der Nervenfasern. APGF sind somit wesentlich besser geeignet das Schwingungsverhalten der Basilarmembran anzunähern als reguläre GTF.

2.3.3 Lineare *One-Zero* Gammaton-Filter

Flanagan (1960b, Gl. 11) modelliert die Schwingung der Basilarmembran über einen APGF 3ter Ordnung mit einer zusätzlichen Nullstelle bei $s = 0$. Diese wurde von Lyon (1997) als Ableitung der APGF Übertragungsfunktion[37] im Laplacebereich (s. a. Katsiamis et al., 2006, 2007) interpretiert. Die sogenannten *Differentiated All-Pole* Gammaton-Filter (DAPGF) besitzen die gleichen Polpaare wie APGF, weisen aufgrund der zusätzlichen Nullstelle jedoch eine wesentlich steilere Filterflanke zu tiefen Frequenzen hin auf. Die Übertragungsfunktion berechnet sich zu (s. a. Anhang A.4):

$$G_{\mathrm{DAPGF},m}(z) = \frac{1 - z^{-1}}{\left(1 - 2r\cos\theta\, z^{-1} + r^2 z^{-2}\right)^m}\,. \tag{2.29}$$

One-Zero Gammaton-Filter (OZGF) unterscheiden sich von DAPGF dadurch, dass die zusätzliche Nullstelle beliebige Werte auf der reellen Achse annehmen kann. OZGF lassen sich sehr einfach aus GTF beliebiger Phase herleiten, indem alle bis auf eine Nullstelle der Übertragungsfunktion weggelassen werden. Sie weisen die gleiche Übertragungscharakteristik wie DAPGF auf. Ein wesentlicher Vorteil besteht jedoch darin, dass sich die Flankensteilheit zu tiefen Frequenzen hin über die Lage der Nullstelle steuern lässt. OZGF können als Verallgemeinerung von DAPGF interpretiert werden. In dieser Arbeit wird nicht zwischen DAPGF

[35] Die Bessel-Funktion erster Art mit halbzahligem Index steht in einem engen Zusammenhang mit der sphärischen Bessel-Funktion erster Art mit ganzzahligem Index (s. a. Abramowitz und Stegun, 1970, 10.1.1). Die entsprechenden Funktionen für ganzzahlige Ordnungen lassen sich relativ einfach über Rekursionsformeln bestimmen und implementieren (siehe z. B. Williams, 1999, Gl. 6.69).

[36] Siehe auch: Ruggero et al. (1997), de Boer und Nuttall (1997), Carney et al. (1999) und Shera (2001).

[37] Das Ableiten der Übertragungsfunktion im Laplacebereich erzeugt eine zusätzliche Nullstelle und entspricht im Zeitbereich einer Multiplikation der Impulsantwort mit t.

und OZGF unterschieden, sondern für beide Filtertypen der einheitliche Begriff OZGF verwendet. Die Berechnung der OZGF Filterparameter wird in Anhang A.4 zusammengefasst.

Abb. 2.14 zeigt die Amplitudengänge zeitdiskreter linearer OZGF unterschiedlicher Ordnung im Vergleich zu den auditiven Filterkurven nach Rosen und Baker (1994). Mit der steileren Flanke des Amplitudengangs zu tiefen Frequenzen hin lassen sich die auditiven Filterkurven wesentlich genauer annähern als mit GTF oder APGF (s. a. Katsiamis et al., 2006, Abb. 6).

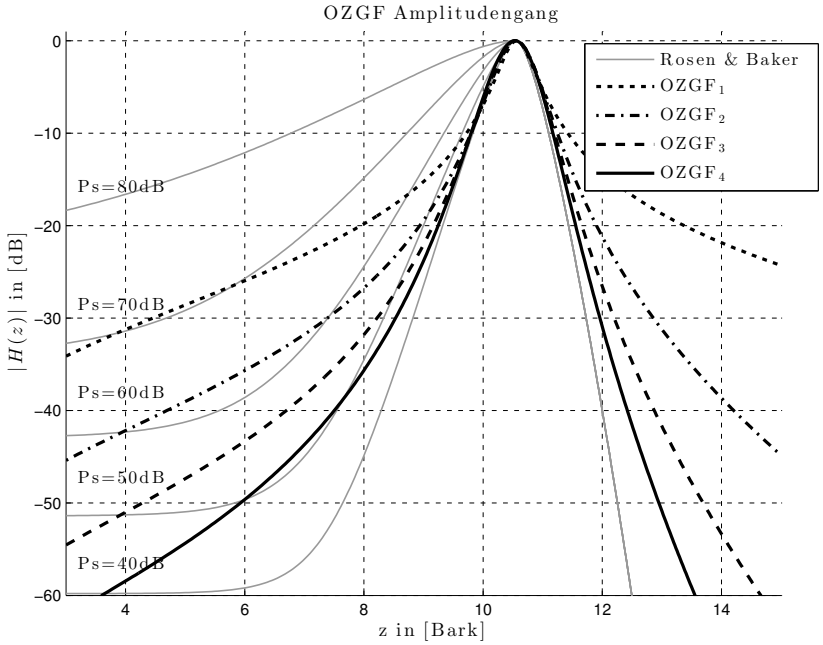


Abbildung 2.14: Amplitudengänge von OZGF (DAPGF) mit einer Nullstelle bei $z = 1$ und unterschiedlicher Ordnung im Vergleich zu den auditiven Filterkurven nach Rosen und Baker (1994) als Funktion des Signalpegels P_S (graue durchgezogene Linien) und einer Abtastfrequenz von 48 kHz.

OZGF sind aufgrund der zusätzlichen Nullstelle analytisch nicht so einfach handhabbar wie APGF, lassen sich jedoch als APGF der Ordnung $(m - 1)$ mit vorgeschaltetem Bandpassfilter realisieren.[38] Die Teilfilter können dabei als Zweipol-Filter ausgeführt werden und lassen sich sehr recheneffizient als digitale oder analoge integrierte Schaltungen implementieren (s. Katsiamis et al., 2006).

[38] Zur Steigerung der Recheneffizienz werden APGF der Ordnung m typischerweise als Kaskade m identischer Tiefpassfilter implementiert (s. a. Kap. 2.3.2).

2.3.4 Lineare *Three-Zero* Gammaton-Filter

Aus den Abbildungen 2.13 und 2.14 ist ersichtlich, dass die APGF/OZGF Sperrdämpfung nicht ausreicht um die auditiven Filterkurven zu hohen Frequenzen hin mit ausreichender Genauigkeit zu modellieren. Lin et al. (2001a, 2002) schlagen zur Erhöhung der Flankensteilheit die Verwendung eines zusätzlichen Kerbfilters 2ter Ordnung knapp über der Filtermittenfrequenz f_c eines OZGF vor. Dadurch ergeben sich zwei zusätzliche Nullstellen. Die Übertragungsfunktion eines sogenannten *Three-Zero* Gammaton-Filters (TZGF) m-ter Ordnung berechnet sich nach Lin et al. (2001a, Gl. 3.1) zu

$$G_{\mathrm{TGZF},m}(z) = \left(1 - r_0 z^{-1}\right) \frac{1 - 2r_1 \cos(\theta_1) z^{-1} + r_1^2 z^{-2}}{\left(1 - 2r_2 \cos(\theta_2) z^{-1} + r_2^2 z^{-2}\right)^m} \,, \tag{2.30}$$

wobei die Nullstellenradien $r_0 = 0.995$ und $r_1 = 0.985$ empirisch ermittelt wurden. Die Mittenfrequenz des Kerbfilters $f_{c_1} > f_c$ entspricht jener Frequenz, bei der die Amplitude des OZGF einen Wert von ungefähr $-60\,\mathrm{dB}$ im Vergleich zur Maximalamplitude annimmt (s. Lin et al., 2001a, Gl. 3.4):

$$f_{c_1} = 117{,}5\,(f_c/1000)^2 + 1135{,}5\,(f_c/1000)^2 + 277\,. \tag{2.31}$$

Die Kreisfrequenzen berechnen sich zu $\theta_1 = 2\pi f_{c_1}/f_s$ bzw. $\theta_2 = 2\pi f_c/f_s$. Die Lage der Polstellen (r_2, θ_2) berechnet sich gleich wie für gewöhnliche APGF/OZGF.

Abb. 2.15 zeigt die Amplitudengänge zeitdiskreter linearer TZGF unterschiedlicher Ordnung im Vergleich zu den auditiven Filterkurven nach Rosen und Baker (1994). Mit der steileren Flanke des Amplitudengangs zu hohen Frequenzen hin lassen sich die auditiven Filterkurven wesentlich genauer annähern als mit OZGF oder APGF. Nach Lin et al. (2001a) wird mit TZGF 4ter Ordnung die beste Übereinstimmung erreicht.

2.3.5 Nichtlineare Gammaton-Filter

Es ist allgemein akzeptiert, dass die Teilbänder einer parallelen linearen Gammaton-Filterbank (GTFB) der Zeit-Frequenzdarstellung auf der Basilarmembran entsprechen (s. Kap. 3). Dies trifft allerdings nur bei relativ moderaten

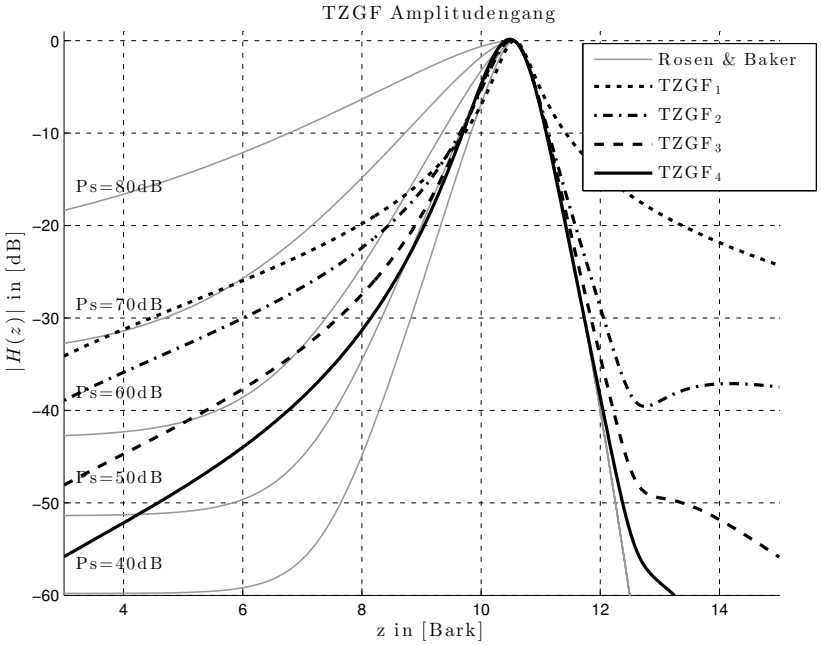


Abbildung 2.15: Amplitudengänge von TZGF unterschiedlicher Ordnung im Vergleich zu den auditiven Filterkurven nach Rosen und Baker (1994) als Funktion des Signalpegels P_S (graue durchgezogene Linien) und einer Abtastfrequenz von 48 kHz.

Signalpegeln zu. Bei sehr geringen Signalpegeln wird die Frequenzselektivität des menschlichen Gehörs stark durch die aktiven Vorgänge in der Cochlea beeinflusst, die zu einer Entdämpfung der Basilarmembranbewegung führen (s. Kap. 2.1.3). Die Resonanzamplitude erreicht bis zu 60 dB und mehr (siehe z. B. Katsiamis et al., 2006, Abb. 3). Mit steigenden Signalpegeln werden die auditiven Filterkurven zunehmend unsymmetrisch und breiter (siehe z. B. Lutfi und Patterson, 1984; Glasberg und Moore, 1990; Rosen und Baker, 1994; Rosen et al., 1998), die charakteristische Frequenz (CF)[39] nimmt mit zunehmender Schallintensität ab (siehe z. B. McFadden und Yama, 1983; Ruggero et al., 1997). Die Studien von Oxenham und Plack (1997) und Plack et al. (2002) zeigen zudem ein stark komprimierendes Verhalten der auditiven Filter.

[39]Die charakteristische Frequenz (CF) oder auch Bestfrequenz (BF) bezeichnet die besonders hohe Empfindlichkeit einer Faser des Hörnervs bei einer ganz bestimmten Frequenz. Bei der CF kann bereits bei sehr geringer Schallintensität eine über der Spontanaktivität liegende Feuerungsrate der Faser festgestellt werden. Umliegende Fasern werden hingegen nur bei höherer Schallintensität mit angeregt (siehe z. B. Møller, 2006, Kap. 3.2).

Lineare Gammaton-Filter können diese Nichtlinearitäten nicht nachbilden. Dies führte zur Entwicklung von nichtlinearen Gammaton-Filtern, bei denen zum Beispiel die Bandbreite (s. Carney, 1993) oder die Lage der Polstellen (s. Pflüger, 1997; Pflüger et al., 1997) dem Signalpegel angepasst werden. Lyon zeigt, dass sich Bandbreite und Mittenfrequenz von APGF und OZGF mit einem einzigen Parameter (der Güte Q) an das nichtlineare Verhalten der Cochlea anpassen lassen (s. Lyon, 1996; Pflüger, 1997, S. 86; Katsiamis et al., 2006).[40] Bei den sogenannten *Passive Gammachirp Filter* (pGCF) wird die Trägerschwingung der Gammaton-Funktion frequenzmoduliert[41] (vgl. Irino und Patterson, 1997; Irino und Unoki, 1999). Eine Weiterentwicklung sind die *Compressive Gammachirp Filter* (cGCF, vgl. Irino und Patterson, 2001; Patterson et al., 2003) und *Dynamic Compressive Gammachirp Filter* (dcGCF, vgl. Irino und Patterson, 2006a), bei denen ein pGCF mit asymmetrischen Hochpassfiltern gekoppelt wird, deren Mittenfrequenzen mit steigendem Signalpegel zunehmen. Dadurch kann eine sehr gute Übereinstimmung mit den Nichtlinearitäten der cochleären Verarbeitung erreicht werden.

Nichtlineare Gammaton-Filter benötigen einen vergleichsweise hohen Rechenaufwand. Aus diesem Grund werden in dieser Arbeit ausschließlich APGF, OZGF und TZGF als Teilfilter einer Analyse-Synthese-Filterbank zur auditiven Signalverarbeitung verwendet. Diese lassen sich sehr recheneffizient implementieren (siehe z. B. Slaney, 1993; Lyon, 1996; Lin et al., 2001a). Zur weiterführenden Studie nichtlinearer Gammaton-Filter sei auf die Literatur verwiesen.

[40] Siehe auch: All-Pol Gammaton-Filter Approximation in Slaney (1993, Kap. 3.5) und OZGF-Modellierung der Schwingung der Basilarmembran in Flanagan (1960b, Gl. 11).

[41] Wie in Møller und Nilsson (1979), Ruggero et al. (1997), de Boer und Nuttall (1997), Carney et al. (1999) und Shera (2001) gezeigt, sind die Impulsantworten der Basilarmembran und Nervenfasern frequenzmoduliert. Gammachirp-Filter lassen sich demzufolge gut auf das Schwingungsverhalten der Basilarmembran abstimmen (s. Irino und Patterson, 2001).

3

Analyse-Synthese Filterbank zur auditiven Signalverarbeitung

Dieses Kapitel befasst sich mit der Frage der recheneffizienten Implementierung von Gammaton-Filterbänken mit möglichst kurzer Systemlatenz. Es wird gezeigt, dass aus der Betrachtung der Phase an den Übergängen benachbarter Teilbänder ein sehr einfaches Verfahren[42] hergeleitet werden kann, welches eine nahezu perfekte Rekonstruktion des Signals durch einfache Summation der Teilbandsignale ermöglicht.

Gammaton-Filterbänke (GTFB) stellen ein rein funktionales Modell der Frequenzauflösung der cochleären Verarbeitung des menschlichen Gehörs dar (siehe z. B. Pflüger, 1997, S. 80). Es ist allgemein akzeptiert, dass die Teilbänder einer parallelen linearen GTFB der Zeit-Frequenzdarstellung auf der Basilarmembran entsprechen (vgl. Patterson und Moore, 1986; Patterson et al., 1987, 1992; Dau et al., 1996a,b; Lopez-Poveda und Meddis, 2001; Cooke, 2006; Katsiamis et al., 2007). Die Leistungen der Teilbandsignale werden dabei als Erregungsverteilung entlang der Basilarmembran interpretiert (siehe z. B. Pflüger, 1997, Kap. 2.7).

[42] Die in diesem Kapitel vorgestellte Methode zum Entwurf einer Analyse-Synthese-Filterbank zur auditiven Signalverarbeitung mit möglichst kurzer Systemlatenz wurde im Rahmen eines Kooperationsprojekts des Instituts für Elektronische Musik und Akustik (IEM) der Universität für Musik und darstellende Kunst Graz mit der AKG Acoustics GmbH Wien entwickelt und ist patentrechtlich geschützt (s. a. Zotter, 2004, Noisternig et al., 2009 bzw. Anhang D).

Allgemein betrachtet, können die Teilfilter einer Gammaton-Filterbank als Atome einer linearen Zeit-Frequenz-Transformation aufgefasst werden (siehe z. B. Hut et al., 2006; Strahl und Mertins, 2009; Venkitaraman et al., 2014). Es sei L^2 der Hilbertraum quadratintegrierbarer Funktionen, versehen mit einem kanonischen Skalarprodukt. Eine lineare Zeit-Frequenz-Transformation

$$Tf(\gamma) = \langle f, \phi_\gamma \rangle = \int_{\mathbb{R}} f(t)\, \overline{\phi_\gamma(t)} dt \tag{3.1}$$

stellt eine Beziehung zwischen einer Funktion $f(t) \in L^2(\mathbb{R})$ und einer Familie $\mathcal{D} = \{\phi_\gamma\}_{\gamma\in\Gamma}$ von Funktionen ϕ_γ, den sogenannten Zeit-Frequenz-Atomen, her. Dabei wird angenommen, dass $\phi_\gamma \in L^2(\mathbb{R})$ und $\|\phi_\gamma\| = 1$ (s. Mallat, 2009, Kap. 4.1). Der Querstrich $\overline{(\cdot)}$ bezeichnet die komplexe Konjugation. Mit dem Satz von Parseval kann gezeigt werden, dass die Energie des Signals bei der Fouriertransformation erhalten bleibt (siehe z. B. Slepian und Pollak, 2013):

$$\int_{\mathbb{R}} f(t)\, \overline{\phi_\gamma(t)} dt = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(\omega)\, \overline{\hat{\phi}_\gamma(\omega)} d\omega\,. \tag{3.2}$$

Dabei bezeichnen $\hat{f}(\omega)$ und $\hat{\phi}_\gamma(\omega)$ die Fouriertransformierten der Funktionen $f(t)$ und $\phi_\gamma(t)$. Die Ausdehnung eines Atoms ϕ_γ in der Zeit-Frequenz-Ebene lässt sich gemäß der Heisenbergschen Unschärferelation über die zugehörige Heisenberg-Box bestimmen (vgl. Küpfmüller, 1924; Heisenberg, 1927; Gabor, 1946a,b,c). Diese beschreibt ein um einen Punkt (u, ξ) der Zeit-Frequenz-Ebene zentriertes Rechteck mit den Seitenlängen σ_t entlang der Zeitachse und σ_ω entlang der Frequenzachse (s. Abb. 3.1). Dabei ist u der Erwartungswert von $|\phi(t)|^2$ im Zeitbereich, ξ der Erwartungswert von $|\hat{\phi}(\omega)|^2$ im Frequenzbereich, $\sigma_t^2(\phi_\gamma) = \sigma(\phi_\gamma)^2/||\phi_\gamma||^2$ die Varianz im Zeitbereich (auch effektive Länge) und $\sigma_\omega^2(\phi_\gamma) = \sigma(\hat{\phi_\gamma})^2/||\hat{\phi_\gamma}||^2$ die Varianz im Frequenzbereich (auch effektive Bandbreite).

Mit dem Produkt der prinzipiellen Unschärfen der Zeit, σ_t, und der Frequenz, σ_ω, kann bei gleichzeitiger Messung beider Größen abgeschätzt werden, wie stark eine Analyse-Synthesefunktion in Zeit und Frequenz streut. Man spricht allgemein auch von der Zeit-Frequenz-Unschärfe der Transformation. Mit der Küpfmüllerschen Unbestimmtheitsrelation (s. Küpfmüller, 1924) gilt für das Produkt der prinzipiellen Unschärfen eines Zeit-Frequenz-Atoms (s. a. Kennard

1927; Gabor 1946a, Gl. 1.26; Mallat 2009, Theorem 2.6) folgende Schranke:

$$\sigma_t \sigma_\omega \geq 0{,}5\,. \tag{3.3}$$

Das Produkt entspricht der Fläche der zugehörigen Heisenberg-Box. Des Weiteren folgt aus Gl. (3.3), dass die prinzipiellen Unschärfen der Zeit und Frequenz nicht gleichzeitig beliebig klein werden können. Die untere Schranke von $\sigma_t \sigma_\omega = 0{,}5$ wird dabei ausschließlich für Gaußfunktionen erreicht. Umgekehrt gilt: Ein Zeit-Frequenz-Atom minimaler Unschärfe stellt sowohl im Zeit- als auch im Frequenzbereich eine Gaußkurve dar.

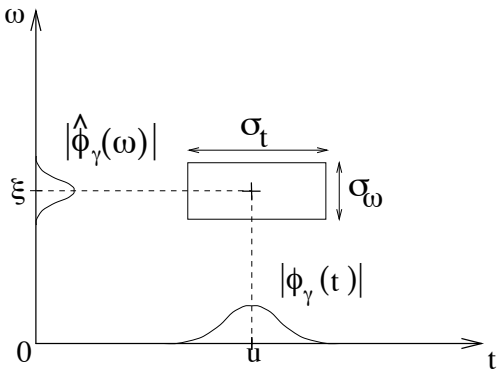


Abbildung 3.1: Heisenberg-Box für ein Zeit-Frequenz-Atom (s. Mallat, 2009, Abb. 4.1)

Nach Hut et al. (2006) kann mit einer linearen GTFB 4ter Ordnung die untere theoretische Schranke der Zeit-Frequenz-Unschärfe annähernd erreicht werden. Die Parameter der Gammaton-Filter wurden dabei so gewählt, dass diese die Impulsantworten des Cochlea-Modells von Diependaal et al. (1987) möglichst gut annähern (vgl. Hut et al., 2006, Gl. 34). Daraus resultieren Gammaton-Filter mit einem von der Frequenz abhängigen Bandbreitenparameter von $b \approx 2\pi \times 0{,}4 \times ERB$ (vgl. Hut et al., 2006, Abb. 8). Die Gammaton-Filter sind also wesentlich schmäler als die von Patterson et al. (1992) mit $b = 2\pi \times 1{,}019 \times ERB$ vorgeschlagenen Filter. Mit der GTFB nach Hut et al. wird eine über den gesamten Hörfrequenzbereich annähernd konstante Zeit-Frequenz-Unschärfe von $\sigma_t \sigma_\omega \approx 0{,}6$ erreicht (vgl. Hut et al., 2006, Abb. 11). Allerdings weicht diese, trotz guter Übereinstimmung der Impulsantworten, sehr stark von der Zeit-Frequenz-Unschärfe des Diependaal-Cochlea-Modells ab, die zu höheren Mittenfrequenzen

hin stark ansteigt. Hut et al. schlussfolgern daraus, dass sich gewöhnliche lineare GTFB zwar sehr gut zur allgemeinen Zeit-Frequenz-Analyse von Audiosignalen eignen, jedoch kein hinreichendes Modell für die auditive Verarbeitung in der Cochlea darstellen. Dies stimmt einerseits mit den Studien von Strahl und Mertins (2009) überein, die mit Hilfe der Frame-Theorie zeigen, dass sich Gammaton-Filter mit $b = 2\pi \times 0{,}5 \times ERB$ sehr gut zur allgemeinen Zeit-Frequenz-Analyse eignen (s. a. Abschnitt 3.1). Andererseits zeigen viele Studien (siehe z. B. Patterson et al., 1987, 1992; Dau et al., 1996a,b), dass durch geeignete Wahl des Bandbreitenparameters mit einer GTFB 4ter Ordnung die psychophysikalischen Abstimmkurven mit ausreichender Genauigkeit modelliert werden können, sofern das nichtlineare und aktive Verhalten der Cochlea im Applikationsentwurf nur eine untergeordnete Rolle spielt (s. a. Kap. 2.3). In Hinblick auf ein möglichst einfach und recheneffizient zu implementierendes Modell der Frequenzselektivität und der Maskierungseigenschaften des menschlichen Gehörs, wird in dieser Arbeit eine parallele Bank linearer Gammaton-Filter (s. a. Kap. 2.3.1) zur Signalanalyse verwendet (s. a. Parallel-Filterbankmodell, Kap. 2.2.5-B). Die Nichtlinearität der auditiven Filter wird nicht berücksichtigt. Das folgende Kapitel befasst sich mit der digitalen Implementierung von Gammaton-Filterbänken und basiert vorwiegend auf den Arbeiten von Pflüger (1997, Kap. 6), Zotter (2004, Kap. 3.2) und Noisternig et al. (2009).

3.1 Gammaton-Filterbank-Implementierung

Erste ausführliche Studien zu Gammaton-Filterbänken und deren Implementierung wurden von Patterson et al. (1987, 1988), Patterson und Rice (1987) und Holdsworth et al. (1988) durchgeführt. Dabei wurde auch der Versuch unternommen GTFB zur auditiven Signalverarbeitung zu standardisieren. Implementierungen nichtlinearer und asymmetrischer GTFB finden sich beispielsweise in den Arbeiten von Carney und Yin (1988), Carney (1993), Pflüger et al. (1997) und Lin et al. (2001a, 2002). Die Anpassung an das pegelabhängige Verhalten der auditiven Abstimmkurven erfolgt hierbei über eine nichtlineare Parametrisierung der Filter. Irino und Unoki (1997, 1998) schlagen zur Modellierung des nichtlinearen Verhaltens die Verwendung von Gammachirp-Filterbänken (GCFB) vor. Diese können

sehr recheneffizient über IIR Filterstrukturen implementiert werden (s. a. Irino und Unoki, 1999, 2001). Komprimierende auditive Filterbänke werden typischerweise als eine Bank paralleler (dc)GCF[43] ausgeführt. Implementierungsvorschläge sind in den Arbeiten von Irino und Patterson (2006a,b) und Unoki et al. (2006) zu finden. Eine umfassende Übersicht über unterschiedliche Methoden zur Implementierung von Gammaton-Filterbänken wird beispielsweise in Lopez-Poveda und Meddis (2001) und Duifhuis (2012) gegeben.

Perfekte Rekonstruktion

Eine der wichtigsten Eigenschaften einer Analyse-Synthese-Filterbank ist die (nahezu) perfekte Rekonstruktion des Eingangssignals am Ausgang der Filterbank. Das heißt, findet keine Verarbeitung der Teilbandsignale statt, kann das breitbandige Originalsignal bis auf eine Verzögerung vollständig aus den Teilbandsignalen rekonstruiert werden. Die Frage der Rekonstruierbarkeit kann dabei allerdings nur unter Berücksichtigung der gewählten Analysefilter beantwortet werden. Dies führt zum Konzept der Synthesefilterung. Im einfachsten Fall der Synthesefilterung wird das Originalsignal durch Summation der Teilbandsignale rekonstruiert. Bei Gammaton-Filtern ist dabei jedoch die mit abnehmender Mittenfrequenz zunehmende Gruppenlaufzeit zu beachten. Dadurch entstehen Phasendifferenzen an den Übergängen benachbarter Teilbänder und es kann bei der Summation zu erheblichen Signalverzerrungen bzw. Signalauslöschungen kommen. Hohmann (2002) kompensiert die unterschiedlichen Signallaufzeiten bevor die Teilbandsignale summiert werden. Dadurch reduziert sich die Welligkeit am Ausgang des Analyse-Synthese-Systems. Eine Weiterführung dieses Ansatzes ist in Zotter (2004), Herzke und Hohmann (2007) und Noisternig et al. (2009) zu finden. In diesen Arbeiten wird die Phase an den Übergängen benachbarter Kanäle berücksichtigt und vor der Summation entzerrt. Kubin und Kleijn (1999a,b) schlagen für die Synthese die Verwendung zeitgespiegelter Gammaton-Filter vor. Um die Kausalität der Synthesefilter zu gewährleisten, ist hierbei eine zusätzliche zeitliche Verzögerung des Signals erforderlich. Die Gesamtverzögerung des Signals entspricht dabei oft nicht den Anforderungen von Echtzeitsystemen. Wie in Irino und Unoki (1998,

[43] *Dynamic Compressive Gammachirp Filter* (s. a. Abschnitt 2.3.5).

2001) gezeigt, lässt sich die Zeitspiegelung auch auf pegelabhängige asymmetrische Gammachirp-Filterbänke anwenden. Bei der Resynthese muss hierbei allerdings der variierende Pegel kompensiert werden. Lin et al. (2001b) leiten über den Ansatz der linearen prädiktiven Entfaltung[44] ein Set von optimalen Resynthesefiltern her. Diese lassen sich recheneffizient und mit einer vergleichsweise geringen Systemlatenz implementieren. Dies ist ein wesentlicher Vorteil gegenüber den zeitgespiegelten Gammaton-Filtern. Slaney et al. (1994) berechnen die Synthesefilter mit Hilfe konvexer Projektion.[45] Feldbauer et al. (2005) und Strahl und Mertins (2009) zeigen, wie sich mit Hilfe der Frame-Theorie[46] Gammaton-Filterbänke mit nahezu perfekter Rekonstruktion herleiten lassen. Um hierbei eine hohe Genauigkeit der Signalrekonstruktion zu erreichen, wird jedoch eine relativ hohe Anzahl sich überlappender Teilbändern benötigt. Nach Strahl und Mertins ergibt sich bei $L = 50$ über einen Frequenzbereich von $[40\,\text{Hz}, 17\,\text{kHz}]$ verteilten Gammaton-Filtern der Ordnung $m = 4$ und $b = 2\pi \times 1{,}019 \times ERB$ ein snug-Frame[47] mit

[44] Bei der prädiktiven Entfaltung (*predictive deconvolution, spiking filtering*) nach Peacock und Treitel (1969) wird das inverse Filter über einen Wiener-Levinson-Ansatz hergeleitet. Die Impulsantwort des zu invertierenden Filters bildet das Eingangssignal eines Wiener-Filters, wobei am Ausgang ein Impuls angenommen wird (siehe z. B. auch Havelock et al., 2008, Kap. 87).

[45] Siehe z. B. Sezan und Stark (1982a,b), Youla und Webb (1982) und Mallat (1989).

[46] Siehe z. B. Christensen (2003, 2008), Mallat (2009) und Vetterli et al. (2011) für eine Einführung in die Frame-Theorie.

[47] Sei V ein Vektorraum über $\mathbb{R}$ und Γ eine abzählbare Indexmenge. Eine Familie $\{\phi_n\}_{n\in\Gamma}$ von Funktionen aus $L^2(\mathbb{R})$ ist dann ein Frame des Vektorraums V, wenn es Konstanten $B \geq A > 0$ gibt, sodass $\forall f \in V$:

$$A \left\|f\right\|^2 \leq \sum_{n\in\Gamma} |\langle f, \phi_n\rangle|^2 \leq B \left\|f\right\|^2 .$$

Für jeden Frame $\{\phi_n\}$ existiert mindestens ein dualer Frame $\{\tilde{\phi}_n\}$, sodass jede Funktion aus V durch diese Frames dargestellt werden kann:

$$f = \sum_{n\in\Gamma} \langle f, \hat{\phi}_n\rangle \phi_n = \sum_{n\in\Gamma} \langle f, \phi_n\rangle \hat{\phi}_n.$$

Das Frameschranken-Verhältnis B/A kann dabei als Stabilitätskonstante aufgefasst werden. Es gibt an wie „gutartig“ ein dualer Frame sein kann. Je kleiner B/A ist, umso besser sind die numerischen Eigenschaften des dualen Frames. Ist $A = B$ lässt sich die Funktion f als lineare Superposition von ϕ_n darstellen. Man spricht in diesem Fall von einem *tight frame* (siehe z. B. Bolcskei et al., 1998). Für $A = B = 1$ ist der *tight frame* eine Orthonormalbasis und man erreicht eine perfekte Rekonstruktion. Existiert für ein Problem kein *tight frame* werden die Funktionen ϕ_n üblicherweise so gewählt, dass sich $B/A \approx 1$ ergibt. Man spricht dann von einem *snug frame*. Hierbei kann, trotz der Redundanz, die Ausgangsfunktion mittels linearer Superposition rekonstruiert werden. Die Energie bleibt dabei nahezu erhalten.

einem Frameschranken-Verhältnis von $B/A = 1{,}109$. Durch Anpassung der Filterordnung und Bandbreite ($m = 11$, $b = 2\pi \times 0{,}85 \times ERB$) verbessert sich das Frameschranken-Verhältnis des snug-Frames auf $B/A = 1{,}020$. Mit $L = 100$ über einen Frequenzbereich von $[60\,\mathrm{Hz}, 17\,\mathrm{kHz}]$ verteilten Gammaton-Filtern ($m = 11$, $b = 2\pi \times 0{,}5 \times ERB$) ergibt sich ein tight-Frame mit $B/A = 1{,}003$. In diesem Fall bilden die Gammaton-Filter eine Orthonormalbasis und erlauben eine perfekte Rekonstruktion des Eingangssignals. Allerdings weichen die Filter aufgrund der hohen Güte stark von den psychophysikalischen Abstimmkurven ab. Die Maskierungseigenschaften des menschlichen Gehörs werden nicht mehr ausreichend genau modelliert. Zudem erschwert der vergleichsweise hohe Rechenaufwand die Implementierung auf digitalen Signalprozessoren. Eine umfassende Übersicht über die Frameschranken-Verhältnisse typischer GTFB-Implementierungen ist in Strahl und Mertins (2009, Tab. 3) zu finden.

Diese Arbeit befasst sich mit der Frage der recheneffizienten Implementierung von Filterbänken zur auditiven Signalverarbeitung mit möglichst kurzer Systemlatenz. Abb. 3.2 zeigt den schematischen Aufbau einer auditiven Analyse-Synthese-Filterbank am Beispiel der breitbandigen Signalaufbereitung. Zur Signalanalyse wird eine parallele Bank sich überlappender Gammaton-Filter verwendet. Das Synthesefilter rekonstruiert das Originalsignal durch Summation der Teilbandsignale. Um dabei die Welligkeit am Ausgang des Analyse-Synthese-Systems so gering wie möglich zu halten, wird ein Kriterium abgeleitet, welches die Notwendigkeit angibt, vor der Summation das Vorzeichen zu wechseln (s. Zotter, 2004; Noisternig et al., 2009). Im Gegensatz zur Methode von Herzke und Hohmann (2007) ist keine Drehung der Phase erforderlich. Anstatt der komplexwertigen Gammaton-Filter, können wesentlich recheneffizientere reellwertige Gammaton-Filter verwendet werden. Um das Transferverhalten des menschlichen Außen- und Mittelohres nachzubilden, wird das Eingangssignal vor der Analyse mit H_{AMF} gefiltert (s. Kap. 2.2.1). Mit dem inversen Filter H_{AMF}^{-1} am Ausgang des Analyse-Synthese-Systems, werden die durch die Außen-Mittelohr-Filterung hervorgerufenen Signalverzerrungen wieder rückgängig gemacht.[48] Die Vorteile des hier vorgestellten Analyse-Synthese-

[48] Die Inversion des Außen-Mittelohr-Filters ist unproblematisch. Um eine stabile Inversion zu gewährleisten, kann der Nullstellenradius in Gl. (2.8) mit $r = 0{,}995$ angenommen werden,

Systems zur auditiven Signalverarbeitung liegen in der signifikanten Reduktion der Systemlatenz und der rechnerischen Effizienz.

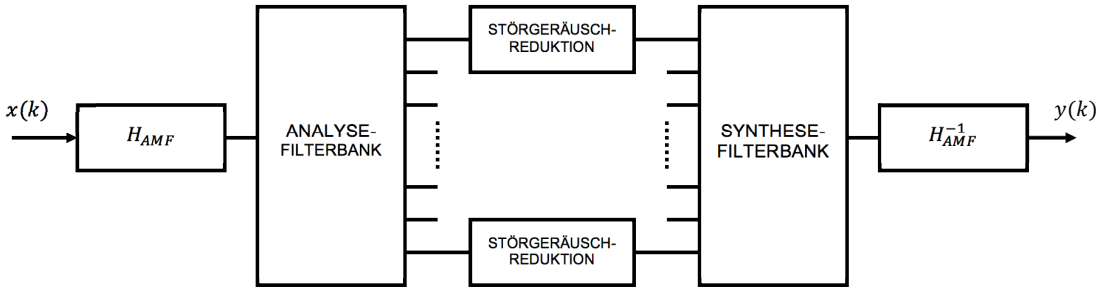


Abbildung 3.2: Schematische Darstellung einer auditiven Analyse-Synthese-Filterbank am Beispiel der breitbandigen Signalaufbereitung.

3.1.1 Analyse-Filterbank

Die Anzahl der Kanäle einer auditiven Analyse-Synthese-Filterbank ist physiologisch nicht vorgegeben und kann beliebig gewählt werden (siehe z. B. Glasberg und Moore, 1990; Agerkvist, 1994). Die Frequenzauflösung lässt sich somit durch die Anzahl der Teilbandfilter verändern. Eine stärkere Überlappung der Kanäle der Filterbank bedeutet allerdings auch einen wesentlich höheren Rechenleistungsbedarf bei der Implementierung. Mit frequenzgruppenbreiten Filtern kann der gesamte Hörfrequenzbereich bei begrenztem Aufwand erfasst werden. Die Aufteilung der Gammaton-Filter über den betrachteten Frequenzbereich folgt den Vorschlägen von Slaney (1993) und Pflüger (1997, Gl. 5.5).[49] Zuerst werden mit Gl. (2.18) die ERB-Raten für die niedrigste Filtermittenfrequenz $e(f_{\mathrm{c,min}})$ und dann für die halbe Abtastrate $e(f_s/2)$ bestimmt. Die ERB-Rate für die gesuchte Mittenfrequenz des ℓ-ten Teilfilters, $e_c(\ell)$, mit $\ell = 0, \ldots L-1$, ergibt sich aus der Teilung des

ohne dabei wesentliche Klangfärbungen des Signals zu erzeugen (s. a. Pflüger 1997, S. 91–94 und Zotter 2004, Kap. 3.1.1).

[49] Nach Pflüger (1997, S. 81 ff.) liefern beide Ansätze die exakt gleichen Ergebnisse. Der einzige Unterschied besteht darin, dass sich bei der mathematisch einfacheren Formulierung nach Pflüger die kleinste Filtermittenfrequenz bei $\ell = 0$ und nicht wie bei Slaney bei $\ell = L$ befindet.

ERB-Raten-Intervalls in L gleiche Teile:

$$e_c(\ell) = e(f_{\mathrm{c,min}}) + \ell \left(\frac{e(f_s/2) - e(f_{\mathrm{c,min}})}{L} \right) . \tag{3.4}$$

Über folgende Gleichung lässt sich sehr einfach die Filtermittenfrequenz (in Hz) des ℓ-ten Teilfilters bestimmen (s. Pflüger, 1997, Gl. 5.6):

$$f_c(\ell) = \frac{(0{,}00437 f_{\mathrm{c,min}} + 1) \left(\frac{0{,}00437 f_s/2 + 1}{0{,}00437 f_{\mathrm{c,min}} + 1} \right)^{\frac{\ell}{L}} - 1}{0{,}00437} . \tag{3.5}$$

Wie in Abschnitt 2.2.4 gezeigt, sind die ERB-Breiten etwas schmäler als die kritischen Bandbreiten (s. Abb. 2.8). Deshalb verwendet Zotter (2004, Kap. 3.2.1) die Tonheitsskala zur Aufteilung der Gammaton-Filter über den betrachteten Frequenzbereich. Die sich daraus ergebenden Teilfilter des Analyse-Synthese-Systems besitzen von allen gehörspezifischen Skalen die geringste Güte und können mit relativ geringer Ordnung implementiert werden. Dies wiederum steigert die Recheneffizienz. Zudem ergibt sich aufgrund der relativ hohen Bandbreiten eine möglichst kurze Systemlatenz. Allerdings lässt sich die mit diesem Ansatz gewonnene Analyse-Filterbank nicht invertieren und das Originalsignal nicht einfach rekonstruieren (siehe z. B. Strahl und Mertins, 2009). Der Entwurf einer geeigneten Synthese-Filterbank wird im folgenden Abschnitt beschrieben.

Die Abbildungen 3.3 bis 3.5 zeigen die Frequenzgänge und Gruppenlaufzeiten verschiedener Gammaton-Filterbänke zur auditiven Signalanalyse (siehe Teilbild a). Die Übertragungsfunktionen wurden für APGF und OZGF 3ter bzw. 4ter Ordnung simuliert. Mit der Tonheitsskala ergeben sich $L = 24$ gleichmäßig über den Hörfrequenzbereich (bis etwas oberhalb von 20 kHz) verteilte Teilbänder. Die Simulationsergebnisse werden in Abschnitt 3.2 ausführlicher diskutiert. Der Rechenleistungsbedarf unterschiedlicher Gammaton-Filterbänke (GTF, APGF, OZG, TZGF; vgl. Abschnitt 2.3) kann Tab. 3.1 entnommen werden.

Tabelle 3.1: Rechenleistungsbedarf einer Filterbank zur auditiven Signalanalyse mit L Kanälen und unterschiedlichen Gammaton-Filtern der Ordnung M (s. Zotter, 2004, Kap. 3.2).

	Additionen	**Multiplikationen**	**Speicherbedarf**	**Koeffizienten**
GTF	$3ML$	$4ML$	$2ML$	$(M+3)L$
APGF	$2ML$	$3ML$	$2ML$	$3L$
OZGF	$2ML+1$	$3ML$	$2ML+1$	$3L$
TZGF	$2(M+1)L+1$	$(3M+2)L$	$2ML+1$	$5L$

3.1.2 Synthese-Filterbank

Wie in den vorhergehenden Abschnitten beschrieben, kann im einfachsten Fall der Synthese das breitbandige Originalsignal durch Summation der Teilbandsignale rekonstruiert werden. Nimmt jedoch die Gruppenlaufzeit wie bei Gammaton-Filtern mit abnehmender Mittenfrequenz zu, kann es bei der Summation aufgrund von Phasendifferenzen an den Übergängen benachbarter Teilbändern zu Signalverzerrungen kommen. Hohmann (2002) kompensiert die unterschiedlichen Signallaufzeiten vor der Summation der Teilbänder. Es kommt zu einer Reduktion der Welligkeit am Ausgang des Analyse-Synthese-Systems. Die zusätzliche Verzögerung des Ausgangssignals steht jedoch im Gegensatz zu der in dieser Arbeit gestellten Forderung nach möglichst kurzer Systemlatenz. In Anhang A.5.3 wird ein Kriterium hergeleitet, welches die Notwendigkeit bestimmt, das Vorzeichen bei der Summation aufeinanderfolgender Teilbänder einer Gammaton-Filterbank (H_k, H_{k+1}) zu wechseln. Der Vorzeichenwechsel begünstigt die konstruktive Überlagerung der Signale an den Bandgrenzen und reduziert die Welligkeit am Ausgang der Synthese-Filterbank. Im Gegensatz zum Ansatz von Hohmann kommt es hierbei zu keiner zusätzlichen Verzögerung des Ausgangssignals. Zur Vereinfachung wird bei der Herleitung des Kriteriums die Annahme getroffen, dass die Amplitude A an der Bandgrenze hauptsächlich von der Distanz zur nächstgelegenen Polstelle abhängt (vgl. Van Compernolle, 1991). Für Gammaton-Filter der Ordnung m kann das

Vorzeichen wie folgt berechnet werden:

$$f_{\mathrm{sgn}}\left(H_{k,k+1}(s)\right) = \begin{cases} 1 & \forall \quad 2m \cdot \arctan\left(\sqrt{\frac{1}{\sqrt[m]{A^2}} - 1}\right) (\mathrm{mod}\ \pi) < \frac{\pi}{2} \\ -1 & \forall \quad 2m \cdot \arctan\left(\sqrt{\frac{1}{\sqrt[m]{A^2}} - 1}\right) (\mathrm{mod}\ \pi) \geq \frac{\pi}{2} \end{cases} \tag{3.6}$$

Zur einfacheren Implementierung kann das Kriterium für den Vorzeichenwechsel durch Umformulieren der Nebenbedingungen in Gl. (3.6) auch in geschlossener Form dargestellt werden:

$$f_{\mathrm{sgn}}\left(H_{k,k+1}(s)\right) = \mathrm{sgn}\left\{\cos\left(2m \cdot \arctan\left(\sqrt{10^{-\frac{C_{\mathrm{dB}}}{10\ m}} - 1}\right)\right)\right\} . \tag{3.7}$$

Dabei wird ein Dämpfungsfaktor C_{dB} eingeführt, der das Verhältnis der Amplitude an der Bandgrenze zur Amplitude bei der Filtermittenfrequenz angibt. Zur Vereinfachung wird zur Berechnung des Dämpfungsfaktors die Amplitude bei der Mittenfrequenz auf 0 dB normiert. Hier soll noch einmal angemerkt werden, dass bei der Summation mit Vorzeichenwechsel nach Gl. (3.7) keine Kompensation der unterschiedlichen Signallaufzeiten durchgeführt werden soll.

3.2 Simulationsergebnisse

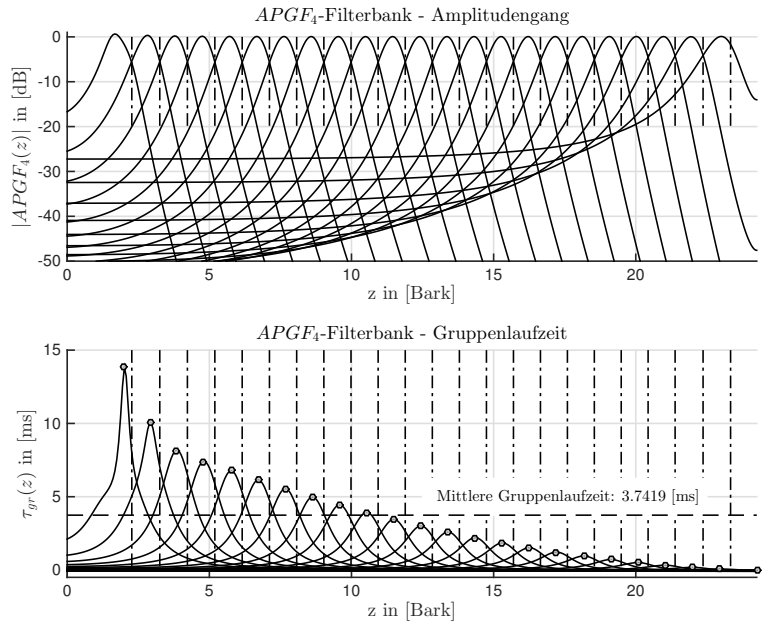
Die Abbildungen 3.3 bis 3.5 zeigen die Frequenzgänge und Gruppenlaufzeiten unterschiedlicher Implementierungen linearer Gammaton-Filterbänke zur auditiven Signalanalyse (Teilbild a), sowie die Welligkeit am Ausgang der zugehörigen Synthese-Filter (Teilbild b). Das Systemverhalten wird am Beispiel einer parallelen Bank von APGF 4ter Ordnung (Abb. 3.3) bzw. OZGF 4ter und 3ter Ordnung (Abb. 3.4 und 3.5) illustriert. Die Ergebnisse lassen sich qualitativ auch auf andere Arten von Gammaton-Filterbänken anwenden. Alle Simulationen wurden mit einer Abtastrate von $f_s = 48\,\mathrm{kHz}$ durchgeführt. Durch die hohe Abtastrate befinden sich viele Polstellen der Gammaton-Filter im ersten Quadranten der z-Ebene nahe dem Einheitskreis (s. a. Pflüger, 1997, S. 107). Um zu vermeiden, dass die Filter bei tiefen Mittenfrequenzen instabil werden, wurden alle Berechnungen mit doppelter Gleitkomma-Genauigkeit durchgeführt. Auf die Implementierung von GTFB auf digitalen Signalprozessoren mit Festkomma-Arithmetik wird in dieser Arbeit nicht

näher eingegangen. Zur Berechnung der Filtermittenfrequenzen wurde das Frequenzband von 20 Hz bis 20 kHz (Hörfrequenzbereich) mit Hilfe der Tonheitsskala in hörspezifische Frequenzgruppen aufgeteilt. Die Bandbreiten der Gammaton-Filter wurden dabei so gewählt, dass sich benachbarte Teilbänder bei $C_{\mathrm{dB}} = -6\,\mathrm{dB}$ (Abb. 3.3 und 3.4) bzw. $C_{\mathrm{dB}} = -4\,\mathrm{dB}$ (Abb. 3.5) überlappen. Die Notwendigkeit, bei der Summation das Vorzeichen zu wechseln, wurde mit Gl. (3.7) berechnet. Um die Reduktion der Welligkeit am Ausgang des Analyse-Synthese-Systems besser beurteilen zu können, wurde die Resynthese auf zwei Arten durchgeführt: (i) durch einfache Summation der Teilbandsignale (siehe Abb. 3.3 bis 3.5, Teilbild b, oben) und (ii) durch Summation der Teilbandsignale mit bedingtem Vorzeichenwechsel (siehe Abb. 3.3 bis 3.5, Teilbild b, unten). Bei der einfachen Summation ohne Vorzeichenwechsel erfolgte keine Kompensation der Signallaufzeiten.

Aus den Abbildungen ist klar ersichtlich, dass bei der Summation mit Vorzeichenwechsel wesentlich geringere Signalverzerrungen auftreten, als bei der direkten Summation der Teilbandsignale. Die Welligkeit am Ausgang der APGF-Filterbank 4ter Ordnung ist $< 5\,\mathrm{dB}$. Am Ausgang der OZGF-Filterbank 4ter Ordnung ergibt sich eine Welligkeit $< 3\,\mathrm{dB}$. Durch Reduktion der Filterordnung lässt sich die Recheneffizienz der auditiven Filterbank erhöhen. Allerdings reicht bei GTF und APGF 3ter Ordnung die Sperrdämpfung nicht aus, um die auditiven Filterkurven mit ausreichender Genauigkeit zu modellieren. Dem kann durch Verwendung von OZGF und TZGF entgegnet werden. Aufgrund der zusätzlichen Nullstelle, besitzen OZGF eine relativ steile Filterflanke zu tiefen Frequenzen hin. TZGF besitzen ein zusätzliches Kerbfilter, welches eine hohe Sperrdämpfung zu höheren Frequenzen hin ermöglicht (s. a. Kap. 2.3). OZGF und TZGF lassen sich zudem relativ recheneffizient mit rekursiven Filterstrukturen implementieren. Beim Applikationsentwurf ist zu beachten, dass eine OZGF-Filterbank der Ordnung M gleich viele Rechenoperationen benötigt wie eine TZGF-Filterbank der Ordnung $(M - 1)$. Ein weiterer Vorteil von Gammaton-Filtern geringerer Güte liegt in den kürzeren Gruppenlaufzeiten. Die Güte lässt sich (i) durch die Filterordnung und (ii) durch den Dämpfungsfaktor C_{dB} (d. h. durch die Amplitude an den Bandgrenzen) steuern. Dies ist vor allem dann von Bedeutung, wenn auditive Filterbänke in Echtzeitsystemen (wie zum Beispiel Freisprechanlagen in Telekommunikationssystemen) verwendet werden.

Abb. 3.5 zeigt den Frequenzgang, die Gruppenlaufzeit und die Welligkeit einer OZGF-Filterbank 3ter Ordnung mit $C_{\mathrm{dB}} = -4\,\mathrm{dB}$. Die Simulationsergebnisse zeigen auch hier eine klare Reduktion der Welligkeit am Ausgang des Analyse-Synthese-Systems, wenn bei der Resynthese eine Summation mit bedingtem Vorzeichenwechsel angewendet wird (s. a. Zotter, 2004; Noisternig et al., 2009).

a)



b)

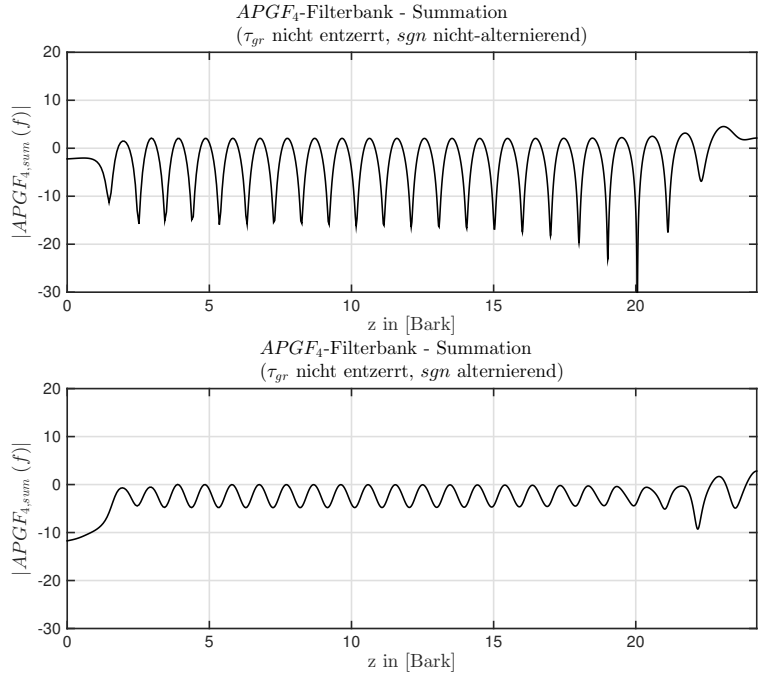
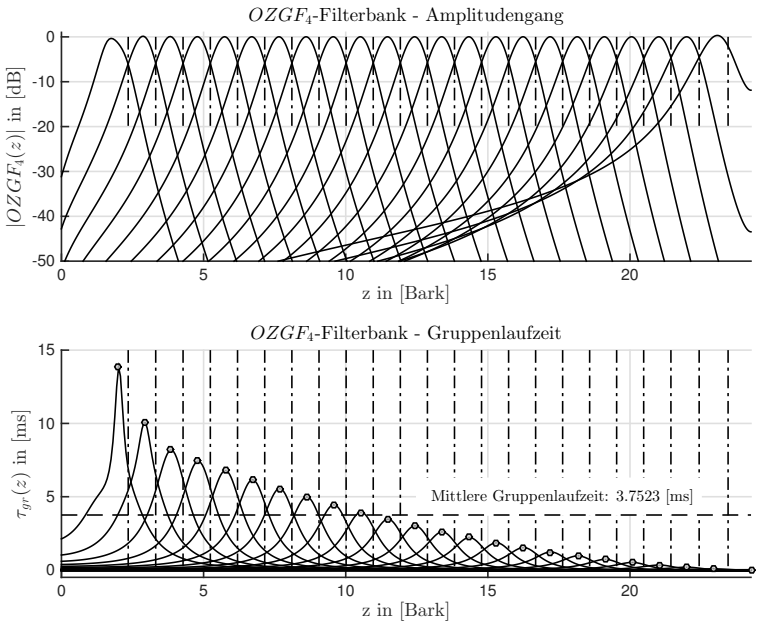


Abbildung 3.3: (a) Frequenzgang, Gruppenlaufzeit und (b) Welligkeit am Ausgang einer auditiven Analyse-Synthese-Filterbank. Die APGF 4ter Ordnung wurden mit der Tonheitsskala über den betrachteten Frequenzbereich verteilt. Die Synthese-Filterbank rekonstruiert das breitbandige Originalsignal durch einfache Summation der Teilbandsignale (Abb. b, oben) bzw. durch Summation mit Vorzeichenwechsel (Abb. b, unten). Die unterschiedlichen Gruppenlaufzeiten der Teilfilter wurden dabei nicht kompensiert.

a)



b)

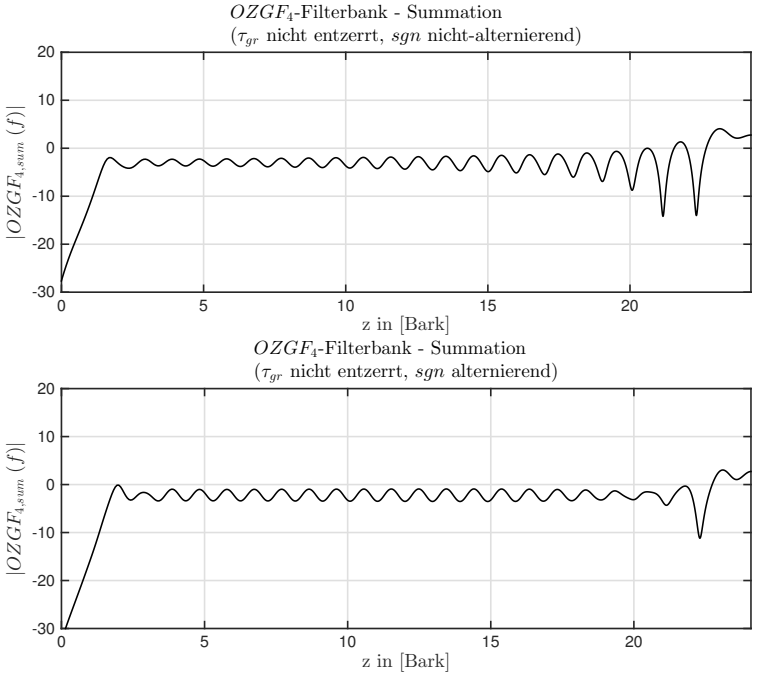
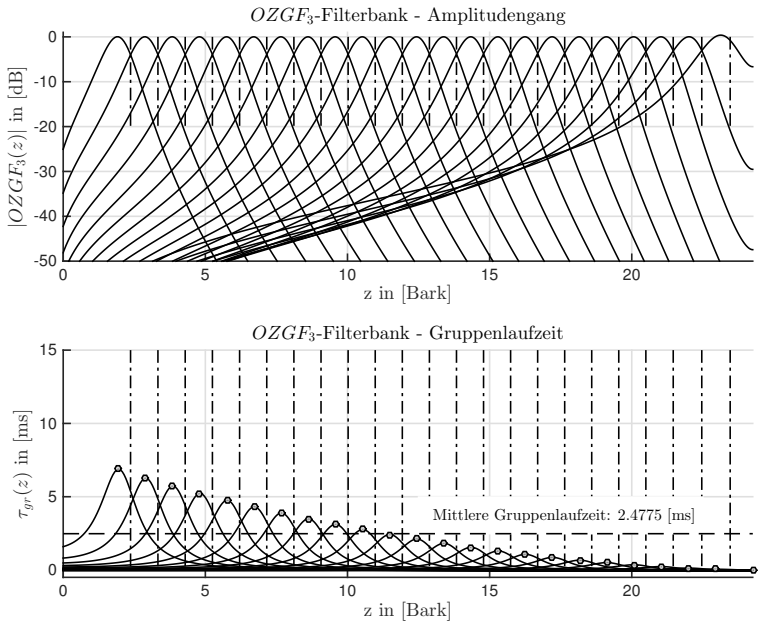


Abbildung 3.4: (a) Frequenzgang, Gruppenlaufzeit und (b) Welligkeit am Ausgang einer auditiven Analyse-Synthese-Filterbank. Die OZGF 4ter Ordnung wurden mit der Tonheitsskala über den betrachteten Frequenzbereich verteilt. Die Synthese-Filterbank rekonstruiert das breitbandige Originalsignal durch einfache Summation der Teilbandsignale (Abb. b, oben) bzw. durch Summation mit Vorzeichenwechsel (Abb. b, unten). Die unterschiedlichen Gruppenlaufzeiten der Teilfilter wurden dabei nicht kompensiert.

a)



b)

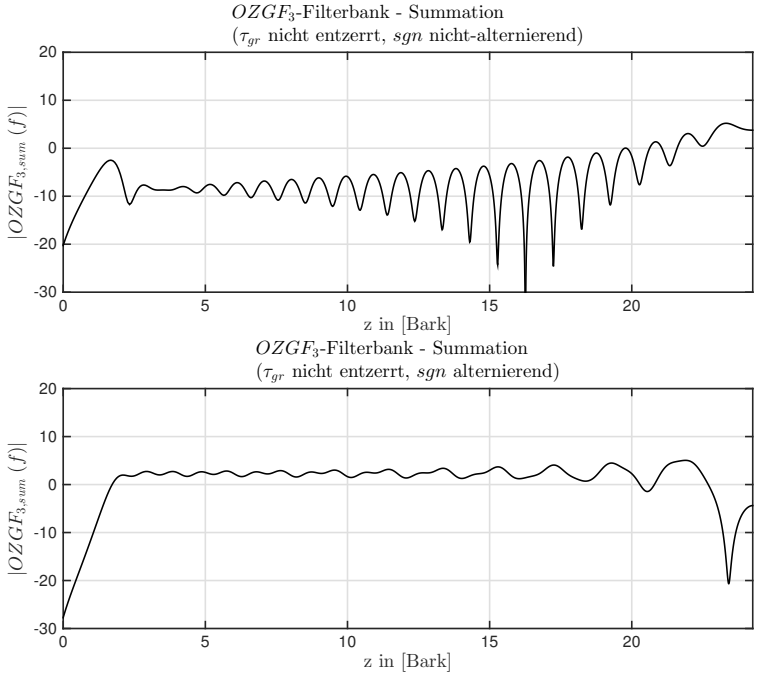


Abbildung 3.5: (a) Frequenzgang, Gruppenlaufzeit und (b) Welligkeit am Ausgang einer auditiven Analyse-Synthese-Filterbank. Die OZGF 3ter Ordnung wurden mit der Tonheitsskala über den betrachteten Frequenzbereich verteilt. Die Synthese-Filterbank rekonstruiert das breitbandige Originalsignal durch einfache Summation der Teilbandsignale (Abb. b, oben) bzw. durch Summation mit Vorzeichenwechsel (Abb. b, unten). Die unterschiedlichen Gruppenlaufzeiten der Teilfilter wurden dabei nicht kompensiert.

4

Breitbandige Signalaufbereitung in einkanaligen Mikrofonanwendungen

Dieses Kapitel diskutiert die Anwendung der spektralen Subtraktion auf die breitbandige Reduktion von Störgeräuschen in einkanaligen Mikrofonanwendungen. Das Hauptaugenmerk liegt auf den Gewichtungsregeln von Ephraim und Malah, die durch Schätzen der unterschiedlichen statistischen und spektralen Eigenschaften der Nutzsignal- und Störsignalkomponenten aus dem gestörten Gesamtsignal bestimmt werden. Die Berechnung wird unter Verwendung eines modifizierten entscheidungsgesteuerten Ansatzes dahingehend optimiert, dass die durch Fehler bei der Schätzung des zeitvarianten Störsignalspektrums auftretenden nichtlinearen Verzerrungen des Nutzsignals minimiert werden.[50]

Das primäre Ziel der breitbandigen Signalaufbereitung liegt in der weitgehenden Unterdrückung von Störgeräuschen und Interferenzsignalen,[51] bei möglichst geringer Verzerrung des Nutzsignals und gleichzeitiger Minimierung der tonalen Ar-

[50] Das in diesem Kapitel vorgestellte Verfahren zur breitbandigen Reduktion von Störgeräuschen in einkanaligen Mikrofonanwendungen wurde im Rahmen eines Kooperationsprojekts des Instituts für Elektronische Musik und Akustik (IEM) der Universität für Musik und darstellende Kunst Graz mit der AKG Acoustics GmbH Wien entwickelt und ist patentrechtlich geschützt (s. a. Zotter, 2004, Noisternig et al., 2009 und Anhang D).

[51] Das heißt, in der Maximierung des Signal-Stör-Verhältnisses (*signal-to-noise ratio*, SNR) bzw. des Signal-Interferenz-Stör-Verhältnisses (*signal-to-interference-plus-noise ratio,* SINR).

tefakte (siehe z. B. Vary et al., 1998, Kap. 12, Vaseghi, 2006, Kap. 11 und 16, Benesty et al., 2009). Das Gütekriterium ist dabei der Höreindruck, d. h. die subjektiv empfundene Qualität des Audiosignals. Prinzipiell kann zwischen einkanaligen und mehrkanaligen Geräuschreduktionsverfahren unterschieden werden:

1. Einkanalige Geräuschreduktionsverfahren beruhen auf der Schätzung der unterschiedlichen statistischen und spektralen Eigenschaften von Nutzsignal- und Störsignalkomponenten aus dem gestörten Gesamtsignal. Dieser Ansatz führt auf die klassischen Verfahren der Optimalfilterung, wie zum Beispiel

   - das Wiener-Filter (s. Wiener, 1942, 1949; Levinson, 1946),
   - die spektrale Subtraktion (s. Boll, 1979) und
   - das Kalman-Filter (s. Kalman, 1960; Kalman und Bucy, 1961).

2. Mehrkanalige Geräuschreduktionsverfahren nutzen, neben den statistischen und spektralen Eigenschaften der Signalkomponenten, auch die räumliche Trennung der Nutz- und Störschallquellen. Mehrkanalige Verfahren lassen sich ganz allgemein in folgende Kategorien einteilen:

   - Geräuschkompensation unter Verwendung eines vom Nutzsignal akustisch entkoppelten Sensors zur Aufnahme des Kompensationssignals in unmittelbarer Nähe der Störgeräuschquelle, wie z. B. die adaptive Geräuschkompensation (*adaptive noise cancellation*, ANC; s. a. Widrow et al., 1975).
   - Geräuschreduktion unter Verwendung einer Gruppe räumlich verteilter Sensoren. Typische Verfahren sind
     - das Beamforming (siehe z. B. Van Veen und Buckley, 1988; Brandstein und Ward, 2001; Van Trees, 2002; Benesty et al., 2008a),
     - das mehrkanalige Wiener-Filter (MWF; vgl. Chen et al., 2006; Cornelis et al., 2011; Yong et al., 2013) und
     - das MIMO[52] Wiener-Filter (MIMO-WF; siehe z. B. Huang et al., 2006, Kap. 3).

[52] *Multiple-input multiple-output* (MIMO).

Wiener Filter. Das klassische Wiener-Filter (Wiener, 1942, 1949; Levinson, 1946) schätzt ein durch additiv überlagertes Rauschen gestörtes Nutzsignal aus dem gestörten Gesamtsignal. Das Wiener-Filter ist in dem Sinne optimal, dass es den Mittelwert des quadratischen Fehlers der Schätzung minimiert (MMSE).[53] Bei der Herleitung des Wiener-Filters wird die Kenntnis der Statistik der Signale vorausgesetzt. Nutzsignal und Störung werden dabei als stationäre, mittelwertfreie Prozesse angenommen, die nicht miteinander korreliert sind. Es ergibt sich ein lineares, zeitinvariantes Optimalfilter (s. a. Kap. 4.1). Beim Wiener-Filter kann es zu erheblichen Verzerrungen des Nutzsignals kommen, die sich teils sehr störend auf den Höreindruck auswirken. Dies limitiert dessen Einsatz in praktischen Anwendungen, wie zum Beispiel ausführlich in Chen et al. (2006) diskutiert wird.

Nichtstationäre Umgebungen. In einer Umgebung, in der sich die Statistik der Signale zeitlich ändert, ist das Wiener-Filter nicht mehr optimal. Ist jedoch die Voraussetzung der Kurzzeitstationarität erfüllt, kann mit Hilfe blockweiser Verarbeitung der Signale eine optimale Lösung erreicht werden. Eine Möglichkeit das Schätzproblem für kurzzeitstationäre Signale zu lösen, ist der Einsatz adaptiver Filter. Diese erlernen die Statistik der beteiligten Signale selbständig und approximieren das Optimalfilter. Eine generelle Einführung in die Theorie adaptiver Filter ist zum Beispiel in Haykin (2002a) zu finden.

Kalman-Filter (Kalman, 1960; Kalman und Bucy, 1961) führen auf eine weitere Lösung des Schätzproblems in nichtstationären Umgebungen. Das Kalman-Filter ist eine Erweiterung der Wiener-Filter-Theorie und schätzt die inneren Zustände eines linearen zeitinvarianten Systems. Zur Schätzung der Systemzustände wird ein Systemmodell benötigt, mit dem die Signale am Ausgang des Systems bewertet werden.[54] Hierbei ist zu beachten, dass der numerische Aufwand kubisch mit der Ordnung des Systemmodells wächst (siehe z. B. Kaps, 2008, Kap. 3.5). Das Kalman-Filter liefert, ausgehend vom gestörten Gesamtsignal, einen Schätzwert für das ungestörte Nutzsignal. Der Schätzwert ist linear und erwartungstreu, der Schätzfehler weist eine minimale Varianz auf (s. Hayes, 1996, Kap. 7.4).

[53] *Minimum mean square error* (MMSE).

[54] Im Gegensatz zum Wiener-Filter, bei dem die Leistungsdichtespektren der Nutz- und Störkomponenten geschätzt werden, benötigt das Kalman-Filter die Parameter der verwendeten Modelle für Nutzsignal und Störung um das Schätzproblem zu lösen.

Ein wesentlicher Vorteil gegenüber dem Wiener-Filter besteht darin, dass das Signal am Ausgang eines Kalman-Filters auch in nichtstationären Umgebungen ohne Einschwingen sofort optimal ist (siehe z. B. Hayes, 1996, Kap. 7.4). Zudem lassen sich Kalman-Filter relativ recheneffizient auf digitalen Signalprozessoren implementieren. Kalman-Filter werden in dieser Arbeit nicht weiter diskutiert, da das Hauptaugenmerk auf dem Verfahren der spektralen Subtraktion liegt. Eine Einführung in die Theorie der Kalman-Filter findet sich beispielsweise in Haykin (2001, 2002a, Kap. 10). Zur weiterführenden Studie von Kalman-Filtern zur Aufbereitung von Sprachsignalen sei auf die Arbeiten von Paliwal und Basu (1987), Gibson et al. (1991), Gannot et al. (1998), Ma et al. (2006), Benesty et al. (2008b, Kap. 8) und Gannot (2012) verwiesen. Der Entwurf eines mehrkanaligen Kalman-Filters wird zum Beispiel in Kaps (2008) diskutiert.

Spektrale Subtraktion. Die spektrale Subtraktion (s. Boll, 1979) ist eng mit dem Wiener-Filter verwandt und gehört zu den am weitesten verbreiteten Methoden zur Unterdrückung von Störgeräuschen in einkanaligen Mikrofonanwendungen. Bei der spektralen Subtraktion wird eine Schätzung des Kurzzeit-Leistungsdichtespektrums des Störsignals vom gestörten Gesamtsignal abgezogen (siehe z. B. Vary et al., 1998, Kap. 12.4). Dabei wird die Unkorreliertheit von Nutzsignal und Störung vorausgesetzt (s. a. Abschnitt 4.2). In der Literatur finden sich zahlreiche Varianten der spektralen Subtraktion. Diese haben meist eine möglichst weitgehende Reduktion der Verzerrung des Nutzsignals bei gleichzeitiger Minimierung der Reststörungen zum Ziel (siehe z. B. Benesty et al., 2008b, Kap. 44). In praxisnahen Anwendungen nehmen die Subtraktionsregeln von Ephraim und Malah (1984, 1985) und die daraus abgeleiteten Verfahren eine Sonderstellung ein, da diese die tonalen Artefakte der spektralen Subtraktion weitestgehend unterdrücken. Das Ephraim-Malah-Filter wird in Abschnitt 4.3 aufgegriffen und dahingehend modifiziert, dass die durch Fehler bei der Schätzung des zeitvarianten Störgeräuschspektrums auftretenden nichtlinearen Verzerrungen möglichst stark reduziert werden (s. a. Zotter, 2004; Noisternig et al., 2009).

Mehrkanalige Geräuschreduktionsverfahren. In der praktischen Anwendung einkanaliger Geräuschreduktionsverfahren führt eine Verbesserung des SNR

nicht unweigerlich zu einer Verbesserung der Sprachverständlichkeit (siehe z. B. Hamacher et al., 2005; Hu und Loizou, 2007). Allerdings konnte in Hörversuchen sowohl eine Verbesserung des Höreindrucks als auch eine Verminderung der Höranstrengung gezeigt werden (vgl. Weiss et al., 1975; Trine und Van Tasell, 2002; Bitzer et al., 2005; Hu und Loizou, 2008). Darüber hinaus kommt es bei Systemen zur automatischen Spracherkennung zu einer deutlichen Steigerung der Erkennungsrate (siehe z. B. Vary et al., 1998, S. 396).

Sind Nutz- und Störschallquellen räumlich getrennt, kann mit mehrkanaligen Geräuschreduktionsverfahren neben dem Höreindruck auch die Sprachverständlichkeit nachweislich verbessert werden (vgl. Hamacher et al., 2005; Chen et al., 2006). In der Praxis finden die aus dem Minimum-Varianz-Ansatz (MVDR/LCMV)[55] abgeleiteten Optimalfilter (vgl. Capon, 1969; Frost, 1972) breite Anwendung. Diese minimieren die Varianz des Signals am Systemausgang, unter der Nebenbedingung unverzerrter Wiedergabe aus Richtung der Nutzschallquelle. Eine robuste Implementierung dieses Ansatzes ist zum Beispiel der *Generalized Sidelobe Canceller* (GSC; vgl. Griffiths und Jim, 1982; Hoshuyama et al., 1999). Sind die Signale schmalbandig, ist die mit dem Minimum-Varianz-Ansatz erreichte SNR-Verbesserung optimal im Sinne einer Maximum-Likelihood (ML) Schätzung. Für breitbandige Signale ist die SNR-Verbesserung nicht mehr optimal (s. a. Monzingo und Miller, 1980; Simmer et al., 2001). Hier lässt sich mit einem mehrkanaligen Wiener-Filter (MWF) eine optimale Lösung im Sinne des MMSE erreichen (vgl. Cornelis et al., 2011). Das MWF kann wiederum als MVDR-Beamformer mit einem in Serie geschalteten einkanaligen Wiener-Filter interpretiert werden (Edelblute et al., 1967).[56] Dies motiviert den in dieser Arbeit verfolgten Ansatz, die residualen Störgeräusche am Ausgang eines Mikrofonarrays durch Nachschalten eines einkanaligen Geräuschreduktionsfilters zu reduzieren.[57] Um die Verzerrung der Transienten bei der Nachfilterung möglichst gering zu halten, wird in Kap. 4.3 ein modifiziertes, schnell ansprechendes Ephraim-Malah-

[55] *Minimum variance distortionless response* (MVDR) und *Linearly constrained minimum variance* (LCMV).

[56] Siehe auch Brooks und Reed (1972), Monzingo und Miller (1980), Simmer et al. (2001), Van Trees (2002, Kap. 6) und Herbordt (2005).

[57] Siehe auch Zelinski (1988), Fischer und Simmer (1996), Marro et al. (1998), Bitzer et al. (1999a, 2001), Simmer et al. (2001), McCowan und Bourlard (2002, 2003) und Hendriks et al. (2009).

Filter vorgestellt. Auf den Entwurf mehrkanaliger Geräuschreduktionsverfahren wird dann in den Kapiteln 5 (robuste Mikrofonarray-Beamformer) und 6 (modale Mikrofonarray-Beamformer) näher eingegangen.

4.1 Einkanaliges Optimalfilter – Wiener-Filter

Die Wiener-Filter-Theorie (Wiener, 1942, 1949) bildet die Grundlage für den Entwurf signalabhängiger Optimalfilter. Die Erweiterung der Theorie auf zeitdiskrete Signale wurde von Levinson (1946) formuliert.[58] Ausgangspunkt der Herleitung des Wiener-Filters[59] ist eine dem Nutzsignal $x(n)$ additiv überlagerte Störung $v(n)$:

$$y(n) = x(n) + v(n). \tag{4.1}$$

Dabei wird angenommen, dass es sich bei $x(n)$ und $v(n)$ um stationäre, mittelwertfreie stochastische Prozesse handelt. Das zeitinvariante Wiener-Filter schätzt das Nutzsignal $\hat{x}(n)$ aus dem gestörten Gesamtsignal $y(n)$. Die Herleitung des zeitdiskreten Wiener-Filters lässt sich durch Umformulieren der Faltungssumme in Matrix-Vektor-Schreibweise sehr anschaulich als einfaches Skalarprodukt

$$\hat{x}(n) = \mathbf{h}^T \mathbf{y}(n) \tag{4.2}$$

des Koeffizientenvektors $\mathbf{h}$ eines Filters der Länge L

$$\mathbf{h} = \left[\begin{array}{cccc} h_0, & h_1, & \ldots, & h_{L-1} \end{array} \right]^T \tag{4.3}$$

mit dem Signalvektor $\mathbf{y}(n)$

$$\mathbf{y}(n) = \left[\begin{array}{cccc} y(n), & y(n-1), & \ldots, & y(n-L+1) \end{array} \right]^T \tag{4.4}$$

[58] Kailath (1974) gibt einen guten Überblick über frühe Arbeiten zur Optimalfilter-Theorie.

[59] Herleitung und Eigenschaften des Wiener-Filters werden beispielsweise in den Arbeiten von Vary et al. (1998, Kap. 12.3), Moschytz und Hofbauer (2000, Kap. 2), Haykin (2002a, Kap. 2), Hänsler und Schmidt (2004, Kap. 5), Vaseghi (2006, Kap. 6) und Benesty et al. (2008b, Kap. 6) ausführlich diskutiert.

darstellen. Da das Wiener-Filter zeitinvariant ist, wird beim Koeffizientenvektor $\mathbf{h}$ der Zeitindex n vorläufig weggelassen.

Als Gütemaß der linearen optimalen Filterung wird der mittlere quadratische Fehler (MSE), der die Abweichung des Schätzwertes $\hat{x}(n)$ vom wahren Wert $x(n)$ beschreibt, herangezogen:

$$\begin{aligned}
\mathrm{E}\left\{e(n)^2\right\} &= \mathrm{E}\left\{\left(x(n)-\hat{x}(n)\right)^2\right\} \\
&= \mathrm{E}\left\{\left(x(n)-\mathbf{h}^T\mathbf{y}(n)\right)^2\right\} \\
&= \mathrm{E}\left\{x^2(n)\right\} - 2\mathbf{h}^T\mathrm{E}\left\{\mathbf{y}(n)x(n)\right\} + \mathbf{h}^T\mathrm{E}\left\{\mathbf{y}(n)\mathbf{y}^T(n)\right\}\mathbf{h} \\
&= \mathbf{r}_{xx}(0) - 2\mathbf{h}^T\mathbf{r}_{\mathbf{y}x} + \mathbf{h}^T\mathbf{R}_{\mathbf{yy}}\mathbf{h}\,. \qquad (4.5)
\end{aligned}$$

$\mathrm{E}\{\cdot\}$ bezeichnet den Erwartungswert (oder das erste Moment) der stochastischen Prozesse und $(\cdot)^H$ die komplex konjugierte (oder auch hermitesche) Transposition. $\mathbf{R}_{\mathbf{yy}} = \mathrm{E}\left\{\mathbf{y}(n)\mathbf{y}(n)^T\right\}$ ist die Autokorrelationsmatrix des Eingangssignals $\mathbf{y}(n)$, wobei für komplexe Signale $\mathbf{R}_{\mathbf{yy}} = \mathrm{E}\left\{\mathbf{y}(n)\mathbf{y}(n)^H\right\}$ gilt. $\mathbf{r}_{\mathbf{y}x} = \mathrm{E}\left\{\mathbf{y}(n)x(n)\right\}$ ist der Kreuzkorrelationsvektor des Nutzsignals $x(n)$ mit dem Eingangssignal $\mathbf{y}(n)$. Für schwach stationäre Signale ist $\mathbf{R}_{\mathbf{yy}}$ eine symmetrische Toeplitz-Matrix, die fast immer positiv definit und damit meist invertierbar ist. Bei einem reellen Signal ist $\mathbf{R}_{\mathbf{yy}}$ symmetrisch (d. h. $\mathbf{R}_{\mathbf{yy}}^T = \mathbf{R}_{\mathbf{yy}}$), bei einem komplexen Signal folglich hermitesch (d. h. $\mathbf{R}_{\mathbf{yy}}^H = \mathbf{R}_{\mathbf{yy}}$). Sind die Signale nicht nur reell und schwach stationär, sondern auch mittelwertfrei, gilt $\mathrm{E}\left\{x^2(n)\right\} = \mathbf{r}_{xx}(0) = \sigma_x^2$. Daraus ergibt sich mit Gl. (4.5) folgende quadratische Fehlerfunktion

$$J(\mathbf{h}) = \mathrm{E}\left\{e(n)^2\right\} = \sigma_x^2 - 2\mathbf{h}^T\mathbf{r}_{\mathbf{y}x} + \mathbf{h}^T\mathbf{R}_{\mathbf{yy}}\mathbf{h}, \qquad (4.6)$$

die für den optimalen Filterkoeffizientenvektor $\mathbf{h}_{\mathrm{opt}}$ ihr Minimum annimmt. $\mathbf{h}_{\mathrm{opt}}$ ergibt sich, wenn der Gradient der Fehlerfunktion nach den Filterkoeffizienten $\mathbf{h}$ zu Null wird

$$\begin{aligned}
\nabla J(\mathbf{h}) = \frac{\partial J(\mathbf{h})}{\partial \mathbf{h}} &= -2\mathrm{E}\left\{\mathbf{y}(n)x(n)\right\} + 2\mathbf{h}^T\mathrm{E}\left\{\mathbf{y}(n)\mathbf{y}^T(n)\right\} \\
&= -2\mathbf{r}_{\mathbf{y}x} + 2\mathbf{h}^T\mathbf{R}_{\mathbf{yy}} = \mathbf{0}, \qquad (4.7)
\end{aligned}$$

$$\mathbf{R}_{\mathbf{yy}}\,\mathbf{h}_{\mathrm{opt}} = \mathbf{r}_{\mathbf{y}x} \qquad (4.8)$$

und die Hesse-Matrix (d. h. die Matrix aller zweifachen partiellen Ableitungen) positiv definit ist.[60] Die Hesse-Matrix von $J(\mathbf{h})$ ergibt sich zu $\mathbf{H_h} = 2\mathbf{R_{yy}}$ und ist damit in der Praxis meist positiv definit.[61] Aufgelöst nach dem Gewichtsvektor ergibt sich aus Gl. (4.8) die Wiener-Hopf-Gleichung für zeitdiskrete Signale:

$$\mathbf{h}_{\text{opt}} = \mathbf{R}_{\mathbf{yy}}^{-1}\, \mathbf{r}_{\mathbf{y}x}. \tag{4.9}$$

Aus der Annahme der Unkorreliertheit der Prozesse $x(n)$ und $v(n)$ folgt für die Korrelationsmatrix des gestörten Signals $\mathbf{R_{yy}} = \mathbf{R_{xx}} + \mathbf{R_{vv}}$ und für den Kreuzkorrelationsvektor $\mathbf{r}_{\mathbf{y}x} = \mathbf{r}_{xx}$. Eingesetzt in Gl. (4.9) ergibt sich für den optimalen Filterkoeffizientenvektor folgender Zusammenhang:

$$\mathbf{h}_{\text{opt}} = \left(\mathbf{R_{xx}} + \mathbf{R_{vv}}\right)^{-1}\, \mathbf{r}_{xx}. \tag{4.10}$$

Das Wiener-Filter $\mathbf{h}_{\text{opt}}$ ist minimalphasig. Wird in der Herleitung von vornherein ein nicht-kausales System angesetzt, indem die Impulsantwort in Gl. (4.3) die Bedingung $h(-i) = h(i)$ mit $i \in \{0, \ldots, L-1\}$ erfüllt, ergibt sich ein nullphasiges Filter (siehe z. B. Vary et al., 1998, Kap. 12.3.1).

4.1.1 Frequenzbereichslösung

Die Frequenzbereichslösung des Wiener-Filters kann sehr einfach über das Wiener-Khintchine-Theorem (Wiener, 1930; Khintchine, 1934) hergeleitet werden. Dieses besagt, dass das Leistungsdichtespektrum (LDS) $\Phi_{XX}(e^{j\Omega})$ und die Autokorrelationsfunktion (AKF) $r_{xx}(k)$ eines stationären Zufallsprozesses $x(n)$ über die Fourier-Transformation zusammenhängen. Das Autoleistungsdichtespektrum $\Phi_{XX}(e^{j\Omega})$ und das Kreuzleistungsdichtespektrum $\Phi_{XY}(e^{j\Omega})$ lassen sich für zeitdiskrete verbundstationäre mittelwertfreie Prozesse wie folgt darstellen:

$$\Phi_{XX}(e^{j\Omega}) = \mathrm{E}\left\{X(e^{j\Omega})\, X^*(e^{j\Omega})\right\} = \sum_{\kappa=-\infty}^{\infty} r_{xx}(k) e^{-j\Omega\kappa}, \tag{4.11}$$

[60] Ist die Hesse-Matrix positiv definit, ist der stationäre Punkt ein isoliertes lokales Minimum. Ist die Hesse-Matrix hingegen positiv semidefinit, ist der Punkt ein lokales Minimum.

[61] In seltenen Fällen ist $\mathbf{R_{yy}}$ positiv semidefinit und es existieren beliebig viele Lösungen, welche die Fehlerfunktion $J(\mathbf{h})$ minimieren.

$$\Phi_{XY}(e^{j\Omega}) = \mathrm{E}\left\{X(e^{j\Omega})\, Y^*(e^{j\Omega})\right\} = \sum_{\kappa=-\infty}^{\infty} r_{xy}(k) e^{-j\Omega\kappa}. \tag{4.12}$$

Die spektrale Darstellung $X(e^{j\Omega})$ einer diskreten Zeitfolge $x(n)$ erfolgt üblicherweise über die zeitdiskrete Fouriertransformation (DTFT)[62]

$$X(e^{j\Omega}) = \sum_{n=-\infty}^{\infty} x(n)\, e^{-j\Omega n} \quad \text{und} \quad x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\Omega})\, e^{j\Omega n} d\Omega. \tag{4.13}$$

Dabei bezeichnet $\Omega = 2\pi f / f_A$ die auf die Abtastrate f_A normierte und im Bereich $-\pi < \Omega \leq \pi$ definierte Frequenzvariable. Bei der Kurzzeit-Fourier-Transformation (STFT)[63] wird das kontinuierliche Spektrum $X(e^{j\Omega})$ durch Diskretisierung approximiert. Dazu werden die zugehörigen zeitdiskreten Signale $x(k)$ in Blöcke der Länge $k = 0, \ldots, K-1$ unterteilt, zur Verhinderung des Leck-Effekts[64] mit einer geeigneten Fensterfunktion $w(k)$ multipliziert und mittels diskreter Fourier-Transformation (DFT) in den Frequenzbereich transformiert.[65] Der Zusammenhang von Kurzzeitspektrum $X_\mu(\ell) := X(e^{j\Omega})|_{\Omega=2\pi\mu/K}$ und diskretem Zeitsignal $x(k)$ lässt sich zeigen als

$$X_\mu(\ell) = \sum_{k=0}^{K-1} w_k\, x(k + \ell R)\, e^{-j\frac{2\pi\mu k}{K}}, \tag{4.14}$$

wobei μ den Frequenzindex, ℓ den Blockindex und R die Sprungweite (*hop size*), welche in einem ganzzahligen Verhältnis zur Blocklänge K stehen soll, bezeichnet. Bei Verwendung von Kurzzeitgrößen kann die Rücktransformation unter Berücksichtigung der Blocklänge und der Blocküberlappung nach dem *Overlap-Save* bzw. nach dem *Overlap-Add*-Verfahren erfolgen. An dieser Stelle sei auf die weiterführende Literatur (z. B. Oppenheim et al. 1998, Kap. 8.7.3; Vary

[62] *Discrete-time Fourier transform* (DTFT); siehe z. B. Allen und Rabiner (1977); Oppenheim et al. (1998, Kap. 2.7).

[63] *Short-time Fourier transform* (STFT); siehe z. B. Oppenheim et al. (1998, Kap. 10.3)

[64] Siehe z. B. Kammeyer und Kroschel (1992, Kap. 7.3.3).

[65] Für Sprache werden meist Blocklängen von $10 - 30$ ms gewählt, in denen die Zufallsprozesse als stationär angenommen werden können (vgl. Rabiner und Schafer, 1978; Cohen, 2005). Werden die Signalblöcke mit Hann-Fenstern multipliziert, wird meist eine Blocküberlappung von 50 %, bei Hamming-Fenstern von 75 % verwendet (vgl. Martin et al., 2004).

und Martin 2006, Kap. 3) verwiesen. In den folgenden Betrachtungen wird auf die Blockverarbeitung nur dann eingegangen, wenn diese für die Realisierung der Algorithmen von besonderer Bedeutung ist.

Unter der Annahme schwach stationärer Signale lässt sich Gl. (4.8) in den Frequenzbereich transformieren:

$$H(e^{j\Omega}) = \frac{\Phi_{XY}(e^{j\Omega})}{\Phi_{YY}(e^{j\Omega})}\,. \tag{4.15}$$

Sind Nutz- und Störsignal nicht miteinander korreliert, folgt aus Gl. (4.10)

$$H(e^{j\Omega}) = \frac{\Phi_{XX}(e^{j\Omega})}{\Phi_{XX}(e^{j\Omega}) + \Phi_{VV}(e^{j\Omega})}\,. \tag{4.16}$$

Die Berechnung des Wiener-Filters setzt demzufolge die Kenntnis der LDS des ungestörten Signals und der Störung voraus. In der Praxis steht als Messgröße allerdings nur das gestörte Gesamtsignal zur Verfügung. Das LDS des ungestörten Signals $\Phi_{\hat{X}\hat{X}}(e^{j\Omega})$ muss geschätzt werden:

$$\begin{aligned}\Phi_{\hat{X}\hat{X}}(e^{j\Omega}) &= \Phi_{YY}(e^{j\Omega}) - \Phi_{VV}(e^{j\Omega}) \\ &= \Phi_{YY}(e^{j\Omega}) \left[1 - \frac{\Phi_{VV}(e^{j\Omega})}{\Phi_{YY}(e^{j\Omega})}\right] \\ &= \Phi_{YY}(e^{j\Omega}) \left|\tilde{H}(e^{j\Omega})\right|^2 .\end{aligned} \tag{4.17}$$

Dabei wird der Term $\left[1 - \Phi_{VV}(e^{j\Omega})/\Phi_{YY}(e^{j\Omega})\right]$ als Betragsquadrat-Frequenzgang des Filters $\tilde{H}(e^{j\Omega})$ interpretiert. Zur Berechnung des Filters ist somit alleine die Schätzung des LDS der Störung notwendig. Durch Ziehen der Wurzel kann aus Gl. (4.17) das Betragsspektrum des Filters bestimmt werden. Dieser Ansatz führt auf die Methode der spektralen Subtraktion, die in Kap. 4.2 ausführlicher behandelt wird.

Umformen obiger Gleichungen führt auf die Gewichte des Wiener-Filters

$$G_{\mathrm{W}}(e^{j\Omega}) = \frac{SNR(e^{j\Omega})}{SNR(e^{j\Omega}) + 1} \tag{4.18}$$

und der spektralen Subtraktion

$$G_{\mathrm{SS}}(e^{j\Omega}) = \sqrt{\frac{SNR(e^{j\Omega})}{SNR(e^{j\Omega})+1}} \tag{4.19}$$

in Abhängigkeit vom Signal-Stör-Verhältnis

$$SNR(e^{j\Omega}) = \frac{\Phi_{XX}(e^{j\Omega})}{\Phi_{VV}(e^{j\Omega})}. \tag{4.20}$$

Die Gewichtsfunktionen $G_{\mathrm{W}}(e^{j\Omega})$ und $G_{\mathrm{SS}}(e^{j\Omega})$ sind in Abb. 4.1 dargestellt. Daraus ist ersichtlich, dass das Wiener-Filter bei geringem SNR Störgeräusche wesentlich stärker unterdrückt als die spektrale Subtraktion.

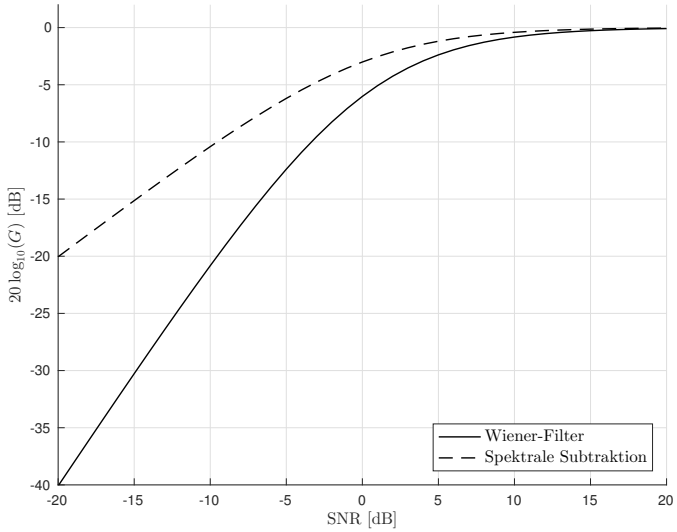


Abbildung 4.1: Spektrale Gewichte des Wiener-Filters (durchgezogene Linie) und der spektralen Subtraktion (gestrichelte Linie) in Abhängigkeit vom SNR.

4.1.2 Periodogramm zur Schätzung des LDS

Das Periodogramm (Welch, 1967) ist eines der bekanntesten nicht parametrischen Verfahren zur Schätzung des LDS. Unter der Voraussetzung schwach stationärer, ergodischer Prozesse, lassen sich die Erwartungswerte in den Gleichungen (4.11) und (4.12) durch Zeitmittelwerte ersetzen. Das Periodogramm $I_{\mu,XX}(\ell)$ berechnet

sich wie folgt aus dem Produkt der DFT-Koeffizienten des Signals

$$\begin{aligned} I_{\mu,XX}(\ell) &= \frac{1}{KU}\,|X_\mu(\ell)|^2 = \frac{1}{KU} X_\mu(\ell)\,X_\mu^*(\ell) \\ &= \frac{1}{KU}\sum_{k_1=0}^{K-1}\sum_{k_2=0}^{K-1} w_{k_1} w_{k_2}\,x(k_1+\ell R)\,x^*(k_2+\ell R)\,e^{-j\frac{2\pi\mu(k_2-k_1)}{K}}, \end{aligned} \tag{4.21}$$

wobei der multiplikative Term $1/KU$ den systematischen Fehler minimiert (vgl. Welch, 1967; Oppenheim et al., 1998) . Durch Normieren der Fensterfunktion

$$\sum_{k=0}^{K-1} |w_k|^2 = 1$$

und geeignete Wahl der Normierungskonstante

$$U = \frac{1}{K}\sum_{k=0}^{K-1} |w_k|^2$$

ergibt sich $1/KU = 1$. Das Periodogramm ist ein asymptotisch erwartungstreuer Schätzer für das LDS eines Signals. Durch rekursive Mittelung zeitlich aufeinander folgender unabhängiger Periodogramme kann die Varianz der Schätzung verringert werden:

$$\tilde{\Phi}_{\mu,XX}(\ell) = \alpha_{\text{avg}}\,\tilde{\Phi}_{\mu,XX}(\ell-1) + (1-\alpha_{\text{avg}}) I_{\mu,XX}(\ell). \tag{4.22}$$

Bei der Bartlett-Methode (s. Bartlett, 1948) werden die sich überlappenden Signalblöcke mit einem Rechteckfenster, beim Welch-Periodogramm (s. Welch, 1967) mit unterschiedlichen Fensterfunktionen bewertet. Der Gedächtnisfaktor der Mittelung $0 \leq \alpha_{\text{avg}} \leq 1$ wird meist als Zeitkonstante τ_{avg} angegeben, mit $\alpha_{\text{avg}} = \exp\left(-1/(\tau_{\text{avg}} \cdot f_A)\right)$. Die Zeitkonstante der Mittelung sollte in etwa dem Zeitraum entsprechen, in dem das Signal als stationär angenommen werden kann.

4.1.3 Schätzung des LDS des Störsignals

Für die Berechnung des Wiener-Filters und der spektralen Subtraktion muss das Störspektrum aus dem gestörten Gesamtsignal geschätzt werden. Im einfachsten

Fall wird diese Schätzung in Segmenten durchgeführt, in denen kein Nutzsignal vorhanden ist. Diese können über eine Signalpausenerkennung (VAD)[66] detektiert werden. Unter der Annahme stationärer bzw. langsam veränderlicher Störsignale gelten die geschätzten Kurzzeitmittelwerte auch in unmittelbarer zeitlicher Umgebung der Schätzung, d. h. in den Segmenten, in denen sowohl Nutz- als auch Störsignal vorhanden sind. In diesen Segmenten kann nun mit Gl. (4.17) das Wiener-Filter berechnet werden. Die Zuverlässigkeit der VAD hat somit direkten Einfluss auf die Qualität der Störgeräuschunterdrückung (vgl. Le Bouquin-Jeannès und Faucon, 1995). Signalpausen lassen sich im einfachsten Fall über adaptive, aus dem Gesamtpegel ermittelte Schwellwerte detektieren (vgl. Doblinger, 1995; Hui, 2000). Unterschreitet der Pegel des gestörten Signals einen unteren Schwellwert, ist im betrachteten Segment ausschließlich Störsignal vorhanden; wird ein oberer Schwellwert überschritten, sind sowohl Nutz- als auch Störsignal vorhanden. Bei zeitlicher Änderung des Störsignalpegels werden die Schwellwerte langsam nachgeführt. Mit globalen, aus dem Gesamtpegel ermittelten Schwellwerten, kann bei geringem SNR nur sehr schwer zwischen Stör- und Nutzsignalperioden unterschieden werden. Es kommt vermehrt zu Fehlern bei der Schätzung des Störspektrums und infolgedessen zu einer Beeinträchtigung des Höreindrucks. Bei Sprache kann durch Schätzen der Signalleistungen über einen nichtlinearen Operator – den sogenannten *Teager-Energy-Operator* (TEO) – die Robustheit der VAD erhöht werden (siehe z. B. Jabloun et al., 1999; Hui, 2000; Junhui et al., 2003; Song et al., 2009). Der TEO bewertet tieffrequente Störungen weniger stark, was vor allem bei Freisprecheinrichtungen in Fahrzeugen[67] zu einer wesentlichen Verringerung der Fehleranfälligkeit der VAD führt (s. a. Hui, 2000). Andere Schätzer berücksichtigen zusätzlich die Wahrscheinlichkeit, ob in einem Segment Nutzsignal vorhanden ist (vgl. Azirani et al., 1996; Malah et al., 1999; Cohen und Berdugo, 2001b), um die Robustheit der VAD und somit den Höreindruck zu verbessern.

Wird das gestörte Gesamtsignal in schmale Frequenzbänder zerlegt, und die VAD lokal in jedem Frequenzband durchgeführt, kann diese den zeitlichen Änderungen des Störspektrums besser folgen. Dies erhöht die Robustheit der VAD.

[66] In der Signalverarbeitung wird, unabhängig davon ob es sich um Sprache handelt oder nicht, meist der Term Sprachaktivitätserkennung (*voice activity detection*, VAD) verwendet.

[67] Hänsler und Schmidt (2008, Kap. 3.2) zeigen den vorwiegend tieffrequenten Charakter typischer Störgeräusche im Fahrzeug (wie z. B. Roll-, Wind- und Motorengeräusche).

Besitzt das Nutzsignal eine hohe Dynamik und ist das Störgeräusch hinreichend stationär, entsprechen die Minima in jedem Frequenzband der spektralen Rauschleistungsdichte. Das Minimum-Statistik-Verfahren (vgl. Doblinger, 1995; Martin, 2001) sucht in einem bestimmten Zeitintervall nach den Minima. Das Störspektrum wird dann durch Mittelung der Minima geschätzt. Das Minimum-Statistik-Verfahren benötigt somit keine VAD und kann die Schätzung auch dann aktualisieren, wenn in einem Segment Nutzsignal vorhanden ist. Das Verfahren gilt als relativ robust, bei vertretbarem Rechenaufwand (vgl. Meyer et al., 1997), kann jedoch in Umgebungen mit nichtstationären Störgeräuschen (wie z. B. Stimmengewirr in einem Café) nicht angewendet werden (siehe z. B. Hänsler und Schmidt, 2004, Kap. 14.1). Ealey et al. (2001) verfolgen einen sehr ähnlichen Ansatz, bei dem die spektrale Rauschleistungsdichte in den „Tälern“ zwischen den Harmonischen eines Sprachsignals geschätzt wird. Daraus leitet sich die Bezeichnung *Harmonic Tunneling* ab. Arslan et al. (1995) schlagen einen weiteren kontinuierlichen Schätzer für das Störspektrum vor, bei dem Anstieg und Abfall der in jedem Teilband geschätzten spektralen Leistungsdichten auf $+3\,\mathrm{dB/s}$ bzw. $-12\,\mathrm{dB/s}$ beschränkt wird. Dadurch folgt die Schätzung dem schnellen Pegelanstieg bei Sprach-Onsets nur langsam, fällt in den Sprachpausen jedoch sehr schnell auf den Wert des Störsignals ab.

Hess (1976) bestimmt die Schwellwerte der VAD aus dem Histogramm der Pegel des gestörten Signals. Dabei wird die Annahme getroffen, dass der Signalpegel mit der größten Häufigkeit dem Störsignal zugeordnet werden kann. Dies trifft vor allem auf Sprache zu. Eine Erweiterung dieses Verfahrens ist in Hirsch und Ehrlicher (1995) zu finden. Diese bestimmen das Histogramm nicht breitbandig, sondern lokal in schmalen Frequenzbändern. Stahl et al. (2000) schlagen eine sehr ähnliche, allerdings auf Quantilen basierende Schätzung des Störspektrums vor. Dabei werden die Amplituden der Kurzzeitspektren zuerst in aufsteigender Reihenfolge sortiert. Der Schwellwert der VAD wird dann über den Medianwert ($q = 0{,}5$) gebildet. Eine Variante dieses Ansatzes (s. Evans und Mason, 2002) berücksichtigt bei der Bestimmung der Schwellwerte auch die Medianwerte benachbarter Frequenzbänder. Dadurch wird eine zusätzliche spektrale Glättung des geschätzten Störspektrums erreicht, die sich positiv auf den Höreindruck auswirkt. Nach Evans et al. (2002) erreichen automatische Spracherkennungssysteme mit quantilen-basierten Ansät-

zen eine höhere Erkennungsrate als zum Beispiel jene mit Harmonic Tunneling. Eine ausführliche Diskussion weiterer Methoden zur Schätzung des Störspektrums findet sich zum Beispiel in Cohen (2003), Vary und Martin (2006, Kap. 11.8), Hendriks et al. (2013, Kap. 6) und Benesty et al. (2008b, Kap. 44.7).

Die Frequenzauflösung der in Kap. 3 vorgestellten auditiven Analyse-Synthese-Filterbank reicht für die sinnvolle Anwendung quantilen-basierter Schätzverfahren bzw. des Harmonic-Tunnelings nicht aus. Aus diesem Grund wird in dieser Arbeit das Störspektrum über eine VAD mit frequenzselektiven adaptiven Schwellwerten geschätzt (vgl. Doblinger, 1995; Hui, 2000). Um bei Sprache die Robustheit zu erhöhen, werden die Signalleistungen über den TEO bestimmt.

4.2 Spektrale Subtraktion

Die spektrale Subtraktion gehört zu den bekanntesten Methoden zur Unterdrückung von Störgeräuschen in einkanaligen Systemen. Der allgemeine Ansatz wurde von Schroeder (1965, 1968) für analoge und etwas später von Boll (1979) für digitale Systeme formuliert.[68] Die spektrale Subtraktion ist sehr eng mit der Optimalfilterung verwandt (siehe z. B. McAulay und Malpass, 1980) und es gelten dieselben Überlegungen wie in Abschnitt 4.1. Ausgangspunkt ist wiederum die Annahme der Unkorreliertheit von Nutz- und Störsignal. Die spektrale Subtraktion schätzt zuerst das Betragsspektrum des ungestörten Signals aus dem gestörten Gesamtsignal. In der Praxis sind die Leistungsdichten durch Kurzzeitschätzwerte zu ersetzen. Dies führt auf das weit verbreitete Verfahren der Kurzzeit-Spektraldämpfung. Dabei werden einzelne Blöcke des Eingangssignals durch eine beliebige Kurzzeit-Spektraltransformation in Teilbänder zerlegt (vgl. Abb. 3.2). Die Teilbandsignale werden zur Korrektur des gestörten Spektralanteils mit einer Gewichtsfunktion gemäß Gl. (4.17) multipliziert. Bei der Rücktransformation wird das geschätzte Betragsspektrum des ungestörten Signals mit der Phase des gestörten Gesamtsignals kombiniert. Neuere Arbeiten schätzen neben

[68] In etwa zeitgleich mit Boll (1979) erschienen die Arbeiten von Berouti et al. (1979), Preuss (1979) und Lim und Oppenheim (1979). Eine gute Zusammenstellung grundlegender Arbeiten zum Thema spektrale Subtraktion findet sich zum Beispiel in Ephraim et al. (2005), Ephraim und Cohen (2006), Vary und Martin (2006, Kap. 11), Vaseghi (2006, Kap. 11), Benesty et al. (2009) und Hendriks et al. (2013).

dem Betragsspektrum auch die Phase des ungestörten Signals.[69] Gerkmann und Krawczyk (2013) zeigen, dass mit diesem Ansatz die Robustheit der Schätzung ungestörter Sprache erhöht werden kann. Wird bei der Rücktransformation die geschätzte Phase verwendet, können jedoch vermehrt unerwünschte Verzerrungen des Sprachsignals auftreten. Sunnydayal und Kumar (2015) schlagen hierzu einen kombinierten Schätzer für das Betragsspektrum und die Phase vor, mit dem sich – unter der Voraussetzung bestimmter Verteilungsdichtefunktionen für Sprache und Störgeräusch – die Qualität des entstörten Sprachsignals wesentlich verbessern lässt. Die Ergebnisse der Hörversuche von Krawczyk-Becker und Gerkmann (2016) zeigen eine klare Präferenz der Versuchspersonen für Verfahren, bei denen die geschätzte Phase mit berücksichtigt wird.

Bei der spektralen Subtraktion kann es durch Fehler bei der Schätzung des Störspektrums zu maßgeblichen Verzerrungen des entstörten Signals kommen. Sehr häufig treten dabei periodische Reststörungen (sog. *Musical Noise*) auf, die typischerweise als sehr störend empfunden werden. Mit der folgenden, aus Hörversuchen abgeleiteten heuristischen Verallgemeinerung von Gl. (4.17), kann die Qualität des entstörten Signals wesentlich besser gesteuert werden (vgl. Vary et al., 1998, S. 387):

$$\begin{aligned}\Phi_{\hat{X}\hat{X}}(e^{j\Omega}) &= \Phi_{YY}(e^{j\Omega}) \left[1 - \delta \left(\frac{\Phi_{VV}(e^{j\Omega})}{\Phi_{YY}(e^{j\Omega})}\right)^{\beta}\right]^{\alpha} \\ &= \Phi_{YY}(e^{j\Omega}) \left|\tilde{H}(e^{j\Omega})\right|^2 .\end{aligned} \tag{4.23}$$

Durch die Wahl der Parameter α und β kann der subtrahierte Spektralanteil überschätzt oder auch unterschätzt werden. Der Parameter δ bestimmt, wie viel Störleistung vom Signal abgezogen wird. Für $\delta > 1$ spricht man im Allgemeinen

[69] Ephraim und Malah (1984) und Cohen (2005) zeigen, dass die optimale MMSE-Schätzung der exponentiellen Phase des ungestörten Signals der exponentiellen Phase des gestörten Signals entspricht. Dabei wird die Phase als gleichverteilt und unabhängig von der Signalamplitude angenommen. Nach Wang und Lim (1982) und McAulay und Malpass (1980), ist die menschliche Hörwahrnehmung bei Sprache zudem relativ unempfindlich gegenüber Phasenverzerrungen. In neueren Arbeiten zur einkanaligen Signalaufbereitung wird der Phase eine stärkere Bedeutung zugemessen (siehe z. B. Lotter und Vary, 2005; Shannon und Paliwal, 2006; Paliwal et al., 2011; Gerkmann und Krawczyk, 2013; Kulmer und Mowlaee, 2015; Mowlaee und Kulmer, 2015; Sunnydayal und Kumar, 2015).

von Übersubtraktion, für $\delta < 1$ von Untersubtraktion. In Perioden mit geringem SNR oder alleinigem Störgeräusch führen die Übersubtraktion und Überschätzung des Störgeräuschspektrums zu einer signifikanten Reduktion des Musical Noise (vgl. Berouti et al., 1979). Allerdings kommt es hierbei oft zu starken Verzerrungen des Nutzsignals, die generell als sehr störend empfunden werden. Dem kann nach Kushner et al. (1989) durch Verwendung einer vom SNR abhängigen adaptiven Gewichtsfunktion entgegengewirkt werden. Diese überschätzt das Störspektrum in Perioden mit geringem SNR und unterschätzt dieses in Perioden mit hohem SNR. Zur Verbesserung des Höreindrucks wird zudem oft eine Beschränkung der erzielbaren Störreduktion in Kauf genommen. Dadurch kann die Verzerrung des Nutzsignals möglichst gering gehalten werden und die Musical Noise wird durch den höheren Grundgeräuschpegel maskiert (siehe z. B. Berouti et al., 1979; Cappé, 1994; Gustafsson et al., 1998; Malah et al., 1999).

Eine weitere Möglichkeit zur Verbesserung des Höreindrucks besteht darin, bei der Berechnung der Spektralgewichte psychoakustische Größen zu berücksichtigen.[70] Eine weit verbreitete Variante dieses Ansatzes nutzt die zeitlichen und spektralen Maskierungseffekte des menschlichen Gehörs, um Störungen nur dort zu unterdrücken, wo sie auch tatsächlich wahrnehmbar sind (vgl. Tsoukalas et al., 1997). Wird zum Beispiel die Spektraltransformation als Filterbank realisiert, deren Kanäle an die Frequenzgruppen des menschlichen Gehörs angepasst werden, lassen sich die Gewichte der spektralen Subtraktion auf die Mithörschwelle begrenzen.[71] Dadurch werden maskierte Störungen weniger stark reduziert, wodurch sich die Verzerrung des Nutzsignals verringert. Die vorliegende Arbeit greift diesen Ansatz auf, wobei die Spektraltransformation mit der in Kap. 3 vorgestellten auditiven Analyse-Synthese-Filterbank durchgeführt wird (s. a. Zotter, 2004; Noisternig et al., 2009). Eine schematische Darstellung der Filterbank findet sich in Abb. 3.2.

[70] Siehe z. B. McAulay und Malpass (1980), Virag (1995, 1999), Tsoukalas et al. (1997), Thiemann (2001), Thiemann und Kabal (2002), Hu und Loizou (2003, 2004), Loizou (2005), Jo und Yoo (2009, 2010).

[71] Siehe z. B. Lorber und Höldrich (1997), Gustafsson et al. (1998, 2002), Irino (1999), Tuffy (1999), Virag (1999), Wolfe und Godsill (2000), Hui (2000), Lin und Ambikairajah (2002) und Lin et al. (2003).

In der Literatur finden sich zahlreiche Weiterentwicklungen der spektralen Subtraktion. In praktischen Anwendungen kommt sehr oft das Ephraim-Malah-Filter zum Einsatz. Dieses wird in den folgenden Abschnitten diskutiert und hinsichtlich der Minimierung nichtlinearer Verzerrungen des Nutzsignals optimiert.

4.3 Das Ephraim-Malah-Filter

Das Ephraim-Malah-Filter (EMSR[72]; vgl. Ephraim und Malah 1983) kann als Wiener-Filter mit zusätzlichem Korrekturterm zur Anpassung der spektralen Gewichte an die lokalen SNR-Verhältnisse interpretiert werden (vgl. Cappé, 1994). Das EMSR schätzt die Amplitude des ungestörten Signals aus dem Kurzzeitspektrum des gestörten Gesamtsignals. Ephraim und Malahs MMSE-Kurzzeit-Amplituden-Schätzer (MMSE-STSA[73]; vgl. Ephraim und Malah 1983, 1984) ist für gaußverteilte, mittelwertfreie und statistisch voneinander unabhängige Nutz- und Störsignale optimal. Zur Berücksichtigung des logarithmischen Lautstärkeempfindens des menschlichen Gehörs kann auch die logarithmierte Amplitude des Nutzsignals geschätzt werden (MMSE-LSA[74]; vgl. Ephraim und Malah 1985). Nach Cohen (2006) ist die Verzerrung des Nutzsignals beim MMSE-LSA-Schätzer wesentlich geringer als beim MMSE-STSA-Schätzer. Der MMSE-LSA-Schätzer erreicht zudem einen wesentlich höheren PESQ-Score[75] als andere Methoden der spektralen Subtraktion.

Eine Weiterentwicklung des MMSE-LSA-Schätzers findet sich in Malah et al. (1999). Dieser modifizierte LSA-Schätzer (MM-LSA)[76] multipliziert die spektralen Gewichte mit der Wahrscheinlichkeit für Sprachpräsenz, wodurch die Entstehung von Musical Noise weitgehend vermieden wird.[77] In Perioden, in denen

[72] *Ephraim and Malah spectral suppression rule* (EMSR).
[73] *MMSE short-time spectral amplitude estimator* (MMSE-STSA).
[74] *MMSE short-time log-spectral amplitude estimator* (MMSE-LSA).
[75] *Perceptual evaluation of speech quality* (PESQ), ITU-T Recommendation P.862 (2001).
[76] *Multiplicatively modified log-spectral amplitude estimator* (MM-LSA).
[77] Die Herleitung der meisten Schätzer beruht auf der Annahme, dass Nutz- und Störsignale gleichzeitig vorhanden sind. In Nutzsignalpausen trifft diese Annahme jedoch nicht zu. Der Schätzer ist nicht mehr optimal und es treten vermehrt hörbare Artefakte auf. Wird in der Herleitung die Auftrittswahrscheinlichkeit des Nutzsignals berücksichtigt, lassen sich diese Artefakte weitgehend vermeiden (siehe z. B. McAulay und Malpass, 1980; Cohen, 2003; Rangachari und Loizou, 2006; Hendriks et al., 2010; Gerkmann und Hendriks, 2012).

kein Nutzsignal vorhanden ist, tritt beim MM-LSA-Schätzer jedoch vermehrt eine unnatürlich klingende Reststörung auf (siehe z. B. Martin et al., 2000). Darüber hinaus kann gezeigt werden, dass der MM-LSA-Schätzer nicht optimal ist. Für den Fall, dass stationäre Störgeräusche (z. B. Hintergrundrauschen) und nichtstationäre Interferenzen (z. B. Raumhall und Echos) gleichzeitig auftreten, schlagen Cohen (2001, 2002) und Cohen und Berdugo (2001a,b) einen optimalen modifizierten LSA-Schätzer (OM-LSA)[78] in Kombination mit einem MCRA-Schätzer[79] vor. Der MCRA-Schätzer bestimmt das Störspektrum über einen Minimum-Statistik-Ansatz (vgl. Doblinger, 1995; Martin, 2001) und glättet die spektralen Gewichte über eine rekursive Mittelung. Dabei wird die Zeitkonstante der Mittelung von der Auftrittswahrscheinlichkeit des Nutzsignals in den jeweiligen Teilbändern abhängig gemacht.[77] Um eine möglichst natürlich klingende Reststörung zu erhalten, werden zudem die spektralen Gewichte durch eine untere Schranke begrenzt. Der von Cohen (2003) vorgeschlagene verbesserte MCRA-Schätzer (IMCRA)[80] ist besonders robust gegenüber transienten Störungen und Umgebungen mit geringem SNR. Nach Cohen (2004b) lässt sich durch die rekursive Mittelung des *a-priori*-SNR (s. Abschnitt 4.3.1) eine Verbesserung des Höreindrucks erreichen. Eine akausale Variante dieses Schätzers kann zudem transiente Störungen unterdrücken (vgl. Cohen, 2004c). Letztere Ansätze sind jedoch zum Teil sehr rechenaufwändig und führen im akausalen Fall zu einer zusätzlichen Verzögerung des Signals. Für den Entwurf der in dieser Arbeit betrachteten Zielanwendung – eines recheneffizienten Verfahrens mit möglichst kurzer Signallatenz – ist vor allem der von Wolfe und Godsill (2001) vorgeschlagene MMSE-Schätzer (MMSE-SP)[81] von Interesse. Der MMSE-SP-Schätzer lässt sich sehr recheneffizient implementieren und weist eine den MMSE-STSA/LSA Schätzern vergleichbare Störgeräuschunterdrückung auf (vgl. Abschnitt 4.3.1).

Wahrscheinlichkeitsdichteverteilung der Spektralkoeffizienten. Bei der Herleitung der spektralen Gewichte wird meist von dem Modell gaußverteilter Spektralkoeffizienten ausgegangen. Diese Annahme stützt sich ganz allgemein auf

[78] *Optimally modified log-spectral amplitude estimator* (OM-LSA).
[79] *Minima controlled recursive averaging* (MCRA).
[80] *Improved minima controlled recursive averaging* (IMCRA).
[81] *MMSE spectral power suppression rule* (MMSE-SP).

den zentralen Grenzwertsatz, der allerdings aufgrund der in der Praxis verwendeten kurzen Blocklängen nur bedingt anwendbar ist (vgl. Vary und Martin, 2006, Kap. 5.11). Manche Studien zeigen für Sprache eine wesentlich bessere Übereinstimmung der Verteilungsdichte der Spektralkoeffizienten mit Super-Gauß-Verteilungen, wie der Laplace- oder der Gamma-Verteilung. Verschiedene MMSE und MAP Schätzer hierzu finden sich in den Arbeiten von Porter und Boll (1984), Martin (2002, 2005), Martin und Breithaupt (2003), Cohen (2006) und Lotter und Vary (2004). Die Verteilungsdichten können auch mit Hilfe von Hidden-Markov-Modellen (HMM) aus den Signalen geschätzt werden (vgl. Ephraim et al., 1989, 1992a, 1992b, 2005). Hierbei hängt die Qualität der Störgeräuschunterdrückung jedoch stark von den in der Lernphase verwendeten Datensätzen ab. Hendriks und Martin (2007) leiten, unter der Annahme einer mehrdimensionalen Normal-Inverse-Gauß-Verteilung (MNIG)[82] der DFT-Koeffizienten des Nutzsignals, einen MAP-Schätzer her. Die MNIG-Verteilung kann eine Vielzahl unterschiedlicher Verteilungsdichten modellieren. Wird nun die tatsächliche Verteilungsdichte aus den Signalen geschätzt, lässt sich der MAP-Schätzer adaptiv anpassen. Experimente zeigen jedoch nur eine geringe SNR-Verbesserung ($< 1\,\mathrm{dB}$) gegenüber Schätzern die unter Annahme einer Normal-, Rayleigh- oder Super-Gauß-Verteilung hergeleitet wurden. Ephraim und Cohen (2006, Kap. 2.3) sehen für praktische Anwendungen kaum Vorteile in der Verwendung unterschiedlicher Verteilungsdichten. Dies wird vorwiegend dadurch begründet, dass für die meisten Verteilungen keine analytische MMSE-Lösung existiert, die allerdings für eine recheneffiziente numerische Implementierung wesentlich ist. Diesem Argument folgend, wird – wiederum hinsichtlich der Zielapplikation – auch in dieser Arbeit von gaußverteilten Spektralkoeffizienten ausgegangen. Eine ausführliche Diskussion der Auswirkung unterschiedlicher Verteilungsdichten auf die spektrale Subtraktion findet sich zum Beispiel in Vary und Martin (2006, Kap. 5).

4.3.1 Berechnung der spektralen Gewichte

Ausgangspunkt ist wiederum eine dem Nutzsignal $x(n)$ additiv überlagerte Störung $v(n)$, siehe Gl. (4.1). Die Signale werden, wie im vorhergehenden Kapitel

[82] Siehe auch Barndorff-Nielsen (1977) und Eriksson et al. (2009).

beschrieben, als statistisch unabhängige, gaußverteilte Zufallsprozesse modelliert. Aufgrund der Eigenschaften der Fouriertransformation, ist diese Annahme auch für die komplexen Koeffizienten der Kurzzeitspektren $\underline{X}_k(m)$ und $\underline{V}_k(m)$ gültig, wobei k den Frequenzindex und m den Blockindex bezeichnet. Komplexe Variablen werden im Folgenden mit einem Unterstrich notiert. $X_k(m)$ bezeichnet somit die Amplitude $|\underline{X}_k(m)|$ der komplexen Variable $\underline{X}_k(m)$. Zur Vereinfachung wird der Blockindex nur dort angegeben, wo dies für das Verständnis notwendig ist. Durch die Orthogonalität der Fourierkoeffizienten kann die Amplitude des ungestörten Signals in einem Frequenzband k unabhängig von den anderen Frequenzbändern bestimmt werden.

Ephraim und Malah (1984) lösen das Bayes-Problem $\hat{X}_k = \mathrm{E}\{X_k|Y_k\}$, um die Amplitude des Nutzsignals aus dem gestörten Gesamtsignal zu schätzen. Zum Formulieren der bayesschen Schätzfunktion wird sowohl das *a-priori*-SNR ξ_k als auch das *a-posteriori*-SNR γ_k benötigt. Diese können wie folgt aus den Varianzen des Nutzsignals $\sigma^2_{x,k}$ und des Störsignals $\sigma^2_{v,k}$ bestimmt werden:[83]

$$\xi_k = \frac{\sigma^2_{x,k}}{\sigma^2_{v,k}}, \tag{4.24}$$

$$\gamma_k = \frac{Y_k^2}{\sigma^2_{v,k}}. \tag{4.25}$$

Das *a-posteriori*-SNR ist eine lokale SNR-Schätzung im jeweils aktuellen Segment der Kurzzeit-Spektraltransformation. Es ist schnell veränderlich und dominiert das Dämpfungsverhalten der spektralen Subtraktion und des Wiener-Filters. Beim Ephraim-Malah-Filter besitzt es nur eine korrigierende Funktion, mit der die Dämpfung zwischen den Gewichten der spektralen Subtraktion und denen des Wiener-Filters variiert wird (vgl. Köhler, 2005, Kap. 11.7). Die Kennlinien sind in Abb. 4.1 dargestellt. Das Dämpfungsverhalten des Ephraim-Malah-Filters wird überwiegend vom *a-priori*-SNR bestimmt. Dieses entspricht im Wesentlichen einem über vergangene Segmente gemittelten SNR (vgl. Cappé, 1994). Ephraim und Malah (1983, 1984) schätzen das *a-priori*-SNR über einen entscheidungsge-

[83] $\sigma^2_{v,k}$ muss aus dem gestörten Gesamtsignal geschätzt werden; $\sigma^2_{x,k}$ ergibt sich implizit aus der $\sigma^2_{v,k}$-Schätzung. Siehe auch Abschnitt 4.1.3.

steuerten Ansatz (DDA)[84]. Mit dem DDA kann ein guter Kompromiss zwischen der Nutzsignalverzerrung und der Störgeräuschreduktion erreicht werden. Durch den geglätteten Verlauf des *a-priori*-SNR wird die Entstehung von Musical Noise weitgehend vermieden (vgl. Cappé, 1994). Dass die Glättung der spektralen Dämpfungsgewichte den Höreindruck positiv beeinflusst, wird zum Beispiel auch in Arslan et al. (1995) und Höldrich und Lorber (1997) festgestellt. Darüber hinaus kann der Höreindruck durch zusätzliches Begrenzen des *a-priori*-SNR nach unten und Übersubtraktion des geschätzten Störspektrums verbessert werden (s. Cappé, 1994; Malah et al., 1999). Als Alternative zum DDA schlägt Cohen (2004a, 2006) einen auf dem verallgemeinerten autoregressiven heteroskedastischen Zeitreihenmodell (GARCH)[85] beruhenden *a-priori*-SNR-Schätzer vor. Der wesentliche Vorteil des GARCH-Modells besteht darin, dass es die besonders stark wahrnehmbaren logarithmischen Verzerrungen verringert. Dadurch wird zum Beispiel ein höherer PESQ-Score[75] (ITU-T, 2001) als beim DDA erreicht. Ein weiterer von Cohen (2004b,c) vorgestellter Ansatz berücksichtigt zudem die zeitliche Korrelation der Spektralkoeffizienten. Dieser Ansatz führt auf einen akausalen *a-priori*-SNR-Schätzer, der zwischen Sprach-Onsets und impulshaften Störgeräuschen unterscheiden kann. Allerdings erfolgt dies wiederum auf Kosten einer erhöhten Signallatenz. Ein äußerst recheneffizienter, modifizierter DDA mit schnellem Ansprechverhalten wird in Abschnitt 4.4 dieser Arbeit vorgestellt.

MMSE-STSA-Schätzer.[73] Aus der Lösung des bayesschen Problems für komplexe, gaußverteilte Spektralkoeffizienten und den Gleichungen (4.24) und (4.25), ergibt sich der MMSE-STSA-Schätzer zu (vgl. Ephraim und Malah, 1984, Gl. 7)

$$\begin{aligned} G_{\text{MMSE-STSA},k} &= \Gamma(1.5)\frac{\sqrt{v_k}}{\gamma_k} M(-0.5, 1, -v_k) \\ &= \Gamma(1.5)\frac{\sqrt{v_k}}{\gamma_k} e^{-\frac{v_k}{2}} \left[(1+v_k)\, I_0\left(\frac{v_k}{2}\right) + v_k I_0\left(\frac{v_k}{2}\right)\right], \end{aligned} \tag{4.26}$$

[84] Siehe Abschnitt 4.3.2.

[85] *Generalized autoregressive conditional heteroscedasticity* (GARCH). Siehe auch Kreiß und Neuhaus 2006, Kap. 14.1.

$\Gamma(\cdot)$ bezeichnet die Gamma-Funktion (wobei $\Gamma(1.5) = \sqrt{\pi}/2$), $I_0(\cdot)$ und $I_1(\cdot)$ sind die modifizierten Bessel-Funktionen 0. und 1. Ordnung und der Faktor v_k ist wie folgt definiert:

$$v_k = \frac{\xi_k}{1+\xi_k}\gamma_k. \tag{4.27}$$

Mit dem MMSE-STSA-Schätzer aus Gl. (4.26) kann die Kurzzeit-Amplitude des Nutzsignals aus der Amplitude des gestörten Gesamtsignals geschätzt werden:

$$\hat{X}_k = G_{\text{MMSE-STSA},k}\, Y_k. \tag{4.28}$$

Ephraim und Malah (1984) und Cohen (2005) zeigen, dass die optimale MMSE-Schätzung der exponentiellen Phase des ungestörten Nutzsignals der exponentiellen Phase des gestörten Gesamtsignals entspricht.[69] Demzufolge lässt sich der MMSE-STSA-Schätzer, Gl. (4.26), auch unmittelbar auf die komplexen Spektralkoeffizienten anwenden:

$$\underline{\hat{X}}_k = G_{\text{MMSE-STSA},k}\, \underline{Y}_k. \tag{4.29}$$

Die spektralen Gewichte des Wiener-Filters lassen sich durch Umformen der Gleichungen (4.15) und (4.24) über das *a-priori*-SNR ausdrücken:

$$G_{\text{W},k} = \frac{\xi_k}{1+\xi_k}. \tag{4.30}$$

Eingesetzt in Gl. (4.26) ergibt sich folgender Zusammenhang zwischen MMSE-STSA-Schätzer und Wiener-Filter:

$$G_{\text{MMSE-STSA},k} = \frac{\sqrt{\pi}}{2} \cdot \sqrt{\frac{G_{\text{W},k}}{\gamma_k}} \cdot M\left(-0.5, 1, -\gamma_k\, G_{\text{W},k}\right). \tag{4.31}$$

MMSE-LSA-Schätzer[74]. Über den Ansatz $\hat{X}_k = \mathrm{E}\{\log X_k | Y_k\}$ lässt sich ein MMSE-Schätzer für die logarithmierte Amplitude herleiten (s. a. Ephraim und

Malah, 1985, Gl. 20):

$$G_{\text{MMSE-LSA},k} = \frac{v_k}{\gamma_k} \cdot \exp\left(\frac{1}{2}\int\limits_{v_k}^{\infty} \frac{e^{-t}}{t} dt\right) \tag{4.32}$$

$$= G_{\text{W},k} \cdot \exp\left(\frac{1}{2}\int\limits_{\gamma_k G_{\text{W},k}}^{\infty} \frac{e^{-t}}{t} dt\right). \tag{4.33}$$

Wird nun das Integral im Exponenten der Gl. (4.33) über eine endliche Reihenentwicklung approximiert (vgl. Zotter, 2004, Kap. 4.2.5; Noisternig et al., 2009)

$$\int\limits_{v_k}^{\infty} \frac{e^{-t}}{t} dt \approx -\ln\left(\frac{v_k}{e^{-C} + v_k}\right), \tag{4.34}$$

vereinfacht sich Gl. (4.33) zu

$$\tilde{G}_{\text{MMSE-LSA},k} = \sqrt{G_{\text{W},k} \cdot \left(G_{\text{W},k} + \frac{e^{-C}}{\gamma_k}\right)}, \tag{4.35}$$

wobei C die Euler-Mascheroni-Konstante[86] ist. Gl. (4.35) lässt sich wesentlich einfacher implementieren als Gl. (4.33), wobei zur Gewährleistung numerischer Stabilität $0 < v_k < \pi$ sein muss.

MMSE-SP-Schätzer.[81] Wolfe und Godsill (2001) leiten über den Ansatz $\hat{X}_k = \mathrm{E}\{X_k^2 | Y_k\}$ einen sehr einfachen Schätzer her

$$G_{\text{MMSE-SP},k} = \sqrt{G_{\text{W},k} \cdot \left(G_{\text{W},k} + \frac{1}{\gamma_k}\right)}, \tag{4.36}$$

mit dem sich, ähnlich wie in Gl. (4.35), der MMSE-LSA-Schätzer approximieren lässt. Der MMSE-SP-Schätzer unterscheidet sich von Gl. (4.35) nur durch einen Skalierungsfaktor e^{-C}. Diese Skalierung des *a-posteriori*-SNR kann als Übersub-

[86] Euler-Mascheroni oder auch Eulersche Konstante: $C = 0{,}577215664901\ldots$ (s. a. Abramowitz und Stegun, 1970, Kap. 1; Bronstein et al., 2013, Tab. 0.1).

Tabelle 4.1: Unterschiedliche Ansätze zur MMSE-Schätzung der Kurzzeit-Amplitude des Nutzsignals aus dem gestörten Gesamtsignal mit zugehörigen Ansatz-, Fehler- und Gewichtsfunktionen (vgl. Scalart und Filho, 1996; Cohen, 2005).

	Ansatz	**Fehlerfunktion**	**Gewichtsfunktion** G_k
STSA	$\hat{X}_k = \mathrm{E}\{X_k \vert Y_k\}$	$\mathrm{E}\left\{(\hat{X}_k - X_k)^2\right\}$	$\Gamma(1.5)\sqrt{\frac{G_{\mathrm{W},k}}{\gamma_k}}M\left(-0.5, 1, -\gamma_k\, G_{\mathrm{W},k}\right)$
LSA	$\hat{X}_k = \mathrm{E}\{\log X_k \vert Y_k\}$	$\mathrm{E}\left\{(\log \hat{X}_k - \log X_k)^2\right\}$	$G_{\mathrm{W},k} \exp\left(\frac{1}{2}\int_{\gamma_k\, G_{\mathrm{W},k}}^{\infty} \frac{e^{-t}}{t} dt\right)$
SP	$\hat{X}_k = \mathrm{E}\left\{X_k^2 \vert Y_k\right\}$	$\mathrm{E}\left\{(\hat{X}_k^2 - X_k^2)^2\right\}$	$\sqrt{G_{\mathrm{W},k}\left(G_{\mathrm{W},k} + \frac{1}{\gamma_k}\right)}$

traktion des geschätzten Störspektrums interpretiert werden. Die beiden Schätzer sind somit äquivalent. In den folgenden Betrachtungen wird vom einfacheren MMSE-SP-Schätzer ausgegangen. Tabelle 4.1 fasst die unterschiedlichen MMSE-Schätzer übersichtlich zusammen.

Die spektralen Gewichte des MMSE-SP-Schätzers lassen sich sehr anschaulich als Fläche über der (ξ_k, γ_k)-Ebene darstellen.[87] Diese wird in den Abbildungen 4.2 und 4.3 gezeigt. Zur einfacheren Interpretation der Funktionsweise des MMSE-SP-Schätzers lässt sich dessen Kennfläche in folgende Bereiche unterteilen:

- $(\gamma_k - 1) \ll 1/\xi_k \Rightarrow G_{\text{MMSE-SP},k} \approx \sqrt{\frac{G_{\mathrm{W},k}}{\gamma_k}}$
- $(\gamma_k - 1) \gg 1/\xi_k \Rightarrow G_{\text{MMSE-SP},k} \approx G_{\mathrm{W},k}$
- $(\gamma_k - 1) = 1/\xi_k \Rightarrow G_{\text{MMSE-SP},k} = \sqrt{2\,\frac{G_{\mathrm{W},k}}{\gamma_k}}$

Der Wertebereich des Wiener-Filters, Gl. (4.30), lässt sich auf ähnliche Weise darstellen und in folgende Bereiche unterteilen:

- $\xi_k \ll 1 \Rightarrow G_{\mathrm{W},k} \approx \xi_k$
- $\xi_k \gg 1 \Rightarrow G_{\mathrm{W},k} \approx 1$

[87] Siehe auch Wolfe und Godsill (2001), Zotter (2004, Kap. 4.2) und Noisternig et al. (2009). Renevey und Drygajlo (2001) und Plapous et al. (2006) stellen in sehr ähnlicher Weise das *a-priori*-SNR als Funktion des *a-posteriori*-SNR dar.

Werden die Teilbereiche des MMSE-SP-Schätzers und des Wiener-Filters wie folgt kombiniert, lässt sich die Kennfläche der spektralen Gewichte in vier Bereiche unterteilen. Diese werden in Abb. 4.2(a) veranschaulicht.

① $(\gamma_k - 1) \ll 1/\xi_k,\ \xi_k \ll 1 \Rightarrow G_{\text{MMSE-SP},k} \approx \sqrt{\xi_k/\gamma_k}$

② $(\gamma_k - 1) \ll 1/\xi_k,\ \xi_k \gg 1 \Rightarrow G_{\text{MMSE-SP},k} \approx \sqrt{1/\gamma_k}$

③ $(\gamma_k - 1) \gg 1/\xi_k,\ \xi_k \ll 1 \Rightarrow G_{\text{MMSE-SP},k} \approx \xi_k$

④ $(\gamma_k - 1) \gg 1/\xi_k,\ \xi_k \gg 1 \Rightarrow G_{\text{MMSE-SP},k} \approx 1$

Bei sehr kleinem *a-posteriori*-SNR – siehe Bereiche ① und ② – entspricht der MMSE-SP-Schätzer einer gewichteten Subtraktion der Leistungsdichtespektren. Der multiplikative Term $1/\sqrt{\gamma_k}$ wirkt dabei einer Verkleinerung der Signalamplitude entgegen. Bei großem *a-posteriori*-SNR – siehe Bereiche ③ und ④ – verhält sich der MMSE-SP-Schätzer hingegen wie das Wiener-Filter. Cappé (1994) erklärt dieses Verhalten wie folgt (s. a. Zotter, 2004, S. 85):

- Ist $(\gamma_k - 1)$ mit dem *a-priori*-SNR ξ_k im Einklang, wird das Leistungsdichtespektrum subtrahiert.

- Ist $(\gamma_k - 1)$ sehr viel kleiner als ξ_k, entspricht dies nicht den statistischen Erwartungen. Es kommt zu einer Anhebung der Signalleistung am Systemausgang.

- Ist $(\gamma_k - 1)$ sehr viel größer als ξ_k, entspricht dies ebenso nicht den statistischen Erwartungen. Es kommt zu einer vergleichsweise starken Dämpfung der Störgeräusche, die dem Wiener-Filter entspricht (vgl. Abb. 4.1).

In den folgenden Abschnitten wird die Funktionsweise des DDA und dessen Zusammenspiel mit dem MMSE-SP-Schätzer anhand der Kennflächen über der (ξ_k, γ_k)-Ebene diskutiert. Zur übersichtlicheren Schreibweise der mathematischen Ausdrücke, werden die MMSE-SP-Gewichte nicht mit $G_{\text{MMSE-SP},k}$ sondern nur kurz mit G_k bezeichnet.

4.3.2 Der entscheidungsgesteuerte Ansatz (DDA)

Der entscheidungsgesteuerte Ansatz (DDA)[88] kombiniert zwei grundlegende SNR-Schätzer zu einem neuen rekursiven Schätzer für das *a-priori*-SNR. Der erste dieser Schätzer bestimmt das instantane SNR

$$(\gamma_k - 1) = \frac{Y_k^2}{\sigma_{v,k}^2} - 1 = \frac{Y_k^2 - \sigma_{v,k}^2}{\sigma_{v,k}^2}, \tag{4.37}$$

welches vor Anwendung der Störgeräuschunterdrückung berechnet wird. In der Regel werden dabei nur positive Werte betrachtet

$$\mathrm{SNR}_{\mathrm{inst}} = \max(\gamma_k - 1, 0). \tag{4.38}$$

Das instantane SNR unterscheidet sich dann vom tatsächlichen SNR, wenn

- Nutzsignal $x(n)$ und Störsignal $v(n)$ stark korreliert sind,
- das Zeitfenster der Analyse im Vergleich zur Stationarität der Signale $x(n)$ und $v(n)$ zu kurz gewählt wird, oder
- nicht-stationäre Störungen auftreten.

Der zweite Schätzer bestimmt das rekonstruierte SNR, welches nach Anwendung der Störgeräuschreduktion berechnet wird:

$$\mathrm{SNR}_{\mathrm{rec}} = \frac{\hat{X}_k^2}{\sigma_{v,k}^2} = \gamma_k \, G_k^2. \tag{4.39}$$

Ist das SNR gering (z. B. $0 < \gamma_k < 2$), weist γ_k kleinere relative zeitliche Schwankungen[89] auf als $(\gamma_k - 1)$. Idealerweise, ergibt sich in diesem Fall eine gleichmäßig hohe Störd ämpfung. Aus Gl. (4.39) ist ersichtlich, dass bei konstanten G_k das $\mathrm{SNR}_{\mathrm{rec}}$ dem zeitlichen Verlauf von γ_k folgt. Demzufolge hat $\mathrm{SNR}_{\mathrm{rec}}$ im Fall eines geringen SNR einen zeitlich glatteren Verlauf als $\mathrm{SNR}_{\mathrm{inst}}$. Der DDA kombiniert nun $\mathrm{SNR}_{\mathrm{rec}}$ und $\mathrm{SNR}_{\mathrm{inst}}$ zu einem neuen rekursiven Schätzer für das

[88] *Decision directed approach* (DDA); s. Ephraim und Malah (1983, 1984).

[89] Die relativen zeitlichen Schwankungen $\Delta\gamma_k = 10 \log(\gamma_k(m)/\gamma_k(m-1))$ sind in der Regel stärker wahrnehmbar als zum Beispiel lineare zeitliche Schwankungen.

a-priori-SNR (vgl. Ephraim und Malah, 1983):

$$\xi_k(m) = (1-\alpha) \cdot \mathrm{SNR}_{\mathrm{inst}}(m) + \alpha \cdot \mathrm{SNR}_{\mathrm{rec}}(m-1). \tag{4.40}$$

Ein glatter zeitlicher Verlauf des geschätzten *a-priori*-SNR $\xi_k(m)$ vermindert das Entstehen von Musical Noise. Der Parameter $0 \ll \alpha < 1$ bestimmt nicht nur die Glättung des *a-priori*-SNR sondern auch das zeitliche Ansprechverhalten des Schätzers. Mit α kann somit das Verhältnis von erreichbarer Stördämpfung zu transienten Verzerrungen des Nutzsignals gesteuert werden. Cappé (1994) verwendet zur Aufbereitung gestörter Sprache $\alpha \approx 0{,}98$ (bei einer FFT Länge von 256 Punkten) und schlägt zudem vor, das *a-priori*-SNR mit einer unteren Schranke $\zeta = -25\,\mathrm{dB}$ zu begrenzen:

$$\xi_k(m) = \max\left[(1-\alpha) \cdot \mathrm{SNR}_{\mathrm{inst}}(m) + \alpha \cdot \mathrm{SNR}_{\mathrm{rec}}(m-1),\, \zeta\right]. \tag{4.41}$$

Weicht die Frequenzauflösung der tatsächlich verwendeten Filterbank von der in Cappé (1994) verwendeten FFT-Länge ab, muss zum Erhalt der Dynamik die untere Schranke ζ dementsprechend angepasst werden.

Um die Verzerrung transienter Klänge (wie z. B. Sprach-Onsets) möglichst gering zu halten, ist ein schnelles Ansprechen des DDA erforderlich. Hui (2000) schlägt zu diesem Zweck vor, den Wert von α adaptiv an das Signal anzupassen. Wird α kurzzeitig vergrößert, bekommt das schnell veränderliche *a-posteriori*-SNR bei der Schätzung von $\xi_k(m)$ mehr Gewicht. Dadurch passen sich die Spektralgewichte schneller an das aktuelle SNR an, wodurch transiente Klänge weniger stark bedämpft werden (d. h. die Transienten passieren das Filter annähernd unverändert). Die Adaption des Parameters α kann dabei zum Beispiel über eine VAD (vgl. Kap. 4.1.3) gesteuert werden.

Cappé (1994) zeigt, dass die dem DDA inhärente Verzögerung zu einem Bias der Schätzung der spektralen Gewichte führt, welches sich negativ auf die Störgeräuschunterdrückung auswirkt. Das von Plapous et al. (2006) vorgeschlagene TSNR-Verfahren[90] wirkt diesem Effekt entgegen, indem die Spektralgewichte in zwei Stufen geschätzt werden. In einem ersten Schritt, wird mit Hilfe des DDA

[90] *Two-step noise reduction technique* (TSNR)

ein vorläufiges *a-priori*-SNR geschätzt. Daraus wird, in einem zweiten Schritt, das *a-priori*-SNR für den Signalblock $(m+1)$ ermittelt, wodurch sich die Ansprechzeit verringert. Dies hat zum Vorteil, dass transiente Klänge unverzerrt übertragen werden und, im Gegensatz zum DDA, keine störenden Nachechos entstehen. Cohen (2005) verwendet ebenso ein zweistufiges Schätzverfahren. Dabei wird, in einem ersten Schritt (*propagation step*), aus den zeitlich aufeinanderfolgenden Signalblöcken ein vorläufiges *a-priori*-SNR, $\hat{\xi}'_k(m)$, geschätzt. Diese Schätzung kann sowohl kausal als auch akausal erfolgen (s. a. Cohen, 2004b). In einem zweiten Schritt (*update step*) wird dann aus dem vorläufigen $\hat{\xi}'_k(m)$ das *a-priori*-SNR geschätzt und die spektralen Gewichte berechnet. Der kausale Schätzer degeneriert zu einem DDA mit zusätzlicher frequenzabhängiger Gewichtung, die mit steigendem $\mathrm{SNR}_{\mathrm{inst}}$ monoton abfällt. Dies führt in Signalpausen zu einer höheren Dämpfung, wodurch das Musical Noise stärker unterdrückt wird; ist Nutzsignal vorhanden verringern sich die spektralen Gewichte und das Signal wird weniger stark verzerrt. Ein weiterer schnell ansprechender DDA wird in Kap. 4.4 vorgestellt.

4.3.3 Kombinierter MMSE-SP-DDA-Schätzer

Wird der MMSE-SP-Schätzer von Wolfe und Godsill, Gl. (4.36), mit dem DDA von Ephraim und Malah, Gl. (4.40), zu einem neuen Schätzer kombiniert, lassen sich dessen Eigenschaften wiederum sehr anschaulich durch den Verlauf der Kennlinien der spektralen Gewichte über der (ξ_k, γ_k)-Ebene darstellen. Durch Einsetzen der Teilbereiche ① bis ④ (siehe S. 104) in Gl. (4.40) ergeben sich die Konturen des DDA auf der Kennfläche des Schätzers. Folgende Wertebereiche sind dabei von besonderem Interesse:

1 $(\gamma_k - 1) \ll 1/\xi_k,\ \xi_k \ll 1 \Rightarrow G_k \approx \sqrt{\xi_k/\gamma_k}$

$$\begin{aligned} \xi_k(m) &\approx (1-\alpha) \cdot \max\left(\gamma_k(m) - 1{,}0\right) + \\ &\quad \alpha \cdot \xi_k(m-1). \end{aligned} \tag{4.42}$$

2 $(\gamma_k - 1) \ll 1/\xi_k,\ \xi_k \gg 1 \Rightarrow G_k \approx \sqrt{1/\gamma_k}$

$$\begin{aligned} \xi_k(m) &\approx (1-\alpha) \cdot \max\left(\gamma_k(m) - 1{,}0\right) + \alpha \\ &\approx \alpha. \end{aligned} \tag{4.43}$$

3 $(\gamma_k - 1) \gg 1/\xi_k,\ \xi_k \ll 1 \Rightarrow G_k \approx \xi_k$

$$\begin{aligned} \xi_k(m) &\approx (1-\alpha) \cdot \max\left(\gamma_k(m) - 1{,}0\right) + \\ &\quad \alpha \cdot \xi_k^2(m-1) \cdot \gamma_k(m-1) \\ &\approx (1-\alpha) \cdot \left(\gamma_k(m) - 1\right). \end{aligned} \tag{4.44}$$

4 $(\gamma_k - 1) \gg 1/\xi_k,\ \xi_k \gg 1 \Rightarrow G_k \approx 1$

$$\begin{aligned} \xi_k(m) &\approx (1-\alpha) \cdot \max\left(\gamma_k(m) - 1{,}0\right) + \\ &\quad \alpha \cdot \gamma_k(m-1) \\ &\approx \alpha \cdot \gamma_k(m-1). \end{aligned} \tag{4.45}$$

5 $(\gamma_k - 1) = 1/\xi_k,\ \xi_k \ll 1 \Rightarrow G_k = \sqrt{2 \cdot \xi_k/\gamma_k}$

$$\begin{aligned} \xi_k(m) &\approx (1-\alpha) \cdot \left(\gamma_k(m) - 1\right) + \\ &\quad 2\alpha \cdot \xi_k(m-1). \end{aligned} \tag{4.46}$$

Die Konturen des DDA auf der Kennfläche des MMSE-SP-Schätzers sind in Abbildung 4.2(b) dargestellt. Die gestrichelten Linien entsprechen der rekursiven Mittelung in den Bereichen 1 und 5, die durchgezogenen Linien der direkten Schätzung in den Bereichen 2 bis 4. Aus dieser Darstellung wird ersichtlich, dass die Konturlinien in den Übergänge zwischen unterschiedlichen DDA-Bereichen nicht stetig verlaufen.

Folgendes Beispiel verdeutlicht den Verlauf der Spektralgewichte auf der Kennfläche: Wird eine dem Nutzsignal additiv überlagerte Störung mit konstantem Signalpegel angenommen, kann mit Gl. (4.25) ganz einfach gezeigt werden, dass γ_k dem zeitlichen Verlauf der Amplitude des Nutzsignals folgt. Die sich daraus ergebenden Spektralgewichte G_k durchlaufen auf der Kennfläche eine Hysterese. Diese ist in Abb. 4.3(b) exemplarisch für ein Sprachsignal dargestellt.

Für den mit einem DDA kombinierten MMSE-SP-Schätzer lassen sich ganz allgemein folgende Eigenschaften ableiten (vgl. Cappé, 1994):

A) Teilbereiche mit rekursiver Mittelung

- ξ-Schätzung durch rekursive Mittelung. Im Teilbereich $\boxed{1}$, Gl. (4.42), wird ξ_k durch rekursive Mittelung des $\mathrm{SNR}_{\mathrm{inst}}$ geschätzt. Die rekursive Mittelung kann dabei durch eine Zeitkonstante τ_{avg} mit $\alpha = \exp\left(-1/(\tau_{\mathrm{avg}} \cdot f_A)\right)$ ausgedrückt werden, wobei $f_A = 1/T_A$ die Abtastrate bezeichnet.
- Konstant-ξ-Effekt. Nimmt ξ_k im Teilbereich $\boxed{1}$ einen konstanten Wert an,[91] können stark wahrnehmbare Artefakte auftreten. Ist $\xi_k \ll 1$ und konstant, wird die Amplitude am Systemausgang konstant gehalten. Dieser Effekt tritt dann auf, wenn $\left(Y_k^2(m)/\sigma_{v,k}^2 - 1\right)$ hinreichend klein ist und lässt sich wie folgt über die Gleichungen (4.42) und (4.30) herleiten:

$$\frac{Y_k^2(m)}{\sigma_{v,k}^2} - 1 \ll \frac{1}{\xi_k} \Rightarrow Y_k^2(m) \ll \frac{\sigma_{v,k}^2}{G_{\mathrm{W},k}} \approx \frac{\sigma_{v,k}^2}{\xi_k}.$$

  Mit den spektralen Gewichten $G_k \approx \sqrt{\xi_k/\gamma_k}$ im Teilbereich ① und Gl. (4.24) ergibt sich für die geschätzte Amplitude des Nutzsignals folgender Ausdruck:

$$\begin{aligned} \hat{X}_k(m) &\approx Y_k(m) \cdot \sqrt{\frac{\xi_k(m)}{\gamma_k(m)}} = Y_k(m) \cdot \sqrt{\frac{\sigma_{x,k}^2}{Y_k^2(m)}} \\ &\approx \sqrt{\sigma_{x,k}^2}. \end{aligned} \tag{4.47}$$

[91] Das *a-priori*-SNR ξ_k nimmt beispielsweise für große Zeitkonstanten τ_{avg} oder an den Grenzen des Wertebereichs annähernd konstante Werte an.

Dies führt mitunter dazu, dass zusätzliche breitbandige Störgeräusche auftreten, die üblicherweise als sehr störend wahrgenommen werden. Wird ξ_k mit einer unteren Schranke ζ begrenzt, tritt der Konstant-ξ-Effekt nur dann auf, wenn $Y_k^2(m) < \sigma_{v,k}^2/\zeta$.

- Instabilität der rekursiven Mittelung. Nimmt der Parameter α im Teilbereich [5], Gl. (4.46), Werte von $\alpha > 1/2$ an, wird die rekursive Mittelung des $\mathrm{SNR}_{\mathrm{inst}}$ instabil. Es kommt zu einem sprunghaften Anstieg von ξ_k.

B) Teilbereiche ohne rekursive Mittelung

- In den Teilbereichen [2], [3] und [4] findet keine rekursive Mittelung statt. Das *a-priori*-SNR wird direkt geschätzt. In Gl. (4.43) nimmt ξ_k einen konstanten Wert an, in Gl. (4.44) entspricht es dem um einen Faktor $(1-\alpha)$ verringerten $\mathrm{SNR}_{\mathrm{inst}}$ und in Gl. (4.45) folgt es mit einem Signalblock Verzögerung γ_k.

Der MMSE-SP-Schätzer mit DDA verhält sich somit nur in den Teilbereichen [1] und [4] entsprechend den Erwartungen. Eine verallgemeinerte Beschreibung der rekursiven Mittelung über eine Zeitkonstante ist ausschließlich für Gl. (4.42) zulässig. Aufgrund des Konstant-ξ-Effekts und der Unstetigkeiten der Konturen des DDA (siehe Abb. 4.2) ist das Verhalten des Schätzers nur schwer kontrollierbar. In Noisternig et al. (2009) wurde ein modifizierter DDA vorgestellt, der eine wesentlich bessere Kontrolle über die Störgeräuschunterdrückung bietet. Dessen Eigenschaften werden im folgenden Abschnitt diskutiert.

4.4 Modifizierter, schnell ansprechender DDA

Durch folgende einfache Modifikation des DDA, kann die Auswirkung der im vorherigen Abschnitt gezeigten Artefakte auf die Unterdrückung von Störgeräuschen minimiert werden:

$$\xi_k(m) = (1-\alpha)\cdot(\rho\cdot\mathrm{SNR}_{\mathrm{inst}}(m)+\zeta) + \alpha\cdot\mathrm{SNR}_{\mathrm{rec}}(m-1). \tag{4.48}$$

Der Parameter ζ kontrolliert den Grundgeräuschpegel (*noise floor parameter*; vgl. Cappé, 1994; Gustafsson et al., 1998). Der Parameter ρ steuert, inwieweit das

$\mathrm{SNR}_{\mathrm{inst}}$ unterschätzt wird. Die Eigenschaften des modifizierten DDA lassen sich, wie in den vorhergehenden Kapiteln gezeigt, wiederum sehr anschaulich durch den Verlauf der spektralen Gewichte auf der Kennfläche über der (ξ_k, γ_k)-Ebene darstellen. Die Kennfläche kann, ähnlich wie in Abschnitt 4.3.1, in folgende vier charakteristische Bereiche unterteilt werden:

$(\gamma_k - 1) \ll 1/\xi_k,\ \xi_k \ll 1,\ G \approx \sqrt{\xi_k/\gamma_k}$

$$\xi_k(m) \approx \rho(1-\alpha) \cdot \max(\gamma_k(m) - 1{,}0) + \alpha \cdot \xi_k(m-1). \tag{4.49}$$

$(\gamma_k - 1) \ll 1/\xi_k,\ \xi_k \gg 1,\ G \approx \sqrt{1/\gamma_k}$

$$\xi_k(m) \approx \alpha. \tag{4.50}$$

$(\gamma_k - 1) \gg 1/\xi_k,\ \xi_k \ll 1,\ G \approx \xi_k$

$$\xi_k(m) \approx \rho(1-\alpha) \cdot (\gamma_k(m) - 1). \tag{4.51}$$

$(\gamma_k - 1) \gg 1/\xi_k,\ \xi_k \gg 1,\ G \approx 1$

$$\xi_k(m) \approx \alpha \cdot \gamma_k(m-1). \tag{4.52}$$

Abbildung 4.3(a) zeigt die Konturen des modifizierten DDA auf der Kennfläche des MMSE-SP-Schätzers. In Abb. 4.3(b) wird am Beispiel gestörter Sprache gezeigt, dass der Verlauf der spektralen Gewichte G_k auf der Kennfläche eine Hysterese ausbildet. In jenem Bereich der Kennfläche, in dem das *a-priori*-SNR über eine rekursive Mittelung geschätzt wird, hängt die Form der Hysterese vom Verlauf der steigenden und fallenden Signalflanken ab. Die charakteristischen Bereiche  bis sind in Abb. 4.4 farblich gekennzeichnet. In Anlehnung an die

Arbeiten von Renevey und Drygajlo (2001) und Plapous et al. (2006), wird in den Abbildungen das *a-priori*-SNR auch als Funktion des *a-posteriori*-SNR dargestellt. Unter der Annahme, dass die Amplituden des Nutz- und des Störsignals bekannt sind, ergibt sich folgender Zusammenhang

$$\begin{aligned} Y_k &= \sqrt{X_k^2 + V_k^2 + 2X_k V_k \cos\phi_k} \quad \Big| \cdot \frac{1}{V_k^2}, \\ \gamma_k &= 1 + \xi_k + 2\sqrt{\xi_k}\cos\phi_k, \end{aligned} \tag{4.53}$$

wobei $0 < \phi_k < \pi$ die Phasendifferenz zwischen X_k und V_k bezeichnet. Beim klassischen Wiener-Filter wird zum Beispiel $\phi_k = \pi/2$ angenommen. In Abb. 4.4 ist die konstruktive Überlagerung von Nutzsignal und Störung ($\phi_k = 0$) mit einer roten, strichpunktierten Linie, die destruktive Überlagerung ($\phi_k = \pi$) mit einer blauen, gestrichelten Linie dargestellt. Diese Linien begrenzen somit jenen Bereich, in dem das 2-Tupel (γ_k, ξ_k) von der tatsächlichen Phasendifferenz ϕ_k abhängt (vgl. Plapous et al., 2006, Abschn. 3). Als Maßstab sind zudem die Konturen der Subtraktion der Leistungsdichtespektren (graue, gestrichelte Linie) und des Wiener-Filters (graue, strichpunktierte Linie) dargestellt.

Beim klassischen MMSE-SP-Schätzer mit DDA (vgl. Kap. 4.3.3) wird die Zeitkonstante der Mittelung, τ_{avg}, typischerweise so gewählt, dass diese in der Größenordnung der Kurzzeit-Stationarität des zu entstörenden Signals liegt. Ein wesentlicher Vorteil des modifizierten DDA Schätzers, Gl. (4.48), besteht darin, dass die Zeitkonstante wesentlich kleinere Werte annehmen kann. Ein typischer Wert für Sprache ist zum Beispiel $\tau_{\text{avg}} = 2\,\text{ms}$. Dies verbessert das dynamische Ansprechverhalten des Schätzers und führt dazu, dass bei der Störgeräuschunterdrückung die Transienten (wie zum Beispiel Sprach-Onsets) weniger stark verzerrt werden. Die geringeren Verzerrungen wirken sich wiederum positiv auf den Höreindruck aus.

Ein weiterer wesentlicher Vorteil des modifizierten DDA besteht in der unabhängigen Kontrolle über die Zeitkonstante der Mittelung und die Unterdrückung des Musical Noise. Der Parameter ρ steuert die Breite der sich auf der Kennfläche des Schätzers ausbildenden Hysterese der spektralen Gewichte und beeinflusst die Dämpfung des Musical Noise. Bei geeigneter Wahl dieses Parameters lassen sich

Unstetigkeiten im Verlauf der Hysterese (siehe Abschnitt 4.3.3) weitgehend vermeiden. Wird das instantane SNR zum Beispiel mit 15 dB unterschätzt, entspricht dies einem Parameter $\rho = 10^{-15/10}$.

Die untere Schranke ζ wird typischerweise so gewählt, dass der Grundgeräuschpegel am unteren Ende der Dynamik des Signals liegt. Ein üblicher Wert für Sprache liegt zum Beispiel bei $-25\,\mathrm{dB}$ (vgl. Lazarus et al., 2007, Kap. 6.1), d. h. einem Parameter $\zeta = 10^{-25/10}$. Der Konstant-ξ-Effekt (siehe Abschnitt 4.3.3) tritt somit nur dann auf, wenn $Y_k^2(m) < \sigma_{v,k}^2/\zeta$.

Aus dem Vergleich der Kennflächen in den Abbildungen 4.2(b) und 4.3(a) ist ersichtlich, dass beim modifizierten DDA die spektralen Gewichte mit steigendem $\mathrm{SNR}_{\mathrm{inst}}$ wesentlich kleiner sind als beim klassische DDA von Ephraim und Malah (siehe gestrichelte Linie).[92] Dies führt, ähnlich wie in Cohen (2005), zu einer stärkeren Unterdrückung des Musical Noise und einer geringeren Verzerrung des Nutzsignals. Allerdings ist beim hier vorgestellten modifizierten DDA kein zweistufiges Schätzverfahren notwendig. Abb. 4.3(b) zeigt die sich auf der Kennfläche des MMSE-SP-Schätzers ausbildende Hysterese der Spektralgewichte G_k.

4.5 Zusammenfassung

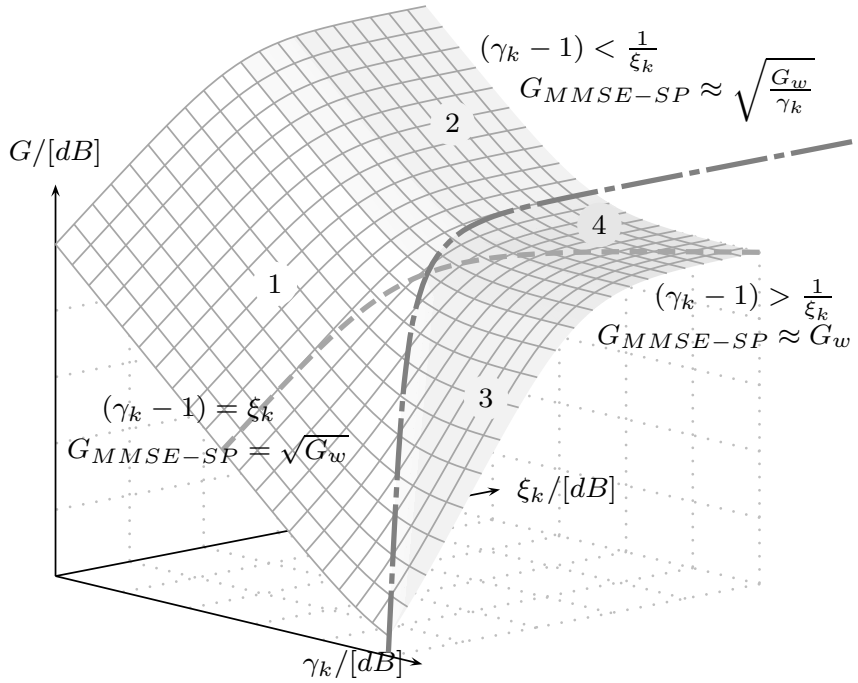
Durch Darstellung der spektralen Gewichte auf der Kennfläche über der (ξ_k, γ_k)-Ebene lassen sich die Eigenschaften der unterschiedlichen Amplitudenschätzer sowie des entscheidungsgesteuerten Ansatzes zum Schätzen des *a-priori*-SNR sehr anschaulich darstellen. Dies ermöglicht neue Einsichten in die grundlegende Funktionsweise des Ephraim-Malah-Filters zur einkanaligen Signalaufbereitung.

Bei den aus der Literatur bekannten Ansätzen der spektralen Subtraktion muss stets ein Kompromiss zwischen der Unterdrückung des Musical Noise und den transienten Verzerrungen des Nutzsignals getroffen werden. Der in dieser Arbeit vorgeschlagene modifizierte DDA erlaubt eine wesentlich flexiblere Handhabung der Unterdrückung des Musical Noise, ein schnelles Ansprechverhalten zur Reduktion der Verzerrungen transienter Klänge sowie eine Verringerung des Konstant-ξ-Effekts. Die Ergebnisse eines informellen Hörtests zeigen eine tendenziell bessere Beur-

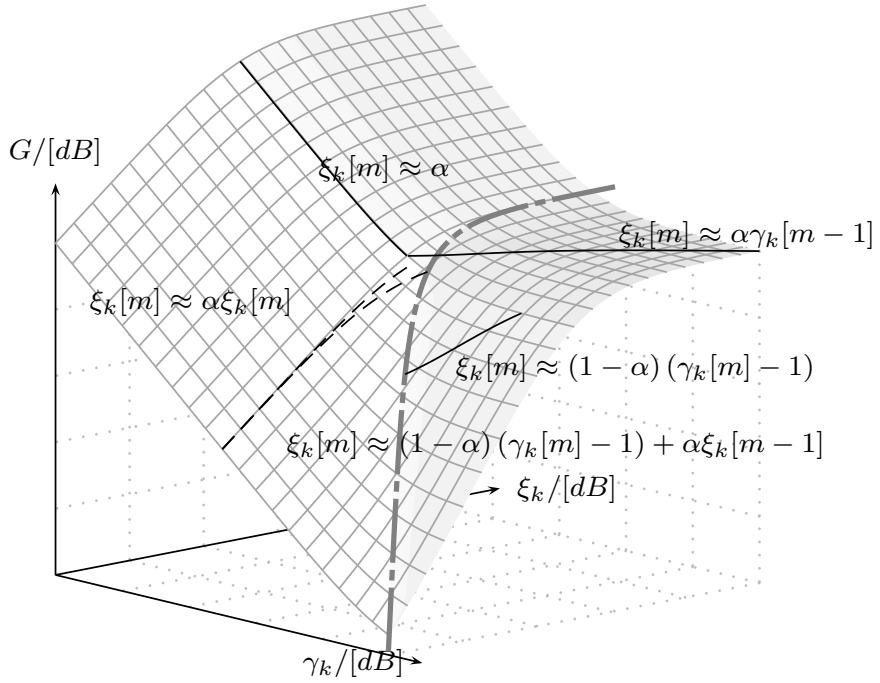
[92] Hierbei ist zu beachten, dass in den Abbildungen das *a-posteriori*-SNR γ_k und nicht wie in der Literatur oft üblich das $\mathrm{SNR}_{\mathrm{inst}} = \max(\gamma_k - 1, 0)$ dargestellt wird.

teilung des Höreindrucks bei Verwendung des modifizierten DDA gegenüber dem DDA von Ephraim und Malah. Zur Spektraltransformation wurde die in Kap. 3 vorgestellte auditive Analyse-Synthese-Filterbank verwendet.

Aktuelle Studien umfassen die Analyse alternativer Schätzverfahren (wie z. B. Cohen, 2004c; Hasan et al., 2004; Lotter und Vary, 2005; Plapous et al., 2006) und den Vergleich mit dem modifizierten DDA.

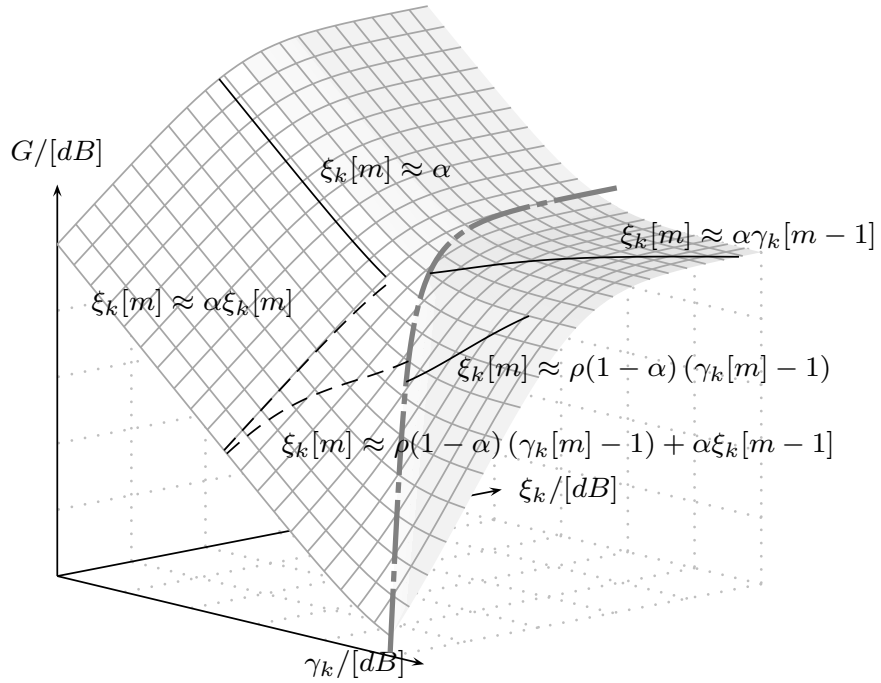


(a) Teilbereiche der Kennfläche des MMSE-SP-Schätzers (siehe S. 104) sowie die Kontur der Subtraktion der Leistungsdichtespektren (gestrichelte Linie) und die Kontur des Übergangs von $G_{\text{MMSE-SP},k} \approx \sqrt{G_{\text{W},k}/\gamma_k}$ zu $G_{\text{MMSE-SP},k} \approx G_{\text{W},k}$ (strichpunktierte Linie).

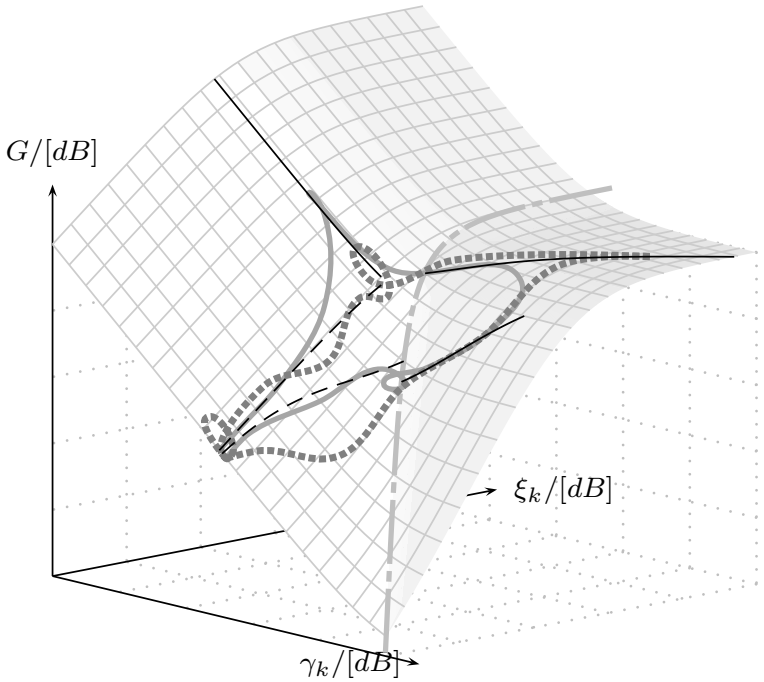


(b) Konturen des DDA auf der MMSE-SP Kennfläche. Mittelwerte der dynamischen (gestrichelte Linien) und der statischen Zusammenhänge (durchgezogene Linien) von γ_k und ξ_k.

Abbildung 4.2: Kennfläche des MMSE-SP-Schätzers in Abhängigkeit von γ_k und ξ_k.



(a) Konturen des modifizierten DDA auf der MMSE-SP Kennfläche. Mittelwerte der dynamischen (gestrichelte Linien) und der statischen Zusammenhänge (durchgezogene Linien) von γ_k und ξ_k.



(b) Schematische Darstellung zweier experimentell ermittelter Hystereseschleifen (gestrichelte und durchgezogene graue Linien).

Abbildung 4.3: Kennfläche des MMSE-SP-Schätzers mit modifiziertem DDA in Abhängigkeit von γ_k und ξ_k.

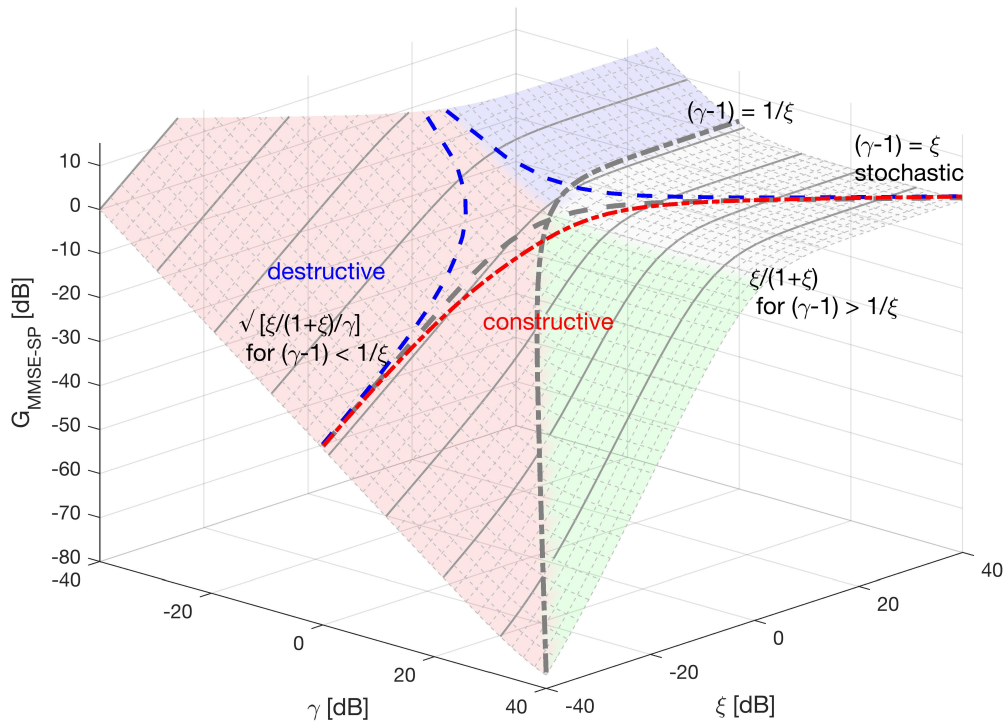


Abbildung 4.4: Kennfläche des MMSE-SP-Schätzers mit modifiziertem DDA in Abhängigkeit von γ_k und ξ_k. Die blaue gestrichelte Linie kennzeichnet die destruktive Überlagerung ($\phi = \pi$), die rote strichpunktierte Linie die konstruktive Überlagerung ($\phi = 0$) von Nutzsignal und Störung. Zudem sind die Konturen der Subtraktion der Leistungsdichtespektren (graue gestrichelte Linie) und des Wiener-Filters (graue strichpunktierte Linie) dargestellt. Die charakteristischen Teilbereiche des modifizierten DDA Schätzers (siehe Abschnitt 4.4) sind farblich gekennzeichnet: ◇1 = rot, ◇2 = grün, ◇3 = blau und ◇4 = grau.

5

Breitbandige Signalaufbereitung in mehrkanaligen Mikrofonanwendungen: Robuste Beamformer

Dieses Kapitel befasst sich mit der breitbandigen Signalaufbereitung in mehrkanaligen Mikrofonanwendungen. Ziel ist eine möglichst hohe und breitbandige richtungsabhängige Verstärkung mit einer kleinen Sensorgruppen-Apertur[93] und einer geringen Anzahl an Sensoren. Zudem soll eine möglichst hohe Robustheit gegenüber Bauteiltoleranzen, Fehlpositionierungen und reflexionsbehafteter Schallwellenausbreitung erreicht werden. Die räumlich-zeitliche Filterung des Schallfeldes erfolgt dabei durch zeitvariante adaptive Beamformer.[94]

Die störsignaldämpfende Wirkung eines Mikrofonarray-Beamformers wird durch Ausrichten des Maximums der Empfindlichkeit (d. h. der Hauptkeule) auf eine bestimmte Vorzugsrichtung erreicht. Störungen und Interferenzen aus anderen Raumrichtungen werden, in Abhängigkeit von ihrer Einfallsrichtung und räumlich-zeitlichen Korrelation, abgeschwächt. Für eine aus Vorzugsrichtung einfallende

[93] Die Apertur (lat. *apertura*: Öffnung) beschreibt die auf die einfallende Wellenfront projizierte wirksame Fläche des Arrays.

[94] Das in diesem Kapitel vorgestellte Verfahren zur breitbandigen Signalaufbereitung (d. h. zur Reduktion von Störgeräuschen) in mehrkanaligen Mikrofonanwendungen wurde im Rahmen eines Kooperationsprojekts des Instituts für Elektronische Musik und Akustik (IEM) der Universität für Musik und darstellende Kunst Graz mit der AKG Acoustics GmbH Wien entwickelt.

Schallwelle werden die Filtergewichte in den Mikrofonkanälen so bestimmt, dass diese die aufgrund der unterschiedlichen Laufzeiten der Schallwellen entstehenden Phasenunterschiede ausgleichen. Die anschließende Summation führt, bei einer bestimmten Frequenz, zu einer zeitlich kohärenten Überlagerung der Mikrofonsignale. Die aus anderen Raumrichtungen einfallenden Schallwellen sowie die nicht bzw. schwach korrelierten Störsignale werden nicht kohärent überlagert und, im Vergleich zum Nutzsignal, weniger stark verstärkt. Dieser Ansatz wird in der Literatur als *Delay-and-Sum*-Beamformer (DAS-BF) bezeichnet (vgl. Van Veen und Buckley, 1988).[95] Bei der Berechnung der Filtergewichte wird meist davon ausgegangen, dass sich die Schallquelle im Fernfeld des Arrays befindet. Mit Gl. (5.5) kann abgeschätzt werden, ob die Fernfeldbedingung erfüllt ist. Die von Abhayapala et al. (2000) und Doclo und Moonen (2003b) vorgestellten Ansätze sind sowohl auf das Nahfeld als auch auf das Fernfeld anwendbar. In der vorliegenden Arbeit liegt das Hauptaugenmerk auf Mikrofonarrays mit möglichst kleiner Apertur. Aus diesem Grund wird in den folgenden Betrachtungen angenommen, dass sich die Schallquellen im Fernfeld befinden. Nahfeldansätze werden nicht diskutiert. Zur weiterführenden Studie der Nahfeldansätze sei auf die Literatur verwiesen.[96] Eine umfassende Übersicht über unterschiedliche Methoden findet sich zum Beispiel in Havelock et al. (2008, Kap. 59).

Die Richtwirkung des DAS-BF ist stark frequenzabhängig, da sich nur schmalbandige Signale exakt überlagern. Beim *Filter-and-Sum*-Beamformer (FAS-BF) werden die Mikrofonsignale vor der Summation mit FIR-Filtern[97] gefiltert. Dies erlaubt einen breitbandigen Entwurf der Richtcharakteristik des Arrays. FAS-BF unterscheiden sich in der Art und Weise, wie die Filterkoeffizienten hergeleitet werden. Hierbei wird ganz allgemein zwischen signalunabhängigen und signalabhängigen Ansätzen unterschieden. Bei den signalunabhängigen Ansätzen werden zur Herleitung der Filterkoeffizienten bestimmte Annahmen betreffend der Eigenschaften des Schallfeldes getroffen (vgl. Abschnitt 5.1.3). Diese Ansätze umfassen breitban-

[95] Siehe auch Schelkunoff (1943), Dudgeon (1977), Flanagan et al. (1985), Flanagan et al. (1991), Brandstein und Ward (2001), Van Trees (2002, Kap. 2)

[96] Siehe auch Maynard et al. (1985), Veronesi und Maynard (1987), Maynard (1997), Williams (1999, Kap. 7), Zheng et al. (2004), Bai et al. (2013).

[97] Nichtrekursive Filter (*finite impulse response, FIR*) sind gekennzeichnet durch eine Impulsantwort mit garantiert endlicher Länge.

dige Beamformer[98], superdirektive Beamformer[99], Least-Squares Ansätze[100] sowie Maximum-SNR-Beamformer[101]. Signalabhängige Beamformer schätzen die statistischen Eigenschaften des Schallfeldes aus den Mikrofonsignalen und adaptieren die Filterkoeffizienten solange, bis sich eine optimale Lösung einstellt. In der Praxis sind der Frost-Beamformer (vgl. Frost, 1972) und der *Generalised Sidelobe Canceller* (GSC, s. Griffiths 1977)[102] am weitesten verbreitet. Frost-Beamformer und GSC können über einen *Linearly-Constrained Minimum Variance* (LCMV) Ansatz (vgl. Capon et al., 1967; Buckley, 1987; Van Trees, 2002, Kap. 6.7) hergeleitet werden und sind, wie Griffiths und Jim (1982) zeigen, unter bestimmten Voraussetzungen äquivalent. Der LCMV-Ansatz minimiert die Varianz des Signals am Ausgang des Systems, unter der Nebenbedingung unverzerrter Wiedergabe des Signals aus Vorzugsrichtung. Beim GSC wird der LCMV-Ansatz in zueinander orthogonalen Unterräumen formuliert. Dies führt auf eine Lösung ohne Zwangsbedingung, die sich in der Praxis relativ einfach implementieren lässt. Dazu wird einem Preprozessor, bestehend aus einem fixen Beamformer (FBF) und einer Blockiermatrix (BM), ein aktiver Störgeräuschunterdrücker (ANC)[103] nachgeschaltet. Theoretisch wird mit dem GSC eine signifikante Reduktion der Störgeräusche bei einer sehr geringen Verzerrung des Nutzsignals erreicht.[104] Aufgrund von Fertigungstoleranzen (wie z. B. Fehlpositionierungen der Mikrofone), Bauteiltoleranzen (wie z. B. Pegel- und Phasenunterschiede) und Raumreflexionen, wird das Nutzsignal in der BM nicht vollständig unterdrückt (siehe z. B. Jablon, 1986a, 1987). Das residuale Nutzsignal in den Referenzkanälen der ANC führt zu einer Fehlanpassung der adaptiven Prozesse, die eine Verzerrung des entstörten Nutzsignals zur Folge hat. Nordebo et al. (1994) und Hoshuyama und Sugiyama (1996)[105] verwenden

[98] Siehe z. B. Sydow (1994), Doclo und Moonen (2003a) und Chen und Ser (2009).

[99] Siehe z. B. Cox et al. (1986), Bitzer und Simmer (2001), Bitzer et al. (2001) und Doclo und Moonen (2007).

[100] Siehe z. B. Algazi und Suk (1975) und Doclo und Moonen (2003b).

[101] Siehe z. B. Araki et al. (2007), Warsitz und Haeb-Umbach (2007), Kolossa et al. (2008) und Tanaka und Shiono (2014).

[102] Siehe auch Griffiths und Jim (1982) sowie Buckley und Griffiths (1986, Kap. 6.7). Der ursprüngliche Ansatz geht auf Applebaum (1976) und Applebaum und Chapman (1976) zurück.

[103] *Active noise canceller* (ANC) oder auch *Multiple input canceller* (MIC).

[104] Siehe z. B. Jablon (1986b), Bitzer et al. (1999b), Benesty et al. (2007) und Benesty et al. (2008b, Kap. 47).

[105] Siehe auch Hoshuyama et al. (1997, 1998, 1999); Hoshuyama und Sugiyama (1996, 1999, 2001).

eine adaptive BM, um das residuale Nutzsignal weitgehend zu unterdrücken, und adaptieren die ANC nur in Perioden, in denen kein Nutzsignal vorhanden ist (s. a. Van Compernolle, 1990). Die Steuerung der Adaption (AMC)[106] hat somit einen wesentlichen Einfluss auf die Robustheit der GSC-Implementierung und wird in Abschnitt 5.2.2 dieser Arbeit ausführlicher diskutiert.

Wie in Kap. 4 (S. 82) erläutert wurde, ist die mit LCMV/MVDR-Beamformern[107] für schmalbandige Signale erreichte SNR-Verbesserung optimal im Sinne einer ML-Schätzung. Sind die Signale hingegen breitbandig, ist die SNR-Verbesserung nicht mehr optimal (vgl. Monzingo und Miller, 1980; Simmer et al., 2001). Hier lässt sich mit einem mehrkanaligen Wiener-Filter (MWF) eine optimale Lösung im Sinne des MMSE erreichen (vgl. Cornelis et al., 2011). Das MWF kann als MVDR-Beamformer mit einem in Serie geschalteten einkanaligen Wiener-Filter interpretiert werden.[108] Dies motiviert den in dieser Arbeit verfolgten Ansatz, die residualen Störgeräusche am Ausgang eines Mikrofonarrays durch Nachschalten eines einkanaligen Geräuschreduktionsfilters zu reduzieren.[109] Um die Verzerrung der Transienten bei der Störgeräuschunterdrückung möglichst gering zu halten, wird zur Nachfilterung das in Kap. 4.3 vorgestellte modifizierte Ephraim-Malah-Filter verwendet.

Bei der Herleitung optimaler Beamformer wird meist vorausgesetzt, dass die Hauptkeule durch zeitliche Vorentzerrung (*pre-steering*) auf die Schallquelle ausgerichtet ist (siehe z. B. McCowan und Bourlard, 2002). Die (adaptiven) Filter in den Mikrofonkanälen dienen somit ausschließlich der Keulenformung (*beamforming*). Eine fehlerhafte Ausrichtung der Hauptkeule führt zu unerwünschten Verzerrungen des Nutzsignals. Generell stehen die zur zeitlichen Vorentzerrung benötigten Verzögerungszeiten in keinem ganzzahligen Verhältnis zur Abtastperiode. Hier kann durch eine fraktionale Verzögerung mit Allpass-Interpolationsfiltern (vgl. Laakso et al., 1996) eine kohärente Überlagerung der Sensorsignale für eine bestimmte Vorzugsrichtung erreicht werden. Der räumlichen Ausdehnung realer Schallquellen und

[106] *Adaptation mode control* (AMC).

[107] *Linearly constrained minimum variance* (LCMV) und *minimum variance distortionless response* (MVDR) Beamformer.

[108] Vgl. Edelblute et al. (1967), Brooks und Reed (1972), Monzingo und Miller (1980), Simmer et al. (2001), Van Trees (2002, Kap. 6) und Herbordt (2005).

[109] Vgl. Zelinski (1988), Fischer und Simmer (1996), Marro et al. (1998), Bitzer et al. (1999a, 2001), Simmer et al. (2001), McCowan und Bourlard (2002, 2003) und Hendriks et al. (2009).

kleineren Bewegungen (wie z. B. Kopfdrehungen eines Sprechers im Fahrzeug) kann mit einer Verbreiterung der Hauptkeule entgegnet werden. Aus diesem Grund wird in dieser Arbeit der von Hoshuyama et al. (1999) vorgeschlagene robuste adaptive Beamformer verwendet, mit dem sich die Breite der Hauptkeule vergleichsweise einfach modifizieren lässt (vgl. Abschnitt 5.2.1). Ist die Position der Schallquelle nicht bekannt, muss diese aus den Mikrofonsignalen geschätzt werden. Algorithmen zur Lokalisation von Schallquellen und zur zeitvarianten Nachführung der Hauptkeule werden in dieser Arbeit nicht behandelt. Es sei an dieser Stelle auf die weiterführende Literatur verwiesen.[110]

Kapitel 5.1 gibt einen kurzen Überblick über die mathematischen Grundlagen zur Beschreibung von Schallfeldern und das dem Mikrofonarray-Beamforming zugrunde liegende Signalmodell. In Kapitel 5.2 werden die Eigenschaften des robusten adaptiven Beamformers diskutiert und es wird gezeigt, wie die Robustheit über eine modifizierte AMC verbessert werden kann.

5.1 Problemformulierung und Signalmodell

Dieser Abschnitt fasst die mathematische Beschreibung von Schallfeldern kurz zusammen und diskutiert das dem Mikrofonarray-Beamforming zugrunde liegende Signalmodell. Zudem werden etablierte und weit verbreitete Maße zur Beurteilung der Leistungsfähigkeit von Mikrofonarrays vorgestellt, die beim Entwurf modaler Beamformer (siehe Kap. 6) aufgegriffen werden. Die folgenden Betrachtungen beruhen hauptsächlich auf den Arbeiten von Ziomek (1994), Van Trees (2002) und Herbordt (2005).

[110] Siehe z. B. Piersol (1981), Brandstein (1995), DiBiase et al. (2001), Di Claudio und Parisi (2001), Strobel et al. (2001), Huang und Benesty (2003), Chen et al. (2003a,b, 2004, 2005), Benesty et al. (2004), Zotkin und Duraiswami (2004), Huang et al. (2006, Kap. 9), Benesty et al. (2008a, Kap. 9) und Benesty et al. (2008b, Kap. 5.1).

5.1.1 Beschreibung der Schallwellenausbreitung

Der durch das Schallfeld eines isotropen[111] Strahlers in einem homogenen, verlustlosen und dispersionsfreien Medium im Aufpunkt $\mathbf{r}$ zum Zeitpunkt t hervorgerufene Schalldruck $p(\mathbf{r},t)$ kann aus der skalaren, linearen Wellengleichung

$$\nabla^2 p(\mathbf{r},t) = \frac{1}{c^2}\frac{\partial^2 p(\mathbf{r},t)}{\partial t^2} \tag{5.1}$$

abgeleitet werden (vgl. Ziomek 1994, Kap. 2; Kuttruff 2000, Kap. 1). Hierbei beschreibt c die Schallgeschwindigkeit und ∇^2 den Laplace-Operator. Ein Punkt $\mathbf{r}$ in einem dreidimensionalen, rechtsseitigen, orthogonalen Koordinatensystem ist durch die kartesischen Koordinaten $\mathbf{r} = [x, y, z]^T$ bzw. die Kugelkoordinaten $\mathbf{r} = [r, \theta, \phi]^T$, mit Radius $0 \leqslant r < \infty$, Azimutwinkel $0 \leqslant \phi < 2\pi$ und Polarwinkel $0 \leqslant \theta \leqslant \pi$, eindeutig definiert. Es gilt: $x = r\sin\theta\cos\phi$, $y = r\sin\theta\sin\phi$, $z = r\cos\theta$ (siehe z. B. Bronstein et al., 2013, Kap. 1.7.9.3).

Die einfachste Lösung der Wellengleichung ist die monochromatische[112] ebene Welle, die über die Gleichung

$$p(\mathbf{r},t) = Pe^{i(\omega_0 t - \mathbf{k}_0^T \mathbf{r})} \tag{5.2}$$

beschrieben werden kann. P ist die Amplitude, ω_0 die Kreisfrequenz der zeitlichen Schwingung und $i = \sqrt{-1}$.[113] Der Wellenzahlvektor $\mathbf{k}_0$ lässt sich wie folgt aus der Wellenlänge λ_0 und dem Einheitsvektor $\mathbf{u}(\theta,\phi)$ in Ausbreitungsrichtung (θ,ϕ) ableiten:

$$\mathbf{k}_0 = k_0\mathbf{u}(\theta,\phi) = \frac{\omega_0}{c}\mathbf{u}(\theta,\phi) = \frac{2\pi}{\lambda_0}\mathbf{u}(\theta,\phi). \tag{5.3}$$

Die Wellenzahl k_0 kann als räumliche Frequenz in die auf den Ursprung des Koordinatensystems bezogene Ausbreitungsrichtung interpretiert werden. Das

[111] Isotropie (von griech. *isos*: gleich und *tropos*: Drehung, Richtung) bezeichnet die Unabhängigkeit einer Eigenschaft von der Richtung.

[112] Monochromatisch (von griech. *mono chromos*: eine Farbe), als Lösung bei Betrachtung nur einer Kreisfrequenz ω_0.

[113] In diesem Abschnitt wird die imaginäre Zahl mit $i = \sqrt{-1}$ bezeichnet, da j für die sphärische Bessel-Funktion verwendet wird. Andere Kapitel dieser Arbeit verwenden hingegen $j = \sqrt{-1}$.

Skalarprodukt $\mathbf{k}_0^T\mathbf{r}$ ist ein Maß für die durch die Laufzeit verursachte zeitliche Verzögerung der Welle im räumlichen Abtastpunkt $\mathbf{r}$.

Die monochromatische Kugelwelle ist eine weitere Lösung der Wellengleichung und beschreibt das Feld einer sich im Koordinatenursprung befindenden isotropen Punktschallquelle:

$$p(r,t) = \frac{P}{r} e^{i(\omega_0 t - k_0 r)}. \tag{5.4}$$

Aus Gl. (5.4) ist klar ersichtlich, dass die Amplitude der Kugelwelle mit der Distanz r des Beobachtungspunktes $\mathbf{r}$ zur Quellposition hyperbolisch abnimmt.

Gl. (5.2) beschreibt das Schallfeld einer Punktschallquelle im Fernfeld eines Sensorarrays, Gl. (5.4) das Nahfeld. Mit folgender Bedingung kann abgeschätzt werden, ob sich eine Schallquelle im Nah- oder im Fernfeld eines Arrays befindet (vgl. Goodman, 1968; Mailloux, 1994):

$$r > \frac{2D^2}{\lambda_0}. \tag{5.5}$$

Die bestimmenden Größen sind der Abstand r der Schallquelle zum Mittelpunkt des Arrays, die größte räumliche Ausdehnung D des Arrays, d. h. die maximale Distanz der Mikrofone zueinander, sowie die Wellenlänge λ_0. Bei den folgenden Betrachtungen unterschiedlicher Mikrofonarray-Beamformer werden, sofern nicht anderweitig angegeben, Fernfeldbedingungen vorausgesetzt.

Durch Transformation der linearen Wellengleichung (5.1) in den Frequenzbereich, ergibt sich die homogene Helmholtz-Gleichung:

$$\nabla^2 p(\mathbf{r},\omega) + k^2 p(\mathbf{r},\omega) = 0, \quad \mathbf{r} \in V, \tag{5.6}$$

wobei, je nachdem ob zwei- oder dreidimensionale Probleme betrachtet werden, V eine Teilmenge von $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ ist. In den folgenden Abschnitten werden die in dieser Arbeit verwendeten Lösungen der Helmholtz-Gleichung kurz zusammengefasst. Eine ausführliche Darstellung ist zum Beispiel in Fazi et al. (2012) zu finden. Das

Gebiet V wird zur Vereinfachung als Kreis bzw. Kugel mit Radius R angenommen und bildet somit die Grundlage der in Kap. 6.1 beschriebenen modalen Beamformer. Zum Entwurf des in Kap. 5.2 vorgestellten robusten adaptiven Beamformers werden hingegen ebene Wellen nach Gl. (5.2) verwendet.

A Integraldarstellung

Befinden sich die Schallquellen im Fernfeld eines Arrays, lässt sich das Schallfeld als gewichtete Überlagerung von aus allen Richtungen $\hat{\mathbf{y}}$ einfallenden ebenen Wellen darstellen. Dieser Vorgang kann als linearer Integraloperator H aufgefasst werden, der wie folgt definiert ist:

$$p(\mathbf{r},\omega) = (H\varphi)(\mathbf{r},\omega) := \int_{\Omega} \exp\left(-i\frac{\omega}{c}\mathbf{r}\cdot\hat{\mathbf{y}}\right)\varphi(\hat{\mathbf{y}},\omega)dS(\hat{\mathbf{y}}). \tag{5.7}$$

Gl. (5.7) wird als Herglotz-Wellenfunktion (HWF) mit Kern bzw. Dichte φ bezeichnet, wobei das Gebiet Ω im 2D-Fall dem Einheitskreis $\mathbb{S}^1$ und im 3D-Fall der Einheitssphäre $\mathbb{S}^2$ entspricht (vgl. Colton und Kress, 2013, Kap. 3.3).

Zur Darstellung von Schallquellen im Nahfeld wird somit ein weiterer Integraloperator benötigt. Der durch

$$p(\mathbf{r},\omega) = (S\psi)(\mathbf{r},\omega) := \int_{\partial V} G(\mathbf{y},\mathbf{r},\omega)\psi(\mathbf{y},\omega)dS(\mathbf{y}) \tag{5.8}$$

definierte lineare Integraloperator S mit Dichte ψ wird Einfachschichtpotential[114] genannt (vgl. Colton und Kress, 2013, Kap. 3.1). Das Einfachschichtpotential kann als auf dem Rand ∂V des Gebiets V kontinuierlich verteilte Linienschallquellen (2D) bzw. Punktschallquellen (3D) interpretiert werden. Die Greensche Funktion in Gl. (5.8) ist wie folgt definiert (siehe z. B. Duffy, 2015, Kap. 6):

$$G(\mathbf{y},\mathbf{r},\omega) = \begin{cases} \exp(i\frac{\omega}{c}|\mathbf{r}-\mathbf{y}|)(4\pi|\mathbf{r}-\mathbf{y}|)^{-1} & \text{3D} \\ \frac{i}{4}H_0(\frac{\omega}{c}|\mathbf{y}-\mathbf{r}|) & \text{2D} \end{cases} \tag{5.9}$$

[114] *Single-layer potential* (SLP).

H_0 ist die Hankel-Funktion erster Art der Ordnung Null. Unter Annahme eines kreis- bzw. kugelförmigen Gebiets V mit Rand ∂V vereinfacht sich Gl. (5.8) zu

$$(S\psi)(\mathbf{r},\omega) = \int_\Omega G(R\hat{\mathbf{y}},\mathbf{r},\omega)\psi(\hat{\mathbf{y}},\omega)R^\alpha dS(\hat{\mathbf{y}}), \tag{5.10}$$

wobei $\alpha = 1$ (2D) bzw. $\alpha = 2$ (3D) gilt. Mit dem Einfachschichtpotential lassen sich alle Schallfelder beschreiben die Gl. (5.6) erfüllen (vgl. Fazi, 2010, Kap. 2). Da die Dichtefunktionen φ und ψ kontinuierlich sind, müssen diese für die praktische Implementierung diskretisiert werden. Dies führt mitunter zu Abbildungsfehlern wie zum Beispiel dem räumlichen Aliasing (vgl. Williams, 1999, Kap. 5.3).

B Reihendarstellung

Das Schallfeld kann auch als lineare Überlagerung stehender Zylinderwellen (2D) bzw. Kugelwellen (3D) der Ordnung $n \in \mathbb{N}$ dargestellt werden:

$$\begin{aligned} \Phi_n(\mathbf{r},\omega) &= j_\nu(kr)Y_\nu^\mu(\theta_r,\phi_r), && \text{3D} \\ \Phi_n(\mathbf{r},\omega) &= J_\nu(kr)\exp(i\nu\phi_r)(2\pi)^{-1/2}. && \text{2D} \end{aligned} \tag{5.11}$$

Dabei ist J_ν die ν-te Bessel-Funktion, j_ν die ν-te sphärische Bessel-Funktion und Y_ν^μ die ν-te harmonische Kugelfunktion, mit $\nu \in \mathbb{N}_0$ und $-\nu \leq \mu \leq \nu$ (s. a. Kap. 6 und Williams 1999, Kap. 4 und 6). Für die Parameter n, ν und μ gilt im 2D-Fall die Beziehung

$$\nu = \begin{cases} (n-1)/2, & n \text{ ungerade} \\ -n/2, & n \text{ gerade} \end{cases} \tag{5.12}$$

und wir erhalten im 3D-Fall

$$\begin{aligned} \nu &= \lceil \sqrt{n} - 1 \rceil, \\ \mu &= n - 1 - \nu - \nu^2, \end{aligned} \tag{5.13}$$

wobei $\lceil \cdot \rceil$ die Aufrundungsfunktion bezeichnet.

Jedes beliebige Schallfeld, welches Gl. (5.6) erfüllt, kann durch folgende Reihe dargestellt werden (s. a. Williams, 1999, Kap. 4 und 6):

$$p(\mathbf{r},\omega) = \sum_{n=1}^{\infty} \Phi_n(\mathbf{r},\omega) c_n(\omega). \tag{5.14}$$

Das Schallfeld im Gebiet V wird durch die Koeffizienten c_n vollständig beschrieben. Die Menge der Koeffizienten ist abzählbar unendlich. Ein Syntheseoperator W auf die Menge der Koeffizienten $\{c_n\}$ kann wie folgt definiert werden:

$$W\{c_n\} = \sum_{n=1}^{\infty} \Phi_n(\mathbf{x},\omega) c_n(\omega). \tag{5.15}$$

C Reihendarstellung der Dichtefunktion

Die Dichtefunktionen φ and ψ lassen sich durch folgende Reihe darstellen (siehe z. B. Müller, 1966; Atkinson und Han, 2012; Dai und Xu, 2013):

$$\varphi(\hat{\mathbf{y}},\omega) = \sum_{n=1}^{\infty} \tilde{\Psi}_n(\hat{\mathbf{y}}) \varphi_n(\omega). \tag{5.16}$$

Die Koeffizienten φ_n der Reihenentwicklung werden dabei über das Skalarprodukt $\langle \Psi_n, \varphi \rangle$ bestimmt

$$\varphi_n(\omega) = \langle \Psi_n, \varphi \rangle := \int_{\Omega} \Psi_n(\hat{\mathbf{y}})^* \varphi(\hat{\mathbf{y}},\omega) dS(\hat{\mathbf{y}}), \tag{5.17}$$

wobei $(\cdot)^*$ wiederum die komplexe Konjugation bezeichnet. Obige Betrachtungen sind in gleicher Weise für ψ gültig. Die Familie der Funktionen $\{\Psi_n\}$ bildet einen Frame des Funktionsraums mit dem zugehörigen dualen Frame $\{\tilde{\Psi}_n\}$.[115] Einen Sonderfall stellen orthogonal normierte Basisfunktionen $\langle \Psi_n, \Psi_m \rangle = \delta_{n,m}$ dar, wobei $\delta_{n,m}$ das Kronecker-Delta bezeichnet. Es gilt: $\tilde{\Psi}_n = \Psi_n$. Ist die Dichtefunktion wiederum auf einem Kreis bzw. einer Kugel definiert und stetig,

[115] Zur weiterführenden Studie der Frame-Theorie sei auf die Arbeiten von Christensen (2003, 2008), Mallat (2009) und Vetterli et al. (2011) verwiesen.

ergibt sich die mit den Koeffizienten

$$\varphi_n(\omega) \quad = \quad \langle \Psi_n, \varphi(\cdot,\omega) \rangle \tag{5.18}$$

formal gebildete Fourierreihe

$$\varphi(\hat{\mathbf{y}},\omega) \quad = \quad \sum_{n=1}^{\infty} \Psi_n(\hat{\mathbf{y}}) \varphi_n(\omega), \tag{5.19}$$

mit

$$\begin{array}{ll} \Psi_n(\hat{\mathbf{y}}) & = Y_\nu^\mu(\theta_y,\phi_y), \qquad\qquad \text{3D} \\ \Psi_n(\hat{\mathbf{y}}) & = \exp(i\nu\phi_y)(2\pi)^{-1/2}, \quad \text{2D} \end{array} \tag{5.20}$$

wobei ν und μ über die Gleichungen (5.12) und (5.13) bestimmt sind. In manchen Anwendungen ist es von Vorteil, anstatt der Basisfunktionen nicht-orthogonale oder überbestimmte Frames (wie zum Beispiel Gabor-Frames, vgl. Balazs et al., 2011) oder Wavelets (vgl. Chambodut et al., 2005; Antoine und Roşca, 2008) zu verwenden. Dadurch lässt sich beispielsweise die Analyse bzw. Synthese sehr einfach auf begrenzte Zonen auf der Kugeloberfläche konzentrieren. Diese Ansätze werden in der vorliegenden Arbeit jedoch nicht weiter behandelt.

Fazi et al. (2012, Abschnitt 2.4) zeigen, wie die unterschiedlichen Darstellungen des Schallfeldes (Abschnitt A - C) mathematisch zusammenhängen. Die sphärisch harmonischen Basisfunktionen (Kugelflächenfunktionen) zur Berechnung modaler Beamformer werden in Kap. 6.1 diskutiert.

5.1.2 Raumimpulsantwort

Die Raumimpulsantwort (RIR)[116] beschreibt die akustische Übertragungsfunktion zwischen einer Schallquelle und einem Empfänger (Mikrofon). Die Impulsantwort $h(\mathbf{r}_m, \mathbf{r}_s, n)$ stellt das an einer Position $\mathbf{r}_m$ aufgenommene Schallfeld mit dem an einer Position $\mathbf{r}_s$ abgestrahlten Schallfeld in einen funktionalen Zusammenhang, wobei n die unter Berücksichtigung des Abtasttheorems mit der Abtastfrequenz

[116] *Room impulse response* (RIR).

f_A diskretisierte Zeitvariable darstellt (siehe z. B. Oppenheim et al., 1998, S. 146). Sind mehrere Schallquellen vorhanden, resultiert das Mikrofonsignal aus der additiven Überlagerung der Faltungsprodukte aller Quellensignale mit den jeweils korrespondierenden RIR. Dieser Ansatz besitzt aufgrund der Linearität der Wellengleichung Gültigkeit (vgl. Allen und Berkley, 1979; Ziomek, 1994; Kuttruff, 2000; Van Trees, 2002).

Neely und Allen (1979) zeigen, dass sich die Nullstellen der z-transformierten[117] RIR gleichförmig sowohl innerhalb als auch außerhalb des Einheitskreises verteilen (s. a. Huang et al., 2006, Kap. 2.2). Somit sind RIR nicht minimalphasig und lassen sich nicht direkt invertieren. Das Abklingverhalten einer RIR wird einerseits durch das Raumvolumen und andererseits durch die Absorptionseigenschaften der das Volumen begrenzenden Materialien bestimmt. In der Raumakustik wird das Abklingverhalten primär durch die Nachhallzeit T_{60} charakterisiert (vgl. Sabine, 1922; Cremer und Müller, 1978), die meist in Terz- oder Oktavbändern angegeben wird. Lange Abklingkonstanten zeigen sich in den geringen Abständen der Nullstellen zum Einheitskreis. Wie in Benesty und Huang (2003, Kap. 3) und Hänsler und Schmidt (2004) ausführlich gezeigt wird, muss in den meisten praktischen Anwendungen zudem von einer hohen zeitlichen Varianz der RIR ausgegangen werden. Die RIR verändern sich zum Beispiel durch Temperaturschwankungen oder durch sich im Raum bewegende Personen. Die mangelnde Minimalphasigkeit und die Zeitvarianz erschweren das Identifizieren und Entzerren von RIR. Dies ist zum Beispiel eine große Herausforderung für den Entwurf von Algorithmen zur Echokompensation in Sprachkommunikationssystemen (siehe z. B. Breining et al., 1999; Herbordt et al., 2003; Huang et al., 2006, Kap. 2.2).

5.1.3 Statistische Beschreibung des Schallfeldes

Wird ein Schallfeld durch Überlagerung einer hohen Anzahl unterschiedlicher Punktquellen gebildet, sind RIR zur Darstellung nur begrenzt anwendbar. Dies trifft zum Beispiel auch auf Schallquellen zu, deren räumliche Ausdehnung nicht vernachlässigbar ist. Hier muss die RIR über das gesamte schallabstrahlende

[117] Eine ausführliche Diskussion der z-Transformation findet sich zum Beispiel in Oppenheim et al. (1998, Kap. 3).

Volumen abgeleitet werden. In beiden Fällen ist eine statistische Beschreibung des Schallfeldes meist zielführender.

Im Allgemeinen besteht in Schallfeldern zwischen den an unterschiedlichen Raumpunkten aufgenommenen Signalen eine gewisse Korrelation. Der Grad der räumlichen Korrelation ist vom Abstand der Beobachtungspunkte zueinander, den akustischen Eigenschaften des Raumes, sowie dem Zeitversatz zwischen den Signalen und somit dem Signalspektrum abhängig. Folglich lassen sich Schallfelder durch räumlich-zeitliche Kreuzkorrelationsfunktionen beschreiben, wobei als bestimmende Größe meist das Betragsquadrat der komplexen räumlichen Kohärenzfunktion Anwendung findet (vgl. Carter et al., 1973; Knapp und Carter, 1976; Carter, 1987; Gardner, 1992). Die statistische Aussage über die Ähnlichkeit der Signale bei variierendem Betrachtungspunkt ersetzt somit die unmittelbare Beschreibung des Signalverlaufs (vgl. Vary et al., 1998).

Statistische Beschreibung im (kontinuierlichen) Zeitbereich. Die Verteilungsdichtefunktion $p_x\left(x(\mathbf{r},t)\right)$ ist der wesentliche Kern folgender Betrachtungen eines Zufallssignals $x(\mathbf{r},t)$ in einem Raumpunkt $\mathbf{r}$ zum Zeitpunkt t. Sie ist die Ableitung der dem jeweiligen Prozess zugrundeliegenden Verteilungsfunktion $P_x\left(x(\mathbf{r},t) \leq u\right)$. Die Verteilungsfunktion gibt die Wahrscheinlichkeit an, mit der die Zufallsvariable $x(\mathbf{r},t)$ kleiner oder gleich dem Wert u ist (s. a. Hänsler, 1983, 1997; Köhler, 2005, Kap. 2.1.5).

Das Zufallssignal $x(\mathbf{r},t)$ lässt sich anhand der statistischen Momente beschreiben (vgl. Papoulis, 1991, Kap. 7). In der Praxis kann jedoch nicht davon ausgegangen werden, dass die Verteilungsfunktion für alle Ordnungen bekannt und damit der stochastische Prozess vollständig beschrieben ist. Vielmehr beschränkt man sich zur Charakterisierung des Schallfeldes auf die statistischen Momente 1. und 2. Grades (vgl. Herbordt, 2005):

$$\eta_x(\mathbf{r},t) = \mathrm{E}\left\{x(\mathbf{r},t)\right\}, \tag{5.21}$$

$$\sigma_x^2(\mathbf{r},t) = \mathrm{E}\left\{x^2(\mathbf{r},t)\right\} - \eta_x^2(\mathbf{r},t), \tag{5.22}$$

$$R_x(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2) = x(\mathbf{r}_1,t_1)\,x(\mathbf{r}_2,t_2), \tag{5.23}$$

$$C_x(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2) = R_x(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2) - \eta_x(\mathbf{r}_1,t_1)\,\eta_x(\mathbf{r}_2,t_2). \tag{5.24}$$

Mit dem Mittelwert $\eta_x(\mathbf{r},t)$, der Varianz $\sigma_x^2(\mathbf{r},t)$, der räumlich-zeitlichen Korrelationsfunktion $R_x(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2)$, sowie der räumlich-zeitlichen Kovarianzfunktion $C_x(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2)$ des stochastischen Prozesses. Der Operator $\mathrm{E}\{\cdot\}$ beschreibt den Erwartungswert (vgl. Köhler, 2005, Kap. 2.1.6). Der Zusammenhang von Kovarianz und Korrelation ist leicht zu zeigen als

$$C_x(\mathbf{r},\mathbf{r},t,t) := \sigma_x^2(\mathbf{r},t) = R_x(\mathbf{r},\mathbf{r},t,t) - \eta_x^2(\mathbf{r},t). \tag{5.25}$$

Die statistischen Parameter hängen somit vom Quellsignal, den Eigenschaften des Ausbreitungspfades zum jeweiligen Beobachtungszeitpunkt und den Mikrofonpositionen ab.

Es lassen sich folgende charakteristische Schallfelder ableiten, die zur Beurteilung von Beamforming-Algorithmen wesentlich sind (vgl. Herbordt, 2005):

- Schwach stationäres Schallfeld: $\eta_x(\mathbf{r},t)$ und $\sigma_x^2(\mathbf{r},t)$ sind unabhängig von t. Somit ist die Korrelationsfunktion $R_x(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2)$ nur von der relativen Verschiebung $\Delta t = t_2 - t_1$ abhängig.

- Homogenes Schallfeld: $\eta_x(\mathbf{r},t)$ und $\sigma_x^2(\mathbf{r},t)$ sind unabhängig von $\mathbf{r}$. Die Korrelationsfunktion $R_x(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2)$ wird neben der zeitlichen Abhängigkeit von t_1,t_2 nur von der relativen Position $\Delta\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ bestimmt.

- Homogenes und isotropes Schallfeld: Die räumliche Abhängigkeit der Korrelationsfunktion $R_x(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2)$ wird ausschließlich durch die Euklidische Distanz $\|\Delta\mathbf{r}\|_2 = \|\mathbf{r}_1 - \mathbf{r}_2\|_2$ der Beobachtungspunkte bestimmt.

Die meisten technischen Prozesse werden als ergodische stochastische Prozesse modelliert, bei denen Zeit- und Scharmittelwerte gleich sind. Dadurch besteht die Möglichkeit durch zeitliche Mittelung nur einer Realisierung (Musterfunktion) Informationen über die Statistik zu gewinnen (vgl. Köhler, 2005, Kap. 2.1.10). Analog dazu kann im funktionalen Raum-Zeit-Modell durch räumlich-zeitliche Mittelung eine statistische Aussage über den Prozess getroffen werden. Aufgrund der endlichen Mittelungslänge stellen die (Raum-)Zeit-Mittelwerte jedoch nur eine Schätzung der Ensemblemittelwerte dar.

Statistische Beschreibung im Frequenzbereich. Die spektrale Darstellung $X(e^{j\Omega})$ der diskreten Zeitfolge $x(n)$ kann nach Gl. (4.13) über die zeitdiskrete Fouriertransformation (DTFT) erfolgen, wobei $\Omega = 2\pi f/f_A$ wiederum die auf die Abtastrate f_A normierte und im Bereich $-\pi < \Omega \leq \pi$ definierte Frequenzvariable bezeichnet. In der Praxis kommt meist die Kurzzeit-Fourier-Transformation (STFT) zur Anwendung, bei der das kontinuierliche Spektrum $X(e^{j\Omega})$ durch Diskretisierung approximiert wird, vgl. Gl. (4.14).

Das Autoleistungsdichtespektrum $S_{x_1x_1}(e^{j\Omega})$ und das Kreuzleistungsdichtespektrum $S_{x_1x_2}(e^{j\Omega})$ für zwei zeitdiskrete, verbundstationäre Prozesse $x_1(n) := x(\mathbf{r}_1,n)$ und $x_2(n) := x(\mathbf{r}_2,n)$ lassen sich darstellen als

$$S_{x_1x_1}(e^{j\Omega}) = \mathrm{E}\left\{\left|X_1(e^{j\Omega})\right|^2\right\}, \tag{5.26}$$

$$S_{x_1x_2}(e^{j\Omega}) = \mathrm{E}\left\{X_1(e^{j\Omega})\, X_2^*(e^{j\Omega})\right\}. \tag{5.27}$$

Unter der Annahme eines schwach stationären Schallfeldes kann das räumlich-zeitliche Kreuzleistungsdichtespektrum $S_{x_1x_2}(e^{j\Omega})$ aus der DTFT der zeitdiskreten Korrelationsfunktion $R_x(\mathbf{r}_1, \mathbf{r}_2, k)$ ermittelt werden (s. a. Kap. 4.1.1)

$$S_{x_1x_2}(e^{j\Omega}) = DTFT\{R_x(\mathbf{r}_1, \mathbf{r}_2, k)\} = \sum_{k=-\infty}^{\infty} R_x(\mathbf{r}_1, \mathbf{r}_2, k)\, e^{-j\Omega k} \tag{5.28}$$

Daraus lässt sich nach Bendat und Piersol (1966) die für zwei verbundstationäre stochastische Prozesse definierte komplexe Kohärenzfunktion

$$\gamma_{x_1x_2}(\Omega) = \frac{S_{x_1x_2}(e^{j\Omega})}{\sqrt{S_{x_1x_1}(e^{j\Omega})\, S_{x_2\,x_2}(e^{j\Omega})}} \tag{5.29}$$

ableiten. Durch die Normierung wird der Wertebereich der Kohärenzfunktion auf $0 \leq |\gamma_{x_1x_2}(\Omega)| \leq 1$ begrenzt (vgl. Knapp und Carter, 1976).

Meist erfolgt die Beschreibung der räumlich-zeitlichen Korrelation der Signale $x_1(n)$ und $x_2(n)$ durch das Betragsquadrat der Kohärenzfunktion (MSC)[118], das

[118] *Magnitude-squared coherence function* (MSC).

nach Carter (1987) wie folgt definiert ist:

$$C_{x_1x_2}(\Omega) = \frac{|S_{x_1x_2}(e^{j\Omega})|^2}{S_{x_1x_1}(e^{j\Omega})\, S_{x_2\,x_2}(e^{j\Omega})}. \tag{5.30}$$

Allgemein gilt: Ist $C_{x_1x_2}(\Omega) = 1$, sind die stochastischen Prozesse $x_1(n)$ und $x_2(n)$ linear voneinander abhängig; ist $C_{x_1x_2}(\Omega) = 0$, sind diese unkorreliert. Bei Werten $0 < C_{x_1x_2}(\Omega) < 1$ sind die Prozesse entweder nicht linear voneinander abhängig oder den Prozessen sind zusätzliche unkorrelierte Signalanteile überlagert (vgl. Bendat und Piersol, 1966).

Korrelations- und Kohärenzmatrix. Die über ein bestimmtes Zeitintervall gebildeten Korrelationsfunktionen der Signale aller möglichen Mikrofonpaarungen $\{x_i(n), x_j(n)\}$, mit $i, j = 1, \ldots, M$, lassen sich in einer Korrelationsmatrix zusammenfassen. Dies führt auf eine sehr kompakte Schreibweise des von einem Mikrofonarray räumlich und zeitlich abgetasteten Schallfeldes.

Werden die letzten N Abtastwerte des am Ausgang des m-ten Sensors anliegenden Signals $x_m(n)$ in dem Vektor

$$\mathbf{x}_m(n) = \Big[\; x_m(n),\;\; x_m(n-1),\;\; \ldots,\;\; x_m(n-N+1) \;\Big]^T \tag{5.31}$$

zusammengefasst und die $\mathbf{x}_m(n)$ für alle $m = 1, \ldots, M$ in einem $MN \times 1$ Stapelvektor

$$\mathbf{x}(n) = \Big[\; \mathbf{x}_1^T(n),\;\; \mathbf{x}_2^T(n),\;\; \ldots,\;\; \mathbf{x}_M^T(n) \;\Big]^T \tag{5.32}$$

gruppiert, lässt sich die räumlich-zeitliche Korrelationsmatrix $\mathbf{R}_{\mathbf{xx}}(n)$ wie folgt ausdrücken

$$\mathbf{R}_{\mathbf{xx}}(n) = \mathrm{E}\left\{\mathbf{x}(n)\, \mathbf{x}^T(n)\right\}, \tag{5.33}$$

wobei $[\cdot]^T$ die Transposition bezeichnet. $\mathbf{R}_{\mathbf{xx}}(n)$ ist eine $MN \times MN$ Blockmatrix, deren Hauptdiagonale aus den Autokorralationsmatriz $\mathbf{r}_{x_mx_m}(n)$ der Signale an den Mikrofonpositionen $\mathbf{r}_m$ besteht. Unter der Annahme reellwertiger Signale gilt $\mathbf{r}_{x_1x_2}(n) = \mathbf{r}^T_{x_2x_1}(n)$. In diesem Fall ist $\mathbf{R}_{\mathbf{xx}}(n)$ blocksymmetrisch.

Die $M \times M$ Kohärenzmatrix $\mathbf{\Gamma_{xx}}(\Omega)$ wird durch die Kohärenzfunktionen $\gamma_{x_1\,x_2}(\Omega)$ der Signale an den Mikrofonpositionen $\{\mathbf{r}_{m_i}, \mathbf{r}_{m_j}\}$, mit $i, j = 1, \ldots, M$, bestimmt:

$$\mathbf{\Gamma_{xx}}(\Omega) = \begin{pmatrix} 1 & \gamma_{x_0x_1}(\Omega) & \ldots & \gamma_{x_0x_{M-1}}(\Omega) \\ \gamma_{x_1x_0}(\Omega) & 1 & \ldots & \gamma_{x_1x_{M-1}}(\Omega) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{x_{M-1}x_0}(\Omega) & \gamma_{x_{M-1}x_1}(\Omega) & \ldots & 1 \end{pmatrix}. \tag{5.34}$$

Aus Gl. (5.34) ist leicht ersichtlich, dass sich für räumlich dekorrelierte Signale (wie z. B. dem Eigenrauschen der Sensoren) die Einheitsmatrix $\mathbf{I}$ ergibt:

$$\mathbf{\Gamma_{xx}}(\Omega) = \mathbf{I}. \tag{5.35}$$

Ist das Schallfeld homogen (d. h. $\forall\, m = 1, \ldots, M$ gilt $S_{x_mx_m} = S_{xx}$) kann für die Leistungsdichtespektren folgende einfache Beziehung hergeleitet werden:

$$\mathbf{S_{xx}}(e^{j\Omega}) = \mathrm{E}\left\{\mathbf{X}(e^{j\Omega})\,\mathbf{X}^H(e^{j\Omega})\right\} = S_{xx}(e^{j\Omega})\,\mathbf{\Gamma_{xx}}(\Omega). \tag{5.36}$$

Hierbei bezeichnet $\mathbf{X}(e^{j\Omega}) = [\ X_0(e^{j\Omega}) \quad X_1(e^{j\Omega}) \quad \ldots \quad X_{M-1}(e^{j\Omega})\ ]^T$ den Vektor der Fourierkoeffizienten der M Mikrofonsignale und $[\cdot]^H$ die komplex konjugierte (auch Hermitesche) Transposition.

$\mathbf{\Gamma_{xx}}(\Omega)$ und $\mathbf{S_{xx}}(e^{j\Omega})$ werden zum Beispiel zur Bestimmung des Richtindex $DI(e^{j\Omega})$ (s. Gl. 5.60) und der Arrayverstärkung G (s. Gl. 5.66) herangezogen (vgl. Abschnitt 5.1.6).

5.1.4 Schätzung der statistischen Kennwerte

Mikrofonarray-Beamformer werden oft unter der Annahme bestimmter statistischer Eigenschaften des Nutz- und Störschallfeldes optimiert. Entspricht das tatsächliche Schallfeld jedoch nicht den getroffenen Annahmen, sind die Algorithmen nicht mehr optimal. Dem kann dadurch entgegnet werden, indem die statistischen Eigenschaften des Schallfeldes kontinuierlich aus den Mikrofonsignalen geschätzt und die Filterkoeffizienten adaptiv angepasst werden.

Unter der Voraussetzung schwach stationärer ergodischer Prozesse, lassen sich, wie im vorhergehenden Kapitel beschrieben, die Erwartungswerte durch Zeitmittelwerte ersetzen. Bei der Kurzzeitmessung werden die Signale $x_1(n)$ und $x_2(n)$ in sich überlappende Blöcke der Länge K unterteilt und mit einer geeigneten Fensterfunktion gewichtet. Das Kreuzleistunsdichtespektrum kann, wie bereits in Kap. 4.1.2 diskutiert wurde, über das Periodogramm geschätzt werden. Die Kurzzeit-Kreuzkorrelation lässt sich wie folgt berechnen:

$$\hat{R}_{x_1x_2}(\nu) = \frac{1}{2K+1} \sum_{k=-K}^{K} x_1(k+\nu)\, x_2(k). \tag{5.37}$$

Einsetzen der über das Periodogramm, Gl. (4.22), geschätzten Leistungsdichtespektren in Gl. (5.30), führt auf die Kurzzeit-MSC-Funktion, $C_{x_1x_2}(\Omega, \ell)$, wobei ℓ den Blockindex bezeichnet (vgl. Gl. 4.14). Nach Bendat und Piersol (1993, S. 167ff) wird durch die rekursive Mittelung der Periodogramme gewährleistet, dass $0 \leq C_{x_1x_2}(\Omega, \ell) \leq 1$.

Jacobson (1962) zeigt für die Kohärenz eines sphärisch isotropen Schallfeldes folgenden Zusammenhang:[119]

$$\gamma_{x_1\,x_2}(\omega) = \frac{\sin\left(\omega \, \left\|\mathbf{r}_2 - \mathbf{r}_1\right\|_2 / c\right)}{\omega \, \left\|\mathbf{r}_2 - \mathbf{r}_1\right\|_2 / c}. \tag{5.38}$$

Typische Beispiele sind das diffuse Schallfeld (wie zum Beispiel der späte Nachhall) sowie das Störgeräuschfeld in Fahrzeuginnenräumen. Abb. 5.1 vergleicht die Kohärenzfunktion eines simulierten diffusen Schallfeldes mit der eines im Fahrzeuginnenraum gemessenen Störgeräuschfeldes.

5.1.5 Signalmodell

Ausgangspunkt der Betrachtungen ist eine sich an der Position $\mathbf{r}_d$ befindliche Punktschallquelle, mit dem Signal $s_d(t)$, und ein aus M Mikrofonen bestehendes Sensorarray. Das von der Schallquelle emittierte Schallfeld erreicht entweder auf direktem Weg (Direktschall) oder zeitlich verzögert über Raumreflexionen die

[119] Siehe zum Beispiel auch Cox (1973b), Cox et al. (1986), Johnson (1993), Bitzer et al. (1999a), Brandstein und Ward (2001, Kap. 2) und Van Trees (2002).

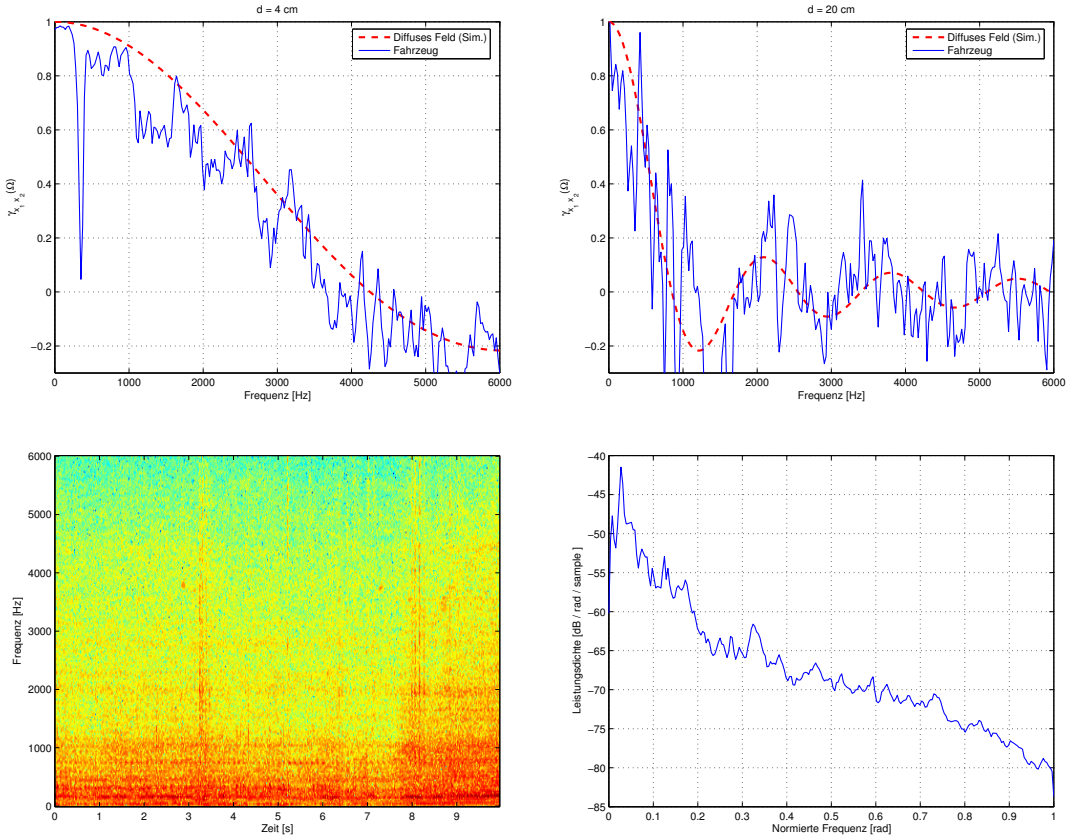


Abbildung 5.1: Kohärenzfunktion eines simulierten diffusen Schallfeldes (gestrichelte rote Linie) und eines im Fahrzeuginnenraum gemessenen Störgeräuschfeldes (durchgezogene blaue Linie) bei einem Mikrofonabstand von (a) $d = 4\,\mathrm{cm}$ und (b) $d = 20\,\mathrm{cm}$, sowie (c) das Spektrogramm und (d) das Leistungsdichtespektrum des Störsignals.

Mikrofone des Arrays an den Positionen $\mathbf{r}_m$ und wird von Störungen und Interferenzen überlagert. Abb. 5.2 zeigt eine schematische Darstellung des Signalmodells.

Der Direktanteil $d_m(t)$ des Nutzsignals am m-ten Mikrofon entspricht dem um die akustische Flugzeit $\tau_m = \|\mathbf{r}_d - \mathbf{r}_m\|_2 / c$ zwischen Schallquelle und Mikrofon verzögerten ungestörten Nutzsignal

$$d_m(t) = s_d(t - \tau_m). \tag{5.39}$$

Die Mikrofone tasten das Schallfeld an den Positionen $\mathbf{r}_m$, mit $m = 1, \ldots, M$, ab. Die zeitdiskreten Mikrofonsignale, $x_m(n)$, ergeben sich aus einer Überlagerung

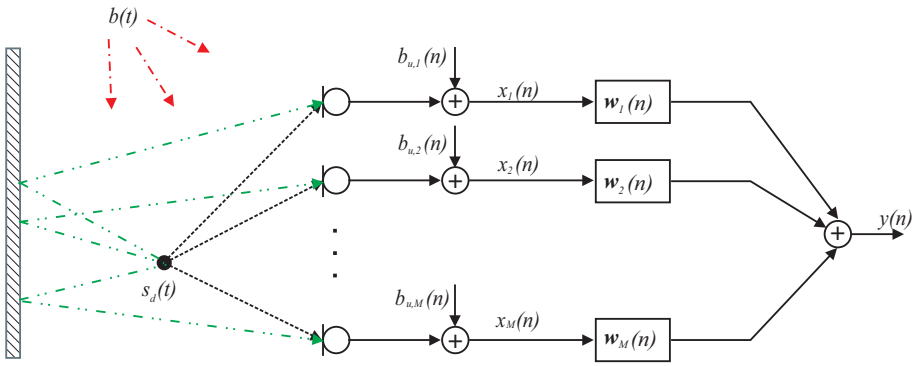


Abbildung 5.2: Signalmodell eines Mikrofonarrays in reflexionsbehafteter Umgebung und einem sich additiv überlagernden Störfeld.

von Nutzsignal, $d_m(n)$, und Störsignalkomponenten, $b_m(n)$, wobei n wiederum die unter Berücksichtigung des Abtasttheorems diskretisierte Zeitvariable darstellt:

$$x_m(n) = d_m(n) + b_m(n). \tag{5.40}$$

Die Signale werden typischerweise als gaußverteilte, mittelwertfreie Zufallsvariablen angenommen. $b_m(n)$ umfasst alle Störungen, wie zum Beispiel Interferenzen und Sensorrauschen.

Interferenzen sind teils klar lokalisierbar (wie z. B. Lüftergeräusche elektronischer Geräte, konkurrierende bzw. störende Sprecher), teils bilden diese ein diffuses Schallfeld aus (wie z. B. der Luftzug einer Klimaanlage). Störgeräusche weisen zudem oft eine starke Klangfärbung auf (wie z. B. Rollgeräusche im Fahrzeug, Sprache oder Musik), wobei hier zwischen stationären, schwach stationären und zeitvarianten spektralen Eigenschaften zu unterscheiden ist. Das Eigenrauschen von Sensoren (wie z. B. Mikrofonen) kann als stationäres, räumlich dekorreliertes, weißes Rauschen modelliert werden. Rückgekoppelte Systeme (wie z. B. Lautsprecher und Mikrofone in Freisprechanlagen) erzeugen zusätzliche Echos, die in Sprachkommunikationssystemen mitunter zu einer erheblichen Beeinträchtigung der Sprachverständlichkeit führen können. Zur Vereinfachung der Notation werden in den folgenden Betrachtungen Raumreflexionen und Echos ebenfalls der Störsignalkomponente $b_m(n)$ zugeordnet.

Das Signal $y(n)$ am Ausgang des Mikrofonarrays wird aus der Summe der gefilterten Mikrofonsignale $x_m(n)$ gebildet (s. Abb. 5.2). Im Schmalbandfall reduzieren sich die Filter in den Mikrofonkanälen zu komplexen Gewichten. Mit den $N \times 1$ Koeffizientenvektoren der M Transversalfilter

$$\mathbf{w}_m(n) = \left[\; w_{0,m}(n), \quad w_{1,m}(n), \quad \ldots, \quad w_{N-1,m}(n) \; \right]^T \tag{5.41}$$

und den $N \times 1$ Vektoren der zugehörigen Mikrofonsignale

$$\mathbf{x}_m(n) = \left[\; x_m(n), \quad x_m(n-1), \quad \ldots, \quad x_m(n-N+1) \; \right]^T \tag{5.42}$$

lässt sich die Faltungssumme als einfaches Skalarprodukt schreiben, und für das Ausgangssignal gilt:

$$y(n) = \sum_{m=1}^{M} w_m(n)^* * x_m(n) = \sum_{m=1}^{M} \mathbf{w}_m(n)^H \mathbf{x}_m(n). \tag{5.43}$$

Dabei bezeichnet $*$ die lineare Faltung. Die (zeitvarianten) Filtergewichtsvektoren $\mathbf{w}_m(n)$ und Signalvektoren $\mathbf{x}_m(n)$ lassen sich zu $MN \times 1$ Stapelvektoren zusammenfassen:

$$\mathbf{w}(n) = \left[\; \mathbf{w}_1^T(n), \quad \mathbf{w}_2^T(n), \quad \ldots, \quad \mathbf{w}_M^T(n) \; \right]^T, \tag{5.44}$$

$$\mathbf{x}(n) = \left[\; \mathbf{x}_1^T(n), \quad \mathbf{x}_2^T(n), \quad \ldots, \quad \mathbf{x}_M^T(n) \; \right]^T. \tag{5.45}$$

Mit den $N \times 1$ Signalvektoren des Direktanteils $\mathbf{d}_m(n)$ und der Interferenzen $\mathbf{b}_m(n)$

$$\mathbf{d}_m(n) = \left[\; d_m(n), \quad d_m(n-1), \quad \ldots, \quad d_m(n-N+1) \; \right]^T, \tag{5.46}$$

$$\mathbf{b}_m(n) = \left[\; b_m(n), \quad b_m(n-1), \quad \ldots, \quad b_m(n-N+1) \; \right]^T, \tag{5.47}$$

die sich wiederum in $MN \times 1$ Stapelvektoren $\mathbf{d}(n)$ und $\mathbf{n}(n)$ zusammenfassen lassen

$$\mathbf{d}(n) = \left[\; \mathbf{d}_1^T(n), \quad \mathbf{d}_2^T(n), \quad \ldots, \quad \mathbf{d}_M^T(n) \; \right]^T, \tag{5.48}$$

$$\mathbf{b}(n) = \left[\; \mathbf{b}_1^T(n), \quad \mathbf{b}_2^T(n), \quad \ldots, \mathbf{b}_M^T(n) \; \right]^T, \tag{5.49}$$

ergibt sich eine sehr kompakte Schreibweise für das Beamformer-Ausgangssignal

$$y(n) = \mathbf{w}(n)^H \mathbf{x}(n) = \mathbf{w}(n)^H \left(\mathbf{d}(n) + \mathbf{b}(n)\right). \tag{5.50}$$

Gl. (5.50) lässt sich im Vektorraum der Filtergewichte $span\{\mathbf{w}(n)\}$ veranschaulichen (vgl. Cox, 1973a). Stehen $\mathbf{w}(n)$ und $\mathbf{x}(n)$ orthogonal zueinander ist $y(n) = 0$, d. h. das Eingangssignal wird vollständig unterdrückt. Ist $\mathbf{x}(n) \in span\{\mathbf{w}(n)\}$, wird das Signal übertragen. Um Interferenzen und Störungen zu unterdrücken, sollte $\mathbf{d}(n) \in span\{\mathbf{w}(n)\}$ und $\mathbf{b}(n)$ orthogonal zu $\mathbf{w}(n)$ sein.

5.1.6 Richtcharakteristik und Bewertungsmaße

Das Raum-Zeit-Signal $s(\mathbf{r},t)$ einer sich im Fernfeld eines Arrays befindenden Schallquelle lässt sich an den Mikrofonpositionen als ebene Schallwelle darstellen (s. a. Abschnitt 5.1.1). Betrachtet man nur den Direktanteil $d_m(n)$ des Mikrofonsignals $x_m(n)$ ergibt sich mit Gl. (5.2)

$$d_m(n) = \hat{D} \exp\left\{j(\Omega\, n - \mathbf{k}^T \mathbf{r}_m)\right\}, \tag{5.51}$$

mit dem Wellenzahlvektor $\mathbf{k}$, dem Abtastintervall T_A, der Signalamplitude $\hat{D}$ und der komplexen Konstante $j = \sqrt{-1}$. Mit der DTFT (siehe Gl. 4.13) lassen sich sowohl das Direktsignal $d_m(n)$

$$D_m(e^{j\Omega}) = DTFT\{d_m(n)\} = \sum_{n=-\infty}^{\infty} d_m(n)\, e^{-j\Omega n} = \hat{D} \exp\left\{-j\mathbf{k}_0^T \mathbf{r}_m\right\} \tag{5.52}$$

als auch die Filterkoeffizenten $\mathbf{w}_m(n)$

$$W_m(e^{j\Omega}) = DTFT\{w_m\} = \sum_{\ell=0}^{N-1} w_m(\ell)\, e^{-j\Omega\ell} \tag{5.53}$$

in den Frequenzbereich transformieren (vgl. Benesty und Huang, 2003, Kap. 6.2.2). Zur Vereinfachung wird im Folgenden – unter Annahme stationärer Bedingungen – der Zeitindex n bei den Filterkoeffizienten weggelassen. Nun werden die Exponentialterme in Gl. (5.52) für alle Mikrofone $\mathbf{r}_m$ in einem $M \times 1$ Vektor

zusammenfasst

$$\mathbf{v}(\Omega,\mathbf{k}_0) = \Big[\ \exp\{j\mathbf{k}_0^T\mathbf{r}_1\},\ \ \exp\{j\mathbf{k}_0^T\mathbf{r}_2\},\ \ \ldots,\ \ \exp\{j\mathbf{k}_0^T\mathbf{r}_M\}\ \Big]^T. \tag{5.54}$$

Der sogenannte *Steering*-Vektor $\mathbf{v}(\Omega,\mathbf{k}_0)$ beschreibt die zeitliche Entzerrung (bzw. Phasenverschiebung), die notwendig ist, um die Hauptkeule des Beamformers bei gegebener Arraygeometrie auf eine aus Richtung $\mathbf{u}(\theta_0,\phi_0)$ einfallende monochromatische ebene Welle mit Wellenzahl k_0 auszurichten. Durch Zusammenfassen der spektralen Filtergewichte in einem $M \times 1$ Vektor

$$\mathbf{w}_F(e^{j\Omega}) = \Big[\ W_1(e^{j\Omega}),\ \ W_2(e^{j\Omega}),\ \ \ldots,\ \ W_M(e^{j\Omega})\ \Big]^T \tag{5.55}$$

und Normierung der Amplitude der zeitlichen Schwingung zu $\hat{D} = 1$, ergibt sich für das Ausgangssignal (vgl. Van Trees, 2002, Kap. 2.6)

$$\mathbf{Y}(\Omega,\mathbf{k}_0) = \mathbf{w}_F^H(e^{j\Omega})\mathbf{v}(\Omega,\mathbf{k}_0). \tag{5.56}$$

$\mathbf{Y}(\Omega,\mathbf{k}_0)$ ist die Systemantwort (*frequency wavenumber response*) des Mikrofonarrays auf eine monochromatische ebene Welle mit Wellenzahl $\mathbf{k}_0$.

Richtcharakteristik. Die Richtcharakteristik $B(\Omega;\theta,\phi)$ eines Mikrofonarrays ergibt sich aus der über alle möglichen Frequenzen und alle möglichen Einfallsrichtungen ($0 \leqslant \theta \leqslant \pi, 0 \leqslant \phi < 2\pi$) berechneten Systemantwort (vgl. Bitzer und Simmer 2001, Kap. 2.2.2; Van Trees 2002, Kap. 2.6):

$$B(\Omega;\theta,\phi) = \mathbf{Y}(\Omega,\mathbf{k}) = \mathbf{w}_F^H(e^{j\Omega})\mathbf{v}(\Omega,\mathbf{k})\big|_{\mathbf{k}:\,\theta\in[0;\pi],\,\phi\in[0;2\pi]}. \tag{5.57}$$

Daraus ergibt sich die auf die Leistungsübertragung bezogene Richtcharakteristik

$$P(\Omega;\theta,\phi) = |B(\Omega;\theta,\phi)|^2. \tag{5.58}$$

Richtwirkung. Der Richtfaktor ist ein Maß dafür, wie stark ein diffuses Schallfeld gegenüber einem Signal in Hauptkeulenrichtung unterdrückt wird. Dieser wird typischerweise als logarithmische Größe, dem sogenannten Richtindex oder

auch Bündelungsmaß (*directivity index*, DI), angeben. Für ein Mikrofonarray mit Vorzugsrichtung (θ_0, ϕ_0) ist der DI wie folgt definiert (vgl. Bitzer und Simmer 2001, Kap. 2.2.3; Van Trees 2002, Kap. 2.6):

$$DI(e^{j\Omega}) = 10\log_{10}\left(\frac{|B(\Omega;\theta_0,\phi_0)|^2}{\frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi}|B(\Omega;\theta,\phi)|^2\sin\theta\,d\phi\,d\theta}\right). \tag{5.59}$$

Durch Einsetzen von Gl. (5.57) in Gl. (5.59) und unter Berücksichtigung der in Gl. (5.34) definierten Kohärenzmatrix $\mathbf{\Gamma}_{xx}(\Omega)$ ergibt sich folgende alternative Schreibweise

$$\begin{aligned} DI(e^{j\Omega}) &= 10\log_{10}\left(\frac{\left|\mathbf{w}_F^H(e^{j\Omega})\mathbf{v}(\Omega,\mathbf{k}_0)\right|^2}{\mathbf{w}_F^H(e^{j\Omega})\mathbf{\Gamma}_{xx}(\Omega)\mathbf{w}_F(e^{j\Omega})}\right) = \\ &= 10\log_{10}\left(\frac{\mathbf{w}_F^H(e^{j\Omega})\mathbf{v}(\Omega,\mathbf{k}_0)\mathbf{v}^H(\Omega,\mathbf{k}_0)\mathbf{w}_F^H(e^{j\Omega})}{\mathbf{w}_F^H(e^{j\Omega})\mathbf{\Gamma}_{xx}(\Omega)\mathbf{w}_F(e^{j\Omega})}\right), \end{aligned} \tag{5.60}$$

wobei zur Berechnung des DI für $\mathbf{\Gamma}_{xx}(\Omega)$ die Kohärenzmatrix eines diffusen Schallfeldes einzusetzen ist.

Arrayverstärkung. Die Arrayverstärkung (*array gain*, AG) ist ein Maß für die Verbesserung des SNR am Ausgang des Mikrofonarrays gegenüber dem SNR an einem Mikrofon (vgl. Bitzer und Simmer 2001, Kap. 2.2.1; Van Trees 2002, Kap. 2.6):

$$G = \frac{SNR_{\text{out}}}{SNR_{\text{in}}}. \tag{5.61}$$

Sind die Signale stationär und Nutz- und Störkomponenten unkorreliert, kann das SNR über das Verhältnis der Leistungsdichtespektren des Nutzsignals, $\Phi_{DD}(e^{j\Omega})$, und des Störsignals, $\Phi_{BB}(e^{j\Omega})$, bestimmt werden (s. a. Kap. 4.1.2):

$$SNR_{\text{in}} = \frac{\Phi_{DD}(e^{j\Omega})}{\Phi_{BB}(e^{j\Omega})}. \tag{5.62}$$

Das Leistungsdichtespektrum am Ausgang des Mikrofonarrays berechnet sich zu

$$\Phi_{YY}(e^{j\Omega}) = \mathbf{w}_F^H(e^{j\Omega})\,\mathbf{S_{xx}}(e^{j\Omega})\,\mathbf{w}_F(e^{j\Omega}). \tag{5.63}$$

$\mathbf{S_{xx}}(e^{j\Omega}) = \mathrm{E}\left\{\mathbf{x}(e^{j\Omega})\,\mathbf{x}^H(e^{j\Omega})\right\}$ ist die Matrix der Leistungsdichtespektren und $\mathbf{x}(e^{j\Omega}) = [\ X_0(e^{j\Omega}),\ \ X_1(e^{j\Omega}),\ \ \ldots,\ \ X_{M-1}(e^{j\Omega})\ ]^T$ der Vektor der Fourierkoeffizienten der M Mikrofonsignale.

Betrachten wir den Fall, in dem ausschließlich Nutzsignal vorhanden ist, ergibt sich aus Gl. (5.63)

$$\Phi_{YY,D}(e^{j\Omega}) = \Phi_{DD}(e^{j\Omega})\left|\mathbf{w}_F^H(e^{j\Omega})\mathbf{v}(\Omega,\mathbf{k}_0)\right|^2. \tag{5.64}$$

Ist hingegen ausschließlich Störsignal vorhanden, ergibt sich mit Gl. (5.36) folgender Ausduck

$$\Phi_{YY,B}(e^{j\Omega}) = \Phi_{BB}(e^{j\Omega})\,\mathbf{w}_F^H(e^{j\Omega})\mathbf{\Gamma_{bb}}(\Omega)\mathbf{w}_F(e^{j\Omega}). \tag{5.65}$$

Mit den Gln. (5.64) und (5.65) kann das SNR am Ausgang des Mikrofonarrays berechnet werden:

$$SNR_{\text{out}} = \frac{\Phi_{DD}(e^{j\Omega})}{\Phi_{BB}(e^{j\Omega})} \cdot \frac{\left|\mathbf{w}_F^H(e^{j\Omega})\mathbf{v}(\Omega,\mathbf{k}_0)\right|^2}{\mathbf{w}_F^H(e^{j\Omega})\mathbf{\Gamma_{bb}}(\Omega)\mathbf{w}_F(e^{j\Omega})}. \tag{5.66}$$

Einsetzen der Gln. (5.62) und (5.66) in Gl. (5.61) führt auf folgenden Ausdruck für die Arrayverstärkung:

$$G = \frac{\left|\mathbf{w}_F^H(e^{j\Omega})\mathbf{v}(\Omega,\mathbf{k}_0)\right|^2}{\mathbf{w}_F^H(e^{j\Omega})\mathbf{\Gamma_{bb}}(\Omega)\mathbf{w}_F(e^{j\Omega})}, \tag{5.67}$$

mit dem sich die Arrayverstärkung sehr einfach für unterschiedliche Störfelder berechnen lässt, da diese meist durch ihre Kohärenzmatrix (vgl. Gl. 5.34) beschrieben werden.

Gewinn für inkohärentes Rauschen. Der Gewinn für inkohärentes Rauschen (*white noise gain*, WNG) ist ein etabliertes Maß, welches angibt, inwieweit ein Mikrofonarray räumlich dekorreliertes Rauschen (wie z. B. das Eigenrauschen von Sensoren) unterdrückt (vgl. Bitzer und Simmer, 2001, Kap. 2.2.5). Es gilt auch als Maß für die Robustheit eines Beamformers.

Wird die Kohärenzmatrix für ein räumlich dekorreliertes Schallfeld $\mathbf{\Gamma_{bb}}(\Omega) = \mathbf{I}$ in Gl. (5.67) eingesetzt, ergibt sich der Gewinn zu

$$\mathrm{WNG}(\Omega) = \frac{\left|\mathbf{w}_F^H(e^{j\Omega})\mathbf{v}(\Omega,\mathbf{k}_0)\right|^2}{\mathbf{w}_F^H(e^{j\Omega})\mathbf{w}_F(e^{j\Omega})}. \tag{5.68}$$

Ist $10\log_{10}\mathrm{WNG}(\Omega)$ positiv, wird das unkorrelierte Rauschen unterdrückt; ist der Wert hingegen negativ, wird dieses verstärkt.

5.2 Entwurf eines robusten Beamformers

Adaptive Beamformer schätzen die statistischen Eigenschaften des Schallfeldes aus den Mikrofonsignalen und adaptieren die Filterkoeffizienten solange, bis sich eine optimale Lösung einstellt. Dadurch wird vor allem bei Arrays kleiner Bauform und geringer Anzahl an Mikrofonen eine wesentlich höhere und breitbandigere richtungsabhängige Verstärkung erreicht, als dies mit signalunabhängigen Ansätzen möglich ist.[120]

In der Praxis ist der GSC (vgl. Griffiths, 1977)[121] am weitesten verbreitet. Dieser löst das LCMV-Problem in zueinander orthogonalen Unterräumen, was auf eine sehr einfache Implementierung ohne Zwangsbedingungen führt. Dazu wird einem Preprozessor, bestehend aus einem fixen Beamformer (FBF) und einer Blockiermatrix (BM), ein aktiver mehrkanaliger Störgeräuschunterdrücker (*multiple input canceler*, MIC) nachgeschaltet. Kommt es aufgrund von Fertigungstoleranzen (wie z. B. Fehlpositionierungen der Mikrofone) und Bauteiltoleranzen (wie z. B. Pegel- und Phasenunterschiede der Mikrofone) zu einer Fehlausrichtung der Hauptkeule (*steering vector errors*), wird das Nutzsignal in der BM nicht vollständig unterdrückt. Dies kann bei der adaptiven Störgeräuschunterdrückung im MIC zu starken Verzerrungen des Nutzsignals führen (vgl. Jablon, 1986a, 1987). Dem Problem kann entweder durch Verbessern der räumlichen Filter der BM (vgl. Claesson und Nordholm, 1992; Er und Cantoni, 1986; Fudge und Linebarger, 1995) oder durch Verringern der Nutzsignalunterdrückung im MIC (vgl Cox et al., 1987; Jablon,

[120] Ein Vergleich signalabhängiger und signalunabhängiger Beamformer ist in zum Beispiel in Pape (2005) zu finden.

[121] Siehe Fußnote 102, S. 121.

1986a; Claesson und Nordholm, 1992) entgegnet werden. Eine sehr robuste Implementierung des GSC wurde von Hoshuyama und Sugiyama (1996)[122] vorgestellt. Das Blockdiagramm dieses robusten adaptiven Beamformers (RAB) ist in Abb. 5.3 dargestellt. Die BM besteht aus adaptiven Filtern mit beschränktem Koeffizientenwachstum (*coefficient constrained adaptive filters*, CCAF), die das Nutzsignal weitestgehend unterdrücken. Konvergieren die CCAF nicht vollständig, verhindern die normbeschränkten adaptiven Filter (*norm constrained adaptive filters*, NCAF) des MIC, dass es aufgrund des residualen Nutzsignals in den Referenzkanälen zu einer wahrnehmbaren Verzerrung des Nutzsignals kommt. Die Adaptionskontrolle (AMC) adaptiert die CCAF der BM nur dann, wenn Nutzsignal vorhanden ist. Ist kein Nutzsignal vorhanden, werden die NCAF des MIC adaptiert (vgl. Van Compernolle, 1990). Die AMC hat somit einen wesentlichen Einfluss auf die Robustheit des RAB. Hoshuyama und Sugiyama bestimmen die Nutzsignalpausen über eine Schätzung des Signal-Interferenz-Verhältnis (*signal-to-interference ratio*, SIR), welches die Leistungen am Ausgang des FBF und der BM ins Verhältnis setzt. In der Praxis kommt es dabei immer wieder zu Fehldetektionen und einer damit verbundenen Fehlanpassung der adaptiven Prozesse. Dies führt zu stark wahrnehmbaren Verzerrungen des entstörten Nutzsignals am Ausgang des RAB. Der in dieser Arbeit vorgestellte modifizierte RAB (vgl. Pape, 2005) minimiert die richtungsabhängigen Schwankungen der SIR-Schätzung und führt ein zusätzliches räumliches Kriterium zur AMC ein, welches die Signalleistungen aus unterschiedlichen Raumrichtungen ins Verhältnis setzt (*steered response power ratio*, SRPR).

Im folgenden Abschnitt wird die Theorie des RAB nach Hoshuyama und Sugiyama kurz zusammengefasst. Daraus wird in Abschnitt 5.2 ein modifizierter RAB abgeleitet, bei dem die Robustheit durch eine verbesserte AMC erhöht wird.

5.2.1 Signalmodell

Abb. 5.3 zeigt das Blockdiagramm des RAB nach Hoshuyama und Sugiyama (1996). Der FBF wird typischerweise als einfacher DAS-BF implementiert. Die adaptive Störgeräuschunterdrücker (CCAF) der BM verwenden das Ausgangssignal

[122] Siehe Fußnote 105, S. 121.

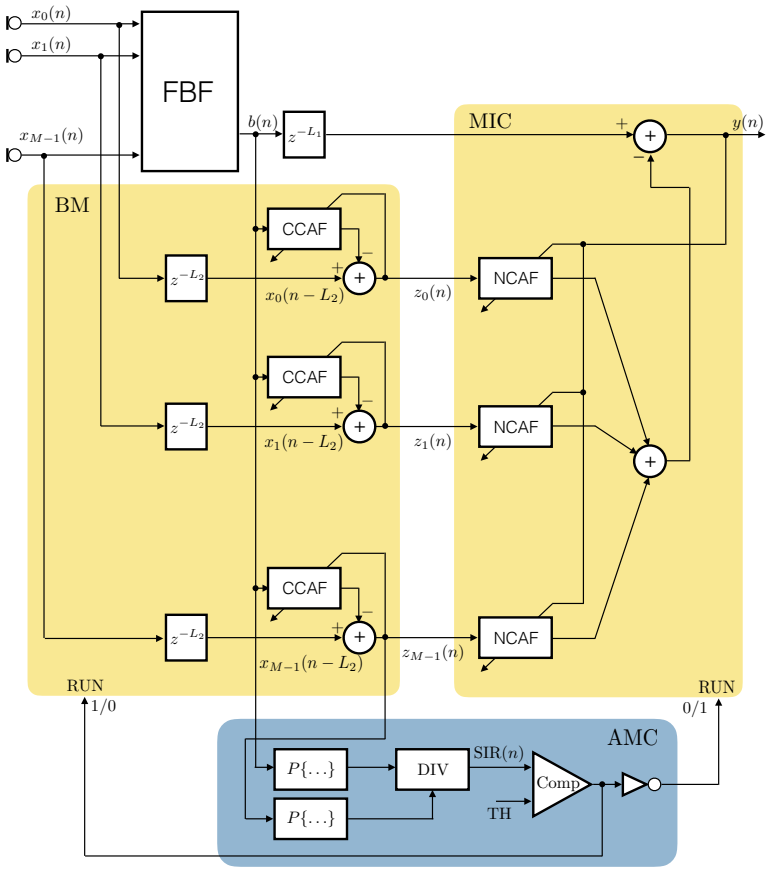


Abbildung 5.3: Blockdiagramm des robusten adaptiven Beamformers (RAB) mit Adaptionskontrolle (AMC) nach Hoshuyama und Sugiyama (1996).

des FBF als Eingangssignal. Dadurch folgt die Nullstelle der Richtcharakteristik der BM der Einfallsrichtung (*direction of arrival*, DOA) des Nutzschalls. Die Filterkoeffizienten der CCAF variieren mit der DOA. Wird das Wachstum der Filterkoeffizienten begrenzt, kann die BM der Nutzquelle nur in einem eingeschränkten Bereich folgen. Der m-te Ausgang der BM ergibt sich zu

$$z_m(n) = x_m(n - L_2) - \mathbf{h}_m^T(n)\mathbf{b}(n), \tag{5.69}$$

$$\mathbf{h}_m(n) = [\ h_{m,0}(n) \quad h_{m,1}(n) \quad \ldots \quad h_{m,K_1-1}(n)\ ]^T, \tag{5.70}$$

$$\mathbf{b}(n) = [\ b(n) \quad b(n-1) \quad \ldots \quad b(n-K_1+1)\ ]^T. \tag{5.71}$$

$\mathbf{h}_m(n)$ ist der Koeffizientenvektor des m-ten CCAF der Länge K_1, $\mathbf{b}(n)$ ist der Vektor der K_1 letzten Abtastwerte des Signals am Ausgang des FBF und L_2 gibt die Anzahl der Abtastwerte an, um die ein Mikrofonsignale $x_m(n)$ verzögert werden muss, damit das System kausal ist. Zur Adaptierung der CCAF wird ein *Normalised-Least-Mean-Squares* (NLMS) Algorithmus verwendet (vgl. Haykin, 2002b, Kap. 9)

$$\mathbf{h}'_m(n+1) = \mathbf{h}_m(n) + \alpha \frac{z_m(n)}{\mathbf{b}(n)^T \mathbf{b}(n) + \delta} \mathbf{b}(n), \tag{5.72}$$

wobei das Wachstum der Koeffizienten nach oben und unten beschränkt wird:

$$\mathbf{h}_m(n+1) = \begin{cases} \mathbf{\Phi}_m, & \text{für} \quad \mathbf{h}'_m(n+1) > \mathbf{\Phi}_m \\ \mathbf{\Psi}_m, & \text{für} \quad \mathbf{h}'_m(n+1) < \mathbf{\Psi}_m \\ \mathbf{h}'_m(n+1), & \text{sonst} \end{cases} \tag{5.73}$$

mit

$$\mathbf{\Phi}_m = [\ \Phi_{m,0} \quad \Phi_{m,1} \quad \dots \quad \Phi_{m,K_1-1}\ ]^T, \tag{5.74}$$

$$\mathbf{\Psi}_m = [\ \Psi_{m,0} \quad \Psi_{m,1} \quad \dots \quad \Psi_{m,K_1-1}\ ]^T. \tag{5.75}$$

Dabei bezeichnet α die Schrittweite und $\delta \ll 1$ die Regularisierungskonstante des NLMS. Abb. 5.4 zeigt die oberen und unteren Schranken $\mathbf{\Phi}$ und $\mathbf{\Psi}$ für die unterschiedlichen Kanäle eines sternförmiges 6-Kanal Mikrofonarrays mit 8 cm Apertur (s. Abb. 5.5) nach vollständiger Adaption der CCAF. Nur wenn die Filterkoeffizienten innerhalb der Schranken liegen, wird das Nutzsignal am Ausgang der BM minimiert.

Die NCAF im MIC subtrahieren diejenigen Signalkomponenten von dem Signal $b(n - L_1)$, die mit den Signalen in den Referenzkanälen $z_m(n)$ korreliert sind. L_1 gibt die Anzahl der Abtastwerte an, um die $b(n)$ verzögert werden muss, damit das System kausal ist. Mit den NCAF der Länge K_2, dem Koeffizientenvektor $\mathbf{w}_m(n)$ und dem Signalvektor $\mathbf{z}_m(k)$ ergibt sich folgende Formulierung für das Signal am

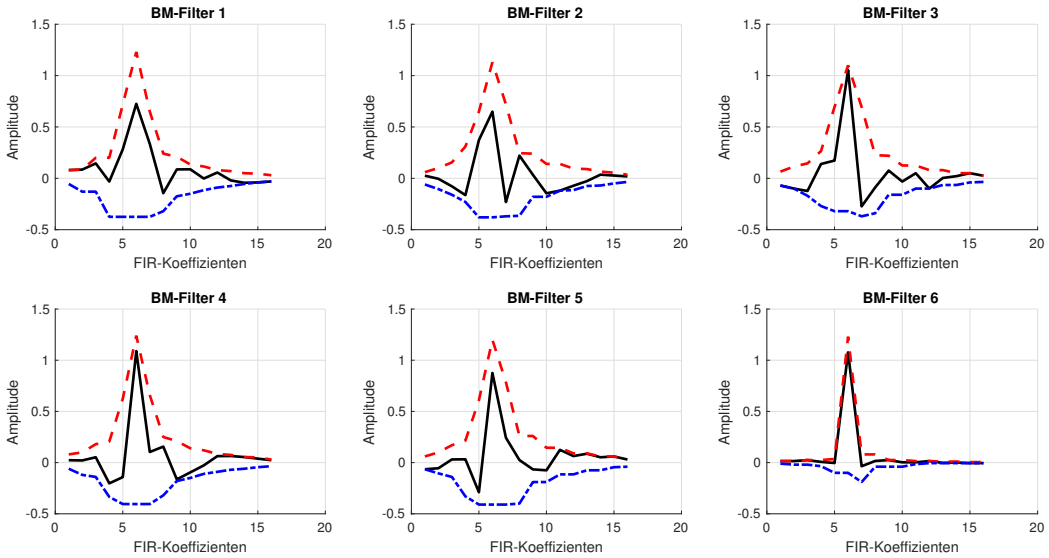


Abbildung 5.4: Koeffizienten der CCAF eines sternförmigen 6-Kanal Mikrofonarrays mit $\pm 15°$ Öffnungswinkel. Die obere Schranke Φ_m ist als rote gestrichelte Linie, die untere Schranke Ψ_m als blaue strichpunktierte Linie dargestellt. Die schwarze durchgezogene Linie zeigt die CCAF Koeffizienten nach vollständiger Adaption des Filters.

Ausgang des RAB:

$$y(n) = b(n - L_1) - \sum_{m=0}^{M-1} \mathbf{w}_m^T(n)\mathbf{z}_m(n), \tag{5.76}$$

$$\mathbf{w}_m(n) = [\ w_{m,0}(n) \quad w_{m,1}(n) \quad \dots \quad w_{m,K_2-1}(m)\ ]^T, \tag{5.77}$$

$$\mathbf{z}(n) = [\ z(n) \quad z(n-1) \quad \dots \quad z(n-K_2+1)\ ]^T. \tag{5.78}$$

Zur Adaptierung der NCAF wird wiederum ein NLMS Algorithmus verwendet

$$\mathbf{w}'_m = \mathbf{w}_m(n) + \beta \frac{y(n)}{\mathbf{z}_m(n)^T\mathbf{z}_m(n) + \delta}\,\mathbf{z}_m(n), \tag{5.79}$$

wobei die quadratische Norm der Filterkoeffizienten

$$\Omega = \mathbf{w}_m'^T\mathbf{w}'_m \tag{5.80}$$

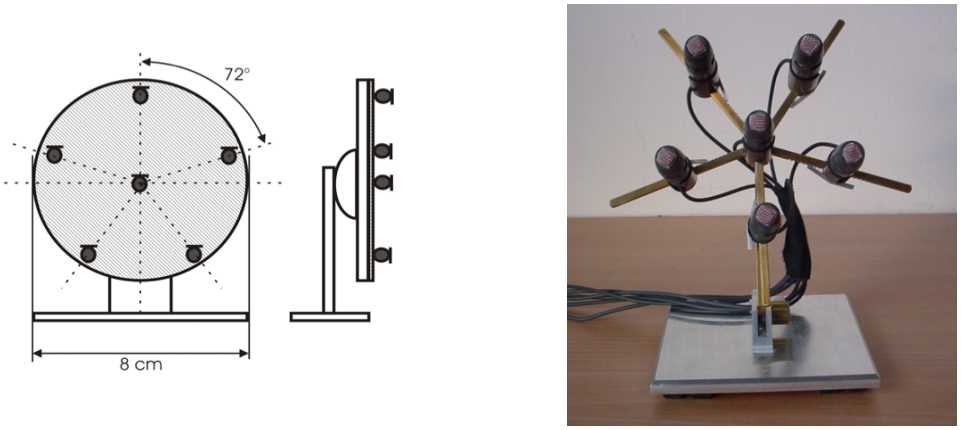


Abbildung 5.5: Sternförmiges 6-Kanal Mikrofonarray mit 8 cm Apertur.

wie folgt beschränkt wird:

$$\mathbf{w}_m(n+1) = \begin{cases} \sqrt{\frac{K}{\Omega}}\,\mathbf{w}'_m, & \text{für} \quad \Omega > K \\ \mathbf{w}'_m, & \text{sonst} \end{cases} \tag{5.81}$$

β bezeichnet die Schrittweite und $\delta \ll 1$ wiederum die Regularisierungskonstante des NLMS. Überschreitet Ω den Schwellwert K, wird $\mathbf{w}_m(n+1)$ mit dem Faktor $\sqrt{K/\Omega}$ skaliert. Dadurch wird ein zu starkes Anwachsen der Filterkoeffizienten verhindert.

Ist in den Ausgangskanälen der BM kein Nutzsignal vorhanden, unterdrückt der MIC die in $b(n-L_1)$ vorhandenen Störanteile. In diesem Fall ist keine Normbeschränkung notwendig. In der Praxis kann die BM das Nutzsignal nicht vollständig unterdrücken (z. B. kommt es durch Raumreflexionen zu Fehlanpassungen der CCAF). Das residuale Nutzsignal in den Ausgangskanälen der BM hat meist einen sehr kleinen Pegel, ist jedoch mit dem Nutzsignal in $b(n-L_1)$ stark korreliert. Ohne Normbeschränkung würde es somit zu unerwünschten Signalauslöschungen kommen. In diesem Fall minimiert die Normbeschränkung die Verzerrungen des entstörten Nutzsignals am Ausgang des RAB.

Ist mindestens ein Eigenwert der Korrelationsmatrix des Signals am Eingang des NCAF gleich Null (d. h. die Korrelationsmatrix ist singulär), konvergiert der NLMS-Algorithmus für den zugehörigen Modus nicht und wird instabil. Die Adaptionseigenschaften des MIC lassen sich in diesem Fall dadurch verbessern,

dass NCAF mit Vergessensfaktor (*leaky norm-constrained adaptive filter*, LNCAF) verwendet werden. Mit Gl. (5.79) ergibt sich

$$\mathbf{w}'_m = (1-\gamma)\mathbf{w}_m(n) + \beta\frac{y(n)}{\mathbf{z}_m(n)^T\mathbf{z}_m(n)+\delta}\,\mathbf{z}_m(n), \tag{5.82}$$

wobei $\gamma \ll 1$ den Vergessensfaktor (*leak factor*) bezeichnet. Der Vergessensfaktor addiert sich zu den Eigenwerten der Korrelationsmatrix und regularisiert dadurch den Adaptionsprozess. Der NLMS-Algorithmus bleibt stabil, konvergiert allerdings nicht mehr vollständig zur optimalen MSE-Lösung. Es entsteht ein Bias. Nach Claesson und Nordholm (1992) kann der Vergessensfaktor auch als additives weißes gaußverteiltes Rauschen am Eingang des Filters interpretiert werden. Dieses führt jedoch zu keinem zusätzlichen Rauschen am Ausgang.

Abb. 5.6 zeigt, wie sich die Breite der Hauptkeule durch die obere und untere Schranke der CCAF beeinflussen lässt. Dabei wurden die Schranken der CCAF so eingestellt, dass sich eine $\pm 15°$ und $\pm 20°$ breite Hauptkeule ergibt. Zusätzlich wurden an den Mikrofonen Pegelschwankungen von $\pm 3\,\mathrm{dB}$ simuliert. Alle Simulationen wurden für das in Abb. 5.5 gezeigte sternförmige Mikrofonarray durchgeführt. Zur Bestimmung der Richtcharakteristik wurde eine ebene Welle in 5°-Schritten von $-\pi/2 \leqslant \phi \leqslant \pi/2$ in der Horizontaleben ($\theta = \pi/2$) um das Array bewegt und die Varianz des Signals am Ausgang des RAB bestimmt. Als Anregungssignal wurde weißes gaußverteiltes Rauschen verwendet, um sicherzustellen, dass die adaptiven Prozesse optimal konvergieren. Die Simulation zeigt, dass der RAB relativ robust gegenüber Pegelschwankungen ist. Die Hauptkeule wird etwas schmäler ($\pm 13°$ anstatt von $\pm 15°$), die Dämpfung der Nebenkeulen bleibt hingegen annähernd gleich. Fehlpositionierungen und Pegelschwankungen der Mikrofone werden durch die adaptiven Prozesse in der BM und im MIC ausgeglichen.

5.2.2 Erhöhung der Robustheit durch verbesserte AMC

Die BM sollte nur dann adaptiert werden, wenn Nutzsignal vorhanden ist. Ist hingegen kein Nutzsignal vorhanden, wird der MIC adaptiert. Die Steuerung über-

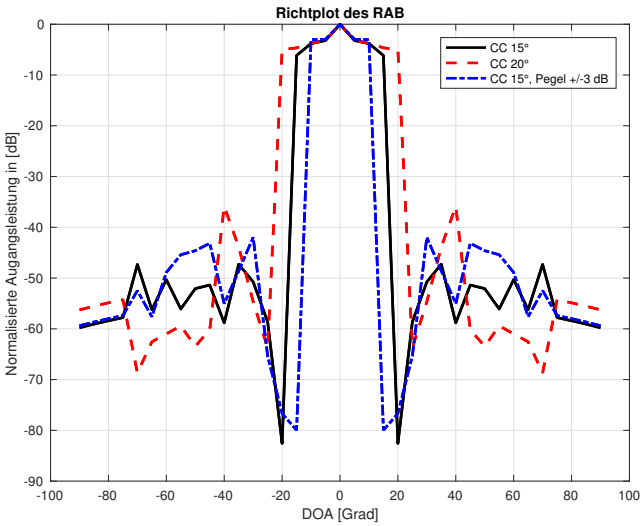


Abbildung 5.6: Richtcharakteristik eines RAB mit CCAF-Schranken für eine ±20° (rote gestrichelte Linie) und ±15° (schwarze durchgezogene Linie) breite Hauptkeule und zufälligen Schwankungen der Mikrofonpegel von ±3 dB (blaue strichpunktierte Linie).

nimmt die AMC. Hoshuyama und Sugiyama verwenden das SIR, um abzuschätzen, ob Nutzsignal vorhanden ist oder nicht. Das SIR lässt sich sehr einfach über eine Schätzung der Leistungen der Signale am Ausgang der BM und am Ausgang des FBF bestimmen:

$$SIR(n) = \frac{S_b(n)}{S_z(n)}, \tag{5.83}$$

$$S_b(n) = (1-\epsilon)S_b(n-1) + \epsilon\, b^2(n), \tag{5.84}$$

$$S_z(n) = (1-\epsilon)S_{z_m}(n-1) + \epsilon\, z_m^2(n), \tag{5.85}$$

mit der Konstante $\epsilon \ll 1$ der zeitlichen Mittelung und $m \in \{0, M-1\}$. In Abb. 5.3 wird zur Schätzung von $S_z(n)$ der Kanal $M-1$ verwendet. Übersteigt das geschätzte SIR einen bestimmten Schwellwert Q, wird die BM adaptiert, ansonsten der MIC:

$$BM_{\text{run}} = \begin{cases} 1, & \text{für} \quad SIR(n) \geq Q \\ 0, & \text{sonst} \end{cases} \tag{5.86}$$

$$MIC_{\text{run}} = \overline{BM}_{\text{run}} \tag{5.87}$$

Abb. 5.7 zeigt die Arbeitsweise der AMC für einen Sprecher aus $\phi = 0°$ (d. h. aus Vorzugsrichtung) und eine Störquelle aus $\phi = -30°$. Beide Quellen befinden sich in der Horizontalebene. Als Störsignal wurde wiederum weißes gaußverteiltes Rauschen verwendet. Der Signal-Interferenz-Abstand beträgt 10 dB. Die Breite der Hauptkeule wurde auf $\pm 15°$ eingestellt. Die Schrittweiten α (siehe Gl. 5.72) und β (siehe Gl. 5.79) wurden mit $\alpha = \beta = 0{,}2$, die Normbeschränkung mit $K = 10$ (siehe Gl. 5.81) und der Schwellwert mit $Q = 5$ angenommen. Das obere Teilbild zeigt das Signal $y(n)$ am Ausgang des RAB, das am Eingang des Mikrofonarrays additiv überlagerte Störsignal $v(n)$ und die Schaltvorgänge der AMC. Der vergrößerte Bildausschnitt zeigt, dass das Störsignal am Ausgang des RAB stark unterdrückt wird. Das mittlere Teilbild zeigt $SIR(n)$ und den Schwellwert Q, das untere Teilbild das Signal $z_m(n)$ am Ausgang der BM.

Der Abbildung ist zu entnehmen, dass unter den simulierten Bedingungen die Pausen des Nutzsignals mit hoher Wahrscheinlichkeit erkannt werden. Das residuale Nutzsignal hat einen um > 20 dB geringeren Pegel als das Nutzsignal am Eingang. Wird die Schrittweite α erhöht, konvergieren die CCAF schneller. Die Signal-Onsets werden besser unterdrückt. Hier muss jedoch ein Kompromiss zwischen Adaptionsgeschwindigkeit und -genauigkeit gefunden werden. Das Nutzsignal soll in der BM soweit unterdrückt werden, dass die Normbeschränkung im MIC zu starke Verzerrungen des entstörten Ausgangssignals vermeidet.

Bei einem Störsprecher aus $\phi = -30°$ mit einem Signal-Stör-Abstand von 3 dB erhöht sich, bei gleichen Arrayparametern, die Fehlerrate der AMC (s. Abb. 5.8). Durch den geringen Signal-Stör-Abstand wird der Störsprecher in den meisten Fällen als Nutzsignal erkannt, obwohl sich dieser außerhalb der Hauptkeule des Arrays befindet. Der MIC wird nicht adaptiert und der Störsprecher, obwohl dieser aufgrund seiner DOA von der BM durchgelassen wird, am Ausgang nicht unterdrückt.

Der SIR-Schätzer weist eine vom DOA der Störquelle abhängige systematische Abweichung auf. Dies hängt vor allem damit zusammen, dass sich die Leistung der Signale am Ausgang der BM, im Vergleich zur Leistung des Signals am Ausgang des FBF, mit der Einfallsrichtung des Störschalls stark ändert. Nähert sich eine Störquelle der Vorzugsrichtung des Arrays, steigt das geschätzte SIR

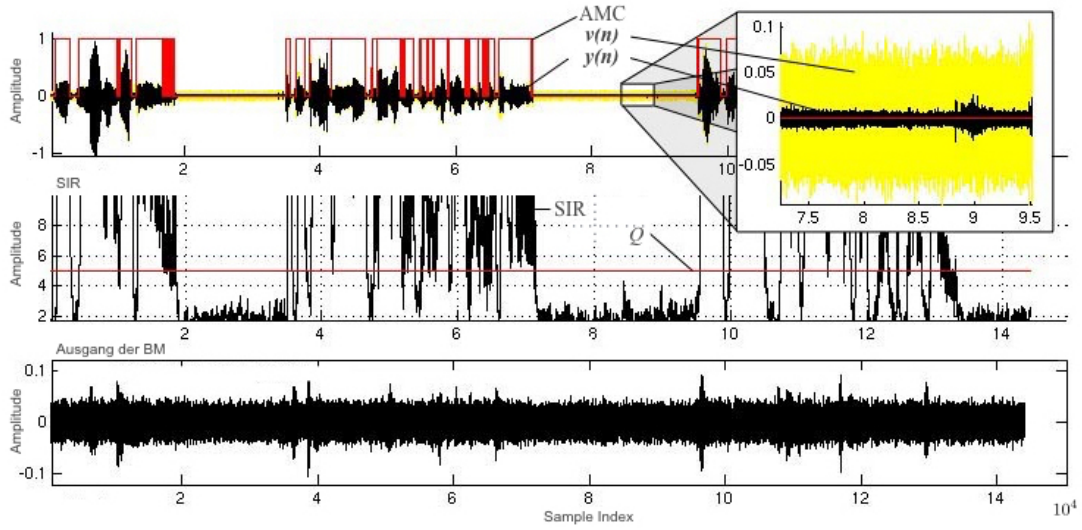


Abbildung 5.7: Arbeitsweise der AMC für ein 6-Kanal Mikrofonarray bei einem Sprecher aus 0°, einer Störquelle (weißes Rauschen) aus −30° und folgenden RAB Parametern: $\alpha = \beta = 0{,}2$, $K = 10$ und $Q = 5$. Der Signal-Interferenz-Abstand beträgt 10 dB. Dargestellt sind das RAB Ausgangssignal $y(n)$, das Störsignal $v(n)$ am Eingang und die Schaltvorgänge der AMC (Teilbild 1); das $SIR(n)$ und der Schwellwert Q (Teilbild 2); das BM-Ausgangssignal $z_m(n)$ (Teilbild 3). Siehe auch Pape (2005, Abb. 4.20).

stark an, sodass der Schwellwert Q permanent überschritten wird. In diesem Fall wird der MIC auch in Nutzsignalpausen nicht adaptiert und das Interferenzsignal nicht unterdrückt. Dem kann dadurch begegnet werden, dass die SIR-Schätzung in Abhängigkeit von $S_z(n)$ korrigiert wird. In Pape (2005) wurde folgender *Limiter/Expander* mit dem Signal $S_z(n)$ am *Sidechain*-Eingang implementiert:

$$SIR'(n) = \frac{S_b(n)}{S_z(n)}, \tag{5.88}$$

$$10\log SIR(n) = \begin{cases} 10\log SIR'(n) - |10\log S_z(n) - T_1|\,, & \text{für } 10\log S_z(n) < T_1 \\ 10\log SIR'(n) + |10\log S_z(n) - T_1|\,, & \text{für } 10\log S_z(n) \geq T_1 \\ 10\log SIR'(n), & \text{für } 10\log S_z(n) < T_2 \end{cases} \tag{5.89}$$

Die Schwellwerte T_1 und T_2 werden dabei in dB angegeben und es gilt $T_2 < T_1$. Unterschreitet $S_z(n)$ den Schwellwert T_1, wird das aus den Signalleistungen

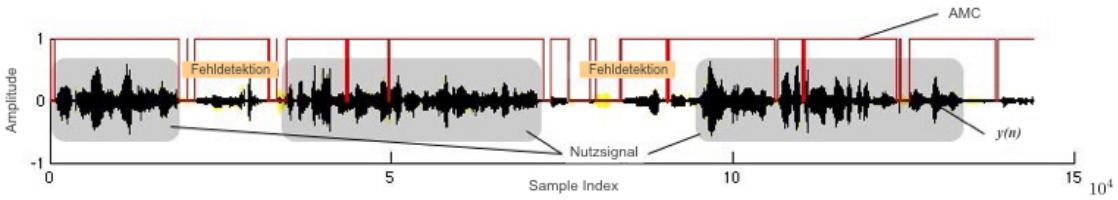


Abbildung 5.8: Arbeitsweise der AMC für ein 6-Kanal Mikrofonarray bei einem Sprecher aus 0°, einem Störsprecher aus −30° und folgenden RAB Parametern: $\alpha = \beta = 0{,}2$, $K = 10$ und $Q = 5$. Der Signal-Interferenz-Abstand beträgt 3 dB. Siehe auch Pape (2005, Abb. 4.22).

geschätzte $SIR(n)$ um die Differenz $|10 \log S_z(n) - T_1|$ nach unten korrigiert. Überschreitet $SIR(n)$ hingegen T_1, wird $SIR(n)$ um die Differenz nach oben korrigiert. Der optionale Schwellwert T_2 legt jenen Grundgeräuschpegel (*noise floor*) fest, bei dem keine aktive Störgeräuschunterdrückung durchgeführt wird, d. h. der MIC wird nicht adaptiert. Wird darüber hinaus der Schwellwerte Q langsam nachgeführt, verringert sich die Fehlerrate der AMC (vgl. Doblinger, 1995; Hui, 2000). Bei Sprache empfiehlt sich zudem, die Signalleistungen über den *Teager-Energy-Operator* zu schätzen. (vgl. Abschnitt 4.1.3).

Bei kleinem Signal-Interferenz-Abstand, lässt sich die AMC über ein zusätzliches Kriterium stabilisieren, welches von der DOA abhängt. Die DOA wird meist über den Zeitversatz der an den unterschiedlichen Mikrofonen eintreffenden Schallwellen geschätzt (*time difference estimation*, TDE). Bei kleiner Arrayapertur und geringer Anzahl an Sensoren ist die TDE sehr fehleranfällig. Mit dem FBF lassen sich H zusätzliche Gitterkeulen erzeugen, die das Schallfeld in einem räumlichen Raster abtasten. Wird nun die Leistung des Signals aus Vorzugsrichtung mit den Leistungen der Signale aus den unterschiedlichen Raumrichtungen ins Verhältnis gesetzt (*steered response power ratio*, SRPR)[123], kann abgeschätzt werden, wie weit die DOA von der Vorzugsrichtung des Mikrofonarrays abweicht:

$$SRPR_\eta(n) = \frac{S_b(n)}{S_\eta(n)}, \quad \text{mit } \eta = 0, 1, \ldots H-1, \tag{5.90}$$

[123]Siehe auch Omologo und Svaizer (1994, 1996), DiBiase et al. (2001), Do et al. (2007).

wobei die Leistung der Signale über eine rekursive Mittelung

$$S_\eta(n) = (1-\varsigma)S_\eta(n-1) + \varsigma\, b_\eta^2(n), \tag{5.91}$$

mit der Konstante $\varsigma \ll 1$ geschätzt wird. $b_\eta(n)$ bezeichnet das Signal der η-ten Gitterkeule des FBF. Tasten die Gitterkeulen das Schallfeld symmetrisch um die Hauptkeule ab, kann ein Schwellwert U festgelegt werden, bei dessen Unterschreiten ein Signal als Interferenz erkannt wird:

$$BM'_{\text{run}} = \begin{cases} 1, & \text{für} \quad SRPR_\eta(n) \geq U \quad \forall\, \eta \in \{0, H-1\} \\ 0, & \text{sonst} \end{cases} \tag{5.92}$$

$$MIC'_{\text{run}} = \overline{BM}'_{\text{run}} \tag{5.93}$$

Die BM wird nur dann adaptiert, wenn das $SRPR_\eta(n)$ den Grenzwert U für alle Gitterkeulen überschreitet und das $SIR(n) \geq Q$ ist, d. h. die Schallquelle sich in dem durch U definierten Vorzugsbereich befindet. Ansonsten wird der MIC adaptiert.

Die SRPR-Schätzung ist nicht immer eindeutig. Starke Raumreflexionen führen zum Beispiel dazu, dass der SRPR-Schätzer abwechselnd zwischen $BM'_{\text{run}} = 1|0$ hin und her schaltet. Dies führt zu einer Fehlanpassung der adaptiven Filter und infolgedessen zu einer Beeinträchtigung des Höreindrucks (d. h. einer Verzerrung des entstörten Nutzsignals und einer starken Färbung des Störgeräusches). Dem kann entgegnet werden, indem der letzte Wert von BM'_{run} über einen gewissen Zeitraum $T_{\max}$ gehalten wird, wenn dieser von 1 auf 0 schaltet, obwohl $SIR(n)$ den Schwellwert Q permanent überschreitet. Dies kann durch eine zusätzliche Haltestufe mit Zeitbegrenzung realisiert werden. Eine schematische Darstellung der modifizierten AMC findet sich in Abb. 5.9.

Die Richtcharakteristik der Gitterkeulen wird durch den FBF bestimmt, welcher typischerweise als DAS-BF ausgeführt wird. Es kommt zu einem systematischen Fehler. Liegt zum Beispiel eine Störquelle in der Nullstelle einer Gitterkeule, wird die Signalleistung $S_\nu(n)$ für diese Richtung falsch geschätzt. Dem kann durch eine breitere Gitterkeule mit kleineren Nebenkeulen entgegnet werden. Dies kann durch

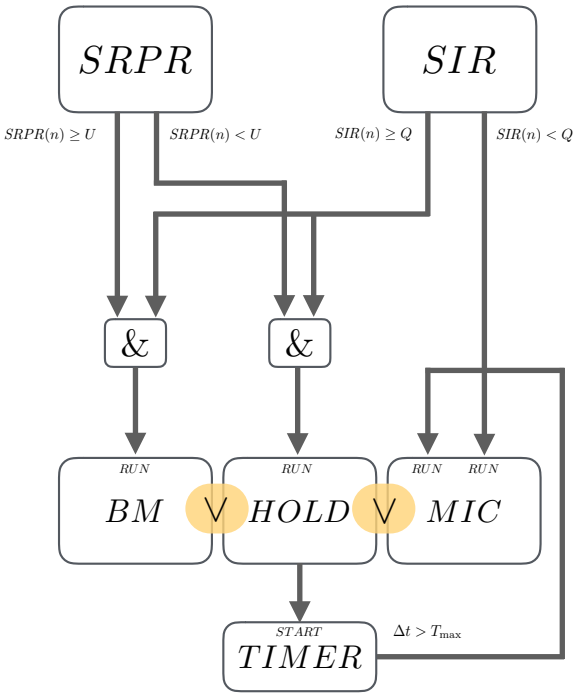


Abbildung 5.9: Schematische Darstellung der modifizierten AMC.

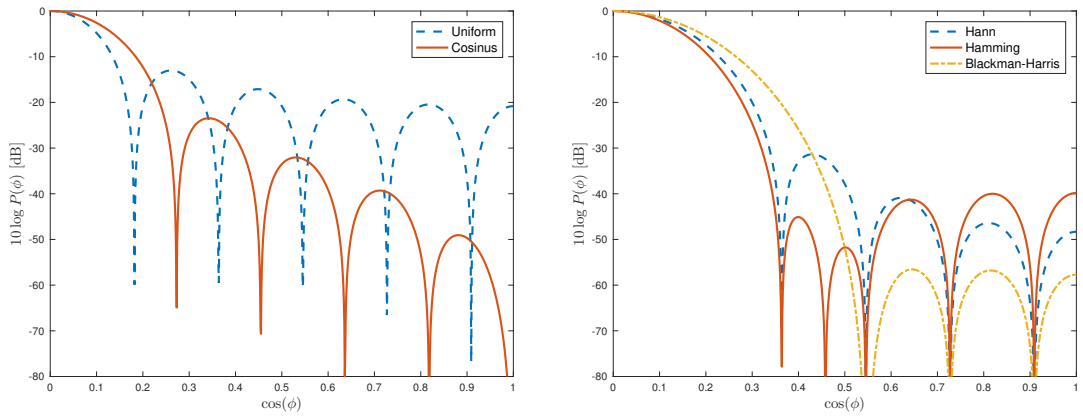


Abbildung 5.10: Einfluss räumlicher Fenster auf die Richtcharakteristik eines DAS-BF. Linkes Teilbild: Rechteckfenster (blau gestrichelte Linie) und Cosinus-Fenster (rote durchgezogene Linie). Rechtes Teilbild: Hann-Fenster (blau gestrichelte Linie), Hamming-Fenster mit $g_0 = 0{,}54$ und $g_1 = 0{,}46$ (rote durchgezogene Linie) und Blackman-Harris-Fenster (orange strichpunktierte Linie).

eine räumliche Fensterung erreicht werden (vgl. Van Trees, 2002, Kap. 3). Der Einfluss unterschiedlicher räumlicher Fenster auf die Richtwirkung eines DAS-BF ist in

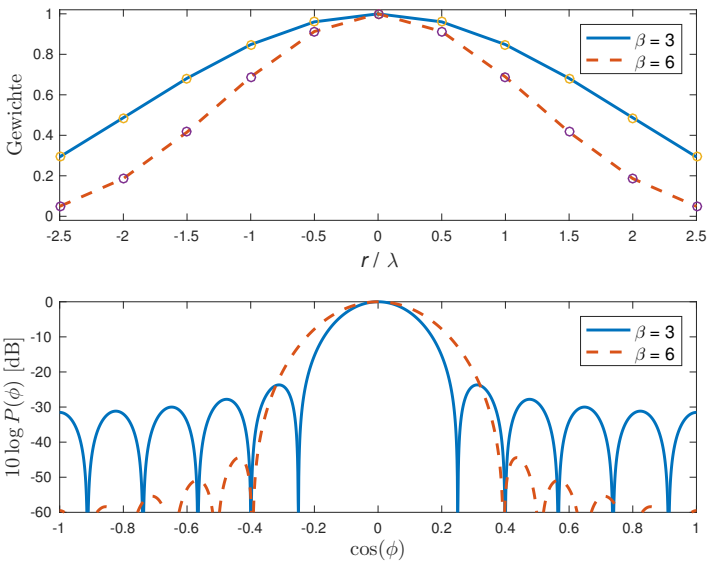


Abbildung 5.11: Einfluss eines Kaiser-Fensters mit unterschiedlichen Steuerparametern auf die Richtcharakteristik eines DAS-BF. Oberes Teilbild: Gewichte der Mikrofonsignale in Abhängigkeit von β. Unteres Teilbild: Richtcharakteristik in Abhängigkeit von β.

den Abbildungen 5.10 und 5.11 dargestellt. Zur besseren Veranschaulichung wurden die Simulationen für ein lineares Mikrofonarray durchgeführt. Die Ergebnisse lassen sich jedoch sehr einfach auf ein sternförmiges Mikrofonarray übertragen, indem die Fenster rotationssymmetrisch um die Hauptachse des Arrays angeordnet werden. Allerdings ist hierzu eine höhere Anzahl an räumlichen Abtastpunkten erforderlich. Aus diesem Grund wurden auch Versuche mit einem sternförmigen 11-Kanal Mikrofonarray mit 8 cm Apertur durchgeführt. Dazu wurde dem in Abb. 5.5 dargestellten Mikrofonarray im Abstand $r = 4\,\text{cm}$ vom Zentrum ein zweiter Ring an Mikrofonen hinzugefügt.

Die Gewichte g_m der in den Abb. 5.10 und 5.11 dargestellten Fenster lassen sich für eine ungerade Anzahl M an Mikrofonen wie folgt berechnen (vgl. Van Trees, 2002, Kap. 3):

- Rechteckfenster:

$$g_m = \frac{1}{M}$$

- Cosinus-Fenster:

$$g(\tilde{m}) = \sin\left(\frac{\pi}{2M}\right)\cos\left(\pi\frac{\tilde{m}}{M}\right), \quad -\frac{M-1}{2} \leq \tilde{m} \leq \frac{M-1}{2}.$$

- Hamming-Fenster:

$$g(\tilde{m}) = g_0 + g_1\cos\left(\frac{2\pi\tilde{m}}{M}\right), \quad \tilde{m} = -\frac{M-1}{2},\ldots\frac{M-1}{2},$$

- Blackman-Harris-Fenster:

$$g(\tilde{m}) = 0{,}42{+}0.5\cos\left(\frac{2\pi\tilde{m}}{M}\right){+}0.08\cos\left(\frac{4\pi\tilde{m}}{M}\right), \quad \tilde{m} = -\frac{M-1}{2},\ldots\frac{M-1}{2}.$$

- Kaiser Fenster:

$$g(\tilde{m}) = I_0\left(\beta\sqrt{1-\left[\frac{2\tilde{m}}{M}\right]}\right), \quad -\frac{M-1}{2} \leq \tilde{m} \leq \frac{M-1}{2},$$

  wobei I_0 die modifizierte Bessel-Funktion nullter Art ist.

Die Richtcharakteristik eines DAS-BF lässt sich auch über einen LS-Ansatz berechnen (vgl. Algazi und Suk, 1975; Okuda et al., 1998; Doclo und Moonen, 2002; Benesty et al., 2008a, Kap. 3.4). Die folgenden Betrachtungen gelten wiederum für ein lineares Mikrofonarray, lassen sich aber sehr einfach auf ein sternförmiges Mikrofonarray übertragen, indem die vorgegebene Systemantwort als rotationssymmetrisch um die Hauptachse des Arrays angenommen wird. Werden die Filterkoeffizienten eines M-Kanal DAS-BF in einem Vektor $\mathbf{w} = [\ w_0,\ \ w_1,\ \ \ldots,\ \ w_{M-1}\ ]^T$ zusammengefasst, berechnet sich die Richtcharakteristik mit Gl. (5.57) zu

$$B(\Omega;\theta,\phi) = \mathbf{w}_F^H(e^{j\Omega})\mathbf{v}(\Omega,\mathbf{k}).$$

In den folgenden Betrachtungen wird $B(\Omega;\theta,\phi)$ vereinfacht als $B(\phi)$ geschrieben. Die Filterkoeffizienten werden nun dahingehend optimiert, dass $B(\phi)$ eine vorgegebene Systemantwort $B_d(\phi)$ möglichst genau annähert (vgl. Abb. 5.12). Die

quadratische Fehlerfunktion lässt sich wie folgt formulieren:

$$e^2 = \int_0^{\pi} \vartheta(\phi)|B(\phi) - B_d(\phi)|^2 d\phi, \tag{5.94}$$

wobei $\vartheta(\phi)$ ist eine reellwertige Gewichtsfunktionen, mit der bestimmte Winkel in der Optimierung bevorzugt werden können. Durch Substitution ergibt sich folgendes Optimierungskriterium

$$e^2 = \mathbf{w}^T\mathbf{Q}\mathbf{w} - 2\mathbf{w}^T\mathbf{p} + \int_0^{\pi} \vartheta(\psi)\,|B_d(\phi)|^2\,d\phi, \tag{5.95}$$

$$\mathbf{Q} = \int_0^{\pi} \vartheta(\phi)\mathbf{v}(\phi)\mathbf{v}^H(\phi)d\phi, \tag{5.96}$$

$$\mathbf{p} = \int_0^{\pi} \vartheta(\phi)\mathrm{Re}\left[\mathbf{v}(\phi)B_d(\phi)\right]d\phi. \tag{5.97}$$

Wird die Fehlerfunktion e^2 nach $\mathbf{w}$ abgeleitet und zu Null gesetzt, ergibt sich für die Koeffizienten des Optimalfilters

$$\mathbf{w}_{LS} = \mathbf{Q}^{-1}\mathbf{p}. \tag{5.98}$$

Die Richtcharakteristik eines linearen 11-Kanal LS-BF, welcher für einen Durchlassbereich von $\phi = \pm 30°$ (rote strichpunktierte Linie) optimiert wurde, ist in Abb. 5.12 dargestellt (blaue durchgezogene Linie). Zum Vergleich wird auch die Richtcharakteristik des DAS-BF (grüne gestrichelte Linie) abgebildet. Durch zeitliche Vorentzerrung kann die Vorzugsrichtung zum Beispiel um $\pm 60°$ gedreht werden, um Gitterkeulen für die SRPR-Schätzung zu erzeugen. Dadurch ergibt sich eine sehr hohe Kanaltrennung zwischen der Hauptkeule und den Gitterkeulen. Zudem kann durch diesen Ansatz der Einfluss der Nullstellen der Gitterkeulen auf die SRPR-Schätzung weitgehend vermieden werden.

Für einen breitbandigen Entwurf kann der FBF als FAS-BF ausgeführt werden. Hierbei werden die optimalen Filterkoeffizienten über den LS-Ansatz für verschiedene Frequenzen bestimmt und daraus Transversalfilter in jedem Mikrofonkanal hergeleitet.

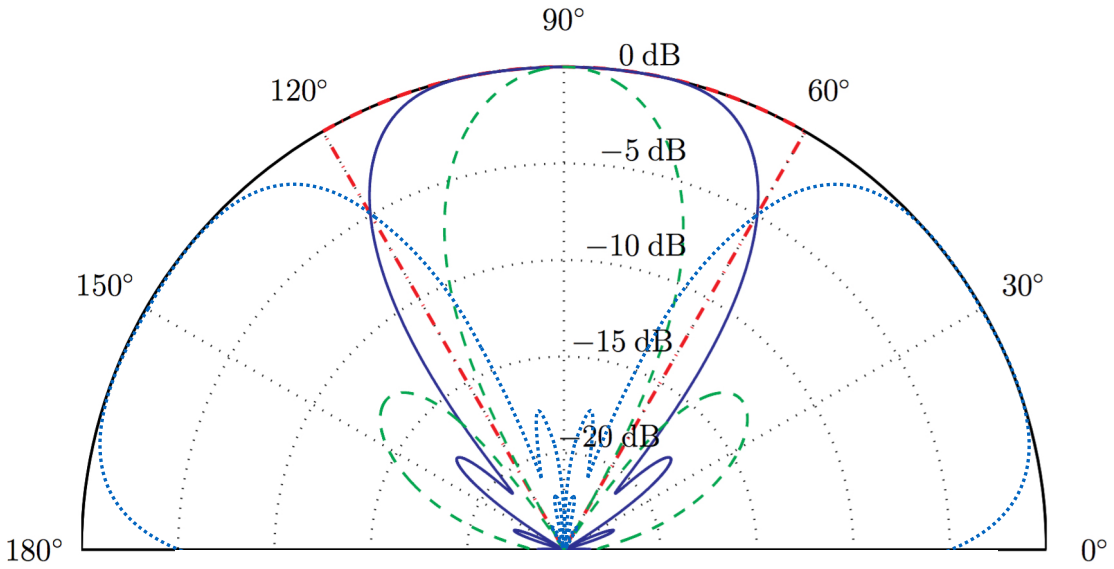


Abbildung 5.12: LS-Optimierung der Richtcharakteristik eines DAS-BF: Vorgegebener Winkelbereich (rote strichpunktierte Linie), LS-BF (blaue durchgezogene Linie), DAS-BF (grüne gestrichelte Linie) und die um ±60° verschobenen Gitterkeulen (blaue punktierte Linien).

5.2.3 Messtechnische Evaluierung

Der in diesem Kapitel vorgestellte RAB mit modifizierter AMC wurde in Pure Data (einer grafischen Programmiersprache zur Echtzeit-Signalverarbeitung)[124] implementiert. Die C++ Bibliotheken wurden als freie Software unter der *GNU General Public License* (GPL)[125] veröffentlicht.

Zur Evaluierung wurde der RAB ausführlich in unterschiedlichen akustischen Bedingungen getestet. Abb. 5.13 zeigt den Messaufbau in einem Büro. Mit einem Kunstkopf mit künstlichem Mund als Nutzquelle, einem kugelförmig abstrahlenden Lautsprecher als Störquelle und einem künstlichen Mund als Störsprecher. Das Büro hat in etwa 0,4 s Nachhallzeit.

Die Richtcharakteristik des RAB wurde in einem schalltoten Raum mit einer unteren Grenzfrequenz von 150 Hz gemessen. Das Mikrofonarray wurde auf einem ferngesteuerten Drehteller montiert der mit einer Schrittweite von 5° betrieben

[124] `https://puredata.info/`
[125] `https://www.gnu.org/`

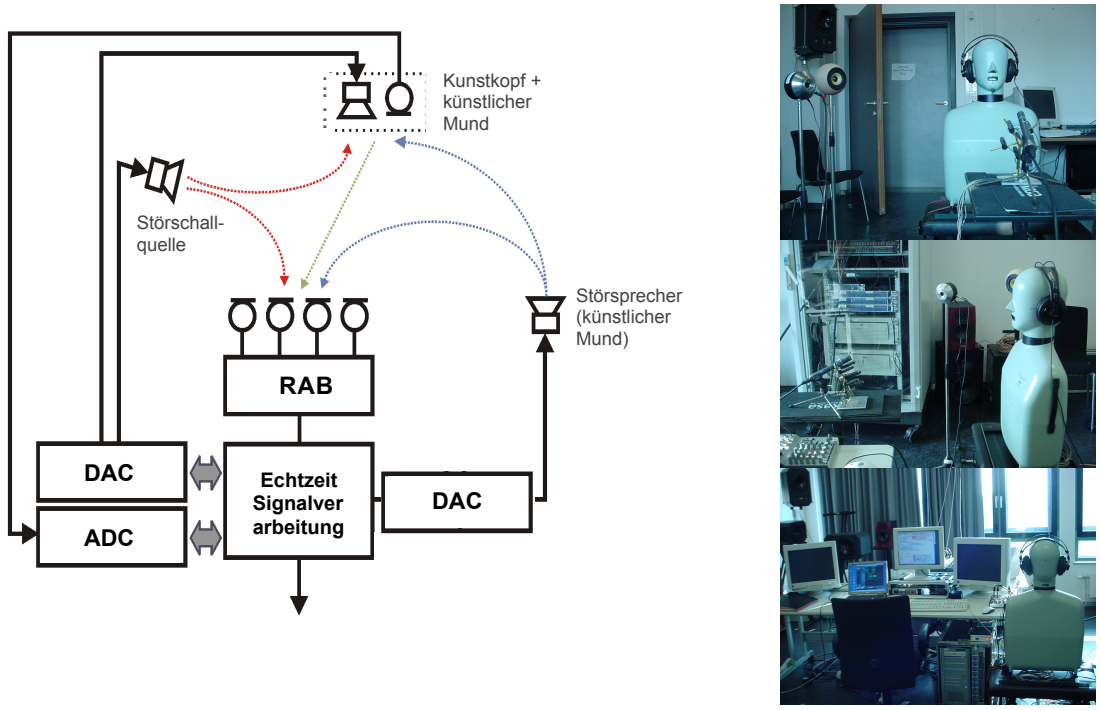


Abbildung 5.13: Messaufbau zur Evaluierung des RAB (Büro mit $T_{60} \approx 0{,}4\,\mathrm{s}$).

wurde. Als Störquellen wurden kugelförmig abstrahlende Lautsprecher verwendet, als Störsprecher ein künstlicher Mund.

Die im schalltoten Raum gemessene Richtcharakteristik des RAB mit modifizierter AMC ist in Abb. 5.15. Eine ausführliche Diskussion der Messergebnisse und zugehörigen Messprotokolle findet sich in Pape (2005, Kap. 4.5.6).

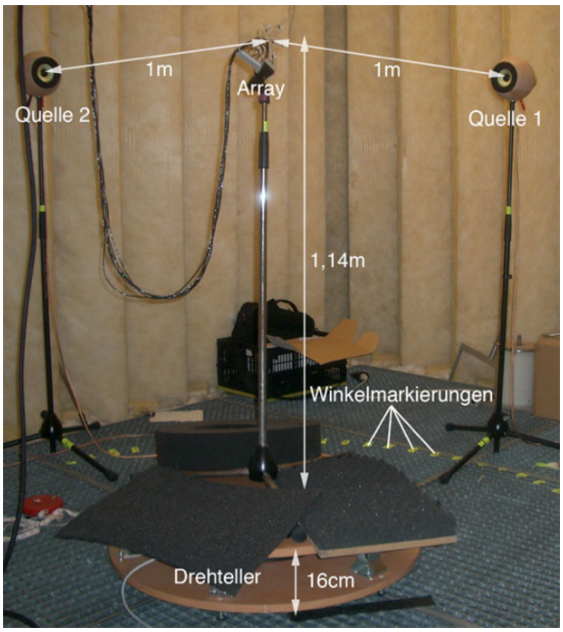


Abbildung 5.14: Messaufbau zur Evaluierung des RAB (schalltoter Raum).

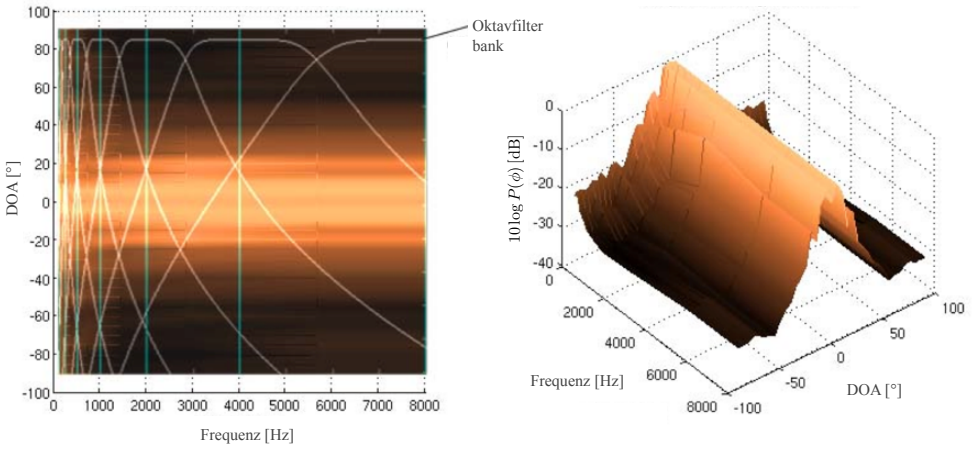


Abbildung 5.15: Gemessene Richtcharakteristik des RAB mit modifizierter AMC.

6

Breitbandige Signalaufbereitung in mehrkanaligen Mikrofonanwendungen: Modale Beamformer

Dieses Kapitel befasst sich mit dem Entwurf modaler Beamformer zur Signalaufbereitung mit möglichst hoher zeitlicher und räumlicher Bandbreite. Neben der Herleitung optimaler modaler Beamformer[126] wird eine neue Methode für den Entwurf robuster Mikrofonarrays vorgestellt, mit dem sich der Interpolationsfehler innerhalb der Kugel mit möglichst wenigen Mikrofonen minimieren lässt.[127]

6.1 Schallfeldbeschreibung in Kugelkoordinaten

Dieser Abschnitt fasst die zur Entwicklung modaler Mikrofonarray-Beamformer benötigte Theorie zusammen. Eine allgemeine Übersicht über Kugelflächenfunktionen (oder auch sphärisch harmonische Funktionen; *spherical harmonics*, SH)

[126]Der Entwurf modaler Beamformer (s. Abschnitt 6.2) wurde im Rahmen des vom französischen Wissenschaftsfonds (*Agence Nationale de la Recherche*, ANR) finanzierten CONTINT Projekt „*Sample Orchestrator 2*“ (SOR2, Projektlaufzeit 2009-2012) in Kooperation mit der Ben Gurion University of the Negev, Israel, durchgeführt.

[127]Die Studien zum Entwurf eines robusten modalen Beamformers (s. Abschnitt 6.5) wurden zum Teil im Rahmen einer Kooperation des *Institut de Recherche et Coordination Acoustique/-Musique* (IRCAM-CNRS-UPMC, UMR 9912 STMS) mit dem Institut für Schallforschung der Österreichischen Akademie der Wissenschaften (ARI/ÖAW) durchgeführt. Die Kooperation wurde einerseits durch das Projekt „*EarToy*“ (Projektnummer ANR RIAM 004 02) und das Projekt „*Amadée*“ (Projektnummer FR 16/2013) des französischen Wissenschaftsfonds ANR und andererseits durch das Projekt „*Wavelets and frames for the space-time-frequency representation of acoustic wave fields*“ des Programms „*Research in Paris*“ der Stadt Paris teilfinanziert.

sowie der Approximationstheorie auf der Einheitskugel ist zum Beispiel in Heine (1861), Müller (1966), Freeden und Schreiner (2009), Atkinson und Han (2012), Dai und Xu (2013) und Michel (2013) zu finden. Der Entwurf und die Herleitung modaler Beamformer wird in Williams (1999, Kap. 6) und Rafaely (2015) ausführlich diskutiert. Eine Zusammenfassung der wichtigsten in dieser Arbeit verwendeten Funktionen findet sich in Anhang B.

Kugelflächenfunktionen. Jeder Punkt $\mathbf{x} \in \mathbb{S}^2$ auf der Einheitskugel $\mathbb{S}^2 := \{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\| = 1\}$ kann wie folgt in Kugelkoordinaten beschrieben werden

$$\mathbf{x} = (x_1, x_2, x_3) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)\,, \tag{6.1}$$

wobei $\theta \in [0, \pi]$ den Polarwinkel und $\phi \in [0, 2\pi)$ den Azimutwinkel bezeichnet. Der Raum $L^2(\mathbb{S}^2)$ der quadratintegrierbaren Funktionen auf $\mathbb{S}^2$ ist (über dem Körper $\mathbb{K}$ mit $\mathbb{K} = \mathbb{R}$ bzw. $\mathbb{K} = \mathbb{C}$) definiert als

$$L^2(\mathbb{S}^2) := \left\{ f : \mathbb{S}^2 \to \mathbb{K} \;\middle|\; \left(\int_{\mathbb{S}^2} |f(\mathbf{x})|^2 \, d\Omega(\mathbf{x}) \right)^{\frac{1}{2}} < \infty \right\}, \tag{6.2}$$

wobei sich das skalare Oberflächenelement auf der Kugel zu $d\Omega(\mathbf{x}) = \sin\theta \, d\theta d\phi$ ergibt. $L^2(\mathbb{S}^2)$ ist ein Banachraum. Ein Banachraum ist ein normierter Vektorraum über dem Körper $\mathbb{K}$, der vollständig bezüglich der induzierten Metrik ist (Alt, 2012, S. 28 ff). Das heißt, dass jede Cauchy-Folge in diesem Vektorraum konvergiert. Ein Hilbertraum $\mathcal{H}$ ist ein Banachraum, der dadurch ausgezeichnet ist, dass die Norm $\|x\| = \sqrt{\langle x, x \rangle}$ für alle $f, g \in L^2(\mathbb{S}^2)$ von einem Skalarprodukt

$$\langle f, g \rangle = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} f(\theta, \phi) \overline{g(\theta, \phi)} \sin\theta d\theta d\phi \tag{6.3}$$

induziert wird (s. Werner, 2011, Definition V.1.4). Der Querstrich $\overline{(\cdot)}$ bezeichnet die komplexe Konjugation. Somit lassen sich alle in Banachräumen definierten Begriffe auf $L^2(\mathbb{S}^2)$ anwenden. Hilberträume gehören zu den wichtigsten Räumen der Funktionalanalysis, insbesondere in der Lösungstheorie partieller Differentialgleichungen (siehe z. B. Christensen, 2010, Kap. 4; Werner, 2011, Kap. 5; Atkinson und Han, 2012, Kap. 1).

Ein Hilbertraum lässt sich in eine Hilbertraumsumme $L^2(\mathbb{S}^2) = \bigoplus_{n=0}^{\infty} H_n$ entwickeln, wobei H_n einen $(2n+1)$-dimensionalen Raum bezeichnet, der von den Kugelflächenfunktionen $Y_n^m(\theta, \phi)$ mit der Ordnung $n \in \mathbb{N}$ und $-n \leq m \leq n$ aufgespannt wird. Die Kugelflächenfunktionen bilden somit ein vollständiges Orthonormalsystem auf der Einheitskugel (s. a. Lebedev 1965, Kap. 7–8; Arfken 1985, Kap. 12; Williams 1999, Kap. 6; Freeden und Schreiner 2009, Kap. 3). Die Vollständigkeit garantiert, dass sich jede quadratintegrierbare Funktion auf der Kugeloberfläche als Linearkombination von SH beschreiben lässt. Die Orthonormalität wiederum bewirkt die einfache Form der Koeffizienten dieser Darstellung. Die Kugelflächenfunktionen

$$Y_n^m(\theta, \phi) = \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta)\, e^{im\phi} \tag{6.4}$$

sind die Eigenfunktionen des Laplace-Operators auf der Kugel und lösen die Helmholtz-Wellengleichung (Gl. 5.6).[128] Sie beschreiben die Schwingungsmoden auf einer Kugeloberfläche. Y_n^m hat m Schwingungen entlang des Breitenkreises. Die assoziierten Legendre-Funktionen $P_n^m(\cos\theta)$ beschreiben stehende Wellen entlang der Längenkreise, deren Amplitude vom Nordpol her abnimmt. Der Vorfaktor ergibt sich aus der Normierung bei Integration über die Kugeloberfläche

$$\int_{\mathbb{S}^2} |Y_n^m|^2 \, d\Omega = 1 \tag{6.5}$$

mit $\Omega = (\theta, \phi)$. Die assoziierten Legendre-Funktionen sowie die in der Literatur am häufigsten verwendeten Normierungen der SH sind in Anhang B zusammengefasst.

Sphärische Fouriertransformation. Die sphärische Fouriertransformation (*spherical harmonic transform*, SHT) einer quadratintegrierbaren Funktion $f(\theta, \phi)$ auf der Einheitskugel, mit den Fourierkoeffizienten f_{nm}, ist gegeben als (vgl. Driscoll und Healy, 1994):

$$f_{nm} = \int_{\mathbb{S}^2} f(\theta, \phi) \overline{Y_n^m(\theta, \phi)} d\Omega = SHT\{f(\theta, \phi)\}, \tag{6.6}$$

[128] Siehe auch Morse und Feshbach (1953, S. 1264 ff), Morse und Ingard (1968, Kap. 7.2), Williams (1999, Kap. 6), Rafaely (2015, Kap. 2).

$$f(\theta,\phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} f_{nm} Y_n^m(\theta,\phi) = SHT^{-1}\{f_{nm}\}. \tag{6.7}$$

Wird $f(\theta,\phi)$ auf der Kugeloberfläche mit Q Sensoren abgetastet, führt dies auf die raumdiskrete sphärische Fouriertransformation (*discrete spherical harmonic transform*, DSHT):

$$\mathbf{f} = \mathbf{Y}\mathbf{f}_{nm}, \tag{6.8}$$

$$\mathbf{f}_{nm} = \mathbf{S}\mathbf{f} = \mathbf{Y}^{\dagger}\mathbf{f}, \tag{6.9}$$

mit dem $Q \times 1$ Vektor der räumlichen Abtastwerte

$$\mathbf{f} = \left[\; f(\theta_1,\phi_1), \quad f(\theta_2,\phi_2), \quad \ldots \quad f(\theta_Q,\phi_Q) \; \right]^T, \tag{6.10}$$

dem $(N+1)^2$ Vektor der sphärischen Fourierkoeffizienten

$$\mathbf{f}_{nm} = \left[\; f_{0,0}, \quad f_{-1,1}, \quad \ldots \quad f_{N,N} \; \right]^T, \tag{6.11}$$

und der $Q \times (N+1)^2$ Matrix $\mathbf{Y}$, welche die sphärischen Fourierkoeffizienten für jeden Abtastpunkt auf der Kugel beinhaltet. Die Fourier-Matrix wird ganz allgemein (unabhängig von der Regularisierung) mit $\mathbf{S}$ bezeichnet.

Lösungen der Wellengleichung. Die Wellengleichung in Kugelkoordinaten kann durch Separation der Variablen gelöst werden:

$$p(r,\theta,\phi,t) = R(r)\,\Theta(\theta)\,\Phi(\phi)\,T(t), \tag{6.12}$$

mit der Radiusfunktion $R(r)$, den zwei Raumwinkelfunktionen $\Theta(\theta)$ und $\Phi(\phi)$ sowie der Zeitfunktion $T(t)$. Dies führt auf die Lösungen für stehende Wellen (vgl. Williams, 1999, Kap. 6.2)

$$p(r,\theta,\phi,\omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \left(A_{nm} j_n(kr) + B_{nm} y_n(kr)\right) Y_n^m(\theta,\phi) \tag{6.13}$$

und fortschreitende Wellen

$$p(r,\theta,\phi,\omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \left(C_{nm} h_n^{(1)}(kr) + D_{nm} h_n^{(2)}(kr) \right) Y_n^m(\theta,\phi), \tag{6.14}$$

wobei $j_n(kr)$ und $y_n(kr)$ die sphärischen Bessel-Funktionen der ersten und der zweiten Art und $h_n^{(1)}(kr)$ und $h_n^{(2)}(kr)$ die sphärischen Hankel-Funktionen der ersten und der zweiten Art bezeichnen (siehe Anhang B). Wellenlänge und Kreisfrequenz hängen über $\omega = k\,c$ zusammen. Die Zeitfunktion ist implizit enthalten und wird erst durch Rücktransformation in den Zeitbereich deutlich:

$$p(r,\theta,\phi,t) = \frac{1}{2\pi} \int_{\infty}^{\infty} p(r,\theta,\phi,\omega) e^{-i\omega t} d\omega. \tag{6.15}$$

Die Lösungen der Wellengleichung in Kugelkoordinaten werden in Morse und Feshbach (1953, S. 1264 ff), Morse und Ingard (1968, Kap. 7.2), Williams (1999, Kap. 6) und Rafaely (2015, Kap. 2) ausführlich diskutiert.

Zerlegung in ebene Wellen. Befinden sich die Schallquellen im Fernfeld des Arrays, lässt sich das Schallfeld als gewichtete Überlagerung von aus allen Richtungen einfallenden ebenen Wellen darstellen (siehe auch Abschnitt 5.1.1). Trifft eine ebene Welle aus Richtung (θ_ℓ, ϕ_ℓ) auf ein Array, ergibt sich der Schalldruck p_ℓ an der Raumposition (r,θ,ϕ) zu (vgl. Rafaely, 2004; Li und Duraiswami, 2007):[129]

$$p_\ell(kr,\theta,\phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} b_n(kr) Y_n^m(\theta_\ell,\phi_\ell)^* Y_n^m(\theta,\phi). \tag{6.16}$$

Der Ausdruck $b_n(kr)$ beschreibt die modalen Amplituden und hängt von der Beschaffenheit des Mikrofonarrays ab. Die modalen Amplituden lassen sich für eine offene Kugel und eine schallharte, geschlossene Kugel mit einem Radius r_m

[129] Im Folgenden wird, wie in der Signalverarbeitung üblich, die komplexe Konjugation mit $(\cdot)^*$ bezeichnet. Zur Erhöhung der Lesbarkeit wird an manchen Stellen auch die in der Mathematik übliche Schreibweise $\overline{(\cdot)}$ verwendet.

über folgenden Ausdruck beschreiben:

$$b_n(kr) = 4\,\pi\,i^n \begin{cases} \left(j_n^{(1)}(kr) - \dfrac{j_n^{(1)\prime}(kr_m)}{h_n^{(2)\prime}(kr_m)} h_n^{(2)}(kr) \right), & \text{schallharte Kugel} \\ j_n^{(1)}(kr), & \text{offene Kugel} \end{cases} \tag{6.17}$$

$j_n(kr)$ und $h_n(kr)$ bezeichnen die sphärischen Bessel- und Hankel-Funktionen, und $j_n'(kr)$ und $h_n'(kr)$ deren Ableitungen. Für den Radius der schallharten Kugel gilt $r_m \leq r$. Die modalen Amplituden (auch holographische Funktionen genannt) sind in Abb. 6.1 dargestellt.

Wird das Schallfeld von einer Punktschallquelle erzeugt, die sich im Nahfeld eines Arrays befindet, lassen sich die modalen Amplituden nach Fisher und Rafaely (2008, 2011) wie folgt berechnen:

$$b_n(kr) = 4\,\pi\,i\,k\,h_n(kr_s) \begin{cases} \left(j_n^{(1)}(kr) - \dfrac{j_n^{(1)\prime}(kr_m)}{h_n^{(2)\prime}(kr_m)} h_n^{(2)}(kr) \right), & \text{schallharte Kugel} \\ j_n^{(1)}(kr), & \text{offene Kugel} \end{cases} \tag{6.18}$$

wobei $r_s \geq r_m$ der Abstand der Schallquelle zum Zentrum der Kugel ist.

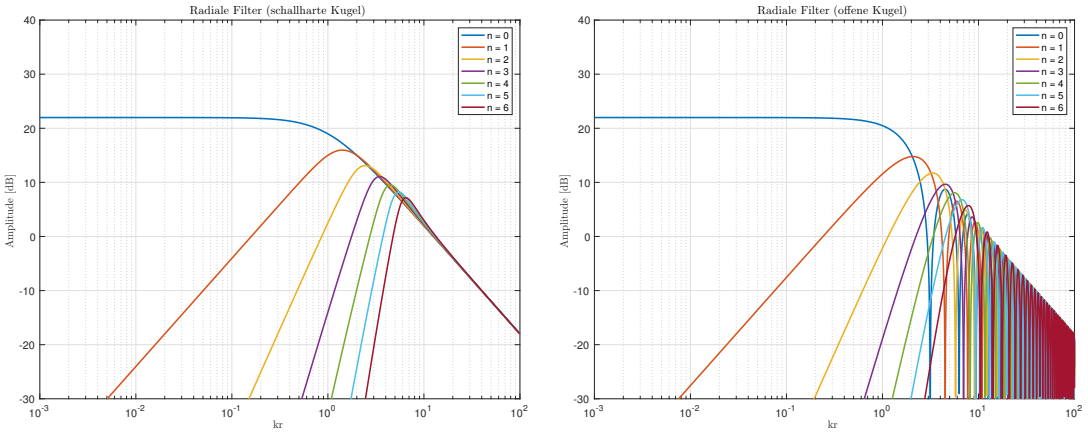


Abbildung 6.1: Modale Amplituden $b_n(kr)$ für eine schallharte Kugel (linkes Teilbild) und eine offene Kugel (rechtes Teilbild) bis zur harmonischen Ordnung $N = 6$.

Durch Anwenden der sphärischen Fouriertransformation (Gl. 6.6) auf (Gl. 6.16), ergeben sich die Fourierkoeffizienten $p_{\ell,nm}$ für eine aus Richtung (θ_ℓ, ϕ_ℓ) einfallende ebene Welle zu

$$p_{\ell,nm}(k) = b_n(kr) Y_n^m(\theta_\ell, \phi_\ell)^*. \tag{6.19}$$

Treffen unendlich viele ebene Wellen aus allen Richtungen $\Omega_\ell = (\theta_\ell, \phi_\ell)$ mit einer Amplitudendichte $\varphi(k, \theta_\ell, \phi_\ell)$ auf ein Array ein, lässt sich der Schalldruck auf der Kugeloberfläche durch Integration von Gl. (6.19) über alle Richtungen Ω_ℓ berechnen:

$$\begin{aligned} p_{nm}(k) &= \int_{\Omega_\ell \in \mathbb{S}^2} \varphi(k, \theta_\ell, \phi_\ell) b_n(kr) Y_n^m(\theta_\ell, \phi_\ell)^* d\Omega_\ell \\ &= \varphi_{nm}(k) b_n(kr), \end{aligned} \tag{6.20}$$

wobei $\varphi_{nm}(k)$ die sphärische Fouriertransformierte der Dichtefunktion $\varphi(k, \theta_\ell, \phi_\ell)$ ist. Aus Gl. (6.20) ergibt sich

$$\varphi_{nm}(k) = p_{nm}(k) \frac{1}{b_n(kr)}. \tag{6.21}$$

Das Schallfeld lässt sich sehr einfach in ebene Wellen zerlegen (*plane wave decomposition*, PWD), indem die sphärischen Fourierkoeffizienten des Schalldrucks, $p_{nm}(k)$, durch die modalen Amplituden, $b_n(kr)$, dividiert werden. Die PWD für drei aus unterschiedlichen Richtungen einfallenden ebenen Wellen ist in Abb. 6.2 dargestellt.

Aus Gl. (6.17) wird ersichtlich, dass die modalen Amplituden einer offenen Kugel proportional zur sphärischen Bessel-Funktion der ersten Art sind. Durch die Nullstellen der sphärischen Bessel-Funktion (siehe Abb. B.1) kommt es bei der Inversion $1/b_n(kr)$ zu numerischen Instabilitäten (vgl. Abhayapala und Ward, 2002; Gover et al., 2004; Rafaely, 2011). Dadurch wird der nutzbare Frequenzbereich kr stark begrenzt. Die Inverse der modalen Amplituden (auch inverse holographische Funktionen) ist in Abb. 6.3 dargestellt.

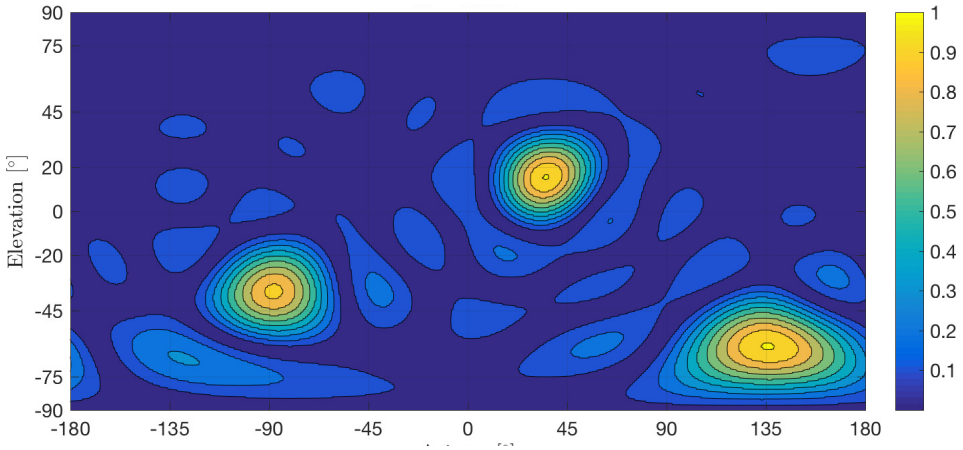


Abbildung 6.2: PWD für drei aus unterschiedlichen Richtungen einfallenden ebenen Wellen ($N = 7$).

Wird anstatt der offenen Kugel eine schallharte Kugel verwendet, können die störenden Resonanzen weitgehend vermieden werden (vgl. Meyer und Elko, 2002). In realen Anwendungen lassen sich schallharte Kugeln allerdings nur mit relativ kleinen Radien fertigen. Bei großen Aperturen werden oft Arrays mit zwei Kugelschalen (schallharte Kugel innen und offene Kugel außen) verwendet (vgl. Balmages und Rafaely, 2007; Parthy et al., 2009; Jin et al., 2014). Diese haben allerdings zum wesentlichen Nachteil, dass im Vergleich zu einem einfachen Array die doppelte Anzahl an Mikrofonen benötigt wird. Das Problem der numerischen Instabilität wird in Abschnitt 6.5 aufgegriffen. Es wird ein neuer Ansatz zur Optimierung der räumlichen Abtastung vorgestellt, mit dem sich der Interpolationsfehler innerhalb der Kugel mit möglichst wenigen Mikrofonen minimieren lässt.

Begrenzung der Verstärkung der modalen Amplituden. Bei höheren harmonischen Ordnungen und niedrigen kr sind die modalen Amplituden sehr klein (siehe Abb. 6.1). Um die verschiedenen Moden nutzen zu können, müssen ihre Beiträge teils erheblich verstärkt werden (siehe Abb. 6.3). Dies kann bei kleinem SNR (wie z. B. hohem Eigenrauschen der Mikrofone) zu einer erheblichen Beeinträchtigung des Ausgangssignals führen. Daher muss die Verstärkung der modalen Amplituden begrenzt und an das verfügbare SNR angepasst werden (vgl.

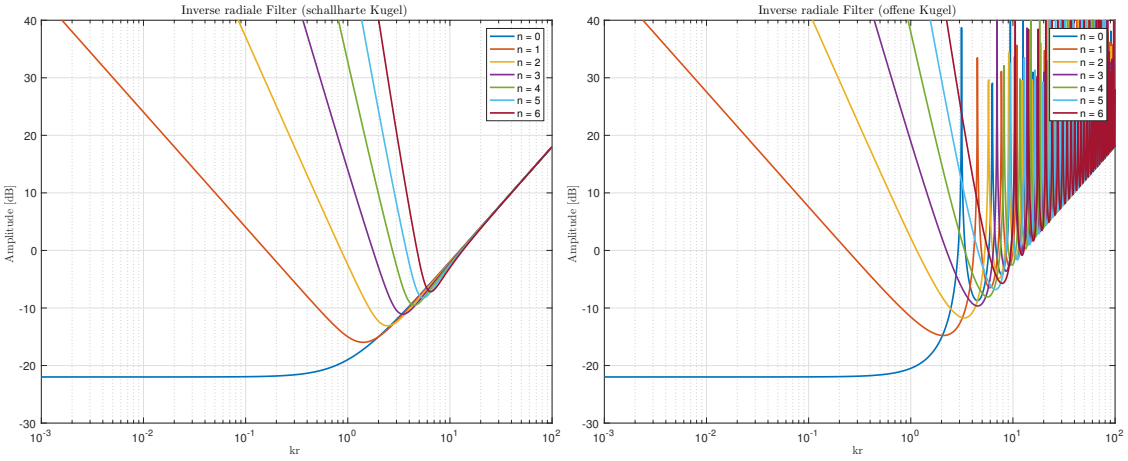


Abbildung 6.3: Inverse der modalen Amplituden $(1/b_n(kr))$ für eine schallharte Kugel (linkes Teilbild) und eine offene Kugel (rechtes Teilbild) bis zur harmonischen Ordnung $N = 6$.

Bernschütz et al., 2011). Die Filterfunktion $d_n(kr)$ lässt sich für eine vorgegebene maximal zulässige Verstärkung $\alpha = 10^{\frac{\alpha\,[\mathrm{dB}]}{20}}$ wie folgt berechnen:

$$c_n(kr) = \frac{2\alpha}{\pi} \frac{|b_n(kr)|}{b_n(kr)} \arctan\left(\frac{\pi}{2\alpha\,|b_n(kr)|}\right). \tag{6.22}$$

Werden nun die inversen modalen Amplituden auf die Fourierkoeffizienten p_{nm} angewendet

$$\tilde{p}_{nm}(k) = p_{nm}(k)\, c_n(kr_m), \tag{6.23}$$

lässt sich das Schallfeld für eine harmonische Ordnung $\tilde{N} \leq N$ wie folgt reproduzieren

$$\tilde{p}(kr) = \sum_{n=0}^{\tilde{N}} \sum_{m=-n}^{n} \tilde{p}_{nm}(k) Y_n^m(\theta, \phi). \tag{6.24}$$

Die Begrenzung der modalen Amplituden für eine maximal zulässige Verstärkung von $\alpha = 30\,\mathrm{dB}$ ist in Abb. 6.4 dargestellt. Durch den weichen Übergang (*soft knee*) wird ein Überschwingen der Arrayantwort im Ortsbereich verhindert.

Räumliches Aliasing. Ein sphärisches Mikrofonarray tastet die kontinuierliche Druckfunktion an diskreten Punkten auf der Kugeloberfläche ab. Durch die raumdiskrete Abtastung kommt es bei Moden höherer harmonischer Ordnung zu

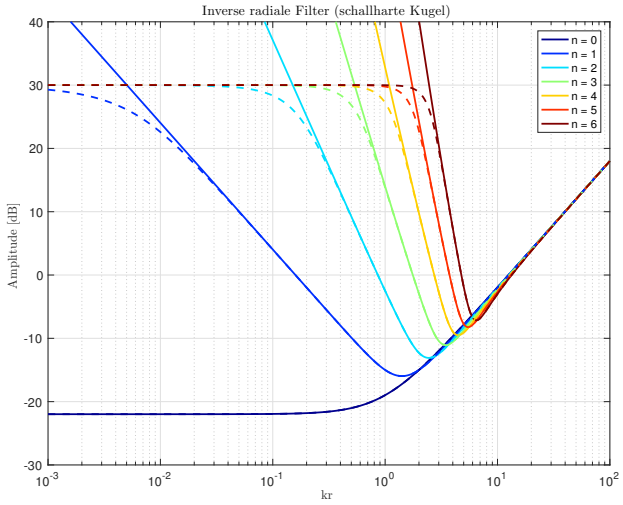


Abbildung 6.4: Beschränkung der Amplitude (*soft limiting*) der inversen modalen radialen Funktionen einer schallharten Kugel für unterschiedliche harmonische Ordnungen, mit einem Schwellwert von 30 dB (gestrichelte Linien).

räumlichem Aliasing, da diese durch die begrenzte Anzahl an Abtastpunkten nicht fehlerfrei abgebildet werden können (vgl. Li und Duraiswami, 2007; Rafaely et al., 2007; Meyer und Elko, 2008; Zotkin et al., 2008). Ist für ein sphärisches Array mit Radius r_m die Bedingung

$$N > kr_m \tag{6.25}$$

erfüllt, werden die Aliasfehler generell als vernachlässigbar angesehen (siehe z. B. Williams, 1999, Kap. 6). Zudem existiert eine Vielzahl mathematischer Lösungsansätze zur optimalen Verteilung der Abtastpunkte. Diese haben zum Ziel, das Integral über $\mathbb{S}^2$ mit einer diskreten Anzahl an Stützstellen bestmöglich numerisch anzunähern.[130] Dieses Problem wird in der Mathematik als Quadratur bezeichnet.

Das folgende Kapitel zeigt die Implementierung verschiedener modaler Beamformer. Bei der Herleitung der unterschiedlichen Algorithmen wird vorausgesetzt, dass die Fourierkoeffizienten das Schallfeld fehlerfrei abbilden. In den Abschnit-

[130] Siehe zum Beispiel Driscoll und Healy (1994), Sloan und Womersley (2004), Rakhmanov et al. (1994), Saff und Kuijlaars (1997), Gorski et al. (2005) und Sneeuw (1994).

ten 6.4 und 6.5 wird die raumdiskrete Abtastung dahingehend optimiert, dass bei einer minimalen Anzahl an Abtastpunkten der Abbildungsfehler minimiert wird.

6.2 Entwurf modaler Beamformer

Die störsignaldämpfende Wirkung eines modalen Beamformers wird durch Ausrichten des Maximums der Empfindlichkeit (d. h. der Hauptkeule) auf eine bestimmte Vorzugsrichtung erreicht. Störungen und Interferenzen aus anderen Raumrichtungen werden, in Abhängigkeit von ihrer Einfallsrichtung, abgeschwächt.

Ein sphärisches Mikrofonarray (*spherical microphone array*, SMA) mit Radius r_m tastet das Schallfeld auf der Kugel mit Q Mikrofonen ab. Das Signal $y(k)$ am Ausgang des Arrays kann entweder im Raumbereich oder im Fourierbereich (d. h. über das sphärische Wellenspektrum)[131] abgeleitet werden. Ein wesentlicher Vorteil das Beamforming im Fourierbereich zu formulieren liegt in der höheren Recheneffizienz, da in der Praxis das Schallfeld meist räumlich überabgetastet wird, d. h. $Q > (N+1)^2$. Die Integration der mit einer Aperturfunktion $w(k, \theta, \phi)$ gewichteten Schalldruckverteilung $p(k, \theta, \phi)$ auf der Kugel ergibt

$$y(k) = \int_0^{2\pi} \int_0^{\pi} w(k, \theta, \phi)^* \, p(k, \theta, \phi) \sin\theta d\theta d\phi. \tag{6.26}$$

Durch Anwenden der sphärischen Fouriertransformation und mit $\Omega = (\theta, \phi)$ lässt sich Gl. (6.26) wie folgt schreiben

$$\begin{aligned} y(k) &= \int_{\mathbb{S}^2} \left(\sum_{n=0}^{\infty} \sum_{m=-n}^{n} w_{nm}^*(k) Y_n^m(\theta, \phi)^* \right) \left(\sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} p_{n'm'}(k) Y_{n'}^{m'}(\theta, \phi) \right) d\Omega \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} w_{nm}^*(k) p_{nm}(k). \end{aligned} \tag{6.27}$$

Zur Vermeidung von Aliasing werden die Moden mit $p_{nm} = 0, \forall n > N$ begrenzt. Dies führt auf folgenden Ausdruck für das Signal am Ausgang des SMA (vgl.

[131] Die sphärischer Fouriertransformation bestimmten Spektralkoeffizienten werden oft auch als sphärisches Wellenspektrum bezeichnet.

Rafaely et al., 2007; Rafaely, 2015, Kap. 5.1)

$$y(k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} w_{nm}^{*}(k) p_{nm}(k) = \mathbf{w}_{nm}^{H} \mathbf{p}_{nm}, \tag{6.28}$$

wobei die $(N+1)^2 \times 1$ Vektoren $\mathbf{w}_{nm}$ und $\mathbf{p}_{nm}$ die (sphärischen) Fourierkoeffizienten der Aperturfunktion und des Schalldrucks auf der Kugel beinhalten.

$$\mathbf{w}_{nm} = \left[\; w_{0,0}(k), \quad w_{1,-1}(k), \quad w_{1,1}(k), \quad \ldots \quad w_{N,N}(k) \; \right]^{T}, \tag{6.29}$$

$$\mathbf{p}_{nm} = \left[\; p_{0,0}(kr), \quad p_{1,-1}(kr), \quad p_{1,1}(kr), \quad \ldots \quad p_{N,N}(kr) \; \right]^{T}. \tag{6.30}$$

Die (zeitliche) Frequenz ist implizit in der Wellenzahl k enthalten. Trifft eine einzelne ebene Welle aus Richtung $\Omega_\ell = (\theta_\ell, \phi_\ell)$ auf das SMA, ergibt sich für das Ausgangssignal

$$y(k) = \sum_{n=0}^{N} b_n(kr_m) \sum_{m=-n}^{n} w_{nm}^{*}(k) Y_n^m(\theta_\ell, \phi_\ell). \tag{6.31}$$

6.2.1 Richtcharakteristik und Bewertungsmaße

Im Fall eines axisymmetrischen[132] modalen Beamformers (vgl. Meyer und Elko, 2002), können die modalen Filtergewicht $w_{nm}^{*}(k)$ wie folgt gewählt werden

$$w_{nm}^{*}(k) = \frac{d_n(k) Y_n^m(\theta_i, \phi_i)}{b_n(kr)}, \tag{6.32}$$

wobei (θ_i, ϕ_i) die Vorzugsrichtung des SMA bezeichnet. Die neuen Filtergewichte, $d_n(k)$ hängen nur von n ab. Einsetzen von Gl. (6.32) in Gl. (6.31) ergibt die Antwort

[132] In Meyer und Elko (2002) wird die Vorzugsrichtung als Symmetrieachse verwendet (s. a. Rafaely, 2015, Kap. 5.2).

des axisymmetrischen modalen Beamformers auf eine ebene Welle

$$\begin{aligned} y(k) &= \sum_{n=0}^{N} d_n(k) \sum_{m=-n}^{n} Y_n^m(\theta_i, \phi_i) Y_n^m(\theta_j, \phi_j) \\ &= \sum_{n=0}^{N} d_n(k) \frac{2n+1}{4\pi} P_n(\cos\Theta), \end{aligned} \tag{6.33}$$

wobei Θ die Winkeldifferenz zwischen (θ_i, ϕ_i) und (θ_j, ϕ_j), gegeben durch $\cos\Theta = \cos\theta_i \cos\theta_j + \cos(\phi_i - \phi_j) \sin\theta_i \sin\theta_j$, bezeichnet. Gl. (6.33) kann als axisymmetrischer modaler Beamformer mit einem *Steering Vector*

$$\mathbf{v} = \frac{1}{4\pi} \left[\; P_0(\cos\Theta), \quad 3P_1(\cos\Theta), \quad \ldots \quad (2N+1)P_N(\cos\Theta) \; \right]^T, \tag{6.34}$$

und einem Gewichtsvektor

$$\mathbf{d}(k) = \left[\; d_0(k), \quad d_1(k), \quad d_1(k), \quad d_1(k), \quad \ldots \quad d_N(k) \; \right]^T \tag{6.35}$$

aufgefasst werden (vgl. Rafaely, 2015, Kap. 5.2). Das Signal am Ausgang des Beamformers berechnet sich zu

$$y(k) = \mathbf{d}(k)^H \mathbf{v}. \tag{6.36}$$

Daraus wird ersichtlich, dass die Filtergewichte $\mathbf{d}$ ausschließlich der Keulenformung (*beamforming*) dienen. Die Ausrichtung der Hauptkeule wird über den Steering-Vektor $\mathbf{v}$ kontrolliert (*beam steering*).

Richtcharakteristik. Die Richtcharakteristik $B(k; \theta, \phi)$ eines SMA ergibt sich aus der über alle möglichen Frequenzen und alle möglichen Einfallsrichtungen $(0 \leqslant \theta \leqslant \pi, 0 \leqslant \phi < 2\pi)$ berechneten Systemantwort:

$$B(k; \theta, \phi) = \mathbf{d}(k)^H \mathbf{v}|_{k;\, \theta \in [0;\pi],\, \phi \in [0;2\pi]}. \tag{6.37}$$

Daraus ergibt sich die auf die Leistungsübertragung bezogene Richtcharakteristik

$$P(k; \theta, \phi) = |B(k; \theta, \phi)|^2 . \tag{6.38}$$

Richtwirkung. Der Richtfaktor ist ein Maß dafür, wie stark ein diffuses Schallfeld vom Mikrofonarray unterdrückt wird. Dieser wird typischerweise als logarithmische Größe, dem sogenannten Richtindex (*directivity index*, DI), angeben. Für einen modalen Beamformer mit Vorzugsrichtung (θ_i, ϕ_i) ist der DI wie folgt definiert:

$$DI = 10\log_{10} \frac{|y(k,\theta_i,\phi_i)|^2}{\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} |y(k,\theta,\phi)|^2 \sin\theta d\theta d\phi} \tag{6.39}$$

$$= 10\log_{10} \frac{\left|\sum_{n=0}^{N}\sum_{m=-n}^{n} w_{nm}(k) b_n(k,r) Y_n^m(\theta_i,\phi_i)^*\right|^2}{\frac{1}{4\pi}\sum_{n=0}^{N}\sum_{m=-n}^{n} |w_{nm}(k) b_n(kr)|^2}, \tag{6.40}$$

wobei sich für einen axisymmetrischen modalen Beamformer folgende Formulierung ergibt

$$DI = 10\log_{10} \frac{\left|\sum_{n=0}^{N} d_n(k) \frac{2n+1}{4\pi}\right|^2}{\frac{1}{4\pi}\sum_{n=0}^{N} \frac{2n+1}{4\pi} |d_n(k)|^2}. \tag{6.41}$$

Gewinn für inkohärentes Rauschen. Der Gewinn für inkohärentes Rauschen (*white noise gain*, WNG) ist ein etabliertes Maß, welches angibt, inwieweit ein Mikrofonarray räumlich dekorreliertes Rauschen unterdrückt. Der WNG ergibt sich aus dem Verhältnis des SNR am Ausgang des Arrays zum SNR an nur einem Mikrofon. Für eine aus Vorzugsrichtung eintreffende ebene Welle mit Einheitsamplitude ergibt sich, unter der Annahme der Unkorreliertheit der Störsignale, das SNR am Eingang des Arrays

$$SNR_{\text{in}} = \frac{1}{\sigma^2}. \tag{6.42}$$

Das SNR am Ausgang des Arrays wird über das Verhältnis der Leistung des Nutzsignals zur Leistung der Störungen bestimmt. Für eine ebene Welle aus Vorzugsrichtung des Arrays berechnet sich die Leistung am Ausgang zu

$$|y|^2 = \left|\mathbf{w}_{nm}^H \mathbf{v}_{nm}\right|^2, \tag{6.43}$$

wobei $\mathbf{v}_{nm}$ ganz allgemein den *Steering*-Vektor bezeichnet. Bei axisymmetrischen Beamformern wird $\mathbf{v}_{nm}$ durch $\mathbf{v}$ (Gl. 6.34) und $\mathbf{w}_{nm}$ durch $\mathbf{d}$ (Gl. 6.32) ersetzt. Für das Störsignal ergibt sich folgender Ausdruck

$$\begin{aligned} \mathrm{E}\left\{|y|\right\}^2 &= \mathbf{w}_{nm}^H \mathrm{E}\left\{\mathbf{p}_{nm}\mathbf{p}_{nm}^H\right\}\mathbf{w}_{nm} \\ &= \mathbf{w}_{nm}^H \mathbf{B}\mathrm{E}\left\{\mathbf{y}_{nm}(\theta_i,\phi_i)\mathbf{y}_{nm}^H(\theta_i,\phi_i)\right\}\mathbf{B}^H\mathbf{w}_{nm}, \end{aligned} \tag{6.44}$$

mit der $(N+1)^2 \times (N+1)^2$ Diagonalmatrix

$$\mathbf{B} = \mathrm{diag}\{\underbrace{b_0(kr)}_{n=0}, \underbrace{b_1(kr), b_1(kr), b_1(kr)}_{n=1}, ..., b_N(kr)\} \tag{6.45}$$

und dem $(N+1)^2 \times 1$ Vektor $\mathbf{y}_{nm}$ der Kugelflächenfunktionen $Y_n^m(\theta_i,\phi_i)^*$. Mit dem Vektor $\mathbf{p}$ des Schalldrucks auf der Kugel und der Annahme unkorrelierter Störsignale folgt

$$\mathrm{E}\left\{\mathbf{p}\mathbf{p}^H\right\} = \sigma^2\mathbf{I}. \tag{6.46}$$

Mit der allgemeinen Schreibweise der diskreten sphärischen Fouriertransformation, $\mathbf{p}_{nm} = \mathbf{S}\mathbf{p}$ (Gl. 6.9), ergibt sich für Gl. (6.44) folgender Ausdruck

$$\begin{aligned} \mathrm{E}\left\{|y|\right\}^2 &= \mathbf{w}_{nm}^H \mathbf{S}\mathrm{E}\left\{\mathbf{p}\mathbf{p}^H\right\}\mathbf{S}^H\mathbf{w}_{nm} \\ &= \sigma^2\mathbf{w}_{nm}^H\mathbf{S}\mathbf{S}^H\mathbf{w}_{nm} \end{aligned} \tag{6.47}$$

und demzufolge für das SNR am Ausgang des Arrays

$$SNR_{\mathrm{in}} = \frac{\left|\mathbf{w}_{nm}^H\mathbf{v}_{nm}\right|^2}{\sigma^2\mathbf{w}_{nm}^H\mathbf{S}\mathbf{S}^H\mathbf{w}_{nm}}. \tag{6.48}$$

Aus den Gleichungen (6.42) und (6.48) lässt sich folgender Ausdruck für den WNG herleiten (vgl. Rafaely, 2015, Gl. 5.34):

$$\begin{aligned} WNG = \frac{SNR_{\mathrm{out}}}{SNR_{in}} &= \frac{\left|\mathbf{w}_{nm}^H\mathbf{v}_{nm}\right|^2}{\mathbf{w}_{nm}^H\mathbf{S}\mathbf{S}^H\mathbf{w}_{nm}} \\ &= \frac{\left|\mathbf{w}_{nm}^H\mathbf{B}\mathbf{y}_{nm}(\theta_i,\phi_i)\right|^2}{\mathbf{w}_{nm}^H\mathbf{S}\mathbf{S}^H\mathbf{w}_{nm}}. \end{aligned} \tag{6.49}$$

Um die *Least Squares*-Lösung zu erhalten, ist für $\mathbf{S} = \mathbf{Y}^{\dagger}$ einzusetzen.

In den folgenden Abschnitten werden modale Beamformer für verschiedene Optimierungsansätze hergeleitet. Die Ergebnisse numerischer Simulationen werden in Abschnitt 6.3 diskutiert. Alle Algorithmen wurden auf dem in Anhang C vorgestellten 64-Kanal Mikrofonarray implementiert. Die messtechnische Evaluierung konnte im Rahmen dieser Arbeit nicht durchgeführt werden.

6.2.2 Modaler Delay-and-Sum Beamformer (DAS-BF)

Beim modalen Delay-and-Sum Beamformer (DAS-BF) werden, für eine aus Vorzugsrichtung einfallende ebene Schallwelle mit Wellenzahl k, die Filtergewichte in den Q Mikrofonkanälen so eingestellt, dass die aufgrund der unterschiedlichen Mikrofonpositionen entstehenden Laufzeiten ausgeglichen werden. Die anschließende Summation führt, bei einer bestimmten Frequenz, zu einer zeitlich kohärenten Überlagerung der Mikrofonsignale. Die aus anderen Raumrichtungen einfallenden Schallwellen sowie nicht bzw. nur schwach korrelierte Störsignale werden hingegen nicht kohärent überlagert und, im Vergleich zum Nutzsignal, geringer verstärkt.

Eine ebene Welle mit dem Wellenzahlvektor $\mathbf{k}$ erzeugt auf einer Kugel mit Radius r eine kontinuierliche Schalldruckverteilung $p(k, r, \theta, \phi)$. Daraus berechnet sich das Signal am Ausgang des SMA:

$$\begin{aligned} y(k) &= \int_0^{2\pi} \int_0^{\pi} w\left(k, \theta, \phi\right)^* p(k, r, \theta, \phi) d\theta d\phi \\ &= \int_0^{2\pi} \int_0^{\pi} w\left(k, \theta, \phi\right)^* e^{-i\mathbf{k}^T\mathbf{r}} d\theta d\phi. \end{aligned} \tag{6.50}$$

Das Skalarprodukt $\mathbf{k}^T\mathbf{r}$ ist ein Maß für die durch die Laufzeit verursachte zeitliche Verzögerung der Welle im räumlichen Abtastpunkt $\mathbf{r}$. Mit den Filtergewichten

$$w^*(k, \omega, \phi) = e^{i\mathbf{k}^T\mathbf{r}} \tag{6.51}$$

werden, bei einer bestimmten Frequenz, die Laufzeitunterschiede für eine ebene Welle aus Richtung (θ_j, ϕ_j) ausgeglichen. Diese können mit Gl. (6.6) in das

sphärische Wellenspektrum transformiert werden

$$w_{nm}^{*} = b_n^{*}(kr) Y_n^m(\theta_j, \phi_j). \tag{6.52}$$

Wird wiederum ein axisymmetrischer Beamformer angenommen, ergibt sich mit Gl. (6.32) folgender Ausdruck für die Filtergewichte

$$d_n(k) = \left| b_n(kr) \right|^2 . \tag{6.53}$$

Nach Gl. (6.36) berechnet sich das Signal am Ausgang des modalen DAS-BF zu

$$y(k) = \mathbf{d}(k)^H \mathbf{v}. \tag{6.54}$$

Abbildung 6.5 zeigt die wichtigsten Kennwerte eines modalen DAS-BF der Ordnung $N = 7$. Die Simulation wurde für das in Anhang C vorgestellte 64-Kanal SMA (schallharte, geschlossene Kugel mit einem Radius von $r = 4{,}2\,\mathrm{cm}$ und kritischer räumlicher Abtastung, vgl. Hyperinterpolations-Grid von Sloan und Womersley 1998). Die Abbildung zeigt die Richtcharakteristik, den Richtindex (DI) sowie den Gewinn für inkohärentes Rauschen (WNG).

6.2.3 Modaler Beamformer mit maximalem Richtindex (MaxDI-BF)

Der Richtindex (Gl. 6.39) beschreibt das Verhältnis der Arrayantwort für eine aus Vorzugsrichtung einfallende ebene Welle zur gemittelten Arrayantwort für aus allen Richtungen einfallende Schallwellen. Der modale Beamformer mit maximalem Richtindex (MaxDI-BF) minimiert die Varianz des Signals am Systemausgang, unter der Nebenbedingung unverzerrter Wiedergabe aus Vorzugsrichtung. Dieser Ansatz führt auf ein Optimierungsproblem mit Nebenbedingung

$$\underset{\mathbf{w}_{nm}}{\text{minimize}}\ \mathbf{w}_{nm}^H \mathbf{B} \mathbf{w}_{nm} \quad \text{mit} \quad \mathbf{w}_{nm}^H \mathbf{v}_{nm} = 1, \tag{6.55}$$

welches sich über das Verfahren der Lagrange-Multiplikatoren lösen lässt:

$$\underset{\mathbf{w}_{nm}}{\text{minimize}}\, \mathbf{w}_{nm}^H \mathbf{B} \mathbf{w}_{nm} + \lambda \left(\mathbf{w}_{nm}^H \mathbf{v}_{nm} - 1 \right) + \lambda^* \left(\mathbf{v}_{nm}^H \mathbf{w}_{nm} - 1 \right). \tag{6.56}$$

Ableiten der Gl. (6.56) nach $\mathbf{w}_{nm}$ und anschließendes Nullsetzen

$$\mathbf{w}_{nm}^H \mathbf{B} + \lambda \mathbf{v}_{nm}^H = 0$$

ergibt folgenden Ausdruck für die optimalen Filterkoeffizienten

$$\mathbf{w}_{nm}^H = \lambda \mathbf{v}_{nm}^H \mathbf{B}^H \left(\mathbf{B}\mathbf{B}^H \right)^{-1}. \tag{6.57}$$

Gleichung (6.57) ist nur dann gültig, wenn alle Diagonalelemente der Matrix $\mathbf{B}$ ungleich Null sind. Durch Multiplizieren mit $\mathbf{v}_{nm}$ von rechts und Einsetzen der Nebenbedingung lassen sich die Lagrange-Multiplikatoren wie folgt berechnen:

$$\lambda = -\frac{1}{\mathbf{v}_{nm}^H \mathbf{B}^H \left(\mathbf{B}\mathbf{B}^H \right)^{-1} \mathbf{v}_{nm}}. \tag{6.58}$$

Substituieren von Gl. (6.58) in Gl. (6.57) führt auf die optimalen Gewichte des modalen MaxDI-BF:

$$\mathbf{w}_{nm}^H = \frac{\mathbf{v}_{nm}^H \mathbf{B}^H \left(\mathbf{B}\mathbf{B}^H \right)^{-1}}{\mathbf{v}_{nm}^H \mathbf{B}^H \left(\mathbf{B}\mathbf{B}^H \right)^{-1} \mathbf{v}_{nm}}. \tag{6.59}$$

Die spektralen Gewichte lassen sich schreiben als:

$$\begin{aligned} w_{nm}^* &= \frac{b_n(kr)^* Y_n^m(\theta_i, \phi_i)}{\left| b_n(kr) \right|^2} \frac{1}{\mathbf{y}_{nm}^H \mathbf{y}_{nm}} \\ &= \frac{4\pi}{(N+1)^2} \frac{Y_n^m(\theta_i, \phi_i)}{b_n(kr)} \end{aligned} \tag{6.60}$$

Wird wiederum eine axisymmetrische Keulenformung angenommen, ergibt sich mit Gl. (6.32) folgender einfache Ausdruck für die Gewichte des modalen MaxDI-BF:

$$d_n = \frac{4\pi}{(N+1)^2}. \tag{6.61}$$

Das Ergebnis stimmt mit Rafaely (2015, Gl. 6.10) überein.

Abbildung 6.6 zeigt die wichtigsten Kennwerte eines modalen MaxDI-BF der Ordnung $N = 7$. Die Simulation wurde für das in Anhang C vorgestellte 64-Kanal SMA (schallharte, geschlossene Kugel mit einem Radius von $r = 4{,}2\,\mathrm{cm}$ und kritischer räumlicher Abtastung, vgl. Hyperinterpolations-Grid von Sloan und Womersley 1998) durchgeführt. Die Abbildung zeigt die Richtcharakteristik, den Richtindex (DI) sowie den Gewinn für inkohärentes Rauschen (WNG).

6.2.4 Modaler Beamformer mit maximalem Gewinn für inkohärentes Rauschen (MaxWNG-BF)

Der Gewinn für inkohärentes Rauschen (*white noise gain*, WNG) gibt an, inwieweit ein sphärisches Mikrofonarray räumlich dekorrelierte Störungen unterdrückt und ist ein etabliertes Maß für die Robustheit eines Arrays (vgl. Abschnitt 6.2.1).

Der modale Beamformer mit maximalem Gewinn für inkohärentes Rauschen (MaxWNG-BF) maximiert den WNG (Gl. 6.49), unter der Nebenbedingung unverzerrter Wiedergabe aus Vorzugsrichtung. Der Ansatz entspricht dem des MaxDI-BF (Gl. 6.55), wobei die Matrix $\mathbf{B}$ durch $\mathbf{S}\mathbf{S}^H$ ersetzt wird. Daraus folgt (vgl. Rafaely, 2015, Gl. 6.22):

$$\mathbf{w}_{nm}^H = \frac{\mathbf{v}_{nm}^H \left(\mathbf{S}\mathbf{S}^H\right)^{-1}}{\mathbf{v}_{nm}^H \left(\mathbf{S}\mathbf{S}^H\right)^{-1} \mathbf{v}_{nm}}. \tag{6.62}$$

Durch Einsetzen der optimalen Gewichte in Gl. (6.49) und mit der Annahme, dass $\mathbf{w}_{nm}^H \mathbf{v}_{nm} = 1$, berechnet sich der maximale WNG zu

$$WNG_{\max} = \frac{1}{\mathbf{w}_{nm}^H \mathbf{B} \mathbf{w}_{nm}}. \tag{6.63}$$

Rafaely (2015, Kap. 6.2) zeigt für den Sonderfall, dass der Schalldruck auf der Kugeloberfläche mit Q Mikrofonen gleichförmig abgetastet wird (wie z. B. der Hyperinterpolation nach Sloan und Womersley), für $\mathbf{B}$ folgende Beziehung:

$$\mathbf{B} = \mathbf{S}\mathbf{S}^H = \frac{4\pi}{Q}\mathbf{I}. \tag{6.64}$$

Daraus ergibt sich der sehr einfach zu implementierende Ausdruck zur Berechnung der Gewichte des modalen MaxWNG-BF

$$w_{nm}^{*} = \frac{Y_n^m(\theta_i, \phi_i) b_n(kr)^*}{\sum_{n=0}^{N} \frac{2n+1}{4\pi} \left|b_n(kr)\right|^2}, \tag{6.65}$$

wobei für den maximalen WNG gilt:

$$WNG_{\max} = \frac{Q}{4\pi} \sum_{n=0}^{N} \frac{2n+1}{4\pi} \left|b_n(kr)\right|^2. \tag{6.66}$$

Unter der Voraussetzung einer axisymmetrischen Keulenformung, ergibt sich mit Gl. (6.32) folgender einfacher Ausdruck für die Gewichte des modalen MaxWNG-BF

$$d_n = \frac{\left|b_n(kr)\right|^2}{\sum_{n=0}^{N} \frac{2n+1}{4\pi} \left|b_n(kr)\right|^2}. \tag{6.67}$$

Abbildung 6.7 zeigt die wichtigsten Kennwerte eines modalen MaxWNG-BF der Ordnung $N = 7$. Die Simulation wurde für das in Anhang C vorgestellte 64-Kanal SMA (schallharte, geschlossene Kugel mit einem Radius von $r = 4{,}2\,\mathrm{cm}$ und kritischer räumlicher Abtastung, vgl. Hyperinterpolations-Grid von Sloan und Womersley 1998) durchgeführt. Die Abbildung zeigt die Richtcharakteristik, den Richtindex (DI) sowie den Gewinn für inkohärentes Rauschen (WNG).

6.2.5 Modaler Beamformer mit maximalem Vorne/Hinten-Verhältnis (MaxFBR-BF)

Der modale Beamformer mit maximalem Vorne/Hinten-Verhältnis (*maximum front/back ratio beamformer*, MaxFBR-BF) maximiert das Verhältnis der Amplitude der Richtcharakteristik des Arrays in Vorzugsrichtung zur Amplitude der Richtcharakteristik in entgegengesetzter Richtung. Zur Vereinfachung der Herleitung wird, ohne Beschränkung der Allgemeinheit, angenommen, dass die Vorzugsrichtung auf den Nordpol ausgerichtet ist. Der Ansatz kann wie folgt

formuliert werden.

$$FBR = \frac{\int_0^{2\pi}\int_0^{\pi/2} |y(\theta,\phi)|^2 \sin\theta d\theta d\phi}{\int_0^{2\pi}\int_{\pi/2}^{\pi} |y(\theta,\phi)|^2 \sin\theta d\theta d\phi}. \tag{6.68}$$

Unter der Voraussetzung einer axisymmetrischen Keulenformung, lässt sich der Zähler des Bruchs in Gl. (6.68) wie folgt umformen (vgl. Rafaely, 2015, Gl. 6.47):

$$\begin{aligned} FBR_{\text{front}} &= \int_0^{2\pi}\int_0^{\pi/2} \sum_{n=0}^{N}\sum_{n'=0}^{N} d_n^* \frac{2n+1}{4\pi} P_n(\cos\theta) d_{n'} \frac{2n'+1}{4\pi} P_n'(\cos\theta)\sin\theta d\theta d\phi \\ &= \frac{1}{8\pi}\sum_{n=0}^{N}\sum_{n'=0}^{N} d_n^*(2n+1) d_{n'}(2n'+1) \int_0^{\pi/2} P_n(\cos\theta) P_n'(\cos\theta)\sin\theta d\theta. \end{aligned} \tag{6.69}$$

Das Integral lässt sich mit $P_n(z) = \sum_{j=0}^{n} p_j^n z^j$ wie folgt berechnen

$$\begin{aligned} \int_0^{\pi/2} P_n(\cos\theta) P_n'(\cos\theta)\sin\theta d\theta &= \int_0^1 P_n(z) P_n'(z) dz \\ &= \sum_{i=0}^{n}\sum_{j=0}^{n'} p_i^n p_j^{n'} \int_0^1 z^{j+i} dz \\ &= \sum_{i=0}^{n}\sum_{j=0}^{n'} \frac{p_i^n p_j^{n'}}{j+i+1}. \end{aligned} \tag{6.70}$$

Mit Gl. (6.70) lässt sich FBR_{front} in Vektor-Matrix-Schreibweise ausdrücken:

$$FBR_{\text{front}} = \mathbf{d}_n^H \mathbf{U} \mathbf{d}_n. \tag{6.71}$$

$\mathbf{U}$ ist eine reellwertige $(N+1)\times(N+1)$ Matrix mit den Elementen

$$U_{n,n'} = \frac{1}{8\pi}(2n+1)(2n'+1)\sum_{i=0}^{n}\sum_{j=0}^{n'} \frac{p_i^n p_j^{n'}}{j+i+1}. \tag{6.72}$$

Für den Nenner (Gl. 6.68) ergibt sich, bis auf das Integral (Gl. 6.70), derselbe Ausdruck wie für den Zähler. Dieses lässt sich wie folgt berechnen:

$$\int_{\pi/2}^{\pi} P_n(\cos\theta) P_{n'}(\cos\theta) \sin\theta d\theta = \int_{-1}^{0} P_n(z) P_{n'}(z) dz = \sum_{i=0}^{n} \sum_{j=0}^{n'} p_i^n p_j^{n'} \int_{=1}^{0} z^{j+i} dz$$
$$= \sum_{i=0}^{n} \sum_{j=0}^{n'} (-1)^{j+i} \frac{p_i^n p_j^{n'}}{j+i+1}, \tag{6.73}$$

wodurch sich FBR_{back} wiederum in Vektor-Matrix-Schreibweise ausdrücken lässt:

$$FBR_{\text{back}} = \mathbf{d}_n^H \mathbf{V} \mathbf{d}_n. \tag{6.74}$$

$\mathbf{B}$ ist eine reellwertige $(N+1) \times (N+1)$ Matrix mit den Elementen

$$V_{n,n'} = \frac{1}{8\pi}(2n+1)(2n'+1) \sum_{i=0}^{n} \sum_{j=0}^{n'} (-1)^{i+j} \frac{p_i^n p_j^{n'}}{j+i+1}. \tag{6.75}$$

Einsetzen der Gleichungen (6.71) und (6.74) in Gl. (6.68) führt auf das Vorne/Hinten-Verhältnis in Vektor-Matrix-Schreibweise

$$FBR = \frac{\mathbf{d}_n^H \mathbf{U} \mathbf{d}_n}{\mathbf{d}_n^H \mathbf{V} \mathbf{d}_n}. \tag{6.76}$$

Gleichung (6.76) kann auch als verallgemeinertes Eigenwertproblem

$$\mathbf{U}\mathbf{d}_n = \lambda \mathbf{V} \mathbf{d}_n \tag{6.77}$$

aufgefasst werden. Der Eigenvektor $\mathbf{d}_n$ zum größten Eigenwert maximiert das FBR, und ergibt die optimalen Gewichte zur Implementierung des MaxFBR-BF.

Abbildung 6.8 zeigt die wichtigsten Kennwerte eines modalen MaxWNG-BF der Ordnung $N = 7$. Die Simulation wurde für das in Anhang C vorgestellte 64-Kanal SMA (schallharte, geschlossene Kugel mit einem Radius von $r = 4{,}2\,\text{cm}$ und kritischer räumlicher Abtastung, vgl. Hyperinterpolations-Grid von Sloan und

Womersley 1998) durchgeführt. Die Abbildung zeigt die Richtcharakteristik, den Richtindex (DI) sowie den Gewinn für inkohärentes Rauschen (WNG).

6.2.6 Modaler Dolph-Chebychev Beamformer (DC-BF)

Mit dem modalen Dolph-Chebychev Beamformer (DC-BF) lässt sich das Verhältnis von Hauptkeulenbreite zu Nebenkeulendämpfung steuern (vgl. Koretz und Rafaely 2009; Rafaely 2015, Kap. 6.6). Dabei wird die Richtcharakteristik des Arrays über ein Chebyshev-Polynom erster Art bestimmt (vgl. Van Trees, 2002, Kap. 3.4.2):

$$y(\theta) = \frac{1}{R} T_\ell \left(x_0 \cos\left(\theta/2\right)\right), \tag{6.78}$$

wobei $\theta \in [-\pi, \pi]$ die Einfallsrichtung bezeichnet. x_0 und R definieren die Breite der Hauptkeule und Dämpfung der Nebenkeulen.

Die Chebyshev-Polynome erster Art, $T_\ell(x)$, gehören zur Klasse der Orthogonalpolynome und sind für alle $x \in \mathbb{R}$ und $\ell \in \mathbb{N}_0$ definiert durch

$$T_\ell(x) = \frac{1}{2}\left(\left(x + \sqrt{x^2-1}\right)^\ell + \left(x - \sqrt{x^2-1}\right)^\ell\right). \tag{6.79}$$

Es gilt (vgl. Van Trees, 2002, Gl. 3.134)

$$T_\ell(x) = \begin{cases} \cos(\ell \cos^{-1} x), & \text{für } |x| \leq 1, \\ \cosh(\ell \cosh^{-1} x), & \text{für } x > 1, \\ (-1)^\ell \cosh(\ell \cosh^{-1} |x|), & \text{für } x < -1. \end{cases} \tag{6.80}$$

Beim Entwurf eines modalen DC-BF wird entweder zuerst die Hauptkeulenbreite oder die Nebenkeulendämpfung festgelegt und der jeweils andere Parameter daraus berechnet (vgl. Van Trees, 2002, Kap. 3.4.2). Wird R festgelegt, ergibt sich daraus

$$x_0 = \cosh\left(\frac{\cosh^{-1}(R)}{\ell}\right) \tag{6.81}$$

und in weiterer Folge

$$\alpha_0 = 2\cos^{-1}\left(\frac{\cos\left(\frac{\pi}{2\ell}\right)}{x_0}\right). \tag{6.82}$$

α_0 entspricht der in den Nulldurchgängen gemessenen Breite der Hauptkeule. Wird hingegen zuerst α_0 festgelegt, ergibt sich daraus

$$x_0 = \frac{\cos\left(\frac{\pi}{2\ell}\right)}{\cos\left(\frac{\alpha_0}{2}\right)} \tag{6.83}$$

und in weiterer Folgt die Dämpfung der Nebenkeulen, welche durch das Haupt/Nebenkeulen-Verhältnis

$$R = \cosh\left(\ell \cosh^{-1}(x_0)\right) \tag{6.84}$$

bestimmt wird.

Sobald die Richtcharakteristik des modalen DC-BF bestimmt ist, können die zur Implementierung benötigten Gewichte d_n hergeleitet werden. Aus dem Vergleich der Gleichungen (6.33) und (6.78) ergibt sich

$$\frac{1}{R}\, T_\ell\left(x_0 \cos\left(\theta/2\right)\right) = \sum_{n=0}^{N} d_n(k) \frac{2n+1}{4\pi} P_n(\cos\theta). \tag{6.85}$$

Durch Substituieren von $z = \cos\theta$, $\ell = 2N$ und $\cos(\theta/2) = \sqrt{(1+\cos\theta)/2}$ in Gl. (6.85) folgt

$$\frac{1}{R}\, T_{2N}\left(x_0 \sqrt{\frac{1+z}{2}}\right) = \sum_{n=0}^{N} d_n(k) \frac{2n+1}{4\pi} P_n(z). \tag{6.86}$$

Nun werden beide Seiten der Gl. (6.86) mit $2\pi P_m(z)$ multipliziert und anschließend über $z \in [-1{,}1]$ integriert. Aufgrund der Orthonormalität der Legendre-Polynome ergibt sich für d_m folgender Ausdruck

$$d_m = \frac{2\pi}{R} \int_{-1}^{1} P_m(z)\, T_{2N}\left(x_0 \sqrt{\frac{1+z}{2}}\right) dz. \tag{6.87}$$

Mit folgenden Definitionen[133]

$$P_m(z) = \sum_{j=0}^{m} p_j^m z^j, \tag{6.88}$$

$$T_{2N}(z) = \sum_{l=0}^{N} t_{2l}^{2N} z^{2l}, \tag{6.89}$$

lässt sich Gl. (6.87) wie folgt umschreiben

$$d_m = \frac{2\pi}{R} \sum_{j=0}^{m} \sum_{l=0}^{N} p_j^m t_{2l}^{2N} \int_{-1}^{1} z^j \frac{(1+z)^l}{2^l} dz. \tag{6.90}$$

Für ein Chebychev-Polynom der Ordnung $2N$ müssen somit nur $N+1$ Elemente summiert werden.

Über die binomische Erweiterung

$$(1+z)^l = \sum_{k=0}^{l} \frac{l!}{k!(l-k)!} z^k \tag{6.91}$$

ergibt sich für Gl. (6.87) folgender Ausdruck

$$\begin{aligned} d_m &= \frac{2\pi}{R} \sum_{j=0}^{m} \sum_{l=0}^{N} \sum_{k=0}^{l} p_j^m \, t_{2l}^{2N} \, 2^{-l} \frac{l!}{k!(l-k)!} \int_{-1}^{1} z^{l+j} dz \\ &= \frac{2\pi}{R} \sum_{j=0}^{m} \sum_{l=0}^{N} \sum_{k=0}^{l} p_j^m \, t_{2l}^{2N} \, 2^{-l} \frac{l!}{k!(l-k)!} \frac{1-(-1)^{l+j+1}}{l+j+1} \end{aligned} \tag{6.92}$$

mit dem sich d_m für alle $0 \leq m \leq N$ berechnen lässt.

Gl. (6.92) kann auch in Vektor-Matrix-Schreibweise formuliert werden (vgl. Rafaely, 2015, Gl. 6.65):

$$\mathbf{d}_n = \frac{2\pi}{R} \mathbf{PACT} \mathbf{x}_0, \tag{6.93}$$

[133] Die hochgestellten Indizes (m und N) bezeichnen die Ordnung des Polynoms und die tiefgestellten Indizes (j und l) den Koeffizientennumber.

wobei

$$\mathbf{x}_0 = \left[\; 1, \;\; x_0^2, \;\; x_0^4, \;\; \ldots, \;\; x_0^{2N} \; \right]^T, \tag{6.94}$$

$$\mathbf{P} = \begin{pmatrix} p_0^0 & 0 & \ldots & 0 \\ p_0^1 & p_1^1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ p_0^N & p_1^N & \ldots & p_N^N \end{pmatrix}, \tag{6.95}$$

$$\mathbf{A} = \begin{pmatrix} 2 & 0 & \ldots & \frac{1-(-1)^{N+1}}{N+1} \\ 0 & \frac{2}{3} & \ldots & \frac{1-(-1)^{N+2}}{N+2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1-(-1)^{N+1}}{N+1} & \frac{1-(-1)^{N+2}}{N+2} & \ldots & \frac{1-(-1)^{2N+1}}{2N+1} \end{pmatrix}, \tag{6.96}$$

$$\mathbf{C} = \begin{pmatrix} 1 & \frac{1}{2} & \ldots & \frac{1}{2^N} \\ 0 & \frac{1}{2} & \ldots & \frac{N}{2^N} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \frac{1}{2^N} \end{pmatrix}, \tag{6.97}$$

$$\mathbf{T} = \operatorname{diag}\left(\; t_0^{2N}, \;\; t_2^{2N}, \;\; \ldots, \;\; t_{2N}^{2N} \; \right). \tag{6.98}$$

Abbildung 6.9 zeigt die wichtigsten Kennwerte eines modalen DC-BF der Ordnung $N = 7$. Die Simulation wurde für das in Anhang C vorgestellte 64-Kanal SMA (schallharte, geschlossene Kugel mit einem Radius von $r = 4{,}2\,\mathrm{cm}$ und kritischer räumlicher Abtastung, vgl. Hyperinterpolations-Grid von Sloan und Womersley 1998) durchgeführt. Die Abbildung zeigt die Richtcharakteristik, den Richtindex (DI) sowie den Gewinn für inkohärentes Rauschen (WNG).

6.2.7 Modale Beamformer mit Standard-Richtcharakteristik

Mit modalen Beamformern kann die Richtcharakteristik von typischen Mikrofonen nachgebildet werden. Dies macht vor allem dann Sinn, wenn die modalen Beamformer für musikalische Aufnahmen verwendet werden. Zudem lassen sich bei der raumakustischen Analyse, neben den 3D Reflektogrammen, auch die standardisierten Bewertungsmaße (wie z. B. die Klarheit und die Durchsichtigkeit) gemessen

werden (vgl. ISO, 2008, 2009a,b). Die Messung der raumakustischen Bewertungsmaße erfordert Mikrofone mit Kugel-, Achter-, Nieren- und Hypernierencharakteristik.

In den folgenden Betrachtungen sind die Koeffizienten $d_n(k)$ unabhängig von der Frequenz. Aus diesem Grund wird, ohne Beschränkung der Allgemeinheit, zur verkürzten Schreibweise die Wellenzahl k weggelassen und $d_n(k)$ durch d_n ersetzt.

A Kugelcharakteristik

Mikrofone mit Kugelcharakteristik sind für Schall aus allen Richtungen gleich empfindlich. Der modale Beamformer simuliert ein ungerichtetes Mikrofon im Zentrum der Kugel. Für d_n gilt:

$$d_n = \begin{cases} 1, & \text{if } n = 0, \\ 0, & \text{if } n \neq 0. \end{cases} \tag{6.99}$$

B Achtercharakteristik (Dipol)

Mikrofone mit Achtercharakteristik sind für frontal und rückwertig einfallenden Schall gleichermaßen empfindlich. Am unempfindlichsten sind sie gegenüber seitlich eintreffenden Schall. Der modale Beamformer simuliert einen Dipol im Zentrum der Kugel. Für d_n gilt:

$$d_n = \begin{cases} 1, & \text{if } n = 1, \\ 0, & \text{if } n \neq 1. \end{cases} \tag{6.100}$$

Abbildung 6.10 zeigt die wichtigsten Kennwerte eines modalen BF der Ordnung $N = 7$ mit Achtercharakteristik. Die Simulation wurde für das in Anhang C vorgestellte 64-Kanal SMA (schallharte, geschlossene Kugel mit einem Radius von $r = 4{,}2\,\text{cm}$ und kritischer räumlicher Abtastung, vgl. Hyperinterpolations-Grid von Sloan und Womersley 1998) durchgeführt. Die Abbildung zeigt die Richtcharakteristik, den Richtindex (DI) sowie den Gewinn für inkohärentes Rauschen (WNG).

C Nierencharakteristik N-ter Ordnung

Mikrofone mit Nierencharakteristik sind am empfindlichsten für frontal eintreffenden Schall, während Seitenschall leiser erfasst und rückwärtiger Schall stark unterdrückt wird. Der modale Beamformer simuliert eine Niere der Ordnung N im Zentrum der Kugel. Für d_n gilt:

$$y(\theta) = (0.5 + 0.5\cos\theta)^N . \tag{6.101}$$

Mit Gl. (6.33) und $z = cos\theta$ ergibt sich

$$\sum_{n=0}^{N} d_n \frac{2n+1}{4\pi} P_n(z) = \frac{(1+z)^N}{2^N}. \tag{6.102}$$

Werden beide Seiten der Gleichung mit $2\pi P_n(z)$ multipliziert und anschließend über $z \in [-1, 1]$ integriert, folgt

$$d_m = \int_{-1}^{1} 2\pi P_m(z) \frac{(1+z)^N}{2^N} dz. \tag{6.103}$$

Mit den erweiterten Legendre-Polynomen (Gl. 6.88) und der binomischen Erweiterung (Gl. 6.91), lässt sich für d_n folgender Ausdruck herleiten:

$$\begin{aligned} d_m &= \int_{-1}^{1} \frac{2\pi}{2^N} \left(\sum_{j=0}^{m} p_j^m z^j \right) \left(\sum_{k=0}^{N} \frac{N!}{k!(N-k)!} z^k \right) dz \\ &= \frac{N!2\pi}{2^N} \sum_{j=0}^{m} \sum_{k=0}^{N} \frac{p_j^m}{k!(N-k)!} \int_{-1}^{1} z^{k+j} dz \\ &= \frac{N!2\pi}{2^N} \sum_{j=0}^{m} \sum_{k=0}^{N} \frac{p_j^m}{k!(N-k)!} \frac{1-(-1)^{k+j+1}}{k+j+1}. \end{aligned} \tag{6.104}$$

Daraus folgt für einen modalen Beamformer mit Nierencharakteristik der Ordnung N

$$d_n = \begin{cases} \frac{N!2\pi}{2^N} \sum_{j=0}^{m} \sum_{k=0}^{N} \frac{p_j^m}{k!(N-k)!} \frac{1-(-1)^{k+j+1}}{k+j+1}, & \text{für } n \leq N, \\ 0, & \text{für } n > N. \end{cases} \tag{6.105}$$

Die Abbildungen 6.11 bis 6.13 zeigen die wichtigsten Kennwerte eines modalen BF der Ordnung $N = 7$ mit Nierencharakteristik der 1. bis 3. Ordnung. Die Simulation wurde für das in Anhang C vorgestellte 64-Kanal SMA (schallharte, geschlossene Kugel mit einem Radius von $r = 4{,}2\,\mathrm{cm}$ und kritischer räumlicher Abtastung, vgl. Hyperinterpolations-Grid von Sloan und Womersley 1998) durchgeführt. Die Abbildung zeigt die Richtcharakteristik, den Richtindex (DI) sowie den Gewinn für inkohärentes Rauschen (WNG).

D Supernierencharakteristik N-ter Ordnung

Ein modaler Beamformer mit Nierencharakteristik N-ter Ordnung entspricht dem modalen MaxDI-BF (Gl. 6.61) der Ordnung N. Für d_n gilt:

$$d_n = \begin{cases} \frac{4\pi}{(N+1)^2}, & \text{für } n \leq N, \\ 0, & \text{für } n > N. \end{cases} \tag{6.106}$$

Die Abbildungen 6.14 bis 6.16 zeigen die wichtigsten Kennwerte eines modalen BF der Ordnung $N = 7$ mit Supernierencharakteristik der 1. bis 3. Ordnung. Die Simulation wurde für das in Anhang C vorgestellte 64-Kanal SMA (schallharte, geschlossene Kugel mit einem Radius von $r = 4{,}2\,\mathrm{cm}$ und kritischer räumlicher Abtastung, vgl. Hyperinterpolations-Grid von Sloan und Womersley 1998) durchgeführt. Die Abbildung zeigt die Richtcharakteristik, den Richtindex (DI) sowie den Gewinn für inkohärentes Rauschen (WNG).

E Hypernierencharakteristik N-ter Ordnung

Ein modaler Beamformer mit Hypernierencharakteristik N-ter Ordnung entspricht dem modalen MaxFBR-BF (Gl. 6.2.5) der Ordnung N. Für d_n gibt es keine analytische Lösung, die Koeffizienten müssen numerisch bestimmt werden.

Die Abbildungen 6.17 bis 6.19 zeigen die wichtigsten Kennwerte eines modalen BF der Ordnung $N = 7$ mit Hypernierencharakteristik der 1. bis 3. Ordnung. Die Simulation wurde für das in Anhang C vorgestellte 64-Kanal SMA (schallharte, geschlossene Kugel mit einem Radius von $r = 4{,}2\,\mathrm{cm}$ und kritischer räumlicher

Abtastung, vgl. Hyperinterpolations-Grid von Sloan und Womersley 1998) durchgeführt. Die Abbildung zeigt die Richtcharakteristik, den Richtindex (DI) sowie den Gewinn für inkohärentes Rauschen (WNG).

6.3 Simulation modaler Beamformer

Die im vorhergehenden Abschnitt abgeleiteten modalen Beamformer wurden für das in Anhang C vorgestellte Array 7. Ordnung simuliert.

Das Array besteht aus einer schallharten, geschlossenen Kugel (Aluminium) mit einem Radius von $r = 4{,}2\,\mathrm{cm}$. Die 64 Mikrofone sind gleichförmig auf der Kugeloberfläche verteilt. Die räumliche Abtastung entspricht der kritischen Abtastung (Hyperinterpolation) nach Sloan und Womersley 1998).

Das Schallfeld wurde in der numerischen Simulation bandbegrenzt, um das Entstehen von räumlichen Aliasing zu vermeiden. Die Amplitude der inversen modalen radialen Funktionen wurden mit einem Schwellwert von 30 dB begrenzt.

Die Abbildungen (6.5) bis (6.19) zeigen die Richtcharakteristik, den Richtindex (DI) sowie den Gewinn für inkohärentes Rauschen (WNG). In den Abbildungen (6.20) und (6.21) wird der Richtindex (DI) und der Gewinn für inkohärentes Rauschen (WNG) verschiedener modaler Beamformer verglichen.

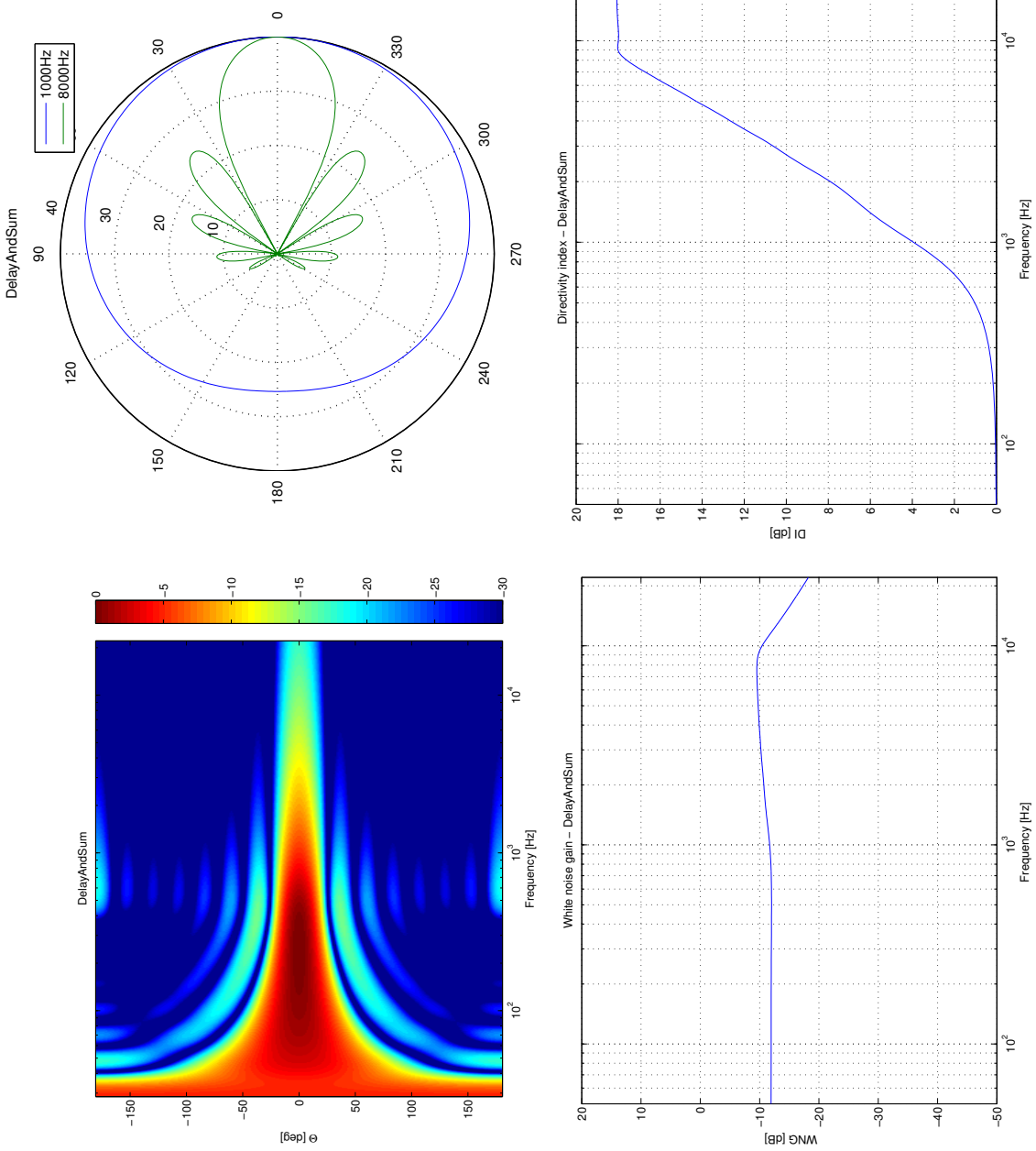


Abbildung 6.5: Kennwerte eines modalen Delay-and-Sum Beamformers (DAS-BF) der Ordnung $N = 7$. Schallhartes 64-Kanal SMA mit $r = 4{,}2\,\mathrm{cm}$, Hyperinterpolation-Abtastung (vgl. Sloan und Womersley, 1998). Richtcharakteristik (links oben), Richtplot (rechts oben), WNG (links unten) und DI (rechts unten).

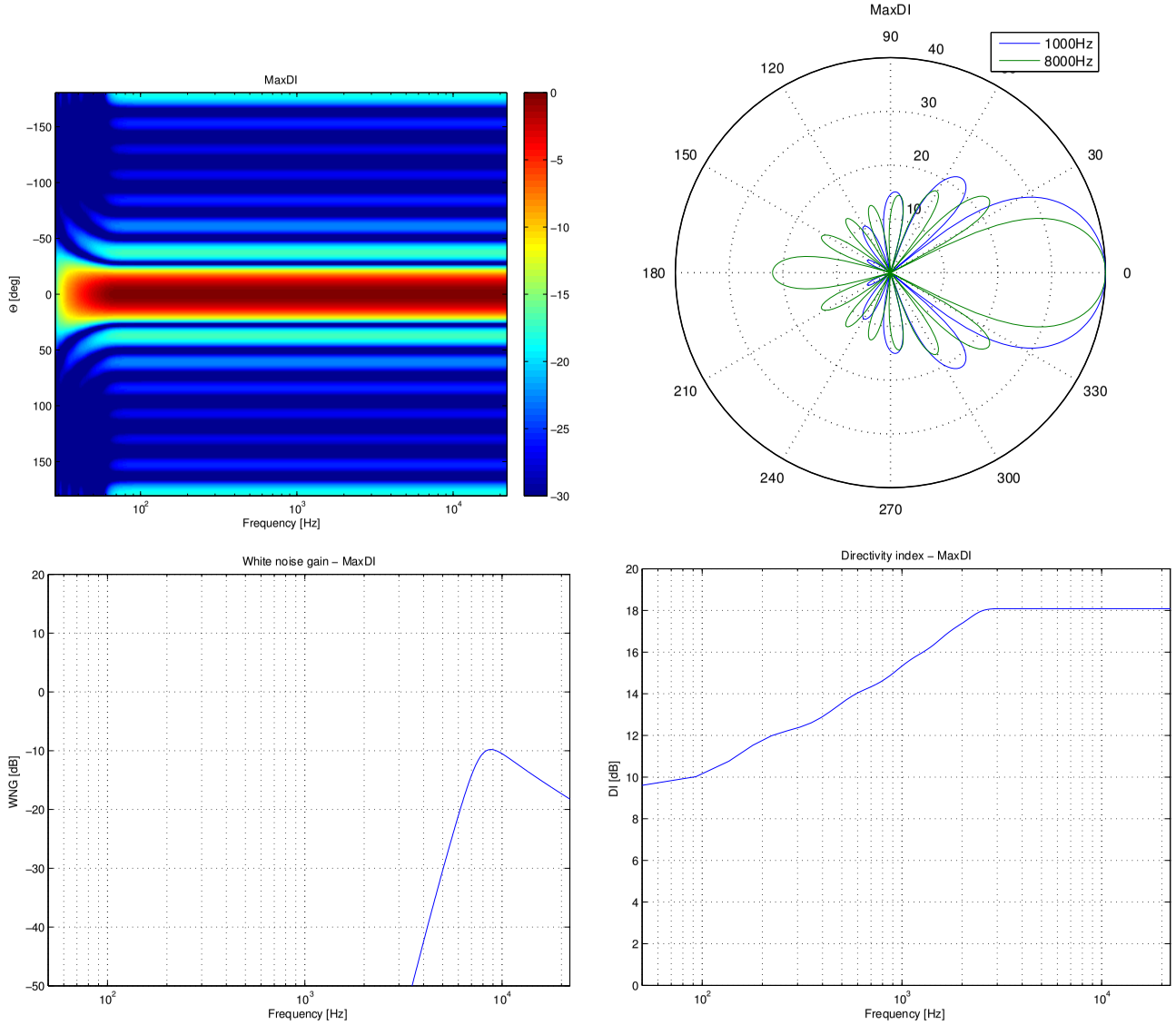


Abbildung 6.6: Kennwerte eines modalen Beamformers mit maximalem Richtindex (MaxDI) der Ordnung $N = 7$. Schallhartes 64-Kanal SMA mit $r = 4{,}2\,\mathrm{cm}$, Hyperinterpolation-Abtastung (vgl. Sloan und Womersley, 1998). Richtcharakteristik (links oben), Richtplot (rechts oben), WNG (links unten) und DI (rechts unten).

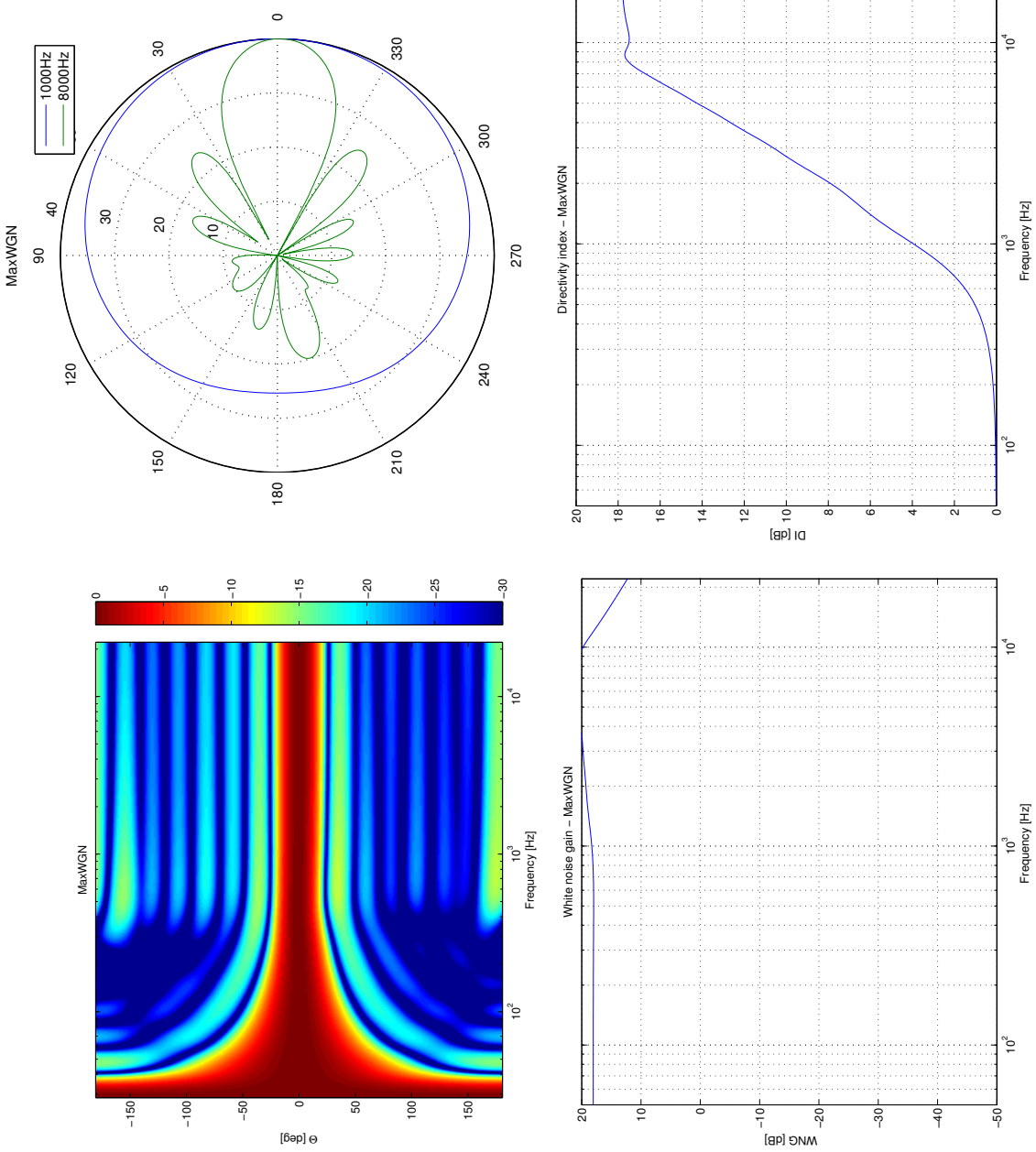


Abbildung 6.7: Kennwerte eines modalen Beamformers mit maximalem Gewinn für inkohärentes Rauschen (MaxWNG-BF) der Ordnung $N = 7$. Schallhartes 64-Kanal SMA mit $r = 4{,}2\,\mathrm{cm}$, Hyperinterpolation-Abtastung (vgl. Sloan und Womersley, 1998). Richtcharakteristik (links oben), Richtplot (rechts oben), WNG (links unten) und DI (rechts unten).

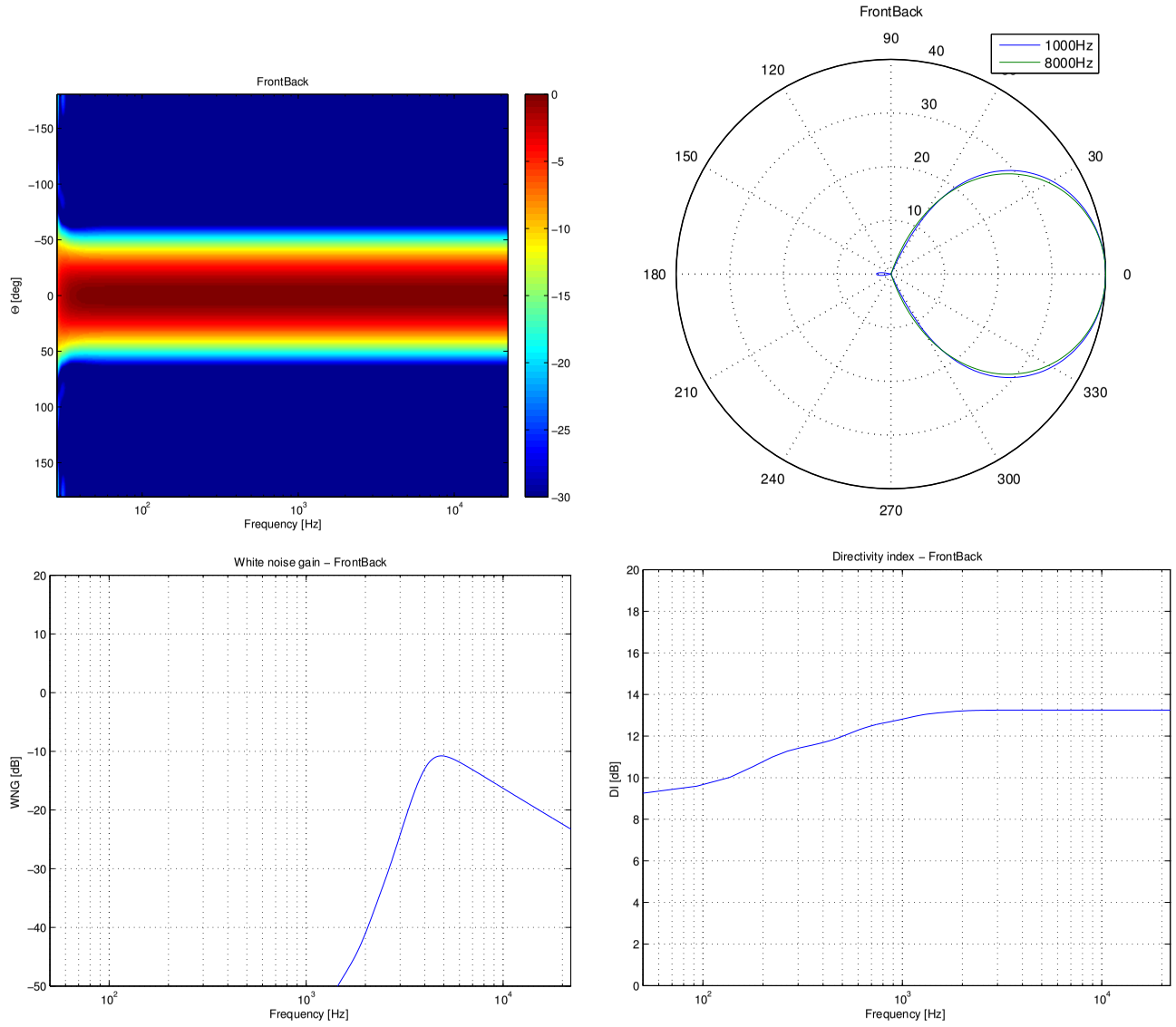


Abbildung 6.8: Kennwerte eines modalen Beamformers mit maximalem Vorne/Hinten-Verhältnis (Max-FBR) der Ordnung $N = 7$. Schallhartes 64-Kanal SMA mit $r = 4{,}2\,\mathrm{cm}$, Hyperinterpolation-Abtastung (vgl. Sloan und Womersley, 1998). Richtcharakteristik (links oben), Richtplot (rechts oben), WNG (links unten) und DI (rechts unten).

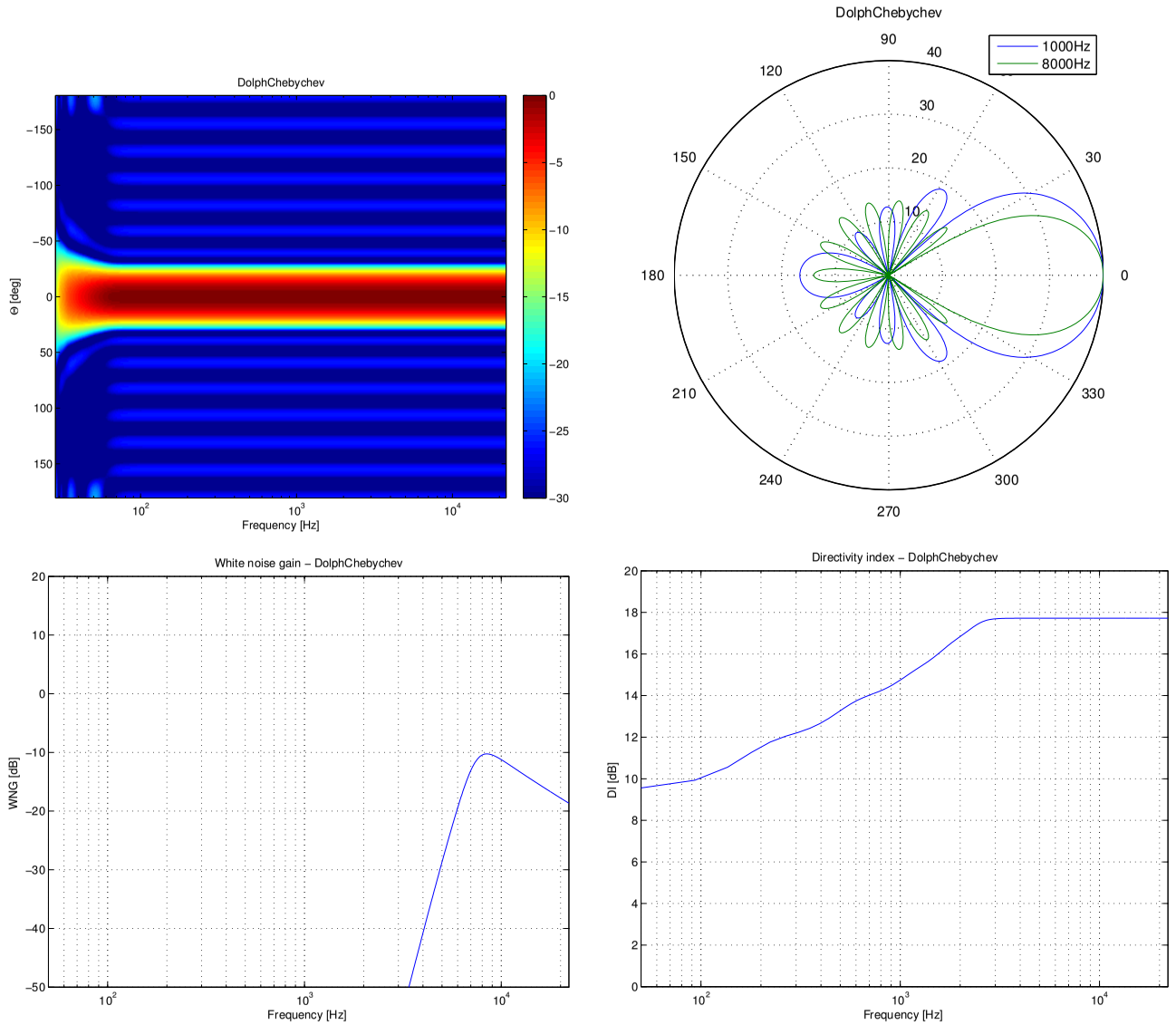


Abbildung 6.9: Kennwerte eines modalen Dolph-Chebychev Beamformers (DC-BF) der Ordnung $N = 7$ mit einer Haupt/Nebenkeulen-Dämpfung von 40 dB. Schallhartes 64-Kanal SMA mit $r = 4{,}2\,\mathrm{cm}$, Hyperinterpolation-Abtastung (vgl. Sloan und Womersley, 1998). Richtcharakteristik (links oben), Richtplot (rechts oben), WNG (links unten) und DI (rechts unten).

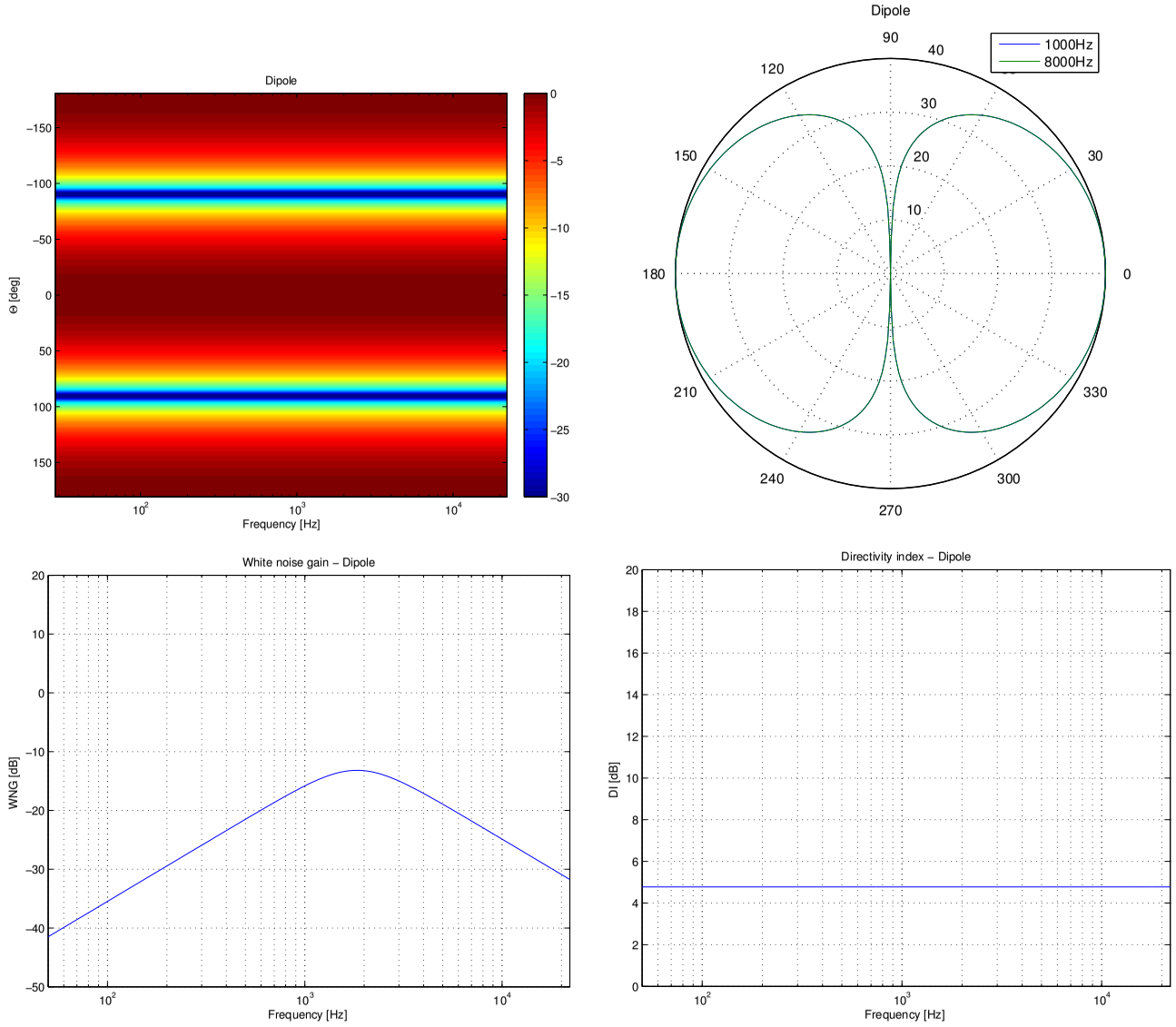


Abbildung 6.10: Kennwerte eines modalen Beamformers der Ordnung $N = 7$ mit Achtercharakteristik. Schallhartes 64-Kanal SMA mit $r = 4{,}2\,\mathrm{cm}$, Hyperinterpolation-Abtastung (vgl. Sloan und Womersley, 1998). Richtcharakteristik (links oben), Richtplot (rechts oben), WNG (links unten) und DI (rechts unten).

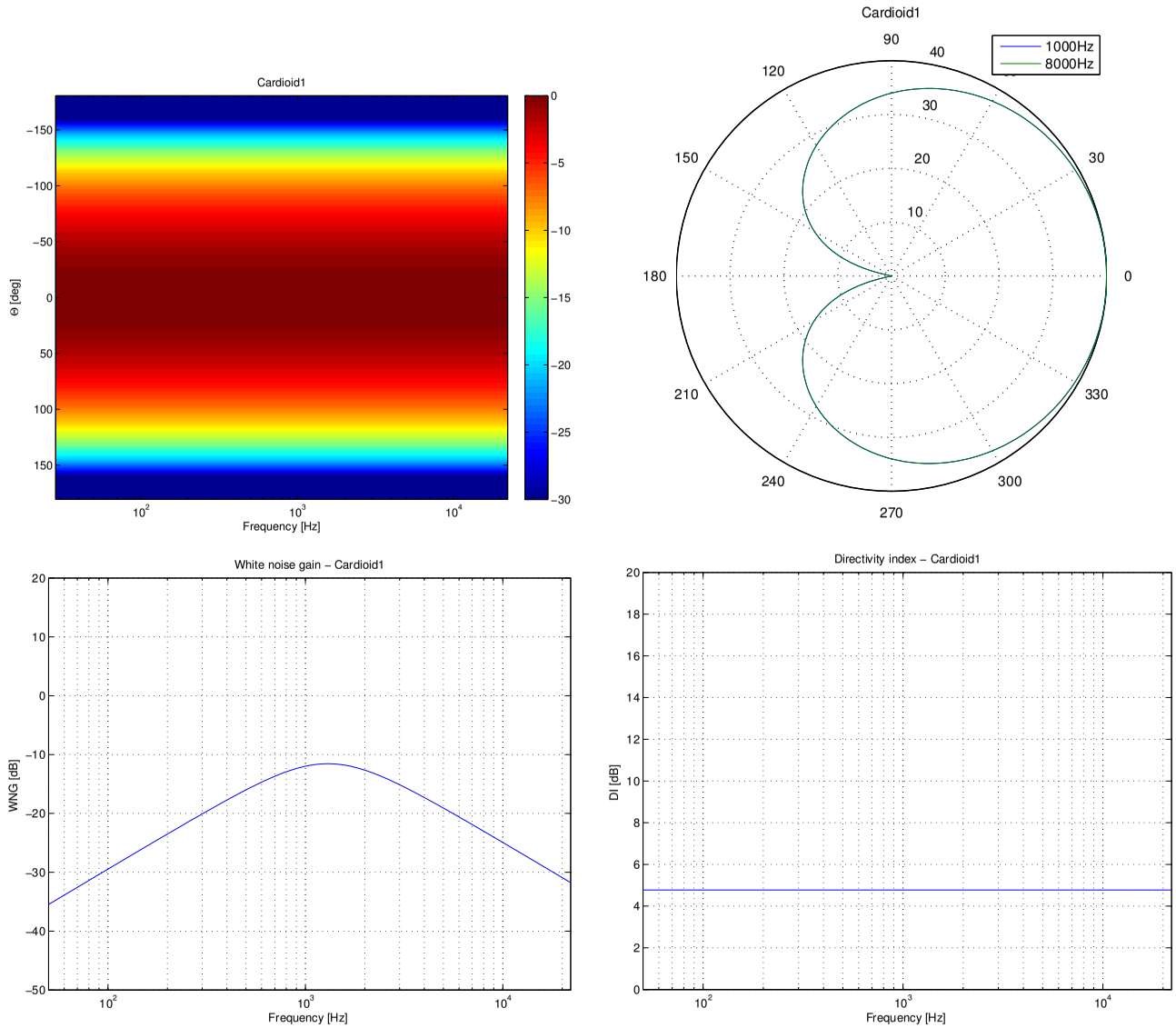


Abbildung 6.11: Kennwerte eines modalen Beamformers der Ordnung $N = 7$ mit Nierencharakteristik 1. Ordnung. Schallhartes 64-Kanal SMA mit $r = 4{,}2\,\mathrm{cm}$, Hyperinterpolation-Abtastung (vgl. Sloan und Womersley, 1998). Richtcharakteristik (links oben), Richtplot (rechts oben), WNG (links unten) und DI (rechts unten).

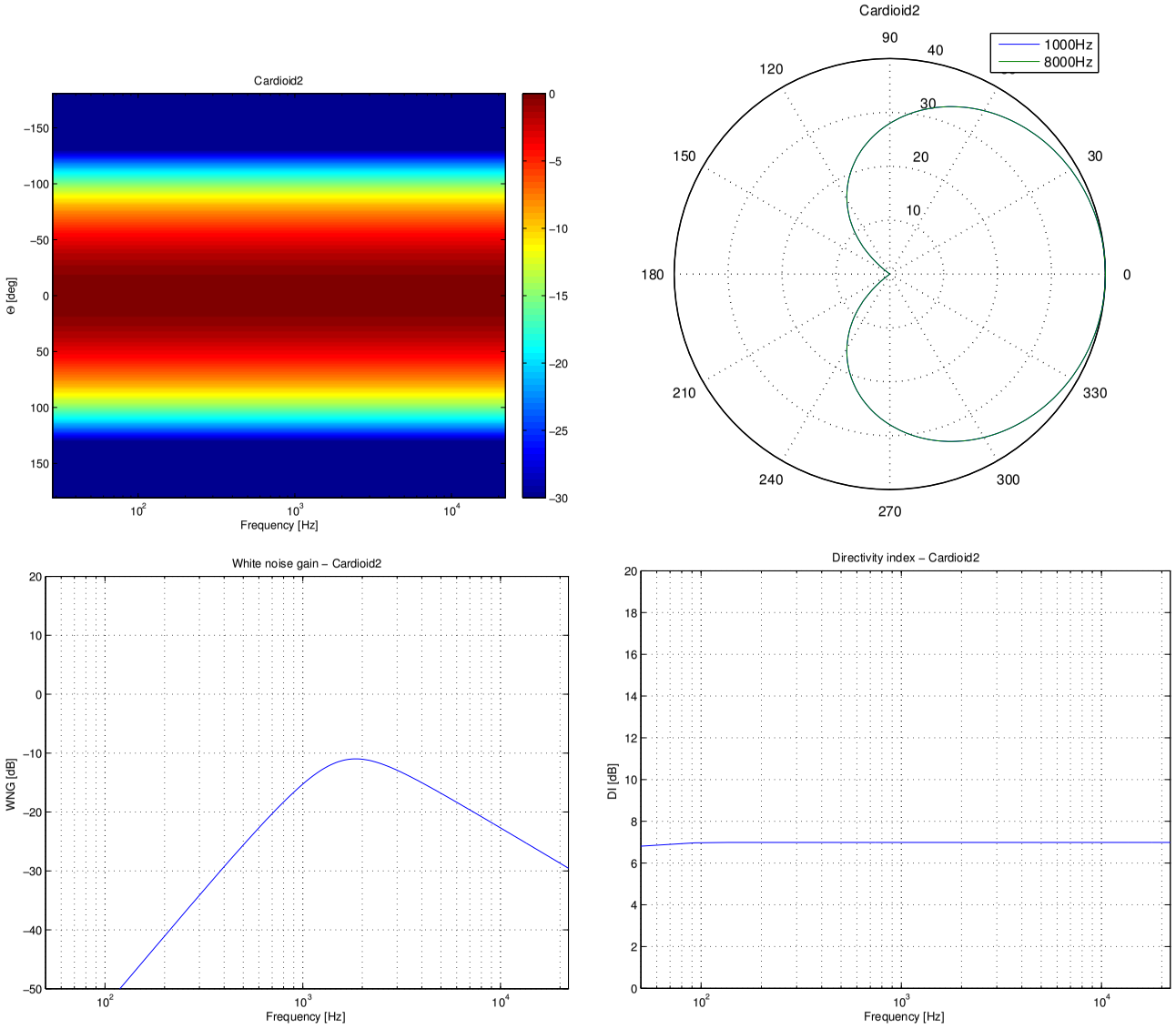


Abbildung 6.12: Kennwerte eines modalen Beamformers der Ordnung $N = 7$ mit Nierencharakteristik 2. Ordnung. Schallhartes 64-Kanal SMA mit $r = 4{,}2\,\mathrm{cm}$, Hyperinterpolation-Abtastung (vgl. Sloan und Womersley, 1998). Richtcharakteristik (links oben), Richtplot (rechts oben), WNG (links unten) und DI (rechts unten).

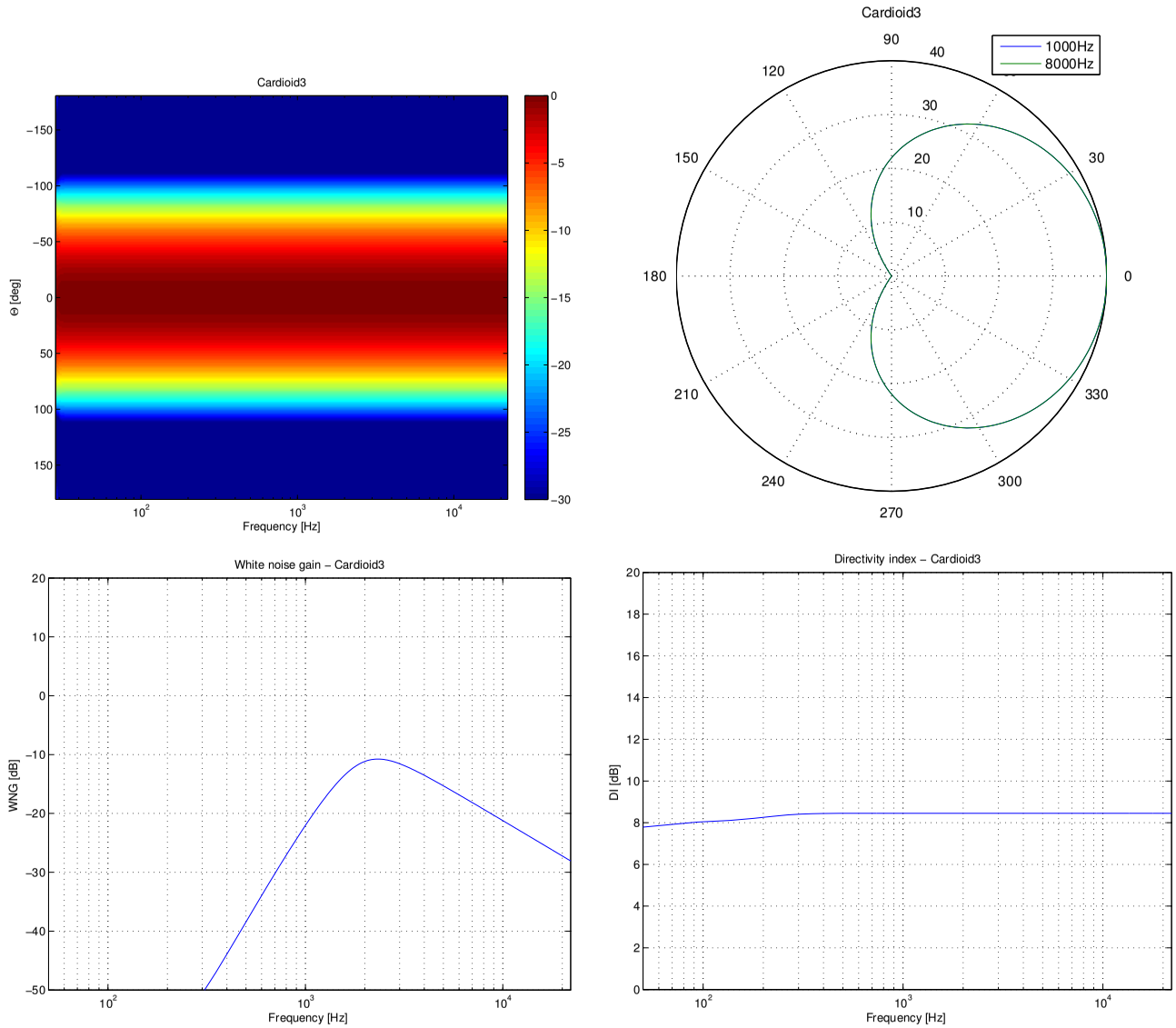


Abbildung 6.13: Kennwerte eines modalen Beamformers der Ordnung $N = 7$ mit Nierencharakteristik 3. Ordnung. Schallhartes 64-Kanal SMA mit $r = 4{,}2\,\mathrm{cm}$, Hyperinterpolation-Abtastung (vgl. Sloan und Womersley, 1998). Richtcharakteristik (links oben), Richtplot (rechts oben), WNG (links unten) und DI (rechts unten).

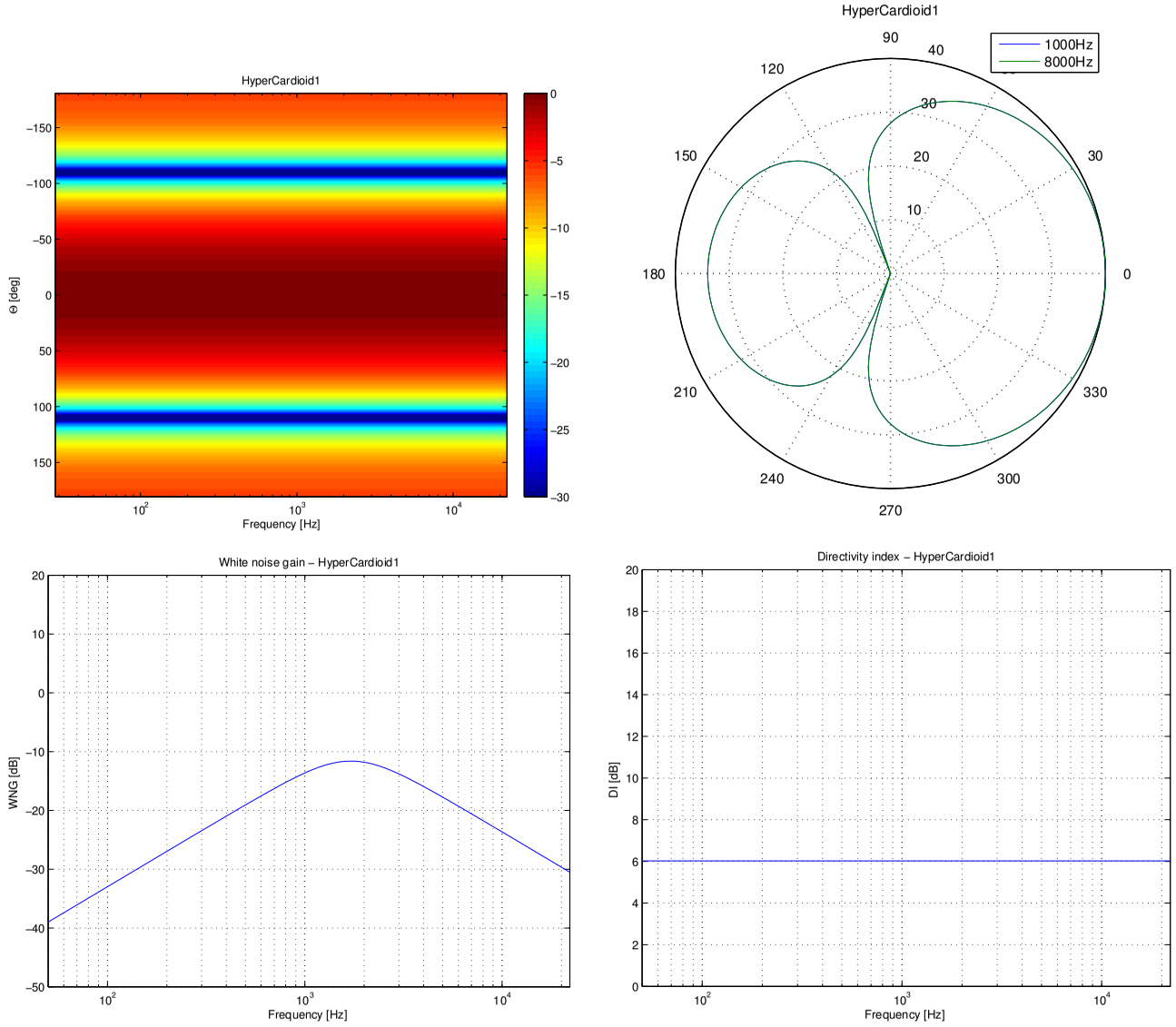


Abbildung 6.14: Kennwerte eines modalen Beamformers der Ordnung $N = 7$ mit Supernierencharakteristik 1. Ordnung. Schallhartes 64-Kanal SMA mit $r = 4{,}2\,\mathrm{cm}$, Hyperinterpolation-Abtastung (vgl. Sloan und Womersley, 1998). Richtcharakteristik (links oben), Richtplot (rechts oben), WNG (links unten) und DI (rechts unten).

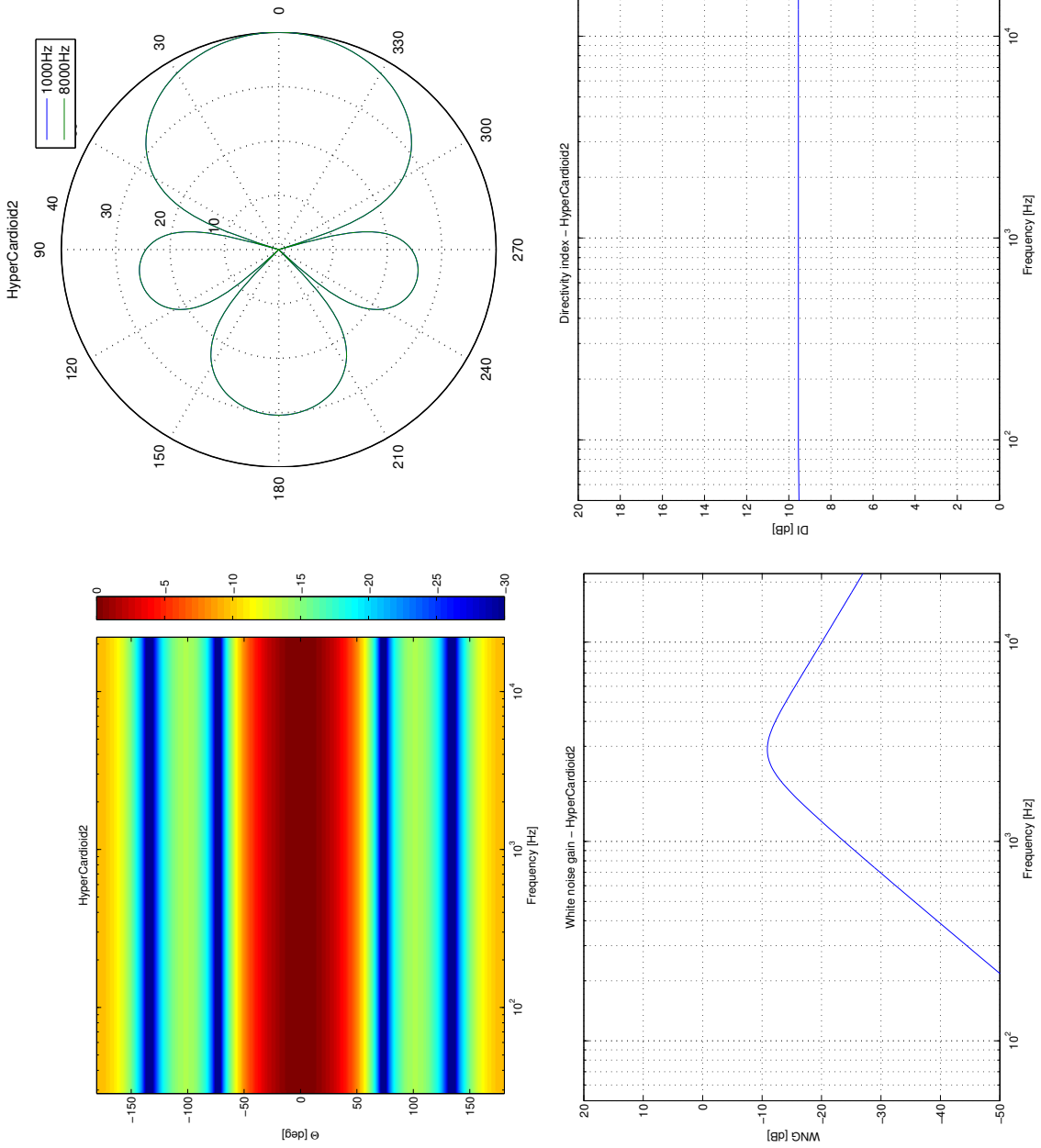


Abbildung 6.15: Kennwerte eines modalen Beamformers der Ordnung $N = 7$ mit Supernierencharakteristik 2. Ordnung. Schallhartes 64-Kanal SMA mit $r = 4{,}2\,\mathrm{cm}$, Hyperinterpolation-Abtastung (vgl. Sloan und Womersley, 1998). Richtcharakteristik (links oben), Richtplot (rechts oben), WNG (links unten) und DI (rechts unten).

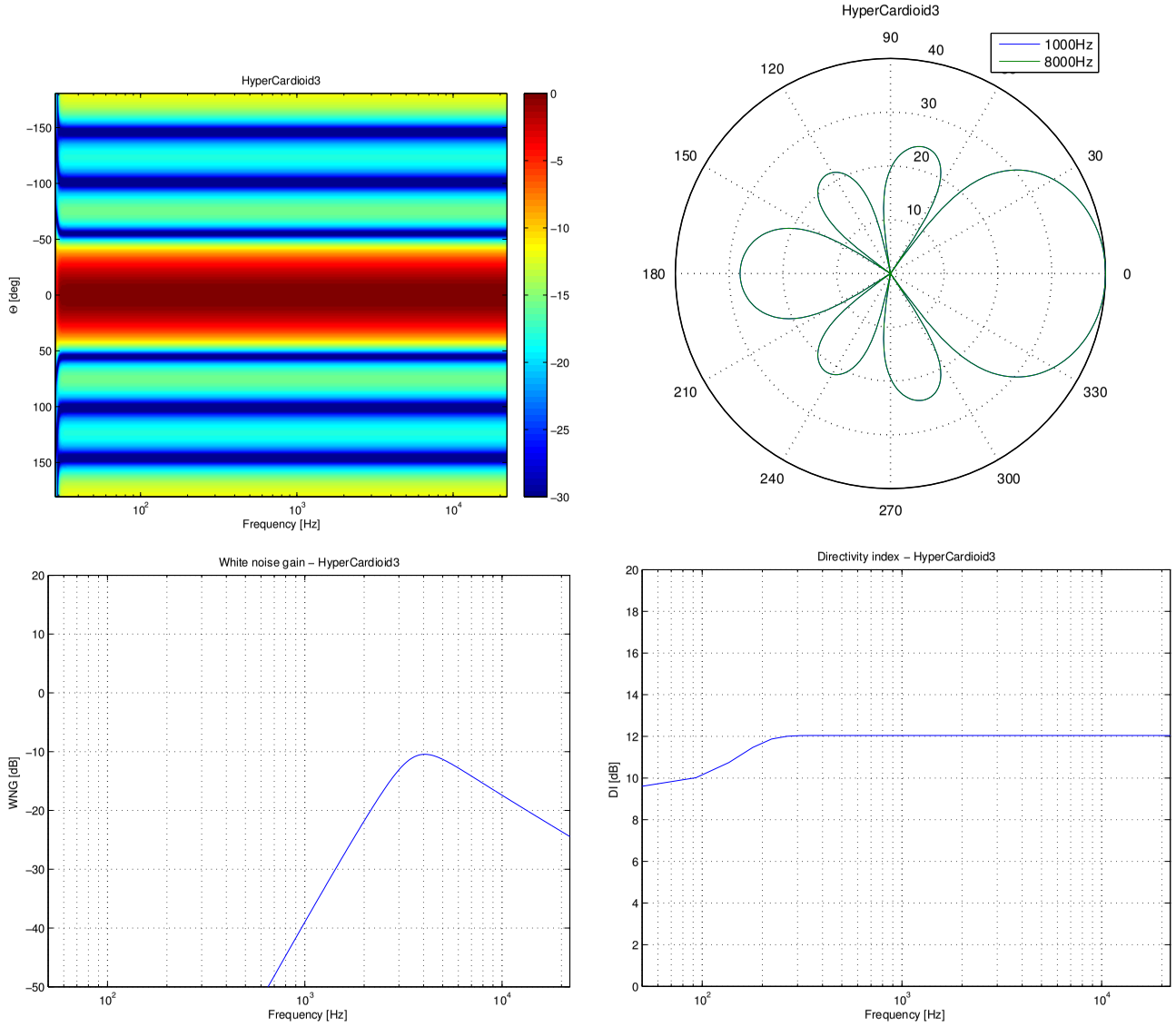


Abbildung 6.16: Kennwerte eines modalen Beamformers der Ordnung $N = 7$ mit Supernierencharakteristik 3. Ordnung. Schallhartes 64-Kanal SMA mit $r = 4{,}2\,\mathrm{cm}$, Hyperinterpolation-Abtastung (vgl. Sloan und Womersley, 1998). Richtcharakteristik (links oben), Richtplot (rechts oben), WNG (links unten) und DI (rechts unten).

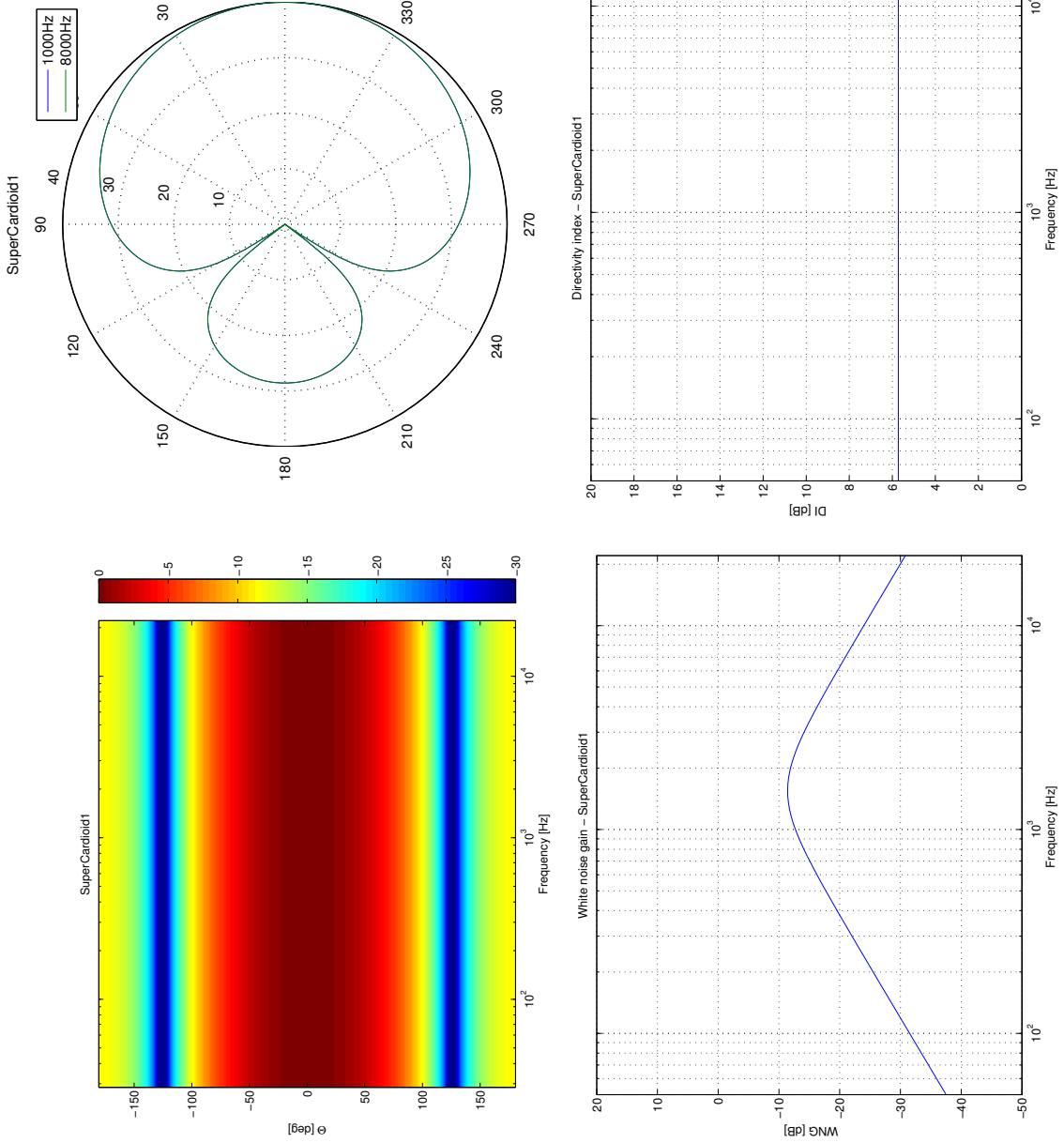


Abbildung 6.17: Kennwerte eines modalen Beamformers der Ordnung $N = 7$ mit Hypernierencharakteristik 1. Ordnung. Schallhartes 64-Kanal SMA mit $r = 4{,}2\,\mathrm{cm}$, Hyperinterpolation-Abtastung (vgl. Sloan und Womersley, 1998). Richtcharakteristik (links oben), Richtplot (rechts oben), WNG (links unten) und DI (rechts unten).

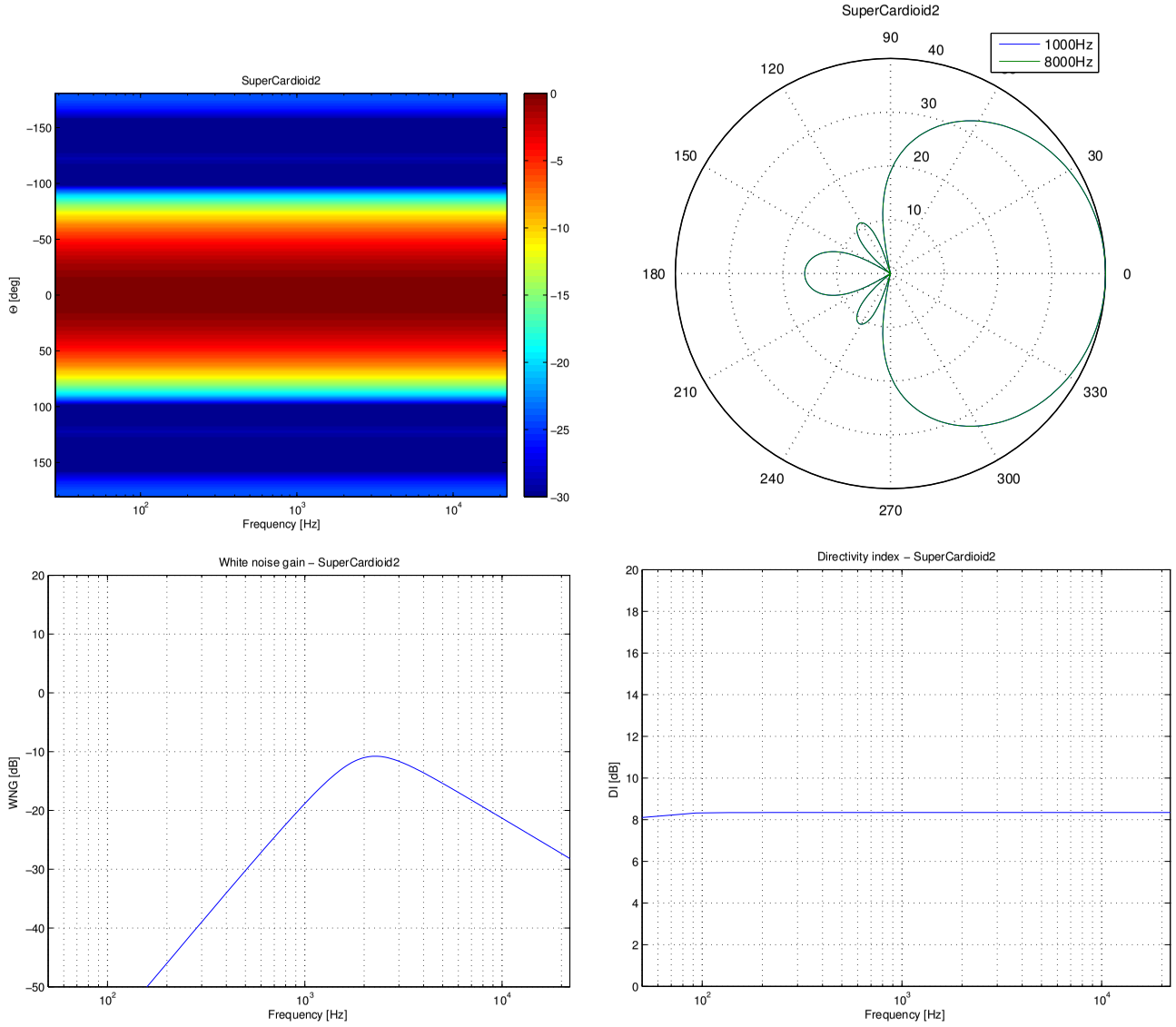


Abbildung 6.18: Kennwerte eines modalen Beamformers der Ordnung $N = 7$ mit Hypernierencharakteristik 2. Ordnung. Schallhartes 64-Kanal SMA mit $r = 4{,}2\,\mathrm{cm}$, Hyperinterpolation-Abtastung (vgl. Sloan und Womersley, 1998). Richtcharakteristik (links oben), Richtplot (rechts oben), WNG (links unten) und DI (rechts unten).

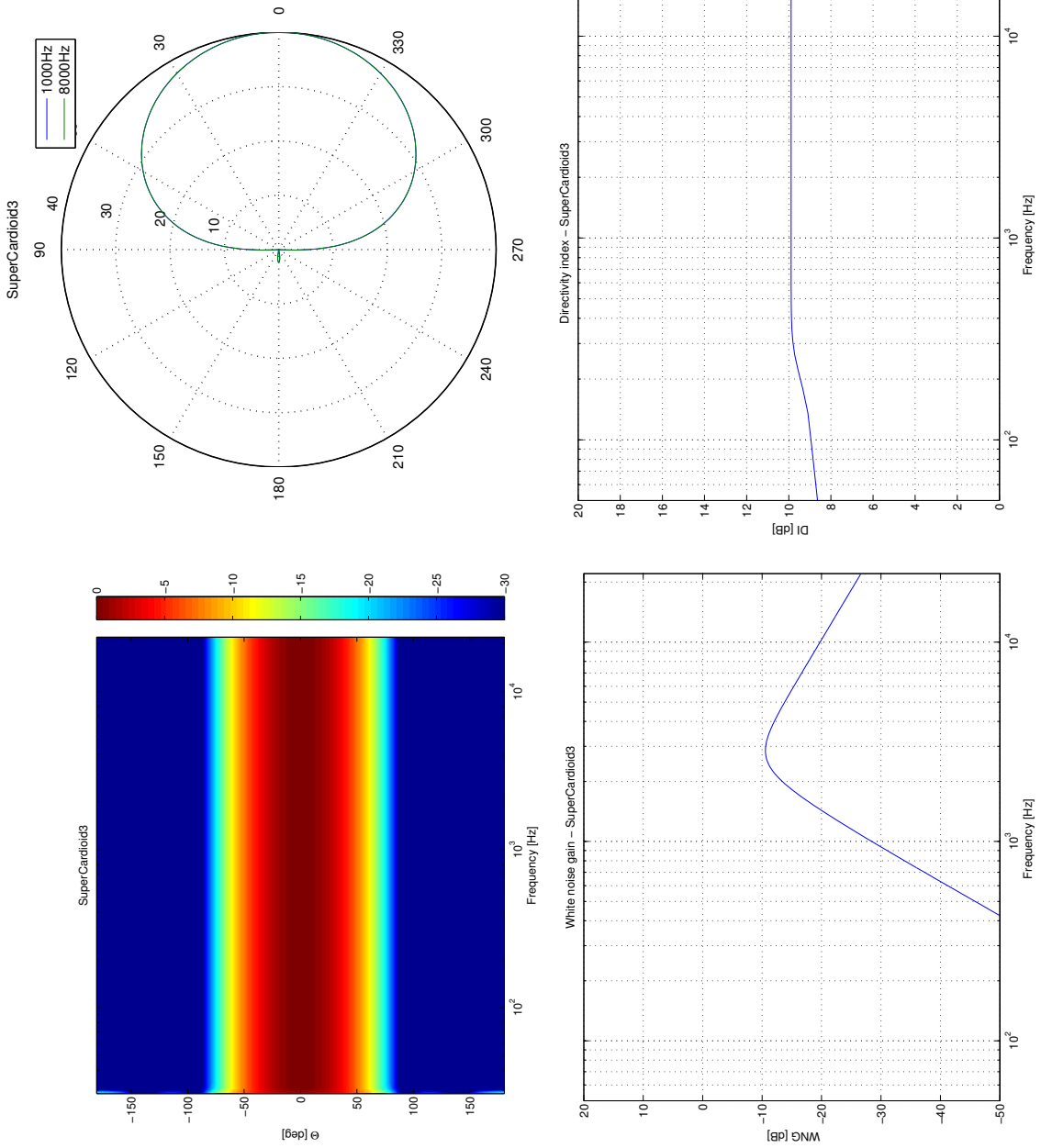


Abbildung 6.19: Kennwerte eines modalen Beamformers der Ordnung $N = 7$ mit Hypernierencharakteristik 3. Ordnung. Schallhartes 64-Kanal SMA mit $r = 4{,}2\,\mathrm{cm}$, Hyperinterpolation-Abtastung (vgl. Sloan und Womersley, 1998). Richtcharakteristik (links oben), Richtplot (rechts oben), WNG (links unten) und DI (rechts unten).

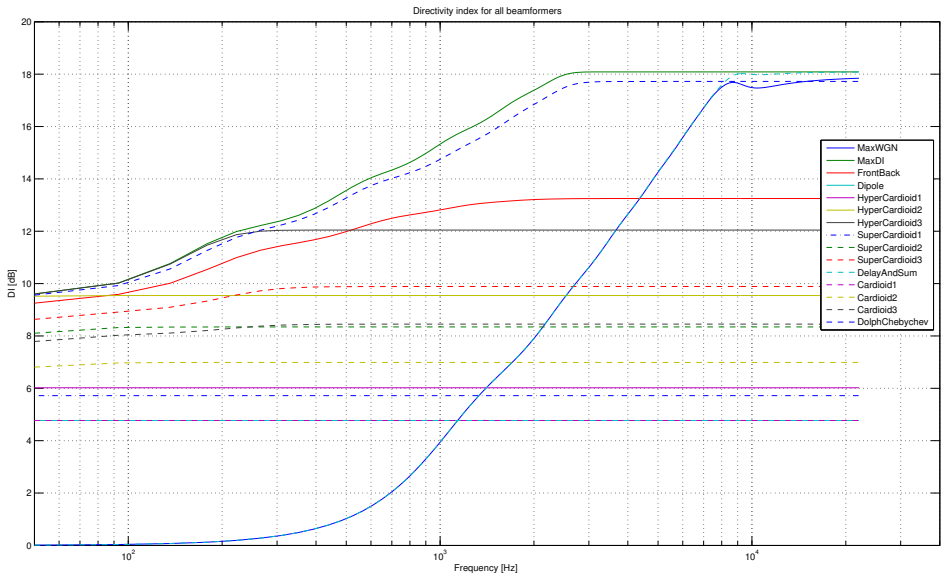


Abbildung 6.20: Vergleich der Richtindizes (DI) verschiedener modaler Beamformer.

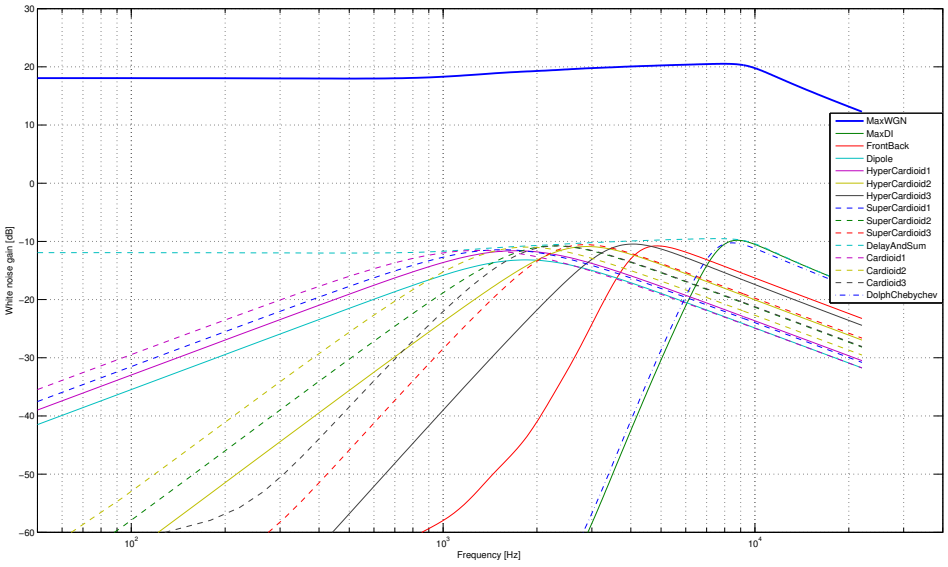


Abbildung 6.21: Vergleich des Gewinn für inkohärentes Rauschen (WNG) verschiedener modaler Beamformer.

6.4 Räumliche Abtastung

Ein sphärisches Mikrofonarray tastet die kontinuierliche Druckfunktion auf der Kugeloberfläche an räumlich diskreten Ortspunkten ab. Dabei muss die harmonische Ordnung N begrenzt werden, da es sonst zu störenden Aliasing-Effekten kommt (vgl. Li und Duraiswami, 2007; Rafaely et al., 2007; Meyer und Elko, 2008; Zotkin et al., 2008). In den folgenden Betrachtungen wird angenommen, dass das Schallfeld auf der Kugel strikt bandbegrenzt ist, sodass kein Aliasing entsteht. Die verwendeten Abtastgitter (d. h. die Verteilung der Abtastpunkte auf der Kugeloberfläche) haben dabei einen wesentlichen Einfluss auf die Fehlertoleranz der Schätzung der Koeffizienten des Wellenspektrums. Es gibt eine Reihe unterschiedlicher mathematischer Ansätze, das Quadraturproblem (d. h. eine numerisch bestmögliche Annäherung an das Integral über die Kugeloberfläche, bei einer möglichst geringen Anzahl an diskreten Stützstellen, zu erreichen) zu lösen (siehe z. B. Atkinson und Han, 2012, Kap. 5; Fornberg und Martel, 2014; Reeger und Fornberg, 2015). Im Folgenden wird das Verhalten verschiedener Abtastgitter auf der Kugeloberfläche anhand der Konditionszahl und der Orthonormalität der Kugelflächenfunktionen gezeigt. Auf die Bestimmung von zusätzlichen Punkten innerhalb der Kugel, mit dem Ziel die Stabilität der Approximation bei irregulären Frequenzen (d. h. bei den Nullstellen der Besselfunktionen) zu erhöhen, wird in Kap. 6.5 eingegangen.

Wird die Druckfunktion auf der Kugeloberfläche mit Q Sensoren abgetastet, können die Koeffizienten p_{nm} des Wellenspektrums über die diskrete sphärische Fouriertransformation berechnet werden (vgl. Abschnitt 6.1). Dabei muss das Integral in Gl. (6.6) mittels geeigneter Quadratur $Q_N(I)$ numerisch berechnet werden. Es ergeben sich die Quadraturknoten $\Omega_\ell = (\theta_\ell, \phi_\ell)$, mit $\ell = 1, \ldots, Q$, auf der Kugel mit den zugehörigen Quadraturgewichten α_ℓ:

$$\int_{\mathbb{S}^2} p(\Omega, k) Y_n^m(\Omega)^* d\Omega \approx \sum_{\ell=1}^{N} \alpha_\ell \, p(\Omega_\ell, k) Y_n^m(\Omega_\ell)^* = Q_N(I), \tag{6.107}$$

wobei N die maximale Ordnung der verwendeten Kugelflächenfunktionen ist. Als Alternative lassen sich die unbekannten Koeffizienten p_{nm} auch über einen *Least-*

Squares-Ansatz bestimmen, sodass gilt:

$$\underset{p_{nm}}{\text{minimize}} \left\| p(\Omega_\ell, k) - \sum_{n=0}^{N} \sum_{m=-n}^{n} p_{nm} j_n(kr) Y_n^m(\Omega_\ell) \right\|^2 . \tag{6.108}$$

Das Abtastgitter wird dabei oft durch eine Quadraturformel berechnet. Die *Least-Squares*-Methode liefert, vor allem bei einer geringen Anzahl an Abtastpunkten, eine genauere Lösung als die Quadratur, ist jedoch wesentlich rechenaufwändiger.

Generell lässt sich die diskrete sphärische Fouriertransformation wie folgt formulieren (s. Gln. 6.8 und 6.9):

$$\mathbf{p}_N = \mathbf{Y}_N \mathbf{p}_{nm}, \tag{6.109}$$

$$\mathbf{p}_{nm} = \mathbf{Y}_N^{-1} \mathbf{p}_N, \tag{6.110}$$

mit dem $Q \times 1$ Vektor der räumlichen Abtastwerte des Schalldrucks

$$\mathbf{p}_N = \left[\; p(\Omega_1), \;\; p(\Omega_2), \;\; \cdots, \;\; p(\Omega_Q) \; \right]^{\mathrm{T}},$$

und dem $(N+1)^2 \times 1$ Vektor der (sphärischen) Fourierkoeffizienten

$$\mathbf{p}_{nm} = \left[\; p_{0,0}(k, r_0), \;\; p_{1,-1}(k, r_0), \;\; p_{1,1}(k, r_0), \;\; \ldots \;\; p_{N,N}(k, r_0) \; \right]^T .$$

Die $Q \times (N+1)^2$ Matrix $\mathbf{Y}_N$ beinhaltet die an den Punkten Ω_ℓ abgetasteten Kugelflächenfunktionen $\mathbf{y}_N(\Omega_\ell)$ der Ordnung N:

$$\mathbf{Y}_N = \left[\; \mathbf{y}_N(\Omega_1), \;\; \mathbf{y}_N(\Omega_2), \;\; \ldots, \;\; \mathbf{y}_N(\Omega_Q) \; \right]^T ,$$

$$\mathbf{y}_N(\Omega_\ell) = \left[\; y_{0,0}(\Omega_\ell), \;\; y_{-1,1}(\Omega_\ell), \;\; \ldots, \;\; y_{N,N}(\Omega_\ell) \; \right]^{\mathrm{T}} .$$

Um $\mathbf{p}_{nm}$ aus $\mathbf{p}_N$ zu berechnen, muss die Matrix $\mathbf{Y}_N$ invertiert werden. $\mathbf{Y}_N$ ist in vielen Fällen (wie z. B. aufgrund einer nicht gleichförmigen Verteilung der Abtastpunkte auf der Kugel) schlecht konditioniert, die Inverse $\mathbf{Y}_N^{-1}$ kann nicht direkt berechnet werden. Ein naheliegender Ansatz ist das Problem nur so gut es geht zu lösen, und gleichzeitig zu verhindern, dass die Norm der Lösung zu

groß wird. Ein weit verbreiteter Ansatz macht sich die Approximationseigenschaft der Singulärwertzerlegung (*singular value decomposition*, SVD) zunutze, um die Pseudoinverse von $\mathbf{Y}_N$ zu berechnen (vgl. Golub und Van Loan 1996, Kap. 2; Golub und Kahan 2006).

Eine Zerlegung der Form $\mathbf{Y}_N = \mathbf{U}\mathbf{S}\mathbf{V}^T$ heißt SVD der Matrix $\mathbf{Y}_N$. Die Einträge der Diagonalmatrix $\mathbf{S}$ sind die Singulärwerte von $\mathbf{Y}_N$, die Spalten der orthogonalen Matrizen $\mathbf{U}$ und $\mathbf{V}$ die linken und rechten Singulärvektoren. Die Singulärwerte sind eindeutig als die positiven Quadratwurzeln der positiven Eigenwerte von $\mathbf{Y}_N\mathbf{Y}_N^T$ (oder $\mathbf{Y}_N^T\mathbf{Y}_N$) bestimmt. Die unitären Matrizen $\mathbf{U}$ und $\mathbf{V}$ sind hingegen nicht eindeutig bestimmt (vgl. Liesen und Mehrmann, 2015, Kap. 19). Werden nun die Diagonalelemente der Matrix $\mathbf{S}$ in absteigender Reihenfolge sortiert und die Matrix $\mathbf{S}$ auf K nicht verschwindende Singulärwerte begrenzt, ergibt sich die abgeschnittene Singulärwertzerlegung (*truncated singular value decomposition*, TSVD).[134] Diese führt auf die Pseudoinverse $\mathbf{Y}_N^\dagger = \tilde{\mathbf{V}}\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{U}}^T$, welche die Inversion regularisiert (vgl. Golub und Kahan, 2006). Angewendet auf die diskrete sphärische Fouriertransformation ergibt sich (siehe auch Gl. 6.9):

$$\tilde{\mathbf{p}}_{nm} = \mathbf{Y}_N^\dagger \mathbf{p}_N. \tag{6.111}$$

Die verallgemeinerte Inversion (Gl. 6.111) hat, in Abhängigkeit von der Dimension der Matrix $\mathbf{Y}_N$ und der Anzahl K der nicht verschwindenden Singulärwerte, folgende Eigenschaften (s. a. Noisternig und Katz, 2009; Noisternig et al., 2011):

1. $K = (N+1)^2 \leq Q \Rightarrow$ Transformation (*discrete spherical harmonic transform*, DSHT):[135] Die Pseudoinverse $\mathbf{Y}_N^\dagger = (\mathbf{Y}_N^T\mathbf{Y}_N)^{-1}\mathbf{Y}_N^T$ invertiert Gl. (6.109) von links. Ist das Schallfeld auf der Kugel strikt bandbegrenzt ($\mathbf{p} = \mathbf{p}_N$), ist die Schätzung der Fourierkoeffizienten exakt ($\tilde{\mathbf{p}}_{nm} = \mathbf{p}_{nm}$). Das Schallfeld wird exakt abgebildet.

2. $K = Q \leq (N+1)^2 \Rightarrow$ Interpolation (*discrete spherical harmonic interpolation*, DSHI):[136] Die Pseudoinverse $\mathbf{Y}_N^\dagger = \mathbf{Y}_N^T(\mathbf{Y}_N\mathbf{Y}_N^T)^{-1}$ invertiert

[134] Siehe z. B. Hansen (1998, Kap. 3) und Vogel (2002).
[135] Die DSHT wird auch als spektrale Analyse bezeichnet (vgl. Boyd, 2000).
[136] Die DSHI wird auch als pseudo-spektrale Analyse bezeichnet (vgl. Boyd, 2000).

Gl. (6.109) von rechts. Ist das Schallfeld auf der Kugel strikt bandbegrenzt ($\mathbf{p} = \mathbf{p}_N$), wird dieses an den Abtastpunkten exakt reproduziert. Zwischen den Abtastpunkten wird das Schallfeld interpoliert.

3. $K < \min\{(N+1)^2, Q\} \Rightarrow$ Approximation (*discrete spherical harmonic approximation*, DSHA): Die Inversion ist weder eine DSHT noch eine DSHI, das Schallfeld wird bestmöglich angenähert.

Die Konditionszahl und die Orthonormalität sind wichtige Bewertungsmaße für die Qualität eines Abtastgitters auf der Kugel. Diese werden im Folgenden kurz erläutert und für unterschiedliche Abtastgitter verglichen.

Konditionszahl. Die Konditionszahl der Matrix $\mathbf{Y}$ ist definiert als (siehe z. B. Köhler, 2005):

$$\mathrm{cond}_p(\mathbf{Y}) = \|\mathbf{Y}\|_p \cdot \left\|\mathbf{Y}^{-1}\right\|_p, \tag{6.112}$$

wobei $\|\cdot\|_p$ die p-Norm bezeichnet. Es gilt:

$$\mathrm{cond}_p(\mathbf{Y}) \geq 1. \tag{6.113}$$

Mit der Konditionszahl lässt sich der Einfluss von Fehlern auf die Robustheit der Inversion $\mathbf{Y}^{-1}$ abschätzen (siehe z. B. Meister, 2015, Kap. 2.3). Matrizen mit kleinen Konditionszahlen werden als gut konditioniert bezeichnet. Kleine relative Fehler in den Daten (wie z. B. Sensorrauschen oder eine Fehlpositionierung der Mikrofone) sind nur als kleine relative Fehler in den Lösungen bemerkbar. Matrizen mit großen Konditionszahlen, werden als schlecht konditioniert bezeichnet. Die Matrix muss für die Inversion regularisiert werden, um den Fehler in den Lösungen auf ein vertretbares Maß zu begrenzen. Die Regularisierung kann zum Beispiel mit Hilfe der TSVD erfolgen.

Abbildung 6.22 zeigt die Existenz und Art der diskreten sphärischen Fouriertransformation für unterschiedliche Abtastgitter auf der Kugel. Die Marker zeigen die Existenz der unterschiedlichen Lösungen (DSHT, DSHI, DSHA) für ein bestimmtes Abtastgitter und eine bestimmte Anzahl an Abtastpunkten. Dabei wird angenommen, dass bei Konditionszahlen $\mathrm{cond}(\mathbf{Y}_N) < 20\,\mathrm{dB}$ die Inverse der Ma-

trix $\mathbf{Y}$ ohne Regularisierung über $\mathbf{Y}^{\dagger}$ bestimmt werden kann (DSHT bzw. DSHI). Konditionszahlen $\text{cond}(\mathbf{Y}_N) \geq 20\,\text{dB}$ erfordern eine Regularisierung (DSHA). Die Simulationen wurden für eine maximale harmonische Ordnung von $N = 9$ und eine unterschiedliche Anzahl von Abtastpunkten Q durchgeführt. Folgende Abtastgitter wurden betrachtet:

1. *Extremal points for hyperinterpolation* (hi), vgl. Sloan und Womersley (2004),
2. *Spiral points* (sp), vgl. Rakhmanov et al. (1994),
3. *Equal-area partitions* (eqa), vgl. Saff und Kuijlaars (1997),
4. *HEALPix* (healpix), vgl. Gorski et al. (2005),
5. *Gauss-Legendre grid* (gl), vgl. Sneeuw (1994),
6. *Equidistant cylindrical partitions* (ecp), vgl. Sneeuw (1994),
7. *Equiangle grid* (equiangle), vgl. Driscoll und Healy (1994),
8. *Cubature grid* (cube), vgl. Fliege (1999) und
9. *Lebedev grid* (leb), vgl. Lebedev (1965, 1976, 1977).

Aus Abbildung 6.22(b) wird ersichtlich, dass einzig die Hyperinterpolation eine exakte und gut konditionierte Lösung der Inversion der Matrix $\mathbf{Y}$ (DSHT und DSHI) bei der kritischen Anzahl der Abtastpunkte $Q = (N+1)^2$ ermöglicht. Die übrigen Abtastgitter erlauben hier nur eine DSHA. Aus diesem Grund wurden bei dem in dieser Arbeit entwickelten Mikrofonarray (siehe Anhang C) die Mikrofone mit einem Hyperinterpolations-Gitter auf der Kugel verteilt.

Orthonormalität. Die Kugelflächenfunktionen sind orthonormal. Es gilt:

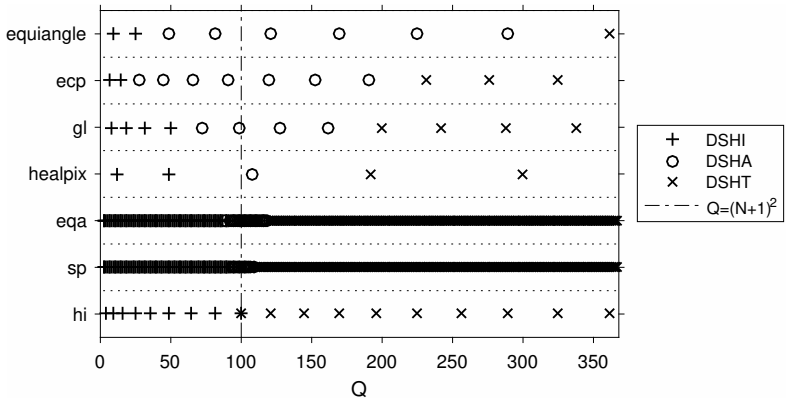
$$\int_{\mathbb{S}^2} Y_n^m(\Omega)^* Y_{n'}^{m'}(\Omega) d\Omega = \delta_{nn'}\delta_{mm'}. \tag{6.114}$$

Das Integral kann mittels geeigneter Quadratur numerisch berechnet werden

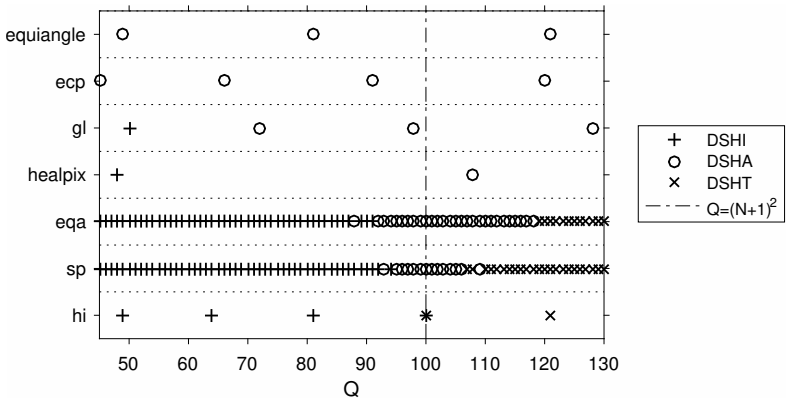
$$\frac{4\pi}{Q} \sum_{\ell=1}^{Q} Y_n^m(\Omega_\ell)^* Y_{n'}^{m'}(\Omega_\ell) \alpha_{n'}^{m'}(\Omega_\ell) = \delta_{nn'}\delta_{mm'}, \tag{6.115}$$

$$\text{mit } 0 \leq n \leq N_{eff} \text{ und } -n \leq m \leq n,$$

$$\text{bzw. } 0 \leq n' \leq N \text{ und } -n' \leq m \leq n'.$$



(a)



(b)

Abbildung 6.22: Existenz und Art der sphärischen Fouriertransformation für unterschiedliche Abtastgitter auf der Kugel. Die Marker zeigen die Existenz der unterschiedlichen Lösungen – DSHT (×), DSHI (+) und DSHA (∘) – für ein bestimmtes Abtastgitter und eine bestimmte Anzahl an Abtastpunkten. Die Konditionszahl wurde zur Regularisierung auf $\mathrm{cond}\{\mathbf{Y}_9\} < 20\,\mathrm{dB}$ begrenzt. Die vertikale strichpunktierte Linie kennzeichnet die Anzahl der zur kritischen Abtastung $Q = (N+1)^2$ benötigten Abtastpunkte auf der Kugel. Die Region um die kritische Abtastung wird in Teilbild (b) detailliert dargestellt, um das einzigartige Verhalten der Hyperinterpolation (hi) zu zeigen.

$\Omega_\ell = (\theta_\ell, \phi_\ell)$, $\ell = 1, \ldots, Q$, sind die Quadraturknoten auf der Kugel und $\alpha_{n'}^{m'}(\Omega_\ell)$ die zugehörigen Quadraturgewichte für $Y_{n'}^{m'}(\Omega_\ell)$. Die effektive Ordnung N_{eff} entspricht nur dann der maximalen Ordnung N des Arrays, wenn diese kleiner als die maximale Ordnung des Schallfeldes ist. Ansonsten entspricht N_{eff} der Ordnung des Schallfeldes.

Gleichung (6.115) ist für überstimmte und rang-beschränkte Matrizen nicht eindeutig bestimmt. Es entsteht ein Orthonormalitätsfehler (*orthonormality error noise*, OEN), der sich wie folgt formulieren lässt:

$$\frac{4\pi}{Q}\sum_{\ell=1}^{Q} Y_n^m(\Omega_\ell)^* Y_{n'}^{m'}(\Omega_\ell)\alpha_{n'}^{m'}(\Omega_\ell) = \delta_{nn'}\delta_{mm'} + \epsilon_{nn'}^{mm'}. \tag{6.116}$$

Li und Duraiswami (2007) zeigen, dass der Orthonormalitätsfehler nur dann vernachlässigt werden kann, wenn

$$\left|\frac{i^n b_n(kr_m)}{i^{n'} b_{n'}(kr_m)}\epsilon_{nn'}^{mm'}\right| \ll 1, \quad \forall\, n, n', m, m' \tag{6.117}$$

erfüllt ist. Daraus lässt sich folgende Bedingung formulieren:

$$\left|\epsilon_{nn'}^{mm'}\right| \ll \left|\frac{b_{n'}(kr_m)}{b_n(kr_m)}\right|, \quad \forall\, n, n', m, m'. \tag{6.118}$$

Die modalen Amplituden b_n klingen zu höheren Ordnung hin sehr schnell ab (vgl. Abb. 6.1). Aus diesem Grund lassen sich die Komponenten höherer Ordnung eines Schallfeldes nur dann mit ausreichender Genauigkeit abbilden, wenn die Bedingung in Gl. (6.118) erfüllt ist. Zudem ist Gl. (6.118) unabhängig von der Amplitude der einfallenden Schallwelle. Werden alle Komponenten des Schallfeldes vom Mikrofonarray erfasst, können diese aufgrund des Orthonormalitätsfehlers trotzdem nicht abgebildet werden (vgl. Li und Duraiswami, 2007).

Der Orthonormalitätsfehler lässt sich, für eine ausgewählte Wellenzahl, sehr anschaulich durch das innere Produkt der numerisch ermittelten Kugelflächenfunktionen darstellen. Abbildung 6.23 zeigt die sich für unterschiedliche Abtastgitter auf der Kugel ergebenden Matrizen in Abhängigkeit von der Ordnung N. Bei analytischer Berechnung über das Integral würde sich die Einheitsmatrix ergeben. Die Anzahl der für ein bestimmtes Abtastgitter benötigten Abtastpunkte variiert mit der Ordnung N. Das Lebedev Gitter benötigt $Q = 1.3\,(N+1)^2$, das Gauß-Legendre Gitter $Q = 2(N+1)^2$, das Equi-Angular Gitter $Q = 2(N+1)^2$, die Hyperinterpolation hingegen nur $(N+1)^2$ Abtastpunkte auf der Kugel (vgl. Abb. 6.22). Es ist leicht zu erkennen, dass die Hyperinterpolation bei höheren Ord-

nungen zum Teil kein Orthonormalsystem bildet und stark von der Einheitsmatrix, die sich bei analytischer Berechnung über das Integral ergeben würde, abweicht. Das Gauß-Legendre Gitter ist in dieser Hinsicht optimal, benötigt aber, wie bereits angesprochen, wesentlich mehr Abtastpunkte. Dies ist in der praktischen Anwendung oft nur schwer umsetzbar.

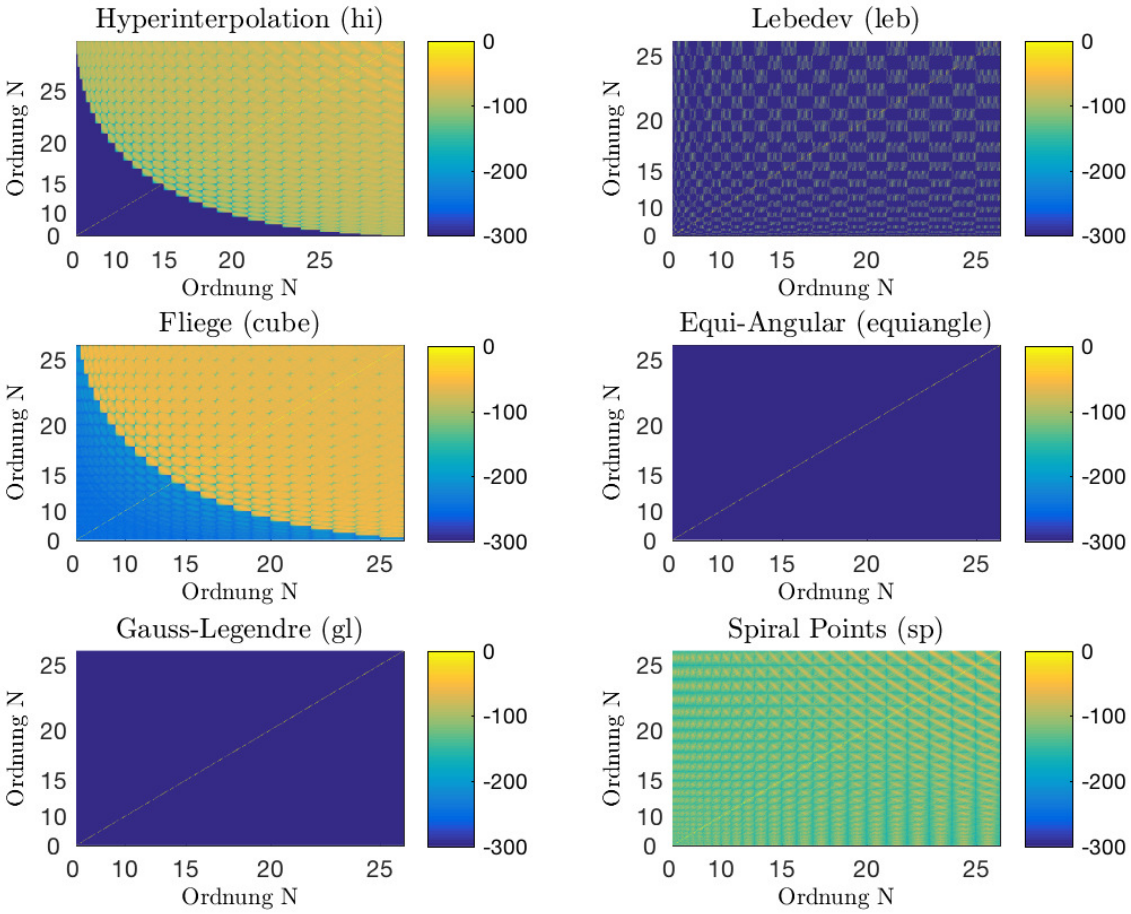


Abbildung 6.23: Orthonormalitätsfehler für unterschiedliche Abtastgitter der Ordnung $N = 20$ auf der Kugel in Abhängigkeit von der Ordnung n der Kugelflächenfunktionen. Die Teilbilder stellen das innere Produkt der numerisch ermittelten Kugelflächenfunktionen bei einer ausgewählten Wellenzahl dar.

Es ist also ein Kompromiss zwischen dem Orthonormalitätsfehler und der Anzahl der benötigten Abtastpunkte zu finden. Ein Array mit möglichst hoher räum-

licher Bandbreite erfordert hohe Ordnungen. Dies motiviert die Entwicklung optimierter Abtastgitter, die auch das Volumen innerhalb der Kugel mit berücksichtigen. Das im Rahmen dieser Arbeit vorgestellte Verfahren zur Optimierung der räumlichen Abtastung wird im folgenden Kapitel diskutiert.

6.5 Entwurf eines robusten modalen Beamformers

Wie in einigen Studien gezeigt wurde (vgl. Abhayapala und Ward, 2002; Gover et al., 2004; Rafaely, 2011), werden sphärische Mikrofonarrays bei einigen Frequenzen instabil. Dies erklärt sich durch die Nullstellen der sphärischen Bessel-Funktionen, die bei der Inversion zu numerischen Instabilitäten führen (vgl. Abschnitt 6.1). Meyer und Elko (2002) verwenden eine geschlossene, schallharte Kugel, um das Problem zu umgehen. In der praktischen Anwendung ist dieser Ansatz jedoch auf relativ kleine Arrayradien beschränkt. Dies führte zur Entwicklung zahlreicher alternativer Arraygeometrien, die im Folgenden kurz zusammengefasst werden. Eine Möglichkeit die Stabilität in der Nähe der Nullstellen der Bessel-Funktionen zu erhöhen besteht darin, anstatt der ungerichteten Mikrofone (Druckempfänger) in Richtung der Flächennormale ausgerichtete Nierenmikrofone zu verwenden (vgl. Rahim und Davies, 1982; Meyer, 2001; Balmages und Rafaely, 2007). Auch dieser Ansatz ist in der Praxis nur schwer umsetzbar, da (i) Nierenmikrofone bei tiefen Frequenzen meist ein relativ hohes Eigenrauschen haben und (ii) sich Fehler in der Ausrichtung und Positionierung der Mikrofone aufgrund der Richtcharakteristik stark bemerkbar machen. Mit Arrays, die aus zwei konzentrischen Kugelschalen mit unterschiedlichen Radien bestehen (*double sphere arrays*), lässt sich das Problem der schlechten Konditionierung lösen (vgl. Balmages und Rafaely, 2007; Parthy et al., 2009; Jin et al., 2014). Auf den beiden Kugelschalen[137] werden meist dieselben Abtastgitter einer bestimmten Ordnung N verwendet. Der wesentliche Nachteil dieses Ansatzes besteht also darin, dass im Vergleich zu einem einfachen Array die doppelte Anzahl an Mikrofonen benötigt wird.

Nicht-kugelförmige Arraygeometrien benötigen meist eine geringere Anzahl an Abtastpunkten, um die Stabilität in der Umgebung der Nullstellen der Bessel-

[137] Die Arrays bestehen entweder aus zwei offene Kugelschalen oder aus einer geschlossenen Kugel (innen) und einer offenen Kugelschale (außen).

Funktionen zu verbessern. Rafaely (2008) zeigt, wie sich bei einem offenen Array (*open spherical shell array*), mit einigen wenigen Mikrofonen im Inneren der Kugel, die Robustheit über einen großen Wellenzahlbereich erhöhen lässt. Die Positionen der innen liegenden Abtastpunkte werden dabei über ein nichtlineares Optimierungsverfahren mit Nebenbedingungen bestimmt, welches die Konditionszahl der Matrix des Produkts der sphärischen Bessel-Funktionen mit den Kugelflächenfunktionen minimiert. Allerdings ist es mit dieser Methode schwierig, die für eine gegebene Ordnung benötigte Anzahl an innenliegenden Abtastpunkten eindeutig zu bestimmen, damit die Stabilität und Konvergenz des Optimierungsverfahrens gewährleistet ist. Abhayapala und Gupta (2009) tasten das Schallfeld mit mehreren kreisförmigen Arrays mit jeweils unterschiedlicher Anzahl an Abtastpunkten ab (*hybrid array*). Unter Ausnutzung bestimmter Eigenschaften der assoziierten Legendre- und sphärischen Bessel-Funktionen, lässt sich das Array über einen erweiterten Frequenzbereich stabilisieren. Das doppelseitige Konus-Array (*double sided cone array*, vgl. Gupta und Abhayapala, 2010) ist ein weiterer Vorschlag zur Lösung des Nullstellenproblems. Dabei wird die radiale Orthogonalität von auf der Oberfläche eines doppelseitigen Konus ausgewerteten Bessel-Funktionen verwendet, um die sphärischen Fourierkoeffizienten über einen möglichst großen Frequenzbereich zu schätzen. Es kann allerdings gezeigt werden, dass eine stabile Schätzung nicht bei allen Frequenzen möglich ist. Diesem Problem kann dadurch begegnet werden, indem das Schallfeld über zwei oder mehrere Konusse abgetastet wird. Allerdings erfordert dies eine relativ hohe Anzahl an Abtastpunkten und ist für die meisten praktischen Anwendungen nicht geeignet. In einem weiteren Vorschlag von Alon und Rafaely (2012) wird das Schallfeld mit einem auf die Oberfläche eines Rotationstorus (*spindle torus array*) projiziertes gleichförmiges Abtastgitter abgetastet, wodurch sich die Robustheit gegenüber Rauschen erhöht. Dieses Verfahren lässt sich sehr einfach mit einem Rastermikrofonarray (*scanning microphone array*)[138] implementieren und kann zum Beispiel zur Messung von 3D Raumimpulsantworten verwendet werden.

Mignot et al. (2014) verwenden ein Array zufällig angeordneter Mikrofone und nähern das Schallfeld über eine Summe ebener Wellen an. Dadurch lassen sich

[138]Beim Rastermikrofonarray wird das Schallfeld mit einem auf einem Roboterarm montierten Druckmikrofon Punkt für Punkt abgetastet.

Raumimpulsantworten bei niedrigen Frequenzen interpolieren. Die Methode zur Bestimmung der Abtastpunkte wird nicht ausreichend theoretisch diskutiert, ist jedoch sehr wahrscheinlich suboptimal. Rafaely (2011) zeigt, wie sich mit einer ungleichförmigen Abtastung zeitlich begrenzter Signale (z. B. Raumimpulsantworten) im Frequenzbereich, Abtastwerte nahe der Nullstellen der Bessel-Funktionen vermeiden lassen. Die ungleichförmige Abtastung wird numerisch ermittelt und minimiert die Konditionszahl der Fouriermatrix und der Diagonalmatrix der sphärischen Besselfunktionen.

In Chardon, Kreuzer und Noisternig (2014b, 2015) wurde eine Methode zum Entwurf optimaler Abtastgitter vorgestellt, mit denen das Schallfeld innerhalb des Volumens eines Arrays mit möglichst hoher Genauigkeit und über einen möglichst breiten Frequenzbereich interpoliert werden kann. Diese Methode wird in den folgenden Abschnitten ausführlich diskutiert. Dazu wird zuerst die Theorie der Approximation und Interpolation von Schallfeldern zusammengefasst. Es wird gezeigt, wie sich die Wahl einer Basis auf die Stabilität der Schätzung der Fourierkoeffizienten, welche die Grundvoraussetzung für eine stabile Interpolation ist, auswirkt. Die numerischen Instabilitäten in der Nähe der Nullstellen der sphärischen Bessel-Funktionen stellen ein besonderes Problem dar. Numerische Simulationen unterschiedlicher Abtastgitter zeigen, dass sich die Interpolation mit einigen wenigen Abtastpunkten im Inneren des Volumens über einen breiten Frequenzbereich stabilisieren lässt. Dies kann neben der Kugel auch ganz allgemein für konvexe und sternförmige Gebiete (wie z. B. ellipsoidische und kubische Arrays) gezeigt werden. Die Simulationen zeigen auch die Effizienz des vorgestellten Optimierungsansatzes, der über den gesamten Frequenzbereich den kleinsten Interpolationsfehler ergibt. Allerdings sind dabei die Abtastpunkte über das gesamte Volumen verteilt, was sich in der Praxis durch mechanische Einschränkungen meist nur schwer implementieren lässt. Aus diesem Grund wird der Ansatz in Folge dahingehend modifiziert, dass zusätzliche Randbedingungen (wie z. B. Einschränkungen in der Arraygeometrie) berücksichtigt werden können. Numerische Simulationen unterschiedlicher Arraygeometrien (*double sphere*, *mixed sphere* und *spindle torus*) zeigen, dass sich mit dem modifizierten Ansatz nicht nur die Robustheit der Interpolation und der Schätzung der Fourierkoeffizienten, sondern auch die praktische Realisierbarkeit eines Abtastgitters erhöhen lässt.

6.5.1 Approximation von Schallfeldern

A Bewertungsmaße

Zur Beurteilung von Mikrofonarrays können unterschiedliche Bewertungsmaße herangezogen werden. Der Gewinn für inkohärentes Rauschen (WNG), die Richtwirkung (DI) und die Konditionszahl der Schätzung der sphärischen Fourierkoeffizienten wurden bereits in Kap. 6.2.1 vorgestellt. Chardon et al. (2014b, 2015) verwenden den bei der Interpolation eines Schallfeldes in einem bestimmten Gebiet Ω (hier innerhalb des von einem Array umschlossenen Volumen) entstehenden Fehler, um die Qualität unterschiedlicher Abtastgitter zu beurteilen. Dies lässt sich folgendermaßen begründen: (i) der Interpolationsfehler lässt sich relativ einfach schätzen, (ii) es existieren zahlreiche theoretische Lösungen für unterschiedliche Gebiete, die den Entwurfsprozess vereinfachen, und (iii) aus dem Interpolationsfehler lassen sich die wichtigsten Bewertungsmaße ableiten. Zudem gilt: Wird das Schallfeld innerhalb des Arrayvolumens mit einem zu vernachlässigenden Interpolationsfehler approximiert, kann jede beliebige Arraygeometrie innerhalb desselben Volumens simuliert werden.

Der Interpolationsfehler ist definiert als die L_2 Norm zwischen dem tatsächlichen Schallfeld p und dem interpolierten Schallfeld $\hat{p}$:

$$\|p - \hat{p}\| = \int_{\Omega} |p - \hat{p}|^2 d\Omega. \tag{6.119}$$

Bei einer endlichen Anzahl an Messungen, wird das Integral mittels geeigneter Quadratur numerisch berechnet. Der Schalldruck $\hat{p}$ in dem Gebiet Ω wird mit Kugelflächenfunktionen approximiert, wobei die Koeffizienten über ein *Least Squares* Verfahren geschätzt werden.

B Approximation von Schallfeldern

Moiola et al. (2011) zeigen, dass sich die Lösungen u der Helmholtz-Wellengleichung

$$\Delta u + k^2 u = 0$$

in einem konvexen, sternförmigen Gebiet Ω über eine lineare Kombination von Kugelflächenfunktionen approximieren lassen:[139]

$$u \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}^{N} j_n(kr) Y_n^m(\theta, \phi), \tag{6.120}$$

wobei j_n die n-te sphärische Bessel-Funktion der ersten Art und Y_n^m die Kugelflächenfunktionen bezeichnet (vgl. Anhang B). Die Lösungen können auch über eine lineare Kombination ebener Wellen angenähert werden (vgl. Abschnitt 5.1.1):

$$u \approx \sum_{j=1}^{J} \beta_j^J \exp(i\mathbf{k}_j \cdot \mathbf{x}), \tag{6.121}$$

wobei der Wellenzahlvektor $\mathbf{k}_j$ auf einer Kugel mit Radius k abgetastet wird. Die Konvergenzgeschwindigkeit der Approximation hängt von der Glattheit von u ab (vgl. Ruzicka, 2004, Kap. A.9). Melenk (1999) zeigt, dass u in den meisten praktischen Anwendungen glatt ist und die Lösungen exponentiell konvergieren. Die Koeffizienten α_{nm}^N and β_j^J hängen von der Ordnung N und J der jeweiligen Approximation ab.

6.5.2 Stabilität der Interpolation von Schallfeldern

Betrachten wir nun ein Schallfeld p, welches mit einer endlichen Anzahl an Abtastpunkten in einem Gebiet Ω gemessen wurde. Dieses kann über die Koeffizienten des approximierten Schallfeldes $\hat{p}$, die über eine *Least Squares*-Schätzung bestimmt werden, interpoliert werden. Die Stabilität der Interpolation ist dann garantiert, wenn der Interpolationsfehler $\|p - \hat{p}\|$ in derselben Größenordnung liegt, wie der kleinste Approximationsfehler $\|p - \tilde{p}\|$, wobei $\tilde{p}$ die beste Approximation von p in dem Gebiet Ω ist. Im Allgemeinen reicht es nicht aus, mehr Abtastpunkte als Freiheitsgrade der Approximation zu haben (vgl. Runge, 1901).[140] Wird das Schallfeld

[139] Ein Gebiet Ω heißt konvex, wenn die Verbindungslinie zwischen zwei beliebigen Punkten aus Ω an keiner Stelle außerhalb von Ω verläuft. Ein Gebiet heißt sternförmig, wenn es zumindest einen Punkt O gibt, dessen Verbindungslinien zu allen anderen Punkten aus Ω an keiner Stelle außerhalb von Ω verlaufen.

[140] Das Phänomen, dass sich die Approximation mit einer steigenden Anzahl an Stützstellen verschlechtert, wird als Runge-Phänomen bezeichnet.

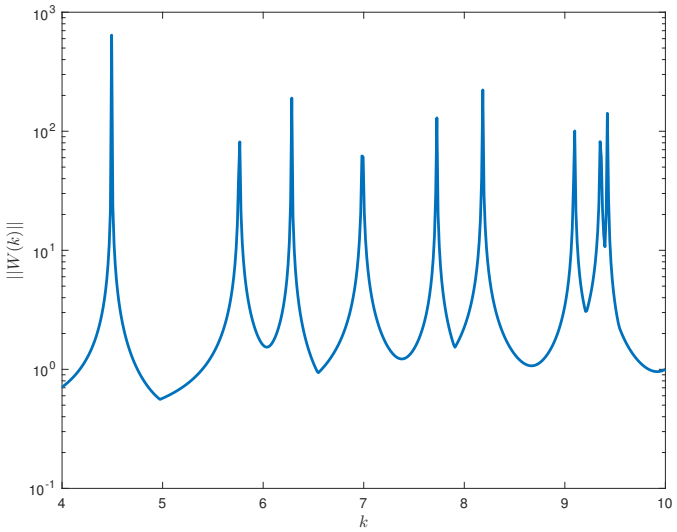


Abbildung 6.24: Norm des Operators $W_k : L_2(\partial\Omega) \to L_2(\Omega)$ für die Einheitskugel. In der Nähe der Eigenfrequenzen des Volumens mit Dirichlet Randbedingung geht die Norm gegen unendlich und die Interpolation des Schallfeldes wird instabil.

zum Beispiel mit Kugelflächenfunktionen approximiert, kann auch bei $Q > (N+1)^2$ Abtastpunkten nicht garantiert werden, dass die Interpolation stabil ist.

Meist wird das Schallfeld am Rand $\partial\Omega$ des Gebiets Ω abgetastet und dann innerhalb von Ω interpoliert.[141] Dabei können bei einigen Frequenzen Instabilitäten auftreten. Dieses Verhalten lässt sich wie folgt formalisieren: Sei $W_k : L_2(\partial\Omega) \to L_2(\Omega)$ ein Operator, der die auf $\partial\Omega$ definierten Funktionen (bei einer bestimmten Wellenzahl k) auf die Lösungen der Helmholtz-Wellengleichung innerhalb von Ω abbildet. Die Norm von W_k lässt sich schätzen und ist für eine Einheitskugel in Abb. 6.24 dargestellt. Nahe der Eigenfrequenzen des Volumens mit einer Dirichlet Randbedingung, strebt die Norm gegen unendlich. Daraus folgt, dass kleine Fehler auf der Oberfläche $\partial\Omega$ (z. B. durch Eigenrauschen der Sensoren oder durch Interpolationsfehler) im Inneren des Gebiets Ω verstärkt werden. Diese Unstetigkeiten hängen ausschließlich von der Form des Gebiets ab und sind unabhängig vom Abtastgitter und dem verwendeten Approximationsverfahren.

In der Literatur existieren zahlreiche Methoden dem Stabilitätsproblem zu begegnen. Besteht ein Array zum Beispiel aus zwei konzentrischen Kugelschalen mit unterschiedlichen Radien, lässt sich das Schallfeld bei einer bestimmten Wellen-

[141] Hier entspricht das von einem Array umschlossene Volumen dem Gebiet Ω der Interpolation.

zahl k meist mit zumindest einer Kugelschale interpolieren (vgl. Balmages und Rafaely, 2007; Parthy et al., 2009; Jin et al., 2014). Allerdings wird hierbei, im Vergleich zu einem einfachen Array, die doppelte Anzahl an Abtastpunkten benötigt. Rafaely (2008) verwendet einige Abtastpunkte innerhalb der Kugel, um eine stabile Interpolation zu gewährleisten. Das dabei verwendete Optimierungsverfahren zur Bestimmung der Abtastpunkte ist jedoch sehr rechenaufwändig und auf die Kugel begrenzt. Weitere Ansätze finden sich in der Einleitung zu diesem Kapitel.

Im folgenden Abschnitt wird ein neuer Ansatz zur räumlichen Abtastung des Schallfeldes vorgestellt. Dieser schätzt die minimale Anzahl an Abtastpunkten, mit denen eine stabile Interpolation bei den Eigenfrequenzen der Kugel gewährleistet ist. Die Schätzung wird etwas später auf allgemeinere Gebiete (wie z. B. Ellipsoid und Kubus) erweitert. Der Ansatz beruht auf dem von Cohen et al. (2013) vorgestellten Kriterium zur stabilen *Least Squares*-Approximation einer an zufällig verteilten Punkten abgetasteten Funktion.

Sei ϱ die Wahrscheinlichkeitsdichte, mit der die Abtastpunkte in einem Gebiet Ω verteilt sind, und (e_i) eine orthogonale Basis (in Bezug auf ϱ) eines endlich-dimensionalen Eigenraums E_λ der Dimension λ, in dem das Schallfeld p approximiert werden soll,[142] dann lässt sich folgendes allgemeine Maß für die Anzahl an benötigten Abtastpunkten berechnen:

$$K(\lambda) = \max_{x \in \Omega} \sum_{j=1}^{\lambda} |e_j(x)|^2 \,. \tag{6.122}$$

Gilt für die Anzahl Q der Abtastpunkte, dass $K(\lambda) < \kappa Q / \log Q$ (wobei κ eine Konstante ist, vgl. Cohen et al., 2013, Gl. 1.4), haben der Schätzfehler und der Approximationsfehler dieselbe Größenordnung (vgl. Cohen et al., 2013, Theorem 2). Demzufolge kann mit $K(\lambda)$ Abtastwerten eine ausreichend stabile Interpolation gewährleistet werden. Der Wert $K(\lambda)$ eines Abtastgitters soll möglichst nahe an der unteren Schranke λ liegen (z. B. für eine Kugel möglichst nahe an $\lambda = (N+1)^2$). In den folgenden Abschnitten wird $K(\lambda)$ für unterschiedliche Abtastgitter und Gebiete Ω numerisch berechnet.

[142] Ein Beispiel ist der von den Kugelflächenfunktionen der Ordnung N, mit $\lambda = (N+1)^2$ aufgespannte Raum.

A Sphärische Arrays

Es kann gezeigt werden, dass wenn alle Abtastpunkte gleichmäßig innerhalb der Kugel verteilt sind, für eine stabile Interpolation mindestens $Q = N^{3/2}$ Abtastpunkte benötigt werden. Befindet sich ein Teil der Abtastpunkte auf der Kugeloberfläche, sind mindestens $Q = N^2/\alpha$ Abtastpunkte erforderlich, wobei $\alpha \in]0{,}1[$ das Verhältnis der Anzahl an Abtastpunkten auf der Kugel zu den Abtastpunkten innerhalb der Kugel beschreibt (vgl. Chardon et al., 2013, 2014a). Abbildung 6.25 zeigt $K(\lambda)$ für folgende Verteilungen bei $kr = 3$:

- gleichverteilt über dem Winkel bei konstantem Radius,
- gleichverteilt innerhab der Kugel,
- gemischte Verteilung (in der Abbildung mit $\alpha = 0.8$).

Daraus ist ersichtlich, dass eine gemischte Verteilung, bei der die meisten Abtastpunkte auf der Oberfläche $\partial\Omega$ des Gebiets Ω (in diesem Fall der Kugel) liegen, am effizientesten ist. Für die gemischte Verteilung $\alpha = 0.8$ sind die Werte von $K(\lambda)$ auch bei höheren Ordnungen nahe an der unteren Schranke $\lambda = (N+1)^2$. Ganz allgemein gilt, liegen mehr Abtastpunkte am Rand des Gebiets als im Inneren, genügen weniger Abtastpunkte, um eine stabile Interpolation zu erreichen. Zudem ist zu beachten, dass bei der Approximation mit Kugelflächenfunktionen $N > kr$ (vgl. Gl. 6.25) sein soll, um Aliasfehler zu vermeiden.

B Nicht sphärische Arrays

Dieser Abschnitt behandelt die Interpolation eines Schallfeldes mit nicht sphärischen Arrays (am Beispiel eines Ellipsoids und eines Kubus). Das Hauptaugenmerk liegt wiederum auf einer gemischten Verteilung, bei der die Mehrzahl der Abtastpunkte auf dem Rand $\partial\Omega$ des Gebiets liegen. Dadurch kommt der Abtastung auf dem Rand des Gebiets eine besondere Bedeutung zu. Im Inneren des Gebiets sollten die Abtastpunkte möglichst gleichverteilt sein. Folgende Abtastgitter wurden simuliert:

- gleichverteilt auf dem Rand $\partial\Omega$ des Gebiets,
- gleichverteilt auf der Kugel und Projektion auf den Rand $\partial\Omega$ des Gebiets (vgl. Alon und Rafaely, 2012),

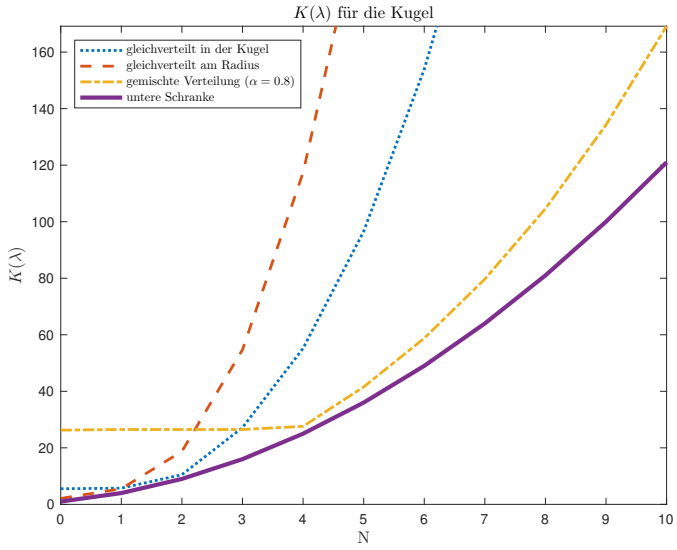


Abbildung 6.25: Minimale Anzahl an Abtastpunkten $K(\lambda)$ für eine Kugel in Abhängigkeit von der Ordnung N der zur Approximation des Schallfeldes verwendeten Kugelflächenfunktionen.

- spezielle Verteilung, die an das jeweilige Gebiet angepasst wird.

Es gibt unterschiedliche Möglichkeiten die Abtastpunkte auf einem Ellipsoid zu verteilen. Nehmen wir an, dass die Abtastpunkte auf der Oberfläche einer zentrierten Kugel optimal verteilt sind. Dann kann die Kugel so weit in Richtung der Halbachsen gestreckt werden, bis ihre Form der des Ellipsoids entspricht. Diese Verteilung ergibt eine höhere Dichte an Abtastpunkten je weiter ein Flächenabschnitt vom Zentrum entfernt ist. Eine weitere Möglichkeit besteht darin, die Abtastpunkte auf der Kugeloberfläche in radialer Richtung auf die Oberfläche des Ellipsoids zu projizieren. Dadurch erhöht sich die Dichte an Abtastpunkten auf Flächenabschnitten die näher am Zentrum sind. Es kann gezeigt werden, dass die radiale Projektion suboptimal ist. Für den Kubus ergibt sich für die Dichte ρ der Abtastpunkte auf der Seitenfläche in der xy-Ebene folgender Ausdruck:

$$\rho = \frac{1}{Z} \frac{1}{\sqrt{1-x^2}\sqrt{1-y^2}}, \tag{6.123}$$

wobei Z eine Konstante ist, die ρ auf auf 1 normiert. Für die anderen Seitenflächen lassen sich sehr einfach ähnliche Ausdrücke herleiten.

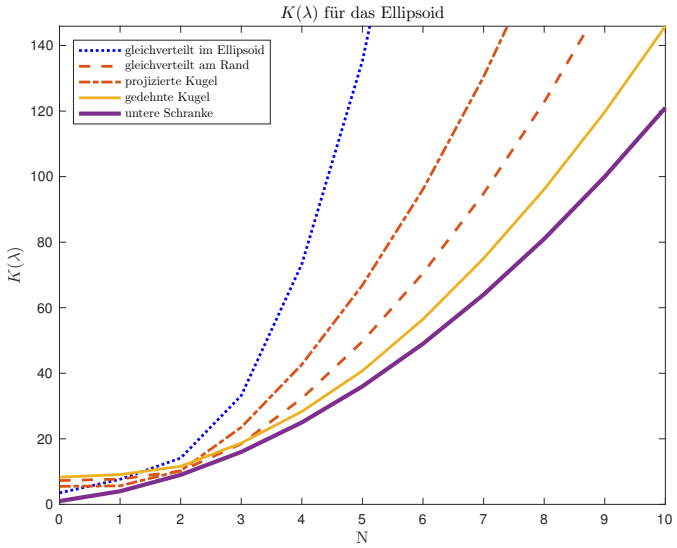


Abbildung 6.26: Minimale Anzahl an Abtastpunkten $K(\lambda)$ für ein abgeplattetes Rotationsellipsoid (mit einem Seitenverhältnis von 0,5) in Abhängigkeit von der Ordnung N der zur Approximation des Schallfeldes verwendeten Kugelflächenfunktionen.

Die Abbildungen 6.26 und 6.27 zeigen $K(\lambda)$ für unterschiedliche Abtastgitter in Abhängigkeit von der Ordnung N der zur Approximation des Schallfelds verwenden Kugelflächenfunktionen. Die Simulationen wurden für Ellipsoid mit einem Seitenverhältnis von 0,5 und einen Kubus durchgeführt. Das Schallfeld wird mit unterschiedlichen Abtastgittern auf der Oberfläche abgetastet. Zusätzlich wurde eine gleichverteilte Abtastung innerhalb des Volumens simuliert. Aus den Ergebnissen wird ersichtlich, dass beim Ellipsoid nur die Streckung des Abtastgitters einer Kugel (siehe oben) Werte nahe der unteren theoretischen Schranke liefert. Beim Kubus liefert eine nicht gleichverteilte Abtastung die besten Ergebnisse. In Kap. 6.5.3 wird gezeigt, wie mit der in dieser Arbeit vorgestellten Methode die Abtastgitter wesentlich besser auf das jeweilige Gebiet angepasst werden können.

C Einfluss der Basisfunktionen

$K(\lambda)$ hängt nicht nur vom Abtastgitter, sondern auch von den Basisfunktionen ab, mit denen das Schallfeld approximiert wird. Numerische Simulationen, die für eine Vielzahl unterschiedlicher Basisfunktionen durchgeführt wurden, zeigen jedoch,

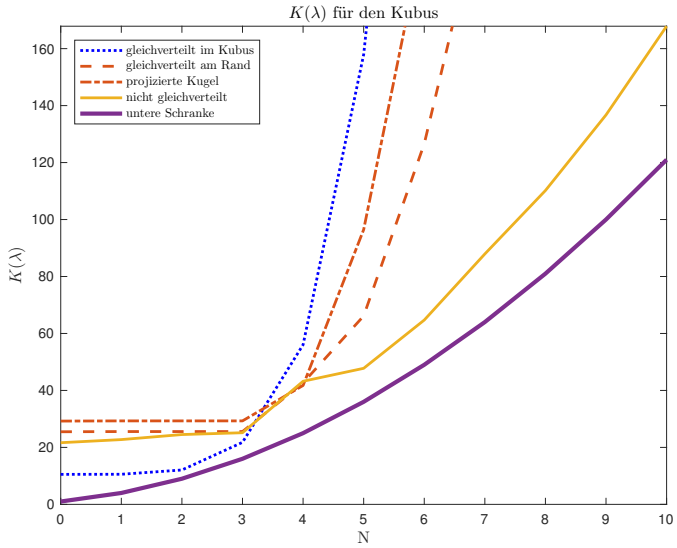


Abbildung 6.27: Minimale Anzahl an Abtastpunkten $K(\lambda)$ für einen Kubus in Abhängigkeit von der Ordnung N der zur Approximation des Schallfeldes verwendeten Kugelflächenfunktionen.

dass $K(\lambda)$ für große λ unabhängig von den Basisfunktionen ist. Dieses Verhalten wird in Abb. 6.28 gezeigt. Die Simulation wurden für ein im Ursprung zentriertes Ellipsoid durchgeführt. Als Basisfunktionen wurden (i) im Ursprung zentrierte Kugelflächenfunktionen, (ii) im Punkt (1,1,1) zentrierte Kugelflächenfunktionen, (iii) ebene Wellen und (iv) fundamentale Lösungen verwendet. Das Abtastgitter auf dem Rand $\partial\Omega$ des Ellipsoids wurde durch Strecken einer Kugel erzeugt, deren Oberfläche mit einem Hyperinterpolations-Gitter (vgl. Sloan und Womersley, 1998; Womersley und Sloan, 2001) abgetastet wird. Das Verhältnis der Anzahl an Abtastpunkten auf der Oberfläche des Ellipsoids, zu den Abtastpunkten innerhalb des Ellipsoids, beträgt $\alpha = 0{,}8$. Die Wellenzahlvektoren der ebenen Wellen werden mit einem Hyperinterpolations-Gitter auf einer Kugel mit Radius $k = 3$ abgetastet. Die Quellen der fundamentalen Lösungen liegen auf einer Kugel mit Radius $k = 2$. Das Abtastgitter entspricht auch hier der Hyperinterpolation. Die Ergebnisse zeigen, dass für alle Basisfunktionen die Werte von $K(\lambda)$ nahe an der unteren Schranke liegen. Aus diesem Grund werden für die weiteren Studie ausschließlich Kugelfunktionen verwendet. Um eine allgemein gültige Aussage treffen zu können, müsste allerdings theoretisch bewiesen werden, dass $K(\lambda)$ für große λ unabhängig von den Basisfunktionen ist.

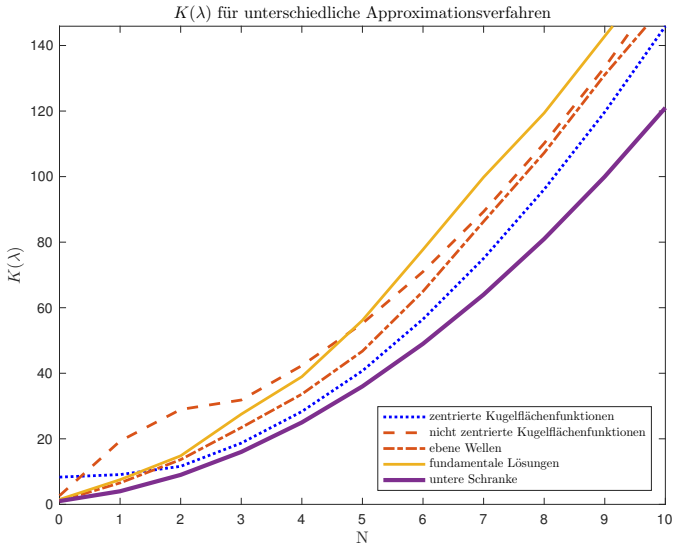


Abbildung 6.28: Minimale Anzahl an Abtastpunkten $K(\lambda)$ für unterschiedliche Approximationsverfahren in Abhängigkeit von der Ordnung N der Approximation, dargestellt für ein Ellipsoid mit einer durch Strecken einer Kugel bestimmten Verteilung der Abtastpunkte auf $\partial\Omega$.

Abschließend lässt sich Folgendes feststellen: Um eine stabile Interpolation des Schallfeldes mit einer minimalen Anzahl an Abtastpunkten zu gewährleisten, sollten (i) nur einige wenige Abtastpunkte innerhalb des Gebiets Ω liegen, (ii) am Rand $\partial\Omega$ des Gebiets Ω eine gut konditionierte Abtastung erfolgen und (iii) die Projektion einer gleichförmigen Abtastung der Kugeloberfläche auf den Rand des $\partial\Omega$ des Gebiets Ω vermieden werden, da diese in der Regel ineffizient ist.

6.5.3 Entwurf eines optimalen modalen Mikrofonarrays

Der Entwurf eines robusten offenen Mikrofonarrays erfordert (i) die Bestimmung eines optimalen Abtastgitters auf dem Rand des betrachteten Gebiets (d. h. des von dem Array umschlossenen Volumens) und (ii) die Festlegung der Abtastpunkte im Inneren des Gebiets, um die Interpolation an den Eigenfrequenzen des Gebiets zu stabilisieren.

A Abtastpunkte am Rand des Gebiets

Um eine stabile Interpolation mit möglichst wenigen Abtastpunkten zu gewährleisten, muss am Rand $\partial\Omega$ des betrachteten Gebiets Ω ein möglichst optimales Abtastgitter verwendet werden. Für die Kugel und das Ellipsoid finden sich in der Literatur zahlreiche optimale Lösungen. In dieser Arbeit wird für die Kugel die Hyperinterpolation nach Sloan und Womersley (1998; 2001) verwendet, die einer kritischen räumlichen Abtastung auf der Kugeloberfläche entspricht. Das heißt, es werden nur so viele Abtastpunkte wie Kugelflächenfunktionen benötigt. Durch Strecken der Kugel, lässt sich das optimale Abtastgitter auf die Oberfläche eines Ellipsoids projizieren.[143]

Die von Maday et al. (2009) vorgeschlagenen „magischen Punkte“ (*magic points method*) können auf allgemeine Oberflächen angewendet werden. Mit dieser Methode lassen sich λ Abtastpunkte bestimmen, die eine stabile Interpolation von Funktionen in einem Vektorraum der Dimension λ gewährleisten. Dabei werden die optimalen Abtastpunkte über ein iteratives Verfahren bestimmt. Sei (e_i) eine Menge von λ Funktionen und $\tilde{e}_j$ die Linearkombination der ersten j Funktionen e_i, die eine Nullstelle bei den ersten $(j-1)$ Punkten x_i aufweisen, dann gilt für den j-ten Abtastpunkt:

$$x_j = \underset{x\in\partial\Omega}{\operatorname{argmax}} \, \tilde{e}_j. \tag{6.124}$$

Ein Mikrofonarray ist dann robust, wenn die Interpolation über einen möglichst großen Wellenzahlbereich stabil ist. Aus diesem Grund wird in dieser Arbeit eine modifizierte Methode zur Bestimmung der magischen Punkte verwendet. Sei (e_i^n) eine Menge von Funktionen, die einer Wellenzahl k_n zugeordnet werden kann,[144] dann können die optimalen Abtastpunkte wie folgt bestimmt werden:

$$x_j = \underset{x\in\partial\Omega}{\operatorname{argmax}} \min_n \tilde{e}_j^n, \tag{6.125}$$

wobei $\tilde{e}_j^n$ wie in Gl. (6.124) definiert ist.

[143] Wie im vorhergehenden Abschnitt gezeigt wurde, ist die direkte (radiale) Projektion der Abtastpunkte auf der Kugel auf die Oberfläche des Ellipsoids suboptimal.

[144] Ein gutes Beispiel ist die Menge der Funktionen $j_n(k_n r)Y_n^m(\theta,\phi)$ für die Wellenzahlen k_n und Ordnungen $0 \le n \le N$.

B Abtastpunkte innerhalb des Gebiets

Dieser Abschnitt stellt eine Methode zur Bestimmung der innerhalb des betrachteten Gebiets liegenden Abtastpunkte vor. Dabei muss vermieden werden, dass Abtastpunkte auf den Knotenpunkten eines bestimmten Eigenmodus liegen, da diese die Interpolation bei der zugehörigen Eigenfrequenz nicht stabilisieren.

Im einfachsten Fall sind die Eigenfrequenzen nicht degeneriert.[145] Wird ein Eigenmodus an einem Punkt abgetastet, wo er eine maximale Amplitude hat, kann dieser möglichst stabil geschätzt werden. Für einen breiteren Frequenzbereich mit mehreren Eigenmoden ist es jedoch sehr unwahrscheinlich, dass diese an der gleichen Stelle ein Maximum aufweisen. Aus diesem Grund wird jener Abtastpunkt bestimmt, bei dem die kleinste Amplitude der Moden p_n im betrachteten Gebiet Ω ein Maximum aufweist:

$$x = \underset{x\in\Omega}{\operatorname{argmax}} \min_{n} |p_n(x)|. \tag{6.126}$$

Der so bestimmte Abtastpunkt x befindet sich in keinem Knotenpunkt, d. h. die Eigenmoden sind in diesem Punkt garantiert nicht Null. Somit lassen sich alle Eigenmoden in diesem Punkt schätzen.

Für die Kugel sind die meisten Eigenfrequenzen degeneriert. In diesem speziellen Fall ist es unmöglich, die Eigenmoden in nur einem Abtastpunkt zu bestimmen. Es werden mindestens $Q = \Lambda$ Mikrophone benötigt, wobei Λ die größte Vielfachheit der Eigenfrequenzen ist. Die Λ Abtastpunkte lassen sich wiederum iterativ bestimmen.

Der erste Punkt wird wie folgt bestimmt:

$$x_1 = \underset{x\in B}{\operatorname{argmax}} \min_{j} \max_{\substack{p\in E_j \\ \|p\|=1}} |p(x)|, \tag{6.127}$$

wobei E_n den zu einer Eigenfrequenz zugehörigen Eigenraum bezeichnet. Der nächste Punkt lässt sich über die Unterräume E_j^1 mit einer Dimension größer als 1 bestimmen, in denen die Eigenmoden im Punkt x_1 einen Knotenpunkt haben

[145] Haben zwei Eigenfunktionen die gleiche Eigenfrequenz, bezeichnet man sie (oder auch diese Frequenz) als entartet.

(d. h. von diesem nicht abgetastet werden können):

$$x_2 = \underset{x\in B}{\operatorname{argmax}} \min_j \max_{\substack{p\in E_j \\ \|p\|=1 \\ p(x_1)=0}} |p(x)|. \tag{6.128}$$

Der n-te Abtastpunkt wird über die Unterräume E_j^{n-1} mit einer Dimension größer als $(n-1)$ bestimmt, in denen die Eigenmoden an den zuvor bestimmten $(n-1)$ Abtastpunkten Null sind. Demzufolge lassen sich die optimalen Abtastpunkte im inneren des Gebiets Ω wie folgt bestimmen;

$$x_i = \underset{x\in B}{\operatorname{argmax}} \min_j \max_{\substack{p\in E_j \\ \|p\|=1 \\ (p(x_j)=0)_{0<j<i}}} |p(x)|. \tag{6.129}$$

Wird nur ein Eigenraum betrachtet, kann das iterative Verfahren in Gl. (6.129) als Variante der magischen Punkte interpretiert werden. Lassen sich die Eigenräume für ein bestimmtes Gebiet nicht einfach berechnen, kann als Alternative die Methode der magischen Punkte (vgl. Gl. 6.125) auf das gesamte Gebiet Ω angewendet werden. Die magischen Punkte lassen sich sehr einfach implementieren und sind aus diesem Grund für viele praktische Anwendungen interessant (vgl. Abschnitt 6.5.5). Nachteilig ist, dass sich mit der Methode der magischen Punkte die minimale Anzahl an benötigten inneren Abtastpunkten nicht bestimmen lässt.

6.5.4 Numerische Simulationen

Dieser Abschnitt diskutiert die Resultate numerischer Simulationen für unterschiedliche Abtastgitter hinsichtlich der Qualität und der Robustheit der Interpolation des Schallfelds innerhalb des betrachteten Gebiets. Als Bewertungsmaß wird der relative Interpolationsfehler ϵ herangezogen:

$$\epsilon = \frac{\|p - \hat{p}\|_{L_2}}{\|p\|_{L_2}}, \tag{6.130}$$

d. h. der auf die Gesamtenergie des Schallfelds innerhalb des Gebiets Ω normierte Fehler zwischen dem tatsächlichen Schallfeld p und dem interpolierten Schallfeld $\hat{p}$ (vgl. Gl. 6.119).

Für die numerischen Simulationen wurde eine Vielzahl ebener Wellen aus unterschiedlichen Raumrichtungen berechnet. Die Robustheit der unterschiedlichen Abtastgitter wird durch einem dem Signal überlagerten Rauschen (SNR 40 dB) getestet. Bei jeder Frequenz wurden 40 Realisierungen des Rauschens und der Einfallsrichtungen der ebenen Wellen berechnet.

A Sphärische Arrays

In diesem Abschnitt wird der relative Interpolationsfehler für eine offene Kugel ($N = 9$) und verschiedene Abtastgitter berechnet. Die Basis für die Approximation des Schallfeldes bilden die mit den sphärischen Bessel-Funktionen multiplizierten Kugelflächenfunktionen (vgl. Gl. 6.120). Abbildung 6.29 zeigt den relativen Interpolationsfehler für folgende Abtastgitter und $k \in]0{,}5; 9[$:

- *open sphere*: Offenes Kugelarray mit Hyperinterpolation auf der Kugel ($N = 9$, $Q = 100$). Keine Abtastpunkte innerhalb der Kugel.
- *open sphere + interior*: Offenes Kugelarray mit Hyperinterpolation auf der Kugel und 9 zusätzlichen Abtastpunkten innerhalb der Kugel ($N = 9$, $Q = 109$). Die 9 zusätzlichen Abtastpunkte wurden über den in Abschnitt 6.5.3-B vorgestellten Ansatz bestimmt.
- *open sphere + Rafaely*: Offenes Kugelarray mit Hyperinterpolation auf der Kugel und 9 zusätzlichen Abtastpunkten innerhalb der Kugel ($N = 9$, $Q = 109$). Die 9 zusätzlichen Abtastpunkte wurden über die von Rafaely (2008) vorgeschlagene Methode bestimmt.
- *magic sphere + interior*: Offenes Kugelarray mit magischen Punkten auf der Kugel und 9 zusätzlichen Abtastpunkten innerhalb der Kugel ($N = 9$, $Q = 109$). Die 9 zusätzlichen Abtastpunkte wurden über den in Abschnitt 6.5.3-B vorgestellten Ansatz bestimmt.
- *magic ball*: Magische Punkte über die gesamte Kugel ($N = 10$, $Q = 121$). Da die magischen Punkte bei einer Ordnung $N = 9$ ($Q = 100$) bereits bei

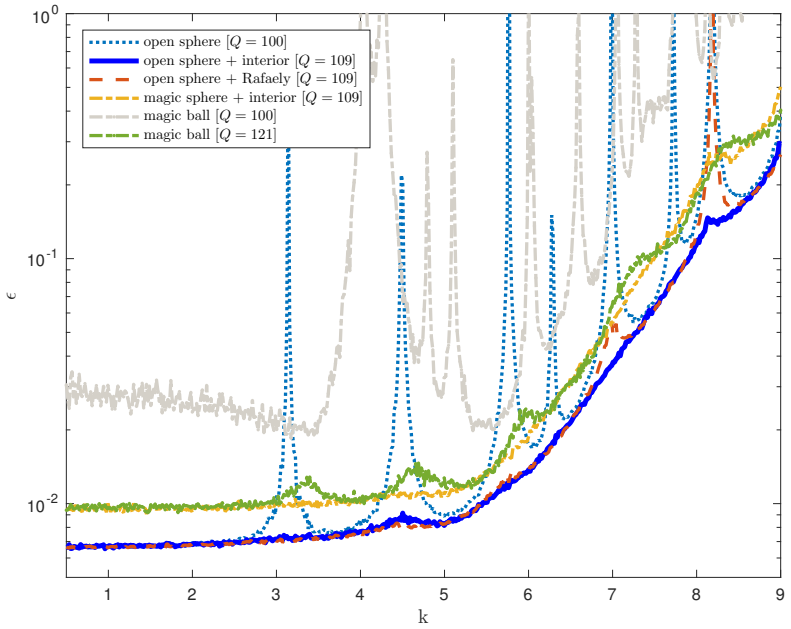


Abbildung 6.29: Relativer Interpolationsfehler ϵ eines offenen sphärischen Arrays in Abhängigkeit vom Abtastgitter (mit Abtastpunkten auf und innerhalb der Kugel) und der Wellenzahl k.

sehr kleinen k starke Resonanzen auftreten, wurde die Ordnung auf $N = 10$ erhöht.

Abb. 6.29 zeigt, dass sich die Stabilität der Interpolation in der Umgebung der Nullstellen der Bessel-Funktionen mit ein paar wenigen Abtastpunkten innerhalb der Kugel verbessern lässt. Der in dieser Arbeit vorgestellte Ansatz zur Bestimmung der inneren Abtastpunkte unterdrückt die Resonanzen bei den Eigenfrequenzen fast vollständig und liefert demzufolge den kleinsten Interpolationsfehler. Der Interpolationsfehler für $k < 5$ ist auf das Messrauschen zurückzuführen, für $k > 5$ begrenzt der Fehler der Approximation des Schallfeldes mit Kugelflächenfunktionen der Ordnung $N = 9$ den Interpolationsfehler.

Zudem ist ersichtlich, dass der von Rafaely (2008) vorgeschlagene Ansatz zur Bestimmung der inneren Abtastpunkte die Resonanz bei $k = 8{,}18$ nicht unterdrückt. Dies lässt sich voraussichtlich auf die gleichverteilte radiale Abtastung des Inneren der Kugel zurückführen. Die magischen Punkte zeigen ein ähnliches Verhalten wie das hier vorgeschlagene optimale Verfahren, weisen jedoch insgesamt einen höheren Interpolationsfehler auf. Lassen sich die Eigenfrequenzen eines

Gebiets nicht einfach bestimmen, sind die magischen Punkte eine gute Alternative um ein optimales Abtastgitter zu berechnen.

B Ellipsoidische Arrays

In diesem Abschnitt wird der relative Interpolationsfehler für ein offenes Ellipsoid (mit Halbachsen $a = 1{,}0$, $b = 0{,}8$ und $c = 0{,}5$) und unterschiedliche Abtastgitter berechnet. Dabei werden die Eigenfrequenzen des Ellipsoids über die Methode von Barnett (2000) und Betcke und Trefethen (2006) bestimmt. Ähnlich dem sphärischen Array, wird auf der Oberfläche ein Abtastgitter der Ordnung $N = 9$ mit $Q = 100$ Abtastpunkten verwendet. Allerdings wird nur ein zusätzlicher Abtastpunkt im Inneren des Ellipsoids benötigt, um die Interpolation bei den Eigenfrequenzen zu stabilisieren. Abbildung 6.30 zeigt den relativen Interpolationsfehler für folgende Abtastgitter und $k \in]0{,}5; 9[$:

- *stretched*: Offenes Ellipsoid mit einem durch Strecken einer Kugel erzeugtem Hyperinterpolations-Gitter ($N = 9$, $Q = 100$). Keine Abtastpunkte innerhalb des Ellipsoids.
- *stretched + interior*: Offenes Ellipsoid mit einem durch Strecken einer Kugel erzeugtem Hyperinterpolations-Gitter und einem zusätzlichen Abtastpunkt innerhalb des Ellipsoids ($N = 9$, $Q = 101$). Der zusätzliche Abtastpunkt wurde über den in Abschnitt 6.5.3-B vorgestellten Ansatz bestimmt.
- *projected + interior*: Offenes Ellipsoid mit einem durch radiale Projektion erzeugtem Hyperinterpolations-Gitter und einem zusätzlichen Abtastpunkt innerhalb des Ellipsoids ($N = 9$, $Q = 101$). Der zusätzliche Abtastpunkt wurden über den in Abschnitt 6.5.3-B vorgestellten Ansatz bestimmt.
- *magic + interior*: Offenes Ellipsoid mit magischen Punkten einem zusätzlichen Abtastpunkt innerhalb des Ellipsoids ($N = 9$, $Q = 101$). Der zusätzliche Abtastpunkt wurden über den in Abschnitt 6.5.3-B vorgestellten Ansatz bestimmt.
- *magic ellipsoid*: Magische Punkte über das gesamte Ellipsoid ($N = 10$, $Q = 121$). Da die magischen Punkte bei $N = 9$ bereits bei sehr kleinen k starke Resonanzen aufweisen, wurde die Ordnung auf $N = 10$ erhöht.

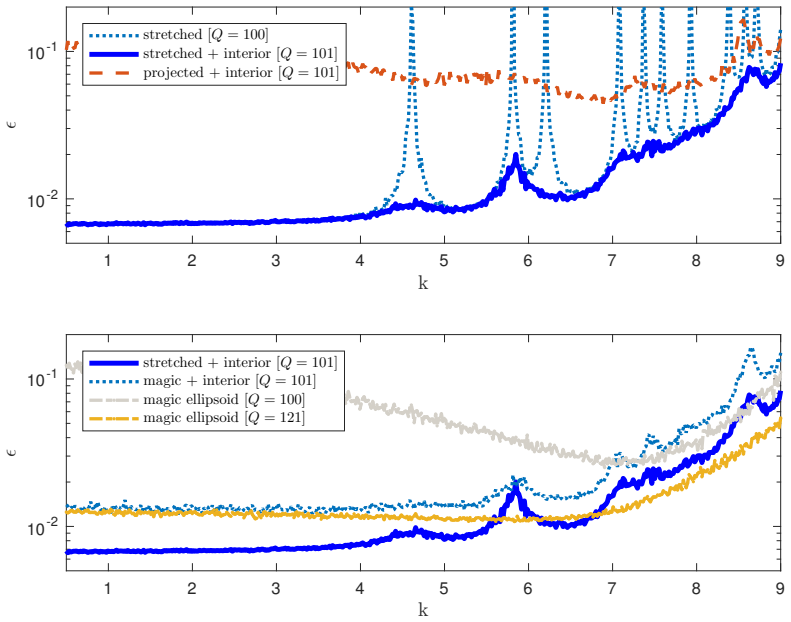


Abbildung 6.30: Relativer Interpolationsfehler ϵ eines offenen Ellipsoids (mit Halbachsen $a = 1.0, b = 0.8$, and $c = 0.5$) in Abhängigkeit vom Abtastgitter (mit Abtastpunkten auf und innerhalb des Ellipsoids) und der Wellenzahl k.

Nach Abbildung 6.30 verhalten sich die unterschiedlichen Abtastgitter sehr ähnlich der Kugel. Wird ausschließlich auf der Oberfläche abgetastet, ergeben sich bei den Eigenfrequenzen des Ellipsoids starke Resonanzen. Im Gegensatz zur Kugel lässt sich die Stabilität der Interpolation mit nur einem einzigen Abtastpunkt im Inneren wesentlich verbessern. Hierbei spielt die Art der Projektion, mit der ein optimales Abtastgitter von der Kugel auf die Oberfläche des Ellipsoids projiziert wird, eine wesentliche Rolle. Wird das Abtastgitter durch Strecken einer Kugel erzeugt, und ein zusätzlicher Abtastpunkt im Inneren hinzugefügt ($Q = 101$), entspricht der relative Interpolationsfehler dem der über das gesamte Volumen bestimmten magischen Punkte der Ordnung $N = 10$ ($Q = 121$). Das in dieser Arbeit vorgestellte Verfahren ist somit wesentlich effizienter hinsichtlich der Anzahl der notwendigen Abtastpunkte. Für generelle sternförmige Gebiete, bei denen die Eigenfrequenzen nicht berechnet werden können, sind die magischen Punkte wiederum eine gute Alternative.

6.5.5 Modifiziertes Optimierungsverfahren mit Nebenbedingungen

Sind die Abtastpunkte über das gesamte Volumen eines Arrays verteilt, lässt sich dieses in der Praxis meist schwer implementieren. Aus diesem Grund wird der Ansatz zur Berechnung optimaler Abtastpunkte dahingehend erweitert, dass Rahmenbedingungen (wie z. B. mechanische Beschränkungen) mit berücksichtigt werden können. In einem ersten Beispiel, werden die Abtastpunkte auf zwei konzentrische Kugelschalen (vgl. *double sphere array*) beschränkt. Dadurch entsteht eine sehr einfach zu implementierende Arraygeometrie. Ein zweites Beispiel beschränkt die optimalen Abtastpunkte auf einen Rotationstorus (vgl. *spindle torus array*).

A Double-Sphere-Array

Das klassische Double-Sphere-Array besteht aus zwei konzentrischen Kugelschalen mit unterschiedlichen Radien und gleichem Abtastgitter. Diese Methode ist sehr ineffizient, da mindestens $Q = 2(N+1)^2$ Mikrofone benötigt werden. Im Folgenden wird gezeigt, wie über den modifizierten Ansatz mit Nebenbedingungen die Anzahl der Abtastpunkte – unter Beibehaltung der sehr einfachen Arraygeometrie – reduziert werden kann. Die Berechnung der optimalen Abtastpunkte wird in zwei Schritten durchgeführt: (i) Berechnung des ersten Abtastpunktes mit Gl. (6.127). Dieser bestimmt den Radius ρ der inneren Kugelschale. (ii) Iterative Berechnung der weiteren Abtastpunkte x_i mit Gl. (6.129), mit der strikten Nebenbedingung, dass diese auf der inneren Kugelschale mit Radius ρ liegen.

Die Ergebnisse der Simulation für ein Double Sphere Array der Ordnung $N = 9$ sind in Abb. 6.31 dargestellt. Diese zeigt den relativen Interpolationsfehler für folgende Abtastgitter und $k \in [0{,}5; 9]$:

- *open sphere*: Offenes Kugelarray mit Hyperinterpolation auf der Kugel ($N = 9$, $Q = 100$). Keine Abtastpunkte innerhalb der Kugel.
- *open sphere + interior*: Offenes Kugelarray mit Hyperinterpolation auf der Kugel und 9 zusätzlichen Abtastpunkten innerhalb der Kugel ($N = 9$, $Q = 109$). Die 9 zusätzlichen Abtastpunkte wurden über den in Abschnitt 6.5.3-B vorgestellten Ansatz bestimmt.

- *double sphere, inner sphere optimized*: Double-Sphere-Array mit Hyperinterpolation auf der äußeren Kugelschale und 9 zusätzlichen Abtastpunkten auf der inneren Kugelschale mit Radius $\rho = 0{,}69$ ($N = 9$, $Q = 109$). Die 9 zusätzlichen Abtastpunkte wurden über den modifizierten Ansatz mit Zwangsbestimmungen bestimmt.
- *double sphere, inner sphere hi-2*: Double-Sphere-Array mit Hyperinterpolation auf der äußeren Kugelschale ($N = 9$, $Q = 100$) und der inneren Kugelschale ($N = 2$, $Q = 9$) mit Radius $\rho = 0{,}69$. Insgesamt $N = 9$ und $Q = 109$.
- *double sphere, inner sphere Balmages*: Double-Sphere-Array nach Balmages und Rafaely ($N = 9$, $Q = 109$) mit einem inneren Radius

$$\rho = 1/\left(1 + \frac{\pi}{2k_{\max}}\right) \approx 0{,}85$$

  (s. Balmages und Rafaely, 2007, Gl. 15).

Das mit dem optimierten Ansatz berechnete Double-Sphere-Array weist einen sehr kleinen Interpolationsfehler bei den Eigenfrequenzen der Kugel auf. Dieser ist nur geringfügig höher als jener des offenen Kugelarrays ohne Zwangsbedingungen. Werden die Abtastpunkte auf der inneren Kugelschale hingegen gleichverteilt (siehe „*double sphere, interior sphere hi-2*“), kann die Resonanz bei $k \approx 8{,}17$ nicht unterdrückt werden. Ein ähnliches Verhalten konnte bereits in Abschnitt 6.5.4-A für die von Rafaely (2008) vorgeschlagene Methode zur Optimierung eines offenen Kugelarrays festgestellt werden. Dies lässt darauf schließen, dass die annähernde Gleichverteilung für die schlechte Unterdrückung dieser Resonanz verantwortlich ist.

Nun wird der modifizierte Ansatz mit Nebenbedingungen auf das von Jin et al. (2014) vorgeschlagene Double-Sphere-Array angewendet. Dieses besteht aus einer schallharten inneren Kugel und einer äußeren Kugelschale. Das Gebiet Ω wird durch die beiden Kugeln begrenzt, wobei auf der äußeren Kugelschale S_o Dirichlet-Randbedingungen und auf der inneren Kugel S_i Neumann-Randbedingungen gelten. Da die Eigenmoden die Neumann-Randbedingungen auf S_i erfüllen, sind sie

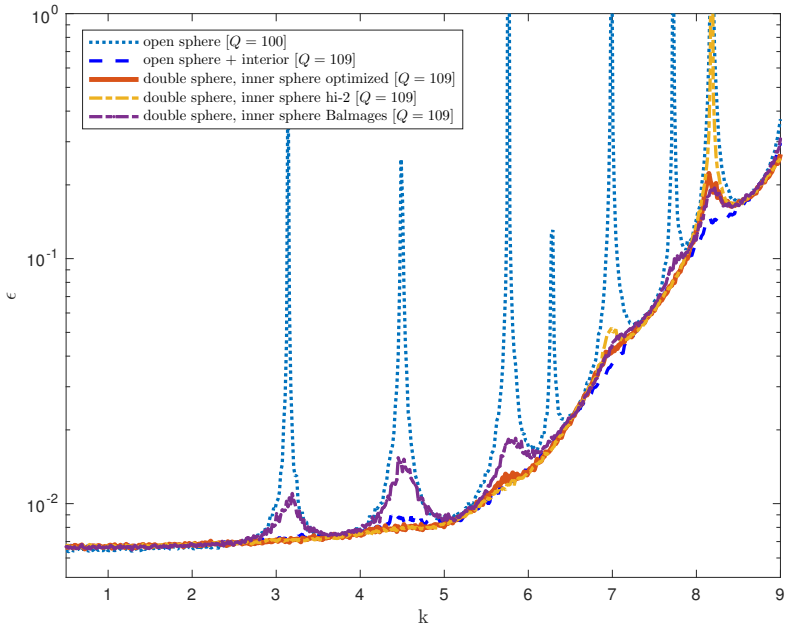


Abbildung 6.31: Relativer Interpolationsfehler ϵ eines offenen Double-Sphere-Arrays in Abhängigkeit vom Abtastgitter und der Wellenzahl k.

fast überall auf der Kugel ungleich Null. Demzufolge lässt sich mit Druckempfängern auf S_i die Interpolation bei den Eigenfrequenzen stabilisieren. Die inneren Abtastpunkte werden mit Gl. (6.129) bestimmt, unter der Nebenbedingung, dass diese auf S_i liegen. Die Eigenräume werden dabei durch das Gebiet und die Randbedingungen bestimmt.

Abbildung 6.32 zeigt den Interpolationsfehler eines Double-Sphere-Arrays (innen schallhart, außen offen) für $N = 9$, $k \in [0{,}5; 9]$ und folgende Abtastgitter:

- *open sphere*: Offenes Kugelarray mit Hyperinterpolation auf der Kugel ($N = 9$, $Q = 100$). Keine Abtastpunkte innerhalb der Kugel.
- *double sphere, inner (closed) sphere optimized*: Double-Sphere-Array mit Hyperinterpolation auf der äußeren Kugelschale und 9 zusätzlichen Abtastpunkten auf der inneren schallharten Kugel mit Radius $\rho = 0{,}5$ ($N = 9$, $Q = 109$). Die 9 zusätzlichen Abtastpunkte wurden über den modifizierten Ansatz mit Nebenbedingungen bestimmt.
- *double sphere, inner (closed) sphere regular*: Double-Sphere-Array mit Hyperinterpolation auf der äußeren Kugelschale ($N = 9$, $Q = 100$) und inneren

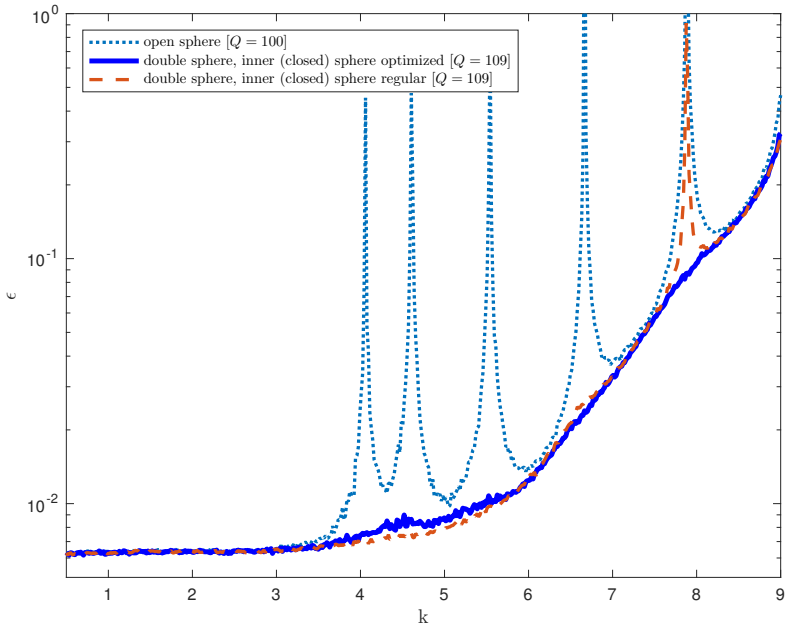


Abbildung 6.32: Relativer Interpolationsfehler ϵ eines Double-Sphere-Arrays (innen schallhart, außen offen) in Abhängigkeit vom Abtastgitter und der Wellenzahl k.

schallharten Kugel ($N = 2$, $Q = 9$) mit Radius $\rho = 0{,}5$. Insgesamt $N = 9$ und $Q = 109$.

Auch hier garantieren die mit dem Optimierungsverfahren bestimmten Abtastpunkte auf der inneren Kugel eine stabile Interpolation über den gesamten Wellenzahlbereich. Es zeigt sich wiederum, dass eine gleichverteilte Abtastung der inneren Kugel die Resonanz bei $k \approx 7.88$ nicht unterdrücken kann. Somit lässt sich ganz allgemein (ohne mathematischen Beweis) feststellen, dass sich eine nicht gleichförmige Abtastung des Schallfelds innerhalb der Kugel (bzw. im Inneren eines konvexen, sternförmigen Gebiets Ω) wesentlich besser zur Stabilisierung des Interpolationsfehlers bei den Eigenfrequenzen der Kugel (bzw. des Gebiets Ω) eignet, als eine annähernde Gleichverteilung der Abtastpunkte.

B Spindle-Torus-Array

Im Folgenden, wird der modifizierte Optimierungsansatz auf das von Alon und Rafaely (2012) vorgeschlagene Spindle-Torus-Array angewendet. Dabei wird das Schallfeld auf einem Rotationstorus abgetastet. Das Abtastgitter wird von einer

konzentrischen Kugel auf die Oberfläche des Rotationstorus projiziert. Das daraus resultierende Abtastgitter ist allerdings nicht effizient (vgl. Abschnitt 6.5.2). Dem kann durch eine auf das Problem adaptierte Berechnung der magischen Punkte (vgl. Gl. 6.5.3) begegnet werden. Hierzu werden die möglichen Positionen der Abtastpunkte entweder (i) auf die Oberfläche des Rotationstorus und/oder (ii) auf die innere Fläche des Rotationstorus beschränkt. Es kann gezeigt werden, dass die inneren Abtastpunkte wiederum die Interpolation bei den Eigenfrequenzen des Gebiets Ω, welches durch die Oberfläche des Rotationstorus begrenzt wird, stabilisieren. Abb. 6.33 zeigt ein Rotationstorus mit den Parametern $R = 0{,}3$ und $r = 0.,$, welches über folgende parametrische Gleichung definiert ist. Es gilt für alle $\phi \in [0, 2\pi[$ und $\theta \in [0, 2\pi[$:

$$\begin{cases} x = (R + r\sin\theta)\cos\phi, \\ y = (R + r\sin\theta)\sin\phi, \\ z = r\cos\theta. \end{cases} \tag{6.131}$$

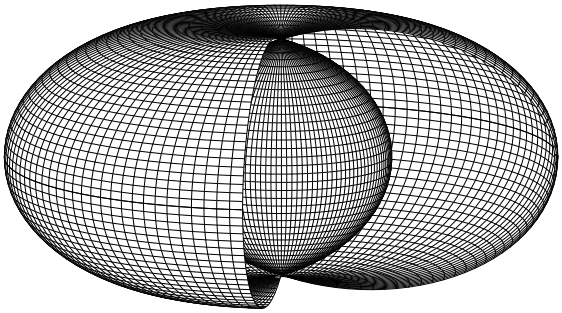


Abbildung 6.33: Rotationstorus (*spindle torus*) mit den Parametern $R = 0{,}3$ und $r = 0{,}7$.

In Abb. 6.34 ist Interpolationsfehler eines Spindle-Torus-Arrays mit den Parametern $R = 0{,}3$ und $r = 0{,}7$ für $N = 9$, $k \in [0{,}5; 9]$ und folgende Abtastgitter dargestellt:

- *projected sphere*: Projektion eines nahezu gleichverteilten Abtastgitters auf der Kugel ($N = 9$, $Q = 100$) auf die Flächen des Rotationstorus. Die eine Hälfte der Abtastpunkte wird auf die Oberfläche, die andere Hälfte auf die innere Fläche projiziert (vgl. Alon und Rafaely, 2012). Die Zuordnung erfolgt zufällig.

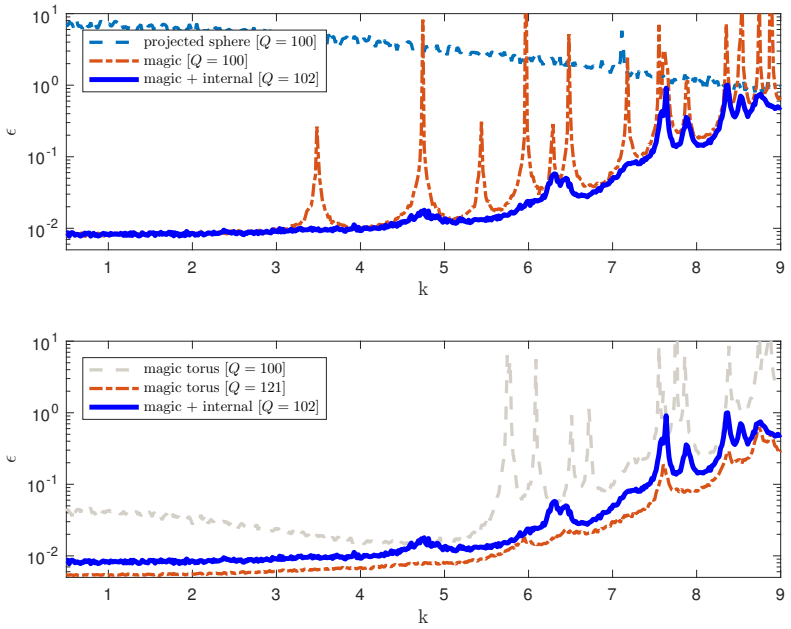


Abbildung 6.34: Relativer Interpolationsfehler ϵ eines Spindle-Torus-Arrays ($R = 0.3$ and $r = 0.7$) in Abhängigkeit vom Abtastgitter und der Wellenzahl k.

- *magic*: Rotationstorus mit magischen Punkten auf der Oberfläche ($N = 9$, $Q = 100$). Die innere Fläche wird nicht berücksichtigt.
- *magic + internal*: Rotationstorus mit magischen Punkten auf der Oberfläche und 2 zusätzlichen Abtastpunkten auf der inneren Fläche ($N = 9$, $Q = 102$). Die 2 zusätzlichen Abtastpunkte wurden über den in Abschnitt 6.5.3-B vorgestellten Ansatz bestimmt.
- *magic torus*: Magische Punkte über den gesamten Torus ($N = 10$, $Q = 121$). Da bei $N = 9$ ($Q = 100$) bereits bei sehr kleinen k starke Resonanzen auftreten, wurde die Ordnung auf $N = 10$ erhöht.

Aus der Abbildung ist ersichtlich, dass die Interpolation bei der von Alon und Rafaely vorgeschlagenen Methode instabil ist. Werden hingegen magische Punkten am Rand $\partial\Omega$ des Gebiets Ω (d. h. das durch die Oberfläche des Rotationstorus begrenzte Volumen) verwendet, ist die Interpolation bis auf die Frequenzen nahe den Eigenfrequenzen von Ω stabil. Auch hier kann wiederum mit Abtastpunkten innerhalb des Gebiets eine Stabilisierung der Interpolation erreicht werden. Wie die Simulation zeigt, ist das in dieser Arbeit vorgeschlagene Optimierungsverfahren

am effizientesten. Es reichen 2 zusätzliche Abtastpunkte ($N = 9$, $Q = 102$) um die Interpolation über den gesamten Wellenzahlbereich zu stabilisieren. Dabei wird auch ein möglichst kleiner Interpolationsfehler erreicht. Werden $Q = 100$ magische Punkte auf die äußere und innere Fläche verteilt, ergeben sich relative starke Resonanzen für $k > 5{,}5$. Wird nun wiederum die Ordnung der magischen Punkte auf $N = 10$ ($Q = 121$) erhöht, ist der Interpolationsfehler über den gesamten Wellenzahlbereich kleiner als bei der hier vorgeschlagenen Methode. Allerdings werden hierbei $Q = 121$ anstatt von $Q = 102$ Abtastpunkten benötigten.

6.5.6 Zusammenfassung

In diesem Kapitel wurden zwei Methoden zum Entwurf robuster modaler Mikrofonarrays vorgestellt. Als Maß für die Robustheit wurde der relative Fehler der Interpolation des Schallfeldes innerhalb des von dem jeweils betrachteten Array umschlossenen Volumen (dem Gebiet Ω) verwendet. Zudem wurde gezeigt, dass eine stabile breitbandige Interpolation des Schallfeldes (und somit eine stabile breitbandige Schätzung der sphärischen Fourierkoeffizienten) nur dann möglich ist, wenn zumindest einige wenige Abtastpunkte im Inneren des Arrays positioniert werden. Das Abtastgitter auf der Oberfläche (dem Rand $\partial\Omega$) sollte dabei auf ein Array optimal angepasst werden.

Die erste Methode bestimmt die exakten Positionen der Abtastpunkte innerhalb des Volumens, um mit möglichst wenigen zusätzlichen Abtastpunkten die Interpolation über einen möglichst großen Frequenzbereich zu stabilisieren. Dazu müssen die Eigenfrequenzen und Eigenmoden des Gebiets Ω berechnet werden, was nur bei sehr einfachen Formen wie zum Beispiel der Kugel möglich ist. Die zweite Methode wurde von der Methode zur Bestimmung der magischen Punkte abgeleitet, die keinerlei *a-priori* Wissen über die Eigenfrequenzen und Eigenmoden des jeweiligen Gebiets Ω benötigt. Es wurde gezeigt, dass die magischen Punkte die störenden Resonanzen bei den Eigenfrequenzen gut unterdrücken, allerdings einen höheren Interpolationsfehler haben als die erste Methode. Trotzdem sind sie eine gute Alternative, für den Falle, dass sich die Eigenfrequenzen und Eigenmoden für ein Gebiet Ω nicht einfach bestimmen lassen.

Um die praktische Realisierbarkeit der optimierten Abtastgitter zu erhöhen, wurde die erste Methode dahingehend modifiziert, dass Nebenbedingungen (wie z. B. vorgegebene Einschränkungen in der Arraygeometrie) mit berücksichtigt werden können. Angewendet auf ein Double-Sphere Array und eine Spindle-Torus Array konnte die Anzahl der für eine stabile Interpolation (und stabile Schätzung der sphärischen Fourierkoeffizienten) benötigten Abtastpunkte im Vergleich zu den in der Literatur vorgeschlagenen Abtastgittern wesentlich reduziert werden.

Einen Sonderfall bildet das Ellipsoid, bei dem mit nur einem einzelnen zusätzlichen Abtastpunkt im Inneren des Arrays die Interpolation über den gesamten Frequenzbereich stabilisiert werden kann.

7

Zusammenfassung

Das primäre Ziel der breitbandigen Signalaufbereitung liegt in der weitgehenden Unterdrückung von Störgeräuschen und Interferenzsignalen, bei möglichst geringer Verzerrung des Nutzsignals. In der vorliegenden Arbeit wurde die Breitbandigkeit nicht nur spektral, sondern auch räumlich verstanden. Eine optimale Lösung kann somit nur dadurch erreicht werden, dass ein- und mehrkanalige Methoden der Signalaufbereitung kombiniert werden. So lassen sich die residualen Störgeräusche am Ausgang eines Mikrofonarrays durch Nachschalten eines einkanaligen Geräuschreduktionsfilters unterdrücken. Die verschiedenen Methoden der Signalaufbereitung in ein- und mehrkanaligen Mikrofonanwendungen wurden ausführlich diskutiert und zu mehreren bis dato offenen Problemen Lösungen vorgeschlagen. Die daraus abgeleiteten Verfahren haben eine möglichst kurze Systemlatenz und einen vergleichsweise geringen Rechenleistungsbedarf und sind für die Implementierung auf digitalen Signalprozessoren in echtzeitfähigen Systemen geeignet.

Die spektrale Subtraktion gehört zu den am weitesten verbreiteten Methoden der Störgeräuschunterdrückung in einkanaligen Mikrofonanwendungen. Bei den aus der Literatur bekannten Ansätzen muss jedoch stets ein Kompromiss zwischen der Verzerrung des Nutzsignals und der Unterdrückung von Störgeräuschen und Interferenzen getroffen werden. Ein äußerst leistungsfähiges Verfahren wurde von Ephraim und Malah vorgestellt. Dieses schätzt die Amplitude des ungestörten Signals aus dem Kurzzeitspektrum des gestörten Gesamtsignals. Um ein grundlegendes Verständnis für die Funktionsweise des Ephraim-Malah-Filters zu bekommen,

wurden die spektralen Gewichte auf einer Kennfläche über der vom *a-priori* und *a-posteriori* SNR aufgespannten Ebene dargestellt. Aus dem Vergleich des Verlaufs der spektralen Gewichte unterschiedlicher Amplitudenschätzer konnte in dieser Arbeit ein schnell ansprechender entscheidungsgesteuerter Ansatz hergeleitet werden. Ein wesentlicher Vorteil dieses neuen Ansatzes sind die unabhängige Kontrolle über die Zeitkonstante der Mittelung und Unterdrückung des Musical Noise sowie das schnelle dynamische Ansprechverhalten des Schätzers. Letzteres führt dazu, dass bei der Störgeräuschunterdrückung die Transienten (wie z. B. Sprach-Onsets) weniger stark verzerrt werden.

Eine weitere Möglichkeit den Höreindruck bei der Störgeräuschunterdrückung zu verbessern besteht darin, die zeitlichen und spektralen Maskierungseffekte des menschlichen Gehörs zu nutzen, um Störungen nur dort zu unterdrücken, wo diese auch tatsächlich wahrnehmbar sind. In der vorliegenden Arbeit wurde die Spektraltransformation als parallele Bank sich überlappender Gammaton-Filter realisiert, deren Kanäle an die Frequenzgruppen des menschlichen Gehörs angepasst wurden. Dadurch lassen sich die Gewichte der spektralen Subtraktion auf die Mithörschwelle begrenzen. Das Synthesefilter rekonstruiert das Originalsignal durch Summation der Teilbandsignale. Da die Gruppenlaufzeit der Gammaton-Filter mit abnehmender Mittenfrequenz zunimmt, kommt es bei der Summation aufgrund von Phasendifferenzen an den Übergängen benachbarter Teilbändern zu teils erheblichen Signalverzerrungen. Aus diesem Grund wurde ein sehr leistungsfähiges Kriterium entwickelt, welches die Notwendigkeit bestimmt, das Vorzeichen bei der Summation aufeinanderfolgender Teilbänder einer Gammaton-Filterbank zu wechseln. Der Vorzeichenwechsel begünstigt die konstruktive Überlagerung der Signale an den Bandgrenzen und reduziert die Welligkeit am Ausgang der Synthese-Filterbank. Es werden keine rechenaufwändigen Synthesefilter benötigt. Im Gegensatz zum Ansatz von Herzke und Hohmann (bei dem die Signallaufzeiten vor der Summation durch Drehung der Phase kompensiert werden) kommt es hierbei zu keiner zusätzlichen Verzögerung des Ausgangssignals. Da keine Drehung der Phase erforderlich ist, können anstatt der komplexwertigen Gammaton-Filter wesentlich recheneffizientere reellwertige Gammaton-Filter verwendet werden.

Glasberg und Moore folgend, lässt sich der Amplitudengang des menschlichen Außen- und Mittelohrs für Frequenzen unterhalb von 1 kHz durch die inverse 100-phon ISO-Kurve, sowie für darüber liegende Frequenzen über den Kurvenverlauf der inversen Ruhehörschwelle nachbilden. Pflüger verwendet eine Kaskade biquadratischer Filter um den Amplitudengang anzunähern. Allerdings wurde hierbei nur ein auf Sprache begrenzter Frequenzbereich betrachtet. Ähnlich dem Ansatz von Pflüger, wird in der vorliegenden Arbeit der Amplitudengang durch ein rekursives Filter 8-ter Ordnung angenähert. Dabei werden die Filterkoeffizienten über einen modifizierten Yule-Walker-Algorithmus bestimmt. Es zeigt sich, dass dieser Ansatz den gewünschten Amplitudenverlauf auch im hohen Frequenzbereich sehr gut annähert. Dies ist vor allem dann von Bedeutung, wenn die Bandbreite des zu verarbeitenden Signals den gesamten Hörbereich umfasst. Zudem wurden im Rahmen dieser Arbeit die von Pflüger vorgeschlagenen parametrischen Filter an die neuere ISO-Norm 226:2003 angepasst.

Die Wirksamkeit der mit der Gammaton-Filterbank kombinierten optimalen Störgeräuschunterdrückung konnte anhand numerischer Simulationen verifiziert werden. Um eine allgemein gültige Aussauge über die tatsächliche Verbesserung des Höreindrucks treffen zu können, müssten in Zukunft Hörversuche durchgeführt werden.

Sind Nutz- und Störquellen räumlich voneinander getrennt, kann mit mehrkanaligen Geräuschreduktionsverfahren eine wesentliche Verbesserung des Signal-Störverhältnisses erzielt werden. Zu diesem Zweck wurde in dieser Arbeit ein robuster adaptiver Beamformer entwickelt, der auf dem Prinzip des von Hoshuyama vorgeschlagenen robusten Generalised Sidelobe Cancellers beruht. Dazu wird einem Preprozessor, bestehend aus einem fixen Beamformer und einer adaptiven Blockiermatrix, ein aktiver Störgeräuschunterdrücker nachgeschaltet. Bei einer fehlerhaften Adaption wird das Signal am Ausgang des Beamformers stark verzerrt. Aus diesem Grund wird die Blockiermatrix nur in denjenigen Perioden adaptiert, in denen Nutzsignal vorhanden ist; der aktive Störgeräuschunterdrücker wird hingegen in nutzsignalfreien Perioden adaptiert. Die Robustheit hängt somit sehr stark von der Steuerung der Adaption ab. Dies ist die Grundlage für den in dieser Arbeit

entwickelten robusten adaptiven Beamformer mit verbesserter Adaptionskontrolle. Am Beispiel eines Mikrofonarrays kleiner Bauform und geringer Mikrofonanzahl wurde gezeigt, dass die Schätzung des Signal-Interferenz-Abstands starken richtungsabhängigen Schwankungen unterliegt. Die Schätzung, ob Nutzsignal vorhanden ist oder nicht, weist einen von der Arraygeometrie abhängigen systematischen Fehler auf. Dem kann durch zeitliches Nachführen der Schwellwerte der Signalpausenerkennung begegnet werden. Darüber hinaus wurde ein räumliches Kriterium eingeführt, mit dem abgeschätzt werden kann, wie weit die Einfallsrichtung von der Vorzugsrichtung des Mikrofonarrays abweicht. Dabei werden die Signalleistungen aus unterschiedlichen Raumrichtungen über zusätzliche Gitterkeulen, die das Schallfeld in einem räumlichen Raster abtasten, geschätzt und ins Verhältnis gesetzt. Messungen und numerische Simulationen zeigen, dass das in dieser Arbeit vorgestellte Verfahren auch bei kleinen Mikrofonarrays und wenigen Mikrofonen eine hohe und breitbandige richtungsabhängige Verstärkung erlaubt. Aufgrund der adaptiven Ausführung ist der Beamformer zudem robust gegenüber Bauteiltoleranzen, Fehlpositionierungen und reflexionsbehaftete Schallwellenausbreitung. Die Robustheit der Adaptionssteuerung kann voraussichtlich dadurch weiter erhöht werden, indem zusätzlich weitere (jedoch meist auch rechenaufwändigere) zur Signalpausenerkennung (wie z. B. das Harmonic Tunneling oder Histogramm-Methoden) implementiert werden.

Die Signalaufbereitung mit modalen Beamformern ermöglicht neben der zeitlichen auch eine hohe räumliche Bandbreite. Dabei haben die verwendeten räumlichen Abtastgitter einen wesentlichen Einfluss auf die Fehlertoleranz der Schätzung der Koeffizienten des Wellenspektrums, welche die Basis für die Implementierung unterschiedlicher Beamformer bilden. Für die Herleitung eines robusten modalen Beamformers wurde zuerst das Verhalten unterschiedlicher Abtastgitter studiert. Diese wurden anhand der Konditionszahl, der Orthogonalität der Basisfunktionen und des Fehlers bei der Interpolation des Schallfeldes innerhalb des Arrays bewertet. Es konnte gezeigt werden, dass sich die Interpolation des Schallfeldes mit einigen wenigen Abtastpunkten im Inneren eines Arrays stabilisieren lässt. In weiterer Folge wurde eine Methode zum Entwurf optimaler Abtastgitter vorgestellt, mit der sich die mindestens benötigte Anzahl an inneren Abtastpunkten

und deren optimale Positionen ermitteln lassen. Numerische Simulationen zeigen, dass sich die Interpolation mit nur einigen wenigen Abtastpunkten im Inneren des Volumens über einen möglichst breiten Frequenzbereich stabilisieren lässt. Dies konnte neben der Kugel auch ganz allgemein für konvexe und sternförmige Gebiete (wie z. B. ellipsoidische und kubische Arrays) gezeigt werden. Die Simulationen zeigen auch die Effizienz des vorgestellten Optimierungsansatzes, der im Vergleich zu anderen Methoden über den gesamten Frequenzbereich den kleinsten Interpolationsfehler ergibt. Ein Nachteil dieses Ansatzes liegt jedoch darin, dass die Abtastpunkte über das gesamte Volumen verteilt sind. Dadurch lassen sich die optimalen Abtastgitter in der Praxis meist nur schwer implementieren. Aus diesem Grund wurde der Ansatz in weiterer Folge dahingehend erweitert, dass bei der Berechnung der optimalen Abtastpunkte zusätzliche Randbedingungen (wie z.B. vorgegebene Arraygeometrien) berücksichtigt werden können. Numerische Simulationen unterschiedlicher Arraygeometrien (double sphere, mixed sphere und spindle torus) zeigen, dass sich mit dem modifizierten Ansatz nicht nur die Robustheit der Interpolation und der Schätzung der Koeffizienten des Wellenspektrums, sondern auch die praktische Realisierbarkeit eines Abtastgitters erhöhen lässt. Ein wesentlicher Nachteil des in dieser Arbeit vorgestellten Optimierungsansatzes ist jedoch, dass zur Bestimmung der optimalen Abtastpunkte die Eigenfrequenzen und Eigenmoden des von einem Mikrofonarray umschlossenen Gebietes berechnet werden müssen. Dies ist für allgemeine Gebiete keine triviale Aufgabe.

A

Berechnungen zu Gammaton-Filtern

Dieser Abschnitt befasst sich mit der Herleitung zeitdiskreter Gammaton-Filter (s. Kap. 2.3) als Teilbandfilter einer Analyse-Synthese-Filterbank zur auditiven Signalverarbeitung (s. Kap. 3). Ausgehend von der analogen Übertragungsfunktion wird die zeitdiskrete Beschreibung von einfachen linearen Gammaton-Filtern (GTF), „All-Pole“ (APGF), „One-Zero“ (OZGF) und „Three-Zero“ (TZGF) Gammaton-Filtern abgeleitet. Die Herleitung folgt vorwiegend den Ausführungen von Slaney (1993), Lyon (1996) und Zotter (2004, Anhang A). Zudem wird aus der Betrachtung der Phase an den Übergangsfrequenzen benachbarter Teilbandfilter ein sehr einfaches Verfahren abgeleitet, welches eine nahezu perfekte Rekonstruktion des Originalsignals durch Summation der Teilbandsignale erlaubt (s. a. Zotter, 2004; Noisternig et al., 2009).

A.1 Beschreibung im Zeitbereich

Die Impulsantwort $g_m(t)$ eines einfachen linearen Gammaton-Filters der Ordnung m kann wie folgt beschrieben werden, siehe auch Gl. (2.23):

$$g_m(t) = a\, t^{m-1} e^{-bt} \cos(\omega_c t + \phi) \quad \forall\, t \geqslant 0, m \geqslant 1\,. \tag{A.1}$$

Der Parameter b bestimmt die Länge der Impulsantwort und somit die Bandbreite des Filters, der Verstärkungsfaktor a dient der Skalierung. Die Filterordnung

m bestimmt zum überwiegenden Teil die Flankensteilheit, $\omega_c = 2\pi f_c$ ist die Kreisfrequenz der Trägerschwingung und ϕ die Anfangsphase.

A.2 Beschreibung im Laplacebereich

Unter Anwendung folgender Transformationspaare (siehe z. B. Bronstein et al., 2013, Kap. 0.8.2 und 1.11.1) kann Gl. (A.1) in den Laplacebereich übergeführt werden,

$$e^{at} \cdot u(t) \quad \circ\!\!-\!\!\bullet \quad \frac{1}{s-a} \tag{A.2}$$

$$e^{j\omega t} x(t) \cdot u(t) \quad \circ\!\!-\!\!\bullet \quad X(s - j\omega) \tag{A.3}$$

$$t^m x(t) \cdot u(t) \quad \circ\!\!-\!\!\bullet \quad (-1)^m \frac{d^m}{ds^m} X(s) \tag{A.4}$$

wobei $u(t)$ die Heaviside-Funktion bezeichnet und somit die Voraussetzung von Kausalität erfüllt ist. Die Einhüllende der Gammaton-Funktion erster Ordnung besitzt die Laplace-Transformierte

$$G_{\mathrm{env},1}(s) = \frac{1}{s+b}\,. \tag{A.5}$$

Unter Anwendung des Ableitungstheorems Gl. (A.4) lässt sich daraus die Laplace-Transformierte der Einhüllenden für beliebige Filterordnungen m ableiten:

$$G_{\mathrm{env},m}(s) = (-1)^{m-1} \frac{(m-1)!}{(s+b)^m} \qquad \forall \left\{ m \mid m \in \mathbb{Z}, m > 0 \right\}. \tag{A.6}$$

Die Kosinusfunktion in Gl. (A.1) lässt sich als Linearkombination imaginärer Exponentialfunktionen darstellen. Zur Vereinfachung wird die Anfangsphase ϕ zu Null gesetzt[146]:

$$\cos(\omega_c t) = \frac{1}{2}\left(e^{j\omega_c t} + e^{-j\omega_c t}\right). \tag{A.7}$$

Mit der Annahme $a = 1$ und unter Anwendung des Modulationstheorems Gl. (A.3) lässt sich mit den Gln. (A.6) und (A.7) die Laplace-Transformierte der Gammaton-

[146] Gammaton-Filter deren Anfangsphase zu Null gesetzt wird werden gemeinhin als reellwertige Kosinus-Phasen GTF (*real-valued cosine-phase GTF*) bezeichnet.

Funktion für beliebige Filterordnungen m ableiten:

$$\begin{aligned} G_m(s) &= (-1)^{m-1}\,\frac{(m-1)!}{2}\left(\frac{1}{(s+b-j\omega_c)^m}+\frac{1}{(s+b+j\omega_c)^m}\right) \\ &= (-1)^{m-1}\,\frac{(m-1)!}{2}\,\frac{(s+b-j\omega_c)^m+(s+b+j\omega_c)^m}{\left((s+b)^2+\omega_c^2\right)^m}\,. \end{aligned} \tag{A.8}$$

Die Laplace-Transformierte in Gl. (A.8) ist für alle $\{m \mid m \in \mathbb{Z}, m > 0\}$ gültig und besitzt m konjugiert komplexe Polstellen bei $s_p = -b \pm j\omega_c$. Die Lage der Nullstellen wird üblicherweise numerisch bestimmt (siehe z. B. Slaney, 1993), kann zur besseren Anschaulichkeit jedoch auch sehr einfach über ein geometrisches Verfahren ermittelt werden (s. Lyon, 1996).

A.3 Beschreibung im z-Bereich

In der Literatur findet man zahlreiche Vorschläge zur Realisierung von Gammaton-Filtern mit digitalen rekursiven Filterstrukturen (siehe z. B. Cooke, 1991; Lyon, 1996; Pflüger, 1997, Kap. 5.1; Hohmann, 2002; Zotter, 2004; Katsiamis et al., 2006). Nach Cooke (1991, Kap. 2.4) stehen für die Transformation vom zeitkontinuierlichen in den zeitdiskreten Bereich im Wesentlichen drei Methoden zur Verfügung[147]: die Pol-Nullstellenübertragung, die Impulsinvarianz-Transformation und die bilineare Transformation. Den Ergebnissen der Studien von Cooke (1991), Slaney (1993) folgend, kann mit der Impulsinvarianz-Transformation die beste Übereinstimmung mit der Impulsantwort bzw. dem Amplituden- und Phasengang analoger GTF erreicht werden (s. Cooke, 1991, Abb. 2.2). Mit den nachfolgenden Ausführungen wird gezeigt, wie sich aus der Laplace-Transformierten analoger GTF die Übertragungsfunktion $G_m(z)$ der digitalen GTF herleiten lässt.

A.3.1 Impulsinvarianz-Transformation

Bei der Impulsinvarianz-Transformation wird die Impulsantwort eines digitalen Filters $h_\mathrm{d}(n)$ durch die gleichförmige Abtastung der Impulsantwort des analogen

[147] Details zu den unterschiedlichen Transformationen vom zeitkontinuierlichen in den zeitdiskreten Bereich sind zum Beispiel in Oppenheim et al. (1998, S. 439 - 463) zu finden.

Filters $h_\mathrm{a}(t)$ in den Zeitpunkten $t = nT$ abgeleitet, wobei die Abtastperiode T dem Kehrwert der Abtastfrequenz f_A entspricht. Die Impulsantwort des zeitdiskreten Systems berechnet sich dabei zu $h_\mathrm{d}(n) = T\, h_\mathrm{a}(nT)$. Angewendet auf Gl. (A.1) ergibt sich die zeitdiskrete Impulsantwort eines GTF der Ordnung m:

$$g_m(n) = a\, T^m n^{m-1} e^{-Bn} \cos(\theta n + \phi)\,, \tag{A.9}$$

mit dem normierten Bandbreitenparameter $B = bT$ und der normierten Kreisfrequenz $\theta = \omega_c T$. Die Übertragungsfunktion $G_m(z)$ ist die z-Transformierte der zeitdiskreten Impulsantwort $g_m(n)$. Durch die Abtastung ergibt sich eine periodische Funktion im Frequenzbereich (siehe z. B. Oppenheim et al., 1998, Kap. 7.1.1). Ist der Frequenzgang des analogen Filters bei der Nyquist-Frequenz nicht vollständig abgeklungen entstehen Aliasing-Artefakte. Cooke (1991) schlussfolgert in einer vergleichenden Studie unterschiedlicher Transformationsverfahren, dass aufgrund der Bandbegrenztheit der Gammaton-Funktion die durch Aliasing entstehenden Abbildungsfehler vernachlässigbar sind (s. a. Slaney, 1993). Die Impulsinvarianz-Transformation ist somit ohne weiteres auf den Entwurf digitaler Gammaton-Filter anwendbar.

Die direkte z-Transformation der zeitdiskreten Impulsantwort in Gl. (A.9) kann nur für $m = 1$ geschlossen analytisch gelöst werden. Für höhere Ordnungen kann numerisch eine Lösung gefunden werden (s. Slaney, 1993). Unter Anwendung folgender Transformationspaare (s. Oppenheim et al., 1998, Tab. 3.1)

$$a^n \cdot u(n) \quad \circ\!\!-\!\!\bullet \quad \frac{1}{1 - az^{-1}} \tag{A.10}$$

$$a^n x(n) \cdot u(n) \quad \circ\!\!-\!\!\bullet \quad X(z/a) \tag{A.11}$$

$$nx(n) \cdot u(n) \quad \circ\!\!-\!\!\bullet \quad -z^{-1}\frac{d}{dz}X(z) \tag{A.12}$$

ergibt sich mit $\phi = 0$, $a = 1$ und $r = e^{-B}$ die Systemantwort für $m = 1$:

$$\begin{aligned} G_1(z) &= \frac{T}{2}\left(\frac{1}{1 - re^{j\theta}z^{-1}} + \frac{1}{1 - re^{-j\theta}z^{-1}}\right) \\ &= T \cdot \frac{1 - r\cos\theta\, z^{-1}}{1 - 2r\cos\theta\, z^{-1} + r^2 z^{-2}}\,. \end{aligned} \tag{A.13}$$

Slaney (1993) zeigt, dass bei direkter Transformation der Pole vom Laplace- in den z-Bereich auch für GTF höherer Ordnung eine geschlossen analytische Lösung für die Impulsinvarianz-Transformation existiert. Dabei werden die Pole $s_p = -b \pm j\omega_c$ des zeitkontinuierlichen Filters über die Transformation $z_p = re^{\pm j\theta}$ auf die Pole des zeitdiskreten Filters übertragen (s. a. Oppenheim et al., 1998, Kap. 7.1.1).

Nach Cooke (1991) lassen sich Gammaton-Filter durch Verschieben eines Tiefpass-Prototypfilters um die jeweilige Filtermittenfrequenz ω_c realisieren. Das digitale TP-Prototypfilter wird dabei als rekursives Filter 4ter Ordnung für komplexwertige Signale implementiert. Slaney hingegen schlägt eine direkte Implementierung mit Filtern 8ter Ordnung und reellen Koeffizienten vor. Der Rechenaufwand für die Filterung bleibt somit gleich. Allerdings kann bei Slaneys Ansatz die Demodulation vor und Modulation nach der Filterung entfallen. In den folgenden Abschnitten werden die Übertragungsfunktionen einfacher linearer GTF bis zur Ordnung $m = 4$ hergeleitet.

A Gammaton-Filter erster Ordnung

Gammaton-Filter höherer Ordnung lassen sich aus einer Kaskade von Filtern zweiter Ordnung zusammensetzen. Ausgehend von der allgemeinen zeitkontinuierlichen Darstellung einer Filterstufe $G(s)$ mit konjugiert komplexen Polen $s_p = -b \pm j\omega_c$ und einer Nullstelle s_z

$$G(s) = \frac{s - s_z}{(s+b)^2 + \omega_c^2}\,, \tag{A.14}$$

kann durch Partialbruchzerlegung

$$G(s) = \frac{1}{j2\omega_c}\left(\frac{s_z + b + j\omega_c}{s + b + j\omega_c} - \frac{s_z + b - j\omega_c}{s + b - j\omega_c}\right)\,, \tag{A.15}$$

Multiplikation mit T und Substitution der Pole mit $z_p = re^{\pm j\theta}$ die z-Transformierte $G(z)$ hergeleitet werden:

$$\begin{aligned} G(z) &= \frac{T}{j2\omega}\left(\frac{s_z + b + j\omega_c}{1 - re^{-j\theta}z^{-1}} - \frac{s_z + b - j\omega_c}{1 - re^{j\theta}z^{-1}}\right) \\ &= T \cdot \frac{1 - r\left(\left(s_z + b\right)\,\omega_c^{-1}\sin\theta + \cos\theta\right)z^{-1}}{1 - 2r\cos\theta z^{-1} + r^2 z^{-2}}\,. \end{aligned} \tag{A.16}$$

Einsetzen von $m = 1$ in Gl. (A.8) führt auf die Laplace-Transformierte eines Gammaton-Filters 1ter Ordnung:

$$G_1(s) = \frac{s+b}{(s+b)^2+\omega_c^2}, \tag{A.17}$$

welche eine Nullstelle bei $s_z = -b$ besitzt. Die z-Transformierte kann durch Einsetzen der Nullstelle in Gl. (A.16) berechnet werden:

$$G_1(z) = T \cdot \frac{1 - r\cos\theta\, z^{-1}}{1 - 2r\cos\theta\, z^{-1} + r^2 z^{-2}}. \tag{A.18}$$

Gl. (A.18) stimmt mit der in Gl. (A.13) durch direkte z-Transformation der zeitdiskreten Gammaton-Impulsantwort hergeleiteten Übertragungsfunktion überein. Die z-Transformierte von GTF höherer Ordnung kann, wie im Folgenden gezeigt wird, in gleicher Weise aus den Gln. (A.8) und (A.16) bestimmt werden. Die grundlegende Schwierigkeit besteht darin, die Nullstellen der Übertragungsfunktion zu bestimmen. Diese werden meist mit Hilfe von numerischen Methoden ermittelt.

B Gammaton-Filter zweiter Ordnung

Die Laplace-Transformierte der Gammaton-Funktion 2ter Ordnung ergibt sich durch Einsetzen von $m = 2$ in Gl. (A.8):

$$\begin{aligned} G_2(s) &= -\frac{1}{2}\left(\frac{1}{(s+b-j\omega_c)^2} + \frac{1}{(s+b+j\omega_c)^2}\right) \\ &= -\frac{s^2+2bs+b^2-\omega_c^2}{\left((s+b)^2+\omega_c^2\right)^2}. \end{aligned} \tag{A.19}$$

Gl. (A.19) besitzt zwei Nullstellen bei $s_{z,1,2} = -b \pm \omega_c$ und kann über den Ansatz der Partialbruchzerlegung in Kaskadenform dargestellt werden:

$$G_2(s) = -\frac{s+b+\omega_c}{(s+b)^2+\omega_c^2} \cdot \frac{s+b-\omega_c}{(s+b)^2+\omega_c^2}. \tag{A.20}$$

Unter Anwendung der Impulsinvarianz-Transformation kann durch Einsetzen der Nullstellen in Gl. (A.16) die z-Transformierte der Gammaton-Funktion 2ter

Ordnung sehr einfach bestimmt werden:

$$\begin{aligned} G_2(z) &= -T^2 \cdot \frac{1 - r\left(\sin\theta + \cos\theta\right) z^{-1}}{1 - 2r\cos\theta z^{-1} + r^2 z^{-2}} \cdot \frac{1 - r\left(-\sin\theta + \cos\theta\right) z^{-1}}{1 - 2r\cos\theta z^{-1} + r^2 z^{-2}} \\ &= -T^2 \cdot \frac{1 - r\sqrt{2}\,\cos\left(\theta - \frac{\pi}{4}\right) z^{-1}}{1 - 2r\cos\theta z^{-1} + r^2 z^{-2}} \cdot \frac{1 - r\sqrt{2}\,\cos\left(\theta + \frac{\pi}{4}\right) z^{-1}}{1 - 2r\cos\theta z^{-1} + r^2 z^{-2}} \,. \end{aligned} \tag{A.21}$$

C Gammaton-Filter dritter Ordnung

Die Laplace-Transformierte der Gammaton-Funktion 3ter Ordnung berechnet sich aus Gl. (A.8) zu

$$\begin{aligned} G_3(s) &= \frac{1}{\left(s + b - j\omega_c\right)^3} + \frac{1}{\left(s + b + j\omega_c\right)^3} \\ &= 2 \cdot \frac{\left(s+b\right)^3 - 3\,\omega_c^2\left(s+b\right)}{\left(\left(s+b\right)^2 + \omega_c^2\right)^3} \\ &= 2 \cdot \frac{\left(s+b\right)\left(s + b + \sqrt{3}\,\omega_c\right)\left(s + b - \sqrt{3}\,\omega_c\right)}{\left(\left(s+b\right)^2 + \omega_c^2\right)^3} \,, \end{aligned} \tag{A.22}$$

und besitzt drei Nullstellen[148] bei $s_{z,1} = -b$ und $s_{z,2,3} = -b \pm \sqrt{3}\,\omega_c$. Durch Einsetzen der Nullstellen in Gl. (A.16) lässt sich wiederum die z-Transformierte bestimmen:

$$\begin{aligned} G_3(z) = 2\,T^3 &\cdot \frac{1 - r\cos\theta z^{-1}}{1 - 2r\cos\theta z^{-1} + r^2 z^{-2}} \cdot \frac{1 - r\left(\sqrt{3}\sin\theta + \cos\theta\right) z^{-1}}{1 - 2r\cos\theta z^{-1} + r^2 z^{-2}} \cdot \\ &\cdot \frac{1 - r\left(-\sqrt{3}\sin\theta + \cos\theta\right) z^{-1}}{1 - 2r\cos\theta z^{-1} + r^2 z^{-2}} \\ = 2\,T^3 &\cdot \frac{1 - r\cos\theta z^{-1}}{1 - 2r\cos\theta z^{-1} + r^2 z^{-2}} \cdot \frac{1 - \frac{r}{2}\,\cos\left(\theta - \frac{\pi}{3}\right) z^{-1}}{1 - 2r\cos\theta z^{-1} + r^2 z^{-2}} \cdot \\ &\cdot \frac{1 - \frac{r}{2}\,\cos\left(\theta + \frac{\pi}{3}\right) z^{-1}}{1 - 2r\cos\theta z^{-1} + r^2 z^{-2}} \,. \end{aligned} \tag{A.23}$$

[148] Die Nullstellen wurden mit Wolfram Mathematica (Version 8 und 9) bestimmt.

D Gammaton-Filter vierter Ordnung

Die Übertragungsfunktion eines GTF 4ter Ordnung folgt aus Gl. (A.8):

$$\begin{aligned} G_4(s) &= \frac{6}{2}\left(\frac{1}{(s+b-j\omega_c)^4}+\frac{1}{(s+b+j\omega_c)^4}\right) \\ &= 6\cdot\frac{s^4+4bs^3+6\left(b^2-\omega_c^2\right)+4\left(b^3-3b\omega_c^2\right)s+b^4+6b^2\omega_c^2+\omega_c^4}{\left((s+b)^2+\omega_c^2\right)^4}\,. \end{aligned} \tag{A.24}$$

Mit den Nullstellen $s_{z,1-4}=-b\pm\left(1\pm\sqrt{2}\right)\omega_c$ ergibt sich die Kaskadenform

$$\begin{aligned} G_4(s) = 6\cdot\frac{s+b+\left(1+\sqrt{2}\right)\omega_c}{(s+b)^2+\omega_c^2}\cdot\frac{s+b+\left(1-\sqrt{2}\right)\omega_c}{(s+b)^2+\omega_c^2}\cdot \\ \frac{s+b-\left(1-\sqrt{2}\right)\omega_c}{(s+b)^2+\omega_c^2}\cdot\frac{s+b-\left(1+\sqrt{2}\right)\omega_c}{(s+b)^2+\omega_c^2}\,, \end{aligned} \tag{A.25}$$

woraus sich die z-Transformierte bestimmen lässt:

$$\begin{aligned} G_4(z) &= 6\,T^4\cdot\frac{1-r\left(\left(1+\sqrt{2}\right)\sin\theta+\cos\theta\right)z^{-1}}{1-2r\cos\theta z^{-1}+r^2z^{-2}}\cdot \\ &\quad\cdot\frac{1-r\left(\left(1-\sqrt{2}\right)\sin\theta+\cos\theta\right)z^{-1}}{1-2r\cos\theta z^{-1}+r^2z^{-2}}\cdot \\ &\quad\cdot\frac{1-r\left(-\left(1-\sqrt{2}\right)\sin\theta+\cos\theta\right)z^{-1}}{1-2r\cos\theta z^{-1}+r^2z^{-2}}\cdot \\ &\quad\cdot\frac{1-r\left(-\left(1+\sqrt{2}\right)\sin\theta+\cos\theta\right)z^{-1}}{1-2r\cos\theta z^{-1}+r^2z^{-2}} \\ &= 6\,T^4\cdot\frac{1-r\sqrt{4+2\sqrt{2}}\,\cos\left(\theta-\frac{3\pi}{8}\right)z^{-1}}{1-2r\cos\theta z^{-1}+r^2z^{-2}}\cdot \\ &\quad\cdot\frac{1-r\sqrt{4-2\sqrt{2}}\,\cos\left(\theta-\frac{\pi}{8}\right)z^{-1}}{1-2r\cos\theta z^{-1}+r^2z^{-2}}\cdot \\ &\quad\cdot\frac{1-r\sqrt{4-2\sqrt{2}}\,\cos\left(\theta+\frac{\pi}{8}\right)z^{-1}}{1-2r\cos\theta z^{-1}+r^2z^{-2}}\cdot \\ &\quad\cdot\frac{1-r\sqrt{4+2\sqrt{2}}\,\cos\left(\theta+\frac{3\pi}{8}\right)z^{-1}}{1-2r\cos\theta z^{-1}+r^2z^{-2}}\,. \end{aligned} \tag{A.26}$$

Zusammenfassend kann Folgendes festgestellt werden: Die Übertragungsfunktion eines digitalen reellwertigen „Kosinus-Phasen“ Gammaton-Filters (mit Anfangsphase $\phi = 0$) besitzt m identische konjugiert komplexe Polpaare bei $z_p = re^{\pm j\theta}$ und m reelle Nullstellen. Die Lage der Nullstellen auf der reellen Achse wird durch die normierte Kreisfrequenz $\theta = \omega_c \cdot T$ und $r = e^{-bT}$ bestimmt. Dies stimmt mit den Ergebnissen der Studien von Lyon und Mead (1988b) überein. Die z-Transformierte eines „Sinus-Phasen“ Gammaton-Filters (mit Anfangsphase $\phi = \pi/2$) besitzt hingegen nur $m - 1$ Nullstellen, bei gleicher Anzahl konjugiert komplexer Polpaare (s. a. Slaney, 1993; Zotter, 2004, Anhang A).

In dieser Arbeit liegt das Hauptaugenmerk auf der recheneffizienten digitalen Implementierung von GTF als Teilbandfilter einer Analyse-Synthese-Filterbank zur auditiven Signalverarbeitung. Aus diesem Grund werden hier ausschließlich Gammaton-Filter bis zur 4ten Ordnung betrachtet, da diese die auditiven Maskierungskurven mit ausreichender Genauigkeit annähern und sehr recheneffizient mit rekursiven Filterstrukturen implementiert werden können (s. Patterson et al., 1992). Die z-Transformierte von Gammaton-Filtern höherer Ordnung kann jedoch in gleicher Weise wie in den vorhergehenden Abschnitten bestimmt werden.

A.3.2 Bilineare Transformation

Bei der bilinearen Transformation wird die linke Hälfte der s-Ebene der Laplace-Transformierten eines zeitkontinuierlichen Systems auf das Innere des Einheitskreises der z-Ebene der z-Transformierten eines zeitdiskreten Systems abgebildet. Die $j\omega$-Achse wird dabei direkt auf den Einheitskreis abgebildet. Die bilineare Transformation ist wie folgt definiert (siehe z. B. Oppenheim et al., 1998, Kap. 7.1.2):

$$s = \frac{2}{T}\frac{1 - z^{-1}}{1 + z^{-1}}\,. \tag{A.27}$$

Im Gegensatz zur Impulsinvarianz-Transformation ist die bilineare Transformation eine umkehrbare Abbildung der s-Ebene auf die z-Ebene. Es treten somit keine störenden Aliasing-Artefakte auf.

Mit Hilfe der bilinearen Transformation kann die z-Transformierte eines Gammaton-Filters der Ordnung m direkt hergeleitet werden, indem s in Gl. (A.8) durch den Ausdruck in Gl. (A.27) substituiert wird:

$$
\begin{aligned}
G_m(z) &= (-1)^{m-1}\frac{(m-1)!}{2}\cdot\frac{\left(\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}}+b-j\omega_c\right)^m+\left(\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}}+b+j\omega_c\right)^m}{\left(\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}}+b-j\omega_c\right)^m\left(\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}}+b+j\omega_c\right)^m} \\
&= (-1)^{m-1}\frac{(m-1)!}{2}\frac{\left(\frac{T}{2}(1+z^{-1})\right)^m}{\left(\frac{1}{\left(1+\frac{2}{T}b\right)^2+\left(\frac{2}{T}\omega_c\right)^2}\right)^m}\cdot \\
&\quad\cdot\left[\left(1+\frac{T}{2}b-\frac{T}{2}j\omega_c+\left(\frac{T}{2}b-\frac{T}{2}j\omega_c-1\right)z^{-1}\right)^m+\right. \\
&\quad\left.+\left(1+\frac{T}{2}b+\frac{T}{2}j\omega_c+\left(\frac{T}{2}b+\frac{T}{2}j\omega_c-1\right)z^{-1}\right)^m\right]\cdot \\
&\quad\cdot\frac{1}{\left(1-\frac{1+\frac{T}{2}j\omega_c-\frac{T}{2}b}{1-\frac{T}{2}j\omega_c+\frac{T}{2}b}z^{-1}\right)^m\left(1-\frac{1-\frac{T}{2}j\omega_c-\frac{T}{2}b}{1+\frac{T}{2}j\omega_c+\frac{T}{2}b}z^{-1}\right)^m}
\end{aligned}
\tag{A.28}
$$

Die Übertragungsfunktion $G_m(z)$ in Gl. (A.28) besitzt m konjugiert komplexe Polstellen bei $z_p = \left(1 \pm \frac{T}{2}j\omega_c - \frac{T}{2}b\right) / \left(1 \mp \frac{T}{2}j\omega_c + \frac{T}{2}b\right)$. Es entsteht kein Aliasing.

Im Vergleich dazu werden bei der Impulsinvarianz-Transformation die m Polpaare auf $z_p = e^{(-b+j\omega_c)T}$ abgebildet. Allerdings sind Gammaton-Funktionen ausreichend bandbegrenzt und die durch das Aliasing auftretenden Abbildungsfehler können vernachlässigt werden. Zudem kann nach Cooke (1991, Kap. 2.4) mit der Impulsinvarianz-Transformation eine wesentlich bessere Annäherung an die gemessenen auditiven Filterkurven erreicht werden als mit der bilinearen Transformation. Letztere hat also keine wesentlichen Vorteile gegenüber der Impulsinvarianz-Transformation und sei hier nur vollständigkeitshalber erwähnt. Im Rahmen dieser Arbeit wird die bilineare Transformation nicht weiter auf den Entwurf zeitdiskreter GTF angewendet.

A.4 All-Pol und One-Zero Gammaton-Filter

Bei den von Slaney (1993) und Lyon (1996) vorgeschlagenen All-Pol Gammaton-Filtern (APGF) werden die Nullstellen der Übertragungsfunktion auf den Koordinatenursprung der z-Ebene verschoben. Dadurch lässt sich die Gammaton-Übertragungsfunktion in Gl. (A.14) auf ein Paar konjugiert komplexer Pole m-ter Ordnung vereinfachen:

$$G_{\mathrm{APGF},m}(s) = \frac{\hat{a}_{\mathrm{APGF},m}}{\left((s+b)^2+\omega_c^2\right)^m}\,, \tag{A.29}$$

wobei der Skalierungsfaktor $\hat{a}_{\mathrm{APGF},m}$ dazu dient, die Gesamtverstärkung bei der Frequenz ω_c auf 0 dB zu normieren. Aus Gl. (A.29) ist ersichtlich, dass sich APGF sehr einfach und recheneffizient durch eine Kaskade von m Zweipol-Filtern[149]

$$G_{\mathrm{APGF},1}(s) = \frac{\hat{a}_{\mathrm{APGF},1}}{(s+b)^2+\omega_c^2} \tag{A.30}$$

implementieren lassen. Die Impulsantwort dieses Filters ist eine Gammaton-Funktion (s. Lyon, 1996)

$$g_{\mathrm{APGF},1}(t) = a_{\mathrm{APGF},1}\, e^{-bt} \sin(\omega_c t)\,. \tag{A.31}$$

mit $a_{\mathrm{APGF},1} = \hat{a}_{\mathrm{APGF},1}/\omega_c$ und Anfangsphase $\phi = -\pi/2$.

Die All-Pol-Filterstruktur ist den Kaskadenmodellen zur Simulation der Wanderwellenausbreitung entlang der Cochlea (vgl. Lyon und Mead, 1988b; Kates, 1993, 1995) sehr ähnlich und stellt somit eine Verbindung zu den physiologisch motivierten Modellen der cochleären Verarbeitung dar. Lyon (1996) hat gezeigt, dass die Asymmetrie der APGF Amplitudengänge (bei Mittenfrequenzen $\omega_c < \pi/2$) wesentlich besser den experimentalpsychologisch gemessenen Daten als die Amplitudengänge regulärer GTF entspricht. Zudem lassen sich APGF einfach parametrisieren und relativ gut and das nichtlineare Verhalten der experimentell ermittelten auditiven Filter anpassen.

[149] IIR-Filter 2ter Ordnung (*second order section*, SOS).

Die Herleitung eines zeitdiskreten APGF kann über die Übertragungsfunktion eines allgemeinen IIR Allpass-Filters mit konjugiert komplexen Polen bei $z_p = re^{\pm j\theta}$ erfolgen:

$$G_{\mathrm{APGF},1}(z) = \frac{1}{(1 - re^{j\theta}z^{-1})\,(1 - re^{-j\theta}z^{-1})}\,. \tag{A.32}$$

Aus der Partialbruchzerlegung der Übertragungsfunktion in Gl. (A.32)

$$\begin{aligned} G_{\mathrm{APGF},1}(z) &= \frac{e^{j\theta}}{e^{j\theta} - e^{-j\theta}}\frac{1}{(1 - re^{j\theta}z^{-1})} - \frac{e^{-j\theta}}{e^{j\theta} - e^{-j\theta}}\frac{1}{(1 - re^{-j\theta}z^{-1})} \\ &= \frac{1 - r\cos\theta\, z^{-1}}{1 - 2r\cos\theta\, z^{-1} + r^2z^{-2}} + \cot\theta\frac{r\sin\theta\, z^{-1}}{1 - 2r\cos\theta\, z^{-1} + r^2z^{-2}} \end{aligned} \tag{A.33}$$

ergibt sich unter Anwendung der inversen z-Transformation die zeitdiskrete Impulsantwort eines APGF erster Ordnung zu

$$\begin{aligned} g_{\mathrm{APGF},1}(n) &= r^n\frac{\sin\left(\theta(n+1)\right)}{\sin\theta}u(n) \\ &= r^n\left(\cos(\theta n) + \cot(\theta)\sin(\theta n)\right)u(n)\,. \end{aligned} \tag{A.34}$$

Aus dem Vergleich von Gl. (A.34) mit Gl. (A.9) für $m = 1$ wird ersichtlich, dass der Phasenterm des APGF von der Filtermittenfrequenz abhängt. Ähnlich den Gammachirp-Filtern (vgl. Irino und Patterson, 1997; Irino und Unoki, 1999) ist die Trägerschwingung des APGF frequenzmoduliert und lässt sich somit besser auf das Schwingungsverhalten der Basilarmembran abstimmen. Dieses Verhalten lässt sich auch durch Zusammenschalten eines Kosinus-Phasen und eines Sinus-Phasen GTF erreichen.

APGF höherer Ordnung können durch eine Kaskade von m Zweipol-Filtern, siehe Gl. (A.32), realisiert werden:

$$G_{\mathrm{APGF},m}(z) = \frac{1}{(1 - 2r\cos\theta\, z^{-1} + r^2z^{-2})^m}\,. \tag{A.35}$$

Die Impulsantwort des Gesamtsystems lässt sich entweder durch Faltung von m zeitdiskreten Impulsantworten, siehe Gl. (A.34), oder durch Partialbruchzerlegung

der Übertragungsfunktion in Gl. (A.35) und inverser z-Transformation ableiten. Für höhere Ordnungen ist meist nur eine numerische Lösung möglich.

Eine Variante der APGF sind die sogenannten One-Zero Gammaton-Filter (OZGF). Durch das Differenzieren eines APGF (d. h. durch das Verschieben einer einzelnen Nullstelle auf $z = 1$) kann eine steilere Flanke des Amplitudengangs zu tieferen Frequenzen hin erreicht werden. Mit OZGF lassen sich die experimentalpsychologisch gemessenen Filterkurven wesentlich besser annähern als mit APGF (s. a. Flanagan, 1960b; Lyon, 1996). Die Übertragungsfunktion eines OZGF m-ter Ordnung kann wie folgt beschrieben werden:

$$G_{\mathrm{OZGF},m}(z) = \frac{1 - z^{-1}}{(1 - 2r\cos\theta\, z^{-1} + r^2 z^{-2})^m}\,. \tag{A.36}$$

OZGF sind analytisch nicht so einfach handhabbar wie APGF und aufwendiger in der Implementierung. Werden OZGF als Teilfilter einer parallelen Filterbank verwendet, kann jedoch die gemeinsame Nullstelle als Vorfilter implementiert werden. Die Teilfilter können dann zur Steigerung der Recheneffizienz als APGF ausgeführt werden.

A.5 Gammaton-Filter Parameter

In diesem Abschnitt werden die wichtigsten Parameter zur Steuerung des Übertragungsverhaltens von Gammaton-Filtern zusammengefasst. Beim Entwurf von Filterbänken ist neben der Bandbreite vor allem die Bestimmung der Phase an den Übergangsfrequenzen benachbarter Teilbänder von Bedeutung. Bei geeigneter Phasenwahl in den sich überlappenden Bereichen der Teilfilter kann mittels einfacher Summation der Teilbänder eine nahezu fehlerfreie Rekonstruktion des Signals erreicht werden.

A.5.1 Bandbreite

Unter der Annahme, dass die Amplitude $|H(e^{j(\theta+\Delta\theta)})|$ eines Gammaton-Filters in der Umgebung $\Delta\theta$ der Resonanzfrequenz θ nur von der Distanz zur nächstgelege-

nen m-fachen Polstelle[150] und einem Verstärkungsfaktor $\hat{a}$ abhängt (vgl. Van Compernolle, 1991), kann diese bei der Frequenz $\theta + \Delta\theta$ wie folgt über die Tangente linear angenähert werden (s. a. Zotter, 2004, Anhang A):

$$|H(e^{j(\theta+\Delta\theta)})| \simeq \frac{\hat{a}}{\prod_{k=1}^{m} \sqrt{(1-r)^2 + \Delta\theta^2}}\,. \tag{A.37}$$

Aus Gl. (A.37) lässt sich die $-3\,\mathrm{dB}$ Bandbreite wie folgt ableiten:

$$|H(e^{j\theta})|^2 \stackrel{!}{=} 2|H(e^{j(\theta+\frac{\Delta\theta_b}{2})})|^2$$

$$\frac{\hat{a}}{\prod_{k=1}^{m}(1-r)^2} \stackrel{!}{=} \frac{2\hat{a}}{\prod_{k=1}^{m}\left[(1-r)^2 + \left(\frac{\Delta\theta_b}{2}\right)^2\right]}$$

$$\prod_{k=1}^{m} \sqrt[m]{2}(1-r)^2 = \prod_{k=1}^{m}\left[(1-r)^2 + \left(\frac{\Delta\theta_b}{2}\right)^2\right]$$

$$r = 1 - \frac{\Delta\theta_b}{2\sqrt{\sqrt[m]{2}-1}}\,, \tag{A.38}$$

$$\Delta\theta_b = 2(1-r)\sqrt{\sqrt[m]{2}-1}\,. \tag{A.39}$$

Mit Gl. (A.39) kann die Bandbreite eines Gammaton-Filters sehr einfach und recheneffizient bestimmt werden. Ein sehr ähnlicher Ausdruck wurde von Van Compernolle (1991, Gl. 2.6c) zur Berechnung der $-3\,\mathrm{dB}$ Bandbreite zeitkontinuierlicher Gammaton-Filter hergeleitet.

A.5.2 Bandgrenze

Beim Entwurf einer Filterbank können die Bandgrenzen (d. h. die Überlappungspunkte benachbarter Teilbänder) auf unterschiedliche Weise definiert werden. Eine einfache Möglichkeit besteht darin das Frequenzband in gleiche Teile aufzuteilen. Dabei kann ein zum $|H(e^{j\theta})| = \frac{1}{2}$ Punkt symmetrischer Frequenzgang gefordert

[150] Zur Bestimmung der Amplitude reicht es den näher gelegenen Pol eines Polpaars zu berücksichtigen, da um die Mittenfrequenz (d. h. im Durchlassbereich des Filters) der Einfluss des weiter entfernten konjugiert komplexen Pols als konstant angenommen werden kann.

werden. Alternativ kann gefordert werden, dass die Summe der Leistungsübertragungsfunktionen $|H(e^{j\theta})|^2$ der Teilfilter konstant ist. Daraus folgt, dass benachbarte Filter zur Übergangfrequenz punktsymmetrische Leistungsübertragungsfunktionen aufweisen (s. Vary et al., 1998, Kap. 4.3.2). Bei Filterbänken zur auditiven Signalverarbeitung wird das Frequenzband hingegen in hörspezifische Frequenzgruppen (Tonheitsskala oder ERB-Skala) aufgeteilt. Dabei müssen die Bandbreiten der Teilfilter an die Frequenzgruppenbreite für die jeweilige Filtermittenfrequenz angepasst werden. Bei gegebenen Mittenfrequenzen lassen sich die Überlappungspunkte zweier APGF, wie folgend dargestellt, aus der Übertragungsfunktion bestimmen.

Analytische APGF. Die Amplitude des i-ten Teilfilters $|H_i(e^{j\omega})|$ wird, wie bereits im vorhergehenden Abschnitt angemerkt, durch die Distanz zur nächstgelegenen (m-fachen) Polstelle $p_i = r_i e^{j\omega_i}$ bestimmt. Aus den Übertragungsfunktionen zweier benachbarter APGF kann nun diejenige Frequenz bestimmt werden, an der die Amplituden $|H_1(e^{j\omega})|$ und $|H_2(e^{j\omega})|$ gleich sind (s. a. Zotter, 2004, Anhang A):

$$\begin{aligned} |H_1(e^{j\omega})|^2 &\stackrel{!}{=} |H_2(e^{j\omega})|^2 \\ \frac{(1-r_1)^{2m}}{\left|e^{j\omega} - r_1 e^{j\omega_1}\right|^{2m}} &= \frac{(1-r_2)^{2m}}{\left|e^{j\omega} - r_2 e^{j\omega_2}\right|^{2m}} \end{aligned} \tag{A.40}$$

$$(1-r_2)^2 \left\{ [\cos(\omega) - \mathrm{Re}(p_1)]^2 + [\sin(\omega) - \mathrm{Im}(p_1)]^2 \right\} = \left\{ [\cos(\omega) - \mathrm{Re}(p_2)]^2 + [\sin(\omega) - \mathrm{Im}(p_2)]^2 \right\} (1-r_1)^2$$

$$(1-r_2)^2 - 2\mathrm{Re}(p_1)(1-r_2)^2\cos(\omega) - 2\mathrm{Im}(p_1)(1-r_2)^2\sin(\omega) + r_1^2(1-r_2)^2 = (1-r_1)^2 - 2\mathrm{Re}(p_2)(1-r_1)^2\cos(\omega) - 2\mathrm{Im}(p_2)(1-r_1)^2\sin(\omega) + r_2^2(1-r_1)^2$$

$$\left[\mathrm{Re}(p_2)(1-r_1)^2 - \mathrm{Re}(p_1)(1-r_2)^2\right]\cos(\omega) + \left[\mathrm{Im}(p_2)(1-r_1)^2 - \mathrm{Im}(p_1)(1-r_2)^2\right]\sin(\omega) = \frac{(1-r_1)^2(1+r_2^2) - (1-r_2)^2(1+r_1^2)}{2}$$

Obige Gleichung lässt sich wie folgt vereinfachen:

$$\mathrm{C}\cos(\omega) + \mathrm{D}\sin(\omega) = \mathrm{B} \qquad \Longrightarrow \qquad \mathrm{A}\cdot\sin(\omega+\phi) = \mathrm{B}\,, \tag{A.41}$$

$$\begin{aligned}
\mathrm{A} &= \sqrt{D^2 + C^2}\,, \\
\mathrm{B} &= \frac{(1-r_1)^2(1+r_2^2) - (1-r_2)^2(1+r_1^2)}{2}\,, \\
\mathrm{C} &= \mathrm{Re}(p_2)(1-r_1)^2 - \mathrm{Re}(p_1)(1-r_2)^2\,, \\
\mathrm{D} &= \mathrm{Im}(p_2)(1-r_1)^2 - \mathrm{Im}(p_1)(1-r_2)^2\,, \\
\phi &= \angle(C, D)\ .
\end{aligned}$$

Daraus berechnen sich die Bandgrenzen zu:

$$\begin{aligned}
\omega_1 &= 3\pi - \arcsin\left(\frac{B}{A}\right) - \phi\,, \\
\omega_2 &= 2\pi + \arcsin\left(\frac{B}{A}\right) - \phi\,.
\end{aligned} \tag{A.42}$$

Für analytische APGF sind die Bandgrenzen in Gl. (A.42) exakt. Bei reellwertigen APGF kann der Einfluss des konjugiert komplexen Pols jedoch nicht vernachlässigt werden. In diesem Fall kann mit folgendem Ansatz eine exakte Lösung hergeleitet werden.

Reellwertige APGF. Die Leistungsübertragungsfunktion $|H_i(e^{j(\omega)})|^2$ eines reellwertigen APGF mit einem konjugiert komplexen Polpaar $p_i = r_i e^{\pm j\theta_i}$ kann wie folgt vereinfacht werden:

$$\begin{aligned}
|H_i(e^{j\omega})|^2 &= \frac{g_i^{2m}}{\left|e^{j\omega} - r_1 e^{j\theta_i}\right|^{2m}\left|e^{j\omega} - r_i e^{-j\theta_i}\right|^{2m}} \\
&= \frac{g_i^{2m}}{\left[\left(\cos(\omega) - \mathrm{Re}(p_i)\right)^2 + \left(\sin(\omega) - \mathrm{Im}(p_i)\right)^2\right]^m \left[\left(\cos(\omega) - \mathrm{Re}(p_i)\right)^2 + \left(\sin(\omega) + \mathrm{Im}(p_i)\right)^2\right]^m} \\
&= \frac{g_i^{2m}}{\left\{\left[(1 + r_i^2 - 2\mathrm{Re}(p_i)\cos(\omega)) - 2\mathrm{Im}(p_i)\sin(\omega)\right]\left[(1 + r_i^2 - 2\mathrm{Re}(p_i)\cos(\omega)) + 2\mathrm{Im}(p_i)\sin(\omega)\right]\right\}^m} \\
&= \frac{g_i^{2m}}{\left((1+r_i^2)^2 - 4\mathrm{Re}(p_i)(1+r_i^2)\cos(\omega) + 4r_i^2\cos^2(\omega) - 4\mathrm{Im}^2\{p_i)\right)^m}
\end{aligned}$$

$$= \frac{g_i^{2m}}{4^m \left(r_i^2 \cos^2(\omega) - (1+r_i^2)\mathrm{Re}(p_i)\cos(\omega) + \mathrm{Re}(p_i) + \frac{\left(1-r_i^2\right)^2}{4} \right)^m}$$

Mit der Bedingung gleicher Amplitude an den Bandgrenzen

$$|H_1(e^{j\omega})|^2 \stackrel{!}{=} |H_2(e^{j\omega})|^2$$

kann wiederum die Übergabefrequenz bestimmt werden:

$$g_2^2 \left(r_1^2 \cos^2(\omega) - (1+r_1^2)\mathrm{Re}(p_1)\cos(\omega) + \mathrm{Re}(p_1) + \frac{(1-r_1^2)^2}{4} \right) = g_1^2 \left(r_2^2 \cos^2(\omega) - (1+r_2^2)\mathrm{Re}(p_2)\cos(\omega) + \mathrm{Re}(p_2) + \frac{(1-r_2^2)^2}{4} \right)$$

$$\left[g_2^2 r_1^2 - g_1^2 r_2^2\right] \cos^2(\omega) + \left[g_1^2(1+r_2^2)\mathrm{Re}(p_2) - g_2^2(1+r_1^2)\mathrm{Re}(p_1)\right] \cos(\omega) + g_2^2 \left(\mathrm{Re}(p_1) + \frac{(1-r_1^2)^2}{4} \right) - g_1^2 \left(\mathrm{Re}(p_2) + \frac{(1-r_2^2)^2}{4} \right) = 0$$

Obige Gleichung lässt sich wie folgt vereinfachen:

$$\mathrm{A}\cos^2(\omega) + \mathrm{B}\cos(\omega) + \mathrm{C} = 0 \tag{A.43}$$

mit

$$\begin{aligned} \mathrm{A} &= g_2^2 r_1^2 - g_1^2 r_2^2 \,, \\ \mathrm{B} &= g_1^2(1+r_2^2)\mathrm{Re}(p_2) - g_2^2(1+r_1^2)\mathrm{Re}(p_1) \,, \\ \mathrm{C} &= g_2^2 \left(\mathrm{Re}(p_1) + \frac{(1-r_1^2)^2}{4} \right) - g_1^2 \left(\mathrm{Re}(p_2) + \frac{(1-r_2^2)^2}{4} \right) . \end{aligned}$$

Die Bandgrenzen ergeben sich aus den reellwertigen Wurzeln der quadratischen Gleichung (A.43) in einem Frequenzbereich $[0, 2\pi[$

$$\omega = \pm \arccos \left[\frac{1}{2\mathrm{A}} \left(-\mathrm{B} \pm \sqrt{\mathrm{B}^2 - 4\mathrm{AC}} \right) \right] . \tag{A.44}$$

Diese Lösung ist für reellwertige APGF exakt.

A.5.3 Übergangsphase an den Bandgrenzen

Mit Hilfe einer Synthese-Filterbank kann das breitbandige Originalsignal aus den Teilbandsignalen rekonstruiert werden. Im einfachsten Fall erfolgt dies über eine Summation der Teilbandsignale. Allerdings kann es dabei aufgrund von Phasendifferenzen an den Teilbandgrenzen zu Signalverzerrungen und Signalauslöschungen kommen. Im Folgenden wird ein Kriterium hergeleitet, welches die Notwendigkeit eines Vorzeichenwechsels bei der Summation aufeinanderfolgender Teilbänder einer Gammaton-Filterbank bestimmt. Dadurch lassen sich die Signalverzerrungen minimieren und es ergibt sich eine nahezu perfekte Rekonstruktion des breitbandigen Originalsignals (s. a. Zotter, 2004; Noisternig et al., 2009).

Zur Vereinfachung wird wiederum angenommen, dass die Amplitude an der Bandgrenze hauptsächlich von der Distanz zur nächstgelegenen Polstelle abhängt. In der Umgebung der Mittenfrequenz lässt sich die Laplace-Transformierte eines Gammaton-Filters erster Ordnung (mit einem konjugiert komplexen Polpaar) durch eine Übertragungsfunktion mit nur einer Polstelle approximieren:

$$H(s) = \frac{b}{s + b - j\omega_c}\,. \tag{A.45}$$

Mit dieser Vereinfachung kann jene Frequenz $s = j\omega$ eines Gammaton-Filters m-ter Ordnung mit einer Mittenfrequenz ω_c berechnet werden, bei der die Amplitude der Übertragungsfunktion einen bestimmten Wert A annimmt:

$$\begin{aligned}
|H(s)|^m &\stackrel{!}{=} A \\
\frac{b^m}{\prod_{k=1}^{m} \sqrt{(\omega - \omega_c)^2 + b^2}} &= A \\
\frac{b^m}{\prod_{k=1}^{m} \sqrt{\Delta\omega_m^2 + b^2}} &= A \\
\frac{b}{\sqrt{\Delta\omega_m^2 + b^2}} &= \sqrt[m]{A} \\
\Delta\omega_m &= b\sqrt{\frac{1}{\sqrt[m]{A^2}} - 1}\,.
\end{aligned} \tag{A.46}$$

Für ein Filter erster Ordnung ergibt sich die Lösung

$$\Delta\omega_1 = b\sqrt{\frac{1}{A^2} - 1}\,. \tag{A.47}$$

Daraus lässt sich die Phasenlage $\angle H^m(s)$ bei der Frequenz $s = j(\omega_c + \Delta\omega_m)$ bestimmen

$$\begin{aligned}
\angle H^m(s) &= m \cdot \angle H(s) \\
&= m \cdot \arctan\left(\frac{\omega - \omega_c}{b}\right)\Bigg|_{\omega=\omega_c+\Delta\omega_m} \\
&= m \cdot \arctan\left(\frac{\omega_c + \Delta\omega_m - \omega_c}{b}\right) \\
&= m \cdot \arctan\left(\frac{b\sqrt{\frac{1}{\sqrt[m]{A^2}} - 1}}{b}\right) \\
\angle H^m(s) &= m \cdot \arctan\left(\sqrt{\frac{1}{\sqrt[m]{A^2}} - 1}\right)
\end{aligned}$$

und daraus die Phasendifferenz an der Bandgrenze zweier sich überlappender Teilbandfilter der Gammaton-Filterbank berechnen:

$$\Delta\angle H_{k,k+1}(s) = \angle\, H_k^m(s)\big|_{\omega=\omega_c+\Delta\omega_m} - \angle\, H_{k+1}^m(s)\big|_{\omega=\omega_c-\Delta\omega_m}$$

$$\Delta\angle H_{k,k+1}(s) = 2m \cdot \arctan\left(\sqrt{\frac{1}{\sqrt[m]{A^2}} - 1}\right)\,.$$

Daraus kann ein einfaches Kriterium, welches die Notwendigkeit eines Vorzeichenwechsels bei der Summation aufeinanderfolgender Teilbänder einer Gammaton-Filterbank bestimmt, abgeleitet werden:

$$f_{\mathrm{sgn}}\left(H_{k,k+1}(s)\right) = \begin{cases} 1 & \forall \quad 2m \cdot \arctan\left(\sqrt{\frac{1}{\sqrt[m]{A^2}} - 1}\right) \pmod{\pi} < \frac{\pi}{2} \\ -1 & \forall \quad 2m \cdot \arctan\left(\sqrt{\frac{1}{\sqrt[m]{A^2}} - 1}\right) \pmod{\pi} \geq \frac{\pi}{2} \end{cases} \tag{A.48}$$

Die Amplitude A an der Bandgrenze lässt sich über den Dämpfungsfaktor C_{dB}, der das Verhältnis von A zur Amplitude der Übertragungsfunktion bei

der Filtermittenfrequenz ω_c darstellt, ausdrücken. Durch Umformulieren der Nebenbedingungen kann das Kriterium für den Vorzeichenwechsel in geschlossener Form dargestellt werden (s. a. Noisternig et al., 2009),

$$f_{\mathrm{sgn}}\left(H_{k,k+1}(s)\right) = \mathrm{sgn}\left\{\cos\left(2m \cdot \arctan\left(\sqrt{10^{-\frac{C_{\mathrm{dB}}}{10m}} - 1}\right)\right)\right\}, \tag{A.49}$$

wobei zur Vereinfachung die Amplitude bei der Mittenfrequenz auf 0 dB normiert wird.

A.5.4 Normierung

Die Verstärkungsfaktoren zur Normierung der Amplitude bei der Mittenfrequenz auf 0 dB können in geschlossener Form aus den Übertragungsfunktionen berechnet werden.

A APGF Verstärkungsfaktor

Der Verstärkungsfaktor $g_{k,\mathrm{APGF}}$ zur Normierung der Amplitude eines APGF bei der Mittenfrequenz θ_k berechnet sich zu:

$$\begin{aligned}
g_{\mathrm{APGF}}^{(k)} &= \frac{1}{|H(e^{j\theta_k})|} \\
&= \left|\left(1 - r_k e^{j\theta_k} e^{-j\theta_k}\right)\left(1 - r_k e^{-j\theta_k} e^{-j\theta_k}\right)\right| \\
&= \left|(1 - r_k)\left(1 - r_k e^{-j2\theta_k}\right)\right| \\
&= (1 - r_k)\sqrt{\left(1 - r_k \cos(2\theta_k)\right)^2 + r_k^2 \sin^2(2\theta_k)} \\
&= (1 - r_k)\sqrt{1 - 2r_k \cos(2\theta_k) + r_k^2}\,.
\end{aligned}$$

B OZGF Verstärkungsfaktor

Der Verstärkungsfaktor $g_{i,\mathrm{OZGF}}$ zur Normierung der Amplitude von OZGF berücksichtigt hingegen die Nullstelle bei $z = 1$:

$$g_{\mathrm{OZGF}}^{(k)} = \frac{g_{\mathrm{APGF}}^{(k)}}{\left((1-\cos\theta_k)^2+\sin^2\theta_k\right)^{\frac{1}{2m}}}$$

$$= \frac{g_{\mathrm{APGF}}^{(k)}}{(2-2\cos\theta_k)^{\frac{1}{2m}}}\,.$$

C GTF Verstärkungsfaktor

Die z-Transformierte linearer Gammaton-Filter besitzt m Nullstellen auf der reellen Achse. Mit steigenden Mittenfrequenzen nähern sich die Nullstellen $z=-1$, mit fallenden Mittenfrequenzen $z=1$. Der empirisch ermittelte Verstärkungsfaktor $g_{\mathrm{GTF}}^{(k)}$ stimmt sehr gut mit den Amplituden von GTF bis zur Ordnung $m=4$ überein. Für höhere Filterordnungen ist dieser jedoch nicht geeignet.

$$g_{\mathrm{GTF}}^{(k)} = \frac{g_{\mathrm{APGF}}^{(k)}}{(2-2\cos\theta_k)^{\frac{1}{5}}\,(2+2\cos\theta_k)^{\frac{1}{4}}}. \tag{A.50}$$

B

Wichtige Funktionen zur Berechnung modaler Beamformer

B.1 Assoziierte Legendre-Funktionen

B.1.1 Definition

Die assoziierten Legendre-Funktionen lassen sich für eine Funktion $x \in [-1, 1]$ und für alle $n, m \in \mathbb{N}$ wie folgt definieren:

$$P_n^m(x) = \begin{cases} (-1)^m (1-x^2)^{\frac{m}{2}} \frac{d^m}{dx^m} P_n(x), & \text{für} \quad n \geq m \\ 0, & \text{für} \quad m > n \end{cases} \tag{B.1}$$

$$P_n^{-m}(x) = (-1)^m \frac{(n-m)!}{(n+m)!} P_n^m(x). \tag{B.2}$$

$P_n(x)$ ist das Legendre-Polynom n-ter Ordnung, der multiplikative Faktor $(-1)^m$ die Condon–Shortley-Phase.

B.1.2 Schmidtsche Halbnormalisierung

Die halbnormalisierten (*Schmidt semi-normalised*, SN3D) assoziierten Legendre-Funktionen lassen sich für eine Funktion $x \in [-1, 1]$ und für alle $n \in \mathbb{N}$ und $m \in \mathbb{Z}$

definieren als:

$$S_n^m(x) = \begin{cases} P_n(x), & \text{für} \quad m = 0 \\ (-1)^m \sqrt{2\frac{(n-m)!}{(n+m)!}} P_n^m(x), & \text{für} \quad m \neq 0 \end{cases} \tag{B.3}$$

B.1.3 Volle Normalisierung

Die voll normalisierten (*fully-normalised*, N3D) assoziierten Legendre-Funktionen lassen sich für eine Funktion $x \in [-1, 1]$ und für alle $n \in \mathbb{N}$ und $m \in \mathbb{Z}$ wie folgt definieren:

$$N_n^m(x) = \begin{cases} \sqrt{2n+1} P_n(x), & \text{für} \quad m = 0 \\ (-1)^m \sqrt{2(2n+1)\frac{(n-m)!}{(n+m)!}} P_n^m(x), & \text{für} \quad m \neq 0 \end{cases} \tag{B.4}$$

B.1.4 Numerische Berechnung der Legendre-Polynome

Um die numerische Stabilität bei der Berechnung assoziierter Legendre-Polynome höherer Ordnungen zu gewährleisten, werden diese rekursiv berechnet. Es gilt die Rekurrenz (Holmes und Featherstone, 2002)

$$N_n^m(x) = a_n^m t N_{n-1}^m(x) - b_n^m N_{n-2}^m(x), \quad \forall 0 < m < n, \tag{B.5}$$

wobei

$$t = \cos x, \tag{B.6}$$

$$a_n^m = \sqrt{\frac{(2n-1)(2n+1)}{(n-m)(n+m)}}, \tag{B.7}$$

$$b_n^m = \sqrt{\frac{(2n+1)(n+m-1)(n-m-1)}{(n-m)(n+m)(2n-3)}}. \tag{B.8}$$

Die Startwerte werden dabei wie folgt angenommen:

$$N_0^0(x) = 1, \tag{B.9}$$

$$N_1^1(x) = \sqrt{3}u \tag{B.10}$$

mit

$$u = \sin x \tag{B.11}$$

und

$$N_m^m(x) = u\sqrt{\frac{2m+1}{2m}}N_{m-1}^{m-1}(x), \quad \forall m > 1. \tag{B.12}$$

B.2 Sphärische Harmonische

B.2.1 Komplexwertige sphärische Harmonische

Die komplexwertigen sphärischen Harmonischen sind für alle $n, m \in \mathbb{N}$ mit $-n \leq m \leq n$ wie folgt definiert:

$$Y_n^m(\theta, \phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(\cos\theta)\, e^{im\phi}, \tag{B.13}$$

wobei P_n^m das (nicht normalisierte) assoziierte Legendre-Polynom bezeichnet. Formeln zur direkten Berechnung von SH niedriger Ordnung finden sich in Williams (1999, Tab. 6.51, S. 192–193). Höhere Ordnungen werden numerisch berechnet. Zudem gilt:

$$Y_n^{-m}(\theta, \phi) = (-1)^m Y_n^m(\theta, \phi)^* \tag{B.14}$$

Die halb normalisierten (SN3D) SH sind wie folgt definiert (vgl. Daniel, 2001):

$$(Y_n^m)^{SN3D}(\theta, \phi) = S_n^m(\cos\theta)e^{im\phi}. \tag{B.15}$$

S_n^m sind die halb normalisierten assoziierten Legendre-Funktionen.

Die voll normalisierten (N3D) SH berechnen sich wie folgt:

$$(Y_n^m)^{N3D}(\theta, \phi) = \sqrt{2n+1}N_n^m(\cos\theta)e^{im\phi}. \tag{B.16}$$

N_n^m sind die voll normalisierten assoziierten Legendre-Polynome.

B.2.2 Konvertierung unterschiedlicher Normierungen

Unter der Annahme numerischer Stabilität aller Berechnungen, lässt sich relativ einfach zwischen unterschiedlichen Normalisierungen $\mathcal{N}_1$ und $\mathcal{N}_2$ konvertieren:

$$(Y_n^m)^{\mathcal{N}_1}(\theta,\phi) = (\alpha_n^m)^{(\mathcal{N}_1\to\mathcal{N}_2)}\,(Y_n^m)^{\mathcal{N}_2}(\theta,\phi)\,, \tag{B.17}$$

mit der Reziprozität

$$(\alpha_n^m)^{(\mathcal{N}_1\to\mathcal{N}_2)} = \frac{1}{(\alpha_n^m)^{(\mathcal{N}_2\to\mathcal{N}_1)}} \tag{B.18}$$

und Transitivität

$$(\alpha_n^m)^{(\mathcal{N}_1\to\mathcal{N}_3)} = (\alpha_n^m)^{(\mathcal{N}_1\to\mathcal{N}_2)}\,\alpha^{(\mathcal{N}_2\to\mathcal{N}_3)}. \tag{B.19}$$

Es gelten folgende Operatoren (vgl. Daniel, 2001, S. 155–157):

- $NN \to SN3D$:

$$m = 0 \Rightarrow (\alpha_n^m)^{(NN\to SN3D)} = \frac{\sqrt{4\pi}}{\sqrt{2n+1}}, \tag{B.20}$$

$$m \neq 0 \Rightarrow (\alpha_n^m)^{(NN\to SN3D)} = (-1)^m\frac{\sqrt{8\pi}}{\sqrt{2n+1}}. \tag{B.21}$$

- $SN3D \to N3D$

$$\forall n,m \Rightarrow (\alpha_n^m)^{(SN3D\to N3D)} = \sqrt{2n+1} \tag{B.22}$$

- $SN2D \to N2D$

$$m = 0 \Rightarrow \left(\alpha_m^{\pm m}\right)^{(SN2D\to N2D)} = 1 \tag{B.23}$$

$$\forall m \neq 0 \Rightarrow \left(\alpha_m^{\pm m}\right)^{(SN2D\to N2D)} = \sqrt{2} \tag{B.24}$$

- $N3D \to N2D$

$$\forall m \Rightarrow \left(\alpha_m^{\pm m}\right)^{(N3D\to N2D)} = \sqrt{\frac{2^{2|m|}\left(|m|!\right)^2}{(2|m|+1)!}} \tag{B.25}$$

B.2.3 Reellwertige sphärische Harmonische

Die reellwertigen SH $\Upsilon_n^m(\theta,\phi)$ lassen sich wie folgt über die komplexwertigen SH $Y_n^m(\theta,\phi)$ bestimmen:

$$\begin{aligned}
\Upsilon_n^m(\theta,\phi) &= \tfrac{1}{2}\left[Y_n^m(\theta,\phi) + Y_n^m(\theta,\phi)^*\right], && \forall n\in\mathbb{N}, \forall m>0,\\
\Upsilon_n^0(\theta,\phi) &= Y_n^0(\theta,\phi), && \forall n\in\mathbb{N},\\
\Upsilon_n^m(\theta,\phi) &= \tfrac{-(-1)^m}{2i}\left[Y_n^m(\theta,\phi) - Y_n^m(\theta,\phi)^*\right], && \forall n\in\mathbb{N}, \forall m<0.
\end{aligned} \tag{B.26}$$

Daraus folgt:

$$\begin{aligned}
\Upsilon_n^m(\theta,\phi) &= \mathrm{Re}\left(Y_n^m(\theta,\phi)\right), && \forall n\in\mathbb{N}, m\geq 0,\\
\Upsilon_n^m(\theta,\phi) &= -(-1)^m \,\mathrm{Im}\left(Y_n^m(\theta,\phi)\right), && \forall n\in\mathbb{N}, m<0.
\end{aligned} \tag{B.27}$$

Die reellwertigen SH bilden kein vollständiges Orthonormalsystem auf der Einheitskugel

$$\|\Upsilon_n^m(\theta,\phi)\|^2 = \int_{\mathbb{S}^2} \Upsilon_n^m(\theta,\phi)^2 d\Omega = \frac{1}{2}\|Y_n^m\|^2. \tag{B.28}$$

(B.29)

Die reellwertige orthonormale Basis (*real orthonormal basis*, RONB) ist wie folgt definiert (vgl. Daniel, 2001, Tab. 3.1, S. 151):

$$\begin{aligned}
\Upsilon_n^m(\theta,\phi)^{RONB} &= \begin{cases}
\frac{1}{\sqrt{2}}\left[Y_n^m(\theta,\phi) + Y_n^m(\theta,\phi)^*\right], & \forall n\in\mathbb{N}, m>0,\\
Y_n^m(\theta,\phi), & \forall n\in\mathbb{N}, m=0,\\
\frac{-(-1)^m}{\sqrt{2}\mathbf{i}}\left[Y_n^m(\theta,\phi) - Y_n^m(\theta,\phi)^*\right], & \forall n\in\mathbb{N}, m<0,
\end{cases}\\
&= \begin{cases}
\sqrt{\frac{2n+1}{2\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta)\cos(m\phi), & \forall n\in\mathbb{N}, m>0,\\
\sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta), & \forall n\in\mathbb{N}, m=0,\\
-(-1)^m\sqrt{\frac{2n+1}{2\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta)\sin(m\phi), & \forall n\in\mathbb{N}, m<0.
\end{cases}
\end{aligned} \tag{B.30}$$

B.2.4 Konvertierung von komplexwertigen zu reellwertigen sphärischen Harmonischen

Ein Vektor $\mathbf{X}_{\mathbb{C}} \in Y$ lässt sich über die Matrix $\mathbf{S}_{\mathbb{C}\to\mathbb{R}}$ als $\mathbf{X}_{\mathbb{R}} \in \Upsilon$ abbilden:

$$\mathbf{X}_{\mathbb{R}} = \mathbf{S}_{\mathbb{C}\to\mathbb{R}}\mathbf{X}_{\mathbb{C}}, \tag{B.31}$$

$$\mathbf{S}_{\mathbb{R}\to\mathbb{C}} = \left(\mathbf{S}_{\mathbb{C}\to\mathbb{R}}\right)^{-1}. \tag{B.32}$$

Die Koeffizienten der Matrizen lassen sich mit

$$\begin{aligned} \Upsilon_n^m(\theta,\phi) &= \mathrm{Re}\left(Y_n^m(\theta,\phi)\right), && \forall m \geq 0, \\ \Upsilon_n^m(\theta,\phi) &= -(-1)^m \,\mathrm{Im}\left(Y_n^m(\theta,\phi)\right), && \forall m < 0, \end{aligned}$$

sehr einfach herleiten. Folgendes Beispiel zeigt die Koeffizienten für $N = 2$.

$$\mathbf{S}_{\mathbb{R}\to\mathbb{C}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.5\mathrm{i} & 0 & -0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.5\mathrm{i} & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5\mathrm{i} & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & -0.5\mathrm{i} & 0 & -0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.5\mathrm{i} & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & -0.5\mathrm{i} & 0 & 0 & 0 & 0.5 \end{pmatrix}, \quad \mathbf{S}_{\mathbb{C}\to\mathbb{R}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathrm{i} & 0 & \mathrm{i} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\mathrm{i} & 0 & 0 & 0 & \mathrm{i} \\ 0 & 0 & 0 & 0 & 0 & \mathrm{i} & 0 & \mathrm{i} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

B.3 Sphärische Bessel- und Hankel-Funktionen

B.3.1 Definition

Die sphärischen Bessel-Funktionen der ersten und zweiten Art sind für alle $x \in \mathbb{R}^{+*}$ und $n \in \mathbb{N}$ wie folgt definiert (vgl. Morse und Feshbach 1953, S. 1465 ff; Williams 1999, Kap. 6):

$$j_n(x) = \left(\frac{\pi}{2x}\right)^{1/2} J_{n+1/2}(x), \tag{B.33}$$

$$y_n(x) = \left(\frac{\pi}{2x}\right)^{1/2} Y_{n+1/2}(x), \tag{B.34}$$

die sphärischen Hankel-Funktionen der ersten und zweiten Art als

$$h_n^{(1)}(x) = j_n(x) + i y_n(x), \tag{B.35}$$

$$h_n^{(2)}(x) = j_n(x) - iy_n(x), \tag{B.36}$$
$$h_n^{(2)}(x) = (h_n^{(1)}(x))^*. \tag{B.37}$$

Die Abbildungen B.1 bis B.4 zeigen die sphärischen Bessel- und Hankel-Funktionen und ihre Ableitungen bis zur Ordnung $N = 8$.

Die sphärischen Hankel-Funktionen der ersten und zweiten Art lösen die Wellengleichung und können als einfallende bzw. auslaufende Schallwelle interpretiert werden.

$$h_n^{(1)}(kr) \propto e^{ikr},$$
$$h_n^{(2)}(kr) \propto e^{-ikr}.$$

Wie in Tourbabin und Rafaely (2015) gezeigt wird, hängt die Interpretation jedoch davon ab, ob die zeitliche Schwingung mit $e^{i\omega t}$ oder $e^{-i\omega t}$ angesetzt wird. Nehmen wir das von einer Punktschallquelle emittierte Schallfeld als Beispiel. Der durch eine Schallquelle an der Position $\mathbf{r}_s$ an der Position $\mathbf{r}$ hervorgerufene Schalldruck lässt sich beschreiben als:

$$P(k,\mathbf{r},\omega,t) = \mathrm{Re}\left\{\frac{e^{i(k|\mathbf{r}_s-\mathbf{r}|-\omega t)}}{|\mathbf{r}_s-\mathbf{r}|}\right\} = \mathrm{Re}\left\{\frac{e^{i(\omega t-k|\mathbf{r}_s-\mathbf{r}|)}}{|\mathbf{r}_s-\mathbf{r}|}\right\}. \tag{B.38}$$

Die Wellenzahl $k = \|\mathbf{r}\|$ ist in diesem Fall ein Skalar, da sich bei kugelförmiger Abstrahlung die Schallwelle in alle Raumrichtungen gleich ausbreitet. Wird die zeitliche Schwingung mit e^{ikr} angesetzt, ergibt sich für die Berechnung des Schalldrucks

$$\frac{e^{-ik|\mathbf{r}-\mathbf{r_s}|}}{|\mathbf{r}-\mathbf{r_s}|} \propto e^{-ikr} \propto h_n^{(2)}(kr). \tag{B.39}$$

Daraus folgt, dass die sphärische Hankel-Funktion der zweiten Art das Schallfeld der Punktquelle repräsentiert. Wird hingegen die zeitliche Schwingung mit e^{-ikr} angesetzt, ergibt sich

$$\frac{e^{ik|\mathbf{r}-\mathbf{r_s}|}}{|\mathbf{r}-\mathbf{r_s}|} \propto e^{ikr} \propto h_n^{(1)}(kr) \tag{B.40}$$

und die sphärische Hankel-Funktion der ersten Art ist die richtige Wahl.

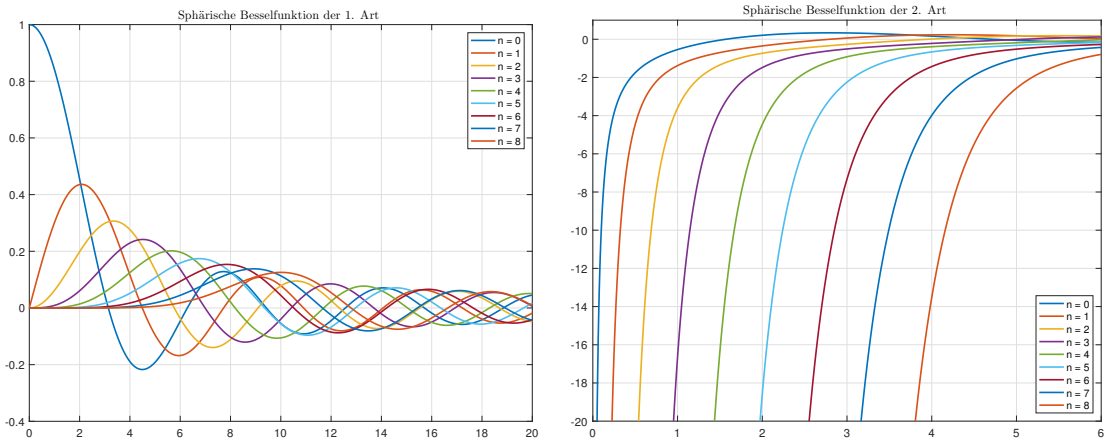


Abbildung B.1: Sphärische Bessel-Funktion der ersten Art, $j_n(x)$, (linkes Teilbild) und der zweiten Art, $y_n(x)$, (rechtes Teilbild) bis zur Ordnung $N = 8$.

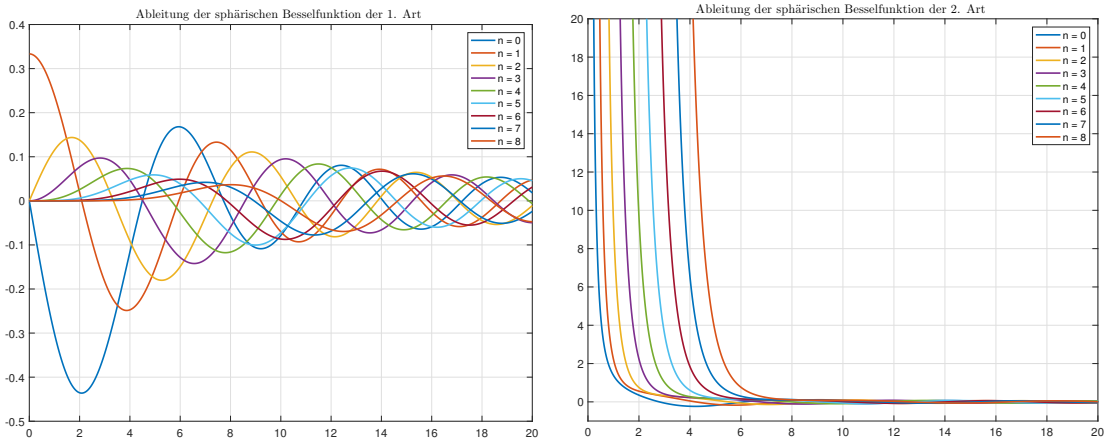


Abbildung B.2: Erste Ableitung der sphärischen Bessel-Funktion der ersten Art, $\frac{dj_n(x)}{dx}$, (linkes Teilbild) und der zweiten Art, $\frac{dy_n(x)}{dx}$, (rechtes Teilbild) bis zur Ordnung $N = 8$.

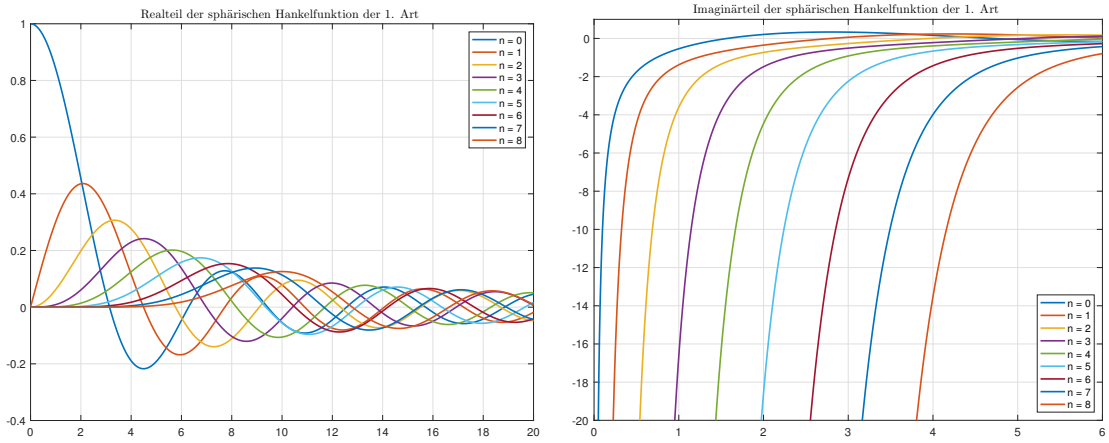


Abbildung B.3: Real- und Imaginärteil der sphärischen Hankel-Funktion der ersten Art $h_n^{(1)}(x)$ (linkes resp. rechtes Teilbild)

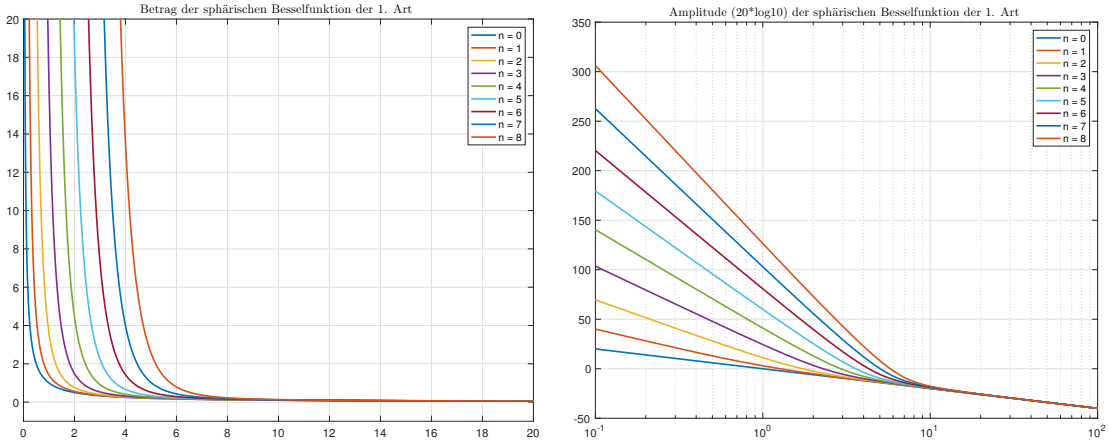


Abbildung B.4: Betrag (linkes Teilbild) und Amplitude (rechtes Teilbild) der sphärischen Hankel-Funktion der ersten Art $h_n^{(1)}(x)$.

B.3.2 Numerische Berechnung der sphärischen Bessel-Funktion

A Sphärische Bessel-Funktion der ersten Art

Für die Berechnung der sphärischen Bessel-Funktion der ersten Art gilt für alle $n \in \mathbb{N}^*$ folgende Rekurrenz (vgl. Barnett, 1996) :

$$j_{n-1}(x) = S_{n+1}(x)\, j_n(x) + j'_n(x), \tag{B.41}$$

$$j'_{n-1}(x) = S_{n-1}(x)\, j_{n-1}(x) - j_n(x), \tag{B.42}$$

wobei $j'_n(x)$ die erste Ableitung $\frac{d}{dx} j_n(x)$ der sphärischen Bessel-Funktion ist und $S_n(x) = \frac{n}{x}$. Die Startwerte der rekursiven Berechnung werden wie folgt angenommen:

$$j_0(x) = \frac{\sin(x)}{x}, \tag{B.43}$$

$$j'_0(x) = \frac{\cos(x)\, x - \sin(x)}{x^2}, \tag{B.44}$$

$$j_1(x) = \frac{\sin(x)}{x^2} - \frac{\cos(x)}{x}, \tag{B.45}$$

$$j'_1(x) = \frac{\cos(x)\, x^2 - \sin(x)\, 2x}{x^4} + \frac{\sin(x)\, x + \cos(x)}{x^2}. \tag{B.46}$$

Die rekursive Berechnung wird für $n >= x$ numerisch instabil. Es gilt folgende Rekurrenz

$$f_n(x) = \frac{j'_n(x)}{j_n(x)} = S_n(x) - \frac{j_{n+1}(x)}{j_n(x)}, \tag{B.47}$$

wobei sich $f_n(x)$ in folgenden Kettenbruch entwickeln lässt

$$f_n(x) = S_n(x) - \frac{1}{T_{n+1}(x) -} \frac{1}{T_{n+2}(x) -} \ldots, \tag{B.48}$$

mit $T_n(x) = \frac{2n+1}{x}$. Die numerische Berechnung von Kettenbrüchen ist ein viel diskutiertes Problem in den Computerwissenschaften. In dieser Arbeit wurde der Algorithmus von Lentz (1976) verwendet.

B Sphärische Bessel-Funktion der zweiten Art

Für die Berechnung der sphärischen Bessel-Funktion der zweiten Art gilt für alle $n \in \mathbb{N}^*$ folgende Rekurrenz:

$$y_0(x) = \frac{-\cos(x)}{x}, \tag{B.49}$$

$$y_1(x) = \frac{-\cos(x)}{x^2} - \frac{\sin(x)}{x}, \tag{B.50}$$

$$y_0'(x) = \frac{\cos(x) + x\sin(x)}{x^2}, \tag{B.51}$$

$$y_1'(x) = \frac{\sin(x)\,x^2 + 2x\cos(x)}{x^4} - \frac{x\cos(x) - \sin(x)}{x^2}. \tag{B.52}$$

Für $n > x$

$$y_n(x) = \frac{n-1}{x} y_{n-1}(x) - y_{n-1}'(x), \tag{B.53}$$

$$y_n'(x) = y_{n-1}(x) - \frac{n+1}{x} y_n(x), \tag{B.54}$$

für $n \leq x$

$$y_n(x) = \frac{2n-1}{x} y_{n-1}(x) - y_{n-2}(x). \tag{B.55}$$

C

Entwicklung eines modularen Systems zur Erfassung von Mikrofonarraydaten in Echtzeit

Die praxisnahe Evaluierung echtzeitfähiger Algorithmen zur breitbandigen Signalaufbereitung in Mehrkanal-Mikrofonanwendungen mit unterschiedlichen Arraygeometrien erfordert ein modulares und möglichst flexibel konfigurierbares System zur Erfassung der Vielzahl von Signalen der räumlich verteilten Sensoren. Dieser Abschnitt gibt einen kurzen Überblick über die im Rahmen dieser Arbeit entwickelten Hardwarekomponenten[151] (Mikrofonvorverstärkung, Signalkonditionierung, A/D-Wandlung, synchrone Datenübertragung über IP-basierte Netzwerke) und diskutiert die Anforderungen an den Aufbau eines 64-Kanal Mikrofonarrays.

Je nach Anwendung besteht ein Mikrofonarray aus einigen wenigen bis zu mehreren hunderten räumlich verteilten Mikrofonen. Die hohe Kanalanzahl und die oft langen Kabelwege limitieren den Einsatz analoger Übertragungstechniken. IP-

[151] Die Hardwarekomponenten wurden im Rahmen eines Kooperationsprojekts des *IInstitut de Recherche et Coordination Acoustique/Musique* (IRCAM-CNRS-UPMC, UMR 9912 STMS) Paris mit der Joanneum Research Forschungsgesellschaft und der Firma xFace in Graz entwickelt und durch das CONTINT Projekt „Sample Orchestrator 2“ (SOR2) des französischen Wissenschaftsfonds (*Agence Nationale de la Recherche*, ANR) teilfinanziert.

basierte Netzwerke[152] haben sich in den letzten Jahren rasant entwickelt und stellen eine gute Alternative zur analogen Audiosignalübertragung dar. Die hohen Datenraten ermöglichen die synchrone digitale Übertragung hochqualitativer, unkomprimierter Mehrkanal-Audiodaten mit minimaler Latenz. Mit Lichtwellenleitern lassen sich aufgrund der geringen Signaldämpfung große Reichweiten (ohne zusätzliche Signalverstärker) realisieren. Um die Vorteile moderner Netzwerktechnologien und Infrastrukturen nutzen zu können, sollte ein Datenprotokoll den IEEE 802.3 Gigabit-Ethernet-Standard (1000BASE-TX bzw. 1000BASE-SX) unterstützen und zu Standard-Hardwarekomponenten (wie zum Beispiel Switches und Router) kompatibel sein. Aufgrund der zur Übertragung mehrkanaliger Audiodaten erforderlichen hohen Datenrate[153] wird der 100-Mbit-Ethernet Standard (100BASE-T) hier nicht weiter betrachtet. Einige Mehrkanal-Mikrofonanwendungen erfordern zudem eine möglichst hohe zeitliche Auflösung, mit Abtastraten bis zu 192 kHz. Proprietäre Formate, wie zum Beispiel CobraNet (Cirrus Logic), Dante (Audinate), Ethersound (Digigram) oder SuperMAC/HyperMAC (Klark Teknik), erlauben bei voller Kanalanzahl meist nur Abtastraten bis 48 kHz. Höhere Abtastraten können durch Reduktion der Anzahl der zu übertragenden Audiokanäle erreicht werden. Brüel & Kjaers LAN-XI Datenerfassungssystem ermöglicht Abtastraten bis zu 262 kHz bei sehr hoher Kanalanzahl und erfüllt die an ein räumlich-zeitlich hochauflösendes akustisches Messsystem gestellten Anforderungen.

In der wissenschaftlichen Forschung ist die Verwendung offener Systeme, die durch einen modularen Aufbau und kontinuierliche Erweiterungen auch an zukünftige Applikationen angepasst werden können, von Vorteil. Aus diesem Grund wurde im Rahmen dieser Arbeit ein modulares Datenerfassungssystem für Mehrkanal-Mikrofonanwendungen entwickelt (s. a. Reitbauer et al., 2012; Reitbauer, 2012, Kap. 5). Die Übertragung der bis zu 64 Audiokanäle (mit 32 Bit Wortbreite und bis zu 192 kHz Abtastrate) erfolgt über ein Standard-Gigabit-Ethernet Netzwerk. Das System unterstützt den IEPE/ICP Interface-Standard[154], welcher die Kom-

[152] Das Internetprotokoll Version 4 (IPv4) stellt Basisdienste für die Übermittlung von Daten in Netzwerken bereit und ist im RFC-791 Standard der *Internet Engineering Task Force* (IETF) spezifiziert. Das neuere Protokoll Version 6 (IPv6) ist im IETF RFC-2460 Standard spezifiziert.

[153] Die Übertragung von 64 unkomprimierten Audiokanälen mit einer Abtastrate von 192 kHz benötigt eine Datenübertragungsrate von 413,2 MBit/s (vgl. Tab. C.3).

[154] IEPE/ICP (*Integrated Electronics Piezoelectric / Integrated Circuit Piezoelectric*) bezeichnet einen Standard für piezoelektrische Sensoren mit eingebautem Vorverstärker (Impedanzwandler)

patibilität mit einer Vielzahl von piezoelektrischen Sensoren und Messmikrofonen unterschiedlicher Hersteller gewährleistet. Dies erleichtert den Aufbau multimodaler Mehrkanalsysteme, die zum Beispiel bei der Schwingungsanalyse mit mehreren Beschleunigungssensoren und Anregung durch einen mit einem Kraftsensor ausgerüsteten Impulshammer zum Einsatz kommen. Ein vergleichbares Datenerfassungssystem mit IP-basierter Übertragung der Mehrkanal-Audiodaten wurde von Pollow et al. (2011) vorgestellt. Dieses System bietet allerdings weder eine offene Hardwarearchitektur noch unterstützt es den IEPE/ICP-Standard. Das vom *National Institute of Standards and Technology* (NIST) entwickelte Mark-III Mikrofonarray[155] bietet eine offene Hardwarearchitektur, ist jedoch für die Verwendung in Systemen zur Sprachsignalverarbeitung mit einer maximalen Abtastrate von 44,1 kHz konzipiert. Die Anpassung der Hardwarekomponenten des NIST-Arrays auf die Verarbeitung unkomprimierter Mehrkanal-Audiodaten mit hohen Abtastraten ist nur bedingt möglich und wesentlich aufwendiger als ein Neuentwurf der elektronischen Schaltkreise.

C.1 Hardwarekomponenten

Das im Rahmen dieser Arbeit entwickelte Mikrofonarray setzt sich aus mehreren Hardwarekomponenten zusammen, die über klar definierte Schnittstellen miteinander verbunden sind (s. Abb. C.1). Dieser durchgehend modulare Ansatz bietet optimale Adaptierbarkeit und hohe Flexibilität hinsichtlich Skalierung und Erweiterbarkeit des Systems. Das Datenerfassungssystem kann mit auf die jeweilige Anwendung optimierten Baugruppen ausgerüstet werden. Das Basissystem ist wie folgt spezifiziert:

Systemspezifikationen:

- 64 IEPE/ICP-kompatible Mikrofon-Eingangskanäle

und ist im IEEE Standard 1451.4 spezifiziert. Das IEPE/ICP-Prinzip erlaubt große Kabellängen bei Verwendung einfacher Verbindungskabel. Zur Versorgung des Vorverstärkers wird ein Konstantstrom von mindestens 2 mA benötigt. Der IEPE/ICP Standard wird von Herstellern oft auch als CCLD (*Constant Current Line Drive*), Isotron, Deltatron oder Piezotron bezeichnet.

[155] Siehe auch: *NIST Smart Space* Projekt (Stanford et al., 2003).

- Digitalisierung mit 24 Bit Pulse-Code-Modulation[156] (PCM)
- Unterstützte Abtastraten: 44,1 kHz, 48 kHz, 96 kHz und 192 kHz
- IP-Protokoll zur synchronen bidirektionalen Übertragung von bis zu 256 Audiokanälen (48 kHz/32 Bit) über 1000Base-TX IP-Netzwerke; höhere Abtastraten erfordern eine Reduktion der Anzahl der zu übertragenden Kanäle (bei 192 kHz/32 Bit können bspw. nur 128 Audiokanäle unidirektional übertragen werden)
- Samplegenaue Synchronisation und minimale Latenz (< 2 ms)[157]
- Signalverarbeitung mit FPGA (Xilinx Spartan-6)[158]
- I^2S/I^2C-Schnittstelle[159] zur Anbindung der A/D-Wandler an das FPGA-Board über eine FMC-Standardverbindung[160]
- UDP-Protokoll[161] zur Übertragung unkomprimierter Mehrkanal-Audiodaten
- UDP/ARP-Protokoll[162] zur Systemkonfiguration und Statusabfrage
- 64 Bit Zeitstempel zur Detektion von Paketverlusten
- Optionale MADI[163] oder "Dante" Interface-Karte

Abb. C.1 zeigt den schematischen Aufbau des entwickelten Datenerfassungssystems für Mehrkanal-Mikrofonanwendungen mit den wesentlichen Elektronikbaugruppen: (i) Elektret-Kondensatormikrofone mit IEPE/ICP-kompatiblen Mikrofonvorverstärkern, (ii) 4-Kanal Interface-Karten mit konstantstromversorgten

[156] Die PCM von Audiosignalen ist in der ITU-T G.711 Richtlinie spezifiziert.

[157] Übertragungszeit eines Datenpakets in einem lokalen Gigabit-Ethernet Netzwerk (*packet delivery time*) plus Verzögerung durch die A/D-Wandlung. Die zur nachfolgenden Audiosignalverarbeitung benötigten Signalbuffer werden hier nicht berücksichtigt.

[158] *Field Programmable Gate Arrays* (FPGA) können Information massiv parallel und mit möglichst kurzen Latenzzeiten verarbeiten.

[159] Die *Inter-Integrated Circuit Sound* (I^2S) Schnittstelle dient der seriellen Übertragung von Audiodaten und wird zur Anbindung von A/D-Wandlern an digitale Signalprozessoren verwendet. I^2S beruht auf dem seriellen *Inter-Integrated Circuit* (I^2C) Datenbus.

[160] Der *FPGA Mezzanine Card* (FMC) Standard zur Anbindung von Tochterkarten an Signalprozessoren ist in der ANSI/VITA 57.1 Spezifikation definiert.

[161] Das *User Datagram Protocol* (UDP) ist im IETF Standard RFC 768 spezifiziert.

[162] Das *Address Resolution Protocol* (ARP) ist im IETF Standard RFC 826 spezifiziert. In IPv6 Netzwerken wurde ARP durch das *Neighbor Discovery Protocol* (NDP) ersetzt, welches im IETF Standard RFC 4861 spezifiziert ist.

[163] Das *Multi Channel Audio Digital Interface* (MADI) ist im AES10-2003 Standard spezifiziert.

IEPE/ICP-Mikrofoneingängen, Signalverstärkern und A/D-Wandlern, (iii) Busleiterplatte zur Aufnahme der A/D-Wandler-Karten und Anbindung des FPGA, (iv) FPGA-Board zur Aufbereitung der Daten für die synchrone Übertragung über IP-basierte Netzwerke, und (v) stabilisiertes Netzteil (in der Abbildung nicht dargestellt).

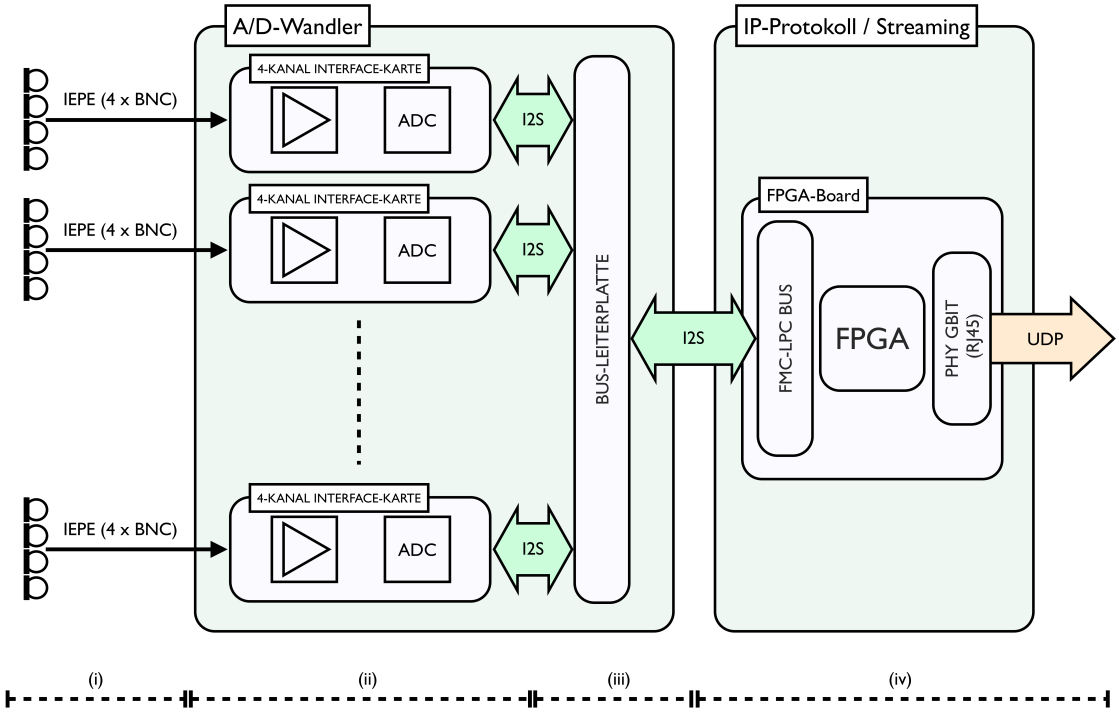


Abbildung C.1: Schematischer Aufbau des Datenerfassungssystems für Mehrkanal-Mikrofonanwendungen.

Beim Entwurf der elektronischen Schaltkreise und der Fertigung der Leiterplatten wurde vor allem Wert auf die Minimierung möglicher Signalverzerrungen und Maximierung des Signal-Störabstands gelegt. Der Klirrfaktor mit Rauschanteil[164] der Gesamtschaltung ohne Mikrofonkapsel soll den Wert von THD+N $\leq -100\,\mathrm{dB}$ nicht überschreiten und das Übersprechen zwischen den Kanälen sollte möglichst gering sein. Zur Verbesserung der elektromagnetischen Verträglichkeit (EMV)[165]

[164] *Total Harmonic Distortions plus Noise* (THD+N).

[165] Die DIN-Norm VDE 0870 definiert die EMV als „die Fähigkeit einer elektrischen Einrichtung, in ihrer elektromagnetischen Umgebung zufriedenstellend zu funktionieren, ohne diese Umgebung, zu der auch andere Einrichtungen gehören, unzulässig zu beeinflussen."

der Baugruppen wurden alle Teilschaltkreise als vierlagige Leiterplatten mit einer flächenhaften Ausführung der Schaltungsmasse (GND) und der Versorgungsspannung (VCC) ausgeführt. Zudem wurden vorwiegend SMD-Bauteile[166] verwendet, da diese geringere parasitäre und antennenähnliche Effekte an den Durchgangsbohrungen hervorrufen und Abstrahleffekte durch Anschlussleitungen vermeiden (siehe z. B. Schwab und Kürner, 2007, Kap. 11). Durch die kleine Bauform der SMD-Bauteile kann eine höhere Packungsdichte mit kürzeren Leiterbahnen erreicht und das Übersprechen zwischen den signalführenden Leiterbahnen minimiert werden. Aufgrund der kapazitiven Belastung durch Steckverbindungen, Leitungslängen und Eingangskapazitäten können Verzerrungen der hochfrequenten Taktsignale entstehen. Um eine möglichst verzerrungsfreie Übertragung dieser Signale zu gewährleisten wurden zusätzliche Stützkondensatoren und impedanzangepasste Signalpuffer (*Clock Driver*) verwendet. Eine weitere Verminderung der Verzerrungen kann durch die Verwendung differentieller Taktsignale und eine symmetrische Signalübertragung erreicht werden. Dabei werden die hochfrequenten Taktsignale über parallele und komplementär zueinander gespeiste Signalleitungen geführt.

C.1.1 Mikrofonvorverstärker und Signalkonditionierung

Die IEPE/ICP-kompatiblen Mikrofonvorverstärker wurden für die Verwendung von vorpolarisierten Sennheiser KE 4-211-2 Elektret-Mikrofonkapseln optimiert. Diese Kapseln sehr kleiner Bauform ($\varnothing$ 4,75 mm) zeichnen sich durch einen besonders linearen Frequenzgang (20 Hz − 20 kHz ± 3 dB bzw. 40 Hz − 15 kHz ± 2,5 dB), eine hohe Empfindlichkeit (10 mV/Pa ± 2,5 dB bei 1 kHz im Freifeld)[167], einen geringen äquivalenten Rauschpegel (27 dB(A)[168] bzw. 38 dB nach ITU-R BS.468-4) und eine hohe Aussteuerungsgrenze (130 dB SPL)[169] bei geringem Klirrfaktor ($\leq$ 1%)[170] aus. Aufgrund ihrer geringen Körperschallempfindlichkeit sind Elek-

[166] *Surface-Mounted Devices* (SMD).

[167] Wird der eingebaute FET nicht als Impedanzwandler sondern im Verstärkerbetrieb verwendet, erhöht sich die Empfindlichkeit um 10 bis 14 dB, d. h. auf 30 mV/Pa bis 50 mV/Pa bei 1 kHz.

[168] Die A-Bewertung des Schalldruckpegels ist in der DIN Norm EN 61672-1 2003-10 (bzw. DIN-IEC 651) spezifiziert.

[169] Der Schalldruckpegel wird auch in der deutschsprachigen Literatur oft mit dem englischen Begriff SPL (*Sound Pressure Level*) abgekürzt.

[170] Bei einer Betriebsspannung von 5 V und einem Anschlusswiderstand von 5,6 kΩ.

tretmikrofone besonders für die Verwendung in Mehrkanal-Mikrofonanwendungen geeignet, da keine aufwendigen Aufhängungen zur mechanischen Entkopplung erforderlich sind. Der typische Frequenzgang einer KE 4-211-2 Mikrofonkapsel ist in Abb. C.2 abgebildet.

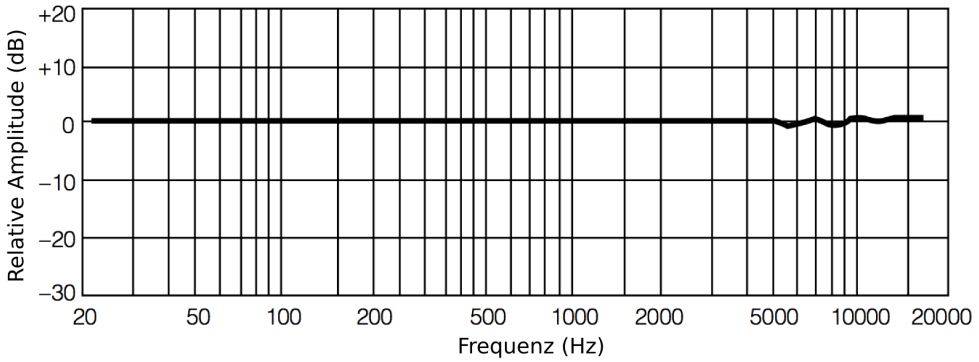
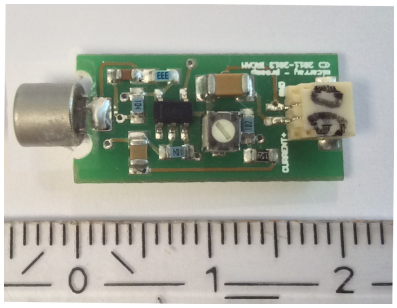


Abbildung C.2: Frequenzgang einer Sennheiser KE 4-211-2 Elektret-Mikrofonkapsel.

Bei Elektret-Mikrofonkapseln wird die Kondensatorvorspannung mit dauerhaft polarisierten Elektretfolien erzeugt. Zum Betrieb der Kapsel wird keine externe Speisespannung benötigt. Elektret-Mikrofonkapseln werden aufgrund ihrer hohen kapazitiven Impedanz stets mit einem nachgeschalteten Impedanzwandler mit sehr hochohmigem Eingang betrieben. Der Impedanzwandler besteht meist aus einem Feldeffekttransistor (FET), der zur Erhöhung der Empfindlichkeit auch im Verstärkerbetrieb verwendet werden kann (s. a. Lerch et al., 2009, Kap. 13.1.3). KE 4-211-2 Mikrofone benötigen für den FET eine Speisespannungen zwischen 0,9 V und 15 V, bei einer Stromaufnahme von ungefähr 250 μA. Die nachfolgende Vorverstärkerschaltung wandelt das störungsempfindliche Signal der Elektret-Mikrofonkapsel in ein Spannungssignal mit niedriger Impedanz am Ausgang um. Aufgrund der niedrigen Ausgangsimpedanz können lange Kabel verwendet werden, ohne dass die Signalqualität wesentlich beeinträchtigt wird. Es werden keine Spezialkabel benötigt. Der IEPE/ICP-kompatible Mikrofonvorverstärker ist in Abb. C.3 dargestellt.

Der Mikrofonvorverstärker wird von der nachgeschalteten Interface-Karte mit einem Konstantstrom von 8 mA[171] gespeist. Der Versorgungsstrom und das Mi-

[171] Beim IEPE/ICP-Protokoll soll der Konstantstrom zwischen 2 mA und 20 mA liegen.



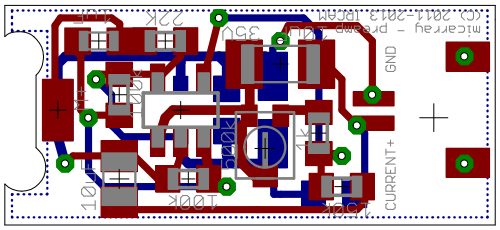


Abbildung C.3: IEPE/ICP-kompatibler Mikrofonvorverstärker mit KE 4-211-2 Elektret-Mikrofonkapsel. Über das Drehpotentiometer können die Mikrofonkapseln kalibriert bzw. auf unterschiedliche akustische Bedingungen angepasst werden.

krofonsignal werden gemeinsam über ein Koaxialkabel übertragen. Die Arbeitspunktspannung ist im IEPE/ICP-Protokoll mit $8 - 12\,\mathrm{V}$, die Aussteuergrenze mit $24 - 30\,\mathrm{V}$ spezifiziert (vgl. Abb. C.4). Das Mikrofonsignal wird der Arbeitspunktspannung als Wechselspannung überlagert und am Mikrofoneingang der Interface-Karte über einen Koppelkondensator zur weiteren Verarbeitung herausgefiltert.

Die im Rahmen dieser Arbeit entwickelte Mikrofonvorverstärkerschaltung ist für einen Schalldruck von $0{,}63\,\mathrm{mPa} - 63{,}25\,\mathrm{Pa}$ ($30 - 130\,\mathrm{dB}$ SPL) dimensioniert. Bei einer Empfindlichkeit von $50\,\mathrm{mV/Pa}$[172] ergibt sich der Scheitelwert der maximalen Mikrofonausgangsspannung zu $4{,}472\,\mathrm{V}$. Die Arbeitspunktspannung wurde mit $10\,\mathrm{V}$ festgelegt. Die Ausgangsspannung liegt somit innerhalb der IEPE/ICP Aussteuergrenzen (vgl. Abb. C.4). Mit einem regelbaren rauscharmen Signalverstärker ($0\,\mathrm{dB}$ bis $+12\,\mathrm{dB}$) können die Mikrofonkapseln kalibriert und für unterschiedliche akustische Bedingungen optimal auf die Signalübertragung angepasst werden.

C.1.2 Datenerfassung und Digitalisierung

Die Interface-Karten des Datenerfassungssystems besitzen vier IEPE/ICP-kompatible Mikrofoneingänge, die jeweils über eine regulierte und temperaturstabilisierte Konstantstromquelle[173] gespeist werden (vgl. Abb. C.5). Der

[172] Anm.: FET in Verstärkerbetrieb.

[173] Die Konstantstromquelle (ST Microelectronics LM234) wird zur Unterdrückung der Schwankungen der Versorgungsspannung über einen Spannungsregler (National Semiconductor LM317 mit einem *Ripple Rejection Ratio* von ca. $80\,\mathrm{dB}$) gespeist.

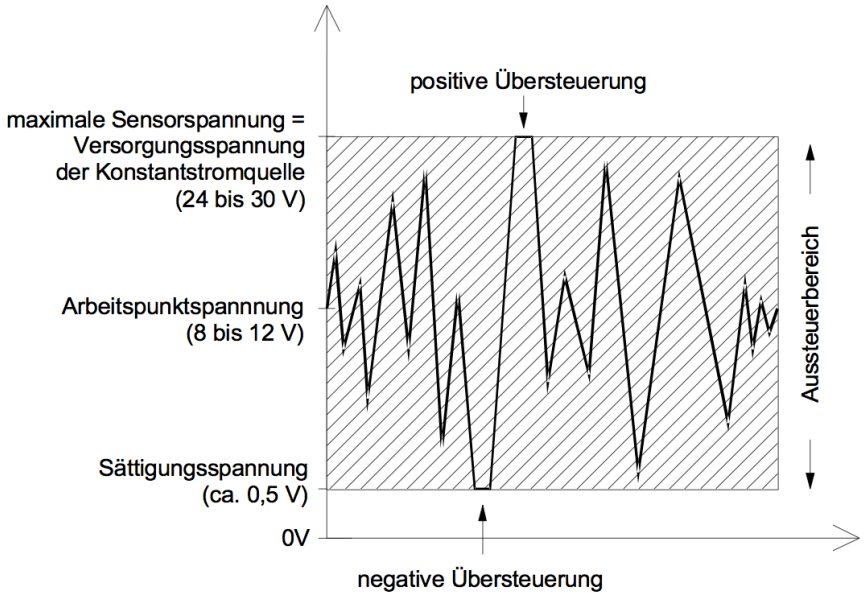


Abbildung C.4: Aussteuergrenzen IEPE/ICP-kompatibler Sensoren.

Eingangsverstärker dient der Skalierung und Symmetrierung der Mikrofonsignale, um diese optimal an die Weiterverarbeitung im nachgeschalteten A/D-Wandler anzupassen. Die A/D-Wandler sind über die Busleiterplatte mit dem FPGA verbunden. Das FPGA-Board stellt neben der stabilisierten Spannungsversorgung auch die Taktsignale für die Digitalisierung und Datenübertragung zur Verfügung. Die Busleiterplatte mit acht 4-Kanal Interface-Karten (s. Abb. C.6) zum Aufbau eines 32-Kanal Mikrofonarrays ist in Abb. C.7 abgebildet.

Beim I²S-Bus werden die Audiodaten und der Takt getrennt voneinander auf separaten Leitungen übertragen. Dadurch können Interface-Jitter und Sampling-Jitter verringert werden. Die I²S-Schnittstelle definiert standardmäßig drei Signale: BCLK (*Bit-Clock*), LRCLK (*Left/Right-Clock*) und DATA. Die BCLK[174] definiert die Taktperiode zur Übertragung der Datenbits; die LRCLK[175] kennzeichnet den Wortanfang und unterscheidet den linken vom rechten Kanal; die Datenleitung DATA dient der Übertragung des zweikanaligen Datenstroms. A/D-Wandler benötigen für die Delta-Sigma-Modulation meist ein zusätzliches Taktsignal, die

[174] Maximale BCLK: $\mathrm{BCLK}_{\max} = 2 \times 32 \times 192\,\mathrm{kHz} = 12{,}288\,\mathrm{MHz}$.

[175] Die LRCLK wird oft auch als *Word-Clock* (WCLK) oder *Frame Synchronization* (FSYNC) bezeichnet; $\mathrm{LRCLK}_{\max} = 192\,\mathrm{kHz}$.

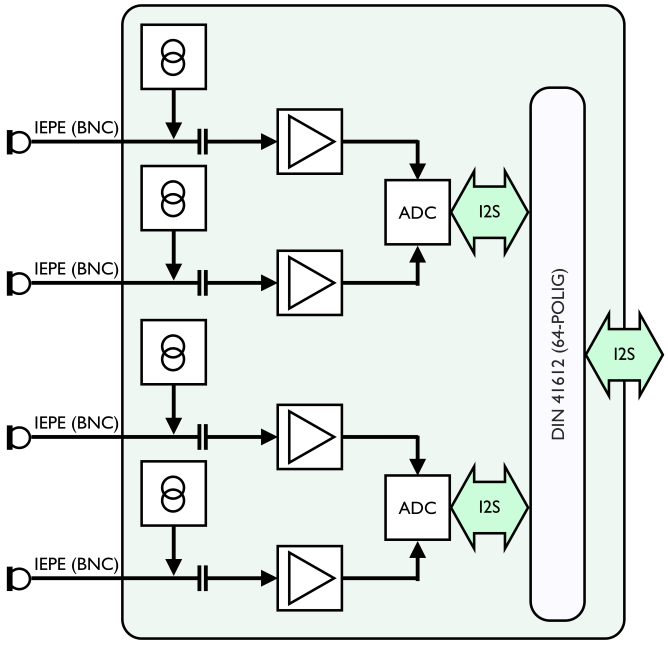


Abbildung C.5: IEPE/ICP-kompatible 4-Kanal Interface-Karte mit Konstantstromquelle, Signalkonditionierer und A/D-Wandlern.

System-Clock (SCLK)[176] mit der 128-fachen bis 512-fachen Rate der den PCM-Daten zugrundeliegenden Abtastrate.

Als Audio-A/D-Wandler kommt ein Texas Instruments PCM 4202 Baustein mit differentiellen analogen Eingangssignalen zum Einsatz. Dieser zeichnet sich durch einen besonders hohen Dynamikbereich von 118 dB und geringe harmonische Verzerrungen (THD+N ≤ -105 dB) aus. Der A/D-Wandler unterstützt 24-Bit PCM Ausgangsdaten mit Abtastraten bis zu 216 kHz, die über die integrierte I^2S/I^2C-Schnittstelle direkt an den nachgeschalteten FPGA angebunden werden. Die digitale Schnittstelle ist kompatibel mit den Logikfamilien für +3,3 V (d. h. kompatibel mit dem Xilinx-FPGA).

[176] Maximale SCLK: $\mathrm{SCLK}_{\max} = 128 \times 192\,\mathrm{kHz} = 24{,}576\,\mathrm{MHz}$.

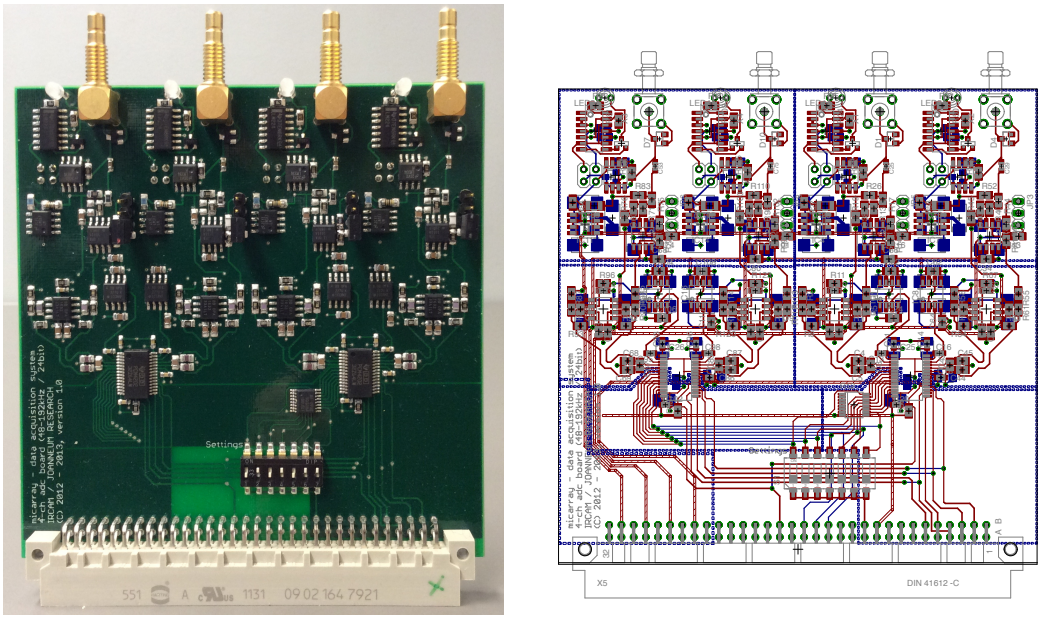


Abbildung C.6: 4-Kanal Interface-Karte mit konstantstromversorgten IEPE-Mikrofoneingängen, Signalverstärkern und A/D-Wandlern.

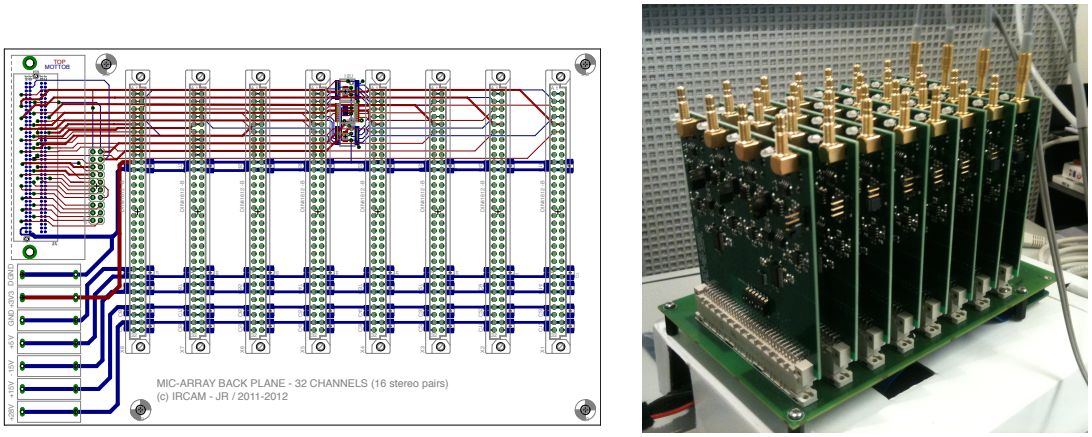


Abbildung C.7: Schema der Busleiterplatte (links) und acht 4-Kanal Interface-Karten zum Aufbau eines 32-Kanal Mikrofonarrays (rechts).

C.1.3 FPGA-Board und IP-basierte Datenübertragung

Das FPGA-Board[177] wandelt die parallel eintreffenden digitalen Audiosignale in einen seriellen Datenstrom (xFaceStream®). Die UDP-Datensegmente werden über

[177] Xilinx Spartan-6 FPGA SP601 EVM.

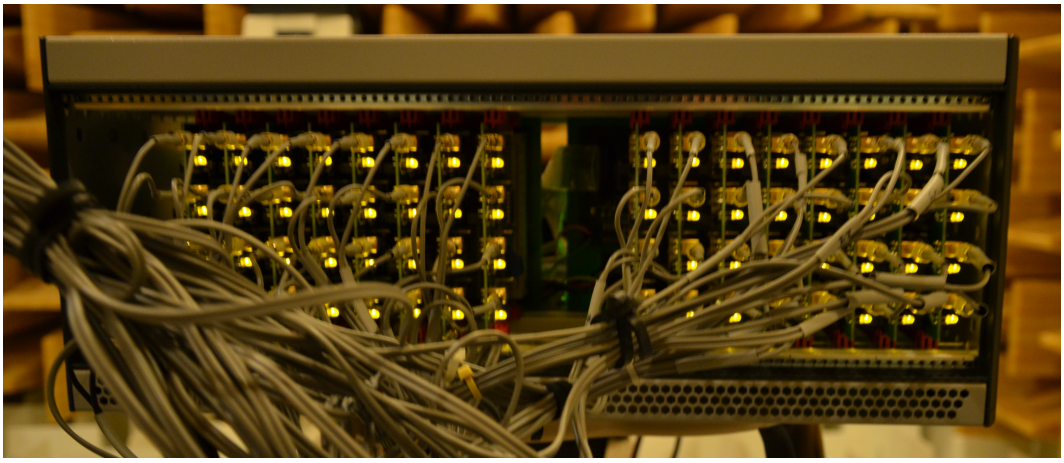


Abbildung C.8: 4-Kanal Interface-Karten im Betrieb eines 64-Kanal Mikrofonarrays.

eine Gigabit-Ethernet Verbindung an den Client-Computer übertragen. Zudem stellt das FPGA-Board die zur A/D-Wandlung und synchronen Datenübertragung benötigten Taktsignale zur Verfügung. Abb. C.9 illustriert den Aufbau des FPGA-Boards mit den einzelnen Signalverarbeitungsstufen.

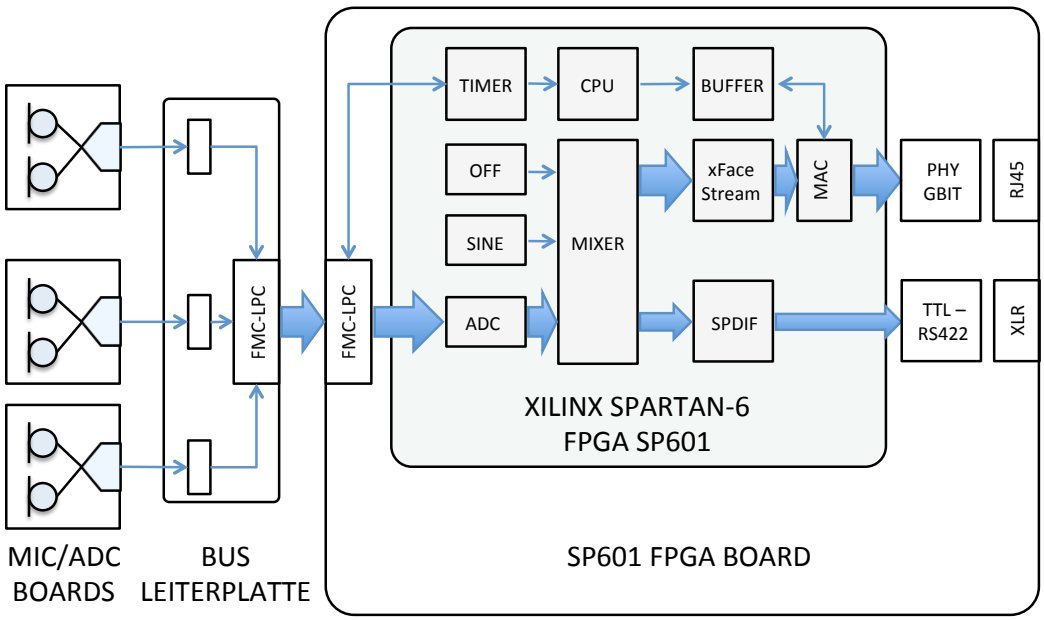


Abbildung C.9: Architektur und Signalverarbeitungsstruktur des FPGA-Boards.

Das xFaceStream-Protokoll (xFSP) ermöglicht eine synchrone digitale Übertragung unkomprimierter Mehrkanal-Audiodaten, mit hohem Datendurchsatz, geringer Netzwerkbelastung und minimaler Latenz. xFSP nutzt zum Versenden der Daten in einem Netzwerk die OSI-Transportschicht. Das OSI-Modell[178] definiert die Kommunikationsarchitektur von IP-basierten Netzwerkprotokollen. Es besteht aus sieben Schichten, wobei jeder Schicht eine klare Schnittstelle und Funktionalität zugeordnet ist (vgl. Tab. C.1). Die Transportschicht (Schicht 4) ist für den gesamten Ende-zu-Ende-Transfer von Anwendungsdaten zuständig und bereitet die Daten auf den Transport durch das Netzwerk vor. Die beiden wichtigsten Protokolle der OSI-Transportschicht sind das TCP (*Transmission Control Protocol*)[179] und das UDP (*User Datagram Protocol*). Diese Protokolle stellen den übergeordneten Schichten zuverlässige Dienste zum Transport der Daten durch ein Netzwerk zur Verfügung.

UDP baut direkt auf dem darunter liegenden IP-Protokoll auf und bietet Grundfunktionen für ein effizientes Zustellen von Datagrammen[180]. Es zeichnet sich besonders durch seinen geringen Protokoll-Overhead und kurze Latenzen aus. UDP ist ein unzuverlässiges[181] und verbindungsloses (d. h. transaktionsorientiertes) Protokoll ohne Flusskontrolle und Fehlerkorrektur. Es garantiert weder den Schutz vor Paketverlust noch die reihenfolgerichtige Zustellung eines Datagramms. Im Gegensatz zu TCP-Verbindungen findet keine wiederholte Übertragung verloren gegangener Datenpakete statt. Der für die verlustfreie Datenübertragung benötigte Bestätigungsmechanismus (Drei-Wege-Handschlag) entfällt. Die Vorteile einer ungesicherten Übertragung sind geringe Schwankungen der Übertragungsverzögerung und eine kurze Verzögerungszeit. UDP wird überall dort eingesetzt, wo eine geringe Verzögerungszeit erforderlich ist, wie zum Beispiel bei der Echtzeitübertragung von Audiodaten.

Das xFaceStream-Protokoll baut auf dem UDP-Protokoll auf, bietet jedoch zusätzliche Kontrollmechanismen. Zur Integritätsprüfung wird für jedes Datenpa-

[178] Das *Open Systems Interconnection* (OSI) Modell ist im ISO/IEC Standard 7498-1:1994 spezifiziert.

[179] Das *Transmission Control Protocol* (TCP) ist im IETF Standard RFC 793 spezifiziert.

[180] Ein Datagramm ist eine in sich geschlossene Dateneinheit mit Nutz- und Steuerdaten.

[181] „Unzuverlässig“ bedeutet in diesem Kontext, dass das Protokoll keinerlei Mechanismen zur Verfügung stellt, um die sichere Zustellung eines Datagramms zu gewährleisten.

Tabelle C.1: Schichtenarchitektur des OSI-Referenzmodells für Netzwerkprotokolle nach ISO/IEC 7498-1:1994.

OSI-Schicht		Kategorie	TCP/IP-Schicht	Protokoll	Einheiten
7	Anwendungen (*Application Layer*)	Anwendungs-orientiert	Anwendung	HTTP(S), UDS, FTP, SMTP, POP, Telnet, LDAP, OPC UA, NCP, SOCKS	Daten
6	Darstellung (*Presentation Layer*)				
5	Sitzung (*Session Layer*)				
4	Transport (*Transport Layer*)	Transport-orientiert	Transport	TCP, UDP, SCTP, SPX	Segmente (TCP), Datagramme (UDP)
3	Vermittlung (*Network Layer*)		Vermittlung	ICMP, IGMP, IP, IPsec, IPX	Pakete
2	Sicherung (*Data Link Layer*)		Netzzugriff	Ethernet, Token Bus, Token Ring, FDDI, ARCNET, IPoAC	Rahmen (*Frames*)
1	Bitübertragung (*Physical Layer*)				Bits, Pakete, Symbole

ket eine Prüfsumme erstellt und gesendet. Die Datensynchronisation erfolgt über einen hochpräzisen 64 Bit Zeitstempel. Mit diesen Protokollmechanismen können Anwendungsprozesse sowohl verloren gegangene Datenpakete als auch Reihenfolgevertauschungen sehr einfach detektieren und geeignet reagieren. Bei der Audioübertragung führt ein etwaiger Paketverlust zu hörbaren Klicks. Diese können zum Beispiel mit Hilfe einer mehrkanaligen Ausfallsverschleierung vermindert werden.

xFSP Datagramme werden mit einer festen Größe von 1076 Byte über das IP-Netzwerk geschickt. Jedes Datensegment enthält neben der Header-Information Nutzdaten mit einer festen Größe von 1024 Byte. Das xFSP kann somit maximal

Tabelle C.2: xFaceStream-Protokoll zur Übertragung von Mehrkanal-Audiodaten.

Byte	IP-Header, Nutzdaten	Datenfelder
0 - 13 (14 Byte)	ETH Header	6 Byte MAC-Adresse Empfänger 6 Byte MAC-Adresse Sender 2 Byte Ethernet [0x0800: IP]
14 - 33 (20 Byte)	IP Header	1 Byte IP-Version + IP-Header Länge (IHL) [0x45] 1 Byte Priorisierung (*Type of Service*, TOS) [0x00] 2 Byte Gesamtlänge [1076] 2 Byte Identifikation [0x0000] 1 Byte Flags 1 Byte Fragment Offset [0x00] 1 Byte Gültigkeitsdauer (*Time to Live*, TTL) [0x40] 1 Byte Protokoll [0x11: UDP] 2 Byte Prüfsumme 4 Byte Quell-IP-Adresse 4 Byte Ziel-IP-Adresse
34 - 41 (8 Byte)	UDP Header	2 Byte Quell-Port [0xface] 2 Byte Ziel-Port [0xface] 2 Byte Datenlänge [1042] 2 Byte Prüfsumme
42 - 51 (10 Byte)	xFSP Header	1 Byte Port [0x00] 1 Byte Format (z. B. [0x5d]: 32 Kanäle, 32 Bit Signed-PCM) 4 Byte Zeitstempel (time_h): Sekunden, Little Endian 4 Byte Zeitstempel (time_l): Samples, Little Endian
52 - 1075 (1024 Byte)	Audiodaten	256×4 Byte Signed-PCM, Little Endian, Left Aligned, 24 Bit Wortbreite

256 Samples mit einer Wortbreite von 32 Bit übertragen. Um eine direkte Verarbeitung der Daten in x86-basierten Systemen zu vereinfachen, werden alle Bytes in *Little Endian* Bit-Reihenfolge, d. h. mit dem niedrigstwertigen Bit (*Least Significant Bit*, LSB) voran, übertragen. Die PCM-kodierten Audiodaten haben eine Wortlänge von 24 Bit und sind im 32 Bit Datenwort links ausgerichtet. Das niedrigstwertige Byte (*Least Significant Byte*) enthält somit keine Audiodaten und ist für zukünftige Erweiterungen, wie z. B. den Einsatz von 32 Bit A/D-Wandlern, reserviert. Die xFSP Protokollstruktur ist in Tab. C.2 zusammengefasst.

Die zur Übertragung der Mehrkanal-Audiodaten erforderliche maximale Datenübertragungsrate in Abhängigkeit von der Kanalanzahl und der Abtastrate kann Tab. C.3 entnommen werden. Für die bidirektionale Übertragung von 64 PCM-kodierten Audiokanälen mit 24 Bit Wortbreite und 192 kHz Abtastrate wird eine Gigabit-Ethernet Verbindung benötigt.

Tabelle C.3: Datenübertragungsrate der digitalen Mehrkanal-Audioübertragung in Abhängigkeit von der Anzahl der Audiokanäle und der Abtastrate.

Abtastrate [kHz]	Anzahl der Audiokanäle	Datenübertragungsrate [MBit/s]
48	32	51,6
96 | 48	32 | 64	103,3
192 | 96	32 | 64	206,6
192	64	413,2

C.2 Anwendungsbeispiele

Typische praktische Anwendungen des modularen Mikrofonarrays umfassen die Nahfeld-Holografie mit einem planaren Mikrofonarray (siehe Abb. C.10) sowie die Messung von MIMO-Raumimpulsantworten mit einer Mikrofonarray-Lautsprecherarray-Anordnung (siehe Abb. C.11). Letztere wird ausführlich in Noisternig et al. (2016) und Morgenstern et al. (2016) beschrieben.

Abbildung C.10: Nahfeld-Holografie mit einem planaren Mikrofonarray.

Abbildung C.11: Mikrofonarray-Lautsprecherarray-Anordnung zur Messung von MIMO-Raumimpulsantworten im Großen Festspielhaus der Salzburger Festspiele (Fotos: © 2016, Johannes Klein, RWTH Aachen).

D

Erfindungsmeldung und Patentschrift

Die folgende Patentschrift umfasst das in dieser Arbeit vorgestellte Verfahren zur einkanaligen Störgeräuschreduktion mit möglichst kurzer Systemlatenz (siehe Kap. 4). Dabei finden die aus den funktionalen Modellen des peripheren Gehörs (siehe Kap. 2.1) abgeleiteten auditiven Gammaton-Filter (siehe Kap. 2.3) als Analyse-Synthese-Filterbank Anwendung (siehe Kapitel 3). Aus der Betrachtung der Phase an den Übergangsfrequenzen benachbarter Teilbandfilter kann ein sehr einfaches und recheneffizientes Verfahren zur Implementierung einer Gammaton-Filterbank hergeleitet werden, welches eine nahezu perfekte Rekonstruktion des Signals durch einfache Summation der Teilbandsignale erlaubt.

An der Erfindung des in der Patentschrift dargestellten Verfahrens waren mehrere Personen beteiligt. Der Anteil der Miterfinder kann der beigelegten Erfindungsmeldung entnommen werden.

UNIVERSITÄT FÜR MUSIK UND DARSTELLENDE KUNST GRAZ · AUSTRIA

ERFINDUNGSMELDUNG

NUR VERSCHLOSSEN VERSENDEN!

An das
Rektorat der KUG
Leonhardstraße 15
8010 Graz

Rektorat der KUG
Eing. Nr. 1870
Eingel. Am: 26. Sep. 2008
Erg. an: Wirtsch. Veranst. ZID Stud.
FiBu AIB Pers Org. Re CoKo
Öffl. Ablage

Titel der Erfindung:	Method and device for low-latency auditory model-based single-channel speech enhancement
Erfinder:	Markus Noisternig, Franz Zotter, Robert Höldrich
Organisationseinheit:	Institut 17 / IEM

Nicht vom Erfinder auszufüllen!

Laufende Nummer: 12

Eingang Rektorat am: 26.9.2008	
Weiterleitung an Verwaltungsabteilungen am: 1.10.2008	
Ende der Aufgriffsfrist: 26.12.2008	

Entscheidung über Inanspruchnahme oder Freigabe:

Die Erfindung wird seitens der Universität verwertet X

Die Erfindung wird seitens der Universität freigegeben O

1.10.2008 _____ Datum, für das Rektorat

_____ Datum, für das Rektorat

→ Kopien der Erfindungsmeldung gehen an die Abteilungen Org./Recht und CoKo.

Die Unterzeichneten melden hiermit der Universität

die im Folgenden beschriebene Erfindung.

Nur eine vollständige und umfassende Darstellung ermöglicht die Prüfung durch die Universität und die Einreichung der Patentanmeldung. Erweiterungen sind innerhalb eines laufenden Patentierungsverfahrens kaum möglich. Ebenso erstreckt sich die allfällige Freigabe einer Erfindung durch die Universität lediglich auf den im Rahmen dieses Meldeformulars spezifizierten Umfang.

Die Unterzeichneten erklären sich damit einverstanden, dass die zur Verfügung gestellten Daten in EDV-Anlagen verarbeitet und gespeichert werden. Sie erteilen darüber hinaus ihr ausdrückliches Einverständnis zum Austausch der für die Bearbeitung der gegenständlichen Erfindungsmeldung erforderlichen Daten zwischen Universität und Dritten (bspw. aws/Tecma), sofern diese im Rahmen der Verwertung durch das Institut oder die Universität herangezogen werden.

ERFINDUNGSMELDUNG

1. Erfindung

a) Beschreibender Kurz-Titel Apparatus for low-latency auditory model-based single channel speech enhancement

b) Beschreibung der Erfindung – bitte kurz und prägnant den erfinderischen Kern beschreiben
(Umfangreichere Beschreibung bitte im Anhang, evtl. in Form von Publikationsmanuskripten oder dergl.)

The present invention relates a method for enhancing wide-band speech audio signals in the presence of background noise and, more particularly to a noise suppression system, a noise suppressing method and a noise suppression program. More specifically, the present invention relates low-latency single-channel noise reduction using sub-band processing based on masking properties of human auditory system.

c) Neuheit - Unterschiede gegenüber dem Stand der Technik

Please see invention reocord attached.

d) Wesentliche Vorteile - verglichen mit herkömmlichen Technologien und dem Stand der Technik

Wide-band speech enhancement with lowest-latency and auditory masking for an improved perceptional quality of signal.

e) Nachteile

f) Reifegrad / vorhandene Daten: ☐ Planungsstadium ☒ Proof of principle (Laborversuch) ☐ Prototyp

Details, evtl. Zeitplan Implementierung in MATLAB, Pure Data, Teilalgorithmen in ANSI C

g) Wurden bereits Schutzrechte angemeldet (Patente, Gebrauchsmuster etc.)? ☒ ja.....☐ nein

Wenn ja, welche und wann? AKG Acoustics GmbH Wien (Kooperationspartner)

h) Gibt es einen Recherche- oder Prüfbericht des Patentamts?	☐ ja.....☒ nein
	(wenn vorhanden bitte beilegen!)

2. Entstehung der Erfindung

Steht die Erfindung im Zusammenhang mit Ihrer fachlichen Tätigkeit an der Universität?

☒ ja ☐ nein

Im Rahmen welchen Projektes ist die Erfindung entstanden?

☐ Diplomarbeit	☒ Forschungsauftrag	
☒ Dissertation	☐ Gefördertes Forschungsprojekt	Projektleiter:
☐ Eigenforschung	☐ Sonstiges, und zwar:	

Die Forschung wurde finanziert von: (Mehrfachnennungen möglich)

☐ Universität, Institut	☒ Firma, und zwar:	AKG Acoustics GmbH, Wien
☐ Ministerium	☐ FFF, FWF	
☐ EU	☐ Sonstiges, und zwar:	

Wurden die Arbeiten mit Universitätsinfrastruktur durchgeführt? ☒ ja ☐ nein

3. Veröffentlichungen

Gab es Präsentationen in der Öffentlichkeit? ☐ ja ☒ nein

(zB: Vorträge, Poster, Messen, Internet, Publikationen, Dissertationen, etc.)

Wenn ja, wo und wann?

Ist eine Veröffentlichung geplant? Termin der Veröffentlichung / Einreichung JA, noch kein Einreichdatum

Art (Paper, Dissertation, Abstract, elektronische Datenbank, etc.) Dissertation, Artikel

4. Bezug oder Weitergabe von Material/Daten

Wurden im Zuge der Erfindung Materialien oder Daten <u>von Dritten bezogen</u>?

☒ ja ☐ nein

Wenn ja, wurden MTAs (Material Transfer Agreements) oder NDAs (Geheimhaltungsabkommen) abgeschlossen? ☒ ja ☐ nein

Wurden im Zuge der Erfindung Materialien oder Daten <u>an Dritte vergeben</u>?

☒ ja ☐ nein

Wenn ja, wurden MTAs (Material Transfer Agreements) oder NDAs (Geheimhaltungsabkommen) abgeschlossen? ☒ ja ☐ nein

5. Verwertung

Gab es bereits Verwertungsversuche?	☒ ja	☐ nein
Wenn ja, welche?	Lizenzierung durch AKG Acoustics GmbH	
Gibt es bereits Interessenten?		

6. Erfinderdaten

	Erfinder 1	Erfinder 2
Titel, Vor- und Zuname	Dipl. Ing. Markus Noisternig	Dipl. Ing. Franz Zotter
Universität/ Firma	Universität für Musik und darstellende Kunst, Graz	Universität für Musik und darstellende Kunst, Graz
Organisations-einheit	Institut für Elektronische Musik und Akustik	Institut für Elektronische Musik und Akustik
Adresse	Inffeldgasse 10/3, 8010 Graz	Inffeldgasse 10/3, 8010 Graz
Telefon und eMail	noisternig@iem.at	+43 316 389 5219, zotter@iem.at
Wohnadresse	32 boulevard Saint-Marcel, 75005 Paris, France	Sandgasse 25C, 8010 Graz
Nationalität	Österreich	Österreich
Stellung zur Universität *)	Universitätsbediensteter, Vertragslehrer neu	Universitätsbediensteter
Anteil an der Erfindung in %)**	35	35
Anzahl der Erfinder	4	Verwertungsanteil der Org.einheit (bis. max 45%):

*) Universitätsbedienstete(r), Bundesbedienstete(r), Gemeindebedienstete(r), StudentIn, DiplomandIn, DissertantIn, Fellow, GastprofessorIn, etc.

**) Ausschlaggebend ist der eigenständige konzeptionelle Beitrag zur Erfindung

Weitere Erfinder sind in der Anlage „Weitere Erfinder" anzuführen!

Ich/wir bestätige/n die Erfindung vollständig und umfassend beschrieben und alle beteiligten Erfinder genannt zu haben. Mir/uns ist bekannt, dass die Erfindung entsprechend Patentgesetz §13 bis zur Entscheidung über einen allfälligen Aufgriff der Geheimhaltung unterliegt und an Außenstehende nur unter der Verpflichtung zur Geheimhaltung weitergegeben werden darf.

Datum, Unterschrift	Datum, Unterschrift

Die Erfindungsmeldung wurde dem Institutsvorstand zur Kenntnis gebracht:

Datum, Unterschrift des Institutsvorstandes / der Institutsvorständin

Folgende Unterlagen sind bitte – so vorhanden – der Erfindungsmeldung beizulegen.

Bitte kreuzen Sie die entsprechenden Boxen bezüglich beigelegter Dokumente an.

Anlagen zur Erfindung:

Bitte legen Sie der Erfindungsmeldung folgende Anlagen bei.

- ☐ Eine möglichst umfassende Beschreibung der Erfindung:
  - ➢ Welche Aufgabe war zu lösen?
  - ➢ Welche bisherigen Lösungsversuche gab es?
  - ➢ Beschreibung der Lösung / Erfindung (wie wird das Problem/die Aufgabe gelöst?)
  - ➢ Ausführungsbeispiele
  - ➢ Zeichnungen
- ☐ Allfällige Manuskripte respektive Entwürfe
- ☐ Manuskript geplanter Veröffentlichungen
- ☐ Schlüsselpublikationen
- ☐ Recherchebericht des Patentamtes

Anlagen zur Rechtslage:

Bitte um Beilage sämtlicher Verträge resp. Vertragsentwürfe, die im Zusammenhang mit dem gegenständlichen Projekt abgeschlossen wurden bzw. sind die in Verhandlung sind

- ☐ Drittmittelverträge
- ☐ Forschungaufträge (zB FWF, Bundesministerium, etc.)
- ☐ MTAs (Material Transfer Agreements)
- ☐ NDAs (Geheimhaltungsverträge)

Anlagen zu weiteren Erfindern:

Sind an der Erfindung mehr als die 2 im Erfindungsmeldebogen genannten Personen beteiligt, so sind diese im Blatt „Weitere Erfinder" anzuführen:

- ☐ Anlage WEITERE ERFINDER

5. Verwertung

Gab es bereits Verwertungsversuche?	☒ ja	☐ nein
Wenn ja, welche?	Lizenzierung durch AKG Acoustics GmbH	
Gibt es bereits Interessenten?		

6. Erfinderdaten

	Erfinder 1	Erfinder 2
Titel, Vor- und Zuname	Dipl. Ing. Markus Noisternig	Dipl. Ing. Franz Zotter
Universität/ Firma	Universität für Musik und darstellende Kunst, Graz	Universität für Musik und darstellende Kunst, Graz
Organisations-einheit	Institut für Elektronische Musik und Akustik	Institut für Elektronische Musik und Akustik
Adresse	Inffeldgasse 10/3, 8010 Graz	Inffeldgasse 10/3, 8010 Graz
Telefon und eMail	noisternig@iem.at	+43 316 389 5219, zotter@iem.at
Wohnadresse	32 boulevard Saint-Marcel, 75005 Paris, France	Sandgasse 25C, 8010 Graz
Nationalität	Österreich	Österreich
Stellung zur Universität *)	Universitätsbediensteter, Vertragslehrer neu	Universitätsbediensteter
Anteil an der Erfindung in %)**	35	35
Anzahl der Erfinder	4	Verwertungsanteil der Org.einheit (bis. max 45%):

*) Universitätsbedienstete(r), Bundesbedienstete(r), Gemeindebedienstete(r), Studentin, Diplomandin, Dissertantin, Fellow, Gastprofessorin, etc.

**) Ausschlaggebend ist der eigenständige konzeptionelle Beitrag zur Erfindung

Weitere Erfinder sind in der Anlage „Weitere Erfinder" anzuführen!

Ich/wir bestätige/n die Erfindung vollständig und umfassend beschrieben und alle beteiligten Erfinder genannt zu haben. Mir/uns ist bekannt, dass die Erfindung entsprechend Patentgesetz §13 bis zur Entscheidung über einen allfälligen Aufgriff der Geheimhaltung unterliegt und an Außenstehende nur unter der Verpflichtung zur Geheimhaltung weitergegeben werden darf.

Datum, Unterschrift	Datum, Unterschrift

Die Erfindungsmeldung wurde dem Institutsvorstand zur Kenntnis gebracht:

Datum, Unterschrift des Institutsvorstandes / der Institutsvorständin

Erfinderdaten

	Erfinder 3	Erfinder 4
Titel, Vor- und Zuname	o. Univ. Prof. Dr. Dipl. Ing. Mag. Robert Höldrich	Dr. Martin Opitz
Universität/ Firma	Universität für Musik und darstellende Kunst, Graz	AKG Acoustics GmbH
Institut	Institut für Elektronische Musik und Akustik	Research and Development (Professional Audio)
Adresse	Inffeldgasse 10/3, 8010 Graz	Lemböckgasse 21-25, 1230 Wien
Telefon und eMail	+43 316 389 3334, hoeldrich@iem.at	+43 1 86654 1387, OpitzM@akg.com
Wohnadresse	Kohlbachgasse 3/12, 8010 Graz	
Nationalität	Österreich	Österreich
Stellung zur Universität *)	Vizerektor	Vertragspartner
Anteil an der Erfindung in %)**	20	10
Anzahl der Erfinder		Verwertungsanteil der Org.einheit (bis. max 45%):

*) Universitätsbedienstete(r), Bundesbedienstete(r), Gemeindebedienstete(r), StudentIn, DiplomandIn, DissertantIn, Fellow, GastprofessorIn, etc.
**) Ausschlaggebend ist der eigenständige konzeptionelle Beitrag zur Erfindung

	Erfinder 5	Erfinder 6
Titel, Vor- und Zuname		
Universität/ Firma		
Institut		
Adresse		
Telefon und eMail		
Wohnadresse		
Nationalität		
Stellung zur Universität *)		
Anteil an der Erfindung in %)**		
Anzahl der Erfinder		Verwertungsanteil der Org.einheit (bis. max 45%):

*) Universitätsbedienstete(r), Bundesbedienstete(r), Gemeindebedienstete(r), StudentIn, DiplomandIn, DissertantIn, Fellow, GastprofessorIn, etc.
*) Ausschlaggebend ist der eigenständige konzeptionelle Beitrag zur Erfindung

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 April 2009 (09.04.2009)

PCT

(10) International Publication Number
WO 2009/043066 A1

(51) International Patent Classification:
G10L 21/02 (2006.01)

(21) International Application Number:
PCT/AT2007/000466

(22) International Filing Date: 2 October 2007 (02.10.2007)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **AKG ACOUSTICS GMBH** [AT/AT]; Lemböckgasse 21-25, A-1230 Wien (AT).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **OPITZ, Martin** [AT/AT]; Hackenberggasse 29/12/3, A-1190 Wien (AT). **HÖLDRICH, Robert** [AT/AT]; Rudolfstrasse 5, A-8010 Graz (AT). **ZOTTER, Franz** [AT/AT]; Inffeldgasse 10/3, A-8010 Graz (AT). **NOISTERNIG, Markus** [AT/FR]; 32, boulevard Saint-Michel, F-75005 Paris (FR).

(74) Agent: **BARGER, PISO & PARTNER**; Mahlerstrasse 9, A-1010 Wien (AT).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(54) Title: METHOD AND DEVICE FOR LOW-LATENCY AUDITORY MODEL-BASED SINGLE-CHANNEL SPEECH ENHANCEMENT

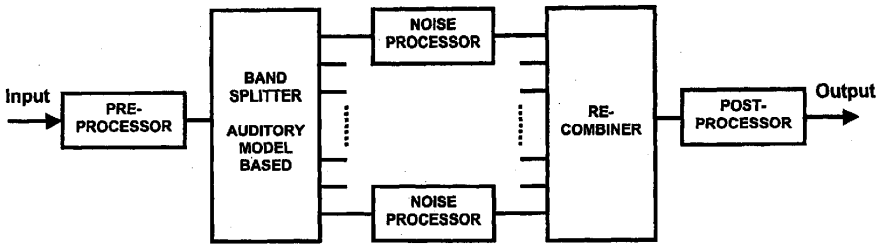


Fig. 1

(57) Abstract: The present invention relates to a method for enhancing wide-band speech audio signals in the presence of background noise and, more particularly to a noise suppression system, a noise suppression method and a noise suppression program. More specifically, the present invention relates to low-latency single-channel noise reduction using sub-band processing based on masking properties of the human auditory system.

WO 2009/043066 A1

METHOD AND DEVICE FOR LOW-LATENCY AUDITORY MODEL-BASED SINGLE-CHANNEL SPEECH ENHANCEMENT

FIELD OF THE INVENTION

The present invention relates to a method for enhancing wide-band speech audio signals in the presence of background noise and, more particularly to a noise suppression system, a noise suppression method and a noise suppression program. More specifically, the present invention relates to low-latency single-channel noise reduction using sub-band processing based on masking properties of the human auditory system.

BACKGROUND OF THE INVENTION

Additive background noise in speech communication systems degrades the subjective quality and intelligibility of the perceived voice. Therefore, speech processing systems require noise reduction methods, i.e. methods aiming at processing a noisy signal with the purpose of eliminating or attenuating the level of noise and improving the signal-to-noise-ratio (SNR) without affecting the speech and its characteristics. In general, noise reduction is also referred to as noise suppression or speech enhancement.

For example, mobile phones are often used in environments with high level of background noise such as public spaces. The use of mobile phones, voice-controlled devices and communication systems in cars has created a great demand for hands-free in-car installations, with the objective to increase safety and convenience; in many countries and regions law prohibits e.g. hand-held telephony in cars. Noise reduction becomes important for these applications, as they often needed to operate in adverse acoustic environments, in particular at low signal-to-noise ratios (SNR) and highly time-varying noise signal characteristics (e.g. rolling noise of cars).

In room teleconferencing applications, such as video-conferencing or speech recognition and querying systems, ambient noise usually arises from fans of computers, printers, or facsimile machines, which can be considered as (long-term) stationary. Conversational noise, emerging from (telephone) talks of colleagues sharing the office, as often referred to as babble noise, contains harmonic components and is therefore much harder to attenuate by a noise reduction

unit.

However, applications within hearing aids and in-car speech communication systems require noise suppression methods, which can be performed in real-time.

Despite, the fast development of the underlying hardware in terms of computing power and storage capacity supports the progress of software implementations.

One of the most widely used methods for noise reduction in real-world applications is referred to in the art as spectral subtraction (see S. F. Boll, "Suppression of Acoustic Noise in Speech using Spectral Subtraction," IEEE Trans. Acoust. Speech and Sig. Proc., vol. ASSP-27, pp. 113-120, Apr. 1979). Generally, spectral subtraction attempts to estimate the short time spectral amplitude (STSA) of clean speech from that of the noisy speech, i.e. the desired speech contaminated by noise, by subtracting an estimate noise signal. The estimated speech magnitude is combined with the phase of the noisy speech, based on the assumption that the human ear is insensitive against phase distortions (see C. L. Wang et al., "The unimportance of phase in speech enhancement," IEEE Trans. Acoust. Speech and Sig. Proc., vol. ASSP-30, pp. 679-681, Aug. 1982). In practice, spectral subtraction is implemented by multiplying the input signal spectrum with a gain function in order to suppress frequency components with low SNR. This SNR-based gain function is formed from estimates of the noise spectrum and noisy speech spectrum assuming wide-sense stationary, zero-mean random signals and the speech and the noise signals to be uncorrelated. These conventional spectral subtraction methods provide significant noise reduction with the main disadvantage of a degradation of the signal quality, acoustically perceptive as "musical tones" or "musical noise". The musical tones emerge from spectrum estimation errors. In the recent years many enhancements to the basic spectral subtraction approach have been developed.

A method to reduce musical tones which is often applied is to subtract an overestimate of the noise spectrum to reduce the fluctuations in the DFT coefficients and prevent the spectral components from going below a spectral floor (see M. Berouti et al., "Enhancement of speech corrupted by acoustic noise," in Proc. IEEE Int. Conf. on Acoust., Speech and Sig. Proc. (ICASSP'79), vol. 4, pp. 208-211, Washington D.C., Apr. 1979). This approach successfully reduces musical tones during low SNR conditions and noise only periods. The main disadvantage is the distortion of the speech signal during voice activity . In practice a tradeoff between speech quality level and residual noise floor level has to be found. Further methods cope with this problem by introducing optimal and adaptive oversubtraction factors for low SNR conditions

and propose underestimation of the noise spectrum at high SNR conditions (see W. M. Kushner et al., "The effects of subtractive-type speech enhancement / noise reduction algorithms on parameter estimation for improved recognition and coding in high noise environments," in Proc. IEEE Int. Conf. Acoustics, Speech and Sig. Proc. (ICASSP'89), vol. 1, pp. 211-214, 1989).

Applying a soft-decision based modification of the spectral gain function (see R. McAulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," in IEEE Trans. Acoust., Speech and Sig. Proc., vol. 28, no. 2, pp. 137-145, 1980) has been shown to improve the noise suppression properties of the enhancement system in terms of musical tone suppression. These soft-decision approaches mainly depend on the a priori probability of speech absence in each spectral component of the noisy speech.

The minimum mean-square error short-time spectral amplitude estimator (MMSE-STSA, see Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time amplitude estimator," IEEE Trans. Acoust. Speech and Sig. Proc., vol. 32, no. 6, pp.1109-1121, 1984) and the minimum mean-square error log spectral amplitude estimator (MMSE-LSA, Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log spectral amplitude estimator," IEEE Trans. Acoust. Speech and Sig. Proc., vol. 33, no. 2, pp.443-445, 1985) minimize the mean squared error of the estimated short-time spectral or log spectral amplitude respectively. It was found that the nonlinear smoothing procedure of the MMSE-STSA/LSA methods (the so-called decision-directed approach) obtains a more consistent estimate of the SNR, resulting in good noise suppression without unpleasant musical tones (see O. Capp, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," IEEE Trans. Speech and Audio Proc., vol. 2, no. 2, pp. 345-349, 1994). Both, Capp and Malah (see E. Malah et al., "Tracking speech-presence uncertainty to improve speech enhancement in non-stationary noise environments," in Proc. IEEE Int. Conf. Acoust., Speech and Sig. Proc. (ICASSP'99), vol. 2, pp. 789-792, 1999) propose a limitation of the a priori SNR estimate to overcome the problem of perceptible low-level musical noise during speech pauses. The so-called a priori SNR represents the information on the unknown spectrum magnitude gathered from previous frames and is evaluated in the decision-directed approach (DDA). As the smoothing performed by the DDA may have irregularities, low-level musical noise may occur. A simple solution to this problem consists in constraining the a priori SNR by a lower bound.

In single-channel spectral subtraction the noise power spectrum is usually estimated during

speech pauses requiring voice activity detection (VAD) methods (see R. McAulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," in IEEE Trans. Acoust., Speech and Sig. Proc., vol. 28, no. 2, pp. 137-145, 1980; and W. J. Hess, "A pitch-synchronous digital feature extraction system for phonemic recognition of speech", in IEEE Trans. Acoust., Speech and Sig. Proc., vol. 24, no. 1, pp. 14-25, 1976). This approach implies stationary noise characteristics during periods of speech. Arslan et al. developed a robust noise estimation method that does not require voice activity detection by recursive averaging with level dependent time constants in each subband (see L. Arslan et al. "New methods for adaptive noise suppression", in Proc. Int. Conf. on Acoustics, Speech and Sig. Proc. (ICASSP-95), Detroit, May 1995). Martin proposes a noise estimation method, which is based on minimum statistics and optimal signal power spectral density (PSD) smoothing (see R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," in IEEE Trans. Speech and Audio Proc., vol. 9, no. 5, pp. 512, July 2001). Further, Ealey et al. present a method for estimating non-stationary noise throughout the duration of the speech utterance by making use of the harmonic structure of the voiced speech spectrum, also referred as harmonic tunnelling (see D. Ealey et al., "Harmonic tunnelling: tracking non-stationary noises during speech," in Proc. Eurospeech Aalborg, 2001). Further, as proposed by Sohn and Sung (see J. Sohn and W. Sung, "A voice activity detector employing soft decision based noise spectrum adaptation," in Proc. IEEE Int. Conf. Acoustics, Speech and Sig. Proc. (ICASSP'98), vol. 1, pp- 365-368, 1998) using soft decision information, the noise spectrum is continuously adapted wheter speech is present or not.

Ephraim and Van Trees propose another important method for noise reduction based on signal subspace decomposition (see Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement", in IEEE Trans. Speech and Audio Proc., vol. 3, pp. 251-266, July 1995). In doing so, the noisy signal is decomposed into a signal-plus-noise subspace and a noise subspace, where these two subspaces are orthogonal. Thus makes it possible to estimate the clean speech signal from the noisy speech signal. The resulting linear estimator is a general Wiener filter with adjustable noise level, to control the trade-off between signal distortion and residual noise, as they cannot be minimized simultanously.

Skoglund and Kleijn point out the importance of the temporal masking property in connection with the excitation of voiced speech (see J. Skoglund and W. B. Kleijn, "On Time-Frequency

Masking in Voiced Speech", in IEEE Trans. Speech and Audio Proc., vol. 8, no. 4, pp. 361-369, July 2000). It is shown that noise between the excitation impulses is more perceptive than noise close to the impulses, and this is especially so for the low pitch speech for which the excitation impulses locates temporal sparsely. Temporal masking is not employed by conventioanl noise reduction methods using frequency domain MMSE estimators. Patent WO 2006 114100 discloses a signal subspace approach taking the temporal masking properties into account.

OBJECT AND SUMMARY OF THE INVENTION

The aim of the present invention consists in providing a single-channel auditory-model based noise suppression method with low-latency processing of wide-band speech signals in the presence of backgound noise. More specifically, the present invention is based on the method of spectral subtraction using a modified decision directed approach comprising oversubtraction and an adjustable noise-level to avoid perceptible musical tones. Further, the present invention uses sub-band processing plus pre- and post-filtering to give consideration to temporal and simultaneous masking inherent to human auditory perception, in particular to minimize perceptible signal distortions during speech periods.

Frequency domain processing is accomplished for the proposed system by using a nonuniform Gammatone filter bank (GTF), which is divided into critical bands, also often referred as Bark bands. This analysis filter bank separates the noisy signal into a plurality of overlapping narrow-band signals, considering spectral (simultaneous) masking properties of human auditory perception.

A pre-processor, which emulates the transfer behaviour of the human outer- and middle ear, is applied to the time-discrete noisy input signal (i.e. the desired speech contaminated by noise and interference).

In each sub-band, the level of the noisy signal is detected and smoothed. These narrow-band level detectors applied to each of the plurality of sub-bands utilize the phase of simple low-order filter sections to provide lowest signal processing delay.

From the smoothed envelope of the sub-band signals the noise level is estimated in each sub-band utilizing a heuristic approach based on recursive Minimum-Statistics.

The instantaneous signal-to-noise-ratio (SNR) in each sub-band is estimated from the envelope

of the noisy signal and the noise level estimate.

The a priori SNR is estimated from the instantaneous SNR by applying the Ephraim-and-Malah Spectral Subtraction Rule (EMSR). In order to minimize the influence of estimation errors an improved decision directed approach (DDA) is proposed, introducing an underestimation parameter and a noise floor parameter.

Temporal masking based on human auditory perception is taken into account by appropriate filtering of the sub-band signals. These non-linear auditory post-masking filters apply recursive averaging to falling slopes of the signal level detected in each sub-band, with the following effects: (a) over-estimating variances of impulsive noise, (b) noise suppression algorithms do not effect signal below the temporal masking threshold, and (c) no additional signal delay is introduced to transient signals, important in speech perception.

A non-linear gain function for each sub-band is derived from the a priori SNR estimates, comprising over-subtraction of the noise signal estimates.

The noisy signal in each sub-band is multiplied by the respective gain in order to suppress the noise signal components.

An optimized nearly perfect reconstruction filter-bank employing a decision criterion for signed summation re-synthesizes the enhanced full-band speech singal.

Finally, a post-processing filter is applied to the enhanced full-band signal to compensate the effect of the pre-processing filter.

NOTES: The noise reduction methods as cited above operate in the frequency domain using the Discrete Time Fourier Transform (DTFT), which is based on block processing of the time-discrete input signals. This block procesisng introduces a signal delay depending on the frame size.

Single channel subtractive-type speech enhancement systems are efficient in reducing back-ground noise; however, they introduce a perceptually annoying residual noise. To deal with this problem, properties of the auditory system are introduced in the enhancement process. This phenomenon is modeled by the calculation of a noise-masking threshold in frequency domain, below which all components are inaudible (see N. Virag, "Single Channel Speech Enhancement Based on Masking Properties of the Human Auditory System", IEEE Trans. on Speech and Audio Proc., vol. 7, no. 2, pp. 126-137, March 1999).

To model auditory masking in subtractive-type speech enhancement systems, filter bank

implementations are especially attractive as they can be adapted to the spectral and temporal resolution of the human ear. The authors propose a noise suppression method based on spectral subtraction combined with Gammatone filter (GTF) banks divided into critcal bands. The concept of critical bands, which describes the resolution of the human auditory systems, leads to a nonlinearly warped frequency scale, called the Bark Scale (see J. O. Smith III and J. S. Abel, "Bark and ERB Bilinear Transforms," IEEE Trans. on Speech and Audio Proc., vol. 7, no. 6, pp. 697-708, Nov. 1999).

The use of Gammatone filter banks outperforms the DTFT based approches in terms of computational complexity and overall system latency. However, the GTF approach allows implementing a low-latency analysis-synthesis scheme with low computational complexity and nearly perfect reconstruction. The proposed synthesis filter creates the broadband output signal by a simple summation of the sub-band signals, introducing a criterion that indicates the necessity of sign alteration before summation. This approach outperforms channel vocoder based approaches as proposed e.g. by McAulay and Malpass (see R. J. McAulay and M. L. Malpass, "Speech Enhancement Using a Soft-Decision Noise Suppression Filter", IEEE Trans. on Acoust., Speech and Sig. Proc., vol. ASSP-28, no. 2, pp. 137-145, April 1980). Within this approach full-band reconstruction of the output signal is performed by the summation of alternately out of phase sub-band signals without considerung the real phase realtions between subbands. This introduces higher distortions in the output signal.

Important note: Sub-band signals without downsampling, as often applied to hearing aid systems, do not require a resynthesis filter bank. Therefore this approach is applicable to low latency speech enhancement systems, but it is computational highly inefficient. The method proposed by the authors allows calculating the output signal from the sub-band signals by simple summation, taking the phase differences into account!

It is worth mentioning, that there are many applications, such as hearing aids or in-car-communication systems, where the computational complexity and signal latency is of utmost importance.

The main advantages of the present invention compared to conventional noise reduction approaches are the significant improvements concerning overall signal latency and computational efficiency.

The invention is not restricted to the following embodiment. It is merely intended to explain

the inventive principle and to illustrate one possible implementation.

According to the invention, the method for low-latency auditory-model based single channel noise suppression and reduction works as an independent module and is intended for installation into a digital signal processing chain, wherein the software-specified algorithm is implemented on a commercially available digital signal processor (DSP), preferably a special DSP for audio applications.

NOTES: With the Ephraim-and-Malah Spectral Subtraction Rule (EMSR) the clean speech signal amplitude is estimated subject to the given amplitude of the noisy signal and the estimated noise variance. To avoid artifacts like musical noise, a modified descision directed approach (DDS) is applied, introducing over-subtraction (under-estimation) of the noise variance plus a noise floor parameter.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG 1 is a schematic illustration of the single-channel sub-band speech enhancement unit of the present invention.

FIG 2 is a schematic ilustration of the non-linear calculation of the gain factor for noise suppression applied to each sub-band.

FIG 3 and 4 show the roof-shaped MMSE-SP attenuation surface dependent on the *a posteriori* (γ_k) and the *a priori* (ξ_k) SNR. To include all values $0 < \gamma_k < \infty$, the x-axis corresponds to γ_k and not $(\gamma_k - 1)$ as in the literature. The dash-dotted line in Fig. 3 marks the transition between the partitions $\sqrt{\frac{G_w}{\gamma_k}}$ and G_w, the dashed line shows the power spectral subtraction contour. The contours of the DDA estimation are plotted in Fig. 4 upon the MMSE-SP attenuation surface. Dashed lines in Fig. 4 show the average of the dynamic relationships between γ_k and ξ_k, solid lines show static relationships.

FIG 5 and 6 are illustrations of the combined (modified) DDA and MMSE-SP estimation behaviour. Dashed lines in Fig. 5 show the average of the dynamic relationships between γ_k and ξ_k, solid lines show static relationships. Two fictitious hysteretis loops of Fig. 6 matching the observations from informal experiments.

FIG 7 shows a block diagram of the overall-system.

FIG 8 shows the over-all system comprising auditory frequency analysis and resynthesis

as front- and back end, and using special low-latency and low-effort speech enhancement in between. A combination of an elaborate noise suppression law with a human auditory model enables high quality performance.

FIG 9 shows an outer- and middle ear filter composed of three second order sections (SOS).

FIG 10 shows an example: Three-Zero Gammatone filter of order 3. The common zero at $z = 1$ is not included in this figure.

FIG 11 shows a familiar way of level-detection. As the signal power is used, the squared amplitude is detected.

FIG 12 shows the Low-Latency FIR level detector

FIG 13 shows a non-linear recursive auditory post-masking filter, responding to falling slopes.

FIG 14 shows a recursive noise level estimator using three time-constant and a counter threshold.

DETAILED DESCRIPTION

In this description new aspects are brought forward concerning the Ephraim and Malah noise suppression rule (EMSR) and the decision directed approach (DDA) for *a priori* signal to noise ratio (SNR) estimation. After partitioning the domain of the amplitude estimator, it becomes clear that the combined DDA estimation obeys an unshaped hysteretic cycle. Introducing a hysteresis width parameter improves the hysteresis shape and reduces musical noise. Eventually, we obtain a more flexible noise suppressor with less dependency on the system sample rate.

I. INTRODUCTION

The Ephraim and Malah amplitude estimator and the Ephraim and Malah decision directed a priori SNR estimate (Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, nr. 6, vol. ASSP-32, pp. 1109–1121, Dec. 1984 and Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, nr.2, vol. ASSP-33, pp. 443-445, Apr. 1985.) are a powerful means of noise suppression in speech

signal processing. Actually there are quite a lot of recently published works on both issues, as the combined algorithm is a powerful tool on the one hand (O. Cappé, "Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor", *IEEE Transactions on Speech and Audio Processing*, nr. 2, vol. 2, pp. 345–349, Apr. 1994), but on the other hand simplifications (P. J. Wolfe and S. J. Godsill, "Simple Alternatives to the Ephraim and Malah Suppression Rule for Speech Enhancement", *Proc. 11th IEEE Signal Processing Workshop*, pp. 496–499, 6–8. Aug 2001) as well as enhancements (I. Cohen and B. Berdugo, "Speech Enhancement for non-stationary noise environments", *Signal Processing*, no. 11, pp. 2403–2418, Elsevier, Nov. 2001; I. Cohen, "Speech Enhancement Using a Noncausal A Priori SNR estimator", *IEEE Signal Processing Letters*, no. 9, pp. 725–728, Sep. 2004; I. Cohen, "Relaxed Statistical Model for Speech Enhancement and A Priori SNR Estimation", Center for Communication and Information Technologies, Israel Institute of Technology, Oct, 2003, CCIT Report no. 443; M. K. Hasan, S. Salahuddin, M. R. Khan, "A Modified A Priori SNR for Speech Enhancement Using Spectral Subtraction Rules", *IEEE Signal Processing Letters*, vol. 11, no. 4, pp 450-453, April 2004) are desirable.

In the amplitude estimation part of the algorithm a signal model is considered in which a noisy signal $y[n]$ consists of speech $x[n]$ and additive noise $d[n]$, at time-index n. The signals $x[n]$ and $d[n]$ are assumed to be statistically independent Gaussian random variables. Due to certain properties of the Fourier transform, the same statistical model can be assumed for corresponding complex short-term spectral amplitudes $\underline{X}_k[m]$ and $\underline{D}_k[m]$ in each frequency bin k, at analysis time m (Underlined variables denote complex quantities here. Therefore, in our notation, $\underline{X}_k[m]$ represents a complex variable. For simplicity of notation $X_k[m]$ shall represent the magnitude $|\underline{X}_k[m]|$.). Given the speech and noise variances $\sigma_{x,k}^2$ and $\sigma_{d,k}^2$, the clean speech amplitude $\underline{X}_k[m]$ can be estimated from the noisy speech $\underline{Y}_k[m]$. An eligible estimator $\hat{\underline{X}}_k[m]$ for the clean speech amplitude is described in section I-A.

The unknown clean speech variance $\sigma_{x,k}^2$ is implicitly determined in the *a priori* SNR estimation part of the algorithm, whereas the noise variance $\sigma_{d,k}^2$ has to be determined in advance, e.g. using Minimum Statistics (P. J. Wolfe and S. J. Godsill, "Simple Alternatives to the Ephraim and Malah Suppression Rule for Speech Enhancement", *Proc. 11th IEEE Signal Processing Workshop*, pp. 496–499, 6–8. Aug 2001), MCRA (I. Cohen and B. Berdugo, "Speech Enhancement for non-stationary noise environments", *Signal Processing*, no. 11, pp. 2403–2418,

Elsevier, Nov. 2001), or Harmonic Tunneling (D. Ealey, H. Kelleher, D. Pearce, "Harmonic Tunneling: Tracking Non-Stationary Noises During Speech", *Proc. Eurospeech*, 2001)

The decision directed estimation described in section I-B determines the *a priori* SNR $\xi_k = \sigma_{x,k}^2/\sigma_{d,k}^2$ in each frequency bin k. Additionally, the noise suppressor utilizes an instantaneous estimate, the so called *a posteriori* SNR, which relates the square of the current noisy magnitude to the noise variance $\gamma_k[m] = Y_k^2[m]/\sigma_{d,k}^2$.

In section II an overview of the combined estimation is given, and its hysteretic shape is presented. Furthermore in section III it is shown how a slight modification can reduce unwanted estimation behaviour and enable a smoother estimation hysteresis.

A. The Ephraim and Malah Suppression Rule (EMSR)

Like mentioned above, the EMSR reconstructs the magnitude of the clean speech signal $\hat{X}_k[m]$ from the noisy observation $Y_k[m]$. As magnitudes at different time-steps m are assumed to be statistically independent, the time index m may be dropped for simplicity of notation.

Ephraim and Malah's MMSE-SA estimator (Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, nr. 6, vol. ASSP-32, pp. 1109–1121, Dec. 1984) solves the Bayesian formula $\hat{X}_k = \mathrm{E}\{X_k|Y_k\}$ to estimate the clean speech magnitude X_k. Applying different distortions to the amplitude, other estimators were derived in similar ways, i.e. the MMSE-LSA (Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, nr.2, vol. ASSP-33, pp. 443-445, Apr. 1985) $\hat{X}_k = e^{\mathrm{E}\{\log X_k|Y_k\}}$, and Wolfe and Godsill's MMSE-SP (P. J. Wolfe and S. J. Godsill, "Simple Alternatives to the Ephraim and Malah Suppression Rule for Speech Enhancement", *Proc. 11th IEEE Signal Processing Workshop*, pp. 496–499, 6–8. Aug 2001) $\hat{X}_k = \sqrt{\mathrm{E}\{X_k^2|Y_k\}}$. For a more detailed description refer to Cohen (I. Cohen, "Relaxed Statistical Model for Speech Enhancement and A Priori SNR Estimation", Center for Communication and Information Technologies, Israel Institute of Technology, Oct, 2003, CCIT Report no. 443).

According to Ephraim and Malah the noisy phase is an optimal estimate of the clean phase.

Thus the reconstruction operator is a real-valued spectral weight $G[m]$:

$$G[m] = \frac{\hat{X}_k[m]}{Y_k[m]} \quad (1)$$

$$\underline{\hat{X}}_k[m] = G[m] \cdot \underline{Y}_k[m]. \quad (2)$$

Because of its simplicity, we have chosen the Wolfe and Godsill MMSE-SP Eq. (3) as the basis of our considerations. The corresponding weighting rule can be expressed as

$$G_{\text{MMSE-SP}}[m] = \sqrt{G_w \cdot \left(G_w + \frac{1}{\gamma_k[m]}\right)}, \quad (3)$$

using the equation of the Wiener Filter

$$G_w = \frac{\hat{X}_k}{\underline{Y}_k}\Big|_{\mathbb{E}\{(\hat{X}_k - X_k)^2\} \to \min.} = \frac{\xi_k}{1+\xi_k}. \quad (4)$$

In order to simplify its application, we partition the reconstruction operator into a few regions

- $(\gamma_k - 1) \ll 1/\xi_k : G_{\text{MMSE-SP}} \approx \sqrt{\frac{G_w}{\gamma_k}}$
- $(\gamma_k - 1) \gg 1/\xi_k : G_{\text{MMSE-SP}} \approx G_w$
- $(\gamma_k - 1) = 1/\xi_k : G_{\text{MMSE-SP}} = \sqrt{G_w \cdot 2/\gamma_k}$.

Additionally, we can approximate the Wiener Filter by

- $\xi_k \ll 1 : G_w \approx \xi_k$
- $\xi_k \gg 1 : G_w \approx 1$.

Combining both, we can divide the MMSE-SP surface into logarithmically flat partitions (see also Fig. 3):

1) $(\gamma_k - 1) \ll 1/\xi_k$, $\xi_k \ll 1 \Rightarrow G_{\text{MMSE-SP}} \approx \sqrt{\xi_k/\gamma_k}$
2) $(\gamma_k - 1) \ll 1/\xi_k$, $\xi_k \gg 1 \Rightarrow G_{\text{MMSE-SP}} \approx \sqrt{1/\gamma_k}$
3) $(\gamma_k - 1) \gg 1/\xi_k$, $\xi_k \ll 1 \Rightarrow G_{\text{MMSE-SP}} \approx \xi_k$
4) $(\gamma_k - 1) \gg 1/\xi_k$, $\xi_k \gg 1 \Rightarrow G_{\text{MMSE-SP}} \approx 1$

Note that in the following sections we use the short form G when we refer to $G_{\text{MMSE-SP}}$.

B. The Decision Directed Approach (DDA)

The DDA combines two basic SNR estimators to a new estimator of the *a priori* SNR ξ_k.

The first estimator is the *instantaneous* SNR $(\gamma_k - 1) = Y_k^2/\sigma_{d,k}^2 - 1 = (Y_k^2 - \sigma_{d,k}^2)/\sigma_{d,k}^2$. Allowing only positive SNR values we get

$$\text{SNR}_{\text{inst}} = \max(\gamma_k - 1, 0), \quad (5)$$

which can be calculated before noise reduction. This *instantaneous* SNR will differ from the true SNR in the following cases:

- when the analysis time-window is too short regarding the stationarity of the signals $x[n]$ and $d[n]$,
- when there is non-stationary noise that can't be identified in detail, or
- when noise and speech signals are highly correlated.

The second estimator describes the *reconstructed* SNR, which is calculated after noise reduction using

$$\mathrm{SNR}_{\mathrm{rec}} = \frac{\hat{X}_k^2}{\sigma_{d,k}^2} = \gamma_k \cdot G^2. \tag{6}$$

In bad SNR conditions, e.g. $0 < \gamma_k < 2$, the *a posteriori* SNR γ_k shows relative variations in time that are smaller than those of $(\gamma_k - 1)$ (Relative variations, e.g. $10 \cdot \log(\gamma_k[m]) - 10 \cdot \log(\gamma_k[m-1])$, are more significant than linear variations regarding human auditory perception.). Ideally, G provides a consistently high attenuation under low SNR conditions. Therefore, the *reconstructed* $\mathrm{SNR}_{\mathrm{rec}}$ will take more consistent values than $\mathrm{SNR}_{\mathrm{inst}}$ in the low SNR case.

Eventually, the DDA for estimation of the *a priori* SNR combines both $\mathrm{SNR}_{\mathrm{inst}}$ and $\mathrm{SNR}_{\mathrm{rec}}$:

$$\xi_k[m] = (1-\alpha) \cdot \mathrm{SNR}_{\mathrm{inst}}[m] + \alpha \cdot \mathrm{SNR}_{\mathrm{rec}}[m-1]. \tag{7}$$

The specific estimation properties can be observed by inserting the suppression gain into the DDA.

II. Combining DDA and EMSR

Using the partitions of the Wolfe and Godsill reconstruction operator $G_{\mathrm{MMSE-SP}}$ from section I-A and inserting them into the Ephraim and Malah DDA (7), the combined *a priori* SNR estimation exhibits the following spheres of action:

1) $\boxed{(\gamma_k - 1) \ll 1/\xi_k,\ \xi_k \ll 1,\ G \approx \sqrt{\xi_k/\gamma_k}}$

$$\begin{aligned} \xi_k[m] \;&\approx\; (1-\alpha) \cdot \max(\gamma_k[m] - 1, 0) + \\ &\quad \alpha \cdot \xi_k[m-1]. \end{aligned} \tag{8}$$

2) $\boxed{(\gamma_k - 1) \ll 1/\xi_k,\ \xi_k \gg 1,\ G \approx \sqrt{1/\gamma_k}}$

$$\begin{aligned} \xi_k[m] &\approx (1-\alpha) \cdot \max(\gamma_k[m] - 1, 0) + \alpha \\ &\approx \alpha. \end{aligned} \tag{9}$$

3) $\boxed{(\gamma_k - 1) \gg 1/\xi_k,\ \xi_k \ll 1,\ G \approx \xi_k}$

$$\begin{aligned} \xi_k[m] &\approx (1-\alpha) \cdot \max(\gamma_k[m] - 1, 0) + \\ &\quad \alpha \cdot \xi_k^2[m-1] \cdot \gamma_k[m-1] \\ &\approx (1-\alpha) \cdot (\gamma_k[m] - 1). \end{aligned} \tag{10}$$

4) $\boxed{(\gamma_k - 1) \gg 1/\xi_k,\ \xi_k \gg 1,\ G \approx 1}$

$$\begin{aligned} \xi_k[m] &\approx (1-\alpha) \cdot \max(\gamma_k[m] - 1, 0) + \\ &\quad \alpha \cdot \gamma_k[m-1] \\ &\approx \alpha \cdot \gamma_k[m-1]. \end{aligned} \tag{11}$$

5) $\boxed{(\gamma_k - 1) = 1/\xi_k,\ \xi_k \ll 1 \Rightarrow G = \sqrt{2 \cdot \xi_k/\gamma_k}}$

$$\begin{aligned} \xi_k[m] &\approx (1-\alpha) \cdot (\gamma_k[m] - 1) + \\ &\quad 2\alpha \cdot \xi_k[m-1]. \end{aligned} \tag{12}$$

The characteristics of the combined approach can be seen in Fig. 4. Considering the magnitude of a speech signal and a constant noise level, i.e. a time-varying *a posteriori* SNR γ_k as input sequence, one can imagine a kind of hysteretic loop evolving on the MMSE-SP surface. Besides the obvious discontinuities in this loop, other properties can be shown (O. Cappé, "Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor", *IEEE Transactions on Speech and Audio Processing*, nr. 2, vol. 2, pp. 345–349, Apr. 1994).

A. Recursive Averaging

1) Expectation by Recursive Averaging: In the above enumeration we can see that in partition 1 the *a priori* SNR estimation corresponds to recursive averaging (Eq. (8)) of the instantaneous $\mathrm{SNR}_{\mathrm{inst}}$ (5). It is feasible to generalize the averaging process by introducing a time-constant τ_{avg} specifying the averaging parameter $\alpha = exp[-1/(\tau_{\mathrm{avg}} \cdot f_s)]$. Here, the sample rate $f_s = 1/T$ denotes the amount of time-frequency transformations per second.

2) The Constant-ξ-Effect: If the *a priori* SNR ξ_k takes a constant value in partition 1, e.g. in case of a large time-constant τ_{avg}, or at the border of ξ_k's value range, the estimator could operate strangely. At a small and constant ξ_k, the system will hold its output magnitude at a constant level. This happens whenever the input is small enough $(Y_k^2[m]/\sigma_{d,k}^2 - 1) \ll 1/\xi_k$ $\Rightarrow Y_k^2[m] \ll \sigma_{d,k}^2/G_w \approx \sigma_{d,k}^2/\xi_k$ (using (8) and its preconditions):

$$\begin{aligned} \hat{X}_k[m] &\approx Y_k[m] \cdot \sqrt{\frac{\xi_k[m]}{\gamma_k[m]}} = Y_k[m] \cdot \sqrt{\frac{\sigma_{x,k}^2}{Y_k^2[m]}} \\ &\approx \sqrt{\sigma_{x,k}^2}. \end{aligned} \tag{13}$$

Under certain circumstances this can lead to annoying additional broad-band noise, which could even be worse when a limitation of ξ_k to a minimum ζ causes a constant output magnitude only for $Y_k^2[m] < \sigma_{d,k}^2/\zeta$.

3) Unstable Recursive Averaging: Following Eq. (12), partition 5 can lead to *a priori* SNR estimation by unstable recursive averaging of $\mathrm{SNR}_{\mathrm{inst}}$ when $\alpha > 1/2$, i.e. ξ_k can increase suddenly in this partition.

B. Partitions Without Recursive Averaging

In the partitions 2, 3, and 4 the recursive averaging interpretation is not useful. Namely, in Eq. (9) the *a priori* SNR estimate ξ_k takes a constant value, and in Eq. (11) ξ_k is determined by a single tap delay. It seems odd that in Eq. (10) the SNR ξ_k is a down-scaled version of $\mathrm{SNR}_{\mathrm{inst}}$.

C. Summary of Properties

Actually, every partition except for 1 and 4 (Eqs. (8) and (11)) exhibits some unexpected behaviour. Defining α by a time-constant, we obtain generalized averaging properties in Eq. (8),

whereas a sample rate dependent behaviour is introduced to the estimation defined by (9)-(12). This form of sample rate dependency rules out a general parameter set suitable for different analysis time-steps and transformation sizes.

Awkward estimation behavior, e.g. the "constant-ξ-effect", and the discontinuities in the hysteresis loop (Fig. 4) give rise to considerations concerning a modification of the DDA and a reconsideration of time-constant and minimum *a priori* SNR quantities.

III. A Modified, Fast Responding DDA

In order to minimize the influence of unexpected estimation performance, we modify the decision directed approach to

$$\xi_k[m] = (1-\alpha)\cdot(\rho\cdot \mathrm{SNR}_{\mathrm{inst}}[m] + \zeta) + \alpha\cdot \mathrm{SNR}_{\mathrm{rec}}[m-1], \tag{14}$$

with ζ being a noise-floor parameter (O. Cappé, "Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor", *IEEE Transactions on Speech and Audio Processing*, nr. 2, vol. 2, pp. 345–349, Apr. 1994) and ρ being an under-estimation parameter of the instantaneous SNR. Similar to the partitions in section II, we can now find:

1) $\boxed{(\gamma_k - 1) \ll 1/\xi_k,\ \xi_k \ll 1,\ G \approx \sqrt{\xi_k/\gamma_k}}$

$$\xi_k[m] \approx \rho(1-\alpha)\cdot\max(\gamma_k[m]-1, 0) + \alpha\cdot\xi_k[m-1]. \tag{15}$$

2) $\boxed{(\gamma_k - 1) \ll 1/\xi_k,\ \xi_k \gg 1,\ G \approx \sqrt{1/\gamma_k}}$

$$\xi_k[m] \approx \alpha. \tag{16}$$

3) $\boxed{(\gamma_k - 1) \gg 1/\xi_k,\ \xi_k \ll 1,\ G \approx \xi_k}$

$$\xi_k[m] \approx \rho(1-\alpha)\cdot(\gamma_k[m]-1). \tag{17}$$

4) $\boxed{(\gamma_k - 1) \gg 1/\xi_k,\ \xi_k \gg 1,\ G \approx 1}$

$$\xi_k[m] \approx \alpha\cdot\gamma_k[m-1]. \tag{18}$$

Regarding the partitions of the new estimator, an over-all estimation scheme can be shown in Fig. 5. Instead of time-constants in the range of speech quasi-stationarity, we now use $\tau_{\mathrm{avg}} = 2$ ms. $\rho = 10^{-15/10}$ ensures that the scale factor in (17) is approximately $\rho(1-\alpha) \approx \rho$, which fixes the discontinuities of the estimation hysteresis. We can choose the noise-floor $\zeta = 10^{-25/10}$ so

small that the maximum attenuation ζ lies at the bottom of the dynamical range of a frequency bin. These measures largely reduce the sample rate dependency described in section II-C and the "constant-ξ-effect" in section II-A.2.

It becomes clear that rising *instantaneous* SNRs are now better attenuated according to Fig. 5 than in Fig. 4. Thus, a stronger attenuation of musical noise, i.e. inconsistently high *instantaneous* SNR, can be provided, while a signal with consistently high SNR is able to pass through the noise suppressor. The two curly loops in Fig. 6 give approximate examples of hysteresis loops occurring during system operation. In the recursive averaging partition the hysteresis path depends on the slope of rising or falling signal amplitude.

The parameter ρ can directly control the suppression hysteresis width and musical noise suppression. Our modification enables a separate control of averaging time-constant and musical noise suppression.

IV. Conclusion

We found a comprehensible way to graphically describe the properties of a combined Wolfe and Godsill spectral amplitude estimation and Ephraim and Malah decision directed *a priori* SNR estimation. This description can similarly be used for other amplitude estimation rules, and provides a new insight into the Ephraim and Malah noise suppressor.

So far the suppression of musical noise has been a trade-off between musical noise suppression and transient distortion. Small modifications in the decision directed estimation rule allow a more flexible handling of musical noise suppression, while reducing dependencies on the analysis time-step and the "constant-ξ-effect". An informal listening test using the modified algorithm with adjustable analysis time/frequency-resolution (filterbank approach) showed useful enhancements in the over-all algorithm.

Our further work will introduce our descriptive methods into the more elaborate estimation approaches of Cohen (I. Cohen, "Speech Enhancement Using a Noncausal A Priori SNR estimator", *IEEE Signal Processing Letters*, no. 9, pp. 725–728, Sep. 2004) or Hasan (M. K. Hasan, S. Salahuddin, M. R. Khan, "A Modified A Priori SNR for Speech Enhancement Using Spectral Subtraction Rules", *IEEE Signal Processing Letters*, vol. 11, no. 4, pp 450-453, April 2004).

APPARATUS FOR LOW-LATENCY SINGLE CHANNEL SPEECH ENHANCEMENT

In the following a preffered embodiment will be described, however the invention is not limited to this embodiment.

The reduction of *musical noise* in noise suppression algrithms still an issue in noise reduction. Although the Ephraim and Malah suppression rule (EMSR) and the decision directed approach (DDA) show a good performance, additional means have to be applied. Moreover, processing delays arising from signal analysis (fast Fourier transform, FFT) pose a problem in real-time applications. Essential improvements in both issues can be achieved by implementing signal analysis and filtering approaches capable of modelling the human auditory perception and latency reduction.

V. Introduction

The major part of this description is dedicated to auditory signal preparation and analysis, using efficient algorithms with low-latency. Our system combines an auditory Gammatone filterbank (R. F. Lyon, "The All-Pole Gammatone Filter and Auditory Models", Proc. Forum Acusticum, Antwerpen 1996; L. Lin, E. Ambikairajah, W. H. Holmes, "Auditory Filterbank Design Using Masking Curves", Proc. EUROSPEECH Scandinavia, 7th European Conference on Speech Communication and Technology, 2001; L. Lin, E. Ambikairajah, W. H. Holmes, "Perceptual Domain Based Speech and Audio Coder", Proc. of the third International Symposion DSPCS 2002, Sydney, Jan. 28-31, 2002) with the Ephraim and Malah noise suppression rule (Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, nr. 6, vol. ASSP-32, pp. 1109–1121, Dec. 1984; Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, nr.2, vol. ASSP-33, pp. 443-445, Apr. 1985; P. J. Wolfe and S. J. Godsill, "Simple Alternatives to the Ephraim and Malah Suppression Rule for Speech Enhancement", *Proc. 11th IEEE Signal Processing Workshop*, pp. 496–499, 6–8. Aug 2001). This combination is newly introduced by the authors, whereas the combination of an auditory Gammatone filterbank with a Wiener noise suppressor is known from (L. Lin,

E. Ambikairajah, "Speech Denoising Based on an Auditory Filterbank", 6th ICSP, International Conference on Signal Processing, (552-555), 26-30 Aug. 2002), and a frequency domain solution is known from WO 00/30264 (International applicatoin No. PCT/SG99/00119). Furthermore, the integration of a time-domain outer and middle ear filter, as well as the integration of a non-linear temporal post-masking filter (G. Stoll, J. G. Beerends, R. Bitto, K. Brandenburg, C. Colomes, B. Feiten, M. Keyhl, C. Schmidmer, T. Sporer, T. Thiede, W. C. Treurniet, "PEAQ - der neue ITU-Standard zur objektiven Messung der wahrgenommenen Audioqualität", RTM - Rundfunktechnische Mitteilungen, die Fachzeitschrift für Hörfunk und Fernsehtechnik, 43. Jahrgang, ISSN 0035-9890 (81-120), Firma Mensing GmbH + Co. KG, Abteilung Verlag, Sept 1999; L. Lin, E. Ambikairajah, W. H. Holmes, "Perceptual Domain Based Speech and Audio Coder", Proc. of the third International Symposion DSPCS 2002, Sydney, Jan. 28-31, 2002) into the noise suppression system is new. Additionally, a narrow band low-latency level detection exploiting the phase of simple first order filters is newly introduced. Finally, we present a simple scheme for signal reconstruction (resynthesis) avoiding band-edge signal cancellation.

- Combination of an auditory Gammatone Filterbank and the EMSR noise suppressor in a time-domain approach
- Integration of outer and middle ear filters into the suppression system in a time-domain approach
- Integration of an auditory post-masking filter
- Low-latency narrow band level-detector
- Low-Effort Wolfe and Godsill signal restauration
- Low-latency up-sampling
- Low-latency resynthesis restraining destructive interference

VI. System Overview

The over-all system is shown in a block diagram in Fig. 7. It can be implemented as analog or digital effect processor or as a part of a software algorithm.

Inside the over-all system there are several subsystems Fig. 8:

- an outer and middle ear filter (H_{OME}),
- a Gammatone filterbank analysis section (GFB),
- the low-latency level detection (LD),

- the auditory post-masking filter (PM),
- a recursive noise spectrum estimation (NE),
- the spectral sutraction weight (EMSR),
- the low-latency upsampling (L ↑)
- the vocoder stage, and
- the inverse outer and middle ear filter (H_{OME}).

VII. Outer and Middle Ear Filter

An outer and middle ear filter consists of three second order sections (SOS) representing the following physiological parts of the human ear (E. Zwicker, H. Fastl, "Psychoacoustics, facts and models", Springer, Berlin Heidelberg, 1999; E. Terhardt, "Akustische Kommunikation", Springer, Berlin Heidelberg, 1998):

1) the high-pass attenuation curve below 1 kHz modelling the 100-Phon curve, which represents the acoustic impedance of the outer ear and the mechanic impedance of the middle ear ossicles,
2) the resonance of the ear channel, and
3) the low-pass attenuation curve above 1 kHz modelling the threshold of hearing.

The latter two filters are optional, whereas the high-pass component is mandatory and reduces the influence of low-frequency noise on the noise suppressor.

In the end, a filter structure providing an appropriate magnitude transfer function could look like Fig. 9. All three filter sections have to be second order sections to provide appropriate slopes. The outer filter skirts can be modelled as second order low- and high-pass shelving filters, whereas the resonance can be modelled as parametric peak-filter (P. Dutilleux, U. Zölzer, "DAFX", Wiley&Sons, 2002).

The filter inversion is straight-forward. If there should be zeros at e.g. $z = 1$ in the z-domain, the inverse filter can't undo this, so perhaps $z = 0.99$ could be a proper choice for a pole location inverting a $z = 1$ zero.

VIII. Frequency Grouping / Auditory Bandwidths

Frequency grouping is an imporant effect in human loudness perception. The perceived loudnesss consists of particular loudnesses associated to individual frequency ranges. An auditory

frequency scale can be used to model this frequency grouping effects, the units of which can be seen as frequency resolution of human auditory loudness perception (E. Zwicker, H. Fastl, "Psychoacoustics, facts and models", Springer, Berlin Heidelberg, 1999). We denote an arbitrary auditory frequency transform with the operator $\mathfrak{B}\{\cdot\}$ and the corresponding inverse frequency transform with $\mathfrak{B}^{-1}\{\cdot\}$. A reasonable frequency scale using a low number of frequency groups can be given by the formula of Traunmüller (E. Terhardt, "Akustische Kommunikation", Springer, Berlin Heidelberg, 1998)

$$\nu/[\text{Bark}] = \mathfrak{B}\{f/[\text{Hz}]\} = \frac{26810 \cdot f}{1960 + f} - 0.53. \tag{19}$$

Accordingly, the inverse transform $\mathfrak{B}^{-1}\{\cdot\}$ is

$$f/[\text{Hz}] = \mathfrak{B}\{\nu/[\text{Bark}]\} = 1960 \cdot \frac{\nu + 0.53}{26.28 - \nu}. \tag{20}$$

The center frequencies f_k of the auditory filterbank can be calculated applying the inverse transform $f_k = \mathfrak{B}^{-1}\{\nu_k\}$ on an equally spaced scale ν_k (with spacing $d\nu$, e.g. $d\nu = 1[\text{Bark}]$) in the Bark-domain. Similarly the bandwidths B_k can be derived from $B_k = \mathfrak{B}^{-1}\{\nu_k + d\nu/2\} - \mathfrak{B}^{-1}\{\nu_k - d\nu/2\}$. Other Bark-scames (e.g. E. Zwicker, H. Fastl, "Psychoacoustics, facts and models", Springer, Berlin Heidelberg, 1999) use smaller bandwidths resulting in auditory filters with more group-delay, thus the above spacing is preferred.

Note that we use ν for the Bark-frequency instead of z in order to avoid confusion with the z-domain variable z.

IX. Auditory Gammatone Filters

Auditory Gammatone filters (R. F. Lyon, "The All-Pole Gammatone Filter and Auditory Models", Proc. Forum Acusticum, Antwerpen 1996) can be efficiently implemented in the time-domain, allowing the separation of a broadband audio signal into auditory band signals. The magnitude response of the Gammatone Filter corresponds to the simultaneous masking properties of the human ear. Plotting the magnitude of this filter along an auditory frequency scale the filter shape remains the same, whatever center frequency the filter is designed to have. The arbitrary form representing the family of Gammatone-filters of the order m is shown below, wherein k is the filterbank channel index. A corresponding z-transform wherein $*GF$ denotes an arbitrary

Gammatone filter (e.g. GF, APGF, OZGF, TZGF), is:

$$H_{*\mathrm{GF},k}(z) = g_{*\mathrm{GF}} \cdot H_{\mathrm{num},k}(z) \cdot \prod_{r=1}^{m} \frac{1}{1 - 2 \cdot r_k \cdot \cos(\theta_k) \cdot z^{-1} + r_k \cdot z^{-2}}. \quad (21)$$

Digital center frequencies θ_k and pole radii r_k are derived from the continuous-time quantities center frequency f_k, bandwidth B_k, the band-edge rejection C_{dB} (e.g. $C_{dB} = -5$[dB]), and the sample rate f_s:

$$\theta_k = 2\pi \cdot \frac{f_k}{f_s}, \qquad r_k = 1 - \frac{2\pi \cdot \frac{B_k}{f_s}}{2 \cdot \sqrt{10^{-\frac{C_{dB}}{10 \cdot m}} - 1}}. \quad (22)$$

An auditory Gammatone Filterbank represents of a set of overlapping Gammatone filters that devide the auditory frequeny scale in equally spaced frequency bands. An order $m = 4$ is frequently used in literature, whereas the order $m = 3$ is proposed to minimize computational cost. The term $g_{*\mathrm{GF}}$ shall be adjusted so that unity gain at the center frequency f_k can be provided. For a special form of gammatone filter the system $H_{\mathrm{num},k}(z)$ has to be adapted suitably as shown in the following sub-sections.

A. Ordinary Gammatone filter

The odinary Gammatone filter (GF; R. F. Lyon, "The All-Pole Gammatone Filter and Auditory Models", Proc. Forum Acusticum, Antwerpen 1996) has to be derived from the continuous-time impulse response using the Laplace- and impulse-invariance transform (A. V. Oppenheim, R. W. Schafer, J. R. Buck, "Discrete-Time Signal Processing", Prentice Hall, 1999):

$$h(t) = t^{m-1} e^{-B_k \cdot t} \cos(2\pi f_k t), \quad (23)$$

which determines the unknown polynomial $H_{\mathrm{num},k}(z)$ in the above equation (21). Due to its shape and computational cost its use is not recommended.

B. All-Pole Gammatone filter

An All-Pole Gammatone filter (APGF) can be obtained when just cancelling the polynomial $H_{\mathrm{num},k}(z) = 1$ in equation (21). It is the most efficient Gammatone filter (R. F. Lyon, "The All-Pole Gammatone Filter and Auditory Models", Proc. Forum Acusticum, Antwerpen 1996).

C. One-Zero Gammatone filter

Just setting $H_{\mathrm{num},k}(z) = (1-z^{-1})$ in equation (21) leads to the so-called One-Zero Gammatone filter (R. F. Lyon, "The All-Pole Gammatone Filter and Auditory Models", Proc. Forum Acusticum, Antwerpen 1996). The One-Zero Gammatone (OZGF) can be efficiently composed of a common "One-Zero" for all channels k before splitting up into k All-Pole Gammatone filters.

D. Three-Zero Gammatone filter

When adding a pair of complex conjugate zeroes $z = r_z \cdot e^{\pm\theta_{z,k}}$ with the digital frequency $\theta_{z,k}$ at 1 Bark above center frequency θ_k, with radius $r_z \approx 0.98$, and one additional zero at $z = 1$, we obtain $H_{\mathrm{num},k}(z) = (1 - 2r_z\cos(\theta_{z,k})z^{-1} + r_z^2 z^{-2}) \cdot (1 - z^{-1})$ for the Three-Zero Gammatone (TZGF) filter with its improved shape (L. Lin, E. Ambikairajah, W. H. Holmes, "Auditory Filterbank Design Using Masking Curves", Proc. EUROSPEECH Scandinavia, 7th European Conference on Speech Communication and Technology, 2001). In comparison, the computational cost of the One-Zero Gammatone filter of order $m+1$ is equal to the cost of the Three-Zero Gammatone Filter of order m, when again, a single "One-Zero" is commonly used by all channels k. Appropriate transforms and digital frequency calculation $\theta_{z,k}$ follow from the equations (19)(20) and (22).

X. Resynthesis

A resynthesis of a broadband signal from the auditory band signals can be implemented as an addition of all signal bands. Unfortunately this can bring destructive signal cancellation at the overlap between neighbouring signal channels. Therefore we derived a simple criterion that indicates the necessity of a sign alteration for every second channel befor signal summation

$$f_{\mathrm{sgn}} = \mathrm{sgn}\left\{\cos\left[2m \cdot \arctan\left(\sqrt{10^{-\frac{C_{\mathrm{dB}}}{10 \cdot m}} - 1}\right)\right]\right\}. \qquad (24)$$

Using this formula, the frequency response of the superposition of all signals lies in the range of approximately $C_{\mathrm{dB}} + 3$[dB] and 0[dB]. Omitting a necessary sign alteration can result in a destructive signal cancellation at the band-edges of adjacent filters.

XI. (Low-Latency) Level Detection

Masking effects modelled by the auditory filterbank cannot be exploited unless the amplitude of the filterbank channels is determined. Suitable ways of level detection are proposed in the following sub-sections.

We propose to use the first simple approach for the high-frequency channels, and the low-latency approach for low-frequency bands.

A. Ordinary Level-Detection with pre-masking

Usually non-linearities like e.g. the absolute value, square, or half-wave rectification are used to transform the signal amplitude into the base band around 0 Hz. Further a smoothing filter removes components at higher frequencies, and in the end the desired amplitude signal is found. Fig. 11 provides an example, which also takes the form-factor F into account.

The commonly used approach of amplitude detection is computationally efficient, but smoothing filters involve group delays in the signal path that have to be compensated for. We recommend to describe the recursive smoothing parameter α by a time-constant τ_{avg} in [s]

$$\alpha = e^{-\frac{1}{\tau \cdot f_s}}. \tag{25}$$

Suitable time-constants match the auditory pre-masking time-constant, which is approximately $\tau_{avg} \approx 2[ms]$ (G. Stoll, J. G. Beerends, R. Bitto, K. Brandenburg, C. Colomes, B. Feiten, M. Keyhl, C. Schmidmer, T. Sporer, T. Thiede, W. C. Treurniet, "PEAQ - der neue ITU-Standard zur objektiven Messung der wahrgenommenen Audioqualität", RTM - Rundfunktechnische Mitteilungen, die Fachzeitschrift für Hörfunk und Fernsehtechnik, 43. Jahrgang, ISSN 0035-9890 (81-120), Firma Mensing GmbH + Co. KG, Abteilung Verlag, Sept 1999).

B. Low-Latency Level detection

Our new method exploits the phase of simple filter sections. This method for level detection is also applicable to other technical fields and not restricted to noise suppression alone.

Using a Hilbert transform, a consistent 90° phase shift can be brought to a broad band signal. Summing up the squares of the original and the shifted signal, squared amplitudes (i.e. signal power) remain while sinusoidal components cancel. But a causal implementation of the Hilbert transform doesn't exist.

Unlike an ideal Hilbert transformator, we only need 90° phase shift in the considered frequency range, i.e. in the corresponding auditory frequency group.

We propose to use the following kinds of filters to provide a 90° phase shift at a frequency θ_k:

- a simple FIR first order section,
- a simple IIR first order all-pass (AP), and
- a simple delay line providing a $\lambda/4$ delay at θ_k

Each of the above mentioned methods can provide a 90° phase shift to a virtually arbitrary frequency θ_k and is therefore suitable.

One can choose between the following properties:

- FIR: numerical not stable around $\theta_k = [0, \pi/2, \pi]$, providing the broadest band featuring a 90° phase.
- AP: numerical unstable around $\theta_k = [0, \pi/2, \pi]$, the 90° phase frequency band is smaller, computational effort bigger.
- $\lambda/4$-delay: numerical stable, the smallest frequency band of 90° phase, computational effort low, more memory needed.

Fig. 12 provides an example for the FIR level detection method. Appropriate parameters can be found using the phase-equations for the corresponding systems, e.g. A. V. Oppenheim, R. W. Schafer, J. R. Buck, "Discrete-Time Signal Processing", Prentice Hall, 1999.

XII. Auditory Post-Masking

Using a non-linear post-masking filter (i.e. recursive averaging only responding to a falling slope) exhibits several benefits:

- impulsive noise variance is slightly over-estimated (over-subtraction) because of the post-masking.
- noise suppression algorithms cannot attenuate signals until the auditory post masking time has elapsed.
- aliasing effects after downsampling or ripples in the amplitude signals are reduced due the post-masking smoothing operation.
- though smoothing is applied, no group delay is brought to the amplitude of important transient signals

We propose a structure that works on the signal power detected in each channel (cf. Fig. 13, L. Lin, E. Ambikairajah, W. H. Holmes, "Perceptual Domain Based Speech and Audio Coder", Proc. of the third International Symposion DSPCS 2002, Sydney, Jan. 28-31, 2002).

The averaging parameter α_k in the channel k has to correspond to human auditory post-masking time-constants at corresponding frequencies f_k. Therefore, we use following equation to derive the averaging parameter α:

$$\alpha_k = e^{-\frac{L}{G \cdot \tau_k \cdot f_s}}. \tag{26}$$

A parameter G can be used to scale the post-masking time-constants if useful.

The time-constant for 1[Bark] is approximately $\tau_{\nu_1} \approx 40$[ms], and for 20[Bark] approximately $\tau_{\nu_{20}} \approx 4$[ms] (G. Stoll, J. G. Beerends, R. Bitto, K. Brandenburg, C. Colomes, B. Feiten, M. Keyhl, C. Schmidmer, T. Sporer, T. Thiede, W. C. Treurniet, "PEAQ - der neue ITU-Standard zur objektiven Messung der wahrgenommenen Audioqualität", RTM - Rundfunktechnische Mitteilungen, die Fachzeitschrift für Hörfunk und Fernsehtechnik, 43. Jahrgang, ISSN 0035-9890 (81-120), Firma Mensing GmbH + Co. KG, Abteilung Verlag, Sept 1999). Following equation can be used to derive τ_k

$$\tau_k/[\text{ms}] = \frac{1}{\frac{1}{\tau_{\nu_1}} + \frac{\frac{1}{\tau_{\nu_{20}}} - \frac{1}{\tau_{\nu_1}}}{\mathfrak{B}^{-1}\{20\} - \mathfrak{B}^{-1}\{1\}} \cdot (f_k - \mathfrak{B}^{-1}\{1\})}}. \tag{27}$$

Alternatively, the equation in above cited reference can be used, but our formula provides a suitable interpolation with longer time-constants.

XIII. Recursive Minimum Statistics

We can use the structure in Fig. 14 to estimate the noise level in each frequency band. Similar approaches can be found in R. Martin, "Noise Power Spectral Estimation Based on Optimal Smoothing and Minimum Statistics", *IEEE Transactions on Speech and Audio Processing*, nr. 5, vol. 9, pp. 504–512, Jul. 2001 or WO 00/30264 (International applicatoin No. PCT/SG99/00119).

This method essentially applies three time-constants of averaging to the signal level. Falling slopes are sligthtly averaged, whereas during rising input slope, the output is held constant (i.e. infinitely large time-constant) during the period of N_w sampling intervals. When N_w sampling intervals are exceeded, the rising signal slope is averaged by a third time constant. The time-constants can be similarly converted to recursive averaging parameters as in equation (25) and (26).

An appropriate counter threshold N_w can be calculated using a continuous time interval T_w

$$N_w = \text{round}(T_w \cdot f_s). \tag{28}$$

Suitable to utterances or words of human speech, this time interval can be chosen e.g. $T_w \approx 1.5s$. The falling slope time-constant can be a scaled version of the post-masking time-constants τ_k, or e.g. constant 200[ms].

The rising slope time-constant defining β can be approximately 700[ms], which corresponds to a velocity of appoximately 6[dB]/[s]. Unlike other time-constants, this one is proposed to be equal for all channels k.

The saturation operation in Fig. 14 can be expressed as:

$$f(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{else.} \end{cases} \tag{29}$$

XIV. Ephraim and Malah Noise Suppression Rule (EMSR)

With the EMSR (Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, nr. 6, vol. ASSP-32, pp. 1109–1121, Dec. 1984; Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, nr.2, vol. ASSP-33, pp. 443-445, Apr. 1985) we can estimate the clean speech amplitude subject to the given noisy speech amplitude and the noise variance. We can e.g. use the Wolfe and Godsill definition of the spectral weight (P. J. Wolfe and S. J. Godsill, "Simple Alternatives to the Ephraim and Malah Suppression Rule for Speech Enhancement", *Proc. 11th IEEE Signal Processing Workshop*, pp. 496–499, 6–8. Aug 2001) and a modified decision directed approach (F. Zotter, M. Noisternig, R. Höldrich, "Speech Enhancement Using the Ephraim and Malah Suppression Rule and Decision Directed Approach: A Hysteretic Process", to appear in IEEE Signal Processing Letters, 2005. First manuscript submitted Jan 24, 2005)

$$g_k[m] = \sqrt{G_{w,k}[m] \cdot \left(G_{w,k}[m] + \frac{1}{\gamma_k[m]} \right)}. \tag{30}$$

The following relations are involved in the above equation:

$$\gamma_k[m] = \frac{\|Y_k[m]\|^2}{\sigma_{d,k}^2[m]} \quad (31)$$

$$G_{w,k}[m] = \frac{\xi_k[m]}{1+\xi_k[m]} \quad (32)$$

$$\xi_k[m] = \alpha \cdot \min(\gamma_k[m]-1, 0) + \rho \cdot (1-\alpha) \cdot \gamma_k[m-1] \cdot g_k^2[m-1] + \zeta \quad (33)$$

$$\alpha = e^{-\frac{L}{\tau_{snr,k} \cdot f_s}} \quad (34)$$

$$m = L \cdot n \quad (35)$$

The noise variance $\sigma_{d,k}^2[m]$ is given by the noise estimation algorithm; m and n are time indices, f_s is the system sample rate and L a down-sampling factor.

According to Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, nr. 6, vol. ASSP-32, pp. 1109–1121, Dec. 1984, $\gamma_k[m]$ is the *a posteriori* SNR, and $\xi_k[m]$ is the *a priori* SNR. $G_{w,k}[m]$ is the spectral weight of a Wiener filter, α is an averaging parameter, defined by an averaging time-constant $\tau_{snr,k}$, which is either approximately 2[ms] (F. Zotter, M. Noisternig, R. Höldrich, "Speech Enhancement Using the Ephraim and Malah Suppression Rule and Decision Directed Approach: A Hysteretic Process", to appear in IEEE Signal Processing Letters, 2005. First manuscript submitted Jan 24, 2005) or derived from the auditory post-masking time-constants.

The "over-subtraction factor" ρ (cf. Zotter *et al*) can be chosen to be $\rho = 10^{-15/10}$, and the noise-floor parameter ζ can be $\zeta = 10^{-40/10}$.

XV. Low-Latency Up-sampling

Usually up-sampling needs either a processing-delay or a group-delay due to the interpolation operation involved. Such a delay is approximately L samples long, using the up-sampling factor L.

We propose to use a special method for up-sampling introducing no additional delays. This can be done if the signal is devided into buffers (perferably with the buffer size of the ADC and DAC).

When in every signal block the last sample of the preceeding block is given, it is possible to linearly interpolate to the following given sample instantaneously. Therefore, the last sample in every block must correspond to a sampling instant at the lower sampling rate.

XVI. Conclusions

Frequency domain solutions using equivalent auditory models require delays in the range of 10 miliseconds, the implementation of our system with 20 frequency bands and the third order TZGF has a mean latency of 3.5 up to 4 miliseconds. The required computational cost is about approximately 8.9 MIPs at $fs = 16$[kHz], which is only slightly more than DFT solutions need (7 MIPs). We also apply a slightly modified Ephraim and Malah suppression rule (EMSR) using the simplified Wolfe and Godsill formula and modified decision directed approach.

The disclosure of all cited publications is included in its entirety into this description.

CLAIMS

1. A method for suppressing noise in an input audio signal (y[n]) which comprises a wanted signal component ($x[n]$) and a noise signal component, the method comprising the steps of
- dividing the input audio signal ($y[n]$) into a plurality of frequency subbands ($y_k[n]$) by means of an analysis band splitter,
- suppressing noise in each of the subbands ($y_k[n]$) by a plurality of noise supressing processors,
- recombining the plurality of subbands ($y_k[n]$) into an output signal ($\hat{x}[n]$) by means of a synthesis filter,
all steps being performed in the time domain.

2. Method according to claim 1, characterized in that the dividing of the input audio signal into a plurality of subbands by means of the analysis band splitter is performed according to human auditory loudness perception.

3. Method according to claim 2, characterized in that the analysis band splitter comprises a Gammatone filter bank (GFB), preferably a nonuniform Gammatone filter bank.

4. Method according to any of claims 1 to 3, characterized in that a pre-processor (H_{OME}) and post-processor (H_{IOME}) perform non-linear filtering to the input audio signal, comprising
a. a pre-processing filter, which emulates the transfer behaviour of the human outer and middle ear applied to the time-discrete noisy input audio signal
b. a post-processing filter applied to the enhanced full-band signal to compensate the effect of the pre-processing filter.

5. Method according to any of claims 1 to 4, characterized in that each noise processor is comprised of a signal level detector (LD), a noise estimator (NE), an auditory masking filter (PM) and a subtraction processor.

6. Method according to claim 5, wherein said signal level detector (LD) exploits the phase of low-order filter sections to generate a quadrature signal and an in-phase signal out of

the sub-band signal $y_k[n]$) and summing up the squared amplitudes of these signals.

7. Method according to claim 5, wherein said noise estimator generates a sub-band noise value by performing smoothing based on Minimum Statistics, more particularly weighted averaging of the previous noise value and the current input value with three different time constants is applied.

8. Method according to claim 5 or 6, wherein said auditory masking filter uses the signal power detected in each sub-channel to generate a temporal masking behaviour based on human auditory perception, more particularly non-linear weighted averaging of the previous signal value and the current sub-band input value is applied only on the falling slope depending on the level detected in each sub-band.

9. Method according to claims 1 to 8, wherein the update of the noise estimator depends on the current input value compared to time-varying, level dependent thresholds, i.e. if the current input value is greater than a predetermined threshold value the current input value is not considered to be noiwe and said noise estimator is not updated.

10. Method according to claims 1 to 9, wherein the noise suppression in each of the subbands is performed using the Ephraim and Malah noise suppression rule (EMSR).

11. Method according to claims 1 to 10, wherein the noise suppression in each of the subbands is performed decision directed approach (DDA).

12. Apparatus for suppressing noise in an input audio signal (y[n]) which comprises a wanted signal component ($x[n]$) and a noise signal component, the apparatus comprising
- an analysis band splitter for dividing the input audio signal ($y[n]$) into a plurality of frequency subbands ($y_k[n]$),
- a plurality of noise supressing processors for suppressing noise in each of the subbands ($y_k[n]$),
- a synthesis filter for recombining the plurality of subbands ($y_k[n]$) into an output signal ($\hat{x}[n]$),
analysis band splitter, noise supressing processors and synthesis filter working in the time domain.

13. Apparatus according to claim 12, characterized in that a level detector (LD) is provided in each of the subbands.

14. Apparatus according to claim 13, characterized in that said signal level detector (LD) exploits the phase of low-order filter sections to generate a quadrature signal and an in-phase signal out of the sub-band signal ($y_k[n]$) and summing up the squared amplitudes of these signals.

15. Apparatus according to claim 14, characterized in that said quadrature signal is generated by FIR first order section provided in the level detector (LD).

16. Apparatus according to claim 14, characterized in that said quadrature signal is generated by FIR first order all-pass (AP) provided in the level detector (LD).

17. Apparatus according to claim 14, characterized in that said quadrature signal is generated by a delay line providing a $\lambda/4$ delay at the digital center frequency (θ_k).

18. Apparatus according to any of claims 12 to 17, characterized in that each noise processor is comprised of a signal level detector (LD), a noise estimator (NE), an auditory masking filter (PM) and a subtraction processor.

19. Apparatus according to any of Claims 12 to 18, characterized in that the analysis band splitter comprises a Gammatone filter bank (GFB), preferably a nonuniform Gammatone filter bank.

20. Apparatus according to any of claims 12 to 19, characterized in that a pre-processor (H_{OME}) and post-processor (H_{IOME}) is provided for performing non-linear filtering to the input audio signal, comprising

a. a pre-processing filter, which emulates the transfer behaviour of the human outer and middle ear applied to the time-discrete noisy input audio signal b. a post-processing filter applied to the enhanced full-band signal to compensate the effect of the pre-processing filter.

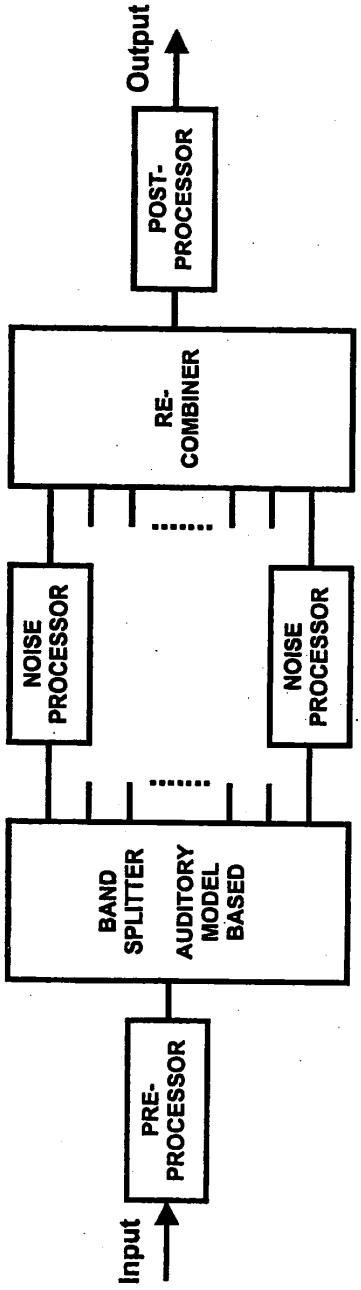

Output
POST-PROCESSOR
RE-COMBINER
NOISE PROCESSOR
NOISE PROCESSOR
BAND SPLITTER AUDITORY MODEL BASED
PRE-PROCESSOR
Input

Fig. 1

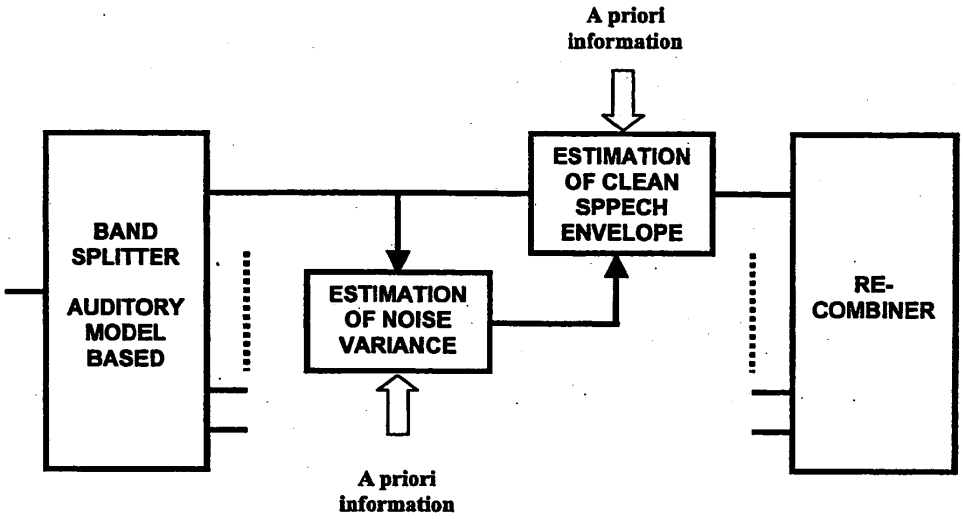

A priori information
ESTIMATION OF CLEAN SPPECH ENVELOPE
BAND SPLITTER
AUDITORY MODEL BASED
ESTIMATION OF NOISE VARIANCE
RE-COMBINER
A priori information

Fig. 2

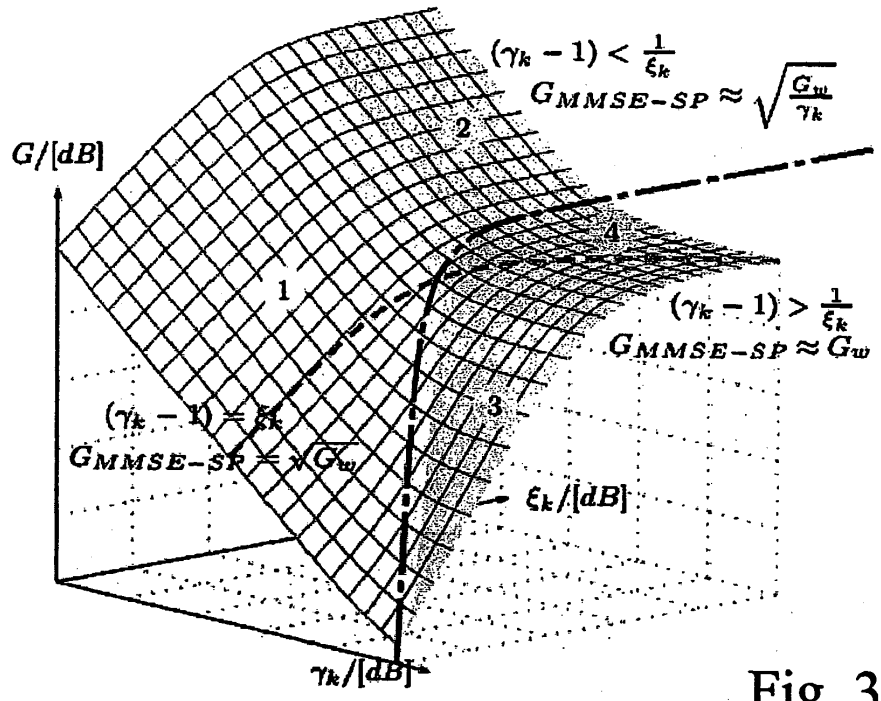

$(\gamma_k - 1) < \frac{1}{\xi_k}$
$G_{MMSE-SP} \approx \sqrt{\frac{G_w}{\gamma_k}}$
$G/[dB]$
2
1
4
$(\gamma_k - 1) > \frac{1}{\xi_k}$
$G_{MMSE-SP} \approx G_w$
$(\gamma_k - 1) = \xi_k$
$G_{MMSE-SP} = \sqrt{G_w}$
3
$\xi_k/[dB]$
$\gamma_k/[dB]$

Fig. 3

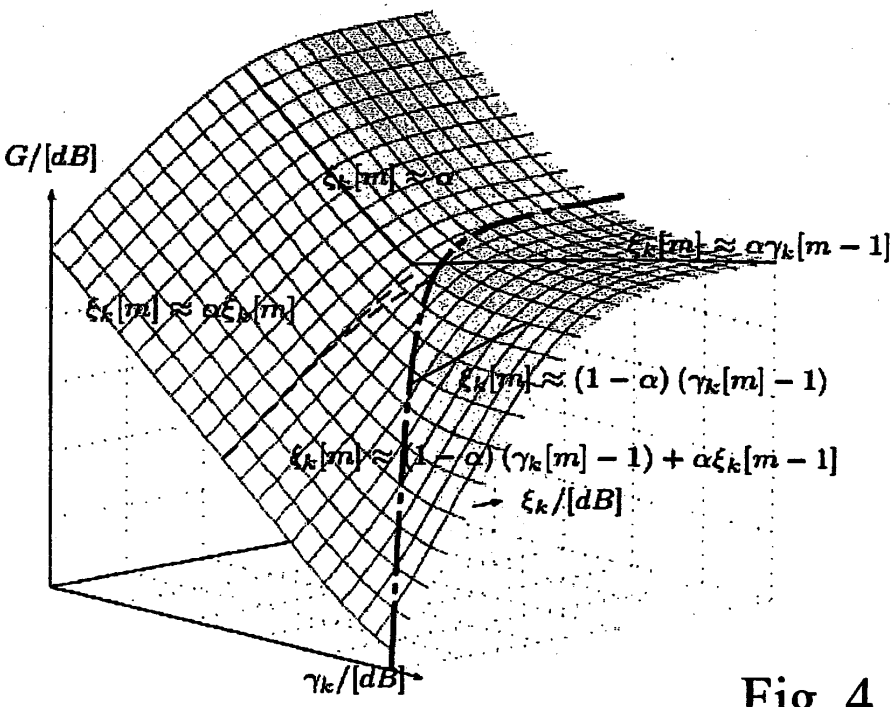

$G/[dB]$
$\xi_k[m] \approx \alpha$
$\xi_k[m] \approx \alpha\gamma_k[m-1]$
$\xi_k[m] \approx \alpha\xi_k[m]$
$\xi_k[m] \approx (1-\alpha)(\gamma_k[m]-1)$
$\xi_k[m] \approx (1-\alpha)(\gamma_k[m]-1) + \alpha\xi_k[m-1]$
$\xi_k/[dB]$
$\gamma_k/[dB]$

Fig. 4

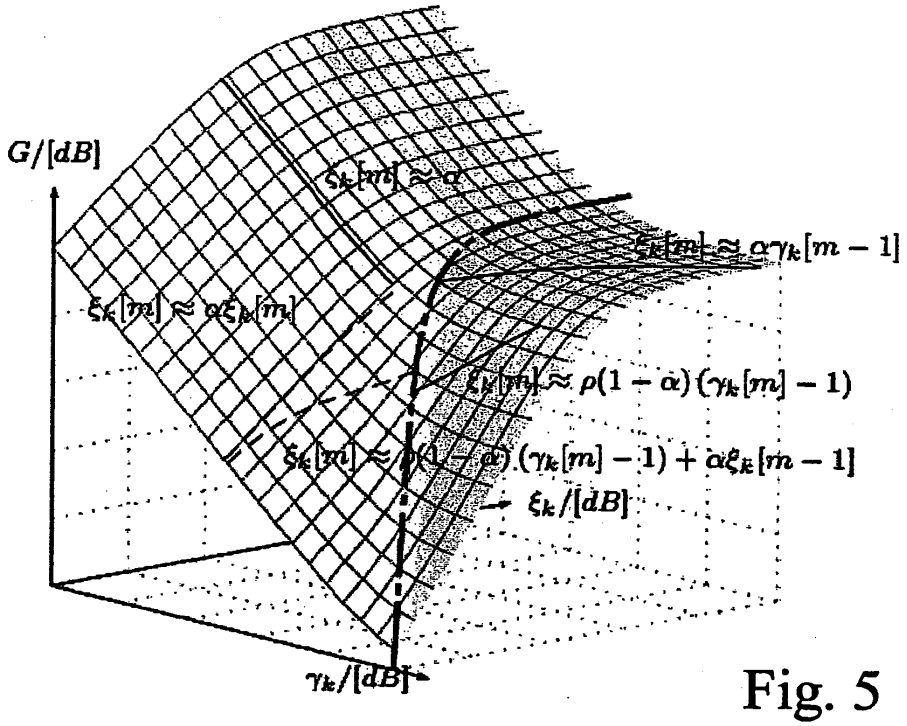

G/[dB]
$\xi_k[m] \approx \alpha$
$\xi_k[m] \approx \alpha\gamma_k[m-1]$
$\xi_k[m] \approx \alpha\xi_k[m]$
$\xi_k[m] \approx \rho(1-\alpha)(\gamma_k[m]-1)$
$\xi_k[m] \approx \rho(1-\alpha)(\gamma_k[m]-1)+\alpha\xi_k[m-1]$
$\xi_k/[dB]$
$\gamma_k/[dB]$

Fig. 5

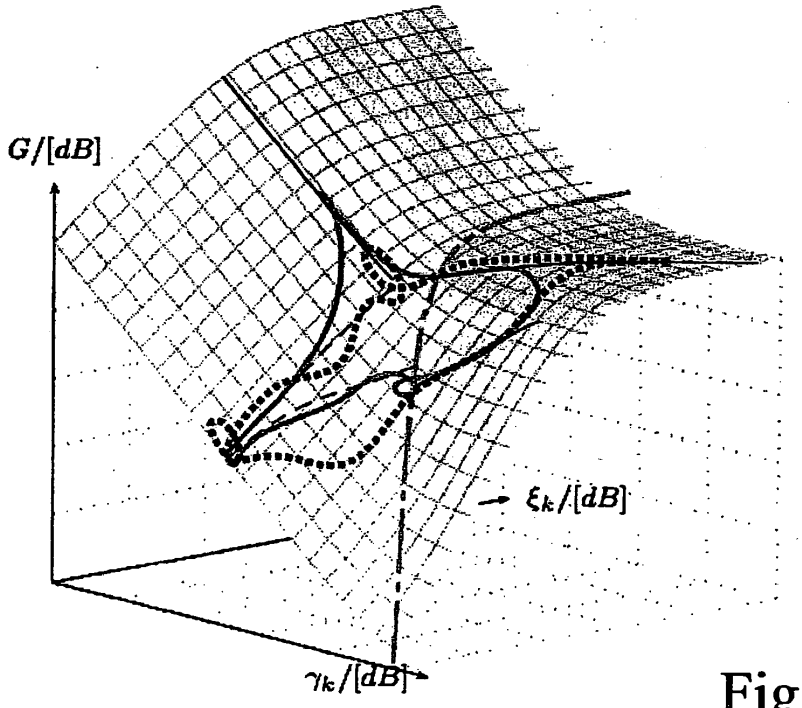

G/[dB]
$\xi_k/[dB]$
$\gamma_k/[dB]$

Fig. 6

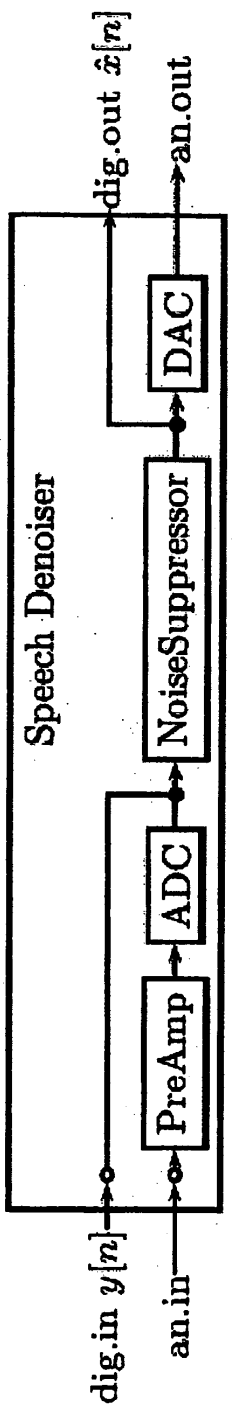

Speech Denoiser
dig.in $y[n]$
an.in
PreAmp
ADC
NoiseSuppressor
DAC
dig.out $\hat{x}[n]$
an.out

Fig. 7

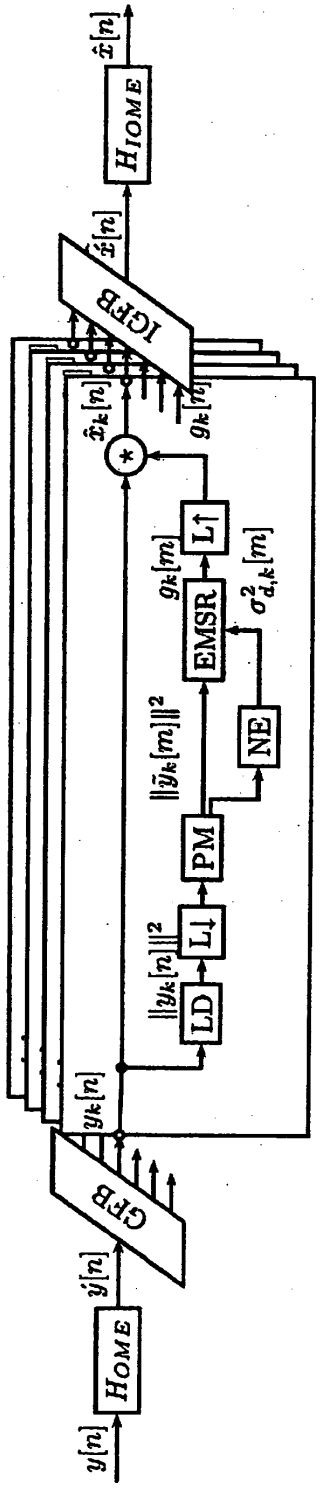

$y[n]$
H_{OME}
$y_k[n]$
GFB
$\|y_k[n]\|^2$
LD
L↓
PM
$\|\bar{y}_k[m]\|^2$
NE
$\sigma^2_{d,k}[m]$
EMSR
$g_k[m]$
L↑
$g_k[n]$
$\hat{x}_k[n]$
IGFB
$\hat{x}[n]$
H_{OME}
$\hat{x}[n]$

Fig. 8

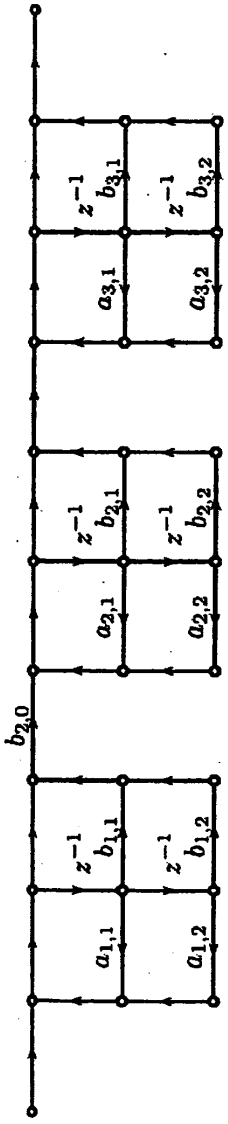

$b_{3,0}$
z^{-1} $b_{3,1}$
z^{-1} $b_{3,2}$
$a_{3,1}$
$a_{3,2}$
$b_{2,0}$
z^{-1} $b_{2,1}$
z^{-1} $b_{2,2}$
$a_{2,1}$
$a_{2,2}$
$b_{1,0}$
z^{-1} $b_{1,1}$
z^{-1} $b_{1,2}$
$a_{1,1}$
$a_{1,2}$

Fig. 9

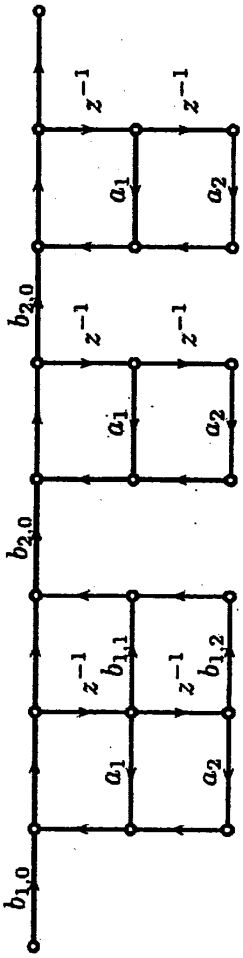

$b_{3,0}$
z^{-1}
z^{-1}
a_1
a_2
$b_{2,0}$
z^{-1}
z^{-1}
a_1
a_2
$b_{2,0}$
z^{-1} $b_{1,1}$
z^{-1} $b_{1,2}$
a_1
a_2
$b_{1,0}$

Fig. 10

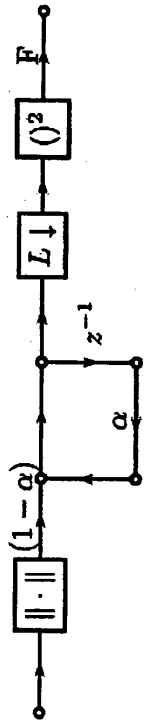

F
$()^2$
$L\downarrow$
z^{-1}
α
$(1-\alpha)$
$\|\cdot\|$

Fig. 11

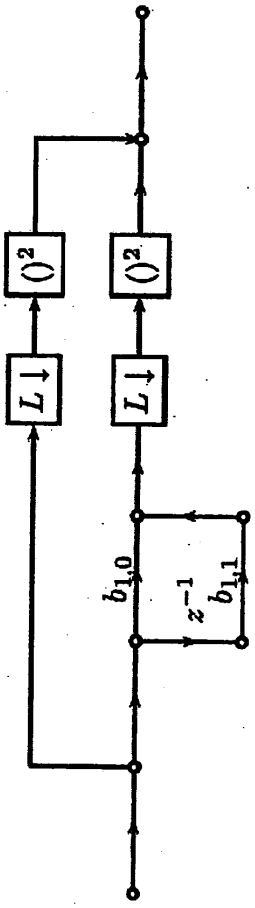

$()^2$
$()^2$
$L\downarrow$
$L\downarrow$
$b_{1,0}$
z^{-1}
$b_{1,1}$

Fig. 12

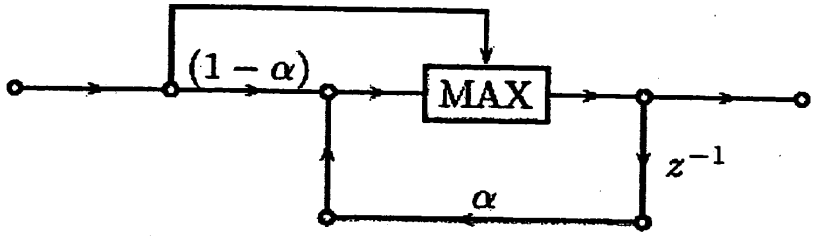


Fig. 13

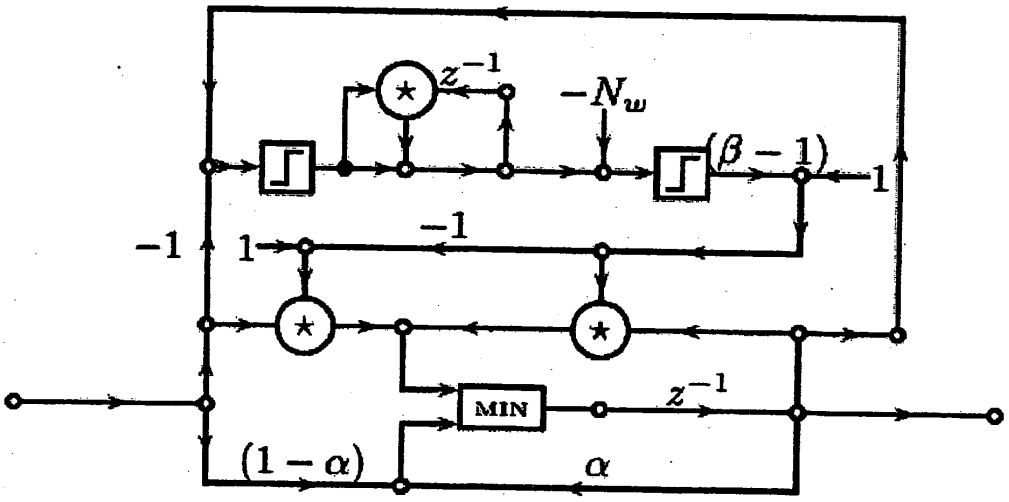


Fig. 14

Literaturverzeichnis

Abbagnaro, L. A.; Bauer, B. B. und Torick, E. L. (1975): „Measurements of diffraction and interaural delay of a progressive sound wave caused by the human head. II." In: *J Acoust Soc Am*, 58(3) S. 693–700.

Abhayapala, T. D. und Gupta, A. (2009): „Alternatives to spherical microphone arrays: Hybrid geometries." In: *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. IEEE, S. 81–84.

Abhayapala, T. D.; Kennedy, R. A. und Williamson, R. C. (2000): „Nearfield broadband array design using a radially invariant modal expansion." In: *J Acoust Soc Am*, 107(1) S. 392–403.

Abhayapala, T. D. und Ward, D. B. (2002): „Theory and design of high order sound field microphones using spherical microphone array." In: *Acoustics, Speech and Signal Processing (ICASSP), 2002 IEEE International Conference on*. S. 1949–1952.

Abramowitz, M. und Stegun, I. A. (1970): *Handbook of Mathematical Functions*. General Publishing Company.

Aertsen, A. M. H. J. und Johannesma, P. I. M. (1980): „Spectro-temporal receptive fields of auditory neurons in the grassfrog." In: *Biol. Cybern.*, 38(4) S. 223–234.

Agerkvist, F. T. (1994): *Time-Frequency Analysis and Auditory Models*. Ph.D. thesis, The Acoustics Laboratory, Technical University of Denmark.

Algazi, V. und Suk, M. (1975): „On the frequency weighted least-square design of finite duration filters." In: *IEEE Trans. Circuits Syst.*, 22(12) S. 943–953.

Algazi, V. R.; Avendano, C. und Duda, R. O. (2001): „Estimation of a Spherical-Head Model from Anthropometry." In: *J Audio Eng Soc*, 49(6) S. 472–479.

Algazi, V. R.; Duda, R. O.; Duraiswami, R.; Gumerov, N. A. und Tang, Z. (2002): „Approximating the head-related transfer function using simple geometric models of the head and torso." In: *J Acoust Soc Am*, 112(5) S. 2053–2064.

Allen, J. B. und Berkley, D. A. (1979): „Image method for efficiently simulating small-room acoustics." In: *J Acoust Soc Am*, 65(4) S. 943–950.

Allen, J. B. und Rabiner, L. R. (1977): „A unified approach to short-time Fourier analysis and synthesis." In: *Proc of the IEEE*, 65(11) S. 1558–1564.

Allen, K.; Alais, D.; Shinn-Cunningham, B. G. und Carlile, S. (2011): „Masker location uncertainty reveals evidence for suppression of maskers in two-talker contexts." In: *J Acoust Soc Am*, 130(4) S. 2043–2053.

Alon, D. L. und Rafaely, B. (2012): „Spindle-Torus Sampling for an Efficient-Scanning Spherical Microphone Array." In: *Acta Acust United Ac*, 98(1) S. 83–90.

Alt, H. W. (2012): *Lineare Funktionalanalysis - Eine anwendungsorientierte Einführung*. 6th. Berlin Heidelberg: Springer-Verlag.

ANSI (2004): „Specification for octave-band and fractional-octave-band analog and digital filters."

Antoine, J.-P. und Roşca, D. (2008): „The wavelet transform on the two-sphere and related manifolds: a review." In: Peter Schelkens; Touradj Ebrahimi; Gabriel Cristóbal und Frédéric Truchetet (Hrgs.) *Photonics Europe*. SPIE, S. 70000B–7000–15.

Applebaum, S. (1976): „Adaptive arrays." In: *Antennas and Propagation, IEEE Transactions on*, 24(5) S. 585–598.

Applebaum, S. und Chapman, D. (1976): „Adaptive arrays with main beam constraints." In: *Antennas and Propagation, IEEE Transactions on*, 24(5) S. 650–662.

Araki, S.; Sawada, H. und Makino, S. (2007): „Blind Speech Separation in a Meeting Situation with Maximum SNR Beamformers." In: *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*. IEEE, S. I–41–I–44.

Arbogast, T. L.; Mason, C. R. und Gerald Kidd, J. (2002): „The effect of spatial separation on informational and energetic masking of speech." In: *J Acoust Soc Am*, 112(5) S. 2086–2098.

Arfken, G. (1985): *Mathematical Methods for Physicists*. 3rd. Academic Press.

Arnold, S. und Burkard, R. (2000): „Studies of interaural attenuation to investigate the validity of a dichotic difference tone response recorded from the inferior colliculus in the chinchilla." In: *J Acoust Soc Am*, 107(3) S. 1541–1547.

Arslan, L.; McCree, A. und Viswanathan, V. (1995): „New methods for adaptive noise suppression." In: *Acoustics, Speech and Signal Processing (ICASSP), 1995 IEEE International Conference on*. Detroit, MI, USA, S. 812–815.

Atkinson, K. E. und Han, W. (2012): *Spherical Harmonics and Approximations on the Unit Sphere: An Introduction*. Lecture Notes in Mathematics. Berlin, Heidelberg: Springer-Verlag.

Azirani, A. A.; Le Bouquin-Jeannès, R. und Faucon, G. (1996): „Speech enhancement using a wiener filtering under signal presence uncertainty." In: *European Signal Processing Conference, 1996. EUSIPCO 1996. 8th*. IEEE, S. 1–4.

Bai, M. R.; Ih, J.-G. und Benesty, J. (2013): „Nearfield Array Signal Processing Algorithms." In: *Acoustic Array Systems*. Singapore: John Wiley & Sons Singapore Pte. Ltd., S. 151–209.

Balazs, P.; Dörfler, M.; Jaillet, F.; Holighaus, N. und Velasco, G. A. (2011): „Theory, implementation and applications of nonstationary Gabor frames." In: *Journal of Computational and Applied Mathematics*, 236(6) S. 1481–1496.

Balmages, I. und Rafaely, B. (2007): „Open-Sphere Designs for Spherical Microphone Arrays." In: *IEEE Trans. Audio Speech Lang. Process.*, 15(2) S. 727–732.

Barndorff-Nielsen, O. (1977): „Exponentially Decreasing Distributions for the Logarithm of Particle Size." In: *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 353(1674) S. 401–419.

Barnett, A. R. (1996): „The Calculation of Spherical Bessel and Coulomb Functions." In: *Computational Atomic Physics*. Berlin, Heidelberg: Springer, Berlin, Heidelberg, S. 181–202.

Barnett, A. H. (2000): *Dissipation in Deforming Chaotic Billiards*. Ph.D. thesis, Harvard University.

Bartlett, M. S. (1948): „Smoothing Periodograms from Time-Series with Continuous Spectra." In: *Nature*, 161(4096) S. 686–687.

Békésy, G. (1928): „Zur Theorie des Hörens. Die Schwingungsform der Basilarmembran." In: *Physik Z*, 29 S. 793–810.

Békésy, G. (1964): „Concerning the pleasures of observing, and the mechanics of the inner ear (Nobel Lecture, December 11, 1961)." In: *Nobel Lectures Physiology or Medicine 1942 - 1962*. Elsevier, S. 722–746.

Békésy, G. und Wever, E. G. (1960): *Experiments in hearing*. McGraw-Hill series in psychology. New York: McGraw-Hill.

Bendat, J. S. und Piersol, A. G. (1966): „Measurements and analysis of random data." John Wiley and Sons Ltd.

Bendat, J. S. und Piersol, A. G. (1993): *Engineering applications of correlation and spectral analysis*. 1st. John Wiley and Sons Ltd.

Benesty, J.; Chen, J. und Huang, Y. A. (2004): „Time-delay estimation via linear interpolation and cross correlation." In: *IEEE Transactions on Speech and Audio Processing*, 12(5) S. 509–519.

Benesty, J.; Chen, J. und Huang, Y. A. (2008a): *Microphone Array Signal Processing*. Springer-Verlag, Berlin, New York, Heidelberg.

Benesty, J.; Chen, J.; Huang, Y. A. und Cohen, I. (2009): *Noise Reduction in Speech Processing*. Springer-Verlag, Berlin, New York, Heidelberg.

Benesty, J.; Chen, J.; Huang, Y. A. und Dmochowski, J. (2007): „On Microphone-Array Beamforming From a MIMO Acoustic Signal Processing Perspective." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(3) S. 1053–1065.

Benesty, J. und Huang, Y. A. (2003): *Adaptive Signal Processing*. Applications to Real-World Problems. Springer Verlag.

Benesty, J.; Sondhi, M. M. und Huang, Y. A. (2008b): *Springer Handbook of Speech Processing*. Springer-Verlag, Berlin, New York, Heidelberg.

Bernschütz, B.; Pörschmann, C.; Spors, S. und Weinzierl, S. (2011): „Soft-Limiting der modalen Amplitudenverstärkung bei sphärischen Mikrofonarrays im Plane Wave Decomposition Verfahren." In: *37th Annual Convention of Acoustics (DAGA)*. Düsseldorf, Germany, S. 661–662.

Berouti, M.; Schwartz, R. und Makhoul, J. (1979): „Enhancement of speech corrupted by acoustic noise." In: *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '79*. Washington, District of Columbia, USA, S. 208–211.

Best, V.; Thompson, E. R.; Mason, C. R. und Gerald Kidd, J. (2013): „Spatial release from masking as a function of the spectral overlap of competing talkers." In: *J Acoust Soc Am*, 133(6) S. 3677–3680.

Betcke, T. und Trefethen, L. N. (2006): „Reviving the Method of Particular Solutions." In: *SIAM Review*, 47(3) S. 469–491.

Bitzer, J.; Kammeyer, K.-D. und Simmer, K. U. (1999a): „An alternative implementation of the superdirective beamformer." In: *Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on*. IEEE, S. 7–10.

Bitzer, J. und Simmer, K. U. (2001): „Superdirective Microphone Arrays." In: *Microphone Arrays*. Springer Verlag, S. 19–38.

Bitzer, J.; Simmer, K. U.; Holube, I. und Schaer, T. (2005): „Some experiments on short-time spectral attenuation (STSA) algorithms and speech intelligibility." In: *International Workshop on Acoustic Echo and Noise Control (IWAENC).* Eindhoven, The Netherlands, S. 193–196.

Bitzer, J.; Simmer, K. U. und Kammeyer, K.-D. (1999b): „Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement." In: *Acoustics, Speech, and Signal Processing, 1999. ICASSP '99. Proceedings., 1999 IEEE International Conference on.* Phoenix, AZ, USA, S. 2965–2968.

Bitzer, J.; Simmer, K. U. und Kammeyer, K.-D. (2001): „Multi-microphone noise reduction techniques as front-end devices for speech recognition." In: *Speech Commun.* Houpert Digital Audio, D-28359 Bremen, Germany, S. 3–12.

Blauert, J. (1974): *Räumliches Hören.* Stuttgart: S. Hirzel Verlag.

Blauert, J. (1996): *Spatial Hearing.* The Psychophysics of Human Sound Localization. The MIT Press.

Bolcskei, H.; Hlawatsch, F. und Feichtinger, H. G. (1998): „Frame-theoretic analysis of oversampled filter banks." In: *Signal Processing, IEEE Transactions on*, 46(12) S. 3256–3268.

Boll, S. (1979): „Suppression of acoustic noise in speech using spectral subtraction." In: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 27(2) S. 113–120.

Boyd, J. (2000): *Chebyshev and Fourier Spectral Methods.* DOVER Publications, New York.

Brandstein, M. S. (1995): *A Framework for Speech Source Localization Using Sensor Arrays.* Ph.D. thesis, Brown University, Brown University.

Brandstein, M. S. und Ward, D. B. (2001): *Microphone Arrays.* Signal Processing Techniques and Applications. Springer Verlag.

Breining, C. et al. (1999): „Acoustic echo control. An application of very-high-order adaptive filters." In: *Signal Processing Magazine, IEEE*, 16(4) S. 42–69.

Brinkmann, K.; Vorländer, M. und Fedtke, T. (1994): „Re-Determination Of The Threshold Of Hearing Under Free-Field And Diffuse-Field Listening Conditions." In: *Acta Acust United Ac*, 80(5) S. 453–462.

Bronstein, I. N.; Semendjajew, K. A.; Grosche, G.; Ziegler, V. und Ziegler, D. (2013): *Springer-Taschenbuch der Mathematik*. Springer Fachmedien Wiesbaden.

Brooks, L. W. und Reed, I. (1972): „Equivalence of the Likelihood Ratio Processor, the Maximum Signal-to-Noise Ratio Filter, and the Wiener Filter." In: *Aerospace and Electronic Systems, IEEE Transactions on*, 8(5) S. 690–692.

Brown, J. C. (1991): „Calculation of a constant Q spectral transform." In: *J Acoust Soc Am*, 89(1) S. 425–434.

Brown, J. C. und Puckette, M. S. (1992): „An efficient algorithm for the calculation of a constant Q transform." In: *J Acoust Soc Am*, 92 S. 2698.

Brungart, D. S. und Rabinowitz, W. (1999): „Auditory localization of nearby sources. Head-related transfer functions." In: *J Acoust Soc Am*, 106(3) S. 1465–1479.

Brungart, D. S. und Simpson, B. D. (2002): „The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal." In: *J Acoust Soc Am*, 112(2) S. 664–676.

Buckley, K. M. (1987): „Spatial/Spectral filtering with linearly constrained minimum variance beamformers." In: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 35(3) S. 249–266.

Buckley, K. M. und Griffiths, L. J. (1986): „An adaptive generalized sidelobe canceller with derivative constraints." In: *Antennas and Propagation, IEEE Transactions on*, 34(3) S. 311–319.

Capon, J. (1969): „High-resolution frequency-wavenumber spectrum analysis." In: *Proceedings of the IEEE*. S. 1408–1418.

Capon, J.; Greenfield, R. und Kolker, R. (1967): „Multidimensional maximum-likelihood processing of a large aperture seismic array." In: *Proceedings of the IEEE*. S. 192–211.

Cappé, O. (1994): „Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor." In: *IEEE Transactions on Speech and Audio Processing*, 2(2) S. 345–349.

Carlyon, R. P. und Shamma, S. (2003): „An account of monaural phase sensitivity." In: *J Acoust Soc Am*, 114(1) S. 333–348.

Carney, L. H. und Yin, T. C. (1988): „Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model." In: *Journal of Neurophysiology*, 60(5) S. 1653–1677.

Carney, L. H. (1993): „A model for the responses of low-frequency auditory-nerve fibers in cat." In: *J Acoust Soc Am*, 93(1) S. 401–417.

Carney, L. H.; McDuffy, M. J. und Shekhter, I. (1999): „Frequency glides in the impulse responses of auditory-nerve fibers." In: *J Acoust Soc Am*, 105(4) S. 2384–2391.

Carter, G. C. (1987): „Coherence and time delay estimation." In: *Proceedings of the IEEE*. S. 236–255.

Carter, G. C.; Knapp, C. und Nuttall, A. H. (1973): „Statistics of the estimate of the magnitute-coherence function." In: *Audio and Electroacoustics, IEEE Transactions on*, 21(4) S. 388–389.

Chambodut, A. et al. (2005): „Wavelet frames: an alternative to spherical harmonic representation of potential fields." In: *Geophysical Journal International*, 163(3) S. 875–899.

Chardon, G.; Cohen, A. und Daudet, L. (2013): „Reconstruction of solutions to the Helmholtz equation from punctual measurements." In: *10th International Conference on Sampling Theory and Applications*. Bremen, Germany, S. 164–167.

Chardon, G.; Cohen, A. und Daudet, L. (2014a): „Sampling and reconstruction of solutions to the Helmholtz equation." In: *Sampl. Theory Signal Image Process.*, 13(1) S. 67–89.

Chardon, G.; Kreuzer, W. und Noisternig, M. (2014b): „Design of a robust open spherical microphone array." In: *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. S. 6810–6814.

Chardon, G.; Kreuzer, W. und Noisternig, M. (2015): „Design of Spatial Microphone Arrays for Sound Field Interpolation." In: *Selected Topics in Signal Processing, IEEE Journal of*, 9(5) S. 780–790.

Chen, H. und Ser, W. (2009): „Design of Robust Broadband Beamformers With Passband Shaping Characteristics Using Tikhonov Regularization." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 17(4) S. 665–681.

Chen, J.; Benesty, J. und Huang, Y. A. (2003a): „Robust time delay estimation exploiting redundancy among multiple microphones." In: *IEEE Transactions on Speech and Audio Processing*, 11(6) S. 549–557.

Chen, J.; Benesty, J. und Huang, Y. A. (2003b): „Robust time delay estimation exploiting spatial correlation." In: *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*. IEEE, S. V–481–4 vol.5.

Chen, J.; Benesty, J.; Huang, Y. A. und Doclo, S. (2006): „New insights into the noise reduction Wiener filter." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(4) S. 1218–1234.

Chen, J.; Huang, N. und Benesty, J. (2004): „An adaptive blind SIMO identification approach to joint multichannel time delay estimation." In: *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*. IEEE, S. –iv–56.

Chen, J.; Huang, Y. A. und Benesty, J. (2005): „Time delay estimation via multichannel cross-correlation [audio signal processing applications]." In: *2001*

IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221). IEEE, S. iii/49–iii/52 Vol. 3.

Chi, T.; Ru, P. und Shamma, S. A. (2004): „Multiresolution spectrotemporal analysis of complex sounds." In: *J Acoust Soc Am*, 118(2) S. 887–906.

Christensen, O. (2003): *An Introduction to Frames and Riesz Bases*. Birkhäuser.

Christensen, O. (2008): *Frames and Bases - An Introductory Course*. Applied and Numerical Harmonic Analysis. Birkhäuser.

Christensen, O. (2010): *Functions, Spaces, and Expansions. Mathematical Tools in Physics and Engineering.* Applied and Numerical Harmonic Analysis. Springer New York Dordrecht Heidelberg London.

Claesson, I. und Nordholm, S. (1992): „A spatial filtering approach to robust adaptive beaming." In: *IEEE Trans. Antennas Propagat.*, 40(9) S. 1093–1096.

Cohen, A.; Davenport, M. und Leviatan, D. (2013): „On the Stability and Accuracy of Least Squares Approximations." In: *Found Comput Math*, 13(5) S. 819–834.

Cohen, I. (2001): „On speech enhancement under signal presence uncertainty." In: *Acoustics, Speech and Signal Processing (ICASSP), 2001 IEEE International Conference on*. S. 661–664.

Cohen, I. (2002): „Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator." In: , 9(4) S. 113–116.

Cohen, I. (2003): „Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging." In: *IEEE Transactions on Speech and Audio Processing*, 11(5) S. 466–475.

Cohen, I. (2004a): „Modeling speech signals in the time–frequency domain using GARCH." In: *Signal Process*, 84(12) S. 2453–2459.

Cohen, I. (2004b): „On the decision-directed estimation approach of Ephraim and Malah." In: *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*. Montreal, Quebec, Canada: IEEE, S. 293–296.

Cohen, I. (2004c): „Speech enhancement using a noncausal a priori SNR estimator." In: , 11(9) S. 725–728.

Cohen, I. (2005): „Relaxed Statistical Model for Speech Enhancement and a Priori SNR Estimation." In: *IEEE Transactions on Speech and Audio Processing*, 13(5) S. 870–881.

Cohen, I. (2006): „Speech spectral modeling and enhancement based on autoregressive conditional heteroscedasticity models." In: *Signal Process*, 86(4) S. 698–709.

Cohen, I. und Berdugo, B. (2001a): „Spectral enhancement by tracking speech presence probability in subbands." In: *IEEE Workshop on Hands Free Speech Communication HSC*. Kyoto, Japan, S. 95–98.

Cohen, I. und Berdugo, B. (2001b): „Speech enhancement for non-stationary noise environments." In: *Signal Process*, 81(11) S. 2403–2418.

Colomes, C.; Lever, M.; Rault, J. B. und Dehery, Y. F. (1995): „*A Perceptual Model Applied to Audio Bit-Rate Reduction*." In: *J Audio Eng Soc*, 43 S. 1–8.

Colton, D. und Kress, R. (2013): *Inverse Acoustic and Electromagnetic Scattering Theory*, vol. 93 von *Applied Mathematical Sciences*. 3rd ed;. Springer New York.

Cooke, M. (1991): *Modelling Auditory Processing and Organisation*. Ph.D. thesis, University of Sheffield.

Cooke, M. (2006): „A glimpsing model of speech perception in noise." In: *J Acoust Soc Am*, 119(3) S. 1562–1573.

Cooke, M.; Beet, S. und Crawford, M. (1993): *Visual Representation of Speech Signals*. John Wiley & Sons.

Cornelis, B.; Moonen, M. und Wouters, J. (2011): „Performance Analysis of Multichannel Wiener Filter-Based Noise Reduction in Hearing Aids Under Second Order Statistics Estimation Errors." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 19(5) S. 1368–1381.

Cox, H. (1973a): „Resolving power and sensitivity to mismatch of optimum array processors." In: *J Acoust Soc Am*, 54(3) S. 771–785.

Cox, H. (1973b): „Spatial correlation in arbitrary noise fields with application to ambient sea noise." In: *J Acoust Soc Am*, 54(5) S. 1289–1301.

Cox, H.; Zeskind, R. und Kooij, T. (1986): „Practical supergain." In: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 34(3) S. 393–398.

Cox, H.; Zeskind, R. und Owen, M. (1987): „Robust adaptive beamforming." In: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 35(10) S. 1365–1376.

Cremer, L. und Müller, H. A. (1978): *Die wissenschaftlichen Grundlagen der Raumakustik. Band I: Geometrische Akustik*. Stuttgart: S. Hirzel Verlag.

Dai, F. und Xu, Y. (2013): *Approximation Theory and Harmonic Analysis on Spheres and Balls*. Springer Monographs in Mathematics. Springer New York Heidelberg Dordrecht London.

Dallos, P. (1992): „The active Cochlea." In: *The Journal of Neuroscience*, 12(12) S. 4575–4585.

Daniel, H. (1997): *Elektrodynamik - relativistische Physik*. Berlin, Germany: Wlater de Gruyter.

Daniel, J. (2001): *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. Ph.D. thesis, Université de Paris 6.

Dau, T.; Püschel, D. und Kohlrausch, A. (1996a): „A quantitative model of the ''effective'' signal processing in the auditory system. I. Model structure." In: *J Acoust Soc Am*, 99(6) S. 3615–3622.

Dau, T.; Püschel, D. und Kohlrausch, A. (1996b): „A quantitative model of the ''effective'' signal processing in the auditory system. II. Simulations and measurements." In: *J Acoust Soc Am*, 99(6) S. 3623–3631.

de Boer, E. (1975): „Synthetic whole-nerve action potentials for the cat." In: *J Acoust Soc Am*, 58(5) S. 1030–1045.

de Boer, E. (1996): „Mechanics of the Cochlea: Modeling Efforts." In: Peter Dallos; Arthur N Popper und Richard R Fay (Hrgs.) *Springer Handbook of Auditory Research*. Springer New York, S. 258–317.

de Boer, E. und de Jongh, H. R. (1978): „On cochlear encoding: Potentialities and limitations of the reverse-correlation technique." In: *J Acoust Soc Am*, 63(1) S. 115–135.

de Boer, E. und Kuyper, P. (1968): „Triggered Correlation." In: *Biomedical Engineering, IEEE Transactions on*, BME-15(3) S. 169–179.

de Boer, E. und Nuttall, A. L. (1997): „The mechanical waveform of the basilar membrane. I. Frequency modulations ("glides") in impulse responses and cross-correlation functions." In: *J Acoust Soc Am*, 101(6) S. 3583–3592.

Di Claudio, E. D. und Parisi, R. (2001): „Multi-Source Localization Strategies." In: Michael S Brandstein und Darren B Ward (Hrgs.) *Microphone Arrays*. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 181–201.

DiBiase, J. H.; Silvemnan, H. F. und Brandstein, M. S. (2001): „Robust Localization in Reverberant Rooms." In: *Microphone Arrays*. Springer Verlag, S. 157–180.

Diependaal, R. J.; Duifhuis, H.; Hoogstraten, H. W. und Viergever, M. A. (1987): „Numerical methods for solving one-dimensional cochlear models in the time domain." In: *J Acoust Soc Am*, 82(5) S. 1655–1666.

Diependaal, R. J. und Viergever, M. A. (1983): „Nonlinear and Active Modelling of Cochlear Mechanics: A Precarious Affair." In: Egbert de Boer und Max A Viergever (Hrgs.) *Mechanics of Hearing*. Springer Netherlands, S. 153–160.

Do, H.; Silverman, H. F. und Yu, Y. (2007): „A Real-Time SRP-PHAT Source Location Implementation using Stochastic Region Contraction(SRC) on a Large-Aperture Microphone Array." In: *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*. IEEE, S. I–121–I–124.

Doblinger, G. (1995): „Computationally efficient speech enhancement by spectral minima tracking in subbands." In: *4th European Conference on Speech Communication and Technology (EUROSPEECH)*. Madrid, Spain, S. 1513–1516.

Doclo, S. und Moonen, M. (2002): „Design of far-field broadband beamformers using eigenfilters." In: *2002 11th European Signal Processing Conference*. IEEE, S. 1–4.

Doclo, S. und Moonen, M. (2003a): „Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics." In: *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, 51(10) S. 2511–2526.

Doclo, S. und Moonen, M. (2003b): „Design of far-field and near-field broadband beamformers using eigenfilters." In: *Signal Process*, 83(12) S. 2641–2673.

Doclo, S. und Moonen, M. (2007): „Superdirective Beamforming Robust Against Microphone Mismatch." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(2) S. 617–631.

Driscoll, J. und Healy, D. (1994): „Computing Fourier Transforms and Convolutions on the 2-Sphere." In: *Adv Appl Math*, 15(2) S. 202–250.

Dudgeon, D. (1977): „Fundamentals of digital array processing." In: *Proc of the IEEE*, 65(6) S. 898–904.

Duffy, D. (2015): *Green's Functions with Applications*. 2nd. Chapman and Hall/CRC.

Duifhuis, H. (2012): *Cochlear Mechanics, Introduction to a Time Domain Analysis of the Nonlinear Cochlea*. Springer New York Dordrecht Heidelberg London.

Ealey, D.; Kelleher, H. und Pearce, D. (2001): „Harmonic tunnelling: tracking non-stationary noises during speech." In: *7th European Conference on Speech Communication and Technology (EUROSPEECH 2001)*. Aalborg, Denmark, S. 437–440.

Ebata, M. (2003): „Spatial unmasking and attention related to the cocktail party problem." In: *Acoustical Science and Technology*, 24(5) S. 208–219.

Edelblute, D. J.; Fisk, J. M. und Kinnison, G. L. (1967): „Criteria for Optimum-Signal-Detection Theory for Arrays." In: *J Acoust Soc Am*, 41(1) S. 199–205.

Ephraim, Y. (1992a): „A Bayesian estimation approach for speech enhancement using hidden Markov models." In: *IEEE Trans. Signal Process.*, 40(4) S. 725–735.

Ephraim, Y. (1992b): „Gain-adapted hidden Markov models for recognition of clean and noisy speech." In: *IEEE Trans. Signal Process.*, 40(6) S. 1303–1316.

Ephraim, Y. und Cohen, I. (2006): „Recent Advancements in Speech Enhancement." In: Richard C. Dorf (Hrg.) *The Electrical Engineering Handbook*. CRC Press, S. 12–26.

Ephraim, Y.; Lev-Ari, H. und Roberts, W. (2005): „A brief survey of speech enhancement." In: Jerry C. Whitaker (Hrg.) *The Electronic Handbook*. Boca Raton: CRC Press, S. 2088–2096.

Ephraim, Y. und Malah, D. G. (1983): „Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator." In: *Acoustics, Speech and Signal Processing (ICASSP), 1983 IEEE International Conference on*. Boston, Massachusetts, USA: Institute of Electrical and Electronics Engineers, S. 1118–1121.

Ephraim, Y. und Malah, D. G. (1984): „Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator." In: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 32(6) S. 1109–1121.

Ephraim, Y. und Malah, D. G. (1985): „Speech enhancement using a minimum mean-square error log-spectral amplitude estimator." In: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 33(2) S. 443–445.

Ephraim, Y.; Malah, D. G. und Juang, B. H. (1989): „On the application of hidden Markov models for enhancing noisy speech." In: *IEEE Trans. Acoust., Speech, Signal Processing*, 37(12) S. 1846–1856.

Ephraim, Y. und Roberts, W. (2005): „Revisiting autoregressive hidden Markov modeling of speech signals." In: *IEEE Signal Processing Letters*, 12(2) S. 166–169.

Er, M. und Cantoni, A. (1986): „An unconstrained partitioned realization for derivative constrained broad-band antenna array processors." In: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 34(6) S. 1376–1379.

Eriksson, A.; Ghysels, E. und Wang, F. (2009): „The Normal Inverse Gaussian Distribution and the Pricing of Derivatives." In: *Derivatives*, 16(3) S. 23–37.

Evans, E. F. (1986): „Cochlear Nerve Fibre Temporal Discharge Patterns, Cochlear Frequency Selectivity and the Dominant Region for Pitch." In: Brian C. J. Moore und Roy D Patterson (Hrgs.) *Nato ASI Series*. Springer US, S. 253–264.

Evans, N. W. D.; Mason, J. S. und Fauve, B. (2002): „Efficient real-time noise estimation without explicit speech, non-speech detection: an assessment on the AURORA corpus." In: *ICDSP-02*. IEEE, S. 985–988.

Evans, N. W. D. und Mason, J. (2002): „Time-frequency quantile-based noise estimation." In: *European Signal Processing Conference (EUSIPCO'02)*. Toulouse, France, S. 539–542.

Farmer-Fedor, B. L. und Rabbitt, R. D. (2002): „Acoustic intensity, impedance and reflection coefficient in the human ear canal." In: *J Acoust Soc Am*, 112(2) S. 600–620.

Fastl, H. und Zwicker, E. (2007): *Psychoacoustics, Facts and Models*. 3rd. Springer-Verlag, Berlin, New York, Heidelberg.

Fazi, F. M. (2010): *Sound Field Reproduction*. Ph.D. thesis, University of Southampton, Southampton, UK.

Fazi, F. M.; Noisternig, M. und Warusfel, O. (2012): „Representation of sound fields for audio recording and reproduction." In: *11ème Congrès Français d'Acoustique and Annual IOA Meeting (Acoustics 2012)*. Nantes, France.

Feldbauer, C.; Kubin, G. und Kleijn, B. W. (2005): „Anthropomorphic Coding of Speech and Audio: A Model Inversion Approach." In: *EURASIP J. Adv. Signal Process.*, 2005(9) S. 1334–1349.

Fischer, S. und Simmer, K. U. (1996): „Beamforming microphone arrays for speech acquisition in noisy environments." In: *Speech Commun*, 20(3-4) S. 215–227.

Fisher, E. und Rafaely, B. (2008): „The Nearfield Spherical Microphone Array." In: *IEEE International Conference on Acoustics, Speech and Signal Processing. In Acoustics, Speech, and Signal Processing (ICASSP), 2008.* IEEE, S. 5272–5275.

Fisher, E. und Rafaely, B. (2011): „Near-Field Spherical Microphone Array Processing With Radial Filtering." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 19(2) S. 256–265.

Flanagan, J. L. (1960a): „Models for Approximating Basilar Membrane Displacement." In: *J Acoust Soc Am*, 32(7) S. 937–937.

Flanagan, J. L. (1960b): „Models for approximating basilar membrane displacement." In: *The Bell System Technical Journal* S. 1163–1192.

Flanagan, J. L.; Berkley, D. A.; Elko, G. W.; West, J. E. und Sondhi, M. M. (1991): „Autodirective Microphone Systems." In: *Acta Acust United Ac*, 73(2) S. 58–71.

Flanagan, J. L.; Johnston, J. D.; Zahn, R. und Elko, G. W. (1985): „Computer-steered microphone arrays for sound transduction in large rooms." In: *J Acoust Soc Am*, 78(5) S. 1508–1518.

Fletcher, H. (1940): „Auditory Patterns." In: *Rev. Mod. Phys.*, 12 S. 47–65.

Fletcher, H. und Munson, W. A. (1933): „Loudness, Its Definition, Measurement and Calculation." In: *J Acoust Soc Am*, 5(2) S. 82–108.

Fliege, J. (1999): „The distribution of points on the sphere and corresponding cubature formulae." In: *IMA Journal of Numerical Analysis*, 19(2) S. 317–334.

Forbes, C.; Evans, M.; Hastings, N. und Peacock, B. (2011): *Statistical Distributions.* 4th ed. Hoboken, New Jersey, USA: John Wiley & Sons, Ltd.

Fornberg, B. und Martel, J. M. (2014): „On spherical harmonics based numerical quadrature over the surface of a sphere." In: *Adv Comput Math*, 40(5-6) S. 1169–1184.

Freeden, W. und Schreiner, M. (2009): *Spherical Functions of Mathematical Geosciences*. A Scalar, Vectorial, and Tensorial Setup. Springer-Verlag.

Freyman, R. L.; Balakrishnan, U. und Helfer, K. S. (2001): „Spatial release from informational masking in speech recognition." In: *J Acoust Soc Am*, 109(5) S. 2112–2122.

Friedlander, B. und Porat, B. (1984): „The Modified Yule-Walker Method of ARMA Spectral Estimation." In: *Aerospace and Electronic Systems, IEEE Transactions on*, AES-20(2) S. 158–173.

Friedlander, B. und Sharman, K. (1985): „Performance evaluation of the modified Yule-Walker estimator." In: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 33(3) S. 719–725.

Frost, O. L. (1972): „An algorithm for linearly constrained adaptive array processing." In: *Proc of the IEEE*, 60(8) S. 926–935.

Fudge, G. L. und Linebarger, D. A. (1995): „Steered response control of the generalized sidelobe canceller." In: *ICASSP-95*. IEEE, S. 3623–3626.

Gabor, D. (1946a): „Theory of communication. Part 1: The analysis of information." In: *Electrical Engineers - Part III: Radio and Communication Engineering, Journal of the Institution of*, 93(26) S. 429–441.

Gabor, D. (1946b): „Theory of communication. Part 2: The analysis of hearing." In: *Electrical Engineers - Part III: Radio and Communication Engineering, Journal of the Institution of*, 93(26) S. 442–445.

Gabor, D. (1946c): „Theory of communication. Part 3: Frequency compression and expansion." In: *Electrical Engineers - Part III: Radio and Communication Engineering, Journal of the Institution of*, 93(26) S. 445–457.

Gannot, S. (2012): „Speech processing utilizing the Kalman filter." In: *IEEE Instrumentation & Measurement Magazine*, 15(3) S. 10–14.

Gannot, S.; Burshtein, D. und Weinstein, E. (1998): „Iterative and sequential Kalman filter-based speech enhancement algorithms." In: *IEEE Transactions on Speech and Audio Processing*, 6(4) S. 373–385.

Gardner, W. A. (1992): „A Unifying View of Coherence in Signal-Processing." In: *Signal Process*, 29(2) S. 113–140.

Gazzaniga, M. S. (Hrg.) (2004): *The Cognitive Neurosciences*. 3rd ed. MIT Press.

Gerkmann, T. und Hendriks, R. C. (2012): „Unbiased MMSE-Based Noise Power Estimation With Low Complexity and Low Tracking Delay." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 20(4) S. 1383–1393.

Gerkmann, T. und Krawczyk, M. (2013): „MMSE-Optimal Spectral Amplitude Estimation Given the STFT-Phase." In: *IEEE Signal Processing Letters*, 20(2) S. 129–132.

Gibson, J. D.; Koo, B. und Gray, S. D. (1991): „Filtering of colored noise for speech enhancement and coding." In: *IEEE Trans. Signal Process.*, 39(8) S. 1732–1742.

Giguere, C. und Woodland, P. C. (1994a): „A computational model of the auditory periphery for speech and hearing research. I. Ascending path." In: *J Acoust Soc Am*, 95(1) S. 331–342.

Giguere, C. und Woodland, P. C. (1994b): „A computational model of the auditory periphery for speech and hearing research. II. Descending paths." In: *J Acoust Soc Am*, 95(1) S. 343–349.

Givelberg, E. und Bunn, J. (2003): „A comprehensive three-dimensional model of the cochlea." In: *J Comput Phys*, 191(2) S. 377–391.

Glasberg, B. R. und Moore, B. C. J. (1986): „Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments." In: *J Acoust Soc Am*, 79(4) S. 1020–1033.

Glasberg, B. R. und Moore, B. C. J. (1990): „Derivation of Auditory Filter Shapes From Notched-Noise Data." In: *Hearing Res*, 47(1-2) S. 103–138.

Glasberg, B. R. und Moore, B. C. J. (2002): „A Model of Loudness Applicable to Time-Varying Sounds." In: *J Audio Eng Soc*, 50(5) S. 331–342.

Glasberg, B. R.; Moore, B. C. J.; Patterson, R. D. und Nimmo-Smith, I. (1984): „Dynamic range and asymmetry of the auditory filter." In: *J Acoust Soc Am*, 76(2) S. 419–427.

Glyde, H.; Buchholz, J. M.; Dillon, H.; Cameron, S. und Hickson, L. (2013): „The importance of interaural time differences and level differences in spatial release from masking." In: *J Acoust Soc Am*, 134(2) S. EL147–EL152.

Gockel, H.; Moore, B. C. J. und Patterson, R. D. (2002): „Asymmetry of masking between complex tones and noise: The role of temporal structure and peripheral compression." In: *J Acoust Soc Am*, 111(6) S. 2759–2770.

Golub, G. und Kahan, W. (2006): „Calculating the Singular Values and Pseudo-Inverse of a Matrix." In: *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, 2(2) S. 205–224.

Golub, G. H. und Van Loan, C. F. (1996): *Matrix Computations*. 3. Baltimore and London: The Johns Hopkins University Press.

Goodman, J. W. (1968): *Introduction to Fourier Optics*. San Francisco, CA: McGraw-Hill Book Company.

Gorski, K. M. et al. (2005): „HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere." In: *The Astrophysical Journal*, 622(2) S. 759–771.

Gover, B. N.; Ryan, J. G. und Stinson, M. R. (2004): „Measurements of directional properties of reverberant sound fields in rooms using a spherical microphone array." In: *J Acoust Soc Am*, 116(4) S. 2138–2148.

Greenwood, D. D. (1961): „Critical Bandwidth and the Frequency Coordinates of the Basilar Membrane." In: *J Acoust Soc Am*, 33(10) S. 1344–1356.

Greenwood, D. D. (1990): „A cochlear frequency-position function for several species—29 years later." In: *J Acoust Soc Am*, 87(6) S. 2592–2605.

Griffiths, L. J. (1977): „An Adaptive Beamformer which Implements Constraints Using an Auxiliary Array Preprocessor." In: *Aspects of Signal Processing With Emphasis on Underwater Acoustics, Part 2*. Dordrecht: Springer Netherlands, S. 517–522.

Griffiths, L. J. und Jim, C. W. (1982): „An alternative approach to linearly constrained adaptive beamforming." In: *Antennas and Propagation, IEEE Transactions on*, 30(1) S. 27–34.

Gumerov, N. A.; O'Donovan, A. E.; Duraiswami, R. und Zotkin, D. N. (2010): „Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation." In: *J Acoust Soc Am*, 127(1) S. 370–386.

Gupta, A. und Abhayapala, T. D. (2010): „Double sided cone array for spherical harmonic analysis of wavefields." In: *Acoustics, Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. S. 77–80.

Gustafsson, S.; Jax, P. und Vary, P. (1998): „A novel psychoacoustically motivated audio enhancement algorithm preserving background noise characteristics." In: *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*. Seattle, WA, USA, S. 397–400.

Gustafsson, S.; Martin, R.; Jax, P. und Vary, P. (2002): „A psychoacoustic approach to combined acoustic echo cancellation and noise reduction." In: *IEEE Transactions on Speech and Audio Processing*, 10(5) S. 245–256.

Hamacher, V. et al. (2005): „Signal Processing in High-End Hearing Aids: State of the Art, Challenges, and Future Trends." In: *EURASIP J. Adv. Signal Process.*, 2005(18) S. 2915–2929.

Hammershøi, D. und Møller, H. (1996): „Sound transmission to and within the human ear canal." In: *J Acoust Soc Am*, 100(1) S. 408–427.

Hansen, P. C. (1998): *Rank-Deficient and Discrete Ill-Posed Problems.* SIAM Monographs,Ä®on Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics.

Hänsler, E. (1983): *Grundlagen der Theorie statistischer Signale.* Berlin: Springer Verlag.

Hänsler, E. (1997): *Statistische Signal.* Berlin: Springer Verlag.

Hänsler, E. und Schmidt, G. (2004): *Acoustic Echo and Noise Control, A Practical Approach.* John Wiley & Sons, Ltd.

Hänsler, E. und Schmidt, G. (2008): *Speech and Audio Processing in Adverse Environments (Signals and Communication Technology).* Springer-Verlag, Berlin, New York, Heidelberg.

Hasan, M. K.; Salahuddin, S. und Khan, M. R. (2004): „A modified a priori SNR for speech enhancement using spectral subtraction rules." In: , 11(4) S. 450–453.

Havelock, D. I.; Kuwano, S. und Vorländer, M. (Hrgs.) (2008): *Handbook of Signal Processing in Acoustics*, vol. 1. Springer.

Hawkins, H. L. (1995): *Auditory Computation.* Springer.

Hayes, M. H. (1996): *Statistical Digital Signal Processing And Modeling.* John Wiley & Sons, Ltd.

Haykin, S. (2001): *Kalman Filtering and Neural Networks.* John Wiley & Sons, Ltd.

Haykin, S. (2002a): *Adaptive filter theory.* Information and System Sciences Series, 4th. Prentice Hall.

Haykin, S. (2002b): *Adaptive Filter Theory (3rd Ed.).* Prentice Hall.

Hebrank, J. und Wright, D. (1974): „Spectral cues used in the localization of sound sources on the median plane." In: *J Acoust Soc Am*, 56(6) S. 1829–1834.

Heine, E. (1861): *Handbuch der Kugelfunctionen.* Berlin: Georg Reimer Verlag.

Heisenberg, W. (1927): „Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik." In: *Z. Physik*, 43(3-4) S. 172–198.

Hendriks, R. C.; Gerkmann, T. und Jensen, J. (2013): „DFT-Domain Based Single-Microphone Noise Reduction for Speech Enhancement: A Survey of the State of the Art." In: *Synthesis Lectures on Speech and Audio Processing*, 9(1) S. 1–80.

Hendriks, R. C.; Heusdens, R.; Kjems, U. und Jensen, J. R. (2009): „On Optimal Multichannel Mean-Squared Error Estimators for Speech Enhancement." In: *IEEE Signal Processing Letters*, 16(10) S. 885–888.

Hendriks, R. C.; Heusdens, R. und Jensen, J. (2010): „MMSE based noise PSD tracking with low complexity." In: *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, S. 4266–4269.

Hendriks, R. C. und Martin, R. (2007): „MAP Estimators for Speech Enhancement Under Normal and Rayleigh Inverse Gaussian Distributions." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(3) S. 918–927.

Herbordt, W. (2005): *Sound Capture for Human / Machine Interfaces: Practical Aspects of Microphone Array Signal Processing*. Lecture notes in control and information series. Berlin Heidelberg: Springer Verlag.

Herbordt, W. et al. (2003): „Full-duplex multichannel communication: real-time implementations in a general framework." In: *Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on*. S. III–49–52 vol.3.

Herzke, T. und Hohmann, V. (2007): „Improved Numerical Methods for Gammatone Filterbank Analysis and Synthesis." In: *Acta Acust United Ac*, 93(3) S. 498–500.

Hess, W. (1976): „A pitch-synchronous digital feature extraction system for phonemic recognition of speech." In: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 24(1) S. 14–25.

Hiipakka, M.; Kinnari, T. und Pulkki, V. (2012): „Estimating head-related transfer functions of human subjects from pressure–velocity measurements." In: *J Acoust Soc Am*, 131(5) S. 4051–4061.

Hirsch, H. G. und Ehrlicher, C. (1995): „Noise estimation techniques for robust speech recognition." In: *Acoustics, Speech and Signal Processing (ICASSP), 1995 IEEE International Conference on*. Detroit, MI, USA, S. 153–156.

Hofman, P.; Van Riswick, J. und Van Opstal, A. J. (1998): „Relearning sound localization with new ears." In: *Nat Neurosci*, 1(5) S. 417–421.

Hohmann, V. (2002): „Frequency analysis and synthesis using a Gammatone filterbank." In: *Acta Acust United Ac*, 88(3) S. 433–442.

Höldrich, R. und Lorber, M. (1997): „Non-Linear Spectral Subtraction with Combined Smoothing Strategies for Broadband Noise Reduction." In: *AES 103rd Convention*. New York, NY, USA.

Holdsworth, J.; Nimmo-Smith, I.; Patterson, R. D. und Rice, P. (1988): *SVOS Final Report (Annex C): Implementing a Gammatone filter bank*. Tech. Rep. APU report 2341, Applied Psychology Unit, Cambridge, UK.

Holighaus, N.; Dörfler, M.; Velasco, G. A. und Grill, T. (2013): „A Framework for Invertible, Real-Time Constant-Q Transforms." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 21(4) S. 775–785.

Holmes, S. A. und Featherstone, W. E. (2002): „A unified approach to the Clenshaw summation and the recursive computation of very high degree and order normalised associated Legendre functions." In: *Journal of Geodesy*, 76(5) S. 279–299.

Hoshuyama, O.; Begasse, B.; Sugiyama, A. und Hirano, A. (1998): „A real time robust adaptive microphone array controlled by an SNR estimate." In: *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*. S. 3605–3608.

Hoshuyama, O. und Sugiyama, A. (1996): „A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters." In: *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*. S. 925–928.

Hoshuyama, O. und Sugiyama, A. (1999): „An adaptive microphone array with good sound quality using auxiliary fixed beamformers and its DSP implementation." In: *Acoustics, Speech, and Signal Processing, 1999. ICASSP '99. Proceedings., 1999 IEEE International Conference on*. S. 949–952.

Hoshuyama, O. und Sugiyama, A. (2001): „Robust Adaptive Beamforming." In: *Microphone Arrays*. Springer Verlag, S. 87–101.

Hoshuyama, O.; Sugiyama, A. und Hirano, A. (1997): „A robust adaptive microphone array with improved spatial selectivity and its evaluation in a real environment." In: *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*. S. 367–370.

Hoshuyama, O.; Sugiyama, A. und Hirano, A. (1999): „A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters." In: *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, 47(10) S. 2677–2684.

Houtgast, T. (1977): „Auditory-filter characteristics derived from direct-masking data and pulsation-threshold data with a rippled-noise masker." In: *J Acoust Soc Am*, 62(2) S. 409–415.

Hu, Y. und Loizou, P. C. (2003): „A perceptually motivated approach for speech enhancement." In: *IEEE Transactions on Speech and Audio Processing*, 11(5) S. 457–465.

Hu, Y. und Loizou, P. C. (2004): „Incorporating a psychoacoustical model in frequency domain speech enhancement." In: , 11(2) S. 270–273.

Hu, Y. und Loizou, P. C. (2008): „Evaluation of Objective Quality Measures for Speech Enhancement." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(1) S. 229–238.

Hu, Y. und Loizou, P. C. (2007): „A comparative intelligibility study of single-microphone noise reduction algorithms." In: *J Acoust Soc Am*, 122(3) S. 1777–1786.

Huang, Y. A. und Benesty, J. (2003): „Adaptive Multichannel Time Delay Estimation Based on Blind System Identification for Acoustic Source Localization." In: *Adaptive Signal Processing*. Berlin, Heidelberg: Springer, Berlin, Heidelberg, S. 227–247.

Huang, Y. A.; Benesty, J. und Chen, J. (2006): *Acoustic MIMO Signal Processing*. Springer-Verlag, Berlin, New York, Heidelberg.

Hudde, H. (1983): „Measurement of the eardrum impedance of human ears." In: *J Acoust Soc Am*, 73(1) S. 242–247.

Hudde, H. (2005): „A functional view on the peripheral human hearing organ." In: Jens Blauert (Hrg.) *Communication Acoustics*. Springer-Verlag Berlin Heidelberg, S. 47–74.

Hudde, H. und Engel, A. (1998a): „Measuring and Modeling Basic Properties of the Human Middle Ear and Ear Canal. Part I: Model Structure and Measuring Techniques." In: *Acta Acust United Ac*, 84(4) S. 720–738.

Hudde, H. und Engel, A. (1998b): „Measuring and Modeling Basic Properties of the Human Middle Ear and Ear Canal. Part II: Ear Canal, Middle Ear Cavities, Eardrum, and Ossicles." In: *Acta Acust United Ac*, 84(5) S. 894–913.

Hudde, H. und Engel, A. (1998c): „Measuring and Modeling Basic Properties of the Human Middle Ear and Ear Canal. Part III: Eardrum Impedances, Transfer Functions and Model Calculations." In: *Acta Acust United Ac*, 84(6) S. 1091–1108.

Hudde, H.; Engel, A. und Lodwig, A. (1999): „Methods for estimating the sound pressure at the eardrum." In: *J Acoust Soc Am*, 106(4) S. 1977–1992.

Hudde, H. und Weistenhofer, C. (1997): „A three-dimensional circuit model of the middle ear." In: *Acustica*, 83(3) S. 535–549.

Hui, K. S. (2000): „Signal Processing Apparatus and Method." World Intellectual Property Organization.

Hut, R.; Boone, M. M. und Gisolf, A. (2006): „Cochlear Modeling as Time-Frequency Analysis Tool." In: *Acta Acust United Ac*, 92(4) S. 629–636.

Irino, T. (1999): „Noise suppression using a time-varying, analysis/synthesis gamma chirp filterbank." In: *Acoustics, Speech, and Signal Processing, 1999. ICASSP '99. Proceedings., 1999 IEEE International Conference on*. Phoenix, AZ, USA, S. 97–100.

Irino, T. und Kawahara, H. (1993): „Signal reconstruction from modified auditory wavelet transform." In: *Signal Processing, IEEE Transactions on*, 41(12) S. 3549–3554.

Irino, T. und Patterson, R. D. (1997): „A time-domain, level-dependent auditory filter: The gammachirp." In: *J Acoust Soc Am*, 101(1) S. 412–419.

Irino, T. und Patterson, R. D. (2001): „A compressive gammachirp auditory filter for both physiological and psychophysical data." In: *J Acoust Soc Am*, 109 S. 2008.

Irino, T. und Patterson, R. D. (2006a): „A Dynamic Compressive Gammachirp Auditory Filterbank." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(6) S. 2222–2232.

Irino, T. und Patterson, R. D. (2006b): „Dynamic, Compressive Gammachirp Auditory Filterbank for Perceptual Signal Processing." In: *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*. S. V–V.

Irino, T. und Unoki, M. (1997): *An Efficient Implementation of the Gammachirp Filter and Its Filterbank Design*. Tech. Rep. TR-H-225, ATR Human Information Processing Research Laboratories, Kyoto, Japan.

Irino, T. und Unoki, M. (1998): „A time-varying, analysis/synthesis auditory filterbank using the gammachirp." In: *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*. Seattle, WA, USA, S. 3653–3656.

Irino, T. und Unoki, M. (1999): „An Analysis/Synthesis Auditory Filterbank Based on an IIR Implementation of the Gammachirp." In: *J Acoust Soc Japan*, 20(5) S. 397–406.

Irino, T. und Unoki, M. (2001): „An Analysis/Synthesis Auditory Filterbank Based on an IIR Gammachirp Filter." In: Steven Greenberg und Malcolm Slaney (Hrgs.) *Computational Models of Auditory Function*. NATO Science Series, IOS Press, S. 47–48.

ISO (1987): „Acoustics - Normal equal-loudness level contours." In: *International Organization for Standardization*, ISO226:1987.

ISO (2003): „Acoustics - Normal equal-loudness level contours." In: *International Organization for Standardization*, ISO 226:2003.

ISO (2008): „Acoustics - Measurement of room acoustic parameters - part 2: Reverberation time in ordinary rooms."

ISO (2009a): „Acoustics - Measurement of room acoustic parameters - part 1: Performance Spaces."

ISO (2009b): „Acoustics - Measurement of room acoustic parameters - part 2: Reverberation time in ordinary rooms."

ITU-T (2001): „ITU-T P.862, Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs." International Telecommunication Union (ITU-T).

Jablon, N. (1986a): „Adaptive beamforming with the generalized sidelobe canceller in the presence of array imperfections." In: *Antennas and Propagation, IEEE Transactions on*, 34(8) S. 996–1012.

Jablon, N. (1986b): „Steady state analysis of the generalized sidelobe canceller by adaptive noise cancelling techniques." In: *Antennas and Propagation, IEEE Transactions on*, 34(3) S. 330–337.

Jablon, N. (1987): „Effect of element errors on half-power beamwidth of the Capon adaptive beamformer." In: *Circuits and Systems, IEEE Transactions on*, 34(7) S. 743–752.

Jabloun, F.; Cetin, A. E. und Erzin, E. (1999): „Teager energy based feature parameters for speech recognition in car noise." In: , 6(10) S. 259–261.

Jacobson, M. J. (1962): „Space-Time Correlation in Spherical and Circular Noise Fields." In: *J Acoust Soc Am*, 34(7) S. 971–978.

Jin, C. T.; Epain, N. und Parthy, A. (2014): „Design, Optimization and Evaluation of a Dual-Radius Spherical Microphone Array." In: *IEEE/ACM Trans. Audio Speech Lang. Process.*, 22(1) S. 193–204.

Jo, S. und Yoo, C. D. (2010): „Psychoacoustically Constrained and Distortion Minimized Speech Enhancement." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 18(8) S. 2099–2110.

Jo, S. und Yoo, C. D. (2009): „Psychoacoustically constrained and distortion minimized speech enhancement algorithm." In: *ICASSP 2009 - 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, S. 4669–4672.

Johannesma, P. I. M. (1972): „The pre-response stimulus ensemble of neurons in the cochlear nucleus." In: *Proceedings of the Symposium on Hearing Theory*. Eindhoven, The Netherlands, S. 58–69.

Johnson, D. H. (1993): *Array Signal Processing*. Concepts and Techniques. Prentice Hall.

Johnstone, B. M.; Patuzzi, R. und Yates, G. K. (1986): „Basilar membrane measurements and the travelling wave." In: *Hearing Res*, 22(1 - 3) S. 147–153.

Junhui, Z.; Jingming, K.; Xiang, X. und Shilei, H. (2003): „Noise suppression based on Teager energy operator for improving the robustness of ASR front-end." In: *International Workshop on Acoustic Echo and Noise Control (IWAENC2003)*. Kyoto, Japan, S. 135–138.

Kailath, T. (1974): „A view of three decades of linear filtering theory." In: *IEEE Trans. Inform. Theory*, 20(2) S. 146–181.

Kalman, R. E. (1960): „A New Approach to Linear Filtering and Prediction Problems." In: *J. Basic Engineering*, 82(1) S. 35.

Kalman, R. E. und Bucy, R. S. (1961): „New Results in Linear Filtering and Prediction Theory." In: *J. Basic Engineering*, 83(1) S. 95.

Kammeyer, K.-D. und Kroschel, K. (1992): *Digitale Signalverarbeitung, Filterung und Spektralanalyse*. 2nd. Teubner Verlag Stuttgart.

Kaps, A. M. (2008): *Mehrkanalige Geräuschreduktion bei Sprachsignalen mittels Kalman-Filter*. Ph.D. thesis, Technische Universität Darmstadt.

Kates, J. (1993): „Accurate tuning curves in a cochlear model." In: *IEEE Transactions on Speech and Audio Processing*, 1(4) S. 453–462.

Kates, J. (1995): „Two-tone suppression in a cochlear model." In: *IEEE Transactions on Speech and Audio Processing*, 3(5) S. 396–406.

Katsiamis, A. G.; Drakakis, E. M. und Lyon, R. F. (2006): „Introducing the Differentiated All-Pole and One-Zero Gammatone Filter Responses and their Analog VLSI Log-domain Implementation." In: *Circuits and Systems, 2006. MWSCAS '06. 49th IEEE International Midwest Symposium on*. S. 561–565.

Katsiamis, A. G.; Drakakis, E. M. und Lyon, R. F. (2007): „Practical Gammatone-Like Filters for Auditory Processing." In: *EURASIP Journal on Audio, Speech, and Music Processing*, 2007(881) S. 1–15.

Katz, B. F. (2001a): „Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation." In: *J Acoust Soc Am*, 110(5) S. 2440–2448.

Katz, B. F. (2001b): „Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements." In: *J Acoust Soc Am*, 110(5) S. 2449–2455.

Katz, B. F. und Noisternig, M. (2014): „A comparative study of interaural time delay estimation methods." In: *J Acoust Soc Am*, 135(6) S. 3530–3540.

Kennard, E. H. (1927): „Zur Quantenmechanik einfacher Bewegungstypen." In: *Z. Physik*, 44(4-5) S. 326–352.

Khintchine, A. (1934): „Korrelationstheorie der stationären stochastischen Prozesse." In: *Mathematische Annalen*, 109(1) S. 604–615.

Kistler, D. J. und Wightman, F. L. (1992): „A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction." In: *J Acoust Soc Am*, 91(3) S. 1637–1647.

Klump, G. (2005): „Evolutionary Adaptations for Auditory Communication." In: Jens Blauert (Hrg.) *Communication Acoustics*. Springer-Verlag Berlin Heidelberg, S. 27–45.

Knapp, C. und Carter, G. C. (1976): „The generalized correlation method for estimation of time delay." In: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 24(4) S. 320–327.

Köhler, B.-U. (2005): *Konzepte der statistischen Signalverarbeitung*. Berlin/Heidelberg: Springer-Verlag.

Kolossa, D. et al. (2008): „Missing feature speech recognition in a meeting situation with maximum SNR beamforming." In: *2008 IEEE International Symposium on Circuits and Systems - ISCAS 2008*. IEEE, S. 3218–3221.

Koretz, A. und Rafaely, B. (2009): „Dolph–Chebyshev Beampattern Design for Spherical Arrays." In: *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, 57(6) S. 2417–2420.

Kramme, R.; Hoffmann, K.-P. und Pozos, R. S. (Hrgs.) (2011): *Springer Handbook of Medical Technology*. Springer Verlag Berlin Heidelberg.

Krawczyk-Becker, M. und Gerkmann, T. (2016): „An evaluation of the perceptual quality of phase-aware single-channel speech enhancement." In: *J Acoust Soc Am*, 140(4) S. EL364–EL369.

Kreiß, J.-P. und Neuhaus, G. (2006): *Einführung in die Zeitreihenanalyse*. Statistik und ihre Anwendungen. Berlin/Heidelberg: Springer Verlag.

Kreuzer, W.; Majdak, P. und Chen, Z. (2009): „Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range." In: *J Acoust Soc Am*, 126(3) S. 1280–1290.

Kringlebotn, M. und Gundersen, T. (1985): „Frequency characteristics of the middle ear." In: *J Acoust Soc Am*, 77(1) S. 159–164.

Kubin, G. und Kleijn, B. W. (1999a): „Multiple-description coding (MDC) of speech with an invertible auditory model." In: *Speech Coding Proceedings, 1999 IEEE Workshop on*. S. 81–83.

Kubin, G. und Kleijn, B. W. (1999b): „On speech coding in a perceptual domain." In: *Acoustics, Speech, and Signal Processing, 1999. ICASSP '99. Proceedings., 1999 IEEE International Conference on*. Phoenix, AZ, USA, S. 205–208.

Kulkarni, A.; Isabelle, S. K. und Colburn, H. S. (1999): „Sensitivity of human subjects to head-related transfer-function phase spectra." In: *J Acoust Soc Am*, 105(5) S. 2821–2840.

Kulmer, J. und Mowlaee, P. (2015): „Phase Estimation in Single Channel Speech Enhancement Using Phase Decomposition." In: *IEEE Signal Processing Letters*, 22(5) S. 598–602.

Kuo, Y.-T.; Lin, T.-J.; Li, Y.-T. und Liu, C.-W. (2010): „Design and Implementation of Low-Power ANSI S1.11 Filter Bank for Digital Hearing Aids." In: *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 57(7) S. 1684–1696.

Küpfmüller, K. (1924): „Über Einschwingvorgänge in Wellenfiltern." In: *Elektr. Nachrichtentechnik*, 1 S. 141–152.

Kushner, W. M.; Goncharoff, V.; Wu, C.; Nguyen, V. und Damoulakis, J. N. (1989): „The effects of subtractive-type speech enhancement/noise reduction algorithms on parameter estimation for improved recognition and coding in high noise environments." In: *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*. IEEE, S. 211–214 vol.1.

Kuttruff, H. (2000): *Room Acoustics*. 4th. Elsevier Science.

Laakso, T. I.; Valimaki, V.; Karjalainen, M. und Laine, U. K. (1996): „Splitting the unit delay [FIR/all pass filters design]." In: *Signal Processing Magazine, IEEE*, 13(1) S. 30–60.

Langendijk, E. H. A. und Bronkhorst, A. W. (2002): „Contribution of spectral cues to human sound localization." In: *J Acoust Soc Am*, 112(4) S. 1583–1596.

Lazarus, H.; Sust, C. A.; Steckel, R.; Kulka, M. und Kurtz, P. (2007): *Akustische Grundlagen sprachlicher Kommunikation*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Le Bouquin-Jeannès, R. und Faucon, G. (1995): „Study of a voice activity detector and its influence on a noise reduction system." In: *Speech Commun.*, 16(3) S. 245–254.

Lebedev, N. N. (1965): *Special Functions and their Application*. Englewood Cliffs, NJ, USA: Prentice Hall.

Lebedev, V. I. (1976): „Quadratures on a sphere." In: *USSR Computational Mathematics and Mathematical Physics*, 16(2) S. 10–24.

Lebedev, V. I. (1977): „Spherical quadrature formulas exact to orders 25–29." In: *Sib Math J*, 18(1) S. 99–107.

Lentz, W. J. (1976): „Generating Bessel functions in Mie scattering calculations using continued fractions." In: *Appl. Opt.*, 15(3) S. 668–671.

Lerch, R.; Sessler, G. M. und Wolf, D. (2009): *Technische Akustik*. Springer Berlin Heidelberg.

Levinson, N. (1946): „The Wiener (Root Mean Square) Error Criterion in Filter Design and Prediction." In: *Journal of Mathematics and Physics*, 25(1) S. 261–278.

Li, Z. und Duraiswami, R. (2007): „Flexible and Optimal Design of Spherical Microphone Arrays for Beamforming." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(2) S. 702–714.

Liesen, J. und Mehrmann, V. (2015): *Lineare Algebra*. Ein Lehrbuch über die Theorie mit Blick auf die Praxis, 2. Wiesbaden: Springer Fachmedien Wiesbaden.

Lim, J. S. und Oppenheim, A. V. (1979): „Enhancement and bandwidth compression of noisy speech." In: *Proc of the IEEE*, 67(12) S. 1586–1604.

Lin, L. und Ambikairajah, E. (2002): „Speech denoising based on an auditory filterbank." In: *Signal Processing, 2002 6th International Conference on*. Bejing, China: IEEE, S. 552–555.

Lin, L.; Ambikairajah, E. und Holmes, W. H. (2001a): „Auditory Filter Bank Design Using Masking Curves." In: *7th European Conference on Speech Communication and Technology*. Aalborg, Denmark, S. 411–414.

Lin, L.; Ambikairajah, E. und Holmes, W. H. (2002): „Perceptual domain based speech and audio coder." In: *6th Int Sym DSP for Communication Systems*. Sydney, Australia, S. 6–11.

Lin, L.; Holmes, W. H. und Ambikairajah, E. (2001b): „Auditory filter bank inversion." In: *Circuits and Systems, 2001. ISCAS 2001. The 2001 IEEE International Symposium on*. S. 537–540 vol. 2.

Lin, L.; Holmes, W. H. und Ambikairajah, E. (2003): „Subband noise estimation for speech enhancement using a perceptual Wiener filter." In: *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*. Hong Kong, Hong Kong, S. 80–83.

Litovsky, R. Y. (2005): „Speech intelligibility and spatial release from masking in young children." In: *J Acoust Soc Am*, 117(5) S. 3091–3099.

Loizou, P. C. (2005): „Speech Enhancement Based on Perceptually Motivated Bayesian Estimators of the Magnitude Spectrum." In: *IEEE Transactions on Speech and Audio Processing*, 13(5) S. 857–869.

Lopez-Poveda, E. A. und Meddis, R. (2001): „A human nonlinear cochlear filterbank." In: *J Acoust Soc Am*, 110(6) S. 3107.

Lorber, M. und Höldrich, R. (1997): „A combined approach for broadband noise reduction." In: *Applications of Signal Processing to Audio and Acoustics, IEEE ASSP Workshop on*. New Paltz, NY , USA, S. 4.

Lord Rayleigh, O. M. (1907): „XII. On our perception of sound direction." In: *Philosophical Magazine Series 6*, 13(74) S. 214–232.

Lotter, T. und Vary, P. (2005): „Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model." In: *Eurasip J Appl Sig P*, 2005(7) S. 1110–1126.

Lotter, T. und Vary, P. (2004): „Noise reduction by joint maximum a posteriori spectral amplitude and phase estimation with super-Gaussian speech modelling." In: *Signal Processing Conference, 2004 12th European*. IEEE, S. 1457–1460.

Lutfi, R. A. und Patterson, R. D. (1984): „On the growth of masking asymmetry with stimulus intensity." In: *J Acoust Soc Am*, 76(3) S. 739–745.

Lyon, R. F. (1982): „A computational model of filtering, detection, and compression in the cochlea." In: *Acoustics, Speech and Signal Processing (ICASSP), 1982 IEEE International Conference on*. S. 1282–1285.

Lyon, R. F. (1996): „The All-Pole Gammatone Filter and Auditory Models." In: *Forum Acusticum*. Antwerpen, Belgium.

Lyon, R. F. (1997): „All-Pole Models of Auditory Filtering." In: E R Lewis; Richard F Lyon; G R Long und P M Narins (Hrgs.) *Diversity in Auditory Mechanics*. Singapore: World Scientific Publishing, S. 205–2011.

Lyon, R. F.; Katsiamis, A. G. und Drakakis, E. M. (2010): „History and future of auditory filter models." In: *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*. S. 3809–3812.

Lyon, R. F. und Mead, C. A. (1988a): „An analog electronic cochlea." In: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 36(7) S. 1119–1134.

Lyon, R. F. und Mead, C. A. (1988b): *Cochlear Hydrodynamics Demystified*. Tech. Rep. Caltech-CS-TR-88-4, Department of Computer Science, California Institute of Technology, Pasadena, CA, USA.

Ma, N.; Bouchard, M. und Goubran, R. A. (2006): „Speech enhancement using a masking threshold constrained Kalman filter and its heuristic implementations." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(1) S. 19–32.

Macpherson, E. A. und Middlebrooks, J. C. (2002): „Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited." In: *J Acoust Soc Am*, 111(5) S. 2219–2236.

Maday, Y.; Nguyen, N.; Patera, A. und Pau, S. (2009): „A general multipurpose interpolation procedure: the magic points." In: *Communications on Pure and Applied Analysis*, 8(1) S. 383–404.

Mailloux, R. J. (1994): *Phased Array Antenna Handbook*. Boston, MA: Artech House.

Majdak, P.; Walder, T. und Laback, B. (2013): „Effect of long-term training on sound localization performance with spectrally warped and band-limited head-related transfer functions." In: *J Acoust Soc Am*, 134(3) S. 2148–2159.

Malah, D. G.; Cox, R. V. und Accardi, A. J. (1999): „Tracking speech-presence uncertainty to improve speech enhancement in non-stationary noise environments." In: *Acoustics, Speech, and Signal Processing, 1999. ICASSP '99. Proceedings., 1999 IEEE International Conference on*. Phoenix, AZ, USA, S. 789–792.

Mallat, S. G. (1989): „Multifrequency channel decompositions of images and wavelet models." In: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(12) S. 2091–2110.

Mallat, S. (2009): *A Wavelet Tour of Signal Processing*. The Sparse Way, 3rd edition. Academic Press.

Marple, S. L. (1987): *Digital spectral analysis with applications*. Signal processing series. Prentice Hall.

Marro, C.; Mahieux, Y. und Simmer, K. U. (1998): „Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering." In: *IEEE Transactions on Speech and Audio Processing*, 6(3) S. 240–259.

Marrone, N.; Mason, C. R. und Gerald Kidd, J. (2008): „The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms." In: *J Acoust Soc Am*, 124(5) S. 3064–3075.

Martin, R. (2001): „Noise power spectral density estimation based on optimal smoothing and minimum statistics." In: *IEEE Transactions on Speech and Audio Processing*, 9(5) S. 504–512.

Martin, R. (2002): „Speech enhancement using MMSE short time spectral estimation with gamma distributed speech priors." In: *Acoustics, Speech and Signal Processing (ICASSP), 2002 IEEE International Conference on*. Orlando, Florida, USA, S. 253–253.

Martin, R. (2005): „Speech enhancement based on minimum mean-square error estimation and supergaussian priors." In: *IEEE Transactions on Speech and Audio Processing*, 13(5) S. 845–856.

Martin, R. und Breithaupt, C. (2003): „Speech enhancement in the DFT domain using Laplacian speech priors." In: *International Workshop on Acoustic Echo and Noise Control (IWAENC2003)*. Kyoto, Japan, S. 87–90.

Martin, R.; Malah, D. G.; Cox, R. V. und Accardi, A. J. (2004): „A noise reduction preprocessor for mobile voice communication." In: *Eurasip J Appl Sig P*, 2004(8) S. 1046–1058.

Martin, R.; Wittke, I. und Jax, P. (2000): „Optimized estimation of spectral parameters for the coding of noisy speech." In: *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on*. Istanbul, Turkey, S. 1479–1482.

Maynard, J. D.; Williams, E. G. und Lee, Y. (1985): „Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH." In: *J Acoust Soc Am*, 78(4) S. 1395–1413.

Maynard, J. D. (1997): „Acoustic Holography." In: *Encyclopedia of Acoustics*. Hoboken, NJ, USA: John Wiley & Sons, Inc., S. 1281–1290.

McAulay, R. und Malpass, M. (1980): „Speech enhancement using a soft-decision noise suppression filter." In: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 28(2) S. 137–145.

McCowan, I. A. und Bourlard, H. (2003): „Microphone array post-filter based on noise field coherence." In: *IEEE Transactions on Speech and Audio Processing*, 11(6) S. 709–716.

McCowan, I. A. und Bourlard, H. (2002): „Microphone array post-filter for diffuse noise field." In: *Proceedings of ICASSP '02*. IEEE, S. I–905–I–908.

McFadden, D. und Yama, M. F. (1983): „Upward shifts in the masking pattern with increasing masker intensity." In: *J Acoust Soc Am*, 74(4) S. 1185–1189.

Meddis, R. und Lopez-Poveda, E. A. (2010): „Auditory Periphery: From Pinna to Auditory Nerve." In: Ray Meddis; Enrique A Lopez-Poveda; Arthur N Popper und Richard R Fay (Hrgs.) *Computational Models of the Auditory System*. Springer New York Dordrecht Heidelberg London, S. 7–38.

Meddis, R.; Lopez-Poveda, E. A.; Popper, A. N. und Fay, R. R. (Hrgs.) (2010): *Computational Models of the Auditory System*. Springer Handbook of Auditory Research. Springer New York Dordrecht Heidelberg London.

Meddis, R.; O'Mard, L. P. und Lopez-Poveda, E. A. (2001): „A computational algorithm for computing nonlinear auditory frequency selectivity." In: *J Acoust Soc Am*, 109(6) S. 2852–2861.

Meddis, R. et al. (2013): „A Computer Model of the Auditory Periphery and Its Application to the Study of Hearing." In: Brian C. J. Moore; Roy D Patterson; Ian M Winter; Robert P Carlyon und Hedwig E Gockel (Hrgs.) *Basic Aspects of Hearing*. New York: Springer, S. 11–20.

Mehrgardt, S. und Mellert, V. (1977): „Transformation characteristics of the external human ear." In: *J Acoust Soc Am*, 61(6) S. 1567–1576.

Meister, A. (2015): *Numerik linearer Gleichungssysteme*. Eine Einführung in moderne Verfahren, 5. Wiesbaden, Germany: Springer Fachmedien.

Melenk, J. M. (1999): „Operator adapted spectral element methods I: harmonic and generalized harmonic polynomials." In: *Numerische Mathematik*, 84(1) S. 35–69.

Metz, O. (1951): „Studies on the contraction of the tympanic muscles as indicated by changes in the impedance of the ear." In: *Acta Otolaryngol.*, 39(5) S. 397–405.

Meyer, J. (2001): „Beamforming for a circular microphone array mounted on spherically shaped objects." In: *J Acoust Soc Am*, 109(1) S. 185–193.

Meyer, J. und Elko, G. W. (2002): „A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield." In: *Acoustics, Speech and Signal Processing (ICASSP), 2002 IEEE International Conference on*. S. 1781–1784.

Meyer, J. und Elko, G. W. (2008): „Handling aliasing in spherical array applications." In: *2008 Hands-Free Speech Communication and Microphone Arrays (HSCMA 2008)*. IEEE, S. 1–4.

Meyer, J.; Simmer, K. U. und Kammeyer, K.-D. (1997): „Comparison of one- and two-channel noise-estimation techniques." In: *5th International Workshop on Acoustic Echo and Noise Control*. London, UK, S. 137–145.

Michel, V. (2013): *Lectures on Constructive Approximation*. Fourier, Spline, and Wavelet Methods on the Real Line, the Sphere, and the Ball. Boston: Birkhäuser Boston.

Middlebrooks, J. C. (1999a): „Individual differences in external-ear transfer functions reduced by scaling in frequency." In: *J Acoust Soc Am*, 106(3) S. 1480–1492.

Middlebrooks, J. C. (1999b): „Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency." In: *J Acoust Soc Am*, 106(3) S. 1493–1510.

Middlebrooks, J. C. und Green, D. M. (1991): „Sound Localization by Human Listeners." In: *Annual Review of Psychology*, 42(1) S. 135–159.

Middlebrooks, J. C.; Makous, J. C. und Green, D. M. (1989): „Directional sensitivity of sound-pressure levels in the human ear canal." In: *J Acoust Soc Am*, 86(1) S. 89–108.

Mignot, R.; Chardon, G. und Daudet, L. (2014): „Low Frequency Interpolation of Room Impulse Responses Using Compressed Sensing." In: *IEEE/ACM Trans. Audio Speech Lang. Process.*, 22(1) S. 205–216.

Minnaar, P.; Plogsties, J.; Olesen, S. K.; Christensen, F. und Møller, H. (2000): „The Interaural Time Difference in Binaural Synthesis." In: *AES 108th Convention*. Paris, France, S. 1–20.

Moiola, A.; Hiptmair, R. und Perugia, I. (2011): „Plane wave approximation of homogeneous Helmholtz solutions." In: *Z Angew Math Phys*, 62(5) S. 809–837.

Møller, A. R. und Nilsson, H. G. (1979): „Inner Ear Impulse Response and Basilar Membrane Modelling." In: *Acta Acust United Ac*, 41(4) S. 258–262.

Møller, A. R. (1961): „Network Model of the Middle Ear." In: *J Acoust Soc Am*, 33(2) S. 168–176.

Møller, A. R. (2006): *Hearing: Anatomy, Physiology, and Disorders of the Auditory System*. 2nd. Academic Press.

Møller, H.; Sørensen, M. F.; Hammershøi, D. und Jensen, C. B. (1995): „Head-Related Transfer Functions of Human Subjects." In: *J Audio Eng Soc*, 43(5) S. 300–321.

Monzingo, R. A. und Miller, T. W. (1980): *Introduction to Adaptive Arrays*. New York: John Wiley and Sons.

Moore, B. C. J. (1993): „Characterisation of simultanous, forward and backward masking." In: *AES 12th International Conference*. Copenhagen, Denmark, S. 22–33.

Moore, B. C. J. (1995): *Hearing*. 2nd. Academic Press.

Moore, B. C. J. (2013): *An Introduction to the Psychology of Hearing*. 6th edition. Leiden, The Netherlands: Brill.

Moore, B. C. J. und Glasberg, B. R. (1983): „Suggested formulae for calculating auditory-filter bandwidths and excitation patterns." In: *J Acoust Soc Am*, 74(3) S. 750–753.

Moore, B. C. J.; Glasberg, B. R. und Baer, T. (1997): „A Model for the Prediction of Thresholds, Loudness, and Partial Loudness." In: *J Audio Eng Soc*, 45(4) S. 224–240.

Moore, B. C. J.; Peters, R. W. und Glasberg, B. R. (1990): „Auditory filter shapes at low center frequencies." In: *J Acoust Soc Am*, 88(1) S. 132–140.

Morgenstern, H.; Klein, J.; Rafaely, B. und Noisternig, M. (2016): „Experimental investigation of multiple-input multiple-output systems for sound-field analysis." In: *ICA*. Buenos Aires, Argentina.

Morse, P. M. und Feshbach, H. (1953): *Methods of Theoretical Physics (Part 2)*. McGraw-Hill Publishing.

Morse, P. M. und Ingard, K. U. (1968): *Theoretical Acoustics*. McGraw-Hill Book Company.

Moschytz, G. und Hofbauer, M. (2000): *Adaptive Filter*. Springer Verlag.

Mowlaee, P. und Kulmer, J. (2015): „Phase Estimation in Single-Channel Speech Enhancement: Limits-Potential." In: *IEEE/ACM Trans. Audio Speech Lang. Process.*, 23(8) S. 1283–1294.

Müller, C. (1966): *Spherical Harmonics*, vol. 17 von *Lecture Notes in Mathematics*. Springer Berlin / Heidelberg.

Necciari, T.; Balazs, P.; Holighaus, N. und Sondergaard, P. L. (2013): „The ERBlet transform: An auditory-based time-frequency representation with perfect reconstruction." In: *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. S. 498–502.

Nedzelnitsky, V. (1980): „Sound pressures in the basal turn of the cat cochlea." In: *J Acoust Soc Am*, 68(6) S. 1676–1689.

Neely, S. T. und Allen, J. B. (1979): „Invertibility of a room impulse response." In: *J Acoust Soc Am*, 66(1) S. 165–169.

Nielsen, L. B. (1993): *An Auditory Model with Hearing Loss*. Tech. Rep. report no. 52, Technical University of Denmark, Lyngby, Denmark.

Nin, F. et al. (2008): „The endocochlear potential depends on two K+ diffusion potentials and an electrical barrier in the stria vascularis of the inner ear." In: *Proc Natl Acad Sci USA*, 105(5) S. 1751–1756.

Nishino, T.; Inoue, N.; Takeda, K. und Itakura, F. (2007): „Estimation of HRTFs on the horizontal plane using physical features." In: *Appl Acoust*, 68(8) S. 897–908.

Noble, W. und Perrett, S. (2002): „Hearing speech against spatially separate competing speech versus competing noise." In: *Perception & psychophysics*, 64(8) S. 1325–1336.

Noisternig, M. und Katz, B. F. (2009): „Reconstructing sound source directivity in virtual acoustic environments." In: *IWPASH*. S. 1–4.

Noisternig, M.; Klein, J.; Berzborn, M.; Recher, A. und Warusfel, O. (2016): „High-Resolution MIMO DRIR Measurements in an Opera Hall." In: *42nd Annual German Congress on Acoustics DAGA*. Aachen, Germany.

Noisternig, M.; Zotter, F.; Höldrich, R. und Opitz, M. (2009): „Method and device for low-latency auditory model-based single-channel speech enhancement." World Intellectual Property Organization.

Noisternig, M.; Zotter, F. und Katz, B. F. (2011): „Reconstructing sound source directivity in virtual acoustic environments." In: Hiroaki Kato; Douglas S Brungart und Yôiti Suzuki (Hrgs.) *Principles and Applications of Spatial Hearing*. World Scientific Publishing Co. Pte. Ltd., S. 357–373.

Nordebo, S.; Claesson, I. und Nordholm, S. (1994): „Adaptive beamforming: Spatial filter designed blocking matrix." In: *IEEE Journal of Oceanic Engineering*, 19(4) S. 583–590.

O'Donovan, J. J. und Furlong, D. J. (2005): „Perceptually motivated time-frequency analysis." In: *J Acoust Soc Am*, 117(1) S. 250–262.

Okuda, M.; Ikehara, M. und Takahashi, S. (1998): „Fast and stable least-squares approach for the design of linear phase FIR filters." In: *IEEE Trans. Signal Process.*, 46(6) S. 1485–1493.

Omologo, M. und Svaizer, P. (1994): „Acoustic event localization using a crosspower-spectrum phase based technique." In: *ICASSP-94*. IEEE, S. II/273–II/276.

Omologo, M. und Svaizer, P. (1996): „Acoustic source location in noisy and reverberant environment using CSP analysis." In: *ICASSP-96*. IEEE, S. 921–924.

Onchi, Y. (1961): „Mechanism of the Middle Ear." In: *J Acoust Soc Am*, 33(6) S. 794–805.

Oppenheim, A. V.; Schafer, R. W. und Buck, J. R. (1998): *Discrete-Time Signal Processing*. 2nd. Prentice Hall.

Otani, M.; Hirahara, T. und Ise, S. (2009): „Numerical study on source-distance dependency of head-related transfer functions." In: *J Acoust Soc Am*, 125(5) S. 3253–3261.

Otani, M. und Ise, S. (2006): „Fast calculation system specialized for head-related transfer function based on boundary element method." In: *J Acoust Soc Am*, 119(5) S. 2589–2598.

Oxenham, A. J. und Plack, C. J. (1997): „A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing." In: *J Acoust Soc Am*, 101(6) S. 3666–3675.

Paliwal, K. K. und Basu, A. (1987): „A speech enhancement method based on Kalman filtering." In: *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '87.* IEEE, S. 177–180.

Paliwal, K. K.; Wójcicki, K. und Shannon, B. (2011): „The importance of phase in speech enhancement." In: *Speech Commun*, 53(4) S. 465–494.

Pape, L. (2005): *Vergleich Robuster Mikrofonarrays.* Masterarbeit, Universität für Musik und darstellende Kunst, Graz.

Papoulis, A. (1991): *Probability, Random Variables, and Stochastic Processing (3rd ed.).* McGraw-Hill Publishing.

Parseihian, G. und Katz, B. F. (2012): „Rapid head-related transfer function adaptation using a virtual auditory environment." In: *J Acoust Soc Am*, 131(4) S. 2948–2957.

Parthy, A.; Jin, C. T. und van Schaik, A. (2009): „Acoustic holography with a concentric rigid and open spherical microphone array." In: *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on.* IEEE, S. 2173–2176.

Pascal, J.; Bourgeade, A.; Lagier, M. und Legros, C. (1998): „Linear and nonlinear model of the human middle ear." In: *J Acoust Soc Am*, 104(3) S. 1509–1516.

Patterson, R. D. (1976): „Auditory filter shapes derived with noise stimuli." In: *J Acoust Soc Am*, 59(3) S. 640–654.

Patterson, R. D. (1987): „A pulse ribbon model of monaural phase perception." In: *J Acoust Soc Am*, 82(5) S. 1560–1586.

Patterson, R. D. (1994): „The sound of a sinusoid: Spectral models." In: *J Acoust Soc Am*, 96(3) S. 1409–1418.

Patterson, R. D. (1995): „Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform." In: *J Acoust Soc Am*, 98(4) S. 1890–1894.

Patterson, R. D. und Henning, G. B. (1977): „Stimulus variability and auditory filter shape." In: *J Acoust Soc Am*, 62(3) S. 649–664.

Patterson, R. D. und Moore, B. C. J. (1986): „Auditory filters and excitation patterns as representations of frequency resolution." In: Brian C. J. Moore (Hrg.) *Frequency Selectivity in Hearing*. Academic Press, S. 123–177.

Patterson, R. D. und Nimmo-Smith, I. (1980): „Off-frequency listening and auditory-filter asymmetry." In: *J Acoust Soc Am*, 67(1) S. 229–245.

Patterson, R. D.; Nimmo-Smith, I.; Holdsworth, J. und Rice, P. (1987): *SVOS Final Report (Annex B): An efficient auditory filterbank based on the Gammatone function* . Tech. Rep. APU report 2341, Applied Psychology Unit, Cambridge, UK.

Patterson, R. D.; Nimmo-Smith, I.; Holdsworth, J. und Rice, P. (1988): *SVOS Final Report (Part A).* Tech. Rep. APU report 2341, Applied Psychology Unit, Cambridge, UK.

Patterson, R. D.; Nimmo-Smith, I.; Weber, D. L. und Milroy, R. (1982): „The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold." In: *J Acoust Soc Am*, 72(6) S. 1788–1803.

Patterson, R. D. und Rice, P. (1987): *SVOS Final Report (Annex A): A preliminary study of the feasibility of a hardware version of the auditory filter bank.* Tech. Rep. APU report 2341, Applied Psychology Unit, Cambridge, UK.

Patterson, R. D.; Unoki, M. und Irino, T. (2003): „Extending the domain of center frequencies for the compressive gammachirp auditory filter." In: *J Acoust Soc Am*, 114(3) S. 1529.

Patterson, R. D. et al. (1992): „Complex sounds and auditory images." In: Y Cazals; L Demany und K Horner (Hrgs.) *9th International Symposium on Hearing*. Oxford, UK, S. 429–446.

Peacock, K. L. und Treitel, S. (1969): „Predictive deconvolution: theory and practice." In: *Geophysics*, 34(2) S. 155–169.

Pflüger, M. (1997): *Modelle des peripheren Gehörs am Beispiel der menschlichen Lautheitsempfindung*. Ph.D. thesis, University of Technology, Graz, Austria.

Pflüger, M.; Höldrich, R. und Riedler, W. (1997): „Nonlinear All-Pole and One-Zero Gammatone Filters." In: *Acta Acust United Ac*, 84(3) S. 513–519.

Philippe, P.; de Saint-Martin, F. M. und Lever, M. (1999): „Wavelet packet filterbanks for low time delay audio coding." In: *IEEE Transactions on Speech and Audio Processing*, 7(3) S. 310–322.

Pichevar, R.; Najaf-Zadeh, H.; Thibault, L. und Lahdili, H. (2011): „Auditory-inspired sparse representation of audio signals." In: *Speech Commun.*, 53(5) S. 643–657.

Pick, G. F. (1980): „Level dependence of psychophysical frequency resolution and auditory filter shape." In: *J Acoust Soc Am*, 68(4) S. 1085–1095.

Pickles, J. (2012): *An Introduction to the Physiology of Hearing*. 4th. Emerald Group Publishing Limited.

Pielemeier, W. J. und Wakefield, G. H. (1996): „A high-resolution time–frequency representation for musical instrument signals." In: *J Acoust Soc Am*, 99(4) S. 2382–2396.

Piersol, A. (1981): „Time delay estimation using phase data." In: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29(3) S. 471–477.

Plack, C. J.; Oxenham, A. J. und Drga, V. (2002): „Linear and Nonlinear Processes in Temporal Masking." In: *Acta Acust United Ac*, 88(3) S. 348–358.

Plapous, C.; Marro, C. und Scalart, P. (2006): „Improved Signal-to-Noise Ratio Estimation for Speech Enhancement." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(6) S. 2098–2108.

Pollow, M.; Dietrich, P.; Krechel, B. und Vorländer, M. (2011): „Unidirektionale mehrkanalige Audioübertragung über Ethernet." In: *37th Annual Convention of Acoustics (DAGA)*. Düsseldorf, Germany, S. 241–242.

Pollow, M. et al. (2012): „Calculation of Head-Related Transfer Functions for Arbitrary Field Points Using Spherical Harmonics Decomposition." In: *Acta Acust United Ac*, 98(1) S. 72–82.

Porat, B. und Friedlander, B. (1985): „Asymptotic analysis of the bias of the modified Yule-Walker estimator." In: *Automatic Control, IEEE Transactions on*, 30(8) S. 765–767.

Porter, J. und Boll, S. (1984): „Optimal estimators for spectral restoration of noisy speech." In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. Institute of Electrical and Electronics Engineers, S. 53–56.

Pralong, D. und Carlile, S. (1994): „Measuring the human head-related transfer functions: A novel method for the construction and calibration of a miniature "in-ear" recording system." In: *J Acoust Soc Am*, 95(6) S. 3435–3444.

Pralong, D. und Carlile, S. (1996): „The role of individualized headphone calibration for the generation of high fidelity virtual auditory space." In: *J Acoust Soc Am*, 100(6) S. 3785–3793.

Preuss, R. D. (1979): „A frequency domain noise cancelling preprocessor for narrowband speech communications systems." In: *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '79.* Washington D.C, USA: IEEE, S. 212–215.

Rabbitt, R. D. und Holmes, M. H. (1988): „Three-dimensional acoustic waves in the ear canal and their interaction with the tympanic membrane." In: *J Acoust Soc Am*, 83(3) S. 1064–1080.

Rabiner, L. R. und Schafer, R. W. (1978): *Digital processing of speech signals*. Prentice Hall.

Rafaely, B. (2004): „Plane-wave decomposition of the sound field on a sphere by spherical convolution." In: *J Acoust Soc Am*, 116(4) S. 2149–2157.

Rafaely, B. (2008): „The Spherical-Shell Microphone Array." In: *IEEE Trans. Audio Speech Lang. Process.*, 16(4) S. 740–747.

Rafaely, B. (2011): „Bessel Nulls Recovery in Spherical Microphone Arrays for Time-Limited Signals." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 19(8) S. 2430–2438.

Rafaely, B. (2015): *Fundamentals of Spherical Array Processing*, vol. 8 von *Springer Topics in Signal Processing*. Berlin Heidelberg: Springer.

Rafaely, B.; Weiss, B. und Bachmat, E. (2007): „Spatial Aliasing in Spherical Microphone Arrays." In: *IEEE Trans. Signal Process.*, 55(3) S. 1003–1010.

Rahim, T. und Davies, D. E. N. (1982): „Effect of directional elements on the directional response of circular antenna arrays." In: *Microwaves, Optics and Antennas, IEE Proceedings H*. S. 18–22.

Rakhmanov, E. A.; Saff, E. B. und Zhou, Y. M. (1994): „Minimal Discrete Energy on the Sphere." In: *Math. Res. Lett.*, 1(6) S. 647–662.

Rangachari, S. und Loizou, P. C. (2006): „A noise-estimation algorithm for highly non-stationary environments." In: *Speech Commun*, 48(2) S. 220–231.

Recio, A. und Rhode, W. S. (2000): „Basilar membrane responses to broadband stimuli." In: *J Acoust Soc Am*, 108(5) S. 2281–2298.

Reeger, J. A. und Fornberg, B. (2015): „Numerical Quadrature over the Surface of a Sphere." In: *Studies in Applied Mathematics*, 137(2) S. 174–188.

Regalia, P. A.; Mitra, S. K. und Vaidyanathan, P. P. (1988): „The digital all-pass filter: a versatile signal processing building block." In: *Proceedings of the IEEE*. S. 19–37.

Reitbauer, C. (2012): *Entwicklung eines echtzeitfähigen Systems zur akustischen Detektion der Fahrtrichtung.* Masterarbeit, University of Technology, Graz, Austria.

Reitbauer, C.; Rainer, H.; Noisternig, M.; Rettenbacher, B. und Graf, F. (2012): „Micarray - A System for Multichannel Audio Streaming over Ethernet." In: *5th Congress of the Alps Adria Acoustics Association.* Petrcane (Zadar), Croatia, S. 1–4.

Renevey, P. und Drygajlo, A. (2001): „Detection of reliable features for speech recognition in noisy conditions using a statistical criterion." In: *Workshop on Consistent and Reliable Acoustic Cues for Sound Analysis.* Aalborg, Denmark, S. 71–74.

Rhode, W. S. (1971): „Observations of the Vibration of the Basilar Membrane in Squirrel Monkeys using the M[o-umlaut]ssbauer Technique." In: *J Acoust Soc Am*, 49(4B) S. 1218–1231.

Rhode, W. S. und Recio, A. (2000): „Study of mechanical motions in the basal region of the chinchilla cochlea." In: *J Acoust Soc Am*, 107(6) S. 3317–3332.

Robinson, D. W. und Dadson, R. S. (1956): „A re-determination of the equal-loudness relations for pure tones." In: *British Journal of Applied Physics*, 7(5) S. 166–181.

Rosen, S. und Baker, R. J. (1994): „Characterising auditory filter nonlinearity." In: *Hearing Res*, 73(2) S. 231–243.

Rosen, S.; Baker, R. J. und Darling, A. (1998): „Auditory filter nonlinearity at 2 kHz in normal hearing listeners." In: *J Acoust Soc Am*, 103(5) S. 2539–2550.

Ruggero, M. A.; Rich, N. C.; Recio, A.; Narayan, S. S. und Robles, L. (1997): „Basilar-membrane responses to tones at the base of the chinchilla cochlea." In: *J Acoust Soc Am*, 101(4) S. 2151–2163.

Runge, C. (1901): „Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten." In: *Zeitschrift für Mathematik und Physik*, 46 S. 224–243.

Ruzicka, M. (2004): *Nichtlineare Funktionalanalysis*. Springer Verlag Berlin Heidelberg New York.

Sabine, W. C. (1922): *Collected Papers On Acoustics*. Cambridge Harvard University Press.

Saff, E. B. und Kuijlaars, A. B. J. (1997): „Distributing many points on a sphere." In: *The Mathematical Intelligencer*, 19(1) S. 5–11.

Scalart, P. und Filho, J. V. (1996): „Speech enhancement based on a priori signal to noise estimation." In: *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*. Atlanta, GA, USA, S. 629–632.

Schelkunoff, S. A. (1943): „A Mathematical Theory of Linear Arrays." In: *Bell Syst Tech J*, 22(1) S. 80–107.

Schofield, D. (1985): *Visualizations of speech based on a model of the peripheral audio system*. Tech. Rep. 62/85, National Physical Laboratory, Teddington, UK.

Schroeder, M. R. (1965): „Apparatus for suppressing noise and distortion in communication signals." In: *United States Patent Office*, US 3180936 A.

Schroeder, M. R. (1968): „Processing of communications signals to reduce effects of noise." In: *United States Patent Office*, US 3403224 A.

Schwab, A. J. und Kürner, W. (2007): *Elektromagnetische Verträglichkeit*. 5. Springer Berlin Heidelberg.

Sezan, M. I. und Stark, H. (1982a): „Correction to Image Restoration by the Method of Convex Projections: Part 2 - Applications and Numerical Results." In: *IEEE Transactions on Medical Imaging*, 1(3) S. 204–204.

Sezan, M. I. und Stark, H. (1982b): „Image Restoration by the Method of Convex Projections: Part 2 - Applications and Numerical Results." In: *IEEE Transactions on Medical Imaging*, 1(2) S. 95–101.

Shackleton, T. M.; McAlpine, D. und Palmer, A. R. (2000): „Modelling convergent input onto interaural-delay-sensitive inferior colliculus neurones." In: *Hearing Res*, 149(1–2) S. 199–215.

Shailer, M. J.; Moore, B. C. J.; Glasberg, B. R.; Watson, N. und Harris, S. (1989): „Auditory filter shapes at 8 and 10 kHz." In: *J Acoust Soc Am*, 88(1) S. 141–148.

Shannon, B. und Paliwal, K. K. (2006): „Role of Phase Estimation in Speech Enhancement." In: *Spoken Language Process ICSLP, International Conference on*. Pittsburgh, PA, USA, S. 1423–1426.

Shaw, E. A. G. (1974a): „The external ear." In: W D Keidel und W D Neff (Hrgs.) *Handbook of Sensory Physiology*. Berlin: Springer, S. 455–490.

Shaw, E. A. G. (1974b): „Transformation of sound pressure level from the free field to the eardrum in the horizontal plane." In: *J Acoust Soc Am*, 56(6) S. 1848–1861.

Shaw, E. A. G. und Teranishi, R. (1968): „Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Point Source." In: *J Acoust Soc Am*, 44(1) S. 240–249.

Shaw, E. A. G. und Vaillancourt, M. M. (1985): „Transformation of sound-pressure level from the free field to the eardrum presented in numerical form." In: *J Acoust Soc Am*, 78(3) S. 1120–1123.

Shera, C. A. (2001): „Frequency glides in click responses of the basilar membrane and auditory nerve: Their scaling behavior and origin in traveling-wave dispersion." In: *J Acoust Soc Am*, 109(5) S. 2023–2034.

Shinn-Cunningham, B. G.; Durlach, N. I. und Held, R. M. (1998): „Adapting to supernormal auditory localization cues. I. Bias and resolution." In: *J Acoust Soc Am*, 103(6) S. 3656–3666.

Simmer, K. U.; Bitzer, J. und Marro, C. (2001): „Post-Filtering Techniques." In: Michael S Brandstein und Darren B Ward (Hrgs.) *Microphone Arrays*. Springer Verlag, S. 39–60.

Simoncelli, E. P.; Pillow, J.; Paninski, L. und Schwartz, O. (2004): „Characterization of neural responses with stochastic stimuli." In: M Gazzaniga (Hrg.) *The Cognitive Neurosciences*. MIT Press, S. 327–338.

Slaney, M. (1988): *Lyon's Cochlear Model*. Tech. Rep. 13, Advanced Technology Group, Apple Computer Inc.

Slaney, M. (1993): *An efficient implementation of the Patterson-Holdsworth auditory filter bank*. Tech. Rep. 35, Advanced Technology Group, Apple Computer Inc.

Slaney, M. und Lyon, R. F. (1993): „On the importance of time-a temporal representation of sound." In: Martin Cooke; Steve Beet und Malcolm Crawford (Hrgs.) *Visual Representations of Speech Signals*. John Wiley and Sons Ltd, S. 95–116.

Slaney, M.; Naar, D. und Lyon, R. F. (1994): „Auditory model inversion for sound separation." In: *Acoustics, Speech and Signal Processing (ICASSP), 1994 IEEE International Conference on*. Adelaide, South Australia, Australia: IEEE, S. 77–80.

Slepian, D. und Pollak, H. O. (2013): „Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty - I." In: *Bell Syst Tech J*, 40(1) S. 43–63.

Sloan, I. H. und Womersley, R. S. (1998): „The Uniform Error of Hyperinterpolation on the Sphere." In: Werner Haußmann; Kurt Jetter und Manfred Reimer (Hrgs.) *Advances in Multivariate Approximation*. Witten-Bommerholz, Germany: Wiley-VCH Verlag, S. 289–306.

Sloan, I. und Womersley, R. (2004): „Extremal Systems of Points and Numerical Integration on the Sphere." In: *Adv Comput Math*, 21(1-2) S. 107–125.

Smith III, J. O. und Abel, J. S. (1999): „Bark and ERB bilinear transforms." In: *IEEE Transactions on Speech and Audio Processing*, 7(6) S. 697–708.

Sneeuw, N. (1994): „Global spherical harmonic analysis by least-squares and numerical quadrature methods in historical perspective." In: *Geophysical Journal International*, 118(3) S. 707–716.

Song, Z.; Zhang, T.; Zhang, D. und Song, T. (2009): „Voice activity detection using higher-order statistics in the teager energy domain." In: *Wireless Communications & Signal Processing, 2009. WCSP 2009. International Conference on*. Nanjing, China: IEEE, S. 1–5.

Stahl, V.; Fischer, A. und Bippus, R. (2000): „Quantile based noise estimation for spectral subtraction and Wiener filtering." In: *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on*. Istanbul, Turkey, S. 1875–1878.

Stanford, V.; Garofolo, J.; Galibert, O.; Michel, M. und Laprun, C. (2003): „The NIST Smart Space and Meeting Room projects: signals, acquisition annotation, and metrics." In: *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*. S. 736–739.

Stinson, M. R. (1985): „The spatial distribution of sound pressure within scaled replicas of the human ear canal." In: *J Acoust Soc Am*, 78(5) S. 1596–1602.

Stinson, M. R.; Shaw, E. A. G. und Lawton, B. W. (1982): „Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal." In: *J Acoust Soc Am*, 72(3) S. 766–773.

Strahl, S. und Mertins, A. (2009): „Analysis and design of gammatone signal models." In: *J Acoust Soc Am*, 126(5) S. 2379–2389.

Strobel, N.; Spors, S. und Rabenstein, R. (2001): „Joint Audio-Video Signal Processing for Object Localization and Tracking." In: *Microphone Arrays*. Berlin, Heidelberg: Springer, Berlin, Heidelberg, S. 203–225.

Sunnydayal, V. und Kumar, T. K. (2015): „Bayesian estimation for speech enhancement given a priori knowledge of clean speech phase." In: *Int J Speech Technol*, 18(4) S. 593–607.

Suzuki, Y. und Takeshima, H. (2004): „Equal-loudness-level contours for pure tones." In: *J Acoust Soc Am*, 116(2) S. 918–933.

Suzuki, Y. et al. (Hrgs.) (2011): *Principles and Applications of Spatial Hearing*. Singapore, Singapore: World Scientific Publishing.

Swets, J. A.; Green, D. M. und Tanner, W. P. (1962): „On the Width of Critical Bands." In: *J Acoust Soc Am*, 34(1) S. 108–113.

Sydow, C. (1994): „Broadband beamforming for a microphone array." In: *J Acoust Soc Am*, 96(2) S. 845–849.

Tanaka, T. und Shiono, M. (2014): „Acoustic Beamforming with Maximum SNR Criterion and Efficient Generalized Eigenvector Tracking." In: *Advances in Multimedia Information Processing – PCM 2014*. Cham: Springer International Publishing, S. 373–382.

Terhardt, E. (1998): *Akustische Kommunikation*. Springer Verlag.

Thiemann, J. (2001): *Acoustic noise suppression for speech signals using auditory masking effects*. Masterarbeit, Department of Electrical & Computer Engineering McGill University, Montreal, Canada.

Thiemann, J. und Kabal, P. (2002): „Low distortion acoustic noise suppression using a perceptual model for speech signals." In: *Speech Coding, 2002, IEEE Workshop Proceedings*. Tsukuba City, Ibaraki, Japan, S. 172–174.

Tourbabin, V. und Rafaely, B. (2015): „On the Consistent Use of Space and Time Conventions in Array Processing." In: *Acta Acust United Ac*, 101(3) S. 470–473.

Traunmüller, H. (1990): „Analytical expressions for the tonotopic sensory scale." In: *J Acoust Soc Am*, 88(1) S. 97–100.

Trine, T. D. und Van Tasell, D. (2002): „Digital hearing aid design." In: *Hearing Journal*, 55(2) S. 36–38.

Tsoukalas, D.; Mourjopoulos, J. und Kokkinakis, G. C. (1997): „Speech enhancement based on audible noise suppression." In: *IEEE Transactions on Speech and Audio Processing*, 5(6) S. 497–514.

Tuffy, M. (1999): *The Removal of Environmental Noise in Cellular Communications by Perceptual Techniques*. Ph.D. thesis, The University of Edinburgh.

Unoki, M.; Irino, T.; Glasberg, B.; Moore, B. C. J. und Patterson, R. D. (2006): „Comparison of the roex and gammachirp filters as representations of the auditory filter." In: *J Acoust Soc Am*, 120(3) S. 1474–1492.

Van Compernolle, D. (1990): „Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings." In: *International Conference on Acoustics, Speech, and Signal Processing*. IEEE, S. 833–836.

Van Compernolle, D. (1991): *Development of a Computational Auditory Model*. Tech. rep., Instituut voor Perceptie Onderzoek (IPO), Eindhoven, The Netherlands.

Van Immerseel, L. und Peeters, S. (2003): „Digital implementation of linear gammatone filters: Comparison of design methods." In: *ARLO*, 4(3) S. 59–64.

van Netten, S. M. und Duifhuis, H. (1983): „Modelling an Active, Nonlinear Cochlea." In: *Mechanics of Hearing*. Dordrecht: Springer Netherlands, S. 143–151.

Van Trees, H. L. (2002): *Optimum Array Processing*. Part IV of Detection, Estimation, and Modulation Theory. Wiley Interscience.

Van Veen, B. D. und Buckley, K. M. (1988): „Beamforming: a versatile approach to spatial filtering." In: *ASSP Magazine, IEEE*. S. 4–24.

Vary, P.; Heute, U. und Hess, W. (1998): *Digitale Sprachsignalverarbeitung*. B. G. Teubner Stuttgart.

Vary, P. und Martin, R. (2006): *Digital Speech Transmission*. John Wiley & Sons, Ltd.

Vaseghi, S. V. (2006): *Advanced Digital Signal Processing and Noise Reduction*. 3rd. John Wiley & Sons, Ltd.

Venkitaraman, A.; Adiga, A. und Seelamantula, C. S. (2014): „Auditory-motivated Gammatone wavelet transform." In: *Signal Process*, 94(1) S. 608–619.

Veronesi, W. A. und Maynard, J. D. (1987): „Nearfield acoustic holography (NAH) II. Holographic reconstruction algorithms and computer implementation." In: *J Acoust Soc Am*, 81(5) S. 1307–1322.

Vetterli, M.; Kovacevic, J. und Goyal, V. K. (2011): *Fourier and Wavelet Signal Processing*. Online Publication under Creative Commons License.

Virag, N. (1995): „Speech enhancement based on masking properties of the auditory system." In: *Acoustics, Speech and Signal Processing (ICASSP), 1995 IEEE International Conference on*. Detroit, MI, USA, S. 796–799.

Virag, N. (1999): „Single channel speech enhancement based on masking properties of the human auditory system." In: *IEEE Transactions on Speech and Audio Processing*, 7(2) S. 126–137.

Vogel, C. R. (2002): *Computational Methods for Inverse Problems*. SIAM.

Vorländer, M. (1991): „Freifeld- und Diffusfeld-Übertragungsmaße von natürlichen Köpfen und von Kunstköpfen." In: *Acta Acust United Ac*, 74(3) S. 192–200.

Wang, D. und Lim, J. S. (1982): „The unimportance of phase in speech enhancement." In: *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 30(4) S. 679–681.

Warsitz, E. und Haeb-Umbach, R. (2007): „Blind Acoustic Beamforming Based on Generalized Eigenvalue Decomposition." In: *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(5) S. 1529–1539.

Wei, C.-W.; Tsai, C.-C.; Chang, T.-S. und Jou, S.-J. (2010): „Perceptual multiband spectral subtraction for noise reduction in hearing aids." In: *Circuits and Systems (APCCAS), 2010 IEEE Asia Pacific Conference on*. S. 692–695.

Weiss, M. R.; Aschkenasy, E. und Parsons, T. W. (1975): *Study and devlopment of the INTEL technique for improving speech intelligibility*. Tech. Rep. RADC-TR-75-108, Nicolet Scientific Corporation, Springfield, USA.

Welch, P. (1967): „The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms." In: *Audio and Electroacoustics, IEEE Transactions on*, 15(2) S. 70–73.

Wenzel, E. M.; Arruda, M.; Kistler, D. J. und Wightman, F. L. (1993): „Localization using nonindividualized head-related transfer functions." In: *J Acoust Soc Am*, 94(1) S. 111–123.

Werner, D. (2011): *Funktionalanalysis*. Springer-Lehrbuch, 7. Berlin, Heidelberg: Springer Berlin Heidelberg.

Wever, E. G. (1962): „Development of Traveling-Wave Theories." In: *J Acoust Soc Am*, 34(9B) S. 1319–1324.

Widrow, B. et al. (1975): „Adaptive noise cancelling: Principles and applications." In: *Proc of the IEEE*, 63(12) S. 1692–1716.

Wiener, N. (1930): „Generalized harmonic analysis." In: *Acta Mathematica*, 55(1) S. 117–258.

Wiener, N. (1942): *The Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications*. Tech. rep., MIT Radiation Lab.

Wiener, N. (1949): *The Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications*. New York: John Wiley.

Wightman, F. L. und Kistler, D. J. (1989a): „Headphone simulation of free-field listening. I: Stimulus synthesis." In: *J Acoust Soc Am*, 85(2) S. 858–867.

Wightman, F. L. und Kistler, D. J. (1989b): „Headphone simulation of free-field listening. II: Psychophysical validation." In: *J Acoust Soc Am*, 85(2) S. 868–878.

Wightman, F. L. und Kistler, D. J. (1992): „The dominant role of low-frequency interaural time differences in sound localization." In: *J Acoust Soc Am*, 91(3) S. 1648–1661.

Williams, E. G. (1999): *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press.

Wolfe, P. J. und Godsill, S. J. (2000): „The Application of Psychoacoustic Criteria to the Restoration of Musical Recordings." In: *AES 108th Convention*.

Wolfe, P. J. und Godsill, S. J. (2001): „Simple alternatives to the Ephraim and Malah suppression rule for speech enhancement." In: *Statistical Signal Processing, 2001. Proceedings of the 11th IEEE Signal Processing Workshop on*. Singapore, S. 496–499.

Womersley, R. und Sloan, I. (2001): „How good can polynomial interpolation on the sphere be?" In: *Adv Comput Math*, 14(3) S. 195–226.

Xie, B.-S. (2013): *Head-Related Transfer Function and Virtual Auditory Display*. 2nd ed. J. Ross Publishung.

Yang, X.; Wang, K. und Shamma, S. A. (1992): „Auditory representations of acoustic signals." In: *Information Theory, IEEE Transactions on*, 38(2) S. 824–839.

Yong, P. C.; Nordholm, S.; Dam, H. H.; Leung, Y. H. und Lai, C. C. (2013): „Incorporating multi-channel Wiener filter with single-channel speech enhancement algorithm." In: *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*. IEEE, S. 7284–7288.

Youla, D. C. und Webb, H. (1982): „Image Restoration by the Method of Convex Projections: Part 1 - Theory." In: *IEEE Transactions on Medical Imaging*, 1(2) S. 81–94.

Youngberg, J. und Boll, S. (1978): „Constant-Q signal analysis and synthesis." In: *Acoustics, Speech and Signal Processing (ICASSP), 1978 IEEE International Conference on*. S. 375–378.

Zelinski, R. (1988): „A microphone array with adaptive post-filtering for noise reduction in reverberant rooms." In: *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*. IEEE, S. 2578–2581.

Zheng, Y. R.; Goubran, R. A. und El-Tanany, M. (2004): „Robust near-field adaptive beamforming with distance discrimination." In: *IEEE Transactions on Speech and Audio Processing*, 12(5) S. 478–488.

Ziomek, L. (1994): *Fundamentals of Acoustic Field Theory and Space-Time Signal Processing*. CRC Press.

Zollner, M. und Zwicker, E. (1993): *Elektroakustik*. Springer Verlag.

Zölzer, U. et al. (2002): *DAFX - Digital Audio Effects*. John Wiley & Sons, Ltd.

Zotkin, D. N. und Duraiswami, R. (2004): „Accelerated Speech Source Localization via a Hierarchical Search of Steered Response Power." In: *IEEE Transactions on Speech and Audio Processing*, 12(5) S. 499–508.

Zotkin, D. N.; Duraiswami, R. und Davis, L. S. (2004): „Rendering localized spatial audio in a virtual auditory space." In: *Multimedia, IEEE Transactions on*, 6(4) S. 553–564.

Zotkin, D. N.; Duraiswami, R.; Grassi, E. und Gumerov, N. A. (2006): „Fast head-related transfer function measurement via reciprocity." In: *J Acoust Soc Am*, 120(4) S. 2202–2215.

Zotkin, D. N.; Duraiswami, R. und Gumerov, N. A. (2008): „Sound field decomposition using spherical microphone arrays." In: *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, S. 277–280.

Zotter, F. (2004): *Unterdrückung hörbarer Störgeräusche in Echtzeitsystemen*. Masterarbeit, University of Music and Performing Arts Graz, Graz.

Zweig, G.; Lipes, R. und Pierce, J. R. (1976): „The cochlear compromise." In: *J Acoust Soc Am*, 59(4) S. 975–982.

Zwicker, E. (1961): „Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen)." In: *J Acoust Soc Am*, 33(2) S. 248–248.

Zwicker, E. (1986): „A hardware cochlear nonlinear preprocessing model with active feedback." In: *J Acoust Soc Am*, 80(1) S. 146–153.

Zwicker, E. und Fastl, H. (1999): *Psychoacoustics, Facts and Models*. 2nd. Springer Verlag.

Zwicker, E.; Flottorp, G. und Stevens, S. S. (1957): „Critical Band Width in Loudness Summation." In: *J Acoust Soc Am*, 29(5) S. 548–557.

Zwicker, E. und Terhardt, E. (1980): „Analytical expressions for critical-band rate and critical bandwidth as a function of frequency." In: *J Acoust Soc Am*, 68(5) S. 1523–1525.

Zwislocki, J. (1946): „Über die mechanische Klanganalyse des Ohrs." In: *Cellular and Molecular Life Sciences*, 2(10) S. 415–417.

Zwislocki, J. (1948): *Theorie der Schneckenmechanik - Qualitative und quantitative Analyse*. Ph.D. thesis, Eidg. Technische Hochshule (ETH), Zürich, Switzerland.

Zwislocki, J. (1962): „Analysis of the Middle-Ear Function. Part I: Input Impedance." In: *J Acoust Soc Am*, 34(9B) S. 1514–1523.