Master's Thesis

Auditory Perception of Spatial Extent in the Horizontal and Vertical Plane

Marian Weger

Graz, March 2, 2016

Institute of Electronic Music and Acoustics

Graz University of Music and Performing Arts

Graz University of Technology

Advisor:
Msc. Ph.D. Georgios Marentakis

Assessor:
O.Univ.-Prof. Mag.art. DI Dr.techn. Robert Höldrich

Abstract

In this thesis, the extent to which spatial sound can be used to represent the horizontal and vertical spatial extent of auditory objects has been investigated. To this end, the perceived spatial extent of horizontally and vertically distributed sound sources has been measured as a function of two spatial extent synthesis algorithms, three stimulus types, and three different spatial distributions of individual loudspeakers.

Two spatial extent synthesis algorithms have been compared to each other. The first was a time-based algorithm which used a spatial distribution of individual grains of a granular synthesis stimulus to represent a spatially extended sound source. The second was a frequency-based algorithm in which the different spectral components of a monophonic input signal were mapped to different spatial locations.

Two of the stimuli were generated by a granular synthesis algorithm, on the on hand with sound material resulting in the impression of strong rain, on the other hand with Dirac impulses as grains, leading to a so-called impulse train stimulus. An additional white Gaussian noise stimulus was used, which was not processed by the time-based algorithm, but instead a spatial distribution of statistically independent noise sources was generated.

In the experiment, participants performed both absolute judgments of spatial extent and pairwise comparisons between representations of different spatial extent. Compared to the literature, smaller targets were used, which are seen as being more practicable for applications in human computer interaction.

Results indicate that the variations of perceived horizontal extent judgments varied systematically with physical extent for all stimuli used in the experiment. The time-based synthesis algorithm resulted in significantly larger judgments of spatial extent irrespective of orientation, compared to the frequency-based algorithm. Perception of vertical extent was not accurate and varied less systematically with actual extent, while judgments largely underestimated the actual vertical extent of sound sources. Finally, the results of the absolute judgments agreed with virtually all information contained in the time-consuming pairwise comparisons.

Kurzfassung

In dieser Arbeit wurde untersucht, inwiefern räumliche Klangwiedergabe dazu beitragen kann, einen Eindruck von horizontaler, sowie vertikaler räumlicher Ausdehnung von Schallereignissen zu erzeugen. Dazu wurde die wahrgenommene räumliche Ausdehnung in Horizontal- sowie Medianebene in Abhängigkeit von zwei Algorithmen, drei Typen von Stimuli, sowie drei verschiedenen räumlichen Verteilungen individueller Lautsprecher ermittelt.

Unterschiedliche Ansätze zur Synthese von räumlich ausgedehnten Schallquellen wurden miteinander verglichen: Einerseits ein zeitbasierter Algorithmus, der auf Granularsynthese basiert, und die einzelnen Grains auf unterschiedliche räumliche Positionen verteilt. Andererseits ein frequenzbasierter Algorithmus, bei dem komplementäre Frequenzbänder eines monophonen Eingangssignals räumlich verteilt werden.

Zwei der verwendeten Stimuli wurden durch einen Granularsynthese Algorithmus generiert: Einerseits mit Klangmaterial, welches zu einem Eindruck von starkem Regen führt, andererseits mit Dirac Impulsen anstelle von Grains, resultierend in einer zeitlichen Abfolge von Impulsen. Als dritter Stimulus wurde weißes Gaußsches Rauschen verwendet, wobei anstelle des zeitbasierten Algorithmus eine räumliche Verteilung von statistisch unabhängigen Rauschsignalen generiert wurde.

Im beschriebenen Experiment wurden separat sowohl absolute Schätzungen der wahrgenommenen räumlichen Ausdehnung, als auch Paarvergleiche zwischen Darstellungen unterschiedlicher räumlicher Ausdehnung durchgeführt. Im Vergleich zu früheren Studien wurden Schallquellen kleineren Maßstabs verwendet, wie sie für eine Anwendung im Bereich gestenbasierter Interaktion oder bei auditorischen Displays als sinnvoller erachtet werden.

Die Ergebnisse zeigen, dass die Verteilungen der Einschätzungen horizontaler Ausdehnung für alle Stimuli des Experiments systematisch mit der physikalischen Ausdehnung variierten. Der zeitbasierte Algorithmus führte, unabhängig von räumlicher Orientierung, zu signifikant größer wahrgenommener Quellbreite, als der frequenzbasierte Algorithmus. Die räumliche Audehnung in der Median-Ebene wurde von den Probanden stark unterschätzt und Einschätzungen der räumlichen Ausdehnung waren ungenau und variierten weniger systematisch mit der physikalischen räumlichen Ausdehnung. Die Ergebnisse der absoluten Schätzungen der wahrgenommenen räumlichen Ausdehnung stimmten mit nahezu sämtlichen Informationen, welche aus den Paarvergleichen gewonnen werden konnten, überein.

Acknowledgments

This work could not have been done without the support by many people to whom I would like to express my gratitude. Unfortunately I am unable to list all their names in detail. However, some of them had particular influence on the outcome of this thesis and stand representative for all those who are not specifically named.

First, I would like to thank my advisor Georgios Marentakis, who guided me through all stages of this thesis with great patience and dedication, and shared his knowledge and experience, especially in experiment design, statistical analysis, and academic writing.

The described investigation is only the result of a long process in which the central research questions were repeatedly modified and refined. During this period of orientation, especially David Pirrò and Gerhard Eckel contributed a lot of time and energy during several informal listening sessions and were always available to give valuable advice.

I would also like to personally thank Robert Höldrich for sharing his comprehensive knowledge and experience, and guiding me through the final stage of this thesis.

Last but not least, I would like to thank my family for their patience and unconditional support during the whole period of my studies.

Contents

List of Figures

List of Tables

1. Introduction

In our everyday acoustic environment, many perceived sounds reveal information about the spatial extent of their origin. In some cases, this information comes from an interpretation of the auditory event, based on acquired knowledge.

Two factors may correspond to spatial extent perception: spatial and non-spatial. The case where the physical dimensions of an imagined source object are guessed from sound characteristics is attributed to non-spatial or source-related factors of auditory perception, which even apply for monaural sounds. On the other hand, if the auditory event is based on an unfamiliar or synthetic sound which can not be intellectually connected to a corresponding physical object, an eventual impression of auditory spatial extent can only be perceived through spatial hearing.

This leads to the assumption that auditory spatial extent may be controlled by both spatial and non-spatial factors of the sound. At this point it is important to distinguish between perceived spatial extent of the auditory event itself, communicated through spatial factors, and the spatial extent of the physical object which produced this sound, which is perceived through non-spatial factors. In the context of this thesis, only the former is investigated. Auditory spatial extent is therefore understood as the impression of an auditory event incorporating a specific size and shape, in contrast to a point-shaped sound source without spatial extent. The term "perceived spatial extent" was proposed by [AS11] and may refer independently to width, height, and potentially also depth. Size describes the one-dimensional spatial extent, sometimes entitled as Auditory Source Width (ASW), spatial blur or spread.

Interactive environments for human-computer interaction (HCI) would benefit from the possibility of giving sound sources a specific spatial extent. However, it is not clear whether current algorithms can be used to create perceived spatial extent of a magnitude relevant for auditory displays, such as deictic interactions with sound, navigation, or spatialized audio graphs.

This could lead to more intuitive auditory virtual environments and also improved congruency in multimodal interaction. Some Virtual or Augmented Reality (VR/AR) applications could benefit by incorporating spatially extended sound sources to display spatially extended objects. At least in the horizontal plane

there is proof that spatial sound scenes which incorporate broad sound sources instead of point sources are generally preferred and judged as sounding more natural [PB03].

It was also shown that auditory feedback with spatial sound improves accuracy of users' gestures in multimodal eyes-free interaction [Bre+03]. Furthermore, eyes-free menu navigation using touch input and auditory feedback with spatial sound outperforms visual techniques after some practice [Zha+07]. Such systems could benefit of spatially extended sound sources, for instance to represent different item sizes.

Imagined are applications in smart rooms for the blind, as described in [Mül+14], in which physical objects give auditory feedback about their current position. As an improvement, objects could not only reveal their current location, but also show their spatial extent. Especially for large objects, like sofas or tables, this would probably lead to a better experience. Also mid-air direct interaction with virtual sound sources could benefit from spatially extended objects. In mobile audio-augmented reality applications, such as the Sound Garden [VAOB12], where different target locations are communicated acoustically, increasing spatial extent could be used to characterize decreasing distance to an auditory object in the proximity zone. Similarly in navigation aid systems with minimal attention interfaces [HMG02] or spatial auditory displays for visually impaired people [Mar+06] the sizes of obstacles could be presented through spatially extended sound sources. By incorporating a hear-through [ML14] or mic-through system [ALS11] the auditory display can be presented without hindering perception of the real environment at the same time.

This thesis is structured in six parts. After this introduction, a literature review on the auditory perception of spatial extent (Chapter 2) is performed. Based on the literature and informal listening tests, hypotheses were created. To evaluate those hypotheses, a controlled experiment is designed and implemented. This process is described in Chapter 3, while the results can be read in Chapter 4. Finally, conclusions are drawn in Chapter 5 and an outlook for the future is presented in Chapter 6.

2. Literature review

As already explained in the introduction, auditory perception of size or spatial extent is influenced by various spatial and non-spatial properties of the sound source. Many of these have been selectively tested in controlled experiments.

The majority of the discussed methods for the creation of auditory spatial extent have in common that a continuous spatially distributed sound source is constructed from a spatial distribution of multiple individual point-shaped sound sources. These sound sources can be identical, partially decorrelated or even completely uncorrelated to form the impression of a coherent auditory object.

2.1. Non-spatial factors affecting size perception

At first, some non-spatial factors for size perception are discussed. They have been tested with monaural sound playback, but are supposed to apply for all types of sound projection.

Likely connected to our personal experience, sounds with high sound pressure level are perceived to be bigger in comparison to softer ones. An increased perceived size with increasing loudness has been verified by [PB82] when headphones are used. The effect was confirmed but found to be weaker when using loudspeakers [CT03]. Similarly, perceived size has been found to increase with signal duration [PB82] in experiments using headphones. Finally, the presence of low frequencies in a signal is often associated with an increase in size when using monaural stimuli, and perceived auditory source width increases monotonically with decreasing stimulus base frequency [MBR05].

2.2. Spectral contributions to shape perception

A significant stream of research was directed towards the question of whether it is possible to identify the source size and shape based on the source spectrum. This question resulted from a series of mathematical publications, in which the singularity of the vibrational modes for one specific shape was investigated

[Kac66]. Although it was shown that spectra are not necessarily specific to geometry, isospectral companions are exceptional [GW96]. At least, there are no isospectral triangles [GM13].

These findings were experimentally investigated in different series of studies [KPT00]. Naive test persons listened to thin suspended plates of same thickness struck by a steel pendulum. In each experiment, three different shapes and materials (steel, wood and plexiglass) were presented.

In the first series, plates of square, non-square and very elongate rectangular shape with same surface area were used. The plates were hanging with the short side up, which means non-square plates were always higher than they were wide. Participants were asked to indicate either height or width by moving two parallel bars in the right positions. The results indicated good performance in shape perception. However, the perceived shapes were always smaller than the actual ones. Material affected the absolute but not the relative perceptual measures of height and width. Also mean perceived heights were always larger than mean perceived widths for the typical and long rectangles, regardless of material. It was concluded that listeners can discern the modal frequencies associated with width from those associated with height. However, as the experiment was performed with real physical plates, it can not be specifically rejected that additional cues from spatial hearing were used for the identification of plate dimensions. It seems obvious that at least for identification of the orientation spatial cues contributed to the decision.

In a second series, the same procedure was repeated with circular, triangular, and square rectangular plates of same surface area. The participants had prior knowledge of the available shape possibilities. For each stimulus, the participants identified the correct shape at a level well above chance. There was a tendency for participants to associate particular materials with particular shapes (wood with circle, steel with triangle, and plexiglass with rectangle). This association of particular materials with particular shapes was also shown by [GM06]. They argued towards a cognitive origin of this bias, which means the identification was at least partially based on an interpretation with reference to the everyday acoustic environment.

In a related experiment [CAKP98], listeners were asked to determine the length of different wooden rods, solely on the basis of the sound these produced when falling on a hard surface. Results indicated that different lengths could be distinguished in correct order, but the perceived length was somehow compressed, meaning short rods were perceived longer while long rods were perceived shorter than they actually were.

The aforementioned experiments led to the conclusion that in general it could

be possible to dictate at least the shape of a given vibrating plate from spectral cues alone. Since the base frequency is, however, dependent on the material, a size perception can only be achieved through additional interpretation based on acquired knowledge, or by additional spatial cues.

2.3. Spatial aspects of auditory extent perception

2.3.1. Binaural cues to Sound Localization

The primary cues for sound source localization in the horizontal plane are the interaural time difference (ITD) and interaural level difference (ILD) [Dic97, p. 118]. In the process of sound source localization, a first mechanism evaluates ITD for signal components below 1.6 kHz [Bla97, p. 164]. Afterwards, a second mechanism, on the one hand, evaluates ILD and on the other hand also interprets time shifts between the envelopes for frequency components starting at about 100 Hz [Bla97, pp. 164,173]. Roughly speaking, the second mechanism dominates for frequency components above 1.6 kHz, while the first mechanism has a stronger effect on frequency components below 800 Hz [Bla97, p. 173]. This implies a transition region in which both mechanisms are ambiguous.

This localization model is described in a simplified way by the Duplex Theory, going back to [Ray07], which implies a dominance of ITD for pure tones with low frequency, compared to a dominance of ILD for high frequencies. This theory applies in a similar way for low-pass and high-pass filtered noise, while for wideband stimuli both ITD and ILD have substantial influence on sound localization [MM02].

For sound sources in the vertical or median plane, there are no interaural signal differences. Here, localization is mainly achieved by direction-dependent filtering effects of the outer ear and upper part of the body, mathematically defined through the individual head-related transfer function (HRTF) of each person [AAD01; RB68; Bla97]. This mechanism can be explained through the extremely simplified concept of directional bands [RB68; Bla97]. It was shown that specific frequency bands are connected to specific directions in the median plane, which arises from direction-dependent resonances of the HRTF. This mechanism for localization in the median plane, however, is mainly based on acquired knowledge and requires a certain bandwidth of the source signal.

The just-noticeable difference (JND) for localization of white noise in the front direction goes down to 2-3° in horizontal and 8° in vertical orientation, and is strongly dependent on sound characteristics [Dic97, pp. 120-121; Bla97, pp. 41-44]. It must be stated that, however, for signals longer than approximately 300 ms,

slight head movements provide sufficient interaural signal differences for improving localization in the median plane [Bla97, p. 95].

2.3.2. Perception of spatially distributed sound sources

Concerning the perceived width of spatial sources, it is well known that interaural cross-correlation (IACC) is an important factor [MBR05].

Consequently, auditory source width is larger for uncorrelated compared to correlated signals, although even correlated signals from different locations, such as the ones used in panning algorithms, can result in an increased width perception [Fra13]. The author presented a reliable model for the prediction of auditory source width depending on the direction vectors and scalar gains of each loudspeaker. Additionally an offset was introduced to account for different room acoustics.

A significant number of studies investigated the perception of spatially distributed uncorrelated noise sources. Overall, it has been found that in general the perception of such a stimulus can but is not always associated with the impression of a single distributed sound source.

In addition, the emerging auditory event in the general case is much narrower in comparison to the real spatial distribution of the loudspeakers. When bandpass noise is used, it appears to be easier to manipulate perceived width when noise bands with high center frequency, corresponding to the region in which ILD perception dominates, are used [SP11; HKH02].

The probability that sound energy will be perceived to be uniformly distributed on the loudspeaker array increases with signal duration [HP08]. This was verified in an experiment in which nine loudspeakers were placed on a circle with a separation of 15° and emitted simultaneous uncorrelated white noise bursts. Increasing signal duration resulted in a more accurate perception of the spatial distribution of the loudspeaker area that resembled the uniform distribution more closely. The perception was mostly point-like when the signal length was under 10 ms. Until 80 ms the perceived spatial width increased with increasing signal length. The authors argue towards a signal length of 40-80 ms for perception of width to build up.

The perception of spatial gaps in the loudspeaker sources comprising the distributed source was investigated by [SP09] using uncorrelated pink noise bursts of 1 s duration. 13 loudspeakers were placed on a circle to form a 184° wide sound source. Different combinations of active loudspeakers were used, while the task was to distinguish which loudspeakers emitted sound. The results showed that small gaps in the sound source were not perceived accurately, while larger gaps

were perceived wider than they actually were, implying that a separation of more than 15° was required to result in a perceivable difference in gap. The spatially extended sound source was always perceived narrower than it actually was. In a succeeding publication [SP11] it was concluded that the perception of the spatial distribution of the sounds was inaccurate when more than three loudspeakers emit sound. In such a case, only the ends of the sound sources were perceived relatively accurately, while in between, perception was unreliable.

2.4. Decorrelation methods for spatial extent synthesis

When trying to create similar phenomena for arbitrary signals, ways are sought to artificially create decorrelated copies of a signal.

In stereophony, such decorrelation techniques are widely-used for the creation of a so-called pseudo stereo signal from a monophonic input signal. Two decorrelated versions of a monophonic signal are generated by two complementary comb-filters, usually implemented through a Lauridsen network [HL56] or all-pass filters [Bau69], and played back from different directions or channels [Sch58; Orb70a; Orb70b]. This comb-filter approach which works for both loudspeaker and also headphone listening, however, can introduce unwanted phasing effects [Ger92]. This method was developed not only as a pseudo-stereo effect but also for adding a specific size to a sound source [Ger92].

Another approach, proposed by [Ken95] incorporates so-called artifical decorrelation, which produces decorrelated versions of a monophonic input signal by applying random phase values for different frequencies. This way, the decorrelated signals differ only in the phase domain and retain an identical and unaltered amplitude spectrum. The approach was further improved by [PB04; BK04]. The pattern recognition experiment described in the previous section [PB03] showed poor performance in pattern recognition for auditory-spatial patterns drawn by decorrelated copies of a monophonic input signal using the artificial decorrelation method with randomized phases. A more recent enhancement used deterministic frequency dependent inter-channel time delays (ICTD), which were implemented as deterministic FIR and IIR allpass structures [Zot+11]. They showed that this approach was capable for controllable phantom source widening.

In contrast to decorrelation in the phase-domain a monophonic input signal can also be split into a number of unique frequency bands which are then distributed to different directions or channels. It was shown that this yields a reduced interaural cross correlation coefficient (IACC) and an increased auditory source width

[Ger92; BL86]. This approach has been found to be working for both panned as well as for real loudspeaker signals, and was also evaluated for loudspeaker arrays with more than two loudspeakers for the creation of auditory spatial extent [HP06a; Lai+12; PSP14].

The way frequencies are distributed on the array affects the perceived center and width of the spatially distributed sources. [HP06a] investigated the center and the perceived with of sounds generated with this technique. Signals were decomposed in frequency bands following the Equivalent Rectangular Bandwidth (ERB) scale. Groups of 9, 18 or 27 ERB bands were rendered sequentially on an array of 9 loudspeakers, so that each loudspeaker played one, two or three ERB bands respectively. In stimuli that contained low frequency ERB bands (below 1.3 kHz) but not for broadband signals, the perceived center of the distributed sound was found to correspond to the discontinuity in spatial frequency distribution that occurred when spatially rotating the frequency bands. Perceived width was found to be always lower than the loudspeaker area and larger for low frequency and broadband signals compared to high frequency signals. A subsequent study [HP06b] additionally revealed that the cases in which adjacent frequency bands were not in neighboring loudspeakers were perceived slightly wider than those in which the frequency bands were spatially placed in order.

A recent enhancement to this technique proposed the use of a Halton sequence [Hal64] to deterministically map frequency bands to directions [PSP14]. Even in this case, the spatialization algorithm was found to be strongly signal-dependent. Generally the approach seemed to work best for signals with wide frequency content and without sudden or impulsive events. Also signals could easily get implausible when spatially extended. Time varying the spatial distribution of frequencies was always less preferred compared to static distributions in pairwise comparisons [PSP14]. A further improvement through additional all-pass decorrelation filters seemed to provide better results but introduced stronger signal coloration.

2.5. Spatial grains

In computer music, a widely-used approach for the creation of auditory spatial extent is granular synthesis [Roa04]. Granular synthesis works by combining short signals (grains) and can provide stimuli usable for design purposes. The grains are usually generated from a single source file and are defined by individual micro-tonal properties. These include position in the source file, duration, envelope and pitch (controlled through playback speed) [Roa04; DS09].

Spatially distributed granular synthesis has been often used by composers and

sound artists to create the impression of spatially distributed sound sources. As grains are in general short and may be designed to have steep envelopes, they can be individually localized with high precision. An early version of this approach was described by [Tru98] who created decorrelated signals through granular synthesis which were then presented as a spatial distribution of point sources via individual loudspeakers. In a further improvement of this method, described by [Bar02], the individual grains themselves are mapped to distinct directions by using Ambisonics for sound spatialization. Another variation was proposed by [DS09], who described a way to synthesize non-point sources of specific shape through spatial grains. Their approach implies a spatial distribution of individual grains of a monophonic source signal. The grains are arranged on a continuous spatial trajectory which defines a specific shape, played back in an Ambisonics environment.

2.6. Auditory-spatial patterns

The perception of auditory spatial extent was also evaluated in a couple of pattern-recognition experiments, in which participants had to distinguish different auditory-spatial patterns.

[PB03] described an experiment, where participants had to distinguish between different spatial distributions of statistically independent noise signals representing simple shapes. The experiment was performed with different types of noise signals (white noise, low-pass noise, and high-pass noise), presented through 7 individual loudspeakers forming a two-dimensional planar array. It was shown that only white noise and high-pass noise provided satisfactory performance in pattern recognition.

Other pattern-recognition experiments were performed by [Lak93] with physical loudspeaker arrays and by [HF94] who used binaural rendering with headphones. Both used 16 sound sources arranged in a shape similar to a seven-segment display for acting as an auditory display for alphanumeric patterns. Patterns were "drawn" acoustically by sequentially feeding the active loudspeakers with a pulse signal in a rate of 120 ms or 8.3 Hz. In both cases the performance in pattern-recognition was well above chance. [Lak93] additionally showed that steeper attack and decay times improved pattern recognition ability. [HF94] additionally tried simpler geometric shapes, using binaural rendering with headphones, which resulted in no significant improvement in pattern recognition performance.

2.7. Model-based sound field synthesis

For sound synthesis based on physical modeling, it seems promising to not only generate the wave form of a musical sound, but also its spatial characteristics. This is especially interesting when using sound field synthesis, such as wave field synthesis (WFS). Both, physical modeling sound synthesis and WFS rely on physical models in the form of partial differential equations, which could be combined.

A proof-of-concept was shown by [MR09]. As an example, they used a physical model of a string and a connected sound board. The sound board is necessary in most musical instruments for impedance matching, which results in higher sound energy. For a simple sound board, a stiff piston model was used here. The velocity of the string was picked up at an arbitrary position and transferred to the center of the piston. The piston could be moved freely in space. Through the direct connection, the stiff piston vibrates in the direction normal to the surface. All points on the disk have the same phase and each of them is modelled as a point source. To be able to create the final driving signals for the loudspeakers, the velocity and sound pressure level at the piston surface was computed numerically and then delivered to the wave field synthesis engine. Informal listening tests showed that location, orientation and motion of the sound source were hearable.

Efforts have also been made to use such physical models and WFS to synthesize and control perceived spatial extent for arbitrary monaural signals. Ahrens and Spors [AS11] created exemplary physical models of a plate of finite size, vibrating in normal direction, and a pulsating sphere, both vibrating in higher modes, to simulate the corresponding sound field. Simulations of interaural coherence and informal listening tests confirmed, that the perception of spatial extent could be evoked by sound source models. However, the perceived spatial extent varies between sound source models: A pulsating sphere produces a sound field closely related to a monopole source, which limits its use for the creation of perceived spatial extent.

2.8. Summary

Some success was found in estimating the size and shape of objects, using acoustic information. However, it is not possible to differentiate whether this reflects the use of spectral or spatial information as these are confounded in the experiments in the literature.

Concerning the experiments which investigated spatially distributed sound sources, although it appears possible to create the impression of perceived spa-

tial extent, the perceived spatial extent seems to underestimate the actual spatial dimensions. This is also the case, if the frequencies comprising sound sources are spatially distributed. For spatial distributions of synthetic noise signals, perceived width increases with increasing bandwidth and center frequency [SP11].

Artificial decorrelation through phase-processing provided good results in phantom-source widening. However, this approach did not work well in a pattern-recognition task with spatially distributed sound sources [PB03].

The results of [SP09] and [PSP14] suggest that dense distributions of uncorrelated noise as well as spatially distributed sound sources are perceived as one coherent sound source, slightly narrower than the actual width. However, the method has only been evaluated in the horizontal plane and with loudspeaker arrays extending up or around 360° around listeners.

Spatially distributed grains seem to be a promising approach for the creation of auditory spatial extent. This method provided good results in artistic applications in computer music as well as in controlled experiments for pattern recognition.

The findings of [HF94] show, that there is a chance that loudspeaker-based algorithms for the creation of spatial extent could also be transferred to binaural rendering using headphones. This would lead to applications for mobile and augmented reality devices.

2.9. Discussion

Go once more through this section and gather the arguments that belong together in the same paragraph or part of the text}

The literature review opens several questions and provides a basis for different hypotheses to be drawn.

First, there is insufficient information on the auditory perception of spatial extent in the median plane. At the time of writing no controlled experiment exists which addresses this question. The possibility of creating more complex shapes is seen to be dependent on the capability of synthesizing the impression of a specific spatial extent in vertical orientation.

As it was shown that spatial distributions of unique frequency-bands of a monophonic input signal are capable of producing perceived spatial extent in the horizontal plane, it is straightforward to try this method for generating vertically extended sound sources. However, since sound localization in the median plane works different than in the horizontal plane, the performance of this method in the median plane is not assured and needs verification through a controlled experiment. The described approach of frequency-dependent mapping to spatial

positions will be referred to as the frequency-based algorithm in the following chapters of this thesis.

The method of spatial grains, on the other hand, is supposed to be a promising candidate for the synthesis of both horizontally and also vertically extended sound sources. This assumption is grounded on the successful applications in computer music and also pattern-recognition experiments. However, no controlled experiments exist which address the synthesis of auditory spatial extent by using this type of algorithm. It is further supposed that the observation that steeper attack times led to better performance in pattern recognition [Lak93] also apply for spatial extent perception in both horizontal and vertical plane. In the following chapters of this thesis, this approach of spatial grains will be named the time-based algorithm.

It is argued here that a similar sequential playback of grains following a spatial trajectory, as described by [Lak93], could also be achieved by a single moving sound source of a granulated stimulus. However, this is just mentioned for completeness and is not further investigated.

Concerning the spatial distributions of statistically independent noise sources, [PB03] obtained good results in a pattern recognition task. The perception of such signals in the horizontal plane has also been well studied in controlled experiments. It is supposed that such spatial distributions of uncorrelated noise signals could also create the impression of vertical spatial extent.

Most investigations in the literature have been done with relatively large spatial distributions. A desired application in human computer interaction (HCI) will require much smaller spatial distributions. It is not known if the mentioned approaches for the creation of auditory spatial extent are suitable for such small-scale implementations.

Some of the approaches described in the literature review have also been confirmed in a series of informal listening tests. Based on these personal observations it was assumed that it was easier to distinguish between two different amounts of spatial extent than dictating the absolute spatial extent of a given sound source. This hypothesis needs verification through a controlled experiment.

Concluding, the literature review led to several questions and hypotheses which need verification through a controlled experiment.

3. Experiment Design and Implementation

The main hypotheses which were articulated above served as a basis for a controlled experiment, which is described in the current chapter. This experiment is designed to verify those hypotheses. The central research question was to understand the amount to which spatial sound can be used to provide information relating to the horizontal and vertical spatial extent of auditory objects. Compared to the literature, smaller targets were used, which are seen as being more applicable for applications in HCI.

3.1. Stimuli

Three types of stimuli were chosen: white Gaussian noise, a stimulus based on granular synthesis, and also a so-called impulse train constructed by a series of Dirac impulses. All sound synthesis was implemented with the graphical programming environment Pure Data (Pd)[1].

3.1.1. White noise

White Gaussian noise was included as the typical signal of choice in such experiments. White noise was preferred to pink noise as it seemed to produce better results in informal pilot tests as well as in [PB03].

3.1.2. Granular synthesis

Based on literature and informal listening tests, a typical granular synthesis algorithm [Roa04] was chosen and implemented in Pd under usage of the Universal Polyphonic Player (UPP)[2] library.

During informal listening tests, different kinds of parameter values and also sound material for the grains were compared. For the experiment, it was decided

[1] Pure Data: `https://puredata.info/`
[2] Universal Polyphonic Player (UPP): `http://grrrr.org/research/software/upp/`

to use a stimulus resulting in the impression of strong rain, which was generated using 48 isolated rain-drop samples, extracted from a recording of rain, and normalized to the same amplitude. Average duration of the grains was 46 ms (standard deviation SD = 18 ms) with an approximate attack time of 2.2 ms (SD = 1.8 ms). To obtain the mentioned attack times from the individual samples, in a first step the discrete-time analytic signal (Hilbert transform) was computed, which describes the envelope of a signal. In a second step, the maximum of this envelope was determined for each individual grain. Finally, the time for each grain's envelope to reach the lowest of all computed envelope maxima was interpreted as the attack time.

The individual grains were played back in a temporal sequence to jointly generate the rain sound, while the next grain was drawn from a uniform distribution. The onset of the next grain relative to the beginning of the currently active grain (inter-onset interval) was sampled from a normal distribution with M = 10 ms and SD = 3 ms. This results in a mean onset frequency of 100 Hz. Occasional negative time delays were mirrored to avoid increasing the probability of a zero delay. This randomized delay helped to avoid a pitched sound when grains were played back in a rate inside the audible frequency range.

The chosen onset frequency is much higher than the one used in the literature by [Lak93; HF94] (100 Hz compared to 8.3 Hz). During informal listening tests, this frequency of 100 Hz was found to be a good compromise between the impression of a coherent spatial distribution of grains, generating a plausible impression of heavy rain, and individually localizable sonic events.

3.1.3. Impulse train

Based on the observation by [Lak93], that steeper attack times led to better performance in pattern recognition, it is assumed that this also applies for the granular synthesis stimulus. It is argued here that grains with maximum steepness, e.g., Dirac impulses, could serve as a good benchmark.

The impulse train stimulus can be seen as a special case of the granular synthesis, with grains of maximum steepness and minimum duration. It is implemented just like the granular signal, but with a Dirac impulse as a grain, played back in a sequence. Such impulse trains are commonly used in spatialization experiments and also assistive technology literature.

3.2. Algorithms

The signals described above were processed by two types of algorithms: a time-based (TB) algorithm, which maps different temporal parts of the signal to different locations, and also a frequency-based (FB) algorithm, which maps different frequency-bands to different locations. The white noise stimulus is treated as an exception and is not processed by the time-based algorithm but instead each active loudspeaker plays statistically independent white Gaussian noise. However, it is still assigned to the time-based algorithm to simplify the condition names of the experiment. A schematic overview of the complete signal chain for white noise and impulse train stimuli is shown in Figure 3.1.

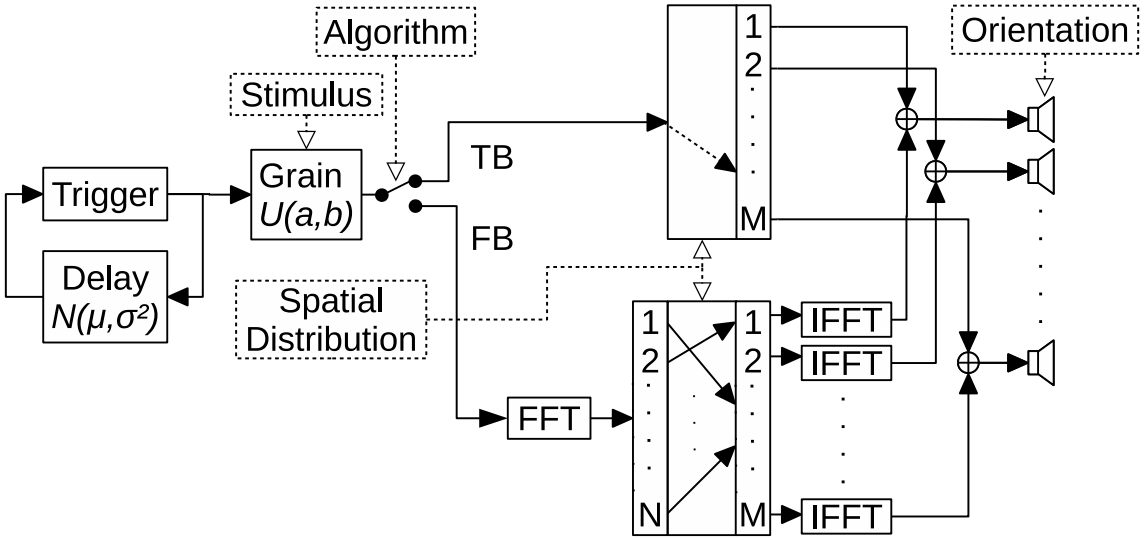


Figure 3.1.: Simplified block diagram of the signal chain for impulse train and granular synthesis stimuli. The independent variables of the experiment are framed in dotted lines.

Scenarios of different physical extent were generated depending on the combination of the used algorithm, stimulus type, and spatial distribution of the loudspeakers. Small, medium, and large spatial distributions were created using 3, 7 and 11 adjacent loudspeakers in both horizontal and vertical orientation.

3.2.1. Frequency-based spatialization

In case of the frequency-based spatialization, the monophonic input signal was decomposed into frequency-bands following the equivalent rectangular bandwidth (ERB) [GM90] scale, according to the algorithm proposed in [HP06a].

The approximation forumula from 1990 by Moore and Glasberg gives the equivalent rectangular bandwidth at a given center frequency [GM90]:

$$ERB = 24.7 \cdot (0.00437 \cdot f + 1) \tag{3.1}$$

Despite the fact that this formula is only valid for moderate sound levels and frequencies between 100 Hz and 10 kHz, it is used for the whole audible frequency-range.

The formula can be rearranged to get the value on the ERB-scale (ERBS) for a given frequency, which equals the number of equivalent rectangular bandwidths below that frequency [MG96]:

$$ERBS = 21.4 \cdot log_{10}(0.00437 \cdot f + 1) \tag{3.2}$$

Both formulas together deliver the center frequencies and bandwidths for the ERB-filters used in the experiment. 38 ERB-bands with center frequencies from 142.5 Hz to 19.7 kHz (see Table 3.1) were chosen. Bands below were considered unnecessary because the output was anyway filtered starting at 200 Hz by using two cascaded 1-pole high-pass filters to protect the loudspeakers from harm. The ERB-bands, however, start below 200 Hz to allow a similar low-frequency roll-off as with the time-based spatialization algorithm. The highest band includes 20 kHz to be sure that no audible information is lost.

Table 3.1.: ERB-scale values, center frequencies and bandwidths of the bandpass-filters used in the experiment.

Nr.	ERBS	f_c [Hz]	ERB [Hz]
1	4.5	142.5	40.1
2	5.5	184.7	44.6
3	6.5	231.7	49.7
4	7.5	284.0	55.4
5	8.5	342.3	61.6
6	9.5	407.1	68.6
7	10.5	479.4	76.4
8	11.5	559.8	85.1
9	12.5	649.4	94.8
10	13.5	749.2	105.6
11	14.5	860.3	117.6
12	15.5	984.0	130.9
13	16.5	1,121.8	145.8
14	17.5	1,275.3	162.4
15	18.5	1,446.1	180.8
16	19.5	1,636.4	201.3
17	20.5	1,848.3	224.2
18	21.5	2,084.3	249.7
19	22.5	2,347.0	278.0
20	23.5	2,639.6	309.6
21	24.5	2,965.5	344.8
22	25.5	3,328.4	384.0
23	26.5	3,732.5	427.6
24	27.5	4,182.5	476.1
25	28.5	4,683.6	530.2
26	29.5	5,241.6	590.5
27	30.5	5,863.1	657.6
28	31.5	6,555.1	732.3
29	32.5	7,325.7	815.4
30	33.5	8,183.9	908.1
31	34.5	9,139.6	1,011.2
32	35.5	10,203.9	1,126.1
33	36.5	11,389.0	1,254.0
34	37.5	12,708.8	1,396.5
35	38.5	14,178.5	1,555.1
36	39.5	15,815.2	1,731.8
37	40.5	17,637.8	1,928.5
38	41.5	19,667.4	2,147.6

The filters were designed with short-time Fourier transform (STFT), performed with Fast Fourier Transform (FFT). The implementation was based on an FFT size of 1024 samples, with Hann-window and 75% (4×) overlap. The FFT size of 1024 was assumed to provide a reasonable compromise between time- and frequency-resolution for the described application [PSP14]. Window type and overlap were chosen to yield perfect reconstruction [Roc03, p. 113]. The frequency-responses of these 38 individual ERB-filters are shown in Figure 3.2. These filters have strong side-lobes, but they are easy to implement, provide great efficiency, and were also used by [PSP14]. The three different magnitude steps for the bands below 1 kHz emerge from the bad frequency resolution of 1-3 frequency bins per band in that range, which results in filters of low steepness. However, in the relevant frequency range above 200 Hz they sum up to approximately zero dB.

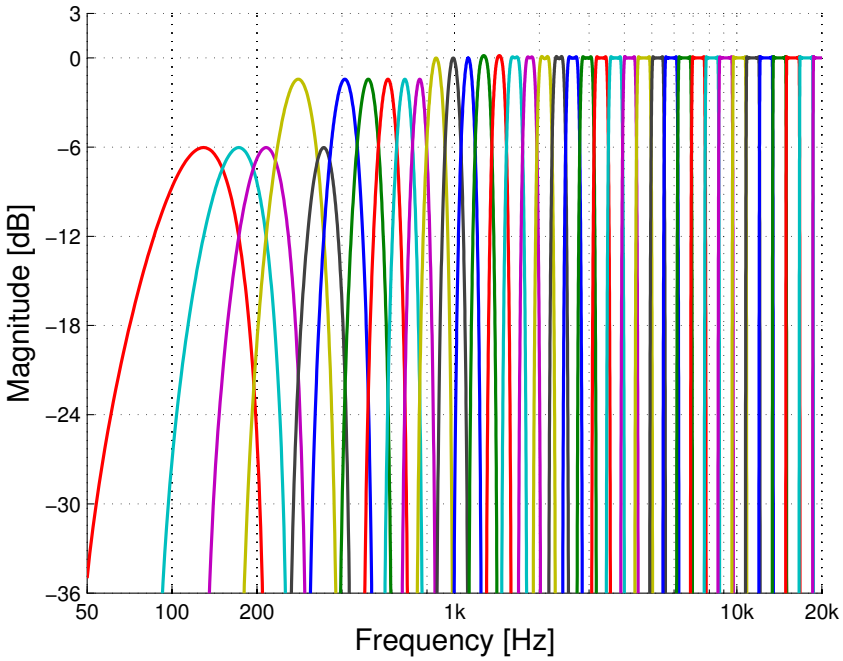


Figure 3.2.: Frequency response of the 38 ERB-filters used in the experiment.

To make sure that the ERB-bands are evenly distributed to all active loudspeakers, and to ensure a reproducible distribution, the output channel of each frequency band was chosen by using a Halton sequence [Hal64], as proposed by [PSP14]. In particular, a long Halton sequence of base 2 (without offset) was precomputed. All elements (having values between '0' and '1') were then multiplied by the maximum amount of active loudspeakers (11) and rounded upwards to the next integer, to represent a sequence of loudspeaker indices. Direct repetitions were removed as well as numbers which would induce a bias towards one loudspeaker being used more often than the others. The first 38 sequence numbers indicated the channels in which the ERB-bands would be rendered.

The other way around, loudspeaker channels and the corresponding ERB-bands are shown in Table 3.2 for small and medium, and Table 3.3 for large spatial distributions respectively. The loudspeakers are numbered from left to right for horizontal and from top to bottom for vertical orientation. Loudspeaker 6 lies in the center and is therefore active in all conditions of the experiment.

Table 3.2.: Mapping of bands to output channels for small (left) and medium (right) spatial distribution.

5	**6**	**7**
2	1	3
4	5	6
7	8	9
10	12	11
14	15	13
16	18	17
20	21	19
23	24	22
26	27	25
28	30	29
32	33	31
34	36	35
38		37

3	**4**	**5**	**6**	**7**	**8**	**9**
4	2	6	1	5	3	7
8	12	10	9	13	11	14
15	17	20	16	19	18	21
22	26	24	27	23	25	28
29	31	34	30	33	32	35
36		38		37		

Table 3.3.: Mapping of bands to output channels for large spatial distribution.

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
8	4	2	9	6	1	5	10	3	7	11
12	16	21	14	18	13	20	17	15	19	22
23	27	25	32	29	24	31	28	26	30	33
34	38		36			35		37		

These sequences led to the complementary filters for the output channels of the three different spatial distributions, which are shown in Figure 3.3 for small (3 channels), Figure 3.4 for medium (7 channels), and Figure 3.5 for large spatial distributions (11 channels).

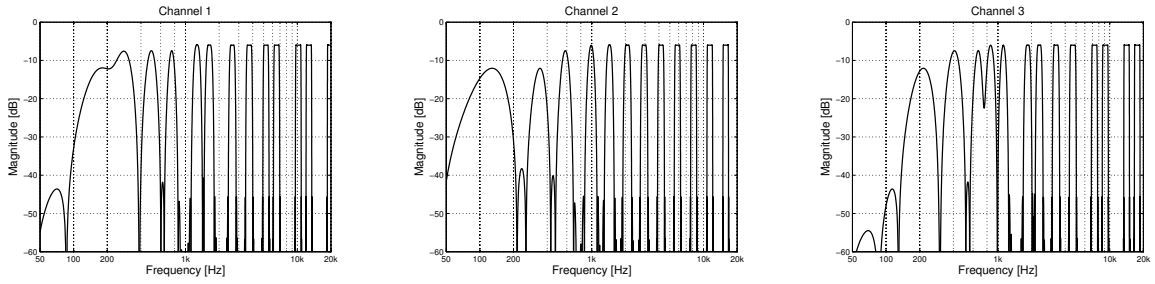


Figure 3.3.: Individual filters for the output channels for small spatial distribution (3 loudspeakers).

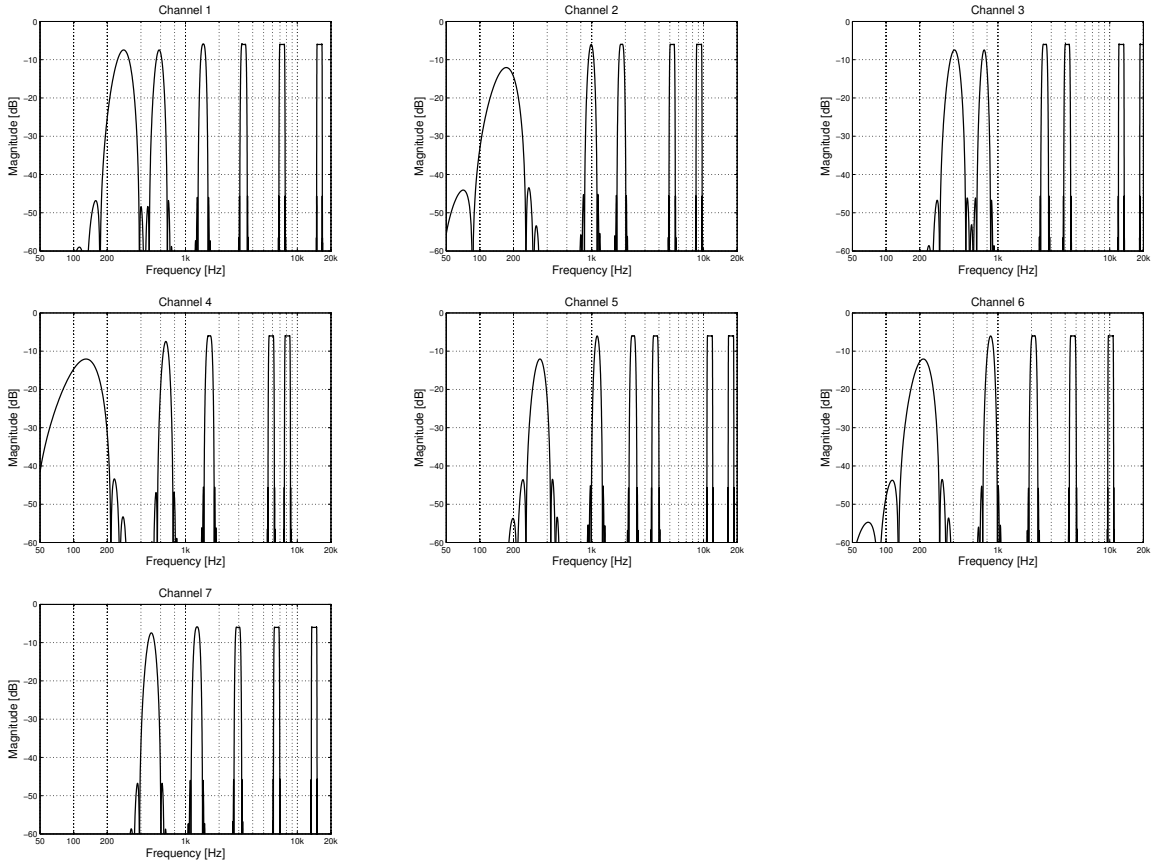


Figure 3.4.: Individual filters for the output channels for medium spatial distribution (7 loudspeakers).

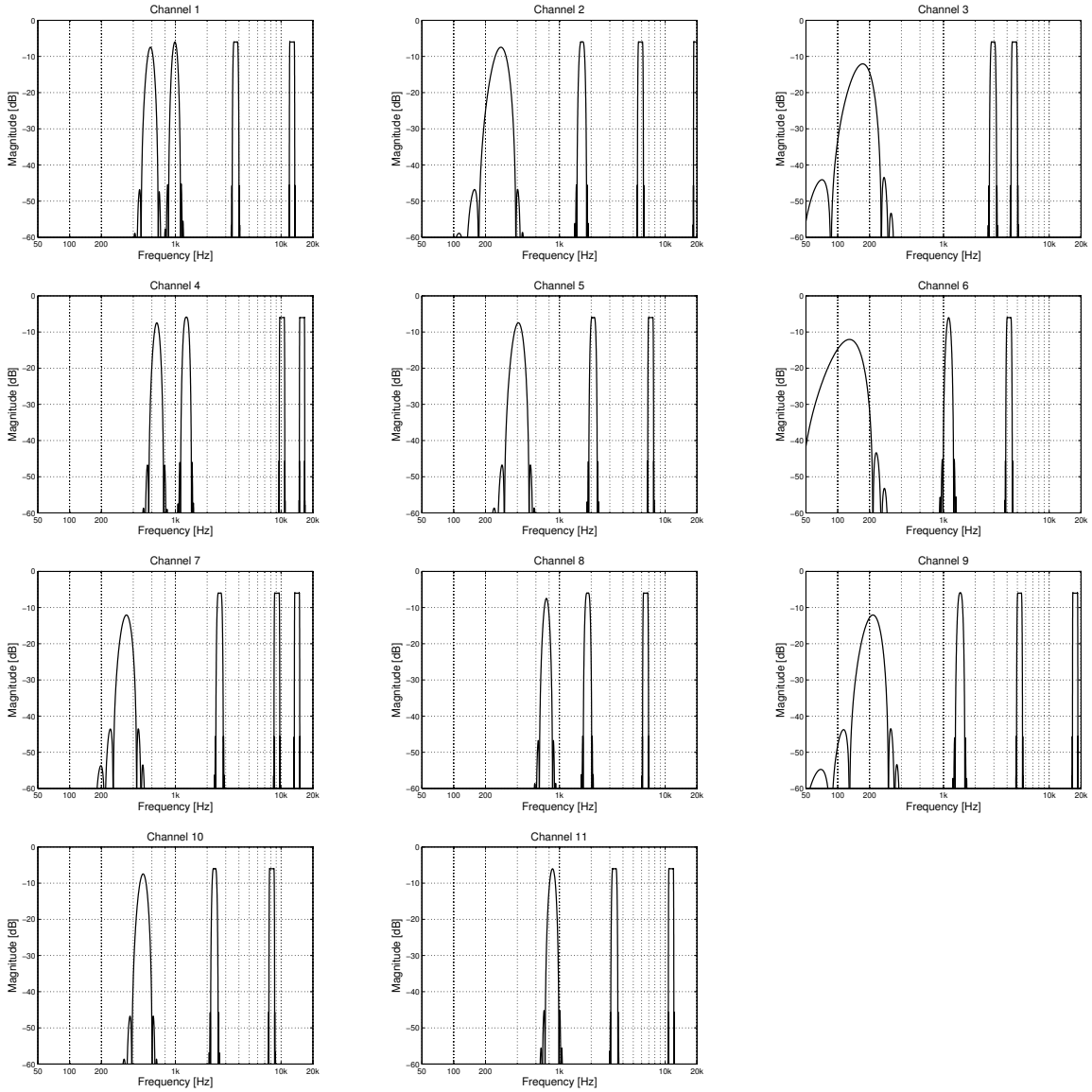


Figure 3.5.: Individual filters for the output channels for large spatial distribution (11 loudspeakers).

The question which arises at this point is how balanced the spatial distributions created by those channel-filters will be. Parseval's Theorem shows that the total energy of a signal can be expressed by the sum of spectral power across frequency [OS09]. The total power of each individual channel can therefore be computed directly from the corresponding channel-filter. However, a perceptual weighting of this power is required to take account of the frequency-dependent auditory perception of the listener. A simplified perceptual weighting is performed through A-weighting [IEC13].

Under consideration of the high-pass filtering and A-weighting [IEC13], the channel-filters results in a theoretical power per loudspeaker, which is shown in Figure 3.6 for all three spatial distributions. For each spatial distribution, the in-

dividual channels in Figure 3.6 sum up to 0 dB. This reference level, however, is an arbitrary decision. It must be noted here that despite the implementation of FFT and IFFT as a transformation with perfect reconstruction, the complementary filters do still not add up to a signal identical to the input signal. This is due to the omitted bands in the low-frequency range, including the DC offset. Furthermore, considering the playback through different loudspeakers under non-optimal acoustic conditions, depending on the listening position the sound waves anyway add up to a signal which is strongly different to the monophonic input signal.

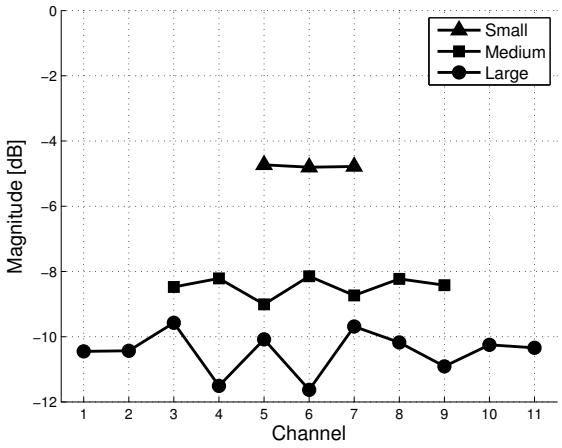


Figure 3.6.: Total power per speaker for the three spatial distributions, under consideration of high-pass filtering and A-weighting.

Even for the worst case of 11 loudspeakers with the large spatial distribution, in which only few ERB bands remain for each channel, there is a maximum power difference of approximately 2 dB between the channels. It is argued here that this provides a well-balanced power distribution which is sufficient for the perception of an evenly spread spatially distributed sound source.

This assumption, however, only applies for broadband signals with high spectral flatness. While white noise and impulse trains both have a flat magnitude spectrum, the spectrum of the rain-sound generated with granular synthesis lacks both low and also high frequencies (see Figure 3.2.1). The plots show a power spectral density estimate for both white gaussian noise and also for the rain stimulus based on granular synthesis. Both spectra include the two cascaded first order high-pass filters.

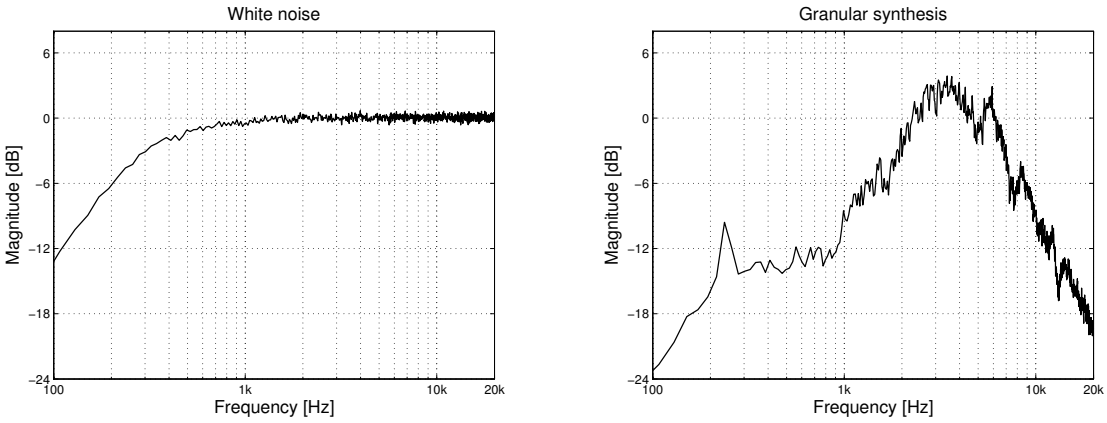


Figure 3.7.: Power spectral density estimate of white noise (left) and granular rain stimulus (right). Both including the two cascaded first order high-pass filters.

3.2.2. Time-based spatialization

The time-based spatialization algorithm works only for grains (including impulses). Each grain is mapped individually to one of the active loudspeakers by using a similar sequence as with the frequency-based algorithm. As described in Section 3.2.1, a long Halton sequence of base 2 (without offset) was precomputed. All elements were then multiplied by the maximum amount of loudspeakers (11) and rounded upwards to the next integer. The final sequences were then generated by taking only the first appearance of every number corresponding to an active channel for the three different spatial distributions (see Table 3.4). The sequence lengths corresponded to the number of active loudspeakers and each loudspeaker number was included once. The fixed sequence for a given spatial distribution was then used to indicate the loudspeaker from which the next grain or impulse would be played, and was applied repeatedly.

Table 3.4.: Pre-computed sequences of the time-based algorithm for the different spatial distributions.

Small	6	7	5								
Medium	6	3	9	7	5	4	8				
Large	6	3	9	2	7	5	10	1	4	8	11

3.3. Apparatus and Materials

In the described experiments, spatially extended sound sources are constructed from spatial distributions of point sources. Each point source is represented by a dedicated physical loudspeaker to provide best possible conditions for sound localization. In future applications these loudspeakers could be replaced by virtual sound sources through sound field synthesis (SFS) techniques such as wave field synthesis (WFS) or Ambisonics, or binaural rendering with headphones.

The loudspeaker array used to create the different spatial distributions in the experiment consisted of 21 custom 2-inch broadband speakers (Peerless PLS-P830983 in closed box) [Bla11]. According to the data sheet the speakers have an effective cone diameter of 4.4 cm [Tym15]. They were arranged in a horizontal and a vertical line of 11 loudspeakers, sharing the center speaker which was placed at a height of 120 cm, roughly corresponding to height of the participant's ears (see Figure 3.8).

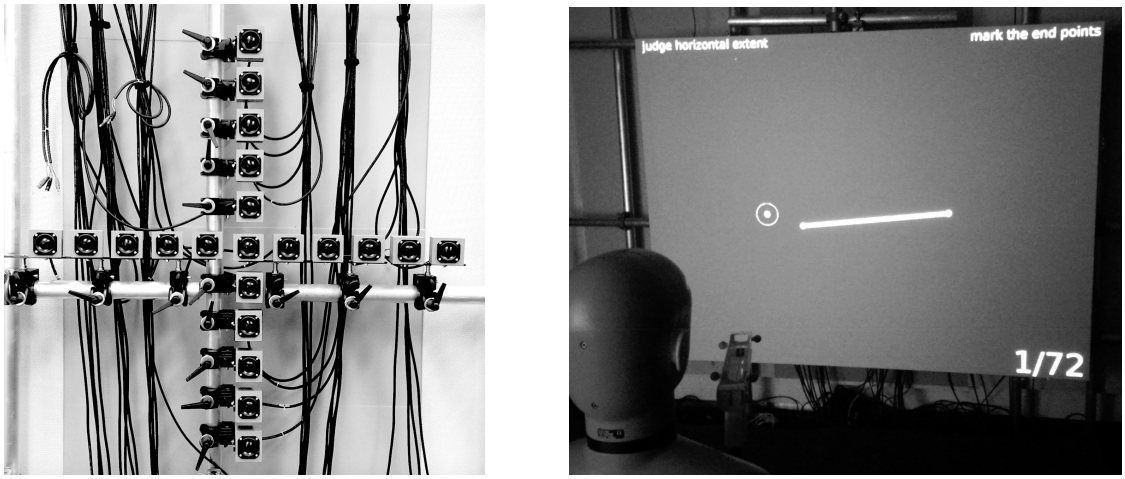


Figure 3.8.: The planar loudspeaker array (left) and the apparatus (right).

The distance between two neighboring loudspeakers was always 10 cm, which results in a maximum distance of 1 m between the two outmost speakers (measured from the membrane centers). Each loudspeaker was aligned to the direction of the listener. Differences in distance to the listener were compensated by individual gain and delay corrections. Loudspeakers were driven by custom 15-W class-D amplifiers (based on Texas Instruments TPA3122) [Bla11], which were connected to Behringer ADA8000 DA-converters running at 44.1 kHz / 24 bit. As already mentioned, all output channels were high-pass filtered at 200 Hz to protect the speakers from harm.

Loudspeakers were hidden using an acoustically transparent projection screen[3] with dimensions of 2 by 1.5 m, which was installed 10 cm in front of the speaker array. A projector was used to display instructions and additional information/graphics, depending on the procedure (absolute or relative judgments), on the projection screen. For correct projection mapping, the screen carried infrared reflective markers to be tracked by an optical motion capture system consisting of 9 NaturalPoint OptiTrack™ Flex13 cameras. The tracking information from the Motive:Body software was sent via Open Sound Control (OSC)[4] to Pd, in which the graphics and logic for the experiment were processed. The graphics and projection mapping were implemented using the Extended View Toolkit [Ven+11]. Both Motive:Body and Pd were running on the same PC in Microsoft Windows 7. All sound synthesis was running on a second PC in Debian GNU/Linux 8.

Participants sat on a chair, facing the loudspeaker array in 2 m distance. According to the approximate physical dimensions of the loudspeakers, one loudspeaker was assumed to roughly form a sound source source of 5 cm width and height. The three different spatial distributions therefore led to approximate physical spatial extents of 25, 65, and 105 cm, or 7.2°, 18.5°, and 29.4° at the listening position. Each stimulus was set to an Equivalent Continuous Sound Level (Leq) of 55 dBA at the listening position (averaged over 10 s; measured with NTi M2210 microphone and XL2 analyzer).

The experiment took place in an acoustically treated room of 4.33 × 6.20 × 3.43 m (width × length × height) size.

3.4. Procedure

During the experiment, participants performed judgments of spatial extent in 36 conditions that manipulated Algorithm (frequency- vs. time-based), Spatial Distribution (small vs. medium vs. large), Orientation (vertical vs. horizontal) and Stimulus (white noise vs. impulse train vs. granular synthesis) (see Table 3.5).

[3] Gerriets OPERA® white perforated (PVC, 390 g/m^2, 7 percent perforation area)
[4] Open Sound Control: `http://opensoundcontrol.org/`

Factor	Levels
Spatial Distribution	small medium large
Algorithm	frequency-based time-based
Stimulus	white noise impulse train granular synthesis
Orientation	horizontal vertical

Table 3.5.: The independent variables in the experiment.

As informal listening tests suggested that relative judgments of spatial extent may be easier than absolute judgments, both absolute and relative judgments were performed in two separate parts of the same experiment. Participants always performed the relative judgments first, because this task was easier to understand and helped accomplishing the second part in which absolute judgments were required. Both parts used the same apparatus, but required different procedures which are explained below.

3.4.1. Relative Judgments

Through pairwise comparison in a two-alternative forced choice (2AFC) task two different variations were presented one after another. The participant had to dictate the stimulus with larger perceived spatial extent by pressing either the first or the second button on the hand-held device (Nintendo Wii Remote™, see Figure 3.9. In each pair, only one of the independent variables Algorithm, Stimulus, or Spatial Distribution was changed, while different pairs were generated by manipulating the other variables. All resulting pairs were presented in both orientations (horizontal and vertical). This led to 6 separate parts, which were counter-balanced between participants with a latin square. As each pair was presented four times (twice in reverse order), each part consisted of 72 (when alternating Stimulus or Spatial Distribution) or 36 (when alternating Algorithm) pairs of stimuli, which were presented in randomized order.

Figure 3.9.: Nintendo Wii Remote™ as used in the pairwise comparisons of the experiment.

Each of the two signals was presented for 750 ms (including 5 ms linear fade-in and fade-out), with a pause of 400 ms between them. The pause was chosen to be just long enough to ensure that the short-term echoic memory would reset [Cow84]. After the second signal and a pause of 850 ms, the whole pair was presented again in an endless loop. When participants had made a choice, after a pause of 800 ms the next pair was presented.

After each part participants were able to take a short break. Before each part, short instructions were projected on the screen, including the information if horizontal or vertical spatial extent had to be judged. During the judgments, this information remained visible at the screen edges, with additional information of remaining judgments for the current part.

3.4.2. Absolute Judgments

After a short break, participants proceeded to the absolute judgments of perceived spatial extent. They went through the 18 signal combinations in a randomized order while being asked to highlight the perceived spatial extent on the projection screen. The participant was equipped with a toy gun (Nintendo Wii Remote™ controller) to indicate the beginning and end point of the perceived sound source.

The gun was equipped with infrared-reflective markers for the optical motion tracking system. An aiming point was permanently shown on screen. By triggering the gun, participants were able to place a permanent straight line through indicating its end points. Afterwards it was possible to correct the entry via drag-and-drop or by clearing all input to restart drawing. When finished, the participant could press another button to proceed to the next trial. Similar to the relative judgments, a pause of 500 ms was inserted between two signals to ensure that

the short-term echoic memory would reset [Cow84]. Each signal was presented continuously, with a 5 ms linear fade-in and fade-out. Horizontal and vertical orientation were tested in separate parts which were presented in counterbalanced order, alternating between participants. Each combination of Algorithm, Stimulus and Spatial Distribution was presented four times. Altogether this sums up to a total of 72 signals per part.

3.5. Participants

18 participants with mixed background, age (M = 26.3 years, SD = 5.4 years), and gender (5 female, 13 male) were tested and received a small financial compensation. None of them had prior knowledge or training in the specific task and all participants reported normal hearing. The task was not restricted in time. Before starting the experiment, participants had to sign a consent form on the utilization of their data (see Appendix A). They were also given detailed instructions in written form (see Appendix B) and through an additional oral explanation. Participants were allowed to pause the experiment to ask questions regarding their task.

4. Results

The results of the experiment are presented separately for the two different tasks. The results of the absolute and relative judgments are presented in Sections 4.1 and 4.2, respectively. Afterwards, both results are compared to each other in Section 4.3. The concluding Section 4.4 of this chapter describes several acoustic measurements, which were performed in order to explain some of the results and verify their validity.

4.1. Results of absolute judgments

4.1.1. Perceived spatial extent

Perceived vs. physical extent in the different conditions in the experiment can be seen in Figure 4.1. The error bars, in this context, indicate the standard error of the mean (SEM) which is defined as the sample estimate of the population standard deviation divided by the square root of the sample size [Bar07, p. 678].

It is evident that overall perceived spatial extent in vertical orientation was smaller than the one in horizontal orientation. In addition, it appears as if the time-based algorithm is outperforming the frequency-based algorithm in this experiment: On the one hand, the time-based algorithm led to larger perceived spatial extent than the frequency-based algorithm. On the other hand, while for the time-based algorithm in horizontal orientation, larger physical extent led to larger perceived spatial extent, this only appears as a small tendency for the frequency-based algorithm in horizontal orientation. Finally, although the results are similar for the white noise and impulse train stimuli, perceived spatial extent for the granular synthesis stimulus appears to be smaller in comparison to both other stimuli.

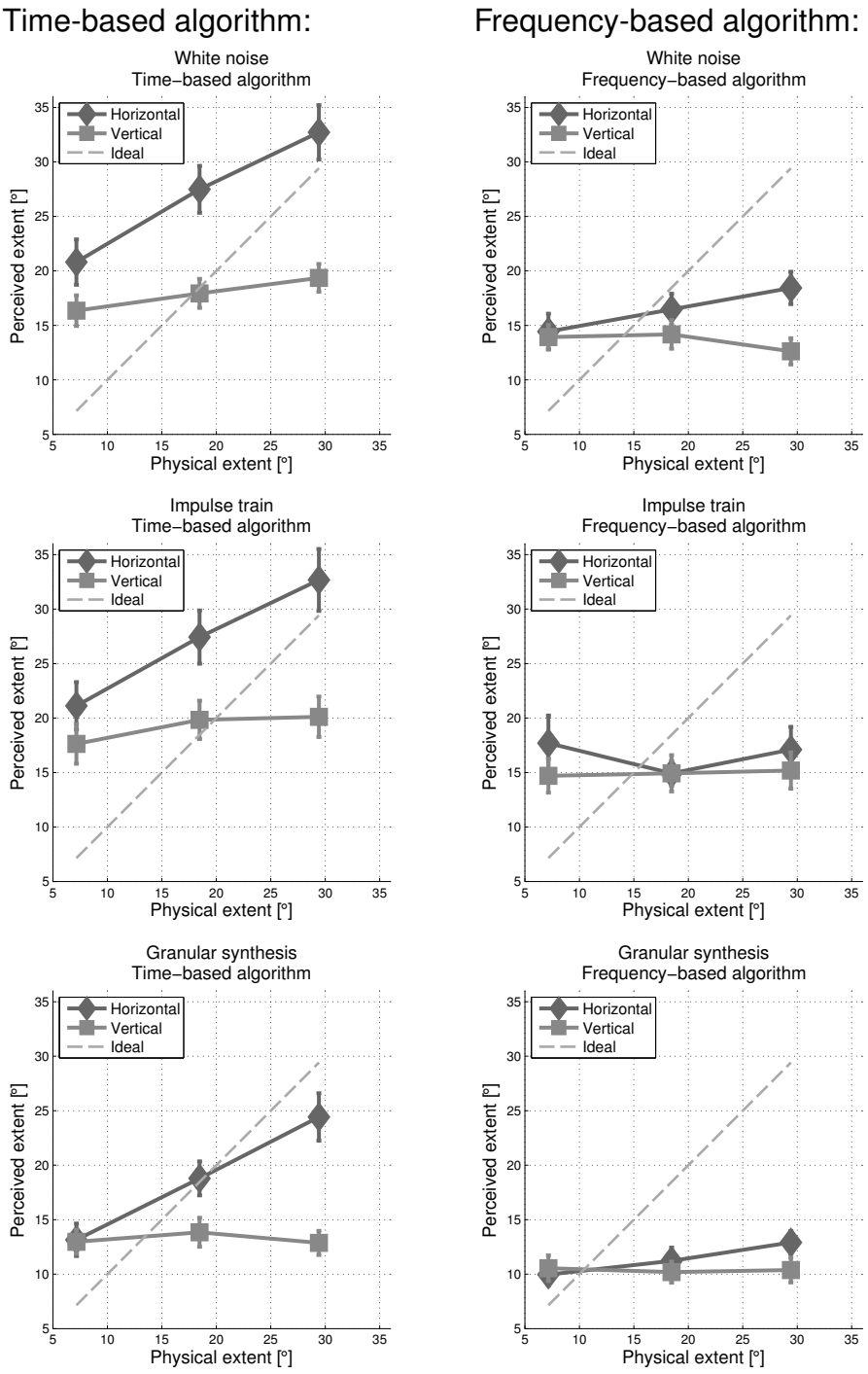


Figure 4.1.: Perceived vs. physical spatial extent in the different conditions of the experiment. Error bars indicate standard error of the mean.

A normal distribution of the judgments of spatial extent was assumed as the Lilliefors test [Lil67] could not reject the null-hypothesis that the data is normally distributed at a significance level of 5 percent. Grubbs' test [Gru50] did not detect any outliers in the data. Therefore a statistical analysis of the results in which a four-way (Stimulus × Algorithm × Spatial Distribution × Orientation) repeated mea-

sures Analysis of Variance (ANOVA) on perceived spatial extent was performed. The results can be found in Table 4.1.

Table 4.1.: The results of a four-way (Stimulus × Algorithm × Spatial Distribution × Orientation) repeated measures ANOVA on perceived spatial extent.

Main effects and interactions			
Stimulus	$F(2,34) =$	20.749 ,	$p < 0.001$
Algorithm	$F(1,17) =$	47.679 ,	$p < 0.001$
Spatial Distribution	$F(2,34) =$	19.550 ,	$p < 0.001$
Orientation	$F(1,17) =$	16.852 ,	$p = 0.001$
Stimulus × Algorithm	$F(2,34) =$	2.890 ,	$p = 0.069$
Stimulus × Spatial Distribution	$F(4,68) =$	0.925 ,	$p = 0.455$
Stimulus × Orientation	$F(2,34) =$	1.188 ,	$p = 0.317$
Algorithm × Spatial Distribution	$F(2,34) =$	24.634 ,	$p < 0.001$
Algorithm × Orientation	$F(1,17) =$	44.511 ,	$p < 0.001$
Spatial Distribution × Orientation	$F(2,34) =$	15.171 ,	$p < 0.001$
Stimulus × Algorithm × Spatial Distribution	$F(4,68) =$	1.794 ,	$p = 0.140$
Stimulus × Algorithm × Orientation	$F(2,34) =$	0.751 ,	$p = 0.480$
Stimulus × Spatial Distribution × Orientation	$F(4,68) =$	1.922 ,	$p = 0.117$
Algorithm × Spatial Distribution × Orientation	$F(2,34) =$	18.897 ,	$p < 0.001$
Stimulus × Algorithm × Spat. Dist. × Orient.	$F(4,68) =$	1.568 ,	$p = 0.193$

In general, p-values of less than or equal to '0.05' are considered as a qualification for significance. The term "marginal significance" is used if the p-value is close to this threshold, even if it is actually exceeding it up to '0.06'. The latter is therefore not considered significant in strict compliance to the binary decision at the threshold.

At the $p \leq 0.001$ level the main effects of Stimulus, Algorithm, Spatial Distribution, and Orientation were significant. In pairwise comparisons using t-tests, white noise and impulse train resulted in significantly larger perceived spatial extent in comparison to the granular synthesis signal ($p<0.001$). Furthermore, the judgments for each spatial distribution were significantly different from each other, small spatial distributions were judged to be significantly narrower than both medium and large, and medium spatial distributions narrower than large ($p<0.01$). Finally, judgments in vertical orientation were significantly narrower than in horizontal orientation, and the time-based algorithm resulted in significantly wider judgments than the frequency-based algorithm. This justifies the observations in Figure 4.1, in which all main effects appeared to have a significant influence on perceived spatial extent.

A number of interactions were significant and are analyzed further. Concerning the interaction between Algorithm and Spatial Distribution, the interaction was

because when the different spatial extents were produced using the time-based algorithm, they resulted in judgments that were always significantly different to each other, i.e., perceived spatial extent with small spatial distribution was significantly smaller compared to both medium and large ones, and perceived spatial extent of medium spatial distribution was significantly smaller compared to the large one ($p<0.001$). In the case of the frequency-based algorithm, Spatial Distribution had no significant effect on the perceived spatial extent. In summary, the described two-way interaction shows the significance of the observation that the time-based algorithm not only led to a significantly larger perceived spatial extent compared to the frequency-based algorithm, but also evoked larger perceived spatial extent for larger spatial distributions compared to smaller ones.

The interaction between Algorithm and Orientation was due to the fact that with the time-based algorithm, judgments for horizontal orientation were significantly wider compared to vertical orientation ($t(17)=5.855$, $p<0.001$), while for the frequency-based algorithm there was no significant difference between the the horizontal and vertical judgments, arguably because they were narrow in both cases. This significant interaction between Algorithm and orientation was also observed in Figure 4.1.

The interaction between Spatial Distribution and Orientation was because of two reasons. The first was that for horizontal orientation, perceived spatial extent was significantly influenced from the actual spatial extent, i.e., perceived spatial extent for the small spatial distribution was significantly smaller compared to both medium and large spatial distribution, and perceived extent for medium spatial distribution was significantly smaller compared to the large one (at least $p<0.01$). However, for vertical orientation, judgments of spatial extent were only marginally significantly different depending on the actual spatial extent, small spatial distribution narrower than medium ($t(17)=-1.839$, $p=0.083$), small narrower than large ($t(17)=-1.865$, $p=0.080$), whereas the difference between judgments for vertical orientation and medium or large spatial distribution was not significant. The second reason relates to the fact that for the small spatial distribution the judgments for horizontal and vertical orientation were not significantly different to each other, whereas both medium and large spatial distribution resulted in significantly larger perceived spatial extent for horizontal compared to vertical spatial extent.

The three-way interaction between Algorithm, Spatial Distribution, and Orientation was also significant as were the two-way interactions that emerge from these three factors. This interaction was significant because irrespective of stimulus, and in particular for the two larger spatial distributions, the perceived spatial extent was significantly larger in the case of the time-based algorithm for horizontal orientation in comparison to vertical orientation. This is evidenced by the fact that

the two-way interaction between Orientation and Spatial Distribution was significant for the time-based but not for the frequency-based algorithm when analyzing the data averaged over Stimulus. This three-way interaction evidences the significance of the above observation, that only for horizontal orientation, the time-based algorithm led to an amount of perceived spatial extent which was ordered according to the individual spatial distributions.

4.1.2. Response time

The individual response times for the different conditions of the experiment are shown in Figure 4.2. It must be noted, that some participants interrupted the experiment. Those obvious outliers were identified during the experiment and replaced by the mean of the other repetitions of the condition. These included the first trial of each of the two parts (horizontal and vertical) for all participants, and a total of three additional trials.

On average it took the participants 11.0 s to complete one trial. Overall there are no big differences, but a tendency that participants took less time for vertical judgments than for horizontal ones can be observed. In the case of horizontal orientation there is also a tendency of longer response time for larger spatial distribution.

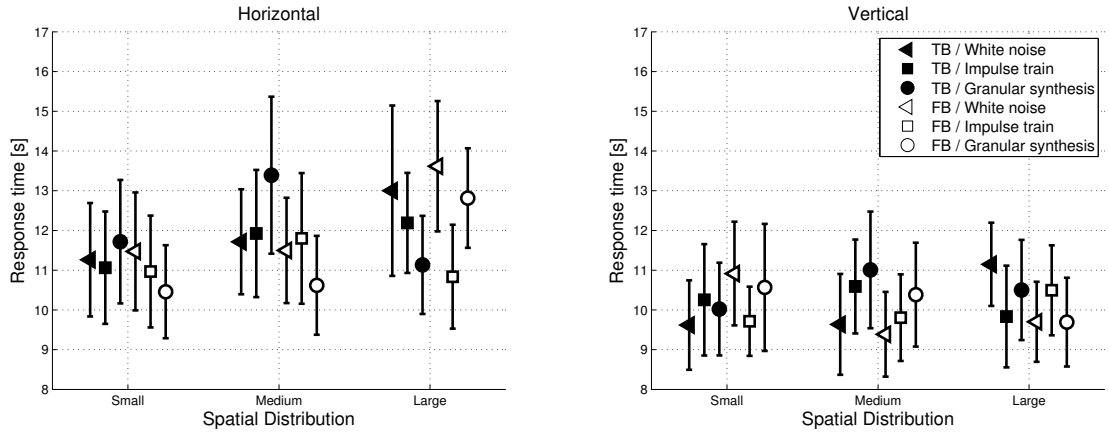


Figure 4.2.: Response times for the different conditions of the experiment. Error bars indicate standard error of the mean.

A normal distribution of the response time is assumed, which was confirmed through the Lilliefors test [Lil67] after a logarithmic transformation. Grubbs' test [Gru50] did not detect any outliers in the data. The results are therefore analyzed statistically by performing a four-way (Stimulus × Algorithm × Spatial Distribution × Orientation) repeated measures ANOVA on the decadic logarithm of the response time. The results can be found in Table 4.2.

According to the statistical analysis, the main effect of Spatial Distribution showed a marginally significant effect on the response time ($F(2,34)=3.492$, $p=0.042$). The main effect of Orientation is just above the threshold and therefore not considered significant ($F(1,17)=4.157$, $p=0.057$). No interactions were significant.

The statistical analysis confirms that the above observations of Figure 4.2 are only small tendencies, which are probably not significant. On the one hand, in horizontal orientation, the response time was marginally significantly longer for larger spatial distributions compared to smaller ones. On the other hand, the response time was marginally significantly longer for horizontal orientation than it was for vertical orientation.

Table 4.2.: The results of a four-way (Stimulus × Algorithm × Spatial Distribution × Orientation) repeated measures ANOVA on the decadic logarithm of the response time.

Main effects and interactions			
Stimulus	$F(2,34) =$	0.539 ,	$p = 0.588$
Algorithm	$F(1,17) =$	0.430 ,	$p = 0.521$
Spatial Distribution	$F(2,34) =$	3.492 ,	$p = 0.042$
Orientation	$F(1,17) =$	4.157 ,	$p = 0.057$
Stimulus × Algorithm	$F(2,34) =$	2.451 ,	$p = 0.101$
Stimulus × Spatial Distribution	$F(4,68) =$	0.809 ,	$p = 0.524$
Stimulus × Orientation	$F(2,34) =$	0.745 ,	$p = 0.482$
Algorithm × Spatial Distribution	$F(2,34) =$	0.418 ,	$p = 0.662$
Algorithm × Orientation	$F(1,17) =$	0.296 ,	$p = 0.594$
Spatial Distribution × Orientation	$F(2,34) =$	2.025 ,	$p = 0.148$
Stimulus × Algorithm × Spatial Distribution	$F(4,68) =$	0.729 ,	$p = 0.575$
Stimulus × Algorithm × Orientation	$F(2,34) =$	0.415 ,	$p = 0.664$
Stimulus × Spatial Distribution × Orientation	$F(4,68) =$	0.320 ,	$p = 0.864$
Algorithm × Spatial Distribution × Orientation	$F(2,34) =$	0.995 ,	$p = 0.380$
Stimulus × Algorithm × Spat. Dist. × Orient.	$F(4,68) =$	2.200 ,	$p = 0.087$

4.1.3. Perceived center

Figure 4.3 shows the center positions, calculated from the beginning and end point of each of the presented sounds, in the different conditions of the experiment.

An offset in the perceived center position can be observed, which is about -0.25° (azimuth) in horizontal and 0.25° (elevation) in vertical orientation. On the one hand, this could be due to inaccuracies in the calibration of the projection mapping. On the other hand it was observed that most participants were drawing

their judgments from left to right, for horizontal, and from top to bottom for vertical orientation, respectively, which could have introduced this drift.

Regarding the error, perceived center positions were much more stable in horizontal than in vertical orientation. While the perceived center position in horizontal orientation seemed to be constant for the time-based algorithm, the frequency-based algorithm seems to introduce an additional offset for medium and large spatial distributions. This could be due to the less flat power distribution of the individual loudspeakers (remember Figure 3.6).

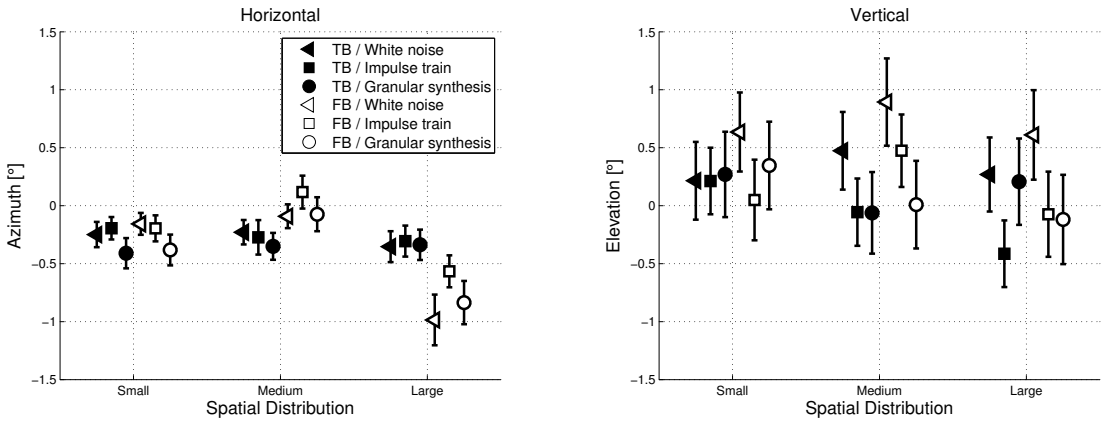


Figure 4.3.: Perceived center in the different conditions of the experiment. Error bars indicate standard error of the mean.

4.1.4. Discussion

The results allow certain conclusions to be made with respect to the possibility of eliciting the perception of either vertically or horizontally extended sounds. While creating the impression of horizontal extension was possible even with the relatively small dimensions used in this experiment, the algorithms used here can only partially create the impression of vertical extent. In addition, irrespective of extension orientation or signal type, the time-based algorithms resulted in significantly larger perceptions of both horizontal and vertical extent.

Horizontal extent impressions created by the time-based algorithm varied systematically with the actual extent irrespective of signal type as evidenced by the fact that the different horizontal extents used in the study resulted in significantly different distributions of perceived spatial extent. Interestingly, actual horizontal extent was overestimated in the perceptual judgments, especially at the smaller actual spatial extents. In the vertical direction however, perceived spatial extent varied less systematically with the actual one. Although a significant increase in the perceived vertical spatial extent with increased actual vertical spatial ex-

tent appeared for the white noise and impulse train stimuli, the difference was consistently significant only when comparing the smallest with largest spatial distribution ($t(17)=3.46$, $p=0.003$ for white noise and $t(17)=2.43$, $p=0.047$ for impulse trains) and no significant differences in perceived vertical extent when using the granular synthesis signal were observed. In addition, judgments of perceived vertical extent underestimated the actual extent by far, pointing to limited applicability in real-world applications. Performance for the white noise and the impulse train signals was similar, both for horizontally and for vertically extended sound sources. For the granular synthesis signal, relatively good identification of spatial extent was obtained in horizontal orientation, but not in the median plane. This might relate to sound design issues that need to be investigated further, such as optimization of the grains to yield as good localization as possible.

Concerning the frequency-based algorithm, although in general perceived horizontal extent increased in proportion to the physical extent, judgments underestimated the actual horizontal extent for large spatial distribution and were significantly smaller than the ones obtained by the time-based algorithm. The results were inferior to those described by [PSP14], however, they used larger spatial distributions than the ones described here. The algorithm fails in representing vertically extended sound sources. This could be attributed to the different mechanisms that operate and determine azimuth and elevation perception. While azimuth perception operates on the basis of interaural time differences, spectral cues and familiarity with source spectrum are mainly responsible for elevation perception [WK97; Bla97]. It appears therefore that while the combination of information from frequencies at different azimuths to yield the impression of coherent spatially extended auditory sources provides a functional basis for the creation of horizontally extended sources, this mechanism fails for vertically extended sources. This can be explained by the fact that presenting signal frequencies at different elevations destroys the consistency with which the signal spectrum is filtered by the outer ear to result in the perception of elevation. This is a fundamental problem when it comes to representing vertical extent by distributing signal frequencies in elevation that might be difficult to overcome.

An aspect worth considering further is the overestimation of the actual spatial extent that occurred for the time-based and to a lesser extent for the frequency-based algorithm. On the one hand, this may be attributed to non-spatial factors pertaining to source-size perception, which are inherently confounded with spatial extent judgments. This may provide an alternative explanation to why the granular synthesis stimulus was always judged to be narrower than the other two signals and indicates that perceptual calibration might be necessary in order to match actual and perceived spatial extent. On the other hand, the overestimation

is possibly induced by the room acoustics in conjunction with the spatial sound radiation of the individual loudspeakers. The smallest perceived spatial extent has been observed for the granular synthesis stimulus with the frequency-based algorithm, being roughly 10°. However, in the literature, this amount of auditory source width already appeared for a single loudspeaker, depending on room acoustics and loudspeaker model, e.g., [Fra13]. This influence of the room acoustics is further investigated in Section 4.4.

The perceived center position was found to be strongly correlated with the perceived spatial extent judgments, which is forced by the task in which participants were drawing their judgments of perceived spatial extent on a projection screen. For perceived center, this confound could have been eliminated by forcing the participants to draw their judgments in alternating direction. However, when relating the offset in perceived center to the corresponding physical spatial extent (7.2° for small, 18.5° for medium, and 29.4° for large spatial distributions), the drift appears low and probably not relevant.

Regarding the response time, two tendencies could be observed. Overall, horizontally extended sound sources led to marginally significantly longer response time than vertical ones. In horizontal orientation, additionally, larger spatial distributions led to marginally significantly longer response time compared to smaller spatial distributions. When comparing these results to the judgments of perceived spatial extent, there appears a strong correlation between both results, meaning larger perceived spatial extent led to marginally significantly longer response time than smaller spatial extent. This seems comprehensible, since it took the participants more time to draw a large line on the projection screen, compared to a smaller one.

4.2. Results of relative judgments

This section presents the results of the pairwise comparisons, gained from relative judgments. In each trial participants compared two different signals generated by the four independent variables Algorithm, Stimulus, Spatial Distribution and Orientation. The presented pairs were structured in three different cases in which always one of the three independent variables Algorithm, Stimulus, or Spatial Distribution was varied. In each of these three cases the levels of a single given independent variable were compared against each other. Participants were asked to judge which of the two presented sounds had a larger perceived spatial extent. This was done separately in both horizontal and vertical orientation.

4.2.1. Perceived spatial extent

The results of the relative judgments of perceived spatial extent are structured by the three independent variables Algorithm, Stimulus, and Spatial Distribution, whose levels were compared against each other.

4.2.1.1. Comparison between algorithms

In the first case, the levels of the independent variable Algorithm (time-based and frequency-based) were compared to each other. The two levels led to only one pairwise comparison for each permutation of the three remaining independent variables Stimulus, Spatial Distribution and Orientation.

Figure 4.2.1.1 shows the probability that signals generated with the time-based algorithm were judged larger than with the frequency-based algorithm. The results are presented in two different plots for horizontal and vertical orientation on the left and right side respectively.

There was an overall tendency that signals generated with the time-based algorithm were judged larger than with the frequency-based algorithm. This tendency seems to be stronger for larger spatial distributions compared to smaller ones. While this tendency towards the time-based algorithm is clear in horizontal orientation, in vertical orientation, however, it seems to be only marginal.

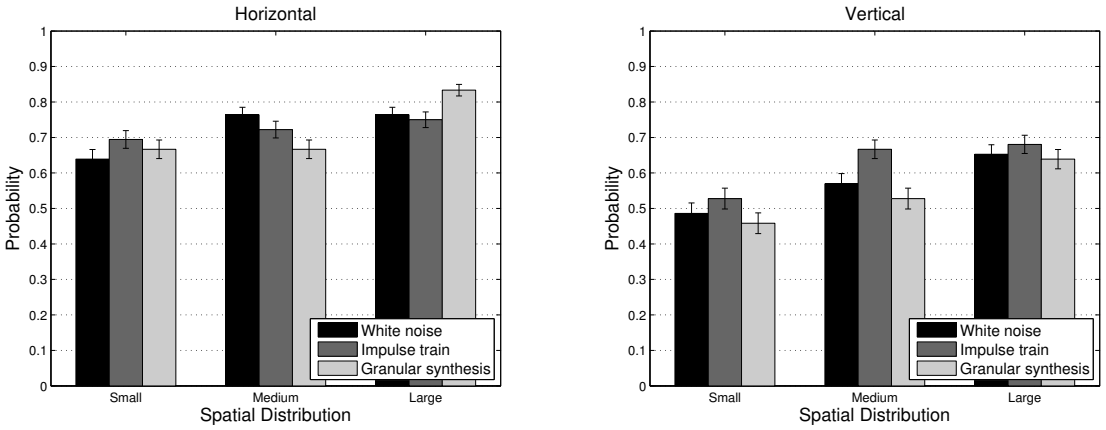


Figure 4.4.: Probability that signals with the time-based algorithm were perceived larger than with the frequency-based algorithm, resulting from pairwise comparison test. Error bars indicate standard error of the mean.

On average, the proportion of trials in which the frequency-based algorithm was judged to provide a larger perceived spatial extent was 34.88 percent. The judgments of each participant (subject) were grouped together and the mean proportion of each participant in all trials in the experiment was calculated. A

single-tailed t-test showed that the distribution of the proportions was significantly less than chance (0.5), $t(17)=-4.31$, $p<0.001$. It is concluded that the time-based algorithm globally resulted in larger perceived spatial extent.

In order to test the effect of the independent variables (Orientation, Spatial Distribution, and Stimulus) on this probability, a generalized linear mixed effects model with a logit link function was used. Orientation, Spatial Distribution, and Stimulus were treated as fixed factors while subjects and repetitions were treated as random factors. Repetitions were nested in subjects. The factors Orientation and Spatial Distribution resulted in regression coefficients that were significantly different from zero (Orientation: $z=5.620$, $p<0.001$, Spatial Distribution: $z=-4.666$, $p<0.001$). There was no effect of factor Stimulus, as indicated by the fact the the regression coefficient for this factor was not significantly different than zero. Accordingly, the aforementioned probability was 64, 67, and 63 percent for the three stimuli. On the contrary, the probability with which the frequency algorithm was found to produce smaller auditory impressions was 72 percent for horizontal orientation, while only 58 percent for vertical orientation. The same probability was 58, 65, and 72 percent for the three spatial distributions, respectively.

The above observations from Figure 4.2.1.1 are all supported by the statistical analysis. Overall, the time-based algorithm resulted in significantly larger perceived spatial extent than the frequency-based algorithm. The tendency of the time-based algorithm resulting in larger perceived spatial extent than the frequency-based algorithm was significantly stronger for horizontal orientation compared to vertical orientation, and also for larger spatial distributions compared to smaller ones.

4.2.1.2. Comparison between stimuli

Here, the results of the second case are presented, in which the three levels of Stimulus (white noise, impulse train, and granular synthesis) were compared to each other. The three levels led to three pairwise comparisons for each permutation of the three remaining independent variables Algorithm, Spatial Distribution and Orientation.

The paired comparison data was analyzed using Thurstone's Case V Model [Thu27; TG11], which generates scale values for the individual levels of the varied independent variable regarding perceived spatial extent. The resulting scales of stimuli are shown in Figure 4.2.1.2. The results are presented in two different plots for horizontal and vertical orientation on the left and right side respectively.

Overall it can be seen that white noise and impulse trains were perceived larger than the granular synthesis stimulus. While for both impulse train and granular

synthesis stimulus the orientation seems to have no effect on the scale values, the scale value of the white noise signal is much lower for vertical orientation than for horizontal orientation.

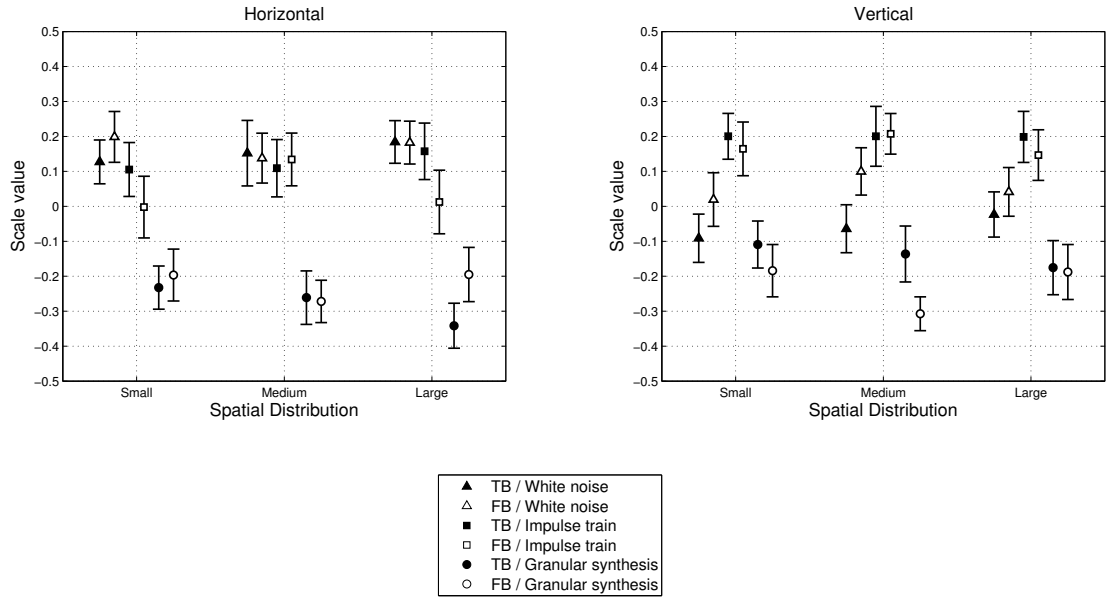


Figure 4.5.: Scales of stimuli, concerning perceived spatial extent, resulting from pairwise comparison test. Error bars indicate standard error of the mean.

No outliers were detected in the data and a normal distribution is assumed. Therefore a statistical analysis of the results by means of a four-way (Stimulus × Algorithm × Spatial Distribution × Orientation) repeated measures ANOVA on the scale-values was performed (see Table 4.3).

Only the main effect of Stimulus was significant ($F(2,34)=14.255$, $p<0.001$). Pairwise comparisons using t-tests showed that scale values for white noise and impulse trains were significantly larger than those for the granular synthesis stimulus ($t(17)=4.080$, $p=0.001$ and $t(17)=5.091$, $p<0.001$), while the difference between white noise and impulse trains was not significant.

There was also a marginally significant interaction of Stimulus and Orientation ($F(2,34)=3.389$, $p=0.045$). This was due to the fact that the scale values of white noise were significantly higher for horizontal orientation than for vertical orientation ($t(17)=3.700$, $p=0.002$), while for both other stimuli the orientation had no significant effect on the scale values.

The statistical analysis gives evidence that both white noise and impulse trains produced significantly larger perceived spatial extent than the granular synthesis stimulus. Also the observation that white noise produced smaller spatial extent in vertical orientation, compared to horizontal orientation, was found to be significant.

Table 4.3.: The results of a four-way (Stimulus × Algorithm × Spatial Distribution × Orientation) repeated measures ANOVA on scale values of stimuli, concerning perceived spatial extent.

Main effects and interactions			
Stimulus	F(2,34) =	14.255 ,	p < 0.001
Algorithm	F(1,17) =	1.495 ,	p = 0.238
Spatial Distribution	F(2,34) =	0.115 ,	p = 0.892
Orientation	F(1,17) =	0.557 ,	p = 0.466
Stimulus × Algorithm	F(2,34) =	1,885 ,	p = 0.167
Stimulus × Spatial Distribution	F(4,68) =	0.886 ,	p = 0.477
Stimulus × Orientation	F(2,34) =	3.389 ,	p = 0.045
Algorithm × Spatial Distribution	F(2,34) =	0.176 ,	p = 0.840
Algorithm × Orientation	F(1,17) =	0.296 ,	p = 0.594
Spatial Distribution × Orientation	F(2,34) =	0.352 ,	p = 0.706
Stimulus × Algorithm × Spatial Distribution	F(4,68) =	0.939 ,	p = 0.447
Stimulus × Algorithm × Orientation	F(2,34) =	2.910 ,	p = 0.068
Stimulus × Spatial Distribution × Orientation	F(4,68) =	0.273 ,	p = 0.894
Algorithm × Spatial Distribution × Orientation	F(2,34) =	0.239 ,	p = 0.788
Stimulus × Algorithm × Spat. Dist. × Orient.	F(4.68) =	0.299 ,	p = 0.878

Additionally, a three-way (Algorithm × Spatial Distribution × Orientation) repeated measures ANOVA on the ranges (distance between maximum and minimum) of the above scale values was performed (see Table 4.4). There was only one significant interaction between Algorithm and Spatial Distribution ($F(2,34)=4.020$, $p=0.027$). This is due to the fact that with the time-based algorithm the range with small spatial distribution was significantly smaller than with medium ($t(17)=-2.419$, $p=0.027$) and marginally smaller than with large spatial distribution ($t(17)=-2.024$, $p=0.059$).

This significant interaction between Algorithm and Spatial distribution can also be observed in Figure 4.2.1.2, in which the range of the scale values in case of the time-based algorithm is larger for horizontal orientation than for vertical orientation.

Table 4.4.: The results of a three-way (Algorithm × Spatial Distribution × Orientation) repeated measures ANOVA on the range of scale values of stimuli, concerning perceived spatial extent.

Main effects and interactions			
Algorithm	F(1,17) =	0.166 ,	p = 0.688
Spatial Distribution	F(2,34) =	1.300 ,	p = 0.286
Orientation	F(1,17) =	1.162 ,	p = 0.296
Algorithm × Spatial Distribution	F(2,34) =	4.020 ,	p = 0.027
Algorithm × Orientation	F(1,17) =	0.634 ,	p = 0.437
Spatial Distribution × Orientation	F(2,34) =	0.365 ,	p = 0.697
Algorithm × Spatial Distribution × Orientation	F(2,34) =	0.288 ,	p = 0.751

4.2.1.3. Comparison between spatial distributions

In the third case the three levels of Spatial Distribution (small, medium, and large) were compared to each other. The three levels led to three pairwise comparisons for each permutation of the three remaining independent variables Algorithm, Stimulus, and Orientation.

As in the previous case scale values regarding perceived spatial extent were generated for the individual levels of the varied independent variable by using Thurstone's Case V Model [Thu27; TG11]. Figure 4.2.1.2 shows the resulting scales of spatial distributions for horizontal and vertical orientation on the left and right side respectively.

The scales of Spatial Distribution can be seen in Figure 4.2.1.3. For the time-based algorithm in horizontal orientation it can be observed that the scale values of larger spatial distributions were always higher compared to smaller ones. A similar tendency exists for the frequency-based algorithm in horizontal orientation. In vertical orientation, no such obvious trend was found.

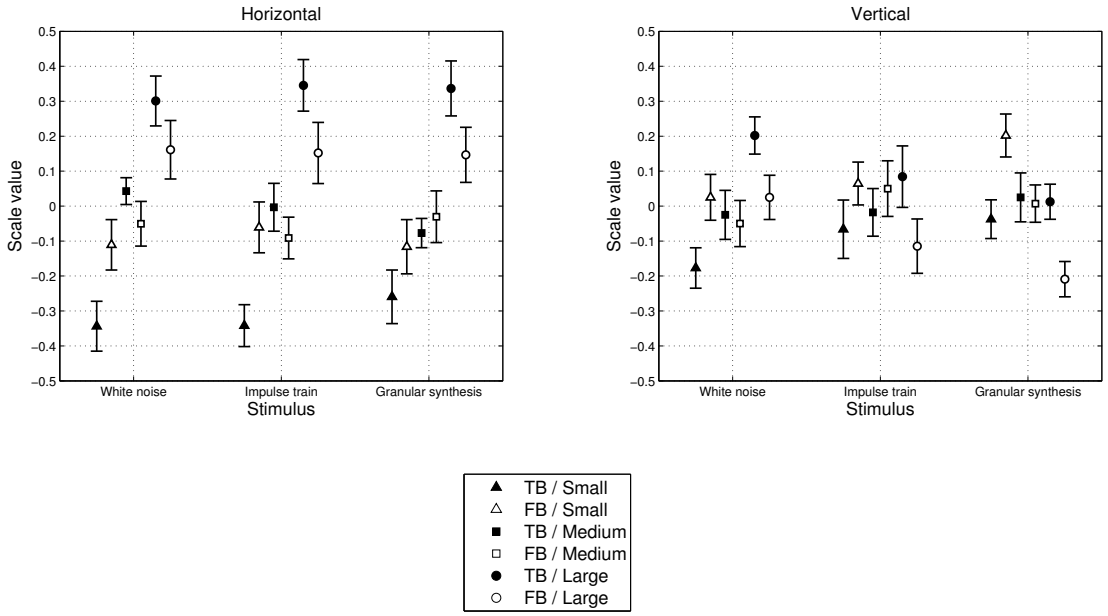


Figure 4.6.: Scales of spatial distributions, concerning perceived spatial extent, resulting from pairwise comparison test. Error bars indicate standard error of the mean.

A normal distribution of the data was assumed and no outliers were detected. For the statistical analysis of the results a four-way (Stimulus × Algorithm × Spatial Distribution × Orientation) repeated measures ANOVA on the scale-values for Spatial Distribution was performed (see Table 4.5). The only significant main effect was Spatial Distribution ($F(2,34)=11.287$, $p<0.001$).

Furthermore, a couple of significant interactions appeared, which need further investigation. The interaction between Algorithm and Spatial Distribution ($F(2,34)=10.185$, $p<0.001$) was due to the fact that Spatial Distribution showed a significant effect on the judgments of perceived spatial extent only for the time-based algorithm but not for the frequency-based algorithm: With the time-based algorithm medium spatial distributions were judged significantly larger than small spatial distributions ($t(17)=-2.603$, $p=0.019$), and large were judged significantly larger than both small and medium spatial distributions ($t(17)=-5.376$, $p<0.001$, res. $t(17)=-4.469$, $p<0.001$). This confirms the observation that the time-based algorithm is capable of creating significantly different perceived spatial extent for different spatial distributions.

The interaction between Spatial Distribution and Orientation ($F(2,34)=10.539$, $p<0.001$) was due to the fact that for horizontal orientation medium spatial distribution was judged significantly larger than small ($t(17)=-2.524$, $p=0.022$), and large spatial distributions were judged significantly larger than both small and medium ones ($t(17)=-4.480$, $p<0.001$, res. $t(17)=-4.259$, $p=0.001$). In vertical orientation, however, none of the two algorithms was able to produce a significantly

different perceived spatial extent for different spatial distributions. The described interaction gives evidence to the observation that for both algorithms in horizontal orientation, larger spatial distributions led to significantly larger perceived spatial extent than horizontal ones.

Table 4.5.: The results of a four-way (Stimulus × Algorithm × Spatial Distribution × Orientation) repeated measures ANOVA on scale values of spatial distributions, concerning perceived spatial extent.

Main effects and interactions			
Stimulus	$F(2,34) =$	0.665 ,	$p = 0.521$
Algorithm	$F(1,17) =$	1.700 ,	$p = 0.210$
Spatial Distribution	$F(2,34) =$	11.287 ,	$p < 0.001$
Orientation	$F(1,17) =$	0.147 ,	$p = 0.707$
Stimulus × Algorithm	$F(2,34) =$	1.277 ,	$p = 0.292$
Stimulus × Spatial Distribution	$F(4,68) =$	2.165 ,	$p = 0.082$
Stimulus × Orientation	$F(2,34) =$	0.855 ,	$p = 0.434$
Algorithm × Spatial Distribution	$F(2,34) =$	10.185 ,	$p < 0.001$
Algorithm × Orientation	$F(1,17) =$	2.125 ,	$p = 0.163$
Spatial Distribution × Orientation	$F(2,34) =$	10.539 ,	$p < 0.001$
Stimulus × Algorithm × Spatial Distribution	$F(4,68) =$	0.355 ,	$p = 0.839$
Stimulus × Algorithm × Orientation	$F(2,34) =$	1.195 ,	$p = 0.315$
Stimulus × Spatial Distribution × Orientation	$F(4,68) =$	1.692 ,	$p = 0.162$
Algorithm × Spatial Distribution × Orientation	$F(2,34) =$	0.184 ,	$p = 0.833$
Stimulus × Algorithm × Spat. Dist. × Orient.	$F(4,68) =$	0.748 ,	$p = 0.563$

An additional three-way (Stimulus × Algorithm × Orientation) repeated measures ANOVA on the ranges of the scale values from above was performed (see Table 4.6).

The main effect of Orientation was significant ($F(1,17)=11.223$, $p=0.004$). This gives further evidence to the result that only in horizontal orientation, larger spatial distributions led to larger perceived spatial extent, when compared to smaller spatial distributions, producing smaller perceived spatial extent.

Algorithm had only a marginally significant effect on the range ($F(1,17)=4.611$, $p=0.46$), which shows once more that the time-based algorithm produced significantly larger perceived spatial extent than the frequency-based algorithm.

Table 4.6.: The results of a three-way (Stimulus × Algorithm × Orientation) repeated measures ANOVA on the range of scale values of spatial distributions, concerning perceived spatial extent.

Main effects and interactions			
Stimulus	F(2,34) =	1.892 ,	p = 0.166
Algorithm	F(1,17) =	4.611 ,	p = 0.046
Orientation	F(1,17) =	11.223 ,	p = 0.004
Stimulus × Algorithm	F(2,34) =	0.789 ,	p = 0.463
Stimulus × Orientation	F(2,34) =	1.608 ,	p = 0.215
Algorithm × Orientation	F(1,17) =	1.268 ,	p = 0.276
Stimulus × Algorithm × Orientation	F(2,34) =	0.285 ,	p = 0.754

4.2.2. Response time

As the presented pairs of signals were played back in an infinite loop until the participant made a choice, the resulting response time varied for each trial. On average, it took participants 6.5 s to complete one trial. 6.3 s for judgments in horizontal and 6.7 s for judgments in vertical orientation.

As with the absolute judgments, obvious outliers of the response times were identified during the experiment and replaced by the mean of the other repetitions of the condition. For all six parts (pair comparisons between algorithms, stimuli and spatial distributions, each in both orientations) the response time for the first judgment was replaced for all participants. Additionally, six individual outliers of different participants were identified and replaced.

The results of the response time are structured by the three independent variables Algorithm, Stimulus, and Spatial Distribution, whose levels were compared against each other.

4.2.2.1. Comparison between algorithms

First, the response times for pairwise comparisons between the two levels of Algorithm (time-based vs. frequency-based) are presented.

Figure 4.7 illustrates, how much time the participants took for their judgments when comparing the two different algorithms. On average, a trial was completed after 6.3 s. There is an overall tendency for decreasing response time with increasing spatial distribution (from small, via medium, to large). Furthermore, participants took less time to judge pairs in horizontal than in vertical orientation. Except for small spatial distributions in horizontal orientation, participants took more time to judge pairs of Algorithm for both white noise and granular synthesis

stimuli than for the impulse train stimulus.

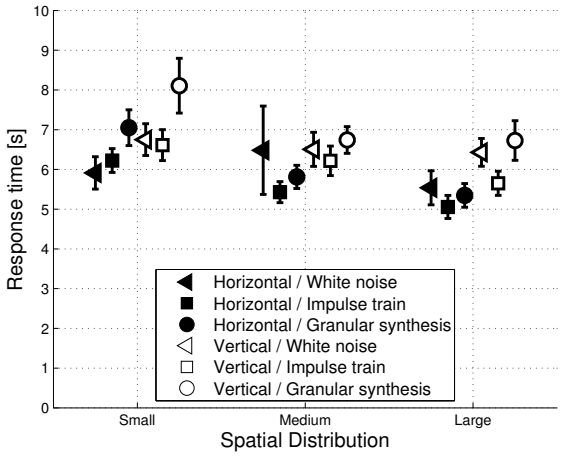


Figure 4.7.: Response time for pairwise comparisons between algorithms. Error bars indicate standard error of the mean.

Grubbs' test did not find any additional outliers, and the data was found to follow a normal distribution by the Lilliefors test, after a square root transformation. Therefore a statistical analysis of the square root of the response time was performed.

The results of the three-way (Stimulus × Spatial Distribution × Orientation) repeated measures ANOVA is shown in Table 4.7. It can be seen that all main effects showed a significant effect on the response time, while no interaction was significant.

The statistical analysis gives evidence to the above observations: First, an increase in spatial distribution led to a significant decrease in response time. Secondly, the response time for vertical orientation was significantly longer than for horizontal orientation. Finally, both white noise and granular synthesis led to a significantly longer response time than the impulse train stimulus.

Table 4.7.: The results of a three-way (Stimulus × Spatial Distribution × Orientation) repeated measures ANOVA on the square root of the response time for pairs of Algorithm.

Main effects and interactions			
Stimulus	F(2,34) =	4.737 ,	p = 0.015
Spatial Distribution	F(2,34) =	5.842 ,	p = 0.007
Orientation	F(1,17) =	7.200 ,	p = 0.016
Stimulus × Spatial Distribution	F(4,68) =	1.674 ,	p = 0.166
Stimulus × Orientation	F(2,34) =	0.367 ,	p = 0.696
Spatial Distribution × Orientation	F(2,34) =	0.367 ,	p = 0.696
Stimulus × Spatial Distribution × Orientation	F(4,68) =	0.389 ,	p = 0.816

4.2.2.2. Comparison between stimuli

Here, the response times for the pairwise comparisons between the different levels of Stimulus (white noise, impulse train, and granular synthesis) are presented.

Figure 4.2.2.2 shows the amount of time it took participants to respond. On average, it took 6.4 s to compare a stimulus pair. When judging horizontal extent in horizontal orientation of medium and large spatial distributions, it can be seen that participants took longer to judge pairs with the frequency-based algorithm than with the time-based algorithm. For vertical orientation it seems that participants took longer to judge pairs of large spatial distribution compared to medium and small spatial distribution.

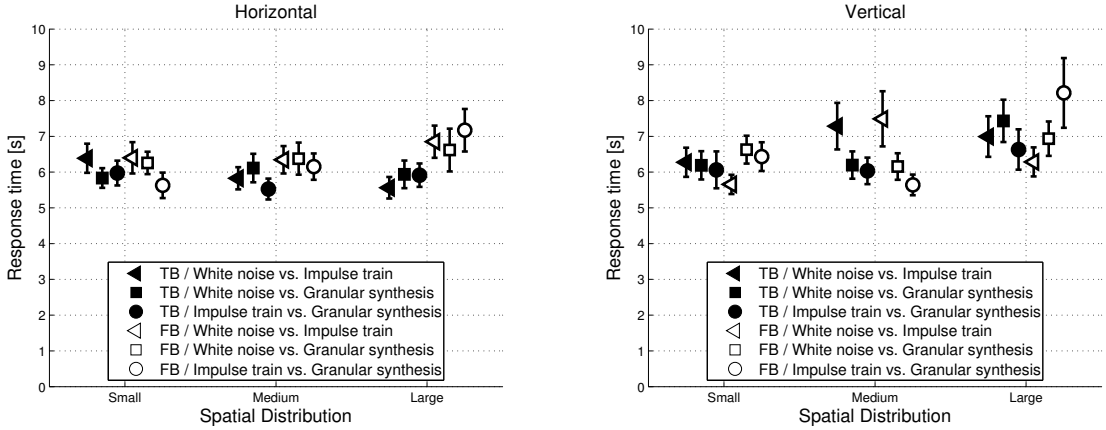


Figure 4.8.: Response time for pairwise comparisons between stimuli. Error bars indicate standard error of the mean.

No additional outliers were detected by Grubbs' test and the Lilliefors test evidenced a normal distribution of the log-transformed response times.

The statistical analysis of the decadic logarithm of the response time is shown in Table 4.8. It can be seen that actually none of the above observations were significant, but two different interactions showed at least marginal significance (Spatial Distribution × Pair: $F(4,68)=2.850$, $p=0.030$; Orientation × Spatial Distribution × Pair: $F(4,68)=2.629$, $p=0.042$). Nevertheless, it is argued here that the differences in response time are small and no satisfactory interpretation of the results can be articulated. The response times for pairwise comparisons between the three levels of Stimulus are therefore not further examined.

Table 4.8.: The results of a four-way (Algorithm × Spatial Distribution × Orientation × Pair) repeated measures ANOVA on the decadic logarithm of the response time for pairs of Stimulus.

Main effects and interactions			
Algorithm	F(1,17) =	2.256 ,	p = 0.151
Spatial Distribution	F(2,34) =	2.978 ,	p = 0.064
Orientation	F(1,17) =	0.570 ,	p = 0.461
Pair	F(2,34) =	0.474 ,	p = 0.627
Algorithm × Spatial Distribution	F(2,34) =	1.018 ,	p = 0.372
Algorithm × Orientation	F(1,17) =	1.059 ,	p = 0.318
Algorithm × Pair	F(2,34) =	0.707 ,	p = 0.500
Spatial Distribution × Orientation	F(2,34) =	1.240 ,	p = 0.302
Spatial Distribution × Pair	F(4,68) =	2.850 ,	p = 0.030
Orientation × Pair	F(2,34) =	0.015 ,	p = 0.986
Algorithm × Spatial Distribution × Orientation	F(2,34) =	1.311 ,	p = 0.283
Algorithm × Spatial Distribution × Pair	F(4,68) =	1.409 ,	p = 0.240
Algorithm × Orientation × Pair	F(2,34) =	1.124 ,	p = 0.337
Orientation × Spatial Distribution × Pair	F(4,68) =	2.629 ,	p = 0.042
Algorithm × Spat. Dist. × Orientation × Pair	F(4,68) =	1.430 ,	p = 0.234

4.2.2.3. Comparison between spatial distributions

Finally, this section presents the resulting response times for relative judgment of perceived spatial extent between the different levels of Spatial Distribution.

Figure 4.2.2.3 shows the amount of time which the participants took for their judgments between pairs of Spatial Distribution. On average the participants took 6.9 s for one trial. It can be seen that participants took a bit longer for pairwise comparisons with the granular synthesis stimulus than with both other stimuli. Participants also took a little more time for judgments with white noise than with the impulse train stimulus. In general, with some exceptions, judgments in horizontal orientation resulted in marginally longer response time than in vertical orientation.

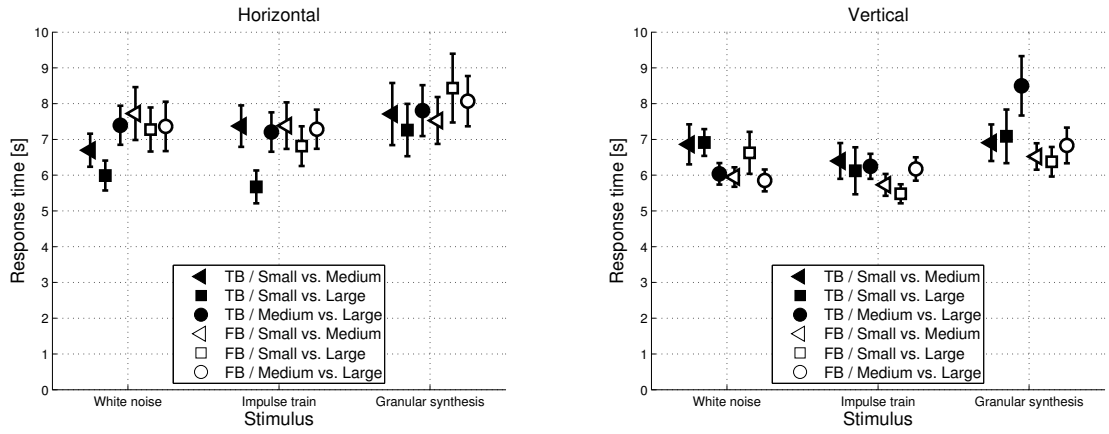


Figure 4.9.: Response time for pairwise comparisons between spatial distributions. Error bars indicate standard error of the mean.

After a logarithmic transform, the response time was assumed to follow a normal distribution and no additional outliers were found in the data.

The statistical analysis of the decadic logarithm of the response time is presented in Table 4.9. Stimulus is the only main effect which showed a significant effect on the response time ($F(2,34)=4.115$, $p=0.025$). Additionally, two interactions were significant (Algorithm × Orientation: $F(1,17)=8.846$, $p=0.009$; Algorithm × Pair: $F(2,34)=3.918$, $p=0.029$). However, since both interactions do not involve any significant main effect, and the time differences are considered relatively small, they are not further investigated.

From the above observations, based on the plots in Figure 4.2.2.3, only the significantly longer response time for the granular synthesis stimulus, compared to both other signals, could be evidenced by the statistical analysis. From the two significant interactions, only the interaction between Algorithm and Orientation could be explained by the fact that in horizontal orientation the frequency-based algorithm led to longer response time than the time-based algorithm, while in vertical orientation it was the other way around. For the interaction between Algorithm and Pair, no interpretation was found.

Table 4.9.: The results of a four-way (Stimulus × Algorithm × Orientation × Pair) repeated measures ANOVA on response time for pairs of Spatial Distribution.

Main effects and interactions			
Stimulus	F(2,34) =	4.115 ,	p = 0.025
Algorithm	F(1,17) =	0.028 ,	p = 0.870
Orientation	F(1,17) =	0.506 ,	p = 0.487
Pair	F(2,34) =	1.781 ,	p = 0.184
Stimulus × Algorithm	F(2,34) =	0.085 ,	p = 0.918
Stimulus × Orientation	F(2,34) =	0.637 ,	p = 0.535
Stimulus × Pair	F(4,68) =	1.824 ,	p = 0.134
Algorithm × Orientation	F(1,17) =	8.846 ,	p = 0.009
Algorithm × Pair	F(2,34) =	3.918 ,	p = 0.029
Orientation × Pair	F(2,34) =	0.533 ,	p = 0.591
Stimulus × Algorithm × Orientation	F(2,34) =	0.097 ,	p = 0.908
Stimulus × Algorithm × Pair	F(4,68) =	0.424 ,	p = 0.791
Stimulus × Orientation × Pair	F(4,68) =	2.179 ,	p = 0.080
Algorithm × Orientation × Pair	F(2,34) =	2.067 ,	p = 0.142
Stimulus × Algorithm × Orientation × Pair	F(4,68) =	0.524 ,	p = 0.718

4.2.3. Discussion

In each trial of the relative judgments of the experiment, the participants were asked to compare two different levels of one of the independent variables Algorithm, Stimulus, Spatial Distribution in terms of perceived spatial extent. Here, the results of this experiment task and the measured response times are summarized and discussed.

The pairwise comparisons between the time-based and the frequency-based algorithm led to the overall result that the time-based algorithm produced significantly larger perceived spatial extent than the frequency-based algorithm. While this overall tendency was only weak (but still significant) in vertical orientation, in horizontal orientation the time-based algorithm clearly outperformed the frequency-based algorithm.

In the case of pairwise comparison between the three different levels of Stimulus (white noise, impulse train, granular synthesis), in horizontal orientation, white noise and impulse train stimuli led to similar and significantly larger perceived spatial extent than the granular synthesis stimulus. It is assumed that this emerges from non-spatial factors affecting the perception of spatial extent. In vertical orientation, the impulse train stimulus resulted in significantly larger spatial extent than both white noise and granular synthesis stimuli. This additionally shows that

the granular synthesis stimulus was judged consistently low for both orientations, which supports the assumption that it is judged based on non-spatial factors. On the other hand, it shows that in vertical orientation the impulse train stimulus clearly outperformed both other stimuli in a pairwise comparison task.

The pairwise comparisons between the three different levels of Spatial Distribution (small, medium, large), are summarized as follows. In general, larger spatial distributions led to significantly larger perceived spatial extent compared to smaller ones. The described tendency was found to be significantly stronger for horizontal orientation compared to vertical orientation, and also for the time-based algorithm compared to the frequency-based algorithm.

Also the response times allow to draw some conclusions. In the case of pairwise comparison between algorithms, significantly increased response time was observed for smaller spatial distributions compared to larger ones. Furthermore, the response time was longer for vertical orientation compared to horizontal orientation. Finally, both white noise and granular synthesis led to significantly longer response time than the impulse train stimulus.

In the case of pairwise comparisons between spatial distributions, the granular synthesis stimulus led to significantly longer response time than both white noise and impulse train stimuli. Additionally, in horizontal orientation, frequency-based algorithm resulted in marginally significantly longer response time than the time-based algorithm, while in vertical orientation it was the other way around.

When linking the aforementioned characteristics of the response times to the relative judgments of spatial extent, the overall impression arises that smaller perceived spatial extent led to significantly longer response time. If it is assumed that participants took longer when it was harder to make a judgment, this leads to the assumption that it was harder to compare pairs of small perceived spatial extent than pairs of large perceived spatial extent.

From the response times of the pairwise comparisons between stimuli, no conclusions can be made. This suggests that it was equally hard for all permutations of Algorithm, Spatial Distribution, and Orientation, to dictate which of two different stimuli invokes a larger perceived spatial extent.

4.3. Comparison of the results from absolute and relative judgments

To effectively compare the results from the two different tasks of the experiment, and to gain more information on their similarities and differences, a more appropriate comparison is needed. This section investigates the hypothesis that rela-

tive judgments of perceived spatial extent are easier than absolute judgments. To achieve this it was chosen to transform the data from the absolute judgments into a comparative form similar to the results of the relative judgments, as a transformation in the opposite direction is not possible.

The transformation of the data is described in Section 4.3.1. Afterwards, the obtained results are structured into three different cases corresponding to the three independent variables Algorithm, Stimulus, and Spatial Distribution, whose levels were compared against each other.

4.3.1. Data transformation

As pointed out in the statistical analysis of the absolute judgments (Section 4.1.1), the judgments of perceived spatial extent in each condition can be assumed to follow a normal distribution. This means, for each combination of Stimulus × Algorithm × Spatial Distribution × Orientation (pooled over all subjects and repetitions), the perceived spatial extent can be interpreted as a Gaussian random variable.

To obtain pairwise comparisons similar to the results of the relative judgments, the probability that a random variable A is larger than another random variable B is computed. According to Equation 4.1 [TG11] this probability is equal to the probability that the difference between A and B is greater than zero.

$$P(A > B) = P(A - B > 0) \tag{4.1}$$

As $(A - B)$ is the difference between to Gaussian random variables, the result also follows a normal distribution. The Cumulative Density Function (CDF) shows the probability that A is smaller than B (see Equation 4.2). The complement of this value therefore describes the probability that A is larger than B.

$$P(A - B > 0) = 1 - CDF(A - B)|_{x=0} \tag{4.2}$$

Now the same pairwise comparisons as performed in the relative judgments of the experiment can be obtained from the absolute judgments of perceived spatial extent. In the case of pairwise comparison between algorithms, the probability that conditions with the time-based algorithm led to larger perceived spatial extent than those with the frequency-based algorithm produces the same data as the one obtained from the relative judgments. In the case of pairwise comparisons between stimuli or spatial distributions, the resulting probabilities were additionally transformed to scale values similar to the results from the relative judgments.

It is mentioned here that the results from both experiments are compared only descriptively, i.e., without statistical analysis.

4.3.2. Comparison between algorithms

In the first case, pairwise comparisons between the levels of the independent variable Algorithm (time-based and frequency-based) were computed from the results of the absolute judgments. The two levels led to only one pairwise comparison for each permutation of the three remaining independent variables Stimulus, Spatial Distribution and Orientation.

Figure 4.3.2 shows the probability that signals with the the time-based algorithm were perceived larger than with the frequency-based algorithm. While the top row shows the results from the absolute judgments of perceived spatial extent, the bottom row recapitulates the results from the relative judgments. The results for horizontal orientation are presented on the left side, while vertical orientation is shown on the right side respectively.

It can be seen that both tasks led to similar results. In particular, there was an overall tendency that the time-based algorithm led to larger perceived spatial extent than the frequency-based algorithm. For both tasks, this tendency was stronger for large spatial distributions compared to smaller ones, and also for horizontal orientation compared to vertical orientation.

In the absolute judgments, this tendency towards the time-based algorithm, concerning larger perceived spatial extent, was even stronger than in the relative judgments. Only the impulse train stimulus with small spatial distribution in horizontal orientation violates this observation, meaning that in this case the preference towards the time-based algorithm was stronger in the relative judgments than in the absolute judgments.

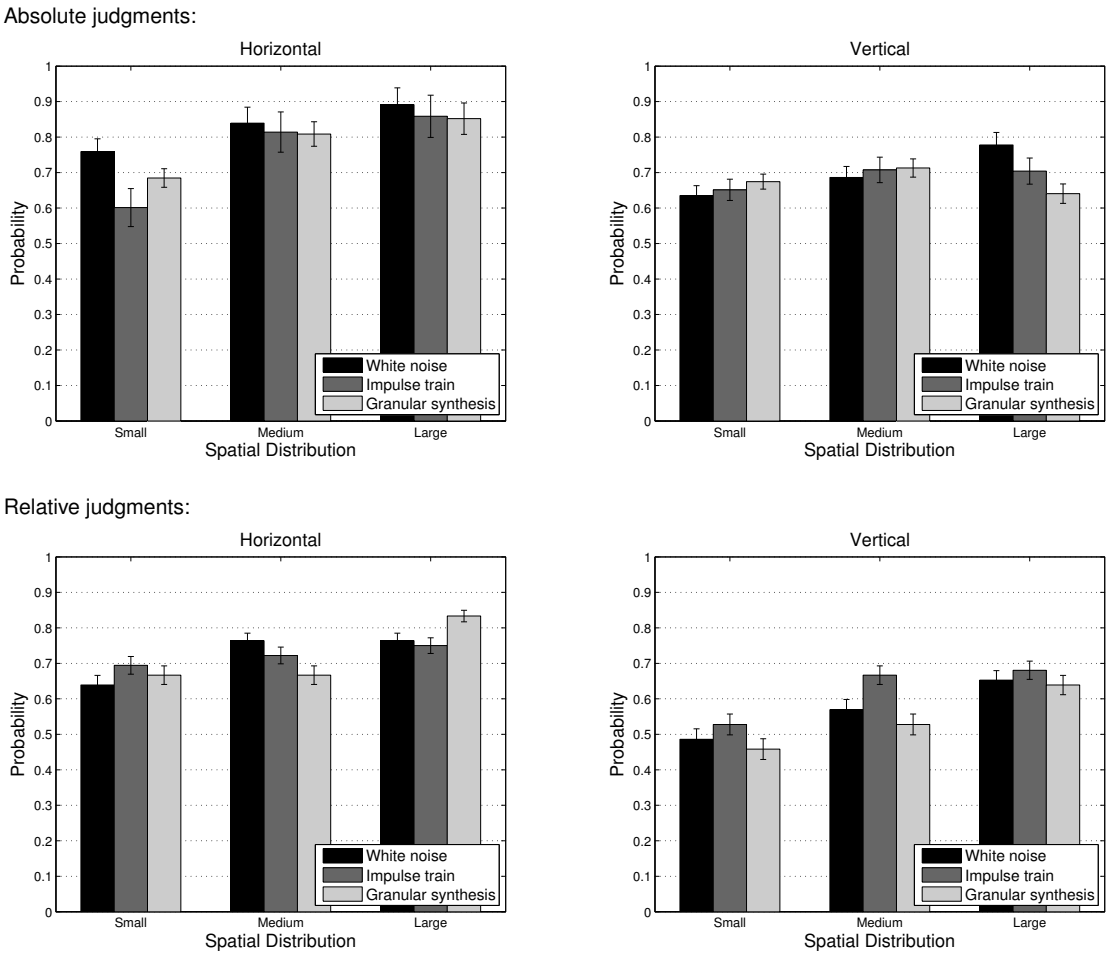


Figure 4.10.: Probability that signals with the time-based algorithm were judged larger than with the frequency-based algorithm, resulting from pairwise comparisons. Top row: transformed data from the absolute judgments. Bottom row: data from the relative judgments. Error bars indicate standard error of the mean.

4.3.3. Comparison between stimuli

Here, the results of the second case are presented. The data of the absolute judgments was transformed to paired comparisons between the three levels of Stimulus (white noise, impulse train, and granular synthesis). The three levels led to three pairwise comparisons for each permutation of the three remaining independent variables Algorithm, Spatial Distribution and Orientation.

Similiar to the results from the relative judgments, scales of Stimuli concerning perceived spatial extent were generated from the pairwise comparisons. The results are shown in Figure 4.3.3. They are presented in two different plots for horizontal and vertical orientation on the left and right side respectively. While the transformed results from the absolute judgments are presented in the top row, the

bottom row recapitulates the results from the relative judgments.

For the absolute judgments, it can be seen that the impulse train stimulus was mostly perceived larger than white noise and both were way more often perceived larger than the granular synthesis stimulus. Both Algorithm and Orientation seem to show no significant effect.

The latter observation differs from the result of the relative judgments in which the white noise signal was perceived significantly smaller in vertical than in horizontal orientation.

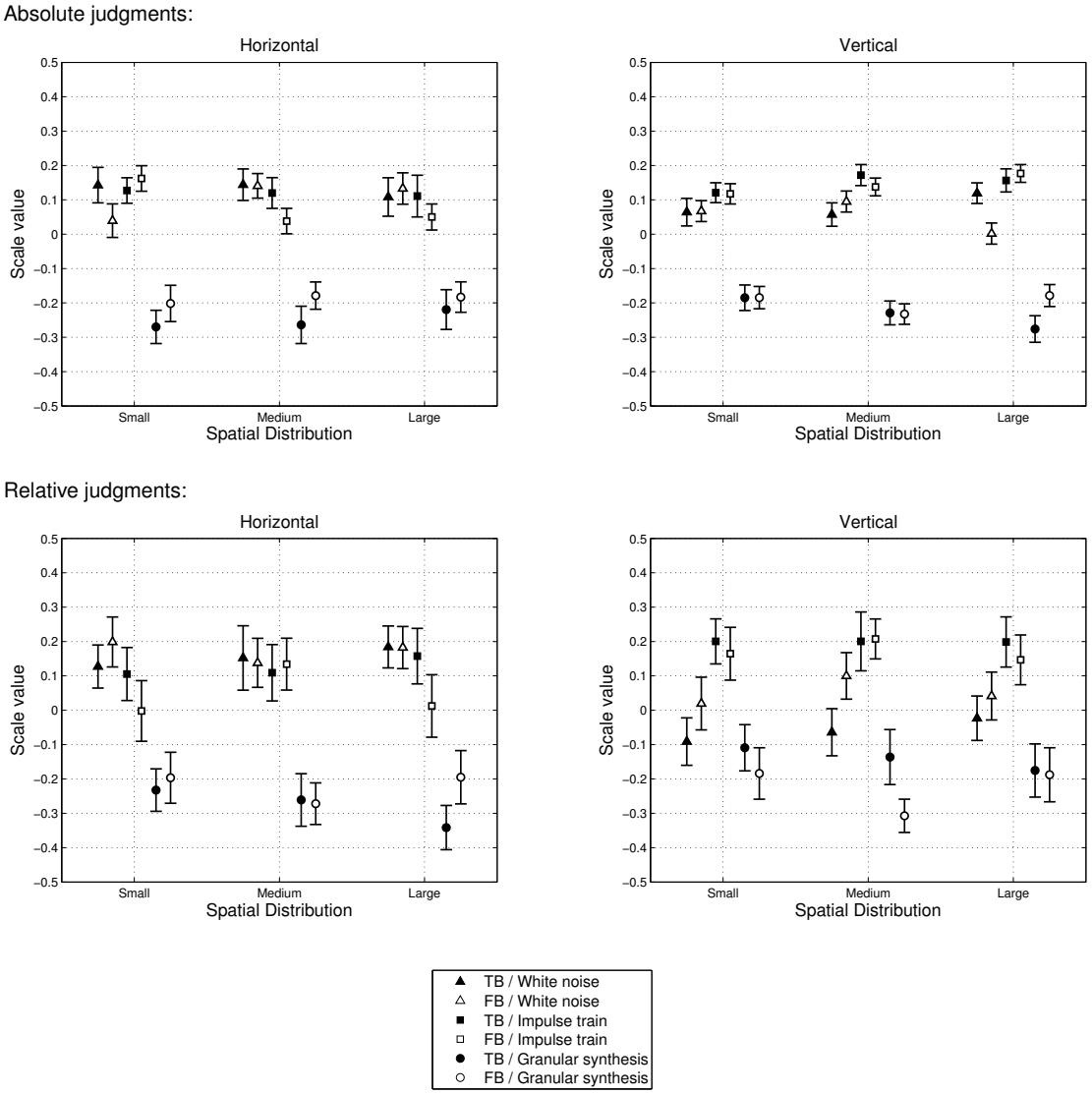


Figure 4.11.: Scales of stimuli, concerning perceived spatial extent, resulting from pairwise comparisons. Top row: transformed data from the absolute judgments. Bottom row: data from the relative judgments. Error bars indicate standard error of the mean.

4.3.4. Comparison between spatial distributions

In the third case, derived from the absolute judgments of perceived spatial extent, the three levels of Spatial Distribution (small, medium, and large) were compared to each other. The three levels led to three pairwise comparisons for each permutation of the three remaining independent variables Algorithm, Stimulus, and Orientation. Out of these, scales of Spatial Distribution were generated according to Thurstone's Case V Model.

The resulting scales of Spatial Distribution, computed from absolute judgments of perceived spatial extent, are shown in top row of Figure 4.3.4. The bottom row recapitulates the results from the relative judgments. While the results for horizontal orientation are presented on the left side, those for vertical orientation are placed on the right side respectively.

In this case, both absolute and relative judgments led to approximately the same results. No relevant differences could be observed.

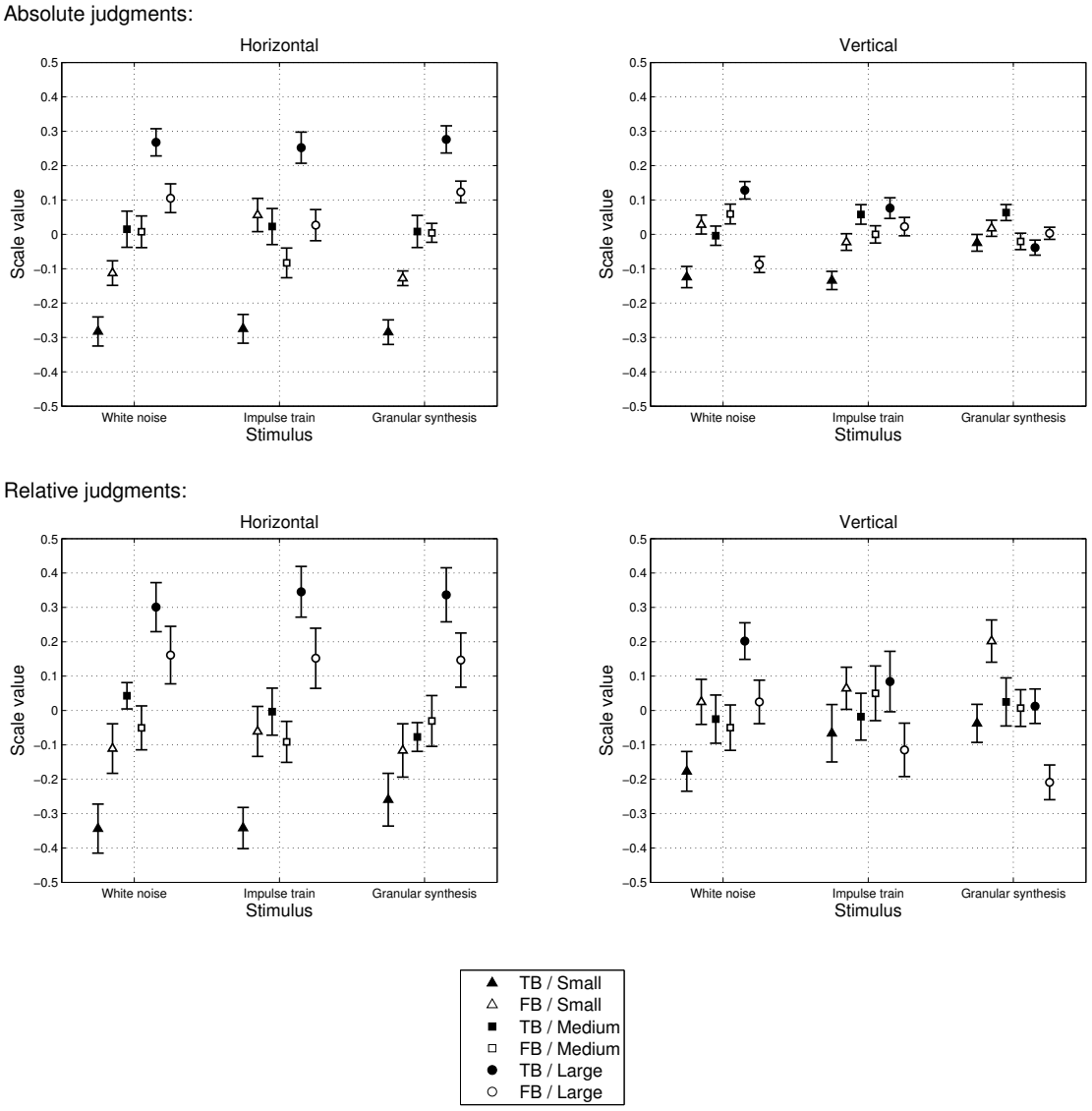


Figure 4.12.: Scales of spatial distributions, concerning perceived spatial extent, resulting from pairwise comparisons. Top row: transformed data from the absolute judgments. Bottom row: data from the relative judgments. Error bars indicate standard error of the mean.

4.3.5. Discussion

Overall, both absolute and relative judgments, led to strongly correlated results in which only small differences could be observed.

In the case of pairwise comparisons between the two levels of factor Algorithm, the overall tendency that the time-based algorithm led to larger perceived spatial extent than the frequency-based algorithm, was found to be stronger for absolute judgments than for relative judgments of perceived spatial extent. However, no interpretation of this difference was found.

The most prominent difference was observed for pairwise comparisons be-

tween the three different levels of Stimulus (white noise, impulse train, granular synthesis). Here, the relative judgments resulted in significantly smaller perceived spatial extent for the white noise stimulus in vertical orientation than in horizontal orientation. In the case of the absolute judgments, no relevant differences between horizontal and vertical orientation could be found for pairwise comparisons between stimuli. This difference could be a consequence of the task during the relative judgments in which participants were forced to decide between two signals with identical physical spatial extent. As it was not made clear to the them, what aspects should be considered for their judgments, this leads to a confound between spatial and non-spatial factors in the auditory perception of spatial extent.

Pairwise comparisons between the three different levels of Spatial Distribution (small, medium, large) led to approximately identical results for both absolute and relative judgments. This observation indicates that absolute and relative judgments of perceived spatial extent are equally hard to accomplish. This leads to the conclusion that the hypothesis of relative judgments being easier than absolute judgments was actually wrong and could not be evidenced by the experiment.

4.4. Acoustic measurements

To gather additional information, some acoustic measurements were performed with the experiment apparatus. The main hypotheses, articulated in Section 2.9 presume that the experiment is performed under far-field conditions, which means that there is no reflected sound. However, the experiment took place in an acoustically treated but not anechoic room. For further evidence that the results and conclusions from the experiment are correct, it was therefore necessary to perform some acoustic measurements. In particular, the reverberation time, critical distance, and lateral energy fraction (LF) are discussed in Sections 4.4.1, 4.4.2, and Section 4.4.3. These measurements were also performed to investigate the overestimation of perceived spatial extent for small spatial distributions.

Additionally, the frequency-responses of the loudspeakers were measured to verify their accuracy in reproduction of the different stimuli (Section 4.4.4).

Finally, the inter-aural cross-correlation (IACC) at the listening position was measured for all conditions of the experiment. This measure is investigated in Section 4.4.5, as according to the literature it should act as a good predictor of perceived spatial extent in horizontal orientation.

4.4.1. Reverberation time

The reverberation time T60, which is the time to drop 60 dB below the original sound pressure level [Dic97, p. 31], was determined with two different methods.

First, the reverberation time was measured directly with an NTi measurement kit consisting of Minirator MR2 signal generator, a dodecahedron loudspeaker, NTi M2210 microphone and NTi XL2 analyzer. The results in Table 4.4.1 show the average of six measurement cycles. The measurements were performed with the interrupted noise method using pink noise.

Additionally, the reverberation time was calculated from impulse response measurements, taken with an omnidirectional microphone (NTi M2210) at the listening position and an exponential sine-sweep played back from the loudspeakers of the experiment. Table 4.4.1 shows the measured reverberation time for the center and outmost speakers of the array in octave bands. The microphone was positioned on-axis with the center loudspeaker. The reverberation time was extrapolated from the time for the energy drop from -5 to -35 dB.

It can be seen that both methods led to similar values for the reverberation time.

Table 4.10.: Measured reverb time RT20 in seconds with a dodecahedron loudspeaker (in 1/3-octave bands), with standard deviations. RT20 means, only the time for the energy drop of 20 db from -5 to -25 dB was actually measured and extrapolated to 60 dB afterwards.

f [Hz]	RT20 [s]	SD [s]
50	0.66	0.28
63	0.41	0.07
80	0.65	0.20
100	0.60	0.17
125	0.42	0.25
160	0.32	0.04
200	0.22	0.06
250	0.22	0.02
315	0.20	0.01
400	0.20	0.04
500	0.23	0.03
630	0.22	0.03
800	0.19	0.02
1,000	0.19	0.01
1,250	0.20	0.03
1,600	0.20	0.04
2,000	0.20	0.02
2,500	0.19	0.01
3,150	0.19	0.02
4,000	0.18	0.01
5,000	0.17	0.01
6,300	0.15	0.01
8,000	0.16	0.02
10,000	0.18	0.01

Table 4.11.: Reverb time RT30 in seconds, computed from impulse responses of the outmost and center loudspeakers.

Octave band	Left	Center	Right	Top	Bottom
250 Hz	0.22	0.20	0.21	0.22	0.20
500 Hz	0.19	0.17	0.20	0.18	0.20
1 kHz	0.20	0.19	0.19	0.22	0.20
2 kHz	0.20	0.19	0.19	0.20	0.19
4 kHz	0.18	0.19	0.18	0.19	0.17
8 kHz	0.17	0.18	0.17	0.17	0.17

Because of the high-pass filtering, only the reverberation time above 200 Hz is considered relevant for the performed experiment. In this range, reverberation

times between 0.16 and 0.22 s were measured. While they are obviously higher than those of an anechoic chamber, at least they are below the suggested reverberation time for speech production studios (0.3 s) [Dic97, p. 35]. However, the reverberation time alone is not seen as a sufficient predictor for the influence of the room acoustics on the auditory event perceived by the listener.

4.4.2. Critical distance and direct-to-reverberant-ratio (DRR)

The influence of reverberation on the auditory event is dependent on the distance between the sound source and the listener. The ratio between direct sound and reverberant sound decreases with increasing distance. At the critical distance, or reverb radius, this ratio equals to '1'.

An approximation of the critical distance can be calculated with Equation 4.3 [Dic97, p. 33]:

$$d_c = 0.057 \cdot \sqrt{\gamma} \cdot \sqrt{\frac{V}{T60}} \tag{4.3}$$

where γ is the directivity factor (equal to '1' for omnidirectional sound sources), V is the room volume and $T60$ the reverb time. For omnidirectional sound sources, this leads to a mean critical distance of about 1.22 m. The participants were always located in 2 m distance to the loudspeaker array. However, the loudspeakers used in the described experiment have a strong directivity, especially for high frequencies, which expands the critical distance. Unfortunately the directivity factor is unknown for the custom loudspeakers used in the experiment.

In this case, the influence of the reverberant sound is better described by the direct-to-reverberant-ratio (DRR) at the listening position, which is computed from the impulse response $h(t)$ as follows in Equation 4.4 [Zah02]:

$$DRR = 10 \cdot \log_{10} \left(\frac{\int_{t_0 - t_c}^{t_0 + t_c} h^2(t)\,\mathrm{d}t}{\int_{t_0 + t_c}^{\infty} h^2(t)\,\mathrm{d}t} \right) \tag{4.4}$$

where t_c is a correction constant, usually set to 2.5 ms [Zah02].

Table 4.12 shows the resulting DRR at the listening position for the four outmost loudspeakers and the center loudspeaker with two different correction constants. The measured DRR was found to be always positive, which means that there was more direct sound than reverberant sound at the listening position. In the case of the standard correction constant (2.5 ms) the reverberant sound is between 7 and 11 dB lower than the direct sound. It is argued here, that this difference is large enough to assume an marginal influence of the reverberation on the perceived

sound.

Table 4.12.: Direct-to-reverberant ratio (DRR) at the listening position in dB for two different correction constants.

correction constant	left	center	right	top	bottom
2.5 ms	7.84	9.49	8.33	9.10	11.21
0.25 ms	4.83	3.05	5.58	6.46	5.02

4.4.3. Lateral energy fraction (LF)

In the literature, the lateral energy fraction (LF) is often seen as a measure for auditory source width. In the context of this thesis, it is therefore used to investigate the auditory source width of a single loudspeaker in the used experimental setup. The lateral energy fraction describes the ratio of the lateral energy to the total energy, and is computed from the impulse responses measured by an omnidirectional microphone $h(t)$ as well as a figure-eight microphone $h_L(t)$, as shown in Equation 4.5 [MA81]:

$$LF = \frac{\int_{t_d}^{t_e} {h_L}^2(t)\,\mathrm{d}t}{\int_{0}^{t_e} h^2(t)\,\mathrm{d}t} \tag{4.5}$$

where t_d is the direct time, usually set to 5 ms [ISO09], and t_e is the early time, usually set to 80 ms [CDL93].

A correction factor α was introduced to calibrate the two microphones relatively to themselves (see Equation 4.6):

$$\alpha = \frac{\int_{t_0-t_c}^{t_0+t_c} h^2(t)\,\mathrm{d}t}{\int_{t_0-t_c}^{t_0+t_c} {h_L}^2(t)\,\mathrm{d}t} \tag{4.6}$$

where t_0 is the time of the greatest peak in the impulse response and t_c is a constant time difference. From the visual waveform representation of the impulse response, the peak was identified to be around 5 ms wide. Therefore t_c was set to 0.25 ms. This leads to the gain-compensated Equation 4.7. The LF for the center and four outmost loudspeakers, measured at the listening position, is shown in Table 4.13. The measurement was performed with an NTi M2210 omnidirectional microphone and a Schoeps type CMC 5 with MK 8 capsule as

a figure-eight microphone, connected to an RME Fireface 400 audio interface running at 44.1 kHz sampling rate / 24 bit resolution.

$$LF = \alpha \cdot \frac{\int_{t_d}^{t_e} h_L(t)\,\mathrm{d}t}{\int_{0}^{t_e} h(t)\,\mathrm{d}t} \tag{4.7}$$

Table 4.13.: Measured LF for the outmost and center loudspeakers.

Orientation	left/top	center	right/bottom
Horizontal	0.10	0.07	0.09
Vertical	0.09	0.08	0.11

For phantom sources in the horizontal plane, it was shown that an adapted version of the LF, with the direct time t_d set to zero, is a better predictor of auditory source width, at least for pink noise signals [Fra13]. Table 4.14 shows the results with the adapted equation.

Table 4.14.: Adapted LF for the outmost and center loudspeakers.

Orientation	left/top	center	right/bottom
Horizontal	0.14	0.07	0.12
Vertical	0.17	0.12	0.17

For an interpretation of these values, the Just-Noticeable Difference (JND) for the LF needs to be considered. According to the literature, the JND for the LF lies in the range of 0.045 to 0.07 [Bla02]. The obtained values for the LF from Tables 4.13 and 4.14 quantified in measures of JND's result in 2-3 JND's for the standard LF and 2-4 JND's for the adapted LF. It is therefore assumed that the lateral energy fraction for a single loudspeaker, measured at the listening position, was perceivable and could have led to an increased auditory source width.

4.4.4. Frequency response of the loudspeakers

Figure 4.13 (left) shows the frequency response of the center speaker, measured at the listening position. The measurement was taken with an omnidirectional microphone (NTi M2210) at the listening position with the exponential sine sweep method and three averages, taking only the direct-part of the impulse response.

The spectrum is averaged in 1/3-octave bands. The microphone was connected to an RME Fireface 400 audio interface running at 44.1 kHz sampling rate / 24 bit resolution.

On the right, Figure 4.13 shows the differences of the four outmost speakers to the frequency response of the center speaker. The magnitude loss towards high frequencies of the outmost loudspeakers, compared to the center, can be explained due to the greater distance to the listening position. Other differences occur as a consequence of the non-ideal room-acoustics and also variances between the loudspeakers themselves.

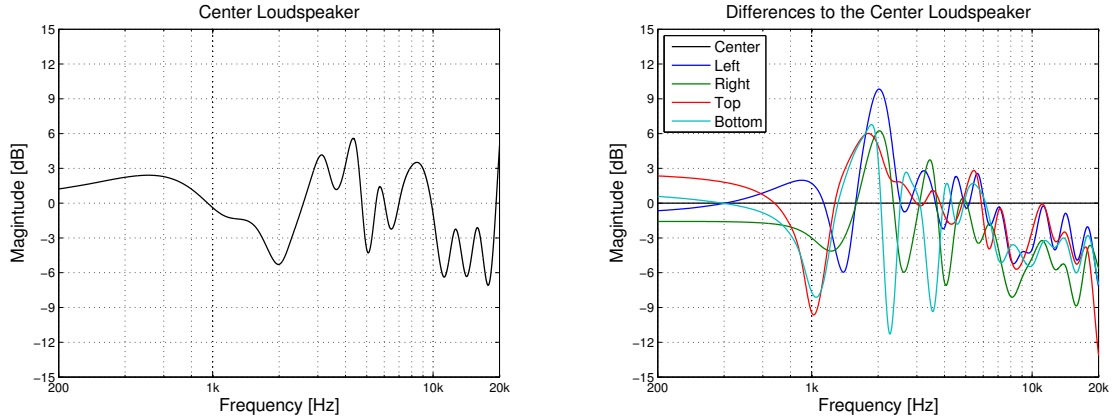


Figure 4.13.: Frequency response of the center loudspeaker (left) and the difference of the four outmost loudspeakers to this reference (right).

The center speaker shows a peak at 20 kHz, which looks like an error, but also appears on the frequency-response of the drivers alone, without enclosure, published by the manufacturer (see Figure 4.14) [Tym15].

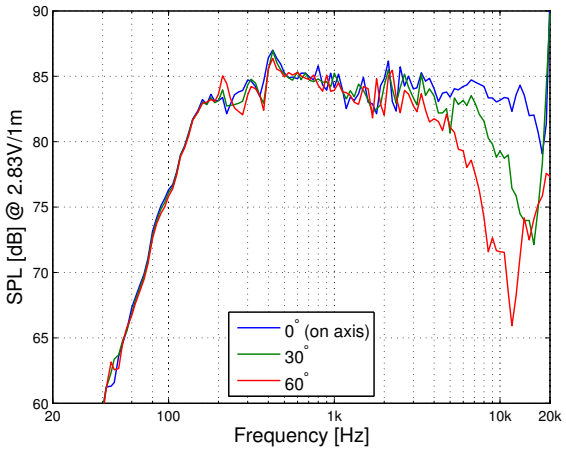


Figure 4.14.: Frequency response of the Peerless PLS-P830983 driver without enclosure, as measured by the manufacturer [Tym15].

4.4.5. Inter-aural cross-correlation (IACC)

In previous studies, the inter-aural cross-correlation (IACC) was shown to be a good predictor for perceived spatial extent in the horizontal plane [MBR05]. However, it seems that the time-based algorithm has not been evaluated yet, concerning its effect on the IACC. Therefore, binaural measurements were performed with a head and torso simulator (HATS, B&K type 4128C), which was placed roughly at the listening position. The microphone outputs were fed through a conditioning amplifier (B&K NEXUS type 2690) set to 1 V/Pa and recorded through an RME Fireface 400 audio interface running at 44.1 kHz sampling rate / 24 bit resolution. The inter-aural cross-correlation coefficient (IACC), which is the maximum of the cross-correlation between the left and right channel of the binaural recording [Bla97], was measured for all combinations of the independent variables Stimulus, Algorithm, Spatial Distribution, and Orientation. The results are shown in Figure 4.15.

For horizontal orientation, as expected, larger spatial distributions produced a lower IACC than smaller ones for all combinations of Stimulus and Algorithm. Also the time-based algorithm resulted in a lower IACC than the frequency-based algorithm. This difference between the time-based and frequency-based algorithm increases with increasing spatial distribution (from small to large). It can also be seen that the granular synthesis signal produced a higher IACC than both white noise and impulse train stimuli. While this tendency is clear for large and medium spatial distributions, it is only marginal for small spatial distributions. The measurement for vertical orientation was only done to verify the results and shows equally high IACC for all conditions.

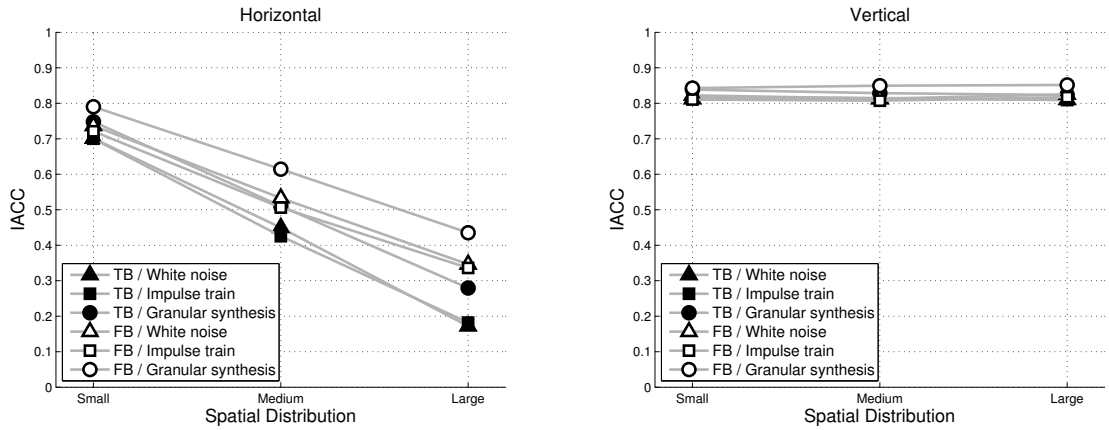


Figure 4.15.: IACC for all combinations of Stimulus × Algorithm × Spatial Distribution × Orientation.

For horizontal orientation, the corresponding time-delays of the cross-correlation

maxima are shown in Table 4.15 for the time-based and in Table 4.16 for the frequency-based algorithm. The time-delays for vertical orientation were always less than or equal to one sample (0.02 ms). All these delays are considered low enough to be irrelevant.

Table 4.15.: Time-delays of the maxima of the IACC plotted in Figure 4.15. Horizontal orientation, time-based algorithm.

	White noise	Impulse train	Granular synthesis
Small	0.00	0.02	0.00
Medium	0.02	0.00	0.00
Large	0.00	0.00	-0.16

Table 4.16.: Time-delays of the maxima of the IACC plotted in Figure 4.15. Horizontal orientation, frequency-based algorithm.

	White noise	Impulse train	Granular Synthesis
Small	0.02	0.02	0.02
Medium	-0.02	-0.02	-0.02
Large	-1.43	-1.43	-1.43

4.4.6. Discussion

According to the results of the acoustic measurements, several conclusions can be drawn.

The influence of the room acoustics can be summarized as follows. The measured reverberation times in the relevant frequency range above 200 Hz lay between 0.16 and 0.22 s. The direct-to-reverberant ratio (DRR) at the listening position was between 3 and 11 dB for the individual loudspeakers and different correction constants. It is argued here that the measured DRR is large enough to assume a low influence of the reverberation on the results of the experiment. It is therefore assumed that the overestimation of perceived spatial extent is not primarily induced by the reverberation.

Concerning the adapted lateral energy fraction (LF) of the individual loudspeakers, values between 0.07 and 0.17 were measured at the listening position, roughly corresponding to 2-4 just-noticeable differences (JND). These are therefore perceivable and could probably have induced an increased auditory source width. In previous studies, comparable amounts of LF resulted in auditory source width which was even larger than the small spatial distribution in the described

experiment. It is argued here that the overestimation of horizontal spatial extent in the absolute judgments of perceived spatial extent could at least partially originate from this relatively high value of lateral energy fraction.

The frequency response of the loudspeakers was found to be by no means flat and there were strong differences of up to ± 10 dB between the spectra of the individual loudspeakers. However, these results are based on only one measurement per loudspeaker, from a distance of 2 m, and were not averaged over several positions in the area of the participant's ears. The differences are therefore probably evoked by room resonances, which are supposed to change with head movements of the participants. In conjunction with the broadband character of the used stimuli, the influence of the low spectral flatness of the loudspeakers on the perceived spatial extent is therefore considered marginal. However, a significant effect on the perceived center can not be excluded.

In compliance with the literature, the IACC was found to be a relatively good predictor for perceived spatial extent in the horizontal plane, as it is highly correlated with the absolute judgments of perceived spatial extent (compare perceived spatial extent in Figure 4.1 and IACC in Figure 4.15). In particular, larger spatial distributions led to a lower IACC than smaller ones, and the time-based algorithm resulted in lower values of the IACC than the frequency-based algorithm. The difference of the IACC between signals with the time-based algorithm and the frequency-based algorithm increases with increasing spatial distribution. Furthermore, also correlated with perceived spatial extent, the granular synthesis stimulus led to always higher IACC than both white noise and impulse train stimuli, who produced similar values. The measurement of the IACC in vertical orientation was only done to verify the results and shows equally high IACC for all conditions. The insignificance of the IACC for vertical localization and therefore also for the perception of vertical spatial extent was already known from the literature.

5. Conclusion

Concluding, the representation of auditory vertical extent and subsequently complex auditory spatial patterns may be difficult to achieve, at least not in conjunction with the small spatial extents tested here. Possible solutions may however emerge within the realm of time-based algorithms for synthesis of spatial extent. Larger extents and listener training may be interesting factors to vary in future experiments and sound design may be a critical factor here.

Concerning the auditory representation of horizontal extent the results are very promising. Participants could differentiate well even in response to the small spatial extents tested here. Designers may therefore start to integrate horizontally extended sounds in virtual and mixed reality applications as well as auditory displays to support distance perception, target acquisition, and maybe even the spatial display of information in auditory graphs. When extents of small magnitude need to be used, time-based extent synthesis algorithms are preferable, as they yield larger impressions of horizontal extent. The use of granular synthesis in this context appears to be a far reaching solution for sound and interface designers.

Overall, the results of this study seem promising for creating interactive systems in which horizontally extended sounds are displayed. These may have applications in different fields related to Sonic Interaction Design (SID). The poor performance of the frequency-based algorithm in the horizontal plane could be due to the relatively small spatial extents compared to the previous experiments [PSP14].

With the tested signals and algorithms the performance in perception of vertical spatial extent seems not to be enough for use in practical applications. However, horizontal spatial extent would suffice for many applications in interactive VR or AR environments. As sound localization works perfectly well in the vertical plane, multiple horizontally extended sound sources distributed in the vertical plane could be useful for data sonification or auditory displays. One possible application could be an auditory display of vertically arranged horizontal bars of different lengths to sonify different kinds of information.

The subjective impression that relative judgments were more easy than absolute judgments could not be verified. Both experiments delivered similar results and therefore could be performed interchangeably to gain the same knowledge.

However, since the participants always performed the relative judgments first, it is assumed that they were effectively trained once they started with the absolute judgments. This could be an explanation for the small differences between the results of both tasks. The performed comparison of the two experiment tasks could be helpful for future investigations also regarding other fields of research whenever a decision about absolute or relative experiment design needs to be made.

The absolute judgments from the first part of the experiment facilitate to compute all possible pairwise comparisons, also those which were not performed in the second part. These would have been very time consuming in a real pairwise comparison task. One absolute judgments in part one took 11.0 s while one pairwise comparison in the second part of the experiment took approximately 6.5 s (pauses between different judgments are included). On average, one pass of 36 absolute judgments therefore took 6 min 36 s, while one pass of the 90 judgments from the second part took 9 min 45 s. All possible pairwise comparisons including the ones not done in the second part (without repetitions) would take more than one hour. The presented results therefore suggest that in some cases absolute judgments can clearly outperform a 2AFC experiment design in terms of expenditure of time and also amount of gathered information. The advantage arises especially with many independent variables.

6. Outlook

This work is seen as a small step into the direction of being able to create sound sources with specific horizontal and vertical spatial extent. Especially in the vertical plane, the knowledge on auditory perception of spatial extent shows still deficits..

Also in the performed experiment, there are still some factors which need further investigation. Spatial distributions of individual point-shaped sound sources were created. However, it is still unknown, how the spatial density of these point sources affects spatial extent perception. There seems to be an upper limit which is given by the just noticeable difference in sound localization, and which is far smaller in the horizontal plane than in the vertical plane. There must exist also a lower limit of sound-source density, where the impression of a coherent spatially distributed sound source vanishes. This could apply for both, time- and frequency-based algorithms. At the time of writing, there seems to be no study which investigated this aspect.

Another factor which needs further research is the inter-onset interval of the granular synthesis with the time-based algorithm. As with the spatial density, there must be a lower limit, where the impression of a coherent spatially distributed sound source breaks. A lower limit could be defined by the just-noticeable difference in sound localization, where a better spatial resolution is not necessary anymore.

It is known from the literature that the bandwidth of the stimulus has a strong influence on the perceived spatial extent. However, there is little scope left for sound design, if only signals with white spectra are chosen. The other way around, sound design is also important for the perceived spatial extent, independent of the used algorithm. Especially regarding non-spatial factors, sound design plays an important role and needs to be considered in the process of creating spatially extended sound sources.

Bibliography

[AAD01] V. Ralph Algazi, Carlos Avendano, and Richard O. Duda. “Elevation localization and head-related transfer function analysis at low frequencies”. In: *The Journal of the Acoustical Society of America* 109.3 (2001), pp. 1110–1122. DOI: http://dx.doi.org/10.1121/1.1349185. URL: http://scitation.aip.org/content/asa/journal/jasa/109/3/10.1121/1.1349185.

[ALS11] R. Albrecht, T. Lokki, and L. Saviola. “A Mobile Augmented Reality Audio System with Binaural Microphones”. In: *Proceedings of IWS*. Stockholm, Sweden, 2011, pp. 7–11.

[AS11] Jens Ahrens and Sascha Spors. “Two Physical Models for Spatially Extended Virtual Sound Sources”. In: *Audio Engineering Society Convention 131*. 2011. URL: http://www.aes.org/e-lib/browse.cfm?elib=16009.

[Bar02] Natasha Barrett. “Spatio-musical composition strategies”. In: *Organised Sound* 7 (03 Dec. 2002), pp. 313–323. ISSN: 1469-8153. DOI: 10.1017/S1355771802003114. URL: http://journals.cambridge.org/article_S1355771802003114.

[Bar07] Hans-Jochen Bartsch. *Taschenbuch Mathematischer Formeln*. 21st ed. Fachbuchverlag Leipzig im Carl Hanser Verlag München, 2007.

[Bau69] Benjamin B. Bauer. “Some Techniques Toward Better Stereophonic Perspective”. In: *J. Audio Eng. Soc* 17.4 (1969), pp. 410–415. URL: http://www.aes.org/e-lib/browse.cfm?elib=1572.

[BK04] Maurice Boueri and Chris Kyriakakis. “Audio Signal Decorrelation Based on a Critical Band Approach”. In: *Audio Engineering Society Convention 117*. 2004. URL: http://www.aes.org/e-lib/browse.cfm?elib=12948.

[BL86] J. Blauert and W. Lindemann. “Spatial mapping of intracranial auditory events for various degrees of interaural coherence”. In: *J. Acoust. Soc. Am.* 79.3 (1986), pp. 806–813.

[Bla02] Matthias Blau. “Difference limens for measures of apparent source width”. In: *Forum Acusticum*. Sevilla, Spain, 2002.

[Bla11] Sebastian Blamberger. *Development of a modular system for speaker array prototyping*. Tech. rep. Institute of Electronic Music, Acoustics, Graz University of Music, and Performing Arts, 2011.

[Bla97] Jens Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. Revised Edition. MIT Press, 1997.

[Bre+03] S. Brewster, J. Lumsden, M. Bell, M. Hall, and S. Tasker. “Multimodal’eyes-free’interaction techniques for wearable devices”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. 2003, pp. 473–480.

[CAKP98] Claudia Carello, Krista L. Anderson, and Andrew J. Kunkler-Peck. “Perception of Object Length by Sound”. In: *Psychological Science* 9.3 (1998), pp. 211–214.

[CDL93] Trevor J. Cox, W. J. Davies, and Yiu W. Lam. “The Sensitivity of Listeners to Early Sound Field Changes in Auditoria”. In: *Acustica* 79.1 (1993), pp. 27–41.

[Cow84] Nelson Cowan. “On short and long auditory stores”. In: *Psychological Bulletin* 96.2 (1984), pp. 341–370.

[CT03] Densil Cabrera and Steven Tilley. “Parameters for auditory display of height and size”. In: *Proceedings of the 9th International Conference on Auditory Display (ICAD)*. Ed. by Eoin Brazil and Barbara Shinn-Cunningham. International Community for Auditory Display. Boston, MA: Georgia Institute of Technology, 2003.

[Dic97] Michael Dickreiter. *Handbuch der Tonstudiotechnik*. Ed. by SRT Schule für Rundfunktechnik. 6th ed. Vol. 1. K.G. Saur, 1997.

[DS09] Etienne Deleflie and Greg Schiemer. “Spatial Grains: Imbuing Granular Particles With Spatial-Domain Information”. In: *Proceedings of the Australasian Computer Music Conference ACMC09*. 2009.

[Fra13] Matthias Frank. “Source Width of Frontal Phantom Sources: Perception, Measurement, and Modeling”. In: *Archives of Acoustics* 38.3 (Oct. 2013), pp. 311–319.

[Ger92] Michael A. Gerzon. “Signal Processing for Simulating Realistic Stereo Images”. In: *Audio Engineering Society Convention 93*. 1992. URL: http://www.aes.org/e-lib/browse.cfm?elib=6711.

[GM06] Bruno L. Giordano and Stephen McAdams. “Material identification of real impact sounds: effects of size variation in steel, glass, wood, and plexiglass plates”. In: *The Journal of the Acoustical Society of America* 119.2 (2006), pp. 1171–1181.

[GM13] Daniel Grieser and Svenja Maronna. “Hearing the Shape of a Triangle”. In: *Notices of the American Mathematical Society* 16.11 (2013), pp. 1440–1447.

[GM90] B. R. Glasberg and B. C. J. Moore. “Derivation of auditory filter shapes from notched-noise data”. In: *Hearing Research* 47.1-2 (1990), pp. 103–138.

[Gru50] Frank E. Grubbs. “Sample Criteria for Testing Outlying Observations”. In: *Ann. Math. Statist.* 21.1 (Mar. 1950), pp. 27–58. DOI: 10.1214/aoms/1177729885. URL: http://dx.doi.org/10.1214/aoms/1177729885.

[GW96] Carolyn Gordon and David Webb. “You Can’t Hear the Shape of a Drum”. In: *American Scientist* 84 (1996), pp. 46–55.

[Hal64] J. H. Halton. *Algorithm 247: Radical-inverse Quasi-random Point Sequence*. New York, NY, USA, Dec. 1964. DOI: 10.1145/355588.365104. URL: http://doi.acm.org/10.1145/355588.365104.

[HF94] Ari J. Hollander and Thomas A. Furness. “Perception of Virtual Auditory Shapes”. In: *Proceedings of the International Conference on Auditory Display*. 1994.

[HKH02] Koichiro Hiyama, Setsu Komiyama, and Kimio Hamasaki. “The Minimum Number of Loudspeakers and its Arrangement for Reproducing the Spatial Impression of Diffuse Sound Field”. In: *Audio Engineering Society Convention 113*. 2002. URL: http://www.aes.org/e-lib/browse.cfm?elib=11272.

[HL56] Franz Schlegel Holger Lauridsen. “Stereofonie und richtungsdiffuse Klangwiedergabe”. In: *Gravesaner Blätter* 5 (1956), pp. 28–50.

[HMG02] Simon Holland, David R. Morse, and Henrik Gedenryd. “AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface”. English. In: *Personal and Ubiquitous Computing* 6 (4 2002), pp. 253–259. DOI: 10.1007/s007790200025. URL: http://dx.doi.org/10.1007/s007790200025.

[HP06a] Toni Hirvonen and Ville Pulkki. “Center and Spatial Extent of Auditory Events as Caused by Multiple Sound Sources in Frequency-Dependent Directions”. In: *Acta Acoustica united with Acoustica* 92 (2006), pp. 320–330.

[HP06b] Toni Hirvonen and Ville Pulkki. “Perception and Analysis of Selected Auditory Events with Frequency-Dependent Directions”. In: *J. Audio Eng. Soc* 54.9 (2006), pp. 803–814. URL: http://www.aes.org/e-lib/browse.cfm?elib=13902.

[HP08] Toni Hirvonen and Ville Pulkki. “Perceived Spatial Distribution and Width of Horizontal Ensemble of Independent Noise Signals as Function of Waveform and Sample Length”. In: *Audio Engineering Society Convention 124*. 2008. URL: http://www.aes.org/e-lib/browse.cfm?elib=14538.

[IEC13] IEC. *61672-1:2013: Electroacoustics - Sound level meters - Part 1: Specifications.* 2013.

[ISO09] ISO. *3382-1:2009: Acoustics - measurement of room acoustic parameters - part 1: Performance spaces.* 2009.

[Kac66] Mark Kac. “Can One Hear the Shape of a Drum?” In: *American Mathematical Monthly* 73.4/2 (1966), pp. 1–23.

[Ken95] Gary S. Kendall. “The Decorrelation of Audio Signals and Its Impact on Spatial Imagery”. In: *Computer Music Journal* 19.4 (1995), pp. 71–87. ISSN: 01489267, 15315169. URL: http://www.jstor.org/stable/3680992.

[KPT00] A. J. Kunkler-Peck and M. T. Turvey. “Hearing shape”. In: *J Exp Psychol Hum Percept Perform* 26.1 (2000), pp. 279–294.

[Lai+12] Mikko-Ville Laitinen, Tapani Pihlajamäki, Cumhur Erkut, and Ville Pulkki. “Parametric Time-frequency Representation of Spatial Sound in Virtual Worlds”. In: *ACM Trans. Appl. Percept.* 9.2 (June 2012), 8:1–8:20. ISSN: 1544-3558. DOI: 10.1145/2207216.2207219. URL: http://doi.acm.org/10.1145/2207216.2207219.

[Lak93] S. Lakatos. “Recognition of Complex Auditory-Spatial Patterns”. In: *Perception* 22 (1993), pp. 363–374.

[Lil67] Hubert W. Lilliefors. “On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown”. In: *Journal of the American Statistical Association* 62.318 (1967), pp. 399–402. ISSN: 01621459. URL: http://www.jstor.org/stable/2283970.

[MA81] Barron M and Marshall AH. “Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure”. In: *Journal of Sound Vibration* 77.2 (1981), pp. 211–232.

[Mar+06] James R. Marston, Jack M. Loomis, Roberta L. Klatzky, Reginald G. Golledge, and Ethan L. Smith. “Evaluation of spatial displays for navigation without sight”. In: *ACM Transactions on Applied Perception* 3 (2006), pp. 110–124.

[MBR05] R. Mason, T. Brookes, and F. Rumsey. “Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli”. In: *J. Acoust. Soc. Am.* 117.3 Pt 1 (2005), pp. 1337–1350.

[MG96] B. C. J. Moore and B. R. Glasberg. “A revision of Zwicker’s loudness model”. In: *Acustica United with Acta Acustica* 82.2 (1996), pp. 335–345.

[ML14] Georgios Marentakis and Rudolf Liepins. “Evaluation of Hear-through Sound Localization”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’14. Toronto, Ontario, Canada: ACM, 2014, pp. 267–270. ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557168. URL: http://doi.acm.org/10.1145/2556288.2557168.

[MM02] E. A. Macpherson and J. C. Middlebrooks. “Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited”. In: *J. Acoust. Soc. Am.* 111.5 Pt 1 (2002), pp. 2219–2236.

[MR09] Alexander Müller and Rudolf Rabenstein. “Physical modeling for spatial sound synthesis”. In: *Proc. of the 12th Int. Conference on Digital Audio Effects (DAFx-09)*. Como, Italy, 2009.

[Mül+14] Jörg Müller, Matthias Geier, Christina Dicke, and Sascha Spors. “The boomRoom: mid-air direct interaction with virtual sound sources”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2014, pp. 247–256.

[Orb70a] Robert Orban. “A Rational Technique for Synthesizing Pseudo-Stereo from Monophonic Sources”. In: *J. Audio Eng. Soc* 18.2 (1970), pp. 157–164. URL: http://www.aes.org/e-lib/browse.cfm?elib=1514.

[Orb70b] Robert Orban. “Further Thoughts on quot;A Rational Technique for Synthesizing Pseudo-Stereo from Monophonic Sourcesquot;” in: *J. Audio Eng. Soc* 18.4 (1970), pp. 443, 444. URL: http://www.aes.org/e-lib/browse.cfm?elib=1484.

[OS09] Alan V. Oppenheim and Ronald W. Schafer. *Discrete-Time Signal Processing*. 3rd. Upper Saddle River, NJ, USA: Prentice Hall Press, 2009. ISBN: 0131988425, 9780131988422.

[PB03] Guillaume Potard and Ian Burnett. “A study on sound source apparent shape and wideness”. In: *Proceedings of the International Conference on Auditory Display*. July 2003, pp. 25–28.

[PB04] Guillaume Potard and Ian Burnett. “Decorrelation techniques for the rendering of apparent sound source width in 3D audio displays”. In: *Proc. of the 7th Int. Conference on Digital Audio Effects (DAFx'04)*. Naples, Italy, 2004, pp. 280–284.

[PB82] D. R. Perrott and T. N. Buell. “Judgments of sound volume: effects of signal duration, level, and interaural characteristics on the perceived extensity of broadband noise”. In: *J. Acoust. Soc. Am.* 72.5 (1982), pp. 1413–1417.

[PSP14] Tapani Pihlajamaki, Olli Santala, and Ville Pulkki. “Synthesis of Spatially Extended Virtual Source with Time-Frequency Decomposition of Mono Signals”. In: *J. Audio Eng. Soc* 62.7/8 (2014), pp. 467–484. URL: http://www.aes.org/e-lib/browse.cfm?elib=17339.

[Ray07] Lord Rayleigh. “XII. On our perception of sound direction”. In: *Philosophical Magazine Series 6* 13.74 (1907), pp. 214–232. DOI: 10.1080/14786440709463595. eprint: http://dx.doi.org/10.1080/14786440709463595. URL: http://dx.doi.org/10.1080/14786440709463595.

[RB68] Suzanne K. Roffler and Robert A. Butler. “Factors That Influence the Localization of Sound in the Vertical Plane”. In: *The Journal of the Acoustical Society of America* 43.6 (1968), pp. 1255–1259. DOI: http://dx.doi.org/10.1121/1.1910976. URL: http://scitation.aip.org/content/asa/journal/jasa/43/6/10.1121/1.1910976.

[Roa04] Curtis Roads. *Microsound*. The MIT Press, 2004. ISBN: 0262681544.

[Roc03] D. Rocchesso. *Introduction to Sound Processing*. Mondo estremo, 2003. ISBN: 9788890112614. URL: https://books.google.at/books?id=F76OxogOaC0C.

[Sch58] Manfred R. Schroeder. “An Artificial Stereophonic Effect Obtained from a Single Audio Signal”. In: *J. Audio Eng. Soc* 6.2 (1958), pp. 74–79. URL: http://www.aes.org/e-lib/browse.cfm?elib=238.

[SP09] Olli Santala and Ville Pulkki. “Resolution of Spatial Distribution Perception with Distributed Sound Source in Anechoic Conditions”. In: *Audio Engineering Society Convention 126*. 2009. URL: http://www.aes.org/e-lib/browse.cfm?elib=14975.

[SP11] Olli Santala and Ville Pulkki. “Directional perception of distributed sound sources”. In: *The Journal of the Acoustical Society of America* 129.3 (2011), pp. 1522–1530. DOI: http://dx.doi.org/10.1121/1.3533727. URL: http://scitation.aip.org/content/asa/journal/jasa/129/3/10.1121/1.3533727.

[TG11] Kristi Tsukida and Maya R. Gupta. *How to Analyze Paired Comparison Data*. Tech. rep. UWEETR-2011-0004. University of Washington, Department of Electrical Engineering, 2011.

[Thu27] L. L. Thurstone. “A law of comparative judgment”. In: *Psychological Review* 34 (1927), pp. 273–286.

[Tru98] Barry Truax. “Composition and diffusion: space in sound in space”. In: *Organised Sound* 3 (02 Aug. 1998), pp. 141–146. ISSN: 1469-8153. DOI: null. URL: http://journals.cambridge.org/article_S1355771898002076.

[VAOB12] Yolanda Vazquez-Alvarez, Ian Oakley, and Stephen A. Brewster. “Auditory display design for exploration in mobile audio-augmented reality”. In: *Personal and Ubiquitous Computing* 16.8 (2012), pp. 987–999.

[Ven+11] Peter Venus, Marian Weger, Cyrille Henry, and Winfried Ritsch. “Extended View Toolkit”. In: *Proceedings of the 4th Pure Data Convention*. 2011, pp. 161–167.

[WK97] F. Wightman and D. Kistler. “Factors affecting the relative salience of sound localization cues”. In: *Binaural and spatial hearing in real and virtual environments* 1 (1997), pp. 1–23.

[Zah02] P. Zahorik. “Direct-to-reverberant energy ratio sensitivity”. In: *J. Acoust. Soc. Am.* 112.5 Pt 1 (2002), pp. 2110–2117.

[Zha+07] Shengdong Zhao, Pierre Dragicevic, Mark Chignell, Ravin Balakrishnan, and Patrick Baudisch. “Earpod: eyes-free menu selection using touch input and reactive audio feedback”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. 2007, pp. 1395–1404.

[Zot+11] Franz Zotter, Matthias Frank, Georgios Marentakis, and Alois Sontacchi. “Phantom source widening with deterministic frequency dependent time delays”. In: *Proc. of the 14th International Conference on Digital Audio Effects (DAFx-11)*. Paris, France, 2011, pp. 307–312.

[Tym15] Tymphany HK LTD. *Peerless PLS-P830983 Transducer Specification Sheet*. `http://www.tymphany.com/transducers/transducer-search-results/transducer-detail/?id=825`. Accessed 2015-12-10. 2015.

Appendices

A. Consent Form

Zustimmungserklärung

Herzlichen Dank, dass Sie an unserer Untersuchung teilnehmen!

Ablauf:

Sie werden gleich an einem ca. 90-minütigen Versuch teilnehmen.
Die von Ihnen während der Bearbeitung der jeweiligen Aufgabenstellung produzierten Daten werden für eine spätere Auswertung aufgezeichnet.
Es erfolgt keine Bild- oder Tonaufnahme.

Datenschutz:

Die Testergebnisse, sowie Ihre persönlichen Daten werden anonymisiert und nicht an Dritte weitergegeben. Sollten Teile der ausgewerteten Daten in verschriftlichter Form veröffentlicht werden, so geschieht dies jedenfalls anonym. Es ist nicht möglich, aus der veröffentlichten Fassung auf Ihre Person rückzuschließen.

Optionale Angabe:

☐ Ich würde gerne über die Ergebnisse dieser Untersuchung informiert werden.

Ich habe obenstehende Informationen gelesen und verstanden und stimme den Bedingungen zu.

Name:

Datum: Unterschrift:

B. Experiment Instructions

Hörversuch zu ausgedehnten Schallquellen

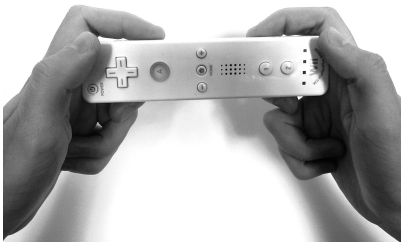
Allgemeine Hinweise

Sie be nden sich auf einem Stuhl, frontal gegenüber einer weißen Leinwand. Von dort werden Ihnen verschiedene Klänge in “Form” von horizontalen bzw. vertikalen Linien präsentiert, die Sie hinsichtlich der wahrgenommenen räumlichen Ausdehnung (Länge) beurteilen sollen.
Kopfbewegungen sind ausdrücklich erlaubt, der Oberkörper sollte jedoch aufrecht und gerade bleiben.

Experiment 1: Paarvergleich

Der Paarvergleich ist in 6 Teile (4 Teile zu je 72 Paaren und 2 Teile zu je 37 Paaren) gegliedert, die in zufälliger Reihenfolge präsentiert werden. Vor jedem Teil wird Ihnen die Orientierung der Linien angezeigt, d.h. ob die horizontale oder vertikale Ausdehnung zu beurteilen ist.

Sie hören in kurzem Abstand zwei unterschiedliche Stimuli.
Bitte drücken Sie auf der Wiimote die Nummer des Stimulus, den Sie als größer emp nden:



Sie emp nden Stimulus 1 größer als Stimulus 2 Taste ❶.
Sie emp nden Stimulus 2 größer als Stimulus 1 Taste ❷.

Die Stimuli werden in kurzen Abständen wiederholt gespielt. Sie können sich jedes Paar beliebig oft anhören. Verlassen Sie sich im Zweifelsfall auf den ersten Eindruck. Sobald Sie eine Antwort gegeben haben wird das nächste Paar präsentiert.

Experiment 2: Absolute Längenbestimmung

Dieses Experiment besteht aus 2 Teilen zu je 72 Stimuli. Vor jedem Teil wird Ihnen die Orientierung der Linien angezeigt, d.h. ob die horizontale oder vertikale Ausdehnung zu beurteilen ist.

Sie hören einen andauernden Klang. Bitte bestimmen Sie die wahrgenommene Ausdehnung, indem Sie Anfangs- und Endpunkt des Klangs mit dem “Wiivolver” auf die Leinwand zeichnen:

Sobald Sie auf die Leinwand zielen, erscheint ein Zielpunkt an der aufgezeigten Position. Durch einmaliges Drücken des Auslösers de nieren Sie den Anfangspunkt der Linie. Durch erneutes Drücken bestimmen Sie den Endpunkt. Die Position der Endpunkte kann im Anschluss jeweils durch Ziehen und Ablegen (drag and drop) korrigiert werden. Mit der Minus-Taste kann die Eingabe komplett gelöscht werden, um erneut mit dem Zeichnen zu beginnen.
Bitte zeichnen Sie die Linie möglichst genau an der Stelle, an der Sie den Klang wahrnehmen.

Zum Bestätigen und Fortfahren zum nächsten Stimulus drücken Sie die A-Taste.

Marian Weger
(Name in Blockbuchstaben)

0673093
(Matrikelnummer)

Erklärung

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

Graz, den ……..

……
Unterschrift der Verfasserin/des Verfassers