DIPLOMA THESIS

Perception and neural representation of suprathreshold signals in the presence of complex maskers

Wahrnehmung und neurale Repräsentation überschwelliger Signale in der Gegenwart von komplexen Maskierern

Katharina Egger

June 2012

Assessor : Prof. Robert Höldrich
Supervisors : Dr. Bastian Epp and Prof. Torsten Dau
Co-Supervisor : Dr. Alois Sontacchi

Institute of Electronic Music and Acoustics
University of Music and Performing Arts, Graz
Graz University of Technology

Centre for Applied Hearing Research
Technical University of Denmark

Technical University of Denmark

Abstract

An important aspect for understanding auditory signal processing is the relation between the physiological (i.e., the neural *internal*) representation of a sound and its corresponding percept. The present study aims to quantify the perceptual attributes of a signal presented in complex maskers which differ in their effectiveness to mask a signal. While the mechanisms and properties of masking in the auditory system have been studied extensively, the perception of only partially masked (i.e., *suprathreshold*) signals has not drawn much attention.

To obtain a metric that relates physical properties of the stimulus to its *internal representation* in the brain, the masked threshold of a signal (target) was modified here by introducing different *cues* that can lead to a release from masking (i.e., a reduced masking effect). Such cues were coherent masker intensity fluctuations across frequency, so-called comodulation, and interaural disparities in the signal. Suprathreshold perception was assessed by varying the signal level relative to the masked threshold. A perceptual measure was found that described the salience of the target signal (i.e., how well the signal can be segregated from its background). While this measure was primarily determined by the signal level above masked threshold, some effects of experimental condition and subject were found. Further, auditory evoked potentials (AEPs) were recorded to investigate the possibility of a electro-physiological correlate to the perceptual measure of the target's salience. Long-latency amplitudes were found to be sensitive to changes in signal level above masked threshold. However, more test subjects are required to further clarify the relation between the AEPs and the perceptual measure of salience.

Overall, the results suggest that the perceptual salience of a signal embedded in a complex masker is strongly dependent on the masking properties of the individual stimuli, such that salience corresponds roughly to the sensation level of the target and not necessarily to the sound pressure level. It was therefore concluded that cues leading to a release from masking contribute to the general perception of signals in suprathreshold hearing.

Kurzfassung

Ein wichtiger Aspekt zum Verständnis der auditorischen Signalverarbeitung ist der Zusammenhang zwischen der physiologischen (internen, neuronalen) Repräsentation eines auditorischen Reizes und dessen Wahrnehmung. Ziel dieser Arbeit ist die Quantifizierung der Wahrnehmung eines überschwelligen Signals in der Gegenwart von komplexen Maskierern, die das Signal in unterschiedlichem Ausmaß maskieren. Während die Funktionsweise und die Eigenschaften von Maskierung im auditorischen System bereits umfassend untersucht wurden, ist die Verarbeitung von überschwelligen Signalen und deren Wahrnehmung bisher nur wenig erforscht.

Um eine Metrik abzuleiten, welche die physikalischen Eigenschaften des Reizes mit der *internen Repräsentation* eines Stimulus in Beziehung setzt, wurden die Mithörschwellen von maskierten Signalen durch systematische Variation der Signaleigenschaften (*Cues*) gezielt verändert. Dazu wurden sowohl kohärente Amplitudenfluktuationen über der Frequenz im Maskierer (Komodulation) als auch interaurale Disparitäten im Signal verwendet. Die Wahrnehmung überschwelliger Signale wurde durch Variieren des Signalpegels relativ zur Mithörschwelle bestimmt. Ein perzeptives Maß für die Salienz des maskierten Signals wurde eingeführt. Es beschreibt wie gut das Signal vom Hintergrund getrennt werden kann. Dieses Maß war in erster Linie vom Signalpegel relativ zur jeweiligen Mithörschwelle (*Sensation Level*) abhängig, jedoch konnten zum Teil auch Einflüsse von Signaleigenschaften und Testpersonen nachgewiesen werden. Darüber hinaus wurden akustisch evozierte Potentiale (AEPe) gemessen, um den Zusammenhang zwischen elektrophysiologischen Daten und dem perzeptiven Maß der Salienz zu untersuchen. Die Amplituden der späten AEPe erwiesen sich als empfindlich gegenüber Änderungen des Signalpegels relativ zur jeweiligen Mithörschwelle. Weitere Testpersonen sind jedoch erforderlich, um den Zusammenhang zwischen der AEPe und der Salienz des maskierten Signals genauer zu bestimmen.

Die Ergebnisse zeigten, dass die Salienz eines maskierten Signals stark von den Eigenschaften der Maskierung der individuellen Stimuli abhängt, was wiederum andeutet, dass die Salienz eher vom Sensation Level als vom physikalischen Pegel bestimmt wird. Daraus lässt sich schließen, dass Cues, die zur Verbesserung der

Mithörschwelle führen, auch bei der Wahrnehmung von überschwelligen Signalen eine wichtige Rolle spielen.

Statutory declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

....

(date) (signature)

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

....

(Datum) (Unterschrift)

Acknowledgements

I would like to thank ...

Bastian Epp & Torsten Dau

for the great supervision, positive motivation and support, interesting discussions, investing time and effort in my project, and giving me the opportunity to be part of CAHR,

Ewen MacDonald

for feedback, advice, and reviewing my thesis in the very busy end of my project,

Kasper Eskelund

for the help getting the EEG setup running,

My mother, Lisi & Harti

for trust, patience, and support,

Sebastian

for being so close to me all along despite the long distance,

Barbara, Federica, Christer & Kristian

for the nice working environment in our office,

Alex, Edd, Nicolas & all the acoustics friends

for tasty dinners, good music, and an enjoyable time in København,

Magdi, the Mongocity crew & my friends in Graz

for keeping in touch and being friends I can count on.

Contents

Nomenclature

Abbreviations

ABR	auditory brainstem response
AEP	auditory evoked potential
ANOVA	analysis of variance
BMLD	binaural masking level difference
CB	critical band
CM	comodulated condition
CMR	comodulation masking release
EC	equalization-cancellation
EEG	electroencephalography
ERP	event-related potential
FFT	fast Fourier transform
fMRI	functional magnetic resonance imaging
IPD	interaural phase difference
IQR	interquartile range
LAEP	late auditory evoked potential
MEG	magnetoencephalography
MLR	middle-latency response
NaN	not a number
REM	rapid-eye-movements
RF	reference condition
SCB	signal centred band
SL	sensation level
SPL	sound pressure level
UN	uncorrelated condition

Chapter 1

Introduction

"How do we recognize what one person is saying when others are speaking at the same time?"

Colin Cherry, 1953.

The *cocktail party problem*, first introduced by Cherry (1953), describes the ability to focus one's listening attention on a single talker while further conversations, music, or background noise are present. One of the main tasks of the auditory system is to distinguish between different sound sources and segregate a sound source (target signal) from its background. Over the last decades, the cocktail party effect has been investigated in various studies (e.g., Bronkhorst, 2000; Arons, 1992; Haykin and Chen, 2005, for a review with different focuses of research). However, the processes in the auditory system that enable listeners to segregate the different sound sources in a natural acoustic environment are still poorly understood.

One way to investigate auditory signal processing is a model-based approach using bottom-up processing, from perceiving sounds up to detecting and identifying a particular sound source. Within this approach, it is assumed that the signal as well as the masker can be represented by their *internal representation* (i.e., the auditory representation of sounds after being processed in order to facilitate the identification of the sounds). The strength of the internal representation is assumed to change, depending on whether the sound is clearly perceived or not. Consider the case of people communicating in noisy environment, talkers will automatically raise their voices in order be better heard. Here, the level of the speech signal within the background noise increases such that the internal representation of the speech signal becomes clearer and the signal is perceived more clearly (i.e., the signal becomes more audible). This can also be visualised by overlapping two waveforms, a sample speech signal and white noise, as shown

in figure 1.1. The waveforms are associated with the internal representations of the speech signal (red) and the noise (blue). The higher the level of the speech, the better its audibility and consequently, the clearer its internal representation, illustrated by the waveform of the speech becoming more and more visible (from left to right).

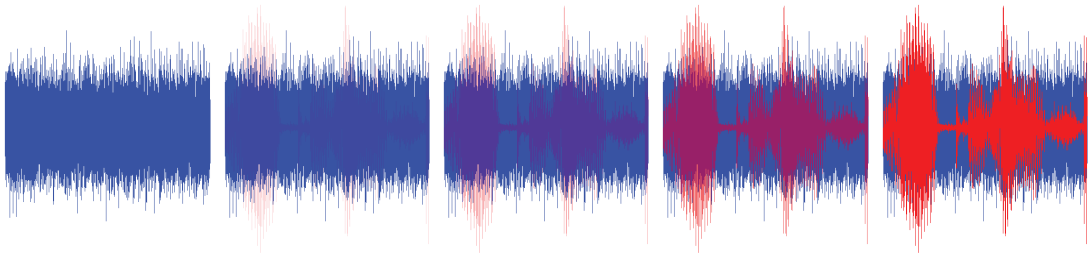

Figure 1.1: Schematic illustration of the idea of an internal representation of sounds with sample waveforms of a speech signal (red) and white noise (blue). The five panels illustrate the waveform of the speech becoming more and more visible (from left to right), corresponding to an improvement in detectability of the speech signal embedded in noise.

The auditory system analyses different aspects of sounds in order to segregate one sound source from others. Among these are monaural cues such as across-frequency information as well as binaural cues such as interaural disparities. The strategies to process this information aim at facilitating signal detection and increasing the audibility of a target signal masked by interfering sound sources. The ability of the auditory system to process such cues is commonly assessed by the measurement of masked thresholds. A characteristic of many natural sound sources is that the signal energy in different frequency bands is modulated coherently in time (Nelken *et al.*, 1999). Such synchronous envelope fluctuations across frequency can facilitate signal detection in noise. It has been shown that masked thresholds of a sinusoidal signal decrease due to the coherent envelope fluctuations of the masker. This phenomenon of enhancement in signal detectability is associated with the term *comodulation masking release* (CMR, first labelled by Hall *et al.*, 1984; Verhey *et al.*, 2003, for a review). Apart from across-frequency information, the auditory system benefits from interaural disparities between the sound waves arriving at both ears due to sounds coming from different directions. It has been demonstrated that differences in interaural disparities between the masker and the signal lower thresholds of masked signals (Hirsh, 1948; Licklider, 1948) and therefore, improve signal detection. The effect of enhancement of signal detectability is referred to as *binaural masking level difference* (BMLD) (Jeffress *et al.*, 1956; van de Par and Kohlrausch, 1999). The improvement of masked thresholds due to the ability of the auditory system to process cues like across-frequency and binaural information available in the perceived sounds is generally

referred to as a *release from masking*. In the presence of such cues, where masked threshold is lower, less signal energy is needed to detect the signal. Considering the idea of the internal representation, the threshold of a masked signal refers to the point where the internal representation of the target signal is strong enough so that the signal can be segregated from the masker. Therefore, it is assumed that a certain internal representation of the target signal is needed to be detected. It is hypothesised that in the presence of the cues (where masked threshold is lower), less energy is required for the same, clear internal representation (i.e., the internal representation of the target is improved compared to conditions where the cues are absent and thus, masked threshold is higher).

While the ability of the auditory system to process such cues is well characterised at masked signal threshold, it remains unclear if such cues contribute to the perception of the signal in suprathreshold hearing (i.e., at levels above masked threshold). The aim of the project is to find a perceptual measure, which quantifies the strength of the internal representation and hence, the audibility of a masked sinusoidal signal in the presence of cues resulting in a masking release.

1.1 Background

Many experiments on masking and signal detection can be accounted for by the power spectrum model of the auditory system proposed by Fletcher (1940). This model assumes that the auditory system and thus the frequency selectivity in the cochlea can be represented as a filter bank of overlapping bandpass filters, known as the *auditory filters*. In a so-called *band-widening experiment*, Fletcher (1940) measured masked thresholds of a sinusoidal signal that is spectrally centred in bandpass noise as a function of bandwidth of the noise. The spectral power density of the noise band was kept constant. Results showed increased thresholds as the noise bandwidth increased up to a certain bandwidth, referred to as critical bandwidth, after which thresholds remained constant. Figure 1.2 schematically illustrates the spectrum of the used stimulus and the behaviour of the obtained results.

Fletcher (1940) concluded that the detectability of the sinusoidal signal depends mainly upon the signal-to-noise ratio within the auditory filter centred at the signal frequency (referred to as *critical band*, CB). As soon as the masking noise band becomes wider than the CB, the masker energy within the CB remains constant and no further threshold changes are observed for larger noise bandwidths. Consequently, the resulting power spectrum model assumes that primarily the CB is taken into account for signal detection. Masked thresholds are determined by the ratio of signal energy to masker energy within this CB

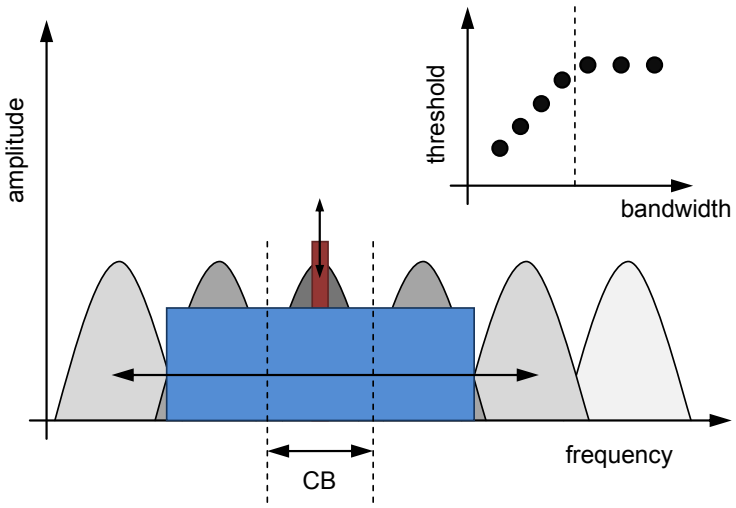

Figure 1.2: Schematic spectrum of the signal and the masker in the band-widening experiment of Fletcher (1940). Masked thresholds of a sinusoidal signal (red) spectrally centred in bandpass noise (blue) as a function of bandwidth of the noise were measured. In the back plane of the spectrum, auditory filters are indicated in grey. The amount of energy of the masking noise within the respective filter is indicated by the different shades of grey, where dark grey corresponds to maximum energy. The upper right panel outlines the behaviour of the detected thresholds in Fletcher (1940) as a function of masker bandwidth. Thresholds increased with increasing noise bandwidth up to a certain, so called critical bandwidth (vertical, dashed line), from which on thresholds remained constant. The critical bandwidth refers to the critical band (CB) in the spectrum (dashed lines).

and only the frequency components of the masking noise lying within the CB have influence on the amount of masking. The model considers the long-term power spectra of the stimuli. Thus, neither the relative phase of the stimuli nor temporal fluctuations of the noise have influence on the model output and hence, masked threshold. Although these assumptions lead to successful predictions in many situations, the model fails in others.

In a band-widening experiment analogues to Fletcher (1940), Hall *et al.* (1984) demonstrated that masked thresholds of a sinusoidal signal with a frequency of 1000 Hz presented in band-limited masking noise centred at the signal frequency decrease for noise bandwidths wider than the critical bandwidth, provided that the temporal envelope fluctuations of the noise across frequencies are coherent. Masked thresholds were measured for random (i.e., uncorrelated) noise and for so-called multiplied noise which was generated by multiplying a wideband of noise by a lowpass noise. This multiplied noise exhibits coherent envelope fluctuations (i.e., comodulation) corresponding to the characteristic of the lowpass noise. Figure 1.3 shows the obtained masked thresholds of Hall *et al.* (1984). For the random noise, thresholds were obtained as expected from the traditional power spectrum model, whereas for the multiplied noise, masked thresholds showed an enhancement of signal detectability of up to about 10 dB relative to the random noise conditions,

for increasing noise bandwidth. This detection advantage for the comodulated noise has been defined as CMR.

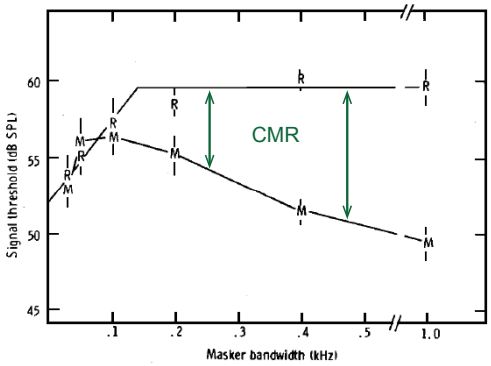

Figure 1.3: Masked thresholds obtained in Hall *et al.* (1984)'s band-widening experiment for random (denoted by R) and multiplied (denoted by M) noise (figure 2 of Hall *et al.*, 1984). Detection thresholds of a sinusoidal signal with a frequency of 1000 Hz, spectrally centred in bandpass noise, were measured as a function of bandwidth of the noise. Thresholds were significantly lower in multiplied (M) noise than in random (R) noise for bandwidths greater than 0.1 kHz. This threshold improvement refers to the CMR. Dashes above and below the data symbols indicate standard deviations.

Apart from band-widening experiments, a second method to demonstrate CMR can be found in the literature (e.g., Hall *et al.*, 1984, 1990; Schooneveldt and Moore, 1987). The *flanking-band paradigm* is a common way to show an effect of CMR by using comodulated narrow bands of noise. In this class of CMR experiments, one noise band is centred at the signal frequency (*signal centred band*, SCB) and then one or more narrow bands (i.e., referred to as flanking bands) are spectrally separated from the signal frequency. The magnitude of CMR can be calculated in two different ways. One way refers to the difference in thresholds between the condition with the SCB only (*reference* condition, RF) and the condition where comodulated flanking bands are present (*comodulated* condition, CM), denoted here as CMR (RF-CM). Alternatively, the benefit of comodulated flanking bands can be considered by comparison to the condition with flanking bands present that have uncorrelated envelope fluctuations (*uncorrelated* condition, UN), denoted as CMR (UN-CM). For example, Hall *et al.* (1990) investigated the magnitude of CMR for a 700 Hz sinusoidal signal as a function of the number and spectral positions of comodulated flanking bands of 20 Hz bandwidth. Figure 1.4 shows the measured thresholds as a function of number of comodulated flanking bands present, when the bands were placed symmetrically (on a linear frequency scale). The results indicated a larger CMR with increasing number of flanking bands. That means that additional masker energy leads to a reduction of masked thresholds. The power spectrum model cannot account for this benefit in signal detectability.

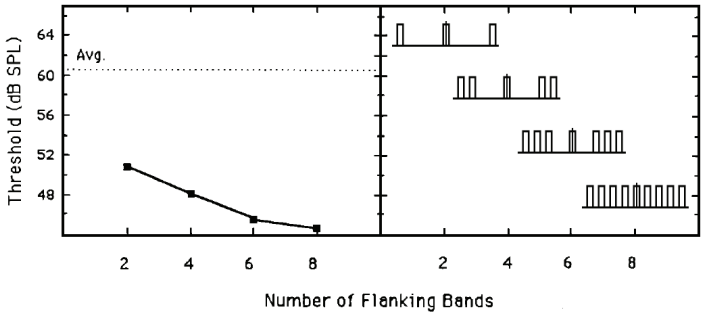

Figure 1.4: Average results over subjects of figure 3 of Hall *et al.* (1990). The left panel indicates signal thresholds as a function of number of comodulated flanking bands present, which were added symmetrically to the SCB. The average threshold for the SCB alone (reference condition, RF) is indicated by the dotted line. For each number of flanking bands present the CMR can be calculated as the difference between the corresponding threshold and the threshold for the SCB alone. The right panel shows the schematic spectra associated with the four corresponding conditions.

Further experiments of Hall *et al.* (1990) showed that noise bands close to the SCB resulted in higher magnitudes of CMR. In literature it has been argued that some amount of CMR can be accounted for by other effects rather than true across-channel processing. It has been suggested that within-channel cues, resulting from beating of components of spectrally close masker bands within one peripheral channel, can at least partly contribute to CMR (McFadden, 1986; Schooneveldt and Moore, 1987). This also accounts for the class of band-widening experiments. Verhey *et al.* (1999) presented a single-channel model in which only the envelope of the peripheral channel tuned to the signal frequency was analysed. The model predicted several aspects of CMR without assuming any across-channel cues. It was suggested that a large amount of CMR in band-widening experiments results from within-channel cues. In contrast, it was assumed that CMR in flanking-band experiments results from *true* across-channel processes, at least when the SCB and the flanking bands are sufficiently separated in frequency.

Two hypotheses have been suggested to account for the effect of CMR. (1) A *dip-listening* model was proposed that assumed the auditory system to be able to use periods of low masker energy (dips) for enhanced signal detection (Buus, 1985). (2) A traditional *equalization-cancellation* (EC) model (Durlach, 1963) was also thought to account for CMR (e.g., Buus, 1985; Richards, 1987). Durlach's EC model was originally developed to interpret data on BMLDs by equalizing and subsequently subtracting the masking components in order to detect the signal. According to this mechanism, the auditory system may use the correlating masker envelope in one or more frequency channels to effectively remove its information from the CB tuned to the signal frequency, achieving a maximal signal-to-noise ratio and reflecting the residual tonal signal. This was supported by a model

proposed by Piechowiak *et al.* (2007) which used an EC mechanism to predict CMR in flanking-band experiments.

Apart from across-frequency information in monaural processing, the auditory system uses binaural cues to segregate sound sources. Differences in interaural disparities between the masker and the signal lead to an enhancement in signal detectability which is commonly referred to as BMLD. Effects of BMLD are typically investigated in threshold detection experiments for sinusoidal signals masked by noise, where an *interaural phase difference* (IPD) has been introduced in either the signal or the masker. Hence, either the signal or the masker is presented dichotically (e.g., interaurally out of phase). The magnitude of the BMLD is then calculated as the difference in thresholds of the diotic condition, where signal and masker are interaurally in phase (the same sound waves reach both ears), and the dichotic condition.

For example, van de Par and Kohlrausch (1999) measured thresholds for sinusoidal signals masked by noise for three binaural configurations: the diotic condition N_0S_0 (both signal and masker interaurally in phase), N_0S_π (masker interaurally in phase and antiphasic signal), and $N_\pi S_0$ (antiphasic masker and signal interaurally in phase). Their results, partly shown in figure 1.5, indicated that BMLD easily reached values up to 20-25 dB but varied substantially depending on the properties of the signal and the masker. Among other parameters, BMLD depended on the signal frequency and the masker bandwidth. Magnitudes of BMLD tended to decrease with increasing masker bandwidth and also with increasing signal frequency.

The combined effect of comodulation (i.e., CMR) and binaural cues leading to BMLD has been investigated in several studies (Hall *et al.*, 1988, 2006, 2011; Cohen and Schubert, 1991; Epp and Verhey, 2009a,b). While Hall *et al.* (1988) and (2006) found that the addition of comodulated masker bands had little to no effect on further enhancement of binaural detection, Epp and Verhey (2009a,b) showed an additive behaviour of the masking releases. Results indicated that, in conditions with both cues present at a time, the total masking release is equal to the sum (in decibels) of CMR and BMLD, provided that the effect of CMR is calculated as the difference between the CM and the UN condition (CMR (UN-CM)). Epp and Verhey (2009b) used a flanking-band paradigm to detect masked thresholds for a sinusoidal signal at three different signal frequencies. While the masker was always presented diotically, interaural disparities in terms of an IPD were gradually introduced to the signal. Thresholds were obtained using the SCB alone (RF condition) or adding either additional uncorrelated (UN condition) or comodulated (CM condition) flanking bands. Results for a signal frequency of 700 Hz and multiplied noise as masker are shown in figure 1.6.

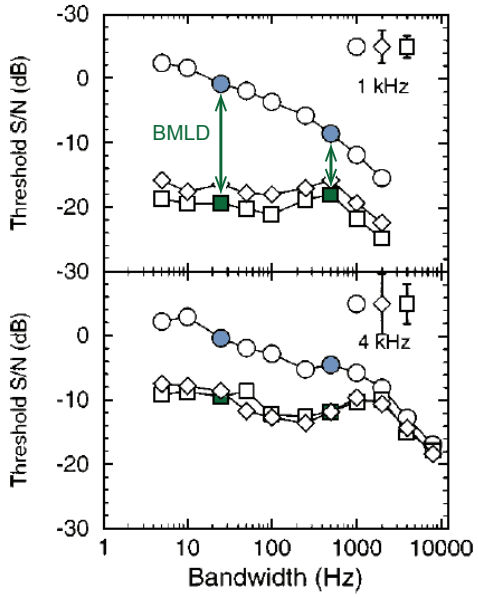

Figure 1.5: Thresholds as a function of masker bandwidth, expressed in signal-to-overall-noise power ratio are shown for the diotic condition N_0S_0 (circles) and the two dichotic conditions N_0S_π (squares) and $N_\pi S_0$ (diamonds), partly taken from figure 1 of van de Par and Kohlrausch (1999). The BMLD can be calculated for instance as the difference in thresholds between N_0S_0 and N_0S_π. Standard deviations across subjects, averaged across masker bandwidth are indicated with the isolated symbols.

Thresholds decreased monotonically as the IPD increased. Adding comodulated flanking bands (CM condition) further improved thresholds, resulting in an almost constant CMR (UN-CM) and a decreasing CMR (RF-CM) with increasing IPD.

Epp and Verhey (2009a) presented a model which supported their findings and reproduced the data by assuming different, serially aligned processing stages that account for the effects of CMR and BMLD. In contrast, the most recent study by Hall *et al.* (2011) indicated a combined masking release larger than the magnitudes of CMR and BMLD obtained in isolation, but *"the effects were less than additive"*. Hall *et al.* (2011) further investigated the effect of presenting the masking noise either continuously or gated (i.e., only during listening intervals within a trial), attempting to explain the different binaural CMR results of Hall *et al.* (2006) and Epp and Verhey (2009a,b). While Epp and Verhey (2009a,b) used gated maskers, Hall *et al.* (2006) presented the masking noise continuously. Hall *et al.* (2011) replicated the finding of Hall *et al.* (2006) that the CMR diminished for N_0S_π when the masking noise was presented continuously. Also for gated noise, CMR was found to be smaller for N_0S_π than N_0S_0, but the decrease was not as strong as for the continuously presented masking noise.

Most studies on masking release have characterised the effects of CMR and BMLD in terms of masked signal thresholds. However, most sound sources in an everyday situation are well above threshold. The question whether masking

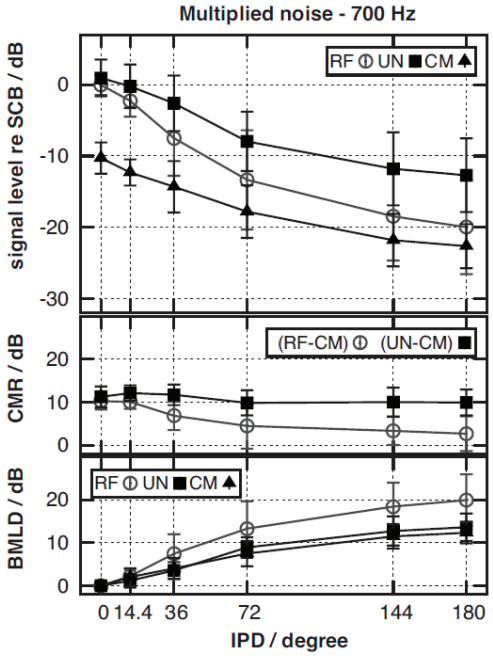

Figure 1.6: Figure 2 of Epp and Verhey (2009b). The upper panel shows the mean detection thresholds over all subjects as a function of IPD, using multiplied noise as masker. Thresholds are plotted relative to the level of the SCB for the RF (circles), UN (squares), and CM (triangles) conditions. The middle panel indicates the average CMR (RF-CM, circles and UN-CM, squares). The average BMLD is plotted in the lower panel for the RF (circles), UN (squares), and CM (triangles) conditions. Error bars indicate ± 1 standard deviation.

releases play a crucial role in suprathreshold hearing, where stimuli are highly detectable, has been investigated only very little. A few studies can be found dealing with the extent to which the presence of either across-frequency information or binaural cues contributes to the processing and perception in suprathreshold situations (e.g., Townsend and Goldstein, 1972; Verhey and Heise, 2012; Hall and Grose, 1995). Studying the effect of BMLD, highest intelligibility of masked speech was found when either the speech signal or the masking noise, but not both, was interaurally out of phase (Licklider, 1948). Rather than speech intelligibility, loudness as the measure of perceived intensity has been used more frequently to assess perception at suprathreshold levels. Several studies indicated that the binaural benefit of BMLD diminished for increasing suprathreshold levels. Investigations on the role of interaural phase at suprathreshold levels showed that an IPD had an effect on loudness of masked, sinusoidal signals for signal levels near threshold. This effect disappeared at higher signal-to-masker ratios (Hirsh and Pollack, 1948). Townsend and Goldstein (1972) studied the binaural unmasking behaviour at suprathreshold levels for masked signals with a frequency of 250 or 500 Hz by a loudness balance procedure. They determined the level difference at equal loudness for a masked diotic signal (masker and signal both interaurally in phase) compared to a masked signal presented dichotically with

an IPD of π. The lines of equal loudness for a masked 500 Hz sinusoidal signal are shown in figure 1.7. They reported a decreasing binaural advantage of the masked antiphasic signal with increasing level above masked threshold, also referred to as *sensation level* (SL). Their findings implied different loudness growth functions for diotic and dichotic signals, i.e., the loudness of the masked signal interaurally out of phase grew slower than the one interaurally in phase.

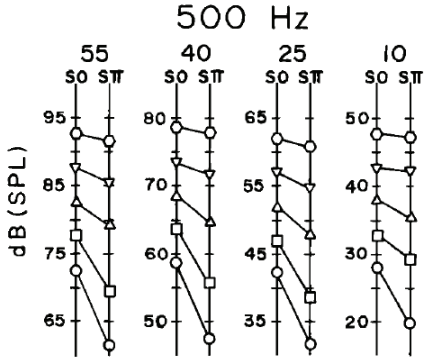

Figure 1.7: Lines of equal loudness between a masked sinusoidal signal with a frequency of 500 Hz presented either diotically (S_0) or dichotically (S_π), right panel of figure 3 of Townsend and Goldstein (1972). The sound pressure level (SPL) necessary for equal loudness for each condition is shown by the ordinates for different levels above masked threshold (i.e., sensation levels, SLs) used (0 dB SL - circles, 5 dB SL - squares, 10 dB SL - triangles, 15 dB SL - inverted triangles, 20 dB SL - hexagons). Loudness matching was performed with four different spectrum levels of the masking noise (55, 40, 24 and 10 dB), indicated by the four columns of equal loudness lines in the plot.

Similar results were found by Soderquist and Shilling (1990), who examined the behaviour of loudness in conditions of BMLD in a loudness matching paradigm. They showed that dichotic signals needed lower signal levels to be perceived as equally loud as the corresponding diotic signals at low sensation levels. However, for higher sensation levels (around 20 dB SL) the binaural benefit vanished and the equal loudness levels for the diotic and the dichotic signals were the same. Furthermore, Fastl and Zwicker (2007) determined levels of a masked, monaurally presented 250 Hz tone, necessary to sound equally loud as binaurally presented 250 Hz tones, where the tone was either presented interaurally in or out of phase. The masking noise was always presented in phase at both ears. Their results indicated that the tone binaurally presented out of phase produced greater loudness than the one presented in phase. This was shown at low signal-to-noise ratios, but also up to 40 dB SL where a residual effect still could be found. The most recent study from Verhey and Heise (2012) can be seen as an expansion of the aforementioned studies considering loudness to quantify suprathreshold perception. In order to assess perception of a tonal component embedded in noise, Verhey and Heise (2012) investigated the magnitude of tonal content (in German

'Tonhaltigkeit') as well as partial loudness of the tone in noise and the change of perception in conditions with masking release. They measured tone levels at equal magnitude of tonal content or at equal loudness of the tone between a diotic tone embedded in unmodulated noise (baseline condition) and a masked tone in conditions of either CMR or BMLD. For both measures, lower signal-to-noise ratios were observed for tones in the masking release conditions than for the baseline condition. The results were in good agreement with the above-mentioned studies showing the most pronounced effect at low levels which diminished towards higher levels. Data obtained for the magnitude of tonal content and partial loudness were in high correlation. In summary, it has been shown that the loudness of a masked sinusoidal signal differed for a signal that was either presented diotically or dichotically. However, the difference in loudness perception diminished for increasing signal level relative to masked threshold.

Considering the effect of CMR in suprathreshold hearing, it has been speculated that the phenomenon of CMR may also be relevant in everyday life where we listen to speech in noisy environments. Communication sounds have highly correlated amplitude envelope fluctuations across frequency (e.g., Nelken *et al.*, 1999). Therefore, it has been suggested that, among other mechanisms, CMR may at least partly contribute to a release from speech masking (e.g., Festen and Plomp, 1990; Festen, 1993). Grose and Hall (1992) investigated the role of CMR for speech stimuli in modulated noise. Their findings demonstrated an enhancement in speech detection tasks attributable to CMR, but no evidence of improved speech recognition was found which could be related to a suprathreshold CMR benefit. This supports subsequent studies which found only little to no benefit of a CMR effect in suprathreshold tasks (e.g., Festen, 1993 for speech recognition; Hall and Grose, 1995; Buss and Hall, 2009 for amplitude discrimination; Hall *et al.*, 1997 for pitch ranking). However, the possibility of a suprathreshold effect of CMR should not be fully excluded and as Kwon (2002) stated, is worth to be reconsidered. Among other things, Kwon (2002) pointed out that in the aforementioned studies on speech recognition (Grose and Hall, 1992; Festen, 1993) it was not able to assess the magnitude of CMR in a traditional way, since the contrast between an uncorrelated and a comodulated condition was not as clear and direct as in conventional flanking-band experiments. This might be a reason why an effect of suprathreshold CMR could not be demonstrated. In contrast, Kwon (2002) found a small but consistent effect of CMR in a consonant recognition task. Furthermore, the study endorsed the dip-listening strategy accounting for CMR at suprathreshold levels rather than the envelope correlation theory. They suggested that the latter might provide cues facilitating signal detection but not necessarily give additional signal information.

The studies mentioned above all considered each cue leading to masking release in isolation, whereas suprathreshold perception for the combined effect of CMR and BMLD has not been investigated so far. It is still unclear to which extent cues leading to a reduction in masked threshold contribute to the processing and perception in suprathreshold situations.

An important aspect to better understand human auditory signal processing is to find physiological correlates to phenomena observed in psychoacoustical experiments. Many physiological studies have focused on the neural mechanisms underlying the processing of cues leading to a release from masking and have attempted to identify which stages of the auditory pathway are involved (e.g., Nelken *et al.*, 1999; Pressnitzer *et al.*, 2001; Neuert *et al.*, 2004 for across-frequency cues; Reale and Brugge, 1990; Thompson *et al.*, 2006; von Kriegstein *et al.*, 2008 for binaural cues). The effect of CMR is not only restricted to be found in human listeners but also in other vertebrates (e.g., in dolphins: Branstetter and Finneran, 2008; in mice: Klink *et al.*, 2010). Thus, animal based studies enable to infer similar processing in humans. For instance, Nelken *et al.* (1999) showed that auditory-cortex neurons in cats were sensitive to CMR and proposed a neuronal correlate to CMR in the auditory cortex. More recent studies reported that also in human listeners a physiological correlate of CMR can be verified at cortical levels (Androulidakis and Jones, 2006 using electroencephalography, EEG; Rupp *et al.*, 2007 using magnetoencephalography, MEG; Ernst *et al.*, 2008, 2010 using functional magnetic resonance imaging, fMRI). Androulidakis and Jones (2006) recorded *auditory evoked potentials* (AEPs) in a CMR paradigm to study the perceptual enhancement of a masked sinusoidal signal at a neurophysiological level. The signal was fixed in level and was either presented in silence or in band-limited masking noise, which was either uncorrelated or amplitude-modulated. The audibility of the signal changed depending on the masker condition (inaudible in uncorrelated but audible in modulated noise). Androulidakis and Jones (2006) recorded N1 and P2 potentials and observed a large N1/P2 response for the sinusoidal signal presented in silence. No N1/P2 response was evoked by the signal in the presence of the uncorrelated masker, but the potentials were elicited again for the signal in modulated noise, even though with a lower amplitude and a longer latency than observed in silence. They proposed the N1/P2 complex as a possible neurophysiological correlate of the perceptual effect of CMR. This is in good agreement with fMRI results from Ernst *et al.* (2010), who indicated *"a spatial dissociation of changes of overall level and signal-to-noise ratio"* in distinct areas of the auditory cortex. Those regions, mainly sensitive to the audibility of a masked target signal, showed stronger activation for modulated than for uncorrelated masking noise. The study of Epp *et al.* (2012) made AEP recordings

of masked sinusoidal signals in conditions of masking release. The audibility of the signal was altered by the introduction of across-frequency and/or binaural cues, yielding a combined effect of CMR and BMLD. The stimuli were presented at different signal-to-masker ratios. EEG measurements were performed in order to determine whether the effect of masking release was reflected in AEPs. It was found that the N1/P2 complex was sensitive to the audibility of the masked signal. More precisely, the N1 component was sensitive to a change in audibility when an interaural disparity was introduced, whereas the potential of P2 represented the increase in audibility when modulated noise was used instead of uncorrelated noise as well as when both cues were present.

1.2 Aim of the project

The aim of this study is to investigate the salience of perceptual cues of signals presented in different types of maskers. While the ability of the auditory system to process such stimuli at masked thresholds is well established, the cues influencing the perception and audibility of a masked signal at suprathreshold levels are not well understood.

In this project, a perceptual measure of the strength of the internal representation is developed, which is primarily determined by its level above masked threshold. It is hypothesised that this measure, representing the audibility of the signal, has equal magnitudes for signal presentations at equal levels above masked threshold, independent of the masking condition or its absolute threshold. In order to obtain a metric that relates the perceptual measure to the audibility of the signal, the masked threshold of the signal to be detected will be modified by introducing a systematic variation of both across-frequency and binaural cues which result in a release from masking (CMR, BMLD). In addition to previous studies, suprathreshold perception will be also investigated for the combined effect of CMR and BMLD. It is hypothesised that in the presence of these cues (where masked threshold is lower) the internal representation of the target is improved compared to conditions where these cues are absent (where masked threshold is higher).

Inspired by the study of Epp *et al.* (2012) it is assumed that the strength of the internal representation can also be assessed by objective measures (e.g., EEG). Here, AEPs are investigated to determine if the effect of masking release can be observed.

Part I

Psychoacoustics

Chapter 2

Experiment I: Signal detection in various conditions of masking release

Masked thresholds of sinusoidal signals in complex maskers were measured in the first experiment. The alternation of masked thresholds was achieved by a variation and combination of the following two cues: adding across-frequency information to the masker (comodulation) and introducing interaural disparities to the signal. Considering the idea of the internal representation of a signal, as introduced earlier in chapter 1, this experiment was carried out in order to determine the signal level that elicits an internal representation which is just strong enough so that the signal can be segregated from the masker. The individually masked thresholds served as a basis for the suprathreshold measure considered in experiment II.

2.1 Method

2.1.1 Stimuli and apparatus

Stimuli and their characteristics within the tested conditions were designed following experiment 1 of Epp and Verhey (2009b) in order to allow a comparison of the results. Masked thresholds were obtained for a sinusoidal signal with a frequency of 700 Hz. The sinusoidal signal had a duration of 250 ms and was temporally centred in the 500 ms-long masking noise, both with 50 ms raised-cosine on- and offset ramps. The signal was presented either diotically (0° IPD) or dichotically (180° IPD), whereas the masker was always presented diotically. The masker consisted of either one (RF condition) or five narrowband noise bands (UN and CM conditions). Each band had a bandwidth of 24 Hz and a level of

60 dB SPL. One noise band was centred at the signal frequency (SCB) and the other four flanking noise bands were arranged symmetrically (on a linear frequency scale) with respect to the SCB at 300, 400, 1000, and 1100 Hz. The remote spectral distance between the SCB and the low and high frequency bands was chosen in order to reduce within-channel cues contributing to CMR (e.g., Cohen, 1991). The conditions with all flanking bands present (also referred to as *multiband* conditions) can be divided into uncorrelated and comodulated noise: The flanking bands had either uncorrelated intensity fluctuations (UN condition) or the same intensity fluctuations (CM condition). Figure 2.1 illustrates examples for the different masker conditions, in the style of figure 1 of Epp and Verhey (2009a).

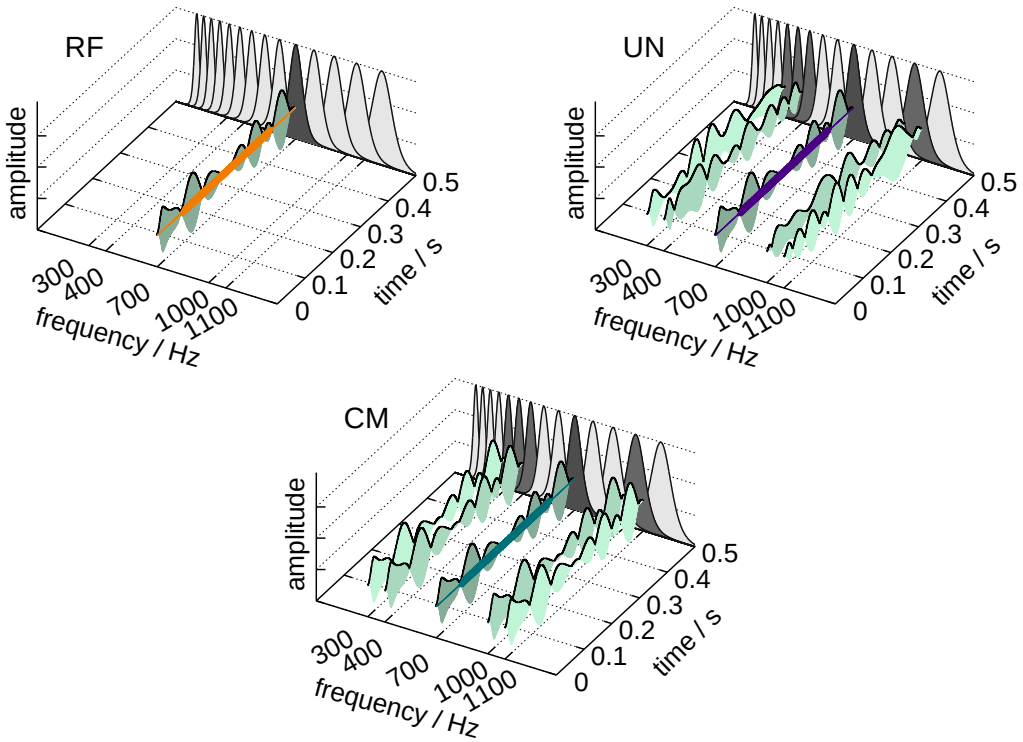

Figure 2.1: Schematic spectograms of the three used masker conditions and the sinusoidal signal with a frequency of 700 Hz. The masker consists of either only the signal centred band (RF condition, upper left panel) or SCB plus the four flanking bands. The latter can be distinguished in uncorrelated (UN condition, upper right panel) and comodulated masker envelopes (CM condition, lower panel). The back plane in each panel represents a gammatone filter bank. The amount of masker energy within the respective filter refers to the different shades of grey, where dark grey corresponds to a high excitation. This should illustrate that practically no masker energy from the flanking bands lies within the SCB. Therefore, the flanking bands cannot contribute to CMR based on within-channel effects.

Multiplied noise and Gaussian noise were used as maskers. These two types of masking noise were chosen because they had been used in previous, related studies (e.g., Hall *et al.*, 1984, 1990, 2011; Epp and Verhey, 2009a,b). Multiplied noise bands were generated by multiplying a random phase sinusoidal carrier at

the desired frequency with a lowpass noise without a DC component. The lowpass noise was generated in the frequency domain by assigning numbers between ± 0.5 out of an uniformly distributed random process to the real and imaginary parts of the frequency components up to 12 Hz. For the RF and UN conditions, independent realisations of the lowpass noise were used for each masker band, while for the CM condition, the same lowpass noise (generated from the same draw of numbers) was used for all five bands. Gaussian noise bands were generated in the frequency domain by assigning numbers drawn from a normally distributed process to the real and imaginary parts of the desired frequency components in each noise band. Different sets of numbers were drawn for each noise band for the RF and UN condition, while for the CM conditions the same set of numbers was assigned to all five masker components. The desired waveform resulted from taking the real part of the subsequent inverse fast Fourier transform (FFT).

Figure 2.2 shows a realisation of time interval for both used masker types. The corresponding envelope amplitude distributions are plotted alongside, since previous studies indicated a dependence of the effect of CMR on the envelope amplitude distribution of the masking noise (e.g Moore *et al.*, 1990; Verhey *et al.*, 2007). The envelope amplitude distribution for the multiplied noise masker corresponds to the positive half of a Gaussian distribution, while the Gaussian envelope amplitude distribution equals a Rayleigh distribution (figure 2 in Epp and Verhey, 2009a).

Maskers for both noise types were newly generated for each interval and each trial. All stimuli were generated digitally in MATLAB with a sampling rate of 44100 Hz and 16 bit resolution, converted from digital to analog (RME DIGI96/8 PAD) and presented via headphones (Sennheiser HD580). The headphones had been calibrated and equalised at the centre frequencies of the five masking noise bands.

2.1.2 Procedure

A three interval, three-alternative forced-choice adaptive procedure was used to determine the masked thresholds. A trial consisted of three masker intervals separated by 500 ms pauses. One randomly chosen interval contained the sinusoidal signal added to the masker. Listeners had to indicate the interval in which the signal was presented by pressing the corresponding key on the keyboard. After each trial, visual feedback was provided. The adaptive signal level adjustment followed a two-down, one-up rule to estimate the 70.7 % point of the psychometric function (Levitt, 1971). The initial step size was 8 dB. After each lower reversal it was halved until it reached the minimum step size of 1 dB. Then the step size

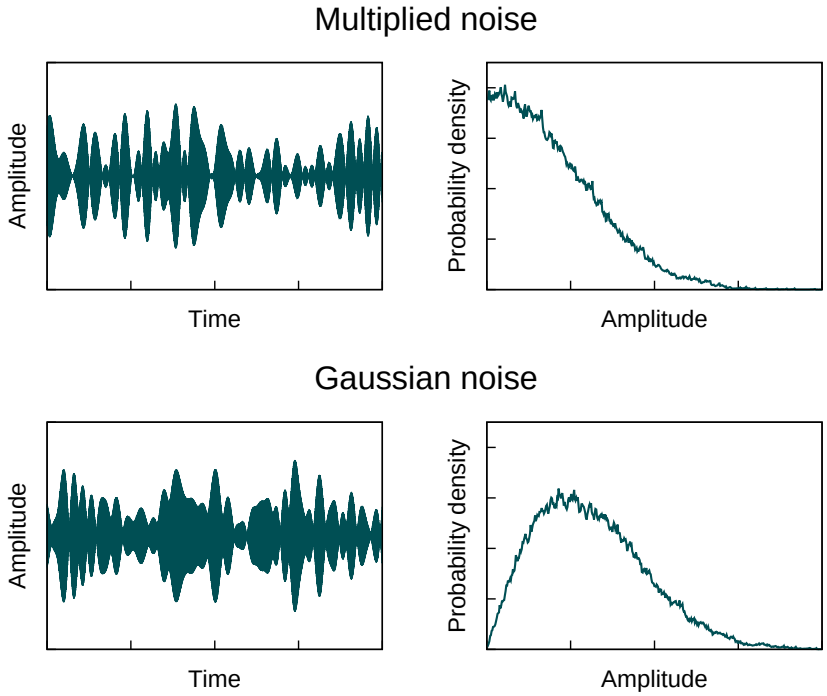

Figure 2.2: One realisation of a 24 Hz wide noise band at 700 Hz for multiplied noise (upper row, left) and Gaussian noise (bottom row, left) with their envelope amplitude distributions respectively (right panels), estimated over 100 seconds signals. The envelope amplitude distribution for the multiplied noise masker (upper, right) corresponds to the positive half of a Gaussian distribution while the Gaussian envelope amplitude distribution equals a Rayleigh distribution (figure 2 in Epp and Verhey, 2009a)

was kept constant for another six reversals. The arithmetic mean of these last six reversals was calculated and taken as the threshold estimate for that run. When the standard deviation of the reversals taken into account exceeded 4 dB, the threshold estimate for this run was rejected and an additional measurement was obtained. Each subject performed five valid threshold measurements for each condition. The threshold estimate for a given condition was determined by the mean of the final four valid runs. The different conditions were presented in random order within blocks, where each condition occurred once. Subjects were allowed to take breaks at any time after completing a run during the experiment.

2.1.3 Listeners

Ten listeners (two female, eight male), aged between 25 and 34, participated in the experiment. One of them was the author, KE. Except subject SG, all participants had previously taken part in psychoacoustic studies. All listeners reported normal hearing in the relevant frequency range from 250 to 2000 Hz. Listeners were seated in a double-walled, sound-attenuating booth during the experiment.

2.2 Results

Individual thresholds for the multiplied noise masker are shown in figure 2.3. In each panel the obtained thresholds for one subject are plotted as a function of IPD. Thresholds are shown in decibels relative to the level of the SCB. In general, the RF (circles) and the UN (squares) masker conditions showed higher thresholds than the CM (triangles) condition. Thresholds for the masked sinusoidal signal with an IPD of 0° were around 0 dB relative to the level of the SCB for the RF and UN conditions, except for subject CV, who had considerably higher thresholds in the UN condition (0° IPD) than the rest of the subjects. For all listeners, thresholds within all masker conditions were lower in the dichotic (IPD = 180°) than in the diotic case (IPD = 0°). This benefit of signal detectability due to the interaural disparity refers to the BMLD. The magnitude of the BMLD differed for the different masker conditions and over all individual listeners. BMLD for the narrowband condition (RF condition) ranged between 10 dB (subject BO) and 27.5 dB (subject NL). Individual differences were also found for the multiband conditions. BMLD varied from approximately 13 dB (subject BO) to approximately 20 dB (subject RD) in the UN condition, and from approximately 7.5 dB (subject BO) to 16.5 dB (subject RD) in the CM condition. The obtained thresholds also show inter-individual differences in the effect of number of flanking bands present on the BMLD. For half of the listeners the largest BMLD could be found in the RF condition. Except of listener RD, BMLD was smallest for the comodulated masker.

A rather large inter-subject variability was also detected in the magnitude of CMR. Comparing the CM condition with the RF and UN conditions, a decrease in CMR was found with increasing IPD (N_0S_π) for most listeners. Only subject RD showed larger CMR when binaural cues and comodulated bands were added simultaneously. The diotic CMR (UN-CM and RF-CM) varied from approximately 7 dB for listener CV to 16 dB for listener AC. For an IPD of 180° the magnitude of CMR (UN-CM and RF-CM) ranged between nearly 12 dB (subject RD) and -2 dB (subject NL). The negative CMR value indicates that in this condition the across-frequency information introduced by the comodulated masker did not result in an enhancement in signal detectability. This occurred for listeners NL and KE, where the effect of CMR quantified in the single-band condition (RF-CM) completely disappeared for an IPD of 180°.

The data for the Gaussian masker type showed similar behaviour (see figure A.1 in appendix A). As seen for the multiplied noise, thresholds within all masker conditions were lower in the dichotic than in the diotic case for all subjects. Highest thresholds were obtained in the RF and UN condition. For six

of the listeners, thresholds in the diotic RF and UN condition were similar. For the other four listeners, thresholds differed up to 5 dB in these two conditions. Subjects CV, NL, and SG showed higher thresholds for the diotic UN condition than the corresponding RF condition, whereas thresholds for subject JK showed the opposite behaviour (diotic RF higher than UN). As for the multiplied masker, subject CV had considerably higher thresholds in the diotic UN condition than all other subjects. In general, thresholds were slightly higher than for multiplied noise compared between the corresponding conditions, respectively.

Large inter-subject variability was found for the magnitudes of CMR and BMLD. BMLD ranged between 10 dB (subject BO) and 22 dB (subject RM) in the RF condition, between 7 dB (subject BO) and approximately 21.5 dB (subject RM) in the UN condition, and varied from approximately 8 dB (subject BO) to 17.5 dB (subject NL) in the CM condition. Comparing the magnitudes of CMR between subjects, except for subjects BO, RD, and SG, a decrease in CMR was found with increasing IPD (N_0S_π). The diotic CMR (UN-CM and RF-CM) varied from approximately 6.5 dB (subject JK) to nearly 12 dB (subject CV), only subject NL showed less effect of CMR (RF-CM) (approximately 3 dB) for an IPD of 0°. For an IPD of 180° the magnitude of CMR (UN-CM and RF-CM) ranged between nearly -2 dB (subject CV) and approximately 11 dB (subject RD).

Mean results over all subjects for both masker types are plotted in figure 2.4. The upper panels provide the arithmetic mean of the thresholds over all subjects as a function of IPD, plotted as in figure 2.3. The middle panels show the respective CMR calculated for each value of IPD. Circles indicate the CMR quantified by the difference in threshold between the single-band and the CM condition (i.e., CMR (RF-CM)) and squares represent the CMR calculated as the difference between the thresholds of both multiband conditions, CMR (UN-CM). In the lower panels BMLDs are plotted which refer to the difference in threshold between the dichotic and its corresponding diotic condition for each masker configuration (RF in circles, UN in squares and CM in triangles). Hence, the BMLD was zero for an IPD of 0° by definition.

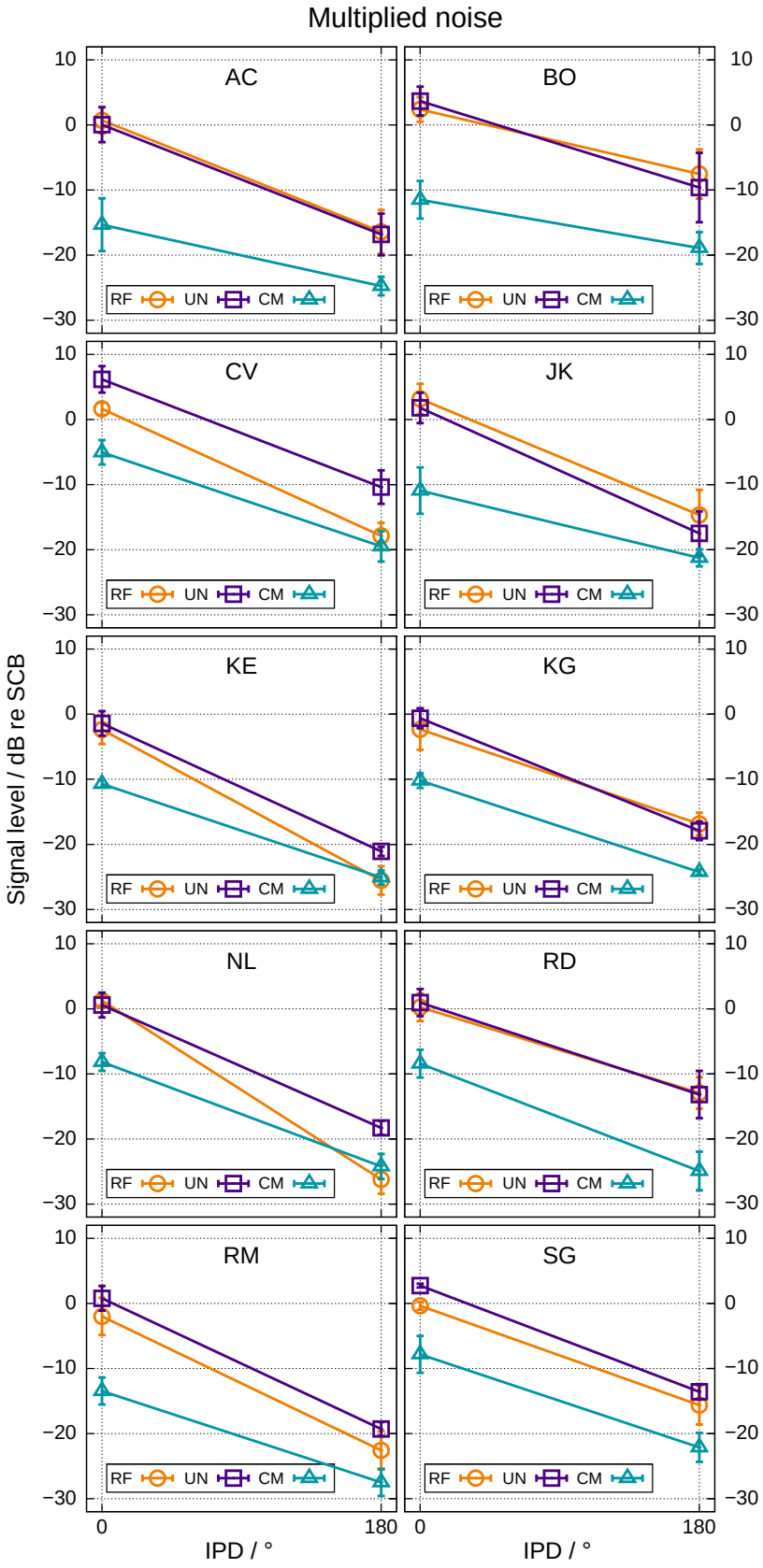

Figure 2.3: Individual masked thresholds for the multiplied noise masker. Mean thresholds over four runs are plotted as a function of IPD. Thresholds are shown relative to the level of the signal centred masker band for the RF (circles), UN (squares), and CM conditions (triangles). Error bars indicate ± 1 standard deviation.

Mean thresholds were higher in diotic conditions (N_0S_0) than in the corresponding dichotic conditions (N_0S_π). As the individual data already indicated, the mean results show that highest thresholds were found in the UN condition and lowest thresholds in the CM condition. CMR showed a decreasing behaviour with increasing IPD for both ways of quantification (RF-CM and UN-CM). CMR reached an average magnitude of approximately 11.5 dB for 0° IPD (multiplied noise) for the definition comparing both multiband conditions (UN-CM) and it decreased to approximately 7.5 dB in the corresponding antiphasic condition. Magnitudes for CMR were slightly lower for RF-CM than for UN-CM (approximately 1 dB difference in the diotic case and up to 2 dB in the dichotic case). BMLD magnitudes reached values up to 17-18 dB in the RF and UN conditions, independent of the masker type. For the CM condition, BMLD was smaller, 13 dB for multiplied noise and 14 dB for Gaussian noise.

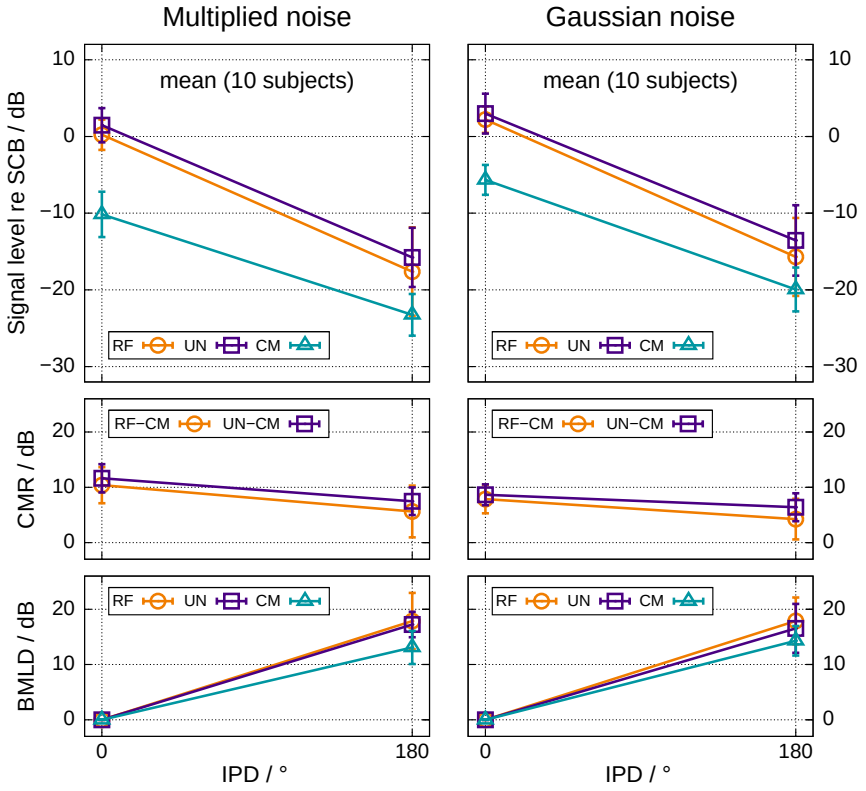

Figure 2.4: Mean results over all subjects for multiplied noise (left column) and Gaussian noise (right column). The upper panels show the mean detection thresholds over all subjects as a function of IPD. Thresholds are plotted relative to the level of the SCB for the RF (circles), UN (squares), and CM (triangles) conditions. The middle panels indicate the average CMR (RF-CM, circles and UN-CM, squares). The average BMLD is plotted in the lower panels for the RF (circles), UN (squares), and CM (triangles) conditions. Error bars indicate ± 1 standard deviation.

Higher thresholds were obtained for Gaussian noise than for multiplied noise in each condition, respectively. Furthermore, Gaussian noise showed a smaller

effect of CMR compared to multiplied noise (up to 3 dB difference for N_0S_0). BMLD reached similar values for both masker types. The variation of the masked thresholds (indicated by the standard deviation) was generally higher for dichotic than for diotic signals. The by far largest standard error was found in the RF condition for a signal IPD of 180°, for both multiplied and Gaussian noise.

Furthermore, a three-way analysis of variance (ANOVA) was performed on the CMR data. The analysis confirmed a significant main effect of signal phase (N_0S_0 or N_0S_π) [$F(1,76) = 29.80$, $p < 0.01$] and masker type [$F(1,76) = 8.7$, $p < 0.01$], but no significant effect of quantification type (RF-CM or UN-CM) [$F(1,76) = 4.96$, $p = 0.028$] was found.

2.3 Discussion

The present study investigated masked thresholds and their alternation by adding across-frequency information to the masker (comodulation) as well as interaural disparities to the signal. The amount of masking release was varied by a variation and combination of both cues. Data showed a threshold improvement in the presence of either one of both used cues as well as when they were introduced simultaneously. In general, results for the measured thresholds are similar to the data found in the literature (e.g., Epp and Verhey, 2009b; Hall *et al.*, 2011).

Considering the cues leading to masking release individually, a magnitude of approximately 10.5 dB CMR (RF-CM) in the diotic condition for the multiplied noise masker is in good agreement with previous studies. For example, Epp and Verhey (2009b) found an average CMR (RF-CM) of 10 dB in the corresponding condition. Moreover, Hall *et al.* (1990) obtained a slightly higher average CMR of approximately 12.5 dB for a signal at 700 Hz embedded in multiplied noise bands of 20 Hz bandwidth that were located at the same centre frequencies as in the present study. The magnitude of CMR for the Gaussian noise masker is also similar to data found in literature (e.g., Hall *et al.*, 2011). Magnitudes of the binaural benefit (BMLD) found in the present results are in good agreement with data obtained in previous studies (e.g., Hall *et al.*, 2011). However, the finding of Epp and Verhey (2009b) that BMLDs were the same for the two multiband conditions (UN and CM) and smaller than the narrow-band condition (RF) cannot be supported. Rather a decrease in BMLD for the CM condition could be found where BMLDs reach up to approximately 4 dB smaller values than in the UN and RF conditions. This coincides with the findings of Hall *et al.* (2011) who showed a significant effect of comodulation on BMLD magnitudes. Further, mean data showed that BMLDs are higher for smaller masker bandwidths. However, the individual data indicate that for only half of the listeners the maximum

BMLD was obtained in the RF condition (multiplied noise masker). In general, large individual differences in the effects of CMR and BMLD was found in the measured data. Such high disparity within subjects had been reported before (e.g., Epp and Verhey, 2009b). Moreover, Bernstein *et al.* (1998) and Buss *et al.* (2007) indicated that the individual differences in binaural detection and thus the large range of BMLDs they obtained using a 500 Hz signal in a narrow-band noise masker resulted from the variability in N_0S_π thresholds across listeners.

The present data show that thresholds for the multiplied noise masker were lower than the corresponding thresholds for Gaussian noise. This supports the findings of Verhey *et al.* (2007) who investigated the influence of envelope distributions of a 100 Hz wide masking noise on signal detection of a 4 kHz tone. They revealed that masked thresholds depend on the envelope distribution of the masker and its change when adding a signal. They also predicted their data with a model which used the change in envelope amplitude distribution due to the addition of the signal as a cue. Lower thresholds for multiplied noise might be related to a bigger change in envelope distribution compared to Gaussian noise when the signal is added (figure 3 in Verhey *et al.*, 2007).

CMR reached higher values for multiplied noise than for Gaussian noise, whereas no influence of the masker statistics on the magnitude of BMLD could be observed. This agrees with the findings of Epp and Verhey (2009a,b). Epp and Verhey (2009b) assumed that the smaller modulation depth of Gaussian noise compared to multiplied noise might account for the difference in CMR. As shown in figure 2.2, the envelope amplitude statistics of multiplied noise obtains higher probability values for zero amplitudes. This might be critical for listening in the dips (Buus, 1985) and thus, taking advantage of the masking noise dropping to low levels more frequently which consequently leads to lower thresholds.

Considering the combined effect of comodulation and interaural differences, contradictory results can be found in literature. First, Hall *et al.* (1988) observed lower magnitudes of CMR (RF-CM) when an IPD was introduced to the signal compared to the diotic condition, even though with substantial individual differences. Cohen and Schubert (1991), Hall *et al.* (2006) as well as Hall *et al.* (2011) found a decrease in CMR when the interaural signal phase was changed from 0 to 180° for both ways of quantification (RF-CM and UN-CM), with a stronger effect noticeable for RF-CM. In contrast, Epp and Verhey (2009b) found that CMR in dichotic conditions clearly depends on the definition of the CMR. Their data showed that the CMR for RF-CM indeed diminished with the introduction of an IPD, while CMR (UN-CM) remained constant, independent of the interaural phase difference in the signal. This supported the hypothesis of Epp and Verhey (2009a) that the effects of CMR and BMLD are additive and thus,

processed independently. Epp and Verhey (2009a) also used a simplified model of the auditory system which assumed an independent and serial processing of across-frequency and binaural cues. Model predictions approved their hypothesis. To explain the inconsistency between their results and those of Hall *et al.* (2006), Epp and Verhey (2009b) suspected that due to the different stimuli characteristics within-channel effects contributed to the larger CMR in the diotic condition of Hall *et al.* (2006) and therefore, led to a reduction in CMR in the diotic condition (CMR for UN-CM dropped from approximately 11.5 dB to 5.5 dB with the introduction of IPD in Hall *et al.*, 2006). In the present study a smaller CMR, both RF-CM and UN-CM, was obtained in the N_0S_π conditions compared to N_0S_0, which is in qualitative agreement with Hall *et al.* (2006, 2011). This finding weakens the assumption of Epp and Verhey (2009b) that within-channel effects contribute to CMR for spectrally close flanking bands in Hall *et al.* (2006). This rather might be only partly valid, as the present data obviously showed an effect of introduced IPD on CMR (RF-CM and UN-CM), but the reduction of CMR in the dichotic condition was not found as steep as in Hall *et al.* (2006). This agrees with Hall *et al.* (2011) who found a significant effect of phase (S_0 or S_π) on CMR, which also is in accord with the previously mentioned effect of comodulation on BMLD magnitudes. Furthermore, Hall *et al.* (2011) related the discrepancy between the results to the different way of presenting the noise stimuli, as Epp and Verhey (2009b) used gated noise whereas Hall *et al.* (2006) used continuous noise.

While the present results do not fully replicate the findings of Epp and Verhey (2009a,b), and therefore cannot support that masking releases resulting from across-frequency information, reflected by CMR (UN-CM), and binaural cues (BMLD) are additive in decibels, it still can be concluded that CMR and BMLD *"are, to some extent, additive"* (Cohen and Schubert, 1991). It was shown that by adding both cues simultaneously thresholds could be lowered further, even when the signal detectability had already been improved due to the presence of either comodulation (CMR) or interaural disparities (BMLD). Both present data as well as previous studies showed large individual differences across subjects. These results could be interpreted as different types of listeners. Analysing the data for the dichotic conditions more precisely reveals that listeners who used the binaural cues most effectively, and therefore, obtained the highest BMLDs, tended to have fairly small or fully absent CMR. For instance, subjects KE and NL, who showed a very strong effect of BMLD in the RF condition (approximately 23 dB and 27.5 dB), did not benefit at all by adding comodulated flanking bands when thresholds had already been reduced by the introduction of binaural cues. For those two listeners, the magnitude of CMR (RF-CM) decreased to -0.5 dB

(subject KE) and -2 dB (subject NL) for an IPD of 180° (multiplied noise). In contrast, for example subjects KG and SG showed moderate BMLDs (approximately 15 dB in the RF condition, multiplied noise), and the corresponding magnitudes of CMR (RF-CM) nearly remained constant with increasing IPD: CMR (RF-CM) only slightly decreased from approximately 8 to 7.5 dB (subject KG) and from approximately 7.5 to 6.5 dB for subject SG. Similar behaviour had also been observed in Hall *et al.* (2011). This might be explained by a kind of saturation effect of threshold improvement which is reached when cues that lead to masking release are introduced simultaneously. It is suggested that listeners who have already benefited largely from the introduction of binaural information are barely able to further improve signal detection by using across-frequency cues.

Chapter 3

Experiment II: Salience rating for signals above masked threshold

The second experiment deals with the assessment of perceptual aspects of a masked signal above masked threshold in both the presence and absence of stimulus properties that lead to a masking release. In experiment I, the individually masked thresholds in different conditions of masking release were measured. Increasing the signal intensity to suprathreshold levels results in a signal that is perceived more easily, suggesting its internal representation is more clear. To quantify the perception of the changing audibility, experiment II attempted to estimate the strength of the internal representation using a perceptual correlate. A perceptual measure, which is referred to as *salience*, was found. It is primarily determined by its signal level above masked threshold and describes how well the tone (signal) stands out of its background (masker). This measure can be related to the audibility of the signal.

3.1 Introduction

In the process of defining a perceptual measure which is related to the strength of the internal represenation and, thus, the audibility of a masked sinusoidal signal, different approaches have been considered. For conditions of BMLD, the measure of perceived intensity (i.e., loudness) was used to assess perception at suprathreshold levels in earlier literature (Townsend and Goldstein, 1972; Soderquist and Shilling, 1990; Fastl and Zwicker, 2007). All of these studies used similar procedures. For instance, Townsend and Goldstein (1972) used a loudness balance procedure in which the loudness impressions of a masked diotic and a masked dichotic sinusoidal signal were compared. Then, the signal level of the dichotic presentation was adjusted in a way such that both sinusoidal signals were per-

ceived equally loud. This was done for different levels above masked threshold (dB SL). Thus, the first approach for experiment II was to adopt this measure to quantify the suprathreshold perception in conditions of CMR and, further, apply the same measure to conditions where a combination of cues is present. However, during pilot testing, the comparison of the partial loudness of a masked diotic versus a masked dichotic sinusoidal signal was found to be rather difficult. A diotic sinusoidal signal that is presented interaurally in phase elicits a very defined and clear sensation of a tone, centred in the middle of the head. In contrast, a dichotic sinusoidal signal is interaurally out of phase, which provokes a spread image of a tone in the head. The tone is perceived differently, coming from one side or even with an unstable, moving impression of location. A masked dichotic sinusoidal signal presented at low to moderate SLs can be clearly perceived, although the loudness impression might be small. Hence, the measure of loudness was not found to be the appropriate choice to assess the audibility of the masked sinusoidal signal.

A perceptual measure, which describes how well the sinusoidal signal can be segregated from its background noise, was seen as being more representative. This measure, which reflects the degree to which the sinusoidal signal stands out of its embedding masking noise, was termed *salience*. It was inspired by the studies of Roberts and Bregman (1991) and Roberts (1998). In one of their experiments, Roberts and Bregman (1991) developed a clarity rating task to study the extent to which a single frequency component (i.e., harmonic) was integrated into a complex tone. Sequences of two tones (A and B) were presented. The complex tone B was preceded by tone A, which consisted of a pure tone with the same frequency as one of the harmonics present in tone B. Subjects were told to listen for a copy of tone A, which was also included in tone B. Subsequently, the subjects were required to rate the perceived clarity of the cued component in the complex tone B on a rating scale (i.e., they judged how salient tone A was within the complex tone B). The derived measure of perceived clarity determined to which extent a single harmonic can be heard out from a complex tone. Roberts (1998) pointed out that this clarity rating procedure had a few drawbacks. Importantly, the method assisted analytical listening, i.e., presenting the pure tone A before the complex tone B cued the corresponding frequency region. In order to avoid this problem and guarantee an estimation of salience of an uncued target component, Roberts (1998) inverted the order of presentation of pure tone and complex tone.

Several types of procedures have been considered to quantify the salience of a masked sinusoidal signal. Comparable studies investigating perceptual measures at suprathreshold levels have used *magnitude estimation* (e.g., Hellman, 1982), *matching paradigms* (e.g Townsend and Goldstein, 1972; Edmonds and Culling,

2009; Verhey and Heise, 2012), or *intensity discrimination tasks* (e.g., Hall and Grose, 1995; Wojtczak and Viemeister, 2008; Buss and Hall, 2009). For example, Hellman (1982) obtained judgements of loudness, annoyance, and noisiness for tones within broadband noise by absolute magnitude estimation, supplemented by a loudness matching task. Hall and Grose (1995) determined amplitude discrimination for masked sinusoidal signals at suprathreshold levels in conditions of masking release. For the present study a magnitude estimation paradigm was chosen to obtain judgements of salience of a masked sinusoidal signal for different levels above masked threshold (dB SL). This procedure was found to be suitable for the task, as it is based on a ratio scale. Therefore, the magnitude of the resulting salience can be easily compared across the different SLs. Furthermore, magnitude estimation is potentially less time consuming than a paired comparison procedure for which pairwise judgements across all stimuli conditions have to be obtained.

3.2 Method

3.2.1 Stimuli and apparatus

Stimuli were identical to experiment I. Stimuli consisted of a 250 ms sinusoidal signal with a frequency of 700 Hz, temporally centred in 500 ms masking noise, both with 50 ms raised-cosine on- and offset ramps. The signal was presented either with 0 or 180° IPD, whereas the masker was always presented diotically. The same three different masker conditions (RF, UN, and CM) as in experiment I were used. For further details, please refer to subsection 2.1.1. Multiplied noise was used for the maskers. Multiplied noise was chosen because experiment I showed stronger effects of masking release, especially CMR, for multiplied noise compared to Gaussian noise.

The perceptual measure was quantified at fixed levels of the masked sinusoidal signal in dB SL. Under the hypothesis that the perceptual measure of salience is related to the audibility of the signal masked by noise, it was assumed that the measure depends solely on the level above masked threshold. In experiment I, masked thresholds were obtained. These thresholds describe the point at which the internal representation of the sinusoidal signal is just strong enough to be identified as such. That means that for all tested conditions, the signal is just perceivable within the noise. The signal becomes more audible (i.e., its internal representation becomes clearer) when the signal level is increased. Hence, the higher the level, the more salient the signal, because it can be segregated more easily from its masking background noise. The level of the sinusoidal signal

was individually adjusted, based on the measured thresholds in each condition. Salience ratings were obtained for six different signal levels: -8, 0, 4, 8, 12 and 16 dB SL (i.e., relative to masked threshold). Since the individual thresholds differed among subjects, actual physical levels measured in dB SPL varied across listeners.

Stimuli were presented paired with a fixed *reference*, which consisted of a diotic sinusoidal signal with a frequency of 700 Hz, presented at a fixed level of 70 dB SPL in all conditions. The purpose of the reference signal was to demonstrate an ideally clear internal representation of the signal alone, with no masking noise present. Thus, the reference signal provided a reference for the maximum possible salience. As only a moderate level above signal threshold was sufficient to achieve this ideal presentation, it was not necessary to adjust its level in dB SL individually for every test subject.

All stimuli were generated digitally in MATLAB with a sampling rate of 44100 Hz and 16 bit resolution, converted from digital to analog (RME DIGI96/8 PAD) and presented via headphones (Sennheiser HD580). The headphones had been calibrated and equalised at the centre frequencies of the five masking noise bands.

3.2.2 Procedure

Tested stimuli were presented in trials, paired with the reference signal. In each trial subjects listened to the reference (sinusoidal signal) as well as the test signal to be judged (sinusoidal signal plus noise). Both could be played as often as the listener wished and were accessed via two separated buttons, either by pressing the corresponding keys on the keyboard or clicking with the mouse. In each condition, the number of times the subject listened to the reference signal as well as the test signal was recorded. The subjects were not made aware of this. The subject's task was to rate the salience of the sinusoidal signal within the test signal (i.e., how well the sinusoidal signal could be segregated from its background noise). Judgements were obtained by magnitude estimation on an endpoint-anchored scale, where the listener placed the rating by moving the slider's position on the given scale. The procedure was programmed in MATLAB. The experiment screen can be seen in figure 3.1.

The scale was anchored on the lower and upper ends by *not audible* and *reference* respectively. This direct grading scale was used to map the subjects' internal perceptual judgement onto a response distribution, in this case, the position of the slider along the scale.

As in Guski (1997) recommended, a scale with labels at the endpoints but no labels in between was used rather than a verbal scale. The latter refers to a scale

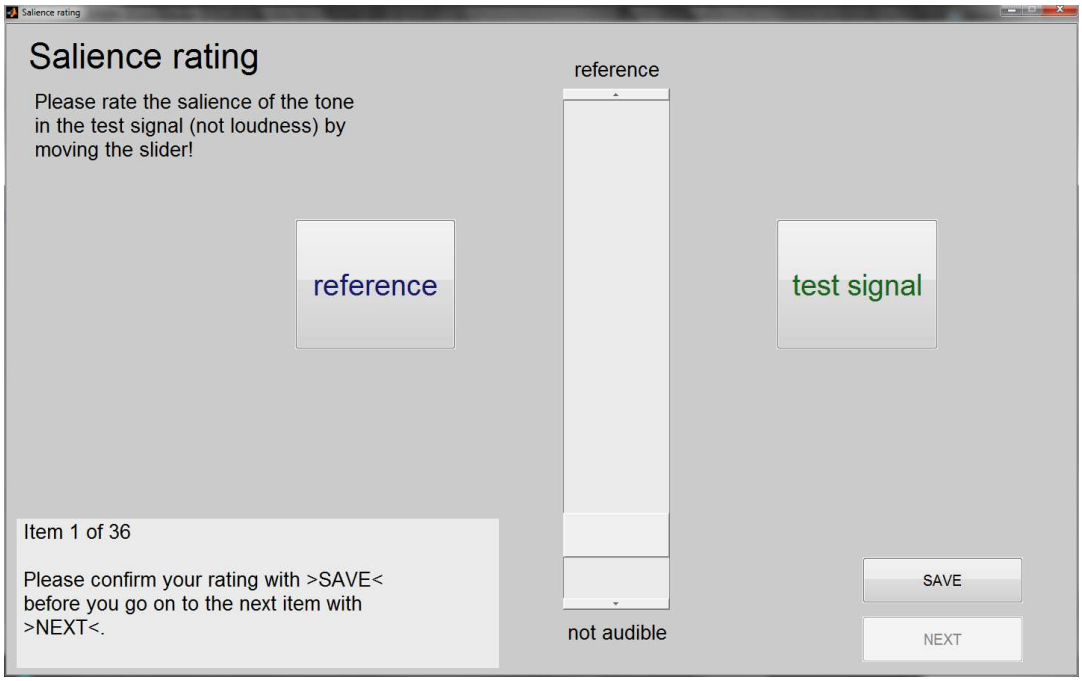

Figure 3.1: Screenshot of the user interface for the salience rating experiment. Reference and test signal could be played by clicking the corresponding buttons. Salience ratings were placed by moving the slider's position on the given scale.

with a set of numbers or quantifying labels at equal intervals. For example, a salience scale could be described by steps named 'not (1), a little (2), fairly (3), highly (4) and extremely (5) salient'. However, such a scale implicitly assumes internal perceptual distances that are equal and map linearly (Zieliński *et al.*, 2008). Unfortunately, we cannot be sure this assumption is correct. Hence, a fully labelled scale might have yielded in responses which would have been interpreted incorrectly. To avoid systematic errors, this type of scale was rejected.

The slider's position was mapped onto a numerical scale between 0 and 10 in the background, which the listeners were not aware of. The lower anchor of '*not audible*' was attributed to the zero value like on a ratio scale. Stimuli could be ranked along the scale according to their salience. The difference in perceived salience between stimuli on the scale could also be calculated. This is an advantage over a paired comparison procedure, where only a ranking of stimuli could have been obtained.

Stimuli were randomised and presented in blocks, where each condition occurred once. In total, each subject performed sixteen blocks of stimuli so that, for each condition, sixteen responses were obtained. The blocks were tested in three sets of blocks in three different sessions. In the first session, six blocks of repetitions were measured. Here, the first two blocks were considered as training and not included in the results. To familiarise the listeners with the task and stimuli, and to avoid learning effects due to this familiarisation process, the first

two runs were discarded for each subject. In the second and third session, five blocks were tested each. As the different sessions were often several days apart, the first response block of each of the two subsequent sessions were excluded from further data analysis. This was motivated by the fact that subjects might have had to readjust their internal scale. Overall, the salience rating for a given condition was derived from twelve responses per subject. Listeners were allowed to take breaks at any time during the experiment.

3.2.3 Task

Listeners were told that the experiment contained the same type of stimuli they had been listening to in the previously performed threshold detection measurements. The task was described as rating the salience of a sinusoidal tone embedded in noise (i.e., how well the tone can be segregated from its background). The following written instructions were displayed on the screen preceding the actual experiment.

> You are going to hear signal pairs consisting of a fixed reference (pure tone) and the test signal (pure tone plus noise). The tone in the test signal (target tone) and the reference have the same frequency, but other parameters may differ.
>
> Your task is to evaluate, how well the target tone can be segregated from its background noise (referred to as 'salience').
>
> It is important that you do not place your judgement only depending on the loudness impression of the test signal. Hence, please fully focus on how well the tone in the test signal stands out from the background.
> It may help to focus on the frequency of the tone in the test signal.
>
> You will be able to listen to both, the reference and the test signal repeatedly and as often as you wish. Please place your rating of the salience of the tone in the test signal by moving the slider's position on the given scale. The scale will be limited on the lower side by 'not audible' and 'reference' on the upper side, which represents the perception of the target tone as salient as the reference.
>
> Before the actual experiment there will be a training session, to help you understand and interpret the task correctly.

Furthermore, it was pointed out that there is *a* version of the reference in the target signal. It was indicated that the reference refers to an ideal that represents the maximum possible salience and lies above the rating scale. Thus, subjects were asked to use the full range of the scale.

To familiarise the listeners with the task, a short training session was provided before the initial run of the experiment. Since it is not commonly used, subjects might not know how to understand the term salience beforehand. Therefore, it was important to introduce the concept of the magnitude of salience of a sinusoidal signal embedded in noise, preferably with the help of stimuli examples. It might be argued that the presentation of the reference signal on its own was sufficient. However, salience describes how well the sinusoidal signal can be segregated from its background noise. Since the reference is a sinusoidal signal, which is presented alone, without masking noise, a magnitude of salience cannot be attributed to this single sinusoidal signal without noise in the first place, without knowing what to listen for (see Vormann *et al.*, 1998; Hansen *et al.*, 2011, about the introduction of the subjectively perceived magnitude of tonal content).

With four buttons, listeners could access four different types of signals. By pressing the first button the reference tone was played. The second button presented a test signal where only the masking noise was present (i.e., there was no pure tone). By clicking this button, a noise sample was randomly drawn from the different masker conditions (RF, UN, or CM). This '*only noise*' test signal was only available during the training session. Button three provided a test signal where the target tone was assumed to be highly salient. In particular, pressing this button played a test signal sample that was randomly drawn from the available six different conditions (0 or 180° IPD of the signal in either RF, UN, or CM noise), in which the sinusoidal signal was 24 dB above masked threshold. Thus, the sensation level of the sinusoid was 8 dB higher than the maximum level of the target tone in the actual experiment, and the tone was clearly perceived within the noise. The fourth button played a test signal with the salience of a random target tone. Here, one of the stimuli used in the subsequent experiment was randomly selected. This allowed the listeners to become familiar with the range of stimuli and helped them to develop an own, internal scale on which they could map their jugdements. As in Zieliński *et al.* (2008) and Bech and Zacharov (2006) suggested, this familiarisation process was used in order to reduce the contraction bias. This bias refers to the response mapping bias which describes the behaviour of listeners normally avoiding the extremes of a scale (Zieliński *et al.*, 2008). This might happen when the stimulus set is unknown as subjects would reserve the extreme values on the given scale for a more extreme, but yet unknown, stimuli presented later in the experiment. Hence, their given ratings *"tend to cluster*

around the centre of the scale" (Bech and Zacharov, 2006). Additionally, direct anchors were applied, in our case the verbal terms and stimuli at the end points of the scale, to further reduce the contraction bias (Poulton, 1989).

3.2.4 Listeners

The listeners were the same as those in experiment I. The individual detection thresholds determined in experiment I provided the basis for the stimuli of the second experiment. Listeners were seated in a double-walled, sound-attenuating booth during the experiment.

3.3 Results

Individual salience ratings for two subjects (KG and BO) are shown in box-and-whisker plots in figures 3.2 and 3.3. Magnitude estimates for the salience of a masked sinusoidal signal were obtained for the RF (upper row), UN (middle row), and CM (bottom row) condition, with either a signal IPD of 0° (indicated by the subscript $_0$) or 180° (indicated by the subscript $_\pi$). For each condition, the ratings are plotted as a function of signal level in decibels relative to the corresponding masked threshold (dB SL). The boxes display the lower and upper quartiles (i.e., the 25^{th} and the 75^{th} percentile, respectively) and the median. Whiskers indicate the lowest occurring response within 1.5 interquartile range (IQR) of the lower quartile and the highest occurring response within 1.5 IQR of the upper quartile. Outliers are marked by asterisks.

Salience ratings for subject KG (figure 3.2) showed a monotonically increasing behaviour with increasing signal level in each condition. No salience was observed (median ratings of 0) in all conditions for the lowest tested level of the sinusoidal signal (-8 dB SL). For the highest signal level measured (16 dB SL), a salience of approximately 8.4 was observed for the diotic UN condition (UN_0), 4.8 for UN_π, 5.5 for CM_0, 4.6 for CM_π, 3.6 for RF_0 and 2.8 for RF_π (all median values). Hence, the smallest increase in salience was obtained in the RF conditions. In contrast, results for subject BO (figure 3.3) showed that salience increased with increasing signal level for the UN and CM conditions but, in the RF conditions, constant high ratings (median salience between approximately 8.7 and 9.3) for all measured signal levels were found. Generally, salience ratings of subject BO for the CM conditions were similar to subject KG, though with higher variances. Salience in the UN conditions increased with increasing signal level. However, the minimum salience in these two conditions was higher than for subject KG. For the lowest tested signal level, a median rating of 4.2 (UN_0) and 3.3 (UN_π)

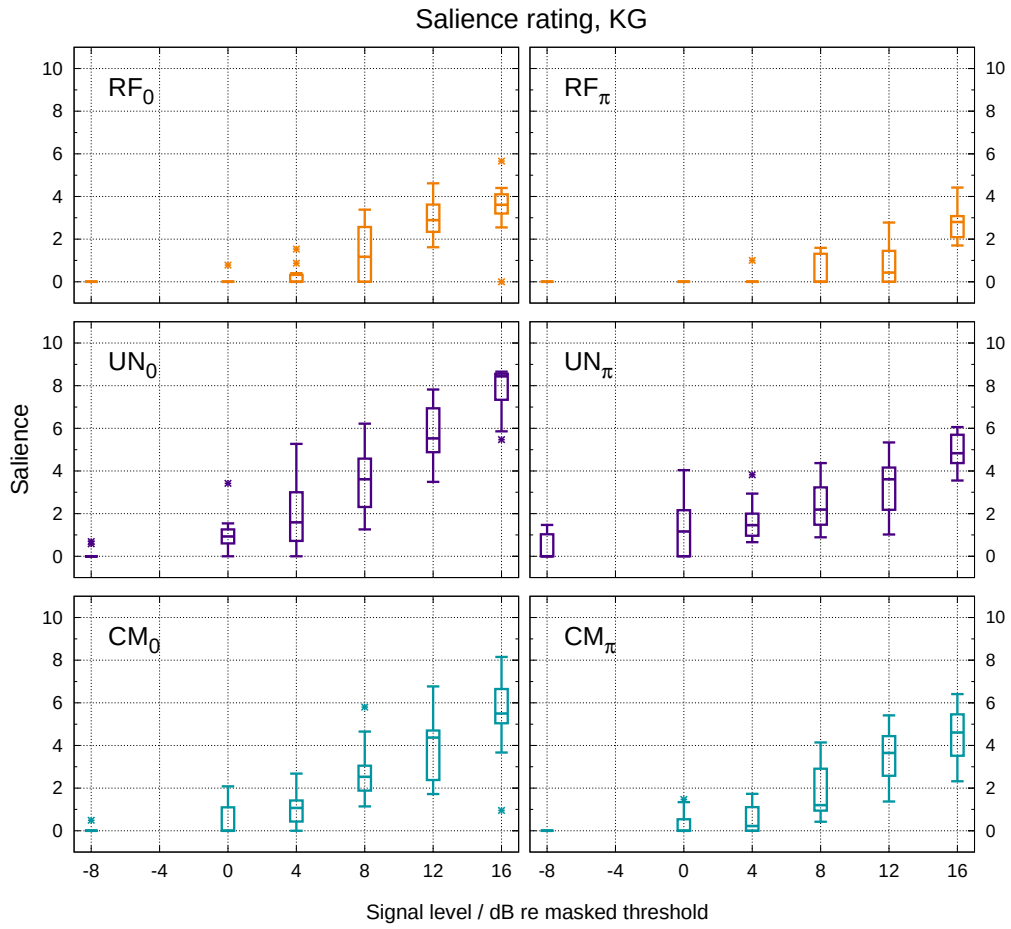

Figure 3.2: Box-and-whisker plot of the obtained salience for subject KG for the RF (upper row), UN (middle row), and CM (bottom row) condition, with a signal IPD of 0° (left panels) or 180° (right panels). The subscripts $_0$ and $_\pi$ indicate the IPD of the masked sinusoidal signal. Ratings are plotted as a function of signal level relative to the corresponding masked threshold. The box ranges from the lower to the upper quartile and is split with a line at the median. Whiskers indicate the lowest occurring response within 1.5 interquartile range (IQR) of the lower quartile and the highest occurring response within 1.5 IQR of the upper quartile. Outliers are marked by asterisks.

was observed.

Analysing all individual data (see appendix B) showed that salience increased with increasing signal level for the CM conditions for all subjects and, with the exception of RM, also in both UN conditions. For the RF conditions, a high inter-subject variability was found. For half of the subjects, the salience in the RF conditions increased with increasing level, and this increase was similar in magnitude to that found in the UN and CM conditions. Generally, a larger increase in salience was observed in the RF_0 than in the RF_π condition. Subjects RD and JK also showed an increase in salience with increasing level in the RF conditions, though the magnitude of this increase was much smaller over the measured range of signal levels. Furthermore, the salience ratings for the lowest tested signal level were rather high for those two subjects (median between 6.2

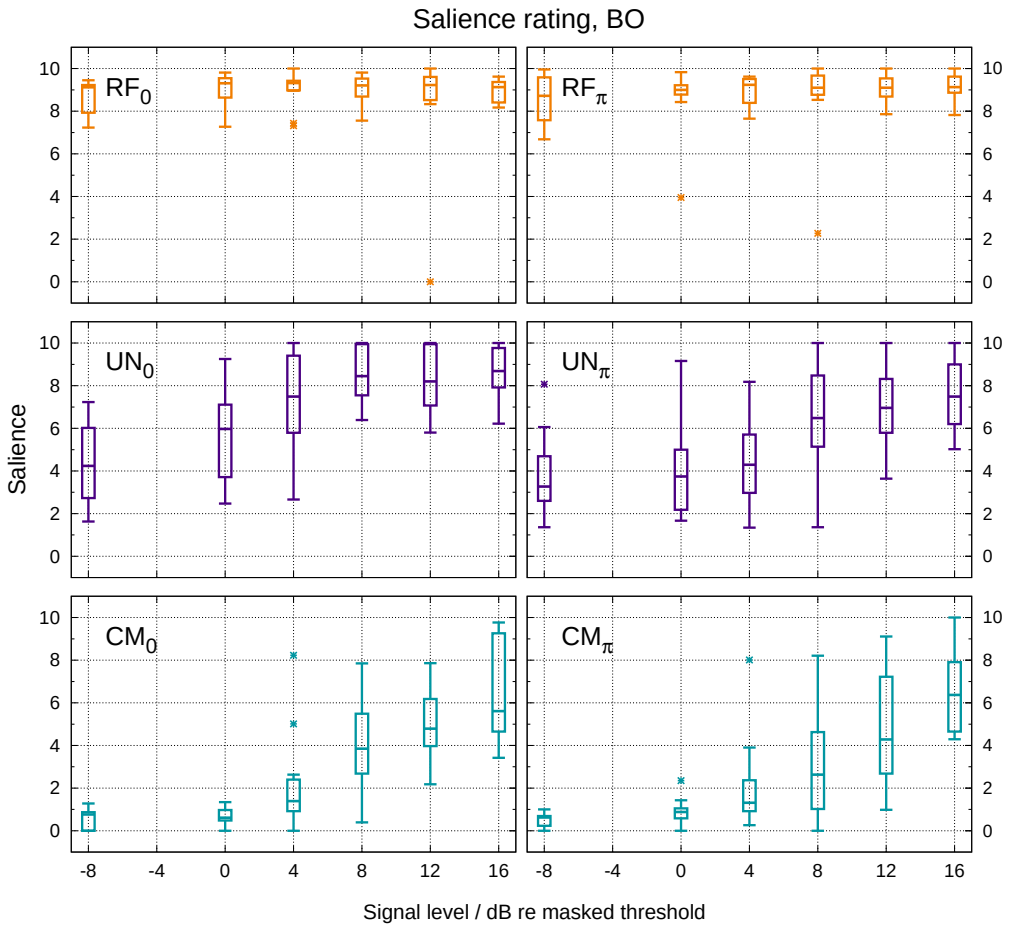

Figure 3.3: Box-and-whisker plot of the obtained salience for subject BO for the RF (upper row), UN (middle row), and CM (bottom row) condition, with a signal IPD of 0° (left panels) or 180° (right panels). Ratings are plotted as a function of signal level relative to the corresponding masked threshold.

and 7.1). For three subjects (BO, CV, and RM) the salience for RF_0 and RF_π was approximately constant for all signal levels. For subjects BO and RM, relatively high ratings were observed (median salience between 8.7 and 9.3 for subject BO, between approximately 8 and 9.6 for subject RM). No salience at all was perceived by subject CV.

Mean results over all subjects are illustrated in figure 3.4. For each condition, the arithmetic mean of the average salience rating of each subject (arithmetic mean over 12 repetitions) is plotted as a function of signal level in dB relative to the corresponding masked threshold. Circles indicate the obtained salience rating for the RF condition, squares for the UN, and triangles for the CM condition. Open symbols correspond to an IPD of 0°, filled symbols to 180° IPD.

Mean salience increased with increasing signal level in each condition. In general, higher ratings were obtained for the RF and UN conditions than in the CM conditions. The highest variability over subjects, indicated by the standard error, was found for both RF conditions (RF_0 and RF_π). Data for the diotic UN

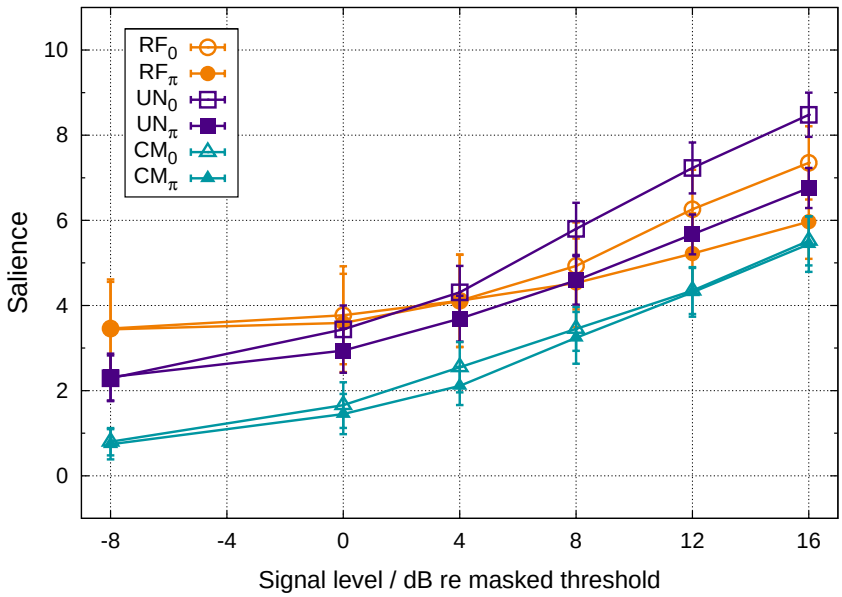

Figure 3.4: Average salience rating over all subjects. The arithmetic mean over all subjects is plotted as a function of signal level relative to masked threshold for the RF (circles), UN (squares), and CM (triangles) condition, with a signal IPD of 0° (open symbols) or 180° (filled symbols). Error bars indicate ± 1 standard error.

condition (UN_0) showed the steepest increase in salience with increasing signal level compared to the other conditions. This condition also obtained the highest absolute salience rating of about 8.5 for the highest signal level tested (16 dB SL). In each condition, the lowest salience rating was found for the lowest tested signal level, which was 8 dB below masked threshold. For this level, a salience of approximately 0.8 was measured in the CM conditions, 2.3 in the UN and 3.5 in the RF conditions. Comparing the obtained salience functions between the different conditions (over all signal levels) indicates that nearly the same salience was obtained for both CM conditions (CM_0 and CM_π), whereas the other conditions diverged. These observations were confirmed by a three-way ANOVA. Earlier on, Bartlett's test was performed in order to reveal differences between the variances in the different conditions. The test showed a significant difference in variance when the salience for all signal levels within all masker conditions were compared [$p < 0.01$], but no significant difference was found within the multiband conditions [$p = 0.994$] and within the RF conditions [$p = 0.998$]. Since the assumption of equality of variances among the tested conditions was not fulfilled, the ANOVA was applied to the mean data of the multiband conditions (UN and CM) and the RF condition separately. The analysis of UN and CM confirmed a significant main effect of level [$F(5,207) = 139.83$, $p < 0.01$], condition [$F(3,207) = 73.90$, $p < 0.01$], and subject [$F(9,207) = 44.849$, $p < 0.01$], but no significant interaction of level and condition [$F(15,207) = 1.00$, $p = 0.454$] was found. In order to determine, which conditions were significantly different from

each other, a post-hoc Tukey HSD test was carried out to compare all possible pairs of conditions. Except for CM_0-CM_π ($p = 0.780$), all variations of CM and UN conditions were significantly different ($p < 0.01$ each). Analysing the RF conditions showed a significant main effect of level [$F(5,99) = 26.93$, $p < 0.01$] and subject [$F(9,99) = 99.83$, $p < 0.01$]. However, no significant effect of condition [$F(1,99) = 6.46$, $p = 0.013$] and no significant interaction of level and condition [$F(5,99) = 1.41$, $p = 0.226$] was found.

In order compare the relative salience and, thus, the slope of the rating curves between the several conditions, the salience ratings were baseline corrected. This was done by subtracting the obtained magnitude of salience for the lowest signal level (-8 dB SL) from the ratings for the higher signal levels, for each condition separately. The resulting relative salience ratings are illustrated in figure 3.5. Comparing the relative salience between conditions indicates that the salience for both CM conditions (CM_0 and CM_π) as well as for UN_π increased with nearly the same slope with increasing signal level (approximately 4.4 - 4.7 increase in salience measured between the lowest and the highest signal level). UN_0 showed the steepest slope (6.2 difference between the salience for 0 and 16 dB SL), whereas the increase in salience was more flat for the RF conditions. The smallest gain in salience over the measured range of signal levels (only about 2.5) was observed for the dichotic RF_π condition. The three-way ANOVA was repeated on the baseline corrected salience ratings, again considering multiband and single-band conditions separately. The analysis of UN and CM showed a significant main effect of level [$F(5,207) = 153.84$, $p < 0.01$], condition [$F(3,207) = 11.93$, $p < 0.01$], and subject [$F(9,207) = 15.30$, $p < 0.01$], while no significant interaction of level and condition [$F(15,207) = 1.10$, $p = 0.355$] was found. Multiple pairwise comparisons using the Tukey HSD test confirmed the observations by showing no significant effect comparing CM_0-CM_π ($p = 0.927$), UN_π-CM_0 ($p = 0.504$), and UN_π-CM_π ($p = 0.864$), whereas UN_0-CM_0, UN_0-CM_π, and UN_0-UN_π were found to be significantly different ($p < 0.01$ each). Analysing the RF conditions showed a significant main effect of level [$F(5,99) = 26.93$, $p < 0.01$] and subject [$F(9,99) = 9.55$, $p < 0.01$], while no significant main effect of condition [$F(1,99) = 5.91$, $p = 0.017$] and no significant interaction of level and condition [$F(5,99) = 1.41$, $p = 0.226$] was found.

Another way to illustrate the salience rating is to plot the measured salience as a function of physical signal level instead of SLs. Therefore, the corresponding mean masked threshold (shown in figure 2.4, left panel) was added to the SL in each condition, respectively. This results in a horizontal shift of the salience rating curves according to the different masked thresholds for the conditions. Since the masked thresholds differed among the tested stimuli conditions, actual physical

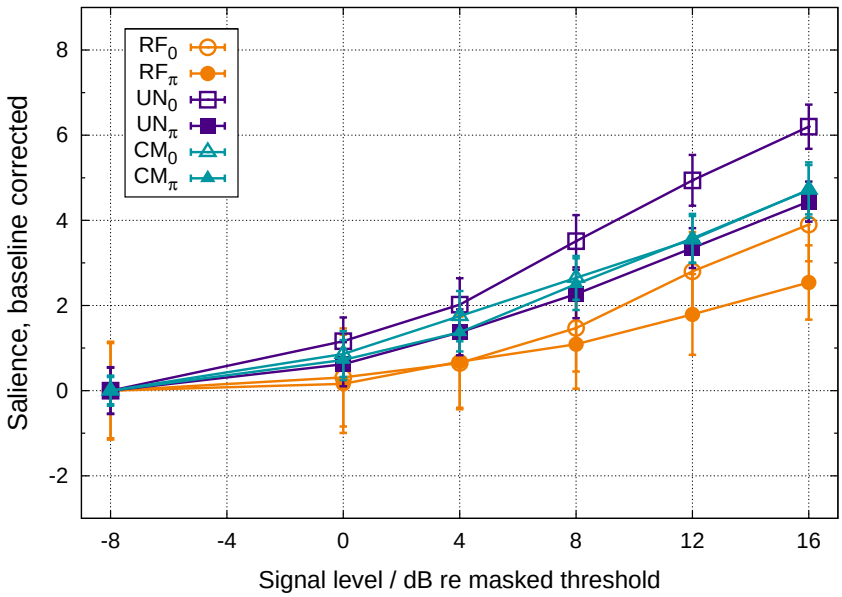

Figure 3.5: Average salience rating over all subjects, baseline corrected. The arithmetic mean over all subjects is plotted as a function of signal level relative to masked threshold for the RF (circles), UN (squares), and CM (triangles) condition, with a signal IPD of 0° (open symbols) or 180° (filled symbols). The magnitude for the obtained salience ratings were baseline corrected, i.e., ratings are displayed relative to the salience obtained at a signal level of -8 dB relative to masked threshold. Error bars indicate ± 1 standard error.

levels measured in dB SPL varied across the different conditions. Figure 3.6 shows the salience as a function of signal level relative to the level of the SCB of the masker.

Salience for the highest physical signal levels were measured for the UN_0 condition since this condition had the highest masked thresholds in experiment I. In contrast, lowest physical signal levels were tested in the CM_π condition as this condition had the lowest masked threshold in experiment I. In order to compare the relative salience as a function of physical signal level, the salience ratings were baseline corrected. The resulting salience displayed relative to the salience rating measured for the lowest signal level in each condition respectively is shown in figure 3.7.

The number of times subjects listened to the reference signal and the test signal (i.e., the sinusoidal signal embedded in masking noise) was counted and then the arithmetic mean over all subjects was taken resulting in the average play count, illustrated in figure 3.8. How often the subjects listened to the reference signal (upper panel) and the test signal (lower panel) is plotted as a function of signal level relative to masked threshold for the RF (circles), UN (squares), and CM (triangles) condition, with a signal IPD of 0° (open symbols) or 180° (filled symbols). Data points have been slightly shifted horizontally to increase readability. The play count for the reference signal was similar in all conditions

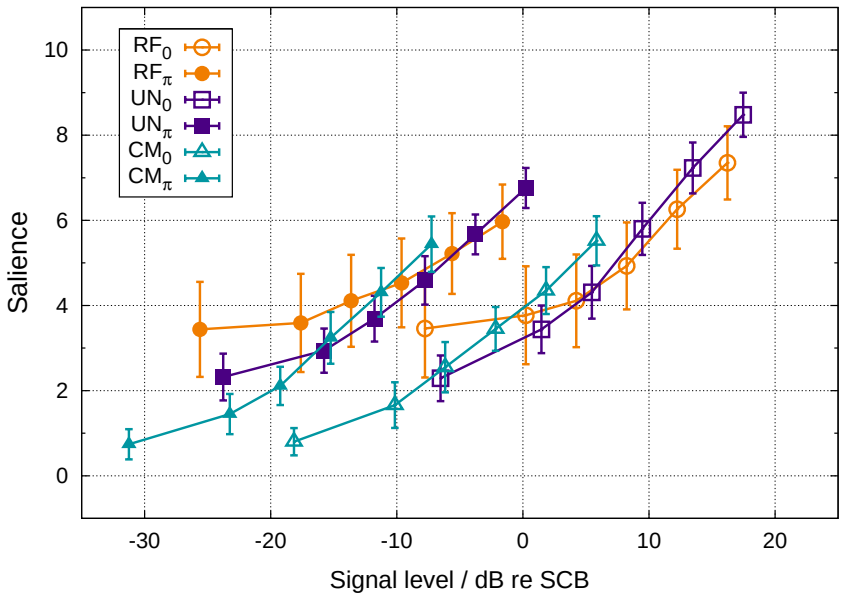

Figure 3.6: Average salience rating over all subjects. The arithmetic mean over all subjects is plotted as a function of signal level relative to the level of the SCB for the RF (circles), UN (squares), and CM (triangles) condition, with a signal IPD of 0° (open symbols) or 180° (filled symbols). Error bars indicate ± 1 standard error.

and for all tested signal levels. Taking the arithmetic mean over all conditions and levels shows that subjects generally listened approximately 2.75 times to the reference signal before placing a final rating for the salience of the sinusoidal signal. Higher play counts and much higher variability were found when analysing the listening behaviour for the test signal. For the test signal, the average play count determined over all conditions and levels was approximately 5.52. However, as illustrated in figure 3.8, for signal levels greater than -8 dB SL, mean data show that the subjects tended to listen more often to dichotic than the corresponding diotic signal presentations when compared within a particular masker condition.

3.4 Discussion

The present study quantified the perception of a masked sinusoidal signal above masked threshold. A perceptual measure of salience was derived that describes how well the sinusoidal signal can be segregated from its background. The measure was related to the strength of the internal representation and consequently, the audibility of the masked signal. The perception of the masked sinusoidal signal was quantified as a function of level above masked threshold. It was hypothesised that the perceptual measure is defined by the signal level above masked threshold (i.e., dB SL) and that equal magnitudes in salience are observed at equal levels above masked threshold, independent of the condition and absolute threshold. Though the results do not fully support this hypothesis, the

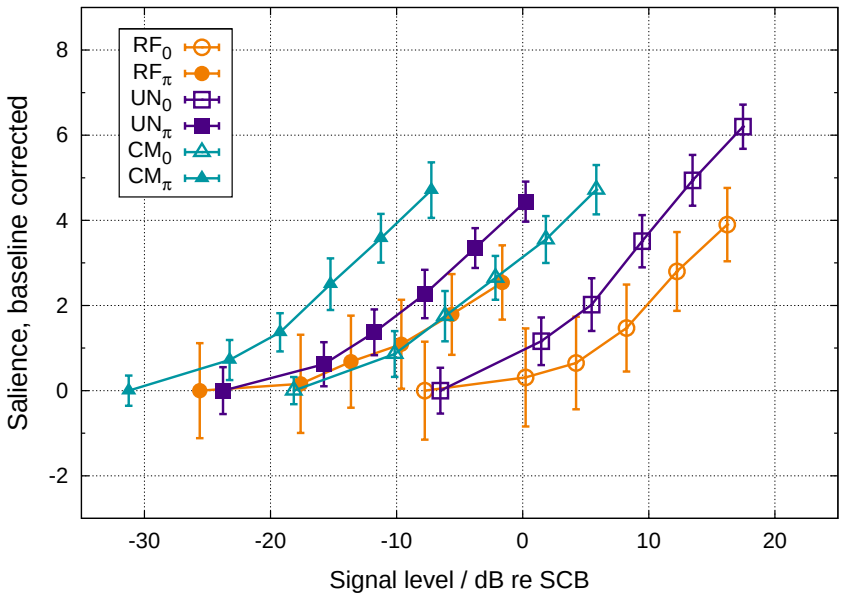

Figure 3.7: Average salience rating over all subjects, baseline corrected. The arithmetic mean over all subjects is plotted as a function of signal level relative to the level of the SCB for the RF (circles), UN (squares), and CM (triangles) condition, with a signal IPD of 0° (open symbols) or 180° (filled symbols). The magnitude for the obtained salience ratings were baseline corrected, i.e., ratings are displayed relative to the salience obtained at the lowest signal level in the corresponding condition, for which salience ratings were measured. Error bars indicate ± 1 standard error.

data show evidence that suprathreshold perception of a masked sinusoid should be assessed as a function of signal level above masked threshold in dB SL rather than physical levels in dB SPL. Mean data showed that salience increased with increasing signal level in each condition. Similar salience functions were found within both CM conditions (CM_0 and CM_π) and within both RF conditions (RF_0 and RF_π), whereas UN conditions as well as other variations of UN, CM, and RF were significantly different. However, analysing this difference in salience rating more carefully provides possible explanations for the observed discrepancies in the data. For the lowest signal level tested, a salience greater zero was observed, even though the sinusoidal signal should not have been audible in the masking noise as it was presented at -8 dB SL. Since the subjects rated a salient sinusoidal signal in the perceived stimuli, something else must have provoked this sensation. It is suggested that the masker itself contains a certain salience (i.e., an intrinsic salience) which could not be distinguished from the salience of the signal. This is supported by the fact that, for the level below threshold within a masker type, the salience was found to be approximately equal in the diotic and the corresponding dichotic condition. This suggests that the obtained rating resulted from a masker-specific property. Furthermore, for the lowest signal level, the highest salience rating was found for the RF masker. This can be related to

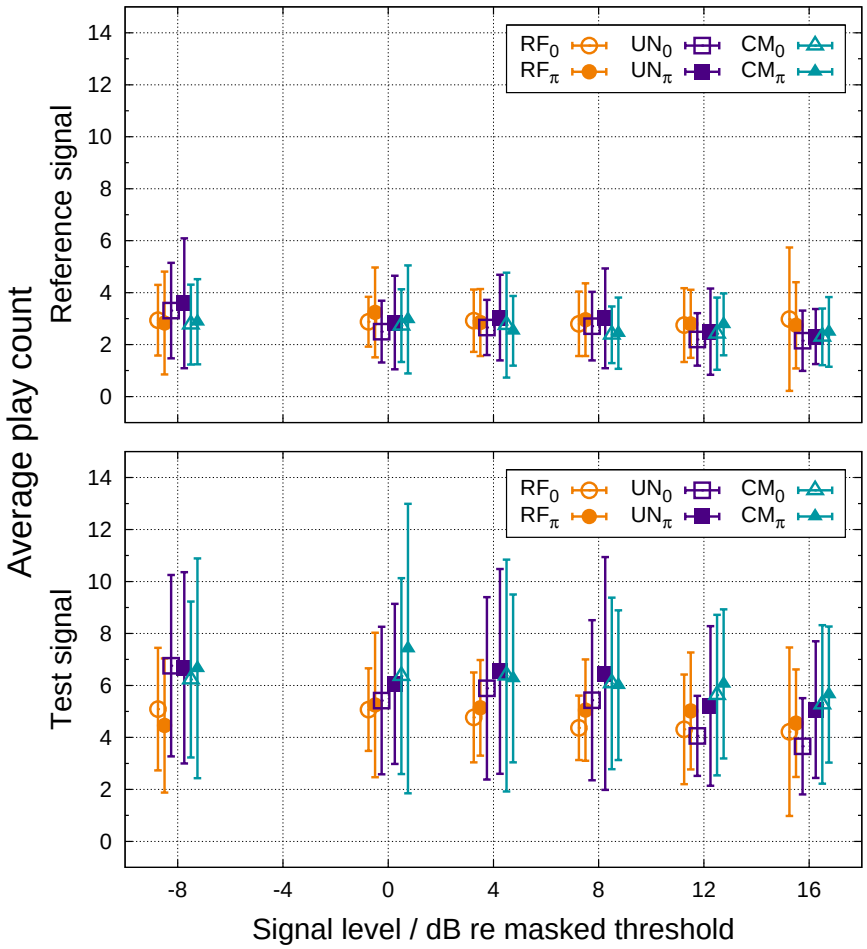

Figure 3.8: Average play count over all subjects. The arithmetic mean over all subjects of the number of times a subject listened to the reference signal (upper panel) and the test signal (lower panel) is plotted as a function of signal level relative to masked threshold for the RF (circles), UN (squares), and CM (triangles) condition, with a signal IPD of 0° (open symbols) or 180° (filled symbols). Data points have been slightly shifted horizontally to increase readability. Error bars indicate ± 1 standard deviation.

the fact that the single-band masker is typically perceived as being more tonal than multiband maskers due to the absence of flanking bands at distant frequencies. Moreover, the high inter-subject variability in the RF conditions observed for all tested levels suggests that subjects perceived the salience of a signal embedded in RF noise in different ways. Individual data suggests that it was more difficult in the RF than UN and CM conditions to differentiate the salience of the masked signal from the intrinsic salience of the masker.

In order to compensate for the intrinsic salience of the masker, the relative salience was considered (see figure 3.5). UN_π, CM_0, and CM_π were found to be dependent on the signal level relative to masked threshold in a similar way, whereas UN_0 was significantly different. The findings that UN_π, CM_0, and CM_π provoked similar salience ratings although their absolute thresholds considerably

differed (approximately 13 dB average BMLD was observed for multiplied noise in experiment I), supports the hypothesis of a correlation between the perceptual measure of salience and the audibility of the masked signal. In contrast, UN_0 showed a substantially steeper increase in salience with increasing signal level. This might be related to an effect of perceived intensity since absolute signal levels were highest for that condition (see figure 3.7). This is supported by the data for RF_0 (which had a high masked threshold similar to UN_0), which started to grow faster for levels above 4 dB SL compared to RF_π. In order to support this hypothesis, a loudness model was used to predict the increase in partial loudness of the masked sinusoidal signal with increasing signal level. An implementation by Hermes (2009) of the loudness model of Moore *et al.* (1997) was adapted to estimate the partial specific-loudness of the masked sinusoidal signal directly from the power spectra of the stimuli presented in experiment II. Loudness models generally calculate their predictions based on the frequency representation of the input sounds (Timoney *et al.*, 2004). Since the model of Moore *et al.* (1997) does not account for comodulation in the masker or interaural disparities within the signal, only UN_0 and RF_0 were modelled. A procedure was programmed in which the model simulated a subject. The partial specific-loudness of the masked sinusoidal signal was predicted for 10 (virtual) subjects with 12 repetitions each for both tested conditions. This was done in order to account for the small differences between spectra that occur in different realisations of the masker due to the non-deterministic way in which they are generated (see section 2.1.1). Figure 3.9 shows the predicted partial specific-loudness of the masked signal (solid lines) plotted as a function of signal level relative to masked threshold for UN_0 (squares) and RF_0 (circles), arithmetically averaged over all (virtual) subjects. In the lower panel, partial specific-loudness is illustrated in comparison with the average, relative salience ratings determined from the listeners (dashed lines, as in figure 3.5). Model predictions show an increased partial specific-loudness with increasing signal level. For UN_0, the partial loudness increased from 0 up to 5.9 sone while for RF_0 a partial loudness of 7.6 sone was reached for the highest tested signal level. In direct comparison with the salience ratings (lower panel), it can be seen that the slopes of the increasing loudness functions were steeper than the gain in salience for all tested conditions. Loudness was smaller than salience for low signal levels and started to grow faster for increasing levels whereas salience increased more linearly. This supports the hypothesis that a perceived intensity effect had influence on the salience rating of the UN_0 condition due to its higher absolute signal levels. However, comparing the baseline corrected salience as a function of absolute signal level (figure 3.7) indicates that, except for UN_0, the physical level does not predict the magnitude in relative salience.

The fact that subjects tended to listen more often to the dichotic signals compared to the corresponding diotic signals suggests that it was more difficult to rate the salience of a dichotic than a diotic masked signal. This might be related to the presentation of the reference signal. Since the reference signal was presented interaurally in phase (0° IPD), the reference signal and a diotic signal elicited the same sensation of a tone and could be compared directly. In contrast, the dichotic signals were perceived differently and it was, perhaps, more difficult to set its perception into relation to the reference signal. Nevertheless, as the results of the salience rating indicated (e.g., compare CM_0 and CM_π), the perception of diotic and dichotic signals was judged in a similar way.

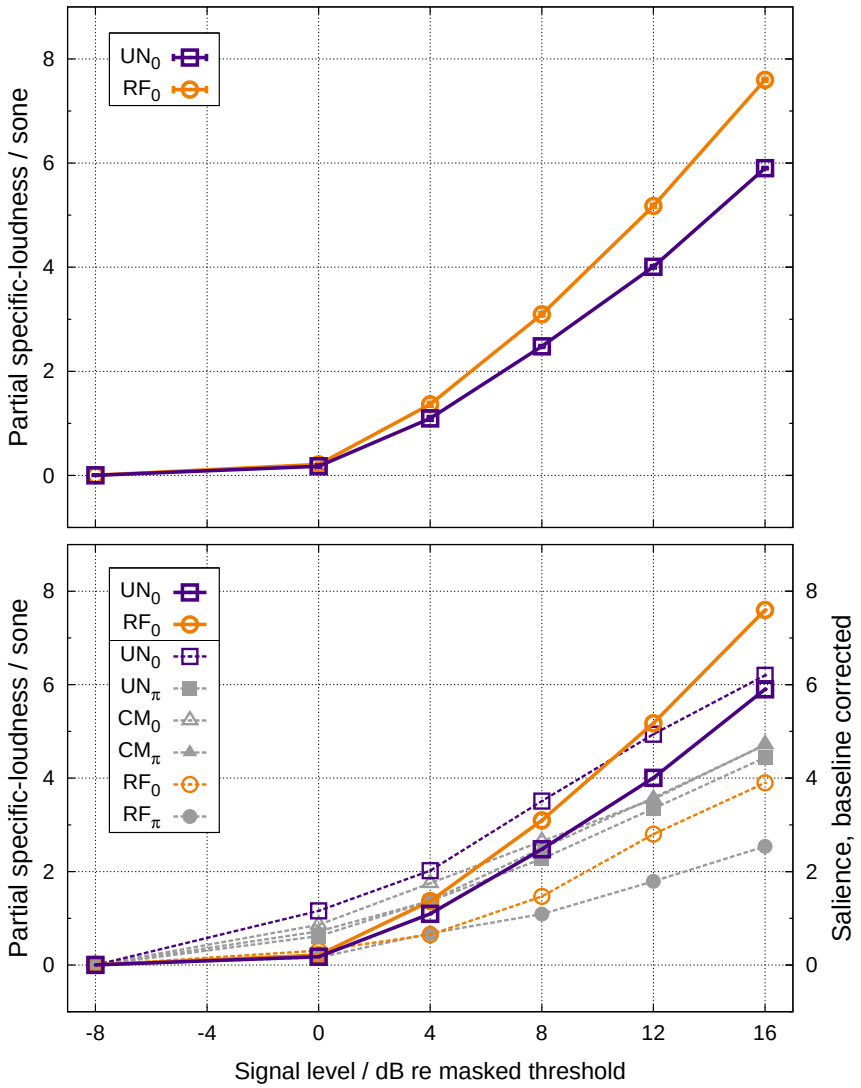

Figure 3.9: Partial specific-loudness (solid lines, both panels), predicted by a loudness model according to Moore *et al.* (1997). Average specific loudness of the masked sinusoidal signal is plotted as a function of signal level relative to masked threshold, for UN_0 (squares) and RF_0 (circles). Error bars indicate $\pm$ 1 standard error (only upper panel, lie within the marker symbols due to a negligible magnitude). The lower panel additionally shows the average, relative salience rating (dashed lines). The arithmetic mean of salience over all subjects is plotted as a function of signal level relative to masked threshold for the RF (circles), UN (squares), and CM (triangles) condition, with a signal IPD of 0° (open symbols) or 180° (filled symbols). The magnitude for the obtained salience ratings were baseline corrected, i.e., ratings are displayed relative to the salience obtained at the lowest signal level in the corresponding condition, for which salience ratings were measured.

Part II

Auditory evoked potentials

Chapter 4

Experiment III: Correlation between signal levels above masked threshold and auditory evoked potentials

The third experiment attempted to find a physiological correlate to the derived perceptual measure of salience in experiment II. Therefore, it was investigated whether the effect of masking release is also observed in objective measures. *Auditory evoked potentials* (AEPs) were recorded in order to assess suprathreshold perception of masked sinusoidal signals in different conditions of masking release. Based on the idea that the internal representation of the masked signal becomes clearer with increasing signal intensity, it was studied whether AEPs were sensitive to changes in signal level (above masked threshold) and thus, can be related to the audibility of the signal.

This was a preliminary study, in which AEPs of four listeners were recorded.

4.1 Introduction

Electrical activity of the human brain can be measured using EEG by placing electrodes on the scalp. An *event-related potential* (ERP) is the neural response elicited by an event (e.g., a sensory stimulus or a mental thought). The resulting electrical potentials can be recorded from the scalp (see Luck, 2005b for a discussion of the origin of the electrical activity in neurons).

The brain response to an auditory stimulus is called AEP. AEPs are generally categorised by their latency, which refers to the time between the stimulus onset and the onset of the neural response. AEPs are normally differentiated into early,

fast responses (up to around 10 ms), referred to as *auditory brainstem responses* (ABRs), *middle-latency responses* (MLRs, 10-50 ms), and *late auditory evoked potentials* (LAEPs, 50-1000 ms). ABRs reflect peripheral and brainstem activity, MLRs represent the activity from the midbrain up to the primary auditory cortex, and LAEPs reflect the activity from the auditory cortex. The same stimulus can elicit responses at all latencies (Picton, 2011a).

Figure 4.1 illustrates the classification of the AEPs with the help of an averaged ERP waveform. The waveform consists of a sequence of voltage deflections with positive and negative polarity, also referred to as components (Luck, 2005b). Those peaks and notches are traditionally labelled by P and N, where the accompanying number indicates the position within the sequence (letters refer to subcomponents of a main P or N wave). ABRs describe early peaks with the same polarity, numbering here is sufficient. Since voltage reflects the difference in electrical potential between two points, AEPs are always recorded between two electrodes. Thus, the polarity depends on the definition of the recording electrodes. The polarity is usually indicated when plotting ERP waveforms (Picton, 2011a).

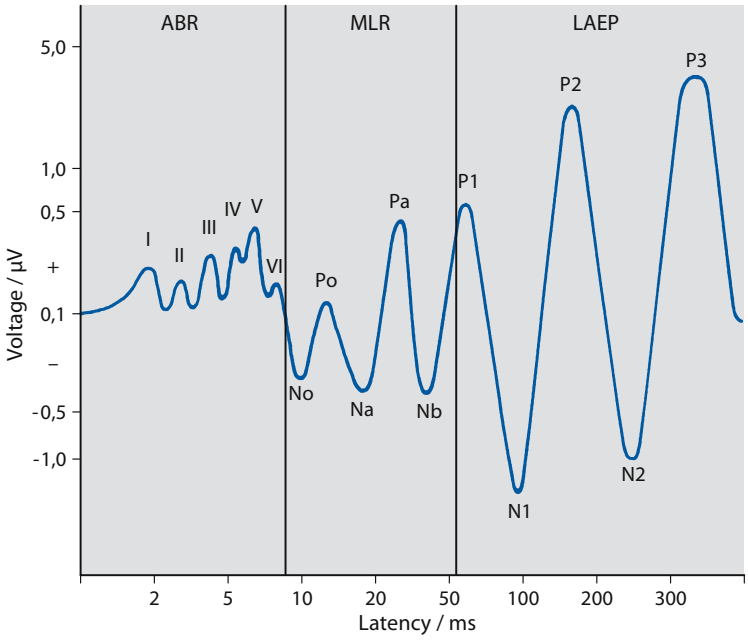

Figure 4.1: Classification of AEPs, figure 2.1a from Maurer *et al.* (2005) (modified). The amplitude of AEPs plotted as a function of latency using double-logarithmic axes. Positive is plotted upwards. This representation illustrates the division into auditory brainstem response (ABR), middle-latency response (MLR), and late auditory evoked potentials (LAEP).

During EEG measurements the electrical activity of the human brain is recorded. Neural oscillations can be observed at any time. For instance, head and eye movements elicit electrical activity visible in the EEG recordings. Neural activity can

also be observed near 10 Hz, originating from the cortex being in an idle state, which reflects the rhythmic and synchronous firing of neurons. This type of brain activity, which can be detected in the frequency range of 8 to 13 Hz, is called alpha waves. The neurons tend to fire more often, and out of sync, when the cortex processes information. In order to get a clear response to a presented stimulus (ERP) and be able to differentiate the ERPs from the background EEG, it is necessary to average ERP waveforms over a sufficient number of trials. This increases the signal-to-noise ratio of the recording (Picton, 2011a).

LAEPs reflect whenever there is an acoustical change. Apart from being elicited by a brief sound, like a short tone or click, presented in silence, LAEPs also occur when an ongoing stimulus changes (Picton, 2011b). N1/P2 potentials can be observed when an ongoing tone varies in intensity or frequency. The N1 wave increases in amplitude and decreases in latency with increasing intensity level (e.g. Näätänen and Picton, 1987; Neukirch *et al.*, 2002). Based on the assumption that the N1/P2 complex is related to the detection and recognition of the sound (i.e. how easy a stimulus is perceived), Androulidakis and Jones (2006) investigated N1/P2 potentials as a possible physiological correlate to CMR. Inspired by the studies of Androulidakis and Jones (2006) and Epp *et al.* (2012), the present experiment focused on recording LAEPs.

To measure LAEPs, it is important that the subjects are awake. LAEPs change significantly when a subject falls asleep (Picton, 2011b). The complex state of sleep is categorised in periods with and without rapid-eye-movements (REM), where the latter can be subdivided into further stages. As sleep goes on in cycles of periods of REM and non-REM, transitions of different brain activity occur during sleep (Picton, 2011b). These brain waves influence the LAEPs, since they lie within the same frequency range.

4.2 Method

4.2.1 Stimuli and apparatus

Stimuli were similar to experiment I and II, but had to be adjusted slightly. Stimuli consisted of a 300 ms sinusoidal signal with a frequency of 700 Hz that was embedded in the last 300 ms of a 1000 ms multiplied-noise masker, and both sinusoid and noise had 20 ms raised-cosine on- and offset ramps. Stimuli duration had to be increased since both the signal and the masker onsets elicit a neural response. In order to obtain clean LEAPs following the signal onset (i.e., with no interference of the response to the masker onset), the duration of the masking noise was doubled and the window, where the signal was present, was increased

by 50 ms and shifted towards the end of the masker presentation. On- and offset ramps were reduced to 20 ms due to the variation of amplitude and latency of LEAPs with stimulus rise time. For instance, the latency of the N1 potential increases with increasing rise time (Picton, 2011b). Picton (2011b) suggested optimal rise- (and fall-) times of 10 to 20 ms for recording N1/P2 waves.

The signal was presented either with 0 or 180° IPD, whereas the masker was always presented diotically. The masker consisted of five narrowband noise bands with centre frequencies at 700 Hz (SCB) and at 300, 400, 1000, and 1100 Hz (flanking bands). Each band had a bandwidth of 24 Hz and a level of 60 dB SPL. The noise bands had either uncorrelated (UN condition) or comodulated (CM condition) intensity fluctuations across frequency. The RF condition with only the SCB present was not considered in this preliminary design. The study aimed to find a physiological correlate of the perceptual measure of salience in the AEPs. Since the results of the salience rating in experiment II showed high inter-subject variability in the RF condition, it was decided to record LAEPs only for UN and CM conditions.

Similar to experiment II, LAEPs were measured at fixed levels of the masked sinusoidal signal in dB SL. The signal level was individually adjusted, based on the masked thresholds for the UN and CM conditions obtained in experiment I. The sinusoidal signal was presented at levels of 0, 8, and 16 dB relative to masked threshold (dB SL). Stimuli were presented in random order, evenly distributed over 400 sweeps (i.e., 400 repetitions per stimulus condition). Stimuli presentations were separated by an inter-stimulus interval of 500 ms.

Two different computers were used to playback the stimuli and record the AEPs. All stimuli were generated digitally in MATLAB with a sampling rate of 44100 Hz and 16 bit resolution using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner *et al.*, 2007), converted from digital to analog (RME Fireface UCX) and presented via two ER-2 insert earphones. The earphones had been calibrated and equalised at the centre frequencies of the five masking noise bands. BIOSEMI's ActiveTwo measurement setup, with active Ag-AgCl electrodes, was used to record AEPs. For active electrodes, the first amplifier stage is integrated in the electrode. As a result, the active electrode system provides low-noise measurements without any skin preparation required (BioSemi, 2012). Subjects wore an elastic cap with plastic, electrode holders (see figure 4.2). The electrodes were placed and named according to the 10/20 layout (standardised by the American Electroencephalographic Society; Niedermeyer and Da Silva, 2005) at approximately equal distances over the scalp. Seven electrode sites were used: Cz (at the top of the head, i.e., the vertex, halfway between nasion and inion), T7/T8 (sites at the temporal bone just above the left/right ear), P9/P10 (the

left/right mastoid), Oz (occipital site just above the inion) and Iz (the inion at the back of the head). The Cz electrode was used as the common reference. Before placing the electrodes, the electrode holders were filled with highly conductive, Signa electrode gel.

Figure 4.2: A 64 channel BIOSEMI headcap was used together with active electrodes to record AEPs. Figure from `http://www.biosemi.com/pics/apply_cap3_large.jpg` .

The electrodes were connected to the ActiveTwo AD-box, which amplified and performed A/D conversion of the measured potentials with a sampling rate of 1024 Hz and 16 bit resolution. The output was connected via optical fibre cable to the USB2 receiver box, which, in turn, was connected to the recording PC. The MATLAB procedure generating and playing the stimuli controlled the timing of the stimulus presentation and sent trigger signals corresponding to the stimulus condition to the receiver box of the recording system through the parallel port. Potentials were recorded with the data acquisition software ActiView (version 6.05), which streamed the continuous EEG to disk in `.BDF` file format. An anti-aliasing lowpass filter with a cut-off frequency of 200 Hz and a highpass filter to reduce the influence of slow, non-neural potentials (e.g. skin potentials; Luck, 2005a) with a cut-off frequency of 0.16 Hz were used on-line, during the recording. Filtering was digital and included in the BIOSEMI setup.

4.2.2 Listeners and procedure

Four of the listeners who had performed in the previous psychoacoustic experiments participated in the AEP experiment: listeners AC, KE, KG, and SG. Listeners were seated in a double-walled, sound-attenuating booth and watched a film with subtitles. Listeners were asked to relax but not fall asleep and to avoid moving during the recordings. The experiment was split up into four blocks of approximately 35 minutes each. Listeners participated in two sessions. In each session, two blocks of stimuli were recorded and were separated by a short break.

The procedure that generated and played the stimuli was implemented in MATLAB.

4.2.3 Post-processing and data analysis

In order to extract the AEPs from the overall EEG and improve the signal-to-noise ratio of the recorded potentials, filtering and averaging are usually applied to the recorded signals (Luck, 2005b). The post-processing was done in MATLAB adapting the scripts `avco3c` and `loadeeg3` (developed by Helmut Riedel, 2009). The signals were digitally filtered with a zero-phase, forward-backward butterworth lowpass filter with a cut-off frequency of 20 Hz. The low cut-off frequency was chosen since the LEAPs lie within this frequency range and faster responses were not of interest. The time interval, which was considered for further processing of the sweeps, was 1500 ms: a 250 ms pre-stimulus and a 1250 ms post-stimulus period. Linear detrending was applied in order to remove any slow drifts of potential. Signals were baseline corrected by subtracting the arithmetic mean over the 250 ms pre-stimulus (silent) period from the waveforms. A linear weighted averaging method was used to average over the sweeps (Riedel *et al.*, 2001). The averaging was applied in two steps. First, two of the four recorded blocks were averaged at a time to get two sub-averages. One sub-average was computed over the first and third block. The other was completed over the second and fourth block. The complete average was then calculated over the sub-averages. Artefact rejection of $\pm$ 100 μV was applied (i.e., sweeps with a higher peak-to-peak voltage than 200 μV were excluded from the average).

Peak amplitudes and latencies for N1 and P2 potentials were extracted for each subject and for each electrode channel separately. A graphical user interface, `scanpeaks1`, that was based on an automatic peak-scan-algorithm, `findpeaks` (both developed by Helmut Riedel, 2000), was used for the analysis. In order to obtain the N1/P2 response elicited by the signal onset, the time window from 700 to 1100 ms was extracted from the averaged sweeps. Baseline correction was applied considering a 100 ms pre-stimulus period (with respect to the onset of the sinusoidal signal at 700 ms). Only the electrodes P9/P10 and Iz (with Cz as reference) were considered further. A first, visual inspection of the plotted AEP data had shown that N1/P2 amplitudes were usually highest in these channels. N1 was measured between 70 and 190 ms and P2 from 180 to 300 ms (with respect to the signal onset), provided that P2 occurred later in time. First, the peak-scan-algorithm was run, which determined N1 and P2 amplitudes and latencies within the given time period. The resulting N1/P2 data were then double-checked by hand. If the polarity of a detected potential was reversed (N1 was assumed to

have negative voltage amplitudes whereas P2 had to be positive), the amplitude and latency values for that peak were discarded and assigned with NaN (i.e., not a number). In this case, and for N1/P2, where no peak amplitude and latency was automatically found, the lowest local minimum (regarding N1) or the highest local maximum (regarding P2) within the corresponding time interval, and with the right polarity, was taken as peak. For some conditions, no valid N1 or P2 could be found, which were further treated as NaN. Figure 4.3 shows the graphical user interface, which was used in order to extract the peak amplitudes and latencies.

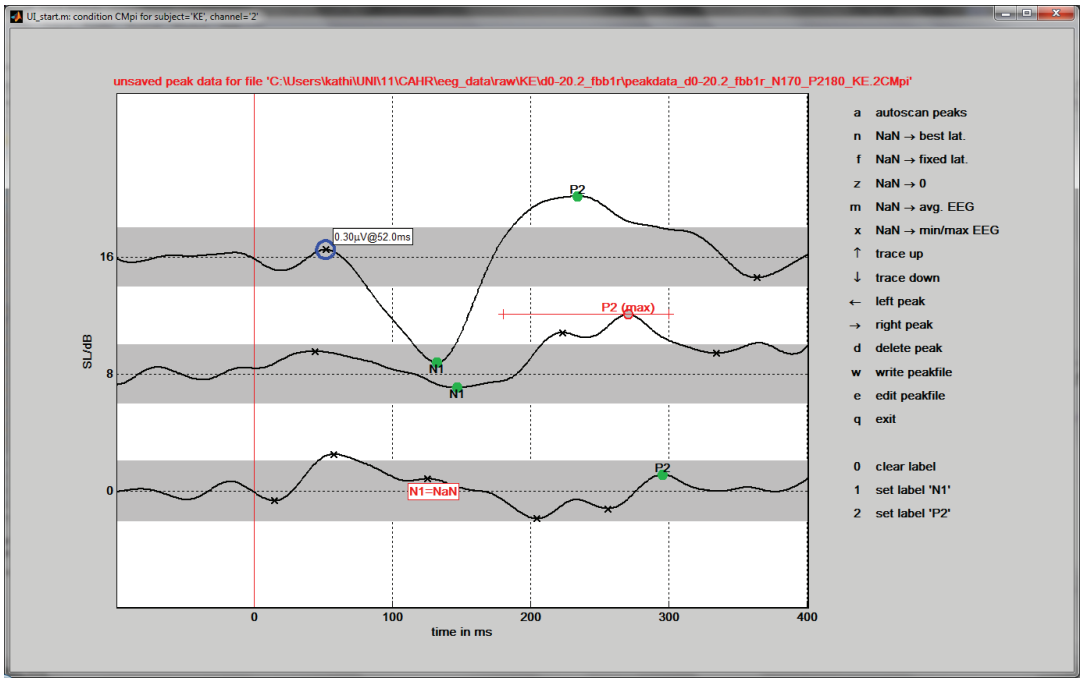

Figure 4.3: Screenshot of the graphical user interface `scanpeaks1`, which was used to extract peak amplitudes and latencies for N1 and P2 potentials, for each subject and for each electrode channel separately. The time interval of 600 to 1100 ms (0 ms on the x-axis corresponded to the sinusoidal signal onset at 700 ms) of the AEP waveforms (dichotic CM condition, electrode P9, subject KE) was plotted for three different levels of the masked sinusoidal signal (0, 8, and 16 dB SL). Green dots indicated N1 and P2 detected by the automatic peak-scan-algorithm. If the algorithm did not find an applicable N1 or P2 component, the lowest local minimum (regarding N1) or highest local maximum (regarding P2) within the corresponding time interval and with the right polarity was assigned to the missing peak (indicated by a red, grey-filled dot). The N1/P2, for which no valid amplitude and latency could be found, were treated as NaN.

AEP waveforms were plotted for the time interval of 600 to 1100 ms (0 ms on the x-axis corresponded to the sinusoidal signal onset at 700 ms) for each condition, electrode channel and subject separately. Automatically detected N1 and P2 potentials were indicated with green dots. N1/P2 which were defined by a local extrema were illustrated by red, grey-filled dots with a horizontal error bar in red representing the time interval where the corresponding potential was measured. NaN indicated that no valid amplitude and latency could be assigned

to the corresponding potential.

For every subject, the average peak amplitude of N1 and P2 potentials was determined as the arithmetic mean over the three considered channels, calculated after removing the NaN values. The grand mean of N1, P2, and the difference between P2 and N1 represents the arithmetic mean over all subjects.

4.3 Results

Recorded AEPs for subject AC, averaged over 400 sweeps are shown in figure 4.4. Waveforms are plotted for the latency interval of -250 (i.e., pre-stimulus) to 1250 ms, in which 0 ms corresponds to the onset of the 1000 ms masking noise and the 300 ms sinusoidal signal was presented between 700 and 1000 ms. The left and right panel illustrate the obtained potentials for the UN and CM condition, respectively, with a signal IPD of 0° (solid lines) or 180° (dashed lines). Waveforms are plotted for three different levels of the masked sinusoidal signal: 0, 8, and 16 dB relative to masked threshold (indicated on the y-axes on the left-hand side). The potentials were baseline corrected by subtracting the arithmetic mean over the 250 ms pre-stimulus period from the waveforms.

Individual data (see also appendix C) showed clear LEAPs due to the masker onset in all conditions for all measured signal levels. All waveforms show similar voltage deflections in the time interval from 0 to 300 ms following the masker onset for each subject. For subject AC, a clear P1/N1/P2 (peak-notch-peak) response was observed. With increasing signal level, LEAPs elicited by the signal onset were found. While, for a signal presented at masked threshold, no specific response in the time interval between 700 and 1000 ms was obtained, a clear N1/P2 complex was observed for a signal level of 16 dB above masked threshold in all conditions. Especially for subjects KE and SG, dichotic signals (UN_π and CM_π) led to stronger LEAPs than diotic signals (UN_0 and CM_0).

The grand mean AEP waveforms, averaged over all four subject, can be seen in figure 4.5. Waveforms are plotted in the latency interval relevant for showing LEAPs elicited by the signal onset (600 to 1100 ms). As before, potentials are illustrated for the UN (left panel) and CM (right panel) condition, with solid lines representing diotic signals and filled symbols dichotic signals. Waveforms are plotted for three different levels of the masked sinusoidal signal: 0, 8, and 16 dB relative to masked threshold. The potentials were baseline corrected using the 100 ms pre-stimulus period (with respect to the onset of the sinusoidal signal at 700 ms). Mean results show increasing N1/P2 potentials for higher signal levels for both UN and CM masking conditions. Data for signal levels above masked threshold (8 and 16 dB SL) illustrate that stronger responses were found in the

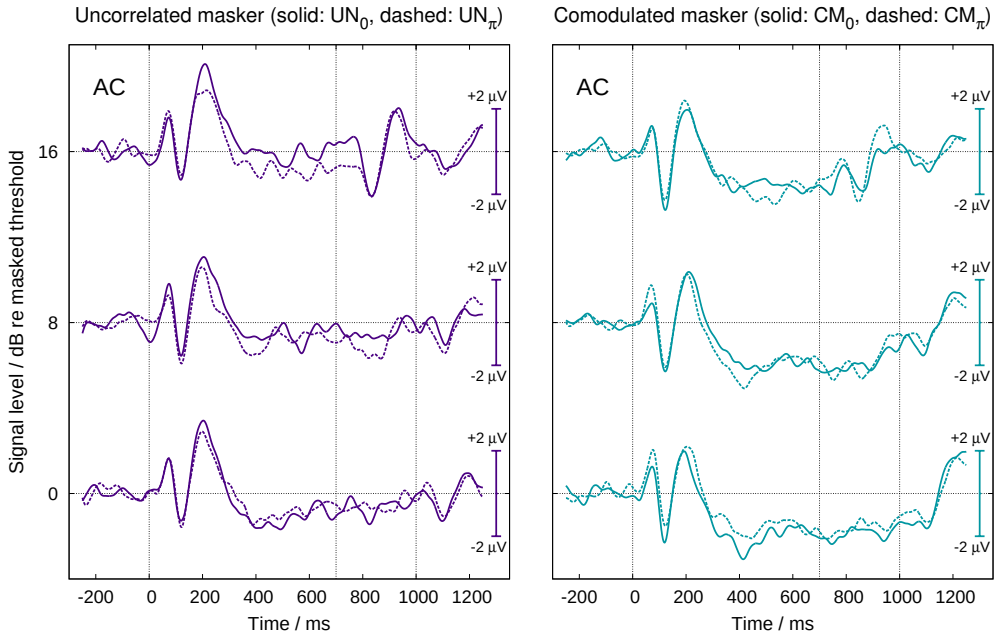

Figure 4.4: Individual AEP data for subject AC, averaged over 400 sweeps. Waveforms are shown for the latency interval of -250 (i.e. pre-stimulus) to 1250 ms. The left and right panel illustrate the obtained potentials for the UN and CM conditions, respectively, with a signal IPD of 0° (solid lines) or 180° (dashed lines). Waveforms are plotted for three different levels of the masked sinusoidal signal: 0, 8, and 16 dB relative to masked threshold (indicated on the y-axes on the left-hand side). The potentials were baseline corrected by subtracting the arithmetic mean over the 250 ms pre-stimulus period from the waveforms. Vertical, dashed lines indicate the masker onset at 0 ms, the signal onset at 700 ms and the stimulus end at 1000 ms.

dichotic conditions (UN_π and CM_π) than in the corresponding diotic conditions (UN_0 and CM_0).

Peaks in the grand mean waveforms might appear to be smaller than those in the individual waveforms for a single subject. This is perfectly reasonable since the latencies of the peak components differed between subjects. For some subjects, voltages were positive at the same time where other subjects showed negative deflections. This results in a grand mean that is somewhat smaller than most of the individual data (Luck, 2005b). Therefore, peak amplitudes and latencies for N1 and P2 potentials were extracted for each subject individually.

Grand mean amplitudes for N1, P2, and the difference P2-N1 are shown in figure 4.6. The amplitudes N1 (left panel), P2 (middle panel), and P2-N1 (right panel) are plotted as a function of signal level relative to masked thresholds in the UN (squares) and CM (triangles) conditions. Open symbols represent the results for the diotic signal, whereas filled symbols show the results for the dichotic signal.

N1 amplitudes for the dichotic conditions decreased with increasing signal level in a similar way, from approximately -0.34 to 2.2 μV for UN_π and from approximately -0.31 to -2.34 μV for CM_π over the measured range of signal levels

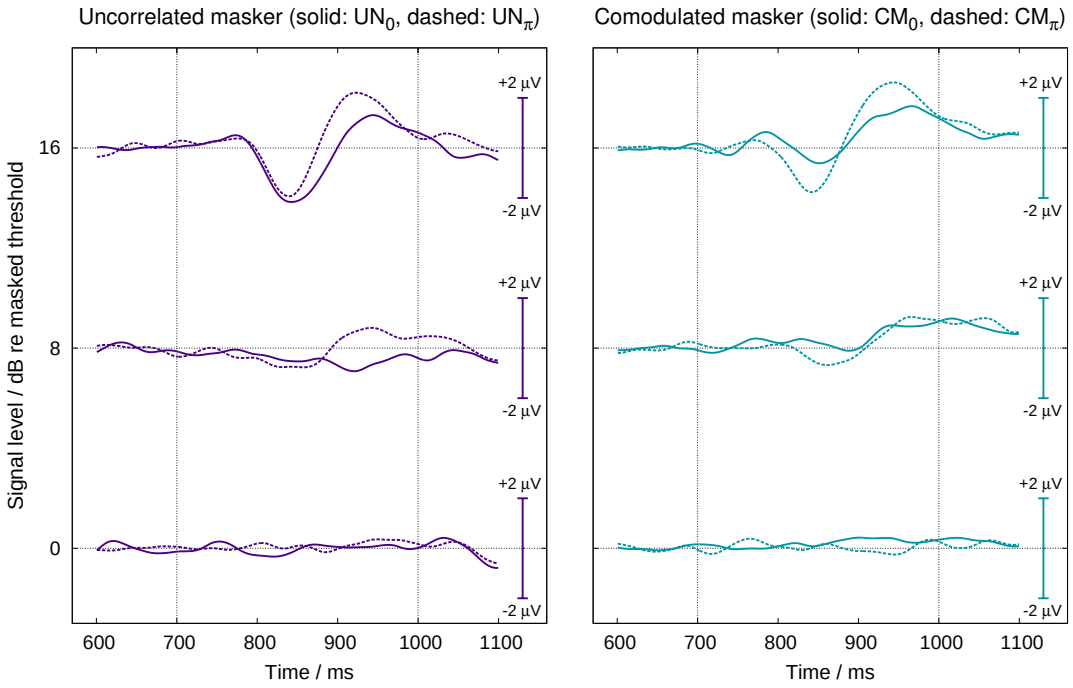

Figure 4.5: Grand mean AEP waveforms for all subjects. Waveforms are shown for the latency interval of 600 to 1100 ms, with an amplitude scale of ± 2 μV. The left and right panel illustrate the obtained potentials for the UN and CM conditions, respectively, with a signal IPD of 0° (solid lines) or 180° (dashed lines). Waveforms are plotted for three different levels of the masked sinusoidal signal: 0, 8 and 16 dB relative to masked threshold (indicated on the y-axes on the left-hand side). The potentials were baseline corrected considering a 100 ms pre-stimulus period (with respect to the onset of the sinusoidal signal at 700 ms). Vertical, dashed lines indicate the signal onset at 700 ms and the stimulus end at 1000 ms.

(0 to 16 dB SL). In contrast, N1 in the UN_0 condition increased slightly and then decreased to a minimum N1 of -2.77 μV. CM_0 showed the highest N1 amplitudes for signal levels above masked threshold with approximately -0.27 μV for 8 dB SL and -0.86 μV for 16 dB SL. P2 amplitudes increased with increasing signal level above masked thresholds for all conditions except UN_0, which showed the smallest amplitude (about 0.4 μV) for a signal level of 8 dB above threshold. Generally higher P2 amplitudes were found in dichotic conditions compared to the corresponding diotic conditions for levels above masked threshold. The highest P2 was observed in the CM_π condition and was approximately 2.81 μV. The difference between P2 and N1 amplitudes tended to be smaller for the dichotic conditions than the corresponding diotic conditions for levels above masked threshold. The highest P2-N1 amplitudes were observed in the CM_π condition, reaching approximately 5.15 μV at a signal level of 16 dB SL.

Another way to illustrate the grand mean amplitudes for N1, P2, and P2-N1 is to plot the amplitudes as a function of physical signal level instead of SLs. Therefore, the corresponding mean masked threshold (shown in figure 2.4, left

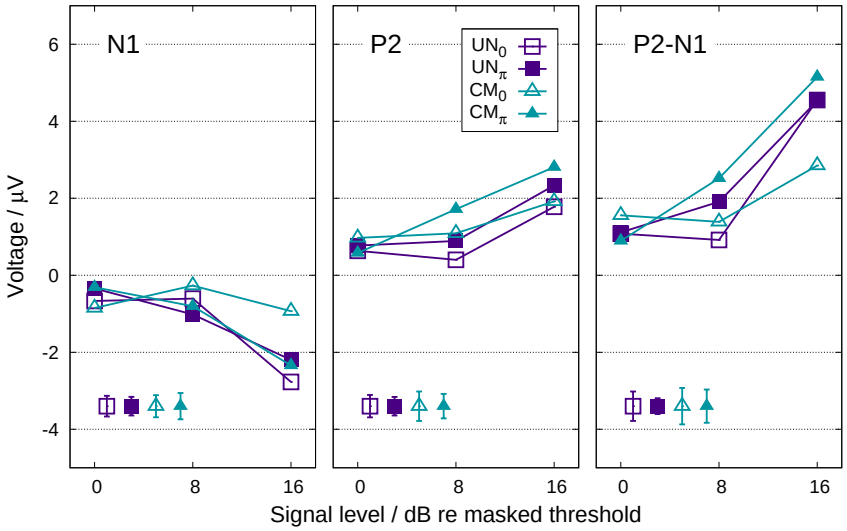

Figure 4.6: Grand mean amplitudes N1 (left), P2 (middle), and the difference between P2 and N1 (right) for all subjects. In each panel, the amplitudes are plotted as a function of signal level relative to masked threshold for the UN (squares) and CM (triangles) conditions. Open symbols represent the results for the diotic signal, whereas filled symbols show the results for the dichotic signal. Error bars in the lower, left corner indicate ± 1 standard error (averaged over all signal levels) for each condition, respectively.

panel) was added to the SLs in each condition, respectively. This results in a horizontal shift of the amplitude curves according to the different masked thresholds for the conditions. Since the masked thresholds differed among the tested stimuli conditions, actual physical levels measured in dB SPL varied across the different conditions. Figure 4.7 shows the grand mean amplitudes as a function of signal level relative to the level of the SCB of the masker. As in figure 4.6, N1 (left), P2 (middle), and P2-N1 (right) are plotted for the UN (squares) and CM (triangles) conditions where open symbols refer to a diotic and filled symbols to a dichotic signal presentation.

4.4 Discussion

The present study investigated the correlation between AEPs and the suprathreshold perception of a masked sinusoidal signal in different conditions of masking release. It was studied whether changes in signal level (above masked threshold) and thus, the audibility of the masked signal could be reflected by LAEPs. Data showed elevated N1/P2 potentials for higher SLs of the masked signal in all tested conditions.

Only a few studies have been conducted investigating the effect of masking release and its suprathreshold perception in AEPs. Androulidakis and Jones (2006) studied whether AEP were sensitive to the perceptual enhancement of a

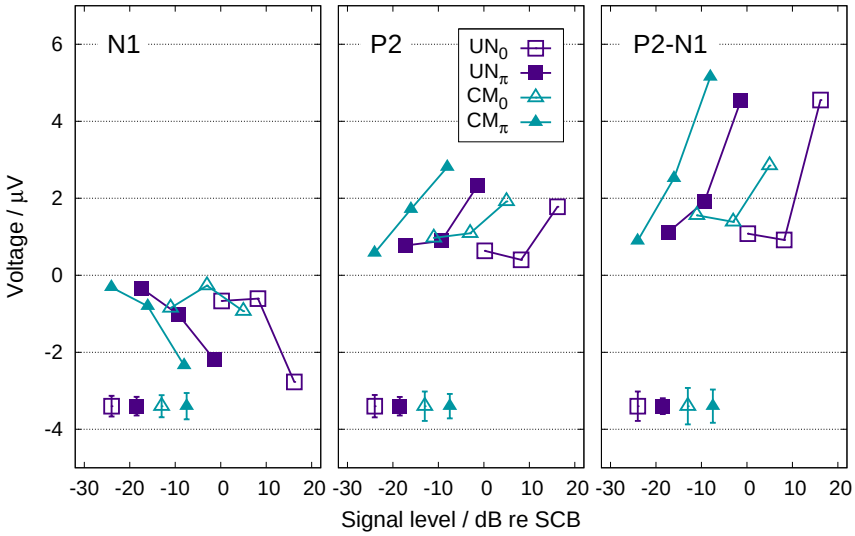

Figure 4.7: Grand mean amplitudes N1 (left), P2 (middle), and the difference between P2 and N1 (right) for all subjects. In each panel, the amplitudes are plotted as a function of signal level relative to the level of the SCB for the UN (squares) and CM (triangles) conditions. Open symbols represent the results for the diotic signal, whereas filled symbols show the results for the dichotic signal. Error bars in the lower, left corner indicate ± 1 standard error (averaged over all signal levels) for each condition, respectively.

masked sinusoidal signal in a CMR paradigm. The present study is not directly comparable since a completely different stimulus paradigm was used. However, Androulidakis and Jones (2006) accessed the audibility of a sinusoidal signal fixed in level by presenting it either in silence, or in band-limited masking noise, which was either uncorrelated or amplitude-modulated. They reported a large N1/P2 response when the signal was presented in silence and diminished N1/P2 amplitudes for a signal masked by modulated noise in which the signal was still audible. In contrast, N1/P2 completely vanished when the signal was presented in uncorrelated noise and was, therefore, inaudible. This is in qualitative agreement with the present results, which showed stronger N1/P2 for clearly audible signals at suprathreshold levels (8 and 16 dB SL) than at masked threshold (0 dB SL), where the masked signal was just perceivable. In order to directly relate the present results to the findings of Androulidakis and Jones (2006), the N1/P2 responses for the UN_0 and CM_0 conditions should be compared at equal signal levels in dB SPL. Since the masked thresholds for the UN and CM conditions differed and AEPs were recorded for equal signal levels relative to masked threshold, actual physical levels measured in dB SPL varied across the conditions. However, grand mean peak data as a function of signal level relative to the level of the SCB in figure 4.7 illustrate that the P2 amplitudes for CM_0 were clearly higher than for UN_0 when compared in the same range of physical signal level in dB SPL. This supports the findings of Androulidakis and Jones (2006).

The most recent study of Epp *et al.* (2012) investigated whether the effect of masking release was observed in AEPs. The stimuli used in the present experiment and in the study of Epp *et al.* (2012) were identical. While Epp *et al.* (2012) presented the stimuli at several fixed signal-to-masker ratios (equal physical signal levels in dB SPL for the different conditions), the present experiment used stimuli with fixed signal levels relative to masked threshold (equal SLs in dB SL). Figure 4.8 shows the grand mean data for the amplitudes N1, P2, and the difference between P2 and N1 of Epp *et al.* (unpublished). Three panels show the N1 (left), P2 (middle), and P2-N1 (right) amplitudes plotted for the UN (squares) and CM (triangles) conditions where grey symbols indicate diotic and coloured symbols indicate dichotic signals. Voltages are plotted as a function of signal level relative to the level of the SCB. Peak data was obtained for signal-to-masker ratios of -15, -5, +5 and +15 dB. Epp *et al.* (2012) found that N1 was sensitive to a change in audibility when an interaural disparity was introduced. P2 represented the increase in audibility when modulated noise was used instead of uncorrelated noise as well as when additionally an IPD was introduced to the sinusoidal signal. The difference P2-N1 varied for each condition and reached highest amplitudes for CM_{π}. Focusing on the P2 amplitudes, the data indicated that for equal physical levels of the masked signal, P2 varied across the tested conditions. More precisely, lowest P2 was observed for the UN_0 condition. When across-frequency information like comodulated masking noise (CM conditions) or an IPD to the signal (dichotic $_{\pi}$ conditions) were introduced (or a combination of both), P2 amplitudes increased. This increase correlated with the amount of masking release due to the comodulation and IPD cues (i.e., the lower the masked threshold, the clearer the signal was perceived and thus, the higher P2). Hence, P2 can be related to the audibility of a masked signal at a certain signal-to-masker ratio.

The present study compared peak amplitudes of LEAPs at equal levels above masked threshold, which implied that the signal was equally audible in all tested conditions. Applying the relation of P2 with the signal audibility, derived from the data of Epp *et al.* (unpublished), to the present experiment suggests that P2 would have been expected to be about the same for all conditions when compared at equal levels (in dB SL). Based on the data of Epp *et al.* (2012) and Epp *et al.* (unpublished), and under the hypothesis that the perceptual measure of salience, derived in experiment II, is related to the audibility of the masked signal, the P2 potential can then be correlated with this perceptual measure. Considering the idea of the internal representation of a signal, as introduced earlier in chapter 1, it can be concluded that P2 also reflects the strength of this internal representation (i.e., whether and how clear the signal is perceivable). However, the present

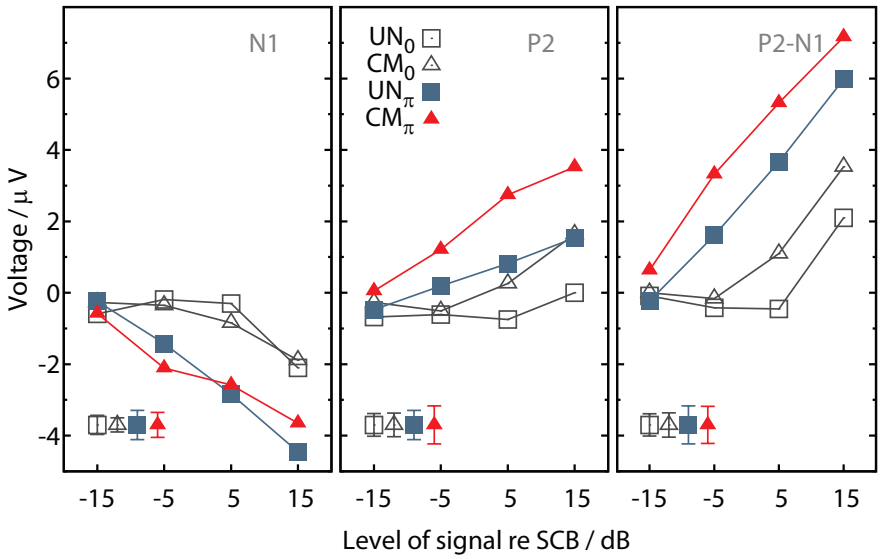

Figure 4.8: Grand mean (for 10 subjects) of N1 (left), P2 (middle) and the difference between P2 and N1 (right), figure of Epp *et al.* (unpublished). Each panel shows four curves, one for each condition. Amplitudes are plotted as a function of signal level relative to the level of the SCB. The uncorrelated and comodulated masker are indicated by triangles (CM condition) and squares (UN condition), for both diotic (grey symbols) and dichotic (filled, coloured symbols) conditions. Error bars in the lower, left corner indicate ± 1 standard error (averaged over all signal levels) for each condition, respectively.

grand mean results for P2 (figure 4.6) do not support this conclusion. P2 amplitudes increased with increasing level, though not in an entirely similar way for all conditions. However, several aspects should be considered while analysing these results. First, LEAPs of only four listeners were recorded. In general, ERP waveforms show high variability between subjects. The same stimulus can elicit N1 and P2 amplitudes which significantly differ among subjects (e.g. Luck, 2005b). Therefore, data from more subjects is required to draw strong conclusions. Further, the range of the presented signal levels was rather small (0 - 16 dB SL) and differed from that used by Epp *et al.* (2012) and Epp *et al.* (unpublished). Epp *et al.* (2012) and Epp *et al.* (unpublished) recorded potentials within a range of ± 15 dB of signal-to-masker ratios. If their data points are transformed to a dB SL scale, the amplitudes for the CM_π condition, which show the clearest response, were obtained well above threshold, whereas in other conditions, such as UN_0, data was collected near to threshold or even below. For low signal levels relative to masked threshold it is difficult to obtain reliable peak data since the LAEPs start to vanish within the noise floor. Thus, results depend considerably on the procedure of extracting the N1/P2 amplitudes for masked signals near threshold. This was observed in the present data for signals at threshold and 8 dB above. Valid and reliable peak data was difficult to obtain for those levels. LAEPs for more and higher signal levels should be recorded in order to obtain

more stable results. For those reasons, the idea of relating P2 to the audibility of a masked signal should not fully be discarded yet. In contrast to Epp *et al.* (2012) and Epp *et al.* (unpublished), the present study did not find an effect of interaural disparity on the peak amplitude of N1. Although the dichotic conditions (UN_π and CM_π) were about the same in magnitude, the diotic conditions (UN_0 and CM_0) differed substantially, especially for 16 dB SL.

Although the present results do not fully replicate the findings of Epp *et al.* (2012) and Epp *et al.* (unpublished), it was shown that LEAPs were sensitive to changes in signal level above masked threshold. A dependence on SL of a masked sinusoidal signal was observed in N1 and P2 amplitudes as well as the difference P2-N1. How exactly the N1/P2 components relate to the stimuli properties and its perception at suprathreshold levels could not been verified due to the small number of test subjects and the limited range of signal levels tested. However, the data indicate that most likely P2 is related to the audibility of the masked signal.

Part III

Conclusion

Chapter 5

General discussion

In Experiment I, thresholds of sinusoidal signals in the presence of complex maskers were measured. The combined effect of CMR and BMLD cues on signal detectability was investigated. A CMR experiment using a flanking-band paradigm was conducted in which the masked signal was presented with an IPD of either 0° or 180°. A release from masking was observed to varying degrees, under conditions where either one or both of these cues were present. A smaller CMR, for both RF-CM and UN-CM, was obtained in the N_0S_π conditions compared to N_0S_0, which is in qualitative agreement with Hall *et al.* (2006, 2011). In contrast to Epp and Verhey (2009a,b), it was concluded that CMR and BMLD were independent (i.e., additive to a certain extent).

In experiment II, suprathreshold perception was assessed by varying the signal level relative to the masked threshold for that condition. A perceptual measure of salience was defined, which describes the extent to which the masked sinusoidal signal can be segregated from its background. The salience of the masked sinusoid was measured at equal levels relative to masked threshold (in dB SL) for all conditions. Salience increased with increasing signal level in each condition, but also depended on the condition and the subject. The largest inter-subject variability was found in the RF conditions. This suggests that, generally, it was more difficult to rate the salience of the masked sinusoidal in the presence of the single-band masker than in the UN and CM conditions. The differences in perceived salience between the RF and the other two conditions (UN and CM) might have resulted from the different masker spectra. It is speculated that the masker itself was perceived differently in the single-band compared to the multi-band conditions. This is supported by the data for the lowest signal level that showed different magnitudes in salience among the different masker conditions. For that signal level, generally high salience ratings were obtained for the RF masker. This speculation is furthermore supported by the large standard error of the masked threshold in the RF_π condition, as found in experiment I. Thus, it

might be difficult to directly compare the RF condition and its perceptional aspects with UN and CM against each other because of their different spectra. This supports the argument of Epp and Verhey (2009b), who questioned whether the RF condition was the right reference condition when investigating the influence of binaural cues on CMR.

To investigate the possibility of a physiological correlate to the derived perceptual measure of salience, AEPs were recorded in experiment III. The results showed that LEAPs were sensitive to changes in signal level above masked threshold. Elevated N1/P2 potentials were found for increased signal levels. Based on the data of Epp *et al.* (2012) and Epp *et al.* (unpublished), it was hypothesised that the P2 amplitudes can be correlated to the perceptual measure. Figure 5.1 compares the average salience rating (upper panel) and the corresponding grand mean P2 amplitudes (lower panel) as a function of signal level relative to masked threshold, for the four listeners who participated in both experiments. UN conditions are illustrated with squares and CM conditions with triangles, where open symbols represent the results for the diotic signal and filled symbols show the results for the dichotic signal. Mean salience increased with increasing signal level in each condition. The diotic UN condition (UN_0) increased faster with increasing signal level than the other UN and CM conditions. It was suggested that the perceived loudness influenced the salience rating for UN_0 due to its higher absolute signal levels. Salience functions for UN_π, CM_0, and CM_π were fairly similar for signal levels above masked threshold. P2 increased with increasing signal level, though the rate at which amplitudes grew was different for each condition. Only weak correlation between the perceptual measure and the peak P2 data was found. However, the high variability for P2 amplitudes for a signal level of 16 dB above masked threshold suggests that data from more subjects are required to draw stronger conclusions regarding the correlation of a perceptional measure and LEAP recordings. In addition, LEAPs at more and higher signal levels should be recorded since valid and reliable peak data were difficult to obtain for signals at or near threshold.

Nevertheless, it was shown that both salience and P2 amplitudes were higher when the signal level was increased equally above threshold for all conditions. Thus, it can be concluded that cues that lead to a masking release at masked signal threshold also dominate the perception of masked signals at supratreshold levels and its neural representation in terms of AEPs. In contrast to Townsend and Goldstein (1972) and Soderquist and Shilling (1990), who found a diminishing benefit of BMLD for increasing suprathreshold levels, similar magnitudes in salience were found for diotic and dichotic conditions for signal levels above masked threshold (compare UN_π, CM_0, and CM_π). Further, LEAP data showed

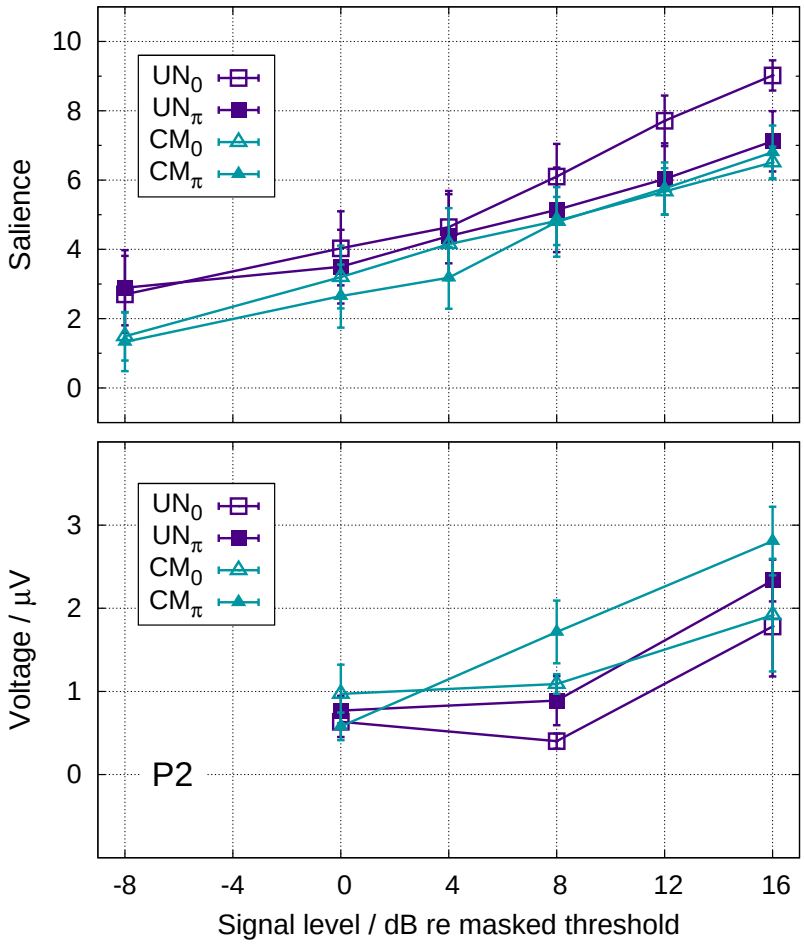

Figure 5.1: Average salience rating (upper panel) and grand mean amplitudes P2 (lower panel) for subjects AC, KE, KG, and SG as a function of signal level relative to masked threshold, for the UN (squares) and CM (triangles) conditions. Open symbols represent the results for the diotic signal, whereas filled symbols show the results for the dichotic signal. Error bars indicate ± 1 standard error.

elevated N1/P2 potentials in each condition for the highest tested signal level of 16 dB above masked threshold. Although previous studies had shown a reduced benefit of BMLD and CMR in suprathreshold tasks (e.g. Townsend and Goldstein, 1972; Hall and Grose, 1995; Verhey and Heise, 2012), the present study found a perceptual measure that is mainly driven by the signal level above masked threshold for several conditions of masking release. It is therefore concluded that cues leading to a release from masking contribute to the general perception of signals above masked threshold (i.e., how well the signal can be segregated from its background), even though the cues might not be beneficial in other, more specific suprathreshold tasks.

Chapter 6

Summary and outlook

The perception and neural representation of a masked sinusoidal signal was investigated as a function of the signal level relative to masked threshold. To obtain a metric that relates the physical properties of the stimulus to the strength of its internal representation and, consequently, the audibility of the signal presented in noise, masked thresholds were modified by introducing a systematic combination of different cues that resulted in a masking release.

Two cues were used to alter masked thresholds: (1) Synchronous amplitude envelope fluctuations across frequency of the masker yielding CMR, and (2) interaural disparities, more precisely an IPD, introduced to the signal resulting in a BMLD. In the first psychoacoustical experiment, thresholds of a masked sinusoidal signal were measured in various conditions.

Based on individuals' masked thresholds, suprathreshold perception was assessed by varying the signal level relative to masked threshold for that condition. A perceptual measure was found that described the salience of a masked sinusoidal tone (i.e., how well the signal can be segregated from its background). Under the hypothesis that the perceptual measure (which quantifies the audibility of a masked sinusoidal signal) is primarily determined by its level above masked threshold, the salience was measured at equal levels relative to masked threshold (in dB SL) for all conditions in the second psychoacoustical experiment.

With increasing signal level relative to masked threshold, an increased salience was found in each tested condition. However, the salience ratings were not fully independent of the masker condition. This discrepancy was attributed to differences in the so-called intrinsic salience of the masker and an effect of perceived loudness for conditions with highly masked thresholds. Furthermore, a significant main effect of the subjects was found. Since the term salience is not commonly used, subjects may have interpreted it differently. Nevertheless, it was concluded that suprathreshold perception of a masked sinusoidal signal is better modelled by the signal level relative to masked threshold (in dB SL) than the absolute

physical level (in dB SPL). However, salience might not be the ideal measure to assess the degree to which a masked sinusoidal signal can be segregated from its background. It is suggested to adapt the perceptual measure of salience to a new measure, which is specified in a more familiar and objective way.

Further, it was investigated whether the audibility of the masked sinusoidal signal can be reflected using objective measures. In the third experiment, AEPs were recorded to investigate the possibility of a physiological correlate of the derived perceptual measure of salience. It was shown that LAEPs, more precisely N1/P2 amplitudes, were sensitive to changes in signal level above masked threshold. However, more test subjects and a wider range of signal levels are required to better define the relation between the N1/P2 components and the stimuli properties and their suprathreshold perception. Hence, an extended experiment to record LAEPs should be conducted to clarify the possibility of a physiological correlate to the perceptual measure of salience. The correlate can then be related to the psychoacoustical phenomena of CMR and BMLD and potentially facilitates interpretations of the processing of sound along the auditory pathway. A possible combination with fMRI would provide further insight into the underlying structure of the neural processing of sound and could possibly demonstrate the location of the regions in the brain reflecting the perception of masked signals at suprathreshold levels.

Bibliography

Androulidakis, A. and Jones, S. (**2006**), "Detection of signals in modulated and unmodulated noise observed using auditory evoked potentials," Clinical Neurophysiology **117**(8), 1783–1793.

Arons, B. (**1992**), "A Review of The Cocktail Party Effect," Journal of the American Voice I/O Society **12**, 35–50.

Bech, S. and Zacharov, N. (**2006**), "Quantification of impression," in *Perceptual Audio Evaluation - Theory, Method and Application* (John Wiley & Sons Ltd), chap. 4.

Bernstein, L. R., Trahiotis, C., and Hyde, E. L. (**1998**), "Inter-individual differences in binaural detection of low-frequency or high-frequency tonal signals masked by narrow-band or broadband noise," The Journal of the Acoustical Society of America **103**(4), 2069–2078.

BioSemi (**2012**), "Active electrodes," URL `http://www.biosemi.com/active_electrode.htm`.

Brainard, D. H. (**1997**), "The Psychophysics Toolbox," Spatial Vision **10**, 433–436.

Branstetter, B. K. and Finneran, J. J. (**2008**), "Comodulation masking release in bottlenose dolphins (Tursiops truncatus)," The Journal of the Acoustical Society of America **124**(1), 625–633.

Bronkhorst, A. W. (**2000**), "The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions," Acta Acustica united with Acustica **86**(1), 117–128.

Buss, E. and Hall, J. W. (**2009**), "Effects of masker envelope coherence on intensity discrimination," The Journal of the Acoustical Society of America **126**(5), 2467–2478.

Buss, E., Hall, J. W., and Grose, J. H. (**2007**), "Individual differences in the masking level difference with a narrowband masker at 500 or 2000 Hz," The Journal of the Acoustical Society of America **121**(1), 411–419.

Buus, S. (**1985**), "Release from masking caused by envelope fluctuations," The Journal of the Acoustical Society of America **78**(6), 1958–1965.

Cherry, E. C. (**1953**), "Some Experiments on the Recognition of Speech, with One and with Two Ears," The Journal of the Acoustical Society of America **25**(5), 975–979.

Cohen, M. F. (**1991**), "Comodulation masking release over a three octave range," The Journal of the Acoustical Society of America **90**(3), 1381–1384.

Cohen, M. F. and Schubert, E. D. (**1991**), "Comodulation masking release and the masking-level difference," The Journal of the Acoustical Society of America **89**(6), 3007–3008.

Durlach, N. I. (**1963**), "Equalization and Cancellation Theory of Binaural Masking-Level Differences," The Journal of the Acoustical Society of America **35**(8), 1206–1218.

Edmonds, B. A. and Culling, J. F. (**2009**), "Interaural correlation and the binaural summation of loudness," The Journal of the Acoustical Society of America **125**(6), 3865–3870.

Epp, B. and Verhey, J. (**2009**a), "Superposition of masking releases," Journal of Computational Neuroscience **26**(3), 393–407.

Epp, B. and Verhey, J. L. (**2009**b), "Combination of masking releases for different center frequencies and masker amplitude statistics," The Journal of the Acoustical Society of America **126**(5), 2479–2489.

Epp, B., Yasin, I., Lüddemann, H., and Verhey, J. L. (**unpublished**), "Interrelation Between Sensation Level and Auditory Evoked Potentials Under Conditions of Masking Release [in preparation]," .

Epp, B., Yasin, I., and Verhey, J. L. (**2012**), "Interrelation Between Sensation Level and Auditory Evoked Potentials Under Conditions of Masking Release [abstract]," in *Abstracts of the 35th MidWinter Meeting of Association for Research in Otolaryngology (ARO)*, pp. 145, abstract nr. 414.

Ernst, S. M., Uppenkamp, S., and Verhey, J. L. (**2010**), "Cortical representation of release from auditory masking," NeuroImage **49**(1), 835–842.

Ernst, S. M., Verhey, J. L., and Uppenkamp, S. (**2008**), "Spatial dissociation of changes of level and signal-to-noise ratio in auditory cortex for tones in noise," NeuroImage **43**(2), 321–328.

Fastl, H. and Zwicker, E. (**2007**), "Binaural Hearing," in *Psychoacoustics: Facts and models* (Springer-Berlin), Springer series in information sciences, chap. 15, 3rd ed.

Festen, J. M. (**1993**), "Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice," The Journal of the Acoustical Society of America **94**(3), 1295–1300.

Festen, J. M. and Plomp, R. (**1990**), "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," The Journal of the Acoustical Society of America **88**(4), 1725–1736.

Fletcher, H. (**1940**), "Auditory Patterns," Rev. Mod. Phys. **12**, 47–65.

Grose, J. H. and Hall, J. W. (**1992**), "Comodulation masking release for speech stimuli," The Journal of the Acoustical Society of America **91**(2), 1042–1050.

Guski, R. (**1997**), "Psychological Methods for Evaluating Sound Quality and Assessing Acoustic Information," Acta Acustica united with Acustica **83**(5), 765–774.

Hall, J. W., Buss, E., and Grose, J. H. (**2006**), "Binaural comodulation masking release: Effects of masker interaural correlation," The Journal of the Acoustical Society of America **120**(6), 3878–3888.

Hall, J. W., Buss, E., and Grose, J. H. (**2011**), "Exploring the additivity of binaural and monaural masking release," The Journal of the Acoustical Society of America **129**(4), 2080–2087.

Hall, J. W., Cokely, J. A., and Grose, J. H. (**1988**), "Combined monaural and binaural masking release," The Journal of the Acoustical Society of America **83**(5), 1839–1845.

Hall, J. W. and Grose, J. H. (**1995**), "Amplitude discrimination in masking release paradigms," The Journal of the Acoustical Society of America **98**(2), 847–852.

Hall, J. W., Grose, J. H., and Dev, M. B. (**1997**), "Signal detection and pitch ranking in conditions of masking release," The Journal of the Acoustical Society of America **102**(3), 1746–1754.

Hall, J. W., Grose, J. H., and Haggard, M. P. (**1990**), "Effects of flanking band proximity, number, and modulation pattern on comodulation masking release," The Journal of the Acoustical Society of America **87**(1), 269–283.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**), "Detection in noise by spectro-temporal pattern analysis," The Journal of the Acoustical Society of America **76**(1), 50–56.

Hansen, H., Verhey, J. L., and Weber, R. (**2011**), "The Magnitude of Tonal Content. A Review," Acta Acustica united with Acustica **97**(3), 355–363.

Haykin, S. and Chen, Z. (**2005**), "The Cocktail Party Problem," Neural Computation **17**(9), 1875–1902.

Hellman, R. P. (**1982**), "Loudness, annoyance, and noisiness produced by single-tone-noise complexes," The Journal of the Acoustical Society of America **72**(1), 62–73.

Hermes, D. J. (**2009**), "Lecture series on Sound Perception: Loudness," URL `http://home.ieis.tue.nl/dhermes/lectures/SoundPerception/05Loudness.html`.

Hirsh, I. J. (**1948**), "The Influence of Interaural Phase on Interaural Summation and Inhibition," The Journal of the Acoustical Society of America **20**(4), 536–544.

Hirsh, I. J. and Pollack, I. (**1948**), "The Role of Interaural Phase in Loudness," The Journal of the Acoustical Society of America **20**(6), 761–766.

Jeffress, L. A., Blodgett, H. C., Sandel, T. T., Wood, C. L., and III (**1956**), "Masking of Tonal Signals," The Journal of the Acoustical Society of America **28**(3), 416–426.

Kleiner, M., Brainard, D., and Pelli, D. (**2007**), "What's new in Psychtoolbox-3?" in *Perception 36 ECVP Abstract Supplement*.

Klink, K. B., Dierker, H., Beutelmann, R., and Klump, G. M. (**2010**), "Comodulation Masking Release Determined in the Mouse (Mus musculus) using a Flanking-band Paradigm," Journal of the Association for Research in Otolaryngology **11**(1), 79–88.

Kwon, B. J. (**2002**), "Comodulation masking release in consonant recognition," The Journal of the Acoustical Society of America **112**(2), 634–641.

Levitt, H. (**1971**), "Transformed Up-Down Methods in Psychoacoustics," The Journal of the Acoustical Society of America **49**(2B), 467–477.

Licklider, J. C. R. (**1948**), "The Influence of Interaural Phase Relations upon the Masking of Speech by White Noise," The Journal of the Acoustical Society of America **20**(2), 150–159.

Luck, S. J. (**2005**a), "Filtering," in *An introduction to the event-related potential technique* (MIT Press), chap. 5.

Luck, S. J. (**2005**b), "An Introduction to Event-Related Potentials and Their Neural Origins," in *An introduction to the event-related potential technique* (MIT Press), chap. 1.

Maurer, K., Lang, N., and Eckert, J. (**2005**), *Praxis der evozierten Potentiale* (Steinkopff Verlag, Darmstadt), 2nd ed.

McFadden, D. (**1986**), "Comodulation masking release: Effects of varying the level, duration, and time delay of the cue band," The Journal of the Acoustical Society of America **80**(6), 1658–1667.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**), "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness," Journal of the Audio Engineering Society **45**(4), 224–240.

Moore, B. C. J., Hall, J. W., Grose, J. H., and Schooneveldt, G. P. (**1990**), "Some factors affecting the magnitude of comodulation masking release," The Journal of the Acoustical Society of America **88**(4), 1694–1702.

Nelken, I., Rotman, Y., and Yosef, O. B. (**1999**), "Responses of auditory-cortex neurons to structural features of natural sounds," Nature **397**(6715), 154–157.

Neuert, V., Verhey, J. L., and Winter, I. M. (**2004**), "Responses of Dorsal Cochlear Nucleus Neurons to Signals in the Presence of Modulated Maskers," The Journal of Neuroscience **24**(25), 5789–5797.

Neukirch, M., Hegerl, U., Kötitz, R., Dorn, H., Gallinat, J., and Herrmann, W. (**2002**), "Comparison of the Amplitude/Intensity Function of the Auditory Evoked N1m and N1 Components," Neuropsychobiology **45**(1), 41–48.

Niedermeyer, E. and Da Silva, F. (**2005**), "EEG Recording and Operation of the Apparatus," in *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields* (Lippincott Williams & Wilkins), Doody's all reviewed collection, chap. 7.

Näätänen, R. and Picton, T. (**1987**), "The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure," Psychophysiology **24**(4), 375–425.

Picton, T. W. (**2011**a), "Introduction: Past, Present, and Potential," in *Human Auditory Evoked Potentials* (Plural Publishing), chap. 1.

Picton, T. W. (**2011**b), "Late Auditory Evoked Potentials: Changing the Things Which Are," in *Human Auditory Evoked Potentials* (Plural Publishing), chap. 11.

Piechowiak, T., Ewert, S. D., and Dau, T. (**2007**), "Modeling comodulation masking release using an equalization-cancellation mechanism," The Journal of the Acoustical Society of America **121**(4), 2111–2126.

Poulton, E. (**1989**), "Contraction Biases," in *Bias in quantifying judgments* (Lawrence Erlbaum), chap. 7.

Pressnitzer, D., Meddis, R., Delahaye, R., and Winter, I. M. (**2001**), "Physiological Correlates of Comodulation Masking Release in the Mammalian Ventral Cochlear Nucleus," The Journal of Neuroscience **21**(16), 6377–6386.

Reale, R. A. and Brugge, J. F. (**1990**), "Auditory cortical neurons are sensitive to static and continuously changing interaural phase cues," Journal of Neurophysiology **64**(4), 1247–1260.

Richards, V. M. (**1987**), "Monaural envelope correlation perception," The Journal of the Acoustical Society of America **82**(5), 1621–1630.

Riedel, H., Granzow, M., and Kollmeier, B. (**2001**), "Single-sweep-based methods to improve the quality of auditory brain stem responses, Part II: Averaging methods," Zeitschrift für Audiologie/Audiological Acoustics **40**, 62–85.

Roberts, B. (**1998**), "Effects of spectral pattern on the perceptual salience of partials in harmonic and frequency-shifted complex tones: A performance measure," The Journal of the Acoustical Society of America **103**(6), 3588–3596.

Roberts, B. and Bregman, A. S. (**1991**), "Effects of the pattern of spectral spacing on the perceptual fusion of harmonics," The Journal of the Acoustical Society of America **90**(6), 3050–3060.

Rupp, A., Las, L., and Nelken, I. (**2007**), "Neuromagnetic Representation of Comodulation Masking Release in the Human Auditory Cortex," in *Hearing - From Sensory Processing to Perception* (Springer), chap. 14, pp. 125–132.

Schooneveldt, G. P. and Moore, B. C. J. (**1987**), "Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," The Journal of the Acoustical Society of America **82**(6), 1944–1956.

Soderquist, D. R. and Shilling, R. D. (**1990**), "Loudness and the binaural masking level difference," Bulletin of the Psychonomic Society **28**(6), 553–555.

Thompson, S. K., von Kriegstein, K., Deane-Pratt, A., Marquardt, T., Deichmann, R., Griffiths, T. D., and McAlpine, D. (**2006**), "Representation of interaural time delay in the human auditory midbrain," Nature Neuroscience **9**(9), 1096–1098.

Timoney, J., Lysaght, T., Schönwiesner, M., and McManus, L. (**October 5-8 2004**), "Implementing Loudness Models in Matlab," in *Proc. of the 7th International Conference on Digital Audio Effects (DAFx04), Naples, Italy.*

Townsend, T. H. and Goldstein, D. P. (**1972**), "Suprathreshold Binaural Unmasking," The Journal of the Acoustical Society of America **51**(2B), 621–624.

van de Par, S. and Kohlrausch, A. (**1999**), "Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters," The Journal of the Acoustical Society of America **106**(4), 1940–1947.

Verhey, J., Pressnitzer, D., and Winter, I. (**2003**), "The psychophysics and physiology of comodulation masking release," Experimental Brain Research **153**(4), 405–417.

Verhey, J. L., Dau, T., and Kollmeier, B. (**1999**), "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model," The Journal of the Acoustical Society of America **106**(5), 2733–2745.

Verhey, J. L. and Heise, S. J. (**2012**), "Suprathreshold Perception of Tonal Components in Noise Under Conditions of Masking Release," Acta Acustica united with Acustica **98**(3), 451–460.

Verhey, J. L., Rennies, J., and Ernst, S. M. A. (**2007**), "Influence of Envelope Distributions on Signal Detection," Acta Acustica united with Acustica **93**(1), 115–121.

von Kriegstein, K., Griffiths, T. D., Thompson, S. K., and McAlpine, D. (**2008**), "Responses to Interaural Time Delay in Human Cortex," Journal of Neurophysiology **100**(5), 2712–2718.

Vormann, M., Schick, A., Meis, M., Klatte, M., and Mellert, V. (**1998**), "Tonhaltigkeit aus psychophysikalischer Sichtweise: Untersuchung zur Versuchspersoneninstruktion bei der Bewertung der Tonhaltigkeit," Fortschritte der Akustik - DAGA '98 , 456–457.

Wojtczak, M. and Viemeister, N. F. (**2008**), "Perception of suprathreshold amplitude modulation and intensity increments: Weber's law revisited," The Journal of the Acoustical Society of America **123**(4), 2220–2236.

Zieliński, S., Rumsey, F., and Bech, S. (**2008**), "On Some Biases Encountered in Modern Audio Quality Listening Tests-A Review," Journal of the Audio Engineering Society **56**(6), 427–451.

List of Figures

Appendix

Appendix A

Experiment I: Additional individual results

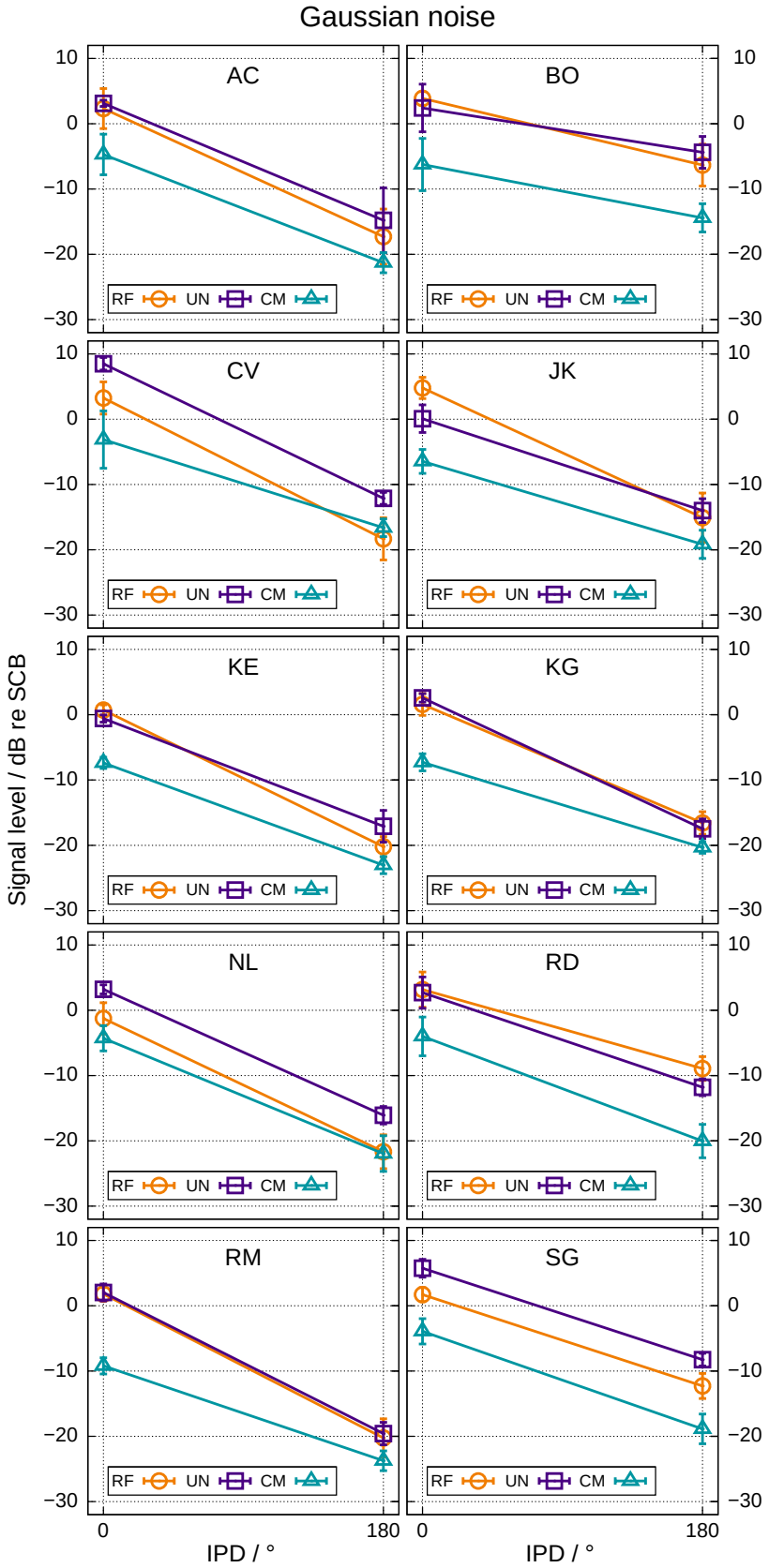

Figure A.1: Individual masked thresholds for the Gaussian noise masker. Mean thresholds over four runs are plotted as a function of IPD. Thresholds are shown relative to the level of the signal centred masker band for the RF (circles), UN (squares), and CM conditions (triangles). Error bars indicate ± 1 standard deviation.

Appendix B

Experiment II: Additional individual results

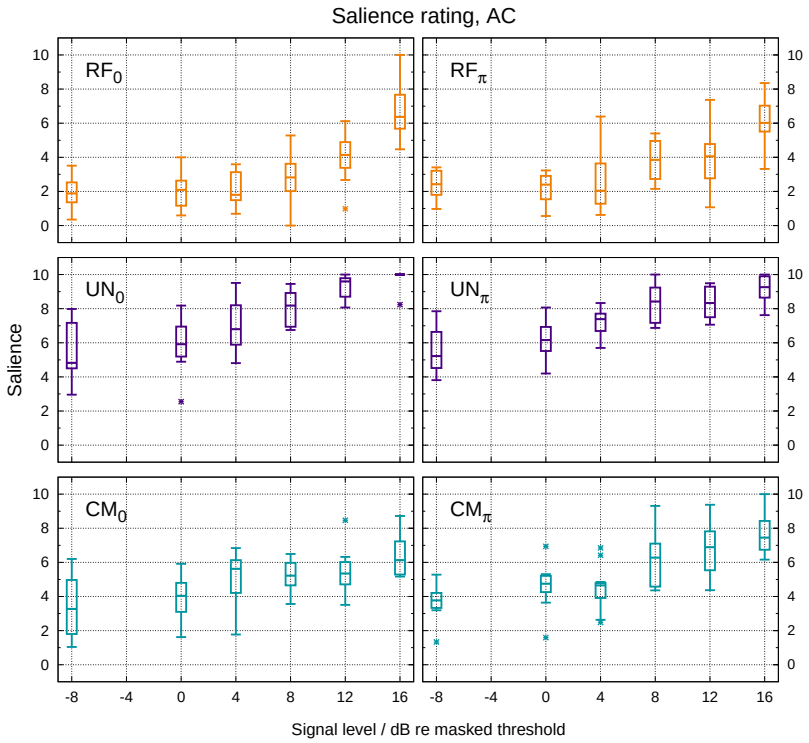

Figure B.1: Box-and-whisker plot of the obtained salience for subject AC for the RF (upper row), UN (middle row), and CM (bottom row) condition, with a signal IPD of 0° (left panels) or 180° (right panels). Ratings are plotted as a function of signal level relative to the corresponding masked threshold.

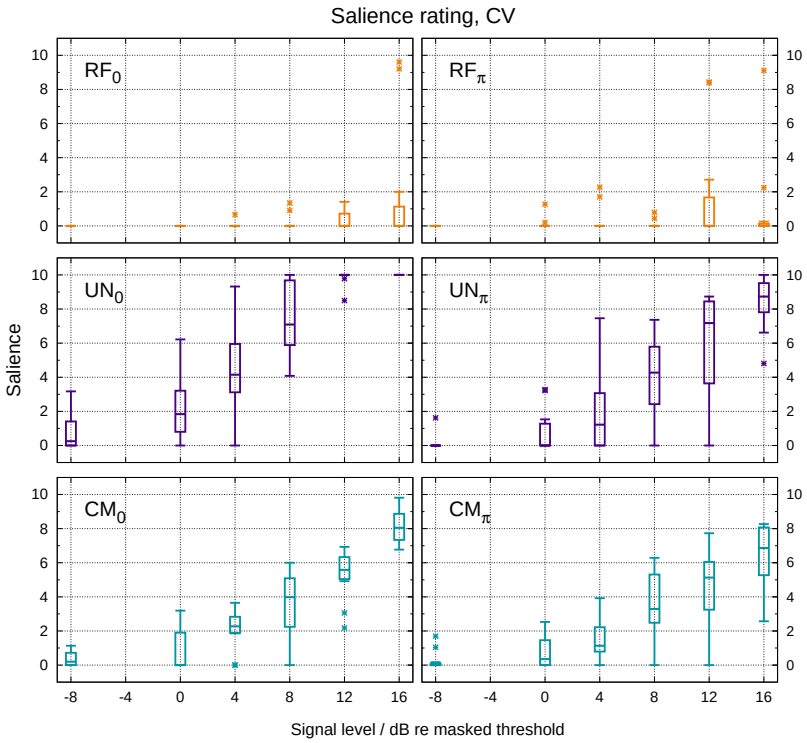

Figure B.2: Box-and-whisker plot of the obtained salience for subject CV for the RF (upper row), UN (middle row), and CM (bottom row) condition, with a signal IPD of 0° (left panels) or 180° (right panels). Ratings are plotted as a function of signal level relative to the corresponding masked threshold.

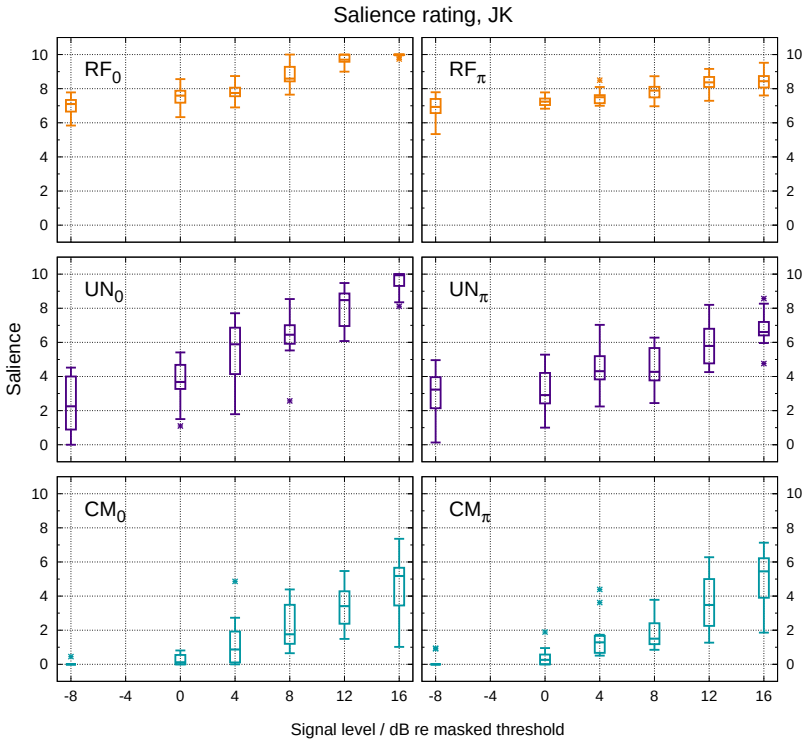

Figure B.3: Box-and-whisker plot of the obtained salience for subject JK for the RF (upper row), UN (middle row), and CM (bottom row) condition, with a signal IPD of 0° (left panels) or 180° (right panels). Ratings are plotted as a function of signal level relative to the corresponding masked threshold.

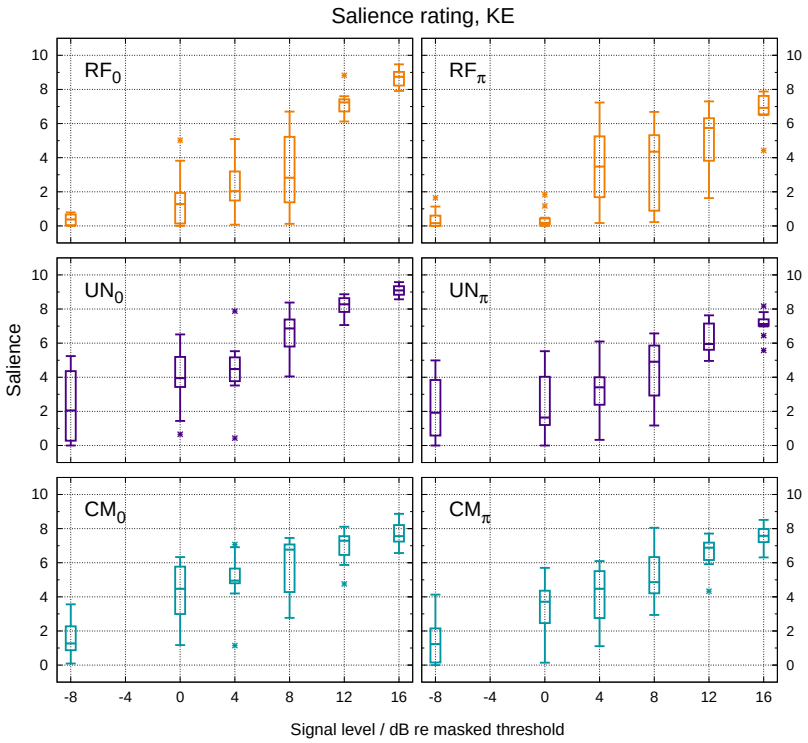

Figure B.4: Box-and-whisker plot of the obtained salience for subject KE for the RF (upper row), UN (middle row), and CM (bottom row) condition, with a signal IPD of 0° (left panels) or 180° (right panels). Ratings are plotted as a function of signal level relative to the corresponding masked threshold.

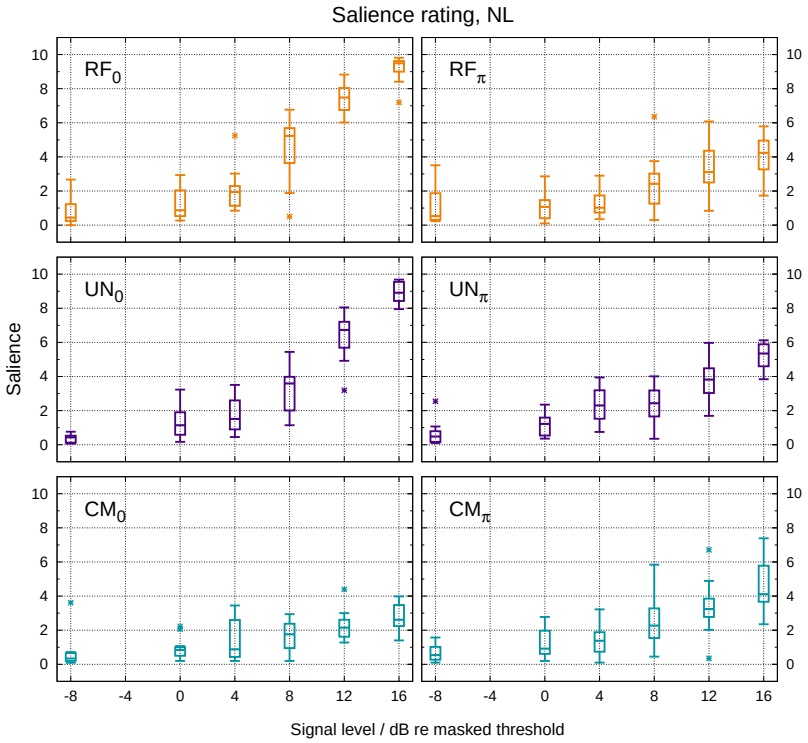

Figure B.5: Box-and-whisker plot of the obtained salience for subject NL for the RF (upper row), UN (middle row), and CM (bottom row) condition, with a signal IPD of 0° (left panels) or 180° (right panels). Ratings are plotted as a function of signal level relative to the corresponding masked threshold.

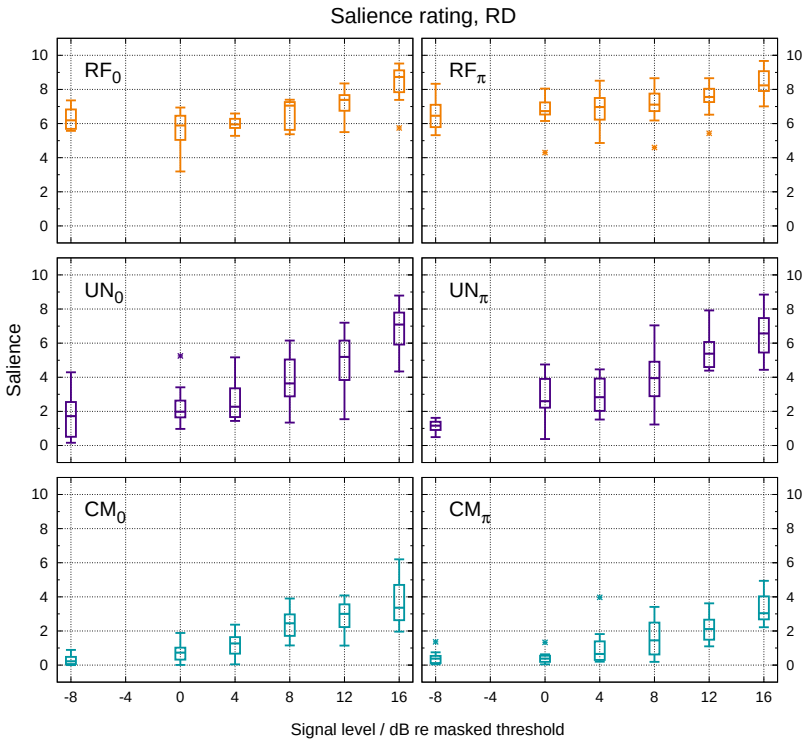

Figure B.6: Box-and-whisker plot of the obtained salience for subject RD for the RF (upper row), UN (middle row), and CM (bottom row) condition, with a signal IPD of 0° (left panels) or 180° (right panels). Ratings are plotted as a function of signal level relative to the corresponding masked threshold.

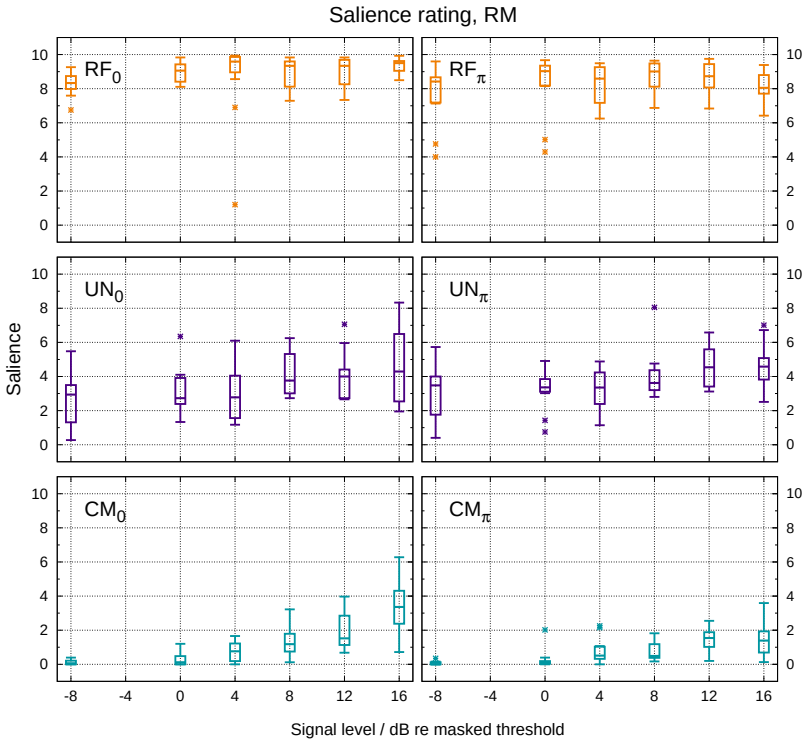

Figure B.7: Box-and-whisker plot of the obtained salience for subject RM for the RF (upper row), UN (middle row), and CM (bottom row) condition, with a signal IPD of 0° (left panels) or 180° (right panels). Ratings are plotted as a function of signal level relative to the corresponding masked threshold.

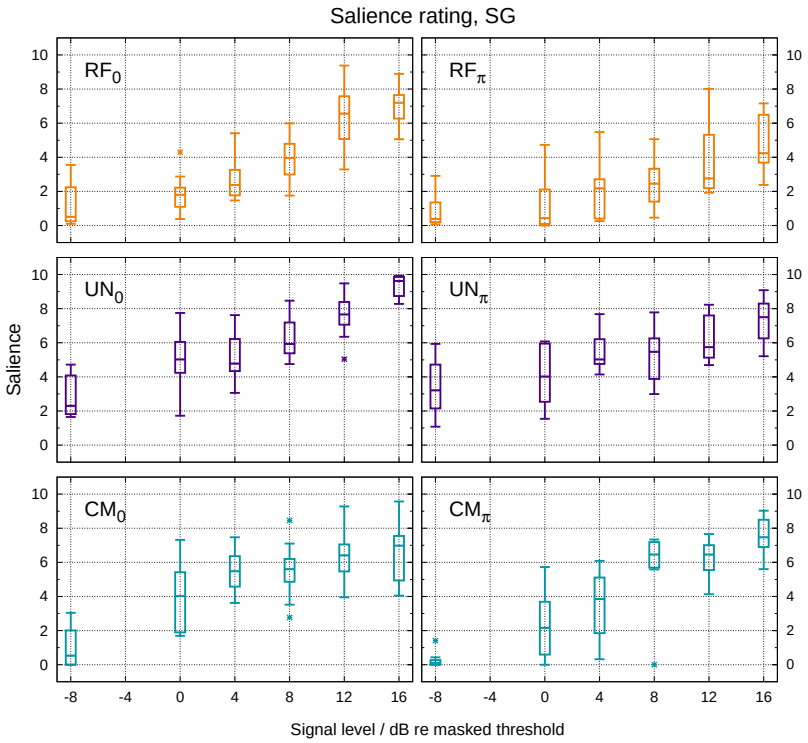

Figure B.8: Box-and-whisker plot of the obtained salience for subject SG for the RF (upper row), UN (middle row), and CM (bottom row) condition, with a signal IPD of 0° (left panels) or 180° (right panels). Ratings are plotted as a function of signal level relative to the corresponding masked threshold.

Appendix C

Experiment III: Additional individual results

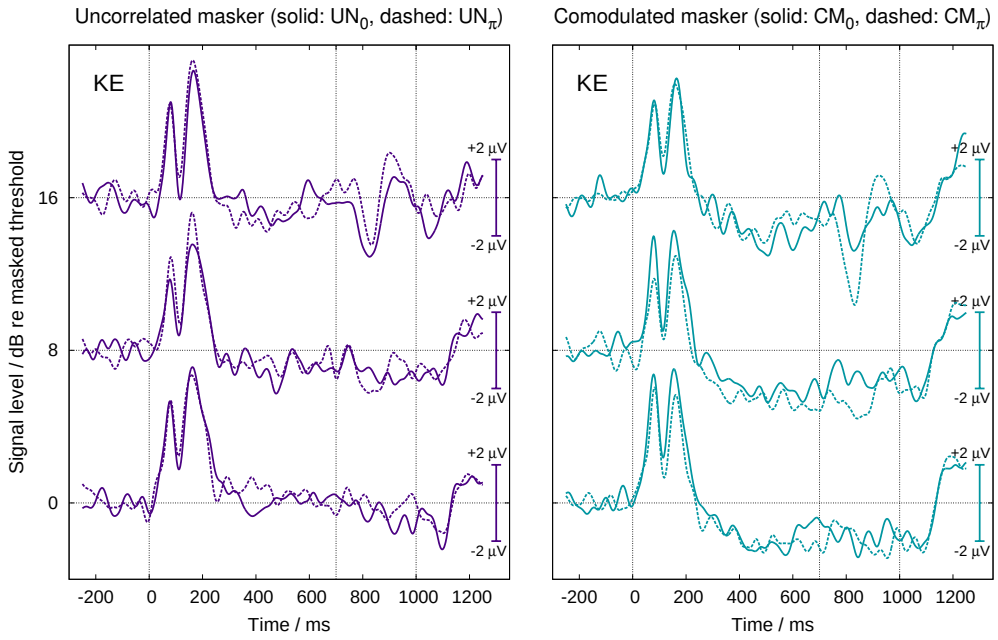

Figure C.1: Individual AEP data for subject KE, averaged over 400 sweeps. Waveforms are shown for the latency interval of -250 (i.e. pre-stimulus) to 1250 ms. The left and right panel illustrate the obtained potentials for the UN and CM conditions, respectively, with a signal IPD of 0° (solid lines) or 180° (dashed lines). Waveforms are plotted for three different levels of the masked sinusoidal signal: 0, 8, and 16 dB relative to masked threshold (indicated on the y-axes on the left-hand side). The potentials were baseline corrected by subtracting the arithmetic mean over the 250 ms pre-stimulus period from the waveforms. Vertical, dashed lines indicate the masker onset at 0 ms, the signal onset at 700 ms and the stimulus end at 1000 ms.

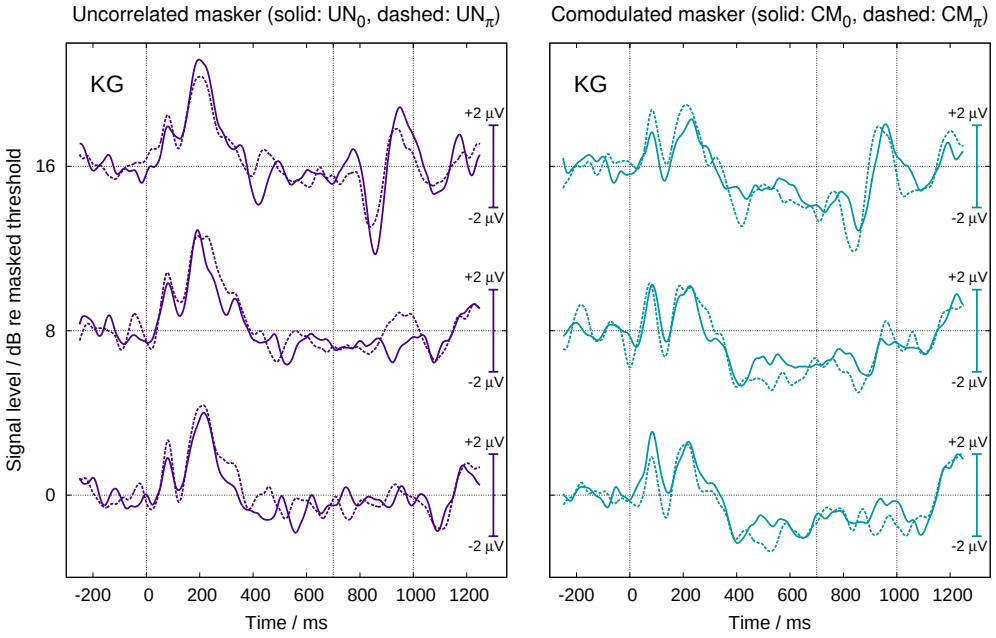

Figure C.2: Individual AEP data for subject KG, averaged over 400 sweeps. Waveforms are shown for the latency interval of -250 (i.e. pre-stimulus) to 1250 ms. The left and right panel illustrate the obtained potentials for the UN and CM conditions, respectively, with a signal IPD of 0° (solid lines) or 180° (dashed lines). Waveforms are plotted for three different levels of the masked sinusoidal signal: 0, 8, and 16 dB relative to masked threshold (indicated on the y-axes on the left-hand side). The potentials were baseline corrected by subtracting the arithmetic mean over the 250 ms pre-stimulus period from the waveforms. Vertical, dashed lines indicate the masker onset at 0 ms, the signal onset at 700 ms and the stimulus end at 1000 ms.

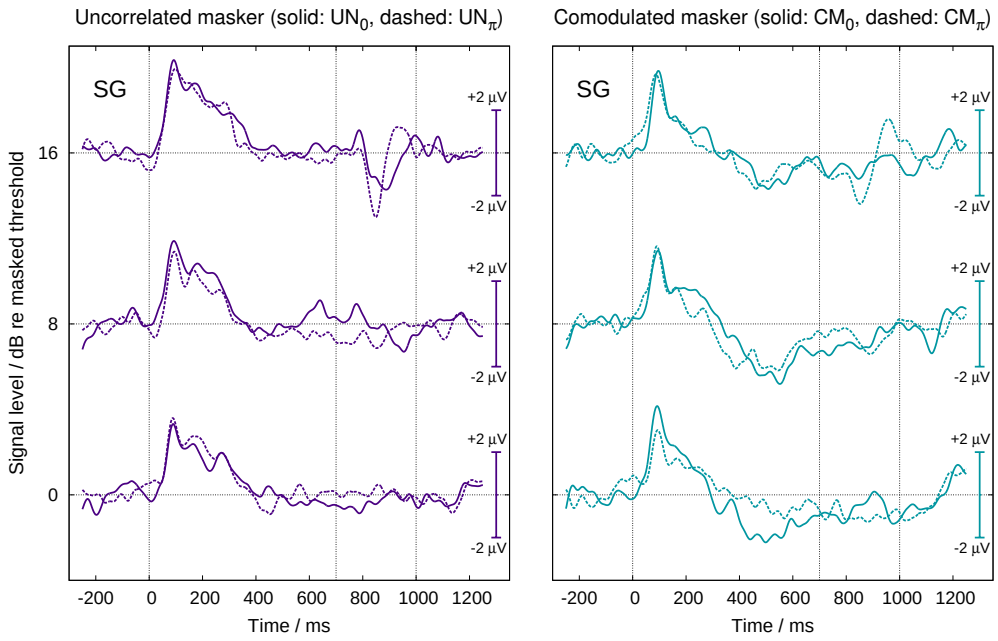

Figure C.3: Individual AEP data for subject SG, averaged over 400 sweeps. Waveforms are shown for the latency interval of -250 (i.e. pre-stimulus) to 1250 ms. The left and right panel illustrate the obtained potentials for the UN and CM conditions, respectively, with a signal IPD of 0° (solid lines) or 180° (dashed lines). Waveforms are plotted for three different levels of the masked sinusoidal signal: 0, 8, and 16 dB relative to masked threshold (indicated on the y-axes on the left-hand side). The potentials were baseline corrected by subtracting the arithmetic mean over the 250 ms pre-stimulus period from the waveforms. Vertical, dashed lines indicate the masker onset at 0 ms, the signal onset at 700 ms and the stimulus end at 1000 ms.