CARL VON OSSIETZKY universität OLDENBURG

TONE-NOISE DICHOTOMY INVESTIGATING TONAL CONTENT MAGNITUDE & PITCH STRENGTH

Von der Fakultät für Mathematik und Naturwissenschaften
der Carl von Ossietzky Universität Oldenburg
zur Erlangung des Grades und Titels eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
angenommene Dissertation

Hans Hansen

geboren am 29. Mai 1979
in Hamburg

Erstgutachter: Prof. Dr. Volker Mellert
Zweitgutachter: Prof. Dr. Jesko Verhey
Tag der Disputation: 4.6.2010

Wherever we are, what we hear is mostly noise. When we ignore it, it disturbs us. When we listen to it, we find it fascinating.
John Cage, Silence.

Abstract

This thesis explores various aspects of the tone-noise dichotomy. Based on existing research literature, two concepts are selected from a multitude and yield a two-fold structure: the magnitude of tonal content and the pitch strength. The magnitude of tonal content represents the relation between two separate objects: the tonal part and the noise part of a sound. These two parts can be related in their magnitude: How much "tone" is in a given sound? It refers to the mix of a given sound. The second dichotomy is within a given auditory object. It relates to the attributes of an auditory object: Is this object more noise-like or more tonal? How large is the pitch strength of this object? Summarizing, in one instance a relation between an acoustical foreground (tonal part) and its background (noise part) is considered, and in the other an attribute of an object is in the focus.

In a cross-cultural study, the perceptual and connotative factors of noise with added tonal components are investigated with a semantic differential in France, Japan, and Germany. The stimuli are varied in spectral content and signal-to-noise ratio. The three factor solution found – an evaluation, a timbre, and a power factor – is stable across cultures. As a second step, an item bias analysis is introduced and conducted. Most of the differences in the items found at *face value* between cultures could be assigned to differential item usage. Lastly, the relevancy of item bias analysis is discussed.

To investigate the hypothesis whether the pitch strength is related to frequency discrimination performance, a segregation experiment is conducted in which the pitch strength of the stimuli was the main cue. Participants had to judge whether a tone and narrowband noise were perceptually fused into one object, a tonal noise, or if they were perceptually separate as a tone and noise. As parameters, i.e., independent variables, the signal-to-noise ratio and the center frequency of the noise with different bandwidths were varied. This performance correlated highly with the frequency discrimination performance. Via this segregation performance based on pitch strength, the pitch strength can be directly linked to the frequency discrimination performance.

At last, the hypothesis is tested whether the judgement of the magnitude of tonal content is related to the tonal parts' partial loudness. Therefore, an adjustment experiment is conducted in which a single tone in noise is compared to a two-tone complex according to both concepts. It was found that the partial loudness was far easier and more intuitive to adjust in a magnitude adjustment experiment than the magnitude of tonal content, even though the concept was explained in detail and with examples. The hypothesis whether partial loudness can be identified with the magnitude of tonal content could thus be confirmed in this experiment. In addition, in the adjustment experiment an asymmetry in the contribution of the tones in the two-tone complex to the partial loudness is discovered. A modulation adjustment experiment explores these asymmetries, additionally supporting these findings.

Zusammenfassung

Diese Arbeit untersucht Aspekte der Dichotomie von Ton und Rauschen. Aus der Literatur wurden unter zahlreichen Begriffen aus verschiedenen Untersuchungszusammenhängen zwei extrahiert: die Tonhaltigkeit und die Ausgeprägtheit der Tonhöhe. Die Tonhaltigkeit zielt als Begriff auf das Größenverhältnis zweier Teile: den tonalen Teil und den rauschhaften Teil eines Geräuschs. Wie viel »Ton« ist in einem Geräusch enthalten? Die zweite Dichotomie ist innerhalb eines Hörobjekts zu finden: Ist das Objekt eher »rauschartig« oder »tonartig« ? Wie groß ist die »Ausgeprägtheit der Tonhöhe« ? Diese Fragen richten sich an ein Attribut, eine Eigenschaft, des Objekts. Tonhaltigkeit bezeichnet also ein Verhältnis von Ton und Rauschen, tonalem Vordergrund und rauschhaften Hintergrund. Ausgeprägtheit der Tonhöhe hingegen beschreibt eine Eigenschaft einen Hörobjekts.

In einer interkulturellen Studie werden die Wahrnehmungs- und konnotativen Dimensionen eines Rauschens mit tonalen Komponenten mit einem semantischen Differential in Japan, Frankreich und Deutschland untersucht. Die Stimuli variieren in ihrem spektralen Gehalt sowie in ihrem Ton-Rausch-Verhältnis. Das semantische Differential beschreibt drei Dimensionen, eine Evaluationsdimension, eine Klangfarbendimension und eine Stärkedimension. Dieses Dimensionen sind in allen untersuchten Ländern zu finden. In einem zweiten Schritt, wird die Itembias-Analyse vorgestellt. Die meisten Mittelwertsunterschiede zwischen den Ländern, können der Art, wie die Skala benutzt wird, zugeordnet werden. Es wird die Relevanz der Itembias-Analyse im Kontext der Klangcharakteruntersuchung diskutiert.

Um die Hypothese zu untersuchen, dass die Ausgeprägtheit der Tonhöhe das perzeptuelle Korrelat der ebenmerklichen Unterschiede in der Frequenz ist, wird ein Segregationexperiment durchgeführt. In diesem dient die Ausgeprägtheit der Tonhöhe als *cue*. Die Teilnehmer müssen jeweils entscheiden, ob ein Ton im Schmalbandrauschen mit den Rauschen verschmolzen ist oder eine »Ton-und-Rauschen« Situation vorliegt. Als Parameter wurde die Mittenfrequenz des Rauschen und mit ihr die Bandbreite des Rauschens sowie das Ton-Rausch-Verhältnis systematisch variiert. Die Daten dieses Versuchs korrelieren hoch mit gemessenen ebenmerklichen Unterschieden in der Frequenzen. Durch das Segregationsexperiment, welches die Teilnehmer auf Basis der Ausgeprägtheit der Tonhöhe durchführen, kann eine direkte Verbindung zu den ebenmerklichen Unterschieden in der Frequenz nachgewiesen werden.

Des weiteren wird die Hypothese untersucht, dass das Urteil der Tonhaltigkeit einem Urteil über die partielle Lautheit des tonalen Anteils entspricht. Hierfür, wurde eine Einstellungsexperiment durchgeführt, in dem ein einzelner Ton in Rauschen mit einem zwei Tonkomplex in Rauschen einmal auf gleiche Tonhaltigkeit und einmal auf

gleiche partielle Lautheit eingestellt wird. Das Ergebnis zeigt, dass es für die Teilnehmer wesentlich leichter war, die partielle Lautheit einzustellen, obwohl die Tonhaltigkeit mit Beispielen und einer Definition vorgestellt wurde. Darüber hinaus ergab sich im Einstellungsexperiment eine Asymmetrie in dem Beitrag der tonalen Komponenten zur partiellen Lautheit. Ein Modulationseinstellungsexperiment, welches diese Asymmetrie weiter untersucht, liefert zusätzliche Hinweise für eine Asymmetrie des Beitrags der tonalen Komponenten.

Contents

List of Tables

List of Figures

1 General introduction

This thesis investigates the tone-noise dichotomy by conceptually separating the magnitude of tonal content and the pitch strength. The investigation by von Helmholtz into the tone sensation serves as a good starting point of the present research into the tone-noise dichotomy. He claims that "the first and principal difference between various sounds experienced by our ear is that between *noises* and *musical tones*" (von Helmholtz, 1954, p. 7). Musical tones have their origin in periodic motion whereas noise has its origin in aperiodic motion (ibid.). Periodic motion enables the extension of the class of musical tones, which in the view of von Helmholtz has their origin in musical instruments, to all sound having a pitch that, at least in principal, can confer a musical melody. The howling of the wind or the squealing of a tram in a curve can be, in this sense, considered as a musical tone. Not only the bipolar concept of sound, tone and noise, but also the attributes by which tones can further be distinguished: the *force*, the *pitch*, and the *quality* (von Helmholtz, 1954, p. 10) are fundamental to this study. It will be those attributes of tone alongside which this dichotomous concept of sound will consistently be revisited.

In the first chapter, two lines of research are brought together to form a consistent concept. The first research line reported is the investigation into the phenomenon of pitch strength, which is an attribute of pitch describing the clarity or strength. In von Helmholtz's concept, this belongs to the sphere of quality. He gives an example to exemplify this sensation. Extremely deep tones, e.g., of an organ, gradually loose their definite pitch: "Here we find that even lower tones of the 16-foot octave, $C_{,}$ to $E_{,}$, begin to pass over into a droning noise, so that it becomes difficult [. . .] to assign their pitch with certainty ; . . . " (von Helmholtz, 1954, p. 175). In this chapter, it will be argued that the pitch strength is to a large extent the result of the frequency discrimination performance in the auditory system. However, in this chapter a distinction is drawn regarding a second line of research in which the magnitude of tonal content is considered. The tonal content as shown in Chapter 2 is viewed as a relation of magnitude between the noise and the tone part. Hence, it is often scaled as a tone-to-noise ratio. This tonal content of a given sound is not only crucial in auditory product design, but also in the studies concerning the impact of noise, as the tonal components often carry information about the source, which contributes to annoyance or, as an opposite, to attributed value. It is in parts governed by national and international standards and regulations (e.g. DIN 45681:2005-03, 2005; ISO 7779:1999, 1999). In the chapter, the limitations of the current standards are shown and options for an improvement, which lie in the use of the partial loudness of the harmonic part and with it the correct estimation of the masking threshold, are discussed. In a last step, the interaction between pitch strength and the magnitude of the tonal content are highlighted.

In Chapter 3, the relation between the tone loudness and the magnitude of tonal content is addressed. From the review of the current literature in Chapter 2, the hypothesis was supported that the loudness of the tonal part of a sound might very well address the scaling of the tonal content in a practically sufficient manner. This would enable the use of models of partial loudness, i.e, the loudness of a part of a sound in presence of another. This hypothesis is directly tested. To test common algorithms providing an estimation of tonal magnitude, the contribution of single tones to the overall partial loudness is investigated. In an modulation adjustment experiment, the qualitative, reciprocal influence of the contributing tones is investigated.

Chapter 4 is concerned with the magnitude of tonal content in noise in a cross-cultural context. Using the semantic differential technique, the connotative meaning is assessed with regard to the tone-to-noise ratio as well as the frequency and number of tones contributing. These connotative meanings are assessed in Japan, France, and Germany leading to similar perceptual dimensions cross-culturally: an evaluation, a metallic (timbre), and a power factor. An item bias analysis is introduced and applied. With this analysis translated items can be analyzed if they are used different in certain cultures. Several of those items are identified. The differences between mean values across cultures are then re-analyzed leaving only negligible differences across cultures on the investigated scales. At last in this chapter, the relevancy of bias analysis is discussed with regard to comparing categorical scaled data across cultures.

As seen in Chapter 2, pitch strength, an attribute of pitch, is independent from the magnitude of tonal content. Furthermore, this chapter provided evidence that pitch strength is inherently related to frequency discrimination performance. Chapter 5 investigates this hypothesis directly. On the one hand, a segregation experiment is conducted in which pitch strength is used as a cue, on the other hand the frequency difference limina of tone-noise complexes are measured. The increase in segregation and the increase in discrimination performance show a high correlation providing evidence for a quantitative Gestalt law governing the segregation performance.

Chapter 6 yields a general conclusion of the thesis. Furthermore, it provides an outlook on the problems yet to be solved, such as the integration of pitch strength and tonal content magnitude into one concept, or the contribution of individual tones to the magnitude of tonal content both for enhancing current noise standards.

2 Pitch strength and the magnitude of tonal content - a review†

Two phenomena are discussed in this review: pitch strength and the magnitude of tonal content. Pitch strength, along with pitch height and pitch chroma, is an attribute of pitch (Fig. 2.1).[1] As an example, the pitch strength of low-pass filtered noise is much weaker than that of a pure tone with the same pitch. Thus, pitch can be ordered on the scale from "faint" to "strong" (e.g., Fastl & Zwicker, 2007). Apart from this attribute of pitch itself, a different focus is the relation of the tonal component to the rest of the sound, especially in the context of noise annoyance or acoustic comfort (Hellman, 1982). In the latter, the relation between the tonal part (acoustic foreground) and the noise part (acoustic background) is under investigation. To describe this magnitude of tonal content the physical measures are often based on signal-to-noise ratio (SNR) approaches (e.g. DIN 45681:2005-03, 2005; ISO 7779:1999, 1999). The review aims at providing a literature overview for the former while providing a guide through the relevant proceedings for the latter. Also, the review attempts to discern the various labels under which both phenomena are researched, and sorts them into both categories. To account for this twofold structure, on the one hand pitch strength as an attribute of pitch and on the other hand the relation between tone and noise portion of the signal yields the structure of the review (Sec. 2.1 and 2.2). In Sec 2.3, the findings are discussed and the common ground of these two concepts is established.

2.1 Pitch Strength

Before addressing the concept of pitch strength in more detail, the attributes of pitch itself will be briefly introduced. Pitch height, often abbreviated as pitch, is the attribute of sounds on which scale they "can be ordered from high to low" (ANSI-S1.1-1994, 1994, p. 34). Pitch height is a major topic since the early days of psychoacoustical research (von Helmholtz, 1954) and there are textbooks entirely devoted to this topic (e.g., Plack et al., 2005). Apart from the dominant *high/low* scale, pitch has other attributes such as pitch strength and pitch chroma (see Fig. 2.1). Fastl & Zwicker (2007) define pitch strength as the scale on which a pitch can be estimated from *faint* to *strong*. This is similar to the *weak/strong* scale of Wightman (1973,

† Submitted as: *H. Hansen, J. Verhey & R. Weber, Pitch strength and the magnitude of tonal content - a review, Acta Acust united Ac.*

[1] The translation of pitch as well as pitch height into German is both "Tonhöhe". Thereby it refers to the most prominent feature of pitch.

pp. 413ff.). It is viewed as independent from the *high/low* scale, which is the general scale of pitch height as a metathetic continuum (Stevens, 1957). Pitch, at least melodic pitch (i.e., pitch which can confer a melody) is bound by an upper and lower limit of pitch height. The upper limit is at about 5 kHz (Semal & Demany, 1990), while the lower limit is at about 30 Hz (Pressnitzer et al., 2001). Close to these limits the pitch becomes weaker, the pitch strength diminishes. Another attribute of pitch is pitch chroma (e.g. Warren et al., 2003). It refers to the timbre similarities of, e.g., octave transposed tones.

In the following sections, different aspects of pitch strength as they appear under various names in different context will be presented to allow for a detailed picture of this sensation. Additionally, Table 2.1 provides an overview of possible translations in French, English, and German.

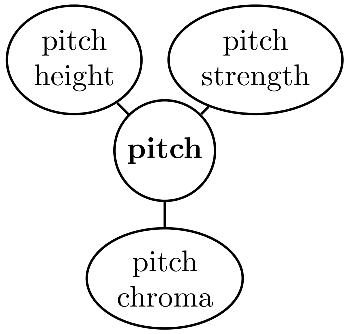

Figure 2.1: *The attributes of pitch: pitch height (ANSI-S1.1-1994, 1994, p. 34), pitch strength (e.g. Fastl & Stoll, 1979) and pitch chroma (e.g. Warren et al., 2003)*

2.1.1 Psychoacoustics

Tonal density

One of the earliest accounts of a report linked to pitch strength, is what Stevens (1934) describes as "auditory density", "compactness", or "concentration". This sensation is obvious comparing pure tones of 200 Hz and 4 kHz, the latter being more "compact". Guirao & Stevens (1964) investigate the auditory density further estimating hypothetical equal tonal density contours in the sound pressure level frequency domain. In their experiment the frequency range is 250–4k Hz. Considering the 1 kHz pure tone as an example, the reported increase in tonal density with sound pressure level corresponds well with data of pitch strength from Fastl (1989, sketched in Fig. 2.2a). A more detailed comparison of the concepts would have been possible, if the tonal density had been measured for frequencies higher than 4 kHz as well. As shown in Fig. 2.2a, the pitch strength decreases in this higher frequency range (> 4 kHz). Weber et al. (2009) measured equal loudness contours and equal pitch strength contours by adjusting the point of subjective equality with regard to a 1 kHz tone for both sensations. They investigated frequencies from 125 Hz to 8 kHz. The contours for loudness and pitch strength show different bandpass characteristics with the pitch strength having far steeper flanks, and showing a decrease in pitch strength beginning at 4 kHz in agreement with Fastl (1989), which is not observed as strong in the equal loudness contours.

Pitch salience – the pitch of complex tones

A harmonic complex tone consists of multiple tone components. But which is the pitch which can be actually heard? If there are multiple pitches, which is the most salient one? The most predominant pitch is the pitch associated with the fundamental of the harmonic complex tone, but it is very well possible to listen to individual harmonics (e.g., Terhardt et al., 1986). If the fundamental is missing the pitch, under certain constrains, is still heard "at" the fundamental. What harmonics contribute predominantly to this phenomenon is tackled with the dominance region concept, a frequency region in which the harmonics contributed mostly to the holistic pitch percept of the harmonic complex tone(Plomp, 1967; Ritsma, 1967; Moore et al., 1985; Dai, 2000). This pitch, however, is far fainter than the pitch of a harmonic complex tone or a corresponding pure tone (Fastl & Stoll, 1979; Fruhmann & Kluiber, 2005).

Kohlrausch & Houtsma (1992) investigated another pitch phenomenon: the pitch related to the edges of broadband signals, i.e., the upper edge of an harmonic complex tone or a low pass noise with a steep spectral skirt. The pitch is close to the highest frequency which is close to the edge. Moore & Glasberg (1989) mention that the quality of the pitch of these two stimuli types is different, i.e., the pitch of the harmonic complex tone is more very similar to the pitch elicited by a sinusoid. Kohlrausch & Houtsma could relate this directly to the fact how well the pitch can be matched. The pitch deviations of the low-pass noise was one order of magnitude higher that in the harmonic complex tone case.

Pitch strength of complex tones and narrowband noises

Magnitude estimation experiment provides a reliable instrument to measure the pitch strength of various classes of sounds, such as pure tones, harmonic complex tones, and narrowband noises, at different frequencies (Fastl & Stoll, 1979). In general, the results indicate that stimuli with a discrete line spectrum such as pure tones and harmonic complex tones with a higher pitch strength than stimuli with a more continuous spectrum such as narrowband noise.

Pure tones were investigated showing a decline of pitch strength toward low and high frequencies (broad maximum 0.75–4 kHz, Fig. 2.2a), small stimulus duration (< 200 *ms*, Fig. 2.2b), and a rather weak dependency on level (10 % per 10 dB SPL, Fig. 2.2b). Various stimuli where investigated further to build a parametrical model of pitch strength (Fruhmann, 2006b). At first, the pitch strength of bandpass noise is considered (Fruhmann, 2004a), yielding a decrease in pitch strength with increasing bandwidth (Fig. 2.2d). In harmonic complex tones with constant sound pressure level, the pitch strength decreases by filtering out the lower harmonics consecutively or if the fundamental is attenuated (Fruhmann & Kluiber, 2005). Thus, the “virtual pitch” has in general a weaker pitch strength than the “spectral pitches”. Furthermore, experiments with harmonic complex tones show a significant decrease in pitch strength with an increasing number of harmonics while, again, the sound pressure level is kept constant (Fruhmann, 2006a).

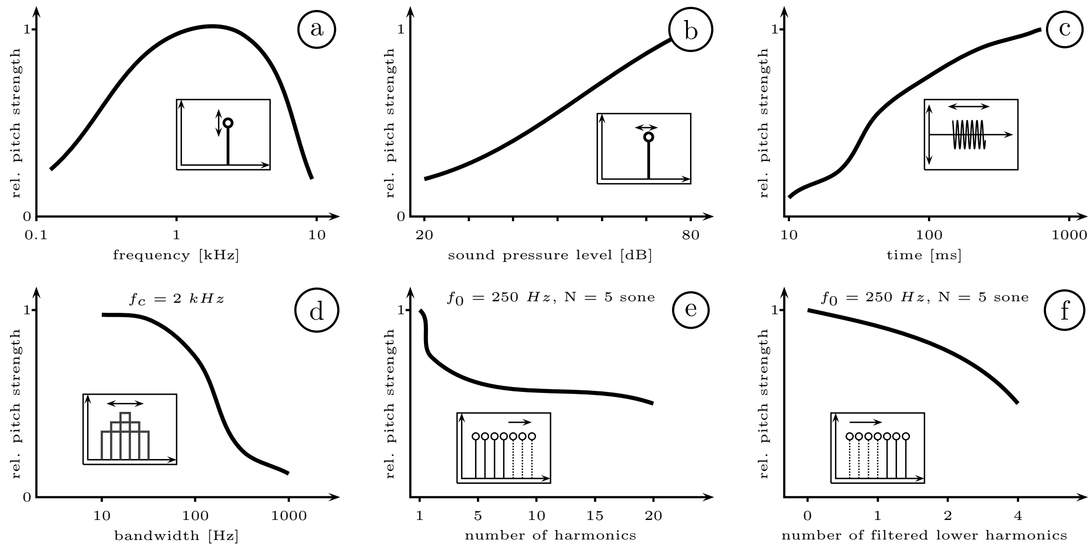

Figure 2.2: *Sketch of various dependencies of relative pitch strength: (a) bandpass characteristics of fundamental frequency, (b) increase with sound pressure level, (c) increase with duration (Fastl, 1989) (d) decrease with bandwidth for narrowband (Fruhmann, 2004a), (e) decrease with decrease in number of harmonics (Fruhmann, 2006a), (f) decrease with number of filtered lower harmonics (Fruhmann & Kluiber, 2005).*

Pitch strength of iterated rippled noise

The pitch strength of rippled or comb-filtered noise has lead to thorough investigations as the pitch strength is adjustable through several parameters. In fact, theses two noises are just special cases of iterated rippled noise. Iterated rippled noise is generated by delaying the noise and adding it with a gain g back onto it after a certain delay d. This is process is repeated n-times. In the *add-original* version the original signal is delayed $1 \cdot d$, $2 \cdot d$,... and added, while in the *add-same* configuration the signal resulting out of the delay-add step is delayed and added (for details refer to Yost, 1996a). Ripple noise is iterated ripple noise with only one iteration. The pitch strength can be influenced by the gain g and the number of iteration. Early on, the pitch strength was compared within this particular class (Yost & Hill, 1978) as the pitch strength of the rippled noise varies by attenuating the delayed feedback. The amount of attenuation is used as an psychophysical parameter for measuring pitch strength. Yost & Hill (1978) found that pitch strength is independent of the overall SPL (25–65 dB SPL, lower than 25 dB SPL the pitch strength drops). Furthermore, the strongest pitches were found within limits of 20 Hz and 2 kHz around 500 Hz. Comparing these "repetition pitches" with other classes, the pitch strength is generally weaker than that of pure tones or harmonic complex tones (Fastl & Stoll, 1979).

The number of iteration has an influence on the pitch strength. The height of the first peak of the signal's autocorrelation function is a good estimator of pitch strength of this stimuli class (Patterson et al., 1996; Yost, 1996b; Yost et al., 1996), even in the case of ambiguous pitch percept (Yost, 1997). Also with iterated rippled noise, the pitch strength of the sounds is found to be smaller than even at a large number of iterations that of harmonic complex tones (Shofner & Selas, 2002; Fruhmann, 2004b). This is the reason why it is not possible to map all possible magnitudes of pitch strength on this easily scalable class. Within the reported investigations of iterated rippled noise's pitch, the focus laid on the temporal aspects of pitch perception, i.e., pitch strength as an indicator for regularity. Following this studies, clicks trains were investigated (Kaernbach & Demany, 1998; Yost et al., 2005), which had an equal overall regularity but varied in temporal fine structure. It was found that the temporal fine structure played an important role. These sound class challenged the long-term autocorrelation or spectra to estimate the pitch strength.

Binaural Pitch

Binaural pitch is a class of pitch phenomena, where pitch is created through binaural interaction. This means presenting the stimuli monaurally there is no pitch sensation, but by presenting the stimuli binaurally a pitch sensation occurs. One of the first account of these pitch phenomena induced by binaural phase shift is by Cramer & Huggins (1958). This "Huggins Pitch" is created by applying the phase shift from 0 to 2π over a narrow frequency region. The pitch heard is equal to a pitch of a tone at the a center frequency of this phase shift. Other inter-aural phase manipulation that elicits a pitch perception are (I) an abrupt change from 0 to 2π (binaural-edge-pitch; Klein & Hartmann, 1981) and (II) a change from 0 to an random phase difference (binaural-coherent-edge-pitch, BCEP; Hartmann & McMillon, 2001).

The pitch strength of these stimuli is often described as faint (e.g. Cramer & Huggins, 1958), but it can be enhanced using multiple phase shifts at harmonics (Bilsen, 1976). The pitch strength of this binaural perception, is weaker than that of a pure tone with corresponding pitch which can be inferred from melody recognition of the three binaural pitches compared to the pure tone case (Akeroyd et al., 2001). The authors show also that the pitch strength of the different binaural pitches differs. In experiments on the performance of binaural model of pitch perception, pitch strength is operationalized the with a frequency discrimination task (Hartmann & Zhang, 2003). In a study by Santurette & Dau (2007), who tested if the binaural pitch was indeed musical, the performance in melody recognition with binaural pitches was consistent with their performance in frequency discrimination. As a general trend, hearing impaired listeners which showed a reduced frequency discrimination performance showed a reduced melody recognition.

Emergence of pitch in noise

The emergence of pitch at threshold or in noise serves as another example for varying pitch strength. To find an analog to the achromatic interval in vision, i.e., the range where a visual stimulus is detected but the color is not determined, tones at threshold are considered (Pollack, 1948, pitch strength is here referred to as tonality). The difference of detection and identification of a tone at hearing threshold was first associated with the frequency difference limen (Harris & Myers, 1949). Several studies report an increase of the frequency difference limen of tones in broadband noise with increasing SNR close to masking threshold (Harris, 1948, 1966; Henning, 1967; Cardozo, 1974; Sinnott & Brown, 1993). For complex tones it was shown that the pitch extraction mechanism is rather robust and that a reduction of frequency difference limen performance due to added noise can be related to the detectability of the tone itself (Gockel et al., 2006). Gockel et al. argue that the noise addition is not a good way to reduce the pitch strength of a complex tone as the noise level necessary to reduce the frequency difference limen performance is in a range where the actual detectability is at stake. However, in a study where a superposition of broadband noise and iterated ripple noise with a large number of iterations has to be compared to iterated noise with various iterations, the pitch strength of the superposition varies (Patterson et al., 1996). The broadband noise and the highly iterated noise form one perceptual object with a certain pitch strength, thus detectability does not become an issue. In the studies mentioned above as well as in the study by Santurette & Dau (2007) reported in Sec. 2.1.1, pitch strength and the frequency difference limen seem to be connected.

2.1.2 Physiology

At various levels of the auditory pathway researcher have investigated pitch and pitch strength. One of the earliest stages in this pathway, is the auditory nerve. Cariani & Delgutte (1996) analyzed the temporal discharge patterns of auditory nerve fibers of cats in response to several different stimuli. To analyze the data, all-order inter-spike interval distributions were pooled across fibers. Interval peak's location and height correspond well with pitch and pitch strength. Nevertheless, the pitch strength of the pure tone is underestimated by this pure temporal approach based on the pooled inter-spike interval distributions.

In one of the highest stages of auditory processing, the primary auditory cortex, highly pitch selective neurons are found (Bendor & Wang, 2005). The authors have shown that their discharge rate is highly correlated to the pitch strength of the stimuli presented. Furthermore, neuro-imaging studies have evidence for a dissociation of regularity processing and the intensity processing in the human brain (Gutschalk et al., 2002, 2004; Ernst et al., 2008). In perceptual terms this would be the dissociation of pitch strength and loudness. Gutschalk et al. (2004) argue that these stages in the auditory system are involved in the cross-frequency channel integration and the calculation of pitch and pitch strength values. Hall & Plack (2007) argue for a number of criteria which have to be satisfied that a brain region should be referred as a *pitch centre*. One among others, the response magnitude should show a "covariation with salience (ibid, p. 83)", which in the terms of this review is dubbed pitch strength. Before referring to the results, it has to be mentioned that the pitch strength has been operationalized by measuring the frequency difference limen. As a major result, the authors stress the fact that a common pitch centre for all different stimuli classes could be found at an individual level, but not across listeners, which is not inline with recent data from Puschmann et al. (2010).

2.1.3 Modeling

Probabilistic model of pitch strength

The viewpoint established so far can be turned around, i.e., not any longer is the external experimenter considered who investigates a known stimulus and the responses thereof, but a living system which has to infer properties out of the acoustical signal reaching the ear (Handel, 2006). Under normal conditions this is not an invertible problem as many stimuli elicit the same response in the system. Thus, the system response's nature is probabilistic.

Following this general idea, Schwartz & Purves (2004) postulate that the pitch of complex tones is determined by the probabilistic relationship between auditory stimuli and their possible natural sources. To determine the distribution of natural sources, they took the TIMIT[2] database of speech sounds, as they claim, that both evolutionarily and developmentally, speech next to music is the predominant source of pitch. The likelihood that an event X is the source of a periodic test stimulus Y is calculated through cross-correlation with the all stimuli of the TIMIT database. The frequency of the maximum correlation is taken as an estimator of the pitch, while the value of the maximum correlation represents the pitch strength. Thus, according to the authors pitch strength turns out to be correlated with the probability that an event can be assigned to a probable source. No empirical weighting function have to be derived as any such weightings are only a counterpart of the stimuli distribution within the environment.

Parametrical model

Fruhmann (2006b) introduces a fully parametrical model based on the findings of Sec. 2.1.1. At first, the "spectral" and "virtual" pitch are extracted using a Fourier-t-transform and followed by the algorithm by Terhardt et al. (1982). Within this functional model framework, the pitch strength is inhibited by harmonics. In this way,

[2] Texas Instruments/Massachusetts Institute of Technology (TIMIT) Acoustic-Phonetic Continuous Speech Corpus

a ratio of fundamental's loudness and all harmonics' loudness is their basic estimator of pitch strength. An example, in the case of a pure tone the loudness of the fundamental and the harmonics is one as there is only one harmonic, the fundamental. Thus, the pure tone has the highest pitch strength, and with the addition of further harmonics this pitch strength is reduced. Again, it is important to note that this is not a common SNR, as the noise is comprised of the energy of the harmonics. All other dependencies, such as bandwidth, total loudness, fluctuation strength, roughness, and fundamental frequency, are introduced via a separate weighting function.

Physiologically motivated models

McLachlan (2009), in a physiologically based pitch strength model, provide a scheme in which spectral and temporal information are combined. After a preprocessing stage consisting out of a Gammatone filterbank, a functional hair cell model, and dendritic filtering, the spike rate is integrated. This output is then used not only to recognize the stimulus, i.e. a pure tone or a harmonic complex tone, but is processed in further stages. Tonotopic inhibition within the cochlear nucleus (CN) and periodotopic inhibition within the inferior colliculus (IC) are modeled to sharpen the neural response. The author argues that tonotopic patterns are necessary for the system to interpret the periodotopic data. Hence, in their pitch strength model, the channel closest to the fundamental is found by correlation of the tonotopic pattern with the harmonic complex tone template, similar to the purely spectral model by Terhardt (1974). In the case of the pure tone, the channel with the maximum spike rate is used. Then, the spike rates above a specific thresholds are summed of all best modulation frequencies of the subharmonics of the specified channel. McLachlan accounts for behavioral data of pitch strength and frequency difference limen of pure with regards to stimulus duration. Again, in the context of their model pitch strength and discrimination are linked to each other.

2.2 Description of tonal content

This section deals with the description of tonal content, i.e., the analysis of the relation between acoustical foreground comprised of tones or other auditory objects having a pitch and a background.

2.2.1 Psychoacoustical Research

Vormann et al. (1999a,b) introduced a new adaptive measurement technique to measure the tonality. This concept basically answers to the question: "Which sound is more tonal?", using the German term "tonhaltig", which refers to a concept of "tone-containing". The concepts clearly aims for the relational aspect of tonal content, i.e. the balance of tone and noise in a sound. Using this method the participants had to complete a alternative forced choice (AFC) procedure in which the a tone in noise was compared to the test stimulus under investigation.[3]

Vormann et al. (1999b) added successively higher harmonics to a pure tone which was embedded in uniform exciting noise (see Fig. 2.3). The overall level was not adjusted. Vormann et al. argue that the increase of the perceptual SNR by 5 dB by adding the first harmonic can neither be explained by energy (3 dB) nor through loudness (10 dB). Vormann et al. (2000a) repeated the experiment to solve this problem, and to check for experimental design issues. In the second experiment, the tonal energy is kept constant, i.e. with doubling the number of harmonics each individual harmonic is reduced by 3 dB. Only the addition of the second harmonic has a major influence on the harmonic complex tone. Additional harmonics add to the perceptual SNR as they did in the model of Aures (1985a). Also, the authors point out that the tonality might be connected to the masking level, i.e. to the physical SNR. They measured the masking threshold which shows a common curvature with the subjective SNR, and thus indicating the energy above masking threshold might be relevant (Vormann et al., 2000a, see their Fig. 5). Note, that this approach is taken by the DIN 45681:2005-03 (2005); ISO 7779:1999 (1999) on the level of single tones. The reduction of masking threshold through multiple harmonics, however, is not included.

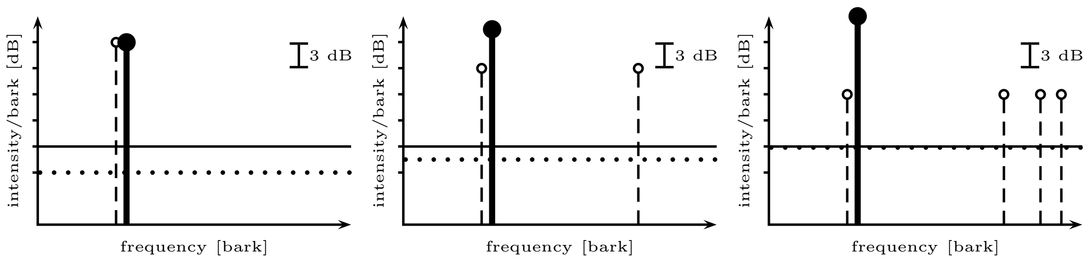

Figure 2.3: *Schematic overview of the experiment of Vormann et al. (2000a). The sinusoidal components (dashed stems) are presented in uniform exited noise (horizontal line). Overall SPL is kept constant, therefore the individual components are reduced by 3-dB for each doubling of components. All components fall into different critical bands. The adjusted value of a single tonal component to match equal tonal magnitude is depicted (bold stems). The magnitude of tonal content increases with increasing masking threshold (Equivalent to the decrease of the threshold of 1.5 dB per doubling the number of components if the masking threshold is expressed in dB per component, Grose & Hall, 1997, here, it is depicted as the masking threshold of the total harmonic complex tone, which thus increases by 1.5 dB per doubling of the components (dashed horizontal line)). Thus, the SNR ratio above the threshold stays constant. On the basis of this result, Vormann et al. (2000a) argues that the SNR above threshold is a valid approach the the magnitude of tonal content.*

[3]Early on, Vormann et al. (1999a,b) use the term "tonality" as an English translation for this estimated magnitude of tonal content.

How can the tonal energy across multiple tones be summed? A simple addition of energies above individual thresholds to describe the tonal content in an adequate manner is also rejected by Hansen & Weber (2009). The correlation between the total level excess above calculated masking threshold and the semantic differential factor describing the tonal content is reduced due to different frequency content, suggesting that an interaction between the tones has to be taken into account. Furthermore, there is evidence that the judgement of tonal content magnitude as suggested by Vormann et al. (1999a,b) is indeed identical with the adjustment of the partial loudness of the tonal part (Hansen & Weber, 2010; Verhey & Heise, 2010).

Hansen & Weber (2008) are broadening the concept of magnitude of tonal content into time by considering the duration and the SNR. In their indirect scaling experiment of the magnitude of the tonal content, the SNR provide a first approach to the measured magnitude of the tonal content. Furthermore, the authors showed that the tonal content judgement is indeed influenced by duration, i.e., the magnitude increases by a constant factor for each doubling of the duration in a range from 250 ms to 2 s.

2.2.2 Physiological findings

The importance of the masking threshold but also the SNR as an estimator of tonal content is supported by brain imaging studies by Ernst et al. (2008, 2009). In their studies a broadband noise with a tone is presented to the listeners. At low SNRs, the tone does not change the overall level, but its partial loudness changes. As soon as 0 dB SNR is reached, the overall loudness starts to increase. Given he hypothesis that there are voxels which respond to the increase below an SNR of 0 dB (SNR voxels) and voxels which response increases at 0 dB SNR (RMS voxels), the authors find almost discrete voxel clusters responding in either way (Ernst et al., 2008). Moreover, looking at the response at SNR levels which were part of the post hoc classification as one voxel or the other, the response to the RMS voxel is at a minimum at SNR levels were the SPL is constant, while the response to SNR voxels show a further increase with increasing SNR. In Ernst et al. (2009), the findings are further investigated. In this study, the masking threshold is lowered by using amplitude modulated noise, similar cortical regions are discovered. Again, the SNR voxel response increase with SNR, and the RMS voxel response is correlated to the overall loudness, even in regions which did not serve as a classification requirement. Interestingly, the SNR voxels in the modulated and unmodulated case seem to show a similar intensity increase, relative to their respective noise condition, thus a modulated noise can shift the SNR by several dB although the tone SPL is kept constant, indication that the threshold is an important parameter in the description of tonal content. Furthermore, the judgement of the tonal content and the tonal part's partial loudness are equally influenced by a shift in masking threshold due to temporal fluctuations of the noise masker.

2.2.3 Modeling and standards

Tonalness

The definition of tonalness refers to Terhardt & Stoll (1981) who define tonalness as the attribute of having more or less pronounced pitches. In fact, they only use a dichotomous scale tonal/not tonal. Aures (1985b) adopts the definition and defines tonalness along a scale between the antipodes of tone and noise. Tonalness as an sound attribute is related to the number and strength of pitches within a sound (Terhardt, 1998, p. 305) and it is viewed analogous to pitch strength as measured by Fastl & Zwicker (e.g. 2007).

At first, the tonalness algorithm as an integrating measure of all pitches within a stimulus has to determine all the parts electing a pitch. Therefore, all prominent tonal components and narrowband components along with their level excess above the noise intensity within one critical band are calculated (Terhardt et al., 1982). The tone-on-tone masking effects are taken into account. Afterwards, these tonal components are removed from the spectrum in order to determine the loudness of the residual noise. The spectral pitch weights of (Terhardt et al., 1982) are summed up quadratically. Additionally, a weight for the spectrally broadened tonal components is introduced. This sum weighted by an exponent is multiplied with an relative loudness weighted also by an exponent to be proportional to the tonalness (Aures, 1985a). The relative loudness is the difference between the calculated loudness with tones N an without tones N_{BG} divided by the total calculated loudness N: $\frac{N-N_{BG}}{N}$ (Aures, 1985a, mod. Eq. 12).

The dependency of tonal content on tonalness is rather complicated as factors co-variate. Especially, the total loudness, the level of the tonal components and the noise loudness. Consider tonal components spectrally centered around 700 Hz (the maximum of the spectral weight (Terhardt et al., 1982)) with the same SPL in uniform-exciting noise (Fastl & Zwicker, 2007). The distance of the components should be not too small to minimize the effect of tone-on-tone masking. Doubling these components would lead to a doubling of the pitch weight sum. In addition, the total loudness increases, while the loudness of the residual noise stays the same. Following the model above, doubling the components would lead to an increase in tonalness. This increase, however, would be far from being double as the result from Vormann et al. (2000a) would suggest. Nevertheless, the algorithm of (Aures, 1985a) was applied with good success for harmonic complex tones in noise, i.e., describing the perceptual tone-to-noise ratio (Fingerhuth & Parizet, 2008).

International and national standards

For the characterization of noise immission, the various industrial norms employ mainly using tone-to-noise ratios (DIN 45681:2005-03, 2005; ANSI S1.13-2005, 2005; ISO 7779:1999, 1999). These tone-to-noise ratios have to be either be reported for a qualitative account of timbre variations or they are added as a penalty to the sound pressure level and are trying to replace subjective evaluations.

According to the DIN 45681:2005-03 (2005) for the evaluation of tonal components, tone level of a single tonal components or a NBN and the background level within a critical band around the component under investigations is calculated. Two phenomena are additionally considered. This tone-to-noise ratio is correct by the masking thresholds (Fastl & Zwicker, 2007). The second phenomenon relates to the fact that the pitch strength of the NBN is fainter than that of a pure tone at the respective center frequency. The DIN 45681:2005-03 (2005) evaluates only those tonal components which have at least 70 % pitch strength of a pure tone (the 70 % criterion). The ANSI S1.13-2005 (2005)'s annex provides two procedures for determining tonal prominence for describing the noise character supplementarily. The first method, the tone-to-noise ratio, is very similar to the DIN 45681. Ratio of the tone SPL and the noise SPL within a critical band surrounding the tone is taken as the tone-to-noise ratio. Then, the tone-to-noise ratio has to exceed a certain level, e.g., 8 dB for tones lower than 1 kHz. Thus, the tone has to exceed the masking threshold by 11 dB to be classified as 'prominent' taking 1 kHz as reference. The ISO 7779:1999 (1999) takes a similar approach to the ANSI S1.13-2005 (2005)'s tone-to-noise method.

Another measure suggested in the same norm is the 'prominence ratio (PR Bienvenue & Nobile, 1991).' This is ratio of the energy within the critical band around the tone to the mean energy of the critical band neighboring the critical band in question. Taking again 1 kHz as the reference, the PR has to exceed 9 dB in order to be classified as prominent.

2.3 Merging Aspects

Various research contexts with regards to pitch strength and the scaling of tonal content have been mentioned. This section contains three subsection using an integrating view of the Sec. 2.1 and Sec. 2.2. So far, two phenomena have been discussed separately, now the possible common ground and interferences are summarized.

2.3.1 Pitch strength and the frequency discrimination performance

This section argues for the hypotheses that the pitch strength, i.e., the notion of "strong, clear" and "faint" pitches are related to their resolvability. Thus, it may, or in fact is, in many cases operationalized as frequency difference limen. In the following, the results from literature are discussed in the light of this hypothesis.

At first, direct accounts of pitch strength, i.e., results from magnitude estimation, are compared to experiments which estimate the frequency difference limen of the respective stimuli. The rapidly decreasing pitch strength's estimate for pure tones at high frequencies (above 4 kHz Fastl & Zwicker, 2007, p. 139) matches well with the rapid increase in frequency difference limen at frequencies above 4 kHz (Moore, 1973b). The same match is found for pure tones at lower frequencies. The correlation between the pitch strength estimates and the frequency difference limen is -0.85.[4] A similar good correlation is found for the increase of pitch strength with level which increases far slower than loudness (ca. 10 % / 10 dB (between 20 and 80 dB SPL); Fastl & Zwicker, 2007, p. 138), which compares very well with data of Wier et al. (1977) showing a similar increase in frequency difference limen for those SPLs. Inline with these findings, is the correlation between the pitch strength estimates of NBN (Fastl & Zwicker, 2007, p. 139) with their frequency difference limen (Michaels, 1957; Moore, 1973a) and the dependency of the estimated pitch strength (Fastl & Zwicker, 2007, p. 138) on duration of the stimulus (Moore, 1973b). The iterated rippled noise stimuli, one of the most thoroughly investigated stimuli class with regard to pitch strength, lacks to the knowledge of the authors extensive frequency difference limen measurement. As an exception, Yost et al. (1978) report the increase of frequency difference limen ($\Delta f/f$) with decreasing pitch strength.

Apart from this indirect comparisons, section 2.1.1 and 2.1.1 provide more direct evidence. The studies in binaural pitch perception with hearing impaired listeners suggest this connection (Santurette & Dau, 2007), as do the decreasing pitch strength and increasing frequency difference limen close to masking threshold (Harris & Myers, 1949). This view is adopted, McLachlan (2009) uses his pitch strength model to account for the frequency difference limen, while Hartmann & Zhang (2003) and Hall & Plack (2007) operationalized the pitch strength psychophysically by measuring the frequency difference limen.

Summing up the evidence and developments, there is good reason to identify the perception of "weak" and "strong" pitches with the discrimination performance of the auditory system.

2.3.2 Magnitude of tonal content

In the description of the magnitude of tonal content, especially when it comes to immission standards (Sec. 2.2.3), the focus lies on technical applicability (DIN 45681:2005-03, 2005, p. 4). It was shown by Vormann and colleagues that several issues cannot be covered using only SNR approaches, even if they are already taking the masking threshold of individual tones of broadband noise into account, e.g., the reduced masking threshold of harmonic complex tones in broadband noise (Vormann et al., 2000a). The tonalness approach by Aures (1985a) is rather complicated and a parameter intensive approach clouds the main dependencies. Though the algorithm's core might not be the summation of Terhardt's pitch weights, the idea to estimate the tonal part of the sound which shows figure-ground-segregation by estimating the tonal part of the total loudness with the help of the loudness of the residual noise. In Aures' tonalness calculation scheme this interpretation is inherent, as the weighting of the pitch weights are far smaller than the weighting of the loudness of the different parts. As this fraction tries to estimate the tonal part, it is not far fetched to take the partial loudness of the tones itself, e.g., by applying a model framework

[4] The data is taken from Moore (1973b), and taken out of Fig. 5.28 in Fastl & Zwicker (2007, p. 139)

of Moore et al. (1997). After identifying the tonal parts and dividing them from the background, the total tonal loudness could provide a reliable measure of the tonal content's magnitude (Hansen & Weber, 2010; Verhey & Heise, 2010).

2.3.3 Common ground

After it was suggested that the frequency difference limen is related to pitch strength and a useful approach to the magnitude of tonal content might be the partial loudness of all tones. The formerly entangled concepts have to be brought together again. How does the pitch influence the magnitude of tonal content? Is an NBN and a tone produce an equal tonal magnitude if their partial loudness is essentially the same? Aures (1985a) and also the DIN 45681:2005-03 (2005) are taking first steps to take the pitch strength into account: Aures by introducing a weighting for NBNs and the DIN 45681:2005-03 (2005) by introducing the mentioned 70 % criterion. How does the rapid decrease in pitch strength of frequencies above 4 kHz influence the magnitude of tonal content at these frequencies albeit the loudness growth function is very similar to those of lower frequencies? There is still no answer and various research opportunities open up.

2.4 Conclusion

The review sets apart the concept of tonal content magnitude and the concept of pitch strength and assign the various concepts to either of them. Auditory objects, as mentioned in the introduction, can provide a conceptualization in which the distinction between these terms is clear cut. Pitch strength is an attribute of a tonal object, while the magnitude of tonal contents focusses on the magnitude relation between figure and ground. The review provides the context of pitch strength research and suggesting that the pitch strength is closely related to the frequency/temporal regularity discrimination of the auditory system. A hypothesis still to be explicitly tested. For the tonal magnitude, the evidence was summed up to argue that it is closely related to the partial loudness of the tonal part. The review provides along with the partial loudness hypothesis the guidelines in which current standards could be improved: To judge the tonal magnitude algorithmically several issues have to be combined, i.e., a correct identification of the tonal part, estimating robustly the masking threshold for this part(s), and then estimating the partial loudness of the tonal components. Furthermore, if several tonal components are present how do they contribute to the overall tonal magnitude? The interaction between tonal content magnitude and pitch strength is yet to be investigated: How much is the tonal magnitude to be decreased to reach the same tonal magnitude if the components have a lower pitch strength. The 70 % criterion of the DIN 45681:2005-03 (2005) can be only a start for the evaluation of tonal components in noise.

Table 2.1: *Translation Table. Concepts in class I relate to a pitch scale "faint-strong". Concepts in class II are considered with the tone-noise dichotomy, i.e., noise and tone are the opposite ends of the scale, thus relating to a perceptual tone/noise ratio. Origin refers to publications using the term, and if possible, providing an appropriate translation. The term sonorité is in brackets as native French speakers questioned the translation.*

Class	*English*	*French*	*German*	origin
I	**pitch strength**	force tonale	Ausgeprägtheit der Tonhöhe	English/German (Fastl & Zwicker, 2007)
	tonal density		Tonale Dichte	English (Stevens, 1934), German translation by the authors
II	**magnitude of tonal content**	rapport signal à bruit perçu	Tonhaltigkeit	German (e.g. DIN 45681:2005-03, 2005)
	perceptual tone/noise ratio			English (Patterson et al., 2000)
	tonality			English translation of the German term by Vormann et al. (1999a)[a]
	tonalness	(sonorité)	Klanghaftigheit	German with English and French translations by Aures (1985a)

[a] The term tonality is already used in English: Music has tonality if it uses the notes of a major or minor scale. This meaning, however, is not implied by Vormann et al. (1999a).

3 Magnitude of tonal content and partial loudness†

Tonal components play a major role in the assessment of noise annoyance, but also in the identification of the sound source. Several national and international noise standards have been developed to assess these tonal components algorithmically (DIN 45681:2005-03, 2005; ANSI S1.13-2005, 2005; ISO 7779:1999, 1999). These standards principally use signal-to-noise (SNR) approaches: After a tone is identified, its energy is calculated and divided by the noise energy within one critical band. Every tone is treated separately even if they are harmonically related. These SNRs are either reported as additional information to describe the noise (ANSI S1.13-2005, 2005; ISO 7779:1999, 1999), or provide the basis for a penalty which is added to the overall noise level choosing the highest calculated SNR (DIN 45681:2005-03, 2005).

The main algorithm used to calculate the magnitude of tonal content within a given noise sample taking all components into account was derived by Aures (1985a). He introduced a way to calculate the magnitude of tonal content based on the *pitch weights* defined by Terhardt et al. (1982).[1] The calculation procedure implements a weighted sum of these pitch weights, which is in turn further weighted by the loudness ratio of the sound with and without the tones. The pitch weights are calculated as the SNR of the tone energy and the energy of the surrounding critical band, which is adjusted by the excitation of other tones present.

Vormann et al. (2000a), however, showed that algorithms based on single tonal components have limits. Using the methodology of magnitude adjustment of tonal content, which adjusts a tonal component in noise to the magnitude of tonal content in the test signal (Vormann et al., 1998), they showed that the distribution of tonal energy over several harmonic tonal components lead to an increase in magnitude of tonal content. The tonal components were presented within uniform excited noise and none of them shared a critical band. The authors measured the masking thresholds of the harmonic tone complexes, and the decrease of the overall threshold matched the increase in tonal content magnitude. The same tendency is observed in data published by Hansen & Weber (2009). The analysis of the semantic differential data revealed inter alia a timbre factor, which was largely influenced by the tonal content magnitude. The factor was plotted against the overall energy above

†Parts to appear in: *H. Hansen & R. Weber, Partial loudness as a criterion of tonal content, Proc Internoise, 2010.*

[1] The pitch weights were originally introduced for calculating spectral and, most of all, virtual pitches, but the authors left it open for a wider range of applications.

masking threshold. This leads to good correlation. However, for tonal components in noise which where harmonically related larger deviations were found. The energy above masking thresholds underestimated the magnitude of tonal content, as the threshold was only estimated for each tone separately. The influence of the masking threshold indicates that separately calculating each component contributing to the overall magnitude of tonal content of a given sound might be too simple. In the examples of Vormann et al. (1998) and of Hansen & Weber (2009), it would have been necessary to measure the masking threshold of the total tonal part.

Given the masking threshold plays an important role, how does the magnitude of tonal content increase above masking threshold? For a single tone, one turns to the loudness growth function of a tone in noise (Fastl & Zwicker, 2007). This function already provides the *magnitude* of the sensation. The magnitude of the tonal content is identified with its loudness, leading to the first hypothesis: The magnitude judgment of tonal content can be identified with the loudness of the harmonic part of a given sound. It is, in fact, the partial loudness of the harmonic part.

To analyze the contribution of individual tones to the magnitude of tonal content, a magnitude adjustment experiment was designed. Two-tone complexes in background noise were compared with a single tone in noise. The participants were requested to adjust the SNR of either the single tone in noise, or one of the two tones in the tone complex to match the two-tone complex as a whole or the single tone in magnitude of tonal content respectively. Therefore, in an interleaved experiment, a fully balanced approach was achieved as all tones were adjusted. Fig. 3.1 depicts all combinations. The magnitude of tonal content judgment was introduced by a stimulus developed by Vormann et al. (1998) on the basis of participant interviews. The stimulus consisted of a tone in noise slowly increasing and decreasing linearly in level. After this, to test the first hypothesis, the experiment was conducted again but with the instruction to adjust the loudness of the tone. In summary, this experiment has two instruction conditions: In the first condition, the magnitude of tonal content instruction is given and the other the partial loudness instruction is given.

The stimuli were designed according to the algorithmic approach to tonal content magnitude by Aures (1985a). To test the summation procedure provided, the experiment's stimuli are designed as follows. The two-tone complex is centered around 700 Hz, which is also the frequency of the single tone. Therefore, the frequency weighting, which is approximately symmetric to 700 Hz on a linear frequency axis, is the same for both tones. As only tones are considered, the narrowband noise weighting can be ignored. Thus for the stimuli chosen, only the weighting of the level excess, which is squared and summed, is relevant (e.g., two tones with the same level excess therefore have the same weight). This was tested in the condition X_{60dB}, in which the tones were 60 dB SPL and therefore about 20 dB above the noise level within its surrounding critical band (Fastl & Zwicker, 2007). Now, to take out the asymmetry due to the excitation pattern, which, according to Terhardt et al. (1982), reduces the level excess of tones reciprocally, the SPL of the tone was reduced by 15 dB (condition X_{45dB}). In this condition, the excitation pattern does not influence the level excess as the excitation pattern at the frequencies is already well below the noise level within the critical band. Thus, in this condition the calculation procedure of Aures (1985a) predicts similar level adjustment for the higher and lower tones.

The second experiment investigates the reciprocal influence of tones on their loudness. The pitch weights based on Terhardt et al. (1982) have a reciprocal, albeit different influence on each other: Basically one masks the other by contributing to the noise energy within the respective critical band. This implies that by modulating an additional tone, the modulator, the target tone would also be modulated as the threshold varies. Following the masking paradigm, the modulation of the target tone would be larger if the tone is lower due to the upward spread of masking. Therefore, the hypothesis is as follows: The modulation of a target tone is larger if the modulator is lower than the target tone compared to a condition were a the modulator is higher than the target tone.

To test this hypothesis a magnitude adjustment paradigm was used. The target tone was modulated. During the presentation of the target tone an additional tone either higher or lower was added: the modulator. This additional tone was also modulated. However, the target ceases to be modulated while the additional tone is present. The participant were asked to adjust the modulation of the target tone within the interval before and after the additional tone to the modulation of the target tone perceived while the additional modulated tone is present. This experiment was conducted at two different frequencies, 700 Hz and 1 kHz. The difference between the target and the modulator was 50 Hz in both frequency conditions.

The current study sets out to test the hypothesis whether the magnitude of tonal content of a given sound is identical with the task of judging the partial loudness of a harmonic part. Furthermore, suggested tone-tone interaction is investigated with regard to the hypothesis of reciprocal masking.

3.1 Tonal content magnitude and the partial loudness of tones – A comparison

3.1.1 Method

In this experiment a tone in noise is compared with a two-tone complex in noise regarding the magnitude of tonal content and the partial loudness of the tone(s). The experiment is fully balanced, i.e., every tone is adjustable. Also, the tones are presented in a reduced level condition (X_{45dB}), which is close to masking threshold.

Stimuli

The stimuli were broadband noise with added tonal components. The broadband noise had a lower cut-off frequency f_{lc} of 200 Hz and a higher cut-off frequency f_{hc} of 10 kHz. The single tonal component f_{st} was added at 700 Hz while the two-tone complex with f_{t1} and f_{t2} at 650 and 750 Hz. The broadband noise was at 60 dB SPL. The tonal components' levels are listed in Table 3.1. The stimuli are also tested with the tone level adjusted by -15 dB. The duration of tone(s) and noise was 800 ms. 50 ms cosine ramps were applied to the signal.

Table 3.1: *Level of the tonal components in every experimental condition. A $\star$ indicates the adjustable component within a condition. A second SPL condition was tested as well with all tonal components reduced by 15 dB (X_{45dB}).*

	X_{60dB} level [dB SPL]			X_{45dB} level [dB SPL]		
Exp. Cond.	f_{st}	f_{t1}	f_{t2}	f_{st}	f_{t1}	f_{t2}
I	$\star$	60	60	$\star$	45	45
II	63	$\star$	57	48	$\star$	42
III	63	57	$\star$	48	42	$\star$

Apparatus

The stimuli were generated digitally with the Matlab software (MathWorks), DA converted (RME-ADI-8 DS), amplified (TDT HB7) and presented diotically through headphones (Sennheiser HD 650). The transfer function was adjusted by a smoothed, inverse headphone transfer function. The participants were situated in a double-walled soundproof room.

Procedure

Magnitude of tonal content adjustment In the first part, the magnitude of tonal content was evaluated. Participants were given an introduction to the concept as proposed by Vormann et al. (1998), i.e., a background noise with an added tone is presented in such a way that the tone increases and then decreases linearly in dB SPL for a duration of 10 s. Then, the point of subjective equality was obtained by an adaptive two-interval, two-alternative procedure using a 1-up-1-down paradigm. The stimuli were presented in an interleaved order, i.e., the configuration changed every trial (Fig. 3.1). The pause between the stimuli was 500 ms. The participant's task was to choose the stimulus which had a larger magnitude of tonal content. At the start of the experiment, the adjustable component had a 75 dB SPL (Type I) and 72 dB SPL (Type II/III). For every choice of the signal interval, the level was reduced by 8 dB. This step size was reduced every odd reversal by a factor of 0.5 to a minimum final step size of 1 dB. The PSE is determined by the mean of the last eight reversals.

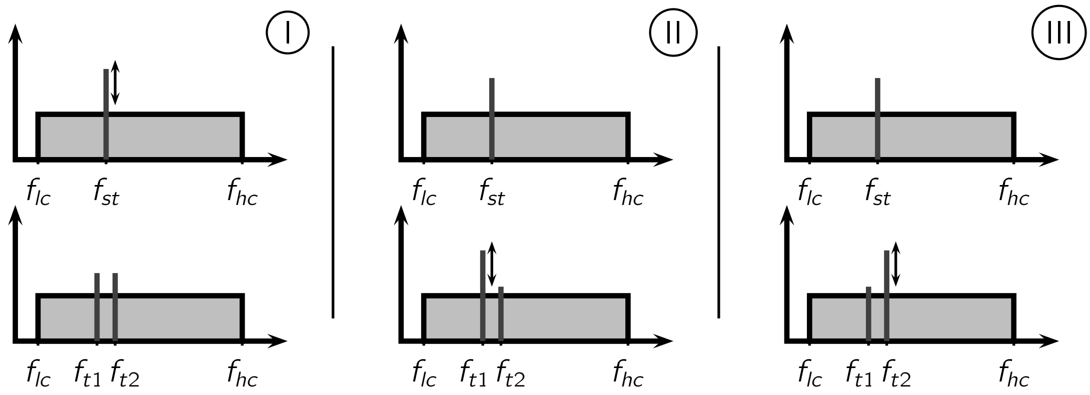

Figure 3.1: *A single tone condition is compared with a two-tone condition. Three different trial configurations (I-III) are obtained to fully balance the experiment. In every trial condition, the test tone to be adjusted is different.*

Partial loudness adjustment After the first part was completed, the participants requested to redo the whole set. The only difference is that they were instructed to choose the stimulus containing the louder tonal part.

Interview After both experiments (duration approx. 45 Min.), there was an open interview during which the participants reported their impressions. This interview focused on the criterion of the two tasks.

Participants

Eight participants voluntarily took part in this experiment (3 female, 5 male). None of them had any history of reported hearing loss. The participants' age ranged from 22 to 41 years with a median age of 28 years.

3.1.2 Results

Reliability The participants exhibited reliable behavior over the trials. This aspect was analyzed by evaluating the standard deviation of the last eight reversals. The mean standard deviation across participants and conditions of the tonal content magnitude adjustment is $\bar{\sigma} = 2.44$ dB SPL, while it is slightly lower for the partial loudness adjustment $\bar{\sigma} = 1.68$ dB SPL. Contrary to the loudness adjustment task, the adjustment of the tonal contents revealed a few high standard deviations. This result is discussed further in the next section, because it is reflected by comments made in the post-experimental interview. The mean standard deviations are similar to the standard errors reported by Buus et al. (1998): 1.6 dB and 2.4 dB over 4 reversals.

Post-experimental interview – A synopsis In the post-experimental interview all of the participants reported that the second task, the loudness adjustment, was easier and did come "more naturally" to them than the judgment of tonal content magnitude. The loudness adjustment seems more straightforward. In the case of the tonal content judgment, as many as half of the participants where unsure about their criterion, and where influenced by timbre difference between the tone and the two-tone condition. This, however, is only reflected in the standard deviation of the last eight reversals in some conditions of two participants. Two participants reported that they were already using the loudness criterion as a criterion for the tonal content.

Tonal content rating and the tone loudness Figures 3.2 and 3.3 depict the experiment's results as box-plots. In all conditions, the interquartile ranges are larger than for the tonal content magnitude task than the loudness adjustment. The same is true for the whiskers. T-tests, adjusted with a Bonferroni method for multiple comparisons, reveal that the differences between these two adjustments are not significantly different from zero ($p = 0.01$).

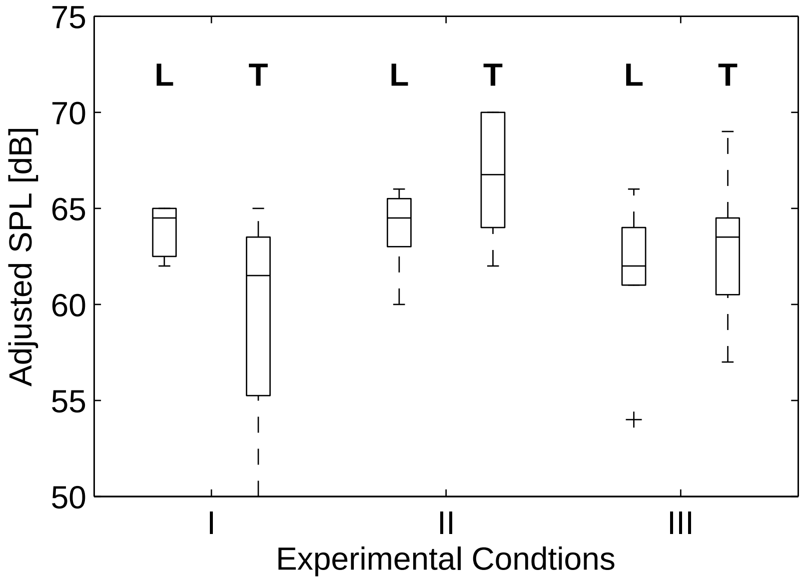

Figure 3.2: *Comparison of the adjusted SPL value between the task to adjust the magnitude of tonal contents (***T***) and the task to adjust the loudness (***L***) of the tonal part within each condition (refer to Fig. 3.1, level range X_{60dB}).*

Tone weighting in tonal content magnitude/partial loudness adjustment Analyzing the results from the loudness adjustment experiment, there is an asymmetry between the adjustment of the lower and the higher frequency tones in Fig. 3.2 and 3.3, even more so in the X_{45dB} condition. Firstly considering the adjustment of

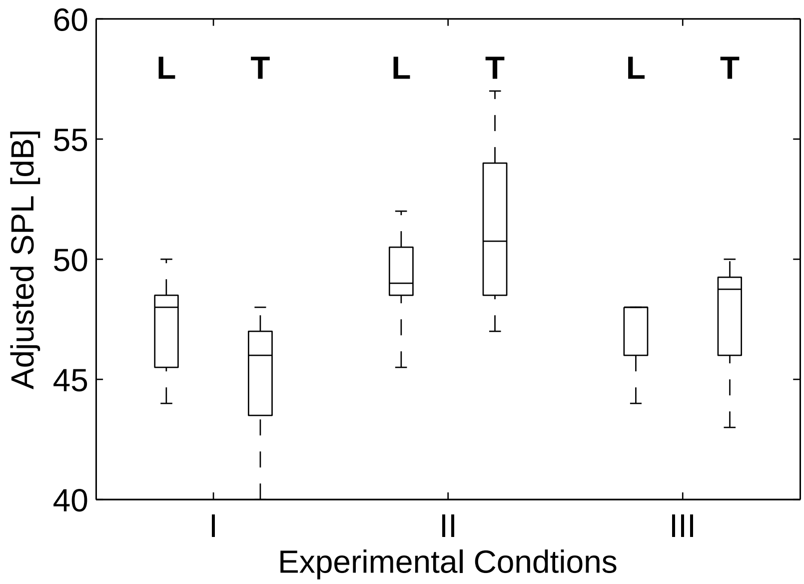

Figure 3.3: *Comparison of the adjusted SPL value between the task to adjust the magnitude of tonal contents (***T***) and the task to adjust the loudness (***L***) of the tonal part within each condition (refer to Fig. 3.1, level range X_{45dB}).*

the single tone, in conditions X_{60dB} and X_{45dB}, the tonal energy seems to be added as an increase of roughly 3 dB is reported. Analyzing the experimental conditions II and III, different contributions of both tones are obtained, and in the case of X_{45dB} are significantly different ($t(7) = 3.20$, $p < 0.05$).

Table 3.2: *Mean of the adjusted SPLs from Table 3.1. None of the differences between the adjustment of the tonal content magnitude and the loudness adjustment are significantly different from zero (t-test, $p = 0.01$, Bonferroni-adjusted).*

factors		tonal content SPL [dB]			loudness SPL [dB]		
level	exp.	mean	n	SE	mean	n	SE
X_{60dB}	I	58.56	8	0.97	63.88	8	0.17
	II	68.06	8	0.75	64.00	8	0.25
	III	62.88	8	0.45	61.75	8	0.46
X_{45dB}	I	45.13	8	0.33	47.25	8	0.27
	II	51.31	8	0.45	49.19	8	0.25
	III	47.63	8	0.31	47.00	8	0.20

In any loudness model (Moore et al., 1997; Fastl & Zwicker, 2007), the energy is summed within one critical band. As the current stimuli are well within one critical band, the results from condition I are expected, while the asymmetry in matching in condition II and III is not supported by these loudness models.

3.2 Modulation adjustment

To further investigate the tone-tone interaction further, a modulation adjustment experiment was conducted. The main idea is that a modulator either higher or lower in frequency than the test tone induces a modulation onto the test tone, again the modulator and the test tone are well within one critical-band. In this experiment, the induced modulation was adjusted and qualitative reports were gathered from participants.

3.2.1 Method

Stimuli

Similar to the first experiment, this experiment was performed with a white noise background (40 dB SPL, f_{lc} = 200 Hz , f_{hc} = 10 kHz). The stimuli are 10 s long and illustrated in Fig. 3.4. Within each stimulus, there are two sinusoids. The first, the signal frequency f_s, is present the whole time. The second frequency, the modulator, which is either lower f_{ml} or higher f_{mh}, is present for 5 s from $t_1 = 3\ s$ until $t_2 = 8\ s$. The modulation, which first is at the signal frequency, is cross-faded to the upper or lower frequency, and then faded back to the signal frequency. The modulation frequency is 3 Hz. At the modulator, the modulation depth is always at -6 dB ($20 \cdot \log_{10} m$). The modulation has a random start phase at the beginning of the signal, at t_1 and t_2. At the time of the cross-fading, the phase is increased by π. The SPL of the sinusoids is 60 dB. The signal frequencies are $f_s = 700\ Hz$ and $f_s = 1\ kHz$. The frequency separation $\Delta f = f_s - f_{l/u}$ is $\Delta f = \pm 50\ Hz$.

Apparatus

The apparatus was identical to that used in the first experiment.

Procedure

The task of the participant was to adjust the modulation depth in interval Δt_{pre} and Δt_{post} to the modulation depth at the signal frequency f_s in the interval Δt_{test}. At the start of the experiment, the adjustable modulation frequency in Δt_{pre} and Δt_{post} had a modulation depth of $-4\ dB$. The initial step size was 4 dB. This step size was reduced every odd reversal by a factor of 0.5 to a minimum final step size of 1 dB. The PSE is determined by the mean of the last eight reversals.

In the beginning of the experiment, the participants listened to the four types of stimuli: both signal frequencies f_s with either an upper or lower additional modulator frequency. Then, the participants were encouraged to describe what they had

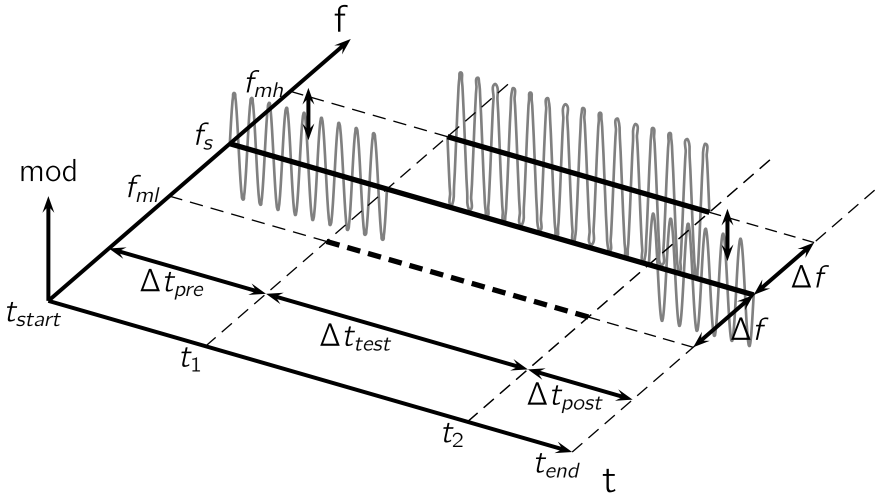

Figure 3.4: *Sketch of the experiment's stimuli. The signal frequency f_s is present in the stimulus from t_{start} until t_{end}. The off-signal frequency, the modulator, either at f_{mh} or f_{ml}, is present from t_1 until t_2. The modulation (depicted as a blue sinusoid) switches from f_s to f_{mh}/f_{ml} with a π-phase shift. The modulation in the t_{pre} and t_{post} is adjusted to match the modulation of the signal frequency f_s*

heard and were asked to report how they perceived the signal frequency f_s when the additional frequency (f_{ml}, f_h) was present. A trial run, using the parameters above, was conducted at one frequency (700 Hz) to help participants to accustom to the experimental procedure. After this, the experiment started as described above; all parameters were fully interleaved. The actual measurement started as soon as the final step size was reached.

Participants

14 participants voluntarily took part in this experiment (4 female, 10 male). None of them had any history of reported hearing loss. The participants' age ranged from 19 to 33 years with a median age of 22 years. Three of the test participants studied music/musicology.

3.2.2 Results

Reliability

The intra-individual reliability can be assessed due to the mean standard deviation across participants and conditions of the modulation depth adjustment. It amounts to $\bar{\sigma}_{700} = 2.2$ dB for the 700 Hz condition and $\bar{\sigma}_{1k} = 2.1$ dB for the 1 kHz condition. This is similar to those of modulation detection (Dau et al., 1997).

However, comparing the adjusted values of the modulation depth, a large standard deviation is obtained across participants (about 8 dB Mod). Therefore, the analysis focuses on pairwise comparisons within participants, i.e., the difference between the condition where the additional tone's frequency is higher and where the additional tone's frequency is lower.

Describing the stimulus - A synopsis

All of the participants where able to describe the stimuli by noticing the additional tone, and the modulation of both tones. However, none of them noticed that there was no modulation of the tone at the signal frequency f_s while the additional tone at $f_{l/h}$ is present. However, after specifically asking them to describe the tone, they noticed that for the condition in which the tone was added below the tone, the tone was perceived as "straight", i.e., not modulated. The music/musicology students noted that at the condition in which the tone was added at f_h, the tone at signal frequency f_s appeared to be "chopped" in the 700 Hz condition.

Quantitative analysis

The differences between the modulation depth adjustment when the tone is added at the lower frequency and at the higher frequency ($\Delta Mod = Mod_{f_h} - Mod_{f_{ml}}$) are shown in Fig. 3.5. With Aures (1985a), the hypothesis for the magnitude of tonal content was that a lower tone would partially mask a higher tone and lead to diminished SNR. The magnitude of tonal content should be fluctuating. This can be rejected as the data in Fig. 3.5 suggest quite the contrary. Taking the results from the first experiment into account, in which the higher tone had a suppressive influence on the loudness of the lower tone, the first and this second experiment conincide. The null hypothesis that the distribution of the median shown in Fig. 3.5 is not larger than zero has to be rejected for the 700 Hz condition ($p > 0.05$). In the 1 kHz condition, the null hypothesis cannot be rejected. Thus, a suppressive influence of the higher frequency tone is, albeit only very weakly, reflected in this experiment.

3.3 General discussion

3.3.1 Tonal content magnitude as partial loudness

The results from the experiments with both forms of instructions, i.e. the introduction of the stimuli by Vormann et al. (1998) including tonal content magnitude adjustment and the partial loudness adjustment of the harmonic part, indicate the following: To a large extent the judgment of the tonal content magnitude as a tone-noise relation can be covered by loudness judgments of the tonal part. Furthermore, result from the post-experimental interview suggest that although a frame of reference for the tonal content magnitude judgment is given, the loudness judgment is far more intuitive to the participants. This might very well be a reason for the lower standard deviations in the loudness adjustment task. Given these results, a feasible

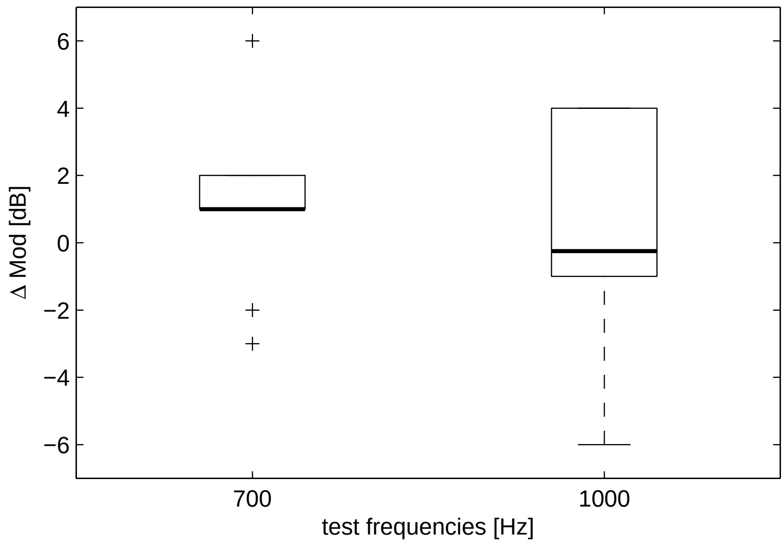

Figure 3.5: *Difference of the adjusted modulation depth between the addition of an modulated tone at higher and lower frequency than the test frequency (f_s = 700/1000 Hz). Outliers higher than 7 dB (3 for each condition) and -7 dB (1 for each condition) are omitted.*

approach to the magnitude of tonal content lies in separating tone and noise parts and then applying algorithms of partial loudness such Moore et al. (1997) do.

Table 3.3: *Calculated level differences between tone and noise of the tone(s) adjusted and the target tones. ΔL_{MT} is the level above calculated masking threshold (Hansen & Weber, 2009). LX is the* level excess *according to Terhardt et al. (1982). In the case of the two tones, the ΔLs are summed energetically. The main difference between the two methods is that the former includes a correction for the masking threshold within noise while the latter includes the reciprocal excitation of both tones. The last calculated measure is partial loudness, i.e., the loudness of the tonal part (Moore et al., 1997). ⋆-ed values are fixed, to-be-adjusted values.*

		ΔL_{MT}		LX		$N_{partial}$	
factors		1 tone	2 tones	1 tone	2 tones	1 tone	2 tones
level	con.	[dB]	[dB]	[dB]	[dB]	[sone]	[sone]
X_{60dB}	I	25.0	⋆ 24.2	22.6	⋆ 17.9	3.2	⋆ 3.3
	II	⋆ 24.1	25.9	⋆ 21.8	21.7	⋆ 2.8	3.7
	III	⋆ 24.1	24.1	⋆ 21.8	17.7	⋆ 2.8	3.5
X_{45dB}	I	8.5	⋆ 9.2	6.0	⋆ 6.6	0.5	⋆ 0.5
	II	⋆ 9.1	11.7	⋆ 6.8	9.2	⋆ 0.5	0.7
	III	⋆ 9.1	9.3	⋆ 6.8	6.8	⋆ 0.5	0.5

3.3.2 Contribution of individual tones to partial loudness

To test the current algorithms for the magnitude of tonal content, two SNR-based measures and the partial loudness were calculated for the adjusted and to-be-adjusted values. The values are reported in Table 3.3. The first value is the SNR taken as a level above calculated threshold ΔL_{MT}, which is basically the procedure published in DIN 45681:2005-03 (2005). As in the data reported by Hansen & Weber (2009), the individual SNR for each tone is added energetically in the two-tone case. The algorithm provides a fairly good estimate for the experimental conditions I and III. In these cases, the difference between the adjusted and to-be-adjusted value is smaller than 1 dB. However, in condition II, in which the lower tone has to be adjusted, the adjusted and the to-be-adjusted value have a difference of about 2 dB.

Can the addition of reciprocal tone masking explain the results? Including the reciprocal tone-on-tone masking, the second approach is the level excess LX by Terhardt et al. (1982), which is again added energetically in the case of multiple tones (Table 3.3). The excitation pattern's flank towards lower frequencies is commonly modeled with an $27 dB/bark$ drop (Terhardt et al., 1982). Therefore, there is no additional decrease in LX for the X_{60dB} condition at the lower frequency as the excitation level of a 750 Hz tone has dropped by 21 dB at 650 Hz, and the excitation level contributed is already well below the noise contribution. For the X_{45dB} condition, there is a decrease of the LX at the lower frequency of about 2 dB in condition I and III, and 1 dB in condition II. The influence of the excitation on the tones at the higher frequency is larger than this. In the X_{60dB} condition, the LX is only reduced in condition II by 1 dB, but in the higher SPL condition X_{45dB}, the LX is reduced by 7, 11, and 4 dB in conditions I, II, and III. Energetically adding the different LX, the data in Table 3.3 shows roughly the same deviations in the X_{45dB} condition for the ΔL_{MT}, while comparing the calculated values for the adjusted and to-be-adjusted values of the X_{60dB} condition yields larger deviations of about 4–5 dB LX. In concluding, the addition of tone-on-tone masking does not improve the weighting for each contributing tone. Firstly, to improve the weighting according to the data, the higher component has to reduce the weighting of the lower. Secondly, as seen in the X_{45dB} condition for the energetically summed LX, the tone-on-tone masking makes it difficult to compare to a single tone.

The last approach listed in Table 3.3 is the calculated partial loudness (Moore et al., 1997). The approach also yields rather large deviations. The condition II in X_{45dB} is overestimated to a large extent; the same holds for the X_{45dB} conditions II and III. As noted in Sec. 3.1.2, the energy of both components falls into the same critical band, thus partial loudness cannot account for adjusted differences in conditions II and III.

The modulation experiment provides further data on the possible contribution of individual tones to the loudness of the harmonic part. Again, contrary to the upward spread of masking, there is no modulation induced by the lower tone. However, if the modulator is higher the signal tone at the frequency f_s is reported to be modulated or interrupted. This is weakly reflected in the quantitative analysis of the data.

The data of the two experiments suggest that there is not an equal contribution of the individual tones to the overall loudness of the tones. In a study investigating the contributions of individual tones to the overall of loudness of multi-tone complexes, it was reported that within one critical band equal perceptual weight is assigned (Leibold et al., 2007). However, in their experiment the lowest and highest tones were two of five tones, so the contrast might not be high enough to detect differences in the contribution to the overall loudness. The question how this asymmetry can be explained remains unanswered, and further research addressing this point is necessary before valid assumptions and conclusions can be made.

3.4 Conclusion

The data support the hypothesis that the magnitude of tonal content roots in the partial loudness of the harmonic part. The overall adjustment with both instructions is the same, while the tonal content magnitude adjustments yields a larger inter-participant deviation. How tonal parts, which have different pitch strength, are integrated into this concept remains an open issue.

The analysis of different algorithmic approaches, the energy above individual threshold, the SNR within a critical band modified by excitation patterns, and the calculated partial loudness of the tones result in a non-congruent picture. The total energy above individual threshold L_{MT} provides, with the given stimuli, a good first approach. However, an underestimation is in principal possible, e.g., the threshold is reduced for harmonic tone complexes. The addition of tone-on-tone masking does not seem to improve the adjustment, especially as it jeopardizes the single tone versus two-tone comparison. The partial loudness provides the best match along with the L_{MT} measure. Its main strength lies in an estimation of the magnitude in *sone* and thus providing a measure comparable across various contexts.

The asymmetric adjustment of the level of the tones along with the results of the second experiment, which suggest a larger influence from higher frequency, suggests perceptual weightings which are not reflected in any of the applied algorithmic measures. This opens up a research opportunity for investigating the contribution of each tone to the overall loudness.

4 Semantic evaluations in Japan, France, and Germany – A cross-cultural comparison†

Knowledge of cultural differences is a valuable resource, be it for economic or social reasons. Differences between cultures have their very start on perceptual dimensions, and when it comes to auditory perception, comparing cultures on percipient and connotative factors in sound quality investigations can lead to significant input on how to optimize product sound design for a heterogeneous and international market. This is, for instance, true for the automotive industry (Hussain et al., 1998), but it also applies for products such as household appliances, multimedia technology, luxury articles, etc. (Guski, 1997; Blauert & Jekosch, 1997). Moreover, investigating how different cultures perceive sound is essential with respect to noise metrics, which are used to establish international noise evaluation standards of high objectivity. These standards recommend reporting and evaluating prominent discrete tones as well as impulsive aspects for a more in-depth noise characterization (e.g. ISO 7779:1999, 1999). But are, for instance, particular sounds evaluated equally across cultures? Sound evaluations are commonly carried out with sound-describing words, or adjectives. They are the basis for measuring perceptual and connotative factors in human beings, i.e., the affective meaning people attribute to objects or stimuli (Osgood et al., 1957). The comprehension and comparability of such adjectives are crucial for assessing product (sound) quality as well as noise annoyance and related reactions to noise.

In the past, several studies on cross-cultural psychoacoustics have been conducted which deal with differences in noise perception (Kuwano et al., 1986; Namba et al., 1991a; Schick & Hoege, 1996) as well as with the sound character of products and musical excerpts (Iwamiya & Zhan, 1997; Kuwano et al., 2006, 2007). In starting with noise issues, Kuwano et al. (1986) performed a study on neighborhood noise in which they evaluated the semantic profiles of loudness, annoyance, and noisiness concepts in three countries. They found that the profiles are stable over years and seem to be equal with one notable difference: the Japanese and German concepts of loudness are rather affectively neutral compared to the English concept. Namba et al. (1991a) investigated the verbal expressions "loud", "noisy" and "annoying" in several countries such as Japan and the USA with the method of selective description,

†Published as: *H. Hansen & R. Weber, Semantic evaluations of noise with tonal components in Japan, France, and Germany - A cross-cultural comparison, J Acoust Soc Am, 2009, 125, 850-862.*

yielding different usages in each country. Schick & Hoege (1996) not only reviewed studies on neighborhood noise but specifically addressed problems commonly found in cross-cultural research such as stimuli selection. Focusing on differences within factorial dimensions of a semantic differential, these researchers found that the German sub-sample showed a more "negatively tuned judgmental structure" (p. 311) while the Japanese participants exhibited a more neutral structure.

To follow with examples dealing with sound character of products and musical excerpts, Iwamiya & Zhan (1997) studied the difference between Japanese and Chinese college students evaluating musical excerpts. Obtaining the factors sharpness, cleanness, and potency for their semantic differential, the Japanese students used the item "pleasant" independently from the three factors. They noted that the same literal meaning does not result in similar usage in the auditory domain. Kuwano et al. (2006) reported data from a cross-cultural car door sound evaluation. In congruence with the previously described findings, their semantic differential revealed a three-dimensional factorial space in both groups. Differences in the usage of adjectives were observed with regards to the adjectives noisy and powerful. In an experiment implementing the semantic differential to evaluate auditory warning signals, Kuwano et al. (2007) found similar results in the USA, Germany and Japan.[1]

A three-factor solution for the evaluation of acoustical stimuli is prevalent in most studies. Similarly, a three-factor solution has been reported in several countries regarding the affective meaning of concepts (Osgood et al., 1975), and they have been labeled, evaluation, power, and activity (EPA-structure). The EPA-structure can be related to the three-factor structure obtained by Namba et al. (1992) in a study to judge artificial sounds: pleasant, power, and metallic/timbre. The evaluation factor corresponds to the pleasant factor, while the power factors correspond directly to each other. The metallic/timbre factor corresponds to the activity factor, because it describes modal features of a sound. In summarizing the studies mentioned above, major differences were obtained mainly with regards to the concept of loudness and noisiness. Also, all of the studies mentioned used culture as an independent variable.[2] Issues of cultural bias and equivalence, which describe the comparability of data across cultures, have yet to be explicitly included in methodological paradigms, although they have been discussed (e.g. Schick & Hoege, 1996). This study incorporates them and they will be introduced in more detail in Sec. 4.1.1.

In the current study, the effect of tonal components in noise on the sound character is investigated, i.e., annoyance, loudness and timbre descriptions in Japan, France and Germany. This class of sounds, i.e., broadband noise containing prominent tones, is widely encountered in the environment, e.g. machine noise, noise in passenger

[1] Moreover, there are studies concerning musical pitch and how intervals, scales, and tuning are perceived in different cultures. Although they offer valuable insight, studies on music cognition are only marginally comparable to the issues in this study, because artificial and stationary stimuli were investigated. For an overview of cross-cultural studies on musical pitch, see Stevens (2004).

[2] The authors are aware that "culture" and "country" are not necessarily the same as Hofstede (1991) pointed out. He admitted, however, that pragmatic reasons often enforce such an operationalization.

compartments of trains and cars, wind turbine noise, and fan noise in electrical appliances. Several national and international norms on noise emissions accommodate the fact that prominent tonal components have a large influence on sound character. The international standard advises to report these components separately (ISO 7779:1999, 1999) in order to describe sound character more exhaustively or an additional penalty is added to the physical level for such sounds (DIN 45681:2005-03, 2005). In order to investigate this class of sounds and to systematize the effect of tonal components on sound character, a set of artificial sounds was generated for the current study. They were used to investigate their effect on sound character, i.e., annoyance, loudness and timbre descriptions using a semantic differential which was carefully translated into every tested culture's language. Schick & Hoege (1996) argued that artificial stimuli in a cross-cultural context with a non-recognizable sound source have two major advantages. First, the listener does not evaluate the source they attribute the sound to. Secondly, they are not attributed to different sources depending on culture. The stimuli are comprised of tonal components, one or two sine signals, which were added to Brownian noise (power density $\propto 1/f^2$). The noise was chosen due to its relatively low overall sharpness. The varying tonal content was implemented by varying the signal level of the sine(s), i.e., the level above masked threshold. The variation of a second parameter, the frequency of the second sine, allowed for the investigation of various ways to integrate tonal components in noise into a signal parameter, which is not covered by the current national and international standards (ISO 7779:1999, 1999; DIN 45681:2005-03, 2005). The stimuli therefore range from broadband noise with barely audible to highly prominent, multiple tonal components.

To be able to cope with issues of cultural bias and equivalence in the current study, a paradigm developed in cross-cultural psychology by v.d.Vijver & Leung (1997) was introduced. A semantic differential, carefully translated into every tested culture's language, was used to explore the perceptual and connotative factors, i.e., the affective meaning, cross-culturally. Then, the method of v.d.Vijver & Leung (1997) was applied to identify bias and to determine the level of equivalence reached in the present investigation.

The current study was designed to research two main goals. The first aim is to test whether the three-factor solution for the affective meaning of sounds can be generalized cross-culturally. Given this basis, an analysis can subsequently be performed in order to relate the magnitude of possible cross-cultural differences on certain scales to the magnitude of bias. This issue is linked to the second goal of the study: the application of acoustical sound descriptors to noise with prominent tones in a cross-cultural context.

4.1 Method

4.1.1 Methodological Background

Using a semantic differential, carefully translated into the respective mother languages, participants from Japan, France and Germany judged the stimuli, consisting of tones in noise. A principal components analysis (PCA) was performed on the obtained semantic differentials and the paradigm reported by v.d.Vijver & Leung (1997) served as a guideline to compare the cross-cultural data from the listening tests. This paradigm is distinguished by two major, opposing concepts: equivalence and bias of the experimental setting.[3] "Data are equivalent when an observed cross-cultural difference on a measurement scale is matched by a corresponding difference on the comparison scale" (Poortinga, 1989, p. 738). Equivalence is jeopardized by various forms of bias, and "scores are equivalent when they are unbiased" (v.d.Vijver & Leung, 1997, p. 7).

The concept of equivalence is characterized by different levels in ascending order: construct equivalence, measurement unit equivalence and scalar equivalence. Analyzing equivalence is hierarchical, and the different levels must be established on the basis of the preceding ones. Optimally, scalar equivalence would be reached for the data to be compared. Once equivalence of the data has been proven in a cross cultural comparison, the remaining differences can be characterized as valid cross-cultural differences. Hence the decisive reason to determine different levels of equivalence is the ability to identify valid cross-cultural differences thereafter. Proven bias destructs equivalence and bias is classified according to its influence on the level of equivalence reached: construct, method, and item bias (Poortinga, 1989, Table 2, p. 745). In the following, the different levels of equivalence along with their vulnerability to bias types are described.

Construct equivalence

Construct equivalence is achieved when the construct measured is identical across the cultures investigated. Deviating from construct equivalence leads to construct bias, e.g. the incomplete overlap of construct definitions, such as the construct *intelligence* in various cultures. The three independent factors obtained in sound evaluation studies, i.e., pleasant, metallic/timbre, and power (e.g. Solomon, 1958; Namba et al., 1992), represent the construct investigated in this study. To test

[3]"Bias refers to the presence of nuisance factors in cross-cultural research. Three types of bias are distinguished, depending on whether the nuisance factor is located at the level of the construct (construct bias), the measurement instrument as a whole (method bias) or the items (item bias or differential item functioning). Equivalence refers to the measurement level characteristics that apply to cross-cultural score comparisons; three types of equivalence are defined: construct (identity of constructs across cultures), measurement unit (identity of measurement unit), and scalar equivalence (identity of measurement unit and scale origin). Bias often jeopardizes equivalence." (v.d. Vijver, 1998, p. 41)

construct equivalence, the semantic differentials for all three countries were analyzed separately by principal component analysis in order to reveal these independent factors. An initial comparison of the factorial structure should reveal any major differences in the construct formation. If the number of perceptual space dimensions is equal in all three cultural sub-samples, target rotation toward an arbitrary factor solution will be performed. A comparison between the rotated and the target structure can show whether and how the perceptual dimensions differ (Watkins, 1989). The target rotation is a necessary step before comparing the perceptual dimensions, because they are not necessarily congruent after factor analysis. The factors are to be named equally for all three cultural groups investigated after determining a common set of marker adjective items, i.e., the adjective items which represent the factor in the best way.

To quantify the relational agreement between the target and the rotated factor solutions, two coefficients are used, Tucker's ϕ and Pearson's correlation coefficient r_{xy}.[4] Tucker's ϕ, the congruence coefficient, while sensitive to a additive constant, is insensitive toward multiplications, i.e., if all factor loadings of one group have a proportional relationship to the factor loadings of the other group, no difference will be detected and Tucker's ϕ is 1.

For Tucker's ϕ no sampling distribution is known (Korth & Tucker, 1975). Nevertheless, the "significance level" can be estimated using Monte Carlo simulations. Korth (1978) reports a "significance level" ($\alpha = 0.05$) of 0.93 (4 factors/10 variables). Korth (1978) stated that Tucker's ϕ is lower for less factors and more variables. A coefficient of 0.93 can be viewed as a conservative estimation for the "significance level" in this study.

The correlation coefficient r_{xy} indicates the strength and direction of a linear relationship between the factors, i.e., it is influenced neither by addition nor multiplication. In this way, Tucker's ϕ and the correlation coefficient quantitatively indicate the construct equivalence.

Another type of bias affecting construct equivalence, the fundamental level of equivalence, is method bias. It involves issues such as differential social desirability, stimulus familiarity, lack of sample comparability, etc., and therewith all issues concerning experimental methods. The different forms of method bias will be discussed in the respective method section.

Measurement unit equivalence and scalar equivalence

After focusing on the perceptual and connotative factors, the analysis will concentrate on the detection of valid cross-cultural differences based on the adjective items in the semantic differential. Therefore, the concepts of item bias and scalar equivalence, as opposites, will be explained in detail before the analysis will be introduced.

With *cross-cultural differences* we imply differences in item means across the three cultural sub-samples tested. Therefore, measurement unit and scalar equivalence

---

[4] Tucker's ϕ: $\phi_{xy} = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum_i y_i^2}}$

have to be established beforehand. After determining construct bias and ruling out method bias, item bias is the major bias source left. In other words, based on equivalent constructs, it will be tested whether the items are used in the same way.

v.d.Vijver & Leung (1997) defined how to determine whether a measure is free of item bias: "... persons with an equal standing to the theoretical construct underlying the instrument should have the same expected score on the item, irrespective of group membership" (p. 69). This, however, does not imply equality of item means across cultures. Instead, on an individual level, this implies that people "with an equal standing" regarding a particular item just score the same. Therefore, a true difference between groups for an item on a bias free scale is the true difference in means for that particular item. Consequently, a bias analysis of the adjective items should precede the analysis of the "impact", because the items used in the cross-cultural context are not necessarily bias-free.

Here is an example to facilitate understanding. It is possible that the groups tested in this study use the adjective item pleasant/unpleasant in different ways. Imagine German participants express a high degree of pleasantness determined by a general high score on all adjectives associated with the pleasant factor. Yet concerning the adjective pair pleasing/unpleasing, they score a medium value, i.e., a 3 on a 1-7 rating scale. Furthermore, imagine Japanese participants also showing a high degree of pleasantness via the overall value, yet also exhibiting a high score on the pleasant/unpleasant item. In this case, comparing the scales at face value for this item would be misleading, because the two groups have an "equal standing" although the Japanese group uses the pleasant/unpleasant item differently, i.e., by rating it higher than the German group.

In the following, an analysis capable of unraveling the depicted types of item biases will be described. The adjective items associated with the factors determined via PCA will be classified according to their strength in representing these factors. The corrected discriminatory power r or item-to-dimension total correlation will be calculated for each item. To generate the total score, all items of a specific dimension are summed, while the item under scrutiny is left out. For each factor, the corrected discriminatory power r is calculated by correlating the corrected total score with the item scores. Discriminatory power r of $(0.5 > r > 0.3)$ is classified as "medium" and $(r \geq 0.5)$ as "high" (e.g. Bortz & Doering, 2002). The homogeneity of the items representing a particular factor will be tested by calculating Cronbach's α in order to qualify the item composition. Values $\alpha > 0.7$ denote a relatively high overall consistency (ibid.). If a strong correlation between the factor and the adjective items is found, i.e., marker items have high internal consistency and are appropriate manifestations of the underlying factors, further analyses will be conducted to test cultural differences between the scores of single adjective items.

To identify item bias in numerical scores, the "equal standing" of participants regarding one perceptual dimension must be operationalized. This is done by using the total factor score as an "equal standing" indicator. The total factor score is the sum of all respective marker items. For each stimulus, each participant as well as each factor, one total factor score is calculated. The range of total factor scores are then divided into sub-ranges containing approximately the same amount of total factor scores for each factor. These sub-ranges are called score levels. They are indicators for "equal standing" of participants for a stimulus with respect to one particular factor. Therefore, two participants show an "equal standing" regarding one stimulus with reference to the perceptual factor if the factor levels are equal in both cases. Item biases can then be detected by running MANOVAS on item means with score levels and culture as the independent variables. The paradigm reported by v.d.Vijver & Leung (1997) is applied with these procedures, and culture as a main effect as well as the interaction between culture and score level indicate cross-cultural differences. On the one hand, culture as a main effect represents whether an item is generally used differently, i.e., higher or lower (uniform bias). On the other hand, the interaction effect indicates a change in the usage or interpretation of an item across score levels (non-uniform bias). Therefore, assuming there is no construct and method bias, measurement unit equivalence only results when non-uniform bias is not found between cultures, because the measurement unit is not affected by a constant offset on an item. However, scalar equivalence, i.e., full score comparability, is only reached if the uniform bias can be excluded, as only then a common scale origin can be assumed (Poortinga, 1989, Table 2, p. 745).

Testing differences between means: The "impact" of culture

If construct equivalence is confirmed and no item bias is found, an additional MANOVA can reveal the influence of culture on the unbiased adjective items. Furthermore, obtained differences on biased items will be related to the magnitude of bias discovered. After excluding all possible bias, the identifying valid cross-cultural differences will be the concluding step in analyzing differences in this cross-cultural study (Berry et al., 2002).

Correlation of perceptual dimensions with acoustical sound descriptors

After identifying different perceptual dimensions, it is of special interest to investigate how they are related to calculated acoustic and psychoacoustic parameters. In this study, a measure describing the tonal contents ΔL, a SPL parameter $L_{Aeq,T}$, and the sharpness S were used. ΔL, the S/N ratio above masked threshold, was investigated, because a S/N approach is suggested by various international and national norms to describe the tonal contents (ISO 7779:1999, 1999; DIN 45681:2005-03, 2005). Furthermore, the $L_{Aeq,T}$, a common level measure in noise evaluations (Namba & Kuwano, 1984; Marquis-Favre et al., 2005), will be correlated with the perceptual dimensions. Sharpness S (von Bismarck, 1984) has been identified as a major source of auditory unpleasantness (Zwicker & Fastl, 1999; Zimmer et al., 2004).[5] Relating and describing the factors with these signal parameters is expected to allow an improved understanding and interpretation of the perceptual dimensions.

4.1.2 Stimuli

The stimuli were various sinusoidal sounds at different frequencies with Brownian noise (red noise, $L_{Aeq,6s}$ = 44.5 dB(A)) added in order to investigate the the effect of tonal components on the perceptual and connotative factors. Brownian noise has lower sharpness resulting in a more pleasant sensation compared with other broadband noises, e.g. uniform masking noise or pink noise (Zwicker & Fastl, 1999). As tonal components often add to annoyance, Brownian noise serves as a adequate starting level.

The first set is comprised of noise with a single component added (f_1 = 500 Hz) . The exact stimuli configurations are shown in Table 4.1. Their salience is described by the measure ΔL. This is the ratio between the sine level and the level of the respective critical band corrected by masking threshold according to DIN 45681:2005-03 (2005), i.e., the level above masking threshold, which is based on Zwicker & Feldtkeller (1967, chapter 18). They use the Békésy method of audiometry, i.e., the calculated threshold estimates the 50 % point on the psychometric function. As the goal is to investigate the effect of the perceptual S/N ratio on sound character, a range of -7–23 dB was chosen for ΔL. 0 dB marks the threshold, and -7 dB can therefore be considered barely audible because it is at the very end of the psychometric function. 13 and 23 dB mark clearly prominent tones that were included to investigate the effect large level changes of dominant tones have on sound character.

Apart from the perceptual S/N ratio, the aim is to investigate the effect of multiple tones on sound character. Therefore, three additional frequencies were chosen. ISO 7779:1999 (1999) gives a proximity limit Δf_{prox} in which the energy should be integrated into a single tone. As in the first set, f_1 was set to 500 Hz. The frequency of the additional tone f_2 is varied. f_2 was set to 530 Hz, 600 Hz or 1000 Hz. 530 Hz lies within the proximity limit ($\Delta f_{prox,}$(500 Hz) = 34 Hz, (ISO 7779:1999, 1999)). 600 Hz lies within one critical band (Zwicker & Fastl, 1999), while 1000 Hz has a harmonic relation and lies outside the critical band. Every frequency condition was presented with six different sine levels. f_1 and f_2 were equal in SPL. The SPL was also equal across frequency conditions, i.e., the level of $f_2(S_8)$ equaled $f_2(S_{14})$. The lowest SPL was chosen to match the signal level above calculated masked threshold of the lowest SPL in the first set (S_{20}, f_2 = -7 dB). The signal levels covered the whole range of perception of prominent tones. The stimuli configurations are summarized in Table 4.2.

[5] The sharpness S is defined as $S = 0.11 \cdot \frac{\int_0^{24Bark} N'g(z)zdz}{\int_0^{24Bark} N'g(z)dz} acum$. The denominator is the total loudness N, while N' is the specific loudness of a critical band (von Bismarck, 1984). The weighting function $g(z)$ is 1 and increases with $z > 16$ $Bark$. Therefore, the sharpness is the first moment of the weighted critical-band rate distribution of the specific loudness (Zwicker & Fastl, 1999).

The stimuli were generated digitally using Matlab (Version 6.0, Mathworks Inc.) with a sampling frequency of 44100 Hz. Every stimulus lasted 6 s each including a 10 ms fade-in and fade-out. Both the noise and the sines had a common onset and offset. The stimuli levels were described by $L_{Aeq,6s}$ measured with a sound level meter (Ono Sokki ONO-LA5110) directly reproduced from a Sony Digital Audio Tape-Coder (TCD-D8). The sharpness S was calculated according to von Bismarck (1984).

Table 4.1: *Stimuli with one sine (frequency f_1 = 500 Hz) with varying signal-to-noise ratios (SNR) embedded in Brownian noise (red noise).*

S_i	f_1 (Hz)	ΔL[6] (dB)	$L_{Aeq,6s}$[7] (dB)	S[8] (acum)
1	-	-	44.5	1.25
2	500	-7	44.5	1.25
3	500	-3	44.7	1.25
4	500	1	44.7	1.24
5	500	5	44.9	1.22
6	500	13	46.7	1.17
7	500	23	53.2	1.06

4.1.3 Semantic Differential

The selection of the adjective compilation is based on semantic differentials reported by Namba et al. (1991a, 1992) and concepts by Schick & Hoege (1996). This enabled an a priori sampling along the dimensions pleasant, metallic, and power, which guided adjective selection. The choice was carried out along these dimensions as the adjective choice alone is a source of method bias. The inclusion of Japanese semantic differentials mitigated this issue (Namba et al., 1992; Kuwano et al., 1994). The pleasant dimension was sampled a priori by adjectives such as *beautiful/annoying/clear* and *pleasant/pleasing*. The difference between pleasant and pleasing seems minimal in English, because they represent an adjective and a verbal adjective from the same word stem. In the languages used in this study, however, they stem from different words. The metallic/timbre factor includes, apart from *metallic/shrill/harmonic/hard* (Namba et al., 1992), adjectives which refer directly to the sound character such as *tonal*. The power factor is represented a priori by the adjectives *loud/powerful/noisy*. According to Schick & Hoege (1996) and Namba et al. (1991a), these adjectives cannot be grouped into a neutral power category in every culture, but in an a priori sampling this coarse approach not only suffices, it is even a requirement for reducing method bias. Finally, the semantic differential chosen to evaluate sound perception consists of 14 adjective pairs each separated by a 7-step rating scale (Osgood et al., 1957). The adjective pairs are presented in

Table 4.2: *Stimuli with two sine signals with varying signal-to-noise ratios (SNR) embedded in Brownian noise (red noise). Both sines have the same SPL.*

S_i	f_1 (Hz)	f_2 (Hz)	ΔL_1[9] (dB)	ΔL_2[10] (dB)	$L_{Aeq,6s}$ (dB)	ΔL_{sum}[11] (dB)	S[12] (acum)
8	500	530	-11	-10.5	44.5	-7.8	1.25
9	500	530	-5	-4.5	44.6	-1.8	1.23
10	500	530	1	1.5	44.8	4.2	1.22
11	500	530	7	7.5	45.7	10.2	1.19
12	500	530	13	13.5	48.2	16.2	1.14
13	500	530	19	19.5	52.3	22.2	1.07
14	500	600	-11	-10	44.5	-7.3	1.25
15	500	600	-5	-4	44.6	-1.3	1.24
16	500	600	1	2	44.8	4.7	1.23
17	500	600	7	8	45.9	10.7	1.19
18	500	600	13	14	48.3	16.7	1.12
19	500	600	19	20	52.6	22.7	1.05
20	500	1000	-11	-7	44.5	-4.5	1.25
21	500	1000	-5	-1	44.6	0.5	1.23
22	500	1000	1	5	44.8	6.5	1.22
23	500	1000	7	11	45.7	12.5	1.19
24	500	1000	13	17	48.1	18.5	1.13
25	500	1000	19	23	52.2	24.5	1.05

Table 4.3 in Japanese, French and German, and an English translation was added for comprehension. The German version served as a reference for all other versions.

They were translated into Japanese, French and German by native speaker professionals with experience in the semantic characterization of sounds. These translators were chosen according to individual proficiency. The translations were made with many precautions. Appropriate choices for translated adjectives were made based on their semantic and not their literal meaning in the target language.

Table 4.3: *Adjective pairs and their translation as used in the experiments. (All Japanese transliterations are written according to the Hepburn system)*

English	Japanese	French	German
powerful/weak	hakuryo no aru/monotarinai	puissant/faible	kräftig/schwach
shrill/calm	kantakai/ochitsu ita	strident/calme	schrill/ruhig
clear/vague	hakkiri shita/bonyari shita	clair/flou	klar/vage
pure/impure	sunda/nigotta	pur/impur	rein/unrein
pleasant/unpleansant	kokosoyoi/fukanai	agréable/désagréable	angenehm/unangenehm
metallic/deep	kinsokuseino/fukami no aru	métallique/profond	metallisch/dumpf
noisy/quiet	yakamashii/shizukana	bruyant/doux	lärmend/still
hard/soft	katai/yawarakai	dur/mou	hart/weich
harmonic/discordant	chouwanotoreta/fuchouwana	harmonique/inharmonique	harmonisch/disharmonisch
tonal/not-tonal	pitchu/pitchu ga hakkirishinai	tonal/pas tonal	tonhaltig/nicht tonhaltig
loud/soft	ooki/chiisai	fort/doux	laut/leise
annoying/not-annoying	urusai/urusaikunai	gênant/pas gênant	lästig/nicht lästig
pleasing/unpleasing	monomashii/monomashikunai	plaisant/déplaisant	gefällig/ungefällig
beautiful/ugly	utsukushii/kitanai	beau/laid	schön/hässlich

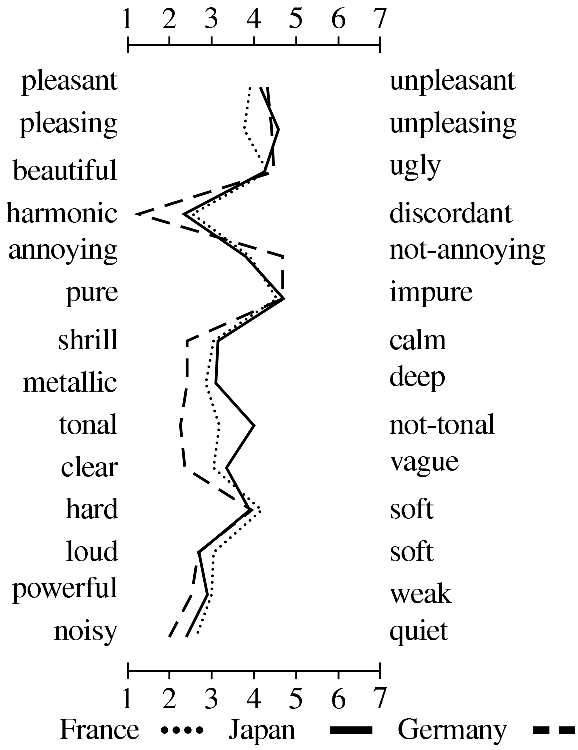

Figure 4.1: *Example of the semantic differential used in the experiment. Mean values of S_{17} are shown for every culture.*

4.1.4 Participants

Although a source of method bias, cross-cultural studies commonly treat culture as an independent variable for reasons of practicability (v.d.Vijver & Leung, 1997, pp. 2-3). Besides culture, socio-demographic and context variables were assessed: age, gender, experience in the field of acoustics, and educational status. 11 Japanese men and 9 Japanese women (average age: $22.9a \pm 1.7a$), 15 French men and 8 French women (average age: $20.5a \pm 1.8a$) as well as 12 German men and 8 German women (average age: $23.7a \pm 2.6a$) with normal hearing abilities participated in this study. Being *Japanese*, *French*, or *German* means participants have the respective mother tongue and are autochthon to the respective country. Furthermore, all were college students at the time of testing and approximately half of them majored in acoustics or related fields such as environmental engineering. As the cultures were matched according to these covariates, method bias was reduced at this stage.

The experiment was conducted in Japan from Nov. 2002 till Jan. 2003, in Germany from Nov. 2003 till Dec. 2003, and in France from Jun. till Oct. 2005.

4.1.5 Apparatus

As a necessary prerequisite for methodical consistency, stimuli presentation had to be equal in the different laboratories used. To guarantee the identical presentation of sound stimuli in all countries, the relevant equipment used was always the same. Failing to do so would have lead to major method bias. The stimuli were presented diotically via the same headphones (STAX SR-Lambda Pro, HA I.1) using the same amplifier (HEAD acoustics HPSIII.2) in sound-proof rooms at Osaka University, Japan, INSA in Lyon, France, and Oldenburg University, Germany. The transfer function of the headphones, measured with an artificial head, had a flat characteristic ($\pm 3dB$) between 200–3000 Hz.

In Japan, the wav-files were recorded on DAT (Sony DTC-ZE700) via a USB-interface (Roland ED UA-30) and were reproduced with a DAT-Recorder (Pioneer D-05). In Germany and France the sounds were recorded on CD-ROM using a CD-player (Philips CD618) for reproduction. The differences in the experimental set-up relating to sound file recording and storage did not cause any significant changes in the stimulus presentation.

4.1.6 Procedure

The instructions and semantic differentials were handed out to participants as paper copies in their respective mother tongue. The participants were instructed to judge each stimulus as a whole according to the adjective pairs on the semantic differential. In order to get accustomed to the stimuli and to establish an adequate frame of reference, each participant first judged 6 test stimuli (S_2, S_6, S_{11}, S_{15}, S_{19}, S_{22}), covering the perceptual range of all stimuli used. This phase also served for clarifying the usage of the adjective pairs in the context of this study. For the main experiment, both the stimulus presentation order and the semantic differential item order were randomized. The 25 stimuli were randomized in three different orders, and each participant listened to one order. Each participant evaluated the set with one of three randomized semantic differentials. This twofold randomization balances position effects, e.g. due to short-term memory. Presenting a stimulus consisted of repeating it three times, and repetitions were separated by a 4 s pause. The participants were allowed to start evaluating the stimulus as soon as its first presentation was over. Only after the participant finished evaluating a stimulus with the semantic differential, the next stimulus was presented. The entire experiment lasted approximately 45 min. At the end of testing, three selected stimuli were replayed and judged again for a supplemental reliability analysis with the French and German samples. The stimuli chosen for repetition (S_{21}, S_1, S_{25}) covered the entire evaluation range: Two extreme and one mid-range stimuli were used.

Procedural differences can lead to method bias. In using the steps described here, the experimental procedure was equal in all cultures. The retest is one exception, but it did not lead to any bias because it was conducted after the experiment. By instructing participants to read the translated introduction without conveying additional information, the experimenter reduced the eminent sources of method bias at this stage.

4.2 Results

The results are based on data obtained with the semantic differential described in Table 4.3 in Japan, France, and Germany. To test reliability, a Pearson correlation was calculated for three stimuli which were evaluated twice by each participant in the French and German samples: 19 out of 20 German and 23 out of 24 French participants were considered reliable judges ($r_{crit,0.01} = 0.40$). Participants who inconsistently judged repeated stimuli were excluded from further analysis. In Japan, no stimuli were repeated after initial testing, rendering a reliability analysis for this sample impossible. This may lead to additional unexplained variance in the Japanese data. Considering the solid reliabilities for the German and French samples as well as several successful semantic differential studies with Japanese participants by (e.g. Kuwano et al., 1994, 2006), the fact that reliability cannot be calculated for the Japanese sample in this study may not jeopardize overall interpretations of the results.

The results section covers four issues: 4.2.1) The equivalence of perceptual dimensions, 4.2.2) Item bias analysis, 4.2.3) Differences in means: the "impact" of culture , and lastly 4.2.4) acoustical sound descriptors in a cross-cultural research.

4.2.1 The equivalence of perceptual dimensions

To analyze the 14 adjective pairs in the semantic differential, factor analyses were computed separately for the French, Japanese, and German samples. In all cases, three factors (Kaiser-Guttmann Criterion, $EV > 1$) were extracted using the Principle Component Analysis (PCA) with Varimax rotation. The Kaiser-Meyer-Olkin measure of sampling adequacy equals 0.86 in the Japanese sample, 0.88 in the French sample and 0.91 in the German sample. As such, the factor analysis was properly applied (Cureton & D'Agostino, 1983, p. 389).

Differences between experimental samples are generally overestimated due to the various orientations of rotated solutions in factorial space. To avoid overestimation, the factor solutions for the French and the German samples were rotated toward the factor solution for the Japanese sample. The rotation procedure was chosen to achieve co-linearity of each the three dimensions, facilitating a comparison between factor loadings for each dimension. The target rotated factor solutions for the French sample and the German sample along with the Japanese Varimax-rotated solution are presented in Figure 4.2.

For each factor, the following marker adjective pairs were identified as common descriptors for all samples:

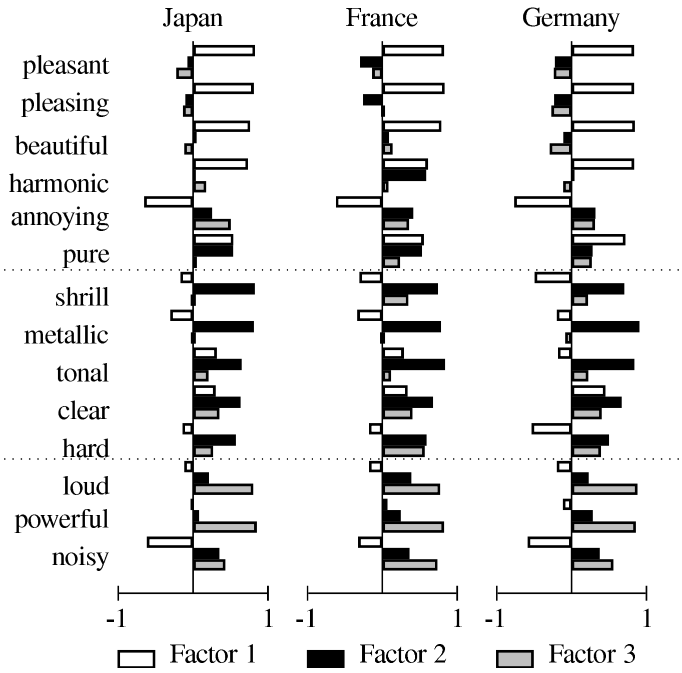

Figure 4.2: *Factor loadings of the Japanese, French, and German sample: Adjective items used in the semantic differential with a 7-step rating scale. The Japanese loadings were rotated with the Varimax rotation. This solution provides the target for the rotation of the French and German sample's factor space.*

Factor I beautiful/ugly, harmonic/discordant, pleasing/unpleasing, pleasant/unpleasant, and not-annoying/annoying

Factor II shrill/calm, clear/vague, metallic/deep, and tonal/not-tonal

Factor III powerful/weak, loud/soft, and noisy/quiet

In considering the semantic quality of these adjectives, the factors were named: pleasant (I), metallic (II), and power (III). The same structure has been reported in other semantic differential studies from Japan (e.g. Namba et al., 1992), which is why the terminology is adopted here.

The similarities and differences between the rotated factor loadings were quantified by computing congruency measures such as Tucker's ϕ and the correlation coefficients between the three matrices. These are relatively high: $\phi \geq 0.91$ and $r \geq 0.85$. The solutions for the Japanese and French samples concerning the pleasant factor loadings are almost proportional to each other ($\phi = 0.98$). This also holds for the comparison between these loadings and the loadings for the German sample's pleasant factor ($\phi = 0.94$). Again, the same holds for metallic factor loadings ($\phi = 0.93$). The French power factor is less noteworthy ($\phi \leq 0.93$) than the Japanese and German power factors. As reported in 4.1.1, $\phi = 0.93$ is a conservative estimation for the significance level. At this value, the power factor loadings can still be considered equivalent.

In commenting two major differences, the adjective pair harmonic/discordant in the French sample does not only load highly on the pleasant factor, but contrary to the other data, it is more closely related to the description of timbre, i.e., the metallic factor (II). Although the translation of this item is considered adequate by scientists with French as their native language, the different use of this item indicates bias. Secondly, the French participants associate the adjective pair noisy/quiet mostly with the powerful factor (III), while in the other groups the pair shows medium loadings on pleasant factor (I) and the powerful factor (III), indicating an evaluative connotation. Moreover, the concept of noisiness combines power and unpleasantness in this case. This may also be due to item bias, meaning that the correctly translated adjective pair is used differently in the French language. The differences measured on the adjective pairs hard/soft and pure/impure were not evaluated due to low communalities in all three groups.

As a major result, the equivalence of all three factors, pleasant, metallic and power was established, even though there are some differences on certain items. This conceptual or construct equivalence is a necessary condition for the following item bias analysis.

4.2.2 Item bias analysis

After exploring the structure of the factorial space, further analyses were conducted to investigate the adjective pairs in more detail. After summarizing the results from the discriminatory power analysis and the internal consistency analysis, i.e., traditional psychometric analyses, the item bias analysis will be reported, followed by a direct comparison of each mean item score between the samples.

As argued in Sec. 4.1.1, the factor-item relation is qualified by computing a discriminatory power analysis and an internal consistency analysis. The former is conducted to inspect how well single adjective pairs represent their associated factor. All characteristic items of the pleasant, metallic, and power factors were tested. Additionally, the items quiet/noisy and pure/impure are included in the analysis for the pleasant factor, because they show high loadings in the German samples and medium loadings in the other samples.

The discriminatory power analysis yielded at least a 'medium' ($0.5 > r > 0.3$) but mostly a 'high' ($r \geq 0.5$) item-to-dimension correlation coefficient. To designate whether an adjective pair represents a factor for all cultures, only items with a 'high' mean item-to-dimension correlation in all cultures and an individual correlation coefficient of at least 'medium' size were selected. Most items are considered to be representative items for their associated factors, because they show high mean discriminatory power. However, the following items had to be excluded: Noisy/quiet did not yield 'medium' discriminatory power in the French sample, while pure/impure showed only 'medium' mean discriminatory power. It turns out that the 12 representative pairs are identical to the 12 marker items identified in Sec. 4.2.1.

The internal consistency (Cronbach's α) is used as a measure to display how consistently the marker items can be aggregated to one scale. This is a necessary procedure for an item bias analysis. All adjective items yielded 'high' internal consistency with their confounding factors pleasant, metallic or power.

As the factor-item relation is within the set limits, the bias was assessed on item level (single adjective pairs). To provide indicators for 'equal standing' (see Sec. 4.1.1), total factor score levels were calculated by summing the scores of all established marker items of each factor. Therefore, for each participant's judgment of a stimulus and for each factor, a total factor score was generated. These total factor scores for each factor were combined in a set comprising the contributions from all three cultures. This set of total factor score levels was subdivided into roughly equally large groups of *score levels*. These roughly equally sized groups of the combined distribution of total factor scores representing a factor were used to derive ten score levels each for the pleasant and metallic factors and five score levels for the power factor. Judgments aggregated on the same score level were assumed to have 'equal standing' regarding the respective sound on the represented factor.[13]

MANOVAs were calculated consecutively for each factor. Score level and culture are the independent variables, the dependent variables are the respective items belonging to a factor ($p = .05$). To quantify the effect size, $\eta^2_{partial}$ was calculated.[17] Table 4.4 shows the results: culture as a main effect (first column) and the interaction effect between culture and score level (second column). These effects can be described as differences between means for the three cultures as measured on a 7-step rating scale. In other words, how large is the bias on significantly biased pairs? Adjective pairs showing bias with an effect size $\eta^2_{partial} \leq 0.6$ have difference between means smaller than 0.5 rating scale divisions between the cultural sub-samples. Examining the item clear/vague in more detail, the Japanese sample shows uniform bias with regard to the French and German samples. At score levels > 2, the French participants used the item about 1 scale division, and the German participants about 0.5 scale divisions lower than the Japanese participants. 'Small' interaction is observed in the German sample. With increasing metallic perception, the German participants cease to describe the stimulus as vague and thus more inline with the Japanese sample, i.e., a non-uniform bias is observed here.

In general, it can be concluded that there are several small biases on the adjective items used. These item scales must be scrutinized while interpreting the results at face value. Nevertheless, several items constituting the pleasant, metallic, and power factors seem to be bias-free, or they show a diminutive bias level and could be used for further level-oriented analysis, e.g. shrill/calm or pleasant/unpleasant (see Table 4.4, $\star$-marked).

[13] The MANOVA is robust against different score level numbers, yielding a similar result to those presented in Table 4.4, as reported by (v.d.Vijver & Leung, 1997). Furthermore, only five score levels were used for the power factor, because a roughly equal number of scores on each level is mandatory.

[17] $\eta^2_{partial} = \frac{SS_i}{SS_i + SS_{error}}$, $SS_i = \sum_m (x_{im} - \bar{x}_i)^2$); i.e., the ratio of the variation induced by a main or interaction effect i (SS_i) to this variation (SS_i) plus the variation left to error (SS_{error}; Cohen, 1973).

Table 4.4: *Effect sizes for the significant ($p = .05$) culture and of culture x score level effects on the item scales constituting the dimensions of perception. The main effect indicates uniform bias, the interaction non-uniform bias.*[15]

		$\eta^2_{partial}$	
Dimension		Culture[16]	Culture × Score Level
Pleasant			
pleasant/unpleasant	⋆	0.02	0.02
harmonic/discordant		0.05	0.02
pleasing/unpleasing	⋆	0.02	0.02
beautiful/ugly	⋆	0.01	0.03
not-annoying/annoying		0.05	0.04
Metallic			
shrill/calm	⋆	-	-
clear/vague		0.07	0.04
metallic/deep	⋆	0.01	0.02
tonal/not-tonal		0.01	0.05
Power			
powerful/weak	⋆	0.01	0.04
noisy/quiet	⋆	0.01	0.02
loud/soft	⋆	0.01	0.03

4.2.3 Differences in means: the "impact" of culture

Apart from investigating the reliability of the semantic differential across cultures, a major goal is to investigate the potential effect of culture, i.e., the 'impact' on the respective adjective items as described in Sec. 4.1.1. Do cultural groups differ in the kinds of judgments they make? Furthermore, do noise stimuli have different effects depending on the cultural group perceiving them?

In order to examine these questions, MANOVAs were calculated with tonal component frequency content f, tonal energy $\overline{\Delta L}$, and culture as independent factors. The S/N-ratio increased in six steps. Stimulus S_1 was excluded because it lacks a tonal component. This analysis yielded inter alia the effects of culture and the interaction between culture and the other independent variables. At this stage, bias-free items must be assumed.

As a result, the tonal frequencies f have an influence on the metallic factor adjectives, which leads to increasing metallic factor scores with growing frequency. $\overline{\Delta L}$ correlates with increasing unpleasant, metallic, and power factor scores. Table 4.5 shows the effect of culture and the interactions between culture and the systematic stimulus variations. The items harmonic/discordant, clear/vague, and tonal/not-tonal have the largest effect sizes. The effects must be analyzed in relation to the findings in 4.2.2, i.e., are the reported changes in means of the same magnitude as the biases?

Table 4.5: *Estimated effect size η^2 for the significant ($p = .05$) main effect culture (cul) and the significant interaction culture x stimulus variation ($f \times cul$, $\overline{\Delta L} \times cul$) on item scales constituting the perceptual factors.*[19]

Adjective Scales	*cul*	$f \times cul$	$\overline{\Delta L} \times cul$
pleasant/unpleasant	0.02	-	0.03
harmonic/discordant	0.02	-	**0.10**
pleasing /unpleasing	0.03	-	0.04
beautiful/ugly	0.01	-	0.05
annoying/not-annoying	0.06	-	0.03
shrill/calm	0.02	-	0.02
clear/vague	**0.06**	-	0.04
metallic/deep	0.01	-	0.02
tona/not-tonal	0.04	-	**0.10**
powerful/weak	-	-	0.04
noisy/quiet	0.04	-	-
loud/soft	0.05	-	0.03

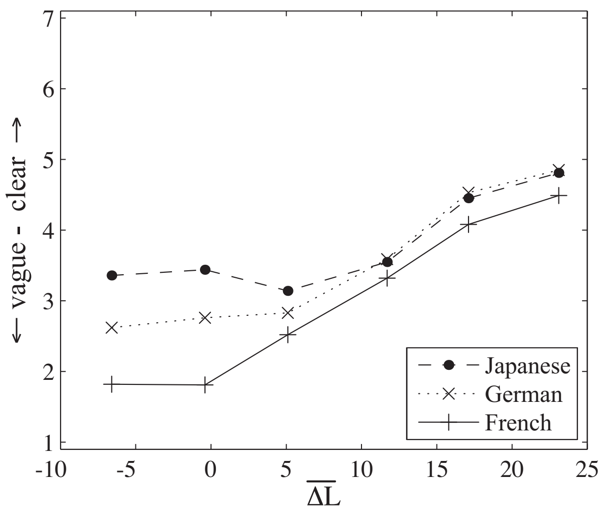

Figure 4.3: *Clear/vague item means for the judgments by the Japanese, French, and German sub-samples versus the $\overline{\Delta L}$. $\overline{\Delta L}$ is the mean value of $\Delta L_{T,sum}/\Delta L$ of the six steps of S/R-ratio increase. The obtained significant difference at lower values of $\overline{\Delta L}$ is mitigated by the fact that the scale is biased. (The standard deviation, $\sigma \approx 1$, was omitted for clarity.)*

Using the example in Fig. 4.3, mean values for the adjective pair clear/vague are plotted against $\overline{\Delta L}$ of stimuli containing two frequencies for each cultural sub-sample at face value. The mean ratings of $\overline{\Delta L}$ = 15 dB and lower must be examined because they have significant differences. Following the concept in 4.1.1, the differences in means cannot be analyzed at face value. The differences in Fig. 4.3 can be explained due to the bias explicitly described in Sec. 4.2.2. Examining the other main and interaction effects in the light of specific bias, the effects are reduced to a minimum yielding no valid significant differences in means across the cultures investigated.

4.2.4 Acoustical sound descriptors in a cross-cultural context

Three objective parametric sound descriptors were calculated to objectively characterize the variance of the stimuli: a level parameter $L_{Aeq,6s}$, a parameter describing the tonal content ΔL_T (DIN 45681:2005-03, 2005), and the sharpness S (von Bismarck, 1984; Zwicker & Fastl, 1999). The background noise (stimulus S_1 in Table 4.1) has the lowest level: 44.5 dBA. For the other stimuli, adding tonal energy increases the level $L_{Aeq,6s}$, and the difference between the lowest and the highest level (53.3 dBA) amounts to 8.7 dBA. Therefore, many of these stimuli have clearly distinguishable loudness differences. The parameter changing the most among the stimuli is the signal-to-noise indicator ΔL_T, which was calculated according to (DIN 45681:2005-03, 2005). This parameter ranges from $-\infty$ to 24.5 dB. With 1.25 acum, stimulus S_1, having no tonal content embedded in the noise, has the highest sharpness S. This sharpness is equivalent to the sharpness of a third octave band centered around 1.25 kHz with a level of 60 dBA. Apparently, adding tones with frequencies lower than 1.25 kHz, i.e., all tonal components in this study, to the background noise decreases the perceived sharpness of stimuli containing tonal energy when compared to stimulus S_1: The more the low frequency tonal content, the lower the sharpness. The addition of tones to the noise affects all three signal parameters, which is represented by high correlation coefficients r between them : $r(\Delta L_T, L_{Aeq,6s})$ = 0.88, $r(S, \Delta L_T)$ = -0.95, and $r(S, L_{Aeq,6s})$ = -0.98. All of these correlations are significant ($p = .05$).

The relation between the objective sound descriptors and the perceptual dimensions is described by correlation coefficients. The marker item score sums for a particular factor dimension (see Sec 4.2.2) were used as scale values of the corresponding stimuli to calculate. The results will be discussed separately for the three dimensions. In following, all reported correlations are significant on a $p = .05$ level.

Power. A level parameter was expected to correlate the most highly with this factor. Yet due to the high correlationbetween all signal parameters, the $L_{Aeq,6s}$ is not better than the other parameters, all showing rather small coefficients. The correlation coefficients for the Japanese sample are significantly lower ($r = 0.24$) than the other samples ($r \approx 0.5$). This difference is a result of the non-uniform bias on the items constituting this factor. As mentioned in Sec. 4.2.2, the non-uniform bias relates to the scale usage of the Japanese sample, i.e., it shows a smaller overall variance.

Thus, the sample shows a smaller correlation. For an integral view, Figure 4.4 shows the relationship between the $L_{Aeq,6s}$ and the average of all subjective scale values for the power factor.

Metallic. Correlations with the objective signal parameters are highest for the metallic dimension ($r \approx 0.8$, Japanese sample $r \approx 0.6$). This factor is best represented by the parameter ΔL_T as the correlations with the level parameter $L_{Aeq,6s}$ and the calculated sharpness are lower. The parameter ΔL_T characterizes the relation between the tonal energy of a stimulus and its noise. The metallic dimension best represents how timbre fluctuates among the stimuli. Again, the difference between the cultures is a result of the lower overall variance in the Japanese sample, which leads to the non-uniform bias on the underlying scale (see Table 4.4).

Figure 4.5 illustrates a broader perspective on the relationship between the parameter ΔL_T and the metallic dimension. The average scale values for the three sub-samples are reported there. A constant 'metallic' perception was observed for ΔL_T below threshold, i.e., < 0. No significant change in the sound character is expected for those values. The relationship becomes linear for higher values. These results are similar to the findings reported by Namba et al. (1992, $\Delta L_T \hat{=} S/N$).

For our data, the correlation coefficient between the metallic or timbre factor and the calculated sharpness is negative in all cultural samples. This is because a more pronounced tonal character corresponds to a decrease in sharpness. Apparently, it is in line with the intercorrelations between the signal parameters.

Pleasant. The correlation between the objective parameters and this factor are the lowest. Even insignificant correlations were found for the French participants, and the highest correlations were found for the German sample ($r \approx 0.4$), while the Japanese sample shows only a weak correlation ($r \approx 0.2$). Due to the fact that the factors are perpendicular to each other as well as to the high correlationbetween the objective parameters, the high correlation found for the metallic factor inevitably results in lower correlations for the other two subjective dimensions. Moreover, this pleasant factor can be interpreted as an evaluation factor. It varies noticeably among participants when they judge the same stimuli.

The parameter ΔL_T, i.e., the sum of all tonal energy above masked threshold, was not the only parameter calculated. Different summation methods for the tonal components were compared, such as taking only the maximal component, adding-up only the frequencies lying in the proximity limit Δf_{prox} (ISO 7779:1999, 1999) or within a common critical band (DIN 45681:2005-03, 2005). The different approaches did not lead to significant differences in correlations for any case ($\Delta r_{crit} \approx 0.13$). The parameter ΔL_T was prone to correlate most highly with the pleasant factor.

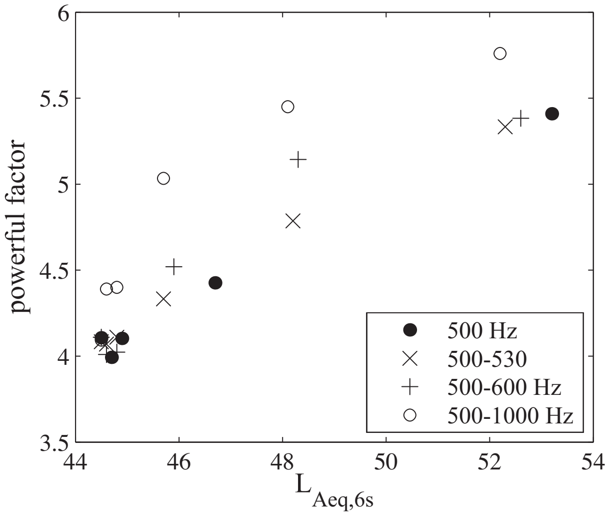

Figure 4.4: $L_{Aeq,6s}$ *is plotted against the power factor, i.e., the mean of the scale values for the underlying adjective items. (The standard deviation,* $\sigma \approx 1$, *is omitted for clarity.)*

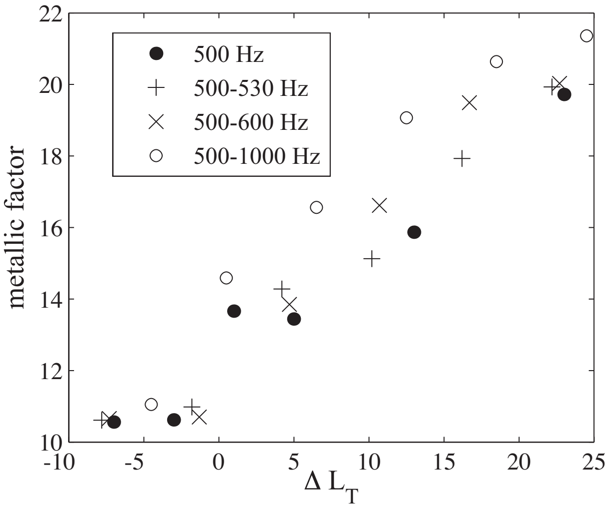

Figure 4.5: Tonal energy ΔL_T *being either* ΔL *or* $\Delta L_{T,sum}$ *is plotted against the* metallic *factor, i.e., the mean scale values for the underlying adjective items. (The standard deviation,* $\sigma \approx 0.8$, *is omitted for clarity.)*

4.3 Discussion

4.3.1 Cross-cultural similarities and differences

According to the results presented in Sec. 4.2.1, the perceptual and connotative factors assessed for each culture are similar, i.e., they feature the same three-factor structure. These factors are labeled pleasant, metallic/timbre, and power after results reported by Namba et al. (1992). The invariance of this factorial solution among the Japanese, French and German samples is comparable to the consistent structure reported by Osgood et al. (1975) regarding the affective meaning of concepts, which was tested in several cultures. In their study, concepts were presented without context to ensure their meaning would not be associated to anything else, and the authors reported a structure they titled EPA (Evaluation, Potency, Activity). The factorial solution found in this study resembles the EPA structure: the evaluation factor corresponds to the pleasantness factor, representing the evaluation of sound, while the power factor coincides with the power factor. It is argued that the metallic/timbre factor represents the activity dimension as the latter is connected to the quality in the sense of modal features. Heise (1969) argues that this structure does not always have to be obtained as it depends on the stimuli investigated. Within certain sound corpora this structure might not be fully evolved as certain dimensions are fused. On the other hand, there may be more dimensions, e.g. Solomon (1958), if for example the sound corpus varies in many aspects of timbre. An example of fused dimension would be if the annoyance is highly coupled with certain characteristics of the sound, such as sharpness or roughness (Zwicker & Fastl, 1999). In this example, the pleasant and metallic/timbre factor would coincide.

Nevertheless, some items – noisy/quiet, harmonic/discordant, clear/vague – show different affective meanings, although the denotative meaning was correctly translated. The Japanese and German samples used the item noisy/quiet is a similar way: Instead of employing it only in a non-evaluative manner to describe power, it is also used along with items loading on the pleasant factor. Concerning the concept of "noise", this is in line with Kuwano et al. (1986), who reported a high correlation between the concepts of "noise" found in Japan and Germany. Another difference was found with harmonic/discordant. Almost solely an indicator of pleasantness in the Japanese and German samples, it refers to the pleasant and timbre factors in the French sample, perhaps eliciting the musical connotation of harmony. By revealing culturally invariant dimensions, item bias is revealed as well. The item bias analysis, e.g. for clear/vague, showed that cross-cultural differences in means for this item correspond to different ways it is used.

The semantic differential used in this study assessed the affective meaning of sounds in a context-free setting. However, the broad similarity in judgments does not imply that potential reactions or annoyance are equal in every culture in a given context. Namba et al. (1991b), for instance, investigated how people deal with neighborhood noise problems by conducting noise evaluations in different cultures. The differences in the countermeasures against noise showed the Japanese sample applied a more defensive approach, i.e., a high percentage responded that they could get used to noise depending on the situation. Hence, in a given context, these results imply differences in the interaction between noise and the context, which influences the formation of affective meanings. For stimuli in a context-free environment, this study shows that the initial judgment of sounds is very similar cross-culturally.

4.3.2 Bias analysis - Considering its relevance

This study employed the bias analysis described by v.d.Vijver & Leung (1997). It provides the possibility to determine the level of equivalence of the construct and the scales. Therefore, the bias analysis serves as more than a mere guideline for critically assessing every step of the experimental procedure regarding the exclusion of method bias: It yields a quantitative analysis to prevent overestimating any face-value difference. This study exemplifies the latter in two ways. At first, contrary to investigations of perceptual dimensions by Schick & Hoege (1996) as well as Iwamiya & Zhan (1997), the factor rotation reduces the differences between the factor solutions obtained in different cultural sub-samples and after this, congruency is estimated quantitatively. Therefore, the bias analysis can draw a clearer picture of the actual similarities and by doing so, the differences in scale usage (e.g. harmonic/discordant in the French sample) are outlined precisely. Secondly, the item bias analysis yielded the prerequisite to analyze the face-value differences. The study showed that the cultural differences relate to the heterogeneous scale usage. Thus, the analysis offers a valuable test procedure indicating to what degree scales are biased. The semantic differential technique, i.e., using multiple scales which can be very close in meaning (e.g. pleasant and pleasing), seems to be a necessary precaution in order to explore the full affective meaning of sounds, although differences in usage are not guaranteed in the end. If semantic differentials, which are believed to cover the essential descriptions, are condensed a priori, it is not possible to obtain the perceptual and connotative factors which the item bias analysis is grounded upon. Hence, the cross- cultural application of single categorical scales, no matter how well translated to reflect their denotative meanings, should be avoided as this procedure is prone to bias errors.

4.3.3 Acoustical sound descriptors

Acoustical sound descriptors were correlated with the perceptual and connotative factors obtained by the factor analyses. The summation of the tonal energy above masked threshold explains the metallic/timbre factors. Therefore, the description of the tonal content can be realized by the tonal energy, if the noise contains distinct tonal features (Hansen & Weber, 2008). However, the tonal energy is summed no matter how the frequencies relate to each other, i.e., whether they are located within one critical band or even within the proximity limit. This suggests that the level above threshold could be the parameter to describe the tonal content. This idea follows an approach by Vormann et al. (2000b), in which a tone in noise adjusted to the same tonal energy of the stimulus was used as an adaptive subjective measure. Furthermore, concerning the items describing the change in tonal timbre, differences, which could not be attributed to differential item usage, were not obtained between the cultural samples.

Namba et al. (1992) have shown that the power factor can be modeled by loudness. Here, too, the power dimension can be attributed to the overall loudness, but it is certainly influenced by the presence of the tonal components. The covariation between the $L_{Aeg,6s}$ and the parameter describing the tonal content makes a distinction impossible.

The pleasant/annoyance factor corresponds only very weakly to any of the sound descriptors, as the items on the factor show a relatively small total variance over all stimuli. In the French sample, the dimension is basically independent from the S/N ratio, contrary to, for example, Hellman (1982). They reported that annoyance increases with an increasing S/N ratio. Oppositely, Zimmer et al. (2004) reported in an indirect scaling experiment examining the auditory unpleasantness of environmental sounds using two predictors: sharpness and roughness. These predictors are valid in groups of similar loudness, and the latter serves as the greatest predictor between groups of different stimuli. These results are in line with a concept proposed by Berglund et al. (1994), in which perceived annoyance is based on three factors: loudness-based annoyance, quality-based intrusiveness (due to roughness and sharpness) and the distortion of information. In the present study, overall pleasantness does not vary much in the sense of auditory unpleasantness (Zimmer et al., 2004), as the underlying parameters roughness/sharpness and loudness do not vary much. Therefore, in a context-free environment the annoyance ratings, which can only be based on the acoustical stimuli, should not vary to a great extent either. If the tonal content had been directly related to auditory unpleasantness, the pleasant and metallic/timbre factor would have fused, which is not the case.

This result does, however, not argue for the independence of annoyance in the sense described by Berglund et al. (1994). Tonal components play a major role in sound identification. Enabling the formation of new emergent meanings, e.g. the sound of a lawn mower is not annoying per se but becomes annoying if, for example, the neighbors are using it in the early morning. The current study has shown there is little difference in the evaluation of tonal sounds within a context-free setting, but adding a context and therewith an emergent meaning might lead to evaluation differences across the investigated cultures. The sound of bells, a typical tonal sound, may illustrate this. Associated with a church in Germany, it is used only as a warning signal in Japan.

4.4 Summary and Conclusions

In this cross-cultural study the perception of noise with tonal components was examined in Japan, France and Germany with the aid of a semantic differential. In order to uncover valid cross-cultural differences, a paradigm from cross-cultural psychology (v.d.Vijver & Leung, 1997; v.d. Vijver, 1998) was introduced and applied to the semantic differential data. Differences in cross-cultural data are not necessarily valid differences as they may be due to different forms of bias that jeopardize the equivalence of the data to be compared. V.d. Vijver and Leung (1997) claim that valid cross-cultural differences can be identified if and only if the different levels of equivalence for the data to be compared had been proven. This was equivalent to the request that the corresponding types of bias had been ruled out after a bias analysis.

The first goal was the evaluation of the semantic differentials in a cross-cultural context. It was necessary to apply a factor rotation in order to correctly assess the difference between the connotative and perceptual factors. After rotating the factors onto each other, almost congruent factors were obtained. These common factors were labeled: pleasant, metallic/timbre, and power. They correspond to three culturally-invariant factors describing the affective meaning of concepts: evaluation, activity, and power (Osgood et al., 1975). Small differences were observed regarding how some adjectives are situated within this three-dimensional space, most notably for the items harmonic/discordant and noisy/quiet. In part, this is attributed to different concepts of noisiness and loudness (Namba et al., 1991a).

Furthermore, the items were examined on the basis of equal perceptual dimensions obtained. It was found that most of the face-value differences were due to bias on the respective adjective items. In conclusion, it is recommended to avoid employing single categorical scales when evaluating certain aspects of sound character, e.g. tonality. No matter how well the denotative meaning is established, a bias evaluation of single items is not possible. Hence, a source of systematic errors in a cross-cultural context would remain.

The second goal, the validation of psychoacoustic parameters regarding tonal contents in a cross-cultural context, leads to the conclusion that cultural bias is responsible for the observed differences in correlations between the metallic/timbre factor and the objective parameters for comparisons between both European sub-samples with the Japanese sub-sample. The S/N-based parameters showed a fairly high correlation with the metallic factor. Thus, they are able to describe this important aspect of sound character. A difference was obtained on the pleasant factor. For the French sample there is no correlation between the pleasant factor and objective parameters investigated, while it is weak in the case of the Japanese and German sample. These results suggest that the tonal content might not have affected auditory unpleasantness (Zimmer et al., 2004). Nevertheless, as the tonal content often confers source information to the recipient of noise, it is prone to have an important role in the context of transient affective meaning, i.e., the change of meaning due to a specific context.

5 Auditory object formation, frequency difference limen and pitch strength†

The auditory sensation pitch has three constituent properties: pitch (height) (ANSI-S1.1-1994, 1994, p. 34), pitch strength (e.g. Fastl & Stoll, 1979) and pitch chroma (e.g. Warren et al., 2003). Often the term pitch is used as a synonym for pitch height as, e.g., in its definition: Pitch is defined as the aspect of auditory sensation that enables to evaluate a sound on a scale from "high" to "low" (ANSI-S1.1-1994, 1994, p. 34). However, pitches differ in clarity, i.e., they are perceived in a range from "faint" to "strong". This sensation is labeled as pitch strength (e.g. Fastl & Stoll, 1979; Yost & Hill, 1978).[1] The present paper investigates the role of pitch strength in auditory object formation. An example from musical acoustics illustrates this adequately. The pitch in some African musical instruments such as a Mbira is sometimes layered on the noise as a separate tone forming a separate object and sometime fused with the noise forming one object: a tonal noise (Fales & McAdams, 1994). How can this phenomenon be understood from the perspective of auditory objects? Kubovy & van Valkenburg (2001) reason that "a *perceptual object* is that which is susceptible to figure-ground segregation (p. 102)," and therefore linking the objecthood to the possibility of figure-ground segregation be it in the visual or the auditory domain. In the example given, the pitch is at one time an attribute of the noise, contrary to the case when the strong pitch (i.e. pitch with a high pitch strength) forms a segregated object, i.e., a figure. In the example of African musical instruments, pitches differ in their strength on a scale from "faint" to "strong", i.e., they differ in clarity. The question arises if this pitch strength sensation alone can give rise to figure-ground segregation. Is faint pitch attributed to the noise leading to a sensation of tonal noise, while strong pitch can become a figure identified as a tone separated from noise? This leads to the first hypothesis: Pitch strength is capable of serving as a segregation cue, i.e., a tone separates from a noise having a sufficiently strong pitch. Therefore, the fusion or segregation has to be measured quantitatively as a function of pitch strength. On the basis of the quantified segregation, other parameters co-varying can be measured to pinpoint

†In preparation: *H. Hansen & R. Weber, Auditory object formation, frequency difference limen and pitch strength, Hearing Res.*

[1] This sensation is also investigated under other labels and in different contexts, such as tonality (e.g. Pollack, 1948; Harris & Myers, 1949), pitch salience (e.g. Terhardt et al., 1982) or tonalness (Aures, 1985b,a).

the psychophysical performance limits underlying the pitch strength sensation. Here, the hypothesis is that the limits of frequency discriminations preformance are the correlate of the pitch differences in clarity or strength.

Artificial stimuli resembling the example of African musical instruments are, e.g., narrowband noise (NBN) superimposed with pure tones at the center frequency of the NBN. With these stimuli, two percepts can occur: A perception of tonal noise where pitch is attributed to the noise and perception of a tone and a noisy background (Fales & McAdams, 1994). In more detail, if the difference between the tone sound pressure level (L_T) and the noise sound pressure level (L_N), i.e., the signal-to-noise ratio (SNR = $L_T - L_N$), is rather small, tone and noise are more likely to be perceived as fused, i.e., they become one auditory object: a tonal noise. If the SNR becomes higher both stimuli are more likely to segregate, and a tone (figure) and a noise (ground) are perceived. For a certain range of small SNRs the perception is ambiguous. This ambiguity is quantifiable in a classification experiment, effectively measuring the amount of grouping/segregation (Kubovy & Wagemans, 1995). Such a classification experiment is described in Section 5.1. The stimuli used, i.e., NBN with an added tone at center frequency, vary considerably in pitch strength depending on the SNR. NBN has a pitch corresponding to the noise's center frequency f_c for small bandwidths (Fastl, 1971), but its pitch strength compared to a pure tone at the same frequency is considerably lower (Fastl & Stoll, 1979). However, two precautions are necessary: If the noise level L_N is held constant and the SNR is increased, then the overall level and consequently the loudness increases. To avoid the loudness to be taken as a cue in the classification task the loudness cue has to be destroyed. Fales & McAdams (1994) made it impossible to use loudness as a cue by varying the overall SPL randomly (roving). Due to this roving, the pitch strength difference is preserved as pitch strength only depends very weakly on the overall level (Fastl & Zwicker, 2007, p. 138). The second cue, apart from the pitch strength, is related to envelope discrimination. The envelope of a NBN is fluctuating while the tone has a flat envelope. As the SNR increases, the participants might use this envelope change as a cue. This cue is controlled by using different NBN bandwidth at different center frequencies f_c: The bandwidth decreases by 50 Hz for each doubling of the center frequency starting at 50 Hz bandwidth at 250 Hz center frequency. The fluctuation strength of the NBN decreases considerably to virtually no fluctuation at 250 Hz bandwidth (Fastl & Zwicker, 2007). However, compared to the influence of the different bandwidth on the envelope, the added tone has a rather small influence, i.e., the envelope fluctuations in the relevant modulation frequency range are changed only very slightly (For additional information refer to Sec. C). Thus, due to the chosen increase of bandwidth with frequency, strong prediction are made if the envelope cue is used: the effect of the SNR on the judgement would be rather small, i.e., the response within a frequency condition would be very similar.The experiment using the tone-NBN stimulus class thus measures the grouping/segregation quantitatively while using pitch strength as an independent variable. The loudness cue is destroyed through roving while the envelope cue is controlled through the experimental parameters, i.e., making strong predictions if this cue is actually used.

Given that combinations of tones and NBN have a large pitch strength range and segregation is measured, it is possible to investigate the psychophysical performance limits co-varying with the pitch strength: The tone's masking threshold and the frequency difference limen. The former investigates the lower limit of a possible influence of the additional tone on pitch strength. The masking threshold of a tone in noise is determined largely by the SNR, but the masker's temporal structure is known to be influential in masking of tones in NBN. Thus, masking models combining spectral and temporal information have been proposed for masking threshold (Kohlrausch et al., 1997; Moore et al., 1998; Derleth & Dau, 2000). Detection of difference, however, is a classification prerequisite. Differences in masking threshold and classification as a separate tone is reported at hearing threshold (Pollack, 1948). This difference is essentially a difference between detection at all and the detected object having a pitch attribute. In the case of detection and classification as a separate tone in NBN, the case is slightly different as NBN already has a pitch. Here, this pitch is associated with the noise. Hence, the detection refers to perceiving a difference between noise and noise plus the added tone. However, the classification refers to labeling the tonal part as being separated from the noise, i.e., the tone segregation. In this way, measuring the masking threshold should provide the lower limit above which the identification of a separate tone becomes possible.

The difference between detection and identification of a tone at hearing threshold was first associated with the frequency difference limen by Harris & Myers (1949). Since then, several studies report an increase of the frequency difference limen of tones in broadband noise with increasing SNR close to masked threshold (Harris, 1948, 1966; Henning, 1967; Cardozo, 1974; Sinnott & Brown, 1993). For complex tones within broadband noise, it was shown that the pitch extraction mechanism is rather robust and that a larger frequency difference limen due to added noise can be related to the detectability of the tone (Gockel et al., 2006). In more detail, with higher noise levels not only did the frequency difference limen increase but the detectability of the pitch in one or both intervals of the two alternative forced-choice paradigm was also reduced. Therefore, the authors reason that the addition of broadband noise is not a good way to reduce the pitch strength as the results have to be interpreted with regard to the loss of detectability. In the studies mentioned, there is an implicit or sometimes explicit connection between the pitch strength and the frequency difference limen. As the frequency difference limen to the knowledge of the author has not been measured for a tone centered in NBN, the experiment to be reported measures the frequency difference limen performance increase with increasing SNR to test the hypothesis that the frequency difference limen might be the underlying performance of the auditory system which relates to the perception of pitch strength.

The current study sets out to measure the amount of grouping/segregation of a tone in NBN with a pitch strength cue quantitatively. As physical parameters the SNR and the frequency are chosen. Two follow-up experiments measure quantitative and qualitative performance limits, tone masking thresholds and the frequency difference limen, to investigate the related psychophysical parameters of pitch strength. Two hypotheses are tested: The first hypothesis is that there is a segregation/grouping process due to pitch strength, which differentiates the perceptual object 'tonal noise' and the perceptual object 'tone in (background) noise'. Furthermore, the second hypothesis is tested that this segregation, which is based upon pitch strength, is grounded on different frequency discrimination performance and therefore links this performance and pitch strength directly.

5.1 Tone in noise classification experiment

In this objective yes/no classification experiment (e.g. Klein, 2001), participants have to decide whether the stimulus belongs either into the category "tonal noise" or "noise and tone". The SPL of the tone is varied, but cannot be used as a cue as it is destroyed by a roving stimulus level paradigm. Thus, if discrimination occurs, only pitch strength remains as a cue. The pitch strength cue, though varying with the SNR, is almost unaffected by the SPL (Fastl & Zwicker, 2007, p. 138), thus not affected by the roving. The yes/no data measuring the qualitative change from tonal noise to tone in noise have then to be analyzed with respect to the SNR and the frequency condition, while the SNR serves now as a measure for pitch strength. The yes/no frequency data measuring the segregation can be connected quantitatively to the frequency condition and the SNR by using a general linear model. Moreover, as data from participants are combined, a population-averaged model estimated by generalized estimating equations links the frequency data with the experimental condition (e.g. Hardin & Hilbe, 2003).

5.1.1 Method

Stimuli

The stimuli are NBNs with an added pure tone at noise center frequency f_c. The noise center frequencies were 250, 500, 1k, 2k, and 4k Hz with bandwidths of 50, 100, 150, 200 and 250 Hz, respectively. The NBN bandwidths chosen have an equal subjective pitch strength according to Fastl & Zwicker (2007, p. 139). As the 4 kHz stimulus' pitch strength is not reported, linear extrapolation yields the chosen bandwidth. The SNRs ($L_T - L_N$ [dB]) used in the experiment are shown in Table 5.1. There are two sets. In informal listening tests, the region of interest, i.e., the separation of tone and noise, was detected to have a lower SNR for the lower frequencies (250/500 Hz), which lead to decrease of -3 dB of the respective SNR conditions. Each of the five frequencies was presented with 8 different SNRs. The SNR ranges from -8 to +4 dB for 250/500 Hz and from -5 to +7 dB in the 1, 2, and, 4 kHz condition with a NBN without any additional pure tone included for each frequency condition. The stimuli were digitally pre-generated using Matlab (vers.: 7.4.0) at a sampling rate of 44.1 kHz with a 16-bit resolution. Each NBN is generated by drawing amplitudes for each frequency bin within the bandwidth out of a Gaussian distribution. Each of these tone in noise sets was generated three

Table 5.1: *Experiments' stimuli. Center frequencies and bandwidth of the experiments are given along with the SNRs used in the classification experiment (Sec. 5.1). In total, there are 40 different stimuli conditions. The ⋆-marked stimuli form the subset for the frequency discrimination experiment (Sec. 5.1).*

$f_c[Hz]$	$\Delta f_{bw}[Hz]$	$SNR[dB]$									
		⋆		⋆		⋆	⋆	⋆		⋆	
250	50	$-\infty$	-8	-5	-3.5	-2	-0.5	1		4	
500	100	$-\infty$	-8	-5	-3.5	-2	-0.5	1		4	
1k	150	$-\infty$		-5		-2	-0.5	1	2.5	4	7
2k	200	$-\infty$		-5		-2	-0.5	1	2.5	4	7
4k	250	$-\infty$		-5		-2	-0.5	1	2.5	4	7

times, to counterbalance the effect of a specific frozen noise sample (Kohlrausch et al., 1997). The total number of trials in one block was therefore 120 stimuli (5 center frequencies $\times 8$ SNR $\times 3$ noises). The duration of every stimulus was 1500 ms including 50 ms cosine ramps. The NBN level was 60 dB SPL. Along with the added tonal energy, the overall level ranged between 62 and 70 dB SPL. A roving stimulus level was used to encourage the judgement due to timbre differences (Fales & McAdams, 1994). Each stimulus was presented with a level deviation which was chosen out of a uniform distribution ranging from ± 6 dB SPL.

Apparatus

The experiment was controlled using the PsiExp v3.4 experimentation environment for the stimulus control and the graphical user interface (Smith, 1995). The stimulus is sent via Max/MSP v4.6 to the Fireface 800 D/A converter. All stimuli were amplified by a P2075 Yamaha Power Amplifier. The stimuli were presented diotically via Sennheiser HD 250 linear II headphones to the participants being seated individually in a double-walled sound attenuating booth.

Participants

16 listeners took part in the initial session of the experiment (10 female, 6 male). All were paid on an hourly basis and had no reported hearing loss. All participants were naive in the sense that they were not affiliated with any psychoacoustic research or laboratory. 11 participants took part in the second session, the re-test (7 female and 4 male). The other five participants did not reply to the re-invitation. The age which were involved in both sessions ranged from 20 to 65 with a median age of 27.

Procedure

To measure the degree of segregation/grouping as function of level and for each frequency condition, a single-interval, randomly interleaved, yes/no classification task was employed. With regard to stimulus presentation frequencies the experiment was fully balanced. The question asked asked was: "Is the tone clearly distinct from the noise?" The answer was "yes" or "no". The answer was given by clicking fields on a screen.

The experiment was divided into two sessions. These two sessions were separated by four months to measure the long-term reliability. Within each session, the participants had to read a written introduction in which they were introduced to the concept of tonal sounds by mentioning typical environmental, tonal sounds such as a piano or a telephone ringing. Thereafter, they listened to the experiment's stimuli, five different SNRs at 1 kHz, starting with NBN (SNR = ∞ dB) followed by a NBN with a distinct pure tone (SNR = 7 dB) for maximum contrast, the test stimuli were at a SNR of -.5, 1 and 2.5 dB. After the introduction, a test block was presented at 500 Hz and 1 kHz. This block consisted of one condition of frozen noise at every SNR step resulting in 16 randomly presented test stimuli. After this, the session started. Each session can be divided into experimental blocks. One block consisted out of 120 trial for every frequency and SNR condition presented at every frozen noise condition (40×3, Table 5.1). In the first session, the participant judged 5 blocks of stimuli (600 trials). This resulted in 15 responses per stimulus per listener. In the second session, the participants had to judge 9 blocks (1080 trials), which were in total 27 responses per stimulus per listener. There was a mandatory 5 min. pause every 3 blocks. In total, each SNR and frequencies condition was evaluated 42 times with the completion of both sessions. The first session took about 30 min. while the second session took about 50 min.

5.1.2 Results

The data obtained in this experiment are 42 yes/no answers for eleven participant for each condition (Table 5.1). Data analysis follows the following roadmap. After reliability analyses, the experiment set out to measure the qualitative change of the classification into the categories "tonal noise" and "tone in noise" depending on different values of pitch strength at five frequency conditions. The pitch strength has been operationalized as SNR. This is possible due to fact that the co-varying loudness has been destroyed through roving, while the pitch strength is largely unaffected by the overall level (Fastl & Zwicker, 2007, p. 138). The roving serves as an experimental filter for the loudness cue. The analysis aims at connecting the yes/no classification data with the SNR at the five frequency conditions.

This paragraph gives a coarse overview of the modeling approach to grasp the qualitative change of classification on the basis of pitch strength operationalized as SNR. At first, the participants data has to be taken into account. The binomially distributed frequency data is transformed to correct for response criterion, i.e., calculating the d' (Klein, 2001). This is done by z-transforming the participants' data for each frequency and subtracting the z-transformed value of the catch trial (NBN-only condition, $SNR = -\infty$, Table 5.1, also refer to Eq. .2). This procedure is done at every frequency condition. These d' values are a perceptual metric for the classification (Macmillan & Creelman, 2005). They are the input of a model describing this classification with regard to the SNR and the frequency condition, quantitatively. This yields the effect of pitch strength on the classification of a spectral component into a separated tone. Sec. 5.1.2 gives the details of the fitting procedure which includes appreciating the distributional properties and the non-independence of the participant's data.

Intra-subjective reliability

An analysis showed that test session and re-test session are in good agreement. The remaining 11 participants' re-test reliability is checked by analyzing the difference between the yes/no frequencies in the first and in the second session on all 40 stimuli (8 SNR $\times$ 5 frequencies). The confidence intervals were calculated using the common Clopper-Pearson method. The differences in means are evaluated applying the Bonferroni-method for multiple comparisons, i.e., correcting the α level to limit the maximum α for all tests combined to $p = .05$. Most participants do not show any significant difference between test and re-test regarding their yes/no rate. Three participants who differ in judgement, differ only in the 4 kHz condition, while one differs randomly at several stimulus conditions.

The next step is to test whether participants differ from making random judgements, i.e., having rates different from 0.5. For this, the data of the two sessions are combined. The calculations of the confidence intervals and the hypotheses testing is done as above. All but one participant deviate from a 0.5 rate significantly on certain stimuli. The participant actually not differing in judgement from a 0.5 rate is identical to the one having random differences between test and re-test. This participant is excluded from further analysis. Thus, in the experiment's analysis the responses of 10 participants will be analyzed.

Quantitative analysis of a qualitative change

There are several problems concerning the analysis of frequency data from an objective yes/no experiment. Thus, the simple fitting procedure to the d' values as suggested in the overview, though viable as an introduction, is not applicable according to a number of reasons. Firstly, to fit a statistical model to the data, the distributional property has to be taken into account, which is the binomial distribution of the frequency data in this case. Secondly, the frequency data has to be transformed into a perceptual metric, as suggested above, to fit a quantitative covariate, in the case the SNR. Yet, the perceptual metric already consists out of two measurements, the hit rate and the false alarm rate. Thirdly, the frequency condition, a qualitative covariate, has to be taken into account. DeCarlo (1998) showed that signal detection theory can be written as a subclass of generalized linear models. Generalized linear models, a general class of regression-like models for continuous and categorical variables, can handle not only the distributional properties of frequency data, but can take into account the two types of variables: The SNR is a continuous while the frequency condition is a categorical variable. With this approach the d' can be estimate with this generalized regression.

At first, the data of each participant is considered individually. In an experiment, the sensitivity d' and the response criterion c have to be estimated from the probability to respond "yes" $p(Y=1\,|\,X)$, which is estimated by the experiment for both experimental conditions the signal trial $X = 1$ and the catch trial $X = 0$. Along with DeCarlo (1998), the estimation can be written as (Appendix A includes the calculation for the "traditional" d'-metric which is used here.):

$$\mathit{probit}\ p(Y=1\,|\,X) = -c + d'X. \tag{5.1}$$

This is an equation typically found in generalized linear models (McCullagh & Nelder, 1997): the link function is the inverse cumulative normal distribution function (*probit*) and the independent variable, $p(Y=1\,|\,X)$, is binomially distributed. The linear predictor includes the parameters c, the response criterion, and d', the sensitivity, while there is only one factor: X. Here, the dummy variate X equals one, if the stimulus is present, and zero if it is the catch trial. Thus, the response criterion c increases with decreasing probability to respond "yes" although there is no signal present $p(Y=1\,|\,X=0)$.

Within the present experiment, the experimental factors are not limited to a single signal and noise trial. Thus, the linear predictor has to be extended to include the SNR S_j, a quantitative covariate, and the frequency f_i, a qualitative covariate:

$$\mathit{probit}\ p(Y=1\,|\,X,S_j,f_i) = \underbrace{\gamma_{f_i}}_{-c} + \underbrace{(\delta_{f_i}^{slope}\cdot S_j + \delta_{f_i}^{0})}_{d'}\cdot X. \tag{5.2}$$

The qualitative covariate (or factor) introduces one set of parameters for each frequency condition (denoted by index f_i). The SNR is implemented by the variable S_j, along with two parameters to account for the $d' = \delta_{f_i}^{slope}\cdot S_j + \delta_{f_i}^{0}$. $\delta_{f_i}^{slope}$ estimates the increase of d' per dB SNR. It estimates the slope. Before looking at $\delta_{f_i}^{0}$, it should be noted that $-c$ is estimated out of the rate at $X = 0$, the catch trial condition ($S_j = -\infty$). The response criterion was set to the mode of the catch trial distribution (see Appendix A). Thus, if $S_j = -\infty$ and $X = 1$ should yield a d' of 0 as it cannot be differentiated from the catch trial trial. This is true for all SNR values until $SNR = \delta_{f_i}^{0}$. This value marks the increase of the d'. In other words, the $\delta_{f_i}^{0}$ estimates the location of the response criterion on the SNR dimension.

To visualize the Eq. 5.2, refer to Fig. 5.1 where the regression $d' = \delta_{f_i}^{slope}\cdot S_j + \delta_{f_i}^{0}$ is plotted. The response rates are given as $\mathit{probit}(p(Y=1\,|\,X,S_j,f_i))$, the left side of Eq. 5.2. Moreover, the vertical line represents the estimated response criterion $-c$ on the SNR dimension over all participants. This values is estimated by the response to the catch trial $S_j = -\infty$. The location of $-c$ was chosen as $d' = 0$ and thus yields the "origin" of d''s linear predictor.

This generalized linear model is applied to data of each participant individually. This preliminary analysis is necessary to exclude participants whose d' are not significantly increasing with the SNR. This means participants are excluded whose slope parameter

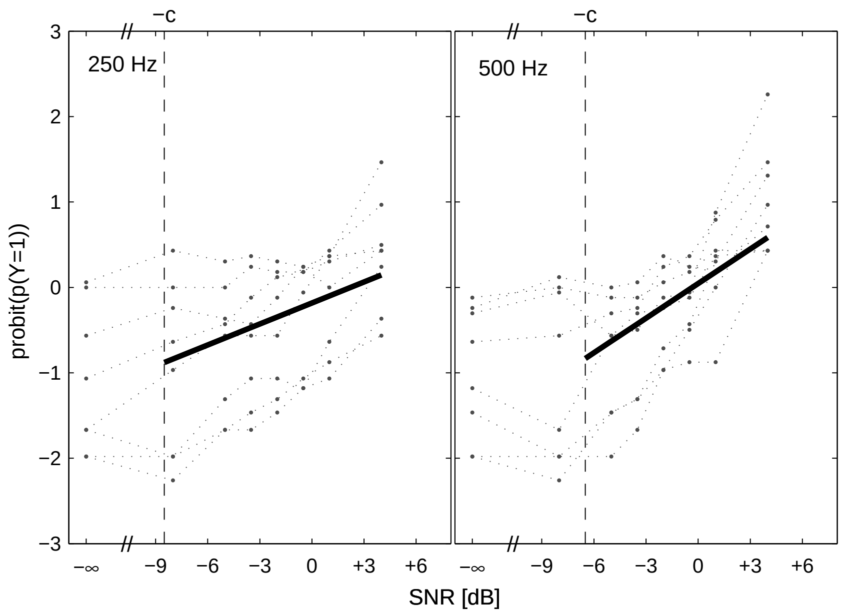

Figure 5.1: *Individual data of eight participants using pitch strength as a cue for the 250 Hz and the 500 Hz condition. The response rates are given as probit*$(p(Y = 1 \mid X, S_j, f_i))$. *The regression via generalized estimating equations (Calculated with Eq. 5.2 using the estimated parameters in Table 5.2) is plotted in a dashed line. This regression estimates the mean increase of* d' *with SNR. Moreover, the vertical line represents the estimated response criterion* $-c$ *on the SNR dimension over all participants. The location of* $-c$ *is chosen with* $d' = 0$.

$\delta_{f_i}^{slope}$ is not larger than zero in the majority of the frequency condition f_i. For two participants this condition was not met. They were responding to the frequency cue only, e.g. responding "yes" at a certain f_i no matter the S_j. As mentioned in the introduction, this performance is predicted if the participants are using the envelope fluctuation as a cue leading to an exclusion of these participants. Therefore, the following population averaged estimations will be based on (8 subjects $\times$ 42 trials) for each experimental condition. The individual data for these 8 participants are plotted in Fig. 5.1 and 5.2. The individual data show a similar increase in d'. However, the response criterion used is rather different within the participants as indicated by different values at the vertical line, which denotes the estimated response criterion location on the SNR dimension across all participants.

Quantitative analysis across participants

After focusing on an individual data analysis to test whether the participants where able to use pitch strength for their classification, the pooled data are analyzed to gain an estimate about how the d' increases with SNR. This indicates quantitatively the

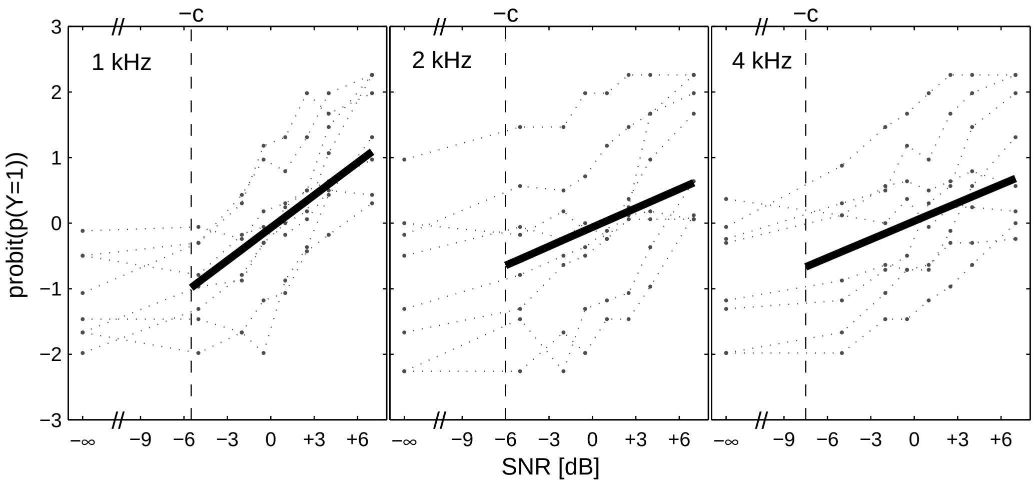

Figure 5.2: *Individual data of eight participants using pitch strength as a cue for the 1 kHz, 2 kHz, and 4 kHz condition. The response rates are given as* $probit(p(Y = 1 \mid X, S_j, f_i))$. *The regression via generalized estimating equations (Calculated with Eq. 5.2 using the estimated parameters in Table 5.2) is plotted in a dashed line. This regression estimates the mean increase of* d' *with SNR. Moreover, the vertical line represents the estimated response criterion* $-c$ *on the SNR dimension over all participants. The location of* $-c$ *is chosen with* $d' = 0$.

influence of the pitch strength on the grouping process. Apart from the estimation of the slope for each frequency (indicated by $\delta_{f_i}^{slope}$ in Eq. 5.2), the location of the response criterion on the SNR dimension can be estimated by $\delta_{f_i}^{0}$ for the pooled data (by setting $d' = 0$). At last, the value of $-c$ which the pooled version give an estimate how the NBN-only condition is judged across participants. Similar to the extension of the ANOVA by the repeated-measure ANOVA, the generalized linear models can be extended to the general estimating equations to deal with within-participant paradigms. This enables to estimate the parameters of Eq. 5.2 even though the data of each participant cannot be assumed as independent (for a more detailed introduction refer to Hardin & Hilbe, 2003, and some brief introductory notes in Appendix B).

The result of this estimation procedure for the parameters in Eq. 5.2 are summarized in Table 5.2.[3] With this parameter values the regression-like lines denoting

[3]The generalized estimating equation can approach the within-participants paradigm differently. In this analysis a population-averaged approach is chosen as there is no hypothesis about differential individual behavior. From the residuals between the estimated model and the data the mean correlation α_x between the trials across participants is estimated as a single parameter estimating the common variance. The approach of estimating the mean correlation α_x between residual and data is just one approach known as the *exchangeable correlation structure*. This structure is chosen as there is no hypothesis about higher correlation between certain residuals, which might be the case for experiments in which the temporal order of presentation plays a role. As this experiment is fully randomized within each participant, assuming another than a mean correlation

Table 5.2: *Estimated parameter values as defined in Eq. 5.2.*

estimated parameter	estimate	SE	χ^2	df	p^2
γ_{250}	-.889	.260	11.700	1	.001**
γ_{500}	-.826	.212	15.118	1	.000**
γ_{1k}	-.979	.221	19.548	1	.000**
γ_{2k}	-.635	.314	4.084	1	.043*
γ_{4k}	-.653	.251	6.741	1	.009**
δ^0_{250}	.707	.186	14.461	1	.000**
δ^0_{500}	.871	.175	24.771	1	.000**
δ^0_{1k}	.907	.205	19.521	1	.000**
δ^0_{2k}	.571	.146	15.267	1	.000**
δ^0_{4k}	.683	.139	24.054	1	.000**
δ^{slope}_{250}	.082	.021	15.319	1	.000**
δ^{slope}_{500}	.135	.031	18.713	1	.000**
δ^{slope}_{1k}	.166	.023	50.202	1	.000**
δ^{slope}_{2k}	.097	.028	12.297	1	.000**
δ^{slope}_{4k}	.093	.027	11.697	1	.001**

the increase of d' with SNR for each frequency shown in Fig. 5.1 and 5.2 can be calculated. At first, the result from the parameter estimation of γ_{f_i} is evaluated. The assumption whether the NBN does not have a pitch producing a separated tone from the noise is supported as the parameters γ_{f_i} are significantly lower than 0. An estimated γ_{f_i} of < 0 indicates a judgement of "distinct tone" in less than 50 % ($z^{-1}(\gamma_{f_i}) < 0.5$). Choosing the response criterion value with DeCarlo (1998) results in the interpretation of the response criterion c as how much the tonal noise itself is perceived as a tone and noise or tonal noise. As shown, the data suggest the latter.

Secondly, analyzing the parameter estimation of the δs from Table 5.2 yields the major result. Eq. 5.2 yields the formula for d' calculated out of the estimated parameters: $d' = \delta^0_{f_i} \cdot S_j + \delta^{slope}_{f_i}$. As the estimated parameters $\delta^0_{f_i}$, defining the slope, are all positive, d' increases with SNR. In Fig. 5.1 and 5.2 the results are depicted in regression-like lines indicating a the different increase in d' for each frequency condition f_i. The steepest increase of d' is at 1 kHz.

The different regression slope address directly the cue apart from pitch strength available in this experiment: the envelope fluctuation. As mentioned in the introduction, the hypothesis that the envelope fluctuation is the cue in the experiment leads to the hypothesis that there is an increase in the slope of the segregation performance with higher bandwidth, i.e. lesser fluctuations. This hypothesis has to be rejected.

α_x between the different conditions within each participant would be speculative. This mean correlation is estimated to $\alpha_x = .29$.

The main result of the experiment is that pitch strength is used in the formation of auditory objects. In NBN the pitch is associated with the noise forming one perceptual object: a tonal noise. With increasing pitch strength, which is in this experiment operationalized as SNR, the noise and the tone form separated auditory objects. The tone becomes distinct from the noise.

5.2 Masking threshold experiment

In the introduction, NBN was introduced as noise having a pitch which is not segregated as a tone. To analyze this assumption, which was based on Fales & McAdams (1994), the masking threshold is measured using an adaptive, three-interval, forced-choice procedure. The masking threshold is the prerequisite of successful tone identification which equals classification into the tone and noise case. At masking threshold, in the forced-choice procedure, only a difference between the signal and the test trial is detected. Yet, the difference does not necessarily lead to a segregated tone, but this threshold can then be compared to the calculated segregation contours.

5.2.1 Method

Stimuli and apparatus

All signal parameters are identical to the values of the first experiment (refer to Sec. 5.1.1). In short, the stimuli are NBN with an added tone at center frequency. The noise center frequencies were 250, 500, 1k, 2k, and 4k Hz with bandwidths of 50, 100, 150, 200 and 250 Hz, respectively. The NBN level was 60 dB SPL. All three intervals of the adaptive, three-interval, forced-choice procedure included the same noise sample and so it was not possible to detect a difference between the intervals due to different instantiations. For each trial, a new noise sample was generated.

The apparatus was again basically identical, except for the program controlling the pre-generated stimuli. The stimuli were digitally generated on an PowerPC using Matlab (vers.: 7.4.0) at a sampling rate of 44.1 kHz with a 16-bit resolution, which was also used conducting the experiment using the PsyLab 2.0 environment (Hansen, 2006). They were converted to analog signals by a Fireface 800 D/A converter. The following sound reproduction stages, the headphone and the sound proof room, are identical to the first experiment (Sec. 5.1.1).

Procedure

The masking thresholds were obtained by an adaptive three-interval, three-alternative, forced-choice procedure using a 1-up-1-down paradigm with feedback. The different frequency condition were presented in an interleaved order, i.e. the frequency changed after every trial. The participant's task was to detect the stimulus which was different within three alternatives. At the start of the experiment, the pure tone had an SNR of 6 dB in the signal interval. For each correct answer the level was initially reduced by 8 dB. This step size was halved every odd reversal to a minimum final step size of 1 dB. The last 6 reversals are averaged. The 1-up-1-down procedure estimates the 50 % point of the psychometric function (Levitt, 1971).

At first, the participants had to read a written introduction in which they were introduced to the concept of tonal sounds by mentioning typical environmental, tonal sounds. Then, the participants were introduced to the stimuli as in the first experiment. A trial experiment with only one frequency (1 kHz) introduced the participant to the experiment. Then, the whole interleaved experiment started. In total, the experiment had a duration of about 30 min.

Participants

The initial 16 listeners of the first experiment took part in the experiment (Sec. 5.1.1).

5.2.2 Results

The masking thresholds for all five frequency conditions of all participants have been measured. The mean intra-individual standard deviation, i.e., the mean deviation of the last six reversals over all participants and conditions, is 1.8 dB. The mean masking thresholds for each condition are shown in Fig. 5.3. Means and 95 % confidence intervals are plotted. The inter-individual confidence intervals range from .9 dB (2 kHz) to 1.8 dB (250 Hz). The thresholds are roughly the same across frequencies, i.e., around -5 dB. The masking threshold for the 250 Hz seems to be slightly elevated.

The results can be compared to other maskings threshold measurements of tone in NBN (Moore et al., 1998, e.g.), though different parameters and procedure foils full comparability. The masking threshold at 1 kHz for three participants with a 65 dB NBN masker (three-interval forced choice, 1-up-three-down rule (79,5 % correct), bandwidth 80 Hz) leads to a masking threshold of 59 dB ± 1 dB (Moore et al., 1998). This leads to a masking threshold of -6 dB, similar to the one reported in Fig. 5.3. The difference of the estimated point on the psychometric function and the different NBN bandwidth have an effect on the reported threshold, this effect, however, is opposed: the higher percentage of correct increases the level necessary to detect the signal interval, while the smaller bandwidth decreases it, as the fluctuations are stronger (Kohlrausch et al., 1997).

5.3 Frequency discrimination experiment

In this experiment, the frequency difference limen of a tone within NBN is measured en bloc. This is done to test the hypothesis that the pitch strength is the perceptual correlate of the frequency difference limen. To ensure comparability with the classification experiment, a subset of the stimuli of the classification experiment is chosen (Table 5.1, $\star$-stimuli). In addition, the performance in each frequency condition is bound by two limits: the frequency difference limen of a tone within NBN cannot be

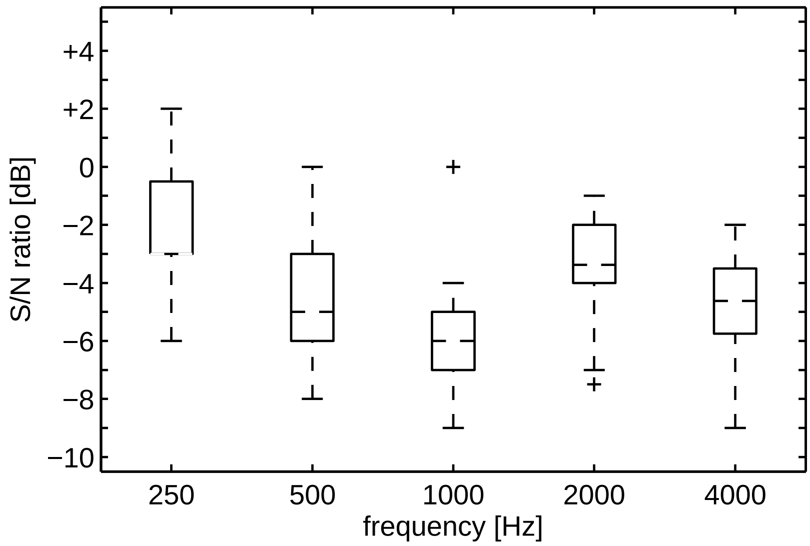

Figure 5.3: *Inter-individual masking thresholds of tones in narrow-band noise obtained by a interleaved 3AFC 1up-1-down procedure at different center frequencies f_c. The respective bandwidths Δf_{BW} are 50, 100, 150, 200 and 250 Hz.*

worse than a condition in which only a NBN is presented, and it cannot be better than the pure tone case. Therefore, these stimuli are measured additionally to the ⋆-stimuli to identify the range of possible frequency difference limina.

5.3.1 Method

Stimuli and apparatus

The parameter of the experiment's stimuli can be found in Table 5.1. They are marked with a ⋆. Furthermore, 60 dB pure tone was tested. In total, there were the five frequency conditions and six SNR conditions. Thus, in total 30 stimuli are measured.

As in the first two experiments, the NBN is generated by drawing amplitudes for each frequency bin within the bandwidth out of a Gaussian distribution. To generate the signal trial, i.e., the trial which is higher in frequency and to be detected, the drawn amplitudes are kept, but are shifted into higher frequency bins. In this way, the temporal structure is kept, which could be otherwise used as a cue. The stimuli were generated digitally with the Matlab software (MathWorks) at a sampling rate of 44.1 kHz with a 16-bit resolution, DA converted (RME-ADI-8 DS), amplified (TDT HB7), and diotically presented for a duration of 500 ms through headphones (Sennheiser HD 650). The participants were situated in a double-walled sound-proof room. The Matlab software was also used to run the experiment using the PsyLab 2.0 environment (Hansen, 2006), which controls the whole measurement procedure.

Procedure

The difference limina were obtained by an adaptive three-interval, three-alternative, forced-choice procedure using a 1-up-2-down paradigm with feedback. All conditions were presented in an interleaved order, i.e., the frequency changed every trial. The participant's task was to detect the interval in which stimulus was different. At the start, the difference between signal and test signal was Δf of 65 cents ($\Delta f/f \approx 3.83 \cdot 10^{-2}$). The initial step size was 16 cent ($\Delta f/f \approx .93 \cdot 10^{-2}$). This step size was reduced every odd reversal by a factor of 0.5 to a minimum final step size of 1 cent ($\Delta f/f \approx .06 \cdot 10^{-2}$). The last six reversals are averaged. With the 1-up-2-down paradigm the 70.7 % of the psychometric function is estimated (Levitt, 1971).

At first, the participants had to read a written introduction in which they were told to select the interval in which the stimulus which differs. A trial experiment with only one frequency (1 kHz) introduced the participant to the experiment's procedure. Thereafter, one of three blocks is tested. Within each block, frequency difference limina for two SNRs at every center frequency f_c are estimated. Every block of each experiment lasts about 45 min.

Participants

1 female and 7 male participants, which were paid on an hourly basis, took part in the experiment. None of them reported ever having suffered from hearing loss. The age of the group ranged from 24 to 37 with an median age of 30.5 years. The participants were not involved in the first two experiments.

5.3.2 Results

The experiment's results, mean and standard error across participants of the frequency difference limen ($\Delta f/f$), are shown in Fig. 5.4. Some general trends can be observed. First, the relative frequency difference limen decreases with increasing SNR. The narrowband noise condition ($-\infty$) and the pure tone (PT) condition serve as the upper and lower limit of the frequency difference limen. Second, there is a general tendency of decreasing frequency difference limen with frequency up to 2 kHz.

These frequency difference limina are analyzed by a two-within factor, repeated measure ANOVA. The factors are the SNR and the frequency condition. It reveals a significant main effect for the SNR condition ($F(1.21, 8.48) = 17.67, p < .01, \epsilon_{GG} = .30$)[4], for center frequency f_c ($F(1.67, 11.71) = 14.08, p < .01, \epsilon_{GG} = .34$), and for the interaction between these two factors, ($F(3.71, 26.12) = 5.39, p < .01, \epsilon_{GG} = .19$). This shows that the observations are confirmed quantitatively.

The inter-individual difference in the estimated frequency difference limina are reflected in the standard errors (Fig. 5.4), which are comparable to the relatively large

[4] Greenhouse-Geissner correction to account for sphericity.

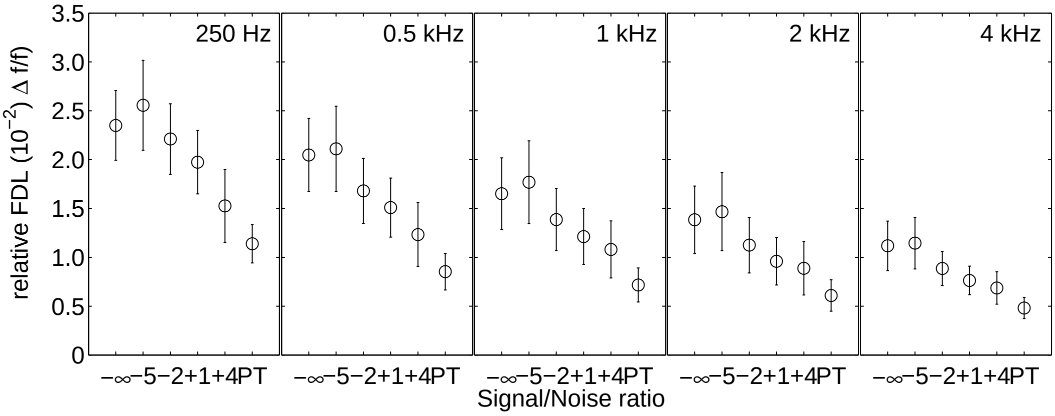

Figure 5.4: *Mean relative frequency difference limen ($\Delta f/f$) along with the standard error (σ_{error}) of eight participants versus the varying SNR conditions are shown. Each frequency condition is reported in a separate subfigure. Two trends can be observed. First, the relative frequency difference limen decreases with increasing SNR. The narrowband noise condition ($-\infty$) and the pure tone (PT) condition serve as the upper and lower limit of the frequency difference limen. Second, there is a general tendency of decreasing frequency difference limen with frequency up to 2 kHz. Furthermore, the frequency difference limen does not increase for SNR values lower than -5 dB.*

inter-individual difference generally reported (Fastl & Zwicker, 2007). The within-participant variation measured by the relative standard error of the last six reversals in the adaptive procedure is found to be lower for the noise case (about 3 %) than in the pure tone case (about 8-10%).

The frequency difference limen of a pure tone has been measured by several authors (Harris, 1966; Moore, 1973b; Sinnott & Brown, 1993; Fastl & Zwicker, 2007). The reported frequency difference limina are comparable to those. For the NBN, frequency difference limen were measured by several authors (Michaels, 1957; Fastl, 1971; Moore, 1973a). The comparability, however, is nearly always lacking due to different, mostly smaller, NBN bandwidths used in their experiments. Data from Moore (1973a) at 2 kHz (64 Hz bandwidth, 75 % point on the psychometric function) show a frequency difference limen of about $\Delta f/f \approx .75 \cdot 10^{-2}$, which is well below the NBN value at 2 kHz of $\Delta f/f \approx 1.4 \cdot 10^{-2}$, which has a bandwidth of 200 Hz.

The mean frequency difference limen is generally best for 2 kHz as reported by Moore (1973b) and Fastl & Zwicker (2007). Furthermore, comparing the data to data of the masking effect on the frequency difference limen, the trend of a decreasing frequency difference limen with the SNR is also observed for pure tones in broadband noise (Harris, 1966). There is in the data of Harris (1966) and in the data reported here a smooth transition in frequency difference limen between the broadband noise respectively the NBN and the tone without masking. This study's data has additionally a lower boundary in the NBN's frequency discrimination performance, which contrary to the broadband noise has a pitch. In the experiment of Harris (1966) the limit is the detectability of the tone in broadband noise.

Is the reduced frequency difference limen performance a mere result from the loss in detectability of the tone? For harmonic complex tones in broadband noise, the reduction of the frequency discrimination performance could largely be attributed to the detectability of the harmonic tone complex (Gockel et al., 2006). Only at a very low SNR, the noise has an additional deteriorative influence on the performance. Quite contrary, in this experiment the detectability does not have a large influence on the the frequency discrimination performance as the masking thresholds (Fig. 5.3) are not reflected in the performance data.[5] The frequency discrimination performance is reduced quite linearly with the SNR, although the detectability for each SNR across different frequency conditions is quite different judging from the respective masking threshold.

5.4 Discussion

5.4.1 Classification and the masking threshold

Comparing the results of the masking threshold experiment (Sec. 5.2) to the data of the classification experiment (Sec. 5.1) gives two main results. First, the comparison of the threshold measured with the SNR, at which cumulative d' equals zero, yields that they are in the same SNR region (well below the 0.75 contour in Fig. 5.5). In this way, the classification into a tone and noise with pitch strength as the only reliable cue is only possible well above the masking threshold. This is inline with the difference between detection and classification as a tone at hearing threshold (Pollack, 1948). The comparison's second result is how the noise is actually classified close to threshold. This result can be inferred from the location of $d' = 0$, on the SNR dimension and the mean value of the response criterion $-c$. At a $d' = 0$ the test and signal trial cannot be told apart with regard to the task. As this is also the location of the response criterion $-c$, the response to the catch trial gives the mean response at $d' = 0$. This value as reported in Sec. 5.1 and is significantly lower that 0. Thus, as the $d' = 0$ contour is close to threshold, at masking threshold the stimulus is classified as tonal noise. This is inline with the experiment of Fales & McAdams (1994), which also showed a classification of NBN into the tonal noise category. The classification experiment in conjunction with the measured masking threshold support the first hypothesis, that there is a tone and noise segregation based on pitch strength, as the NBN is perceived as tonal noise and only increasing pitch strength yields a classification into tone and noise.

[5] Furthermore, in Sec. C the relative standard deviation of the instantaneous frequency estimation is shown in Fig. C.3. The deviation is remarkably reduced with SNR hinting a deteriorative influence of the narrowband noise.

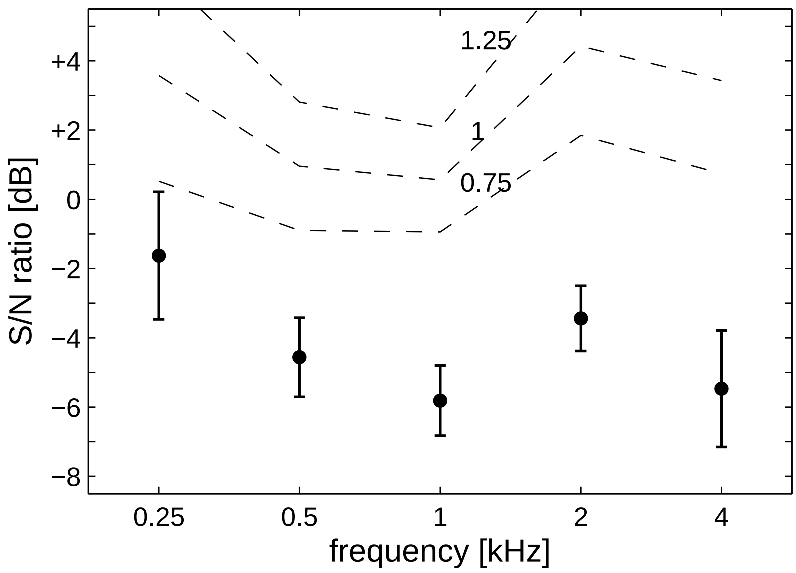

Figure 5.5: *Visualization of model's result through estimated equal cumulative d' contours (segregation contours, data from Table 5.2). The parameter of the set of curves is d'. The bandwidths Δf_{BW} at their respective center frequencies f_c are 50, 100, 150, 200 and 250 Hz*

5.4.2 Classification and the frequency difference limen

The frequency difference limen was measured to test the hypothesis that the frequency discrimination performance might underlie the percept of pitch strength on which the segregation and hence the classification experiment was based. Fig. 5.6 shows the estimated cumulative d' versus the mean frequency difference limen relative to the mean pure tone frequency difference limen at the appropriate frequency ($\Delta f/\Delta f_{PT}$). This is necessary to compare the data across frequency conditions as the frequency difference limen of a pure tone is different for each frequency and the interest lies only in the pitch strength's increase within one frequency condition compared to the increase in segregation. The cumulative d' is linearly related to the relative increase in frequency discrimination performance within the respective frequency condition ($r = -.89, p < .01$). In other words, the classification of a given stimulus into the 'tone and noise' category, a higher d', is more probable with higher frequency discrimination performance, a lower $\Delta f/\Delta f_{PT}$.

This result is supported by comparing other pitch strength data with frequency difference limen data. The rapidly decreasing pitch strength estimate for pure tones at high frequencies (above 3000 Hz Fastl & Zwicker, 2007, p. 139) matches well with the rapid increase in frequency difference limen at frequencies above 3000 Hz (Moore, 1973b). The same holds for the pure tone at lower frequencies. The increase of pitch strength with level, which increases far slower than loudness (ca. 10 % / 10 dB (between 20 and 80 dB SPL) Fastl & Zwicker, 2007, p. 138), relates very well to the data of Wier et al. (1977) showing a similar increase in frequency

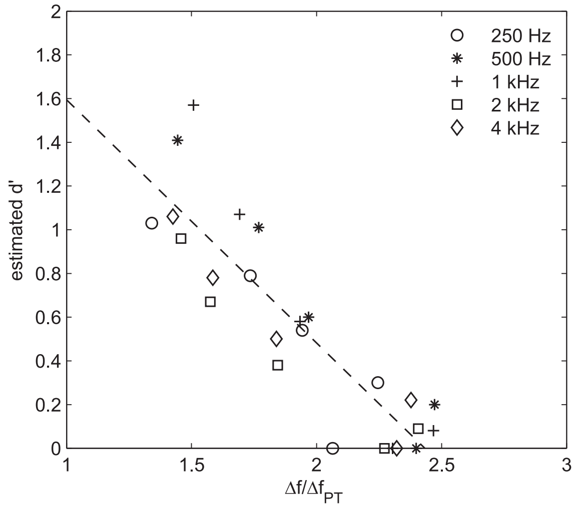

Figure 5.6: *The mean frequency difference limen relative to the median pure tone frequency difference limen at the appropriate frequency ($\Delta f/\Delta f_{PT}$) is plotted versus the estimated d'. This d' is calculated using the parameter estimates of Table 5.2. The correlation is highly significant ($r = -.89, p < .01$). The standard error is omitted for clarity (see Fig. 5.4).*

difference limen for those SPLs. Similar support can be raised by comparing the pitch estimates of NBN (Fastl & Zwicker, 2007, p. 139) with their frequency difference limen (Michaels, 1957; Moore, 1973a) as well as the dependency of the estimated pitch strength (Fastl & Zwicker, 2007, p. 138) on duration of the stimulus with the increase of frequency difference limen with decreasing duration (Moore, 1973b). Melody recognition performance decreases with decreasing pitch strength. At the lower frequency end of melody recognition at about 30 Hz (Pressnitzer et al., 2001) this can be heard at large organ pipes. The pitch of large 16'-32' organ pipes is described as faint/weak (von Helmholtz, 1954, p. 174-179). By connecting the melody recognition to pitch strength, recent results of experiments on binaural pitch perception in normal hearing and hearing impaired listeners suggest a connection of melody recognition and frequency discrimination performance further support the present results (Santurette & Dau, 2007). At last, in recent attempts to physiologically plausible modeling of the pitch strength, the operational definition of pitch strength is the *certainty of pitch* (McLachlan, 2009). This view is very well supported by the result of this study and, in conclusion, this view is tested experimentally by relating a segregation task possible on the basis of pitch strength with explicit frequency difference limen measurements.

5.4.3 Quantification of qualitative change

The analysis of Fig. 5.6 show a quantitative relation between the frequency difference limen and either the grouping of a pitch with the noise or the formation of a separate tone. This quantitative relation has an analogue in vision: the grouping in dot lattices by proximity (Kubovy & Wagemans, 1995). In the grouping process of dot lattices there are four-alternative choices contrary to the two categories in the reported experiment. In the analysis of this grouping data, van Valkenburg & Kubovy plot a log-odd measure versus the physical attribute (van Valkenburg & Kubovy, 2004, p. 133). In terms of a general linear models, van Valkenburg & Kubovy link the probabilities of choosing a certain direction of the dot lattice to linear predictor which includes the relative dot distance as a physical attribute. The link function is a multinomial logit function. Van Valkenburg & Kubovy (2004) interpret this as a *quantitative Gestalt law* able to predict the grouping processes in dot lattices due to proximity probabilistically. Along these lines, the relation found in Fig. 5.6 between the frequency difference limen and the grouping can be viewed as a quantitative Gestalt law governing the grouping due to pitch strength again in a probabilistic way.

Van Valkenburg & Kubovy (2004) offer a theoretical framework to discuss the results. Early process extract elements and group them together into perceptual organizations. One of those perceptual organizations will be, via attention, the figure, the others will form the ground. The ground is normally not available to the listener in an as differentiated fashion as the figure (Brochard et al., 1999). Within this framework the results can be analyzed as follows. The pitch is at one time an attribute of the noise forming *one* perceptual object. If the pitch strength is large enough this changes, and the attributes form *two* perceptual objects, a tone and the noise. One question which remains open though is what actually happens to the background's pitch when it is separated as a tone.

5.5 Conclusion

The perception of a tone in the presence of tonal noise was investigated. According to the results of experiments, pitch can be fused with a noise or the pitch can lead to a tone separate from the noise. The hypothesis that this segregation performance is based on timbre aspects, namely pitch strength, is experimentally verified as all other possible cues were destroyed. Furthermore, the identification of a separate tone is in general above the masking threshold. The comparison of the results of the second and the third experiment yields that at all other frequency conditions, the masking thresholds are inline with the improvement in frequency difference limen. This makes the transition to the second hypothesis. In the first experiment, the amount of grouping/segregation measured with an objective yes/no experiment quantifies the strength in grouping of the tone and the noise. On the one hand, the linear relation of the frequency difference limen, i.e., the frequency discrimination performance, and the grouping provides direct evidence of the pitch strength being related to this discrimination performance. On the other hand, the linear relationship can be interpreted as a quantitative Gestalt law governing the grouping due to pitch strength.

6 General conclusions & outlook

This thesis has explored various aspects of the tone-noise dichotomy. Based on research literature, two concepts were extracted from a plethora of existing concepts and research contexts: the magnitude of tonal content and the pitch strength. As it appears, the tone-noise dichotomy is two-fold. On the one hand, the relation between two separate objects was considered: the tonal part and the noise part of a sound. These two parts can be related in magnitude: Does the noise part or the tonal part dominate? The second dichotomy is within a given auditory object, relating to the question: Is this object more noise-like or more tonal? How large is the pitch strength of this object? Therefore, in one instance a relation between an acoustical foreground (tonal part) and background (noise part) is considered and in the other an attribute of an object is in the focus.

With regard to the tone-noise-relation, it has been shown that perceptual and connotative factors are stable across cultures. In the wake of this study, most cross-cultural difference in the judgement could be related to item bias, i.e., different groups using the items differently. Furthermore, the idea has been proposed that the judgement of the magnitude of tonal content is related to the tonal parts' partial loudness. It was found that the partial loudness was far easier and more intuitive to adjust in a magnitude adjustment experiment than the magnitude of tonal content even though the concept was explained in detail and with examples. The hypothesis whether partial loudness can be identified with the magnitude of tonal content could thus be confirmed in this first experiment. Given this result, the shortcomings of existing standards such as the DIN 45681:2005-03 can be pointed out. They treat all components as independent, thus neglecting the increase of the magnitude of tonal content due to a reduced masking threshold of harmonic complex tones. Furthermore, using SNR approaches the actual perceptual magnitude is not obtained. Currently, no standard offers any method, apart from reporting all components or the strongest one, how to add different tonal components. For these issues, the partial loudness approach offers solutions eo ipso.

The second main concept, pitch strength, has been investigated according to the hypothesis whether it is related to the frequency discrimination performance of the human auditory system. The frequency difference limen correlated strongly with segregation performance in an experiment in which the pitch strength of the stimuli was the main cue. Therefore, the hypothesis could be confirmed.

Some open issues still remain. First, with regard to contributions of single tones of a two tone complex to the overall partial loudness, an asymmetry has been found. Although all tones are within one critical band and should therefore be energetically added, the high and low tones contribute differently. The effect is found to be larger with small SNR. This result suggests the need for experiments which vary the contribution of higher and lower tones systematically. In the reported experiments, one tone of the two-tone singles tones were adjusted, however for maximum contrast two tone complexes with different spectral content should compared directly.

A second open question is the combination of the two concepts presented. How does the pitch strength reduce the magnitude of tonal content? The hypothesis of the identification of the magnitude of tonal content with the partial loudness of the tonal part relies on the assumption that they have more or less equal pitch strength. The DIN 45681:2005-03 (2005) specifically excludes tonal components which do not have a pitch strength of 70 % or more. Therefore, the next step would be to merge those aspect generating a reliable measure of magnitude of tonal content for every application context, be it in the correct characterization of noise immission or the optimal target sound in sound design processes.

Appendix

A Signal detection and generalized linear models

DeCarlo (1998) showed that signal detection theory can be formulated as a subclass of generalized linear models. In this section this is shown for d', i.e., the signal detection theory in which the underlying distributions of the responses are normal-distributed and the variance $\sigma = 1$.

The signal detection theory is characterized by two main ideas. The first idea is the assumption that the effect of an event, such as signal and noise, can be represented by an underlying probability distribution. These distributions only vary in their location characterized by their distribution modes Ψ_s and Ψ_n. The second idea is that the participants are using a criterion c which is used to decide, e.g. whether the event is signal or noise or in which category the event is classified.

The sensitivity parameter d can be written as the difference between both distributions' modes (Ψ_s,Ψ_n) divided by a scale parameter τ.

$$d = \frac{\Psi_s - \Psi_n}{\tau} \tag{.1}$$

For the normal distributed models $\tau = \sigma$, i.e., the standard deviation. In signal detection theory the *hit* (H) and the *false alarm* (F) rate are transformed via a *z transformation* into standard deviations units. The z transformation is the inverse of the normal distribution function (Macmillan & Creelman, 2005, p. 8). The standard deviation is $\sigma = 1$. Hence, Eq. .1 is now

$$d' = z(H) - z(F) \tag{.2}$$

for the normal distributed case. From now on, Ψ_s and Ψ_n, the modes of the distribution, will be referred to as $z(S)$ and $z(N)$.

The two parameters, i.e. sensitivity d' and response criterion c, the location of the criterion, are able to separate the sensory and decisional aspects of the tasks, respectively (DeCarlo, 1998, p. 187). To analyze the conceptual similarities between the signal detection theory and the generalized linear models, the probabilities of responding "yes" are analyzed. The area under the signal distribution $z(S)$ and *left* of the criterion gives the probability of a responding "yes" if the signal is presented $p(Y=1|S)$. For the normal distributed case, it is the cumulative normal distribution Φ which allows to calculate this probability:

$$p(Y=1|S) = 1 - \Phi(c - z(S)) = \Phi(-(c - z(S))). \tag{.3}$$

For the noise presentation the equation is similar:

$$p(Y = 1 | N) = 1 - \Phi((c - z(N))) = \Phi(-(c - z(N))). \qquad (.4)$$

First, Eq. .3 and Eq. .4 have to be inverted. Therefore, the *probit* function is introduced:

$$probit\ p = \Phi^{-1}(p). \qquad (.5)$$

Than, Eq. .3 and Eq. .4 can be summarized by introducing a "dummy variable" X, which is 0 for noise and 1 for signal trials.

$$\begin{aligned} probit\ p(Y = 1 | X) &= (z(N) - c) + (z(H) - z(N)) \cdot X \\ &= (z(N) - c) + d'X \end{aligned} \qquad (.6)$$

Now, $z(N)$ is arbitrarily set to 0 and the equation relates the signal detection parameters c and d' to the intercept and slope of a generalized linear model:

$$probit\ p(Y = 1 | X) = -c + d'X \qquad (.7)$$

As already stated by DeCarlo (1998), now the signal detection parameters can be estimated with the full utility of the generalized linear models.

B Generalized estimating equations – a motivation

In case of balanced experiments, i.e., experiments in which every participant takes part in every trial, the generalized linear model has to be extended. This extension, similar to extending the ANOVA with the repeated-measure ANOVA, is called generalized estimating equations. Here, the correlation between the participants is taken into account. The following introduce the concept following closely the notation of Hardin & Hilbe (2003, chapter 3).

There are two general approaches to the problem of non-independent data due to participants taking part in multiple conditions. The first approach is to model the conditional expectation $\mu_{it}^{PS} = E(y_{it}|\nu_i)$, for y_{it} being the outcome, $\underline{x}_{it}$ being the covariate vector corresponding to the parameter vector $\underline{\beta}$. The *PS* superscript stands for *participant-specific* model. Now, an additional covariate vector $\underline{z}_{it}$ is introduced with random effect $\underline{\nu}_i$. For a given panel (or in our case participant):

$$g(\mu_{it}^{PS}) = \underline{x}_{it}\underline{\beta}^{PS} + \underline{z}_{it}\underline{\nu}_i \qquad (.8)$$

Instead of focusing on the distribution of the random effects as the source for lacking independence, one can also consider the marginal expectation (population averaged (PA)) of the outcome, i.e., $\mu_{it}^{PA} = E(E(y_{it}|\nu_i))$. The responses are given by:

$$g(\mu_{it}^{PA}) = \underline{x}_{it}\underline{\beta}^{PA} \qquad (.9)$$

Both sets of parameters $\underline{\beta}^{PS}$ and $\underline{\beta}^{PA}$ have a distinct interpretation. In first case, there is a subject-specific interpretation of $\underline{\beta}$, while in the latter case, the general effect is modeled. For the estimation of the overall d' values the population average has to be estimated, and no full parameterization of the population distribution is necessary. The result focuses on the population-averaged view as the effect of the signal parameters on the overall subjective responses is the main issue (Gardiner et al., 2008)

Using the population-averaged model the estimator μ_{it} is linked to the predictor $\eta_{it} = x_{jit}\beta_j$ by $g(\mu_{it}) = \eta_{it}$ with $j = 1, \dots, p$ being the number of parameters, $i = 1, \dots, n$ being the number of panels, and $t = 1, \dots, n_i$ being the number of correlated observations within one panel. Now, the generalized estimating equations are set up. Starting from a quasi-likelihood:

$$\mathcal{Q}(y;\mu) = \int^{\mu} \frac{y-\mu^*}{V(\mu^*)a(\phi)} d\mu^*. \tag{.10}$$

To estimate the parameter set $\underline{\beta}$ the estimating equations are given by:

$$\Psi(\beta_i) = \frac{\partial \mathcal{Q}}{\partial \beta_i} = 0 \tag{.11}$$

By using the chain rule,

$$\frac{\partial \mathcal{Q}}{\partial \beta} = \frac{\partial \mathcal{Q}}{\partial \mu}\frac{\partial \mu}{\partial \eta}\frac{\partial \eta}{\partial \beta}, \tag{.12}$$

Eq. .11 becomes

$$\Psi(\underline{\beta}) = \left\{ \sum_{i=1}^{n} \sum_{t=1}^{n_i} \frac{y_{it}-\mu_{it}}{V(\mu_{it})a(\phi)} \left(\frac{\partial \mu}{\partial \eta}\right)_{it} x_{jit} \right\}_{j=1,\dots,p} = \underline{0}. \tag{.13}$$

To get the main idea the equations can be re-written in matrix notation:

$$\Psi(\underline{\beta}) = \left\{ \sum_{i=1}^{n} \begin{pmatrix} x_{ji1} \\ \vdots \\ x_{jin_i} \end{pmatrix}^{\dagger} \begin{pmatrix} \frac{\partial \mu}{\partial \eta} & & \\ & \ddots & \\ & & \frac{\partial \mu}{\partial \eta} \end{pmatrix}_i \left[\begin{pmatrix} V(\mu_{i1}) & & \\ & \ddots & \\ & & V(\mu_{in_i}) \end{pmatrix} \right]^{-1} \begin{pmatrix} \frac{y_{i1}-\mu_{i1}}{a(\phi)} \\ \vdots \\ \frac{y_{in_i}-\mu_{in_i}}{a(\phi)} \end{pmatrix} \right\}_{j=1,\dots,p} = \underline{0} \tag{.14}$$

The matrix $V(\mu_i)$ is clearly a diagonal matrix, which can be written:

$$V(\underline{\mu_i}) = \left[\begin{pmatrix} V(\mu_{i1}) & & \\ & \ddots & \\ & & V(\mu_{in_i}) \end{pmatrix} \right]^{1/2} \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} \left[\begin{pmatrix} V(\mu_{i1}) & & \\ & \ddots & \\ & & V(\mu_{in_i}) \end{pmatrix} \right]^{1/2} \tag{.15}$$

Now, it is obvious why the calculation treats every observation in each variable as independent within a panel: There are no off-diagonal elements in the variance matrix. Therefore, it is called the *independence model*. Now, this model is adjusted by replacing the identity matrix with a more general term. This is the matrix $\underline{\underline{R}}(\alpha)$, which is called the *working correlation matrix*. On the diagonal this matrix is still 1, while the off-diagonal elements can be adjusted as one sees fit. These are the parameters $\underline{\alpha}$. These values have to be estimated by additional estimating equations.

At last, one possible approach to the parameterization of the working correlation matrix is given. This is actually the approach used in this thesis:

$$R_{uv} = \begin{cases} 1 & \text{if u = v} \\ \alpha & \text{else} \end{cases} \tag{.16}$$

This parameterization is used for repeated measurements in which no explicit time dependency and any permutation of the experiment is valid. Only one parameter needs to be estimated which is the mean correlation in one panel.

C Signal analysis of a tone in narrowband noise

In the experiments in Chapter 5 tones in narrowband noise are investigated. Here, additional signal analyses are provided for the stimuli in Sec. 5.1. First, the envelope fluctuations of the signals are investigated. In a second step, the variations of the instantaneous frequencies are explored. The parameters of the stimuli are provided in Table 5.1. The values depicted in the Fig. C.1, C.2, and C.3 represent mean values off the three noise instantiations.

Envelope fluctuations

The envelopes' fluctuations are characterized by the relative standard deviation. Fig. C.1 depicts the envelopes' relative standard deviation for each stimulus used in the first experiment of Chapter 5. Almost the same relative standard deviation is obtained at each SNR ratio.

However, according to Fastl & Zwicker (2007) fluctuations above ca. 20 Hz have a minimal fluctuations strength. Therefore, the envelope fluctuations are reanalyzed with a 20 Hz low-pass filter applied to the envelope. In this way, only relevant fluctuations are taken into account – the envelope is smoothed. The results are shown in Fig. C.2. Taken the $-\infty$-condition as an example, the perceptual difference in fluctuation strength is better reflected in this filtered data than in the unfiltered data in Fig. C.1: 250 Hz NBN with a 50 Hz bandwidth has a higher perceived fluctuation than the 4 kHz stimulus with a 250 Hz bandwidth (Fastl & Zwicker, 2007). Comparing these differences between the different frequency condition to the differences due to different SNR, the influence of the tone on the fluctuation is rather small.

The values in Fig. C.2 can be compared to sinusoidal amplitude modulations of a sine carrier. A 4 Hz amplitude modulated tone with a modulation depth of -3 dB ($20 \cdot \log_{10}(m)$) has a relative standard deviation of its envelope of 0.50, while at -6 dB this value drops to 0.35. Comparing these values to the data in Fig. C.2 this depicts

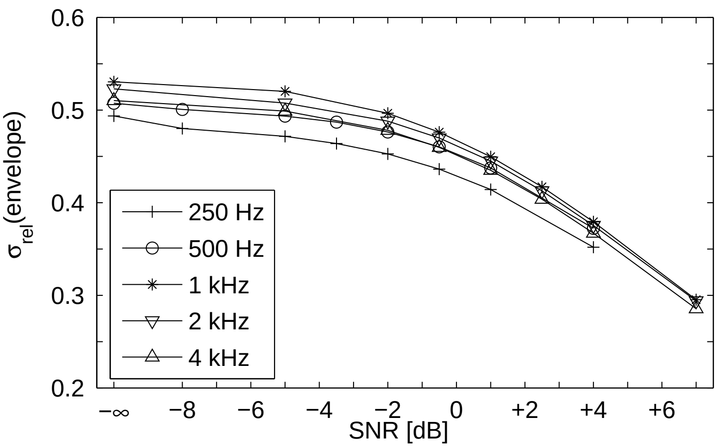

Figure C.1: *Envelopes' relative standard deviation of the stimuli from Table 5.1 over the SNR. All in all, the same standard deviations for each SNR condition is obtained.*

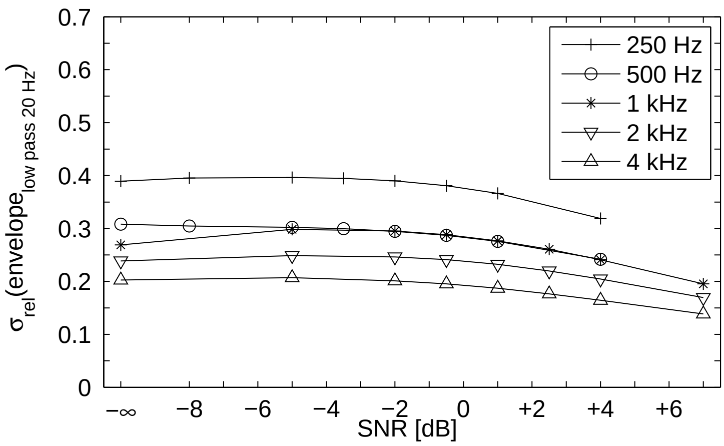

Figure C.2: *20 Hz low-pass-filtered envelopes' relative standard deviation of the stimuli from Table 5.1 over the SNR. Comparing the difference between the frequency conditions to the difference between different SNR conditions within one frequency condition, the influence of the SNR is small.*

roughly the range of the obtained relative standard deviations. Thus, difference in modulation depth of about 3 dB is obtained comparing the 250 Hz and the 4 kHz data. Again, the difference in *relevant* envelope fluctuation within one frequency condition is far smaller than this 3 dB modulation depth.

Instantaneous frequency fluctuations

The stimuli of the experiment described in Sec. 5.1 co-vary in center frequency f_c and bandwidth Δf. The center frequency f_c is 250, 500, 1k, 2k, and 4k Hz with a bandwidth Δf of 50, 100, 150, 200, and 250 Hz, respectively. The relative instantaneous frequency fluctuations vary not only with center frequency f_c, respectively bandwidth Δf, but also with the SNR (Fig. C.3). With larger bandwidth Δf, lower center frequency f_c, the fluctuations are increasing. With lower SNR, the fluctuations increase as the noise has a deteriorative influence on the estimation of the instantaneous frequency. Furthermore, within each frequency condition, the frequency fluctuations decrease by a factor of 3 up to a factor of 7.

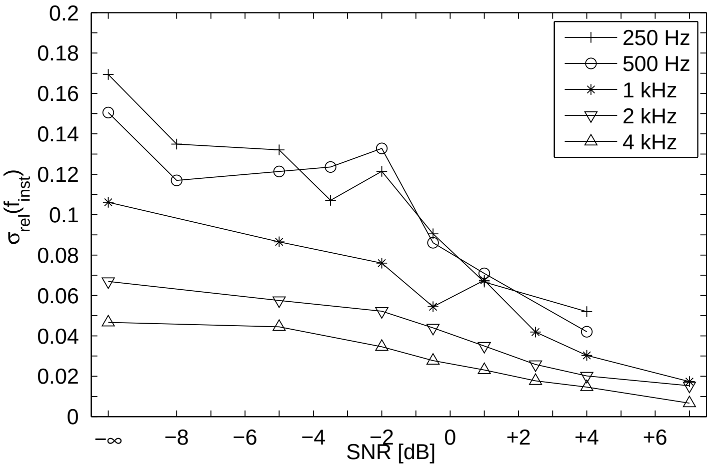

Figure C.3: *Instantaneous frequency fluctuations of the stimuli from Table 5.1 over the SNR. The measure is the relative standard deviation of the instantaneous frequency.*

Bibliography

Akeroyd, M. A., Moore, B. C. J., & Moore, G. A. (2001). Melody recognition using three types of dichotic-pitch stimulus. *J Acoust Soc Am*, 110(3), 1498–1504.

ANSI-S1.1-1994 (1994). Acoustical terminology. ASA, New York. American national standard.

ANSI S1.13-2005 (2005). Measurement of sound pressure levels in air. ASA, New York. American national standard.

Aures, W. (1985a). Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale [A model for calculating the sensory euphony of various sounds]. *Acustica*, 59, 130–41.

Aures, W. (1985b). Der sensorische Wohlklang als Funktion psychoakustischer Empfindungsgrößen [Sensory pleasantness as a function of psychophysical parameters]. *Acustica*, 58, 282–290.

Bendor, D. & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, 436, 1161–1165.

Berglund, B., Harder, K., & Preis, A. (1994). Annoyance perception of sound and information extraction. *J Acoust Soc Am*, 95(3), 1501–1509.

Berry, J., Poortinga, Y., Segall, M. H., & Dasen, P. R., Eds. (2002). *Cross-Cultural Psychology, Research and Applications*. Cambridge, UK: Cambrigde University Press, 2nd edition.

Bienvenue, R. & Nobile, M. A. (1991). Prominence ratio for noise spectra with discrete tones. In *Proc inter-noise '91* (pp. 53–56). Sydney, Australia.

Bilsen, F. A. (1976). Pronounced binaural pitch phenomenon. *J Acoust Soc Am*, 59(2), 467–468.

Blauert, J. & Jekosch, U. (1997). Sound-quality evaluation – a multi-layered problem. *Acta Acust united Ac*, 83, 747–753.

Bortz, J. & Doering, N. (2002). *Forschungsmethoden und Evaluation [Methods and evaluation]*. Berlin, Germany: Springer, 3rd edition.

Brochard, R., Drake, C., Botte, M., & McAdams, S. (1999). Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation. *J Exp Psychol Human*, 25(6), 1742–1759.

Buus, S., Müsch, H., & Florentine, M. (1998). On loudness at threshold. *J Acoust Soc Am*, 104(1), 399–410.

Cardozo, B. L. (1974). Some notes on frequency discrimination and masking. *Acustica*, 31, 330–336.

Cariani, P. A. & Delgutte, B. (1996). Neural correlates of the pitch of complex tones. i. pitch and pitch salience. *J Neurophysiol*, 76, 1698–1716.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor anova designs. *Educ Psychol Meas*, 33, 107–112.

Cramer, E. M. & Huggins, W. H. (1958). Creation of pitch through binaural interaction. *J Acoust Soc Am*, 30(5), 413–417.

Cureton, E. E. & D'Agostino, R. B. (1983). *Factor Analysis: An Applied Approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dai, H. (2000). On the relative influence of individual harmonics on pitch judgment. *J Acoust Soc Am*, 107(2), 953–959.

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. i. detection and masking with narrow-band carriers. *J Acoust Soc Am*, 102(5), 2892–2905.

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychol Methods*, 3(2), 186–205.

Derleth, R. P. & Dau, T. (2000). On the role of envelope fluctuation processing in spectral masking. *J Acoust Soc Am*, 108(1), 285–296.

DIN 45681:2005-03 (2005). Acoustics – Detection of tonal components of noise and determination of tone adjustment for the assessment of noise immisions. Beuth Verlag, Berlin, Germany. German industry standard.

Ernst, S. M., Uppenkamp, S., & Verhey, J. L. (2009). Cortical representation of release from auditory masking. *NeuroImage*, (pp. (in press)).

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

Fales, C. & McAdams, S. (1994). The fusion and layering of noise and tone: Implications for timbre in african instruments. *Leonardo Music J*, 4, 69–77.

Fastl, H. (1971). Über Tonhöhenempfindungen bei Rauschen [On pitch of noise]. *Acustica*, 25(6), 350–354.

Fastl, H. (1989). Pitch strength of pure tones. In *13th international congress on acoustics* (pp. 11–14). Yugoslavia.

Fastl, H. & Stoll, G. (1979). Scaling of pitch strength. *Hearing Res*, 1(4), 293–301.

Fastl, H. & Zwicker, E. (2007). *Psychoacoustics: Facts and Models*. Berlin, Germany: Springer, 3rd edition.

Fingerhuth, S. & Parizet, E. (2008). Tonalness perception of harmonic complex tones (A). *J Acoust Soc Am*, 5(5), 3685.

Fruhmann, M. (2004a). On the pitch strength of bandpass noises. In *18th International congress on acoustics, ICA 2004* (pp. 1791–1794).

Fruhmann, M. (2004b). On the pitch strength of harmonic complex tones and combfilter noises. In *Fortschritte der Akustik - CFA/DAGA '04* (pp. 757–758).

Fruhmann, M. (2006a). Ein Modell zur Ausgeprägtheit der Tonhöhe in Theorie und Praxis [A pitch strength model in theory and application]. In *Fortschritte der Akustik - DAGA '06* (pp. 315–316).

Fruhmann, M. (2006b). *Ein Modell zur Wahrnehmung der Ausgeprägtheit der Tonhöhe von Schallen [A pitch strength model]*. PhD thesis, Technische Universität München.

Fruhmann, M. & Kluiber, F. (2005). On the pitch strength of complete and incomplete harmonic complex tones. In *Fortschritte der Akustik - DAGA '05* (pp. 467–468).

Gardiner, J. C., Luo, Z., & Roman, L. A. (2008). Fixed effects, random effects and GEE: What are the differences? *Statist Med*, 28, 221–239.

Gockel, H., Moore, B. C. J., Plack, C. J., & Carlyon, R. P. (2006). Effect of noise on the detectability and fundamental frequency discrimination of complex tones. *J Acoust Soc Am*, 120(2), 957–965.

Grose, J. H. & Hall, J. W. (1997). Multiband detection of energy fluctuations. *J Acoust Soc Am*, 102(2), 1088–1096.

Guirao, M. & Stevens, S. S. (1964). Measurement of auditory density. *J Acoust Soc Am*, 36, 1176–1182.

Guski, R. (1997). Psychological methods for evaluating sound quality and assessing acoustic information. *Acta Acust united Ac*, 83, 765–774.

Gutschalk, A., Patterson, R. D., Rupp, A., Uppenkamp, S., & Scherg, M. (2002). Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. *NeuroImage*, 15, 207–216.

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., & Rupp, A. (2004). Temporal dynamics of pitch in human auditory cortex. *NeuroImage*, 22, 755–766.

Hall, D. & Plack, C. (2007). Searching for a pitch centre in human auditory cortex. In B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, & J. Verhey (Eds.), *Hearing - from sensory processing to perception* (pp. 83–89). Berlin, Germany: Springer.

Handel, S. (2006). *Perceptual Coherence: Hearing and Seeing*. Oxford University Press.

Hansen, H. & Weber, R. (2008). The influence of tone length and S/N-ratio on the perception of tonal content: An application of probabilistic choice models in car acoustics. *Acoust Sci & Tech*, 29(2), 156–166.

Hansen, H. & Weber, R. (2009). Semantic evaluations of noise with tonal components in Japan, France, and Germany - A cross-cultural comparison. *J Acoust Soc Am*, 125(2), 850–862.

Hansen, H. & Weber, R. (2010). Zum Verhältnis von Tonhaltigkeit und der partiellen Lautheit der tonalen Komponenten in Rauschen [The relationship of magnitude of tonal contents and partial loudness of tonal component in noise] . In *Fortschritte der Akustik - DAGA '10*.

Hansen, M. (2006). Lehre und Ausbildung in Psychoakustik mit psylab: freie Software für psychoakustische Experimente. In *Fortschritte der Akustik - DAGA '06* Oldenburg, Germany: DEGA.

Hardin, J. W. & Hilbe, J. M. (2003). *Generalized Estimating Equations*. Boca Raton, FL: Chapman & Hall/CRC.

Harris, J. D. (1948). Pitch-discrimination under masking. *Amer J Psychol*, 61(2), 194–204.

Harris, J. D. (1966). Masked DL for pitch memory. *J Acoust Soc Am*, 40(1), 43–46.

Harris, J. D. & Myers, C. K. (1949). The emergence of a tonal sensation. *J Exp Psych*, 39(2), 228–237.

Hartmann, W. M. & McMillon, C. D. (2001). Binaural coherence edge pitch. *J Acoust Soc Am*, 109(1), 294–305.

Hartmann, W. M. & Zhang, P. X. (2003). Binaural models and the strength of dichotic pitches. *J Acoust Soc Am*, 114(6), 3317–3326.

Heise, D. R. (1969). Some methodological issues in semantic differential research. *Psychol Bull*, 72(6), 406–422.

Hellman, R. P. (1982). Loudness, annoyance, and noisiness produced by single-tone-noise complexes. *J. Acoust. Soc. Am.*, 72(1), 62–73.

Henning, G. B. (1967). Frequency discrimination in noise. *J Acoust Soc Am*, 41(4A), 774–777.

Hofstede, G. (1991). Empirical models of cultural differences. In N. Bleichrodt & P. Drenth (Eds.), *Contemporary issues in cross-cultural psychology* (pp. 4–20). Amsterdam: Swets & Zeitlinger.

Hussain, M., Pflüger, M., Brandl, F., & Biermayer, W. (1998). Intercultural differences in annoyance response to vehicle interior noise. In *Proc euronoise 98* (pp. 521–526). Munich, Germany.

ISO 7779:1999 (1999). Acoustics – measurement of airborne noise emitted by information technology and telecommunications equipment. ISO, Genève, Switzerland. International standard.

Iwamiya, S. & Zhan, M. (1997). A comparison between Japanese and Chinese adjectives which express auditory impressions. *J. Acoust. Soc. Jpn. (E)*, 18(6), 319–323.

Kaernbach, C. & Demany, L. (1998). Psychophysical evidence against the autocorrelation theory of auditory temporal processing. *J Acoust Soc Am*, 104(4), 2298–2306.

Klein, M. A. & Hartmann, W. M. (1981). Binaural edge pitch. *J Acoust Soc Am*, 70(1), 51–61.

Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Percept Psychphys*, 63(8), 1421–1455.

Kohlrausch, A., Fassel, R., Heijden, M. v. d., Kortekaas, R., Par, S. v. d., Oxenham, A. J., & Puschel, D. (1997). Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations. *Acta Acust united Ac*, 83(11), 659–669.

Kohlrausch, A. & Houtsma, A. (1992). Pitch related to spectral edges of broadband signals. *Phil Trans R Soc Lond*, 336, 375–382.

Korth, B. (1978). A significance test for congruence coefficients for cattell's factors matched by scanning. *Multivar Behav Res*, 13, 419–430.

Korth, B. & Tucker, L. (1975). The distribution of chance congruence coefficients from simulated data. *Psychometrika*, 40(3), 361–372.

Kubovy, M. & van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, 80, 97–126.

Kubovy, M. & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: a quantitative gestalt theory. *Psychol Sci*, 6(4), 225–234.

Kuwano, S., Fastl, H., Namba, S., Nakamura, S., & Uchida, H. (2006). Quality of door sounds of passenger cars. *Acoust Sci & Tech*, 27(5), 309–312.

Kuwano, S., Namba, S., Kurakata, K., & Kikuchi, Y. (1994). Evaluation of broad-band noise mixed with amplitude-modulated sounds. *J Acoust Soc Jpn (E)*, 15(3), 131–142.

Kuwano, S., Namba, S., & Schick, A. (1986). A cross-cultural study on noise problems. In *Contributions to Psychological Acoustics* Oldenburg, Germany: BIS.

Kuwano, S., Namba, S., Schick, A., Hoege, H., Fastl, H., Filippou, T., & Florentine, M. (2007). Subjective impression of auditory danger signals in different countries. *Acoust Sci & Tech*, 28, 360–362.

Leibold, L. J., Tan, H., Khaddam, S., & Jesteadt, W. (2007). Contributions of individual components to the overall loudness of a multitone complex. *J Acoust Soc Am*, 121(5), 2822–2831.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*, 49(2), 467–477.

Macmillan, N. A. & Creelman, C. D. (2005). *Detection theory: a user's guide*. Hillsdale, NJ: Erlbaum, 2nd edition.

Marquis-Favre, C., Premat, E., Aubrée, D., & Vallet, M. (2005). Noise and its effects - a review on qualitative aspects of sound. part I: Notions and acoustic ratings. *Acta Acust united Ac*, 91, 613–625.

McCullagh, P. & Nelder, J. A. (1997). *Generalized Linear Models*. Monographs on Statistics and Applied Probability 37. London: Chapman & Hall, 2nd edition.

McLachlan, N. (2009). A computational model of human pitch strength and height judgments. *Hearing Res*, 249, 23–35.

Michaels, R. M. (1957). Frequency difference limens for narrow bands of noise. *J Acoust Soc Am*, 29(4), 520–522.

Moore, B. C. J. (1973a). Frequency difference limens for narrow bands of noise. *J Acoust Soc Am*, 54(4), 888–896.

Moore, B. C. J. (1973b). Frequency difference limens for short-duration tones. *J Acoust Soc Am*, 54(3), 610–619.

Moore, B. C. J., Alcántara, J. I., & Dau, T. (1998). Masking patterns for sinusoidal and narrow-band noise maskers. *J Acoust Soc Am*, 104(2), 1023–1038.

Moore, B. C. J. & Glasberg, B. R. (1989). Difference limens for phase in normal and hearing-impaired subjects. *J Acoust Soc Am*, 86(4), 1351–1365.

Moore, B. C. J., Glasberg, B. R., & Baer, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *J Audio Eng Soc*, 45(4), 224–240.

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *J Acoust Soc Am*, 77(5), 1853–1860.

Namba, S. & Kuwano, S. (1984). Psychological study on L_{eq} as a measure of loudness of various kinds of noises. *J Acoust Soc Jpn (E)*, 5(3), 135–148.

Namba, S., Kuwano, S., Hashimoto, T., Berglund, B., Rui, Z., Schick, A., Hoege, H., & Florentine, M. (1991a). Verbal expression of emotional impression of sound: A cross-cultural study. *J Acoust Soc Jpn (E)*, 12(1), 12–29.

Namba, S., Kuwano, S., Kinoshita, K., & Kurakata, K. (1992). Loudness and timbre of broad-band noise mixed with frequency-modulated sounds. *J Acoust Soc Jpn (E)*, 13(1), 49–58.

Namba, S., Kuwano, S., Schick, A., Aclar, A., Florentine, M., & Rui, Z. (1991b). A cross-cultural study on noise problems: Comparison of the results obtained in Japan, West Germany, the U.S.A., China and Turkey. *J Sound Vib*, 151(3), 471–477.

Osgood, C. E., May, W. H., & Miron, M. S. (1975). *Cross-cultural universals of affective meaning*. IL: University of Illinois Press.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The Measurement of Meaning*. IL: University of Illinois Press, new edition.

Patterson, R. D., Handel, S., Yost, W. A., & Datta, A. J. (1996). The relative strength of the tone and noise components in iterated rippled noise. *J Acoust Soc Am*, 100(5), 3286–3294.

Patterson, R. D., Yost, W. A., Handel, S., & Datta, A. J. (2000). The perceptual tone/noise ratio of merged iterated rippled noises. *J Acoust Soc Am*, 107(3), 1578–1588.

Plack, C. J., Oxenham, A. J., Fay, R. R., & Popper, A. N., Eds. (2005). *Pitch: Neural Coding and Perception*. Springer Handbook of Auditory Research. New York: Springer.

Plomp, R. (1967). Pitch of Complex Tones. *J Acoust Soc Am*, 41(6), 1526–1533.

Pollack, I. (1948). The atonal interval. *J Acoust Soc Am*, 20(2), 146–148.

Poortinga, Y. (1989). Equivalence of cross-cultural data: An overview of basic issues. *Int. J. Psychol.*, 24, 737–756.

Pressnitzer, D., Patterson, R. D., & Krumbholz, K. (2001). The lower limit of melodic pitch. *J Acoust Soc Am*, 109(5), 2074–2084.

Puschmann, S., Uppenkamp, S., Kollmeier, B., & Thiel, C. M. (2010). Dichotic pitch activates pitch processing centre in heschl's gyrus. *Neuroimage*, 49, 1641–1649.

Ritsma, R. J. (1967). Frequencies dominant in the perception of the pitch of complex sounds. *J Acoust Soc Am*, 42(1), 191–198.

Santurette, S. & Dau, T. (2007). Binaural pitch perception in normal-hearing and hearing-impaired listeners. *Hearing Res*, 223(1-2), 29–47.

Schick, A. & Hoege, H. (1996). Cross-cultural psychoacoustics. In *Recent Trends in Hearing Research, Festschrift for Seiichiro Namba* chapter 12, (pp. 287–315). Oldenburg, Germany: BIS.

Schwartz, D. & Purves, D. (2004). Pitch is determined by naturally occurring periodic sounds. *Hearing Res*, 194, 31–46.

Semal, C. & Demany, L. (1990). The upper limit of musical pitch. *Music perception*, 8, 165–176.

Shofner, W. P. & Selas, G. (2002). Pitch strength and Stevens's power law. *Percept Psychphys*, 64(3), 437–450.

Sinnott, J. M. & Brown, C. H. (1993). Effects of varying signal and noise levels on pure-tone frequency discrimination in humans and monkeys. *J Acoust Soc Am*, 93(3), 1535–1540.

Smith, B. (1995). PsiExp: an environment for psychoacoustic experimentation using the IRCAM musical workstation. In *Proc Society for Music Perception and Cognition Conference '95* Berkeley, CA: University of California.

Solomon, L. (1958). Semantic approach to the perception of complex sounds. *J Acoust Soc Am.*, 30(5), 421–425.

Stevens, C. (2004). Cross-cultural studies of musical pitch and time. *Acoust Sci & Tech.*, 25(6), 433–438.

Stevens, S. S. (1934). The attributes of tones. *Proc Natl Acad Sci USA*, 20, 457–459.

Stevens, S. S. (1957). On the psychophysical law. *Psych Rev*, 64, 153–181.

Terhardt, E. (1974). Pitch, consonance, and harmony. *J Acoust Soc Am*, 55(5), 1061–1069.

Terhardt, E. (1998). *Akustische Kommunikation [Acoustic Communication]*. Berlin, Germany: Springer-Verlag.

Terhardt, E. & Stoll, G. (1981). Skalierung des Wohlklangs von 17 Umweltschallen und Untersuchung der beteiligten Hörparameter. *Acustica*, 48, 247–253.

Terhardt, E., Stoll, G., Schernbach, R., & Parncutt, R. (1986). Tonhöhenmehrdeutigkeit, Tonverwandschaft und Identifikation von Sukzessivintervallen [Pitch ambiguity, tone affinity, and identifitazion of successive intervals]. *Acustica*, 61, 57–66.

Terhardt, E., Stoll, G., & Seewann, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *J Acoust Soc Am*, 71(3), 679–688.

van Valkenburg, D. & Kubovy, M. (2004). From Gibson's fire to Gestalts – a bridge-building theory of perceptual objecthood. In J. G. Neuhoff (Ed.), *Ecological Psychoacoustics*. San Diego, CA: Elsevier.

v.d. Vijver, F. (1998). Towards a theory of bias and equivalence. In J. Harkness (Ed.), *Cross-Cultural Survey Equivalence* (pp. 41–65). Mannheim, Germany: ZUMA.

v.d.Vijver, F. & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.

Verhey, J. & Heise, S. (2010). Einfluss der Zeitstruktur des Hintergrundes auf die Tonhaltigkeit und Lautheit des tonalen Vordergrundes [The influence of the noise background's temporal stucture on the judgement of tonal content's magnitude and the loudness of the tonal foreground]. In *Fortschritte der Akustik - DAGA '10*.

von Bismarck, G. (1984). Sharpness as an attribute of timbre of steady sounds. *Acta Acust*, 30, 159–172.

von Helmholtz, H. (1885/1954). *On the sensation of tone*. New York, NY: Dover. (A. J. Ellis, Trans., orginal work published 1877).

Vormann, M., Meis, M., Mellert, V., & Schick, A. (1999a). A new approach for the evaluation of tonal noise (tonality). In T. Dau, V. Hohmann, & B. Kollmeier (Eds.), *Psychophysics, Physiology and Models of Hearing* (pp. 109–112). Singapore, Singapore: World Scientific, 1. edition.

Vormann, M., Schick, A., Meis, M., Klatte, M., & Mellert, V. (1998). Tonhaltigkeit aus psychophysikalischer Sichtweise: Untersuchung zur Versuchspersoneninstruktion bei der Bewertung der Tonhaltigkeit. In *Fortschritte der Akustik - DAGA '98* Oldenburg, Germany: DEGA.

Vormann, M., Verhey, J., Mellert, V., & Schick, A. (1999b). Factors influencing the subjective rating of tonal components in noise. In *Fortschritte der Akustik - DAGA '99* Oldenburg, Germany: DEGA.

Vormann, M., Verhey, J., Mellert, V., & Schick, A. (2000a). Ein adaptives Verfahren zur Bestimmung der subjektiven Tonhaltigkeit. In *Fortschritte der Akustik - DAGA 2000* Oldenburg, Germany: DEGA.

Vormann, M., Verhey, J. L., Mellert, V., & Schick, A. (2000b). Subjective ratings of tonal components in noise with an adaptive procedure. In A. Schick, M. Meis, & C. Reckhardt (Eds.), *Contributions to Psychological Acoustics, Results of the 8th Oldenburg Symposium on Psychological Acoustics* (pp. 145–153). Oldenburg: BIS-Verlag.

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proc Natl Acad Sci USA*, 100(17), 10038–10042.

Watkins, D. (1989). The role of confirmatory factor analysis on cross-cultural research. *Int J Psychol*, 24, 685–701.

Weber, R., Brosig, D., & Verhey, J. L. (2009). The tonality salience of pure tones as function of frequency. In *Fortschritte der Akustik - DAGA '09*.

Wier, C. C., Jesteadt, W., & Green, D. M. (1977). Frequency discrimination as a function of frequency and sensation level. *J Acoust Soc Am*, 61(1), 178–184.

Wightman, F. L. (1973). Pitch and stimulus fine structure. *J Acoust Soc Am*, 54(2), 397–406.

Yost, W. A. (1996a). Pitch of iterated rippled noise. *J Acoust Soc Am*, 100(1), 511–518.

Yost, W. A. (1996b). Pitch strength of iterated rippled noise. *J Acoust Soc Am*, 100(5), 3329–3335.

Yost, W. A. (1997). Pitch strength of iterated rippled noise when the pitch is ambiguous. *J Acoust Soc Am*, 101(3), 1644–1648.

Yost, W. A. & Hill, R. (1978). Strength of the pitches associated with ripple noise. *J Acoust Soc Am*, 64(2), 485–492.

Yost, W. A., Hill, R., & Perez-Falcon, T. (1978). Pitch and pitch discrimination of broadband signals with rippled power spectra. *J Acoust Soc Am*, 63(4), 1166–1173.

Yost, W. A., Mapes-Riordan, D., Shofner, W., Dye, R., & Sheft, S. (2005). Pitch strength of regular-interval click trains with different length "runs" of regular intervalls. *J Acoust Soc Am*, 117(5), 3054–3068.

Yost, W. A., Patterson, R., & Sheft, S. (1996). A time domain description for the pitch strength of iterated rippled noise. *J Acoust Soc Am*, 99(2), 1066–1078.

Zimmer, K., Ellermeier, W., & Schmid, C. (2004). Using probabilistic choice models to investigate auditory unpleasantness. *Acta Acust united Ac*, 90, 1019–1028.

Zwicker, E. & Fastl, H. (1999). *Psychoacoustics - Facts and Models*. Berlin, Germany: Springer, 2nd updated edition.

Zwicker, E. & Feldtkeller, R. (1967). *Das Ohr als Nachrichtenempfänger*. Stuttgart, Germany: Hirzel Verlag.

Danksagung

Bei den Mitgliedern der Prüfungskommission, Prof. Volker Mellert, Prof. Jesko Verhey und Prof. Martin Holthaus bedanke ich mich zuerst. Herausheben möchte ich an dieser Stelle Prof. Martin Holthaus, der sehr kurzfristig trotz knapper Zeit für die Disputation zur Verfügung stand.

Besonderer Dank geht an Prof. Volker Mellert und Prof. Martin Holthaus, die maßgeblich dafür verantwortlich waren, dass ich bei der Studienstiftung des dt. Volkes als Promotionsstipendiat aufgenommen wurde, welches eine ortsunabhängige Finanzierung der Promotion möglich machte.

Mein herzlicher Dank geht an die Mitglieder der Arbeitsgruppe Akustik, im Speziellen an Dr. Reinhard Weber für all die Diskussionen und die kritischen Kommentare zu meinen Textproduktionen, an meine »ZellengenossInnen« Dr. Michael Bellmann, Jaqueline Rausch und Christopher Haut für die gute Zeit, und an Nils Freese für die Einführung in alles was mit Computern, Formaten und IT-Sicherheit zusammenhängt. Aus der Gruppe Neuroakustik geht mein Dank an Prof. Jesko Verhey für die interessanten Diskussionen und an Dr. Bastian Epp für die Organisation der Sommerschule »Two Ears, two eyes, one system?« und für die Zusammenarbeit bei der Gründung der DEGA-Studentengruppe sowie der EAA-Studentengruppe.

Ein Dank geht an alle VersuchsteilnehmerInnen, die so tapfer an den Versuchen teilgenommen haben.

Ein ganz herzlicher Dank geht auch an die Mitglieder der Arbeitsgruppe »Design et Perception sonore« für die Aufnahme in die Gruppe und die schöne, interessante Zeit am IRCAM. Speziell möchte ich mich bei Dr. Guillaume Lemaitre für die Hilfe bei den Experimenten und den Diskussionen über Statistik bedanken.

Bedanken möchte ich mich auch bei Cara H. Kahl, für die Einführung in die Psychologie, das Korrekturlesen, zahlreiche Diskussionen und die freundschaftliche Unterstützung, sowie die intensive Zusammenarbeit bei unseren Projekten, sei es der Kita-Artikel oder die Kreativitätsforschung.

Prof. Sonoko Kuwano und Prof. Seiichiro Namba für die Aufnahme in ihr Labor, welches den interkulturellen Vergleich erst möglich machte. In diesem Zusammenhang geht der Dank auch an Prof. August Schick, der die Anregung gab, nach Japan zu gehen.

Am Schluss möchte ich mich bei meiner Familie bedanken, bei Berit für die liebevolle Unterstützung und bei Willem für die Einhaltung seiner ersten Deadline, bei meinen Eltern, welche mich in meiner Ausbildung so großzügig in aller Freiheit unterstützt haben und nicht zuletzt bei meinem Bruder.

Publikationen

Veröffentlichungen

H. Hansen & R. Weber (eingereicht). *Partial loudness as a measure of the magnitude of tonal content* Acoust. Sci. & Tech.
H. Hansen, J.L. Verhey & R. Weber (eingereicht). *Pitch strength and the magnitude of tonal content - a review.* Acta Acust. united Ac.
C. Kahl & H. Hansen (in Vorbereitung). Lärm als Gesundheitsbelastung. In T. Kliche & J. Töppich (Hrsg.), *Handbuch der Prävention und Gesundheitsförderung in Kitas. Stand, Grundlagen und Praxisansätze.*
H. Hansen & R. Weber (2009). *Cross-cultural Evaluation of Perceptual Dimensions of Noises with Tonal Components in Japan, France, and Germany.* J. Acoust. Soc. Am., 125, 850-862.
H. Hansen & R. Weber (2008). *The influence of tone length and S/N-ratio on the perception of tonal content: An application of probabilistic choice models in car acoustics.* Acoust. Sci. & Tech., 29, 156-166.

Vorträge & Konferenzbeiträge

C. Kahl & H. Hansen (2010). *Modeling Creativity from a Systems Perspective: CRESY-I* 3rd World Congress on Social Simulation, Kassel, Deutschland.
H. Hansen & R. Weber (2010) *Partial loudness as a tonal content measure.* Internoise '10, Lissabon, Portugal.
H. Hansen & R. Weber (2010) *Zum Verhältnis von Tonhaltigkeit und der partiellen Lautheit der tonalen Komponenten in Rauschen* DAGA '10, Berlin.
H. Hansen & R. Weber (2010) *Die Ausgeprägtheit der Tonhöhe, das Aufösungsvermögen des Gehörs & und die Bildung auditorischer Objekte* DAGA '10, Berlin.
H. Hansen & R. Weber (2009) *Pitch strength, tone segregation, and frequency difference limen (A).* J. Acoust. Soc. Am., 2009, 126, 2242.
H. Hansen & R. Weber (2009). *Item bias in the cross-cultural evaluation of tonal content.* Internoise '09, Ottawa, Kanada.
H. Hansen & R. Weber (2009). *Tone Identification and Detection in Narrowband Noise.* NAG/DAGA '09, Rotterdam, Holland.
C. Kahl & H. Hansen (2008). *Modeling an Environment to Facilitate Creativity.* Proc 39th European Mathematical Psychology Group Meeting (EMPG), Graz, Österreich.
R. Weber & H. Hansen (2008). *Intercultural comparison of the perception of tonal components in noise.* Internoise '08, Istanbul, Türkei.

H. Hansen & R. Weber (2008). *Fusion and masking threshold of a tone in narrow-band noise (A).* J. Acoust. Soc. Am., 123, 3162.

R. Weber & H. Hansen (2008). *Sound character ratings by experienced and naive listeners (A).* J. Acoust. Soc. Am., 123, 3244.

H. Hansen & R. Weber (2007). *Comparison of Perceptual Dimension in Japan, France, and Germany (Poster).* 2nd International VDT Symposium – Sound Design, Stuttgart.

H. Hansen, R. Weber & V. Mellert (2006). *Pitch salience of weak tonal components with variable durations and levels in noise.* Internoise 2006, Honolulu, USA.

R.Weber & H. Hansen (2006). *Perception of noises with tonal components in Japan, France, and Germany.* Internoise '06, Honolulu, USA.

H. Hansen, R. Weber, U. Letens (2005). *Quantifying tonal phenomena in interior car sounds.* Forum Acusticum 2005, Budapest, Ungarn.

H. Hansen, R. Weber, U. Letens (2005). *Subjektive Bewertung der Tonhaltigkeit in der Fahrzeugakustik.* DAGA '05, München.

H. Hansen, R. Weber, T. Matsui, S. Kuwano, S. Namba (2004). *Evaluation of sound quality affected by tonal components in Germany and Japan.* CFA/DAGA '04, Strassburg, Frankreich.

M. Morinaga, H. Hansen, R. Weber, T. Matsui, S. Kuwano, S. Namba (2004). *The effect of virtual pitch on sound quality.* The proceedings of the autumn meeting of the Acoustical Society of Japan, 445-446.

H. Hansen, R. Weber, T. Matsui, S. Kuwano, S. Namba (2003). *The effect of tonal components on sound quality.* Acoustic Society of Japan, Research Council of Noise and Vibration, N-2003-08.

Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen verwendet habe.

Oldenburg, den 1.4.2010

Lebenslauf

gemäß Paragraph 16, Absatz 3, der Fakultät für Mathematik und Naturwissenschaften der Carl von Ossietzky Universität Oldenburg (11.12.2003)

Name:	Hans Hansen
geboren am:	29.05.1979
in:	Hamburg
Staatsangehörigkeit:	Deutsch

Bildungsgang:

08/1990 – 07/1998	Friedrich-Ebert-Gymnasium, Hamburg
3.07.1998	Erlangung der Allgemeinen Hochschulreife
07/1998 – 07/1999	Zivildienst
10/1999 – 10/2001	Studium der Physik an der Fakultät für Physik der Julius-Maximilians-Universität, Würzburg (Abschluss Vordiplom)
10/2001 – 09/2005	Studium der Physik am Institut für Physik der Carl von Ossietzky Universität, Oldenburg
10/2002 – 08/2003	Austauschstudent an der Osaka Universität (AIEJ Stipendium, Labor Prof. S. Kuwano)
06/2004 – 06/2005	Diplomarbeit in Kooperation mit der Daimler AG (Dr. Letens) zur Bewertung tonhaltiger Geräuschanteilen im Fahrzeuginnenraum
10/2005 – 12/2005	Freiwilligenarbeit im Chiiori-Projekt, Shikoku, Japan zur Erhaltung kultureller Techniken und Lebensweisen des ländlichen Japans
03/2006 – 09/2006	Wissenschaftlicher Mitarbeiter im Kato Tp 04 Projekt zu Flugzeugkabinenkomfort
10/2006 – 06/2010	Promotion am Institut für Physik der Carl von Ossietzky Universität, Oldenburg
10/2006 – 12/2006	Promotionsstipendiat der *International Graduate School of Neurosensory Science and Systems* (derzeit assoziiertes Mitglied)
01/2007 – 09/2009	Promotionsstipendiat der *Studienstiftung des deutschen Volkes*
08/2007 – 08/2008	Forschungsaufenthalt in der Gruppe *Perception et Design Sonores*, IRCAM, Paris
seit 9/2008	Freiwilliger Projektmitarbeiter am Lehrstuhl Sozialpsychologie, Universität Hamburg: Agentenbasierte Modellierung von Kreativität
ab 10/2009	Studium M.A. Philosophie – Philosophie im europäischen Kontext (Fernuniversität Hagen, Teilzeit).