Stephan Töpken

Sound Characteristics of Multi-Tone Sounds

Measurement Concepts and Model Approach

BIS-Verlag der Carl von Ossietzky Universität Oldenburg

Oldenburg, 2017

Verlag / Druck / Vertrieb

BIS-Verlag
der Carl von Ossietzky Universität Oldenburg
Postfach 2541
26015 Oldenburg
E-Mail: bisverlag@uni-oldenburg.de

ISBN 978-3-8142-2356-8

Preface

This book presents a series of studies on the quality and perception of multiple pitched complex tones. Besides being of basic scientific interest, these types of signals are relevant in various industrial applications as well as in aircraft engines equipped with counter rotating open rotors (CROR). The signature of this CROR "noise" can be characterized as such a multiple pitched complex tone which includes also interaction tones. With such technical applications becoming more relevant, the question also arises how such sounds will be appreciated by users. Although multiple pitch signals have been investigated extensively in the context of music, much less is known about contexts in which these signals are in principle un-desired side effects of the technical application. In this book a number of studies are presented that give insight in the perceptual effects of such multiple pitched tone complexes, both in terms of perceptual attributes as well as in terms of how their preference depends on specific signal parameters such as pitch ratio and spectral envelope. In these studies several methodologies are used to gain this insight. In the first experiments, using a semantic differential, the new finding emerges that preferably the ratio of fundamental frequencies of the two complex tones should not be a ratio of low integer number; rather is it preferred to have a deviating frequency ratio that leads to a signal that inherently has a long time span in which it repeats itself (repetition time). Of specific interest is also the methodology that is used to measure preferences of the listeners. Using a method that adaptively adjusts levels of different signals that are compared, until they are equality preferred (at the Point of Subjective Equality) allows to quantify subjective preference in decibels. Thus this method allows to make a direct link between the technical domain, where signals can be quantified according to their physical sound pressure level and the subjective domain of preference as experienced by the listeners. The book concludes with a study that presents a metric based on a predictor for perceived sharpness and the repetition time which is able to model a wide range of subjective preference data obtained with various parameter manipulations. I am confident that the reader who is interested in the perception and quality of complex sounds will find numerous interesting results and insights in this book.

Prof. Dr. Ir. Steven van de Par, November 2016

Summary

Tonal sounds are a part of the environmental noise that humans perceive every day. They often originate from rotating parts in machinery like for example fans, turbines and engines. In general, noise containing stationary audible tonal components is perceived louder than without tonal components and, in addition to that, a higher judged annoyance or unpleasantness is often observed. This increased nuisance from noise containing distinct tonal components is also addressed by national and international standards on noise immission by level adjustments. However, if the number of the tonal components is strongly increased, then the quality of the resulting sound changes considerably compared to a pure single tone and it is unclear how the resulting complex sound character and the appraisal of such sounds is affected. In addition, the identification of relevant signal parameters becomes a challenge in the considerably increased parameter space, especially for signals consisting of aperiodically spaced partials, e.g. due to non-linearities in a sound generation mechanism. In order to reduce adverse effects of such sounds, it is necessary to understand the underlying perceptual aspects driving the unpleasantness and annoyance of multi-tone sounds.

The overall aim of this PhD-thesis is the characterization and quantification of the sound character and the appraisal resulting from temporal and spectral characteristics of a particular type of multi-tone sounds. The sounds of interest consist of two complex tones and additional combination tones and they can be regarded as generic representatives of signals produced by two sound sources which are coupled by a non-linear interaction. Such signals are found in machinery noise of counter-rotating open rotors and also as so called multiphonics in the sounds of music instruments. In this PhD-thesis, the influence of specific signal parameters, having a direct equivalent in a potential technical application, on the assessment of the sounds is addressed.

In a first study, the perceptual space of multi-tone sounds is explored for a variation of the frequency ratio between the two fundamentals of the complex tones, which modifies the spectral spacing of all higher partials and influences the overall character of the sound. The factorial analysis of the results from a semantic differential gives insight into the structure of the perceptual space and the underlying perceptual dimensions. The factor explaining most of the variance is found to be related

to the pleasantness and the loudness of the sounds and a link to the perceived temporal structure becomes apparent. It turns out that the repetition rate of the signals, which is the inverse of the signal's periodicity, is a suitable descriptor for the temporal signature of the multi-tone sounds which can be related to the pleasant dimension of the semantic differential. Low values of the repetition rates, which are equivalent to long periodic times, are judged considerably more pleasant than high repetition rates or short periodic times.

In a follow-up study, a special measurement method was used for a detailed quantification of the preference relevant sound characteristics for a broader variety of signal parameters. In extensive listening tests, the loudness and the preference were each measured as levels of the multi-tone sounds at the points of subjective equality (PSE) compared to the same, fixed reference sound. In this way, the closely related loudness judgments and preference evaluations can be quantified and differentiated on a dB-scale. The level difference between the two PSEs is attributed to the sound character and it is used as a quantitative measure describing the supplementary contribution of the sound character to the preference evaluation, which comes on top of the loudness judgment.

The variability in the PSEs for preference resulting from differences in the spectral composition can be related to the psychoacoustic sharpness according to the DIN 45692 standard. High sharpness values are linked to low PSEs for preference, meaning that the multi-tone sounds with high sharpness values need to be considerably reduced in level to become equally preferred. The influence of the temporal structure of the signals on the preference evaluations are effectively covered by the repetition rate of the time signals, which was already identified in the semantic differential experiment as an important factor related to the (un-)pleasantness. A prediction model for the preference evaluations, which describes a plane of equal preference as a function of these two descriptor variables, explains 87 % of the variance in the PSE values. The fitted parameters of the model disclose the quantitative relationships between the preference equivalent dB(A)-level and the sound character of the multi-tone sounds, which is characterized by the sharpness and the repetition rate of the signals.

Zusammenfassung

In vielen Geräuschen des täglichen Lebens finden sich tonale Komponenten, die ihren Ursprung häufig in rotierenden Maschinenteilen, wie Ventilatoren, Turbinen, Motoren und Triebwerke haben. Im Allgemeinen tragen stationäre tonale Komponenten in Umweltgeräuschen zu einer erhöhten Unangenehmheit oder Lästigkeit aufgrund der Geräusche bei. Diese erhöhte Störwirkung von tonalen Komponenten ist entsprechend auch in Normen zur Bewertung von Schallimmisionen in Form von Pegelzuschlägen berücksichtigt. Für Geräusche mit einer Vielzahl von tonalen Komponenten ändert sich der Klangcharakter im Vergleich zu einem einzelnen Ton jedoch deutlich und es ist unklar inwiefern die Beurteilung eines solchen Geräusches beeinträchtigt wird. Hinzu kommt, dass eine Identifizierung bewertungsrelevanter Signalparameter innerhalb eines deutlich größeren Parameterraums eine Herausforderung darstellt, insbesondere für Signale deren Teiltöne aperiodisch im Spektrum verteilt sind. Um potentielle negative Auswirkungen solcher Geräusche reduzieren zu können, ist es notwendig die der Unangenehmheit oder Lästigkeit zu Grunde liegenden perzeptiven Aspekte zu verstehen.

Das Ziel dieser Arbeit ist es, den Klangcharakter einer speziellen Art von Multitonsignalen zu charakterisieren und den Einfluss von präferenzrelevanten, zeitlichen und spektralen Aspekten auf den Klangcharakter zu quantifizieren. Die untersuchten Multitonsignale bestehen aus zwei Klängen und zusätzlichen Kombinationstönen und sind somit representativ für Signale zweier harmonischer Klangquellen, welche über eine nichtlineare Interaktion miteinander gekoppelt sind. Beispiele für Signale dieser Art finden sich sowohl bei Maschinen mit gegenläufigen offenen Rotoren als auch bei Musikinstrumenten in Form von sogenannten Mehrklängen ("Multiphonics"). Im Rahmen dieser Arbeit wird insbesondere der Einfluss technisch relevanter Signalparameter auf die Beurteilung der Geräusche betrachtet.

In einer ersten Studie wurde der Wahrnehmungsraum von Multitonsignalen für eine Variation des Frequenzverhältnisses zwischen den Grundtönen der Klänge, welches sämtliche spektralen Abstände aller höheren Teiltöne beeinflussen, exploriert. Die faktoranalytische Auswertung der Daten aus einem semantischen Differential ermöglicht einen Einblick in die Struktur des Wahrnehmungsraums und der ihm zugrundeliegenden, latenten Wahrnehmungsdimensionen. Es zeigt sich, dass der Faktor mit

der höchsten Varianzaufklärung im Beziehung zu der Angenehmheit und der Lautheit der Geräusche steht, wobei eine Verbindung zu einer weiteren Wahrnehmungsdimension, die die zeitliche Struktur beschreibt, erkennbar wird. Für die Beschreibung der zeitlichen Signatur der Multitonsignale eignet sich die Wiederholrate des Zeitsignals, also die Inverse der Periodizität, welche sich in Beziehung mit der Angenehmheitsdimension aus dem semantischen Differential setzen lässt. Niedrige Werte der Wiederholrate, welche äquivalent zu langen Periodendauern sind, wurden als deutlich angenehmer bewertet als hohe Wiederholraten bzw. kurze Periodendauern.

In einer Folgestudie wurde ein spezielles Messverfahren für die detaillierte Quantifizierung präferenzrelevanter Klangcharackteristiken für eine größere Bandbreite variierter Signalparameter eingesetzt. In umfangreichen Hörexperimenten, wurden die Lautheit und die Präferenz der Multitonsignale als Punkte subjektiver Gleichheit (englisch: Point of subjective equality, PSE), jeweils im Vergleich zu dem selben, konstanten Referenzgeräusch gemessen und quantitativ unterscheidbar auf einer dB-Skala abgebildet. Der Pegelunterschied zwischen den PSEs der zwei Urteile wird verstanden als der zusätzlichen Beitrag des Klangcharakters zu dem Präferenzurteil, welcher über die reine Lautheit hinausgeht. Entsprechend stellt dieser Pegelunterschied ein quantitatives Einzahlmaß für den Einfluss des Klangcharakters auf das Präferenzurteil dar.

Die Variabilität der Präferenzurteile, die aus Unterschieden in der spektralen Zusammensetzung der Signale hervorgeht, wird von der psychoakustischen Schärfe gemäß der DIN 45692 am besten wiedergespiegelt. Hohe Werte der Schärfe sind verknüpft mit niedrigen PSEs, was bedeutet, dass Geräusche mit hoher Schärfe deutlich im Pegel reduziert werden müssen, um gleich präferiert zu werden. Der Einfluss der zeitlichen Struktur auf die Präferenzurteile wird effektiv von der Wiederholrate abgebildet, die schon mit dem Semantischen Differential als potentielle Deskriptorvariable identifiziert wurde. Ein Vorhersagemodell der Präferenzurteile, welches eine Fläche gleicher Präferenz als Funktion dieser zwei Deskriptorvariablen beschreibt, erklärt 87 % der Urteilsvarianz. Die angepassten Parameter des Modells offenbaren dabei die quantitativen Zusammenhänge zwischen dem präferenzäquivalenten dB(A)-Pegel und dem Klangcharakter der Multitonsignale, beschrieben durch die psychoakustische Schärfe und die Wiederholrate.

Contents

1 Introduction

Sounds in everyday life often contain tonal components. Besides voices and music instrument, there are a lot of technical devices emitting rather stationary tonal sounds. Examples can be found in nearly all machines with rotating parts, like for example fans, compressors, internal combustion engines, gears, turbines, etc.

When perceived by human listeners, the resulting rather stationary tonal sounds are often found to be related to a higher perceived loudness, noisiness and annoyance, compared to broadband sounds without salient tonal components. This was found for single tones in background noise[(105,133,178)] as well as for sounds with multiple tonal components[(178,221)]. This increased nuisance effect of noise containing discrete tonal components is also addressed in German and international standards by tone adjustments to be included in the rating levels of noise immission[(52,53,120)].

However, for sounds containing a large number of tonal components, the resulting sound character is quite different compared to a single pure tone. Already the change of the spectral shape of a single complex tone, consisting of a fundamental frequency and higher harmonics, creates many degrees of freedom in varying the timbre of the complex tone. This is illustrated by several studies that systematically varied specific aspects of harmonic complex tones, like the overall spectral shape[(141,237)], the cut-off frequency of low-pass filtered harmonic complex tones[(19)] and the partial content of bandpass filtered harmonic complex tones[(20)]. Due to the high number of tonal components, it also becomes rather difficult to differentiate whether a signal parameter is related to the sound character, the loudness or contributing to both aspects of a sound.

Even more intricate are multi-tone signals which have an aperiodic spacing of the partials due to a non-linear interaction between two harmonic sound sources. Such sounds, consisting of two complex tones and additional combination tones, occur for example in machinery noise of counter-rotating open rotors [96]. Counter-rotating open rotors were investigated as aircraft engines during the 1980s, because of the high aerodynamic efficiency and potential reduction of fuel burn. At that time, the emitted noise was a major problem and numerous studies on the annoyance effects of simulated fly-over noise were conducted [148–151]. Nowadays, technologies for the computationally demanding acoustic optimization of the sound sources have improved and the inherently high aerodynamic efficiency renewed the interest in this type of engines [255]. However, a general understanding of the acoustic factors affecting the perception and the appraisal of the sounds is not available and target values for the optimization of the resulting sounds in the cabin and the acceptability of fly-over noise need to be defined.

The particular spectral signature, consisting of two complex tones and combination tones, can also be found in a musical context. Even though the context, the expectations and the overall appraisal of complex tones are quite different in the context of music, the perception of timbre from a musical perspective might help for the characterization of the particular sound character of these sounds. Especially for wind instruments, a rise in excitation pressure can lead to so called multiphonics in the transition domain from periodic oscillations to a chaotic regime [82], which resembles the above mentioned spectral structure. However, the sound generating mechanisms and the perceived timbre of saxophone multiphonics are not fully understood yet and still subject to current research [18,160,194].

This PhD-Thesis investigates the perception and appraisal of generic multi-tone sounds which are representative for a combination of two harmonic sound sources and a non-linear coupling between the two. The investigated signals consist of two complex tones and additional combination tones, which result from the non-linear interaction. The high number of tonal components and the complex spectral structure of the occurring frequency components results in a large number of signal parameters for this type of multi-tone sounds. It is unclear, which signal parameters are primarily relevant for the perceived sound character and how the sound character can be described by (psycho-)acoustic descriptors. It is further of interest to understand how the loudness and the sound character contribute to the overall appraisal of this type of multi-tone sounds.

1.1 Aim and scope of the work

The overall aim of this work is a characterization of the perceptual effects, resulting from temporal and spectral sound characteristics of multi-tone sounds, which consist of two harmonic complex tones and additional combination tones, and to understand how these characteristics influence the appraisal of such sounds.

The intricate spectral signature of this type of multi-tone sounds is essentially defined by the two fundamental frequencies of the two complex tones. While their absolute frequency values are anticipated to be mainly related to the resulting pitch, the frequency ratio between the two fundamentals influences the frequency spacing of all higher partials of the complex tones and of the combination tones. Therefore, the ratio between the fundamental frequencies determines the temporal fine structure as well as the details in the pattern of the spectral content. Closely spaced partials will lead to beats with a beat frequency equal to the frequency difference of the two neighboring components, while exactly matching partials will result in a single superimposed component with an overall amplitude depending on the phases of the original components before superposition. Hence, the first objective is an exploration of the structure of the perceptual space which results from a variation of the frequency ratio between the fundamental frequencies of the complex tones by means of a semantic differential. Especially, the perceptual dimension related to the appraisal of the sounds is of interest. Linked to this is the identification of a signal parameter which translates the frequency ratios between the fundamentals into a continuous descriptor variable that can explain a considerable amount of the variance in the judgments.

The second objective is a detailed determination of preference relevant sound characteristics for a wider variation of generic signal parameters and to understand their quantitative impact on overall loudness judgments and preference evaluations in terms of equivalent dB-values. Therefore, the loudness and the preference are determined as points of subjective equality (PSEs) compared to the same reference sound. The PSEs are quantified as the levels at which the test sounds are judged to have the same loudness or preference as the reference sound, which is fixed in level. The frequency ratio between the fundamentals of the complex tones is one independent parameter also in this part of the study. Beyond that, it is of interest to know how the different elements of the multi-tone signals, the two complex tones and the combination tones, contribute to the combined perception and how the mixing ratio between these elements influences the assessments. This knowledge helps to clarify whether reductions of particular sound elements are beneficial in terms of loudness or preference. Moreover, also the impact of different shapes of the spectral envelope and the effect of level reductions of particular frequency ranges on the loudness judgments and the preference evaluations are investigated.

The third objective is the development of a prediction model for the preference evaluations based on psychoacoustic descriptors. The model is intended to generalize the experimentally determined relationships between the preference evaluations and the multiple, generically varied signal parameters towards a description by (psycho-) acoustic variables, which can be directly derived from the time signals of the sounds. The mathematical formulation of the model is planned to depict an equal preference plane which links the sound characteristics, quantified by the descriptor values, to the levels at the PSEs for preference.

Hereinafter, the background of this PhD-thesis is presented. Section 1.2 introduces the particular multi-tone sounds that are investigated in the entire work and presents examples of their occurrence in machinery noise as well as in musical sounds. Section 1.3 presents the concepts of perceptual measurements and appraisal as well as the quantitative and qualitative aspects in both domains. These concepts build the basis for the description and definition of fundamental auditory sensations, associated psychoacoustic descriptors and also attributes which are used to express the appraisal of sounds in Sec. 1.4. An overview of psychoacoustic listening test methods, especially those used for the subjective evaluation in this PhD-thesis, is given in Sec. 1.5. Section 1.6 provides a literature review of the relationships between quantitative and qualitative acoustic variables and the appraisal of sounds, leading to a measurement approach in Sec. 1.7, which was used extensively in a study described in Chap. 3. Section 1.8 concludes this introduction and gives an outline over the following chapters of this thesis.

1.2 Multi-tone signals with biperiodic spectra

In this work, we will concentrate on multi-tone signals consisting of two complex tones and additional combination tones. Such spectra are of interest in a number of practical situations, especially in machinery noise where non-linear interactions between different sound generating processes might lead to the appearance of combination tones.

One example is the noise from aircraft engines with Counter-Rotating Open Rotors (CROR)[(67)]. Due to the high aerodynamic efficiency and potential reduction of fuel burn, the counter-rotating rotor configuration was investigated for aircraft engines in the 1980s, after the oil crisis and rising fuel costs. This engine configuration recently regained interest for aircraft usage, because it still offers a considerable potential in the reduction of specific fuel consumption, reducing the environmental impact of air traffic, and it is a substantial part of the European projects *CleanSky* and its successor *CleanSky II* [1].

Due to the high efficiency, counter-rotating configurations are also investigated for axial fans with a considerably smaller rotor diameter, which are used to cool electronic devices[(172,247,248)]. The main differences to the counter-rotating open-rotor engines are considerably higher fundamental frequencies (blade passing frequencies) for the smaller fans and a reduced amount of interaction tones, due to a smaller non-linear interaction.

In such technical systems, the two rotors can, in principle, run at a different speed (N_1 and N_2) and have different blade numbers (B_1 and B_2), resulting in two fundamental frequencies f_{10} ($= B_1 \cdot N_1$) and f_{01} ($= B_2 \cdot N_2$). In general, the non-linear interaction of the two primary input signals leads to combination tones which can be categorized into summation tones with frequencies

$$m \cdot f_{10} + n \cdot f_{01} \tag{1.1}$$

and difference tones with frequencies

$$m \cdot f_{10} - n \cdot f_{01} \tag{1.2}$$

with m and n being positive integers[(97)]. However, in counter rotating open rotors, only the summation tones with frequency components

[1]http://www.cleansky.eu/

described in equation 1.1 dominate the emitted sound spectrum, due to a reduced radiation efficiency of the difference tones [67,96]. Such a spectrum, consisting of summation tones only, is also called a biperiodic spectrum, because it is decomposable into multiples of two basis frequencies (here f_{10} and f_{01}) [127].

Further details on the technical parameters influencing the sound generation and also on the annoyance from open rotor noise are presented in section 1.2.1. Besides the occurrence of such multi-tone sounds with biperiodic spectra in the machinery noise of counter-rotating open rotors, they can also be found in music instruments as so called "multiphonics" [18] which are described further in section 1.2.2, below.

1.2.1 Examples from machinery noise

The tonal components of two counter-rotating open rotors results from a complex aerodynamic interaction of the air-flow and the two rotors. In this interaction, the upstream rotor (forward) produces an instationary airflow that interacts with the second (aft) rotor downstream. Four principal noise generation mechanism have been identified [142]:

1. Steady loading and thickness of each rotor
2. Unsteady loading noise generated by the aft rotor due to an interaction with the wake shed by the forward rotor
3. Unsteady loading noise generated by the aft rotor due to an interaction with the tip vortex of the forward rotor
4. Unsteady loading noise (generated by either rotor) due to the presence of a rotating potential field of the other rotor.

Steady loading and thickness noise leads to tonal components at the blade passing frequency (BPF), which is determined by the product of the number of blades B and the rotation frequency N, and higher harmonics [90]. Even though the front rotor tip vortex can influence blade passing tone produced by the aft rotor, the front rotor wake is primarily responsible for the generation of interaction tones [255]. The unsteady loading of the fan blades by a harmonic distortion leads to all combination frequencies

$$f_{m,k} = m \cdot B_1 \cdot N_1 + k \cdot B_2 \cdot N_2 = m \cdot BPF_1 + k \cdot BPF_2 \qquad (1.3)$$

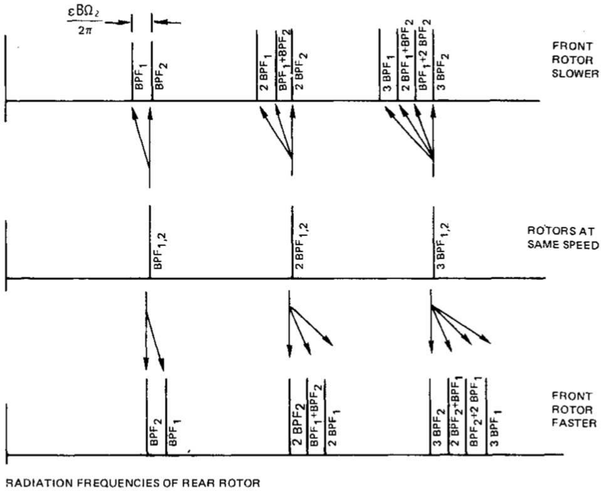


Figure 1.1: *Schematic spectrum of a counter-rotating open rotor and the influence of nonsynchronous rotors on the occuring frequency components (from Hanson and McColgan, 1985*(95)*).*

with the number of blades on the rotors B_1, B_2, the rotational frequencies of the rotors N_1, N_2 ($N_1 = \Omega_1/2\pi$) and where m (the order of sound harmonic) and k (the order of load harmonic) take on all integer values from $-\infty$ to $+\infty$(95,96). Whenever m and k are of different sign, the radiation efficiency of the corresponding modes is low(66,67,95,179). Hence, only the summation tones are relevant for the generated far field noise and only frequencies as defined in formula 1.1 play a considerable role for the perception.

For the general case of non-synchronous rotors, a splitting of the BPF harmonics as shown in Fig. 1.1 (from Hanson and McColgan(95)) occurs. In between the highest and the lowest peak from the steady sources, peaks of the unsteady aerodynamic interaction emerge. Figure 1.2 shows a schematic noise spectrum of a counter-rotating propeller with different blade passing frequencies (BPFs) for the two rotors(56), Fig. 1.3 shows a spectrum from wind tunnel measurements with a scale model, with essentially the same frequency structure(58). In both cases the summation tones are clearly visible between the higher harmonics of the two

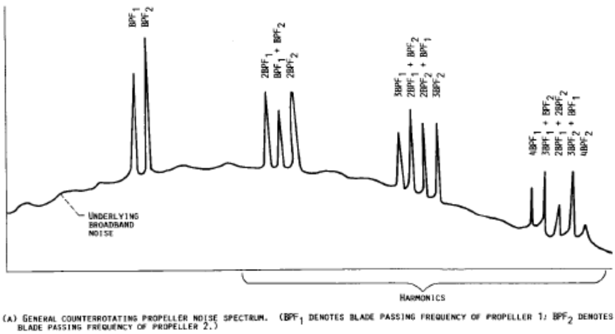


Figure 1.2: *Schematic spectrum of a counter-rotating open rotor (from Dittmar, 1986[(56)]).*

fundamental blade passing frequencies.

1.2.1.1 Examples of acoustically relevant technical parameters and their influence on the spectral composition

The influence of different technical parameters on the amount of interaction tones has been investigated quite extensively for the open-rotor engines during the 1980s, even though computational tools for the simulation and optimization of the rotors were barely available. The main interest was the identification of the origin and a determination of the amount of interaction tones for different rotor configurations. The rotor to rotor distance was found to influence the strength of interaction tones. Depending on the axial mach number M the levels of the first three interaction tones decrease with increased rotor to rotor spacing ($M = 0.8$), while at $M = 0.76$ and $M = 0.72$ they increased or remained the same for some conditions[(57)]. A reduction of the aft-propeller diameter of 14% yields a reduction of 7.5 dB in the blade passing tone and 15 dB in the harmonics of the aft-rotor alone tones. Additionally, also the interaction tones were reduced - mainly the first interaction tone[(58)]. A combination of the front to rear propeller spacing and a reduction of the aft rotor diameter yields reduced interaction tones as long as long as the two rotors are aerodynamically coupled. A further increase of the rotor

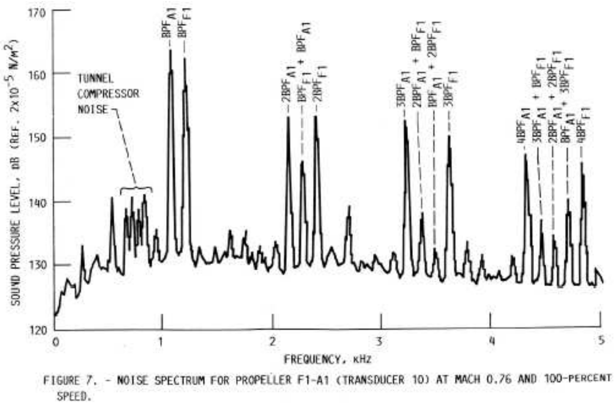


Figure 1.3: *Measured spectrum of a counter-rotating open rotor model (from Dittmar and Stang, 1987[58]).*

to rotor distance does not lead to an additional decrease in interaction tones[59].

Newer studies showed that the wake of the front rotor is more important than the tip vortex for the interaction tone noise. Thus, the whole aft propeller blade span is involved in the noise generation and needs to be considered for further noise reduction measures[255]. Methods like front rotor trailing edge blowing[215] or trailing edge serrations[251] have been found to reduce the overall noise by up to 3 dB. To reduce the sound power of the tonal noise inside the cabin, also active noise control measures were investigated, which offer a reduction of the first CROR frequencies by more than 10 dB[91,92].

The installation of the engine and the disturbance of the upstream flow in a pusher configuration increases the noise in the propeller plane by up to 5 dB and the upstream noise by up to 15 dB[37], which is mainly due to increased levels of the first front rotor harmonics[193]. This effect is depending on the pylon to rotor gap and has a pronounced azimuth directivity which is coupled with the sense of rotation of the front rotor. However, the effect can be largely compensated by trailing edge blowing, which makes the pylon acoustically transparent[193].

In addition, the radiation pattern of the steady loading noise of a single rotor and that from the interaction tones are different. The noise of a single rotor is radiated mainly in the rotating plane, whereas the interaction tones tend to radiate more in up- and downstream directions(255). In the far-field noise propagation, the interaction tones were found to be more effectively propagated than the individual rotor tones and thus are the main contributors to fly-over noise.

1.2.1.2 Annoyance from open rotor noise

In light of aircraft flyover noise, the annoyance of the sounds was investigated in several studies of McCurdy. In all studies, the annoyance of synthesized flyover sounds from different rotor configurations and recorded sounds from conventional jet aircraft were judged by participants in laboratory experiments. The stimuli included sounds resulting from equal numbers of blades on each rotor (148), different numbers of blades on each rotor(149), noise from single-rotating propeller configurations (fundamental frequencies: 67.5 Hz ... 292.5 Hz)(150) and noise from a counter-rotating propeller configuration (fundamental frequencies 112.5 Hz ... 292.5 Hz)(151). A prediction of the judged annoyance based on the A-weighted sound pressure level could be improved by a duration and a tone correction. Furthermore, McCurdy *et al.* identified a complex influence of the interaction of fundamental frequency and tone-to-broadband noise ratio, and the interaction of tone-to-broadband noise ratio and level on the annoyance judgments. No influence was found for the frequency envelope and the difference in blade numbers between front and aft rotor in their studies.

Besides the effect of fly-overs on the community noise, the pronounced directivity of the propeller harmonics and the presence of many discrete high-frequency noise components were identified as possible problems for a passenger airplane with respect to the acoustic comfort inside the airplane(135). However, detailed insight on the contribution of the various signal parameters to the unpleasantness of this particular multi-tone noise signature is not available in the literature so far.

1.2.2 Examples from music instruments

Some fingerings of woodwind instruments lead to tones with a special quality - so called multiphonics. They are perceived as having two or more pitches sounding simultaneously or as a tone with a rough and beating quality(18). Apart from two independent main frequency components A and B, Backus identified additional heterodyne components $|A-B|$, $|A+B|$, $|2A-B|$, $|2A+B|$, generated by a non linear reed system and a paper cone, a bassoon, a clarinet, an oboe and a flute(15). Fletcher attributed the occurrence of these frequencies to the mode coupling between the two oscillators (a generator and a resonator) in musical instruments(75). A representation of clarinet sounds in phase space allowed a characterization of the biperiodic multiphonics as an intermediate regime, which occurs in the transition from periodic oscillations to a chaotic regime, resulting from a rise in excitation pressure(82).

A reanalysis of the Backus' data by Schumacher(203) revealed that all frequencies $f(A,B)$ found by Backus could be represented in a form

$$f(A,B) = m \cdot C + n \cdot D \tag{1.4}$$

with positive integers n, m and hidden low frequency subharmonics C, D. Apparently, this structure is equivalent to the formulation of the combination tones in formula 1.1. Locking modes from two inharmonically related oscillators, which are characterized by ratios between the two frequencies C and D equal to ratios of small integer numbers, can be distinguished from non-mode locking where the two frequencies C and D are incommensurate(127). Figure1.4 shows an exemplary multiphonic spectrum from Schumacher (1981)(203). Here, the hidden subharmonics are equal ($C = D$), which could be regarded as a simplified case of the general formulation in equation 1.4.

1.2.2.1 Perception of multiphonic sounds

Riera et al. investigated the perceptual effects and implications of multiphonics on the timbre of alto saxophone sounds(194). Based on dissimilarity ratings of 15 representative recorded saxophone sounds and a multi-dimensional scaling technique the perceptual space was explored. The proposed four classes of sound could be segregated into the obtained two dimensional perceptual space. The spectral centroid and the modulation frequency were found as suitable physical correlates of the two perceptual dimensions, respectively.

Recently, multiphonics are under investigation with respect to their independent structural potential in a musical piece(160) and as a funda-

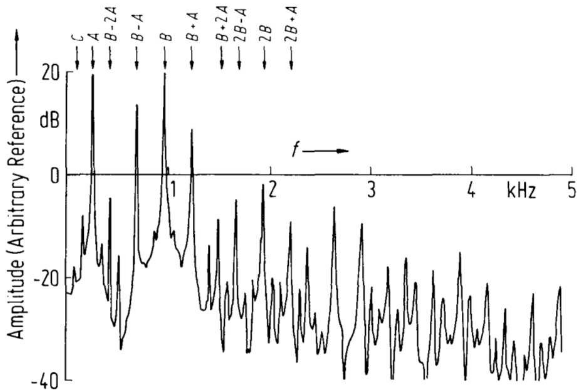


Figure 1.4: *Multiphonic spectrum with the heterodyne frequencies based on the fundamentals A and B and one underlying subharmonic fundamental C (from Schumacher, 1981*[203] *).*

mental element in Solo improvisations[207].

1.3 Quantitative and qualitative aspects in judgments and in evaluations of sounds

The judgment of a sound by a listener is often described in perceptual models as a two-stage process[17,181,242]. The main difference between the two stages lies in the amount of cognitive factors involved in the judgment processes on each level. A general model view, is schematically shown in Fig. 1.5.

In this model, the physical sound wave, which originates from a sound source, is transformed by the ear and the auditory system into an neural auditory event. The ear and the auditory system act as filter ("Filter 1") on the incoming acoustic information. This auditory event is further transformed by higher cognitive processing ("Filter 2") including context, mood, emotion, memories, attention, etc., into a cognitive percept (so called bottom-up processing), which is evaluated . Equally, cognitive

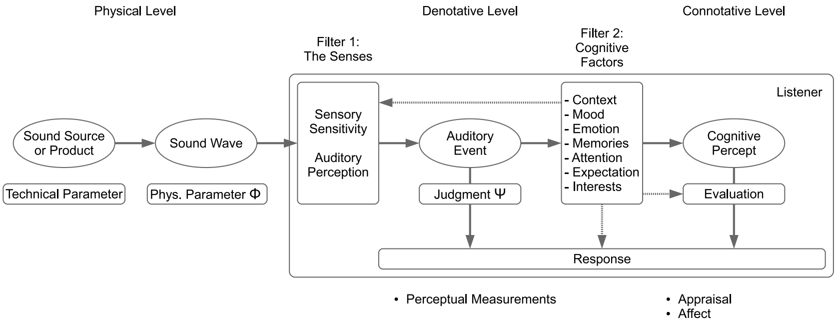


Figure 1.5: *General schematic illustrating the process of judgment on the denotative level and the evaluation of a sound on the connotative level.*[2]

factors can also have an effect on the auditory system ("Filter 1") and on the way a response is given by a listener (so called top-down processes).

Judgments on the first stage are thought of as being rather direct and neutral descriptions of sounds on a denotative level. The evaluation of sounds on the second stage explicitly includes cognitive factors on the connotative level and goes beyond the pure description. In this view, the scaling of auditory events in perceptual measurements is located on the denotative level. The evaluation, the appraisal of sounds and the concept of affect take place on the connotative level.

1.3.0.2 Quantitative and qualitative aspects on the denotative level

The perceptual measurement is residing on the denotative level. The overall aim of perceptual measurements is a characterization of the auditory system ("Filter 1") by determining the relationships between a physical variable Φ and a corresponding sensory continuum Ψ, also called basic auditory sensation. This stimulus-response relationship between the two is described by the psychophysical law or function f:

$$\Psi = f(\Phi) \tag{1.5}$$

[2]A similar view is also shared by Bech and Zacharov (17), Blauert (34), Pedersen and Fog (181), and Västfjäll (242).

Stimulus-response theory focuses on the uni-dimensional determination of the relationship between a stimulus and the response of a listener in an operationalized process. The details of the transformation from the physical sound wave to the auditory event, are not of interest because they are not easily accessible to the investigator. It is the measurement itself and, more precisely, the question to the participant which defines what is measured. The question, whether it is possible at all to give an answer without any involvement of cognitive processes, is not posed (or discussed).

In such measurements, effort is taken to minimize potential influences of cognitive factors, which might interfere with the intended research problem, by controlled laboratory conditions. In addition, for the measurement of sensory thresholds, experimental methods were developed which allow a distinction between the observer's sensitivity (performance) and bias due to individual response behavior. Examples are forced choice methods and the application of signal detection theory (81).

In listening experiments, the response of the observer might be given in the form of in-/equality judgments or by quantitatively expressed responses (e.g., numeric values for intervals, ratios). In the first case, an expression of the judgment e.g. in terms of a numeric response is not required and the quantification of the equality judgment is realized by a physical measure of the investigated stimulus property. In the latter case, the human listener is asked to give a numerical response and the transformation of the judgment into the numeric response is included in the measurement process. Basic auditory sensations measured in this way are, e.g., loudness, sharpness, pitch, roughness and fluctuation strength further described in section 1.4.1.

According to Stevens, sensory continua can be divided into two different classes: one related to quantity and one linked to quality (214). Stevens called the quantity or intensity related continua *prothetic* and the quality related ones *metathetic*. The *prothetic* continua are understood to have an order in terms of degrees of magnitude or sensed intensity, while the *metathetic* continua describe a change in position or quality. In Stevens view of *prothetic* continua, a higher sensed intensity is also thought to result in a higher physiological excitation (higher nerve firing rate), whereas the physiological excitation can remain constant in magnitude, but different neurons are excited for changes on *metathetic* continua. Loudness is seen as a *prothetic* continuum on which sound can be less intense (softer) or more intense (louder). Changes in the sound character take place on the *metathetic* continua - the intensity of a sound can remain the same while the character of a sound in terms of sharpness, roughness or timbre changes.

In terms of measurements it turned out that the two classes of continua behave differently. For the *prothetic* continua, Stevens found that equal stimulus ratios produce equal subjective ratios, which would mean that the just noticeable differences (JNDs) are proportional to the relative change of the physical magnitude[(214)]:

$$\frac{\delta\Psi}{\Psi} = C \cdot (\frac{\delta\Phi}{\Phi}) \tag{1.6}$$

Integrating the formula yields a linear relationship between the logarithms of the physical and the perceptual scale:

$$ln\,\Psi = C \cdot ln\,\Phi + k \tag{1.7}$$

which is identical to a power-law between stimulus and sensation magnitude:

$$\Psi = K \cdot \Phi^C \tag{1.8}$$

Stevens successfully identified power-law relationships for a variety of sensory continua and determined different exponents C for different sensory modalities[(214)]. For example, for the relationship between the sound pressure and the loudness sensation, an exponent $C = 0.6$ was determined, which means that a change in sound pressure level of 10 dB corresponds to a doubling in perceived loudness (for a 1 kHz tone and levels above 40 dB SPL)[(263)].

For *metathetic* continua, on the other hand, logarithmic relationships between stimulus Φ and sensation magnitude Ψ were found:

$$\Psi = c \cdot log(\Phi/\Phi_0) \tag{1.9}$$

This logarithmic relationship is also known as the Weber-Fechner Law, which assumes the JNDs to be a constant fraction of the physical stimulus intensity:

$$\delta\Psi = C \cdot (\frac{\delta\Phi}{\Phi}) \tag{1.10}$$

1.3.0.3 Sound character

In the distinction between prothetic and metathetic continua, the term quality was used to describe qualitative aspects of sound in contrast to quantitative aspects in perception. More recently, qualitative differences between sounds (on a denotative level) are described as differences in *sound character*, a concept that is not to be confused with the concept of *sound quality* (described in more detail in Sec. 1.4.2)(35,256).

Sound character is understood as the comprehensive description of a sound by (expert) listeners, which is preferably specifiable by a profile of acoustic and/or psychoacoustic parameter values(35,206). Following this definition, sound character is seen as a multidimensional description of a sound itself, which relates primarily to the neutrally, not hedonistically judged, descriptive, sensory properties of a signal(256).

In this understanding, sound character can be described as a multidimensional profile of discriminable aspects of a sound (also called basic auditory sensations, examples relevant in this work are given in Sec. 1.4.1). An alternative description for the dimensions of the multi-dimensional profile are psychoacoustic parameters or descriptors which are derived algorithmically from the sound signal(35) (exemplary descriptors are given in Sec. 1.4.3).

For the perceptual characterization of musical sounds, the attribute *timbre* is more commonly used(132). Timbre is defined as the "attribute of auditory sensation which enables a listener to judge that two non-identical sounds, similarly presented and having the same loudness and pitch, are dissimilar"(5). In this sense, timbre is subtractively defined as everything allowing a discrimination between two sounds, when loudness and pitch of the sounds are equal. This definition shows the metathetic nature of timbre and, thus, also sound character. For a description of sound character, multidimensional measurement methods like multidimensional scaling (MDS)(218) or semantic differentials (explained in more detail in Sec. 1.5 and applied in Chap. 2) are used.

The multidimensional measurement of the sound character showed that different attributes might interfere with each other in complicated ways. Stevens identified the covariation of four perceptual attributes - loudness, volume, density and pitch - for pure tones varied in stimulus intensity and frequency[212]. Depending on the exact task to the participants, they could deliver distinct answers with respect to the different attributes with a high degree a subjective certainty. All four attributes could be represented as a function of the two physical stimulus parameters intensity (in dB) and the frequency (in Hz). Melara showed that the three perceptual dimensions loudness, pitch and timbre are processed jointly and each dimension suffers interference from variations along irrelevant dimensions[153]. Depending on the relevant and irrelevant dimension, different types of complex interaction emerged. Thus, a clear distinction between prothetic and metathetic continua is by no means unambiguously possible.

Sound descriptions may also rely on the experience and memory of a listener or the identification of a sound by a listener, which obviously involves cognitive factors. Therefore it is not always clear where to draw a line between the denotative and the connotative domain. To reduce the influence of cognitive factors on measurements of the sound character, signal processing techniques neutralizing the meaning and inhibiting the identifiability of sounds were sometimes used[70]. It was found that loudness judgments were barely affected by neutralizing the sounds[63], whereas the identifiability played a role in terms of annoyance judgments[62].

1.3.0.4 Quantitative and qualitative aspects on the connotative level

Evaluative judgments like the pleasantness or unpleasantness, annoyance and the appraisal of sounds go beyond the neutral, denotative description of a sound and are, thus, attributed to the connotative level. They clearly involve cognitive factors where the evaluation is based on experience, expectation and further context factors. A pure psychoacoustical approach with an operationalized measurement, which aims at a functional description of the relationships between a given stimulus (input) and a response (output), while minimizing and controlling for cognitive factors, does not provide the desired insight on this level. In contrast to the psychoacoustical stimulus-response theory, cognitive factors need to be included in the measurement process. In addition, also the way, how the judgment process takes place and how cognition, action, emotion and context influence a judgment are often included in the research question.

The major facets of the connotative aspect of appraisal of sounds are presented in Sec. 1.4.2 in further detail.

The concept of affect, dealing with the characterization of emotional states, and the measurement of affect dimensions also involves cognitive processes. Similar to the *prothetic* and *metathetic* types of sensory continua on the denotative level, Russel identified two dimensions of core affect in his circumplex model of affect (197). The two dimensions were originally denoted *pleasure* and *arousal*, but the terms *valence* and *activation* are also used synonymously for them. Västfjäll *et al.* largely confirmed the circumplex model by Russel and developed a set of twelve adjective scales which can be used for a selective measurement of particular affect dimensions (240). They further applied the circumplex model of affect to judgments of sounds. Over several studies they determined relationships between perceptual aspects, affective appraisal and overall preference for aircraft cabin noise. These studies are further described in section 1.6.2.4 after the introduction of the basic auditory sensations (241,243–246).

1.4 Basic auditory sensations, appraisal of sounds and psychoacoustic descriptors

In the following, a selection of basic auditory sensation, commonly used attributes of the appraisal of sounds and (psycho-)acoustic descriptors are introduced.

Basic auditory sensations can be defined as the attributes of discriminable and independent aspects between two auditory events while ignoring all other aspects (49). In this sense, loudness, pitch and certain aspects of timbre like for example sharpness, roughness and fluctuations strength are generally understood as basic auditory sensations that can be scaled independently from each other. For the multi-tone sounds investigated in this PhD-thesis, which are based on two fundamental frequencies, also the concepts of consonance and harmonicity naturally come into play. In this context also a close link between slow modulations and the appraisal of sounds becomes apparent.

The appraisal of a sound or the description of the mood of the listener hearing a sound can be expressed by different attributes. Especially the term *annoyance* and the psychological concept of *noise annoyance* are outlined in more detail because of the extensive usage of these terms in the literature. Furthermore, the meaning and the usage of the terms *loudness* (here from the connotative viewpoint), *noisiness*, *pleasantness*, *preference* and *sound quality* are explained.

To predict the human response of basic auditory sensations from listening tests, different psychoacoustic descriptor algorithms (also called psychoacoustic metrics or parameters) were developed. Some of the descriptors are standardized in national and international standards (e.g. loudness, sharpness), others (e.g. roughness, tonality, fluctuation strength) are not standardized yet. Those descriptors which are later used in the correlation analysis and modeling of the preference evaluations (Chap. 4) are presented here. In addition to the descriptors of the sound character, metrics referring to speech intelligibility are also introduced, because they are commonly used in the context of aircraft cabin noise.

1.4.1 Basic auditory sensations

Due to the specific structure of the investigated multi-tone sounds, consisting of a high number of tonal components, many basic auditory sensations are anticipated to be involved in the perception and they are potentially relevant for the appraisal of the sounds. These include sensations related to the physical intensity as loudness, to the spectral distribution of energy as sharpness, to the temporal structure of the signals as roughness and fluctuation strength, and also to musical aspects like pitch, consonance and harmonicity. In the following the different basic auditory sensations and their general relationships with physical signal parameters are presented in more detail.

1.4.1.1 Loudness

Loudness is the "attribute of auditory sensation of which sounds may be ordered on a scale extending from *soft* to *loud*"(5). It is most directly related to the perceived intensity of a sound and it is linked to the physical intensity of a sound in a sophisticated manner. The sound pressure level is a major physical factor influencing the loudness of a sound(263). Furthermore, loudness is depending on various stimulus properties, like the spectral bandwidth(115,200,263), spectral content(138,199), the duration(77,232), and the temporal signature(86,159). Especially for sounds which strongly differ in terms of spectral content, it may occur that an increase of the A-weighted sound pressure level can even produce a decrease in loudness (e.g. for increasing values of a tone-to noise ratio)(104).

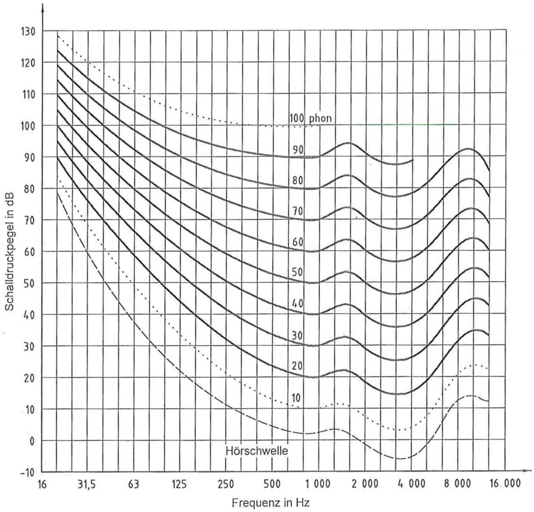


Figure 1.6: *Equal loudness level contours for pure tones (from ISO 226:2003(118)). The absolute hearing threshold measured under free-field conditions is plotted as dashed line. The 10 and the 100 phone isophone are plotted as dotted lines because the lack of data for these curves.*

Different approaches to measure loudness are summarized by Marks and Florentine(143).

One way to characterize the loudness of a sound is the loudness level which is measured with the unit phon. A loudness level of a sound of N phon is defined as the sound pressure level of a 1 kHz reference tone (of N dB SPL) adjusted to equal loudness in comparison to the sound of interest(72). This definition leads to the frequency and level dependent equal loudness level contours (isophone curves) shown in Fig. 1.6. The contours show that the human ear is most sensitive in the frequency region around 3 to 4 kHz, which is owed to the $\lambda/4$-resonance of the ear canal. Tones at very low and very high frequencies can be considerably higher in level than these mid frequency tones, while being equally loud. Mean equal loudness level contours based on the data from several labo-

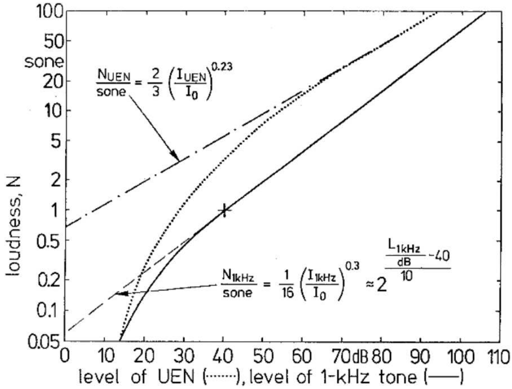


Figure 1.7: *Loudness function of uniform exciting noise (UEN, dotted line) and a 1 kHz pure tone (solid line) as a function of sound pressure level and fitted power laws (from Zwicker and Fastl, 2007[263]).*

ratories are the basis for the ISO 226 standard[118,222], which is currently under discussion due to discrepancies in the low frequency region.

Another way to characterize loudness is to determine the changes in sound pressure level (dB SPL) necessary to double or half the perceived loudness by ratio scaling. In this way, the loudness as a function of the SPL is measured on the so called sone-scale[143]. Figure 1.7 shows the loudness function of a 1 kHz pure tone and uniform exciting noise (UEN) as a function of sound pressure level. The sone-scale is based on a 1 kHz tone of 40 dB SPL (which equals 40-phon) as a reference, which is defined as having a loudness of 1 sone. On this scale, loudness ratios can be directly expressed and a doubling of the loudness corresponds to a doubling in sone values. For a 1 kHz tone above 40 dB SPL a loudness doubling is equivalent to an increase of about 10 dB SPL or 10 phon[263].

1.4.1.2 Pitch

The pitch of a sound is defined according to the ANSI acoustical terminology standard as "the attribute of auditory sensation in terms of which sounds may be ordered on a scale from *low* to *high*"[5]. This definition refers to the perhaps most dominant quality of pitch, which is pitch height. For single tones, pitch height is proportionally related to the physical frequency, but it is not exactly the same.

Additional qualities of pitch are pitch strength and chroma[93]. For narrow-band noise the pitch strength decreases with increasing bandwidth of the stimulus and similarly it decreases with an increase of higher harmonics for complex tones. If the fundamental or lower harmonics are removed from a complex tone, then the pitch height remains rather stable, even though there is no energy at the frequency of the perceived residual pitch[263]. Similarly, for inharmonic tones the removal of lower components leads to a virtual pitch which relates to the periodicity of the signal[253]. Several further aspects of pitch have been described by Terhardt[227].

1.4.1.3 Sharpness

Von Bismarck used a semantic differential to investigate the multidimensional aspect of timbre for harmonic complex tones with different spectral shapes, while being fixed in pitch and having equal loudness[237]. Based on a factorial analysis of semantic differential data, the attribute *sharp-dull (sharpness)* explained most of the variance. He showed that the first moment of the specifically weighted specific loudness distribution could serve as a measure for sharpness. In a follow-up study, von Bismarck measured the sharpness for a variety of spectrally shaped noises and complex tones[238]. He found that the spectral fine structure only had a small effect on the judged sharpness, more important is the overall spectral envelope. The sharpness is measured in the unit acum in comparison to a reference sound, which is a critical band wide narrow-band noise (Δf = 160 Hz), centered at 1 kHz, with a level of 60 dB SPL defined as 1 acum[263].

1.4.1.4 Roughness and fluctuation strength

Fluctuation strength and roughness are both sensations which are related to the temporal structure of a sound and depend mainly on the modulation frequency and the degree of modulation (for amplitude modulated sounds) or the frequency modulation index in the case of frequency modulated sounds. Both sensations rise with an increase in the degree of modulation or the frequency modulation index, respectively. Both sensations are linked to specific regions in the modulation frequency spectrum.

Fluctuation strength is a sensation that is evoked by amplitude or frequency fluctuations which are slow enough for a listener to follow them. Fluctuation strength is maximal around a modulation frequency of 4 Hz, which equals roughly the syllable rate in fluent speech of 4 to 5 syllables per second[137]. The fluctuation strength elicited from a 60 dB SPL 1 kHz tone, 100% amplitude modulated with a modulation frequency of 4 Hz is defined as a fluctuation strength of 1 vacil[263].

At higher modulation frequencies of about 20 Hz to 30 Hz the sensation of fluctuation transitions smoothly into a roughness impression[224]. Maybe von Helmholtz was one of the first to characterize faster beats between neighboring tones as rough (German: rau) and creaky (German: knarrend) sounds[239]. For 100% amplitude modulated tones, roughness shows a bandpass characteristic as a function of modulation frequency. The maximum roughness occurs at a modulation frequency of 30 Hz for a carrier of 125 Hz and then rises to 70 Hz for carriers of 1 kHz and higher. The roughness elicited from a 1 kHz tone, 100% amplitude modulated with a modulation frequency of 70 Hz is defined as a roughness of 1 acum[263]. The upper limiting modulation frequency for roughness is about 100 Hz for carrier frequencies up to 500 Hz and then rises up to 200 Hz for carriers above 2 kHz - above this modulation frequency roughness vanishes[224]. The overall roughness of a sound is composed by partial roughnesses of the different critical bands. Antiphasic fluctuations of partial roughnesses decrease the overall roughness but do not cancel each other completely. In-phase fluctuations increase the roughness, especially when occurring in one frequency band[225].

1.4.1.5 Consonance, harmonicity and frequency ratios

The classical studies on consonance theory by von Helmholtz[239], and Kameoka and Kuriyagawa[124] state that musical intervals with ratios between the fundamental frequencies equal to ratios of small numbers (e.g. 4:3) are consonant and elicit less roughness than deviations from exact ratios of small integers. Another study of Kameoka and Kuriyagawa[125] already indicates that the consonance of complex tones is also depending on the number and the relation of even and odd partials and cannot only be derived from classically known intervals equal to ratios of small integers. Nevertheless, Terhardt concluded that the concept of psychoacoustic consonance, defined by the absence of roughness is only fulfilled by the just scale based on mathematically exact ratio of small integers [226].

However, some studies showed quite the opposite. The study of Miskiewicz et al. finds that a superposition of two complex tones with perfect musical intervals (mathematically exact ratios of small integers between the fundamental frequencies) are judged higher in terms of roughness than equally tempered musical intervals which slightly deviate from the mathematically exact ratios of small integer numbers [157]. They explained the reduction in perceived roughness by the occurrence of slow fluctuations in the sounds with beat rates below 20 Hz. The lowest roughness values were found for those sounds with the smallest deviations from the exact integer ratio yielding the lowest beat rates. A very similar relationship between frequency ratios and the judged pleasantness, which can be meaningfully related to the repetition rate of the time signals, is presented in Chap. 2 of this thesis.

Fletcher et al. investigated the quality of piano tones [73] and organ tones [74]. For synthesized piano tones below middle c ($f_0 = 261.6$ Hz), "the lack of being harmonic gives rise to the peculiar quality known as piano quality, namely, the live-ness or warmth" [73]. Similarly, beatings between neighboring partials contributed to the warmth of organ tones [74]. In both cases, a mistuning of some partials in the re-synthesis led to a warmer sound which made it more difficult for listeners to distinguish the synthesized sounds from original recordings. The correct reconstruction of the "warmth" of the piano tones was preferred over frequencies made harmonic and added warmth for the organ tones also was preferred over a mathematically exact harmonic structure. Van de Geer and coauthors studied the perceptual space of 23 different musical intervals with a semantic differential of ten adjective pairs [80]. Three perceptual dimensions were identified: pitch, evaluation and fusion. In their study the adjective pair *consonant - dissonant* also showed to be purely evaluative. In a study with natural sounds also the sounds with a non-harmonic structure were judged to be more pleasant than those with a purely harmonic structure [31]. Similarly, for synthesized guitar sounds, short melodies based on inharmonic sounds were slightly preferred over strictly harmonic ones [71,235].

From all of the mentioned literature in this subsection, it becomes clear that consonance and harmonicity are closely linked to other auditory sensations like roughness and also the appraisal of tonal sounds.

Beats and slow modulations which are based on small deviations from mathematically exact ratios of small integers between partials seem to considerably contribute to the pleasantness of sounds with multiple tonal components.

1.4.2 Appraisal of sounds

Various attributes are used to express the evaluation of sounds and the moods evoked by sounds. A lot of methodologically relevant literature for this PhD-Thesis has been carried out in the research fields of community and environmental noise, where attributes like *annoyance*, *noise annoyance* and *noisiness* are frequently used. In sound quality studies, attributes expressing the *pleasantness* or *preference* are more commonly found. In the following some exemplary definitions of the particular attributes are presented and also the concepts of *sound quality* and *product sound quality* are explained.

1.4.2.1 Annoyance and noise annoyance

Annoyance and noise annoyance are commonly used attributes from everyday life and they are also used as judiciary terms in noise legislation to describe the nuisance effects of sounds. This might be one reason for the frequent usage of the attribute in the evaluation of sounds also in laboratory settings. However, the psychological concept of annoyance is a rather difficult one.

Based on expert survey data Guski *et al.* summarized that noise annoyance can be regarded as a multifaced psychological concept. It is mainly related to immediate behavioral noise effects (like disturbance and interference with intended activity) and evaluative aspects (like nuisance and unpleasantness)[(87)]. But even though "annoyance judgments often do covary with acoustic variables, noise annoyance is not just reflecting acoustic characteristics."[(87)] It can further be based on the emotions evoked by sounds, attitudes towards sound sources and knowledge about sounds. In this sense, noise annoyance can be compared to the concept of product sound quality, which is even though often highly correlated to acoustic variables, not solely determined by acoustic factors. Marquis-Favre *et al.*[(146)] compiled a review of acoustic and non-acoustic factors linked to noise annoyance. Namba and Kuwano also stress that many factors like the subjective meaning, value of sounds or sound sources, individual and social situations, and individual noise sensitivity can affect the degree of *annoyance* besides the nuisance effect. To measure *annoyance* in a laboratory, these additional factors would need to be controlled for, which makes it difficult to measure "true" annoyance in a laboratory setting [163].

For laboratory measurements, Berglund *et al.* proposed a definition of *annoyance* as "the nuisance aspect of noise in an imaginary situation". They used the phrase, "After a hard day's work you have been comfortably seated in your chair and intend to read your newspaper." and asked the participants "how annoyed [they] would feel when exposed to the given ... noises." [25] In some studies, the term *short term annoyance* is used to indicate that *annoyance* was judged in a laboratory setting and based on immediate judgments of rather short sound samples [1,2,230].

Even if participants are instructed to put their self into an imagined situation, it will only in parts be successfully accomplished by participants and laboratory experiments on noise annoyance remain limited in controlling context factors which might affect the judgment [180].

1.4.2.2 Loudness and noisiness

In the psychoacoustic context, *loudness* is clearly understood as the sensation related to the perceived intensity of a sound [163] (see also Sec. 1.4.1.1). In listening tests with non-expert listeners, it is sometimes also defined as "the perceptual aspect ... that is changed by turning the volume knob on a radio set" [25].

In an exploration of the meaning of different verbal concepts, the adjective *loud* was found to have a rather negative connotation in Germany, UK and in the U.S. and is rather close to *annoying*, while it is seen rather neutral in Japan, Sweden and China [164,169]. Thus, a close link between *loudness* and *annoyance* may be in parts already given by the culturally based negative connotation of *loudness* and the usage of the term *loudness* as an evaluative attribute. To reduce a possible confusion of listeners, Zwicker proposed to ask for "the loudness, not the annoyance" or "the annoyance, not the loudness" to clarify the meaning of the requested attribute in listening tests [261].

The term *noisiness* is meant to describe the unwantedness of a sound based on auditory factors, whereas *annoyance* expresses the unwantedness by non-auditory factors [167]. Berglund *et al.* defined *noisiness* as "the quality of the noise" [25]. They gave the examples that "the sound from a jackhammer may be more or less noisy than that from a motorbike even if they are considered equally loud" and that "music may be loud but still not be perceived as noisy." Similarly, Namba and Kuwano defined *noisiness* "as unpleasant quality of noise" and "subjects can judge which sound is better or worse in sound quality than the other" [163]. Following these descriptions, a German translation of *noisiness* would be "Lärmhaftigkeit" rather than "Rauschhaftigkeit."

1.4.2.3 Pleasantness and preference

The term *pleasantness* (or unpleasantness) can be used to express the appraisal of a sound ("how pleasant the sound is") and as an affective reaction of "how pleasant an observer feels" [241]. Most commonly used is the former definition for the description of the sound quality or the sound source, for example in the assessment of everyday sounds [257], for speech masking noise [113], as an adjective of a semantic differential [134,165], and also for vehicle sounds [3,241,245].

Preference can be can be requested from listeners as a decision between two alternative choices, e.g. in a preference paired comparisons of different sounds [29,234]. Thus, preference evaluations are always relative comparisons against a reference. The reference is in most cases given as an external reference (or alternative) and not referring to an internal reference in the mind of a listener. Sometimes also non-preference is used in listening tests by requesting from participants which of two presented alternatives they would have preferred to switch off [24]. Preference can also be granted towards affective emotional reactions [241]. In this context, Västfjäll *et al.* suggested that it is useful to decompose emotional reactions to sound into the two core affect dimensions [243] and combine valence and activation measures with preference evaluations [244]. Preference (from paired comparison experiments) and pleasantness judgments (judged by magnitude estimation) were also found to be highly correlated to each other[3] [154].

[3]The stimuli used in that study are similar to those used in the current work. The stimuli were complex tones consisting of harmonic and inharmonic partials which were based on a fundamental frequency of 125 Hz and higher partials up to 1.5 kHz

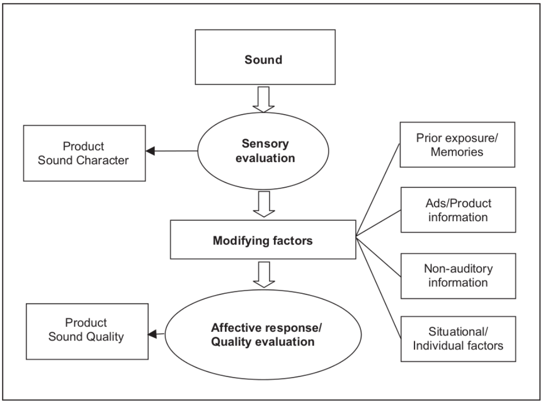


Figure 1.8: *Model representation differentiating between sound character and sound quality (from Västfjäll, 2004[242]).*

1.4.2.4 Sound quality

The term *sound quality* is thought of as referring to superiority, satisfaction and dissatisfaction[139] and involving a comparison of perceived auditory nature and desired auditory nature which is defined by the internal reference of an observer[121]. In this sense sound quality is not a property of the sound itself but it exists in the interplay between the sound coming from a source and a listener perceiving an auditory event and interpreting it, based on the experience, knowledge, emotion and also the attitude of the individual listener as context factors. This means that the frame of reference for a judgment of sound quality is the internal reference of the individual listener. Guski similarly concluded that sound quality includes the suitability, the pleasantness and the identifiability of sounds or sound sources[88].

Västfjäll proposes two stages in the process of *product sound quality* judgments (Fig. 1.8)[242]. On the first stage, sounds are judged on a sensory level, which is also the basis for the description of the product sound character. The second stage is responsible for the affective response and the quality evaluation, from which product sound quality is formed. A filter between the two stages modifies the information passed to the second stage. It encompasses memories, non-auditory information and other individual factors relevant for the formation of a quality

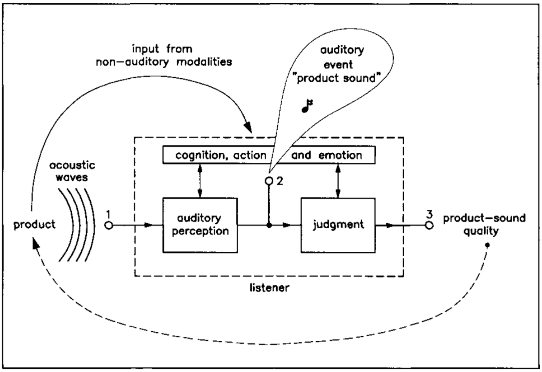


Figure 1.9: *The product sound quality circle according to Blauert and Jekosch, 1997(34).*

evaluation. A similar view is also held by Pedersen and Fog(181) and by Bech and Zacharov(17).

Blauert and Jekosch went even further by including the product as the source of the sound instead of just the acoustic stimulus and also the interaction with the product into their "sound quality circle" (See Fig. 1.9). They defined *product sound quality* as "... a descriptor of the adequacy of the sound attached to a product. It results from judgments upon the totality of auditory characteristics of the said sound - the judgments being performed with reference to the set of those desired features of the product which are apparent to the users in their actual cognitive, actional and emotional situation(34)". They further proposed a structure of product sound quality based on different degrees of abstraction(36).

1.4.3 Acoustic and psychoacoustic descriptors

The various acoustic and psychoacoustic descriptors can be divided into two main classes - metrics aiming at a prediction of human responses to sound (like the above described basic auditory sensations) and purely technical metrics providing single values for the description of sound features, which do not necessarily have a direct link to perception. Metrics of the first class are often based on a physiologically inspired, functional model of human hearing and deliver a single-number value (e.g., loud-

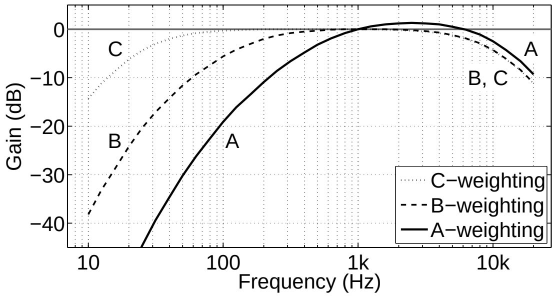


Figure 1.10: *Weighting functions used in noise measurements plotted as gain over frequency (according to the ANSI S1.42-2001[8]).*

ness, sharpness and roughness algorithms). With laboratory listening experiments, the model outputs are calibrated to match the results of the listening tests best. The test sounds for the model calibration are, similar like the sounds defining the reference of an auditory sensation often rather generic sounds like broadband or bandpass filtered noise, single tones or mixtures of them. As a result, such models work quiet well for rather simple synthetic stimuli. For more complex sounds like real environmental or technical sounds the applicability of a metric has to be tested in each case. The more technical metrics of the second class are often based on a FFT-based signal analysis and algorithms which provide a single value (e.g., tonality, spectral centroid) for a sound feature. Such metrics can be highly correlated with human responses in listening tests, but they are usually not calibrated to perceptual scales, e.g. the spectral centroid is often highly correlated with sharpness values, but it is not calibrated to the acum-scale. In any case, the output of psychoacoustic descriptor algorithms should not be confused with the auditory sensation itself which always involves a measurement by a human listener. The descriptors can be regarded as a prediction of a respective auditory sensations with a limited range of generalizability[33].

1.4.3.1 Frequency sensitivity weighting of the sound pressure level

The physical measure for the strength of sound is the sound pressure level L, defined as:

$$L = 20 \cdot log_{10} \left(\frac{\widetilde{p}}{p_0} \right) = 10 \cdot log_{10} \left(\frac{\widetilde{p}^2}{p_0^2} \right) \text{ dB} \tag{1.11}$$

with $\widetilde{p}$ being the root-mean-square (RMS) sound pressure[4] in Pascal and a reference sound pressure of $p_0 = 20 \cdot 10^{-6}$ Pa. As a first approximation to the frequency dependence of human loudness perception, different weighting filters are implemented in sound level measurements[(8)]. The gain factors of the A-, B- and C-weighting are shown in Fig. 1.10. The A-weighting is a filter approximating the shape of the 40 phone isophone contour with a considerable attenuation of low and very high frequencies. The B- and C- Weighting approximate the 60 and 80 phone isophone contours, respectively. The A-weighted sound pressure level is defined as

$$L_A = 20 \cdot log_{10} \left(\frac{\widetilde{p}_A}{p_0} \right) \text{ dB(A)} \tag{1.12}$$

with $\widetilde{p}_A$ being the A-weighted RMS pressure in Pascal. To describe long term measurements of non-stationary sounds with a single number value over a time period T, the energy equivalent A-weighted sound pressure level L_{Aeq} is used:

$$L_{Aeq} = 10 \cdot log_{10} \left(\frac{1}{T} \int_0^T 10^{0.1 \cdot L_A(t)} \right) dt \text{ dB(A)} \tag{1.13}$$

with $L_A(t)$ being the time course of the A-weighted sound pressure level[(161)]. In this sense, the L_{Aeq} is the A-weighted, RMS sound pressure level with the measurement duration T used as averaging time[(30)]. The A-weighting is prescribed in most standards for noise immission measurements, even for noise situations where the loudness level is clearly above 40 phon. In such instances the A-weighting underestimates the low and

[4] RMS sound pressure: $\widetilde{p} = \sqrt{\frac{1}{T} \int_0^T p^2(t) \, dt}$ Pa

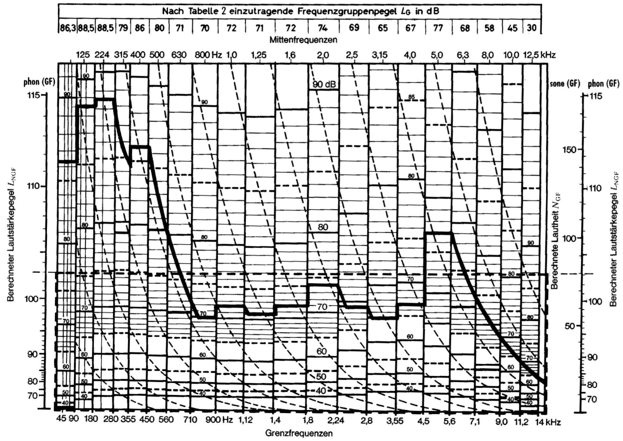


Figure 1.11: *Loudness pattern of an exemplary sound on the curvature template from Zwicker according to the DIN 45631:1991(from the DIN standard(50)).*

high frequency content.

1.4.3.2 Loudness

Instead of the A-weighted sound pressure level, loudness models often provide better estimates for the intensity sensation of sounds. The calculation of the loudness for stationary sounds is standardized in an international standard. The ISO 532:1975 loudness standard includes the work of Stevens (532-A) and Zwicker (532-B)(119). An updated version of Zwicker's method described in part B of the ISO 532:1975 is also implemented in a German national standard DIN 45631:1991(50). According to the German DIN 45631 standard, the specific loudness N' per critical band (from 0 to 24 Bark) is calculated with a curvature template, based on excitation patterns which were measured in masking experiments. Figure 1.11 shows the curvature template with an exemplary loudness pattern derived from third octave band sound pressure levels. The area under the resulting curve in each frequency band equals the specific loudness N' of the sound and an integration over the specific loudnesses of all 24 critical bands (which is equal to the area under the complete curve) yields the overall loudness N [260]:

$$N = \int_{z=0}^{z=24\text{Bark}} N'(z)\ dz \tag{1.14}$$

The latest version of the DIN standard includes the appendix DIN 45631/ A1:2008 [51] extending the calculation method to non-stationary sounds, while remaining compatible with the existing standard for stationary sounds.

A model approach of Moore and Glasberg led to the American standard ANSI S3.4-2007 [9]. In the next revision of the ISO 532 standard, the Zwicker method from the current German standard and the Moore and Glasberg Method according to the ANSI S3.4-2007 will be included into the two parts ISO 532-1 and ISO 532-2 of the revised international standard, respectively [201].

More recent models for the calculation of loudness for time varying sounds are the dynamic loudness model (DLM) after Chalupper and Fastl [46], the time varying loudness (TVL) model of Glasberg and Moore [83]. The optimization of these models especially with respect to covering effects of spectral loudness summation [191] and a better description of technical sounds [192] is subject of current research.

1.4.3.3 Tonality standards

The current standards on tonality are quite different from other psychoacoustic standards like those on loudness or sharpness, because they do not aim at a detailed model of the tonality sensation. The objective of current tonality standards is the description of the disturbing effect and the noise annoyance resulting from tonal components by the prescription of tone adjustments.

Tonality is described in the German national standard DIN 45681:2005 by the tone to noise ratio of a single tone compared to the critical band level, both based on the A-weighted level [54]. Whenever this ratio exceeds 6 dB, it is assumed that a tone is detectable. Depending on the prominence of a tonal component a level penalty (tone adjustment) of up to 6 dB is added to the measured A-weighted sound pressure level.

Another way to describe tonal components is the method of prominence ratio according to the ECMA-74 [60] standard and as an informative appendix in the ANSI S1.13-2005 standard (6). A tonal component is identified as prominent tone, if the level of the critical band centered around the tonal component is more than 9 dB above the average level of the two adjacent (lower and higher) critical bands for frequencies above 1000 Hz. For lower frequencies higher threshold values are applied.

1.4.3.4 Sharpness models and spectral centroid

After having found the sharpness as a discriminable attribute of timbre (237) von Bismarck investigated its dependency from a multitude of signal parameters and gave a first formulation for a simple model of sharpness formation (238). The sharpness according to von Bismarck can be calculated by[5]:

$$S_B = 0.11 \cdot \frac{\int_{z=0}^{z=24\text{Bark}} N'(z) \cdot g_B(z) \cdot z/\text{Bark}\; dz}{\int_{z=0}^{z=24\text{Bark}} N'(z) dz} \text{acum} \tag{1.15}$$

The sharpness S_B is determined by the integral of the specific loudnesses $N'(z)$ over all 24 critical bands which are weighted by a weighting function $g_B(z)$ and normalized by the overall loudness. The weighting function is defined as:

$$g_B(z) = 1 \qquad \text{für } z \leq 15 \text{ Bark} \tag{1.16}$$

$$g_B(z) = 0.2 \cdot e^{(0.308(z/\text{Bark}-15)} + 0.8 \qquad \text{für } z > 15 \text{ Bark} \tag{1.17}$$

Aures extended the model of von Bismarck with a loudness depending normalization factor to cover data with loudness values in the range from 2 sone to 28 sone. The weighting function is approached with an exponential Ansatz $g_A(z)$ (14). The sharpness according to the Aures method can be calculated by:

$$S_A = 0.11 \cdot \frac{\int_{z=0}^{z=24\text{Bark}} N'(z) \cdot g_A(z) \cdot z/\text{Bark}\; dz}{\int_{z=0}^{z=24\text{Bark}} N'(z) dz} \text{acum} \tag{1.18}$$

[5] given here are the mathematical formulations mentioned in the appendix of the DIN 45692 standard (55) allowing for an easy comparison of the three different formulas

with the overall loudness N given by:

$$N = \int_{z=0}^{z=24\text{Bark}} N'(z)\ dz \tag{1.19}$$

with a weighting function:

$$g_A(z) = 0.078 \cdot \frac{e^{0.171 \cdot z/\text{Bark}}}{z/\text{Bark}} \cdot \frac{N/\text{sone}}{ln(0.05 \cdot N/\text{sone} + 1)} \tag{1.20}$$

The sharpness is standardized in a German standard DIN 45692:2009 [55] which follows the general approach of von Bismarck, being independent of the overall loudness value. The sharpness as defined by the DIN standard uses a slightly different weighting function $g(z)$ which can be calculated by:

$$S = k \cdot \frac{\int_{z=0}^{z=24\text{Bark}} N'(z) \cdot g(z) \cdot z/\text{Bark}\ dz}{\int_{z=0}^{z=24\text{Bark}} N'(z)\ dz} \text{acum} \tag{1.21}$$

with the weighting function

$$g(z) = 1 \qquad \text{für } z \leq 15.8 \text{ Bark} \tag{1.22}$$

$$g(z) = 0.15 \cdot e^{(0.42(z/\text{Bark}) - 15.8)} + 0.85 \qquad \text{für } z > 15.8 \text{ Bark} \tag{1.23}$$

and the standardization constant k ($0.105 \leq k < 0.115$) for an adjustment of the calculated sharpness value for the reference sound to 1 acum[6].

An alternative descriptor of the brightness sensation of a sound is the spectral centroid SC (sometimes also called spectral center of gravity SCG). It describes the balance between high and low frequency content in terms of the "center of gravity" of the spectrum of a sound [16]:

$$SC = \frac{\sum_{n=1}^{N} f(n) \cdot x(n)}{\sum_{n=1}^{N} x(n)} \tag{1.24}$$

[6] The reference sound with a sharpness of 1 acum is a critical band wide narrow-band noise ($\Delta\ f\ =\ 160$ Hz), centered at 1 kHz, with a level of 60 dB SPL

In this formulation N is the number of spectral lines, $x(n)$ the magnitude of line number n at the frequency $f(n)$. The spectral centroid is often used as a descriptor of timbre in the context of music, where sharpness is not that commonly used(42,65,147).

1.4.3.5 Roughness and fluctuation strength

Based on the subjective measurements of roughness by Terhardt, Vogel, Fastl and his own measurements, Aures proposed a model for the calculation of roughness(12). The model, schematically shown in Fig. 1.12, is based on a functional model of the ear, consisting of an outer and middle ear filtering, a Bark-scale filter bank (representing the function of the basilar membrane) and a signal rectification by calculation of the absolute value to obtain the envelope of the signal. The signals from the 24 Bark channels are then filtered by a bandpass h_{BP} extracting those envelope modulations relevant for the roughness perception (from 20 Hz to 300 Hz). A normalization of the signals to their direct component leads to a level independent measure comparable to the degree of modulation m^* which is weighted by a spectral weighting factor $g_r(z_i)$ and then squared to match the found relationship of roughness being proportional to the square of the degree of modulation, leading to the partial roughness of each filter. The overall roughness is calculated as the weighted sum of all 24 partial roughness values. The weighting is based on the correlation coefficients between the envelope fluctuations of each filter channel and the two neighboring filter channels correcting for an accumulation of partial roughness values for broadband noise signals.

Newer approaches to model roughness were proposed by Sotteck(210) and by Daniel and Weber(48). A modified version of Sotteck's approach and the model approach by Oetjen *et al.*(173,174) are currently subject to research in the development of a German standard for roughness calculation.

Most recently, Perakis *et al.* proposed the inclusion of the psychoacoustic roughness into a better description of the noise produced by future aircraft engines(184). They proposed a simplified model for a direct estimation of the roughness sensation from the narrow-band frequency spectrum, to facilitate estimates of noise judgments in early technical development stages of aircraft engines. However, their roughness estimation under-predicted the experimental data of a listening experiment in which amplitude modulated tones were compared with tone pairs. They

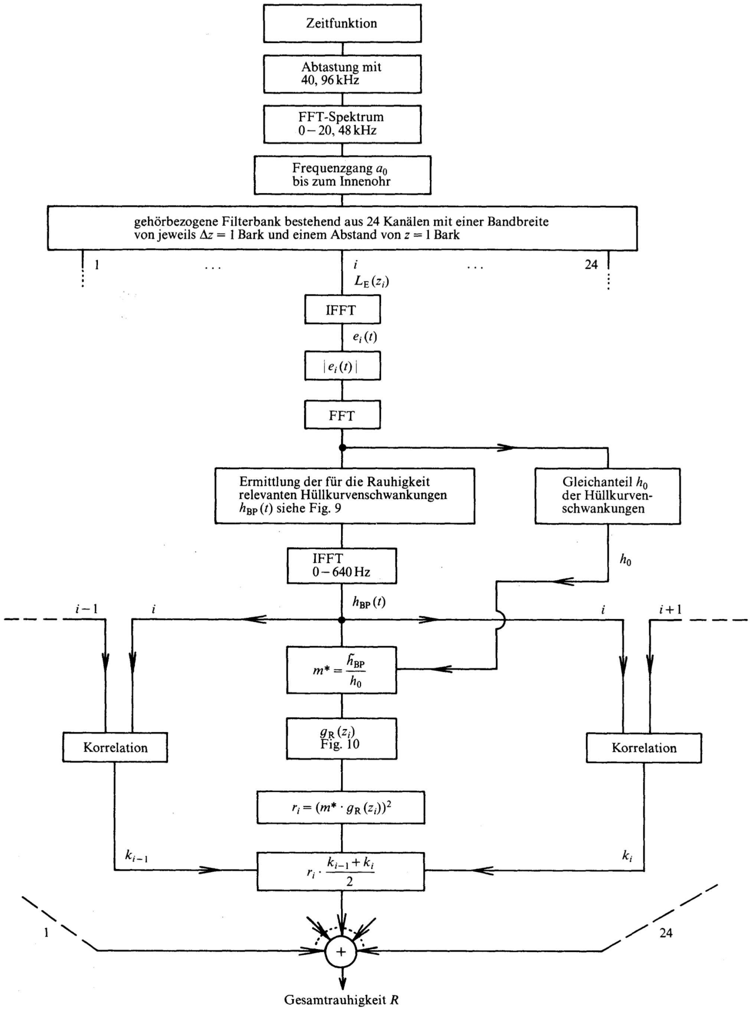


Figure 1.12: *Scheme of the roughness model by Aures (from Aures, 1985[12]).*

request further work on the roughness of sounds consisting of multiple tonal components especially when the tones are spread over several critical bands.

For the sensation of fluctuation strength, Zwicker and Fastl proposed a simple model based on the temporal variation of the masking pattern[(263)]. In their formulation (eq. 1.25) the fluctuation strength F is proportional to the level difference between maximum and minimum of the temporal masking pattern called temporal masking depth ΔL and the modulation frequency f_{mod}. Instead of a global masking depth ΔL, they propose an integration over the temporal masking depths over frequency along the critical-band rate. The denominator in equation 1.25 is responsible for a peak of the curve at a modulation frequency of $f_{\mathrm{mod}} = 4$ Hz which reflects the maximum of the fluctuation strength found at this point.

$$F = \frac{0.008 \cdot \int_{z=0}^{z=24\mathrm{Bark}} (\Delta L/\mathrm{dB\ Bark})\ dz}{(f_{\mathrm{mod}}/4\ \mathrm{Hz}) + (4\ \mathrm{Hz}/f_{\mathrm{mod}})}\ \mathrm{vacil} \tag{1.25}$$

Recently, a new model approach, which is inspired by a hearing model, was proposed by Zhou[(259)]. None out of the proposed models for fluctuation strength are standardized yet.

1.4.3.6 Speech interference level (SIL) and speech intelligibility index (SII)

The speech interference level (SIL) is an acoustic parameter, primarily intended to characterize the interference of noise with speech communication[(21)]. It describes the level of the noise in the frequency range which is most relevant for speech. Three different variants exist - SIL-3, SIL-4, and P-SIL (preferred SIL) - which are based on the mean sound pressure level over three or four octave bands:

$$\text{SIL-3} = \frac{L_{p1000} + L_{p2000} + L_{p4000}}{3}\ dB \tag{1.26}$$

$$\text{SIL-4} = \frac{L_{p500} + L_{p1000} + L_{p2000} + L_{p4000}}{4}\ dB \tag{1.27}$$

$$\text{P-SIL} = \frac{L_{p500} + L_{p1000} + L_{p2000}}{3}\ dB \tag{1.28}$$

where L_{pX} is the unweighted sound pressure level of the octave band centered at X Hz. The current ANSI S12.65-2006 standard is defined as the SIL-4 which is based on four octave bands(7). Since the 1950s, the SIL measures are used in the aviation industry to characterize speech degradation by interfering noise, especially in situations where speech communication is essential in terms of security (e.g. in the cockpit)(21,41). In addition, the SIL measures are also used for a characterization of the cabin noise and its potential interference with the desired communication between passengers.

The speech intelligibility index (SII) is an alternative single value measure predicting the percentage of available speech cues for a listener in noise. A value of 1 (100%) indicates that all speech cues are available, no speech cues are available for a value of 0 (0%). The calculation is based on the amount of speech energy, remaining after hearing thresholds, masking effects and distortion by the noise, summed over the critical bands. Different calculation methods and standard speech levels (for different amounts of vocal efforts) are specified in the ANSI S3.5-1997 standard(10).

1.4.3.7 Universal estimates for the appraisal of sounds from acoustic variables

Based on the different findings about how basic auditory sensations generally relate to the unpleasantness of sounds, which are described in more detail in section 1.6 of this thesis, attempts were made to derive universal estimates of the appraisal based on psychoacoustic descriptors. Examples are the "sensory euphony"[7] proposed by Aures(13), "sensory pleasantness" by Zwicker and Fastl(263), and "unbiased annoyance" by

[7] German: Sensorischer Wohlklang

Zwicker[262]. Due to a lot of other additional factors like emotions, attitude, knowledge or experience contributing to noise annoyance the role of acoustic variables is seen as being rather small[87] and especially the attempt to derive an unbiased annoyance is also critically seen[89]. To establish a sound relationship between audio descriptors and the evaluation of sounds it is necessary to find suitable descriptors that reflect the perceptual dimensions of a particular sound or class of sounds, which considerably limits such estimation models[35]. Some examples of very specific sound descriptors, which were found to be related to the (un-) pleasantness (or annoyance) of sounds, are presented in section 1.6.2.

1.5 Methods for the rating of sounds

In psychoacoustics, different methods are used to determine the relationship between physical signal properties and the corresponding subjective sensation or appraisal of a sound. For the measurement of thresholds and just noticeable differences, the method of constant stimuli[81] and adaptive procedures[131,140] are examples of commonly used uni-dimensional methods. Examples for supra-threshold methods are the method of magnitude estimation[144,263], paired comparisons[177] and multi-dimensional scaling techniques[218].

In the following those supra-threshold methods which are used throughout this PhD thesis will be presented in detail: Categorical rating scales, their application in the multi-dimensional technique of the semantic differential and adaptive matching methods, which are used to determine the point of subjective equality (PSE).

1.5.1 Categorical rating scales

Categorical scales are a simple way to measure sensory attributes on an uni-dimensional rating scale by assigning them to categories which are visually presented to a participant[38]. Commonly used are scales with 3 to 11 categories. The number of category scale points has to be chosen by the investigator based on the anticipated abilities of an observer to differentiate between objects and the demanded degree of differentiation. The upper limit of categorical scale points is given by the capacity of humans for processing information. Miller identified the capacity to differentiate between objects along a one dimensional scale (e.g. loudness) as "the magic number seven, plus or minus two"[156]. This means that, if observers are presented with a multitude of stimuli, which they are requested to assess, one can expect them to reliably differentiate only about seven (plus minus two) different categories. Besides this limitation, the number of categories also expresses the degree of differentiation between objects which is requested by the researcher[195].

To obtain results on an interval scale, the scale categories need to represent equally spaced intervals which are interpreted and used by the observer as such[38,81]. In this sense, the method of categorical scaling is classified as a special case of partition scales[144]. To convey the equal spacing to the participant, rating scales are verbally, numerically and/or graphically symbolized and visually presented to the observer. Common verbal labels, that denote steps which are interpreted by the observer as equally spaced, are the so called Rohrmann attributes[195]. Numerical labels can be used to indicate the uni- (e.g. from 1 to 10) or bi-polarity (e.g. from -3 to 3) of a scale. An example for a purely graphical symbolized scale, is the self assessment manikin (SAM[39]). Figure 1.13 shows an exemplary 5-step bi-polar scale with numerical labels.

For the scaling of loudness over a broad range of stimulus intensities, especially in the audiometry context, category subdivision scales are used[103]. Category subdivision scales consist of 5 verbally distinguished main categories that are each subdivided by a ten-point partitioning scale. The assessment using these scales follows a two-step procedure - first on the raw verbal scale and then on the fine numerical scale. Besides loudness scaling, subdivision scales also proved to be useful in noise assessments[101,102].

In the direct and absolute measurement on a category scale, a variety of judgment errors and bias might occur[38,187,252]. Examples are:

— Ceiling- and Floor-Effects: Due to the lack of knowledge about the distribution of a characteristic over the stimulus set, sounds with extreme characteristics are all put into the highest (Ceiling-Effect) or lowest (Floor-Effect) category, although they differ with respect to that characteristic.

— Effect of Central Tendency: Extreme values of the scales are avoided - they are reserved for potential objects with extreme characteristics

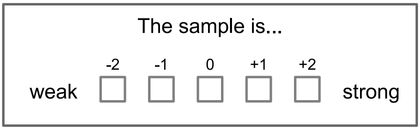


Figure 1.13: *Exemplary 5-point categorical rating scale for the adjective pair* weak - strong *with numbered categories*

— Halo-Effect: Observers are not willing or able to distinguish between different scales and rate objects based on a underlying sweeping judgment.

— Primacy-Regency-Effect: The sequential presentation of the object affects the judgments. Objects presented earlier influence the judgment of later presented ones.

To reduce unavoidable context effects, a uniform frame of reference should be realized for all participants by using the same written instructions and an orientation phase with a presentation of all stimuli prior to the evaluation of the sounds [64,256]. Furthermore, the composition of the stimulus set [256], the stimulus intensity, the frequency of presentation for different stimuli should be carefully considered [144], and a verbal fixation of categories can help to establish a uniform frame of reference [195].

Taking these factors into account, categorical ratings scales can provide an insight into the overall level of a rating compared to the method of absolute magnitude estimation which better captures the rate of increase for a sensation of interest [102]. In addition, categorical rating scales are an economical method of scaling which can be easily understood by participants. Therefore, categorial scales are also used in the multi dimensional measurement method of the semantic differential, which requests a large number of judgments on multiple scales from the participants.

1.5.2 Semantic differential

The semantic differential is a multi-dimensional technique which was initially developed to measure personality traits and the connotative meaning of linguistic concepts [175]. The concepts are judged with respect to different oppositional (bi-polar) adjective pairs (or antonym pairs, often also called items) on multiple 5- or 7-point categorial rating scales.

In specific applications, semantic differentials are often composed differently than the universal semantic differential proposed by Osgood and include context specific items that are closely related to the particular investigated objects[38,47]. These specific items are in most cases included as uni-polar items, if direct antonyms are hardly available for them. For the assessment of sounds, items characterizing the meaning (e.g. weak - strong, functioning - not functioning), items describing sound characteristics (e.g. soft - loud, low - high, dull - sharp), and items used by expert listeners (e.g. tonal - noisy) are used together with evaluative items (e.g. pleasant-unpleasant)[3,64,237]. Multiple items can be arranged together for the assessment of one sound at a time or presented solely to the observer for a successive assessment of multiple sounds. Parizet found that both methods deliver similar results but the latter method proved to be faster and more accurate than the other, because listeners are more concentrated on the respective scale[176].

After a factor analysis of the multiple ratings on the different adjective scales, a limited number of factors is extracted which explain the majority of the underlying variance. The three major factors found in the initial studies of Osgood were labeled "Evaluation", "Potency" and "Activation."[175] Since the first usages of the semantic differential, this EPA-structure was found in most of the analyses of semantic differential data over many different studies and with a wide variety of judged concepts [22,100]. Even the order and the relation of the explained variance shown in Fig. 1.14 is rather robust over studies. The "Evaluation" factor often appears first, accounting for about half to three-quarters of the extractable variance. The second ("Potency") and the third ("Activation") factor typically account for approximately half of the variance of the first factor, each. The following further factors, if apparent, account for less than half of the variance of the second or third factor[175].

For acoustic stimuli, three factor solutions of the factorial analysis are also quite common. The three factors can often be interpreted as one factor related to the pleasantness of the sounds ("Evaluation"), one factor describing the powerfulness of the sound ("Potency") and one factor related to the timbre ("Activity")[134,165]. The *pleasant* factor usually covers the items related to the evaluation of the sounds like e.g. pleasant - unpleasant, annoying - not annoying. The *power* factor often shows high loadings on adjectives describing a sound as e.g. loud, powerful or dominant. Due to high loadings of the item pair deep - metallic on the

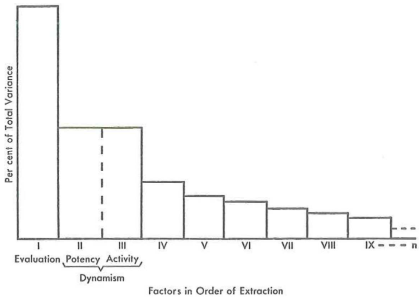


Figure 1.14: *General occurrence of factors from semantic differential data and the relative percentage of variance explained (from Osgood* et al.*, 1957*[175]*)*

timbre factor, this factor is sometimes also called *metallic* factor[165]. Similarly, von Bismarck found sharpness as an important attribute of the timbre for steady state sounds[238].

The three perceptual dimensions *pleasant*, *power* and *metallic/timbre* were found similarly for a variety of sounds including music, vehicle noise, aircraft noise, octave noise bands, pure tone complexes[128], musical sounds, differing in timbre[130], broad-band noise mixed with frequency-modulated sounds[165], broad-band noise mixed with amplitude modulated sounds[134] and also for tonal components embedded in red noise[94].

Sköld *et al.* found the four factors *quality*, *audibility*, *tonal content* and *modulation* for car interior sounds during slow acceleration[206]. Here the timbre factor is divided into two factors - one factor related to spectral aspects and the other related to the temporal structure. This is also the case in the following studies.

Klein *et al.* found overall four factors covering the *hedonic judgment*, the *temporal characteristics*, the *spectral features* and the *activity* for single pass-by recordings from traffic noise. Similarly, the factors *evaluation*, *timbre*, *power* and *temporal change* were also found for a broad range of environmental sounds[258]. Slightly different factors were found for HVAC sounds from cars by Hohls *et al.*[111]. They identified three factors: Factor one describes the *power, pleasantness and quality*, factor two the *timbre and pitch* and factor three *modulation and tonality.* In their study, the *pleasantness* and the *power* seem to be very closely related, forming a joint factor, while the two other factors refer to different aspects of the sound character. Similar factors were also obtained for aircraft cabin noise, evaluated with the same semantic differential in a acoustic laboratory and in an aircraft cabin simulator[(183)]. In both experimental environments, Pennig *et al.* found three perceptual dimensions related to the *pleasantness*, the *variation* and the *brightness* of the sounds, whereas the pleasantness factor included items like *quiet - loud* and *unobtrusive - obtrusive* which are usually loading highly on the *power* factor.

1.5.3 Point of subjective equality (PSE) and matching methods

Matching methods are a way to determine points of subjective equality (PSEs)[(208)]. A PSE is usually determined by a variation of a physical signal parameter of a reference sound until subjective equality with a test sound is reached. The matching of the physical parameter can be realized as an adjustment, which is controlled by the listener, or in an adaptive procedure in which the physical parameter is varied based on choices made by a participant in paired comparisons. In a comparison of the test and the reference sound, the PSE characterizes the situation where a listener experiences both sounds as identical with respect to the attribute of interest. Hence, in a decision task the probability to choose one of the two sounds is balanced - 50 percent for each of the two possible choices - and the PSE represents the 50 percent-point of the psychometric function. The value of the varied physical signal parameter of the reference sound at the PSE is used as the quantitative measure to describe the test sound. The measurement of a PSE can also be realized as an adjustment of the test sounds which are compared against a fixed reference.

Perhaps the most common example for PSE measurements is the matching of the loudness by a variation in sound pressure level, which was used as a basis for the equal loudness level contours[(72)]. Here the definition of the loudness level of a sinusoid measured in phon is based on a loudness matching of the sinusoid and the 1 kHz reference sound. A loudness level of N phon is defined as the loudness of a 1-kHz reference sinusoid at N-dB SPL which functions as the common currency to measure loudness[(143)].

In recent literature, direct matching by an adaptive procedure is often used to determine loudness matches[68,77,114,115,122,143]. This way to determine the loudness of a sound by matching it to another is also seen as "the gold standard to which results obtained by other methods must conform"[143]. One advantage of matching methods is that participants are asked to make a comparison between two sounds, one being a given (external) reference, instead of comparing the test stimulus to their internal reference in absolute judgments (like,e.g., in categorial scaling, magnitude estimation). In this way, direct matching methods do not rely on the ability to translate sensation magnitudes into responses expressed by numerical values and they seem to be rather easily performed by participants[216]. Furthermore, only a very small amount of short term memory is required for the matching process compared to an absolute categorical scaling, where the mapping of stimulus magnitudes to categories has to be kept in mind over the whole duration of the scaling process. For the matching procedure, only the impression of a preceding sound needs to be memorized for a comparison with actual sound.

This way of subjective measurements, which is based on the internal matching of sensation magnitudes without using any numbers in the judgment process, was called *natural measurement* by Zwislocki[264]. In his understanding, numbers are introduced automatically to quantify the subjective judgments and an introduction of physical units then makes a characterization of the judgments as a function of associated physical variables possible. Hence, the usage of a physical, stimulus bound scale to quantify subjective judgments offers the advantage that it directly has a meaningful unit like,e.g., the dB-scale[32,171].

The point of subjective equality can be measured with adaptive procedures (including the simple staircase)[140] or also the method of adjustment[221]. Sometimes, PSE values are indirectly determined by the method of constant stimuli[78,209], from rating scale data (based on mean relationships between scale values and levels of a reference sound)[109,198], from magnitude estimation data (by fitting a power function to transform loudness into corresponding sound levels)[110,168] or by an individual master scaling transformation[23,171]. Even though the indirect determination of loudness PSEs from magnitude estimation are often in good agreement with direct matches[166], data gained from magnitude estimation can be affected by significant bias effects (like, e.g., the range effect)[145]. Reckhardt[190] showed that the influence of start levels on the PSEs is smaller for an adaptive procedure than with the method of constant stimuli. In their study, the consideration of individual differences in the absolute threshold also led to a reduction in inter-individual differences.

Stevens already showed, that also cross-modal matches between the loudness of a sound and other sensory continua are possible[213]. Pollack determined several different aspects of sounds (like volume, annoyance, density and force) as sound pressure levels at the respective points of subjective equality and distinguished between the judged attributes by a comparison with the PSEs for loudness[186]. Similarly, combined measurements of the PSEs for loudness and for annoyance[170] or loudness and unpleasantness[196] of different types of noise signals were determined using the method of adjustment.

1.6 Quantitative and qualitative acoustic variables and their relationship to the appraisal of sounds

Even though acoustic variables are limited in explaining the variance found in product sound quality or noise annoyance studies (see also Sec. 1.4.2), still a lot of acoustic factors have been found to explain a considerable part of the variance found in sound quality or annoyance judgments. Especially, quantitative acoustic variables, like the A-weighted sound pressure level and the loudness sensation have been found to be closely linked to the appraisal of sounds, especially if sounds are unwanted. This is not unexpected - an increase in sound pressure level, usually yields a higher loudness which often results in a higher disturbance or nuisance effect. Examples showing the close link and the particular differences between the physical dB-measures, the perceptual attribute loudness and different attributes used to express the appraisal of sounds are given in Sec. 1.6.1.

Differences in the appraisal are also found for sounds which have the same sound pressure level or equal loudness. In such cases, sounds are qualitatively dissimilar due to differences in the sound character. Thus, qualitative acoustic variables, like the tonality, the sharpness and temporal descriptors were also found to be linked to the appraisal, which is described in Sec. 1.6.2 in more detail.

Differences in sound character are often hidden below the dominant effect of quantitative acoustic variables like the loudness. Therefore, measurements of the sound character are usually carried out with loudness equalized stimuli. This opens a gap between the measurement of quantitative and qualitative sensory continua. The quantification of relationships between differences in terms of the sound character and differences in intensity becomes a challenging task. However, in a few studies the decibel is used as a common currency for both aspects - a view which is also followed in this PhD-thesis and explained in Sec. 1.6.3.

1.6.1 Quantitative acoustic variables related to the appraisal of sounds - The close link between sound pressure level, loudness and appraisal of sounds

In many studies a close relationship between physical measures of the intensity of a sound, like the sound pressure level, the loudness sensation, reflecting the perceived intensity of a sound, and the appraisal of sounds was found. Examples are high correlation coefficients between noise level metrics and different evaluative attributes determined between

— laboratory annoyance of different traffic noises judged on categorical rating scales and L_{eq} values[(116)],

— short term annoyance of industry noises judged in a free categorization task and L_{Aeq} values[(1)],

— noisiness of repeated impact sounds rated as PSEs with the method of constant stimulus and L_{pE} values[(209)].

Due to the major influence of sound pressure level on the loudness sensation, also tight links between the loudness sensation and the appraisal of sounds are found. High positive correlation coefficient were obtained between annoyance and loudness judgments

— for single and combined community noises judged with magnitude estimation[(27)],

— for indirectly determined PSEs for environmental sounds[(109)],

— and for categorical ratings of gear noise[(202)].

Similarly, high negative correlation coefficients were determined between loudness and preference/pleasantness judgments:

— for high intensity everyday sounds rated on categorical scales [257]

— for noise with different slopes used in speech masking judged on a visual analogue scale [113]

— for laundry dryers judged with a semantic differential [211]

— for exterior car sounds [241]

— interior aircraft sounds [245].

In all these examples, an increase in loudness was related to a an increase in annoyance or a decrease in preference or pleasantness judgments, respectively.

A possible explanation for the close link between loudness judgments and the evaluation of sounds might be that, even though carefully instructed, participants have a hard time separating the loudness and the annoyance dimension and sometimes include feelings of annoyance into their loudness judgment for complex sounds [98]. Namba similarly reports that loudness judgments may be affected by the impression of noisiness or annoyance, leading to high correlation coefficients between loudness and annoyance judgments, especially for real sounds [166]. Zeitler also found significant differences in the loudness judgments of warning and alarm sound between original sounds and a version which was neutralized with respect to the identifiability of the sounds [256]. The neutralized sounds maintained the loudness and the spectral envelope of the originals while being unidentifiable by the listeners. The differences in the loudness judgments between the two conditions were significantly correlated with the evaluation of the sounds on the scale unpleasant - pleasant of a semantic differential. Another potential explanation is given by Völk [236]. In their study, with various environmental sounds, extreme annoyance judgments occurred whenever the natural loudness of a sound was exceeded. Participants of the listening experiments stated that already small increases of loudness above the "natural" loudness, made the participants judge the sounds to be rather annoying, even though they were judged rather neutral at their "natural" loudness.

On closer inspection, there is also evidence for differences between the judgments of the perceptual and evaluative attributes beyond high correlation coefficients between them. Especially for sounds with salient sound characteristics, the differences between the particular descriptive or evaluative attributes become clear.

Hellmann determined loudness, noisiness and annoyance judgments for a large variety of stimuli, like single tones in noise [105], single tones in low- and high-pass filtered noise [106], and two tone noise complexes [107]. Using the method of absolute magnitude estimation she found that annoyance is most closely related to loudness but the annoyance-to-loudness ratio was not constant. Depending on the particular tone to noise configuration ratios around unity, larger than unity and level dependent ratios were found [105,107]. This means that, with increasing sound pressure level, the growth of the loudness and the annoyance judgments is different from each other. Similarly, Berglund *et al.* found ratios between annoyance and loudness scale values greater than unity for the community sounds [26,28] and aircraft flyover noises [25]. They used a master scaling method to express loudness and annoyance judgments as pink noise equivalents on a sone scale [28].

The different growth of the particular perceptual and evaluative attributes with overall sound pressure level enables a distinction between them. This distinction becomes even more directly visible, when judgments are quantified as levels at the point of subjective equality (PSE). Sone *et al.* investigated the relationship between sound exposure level of repeated impact sounds and the PSEs of a continuous comparison stimulus. The judged attributes *loudness* and *noisiness* could be distinguished by different values for the slopes, each [209]. Hiramatsu *et al.* directly compared the PSEs of the attributes loudness, noisiness and annoyance [109]. In their laboratory study, the PSE levels of the sounds were calculated by comparing the categorial judgment of the 59 test sounds with the categorial judgments of white noise signals at different levels. The test sounds included natural sounds, human sounds, mechanical sounds, indicator sounds and white noise at different dB(A)-levels. Based on the linear relationships between the level and loudness/noisiness/annoyance ratings of the white noise sounds, the PSE levels for all other sounds were calculated. Figure 1.15 shows the three pairs of PSEs plotted over each other. In addition to the overall high

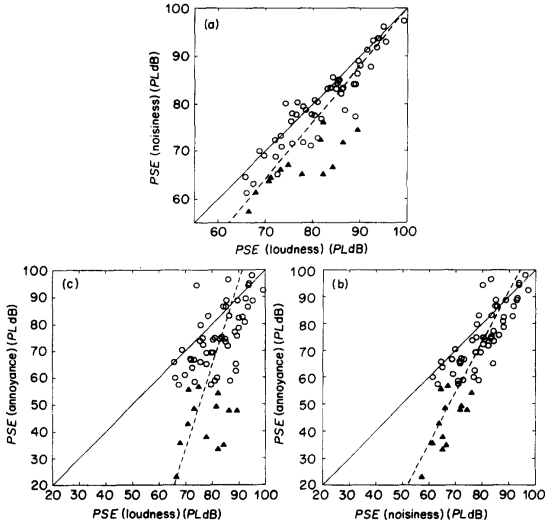


Figure 1.15: *Comparison of the indirectly determined PSEs of annoyance, noisiness and loudness from Hiramatsu* et al.*, 1988*(109)*. The relationship between the PSEs for annoyance and loudness are shown in subplot (c), which shows a stronger increase of the PSEs for annoyance than for loudness.*

correlation coefficients between all three attributes, differences in the slopes of the linear regressions between them were found. A value of 3.11 was found for the slope of the linear regression between the PSEs for annoyance and loudness. This means a change in the loudness PSEs of 10 dB equals a change in the annoyance PSEs of about 30 dB.

From all these findings it can be concluded that, even though the sensation of loudness and different attributes expressing the appraisal of sounds are often highly correlated with each other, the growth rates with sound pressure level are different between them. Changes in sound pressure level yield a different growth of the subjective magnitude for the particular attributes. Depending on the experimental methods used, these differences become more or less apparent and sometimes only high correlation coefficients and not the proportionality constants are given in literature data. Approaches to collect data of different perceptual aspects allow for more detailed insights into the relationships between the physical measure of sound pressure level, the loudness sensation and the appraisal of sounds on the connotative level.

1.6.2 Qualitative acoustic variables related to the appraisal of sounds - The influence of sound character on the appraisal of sounds

For sounds having the same dB(A)-level, many studies have shown that the sound character together with the loudness of the sounds contributes to its pleasantness or annoyance(188,189). Different sound characteristics, which can be described by psychoacoustic metrics, were found to be relevant in terms of the evaluation of a sound. In the following some specific examples of sound characteristics, which are relevant in terms of evaluation, are presented.

1.6.2.1 The influence of tonal components on the evaluation of sounds

Many studies found that prominent tonal components contribute to the annoyance or unpleasantness of a sound. Examples are single tonal components embedded in low-pass and high-pass noise(106) and embedded in a flat broadband spectrum(108) (inspired by the broadband spectrum of aircraft and machinery noise) or single pure tones embedded in white noise, low-pass filtered noise, and fan noise(155). Similarly, tonality (together with the Zwicker loudness) was found to influence the annoyance from steady-state aircraft interior noise containing tonal components(4). For tramway noise, the total energy of tonal components (TETC) above 12 Bark contributed to the short-term annoyance(230). In contrary to these results, an increase in tonality was also found to be related to a higher judged sensory pleasantness(263). Thus, from this examples it becomes clear that it depends on the context whether tonal sounds have a positive or a negative connotation.

1.6.2.2 The influence of spectral aspects on the evaluation of sounds

With respect to spectral aspects, sounds dominated by mid frequency and high frequency content are in most cases judged to be the rather dissatisfactory (e.g. in speech masking applications[113]) or for thresholds of discomfort in hearing aid fitting with environmental sounds[249]. Such findings can often be rather well described by the psychoacoustic sharpness, and, in general, pleasantness is found to decrease with an increase in sharpness[13,263]. Further examples are the sharpness as a factor related to the perceived annoyance of complex community sounds[28], for industrial noise[40]. For car interior HVAC noise[8] with equal loudness, the sharpness was also found to be the best descriptor of preference evaluations and lower sharpness was preferred[112]. The psychoacoustic sharpness was also proposed as a basis for a quantitative measure of the intrusiveness of sounds[189].

Similar to the sharpness, the spectral centroid and spectral spread have been found to explain a considerable amount of the variance in the annoyance judgments for sounds with equal L_{Aeq}[126]. Beside this general increase of unpleasantness with increasing sharpness, Weber and Eilers found a minimum in unpleasantness judgments of amplitude modulated, bandpass filtered noise for a sharpness in the range from 0.6 to 1 acum[250].

1.6.2.3 The influence of temporal aspects on the evaluation of sounds

In terms of temporal factors, Hiramatsu *et al.* found that the fluctuation frequency of sinusoidally, saw-tooth and randomly modulated noise has little effect on the judged annoyance below 5 Hz. For higher fluctuation frequencies a tendency of increasing annoyance with rising fluctuation frequency was found[110]. In extension to higher modulation frequencies, the psychoacoustic roughness was found to be one major temporal factor influencing the pleasantness or annoyance in several studies. An increase in roughness was found to decrease the relative "euphony"[13] and sensory pleasantness[263].

Traffic noise annoyance was found to be depending on the spectral content, and the temporal structure of the noise. For single vehicle pass-by noises, the *sputtering* and *nasal* sound character, driving the annoyance, could be effectively described by indices characterizing the fluctuations of the envelope in different modulation frequency ranges[129]. *Sputtering* is located in the modulation frequency range from 2 Hz to 100 Hz and

[8]HVAC: Heating Ventilation and Air Conditioning

modulation frequencies in the range from 100 Hz to 200 Hz contribute to the *nasal* aspect.

1.6.2.4 Relationships between perceptual and affective dimensions in the judgment of aircraft cabin noise

Västfjäll *et al.* investigated the relationship between perceptual dimensions and affective evaluations and affect state for aircraft cabin noise over several studies (241,243–246). They found a link between the five perceptual factors determined in a semantic differential and the two core affect dimensions (valence and arousal), judged in a separate experiment. The valence (pleasantness) dimension was negatively correlated with the loudness factor and positively with the naturalness factor of the semantic differential. The affect dimension of arousal (activation) was positively linked to the sharpness factor and negatively to the tonal content factor (245). They further found that the two core affect dimensions differ in their relationship to psychoacoustic descriptors. Activation was highly correlated to the qualitative, sound character describing variables like roughness and tonality (241) or to the sharpness metric (243,245). Valence was found to be mainly linked to intensity related psychoacoustic descriptors like loudness (241,243) or loudness and roughness (245). Changes in overall stimulus level over a range of 40 dB were found to mainly affect the valence dimension, but also the activation dimension was slightly influenced (243). In terms of overall preference towards affective reactions, or affective evaluations, they found a dependence on both of the two affective dimensions valence and activation (241,244). In summary, their finding suggest that it might be feasible to relate the contributions of perceived intensity and sound characteristics to preference evaluations by adjustment of the overall level, each.

1.6.3 A common currency for quantitative and qualitative aspects of sounds

As was shown in the previous two subsections 1.6.1 and 1.6.2, the evaluation of sounds is often linked to the quantity of a sound (in terms of its loudness or the dB(A)-level) and the qualitative differences in sound character. However, the tight link between the loudness and the appraisal of a sound yields a dominant influence of loudness differences on evaluative judgments[28,40]. Whenever prominent loudness cues are apparent, it might be rather difficult for a listener to identify more subtle sound character differences e.g. for preference evaluations of car interior sounds[217]. Therefore, to enable the emergence of attributes related to the sound character or timbre, an elimination of the dominant influence of the loudness by a loudness equalization is often recommended[219]. In compliance with the definition of timbre according to the ANSI standard, which refers to dissimilarity between sounds, having equal loudness (and equal pitch)[5], the assessment of timbre is in most cases also based on loudness equalized stimuli, to eliminate loudness as a factor[85].

However, the perceptual representation of sounds, in terms of the dimensions of the perceptual space obtained from distance scaling and an MDS analysis, was found to differ between loudness-equalized and non-equalized sounds[220]. This raises concerns in the applicability of sound characteristics obtained for loudness equalized sounds to real-life applications where sounds differ with respect to intensity (e.g. sound level or loudness) and sound character. The equalization of the loudness, which is intended to separate dominant loudness cues from differences in the sound character, opens a gap between the characterization of quantitative and qualitative aspects of sounds, because a combined measurement of both aspects at the same time seems to become incompatible with each other. Similarly, in engineering problems, alternative sounds do in most cases differ with respect to both aspects - loudness and sound character. Thus, for a successful consideration of alternatives the relationships between between loudness and sound character are required to bridge the gap between quantity and quality. Especially, knowledge about the conversion ratio between quantitative and qualitative aspects is desired.

1.6.3.1 The decibel as a common currency for the measurement of quantitative aspects and also sound character

The close link between the sound pressure level and the appraisal of sounds can be used for a measurement of a particular attribute by matching a reference sound to an equality judgment by means of a level adjustment. In this way, equal annoyance matches of broadband stimuli with dominant low frequency noise[84], equal unpleasantness contours of low frequency pure tones[117,196] and equal preference matches for the evaluation of speech transmission circuits (162) were successfully determined. The direct adjustment of stimulus levels was also used to determine equal-aversion levels for pure tones of different frequencies (158). In summary, it seams feasible to express qualitative differences between sounds in terms of sound level differences and use the decibel as a "common currency" which is also similarly used for the perceptual scaling of loudness.

Cardozo *et al.* aimed at measuring the contribution of sound character to the annoyance for a wide variety of sounds (43–45). They followed the idea to decompose the factors affecting the appraisal of a sound (e.g. the annoyance) by projecting all intensity related factors to one dimension defined by the A-weighted sound pressure level and to summarize all other factors by a second dimension denoted as sound character. Thus, sound character was defined by them as everything that differentiates two sounds with the same A-weighted sound pressure level in terms of their annoyance from each other. For two sound sources having the same A-weighted L_{eq}, the difference between the annoyance ratings is understood as a measure of the difference in sound character. Based on annoyance judgments of multiple stimuli at two fixed levels, they derived levels of equally annoyance for the sounds (43). Similarly, Preis proposed to extend an existing noise rating standard (e.g. L_{eq}) by adding a correction term, describing the intrusiveness of the sounds which would not be reflected correctly by the current rating standard alone (188). This is also one underlying idea of the concept of rating level, which is used in the assessment of noise immissions (described in more detail in Sec. A.1 in the appendix).

However, only three approaches were found in the literature which aimed at an explicit distinction of quantitative perceptual measurements and the appraisal of sounds to unscramble their relationship based on a measurement of points of subjective equality (PSEs). Niese measured the loudness and the annoyance of different test sounds by the method of adjustment, both with a variation of sound pressure levels (170). In separate experiments participants were asked to adjust the level of a reference sound to equal loudness and to equal annoyance compared to the test sound. The level differences between the results from the annoyance and the loudness matching were used to characterize the annoyance of the test sounds. The test sounds (howling noise, German: Heulrauschen) were found to be 4 dB more annoying than loud, compared to broadband stimuli. Similarly, Ronnebaum and Weber determined PSEs for loudness and unpleasantness for single tones and also narrow-band noises for center frequencies between 20 Hz to 200 Hz against a reference located at 100 Hz which was fixed in level [196]. The PSEs for unpleasantness were below the PSEs for loudness especially for the very low center frequencies, which means that equally loud sounds would have been considerably more unpleasant. The differences between the loudness and the unpleasantness judgments were about 5 dB for the narrow-band noise signals in their study. Pollack showed that it is possible to specify several different aspects of sounds (like volume, annoyance, density and force) as sound pressure levels at the respective points of subjective equality compared to an equal loudness level [186].

These studies show that, in principal, it is possible to quantify qualitative aspects in the form of an associated level of the test sound or the level of a reference sound on a dB-scale. In addition to that, Pollack and Niese explicitly used the level difference between the PSE of annoyance (on the connotative level) and the PSE for loudness (on the denotative perceptual level) as a quantitative measure, describing the contribution of sound characteristics to the appraisal of the sounds.

1.7 Method for a quantification of preference relevant sound character differences

In this thesis, the influence of the multi-dimensional aspect of sound character on the appraisal of sounds is measured by a projection onto the dB-scale[9]. The measurement is carried out indirectly by a combined determination of the quantitative perceptual attribute *loudness* and the appraisal of the sounds in terms of the *preference*, both as levels at the points of subjective equality (PSE, described in subsection 1.5.3).

Given the observation that a decrease in level reduces the loudness and the annoyance or unpleasantness of a sound, the following three steps are taken to quantify preference relevant differences of the sound character in an adaptive paired comparison procedure:

[9] After having successfully applied this idea, the author became aware of three similar approaches in the literature that were presented in detail in the previous section 1.6.3.

1. Choice of a suitable reference sound with a fixed dB(A)-level L_{ref}.

2. Determination of the dB(A)-level L_{loud} for the test sound at which the test sound is equally loud as the reference sound (PSE for loudness).

3. Determination of the dB(A)-level L_{pref} for the test sound at which the test sound is equally preferred as the reference (PSE for preference).

The *preference* is measured as the level of the multi-tone test sound L_{pref} at which it is equally preferred as the fixed reference sound. Due to the close link between the sound pressure level and the evaluation of a sound, which was extensively presented in the previous section 1.6, it is assumed that, if the level of the multi-tone sound is considerably attenuated compared to the fixed level of the reference sound, then it will be preferred, simply because the reduced level facilitates the preference. If the level of the multi-tone sound is set higher than the reference, then the reference will be preferred. Accordingly, the level of a multi-tone test sound is reduced, every time it is not preferred and raised if it is preferred in the adaptive procedure. In between these two conditions lies the PSE for preference.

The loudness is similarly measured like in classic loudness matching experiments, which were presented in Sec. 1.5.3, with an adaptive procedure converging at the PSEs for loudness L_{loud}. The level of a multi-tone test sound is decreased every time it is judged to be louder than the fixed reference sound and it is increased, if the reference sound is judged to be the louder one.

The current method (schematically shown in Fig. 1.16) further follows the general finding that the appraisal of a sound (here in terms of preference) is related to the two core affect dimensions - valence and arousal - which were found to be linked to the quantitative (loudness) and qualitative attributes (sound characteristics) of the sounds, respectively[(241,244)]. It is assumed that the preference PSE L_{pref} can be expressed by the quantitative loudness judgment L_{loud} and an additional term which subsumes all qualitative aspects related to the sound character that are not already covered by the loudness judgment. This additional term is

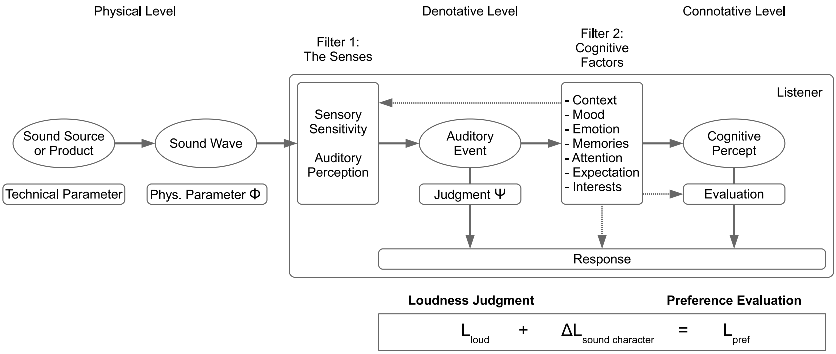


Figure 1.16: *Schematic of the method with a combined measurement on the denotative and the connotative level. The loudness judgments* L_{loud} *and preference evaluations* L_{pref} *are each determined as levels at the PSE.*

denoted $\Delta L_{\text{sound character}}$.

$$L_{\text{pref}} = L_{\text{loud}} + \Delta L_{\text{sound character}} \tag{1.29}$$

In this way, sound character is defined as everything that differentiates two equally loud sounds with respect to their preference. If there is no such difference, then equal loudness is equivalent to equal preference and

$$\Delta L_{\text{sound character}} = 0 \text{ dB}.$$

On the other hand, if equal preference is not equivalent with equal loudness due to preference relevant differences between the two sounds in terms of the sound character, then

$$\Delta L_{\text{sound character}} \neq 0 \text{ dB}.$$

Figure 1.17 gives a schematic plot of the two PSEs and the level difference between them. Due to the rather unpleasant sound character of the particular multi-tone sounds investigated here, the PSEs for preference are expected to be lower than the PSEs for loudness, which results in negative values for $\Delta L_{\text{sound character}}$. This means that level reductions are necessary to shift from equal loudness to equal preference for the

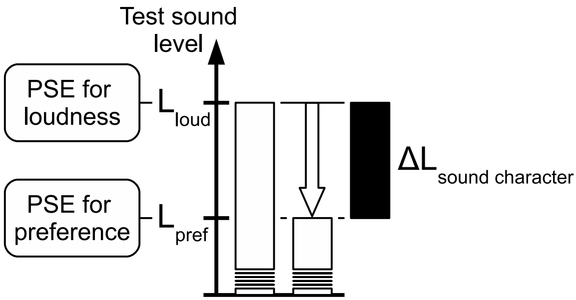


Figure 1.17: *Level scheme: The reference sound is kept constant in level while the multi-tone test sound level is varied by an adaptive procedure on the one hand converging at the PSE for loudness (loudness task, left bar) and on the other hand at the PSE for preference (preference task, middle bar). The resulting levels* L_{loud} *and* L_{pref} *represent the points at which the test sound is equally loud and equally preferred as the reference sound at* L_{ref} *(not shown here). The difference between* L_{pref} *and* L_{loud} *is attributed to the sound character and thus defined as* $\Delta L_{\mathrm{sound\ character}}$ *(right bar).*

tested multi-tone stimuli. Negative values of $\Delta L_{\mathrm{sound\ character}}$ can be interpreted as level penalties and positive value of $\Delta L_{\mathrm{sound\ character}}$ would refer to a level bonus in comparison to a condition of equal loudness with the reference sound, respectively.

Thus, the level difference between L_{pref} at the PSE for preference and L_{loud} at the PSE for loudness can be used as a quantitative distance measure for the contribution of the sound character to the preference evaluation:

$$\Delta L_{\mathrm{sound\ character}} = L_{\mathrm{pref}} - L_{\mathrm{loud}} \tag{1.30}$$

The current method differs from the above cited studies in the definition of the sound character regarding some details. In the current case, the individual level difference between the PSE for preference and the PSE for loudness is attributed to the sound character. In this understanding, sound character is defined as the entity of preference relevant sound properties that are able to describe differences between two sounds which have the same loudness. Thus, the individual, equal loudness impression is used as the base value (similar to the studies of Pollack [186], Niese [170] and Ronnebaum and Weber [196]) instead of an equal dB(A)-level (like in the studies of Cardozo and van Lieshout [43,45], Preis [188] and the idea of rating level presented in appendix A.1).

Furthermore, the current study chooses relative preference evaluations within a paired comparison over absolute annoyance judgments which were used in many previous studies. The relative preference task avoids (at least in parts) problems occurring with the definition and the absolute assessment of annoyance, which relies on an internal reference of the individual listener, especially in a laboratory context (which were presented in Sec. 1.4.2.1). The presentation of a common acoustic frame of reference, in terms of the reference noise, presented in each trial, is intended to reduce inter-individual differences, which would result from differences in terms of individual internal references, and to make it a rather easy task for the participants. The usage of a question for the preference is further supported by the close relationship between preference and pleasantness judgments [244] and unpleasantness being one major underlying aspect of noise annoyance [87,236]. Furthermore, the measurement of two attributes - loudness and preference - has the advantage, that the listeners are given the possibility to differentiate between the two aspects [261].

1.8 Conclusion and outline of the thesis

The overall aim of this PhD-thesis is a characterization and the quantification of the sound character and the appraisal resulting from temporal and spectral characteristics of a particular type of multi-tone sounds. The sounds of interest consist of two complex tones and additional combination tones and they can be regarded as generic representatives of signals produced by two sound sources with a non-linear interaction between them. Such signals are found in machinery noise of counter-rotating open rotors and also as so called multiphonics in the sounds of music instruments.

From the literature on saxophone multiphonics (reviewed in Sec. 1.2.2) it becomes clear that the frequency ratio between the two fundamentals (of the complex tones each) is a major underlying variable which influences the spacing of all higher partials of the complex tones and the combination tones resulting in changes of the timbre of the sounds. In the technical application of counter-rotating open rotors, this frequency ratio is defined by the ratio of the blade passing frequencies. However, acoustic descriptor variables, found to be relevant for differences in timbre in the musical context (like the spectral centroid and the modulation frequency), were not identified in available aircraft noise annoyance studies. It is remains rather unclear how the frequency ratio between the fundamentals changes the sound character of the multi-tone sounds and also how the appraisal of the sounds is affected by this signal parameter.

Therefore, in Chap. 2, the perceptual space resulting from a variation of the ratio between the fundamental frequencies is explored with a semantic differential. The factorial analysis of the results from the semantic differential provides insight in the underlying perceptual dimensions and their contribution to the pleasantness of the multi-tone sounds, which allows an identification of an effective descriptor variable for the effect of the frequency ratio on the judgment of the sounds.

In the context of machinery noise, the resulting multi-tone sound from the two rotor tones and the interaction tones depends on many technical factors in the sound generation mechanisms, the radiation patterns of the particular sound sources and noise control measures on the transmission path to a listening position, as was shown in section 1.2.1.1. In addition to the spectral spacing of the partials, technical parameters mainly influence the resulting sound in terms of the mixing ratio between the complex tones and the combination tones, and the overall spectral envelope of the tonal components.

Based on the literature on basic auditory sensations, introduced in Sec. 1.4, the variation of these signal parameters is expected to produce differences in terms of quantitative (like loudness) and qualitative psychoacoustic variables (like sharpness or roughness), which were both found to be related to the appraisal of the sounds (see Sec. 1.6). In detail, for the intricate multi-tone signals investigated here, it is unclear, whether the variation of a signal parameter influences the loudness only, or if the overall appraisal of the sound is affected beyond the loudness effect. The usage of different measurement methods for the unidimensional loudness and the multidimensional sound character, which are outlined in Sec. 1.5, and the common implementation of loudness equalization in measurements of the sound character with a multi-dimensional methods are another reason for the knowledge gap regarding this question. Accordingly, in Sec. 1.7, an approach for the combined measurement of the loudness and the preference as levels at the points of subjective equality (PSEs) is established, which is intended to distinguish between the influence of the loudness and the additional contribution of the sound character to the overall preference of a sound.

Consequently, in Chap. 3, the loudness and the preference of multi-tone sounds are determined in extensive listening tests as levels at the points of subjective equality compared to a fixed reference noise. The relationship between the loudness judgments and the preference evaluations, and the difference between the two, substantiates the capability of the measurement method to clearly distinguish between loudness and preference, even though both are measured in terms of levels at the PSEs. The results give an insight into the relevance and the quantitative contribution of different signal parameters to the listener's judgments and evaluations in the form of level differences on a dB-scale. The difference between the raw preference and loudness PSEs is used as a quantitative measure for the additional contribution of the sound character to the preference evaluation, which turns out to deliver even clearer relationships with the signal parameters than the raw preference data alone.

The results of Chap. 3 provide insight into the specific relationships between particular signal parameters and the loudness judgments and preference evaluations. In order to generalize the discovered connections, more general relationships between psychoacoustic predictor variables and the judgments and evaluations would be desirable, which could help to transfer the gained knowledge even further. Even though high correlation coefficients between quantitative and qualitative psychoacoustic descriptors and the evaluation of sounds have already often been identified, as shown in Sec. 1.6, the specific quantitative relationships between the two are not known for the particular multi-tone sounds investigated in this PhD-thesis.

In Chap. 4 the relationships between the judgments obtained in the PSE-listening tests from Chap. 3 and selected psychoacoustic parameters, which were introduced in Sec. 1.4, are analyzed to identify suitable descriptor variables. A descriptor variable found in the semantic differential study (Chap. 2), reflecting the ratio between the fundamental frequencies, is verified with the PSE results. Two prediction models are set up to estimate the PSE for preference and the sound character difference based on the combination of only two descriptor variables. The relationship between the descriptors of the sound character and the PSEs for preference bridge the gap between the qualitative aspect of the sound character and the quantitative aspect of a preference equivalent sound pressure level. The model, described in Sec. 4.2 provides a link between the descriptor value units and the PSEs for preference on a dB-scale.

2 Perceptual space of multi-tone sounds and its relationship to the signal's periodicity[1]

Multi-tone sounds with a biperiodic spectrum can be described as a composition of two complex tones and additional combination tones. As shown in Sec. 1.2 of the introduction, all three elements depend on the underlying fundamental frequencies of the two complex tones (defined as f_{10} and f_{01}). Changes in these fundamental frequencies entail frequency shifts of all higher partials, also those of the combination tones. This principally influences the spectral composition from low to high frequencies and the resulting timbre. With regard to beats and modulations between neighboring partials, the relative ratio between the fundamental frequencies plays a major role and it is thus chosen as experimental parameter $\rho = f_{01} : f_{10}$. For sounds with a high number of components, it is unclear how the timbre evolves as a function of this frequency ratio ρ and which sound characteristics constitute the perceptual space of the resulting sounds. In order to reduce potential adverse effects of such multi-tone sounds, especially in light of prospective technical applications, it is necessary to understand which perceptual aspects contribute to the pleasantness or unpleasantness evaluation and also which signal parameters are related to it.

Therefore, the objectives of this study are an exploration of the perceptual space for multi-tone sounds consisting of a bi-periodic spectrum, a characterization of the influence of frequency ratio ρ on the sound character, and an identification of potential sound descriptors associated with

[1]This chapter is based on, and embodies parts of:
S. Töpken *et al.* Perceptual space, pleasantness and periodicity of multi-tone sounds, J. Acoust. Soc. Am. 138, 288 (2015) (229).

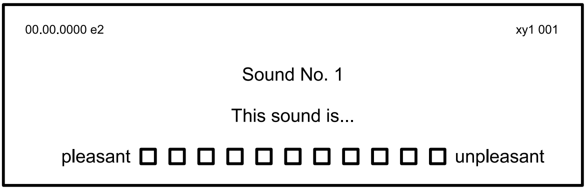


Figure 2.1: *Exemplary categorial 11-point scale used in the semantic differential - in this case for the sound no. 1 and the adjective pair* pleasant - unpleasant *(shown is the translation of the originally German version).*

the evaluative perceptual dimension. To have precise control over the frequency components, 15 synthetic sounds consisting of two harmonic complex tones and additional combination tones are evaluated with a semantic differential consisting of 16 different adjective scales. The experimental procedure is described in Sec. 2.1. The results for the perceptual space are given in Sec. 2.2 and the relationship between judgments and acoustic descriptors is presented in Sec. 2.3. Section 2.4 concludes this chapter.

2.1 Realization of the measurements

2.1.1 Procedure

To determine the perceptual dimensions, 15 sounds were judged with respect to 16 adjective pairs on individual 11-point categorial scales each (see Fig. 2.1). Seven adjective scales were unipolar, nine were bipolar. The adjectives have been composed on the basis of a pre-test including a free verbalization of appropriate adjectives(228). To reveal the relationship between sound characteristics and the appraisal of the sounds, the adjectives comprised sound-describing as well as evaluative items. The sound-describing adjectives included items directly referring to psychoacoustic sensations (e.g. *dull-sharp* corresponding to psychoacoustical sharpness) enabling a possible link to psychoacoustic descriptors. The adjective pair *harmonic-discordant* was included to facilitate a comparison with respect to consonance theory. Additional adjective pairs were derived from specific comments which were mentioned for some sounds in

Table 2.1: *English translations of the adjective scales in the chronological order of the evaluation. Two adjective scales are repeatedly measured at the end of the evaluation (No. 17 & 18). The originally used scales in German language are given in brackets below each English adjective pair.*

No.	adjective scale		
1	pleasant	-	unpleasant
	(angenehm	-	unangenehm)
2	smooth	-	rough
	(glatt	-	rau)
3	noisy	-	tonal
	(rauschhaft	-	klangartig)
4	dull	-	sharp
	(stumpf	-	scharf)
5	not fluctuating	-	fluctuating
	(nicht schwankend	-	schwankend)
6	soft	-	hard
	(weich	-	hart)
7	harmonic	-	discordant
	(wohlklingend	-	missklingend)
8	not loud	-	loud
	(nicht laut	-	laut)
9	low	-	high
	(tief	-	hoch)
10	clean	-	dirty
	(sauber	-	dreckig)
11	not functional	-	functional
	(nicht funktionstüchtig	-	funktionstüchtig)
12	not intrusive	-	intrusive
	(nicht aufdringlich	-	aufdringlich)
13	vague	-	clear
	(trüb	-	klar)
14	not dominating	-	dominating
	(nicht dominant	-	dominant)
15	not hammering	-	hammering
	(nicht hämmernd	-	hämmernd)
16	not threatening	-	threatening
	(nicht bedrohlich	-	bedrohlich)
17	pleasant	-	unpleasant (rep.)
	(angenehm	-	unangenehm (wdh.))
18	soft	-	hard (rep.)
	(weich	-	hart (wdh.))

the pre-test phase. The item pair *soft - hard* was included as descriptive pair and the adjectives referring to functionality, threat, intrusiveness and dominance served as evaluative items. Two scales (*pleasant - unpleasant* and *soft - hard*) were repeated at the end of each session to be able to check the ratings of the participants in terms of reliability.

Table 2.1 shows English translations and the originally used German adjectives in the chronological order of the assessments. The assessments were performed in such a way that all sounds were judged with respect to one assessment criterion, e.g. one adjective scale, before switching to the next criterion / adjective scale(176). The scales were arranged in 18 booklets (16 adjective scales plus the two repeated scales), with single scales on 15 pages, one page for each sound. The direction of the scales and the order of presentation of the sounds were randomized independently for each scale, but the procedure was the same for all participants. An assessment session started by handing out written instructions to all participants[2]. After having clarified open questions, an orientation phase to establish a frame of reference in the participants followed in which all 15 sounds were played back twice in random order. Then the scaling of the sounds with the semantic differential started with the first scale booklet. Within the software individual tracks for each of the 18 scales were arranged with different randomization of the 15 test sounds. Each sound was preceded by a recorded announcement with the corresponding number of the sound (e.g. "Sound 1"). This was done to ensure that the right page of the booklet is used. The sounds had a duration of 5 seconds with a pause of 5 seconds before the next announcement started to allow for making the assessment on the adjective scale.

2.1.2 Stimuli

The choice of the stimuli was aimed at the investigation of the effects of different frequency ratios on the perception of multi-tone sounds. The 15 stimuli were synthesized superpositions of two harmonic complex tones (CX1 and CX2) with fundamental frequencies f_{10} and f_{01} ($f_{10} < f_{01}$) and 29 higher harmonics of these fundamentals as well as additional combination tones (CTs) with partials f_{ij}. The combination tones are intended to simulate a non-linear interaction in the generating mechanisms of the sounds which can occur in machinery as well as musical instruments (see

[2] The written instruction can be found in Sec. A.2 in the appendix.

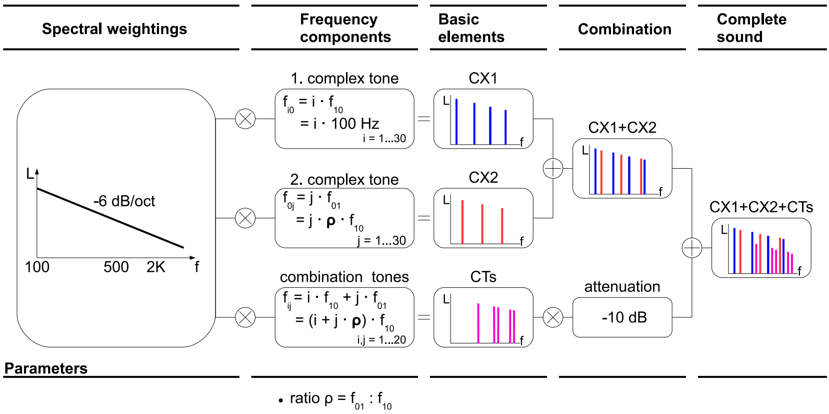


Figure 2.2: *Block diagram of the stimulus generation and definition of the experimentally varied parameters.*

Sec. 1.2 for more background information on the signals) and depend on both fundamentals f_{10} and f_{01}. In detail, the frequency components of the test stimuli were as follows:

$$\begin{aligned}
\text{1st complex tone CX1:} \quad f_{i0} &= i \cdot f_{10} && (2.1)\\
& \quad i = 1, 2, 3, \ldots, 30 \\
\text{2nd complex tone CX2:} \quad f_{0j} &= j \cdot f_{01} && (2.2)\\
& \quad j = 1, 2, 3, ..., 30 \\
\text{combination tones CTs:} \quad f_{ij} &= i \cdot f_{10} + j \cdot f_{01} && (2.3)\\
& \quad i, j = 1, 2, 3, ..., 20
\end{aligned}$$

The generation of the test stimuli is schematically shown in Fig. 2.2. Each test sound consisted of 460 partial tones. All partials had a random starting phase, taken from a uniform distribution of random numbers between 0 and 2π. The same phase values were used for the generation of all signals to avoid changes in timbre due to different phase information[185]. All partials decreased by -6 dB/octave in level towards high frequencies. The combination tones were attenuated by 10 dB compared to the two complex tones before the superposition of all partials. All stimuli had a duration of 5 seconds. The lower fundamental frequency

Table 2.2: *Signal parameters of the 15 stimuli: fundamental frequencies f_{10}, f_{01} (rounded to two decimal places), detunings Δf_{01} with respect to the corresponding exact ratio of small integer numbers and the resulting ratios of the fundamental frequencies $\rho = f_{01} : f_{10}$.*

f_{10} (Hz)	f_{01} (Hz)	Δf_{01} (Hz)	$\rho = f_{01} : f_{10}$
100	128.57	0	9:7
100	129.30	-0.7	1293:1000
100	130.00	0	13:10
100	130.70	+0.7	1307:1000
100	132.33	-1.0	397:300
100	132.66	-0.66	199:150
100	133.08	-0.25	1597:1200
100	133.33	0	4:3
100	133.58	+0.25	1603:1200
100	134.00	+0.66	67:50
100	134.33	+1.0	403:300
100	135.86	-0.5	2989:2200
100	136.36	0	15:11
100	136.86	+0.5	3011:2200
100	137.50	0	11:8

f_{10} was fixed to $f_{10} = 100$ Hz. The ratio between the fundamental frequencies $\rho = f_{01} : f_{10}$ (and therewith the higher fundamental f_{01}) was varied as a parameter. The ratio values included five ratios ρ equal to ratios of small integers ($\rho = 9:7$, $13:10$, $4:3$, $15:11$ and $11:8$) as well as ten detuned ratios, constructed by slight variations of the upper fundamental frequency Δf_{01}, resulting in ratios of large integers (e.g. $\rho = 1293:1000 \approx 13:10$ or $\rho = 134:100 = 67:50 \approx 4:3$). Table 2.2 shows the fundamental frequencies, the detuning Δf_{01} in Hz (if applicable), and the ratio ρ of all 15 stimuli. Since the lower fundamental was always 100 Hz, the sounds will be identified by their upper fundamental frequency f_{01} or the corresponding ratio of integer numbers $\rho = f_{10} : f_{01}$ in the following. All stimuli were generated digitally using Matlab, stored on a computer with a sampling rate of 22050 Hz (16 bits) and are presented at the same fixed level of 70 dB(A).

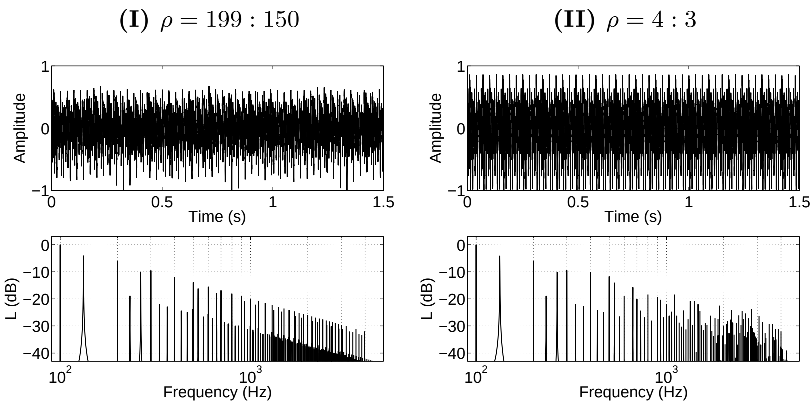


Figure 2.3: *Overview over exemplary time series (top) and amplitude spectra (bottom) for two sounds with different ratios ρ between the fundamentals. The differences between the signals with two different values for ρ are clearly visible as (ir-)regularities in the envelope of the time series and the spectrum.*

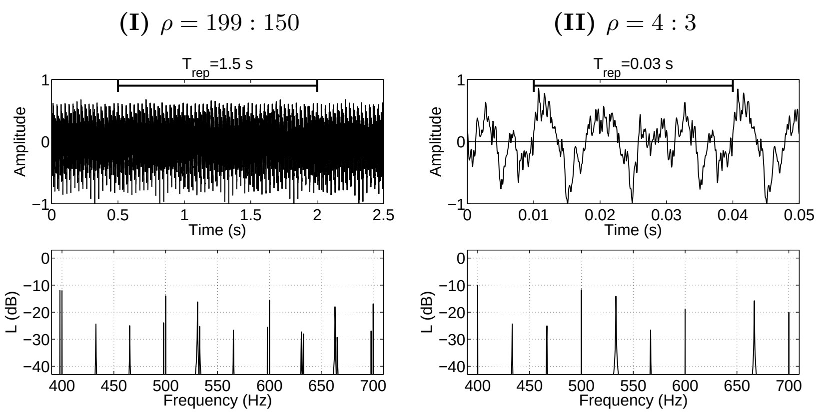


Figure 2.4: *Detail of the time series (top) and amplitude spectra (bottom) of two signals with different ratios ρ between the fundamentals. The two values for ρ yield distinct spectral spacings of the partials which lead to different periodic times.*

Figure 2.3 shows the time series (top) and the amplitude spectrum (bottom) of two exemplary stimulus signals with $\rho = 199 : 150$ and $\rho = 4 : 3$ differing in the upper fundamental frequency f_{01} by only $\Delta f_{01} = 0.66$ Hz. For $\rho = 199 : 150$ (Fig. 2.3(I)) the time series shows some modulations, while the amplitudes of the tonal components fall of rather regularly with frequency in the amplitude spectrum. The ratio of small integer numbers ($\rho = 4 : 3$) shown in Fig. 2.3(II)) yields a different picture. Here the time series shows a clear periodic structure whereas the amplitudes of the higher partials are irregular. This is due to the superposition of partials at those frequencies, which can be derived as multiples from both of the two fundamentals. Except for some missing lower partials, this leads to an equidistant spectral spacing of the tonal components (see Fig. 2.4(II)). The spectral spacing of components for $\rho = 199 : 150$ is more irregular, with closely spaced neighboring partials, shown in Fig. 2.4(I) (bottom). If ρ has a value of large integer numbers, superpositions of partials at frequencies which are derivable from both fundamentals do not occur until rather high partial numbers. The resulting beats and modulations are seen in the time series of Fig. 2.4(I) (top). Due to the rather low missing fundamental, the resulting periodic time T_{rep} of a signal is large for ratios of large integer numbers like e.g. $\rho = 199 : 150$ compared to the periodic time, resulting from ratios of small integer numbers (e.g. $\rho = 4 : 3$), which are rather short.

2.1.3 Aparatus

The experiments took place in a non-rectangular seminar room (room size: $A \approx 40\ m^2$, reverberation time: $T_{60} = 0.3\ s$) as group listening tests. The investigator and up to five participants sat around a hexagonal table. The participants were facing an active loudspeaker (Mackie, HR 824) which was connected to an external sound card (M-Audio, Fast Track Pro). The sounds were played back with an audio software (Adobe, Audition) from a computer. The listening setup was calibrated with a handheld sound level meter (B&K 2226) at the listening positions above the empty chairs. The stimulus playback level of 70 dB(A) was checked before and after each experiment.

2.1.4 Participants

Overall 37 paid volunteers (21 female, 16 male) with a median age of 21 (min=19 years, max=62 years) participated in the experiment. All participants reported no hearing problems.

2.2 Results - Perceptual space

The judgments of the 37 participants on 18 adjective scales for the 15 sounds are the basis for two factor analyses. Out of the overall 18 adjective scales, the repeated measurement of two scales allows the evaluation of the participants' ratings in terms of their reliability throughout the listening session which is presented in Sec. 2.2.1. One principal component analysis (PCA) is conducted on the adjectives (defined by the ratings of all participants for all sounds) to determine the dimensions spanning the perceptual space for this type of sounds (Sec. 2.2.2). Another PCA is conducted on the sounds (defined by the ratings of all participants on all scales) to reveal underlying factors common for sub-groups of sounds (Sec. 2.2.3). Semantic profiles of the factors found in the PCA over the sounds provide an insight how the sub-groups of sounds are judged in terms of the perceptual dimensions from the first PCA (Sec. 2.2.4). In Sec. 2.2.5 the found perceptual dimensions are compared with literature data.

2.2.1 Reliability of the judgments

The scale *pleasant - unpleasant* was initially used as the first of the 18 scales and assessed repeatedly as the second last scale. The scale *soft - hard* was initially utilized after five other scales and employed repeatedly as the very last scale. The order of the sounds and the scale direction for each sound was randomized differently for both adjective scales in the test and in the retest condition.

The detailed results of test and retest for both adjective scales are given in Fig. 2.5 for the individual data (15 sound, each judged by 37 participants). The correlation coefficients between the test and the retest condition based on the individual data for all 15 sounds are $r = 0.52$ (Spearman, $p < 0.001$) for the adjective scale *pleasant-unpleasant* and $r = 0.60$ (Spearman, $p < 0.001$) for the adjective scale *soft - hard*. An explanation for the occasionally extreme rating differences between the

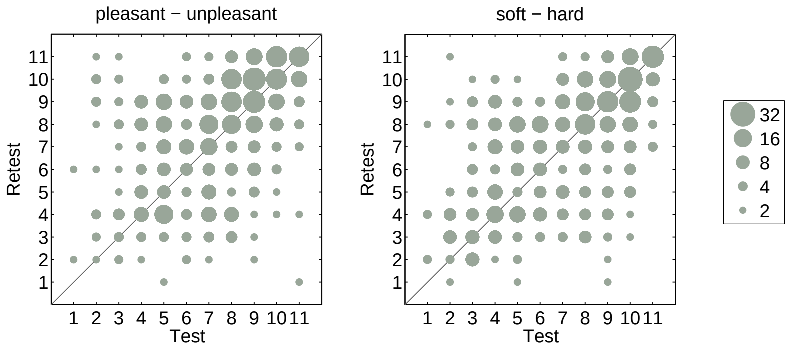


Figure 2.5: *Detailed results of the adjective pairs* pleasant - unpleasant *and* soft - hard *which were judged twice within the listening sessions for all 15 sounds by all 37 participants. Plotted are absolute numbers of occurrence (indicated by the marker size) for the individual judgments of the repetition (Retest) over the initial judgment (Test) on the scales from 1 to 11.*

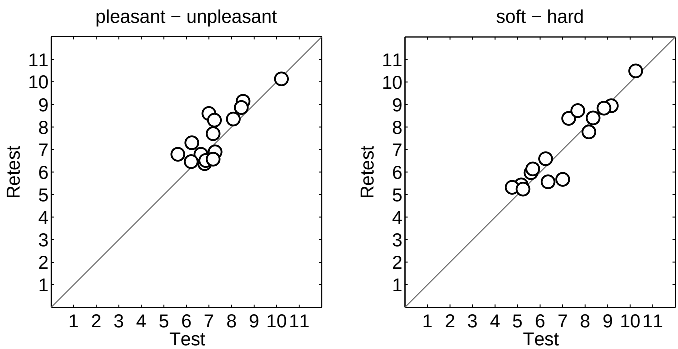


Figure 2.6: *Averaged results of the adjective pairs* pleasant - unpleasant *and* soft - hard *which were judged twice within the listening session by all 37 participants. Plotted are mean scale values of the repetition (Retest) over the initial judgment (Test) for all 15 multi-tone sounds.*

two conditions, which occurred for both adjective scales, might be the randomization of the scale direction. This random change of the verbal anchors of the scales might have been confusing for the participants in some cases and resulted in random commutation of scale values[3].

The mean judgments for the 15 different sounds, shown in Fig. 2.6, are reproduced within a range of 1.5 scale units for both adjective scales. The correlation coefficients between the test and the retest condition of $r = 0.76$ (Spearman, $p = 0.001$) in terms of the mean values for the scale *pleasant-unpleasant* and of $r = 0.92$ (Spearman, $p < 0.001$) for the mean values of the scale *soft* - *hard* further manifest the robustness of the mean judgments throughout the listening session.

2.2.2 Dimensions of the perceptual space

To explore the dimensions of the perceptual space, a PCA was conducted on the adjective scales over all 15 sounds judged by all 37 participants. As a prerequisite the Kaiser-Meyer-Olkin measure of sample adequacy ($KMO = 0.823$, 'meritorious'[(123)]) and the Bartlett test of sphericity ($\chi^2(153) = 3629.12, p \leq 0.0001$) prove the adequacy of the data set for the factor analysis. Instead of the scree criterion, which is slightly ambiguous for this data set, a factor extraction based on the Kaiser criterion is chosen. This solution delivers four, well interpretable dimensions which in combination explain 53% of the total variance. The resulting varimax rotated component matrix is shown in Tab. 2.3.

The first factor (i) shows high loadings of the adjective scales *pleasant - unpleasant*, *harmonic - discordant*, not *intrusive - intrusive*, *not loud - loud*, *soft - hard* and *dull - sharp*. It is denoted as the "pleasant" factor after the highest loading adjective pair.

The second factor (ii) can be interpreted as a "power" factor with high loadings of the adjective scales *vague - clear*, *not dominant - dominant* and *clean - dirty*.

The third factor (iii) represents the temporal structure of the sounds with high loadings of the adjective scales *not fluctuating - fluctuating*, *not functional - functional*, *not hammering - hammering* and *smooth - rough*, denoted "temp" in Tab. 2.3. The fourth factor (iv) shows high

[3]If a participant was not aware of the current scale direction, then a scale value of 11 ("unpleasant") in the test condition of one sounds might become a value of 1 ("pleasant") in the retest of this sound .

loadings of the items *not threatening - threatening*, *low - high* and *noisy - tonal*, describing the spectral content ("spec" in Tab. 2.3).

It is worth noting the connotative meaning of the adjectives found on the sound character describing third and fourth factor. With respect to the temporal structure of the sounds a *not fluctuating*, *hammering* and *rough* sound is linked to a *functional* impression. A *fluctuating*, *not hammering* and *smooth* sound evokes a connotation of being *not functional* accordingly. Regarding the spectral content the adjectives *high* and *tonal* are associated to the adjective *threatening*, whereas the items *low* and *noisy* are *not threatening*.

2.2.3 Grouping of the sounds

In order to reveal common underlying factors in the judgments of sub-groups of sounds another PCA is conducted on the 15 sounds over all 18 item pairs and all 37 participants. The measure of sample adequacy ($KMO = 0.817$, ‘meritorious’) and the Bartlett test of sphericity ($\chi^2(105) = 3313.55, p \leq 0.0001$) confirm the adequacy of the data set for the factor analysis. The Kaiser criterion as well as the scree criterion suggest an extraction of three factors with an explained total variance of 58%. Table 2.4 shows the resulting varimax rotated component matrix. The first factor shows high loadings of the sounds with ratios $\rho = 1293 : 1000$, $\rho = 2989 : 2200$, $\rho = 3011 : 2200$ and $\rho = 1307 : 1000$. These sounds will be referred to as sound group (a) in the following. The five sounds with ratios ρ based on ratios of small integers (e.g. $\rho = 4 : 3$, $\rho = 9 : 7$) load high on the second factor (sound group b). The third factor shows high loadings of the six sounds which were constructed as detunings from the ratio $\rho = 4 : 3$ (e.g. $\rho = 199 : 150$, $\rho = 67 : 50$, sound group c).

2.2.4 Mean semantic profiles of the sound sub-groups

To get an insight into the relevance of the perceptual dimensions for the differentiation of multi-tone sounds, semantic profiles of the three sounds groups (a,b,c) are shown in Fig. 2.7. Plotted are the mean values over the sounds loading to each of the three factors for the 18 adjective scales. The adjective list is ordered corresponding to the occurrence of the adjectives in the factor analysis on the perceptual dimensions. The

Table 2.3: *Results of the PCA on the 18 adjective pairs (varimax-rotated component matrix, English translations of the adjectives) showing four perceptual dimensions describing the (i) the pleasantness, (ii) the power, (iii) the temporal structure and (iv) the spectral content. Factor loadings* ≤ 0.4 *are omitted for clarity reasons.*

	adjective scales	component i	ii	iii	iv
pleasant	pleasant - unpleasant (rep.)	0.77			
pleasant	harmonic - discordant	0.73			
pleasant	not intrusive - intrusive	0.71			
pleasant	pleasant - unpleasant	0.67			
pleasant	not loud - loud	0.63			
pleasant	soft - hard (rep.)	0.61			
pleasant	soft - hard	0.61			
pleasant	dull - sharp	0.46			0.42
power	vague - clear		0.92		
power	not dominant - dominant		0.91		
power	clean - dirty		-0.55		
temp	not fluctuating - fluctuating			-0.67	
temp	not functional - functional			0.55	
temp	not hammering - hammering	0.46		0.52	
temp	smooth - rough			0.42	
spec	not threatening - threatening				0.62
spec	low - high				0.61
spec	noisy - tonal		0.46		0.49
	explained variance:	23%	13%	9%	8%

Table 2.4: *Results of the PCA on the 15 sounds (varimax-rotated component matrix) showing three groups of sounds: a, b and c. Factor loadings* ≤ 0.5 *are omitted for clarity reasons.*

	f_{01} (Hz)	$\rho = f_{01} : f_{10}$	component a	component b	component c
sound group (a)	129.30	1293:1000	0.73		
	135.86	2989:2200	0.71		
	136.86	3011:2200	0.67		
	130.70	1307:1000	0.61		
sound group (b)	133.33	4:3		0.80	
	128.57	9:7		0.74	
	137.50	11:8		0.69	
	136.36	15:11		0.64	
	130.00	13:10		0.60	
sound group (c)	132.66	199:150			0.74
	134.00	67:50			0.72
	132.33	397:300			0.71
	134.33	403:300			0.69
	133.58	1603:1200			0.63
	133.08	1597:1200			0.56
		explained variance:	20%	19%	19%

limits of the four perceptual dimensions (i) pleasant, (ii) power, (iii) temp and (iv) spec are indicated by horizontal lines.

The mean values over the sound group (b) have a considerable different profile compared to the mean profiles of the sound groups (a) and (c), particularly for the "pleasant" dimension (i) and the factor describing the temporal structure of the sounds (iii). With respect to the "pleasant" dimension (i) the sounds of group (b) are notably more unpleasant, more discordant, more intrusive and harder and regarding the "temporal" factor (iii) the sounds are less fluctuating and more hammering than the profiles of the sound groups (a) and (c). The mean profiles of the sound groups (a) and (c) are judged rather neutrally on all four perceptual dimensions. The mean profile of sound group (a) is characterized by being the most dull, the most vague and the most noisy one and the mean

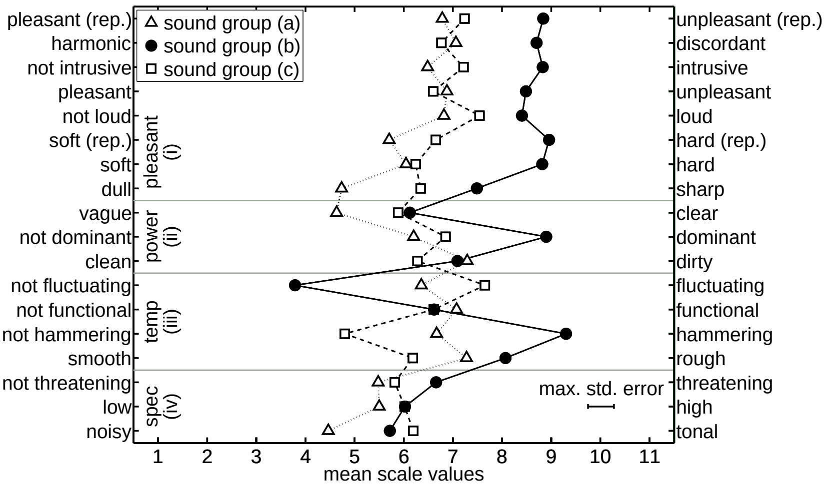


Figure 2.7: *Semantic profiles based on the mean values over the groups of sounds loading to one of the three factors (a, b and c) from the PCA over the sounds (see Tab. 2.4). Roman numerals (i, ii, iii and iv) indicate the mapping of the adjective scales to the four perceptual dimensions shown in Tab. 2.3. The open triangles mark the mean value of the sounds loading to factor (a), which are rather pleasant, soft, dull, vague and noisy. The mean value over the sounds of factor (b), indicated by filled circles are rather unpleasant, intrusive, loud, hard, sharp, dominant, not fluctuating, hammering and rough sounds. The mean values of the third group of sounds (factor c, open squares) are if not in between the two other groups rather fluctuating, not hammering and smooth.*

profile of sound group (c) is distinguished from the others in being the most fluctuating, the least hammering and smoothest. The results show the relative importance of the temporal structure on the differentiation between the three groups of sounds and indicate a relationship between the temporal structure and the "pleasant" dimension.

2.2.5 Comparison of the discovered perceptual space with other semantic differential studies and its relation to consonance theory

The PCA on the 18 adjective scales leads to four well interpretable factors: (i) pleasant, (ii) power, (iii) temporal structure, (iv) spectral content. The derived four perceptual dimensions can be well related with the three factorial timbre structure found by Namba et al.[(165)] for broadband noise mixed with frequency modulated sounds and found by Hansen and Weber[(94)] for red noise containing a single sinusoidal component. The first two factors of the current study (pleasant (i) and power(ii)) are both similarly found by Namba et al. and Hansen and Weber. The third dimension from the literature data ("metallic"), describing the sound character, is in the current case split into two factors - one factor describing the temporal structure (iii) and one factor related to the spectral content (iv).

The perceptual dimensions found in the present study also resemble the results of a study on car sounds by Sköld et al.[(206)]. The first four dimensions found in their study (quality, audibility, tonal content, modulation) can be directly related to the four factors of the current study (pleasant, power, temporal structure, spectral content), when exchanging the third and fourth factor. The fifth dimension in the study of Sköld et al. labeled safety is in the present study included in the spectral content factor by the adjective pair *not threatening - threatening*. In the study of Sköld et al. a rise of tonal prominence by an increase in tone level is associated to judgments of insecurity and threat. This is in good agreement with the positive factor loadings of the adjectives *not threatening - threatening* and *noisy - tonal* on the fourth factor (Tab. 2.3) of the current study.

The high loadings of the adjective scales *pleasant - unpleasant* and *soft - loud* on the evaluative first factor is in good agreement with the studies of Penning *et al.*[(183)] and Hohls *et al.*[(111)] where a similar fusion of *evaluation* and *power* related scales on one factor was found. However, in their studies, neither the sound pressure level nor the loudness were kept constant. This might have facilitated the close link between the scales. In the current study, the A-weighted level was kept constant over all sounds which might be an explanation of the slightly lower loadings of the two scales on the first factor (between 0.6 and 0.7) compared to those determined in the study of Penning *et al.* of about 0.9.

In the semantic differential test of the current study the adjective pair *harmonic - discordant* is highly loading on the evaluative factor (i) which is related to the pleasantness of the sounds. This is in good agreement with the study of van de Geer and coauthors[(80)]. In the current study, however, the sounds based on ratios between the fundamentals equal to ratios of small integers are judged rather discordant and unpleasant, which is in contrast to the study of van de Geer and coauthors and it also contradicts classical consonance theory[(124,239)].

The more recent study of Miskiewicz et al. shows that a superposition of two complex tones with perfect musical intervals (mathematically exact ratios of small integers) are judged higher in terms of roughness than equally tempered musical intervals which are a slight deviation from the mathematically exact ratio[(157)]. Their set of stimuli is nearest to the set used in the current study in terms of lower fundamental frequency, number of partials and fundamental frequency ratios. In their experiments, the stimuli included an exact ratio of 4:3 and corresponding small departures from integer frequency ratios for a superposition of two complex tones consisting of 10 partials each and a lower fundamental frequency of 261.6 Hz. The stimuli investigated in other studies are all based on rather high fundamental frequencies and low numbers of partials[(79,204,231)]. Due to the likely differences of in terms of overall pitch and timbre, it is difficult to compare these studies to the current one.

Combining the results from the study of Miskiewicz[(157)] with the conclusion of Terhardt[(226)] that psychoacoustic consonance, defined by the absence of roughness accounts for a sound's pleasantness, is in good agreement with the results of the current study. Here, the least pleasant sounds (sound group b) with exact ratios of small integers between the fundamentals are the most rough, most unpleasant sounds, clearly hammering and hardly fluctuating in the perceptual space (Fig. 2.7, filled circles). The other two groups of sounds (a and c) based on deviations from ratios of small integers are rather pleasant and judged more in the middle of the adjective scales *smooth - rough* and *not fluctuating - fluctuating*. This result is also in good agreement with the findings of Fletcher et al.(73,74), Björk(31), Swallowe *et al.*(223) and Fastl and Völk(71) which all found that inharmonicity in the tonal structure leads to more pleasant sounds. Examples for inharmonicities are for example sounds of gongs and bells, which are often described as pleasant or musical(223) and sounds of electric guitars(71). Furthermore, slight inharmonicities are also used in the well tempered tuning of music instruments, which uses slight mistunings to make different keys playable and transposable with the same twelve notes per octave while maintaining a perception of being in tune.

2.3 Results - Sound descriptors

Different classical psychoacoustic metrics have been calculated to test them as potential descriptors for the "pleasant" dimension. In Sec. 2.3.1 it will be shown that the classical psychoacoustic metrics calculated by a commercial software do not reflect the subjective judgments very well for this set of stimuli. Section 2.3.2 analyzes the autocorrelation function of the stimuli signals and shows that the cycle duration or its inverse, the repetition rate of the signal's time series, is a suitable way to characterize the stimuli which can be related to the "pleasant" dimension.

2.3.1 Relationship between adjective scales and classical psychoacoustic descriptors

The variation of the ratio between the fundamental frequencies ρ leads to changes in the spectral envelope and also in the temporal structure of the stimuli. It could also be shown that differences in the temporal structure are related the "pleasant" dimension. Therefore values of the psychoacoustic sharpness, fluctuation strength and roughness have been calculated for all 15 sounds to identify their potential to reflect the respective sound describing adjective scales from semantic differential[4].

Figure 2.8 shows the mean ratings of the 15 sounds with respect to the adjective pairs *dull - sharp*, *not fluctuating - fluctuating* and *smooth - rough*, each plotted over the calculated values for the psychoacoustic

[4]calculations are carried out by a commercial software, HEAD acoustics ArtemiS 11

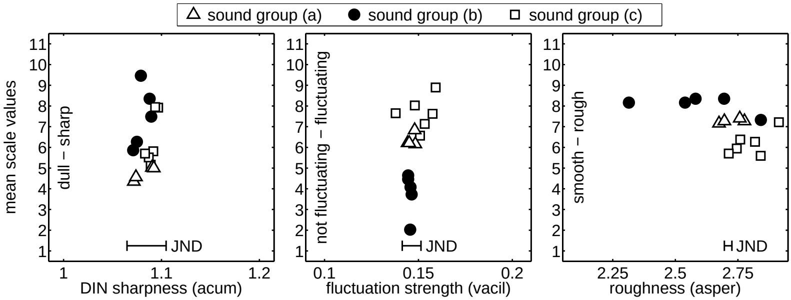


Figure 2.8: *Mean scale values of three different adjective pairs of the semantic differential experiment (N=37) plotted over the calculated values of the respective psychoacoustic metrics. Symbols show the link of each stimulus to one of the three sound groups found in the factorial analysis. Bars in the bottom of the plot indicate one JND*[182,254].

metric that is expected to be related to them. The differences in spectral envelope between the sounds yield sharpness values from 1.07 acum to 1.1 acum, calculated with the DIN 45692 standard. The range of sharpness values of only 0.03 acum is below the just-noticeable difference (JND) of sharpness which was found as $JND_{\text{sharpness}} = 0.04$ acum[182]. The calculated values of the fluctuation strength cover a range of smaller than 0.025 vacil (from 0.135 vacil to 0.16 vacil) which is just above two JNDs for fluctuation strength $JND_{\text{fluctuation strength}} = 0.01$ vacil[254]. For the roughness a range of 0.7 asper (from 2.3 asper to 3.0 asper) is covered which is clearly above the JNDs for roughness ($JND_{\text{roughness}} =$ 0.03 asper[254]). However, the sounds with high judged roughness yield low calculated roughness values, while for the rather smooth sounds high roughness values are calculated. The reason for this discrepancy is unclear. One possible explanation is that the calculated roughness values do not predict the right order of roughness for the stimuli considered here.

All three calculated classical psychoacoustic metrics do not reflect the subjective scalings obtained in the semantic differential test correctly. Thus, all three metrics fail as consistent descriptors of the spectral and also temporal sound characteristics. In particular, the fluctuation strength and the roughness are not suitable as descriptors of the temporal factor which was found to be linked to the "pleasant" dimension.

2.3.2 Relationship between the pleasant dimension and the periodicity of the signal's time series

The variation of the frequency ratio ρ between the two fundamentals directly affects the time series of the signals in terms of the periodic time and the repetition rate [5]. An analysis of the autocorrelation function ($ACF(\tau)$) reveals considerable differences in the periodicity of the stimuli used in terms of repetition rates. Figure 2.9 shows the autocorrelation function $ACF(\tau)$ and the distance $\Delta\tau$ between zero lag and the first peak of the ACF for three of the 15 stimuli used in the present study. Each one of the three sounds is a representative of one of the three sound groups (a, b and c) found in the semantic differential (Tab. 2.4). In the current case with sounds consisting entirely of multiple pure tones the repetition rate can directly be calculated as integer fraction of each fundamental frequency as shown in Eq. 2.4:

$$f_{\mathrm{rep}} = \frac{\text{fundamental frequency}}{\text{associated integer number}} \tag{2.4}$$

All higher partials of the complex tones and the combination tones then automatically fit this repetition rate. For a frequency ratio of $\rho = f_{01} : f_{10} = 4 : 3$ the fundamental frequencies $f_{10} = 100$ Hz and $f_{01} = 133.33$ Hz yield the following repetition rate f_{rep} and periodic time T_{rep}:

$$f_{\mathrm{rep}} = \frac{133.33\ \mathrm{Hz}}{4} = \frac{100\ \mathrm{Hz}}{3} \approx 33.33\ \mathrm{Hz} \tag{2.5}$$

$$T_{\mathrm{rep}} = \frac{1}{f_{\mathrm{rep}}} = \frac{1}{33.33\ \mathrm{Hz}} \approx 0.03\ \mathrm{s} \tag{2.6}$$

Table 2.5 shows the relationship between the frequency ratio ρ of the fundamental frequencies, the resulting periodic times and repetition rates f_{rep} for all fifteen stimuli in descending order of the repetition rate

[5] The repetition rate describes the overall repetition rate of the time signal which can also be found as the lowest component in the modulation spectrum of the signal. Other audible modulations at higher modulation frequencies can also occur for this type of signals.

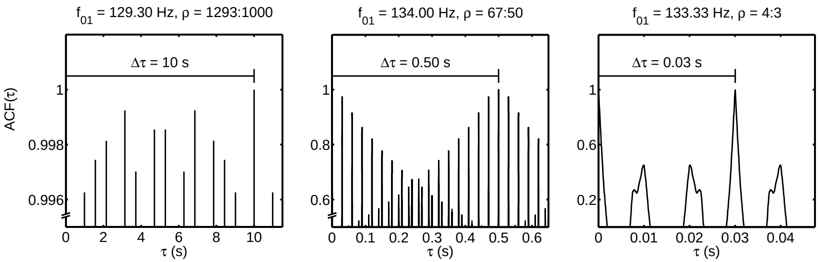


Figure 2.9: *Normalized autocorrelation function ACF as a function of the lag (τ) for three exemplary stimuli from the three sound groups (sound group a: $\rho = 1293 : 1000$, sound groupc c: $\rho = 67 : 50$ and sound group b: $\rho = 4 : 3$). Note the different scale factors (in time and ACF magnitude) necessary for the identification of the repetition rate as $\Delta\tau$ between zero lag and the first unity peak of the ACF. A ratio of the fundamental frequencies of $\rho = 1293 : 1000$ leads to a first unity peak in the ACF after a lag of $\Delta\tau = 10$ seconds corresponding to a rather low repetition rate of $f_{\text{rep}} = 0.1$ Hz (left plot). For $\rho = 4 : 3$ (right plot) the first peak occurs already after a lag of $\Delta\tau = 0.03$ seconds resulting in a rather high repetition rate of $f_{\text{rep}} = 33.33$ Hz.*

Table 2.5: *Temporal properties of the 15 stimuli: fundamental frequencies f_{10} and f_{01}, integer ratio of the fundamental frequency ρ, periodic time T_{rep} and repetition rate f_{rep}. Values of ρ equal to ratios of small integers lead to small periodic times and large repetition rates.*

f_{10} (Hz)	f_{01} (Hz)	$\rho = f_{01} : f_{10}$	T_{rep} (s)	f_{rep} (Hz)
100	133.33	4:3	0.03	33.33
100	128.57	9:7	0.07	14.29
100	137.5	11:8	0.08	12.5
100	130	13:10	0.10	10.0
100	136.36	15:11	0.11	9.09
100	134	67:50	0.5	2.0
100	132.66	199:150	1.5	0.67
100	132.33	397:300	3.0	0.33
100	134.33	403:300	3.0	0.33
100	129.3	1293:1000	10.0	0.1
100	130.7	1307:1000	10.0	0.1
100	133.08	1597:1200	12.0	0.083
100	133.58	1603:1200	12.0	0.083
100	135.86	2989:2200	22.0	0.046
100	136.86	3011:2200	22.0	0.046

(increasing order of the periodic time). Values of ρ equal to ratios of small integers lead to small periodic times T_{rep} which are equivalent to large repetition rates f_{rep}. Slight changes of the ratio ρ between the fundamental frequencies lead to rather large changes in the repetition rates of the sounds and hence in the periodic times of the stimuli. The directly calculated values for the periodic times T_{rep} in Tab. 2.5 exactly match the peak to peak distance $\Delta\tau$ found by the ACF analysis shown exemplarily for three sounds in Fig. 2.9.

The relationship between the factor values of the "pleasant" factor from the PCA and the repetition rate is shown in Fig. 2.10. The adjective scale *pleasant - unpleasant* is loading positively to the pleasant factor, meaning that low factor values are associated to pleasantness and high factor values to unpleasantness. The sounds are judged more unpleasant for higher repetition rates than for lower ones. The correlation coefficient between the pleasant factor and the (linear) repetition rate of $r = 0.84$ is significant ($p < 0.05$). Similar result were also found in a different study

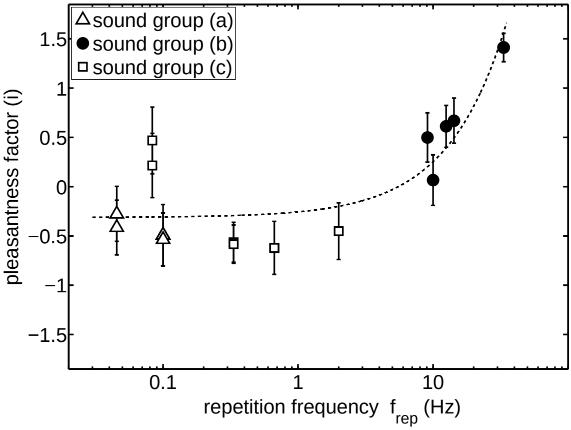


Figure 2.10: *Standardized factor values of the "pleasant" factor (i) over the repetition frequency f_{rep} (on a log scale) of the stimuli time series. Errorbars indicate the 95 percent confidence interval (1.96 standard errors). Symbols show the link of each stimulus to one of the three sound groups (a, b, c) from the factor analysis on the sounds (TAB. 2.4). A significant correlation coefficient between repetition frequency and unpleasantness judgments is found ($r = 0.84$, $p < 0.05$). The result of a linear regression is shown by the dashed line which is curved due to the logarithmic frequency axis.*

using a complete paired comparison to measure the (un-)pleasantness of the sounds (229).

2.4 Conclusion

The perceptual space of multi-tone sounds with a biperiodic spectrum was investigated in listening tests with a semantic differential. The stimuli consisted of two complex tones and additional combination tones, all based on summed up integer multiples of two fundamental frequencies f_{10} and f_{01}. The frequency ratio between the fundamental frequencies $\rho = f_{01} : f_{10}$ was varied as an experimental parameter. Three different groups of sounds could be identified in a perceptual space which is spanned by four independent dimensions, describing (i) the pleasantness, (ii) the power, (iii) the temporal structure and (iv) the spectral content of the sounds. Generally, a close link between the loudness and the pleasantness judgments was found. The adjective pairs *pleasant - unpleasant* and *not loud - loud* both loaded highly on the evaluative perceptual dimension (i) which also explained most of the variance.

The variation of the frequency ratio ρ changed the timbre and considerably affected the perceived pleasantness of the sounds. All sounds based on values of ρ equal to ratios of small integers were accumulated in one category of sounds, judged as rather unpleasant, while the other two sound categories contained sounds based on ratios of large integer numbers, which were judged more neutral on the "pleasant" dimension. The biggest timbre differences between the sound categories were found on the dimension describing the temporal structure and the adjective pair *not dominant - dominant* from the power dimension while differences on the dimension describing the spectral content turned out to be rather subtle. In terms of the pleasantness, already small shifts of the upper fundamental of less than 1 Hz, which led to ratios of large integer numbers (e.g. $\rho = 199 : 150$), were perceived considerably more pleasant than ratios of small integer numbers (e.g. $\rho = 4 : 3$). This result is counter intuitive from classical consonance theory but in good agreement with newer literature data. It turns out that classical psychoacoustic descriptors (like sharpness, fluctuation strength and roughness) are not able to describe the perceptual differences found in the assessment of the scaled sounds.

An effective way to describe the signature of the multi-tone sounds resulting from a variation of the frequency ratio ρ, is the characterization by the repetition rate of the time signal, which is the inverse of the peak to peak distance in the autocorrelation function of the signal's time series. For the multi-tone sounds used in the present study, the repetition rate can directly be derived from the integer ratio between the fundamentals and the absolute frequency values. High values of the repetition rate (equivalent to short periodic times) are assessed considerably more unpleasant than low values of the repetition rate (equivalent to long periodic times). The results indicate that the application of suitable slight frequency shifts may help making sounds with a biperiodic structure more pleasant.

3 Quantification of sound character differences by a measurement of loudness and preference

In the previous Chapter 2, the perceptual space of multi-tone sounds was explored using a semantic differential. The factorial analysis of the results from the semantic differential provided a general insight into the underlying dimensions of the perceptual space for multi-tone sounds consisting of two complex tones and additional combination tones, resulting from a variation of the frequency ratio ρ between the two fundamentals. The adjective pairs *not loud - loud* and *pleasant - unpleasant* both loaded highly onto the first factor, which explained the largest part of the variance. In addition, a link between this "pleasant"-factor and a factor describing the temporal structure of the multi-tone sounds could be revealed.

Thus, with respect to the overall appraisal of the sounds, besides the loudness also the sound character is expected to play a role, which is in line with findings of the literature reviewed Sec. 1.6. The specific, quantitative contribution, of the loudness on the one hand and the sound character on the other hand, to the appraisal of the sounds is not disclosed by the results of the semantic differential, presented in Chapter 2. Even though the intervals of the categorical rating scales are designed to reflect equal distances, the subjective meaning of the particular scale units for the different adjective pairs is defined individually by each participant. It is therefore neither possible to compare the different scales units one-on-one, nor is it possible to derive conversion ratios between the units, which would be necessary for a quantitative assessment.

The aim of the study presented in this chapter is to get a more detailed insight into the the contribution of the loudness and the sound characteristics to an overall preference decision in the specific application of aircraft cabin interior noise. The evaluative perceptual dimension, which was found in the semantic differential test, is investigated in more detail, for a wider range of signal parameters, whose quantitative effects on the appraisal of the sounds is unclear so far. The signal parameters are chosen to be conceivably linked to the perceptual dimensions, identified with the semantic differential (Sec. 2.2), while also being related to technically relevant design parameters (presented in Sec. 1.2.1).

In the following, a measurement method is used, which determines levels at the points of subjective equality (PSEs) for loudness and for preference each compared to the same fixed reference sound. This measurement method, established in Sec. 1.7 of the introduction, is based on the assumption that the loudness and the unpleasantness increases with rising sound pressure level. The influence of signal parameters on the denotative loudness judgments and on the connotative preference evaluations, are quantified on the same dB-scale. This combined measurement of the loudness and the preference allows for a comparison and a distinction between the two in terms of dB-values. The level difference between the PSEs for preference and for loudness is attributed to the particular sound character, which in return becomes quantifiable as a single value on a dB-scale.

The multi-tone signals investigated here have the same spectral frequency structure as those which were used in the semantic differential, but further signal parameters affecting the composition of the sound elements and the spectral envelope of the sounds were varied. The realization of the measurements is described in Sec. 3.1. The results of the listening tests are divided into three sections. Section 3.2 gives a general insight into the relationship and the difference between the loudness judgments and the preference evaluations. The relationships substantiate the idea of the measurement approach that the PSEs for preference and especially the level difference attributed to the sound character reflect an evaluative aspect that is not covered by the loudness judgments alone. In Sec. 3.3 the results of repeated measurements, within participants and between disjunct groups of participants, are presented to verify the reliability and reproducibility of the results obtained from the PSE measurements. The influence of the experimentally varied signal parameters on the loudness judgments and preference evaluations is presented in Sec. 3.4 in detail. The conclusion 3.5 closes this chapter.

3.1 Realization of the measurements

3.1.1 General Procedure

In two separate matching experiments the points of subjective equality (PSEs) for preference and for loudness were measured with an adaptive level varying technique for multi-tone signals. Since the PSE is identical to the 50%-point (P50) of the psychometric curve, an adaptive staircase method (2 interval, 2 AFC) with a simple 1-up, 1-down rule was used [(140,152)]. In both tasks (preference and loudness) the presentation order of the reference and the multi-tone test sound was randomized within each trial. The level of the reference sound was always kept constant and only the level of the test sounds was varied [(76)]. Pilot experiments with similar signals have shown that level reductions are necessary for the multi-tone test sounds to achieve equal loudness or equal preference compared to the reference sound. Thus, by adjusting the multi-tone test sound level with the adaptive paradigm, the participants were secured from reaching harmful levels resulting from the adaptive level variation. In addition to that, the experimental setup was configured such that sound pressure levels higher than 81 dB(A) led to an abort of the experiment. About the half of the participant carried out the preference task first followed by the loudness task in each measurement session. The other half of the participants always did the two experiment in the opposite order. Each experiment started by handing out written instructions to the participant[1].

Preference task
In a paired comparison of the multi-tone sounds and the aircraft cabin reference noise the participants were asked "Which sound do you prefer as cabin interior noise?" Depending on the answer of the participants, the level of the multi-tone sound was varied adaptively. It was reduced, every time the test sound was not preferred and raised, if the test sound was preferred. At the beginning of each adaptive track the multi-tone test sound had the same A-weighted level as the reference

[1] The written instructions can be found in Sec. A.4 in the appendix.

sound: $L_{\mathrm{ref}} = 74$ dB(A). The level of the multi-tone test stimulus was initially varied with a step size of 6 dB which was halved after each upper reversal of the level curve until the minimum step size of 1.5 dB had been reached, where the measurement phase was started. The duration of the stimuli was 5 seconds each, with a separation of the stimuli by a pause of 1 second. A pause of 2 seconds was introduced between the answer of the participant and the presentation of the next stimulus pair.

Loudness task

The measurement of the loudness was carried out with a similar paradigm, but using the question "Which sound is louder?" for the loudness matching procedure. The step size of the adaptive procedure was in this case 3 dB at the beginning, which was halved after the second upper reversal of the level curve down to a step size of 1.5 dB (measurement phase). For the loudness matches the stimulus duration was 1.5 seconds each, with a pause of 0.5 seconds between the sounds of a pair. Here, a pause of 1 second after the answer of the participant separated the judgment pairs.

The individual PSEs are described as levels L_{pref} and L_{loud}, each calculated as mean value over 6 level reversal-points of the adaptive track during the measurement phase (with a step size of 1.5 dB). The level difference between the PSE for preference and the PSE for loudness was calculated individually, for each participant and each stimulus, and it is denoted:

$$\Delta L_{\mathrm{sound\ character}} = L_{\mathrm{pref}} - L_{\mathrm{loud}}$$

3.1.2 General properties of the multi-tone stimuli

In this study, similar multi-tone sounds as in the semantic differential test were used as test sounds (see also Sec. 2.1.2). The synthesis of the stimuli, consisting of two complex tones (CX1 and CX2) and additional combination tones (CTs) is schematically shown in Fig. 3.1. The test signals included representative sounds from each of the three sound groups identified with the semantic differential. They also included the most pleasant and the most unpleasant sounds for the variation of the frequency ratio ρ between the fundamentals from the semantic differential experiment. The main differences compared to the signals used in the semantic differential are:

— The partials were shaped by a triangular spectral envelope for which the parameters of the rising ($\mathrm{slope}_{\mathrm{UP}}$) and the falling slope ($\mathrm{slope}_{\mathrm{DN}}$) and in addition also the peak frequency (f_{peak}) were varied.

— The attenuation of three frequency ranges (low/mid/high) by 10 dB and the attenuation δ of the combination tones compared to the complex tones were varied as experimental parameters.

— Selected frequency components were combined to basic elements consisting of each of the two complex tones alone (CX1 and CX2) and the combination tones alone (CTs). A combination of the two complex tones without the combination tones (CX1+CX2) was also used as stimulus.

— The level of the multi-tone sounds was not constant, as in the semantic differential experiment of Chapter 2. Instead, the level needed to be varied in the adaptive procedure to measure the points of subjective equality.

In general, the stimuli were superpositions of two harmonic complex tones (CX1 and CX2) with fundamental frequencies $f_{10} = 100$ Hz and $f_{01} = \rho \cdot f_{10}$ ($f_{10} < f_{01}$) and 29 higher harmonics of these fundamentals each, as well as additional combination tones CTs with f_{ij}. In detail, the frequency components of the test stimuli were as follows:

1. complex tone CX1: $$f_{i0} = i \cdot f_{10}, \quad i = 1, 2, 3, \ldots, 30 \tag{3.1}$$

2. complex tone CX2: $$f_{0j} = j \cdot f_{01}, \quad j = 1, 2, 3, ..., 30 \tag{3.2}$$

combination tones CTs: $$f_{ij} = i \cdot f_{10} + j \cdot f_{01}, \quad i, j = 1, 2, 3, ..., 20 \tag{3.3}$$

The partials of each element (CX1, CX2 and CTs) were multiplied by spectral weightings which are defined by a peak frequency f_{peak}, a rising lower slope $\mathrm{slope}_{\mathrm{UP}}$, up to the peak frequency, and a falling upper slope $\mathrm{slope}_{\mathrm{DN}}$, above the peak frequency. Due to a lowest frequency component of 100 Hz, a peak frequency of $f_{\mathrm{peak}} = 100$ Hz led to partials decreasing

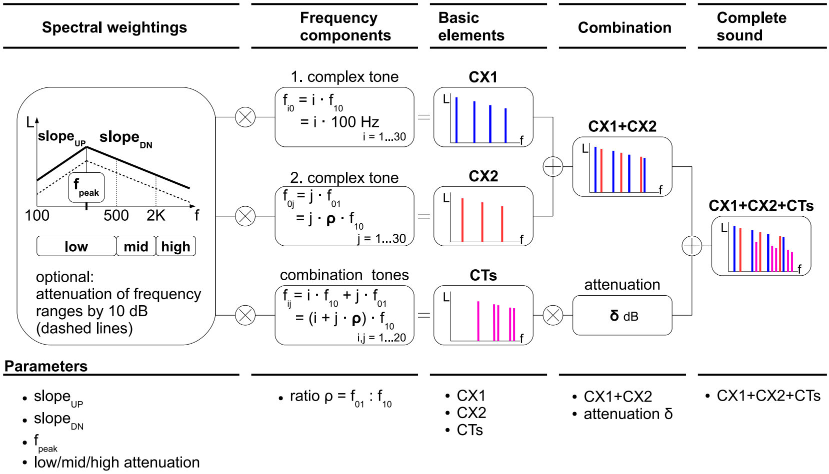


Figure 3.1: *Block diagram of the stimulus generation and definition of the stimulus parameters.*

in level with rising frequency only. For peak frequencies higher than 100 Hz, the partials initially rose in level with the slope slope_{UP} up to the peak frequency f_{peak} and then decreased in level by the value of slope_{DN} for frequencies above f_{peak}. The mixing ratio between the superposition of the two harmonic complex tones (CX1+CX2, 60 partials) and the combination tones (CTs, 400 partials) in a superposition of all three components (CX1+CX2+CTs, 460 partials) was specified by the attenuation factor δ. A value of $\delta = 0$ dB means that the levels of the partials from the complex tones (CX1+CX2) and from the combination tones (CTs) are derived from the same spectral envelope. A value of $\delta = -10$ dB means that the overall level of the combination tones (CTs) is reduced by 10 dB, which would correspond to a downward shift of the spectral envelope for the CTs (schematically shown in Fig. 3.1 as a dotted line in the spectral weightings). Besides the baseline configuration (CX1+CX2+CTs) also the complex tones alone(CX1 and CX2), the combination tones alone (CTs) and the blend of the two complex tones (CX1+CX2) were investigated as test stimuli. These sound elements can be regarded as extreme values of the parameter $\delta = +\infty$ dB (CTs only) and $\delta = -\infty$ dB (CX1+CX2).

A fractional factorial design with overall 55 test sounds was used. An overview of the signal parameter combinations of the 55 multi-tone stimuli can be found in Tab. A.1 in the appendix. The signals are identified by single letters (e.g. A) or combinations of a letter and a number (e.g. D2). The baseline conditions of the test sounds are identified by letters (A to E) which are reflecting specific values of the frequency ratio ρ between the fundamentals. Based on the finding that the frequency ratio ρ between the fundamentals of the two complex tones ($\rho = f_{01} : f_{10}$) considerably changes the sound character and also influences the judged pleasantness (found in the semantic differential test, Chap. 2), all parameter variations other than the frequency ratio ρ were derived from two baseline sounds, with $\rho = 199 : 150$ (sound D) and $\rho = 4 : 3$ (sound E), which were the most pleasant and the most unpleasant stimuli in the semantic differential experiment. Sub-elements, derivatives or repeated measurements of the two baseline conditions (D and E) are identified by additional numbers (e.g. E2). The two complex tones alone (CX1 and CX2), which were also used as separate signal conditions, are each based on only one fundamental frequency, identified by the letters F and G.

The reference sound was an aircraft cabin interior signal with a spectral slope of approximately -6 dB per octave up to 1 kHz and -12 dB per octave above 1 kHz and a fixed sound pressure level of 74 dB(A). The reference sound was exactly the same in the loudness and in the preference task. All stimuli were generated and stored in a computer with a sampling rate of 22050 Hz and a resolution of 16 bits. All stimuli have a duration of 5 seconds in the preference task and 1.5 seconds in the loudness task.

3.1.3 Aparatus

The experiments were conducted in the anechoic chamber of the University of Oldenburg (lower limiting frequency: 50 Hz). The task itself was implemented as a Matlab (The Mathworks) program on a computer [(69)]. An external audio interface (M-Audio, Fast Track Pro) supplied the audio signals to an active loudspeaker (Mackie, HR 824), positioned in front of the participant, seated inside the anechoic chamber. The experimental routine was operated by the participant via a computer keyboard, used as an answerbox (Fig. 3.2), and the TFT-screen, placed underneath the

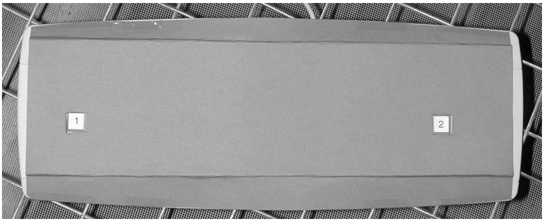


Figure 3.2: *Covered computer keyboard used as answerbox by the participants during the adaptive paired comparison experments.*

loudspeaker. The experimental setup in the anechoic chamber is shown in Fig. 3.3 as a schematic view and as a photo in Fig. 3.4. The setup was calibrated with a handheld sound level meter (B&K 2226) at the listening position above the empty chair. The calibration was checked before the first and after the last experiment each day, throughout each of the three measurement stages.

3.1.4 Measurement stages

The measurements of the different signal parameter variations were carried out in three measurement stages with overall 103 individual participants. However, some of the participants took part in multiples measurement stages. Six distinct groups of participants (P1-P6) are bundled together to allow for a segregation between the participation in the three measurement stages. Figure 3.5 shows an overview of the three measurement stages, the signal parameters varied in each stage and the participant structure. The panel labeled "Parameter Variations" outlines the various parameters that were varied in the three stages of the study (expanded along the horizontal direction). Each of the resulting stimuli were evaluated once by each participant in terms of loudness and preference.

Throughout the three measurement stages, repeated measurements were carried for two values of the frequency ratio ρ with different participant groups. These two values of the frequency ratio ρ were the most pleasant and most unpleasant conditions identified in the semantic differential experiment, presented in Chap. 2. The test sounds D, D5, D7 and D14 ($\rho = 199 : 150$) are identical in terms of the underlying signal

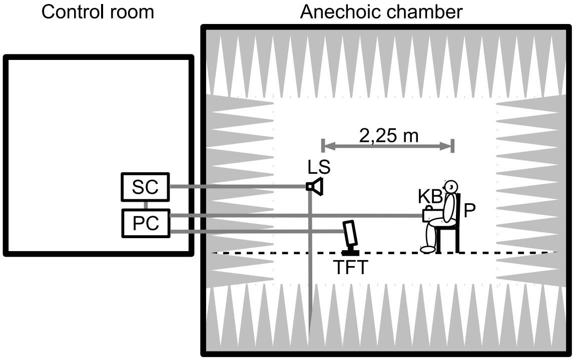


Figure 3.3: *Schematic view on the experimental setup in the anechoic chamber:* **SC** *soundcard,* **PC** *computer,* **LS** *loudspeaker,* **TFT** *flatscreen,* **KB** *keyboard (answerbox),* **P** *participant.*

Figure 3.4: *Experimental setup in the anechoic chamber*

but are measured in different measurement stages by different participant groups. The same applies to the sounds and E, E5, E7 and E14 ($\rho = 4 : 3$).

In order to assess the intra-individual reliability, a total of 47 participants, consisting of the participant groups P1, P2, and P3, measured two sounds denoted as conditions D and E repeatedly (the repeats are denoted as conditions D5 and E5), but separated by about 2 weeks. This assessment will be presented in Sec. 3.3.1 of the results. The same signals (D and E) were also included in stage 2 (as condition D7 and E7) and stage 3 (as condition D14 and E14). In these latter two stages, two additional groups of participants P5 and P6, which were disjunct from the group P3, also judged the two sounds. This repeated measurement by three disjunct groups of participants enables an analysis of the stability of the three group means for the two sounds, which is presented in Sec. 3.3.2.

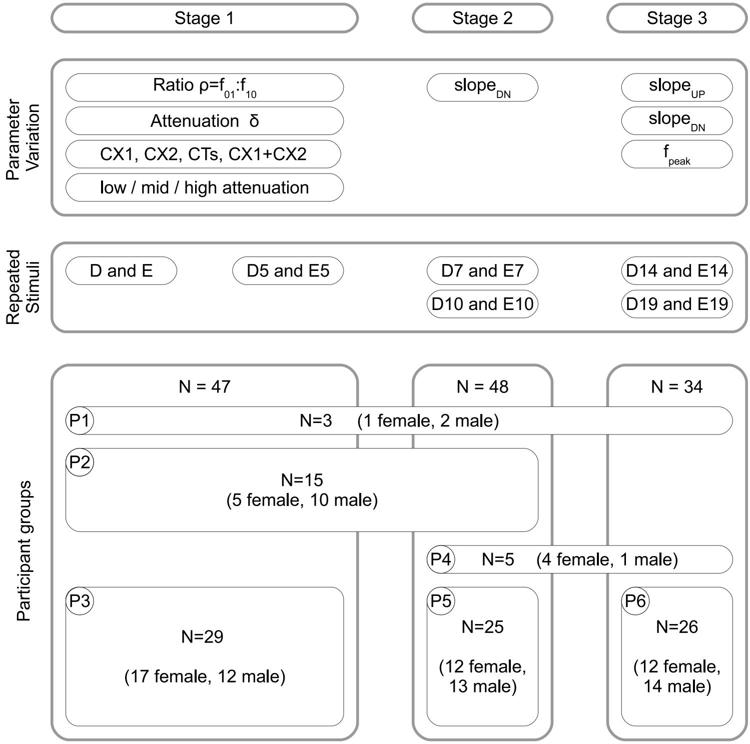


Figure 3.5: *Overview of the three measurement stages: Different signal parameters were varied in the three stages as depicted in the panel labeled "Parameter Variation". The sounds representing the two baseline conditions D and E were repeatedly measured by the same participants in stage 1 (as D5 and E5) and also included into the other two measurement stages (as D7 and E7 in stage 2 and D14 and E14 in stage 3) which allows for an analysis of the reliability of the measurements (presented in Sec. 3.3). Two further sounds of stage 2 (D10 and E10) were also included in stage 3 (D19 and E19). The lower part of the figure gives an overview of the participant structure and their affiliation to six participant groups P1 to P6 over the three measurement stages.*

3.2 Results - Relationships between the loudness judgments, the preference evaluations and the level difference attributed to the sound character

The results of the listening tests provide three measures for each test stimulus and each participant:

1. The measured point of subjective equality for loudness: L_{loud}
2. The measured point of subjective equality for preference: L_{pref}
3. The individually calculated level difference between the two PSEs which was attributed to the sound character in the definition of the measurement approach: $\Delta L_{\text{sound character}} = L_{\text{pref}} - L_{\text{loud}}$

In this section, the relationships between the loudness judgments and the preference evaluations, each determined as level at the point of subjective equality (PSE), and the level difference $\Delta L_{\text{sound character}}$ attributed to the sound character are presented. The relationships between these three measures reveal the meaning of $\Delta L_{\text{sound character}}$ and verify that L_{pref} and $\Delta L_{\text{sound character}}$ do actually reflect evaluative aspects which are not covered by the loudness judgment.

The relationship between the preference and the loudness judgments is shown in a scatter plot of the individual data as well as the mean data of all 55 sounds investigated in Fig. 3.6. When looking at the mean values, the dynamic range of the PSEs for loudness is about 12 dB over all signal parameters variations. The variability in the PSEs for preference is larger, covering a range of about 30 dB. The different level ranges covered by the results from the two tasks suggest that the participants judged differently in the particular tasks. Nevertheless, a statistically significant correlation coefficient of $r_{ind} = 0.59$ ($p < 0.001$) is found between L_{loud} and L_{pref} for the individual data and $r_{mean} = 0.86$ ($p < 0.001$) for the mean values. This in good agreement with high correlation coefficients between loudness and unpleasantness or loudness and annoyance judgments found in the literature, which were reviewed in Sec. 1.6. It is also in accordance with the results from the semantic differential test (Chap. 2), where the adjectives *not loud - loud* and *pleasant - unpleasant* both loaded highly on the factor explaining most of the variance. The assessment of only one class of sounds (in this case

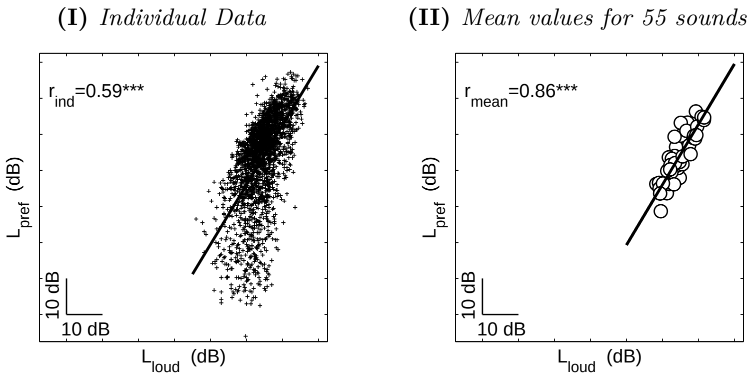


Figure 3.6: *Relationship between preference and loudness judgments: Individual data (crosses, left figure) and mean values (circles, right figure) of L_{pref} plotted over L_{loud} for all 55 sounds. The results of the two tasks span different level ranges. Considerably lower levels are needed to achieve equal preference than to achieve equal loudness.*

multi-tone sounds)(128), and the assessment in laboratory experiments including a level variation of the sounds under test might have further facilitated this close relationship(89).

Included in Fig. 3.6 are also lines from linear regressions. The slope of the linear regression between L_{pref} and L_{loud} is 1.65 for the individual values (Fig.3.6(I)) and 1.67 for the mean values (Fig.3.6(II)). The effect of the signal parameter variations influences the PSEs for preference stronger than the PSEs for loudness. This qualitative finding is in good agreement with the relationship between PSEs for annoyance and loudness found by Hiramatsu *et al.*(109). However, the slope found in the current study is about half the slope found by Hiramatsu *et al.* [2]. Possible explanations for these differences are the different types of tasks and contexts of the two studies (preference in the context of aircraft cabin noise vs. annoyance of everyday sounds), the indirect determination of the PSEs in the study of Hiramatsu *et al.* and the considerable spread in their data.

[2]The data from the Hiramatsu *et al.* study is shown in Fig. 1.15 in Sec. 1.6

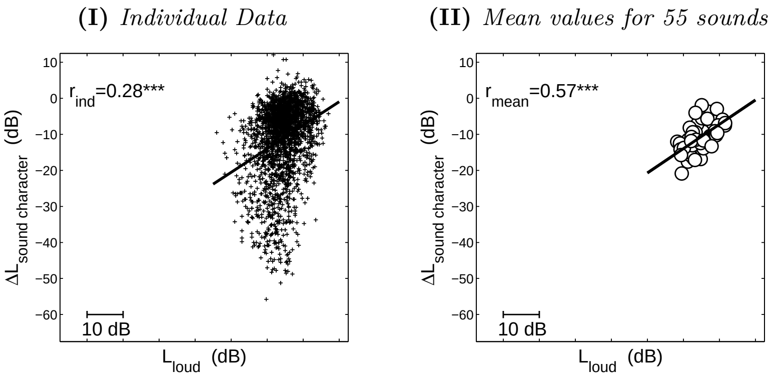


Figure 3.7: $\Delta L_{\text{sound character}}$ *plotted over* L_{loud}. *Individual data (crosses, left figure) and mean values (circles, right figure) for all 55 sounds. A rather low correlation coefficient* $r_{\text{ind}} = 0.28$ *is found between the individual data for* $\Delta L_{\text{sound character}}$ *and* ΔL_{loud}*, which indicates a rather weak relationship between the loudness and sound character assessments.*

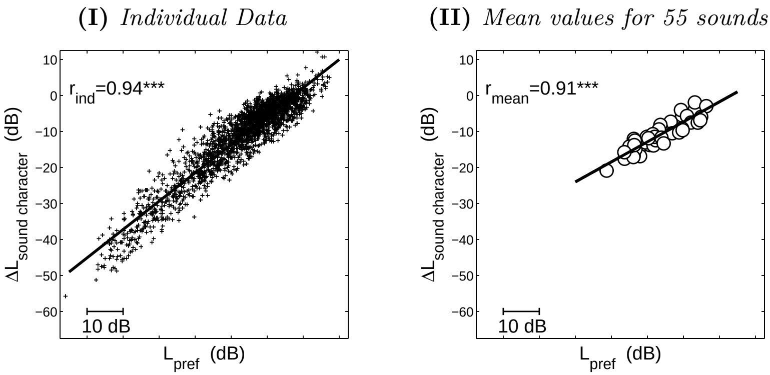


Figure 3.8: $\Delta L_{\text{sound character}}$ *plotted over* L_{pref}. *Individual data (crosses, left figure) and mean values (circles, right figure) for all 55 sounds. The correlation coefficient between* L_{pref} *and* $\Delta L_{\text{sound character}}$ *is very high (*$r_{\text{ind}} = 0.94$ *and* $r_{\text{mean}} = 0.91$*), showing that the level difference* $\Delta L_{\text{sound character}}$ *is closely related to the preference evaluation.*

The relationship between the level difference attributed to the sound character $\Delta L_{\text{sound character}}$ and the loudness judgments L_{loud} is displayed in Fig. 3.7. The mean values of the level difference $\Delta L_{\text{sound character}}$ range from nearly 0 dB to -20 dB. A value of $\Delta L_{\text{sound character}} = 0$ dB means that equal loudness is equivalent to making the sound also equally preferred and no prominent sound character differences between test and references sound were noticeable. In the case of $\Delta L_{\text{sound character}} = -20$ dB a test sound is considerably different compared to the reference sound while already being equally loud. An additional level reduction of the test sound by 20 dB is necessary to compensate the apparent difference in sound character to make both sounds equally preferred.

A rather low correlation coefficient of $r_{ind} = 0.28$ ($p < 0.001$) between the individual data of $\Delta L_{\text{sound character}}$ and L_{loud} is found. For the mean values of the 55 sounds, the correlation coefficient is $r_{\text{mean}} = 0.57$ ($p < 0.001$). Even though the correlation coefficients are statistically significant, the resulting shared variance of only about 8% (individual data) and 32% (mean values) stresses the rather weak relationship between the sound character difference measure and the PSEs for loudness. Instead, Fig. 3.8 shows the close relationship between $\Delta L_{\text{sound character}}$ and the PSEs for preference. This relationship between the level difference defined as $\Delta L_{\text{sound character}}$ and the preference evaluations is supported by correlation coefficients of $r_{\text{ind}} = 0.94$ and $r_{\text{mean}} = 0.91$, which are both highly significant ($p < 0.001$).

Figure 3.9 summarizes the relationships between the three measures and gives an overview of the shared variances between them. The PSEs for preference and for loudness are significantly correlated with each other, but the different level ranges covered by the results from the two tasks indicate that the preference evaluations were clearly distinguished by the participants from the loudness judgments. Even though the loudness of the sounds is varied throughout the adaptive level-varying procedure of the preference measurement, the data suggests that the participants were able to state distinct preference evaluations.

The dual measurement of two different aspects — loudness and preference — and the calculation of the relative measure $\Delta L_{\text{sound character}}$ allows to distinguish between the contributions of loudness and other factors, which are subsumed under the term *sound character*, to the preference evaluation. Due to the measurement of preference and loudness as level dependent PSEs against one common reference, it is also possible to

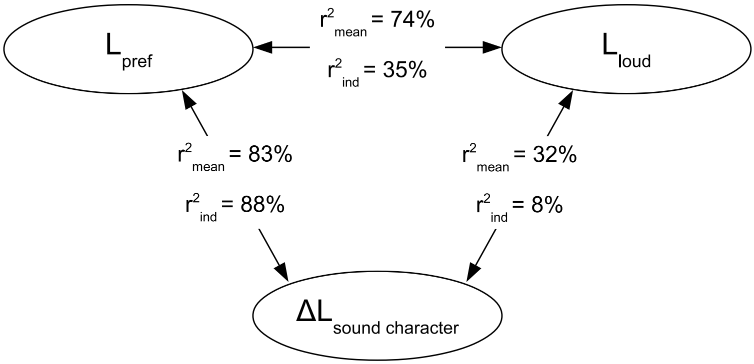


Figure 3.9: *Overview of the shared variances between the PSEs for loudness L_{loud}, for preference L_{pref} and the difference measure $\Delta L_{\mathrm{sound\ character}}$ based on the individual and the mean data.*

quantify the difference between the two measures as a level difference and to express how much more unpleasant one sound is compared to another equally loud sound in terms of dB-values. The individually determined level difference between the PSE for preference and the PSE for loudness which has been defined in the measurement approach as $\Delta L_{\mathrm{sound\ character}}$ shares a rather small amount of variance with the loudness judgments and a considerably higher amount of variance with the preference evaluations. This means that the level differences $\Delta L_{\mathrm{sound\ character}}$ are only weakly related to the loudness judgments and may be reasonably attributed the additional influence of differences in the sound character, which are relevant in terms of preference.

3.3 Results - Reliability of the judgments

In this section, the results of repeated measurements are presented to verify the reliability of the loudness judgments and preference evaluations gained in the listening tests over three larger measurement stages. Details on the participant groups and measurement stages can be found in Fig. 3.5 in Sec. 3.1.4 above. In the following, the two repeatedly measured sounds will be always be identified as D and E and reference to the (re-)test condition and participant group will be given instead of the numerical identifiers for the sounds.

The intra-individual reproducibility of the results by 47 participants are presented in Sec. 3.3.1. The inter-group stability of the measurements from three disjunct participant groups are given in Sec. 3.3.2.

3.3.1 Intra-individual reproducability

3.3.1.1 Individual results of the paired repeated measurements

The individual results of the same 47 participants (consisting of participant groups P1, P2 and P3) for sound D ($\rho = 199 : 150$) are shown in Fig. 3.10 and for sound E ($\rho = 4 : 3$) in Fig. 3.11. The individual data of the repeated measurement (retest) is plotted over the initial evaluation (test) for the preference (left) and the loudness task (right) in both figures.

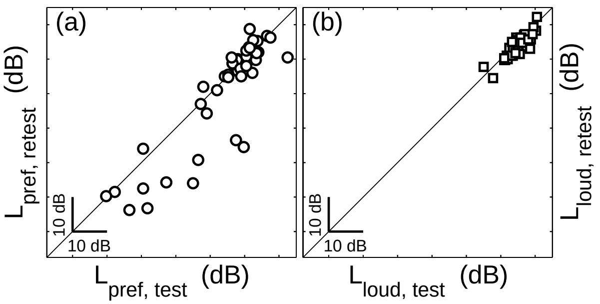


Figure 3.10: *Individual results of the preference (a, left) and the loudness task (b, right) for the sound D ($\rho = 199 : 150$). Shown are the results of the repeated measurement (retest) over the initial judgments (test) from the same participant sample ($N = 47$). The bisecting line indicates identity. A considerable inter-individual variability can be seen for the preference evaluations and especially big attenuation values are individually not well reproduced. The loudness is judged more uniformly over the 47 participants and the judgments are also individually better reproduced.*

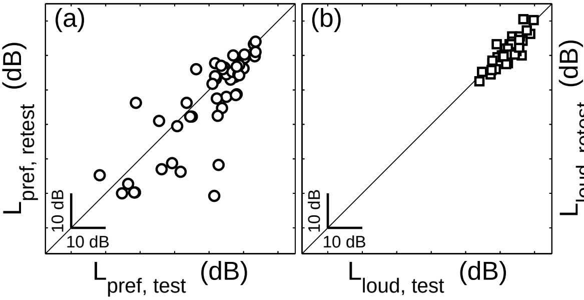


Figure 3.11: *Same as Fig. 3.10 but for sound E ($\rho = 4:3$). The inter-individual spread of the preference evaluations and the more uniform judgment of the loudness is similar like for sound D, shown in Fig. 3.10.*

For both sounds the results of the preference task (Fig. 3.10a and 3.11a) show considerably more inter-individual spread (in the sense of overall dynamic range) and also more intra-individual variability (in terms of deviation from equality line) than the results of the loudness task (Fig. 3.10b and 3.11b). The loudness is judged more reliably by each participant and more uniformly by the 47 participants than the preference. This finding and also the extent of the inter-individual variability is in agreement with literature data using a similar experimental procedure[(233)].

The correlation coefficients between the individual test and retest results of the preference and the loudness judgments are given in Tab. 3.1 for the two sounds. Statistically significant correlation coefficients are found between the test and retest conditions for both sounds based on the individual results of 47 participants.

3.3.1.2 Mean values of the paired repeated measurements

The mean values of 47 participants for the repeated measurements of the preference and the loudness task are shown in Fig. 3.12. The error bars, indicating the standard error of the calculated mean values, are

Table 3.1: *Correlation coefficients r between the individual results of test and retest judgments for the sound D and E with N=47 each.*

sound	$\rho = f_{01} : f_{10}$	measure	Correlation coefficient
D	199 : 150	L_{loud}	$r = 0.86$, $p < .001$
		L_{pref}	$r = 0.89$, $p < .001$
E	4 : 3	L_{loud}	$r = 0.85$, $p < .001$
		L_{pref}	$r = 0.81$, $p < .001$

overlapping for test and retest condition, each. The results of a statistical test (T-Test, rep. meas.) are given in Tab. 3.2.

For the loudness judgments differences of about 1 dB between the mean values of the test and the retest are already statistically significant for both test sounds. These differences being slightly above the just noticeable differences in level ($JNDL \approx 1$ dB[(263)]) suggests that they occur on a statistical basis due to the high number of participants ($N = 47$) rather than being considerable differences in perceived loudness. For the preference evaluations, statistically significant differences between the mean values of test and retest of about 3 dB are found for both sounds. A possible explanation for the difference between the mean values are some extreme evaluations in the retest condition by a few participants indicating a possible raised awareness for the unpleasant character of the sounds. An exclusion of these unreliably and extreme judging participants would render the test-retest differences from the preference task insignificant. For the level difference attributed to the sound character differences smaller than 2 dB are found between test and retest. The null-hypothesis of equal mean values ($H_0 : \mu_{test} = \mu_{retest}$) can not be refused and a statistically significant difference is not observed. In the following the mean values of the initial test based on the data of all 47 participants are used.

3.3.2 Inter-group stability of mean judgments

Two values of the frequency ratio $\rho = 199 : 150$ and $\rho = 4 : 3$ were included as signal parameters into each of the three measurement stages. The results from the three measurement stages give an insight into the reproducibility of the mean judgments by disjunct groups of participants and an indication of the equivalence of the results from the three groups. Figure 3.13 shows the mean values of the loudness, the preference and

Table 3.2: *Results of the repeated T-Test for the paired measurements of the sound D and E with $N = 47$ each.*

sound	$\rho = f_{01} : f_{10}$	measure	T-test (rep. meas.)
D	199 : 150	L_{loud}	$T(1,46) = 4.04$, $p = .000$
		L_{pref}	$T(1,46) = 2.74$, $p = .009$
		$\Delta L_{\text{sound character}}$	$T(1,46) = 1.89$, $p = .066$
E	4 : 3	L_{loud}	$T(1,46) = 3.58$, $p = .001$
		L_{pref}	$T(1,46) = 2.50$, $p = .016$
		$\Delta L_{\text{sound character}}$	$T(1,46) = 1.69$, $p = .097$

the sound character judgments for three groups of participants. Table 3.3 shows the results of the corresponding ANOVA statistics.

Statistically significant difference are found for the loudness judgments of the three groups for sound D (one-way ANOVA, $F(2,77) = 6.2$, $p < 0.01$) and also for sound E (one-way ANOVA, $F(2,77) = 10.7$, $p < 0.05$). Post-hoc tests indicate significant differences between group P3 and P6 for sound D and between the participant group P3 and each of the two other groups (P5 and P6) for sound E. No statistically significant differences were found between the preference evaluations of the three participant groups. For sound D the absolute differences are smaller than 2 dB and for sound E the differences are smaller than 2.5 dB between the three groups of participants. The values of the level difference $\Delta L_{\text{sound character}}$ between the the PSE for preference and for loudness also remain considerably stable with absolute differences smaller than 3 dB between the three groups, which are statistically not significant.

The distance between two sounds (X and Y) in terms of sound character may be quantified by the difference between the respective values for $\Delta L_{\text{sound character}}$, defined here as $d_{\text{X,Y}}$:

$$d_{\text{X,Y}} = \Delta L_{\text{sound character}}(\text{Y}) - \Delta L_{\text{sound character}}(\text{X}) \qquad (3.4)$$

The mean values of the three participant groups for the distance $d_{\text{D,E}}$ between the sounds D and E are given in Tab. 3.4. The Nullhypothesis of equal mean values from a one-way ANOVA can not be rejected ($F(2,77) = 0.2$, $p = 0.85$), indicating no significant differences between

[3]See Fig. 3.5 for details on the participant groups and measurement stages.

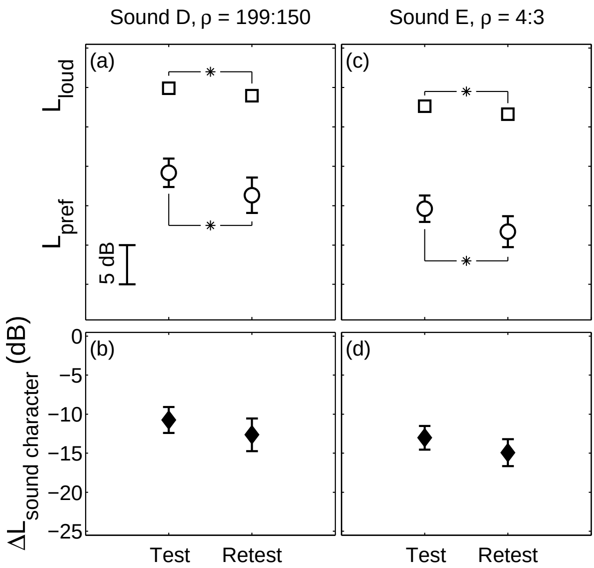


Figure 3.12: *Mean values for the repeated measurements of the preference, the loudness task and the sound character for the sound D ($\rho = 199:150$) and E ($\rho = 4:3$). Error bars indicate the standard error of the calculated mean ($N = 47$). For the loudness judgments the error bars are hidden behind the markers. Results of a T-test (rep. meas., paired samples) are given in Tab. 3.2. Statistically significant differences are indicated by * in this figure.*

the three groups of participants for the mean distance in sound character between sound D and E. Overall, the mean absolute value of the sound character difference is 2.8 dB bigger for sound E compared to sound D.

3.3.3 Reliability of the measurement method - Conclusion

For the mean loudness judgments, statistically significant differences were found between the results of the initial and the repeated measurements within a participant group. Even though statistically significant,

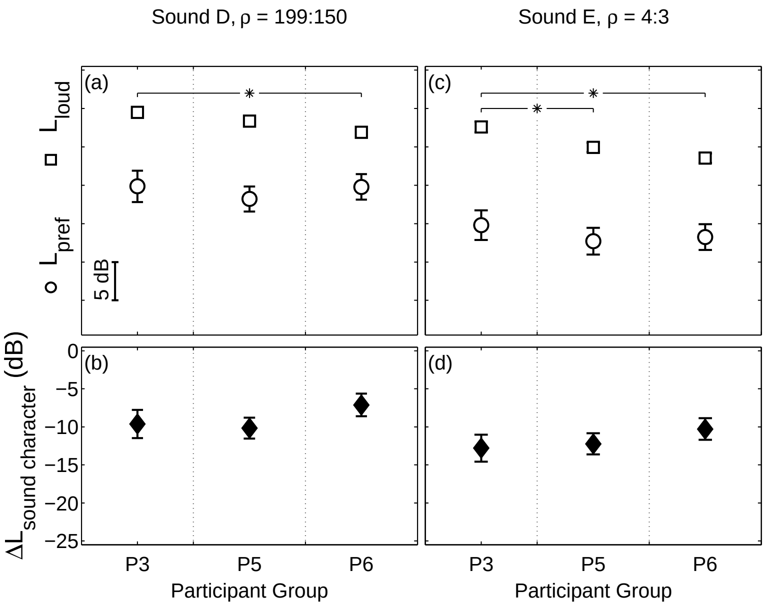


Figure 3.13: *Mean PSEs for loudness and for preference and the level difference attributed to the sound character for two sounds measured by three disjunct groups of participants, group P3 ($N = 29$) , group P5 ($N = 25$) and group P6 ($N = 26$)[3]. Error bars indicate the standard error of the calculated mean. ANOVA results are given in Tab. 3.3. Statistically significant differences are indicated by * in this figure.*

the differences between the mean values within the participant group are just above the just noticeable differences in level ($JNDL \approx 1$ dB[263]) and should not be misinterpreted as considerable loudness differences. Slightly bigger differences of up to 4 dB between the mean loudness values of three distinct groups of participants were found (which were also statistically significant). Nevertheless, these differences between the mean values of the three groups lie within the observed inter-individual standard deviations of about 3-4 dB, which are also comparable to those described in the literature[78,86,99,205].

With regard to the mean preference evaluationss, statistically significant differences of about 3 dB were found for the paired within subjects

Table 3.3: *Results of the ANOVA for the sound D and E, each judged by three disjunct groups of participants P3 ($N = 29$), P5 ($N = 25$) and P6 ($N = 26$).*

sound	$\rho = f_{01} : f_{10}$	measure	statistics
		L_{loud}	$F(2,77) = 6.22$, $p = 0.003$
D	199 : 150	L_{pref}	$F(2,77) = 0.254$, $p = 0.777$
		$\Delta L_{sound\ character}$	$F(2,77) = 0.983$, $p = 0.379$
		L_{loud}	$F(2,77) = 10.69$, $p = 0.000$
E	4 : 3	L_{pref}	$F(2,77) = 0.369$, $p = 0.692$
		$\Delta L_{sound\ character}$	$F(2,77) = 0.707$, $p = 0.496$

Table 3.4: *Mean values of the individually calculated distance $d_{D,E}$ between the level differences $\Delta L_{\text{sound character}}$ of the two sounds D and E. No significant differences between the three disjunct participant groups are found.*

participant group	sound character distance $d_{\mathrm{D,E}}$ (dB)	std. error
P3 ($N = 29$)	-3.17	2.08
P5 ($N = 25$)	-2.07	0.79
P6 ($N = 26$)	-3,17	1,24
overall ($N = 80$)	-2.83	0.88

samples, which become insignificant when excluding unreliably judging participants with extreme judgments in the retest condition. No statistically significant differences were found between the mean values of three disjunct groups of participants which were smaller than 2 dB, indicating the reproducibility of the preference evaluations.

In terms of the level difference attributed to the sound character, absolute differences in the mean values smaller than 2 dB (within a participant group) and of about 3 dB (between disjunct participant groups) were observed. Neither for the repeated (within subject) measurements, nor for the comparison of three disjunct groups of participants (between subjects) statistically significant differences were found. The findings support the robustness of the level difference $\Delta L_{\text{sound character}}$ in terms of the intra-individual reproducibility and the equivalence of the results between participant groups.

The relative distances between the two test sounds for the sound character distance $d_{\text{D,E}}$ are remarkably stable with a standard deviation smaller than 1 dB. Similarly, also the relative differences between the two repeatedly measured sounds with respect to the loudness judgments are only about 1 dB and smaller than 2 dB for the preference evaluations for each of the three participant groups.

3.4 Results - Influence of signal parameters on the judgments of loudness, evaluations of preference and sound character differences for multi-tone sounds

In this section the influence of the experimentally varied signal parameters, which served as independent variables, on the judgments is presented in the following subsections in detail:

— Contribution of basic sound elements like the two complex tones and the combination tones to judgments of a multi-tone composition, Sec. 3.4.1, p. 115

— Influence of the frequency ratio ρ between the fundamental frequencies on the judgments, Sec. 3.4.2, p. 120

— Influence of the mixing ratio between complex tones and combination tones on the judgments, Sec. 3.4.3, p. 124

— Influence of the spectral envelope on the judgments - falling upper slope slope_{DN}, Sec. 3.4.4, p. 128

— Influence of the spectral envelope - peak frequencies f_{peak} for three combinations of rising lower slope (slope_{UP}) and falling upper slope (slope_{DN}), Sec. 3.4.5, p. 132

— Relevance of frequency ranges for the judgments, Sec. 3.4.6, p. 140

In the following, the PSEs for loudness L_{loud} and for preference L_{pref} are each presented as relative values and the resulting $\Delta L_{\text{sound character}}$

is given as the absolute difference between the PSE for preference and the PSE for loudness. The mean values of the level difference $\Delta L_{\text{sound character}}$ were found to be negative for all test sounds, which means that the PSEs for preference were always lower than the respective PSEs for loudness. Apparently, equal loudness was not sufficient to make the the multi-tone test sound equally preferred as the reference and differences in terms of sound character are present for all multi-tone sounds compared to the reference sound. Additional level reductions were necessary to render the unpleasant multi-tone sounds equally preferred as the fixed reference stimulus. The more negative the value for $\Delta L_{\text{sound character}}$, the bigger is the difference in terms of sound character, which renders the multi-tone test sound unpleasant and, thus, the less preferred would a sound be, when equally loud as the reference.

3.4.1 Contribution of basic sound elements to judgments and evaluations of a multi-tone composition

3.4.1.1 Stimuli

The contribution of the basic sound elements to the subjective judgments was investigated for the two frequency ratios $\rho = 199 : 150$ and $\rho = 4 : 3$ to identify the contribution of the basic elements to the overall assessment of the complete signals. The superposition of the basic elements yields different characteristics for the two values of the frequency ratio. A value of $\rho = 4 : 3$ leads to matched partials from the third harmonic which leads to a rather irregular spectral envelope due to the superposition of multiple partials at dedicated frequencies. For a value of $\rho = 199 : 150$ partials matching in frequency do not occur until rather high frequencies, but partials at low frequencies are closely spaced which leads to beats and modulations depending on the combination of elements. The contribution of the complex tones CX1 (sound F) and CX2 (sound G) was only investigated for one fixed frequency ratio between the fundamental frequencies ($\rho = 199 : 150$), because the lower complex tone CX1 is the same for both values of ρ and, despite a slightly higher pitch, no big differences for the higher complex tone alone were expected for $f_{01} = 133.33$ Hz ($\rho = 4 : 3$) instead of $f_{01} = 132.66$ Hz ($\rho = 199 : 150$). Table 3.5 gives an overview of the signal parameters for the eight stimuli.

Table 3.5: *Parameters for the composition of the eight multi-tone test stimuli based on three constituting basic elements:*

ID	slope_{UP} (dB/oct)	slope_{DN} (dB/oct)	f_{peak} (Hz)	ratio $\rho = \frac{f_{01}}{f_{10}}$	basic multitone elements	attenuation δ (dB)	freq. range attenuated by 10 dB
D	—	-6	100	199:150	CX1+CX2+CTs	-10	—
E	—	-6	100	4:3	CX1+CX2+CTs	-10	—
F	—	-6	100	199:150	CX1	—	—
G	—	-6	100	199:150	CX2	—	—
D2	—	-6	100	199:150	CX1+CX2	—	—
E2	—	-6	100	4:3	CX1+CX2	—	—
D3	—	-6	100	199:150	CTs	—	—
E3	—	-6	100	4:3	CTs	—	—

3.4.1.2 Participants

The experiments were carried out by 47 paid volunteers (23 female, 24 male) mainly from the university. The mean age of the participants was 23 years (min=19 years, max=31 years). Approximately one half of the participants already participated in other previous psychoacoustic experiments (10 female, 13 male), the other half was inexperienced with psychoacoustic tests (13 female, 11 male). All participants reported no hearing difficulties.

3.4.1.3 Results

The mean results of the loudness, the preference and the level difference attributed to the sound character are shown in Fig. 3.14 for the basic sound elements and combinations of them.

Loudness
A significant main effect of the five different elements as conditions is found for the loudness judgments (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.862, 131.644) = 39.451, p = 0.000$). Post-hoc tests (Bonferroni corrected) on the loudness judgments show significant differences between the two frequency ratios $\rho = 199 : 150$ and $\rho = 4 : 3$, the latter requiring lower levels for the loudness PSEs, which means that a signal with $\rho = 4 : 3$ would yield a higher loudness at an equal dB(A) level than a value of $\rho = 199 : 150$. Furthermore, the two complex tones alone (CX1 and CX2, each) require up to 5 dB more attenuation than the combination tones (CTs) and the combined sound elements

(CX1+CX2+CTs and CX1+CX2), which do not differ significantly with respect to the loudness PSEs for each of the two values of the frequency ratio ρ. The complex tone with the higher fundamental frequency (CX2) is also significantly different from the complex tone with the lower fundamental frequency (CX1).

Preference

A significant main effect is found for the preference evaluations (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(3.526, 162.185) = 23.429$, $p = 0.000$). Post-hoc tests (Bonferroni corrected) reveal that the mean difference between the combinations CX1+CX2+CTs and CX1+CX2 are not statistically significant for both values of the frequency ratio ρ. Accordingly, it makes no difference with respect to the preference evaluation whether the combination tones CTs are added to a superposition of the two complex tones CX1+CX2 or not. For the combination tones alone (CTs), no statistically significant difference is found between the two values of ρ. The complex tones alone (CX1, CX2) need more attenuation than the combinations tones alone (CTs) and the complex tone with the higher fundamental frequency (CX2) needs the most attenuation and, thus, would be the least preferred sound of this study at equal A-weighted level.

Sound character

The influence of the constituting basic sound elements on the level difference $\Delta L_{\text{sound character}}$ is shown in Fig. 3.14(b). Values of $\Delta L_{\text{sound character}}$ from -10.7 dB to -20.9 dB reflect the considerable gap between equal preference and equal loudness with regard to the reference sound. In addition, differences of more than 10 dB in between the different conditions of the multi-tone test sounds are observed. In both cases, this gap is ascribable to differences in the sound character. A range of about 10 dB for $\Delta L_{\text{sound character}}$ in between the multi-tone test sounds leads to statistically significant differences (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.934, 180.976) = 13.198, p = 0.000$). The results of a post-hoc paired comparison yields similar relationships like for the results of the preference evaluations. The basic elements alone (CX1, CX2 and CTs) yield considerably more negative sound character values than the baseline consisting of all elements (CX1+CX2+CTs). In terms of $\Delta L_{\text{sound character}}$, no statistically significant difference is found

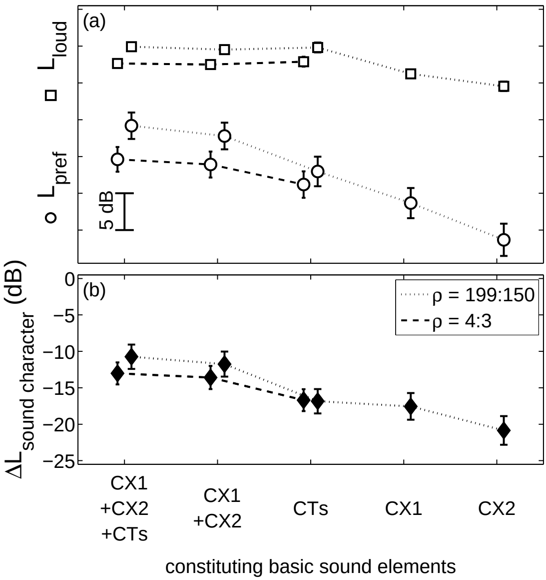


Figure 3.14: *Mean PSEs for loudness (a, open squares) and for preference (a, open circles), each plotted as relative values, and the level difference attributed to the sound character (b, filled diamonds) plotted for the complete multi-tone sound (CX1+CX2+CTs), the mix of the two complex tones (CX1+CX2), the combination tones only (CTs), and the two complex tones (CX2, CX1). Frequency ratios between the fundamental frequencies are indicated by dotted ($\rho = 199 : 150$) and dashed ($\rho = 4 : 3$) lines. Errorbars indicate the standard deviation of the calculated mean ($N = 47$, sometime even smaller than the symbols). The basic elements need considerably more attenuation to become equally loud and equally preferred as the combinations of these elements.*

between the baseline and the combination of the two complex tones (CX1+CX2). Like in the preference evaluations, it makes no difference for $\Delta L_{\text{sound character}}$, whether the combination tones are added to a multi-tone composition or not. No statistically significant influence of ρ was found for the stimulus baseline (CX1+CX2+CTs), the combination of the complex tones (CX1+CX2) and the combination tones alone (CTs).

3.4.1.4 Key findings

— The basic elements (CX1, CX2 and CTs) need to be considerably lower in level to become equally preferred as the combinations (CX1 +CX2 and CX1+CX2+CTs) of them. Thus, at equal dB(A)-levels, the basic elements would be considerably less preferred than the combined sounds.

— The lower the overall number of partials, the lower are the PSEs for loudness and preference of the sounds. Hence, a higher number of partials would presumably be more preferred at equal dB(A)-levels.

— The PSEs for loudness and preference are lower for the complex tones with the higher fundamental frequency (CX2, $f_{01} = 132.66$ Hz) than for the complex tone with the lower fundamental frequency (CX1, $f_{01} = 100$ Hz).

— No significant differences are found between the sounds with (CX1 +CX2+CTs) or without (CX1+CX2) the combination tones (CTs) for the preference PSEs and the level difference $\Delta L_{\text{sound character}}$ for each of the values of the frequency ratio ρ.

3.4.2 Influence of the frequency ratio between the fundamentals on the judgments and evaluations

3.4.2.1 Stimuli

In this experiment, the frequency ratio between the fundamental frequencies $\rho = f_{01} : f_{10}$ and therewith the higher fundamental f_{01} was varied for complete test sounds. The varied signal parameter is identical to the semantic differential, but only five values of the frequency ratio ρ were investigated. The values of the frequency ratio ρ included three ratios equal to ratios of small integer numbers ($\rho = 9 : 7$, $13 : 10$, $4 : 3$) as well as two different ratios, detuned by Δf_{01}, resulting in ratios of rather large integers ($\rho = 1293 : 1000$ and $\rho = 199 : 150$). The signals used here, cover the most unpleasant and the most pleasant sounds from the semantic differential study described in Chap. 2. In addition, also representatives of the sound groups identified in the semantic differential study (sound group (a): $\rho = 1293 : 1000$, sound group (b): $\rho = 4 : 3$, and sound group (c): $\rho = 199 : 150$) were included. Table 3.6 gives the details of the fundamental frequencies and the other signal parameters for the five stimuli, which consisted of all basic elements (CX1+CX2+CTs).

Table 3.6: *Signal parameters of the five multi-tone test stimuli with a variation of the frequency ratio ρ.*

ID	slope_{UP} (dB/oct)	slope_{DN} (dB/oct)	f_{peak} (Hz)	ratio $\rho = \frac{f_{01}}{f_{10}}$	basic multitone elements	attenuation δ (dB)	freq. range attenuated by 10 dB
A	—	-6	100	9:7	CX1+CX2+CTs	-10	—
B	—	-6	100	1293:1000	CX1+CX2+CTs	-10	—
C	—	-6	100	13:10	CX1+CX2+CTs	-10	—
D	—	-6	100	199:150	CX1+CX2+CTs	-10	—
E	—	-6	100	4:3	CX1+CX2+CTs	-10	—

3.4.2.2 Participants

The experiments were carried out by 47 paid volunteers (23 female, 24 male) mainly from the university. The mean age of the participants was 23 years (min=19 years, max=31 years). Approximately one half of the participants already participated in other previous psychoacoustic experiments (10 female, 13 male), the other half was inexperienced with psychoacoustic tests (13 female, 11 male). All participants reported no hearing difficulties.

3.4.2.3 Results

The results of the loudness judgments and the preference evaluations for a variation of the frequency ratio ρ between the fundamental frequencies is shown in Fig. 3.15(a). The data presented for a ratio $\rho = 199 : 150$ (sound D) and for $\rho = 4 : 3$ (sound E) is exactly the same as for the complete stimulus (CX1+CX2+CTs) shown in Fig. 3.14 of the previous section.

Loudness
A significant main effect of the frequency ratio ρ between the fundamentals is found for the loudness judgments (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(3.280, 150.865) = 31.821, p = 0.000$). Post-hoc tests reveal that the sound with a value of $\rho = 4 : 3$ is statistically different from all other sounds and thus would be considerably louder at equal dB(A)-level than all others. The difference of about 2 dB between the two sounds which are based on ratios of large integer numbers ($\rho = 1293 : 1000$ and $\rho = 199 : 150$) and all others is also statistically significant. Hence, the sounds based on ratios of small integer numbers need about 2 dB lower levels to be equally loud as sounds which are based on frequency ratios of rather large integer numbers.

Preference
A significant main effect of the frequency ratio ρ between the fundamentals is found for the preference evaluations (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.182, 100.358) = 7.639, p = 0.001$). Similar like in the loudness judgments, the two sounds which are based on ratios of large integer numbers ($\rho = 1293 : 1000$ and $\rho = 199 : 150$) can be about 4 dB higher in A-weighted level while being equally preferred as the other sounds, based on ratios of small integer numbers ($\rho = 9 : 7$, $\rho = 13 : 10$ and $\rho = 4 : 3$).

Sound character
Figure 3.15(b) shows the influence of the frequency ratio ρ between the fundamentals on the level difference attributed to the sound character $\Delta L_{\text{sound character}}$. Values of $\Delta L_{\text{sound character}}$ from -10.5 dB to -13.5 dB

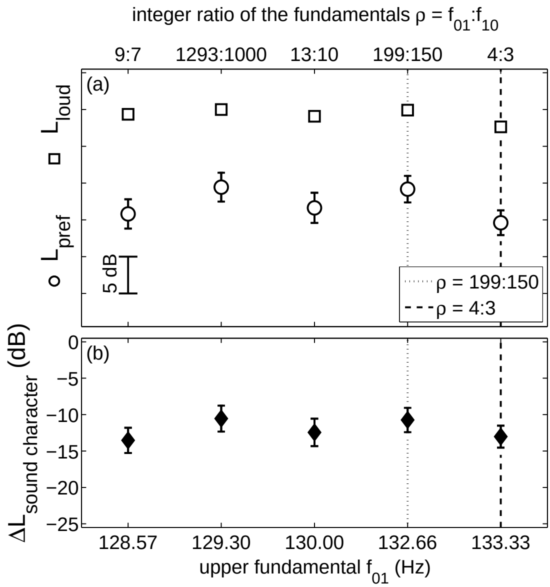


Figure 3.15: *Mean PSEs for loudness (a, open squares) and for preference (a, open circles), each plotted as relative values, and the level difference attributed to the sound character (b, filled diamonds) plotted for the five ratios of the fundamental frequencies* ρ*. Dotted and dashed lines highlight the frequency ratios (*$\rho = 199 : 150$*) and (*$\rho = 4 : 3$*). The resulting upper fundamental frequencies are given at the bottom. Errorbars indicate the standard deviation of the calculated mean (*$N = 47$*, sometime even smaller than the symbols).*

are observed. Even though a statistically significant main effect is found (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.352, 108.173) = 3.355, p = 0.031$), only the difference of about 3 dB between a ratio of $\rho = 9 : 7$ and $1293 : 1000$ and the difference of about 2 dB between $\rho = 13 : 10$ and $1293 : 1000$ are statistically significant. Overall, the spread in between the five multi-tone test sounds of about 3 dB is rather small compared to e.g. the differences between the sound elements which were presented in the previous section.

3.4.2.4 Key findings

— Sounds based on frequency ratios of small integer numbers (e.g., $\rho = 4 : 3$) need 2 dB lower levels than sounds based on frequency ratios of large integers (e.g., $\rho = 1293 : 1000$) to be equally loud.

— For equal preference, the sounds based on frequency ratios of small integer numbers require 4 dB lower levels than the sound based on frequency ratios of large integers.

3.4.3 Influence of the mixing ratio between complex tones and combination tones on the judgments and evaluations

3.4.3.1 Stimuli

The mixing ratio between the two combined complex tones and the combination tones is determined by the attenuation parameter δ. A value $\delta = -\infty$ is represented by the two complex tones without any combination tones (CX1+CX2 only) and $\delta = +\infty$ would refer to a condition of the combination tones without the two complex tones (CTs only). Table 3.7 gives an overview of the investigated signal parameters. The sounds D5 and E5 reflect the baseline configuration in which the combination tones (CTs) were attenuated by $\delta = 10$ dB compared to the combined complex tones (CX1+CX2).

Table 3.7: *Parameters of the ten multi-tone test stimuli with a variation of the attenuation parameter δ, which determines the mixing ratio of the two complex tones together (CX1+CX2) and the combination tones (CTs).*

ID	slope_{UP} (dB/oct)	slope_{DN} (dB/oct)	f_{peak} (Hz)	ratio $\rho = \frac{f_{01}}{f_{10}}$	basic multitone elements	attenuation δ (dB)	freq. range attenuated by 10 dB
D3	—	-6	100	199:150	CTs	—	—
E3	—	-6	100	4:3	CTs	—	—
D4	—	-6	100	199:150	CX1+CX2+CTs	-5	—
E4	—	-6	100	4:3	CX1+CX2+CTs	-5	—
D5	—	-6	100	199:150	CX1+CX2+CTs	-10	—
E5	—	-6	100	4:3	CX1+CX2+CTs	-10	—
D6	—	-6	100	199:150	CX1+CX2+CTs	-15	—
E6	—	-6	100	4:3	CX1+CX2+CTs	-15	—
D2	—	-6	100	199:150	CX1+CX2	—	—
E2	—	-6	100	4:3	CX1+CX2	—	—

3.4.3.2 Participants

The experiments were carried out by 47 paid volunteers (23 female, 24 male) mainly from the university. The mean age of the participants was 23 years (min=19 years, max=31 years). Approximately one half of the participants already participated in other previous psychoacoustic experiments (10 female, 13 male), the other half was inexperienced with psychoacoustic tests (13 female, 11 male). All participants reported no hearing difficulties.

3.4.3.3 Results

Figure 3.16 (a) shows the the results of the loudness and the preference task, 3.16 (b) shows the level difference attributed to the sound character as a function of the attenuation δ of the combination tones.

Loudness
The resulting mean loudness PSEs are rather constant for a variation of the attenuation parameter δ, but a consistent offset of about 2.5 dB between the two values of the frequency ratio $\rho = 199 : 150$ and $\rho = 4 : 3$ can be seen. This is manifested by a significant main effect for the factor ρ (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1, 46) = 156.706, p = 0.001$), while the mixing ratio determined by the attenuation factor δ is not statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.624, 120.705) = 2.516, p = 0.069$). Also the interaction of the two factors ($\rho * \delta$) is not statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(3.312, 152, 366) = 1.223, p = 0.304$). Because of the high number of partials contained in the complex tones (with 30 partials each), it seems as if the loudness is rather unaffected by additional partials in terms of the combination tones as long as the A-weighted level and the resulting temporal structure remains the same.

Preference
In general the levels at the PSEs for preference slightly decrease with increasing attenuation factor δ which means that with an increasing portion of the combination tones the sounds become more unpleasant and lower levels are required to render the multi-tone sounds equally preferred in comparison to the common reference. Significant main effects are found for the attenuation parameter δ (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(3.095, 142.378) = 6.609, p = 0.000$), for the frequency ratio ρ (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1, 46) = 31.965, p = 0.000$) and also for the interaction of the two factors (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(3.525, 162.160) = 4.141, p = 0.005$).

Sound character
The level difference attributed to the sound character becomes more negative with rising values of δ and a higher amount of combination tones

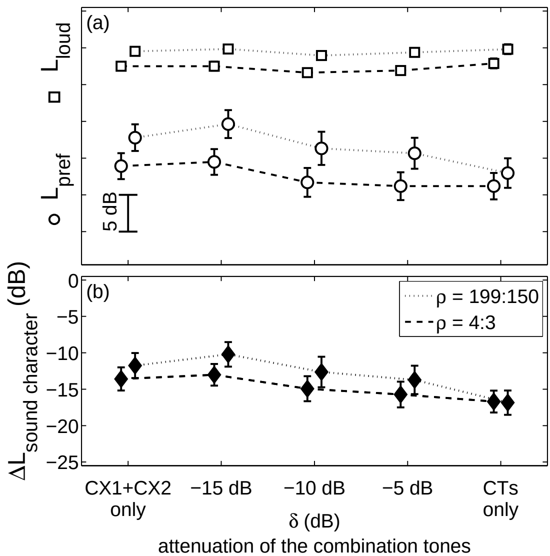


Figure 3.16: *Mean PSEs for loudness (a, open squares) and for preference (a, open circles), each plotted as relative values, and the level difference attributed to the sound character (b, filled diamonds) plotted for the five levels of the attenuation parameter δ. The two frequency ratios between the fundamentals are indicated by dotted ($\rho = 199 : 150$) and dashed ($\rho = 4 : 3$) lines. Errorbars indicate the standard deviation of the calculated mean ($N = 47$, sometime even smaller than the symbols).*

leads to bigger level differences between the PSEs for preference and for loudness. The effects of the mixing parameter δ (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(3.134, 144.181) = 7.841, p = 0.000$), the frequency ratio ρ (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1, 46) = 6.968, p = 0.011$) and also the interaction between the two (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(3.664, 168.561) = 3.064$, $p = 0.018$) are statistically significant. However, for a configuration with combination tones only (CTs only), the level difference becomes the same for both values of the frequency ratio ρ.

3.4.3.4 Key findings

— No influence of the attenuation of the combination tones δ on the loudness judgments.

— The higher the amount of the combination tones, the lower the PSEs for preference. Thus, for the CTs only condition, the levels of the multi-tone signals have to be reduced by about 2 dB (for $\rho = 4:3$) and 6 dB (for $\rho = 199:150$) to remain equally preferred as the combination of the two complex tones without the combination tones (CX1+CX2).

— The level difference $\Delta L_{\text{sound character}}$ becomes more negative with increasing amount of combination tones - the effect is only about 3 dB.

3.4.4 Influence of spectral envelope on the judgments and eveluations - falling upper slope

3.4.4.1 Stimuli

The falling upper slope, denoted slope_{DN}, which determines the decrease of the partial's levels with increasing frequency, was varied for two values of the frequency ratio between the fundamentals ($\rho = 199 : 150$ and $\rho = 4 : 3$). An overview of the investigated parameter combinations is given in Tab. 3.8.

Table 3.8: *Parameters of the eight multi-tone test stimuli with a variation of the parameter* slope_{DN}*, which describes the decrease of the partials' amplitude with rising frequency.*

ID	slope_{UP} (dB/oct)	slope_{DN} (dB/oct)	f_{peak} (Hz)	ratio $\rho = \frac{f_{01}}{f_{10}}$	basic multitone elements	attenuation δ (dB)	freq. range attenuated by 10 dB
D7	—	-6	100	199:150	CX1+CX2+CTs	-10	—
E7	—	-6	100	4:3	CX1+CX2+CTs	-10	—
D8	—	-9	100	199:150	CX1+CX2+CTs	-10	—
E8	—	-9	100	4:3	CX1+CX2+CTs	-10	—
D9	—	-12	100	199:150	CX1+CX2+CTs	-10	—
E9	—	-12	100	4:3	CX1+CX2+CTs	-10	—
D10	—	-15	100	199:150	CX1+CX2+CTs	-10	—
E10	—	-15	100	4:3	CX1+CX2+CTs	-10	—

3.4.4.2 Participants

The experiments were carried out by 48 paid volunteers (22 female, 26 male), which were students or other university members. A subgroup of 18 participants (6 female, 12 male) already participated in the experiments of the first measurement stage, in which other parameter variations were investigated. The mean age of the 48 participants was 24 years (min=20 years, max=36 years). All participants reported no hearing difficulties.

3.4.4.3 Results

The influence of the upper slope on the PSEs for loudness, for preference and the level difference attributed to the sound character is shown in Fig. 3.17.

Loudness

For the loudness judgments the PSEs rise monotonously for a variation of the falling upper slope from $\text{slope}_{\text{DN}} = -6$ dB/octave to $\text{slope}_{\text{DN}} = -15$ dB/octave by about 5 dB for each of the two values of the frequency ratio ρ. This means that sounds with a steep decline towards high frequencies and an upper slope of -15 dB/octave can be 5 dB higher in A-weighted sound pressure level than equivalent sounds with a slope of -6 dB/octave while remaining constant in perceived loudness compared to the reference sound. This effect is statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1.487, 69.890) = 56.874, p = 0.000$). The overall offset in the loudness values of about 2.5 dB between the two values of the frequency ratio ρ is also statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1, 47) = 166.53, p = 0.000$) as well as the interaction of the two factors slope_{DN} and ρ (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.829, 132.981) = 16.514, p = 0.000$).

Preference

For the preference evaluations, the monotonous rise of the PSEs with increased steepness of the falling upper slope is even more pronounced than for the loudness judgments. Here, the sounds with a steeper decline towards high frequencies of $\text{slope}_{\text{DN}} = -15$ dB/octave, and thus less high frequency content, can be up to 15 dB higher in A-weighted level compared to a multi-tone sound with a falling upper slope of $\text{slope}_{\text{DN}} = -6$ dB/octave, which is significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1.410, 66.278) = 100.654, p = 0.000$). Again the two values of frequency ratio ρ lead to a significant offset in the PSEs - all sound with $\rho = 199 : 150$ can be up to 8 dB higher in level than those with $\rho = 4 : 3$ and still be equally preferred (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1, 47) = 73.639, p = 0.000$). The differences between the preference evaluations for the two values of the frequency ratio ρ become smaller for steeper slopes slope_{DN} – this interaction effect is also significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.559, 120.296) = 5.532, p = 0.002$).

Sound character

The results of the level difference $\Delta L_{\text{sound character}}$ attributed to the sound character are shown in Fig. 3.17(b)). The level difference

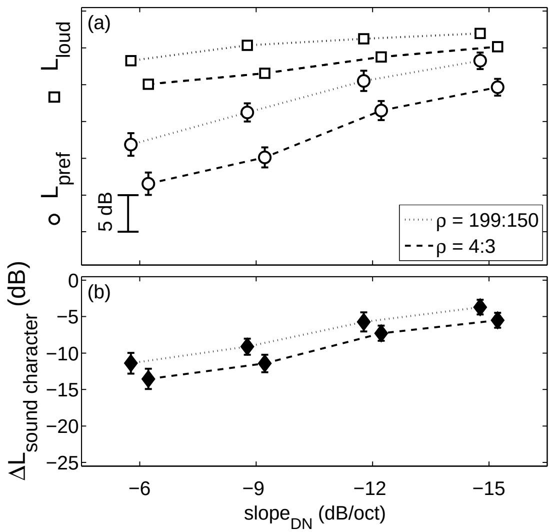


Figure 3.17: *Mean PSEs for loudness (a, open squares) and for preference (a, open circles), each plotted as relative values, and the level difference attributed to the sound character (b, filled diamonds) plotted as a function of four different slopes* $\mathrm{slope}_{\mathrm{DN}}$ *towards higher frequencies. frequency ratios between the fundamentals are indicated by dotted* ($\rho = 199 : 150$) *and dashed* ($\rho = 4 : 3$) *lines. Errorbars indicate the standard deviation of the calculated mean* ($N = 48$, *sometime even smaller than the symbols).*

$\Delta L_{\text{sound character}}$ becomes monotonously smaller with increasing steepness of the upper slope. The difference of about -12 dB for an upper slope of $\text{slope}_{\text{DN}} = -6$ dB/octave reduces to -4 dB for an upper slope of $\text{slope}_{\text{DN}} = -15$ dB/octave for a frequency ratio of $\rho = 199 : 150$. Similarly, a decrease in the level difference from -14 dB to -6 dB is found for $\rho = 4 : 3$. For both frequency ratios ρ the dynamic range in terms of $\Delta L_{\text{sound character}}$ for this parameter variation is about 8 dB. This effect of the slope is significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1.511, 71.022) = 43.753, p = 0.000$). The difference between the two values of the frequency ratio ρ is also significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1, 47) = 13.241, p = 0.001$), while the interaction of slope_{DN} and ρ is not (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.747, 129.091) = 0.401, p = 0.735$), suggesting a rather constant offset between the values for the two frequency ratios ρ.

3.4.4.4 Key findings

— The PSEs for loudness and for preference rise with increasing steepness of the upper falling slope - the effects are about 5 dB for loudness and 12 dB for preference. Transferred to equal dB(A)-levels, sounds with less high frequency content would be preferred over sounds with more high frequency content.

— The level difference $\Delta L_{\text{sound character}}$ becomes smaller with increasing steepness of the falling upper slope, while the differences between the two values of the frequency ratio ρ remain rather constant over all tested slope values.

3.4.5 Influence of spectral envelope on the judgments and evaluations - peak frequencies

3.4.5.1 Stimuli

The influence of the peak frequency was determined for three combinations of rising lower slope (slope_{UP}) and falling upper slope (slope_{DN}). The extreme values for the falling upper slope of −6 dB/octave and −15 dB/octave, described in Sec. 3.4.4 were booth combined with a rising lower slope of +6 dB/octave for five different peak frequencies f_{peak} from 100 Hz up to 500 Hz in 100 Hz steps. In addition, the falling upper slope $\text{slope}_{\text{DN}} = -15$ dB/octave was also combined with a rising lower slope $\text{slope}_{\text{UP}} = +15$ dB/octave for two different peak frequencies (300 Hz and 500 Hz). An overview of the stimulus configurations can be found in Tab. 3.9. The values of the rising lower slope slope_{UP} are given in brackets for all stimuli with a peak frequency of 100 Hz, because this parameter does not directly apply in these cases, due to the lack of partials below 100 Hz.

3.4.5.2 Participants

The experiments are carried out by 34 participants (17 female and 17 male). Approximately 53% of the participants (9 female, 9 male) are familiar with psycho-acoustical experiments in general, participated in previous listening tests dealing with aircraft interior sounds or previous measurement stages of this study. The other 47% (8 female, 8 male) had no prior experience. The mean age of the 34 participants was 23 years (min=18 years, max=30 years). All participants reported no hearing difficulties.

3.4.5.3 Results

The results for a variation of the peak frequency f_{peak} for the three combinations of rising slope (slope_{UP}) and falling slope (slope_{DN}) are presented in three figures:

(I) $\text{slope}_{\text{UP}} = 6$ dB/oct, $\text{slope}_{\text{DN}} = -6$ dB/oct in Fig. 3.18(I)

(II) $\text{slope}_{\text{UP}} = 6$ dB/oct, $\text{slope}_{\text{DN}} = -15$ dB/oct in Fig. 3.19(II)

(III) $\text{slope}_{\text{UP}} = 15$ dB/oct, $\text{slope}_{\text{DN}} = -15$ dB/oct in Fig. 3.19(III)

Table 3.9: *Parameters of the 24 multi-tone test stimuli with a variation of the peak frequency f_{peak} and the slopes* $\mathrm{slope}_{\mathrm{UP}}$ *and* $\mathrm{slope}_{\mathrm{DN}}$*, describing the spectral envelope of the partials.*

ID	$\mathrm{slope}_{\mathrm{UP}}$ (dB/oct)	$\mathrm{slope}_{\mathrm{DN}}$ (dB/oct)	f_{peak} (Hz)	ratio $\rho = \frac{f_{01}}{f_{10}}$	basic multitone elements	attenuation δ (dB)	freq. range attenuated by 10 dB
D14	(+6)	-6	100	199:150	CX1+CX2+CTs	-10	—
E14	(+6)	-6	100	4:3	CX1+CX2+CTs	-10	—
D15	+6	-6	200	199:150	CX1+CX2+CTs	-10	—
E15	+6	-6	200	4:3	CX1+CX2+CTs	-10	—
D16	+6	-6	300	199:150	CX1+CX2+CTs	-10	—
E16	+6	-6	300	4:3	CX1+CX2+CTs	-10	—
D17	+6	-6	400	199:150	CX1+CX2+CTs	-10	—
E17	+6	-6	400	4:3	CX1+CX2+CTs	-10	—
D18	+6	-6	500	199:150	CX1+CX2+CTs	-10	—
E18	+6	-6	500	4:3	CX1+CX2+CTs	-10	—
D19	(+6)	-15	100	199:150	CX1+CX2+CTs	-10	—
E19	(+6)	-15	100	4:3	CX1+CX2+CTs	-10	—
D20	+6	-15	200	199:150	CX1+CX2+CTs	-10	—
E20	+6	-15	200	4:3	CX1+CX2+CTs	-10	—
D21	+6	-15	300	199:150	CX1+CX2+CTs	-10	—
E21	+6	-15	300	4:3	CX1+CX2+CTs	-10	—
D22	+6	-15	400	199:150	CX1+CX2+CTs	-10	—
E22	+6	-15	400	4:3	CX1+CX2+CTs	-10	—
D23	+6	-15	500	199:150	CX1+CX2+CTs	-10	—
E23	+6	-15	500	4:3	CX1+CX2+CTs	-10	—
D24	+15	-15	300	199:150	CX1+CX2+CTs	-10	—
E24	+15	-15	300	4:3	CX1+CX2+CTs	-10	—
D25	+15	-15	500	199:150	CX1+CX2+CTs	-10	—
E25	+15	-15	500	4:3	CX1+CX2+CTs	-10	—

Loudness (I)

The results for the shallow falling upper slope $\mathrm{slope}_{\mathrm{DN}} = -6$ dB/oct (and a rising slope of $\mathrm{slope}_{\mathrm{UP}} = 6$ dB/oct) in Fig. 3.18(I) show a clear difference of about 3 dB between the PSEs for loudness for the two values of the frequency ratio ρ, which is statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1,33) = 125.681$, $p < 0.001$). Additionally, a slight local maximum of the loudness PSEs at $f_{\mathrm{peak}} = 400$ Hz compared to lower and higher peak frequencies can be seen. This influence of f_{peak} on the loudness PSEs is also statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.44, 80.508) = 4.527, p = 0.009$). The interaction effect between both factors ($\rho * f_{\mathrm{peak}}$) is not significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(3.618, 119.405) = 1.256, p = 0.292$), indicating a rather constant offset between the PSEs for the two values of the fre-

(I)

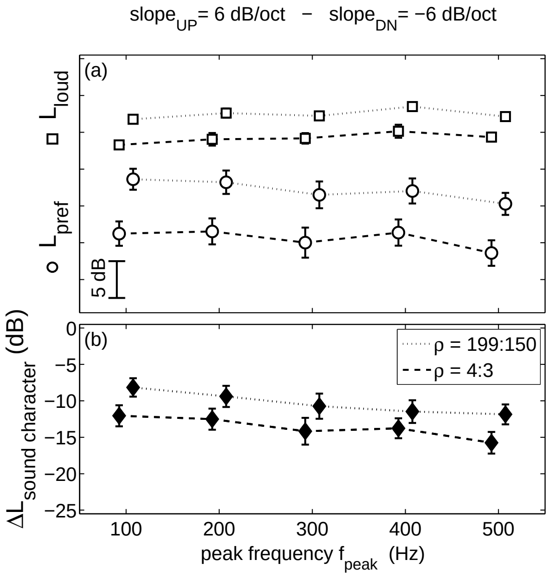


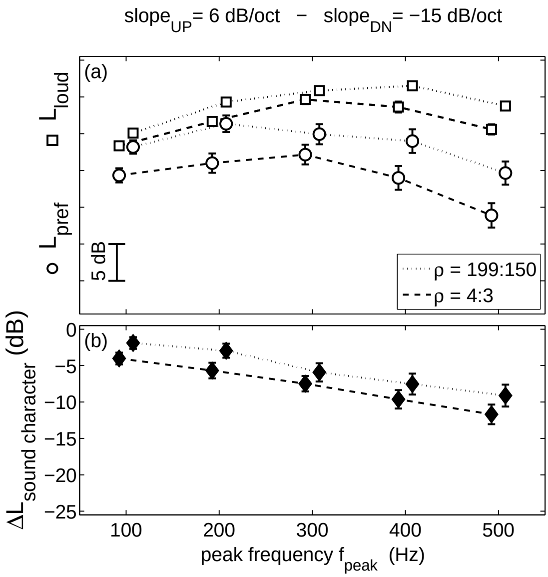
Figure 3.18: *Mean PSEs for loudness (a, open squares) and for preference (a, open circles), each plotted as relative values, and the level difference attributed to the sound character (b, filled diamonds) plotted as a function of five different peak frequencies f_{peak} for a rising slope of* $\mathrm{slope_{UP}} = +6$ *dB/oct. and a falling slope of* $\mathrm{slope_{DN}} = -6$ *dB/oct. The frequency ratios between the fundamental frequencies are indicated by dotted ($\rho = 199:150$) and dashed ($\rho = 4:3$) lines. Errorbars indicate the standard deviation of the calculated mean ($N = 34$, sometime even smaller than the symbols).*

quency ratio ρ.

Loudness (II)

For the more steeply falling upper slope $\mathrm{slope_{DN}} = -15$ dB/oct (and a rising slope of $\mathrm{slope_{UP}} = 6$ dB/oct) in Fig. 3.19(II), a more pronounced local maximum of the PSEs for loudness than for the shallower falling upper slope (I) can be seen. This influence of f_{peak} on the loudness PSEs

(II)



(III)

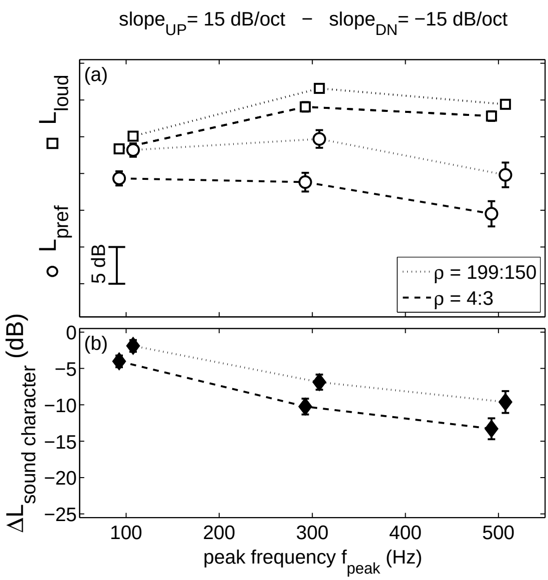


Figure 3.19: *Same as figure 3.18(I), but for different slope configurations*

is statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.562, 84.543) = 50.246, p = 0.000$). The maxima are located at $f_{\text{peak}} = 300$ Hz for $\rho = 4 : 3$ and $f_{\text{peak}} = 400$ Hz for $\rho = 199 : 150$. This effect of the frequency ratio ρ (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1, 33) = 56.921, p = 0.000$) and the interaction effect between both factors ($\rho * f_{\text{peak}}$) are also significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(3.747, 123.659) = 6.766, p = 0.000$).

Loudness (III)
For a steep rising lower ($\text{slope}_{\text{UP}} = 15$ dB/oct) and a steep falling upper slope ($\text{slope}_{\text{DN}} = -15$ dB/oct) shown in Fig. 3.19(III) the PSEs for loudness indicate a maximum at $f_{\text{peak}} = 300$ Hz for each of the two values of the frequency ratio ρ. This result is similar to that found for the shallower rising slope $slope_{UP} = 6$ dB/octave, shown in Fig. 3.19(II). For the loudness PSEs, significant effects of the peak frequency f_{peak} (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1.47, 48.498) = 67.695, p = 0.000$) and of the frequency ratio ρ (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1, 33) = 45.910, p = 0.000$) are found. The interaction effect is statistically not significant (ANOVA, rep. meas., Greenhouse-Geisser corrected, loudness: $F(1.978, 65.281) = 2.863, p = 0.065$), which indicates a rather constant offset of the loudness PSEs between the two values of the frequency ratio ρ over all peak frequencies.

Preference (I)
For a rising slope of $\text{slope}_{\text{UP}} = 6$ dB/oct) and a falling upper slope $\text{slope}_{\text{DN}} = -6$ dB/oct (shown in Fig. 3.18(I)), the PSEs for preference decrease with increasing peak frequency f_{peak}, which is statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.213, 73.026) = 3.428, p = 0.033$). The difference between the PSEs for the two values of the frequency ratio ρ of about 8 dB is also statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1, 33) = 35.278, p = 0.000$). The interaction of the two parameters is not significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.698, 89.035) = 0.844, p = 0.463$), indicating a constant offset between the two values of the frequency ratio ρ.

Preference (II)

For a steeper decline of high frequencies ($\text{slope}_{\text{DN}} = -15$ dB/oct. (and $\text{slope}_{\text{UP}} = 6$ dB/oct., Fig. 3.19(II)) the PSEs for preference have a pronounced maximum as a function of the peak frequency f_{peak} in the spectrum. This maximum reflects a local optimum configuration, because the level of the multi-tone sound can be higher for this particular peak frequency than for lower or higher peak frequencies. This effect of the peak frequency f_{peak} is statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(3.096, 102.177) = 15.759, p = 0.000$). The PSEs are considerably different for the two values of frequency ratio ρ over all values of the peak frequency, yielding statistical significance (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1, 33) = 42.592, p = 0.000$). For a frequency ratio of $\rho = 199 : 150$ the maximum is at $f_{\text{peak}} = 200$ Hz and for $\rho = 4 : 3$ at $f_{\text{peak}} = 300$ Hz. The interaction effect of the frequency ratio and the peak frequency ($\rho * f_{\text{peak}}$) is statistically not significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.633, 86.903) = 2.572, p = 0.067$).

Preference (III)

The results for a rising slope $\text{slope}_{\text{UP}} = 15$ dB/oct. (in Fig. 3.19(III)) are similar to those for the more shallow rising slope of $\text{slope}_{\text{UP}} = 6$ dB/oct. (in Fig. 3.19(II)) with a local optimum for peak frequencies around 300 Hz. The effect of the peak frequency f_{peak} (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1.360, 44.869) = 8.458, p = 0.003$) and the effect of the frequency ratio ρ are statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1, 33) = 43.414, p = 0.000$). The interaction effect is statistically not significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1.546, 51.012) = 1.361, p = 0.262$).

Sound character (I)

For all slope combinations, the level difference $\Delta L_{\text{sound character}}$ becomes monotonously more negative with increasing peak frequency, meaning that that the difference in terms of sound character compared to the reference increases. A possible explanation for this is the increasing high frequency content with rising peak frequency. For a rising slope of $\text{slope}_{\text{UP}} = 6$ dB/oct) and a falling upper slope $\text{slope}_{\text{DN}} = -6$ dB/oct (shown in Fig. 3.18(I)b), the effect of f_{peak} is statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.611, 86.157) =$

$4.885, p = 0.005$). Additionally, a rather constant offset in the level difference attributed to the sound character is found between the two values of the frequency ratio ρ which is also statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1,33) = 9.726, p = 0.004$). The interaction effect of the two factors ($\rho * f_{\text{peak}}$) is statistically not significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.849, 94.005) = 0.897, p = 0.442$).

Sound character (II)

For a falling upper slope of $\text{slope}_{\text{DN}} = -15$ dB/oct (shown in Fig. 3.19(II)b), the level difference attributed to the sound character $\Delta L_{\text{sound character}}$ gets monotonously more negative for a rise in peak frequency from $f_{\text{peak}} = 100$ Hz to $f_{\text{peak}} = 500$ Hz by about 8 dB. This effect of the peak frequency is significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(3.001, 99.047) = 21.366, p = 0.000$). Additionally the sound character difference is significantly more negative for a frequency ratio $\rho = 4:3$ than for $\rho = 199:150$ (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1,33) = 12.524, p = 0.001$). A non-significant interaction effect (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.588, 85.414) = 0.330, p = 0.774$) indicates that the curves for the two values of the frequency ratio ρ are offset by a constant shift of about 3 dB.

Sound character (III)

For the steep rising and falling slope (shown in Fig. 3.19(III)b), the level difference $\Delta L_{\text{sound character}}$ becomes monotonously more negative with increasing peak frequency f_{peak} which is statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1.523, 50.269) = 30.622, p = 0.000$) like for the other two slope configurations (I and II). A rather constant offset between the results for the two values of ρ is supported by a statistically significant effect of the frequency ratio ρ (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1,33) = 21.229, p = 0.000$) and no significant interaction effect between the two parameters (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1.668, 55.037) = 0.760, p = 0.450$).

3.4.5.4 Key findings

— Sounds with a shallow falling upper slope of $\text{slope}_{\text{DN}} = -6$ dB/oct would be considerably more unpleasant at equal dB(A)-levels and, thus, require lower levels to become equally preferred than corresponding sounds with a steep falling upper slope of $\text{slope}_{\text{DN}} = -15$ dB/oct.

— For the sounds with a steep falling upper slope, local maxima in the PSEs for loudness and for preference are found at peak frequencies between 200 Hz and 400 Hz, which reflect a local optimum in terms of highest A-weighted levels while remaining equally loud/preferred compared to lower or higher peak frequencies.

— The level difference attributed to the sound character becomes monotonously more negative with increasing peak frequency for all slope configurations investigated. The rate of change is slightly more pronounced for the sounds with the steep upper slope of $\text{slope}_{\text{DN}} = -15$ dB/oct than for those with the more shallow upper slope of $\text{slope}_{\text{DN}} = -6$ dB/oct.

3.4.6 Relevance of frequency ranges for the judgments and evaluations

3.4.6.1 Stimuli

To identify the relevance of frequency ranges to the overall judgments, the amplitudes of the partials within three different frequency ranges were attenuated by 10 dB, based on two values of the frequency ratio ρ between the fundamental frequencies. The frequency ranges are defined as:

— Low: only the low frequency content up to 500 Hz was attenuated by 10 dB

— Mid: only the mid frequency content between 500 Hz and 2 kHz was attenuated by 10 dB

— High: only the high frequency content above 2 kHz was attenuated by 10 dB

Table 3.10 gives an overview of the stimulus conditions.

Table 3.10: *Parameters of the six multi-tone test stimuli with attenuated frequency ranges and the two baseline configurations (D and E) without attenuations.*

ID	slope_{UP} (dB/oct)	slope_{DN} (dB/oct)	f_{peak} (Hz)	ratio $\rho = \frac{f_{01}}{f_{10}}$	basic multitone elements	attenuation δ (dB)	freq. range attenuated by 10 dB
D11	—	-6	100	199:150	CX1+CX2+CTs	-10	low
E11	—	-6	100	4:3	CX1+CX2+CTs	-10	low
D12	—	-6	100	199:150	CX1+CX2+CTs	-10	mid
E12	—	-6	100	4:3	CX1+CX2+CTs	-10	mid
D13	—	-6	100	199:150	CX1+CX2+CTs	-10	high
E13	—	-6	100	4:3	CX1+CX2+CTs	-10	high
D	—	-6	100	199:150	CX1+CX2+CTs	-10	—
E	—	-6	100	4:3	CX1+CX2+CTs	-10	—

3.4.6.2 Participants

The experiments were carried out by 47 paid volunteers (23 female, 24 male). The mean age of the participants was 23 years (min=19 years, max=31 years). Approximately one half of the participants already participated in other previous psychoacoustic experiments (10 female, 13 male), the other half was inexperienced with psychoacoustic tests (13 female, 11 male). All participants reported no hearing difficulties.

3.4.6.3 Results

Figure 3.20 shows the mean PSEs for loudness and for preference and the level difference attributed to differences in the sound character for an attenuation of the low (below 500 Hz), the mid (between 500 Hz and 2 KHz) and the high frequencies (above 2 KHz). In addition the results of the respective sounds without attenuated frequency ranges are shown as a baseline on the right side of the plots.

Loudness
For both values of the frequency ratio ρ, the PSEs for loudness have a local minimum for an attenuation of the mid frequencies (between 500 Hz and 2 kHz) and a local maximum for an attenuation of the high frequencies above 2 kHz. This effect of the attenuated frequency range is statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.411, 110.911) = 50.057, p < 0.001$). The attenuation of high frequencies renders the multi-tone stimuli softer than the two other attenuated frequency ranges and the baseline configuration, allowing higher A-weighted levels for while remaining equally loud. A constant offset between the PSEs for the two values of the frequency ratio ρ is supported by a statistically significant effect of the ratio ρ (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1, 46) = 133.846, p < 0.001$, while the interaction between attenuation ranges and ρ is not significant (ANOVA, rep. meas., Greenhouse-Geisser corrected, loudness: $F(2.908, 133.749) = 1.579, p = 0.199$).

Preference
Similarly like for the loudness judgments, the PSEs for preference also have a local maximum for an attenuation of the high frequencies above 2 kHz, which is statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.264, 104.122) = 11.967, p = 0.000$). The attenuation of the high frequencies renders the multi-tone sounds more pleasant than the baseline and the other attenuated frequency ranges and allows higher A-weighted levels while remaining equally preferred. The attenuation of the low frequency range (below 500 Hz) yields lower PSEs than the other two conditions and also the baseline configuration. This means

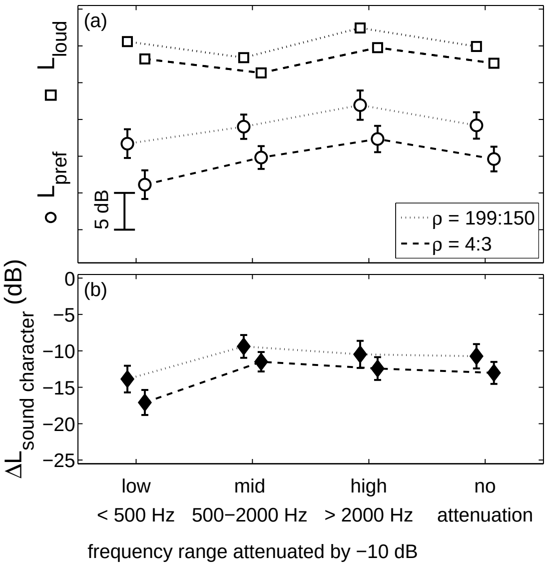


Figure 3.20: *Mean PSEs for loudness (a, open squares) and for preference (a, open circles), each plotted as relative values, and the level difference attributed to the sound character (b, filled diamonds) plotted for three conditions with attenuation of frequency ranges by 10 dB and without attenuation (baseline condition). Frequency ratios between the fundamentals are indicated by dotted ($\rho = 199 : 150$) and dashed ($\rho = 4 : 3$) lines. Errorbars indicate the standard deviation of the calculated mean ($N = 47$, sometimes even smaller than the symbols).*

that an attenuation of the low frequencies would be less preferred than any of the other conditions when having the same dB(A)-level. A statistically significant influence of the frequency ratio ρ (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1, 33) = 23.796, p = 0.000$) and no significant interaction effect (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.519, 115.867) = 0.775, p = 0.490$) speak for a constant offset between the PSEs for the two values of the frequency ratio ρ.

Sound character

In terms of the level difference attributed to the sound character, an attenuation of the low frequencies yields a considerably more negative level difference $\Delta L_{sound\ character}$ than a reduction of the mid or the high frequencies, which yield values on the same level as the baseline condition, without any attenuated frequency range. This effect of the frequency ranges is statistically significant (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.267, 104.303) = 11.793, p = 0.000$). A statistically significant influence of the frequency ratio ρ (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(1, 46) = 6.681, p = 0.013$) and no significant interaction effect (ANOVA, rep. meas., Greenhouse-Geisser corrected: $F(2.671, 122.880) = 0.719, p = 0.527$) support a constant offset between the results for the two values of ρ.

3.4.6.4 Key findings

— An attenuation of the low frequency range (below 500 Hz) is similarly loud as the baseline configuration but less preferred, requiring about 3 dB lower levels at the PSE for preference than the baseline condition.

— An attenuation of the mid frequency range (between 500 Hz and 2 kHz) is comparable to the baseline condition in terms of loudness and preference.

— An attenuation of the high frequencies range (above 2 kHz) allows about 3 dB higher levels at the PSEs for loudness and also for preference compared to the baseline condition.

3.5 Conclusion

The preference and the loudness of 55 multi-tone test sounds were measured as levels at the points of subjective equality (PSEs), in the context of aircraft cabin interior noise. The PSEs for loudness (L_{loud}) and preference (L_{pref}) were determined in comparison to a fixed reference noise, based only on the reasonable assumption that loudness and unpleasantness increase with rising dB(A)-level. The difference between the preference and the loudness PSEs was ascribed to the particular multi-tone sound character, quantified as $\Delta L_{\text{sound character}} = L_{\text{pref}} - L_{\text{loud}}$.

In general, all test sounds yielded lower levels for the preference PSEs than for the loudness PSEs, resulting in negative values for the level difference $\Delta L_{\text{sound character}}$. This means that, for all stimulus configurations, the adjustment to equal loudness was not sufficient, and additional level reductions were required, to make the multi-tone sounds equally preferred as the rather broadband reference sound. It could be shown that the level difference $\Delta L_{\text{sound character}}$ was closely linked to the raw preference evaluations (L_{pref}) and only weakly related to the loudness judgments (L_{loud}). Thus, the differentiated loudness judgments and preference evaluations, which were delivered by the participants, allow a quantitative distinction between the two tasks. The standard deviations of the calculated mean values were smaller than 1 dB for the loudness judgments and smaller than 2.5 dB for the preference evaluations. In addition, no statistically significant differences were found between three disjunct groups of participants in terms of the mean preference PSEs and also for $\Delta L_{\text{sound character}}$, which further supports the reliability and reproducibility of the obtained results.

In comparison to the most directly applicable literature, the inter-individual standard deviations observed in the present study are of similar magnitude as in the data of Pollack[(186)] for bandpass filtered noise, in the data Niese[(170)] for howling noise (German: "Heulrauschen") and in the data from Ronnebaum and Weber for pure tones and narrow-band noise[(196)]. However, the overall extent of the level difference $\Delta L_{\text{sound character}}$ between the preference and the loudness judgments, presented in this chapter, is considerably larger than in the results found in the literature. The multi-tone signals investigated here exhibit considerable differences in terms of the sound character compared to the reference sound and also in between different multi-tone configurations. Possible explanations are the particular multi-tone signals investigated in the current study, which differ from the noise like signals in the studies of Pollack, Niese and Ronnebaum and Weber. Further differences are the particular tasks (annoyance or unpleasantness in the literature, preference in the current study) and contexts in which the judgments took place.

The biggest effects of the experimentally varied signal parameters on the loudness judgments and the preference evaluations were found for a variation of the spectral envelope towards high frequencies. The steeper the overall decrease of tonal energy towards high frequencies, the higher the PSEs for loudness and also for preference. Furthermore, differences among the multi-tone signals of up to 10 dB were found between the extreme slope configurations with regard to the level difference $\Delta L_{\text{sound character}}$.

The variation of the frequency ratio ρ between the fundamental frequencies, which was also varied in the semantic differential, led to rather small differences between the multi-tone sounds of only about 2 dB for the loudness PSEs, 5 dB for the preference PSEs, and 3 dB for $\Delta L_{\text{sound character}}$. The complex tones alone, which were a basic element of the composed multi-tone sounds, required the lowest PSE levels in the preference task and also the most negative values for the level difference $\Delta L_{\text{sound character}}$ (of about -20 dB) which means that these sounds would be the least preferred sounds when equally loud as the other multi-tone sounds.

Apparently, the complex relationships between signal parameter variations and the preference evaluations stem from an interference with the loudness judgments, because the individual subtraction of the loudness PSEs from the preference evaluations reduced the complexity considerably. The individual level difference $\Delta L_{\text{sound character}}$ yielded an overall clearer picture of the results with generally more monotonous or rather linear relationships with the parameter variations. These simpler relationships suggest that the level difference $\Delta L_{\text{sound character}}$ might be a more useful basis for a prediction model than the raw PSEs for loudness or preference and their rather complex relationships with parameter variations.

4 Description and prediction of the judgments and evaluations for multi-tone sounds by (psycho-)acoustic descriptors

For the assessment of technical sounds in an early development stage, prediction models are often desirable, in order to estimate the impact of changes in terms of technical design parameters on judgments of the resulting sounds by humans. Such models are usually based on a regression analysis between a number of acoustic descriptor variables and subjective judgments in the form of listening test data. The challenge in the development of such models lies in the identification of acoustic descriptor variables, which share a considerable amount of variance with the judgments and at the same time can be assumed to be related to the perceived sound characteristics. Ideally, a high number of varied signal parameters is summarized by a small number of suitable descriptor variables, effectively depicting the characteristics of the particular sounds, which are of interest. Taking the limitations of such models in terms of their generalizability for granted, they can help in the optimization of prospective sounds by disclosing the relationships between descriptor variables and judgments for a specific context and within a restricted range of validity.

In the semantic differential test of Chap. 2, the repetition rate of the multi-tone signals was identified as a consistent descriptor of the periodicity, resulting from a variation of the frequency ratio ρ between the two fundamental frequencies, which was related to the pleasantness of the sounds. For the overall 55 stimuli of the PSE measurements (described in Chap. 3), resulting from a broader range of varied signal parameters, further potential descriptors need to be identified to cover the extended variety of the sound characteristics, especially in terms of the composition of the spectral content. The majority of the test sounds in the PSE measurements were derived from the most pleasant and the most unpleasant values of the frequency ratio ρ determined in the semantic differential test. Therefore, the repetition rate will also be tested as a potential descriptor variable of the PSE data.

In the following, Sec. 4.1 presents the relationships between common psychoacoustic descriptor variables (which were introduced in Sec. 1.4.3) and the three subjective PSE measures (L_{loud}, L_{pref} and $\Delta L_{\mathrm{sound\ character}}$), obtained in the listening tests of the previous chapter 3, which cover a wide range of signal parameter variations. Based on the relationships between the psychoacoustic descriptors and the judgments from the PSE measurements a prediction model for $\Delta L_{\mathrm{sound\ character}}$ and for the preference PSEs is set up, each. The selection of the descriptors, their combination into the prediction models and an estimation of the prediction error is described in Sec. 4.2.

4.1 Relationships between judgments, evaluations and psychoacoustic descriptors

In this section, the results of the listening tests for the loudness PSEs L_{loud}, for the preference PSEs L_{pref} and the level difference attributed to the sound character $\Delta L_{\mathrm{sound\ character}}$ are each related to a set of common psychoacoustic descriptors. The relationships between the judgments and the descriptor values are examined separately for the two values of the frequency ratio ρ, which resulted in the most pleasant ($\rho = 199 : 150$) and the most unpleasant sound ($\rho = 4 : 3$) in the semantic differential experiment.

4.1.1 Relationships between loudness judgments and outputs of loudness models as well as other psychoacoustic descriptors

The loudness of the sounds was determined in the listening tests as the point of subjective equality (PSEs) compared to a reference sound. As a result, all data points are judged to be equally loud as the reference sound and, thus, also equally loud among each other. Figure 4.1 shows

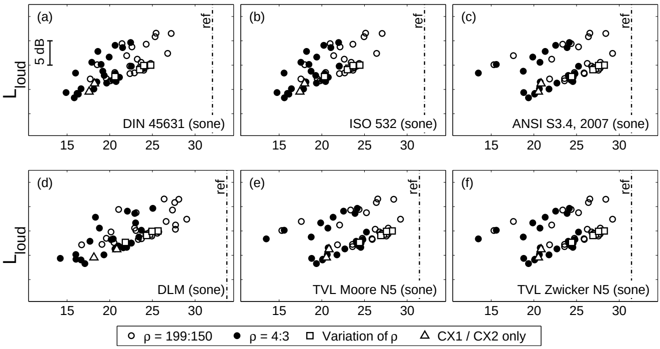


Figure 4.1: *Mean values of the PSEs for loudness L_{loud} of the 55 multi-tone sounds plotted over the calculated loudness values by six different metrics (a, b, c for stationary sounds, d, e, f for time varying sounds). Markers indicate the value of the corresponding frequency ratio ρ. The calculated loudness of the constant reference sound is indicated by the vertical dash-dotted line in the plot of each metric. No loudness metric reasonably reflects the equal loudness judgments.*

the mean values of the PSEs for loudness L_{loud} plotted over the respective loudness values, calculated with six different loudness metrics. If the loudness metrics would reflect the loudness judgment well, then the data points should vary in ordinate (different levels at the PSE for loudness) but be close together in abscissa (calculated loudness values) in the vicinity of the reference sound (vertical dash-dotted line). In contrast, all loudness metrics indicate lower loudness values for all of the multi-tone test sounds than for the reference sound and the judged equal loudness with the reference sound (indicated by the vertical line) is not reflected by any metric. In addition, a considerable spread in the calculated loudness values (abscissa) can be seen for all six loudness metrics which means that the judged equal loudness among the multi-tone sounds is also not well reflected by any of the six loudness metrics.

An explanation for the failure of the tested loudness metrics in terms of a correct representation of the equal loudness judgments might be the close relationship between the loudness judgments and the preference evaluations, which was discussed in more detail in Sec. 3.2. Even though the participants are instructed to judge the loudness of the sounds, it

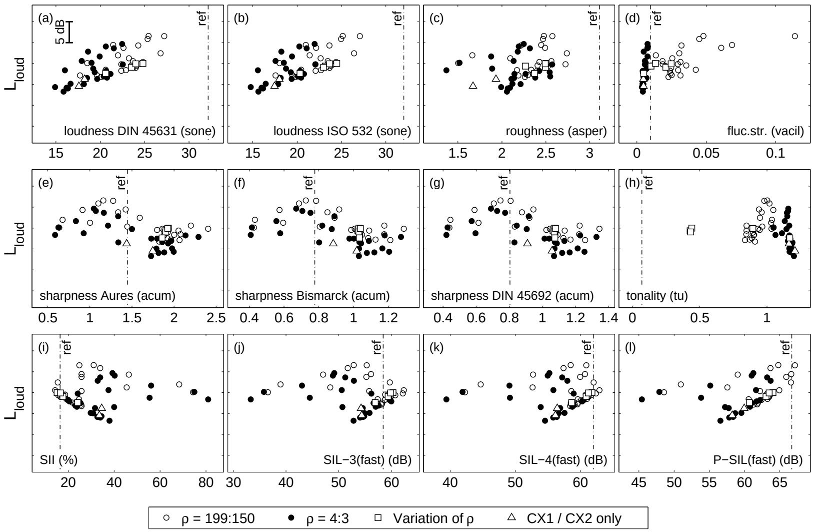


Figure 4.2: *Mean values of the PSEs for loudness of 55 sounds plotted over the values of 12 psychoacoustic descriptors calculated for the respective PSE levels L_{loud}. Markers indicate the value of the corresponding frequency ratio ρ. The value of the reference sound is indicated by the vertical dash-dotted line in the plot of each metric.*

cannot be totally excluded that their judgments are influenced by the rather unpleasant sound character of the multi-tone sounds which in turn causes a preference bias towards the reference sound. Such a preference bias towards the reference, would lead to measurement results below the actual point of equal loudness to partially compensate for the unpleasant multi-tone sound character. The resulting lower levels at the PSEs for loudness then yield lower calculated loudness values by the loudness algorithms. This potential bias in the loudness PSEs, which partially includes the unpleasantness of the multi-tone sounds, is of course not correctly reflected by the loudness models.

The relationship between the PSEs for loudness L_{loud} and 10 further psychoacoustic descriptors is shown in Fig. 4.2. The two included loudness metrics (DIN 45631, ISO 532) are exactly the same as in Fig. 4.1 (a) and (b). For the three sharpness metrics (Fig. 4.2, subplot (e), (f) and (g)) a peak in the PSEs for loudness L_{loud} is apparent around 1.3 acum (Aures method, e) and around 0.7 acum (von Bismarck method and DIN 45692 standard, f and g). This is also about the sharpness value of the reference sound which is indicated by the dashed-dotted line. This means that sounds having a sharpness values of about this magnitude can have higher levels than other sounds with lower or higher sharpness values, while being judged to be equally loud as the reference sound. In this sense a sharpness value of about 0.7 acum (von Bismarck method, DIN 45692 standard) or about 1.3 acum (Aures method) represents an optimum by allowing higher A-weighted levels while maintaining equal loudness compared to the reference sound. For all other psychoacoustic descriptors, the relationships between judged loudness and descriptor values are not that clear and distinct links between the descriptors and the PSEs for loudness are not apparent, as would be expected.

4.1.2 Relationships between preference evaluations and psychoacoustic descriptors

The relationships between the mean values of the PSEs for preference L_{pref} and 12 psychoacoustic descriptors are shown in Fig. 4.3. The correlation coefficients between each of the descriptors and the preference evaluations are given in Tab. 4.1, for the two values of the frequency ratio ρ, which were judged as most pleasant and most unpleasant in the semantic differential test, separately. For level dependent descriptors (e.g. the loudness metrics), the correlation coefficients are to some extent intrinsically given because of the quantification of the preference evaluations as levels at the point of subjective equality. Thus, care needs to be taken in the interpretation of these relationships.

The highest correlation coefficients are found for the two loudness metrics calculated according to the DIN and the ISO standard, followed directly by the three sharpness metrics. For the loudness metrics (shown in Fig. 4.3 (a) and (b)), high values of the PSEs for preference L_{pref} yield high calculated loudness values. This means that if the sound character of a multi-tone sound is judged to be rather pleasant, then this sound yields a rather high PSE for preference which is reflected by the loudness metrics as a rather high loudness value. If, on the other hand, a multi-tone sound is rather undesirable due to its unpleasant sound character, then it needs a considerable level reduction to become equally preferred as the reference and the resulting low A-weighted level at the PSE for

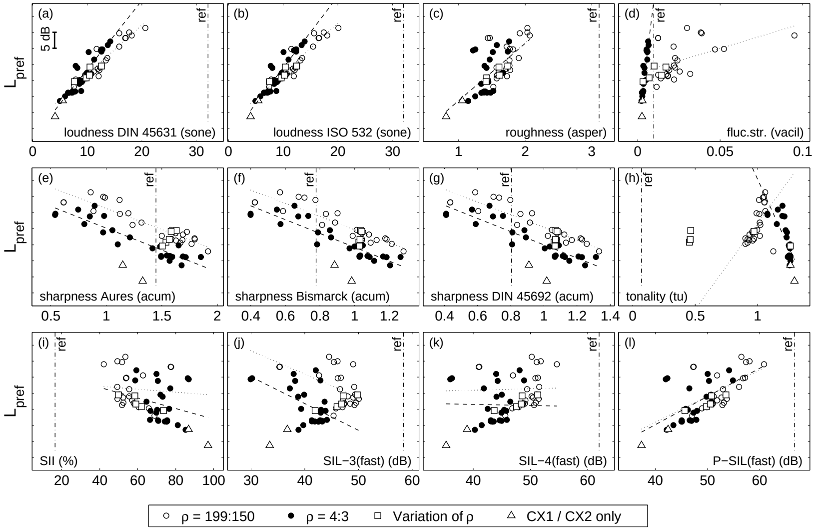


Figure 4.3: *Mean values of the PSEs for preference L_{pref} of 55 sounds plotted over the values of 12 different psychoacoustic descriptors calculated for a sound pressure level corresponding to the respective PSE for preference L_{pref}. Markers indicate the value of the corresponding frequency ratio ρ. The value of the reference sound is indicated by the vertical dash-dotted line in the plot of each metric.*

preference results in a low loudness value. In this sense, the found close relationship between the loudness metrics and the PSEs for preference seem to reflect the level dependency of the loudness metrics and should not be mistakenly interpreted as an increase in loudness being beneficial for the preference towards a sound.

In terms of the three sharpness metrics (shown in Fig. 4.3 (e), (f) and (g)), a decrease of the PSEs for preference L_{pref} for increasing sharpness values is observed in general. In the case of the level independent sharpness metrics (according to von Bismarck's method and the DIN 45692 standard), the statistically significant correlation coefficients between the sharpness metrics and the PSEs for preference can not be traced back to the level dependencies of the descriptors which suggests "true" relationships between the preference evaluations and the metrics.

Table 4.1: *Pearson correlation coefficients and p-values between 12 psychoacoustic descriptor variables and the PSEs for preference, separated for the two values of the frequency ratio ρ. The psychoacoustic descriptor values were calculated with a sound pressure level equal to the PSE for preference. The characters (a) to (l) resemble those used for the subplots of Fig. 4.3. The correlation coefficients reflect the linear relationships indicated by the dotted ($\rho = 199 : 150$) and the dashed ($\rho = 4 : 3$) lines in that figure.*

		$\rho = 199 : 150$		$\rho = 4 : 3$	
	descriptor	Pearson's r	p-value	Pearson's r	p-value
(a)	loudness DIN 45631	0.93	<0.001	0.93	<0.001
(b)	loudness ISO 532	0.93	<0.001	0.92	<0.001
(c)	roughness	0.59	0.002	0.45	0.023
(d)	fluctuation strength	0.50	0.011	0.93	<0.001
(e)	sharpness Aures	-0.92	<0.001	-0.92	<0.001
(f)	sharpness Bismarck	-0.91	<0.001	-0.92	<0.001
(g)	sharpness DIN 45692	-0.91	<0.001	-0.92	<0.001
(h)	tonality	0.57	0.003	-0.78	<0.001
(i)	SII	-0.07	0.737	-0.20	0.335
(j)	SIL-3 (fast)	-0.49	0.012	-0.53	0.006
(k)	SIL-4 (fast)	0.03	0.903	-0.02	0.933
(l)	P-SIL (fast)	0.53	0.006	0.51	0.009

The relationship between the PSEs for preference L_{pref} and the sharpness according to the DIN 45692 standard is shown in Fig. 4.4 in more detail. In this figure, the variation of the frequency ratio ρ between the fundamental frequencies of the multi-tone sounds is highlighted by square symbols. The variation of the frequency ratio leads to rather constant values for the sharpness metric even though the amplitude spectra of the signals differ from each other in detail (see Fig. 2.3(I) and Fig. 2.3(II) in section 2.1.2). The changes in the temporal structure induced by a variation of the frequency ratio ρ do not strongly co-vary with the sharpness compared to variability in the sharpness values for the overall 55 test sounds. A rather constant offset between the preference PSEs for $\rho = 199 : 150$ and $\rho = 4 : 3$, which were also found to be related to separate sound groups in the semantic differential test, can be observed. Moreover, for each of the two values of the ratio ρ, significant correlation coefficients indicate linear relationships with the sharpness metric, each. However, the two sounds consisting of a single complex

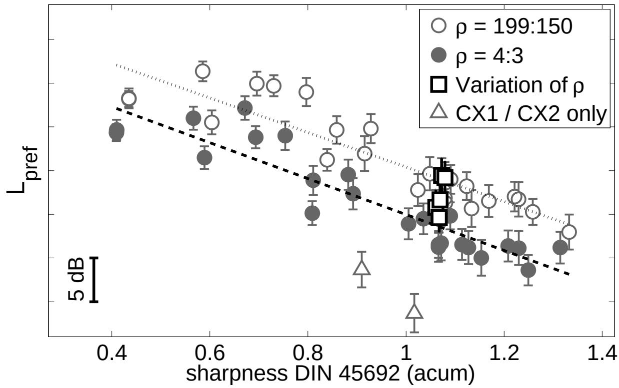


Figure 4.4: *Mean values of the PSE for preference L_{pref} of 55 sounds plotted over the sharpness values according to the DIN 45692 standard. Errorbars indicate the standard error of the calculated mean values. Square symbols indicate the variation of the frequency ratio ρ which leads only to slight differences in the sharpness values. Linear regression lines are given for $\rho = 199:150$ (dashed line, open circles) and $\rho = 4:3$ (solid line, filled symbols). Values of the two sounds consisting of the complex tones only, which are not well described by the regression lines, are shown as open triangles.*

tone each (CX1 and CX2 only) are not well covered by the overall linear trend between the preference PSEs and the sharpness values.

Besides the overall linear trend, indicated by the significant correlation coefficients for the three sharpness metrics (Fig. 4.3 (e), (f) and (g)), the PSEs for preference peak at about 0.8 acum (Aures method) and between 0.5 acum and 0.8 acum (von Bismarck method, DIN 45692 standard). The maxima observed here are not as peaky as those found for the loudness judgments (shown in Fig. 4.2 (e), (f) and (g)), but their position in terms of the sharpness values is essentially the same. Multi-tone sounds with sharpness values in this range can have a higher A-weighted level than other sounds with lower or higher sharpness values, while being equally preferred as the reference sound. In this sense these peaks reflect an optimum in terms of preference for the multi-tone sounds.

4.1.3 Relationships between the level difference attributed to the sound character and psychoacoustic descriptors

The level difference attributed to the sound character $\Delta L_{\text{sound character}}$, is not bound to an absolute sound pressure level due to the calculation as the difference between the PSE for preference and the PSE for loudness. Thus, for level dependent acoustic descriptors, the level basis for the calculation is not defined and only level independent descriptors can be meaningfully related to the level difference[1]. The sharpness calculated according to the DIN 45692 standard is level independent because of the loudness independent weighting function and the loudness normalization in the algorithm. In addition, out of the 12 psychoacoustic descriptors, which were tested as potential descriptors of the level difference, the highest correlation coefficients with the level difference are obtained for the DIN sharpness (Tab. A.3 in the appendix). The relationship between the level difference $\Delta L_{\text{sound character}}$ attributed to the sound character and the level independent DIN sharpness, is shown in Fig. 4.5.

In general, the level difference $\Delta L_{\text{sound character}}$ becomes more negative with increasing sharpness values which means that the difference between the PSEs for preference and the PSEs for loudness increases with a rise in sharpness values. A statistically significant correlation coefficient of $r = -0.91$ is reached for $\rho = 199 : 150$ and $r = -0.94$ for $\rho = 4 : 3$. Even though the correlation coefficients found here are of similar magnitude as those found for the preference PSEs (in Tab. 4.1), the small residuals between the linear regression lines indicated by the dotted and the dashed lines in Fig. 4.5 and the measured data are striking.

The maxima which were found for sharpness values in the range between 0.5 acum and 0.8 acum for the loudness PSEs (in Fig. 4.1, e and f) and similarly for the preference PSEs (in Fig. 4.4) are not apparent in the current case for the level difference $\Delta L_{\text{sound character}}$. The absence of this peak for the level difference $\Delta L_{\text{sound character}}$ suggests that it originates from a common latent feature underlying the PSEs for loudness and the

[1] For the sake of completeness, the relationships between the level difference attributed to the sound character and the 12 descriptors (including level dependent ones) which were also tested as potential descriptors for the loudness and the preference PSEs can be found in the appendix A.6

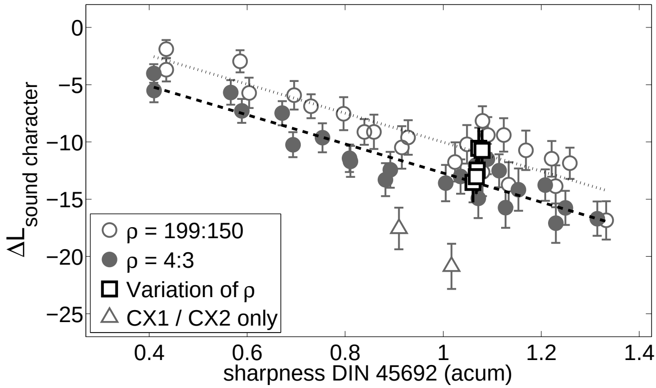


Figure 4.5: *Mean values of $\Delta L_{\text{sound character}}$ for 55 sounds plotted over the sharpness values according to the DIN 45692 standard. Errorbars indicate the standard error of the calculated mean values. Square symbols indicate the variation of the frequency ratio ρ which leads only to slight differences in sharpness values. Linear regression lines are given for $\rho = 199 : 150$ (dashed line, open circles) and $\rho = 4 : 3$ (solid line, filled symbols). Values of the two sounds consisting of the complex tones only (CX1, CX2), which are not covered well by the regression lines, are shown as open triangles.*

PSEs for preference in the same way. The tightness of the peak, which was more distinct for the loudness PSEs than for the preference PSEs might be an indication that this common latent feature is more closely related to the loudness judgment then to the preference evaluation.

Variations of the frequency ratio ρ between the fundamentals, resulting in differences of the temporal structure which can be described by the repetition rate of the time signals, remain rather constant in terms of sharpness at values of about 1 acum (square symbols in Fig. 4.5). This was also similarly found for the stimuli used in the semantic differential which covered 15 different frequency ratios, shown in Fig. 2.8. The interference of matching partials, which occurs for values of the frequency ratio ρ equal to ratios of small integer numbers, barely affects the sharpness values. However, the two test sounds CX1 and CX2, each consisting of one complex tone only, have sharpness values of about 0.9 acum (CX1)

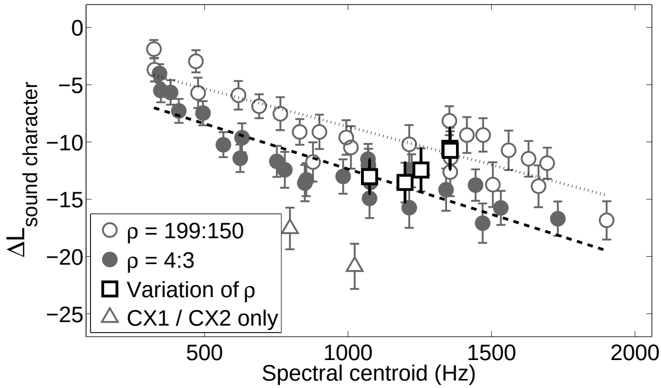


Figure 4.6: *Mean values of $\Delta L_{\text{sound character}}$ for 55 sounds plotted over the values of the spectral centroid. Errorbars indicate the standard error of the calculated mean values. Square symbols indicate the variation of the frequency ratio ρ which yields considerable variability in terms of spectral centroid values. Linear regression lines are given for $\rho = 199 : 150$ (dashed line, open circles) and $\rho = 4 : 3$ (solid line, filled symbols). A co-variation between the frequency ratio ρ (indicated by open squares) and the sharpness values becomes apparent.*

and 1 acum (CX2) which are in the same medium range of sharpness values as the complete multi-tone sounds (CX1+CX2+CTs, indicated as square symbols in Fig. 4.5). Due the exceptionally low PSEs for preference, which were determined for these two test sounds, also the most negative values for $\Delta L_{\text{sound character}}$ are reached by them. As a result, the values of the level difference $\Delta L_{\text{sound character}}$ for these two multi-tone sounds are not covered very well by the almost perfect linear relationships observed for the two values of the frequency ratio ρ, indicated by the dotted and the dashed line in Fig. 4.5.

Another potential level independent descriptor for the spectral content of a signal is the spectral centroid, which is more commonly used for the description of musical timbre. The relationship between the level difference $\Delta L_{\text{sound character}}$ and the spectral centroid is shown in Fig. 4.6[2]. In comparison to the sharpness, shown in Fig. 4.5, the level difference $\Delta L_{\text{sound character}}$ is not decreasing as monotonously with increasing values of the spectral centroid as it does with the sharpness, yielding slightly lower correlation coefficients of $r = -0.88$ (for $\rho = 199 : 150$) and $r = -0.89$ (for $\rho = 4 : 3$). In addition to this, the spectral centroid is stronger influenced by the detailed spectral composition of the multi-tone sounds than the sharpness. This results in a considerable spread in the spectral centroid values for a variation of the frequency ratio ρ. Thus, in the following, the sharpness according to the DIN standard will be used to characterize the spectral characteristics of the multi-tone sounds in the prediction model.

4.1.4 Description of the influences of the temporal structure on the judgments and evaluations by the signal's repetition rate

In the semantic differential test, it could be shown that a description of the temporal periodicity by the repetition rate of the signals' time series is highly correlated to the pleasant dimension of the perceptual space. In the following, it is tested whether the repetition rate is also suitable for the description of the results gained in the PSE measurements, especially for the two sounds CX1 and CX2, which were not included in the semantic differential test, and which have repetition rates of at least one order higher than the complete multi tone sounds (CX1+CX2+CTs).

The periodic time of the multi-tone signals T_{rep} and the resulting repetition rates f_{rep} can be obtained by the lag of the first unity peak in the auto-correlation function of the signals' time series, which is shown in Fig. 4.7. For the particular multi-tone signals investigated here, which consist of tonal components only, the repetition rate can also be directly calculated from the integer ratio of the underlying fundamental frequencies and the absolute values of the fundamental frequencies[3]. For the two

[2] The spectral centroid is calculated with the implementation of the MIR-toolbox [(136)].

[3] Details on the calculation are given in Sec. 2.3.2.

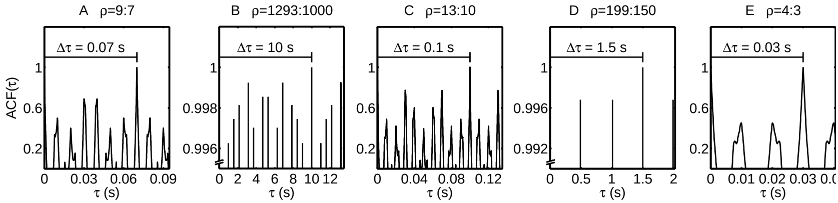


Figure 4.7: *Auto-correlation function (ACF) as a function of signal lag for the five multi-tone signals A - E. The periodic times T_{rep} can be clearly identified by the lag $\Delta\tau$ of the first unity peak. The found periodic times range over four magnitudes from 0.03 seconds up to 10 seconds.*

sounds F and G, each consisting of one complex tone only, the repetition rate is directly determined by the underlying fundamental frequency, each.

Table 4.2 summarizes the periodic times T_{rep} and the repetition rates f_{rep} for the five complete multi-tone sounds (CX1+CX2+CTs), for which the frequency ratio ρ between the fundamental frequencies was varied, and the two sound sounds F and G. A variation of the frequency ratio ρ yields periodic times T_{rep} ranging over four orders of magnitudes from 10 seconds (sound B) down to 0.03 seconds (sound E). For the two sounds F (CX1 only) and G (CX2 only) the periodicity is even shorter, going down to 0.0076 seconds for the sound with the higher fundamental frequency (G, CX2). Thus, the resulting repetition rates f_{rep}, which are the inverse of the periodic times T_{rep}, are one order of magnitude higher for the sounds F and G compared to the other multi-tone compositions consisting of two complex tones and combination tones.

Figure 4.8 shows the PSEs for loudness L_{loud}, the PSEs for preference L_{pref} and the level difference attributed to the sound character $\Delta L_{\mathrm{sound\ character}}$ plotted over the repetition rates for the seven sounds described in Tab. 4.2. All three judgments decrease with an increase of the repetition rate. This means that lower levels are necessary for sounds with high repetition rates to render them equally loud and equally preferred as sounds with low repetition rates. Additionally also the difference between the preference and the loudness PSEs, $\Delta L_{\mathrm{sound\ character}}$, becomes more negative with increasing repetition rate. Surprisingly, lin-

Table 4.2: *Properties of the seven stimuli varying mainly in signal periodicity, sorted by repetition rate f_{rep}: ID of the sounds, integer frequency ratio between the fundamental frequencies ρ, constituting sound elements, periodic time T_{rep} and repetition rate f_{rep}. Values of ρ equal to ratios of small integers lead to small periodic times and large repetition rates.*

ID	ratio $\rho = \frac{f_{01}}{f_{10}}$	basic multitone elements	T_{rep} (s)	f_{rep} (Hz)
B	1293:1000	CX1+CX2+CTs	10.0	0.1
D	199:150	CX1+CX2+CTs	1.5	0.67
C	13:10	CX1+CX2+CTs	0.10	10.0
A	9:7	CX1+CX2+CTs	0.07	14.29
E	4:3	CX1+CX2+CTs	0.03	33.33
F	199:150	CX1	0.01	100.0
G	199:150	CX2	0.0076	132.33

ear regressions between repetition rate f_{rep} and the three judgments indicated by the continuous lines do not only cover the five multi-tone compositions (A-E). Also the two complex tones F (CX1 only) and G (CX2 only) can be included seamlessly into the linear regressions, which suggests a usage of the repetition rate as a descriptor for the temporally periodic signature in the prediction model. The determination coefficients between the judgments and the repetition rate over all seven sounds varying in terms of periodicity are $r^2 = 0.96$ for L_{loud}, $r^2 = 0.98$ for L_{pref}, and $r^2 = 0.96$ for $\Delta L_{\text{sound character}}$.

4.1.5 Discussion of the found relationships between judgments, evaluations and psychoacoustic descriptors

Clear relationships were disclosed between two out of the tested various psychoacoustic descriptors and the judgments. The level independent sharpness according to the DIN 45692 standard and the repetition rate of the time signals each share a considerable amount of variance with the judgments, for a separate set of signal parameter variations each.

The variability in the judgments resulting from signal parameters which mainly influence the spectral content is best reflected by the sharpness metric. Statistically significant correlation coefficients between the

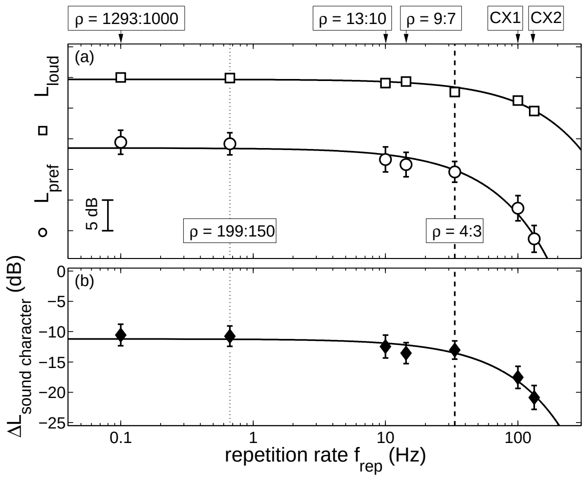


Figure 4.8: *Mean PSEs for loudness (a, open squares) and for preference (a, open circles) and the level difference attributed to the sound character (b, filled diamonds) for the five ratios between the fundamental frequencies ρ and the two complex tones alone (CX1, CX2) plotted over the calculated repetition rate f_{rep} (on a logarithmic axis). Dotted and dashed lines highlight the frequency ratios $\rho = 199 : 150$ and $\rho = 4 : 3$. Errorbars indicate the standard deviation of the calculated mean ($N = 47$, sometime even smaller than the symbols). Results of linear regressions are given as continuous lines, which are curved due to the logarithmic frequency axis.*

sharpness values and the PSEs for preference indicate an overall decrease of the PSEs for increasing sharpness values. This means that sounds with high sharpness values would be less preferred than sounds with low sharpness values, when having the same A-weighted level. This is in good agreement with relationships found in the literature. Zwicker found an increase in "unbiased annoyance" for an increase of the sharpness[(262)]. Similarly, Aures found a decrease of the so called "sensory euphony"[4] for an increase in relative sharpness[(13)] and Zwicker and Fastl found a decrease in sensory pleasantness for an increase in relative sharpness[(263)].

[4]German: Sensorischer Wohlklang

Upon closer inspection of the relationships, an optimum in terms of sharpness values in the range from 0.5 acum to 0.8 acum (calculated according to the DIN 45692 standard) is noticeable. The peak of the maximum is more pronounced for the loudness PSEs than for the preference PSEs. This optimum of the sharpness in terms of preference is in good agreement with the work of Eilers and Weber [61,250]. They found a dependency between the assessed preference and the sharpness and roughness of narrow-band noises. In their work a sharpness value between 0.6 acum (for degrees of modulation between $m = 0.2$ and $m = 0.6$) and 1 acum (for a degree of modulation $m = 1$) was optimal in terms of preference.

For a variation of those signal parameters that affect the temporal structure of the signals, statistically significant correlation coefficients are obtained between the repetition rate and each of the three judgments - the PSEs for loudness, the PSEs for preference and the level difference $\Delta L_{\text{sound character}}$. Sounds with a high repetition rate require considerably lower levels to be equally loud and also equally preferred than sounds with a low repetition rate and the effect increases for higher repetition rates. Furthermore, the level difference $\Delta L_{\text{sound character}}$ between the PSEs for preference and for loudness, which was attributed to the sound character, becomes more negative for an increase in repetition rate. This is in good agreement with the results of the semantic differential experiment (described in Sec. 2.3.2), where statistically significant correlation coefficients between the repetition rate and the factor values of the pleasantness factor were found.

This finding is also in overall agreement with a study of Hiramatsu *et al.* [110]. They found that fluctuation frequencies up to 5 Hz do not significantly influence annoyance judgments expressed as equivalent sound level for sinusoidally, saw-tooth and randomly modulated noise. For frequencies above 5 Hz they found a tendency of increased equivalent sound levels (equivalent to an increase in annoyance) for the saw-tooth modulated noise.

4.2 Prediction models

Two descriptor variables could be identified, which reasonably reflect the PSEs for preference L_{pref} and also the level difference attributed to the sound character $\Delta L_{\text{sound character}}$. Significant correlation coefficients were obtained between the sharpness (according to the DIN 45692 standard) and each of the judgments for a variation of those signal parameters that change the spectral composition of the multi-tone sounds. The variability in the judgments related to changes in the temporal structure is very well reflected by the repetition rate of the time signals, again for the preference PSEs and also the level difference $\Delta L_{\text{sound character}}$. It further turns out that changes in the temporal structure barely influence the sharpness of the multi-tone sounds. The sharpness values for a variation of those signal parameters that affect the temporal structure remain rather constant around 1 acum. Due to the resulting rather small amount of shared variance between the repetition rate and the sharpness, these two descriptor variables are used as the basis of the prediction models for $\Delta L_{\text{sound character}}$ and also for a model of the preference PSEs in the following.

4.2.1 Prediction model for the level difference attributed to the sound character

The relationship between the level difference attributed to the sound character $\Delta L_{\text{sound character}}$ and the values of the repetition rate f_{rep}, as a descriptor for the temporal signature, and the sharpness S, as descriptor of the spectral characteristics is shown in Fig. 4.9. In general, the level difference $\Delta L_{\text{sound character}}$ becomes monotonously more negative with increasing repetition rate f_{rep} and increasing sharpness S. A linear regression approach based on the two descriptors - repetition rate and sharpness - is given in equation 4.1 and the fitted parameters can be found in Tab. 4.3.

$$\Delta L_{\text{sound character}}(f_{\text{rep}}, S) = a + b \cdot f_{\text{rep}} + c \cdot S + d \cdot f_{\text{rep}} \cdot S \qquad (4.1)$$

The purely linear regression without an interaction term ($d = 0$) yields an adjusted determination coefficient of $r^2 = 0.89$ which means that 89 percent of the variance is explained ($F(2, 52) = 219.6, p \leq 0.001$). The re-

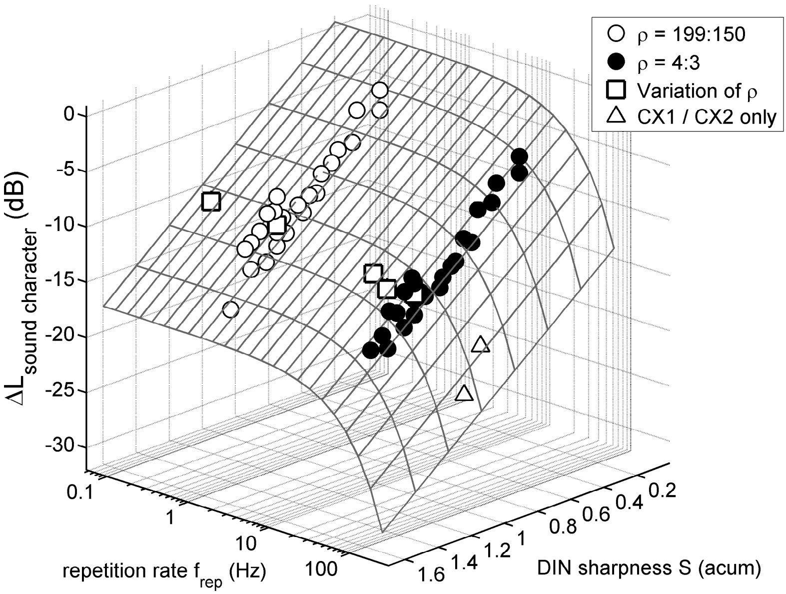


Figure 4.9: *Mean values of the level difference attributed to the sound character $\Delta L_{\text{sound character}}$ for the 55 multi-tone sounds plotted over their respective values for the repetition rate f_{rep} and the sharpness S (DIN 45692). Markers indicate the frequency ratio ρ between the fundamentals. A purely linear regression model (indicated by the mesh grid) yields an adjusted $r^2 = 0.89$.*

Table 4.3: *Fitted parameters a, b, c and d of a linear regression model described in equation 4.1 without ($d = 0$) and with an interaction term ($d \neq 0$) and the corresponding values of the adjusted r^2.*

model	a	b	c	d	adj. r^2
linear	3.0 dB	-0.079 $\frac{\text{dB}}{\text{Hz}}$	-12.9 $\frac{\text{dB}}{\text{acum}}$	—	0.89
linear + interaction	3.1 dB	-0.085 $\frac{\text{dB}}{\text{Hz}}$	-13.1 $\frac{\text{dB}}{\text{acum}}$	0.0063 $\frac{\text{dB}}{\text{Hz}\cdot\text{acum}}$	0.89

gression approach including an additional interaction term ($d \neq 0$) does not increase the amount of explained variance ($r^2 = 0.89,\ F(3, 51) = 143.7, p \leq 0.001$). Thus the inclusion of an interaction term is rejected because it does not yield a higher amount of explained variance. The purely linear regression model without the interaction term is indicated by the mesh grid in Fig. 4.9.

4.2.1.1 Cross validation

To obtain an estimate of the prediction error made by the regression model, different cross validations are carried out rather than directly evaluating the residuals of the regression[(11)]. The procedure of the cross validation is schematically visualized in Fig. 4.10.

The overall data consisting of the judgments and corresponding values of the descriptor variables for all 55 sounds is split into K parts of equal size. Instead of splitting the data into only one training and one test data set, each part of the data is used for testing and training. One of the K parts (identified by i_{Test}) is left out at a time, to test a model which is trained by all other parts (identified by i_{Train}). In figure 4.10 the first part ($i_{\text{Test}} = 1$) is exemplarily left out for testing, but the procedure is carried out K-times and each of the K parts is used as the test part.

The $K - 1$ training parts, identified by the indices i_{Train}, are used to fit the parameters a, b and c of the linear regression between the two predictor variables ($S(i_{\text{Train}})$ and $f_{\text{rep}}(i_{\text{Train}})$) and the values of $\Delta L_{\text{sound character}}(i_{\text{Train}})$. Based on this parameter set, predictions of the judgments are derived from the descriptor variables for the left out part ($S(i_{\text{Test}})$ and $f_{\text{rep}}(i_{\text{Test}})$). The predictions $\Delta L_{\text{sound character}}(S(i_{\text{Test}}), f_{\text{rep}}(i_{\text{Test}}))$ are then compared to the actually measured data $\Delta L_{\text{sound character}}(i_{\text{Test}})$.

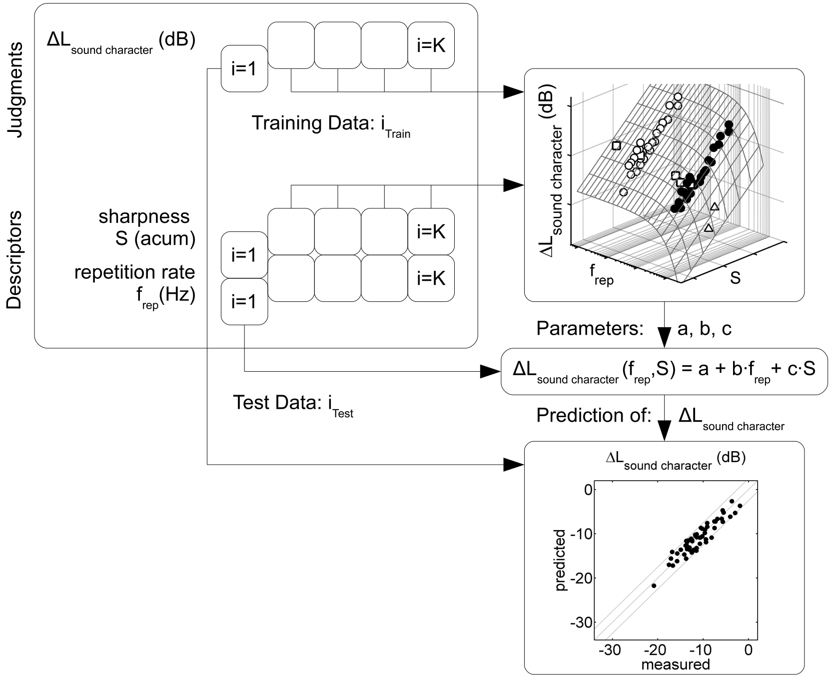


Figure 4.10: *Scheme of the K-fold cross validation for the estimation of the sound character difference* $\Delta L_{\text{sound character}}$.

Figure 4.11 shows the results of the 5-fold cross validation ($K = 5$) as a scatter plot of the predicted and the measured values for $\Delta L_{\text{sound character}}$ and a histogram of the difference between predicted and measured values. Detailed results of two further cross validations (leave one out (55-fold) and 11-fold), each showing similar results, can be found in the appendix A.7.

The resulting mean squared errors (MSEs) of three tested cross validations (given in Tab. 4.4) are between 1.75 dB^2 for the 55-fold (LOOCV) and 1.83 dB^2 for the 11-fold cross-validation. This common order of magnitude of the MSEs for the different amounts of validation data (from one left out data point up to 20% left out) supports the robustness of the prediction model. The prediction error, estimated as a root mean squared error (RMSE), is only about 1.3 dB, for a regression model, which is based on only three fitted parameters. This value is of the order of the standard error of the measured mean values for $\Delta L_{\text{sound character}}$ and also of similar magnitude as the just noticeable difference in level (JNDL) for broadband sounds[263].

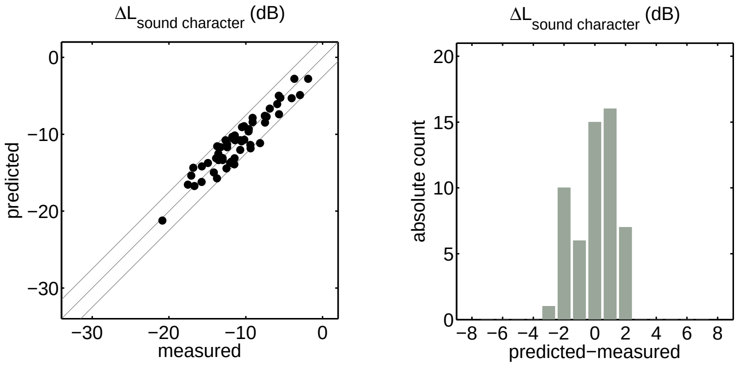


Figure 4.11: *Results of the 5-fold cross validation. Left: predicted values of $\Delta L_{\text{sound character}}$ plotted over the measured data. Continuous lines indicate identity ± 2.5 dB. Right: Histogram of the difference between predicted and measured values of $\Delta L_{\text{sound character}}$.*

Table 4.4: *Mean squared error (MSE) and root mean squared error (RMSE) determined by the three different K-fold cross-validations for the prediction of $\Delta L_{\text{sound character}}$*

K-fold	MSE	RMSE
55-fold (LOOCV)	1.75 dB^2	1.32 dB
11-fold	1.83 dB^2	1.35 dB
5-fold	1.80 dB^2	1.34 dB

4.2.2 Prediction model for the preference evaluations

The relationships between the psychoacoustic descriptors and the PSE for preference revealed a general decrease of L_{pref} for increasing values of the repetition rate f_{rep}, which originated from a variation of the temporal structure of the signals, and an overall decrease of L_{pref} for increasing values of the sharpness S (DIN 45692), which resulted from differences in the spectral composition. Therefore, a direct prediction of the PSEs for preference as a function of the sharpness and the repetition rate, similar to the prediction model for $\Delta L_{\mathrm{sound\ character}}$, is tested.

An analysis of the relationships between the judgments in Sec. 3.2 showed that the PSEs for preference and the level difference attributed to the sound character $\Delta L_{\mathrm{sound\ character}}$ are closely related to each other, which was reflected by a significant correlation coefficient of $r_{\mathrm{mean}} = 0.91$, meaning a shared variance of 83% between the two variables. Based on the prediction of $\Delta L_{\mathrm{sound\ character}}$ (presented in the preceding Sec. 4.2.1) with an estimated root mean squared error of only about 1.3 dB and the close relationship between the two, an indirect prediction of the PSE for preference from $\Delta L_{\mathrm{sound\ character}}$ seems also possible.

In the following, the direct prediction approach $L_{\mathrm{pref}}(f_{\mathrm{rep}}, S)$ is compared to the indirect approach $L_{\mathrm{pref}}(\Delta L_{\mathrm{sound\ character}}(f_{\mathrm{rep}}, S))$ which is based on an intermediate calculation of the level difference $\Delta L_{\mathrm{sound\ character}}$.

4.2.2.1 Direct prediction approach

The relationship between the mean values of the PSEs for preference L_{pref} on the one hand and the repetition rate f_{rep} and the sharpness S of the sounds on the other, are presented in Fig. 4.12. In general, the values for L_{pref} decreases for an increase of the repetition rate and also an increase of the sharpness values. This relationship is similar to that found for the sound character difference measure $\Delta L_{\mathrm{sound\ character}}$ (in Fig. 4.9), but the decrease progresses more steeply for L_{pref} than for $\Delta L_{\mathrm{sound\ character}}$.

A linear regression, based on the repetition rate and the sharpness as independent variables, following equation 4.2 yields an adjusted $r^2 = 0.87$ independent of an inclusion of an interaction term. The fitted values for the regression parameters (given in Tab. 4.5) further support the in-

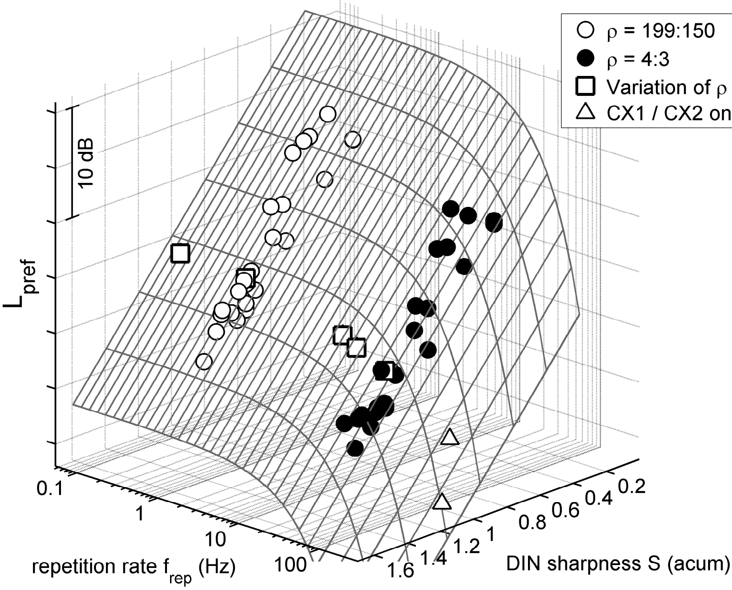


Figure 4.12: *Mean values of the PSE for preference L_{pref} for the 55 multi-tone sounds plotted over their respective values for the repetition rate f_{rep} and the sharpness S (DIN 45692). Markers indicate the ratio ρ between the fundamental frequencies. The direct prediction with a linear regression model $L_{\mathrm{pref}}(f_{\mathrm{rep}}, S)$ is indicated by the mesh grid.*

significance of the interaction term, because the parameters a - c remain rather stable with or without inclusion of the interaction term. Due to no added gain in explained variance by the regression with an interaction term, only the purely linear regression without interaction term is considered further. It is shown in Fig. 4.12 as a mesh grid.

The grid surface in Fig. 4.12 defined by the regression model represents an equal preference plane of the tested multi-tone sounds with respect to the reference sound. The fitted parameters of the model (in Tab. 4.5) disclose the relationships between the sound characteristics, quantified by the two descriptor variables, and the overall dB(A)-level of the sounds.

Presuming the validity of the model, moving on this plane gives combinations of parameter values for the overall A-weighted level, the repetition rate and the sharpness, which yield equally preferred sounds. According to the regression coefficients of the model, a level increase can be compensated with a reduction of the sharpness, the repetition rate or the combination of both.

$$L_{\mathrm{pref}}(f_{\mathrm{rep}}, S) = a + b \cdot f_{\mathrm{rep}} + c \cdot S + d \cdot f_{\mathrm{rep}} \cdot S \tag{4.2}$$

Table 4.5: *Fitted parameters a, b, c and d of a linear regression model described in equation 4.2 without ($d = 0$) and with an interaction term ($d \neq 0$) and the corresponding values of the adjusted r^2.*

model	a	b	c	d	adj. r^2
linear	const.	-0.14 $\frac{\mathrm{dB}}{\mathrm{Hz}}$	-20.1 $\frac{\mathrm{dB}}{\mathrm{acum}}$	—	0.87
linear + interaction	const.	-0.14 $\frac{\mathrm{dB}}{\mathrm{Hz}}$	-20.0 $\frac{\mathrm{dB}}{\mathrm{acum}}$	-0.0024 $\frac{\mathrm{dB}}{\mathrm{Hz \cdot acum}}$	0.87

4.2.2.2 Indirect prediction approach

The indirect prediction approach for the preference PSEs with an intermediate calculation of the level difference $\Delta L_{\mathrm{sound\ character}}$ is schematically shown in Fig. 4.13. In this model approach, a predicted value of the level difference $\Delta L_{\mathrm{sound\ character}}$ is used as the input for the estimated relationship between the level difference $\Delta L_{\mathrm{sound\ character}}$ and the preference PSEs L_{pref}.

The data set of the judgments and the descriptor variables (left part in Fig. 4.13) is randomly assigned into K equally sized subsets. Each of the K subsets is used as test data, identified by the index i_{Test}, for a model which is trained by all other subsets, identified by the indices i_{Train}. In figure 4.13 the first part ($i_{\mathrm{Test}} = 1$) is exemplarily left out for testing, but the procedure is carried out K-times and each of the K parts is used as the test part.

From the training data subsets (i_{Train}), the parameters of two separate regressions are estimated which are the parameters (a, b, c) in the function $\Delta L_{\mathrm{sound\ character}}(f_{\mathrm{rep}}(i_{\mathrm{Train}}), S(i_{\mathrm{Train}}))$ and the parameters (e, f) in the function $L_{\text{pref}}(\Delta L_{\text{sound character}}(i_{\text{Train}}))$ (Fig. 4.13, right part, middle).

Based on the entire parameter set (a, b, c, e, f)[5] and the test data of the descriptors (i_{Test}), a prediction of $\Delta L_{\text{sound character}}(f_{\text{rep}}(i_{\text{Test}}), S(i_{\text{Test}}))$ is calculated (Fig. 4.13, right part,bottom). The predicted values of $\Delta L_{\text{sound character}}$ form the basis for the prediction of $L_{\text{pref}}(\Delta L_{\text{sound character}}$ $(f_{\text{rep}}(i_{\text{Test}}, S(i_{\text{Test}})))$ which are compared to the corresponding values of the test data of $L_{\text{pref}}(i_{\text{Test}})$, measured in the listening tests (Fig. 4.13, right part, top).

4.2.2.3 Cross validation

Three cross validations, the 55-fold ($K = 55$, leave one out cross validation, LOOCV), the 11-fold ($K = 11$) and the 5-fold ($K = 5$), were calculated for both prediction approaches to assess their predictive powers based on a comparison of their mean squared errors. The calculation of the cross validation for the direct prediction approach is similar to that one used for the direct prediction of the sound character difference measure explained in Fig. 4.10.

[5] The parameter d is missing, because the interaction term is not included in the model for the level difference $\Delta L_{\text{sound character}}$, due to no added explained variance.

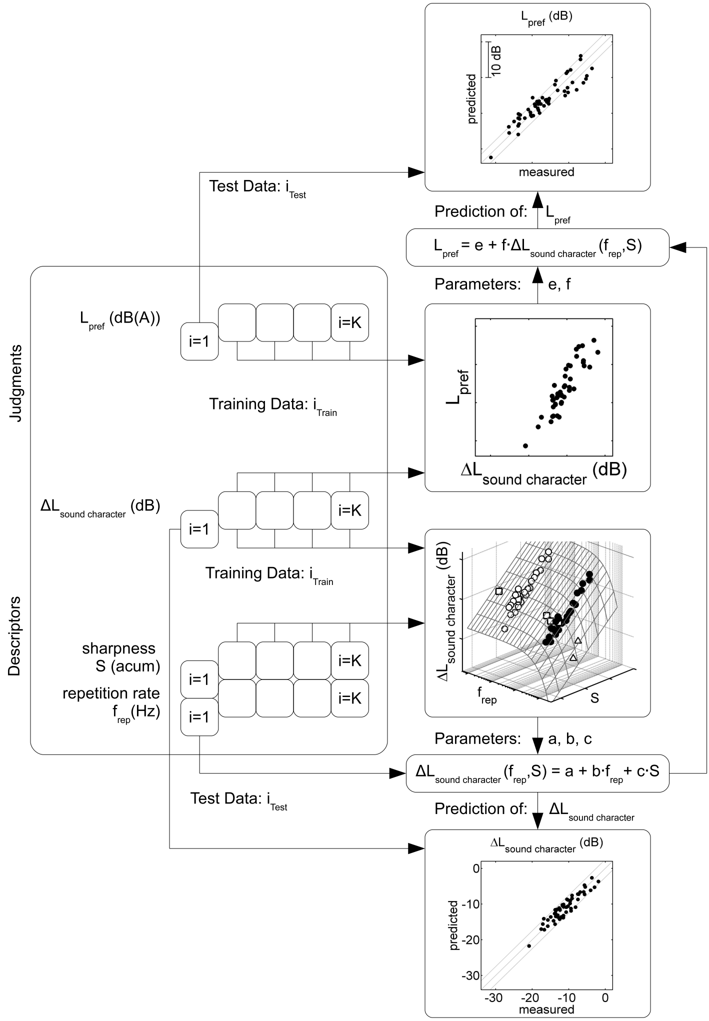


Figure 4.13: *Scheme of the indirect estimation of the preference evaluations $L_{\text{pref}}(\Delta L_{\text{sound character}}(f_{\text{rep}}, S))$ and the data splitting for the K-fold cross validations. Based on a set of training data, the linear regression between the two descriptor variables, sharpness S and repetition rate f_{rep}, and $\Delta L_{\text{sound character}}$ and the linear regression between $\Delta L_{\text{sound character}}$ and L_{pref} are fitted. The separate test data of the descriptor variables is then used for a prediction of the preference PSEs which are compared with the test data of the actually determined evaluations.*

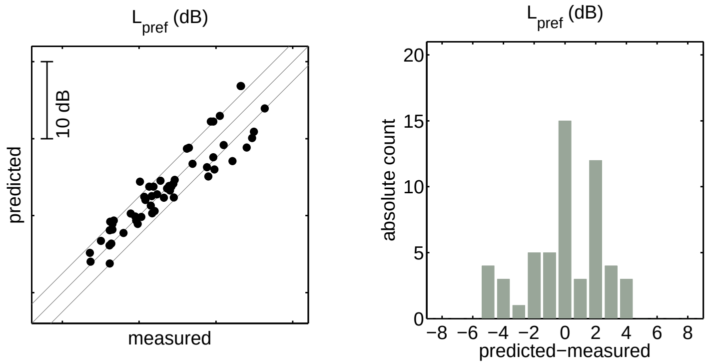
The results of the 5-fold cross validations for the direct (I) and the indirect (II) prediction approach are shown in Fig. 4.14 as a scatter plot of the predicted and the measured values for L_{pref}, each . The scatterings of the predicted and the measured data indicate a bigger spread between measured and predicted data for high values of L_{pref} than for low values. The predictions tend to underestimate the measured values, similarly for both prediction approaches. This can also be seen in the skewness of the histograms for the difference between predicted and measured values towards negative values. The figures of the two other cross validations (the leave one out (55-fold) and the 11-fold), both showing similar results, can be found in the appendix A.8.

The mean squared and root mean squared errors resulting from the cross validation are given in Tab. 4.6. Over all three cross validation procedures, comparable mean squared errors are found for both approaches, resulting in a root mean squared error of about $RMSE = 2.4$ dB in both cases. Hence, in terms of predictive accuracy it makes no difference whether the PSEs for preference are predicted directly from the descriptor values, or indirectly, based on an intermediate calculation of $\Delta L_{\mathrm{sound\ character}}$, which is itself predicted from the descriptor values. Taking into account the number of regression parameters necessary for the prediction, the direct prediction with only three parameters is preferable over an indirect prediction with five fitted parameters.

Table 4.6: *Mean squared error (MSE) and root mean squared error (RMSE) determined by the three different K-fold cross-validations for the direct prediction $L_{\mathrm{pref}}(f_{\mathrm{rep}}, S)$ and the indirect prediction $L_{\mathrm{pref}}(\Delta L_{\mathrm{sound\ character}}(f_{\mathrm{rep}}, S))$.*

	direct		indirect	
K-fold	MSE	RMSE	MSE	RMSE
55-fold (LOOCV)	5.83 dB^2	2.41 dB	5.80 dB^2	2.41 dB
11-fold	5.74 dB^2	2.40 dB	5.75 dB^2	2.40 dB
5-fold	5.65 dB^2	2.38 dB	5.83 dB^2	2.41 dB

(I) *direct prediction* $L_{\mathrm{pref}}(f_{\mathrm{rep}}, S)$



(II) *indirect prediction* $L_{\mathrm{pref}}(\Delta L_{\mathrm{sound\ character}}(f_{\mathrm{rep}}, S))$

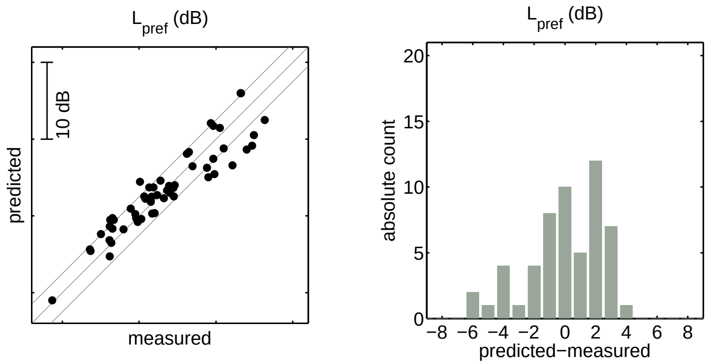


Figure 4.14: *Results of the 5-fold cross validation of the direct (I) and the indirect (II) prediction approach. Left: predicted values of* L_{pref} *plotted over the measured data. Continuous lines indicate identity* ± 2.5 *dB. Right: Histogram of the difference between predicted and measured values of* L_{pref}.

4.3 Conclusion

The relationships between the subjective judgments from Chap. 3 and psychoacoustic descriptors were investigated in order to identify descriptor variables that explain a considerable amount of the variance in the subjective judgments while effectively describing the sound characteristics resulting from the broad variety of technically relevant signal parameter variations. The identified descriptors are combined into regression models which aim at a prediction of the judgments in terms of levels at the points of subjective equality.

The PSEs for loudness L_{loud} are not reasonably reflected by any of the tested loudness models. A possible explanation for this might be the revealed close relationship between the PSEs for loudness and the PSEs for preference which might have resulted in a bias of the loudness judgments by partially compensating the unpleasantness of the sounds. No other psychoacoustic descriptors could be identified to reflect the loudness judgments reasonably, as was expected.

The PSEs for preference L_{pref} and also the level difference attributed to the sound character $\Delta L_{\text{sound character}}$ can both be similarly related to the same two descriptors. The variability in the judgments resulting from differences in the spectral composition are best reflected by the sharpness according to the DIN 45692 standard. The variability in the judgments originating from differences in the temporal signal structure are covered well by the repetition rate of the time signals. In general, the values of L_{pref} and $\Delta L_{\text{sound character}}$ become more negative for an increase of the sharpness and also for an increase of the repetition rate. Due to the small amount of shared variance between the sharpness and repetition rate, these two descriptors were chosen as the basis of the prediction models.

The combination of the repetition rate and the sharpness into a purely linear regression model for the level difference attributed to the sound character $\Delta L_{\text{sound character}}$ accounts for 89 % of the variance (adjusted $r^2 = 0.89$). The model predicts the level difference with a root mean squared error of $RMSE = 1.3$ dB with only three fitted parameters. For the preference PSEs, a direct prediction of the equal preference contour, based on the repetition rate and the sharpness, yields an adjusted

$r^2 = 0.87$ and a prediction error of $RMSE = 2.4$ dB, also with only three fitted parameters. This model describes a plane of equal preference and the fitted parameters of the model disclose the quantitative relationships between the overall dB(A)-level of the multi-tone sounds and the sound characteristics, reflected by the two descriptor variables.

5 General conclusion

Tonal sounds, originating from rotating machinery parts, are a part of everyday noise, which is known to be evaluated as rather unpleasant and having a nuisance effect compared to broadband noise without salient tonal components. However, for sounds consisting of multiple tonal components, the sound character is far more complex compared to single tones and also the signal parameter space is increased in complexity. Such multi-tone sounds, consisting of two complex tones and additional combination tones, occur for example in machinery noise of aircraft engines with counter-rotating open rotors or counter-rotating fans, where non-linearities in the sound generating process lead to the combination tones. In particular, it is unclear, how signal parameter variations, which are relevant in technical applications, influence the loudness and the sound character of such sounds. In terms of the overall appraisal of such sounds, it is furthermore of interest to understand the contribution of the loudness and the sound character to the preference of such sounds in the context of aircraft cabin noise.

The overall aim of this thesis was a fundamental understanding of the signal parameters and the acoustic variables influencing the sound character and the appraisal of multi-tone sounds consisting of two harmonic complex tones and additional combination tones. The specific objectives were (I) an exploration of the perceptual space of multi-tone sounds resulting from a variation of the frequency ratio between the fundamental frequencies of the two complex tones, (II) a determination of the quantitative impact of relevant sound characteristics on loudness judgments and preference evaluations in terms of equivalent dB(A)-values, and (III) a prediction model for the preference evaluations, which provides a link between psychoacoustic variables, describing the sound characteristics, and preference equivalent dB-values.

In Chap. 2, the perceptual space, resulting from a variation of the frequency ratio $\rho = f_{01} : f_{10}$ between the fundamental frequencies, was explored. The factorial analysis of the semantic differential data revealed four perceptual dimensions which describe the pleasantness, the power, the temporal structure and the spectral content of the sounds. It could be shown that the frequency ratio ρ changed the sound character and significantly affected the judged pleasantness of the multi-tone sounds. Mathematically exact ratios of small integers were found to be considerably more unpleasant than ratios of large integer numbers, which can be achieved by slightly shifting one fundamental frequency by less than 1 Hz. The mean semantic profiles of three groups of sounds showed that the differences in pleasantness, resulting from a variation of the frequency ratio ρ, are mainly related to the temporal structure of the multi-tone sounds. It turned out that the repetition rate of the signal's time series, which is equivalent to the inverse of the periodicity, is a suitable descriptor, sharing 70.5 percent of the variance with the factor values of the pleasant dimension. High values of the repetition rate (equivalent to short periodic times) are judged to be considerably more unpleasant than low values of the repetition rate (equivalent to long periodic times). This result suggests that adjusting the frequency ratio in such a way that the repetition rate is reduced, and the periodicity increased, might help to render sounds containing two complex tones and combination tones more pleasant.

The determination of loudness judgments and preference evaluations as levels at the point of subjective equality (PSE) are presented in Chap. 3. The PSEs were determined with an adaptive level varying procedure, converging at the 50 % point of the psychometric function compared to a reference sound, which was fixed in level. For all tested multi-tone sounds, the PSEs for preference were below the PSEs for loudness, indicating that an additional level reduction was necessary to compensate for the unpleasant sound character of the multi-tone sounds and to achieve equal preference. The level difference between the PSE for preference L_{pref} and the PSE for loudness L_{loud} is attributed to the sound character ($\Delta L_{\text{sound character}} = L_{\text{pref}} - L_{\text{loud}}$) and used as a quantitative measure for it. It turned out that the level difference $\Delta L_{\text{sound character}}$ is closely related to the preference PSEs ($r^2_{\text{mean}} = 83$ %) and only weakly related to the loudness PSEs ($r^2_{\text{mean}} = 32$ %). This indicates that the participants actually differentiated between the loudness judgments and the preference evaluations.

The variation of the slope of the spectral envelope towards high frequencies results in differences of about 10 dB with respect to $\Delta L_{\text{sound character}}$. Similarly, the variation of temporal sound characteristics in terms of the periodicity of the signals covers a range of about 10 dB in terms of $\Delta L_{\text{sound character}}$. The variation of the frequency ratio ρ, which was also the experimental parameter in the semantic differential experiment, led to differences of only about 4 dB in terms of the preference equivalent levels. The two complex tones alone led to the lowest loudness and preference PSEs of all investigated sounds and would be the loudest and least preferred sounds when presented at equal dB(A)-level. Increasing the number of tonal components from a complex tone to a mixture of two complex tones was found to be beneficial in terms of preference equivalent dB(A)-levels - a finding which was rather unexpected. Over the whole study, the level difference $\Delta L_{\text{sound character}}$ provided clearer relationships with the signal parameters, which were easier to interpret than the raw loudness judgments and the raw preference evaluations. This difference measure was facilitated by a combined measurement of one attribute on the denotative level (loudness) and one on the connotative level (preference) with a method, allowing for a quantitative comparison of the two.

The results of the extensive listening tests presented in Chap. 3 are the basis for two prediction models described in Chap. 4. The relationships between the loudness and preference PSEs and selected (psycho-)acoustic parameters were analyzed to identify suitable descriptor variables that reflect the differences of the multi-tone sounds in terms of their sound character. It turned out that current loudness models were not able to satisfactorily predict the loudness judgments obtained in the listening tests. The high correlation coefficient found between the loudness judgments and the preference evaluations ($r^2_{\text{mean}} = 74$ %), indicating a link between the preference evaluation and the loudness judgment, might be an explanation for this. The variability in the preference evaluations due to the specific signal parameters that mainly affect the spectral content of the multi-tone sounds are best reflected by the sharpness metrics. Differences in the temporal structure of the multi-tone sounds with respect to the periodicity were best reflected by the repetition rate, which was already identified as an effective descriptor in the semantic differential study, presented in Chap. 2. Given that these two descriptor variables, sharpness and repetition rate, have very little shared variance, they are used as a complementary basis for the prediction of the preference evaluations. The level independence of both descriptors further excludes an intrinsic co-variation with the judgments, which are expressed as levels at the PSEs, and avoids spurious correlation between the two. The evident relationships between the level difference attributed to the sound character $\Delta L_{\text{sound character}}$ and the signal parameter variations revealed in Chap. 3 are also found to result in rather clear relationships with the sharpness and the repetition rate.

The prediction model for the preference evaluations, resulting from a linear regression based on the two descriptors, depicts an equal preference plane which links the descriptor variables and the preference equivalent dB(A)-level for the investigated type of multi-tone sounds. In this way, the model allows to relate differences in sound character (e.g differences in sharpness or periodicity) to preference equivalent dB(A)-levels. The other way round, level differences connected by the model plane can be interpreted as level penalties (or level boni) resulting from sound character differences, which are precisely characterized by the two descriptors. This knowledge can be used to estimate the effects of technical engine parameters or noise control measures for an optimization of multi-tone sounds in terms of preference.

All in all, the outcomes of the semantic differential and the PSE measurements provide a better understanding of the acoustic factors, influencing the sound character of multi-tone sounds, which consist of two complex tones and combination tones. Furthermore, the findings provide a quantitative link between sound characteristics, the overall level and the appraisal of multi-tone sounds in the context of aircraft cabin noise, which was not available before. For the investigated set of multi-tone sounds, those stimuli having low repetition rates and sharpness values of about 0.7 acum allow for the highest dB(A)-levels while remaining equally preferred as the rest of the stimulus set at lower dB(A)-levels.

Even though the particular results gained in this thesis might be limited in terms of the specific signal structure and evaluation context investigated, the used measurement method may be advantageous for future research questions. Whenever *quantitative* links between sound characteristics and the appraisal of sounds are demanded, a common currency for both domains is required. The results of this thesis show, that it is in principle possible to measure loudness judgments (on the denotative level) and preference evaluations (on the connotative level) on the same dB-scale. The combined measurement of both aspects opens the possibility to quantify the influence of sound characteristics on the overall appraisal in the form of the relative difference between the two, also on a dB-scale.

Appendix

A.1 Accounting for annoyance beyond the dB(A): The concept of rating level

In noise legislation and acoustic standards, sounds are so far almost exclusively characterized by the single number value of the A-weighted sound pressure level which are based on the found high correlation coefficients between the annoyance and the A-weighted level. However, at a closer look, the judgments of sounds with the same A-weighted sound pressure level can also vary considerable in their annoyance effect, due to differences in the sound character or the emitting sound source. To compensate for this shortcoming of the dB(A) in correctly reflecting the annoyance effects of sounds, level adjustments (in most cases penalties) are introduced. The combination of the measured A-weighted level and the level adjustment is called rating level[1]. Thus, for noise types which are were found to be more annoying than others (e.g. due to tonal components) the level adjustment acts as a penalty in the instrumental assessment and noises with equal rating level are intended to represent a comparable annoyance effect.

In detail, the rating level L_r consists of the A-weighted equivalent continuous level L_{Aeq} and additional level adjustments k:

$$L_r = L_{Aeq} + k \text{ dB(A)}^2 \tag{A.1}$$

The level adjustments k are prescribed for different sound sources (which are also politically driven), time periods and also for sound characteristics, which are based on the observation that impulsive and tonal sounds

[1] German: Beurteilungspegel

[2] This is the simplified formulation, not taking into account different level adjustment for partial times.

yield a higher annoyance. Figure A.1 gives an overview of level adjustments after the ISO 1996-1:2003 standard[(120)].

The rating level is used in the rating of environmental noise immissions (DIN 45645-1:1996[(52)] and ISO 1996-1:2003[(120)]) and also noise at the working place (DIN 45645-2:1996[(53)]). Except for railway noise, the rating level is usually higher than the technically measured sound pressure level. In the measurement of noise immissions in the field, the level adjustments are usually prescribed on the basis of the subjective impression of a consultant who carries out the measurement. Only in some application cases the penalties can be objectively determined by an algorithm, e.g. for the tonality described in the DIN standard 45681[(54)].

Type	**Specification**	**Level adjustment** dB
Sources of sound	Road traffic	0
	Aircraft	3 to 6
	Railway[a]	−3 to −6
	Industry	0
Source character	Regular impulsive[b]	5
	Highly impulsive	12
	High-energy impulsive	See Annex B
	Prominent tones[c]	3 to 6
Time period	Evening	5
	Night	10
	Weekend daytime[d]	5

[a] The railway adjustments do not apply to long diesel trains or to trains travelling in excess of 250 km/h.

[b] Some countries apply objective prominence tests to assess whether sound sources are regular impulsive.

[c] If the presence of prominent tonal content is in dispute, then ISO 1996-2 provides measurement procedures that should be used to verify its presence.

[d] The weekend daytime adjustment is added to L_d as defined by the corresponding authority (see 6.5).

Figure A.1: *Level adjustments k for different sound sources, sound characteristics and time periods according to the ISO 1996-1:2003 standard[(120)].*

A.2 Sematic differential test – Instructions to the participants

Versuche zur Bestimmung eines optimalen Geräusches

Geräusche von Geräten des alltäglichen Lebens können sehr unterschiedliche klangliche Eigenschaften haben, selbst wenn sie vom gleichen Gerätetyp stammen.
Unter verschiedenen Geräuschvarianten soll nun ein optimales Geräusch gefunden werden. Dazu werden diese Hörversuche durchgeführt.

Zur Orientierung werden zunächst alle 15 Geräusche vorgespielt. Zu diesen Zeitpunkt geht es darum, einen Eindruck von den später zu bewertenden Geräuschen zu bekommen.

Danach werden die Geräusche einzeln nacheinander hinsichtlich 16 verschiedener Eigenschaften beurteilt. Dazu werden jeweils Skalen ausgegeben.
Die Geräusche werden in einem ersten Durchgang bezüglich einer ersten Eigenschaft bewertet. Im nächsten Durchgang wird eine zweite Eigenschaft beurteilt, usw...

Nach jeweils sechs Skalen bitten wir Sie, ihre Eindrücke schriftlich festzuhalten.

Nach Abschluss aller Bewertungen tauschen wir uns über die Eindrücke in einer Diskussion aus.

Für den gesamten Versuch haben wir einen Zeitraum von etwa 2 Stunden vorgesehen.

Vielen Dank für die Teilnahme an diesem Versuch!

A.3 Stimulus properties

Table A.1: *Overview of the properties of all 55 multi-tone stimuli used in the listening tests with the PSE-method, described in Chap. 3.*

	ID	slope_{UP} (dB/oct)	slope_{DN} (dB/oct)	f_{peak} (Hz)	ratio $\rho = \frac{f_{01}}{f_{10}}$	basic multitone elements	attenuation δ (dB)	freq. range attenuated by 10 dB
stage 1	A	—	-6	100	9:7	CX1+CX2+CTs	-10	—
	B	—	-6	100	1293:1000	CX1+CX2+CTs	-10	—
	C	—	-6	100	13:10	CX1+CX2+CTs	-10	—
	D	—	-6	100	199:150	CX1+CX2+CTs	-10	—
	E	—	-6	100	4:3	CX1+CX2+CTs	-10	—
	F	—	-6	100	199:150	CX1	—	—
	G	—	-6	100	199:150	CX2	—	—
	D2	—	-6	100	199:150	CX1+CX2	—	—
	E2	—	-6	100	4:3	CX1+CX2	—	—
	D3	—	-6	100	199:150	CTs	—	—
	E3	—	-6	100	4:3	CTs	—	—
	D4	—	-6	100	199:150	CX1+CX2+CTs	-5	—
	E4	—	-6	100	4:3	CX1+CX2+CTs	-5	—
	D5	—	-6	100	199:150	CX1+CX2+CTs	-10	—
	E5	—	-6	100	4:3	CX1+CX2+CTs	-10	—
	D6	—	-6	100	199:150	CX1+CX2+CTs	-15	—
	E6	—	-6	100	4:3	CX1+CX2+CTs	-15	—
stage 2	D7	—	-6	100	199:150	CX1+CX2+CTs	-10	—
	E7	—	-6	100	4:3	CX1+CX2+CTs	-10	—
	D8	—	-9	100	199:150	CX1+CX2+CTs	-10	—
	E8	—	-9	100	4:3	CX1+CX2+CTs	-10	—
	D9	—	-12	100	199:150	CX1+CX2+CTs	-10	—
	E9	—	-12	100	4:3	CX1+CX2+CTs	-10	—
	D10	—	-15	100	199:150	CX1+CX2+CTs	-10	—
	E10	—	-15	100	4:3	CX1+CX2+CTs	-10	—
stage 1	D11	—	-6	100	199:150	CX1+CX2+CTs	-10	low
	E11	—	-6	100	4:3	CX1+CX2+CTs	-10	low
	D12	—	-6	100	199:150	CX1+CX2+CTs	-10	mid
	E12	—	-6	100	4:3	CX1+CX2+CTs	-10	mid
	D13	—	-6	100	199:150	CX1+CX2+CTs	-10	high
	E13	—	-6	100	4:3	CX1+CX2+CTs	-10	high
stage 3	D14	(+6)	-6	100	199:150	CX1+CX2+CTs	-10	—
	E14	(+6)	-6	100	4:3	CX1+CX2+CTs	-10	—
	D15	+6	-6	200	199:150	CX1+CX2+CTs	-10	—
	E15	+6	-6	200	4:3	CX1+CX2+CTs	-10	—
	D16	+6	-6	300	199:150	CX1+CX2+CTs	-10	—
	E16	+6	-6	300	4:3	CX1+CX2+CTs	-10	—
	D17	+6	-6	400	199:150	CX1+CX2+CTs	-10	—
	E17	+6	-6	400	4:3	CX1+CX2+CTs	-10	—
	D18	+6	-6	500	199:150	CX1+CX2+CTs	-10	—
	E18	+6	-6	500	4:3	CX1+CX2+CTs	-10	—
	D19	(+6)	-15	100	199:150	CX1+CX2+CTs	-10	—
	E19	(+6)	-15	100	4:3	CX1+CX2+CTs	-10	—
	D20	+6	-15	200	199:150	CX1+CX2+CTs	-10	—
	E20	+6	-15	200	4:3	CX1+CX2+CTs	-10	—
	D21	+6	-15	300	199:150	CX1+CX2+CTs	-10	—
	E21	+6	-15	300	4:3	CX1+CX2+CTs	-10	—
	D22	+6	-15	400	199:150	CX1+CX2+CTs	-10	—
	E22	+6	-15	400	4:3	CX1+CX2+CTs	-10	—
	D23	+6	-15	500	199:150	CX1+CX2+CTs	-10	—
	E23	+6	-15	500	4:3	CX1+CX2+CTs	-10	—
	D24	+15	-15	300	199:150	CX1+CX2+CTs	-10	—
	E24	+15	-15	300	4:3	CX1+CX2+CTs	-10	—
	D25	+15	-15	500	199:150	CX1+CX2+CTs	-10	—
	E25	+15	-15	500	4:3	CX1+CX2+CTs	-10	—

A.4 Loudness and preference matchings – Instructions to the participants

Lautstärke von Flugzeugkabinengeräuschen

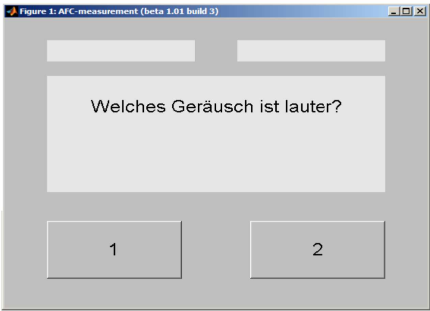
In diesem Versuch geht es um die Lautstärke von Flugzeugkabinengeräuschen.

Die Geräusche sind einerseits Kabinengeräusche aus einem aktuellen Flugzeug und zum anderen futuristische Flugzeugkabinengeräusche.

Sie hören die Geräusche in unserem sog. *reflexionsarmen Raum*, in dem die Wände den Schall „schlucken“ und so eine Färbung der Geräusche durch Raumeinfluss vermeiden.

Ihnen werden jeweils zwei Geräusche angeboten und danach sollen Sie entscheiden, welches Geräusch lauter ist. Beim Abspielen des ersten Geräusches leuchtet auf dem Bildschirm vor Ihnen links das Feld mit der „1“. Nach einer kurzen Pause wird das zweite Geräusch, während Feld „2“ leuchtet, vorgespielt.

Bitte entscheiden Sie nun die Frage:



und drücken Sie als Antwort auf der **Tastatur** entweder die „1“ oder die „2“.

Nach einer kurzen Pause werden Ihnen die nächsten beiden Geräusche vorgespielt, Sie entscheiden sich wieder für das Lautere, drücken die entsprechende Taste, usw. ...

Der Versuch dauert etwa 15 Minuten und endet mit folgender Bildschirmmeldung

„Das Experiment ist abgeschlossen. Herzlichen Dank!“

Bitte beachten Sie weiter Folgendes:

Wenn Sie aus irgendwelchen Gründen auch immer, den Versuch vorzeitig beenden möchten, können Sie dies jederzeit tun, aufstehen und den Raum verlassen.

Die Geräusche, die Ihnen in diesem Versuch angeboten werden, sind nie lauter als Flugzeugkabinengeräusche in modernen Düsenflugzeugen.

Zum Schluss möchten wir Sie dringend bitten, bis zum Ende diesen Jahres, wenn diese Versuchserie zu Ende sein wird, nicht über Ihre Erfahrungen in diesen Versuchen zu sprechen, um nicht andere potentielle Versuchsteilnehmer oder Versuchsteilnehmerinnen zu beeinflussen.

Figure A.2: *Instruction used for the loudness matching experiments.*

CARL VON OSSIETZKY universität OLDENBURG acoustics GROUP

<u>Beurteilung von Flugzeugkabinengeräuschen</u>

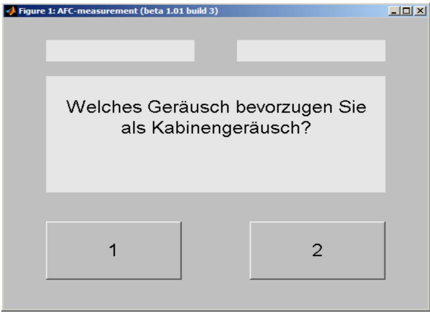
In diesem Versuch geht es um die Angenehmheit bzw. Unangenehmheit von Geräuschen.

Die Geräusche sind einerseits Kabinengeräusche aus einem aktuellen Flugzeug und zum anderen futuristische Flugzeugkabinengeräusche.

Sie hören die Geräusche in unserem sog. *reflexionsarmen Raum*, weil die Wände den Schall „schlucken" und so eine Färbung der Geräusche durch Raumeinfluss vermeiden.

Ihnen werden jeweils zwei Geräusche angeboten und danach sollen Sie entscheiden, welches Sie bevorzugen. Beim Abspielen des ersten Geräusches leuchtet auf dem Bildschirm vor Ihnen links das Feld mit der „1". Nach einer kurzen Pause wird das zweite Geräusch, während Feld „2" leuchtet, vorgespielt.

Bitte entscheiden Sie nun die Frage:



und drücken Sie als Antwort auf der **Tastatur** entweder die „1" oder die „2".

Nach einer kurzen Pause werden Ihnen die nächsten beiden Geräusche vorgespielt, Sie entscheiden sich wieder für das Angenehmere, drücken die entsprechende Taste, usw. ...

Der Versuch dauert etwa eine halbe Stunde und endet mit folgender Bildschirmmeldung

„Das Experiment ist abgeschlossen. Herzlichen Dank!"

Bitte beachten Sie weiter Folgendes:

Wenn Sie aus irgendwelchen Gründen auch immer, den Versuch vorzeitig beenden möchten, können Sie dies jederzeit tun, aufstehen und den Raum verlassen.

Die Geräusche, die Ihnen in diesem Versuch angeboten werden, sind nie lauter als Flugzeugkabinengeräusche in modernen Düsenflugzeugen.

Zum Schluss möchten wir Sie dringend bitten, bis zum Ende diesen Jahres, wenn diese Versuchserie zu Ende sein wird, nicht über Ihre Erfahrungen in diesen Versuchen zu sprechen, um nicht andere potentielle Versuchsteilnehmer oder Versuchsteilnehmerinnen zu beeinflussen.

Figure A.3: *Instruction used for the preference matching experiments.*

A.5 Influence of the spectral envelope on the judgments and evaluations – compilation of the results for a variation of the spectral decline and the peak frequency

The measurements of the loudness judgments and the preference evaluations for a variation of the the peak frequency f_{peak} were carried out in measurement stage 3 ($N = 34$) for two values of the falling upper slope slope_{DN} which was investigated in detail and parametrically varied in measurements stage 2 ($N = 48$). The influence of the peak frequency f_{peak} was determined for the most shallow $\text{slope}_{\text{DN}} = -6$ dB/oct and the steepest declines towards high frequencies $\text{slope}_{\text{DN}} = -15$ dB/oct, each for the two values $\rho = 199 : 150$ and $\rho = 4 : 3$.

In the following figures, the results from the measurement stages 2 and 3 were compiled into single plots showing the subjective judgments as a function of the two variables: falling upper slope slope_{DN} and spectral peak frequency f_{peak}.

Consistent with the findings on the reproducibility of the judgments between different groups (in Sec. 3.3.2), the mean values of the loudness and preference PSEs, measured in both measurement stages (in the corners of the plots), by mostly disjunct groups of participants, are in good agreement with each other. The slight differences apparent between the mean values of $\Delta L_{\text{sound character}}$ from the two measurement stages are comparable to those which were found to be not statistically significant in the comparison of the three disjunct groups of participants.

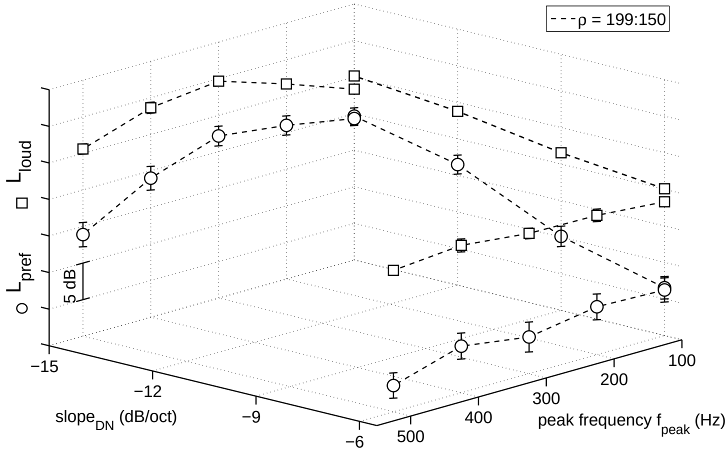


Figure A.4: *Mean values of the PSEs for loudness (open squares) and for preference (open circles) plotted as a function of peak frequency and falling upper slope for* $\rho = 199 : 150$*. Errorbars indicate the standard deviation of the calculated mean values (sometime even smaller than the symbols).*

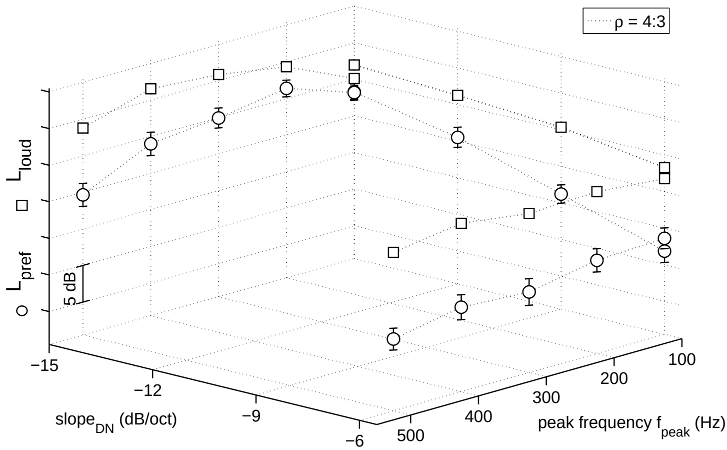


Figure A.5: *Same as figure A.4, but for* $\rho = 4 : 3$*.*

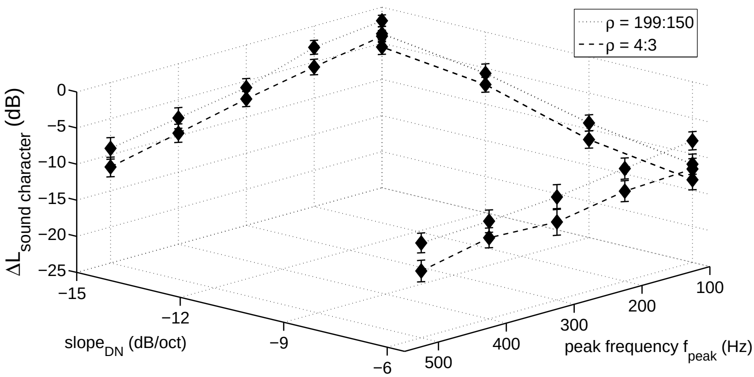


Figure A.6: *Mean values of the level difference $\Delta L_{\text{sound character}}$ plotted as a function of peak frequency and falling upper slope for the two values $\rho = 199 : 150$ (dotted lines) and $\rho = 4 : 3$ (dashed lines). Errorbars indicate the standard deviation of the calculated mean values (sometime even smaller than the symbols).*

A.6 Relationships between the level difference attributed to the sound character and psychoacoustic descriptors

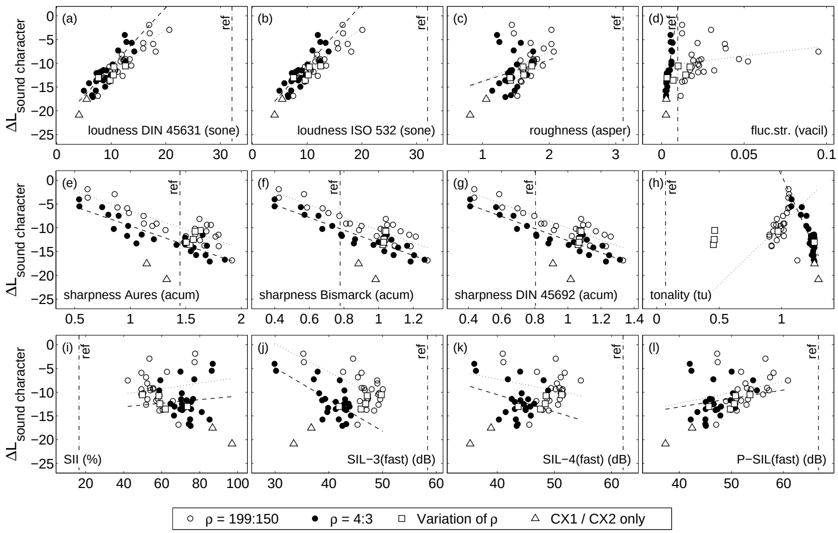


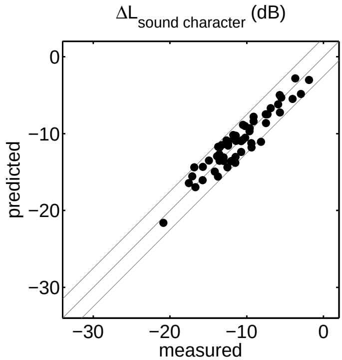
Figure A.7: *Mean values of the level difference* $\Delta L_{\text{sound character}}$ *for 55 sounds plotted over the values of 12 different psychoacoustic descriptors. The value of each psychoacoustic descriptor for the reference sound is indicated by the vertical dash-dotted line in the plot of each metric. The metric values are calculated based on a sound pressure level corresponding to the respective PSE for preference* (L_{pref})*. Due to the independence of* $\Delta L_{\text{sound character}}$ *from an absolute level, the relationships with level dependent descriptors have to be regarded with care.*

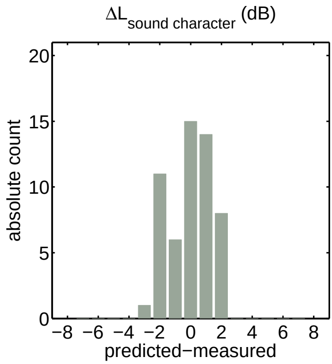
Table A.3: *Pearson correlation coefficients and p-values between 12 psychoacoustic descriptor variables and $\Delta L_{\text{sound character}}$, separated for the two values of the frequency ratioρ. The psychoacoustic descriptor values were calculated with a sound pressure level equal to the PSE for preference (L_{pref}). The characters (a) to (l) resemble those used for the subplots of Fig. A.7. The correlation coefficients reflect the linear relationships indicated by the dotted ($\rho = 199 : 150$) and the dashed ($\rho = 4 : 3$) lines in that figure. Due to the independence of $\Delta L_{\text{sound character}}$ from the absolute level, the relationships with level dependent descriptors have to be regarded with care.*

		$\rho = 199 : 150$		$\rho = 4 : 3$	
	descriptor	Pearson's r	p-value	Pearson's r	p-value
(a)	loudness DIN 45631	0.89	<0.001	0.89	<0.001
(b)	loudness ISO 532	0.88	<0.001	0.88	<0.001
(c)	roughness	0.32	0.121	0.19	0.368
(d)	fluctuation strength	0.21	0.316	0.88	<0.001
(e)	sharpness Aures	-0.89	<0.001	-0.90	<0.001
(f)	sharpness Bismarck	-0.91	<0.001	-0.94	<0.001
(g)	sharpness DIN 45692	-0.91	<0.001	-0.94	<0.001
(h)	tonality	0.42	0.036	-0.87	<0.001
(i)	SII	0.14	0.519	0.07	0.717
(j)	SIL-3 (fast)	-0.62	<0.001	-0.70	<0.001
(k)	SIL-4 (fast)	-0.21	0.314	-0.30	0.139
(l)	P-SIL (fast)	0.23	0.259	0.17	0.412

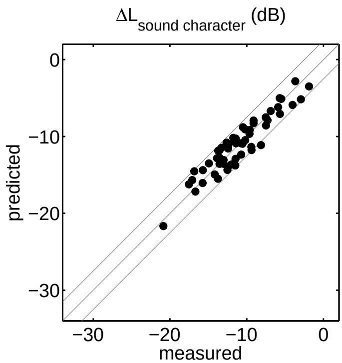
A.7 Cross validations of the prediction model for the level difference attributed to the sound character

(I) *Leave one out cross validation (LOOCV, 55-fold)*





(II) *11-fold cross validation*



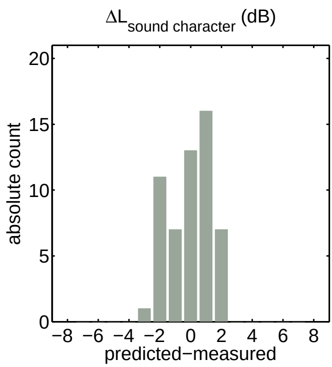
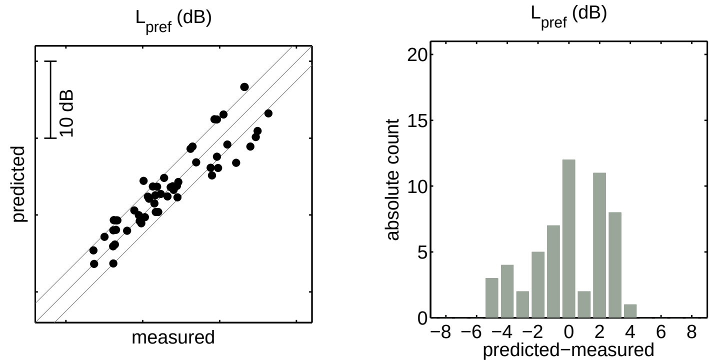


Figure A.8: *Results of (I) the leave one out (LOOCV, 55-fold) and (II) the 11-fold cross validation. Left: predicted values of* $\Delta L_{\text{sound character}}$ *plotted over the measured data. Continuous lines indicate identity* ± 2.5 *dB. Right: Histogram of the difference between predicted and measured values of* $\Delta L_{\text{sound character}}$.

A.8 Cross validations of the prediction model for the preference evaluations

(I) *Leave one out cross validation (LOOCV, 55-fold)*



(II) *11-fold cross validation*

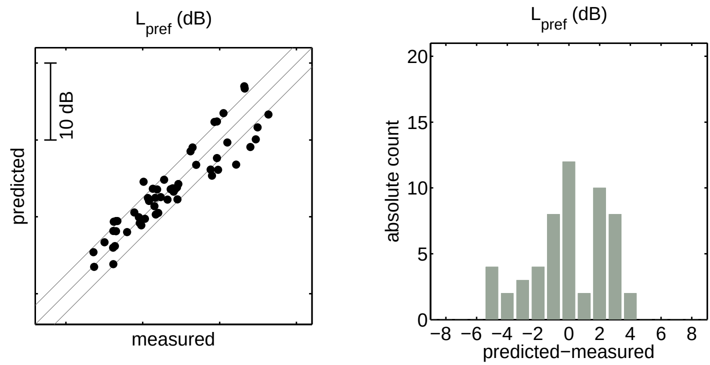
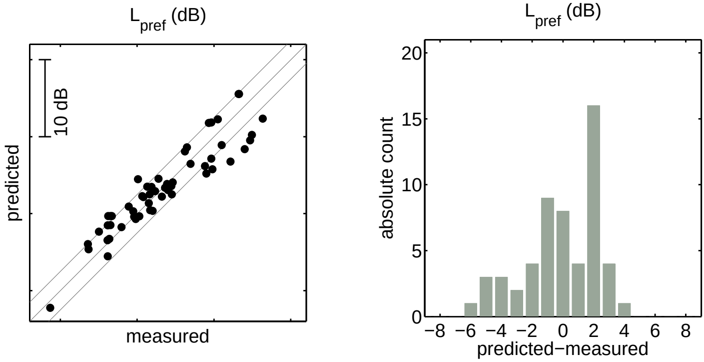


Figure A.9: *Results of (I) the leave one out (LOOCV, 55-fold) and (II) the 11-fold cross validation. Left: directly predicted values of L_{pref} plotted over the measured data. Continuous lines indicate identity ± 2.5 dB. Right: Histogram of the difference between predicted and measured values of L_{pref}.*

(I) *Leave one out cross validation (LOOCV, 55-fold)*



(II) *11-fold cross validation*

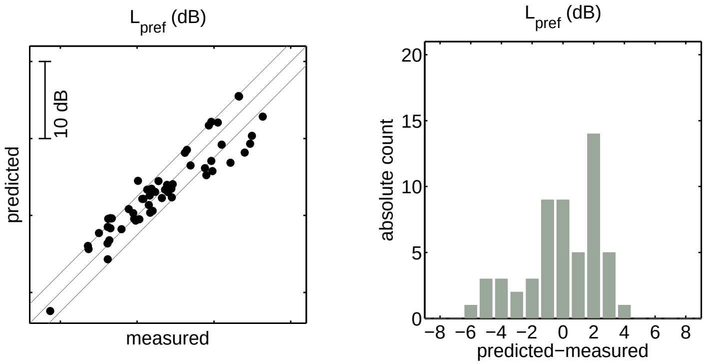


Figure A.10: *Results of (I) the leave one out (LOOCV, 55-fold) and (II) the 11-fold cross validation. Left: indirectly predicted values of* L_{pref} *plotted over the measured data. Continuous lines indicate identity* ± 2.5 *dB. Right: Histogram of the difference between predicted and measured values of* L_{pref}*.*

Bibliography

[1] M. Alayrac, C. Marquis-Favre, S. Viollon, J. Morel, and G. L. Nost. Annoyance from industrial noise: Indicators for a wide variety of industrial sources. *J. Acoust. Soc. Am.*, 128(3):1128–1139, 2010. DOI: 10.1121/1.3466855.

[2] M. Alayrac, C. Marquis-Favre, and S. Viollon. Total annoyance from an industrial noise source with a main spectral component combined with a background noise. *J. Acoust. Soc. Am.*, 130(1): 189–199, 2011. DOI: 10.1121/1.3598452.

[3] M. E. Altinsoy and U. Jekosch. The semantic space of vehicle sounds: Developing a semantic differential with regard to customer perception. *J. Audio Eng. Soc.*, 60(1/2):13–20, 2012.

[4] J. Angerer, D. McCurdy, and R. Erickson. Development of an annoyance model based upon elementary auditory sensations for steady-state aircraft interior noise containing tonal components. Technical Memorandum 104147, NASA, 1991.

[5] ANSI S1.1-1994. Acoustical terminology. *Amercian National Standards Institute*, pp. 1–58, 1994.

[6] ANSI S1.13-2005. Measurement of sound pressure levels in air. *Amercian National Standards Institute*, pp. 1–52, 2005.

[7] ANSI S12.65-2006. Rating noise with respect to speech interference (Revision of ANSI S3.14-1977). *Amercian National Standards Institute*, pp. 1–20, 2006.

[8] ANSI S1.42-2001. Design response of weighting networks for acoustical measurements. *Amercian National Standards Institute*, pp. 1–29, 2001.

[9] ANSI S3.4-2007. Procedure for the computation of loudness of steady sounds. *Amercian National Standards Institute*, 2007.

[10] ANSI S3.5-1997. Methods for calculation of the speech intelligibility index. *Amercian National Standards Institute*, pp. 1–31, 1997.

[11] S. Arlot and A. Celisse. A survey of cross-validation procedures for model selection. *Statist. Surv.*, 4:40–79, 2010. DOI: 10.1214/09-SS054.

[12] W. Aures. Ein Berechnungsverfahren der Rauigkeit (A procedure for calculating auditory roughness). *Acustica*, 58(5):268–281, 1985.

[13] W. Aures. Der sensorische Wohlklang als Funktion psychoakustischer Empfindungsgrößen (The sensory euphony as a function of auditory sensations). *Acustica*, 58(5):282–290, 1985.

[14] W. Aures. Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale (A model for calculating the sensory euphony of various sounds). *Acustica*, 59(2):130–141, 1985.

[15] J. Backus. Multiphonic tones in the woodwind instruments. *J. Acoust. Soc. Am.*, 63(2):591–599, 1978.

[16] J. W. Beachamp. Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones. *J. Audio Eng. Soc.*, 30(6):396–406, 1982.

[17] S. Bech and N. Zacharov. *Perceptual audio evaluation: theory, method and application.* John Wiley & Sons Ltd., 2006. ISBN: 0470-86923-2.

[18] A. H. Benade. *Fundamentals of musical acoustics.* New York, Oxford University Press, 1976.

[19] K. Benedini. Klangfarbenunterschied zwischen tiefpaßgefilterten harmonischen Klängen (Timbre differences between lowpass-filtered harmonic complex tones). *Acustica*, 44:129–134, 1980.

[20] K. Benedini. Messung der Klangfarbenunterschiede zwischen schmalbandigen harmonischen Klängen (Measurement of timbre diferences between complex tones). *Acustica*, 44:188–193, 1980.

[21] L. L. Beranek and H. W. Rudmose. Sound control in airplanes. *J. Acoust. Soc. Am.*, 19(2):357–364, 1947. DOI: 10.1121/1.1916491.

[22] R. Bergler. *Das Eindruckdifferential - Theorie und Technik*. Verlag Hans Huber, Bern, 1975.

[23] B. Berglund and A. Preis. Is perceived annoyance more subject-dependent than perceived loudness? *Acta Acust. United Acust.*, 83(2):313–319, 1997.

[24] B. Berglund, P. Hassmén, and A. Preis. Loudness and sharpness as determinants of noise similarity and preference. In F. A. Hil and R. Lawrence, editors, *Proceedings Internoise 1996 - Liverpool*, volume 4, pp. 2099–2104. Institute of Acoustics, 1996.

[25] B. Berglund, U. Berglund, and T. Lindvall. Scaling loudness, noisiness, and annoyance of aircraft noise. *J. Acoust. Soc. Am.*, 57(4): 930–934, 1975. DOI: 10.1121/1.380535.

[26] B. Berglund, U. Berglund, and T. Lindvall. Scaling loudness, noisiness, and annoyance of community noises. *J. Acoust. Soc. Am.*, 60(5):1119–1125, 1976. DOI: 10.1121/1.381212.

[27] B. Berglund, U. Berglund, M. Goldstein, and T. Lindvall. Loudness (or annoyance) summation of combined community noises. *J. Acoust. Soc. Am.*, 70(6):1628–1634, 1981. DOI: 10.1121/1.387229.

[28] B. Berglund, A. Preis, and K. Rankin. Relationship between loudness and annoyance for ten community sounds. *Env. Int.*, 16:523–531, 1990.

[29] P. Bergman, A. Sköld, D. Västfjäll, and N. Fransson. Perceptual and emotional categorization of sound. *J. Acoust. Soc. Am.*, 126 (6):3156–3167, 2009. DOI: 10.1121/1.3243297.

[30] P. Bernard. Brüel & Kjaer application notes - Leq, SEL - What? Why? When? Technical report, Brüel & Kjaer, 1985.

[31] E. A. Björk. The perceived quality of natural sounds. *Acta Acust. United Acust.*, 57(3):185–190, 1985.

[32] H. Blanton and J. Jaccard. Arbitrary metrics in psychology. *Am. Psychol.*, 61(1):27–41, 2006. DOI: 10.1037/0003-066X.61.1.27.

[33] J. Blauert and R. Guski. Critique of pure psychoacoustics. In *NAG/DAGA 2009, International Conference on Acoustics, Rotterdam*, pp. 1518–1519. DEGA e.V., Berlin, ISBN: 978-3-9808659-6-8, 2009.

[34] J. Blauert and U. Jekosch. Sound-Quality Evaluation - A Multi-Layered Problem. *Acta Acust. United Acust.*, 83(5):747–753, 1997.

[35] J. Blauert and U. Jekosch. Concepts behind sound quality: Some basic considerations. In *Proceedings of Internoise 2003, Seogwipo, Korea*, pp. 73–79, 2003.

[36] J. Blauert and U. Jekosch. A layer model of sound quality. *J. Audio Eng. Soc.*, 60(1/2):4–12, 2012.

[37] P. Block. Experimantal study of the effects of installation on single- and counter-rotation propeller noise. Technical paper 2541, NASA, 1986.

[38] J. Bortz and N. Döring. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler.* Springer-Verlag Berlin Heidelberg, 4 edition, 2006. DOI: 10.1007/978-3-540-33306-7.

[39] M. M. Bradly and P. J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry*, 25(1):49–59, 1994. DOI: 10.1016/0005-7916(94) 90063-9.

[40] W. Brennecke and H. Remmers. Physikalische Parameter bei der Bewertung der Lästigkeit von Industriegeräuschen (Physical parameters at the valuation of the annoyance of industrial noises). *Acta Acust. United Acust.*, 52(5):279–289, 1983.

[41] T. Bucak, E. Bazijanac, and B. Juricic. Correlation between SIL and SII in a light aircraft cabin during flight. In *Proc. 14th Int. Congress on Sound and Vibration ICSV*, pp. 1–8, 2007.

[42] A. Caclin, S. McAdams, B. K. Smith, and S. Winsberg. Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *J. Acoust. Soc. Am.*, 118(1):471–482, 2005. DOI: 10.1121/1.1929229.

[43] B. L. Cardozo and R. A. J. M. van Lieshout. Splitting sound annoyance into two components: level and sound character. Annual Progress Report 16, IPO - Instituut voor Perceptie Onderzoek - Institute for Perception Research, 1981.

[44] B. Cardozo and K. van der Veen. Estimation of annoyance due to low level sound. *Appl. Acoust.*, 12(5):389 – 396, 1979. DOI: 10.1016/0003-682X(79)90017-3.

[45] B. Cardozo and R. van Lieshout. Estimates of annoyance of sounds of different character. *Appl. Acoust.*, 14(5):323 – 329, 1981. DOI: 10.1016/0003-682X(81)90050-5.

[46] J. Chalupper and H. Fastl. Dynamic Loudness Model (DLM) for Normal and Hearing-Impaired Listeners. *Acta Acust. United Acust.*, 88(3):378–386, 2002.

[47] N. Chouard and T. Hempel. A semantic differential design especially developed for the evaluation of interior car sounds. In *Collected Papers from the Joint Meeting "Berlin 99" 137th regular meeting of the Acoustical Society of America 2nd convention of the EAA: Forum Acusticum - integrating the 25th German Acoustics DAGA Conference*, pp. 1–4. ISBN 3-9804568-5-4, Deutsche Gesellschaft für Akustik (DEGA), 1999.

[48] P. Daniel and R. Weber. Psychoacoustical roughness: Implementation of an optimized model. *Acta Acust. United Acust.*, 83(1): 113–123, 1997.

[49] DIN 1320. Acoustics - terminology. *Beuth, Berlin, Germany, German industry standard*, pp. 1–35, 1997.

[50] DIN 45631:1991. Procedure for calculationg loudness level and loudness. *Beuth, Berlin, Germany, German industry standard*, 1991.

[51] DIN 45631/A1:2008-01. Calculation of loudness level and loudness from the sound spectrum - Zwicker method - Amendment 1: Calculation of the loudness of time-variant sound. *Beuth, Berlin, Germany, German industry standard*, 2008.

[52] DIN 45645-1:1996. Determination of rating levels from measurement data - part 1: Noise immission in the neighbourhood. *Beuth, Berlin, Germany, German industry standard*, pp. pp. 1–9, 1996.

[53] DIN 45645-2:1997. Determination of rating levels from measurement data - part 2: Noise immissions at the workplace. *Beuth, Berlin, Germany, German industry standard*, pp. pp. 1–12, 1997.

[54] DIN 45681:2005–03. Acoustics—Detection of tonal components of noise and determination of tone adjustment for the assessment of noise immisions. *Beuth, Berlin, Germany, German industry standard*, 2005.

[55] DIN 45692:2009-08. Measurement technique for the simulation of the auditory sensation of sharpness. *Beuth, Berlin, Germany, German industry standard*, 2009.

[56] J. Dittmar. Cruise noise of counterrotating propeller at angle of attack in wind tunnel. Technical Memorandum TM 88869, NASA, 1986.

[57] J. Dittmar. The effect of front-to-rear- propeller spacing on the interaction noise of a modell counterrotating propeller at cruise conditions. Technical Memorandum TM 100121, NASA, 1987.

[58] J. Dittmar and D. Stang. Noise reduction for model counterrotating propeller at cruise by reducing aft-propeller diameter. Technical Memorandum TM 88936, NASA, 1987.

[59] J. Dittmar, E. Gordon, and R. Jeracki. The effect of front-to-rear-propeller spacing on the interaction noise at cruise conditions of a modell counterrotating propeller having a reduced diameter aft propeller. Technical Memorandum TM 101329, NASA, 1988.

[60] ECMA-74. Measurement of airborne noise emitted by information technology and telecommunications equipment. *European Computer Manufacturers Association, Standard*, 2012.

[61] R. Eilers. Abhängigkeit der beurteilten Unangenehmheit von der spektralen Lage und dem Modulationsgrad bei amplitudenmoduliertem Bandpassrauschen (Dependency of the judged unpleasantness from the spectral location and the degree of modulation

for amplitude modulated band-pass noise). Studienarbeit, Universität Oldenburg, Akustik, August 2007. Betreuung: R. Weber.

[62] W. Ellermeier, A.Zeitler, and H.Fastl. Predicting annoyance judgments from psychoacoustic metrics: Identifiable versus neutralized sounds. In *Proc. Internoise 2004, Prague, Czech republic*. I-INCE, 2004.

[63] W. Ellermeier, A. Zeitler, and H. Fastl. Impact of source identifiability on perceived loudness. In *Proceedings ICA 2004*, 2004.

[64] W. Ellermeier, J. Hellbrück, A. Kohlrausch, and A. Zeitler. *Kompendium zur Durchführung von Hörversuchen in Wissenschaft und industrieller Praxis (Compendium for the realization of listening tests in science and industrial practice)*. Deutsche Gesellschaft für Akustik e.V., 2008.

[65] T. M. Elliott, L. S. Hamilton, and F. E. Theunissen. Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *J. Acoust. Soc. Am.*, 133(1):389–404, 2013. DOI: 10.1121/1.4770244.

[66] E. Envia. Open rotor aeroacoustic modelling. Technical Report NASA/TM—2012-217740, NASA Glenn Research Center, Cleveland, Ohio, 2012. Prepared for the 15th International Conference on Fluid Flow Technologies (CMFF'12) sponsored by the Budapest University of Technology and Economics (BME) Budapest, Hungary, September 4–7, 2012.

[67] E. Envia. Aeroacoustic analysis of a high-speed open rotor. *Int. J. Aeroacoust.*, 14(3-4):569–606, 2015. DOI: 10.1260/1475-472X.14.3-4.569.

[68] M. Epstein and M. Florentine. A test of the equal-loudness-ratio hypothesis using cross-modality matching functionsa). *J. Acoust. Soc. Am.*, 118(2):907–913, 2005. DOI: 10.1121/1.1954547.

[69] S. D. Ewert. AFC - a modular framework for running psychoacoustic experiments and computational perception models. In *AIA-DAGA 2013, International Conference on Acoustics, Merano*, pp. 1326–1329. DEGA e.V., Berlin, ISBN: 978-3-939296-05-8, 2013.

[70] H. Fastl. Neutralizing the meaning of sound for sound quality evaluations. In *Proceedings of the 17th International Congress on Acoustics, ICA 2001, Rome, italy*, pp. 1–2, 2001.

[71] H. Fastl and F. Völk. Inharmonicity of sounds from electric guitars: Physical flaw or musical asset? In *Proc. ICMPC 2008, Sapporo, Japan*, pp. 1–5, 2008.

[72] H. Fletcher and W. A. Munson. Loudness, its definition, measurement and calculation. *J. Acoust. Soc. Am.*, 5(2):82–108, 1933. DOI: 10.1121/1.1915637.

[73] H. Fletcher, E. D. Blackham, and R. Stratton. Quality of piano tones. *J. Acoust. Soc. Am.*, 34(6):749–761, 1962.

[74] H. Fletcher, E. D. Blackham, and D. A. Christensen. Quality of organ tones. *J. Acoust. Soc. Am.*, 35(3):314–325, 1963.

[75] N. H. Fletcher. Mode locking in nonlinearly excited inharmonic musical oscillators. *J. Acoust. Soc. Am.*, 64(6):1566–1569, 1978.

[76] M. Florentine, S. Buus, and P. Bonding. Loudness of complex sounds as a function of the standard stimulus and the number of components. *J. Acoust. Soc. Am.*, 64(4):1036–1040, 1978. DOI: 10.1121/1.382062.

[77] M. Florentine, S. Buus, and T. Poulsen. Temporal integration of loudness as a function of level. *J. Acoust. Soc. Am.*, 99(3):1633–1644, 1996. DOI: 10.1121/1.415236.

[78] B. Gabriel, B. Kollmeier, and V. Mellert. Influence of individual listener, measurement room and choice of test-tone levels on the shape of equal-loudness level contours. *Acta Acust. United Acust.*, 83(4):670–683, 1997.

[79] J. M. Geary. Consonance and dissonance of pairs of inharmonic sounds. *J. Acoust. Soc. Am.*, 67(5):1785–1789, 1980. DOI: 10.1121/1.384307.

[80] J. V. D. Geer, W. Levelt, and R. Plomp. The connotation of musical consonance. *Acta Psychol. (Amst.)*, 20(0):308 – 319, 1962. DOI: 10.1016/0001-6918(62)90028-8.

[81] G. A. Gescheider. *Psychophysics: the fundamentals, 3rd. ed.* Lawrence Erlbaum Associates, Inc., Mahwah, NJ, 1997.

[82] V. Gibiat. Phase space representations of acoustical musical signals. *J. Sound Vib.*, 123(3):529 – 536, 1988. DOI: 10.1016/S0022-460X(88)80168-8.

[83] B. R. Glasberg and B. C. J. Moore. A model of loudness applicable to time-varying sound. *J. Audio Eng. Soc.*, 50:331–341, 2002.

[84] M. Goldstein and A. Kjellberg. Annoyance and low frequency noise with different slopes of the frequency spectrum. *J.Low Freq. Noise Vib.*, 4(2):43–51, 1985.

[85] J. M. Grey. Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am.*, 61(5):1270–1277, 1977. DOI: 10.1121/1.381428.

[86] G. Grimm, V. Hohmann, and J.L.Verhey. Loudness of fluctuating sounds. *Acustica*, 88:359–368, 2002.

[87] R. Guski, U. Felscher-Suhr, and R. Schuemer. The concept of noise annoyance: How international experts see it. *J. Sound Vib.*, 223 (4):513 – 527, 1999. DOI: 10.1006/jsvi.1998.2173.

[88] R. Guski. Psychological methods for evaluating sound quality and assessing acoustic information. *Acustica*, 83:765–774, 1997.

[89] R. Guski and H.-G. Bosshardt. Gibt es eine "unbeeinflußte" Lästigkeit? - Does "unbiased annoyance" exist? *Z. Lärmbekämpfung*, 39:67–74, 1992.

[90] V. Gutin. On the sound field of a rotating propeller. Technical Memorandum 1195, NACA, 1948. Translation of "Über das Schallfeld einer rotierenden Luftschraube", Phys. Sowjetunion, Band 9, Heft 1, 1936.

[91] T. Haase and O. Unruh. Active control of counter-rotating open rotor interior noise in a dornier 728 experimental aircraft: Optimised sensor placement. *Acta Acust. United Acust.*, 102(2):361–372, 2016. DOI: 10.3813/AAA.918952.

[92] T. Haase, S. Algermissen, O. Unruh, and M. Misol. Experiments on active control of counter-rotating open rotor interior noise. *Acta Acust. United Acust.*, 100(3):448–457, 2014. DOI: 10.3813/AAA.918725.

[93] H. Hansen. *Tone-noise dichotomy - investigating tonal content magnitude & pitch strength.* PhD thesis, Carl von Ossietzky Universität, Oldenburg, 2010.

[94] H. Hansen and R. Weber. Semantic evaluations of noise with tonal components in japan, france, and germany: A cross-cultural comparison. *J. Acoust. Soc. Am.*, 125(2):850–862, 2009. DOI: 10.1121/1.3050275.

[95] D. B. Hanson and C. J. McColgan. Noise of counter-rotation propellers with nonsynchronous rotors. *J. Aircraft*, 22(12):1097–1099, 1985.

[96] D. Hanson. Noise of counter-rotation propellers. *J. Aircraft*, 22(7):609–617, 1985.

[97] W. M. Hartmann. *Signals, Sound, and Sensation.* Springer-Verlag Inc., New York, NY, 1998.

[98] P. Hassmen. Note on the relationship between loudness and annoyance. *Percept. Mot. Skills*, 79(3):1325–1326, 1994. ISSN 0031-5125. DOI: 10.2466/pms.1994.79.3.1325.

[99] W. Heeren, J. Rennies, and J. L. Verhey. Spectral loudness summation of nonsimultaneous tone pulses. *J. Acoust. Soc. Am.*, 130(6):3905–3915, 2011. DOI: 10.1121/1.3652866.

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

[101] J. Hellbrück. *Recent trends in hearing research. Festschrift for Seiichiro Namba*, chapter Category-subdivision scaling – a powerful tool in audiometry and noise assessment, pp. 317–336. BIS, Oldenburg, Germany, 1996.

[102] J. Hellbrück and W. Ellermeier. *Hören: Physiologie, Psychologie und Pathologie.* Hogrefe, Göttingen, 2 edition, 2004.

[103] O. Heller. Hörfeldaudiometrie mit dem Verfahren der Kategorienunterteilung (KU) (Audiometry of the hearing field by the method of category subdivision scaling (CS)). *Psychol. Beitr.*, 27: 478–493, 1985.

[104] R. Hellman and E. Zwicker. Why can a decrease in dB(A) produce an increase in loudness? *J. Acoust. Soc. Am.*, 82(5):1700–1705, 1987. DOI: 10.1121/1.395162.

[105] R. P. Hellman. Loudness, annoyance, and noisiness produced by single-tone-noise complexes. *J. Acoust. Soc. Am.*, 72(1):62–73, 1982. DOI: 10.1121/1.388025.

[106] R. P. Hellman. Growth rate of loudness, annoyance, and noisiness as a function of tone location within the noise spectrum. *J. Acoust. Soc. Am.*, 75(1):209–218, 1984. DOI: 10.1121/1.390397.

[107] R. P. Hellman. Perceived magnitude of two-tone-noise complexes: Loudness, annoyance, and noisiness. *J. Acoust. Soc. Am.*, 77(4): 1497–1504, 1985. DOI: 10.1121/1.392044.

[108] R. Hellman. Contribution of tonal components to the overall loudness, annoyance, and noisiness of noise. Contractor Report 3892, NASA, 1985.

[109] K. Hiramatsu, K. Takagi, and T. Yamamoto. A rating scale experiment on loudness, noisiness and annoyance of environmental sounds. *J. Sound Vib.*, 127(3):467 – 473, 1988. DOI: 10.1016/0022-460X(88)90371-9.

[110] K. Hiramatsu, K. Takagi, and T. Yamamoto. Experimental investigation on the effect of some temporal factors of nonsteady noise on annoyance. *J. Acoust. Soc. Am.*, 74(6):1782–1793, 1983. DOI: 10.1121/1.390263.

[111] S. Hohls, T. Biermeier, R. Blaschke, and S. Becker. Psychoakustische Bewertung von Fahrzeugklimatisierungsgeräuschen anhand von Probandentests (Psychoacoustic evaluation of HVAC noise in cars by jury tests). In *Fortschritte der Akustik - DAGA 2014, Oldenburg, Deutschland*, pp. 323–324. DEGA e.V., Berlin, ISBN: 978-3-939296-06-5, 2014.

[112] S. Hohls, T. Biermeier, R. Blaschke, and S. Becker. Psychoacoustic analysis of HVAC noise with equal loudness. In *Proc. Internoise 2014, Melbourne, Australia*, pp. 1–7. Austrian Acoustical Society, ISBN: 978-0-909882-04-4, 2014.

[113] V. Hongisto, D. Oliva, and L. Rekola. Subjective and objective rating of spectrally different pseudorandom noises - implications for speech masking design. *J. Acoust. Soc. Am.*, 137(3):1344–1355, 2015. DOI: 10.1121/1.4913273.

[114] J. Hots, J. Rennies, and J. L. Verhey. Loudness of sounds with a subcritical bandwidth: A challenge to current loudness models? *J. Acoust. Soc. Am.*, 134(4):EL334–EL339, 2013. DOI: 10.1121/1.4820466.

[115] J. Hots, J. Rennies, and J. L. Verhey. Loudness of subcritical sounds as a function of bandwidth, center frequency, and level. *J. Acoust. Soc. Am.*, 135(3):1313–1320, 2014. DOI: 10.1121/1.4865235.

[116] E. Öhrströhm, M. Björkman, and R. Rylander. Laboratory annoyance and different traffic noise sources. *J. Sound Vib.*, 70(3):333 – 341, 1980. DOI: 10.1016/0022-460X(80)90303-X.

[117] Y. Inukai, N. Nakamura, and H. Taya. Unpleasantness and acceptable limits of low frequency sound. *J. Low Freq. Noise V. A.*, 19 (3):135–140, 2000. DOI: 10.1260/0263092001492895.

[118] ISO 226:3003. Acoustics - Normal equal-loudness level contours. *International Organisation for Standardisation*, pp. 1–23, 2003.

[119] ISO 532-1977. Acoustics - Method for calculating loudness level. *International Organisation for Standardisation*, 1977.

[120] ISO1996-1:2003. Acoustics - description, measurement and assessment of environmental noise - part1: Basic quantities and assessment procedures. *International Organisation for Standardisation*, pp. 1–32, 2003.

[121] U. Jekosch. Basic concepts and terms of "quality", reconsidered in the context of product-sound quality. *Acta Acust. United Acust.*, 90(6):999–1006, 2004.

[122] W. Jesteadt. An adaptive procedure for subjective judgments. *Percept. Psychophys.*, 28(1):85–88, 1980.

[123] H. F. Kaiser and J. Rice. Little jiffy, mark iv. *Educ. Psychol. Meas.*, 34(1):111–117, 1974. DOI: 10.1177/001316447403400115.

[124] A. Kameoka and M. Kuriyagawa. Consonance theory part i: Consonance of dyads. *J. Acoust. Soc. Am.*, 45(6):1451–1459, 1969. DOI: 10.1121/1.1911623.

[125] A. Kameoka and M. Kuriyagawa. Consonance theory part ii: Consonance of complex tones and its calculation method. *J. Acoust. Soc. Am.*, 45(6):1460–1469, 1969. DOI: 10.1121/1.1911624.

[126] C. H. Kasess, A. Noll, P. Majdak, and H. Waubke. Effect of train type on annoyance and acoustic features of the rolling noise. *J. Acoust. Soc. Am.*, 134(2):1071–1081, 2013. DOI: 10.1121/1.4812771.

[127] D. H. Keefe and B. Laden. Correlation dimension of woodwind multiphonic tones. *J. Acoust. Soc. Am.*, 90(4):1754–1765, 1991. DOI: 10.1121/1.401656.

[128] J. S. Kerrick, D. C. Nagel, and R. L. Bennett. Multiple ratings of sound stimuli. *J. Acoust. Soc. Am.*, 45(4):1014–1017, 1969. DOI: 10.1121/1.1911487.

[129] A. Klein, C. Marquis-Favre, R. Weber, and A. Trollé. Spectral and modulation indices for annoyance-relevant features of urban road single-vehicle pass-by noises. *J. Acoust. Soc. Am.*, 137(3): 1238–1250, 2015. DOI: 10.1121/1.4913769.

[130] M. Kloppenburg, H.-J. Maempfel, and S. Weinzierl. The appropriateness of the psychoacoustic measures sharpness and roughness for the prediction of aesthetic impression caused by sound. In *24th Tonmeistertagung, Leipzig, Germany*, pp. pp. 1–8, 2006.

[131] B. Kollmeier, R. H. Gilkey, and U. K. Sieben. Adaptive staircase techniques in psychoacoustics: A comparison of human data and a mathematical model. *J. Acoust. Soc. Am.*, 83(5):1852–1862, 1988. DOI: 10.1121/1.396521.

[132] C. Krumhansl. Why is musical timbre so hard to understand? In S. Nielzen and O. Olsson, editors, *Structure and perception of electroacoustic sound and music*, volume 9, pp. 43–53. Amsterdam: Elsevier, 1989.

[133] K. D. Kryter and K. S. Pearsons. Judged noisiness of a band of random noise containing an audible pure tone. *J. Acoust. Soc. Am.*, 38(1):106–112, 1965. DOI: 10.1121/1.1909578.

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

[135] V. Kuznetsov. Noise control problems of passenger airplanes (a review). *Acoust. Phys.*, 49(3):241–262, 2003.

[136] O. Lartillot and P. Toiviainen. A matlab toolbox for musical feature extraction from audio. In *Proc. of the 10th Int. Conference on Digital Audio Effects*, pp. 1–8, 2007.

[137] J. Laver. *Principles of phonetics.* Cambridge Univ. Press, Cambridge, UK, 1995.

[138] L. J. Leibold, H. Tan, S. Khaddam, and W. Jesteadt. Contributions of individual components to the overall loudness of a multitone complex. *J. Acoust. Soc. Am.*, 121(5):2822–2831, 2007. DOI: 10.1121/1.2715456.

[139] T. Letowski. Sound quality assessment: Concepts and criteria. In *AES Convention 87*, Oct 1989.

[140] H. Levitt. Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.*, 49(2B):467–477, 1971. DOI: 10.1121/1.1912375.

[141] W. H. Lichte. Attributes of complex tones. *J. Exp. Psychol.*, 28 (6):455–480, 1941.

[142] R. Majjigi, K. Uenishi, and P. Gliebe. An investigation of counterrotating tip vortex interaction. Contract Report 185135, NASA, 1989.

[143] L. E. Marks and M. Florentine. *Springer Handbook of Auditory Research*, volume 37 - Loudness. Springer Science+Business Media, LLC, 2011.

[144] L. E. Marks and G. A. Gescheider. *Steven's Handbook of Experimental Psychology*, volume 4, chapter Psychophysical Scaling, pp. 91–138. John Wiley & Sons Inc., New York, NY, 3 edition, 2002.

[145] L. Marks. Magnitude estimation and sensory matching. *Percept. Psychophys.*, 43(6):511–525, 1988. DOI: 10.3758/BF03207739.

[146] C. Marquis-Favre, E. Premat, and D. Aubree. Noise and its effects a review on qualitative aspects of sound. part ii: Noise and annoyance. *Acta Acust. United Acust.*, 91(4):626–642, 2005.

[147] S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff. Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychol. Res.*, 58(3):177–192, 1995. DOI: 10.1007/BF00419633.

[148] D. McCurdy. Advanced turboprop aircraft flyover noise: annoyance to counter-rotating-propeller configurations with an equal number of blades on each rotor – preliminary results. Technical Memorandum 100612, NASA, 1988.

[149] D. McCurdy. Advanced turboprop aircraft flyover noise: annoyance to counter-rotating-propeller configurations with a different number of blades on each rotor – preliminary results. Technical Memorandum 100638, NASA, 1988.

[150] D. McCurdy. Annoyance caused by advanced turboprop aircraft flyover noise - single-rotating propeller configuration. Technical Paper 2782, NASA, 1988.

[151] D. McCurdy. Annoyance caused by advanced turboprop aircraft flyover noise - counter-rotating propeller configuration. Technical Paper 3027, NASA, 1990.

[152] T. S. Meese. Using the standard staircase to measure the point of subjective equality: A guide based on computer simulations. *Percept. Psychophys.*, 57(3):267–281, 1995.

[153] R. D. Melara and L. E. Marks. Interaction among auditory dimensions: Timbre, pitch and loudness. *Percept. Psychophys.*, 48(2): 169–178, 1990.

[154] S. Meunier, D. Habault, and G. Canevet. Auditory evaluation of sound signals radiated by a vibrating surface. *J. Sound Vib.*, 247 (5):897–915, 2001. DOI: 10.1006/jsvi.2001.3791.

[155] S. Meunier, G. Rabau, and E. Friot. Annoyance and loudness of pure tones in noise: application to active control of fan noise. In D. Cassereau, editor, *Proc. of the Joint Congress CFA/DAGA 2004*, pp. 753–754. DEGA e.V., Berlin, ISBN: 2-9521105-2-2, 2004.

[156] G. A. Miller. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psych. Review*, 63(2):81–97, 1956.

[157] A. Miskiewicz, T. Rogala, and J. Szepanska-Antosik. Perceived roughness of two simoultaneous harmonic complex tones. *Arch. Acoust.*, 32(3):737–748, 2007.

[158] J. Molino. Measuring human aversion to sound without verbal descriptors. *Percept. Psychophys.*, 16(2):303–308, 1974. ISSN 0031-5117. DOI: 10.3758/BF03203947.

[159] B. C. J. Moore, D. A. Vickers, T. Baer, and S. Launer. Factors affecting the loudness of modulated sounds. *J. Acoust. Soc. Am.*, 105(5):2757–2772, 1999. DOI: 10.1121/1.426893.

[160] K. Moore. *A Multiphonic Reappraisal and the Alto Saxophone Concerto Radial*. PhD thesis, Columbia University, 2014.

[161] M. Möser. *Technische Akustik (Technical Acoustics)*. Springer Berlin, 7 edition, 2007.

[162] W. A. Munson and J. E. Karlin. Isopreference method for evaluating speech-transmission circuits. *J. Acoust. Soc. Am.*, 34(6): 762–774, 1962. DOI: 10.1121/1.1918193.

[163] S. Namba and S. Kuwano. On the definition of annoyance in social surveys of noise problems in the case of cross-cultural comparison. In *Proc. Internoise 2000, Nice, France*, pp. 1–6, 2000.

[164] S. Namba, S. Kuwano, T. Hashimoto, B. Berglund, Z. D. Rui, A. Schick, H.Hoege, and M. Florentine. Verbal expression of emotional impression of sound: A cross cultural study. *J. Acoust. Soc. J. (E)*, 12(1):19–29, 1991.

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

[166] S. Namba. *Recent Topics in Environmental Psychoacoustics*, chapter The Evaluation of Sound Environment and Psychological Methods, pp. 1–32. Osaka Universty Press, 2007.

[167] S. Namba. On the psychological measurement of loudness, noisiness and annoyance: A review. *J. Acoust. Soc. J. (E)*, 8(6):211–222, 1987.

[168] S. Namba and S. Kuwano. Psychological study on leq as a measure of loudness of various kinds of noises. *J. Acoust. Soc. J. (E)*, 5(3): 135–148, 1984.

[169] S. Namba, S. Kuwano, and A. Schick. A cross-cultural study on noise problems. *J. Acoust. Soc. J. (E)*, 7(5):279–289, 1986.

[170] H. Niese. Beitrag zur Relation zwischen Lautstärke und Lästigkeit von Geräuschen (A contribution to the relationship between loudness and annoyance of noises). *Acustica*, 15:236–243, 1965.

[171] M. E. Nilsson. A-weighted sound pressure level as an indicator of short-term loudness or annoyance of road-traffic sound. *J. Sound Vib.*, 302(1-2):197–207, 2007. ISSN 0022-460X. DOI: 10.1016/j.jsv.2006.11.010.

[172] H. Nouri, F. Ravelet, and F. Bakir. Étude expérimentale de l'écoulement et de l'interaction entre deux rotors contrarotatifs subsoniques (Experimental study of the flow and the interaction in-between two counter-rotating subsonic axial-flow fans). *La Houille Blanche*, 2014(3):13–26, 2014.

[173] A. Oetjen, R. Weber, and J. L. Verhey. Rauigkeitsberechnung unter Berücksichtigung der Einhüllendenform. In *Fortschritte der Akustik - DAGA 2012, Darmstadt, Germany*, pp. 581–582. Dega e.V., Berlin, ISBN: 978-3-939296-04-1, 2012.

[174] A. Oetjen, U. Letens, S. van de Par, J. L. Verhey, and R. Weber. Roughness calculation for randomly modulated sounds. In *AIA-DAGA 2013, International Conference on Acoustics, Merano*, pp. 1122–1124. DEGA e.V., Berlin, ISBN: 978-3-939296-05-8, 2013.

[175] C. Osgood, G. Suci, and P. Tannenbaum. *The measurement of meaning*. University of Illinois Press, Champaign IL, 8 edition, 1971.

[176] E. Parizet. Comparison of two semantic differential test methods. *J. Acoust. Soc. Am.*, 105(2):1189–1189, 1999. DOI: 10.1121/1.425609.

[177] E. Parizet. Paired comparison listening tests and circular error rates. *Acta Acust. United Acust.*, 88(4):594–598, 2002.

[178] E. Parizet and S. Fingerhuth. Subjective evaluation of tonalness and relation between tonalness and unpleasantness. In *Proc. Internoise 2009, Ottawa, Canada*, pp. 1–8, 2009.

[179] A. B. Parry and D. Crighton. Prediction of counter-rotation propeller noise. In *Proc. 12th AIAA Aeroacoustic Conference*, 1989.

[180] R. Paulsen and A. Schmidt. Lästigkeit, Gestörtheit und Unangenehmheit - im Labor nur Lautheitsurteile? (Annoyance, disturbance and unpleasantness - in the laboratory only loudness judgments?). In *Fortschritte der Akustik - DAGA1998, Zürich, Swiss*, pp. 84–85. DEGA e.V., Berlin, ISBN: 3-9804568-3-8, 1998.

[181] T. H. Pedersen and C. L. Fog. Optimisation of perceived product quality. In H. Fastl and J. Scheuren, editors, *Proc. Euronoise 1998, Munich, Germany*, volume 2, pp. 633–638. DEGA, 1998.

[182] F. Pedrielli, E. Carletti, and C. Casazza. Just noticeable differences of loudness and sharpness for earth moving machines. *J. Acoust. Soc. Am.*, 123(5):3164–3164, 2008. DOI: 10.1121/1.2933219.

[183] S. Pennig, J. Quehl, and V. Rolny. Effects of aircraft cabin noise on passenger comfort. *Ergonomics*, 55(10):1252–1265, 2012.

[184] G. J. Perakis, I. H. Flindell, and R. H. Self. Towards roughness as an additional metric for aircraft noise containing multiple tones.

Acta Acust. United Acust., 99(5):828–835, 2013. DOI: 10.3813/AAA.918661.

[185] R. Plomp and H. J. M. Steeneken. Effect of phase on the timbre of complex tones. *J. Acoust. Soc. Am.*, 46(2B):409–421, 1969. DOI: 10.1121/1.1911705.

[186] I. Pollack. Loudness as a discriminable aspect of noise. *Am. J. Psychol.*, 62(2):285–289, 1949.

[187] E. C. Poulton. *Bias in quantifying judgements*. Lawrence Erlbaum Associates, 1989. ISBN 0-86377-105-X.

[188] A. Preis. Intrusive sounds. *Appl. Acoust.*, 20:101–127, 1987.

[189] A. Preis. Measurement of annoyance components. In F. A. Hil and R. Lawrence, editors, *Proceedings Internoise 1996 - Liverpool*, volume 5, pp. 2289–2294. Institute of Acoustics, 1996.

[190] C. Reckhardt, V. Mellert, and B. Kollmeier. Bestimmung von Isophonen mit einem Adaptiven Verfahren - Einfluss experimenteller Parameter auf die Ergebnisse (Determination of isophone curves with an adaptive procedure - Influence of experimental parameters on the results). In *Fortschritte der Akustik - DAGA1998, Zürich, Swiss*, pp. 450–451. DEGA e.V., Berlin, ISBN: 3-9804568-3-8, 1998.

[191] J. Rennies, J. L. Verhey, J. Chalupper, and H. Fastl. Modeling temporal effects of spectral loudness summation. *Acta Acust. United Acust.*, 95(6):1112–1122, 2009. DOI: doi:10.3813/AAA.918243.

[192] J. Rennies, M. Wächtler, J. Hots, and J. Verhey. Spectro-temporal characteristics affecting the loudness of technical sounds: Data and model predictions. *Acta Acust. United Acust.*, 101(6):1145–1156, 2015. DOI: 10.3813/AAA.918907.

[193] J. Ricouard, E. Julliard, M. Omais, V. Regnier, A. B. Parry, and S. Baralon. Installation effects on contra-rotating open rotor noise. In *16th AIAA/CAES Aeroacoustics Conference*, pp. 1–8, 2010.

[194] P. E. Riera, M. Proscia, and M. C. Eguia. A comparative study of saxophone multiphonics: Musical, psychophysical and spectral analysis. *J. New Music Res.*, 43(2):202–213, 2014.

[195] B. Rohrmann. Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung (Empirical studies of rating scales for social science research). *Zeitschrift für Sozialpsychologie*, 9:222–245, 1978.

[196] T. Ronnebaum and R. Weber. Lautstärke und Unangenehmheit von Schallen im Tieffrequentem Bereich (Loudness and unpleasantness of low frequency sounds). In *Fortschritte der Akustik - DAGA 1996, Bonn, Germany*, pp. 350–351. DEGA e.V., Berlin, ISBN: 3-9804568-1-1, 1996.

[197] J. A. Russel. A circumplex model of affect. *J. Pers. Soc. Psychol.*, 39(6):1161–1178, 1980.

[198] T. Saeki, S. Yamaguchi, T. Tamesue, and K. Oimatsu. A method to predict psychological response of noisiness to an acoustical stimulus with a prominent frequency component. *J. Sound Vib.*, 277 (1-2):299 – 307, 2004. DOI: 10.1016/j.jsv.2003.08.023.

[199] B. Scharf. Loudness of complex sounds as a function of the number of components. *J. Acoust. Soc. Am.*, 31(6):783–785, 1959. DOI: 10.1121/1.1907785.

[200] B. Scharf. Loudness summation and spectrum shape. *J. Acoust. Soc. Am.*, 34(2):228–233, 1962. DOI: 10.1121/1.1909174.

[201] J. Scheuren. Iso 532 – living and working with alternative loudness standards. In *Proc. Internoise 2014, Melbourne, Australia*, pp. 1–6. Austrian Acoustical Society, ISBN: 978-0-909882-04-4, 2014.

[202] J. Schlittenlacher and W. Ellermeier. Psychoacoustic evaluation of gear noise using category ratings of multiple attributes. In *Proc. Internoise 2013, Innsbruck, Austria*, pp. 1–5, 2013.

[203] R. T. Schumacher. Ab initio calculations of the oscillations of a clarinet. *Acustica*, 48(2):71–85, 1981.

[204] W. A. Sethares. Local consonance and the relationship between timbre and scale. *J. Acoust. Soc. Am.*, 94(3):1218–1228, 1993. DOI: 10.1121/1.408175.

[205] I. Silva and M. Florentine. Effect of adaptive psychophysical procedure on loudness matches. *J. Acoust. Soc. Am.*, 120(4):2124–2131, 2006. DOI: 10.1121/1.2336747.

[206] A. Sköld, D. Västfjäll, and M. Kleiner. Perceived sound character and objective properties of powertrain noise in car compartments. *Acta Acust. United Acust.*, 91(2):349–355, 2005.

[207] T. Snekkestad. The poetics of a multiphonic landscape. June 2015. URL http://torbensnekkestad.com/research/.

[208] J. G. Snodgrass. *Experimental Sensory Psychology*, chapter Psychophysics, pp. 17–67. Scott, Foresman and Company, IL, 1975.

[209] T. Sone, K. Izumi, S. Kono, Y. Suzuki, Y. Ogura, M. Kumagai, H. Miura, H. Kado, H. Tachibana, K. Hiramatsu, S. Namba, S. Kuwano, O. Kitamura, M. Sasaki, M. Ebata, and T. Yano. Loudness and noisiness of a repeated impact sound. *J. Acoust. Soc. J. (E)*, 8(6):249–261, 1987.

[210] R. Sotteck. Gehörgerechte Rauigkeit (Psychoacoustically adequate roughness). In *Fortschritte der Akustik - DAGA 1994*, pp. 1–4, 1994.

[211] J. Steffens. *Wieviel Realität braucht der Mensch? Untersuchungen zum Einfluss der Versuchsumgebung auf die Geräuschbeurteilung von Haushaltsgeräten.* PhD thesis, TU Berlin, Fak. 5, 2013.

[212] S. S. Stevens. The attributes of tones. *Proceedings of the National Academy of Sciences*, 20(7):457–459, 1934.

[213] S. S. Stevens. Matching functions between loudness and ten other continua. *Percept. Psychophys.*, 1(1):5–8, 1966.

[214] S. S. Stevens. *Psychophysics: Introduction to its perceptual, neural and social prospects.* John Wiley & Sons Inc., NY, 1975.

[215] A. Stürmer and R. A. D. Akkermans. Multidisciplinary analysis of CROR propulsion systems: DLR activities in the JTI SFWA project. *CEAS Aeronautical Journal*, 5:265–277, 2014. DOI: 10.1007/s13272-014-0105-4.

[216] P. Susini. Psychophysical validation of a proprioceptive device by cross-modal matching of loudness. *Acta Acust. United Acust.*, 86 (3):515–525, 2000.

[217] P. Susini, S. McAdams, and S. Winsberg. A multidimensional technique for sound quality assessment. *Acta Acust. United Acust.*, 85(5):650–656, 1999.

[218] P. Susini, S. McAdams, S. Winsberg, I. Perry, S. Vieillard, and X. Rodet. Characterizing the sound quality of air-conditioning noise. *Appl. Acoust.*, 65(8):763 – 790, 2004. DOI: 10.1016/j.apacoust.2004.02.003.

[219] P. Susini, G. Lemaitre, and S. McAdams. Psychological measurement for sound description and evaluation. In B. Berglund, G. B. Rossi, J. T. Townsend, and L. R. Pendrill, editors, *Measurement with Persons: Theory, Methods, and Implementation Areas*, chapter 11, pp. 227–253. New York: Psychology Press, 2011.

[220] P. Susini, O. Houix, and G. Saint Pierre. The effect of loudness on the perceptual representation of sounds with similar timbre. *Acta Acust. United Acust.*, 101(6):1174–1184, 2015. DOI: 10.3813/AAA.918910.

[221] Y. Suzuki, S. Kono, and T. Sone. An experimental consideration of the evaluation of environmental noise with tonal components. *J. Sound Vib.*, 127(3):475 – 484, 1988. DOI: 10.1016/0022-460X(88)90372-0.

[222] Y. Suzuki and H. Takeshima. Equal-loudness-level contours for pure tones. *J. Acoust. Soc. Am.*, 116(2):918–933, 2004. DOI: 10.1121/1.1763601.

[223] G. M. Swallowe, R. Perrin, G. Sattar, A. M. Colley, and D. J. Hargreaves. On consonance: Pleasantness and interestingness of four component complex tones. *Acta Acust. United Acust.*, 83(5):897–902, 1997.

[224] E. Terhardt. Über die durch amplitudenmodulierte Sinustöne hervorgerufene Hörempfindung (On the hearing sensation evoked by amplitude modulated sinusoidal tones). *Acustica*, 20(4):210–214, 1968.

[225] E. Terhardt. On the perception of periodic sound fluctuations (roughness). *Acustica*, 30(4):201–213, 1974.

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

[227] E. Terhardt. *Akustische Kommunikation.* Springer, Berlin, 1998.

[228] S. Töpken, J. L. Verhey, and R. Weber. Psychoakustische Bewertung von superponierten Tonkomplexen (Psychoacoustic evaluation of superimposed complex tones). In *Fortschritte der Akustik - DAGA 2010, Berlin, Germany*, pp. 589–590. DEGA e.V., Berlin, ISBN: 978-3-9808659-8-2, 2010.

[229] S. Töpken, J. L. Verhey, and R. Weber. Perceptual space, pleasantness and periodicity of multi-tone sounds. *J. Acoust. Soc. Am.*, 138(1):288–298, 2015. DOI: 10.1121/1.4922783.

[230] A. Trollé, C. Marquis-Favre, and A. Klein. Short-term annoyance due to tramway noise: Determination of an acoustical indicator of annoyance via multilevel regression analysis. *Acta Acust. United Acust.*, 100(1):34–45, 2014. DOI: 10.3813/AAA.918684.

[231] J. B. Tufts, M. R. Molis, and M. R. Leek. Perception of dissonance by people with normal hearing and sensorineural hearing loss. *J. Acoust. Soc. Am.*, 118(2):955–967, 2005. DOI: 10.1121/1.1942347.

[232] J. L. Verhey and B. Kollmeier. Spectral loudness summation as a function of duration. *J. Acoust. Soc. Am.*, 111(3):1349–1358, 2002. DOI: 10.1121/1.1451065.

[233] J. L. Verhey, J. Hots, T. Luft, S. Ringwelski, and U. Gabbert. A psychoacoustical evaluation of active and passive methods for noise reduction in automotive engineering. In *AIA-DAGA 2013, International Conference on Acoustics, Merano*, pp. 1772–1775. DEGA e.V., Berlin, ISBN: 978-3-939296-05-8, 2013.

[234] P. C. Vitz. Preference for tones as a function of frequency (Hertz) and intensity (Decibels). *Percept. Psychophys.*, 11(1):84–88, 1972.

[235] F. Völk, H. Fastl, M. Fruhmann, and S. Kerber. Psychoakustische Untersuchungen zu Inharmonizitäten von Gitarrenseiten (Psychoacoustic investigation of inharmonicities in guitar strings). In

Fortschritte der Akustik - DAGA 2006, Braunschweig, Germany, pp. 743–744. DEGA e.V., Berlin, ISBN: 3-9808659-2-4, 2006.

[236] W. Völk. Lautheit und Lästigkeit von Geräuschen (Loudness and annoyance of noises). In A. Schick, editor, *Akustik zwischen Physik und Psychologie*, pp. 63–67. Klett-Cotta, Stuttgart, 1981.

[237] G. von Bismarck. Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acustica*, 30:146–159, 1974.

[238] G. von Bismarck. Sharpness as an attribute of the timbre of steady state sounds. *Acustica*, 30:159–172, 1974.

[239] H. von Helmholtz. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik (On the sensations of tone as a physiological basis for the theory of music).* Friedrich Vieweg und Sohn, 1896.

[240] D. Västfjäll, M. Friman, T. Gärling, and M. Kleiner. The measurement of core affect: a swedish self-report measure derived from the affect circumplex. *Scand. J. Psychol.*, 43(1):19–31, 2002.

[241] D. Västfjäll, M. A. Gulbol, M. Kleiner, and T. Gärling. Affective evaluations of and reactions to exterior and interior vehicle auditory quality. *J. Sound Vib.*, 255(3):501 – 518, 2002. ISSN 0022-460X. DOI: 10.1006/jsvi.2001.4166.

[242] D. Västfjäll. Contextual influences on sound quality evaluation. *Acta Acust. United Acust.*, 90(6):1029–1036, 2004.

[243] D. Västfjäll. Emotional reactions to sounds without meaning. *Psychology*, 3(8):606–609, 2012. DOI: 10.4236/psych.2012.38091.

[244] D. Västfjäll. Emotional reactions to tonal and noise components of environmental sounds. *Psychology*, 4(12):1051–1058, 2013. DOI: 10.4236/psych.2013.412153.

[245] D. Västfjäll, M. Kleiner, and T. Garling. Affective reactions to interior aircraft sounds. *Acta Acust. United Acust.*, 89(4):693–701, 2003.

[246] D. Västfjäll, T. Gärling, and M. Kleiner. Preference for current mood, anticipated emotional reaction, and experienced emotional reaction. *Scand. J. Psychol.*, 45(1):27–36, 2004. DOI: 10.1111/j.1467-9450.2004.00375.x.

[247] J. Wang, F. Ravelet, and F. Bakir. Experimental comparison between a counter-rotating axial-flow fan and a conventional rotor-stator stage. In *10th European Turbomachinery Conference, Lappeenranta, Finland*, pp. 1–10, 2013.

[248] J. Wang, F. Ravelet, and F. Bakir. Influence of design parameters on the global performances of low-speed counter-rotating axial-flow fans. In *Proceedings of the ASME 2014 4th Joint US-European Fluids Engineering Division Summer Meeting and 11th International Conference on Nanochannels, Microchannels, and Minichannels*, pp. 1–9, 2014.

[249] R. L. Warner and R. A. Bentler. Thresholds of discomfort for complex stimuli: Acoustic and sound-quality predictors. *J. Speech Lang. Hear. Res.*, 45(5):1016–1026, 2002. DOI: 10.1044/1092-4388(2002/082).

[250] R. Weber and R. Eilers. Combined contribution of roughness and sharpness to the unpleasantness of modulated bandpass noise. In *Fortschritte der Akustik - DAGA 2007 - Stuttgart*, pp. 565–566, 2007.

[251] C. Weckmüller and S. Guérin. On the influence of trailing-edge serrations on open rotor tonal noise. In *DGLR- und DEGA- Fachausschusssitzung*, 2011.

[252] E. G. Wever and K. E. Zener. The method of absolute judgment in psychophysics. *Psychological Review*, 35(6):466–493, 1928.

[253] W. A. Yost. *Fundamentals of Hearing - An Introduction*, volume 3. Academic Press Inc., San Diego, CA, 1994.

[254] J. You and J. Y. Jeon. Just noticable differences of sound quality metrics of refrigerator noise. In *19th International Congress on Acoustics, Madrid*, 2007.

[255] D. E. V. Zante and E. Envia. Prediction of the aero-acoustic performance of open rotors. Technical Report NASA/TM—2014-218132, NASA, Glenn Research Center, Cleveland, Ohio, 2014.

[256] A. Zeitler. Kognitive Faktoren bei der Skalierung von Höreindrücken (Cognitive factors in the scaling of auditory sensations). In *Fortschritte der Akustik - DAGA 2007, Stuttgart, Germany*, pp. pp. 1–4. DEGA e.V., Berlin, ISBN: 978-3-9808659-3-7, 2007.

[257] A. Zeitler and J. Hellbrück. Sound quality assessment of everyday-noises bay means of psychophysical scaling. In S. Glegg and Y. Yong, editors, *Proc. Internoise 1999, Ft. Lauderdale*, pp. 1–6. INCE, USA, 1999.

[258] A. Zeitler and J. Hellbrück. Semantic attributes of environmental sounds and their correlations with psychoacoustic magnitudes. In *Proc. ICA 2001,Rome*, pp. 1–2. ISBN 88-88387-03-X, 2001.

[259] T. Zhou, M. Zhang, and C. Li. A model for calculating psychoacoustical fluctuation strength. *J. Audio Eng. Soc.*, 63(9):713–724, 2015. DOI: 10.17743/jaes.2015.0070.

[260] E. Zwicker. Ein Verfahren zur Berechnung der Lautstärke (A Method to calculate loudness levels). *Acta Acust. United Acust.*, 10(4):304–308, 1960.

[261] E. Zwicker. Ein Beitrag zur Unterscheidung von Lautheit und Lästigkeit (A contribution towards the distinction between loudness and annoyance). *Acustica*, 17:22, 1966.

[262] E. Zwicker. Ein Vorschlag zur Definition und zur Berechnung der unbeeinflussten Lästigkeit (A proposal for a definition and calculation of unbiased annoyance). *Z. Lärmbekämpfung*, 38:91–97, 1991.

[263] E. Zwicker and H. Fastl. *Psychoacoustics: Facts and Models 3rd ed.* Springer, Berlin/Heidelberg, 2007.

[264] J. J. Zwislocki. *Ratio Scaling of Psychological Magnitude - In Honor of the Memory of S. S. Stevens*, chapter Natural Measurement, pp. 18–26. Lawrence Erlbaum Associates Inc., NJ, 1991.

Acknowledgements

I would like to thank some people that contributed to the completion of this thesis.

First, i would like to thank Reinhard Weber for the continuous support and supervision of this PhD-project. Thank you for the many discussions we had, often starting from a "...may i briefly ask you something?" I hope you were not too desperate whenever i already started to think about potential consequences of a line of thought instead of just letting thoughts flow freely. Looking back, i am blessed to having had the chance to learn from your broad professional experience and vision. I hope, i can take your ideas further. I also like to thank you for always encouraging me to visit conferences and i happily look back to the trips and experiences we had together in Lisbon, in Osaka and in New York.

I further like to thank Prof. Steven van de Par for taking over the supervision. You started your position in the Acoustics group just after the start of my PhD-phase and the research area "Applied Psychoacoustics" was not your primary topic at that time. This has considerably changed during the last years and i appreciate the exchange of thoughts with you and the new ideas you brought into my work and also into the Acoustics group. I like to thank you for the mutual trust we share and cherish your persistent relaxed mood.

I also would like to thank Prof. Hugo Fastl from TU Munich for accepting the external review of my thesis and together with Prof. Volker Hohmann for the participation in the defense committee.

I would like to thank Henning Scheel from Airbus Operation for the trustworthy collaboration and the financial support in the several projects we did together. It is exciting to see how an engineering problem can evolve to a better and deeper general understanding in the field of applied psychoacoustics. Special thanks also go to those colleagues from the Acoustics group which were directly involved in these projects at that time: Thomas Biberger, Jens Alrik Adrian and my longterm office mate Andreas Escher. Andreas, i highly appreciate your engagement throughout the several measurement campaigns we carried out together. Your excellent work was a major contributor to the sound data basis of chapter 3 of this PhD-thesis.

Further thanks go to the whole Acoustics group and many colleagues from the neighboring working groups of the University of Oldenburg for the creative, multi-faced working environment which encouraged and inspired me through the sometimes stressful times. I also appreciate the diverse leisure activities with many colleagues beyond the daily work in the university.

I also would like to explicitly thank our technical assistant, Christoph Scheicht, for his continuous and invaluable technical support of our group. Your approaches on solving problems still astound me. Sometimes i have the impression as if you guessed or knew about some problems already and prepared solutions in advance. Nevertheless, we will continue discussing the "‘correct way"’ of handling cables.

I would like to thank my long-term personal friends. Thank you for providing the creative and musical playgrounds apart from the university environment and for sharing knowledge and creativity from which i definitely benefited and still benefit in the daily university work, especially *DLTFY* and *City Slickers*.

I would like thank my parents and my sister for giving me the most valuable and stable support i could imagine. Thank you for being the continuous foundation through the very different phases of my life from school, to apprenticeship, to finally studying and now concluding my PhD. I highly appreciate you and your contribution to making me the person i am today.

Last but not least i would like to thank Christina for the scientific discussions we had and have, the many journeys we already undertook and every private moment we share. Thank for your vivid nature and everything you bring into my life.