Zur Homepage der Dissertation

Timbre perception and object separation with normal and impaired hearing

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

Suzan Selma Emiroğlu
geboren am 24. Juni 1976
in Dachau

Gutachter: Prof. Dr. Dr. Birger Kollmeier
Zweitgutachter: Jun.-Prof. Dr. Jesko Verhey
Tag der Disputation: 18. Juli 2007

Abstract

Timbre is a combination of all auditory object attributes other than pitch, loudness and duration, and is used to distinguish different musical instruments or voices. People with sensorineural hearing loss often have problems with timbre distortion. Even for modern hearing aids it is difficult to provide good audio quality for speech intelligibility while preserving the natural timbre. This not only affects music perception, but may also influence object recognition in general. The present study aims to quantify differences in object segregation and timbre discrimination between normal-hearing and hearing-impaired listeners with a sensorineural hearing loss. In order to improve auditory models and hearing aids, a new method for studying timbre perception was developed. Using cross-faded (morphed) instrument sounds in psychoacoustic measurements, the subtle timbre perception differences between listener groups are studied.

In order to characterize timbre perception differences, rating measurements were performed, in which normal-hearing and hearing-impaired subjects judged the similarity of the presented morphed sounds. When stimuli were amplified to provide intermediate loudness impressions in all subjects, most hearing-impaired subjects gave ratings similar to those of normal-hearing listeners. Only a few subjects showed distinct rating deviations from normal-hearing listeners. In order to verify subtle perception differences, the morphed stimuli were combined with discrimination measurements. Experiments with normal-hearing musicians and non-musicians showed that the new method enables objective determination of a value that provides a comparison between different subject groups and timbres: a just noticeable difference (JND) of timbre. For the discrimination measurements, the attack portion of the sound was cut off, which minimizes recognition of the sounds and thus makes the method independent of subjects' previous knowledge.

Discrimination measurements with normal-hearing and hearing-impaired listeners aim to quantify differences in object segregation and timbre discrimination, investigating timbre JNDs in silence and different background-noise conditions, on different sound levels and in subjects with different hearing loss configurations. The results indicate that at intermediate levels JNDs of subjects with flat or diagonal hearing loss are similar to those of normal-hearing listeners, when an appropriate linear sound amplification is provided. This contradicts the common hypothesis that hearing-impaired people generally have more problems in distinguishing different timbres, for example, due to reduced frequency selectivity. However, subjects with a steep hearing loss show significantly higher JNDs than normal-hearing listeners, both in silence and in noise. In the condition testing transferability from silence to

noise, no significant JND differences across listener groups were found, which contradicts the hypothesis that hearing-impaired listeners generally have more problems in object segregation than normal-hearing listeners.

JNDs and similarity ratings of all subjects show distinct variation across instrument continua, which is discussed in the context of common timbre models. Using spectro-temporal timbre descriptors, measurement results can be explained by primary factors involved in sensorineural hearing loss, that is attenuation and loss of compression. On one hand, insufficient sound amplification and severe hearing-loss at frequencies above 2 kHz may cause problems in distinguishing the attack and spectral centroid of the sound. On the other hand, due to compression loss, enhanced internal intensity differences may lead to enhanced perceptual differences of the spectral centroid in hearing-impaired listeners. In order to objectively predict the results independent from percept, the psychoacoustic measurements of the present study are simulated with the Perception Model *PeMo* for the normal and impaired hearing system, which had been evaluated for nearly all basic psychoacoustic experiments. Simulations with this effective model confirm quantitatively the effects of hearing loss and different timbres on discrimination thresholds. However, a crucial factor for both the perception-descriptive timbre model and the effective computer model, seems to be the unclear perceptual weighting of temporal and spectral changes in the sound. Approaching this unsolved question may be an important task for future studies modeling timbre perception.

The present study shows that, as opposed to reduced ability of hearing-impaired listeners to separate natural objects due to a reduction in time and frequency resolution, certain timbre dimensions seem to not be degraded by compression loss and might provide hearing-impaired listeners with cues for separating objects when linear sound amplification is provided. Lowering the distortion connected to non-linear amplification in hearing aids may not only enhance the pleasure of listening to music but also support the user's ability to separate objects.

Zusammenfassung

Klangfarbe verbindet alle Hörobjektmerkmale, die nicht Tonhöhe, Lautheit und Länge sind, und dient dazu, Musikinstrumentenklänge oder Stimmen zu unterscheiden. Menschen mit einem sensorineuralen Hörverlust haben oft Probleme mit einer Klangfarbenverzerrung. Sogar mit modernen Hörgeräten ist es schwierig, eine gute Audioqualität für die Sprachverständlichkeit zu erreichen und gleichzeitig die natürliche Klangfarbe zu erhalten. Dies hat nicht nur Auswirkungen auf die Musikwahrnehmung, sondern kann auch die Objekterkennung im Allgemeinen beeinflussen. Ziel der vorliegenden Studie ist es, Unterschiede in der Objekttrennung und Klangfarbenunterscheidung zwischen Normalhörenden und Schwerhörenden mit sensorineuralem Hörverlust zu quantifizieren. Im Hinblick auf die Verbesserung von Hörmodellen und Hörgeräten wird eine neue Methode entwickelt, um die Klangfarbenwahrnehmung zu untersuchen. Unter der Verwendung von übergeblendeten (gemorphten) Instrumentenklängen in psychoakustischen Messungen werden die feinen Klangfarbenwahrnehmungsunterschiede zwischen den Hörergruppen untersucht.

Um die Klangfarbenwahrnehmungsunterschiede zu charakterisieren wurde ein Paarvergleich durchgeführt, in dem normal- und schwerhörende Probanden die Ähnlichkeit der gemorphten Klänge bewerteten. Dabei wurden die Stimuli so verstärkt, dass bei allen Probanden ein mittlerer Lautheitseindruck entstand. Die meisten schwerhörenden Probanden gaben ähnliche Wertungen wie die normalhörenden an. Nur wenige Probanden gaben Bewertungen ab, die deutlich von denen der Normalhörenden abwichen. Um die geringfügigen Wahrnehmungsunterschiede zu erfassen, wurden die gemorphten Signale mit Diskriminationsmessungen kombiniert. Wie Experimente mit normalhörenden Musikern und Nichtmusikern zeigten, lässt die neue Methode objektiv eine Gröe bestimmen, die einen Vergleich zwischen unterschiedlichen Probandengruppen und Klangfarben ermöglicht: ein gerade-noch-wahrnehmbarer Unterschied (JND) der Klangfarbe. Für die Diskriminationsmessungen wurde der Einschwingvorgang der Klänge abgeschnitten, was die Bestimmung des Instruments minimiert und so die Methode unabhängig vom Vorwissen der Probanden macht.

Diskriminationsmessungen mit Normal- und Schwerhörenden sollen die Unterschiede in Objekttrennung und Klangfarbenunterscheidung quantifizieren, indem sie Klangfarben-JNDs in Ruhe und unterschiedlichen Störgeräuschbedingungen, bei unterschiedlichen Pegeln und von Probanden mit unterschiedlichen Hörverlustkonfigurationen untersuchen. Die Ergebnisse weisen darauf hin, dass die JNDs der Probanden mit flachem oder diagonalem Hörverlust bei mittleren Pegeln ähnlich zu

denen der Normalhörenden sind, wenn eine angemessene lineare Verstärkung angeboten wird. Dies widerspricht der verbreiteten Hypothese, dass schwerhörende Menschen, beispielsweise verursacht durch eine verringerte Frequenzauflösung, allgemein mehr Probleme haben, unterschiedliche Klangfarben zu unterscheiden. Probanden mit einen steilen Hörverlust zeigen jedoch sowohl in Ruhe als auch im Störgeräusch signifikant höhere JNDs als Normalhörende. In der Messbedingung, die die Transferleistung von Ruhe ins Störgeräusch prüft, werden keine signifikanten Unterschiede zwischen den Probandengruppen gefunden. Dies widerspricht der Hypothese, dass Schwerhörende generell mehr Probleme mit der Objekttrennung als Normalhörende haben.

Die JNDs und Ähnlichkeitbewertungen aller Probanden unterscheiden sich deutlich zwischen den Instrumentenkontinua, was im Kontext der allgemeinen Klangfarbenmodelle diskutiert wird. Unter der Verwendung von spektro-temporalen Klangfarben-"Deskriptoren" können die Messergebnisse durch Primärfaktoren für sensorineuralen Hörverlust erklärt werden, d.h. Intensitätsabschwächung und Dynamikkompressionsverlust. Auf der einen Seite können ungenügende Klangverstärkung und starker Hörverlust oberhalb von 2kHz Probleme verursachen, den Einschwingvorgang und den spektralen Schwerpunkt zu unterscheiden. Andererseits können erhöhte interne Intensitätsunterschiede zu erhöhten wahrgenommenen Unterschieden des spektralen Schwerpunkts in Schwerhörenden führen. Um die Ergebnisse objektiv und unabhängig vom Perzept vorhersagen zu können, werden die psychoakustischen Messungen der vorliegenden Studie mit dem Perzeptionsmodell *PeMo* für das normale und beeinträchtigte Hörsystem simuliert, welches für nahezu alle grundlegenden psychoakustischen Experimente evaluiert ist. Simulationen mit diesem Effektivmodel bestätigen quantitativ die Auswirkungen von Hörverlust und unterschiedlichen Klangfarben auf die Diskriminationsschwellen. Ein entscheidender Faktor sowohl für das Perzept-beschreibende Model als auch für das effektive Computermodel scheint die unklare Gewichtung der zeitlichen und spektralen Veränderungen im Klang zu sein. Für zukünftige Studien zur Klangfarbenmodellierung kann es eine wichtige Aufgabe sein, diese ungelöste Frage zu verfolgen.

Im Gegensatz zur eingeschränkten Fähigkeit von Schwerhörenden, aufgrund von (verringerter) Zeit- und Frequenzauflösung natürliche Objekte zu trennen, scheinen bestimmte Klangfarbendimensionen nicht durch Kompressionsverlust beeinträchtigt zu sein, wie die vorliegende Studie zeigt. Bei linearer Verstärkung könnten diese Klangfarben Schwerhörenden helfen, Objekte zu trennen. Eine Reduktion der Verzerrung, die durch nicht-lineare Verstärkung in Hörgeräten verursacht wird, könnte nicht nur die Hörfreude an Musik fördern, sondern auch den Hörgeräteträger bei der Objekttrennung unterstützen.

Contents

Chapter 1

General introduction

Timbre is not only a colourful sound attribute that gives joy to music perception, but is also used to distinguish acoustical objects like musical instrument sounds and different voices. People with sensorineural hearing loss often have problems with timbre distortion, which affects not only music perception, but also object recognition in general. Even for modern hearing aids it is difficult to provide good audio quality necessary for speech intelligibility by preserving the natural timbre. Special features like noise-reduction algorithms, which are doubtless necessary, inevitably distort the timbre of a sound. The present study aims to quantify differences in object segregation and timbre discrimination between normal-hearing listeners and people with a sensorineural hearing loss. In order to improve auditory models and hearing aids, a new method to study timbre was developed. Using cross-faded ("morphed") instrument sounds in similarity rating and discrimination experiments, the subtle timbre perception differences between listener groups are studied. In correlation with previous studies, the newly established method is brought into the context of common timbre models and the measurement results are discussed in the context of existing theories on timbre and hearing loss.

Timbre and perception-descriptive timbre models

The label *timbre* combines all auditory object attributes other than pitch, loudness, duration, spatial location and reverberation environment. The *physical* timbre space is made up of frequency, time and amplitude of sound, which are the fundamental measures of acoustics, while the timbre *perception* is multidimensional with descriptions like brightness, roughness and noisiness. Previous timbre studies tried to find a timbre *model* by connecting physics and perception, that is, to find psychophysical quantities that represent timbre. Similarity rating measurements and subsequent

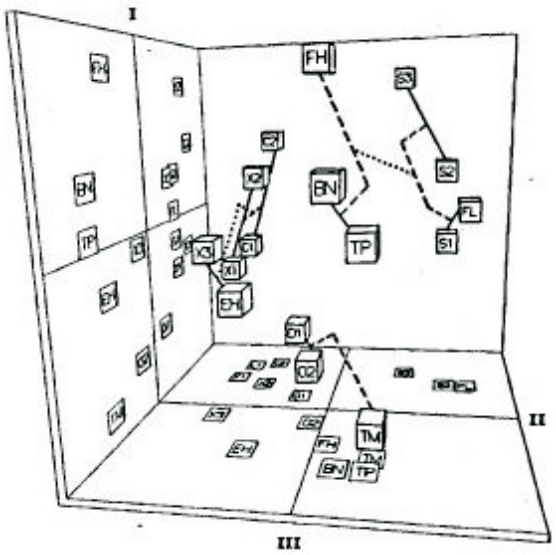


Figure 1.1: Multidimensional scaling (MDS) of timbre ratings of 16 musical instruments (Grey, 1977). FH: French horn, TM: trombone, S1-S3: cellos as string instruments, X1-X3: saxophones, FL: flute, TP: trumpet, EH: English horn, C1-C3: clarinets, O1-O2: oboes, BN: bassoon. Interpretation of dimensions: spectral centroid (axis towards up, I), overtone synchronicity (axis towards right, II), high-energy in the attack segment (axis towards reader, III).

multidimensional scaling (MDS) can identify timbre dimensions that dominate our perception. Figure 1.1 shows a 3-dimensional MDS space, in which distances account for perceived similarity differences of 16 musical timbres. For 30 years MDS studies have revealed various timbre dimensions of musical instrument sounds (e.g., Grey, 1977; Grey & Gordon, 1978; Krumhansl, 1989; Iverson & Krumhansl, 1993; McAdams et al., 1995; Lakatos, 2000). The perceptual dimensions are represented by spectro-temporal timbre descriptors, which are linear combinations of physical fundamental measures. Possible descriptors of timbre dimensions are shown in the blocks of Figure 1.2. However, the exact definition of physical dimensions used within the MDS varies considerably across studies. The descriptors shown in Figure 1.2 are not independent from each other and, for example, a set of orthogonal timbre dimensions other than those describing the axis in Figure 1.1 may be able to explain the results of Grey's (1977) experiment. However, the low number of instruments used in common studies allows only an approximation of the timbre cues used by subjects, for example, the 16 instruments in Grey's (1977) MDS study allowed 3 timbre dimensions with a residual MDS tension. Since timbre is multidimensional, a different set of stimuli may lead to a different set of descriptors that dominate the ratings. In an attempt to re-interpret the results with a uniform set of acoustic descriptor families, McAdams et al. (1995), McAdams & Winsberg (2000) and Levitin et al. (2002) collected old and new data and applied appropriate measures dependent on instrument (family) and subject classes.

A timbre model can be described as a combination of weighted spectro-temporal descriptors, whereby weighting may depend on instrument group and subject class. Figure 1.2 sketches a possible model with the common descriptors as building blocks,

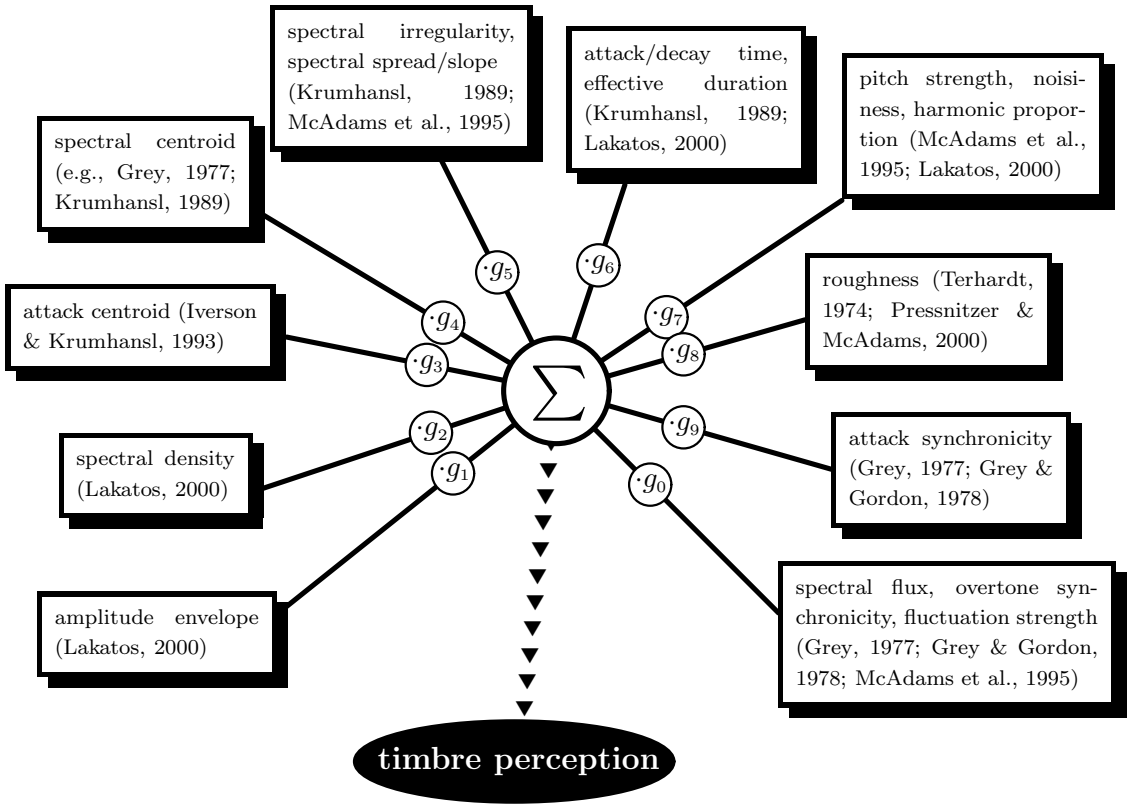


Figure 1.2: A hypothetical timbre model combining all g_k-weighted common spectro-temporal timbre descriptors from the literature in order to predict similarity ratings across sounds that differ in timbre.

which are summed up in individual g_k-weights and transferred into an objective rating matrix or discrimination threshold value. (Note that the hypothetical model in Figure 1.2 would need some weights be set to g_k=0, because blocks are not independent from each other.) An optimal model that simultaneously accounts for all musical instruments would predict similarity ratings and discrimination thresholds of timbre measurements, both for normal-hearing and hearing-impaired subjects. Since no unique model has been established, in the present study the most common timbre descriptors known from the literature are used to interpret the measurement results, particularly with respect to differences between normal-hearing and hearing-impaired subjects (Chapters 2).

Morphing as a new method for timbre rating and discrimination

In order to improve auditory models and hearing aids, a new method to study timbre was developed. By linear interpolation of spectral parameters, sounds of musical instruments are morphed, thus generating stimulus continua between natural instruments. Using the morphed stimuli in timbre rating experiments, Chapter 2 presents measurements, in which the subjects judged the similarity of the presented sounds. Chapter 3 gives a detailed description of the morphing method and evaluates it as a method for timbre *discrimination* studies measuring just noticeable differences (JND) along continua of morphed musical instruments. While Chapter 2 uses the entire sounds, for Chapter 3 as well as Chapter 4, the attack portion of the sounds was cut off, which minimizes recognition of the sounds. The experiments in Chapter 3 determine JNDs of timbre in normal-hearing subjects with different musical experience. By measuring the JNDs along different timbre dimensions and relating the JND variation to spectro-temporal descriptors, the newly established method is brought into the context of common timbre models.

Object separation in normal-hearing and hearing-impaired listeners

A common hypothesis argues that the reduced frequency selectivity in hearing-impaired people leads to a reduced ability to distinguish timbre and, hence, sounds of musical instruments (Moore, 2003). It is still unproven whether and for which conditions this statement holds true; that is, if and how the ability changes for different types and severities of hearing loss, with different sound types, and in the presence of other sounds. Since timbre is an object attribute used to distinguish acoustical objects, a reduced ability in timbre discrimination may also affect object separation in general. While the negative influence of a sensorineural hearing impairment has been proved in nearly all psychoacoustically ascertainable hearing functions (e.g. Festen & Plomp, 1983; Moore, 1998), the influence of the disturbed psychoacoustic functions on speech intelligibility in silence and in noise and on general object segregation is not yet resolved unambiguously. It is commonly accepted, however, that the alteration in the compressive nonlinearity caused by outer hair cell loss is a major cause of most of the perceptual changes observed in cochlear hearing loss (Bacon et al., 2004). In order to study the consequences of the compressive non-linearity that is altered in hearing-impaired listeners with regards to object segregation, in the present study psychoacoustic measurements are performed with the object feature *timbre*, which is also used to separate auditory objects (Iverson, 1995). While Chapter 2 characterizes coarse timbre perception differences between normal-hearing and hearing-impaired listeners using rating experiments, Chapter 4 quantifies differences in object segregation and timbre discrimination. The exper-

iments in Chapter 4 investigate timbre JNDs in silence and different background-noise conditions, on different sound levels and in subjects with different hearing loss configurations.

Computer model

In Chapters 2, 3 and 4 the spectro-temporal timbre descriptors of the common timbre models are used to interpret the measurement results. Although these models are able to successfully describe certain timbre dimensions that influence perception (e.g., Grey & Gordon, 1978; Krumhansl, 1989; Iverson & Krumhansl, 1993), no uniform set of timbre measures seems to account for all instruments (McAdams & Winsberg, 2000; Levitin et al., 2002). The new method using morphed sounds for timbre measurements and analysis may help future studies to find the optimal set of timbre descriptors. However, in the present study, Chapter 5 approaches a timbre model from the physical side. Instead of describing the timbre percept, which may be different for individual subjects and hearing losses, timbre discrimination thresholds are modeled by fundamental physical measures. Using a model that is validated for basic perception limens and implementing only primary factors of hearing loss, Chapter 5 aims to predict physiological limits of timbre discrimination independent of any categories of the percept. In order to predict subjective timbre similarity ratings and discrimination thresholds with a computer model, in Chapter 5 the psychoacoustic measurements of Chapters 2, 3 and 4 are simulated using a modified version of the effective Perception Model *PeMo* for the normal and impaired hearing system (Dau et al., 1996; Derleth et al., 2001).

Summary

This study aims to characterize differences in timbre perception and object separation between normal-hearing and sensorineural hearing-impaired listeners. In order to improve auditory models and hearing aids, a new method to study timbre is presented using morphed instrument sounds in psychoacoustic measurements. While Chapter 2 presents timbre rating measurements which characterize perception differences between normal-hearing and hearing-impaired subjects, Chapter 3 gives a detailed description of the morphing method and evaluates it as a new method for timbre *discrimination* studies. In an attempt to quantify differences in object segregation and timbre discrimination between normal-hearing and hearing-impaired listeners, the experiments in Chapter 4 investigate timbre JNDs in silence and different background-noise conditions, on different sound levels and in subjects with different hearing loss configurations. While in Chapters 2, 3 and 4, results are dis-

cussed in the context of common timbre models that describe the timbre percept, in Chapter 5 the psychoacoustic measurements of Chapters 2, 3 and 4 are simulated using an effective auditory computer model for the normal and impaired hearing system.

Amongst the scientific goals, the present thesis aims to provide insight and better understanding of the complex sound attribute *timbre*. Therefore, endnotes and an appendix contain additional ideas and explanations that are beyond the main argumentation line but may satisfy curiosity and may be helpful for those who work in related research. In Appendix A timbre and its dimensions found in previous studies are introduced in detail and brought into the context of the morphed sounds used in Chapter 2. Appendix B tries to explain the results of the measurements in Chapter 4 in the context of compression (loss) and its secondary effects. Appendix C illustrates internal representations of the simulations in Chapter 5 and Appendix D contains extra notes marked with superscript numbers[1] in the text.

Chapter 2

Similarity rating on timbre perception in hearing-impaired and normal-hearing listeners

Abstract

People with sensorineural hearing loss often have problems with timbre distortion. In an attempt to characterize differences in perception between normal-hearing and hearing-impaired listeners in terms of spectro-temporal dimensions, timbre rating experiments, in which the subjects judged the similarity of the presented sounds, were performed with both groups of listeners. By linear interpolation of spectral parameters, sounds of musical instruments were cross-faded ("morphed"), whereby stimulus continua between natural instruments were generated for the different instrument pairs. Timbre variance along the first continuum is dominated by spectral centroid, the second continuum mainly varies by the attack, while spectral flux, noisiness and attack vary along the third continuum. Rated distance by normal-hearing subjects depends mainly on the physical distance of the presented sound pair. However, in continua in which the crucial timbre dimension changes along the continuum, rated distance - compared to physical distance - varies slightly along the continuum. Most hearing-impaired subjects, for which the presentation level was 3-30 dB higher, show in all instrument continua similar judgments to normal-hearing subjects. However, along the first and third continua the mean rated distance of the hearing-impaired subjects shows a higher variance compared to the physical distance. Two hearing-impaired subjects show problems distinguishing the stimuli of the second continuum. Differences between listener groups seem to be connected to reduced ability to use high-frequency energy present in the attack as discrimination cue. In addition, perceptual differences of spectral centroid appears to be enhanced in hearing-impaired listeners.

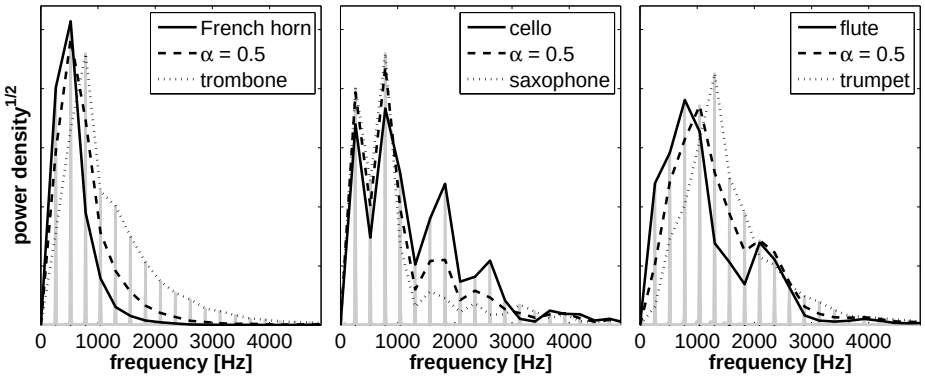


Figure 2.1: Spectral energy distribution. Mean spectra of the natural instruments and an intermediate hybrid instrument in the horn-trombone (left), cello-sax (center) and flute-trumpet (right) continua. The spectra are shown in grey, while the spectral peaks are connected by black lines indicating the instrument (see legend).

2.1 Introduction

Timbre perception is not yet sufficiently understood in a quantitative way, neither for normal-hearing nor for hearing-impaired listeners. The label *timbre* combines all auditory object attributes other than pitch, loudness, duration, spatial location and reverberation environment. Timbre is the multidimensional parameter that is used, for example, for distinguishing musical instruments or different voices. The physical properties that are connected to the determination of timbre can be divided into spectral, temporal and spectro-temporal timbre descriptors (Appendix A). A French horn, for example, sounds dull due to the high amount of low-frequency energy in the spectrum, whereas a trumpet with its high-frequency energy sounds bright when playing the same note (Figure 2.1). In a temporal dimension, hit and hammered instruments like drum and piano can be distinguished from string and wind instruments like violin and flute. While the drum shows an immediate maximal excitation followed by a decay, a violin has a smooth start and reaches maximal level after 80 ms to 2 s (Figure 2.2).

Similarity rating measurements and subsequent multidimensional scaling (MDS) can identify timbre dimensions that dominate our perception. For 30 years MDS studies have revealed various timbre dimensions of musical instrument sounds (Grey, 1977; Grey & Gordon, 1978; Krumhansl, 1989; Iverson & Krumhansl, 1993; McAdams et al., 1995; Lakatos, 2000). All studies agree in the finding that the spectral centroid and the attack dominate the timbre impression of tonal instruments, while the measures for these two dimensions remain inconclusive, and specifically,

the most appropriate way of estimating the perceptual correlates of psychophysical timbre dimensions is not clear (Appendix A). Further spectro-temporal parameters that are important for timbre perception are discussed in a nonconclusive way. Possible spectro-temporal descriptors of timbre dimensions are:

spectral

- spectral centroid (with various ways of calculation)
- spectral deviation, irregularity, spread, or slope (Krumhansl, 1989; McAdams et al., 1995)
- spectral density (Lakatos, 2000)
- pitch strength, noisiness, or harmonic proportion (McAdams et al., 1995; Lakatos, 2000)

temporal

- attack/decay time or effective duration (Krumhansl, 1989; Lakatos, 2000)
- amplitude envelope (Lakatos, 2000)

spectro-temporal

- spectral flux, overtone synchronicity, or fluctuation strength (Grey, 1977; Grey & Gordon, 1978; McAdams et al., 1995)
- attack synchronicity (Grey, 1977; Grey & Gordon, 1978)
- attack centroid (Iverson & Krumhansl, 1993)
- roughness (Terhardt, 1974; Pressnitzer & McAdams, 2000)

The exact definition of physical dimensions used within the MDS varies considerably across studies, because the underlying model assumptions are not clear.[2] In MDS different methods for calculating the timbre descriptors may distinctly change the instrument distribution in the timbre space and change the results of dominating timbre dimensions. Hence, calculation method and signal processing parameters are major factors in the search for perceptual timbre dimensions and make the timbre studies even more complex and difficult to interpret than the multidimensional timbre perception already assumed.

In an attempt to re-interpret the results with a uniform set of acoustic descriptor families, McAdams et al. (1995), McAdams & Winsberg (2000) and Levitin et al.

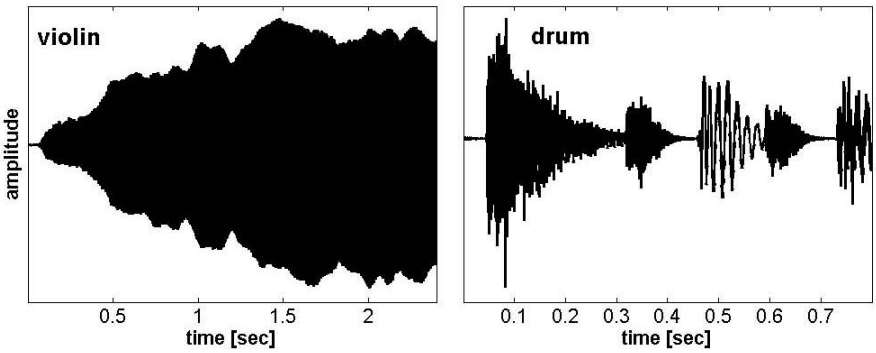


Figure 2.2: Temporal envelope of violin (left) and drum (right)

(2002) have collected old and new data and applied appropriate measures dependent on instrument (family) and subject classes. However, the number of possible timbre dimensions is large and the dimensions vary with instruments and subjects, which makes conclusions difficult. Every new natural instrument or new recording of an instrument used as a stimulus in measurements may add another dimension. Therefore, a new method is presented in the present study. This method produces new hybrids of timbres that were already used in timbre perception experiments and, hence, adds new timbres and timbre distances without adding new dimensions to the timbre space. By linear interpolation of spectral parameters, sounds of musical instruments were cross-faded ("morphed") along spectro-temporal dimensions, whereby stimulus continua between natural instruments were generated. As a pilot experiment using these morphed sounds, the present study shows a pair wise comparison with an 8-step scale. In order to verify coarse differences in timbre perception between hearing-impaired and normal-hearing listeners, subjects with and without hearing loss were requested to judge the similarity of the morphed sounds. In comparison with the results from MDS studies from the literature, a mapping between common psychophysical timbre dimensions and the model parameter α is provided. Thus, applied within MDS studies, the presented method may supplement earlier methods to verify a uniform set of timbre descriptors for musical instruments.

2.2 Stimulus preparation

Three pairs of musical instruments were chosen in a way such that each pair was very dissimilar in one timbre-dominating dimension of Grey's (1977) MDS space and similar in the other dimensions:

a) trombone and French horn, which show different spectral centroids ("brightness")

b) saxophone and cello, which, according to Grey (1977), differ mainly in spectral flux ("brightness fluctuation" or sometimes "roughness")

c) flute and trumpet, which differ mainly in the attack segment ("smooth or noisy attack" vs. "percussive attack")

First, acoustic recordings (Fritts, 2002) of these instruments pitched at C4 ($f_0 \approx 262\ Hz$) were synthesized using the DAFX toolbox (Amatriain et al., 2002). The synthetic signals were then equalized in pitch and level, and faded out with linear flanks of 150 ms to produce signals of 1.8 s length.

By linear interpolation of spectral parameters, sounds were then pair-wise cross-faded ("morphed"), whereby three stimulus continua, one between trombone and French horn (horn-trombone continuum), another between cello and saxophone (cello-sax continuum), and the third between flute and trumpet (flute-trumpet continuum) were generated. The morphing used an overlap-add analysis-synthesis algorithm based on a sinusoidal plus residual model (Amatriain et al., 2002) and interpolated frequency, amplitude and phase of the sinusoidal part (i.e. the harmonic sound partials) as well as the amplitudes of the residuum (i.e. remaining noise portion of the sound). A more detailed description of the morphing method can be found in Chapter 3. Note that for the present study the attack portion was not removed as in Chapter 3.

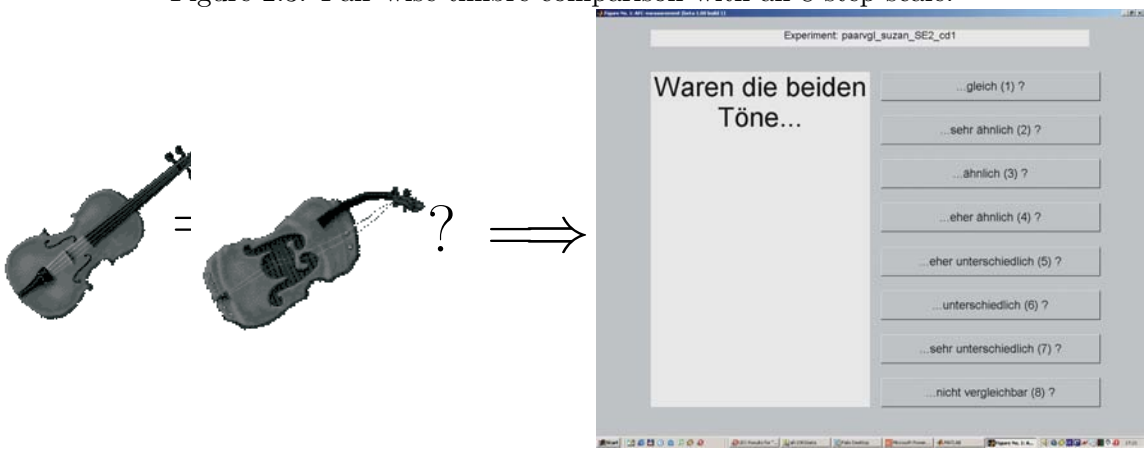
In this way, three instrument continua were generated and used in the psychoacoustic measurements described below. In the following, the labels "instrument continuum" and "stimulus continuum" are used synonymously. The morphed stimuli were named by their morphing-parameter α, which corresponds to the ratio of one of the original instruments to the original sounds. Hence, α ranges between 0 (corresponding to the sound of the original French horn, cello or flute) and 1 (trombone, saxophone or trumpet), where a spacing of 0.1 was used.

2.3 Experiments

2.3.1 Experimental setup

The sounds were presented diotically through ear phones (Sennheiser HD580) in a soundproof booth. The length of the signals was 1.8 s, separated by a silent interval of 0.5 s. All signals were digitally generated on a PC prior to the measurements, output via a digital I/O-card (RME Digi96 PAD) and optically passed to a 24 bit DA-converter (RME ADI-8 PRO). The presentation level was calibrated to and

Figure 2.3: Pair wise timbre comparison with an 8-step scale.



played at 65 dB SPL for the normal-hearing subjects. For the hearing-impaired subjects, the sound level was amplified linearly and broad-band to 68-95 dB SPL until the sound was perceived with a "comfortable and intermediate" loudness. The stimuli had thus a level of 68 dB SPL for subject iUL, 95 dB for subject iGM, and 80 dB for the remaining hearing-impaired subjects.

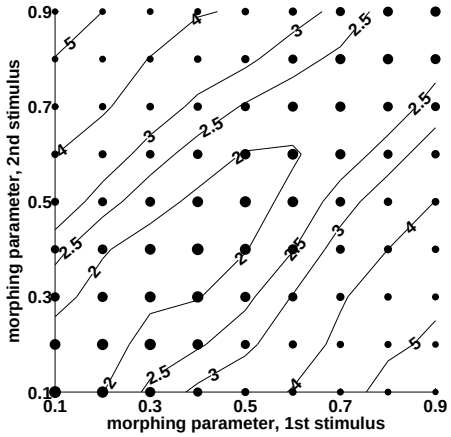
In a pair wise comparison with an 8-step scale, two signals of the same instrument continuum were presented in each trial. The subjects' task was to rate the similarity and to indicate whether the sounds were "equal" (1), "very similar" (2), "similar" (3), "rather similar" (4), "rather different" (5), "different" (6), "very different" (7), or "not comparable" (8) (Figure 2.3). No feedback was given.

All 11 stimuli per instrument continuum were compared with each other, resulting in 121 stimulus pairs (trials) in each continuum. The trials of the three instrument continua were presented interleaved in a random order. After 60 practice trials of randomly selected sound pairs, each subject had to rate all 363 different stimulus pairs.

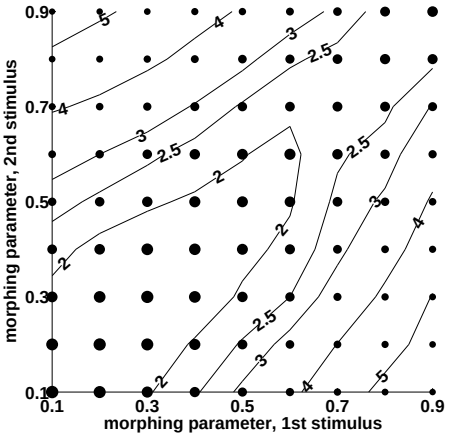
2.3.2 Subjects

7 normal-hearing subjects aged between 21 and 45 years and 6 hearing-impaired subjects aged between 33 and 62 years took part in the experiments and were paid for participation.

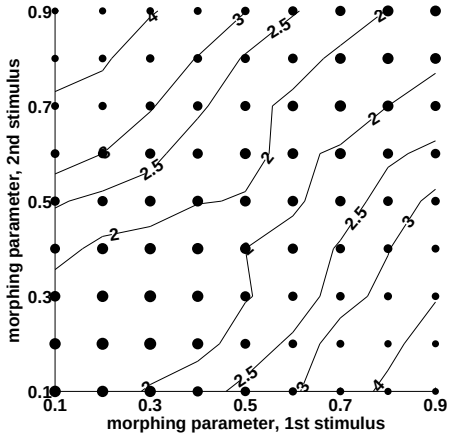
The subjects were interviewed for their musical background. 1 hearing-impaired subject (iUL) was a professional music and instrument teacher. 5 of the normal-hearing subjects (nJF, nNG, nMN, nKP, nKS) and 3 of the hearing-impaired subjects (iGH, iFL, iEW) were amateur musicians, had had more than 4 years of regular experience in learning and practising an instrument or singing, and were still ac-



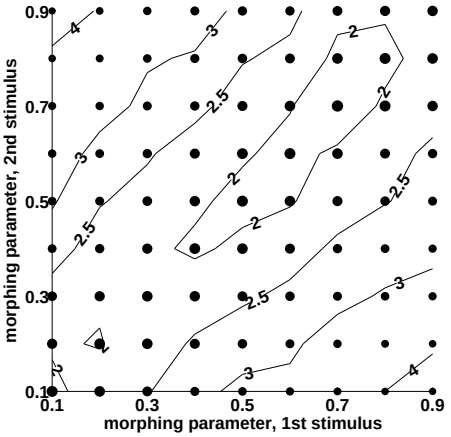
(a) NH horn-trombone continuum



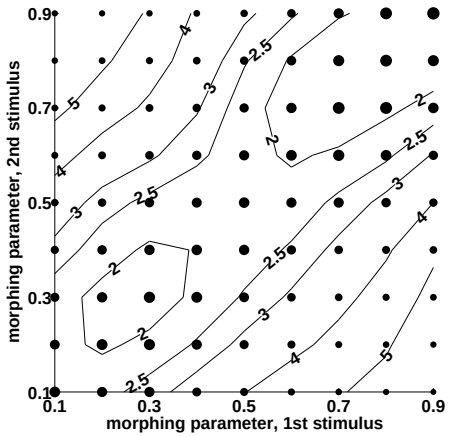
(d) HI horn-trombone continuum



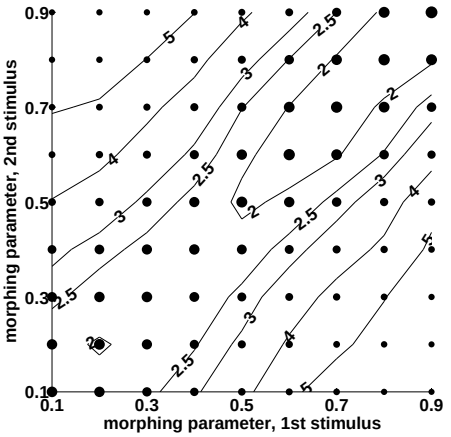
(b) NH cello-sax continuum



(e) HI cello-sax continuum



(c) NH flute-trumpet continuum



(f) HI flute-trumpet continuum

Figure 2.4: Similarity ratings of (a-c) 7 normal-hearing and (d-f) 6 hearing-impaired subjects in the three instrument continua. Axes indicate morphing-parameter α of presented stimuli in the respective continuum. The dot size represents the amount of similarity between stimuli. For a clearer view, the similarity ratings were smoothed by a running mean of 3 stimuli with adjacent morphing-parameter α. Additional contour lines show equi-rating levels.

tively practising music at the time of the experiment. 2 normal-hearing (nRM, nRW) and 2 hearing-impaired subjects (iDL, iGM) reported having no experience playing musical instruments or only little musical practice in the past.

2.3.3 Results of normal-hearing subjects

Figures 2.4(a), (b) and (c) show the mean "amount of similarity" given by the 7 normal-hearing subjects in the three instrument continua. The larger the symbol size in Figure 2.4, the "more similar" the stimulus pair was rated. The symbol sizes and isolines in Figures 2.4(a)-(c) look nearly symmetric around the axis diagonal, which indicates that the presentation order of the stimuli was of low importance. The isolines in Figures 2.4(a)-(c) are nearly parallel to the axis diagonal, which indicates that similarity ratings depended mainly on the morphing-parameter *difference* between the two rated sounds. However, the isolines differ slightly from diagonal parallels. This indicates that the similarity ratings were slightly dependent on absolute morphing-parameters α of the stimuli.

Statistical tests with results

The minor dependence of the ratings on stimulus order was also confirmed by the Wilcoxon rank sum test, which did not show any significant difference ($p>0.05$) between ratings for positive and ratings for negative morphing-parameter difference. Neglecting the order, a variance analysis ANOVA was conducted for the two factors "absolute morphing-parameter difference" $\Delta\alpha$ (>0) and "morphing-parameter α of the first stimulus". The first factor was highly significant ($p<0.001$) and the second factor was only significant in the flute-trumpet continuum ($p=0.33$, 0.07 and 0.04 for horn-trombone, cello-sax and flute-trumpet respectively). Hence, while in the flute-trumpet continuum ratings were slightly dependent on absolute morphing-parameters α, the main determining parameter for the similarity rating was the absolute value of the morphing-parameter distance between the two pairwise presented stimuli in all continua. This parameter difference will therefore be used as the comparison parameter in the following sections.

Similarity rating vs. morphing-parameter difference $\Delta\alpha$

The mean ratings of all normal-hearing subjects against morphing-parameter distance are shown as circles in Figure 2.5 for the three instrument continua. All rating curves increase monotonically, which indicates that the applied morphing algorithm also monotonically interpolates in a perceptive space.

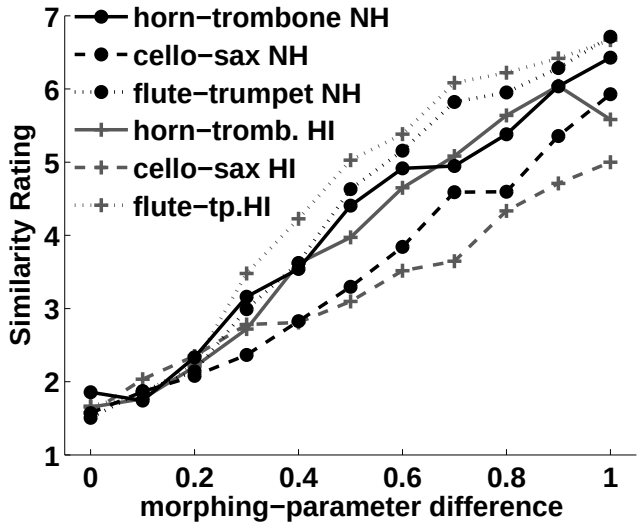
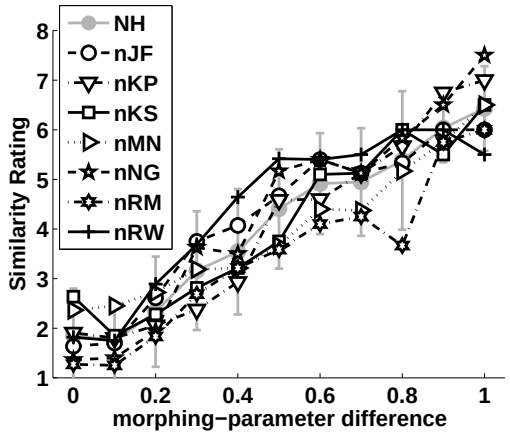


Figure 2.5: Comparing similarity ratings across listener groups and instrument continua. Abscissa indicates morphing-parameter difference $\Delta\alpha$ of the presented stimulus pair. Ordinate indicates rating within respective subject group and instrument continuum.

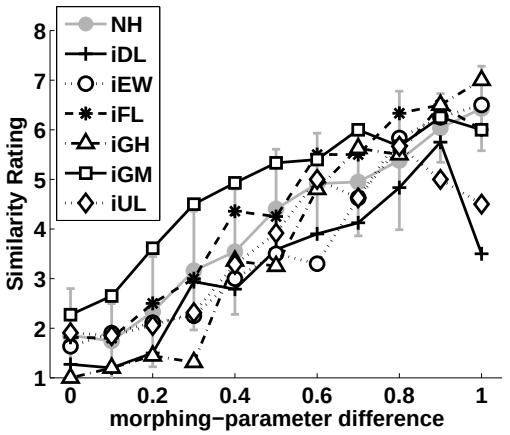
During the measurements, the instrument pairs of all three continua were presented and rated in an interleaved way without informing the subjects about different instrument continua. As a sole constraint on the rating scale, subjects were asked to use the rating 7 (very different) or 8 (not comparable) at least once in the measurement. Before the evaluated measurement, subjects heard and rated 70 training pairs demonstrating the variance of the stimuli used. Hence, comparing (maximal) rating results across continua may indicate different perceptual weighting of the timbre dimensions represented by the continua. Slight differences can be observed between the curves of the different instrument continua (Figure 2.5). The ratings of the maximal $\Delta\alpha$ in the three instrument continua are 6.4, 5.9 and 6.7, respectively. Hence, flute and trumpet (3rd continuum) seem to be perceived as more different than trombone and French horn, which are again perceived as more different than saxophone and cello.[3]

2.3.4 Results of hearing-impaired subjects

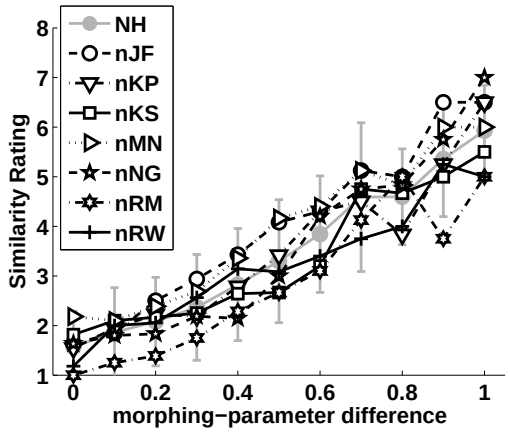
The rating responses of the 6 hearing-impaired subjects are shown in Figures 2.4(d)-(f) in a similar way as for the 7 normal-hearing subjects (Figures 2.4(a)-(c)). Symbol sizes and isolines in the horn-trombone continuum seem to be similar in both listener groups. Small differences can be seen in the flute-trumpet continuum, whereas symbol sizes and isolines differ visibly in the cello-sax continuum.



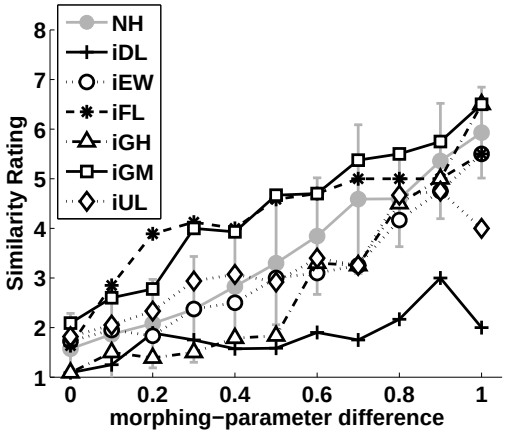
(a) NH horn-trombone continuum



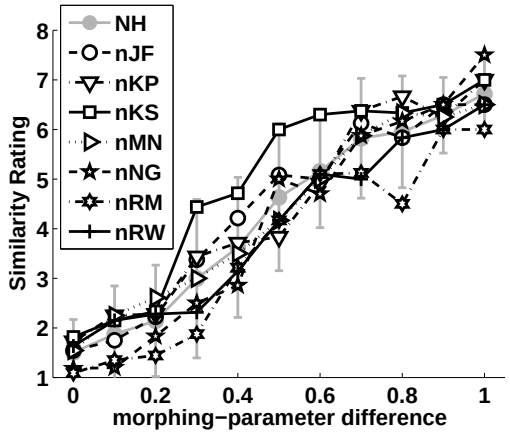
(d) HI horn-trombone continuum



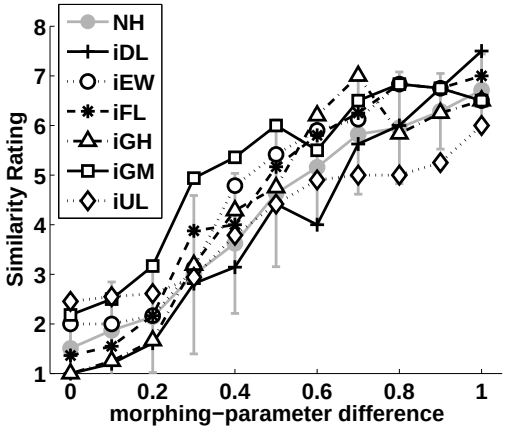
(b) NH cello-sax continuum



(e) HI cello-sax continuum

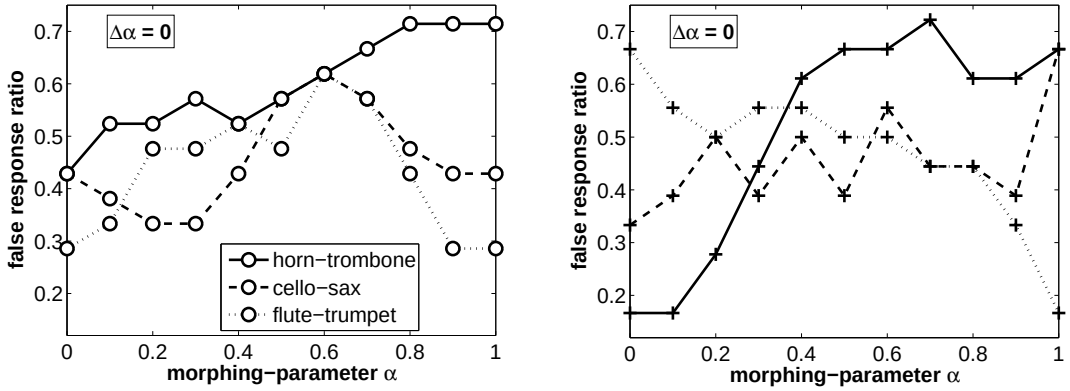


(c) NH flute-trumpet continuum



(f) HI flute-trumpet continuum

Figure 2.6: Individual similarity ratings of the 7 normal-hearing (a-c) and 6 hearing-impaired (d-f) subjects. Abscissa indicates morphing-parameter distance $\Delta\alpha$ of the presented stimulus pairs and ordinate indicates the rating. For comparison the mean ratings of the normal-hearing subjects are plotted in grey with inter- and intra-individual standard deviations.



(a) normal-hearing (b) hearing-impaired

Figure 2.7: Rating dependency on morphing-parameter α. False responses (i.e., summed ratings of "very similar" (2) through "not comparable" (8)) as a proportion of all responses when identical stimuli with morphing-parameter distance $\Delta\alpha = 0$ were presented. Mean false response ratio of (a) 7 normal-hearing and (b) 6 hearing-impaired subjects in the horn-trombone (continuous line), cello-sax (dashed line) and flute-trumpet (dotted line) continua. For clarity, the ratio was smoothed by a running mean of 3 adjacent αs.

Similarity rating vs. morphing-parameter difference $\Delta\alpha$

Figure 2.5 shows the mean ratings of all hearing-impaired subjects against absolute morphing-parameter distance $\Delta\alpha$ in the three instrument continua. The difference of the averaged ratings between hearing-impaired and normal-hearing subjects is neither high nor significant ($p>0.05$ in ANOVA). The difference between instrument continua is higher than the difference between listener groups. Due to the variety of hearing loss types across subjects, the individual abilities can vary distinctly. Therefore, the individual data is shown in Figures 2.6(d)-(f). For better comparison, the mean data of the normal-hearing subjects is added in grey in the background of Figures 2.6(d)-(f), and the individual results of the normal-hearing listeners is shown in Figures 2.6(a)-(c).[45]

2.3.5 Rating dependency on morphing-parameter

Since Chapter 3 showed a dependency of JND results on absolute morphing-parameter, the effect of the stimuli's morphing-parameter on distinguishability are analyzed in Figure 2.7 as a false response ratio. Figure 2.7 shows the incorrect responses given to identical stimuli, that is, the number of ratings of "very similar" (2) through "not comparable" (8) as proportion of all responses. It represents the "uncertainty" of the ratings as function of α. For both, normal-hearing and

hearing-impaired subjects, false response ratio is not constant along all three continua indicating a certain rating dependency on morphing-parameter α in all continua. For both subject groups, the horn-trombone continuum (continuous lines) shows an increasing false response ratio (i.e. increasing uncertainty) with increasing α. In the horn-trombone (continuous lines) and flute-trumpet (dotted lines) continuum, the ratio varies for the hearing-impaired subjects stronger along the respective continuum than for the normal-hearing subjects.

2.4 Discussion

The main findings of the measurements can be summarized as follows:

- The similarity ratings mainly depended on morphing-parameter distance $\Delta\alpha$, while the order of pairwise-presented stimuli was of low importance for the ratings.
- Ratings were slightly dependent on absolute morphing-parameter α, in particular in the flute-trumpet continuum.
- Most hearing-impaired subjects, who were provided with adequate broad-band amplification of the stimuli, showed in all instrument continua similar ratings to normal-hearing subjects.
- Two hearing-impaired subjects showed distinctly lower rating values than normal-hearing subjects in the cello-sax continuum.
- Along the horn-trombone and flute-trumpet continuum the mean rating of the hearing-impaired subjects varied more strongly with absolute morphing-parameter α than that of normal-hearing subjects.

In order to understand the experimental findings, the correlation of spectro-temporal timbre descriptors with the results will be discussed. It is assumed that the rating results are correlated with objective timbre descriptors found in previous studies on musical timbre rating. Hence, a correlation between similarity rating results and distances between appropriate timbre descriptors that is consistent with the MDS relation between ratings and descriptors from the literature will demonstrate the usability of the morphing method employed here to explore and sample the complex timbre space. It will also help to allocate any differences between normal-hearing and hearing-impaired subjects in the MDS space known from literature.

Table 2.1: Correlation of timbre descriptor differences with rating results in the three continua. Bold numbers indicate descriptors that were found in Appendix A to vary distinctly along the continuum. Symbols ($^{\natural}$,$^{\flat}$,$^{\sharp}$) indicate strong cross-correlation (p>0.95) between "dominating" descriptors, which may be dependent (see also Appendix A).

	horn-trombone	cello-sax	flute-trumpet	
α	0.78	0.72	0.79	morphing-parameter
Fc	**0.76**$^{\sharp}$	0.58	0.79	spectral centroid
spIrr	0.75$^{\sharp}$	0.31	0.68	spectral irregularity
OS_{stat}	0.36	0.51	**0.79**$^{\flat}$	overtone synchronicity >840 Hz
Fc_{atk}	0.77$^{\sharp}$	**0.72**	0.45	attack: centroid
OS_{atk}	0.63	**0.52**	0.76$^{\flat}$	attack: overtone synchronicity
rise-time	0.38	**0.51**	**0.44**	attack: log-rise-time
E-high	0.58	**0.64**	0.68$^{\flat}$	attack: high-frequency energy
E-noise	0.62	0.70$^{\natural}$	**0.39**	inharmonic energy
$E\text{-noise}_{atk}$	0.53	0.70$^{\natural}$	**0.63**	inharmonic energy during attack
$E\text{-noise}_{stat}$	0.61	0.69$^{\natural}$	**0.78**	inharmonic energy after attack

2.4.1 Spectro-temporal timbre descriptors

While the variable α used in the present study makes up a physical correlate of timbre space independent from perception, previous studies found psychophysical timbre dimensions to coincide with "spectro-temporal timbre descriptors", that is to say parameters extracted from the sound signal that represent the perceptual timbre dimensions in a physical space (Section 2.1). Appendix A provides the mapping between α and common timbre descriptors from the literature and shows which descriptors vary across the stimuli of the present study. High variation of a descriptor along a continuum is an indication of parameters that may be used for similarity ratings. The main findings of Appendix A can be summarized as follows:

① Variation in most timbre dimensions were observed along all instrument continua, but only up to three independent descriptors vary along each continuum to a large extent.

② Spectral centroid varied most in the horn-trombone continuum.

③ The attack (i.e. attack's centroid, overtone synchronicity, high-frequency energy and log-rise time) varied distinctly in the cello-sax continuum.

④ Variation of spectral flux, log-rise time and inharmonic content were high in the flute-trumpet continuum.

⑤ In the trumpet, the inharmonic energy was tonal and present only during the

attack, while in the flute it was noise-like throughout the entire stimulus.

⑥ Some descriptors did not seem to be independent from others. E.g, spectral irregularity seemed to depend on spectral centroid in the horn-trombone continuum, and the attack's overtone synchronicity as well as the high-frequency energy during attack seemed to dependent on spectral flux and spectral centroid in the flute-trumpet continuum.

In order to verify the spectro-temporal descriptors that may have actually been *used* by subjects to distinguish and rate the stimuli, the rating results of the normal-hearing subjects were correlated (using the Pearson product, Equation 5.3, p.68) with the timbre descriptor differences of the corresponding stimulus pairs. The correlation coefficients for the three instrument continua are shown in Table 2.1. In confirmation of the high variation of descriptor value along the continua, spectral centroid seems to be a main factor in the horn-trombone continuum (p=0.76 in agreement with ②), attack's centroid dominates the cello-sax continuum (p=0.72, ③), and overtone synchronicity (or spectral flux) is a dominating factor in the flute-trumpet continuum (p=0.79, ④). The attack's overtone synchronicity in the cello-sax continuum and the log-rise-time in the flute-trumpet and cello-sax continua do not seem to correlate significantly with the results, although these timbre descriptors varied distinctly in these continua, respectively (③ and ④). On the other hand, the spectral centroid may be used as a cue in the flute-trumpet continuum and the inharmonic content may be a cue in the cello-sax continuum. However, since the variance of these descriptors is not as distinct as in other continua (④ compared to ②, and ③ compared to ④), the spectral centroid and the noise content may only be minor cues in the flute-trumpet and cello-sax continua, respectively. In the flute-trumpet continuum, the correlation of the results with inharmonic energy is low, although this timbre descriptor varied distinctly in this continuum (④). This may be due to the different kinds of inharmonic content in the flute and trumpet sounds (⑤). Correlation of the results with inharmonic energy for the attack and stationary segment separately show high coefficients, which indicates that the noise content after the attack (p=0.78) was used particularly as a distinction cue in the flute-trumpet continuum. The other descriptors with which the results were highly correlated (see Table 2.1) may be dependent on other dominating factors (⑥).

Hence, this section indicates perceptual relevance of the objective timbre descriptors in the rating experiments of the present study. The variance of the spectro-temporal descriptors along the continua (①-⑥) and correlation of the rating results with descriptor differences (Table 2.1) suggest that the following dimensions are used as major (underlined) and minor discrimination cues by subjects:

- horn-trombone continuum: spectral centroid
- cello-sax continuum: attack's centroid , noise content
- flute-trumpet continuum: spectral flux , inharmonic content , spectral centroid

Hence, the instrument continua represent different timbre dimensions, and rating differences between instrument continua may result from different perception or discrimination abilities of the corresponding timbre dimension.

2.4.2 Rating dependency on morphing-parameter

ANOVA and Wilcoxon tests showed that the order of pairwise-presented stimuli is of low importance for similarity ratings, but that the ratings are biased by absolute morphing-parameter α. The dependency of ratings on α may be due to different reasons in the different continua.

In the horn-trombone continuum, the false response ratio increases with α (Figure 2.7); that is to say, the uncertainty in detecting identical stimuli increases with α in this continuum. Spectral centroid, which is the dominant distinction cue in this continuum, also increases distinctly with α (Appendix A). JND of spectral centroid commonly increases with increasing frequency according to Weber's law (Chapter 3), which suggests that the increasing uncertainty and the rating dependency on α may be due to increasing centroid in this continuum.

In the flute-trumpet continuum, different factors seem to dominate the ratings, and various cues that can be used to distinguish the stimuli are clearly audible when listening to the stimuli. Noise content decreases from flute to horn; in other words, sound becomes less soft and more tonal. Simultaneously, the percussive/inharmonic attack of the trumpet becomes more and more audible/distinct along the continuum. And while the brightness on the flute end increases during the duration of the sound (high spectral flux), on the trumpet end, brightness is rather constant. The salience of these cues varies along the continuum, and the dominating timbre dimension which leads to the ratings changes along the continuum. This may lead to a dependency in ratings on absolute morphing-parameter.

In the cello-sax continuum, the shape of the false response ratio (Figure 2.7, p.17) is distinctly different than the u-shaped JND-to-α_{ref} curve in Chapter 3 (Figure 3.1, p.31). Since the false response ratio indicates the uncertainty in detecting identical stimuli, it is assumed to be correlated with JND. In Chapter 3 the same stimuli as in the present study were used, but the attack of the stimuli was removed, which may have changed the dominating timbre cues used to distinguish the stimuli in the

cello-sax continuum. This confirms that *attack* descriptors dominate the ratings in the cello-sax continuum in the present study, but that other timbre dimensions like noise content or flux may be used as additional distinction cues.

2.4.3 Hearing-impaired subjects

Comparison between hearing-impaired and normal-hearing subjects shows that most hearing-impaired subjects gave similar ratings to normal-hearing listeners. Of 6 hearing-impaired subjects, only subjects iDL and iGH showed distinct deviation from normal-hearing subjects and higher discrimination thresholds in all continua. In the cello-sax continuum, subject iDL's responses did not exceed rating 3 ("similar") for any rated pair, whereas in the other continua he used the rating range up to 7 ("very different"). iDL's and iGH's frequency-dependent threshold configurations with distinctly higher loss above 2 kHz and around 1 kHz, respectively[6], seemed to obscure certain timbre variations, in particular in the cello-sax continuum. Also for other hearing-impaired subjects, the highest deviations in the results were observed in the cello-sax continuum. Normal-hearing listeners did not show higher variance of rating in this continuum. Hence, the reason for the higher variance in hearing-impaired results may lie in a hearing loss sensitive timbre dimension, that is to say, timbre differences that are sensitive to elevated threshold, compression loss and/or distortion. The main timbre variation in this continuum lies in the attack segment, that is overtone synchronicity and high-frequency energy during the first 200-500 ms. This leads to the assumption that the hearing-impaired subjects had problems in perceiving and distinguishing the high-frequency energy during the attack present in the saxophone hybrids. The cello-sax continuum is also the only instrument continuum with jagged harmonics amplitudes leading to high (harmonic-collective) amplitude modulations (Appendix A). A high amplitude fluctuation can distract from discrimination tasks, if it cannot be used as a discrimination cue (spectral irregularity in the cello-sax continuum is high, but irregularity differences are low, Figure A.1(b), p.90). Since hearing-impaired listeners may perceive an enhanced internal variance of inherent amplitude fluctuation (Oxenham & Bacon, 2003, and Appendix B), a "masking" fluctuation or irregularity may be more of a disadvantage for hearing-impaired than for normal-hearing subjects.

The false response ratio for the hearing-impaired subjects in the horn-trombone continuum increases with increasing α as for normal-hearing subjects (Figure 2.7(b)). At the horn side, hearing-impaired subjects show distinctly fewer wrong responses[7], the ratio increases with α faster than for normal-hearing subjects, and maximal uncertainty is reached earlier (for $\alpha \geq 0.5$). Hence, the uncertainty func-

tion shows a "recruitment phenomenon" from dull horn to bright trombone; the internal brightness variance may be enlarged by compression loss. Better timbre discrimination skills in hearing-impaired subjects were also observed in Chapter 4.

On the other hand, in the flute-trumpet continuum (dotted line) hearing-impaired subjects show higher uncertainty at the flute end than at the trumpet end; at the flute end they show a higher uncertainty than normal-hearing subjects. The high noise content in the flute sound may explain this finding.[8]

2.5 Conclusion

The present study measured similarity ratings of musical instrument sounds along three timbre continua. Timbre is a multidimensional psychoacoustical attribute. While previous studies found spectro-temporal parameters that physically describe the perceptual timbre dimensions, the variable α used in the present study allows a continuous path within the timbre space. Appendix A provided the mapping between α and common spectro-temporal timbre descriptors from the literature and showed which descriptors vary across the stimuli of the present study. The present study produced evidence for the perceptual relevance of the objective timbre descriptors for the results of the rating experiments conducted here. The main findings of this study can be summarized as follows:

- The order of pairwise-presented stimuli was of low importance for the ratings. Ratings were slightly dependent on absolute morphing-parameter α, if spectro-temporal timbre descriptors vary distinctly along the continuum and if the crucial timbre dimension changes along the continuum. However, the similarity ratings depended mainly on morphing-parameter distance $\Delta\alpha$.

- In each instrument continuum, similarity ratings were based on other dominating timbre descriptors. Rating in the horn-trombone continuum seemed to be mainly based on spectral centroid (i.e. a brightness percept). In the cello-sax continuum, the attack seemed to be the dominating rating cue, in particular the high-frequency energy and noise present during the attack. In the flute-trumpet continuum, the inharmonic content and spectral flux seemed to concur for the rating; these are perceived as noise content over the duration of the sound in the flute hybrids, inharmonic percussive attack in the trumpet hybrids, and temporally varying brightness in the flute hybrids.

- Timbre ratings of musical instrument sounds are not necessarily affected by a hearing loss if an adequate linear amplification is used. Some hearing-

impaired subjects with moderate hearing loss gave similarity ratings similar to normal-hearing listeners when stimuli were (broad-band) amplified to provide approximately the same loudness impression (i.e. MCL or intermediate) as for normal-hearing listeners. However, certain hearing losses may obscure or distort timbre variation along certain timbre dimensions by elevated hearing threshold and compression loss. In particular, high-frequency energy during the attack may be dismissed. On the other hand, internal brightness variation (i.e. brightness differences and brightness uncertainty) may be enlarged by hearing loss, which may compensate for some of the deficits listed above.

- The correlation between the rating results and the appropriate timbre descriptors demonstrated the usability of the morphing method employed here to explore the complex timbre space. By adding new timbres and timbre distances without adding new dimensions to the timbre space, the presented method, for example in combination with MDS, may supplement earlier methods to verify a uniform set of timbre descriptors for musical instruments.

Chapter 3

Timbre discrimination of morphed sounds

Abstract

*In the present study morphing (that is, the continuous sound transformation of one instrument into another) is introduced as a new method for timbre perception studies. This technique interpolates the timbre between natural instruments and thus makes it possible to analyze the discrimination of similar, quasi-natural timbres and small perception differences. Combined with just noticeable difference (JND) measurements, morphing allows for the objective determination of a value that provides comparison between different subject groups and timbres. The present study shows that the measurement method, which is independent of the subjects' previous knowledge (e.g., knowledge of instrument names) and the instruction about the sound to be expected, can reveal differences in timbre perception between subjects with different levels of musical experience. The exemplary JND measurements reveal a systematic change in timbre JND with reference stimulus, which is correlated with the stimuli's spectral centroid (Fc) and, in one continuum, additionally with a spectro-*temporal *dimension. Fc difference at threshold increases with Fc, in conformance with Weber's law. The results indicate that Fc is a dominant distinction cue for both of the instrument continua used in the study, while an additional cue, such as spectral flux, may dominate the perceptual differences in one of the continua.*

3.1 Introduction

People with sensorineural hearing loss (including hearing aid users) often have problems with timbre distortion and, as a consequence, with music perception, which is not yet well understood and can not be compensated with hearing aids. Therefore a better understanding of timbre perception can help to improve auditory models and hearing aids. For more than 30 years, studies on musical timbre were done on normal-hearing and hearing-impaired listeners, mostly as similarity ratings of different musical instruments and multidimensional scaling (MDS) of the judgements (e.g. Plomp, 1970, 1975; Grey, 1977; Wessel, 1979; Krumhansl, 1989; McAdams & Cunibile, 1992; Iverson & Krumhansl, 1993; McAdams et al., 1995) or as recognition tasks (e.g. Gfeller et al., 2002b,a). Both methods are important for identifying perceived timbre dimensions and indicating differences in timbre perception between normal-hearing and hearing-impaired listeners. Using natural musical instruments, these methods are quite coarse, so that after sufficient training even cochlea-implant recipients can achieve recognition scores equivalent to those of normal-hearing people (Gfeller et al., 2002a). The perceptual differences in timbre perception between acoustical hearing-impaired and normal-hearing people are even smaller (Chapters 2 and 4) and may not therefore be characterized with these methods. In addition, similarity scaling is a subjective method, and the number of dimensions found with MDS is limited by the number of instruments used. Recognition tasks depend on the subjects' earlier knowledge, and hence is rather a measure for the ability of hearing-impaired people to transfer the knowledge from when they still had normal hearing to the cues of the impaired hearing.

Other studies (Grey, 1978; McAdams et al., 1999) used resynthesized sounds with simplified spectro-temporal parameters to measure timbre discriminability near subjects' perception thresholds. This method is well suitable to studying small discrimination differences of different subject groups along certain physically determined parameters. However, the method only measures timbre perception differences along dimensions that are determined by the signal-processing simplification applied. Timbre dimensions are not yet sufficiently understood, and certain minor dimensions might be overlooked by the method. Since the method only removes, smoothes or reduces certain parameters, a combination of large and small timbre differences, as would be used, for example, for measuring timbre perception limits and their variation between natural musical instrument groups, is not possible.

Therefore a new psychophysical paradigm for assessing timbre perception is needed that should build on an interpolation between natural instruments in order to increase the number of stimuli between two natural instruments and to pro-

duce new stimuli with small timbre distances. Such a method should be capable of demonstrating any differences in timbre perception between normal-hearing and hearing-impaired listeners. Ideally, this method should be independent of earlier knowledge, analyzing primarily the subject's present perception and the bottom-up processes involved. A common psychophysical measure is the just noticeable difference (JND). JNDs have been determined for many acoustic parameters, such as intensity level, sound location, and frequency, in people with and without hearing impairment. JND measurements have also been applied for certain timbre dimensions using artificial complex tones, for example, studying the perception of spectral energy distribution in profile analysis (Green, 1988b,a). If used in trained subjects, the JND is usually rather independent from higher cortical processes, and well comparable between different listener groups and auditory models. So a JND of timbre could be used to quantify differences with respect to timbre dimensions between people with hearing loss and people with normal hearing.

Timbre is a combination of all acoustical attributes that are not exclusively assigned to the perception of pitch, loudness or length (American Standard Association, 1960; Plomp, 1970); that is to say, timbre is a multidimensional perception measure. Dominating timbre dimensions of musical instruments are the spectral energy distribution (spectral centroid, spectral irregularity), the amount of spectral flux within the tone over time (or presence of synchronicity in the upper harmonics), and the initial attack segment (presence of low-amplitude, high-frequency energy in the attack segment and logarithm of the rise time[9])(Grey, 1977; Krumhansl, 1989; Krimphoff et al., 1994). In general, the dimensions can be divided into spectral and spectro-temporal timbre descriptors.

By linear interpolation of spectral parameters within an overlap-add analysis and synthesis, sounds of musical instruments can be cross-faded ("morphed") along these dimensions, whereby stimulus continua between natural instruments are generated. This can be used to produce sounds for similarity judgments (Chapter 2) and multidimensional scaling, filling the gaps between natural instruments and so avoiding clustering of the stimuli in the MDS space. This morphing as a method to determine timbre JND will be described below. However, it is unclear how the physically defined morphing-parameter relates to perceptual differences and whether this method depends on different stimulus types or individual subjects. Therefore, a series of measurements will be presented that evaluate the method and show that these JNDs can characterize differences in perception between different subject groups in terms of different stimuli. Finally it is shown how the morphing-parameter relates to spectro-temporal timbre descriptors and which of the common dominating timbre dimensions have effects onto the JND results.

3.2 Morphing method

Morphing is a transformation that generates new elements with hybrid properties from two or more elements (Amatriain et al., 2002). The morphing algorithm by Amatriain et al. (2002) is imbedded in the *Digital Audio Effects* (DAFX) framework by Zölzer and colleagues (Zölzer, 2002) and is the source of the algorithm described below.

An analysis-synthesis algorithm based on a sinusoidal plus residual model is used, in which the input sound s(t) is modelled by

$$s(t) = \sum_{r=1}^{ref} A_{ref}(t) cos[\vartheta_{ref}(t)] + e(t) \tag{3.1}$$

where $A_{ref}(t)$ and ϑ(t) are the instantaneous amplitude and phase of the r^{th} sinusoid, respectively, and $e(t)$ is the noise component at time t (in seconds).

3.2.1 Analysis

The signal is analyzed with a short-time Fast Fourier Transform in overlapping windows. For the present study, the window length (w) was set to 1024 samples at a sampling frequency (Fs) of 44100 Hz. In each window the spectral peaks, namely up to 50 amplitude maxima of the spectrum, are detected, the pitch is calculated, and each peak is joined to a windows-continuing track (representing approximately the partial number). The frequency, phase and amplitude of each spectral peak is saved, as well as the residual, which is the difference between the signal's spectrum and the spectrum of the detected sinusoids in the representation of Equation 3.1.

3.2.2 Morphing

After two sounds are analyzed, the frequency and amplitude of each sinusoidal component of both sounds are interpolated linearly in each track:

$$X_{new} = \alpha \cdot X_{old1} + (1 - \alpha) \cdot X_{old2} \tag{3.2}$$

where X is the frequency or amplitude of the spectral maxima. In the first time window of a new track, the instantaneous phase is also interpolated according to Equation 3.2. In any succeeding window at time t, that is to say, if a track for the analyzed partial already exists, the new instantaneous phase is taken to be the integral of the instantaneous frequency and calculated by

$$\vartheta_{new}(t) = \vartheta_{new}(t - \Delta t) + \Delta t \cdot f_{new}(t - \Delta t) \tag{3.3}$$

with the previous window's interpolated frequency f_{new} and the window hop size (or frame sampling period) Δt. In this way, the sinusoid continues in each track consistently from the previous window without phase jump. Since all natural timbre dimensions, including the noise part of a sound, are subject to this morphing scheme, the residuum is also interpolated. The morphing-parameter (interpolation factor) α controls the amount of the first sound in the resulting morph. Subsequently, the morphed sound is synthesized with the new frequencies, phases and amplitudes with the analysis algorithm in reversed order.

For the purpose of the psychoacoustic experiments described below, Amatriain's morphing algorithm (Amatriain et al., 2002) was modified as follows. An interpolation of the residual component of the sound was added:

$$e_{new} = \alpha \cdot e_{old1} + (1 - \alpha) \cdot e_{old2} \tag{3.4}$$

where α is the morphing-parameter and e the residual FFT vector. Furthermore, at the beginning of each partial track, the phase was interpolated and used for the synthesis of the morphed sound:

$$\vartheta_{new} = (\alpha \cdot \vartheta_{old1} + (1 - \alpha) \cdot \vartheta_{old2}) mod(2\pi) \tag{3.5}$$

where ϑ is the instantaneous phase.

3.2.3 Stimuli preparation

Two pairs of musical instruments were chosen in a way that one pair (trombone and French horn) differed greatly in their spectral centroid and the other (saxophone and cello) in their spectral flux; all other physical parameters were similar within each pair (Grey, 1977). First, acoustic recordings (Fritts, 2002) of these instruments pitched at C4 ($f0 \approx 262\ Hz$) were low-pass filtered at 10 kHz using an 800th order linear-phase FIR filter; this was done for better audiological comparison of hearing-impaired with normal-hearing listeners. The attack sequence was cut off, because the perceived length of the sound depends on the attack length (McAdams et al., 1995). An approximately stationary section of 0.9 s of the remaining signal was used and equalized in pitch with the other signals by "pitch discretization to temperate scale" using the DAFX toolbox (Amatriain et al., 2002), in which the pitch frequency is the common divisor of the spectral peaks in the analyzed frame and derived by the two-way mismatch procedure proposed by Maher & Beauchamp (1994). Then the signals were equalized in level by normalizing the signals' root-mean-square values, and morphed pair-wise (see Sections 3.2.1 and 3.2.2). The output signals were again low-pass filtered at 10 kHz as described above

to avoid high-frequency artifacts from the analysis/synthesis. Finally, a centered section of 0.7 s of each output signal was used and faded in and out with cosine flanks of 0.1 s each.

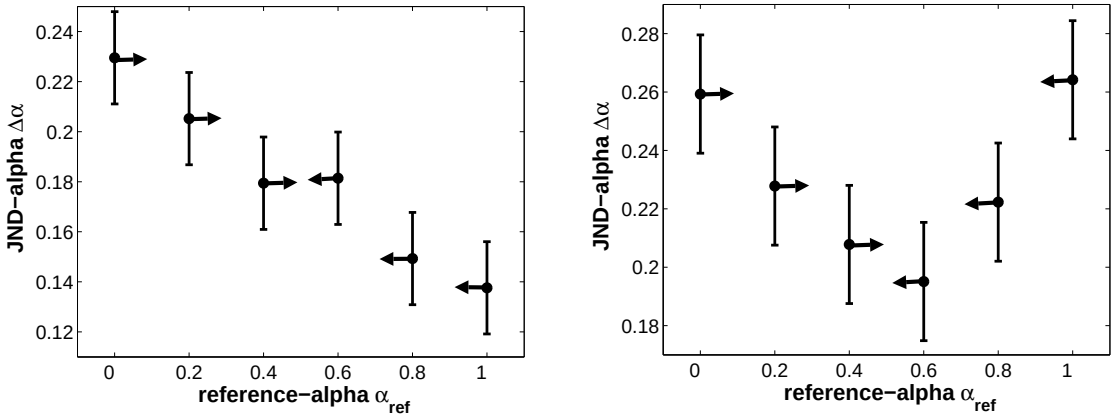
In this way, two instrument continua were generated ("horn-trombone" and "cello-sax") and used in the psychoacoustic measurements described below. The morphed stimuli were defined by their morphing-parameter α, which ranged between 0 (corresponding to the sound of the original French horn or cello) and 1 (trombone or saxophone), with a spacing of 0.01.

3.3 Psychoacoustic JND measurements

3.3.1 Experimental setup

The sounds were presented diotically through ear phones (Sennheiser HD580) in a soundproof booth. The length of test and reference signals was 0.7 s, separated by a silent interval of 0.5 s. All signals were digitally generated on a PC prior to the measurements, output via a digital I/O-card (RME Digi96 PAD) and optically passed to a 24-bit DA converter (RME ADI-8 PRO). The presentation level was calibrated to 65 dB SPL.

In an adaptive 3-alternative forced-choice discrimination experiment, two identical reference signals with morphing-parameter $\alpha = \alpha_{ref}$ and a test signal with adapting $\alpha = \alpha_{test}$ were presented, whereby $\alpha_{test} > \alpha_{ref}$ for $\alpha_{ref} < 0.5$ and $\alpha_{test} < \alpha_{ref}$ for $\alpha_{ref} > 0.5$. The experiment was measured in 12 different conditions with $\alpha_{ref} \in \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$ in two instrument continua. The subjects' task was to indicate which of the three presented signals differed in timbre from the remaining two. By an interleaved 1-up-2-down adaptive tracking procedure, the value of $\Delta\alpha = |\alpha_{ref} - \alpha_{test}|$, at which the test stimulus was chosen correctly with 71% probability (Levitt, 1970), was determined in each condition for 23 normal-hearing subjects (11 female, 12 male) aged between 20 and 46 years. Hence this $\Delta\alpha$ represents a timbre JND. One block consisted of trials of three conditions with $\alpha_{ref} = 0.0$, 0.4 and 0.8 or $\alpha_{ref} = 0.2$, 0.6 and 1.0. Thus all stimuli in one block were from the same instrument continuum, in order to avoid subjects' confusion of detection cues. Trials of the three α_{ref} conditions were presented alternately and in random order within a block. The order of the condition blocks was permuted randomly for each subject. Each subject performed the measurements twice in every condition, and all sessions had an interleaved training block preceding the valued measurements.



(a) horn-trombone continuum

(b) cello-sax continuum

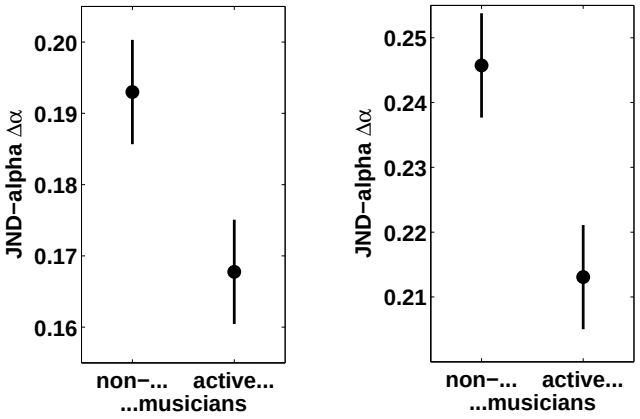
Figure 3.1: Timbre JND (expressed as JND of the morphing-parameter α) as a function of the morphing-parameter α_{ref} of the reference stimulus in the horn-trombone (a) and cello-sax (b) continua. Each plotted data point is the mean of 23 normal-hearing subjects and the 95% confidence interval. The morphing-parameter α represents the ratio of one of the original sounds to the original sounds of the continuum. The arrows indicate in which direction from the reference stimulus the JND was measured; that is, measurements of the center two points had similar stimulus pairs at threshold.

The subjects were interviewed for their musical background and thus divided into two groups: 14 "non-musicians" did not have any experience playing musical instruments or (in the case of 3 subjects) had musical practice in the past but had not actively practised music for at least 2 years prior to the experiment. The 9 "active musicians" were amateur musicians, had at least 4 years of regular experience learning and practising an instrument/singing, and were still actively practising music at the time of the experiment.

4 non-musician subjects performed two extra sessions to verify training effects: The condition with $\alpha_{ref} = 0.0$, 0.4 and 0.8 was measured four additional times in each instrument continuum.

3.3.2 Experimental results

The average results of the timbre JND values for normal listeners are plotted in Figure 3.1. To detect any significant effects of the reference stimulus, subjects' musical experience, and repetition, an analyses of variance (ANOVA) was performed for each instrument continuum with the three factors morphing-parameter α_{ref} (6 levels), musical background (2 levels) and test/retest (2 levels). The main effects of α_{ref} and musical background were highly significant ($p < 0.001$), but the main effect



(a) horn-trombone (b) cello-sax

Figure 3.2: Mean timbre JNDs with 95% confidence intervals in the horn-trombone (a) and cello-sax (b) continua, as a function of the musical background of subjects. For definition of musical background see text.

of test-retest (or repetition) was not significant. No significant interaction effects were observed.

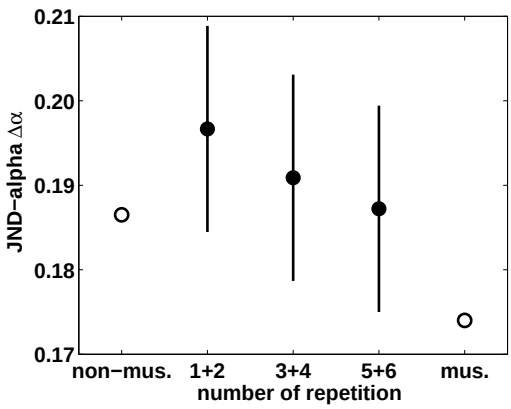
In the horn-trombone continuum (Figure 3.1(a)) the timbre JND was found to decrease stepwise from the horn end ($\alpha = 0$) to the trombone end of the scale ($\alpha = 1$). Specifically, the minimum of $\Delta\alpha$ at $\alpha_{ref} = 1.0$ was significantly smaller ($p < 0.05$) than at $\alpha_{ref} = 0.4$ and 0.6, which was again significantly smaller ($p < 0.05$) than the maximum at $\alpha_{ref} = 1.0$. In the cello-sax continuum (Figure 3.1(b)), the timbre JND showed maxima at either end of the scale ($\alpha_{ref} = 0.0$ and 1.0) and the minimum at the intermediate morphed stimuli ($\alpha_{ref} = 0.6$ and 0.4). In both instrument continua, the minimum was highly significantly ($p < 0.001$) smaller than the maximum. The average JND differences in α between test and retest results was $\alpha = 0.006$ in the horn-trombone and $\alpha = 0.001$ in the cello-sax continuum and did not deviate significantly from zero.

Figure 3.2 shows the JND for each continuum averaged over all α_{ref} and separated by the subjects' musical background. In both instrument continua, the JNDs of active musicians were smaller than those of non-musicians - this difference was highly significant ($p < 0.001$).

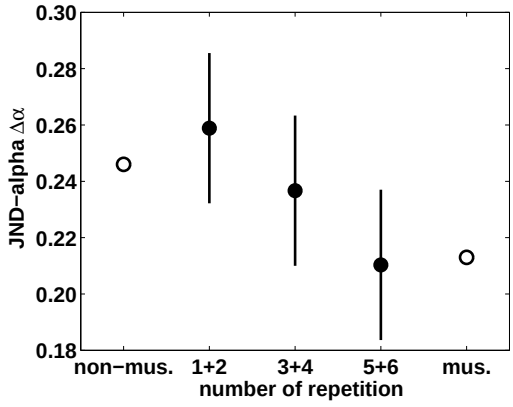
The results of additional training of the non-musician subjects showed distinct but non-significant trends of decreasing JND in both instrument continua (Figure 3.3).

3.4 Effect of spectro-temporal timbre descriptors

Spectral and spectro-temporal dimensions were selected from the literature (Plomp, 1970; Terhardt, 1974; Plomp, 1975; Grey, 1977; Grey & Gordon, 1978; Wessel, 1979;



(c) horn-trombone



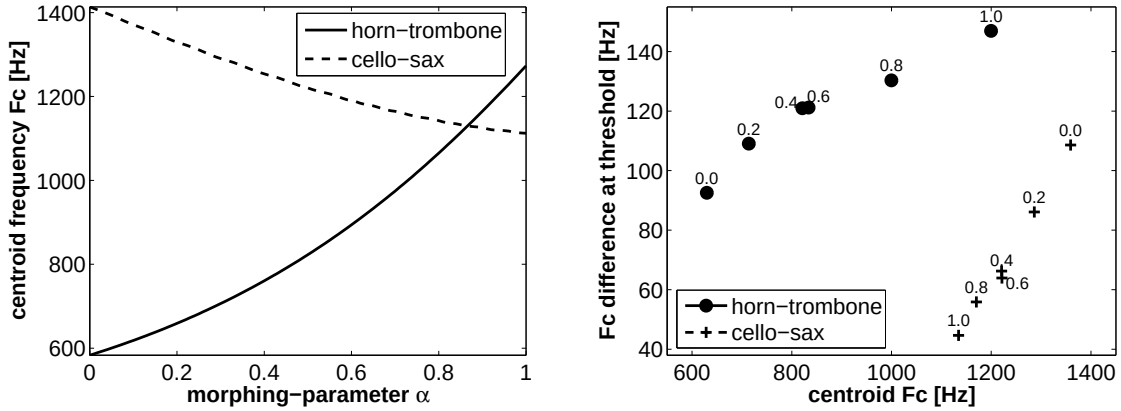
(d) cello-sax

Figure 3.3: Mean timbre JNDs and 95% confidence intervals of 4 non-musician subjects (filled circles) as a function of training. Only the conditions with $\alpha_{ref} = 0.0$, 0.4 and 0.8 were measured. For comparison, open circles show the mean JNDs of conditions $\alpha_{ref} =$ 0.0, 0.4 and 0.8 for non-musicians and musicians.

Krumhansl, 1989; Iverson & Krumhansl, 1993; Krimphoff et al., 1994; McAdams et al., 1995; Kendall et al., 1999; Lakatos, 2000; Pressnitzer & McAdams, 2000; McAdams & Winsberg, 2000; Levitin et al., 2002) to verify which physical parameters vary along the instrument continua employed here. High variation of a relevant timbre descriptor along a continuum and distinct trends of the descriptor difference with the respective descriptor value for stimuli at threshold may indicate which of the common dominating timbre dimensions influence discrimination of stimuli in the JND measurements. Since the attack segment of the stimuli was cut off for the present study, and since the stimuli are tonal sounds, the following spectro-temporal timbre descriptors seem to be relevant for the present study: spectral energy distribution, which is measured in terms of spectral centroid and spectral irregularity, and spectral flux, which is measured in terms of both temporal deviation of spectral centroid and temporal correlation of spectrum (e.g. Grey & Gordon, 1978; Wessel, 1979; Krumhansl, 1989; Iverson & Krumhansl, 1993; Krimphoff et al., 1994; McAdams et al., 1995; Kendall et al., 1999). A detailed spectro-temporal analysis of the morphed stimuli and more information on the timbre descriptors can be found in Appendix A.

3.4.1 Effect of spectral centroid

Since the centroid of the spectrum has been shown to be strongly connected with timbre discrimination (Grey, 1977; Grey & Gordon, 1978; Wessel, 1979; Krumhansl, 1989; Iverson & Krumhansl, 1993; Kendall et al., 1999), in a first step the morphing-



(a) Fc vs. morphing-parameter

(b) Fc difference at threshold

Figure 3.4: Relation between spectral centroid, morphing-parameter α, and obtained timbre JND values: (a) Spectral centroid Fc vs. morphing-parameter α and (b) centroid difference ΔFc vs. centroid Fc of stimuli at threshold for the horn-trombone and the cello-sax continuum. Numbers indicate the morphing-parameter α_{ref} of the reference stimulus of the respective stimulus pair.

parameter α is mapped into the resulting spectral centroid Fc and the effect of Fc on the above results is analyzed. The spectral centroid Fc is defined here as:

$$Fc = \frac{\sum_{k=1}^{N}(A_k \cdot f_k)}{\sum_{k=1}^{N} A_k}, \tag{3.6}$$

where A_k is the amplitude and f_k the frequency of partial k, and N is the total number of partials (e.g. Krimphoff et al., 1994; McAdams & Winsberg, 2000)[10]. Figure 3.4(a) shows that Fc increases monotonically from French horn to trombone and decreases monotonically from cello to saxophone.

The horn-trombone continuum had an Fc range of 690 Hz (or 1.2 octaves from lowest Fc), which is distinctly higher than the range of 300 Hz (or 0.3 octaves) seen in the cello-sax continuum.

Using Equation 3.6 to determine Fc for each of the stimuli employed in this study, the morphing-parameters α_{ref} of the reference stimuli are mapped into centroid frequency Fc_{ref}. The JND results $\Delta\alpha$ measured in units of morphing-parameter (Figure 3.1) are subsequently translated into centroid difference ΔFc at threshold by subtracting Fc of the test stimulus (*test*) from Fc of the reference stimulus (*ref*) at threshold:

$$\Delta Fc = \left| \frac{\int A_{test} \cdot f_{test} df}{\int A_{test} df} - \frac{\int A_{ref} \cdot f_{ref} df}{\int A_{ref} df} \right| \tag{3.7}$$

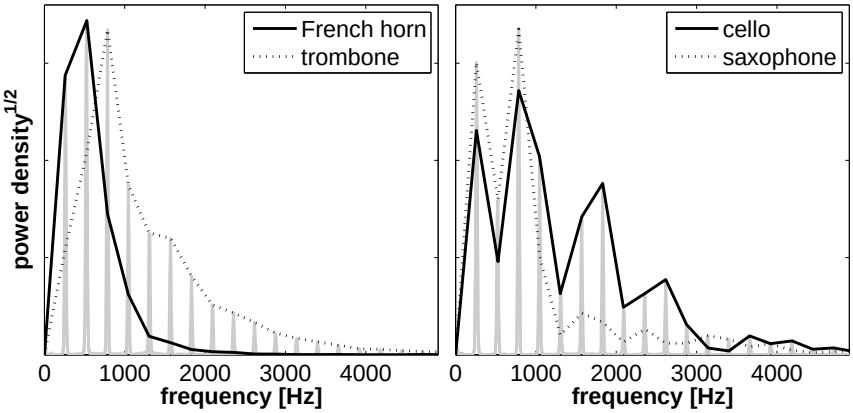


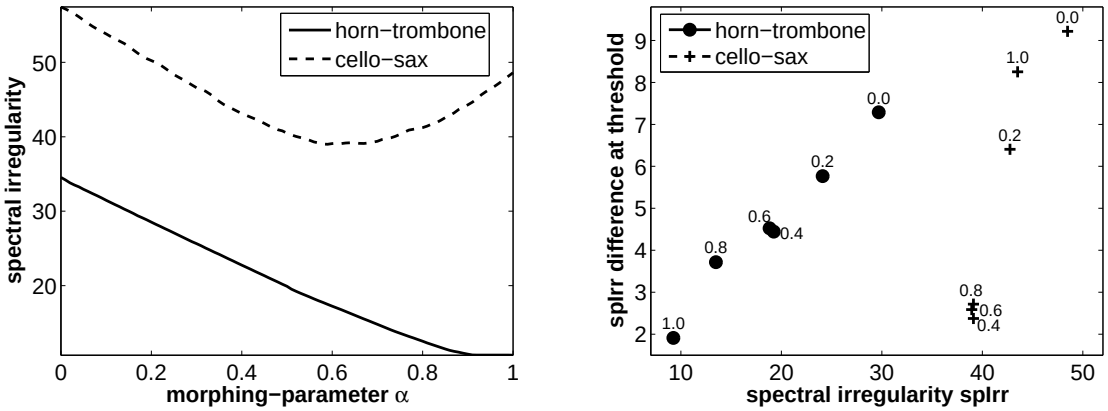
Figure 3.5: Mean spectra of the natural instruments in the horn-trombone (left) and cello-sax (right) continua. The spectra are shown in grey, while the spectral peaks are connected by black lines indicating the instrument (see legend).

Centroid difference ΔFc at threshold as a function of the mean centroid Fc of the stimuli at threshold is shown in Figure 3.4(b) for the horn-trombone (filled circles) and cello-sax (cross symbols) continua.

In contrast to JND-α, ΔFc at threshold increases with increasing Fc in both continua (compare Figures 3.1 and 3.4(b)). The symbols of the cello-sax continuum in Figure 3.4(b) lie below and to the right of the horn-trombone symbols; and in the range Fc=1112-1273 Hz, where thresholds were measured in both instrument continua, ΔFc in the cello-sax continuum are lower than in the horn-trombone continuum. Figure 3.4(b) also shows that the ΔFc growth, i.e. $\frac{\Delta Fc}{Fc}$, in the cello-sax continuum is higher than in the horn-trombone continuum. Both findings are consistent with the notion that Fc differences are the salient cues for detecting changes in the horn-trombone continuum but only minor cues in the cello-sax continuum (see discussion).

3.4.2 Effect of spectral irregularity

In addition to the centroid of the spectrum, the spectral irregularity was found to be a timbre dimension perceived in musical instruments (Krimphoff et al., 1994). For instance, spectra of clarinet sounds show mainly harmonics with odd partial numbers, whereas the even-numbered partials are missing or of low amplitude. Hence, clarinets show a high spectral irregularity. A measure for the spectral irregularity *spIrr* is the logarithm of the (spectral) deviation of component amplitudes from a global spectral envelope derived from a running mean of the amplitude of three adjacent harmonics (Krimphoff et al., 1994).



(a) spIrr vs. morphing-parameter

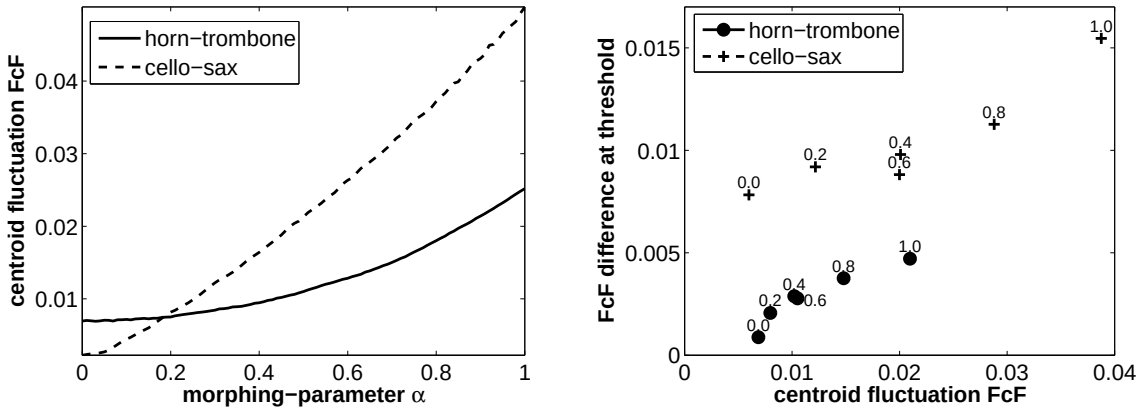
(b) spIrr difference at threshold

Figure 3.6: Relation between spectral irregularity, morphing-parameter α and obtained timbre JND values: (a) Spectral irregularity spIrr vs. morphing-parameter α and (b) spectral irregularity difference ΔspIrr vs. spIrr of stimuli at threshold for the horn-trombone and the cello-sax continuum. Numbers indicate the morphing-parameter α_{ref} of the reference stimulus of the respective stimulus pair.

Figure 3.5 shows the spectra of the "end stimuli", i.e., spectra of French horn, trombone, cello and saxophone. Similar to a clarinet sound, the spectra in the cello-sax continuum show irregular harmonic amplitudes, while the spectra in the horn-trombone continuum show smooth envelopes (Figure 3.5). Spectral irregularity in the horn-trombone continuum is lower, but variance of spectral irregularity along stimuli is higher than in the cello-sax continuum (see Figure 3.6(a) and Appendix A for detail). JND-α results are translated into spectral irregularity, as was done for Fc in Equation 3.7, and spectral irregularity difference $\Delta spIrr$ at threshold is shown in Figure 3.6(b). In the horn-trombone continuum, $\Delta spIrr$ increases with $spIrr$ at threshold. Note that this relation might be predicted from ΔFc as a function of Fc (Figure 3.4), because $spIrr$ is inversely correlated to the Fc trend (Figure 3.6(b), see also Chapter 2, Table 2.1). In the cello-sax continuum, no distinct trend can be observed (Figure 3.6(b)).[11]

3.4.3 Effect of spectral flux

Spectral flux and the presence of synchronicity in overtones can be thought of as two views of the same spectro-temporal dimension. This dimension has been shown to be connected with timbre rating and discrimination, although it is discussed controversially as possibly not being independent from other dominating timbre dimensions (Grey, 1977; Grey & Gordon, 1978; Wessel, 1979; Krumhansl, 1989;



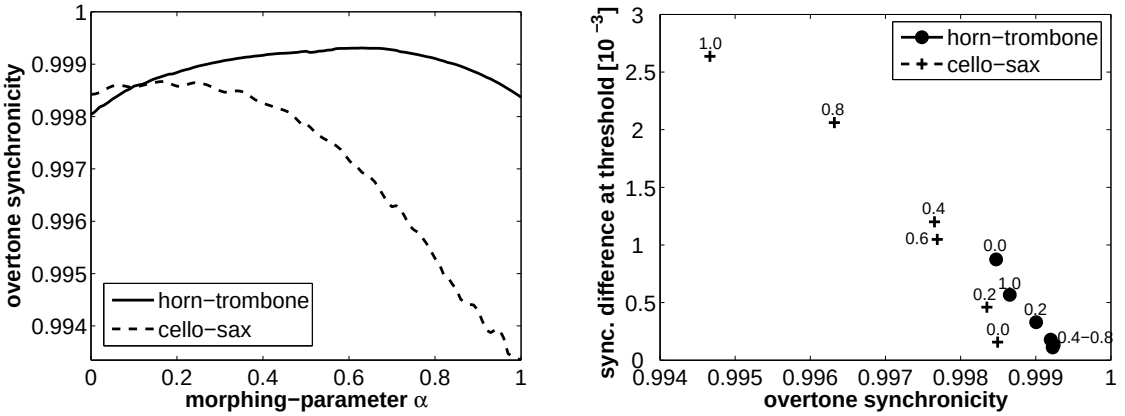
(a) Fc fluctuation vs. morphing-parameter (b) Fc fluctuation difference at threshold

Figure 3.7: Relation between spectral flux measured in terms of centroid fluctuation., morphing-parameter α, and obtained timbre JND values: (a) Standard deviation of the centroid over running 93 ms windows ($\hat{=}$ 11Hz sampling rate) vs. morphing-parameter α, and (b) standard-deviation difference vs. standard deviation of stimuli at threshold for the horn-trombone and the cello-sax continuum. Standard deviation is shown as a ratio to mean centroid. Numbers indicate the morphing-parameter α_{ref} of the reference stimulus of the respective stimulus pair.

Iverson & Krumhansl, 1993; Kendall et al., 1999).

One measure correlated with the spectral flux is the temporal fluctuation of the spectral centroid (McAdams et al., 1999). To test the degree to which the centroid fluctuation (FcF) varies along the instrument continua employed here, we calculate FcF as the standard deviation of spectral centroids along a running time window of 93 ms over the duration of the stimulus and normalize it by the mean spectral centroid Fc. The spectral centroid in each window is calculated using Equation 3.6. Figure 3.7(a) shows the variation of FcF along the instrument continua, which increases with α in both continua. The FcF range in the horn-trombone continuum is less than half of the FcF range in the cello-sax continuum.

In the same way as described above with Fc (Equation 3.7), the morphing-parameters α of the stimuli and the JND results $\Delta\alpha$ (Figure 3.1) are translated into centroid-fluctuation measures (ΔFcF, FcF$_{ref}$ and FcF$_{test}$). Centroid-fluctuation difference ΔFcF at threshold is shown in Figure 3.7(b) as a function of the mean centroid fluctuation of stimuli at threshold. Both continua show at threshold an increase of ΔFcF with increasing FcF. All ΔFcF at threshold in the horn-trombone continuum are lower than those in the cello-sax continuum, in particular at equal FcF. This is consistent with the notion that spectral flux differences are salient cues for the timbre discrimination in the cello-sax continuum, while this cue is irrelevant



(a) synchronicity vs. morphing-parameter (b) synchronicity difference at threshold

Figure 3.8: Relation between spectral flux measured in terms of overtone synchronicity, morphing-parameter α, and obtained timbre JND values: (a) Synchronicity of overtone spectra in adjacent 46 ms windows vs. morphing-parameter α, and (b) synchronicity difference vs. synchronicity at threshold for the horn-trombone and the cello-sax continuum. The synchronicity was calculated by the Pearson product of the harmonic spectra between 2 and 5 kHz. Numbers indicate the morphing-parameter α_{ref} of the reference stimulus of the respective stimulus pair.

for the trombone-horn continuum (see discussion).

Another measure for spectral flux is the average of the correlations between amplitude spectra in adjacent time windows (Krimphoff et al., 1994) (see Appendix A for details). Since Grey (1977) already described the flux dimension as the "presence of synchronicity in the upper harmonics" the analysis was done only correlating the spectra above 840 Hz (without the lowest 3 harmonics) and above 2 kHz (without the lowest 7 harmonics). The literature is not clear about which harmonic numbers contain the crucial synchronicity that is perceived as spectral flux cue. However, the overtone synchronicity for spectra above 840 Hz showed the same trend as for spectra above 2 kHz, but to a lower extent (not shown). Therefore, Figure 3.8 shows the overtone synchronicity for harmonic numbers 8 to 19. The horn-trombone continuum shows high synchronicity and hence, low spectral flux, while synchronicity decreases and, hence, spectral flux increases from cello to saxophone (Figure 3.8(a)). For stimuli at threshold in the cello-sax continuum, synchronicity difference decreases with increasing synchronicity, and hence, flux difference decreases with decreasing flux.

3.5 Discussion

The measurements show that JND depends significantly on the morphing-parameter of the reference sound and that the function JND(α_{ref}) differs between the two instrument continua. This indicates a non-uniform mapping between the physical parameter change (as given by the parameter α) and the perceptual continuum between the respective endpoints. While this mapping is monotonic between French horn and trombone, which are characterized primarily by differences in spectral centroid, an u-shaped mapping results between cello and saxophone, which are characterized by differences in the time domain rather than in a distinct spectral change. Hence, the different JND-to-α_{ref} relations may reflect different roles of temporal and spectral cues in timbre perception for the stimuli employed here. Another possible explanation for the u-shaped relation between JND and morphing-parameter α might be categorical perception at the endpoints of the α-scale, that is to say, a percept "attraction" to the endpoint, which makes any physical deviation from the end less easily perceived. However, the attack-reduced stimuli make instrument recognition difficult (Taylor, 1992; Cook, 1999; Levitin et al., 2002), which makes categorical perception in the cello-sax continuum unlikely.

With the method described here, JND results could be measured in a reproducible way with high precision, so that significant differences due to instrument group, reference sounds and subjects' musical background can be observed. This allows the use of JND measurements for comparing differences in timbre perception between different listener groups and different stimuli. In addition, Chapter 2 showed that the method is also applicable (with restrictions) to stimuli with their natural attack segment, which was cut off in the present study. However, in both instrument continua, active musicians showed significantly lower JNDs than non-musicians. This must be taken into account when using the method described here for studying timbre discrimination in different subject groups, for example non-musicians or hearing-impaired listeners.

Using morphed stimuli instead of artificial tone complexes enables the study of timbre perception and JND of natural objects, such as musical instrument sounds. However, when searching for perceivable timbre dimensions and their JNDs in terms of physical aspects (frequency, time, amplitude and phase), some care has to be taken in order not to miss any dimension. Some timbre characteristics might be missing in the stimuli, because they were diluted by the analysis/morphing/synthesis: the limits of any psychoacoustic method using synthesized stimuli are connected to the limitations of the frequency and time resolution of the short-time Fourier analysis and synthesis employed in preparing the stimuli. In theory, this would not be

problematic since an exact reconstruction of the original signal would be achieved if the original phase information is preserved. In the current morphing algorithm, however, the phase is determined by the underlying reconstruction method (see Section 3.2.2), which constructs the phase in each time frame for those frequency components that have been selected as spectral peaks. This phase reconstruction is limited by the time resolution of the analysis/synthesis because

1. the instantaneous frequency averaged in each time frame is used to set the frequency of the reconstructed spectral peak across each whole time frame, and

2. the initial phase (of each spectral peak) in each reconstructed time frame is determined by the phase at the end of the previous time frame (see Equation 3.3). This effect even accumulates across time frames and yields differences between reconstructed and original signals.

If these differences become too large (for example, if the analysis/synthesis frame rate is too low), they may become audible. The reconstructed morphed signal might therefore not convey all cues normally accessible for discriminating the original sounds and, hence, may be missing some timbre cues. On the other hand, the reconstruction method employed is not necessarily limited by the frequency resolution of the short-time Fourier analysis. The maximal frequency resolution given by the sampling frequency (Fs=44100 Hz) and FFT length (w=1024) would be only 43 Hz. However, zero padding of the analyzed signal windows and an additional interpolation of the spectrum around the spectral peaks enables detection of their instantaneous frequency in an accurate way, so that even the small natural frequency fluctuations of partial tones are preserved in the current analysis/synthesis scheme quite well. Hence, the analysis/synthesis method underlying the morphing algorithm employed here provides sufficient accuracy for performing psychoacoustic experiments, if a sufficiently high frame rate (> 200 Hz) is selected.

Physical parameters and timbre dimensions

Many studies have verified the primary timbre dimensions that provide the physical basis of rating and discrimination of musical instrument sounds (e.g. Grey & Gordon, 1978; Wessel, 1979; Krumhansl, 1989; Iverson & Krumhansl, 1993; Krimphoff et al., 1994; McAdams et al., 1995; Kendall et al., 1999). Here, the spectro-temporal timbre descriptors found in the literature are assumed to provide a complete representation of primary timbre dimensions. Using these timbre dimensions, Appendix A gives a detailed analysis of the morphed stimuli with attack and provides the basis for the

present study. Thus, Section 3.4 analyzes all of the literature's spectro-temporal timbre descriptors that seem to be relevant for the tonal stimuli without attack segment used in the present study.[12] High variation of a certain timbre descriptor along a continuum and low variation of the remaining descriptors may imply that this descriptor provides a salient cue for distinguishing the stimuli in the respective continuum. A distinct descriptor difference at threshold for the respective descriptor (Section 3.4) suggests that the descriptor has been used by subjects as a cue for distinguishing the stimuli in the respective continuum. A distinct trend of descriptor difference with descriptor value for stimuli at threshold, for example a trend according to Weber's law, may confirm the assumption.

In the literature, the dimensions spectral flux and spectral irregularity are discussed controversially as dominant and independent timbre dimensions. Some other dimensions, such as graininess, inharmonicity or presence of a clunk, seem to only be present or crucial for specific instruments or subjects (McAdams et al., 1995). However, the centroid of the spectrum, which describes the brightness percept, and the initial attack have commonly been shown to be strongly correlated with the most prominent dimensions of multidimensional-scaling representations of timbrel differences (Grey & Gordon, 1978; Wessel, 1979; Krumhansl, 1989; Iverson & Krumhansl, 1993; Kendall et al., 1999). Since the attack segment of the stimuli was cut off for the present study, one might conjecture that a listener's ability to detect stimulus differences is due to detection of centroid differences rather than other timbre dimensions.

The stimulus pairs were chosen in a way that the instruments saxophone and cello had distinctly different spectro-*temporal* parameters, whereas the spectral centroid was similar (Section 3.2.3 and Appendix A). Although in synthesized tones spectral centroid can be controlled independently of other spectral-amplitude modifications, they are not necessarily separable in musical instrument sounds. Hence, spectral centroids are similar but not identical in the cello-sax continuum. This could be perceived as slight brightness differences between stimuli in the cello-sax continuum, whereas brightness differences in the horn-trombone continuum were distinctly perceivable.

The main findings of the measurements with regards to **spectral centroid** can be summerized as follows:

1. In contrast to JND-α, centroid difference ΔFc at threshold increased with increasing centroid Fc in both continua (compare Figures 3.1 and 3.4(b)).

2. For comparable Fcs, ΔFc in the cello-sax continuum were lower than in the horn-trombone continuum; in the cello-sax continuum, subjects could distin-

guish stimuli with lower Fc differences than in the horn-trombone continuum.

3. ΔFc growth at threshold in the cello-sax continuum was higher than in the horn-trombone continuum.

According to Weber's law, the frequency difference limen Δf increases with frequency f. Hence, finding 1 in combination with finding 2 suggests that spectral centroid was a dominant distinction cue at least for the horn-trombone continuum. In the cello-sax continuum, it may be a distinctive cue for low α values, while the ΔFc values may have been below discrimination thresholds for high values of α.

Finding 2 indicates that either subjects used an additional cue to distinguish stimuli in the cello-sax continuum, or some stimulus feature distracted subjects from distinguishing centroids in the horn-trombone continuum. Spectral centroid varied in the horn-trombone continuum by approximately 690 Hz from 583 to 1273 Hz, but in the cello-sax continuum by only approximately 300 Hz from 1112 to 1414 Hz. In Grey's (1977) study, the perceptual difference between saxophone and cello stimuli was influenced mainly by spectral flux and less by spectral centroid, while the reverse was true for trombone and French horn. In any case, finding 1 suggests that spectral centroid also played a role in distinguishing the stimuli in the cello-sax continuum, whereas Grey's (1977) MDS makes it most likely that the lower ΔFcs (2) and higher ΔFc growth (3) resulted from an additional cue (along with Fc) dominating the perceptual differences in the cello-sax continuum.

In order to verify the influence of the remaining timbre dimensions on the results, dimensions were selected that seemed to strongly influence timbre rating and discrimination in previous studies: spectral irregularity (another spectral dimension) and spectral flux (a spectro-temporal dimension).

The **spectral irregularity** as defined by Krumhansl (1989) varies in the horn-trombone continuum more than in the cello-sax continuum (Figure 3.6(a)). However, as the spectra show, no "real" irregularity, that is to say altering partial amplitudes, can be observed in the horn-trombone continuum but rather a narrow spectral peak in the horn in contrast to a broader spectral distribution in the trombone (Figure 3.5). While the systematic trend in the horn-trombone continuum is probably correlated with the centroid shift (see Appendix A, Table 2.1), no systematic correlation was found for stimuli at threshold with respect to spectral irregularity in the cello-sax continuum. This indicates that spectral irregularity does not dominate the discrimination task.

Spectral flux is discussed controversially as a dominant timbre dimension independent from other dominating timbre dimensions (Grey, 1977; Grey & Gordon,

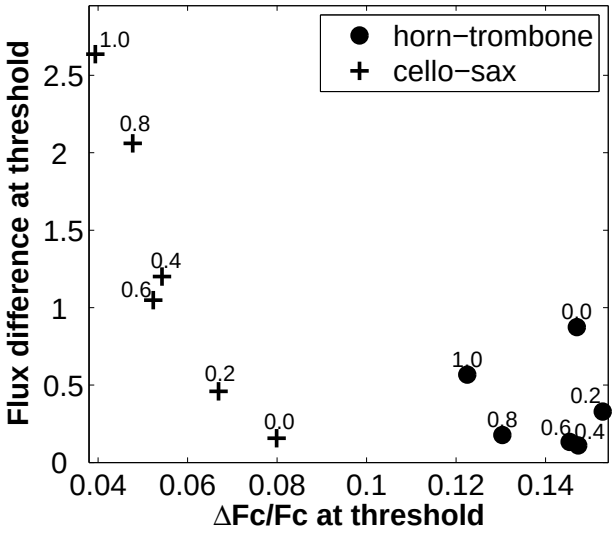


Figure 3.9: Spectral vs. spectro-*temporal* cues: Spectral flux in measures of overtone synchronicity [10^{-3}] vs. Weber fraction of spectral centroid for the horn-trombone and cello-sax continuum. Numbers indicate morphing-parameter α_{ref} of reference stimulus.

1978; Wessel, 1979; Krumhansl, 1989; Iverson & Krumhansl, 1993; McAdams et al., 1995; Kendall et al., 1999). In this study, the centroid fluctuation and overtone synchronicity (i.e. spectral correlation of adjacent time windows) were chosen as measures for spectral flux.[13] In the horn-trombone continuum, both centroid fluctuation and overtone synchronicity showed distinctly lower variation than in the cello-sax continuum, while the centroid Fc showed distinctly higher variation (compare Figures 3.7(a), 3.8(a) and 3.4(a)). This confirms the starting point that the pairs of musical instruments were chosen such that one pair (trombone and French horn) differed greatly in their spectral centroids and the other (saxophone and cello) in their spectral flux, whereas the other timbre descriptors were similar within each pair. The high variance of the spectral flux along the cello-sax continuum, the high values of the spectral flux (i.e., high centroid fluctuation and low overtone synchronicity) and the high flux difference at threshold for stimuli with α_{ref}=0.4-1.0 indicate that spectral flux influences discrimination in the cello-sax continuum (Figures 3.7(a) and 3.8(a)).

However, spectral flux difference at threshold increases distinctly from α_{ref}=0.0 to α_{ref}=1.0 in the cello-sax continuum, which may reflect increasing importance of the spectro-*temporal* cue compared to the spectral centroid shown above. To illustrate the different interference of centroid and flux in the two instrument continua, Figure 3.9 shows the flux difference (in measures of overtone synchronicity) as a function of the centroid difference ΔFc for stimuli at threshold, whereby the centroid difference is given as a ratio of ΔFc to Fc. In the horn-trombone continuum, spectral flux differences are low and the Fc Weber fraction is relatively constant along the continuum ($\frac{\Delta Fc}{Fc} = 0.12 \pm 0.02$), while Fc varies distinctly and systematically (Figures 3.9 and 3.4). In the cello-sax continuum, $\frac{\Delta Fc}{Fc}$ is distinctly smaller and

decreases with α_{ref}, while the flux difference simultaneously increases (Figures 3.9). The distinction cue used for discriminating the stimuli at threshold seems to successively shift along the cello-sax continuum from spectral to spectro-*temporal* timbre descriptors (Figure 3.9).

Hence, while the horn-trombone continuum seems to be nearly exclusively varied by the spectral centroid, signal variation along and timbre discrimination in the cello-sax continuum seem to be influenced by multiple cues like spectral centroid and spectral flux. The concurring importance of spectral and spectro-*temporal* cues used to distinguish the stimuli may also explain the u-shaped relation between JND-α and morphing-parameter α_{ref} (Figure 3.1(b)). The Weber fraction of centroid frequency for the stimuli in the cello-sax continuum that are discriminated mainly by spectral cues is here $\frac{\Delta Fc}{Fc}$=0.06-0.08 (α_{ref}=0.0-0.2), which is distinctly lower than $\frac{\Delta Fc}{Fc} = 0.12$ found in the trombone-horn continuum. Both continua show distinctly higher Weber fractions than for frequency discrimination of pure tones, which is $\frac{\Delta f}{f}$=0.002 to 0.01 (for frequencies below 4 kHz; Moore, 2003). The Weber fraction of centroid frequency for complex tones seems to depend on spectral content and present spectral flux, even if no other cues can be used for discrimination.

3.5.1 Conclusion

The present study showed that stimuli generated with the morphing method described above can be used to study timbre discrimination independently from subjects' previous acquaintance with the stimulus names. The JND measurement method enables the comparison of timbre perception of different listener groups, such as subjects with different musical experience. Even untrained subjects give results that reveal small systematic stimuli differences. However, the detected JND difference between active musicians and non-musicians must be taken into account when studying timbre perception in different subject groups. The observed training effects in non-musicians require sufficient training of subjects preceding the measurements.

The merit of the new technique over previous methods includes the following advantages and possibilities:

- Using quasi-natural timbres preserves all natural timbre attributes, ensures that subjects are familiar with the stimuli and makes results applicable to real objects.
- The measurements are independent of earlier knowledge, analyzing primarily the subject's bottom-up processes.

- The technique allows one to measure small perception differences, for example between normal-hearing and hearing-impaired subjects, by measuring discrimination differences in JND measurements as well as by verifying dimension differences using morphed sounds for similarity rating tasks and MDS.

- The method enables the measurement of perception limits along large timbre distances, for example measuring and comparing perception thresholds in and between instrument families.

- The morphing method makes it possible to verify timbre dimensions that are only present and specific in certain instruments (McAdams et al., 1995).

The systematic change of timbre JND with reference stimulus could be explained with spectro-temporal timbre descriptors. In the horn-trombone continuum, centroid difference at threshold increases proportionally with stimulus centroid in conformance with Weber's law, while no distinct systematic change of the other descriptors was found. This suggests that spectral centroid is the dominant distinction cue for the stimuli in the horn-trombone continuum. The Weber fraction found for this continuum is $\frac{\Delta Fc}{Fc} = 0.12 \pm 0.02$, which is distinctly higher than the Weber fraction for pure tones ($\frac{\Delta f}{f} < 0.01$ for frequencies below 4 kHz). In the cello-sax continuum, the cue used for discriminating the stimuli at threshold seems to shift successively along the continuum from spectral to spectro-*temporal* descriptors. While discrimination seems to be strongly influenced by spectral centroid at the cello end, spectral flux seems to play a major role at the saxophone end of the continuum.

Chapter 4

Timbre discrimination in normal-hearing and hearing-impaired listeners under different noise conditions

Abstract

In an attempt to quantify differences in object segregation and timbre discrimination between normal-hearing and hearing-impaired listeners with a moderate sensorineural hearing loss of two different configurations, psychoacoustic measurements were performed with a total of 50 listeners. The experiments determined just noticeable differences (JND) of timbre in normal-hearing and hearing-impaired subjects along continua of "morphed" musical instruments and investigated the variance of JND in silence and different background-noise conditions and on different sound levels. The results show that timbre JNDs of subjects with a steep hearing loss are significantly higher than of normal-hearing subjects, both in silence and noise, whereas timbre JNDs of flat/diagonal hearing-impaired subjects are similar to JNDs of normal-hearing subjects for signal levels above 55 dB (plus appropriate amplification for hearing-impaired). In noise (SNR=+10 dB) timbre JNDs of all subject groups are significantly higher than in silence. In the condition testing transferability from silence to noise, no significant JND differences across listener groups were found. The results can be explained by primary factors involved in sensorineural hearing loss and contradict the hypothesis that hearing-impaired people generally have more problems in object discrimination than normal-hearing people.

4.1 Introduction

People with sensorineural hearing loss including hearing aid users often have problems with timbre distortion. This affects not only music perception, but may also influence object recognition in general. A common hypothesis argues that the reduced frequency selectivity in hearing-impaired people leads to a reduced ability in distinguishing timbre and, hence, sounds of musical instruments (Moore, 2003). It is still unproven, whether and for which conditions this statement holds true; that is, if and how the ability changes for different types and severities of hearing loss, with different sound types, and in the presence of other sounds. The present study aims, amongst others, to test this hypothesis.

The interpretation of auditory scenes in a real acoustical environment is a complex, poorly understood task of the auditory system, which is distinctly disturbed in hearing-impaired people. While the negative influence of a sensorineural hearing impairment has been proved in nearly all psychoacoustically ascertainable hearing functions (e.g. Festen & Plomp, 1983; Moore, 1998), the systematic interdependence of the individual hearing functions among each other is still quite unclear. That is, a chain of cause and effect between primary influencing factors and secondary factors that result from the disturbance of the primary factors is controversial. Physiologically identified primary factors are the loss of inner and outer hair cells (IHC and OHC) causing primarily loss in sensitivity and compression, respectively. Loss of binaural interaction and a larger internal noise may also be candidates for primary factors (Kollmeier, 1999), while the role of time and frequency resolution is discussed controversially (Launer et al., 1997; Moore, 1998). Likewise, the influence of the psychoacoustic functions that are disturbed with hearing impairment on speech intelligibility in silence and in noise, and on general object segregation, is not yet resolved unambiguously. It is commonly accepted, however, that the alteration in the compressive nonlinearity caused by OHC loss is a major cause of most of the perceptual changes observed in cochlear hearing loss (Bacon et al., 2004).

In order to study the consequences of the compressive non linearity that is altered in hearing-impaired listeners with regards to object segregation, psychoacoustic measurements are performed in the present study with the object feature “timbre”, which is also used to separate auditory objects (Iverson, 1995). Timbre is a combination of all acoustical attributes that are not exclusively assigned to the perception of pitch, loudness or length (American Standard Association, 1960; Plomp, 1970); that is to say, timbre is a multidimensional perception measure. Most of previous studies quantified the dominating acoustical attributes in normal-hearing listeners by similarity rating experiments (e.g. Grey, 1977) and studied the recognition of

musical instruments in cochlea implant users (e.g. Gfeller et al., 2002b), but did not consider the acoustical hearing-impaired listeners, especially hearing aid users. For determining the small perception differences between acoustical hearing-impaired and normal-hearing listeners, a method to determine an objective comparison measure of different timbres has been developed (Chapter 3), which is applied in the present study.

In order to characterize object discrimination abilities, subjects were asked to distinguish timbres of natural (musical) but unidentifiable sounds in the presence of different background noises. In an attempt to quantify differences in timbre perception between normal-hearing and hearing-impaired listeners in terms of temporal and spectral dimensions (e.g. spectral centroid, spectral fluctuation), psychoacoustic measurements on timbre perception were performed with both groups of listeners. By linear interpolation of spectral parameters, sounds of musical instruments were cross-faded ("morphed") along these dimensions, thus generating stimulus continua between natural instruments. The experiments determined just noticeable differences (JNDs) of timbre in normal-hearing listeners and hearing-impaired listeners along these instrument continua. The JNDs were measured in silence, in different background noises and on different sound levels.

The control measurements in silence test Moore's (2003) hypothesis that hearing-impaired listeners have problems distinguishing timbres due to reduced frequency selectivity. If Moore's hypothesis is correct, timbre discrimination thresholds in hearing-impaired listeners shall be higher than in normal-hearing listeners at the same sensation level, because the frequency resolution in sensorineural hearing-impaired listeners is supposed to be lower than for normal-hearing listeners. This difference should get larger with decreasing sensation level. In noise measurements subjects had to use timbre to segregate the tonal stimuli from the noise background before distinguishing the instrumental sounds. If compression loss reduces the ability to separate simultaneous auditory objects, JNDs of flat and steep hearing-impaired listeners shall be higher than JNDs of normal-hearing listeners. Hence, the noise condition tests the hypothesis that hearing-impaired listeners listeners exhibit a poorer object separation ability even at the same sensation level as normal-hearing listeners.

The crucial *Transfer* condition in this study examines the ability to deduce from how the stimulus is heard in silence to how it would sound in noise. This experiment approaches the hypothesis of an "object invariance" or even "object linearity" in normal-hearing listeners, i.e. whether the percept of one auditory object changes when adding a second auditory object. If object perception is linear and cross-masking effects could be neglected, then subjects should be able to transfer the

percept of the signal into noise and decide which of the signals heard in noise was equal to the signal heard in silence. Due to masking effects of the noise, object invariance can only be assumed to a certain degree. However, if the hypothesis of object invariance is true, JNDs of the *Transfer* condition should be comparable to JNDs of the condition where the reference sound was also heard in noise. By comparing the respective performance in the *Transfer* condition, we can test the hypothesis that hearing-impaired listeners generally have more problems in object segregation than normal-hearing listeners, due to, for example, loss in object linearity.

4.2 Psychoacoustic measurements

4.2.1 Stimuli

Two pairs of musical instruments were chosen such that one pair (trombone and French horn) had very different spectral centroids and the other (saxophone and cello) different temporal flux; all other physical timbre descriptors were similar in each pair (Grey, 1977). For Experiment A (see Section 4.2.2), in addition a pair with variation in both spectral and temporal aspects (flute and trumpet) was chosen. First, acoustic recordings (Fritts, 2002) of these instruments pitched at C4 ($f_0 \approx 262\ Hz$) were low-pass filtered at 10 kHz using a linear-phase FIR filter; this was done for better audiological comparison of hearing-impaired with normal-hearing listeners. In order to avoid recognition of the instruments and, hence, categorization of the sounds, and because the perceived length of the sound depends on the attack length (McAdams et al., 1995), the attack sequence was cut off. An approximately stationary section of 0.7 s of the remaining signal was used and equalized in pitch and level with the other signals. By linear interpolation of spectral parameters, sounds of musical instruments were then pair-wise cross-faded ("morphed") within the respective instrument group. Three stimulus continua, one between trombone and French horn ("horn-trombone"), another between saxophone and cello ("cello-sax"), and the third between flute and trumpet ("flute-trumpet") were generated. The morphing used an overlap-add analysis-synthesis algorithm based on a sinusoidal plus residual model (Amatriain et al., 2002) and interpolated instantaneous frequency, amplitude and phase of the sinusoidal part (i.e. the tonal partials) as well as the amplitudes of the residuum (i.e. remaining noise portion of the sound). The analysis/synthesis window used extended 23 ms plus 23 ms zero-padding. The output signals were again low-pass filtered at 10 kHz. Finally, a centered section of 0.5 s of each output signal was used which faded in and out with cosine flanks of 50 ms each. A more detailed description of the morphing method can be found in

Chapter 3.

In this way, three instrument continua were generated and used in the psychoacoustic measurements described below. The morphed stimuli were named by their morphing-parameter α, which corresponds to the ratio of the second (original) instrument in the morphed sound. Hence, α ranges between 0 (e.g. corresponding to the sound of the original French horn) and 1 (e.g. trombone), with a spacing of 0.01.

4.2.2 Experimental setup

The sounds were presented diotically through ear phones (Sennheiser HD580) in a soundproof booth. The length of test and reference signals was 0.7 s in Experiment A and 0.5 s in Experiment B, in both experiments separated by a silent interval of 0.5 s. All signals were digitally generated on a PC prior to the measurements, output via a digital I/O-card (RME Digi96 PAD) and optically passed to a 24 bit DA-converter (RME ADI-8 PRO). For Experiment A, the presentation level of the instrumental sounds without noise (see below) was calibrated to 65 dB SPL and for Experiment B to 57.5 dB SPL for the normal-hearing subjects.

Experiment A

Experiment A determined the timbre JND at the end-points of the morphing continua. For the hearing-impaired subjects the signal level of 55 dB SPL was amplified (and shaped in frequency) by half of the respective subject's frequency-dependent hearing loss ($+\mathrm{H}_f$). This resulted in an approximately loudness-matched presentation.

An adaptive 3-alternative forced-choice discrimination experiment was performed. Two identical reference signals with morphing-parameter $\alpha = 0$ and a test signal with an adapting parameter $\alpha = \alpha_{test}$ were presented. The subjects' task was to indicate which of the three presented signals "sounded different" than the remaining two. Feedback[14] was given throughout the entire experiment. A 1-up-2-down tracking rule was used to adaptively approach the discrimination threshold. After three training runs, each subject performed the measurements twice in each instrument continuum. The order of the instrument continua in each repetition was permuted randomly for every subject.

During the first 6 reversals the individual threshold region was approached, while the tracking variable $\Delta\alpha = \alpha_{test}$ was adapted by adding/subtracting 0.10 (up to the 2nd reversal), 0.04 (up to the 4th reversal) or 0.01 (after the 4th reversal). By

the 1-up-2-down adaptive tracking procedure, the value of $\Delta\alpha$, at which the test stimulus was chosen correctly with 70.7% probability (Levitt, 1970), was determined by averaging the tracking variables for the 5th to 12th reversals. Hence, this $\Delta\alpha$ represents a timbre JND in the respective timbre continuum.

Experiment B

Experiment B determined the timbre JND at the mid-point of the respective timbre continuum. An adaptive 3-interval 2-alternative forced-choice discrimination experiment was performed: two identical reference signals with morphing-parameter $\alpha = \alpha_{ref}$ and a test signal with $\alpha = \alpha_{test}$ were presented.

$$\Delta\alpha = \alpha_{test} - \alpha_{ref} \tag{4.1}$$

was the parameter, whereby

$$\alpha_{test} - 0.5 = 0.5 - \alpha_{ref}. \tag{4.2}$$

In every trial it was randomly defined whether $\alpha_{test} > 0.5 > \alpha_{ref}$ or $\alpha_{test} < 0.5 < \alpha_{ref}$. The test sound to be detected was presented in either the 2nd or 3rd interval.

In some conditions, background noise with a signal-to-noise ratio (SNR) of 10 dB was presented in addition to the signals; the noise was faded in and out with cosine flanks synchronous with the signal. The experiment was performed using 4 different noise conditions, employing running noise, and on two different levels (see Table 4.1). In conditions *Quiet* and *Quiet-Low* the signals were presented in silence, that is without any background noise, and in conditions *Pink* and *Pink-Low* an additional pink noise was presented. In condition *Shaped*, a noise was used that had the same spectral envelope as the average of all signals employed. To create this noise, for all stimuli the local amplitude maxima were calculated across frequency (within 200 Hz-wide windows around harmonic frequencies $n \cdot 261.6\ Hz$), for each harmonic partial number the maximum of the local amplitude maxima was calculated across stimuli, and the partial amplitudes of this maximum spectrum were linearly interpolated and used as a filter for white noise. In condition *Transfer*, the signal in the first interval of a trial was presented in silence, whereas the signals in the second and third interval were presented in pink noise.

The mean signal level was 57.5 dB SPL + H_f in conditions *Quiet*, *Pink*, *Shaped* and *Transfer*, and 50 dB SPL + H_f in conditions *Quiet-Low* and *Pink-Low*. For the hearing-impaired subjects, the spectral shape and sound level of signal and noise (as described above for normal-hearing listeners) was amplified by half of their individual frequency-dependent hearing loss (+H_f). For all subjects, the signal level of each

Table 4.1: Noise and levels used in the 5 noise/level conditions of Experiment B (2nd + 3rd column) and results obtained by ANOVA and subsequent multiple comparison test with factors "subject group" (4th column) and "noise/level condition" (5th column). Condition: condition name used in text; background: presence and spectral envelope of noise; SNR: signal-to-noise ratio. The signal level in all conditions was roved by 2.5 dB and amplified for hearing-impaired listeners by a half gain (H_f). Subject groups: NH (normal-hearing), fHI (flat hearing loss), sHI (steep hearing loss). Instrument continua: H-T (horn-trombone), C-S (cello-sax). 4th column indicates which subject groups have significantly higher (>) JNDs than other subject groups in the respective condition. 5th column indicates whether JNDs in the respective condition are significantly higher than in another condition. See Section 4.2.4 for more detail of tests' procedure and results.

condition	**background (SNR =+10dB)**	**signal level [dB SPL]**	**significant JND differences** ($p<0.05$) across **subject groups**	**conditions**
Quiet	none	$57.5\pm2.5+H_f$	sHI > NH, fHI	
Quiet-Low	none	$50.0\pm2.5+H_f$	sHI > fHI $\overset{*}{>}$ NH (*p=0.06 for fHI,H-T)	> *Quiet* for fHI in H-T
Pink	pink noise	$57.5\pm2.5+H_f$	sHI > NH, fHI	> *Quiet* (p=0.06 for fHI in C-S)
Pink-Low	pink noise	$50.0\pm2.5+H_f$	sHI > NH, fHI	> *Pink* for sHI in C-S
Shaped	noise with signal's spectr. envelope	$57.5\pm2.5+H_f$	sHI > NH, fHI	> *Quiet*, but none vs. *Pink*
Transfer	1^{st} interval: none 2^{nd},3^{rd}: pink noise	$57.5\pm2.5+H_f$	none	> *Pink* for NH in C-S

presentation was randomly roved by ±2.5 dB in order to avoid an identification of the test signal by means of loudness. (Pilot experiments showed that loudness matching of the signals used, which have equal root-mean-square (rms) levels, resulted in up to 5 dB inter-individual variance across subjects.)

All conditions (Table 4.1) were measured in both instrument continua (horn-trombone and cello-sax), totaling 12 different measurements. The order of the measurements was permuted randomly for each subject and each session. The subjects' task was to indicate whether the second or third presented signal differed in timbre from the first signal. They were asked to actively ignore the noise in the background if present, to ignore loudness differences and to base the decision only on the timbre. They were instructed that this timbre could be any "quality" of sound, for example brightness or roughness, and that they should decide without knowing what the difference was. It was explained that a repeated presentation of the same trial was not possible and that they must guess if all sounds sounded equal.

Feedback[14] about the correctness of the response was only provided at the beginning of each track until the third reversal of the adapting parameter occurred. No feedback was provided during the actual data collection phase following the 7th reversal in each track. During the initial 6 reversals of each track, the tracking variable $\Delta\alpha$ (Equation 4.1) was adapted by multiplying/dividing by 1.45 (up to the 2nd reversal), 1.35 (up to the 4th reversal) and 1.25 (afterwards). By a 1-up-3-down adaptive tracking procedure, the value of $\Delta\alpha$, at which the test stimulus was chosen correctly with 79.4% probability (Levitt, 1970), was determined by averaging the tracking variables at the 7th to 12th reversal.

4.2.3 Subjects

Experiment A

34 subjects (20 normal-hearing listeners, 14 hearing-impaired listeners) aged between 24 and 76 years fulfilled one session of less than 60 minutes including a break. Most of the subjects (16 normal-hearing listeners, 12 hearing-impaired listeners) had experience in psychophysical experiments from different studies. The first three runs of the session were training measurements.

Experiment B

19 subjects aged between 23 and 68 years started the measurement series. Of the initially selected 19 subjects only 16 (6 normal-hearing listeners, 10 hearing-impaired listeners) were able to perform the whole measurement set and were therefore included in this study. Each of the 16 subjects fulfilled 12 to 17 sessions of 75 minutes and 9 to 12 measurements each. After every three measurement tracks, that is after every 15-20 minutes, a break was taken. For each subject, the first two sessions as well as the first measurement of each session were training sessions. Hearing-impaired subjects were requested to perform 9 non-training repetitions of each measurement condition, and normal-hearing subjects performed 6. In order to balance out any difference in training across normal-hearing and hearing-impaired subjects, the normal-hearing subjects did 4 additional repetitions of similar measurements (not shown here).

The subjects were interviewed for their musical background. 14 subjects reported having no experience playing musical instruments, or had musical practice in the past but had not actively practised music for at least 4 years prior to the experiment. Only one hearing-impaired and one normal-hearing subject were amateur musicians,

had more than 4 years of regular experience learning and practising an instrument, and were still actively practising music at the time of the experiment.

Subjects were tested for their pure-tone hearing threshold, in which all normal-hearing listeners showed thresholds $\leq$ 10 dB HL for frequencies up to 4 kHz and $\leq$ 15 dB HL at 6-8 kHz. The hearing-impaired subjects were divided into two groups according to the configuration of their audiometric hearing loss. 5 subjects showed a "flat" or "diagonal" configuration and hearing thresholds of 45-80 dB HL for frequencies between 1 and 8 kHz. The other 5 subjects showed a "steep" configuration, exhibiting a slope of more than 20 dB/octave for at least one octave between 0.5 and 4 kHz and hearing thresholds of 60 dB HL or more for frequencies above 4 kHz. All hearing-impaired subjects performed additional tests in categorical loudness scaling and showed recruitment at frequencies with hearing loss. Speech intelligibility was tested for subjects with flat hearing loss using the "Oldenburger Satztest" (Wagener et al., 1999c,a,b) in noise with a linear half-gain amplification. The subjects showed a mean increase in speech reception threshold (SRT) of 3.1 dB in stationary background noise.

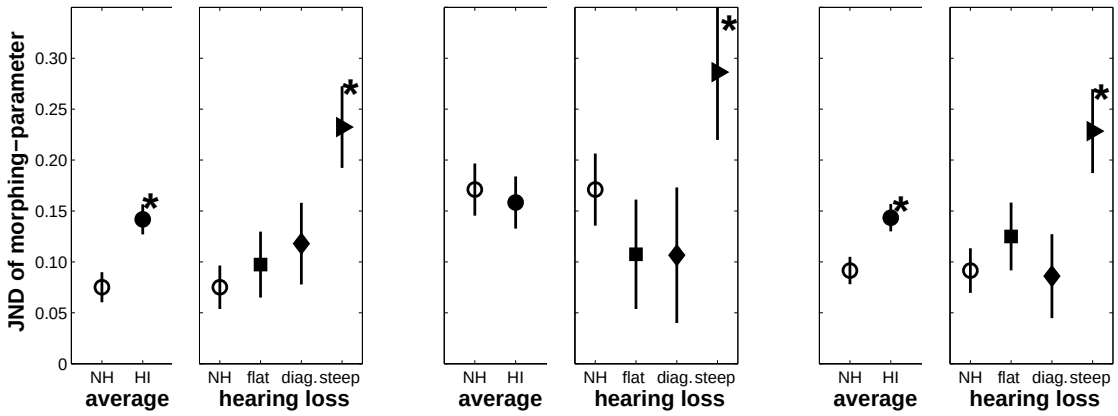
4.2.4 Experimental results

Experiment A

The average timbre JND values are plotted in Figure 4.1. The Wilcoxon rank sum test showed that the medians of the normal-hearing listeners, in both the horn-trombone and flute-trumpet continua, were significantly lower than those of hearing-impaired listeners ($p<0.05$). In the cello-sax continuum, the test did not show a significant difference.

The hearing-impaired subjects were then divided into three groups according to the configuration of their audiometric hearing loss: there were 6 subjects with "flat", 4 with "diagonal" and 4 with "steep" (that is, exhibiting a slope of more than 20 dB/octave for at least one octave between 0.5 and 4 kHz) hearing loss. The respective results according to this division are given in Figure 4.1.

The Wilcoxon rank sum test yielded that, in all instrument continua, the medians of hearing-impaired listeners with steep hearing loss were significantly higher than those of hearing-impaired listeners with flat and diagonal hearing loss, as well as those of normal-hearing listeners ($p<0.05$). In the cello-sax continuum, the JND medians of the flat and diagonal hearing-impaired listeners are slightly, but not significantly smaller than of the normal-hearing listeners.



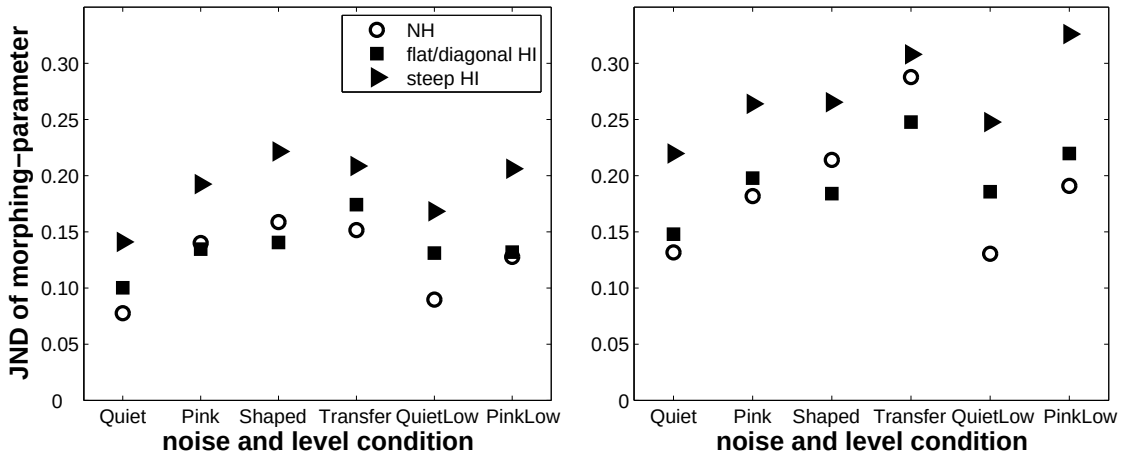
(a) horn-trombone (b) cello-sax (c) flute-trumpet

Figure 4.1: Medians of timbre JNDs (expressed as JND of the morphing-parameter $\Delta\alpha$) with 95% confidence intervals in the French horn-trombone (a), cello-sax (b) and flute-trumpet (c) continua. The left figures compare the average results of normal-hearing listeners (NH; n=20, open symbols) and hearing-impaired listeners (HI; n=14, filled symbols) for each continuum. In the right figures, hearing-impaired subjects are further divided into flat (n=6, squares), diagonal (n=4, diamonds) and steep (n=4, triangles) hearing loss. The morphing-parameter $\Delta\alpha$ represents the JND (70.7%) of timbre as a ratio of the original sounds in the respective continuum. Stars indicate JND medians with significant deviation from other subject groups in the respective comparison.

Experiment B

The average results of the timbre JND values are plotted in Figure 4.2, separated into the two instrument continua "horn-trombone" (4.2(a)) and "cello-sax" (4.2(b)). The abscissae show the 6 noise/level conditions (see Table 4.1), and in each condition three symbols are shown according to the three subject groups normal-hearing listeners (open circles), flat/diagonal hearing-impaired listeners (filled squares) and steep hearing-impaired listeners (filled triangles). Hence, each symbol shows the JND mean of all subjects (6 for the normal-hearing, 5 for flat hearing-impaired, and 5 for the steep hearing-impaired group) and repetitions (6 for the normal-hearing and 9 for the hearing-impaired groups) within one group.

To detect any significant effects of hearing loss or of the conditions (i.e. effects of level, noise presence, noise shape and silence-to-noise transfer), an analysis of variance (ANOVA) was performed for each instrument continuum with the factors "noise/level condition" and "subject". The (unbalanced) 2-way ANOVA showed that both factors had significant ($p<0.05$) main and interaction effects on the results, in both instrument continua. Further, a multiple comparison test[15] was applied in



(a) horn-trombone (b) cello-sax

Figure 4.2: Results of Experiment B, timbre JND measurements in different noise/level conditions, for the instrument continua (a) horn-trombone and (b) cello-sax. The abscissae show the 6 different noise/level conditions (see Table 4.1 and text for detail) and the ordinates show the means of timbre JND (expressed as JND of the morphing-parameter $\Delta\alpha$). Different symbols represent the JND means of the three listener groups of normal-hearing listeners (open circles), flat/diagonal hearing-impaired listeners (filled squares) and steep hearing-impaired listeners (filled triangles). For clearer visibility, standard deviations are not shown; significantly different conditions and subject groups are listed in Table 4.1. The morphing-parameter $\Delta\alpha$ represents the JND (79.4%) of timbre as ratio of the original sounds in the respective instrument continuum.

order to verify *which* pairs of subject groups have significantly different JNDs (at the 0.05-level) than other subject groups in a respective condition. Accordingly, within each subject group a multiple comparison test verified, whether JNDs in one condition are significantly different than in another condition for the respective subject group. The following significant outcomes of the ANOVA and the multiple comparison tests are also listed in Table 4.1 (p.53).

Effect of hearing loss: Even though both the intra- and inter-individual variation was quite large within each group of subjects (normal-hearing, flat/diagonal hearing-impaired, steep hearing-impaired), significant differences between the listener groups were observed. In both instrument continua and in all noise/level conditions, the JND medians of the group of steep hearing-impaired listeners were higher than those of the normal-hearing listeners and of the group of flat/diagonal hearing-impaired listeners (Figure 4.2). In both instrument continua these differences were significant for all noise/level conditions with the exception of condition *Transfer*. In condition *Transfer*, a JND comparison across listener groups did not produce any

significance. Significant differences between flat/diagonal hearing-impaired listeners and normal-hearing listeners were only observed in condition *Quiet-Low* of the cello-sax continuum, where the JND of flat/diagonal hearing-impaired listeners were significantly higher than those of the normal-hearing listeners; in the same condition in the horn-trombone continuum, the difference was nearly significant (p=0.06).

Effect of levels: Comparison between the silence conditions *Quiet* (at 57.5 dB+H_f) and *Quiet-Low* (at 50 dB+H_f) showed that a presentation level reduction of 7.5 dB increases the JNDs for both hearing-impaired groups in both instrument continua (Figure 4.2). This JND difference was significant for the flat/diagonal hearing-impaired listeners in the horn-trombone continuum. For the normal-hearing listeners a reduction of level caused very little change in JND. As a consequence, flat/diagonal hearing-impaired listeners showed almost significantly higher JNDs than normal-hearing listeners in condition *Quiet-Low*, although the JNDs were similar in all other conditions, as mentioned above.

In pink noise (condition *Pink* at 57.5 dB+H_f and condition *Pink-Low* at 50 dB+H_f, Figure 4.2), a reduction in level only caused a significant JND increase for the steep hearing-impaired listeners in the cello-sax continuum.

Effect of presence of noise: For all subject groups and in both instrument continua, JNDs increased when pink noise (*Pink* compared to *Quiet*) or noise with the signals' spectral envelope (*Shaped* compared to *Quiet*) was added to the signals, that is to say, to *all* presentation intervals (Figure 4.2). In all cases but one, this difference was significant ($p<0.05$). For the flat/diagonal hearing-impaired listeners in the cello-sax continuum, when pink noise was added, p=0.06 due to the large confidence intervals.

Effect of noise shape: When comparing JNDs between the different background noises used in the experiment (*Pink* compared to *Shaped*, Figure 4.2), no significant difference was found. However, normal-hearing listeners showed slightly higher JNDs when the instrument stimuli were embedded in noise with a spectral envelope of the tonal stimuli than when they were embedded in pink noise (see Section 4.2.2 for noise descriptions).

Effect of transfer: Condition *Transfer* – in comparison to condition *Pink* – tested the ability to imagine how the stimulus heard in silence would sound in noise. In both conditions pink noise was used as background. While in condition *Pink* the noise was added to all intervals, in condition *Transfer* the reference interval was played in silence. In all instrument continua and listener groups, JNDs were higher in the transfer situation (*Transfer* compared to *Pink* in Figure 4.2), even though significance was only found for normal-hearing listeners in the cello-sax continuum.

Since the effect of noise, or the transfer from silence to noise, is of interest to the main problems of hearing-impaired listeners (as in the cocktail party effect), the individual relative changes from the condition *Pink* to the *Transfer* condition will be considered in more detail.

All JND results of condition *Transfer* were normalized by the JNDs of condition *Pink*, that is, for each subject the JNDs of *Transfer* were divided by the JNDs of condition *Pink* in the respective instrument continuum in the respective session (or in one of the adjacent sessions, if condition *Pink* was not measured in the session). In both continua, the mean JND ratio was higher for normal-hearing listeners than for hearing-impaired listeners. A variance analysis, which was applied to the individual JND ratios, showed that in the cello-sax continuum this difference was significant ($p<0.05$).

4.3 Discussion

The main outcomes of the measurements can be summarized as follows:

1. Timbre JNDs of subjects with a **steep hearing loss** are significantly higher than those of normal-hearing listeners, both in silence and in noise.

2. Timbre JNDs of **flat/diagonal hearing-impaired** listeners are for stimulus levels above 55 dB + H_f (i.e., with appropriate amplification for hearing-impaired listeners) similar to JNDs of normal-hearing listeners. Slight JND differences vary across instrument continua and can also be negative.

3. Reducing the mean **level** affects timbre JNDs of normal-hearing listeners only slightly. It increases JNDs of flat hearing-impaired listeners in silence, and increases JNDs of steep hearing-impaired listeners in all noise/level conditions.

4. Timbre JNDs of all subject groups are higher **in noise** than in silence; in particular for normal-hearing listeners this is significant. (No significant JND difference between the different noise spectra were found.)

5. In condition ***Transfer***, which tested the ability to imagine how the stimulus heard in silence would sound in noise, no significant JND differences across listener groups were found and JNDs are higher than in condition *Pink*.

6. **JND increases from noise to the transfer** condition less for hearing-impaired listeners than for normal-hearing listeners.

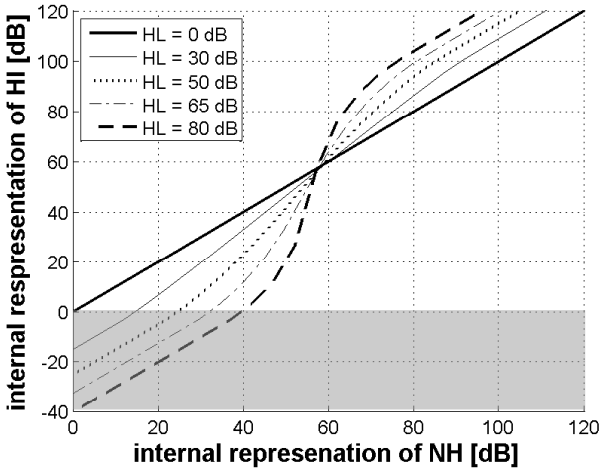


Figure 4.3: Input-output function of linear amplification (by 50% HL), linear attenuation (by 20% HL), and expansion (by 80% HL) for different assumed hearing losses (HL). The function outlines the hypothetical relation of perceived (partial) intensity in subjects with flat hearing loss (with 80%/20% due to OHC/IHC loss) to the corresponding perceived intensity in normal-hearing listeners for the present study.

4.3.1 Compression loss, attenuation and amplification

One explanation of the observations reported here can be pursued by primary factors of the hearing impairment (i.e. compression loss and loss in sensitivity, or loss of OHC and IHC, respectively) and by the fact that the (linear!) half-gain amplification only equalized overall loudness in hearing-impaired and normal-hearing listeners.[16] Since the compression loss component in sensorineural hearing-impaired listeners leads to a distorted mapping between the presented and the "internal" level of the stimulus components (Figure 4.3), two effects can be observed:

a) For flat hearing-impaired listeners[17], a half-gain amplification may lift many sound parts into the hearing range and even over-amplify sound components that are above the average sound level, making subtle intensity differences more audible (outcome 2, p. 59). This may be the reason for the slightly lower JNDs of hearing-impaired listeners with flat/diagonal hearing loss than of normal-hearing listeners in the cello-sax continuum in Experiment A.

b) Reducing the sound level by 7.5 dB in silence (i.e. from condition *Quiet* to *Quiet-Low*) represents a higher loudness reduction for hearing-impaired listeners than for normal-hearing listeners (Figure 4.3). Reducing the level may cause more sound parts to fall below hearing threshold for hearing-impaired listeners than for normal-hearing listeners. In particular for flat hearing-impaired listeners, more crucial sound parts that were still available at 57 dB + H_f may become inaudible. This decreased audibility might result in an increase in JND (outcome 3, p. 59).

4.3.2 Steep hearing loss

In contrast to the flat hearing-impaired listeners, for steep hearing-impaired listeners [18] low-level high-frequency partials, which are crucial for timbre differences[19], may fall below their hearing threshold even after amplification using the half-gain rule (outcome 1, p. 59).

An additional reason that causes the same effect (i.e. outcome 1 in contrast to outcome 2, p. 59), may be dead regions and distortion at and above the steep flank in steep hearing-impaired listeners. Cochlear dead regions, which are likely to be present with a hearing loss above 80 dB, reduce intensity resolution and distort the timbre in steep hearing-impaired listeners due to off-frequency listening (Moore et al., 2000).[20]

4.3.3 Masking effects

Outcome 4 (p. 59) can be explained by the masking effect of the background noise, because the noise and signal spectra overlap. Hence, noise with a level that is only 10 dB lower than the signal may mask crucial harmonics that are in the *Quiet* condition used for discrimination.[21] This highlights the strong role of low-level components for timbre discrimination that are already masked at the favourable SNR (+10 dB) employed here.

4.3.4 Frequency selectivity and temporal resolution

Two important factors involved in sensorineural hearing loss are the reduced frequency selectivity and poorer temporal resolution, for which basilar membrane compression loss as the primary factor can account (Oxenham & Bacon, 2003, and Appendix B).[22] A common hypothesis (Moore, 2003) indicates that the reduced frequency selectivity in hearing-impaired listeners leads to a reduced ability in distinguishing different timbres. Outcome 2 (p. 59) contradicts this hypothesis. Frequency selectivity as well as temporal resolution do not seem to play an important role in distinguishing the timbres of the present study.

4.3.5 Object separation

The large number of possible timbre dimensions (McAdams et al., 1995; Terasawa et al., 2005) and the uncommon task to use timbre to discriminate objects[23], leads to a large inter-individual scatter of the experimental results. But on the other hand,

the stimuli used still sound natural (in contrast to pure tones or complex tones) and the lack of categories is an advantage for using timbre to measure subjects' bottom-up abilities in separating objects without learned categories.

In most "natural" cases, object segregation (i.e. separation of objects at similar levels) is observed as being slightly problematic for normal-hearing listeners and more difficult for hearing-impaired listeners, which speech measurements in noise and quiet for example show (Pekkarinen et al., 1990; Beattie et al., 1997). This suggests object invariance for normal-hearing listeners and object non-linearity for hearing-impaired listeners. A common hypothesis postulates that auditory processing is effectively linear for normal-hearing listeners and that non-linearity (e.g. due to additional cochlear distortion) causes worse object separation in hearing-impaired listeners (Kollmeier & Derleth, 2001). Outcome 4 (p. 59) conflicts with the hypothesis that an acoustical object is invariant for normal-hearing listeners when adding a second independent object; normal-hearing listeners cannot fulfill optimal segregation of the tonal stimuli from background noise, and hence, distinguishability of stimuli in noise is significantly worse than in silence. Outcome 6 (p. 59) indirectly contradicts the hypothesis that hearing-impaired listeners have specifically more problems in object segregation than normal-hearing listeners with respect to recognition and transfer.

Chapter 3 showed that the horn-trombone continuum is almost exclusively varied by the timbre dimension "brightness", that is to say, changes in spectral shape of the tonal content caused by intensity changes of the harmonics. The cello-sax continuum is varied by spectral shape and a spectro-*temporal* dimension such as spectral flux, that is, temporal variation of the spectral shape. If temporal resolution were decisive for distinguishing the spectral flux or the spectral resolution for distinguishing the the spectral shape, then hearing-impaired subjects with a flat hearing loss should show higher timbre JNDs. If temporal or spectral resolution played a crucial role in segregating the tonal stimuli from the background noise, then hearing-impaired subjects with a flat hearing loss should show higher timbre JNDs in the noise conditions and a stronger JND increase from condition *Pink* to *Transfer*. Since neither is the case, intensity discrimination seems to play the major role in discriminating the stimuli of the present study. Since intensity resolution seems not to be distinctly degraded by compression loss, hearing-impaired listeners do not show more problems in discriminating objects using intensity differences.[24]

4.4 Summary

1. The present study showed that timbre JNDs of subjects with a steep hearing loss are significantly higher than those of normal-hearing listeners, both in silence and noise, whereas timbre JNDs of flat/diagonal hearing-impaired listeners are, for signal levels above 55 dB + H_f, similar to JNDs of normal-hearing listeners if appropriate amplification is used for hearing-impaired listeners. This contradicts Moore's (2003) hypothesis that hearing-impaired listeners have problems in distinguishing timbre (due to reduced frequency selectivity) for the tested instruments. All results could be explained by primary factors involved in sensorineural hearing loss, such as attenuation and loss of compression for flat/diagonal hearing-impaired listeners and additional distortion and loss of intensity resolution for steep hearing-impaired listeners.

2. While compression loss can lead to a reduced time and frequency resolution and thus might reduce the ability of hearing-impaired listeners to separate objects by attributes connected to time and/or frequency, for example onset, pitch and location, the non-degraded intensity resolution in hearing-impaired listeners seems to dominate over poor frequency and time resolution for certain timbre discriminations near threshold, for example spectral shape or spectral flux.

3. In noise, timbre JNDs of all subject groups are significantly higher than in silence, and JND increases for normal-hearing listeners from silence to noise more than for hearing-impaired listeners. In the condition *Transfer*, which tested the ability to imagine how the stimulus heard in silence would sound in noise, no significant JND differences across listener groups are found and JNDs are higher than in condition *Pink*. This contradicts the hypothesis that hearing-impaired listeners generally have more problems in object segregation than normal-hearing listeners, if an appropriate amplification is provided.

Chapter 5

Modeling timbre discrimination of the normal and impaired auditory systems

Abstract

In an attempt to predict subjective timbre similarity ratings and discrimination thresholds, psychoacoustic measurements were simulated using a modified version of the effective Perception Model PeMo *for the normal and impaired hearing system (Dau et al., 1996; Derleth et al., 2001). The model preprocesses signals accounting for the peripheral and cortical auditory processing, correlates the preprocessing output of different signals and calculates an "internal stimulus distance". Crucial model parameters are tested including the number of modulation filters, the correlation of the preprocessing output and the perceptual weighting across time, frequency and modulation dimension. Since the multidimensional timbre variation of stimuli in the present study assumes attentional processes and memory limits involved in the measurements, different model settings account for different stimulus and noise conditions. Specifically, the results of the timbre measurements of the previous chapters could be predicted for spectral-dominated timbre differences by averaging the preprocessing output over time, while time-step-wise comparison was required for temporal differences. Moreover, the effect of hearing loss could be predicted by processing stages accounting for attenuation and reduced compression. However, the difference between flat and steep hearing impairment could only be predicted at low levels. While* PeMo *is limited to bottom-up processes and cannot simultaneously account for multidimensional timbre variation, it still provides predictive power for unidimensional tasks within the perceived timbre differences. Further modeling efforts are required to clarify the weighting in the internal-representation space, to substitute the gammatone filter bank with a more physiological filter bank, or to introduce a model stage that accounts for the across-channel processing which is crucial for timbre perception.*

5.1 Introduction

All the sounds that we hear are the result of processing evoked by external acoustical signals on the way from the outer ear to higher cortical regions. The morphology of the peripheral auditory system is well known, but the hearing sense with its detailed peripheral signal processing is not yet fully understood. In particular, the mechanism that leads to the performance of object separation in normal-hearing listeners in contrast to sensorineural hearing-impaired listeners (see Appendix B) is not satisfactorily understood. Analytical and numerical models are constructed in order to find the mechanisms of perception. Since the auditory system is complex and nonlinear, computer models have been used to simulate the processing numerically. By substituting the physiologically or psychoacoustically verified processing stages with signal processing modules, a hypothetical internal representation of the digitalized external sound can be calculated. These internal representations should account for perceptual distances found in psychoacoustic measurements, such as just noticeable differences (JNDs) and similarity ratings. An optimal model would simultaneously account for perceptual distances in all distinguishable sound parameters like, for example, pure tone intensity or frequency, sound location or timbre. A model that was validated for most "basic" psychoacoustic functions is the Perception Model (*PeMo*) from Dau et al. (1996), which shall be used in this study to simulate the timbre measurements of the previous chapters and predict similarity ratings and JND results. To what extent can the performance found in Chapters 2, 3 and 4 be derived from simple model assumptions? In other words, can the auditory preprocessing steps considered here account for the experimental findings (i.e., rating and threshold differences depending on stimuli and hearing loss), even if higher cognitive processes are neglected?

5.1.1 *PeMo* preprocessing

In the effective and physiologically-motivated Perception Model *PeMo*, the processing steps (Figure 5.1) are attributed to physiological stages. A pre-filtering accounts for the spectral shaping by pinna, ear channel and middle ear. A filter bank represents the frequency-place transformation on the basilar membrane. The signal is processed further as parallel and separate time signals in frequency-overlapping frequency channels. Each channel's signal is subsequently half-wave rectified and low-pass filtered with a cutoff frequency of approximately 1 kHz. This represents the transformation of the mechanical oscillations of the basilar membrane into electrical potentials in the inner hair cells; for high carrier frequencies, the temporal

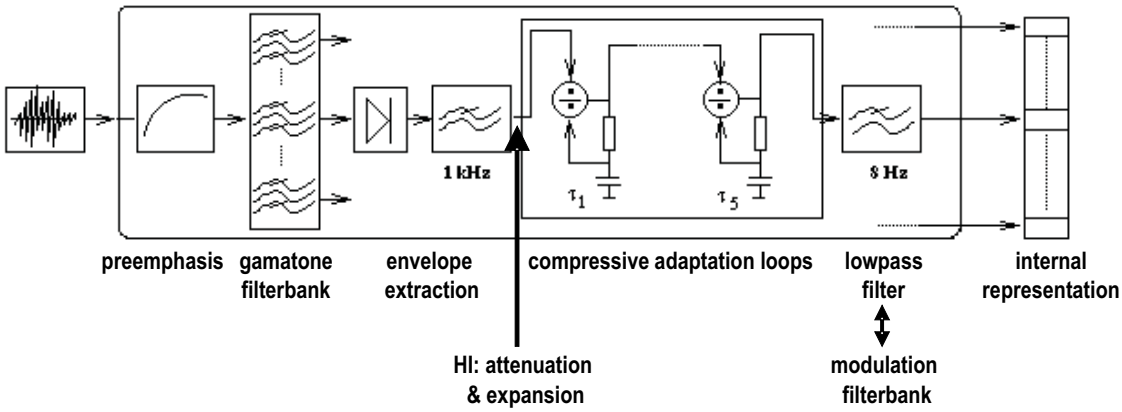


Figure 5.1: Exemplary block diagram of the acoustic signal processing in the Oldenburg Perception Model *PeMo*. See text for description.

envelope is preserved. For the impaired hearing system an instantaneous attenuation and expansion, which represent the inner and outer hair cell loss, respectively, are included (Derleth et al., 2001).[25] After the hair cell stage, five compressive adaptation loops account for peripheral and central temporal adaptation as well as for dynamic compression. In the following stage, the signals are low-pass filtered with a cutoff frequency of almost 8 Hz accounting for effects of temporal integration. Instead of this low-pass filter (Dau et al., 1996), a linear modulation filter bank may be used to further analyze the amplitude changes of the envelope (Dau et al., 1997a). The output of the preprocessing stages is a 2-dimensional (only using the low-pass filter) or 3-dimensional (with modulation filter bank) time-varying activity pattern, which is here called "internal representation" (IR)[26]. (See Appendix C for illustration of IR differences.)

5.1.2 *Optimal detector* and IR distances

Subsequent to the preprocessing, the internal representations of signals are compared, for which three different methods are published. The original version of the *PeMo* uses the ***optimal detector***, which combines all filter outputs linearly and uses an "optimal" weighting of the channels according to a template (Dau et al., 1996). Using this *optimal detector* to simulate detection experiments, the IR increment that varies between presented intervals is weighted with a normalized super-threshold signal increment called "template"[27]. That is, the *difference* of test interval and reference interval is weighted with the template as follows:

$$\mu = \frac{1}{\sqrt{T{\cdot}F{\cdot}M}} \cdot \sum_{t,f,m} (IR_{S+R}(t,f,m) - IR_R(t,f,m)) \cdot (IR_{TP+R}(t,f,m) - IR_R(t,f,m)) \quad (5.1)$$

where t, f and m are time step, frequency channel number and modulation channel number, respectively; T, F and M are the total numbers of time steps, frequency channels and modulation channels; S, R and TP refer to the IR of the test signal, reference signal and template signal, respectively. This is commonly used in detection experiments, such as pure tone detection in noise, where the test signal is a pure tone, the reference signal is noise and the template is a pure tone with amplitude above threshold. The *optimal detector* is restricted to conditions under which all channels can assumed to be independent observations (Dau et al., 1997b). If the template in Equation 5.1 is substituted by the test signal, μ will be related to the **Euclidean distance**:

$$IRdist = \sqrt{\frac{1}{T \cdot F \cdot M} \cdot \sum_{t,f,m} (IR_S(t,f,m) - IR_R(t,f,m))^2} \tag{5.2}$$

This is commonly used in *PeMo* variants for automatic speech recognition (Holube & Kollmeier, 1996). Instead of correlating the IR *difference*, the *PeMo-Q* correlates the entire IRs by a **cross-correlation** using the Pearson product (Huber & Kollmeier, 2006):

$$IRcorr = \frac{\sum_{t,f,m} (IR_S(t,f,m) \cdot IR_R(t,f,m))}{\sqrt{(\sum_{t,f,m} IR_S^2(t,f,m)) \cdot (\sum_{t,f,m} IR_R^2(t,f,m))}} \tag{5.3}$$

The *PeMo-Q* is mainly used in objective sound quality assessment.

5.1.3 *PeMo* for modeling timbre rating and discrimination

As mentioned above, *PeMo* (Dau et al., 1996) was designed primarily for detection experiments. Some implementations are based on presumptions that are fulfilled in measurements in which basic sound attributes should be detected, but not necessarily in timbre comparison. These restrictions concern, for example, the weighting and interaction of frequency bands. In a pure-tone detection task primarily one frequency channel contains the discrimination cue, whereas for timbre perception the interaction and weighting of different frequencies and time steps are important factors.

A related problem is the ability to memorize and compare independently perceived attributes[28]. If, in the most extreme case, two presentations of random noise are compared, an optimal detector would be able to combine the information of 30 frequency channels, 6 modulation channels and 50 time steps - a total of 9000 independent elements - whereas a human does not have the capacity to memorize

and compare this many independent observations (Goossens et al., 2006). Since timbre is a multidimensional sound attribute, memory limits may also prove important factors in predicting timbre discrimination thresholds.

For speech intelligibility prediction, other presumptions lead to model adaptations such as the Euclidean distance (Section 5.1.2) and a "dynamic time warp", which optimally stretches the temporal dimension of internal representations (Holube & Kollmeier, 1996). Some timbre dimensions are the result of a *simultaneous* comparison process in one stimulus (e.g. of frequency distribution or overtone synchronicity) and only thereafter is a successive comparison made (Green, 1983). This assumes that, as for speech intelligibility prediction, time-step-wise comparison of IRs may mislead and give false information for some timbre dimensions, while for other dimensions (such as spectral flux; Chapter 2) a temporal comparison may be crucial.

Of the three *PeMo* variants, *PeMo-Q*, which deals with sound quality assessment (Huber & Kollmeier, 2006), may be most closely related to timbre comparison. However, *PeMo-Q* is not designed for discrimination experiments at threshold, but rather for evaluating sound distances above threshold. Hence, the following problems and questions arise when adapting *PeMo* to timbre rating and discrimination, and shall be approached with this study:

1. **What is the optimal weighting of IR observations?** Which weighting of IR channels and IR regions can predict discrimination or similarity judgements best, if all dimensions (time, frequency, modulation frequency) vary simultaneously in the stimuli? Some dimensions carry crucial information, while others may contain distinct differences that are not perceived distinctly or that are perceived but cannot be used as distinction cues. It is unclear whether spectral, temporal and spectro-temporal differences are perceived equally or whether one dimension dominates the percept. Certain time periods, frequency regions or modulation frequency regions may dominate the percept; for example differences at edges may be perceived more easily.

2. **What is the optimal measure of IR distance?** The "internal distance" of timbres, measured for example as cross-correlation or with Euclidean distance, may be different for timbre differences above threshold than for those at threshold.

3. **Can the results of the timbre measurements in the previous chapters be predicted with the *PeMo*?** It is unclear whether one set of model parameters can simultaneously cope with the effects of timbre dimensions,

reference stimuli, background noise, signal level, and the subject's hearing impairment, on both similarity ratings of timbre differences above threshold and discrimination thresholds in normal-hearing and hearing-impaired subjects.

5.2 Simulation and results

In the previous chapters timbre perception in normal-hearing and hearing-impaired subjects was studied with psychoacoustic measurements. Similarity ratings verified the perceptual distances of instrumental sounds along and across different instrument continua (Chapter 2). Discrimination measurements verified the just noticeable timbre differences with respect to different reference stimuli (Chapter 3), level and background noise conditions (Chapter 4). The difference between hearing-impaired and normal-hearing subjects was measured both for large timbre differences (Chapter 2) and at discrimination threshold (Chapters 4). In the previous chapters, the results were discussed in the context of common timbre models that describe the cues probably used by the subjects to perform the judgements on their timbre percepts. In the present study, however, these psychoacoustic measurements are simulated by a numerical model that selects the cues automatically, i.e. without relying on the absolute timbre percept. Simulations aim to objectively predict effects of stimulus variation, background noise and hearing loss on similarity ratings and discrimination thresholds.

In the following, the stimuli of these measurements are preprocessed using the Perception Model introduced above (Figure 5.1). For the preprocessing, the standard gammatone filter bank was used with ERB-wide filters at center frequencies from 170 Hz to 15 kHz, which are 2/3 of the fundamental frequency and 1.5 times the cut-off frequency of the stimuli, respectively.[29] With regard to the further processing, 10 model versions with different parameter settings are compared in the present study (Table 5.1). Model version M1 contains standard settings of *PeMo-Q* without modulation filter bank, in which the internal representations of the presented stimuli are compared by a cross-correlation (Equation 5.3). M2 uses 6 modulation channels with modulation-filter center frequencies of up to 46.3 Hz. This limit was chosen for all audio-frequency channels, because the lowest gammatone filter, centered at 170 Hz, does not allow for higher modulation frequencies. As mentioned above, time-step-wise comparison of IRs may give false information, therefore in M3 and M4 the temporal dimension (of IRs in M1 and M2, respectively) is removed before cross-correlating the IRs. Since Iverson & Krumhansl (1993) showed that the crucial information for timbre similarity ratings is contained in the spectral information of

model version	number of mod. filters	temporal reduction of preproc. output	resulting IR dimensions	distance measure
M1	0	–	2-dim IR (t,f)	cross-corr.
M2	6	–	3-dim IR (t,f,mod)	cross-corr.
M3	0	temporal mean	1-dim IR (f)	cross-corr.
M4	6	temporal mean	2-dim IR (f,mod)	cross-corr.
M5	0	temp. attack mean	1-dim IR (f)	cross-corr.
M6	0	–	2-dim IR (t,f)	Euclidean
M7	6	–	3-dim IR (t,f,mod)	Euclidean
M8	0	temporal mean	1-dim IR (f)	Euclidean
M9	6	temporal mean	2-dim IR (f,mod)	Euclidean
M10	0	temp. attack mean	1-dim IR (f)	Euclidean
M8*	0	temporal mean	1-dim IR (f)	p-weighted

Table 5.1: Model versions with parameter settings that are varied in the present study. The 1st column shows the model version number. The 2nd column shows the number of modulation filters used in the preprocessing. The 3rd column displays the model versions in which the internal representations (IR) are time-averaged using the entire stimulus (temporal mean) or only the first 300ms of the IR (temp. attack mean). The 4th column shows which dimensions of time (t), frequency (f) and modulation frequency (mod) channels are correlated by the distance measures shown in the 5th column (cross-corr.: Equation 5.3, Euclidean: Equation 5.2, p-weighted: Equation 5.4.)

the stimuli's attack segment, in model version M5 the IR (of version M1) is reduced to one time step ("time-averaged") using only the first 300ms. M6-M10 are identical to M1-M5 except that the Euclidean distance (Equation 5.2) is used instead of the cross-correlation as the IR distance measure. (The distance measure of the *optimal detector* is not used in the present study, because it uses a super-threshold signal, while timbre ratings and thresholds were shown to depend on the distance above threshold (i.e. morphing-parameter α; see Chapters 2 and 3).)

All measurements were simulated using the model as the "subject" in automatic alternative-forced-choice measurements. The thus predicted measurement results were compared to the subjective results of the previous chapters.

5.2.1 Predicting similarity rating

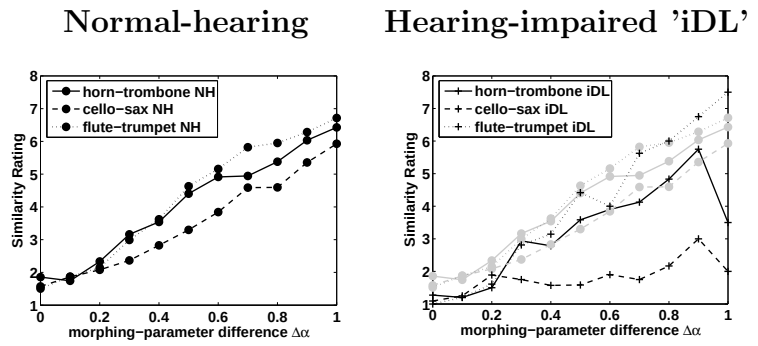
Using the model versions M1-M10, IR distances between the stimuli of the similarity rating measurements in Chapter 2 were calculated. The IR distances were subse-

model number:	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**	**M10**
horn-trombone	-0.75	-0.77	-0.74	-0.75	-0.74	0.78	0.77	**0.78**	0.77	**0.78**
cello-sax	-0.69	-0.71	-0.72	-0.72	-0.72	0.66	0.68	**0.72**	0.68	**0.72**
flute-trumpet	-0.76	-0.77	-0.75	-0.76	-0.75	0.78	0.78	**0.79**	0.77	**0.79**
all continua	-0.66	-0.63	-0.60	-0.58	-0.70	**0.71**	0.68	0.57	0.55	**0.77**

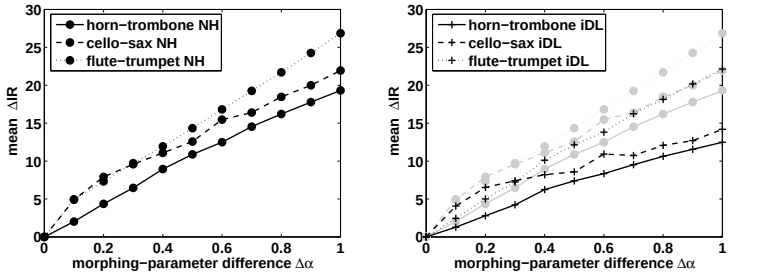
Table 5.2: Correlation coefficients of subjective and objective results of the timbre similarity rating experiments. The values indicate correlation of rating results with internal representation differences ΔIR for the different model versions (see Table 5.1 and text for parameter settings). Bold numbers indicate the model versions with the highest correlation coefficients for the individual continuum correlations (M8, M10) and the simultaneous correlation of all continua (M6, M10).

quently correlated (using the Pearson product) with the similarity ratings given by the normal-hearing subjects (Chapter 2). Correlation coefficients of subjective and objective results are shown in Table 5.2. The correlation was done for the three instrument continua separately, and for all continua simultaneously. For the individual continuum correlations (upper 3 rows in Table 5.2), correlation coefficients, which range from 0.66 to 0.79, did not vary distinctly. However, the highest correlation coefficients were found for versions M8 and M10, which use preprocessing without a modulation filter bank, time-average the IR (using the entire length of the stimulus or only the first 300ms, respectively) and using Euclidean distance as a measure of IR distance. In the simultaneous correlation of all continua (lower row in Table 5.2), coefficients varied to a higher extend. Coefficients were highest for versions M6 and M10, which use preprocessing without a modulation filter bank, correlate the entire IR or only the attack mean, and use Euclidean distance as IR distance measure.

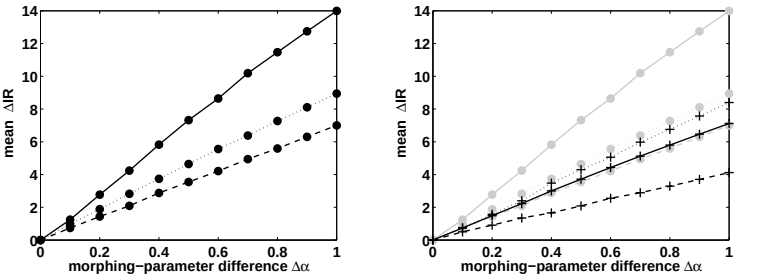
For the model versions M6, M8 and M10, which produced the highest correlation with the subjective results, and version M9, which includes the modulation filter bank, the left panels of Figure 5.2 show mean IR distances as a function of the morphing-parameter difference $\Delta\alpha$. (Note that version M7, which compares the 3-dimensional IRs including 6 modulation channels across stimuli and also produced high correlation with the subjective results, is not shown. M7 produced results almost equal to those of M6 without modulation analysis (data not shown).) For comparison, experimental data is shown in Figure 5.2(a) (left). Additionally, IR distance using the model stage for hearing-impaired processing was calculated using the audiogram of subject iDL (Figure 5.2, right), whose experimental results showed the highest deviations from normal-hearing subjects (Figure 5.2(a) right). All rating curves increase monotonically in agreement with the experimental data (Fig-



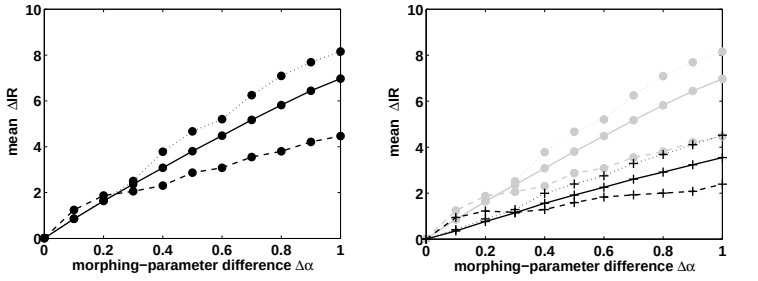
(a) experimental results



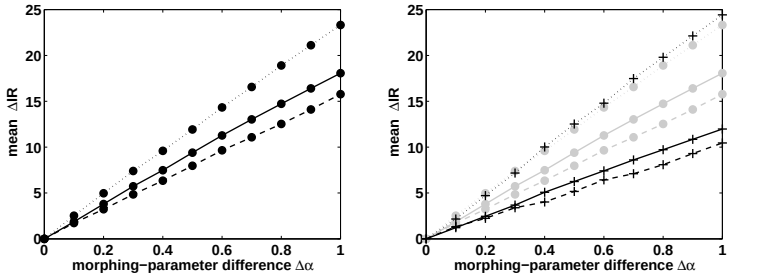
(b) model version M6



(c) model version M8



(d) model version M9



(e) model version M10

Figure 5.2: Mean modeled similarity (b-e) of stimuli in timbre rating experiments as function of morphing-parameter difference $\Delta\alpha$ using model versions M6, M8, M9 and M10 (see Table 5.1 and text for parameter settings). Circles and crosses represent results of normal-hearing listeners (left panels) and hearing-impaired subject iDL (right panels), respectively. For comparison, experimental results are shown in (a). In the right panels, results for normal-hearing listeners in the respective model version or experiment are additionally shown as grey lines and circles.

ure 5.2(a)). The main difference between the different model versions and the main difference between objective and subjective data seem to lie in the different curve order, that is differences in slope between the continua (compare Figures 5.2(a)-(e)). The individual curves show a rather constant ΔIR growth with $\Delta\alpha$, in particular for $\Delta\alpha > 0.2$. This indicates that within each continuum, the predicted similarity shows the same variation with relative morphing-parameter difference $\Delta\alpha$ as the experimental data (Figure 5.2(a)). (Note that here no internal noise, which is used to simulate the discrimination threshold in the discrimination measurements shown below, was used in the rating simulations.[30] Therefore, results of simulations near threshold will not be discussed here but in the following sections.)

Hence, perceptive timbre distances above threshold could be predicted for stimuli in the same instrument continuum. For the individual continuum correlations, correlation coefficients do not vary distinctly across model versions. Predictability does not depend distinctly on IR distance measure, number of modulation filters used or presence of temporal IR dimension. This is probably due to the main variation of subjective and objective data with *relative* morphing-parameter *difference* $\Delta\alpha$, which is well represented by the relative IR difference for all model versions (see Figure 5.2, which shows that IR distances ΔIR and subjective ratings increase monotonically with $\Delta\alpha$).

When correlating all continua simultaneously, coefficients varied to a higher extent. Different model versions lead to different ΔIR growth with $\Delta\alpha$. The crucial difference between objective and subjective data seems to lie in the relative slopes (of the ΔIR-$\Delta\alpha$ function) across instrument groups. Since the different instrument groups vary in spectro-temporal parameters (Chapter 2), this reflects the lack of clarity in perceptual weighting of temporal and spectral differences in rating an overall difference. For the objective approach, this indicates the difficulty of weighting IR differences along time, frequency and modulation frequency channels. The model predicted incorrectly the order of instrument continua when comparing IR time-step-wise (compare Figure 5.2(b) and (a)), and too shallow an increase in the cello-sax curve when using a modulation filter bank (compare Figure 5.2(d) and (a)). Comparing only the IR's temporal mean of the attack (M10) seems to best predict the subjective data for normal-hearing subjects (compare Figure 5.2(e) to (a)). Iverson & Krumhansl (1993) showed that the crucial information for similarity ratings (including spectro-*temporal* cues) of natural instruments is contained in the spectrum of the attack. Since the time-step-wise comparison in the model may introduce non-perceived differences (see below), comparing the attack's spectrum may be a good compromise to account for perceived spectral and temporal timbre differences.

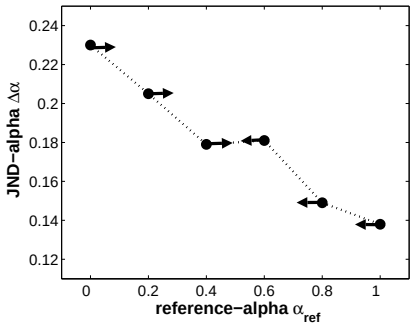
Depending on the model version, the model stage simulating the hearing loss shows different effects in the different instrument continua. Using no modulation filter bank and time-averaging the IR (M8) leads to the best prediction of subjective results.

5.2.2 Predicting JND against morphing-parameter

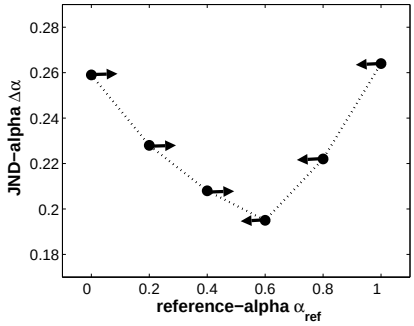
Chapter 3, in which timbre discrimination experiments were performed, showed slight but significant JND dependency on morphing-parameter of the reference stimulus within one instrument continuum. This seemed to be due to spectral differences (such as difference in spectral centroid) in the horn-trombone continuum, but due to a combination of spectral and spectro-*temporal* differences (such as the spectral flux) in the cello-sax continuum. With the same AFC procedure as described in Chapter 3, the morphing-parameter difference $\Delta\alpha$ was adapted to a predicted JND value in the present study. Automatic comparison of the three presented intervals was made by the largest pair-wise IR distance between the stimuli. If IR distance was below a given internal noise level, the test stimulus was selected as containing the "different" stimulus, in the other case a random interval was assigned[31]. The value of the model decision threshold ("internal noise") was set such that, in the horn-trombone continuum in the condition with α_{ref}=0, the predicted JND was equal to the experimental JND result of $\Delta\alpha$(JND)=0.23. The other 11 conditions were subsequently simulated with the same internal noise value.

Figure 5.3 shows the JND results of the simulated measurements for model versions M6, M8 and M9. (Note that version M10, which compares the attack's time-averaged IR across stimuli and accurately predicted similarity rating results in all continua, is not shown for the discrimination experiments, because the natural attack was removed from the stimuli for the JND measurements. In all discrimination measurements of the present study, M10 produced JND results almost identical to those of M8, in which the IR is time-averaged along the entire stimulus (data not shown).) The 1-dimensional IR of version M8, that is, without modulation filter bank and time-averaged, leads to the best predictions in the horn-trombone continuum (Figure 5.3(e) and (f)), but for the cello-sax continuum, predicted JNDs are distinctly higher than the experimental values. Using 6 modulation filters and time-averaging (M9) does not result in any improvement in the predictions in any continuum (Figure 5.3(g) and (h)). When using the time-step-wise comparison of stimuli (M6), predicted JNDs in the cello-sax continuum are lower and predict experimental results better, whereas the relative JND trend in the horn-trombone continuum resembles experimental data less accurately than when the IR is time-

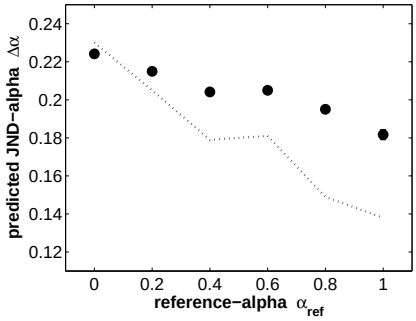
Figure 5.3: Simulated timbre JND as a function of morphing-parameter α_{ref} of the reference stimulus for model versions M6, M8 and M9. For comparison, (a) and (b), as well as the dotted lines in (c)-(f), show experimental results, which are out of range in (f) and (h). Note that in the right panels, the ordinates show different ranges. The arrows in (a) and (b) indicate the direction of the reference stimulus in which the JND was measured.



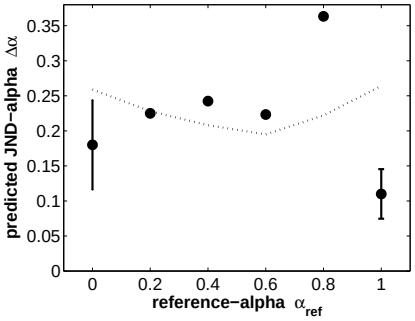
(a) horn-trombone, experiment



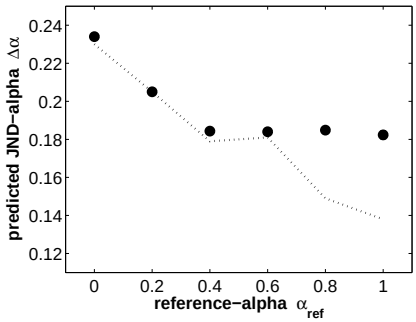
(b) cello-sax, experiment



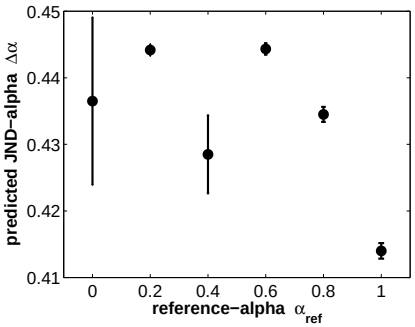
(c) horn-trombone, M6



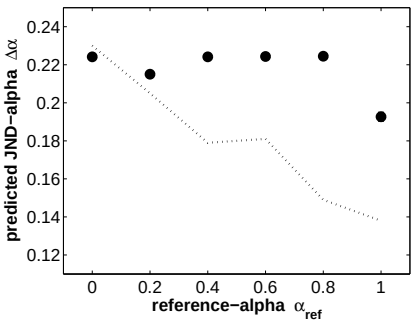
(d) cello-sax, M6



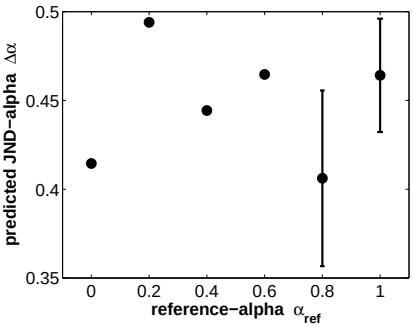
(e) horn-trombone, M8



(f) cello-sax, M8



(g) horn-trombone, M9



(h) cello-sax, M9

averaged (Figure 5.3(c) and (d)). This is an indication that spectral cues are used for discrimination in the horn-trombone continuum, and temporal cues are used in the cello-sax continuum. This may be the reason that a time-averaged model version without modulation analysis (M8) predicts the horn-trombone continuum best while a temporally resolved model version (M6) performs best for the cello-sax continuum.

However, the predictive power of the models used here is very limited: Even the best predictions achieved for the horn-trombone transition for M8 is only achieved for the range of α_{ref}=0 to 0.6 (Figure 5.3(e)). Likewise, for the cello-sax continuum, the best prediction with M6 does not cover the range for α_{ref}=0.8 to 1 well (Figure 5.3(d)). One reason for the limited accuracy of the models might be that the underlying assumption of an Euclidean distance and a uniform weighting in the IR space is not fulfilled.[32] Hence, the option for a different weighting is explored in the following.

Weighting of time and frequency

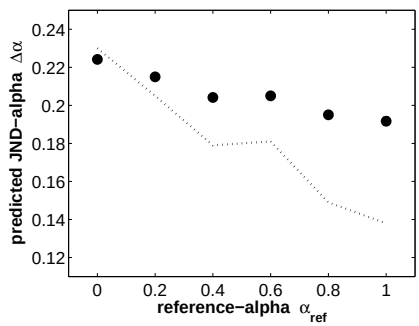
In order to improve prediction of the relative JND trend in both continua, a modification was tested of the way in which information across frequency, modulation frequency and time is combined to achieve the final distance of the internal representations. Hence, JNDs were predicted using different "p-weights", with which ΔIR was averaged along the frequency channels and time steps (M8*, Figure 5.4). In this modification of model version M8, after the preprocessing without a modulation filter bank,

$$IRdist = (\frac{1}{T \cdot F \cdot M} \cdot \sum_{t,f,m} |IR_1(t,f,m) - IR_2(t,f,m)|^p \ )^{1/p} \qquad (5.4)$$

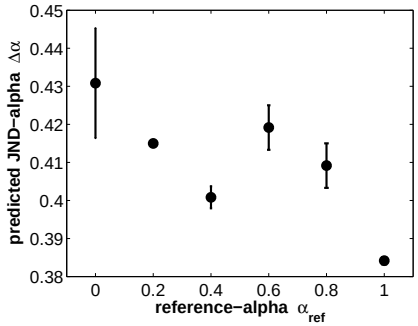
with p=0.5 and p=10, was used as a distance measure instead of the Euclidean distance (for which p=2, see Equation 5.2). As in M8, IR was time-averaged before calculating ΔIR. Figure 5.4 shows that in the cases of p=10 and p=0.5 in the horn-trombone continuum, the relative JNDs show the same decreasing trend with α_{ref} as the experimental results. For p=0.5, JND for α_{ref}=1 is predicted better, but JNDs for α_{ref}=0.4 and 0.6 are predicted more poorly.

Hence, for p-weights p=10 and p=0.5 the simulations with model version M8* predict a decreasing JND trend with α_{ref}, which was observed in the experimental results, along the entire horn-trombone continuum, but JNDs at α_{ref}=0.4 and 0.6 are predicted less accurately than for p=2 in version M8. Predictions at α_{ref}=1 are better for p=0.5 (in M8*) than for p=2 (in M8). Profile analysis studies

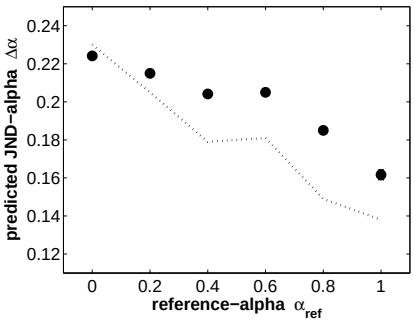
Figure 5.4: Effect of p-weighting in the IR-distance measure for simulation with a modification of model version M8. Simulated timbre JND (expressed as JND of the morphing-parameter α) as a function of the morphing-parameter α_{ref} of the reference stimulus in the horn-trombone and cello-sax continua. For comparison, dotted lines indicate experimental results, which are out of range in (b) and (d).



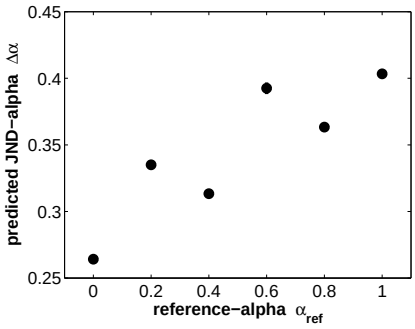
(a) horn-trombone, M8*, p=10



(b) cello-sax, M8*, p=10



(c) horn-trombone, M8*, p=0.5



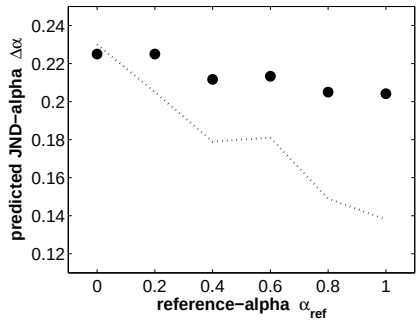
(d) cello-sax, M8*, p=0.5

showed that intensity increments of complex tones are detected more easily in certain frequency regions than in others (Green, 1988b; Lentz & Leek, 2003) and peripheral suppression effects may lead to different weighting of frequency regions[33]. For timbre discrimination of harmonic complex tones as in the present case, interaction of frequency regions on the basilar membrane may not be negligible (Appendix B). Hence, an additional weighting of frequency channels or substituting the gammatone filter bank with a more physiological filter bank[34] may improve prediction of subjective data. However, these modifications are out of the scope of the current study.

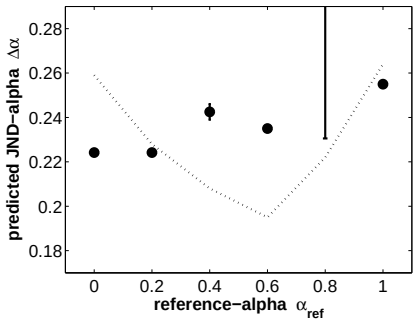
Cross-correlation as a distance measure

Figure 5.5 shows results of simulations with model versions M1 and M3, i.e., using cross-correlation (Equation 5.3) as a distance measure. When the IR was time-

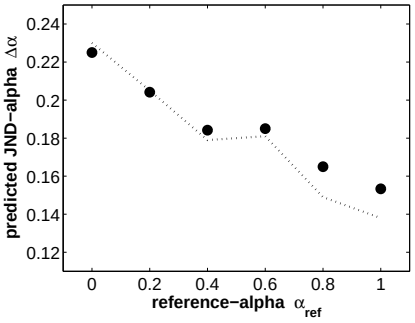
Figure 5.5: Simulated timbre JNDs as a function of the morphing-parameter α_{ref} of the reference stimulus using cross-correlation as the distance measure. For comparison, dotted lines indicate experimental results, which are out of range in (d).



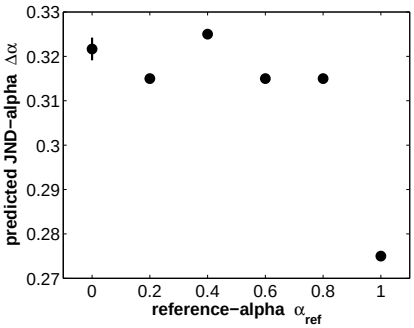
(a) horn-trombone, M1



(b) cello-sax, M1



(c) horn-trombone, M3



(d) cello-sax, M3

averaged (M3), cross-correlation of the spectrum accurately predicts the threshold in the horn-trombone continuum (Figure 5.5(c)). In the cello-sax continuum, prediction was improved by time-step-wise comparison of stimuli (M1), whereas in the horn-trombone continuum the relative JND trend resembled experimental data more poorly (compare Figures 5.5(a)-(b) to (c)-(d)).

Hence, using cross-correlation as the distance measure of the time-reduced IR (M3) succeeded in generating the best prediction of the results in the horn-trombone continuum[35]. However, for any distance measure and weighting using the temporal IR mean (M3, M8, M8*, M9), JND predictions in the cello-sax continuum are too high (Figure 5.3). Better prediction of results in the cello-sax continuum requires a time-step-wise comparison of stimuli (M1, M6), which confirms the spectro-*temporal* discrimination cues that were found to likely be exploited in this continuum (Chapter 3). On the other hand, in the horn-trombone continuum the time-step-wise IR comparison worsened prediction results, and in the cello-sax continuum neither distance measure nor p-weighting could predict the relative JND

trend observed in the experimental results. One explanation may be a different p-weighting of time steps than of frequency channels, because concurring temporal and spectral cues are used to distinguish stimuli in the cello-sax continuum (Chapter 3). Additionally, a dynamic time warp may improve predictions of spectro-*temporal* differences (see below).

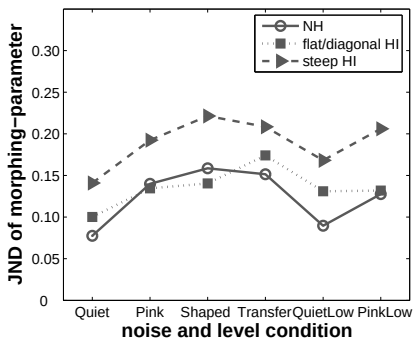
5.2.3 Predicting JND against level and background noise

Chapter 4 showed timbre discrimination measurements with different level and noise conditions with normal-hearing and hearing-impaired subjects. Figure 5.6 shows the predicted JND results simulating these measurements, for which a preprocessing without a modulation filter bank was used (M6, M8, M8*). For Figures 5.6(e)-(h), the time-averaged version of the IR was used (M8, M8*). For Figures 5.6(c)-(f), the common Euclidean distance (p=2) was used as the distance measure (M6, M8), while for Figures 5.6(g) and (h) a p-weight with p=10 (Equation 5.4) was used to average across IR channels (M8*). In the same AFC procedure as described in Chapter 4, the morphing-parameter difference $\Delta\alpha$ was adapted to a predicted JND value. In contrast to the experiment, in the noise conditions the same ("frozen") noise was used for all intervals in a trial; the noise was generated randomly for every trial. Before calculating the IR distance, IR of the frozen noise was subtracted from IR of the intervals. For automatic decision, the IR distance between test intervals and reference interval was calculated. If the difference between the two IR distances exceeded the internal noise, the interval with maximal IR distance was selected to contain the test signal; in the other case a random interval was selected[36]. The value of the model decision threshold (internal noise) was set such that in the horn-trombone continuum in the *Quiet* condition, the predicted JND was equal to the experimental JND result of $\Delta\alpha$(JND)=0.08.

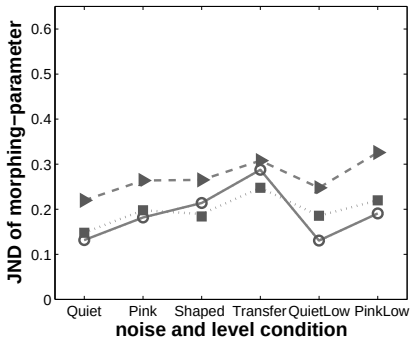
In the *Transfer* condition, in which the reference stimulus was heard in quiet and the test stimuli in noise, the simulated adaptive measurements did not converge in any simulation. This is due to the non-additivity of IRs; for example, IR of noise plus IR of signal does not result in IR of signal and noise. Or in other words, the model that does not include cognitive grouping processes cannot predict object separation. A model that compares perceptive timbre descriptors, for example the spectral centroid of the tonal harmonic sound content, may be able to simulate the measurements of the *Transfer* condition. Hence, in the following, only results of the remaining 5 conditions will be presented.

For normal-hearing subjects in the horn-trombone continuum, model version M8 led to a realistic prediction of the subjective data (compare Figure 5.6(e) to

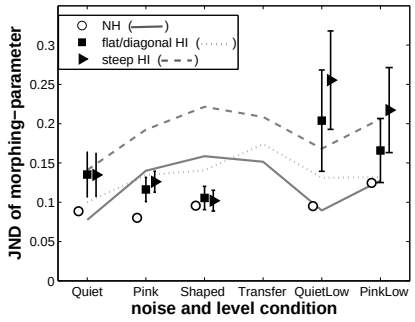
Figure 5.6: Simulated results of timbre JND measurements in different noise/level conditions (see Table 4.1 and text in Chapter 4 for details) For comparison, (a) and (b) as well as grey symbols and lines in (c)-(h) indicate experimental results. Different symbols represent the JND means of the three listener groups of normal-hearing listeners (circles), flat/diagonal hearing-impaired listeners (squares) and steep hearing-impaired listeners (triangles).



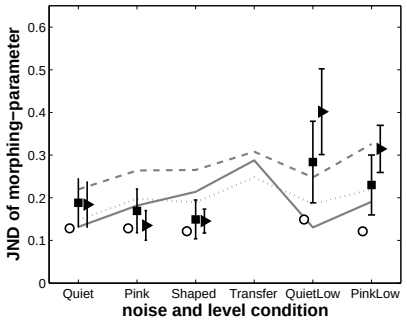
(a) horn-trombone, experiment



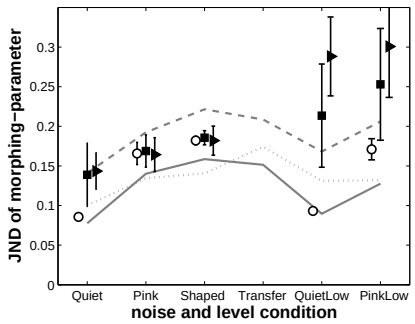
(b) cello-sax, experiment



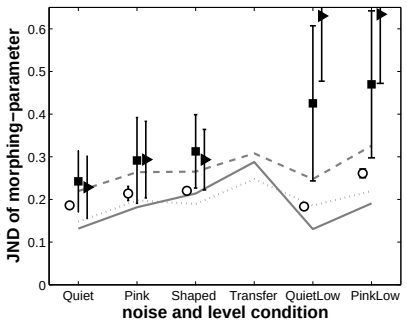
(c) horn-trombone, M6 (p=2)



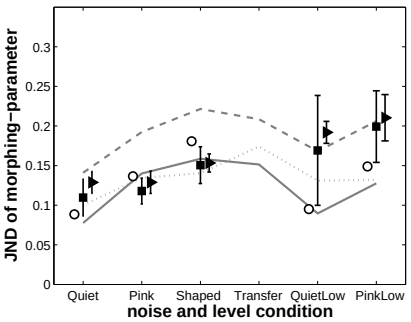
(d) cello-sax, M6 (p=2)



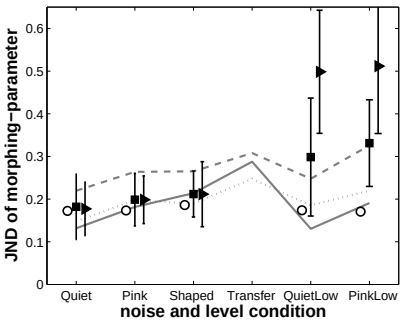
(e) horn-trombone, M8 (p=2)



(f) cello-sax, M8 (p=2)



(g) horn-trombone, M8*, p=10



(h) cello-sax, M8*, p=10

(a)). Specifically, the masking effect of the noise led to a distinct increase in JND, whereas reducing the level had little effect on JND. In the cello-sax continuum, neither JND in quiet nor the noise effect on JND was predicted as well. In quiet, the JND predictions of flat hearing impairment are too high in the horn-trombone continuum, and predicted JNDs of steep hearing impairment are too low in the cello-sax continuum. In noise, predicted JNDs of steep hearing impairment are too low in the horn-trombone continuum. However, for all hearing impairments a reduction of level led to a predicted JND increase, which is in agreement with the experimental results, although the predicted level effect is too high. The main difference between objective and subjective data with respect to hearing impairment seems to be the effect of threshold configuration. In the experiment, subjects with a steep hearing loss showed significantly higher JNDs than subjects with a flat hearing loss, whereas in the simulation no distinct differences for the high-level conditions can be observed.

Using the time-step-wise IR difference for calculating the IR distance yields a more accurate prediction for normal-hearing subjects in the cello-sax continuum than using the time-averaged version; that is, model version M6 predicted experimental JND results better than M8 (compare Figure 5.6(c) and (d) to (a) and (b), respectively). However, effects of noise on JND are predicted more poorly by M6 than when using the temporal IR mean in M8. Model versions M1 and M3, which use cross-correlation as the distance measure, led to similar results as M6: they predicted almost no noise and level effects for normal-hearing listeners (not shown here).

Using other p-weights to calculate IR distance (M8*, Equation 5.4) instead of Euclidean distance improved JND prediction of normal-hearing subjects in the cello-sax continuum (Figures 5.6(g)-(h)). On the other hand, JND difference between conditions *Pink* and *Shaped*, that is, the effect of the background noise *spectrum*, in both continua, as well as the effect of noise in the cello-sax continuum, are predicted with less accuracy (compare Figure 5.6(g) and (h) to (a) and (b), respectively). In some noise conditions, using a p-weighting of p=10 led to a higher JND for normal-hearing subjects than for hearing impaired.

Hence, for normal-hearing subjects in the horn-trombone continuum, the masking effect of noise, as well as low level effect, is predicted best using the Euclidean distance of time-reduced IRs (M8). In this case a slight masking effect is also predicted in the cello-sax continuum[37]. However, no model version could predict the masking effect of noise sufficiently in the cello-sax continuum. When using other p-weights (M8*), time-step-wise IR comparison (M6) or cross-correlation as a distance measure (M1, M3), in both continua either the predicted masking effect of noise is too low or a distinct difference between noise spectra is predicted, which

was not observed in the experiment. The low noise effect may be due to the *frozen* noise used in the simulations. Using random noise and subtracting the mean IR of 20 noise presentations may improve predictability[38]. However, this modification is beyond the scope of the present study.

In the simulation of the hearing-impaired system, the level effect on hearing-impaired subjects can be predicted, although predicted effect exceeds the effect observed in the experiment. However, the main discrepancy between simulated and experimental data is the difference between flat and steep hearing impairment. In the experiment, JNDs of steep hearing-impaired subjects were significantly higher than JNDs of flat hearing-impaired subjects, whereas no distinct difference was predicted by the model. Different p-weighting (M8*) or time-step-wise IR comparison (M6) succeeded in some difference between flat and steep hearing loss. An optimal p-weighting of temporal and spectral dimension may lead to better predictions. Another possibility is a different weighting along the spectral dimension for flat and steep hearing impairment according to psychoacoustical profile studies with normal-hearing and hearing-impaired subjects. Green (1988b), Doherty & Lutfi (1999) and Lentz & Leek (2003) verified that normal-hearing listeners and people with different hearing loss rely on different components of the spectrum: Normal-hearing listeners rely more on the central components of the spectrum, whereas hearing-impaired listeners are more likely to use the edge sound components (Lentz & Leek, 2003). In addition, hearing-impaired listeners with a steep hearing loss weigh the region of their hearing loss more efficiently than normal-hearing listeners (Doherty & Lutfi, 1999).

5.3 Discussion

The results show that simulating timbre rating and discrimination experiments of morphed stimuli with equal fundamental frequency is possible with the Perception Model *PeMo* with regard to morphing-parameter α, background noise and signal level. For normal-hearing and hearing-impaired listeners, *PeMo* can predict the variation of timbre discrimination threshold across different spectro-temporal dimensions, which are represented by the different instrument continua, as well as the JND increase caused by a noise masker and a lower signal level. However, depending on instrument group, noise condition and hearing loss, a different model version is required:

- **Timbre similarity ratings** *within* the continua were predicted well by all model versions and in all continua, whereby the versions M8 and M10 led to

the best cross-correlation with subjective results. Across instrument continua, version M10, followed by M6, led to the best predictions for normal-hearing listeners, whereas version M8 was best for hearing-impaired listeners.

- Predicting the variation of **timbre discrimination thresholds in normal-hearing listeners** with morphing-parameter (α_{ref}) required the model versions M3, M8 or M8* in the horn-trombone continuum and M1 or M6 in the cello-sax continuum, whereby M3 and M1 led to the best results, respectively. JNDs in noise were best predicted by M8 or M8* in both continua, whereas noise effects predicted by M1, M3 and M6 were too low.

- **Timbre discrimination in hearing-impaired listeners** in quiet and noise was best predicted by versions M8 and M8*. However, all model versions overestimated the level effects and underestimated the differences between flat and steep hearing impairment.

- **Object separation**, tested by the *transfer* condition, could not be simulated due to non-additivity of IRs.

The model versions tested in the present study differ in the number of modulation filters used, the number and length of time steps compared across stimuli, and the distance measure used between stimuli. None of the model versions that led to good predictions (M1, M3, M6, M8, M8* and M10) uses the modulation filter bank. Instead, these versions only compare 1- (frequency) or 2-dimensional (frequency $\times$ time) IRs. M1 and M6 compare IRs time-step-wise across stimuli, whereas M3, M8 and M8* time-average the IRs prior to comparison, and M10 time-averages the IRs using only the length of the attack (first 300 ms). While M1 and M3 use a cross-correlation (Equation 5.3) as the distance measure, M6, M8 and M10 calculate the Euclidean distance (Equation 5.2) and M8* calculates a p-weighted distance (Equation 5.4).

Model settings dependent on timbre dimension, noise condition and hearing loss

Hence, similarity ratings of stimuli with natural attack segment seem to be best predicted by comparing the energy distribution along IR frequency channels during the 300 ms attack segment (M10), which confirms the findings of Iverson & Krumhansl (1993) that the crucial information for timbre similarity ratings is contained in the spectral information of the stimulus attack segment. However, for the discrimination experiments, the attack was removed. Discrimination differences

along the horn-trombone continuum are best predicted without a modulation filter bank and by time-averaging the IRs (model version M3 or M8), that is by comparing the energy distribution along frequency channels of the 1-dimensional IR. Chapter 3 showed that the horn-trombone continuum is mainly varied by spectral differences. A time-step-wise comparison of IRs may add temporal differences that subjects did not use to distinguish the stimuli in this continuum. On the other hand, the perceived timbre differences along the cello-sax continuum can only be predicted sufficiently by time-step-wise comparison of IRs (M1, M6). Chapter 3 showed that the variation along the cello-sax continuum is dominated by a spectro-*temporal* cue, which requires temporal comparison of stimuli. Hence, the optimal model settings agree with the psychoacoustical results of perceptual cues found in previous chapters.[39] However, **in noise**, a time-step-wise comparison (M6) produced threshold predictions that were too low for both continua; thus, simulating the masking and distracting effects of noise requires a temporal average of IR (M8). Another possibility would be to use random noise instead of the frozen noise used here, and to subtract the mean IR of 20 noise presentations and then compare time-step-wise (M6).

Modulation filters and time-averaging the IR (M9) do not seem to extract the temporal stimulus differences in the cello-sax continuum. This may be due to the modulation phase that was removed by the temporal mean, even though the amount of (modulation phase) synchronicity between the frequency channels is perceived by subjects (see Chapter 3 and Appendix A). Including modulation filters in the time-step-wise comparison of IR does not seem to extract more temporal stimulus differences than excluding modulation analysis. This may be because the underlying assumption of a uniform weighting in the IR space is not fulfilled. An additional weighting of modulation channels and a time-step-wise comparison using larger time windows may improve prediction of subjective data. However, since timbre is the result of a *simultaneous* comparison within one stimulus (e.g. across frequency channels, Green, 1983), an additional model stage that accounts for the simultaneous comparison across frequency channels may improve simulation of timbre experiments. Further modeling efforts are required to test these modifications.

Various **distance measures** as introduced above were tested in the present study and compared to the common Euclidean distance. A p-weighted distance (M8*), for example with p=0.5 or p=10, led to slight improvement of simulation of spectral timbre discrimination. Using cross-correlation as the distance measure (M1, M3) distinctly improved predictability of timbre JNDs in silence. That is, model version M3 predicts well spectral discrimination thresholds in dependency of spectral distribution, and M1 predicts well the spectro-*temporal* discrimination thresholds in

silence. However, cross-correlation cannot predict noise and level effects on JNDs.

Effects of hearing loss (a distinct increase of JNDs at low levels and slight increase of JNDs at intermediate levels) could be predicted by additional model stages simulating compression loss and attenuation. Note that it was not necessary to increase the internal noise at the decision stage of the model as was done in the original version of *PeMo* for the impaired hearing system (Derleth et al., 2001). However, the difference between flat and steep hearing impairments that was observed in the experiment could only be predicted at low levels. The reason may be the lack of simulation for dead regions (Chapter 4), such as an increased internal noise for hearing loss above 70 dB. Another reason may be a different perceptual weighting dependent on hearing loss configuration (Green, 1988b; Doherty & Lutfi, 1999; Lentz & Leek, 2003), which was not modeled here.

Model limits

If two sounds differ in various independent components above perception threshold, perceptual weighting of these components to an overall similarity may significantly depend on subject and stimuli (McAdams et al., 1995). Since timbre is a multidimensional sound attribute, modeling timbre rating is limited to tasks that do not incorporate such (conscious or unconscious) attentional weighting of synchronous variation of sound attributes. Hence, in the present study, similarity ratings *within* a continuum that is varied along one spectro-temporal dimension could be predicted well, but ratings across continua could not.

Attention and memory limits may play a role in predicting timbre discrimination thresholds. If a subject knows *which* sound feature is varied along the presentation intervals, he/she can attend to this feature and his/her measurement results approach his/her physiological threshold. If the subject does not know which feature is varied, attention is spread along all sound features that may be changed and JND results may be higher than in a unidimensional task. Since capacity of memory for neurosensory tasks is limited, and since timbre may vary in a large number of independent percepts, some changes may be dismissed even by trained subjects. *PeMo*, which does not incorporate attentional model stages, may predict perception thresholds that are too low, if various independent sound features are varied or if the feature that is varied is not specified to the subject. This may be the case for synchronous spectral and spectro-*temporal* variation in the present study.

PeMo is limited to bottom-up processes. Hence, all tasks including top-down processes like cognitive grouping mechanisms for object separation (e.g. for separating tonal sound objects from noise) or subjective perceptual weighting of frequency

regions. Since timbre is the result of a *simultaneous* comparison (i.e. grouping) in one stimulus (Green, 1983), this may also cause problems in timbre experiments with natural timbres. For example, for the percept and discrimination of overtone synchronicity (Appendix A), the time dimension is crucial for the comparison process in the stimulus along overtones, while time-step-wise comparison of IRs between stimuli may mislead and give false information. In the present study, perceptual and physical variation along the morphed continua show parallel trends due to the morphing algorithm. If, for example, overtones become asynchronous along the continuum, the spectrum at each time step changes gradually along the continuum.

5.3.1 Summary

The present study simulated timbre rating and discrimination experiments performed in the previous chapters with the effective Perception Model *PeMo* (Dau et al., 1996, 1997a) and some modifications with respect to preprocessing for the hearing-impaired system (Derleth et al., 2001), the distance measure (Holube & Kollmeier, 1996; Huber & Kollmeier, 2006) and the weighting of time, frequency and modulation dimension. While simulation of none of the timbre experiments required the modulation filter bank, changes in other model settings improved predictions. The main outcome of the study can be summarized as follows:

- The rating results of psychoacoustic measurements could be predicted in terms of timbre similarity dependence on morphing-parameter difference. The best correlation of objective and subjective results was found by reducing the internal representation (IR) to the frequency distribution of the first 300 ms of the stimuli. This confirms the findings of Iverson & Krumhansl (1993) that the crucial information for timbre similarity ratings is contained in the spectral information of the stimulus attack segment.

- Discrimination thresholds of normal-hearing subjects for stimuli without attack could be predicted for the spectrally dominated instrument continuum horn-trombone, when reducing IR to purely spectral distribution. On the other hand, the spectro-*temporal* variations along the cello-sax continuum required a time-step-wise IR comparison and could not be extracted by the modulation filter bank.

- Since the multidimensional timbre variation of stimuli in the present study assumes some cognitive attention processes of the subjects involved in the measurements, while predictions by *PeMo* are limited to bottom-up processes, no

consistent set of model parameters could be found that accounts for simultaneous changes of spectral and (spectro-)temporal timbre dimensions. The high number of timbre percepts and, hence, high number of possible stimulus variations in the experiment may also limit predictions due to memory limits of subjects. However, different p-weighting along and across frequency channels and time steps may account for perceptual weighting and may improve predictions in unidimensional timbre tasks as shown in this study.

- While cross-correlation as a distance measure accounts for threshold variation with morphing-parameter of the reference stimulus, using Euclidean distance is more applicable for level and background noise effects. Although the Euclidean distance is restricted to independent observations, it seems (with restrictions) to be applicable to the complex stimuli with multidimensional variation of the present study. In contrast to sound quality studies (Huber & Kollmeier, 2006), which uses cross-correlation, here the Euclidean distance seems to be more appropriate for stimulus comparison above discrimination threshold.

- When modeling the impaired hearing system with the instantaneous attenuation and expansion suggested by Derleth et al. (2001), some experimental results could be predicted without increasing the internal noise, which was necessary in the studies of Derleth et al. (2001). The significant differences between flat and steep hearing impairment, which were observed in the experiments, could only be predicted at low levels.

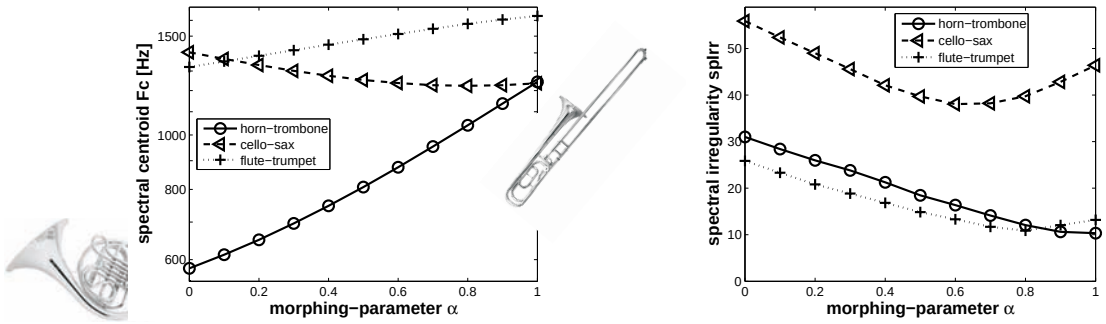
- In order to account for the multidimensionality of timbre, further modeling efforts are required to clarify the weighting in the IR space or to substitute the gammatone filter bank by a more physiological filter bank. Since timbre is the result of a *simultaneous* comparison in one stimulus (Green, 1983), an additional model stage that accounts for the across-channel processing may further improve simulation of timbre experiments for normal-hearing and hearing-impaired listeners. With respect to the processing of the impaired auditory system, future studies should consider an increased frequency-dependent internal noise that depends on the individual hearing loss. This might represent the individual variability in the loss of inner hair cells, which has to be included in order to model the differences between flat and steep hearing losses.

Appendix A

Timbre

The label *timbre* combines all auditory object attributes other than pitch, loudness, duration, spatial location and reverberation environment. The *physical* timbre space is made up of frequency, time and amplitude of sound, which are the fundamental measures of acoustics, while timbre *perception* is multidimensional with descriptions like brightness, roughness and noisiness. Previous timbre studies tried to find a timbre *model* by connecting physics and perception, that is, finding linear combinations ("spectro-temporal timbre descriptors") of fundamental physical measures that represent the perceptual timbre dimensions or psychophysical quantities that represent timbre. These studies have shown that timbre dimensions are connected to certain spectro-temporal descriptors that do not necessarily combine all three fundamental measures linearly and with equal weights. For example, the descriptor "spectral centroid", which is a physical representative for the brightness percept, is a combination of amplitude and frequency, while the optimal way of calculating the spectral centroid is unsolved (e.g. Grey, 1977; McAdams et al., 1995).

The measurements in Chapter 2 use three stimulus continua, which are generated by "morphing" natural instruments pair-wise. The musical instruments were chosen in a way that each pair was very dissimilar in one timbre-dominating dimension of Grey's (1977) MDS space and similar in the other dimensions: Trombone and French horn show different spectral centroids, saxophone and cello that according to Grey (1977) differ mainly in spectral flux, and flute and trumpet that differ mainly in the attack segment. By linear interpolation of spectral parameters, sounds were then pair-wise cross-faded (morphed), whereby continua between trombone and French horn (horn-trombone continuum), cello and saxophone (cello-sax continuum), and flute and trumpet (flute-trumpet continuum) were generated. (A more detailed description of the morphing method can be found in Chapter 3.) The morphed stimuli are named by their morphing-parameter α, which corresponds to



(a) spectral centroid (b) spectral irregularity

Figure A.1: Spectral energy distribution. (a) Spectral centroid Fc (Equation A.1) and (b) spectral irregularity *spIrr* (Equation A.2) vs. morphing-parameter α for horn-trombone (circles), cello-sax (crosses) and flute-trumpet (triangles) continua. Note that the ordinate in (a) showing Fc is plotted on a logarithmic scale.

the ratio of one of the original instruments to the original sounds. Hence, α ranges between 0 (corresponding to the sound of the original French horn, cello or flute) and 1 (corresponding to the sound of the original trombone, saxophone or trumpet), where a spacing of 0.1 was used. The morphing-parameter α is a measure that combines the linear frequency scale with the linear amplitude and linear time scale (Section 2.2). In order to link the common timbre models with the measure α used in the present thesis, in the following the spectro-temporal descriptors that are commonly correlated with timbre similarity ratings are calculated for the three instrument continua used in Chapter 2.

A.1 Spectral energy distribution

The centroid of the spectrum, which represents the percept of brightness, has commonly been shown to be strongly correlated with the most prominent dimensions of multidimensional-scaling representations of musical timbre differences (Grey & Gordon, 1978; Wessel, 1979; Krumhansl, 1989; Iverson & Krumhansl, 1993; Kendall et al., 1999). While there are many ways of calculating the spectral centroid[40], here the spectral centroid Fc is defined as

$$Fc = \frac{\sum_{k=1}^{N} (A_k \cdot f_k)}{\sum_{k=1}^{N} A_k}, \tag{A.1}$$

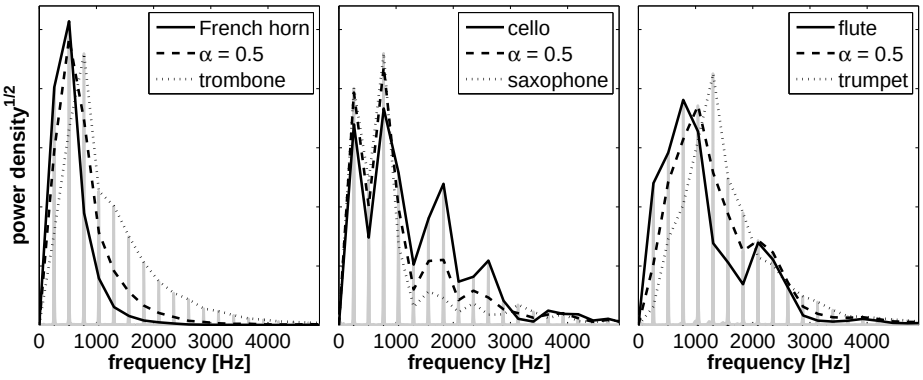


Figure A.2: Spectral energy distribution. Mean spectra of the natural instruments and an intermediate hybrid instrument in the horn-trombone (left), cello-sax (center) and flute-trumpet (right) continua. The spectra are shown in grey, while the spectral peaks are connected by black lines (see legend).

where A_k is the amplitude and f_k the frequency of partial k, and N is the total number of partials (e.g. Krimphoff et al., 1994; McAdams & Winsberg, 2000). Note that only the tonal part of the instrument sounds contribute to the centroid.[41] To test the degree to which the centroid varies along the instrument continua, Figure A.1(a) shows Fc as a function of morphing-parameter α in the three instrument continua. The figure shows that Fc variance in the horn-trombone continuum is higher than in the other continua, which is in agreement with Grey's (1977) study.

Not only the centroid of the spectrum, but also the spectral irregularity or spectral smoothness was found to be a perceived timbre dimension in musical instruments (Krimphoff et al., 1994). For instance, spectra of clarinet sounds show mainly harmonics with odd partial numbers, whereas the even-numbered partials are missing or of low amplitude. Hence, clarinet spectra are jagged and have a high spectral irregularity. Figure A.2 shows the spectra of the stimuli with α=0, 0.5 and 1, which are the spectra of French horn, trombone, cello, saxophone, flute, trumpet, and the 50% hybrid of each continuum. Similar to clarinet sounds and in contrast to the other continua, the spectra in the cello-sax continuum show irregular harmonic amplitudes. The horn as well as the trombone spectrum shows a high and narrow maximum, while the energy in the flute-trumpet continuum is broadly distributed, but both continua show relatively smooth spectra. A measure for the spectral irregularity is the logarithm of the (spectral) deviation of component amplitudes from a global spectral envelope derived from a running mean of the amplitude of three

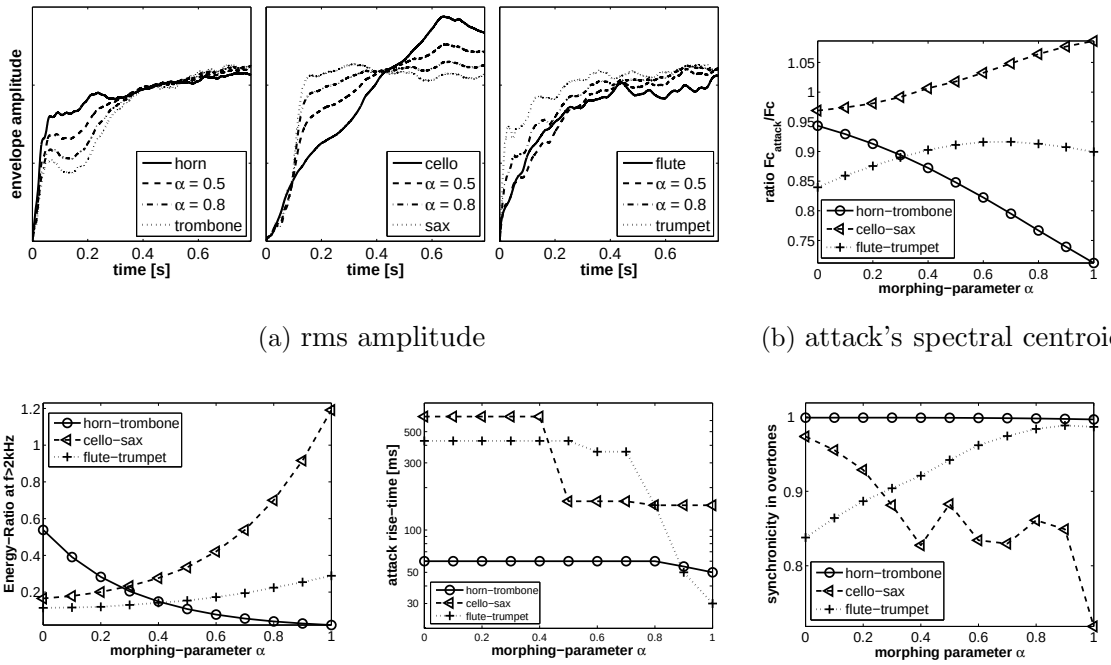
adjacent harmonics (Krimphoff et al., 1994):

$$spIrr = 20 \cdot \sum_{k=1}^{N} |log_{10}(A_k) - \frac{log_{10}(A_{k-1}) + log_{10}(A_k) + log_{10}(A_{k+1})}{3}|, \qquad (A.2)$$

where A_k is the amplitude of partial k, and N is the total number of partials. To test the degree to which the spectral irregularity varies along the instrument continua, spIrr is calculated for all stimuli and shown in Figure A.1(b). As the spectra (Figure A.2) assume, the spectral irregularity spIrr in the cello-sax continuum is twice as high as that of the two other continua (Figure A.1(b)). But, although spIrr in the horn-trombone continuum is lower, the difference in spectral irregularity between horn and trombone is higher than between cello and saxophone (Figure A.1(b)). French horn shows a narrow and high spectral peak at low frequencies, which causes relatively high spIrr without having a "zigzag" irregularity, while the trombone's spectrum is shallower and broader (Figure A.1(a)). Figure A.1(b) also shows that spectral irregularity (as defined by Equation A.2) is not mapped monotonously by α in all continua. Both findings assume that spIrr is an unstable timbre descriptor which seems to be dependent on spectral centroid. This may be a reason that many previous studies did not include spIrr as an independent timbre dimension.

A.2 Attack: rise time vs. high-frequency energy

While the role of spectral irregularity and spectral flux (Section A.3) in timbre perception is discussed controversially, the initial attack has commonly been shown to be correlated with the most prominent dimensions of multidimensional-scaling (MDS) representations of timbre differences (Grey & Gordon, 1978; Wessel, 1979; Krumhansl, 1989; Iverson & Krumhansl, 1993; Kendall et al., 1999). Grey (1977) described this dimension as the presence of high-frequency, low-amplitude, often inharmonic energy preceding the stable harmonics in the attack segment. This inharmonicity often gives the attack a noise-like character, the attack becomes "buzzlike" or slightly grating, while the absence of this high-frequency energy leads to a clean attack (Grey & Gordon, 1978). If much of this high-frequency energy is present, the attack is longer and softer than if it is absent (Wessel, 1979). This led Krimphoff et al. (1994) to the definition of the log-rise-time for this MDS dimension, that is the logarithm of the time that a sound takes to reach maximum rms amplitude, which often correlates for natural musical instruments with Grey's (1977) definition. But note that the converse of Wessel's (1979) conclusion can not be drawn: if the attack is long, there is not necessarily high-frequency energy present in the attack. For instance, a long rise time may also result from a synchronous increase of all partials,



(a) rms amplitude (b) attack's spectral centroid

(c) attack's high-freq. energy (d) attack's log-rise-time (e) attack's overtone synchr.

Figure A.3: Attack segment. (a) Temporal envelope of the first 800ms for instruments with α=0, 0.5, 0.8 and 1 (continuous, dashed, dash-dotted and dotted lines, respectively) in the horn-trombone (left), cello-sax (center) and flute-trumpet (right) continua. The envelope was calculated by the root-mean-square amplitude of a running window of 10 ms. (b) Spectral centroid of the first 300 ms normalized by the mean centroid of the stimulus, (c) the amount of high-frequency energy (above 2 kHz) in the attack segment (first 300 ms) normalized by the high-frequency energy during the stationary portion of the stimulus (after 600 ms), (d) rise time plotted on a log scale, and (e) synchronicity in upper overtones during the first 300 ms of the stimulus in the horn-trombone (circles), cello-sax (triangles) and flute-trumpet (crosses) continua.

when sound level increases monotonously during a smooth crescendo. The overtone synchronicity (Equation A.3, Section A.3) during the attack was analyzed by Grey & Gordon (1978). Hence, care must be taken when comparing results of different studies, which use the following definitions as a measure for the attack segment:

① presence of low-amplitude, high-frequency energy in the attack (Grey, 1977)

② overtone synchronicity during the attack (Grey & Gordon, 1978)

③ spectral centroid of the attack (Iverson & Krumhansl, 1993)

④ log-rise-time of the attack envelope (Krimphoff et al., 1994)

Since the four definitions of the attack descriptors are distinctly different, all attack measures will be calculated here for the stimuli of Chapter 2. Figures A.3(b) and (c) show the ratio of the attack centroid to the mean centroid (③) and the ratio of

high-frequency energy (above 2 kHz) in the attack segment to high-frequency energy in the stationary portion (①), respectively. In both cases and for all instruments, the attack segment was defined as the first 300 ms, while for the stationary segment the first 600 ms have been removed. Overtone synchronicity (②) in stimuli of the present study is shown in Figure A.3(e), and a rough estimate of stimuli's log-rise-time (④) is shown in Figure A.3(d). The rms envelope in Figure A.3(a) shows that even for some of the 6 natural, non-hybrid instruments the attack's end is ambiguous[42]. However, for this study the (eye-picked) most distinct bend in the rms envelope curve was defined as the end of the attack and used for Figure A.3(d).

Figure A.3(e) shows that the **saxophone** in the present study has asynchronous overtones during the attack. Figure A.3(b) and (c) show that only the saxophone sound exhibits high-frequency energy in the attack segment (ratio >1) and a high attack centroid. On the other hand, the rise time of the saxophone in the present study seems shorter than those of cello and flute (Figures A.3(a) and (d)); Wessel's (1979) hypothesis would assume a longer rise time due to the additional high-frequency energy in the saxophone's attack. However, the saxophone's rise time is far longer than those of the brass instruments, which agrees with the hypothesis. The short rise times of **trumpet**, **trombone** and **horn** (Figures A.3(a) and (d)) are in agreement with the low attack centroid (Figure A.3(b)), lack of high-frequency energy (Figure A.3(c)) and high overtone synchronicity during the attack (Figure A.3(e)) found in these instruments. The long rise time of the **flute** (Figures A.3(a) and (d)) confirms the frequently perceived noisy and smooth attack in flutes (Grey & Gordon, 1978) noted also in the experiment, and agrees with the asynchronous overtones found in the flute's attack (Figure A.3(e)). But the low attack centroid (Figure A.3(b)) and the absence of high-frequency energy (Figure A.3(c)) contradict the equivalence of the attack measures ①, ③ and ④. An explanation for the contradiction may be the noise content in the flute sound of the present study, which is present not only during the attack, but during the entire stimulus (compare Figures A.7(a) and (b) Section A.4). Hence, normalizing the high-frequency energy of the attack (by the high-frequency energy of the stationary portion) may have obscured the real inharmonic high-frequency content in Figure A.3(c)[43]. In the **cello** the long rise time (Figure A.3(d)) is also in contradiction with the low high-energy content during the attack (Figure A.3(c)) and the rather synchronous overtones (Figure A.3(e)).

Hence, care must be taken when using the different measures for the attack. If not normalized by the centroid frequency, Grey's (1977) and Iverson & Krumhansl's (1993) high-frequency energy and attack centroid may be influenced by a high mean centroid of the harmonics, and if normalized, by high-frequency noise energy present

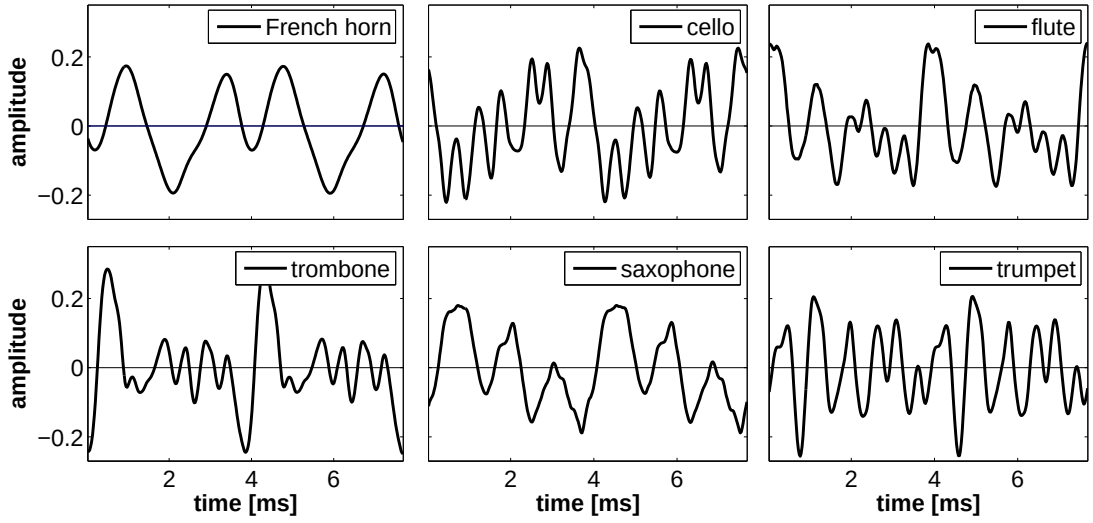
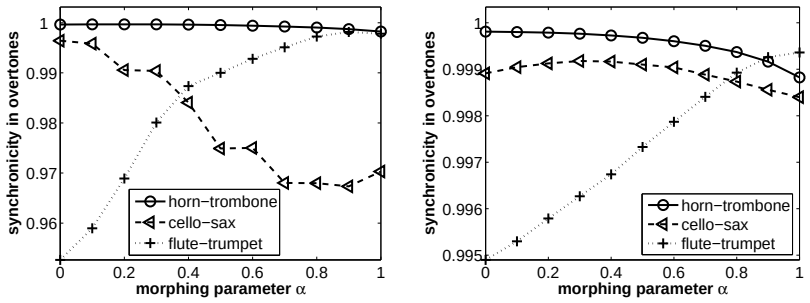


Figure A.4: Temporal wave form. Sound pressure of two cycles for instruments with α=0 (upper panels) and α=1 (lower panels) in the horn-trombone (left), cello-sax (center) and flute-trumpet (right) continua.

throughout the entire stimulus. Grey & Gordon's (1978) definition of overtone synchronicity is dependent on the frequency region being analyzed and the window size being correlated (see also next section); hence, it may be influenced by a dominant fundamental partial or by a high noise content. The log-rise-time of Krimphoff et al. (1994) is difficult to determine due to the unclear end of the attack, and this measure does not necessarily correlate with the other measures. Although the attack doubtless influences timbre perception, none of the common attack measures seems to be optimal; none is independent from the other timbre measures and applicable to all instruments. In Chapter 2, all 4 measures will therefore be correlated with the measurement results to find the descriptors that give evidence to rating cues. However, dependencies across timbre descriptors will carefully be considered for the attack measures that show high correlation with subjective data.

A.3 Spectral flux or overtone synchronicity

The spectral distribution together with the phase correlation of the partials is determined by the temporal wave form of the sound, or the wave fine structure on a time scale up to one cycle of the fundamental frequency (here F0≈262 Hz corresponding to a cycle of 4 ms). The above mentioned clarinet sound, which lacks even-numbered partials, shows a waveform with a square-like shape. The waveforms



(a) entire stimulus (b) stationary portion

Figure A.5: Synchronicity of the upper overtones, i.e., "inverse spectral flux" of 4th to 19th harmonic partials, for the entire stimulus (a) and for the quasi-stationary segment (b).

of the instruments in the present study are shown in Figure A.4. The cello shows the typical sawtooth wave form of strings, the dull French horn sound does not show high-frequency fluctuation, and the sounds of flute and trumpet show quite irregular shapes.

On the other hand, spectral flux is a measure on a larger time scale and visible in the wave form *change* over several cycles as well as in the fluctuation of the temporal envelope. Spectral flux can be measured by correlating the spectral distribution in adjacent time windows (Krimphoff et al., 1994). Grey (1977) called this dimension "synchronicity in overtones", while McAdams et al. (1999) used the temporal deviation of the spectral centroid as a measure of flux. Low spectral flux is a reflection of high overtone synchronicity and high spectral correlation of adjacent time windows. This often leads to high fluctuation of both spectral centroid and temporal envelope. On the other hand, asynchronous overtones produce high spectral flux and low spectral correlation.

In the present study spectral flux is estimated by correlating the spectra of adjacent time windows. Here, the synchronicity in overtones (OS), or the inverse spectral flux, is defined as the average of the Pearson product between amplitude spectra in adjacent time windows of 46 ms length.[44] Since the synchronicity in *overtones* shall be quantified, Krimphoff's (1994) relation will only be applied to the spectra of the sound's harmonics[45]:

$$OS = \frac{1}{N} \cdot \sum_{t=2}^{N} |r(A_{t-1}(k), A_t(k))|, \tag{A.3}$$

where $A_t(k)$ is the square-root of the power density of partial k in the time window t, N is the total number of adjacent windows along the stimulus, and r is the Pearson product (Equation 5.3, p.68). Since synchronicity of the *upper* harmonics was found to be relevant (Grey, 1977), only partials with k>3 are used. Above the 19th partial, partial energy cannot be sufficiently separated from the high noise energy in some

(a) flute (b) saxophone

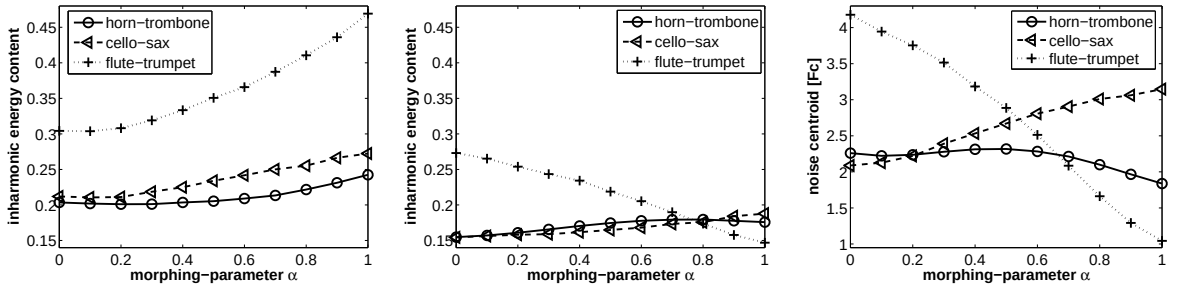
Figure A.6: Amplitude line spectrogram for the first 6 harmonics in the flute (a) and saxophone (b) sounds. Displayed are the envelopes of the harmonics normalized to the respective harmonic maxima. Thick lines indicate partials which dominate the mean spectrum, or those, with the higher mean amplitude.

instruments. Hence the partials between $4 \leq k \leq 19$, or the spectrum between 840 and 5000 Hz, are used to calculate OS.

To test the degree to which the spectral flux varies along the instrument continua, OS is calculated for all stimuli using Equation A.3 (Figure A.5(a)). Due to the increased dynamics during the attack, mean spectral flux may be dominated by the attack segment (compare Figure A.3(e) to Figure A.5(a)). Therefore, OS is additionally calculated using only the stationary portion (Figure A.5(b)) of the sounds. Note that the range of the ordinate in Figure A.3(e) is about 50 times larger than that in Figure A.5(b); saxophone and flute show a distinctly higher flux during the attack segment than during the stationary portion.

OS of the stationary portion increases along the flute-trumpet continuum (Figure A.5(b)), and hence, spectral flux decreases from flute to trumpet. Spectral flux of the flute and spectral flux variation in the flute-trumpet continuum are distinctly higher than in the other two continua. The absence of overtone synchronicity in the flute is illustrated in Figure A.6(a), showing normalized amplitude fluctuation for the lowest 6 partials. During the attack, spectral flux in the flute is also high, but that in the saxophone is even higher, and the cello-saxophone continuum has the highest flux variation of all the continua (Figure A.3(e)). After the attack, the saxophone's harmonics seem to become more synchronous, as Figure A.6(b) shows for the lowest 6 partials. In comparison to the flute (Figure A.6(a)), the saxophone's harmonics fluctuate more synchronously during the stationary portion. However, during the first 300 ms, some asynchronous amplitude dips can be observed in the saxophone (Figure A.6(b)). Both the mean flux and the flux during the stationary portion in the 3 brass instruments (trombone, horn, trumpet) and the cello are low compared to those of the saxophone and flute (Figure A.5).

Spectral flux, which is a spectro-temporal dimension, is difficult to measure,



(a) during attack (b) after attack (c) spectral centroid of noise

Figure A.7: Content of inharmonic energy during attack (first 300 ms) (a) and stationary portion (without first 600 ms) (b), (c) spectral centroid of inharmonic energy as ratio of the harmonics' centroid in the horn-trombone (circles), cello-sax (triangles) and flute-trumpet (crosses) continua.

because any measure is more strongly influenced by analysis parameters such as window length than a pure temporal or spectral dimension like attack rise time, spectral centroid or spectral irregularity. This was also found in the present study (see Endnote[44]). Additionally, the influence of the dominating and varying spectral shape or centroid cannot be segregated from spectral flux analysis: the interference of spectral and temporal dimensions was shown by other studies (e.g. Green, 1988a; Moore et al., 2006). These difficulties may partly explain the controversial results and discussions about spectral flux over the past 30 years, whereas pure temporal dimensions, spectral dimensions or dimensions comparing a spectral parameter across two longer time windows (e.g., comparing Fc between attack segment and stationary portion as done by Iverson & Krumhansl (1993)) seem to be more stable. Note also that due to higher dynamics during the attack, spectral correlation (Equation A.3) in the attack segment is generally distinctly lower than during the stationary portion. Hence calculation of synchronicity or spectral flux for the entire stimulus may be dominated by the flux during the attack (compare Figures A.5(a), (b) and (c)) and perceived fluctuation is often dominated by the attack. This may be another reason that previous studies (Iverson & Krumhansl, 1993, e.g.) often found spectral flux to be a minor timbre dimension compared to the attack: only in instruments without a prominent attack could the spectral flux (in the stationary portion) be distinctly perceived. This can be confirmed by comparing results of Chapters 2 and 3: The stimuli that are analyzed here, which include the attack, are used in the measurements in Chapter 2, while for the measurements in Chapter 3 the attack was removed from the stimuli. While in the measurements of Chapter 2 the flux does not seem to be a major discrimination cue in the cello-sax continuum, it seems to dominate this continuum in Chapter 3.

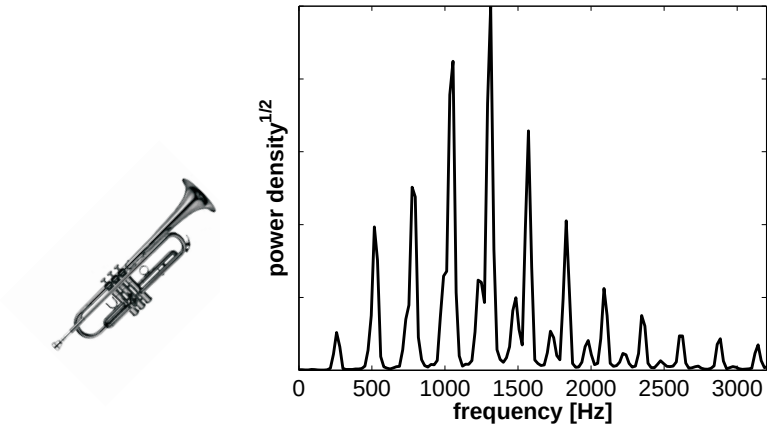


Figure A.8: Amplitude spectrum of trumpet's attack displayed on a linear scale.

A.4 Inharmonic energy

In wind instruments, noise or inharmonic tonal energy may be clearly perceivable, as is the case for the flute (in which noise is perceivable) and trumpet during the attack (inharmonic tonal energy). Figure A.8 depicts the spectrum of the first 300 ms in the trumpet sound, which shows a relatively high inharmonic contribution (compare also to spectrum of entire sound in Figure A.2). Lakatos (2000) and McAdams et al. (1995) described the corresponding inharmonic timbre dimensions as noisiness, pitch strength or harmonic proportion. Figure A.7 shows the noise and inharmonic energy content for the stimuli used in the present study. During the attack (Figure A.7(a)) the trumpet's inharmonic energy is the highest of all the instruments; hence, the trumpet's tonal inharmonic energy exceeds any noise energy present in the other instruments. After the attack, as soon as the tonal inharmonicity disappeares, the inharmonic content in the trumpet is as low as in the other brass instruments (Figure A.7(b)). As mentioned in Section A.2, the noise content in the flute remains high over the entire stimulus (compare Figures A.7(a) and (b)). Figure A.7(c) shows that the flute's noise has a distinctly higher centroid frequency than the harmonics.

Appendix B

Object binding by compression and co-modulation

Abstract

For normal-hearing listeners, compression is a great help by the periphery to bind natural acoustical objects (with an inherent co-modulation) and separate them from other objects. Previous studies showed that compression leads to a great deal of secondary effects, for example, sharp tuning, good temporal resolution, suppression and co-modulation detection difference, which are strongly connected to grouping and discrimination of natural objects. The present chapter shows theoretically and partly hypothetically how compression and the inherent co-modulation in natural sounds lead to an enhanced ability to separate objects even in disadvantageous situations, when the noise level is higher than the object levels. Compression loss, on the other hand, can theoretically explain the reduced ability of hearing-impaired listeners to separate natural objects by means of its secondary effects such as reduced time and frequency resolution. On the other hand, certain timbre dimensions seem to not be degraded by compression loss and might provide hearing-impaired listeners help in separating objects. Hence, lowering the distortion in hearing aids and training hearing-impaired listeners to distinguish timbre may enhance their ability to separate objects.

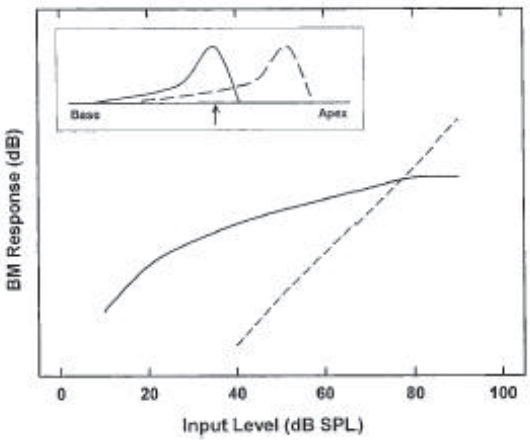


Figure B.1: Basilar membrane input-output function for a tone at CF (10 kHz; solid line) and a tone one octave below CF (5 kHz; dashed line). The inset is a cartoon of the traveling wave envelopes for the two tones. The arrow indicates the measurement location. (From Oxenham & Bacon (2003).)

B.1 Introduction

In the present chapter, timbre discrimination in quiet and noise will be discussed in the context of object separation and cochlear compression. What is the reason for studying timbre in this context? Hearing-impaired people have problems in separating objects, for example, separating the speech of a discussion partner from the loud background noise. Normal-hearing people, on the other hand, are more able to separate objects. They can understand their discussion partner even if he or she is speaking more quietly than the music from a nearby loudspeaker. This excellent performance of separating objects is done by discriminating the object attributes: sound fractions with equal onset, frequency region or direction are assigned to one object. Sound fractions that differ in one or more of these attributes are assigned to different objects. What is responsible for this ability in normal-hearing listeners and (in the worst case) disability in hearing-impaired listeners? In other words, why do normal-hearing listeners show the so-called "cocktail party effect", while hearing-impaired listeners do not? Hearing-impaired listeners show deficits in various discrimination tasks and low resolution in various acoustical parameters, for example, time, localization and frequency. What are the primary and secondary factors for these deficits in hearing-impaired listeners?

In contrast to the normal-hearing listeners' good object separation abilities and hearing-impaired listeners' deficits, the preceding chapters showed that normal-hearing listeners cannot optimally segregate noise from objects using timbre as a discrimination cue. Timbre discrimination ability in noise is also poorer than in quiet for normal-hearing listeners, despite a high signal-to-noise ratio (SNR) of +10 dB. In comparison, speech intelligibility tests produce a score of 100% for SNR down to 0 dB.[21] Furthermore it was shown that hearing-impaired listeners with flat hearing loss show timbre discrimination thresholds similar to normal-hearing listeners,

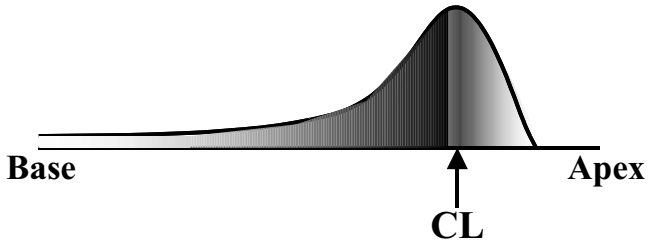


Figure B.2: Schematic basilar membrane excitation pattern for a pure tone with characteristic frequency CF at characteristic location CL.

if the hearing loss was compensated by a linear, hearing loss adequate amplification of sound level. In contrast, thresholds of many other object attributes such as frequency discrimination or localization are higher in hearing-impaired listeners even if compensated by sound level.[46] Thus, hearing-impaired listeners are better in timbre discrimination than expected, whereas normal-hearing listeners are worse in timbre segregation than in segregation of other attributes. What is the reason that normal-hearing listeners are able to cope with other segregation tasks better, and why are hearing-impaired listeners worse at handling other object separation tasks?

Common answers to these questions and established explanations for the deficits of hearing-impaired listeners in object separation are sound attenuation, cochlear compression loss, broader frequency bands per se, higher internal noise, cortical deficits, and deficits in binaural coupling. Of these deficits, only the first two can be verified physiologically as primary factors: Loss of inner hair cells (IHC) leads to sound attenuation, that is higher hearing thresholds, and loss of outer hair cells (OHC) leads to compression loss. Most psychoacoustic thresholds were measured at various levels and amplified for hearing-impaired listeners, that is especially compensated for the IHC attenuation, which nevertheless resulted in higher thresholds in hearing-impaired listeners than in normal-hearing listeners. The present chapter discusses the secondary effects of OHC loss and points out how much cochlear compression loss can account for hearing-impaired listeners' deficits. Can peripheral compression loss as a primary factor explain the poorer object separation in hearing-impaired listeners? And is it not a paradox that an (object) non-linearity (in normal-hearing listeners) works better for object separation than a linear "hifi" system?

B.2 Non-linearity on the basilar membrane

The basilar membrane (BM) shows two physiologically and psychoacoustically measured features which are dependent on the presence of a hearing loss:

1a. On-frequency[47] dynamics in normal-hearing listeners: Due to healthy outer hair cells, incoming sound levels get transferred non-linearly (compressed) at the characteristic location[48] (CL) for intermediate levels. In other words, for a pure-tone excitation, the BM response intensity (BMR) at the CL behaves compressively to the incoming level (L):

$$BMR_f(L) = \text{compressive, for approximately } 20\ dB < L < 80\ dB.$$

As observed in physiological and psychophysical measurements, the input-output function of the cochlea for on-frequency stimulation shows a linear growth for very low and high levels and a compressive increase for levels between 20 and 80 dB (Figure B.1).

1b. On-frequency dynamics in hearing-impaired listeners: On the other hand, a complete loss of outer hair cells in hearing-impaired listeners would result in a linear BM input-output function for all levels. That is to say, for a pure-tone excitation, BMR at CL behaves linearly to the incoming level (L) for all sound levels:

$$BMR_f(L) = \text{linear, for } 0\ dB < L < 120\ dB.$$

2. Off-frequency[49] dynamics: Due to the non-linear interconnection between BM frequency regions, a single frequency excites the entire BM and produces a non-linear location-excitation pattern along the BM (Figure B.2). In other words, for off-frequency excitation with a pure-tone, the BM response intensity is related non-linearly to characteristic frequency[47] (CF) distance:

$$BMR_L(\Delta f) = \text{non} - \text{linear}.$$

Because of the non-linear excitation pattern along the BM, the basilar membrane shows different behaviour for on- and off-frequency excitation (Figure B.1). Instead of the non-linear (compressive) BM response "on frequency", stationary off-frequency excitation shows a linear BM response for signal frequencies far from the CF.[50]

For normal-hearing listeners, these characteristics interact in an advantageous way such that object segregation becomes possible and enhanced by the periphery; this will be described in detail in this section.

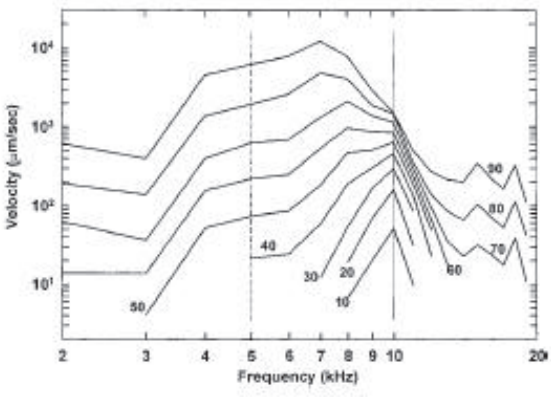


Figure B.3: The response of the chinchilla BM at a CF of 10 kHz in response to a fixed-level tone with a frequency represented along the abscissa. The level of the tone varied from 10 to 90 dB SPL. Vertical lines mark the responses to tones at 5 and 10 kHz. (From Oxenham & Plack (1997) with data from Ruggero et al. (1997).)

B.2.1 Frequency selectivity and temporal resolution

As Oxenham & Bacon (2003) very nicely described with resumed studies of psychoacoustic and physiological measurements with normal-hearing and hearing-impaired listeners, BM compression or the loss of compression, respectively, can account for many results observed in these measurements, in particular for the differences between the listener groups.

Figure B.3 shows the response at one point along the chinchilla BM to tones of various frequencies (Ruggero et al., 1997). It demonstrates the two features of a healthy BM, which were described above (p. 103). The response growth at the CF of 10 kHz (on-frequency) is highly compressive. For on-frequency excitation, low signal levels of up to 20-30 dB SPL cause a relatively high excitation, whereas the excitation increases little at high signal levels, for example from 70 to 80 dB SPL. On the other hand, the response growth is roughly linear at CF of 5 kHz (off-frequency). Low levels of up to 30 dB SPL do not cause any measurable excitation effect, whereas excitation increases rapidly and linearly from 40 to 90 dB SPL. At high levels the excitation difference between the on- and off-frequency conditions is very low, which results in a broad response region at high levels (see the upper-most curve in Figure B.3). At low levels the excitation difference between on- and off-frequency conditions is higher than at high levels due to the compression in the on-frequency condition. This leads to a sharp tuning at lower levels, which is visible in Figure B.3 at the lower-most curves around 10 kHz signal frequency. A BM compression loss decreases the difference between off- and on-frequency excitation at low and intermediate levels. The tuning found in damaged cochleas often resembles the broad tuning found at the highest sound levels in the normal cochlea (see the upper-most curve in Figure B.3). Thus, it is likely that the poorer frequency selectivity found in hearing-impaired listeners is caused by broader cochlear tuning.

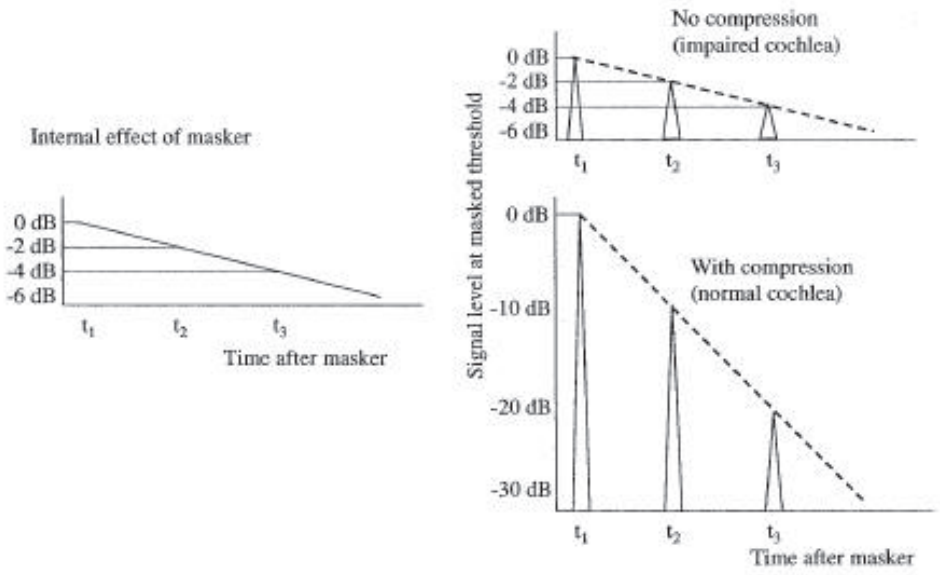


Figure B.4: Forward masking (in the on-frequency condition). The left panel shows how the internal representation of the masker decays after the masker has been turned off. The upper and lower right panels show the changes in signal levels needed to match the decay of masker excitation for impaired and healthy cochleas, respectively. In the impaired case (upper panel), the signal is processed linearly and so a 2 dB change in masker excitation is matched by a 2 dB change in signal level. In the normal case (lower panel), the signal is compressed and so a 10 dB change in signal level is required to match the 2 dB change in masker excitation. (From Oxenham & Bacon (2003).)

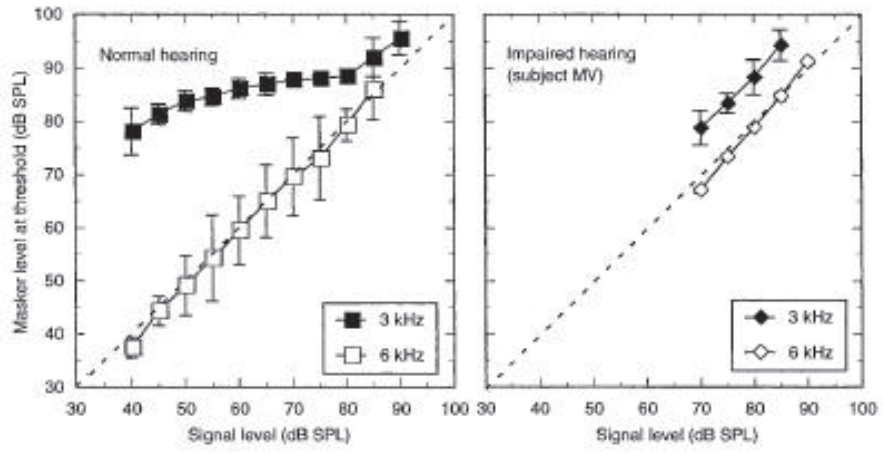


Figure B.5: Forward masking difference between on- and off-frequency masker. The level of a masker required to mask the 6 kHz signal as a function of signal level for normal-hearing (left) and hearing-impaired listeners (right). Dashed lines denote linear growth of masking. (From Oxenham & Plack (1997).)

A compressional or non-compressional input-output function has also consequences on temporal resolution. The temporal decay of a signal (namely the decay of the signal's internal representation after the signal has been turned off) is linear and independent of frequency, level and BM compression. That is to say, the internal representation of a signal decays by a constant rate $\Delta IR/\Delta t$ in dB/s (Oxenham & Bacon, 2003). If the BM input-output is compressive, an intensity difference of the internal representation of, for example, $\Delta IR = 2\ dB$ is matched by a bigger external level difference (e.g. $\Delta L = 10\ dB$) of the signal than in the non-compressive case (e.g. $\Delta L = 2\ dB$). Hence, compression reduces the forward-masking effect in normal-hearing listeners compared to hearing-impaired listeners (Figure B.4), which implies poorer temporal discrimination in hearing-impaired listeners.[51]

In forward-masking experiments, compression also leads to different masker-to-signal-level ratios at threshold for on-frequency maskers compared to off-frequency maskers. In the on-frequency condition in normal-hearing listeners, the masker and signal are *both* processed by the same amount of compression. Hence, a 10 dB change in signal level requires a 10 dB change in masker level (although it is matched by a 2 dB change of internal representation). In forward-masking experiments with an off-frequency masker, in normal-hearing listeners the signal is compressed, while the masker with a frequency far from signal's frequency is processed approximately linearly (at the signal's characteristic frequency on the BM). Hence, a 10 dB change in signal level matches a 2 dB change of internal representation, which requires only a 2 dB change of external masker level. How linearly the masker is processed at the signal's frequency depends on the frequency distance (between signal and masker) and on masker level. In the on-frequency condition, the level at threshold increases approximately linearly with signal level (Oxenham & Plack, 1997). For an off-frequency masker, on the other hand, the signal-to-masker level at threshold shows a strongly non-linear growth in normal-hearing listeners (Figure B.5).

As shown in this section, the BM input-output functions of normal-hearing listeners are able to account for the sharp frequency selectivity and good temporal resolution in forward masking experiments. On the other hand, loss of compression can explain not only the higher hearing thresholds and recruitment in loudness growth, but also both the worse frequency selectivity and temporal discrimination in forward masking experiments.

B.2.2 Suppression

An obvious consequence of a non-linear compressive dynamic and non-linear excitation pattern along the BM is that the presence of one sound can influence the

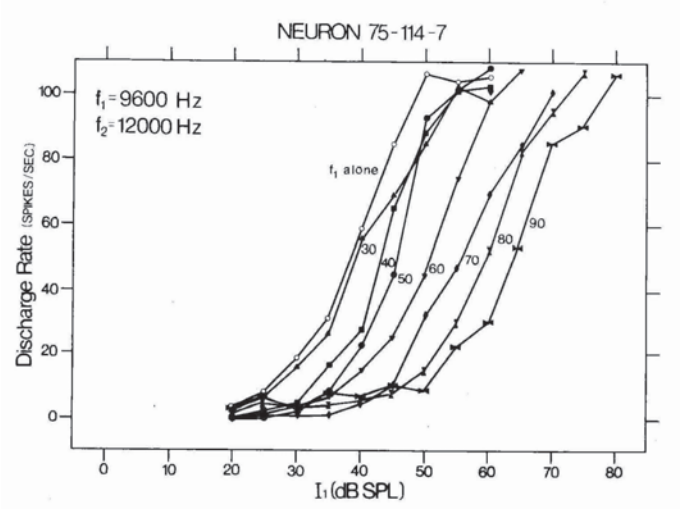


Figure B.6: Two-tone suppression in auditory nerve fibers. A rate versus level function shifts as a result of 2-tone suppression. f1=CF=9.6 kHz; f2=12 kHz. Numbers adjacent to the f1+f2 rate-intensity functions (filled symbols) indicate the values of I2 used to obtain the functions. (From Javel et al. (1978).)

physiological response to another in a non-linear way. One class of this phenomenon is known as "two-tone suppression": the response to one signal can be reduced by the addition of a second suppressor sound (Oxenham & Bacon, 2004). This effect is shown in Figure B.6 with physiological data by Javel et al. (1978).

Non-linear processing can lead to effects that sound paradoxical. The non-linear on-frequency dynamic combined with the non-linear frequency connection in the cochlea (Section B.2) leads to non-linear interconnection of physical, BM-physiological and perceptual space. Hence, addition in physical space does not imply addition in perceptual space. In the following I want to list some ideas that may help to solve this paradox[52]:

- A subject cannot distinguish whether a frequency region on the BM is excited or not, if excitation is below detection threshold. For example, a pure tone excites off-frequency regions on the BM, while only the pure-tone frequency is perceived.

- Pure tones with equal frequency are perceived at equal tone height independent from signal level, while the location of maximal excitation on the BM varies with level. Excitation of a tone does not spread homogenously around the characteristic frequency on the BM, but is sharply tuned at the characteristic frequency for low levels and more broadly for higher levels.

- An off-frequency tone added to a signal tone changes the excitation pattern around the characteristic location in a different way than an on-frequency tone. Hence, by adding an off-frequency tone, the excitation caused by the signal becomes spread over a broader region. Hence, the signal's detection location (BM excitation location which results in signal detection by the subject) of a

signal tone alone may be different from conditions in which an off-frequency tone is added or an on-frequency tone is added. Note that even adding a sound with a level below the hearing threshold changes the excitation pattern, amplitude and detection location.

How may peripheral suppression influence object segregation? As a direct effect of the instantaneous compression in the cochlea, the suppression is almost instantaneous; it even varies periodically within the individual cycles of sufficiently low-frequency suppressor tones, and, hence, with the modulation cycles of amplitude-modulated suppressor tones (Cooper, 2004). Shannon (1976) studied this phenomenon in a psychoacoustic forward-masking experiment. The masker contained two frequency components with one at the same frequency as the signal and the other at a variable frequency. For certain frequencies and intensities of the variable masker component, the threshold of the signal was lower than if that component were not present, in both the simultaneous and successive forward-masking situations. In other words, the variable off-frequency masker with the same temporal envelope as the on-frequency masker increases the probability for the signal to be detected. Natural objects show inherent synchronous amplitude fluctuations of different frequency regions. In the same way as in the above-mentioned psychoacoustic measurement, the different frequency regions "suppress" each other. Specifically, suppression reduces masking effects onto other objects, which show amplitude fluctuations that are not correlated to the first object. This effect may be related to the Comodulation Detection Difference (Section B.2.3). Two-tone suppression only appears to affect probe tones that undergo amplification and compression on the basilar membrane (Cooper, 2004). Hence, hearing-impaired listeners lack any advantage that normal-hearing listeners get from suppression.

B.2.3 Co-modulation

Co-modulation (that is, two or more frequency bands with equal temporal envelope) is a common phenomenon observed in natural stimuli. All natural sounds show a broadband spectrum, often with different spectral regions fluctuating synchronously (or nearly synchronously) in intensity. Co-modulation was shown to reduce masking effects and enhance detectability of a signal by Comodulation Masking Release (CMR) (Hall et al., 1984; Verhey et al., 2003) or Comodulation Detection Difference (CDD) (McFadden, 1987). Recently it has been discussed that parts of the effect can be explained in terms of suppression (Section B.2.2), and, hence, depending on the degree of compression present in the cochlea (Ernst & Verhey, 2006, submitted; Buschermöhle et al., 2006, accepted).

Considering the effects described in the previous sections, the reduced masking effect of a co-modulated masker or the increased "object binding" effect of a co-modulated signal in a compressive system may be qualitatively explained as follows. Any signal at one frequency excites the whole BM and the excitation along the BM shows the same amplitude modulation on all locations. If a masker fluctuates with the same amplitude modulation as the signal, at any frequency location along the BM the amplitude maxima of the masker coincide temporally with the signal's maxima. Hence, the more correlated masker and signal are, the higher is the variance of excitation fluctuation at a given BM location. As a statistical effect, the temporal mean of the fluctuation depends on the correlation of the signals if the system has a compressive dynamic. The temporal mean is higher in the uncorrelated case than in the correlated, which is the reason for the CDD effect (Buschermöhle et al., 2006, accepted): The ability to detect an amplitude-modulated narrow-band sound signal in the presence of one or several masking noise bands is best if all maskers share the same time course of amplitude modulation while the signal band's envelope fluctuates independently. If using a simple one-band energy model, a signal is detectable within a masker if the temporal integration of the Hilbert envelope changes by the JND or more when adding the signal to the masker. In other words the CDD effect can be described by

$$\int_t M_c(t) + S_u(t)dt - \int_t M_c(t)dt \quad > \quad \int_t M_c(t) + S_c(t)dt - \int_t M_c(t)dt, \qquad \text{(B.1)}$$

where t is the temporal envelope of the correlated masker bands M_c, S_c is the signal band correlated to the masker, and S_u is the signal band uncorrelated to the masker. Buschermöhle et al. (2006, accepted) showed that this model can explain the observed results only when the temporal envelope is compressed prior to temporal integration. Hence, the perceived difference between masker and signal decreases with increasing correlation of amplitude modulations, if the sounds are processed compressively.

B.3 Grouping

As depicted above, compression leads to various secondary effects, such as coarser intensity resolution, sharp tuning, good temporal resolution, suppression effects and higher separability of co-modulated from uncorrelated sounds. Compression loss, on the other hand, directly leads to reduction of these features. But how is this connected to object separation? Central grouping, which is the central mechanism that binds sound parts to one object, uses the

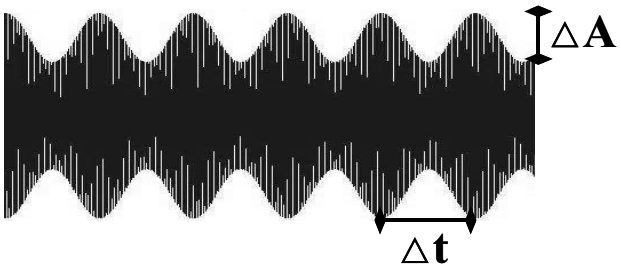


Figure B.7: Time and intensity resolution are connected to amplitude modulation detection.

- time domain (e.g. for object separation by onset, length, localization)
- frequency domain (e.g. by pitch, harmonic structure)
- intensity (e.g. by loudness, localization)

The concurring demands between time-, frequency- and intensity-resolution decide whether object separation improves by compression. Discriminability of object attributes that are only connected to the time and/or frequency domain should theoretically improve by compression, because both time and frequency resolution improve by compression. Hence, object separation by onset and length (both connected to time resolution) or by pitch (connected to frequency and time resolution) should improve by cochlear compression and degrade by compression loss. On the other hand, if a certain object attribute is connected to intensity and time or frequency domains, relative importance of the concurring domains for the attribute decide whether the attribute improves. An example for such an object attribute is amplitude modulation, for which intensity and time resolution are of importance (Figure B.7). Depending on modulation frequency and modulation degree, intensity or time will be more or less important for modulation detectability. At first glance, it is unclear whether the trade-off between (theoretically worse) intensity and (better) time resolution of the compressive cochlea (compared to a linear system) results in improved detectability. Only psychoacoustic measurements can reveal in which situations hearing-impaired subjects have worse modulation detectability than normal-hearing subjects, which also depends on the sound level amplification applied for the hearing-impaired subjects. In the same way as for modulation detection, localization is connected to time and intensity resolution. Here, apparently the trade-off is positive for compression, that is, after a compression loss, localization degrades. On the other hand, the timbre mentioned above and in previous chapters does not seem necessarily to degrade by compression loss. For example, for the timbre dimension “brightness”, intensity and frequency resolution are the important domains. In subjects with a flat hearing loss, equal intensity resolution seems to play the major role. The reduced frequency resolution seems to play a minor role (or may even be

counteracted by an increased perceptual intensity difference), so that a compression loss does not significantly affect brightness discriminability.

B.4 Summary and discussion

In a healthy cochlea, the amount of dynamic compression applied to the incoming sound at one characteristic location on the BM depends on the characteristic frequency and the frequency content of the sound. As depicted above, this compression is responsible for a series of secondary effects in normal-hearing listeners, such as:

- ✓ **Good temporal resolution**: Temporal separability of non-simultaneous acoustical signals or objects improves by compression.
- ✓ **Sharp tuning**: Frequency resolution and spectral separability of acoustical objects improve by compression.
- ✓ **Suppression and co-modulation detection difference**: Compression may reduce masking effects of different objects.

All these effects help normal-hearing listeners to separate sound objects. In particular, separability of natural objects, which show a broad band spectrum and comprise identical amplitude fluctuations of different frequency regions (e.g., co-modulation of the overtones in speech), improves by compression. Even in disadvantageous situations like cocktail parties, a dialog partner can be understood when the surrounding noise level is higher than his or her voice. A loss of compression directly results in reduced temporal discrimination, reduced frequency resolution, reduced suppression effects and a reduced advantage of binding co-modulated objects. The increase of perceptual intensity differences, which is caused by compression loss, may lead to additional disadvantageous side effects. For example, perceived variance of (inherent) fluctuation of noise may increase due to compression loss (Oxenham & Bacon, 2003). Hence, while the compressive properties of the cochlea may be responsible for the good object segregation abilities observed in normal-hearing listeners, compression *loss* may explain the reduced ability to discriminate objects in hearing-impaired listeners.

Separating natural sounds, which comprise co-modulation, seem to be strongly connected to the above-depicted secondary effects of compression. Hence, hearing-impaired listeners show a reduced ability to discriminate objects compared to normal-hearing listeners, even if sounds are amplified by hearing aids. On the other hand, discriminability of certain timbre dimensions does not seem to be degraded by

reduced compression. Hearing-impaired listeners seem to be able to take advantage of the non-degraded intensity resolution, whereas the reduced abilities (e.g., frequency selectivity for distinguishing brightness) play a minor role.[53] The potential of using timbre as an object separation cue does not seem to be fully exploited by non-musicians and may be improved by actively playing music or by psychoacoustic training (Chapter 3). Hence, training hearing-impaired listeners to listen more to the timbre of a sound and to use it as an object cue may help hearing-impaired listeners to separate objects.

Unfortunately, special features in modern hearing aids, for example noise-reduction algorithms, inevitably distort the timbre of a sound. These features are doubtless necessary for intelligibility enhancement.[54] However, the optimal compromise should be carefully sought between the advantages of these features and the naturalness of timbre.[55] Timbre may not only be useful for object segregation, but it is definitely a ❁*beautiful*❁ sound feature.

Appendix C

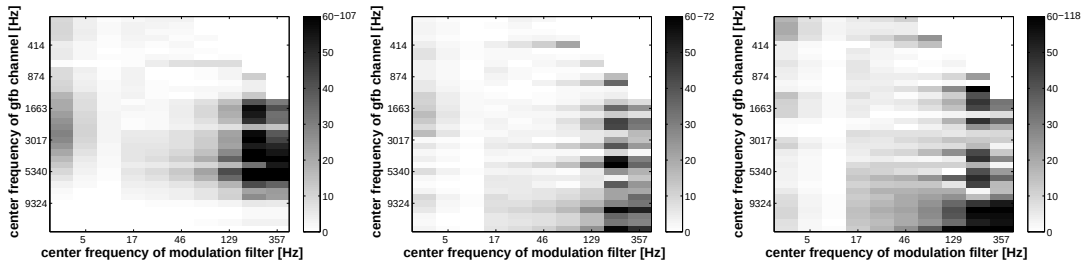
Internal representations

This appendix illustrates internal representations (IR) of the simulations in Chapter 5.

Figure C.1 shows the IR difference between stimuli with α=0 and α=1 in the similarity rating measurements (Chapter 2), i.e. ΔIR of horn and trombone, cello and saxophone, and flute and trumpet, respectively. Before calculating the difference, IRs were reduced to one time step (Figures C.1(a)-(c)), to the temporal attack mean (Figures C.1(d)-(f)), and to the mean of the stationary portion (Figures C.1(g)-(i)). Hence, Figure C.1 shows the energy (difference) distribution along frequency and modulation frequency channels. Note the increasing energy with increasing modulation frequency up to modulation filter with center frequency of approximately 262 Hz, which is the fundamental frequency of the stimuli. This shows, how the "temporal pitch" is represented in the IRs. The high energy in the upper modulation channels dominate the stimulus distance even in the spectral-dominated trombone-horn continuum.

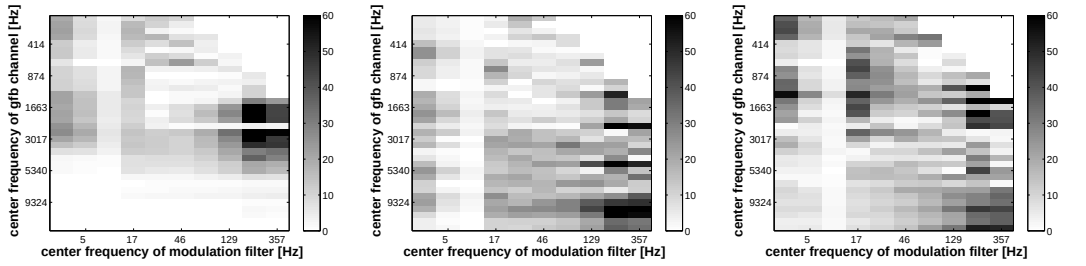
Figure C.2 shows IR differences for 3 hearing-impaired subjects, namely with a higher presentation level and a preprocessing with an additional attenuation and expansion stage according to the individual audiogram.

Figure C.3 shows IR difference distribution along frequency for the JND measurements in quiet (Experiment A of Chapter 4). In contrast to the stimuli in Figure C.2, the natural initial attack was removed from the stimuli. Note that for normal-hearing subjects IRs differ distinctly at frequencies above 1-2 kHz, both with and without modulation filters. Hence, for timbre discrimination, frequencies above 1-2 kHz seem to be crucial.



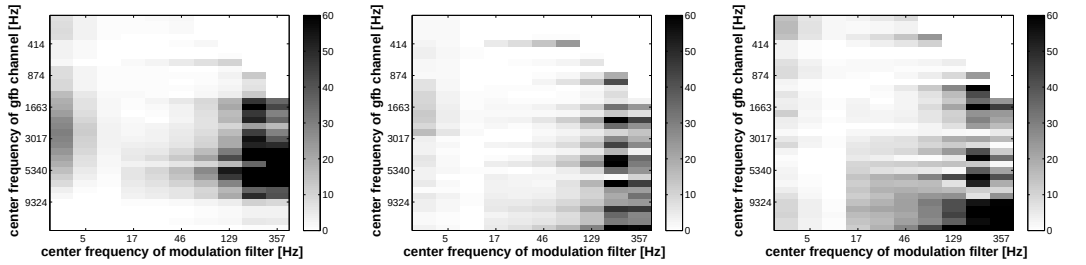
(a) horn-trombone (b) cello-sax (c) flute-trumpet

Only attack, i.e. first 300 ms:



(d) horn-trombone (e) cello-sax (f) flute-trumpet

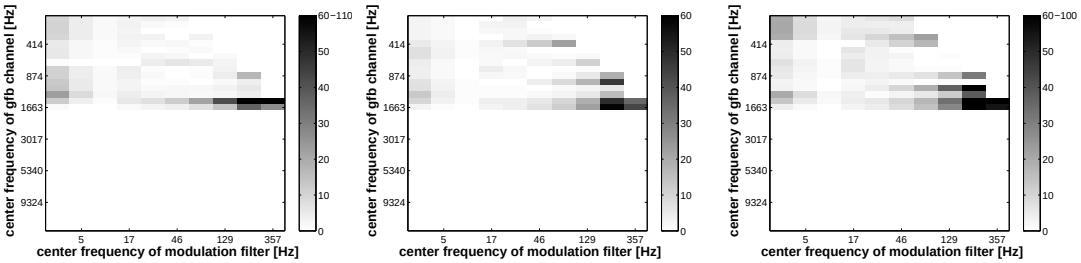
Without attack, i.e. 0.6-1.5 s:



(g) horn-trombone (h) cello-sax (i) flute-trumpet

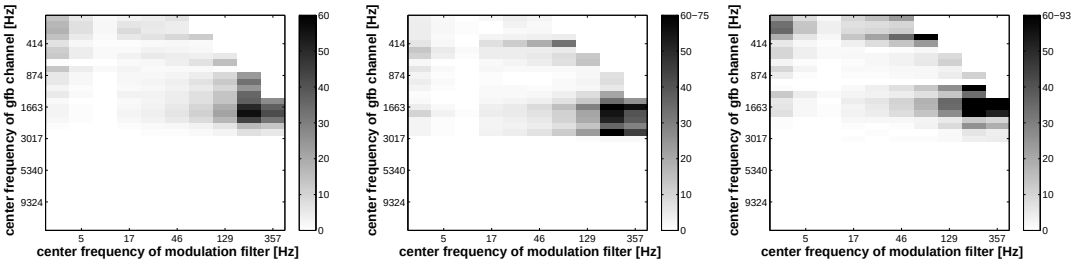
Figure C.1: Internal representation difference ΔIR of stimuli with α=0 and 1, that is ΔIR between horn and trombone, cello and sax, and flute and trumpet, respectively. For (a)-(c) the entire stimulus was analyzed, whereas for (d)-(f) only the first 300 ms of the respective stimulus were analyzed and for (g)-(i) only the stationary portion of IR was used.

Hearing-impaired subject 'iDL' (steep hearing loss, stimulus level = 80 dB):



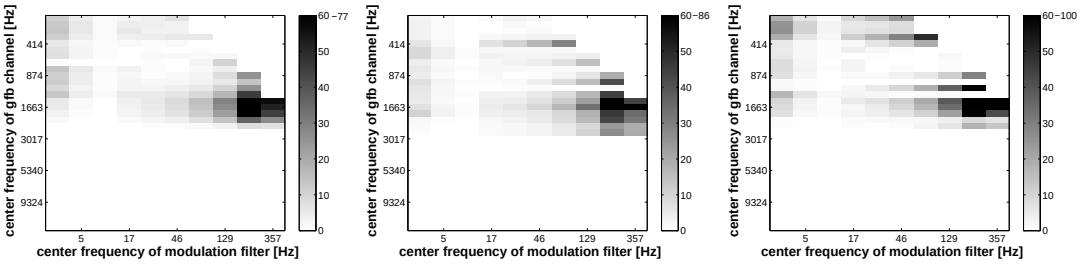
(a) horn-trombone (b) cello-sax (c) flute-trumpet

Hearing-impaired subject 'iGM' (steep hearing loss, stimulus level = 95 dB):



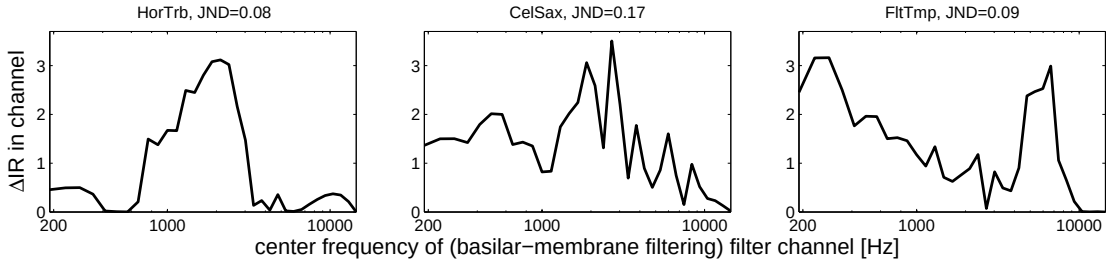
(d) horn-trombone (e) cello-sax (f) flute-trumpet

Hearing-impaired subject 'iFL' (flat hearing loss, stimulus level = 80 dB):

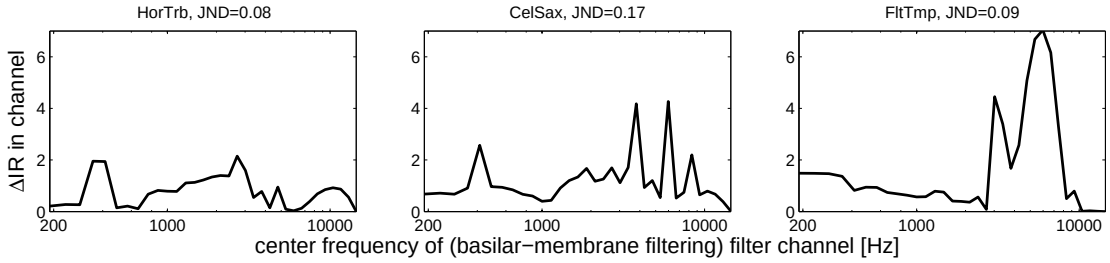


(g) horn-trombone (h) cello-sax (i) flute-trumpet

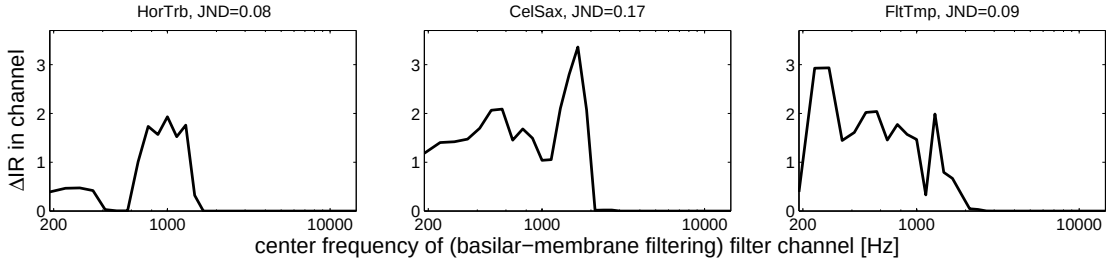
Figure C.2: ΔIR for three hearing impairments. That is, same processing as for Figure C.1(a)-(c) but with a 15-30 dB higher stimulus level and an additional attenuation and expansion model stage according to subject's audiogram. (a)-(c) show IR for subject 'iDL' whose hearing threshold increases steeply from 30 dB HL at 1 kHz to 75 dB at 2-3 kHz.(d)-(f) show IR for subject 'iGM' showing a hearing loss of 75 dB at 1 kHz and above.(g)-(i) show IR for subject 'iFL' showing a diagonal to flat hearing loss and a hearing threshold of 55-60 dB between 1-8 kHz. Note that for better comparison, the colour-bar was not scaled, that is, equal grey-colour indicate equal ΔIR values for all subjects, but for some hearing-impaired subjects black may also indicate higher ΔIR values than 60 as indicated by colour bars.



(a) with 8 Hz low-pass filter



(b) with 6 modulation filters



(c) hearing-impaired 'iDL'

Figure C.3: Energy distribution of IR differences ΔIR along frequency between stimuli at normal-hearing subjects' mean threshold (α_{ref}=0, Experiment A of Chapter 4). IR was reduced to one time step before calculating ΔIR. Preprocessing was done without modulation filter bank (a,c) and with 6 modulation filters (b). For (b) IR *differences* were averaged quadratic (i.e. by Euclidean mean) across modulation channels. Preprocessing for (c) contained the hearing-impairment stage of attenuation and expansion according to the subject's pure-tone threshold, which increases steeply from 30 dB HL at 1 kHz to 75 dB at 2-3 kHz.

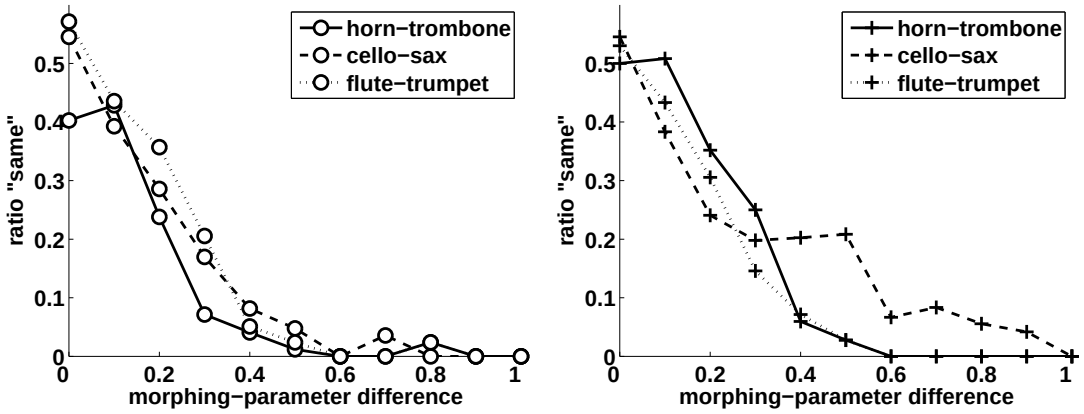
Appendix D

Notes, hypotheses and blabla

[1]This is an endnote.

[2]The physical dimensions are defined by certain signal analysis tools which make the decision for the "real" dimension even more problematic. For example, for the Fourier and Hilbert transformation, the time-vs-frequency trade-off make decisions about the window settings difficult. To compensate for the physiological periphery and logarithmic perception, sound may be preprocessed by spectral weighting (A,B,C) or complex auditory models, and amplitude and frequency can be calculated on a linear or logarithmic scale. For some dimensions only the harmonic, noise or high-frequency part of the sound can be of relevance, instead of using the full spectrum. Since the peripheral preprocessing and timbre perception is strongly dependent on the spectro-temporal content in the sound and the psychoacoustic task, decisions for the analytical tools cannot easily be taken from basic psychoacoustic measurements with pure tones and artificial complex tones.

[3] **Normal-hearing subjects' same/different results:** In the measurements, the rating scale used non-metric rating words ("very similar", "similar",...) that are not "calibrated" and probably not equi-distant on a perceptive scale. This is probably one reason that multidimensional scaling (not shown) elicited no results for the low number of subjects. In order to perform an ANOVA analysis with a sufficient high number of data entries, the ratings were redefined and categorized into dual-ratings "same" (previous similarity rating 1) and "different" (previous similarity ratings 2-8). The dual-ratings were then averaged over all normal-hearing listeners. Again, stimulus order was tested by Wilcoxon rank sum test, which showed no significant differences ($p>0.05$). A 2-way ANOVA was applied with the factors "absolute morphing-parameter distance $\Delta\alpha$ " (>0, i.e. morphing distance between stimuli) and "morphing-parameter α of first stimulus". In all instrument continua, dependency on absolute morphing-parameter distance $\Delta\alpha$ was highly significant ($p<0.001$). In contrast to the ANOVA with similarity ratings (Section 2.3.3), here only the cello-sax continuum elicited significant differences ($p<0.05$) in morphing-parameter α of the first stimulus, which will be further analyzed in Section 2.3.5. However, since the results depend mainly on the absolute morphing-parameter distance $\Delta\alpha$, the role of both order and morphing-



(a) normal-hearing (b) hearing-impaired

Figure D.1: Same/different ratings of 7 normal-hearing (crosses) and 6 hearing-impaired (circles) subjects, in the three instrument continua horn-trombone (solid line), cello-sax (dashed line) and flute-trumpet (dotted line). Abscissa indicates absolute values of morphing-parameter distance $\Delta\alpha$ of the presented stimulus pairs. The ordinate shows the relative number of responses of "same" across subjects and ratings

parameter α of first stimulus will be neglected for analysis in the following sections. Figure D.1(a) shows the mean same/different ratings of the normal-hearing subjects across absolute morphing-parameter distance $\Delta\alpha$ of the stimulus pair. Since the dual same/different rating corresponds to an inverted "discriminability", this figure can be also seen as a psychometric function. In the cello-sax and flute-trumpet continua, the curves show a smooth, flat, continuous decrease, whereas in the horn-trombone continuum the psychometric function shows a steeper decrease between $\Delta\alpha = 0.1$ and $\Delta\alpha = 0.3$ than in the other continua.

[4]Here the **individual ratings of the hearing-impaired listeners** shall be analyzed in detail. The two hearing-impaired subjects iEW and iUL show ratings that do not differ visibly from the results and spread of normal-hearing subjects. These two subjects show the most moderate hearing loss of all hearing-impaired subjects. iUL's pure tone threshold is normal (20 dB HL) up to frequencies of 2 kHz, and increases for higher frequencies up to 60 dB at 6 kHz. iEW's threshold lies around 35 dB HL for frequencies up to 3-4 kHz, increases for higher frequencies up to 70 dB at 6 kHz. Subject iGM shows in all instrument continua a more convex curve compared to the other subjects; he rated $\Delta\alpha$ between 0.1 and 0.9 with a higher difference than normal-hearing listeners did. iGM shows the highest hearing threshold of all hearing-impaired listeners of around 75 dB HL at 1 kHz and above. This subject chose by far the highest presentation level (95 dB SPL) of all hearing-impaired listeners (80 dB SPL and lower) to perceive a "comfortable loudness". In categorical loudness scaling he rated the highest levels of 100 dB HL at "intermediate". Hence, the high physical amplification of certain stimulus contents and stimulus distortion and/or physiological distortion due to high compression and/or dead regions may lead to a larger difference in perception. Subject iFL shows similar ratings to normal-hearing listeners in the first and third continua, but a convex curve in the second continuum. Curves of subject iGH show "recruitment" behaviour in the first and second continua, that is to say, lower ratings than normal-hearing listeners for low $\Delta\alpha$,

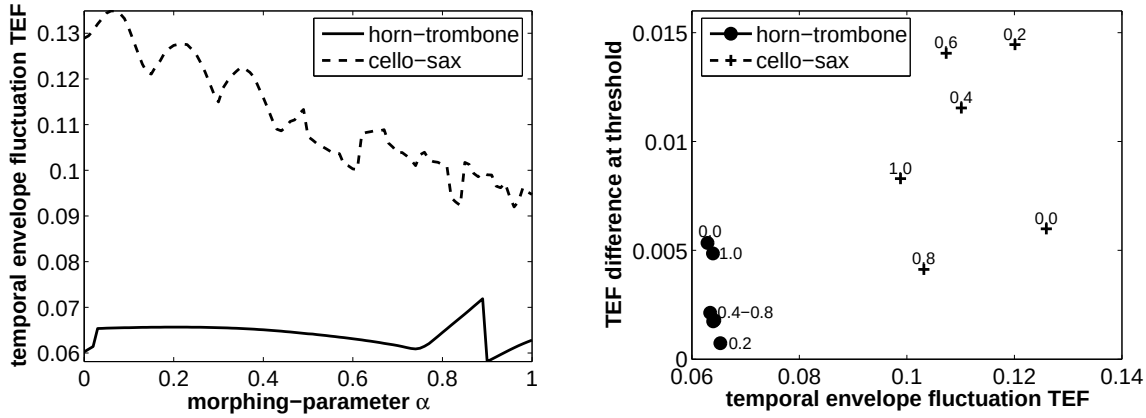
but similar ratings for high $\Delta\alpha$. iDL shows little recruitment in the first continuum and a flat curve in the second continuum. iFL shows a nearly flat hearing loss with flat threshold of 55-60 dB HL at 1 kHz and above. iGH's threshold of 30-40 dB HL is flat from 0.125 to 8 kHz, but shows 20 dB more hearing loss at 0.75-2 kHz. iDL's threshold increases steeply from 30 dB HL at 1 kHz to 75 dB at 2-3 kHz. Hence, subjects iFL, iGH and iDL show moderate hearing losses of different kinds and all chose presentation levels of 80 dB SPL. The various hearing thresholds, that is, various compression loss and attenuation, apparently distort timbre in various ways, also depending on music instrument continuum and, hence, depending on timbre dimension. For instance, the flat similarity curve of iDL in the cello-sax continuum may result from the significant cue in the upper frequencies, in agreement with the high variation in high-frequency energy during the attack found in this continuum (Section A.2). Hence, the high frequencies of the broad-band amplified stimulus at 80 dB SPL might have been inaudible for subject iDL, whereas for subject iGM, who has a similar hearing loss at the high frequencies but listened to the stimuli at a higher level of 95 dB SPL, differences in high-frequency energy may have been above hearing threshold. On the other hand, subject iFL, who does not regularly wear hearing aids, may not be used to high frequencies and may have perceived the varying high-frequency parts as dominating or even disturbing.

[5]**Hearing-impaired subjects' same/different results:** As described in Endnote 3, the similarity ratings were subsequently redefined and categorized to dual-ratings "same" and "different". Figure D.1(b) shows the mean same/different results of the hearing-impaired subjects across morphing-parameter difference $\Delta\alpha$. Similar to the normal-hearing subjects (Figure D.1(a)) the psychometric function of the horn-trombone continuum shows a steeper decrease between $\Delta\alpha$=0.1 and 0.4 than in the other two continua. For normal-hearing subjects, same/different ratings show a steeper "psychometric slope" in the horn-trombone continuum than in the other continua, which may be due to better differentiability of spectral centroid (brightness) than of the other timbre dimensions such as spectral flux, initial attack or noise content. Brightness is the only timbre cue that was found in all studies using tonal instruments. Subjects, in particular non-musicians, seem to be more accustomed to the clear characteristics of brightness than to other timbre dimensions. Brightness is commonly used when distinguishing vowels in speech, hence, it seems easier to identify, classify and memorize than the fuzzy (spectro-) temporal parameters such as spectral flux or initial attack. In the cello-sax continuum, where similarity ratings (Section 2.3.4) show the most differences between hearing-impaired and normal-hearing subjects, the same/different ratings also show a distinctly shallower decline for $\Delta\alpha \geq 0.4$.

[6]iDL's hearing threshold increases steeply from 30 dB HL at 1 kHz to 75 dB at 2-3 kHz. iGH's threshold of 30-40 dB HL is flat from 0.125 to 8 kHz, but shows a 20dB higher hearing loss at 0.75-2 kHz.

[7]The higher number of wrong responses by normal-hearing subjects for stimuli with α=0 compared to hearing-impaired subjects is not an artifact of the smoothing applied for Figure 2.7: the end points remained identical.

[8]The inherent fluctuation of noise may be perceived as stronger due to a compression loss (Appendix B), that is to say, noise content may be perceived as more disturbing by the hearing-



(a) envelope fluctuation vs. morphing-parameter (b) envelope fluctuation difference at threshold

Figure D.2: Relation between temporal envelope fluctuation, morphing-parameter α, and obtained timbre JND values: (a) Temporal envelope fluctuation TEF (standard deviation of temporal envelope) vs. morphing-parameter α, and (b) temporal envelope fluctuation difference ΔTEF (standard deviation difference) vs. temporal envelope fluctuation TEF of stimuli at threshold for the horn-trombone and the cello-sax continua. Numbers indicate the morphing-parameter α_{ref} of the reference stimulus of the respective stimulus pair.

impaired listeners, which increases uncertainty. However, this is only speculative, and cues used by hearing-impaired listeners may be different from those used by normal-hearing subjects.

[9]Note that both measures are, although often correlated, not the same (Appendix A)!

[10]More information on calculating Fc is found in Appendix A.

[11]**Effect of temporal envelope fluctuation:** While the previous paragraphs analyzed the effect of spectral dimension on the results, here the physical time dimension alone will be correlated with the stimuli and JND results. Although neglecting the spectral content seems unphysiological, the first approach to perceived spectro-temporal dimensions is the fluctuation of the temporal envelope of the sound, and hence, a temporal dimension.
To test the degree to which the temporal envelope fluctuation (TEF) varies along the instrument continua, we calculate TEF as the standard deviation of the temporal envelope over the signal duration, that is, peaks of the instantaneous magnitude of the analytical signal within a running time window of 6 ms:

$$TEF = std(\max_{[t_0, t_0+6\text{ms}]}(abs(HT(A(t))))), \tag{D.1}$$

where HT is the Hilbert transformation (note that $F0 \approx 262\ Hz$, corresponding to a period of 4 ms). Figure D.2(a) shows the variation of TEF along the instrument continua. For the cello-sax continuum a trend of decreasing TEF with increasing α can be observed, and both TEF and TEF range are higher in the cello-sax than in the horn-trombone continuum.
Using Equation D.1, the morphing-parameters α_{ref} of the reference stimuli and the JND results

$\Delta\alpha$ (Figure 3.1) are translated into temporal-envelope-fluctuation measures (ΔTEF, TEF$_{ref}$ and TEF$_{test}$), as described above with Fc (Equation 3.7). Temporal-fluctuation difference ΔTEF at threshold as a function of the mean temporal fluctuation TEF of the stimuli at threshold is shown in Figure D.2(b) for the horn-trombone (circles) and cello-sax (crosses) continua. Figure D.2(b) shows that all TEF at threshold and most ΔTEF at threshold are higher in the cello-sax continuum than in the horn-trombone continuum.

Another measure that is connected to the temporal fluctuation is the peak factor (PF) of the wave form, that is, the ratio of the peak amplitude to the root-mean-square value. The PF is often used as a measure for the fluctuation in complex tones depending on the phase relation of cosine components (e.g. Moore et al., 2006). Hence, the peak factor of a non-stationary signal combines temporal envelope fluctuation and the fine structure of the wave. We define the peak factor PF as:

$$PF = \frac{\max(|A(t)|)}{\sqrt{\frac{1}{N} \cdot \sum_{t=1}^{N} A^2(t)}}, \tag{D.2}$$

where A(t) is the wave amplitude over time t. To test the degree to which the peak factor varies along the instrument continua, we calculate PF of all stimuli. The trend in the instrument continua are similar to TEF (Figure D.2(a)), but in the cello-sax continuum PF shows a high fluctuation along α (data not shown). This is probably a result of a flaw in the morphing algorithm used for the present study, in which phases were not appropriately morphed. The code version used for morphing the stimuli for the present study did not use Equation 3.3, but starting phases $\vartheta_{new}(t)$ (Equation 3.3) were set to zero. Since zero-phase correlation of partials in a complex tone maximizes wave fluctuation, an inappropriate change of phase to zero may increase the fine-structure fluctuation and, hence, increase the variance of the peak factor of the signal in different time frames, or here increase peak factor variance along α. For the following studies, this flaw was removed from the code and PF fluctuation along α disappeared. Using Equation D.2, the morphing-parameters α_{ref} of the reference stimuli and the JND results $\Delta\alpha$ (Figure 3.1) are translated into peak factor measures (ΔPF, PF$_{ref}$ and PF$_{test}$) as described above with Fc (Equation 3.7). As in TEF and in contrast to Fc, no systematic trend can be observed in either continuum. However, both peak factors and peak factor variation are higher in the cello-sax continuum than in the horn-trombone continuum, and so are all ΔPF at threshold (data not shown).

Discussion on the purely temporal timbre descriptors: The analysis of **temporal envelope fluctuation** and of peak factor showed similar results, which is in conformance with their being connected to the same signal characteristic. Both measures show higher fluctuation in stimuli of the cello-sax continuum as well as a higher variance of fluctuation along the continuum in comparison with the horn-trombone continuum. Since the cello-sax pair was chosen to reflect changes in the spectral flux, which strongly influences the temporal envelope, this is in conformance with expectations. At threshold, stimuli in the cello-sax continuum also exhibit higher fluctuation difference than in the horn-trombone continuum, which suggests that the present fluctuation differences influence discriminability in the cello-sax continuum.

[12]Note that Appendix A amongst others showed that the stimuli used in the present study do not have a distinct or crucial inharmonic content.

[13]Spectral flux, a spectro-temporal dimension, is difficult to measure, because any measure is influenced more strongly by analysis parameters such as window length than a pure temporal or

spectral dimension (such as attack rise-time, spectral centroid or spectral irregularity). Additionally, the influence of the dominating and varying spectral shape or centroid cannot be segregated from the spectral flux analysis; the interference of spectral and temporal dimension was shown by various other studies (e.g. Green, 1988a; Moore et al., 2006). In particular, the low-frequency harmonics influence the measures when analyzing the spectral flux of the whole spectrum, whereas cut-off frequency and normalization factor determine the measures when analyzing synchronicity of only the upper harmonics. These difficulties may partly explain the controversial results and discussions about spectral flux over the past 30 years, whereas pure temporal dimensions, spectral dimensions or dimensions comparing a spectral descriptor across two long time windows (attack segment vs. stationary segment; Iverson & Krumhansl, 1993) seem to be more stable (see McAdams et al., 1995, for comparison of timbre dimensions found in previous studies).

[14]Feedback indicated whether the subject's response was correct. Feedback was given throughout all runs in Experiment A, while only in the beginning of every run in Experiment B.

[15]The *multiple comparison procedure* provides an upper bound (p) on the probability that *any* comparison (of all group pairs to compare) will be incorrectly found significant (MATLAB's Help browser).

[16]Combining this loudness distortion caused by compression loss with the loudness shift caused by attenuation and the linear level amplification used in the present study, the intensity change due to these effects can be sketched. Figure 4.3 shows the input-output function of compression loss by 80% (due to OHC damage), linear attenuation by 20% (due to IHC damage) (Moore & Glasberg, 1997), and amplification by 50% of the total hearing loss, for "common" flat hearing losses of various degrees from 0 to 80 dB.
Imagine a **hypothetical** hearing-impaired subject whose ears show the modeled input-output function with 80% compression loss and 20% attenuation. Then Figure 4.3 would display schematically the internal intensity representation (or "partial loudness") which this hearing-impaired listener perceived in the experiment as a function of the partial loudness that normal-hearing listeners perceived in the experiment when listening to the same stimulus. Note that Figure 4.3 does not consider additional effects due to auditory processing, for example cochlear suppression. Any acoustic signal (or stimulus harmonic with equivalent amplitude density) with an internal representation below 0 dB (grey-shaded area in Figure 4.3) would not be audible. Figure 4.3 shows that in the present experiments, a hearing-impaired subject could not perceive signals or stimulus parts which normal-hearing listeners perceived with low loudness. Hence, hearing-impaired listeners could not use low-level harmonics up to a certain amplitude density (depending on hearing loss) that normal-hearing listeners could use for discriminating stimuli. On the other hand, loudness differences at intermediate signal levels, from hearing threshold up to more than 57 dB (crossing point of curves in Figure 4.3), are for hearing-impaired listeners higher than for normal-hearing listeners. Hence, if a sound partial was audible for a hearing-impaired subject, he/she perceived a larger partial loudness difference between two different stimuli than the normal-hearing listeners in the same stimulus pair. This hypothesis is supported by amplitude modulation studies on hearing-impaired listeners (Moore et al., 1996) and with intensity studies by Florentine et al. (1993) and Schroder et al. (1994), who observed that the Weber fraction (delta I/I) is sometimes smaller in cochlear-impaired than in normal-hearing listeners, when compared at the same SL.

Florentine et al. (1993) showed that hearing-impaired listeners' intensity resolution in comparison with normal-hearing listeners is highly dependent upon presentation level and whether measurements are carried out at equal SPL, equal SL, or equal loudness level. For low and intermediate levels, the hearing-impaired listeners usually perceive a stimulus as quieter than normal-hearing listeners if it is presented at equal SPL, and as louder at equal SL. On one hand, intensity difference perceived by hearing-impaired listeners may be higher than by normal-hearing listeners, if level amplification overcompensates for compression and hearing loss (i.e. at equal SL). On the other hand, intensity difference perceived by hearing-impaired listeners may be equal to or lower than that of normal-hearing listeners if presentation level does not compensate (i.e., at equal SPL).

[17]All subjects with flat hearing loss show a hearing loss < 60 dB for almost all frequencies.

[18]All subjects with steep hearing loss show a hearing loss > 60 dB at frequencies > 4 kHz.

[19]Chapter 5 and Appendix C show that stimuli at timbre discrimination threshold show distinct differences in the upper frequencies above 1-2 kHz

[20]Timbre distortion is probably also caused by the steep flank and asymmetry of the hearing loss, which may cause an additional frequency distortion compared to just the intensity/dynamic distortion in flat hearing losses. The difference between flat and steep hearing thresholds on timbre was shown, for example, by Doherty & Lutfi (1999) and Lentz & Leek (2003) in profile analyses. While subjects with a flat hearing loss are more likely to use the sound parts at the edge frequencies of the complex tone, subjects with a steep hearing loss weigh the region of their hearing loss more efficiently.

[21]In contrast to timbre discrimination, speech intelligibility in normal-hearing listeners does not degrade down to a negative SNR, because speech is an over-determined object with salient features.

[22]Oxenham & Bacon (2003) showed that frequency selectivity and temporal resolution in hearing-impaired listeners is lower than in normal-hearing listeners at equal sound level. Additionally, frequency selectivity decreases with sound level. Hence, at equal intermediate sound levels, at equal sensation level or at equal loudness, frequency selectivity and temporal resolution in hearing-impaired listeners are poorer than in normal-hearing listeners (see also Appendix B).

[23]In previous timbre experiments (Chapter 3) significant JND differences between amateur musicians and non-musicians, and significant training effects in subjects without musical experience were observed. In everyday life, many non-musicians do not consciously pay attention to timbre. They are not used to telling the difference between similar musical timbres and do not consciously use timbre as a separation cue. Hence, in the experiment it is difficult to instruct the subjects, exactly what they have to listen to; different subjects use different timbre cues for distinction, and not everyone finds the optimal cue. This difficulty results from the large number of possible dimensions in which timbre can vary (McAdams et al., 1995), the small number of descriptive words for them, and the lack of categories assigned to real objects (e.g. musical instruments or

human voices).

[24]See also Appendix B for detailed explanation and discussion.

[25]According to subject's individual pure-tone audiogram, 80% of the hearing loss at the filter channel's center frequency is attributed to outer hair cell loss, but maximal 55 dB for frequencies up to 2 kHz or 65 dB for frequencies higher than 2 kHz.

[26]To model the limits of resolution, an internal noise is added to the output of the preprocessing stages. However, for this study, the transformed signal without adding an internal noise is called "feature vector" or "internal representation" IR (Figure 5.1).

[27]In another way of interpreting Equation 5.1, the *optimal detector* uses the Pearson product to correlate the IR signal increment with a template (compare Equations 5.1 and 5.3). Instead of normalizing the correlation coefficient by the signal energies (denominator in Equation 5.3) the *optimal detector* only normalizes by the size of IR matrix (pre-factor in Equation 5.1).

[28]The *optimal detector* assumes that all frequency channels, modulation channels and time steps are independent "observations" (Dau et al., 1997b).

[29]The ERB-wide filters are spaced at 1 per ERB as in the original PeMo version. The first channel of the modulation filter bank is a low-pass filter with cutoff frequency of 2.5 Hz and the other channels are band-pass filters centered at 5, 10, 16.7, 27.8, 46.3, 77.2, 128.6, 214.3 and 357.2 Hz.

[30]Including a threshold would lead to an asymptotic ΔIR curve and flatter slopes for low $\Delta\alpha$ in all continua.

[31]Note that decision procedure is somewhat different to the one used by Dau et al. (1996). The way with subtracting the IRs of test interval and reference sound and subsequently comparing it to a stored template is not adaptable here. In the present case, spectrum of test interval is not necessarily a (positiv) sum of reference interval and increment, as well as no template can be calculated by a "super-threshold signal".

[32]A p-weighting (Equation 5.4) with p=2 (Euclidean) weighs IR *intensity differences* quadratically, with p$\gg$2 (or using a one-channel model) detects timbre differences only by the channel with the highest intensity difference, and with p=1 equally integrates all channels. On the other hand, a weighting of certain frequency channels simulates edge effects or perceptual grouping, for example of the harmonic frequencies $f = n \cdot F0$.

[33]Both effects may be related.

[34]Physiological filters would be asymmetric and compressive and would be able to predict peripheral suppression effects and comodulation-masking-release.

[35]Simulations of the experiments in noise using cross-correlation did not lead to useful results.

[36]Note difference to Dau et al. (1996) and Derleth et al. (2001), who used the *optimal detector* as decision device, which uses a normalized template IR (i.e. equal absolute template energy for normal-hearing and hearing-impaired listeners) and weighs the IR channels according to the energy distribution in the template (note also footnote 31, p.126).

[37]In the cello-sax continuum, the fluctuating background noise seems to mask temporal cues of the signal, hence, spectral cues may become more important.

[38]In simulations using random noise and subtracting the IR of a random noise presentation, the adaptive measurements diverged due to the high noise fluctuation (not shown here).

[39]***PeMo* and timbre descriptors:** Comparing Table 5.2 with Table 2.1 shows that highest coefficients of correlating IR distances with subjective ratings are similar to highest coefficients of correlating spectro-temporal timbre descriptors with subjective ratings in every instrument continuum, respectively. Hence, similarity ratings seem to be predicted by *PeMo* to the same extent as by common spectro-temporal timbre descriptors (see Appendix A). However, for detecting differences along a certain timbre dimension another processing and weighting in *PeMo* may be optimal than for another dimension. For example, spectral differences (spectral centroid, spectral irregularity, tonal inharmonic energy) are well detected without modulation filter bank and only comparing the temporal IR mean across stimuli. Time-step-wise comparison may even add here non-perceived differences, for example if the same amount of random noise is present in the two stimuli. Spectral flux differences (or "synchronicity in the overtones") seem to require time-step-wise comparison of IRs. Spectral flux differences were even not extracted using modulation filters and the temporal IR mean (Figure 5.3(f)), which may be due to the modulation *phase* that was removed by the temporal mean. Not the modulation phase is perceived by subjects but the amount of (modulation phase) synchronicity. A dynamic time-warp may improve the simulations, for example by using the maximum of the correlation function. Differences in attack descriptors (attack's centroid, overtone synchronicity, log-rise-time, high-frequency energy and inharmonic content) may be detected by a temporal IR mean of the attack duration. Since sound features during the attack are perceived dominating even if low in amplitude (Chapter 2), comparing entire stimuli of 1.8 seconds may obscure the crucial differences by unperceived random differences during the stationary portion of stimulus. Hence, a higher weighting of the first 300-500 ms may be appropriate here. Varying noise content may be difficult to detect, because low inharmonic amplitude can be perceived distinctly by subjects, whereas *PeMo* does not know anything about "Gestalt" and harmonicity.

[40]Instead of using Equation A.1 (p. 90), the spectral centroid can also be calculated using the power-spectrum in dB (i.e. $10 \cdot log_{10}(A^2)$ instead of A) and logarithmic frequency (i.e. $\log_{10}$(f)

instead of f) according to Weber's law, or using the entire spectrum instead of the line spectrum of the harmonics, or using a loudness model and calculating the centroid of an internal spectral representation (Zwicker & Scharf, 1965; Grey & Gordon, 1978).

[41]Krimphoff et al. (1994) calculated the spectral centroid as the average of the instantaneous spectral centroids within a running time window of 12-16 ms. The short time span makes it difficult to extract the tonal sound part from the inharmonic energy, due to the low fundamental frequency of 262 Hz and the coarse spectral resolution of 83 Hz, when using an FFT length of 12 ms. Therefore Welch's (1967) overlap-add SFFT analysis was used to optimize the spectral information, while the centroid was calculated of the time-averaged spectrum.

[42]In some instruments the steepest increase is finished early before reaching absolute amplitude maximum. And furthermore is it indecisive, if the increasing amplitude after the first local maximum in the trumpet (dotted line in the left graph of Figure A.3(a)) still belongs to the attack or is rather an exciting crescendo after the attack. The local dip in the trombone sound (dotted line in the right graph of Figure A.3(a)) at around 150 ms could also be either the end of the attack or an air-jam belonging to the attack.

[43]If high-frequency content as well as attack centroid were not normalized (by the equivalent measures of the stationary portion) the measures would be strongly correlated with and not separable from mean spectral centroid Fc.

[44]The window length for calculating the overtone synchronicity OS shall be the low as possible to measure fast spectral fluctuation. However, due to the low fundamental frequency of 262 Hz, a SFFT length of 46 ms is necessary for a resolution of 22 Hz to extract the tonal sound part from the inharmonic energy.

[45]If correlating the entire spectrum, correlation value is partly dependent on window size due to inharmonic but non-noise energy, as in the trumpet. With increasing FFT length inharmonic content and noise becomes analyzed with increasing resolution.

[46]Subjects of the study in Chapter 4 with a flat hearing loss showed a mean increase in speech reception threshold (SRT) of 3.1 dB in stationary background noise.

[47] On-frequency refers to frequencies near the characteristic frequency. The characteristic frequency (CF) describes the frequency of a pure tone that excites a given location along the basilar membrane maximally at low levels.

[48] Characteristic location (CL) describes the location on the basilar membrane that is excited maximally by a given pure tone of frequency CF at low levels.

[49] Off-frequency refers to frequencies far from characteristic frequency

[50] In a certain distance from the best frequency, the non-linear placement of excitation along the BM seems to countervail the on-frequency non-linearity for stationary signals, which results in the above-mentioned linear BM response for stationary off-frequency excitation.

[51]If the signal is processed linearly (as in hearing-impaired listeners), a 2 dB decay of internal representation after the time instance Δt is matched by a 2 dB change of external level. That is, after Δt, an external masker signal can still mask an external signal with a level 2 dB lower than the masker. In the compressive case (for normal-hearing listeners in the on-frequency condition), a 2 dB change of internal representation is matched by a 10 dB change of masker level. Hence, after Δt, an external masker can only mask an external signal with a level 10 dB lower than the masker.

[52]Note that effects in this section are based on scientific findings, but whether they are connected to peripheral suppression effects as described in this section is speculative.

[53]Theoretically, spectral centroid (perceived as brightness) alone cannot be used as an object segregation cue, because sound spectra superpose to one spectrum (comparable with colours in the visual system). However, if the two sounds can be separated by other cues, such as by onset or pitch (e.g, by fundamental frequency or separable frequency bands), brightness difference may help to discriminate the sounds.

[54]A hearing loss cannot be counterbalanced by an inverse function as short-sightedness is counterbalanced by glasses. Hence, dynamic features such as automatic gain control and spectro-temporal features such as noise reduction must be used to design a good crutch instead of an optimal prothesis.

[55]Research on new hearing aid features is often done by normal-hearing listeners who test the improvement for object separation and speech intelligibility. And artifacts of hearing aid processing, in particular distortion, are often evaluated as being more or less disturbing by normal-hearing researchers or subjects. Evaluation by hearing-impaired listeners may be different, in particular after a timbre training.

Bibliography

Amatriain, X., Bonada, J., Loscos, A., & Serra, X. (**2002**). "Spectral processing". In *DAFX - Digital Audio Effects*, edited by U. Zölzer. John Wiley & Sons, Ltd.

American Standard Association (**1960**). *Acoustical Terminology, S1.1-1960*. New York: American Standard Association.

Bacon, S., Fay, R., & Popper, A., eds. (**2004**). *Compression - From Cochlea to Cochlear Implants*. New York: Springer.

Beattie, R., Barr, T., & Roup, C. (**1997**). "Normal and hearing-impaired word recognition scores for monosyllabic words in quiet and noise". *Br J Audiol*, **31**(3), 153–164.

Buschermöhle, M., Feudel, U., Freund, J., Klump, G., & Beey, M. (**2006, accepted**). "Signal detection enhanced by comodulated noise". *Fluctuation and Noise Letters*, **6**(4).

Cook, P. (**1999**). *Music, Cognition, and Computerized Sound*. Cambridge, MA: MIT Press.

Cooper, N. (**2004**). "Compression in the peripheral auditory system". In *Compression - From Cochlea to Cochlear Implants*, edited by S. Bacon, R. Fay, & A. Popper. New York: Springer.

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

Dau, T., Kollmeier, B., & Kohlrausch, A. (**1997**b). "Modeling auditory processing of amplitude modulation: II. Spectral and temporal integration in modulation detection". *J Acoust Soc Am*, **102**, 2906–2919.

Dau, T., Puschel, D., & Kohlrausch, A. (**1996**). "A quantitative model of the "effective" signal processing in the auditory system. I. Model structure". *J Acoust Soc Am*, **99**, 3615–3622.

Derleth, R.-P., Dau, T., & Kollmeier, B. (**2001**). "Modelling temporal and compressive properties of the normal and impaired auditory system". *Hearing Research*, **159 (1-2)**, 132–149.

Doherty, K., & Lutfi, R. (**1999**). "Level discrimination of single tones in a multitone complex by normal-hearing and hearing-impaired listeners". *J Acoust Soc Am*, **105**, 1831–1840.

Dueck, G. (**2005**). *Wild Duck*. Berlin: Verlag Markus Kaminski.

Ernst, S., & Verhey, J. (**2006, submitted**). "Role of suppression and retro-cochlear processes in comodulation masking release". *J Acoust Soc Am*.

Festen, J., & Plomp, R. (**1983**). "Relations between auditory functions in impaired hearing". *J Acoust Soc Am*, **73**, 652–662.

Florentine, M., Reed, C., Rabinowitz, W., Braida, L., Durlach, N., & Buus, S. (**1993**). "Intensity perception. XIV. Intensity discrimination in listeners with sensorineural hearing loss". *J Acoust Soc Am*, **94**, 2575–2586.

Fritts, L. (**2002**). "Data base by Electronic Music Studios, Iowa University". URL: http://theremin.music.uiowa.edu/MIS.html.

Gfeller, K., Witt, S., Adamek, M., Mehr, M., Rogers, J., Stordahl, J., & Ringgenberg, S. (**2002**a). "Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients". *J Am Acad Audiol*, **13**, 132–145.

Gfeller, K., Witt, S., Woodworth, G., Mehr, M. A., & Knutson, J. (**2002**b). "Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal". *Ann Otol Rhinol Laryngol*, **111**, 349–356.

Goossens, T., van de Par, S., & Kohlrausch, A. (**2006**). "Discriminality of statistically independent gaussian noise tokens and random tone-burst complexes". In *14th International Symposium on Hearing*. Cloppenburg: Springer Verlag.

Green, D. (**1983**). "Profile analysis - a different view of auditory intensity discrimination". *American Psychologist*, 133–142.

Green, D. (**1988**a). "Auditory profile analysis: Some experiments on spectral shape discrimination". In *Auditory Function - Neurobiological Bases of Hearing*, edited by G. Edelman, W. Gall, & W. Cowan. New York: Wiley.

Green, D. (**1988**b). *Profile Analysis. Auditory Intensity Discrimination.* New York: Oxford University Press.

Grey, J. (**1977**). "Multidimensional perceptual scaling of musical timbres". *J Acoust Soc Am*, **61**, 1270–1277.

Grey, J. (**1978**). "Timbre discrimination in musical patterns". *J Acoust Soc Am*, **64**, 467–472.

Grey, J., & Gordon, J. (**1978**). "Perceptual effects of spectral modifications on musical timbres". *J Acoust Soc Am*, **63**, 1493–1500.

Hall, J., Haggard, M., & Fernandes, M. (**1984**). "Detection in noise by spectro-temporal pattern analysis". *J Acoust Soc Am*, **76**, 50–56.

Holube, I., & Kollmeier, B. (**1996**). "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model". *J Acoust Soc Am*, **100**, 1703–1716.

Huber, R., & Kollmeier, B. (**2006**). "Pemo-q. a new method for obejctive audio assessment using a model of auditory perception". *IEEE Transactions on Audio, Speech and Language Processing*, **14**, 1902–1911.

Iverson, P. (**1995**). "Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes". *Journal of Experimental Psychology*, **21**, 751–763.

Iverson, P., & Krumhansl, C. (**1993**). "Isolating the dynamic attributes of musical timbre". *J Acoust Soc Am*, **94**, 2595–2603.

Javel, E., Geisler, C., & Ravindran, A. (**1978**). "Two-tone suppression in auditory nerve of the cat: Rate intensity and temporal analyses". *J Acoust Soc Am*, **63**(4), 1093–1104.

Kendall, R., Carterette, E., & Hajda, J. (**1999**). "Perceptual and acoustical features of natural and synthetic orchestral instrument tones". *Music Perception*, **16**, 327–364.

Kollmeier, B. (**1999**). "On the four factors involved in sensorineural hearing loss". In *Psychophysics, Physiology and Models of Hearing*, edited by T. Dau, V. Hohmann, & B. Kollmeier. Singapore: World Scientific Publishing.

Kollmeier, B., & Derleth, R.-P. (**2001**). "How linear is the auditory system? a model of compression and expansion based on psychoacoustics for normal and hearing impaired listeners". In *Physiological and Psychophysical Bases of Auditory Function, Proceedings of the 12th International Symposium on Hearing*, edited by Breebart, Houtsma, Kohlrausch, Prijs, & Schoonhoven. Maastricht: Shaker.

Krimphoff, J., McAdams, S., & Winsberg, S. (**1994**). "Caracterisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophasique". *Journal de Physique*, **4**, 625–628.

Krumhansl, C. (**1989**). "Why is musical timbre so hard to understand?" In *Structure and Perception of Electroacoustic Sound and Music*, edited by Nielzen, & Olsson. Amsterdam: Elsevier.

Lakatos, S. (**2000**). "A common perceptual space for harmonic and percussive timbres". *Perception and Psychophysics*, **62**, 1426–1439.

Launer, S., Hohmann, V., & Kollmeier, B. (**1997**). "Modeling loudness growth and loudness summation in hearing-impaired listeners". In *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt, & N. Mahwah. Lawrence Erlbaum Assoc.

Lentz, J., & Leek, M. (**2003**). "Spectral shape discrimination by hearing-impaired and normal-hearing listeners". *J Acoust Soc Am*, **113**, 1604–1616.

Levitin, D., McAdams, S., & Adams, R. (**2002**). "Control parameters for musical instruments: a foundation for new mappings of gesture to sound". *Organized Sound*, **2**, 171–189.

Levitt, H. (**1970**). "Transformed updown methods in psychophysics". *J Acoust Soc Am*, **49**, 467–477.

Maher, R. C., & Beauchamp, J. W. (**1994**). "Fundamental frequency estimation of musical signals using a two-way mismatch procedure". *J Acoust Soc Am*, **95**, 2254–2263.

McAdams, S., Beauchamp, J., & Meneguzzi, S. (**1999**). "Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters". *J Acoust Soc Am*, **105**, 882–897.

McAdams, S., & Cunibile, J.-C. (**1992**). "Perception of timbral analogies". *Philosophical Transactions of the Royal Society, London, Series B*, **336**, 383–389.

McAdams, S., & Winsberg, S. (**2000**). "Psychophysical quantification of individual differences in timbre perception". In *Contributions to Psychological Acoustics*, edited by A. Schick, M. Meis, & C. Reckhardt, vol. 8. Oldenburg: BIS.

McAdams, S., Winsberg, S., Donnadieu, S., Soete, G. D., & Krimphoff, J. (**1995**). "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes". *Psychol Res*, **58**, 177–192.

McFadden, D. (**1987**). "Comodulation detection differences using noise-band signals". *J Acoust Soc Am*, **81**, 1519–1527.

Moore, B. (**1998**). *Cochlear hearing loss*. London: Whurr Publishers.

Moore, B. (**2003**). *An Introduction to the Psychology of Hearing*. Academic Press, 5th ed.

Moore, B., & Glasberg, B. (**1997**). "A model of loudness perception applied to cochlear hearing loss". *Auditory Neuroscience*, **3**, 289–311.

Moore, B., Glasberg, B., & Hopkins, K. (**2006**). "Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure". *Hearing Research*, **222**, 16–27.

Moore, B., Huss, M., Vickers, D., Glasberg, B., & Alcntara, J. (**2000**). "A test for the diagnosis of dead regions in the cochlea". *British Journal of Audiology*, **34**, 205–224.

Moore, B., Wojtczak, M., & Vickers, D. (**1996**). "Effect of loudness recruitment on the perception of amplitude modulation". *J Acoust Soc Am*, **100**, 481–489.

Oxenham, A., & Bacon, S. (**2003**). "Cochlear compression: Perceptual measures and implications for normal and impaired hearing". *Ear & Hearing*, **24**(5), 352–366.

Oxenham, A., & Bacon, S. (**2004**). "Psychophysical manifestation of compression: Normal-hearing listeners". In *Compression - From Cochlea to Cochlear Implants*, edited by S. Bacon, R. Fay, & A. Popper. New York: Springer.

Oxenham, A., & Plack, C. (**1997**). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing". *J Acoust Soc Am*, **101**, 3666–3675.

Pekkarinen, E., Salmivalli, A., & Suonpaa, J. (**1990**). "Effect of noise on word discrimination by subjects with impaired hearing, compared with those with normal hearing". *Scand Audiol*, **19**, 31–36.

Plomp, R. (**1970**). "Timbre as multidimensional attribute of complex tones". In *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp, & G. Smoorenburg. Leiden: Sijthoff.

Plomp, R. (**1975**). "Auditory analysis and timbre perception". In *Auditory Analysis and Perception of Speech*, edited by G. Fant, & M. Tatham. New York: Academic Press.

Pressnitzer, D., & McAdams, S. (**2000**). "Acoustics, psychoacoustics and spectral music". *Contemporary Music Review*, **19**, 33–59, 139–143.

Ruggero, M., Rich, N., Recio, A., Narayan, S., & Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea". *J Acoust Soc Am*, 2151–2163.

Schroder, A., Viemeister, N., & Nelson, D. (**1994**). "Intensity discrimination in normal-hearing and hearing-impaired listeners". *J Acoust Soc Am*, **96**(5), 2683–2693.

Shannon, R. (**1976**). "Two-tone unmasking and suppresion in a forward-masking situation". *J Acoust Soc Am*, **59**(6), 1460–1470.

Taylor, C. (**1992**). *Exploring Music: the Science and Technology of Tones and Tunes*. Bristol, UK; Philadelphia, PA: Institute of Physics Publishing.

Terasawa, H., Slaney, M., & Berger, J. (**2005**). "The thirteen colors of timbre". In *IEEE Workhshop on Applications of Signal Processing to Audio and Acoustics*.

Terhardt, E. (**1974**). "On the perception of periodic sound fluctuations (roughness)". *Acustica*, **30**, 201–213.

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

Wagener, K., Brand, T., & Kollmeier, B. (**1999**a). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache II: Optimierung des Oldenburger Satztests". *Zeitschrift für Audiologie/Audiological Acoustics*, **38**, 44–56.

Wagener, K., Brand, T., & Kollmeier, B. (**1999**b). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests". *Zeitschrift für Audiologie/Audiological Acoustics*, **38**, 86–95.

Wagener, K., Kühnel, V., & Kollmeier, B. (**1999**c). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests". *Zeitschrift für Audiologie/Audiological Acoustics*, **38**, 4–15.

Welch, P. (**1967**). "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms". *IEEE Trans Audio Electroacoust*, **AU-15**, 70–73.

Wessel, D. (**1979**). "Timbre space as a musical control structure". *Computer Music Journal*, **3**, 45–52.

Zölzer, U., ed. (**2002**). *DAFX: Digital Audio Effects*. John Wiley & Sons, Ltd.

Zwicker, E., & Scharf, B. (**1965**). "A model of loudness summation". *Psychological Review*, **72**, 3–26.

Danksagung und Nachwort


Prof. Dr. Dr. Birger Kollmeier danke ich für die Ermöglichung dieser Arbeit in einer Arbeitsgruppe mit hervorragenden Arbeitsvoraussetzungen, für die Hilfe zur Einbettung der Arbeit in einen "wissenschaftlichen Rahmen" (s. auch Nachwort) und für inspirierende wissenschaftliche Diskussionen.


Jun.-Prof. Dr. Jesko Verhey möchte ich für die kritischen Anregungen und die Übernahme des Korreferates danken.


Finanziell unterstützt wurde diese Arbeit von der *Deutschen Forschungsgemeinschaft* (DFG) im Rahmen des Graduiertenkollegs *Neurosensorik* und des SFB/TR31 *Aktives Gehör* sowie von der Universität Oldenburg durch das Graduiertenförderungsgesetz. Besten Dank dafür.


Herzlichen Dank an Dr. Stefan Ewert und Dr. Rainer Huber für die stete Hilfe in Fragen um das *Perzeptions-Modell*, besonders in Fällen, da ein "Verbiegen" nötig war, um das Modell für meine Zwecke nutzbar zu machen. Vielen Dank an meinen Bläschen-Betreuer Dr. Thomas Brand für die stete Hilfsbereitschaft, sich mit meinem Thema auseinanderzusetzen, obwohl es meist sein Fachgebiet sprengte und meine Diskussionsart recht kritisch und temperamentvoll war. Dr. Manfred Mauermann möchte ich danken für Motivation und "periphere" Diskussionen, Stefan Strahl für die prompten Hilfen beim Lösen vieler Softwareprobleme, und Jennifer Shelley für die gründlichen und ästhetisch bereichernden Sprachkorrekturen. Da das bearbeitete Thema "Waise" in der Arbeitsgruppe war bzw. auf allen Stühlen saß, war ich auf die mannigfaltige Hilfe von vielen Kollegen in der Arbeitsgruppe, im Graduiertenkolleg und darüber hinaus angewiesen. Allen einen lieben Dank sowohl für die wissenschaftliche Hilfe als auch für die freundschaftliche und lustige Arbeitsathmosphäre!


Unserem unersätzlichen Systemadministrator Frank Grunau und den 3 Engeln aus Medi-Geschäftsstelle und Labor, Susanne Garre, Ingrid Wusowski und Anita Gorges, danke ich sehr herzlich für die Grundbedürfnisse: sowohl für die stets reibungslosen (verwaltungs)technischen Hilfen als auch für die persönlichen Aufmunterungen in Mitten der Wissenschaftler.


Bei Dr. Daniel Pressnitzer und Dr. Arne Lejon möchte ich mich für die anregenden Kurzaufenthalte in Paris und Stockholm bedanken und bei Alain de Cheveigné für die kritischen Gespräche zu Klangfarbe und Objektlinearität.


Dr. Birgitta Gabriel und den Mitgliedern des Hörzentrums Oldenburg danke ich für ein sehr lehrreiches Praktikum.


Insbesondere gilt mein Dank den 86 Personen, die an den oft langwierigen und anstrengenden Messungen teilgenommen haben.


Rapalje (www.rapalje.com) und Trio Mio (www.triomio.dk) herzlichen Dank für die gemorphten  und "verzerrten"  Instrumente zur Illustration, und William, Jens, Kristine und Nikolaj für die Musik, Wärme und Lebensfreude, die mir immer wieder Kraft zum Durchhalten gaben.


Melanie Zokoll möchte ich für die persönliche Unterstützung während des Arbeitsalltags und darüber hinaus ganz lieb danken! Ich bin keine Wissenschaftlerin im heutigen Sinn, ich bin neugierig wie ein kleines Kind, produziere nicht-endend-wollend Ideen und liebe es Zusammenhänge zu verstehen und sie anderen verständlich zu machen, aber komme nur suboptimal mit der Rigidität des bewussten Denkens, mit Trockenheit und kurzen einseitigen Erklärungen der "hypothesengetriebenen Publikationen", und mit dem (Bewertungs-) Druck der Brötchengeber klar. Deshalb wäre diese Arbeit ohne das Dasein von Uljana, Rike, Melli, Jutta und unserem ganzen Frauenseptet, ohne die Geige und die Liebe, die mir von meinen Eltern geschenkt wurden, und ohne die Zwischentöne von den anderen Freunden und lieben Menschen nicht zustande gekommen. Vielen lieben Dank dafür und vor allem dass Ihr mich (auch in der nervenaufreibenden Endphase) immer so nahmt wie ich war! Euch, Wilfried Menke, Andreas Burzik, der Folkmusik und Schweden ein besonderer Dank für die Welt "beyond science", das Aufzeigen neuer ganzheitlicher Wege und den Zuspruch, dass diese möglich sind!

...und weiter geht es auf neuen Wegen, auf denen ich als ganze Suzan wandern "darf" und gebraucht werde. Das analytische Denken, das den Baustein der Wissenschaft darstellt, macht mir immernoch Spaß, ist ein gutes Steuerrad und auch ein guter Gefühlsdosierer, wenn es innerlich zu turbulent wird; aber es ist eben nur ein Teil von mir. Ich verstehe, dass wissenschaftliche Publikationen einem gewissen Rahmen genügen müssen. Auch wenn mir die damit verbundene Kritik nicht immer schmecken wollte, habe ich dadurch während der letzten 4 Jahre auch viel für mich gelernt, wie z.B. objektive Sachverhalte von subjektiven Interpretationen zu trennen oder Missverständnisse, die durch Sprache und unterschiedliche Sichtweisen entstehen, zu entdecken und aufzuklären. Schade finde ich allerdings immernoch, dass sich Kreativität und Bildsprache einerseits und Glaubwürdigkeit in der Wissenschaft andererseits anscheinend ausschließen. Sofern deutlich ersichtlich ist, was schlicht zur Motivation oder Veranschaulichung dient, sollten meiner Meinung nach verdeutlichende Bilder, "aus dem Leben gegriffene" einleitende Worte oder direkte Fragen, die provozierend an den Leser gerichtet sind, auch in Publikationen gestattet sein. Lenken beispielsweise die Instrumenten-Fotos in Abbildung A.6 (Seite 97) vom wissenschaftlichen Ergebnis ab? Wissenschaftliche Durchbrüche entsprangen oft aus

Köpfen (und damit meine ich nicht meinen!), die nicht nur Denkpotential hatten, sondern auch einen gewissen Sinn für Ästhetik und verspielte Kreativität - hypothetisch behaupte ich an dieser Stelle, dass eingeschränkte kreative Freiheit neben Motivation und Glück auch die wissenschaftliche Leistung hemmt (s. auch Dueck, 2005). Die aalglatt dargestellten Ergebnisse in heutigen Publikationen lassen vermuten, dass sich die Welt durch unikausale und unidirektionale Zusammenhänge erklären lässt. Das mag beruhigend sein, wenn mann sich mit "verwirrenden" weil multiplen Begründungen schwertut. Ich glaube allerdings, dass die einseitige, trockene und kopflastige Weise, in der Wissenschaft publiziert werden muss, eine unerwünschte Selektion der Wissenschaftler und damit ein Ungleichgewicht hervorruft: Auch wenn Uni-Alltag, Forschung und Diskussionen sehr lebendig und inspirierend sind und viel Raum für Querdenker und Spinnereien lassen, suchen sich viele Menschen mit ganzheitlicher Sichtweise lieber andere Betätigungsfelder, und die wissenschaftlichen Erkenntnisse der übrigen Wissenschaftler befriedigen nicht unbedingt die Bedürfnisse eines Otonormalverbrauchers. Praktisch ausgedrückt, wenn Hörgeräte nur auf Sprache optimiert werden, da für die verbleibenden Wissenschaftler Information um so viel wichtiger ist als die Zwischentöne, haben schwerhörende Musiker schlechte Karten. Dazu kommt auch noch die Abhängigkeit von Geld und Ruhm, wodurch beispielsweise alte Entwicklungen wie die analogen Hörgeräte oder digitale Hörgeräte mit simpler linearer Verstärkung vom Markt genommen werden bzw. dem Kunden nicht mehr angeboten werden, obwohl sie die neuen Errungenschaften (von der Verbraucher Sicht aus) gut ergänzen würden. Sollte Wissenschaft nicht eher der Menschheit als der Industrie dienen? Natürlich ist Einseitigkeit nicht nur hier zu beobachten, und auch wenn die Wissenschaft in unserer Gesellschaft hoch angesehen ist, ist sie eben auch nur ein Arbeitsbereich und eine Spielwiese von Menschen. Genauso wie die Musikbranche...

...wie schön dass die Musik selbst unbeeindruckt davon bleibt! Ist es nicht klasse, dass man Musik und Klangfarben nicht auf Papier festhalten kann? Dass Musik weder falsch noch richtig ist, dass man sie nicht beweisen muss, ja dass man sie eigentlich gar nicht mit Denken greifen kann? Ist es nicht herrlich, dass man mit Musik "Ottonormalverbraucher" erreichen kann ohne sich selbst einzuschränken? Und dass Musik die Verbindung zu uns selbst und zu anderen Menschen, die uns durch unsere Körper und Köpfe genommen wurde, wieder herstellt! Aber weiter philosophiere ich darüber lieber auf meinem Geigchen. An dieser Stelle, nur noch eine Bitte an Euch; nachdem die Danksagung zu den meist-gelesenen Teilen einer Diss gehört, ist es quasi die Quintessenz der vorliegenden Studie: Passt bitte auf Eure Ohren auf, so dass Ihr bei Tanzabenden, Konzerten oder Sessions vorbeihören und Euch noch lange an farbigen Klängen erfreuen könnt!

Wissenschaftlicher Lebenslauf von Suzan Selma Emiroğlu

24. Juni 1976	geboren in Dachau (D), Staatsangehrigkeit: deutsch
1983 - 1987	Besuch der *Grundschule Parksiedlung* in Oberschleißheim
1987 - 1995	Besuch des *Gerhardinger-Gymnasiums* (musischer Zweig) in München
20.06.1995	Abitur ◦ Prüfungsfächer: Mathematik, Physik, Englisch, Religion
1995 - 1997	Studium der Mathematik (Diplom) an der Universität München
Nov'95 - Aug'99	Aushilfstätigkeit als Planetariumsvorführer im *Forum der Technik* in München, inkl. live-Vorführungen, Dia- und Bildbearbeitung und Organisation im Rahmen von Show-Produktionen, 8-40 Std/Wo
1997 - 2003	Studium der Geophysik (Diplom) an der Universität München ◦ hilfswissenschaftliche Arbeiten in paläomagnetischen und seismischen Messungen und der Bearbeitung der Instituts-Internetseite ◦ 7-wöchiges Praktikum beim *British Geological Survey* (Edinburgh, GB) zur Simulation seismischer Wellenausbreitung ◦ Organisation einer geo$\frac{\text{physikal}}{\log}$ischen Exkursion nach Neuseeland in Kooperation mit der *Victoria University* (Wellington, NZ)
03.05.1999	Vordiplom in Geophysik ◦ Prüfung in exp. Physik, theor. Physik, Mathematik, Mineralogie
Jul'00 - Feb'03	gefördertes Mitglied der *Studienstiftung des deutschen Volkes*
17.03.2003	Diplom in Geophysik (Studiendauer: 11 Fachsemester) ◦ Prüfung in Geophysik, Geologie, exp. Physik, Kristallographie ◦ Diplomarbeit mit dem Titel *Depth dependency of magnetic properties in recent sediments of the Ría de Arousa, Spain*, 12 Monate einschließlich 6 Monate an der Universität Vigo, Spanien
seit 01.09.2003	Promotionsstudent an der Universität Oldenburg ◦ Betreuer: Prof. Dr. Dr. Birger Kollmeier (*Medizinische Physik*) ◦ 10-wöchiges Praktikum am *Hörzentrum Oldenburg* ◦ Anleitung von Studenten im Rahmen der Praktika *Psychoakustik* und *Digitale Signalverarbeitung* ◦ Graduiertensprecherin im Graduiertenkolleg *Neurosensorik* (1 Jahr) ◦ dezentrale Frauenbeauftragte des Instituts für Physik (2 Jahre)
Sep'03	Stipendiatin und ...
seit Okt'03	... assoziiertes Mitglied des Graduiertenkollegs *Neurosensorik*, gefördert von der *Deutschen Forschungsgemeinschaft* (DFG)
Okt'03 - Aug'05	Stipendiatin der Universität Oldenburg durch das Graduiertenförderungsgesetz
seit Sep'05	wissenschaftliche Mitarbeiterin im Sonderforschungsbereich *Das aktive Gehör*, gefördert von der *Deutschen Forschungsgemeinschaft* (DFG)

Erklärung

Hiermit versichere ich, Suzan Emiroğlu, dass ich die vorliegende Doktorarbeit selbständig verfasst habe. Es wurden keine anderen Quellen und Hilfsmittel außer den hier aufgeführten verwendet. Die Dissertation ist noch nicht in Gänze oder Teilen veröffentlicht. Kapitel 4 wurde am 11.05.2007 mit dem Titel "Timbre discrimination in normal-hearing and hearing-impaired listeners under different noise conditions" zur Veröffentlichung bei der Zeitschrift *Brain Research* eingereicht.

Oldenburg, den 30.05.2007

.....

Timbre is a combination of all auditory object attributes other than pitch, loudness and duration, and is used to distinguish different musical instruments or voices. People with sensorineural hearing loss often have problems with timbre distortion. Even for modern hearing aids it is difficult to provide good audio quality for speech intelligibility while preserving the natural timbre. This not only affects music perception, but may also influence object recognition in general. The present study aims to quantify differences in object segregation and timbre discrimination between normal-hearing and hearing-impaired listeners with a sensorineural hearing loss. In order to improve auditory models and hearing aids, a new method for studying timbre perception was developed. Using cross-faded (morphed) instrument sounds in psychoacoustic measurements, the subtle timbre perception differences between listener groups are studied. The results of the similarity rating and discrimination experiments are discussed in the context of common timbre models and simulated using an effective auditory computer model for the normal and impaired hearing system. The present study shows that, as opposed to reduced ability of hearing-impaired listeners to separate natural objects due to a reduction in time and frequency resolution, certain timbre dimensions seem to not be degraded by compression loss and might provide hearing-impaired listeners with cues for separating objects when linear sound amplification is provided. Lowering the distortion connected to non-linear amplification in hearing aids may not only enhance the pleasure of listening to music but also support the user's ability to separate objects.