Zur Homepage der Dissertation

Loudness Models for rehabilitative Audiology

Vom Fachbereich Physik der Universität Oldenburg
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
angenommene Dissertation

Jens-Ekkehart Appell
geboren am 25. September 1966
in Rotenburg an der Fulda

Erstreferent:	Prof. Dr. Dr. Birger Kollmeier
Korreferent:	Prof. Dr. Volker Mellert
Tag der Disputation:	21. Februar 2002

Zur Homepage der Dissertation

Loudness Models for rehabilitative Audiology

Vom Fachbereich Physik der Universität Oldenburg
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
angenommene Dissertation

Jens-Ekkehart Appell
geboren am 25. September 1966
in Rotenburg an der Fulda

Erstreferent:	Prof. Dr. Dr. Birger Kollmeier
Korreferent:	Prof. Dr. Volker Mellert
Tag der Disputation:	21. Februar 2002

Contents

Chapter 1

Introduction

Rehabilitative audiology is a quiet young, interdisciplinary field of research, which is concerned with the diagnosis of hearing impairment and its rehabilitation with hearing instruments. Due to the introduction of digital technology in modern hearing instruments, major advancements have been made in this field in recent years. However, these advances primarily are concerned with technical issues, such as, e.g., advanced signal processing techniques for noise reduction, suppression of feedback and increase of listening as well as handling comfort. The techniques for restoration of a patients individual deficiencies, on the other hand, are still limited by our limited knowledge about the effect of hearing impairment on the individual listener. Hence, a major research effort is needed for a better understanding of the individual hearing impairment in order to derive better processing techniques for the compensation of the individual listeners' hearing handicap. This thesis is directed towards this aim.

For the diagnosis and treatment of hearing disorders, hearing abilities of hearing–impaired listeners are classically characterized by the audiogram, i.e., the loss of hearing sensitivity as a function of frequency. However, it is widely known that people suffering from cochlear hearing loss exhibit alterations in several auditory functions, such as loudness perception, intensity discrimination, frequency selectivity, temporal resolution and speech perception [for an overview see Moore, 1995]. Hence, supra–threshold hearing might be as important for the treatment of the individual sensorineural hearing–impaired listener as the sensitivity loss measured by the audiogram. The probably most important quantity for characterizing supra–threshold perception has been described with the term *loudness*, i.e., the subjective sensation corresponding to the physical level of the acoustical stimulus which is also influenced by a number of other issues (e.g., temporal and spectral structure of the signal). Loudness and loudness perception is a major issue in psychophysics which can be quantified for normal listeners in terms of *loudness models* [e.g., Fletcher and Munson, 1933; Moore and Glasberg, 1996; Stevens, 1956; Zwicker, 1958; Zwislocki, 1965]. Thereby, loudness models aim at predicting perceived loudness from the physical properties of the sound by considering psychophysical principles (such as the pioneering assumptions of Fechner [1860], Weber [1905] and Stevens [1957]) as well as physiological and psychoacoustical findings.

The finding of an altered loudness perception in sensorineural hearing–impaired listeners, i.e., the so called *recruitment phenomenon* [Brunt, 1994; Fowler, 1936; Steinberg and Gardner, 1937] (the phenomenon that once the level of a sound is increased above threshold, the loudness increases more rapidly than in normal–hearing listeners), has also been described and discussed in the literature for many years. Based on the classical model considerations by Fletcher [Fletcher and Munson, 1933; Fletcher and Steinberg, 1924] and Stevens [Stevens, 1953, 1956], Zwicker [Zwicker, 1958; Zwicker and Fastl, 1990] developed a loudness model which has been the basis for a variety of modifications and improvements both to predict recent data and to predict the effect of hearing impairment [Chalupper, 2000; Florentine and Zwicker, 1979; Launer, 1995; Launer *et al.*, 1996, 1997; Marzinzik *et al.*, 1996b; Moore and Glasberg, 1996, 1997; Moore *et al.*, 1996, 1997, 1999b, 2000, 2001; Paulus and Zwicker, 1972]. Since any compensation for distorted loudness perception in hearing–impaired listeners provided by a hearing aid should be based on a solid understanding of the nature of loudness perception, the development and evaluation of appropriate loudness models is of major importance in rehabilitative audiology. This motivated the work in the current thesis.

The thesis is organized as follows: It starts with a description of a field–test performed with a prototype digital signal processing hearing aid, which aims at restoring loudness perception in the individual hearing–impaired listeners in a limited number of frequency channels with different algorithmic approaches (mainly with respect to dynamic compression algorithms, chapter 2). Since the limitations of the assessment of loudness perception in this field–test and its compensation by the hearing aid is obvious, the next chapter (chapter 3) is devoted towards appropriate measurement methods for assessing loudness perception in normal– and hearing–impaired listeners. Specifically the loudness matching task (i.e., adjustment of the level of two different stimuli to yield the same loudness impression) is compared to categorical loudness scaling (i.e., direct numbering of the perceived loudness impression by the subject with given categorical units). Although both methods are found to be quite inaccurate in quantifying loudness summation for individual listeners, the loudness scaling method has the advantage of a higher applicability in the clinics. Hence, the reminder of the thesis is devoted towards modeling loudness perception for stationary sounds in normal– and hearing–impaired listeners based on data derived from the method of loudness scaling. While chapter 4 gives a thorough review of the literature on different versions of the loudness models based on the approach by Zwicker, a new loudness model (The "Oldenburg Loudness Model") is introduced in chapter 5, which has the advantage of better predicting the individual data of hearing–impaired listener and taking into consideration the most recent data on equal–loudness level contours. Finally chapter 6 performs an evaluation of this model in comparison with two other models known from the literature.

Chapter 2

Evaluation of different 3–Channel Dynamic Compression Schemes in a Field–test[1]

Abstract

Three different 3–channel dynamic compression schemes (automatic volume control, syllabic compression and compression limiting) which have previously been tested in the laboratory by Appell *et al.* [1995] were investigated in everyday life with five hearing–impaired subjects using a prototype digital hearing–aid. A battery of tests was performed containing categorical loudness scaling with narrow- and broadband stimuli, speech intelligibility in noise using the Göttingen Sentence Test and quality assessment by paired comparison tests as well as a questionaire and an informal interview. Due to the carefully selected control conditions (i.e., unaided situation at roughly the same perceived loudness as in the aided situations, algorithms fitted with the same fitting rationale, same frequency response for all algorithms for a medium input level with a speech–shaped spectra) the differences across algorithms were only very small. Hence, no overall 'winner algorithm' can be derived from the current data. However, it is found that subjects with a small residual dynamic range and high speech reception thresholds (SRT) showed best performance in quality and speech intelligibility with dynamic compression whereas no clear–cut preference is found in the other subjects. From the current study it can be suggested that for low input levels a slow acting compression with a high compression ratio (i.e., automatic volume control) should be used to provide audibility at this input level range, whereas syllabic compression (small compression ratio) or even linear amplification seems to be beneficial at medium to high input levels. In any case compression limiting should be provided to prevent from high level signal peaks.

[1]This chapter has been submitted by Jens-E. Appell, Volker Hohmann, Birgitta Gabriel and Birger Kollmeier for publication in *Speech Communication, special issue on speech processing for hearing aids.*

2.1 Introduction

One of the most common principles in modern digital hearing–aids is the dynamic range compression. It is introduced in order to compensate for the so–called recruitment phenomenon [Steinberg and Gardner, 1937], i.e. the deteriorated loudness perception in sensorineural hearing–impaired patients. While the classical solution is a broadband automatic volume control (AVC, i.e., gain adjustment of the hearing–aid across the full frequency range according to the input level averaged over a certain time window), more advanced multiband dynamic compression systems have been suggested and evaluated in the past [i.e., Bustamente and Braida, 1987; de Genaro *et al.*, 1986; Festen, 1999; Fröhlich, 1993; Hohmann and Kollmeier, 1995a; Lippmann *et al.*, 1981; Marzinzik *et al.*, 1997; Moore *et al.*, 1999c; Nábělek, 1983; Plomp, 1988; Stone *et al.*, 1999; Tejero–Calado J.C., Rutledge J.C., and Nelson P.B.; Walker and Byrne, 1984; White, 1986]. A comprehensive overview is given by Working Group on Communication Aids for the Hearing–Impaired [1991], Kollmeier [1997a] and Verschuure and Dreschler [1996]. The aim of these compression algorithms is to restore a maximum number of the impaired auditory functions found in hearing–impaired patients in an optimum way. Given the limited signal processing capabilities available in hearing–aids and our limited knowledge about the ”effective” signal processing of the normal and impaired auditory system, this restoration can only be incomplete. Therefore, an important partial goal is to at least compensate for the altered loudness impression in hearing–impaired listeners and/or to provide the optimum presentation level of the input signal. This should help the hearing–impaired listener to process speech in an optimum way in order to optimize speech intelligibility.

In any case, linear frequency shaping is required, typically with a fine spectral resolution, in order to equalize the frequency response independently from the input level. In addition, a nonlinear compression component is required in order to compress the large dynamic range of input signals to the limited dynamic range of the impaired ear. Typically, the latter operation can be performed with a lower frequency resolution than the linear frequency shaping. Several multi–band dynamic compression schemes have been suggested that usually perform a dynamic compression in several frequency bands independently. In general, the sound quality (and in most cases also the performance in terms of restoring speech intelligibility in quiet and in noise) deteriorates with increasing number of frequency channels and with decreasing time constants [de Genaro *et al.*, 1986; Festen, 1999; Goedegebure *et al.*, 1996a; Hansen, 2000; Nábělek, 1983; Neuman *et al.*, 1995; Plomp, 1988]. However, typically the evaluation studies on dynamic compression hearing–aids have the following shortcomings:

— The linear frequency shaping (which should be applied for a medium input level and should provide a frequency shaping normally introduced by fitting rules for linear hearing–aids such as NAL [Byrne, 1986]) is typically not separated from the nonlinear component of the level correction (i.e., compression in few bands). Hence, when comparing different compression schemes with each other it should be secured that all algorithms provide the same frequency shaping for a certain input signal (such as, e.g., speech–simulating noise) at a certain input level (such as, e.g., medium conversation level).

— The evaluation is usually only based on laboratory studies with a limited set of acoustical situations and a very limited set of input levels. Under daily life conditions, however, dynamic compression algorithms have to perform in a variety of acoustical conditions and presentation levels. Hence, a more comprehensive comparison of dynamic compression systems should be performed that encompasses a wide range of conditions and levels as well as a field test.

— Most studies have concentrated on comparatively few items to be assessed by the hearing–impaired listeners (such as, e.g., speech intelligibility in quiet and in noise). However, other important items should be considered that characterize the hearing–aid performance in real–life situations (such as, e.g., loudness compensation for a variety of input levels and bandwidths, subjective assessment of the hearing–aid and overall quality rating).

Hence, the current study tries to perform a valid comparison of different compression schemes for multi–channel dynamic compression hearing–aids (syllabic compression, automatic volume control, and compression limiting) while concentrating on a variety of different evaluation criteria and using a wearable prototype signal processing aid in a field test. The aim of the study is to find out — under controlled experimental conditions both in the laboratory and in the field — if there are any consistent differences across compression rationales and fitting rules.

2.2 Processing Schemes

Within this study, different 3–channel dynamic compression schemes were investigated in everyday life which have previously been tested in the laboratory by Appell *et al.* [1995]. The main parameters of the processing schemes were two cutoff frequencies separating the three frequency channels, the attack and release time constants of the input level estimators and the input–output–characteristics (*I/O–characteristic*) of the dynamic compression. A schematic overview of the 3–channel–master–hearing–aid algorithm employed in this study is given in Figure 2.1.

The algorithms were implemented on a wearable digital hearing–aid device, the so–called *DASi–2* (**D**igital **A**uditory **Si**gnalprocessor, version 2) which had been developed by Raß and Steeger [2000]. The hardware is described in detail in appendix A.1. The signal processing framework implemented on the DASi provides signal processing in the frequency domain using an Overlap–Add processing scheme (18.9 kHz sampling frequency, 180 sample Hanning windowing with 50% overlap, zero padding by 76 sample resulting in 256 point FFT). For more details see appendix A.2).

The algorithms were implemented as follows: The low–pass, band–pass and high–pass signals are obtained by summing up the intensities $I_n(f)$ of the FFT bins $f_{start}(c)$ to $f_{end}(c)$ belonging to the respective channel c:

$$I_n(c) = \sum_{f_{start}(c)}^{f_{end}(c)} I_n(f) \qquad (2.1)$$

where n is used as the sample index. The two cutoff frequencies f_{LowCut} and $f_{HighCut}$ between the three channels are adjustable by selecting the corresponding FFT bins:

$$\begin{aligned} f_{LowCut} &= f_{end}(c=1) = f_{start}(c=2) - 1 \\ f_{HighCut} &= f_{end}(c=2) = f_{start}(c=3) - 1 \end{aligned} \qquad (2.2)$$

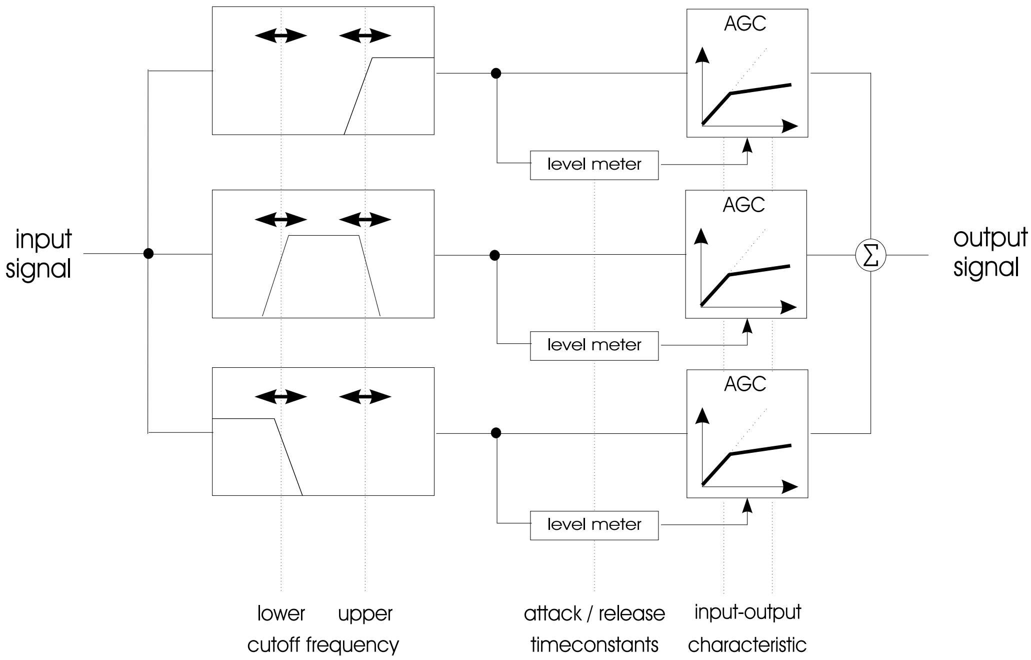

Figure 2.1: *Schematic overview of the 3–channel dynamic compression master–hearing–aid.*

The estimated band signal levels are then calculated from the respective summed intensities by applying a temporal first–order recursive averaging filter and subsequently transforming it into the dB–scale. The averaging filter allows for the definition of different attack and release time constants for each AGC channel, i.e., different adaptation times to rising and falling input levels as follows: In the first step, a peak hold with decay applies the release time constant $\tau_{rel}(c)$ when the input level decreases:

$$L_n(c) = \begin{cases} I_n(c) & \text{for} \quad I_n(c) \geq \tau_{rel}(c) \cdot L_{n-1}(c) \\ \tau_{rel}(c) \cdot L_{n-1}(c) & \text{for} \quad I_n(c) < \tau_{rel}(c) \cdot L_{n-1}(c) \end{cases} \tag{2.3}$$

Then gain $G_n(c)$ is calculated from an I/O–characteristic (see below) using $L_n(c)$ as the input parameter. Subsequently, gain $G_n(c)$ is smoothed by a first order IIR–filter[2] applying the attack time constant $\tau_{att}(c)$:

$$\overline{G_n(c)} = \tau_{att}(c) \cdot \overline{G_{n-1}(c)} + (1 - \tau_{att}(c)) \cdot G_n(c) \tag{2.4}$$

Within each frequency channel, an I/O–characteristics is defined, which prescribes the desired output level as a function of the estimated signal level on a log–log scale. The current gain in each band is then calculated from the respective input level using the I/O–characteristics and applied to the band signals. The output signal is formed by summing up the modified band signals. In this way, the frequency channels are compressed independently from each other.

[2] IIR: infinite impulse response

The I/O–characteristics uses 3 piecewise linear sections, i.e., 2 knee points. It is implemented as a table lookup with linear interpolation between the table entries. Figure 2.2 shows examples of the compression characteristics for different compression schemes.

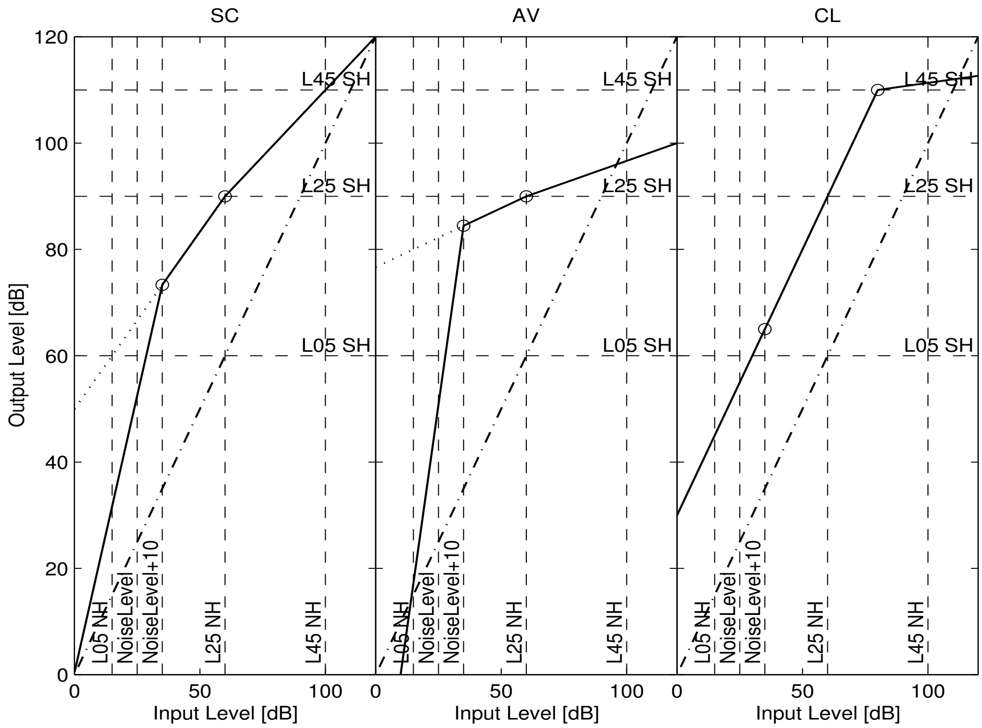

Figure 2.2: *Example of the I/O–characteristics (prescribed output level as a function of input level on a dB–scale) for one frequency channel of the dynamic compression algorithms SC, AV and CL. The solid lines show the compressive and expansive parts of the I/O–characteristics, kneepoints are shown by circles (○). The dotted lines indicate the continuation of the I/O–characteristic when no expansion was implemented. The dash–dotted line shows a linear system without amplification. The vertical and horizontal lines show the levels corresponding to 'very soft', 'medium' and 'very loud' (L05, L25 and L45, respectively) for normal–hearing subjects (NH, vertical lines) and hearing–impaired listeners (SH, horizontal lines), as well as the definition of the lower kneepoint at 10 dB above the noise floor of the device.*

Based on the 3-channel master compression algorithm described above, 4 different compression schemes (i.e., settings of its parameters) where defined, which represent fundamentally different approaches for dynamic range reduction. Note that all algorithms were adjusted in such a way that they should provide the same (average) amplification for a speech spectrum shaped signal at 65 dB SPL, (i.e., the level that corresponds to the average MCL for normal listeners for speech signals). The different schemes are described as follows:

Linear amplification (*LIN*)

The linear gain applied by the linear amplification scheme denoted as *LIN* is based on a mapping of the unaided *MCL* (most comfortable level, category 'medium' in the loudness scaling, see 2.5.1) to the average MCL of normal–hearing listeners. This linear frequency shaping was performed in 20 non–overlapping frequency bands with a bandwidth of 1 critical band (1 Bark). Since this scheme does not include dynamic compression, the patient is not protected from very loud sounds at high input levels. Therefore this scheme was used only within the laboratory experiments as a reference.

Compression limiting (*CL*)

This algorithm is based on algorithm *LIN* but additionally provide compression limiting at high input levels and therefore can protect the patient from too loud signals (see Figure 2.2, right panel). Compression limiting was realized using a compression ratio (i.e., the inverse of the slope of the I/O–characteristics) of 15, attack time constants of 1 msec and release time time constants of 50, 7 and 3 msec in the low-, band- and high–pass channel, respectively, and a compression kneepoint corresponding to the output level that matched the judgement 'too loud' in the loudness scaling experiment. This algorithm was one program selected for the tests in everyday life situations.

Syllabic compression (*SC*)

The syllabic compression used the same linear frequency shaping as *LIN* and *CL*, i.e., the same mapping of the unaided MCL to normal–hearing MCL. In addition, a compressive I/O–characteristic was implemented which matches for loudness at low (category 'very soft') and high (category 'very loud') levels. Since the high gain produced by the dynamic compression at low input levels would make the noise floor of the audio hardware audible to the patient, an expansive part was implemented in the I/O–characteristic at very low input levels (see Figure 2.2, left panel). The kneepoint for the expansive characteristics was set to be 10 dB higher than the noise level of the audio hardware in each channel. In addition, the compression ratio was set so, that input levels equal to the noise level of the audio hardware have the same amplification as the overall gain at normal–hearing MCL (L25 NH in Figure 2.2). Frequency–channel–dependent time constants were used which are short enough to follow the frequency of syllables (envelope compression) and long enough not to distort the signal's waveform. Therefore, the same attack and release time constants of 50, 7 and 3 msec in the low-, band- and high–pass channel, respectively, were chosen as in the compression limiting case. This algorithm was one program selected for the tests in everyday life situations.

Automatic volume control (*AV*)

The aim of this specific compression algorithm was to compress the input signal 'effectively' by a similar amount as algorithm *SC* but not to compress the temporal structure of the input signals in order to preserve the temporal contrasts. Since the 'effective' compression of speech signals is a function of the modulation transfer function of the compressive system and the modulation spectrum of speech [c.f., Hohmann and Kollmeier, 1995a; Plomp, 1988; Verschuure and Dreschler, 1996; Villchur, 1989], larger time constants have to be counteracted by a higher compression ratio in order to still yield the same 'effective' compression. Hence, the compression ratios individually obtained for each subject by algorithm *SC* were increased by a factor of 3 and longer attack and release time constants of 200 msec were used in all channels. The other parameters were chosen as for algorithm *SC*. This algorithm was one program selected for the tests in everyday life situations.

2.3 Subjects

Five sensorineurally hearing–impaired patients with mild to moderate sloping hearing loss were selected for the tests. The difference between air and bone conduction was less than 15 dB at all frequencies tested (0.5, 1, 2, 3 and 6 kHz). All patients are experienced hearing–aid users and were motivated as well as skilled enough to handle the prototype DSP hearing–aid. They received a nominal fee for their participation in the study.

At the beginning of the field test each subject participated in a complete routine audiological examination including pure tone audiogram, bone conduction hearing loss, determination of Uncomfortable Loudness Level (*UCL*) and impedance audiometry. The pure tone audiogram (*PTA*), the bone conduction hearing loss and the UCL was measured using an Interacoustics DA930 audiometer. Air conduction threshold was measured for each ear at frequencies of 0.125, 0.25, 0.5, 1, 2, 3, 4, 6 and 8 kHz. Bone conduction threshold and UCL was determined at 0.5, 1, 2, 3 and 4 kHz. Table 2.1 shows the data of the five participating subjects.

Subj	Ear	250 Hz	500 Hz	1000 Hz	2000 Hz	3000 Hz	4000 Hz	6000 Hz
BD	right	45/ -/ -	50/50/105	50/50/110	30/30/110	40/40/110	55/45/110	65/ -/ -
BD	left	55/ -/ -	55/55/110	55/50/105	45/45/105	70/55/105	70/60/110	80/ -/ -
EJ	right	45/ -/ -	55/55/nm	55/60/120	55/60/115	50/50/nm	50/50/nm	50/ -/ -
EJ	left	55/ -/ -	60/50/nm	60/60/120	45/60/110	50/50/110	45/55/nm	45/ -/ -
GH	right	75/ -/ -	70/55/105	70/55/105	70/65/110	70/70/115	75/70/115	90/ -/ -
GH	left	50/ -/ -	40/35/ 95	40/35/100	35/25/ 95	35/30/ 95	40/30/100	45/ -/ -
HM	right	25/ -/ -	35/40/ 90	55/50/ 95	55/55/ 90	45/45/ 85	60/60/ 90	65/ -/ -
HM	left	20/ -/ -	35/30/ 90	50/45/ 85	50/50/ 85	55/50/ 90	60/55/ 90	70/ -/ -
MW	right	50/ -/ -	55/45/ 95	70/60/100	80/65/100	70/55/100	70/55/100	70/ -/ -
MW	left	55/ -/ -	55/45/ 95	65/50/100	75/70/100	75/65/100	80/65/105	80/ -/ -
Mean		48/ -/ -	51/46/ 98	57/52/104	54/53/102	56/51/101	61/55/103	66/ -/ -
STD		16/ -/ -	11/ 9 / 8	9 / 8 / 11	17/15/ 10	14/11/ 10	13/11/ 9	15/ -/ -

Table 2.1: *Pure tone audiogram (air conduction thresholds), bone conduction hearing loss and uncomfortable loudness level in dB HL for all subjects. The mean value and standard deviation are given in the last two rows for all frequencies, respectively. Data denoted with nm could not be observed due to limitations of the audiometers output level.*

The mean hearing loss across all frequencies in table 2.1 was about 55 dB HL. It increased slightly with frequency and did not vary much among subjects. The UCL of all subjects was at about 100 dB HL for all frequencies between 0.5 and 4 kHz, which corresponds to the UCL of normal–hearing listeners. Hence, the hearing–impaired listeners showed recruitment with a residual dynamic range of about 40 to 70 dB.

All five subjects participated in the laboratory tests, but only four subjects (i.e., subject EJ excluded) participated in the field test.

2.4 Fitting of the Compression Schemes

In the first step of the fitting procedure, the linear gain is adjusted to obtain a mapping of the unaided MCL to the average MCL of normal–hearing listeners. The data required for this prescriptive step of the fitting was measured by monaural loudness scaling using narrow–band noise signals (see section 2.5.1). Figure 2.3 shows schematically the components of the wearable hearing–aid which determine the amplification of the device. The microphone

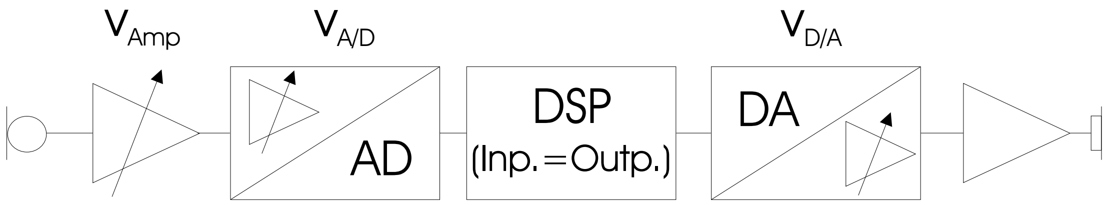

Figure 2.3: *Block diagram of signal flow in the DASi–2 [Raß and Steeger, 2000].*

amplifier of the *Cosmea M* ITE hearing–aids was set to its maximum value (vAmp = 17 dB) in order to obtain a maximal signal level on the rather long wires to the digital device. The analog attenuator $V_{A/D}$ before the *ADC* (analog to digital converter) was adjusted to yield a digital input level of -12 dB relative to the overload level of the ADC for a 90 dB SPL speech simulating noise input signal. This was done in order to shift the effective dynamic range of the ADC (approximately 90 dB) to common signal levels of everyday life situations. A first coarse fitting to an individual hearing loss was achieved by setting the attenuation of the DAC ($V_{D/A}$) to a value which ensures the desired amplification in the speech–relevant range of 500 to 4000 Hz. Fine tuning was done by setting further frequency–dependent attenuation in the digital domain in each frequency channel (critical band) or — if required to achieve the prescribed gain — on the Basis of FFT Bins. Afterwards, loudness scaling of narrowband noises was performed in the aided condition (center frequencies 0.5, 1 and 3 kHz) in order to verify the gain setting. When the MCL in the aided condition differed more than 10 dB from the average MCL of normal–hearing listeners, the amplification in the corresponding frequency channel was readjusted and the loudness scaling was repeated[3]. Finally, a loudness scaling with speech–shaped noise was carried out to adjust the overall level for broadband signals. It was set such that MCL in the aided condition is achieved at 65 dB SPL. Because broadband signals in general are judged louder than narrow band signals at the same physical level (due to spectral loudness summation), this correction led to a small attenuation of a few Decibels towards lower levels and was performed by attenuating the output of the speech processor digitally.

Within the field test the volume control switch on the digital device was programmed in a way that the user could increase the prescribed amplification by up to 6 dB and decrease it down to -18 dB. In all laboratory experiments the volume control switch was set to 0 dB.

In order to fit the dynamic compression algorithms described in section 2.2 to the individual hearing loss, the I/O–characteristics in the three frequency channels and the cutoff frequencies between these channels had to be adjusted. Since all subjects showed a relatively

[3]In general no readjustments had to be made here.

flat audiogram, the cutoff frequencies could be set to the same values of 703 Hz and 1828 Hz for all subjects. These cutoff frequencies provide a nearly equal distribution of the input (speech) energy to each of the three channels and were suggested by Kießling and Steffens [1991] for other three–channel dynamic compression systems.

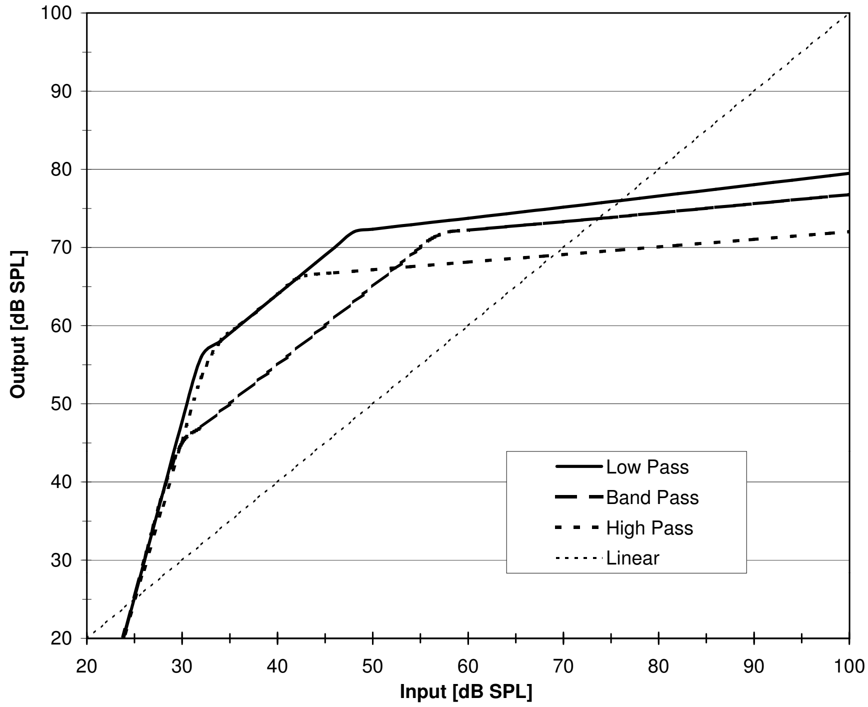

Figure 2.4: *Example of an actually implemented I/O–characteristics for algorithm AV in the three frequency channels adjusted for one test subject (MW, right ear). The I/O–characteristics for the low- and high–pass channel had to be limited to a maximum of 24 dB (maximum possible gain given by the implementation of the gain Table). The maximum gain of the band pass signal had to be further reduced to avoid feedback.*

The I/O–characteristics of the dynamic compression algorithms under study were determined by the loudness scaling data. However, because of the fixed point representation of the signal within the DASi's DSP, the maximum gain in the gain Table defining the I/O–characteristic could not exceed 24 dB. Therefore, the dynamic compression algorithms — especially algorithm AV – could not be fitted exactly according to their prescription. In addition, it turned out that the AV algorithm had a tendency to produce feedback at low input levels due to its high gain. In both cases, the maximum gain of the I/O–characteristics had to be reduced. Figure 2.4 shows an example where extreme limitations had to be done. It should be pointed out that these restrictions had not to be made in the laboratory study by Appell *et al.* [1995], where the same algorithms were tested on a stationary signal processing system.[4]

[4] In the laboratory study by Appell *et al.* [1995] the input levels of the stimuli were in any case well known, so the dynamic range of the system could be adjusted accordingly. In addition, all stimuli were presented via headphones which prevents feedback in the system.

Throughout the field test, four programs were stored on the DASi. Thus, the subject could easily switch between four hearing–aid algorithms. Table 2.2 lists the algorithms assigned to the four programs. In addition to the 3 processing schemes described here, the field test was carried out with one additional algorithm providing a combination of dynamic compression with a noise suppression algorithm. The results from this algorithm are reported elsewhere [Appell *et al.*, 1999].

Program Number	***Algorithm***
A	*CL*, compression limiting
B	*SC*, syllabic compression
C	*AV*, automatic volume control
D	*NR & DC*, combination of the 'best' noise reduction algorithm with the 'best' dynamic compression scheme.

Table 2.2: *Assignments of the DASi–2 program number and the hearing–aid configuration used for the field test.*

2.5 Assessment Methods

A battery of audiological tests was used in order to measure speech intelligibility, loudness perception and the system's sound quality as well as acclimatization effects. For the audiological classification of the subjects, at first a set of tests were performed for the unaided condition which includes standard audiometry, categorical loudness scaling experiments and speech intelligibility tests in noise. In the subsequent first laboratory test, the DASi–2 prototype hearing–aid was fitted to the individual subject's hearing loss as described in section 2.4. A series of loudness scaling and speech intelligibility tests were then performed for the aided condition. After these measurements, subjects tested the device in everyday life for at least 14 days in order to get accustomed to the 'new' hearing–aid. After this period, loudness scaling experiments with broadband signals and intelligibility tests were repeated in the laboratory. In addition, paired comparison quality tests were conducted and the subjects had to answer a questionnaire concerning their everyday experience with the hearing–aids in real life conditions.

2.5.1 Categorical Loudness Scaling

Loudness perception was individually measured with a categorical loudness scaling procedure described by Hohmann and Kollmeier [1995b] and Kollmeier [1997b] using a 11–category scale (i.e., 5 main categories 'very soft', 'soft', 'medium', 'loud' and 'very loud', plus 4 intermediate categories between them, plus the 2 limiting categories 'inaudible' and 'too loud'). The procedure estimates the loudness given in categories as a function of the signal level and was used for estimating loudness perception of several stimuli for unaided as well as for the aided condition. At first, loudness scaling was performed for prescribing linear amplification and compression I/O–characteristics. These unaided measurements used narrow band signals which were presented monaurally via headphones (Sennheiser HDA 200). All other loudness scaling experiments, including the fine tuning of the amplification, were carried out aided under free field conditions in a sound proof booth using one loudspeaker (FAR CR10–S) at a distance of 1 meter directly in front of the subject. For these measurements the ITE of the contralateral ear was switched off in order to achieve monaural testing. These latter experiments were carried out with several narrowband stimuli as well as with a broadband noise (speech–shaped noise).

2.5.1.1 Narrowband Noise Signals

The stimuli used for the narrowband loudness scaling were third octave–band noises centered at 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4 and 6 kHz of 2 second duration, including cosine ramps of 50ms. These stimuli were presented monaurally via headphones to individually obtain the data for the parameter prescription of the respective algorithms.

To evaluate and verify the effect of the different dynamic compression schemes on loudness perception, loudness scaling with the narrowband noise signals centered at 0.5, 1.5 and 3 kHz was performed directly after fitting of the processing schemes and were repeated at the end of the field test period.

2.5.1.2 Broadband Noise Signals

To evaluate the effect of the different dynamic compression algorithms on loudness perception of broadband signals, loudness scaling was performed with speech–shaped noise of 2 second duration, which was taken from the Göttingen Sentence Test material [Kollmeier and Wesselkamp, 1997]. Scaling was performed directly after fitting the parameters and repeated at the end of the field test period.

2.5.2 Adaptive Sentence Test

An adaptive sentence test (Göttingen Sentence Test) was used to measure speech intelligibility in speech–shaped noise. The sentence test is described in detail in Wesselkamp *et al.* [1992], Brand and Kollmeier [1996] and Kollmeier and Wesselkamp [1997].

In this study all speech intelligibility tests were performed under free field condition. The noise level in the aided conditions was set to 65 dB SPL which corresponds to the (signal–specific) average MCL of normal–hearing listeners. Note that for each subject all algorithms were adjusted to give the same loudness for this signal at this level. The noise level in the unaided condition was set individually to the subjects individual MCL, which was derived from the broadband loudness scaling in the unaided condition. For each condition to be measured, 20 sentences (two test lists) were used to determine the speech reception threshold (SRT, which is defined as the signal–to–noise ratio at which 50% word score is obtained). During the adaptive tests, the speech level was varied whereas the noise level was kept fixed.

2.5.3 Quality Measurements

The quality measurements should assess the preference across the processing schemes with respect to the subjectively assessed quality in a specific acoustical environment. The assessment was performed for speech in quiet (presentation levels 45 and 65 dB SPL), speech in cafeteria noise (presentation levels 45 and 65 dB SPL) and music (i.e., a segment of a pop–song at presentation levels of 45, 65 and 80 dB SPL).

The subjectively perceived quality of the processing schemes was measured in a complete paired comparison experiment, i.e., each scheme was compared to each other. The subject was asked to compare the 'overall quality impression' for two schemes on a verbal scale (for details on the instructions given to the subjects see appendix A.3 Figure A.4). The answers were transformed into scores according to Table 2.3.

verbal judgement	***score***
'A is much poorer than B'	2 points for B
'A is poorer than B'	1 point for B
'A and B are the same'	no points
'A is better than B'	1 point for A
'A is much better than B'	2 points for A

Table 2.3: *Verbal categories for rating the difference in overall quality between two processing schemes. For further analysis, these categories were transformed into the numerical scores given in this Table.*

For each comparison, the two programs to be compared were stored at two neighboring positions of the program switch of the DASi–2 prototype hearing–aid. The subject was allowed to listen to the test stimuli and to switch between the two programs at will.

2.5.4 Questionnaire and Interview

After long–term testing, the subjects had to answer a questionnaire concerning their everyday experiences while using the processing schemes in real life conditions. The questionnaire focused on speech intelligibility, sound quality and loudness. The questions were answered for each processing scheme that had been tested in everyday life. Each question concerned a specific subjective sound impression (like naturalness) and could be answered by choosing out of 5 response alternatives that range from a positive to a negative rating. The questionnaire is given in appendix A.3 Figure A.5.

In addition to the questionnaire, the subjects were informally interviewed about their experiences with the hearing–aid algorithms during the field test. The diary, which the subjects were instructed to keep for this purpose, was used as a starting point for the interview.

2.6 Results and Discussion

2.6.1 Loudness Scaling

Figure 2.5 shows the results of the narrowband loudness scalings measured shortly after the fitting of the processing parameters to the hearing–impaired subjects. Plotted are the mean values averaged across the five subjects (10 ears) of the equal–loudness levels at the categories 'very soft', 'medium' and 'very loud'. Average data of normal–hearing listeners taken from Hohmann and Kollmeier [1995b] are included as a reference. As expected from the fitting procedure, all algorithms do restore normal loudness at input levels corresponding to the normal MCL (loudness impression 'medium' for normal–hearing subjects). However, the two processing schemes exhibiting a linear characteristic within the relevant range of input levels (*LIN* and *CL*) were not able to restore normal loudness at low input levels. The full range dynamic compression schemes (*SC* and *AV*), however, could restore normal loudness for low and medium input levels. It can be seen from Figure 2.5 that algorithms *SC* and *AV* gave the same results in this experiment. Obviously, the effective compression for the signals presented was the same for both algorithms, i.e., the higher compression ratio of scheme *AV* is compensated for by the longer time constants. Hence, the design goal of approximately equating the effective compression ratio for algorithms *SC* and *AV* has been achieved.

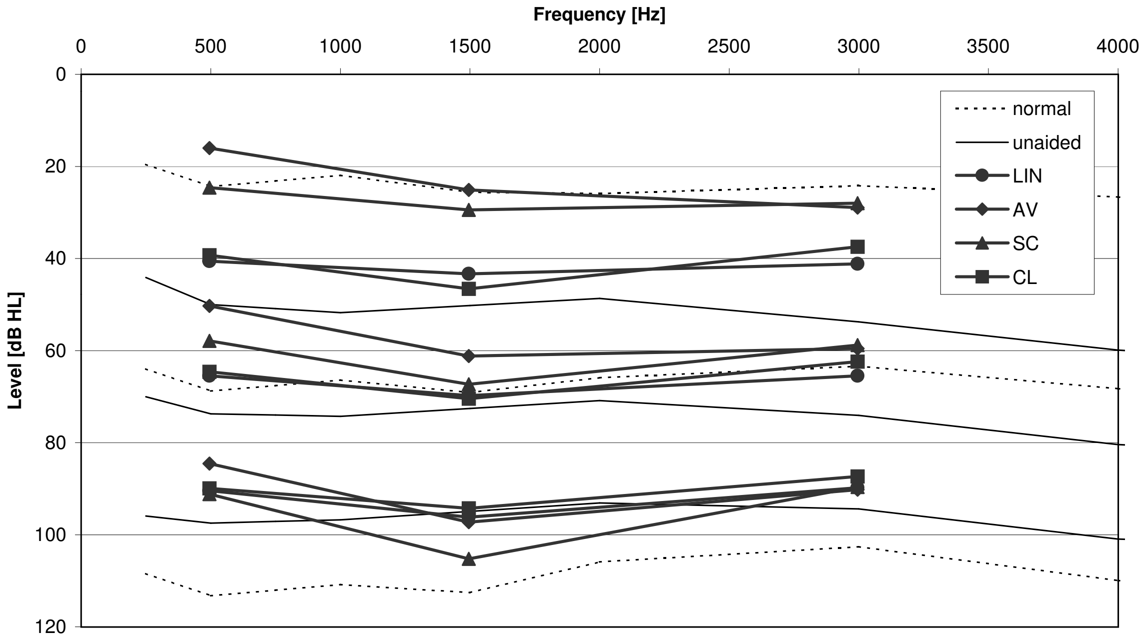

Figure 2.5: *Mean results of the narrowband loudness scalings with algorithms LIN, AV, SC and CL as well as in the unaided condition. Curves of equal–loudness for the loudness categories 'very soft', 'medium' and 'very loud' are plotted, respectively (mean values over the five hearing–impaired subjects). The dotted lines show respective normal–hearing data.*

Loudness scaling measurements according to those shown in Figure 2.5 were repeated for algorithms *CL*, *SC* and *AV* at the end of the field test (see appendix A.5, Figure A.6). The repeated scalings showed a consistent shift towards higher levels for all algorithms, i.e., the same signals were generally judged softer at the end of the field test than at the beginning of the field test. For low and medium levels this shift was about 5 dB. It amounts to 10 dB for high levels. This is most probably an effect of acclimatization.

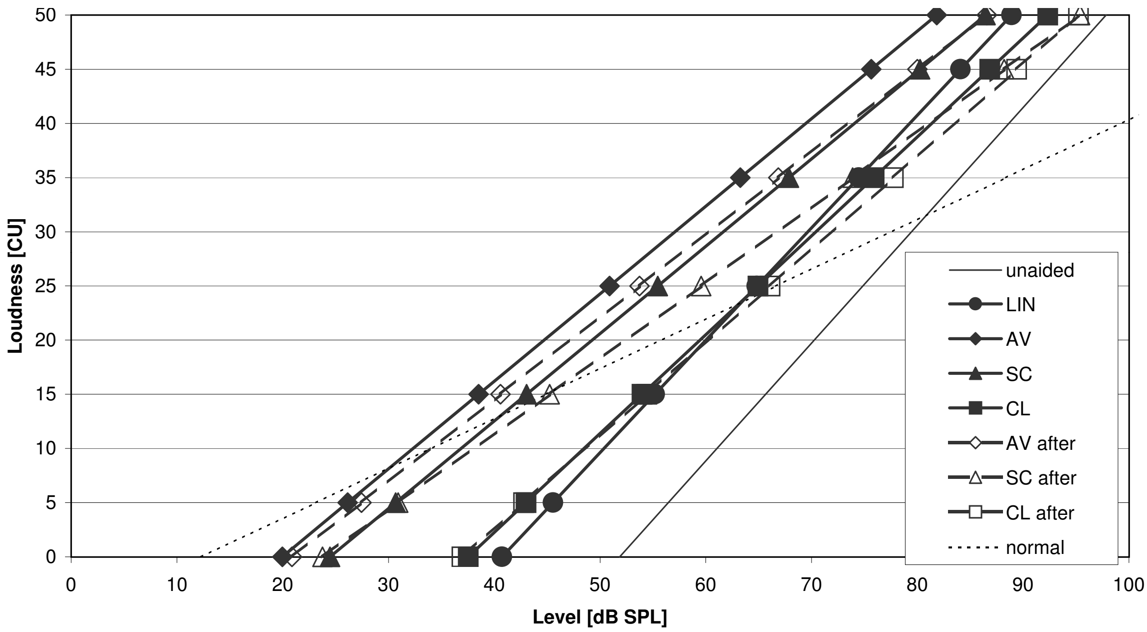

Figure 2.6: *Results of the broadband loudness scalings with algorithms LIN, AV, SC and CL as well as unaided. Each curve represents the mean of the individual linear functions fitted to the individual measurement points of the four hearing–impaired subjects before (solid lines, filled symbols) and after (dashed lines, open symbols) the field test. Symbols are shown for 0 CU='not heard', 5 CU='very soft', 25 CU='medium', 45 CU='very loud' and 50 CU='too loud'. The dotted lines show the respected normal–hearing data.*

The broadband loudness scalings (Figure 2.6) show a similar behaviour as the narrowband scalings. The prescription goal for algorithms *LIN* and *CL* is perfectly met, i.e., the match between normal–hearing and hearing–impaired MCL's. However, *LIN* and *CL* are only able to shift the hearing–impaired listeners' dynamic range without extending it. In contrast, *SC* and *AV* partially achieve their respective prescriptive goal to extend the hearing–impaired listeners' dynamic range. However, hearing–impaired listeners still perceived high levels too loud compared to normal–hearing listeners. The repetition of the broadband loudness scalings at the end of the field test (dashed lines with open symbols in Figure 2.6) again showed an effect of acclimatization especially at high levels. The data shows an increase of the subjects dynamic range by 5 dB for algorithm *AV* and 10 dB for algorithm *SC*, respectively. It is not quite clear why for algorithm *SC* this slightly larger acclimatization effect occurs.

2.6.2 Adaptive Sentence Test

The speech intelligibility measurements were carried out in speech simulating noise at a presentation level corresponding to the subjects individual MCL. Therefore, the presentation level was set to 65 dB SPL in the aided conditions and was adjusted to the subjects individual MCL (derived from the broadband loudness scaling) in the unaided condition. Hence, speech intelligibility was tested at the same overall loudness. The results are shown in Figure 2.7. Except for subject HM, all test subjects performed better unaided (i.e., without any change of the frequency spectrum and dynamic range) than aided with any of the algorithms. Because the test stimuli were presented at the same overall loudness and because of the comparatively flat hearing loss for most subjects, this could more or less be expected because the compression and the linear frequency shaping in the hearing–aid will not provide any additional benefit. On the contrary, the occlusion of the ear, the restricted receiver transmission quality (limited frequency range, nonlinear distortion) as well as hearing–aid noise (originating from the microphone, analogue circuitry and quantization noise) will limit the performance with hearing–aid considerably. Hence, speech intelligibility was best without hearing–aid at appropriate input levels. Only subject HM, who had the most sloping hearing loss (cf. Table 2.1), shows a benefit in speech intelligibility in the aided conditions.

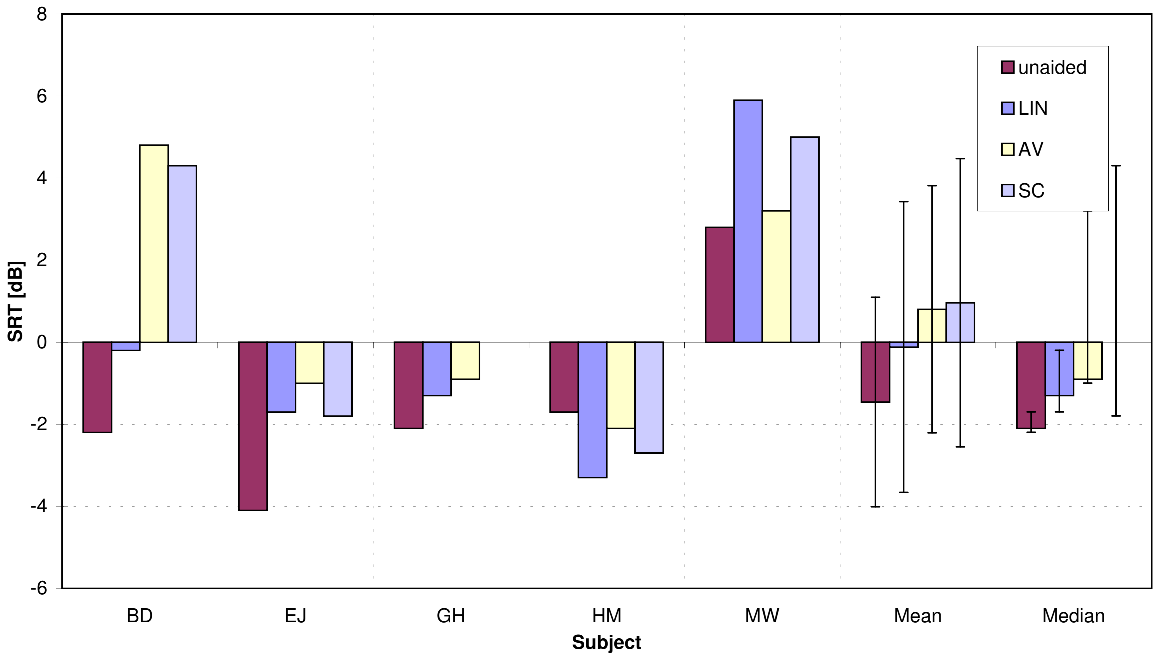

Figure 2.7: *Results of the speech intelligibility measurements in noise with processing schemes LIN, AV and SC as well as in the unaided condition. Given are the speech reception thresholds (SRT) for each subject with each algorithm. The two right columns show the mean and median data averaged across subjects, error bars denote plus minus one standard deviation and the interquartile range, respectively.*

For these reasons, the processing schemes were compared to each other and not to the unaided condition. Figure 2.8 shows the data from Figure 2.7 for schemes *AV* and *SC* relative to scheme *LIN*. No clear trend could be found in the performance of schemes *AV* and *SC* as compared to the reference. Subject BD performed best with *LIN*, subject MW performed best with *AV* and all other subjects performed similar with any of the processing schemes. This was more or less expected because all schemes were fitted to give the same amplification for the 65 dB SPL speech simulating noise presented in this experiment. It can

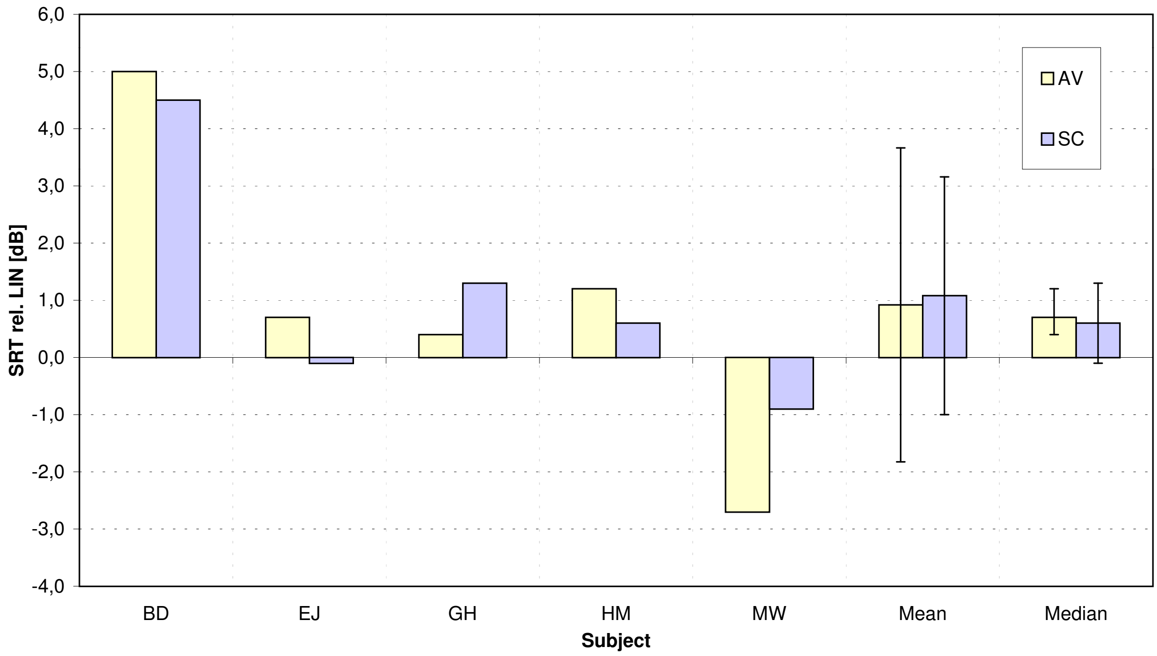

Figure 2.8: *SRT's for algorithms AV and SC plotted relative to the SRT observed with algorithm LIN. The data were calculated from data shown in Figure 2.7.*

be concluded that the dynamic compression algorithms does neither improve nor deteriorate speech intelligibility in noise at medium levels as compared to linear processing. On the other hand, at low signal levels, the dynamic compression algorithms certainly will outperform algorithm *LIN* just due to audibility.

2.6.3 Quality Measurements

The perceived quality of the processing schemes was measured in a complete paired comparison experiment. Quality scores were calculated as a sum of the numerical values according to Table 2.3 across all paired comparisons for each condition, respectively. The results are given in Figures 2.9 to Figure 2.11. The median values are shown together with the interdecile ranges I_{80}, which are a dispersion measure covering 80% of the distribution of scores. Because each of the four processing schemes was compared to each other, a maximum score of 6 could be reached for a scheme that was judged as 'much better' in all comparisons (3 comparisons * 2 score per comparison), whereas a score of zero would indicate that this scheme was never judged better in quality as compared to any other scheme.

Figure 2.9 shows the results sorted by presentation Level and averaged across subjects and test stimuli. It can be seen that the dynamic compression schemes (*AV* and *SC*) get higher scores as compared to the linear schemes (*LIN* and *CL*) at low and medium input levels. This holds especially for low levels and moreover for *AV*, whereas the differences among the schemes become smaller with increasing levels. At high levels only one condition was tested ('music' at 80 dB SPL) and no clear difference between the algorithms is observed.

A more distinguishing view on the results is provided by Figure 2.10, which shows the results for each of the test stimuli 'speech in quiet', 'speech in cafeteria noise' and 'music' averaged across subjects and presentation levels. Here *SC* showed good performance for all

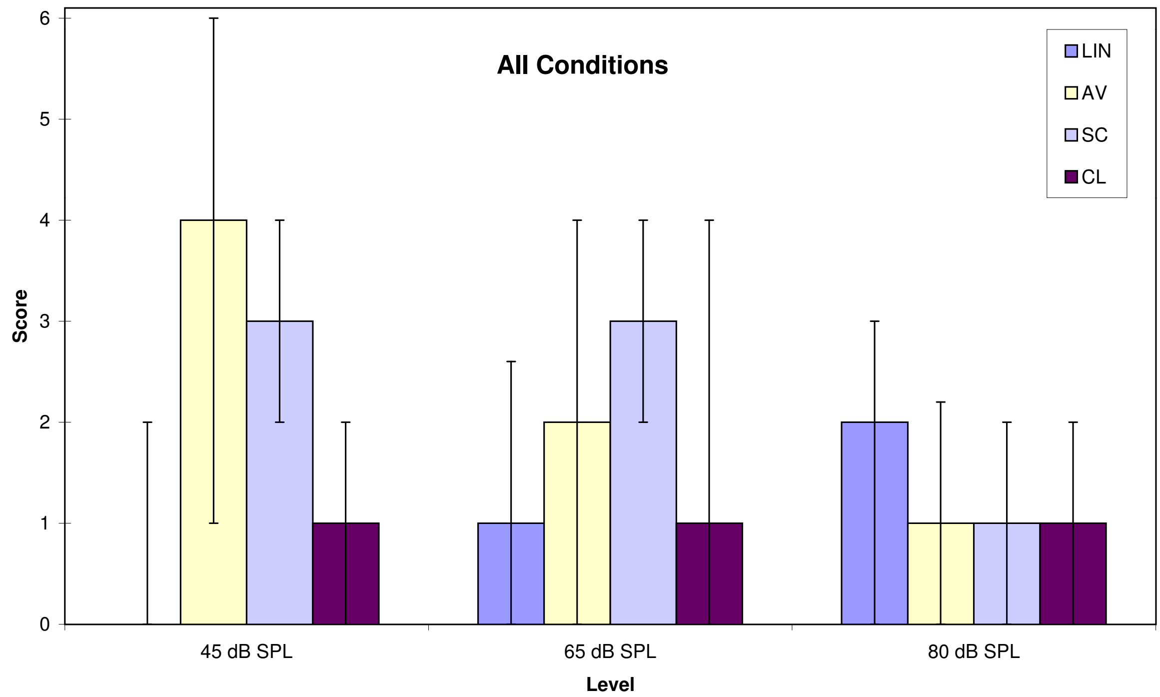

Figure 2.9: *Results of the paired comparison test sorted by presentation level and averaged across subjects and test stimuli. Given are the median scores with the interdecile range I_{80} of processing schemes LIN, AV, SC and CL for each presentation level: 45 dB SPL (15 observations), 65 dB SPL (15 observations) and 80 dB SPL (5 observations).*

conditions, whereas *AV* reached very good scores for 'speech in quiet' and 'music' at low levels but its performance decreased for stimuli which employ impulsive noise bursts ('speech in cafeteria noise') at higher levels. Focusing on the linear schemes *LIN* and *CL* it can be seen, that their performance in general is poorer as compared to the dynamic compression schemes. An exceptional condition is 'music' at high levels, where *LIN* gets the highest score. The reason for this might be the subjects acceptance of high levels in this specific situation and that *LIN* provided the highest output level in this situation. Another exception to the in general poorer performance of the linear schemes is given for *CL* in the 'speech in cafeteria noise' situation at the highest level. Here, the ability of *CL* to effectively limit high level peaky sounds seems to be advantageous for some subjects.

In order to analyze individual differences in the quality judgements, Figure 2.11 shows the results for each subject averaged across presentation levels and test stimuli. It can be seen that subjects BD, EJ and HM did not show a clear preference for any of the processing schemes, whereas subject GH shows a clear preference for *AV* and subject MW for both dynamic compression schemes (*AV* and *SC*). The latter may be influenced by the restrictions which had to be made with regard to the I/O–characteristic of algorithm *AV* for subject MW, i.e., subject MW might have given higher scores for algorithm *AV* if the I/O–characteristic

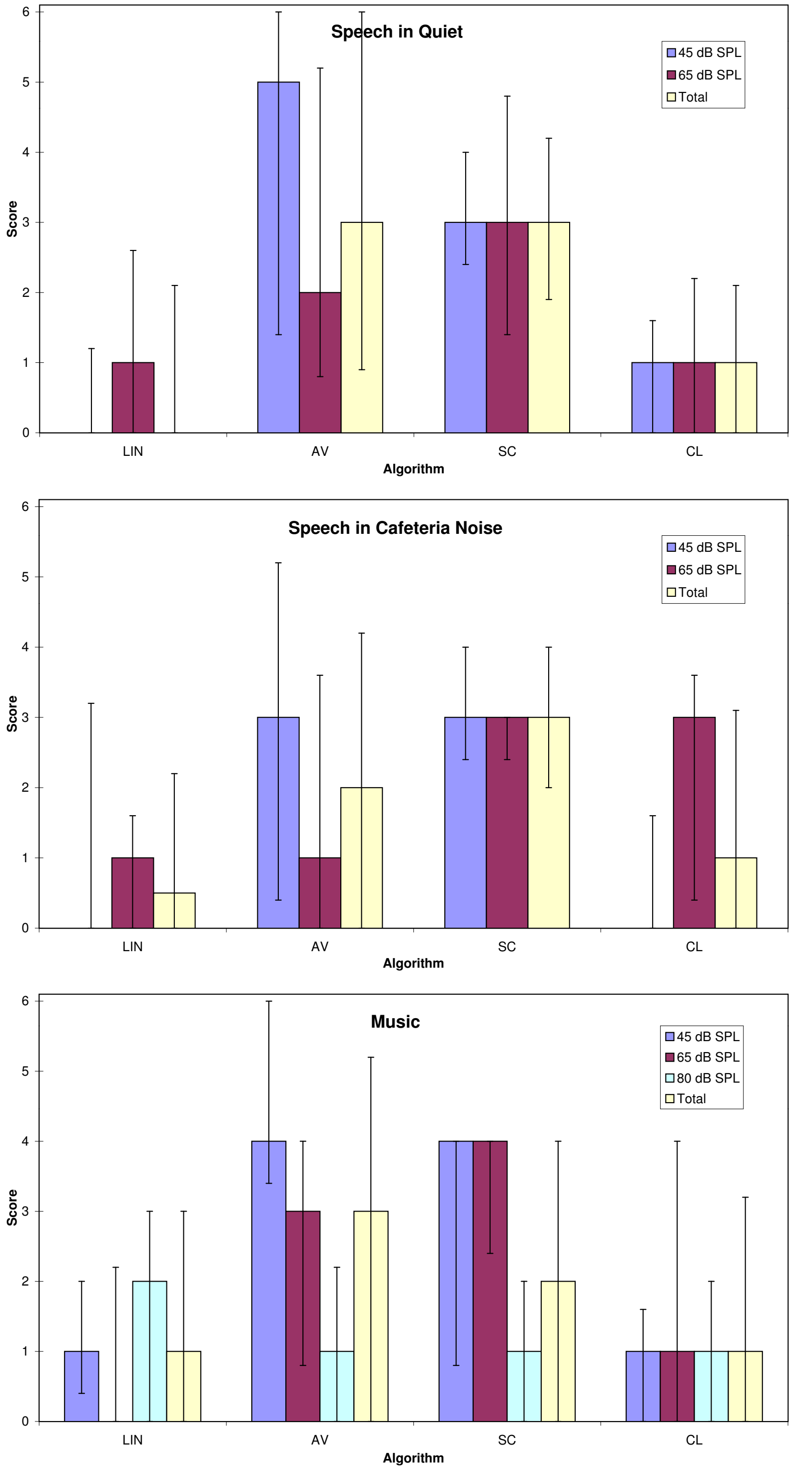

Figure 2.10: *Results of the paired comparison test for 'speech in quiet' (top), 'speech in cafeteria noise' (middle) and 'music' (bottom). Displayed are the median scores with interdecile range I_{80} across the 5 subjects for the processing schemes LIN, AV, SC and CL, respectively, subdivided into the presentation levels. For each scheme the rightmost bar denotes the median across all subjects and all presentation levels.*

could have been fitted exactly to its prescription (see Figure 2.4). The individual results for subjects GH and MW can be explained by the fact that their residual dynamic range was the smallest within the group of subjects (c.f., Table 2.1). The effect of compression was therefore expected to be most pronounced in these subjects.

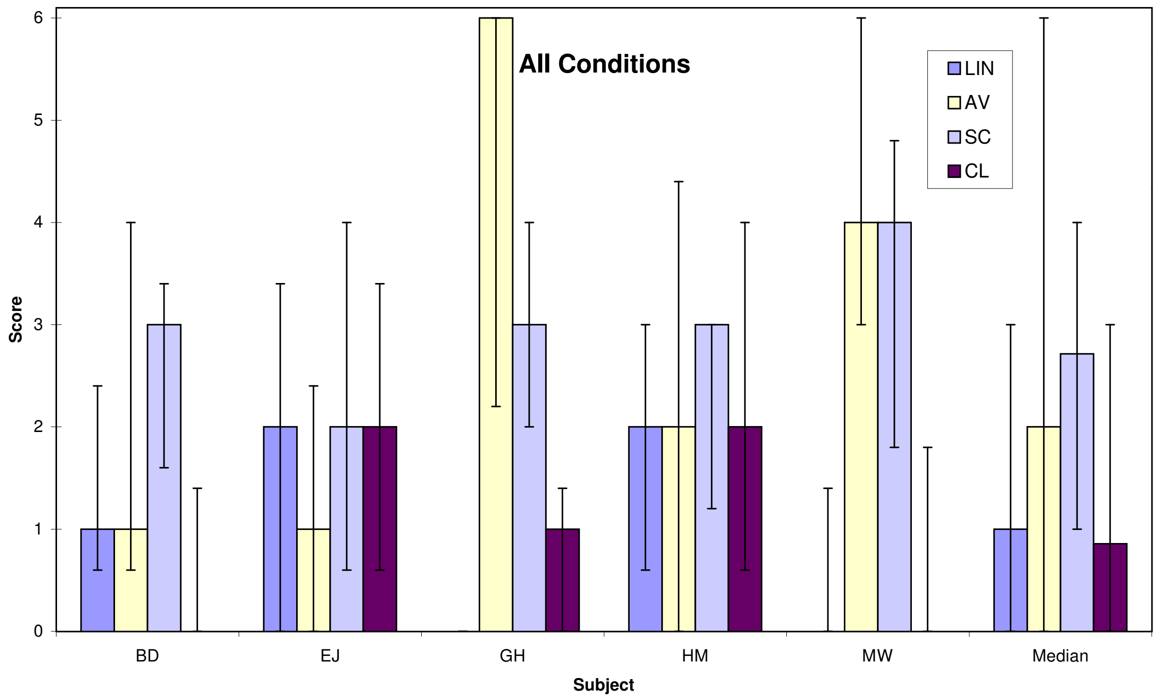

Figure 2.11: *Individual results of the paired comparison test. Given are the median scores with interdecile range I_{80} averaged across all stimuli and level combinations (7 observations) for processing schemes LIN, AV, SC and CL(7 observations per algorithm). The rightmost bars denote the median over stimuli, levels and subjects for each algorithm (35 observations per algorithm).*

Taking together the data for all stimuli, levels and subjects (Figure 2.11, bars on the right) only small differences between the algorithms could be found.

2.6.4 Questionnaire

After having tested the different processing schemes in real life conditions the subjects filled out the questionnaire described in section 2.5.4. To evaluate the results, the judgements were transposed from the verbal scale into scores varying from 0 to 4, where 0 was the most negative and 4 the most positive rating of the algorithm in each question. The results are given in Figures 2.12 to 2.15.

Subjective speech intelligibility was generally judged better for the dynamic compression schemes (AV and SC) than for scheme CL (Figure 2.12). The best results were obtained for AV. This is found consistently for all listening conditions. Generally the rating for ‘speech in noise’ decreased slightly as compared to the other situations for all schemes, which can be

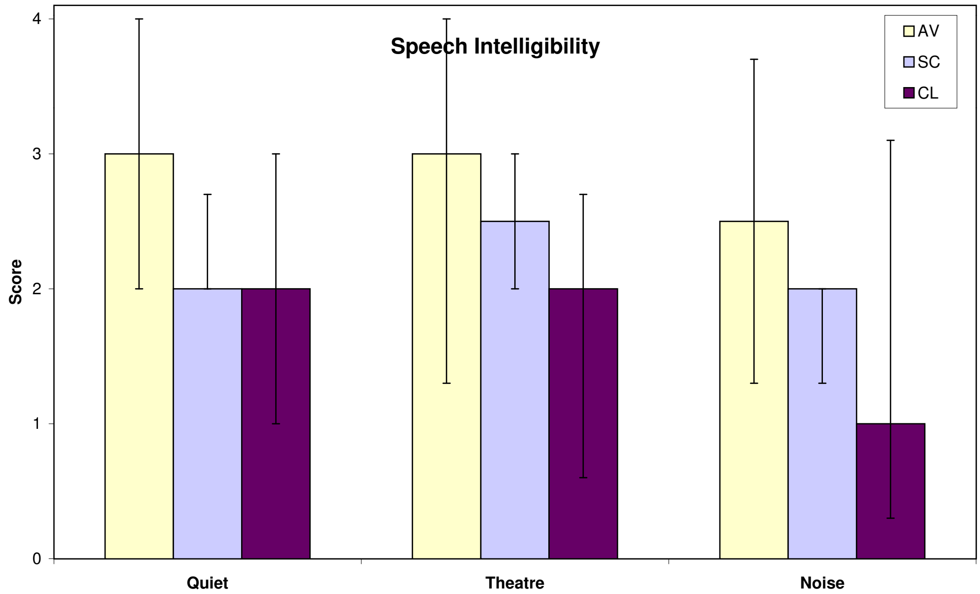

Figure 2.12: *Subjective judgements concerning speech intelligibility in real life conditions regarding the distinguished criteria 'Speech in quiet', 'Speech in theater or lecture' and 'Speech in noise' (questionnaire assessment). For each situation the median scores across subjects for the processing schemes AV, SC and CL are plotted together with their interdecile range I_{80}.*

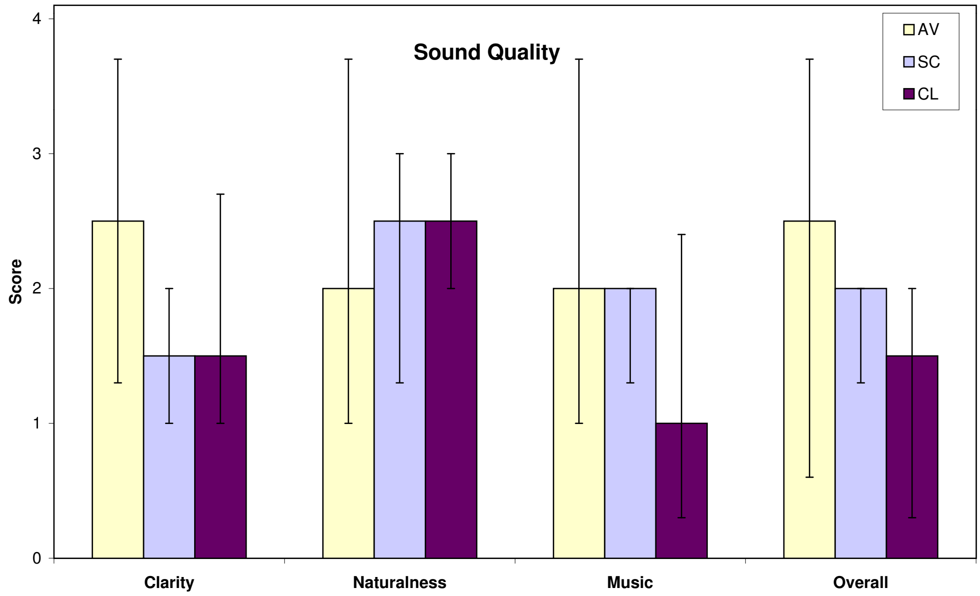

Figure 2.13: *Subjective judgements concerning sound quality in real life conditions regarding the distinguished criteria 'naturalness of sound', 'clarity of sound', 'sounding of music' and 'overall sound quality' (questionnaire assessment). For each situation the median scores across subjects for the processing schemes AV, SC and CL are plotted together with their interdecile range I_{80}.*

explained by the hearing–impaired's principle difficulty to understand speech in background noise.

The judgements concerning sound quality (Figure 2.13) show no clear preference for one of the processing schemes but there is a tendency for higher ratings for dynamic compression schemes, especially for *AV*. However, the data show good performance for *CL* in the rating of sound naturalness. Thus, it can be assumed that linear processing comes closest to the subjects impression of how their acoustical environment normally sounds. But the better results for *AV* in sound clarity show that the processing scheme providing the most natural sound is subjectively not automatically the scheme that gives the highest sound clarity and moreover the best speech intelligibility as can be seen in Figure 2.12.

Figure 2.14: *Subjective judgements concerning loudness in real life conditions (questionnaire assessment). Loudness values -2, 1, 0, 1 and 2 correspond to the loudness impression 'much too soft', 'too soft', 'ok', 'too loud' and 'much too loud', respectively. The rightmost bars denote the median across all subjects.*

Figure 2.14 gives the median subjective loudness judgements from the questionaire for all subjects and processing schemes. It reveals that the loudness judgements at the end of the field test agree with the results of the loudness scaling (see 2.6.1). Algorithm *CL* was judged rather too soft, *AV* rather too loud, whereas *SC* was judged as comfortable. On the other hand there are clear differences in the individual judgements. Algorithm *AV* provided the correct amplification for subjects GH and MW, whereas *CL* was best for subjects BD and HM. Like it was argued in section 2.6.3, the preference of subjects GH and MW for processing scheme *AV* can be explained by the fact that their residual dynamic range was the smallest within the group of subjects and as it can be observed from their individual loudness scaling data, algorithm *AV* is best capable to widen their dynamic range to the dynamic range of normal hearing.

These individual differences were also found when taking the median across all speech

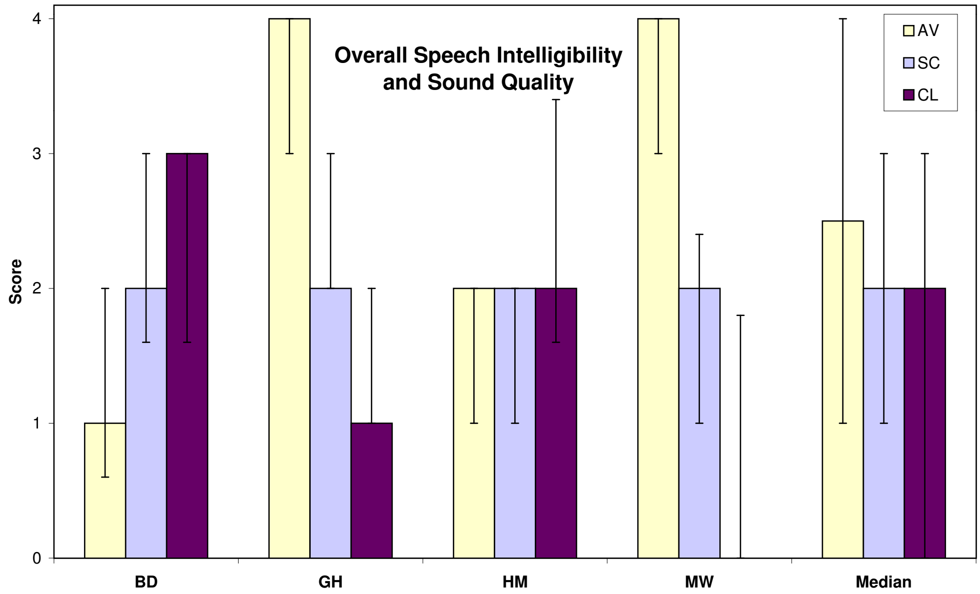

Figure 2.15: *Median values across all speech intelligibility and sound quality judgements for each subject and each processing scheme (questionnaire assessment). The rightmost bars denote the median over all subjects.*

and sound quality judgements for each subject individually[5]. Again, algorithm *AV* was best for subjects GH and MW, whereas *CL* is best for subjects BD and might be best for subject HM.

Taken together, the results of the questionnaire assessment suggests, that scheme *SC* is a good compromise between the schemes *CL* and *AV* (Figures 2.12 and 2.13). But the individual data shown in Figures 2.14 and 2.15 reveal that there are large inter–individual differences between the subjects. It can be concluded that some subjects in general prefer algorithm *AV* whereas other subjects prefer *CL*. This finding is supported by the observation of larger error–bars in the results for algorithm *AV* and *CL* as compared to the error–bars for algorithm *SC*.

2.6.5 Subjective Assessment (Interview)

In the following, some of the results of the interviews with the test subjects after the field trial period are summarized. Most subjects reported that they had no problems handling the device but complained about the inconvenience caused by wearing the DASi-2 prototype hearing–aids. Especially the cable connections between the ITE devices and the speech processor are disliked.

In accordance with the results given so far, subject BD liked most scheme *CL*, whereas scheme *AV* was judged as being too loud and exhibiting too many background noises. Scheme

[5]It is critical to take the median across data coming from questions pointing at quite different attributes, which –in addition– are answered on different verbal scales. But here we just want to clarify if there are inter–individual differences in the judgements which are consistent among all these attributes.

SC was preferred in case of unintelligible speech in television. Subject HM did not like the DASi-2 prototype hearing–aids at all. Among all schemes HM liked *CL* most because other voices were most intelligible. Like subject BD, she complained about loud background noises produced by scheme AV. In addition, she was annoyed about the tendency to feedback exhibited with scheme AV.

On the contrary, subjects GH and MW remarked that the loudness of scheme *CL* was much too soft for most real life situations. Both subjects liked AV most and preferred the use of AV especially in theater and cinema. Subject GH denoted the excellent sound of AV which was quiet similar to his own hearing–aids. Subject MW preferred AV because it gave the best speech indivisibility. MW voted *SC* for giving the best sound quality because its sound was the softest.

2.7 Summary and Discussion

A battery of tests was performed with five hearing–impaired subjects using a prototype digital hearing–aid in order to compare different 3–channel dynamic compression algorithms. Due to the carefully selected control conditions (i.e., unaided situation at roughly the same presentation level as in the aided situations, same frequency response for all algorithms for a medium input speech level with a speech–shaped input signal) the differences across algorithms were only very small. Hence, no 'winner algorithm' can be derived from the current data. However, the controlled experimental design (battery of tests with a set of algorithms fitted with the same fitting rationale) provides information for a variety of aspects that may be relevant for other dynamic compression schemes as well:

Loudness scaling

Loudness scaling data with narrowband stimuli revealed that all algorithms achieved their main fitting goal, i.e., approximately restoring the loudness contour 'medium' for narrowband signals across different frequencies. For low input levels, the linear algorithms (including the compression limiting algorithm) were not able to provide sufficient amplification. For high input levels, on the other hand, all algorithms provided too much amplification, even though the compression algorithms in principle should exhibit less amplification here than the linear algorithms. A similar finding can be derived for the broadband loudness scaling (cf. Figure 2.6): While the algorithms provide their correct gain at medium levels, they do on the average not provide enough compression, i.e., they show too little amplification at low input levels and too high amplification at high input levels. Although the compression algorithms perform better in this respect (by providing an enlarged input dynamic range of approximately 15 dB immediately after fitting and 5–10 dB more after acclimatization), they provide too much amplification at medium levels and do not provide enough dynamic compression to map the whole dynamic range of normal listeners into the remaining dynamic range of the hearing–impaired patients considered here. One reason for the insufficient amplification of the dynamic compression algorithms for low levels is the feedback problems encountered in the real–time processing at high insertion gain values, which forced to limit the maximum achievable gain by the wearable device. In this respect, the study with the wearable device differs considerably from the original experiments performed by Appell *et al.* [1995] using the same dynamic compression algorithms on a laboratory computer setup.

As already noted above, a certain acclimatization effect [according to Gatehouse, 1992] was observed after 4 to 6 weeks of using the wearable prototype hearing–aid in daily life for the compression algorithms. It yields a 5–10 dB extension of the input dynamic range. However, it is not clear why the syllabic compression algorithm received a higher acclimatization effect than the automatic volume control algorithms (AV), even though this difference is comparatively small (5 dB).

Sentence test

In general, no improvement in sentence intelligibility in noise could be achieved with any of the processing schemes (including the linear amplification) implemented on the wearable device. This finding is in line with most reports from literature [Harten-de Bruijn *et al.*, 1996; Festen, 1999; Goedegebure *et al.*, 1996b; Hohmann and Kollmeier, 1995a; Lunner *et al.*, 1998; Stone *et al.*, 1997; Verschuure *et al.*, 1998; Walker and Byrne, 1984], including own laboratory experiments, and in disagreement with few reports about intelligibility improvements for certain dynamic range compression algorithms [Benson *et al.*, 1992; Moore *et al.*, 1999c]. The main reason for this absence of a measurable benefit is the carefully selected reference condition that provides some overall level adjustment to compensate for the 'attenuation' component of the hearing loss (see Results section). Even if the implemented algorithms are compared against each other, no systematic improvement or deterioration could be achieved with the dynamic compression algorithm in comparison with the linear amplification. This is probably due to the fact that the compression algorithms did not provide any further audibility of signal components in comparison with the linear condition. This contrasts with some of the studies from the literature that report a positive effect on speech reception thresholds both in quiet and in noisy conditions. A larger difference might have occurred at lower presentation levels (where audibility limits the maximum performance both in quiet and to a lesser degree in noise) and for comparatively high presentation levels (where distortions introduced by the hearing–aid system and by peak clipping processing may limit the maximum intelligibility). However, the amount of intelligibility measurement blocks was limited by the number of available test sentences. Hence, a systematic intelligibility evaluation could not be performed across a large dynamic speech range within the current study. This contrasts to the previous study Appell *et al.* [1995] where a larger range of test levels was employed and in general algorithm AV performed better than the others.

One effect that cannot be avoided in the current experiments is that both the target speech and the interfering noise signal control the dynamic compression circuit. Hence, the detrimental effect of the noise may even be enhanced for low signal–to–noise ratios (SNR's) since the overall gain may fluctuate synchronously with the noise fluctuations, while this effect is negligible at high (positive) SNR's. As a consequence and already noted by Verschuure *et al.* [1998], the SNR corresponding to the individual subject's unaided speech reception threshold (SRT) influences the amount of this detrimental effect and may hence directly influence the potential benefit obtained with a dynamic compression algorithm in speech intelligibility tests. This may be the reason why Subject MW (who showed the highest unaided SRT, the smallest residual dynamic range and required the highest compression) performs systematically better with the compression algorithms when compared to algorithm *LIN*.

Quality judgements

The paired comparison test performed with various acoustic materials at three different levels clearly shows that compression (especially with algorithm *AV*) is preferable for low input levels, while at medium levels the syllabic compression seems to be advantageous. At higher levels no differences were found. Of course, any of these algorithms (especially the linear algorithm) would need a peak clipping or compression limiting algorithm to prevent the user from too high output signals or sound peaks.

If the quality judgements are compared across subjects, it is striking that the subjects with the smallest dynamic ranges (i.e., Subjects GH and MW) did profit most from the dynamic compression algorithms. For them, algorithm *AV* gave the best overall scores, although it reacts comparatively slowly and hence cannot protect the user against sudden loud sounds in a very efficient way. The results from the sound quality judgements (obtained by paired comparisons) are consistent with those from the questionnaire, where again *AV* performed best for speech (but no clear preference is given with respect to sound quality). On the other hand, the questionnaire results with respect to loudness, behave like the loudness scaling data: While the algorithm *CL* is judged to be rather too soft, *AV* is judged to be a bit too loud. Again, the differences are quite small, especially if the large individual differences across subjects are considered that can be viewed from the error bars (cf. Figure 2.14). These individual differences are at least partly due to the differences in the residual dynamic range: As above, subjects with the smallest residual dynamic range (GH and MW) judged algorithm *AV* to be best, while the other subjects preferred the compression limiting algorithm. The results from the informal interview in general coincide with the data from the other experiments.

In general, the results of the quality comparison show that for a real–world hearing–aid high amplification at low input levels (as provided by algorithm *AV*) is advantageous and should be combined with a strategy with a better compensation of the individual loudness deficits for medium and high levels using a compression strategy such as the syllabic compression. In addition, an effective reduction of sudden noise bursts at higher levels (provided by the compression limiting system) seems to be advantageous. Especially patients with a strongly reduced dynamic range appear to profit from dynamic compression and judge these algorithms on the average to provide a higher signal quality than the linear amplification.

Although the general approach and fitting rationale behind the algorithms were to compensate for abnormal loudness perception in hearing–impaired listeners (at least within an intermediate level range), loudness compensation (most thoroughly provided perhaps by algorithm *SC*) does not necessarily lead to a higher speech intelligibility when compared to linear amplification and to automatic volume control at a certain speech level. This effect may well be due to our limited knowledge about loudness perception in hearing–impaired listeners and its relation to speech intelligibility in quiet and noise which calls for further research in this area. Another effect is that the field test results obtained here cannot directly be compared with laboratory tests due to several restriction of a real–world hearing–aid (such as, e.g., limited feedback margin, limited frequency range and distortions and noise introduced by the hearing–aid setup). However, since the ultimate goal of dynamic compression algorithms is to provide a user benefit in wearable hearing–aids in the real world, field tests with such wearable devices are indicated for testing the performance of new algorithms.

Even though the results from this study are not clear–cut, it still provides interesting general findings that may be usable for the design of future hearing–aids:

— No clear–cut preference for a single dynamic compression algorithm against its competitors is observable at medium input levels as long as some basic requirements are met (i.e., match of overall frequency shape and overall gain at intermediate levels). Hence, no strong arguments can be made about the benefit of one algorithm over the other, which may explain why there has not evolved a single, optimum solution for commercial digital hearing aids so far. However, from the current study it can be suggested that for low input levels a slow acting compression with a high compression ratio (i.e., automatic volume control) should be used to provide audibility at this specific input level range, whereas syllabic compression (small compression ratio) or even linear amplification seems to be beneficial at medium to high input levels [comparable results are also found by Maré *et al.*, 1992]. Such a reduction of the compression ratio with increasing input level is supported by the fact that sensorineural hearing–impairment is combined with a loss of the compressive nonlinearity (outer–haircell damage) on the basilar membrane which especially compresses weak sounds. In addition I/O–characteristics derived from loudness models for stationary sounds do show the same effect (see chapter 6). However, in any case compression limiting should be provided to prevent from high level signal peaks.

— 'Individual' differences across subjects (such as, e.g. different residual dynamic range and other audiological features of the hearing loss) as well as some practical limitations of the hearing–aid tested (such as, e.g., limitation in the maximum gain and frequency response and fidelity of the transducer) may play a more important role in some cases than the exact choice of the dynamic compression algorithm. Since in the current study the effect of these additional factors were eliminated to a certain degree (i.e., the same hardware was employed with the same subjects and a very similar fitting rationale), some effects of the dynamic compression algorithm could be detected. However, as soon as the balancing of the additional parameters becomes incomplete (i.e., if two hearing–aids from different manufacturers are compared with each other), the effect of the algorithm implemented in the respective hearing–aids can no longer be studied in isolation from the other factors. This clearly limits the validity of comparative field tests that try to compare the user benefit obtainable with different hearing–aid hardware and fitting rationales.

Chapter 3

Comparison of Loudness Matching and Loudness Scaling for the Measurement of Spectral Loudness Summation

Abstract

Normative data for spectral loudness summation (i.e., the level difference for a narrowband and a broadband sound that are judged to have the same loudness) is important both for perception theories (e.g., loudness models) and loudness restoration schemes in hearing aids. In an attempt to obtain bias–free reference data, the current work tests the method of loudness matching and the method of loudness scaling to determine reference data of normal–hearing listeners for spectral loudness summation. Although the results of the loudness matching experiments show small intra–individual differences within a single measurement setup, the subjects responses depend markedly on procedural details. Furthermore, the inter–individual differences observed in the loudness matching and the loudness scaling experiments were only slightly smaller than the measured quantity (i.e., the amount of spectral loudness summation) whereas the intra–individual variability was small. It is shown that the inter–individual differences in the perception of spectral loudness summation are to some extent independent from the measurement method. Thus, we conclude that loudness summation strongly depends on the individuum, which has to be assumed for hearing–impaired as well. Hence, normative data for spectral loudness summation is only meaningful if a measurement method without a bias is employed and if the individual variability is reported in addition to the mean value. The applicability of normative loudness summation data derived in such a way, e.g., for the individual fitting of a hearing aid, will have to consider the variability in the data.

3.1 Introduction

Loudness perception is a very important issue both for theories on hearing and for practical applications in rehabilitative audiology.

Most people suffering from cochlear hearing loss show loudness recruitment [Brunt, 1994; Fowler, 1936; Steinberg and Gardner, 1937]: The absolute threshold for detecting sounds is higher than normal and once the sound level is increased above threshold, loudness increases more rapidly than in normal–hearing listeners with the consequence that the perceived loudness is about the same at high levels when compared to normal–hearing listeners [Moore, 1995]. If the alterations in loudness perception found in hearing-impaired listeners should be quantified and compensated for by hearing aids, two questions are of interest. First: which measurement method is able to quantify the effect? Because such a method may also be of interest for hearing–aid fitting, it should be time efficient and easy to handle for the subjects. Second: Is the measurement method capable of generating reference data for the normal–hearing system when normal–hearing subjects are under test? The latter is important because in relation to hearing–aid evaluation and fitting, loudness data obtained from hearing–impaired people is compared relative to normal reference data, either by directly comparing unaided and aided loudness data to the normative reference or indirectly by setting up a fitting target based on normal reference data. Elberling [1999] even argues that the individual differences in loudness perception are so large that it is useless to measure the individual hearing–impaired loudness impression for a prescriptive hearing–aid fitting.

The current work focuses on both questions in the context of spectral integration of loudness (loudness summation). The amount of spectral loudness summation, i.e., the level difference between two stimuli with different frequency bandwidth that produce the same loudness impression, is investigated with normal–hearing listeners. Two loudness measurement methods are tested: the method of loudness matching and the method of categorical loudness scaling. The experiments are conducted with two sets of bandlimited noise stimuli: A uniform exciting noise [Zwicker and Fastl, 1990], which due to its spectral shape should produce a higher effect of loudness summation than bandlimited white noise and low–noise noise [Hartmann and Pumplin, 1988; Pumplin, 1985] which is optimized for low temporal fluctuations.

The method of *loudness matching* asks for a direct adjustment of the equal loudness–level for one stimulus (test stimulus) when compared to another stimulus (reference stimulus). If both stimuli differ in bandwidth, the amount of loudness summation is then calculated from the difference between the adjusted equal loudness–level and the level of the reference stimulus. Several authors found their results in loudness measurements influenced by methodological factors [see for a review Verhey, 1998]. Biases are reported especially when the difference among the physical parameters of the two sounds in comparison is increased. For example Gabriel [1996] and Marks [1994] found their results affected by the range of levels employed for the test stimulus.

In the context of spectral loudness summation, the difference in the physical parameters of the two sounds in comparison increases when the difference in bandwidth between the test and the reference stimulus is enlarged. However, only few publications considered the influence of the reference bandwidth in loudness matching experiments with band–limited noise signals. Verhey [1998] for example, tested the influence of the reference bandwidth by selecting the bandwidths of the two reference stimuli close to the edges of the test stimuli bandwidths, with the consequence that the bandwidth of the reference and test stimulus differs markedly for some combinations of reference and test stimuli. He found that some subjects showed a dependency on the choice of the reference bandwidth, whereas on average across subjects no dependency was found. However, to reduce the difference in the physical properties between reference and test stimulus — and therefore to reduce potential biases — a reference stimulus should be employed that is most easily comparable to all the test stimuli under investigation. In the context of quantifying spectral loudness summation this should be the case for a medium reference bandwidth. Whether loudness summation depends on the selection of the reference bandwidth is tested in experiment 1 and 2.

Another potential bias concerns the presentation order of the stimuli. In case of loudness matching experiments, several authors reported an advantage of interleaved track presentation to reduce several bias effects concerning the choice of reference, the test stimulus starting level and the subjects response behavior [Buus *et al.*, 1999; Florentine *et al.*, 1996; Hübner and Ellermeier, 1993; Jesteadt, 1980; Levitt, 1968, 1971; Verhey, 1998]. For example Jesteadt [1980] stated, in accordance with Levitt [1971], that interleaved track presentation eliminates the certainty that the subsequent stimulus will be similar to the current one or that any given response will have a direct influence on the next stimulus. Hence, the principle sources of a sequential response bias[1] are eliminated. However, in most cases the authors preference for interleaved track presentation was based on psychological assumptions. The only known study which directly compared interleaved versus subsequent track presentation in loudness comparisons is Verhey [1998]. He found his results less influenced by the selected stimulus starting level when interleaved track presentation was used. However, the size of the effect was small and was only found for signals of long duration whereas no effect was found for signals of short duration. Whether loudness summation depends on the stimuli presentation order or not is investigated experiment 3.

In contrast to loudness matching procedures the procedure of *loudness scaling* [Allen *et al.*, 1990; Heller, 1985; Hohmann and Kollmeier, 1995b; Pascoe, 1978] is a direct measure of the *loudness function*, i.e., the loudness of a stimulus is measured as a function of the physical level. The stimulus is presented at a wide range of levels and the loudness of each presentation is rated on a verbal scale. In several studies the method of loudness scaling has been used for the fitting of the compression characteristic of hearing aids [Allen *et al.*, 1990; Appell *et al.*, 1995; Hohmann, 1993; Kießling, 1996; Kollmeier, 1997b; Moore *et al.*, 1992]. Some studies also tested the success in restoring normal loudness with that method by comparing the hearing–impaired listeners' aided loudness function with a normal reference. However, none of the studies focused on the effect of spectral loudness summation by comparing the amount of loudness summation of the aided impaired ear with normal reference data. In experiment 4 two methods of loudness scaling are investigated which differ in the way how the range of presentation levels is determined, i.e., whether the range of levels is estimated by a pre–measurement of the subjects dynamic range (*Oldenburg Loudness Scaling Procedure*) or whether it is estimated adaptively during the measurement (*Oldenburg–ACALOS*, Oldenburg

[1] Throughout this paper, the term *sequential response bias* is used for any bias were the previous stimulus presentation and/or the subjects response to the previous stimulus influences the subjects response to the current stimulus. In the literature [e.g., Poulton, 1989], the term *sequential contraction bias* is also used to describe a sequential response bias.

– Adaptive CAtegorical LOudness Scaling). The latter method has been suggested by Brand and Hohmann [2001a] for simplifying the subjects task and to reduce measurement time and therefore is of special interest in measurements with hearing–impaired subjects.

3.2 Methods

3.2.1 Stimuli

All stimuli were generated digitally by using a sample of white noise with Gaussian amplitude statistics, transforming the sample into the frequency domain, setting the magnitude of the Fourier coefficients to the desired spectrum and transforming back the spectrum into the time domain. Two sets of stimuli were used: bandlimited uniform exciting noise [Zwicker and Fastl, 1990] and bandlimited low–noise noise stimuli [Hartmann and Pumplin, 1988; Pumplin, 1985].

The characteristic of uniform exciting noise (*UEN–stimuli*) is that it produces the same intensity in each critical band. Therefore it should produce the greatest effect of spectral loudness summation when the stimulus bandwidth is increased. The UEN–stimuli were geometrically centered at 10.5 Bark. The bandwidths of the signals were 1, 3, 5, 9, and 17 Bark.

Bandlimited low–noise noise (*LNN–stimuli*) was used as a second set of stimuli. This type of stimulus has the advantage of lower temporal envelope fluctuations (lower *crest factor*, i.e., lower ratio between maximum peak level and signal RMS–level) as compared to Gaussion noise. In the present study, a simplified algorithm proposed by Kohlrausch *et al.* [1997] is used to generate the LNN–stimuli: After bandlimiting the Gaussion noise by setting the magnitude of the Fourier coefficients outside the desired frequency region to zero, the Hilbert envelope was calculated and the time waveform was divided by this envelope on a sample–by–sample basis and then again restricted to its original bandwidth. In contrast to the method proposed by Kohlrausch *et al.* [1997], these steps were applied only once, whereas Kohlrausch used several iterations. The number of iterations was reduced to prevent the resulting signals from having a tonal characteristic, while still reducing the crest factor significantly (e.g., from about 4.1 to about 2.3 for the bandlimited Gaussion noises). The LNN–stimuli were arithmetically centered at 2000 Hz. The bandwidths of the signals were 200, 400, 800, 1600, 3200, and 6400 Hz.

Table 3.1 summarizes the stimulus parameters. It can be seen that the signal center frequencies and the range of bandwidths are approximately the same for both noise types.

3.2.2 Loudness Matching Procedures

Stimuli with different bandwidths were matched in loudness to a reference signal with a fixed bandwidth using an adaptive 2–interval, 2–alternative forced choice procedure. In each trial the listener heard two sounds, the reference and the test signal, which were separated by a 500 ms silent interval. The listener indicated which signal was louder by pressing the corresponding key on a keyboard or on a hand–held computer. The reference level was fixed and the level of the test signal was varied according to a simple 1–up 1–down procedure,

	UEN (uniform exciting noise)					LNN (low–noise noise)					
bandwidth [*Bark*] bandwidth [*Hz*]	1 210	3 640	5 1080	9 2070	17 5100	0.67 200	1.33 400	2.67 800	5.19 1600	9.45 3200	14.94 6400
center frequency center frequency	10.5 Bark 1370 Hz					13.0 Bark 2000 Hz					
signal duration	2000 msec					1000 msec					
Hanning window ramps	50 msec					10 msec					
sampling frequency	44100 Hz					32000 Hz					
free–field equalization	yes					no					
noise statistic	gaussian					gaussian, one low–noise noise iteration					
spectral shape	bandlimited uniform exciting noise					bandlimited white noise					

Table 3.1: *Signal parameters of the noise stimuli. Left column shows the parameters for the UEN–stimuli, right column for the LNN–stimuli.*

which converges at the 50 % point of the psychometric function [Levitt, 1971]. If the listener indicated that the test signal was the louder one, its level was reduced in the next trial, otherwise it was increased. Each stimulus configuration was tested for at least two different initial level differences between test and reference signal (starting level difference). Each of this *tracks* consisted of an adaptation and a data acquisition phase. During the adaptation phase the step size of the level change was reduced when a reversal occurred, i.e., when a ‘softer’ decision followed a ‘louder’ decision or vice versa. During the following data acquisition phase the step size was maintained and the track finished after 4 reversals. The equal loudness–level for each track was determined by calculating the median level across these 4 reversals.

Throughout this study two methodological paradigms reflecting the trial presentation order within tracks were used. In the *subsequent paradigm*, the tracks for the different stimulus pairs were performed subsequently. Hence, the track for one stimulus pair has to be finished before the track for the next stimulus pair starts. In the *interleaved paradigm*, the tracks run concurrently in a series of trials to simultaneously obtain loudness matches for all stimulus pairs under test. Both paradigms were tested in two different configurations concerning the apparatus, the stimuli and some other methodological parameters. The resulting four methodological paradigms are summarized in Table 3.2.

3.2.3 Loudness Scaling Procedures

Two different loudness scaling methods, one constant stimulus and one adaptive method, were used in this study.

The non–adaptive *Oldenburg Loudness Scaling Procedure* [Brand, 2000; Hohmann and Kollmeier, 1995b] consists of two phases. In the first phase, the auditory dynamic range of the subject is estimated by presenting an ascending level sequence. The subject’s task is to press the response button as soon as the stimulus is audible. After that, the listener

	LM–METHOD–1	*LM–METHOD–2*	*LM–METHOD–3*	*LM–METHOD–4*
track presentation	subsequent	interleaved	subsequent	interleaved
stimuli	UEN	UEN	LNN	LNN
level step size in adaptation phase	10, 5, 2, 1 dB	10, 5, 2, 1 dB	8, 4, 2 dB	8, 4, 2 dB
level step size in acquisition phase	1 dB	1 dB	2 dB	2 dB
starting level difference	random in range ± 15 dB	random in range ± 15 dB	-10, 0, 10 dB	-10, 0, 10 dB
trial presentation order	reference first	reference first	random	random
presentation	monaurally	monaurally	diotic	monaurally
number of tracks per subject	4	4	3	6
subject groups	A,B,D	D	C	A,B,C

Table 3.2: *Variations in the measurement parameter of the loudness matching methods investigated.*

is asked to press the response button immediately when the stimulus is perceived as being too loud. In case that the listener does not press the response button, the sequence stops at a maximum level of 120 dB SPL. In the second phase, the loudness function is assessed by presenting stimuli covering the predetermined individual auditory dynamic range with a uniform distribution of presentation levels. In this phase, the stimuli are presented twice at each of 7 different levels. In order to avoid context effects which are due to the tendency of many listeners to rate the current stimulus relatively to the previous stimulus, the stimuli are presented in pseudo–random order in a way that the maximum difference of subsequent presentation levels is smaller than half of the dynamic range of the sequence. The listener's task is to rate the loudness of the stimuli on the verbal scale shown in Figure 3.1 (i.e., 5 main verbal categories, 4 intermediate categories and 2 limiting categories).

The adaptive *Oldenburg — Adaptive CAtegorical LOudness Scaling* (*Oldenburg–ACALOS*) was proposed by Brand [Brand, 2000; Brand *et al.*, 1997a,b,c]. In contrast to the *Oldenburg Loudness Scaling Procedure*, the subjects individual dynamic range is estimated from the subjects response to the stimuli. In this way, the procedure adaptively spans the dynamic range of the subject within a small number of about 5–6 trials. The pre–measurement is omitted. Throughout the whole procedure, the subjects only task is to rate the loudness on the categorical scale shown in Figure 3.1. As in the *Oldenburg Loudness Scaling Procedure* the stimuli levels are presented in randomized order. Brand and Hohmann [2001a] found that the adaptive version is more efficient than the non–adaptive version, i.e., the same accuracy is achieved with fewer trials because the pre–measurement is omitted. Furthermore, the training of the subjects is easier as there is no distinction between the pre–measurement and the main phase of the procedure.

Table 3.3 summarizes the measurement parameters under investigation. *LS–METHOD–1* and *LS–METHOD–3* were performed according to the *Oldenburg Loudness Scaling* procedure, whereas *LS–METHOD–2* and *LS–METHOD–4* were carried out according to the *Oldenburg–*

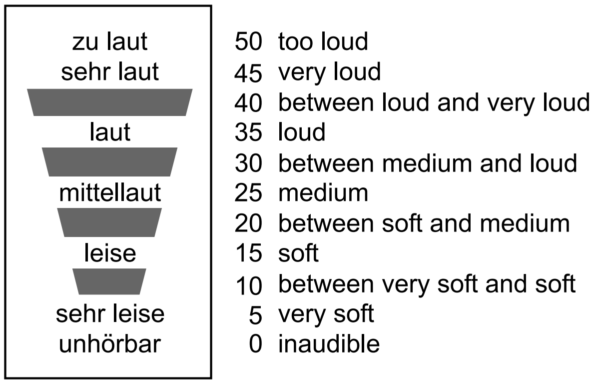

Figure 3.1: *Category scale with 11 response alternatives used by the subjects to rate the loudness. The alternatives presented to the subjects on the screen of the handheld computer are visualized within the box. The numbers next to the box show the corresponding categorical units (CU) which are used for data storage and analysis. An english translation of the german categories displayed on the screen is given in the right column. The categorical units and the english translation were not visible to the listener.*

ACALOS procedure. Both procedures were investigated for the UEN–stimuli (*LS–METHOD–1*, *LS–METHOD–2*) as well as for the LNN–stimuli (*LS–METHOD–3*, *LS–METHOD–4*).

	LS–METHOD–1	*LS–METHOD–2*	*LS–METHOD–3*	*LS–METHOD–4*
adaptive	no	yes	no	yes
stimuli	UEN	UEN	LNN	LNN
subject groups	A,B	A,B	A,B	A,B

Table 3.3: *Variations in the measurement parameter of the loudness scaling methods investigated.*

After completion of the loudness scaling measurements, the loudness function for each stimulus is approximated by a model function consisting of two straight lines smoothed by a Bezier curve around their kneepoint [Brand *et al.*, 1998, for details see appendix B.2]. The amount of loudness summation between two stimuli is then calculated at a specific loudness value from the stimuli's loudness functions by taking the level difference between the two levels that correspond to the same specific loudness category.

The advantage of the loudness scaling procedure over the loudness matching procedure is, that it produces an absolute measure of loudness, i.e., the loudness functions of the stimuli can be easily compared at any level or categorical loudness. Therefore, loudness scaling of two stimuli provides more information than a loudness match (relative measure) between these two signals and requires about the same measurement time.

3.2.4 Apparatus

All stimuli were generated on a silicon graphics workstation (INDY) using the software package *SI* which was developed at the University of Göttingen.

The loudness matching experiments for the methodological paradigms *LM–METHOD–3* and *LM–METHOD–4* were investigated using a procedure that was completely controlled by the *SI* software. The stimuli were D/A converted (16 bits), and then amplified with a computer controlled audiometric amplifier. The stimuli were presented via headphones (Sennheiser HD 25). The subject's response was collected using a personal computer showing the response alternatives. The personal computer was connected to the workstation via TCP/IP network. The subjects were seated in a sound–insulated booth.

For the other two matching procedures (*LM–METHOD–1* and *LM–METHOD–2*) and for all loudness scaling experiments (*LS–METHOD–1* to *LS–METHOD–4*) a computer–controlled audiometry workstation was used which was developed within a German joint research project on speech audiometry [Kollmeier, 1996]. A personal computer with a coprocessor board (Ariel DSP 32C) with 16–bit stereo A/D–D/A converters controlled the complete experiment as well as stimulus presentation and recording of the subject's responses. The stimulus levels were adjusted by a computer–controlled custom–designed audiometer comprising attenuators, anti–aliasing filters and headphone amplifiers. Signals were presented monaurally to the subjects with Sennheiser HD 25 headphones. The subjects response was collected using an Epson EHT 10S handheld computer with a LCD touchscreen showing the response alternatives. The handheld computer was connected to the personal computer via serial interface. The listeners were seated in a sound–insulated booth.

3.2.5 Subjects

Four groups of subjects participated in the experiments. All subjects had normal hearing (absolute threshold in quiet $\leq$ 15 dB HL at 250, 500, 1000, 2000, 4000 and 6000 Hz) and no previous history of any hearing problems. They ranged in age from 26 to 33. All subjects had extensive experience in psychoacoustic experiments. The number of subjects in group A was 3, in group B 3, in group C 4 and in group D 6. Groups A and B participated in the same experiments, i.e., experiment 1 (section 3.3.1) and 2 (section 3.3.3) as well as in the loudness scaling experiments (section 3.3.4). The 6 subjects from group A and B were separated after the measurements in these two subgroups for subsequent data analysis, because it was found that 3 subjects (group A) showed on average across all experiments significantly lower amount of loudness summation than the other 3 subjects (group B). Groups C and D participated in experiment 3 (section 3.3.2). Some of the subjects in group A and B also participated in the measurements with group C and D (for details see appendix B.1).

3.3 Results

3.3.1 Experiment 1: Subsequent Track Presentation

In the first experiment the influence of the reference bandwidth is investigated using the non–interleaved paradigm *LM–METHOD–1* and a fixed reference level of 75 dB SPL for all reference bandwidths under test.

Figure 3.2 shows the results for reference bandwidths of 210 Hz (1 Bark), 1080 Hz (5 Bark) and 5100 Hz (17 Bark) in the left, mid and right panel, respectively. Each panel shows the

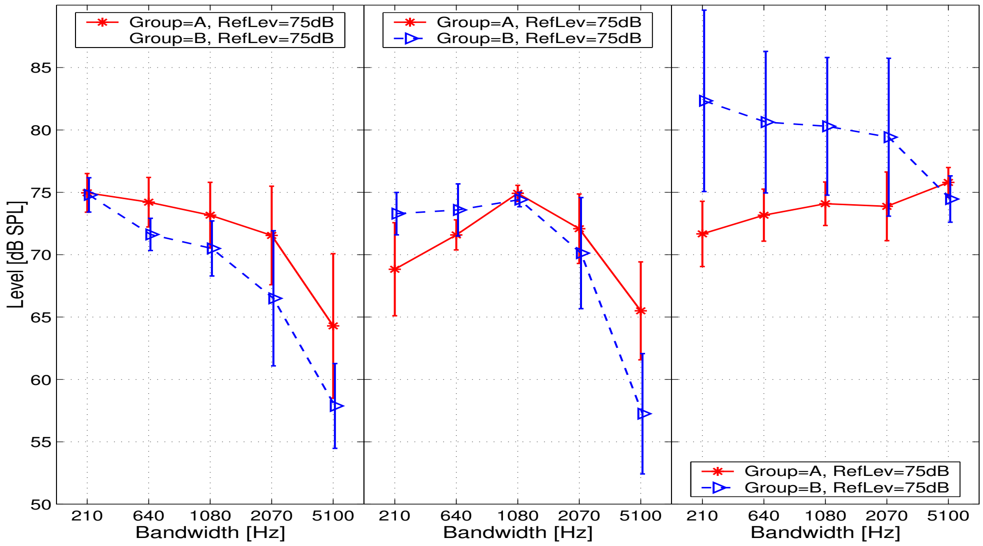

Figure 3.2: *Equal–loudness levels as a function of the test–signal bandwidth measured with non–interleaved tracks (paradigm LM–METHOD–1). Three reference bandwidths were tested: 210 Hz (1 Bark, left panel), 1080 Hz (5 Bark, mid panel) and 5100 Hz (17 Bark, right panel). The level of the reference signals was fixed at 75 dB SPL. Each panel shows the averaged data across three subjects each in group A (∗, solid lines) and B (▷, dashed lines). The vertical bars show plus minus one (inter–individual) standard deviation of the individual mean value across subjects.*

levels corresponding to equal–loudness averaged data across subjects in group A (∗, solid lines) and B (▷, dashed lines, individual data is presented in appendix B.3 Figure B.1). A level below the reference level indicates a lower test signal level than the reference at equal loudness. The error bars represent the amount of variability across subjects, calculated as plus minus one (inter–individual) standard deviation. As expected, the data shows no level difference to the 75 dB SPL reference and the smallest error when test and reference signal are the same. Increasing the bandwidth difference between test and reference signal leads to a larger effect and to a larger uncertainty in the judgements. In all conditions, subjects in group A (∗, solid lines) show a smaller amount of loudness summation than subjects in group B (▷, dashed lines). However, only the data obtained with the smallest reference bandwidth (Figure 3.2, left panel) is consistent with the literature [e.g., Port, 1963; Zwicker *et al.*, 1957], i.e., the level of the test stimulus decreases monotonously with increasing bandwidth. Increasing the reference bandwidth leads to a deviation from these data. For the 5100 Hz reference bandwidth (Figure 3.2, right panel) the effect of loudness summation is reduced for subjects in group B (▷, dashed lines). Moreover, it vanishes or is even slightly negative for subjects in group A (∗, solid lines).

It is noteworthy that some subjects complained about difficulties in comparing the narrowband signals which exhibits audible modulations to the broadband signals. However, this perceptual difference within the UEN–stimuli is assumed to affect the results independently from the choice of the reference bandwidth. It therefore can not explain the dependency on the reference bandwidth in this experiment. Instead the subjects — especially from group A

— tend to lower the test stimulus level, whenever the perceptual difference to the reference increases. This might be explained by the subjects tendency to adjust the test stimulus level towards their impression of 'normality', i.e., a level at or below 70 dB SPL corresponding to a 'comfortable loud'. This type of bias was reported by several authors [e.g., Florentine *et al.*, 1996; Port, 1963; Scharf, 1961; Stevens, 1955] and may influence the results especially when the test signal levels required to produce equal loudness become high enough to produce annoyance. Such high test signal levels did primarily occur for the largest reference bandwidth.

3.3.2 Experiment 2: Interleaved Track Presentation with Level Correction

To exclude the influence of different loudness levels of the reference signal, experiment 2 was conducted with the same subjects as in experiment 1 using the paradigm *LM–METHOD–4*. The reference level for the reference bandwidths 200 Hz, 1600 Hz and 6400 Hz were set to 78 dB SPL, 75 dB SPL and 60 dB SPL, respectively, according to an a–priori estimate of the amount of loudness summation from experiment 1 and the literature [i.e., Launer, 1995; Verhey, 1998]. In addition, low–noise noise stimuli are used to avoid audible modulations in the narrowband stimuli and the stimuli tracks are presented in an interleaved way to avoid a possible sequential response bias.

The results are presented in Figure 3.3 (individual data is presented in appendix B.3 Figure B.2). The general shape of the curves and the amount of loudness summation observable in Figure 3.3 corresponds much better with the literature than in experiment 1 (Figure 3.2). As in experiment 1, the subjects in group A ($*$, solid lines) show a smaller amount of loudness summation than the subjects in group B ($\triangleright$, dashed lines) for all reference bandwidth. Moreover, the difference between the data of group A and B is approximately the same in both experiments.

A comparison with experiment 1 further shows that there is only a small tendency in the subjects response behavior to adjust towards a lower test stimulus level when the perceptual difference to the reference increases. As in experiment 1, this type of bias seems to affect the data of group A more than the data of group B. However, the influence of the reference bandwidth is strongly reduced. Figure 3.4 shows the remaining influence of the reference levels on the data for two individual subjects. The left panel shows the data for a subject with a small amount of loudness summation (BG, group A), the right panel shows the data for a subject with a large amount of loudness summation (JV, group B). It is obvious that the reference levels selected in this experiment could be optimized in order to bring the three curves into agreement. For example the curve for the 200 Hz reference at 78 dB SPL of JV (right panel) would have been better reproduced by the data from the 1600 Hz and 6400 Hz reference, if their levels were adjusted to about 70 dB SPL and 50 dB SPL, respectively. In the same way levels of about 78 dB SPL for the 1600 Hz and 70 dB SPL for the 6400 Hz reference could have been more appropriate for BG (left panel). This shows the basic problem in setting the appropriate reference levels in this experiment: the individual data required to appropriately set the measurement parameters individually requires the data of the measurement itself. This would require an iterative process of setting the measurement parameters if the first

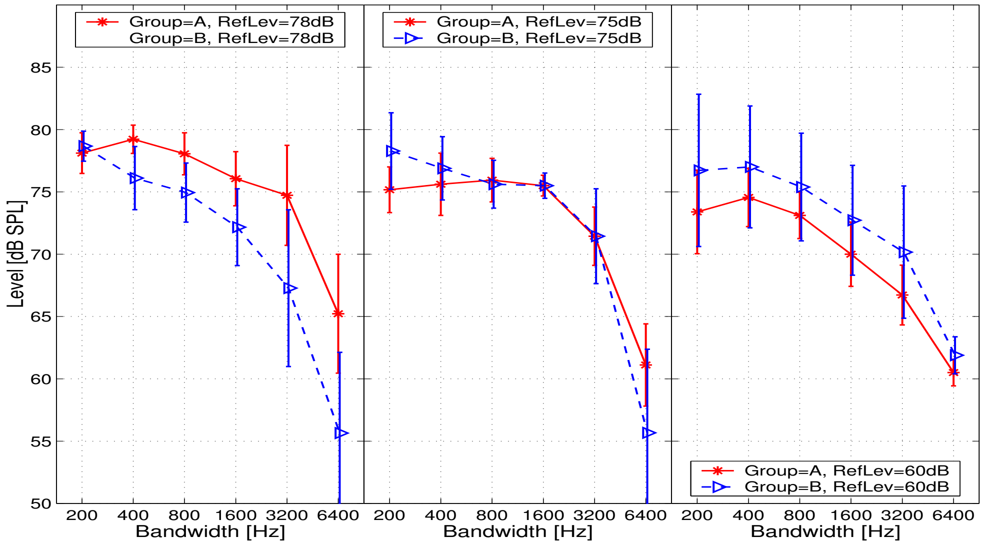

Figure 3.3: *Equal–loudness levels as a function of the test–signal bandwidth measured with interleaved tracks (paradigm LM–METHOD–4). Three reference bandwidths were tested. The presentation level of the reference signals was a–priori corrected according to the expected amount of loudness summation: 200 Hz at 78 dB SPL (left panel), 1600 Hz at 75 dB SPL (mid panel) and 6400 Hz at 60 dB SPL (right panel). As in Figure 3.2 each panel shows the averaged data across subjects in group A (∗, solid lines) and B (▷, dashed lines). The vertical bars show plus minus one standard deviation of the inter–individual mean.*

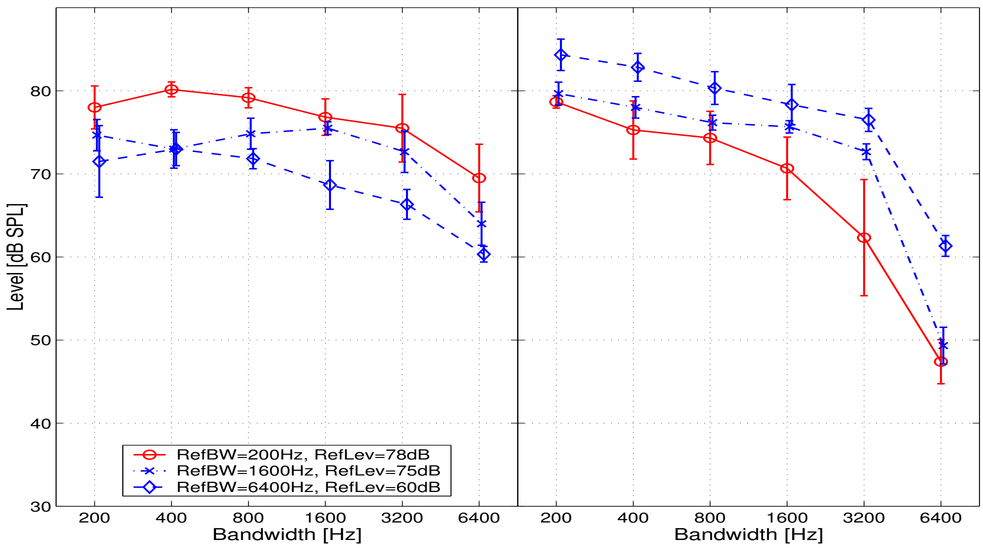

Figure 3.4: *Equal–loudness levels as a function of the test–signal bandwidth for two subjects measured with paradigm LM–METHOD–4. Left panel shows the data for subject BG (group A), right panel shows the data for subject JV (group B). Three reference bandwidths were tested: 200 Hz at 78 dB SPL (○, solid lines), 1600 Hz at 75 dB SPL (×, dash–dotted lines), and 6400 Hz at 60 dB SPL (◇, dashed lines). The vertical bars show plus minus one standard deviation of the intra–individual mean.*

a–priori estimate of the reference levels leads to inconsistent data. As a consequence, more measurement efforts would have to be spent to produce bias–free measurement results.

3.3.3 Experiment 3: Interleaved and Subsequent Track Presentation

To investigate if the different result between experiment 1 and experiment 2 were primarily due to the interleaved procedure employed in experiment 2 but not in experiment 1, another experiment is performed: the subsequent track presentation in experiment 1 (*LM–METHOD–1*) is compared with interleaved track presentation (*LM–METHOD–2*) and the interleaved track presentation (*LM–METHOD–3*) in experiment 2 is compared with subsequent track presentation (*LM–METHOD–4*). In both comparisons, all other methodological parameters are kept constant.

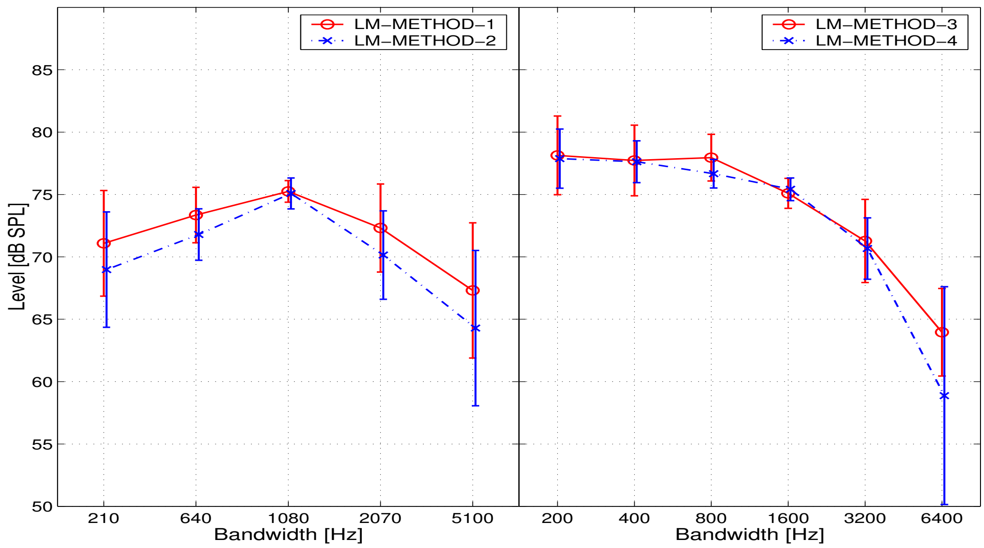

Figure 3.5: *Equal–loudness levels as a function of the test–signal bandwidth measured with subsequent (○, straight lines) and interleaved (×, dash–dotted lines) tracks. Averaged data across subjects with inter–individual standard deviations is shown. Left panel shows the results for group D and methodological paradigms LM–METHOD–1 and LM–METHOD–2, respectively. The right panel shows the results for group C and the methodological paradigms LM–METHOD–3 and LM–METHOD–4. The bandwidths of the reference signals in the left and right panel was 1600 Hz and 1080 Hz, respectively. The level of the reference signals was fixed at 75 dB SPL.*

Figure 3.5 shows the equal loudness levels as a function of the test–signal bandwidth with a reference signal presented at 75 dB SPL (individual data is presented in appendix B.3 Figure B.3). Note that a different set of subjects (6 subjects, group D) was employed in the comparison between paradigms *LM–METHOD–1* and *LM–METHOD–2* (left panel in Figure 3.5, both employing UEN as stimulus) than in experiment 1 and a different set of subjects (4 subjects, group C) was employed in the comparison between paradigms *LM–METHOD–3* and *LM–METHOD–4* (right panel in Figure 3.5, both employing LNN as stimulus) than in experiment 2. Nevertheless, the shape of the results in the left panel in Figure 3.5 is very similar to the corresponding results from experiment 1 (*LM–METHOD–1*, cf. mid panel in Figure 3.2) whereas the results in the right panel of Figure is very similar to the corresponding results from experiment 2 (*LM–METHOD–4*, cf. mid panel in Figure 3.3). The difference between the interleaved and the subsequent procedure in both panels of Figure 3.5 is negligible. This indicates that the reduced bias effect observed in experiment 2 is not due to interleaved track presentation. It can be concluded, that the presentation order plays a minor role, while the level correction introduced in experiment 2 influences the results in a significant way.

3.3.4 Experiment 4: Loudness Scaling

Figures 3.6 and 3.7 show the equal loudness levels across bandwidth calculated from the individual loudness functions for each subject (individual data is presented in appendix B.3 Figure B.4). To derive this data, the individual loudness at the desired level (here 50 and 75 dB SPL) from the loudness function of the signal with the smallest bandwidth (reference) was used to read out the levels from the respective loudness functions for the stimuli with the other bandwidths. Subsequently, the mean and standard deviation across subjects were calculated. Because this calculation produces the same level for the reference bandwidth for all subjects, the data points for 210 Hz show no inter–individual error by definition.

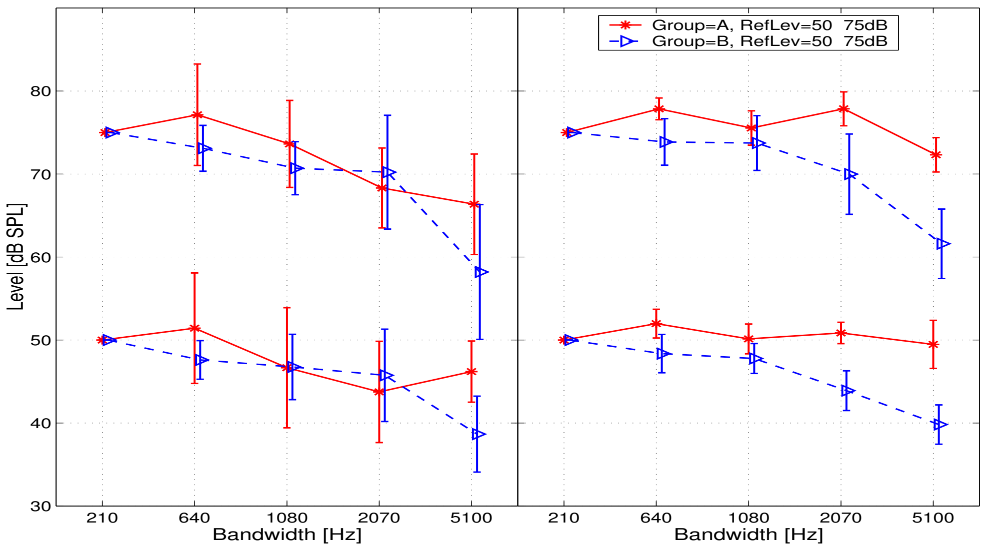

Figure 3.6: *Equal–loudness levels across bandwidth calculated from loudness scaling data for the UEN–stimuli. Each panel shows the averaged data across subjects in group A (∗, solid lines) and B (▷, dashed lines) calculated relatively to the signal with the smallest bandwidth at 50 and 75 dB SPL. The left panel shows the results for the non–adaptive procedure LS–METHOD–1, the right panel shows the results for the adaptive procedure LS–METHOD–2. The vertical bars show plus minus one standard deviation of the inter–individual mean.*

Figure 3.6 shows the data for the non–adaptive measurement paradigm *LS–METHOD–1* (left panel) and the adaptive paradigm *LS–METHOD–2* (right panel), both employing UEN–

stimuli. As in experiments 1 and 2, the subjects are grouped into group A ($*$, solid lines) and group B ($\triangleright$, dashed lines). Figure 3.7 shows the data for the non–adaptive measurement paradigm *LS–METHOD–3* (left panel) and the adaptive paradigm *LS–METHOD–4* (right panel), both employing LNN–stimuli.

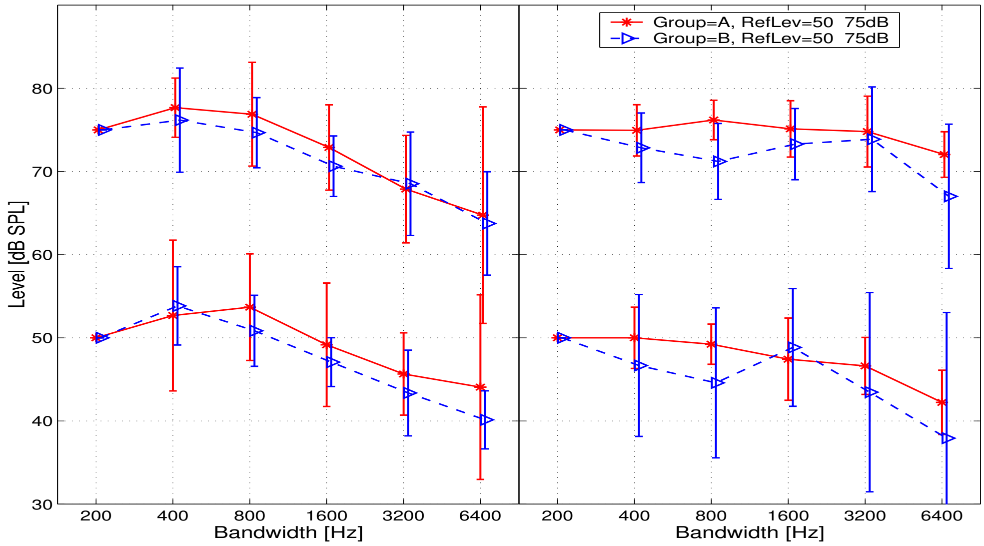

Figure 3.7: *Equal–loudness levels across bandwidth calculated from loudness scaling data for the LNN–stimuli. Each panel shows the averaged data across subjects in group A ($*$, solid lines) and B ($\triangleright$, dashed lines) calculated relatively to the signal with the smallest bandwidth at 50 and 75 dB SPL. The left panel shows the results for the non–adaptive procedure LS–METHOD–3, the right panel shows the results for the adaptive procedure LS–METHOD–4. The vertical bars show plus minus one standard deviation of the inter–individual mean.*

When the results of the non–adaptive paradigms (left panels in 3.6 and 3.7) are compared with the results of the adaptive paradigms (right panels in 3.6 and 3.7), no significant difference is found between the non–adaptive and the adaptive procedure in view of the large variability in the data. The data calculated for the 50 and the 75 dB SPL reference also show no clear difference. This finding was expected from the literature [Florentine and Zwicker, 1979; Launer, 1995; Zwicker and Fastl, 1990], i.e., no or only a small change in the amount of loudness summation with level in the investigated region of levels. In accordance with experiment 1 and 2, the results for the loudness scaling experiments show a smaller amount of loudness summation for group A than for group B.

3.4 Discussion and Conclusions

The effect of spectral loudness summation was investigated using loudness matching and loudness scaling experiments. The results of the loudness matching experiments largely depend on the measurement procedure. In experiment 1 it was found, that the subjects tend to adjust towards a lower test stimulus level whenever the perceptual difference to the reference increases. This was not found when the perceptual difference between the two stimuli was small, i.e., when the difference in bandwidth between the two stimuli was small.

The results of experiment 2 show the loudness summation effect as expected from the literature while avoiding the bias effect from experiment 1 due to several changes in the measurement setup: First, the potential problem of temporal modulations in the narrowband Gaussian noise stimuli was eliminated by using low–noise noise stimuli. Second, the sequential response bias [Levitt, 1968] possibly arising from subsequent track presentation was avoided by using interleaved track presentation and third, the influence of different loudness levels of the reference signal was reduced by adjusting the reference level to an a–priori estimate of the amount of loudness summation. The question is, which of these changes in the measurement setup led to a reduction in the bias found in experiment 1. The choice of test stimulus (Gaussian uniform–exciting noise vs. low–noise noise) can be excluded, because temporal modulations in the UEN–stimuli would influence the results independently from the choice of the reference bandwidth, which was not found in experiment 1.

Experiment 3 shows that the effect of presenting interleaved tracks as compared to a subsequent track presentation is very small (or even vanishes). It is definitely smaller than the inter–individual differences found in experiments 1 and 2. Therefore, the reduced bias found in experiment 2 is not due to the presentation order of the stimuli, which would have indicated a sequential response bias. Although this holds for the experiments introduced here, it might not generally hold for experiments with slightly different parameters. Verhey [1998], for example, tested two procedures similar to paradigms *LM–METHOD–3* and *LM–METHOD–4*. He found that interleaved tracks should be used to measure loudness summation particularly when temporal aspects of loudness summation are investigated and that interleaved tracks reduce the influence of the test signals starting level. Therefore it can be stated that as long as experiments may be influenced by sequential biases it is advisable to use interleaved track presentation to avoid them. Alternatively, one could use a presentation order that quantifies this kind of bias [Levitt, 1968].

The bias found in experiment 1 were reduced when the levels of the reference stimuli are set to the same expected loudness and thereby lower presentation levels were achieved (experiment 2). Because in experiment 1 the level of all reference stimuli were relatively high (75 dB SPL), and even higher levels for the test stimuli are expected from the effect of loudness summation when a narrowband test stimulus is compared to the reference with the broadest bandwidth, the tendency to adapt towards lower test stimulus levels corresponds to a migration towards a comfortable level. This finding is in accordance with Florentine *et al.* [1996] who found the subjects biasing their judgements such that the variable stimulus migrates towards a comfortable level. In general, the tendency to judge a stimulus too close to the internal standard (*stimulus contraction bias*, according to Poulton, 1989) is also known from other experiments [e.g., Port, 1963; Scharf, 1961; Stevens, 1955]. However, it is also conceivable that other perceptual categories like annoyance may have influenced the results in the observed way [Appell and Hohmann, 1998; Zwicker, 1966]. For experiment 1 it can be concluded that the subjects response behavior changes with an increasing difference in bandwidth between reference and test stimulus. If the difference is small, they are able to compare the two stimuli in loudness but when the perceptual difference is increased, the subjects tend to adjust lower test stimulus levels.

It was found in experiment 2 that the remaining dependency on the reference bandwidth may be further reduced, if the reference level is adjusted individually. However, this would require additional data to set up the measurement parameters that can only be derived from the measurement itself. This would result in an iterative process in which the reference level is set on the basis of the results of a pre–measurement. Unfortunately, such a procedure is in contrast to the aim of this study, i.e., to find a time efficient method for measuring loudness summation. It is also noteworthy that an individual adjustment of the reference level would reduce the intra–individual difference obtained for the three reference bandwidths, but that the inter–individual difference among subjects would not be affected by an individual level correction.

The fact that the data showed large inter–individual differences was emphasized by dividing the subjects into two groups. One of them (Group B) shows a larger amount of loudness summation and their results were less affected by the measurement setup than the other (Group A). Figure 3.8 shows the rank order of the amount of loudness summation across

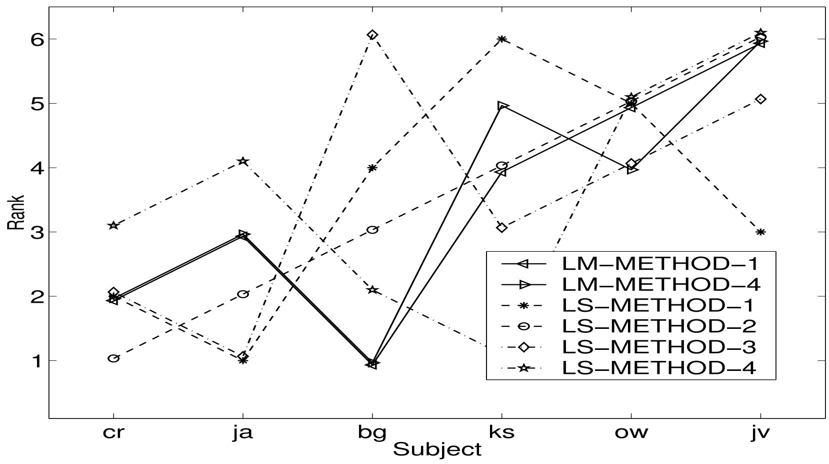

Figure 3.8: *Amount of loudness summation between the stimulus with the smallest and the largest bandwidth is plotted in terms of the rank order for each measurement method over subjects. Subjects cr, ja and bg belong to group A and subjects ks, ow and jv belong to group B. A lower rank indicates a smaller amount of loudness summation.*

subjects for each measurement paradigm that was carried out with subjects in group A and B (the corresponding data of experiment 3 carried out with groups D and C can be found in appendix B.4). The amount of loudness summation was calculated by the difference in equal–loudness level between the signal with the smallest and the signal with the largest bandwidth under test. The ranks of the subjects belonging to group A are plotted in the left half of the Figure, the other subjects belong to group B. Note, that because of the large differences found in the amount of loudness summation across methods, Figure 3.8 shows the rank order of the amount of loudness summation instead of its absolute amount in dB.

From Figure 3.8 it can be concluded that the loudness matching paradigms *LM–METHOD–1* and *LM–METHOD–4* (solid lines) produce almost the same rank order (this also holds for the data of experiment 3, refer to Figures B.5 and B.6 in appendix B.4). This is interesting, because the absolute results for paradigms *LM–METHOD–1* (Figure 3.2) and *LM–METHOD–4* (Figure 3.3) differ significantly. It can be concluded, that inter–individual differences are consistent among the loudness matching paradigms even if the absolute results depend markedly on the method used. The rank order for the loudness scaling experiments (broken lines in Figure 3.8) do not show the same rank order for all paradigms. Only a tendency to a greater amount of loudness summation (higher rank in Figure 3.8) is found when comparing the ranks of the subjects on the left half of Figure 3.8 (Group A) with the subjects on the right half (Group B). However, this can be expected from the larger variance found in the loudness scaling experiments. Hence, the data for groups A and B is not as clearly distinguishable from each other as it is the case for the loudness matching data.

This variability in the results is further investigated by Table 3.4 where the mean standard deviations found in the loudness matching experiments are compared with those from the loudness scaling experiments. As can be seen, the mean standard deviation is slightly larger in the loudness scaling experiments in both, the inter–individual and the intra–individual case. It is noteworthy that the standard deviations given in Table 3.4 are smaller than the measured effects, i.e., the amount of spectral loudness summation as well as the combination of loudness summation and bias effect found in experiment 1.

Method	InterStd	IntraStd	Quotient
Matching	3.04	1.93	1.58
Scaling	3.49	2.48	1.40

Table 3.4: *Relation between inter–individual and intra–individual variability for the method of loudness matching and loudness scaling calculated from the data of the subjects in group A and B. Column 'InterStd' shows the mean of the inter–individual standard deviations in dB at each combination of reference- and test–signal. Column 'IntraStd' shows the mean of the intra–individual standard deviations in dB for each combination of subject, reference- and test–signal. Column 'Quotient' shows the quotient between 'InterStd' and 'IntraStd'.*

Table 3.4 clearly shows a smaller intra–individual than inter–individual standard deviation for both measurement methods. This indicates that both methods strongly depend on the individual perception of loudness even when normal–hearing subjects are tested. It can be suggested from the consistent rank order of the amount of loudness summation across methods (Figure 3.8) that the perception of loudness and the effect of loudness summation in principle differs across subjects somehow independent from the measurement method. The data known from the literature obviously can only be achieved if loudness summation is calculated as the mean across normal–hearing subjects. Even Though a large interindividual variability exists, the results of this study clearly demonstrates the effect of loudness summation is present without doubt. Whereas in experiment 1 loudness summation can not be clearly separated from the observed bias effect, a loudness summation effect of approximately 10 to 15 dB for the LNN-stimuli and an effect of 5 to 10 dB for the UEN-stimuli is observed across the other experiments.

In general it is concluded from this work that the method of loudness matching may be largely influenced by bias effects when the perceptual difference between the stimuli under comparison is increased (experiment 1). This may invalidate this method for the usage in certain applications, especially in audiology. It is further shown that loudness scaling allows for a quantification of spectral loudness summation although the intra– and inter–individual standard deviations are slightly higher than for loudness matching. Larger inter–individual than intra–individual differences were found for the apparent amount of loudness summation in both methods. This clearly limits the usage of normative reference data on loudness summation in audiological applications where an assessment of the individual hearing–impaired listener's loudness summation effect is required as a basis for loudness compensation. Hence, further research is necessary on the factors influencing the individual's loudness perception and loudness summation, the effect of hearing impairment on these factors as well as its possible compensation by hearing aids.

Chapter 4

Loudness Models based on the Model proposed by Zwicker

Abstract

An overview of the general structure of Zwicker's loudness model [Zwicker, 1958; Zwicker and Fastl, 1990] is given and modifications are described with respect to several implementation details by different authors [i.e., Blum, 1999; Launer, 1995; Marzinzik, 1996; Marzinzik *et al.*, 1996b; Moore and Glasberg, 1996]. While all models use the same overall structure (i.e., pre–filtering to account for outer and middle ear transmission, separation into critical bands, construction of excitation patterns, compression of the excitation into specific loudness, summation of the specific loudness across frequency), and predict the same loudness function for normal listeners at high input levels, they differ with respect to the pre–filtering, the definition of the frequency scale (BARK versus ERB scale), the reference level for the excitation pattern, and the way in which specific loudness is computed at low input levels. Other differences refer to how the model is modified in order to predict loudness perception in hearing-impaired listeners and how the model output can be transformed to predict loudness on a different scale (e.g., categorical loudness scale). To facilitate comparisons across the model versions, a common implementation of the models has been implemented (DOS–Program). The details of the models are reviewed and a comparison of the model predictions for a limited set of input sounds is given.

4.1 Introduction

The perception of loudness and its relation to physical parameters such as intensity, spectral and temporal properties has been a major issue in psychophysical research over many decades. While the famous Weber–Fechner law relates the just noticeable difference in sound intensity to the absolute magnitude of sound intensity and hence infers a logarithmic loudness perception as a function of intensity, further work by Fletcher [e.g., Fletcher, 1995], Stevens [Stevens, 1957] and Zwicker established the power-law relation between sound intensity and perceived loudness in *sone*[1]. Moreover, Zwicker [Paulus and Zwicker, 1972; Zwicker, 1958, 1960; Zwicker and Fastl, 1990; Zwicker and Scharf, 1965] derived a sophisticated model that predicts loudness perception (in sone) not only as a function of intensity, but also depending on the spectral shape of a stationary sound using ideas both from physiological acoustics (such as filtering in the outer and middle ear, cochlea–like frequency scale, excitation patterns) and psychoacoustics (i.e., psychoacoustic frequency scale, masking patterns, power-law compression, additivity of loudness in different frequency bands). Later on, this model was extended to cover temporally varying sounds [Chalupper, 2000; Zwicker and Fastl, 1999] and to incorporate more recent findings in psychoacoustics [Moore and Glasberg, 1996; Moore *et al.*, 1997]. In order to extend the model for loudness prediction in hearing-impaired listeners, two major strategies have been proposed. The 'one–component approach' assumes that alterations in the perception of loudness (e.g., a raised absolute threshold, a reduced dynamic range) can be modeled by a single parameter describing the individual hearing loss [Florentine and Zwicker, 1979; Moore, 1995]. In contrast Launer [1995], based on data presented by Hohmann [1993], Kießling *et al.* [1994], Kießling [1995], Launer [1995] and Launer *et al.* [1996], argued that it is not appropriate to predict the reduced dynamic range from elevated hearing threshold alone. Instead he proposed a 'two–component model' that accounts for the raised threshold and the reduced dynamic range independently. Other authors followed Launer's argument and refined his model [Marzinzik *et al.*, 1996b] or developed a model that relates raised hearing threshold and reduced dynamic range to a loss of inner and outer hair cells, respectively [Moore and Glasberg, 1997; Moore *et al.*, 1996, 1999b].

Since all of these authors used the same general structure of the model and the same processing principles, the different approaches introduced in the literature should in principle be comparable. However, the approaches differ in several implementation details, such as, e.g., the reference level definition, the frequency scale and critical bandwidth employed, as well as the exact definition of the parameters to adjust the model. This paper therefore gives an overview of these different modifications and compares the models on the basis of a generic implementation (DOS Program). The aim of this generic implementation is to serve as a platform for further modifications of the model to compensate for deficiencies of the model versions described here (see Chapters 5 and 6).

[1] The sone–scale is derived from loudness comparison and production techniques. An appropriate requirement therefore is that a doubling in perceived loudness corresponds to a doubling of the sone value, i.e., the sone–scale describes loudness in relation to a reference (in general 1 sone should correspond to a sinusoidal input signal of 40 dB SPL at 1 kHz).

4.2 Steven's Power Law

The loudness models discussed in this chapter are based on *Steven's power law* [Stevens, 1957] relating subjective loudness N to the stimulus intensity I, by

$$N = C \cdot I^{\alpha}, \tag{4.1}$$

where α is a compressive exponent equal to about 0.3. This yields a doubling of the loudness with an 10 dB increase of the input level. Factor C is a scaling factor shifting the loudness function when loudness is plotted against sound intensity in a log–log plot. In general, C is adjusted such that $N = 1$ *sone* is achieved for a sinusoidal input signal at 1 kHz at a level of 40 dBSPL.

However, Equation 4.1 does not reflect the observation that loudness N becomes zero at absolute threshold and that the change in loudness with level is more rapid near absolute threshold than predicted by this simple power law. Therefore, different modifications of Steven's power law have been proposed [for a review see Buus *et al.*, 1998; Humes and Jesteadt, 1991], which differ in the way how the absolute threshold is taken into account. Often, the absolute threshold is assumed to be generated by an internal (inaudible) noise which masks stimuli at very low levels. Hearing impairment might thus be modeled as a raised level of this internal noise.

The simplest way of taking absolute threshold into account is to apply a linear correction to the power law [*uncompressed internal noise power law*, Stevens, 1966]:

$$N = C \cdot (I - I_{ThQ})^{\alpha}, \tag{4.2}$$

where I_{ThQ} is the intensity of sound at absolute threshold. Here, the intensity at absolute threshold is simply subtracted from the stimulus intensity before compression, i.e., the correction applies before the nonlinearity. This way of accounting for threshold fulfills the constraint of zero loudness at absolute threshold but generates a too steep slope near threshold.

A second way of taking absolute threshold into account is to subtract the two intensities after they have been compressed separately [*compressed internal noise power law*, Humes and Jesteadt, 1991]:

$$N = C \cdot (I^{\alpha} - I^{\alpha}_{ThQ}). \tag{4.3}$$

The advantage of Equation 4.3 over Equation 4.2 is the more appropriate slope of the loudness function near absolute threshold.

The third way, proposed by Zwislocki [1965], is to calculate the 'overall intensity' (stimulus + internal noise) first before calculating the loudness in the same way as for the compressed internal noise model, i.e., by subtracting the compressed internal noise from the compressed total intensity (*mixed internal noise power law*). Again it is assumed that the level of the internal noise is equal to I_{ThQ}:

$$N = C \cdot ((I + k_1 \cdot I_{ThQ})^{\alpha} - (k_2 \cdot I_{ThQ})^{\alpha}). \tag{4.4}$$

Variations in the coefficients k_1 and k_2 allow for a great variety in the shape of the predicted loudness functions especially near threshold. In general, these parameters will depend on the bandwidth and the center frequency of the stimulus but they also can be adapted in order to account for spectral loudness summation or alterations in loudness perception caused by hearing impairment [Hellman and Meiselman, 1990].

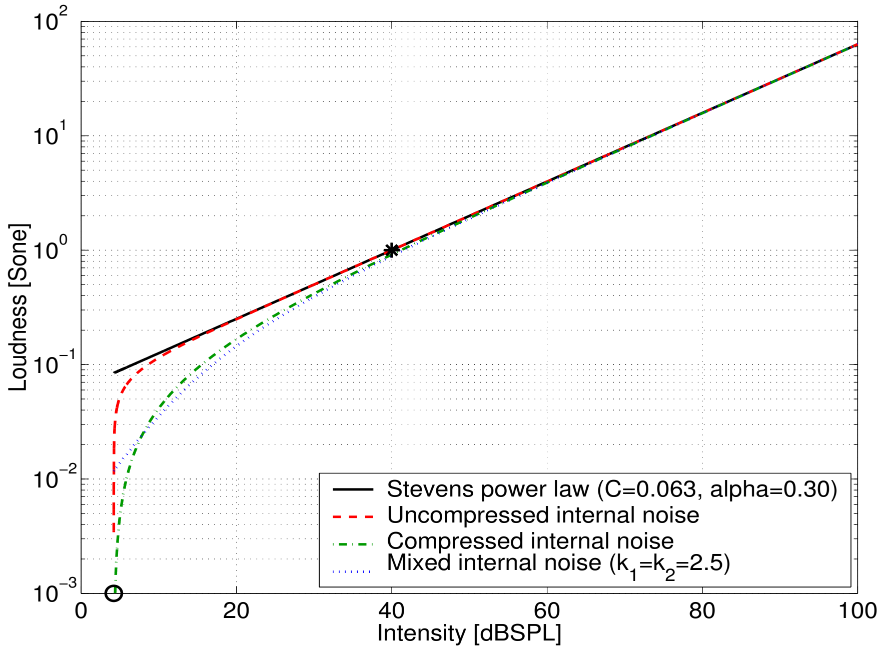

Figure 4.1: *Principle shape of the loudness functions described by the Equations 4.1 to 4.4. The symbols denote the two constraints imposed on the loudness functions, i.e., loudness becomes zero at absolute threshold (circle) and loudness equals 1 sone at an input level of 40 dBSPL.*

Figure 4.1 shows the shape of the loudness functions as described by the Equations 4.1 to 4.4. The parameters are chosen as follows: α is set to 0.3 providing a doubling of the loudness according to Equation 4.1 for a 10 dB increase of the input level, I_{ThQ} equals 4.2 dBSPL according to the *minimum audible field* [*MAF*, for a definition see ISO 226(E), 1987] at 1 kHz, C is adjusted to produce a loudness of 1 sone for a 1 kHz tone with an input level of 40 dBSPL according to Equation 4.1, and k_1 as well as k_2 in Equation 4.4 were set 2.5 assuming a pure tone input stimulus as in Hellman and Meiselman [1990].

As expected, Equations 4.1 to 4.4 show a similar shape for the loudness functions at high input levels (above about 40 dBSPL), i.e., a linear increase in loudness with level on a log–log scale with a slope according to the value of exponent α. The different ways to account for absolute threshold do only change the shape of the loudness function at low levels.

However, Stevens power law and its modifications describe the general dependence of loudness from level, but do not account for the effects of the signals spectrum on loudness. A more sophisticated model that allows for predicting loudness of arbitrary stationary sounds is discussed in the next section.

4.3 Loudness Models

Several loudness models were described in the literature that extend the formulas based on Steven's power law by combining them with a peripheral frequency analysis so that the loudness of arbitrary signals can be modeled. Most of these approaches are restricted to stationary sounds, i.e., sounds that are completely defined by their frequency spectrum. All models considered here are based on a model proposed by Zwicker [Zwicker, 1958; Zwicker and Fastl, 1990]. In addition to the basic transformations reviewed in section 4.2, a loudness model for arbitrary sounds should in principle account for the following psychophysical facts: hearing threshold, the change in loudness with level, spectral masking of frequency components and the effect of spectral loudness summation. For modeling hearing impairment the model additionally has to account for alterations in the perception of loudness in hearing–impaired people, such as the raised hearing threshold in quiet, loudness recruitment and a reduced spectral loudness summation.

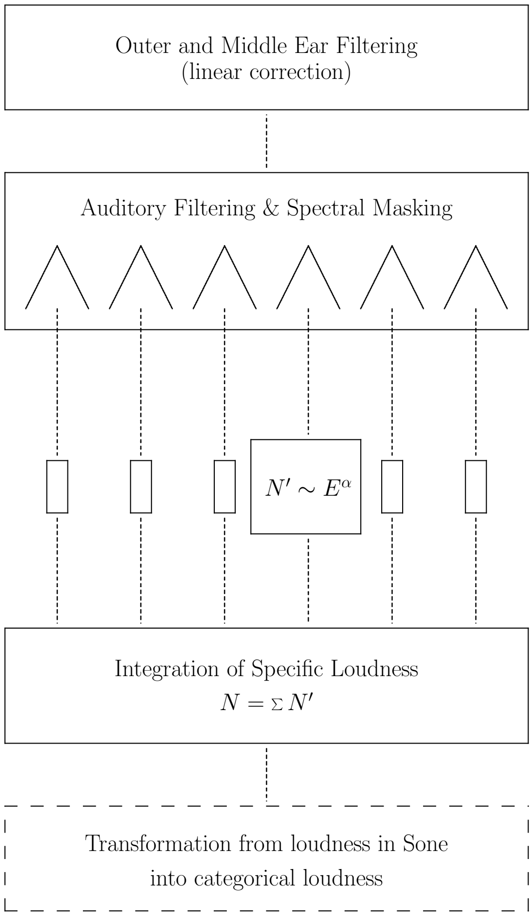

Figure 4.2: *General structure of a loudness model based on the loudness model proposed by Zwicker. It consists of a fixed filter representing the transfer through the outer and middle ear, calculation of an excitation pattern, transformation of the excitation level into specific loudness and integration of the specific loudness to the total loudness in sone and optionally a transformation into other loudness scales.*

The basic structure of the model proposed by Zwicker is illustrated in Figure 4.2. In the first stage a frequency–dependent linear correction is applied to the input spectrum. This stage accounts for the transformation of the sound through the outer and middle ear. The next stage accounts for spectral masking by estimating the masking level that each of the input components produce in the respective other frequency channels. The excitation level $E(f)$ in each frequency channel is then derived by taking the maximum of the calculated masking levels in the different frequency channels. In the next stage the *specific loudness* $N'(f)$ is calculated separately in each frequency channel assuming a power law relationship, i.e., $N'(f) \sim (E(f))^{\alpha}$. In stage 4 the *total loudness* N of the input sound in sone is calculated by integrating the specific loudness values across frequency. To quantify loudness on scales that differ from the sone–scale, a final stage may be added which transforms the sone–scale to other loudness scales such as the categorical loudness scale.

Although several authors have proposed the same principle stages for their loudness model as Zwicker, their implementations differ considerably in detail. The differences are discussed in more detail in the following sections.

4.3.1 Stage 1: Outer and Middle Ear Filtering

The first stage of the model accounts for the transmission through the outer and middle ear by applying a linear frequency–dependent correction to the input spectrum. Figure 4.3 shows the linear correction applied by the loudness models proposed by Zwicker [1958], Moore and Glasberg [1996] and Launer [1995].

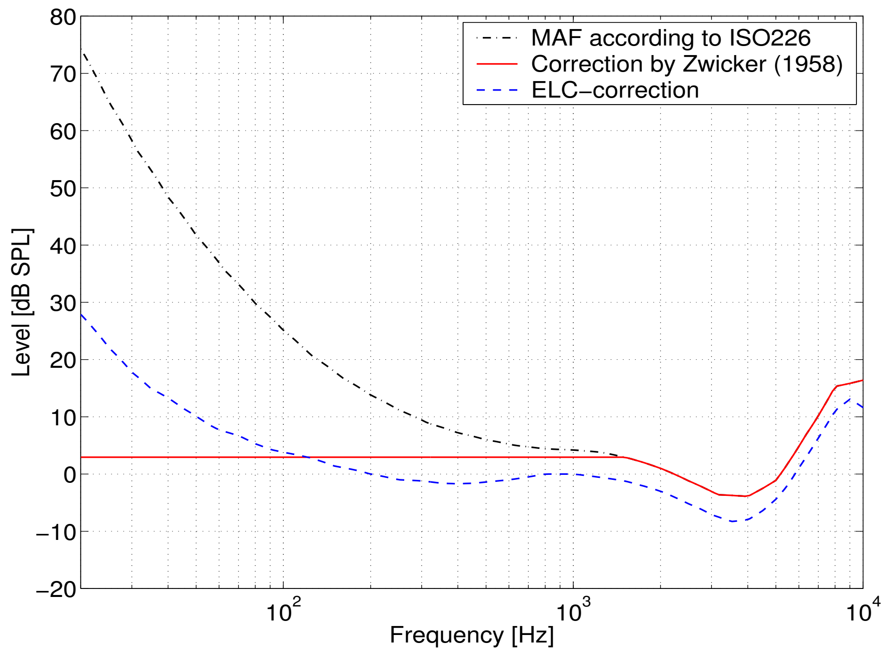

Figure 4.3: *ELC–correction proposed by Moore and Glasberg [1996], i.e., 100 phon equal–loudness level contour shifted to 0 dB at 1 kHz (dashed line). For comparison, minimum audible field (MAF) according to ISO 226(E) [1987] (dash–dotted line) and the outer and middle ear correction proposed by Zwicker [1958] are shown (solid line).*

Zwicker [1958] assumed a linear transmission through the outer and middle ear for frequencies below 1500 Hz and argued that the increase in threshold for these frequencies is produced by (inaudible) internal noise (e.g., blood flow) which therefore has to be accounted for in stage 3 of the loudness model (see 4.3.3). Above 1500 Hz he proposed that the input signal is attenuated through the outer and middle ear according to the MAF, i.e., for frequencies above 1500 Hz the MAF is subtracted from the input spectrum, whereas the input spectrum below 1500 Hz is changed by a constant value (MAF at 1500 Hz).

Moore and Glasberg [1996], following publications of Rosowski [1991] and Zwislocki [1975], found it to be unrealistic to ascribe the whole increase in absolute threshold at low frequencies to internal noise. They proposed a correction derived from the equal–loudness level contour (*ELLC*) at a loudness level of 100 phon by preserving the shape of this contour but shifting it such that it amounts to 0 dB at 1 kHz. They referred to this correction as the *ELC–correction*. This modification to the correction proposed by Zwicker applies mainly to frequencies below 1500 Hz, whereas for frequencies above 1500 Hz the ELC–correction does not differ significantly from the MAF. For the models considered in the following, we will adhere to the argumentation of Moore and Glasberg [1996] and will account for transmission through the outer and middle ear by the ELC–correction.

4.3.2 Stage 2: Auditory Filtering and Spectral Masking

Zwicker [1958] suggested that loudness is not directly related to stimulus intensity but is related to the spread of excitation evoked by the stimulus along the basilar membrane. Consequently, this *excitation pattern* is frequency dependent and in general is represented on a physiologically motivated frequency scale such as the Bark–scale [Zwicker, 1958; Zwicker and Terhardt, 1980] or the equivalent rectangular bandwidth (*ERB*) scale [Moore and Glasberg, 1996], that relates to frequency f in Hz by

$$f_{ERB} = 21.4 \cdot \lg(0.00437 \cdot f_{Hz} + 1). \tag{4.5}$$

Zwicker originally derived excitation patterns from masking patterns of pure tones masked by narrowband noise signals. Moore [1993] argued that masking patterns deduced from narrowband noise maskers may be influenced by several factors such as off–frequency listening, beat detection and combination products. He proposed a method for constructing excitation patterns from the output of *auditory filters* using broadband masking data with a spectral notch ('*notched–noise*' technique) [Patterson and Moore, 1986; Patterson *et al.*, 1987]. In this approach, the excitation pattern is calculated by integrating the intensity of each input spectrum component filtered by the respective auditory filter [for details refer to Glasberg and Moore, 1990; Moore and Glasberg, 1987].

Figure 4.4 shows the excitation patterns for a 1 kHz sinusoid at various input levels. The data is plotted relative to the ELC–correction applied in stage 1 (see Figure 4.3). As expected, the spread of excitation increases with increasing level of the input stimulus. The dash–dotted line in Figure 4.4 shows the MAF relative to the ELC–correction. It can be expected that excitation levels below the MAF are inaudible and therefore should not contribute to specific loudness, which is calculated in the next stage of the model.

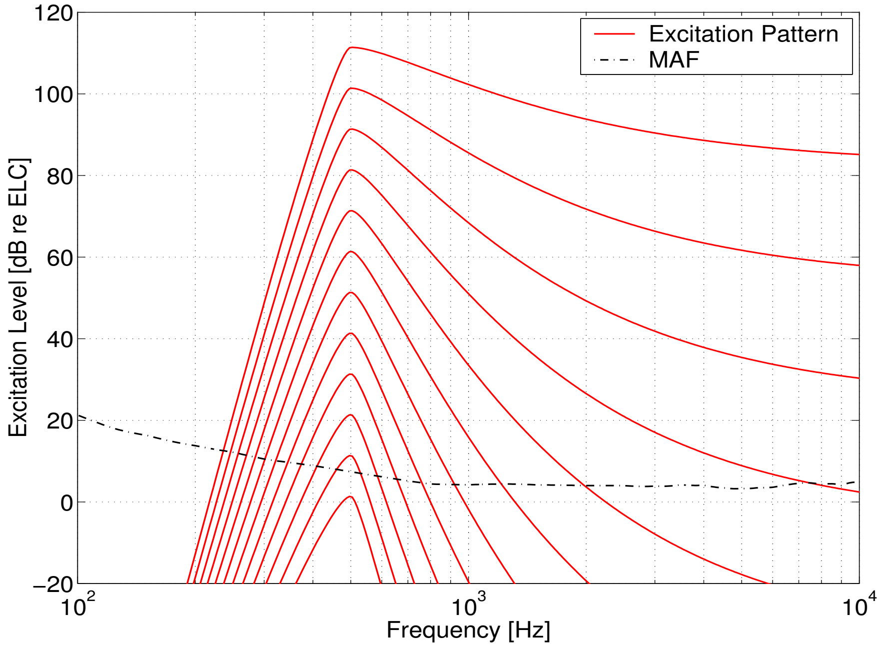

Figure 4.4: *Level of excitation relative to the ELC–correction for a 1 kHz sinusoid at input levels from 0 to 110 dBSPL in steps of 10 dB. The dash–dotted line shows the MAF (also relative to the ELC–correction).*

4.3.3 Stage 3: Calculation of Specific Loudness

In the third stage of the model, *specific loudness* $N'(f_{ERB})$, as the loudness per critical band, is calculated from the excitation pattern $E(f_{ERB})$ as a function of frequency[2]. Accordingly, we formulate the power law

$$N' = C \cdot E^{\alpha}, \tag{4.6}$$

where C and α are independent from frequency. In a log–log plot a variation in C allows for a shift of the function along the loudness axis. An exponent $\alpha < 1$ accounts for the nonlinear, compressive relationship between excitation level and specific loudness and defines the slope of the function in a log–log plot. The values for C and α result from the definition of the sone–scale.

As mentioned above, a steepening of the loudness function is observed at levels near the MAF. In the model proposed by Zwicker, as well as in the models based on it, this is accounted for by modifying Equation 4.6 in a similar way as discussed in section 4.2, i.e., the way in which the MAF is taken into account. The different models will be introduced in the next sections.

4.3.3.1 Specific Loudness according to Zwicker

Zwicker [1958] assumed that hearing threshold at high frequencies (above 1.5 kHz) is defined by the transmission function of the outer and middle ear which is represented in the model

[2]For simplicity $N'(f_{ERB})$ and $E(f_{ERB})$ are denoted by N' and E, respectively, throughout this chapter even if they are frequency dependent quantities.

by stage 1. At lower frequencies (below 1.5 kHz) the excitation produced by an inaudible internal noise E_{Thq} limits audibility by masking the excitation E produced by the stimulus. Zwicker and Fastl [1990] calculate specific loudness according to:

$$N' = C \cdot \left(E_{\substack{Thq\\[ELCC]}}\right)^{\alpha} \cdot \left(\left(\frac{1}{2} + \frac{E_{[ELCC]}}{2 \cdot E_{\substack{Thq\\[ELCC]}}}\right)^{\alpha} - 1\right) \tag{4.7}$$

Note that this formula differs from the original version in that the excitation level E corresponds to the excitation pattern as proposed by Moore [1993]. This is necessary to allow for a direct comparison of the loudness models described in this chapter in a consistent way, i.e., all models differ only in the way the specific loudness is calculated (stage 3). As a consequence, in Equation 4.7 the excitation at hearing threshold $E_{\substack{Thq\\[ELCC]}}$ and the excitation produced by the stimulus $E_{[ELCC]}$ are taken relatively to the ELC–correction (as indicated by the index '$_{[ELCC]}$')[3] instead of the correction originally proposed by Zwicker [1958].

4.3.3.2 Specific Loudness according to Moore and Glasberg

In agreement with Zwicker, Moore and Glasberg [1996] account for the threshold by assuming an inaudible internal noise masking the input stimulus. Similar to Zwicker, they simply assume additivity between the specific loudness $N'_{Stimulus}$ produced by the stimulus

$$N'_{Stimulus} = C \cdot \left(E_{[ELCC]}\right)^{\alpha} \tag{4.8}$$

and specific loudness produced by the internal noise

$$N'_{InternalNoise} = C \cdot \left(E_{\substack{Thq\\[ELCC]}}\right)^{\alpha} \tag{4.9}$$

Hence, specific loudness equals to

$$\begin{aligned} N' &= N'_{Stimulus} - N'_{InternalNoise} \\ N' &= C \cdot \left(\left(E_{[ELCC]}\right)^{\alpha} - \left(E_{\substack{Thq\\[ELCC]}}\right)^{\alpha}\right). \end{aligned} \tag{4.10}$$

In cases where the excitation $E_{[ELCC]}$ produced by the stimulus is below the excitation at threshold ($E_{\substack{Thq\\[ELCC]}}$) Moore and Glasberg set the specific loudness N' to zero assuming that specific loudness can not be negative.

For modeling hearing impairment it seems plausible to explain the raised hearing threshold by increasing the level of the (inaudible) internal noise [Florentine and Zwicker, 1979; Hellman and Meiselman, 1990; Moore, 1995]. Hence, these authors predicted the loudness in hearing impaired listeners based on the individual's audiogram alone. Actually such an approach steepens the loudness functions derived from Equations 4.7 and 4.10 at low levels, whereas loudness catches up with normal loudness at high levels [Launer, 1995]. Thus, such

[3] Throughout this paper a quantity X taken relatively to a reference Y, where X and Y are given in units of sound pressure level, will be written as $X_{[Y]}$, where $X_{[Y]} \equiv \frac{X}{Y}$ has no dimension.

an *one–component approach*[4] is capable to account for both: increased hearing threshold and recruitment. However, such an approach is based on the assumption that there is a fixed relation between raised hearing threshold and the steepening of the loudness function. Such a strong relation was found to be justified when considering mean data across hearing–impaired subjects having similar hearing thresholds [Florentine and Zwicker, 1979; Hellman and Meiselman, 1990; Moore, 1995]. On the contrary, several publications [Hohmann, 1993; Kießling, 1995; Kießling *et al.*, 1994; Launer, 1995; Launer *et al.*, 1996] showed that it is not appropriate to predict the slope of the loudness function from measurements of the hearing threshold when the individual loudness perception is under examination, as it is in case of hearing aid fitting.

4.3.3.3 Specific Loudness according to Launer

As pointed out by Launer [1995], the one–component approach can not model individual loudness data of hearing–impaired listeners in such a way that individual differences in the perception of loudness (e.g., differences in the UCL in listeners with the same audiogram) are represented in an appropriate way. In particular, the slope of the loudness function can not be adjusted independently from hearing threshold. Therefore Launer proposed a *two–component approach* allowing for an individual adjustment of the parameter α that defines the slope of the loudness function and the audiometric threshold $E_{\underset{[HL]}{Thq}}$ (Excitation relative to average normal hearing threshold):

$$N' = C \cdot \left(\frac{E_{[ELCC]}}{E_{\underset{[HL]}{Thq}}} \right)^{\beta \cdot \alpha} . \tag{4.11}$$

In Equation 4.11[5] the slope of the loudness function at high levels in a log–log plot is given by the exponent $\beta \cdot \alpha$. The factor β describes the frequency dependence of the exponent. Since a smooth variation of β can be assumed across frequencies, a polynomial fit of second order is used for β:

$$\beta = \beta(f_{ERB}) = a \cdot (f_{ERB})^2 + b \cdot f_{ERB} + c \tag{4.12}$$

f_{ERB} denotes the frequency in *ERB* and parameters a, b and c are fitted to the individual loudness scaling data for narrowband stimuli [least–squares technique, Press *et al.*, 1992].

It is noteworthy that Launer's approach has its main focus on modeling categorical loudness scaling data of hearing–impaired patients. He found in his data that the loudness functions of hearing–impaired patients show almost a constant slope when plotted against level and that threshold effects are negligible. As a consequence, he found his data well approximated by Equation 4.11 where — compared to Equations 4.7 and 4.10 — the subtractive threshold term, that is responsible for the steepening of the loudness function at low levels, is left out. One drawback of Equation 4.11 is that loudness does not become zero

[4]This terminology relates to the fact that hearing impairment is accounted for by changing only one parameter of the model and was introduced by Launer [1995]

[5]Equation 4.11 was taken from the authors C–source–code and deviates from the formula given in Launer [1995], i.e. Equation 4.7. The difference between both formulas mainly affects loudness predictions for normal–hearing listeners at low stimulus levels.

when the excitation level is below hearing threshold. Therefore Launer sets N' to zero in cases where $E_{[ELCC]}$ is below or equal to $E_{\underset{[HL]}{Thq}}$. In case of normal hearing ($E_{\underset{[HL]}{Thq}} = 1$), this corresponds to zeroing of N' when the excitation evoked by the stimulus is below or equal to the ELC–correction ($E_{[ELCC]} \leq 1$). Because of the difference between the ELC–correction and the MAF (see Figure 4.3), his approach is not able to accurately predict normal hearing thresholds in quiet. The second drawback of Launer's approach is that Equation 4.11 is not able to model the steepening of the loudness functions for the normal–hearing system at low levels.

4.3.3.4 Specific Loudness according to Marzinzik

To overcome the drawbacks connected with Equation 4.11 Marzinzik *et al.* [1996b] proposed the following modification[6]:

$$N' = C \cdot \left(\left(\frac{E_{[ELCC]}}{E_{\underset{[HL]}{Thq}}} \right)^{\beta \cdot \alpha} - \left(E_{\underset{[SPL]}{MAF}} \right)^{\beta \cdot \alpha} \right). \tag{4.13}$$

This approach is very similar to Moore's approach (Equation 4.10) but differs in the way hearing threshold is applied, i.e., the way in which the formula accounts for normal hearing threshold and the change in hearing threshold for hearing impaired. Whereas Moore [1995] included raised hearing threshold by modifying $E_{\underset{[ELCC]}{Thq}}$ providing a steepening of the loudness function at low levels, Marzinzik *et al.*, according to Launer, "attenuate" the excitation $E_{[ELCC]}$ by the amount of hearing loss and provide a steepening of the loudness function by fitting the pre–factor β. In addition, Marzinzik *et al.* subtracted the MAF in units of sound pressure level, whereas Moore and Glasberg [1996] subtract the MAF relative to the ELC–correction. Because $E_{[ELCC]}$ and $E_{\underset{[SPL]}{MAF}}$ in the equation proposed by Marzinzik *et al.* differ in units, zero loudness at threshold is not exactly matched in the prediction of loudness data for normal hearing ($E_{\underset{[HL]}{Thq}} = 1$).

Furthermore, Marzinzik *et al.* [1996b] presented only a brief evaluation of their model based on mean loudness data for a 1 kHz tone presented by Hellman and Meiselman [1990] for normal and impaired listeners. In addition, they did not use a polynomial fit for the pre–factor β, as suggested by Launer [1995]. Instead, they allowed for a free adjustment of the pre–factor across frequency allowing for a great variety in possible predictions. However, this might not be necessary in order to predict individual loudness perception. As stated by Launer [1995], he obtained good model predictions of the individual impaired loudness perception by fitting the pre–factor β by a first order polynomial fit for almost all of the subjects data and it can be assumed that this will also hold for the modified model by Marzinzik *et al.* [1996b].

[6] The exact formula was taken from the authors MATLAB–source–code because the reference level for $E_{[ELCC]}$ and $E_{\underset{[SPL]}{MAF}}$ were not clearly defined in Marzinzik *et al.* [1996b].

4.3.4 Stage 4: Transformation from Specific Loudness to Total Loudness

In the last stage of the loudness model proposed by Zwicker [1958] total loudness N in sone is calculated by integrating the specific loudness per critical band N' across frequencies:

$$N = \int_0^\infty N'(f_{ERB})\, \mathrm{d}f_{ERB}. \tag{4.14}$$

Equation 4.14 holds for all of the models described above. The only difference to the Equation proposed by Zwicker is in the frequency scale used for the integration. Zwicker used the Bark–scale, whereas the models described here are based on calculating excitation pattern dependent on the ERB–scale.

4.3.5 Stage 5: Transformation from Loudness in Sone to Loudness in Categorical Units

Most experimental setups for quantifying loudness perception, such as absolute–magnitude–estimation [e.g., Blum *et al.*, 1998; Gescheider and Hughson, 1991], absolute–magnitude–production [e.g., Hellman and Meiselman, 1990; Serpanos *et al.*, 1997], cross–modality–matching [e.g., Hellman and Meiselman, 1988] or loudness scaling [e.g., Allen *et al.*, 1990; Brand and Hohmann, 2001a; Hellbrück, 1991], do not directly observe loudness on a relational (or ratio) scale, i.e., in units of sone. In general a transformation between the observed loudness scale and the sone–scale is required when comparing experimental data with the loudness models.

Especially the method of categorical loudness scaling has become a common method for measuring loudness perception in the clinic [Kießling *et al.*, 1994; Kollmeier, 1997b] as well as for hearing aid fitting [Kießling, 1996; Pascoe, 1978]. Hohmann [1993], Launer [1995] as well as Blum [1999] have proposed a transformation between the loudness in sone and the loudness in categorical units (*CU–scale*) for the 'Oldenburg loudness scaling' procedure. The subjects task in this procedure is to rate the loudness of the presented sound on a verbal scale with a total of 11 categories covering the full auditory dynamic range including the limiting categories 'inaudible' and 'too loud'. The subject's verbal ratings are then assigned to a numerical scale ranging from 0 to 50, representing the CU–scale.

Launer [1995] proposed the following transformation between the loudness N in sone as it is predicted by his loudness model and the loudness CU in categorical units[7]:

$$CU = 17.33 \cdot \lg(3.6 \cdot N). \tag{4.15}$$

This transformation is based on the assumption that the relation between the loudness of a sound in categorical units and the sound pressure level in dB can be modeled by a straight line. Brand *et al.* [1998] showed that the relation between categorical loudness and sound pressure level is more accurately reproduced by a model function consisting of two straight

[7] The earlier proposed transformation by Hohmann [1993] had the same structure as the one proposed by Launer but differs in its constant parameters, i.e., Equation 4.15 was given by $CU = 18 \cdot \lg(2.5 \cdot N)$.

lines smoothed by a Bezier curve around their kneepoint (for details see appendix B.2). Blum [1999] proposed the transformation

$$CU = 14.5 + \max(8.2; 4.2 \cdot \lg(N) + 8.2) \cdot \lg(N). \tag{4.16}$$

which accounts for both: the changes implied by the modified model function and the loudness model as it is proposed by Marzinzik *et al.* [1996b]. However, Blum [1999] tested his transformation within the 'complete' model, i.e., β was fitted to give best model predictions for measured individual loudness scaling data. Therefore, it is not possible to clearly separate between the relative contribution of the fitting of β to the loudness scaling data and the optimized transformation. Figure 4.5 shows his transformation in comparison with the transformation earlier proposed by Launer [1995].

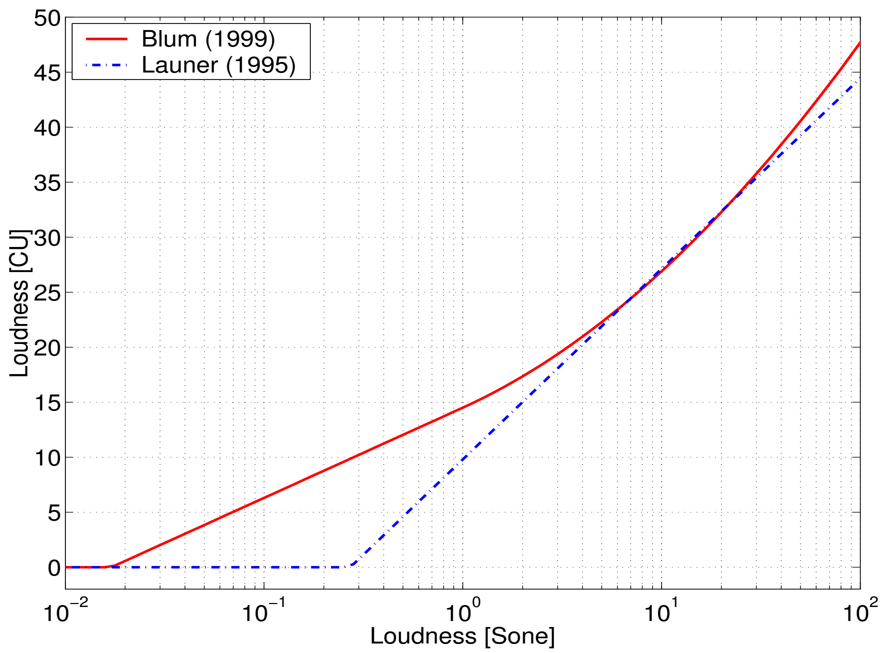

Figure 4.5: *Transformations between loudness in sone and loudness in categorical units. The dash–dotted line represents the transformation proposed by Launer [1995] (Equation 4.15) and the solid line the transformation proposed by Blum [1999] (Equation 4.16).*

4.4 Adjustment and Predictions of the Models

In order to directly compare the different model approaches with each other using the same input signals, an adjustment of the parameters in the formulas for the calculation of specific loudness, C and α, is required. Specifically, the parameters were set so that a loudness of 1 sone is achieved for a 1 kHz tone at a level of 40 dBSPL, and a doubling in loudness is predicted whenever the level of this tone is increased by 10 dB at high input levels. In accordance with these constraints, both parameters were fitted simultaneously by a least–squares technique to achieve $N = 1, 2, 4$ and 8 sone for input level intensities of $I = 40, 50, 60$

Parameter	Zwicker (Equation 4.7)	Moore (Equation 4.10)	Launer (Equation 4.11)	Marzinzik (Equation 4.13)
C	0.0960	0.0730	0.0401	0.0755
α	0.2083	0.2159	0.2522	0.2122

Table 4.1: *Summary of the parameters fitted to the models. Rows C and α show the respective parameters fitted to the models discussed above.*

and 70 dBSPL of the 1 kHz tone, respectively. Table 4.1 shows the respective parameters for the models discussed above.

Figure 4.6 shows loudness functions for sinusoids of different frequencymodeled for the normal–hearing system, i.e., without any adaptation to the individual hearing threshold. As can be seen, the predictions obtained with Zwicker's and Moore's models show almost no difference. Near threshold the predicted slope is slightly steeper in Moore's model when compared to Zwicker's model. As specified, N' approaches zero asymptotically in both models when the level of the tone approaches hearing threshold. This is not the case for the predictions obtained with Launer's and Marzinzik's model. Whereas Launer's model in general predicts too low hearing thresholds, Marzinzik's model better accounts for the MAF even though it still not accurately predicts the MAF.

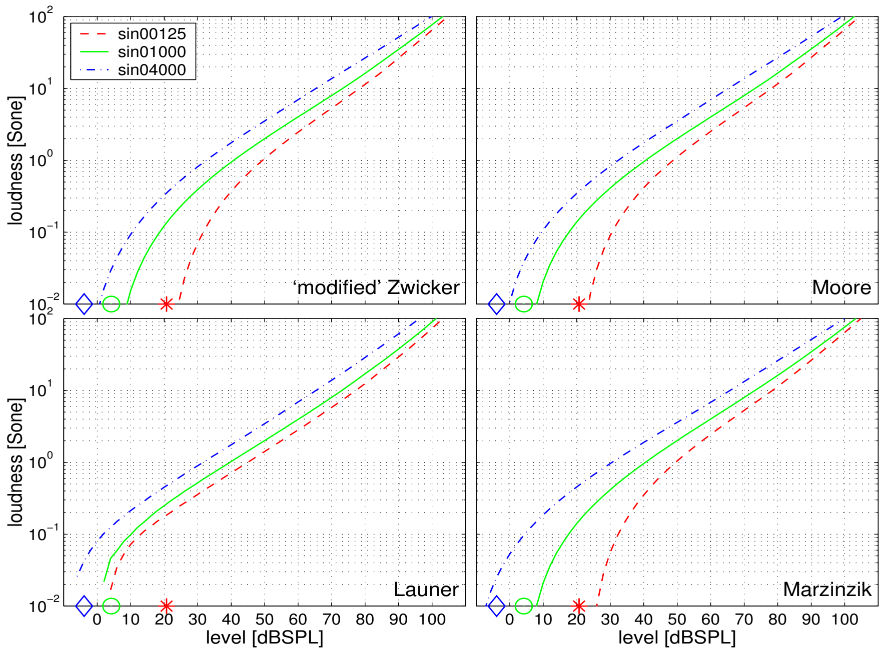

Figure 4.6: *Loudness functions for sinusoids at 125, 1000 and 4000 Hz (solid, dashed and dash–dotted lines, respectively). From left to right and top to bottom the panels show the predictions for the model proposed by Zwicker, Moore, Launer and Marzinzik. The symbols on the abscissa denote the MAF at frequencies of 125 Hz (stars), 1000 Hz (circles) and 4000 Hz (diamonds).*

As expected from the model's respective formula for calculating specific loudness, the models after Launer and Marzinzik differ markedly near threshold. The way in which the formulas account for threshold define the absolute threshold as well as the slope of the loudness function. The slopes derived from the model proposed by Marzinzik show only small differences from the slopes predicted by Zwicker's and Moore's model, whereas the predictions obtained with Launer's model deviate markedly. As expected from Equation 4.11 no steepening of the loudness functions at low levels is observed in the Launer model. The steepening is only obtained at levels very close to the threshold and can be explained by the way Launer limits specific loudness in his approach (i.e., N' is set to zero when $E_{[ELCC]}$ is equal or below $E^{Thq}_{[ELCC]}$).

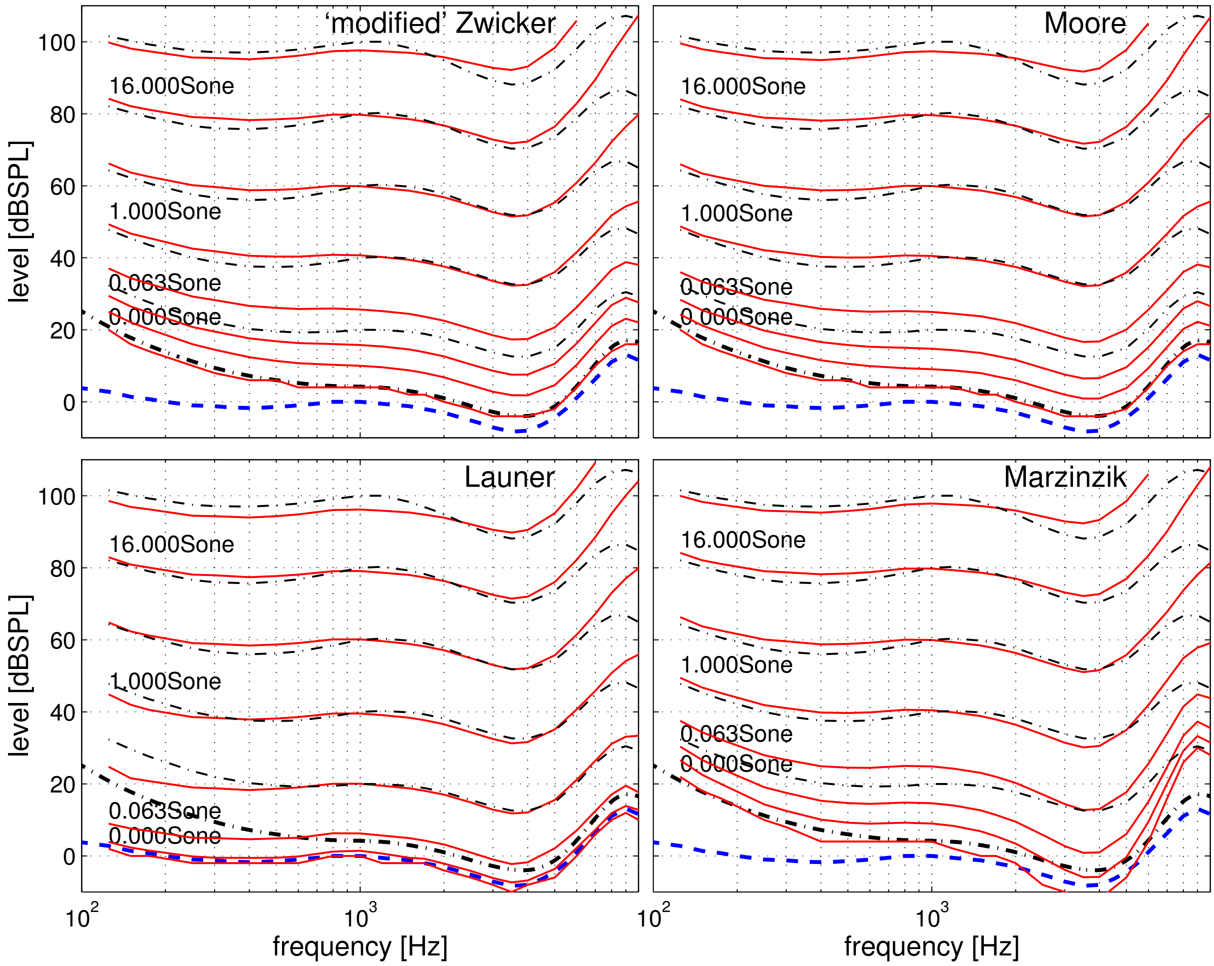

Figure 4.7: *Equal–loudness level contours in sone (ELLC–S) for 0 (absolute threshold), 1/16, 1/4, 1, 4, 16 and 64 sone according to the predictions made by the loudness models. From left to right and top to bottom the panels show the predictions for the model proposed by Zwicker, Moore, Launer and Marzinzik. The thick dash–dotted line corresponds to the MAF. The thick dash line corresponds to the ELLC.The dash–dotted lines in each panel show equal–loudness level contours in phon (ELLC–P) according to ISO 226(E) [1987] for the range of 20 to 100 phon in steps of 20 phon.*

In Figure 4.7 the predictions of the models are tested by comparing the predictions with the equal–loudness level contoursin phon (*ELLC–P*) taken from ISO 226(E) [1987]. This is done by plotting ELLC in sones (*ELLC–S*) calculated by the loudness models for tones at a variety of frequencies. Again, a mismatch between the data and the predicted MAF is observed for Launer's and Marzinzik's approaches. However, all models show the expected doubling in loudness when the level of a 1 kHz tone is increased by 10 dB for levels above 40 dBSPL. In addition, there is almost no difference in the predictions at very high levels. This was expected because the same linear correction (i.e., the ELC–correction) is applied in all models. Moreover, because the ELC–correction and the 100 phon ELLC–P have the same shape, the model predictions show good accordance with the shape of the 100 phon ELLC–P at all frequencies.

4.5 Summary and Discussion

The principle relation between the intensity of a stationary sound and its loudness was discussed and several alternative extensions of the simple power law to account for the steepening of the loudness function near absolute threshold were reviewed. Then, the more generalized loudness models for predicting stationary loudness of arbitrary sounds based on the model proposed by Zwicker [1958] are examined. The revision of Zwicker's loudness model proposed by Moore and Glasberg [1996], as well as the models proposed by Launer [1995] and Marzinzik *et al.* [1996b], use the same processing stages but differ in the linear correction applied in the first stage of the model, the calculation of excitation and the way how specific loudness is calculated from the excitation pattern. Moore and Glasberg showed that there are several advantages of their implementation over Zwicker's original implementation [for details refer to Moore and Glasberg, 1996]. In the present study, we introduced a common implementation of the models and performed a brief evaluation of all four models so that the only difference between the different models tested is in the way how specific loudness is calculated from excitation. While the models by Launer [1995] and Marzinzik *et al.* [1996b] were used in their respective original version, the Zwicker model had to be adapted to the new framework. From the predictions for normal hearing (Figures 4.6 and 4.7) it can be concluded that this adaptation results in almost equal predictions when compared with the predictions made with the model proposed by Moore and Glasberg [1996]. Both approaches were able to give a good prediction of the ISO 226(E) [1987] equal–loudness level contours (ELLC) at all levels. In contrast, the two models proposed by Launer [1995] and Marzinzik *et al.* [1996b] show significant deviations for tones at low stimulus levels from the ELLC's proposed by ISO 226(E) [1987], whereas the two models show good agreement with ISO 226 at high levels. It is obvious that the models proposed by Launer and Marzinzik *et al.* should be improved in order to predict loudness near threshold more accurately.

For predicting individual loudness data for a hearing–impaired subject, the models proposed by Launer [1995] and Marzinzik *et al.* [1996b] rely on data obtained from a loudness scaling procedure and therefore require a transformation between the loudness in sone and the loudness in categorical units. It was found by Blum [1999] that the transformation suggested by Launer [1995] leads to good results in the framework of the Launer model but not within the framework of the Marzinzik model. Therefore Blum introduced a new transformation. However, because of the non critical setup he used in his evaluation, this transformation requires further evaluation.

Finally, it should be pointed out, that the exact form of the ELLC–P at low frequencies published in ISO 226(E) [1987] is a matter of controversy. Several studies [Betke, 1991; Fastl *et al.*, 1990; Gabriel *et al.*, 1994; Reckhardt, 2000; Suzuki *et al.*, 1989; Watanabe and Møller, 1990] have been published suggesting higher levels than the standard values of ISO 226(E) [1987] for all frequencies below 1 kHz. Therefore it seems reasonable to consider a modified loudness model that accounts for the raised ELLC–P's at low frequencies based on recent data.

Chapter 5

The Oldenburg Loudness Model: An extension of Zwicker's loudness model to predict categorical loudness scaling in normal and hearing–impaired listeners

Abstract

A loudness model for stationary sounds based on the model proposed by Zwicker [Zwicker, 1958; Zwicker and Fastl, 1990] is introduced. The model is designed for predicting loudness perception in normal–hearing and hearing–impaired listeners and provides a transformation between the loudness given in sone and the loudness given in categorical units (CU) as it results from loudness scaling experiments. Earlier publications by Hohmann and Kollmeier [1995b], Launer [1995], Marzinzik *et al.* [1996b] and Blum [1999] are assembled and revised in the model, concerning 1) a more accurate prediction of loudness perception near absolute threshold, 2) recently suggested modifications to the equal–loudness level contours as they are standardized in the standard ISO 226(E) [1987] and 3) the transformation between loudness in sone and loudness in CU. The modifications are based on loudness scaling data for narrowband stimuli from 84 normal–hearing subjects. All modifications were introduced in order to achieve improved prediction of loudness perception for the normal hearing, whereas the principle way of accounting for hearing impairment by adapting the compressive exponent in the transformation between excitation level and specific loudness — as it was proposed by Launer [1995] — remains unchanged. An evaluation of the model with normal–hearing and hearing–impaired listeners is given in chapter 6.

5.1 Introduction

The loudness model proposed by Zwicker [Zwicker, 1958; Zwicker and Fastl, 1990] accounts for predicting the loudness of arbitrary steady sounds and has been used in many practical applications. In Zwicker's model, loudness is calculated from the frequency spectrum of a sound in four subsequent stages: The first stage is a fixed filter that accounts for the transmission of the sound through the outer and middle ear. In the second stage the spectrum is transformed into an excitation pattern, which he derived from the spectral masking produced by the input spectrum. In stage three, specific loudness is calculated from the excitation pattern and is then integrated across frequency in the last stage reflecting the total loudness in sone. Moore and Glasberg [1996] followed the principle structure of Zwicker's model but made several modifications, that allow for a more accurate prediction of equal–loudness level contours. They used a linear correction applied in stage 1 of the model corresponding to the shape of the 100 Phon equal–loudness level contour (*ELC–correction*, see Figure 4.3). In addition, they used the concept of auditory filters [for details refer to Glasberg and Moore, 1990; Moore and Glasberg, 1987] based on the equivalent rectangular bandwidth (*ERB*) scale [Patterson and Moore, 1986; Patterson *et al.*, 1987] for calculating the excitation pattern. This allows for the calculation of the excitation from analytical formulae.

Launer [1995] followed these two modifications and, additionally, extended the model to account for hearing impairment. For calculating specific loudness N' from the excitation pattern, he used the ELC–corrected level $E_{[ELCC]}$ in the simplified equation

$$N' = C \cdot \left(\frac{E_{[ELCC]}}{E_{\substack{Thq \\ [HL]}}} \right)^{\beta \cdot \alpha} , \tag{5.1}$$

where $E_{\substack{Thq \\ [HL]}}$ equals the excitation at audiometric threshold relative to average normal hearing threshold[1]. C and α are frequency–independent scaling factors providing a loudness of $N = 1$ sone for a sinusoidal input signal at 1 kHz at a level of 40 dBSPL and a doubling of the loudness per 10 dB increase of the input level for input levels above 40 dBSPL. Equation 5.1 accounts for hearing impairment in two ways: First, the raised threshold is simulated by attenuating the excitation level $E_{[ELCC]}$ within each auditory filter by the audiometric threshold $E_{\substack{Thq \\ [HL]}}$. This is done before the power law is applied. Second, loudness recruitment is accounted for by increasing the compressive exponent α by a pre–factor β, such that loudness increases more rapidly with increasing excitation level. This second component is independent from the attenuation loss and reflects the loss in compression. The factor β describes the frequency dependence of the exponent. Since a smooth variation of β can be assumed as a function of frequency, a polynomial fit of second order is used for β:

$$\beta = \beta(f_{ERB}) = a \cdot (f_{ERB})^2 + b \cdot f_{ERB} + c \tag{5.2}$$

f_{ERB} denotes the frequency in ERB and the parameters a, b and c are fitted to the individual loudness scaling data for narrowband stimuli using a least–squares method [Press *et al.*, 1992].

[1] Troughout this paper, the excitation E (output of the excitation pattern model) is a unitless quantity that takes an index in square brackets describing the calibration of the input to the excitation pattern model. E.g., $E_{[ELCC]}$ is the excitation calculated from a signal that is referenced to the predefined (frequency-dependent) ELC–correction.

5.1.1 Calculating Specific Loudness according to Marzinzik

Marzinzik *et al.* [1996b] noted that the loudness model of Launer does not predict correctly the total loudness in normal hearing listeners ($\beta = 1$ and $E_{\underset{[HL]}{Thq}} = 1$ in Equation 5.1). They showed that the formula proposed by Launer does not allow for a correct prediction of the normal hearing threshold (minimum audible field, MAF), i.e., N' does not approach zero if E approaches threshold. Moreover, Launer's approach does not show the typical steepening of the loudness function at low levels. To account for these drawbacks, Marzinzik *et al.* proposed a formula that is similar to the one proposed by Moore and Glasberg [1996], which accounts for the hearing threshold by assuming an inaudible internal noise providing a certain specific loudness that has to be subtracted from Equation 5.1. The formula proposed by Marzinzik *et al.* [1996b] is

$$N' = \begin{cases} C \cdot \left(\left(\dfrac{E_{[ELCC]}}{E_{\underset{[HL]}{Thq}}} \right)^{\beta \cdot \alpha} - \left(E_{\underset{[SPL]}{MAF}} \right)^{\beta \cdot \alpha} \right) & \text{for} \quad \dfrac{E_{[ELCC]}}{E_{\underset{[HL]}{Thq}}} > E_{\underset{[SPL]}{MAF}} \\ 0 & \text{for} \quad \dfrac{E_{[ELCC]}}{E_{\underset{[HL]}{Thq}}} \leq E_{\underset{[SPL]}{MAF}}, \end{cases} \tag{5.3}$$

with $E_{\underset{[SPL]}{MAF}}$ being equal to the MAF in units of sound pressure level (SPL). Note, that the excitation level $E_{[ELCC]}$ in Equation 5.3 is given relatively to the ELC–correction applied in the first stage of the model, i.e., $E_{[ELCC]}$ is already corrected by a certain portion of the MAF. However, Marzinzik *et al.* subtracted the quantity $E_{\underset{[SPL]}{MAF}}$ which equals the MAF according to ISO 226(E) [1987], i.e., a portion of the MAF is considered twice (for illustration see Figure 4.3). As a consequence of this, the specific loudness in Equation 5.3 does not exactly approach zero when $E_{[ELCC]}$ approaches the absolute threshold. In section 5.2.1 a modified version of Equation 5.3 will be introduced that gives a better prediction of the MAF.

5.1.2 Equal–Loudness Level Contours

The shape of the equal–loudness levels in phon ($ELLC$–P) suggested by ISO 226(E) [1987] has recently been shown to be not correct. Several studies suggested higher sound pressure levels than the standard values given in the ISO 226(E) [1987] at fixed phon values for frequencies below 1 kHz [Betke, 1991; Fastl *et al.*, 1990; Gabriel *et al.*, 1994; Suzuki *et al.*, 1989; Watanabe and Møller, 1990; for a review see Gabriel, 1996; Reckhardt, 2000]. This mismatch was investigated by comparing a test tone in loudness with a reference using several different experimental set–ups [constant stimuli procedure by Betke, 1991; Reckhardt, 2000; Suzuki *et al.*, 1989; 2–alternative forced–choice procedure by Fastl *et al.*, 1990; Reckhardt, 2000; method of adjustment by Betke, 1987 and method of limits by Suzuki *et al.*, 1989; Watanabe and Møller, 1990]. A revision of the ISO 226 based on the data listed above is currently in progress [Committee Draft ISO/CD 226, 2000, also see Figure C.1 in Appendix C].

An alternative procedure to derive equal–loudness level contours uses loudness scaling data for narrowband stimuli. In the present study, data from 84 normal–hearing subjects (aging from 20 to 41 years, hearing threshold below 10 dB HL for frequencies between 125 Hz and 4 kHz and below 15 dB for 6 to 8kHz) were evaluated using the non–adaptive *Oldenburg*

Loudness Scaling Procedure and were compared with the ISO 226 [ISO 226(E), 1987 and Committee Draft ISO/CD 226, 2000]. The categorical loudness procedure is described in detail by Hohmann and Kollmeier [1995b] and Brand [2000] and is also summarized in section 3.2.3. Third–octave bands of noise with center frequencies of 250, 500, 1000, 2000, 3000, 4000, 6000 and 8000 Hz were presented monaurally to the subjects via headphones (Sennheiser HDA 200). After completing the loudness scaling measurements, the individual loudness function for each stimulus, i.e., the loudness in categorical units (CU) as a function of the sound pressure level, was approximated for each stimulus using the model function proposed by Brand *et al.* [1998] (for details see appendix B.2). Finally, ELLC's at categorical loudness values of 5, 15, 25, 35 and 45 CU (corresponding to the verbal response categories 'very soft', 'soft', 'medium', 'loud' and 'very loud', respectively) were calculated for all frequencies from the fitted model function.

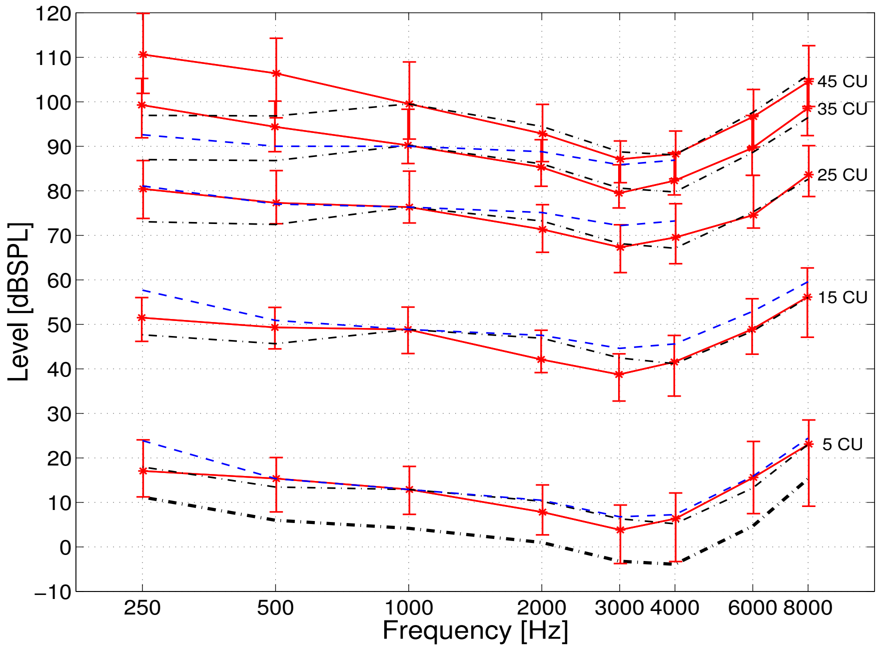

Figure 5.1: *Median equal–loudness level contours calculated from loudness scaling data (ELLC–CU's) for narrowband noise stimuli. The error bars indicate the interindividual 25 % and 75 % percentiles. The thick dash–dotted line corresponds to the MAF according to ISO 226(E) [1987]. The thin dash–dotted lines show ELLC's in phon (ELLC–P's) according to ISO 226(E) [1987]. The dashed lines show ELLC–P's according to the Committee Draft ISO/CD 226 [2000] in which ELLC–P's are only given up to 90 Phon (only up to 70 Phon for frequencies above 4 kHz).*

Figure 5.1 shows the interindividual median results obtained from the 84 normal listeners. The dash–dotted lines show ELLC–P's according to the ISO 226(E) [1987]. In order to compare the data, the levels obtained for the respective CU at 1 kHz were taken as the phon level of the respective ELLC–P. If both data sets would represent the same general dependence of loudness levels across frequency, it is expected that the equal–loudness level contours calculated from categorical loudness scaling data (*ELLC–CU*) for a certain categorical unit and its respective ELLC–P show the same shape across frequencies. As can be seen from Figure 5.1, this is certainly not the case. Figure 5.1 shows that in the low frequency region the ELLC–CU have higher values than the corresponding ISO 226 curves (dash–dotted lines). The largest deviation of about 15 dB is observed at very high levels and decreases towards lower levels. No deviation is observed near threshold. This general finding is in agreement with the literature cited above. The amount of the deviation is very similar to the data presented by Fastl *et al.* [1990] and Suzuki *et al.* [1989]. The dashed lines in Figure 5.1 also show ELLC–P's, but this data is taken from the tables of the recent draft version of the ISO 226 [Committee Draft ISO/CD 226, 2000] in which ELLC–P's are only given up to 90 Phon (only up to 70 Phon for frequencies above 4 kHz). Comparing the ELLC–CU data with the draft version of the ISO 226 [Committee Draft ISO/CD 226, 2000], the categorical loudness scaling data shows a better agreement with the preliminary new ELLC–P's (that are not standardized yet) than with the current standardized ISO 226 data. It can be concluded, however, that any new version of a loudness model should be consistent with the preliminary new ELLC–P's and the ELLC–CU's presented here.

It can be seen in Figure 5.1 that the distance between the ELLC's obtained from the loudness scaling experiments is approximately constant across frequency and that the shape of both, the ELLC–CU's and the new ELLC–P's is similar to the MAF. One possible consequence of this finding is that the general shape of the MAF can not only be used to predict loudness close to absolute threshold but also to predict loudness at higher levels. This finding will be used in section 5.2.2 for modifying the loudness model in a way that it accounts for ELLC's that are similar to the ELLC's obtained from the loudness scaling experiments.

5.1.3 Transformation from Loudness in Sone to Loudness in Categorical Units

As mentioned above, the loudness model of Marzinzik *et al.* [1996b] — as well as its predecessor, the model proposed by Launer [1995] — accounts for the reduced dynamic range of sensorineuraly impaired hearing by fitting the pre–factor β to measured individual loudness scaling data for narrowband stimuli. This is done by an iterative least–squares technique [Press *et al.*, 1992] in which the model predictions are compared with the measured data. For this, the model predictions in sone have to be transformed into the loudness in categorical units as it is measured in the loudness scaling procedure.

Two transformations have been suggested in the past and were introduced in section 4.3.5, the transformation proposed by Hohmann [1993] and Launer [1995] (Equation 4.15) and the transformation by Blum [1999] (Equation 4.16). Whereas the transformation proposed by Launer [1995] was optimized for his loudness model, the transformation proposed by Blum [1999] was optimized for the model proposed by Marzinzik *et al.* [1996b]. In section 5.2.3 we will propose a new transformation between the sone and the CU–scale that seems to be more appropriate for the modified model introduced here.

5.2 Model Extensions

5.2.1 Calculation of Specific Loudness

In order to correct for the inaccuracy of the model proposed by Marzinzik *et al.* [1996b], we propose to replace the quantity $E_{\underset{[SPL]}{MAF}}$ (MAF in units of sound pressure level) in Equation 5.3 by the quantity $E_{\underset{[ELCC]}{MAF}}$, were $E_{\underset{[ELCC]}{MAF}}$ equals the MAF according to ISO 226(E) [1987] given in dB relative to the ELC–correction. Equation 5.3 becomes

$$N' = \begin{cases} C \cdot \left(\left(\frac{E_{[ELCC]}}{E_{\underset{[HL]}{Thq}}} \right)^{\beta \cdot \alpha} - \left(E_{\underset{[ELCC]}{MAF}} \right)^{\beta \cdot \alpha} \right) & \text{for} \quad \frac{E_{[ELCC]}}{E_{\underset{[HL]}{Thq}}} > E_{\underset{[ELCC]}{MAF}} \\ 0 & \text{for} \quad \frac{E_{[ELCC]}}{E_{\underset{[HL]}{Thq}}} \leq E_{\underset{[ELCC]}{MAF}} \, . \end{cases} \tag{5.4}$$

Note, that for normal hearing, when $E_{\underset{[HL]}{Thq}}$ and β are equal to one, Equation 5.4 equals the formula proposed by Moore and Glasberg [1996]. For impaired hearing, the raised threshold $E_{\underset{[HL]}{Thq}}$ — as already considered by Launer [1995] — produces a linear attenuation of the excitation level $E_{[ELCC]}$. This can be interpreted as the *attenuation component* of the hearing loss. Note however, that an individual adjustment of the pre–factor β produces a steepening of the loudness function which can be interpreted as the *compression–loss component*. In the current model, this is an independent component not determined by the first component. It is motivated by loudness scaling data from a large number of hearing–impaired subjects [Kießling *et al.*, 1994; Launer *et al.*, 1996]. The model is in contrast to the One–component approach by Moore [1995] as well as to the 'standard set–up' of the Two–component approach by Moore and colleagues [Moore and Glasberg, 1997; Moore *et al.*, 1996, 1999b].

A comparison of the model predictions according to Equations 5.3 and 5.4 with the standard equal–loudness level contours in phon (*ELLC–P*) as proposed by ISO 226(E) [1987] is given in Figure 5.2. The predictions show data for the normal–hearing system ($\beta = 1$ and $E_{\underset{[HL]}{Thq}} = 1$). The constant parameters C and α were set to 0.0752 and 0.2126 in Equation 5.3 and to 0.0730 and 0.2159 in Equation 5.4, respectively. C and α were fitted using a least–squares technique [Press *et al.*, 1992] such that a 40 dBSPL 1 kHz tone produced a loudness of 1 sone and a 10 dB increase in level of a 1 kHz tone led to a doubling of the predicted loudness for a level range from 40 to 80 dBSPL.

The predictions of the two models show noticeable differences only at low levels. In Marzinzik's original model the zero sone ELLC differs from the expected MAF–curve (thick dash–dotted line). This is because the ELC–correction (thick dashed line) is subtracted from the stimulus spectrum before the excitation level is calculated (first stage of the model) and, in addition, the (compressed) MAF given in units of sound pressure level is subtracted from the (compressed) excitation level in stage three of the model (Equation 5.3), hence the zero sone ELLC equals approximately the sum of the ELC–correction and the MAF. The modified model predicts the MAF with higher accuracy than Marzinzik's original model and in general predicts the shape of the ELLC–P's suggested by the ISO 226(E) [1987] very well. As expected, the models show very little differences at high levels, i.e., when E_{MAF} in Equations 5.3 and 5.4 is negligible.

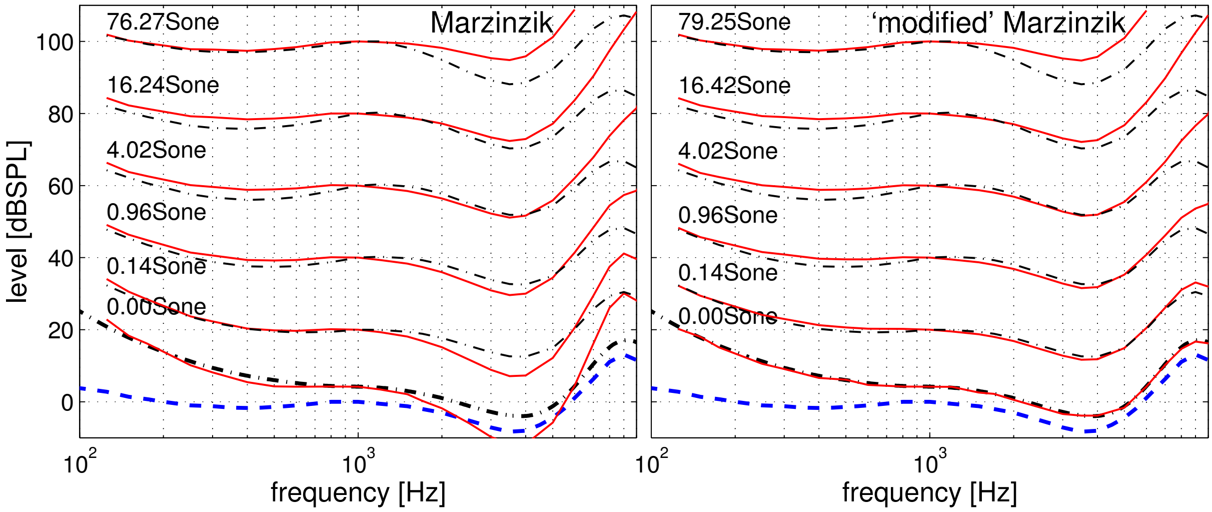

Figure 5.2: *Equal–loudness level contours in sone (ELLC–S, solid lines) according to the predictions made by the model proposed by* Marzinzik et al. *[1996b] (left panel; Equation 5.3) and the modified model (right panel; Equation 5.4). The dash–dotted lines in each panel show equal–loudness level contours in phon (ELLC–P) according to ISO 226(E) [1987] for the range of 20 to 100 phon in steps of 20 phon. The thick dash–dotted line corresponds to the MAF, whereas the thick dashed line shows the ELC-correction.*

5.2.2 Adaptation to Measured ELLC's

The comparison between the modified loudness model and the ISO 226(E) [1987] (Figure 5.2) showed general agreement. However, a revision of the ISO 226 concerning raised ELLC's at low frequencies and high levels [as already suggested in Committee Draft ISO/CD 226, 2000] will require a modification of the loudness model. The revision of the ISO 226 has the following implications on the loudness models: First, the MAF remains unchanged. Second, the ELC–correction (which is derived from ELLC at 100 phon) has to be replaced by a correction that is similar in shape to the ELLC–CU observed in the loudness scaling experiment at high levels. Note, that the shape of the ELLC's in sone predicted by the model at high levels is similar to the shape of the linear correction applied in the first stage of the model, because for high stimulus levels the subtractive term in Equation 5.4 plays only a minor role. Replacing the ELC–correction with a curve shaped like the ELLC–CU at high input levels will therefore force the model to produce equally shaped ELLC's in sone at high levels as well. Due to the small number of frequencies tested in the loudness scaling experiment, it seems not appropriate to directly use the loudness scaling data. Instead, we assume that the ELLC at 45 CU can be approximated by the shape of the MAF itself. Consequently we replace the former ELC–correction with the *MAF–correction* by shifting appropriately the MAF. Hence, the correction is equal to 0 dB at 1 kHz, as it was the case for the ELC–correction. Figure 5.3 shows the relation between the ELC–correction (dashed line), the MAF–correction (dash–dotted line) and the ELLC–CU at 45 CU (solid line).

Figure 5.3 shows that the MAF–correction differs from the ELC–correction only at low levels and that it can be understood as a moderate approximation towards the 45 CU ELLC

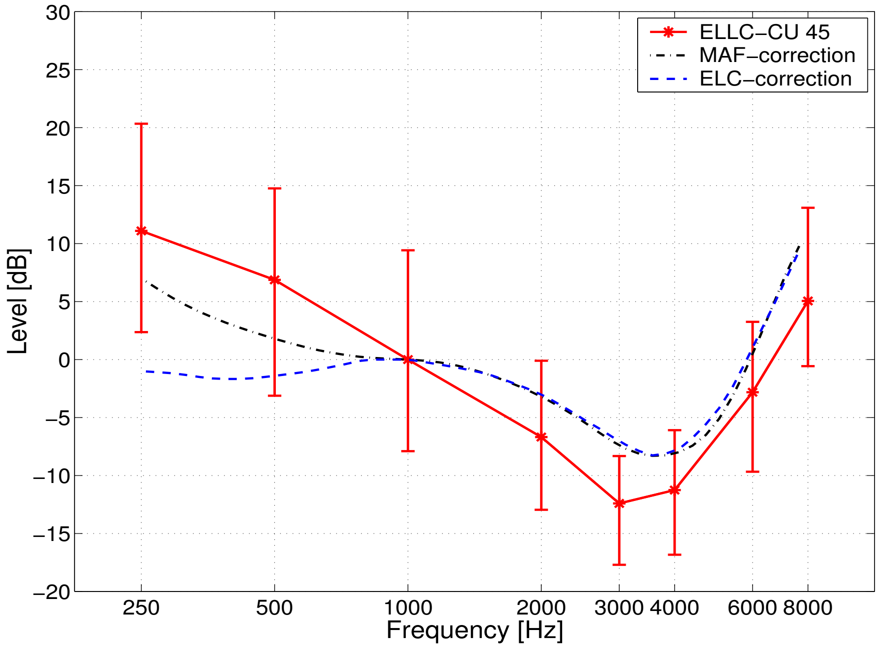

Figure 5.3: *Median ELLC's for the loudness category 45 CU. Error bars show the interindividual 25 % and 75 % percentiles. The dash–dotted line shows the MAF. The dashed line shows the ELC–correction after Moore and Glasberg [1996]. All curves are shifted to match 0 dB at 1 kHz.*

observed in the loudness scaling experiments. Replacing the ELC–correction in Equation 5.4 with the MAF–correction results in

$$N' = \begin{cases} C \cdot \left(\left(\dfrac{E_{[MAFC]}}{E_{\substack{Thq \\ [HL]}}} \right)^{\beta\cdot\alpha} - \left(E_{\substack{MAF \\ [MAFC]}} \right)^{\beta\cdot\alpha} \right) & \text{for} \quad \dfrac{E_{[MAFC]}}{E_{\substack{Thq \\ [HL]}}} > E_{\substack{MAF \\ [MAFC]}} \\ 0 & \text{for} \quad \dfrac{E_{[MAFC]}}{E_{\substack{Thq \\ [HL]}}} \leq E_{\substack{MAF \\ [MAFC]}}, \end{cases} \tag{5.5}$$

i.e., the excitation level $E_{[MAFC]}$ and the excitation at normal hearing threshold $E_{\substack{MAF \\ [MAFC]}}$ are given relative to the MAF–correction. Note, that this modification results in $E_{\substack{MAF \\ [MAFC]}}$ being independent of frequency, so that ELLC's in sone predicted by the model using Equation 5.5 will have the same shape across frequencies. This is shown in Figure 5.4.

Figure 5.4 shows the loudness scaling data for narrowband noise stimuli (solid lines) as already shown in Figure 5.1. The dashed and dash–dotted lines show ELLC's in sone (ELLC–S) for the same stimuli predicted by the model using the ELC–correction (Equation 5.4) and the MAF–correction (Equation 5.5), respectively. As expected from the difference between the ELC– and the MAF–correction (Figure 5.3), both model predictions are identical for frequencies above 1 kHz. Below 1 kHz the model predictions using the MAF–correction are in much better agreement with the ELLC's derived from the loudness scaling experiment and (as can be seen from Figure 5.1) with the preliminary draft version of the ISO 226 [Committee Draft ISO/CD 226, 2000] when compared to the predictions made with the ELC–correction.

However, the predictions of the model according to Equation 5.5 still show a small deviation from the loudness scaling data at high levels because the applied MAF–correction still shows some deviation from the ELLC at 45 CU obtained from the loudness scaling experiment. Hence, a further improvement of the correction accounting for the transmission through the outer and middle should be implemented, when new equal–loudness level contours are published by the International Organization for Standardization.

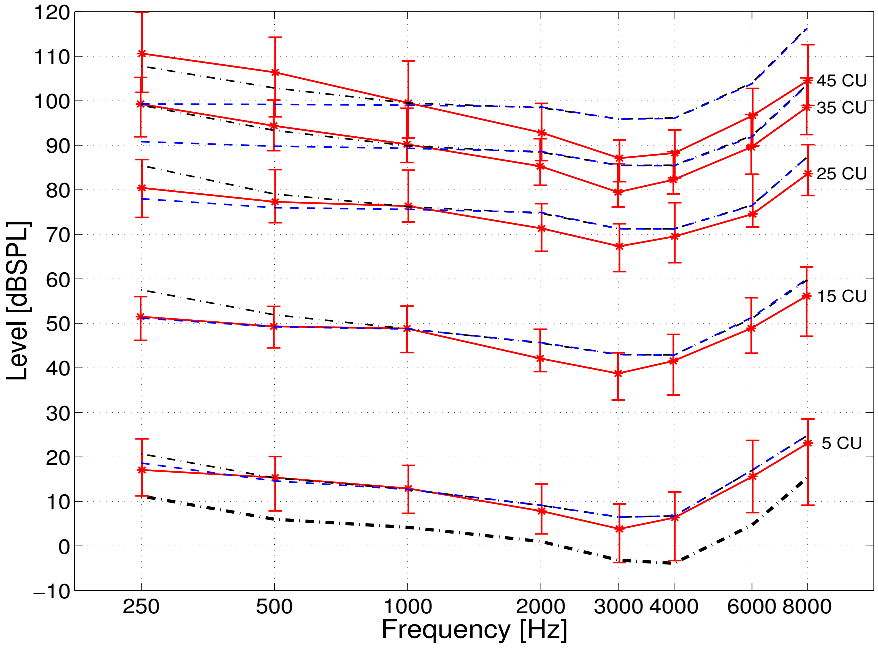

Figure 5.4: *Median ELLC–CU calculated from loudness scaling data for narrowband noise stimuli (solid lines). The error bars indicate the interindividual 25 % and 75 % percentiles. The thick dash–dotted line shows the MAF. The dashed and dash–dotted lines show ELLC's in sone (ELLC–S) for the same stimuli predicted by the model using the ELC–correction and the MAF–correction, respectively.*

5.2.3 Optimized Transformation from Loudness in Sone to Loudness in Categorical Units

In this section a new transformation is introduced that accounts for the modifications of the loudness model as described in the previous sections and which is directly derived from a cubic polynomial fit to the loudness scaling data and the respective predicted loudness in sone. Loudness scaling data were taken from the experiment described in section 5.1. The fitting process included the following steps. First, the mean presentation level in dBSPL was taken at each loudness category and for each center frequency of the narrowband noises as average across all of the subjects responses. Second, the model function proposed by Brand *et al.* [1998] was fitted to the mean data calculated in step one for each center frequency. Third, this model functions were used for calculating loudness functions in CU, while the loudness functions in sone for the respective stimuli were predicted by the loudness model. In

Figure 5.5 these loudness functions are plotted against each other for the respective levels and for each stimulus. In the last step, a cubic polynomial fit to these functions was performed. The resulting transformation between loudness N in sone and loudness CU in categorical units is

$$CU' = 0.3942 \cdot \lg(N)^3 + 3.2018 \cdot \lg(N)^2 + 9.7334 \cdot \lg(N) + 11.2097 \quad (5.6)$$

$$CU = \max(0, CU'). \quad (5.7)$$

This transformation is shown in Figure 5.5 (thick solid line). Thin solid lines show the relation between the measured loudness in CU and the model predictions in sone for each center frequency of the narrowband noises. As can be seen from Figure 5.5, the data is

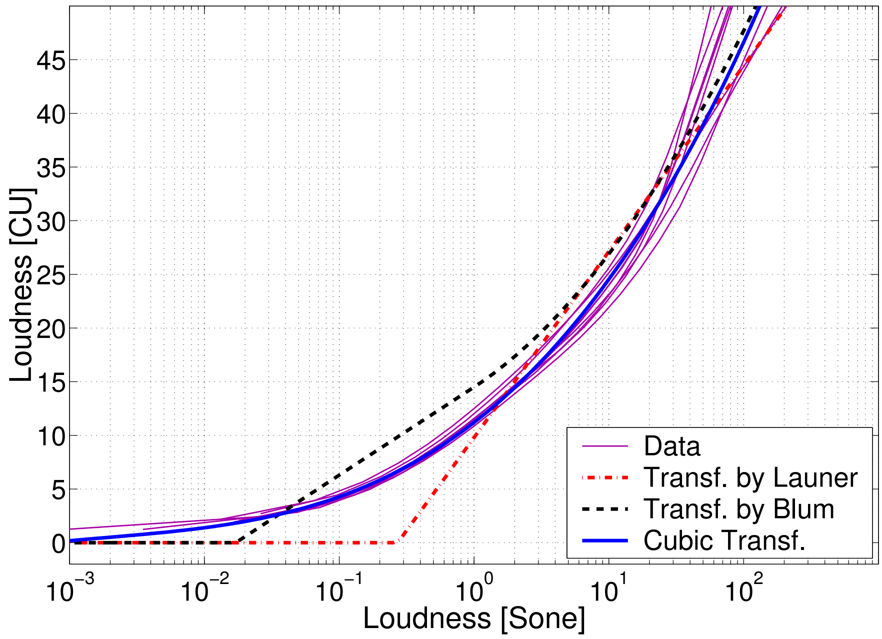

Figure 5.5: *Transformations between loudness in sone and loudness in CU. Thin lines show interindividual mean CU–loudness data for narrowband noise stimuli versus model predictions for the respective stimuli in sone. The thick line shows the cubic polynomial fit to this data (Equation 5.6). The dash–dotted and dashed lines show the transformations proposed by Launer [1995] and Blum [1999], respectively.*

well consistent across frequencies indicating that the transformation between the CU–scale (loudness scaling data) and the sone–scale (model prediction) is not frequency dependent. This consistency is due to the modification of the model described in the previous section, i.e., the replacement of the ELC–correction by the MAF–correction, which improved the similarity of the ELLC shapes predicted by the model with the ELLC's calculated from loudness scaling data. Without this modification, the deviation between the transformation functions at different frequencies would be larger, i.e., the transformation between the CU–scale and the sone–scale would be partly dependent on frequency.

For comparison, Figure 5.5 also shows the transformations proposed by Launer [1995] (dash–dotted line) and the transformation by Blum [1999] (dashed line). Whereas the transformation proposed by Launer fails in predicting categorical loudness at low loudness values, the transformation proposed by Blum [1999] fits the data comparatively well. However the new transformation (Equation 5.6) better approximates the relation between loudness in categorical units and loudness in sone over the whole range of loudness values.

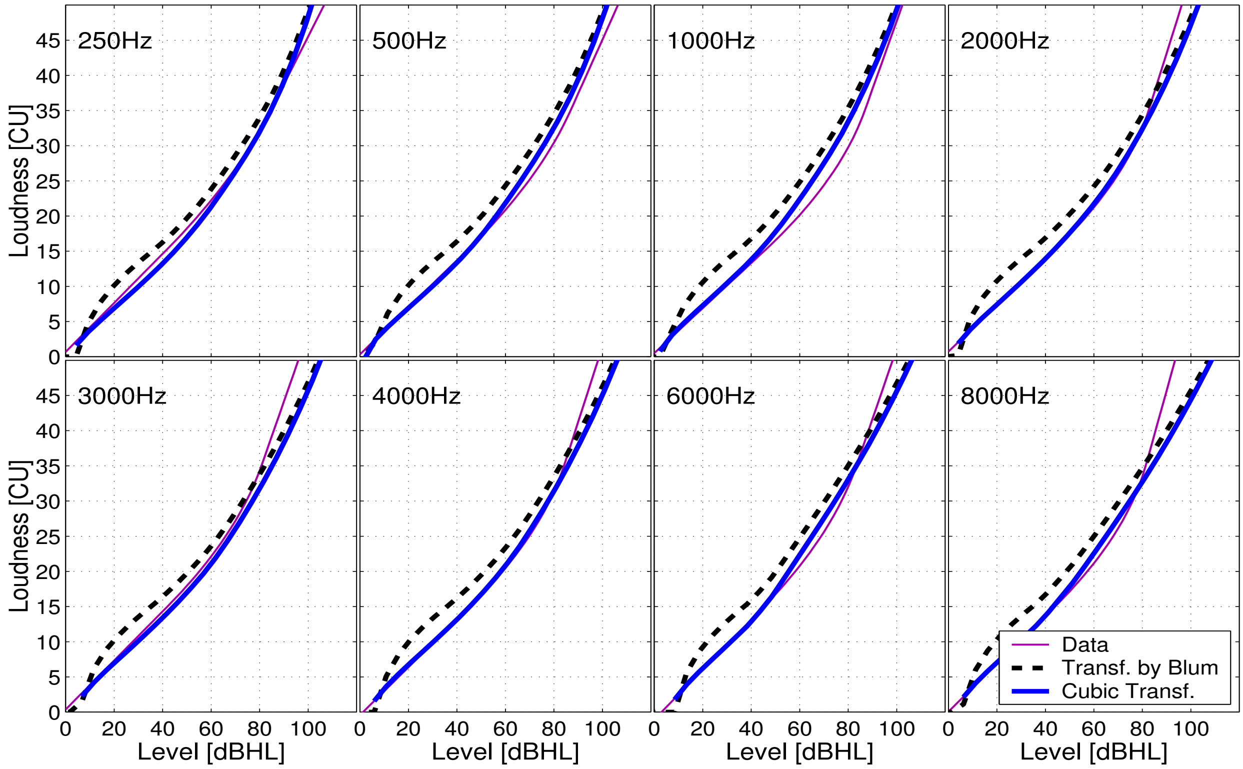

Figure 5.6: *Loudness of narrowband noise stimuli with center frequencies of 0.25, 0.5, 1, 2, 3, 4, 6 and 8 kHz in categorical units (CU) versus level relative to hearing level. Each panel shows the data for the respective center frequency. The thin line shows data obtained from the loudness scaling experiment. The dashed line and the thick solid line show predictions made by the loudness model using the transformation proposed by Blum [1999] and the cubic transformation (Equation 5.6), respectively.*

Figure 5.6 shows loudness functions obtained from the loudness scaling experiment with narrowband noise stimuli (thin lines). The dashed and the thick solid lines show predictions made by the loudness model using the transformation proposed by Blum [1999] and the cubic transformation (Equation 5.6), respectively. Whereas the model predictions using any of the two transformations shows similar agreement with the experimental data at high levels, the model predictions at medium and low levels are more accurate in combination with the cubic transformation introduced here.

5.3 Summary and Discussion

In this chapter two modifications of the loudness model proposed by [Marzinzik *et al.*, 1996b] were introduced. The first modification allows for an accurate prediction of the normal–hearing threshold (minimum audible field, MAF) by modifying the threshold–related subtractive term in the equation for the calculation of the specific loudness (compare Equation 5.4 with Equation 5.3). This modification already results in identical predictions of our model to those of the model proposed by Moore and Glasberg [1996], when the models are applied to normal hearing. The second modification allows for a better prediction of normal equal–loudness level contours (ELLC's) below 1 kHz by proposing a linear correction accounting for transmission through the outer and middle ear that is equal in shape to the MAF (MAF–correction). This correction replaces the ELC–correction first proposed by Moore and Glasberg [1996]. The reference ELLC's, that motivated this modification, were derived from loudness scaling experiments with 84 normal–hearing subjects using 1/3–octave bands of noises. Our data show a general agreement with the results of recent studies investigating the shape of the ELLC's [Betke, 1991; Fastl *et al.*, 1990; Reckhardt, 2000; Suzuki *et al.*, 1989; Watanabe and Møller, 1990]. These studies have shown that the standard ELLC's — as suggested in the ISO 226 standard [ISO 226(E), 1987] — may require a revision for frequencies below 1 kHz, i.e., they suggest higher sound pressure levels at fixed phon values for frequencies below 1 kHz than the ISO 226 standard. In fact, the recent preliminary version of the ISO 226 standard [Committee Draft ISO/CD 226, 2000] adapts these results.

The modification of the loudness model concerning the linear correction applied to the input spectrum (and thereby accounting for the transmission through the outer and middle ear) has several implications. The ELLC's predicted by the model become similar in shape across all frequencies (see Figure 5.4). This indicates that in contrast to the predictions of the model proposed by Moore and Glasberg [1996] and the ISO 226 standard, the dynamic range between the MAF and the ELLC's at high equal–loudness levels is not reduced below 1 kHz when compared to the dynamic range above 1 kHz. This is because the subtractive term in Equation 5.5 is constant across frequencies, whereas the subtractive term used by Moore and Glasberg is constant above 1 kHz but increases below 1 kHz[2]. Moore and Glasberg argued, that below 1 kHz the rise in the MAF with decreasing frequency is accounted for by a linear component (ELC–correction) and a component that relates to an inaudible internal noise being present at low frequencies (e.g., blood flow). As a consequence, they assumed that the inner ear is equally sensitive at all frequencies above 1 kHz. In a later publication Moore *et al.* [1997] applied several modifications to their model. One of it is that they assumed equal sensitivity of the inner ear at all frequencies above 500 Hz. From our model and based on the loudness scaling data presented here, we even assume that the inner ear is equally sensitive at about the entire audible frequency range. This finding is in accordance with recent physiological data from Puria *et al.* [1997] and Aibara *et al.* [2001] who measured transmission through the middle–ear in human cadavers. Especially in the frequency range from 150 Hz to 1 kHz, where the MAF–correction differs from the former ELC–correction (see Figure 5.3), they found a strong correlation between absolute threshold and constant

[2]Being more precise: In the formula for calculating specific loudness, Moore and Glasberg [1996] used the difference between the MAF and the ELC–correction (see Figure 5.2). Their formula is equal to Equation 5.4 for the normal hearing case.

inner–ear pressure.

The modifications to the loudness model showed improved prediction of normal loudness perception based on the sone–scale. To allow for a prediction of loudness in terms of categorical loudness (CU), as it is obtained from loudness scaling experiments, a transformation is required that relates predicted loudness in sone with loudness in CU. Because such a transformation depends on the output of the loudness model in sone, it was tested whether the transformation proposed by Blum [1999] is still appropriate for the modified model. It was found that model predictions using a cubic transformation between the loudness in sone and the loudness in CU (polynomial fit of third degree between the model predictions in sone and the experimental data in CU) better predicts the experimental data than the transformation proposed by Blum [1999] and we therefore recommend this transformation.

The extensions to the Zwicker loudness model described here (which we term the "*Oldenburg Loudness Model*") showed improvements in the prediction of loudness perception in the normal–hearing system. The results of experiments with hearing impaired and the respective model predictions are presented in the next chapter (chapter 6).

Chapter 6

Evaluation of the Oldenburg Loudness Model

Abstract

Three loudness models are tested and compared that extend the loudness model proposed by Zwicker [Zwicker, 1958; Zwicker and Fastl, 1990] to account for hearing impairment: the models proposed by Moore *et al.* [1999b, *MODEL–1*], by Marzinzik *et al.* [1996b, *MODEL–2*], and the model introduced in chapter 5 (*MODEL–3*). Although the models are based on similar assumptions, they differ in several relevant details (i.e., the way to account for normal–hearing threshold, the compatibility with recently revised equal–loudness level contours [Committee Draft ISO/CD 226, 2000], the transformation relating the sone-scale with the categorical loudness scale and the way to account for hearing impairment). The predictions of the loudness models are compared to loudness scaling data for narrow– and broadband stimuli measured with normal–hearing and hearing–impaired subjects (monaural measurement, 12 ears per group). As a result, *MODEL–1* gives better predictions of the loudness perception in the normal–hearing subjects when compared to *MODEL–2*, whereas *MODEL–2* better predicts loudness perception for the group of hearing–impaired subjects. In most conditions, however, *MODEL–3* yields better predictions than *MODEL–1* and *MODEL–2* indicating that this extended version of the Zwicker model (denoted as *Oldenburg Loudness Model*) should be preferred over the other two models.

6.1 Introduction

People suffering from cochlear hearing loss exhibit alterations in several auditory functions, such as loudness perception, intensity discrimination, frequency selectivity, temporal resolution and speech perception [for an overview see Moore, 1995]. Many models have been proposed for the prediction of these alterations [for an overview see Jesteadt, 1997]. In particular, models for the prediction of impaired loudness perception have a great practical impact in the field of hearing aid development [e.g., Bray *et al.*, 1998; Fröhlich, 1993; Hohmann and Kollmeier, 1996; Hohmann *et al.*, 1997; Marzinzik *et al.*, 1996a] and hearing aid fitting [e.g., Moore, 1999, 2000; Moore and Glasberg, 1998; Moore *et al.*, 1999a].

Loudness models for stationary sounds, as considered here, have to account for two distinct aspects of cochlear hearing impairment: First, the raise in absolute threshold (*attenuation component*) and second, *loudness recruitment*, i.e., the more rapid increase of loudness with increasing stimulus level than in normal–hearing listeners due to a reduction of the compressive characteristic of the auditory system. This *compression–loss component* is part of the '*distortion component*' introduced by Plomp and Mimpen [1979]. It may be explained by a damage of outer hair cells (*OHC*) in the cochlea yielding a less compressive basilar membrane input–output function. On the other hand, both a loss of OHC and a loss of inner hair cells (*IHC*) contribute to the attenuation component [for a review on physiological correlates of hearing impairment, see Launer, 1995].

Several studies have shown that the elevated hearing threshold (loss of sensitivity) and the reduced dynamic range (loss of compression) are more or less independent from one another [Hohmann, 1993; Kießling, 1995; Kießling *et al.*, 1994; Launer, 1995; Launer *et al.*, 1996; Moore *et al.*, 1999d]. Therefore it seems likely that a loudness model for the prediction of loudness perception in hearing–impaired listeners should account for these two aspects independently ('*two–component approach*'). In this study, three two–component approaches for predicting normal and impaired loudness perception of steady sounds are compared: proposed Moore *et al.* [1999b] (*MODEL–1*), the model proposed by Marzinzik *et al.* [1996b] (*MODEL–2*), which is a refinement of the first two–component model proposed by Launer [Launer, 1995; Launer *et al.*, 1997] and the "Oldenburg loudness model" introduced in chapter 5 (*MODEL–3*) (See section 6.2 for a detailed description of the differences between the three models).

The predictions of the three loudness models are compared with loudness scaling data for narrowband and broadband stimuli measured with normal–hearing and hearing–impaired subjects (monaural measurement, 12 ears per group). Hence, the coincidence between empirical and theoretically predicted loudness is tested over the complete dynamic range of hearing for a wide range of different stimuli both for normal and hearing–impaired listeners.

6.2 Description of the three Models

The models follow, in principle, the processing stages of the loudness model proposed by Zwicker [Zwicker, 1958; Zwicker and Fastl, 1990], which is designed for the prediction of loudness perception in normal–hearing listeners: fixed filter representing the transfer through the outer and middle ear, calculation of the excitation pattern, transformation into specific loudness and finally integration of the specific loudness to the total loudness, given in sone. In all models used in this study, the excitation pattern is calculated using the concept of auditory filters [Moore and Glasberg, 1987; Patterson and Moore, 1986; Patterson *et al.*, 1987], i.e., the level of excitation at a specific center frequency of the auditory filter is calculated from the amount of energy falling in the respective auditory filter. The variation of the auditory filter bandwidth and shape with frequency and level follows the formulae proposed by Glasberg and Moore [1990] when loudness data for normal–hearing listeners are predicted. However, the models differ in the way they account for broadened auditory filters in the case of cochlear hearing impairment[1]. In *MODEL–1* it is assumed that the loss of OHC has a direct influence on the slopes of the auditory filters. In this model, the excitation pattern becomes broader with increasing hearing loss, while at the same time the pattern changes less with level [Moore *et al.*, 1999b]. In *MODEL–2* and *MODEL–3* [based on the assumptions made by Launer, 1995] it is assumed that auditory filter bandwidth in hearing–impaired listeners is the same as that of normal–hearing listeners at the same sound pressure level (dB SPL) but differs when compared at the same sensation level (dB SL). Consequently, these models calculate the excitation pattern for hearing impaired in the same way as for normal–hearing subjects, while accounting for the broadening of auditory filters (and reduced frequency selectivity) indirectly. However, Launer [1995] as well as Moore [1995] showed that incorporating reduced frequency selectivity in a loudness model will only have a small influence on the predicted slopes of the loudness growth functions, but might be substantial when there is a "*dead region*" [2] of the cochlea [Moore *et al.*, 1996].

In the third modeling stage, a transformation is made that relates the excitation pattern $E(f_{ERB})$ to the specific loudness pattern $N'(f_{ERB})$[3], where f_{ERB} is the frequency on ERB–scale. The most substantial differences between the three models are reflected in this transformation. The formula proposed by Marzinzik *et al.* [1996b] (*MODEL–2*) is

$$N' = \begin{cases} C \cdot \left(\left(\frac{E_{[ELCC]}}{E_{\underset{[HL]}{Thq}}} \right)^{\beta \cdot \alpha} - \left(E_{\underset{[SPL]}{MAF}} \right)^{\beta \cdot \alpha} \right) & \text{for } \frac{E_{[ELCC]}}{E_{\underset{[HL]}{Thq}}} > E_{\underset{[SPL]}{MAF}} \\ 0 & \text{for } \frac{E_{[ELCC]}}{E_{\underset{[HL]}{Thq}}} \leq E_{\underset{[SPL]}{MAF}}, \end{cases} \tag{6.1}$$

where the excitation produced by the stimulus $E_{[ELCC]}$ is expressed relative to the linear attenuation applied in the first stage of the model, i.e., the ELC–correction (as indicated by the index '$_{[ELCC]}$')[4]. The constant values C and α are frequency–independent scaling factors.

[1] Broadened auditory filters in sensorineural impaired listeners are assumed because of their poorer than normal ability to discriminate tones in masking noise (reduced frequency selectivity, for a review see Moore, 1995; Tyler, 1986).

[2] A region of the cochlea where with a complete loss of IHC (or nonfunctional neurons), that can be described in terms of the range of characteristic frequencies that would normally be associated with that region [Moore *et al.*, 2001].

[3] To simplify the formulas given below, $N'(f_{ERB})$ and $E(f_{ERB})$ are denoted by N' and E, respectively, even if they are frequency dependent quantities.

[4] Troughout this paper, the excitation E (output of the excitation pattern model) is a unitless quantity that takes an index in square brackets describing the calibration of the input to the excitation pattern model. E.g., $E_{[ELCC]}$ is the excitation calculated from a signal that is referenced to the predefined (frequency-dependent) ELC–correction.

They are determined such that a loudness of $N = 1$ sone results for a sinusoidal input signal at 1 kHz at a level of 40 dBSPL, and that a doubling of the loudness per 10 dB increase of the input level results for input levels above 40 dBSPL when normal loudness perception is predicted. The quantity $E_{\substack{MAF\\[SPL]}}$ is equal to the excitation produced by the *minimum audible field* [*MAF*, for a definition see ISO 226(E), 1987] in units of sound pressure level. As in Launer [1995], *MODEL–2* accounts for hearing impairment in two ways: First, the raised threshold (attenuation component) is simulated by attenuating the excitation level $E_{[ELCC]}$ within each auditory filter by the audiometric threshold $E_{\substack{Thq\\[HL]}}$. This is done before the power law is applied. Second, loudness recruitment is accounted for by increasing the compressive exponent α by a pre–factor β, such that loudness increases more rapidly with increasing excitation level. This second component is independent of the attenuation loss and reflects the loss in compression. The factor β describes the frequency dependence of the exponent and has to be fitted individually to the loudness scaling data for narrowband stimuli.

In chapter 5 it was pointed out that Equation 6.1 (*MODEL–2*) fails to predict equal–loudness levels contours at low frequencies as they are suggested in several recent studies [for a review see Gabriel, 1996; Reckhardt, 2000] and as they are suggested by the recent draft version of the ISO 226 [Committee Draft ISO/CD 226, 2000]. In addition, it was found that *MODEL–2* does not correctly predict normal–hearing loudness perception near absolute threshold accurately, because specific loudness does not approach zero when the excitation $E_{[ELCC]}$ approaches the absolute threshold. This is because the $E_{[ELCC]}$ and $E_{\substack{MAF\\[SPL]}}$ values are taken relative to different reference values. Therefore, Equation 6.1 is changed in *MODEL–3* so that the linear correction applied in the first stage of the model has the curvature of the MAF, but is shifted to match 0 dB at 1 kHz (referred to as the *MAF–correction*). This has the consequence that the excitation invoked by the stimulus equals to $E_{[MAFC]}$. To ensure that specific loudness becomes zero at absolute threshold, the excitation produced by the MAF taken relative to the MAF–correction $E_{\substack{MAF\\[MAFC]}}$, is subtracted from $E_{[MAFC]}$. As discussed in chapter 5, this has the following consequences for the prediction of normal–hearing loudness perception: First, $E_{\substack{MAF\\[MAFC]}}$ is constant across frequency and corresponds to the absolute threshold at 1 kHz. Second, loudness functions for tones at different frequencies have the same curvature when plotted against the tone level in dB relative to the MAF. Therefore normal–hearing equal–loudness level contours predicted by *MODEL–3* have the same curvature over frequency as the MAF but are shifted to their respective equal–loudness level. The implications of these modifications are discussed in chapter 5 and will be further discussed later. Applying these alteration to Equation 6.1 the equation relating the excitation pattern $E_{[MAFC]}$ to specific loudness N' in *MODEL–3* becomes

$$N' = \begin{cases} C \cdot \left(\left(\dfrac{E_{[MAFC]}}{E_{\substack{Thq\\[HL]}}} \right)^{\beta\cdot\alpha} - \left(E_{\substack{MAF\\[MAFC]}} \right)^{\beta\cdot\alpha} \right) & \text{for} \quad \dfrac{E_{[MAFC]}}{E_{\substack{Thq\\[HL]}}} > E_{\substack{MAF\\[MAFC]}} \\ 0 & \text{for} \quad \dfrac{E_{[MAFC]}}{E_{\substack{Thq\\[HL]}}} \le E_{\substack{MAF\\[MAFC]}} . \end{cases} \tag{6.2}$$

In conjunction with *MODEL–2* and *MODEL–3*, the formula relating the excitation pattern with specific loudness proposed by Moore and colleagues [Moore and Glasberg, 1997; Moore *et al.*, 1996, 1999b] (*MODEL–1*) accounts for hearing impairment by assuming that the overall hearing loss HL_{TOTAL} can be partitioned at a given frequency into a component due to OHC damage (HL_{OHC}) and a component due to IHC (and neural) damage HL_{IHC}[5]:

$$HL_{TOTAL} = HL_{OHC} + HL_{IHC}. \tag{6.3}$$

By introducing the quantities

$$\begin{aligned} f_{IHC} &= 10^{\frac{HL_{IHC}}{10}} \\ f_{OHC} &= 10^{\frac{HL_{OHC}}{10}} \end{aligned}$$

their formula is

$$N' = \begin{cases} C \cdot \left(\left(\frac{E_{[ELCC]}}{f_{IHC}} \right)^{\alpha} - \left(f_{OHC} \cdot E_{\substack{MAF \\ [ELCC]}} \right)^{\alpha} \right) & \text{for} \quad \frac{E_{[ELCC]}}{f_{IHC}} > f_{OHC} \cdot E_{\substack{MAF \\ [ELCC]}} \\ 0 & \text{for} \quad \frac{E_{[ELCC]}}{f_{IHC}} \leq f_{OHC} \cdot E_{\substack{MAF \\ [ELCC]}} . \end{cases} \tag{6.4}$$

In this formula the level of excitation $E_{[ELCC]}$ and the excitation at normal–hearing absolute threshold $E_{\substack{MAF \\ [ELCC]}}$ are taken relative to the ELC–correction applied in the first stage of the model, so it is ensured that loudness becomes zero for the MAF. As can be seen from Equation 6.4, damage of IHCs is modeled by attenuating the excitation pattern by the quantity f_{IHC} and therefore contributes to the attenuation component of the hearing loss, whereas damage of OHCs is modeled by increasing $E_{\substack{MAF \\ [ELCC]}}$ by the quantity f_{OHC}, which results in a steeper growth of the specific loudness with increasing level of excitation and therefore contributes to the loss in compression.

	MODEL–1	*MODEL–2*	*MODEL–3*
Broadening of auditory filters depends on hearing loss	yes	no	no
Accurate prediction of hearing threshold	yes	no	yes
Accounts for new equal loudness level contours [ISO draft]	no	no	yes
Linear correction accounting for outer and middle ear	ELCC	ELCC	MAFC

Table 6.1: *List differences between the three loudness models. See text for details.*

For an overview, the differences between the three models are summarized in table 6.1. The way the three models were adjusted for modeling individual loudness perception is described in detail in section 6.4.1. Note, that the constant parameters C and α in equations 6.1, 6.2 and 6.4 were set for the normal hearing case such that a loudness of 1 sone is

[5]As mentioned in Moore *et al.* [1999b] this does not mean that the proportion of HL_{OHC} and HL_{IHC} is the same as the proportion of OHCs and IHCs that are damaged or lost

achieved for a 1 kHz tone at a level of 40 dBSPL and a doubling in loudness is predicted whenever the level of this tone is increased by 10 dB at high input levels. The respective settings are summarized in table 6.2.

Parameter	*MODEL–1* (Equation 6.4)	*MODEL–2* (Equation 6.1)	*MODEL–3* (Equation 6.2)
C	0.0872	0.0755	0.0730
α	0.2040	0.2122	0.2159

Table 6.2: *Constant parameter setting of the models. Rows C and α show the respective parameters fitted to the models.*

Section 6.3 will present the loudness scaling data measured with normal–hearing and hearing–impaired subjects (monaural measurement, 12 ears per group). Whereas the data obtained for the narrowband stimuli are required for the adjustment of the parameters of *MODEL–2* and *MODEL–3*, the data obtained for the narrow– and broadband stimuli will be compared to the predictions of the loudness models in section 6.4.

6.3 Loudness Scaling Experiments

6.3.1 Method

Loudness data of narrowband and broadband stimuli were measured monaurally with the non–adaptive *Oldenburg Loudness Scaling Procedure* [Brand, 2000; Hohmann and Kollmeier, 1995b]. This procedure consists of two phases. In the first phase, the auditory dynamic range of the subject is estimated by presenting an ascending level sequence. The subject's task is to press the response button as soon as the stimulus is audible. After that, the listener is asked to press the response button immediately when the stimulus is perceived as being too loud. In case that the listener does not press the response button, the sequence stops at a maximum level of 120 dB SPL. In the second phase, the loudness function is assessed by presenting stimuli covering the predetermined individual auditory dynamic range with a uniform distribution of presentation levels. In this phase, the stimuli are presented twice at each of 7 different levels. In order to avoid context effects which are due to the tendency of many listeners to rate the current stimulus relatively to the previous stimulus, the stimuli are presented in pseudo–random order in a way that the maximum difference of subsequent presentation levels is smaller than half of the dynamic range of the sequence. The listeners task is to rate the loudness of the stimuli on the verbal (i.e., 5 main verbal categories, 4 intermediate categories and 2 limiting categories; see Figure 3.1 in chapter 3).

When the loudness scaling measurements are completed, the loudness function for each stimulus is approximated by a model function consisting of two straight lines smoothed by a Bezier curve around their kneepoint [Brand *et al.*, 1998, for details see appendix B.2].

6.3.2 Subjects

6 normal–hearing listeners (5 female, 1 male; aged 25–30 years; median 27 years) and 6 hearing–impaired listeners (5 female, 1 male; aged 22–75 years; median 60 years) participated in the experiments. The hearing threshold of the normal–hearing listeners was better than 15 dB HL at the standard audiometric frequencies from 125 Hz to 8 kHz. Two of the normal–hearing listeners were members of the research group. The other listeners had little experience in psychoacoustical experiments and were paid on an hourly basis. The hearing–impaired subjects showed moderate sloping high frequency hearing losses of cochlear origin. Mean hearing thresholds at .5, 1, 2, 3, 4 and 6 kHz were 37, 44, 53, 54, 58 and 62 dB HL, respectively. The pure tone audiograms of all hearing–impaired subjects are given in Table 6.3 and in Appendix D.1 (Figure D.1). Note, that all hearing–impaired subjects exhibit a sloping high–frequency hearing loss. Hence, the interindividual standard deviation is comparatively low (below 10 dB) for most audiometric frequencies.

		Frequency [Hz]									
Subjects	**Ear**	**125**	**250**	**500**	**1000**	**1500**	**2000**	**3000**	**4000**	**6000**	**8000**
BU	left	40	40	40	45	45	45	60	55	65	60
BU	right	30	30	30	40	55	60	65	65	65	60
HH	left	15	20	35	40	35	35	45	45	60	60
HH	right	5	15	30	35	30	30	40	45	45	65
HM	left	25	25	30	40	45	60	60	65	65	80
HM	right	25	25	35	35	40	50	60	70	65	70
MH	left	20	25	35	45	55	50	50	50	65	70
MH	right	20	30	50	60	60	60	45	45	60	55
MW	left	25	30	45	50	55	70	65	60	65	50
MW	right	35	40	50	60	65	75	55	60	70	55
WH	left	30	25	30	35	40	50	50	75	55	50
WH	right	30	25	35	40	50	55	55	65	60	50
Mean		25	28	37	44	48	53	54	58	62	60
Std		9	7	8	9	11	13	8	10	7	9

Table 6.3: *Pure tone audiogram (air conduction thresholds) in dB HL for the hearing–impaired subjects. The mean value and standard deviation are given in the last two rows for all frequencies, respectively.*

6.3.3 Stimuli

All stimuli (except the speech simulating noise) were generated digitally at a sample rate of 44.1 kHz by generating a sample of white noise of 5 s duration with Gaussian amplitude statistics, transforming the sample into the frequency domain, setting the magnitude of the Fourier coefficients to the desired spectrum and transforming back the spectrum into the time domain. A segment of 2 s duration was selected randomly and windowed with 100 ms $\cos^2$ ramps. Two sets of stimuli were used: narrowband noises and broadband stimuli. The narrowband stimuli were third–octave narrowband noises centered at 0.25, 0.50, 1, 2, 3, 4, 6 and 8 kHz.

The broadband stimuli were bandlimited uniform exciting noise [Zwicker and Fastl, 1990] as well as a speech simulating noise [Kollmeier *et al.*, 1992]. The characteristic of the uniform exciting noise stimuli (*UEN–stimuli*) is that it produces the same intensity in each critical band. Therefore it should produce the greatest effect of spectral loudness summation when the stimulus bandwidth is increased. The UEN–stimuli were geometrically centered at 10.5 Bark. The bandwidths of the signals were 1, 3, 5, 9, and 17 Bark. Table 6.4 shows the corresponding quantities in Herz (more details on the UEN–stimuli are given in Table 3.1).

center frequency [$Bark$]	10.5				
center frequency [Hz]	1370				
bandwidth [$Bark$]	1	3	5	9	17
bandwidth [Hz]	210	640	1080	2070	5100

Table 6.4: *Center frequency and bandwidths of the UEN–stimuli.*

6.3.4 Apparatus

A computer–controlled audiometry workstation was used which was developed within a German joint research project on speech audiometry [Kollmeier *et al.*, 1992]. A personal computer with a coprocessor board (Ariel DSP 32C) with 16–bit stereo AD–DA converters was used to control the complete experiment as well as stimulus presentation and recording of the listener's responses. The stimulus levels were adjusted by a computer–controlled custom–designed audiometer comprising attenuators, anti–aliasing filters and headphone amplifiers. The signals were presented monaurally to the listeners via headphones (Sennheiser HDA 200). The listeners were seated in a sound–insulated booth. Their task was to rate the loudness of each stimulus presented using a handheld computer (Epson EHT 10S) with a LCD touchscreen showing the response scale. The handheld computer was connected to the personal computer via serial interface. The listeners' loudness ratings for each stimulus were stored for later statistical analysis.

6.3.5 Results and Discussion

Figure 6.1 to 6.4 show the pooled responses of the subjects in the loudness scaling experiments (circles) together with the approximated model function (solid line). This function was fitted to the mean categorical loudness values taken across all data points for the respective stimulus. The shaded areas show plus minus one standard deviation of the mean across the subjects responses at fixed categorical loudness.

The data from the loudness scaling experiments with narrowband stimuli are shown in Figures 6.1 and 6.2. As expected, the narrowband loudness functions observed for the group of normal–hearing subjects (Figure 6.1) essentially show no differences for the different center frequencies, whereas the data for the hearing–impaired subjects (Figure 6.2) show increased

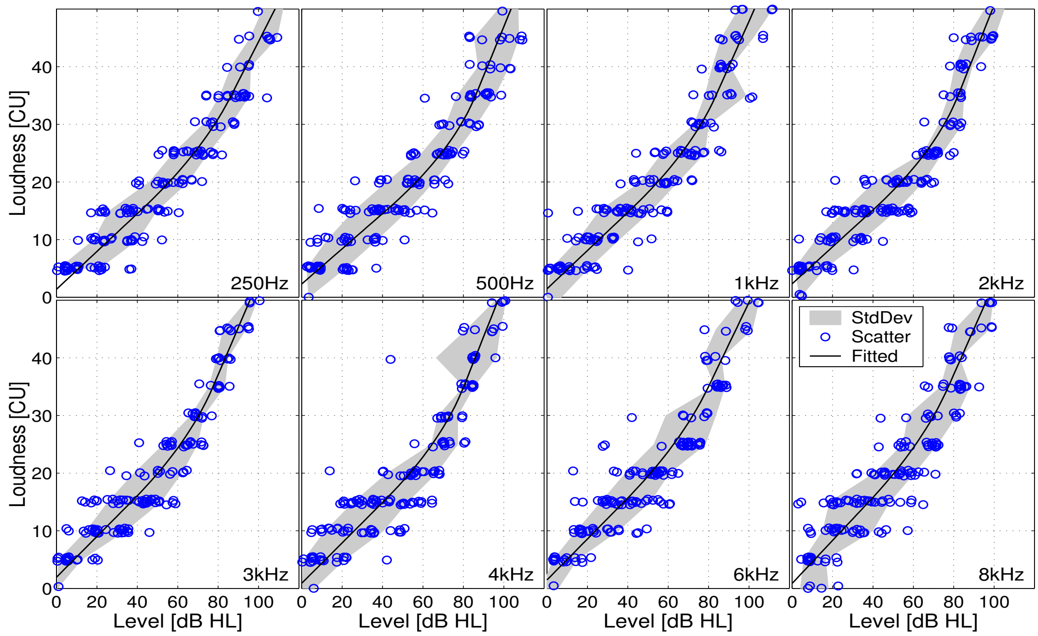

Figure 6.1: *Pooled responses of the normal–hearing subjects in the loudness scaling experiments carried out with the narrowband stimuli (circles). Solid lines show the model function fitted to the mean categorical loudness values taken across all data points for the respective stimulus. The shaded areas show plus minus one standard deviation of the mean across the subjects responses at fixed categorical loudness. Each panel shows the results for center frequencies at 0.25, 0.5, 1, 2, 3, 4, 6 and 8 kHz (left to right and top to bottom, respectively).*

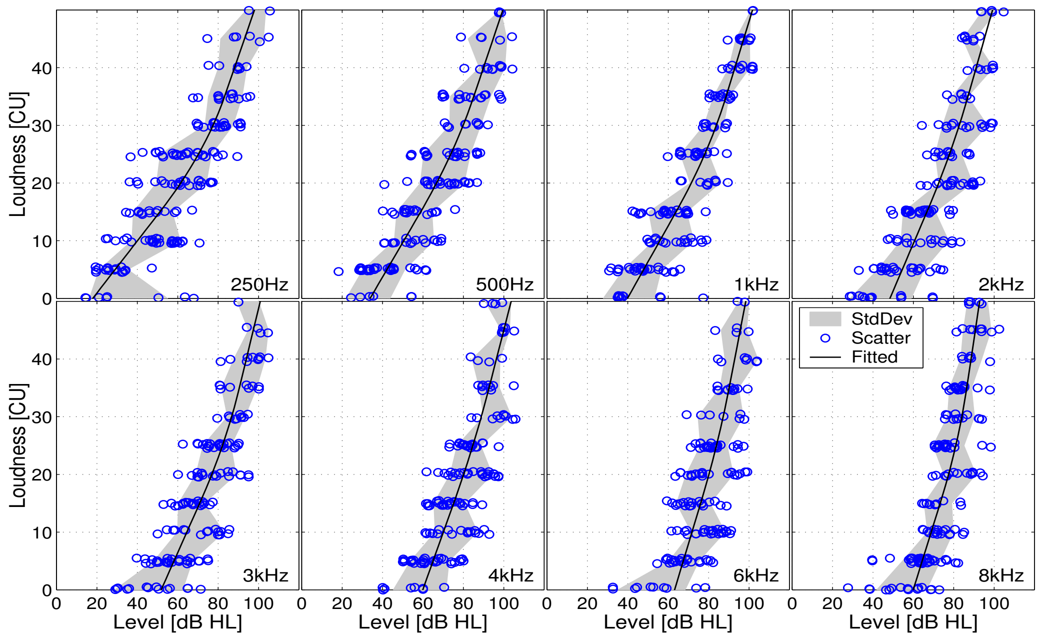

Figure 6.2: *Same as Figure 6.1 for the group of hearing–impaired subjects.*

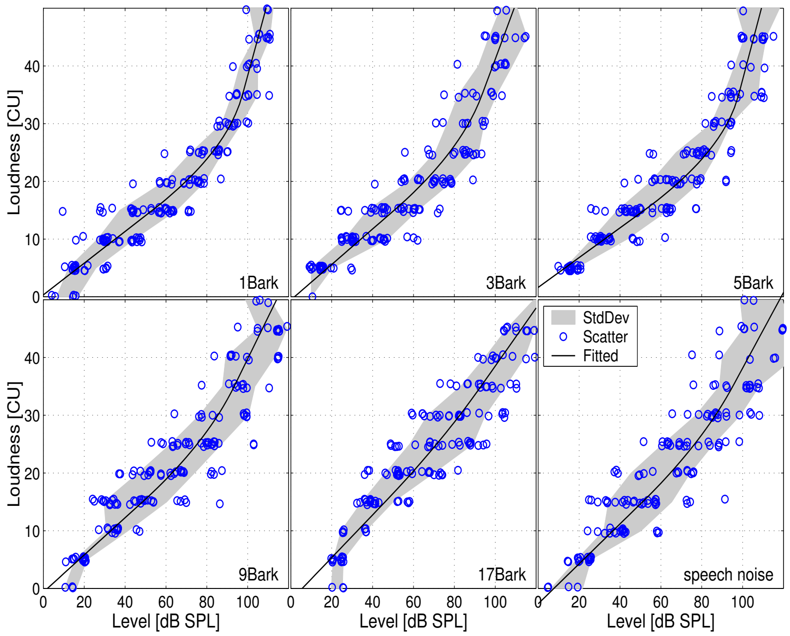

Figure 6.3: *Pooled responses of the normal–hearing subjects in the loudness scaling experiments carried out with the broadband stimuli (circles). Solid lines show the approximated model function. The shaded areas show plus minus one standard deviation of the mean across the subjects responses at fixed categorical loudness. Each panel shows the results for 1, 3, 5, 9 and 17 Bark wide uniform exciting noise and speech simulating noise. (left to right and top to bottom, respectively).*

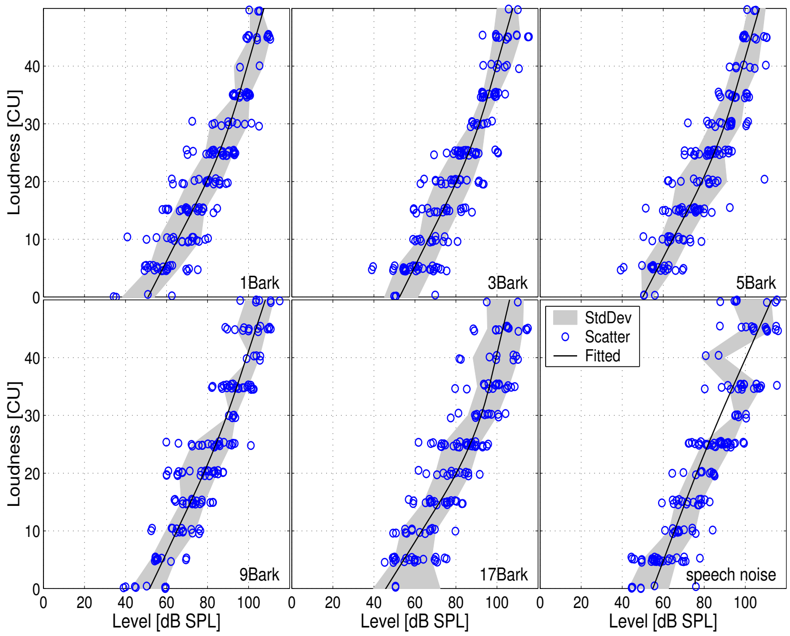

Figure 6.4: *Same as Figure 6.3 for the group of hearing–impaired subjects.*

thresholds and a reduced dynamic range (*loudness recruitment*) especially at high frequencies. Generally, the mean of the responses of the normal–hearing subjects, as well as the respective approximated model function, show an upwardly concave slope. This also holds for the group of the hearing–impaired subjects at frequencies where the subject's hearing thresholds are close to normal. More straightened loudness functions are observed when the subject's hearing thresholds increase. In general, the change of the loudness functions with increasing hearing threshold mainly affects the slope of the loudness functions at low to moderate stimulus levels, indicating that recruitment is most prominent in the lower loudness range.

The data from the loudness scaling experiments with broadband stimuli are shown in Figure 6.3 and 6.4. For the group of normal–hearing subjects (Figure 6.3) it is observed, that the positive curvature of the loudness functions decreases with increasing stimulus bandwidth, while at the same time the stimuli are rated louder, especially at medium levels. This can be explained by spectral loudness summation which is known to be most prominent at medium levels and less prominent at high and low levels. The loudness function observed for the speech simulating noise (bottom–right of Figure 6.3) is in between the loudness function for the two broadest uniformly exciting noises. However, at very high loudness categories loudness summation seems to be strongly reduced or even negative, i.e., at high stimulus levels the subjects rated the loudness of the noises with bandwidths above 5 Bark to be softer in comparison to the loudness of the more narrowband stimuli. It might be suggested that the subject's loudness ratings are influenced by other criteria such as annoyance when high levels are presented and that stimuli with a larger bandwidth might be better tolerated when judged in terms of annoyance than narrowband stimuli[6]. However, due to the principles of the loudness models this can not be predicted by these models. Figure 6.3 also shows that the subjects responses (circles) to presentation levels near hearing threshold indicate a steepening of the loudness functions when the bandwidth increases. However, due to the limitations of the model function (straight lines), this can not be fitted properly.

The loudness functions of the hearing–impaired subjects (Figure 6.4) show a similar dependency on the stimulus bandwidth as for the normal–hearing subjects. However, as expected from the literature [e.g., Launer *et al.*, 1997; Moore, 1995], the effect of spectral loudness summation is strongly reduced. In general, the results of the loudness scaling experiments are in agreement with results presented by Brand and Hohmann [Brand, 2000; Brand and Hohmann, 2001b], who compared individual loudness functions of third–octave narrowband noises with speech simulating noise for normal–hearing and hearing–impaired listeners using a different group of subjects.

In order to compare the dynamic range of the normal–hearing and hearing–impaired group of subjects, Figure 6.5 shows the equal–loudness level contours ($ELLC$) derived from the loudness scaling experiments with the narrowband stimuli. In this case, the model functions were fitted individually and the level corresponding to the respective loudness categories were averaged. In accordance with the data presented in chapter 5 (Figure 5.1), the dynamic range of the normal–hearing subjects (solid lines) depends only little on frequency, i.e., the ELLC's for the different loudness categories are almost shifted parallel on the level axis. The ELLC's for the group of hearing–impaired subjects (dashed lines) show a smaller

[6] A similar suggestion was discussed in more detail in chapter 3

dynamic range than the normal–hearing group. It decreases with increasing frequency and can be explained by the subject's sloping high frequency loss of sensorineural origin.

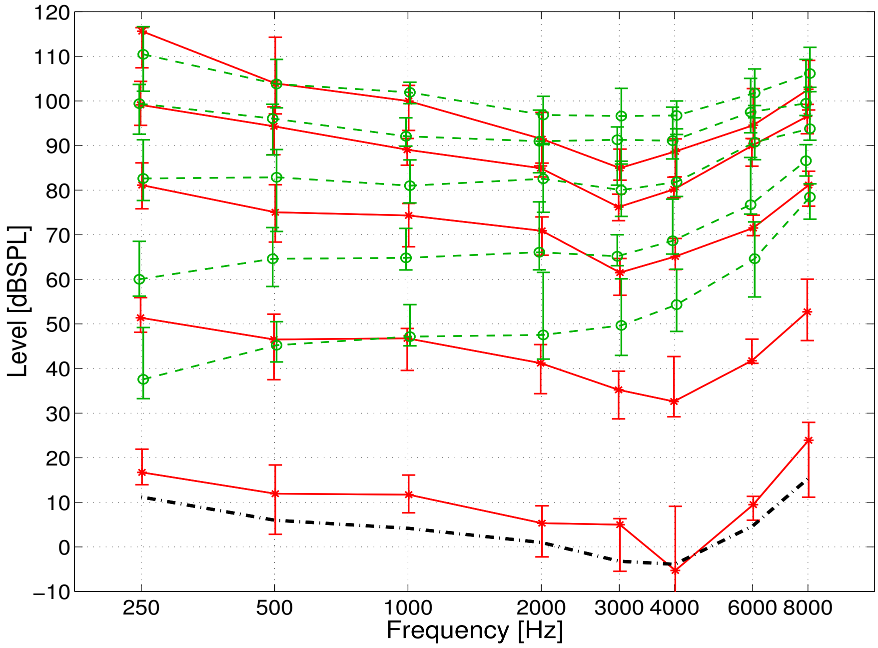

Figure 6.5: *Median equal–loudness level contours for normal–hearing (solid lines) and hearing–impaired listeners (dashed lines) calculated from loudness scaling data (ELLC–CU's) for narrowband noise stimuli at loudness categories 5, 15, 25, 35 and 45 CU. Error bars indicate the interindividual 25 % and 75 % percentiles. The dash–dotted line corresponds to the MAF according to ISO 226(E) [1987].*

6.4 Model Predictions

6.4.1 Individual Adjustment of the Models

The loudness model proposed by Moore *et al.* [1999b][7] (*MODEL–1*) allows for an individual adjustment of the overall hearing loss, the definition of frequency limits defining the 'dead regions' and the partitioning of the overall hearing loss into a component due to OHC damage and a component due to IHC damage. Whereas the overall hearing loss (HL_{TOTAL}) can be obtained from the subjects pure tone audiogram, the partitioning between OHC and IHC loss is restricted by Equation 6.3, so that either HL_{OHC} or HL_{IHC} is a free parameter of the model[8]. Since Moore and colleagues suggested no rigorous way for the fitting of HL_{OHC}

[7]For this study the loudness model proposed by Moore and colleagues was implemented as a MATLAB script replicating the DOS program loud.exe in version 3.2 (1999) that is downloadable from the website of the cambridge group (`http://hearing.psychol.cam.ac.uk/Demos/demos.html`). It differs from the previous version 3.0 (1996) in the way auditory filters broaden with level.

[8]In the publication by Moore *et al.* [1999b] these parameters were adapted in the following way: the overall hearing loss was determined by the subjects audiogram, HL_{OHC} (and thereby HL_{IHC}) was fitted

and the frequency limits of any 'dead region', we used the default settings for their model for each individual listener as follows[9]:

- The proportion accounting for OHC loss (HL_{OHC}) is set to 80% of the total hearing loss HL_{TOTAL}. In cases when this results into HL_{OHC} being larger than 55 dBHL for frequencies below or equal to 1 kHz or larger than 65 dBHL for frequencies above 1 kHz, HL_{OHC} is limited to 55 dBHL or 65 dBHL, respectively.
- The proportion accounting for IHC loss (HL_{IHC}) is always set in accordance with Equation 6.3.
- dead regions are disregarded.

Note, that the usage of this standard setup strongly reduces the number of free parameters of the model.

The individual adjustment of *MODEL–2* and *MODEL–3* allows for the setting of two frequency–dependent parameters: First, the overall hearing threshold is determined directly by the subjects audiogram. Second, the pre–factor β in Equations 6.1 and 6.2 is fitted to the individual loudness scaling data for narrowband stimuli using a least–squares method [Press *et al.*, 1992]. To do so, the loudness in sone as it is predicted by the models has to be transformed to the loudness in categorical units (CU). For this purpose the transformation by Blum [1999] and the transformation introduced in chapter 5 section 5.2.3 have been used for *MODEL–2* and *MODEL–3*, respectively. The frequency dependence of β is modelled as a polynomial on the ERB–scale:

$$\beta = \beta(f_{ERB}) = a \cdot (f_{ERB})^2 + b \cdot f_{ERB} + c \tag{6.5}$$

The free parameters a, b and c were fitted to the data for narrowband stimuli so that the squared distance between data points and model predictions was minimized using a simplex method [Press *et al.*, 1992]. As initial parameter settings, $c = 1$ and $c = 2$ was used for the group of normal–hearing and the group of hearing–impaired subjects, respectively, whereas a and b were set to zero in all cases.

Figure 6.6 shows the interindividual median values of the β–factor as they result from this fitting process. Upper and lower panels show the respective data for *MODEL–2* and *MODEL–3*, respectively. The individually fitted coefficients of the β–polynomial are summarized in Appendix D.2. Generally a value of $\beta = 1$ across frequency is expected for the group of normal–hearing subjects (left panels), because this corresponds to the standard setting of the models that should be directly applicable to normal loudness perception. As can be seen from Figure 6.6 this is exactly achieved for *MODEL–3* when the degree of the polynomial is restricted to zero (stars in Figure 6.6, lower left panel). This indicates, that the loudness scaling data for narrowband stimuli with different center frequencies are on average very well predicted by the standard setting of *MODEL–3*. Note, that the standard settings of *MODEL–*

'by hand' to give best predictions for the respective individual loudness matching data of pure tones and frequency limits of any 'dead region' were adjusted so that the model predicted with reasonable accuracy the absolute threshold at frequencies within any 'dead region'.

[9] In a more recent work, Moore *et al.* described a method for the diagnosis of dead regions and the definition of their frequency limits [Moore *et al.*, 2000]. This method could not be included in this work, because it was not available when our data was collected. However, from the relatively smooth audiograms of our subjects it seems likely that no dead regions have to be assumed in this case, but as can be seen from the results in Moore *et al.* [2000], this assumption is not clear cut.

3 have been achieved by a completely different set of data, indicating that *MODEL–3* yields a robust prediction for this kind of experiment. For *MODEL–2* (Figure 6.6, upper left panel) the fitting of β generally results in values smaller than one. Note, that decreasing β leads to a stronger compression of the excitation level, so that loudness increases less with increasing stimulus level (shallower loudness function) resulting in a larger dynamic range. On the contrary, increasing β leads to steeper loudness functions and a smaller dynamic range. Therefore, it can be concluded that the standard setting of *MODEL–2* in general predicts too steep loudness functions for normal–hearing subjects and that the fitting of β aims at compensating for this.

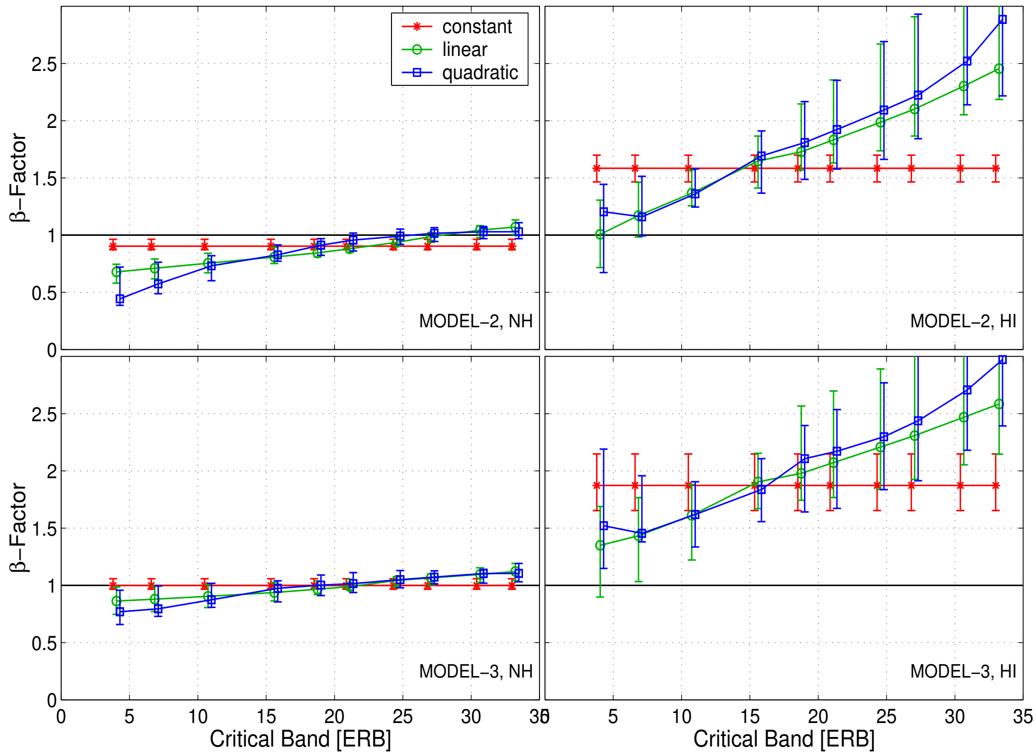

Figure 6.6: *Interindividual median values of the fitted β–factors as a function of frequency in units of ERB for the group of normal–hearing subjects (left panels) and the group of hearing–impaired (right panels). Polynomials of degrees zero (constant), one (linear) and two (quadratic) are indicated by stars, circles and squares, respectively. Upper and lower panels show the β–factors for MODEL–2 and MODEL–3, respectively. Error bars show the interquartile range across subjects.*

However, when β is fitted by a polynomial of higher degree, it is found to be slightly smaller than one for frequencies below about 22 ERB (approximately 2 kHz) and slightly larger for frequencies above when *MODEL–3* is considered. This indicates that *MODEL–3* slightly underestimates the dynamic range of the normal–hearing subjects at low frequencies and vice versa slightly overestimates it at high frequencies. This can be explained by considering the ELLC–CU's of the normal–hearing subjects shown by the solid lines in Figure 6.5. There (as well as in Figure 5.4 for another group of normal–hearing subjects) it is found that the dynamic range of the normal–hearing subjects seems to slightly decrease with increasing frequencies, whereas *MODEL–3* approximately predicts a constant dynamic range. For *MODEL–2*, however, it can be seen from Figure 6.6 (upper left panel) that the frequency dependence of the β–values is more pronounced than for *MODEL–3*. This difference is due to the different corrections used in the two models, i.e., the ELC–correction in *MODEL–2* and the MAF–correction in *MODEL–3*. It can be concluded, that for the prediction of ELLC's the introduction of the MAF–correction is a step into the right direction. However, as it was already mentioned in chapter 5.2.2, the MAF–correction was introduced in *MODEL–3* as a *moderate* adaptation of the model to partially predict the higher ELLC's at low frequencies than standardized in ISO 226(E) [1987] as found in recent studies [refer to Betke, 1991; Fastl *et al.*, 1990; Gabriel *et al.*, 1994; Suzuki *et al.*, 1989; Watanabe and Møller, 1990; for a review see Gabriel, 1996; Reckhardt, 2000]. Note that these higher ELLC's are also in line with the results of the loudness scaling experiments presented in chapter 5 as well as with the current results. Possibly, any further adaptation of the model towards the new ELLC–data will resolve the small discrepancies found while fitting the β–polynomial of orders 2 and 3 for *MODEL–3*.

The right panels in Figure 6.6 show that the fitting of the β–polynomial for the group of hearing–impaired subjects results in β–values greater than one and that the fitted β–values increase with frequency. This can be explained by the subjects sloping high frequency hearing loss of cochlear origin with the consequence of a reduced dynamic range especially at high frequencies. It is observed for *MODEL–2* and *MODEL–3*, that the linear fitting of the β–polynomial in general leads to similar results as the quadratic fit. However, this might be due to the limited variability in the type of hearing losses investigated here. Hence, we retain the assumption made by Launer [1995] that a quadratic polynomial should provide enough flexibility for the frequency dependent adjustment of β.

In general it is questionable whether a fit of β to loudness scaling data of narrowband stimuli is necessary, or whether β can be estimated directly from the subject's audiogram. Table 6.5 shows β values over frequency as they are fitted within *MODEL–3* (third row in Table 6.5, same data as shown Figure 6.6, lower right panel, square symbols) together with β–values simply predicted from the mean audiogram taken across the hearing thresholds of the hearing–impaired subjects (bottom row in Table 6.5). These estimates were derived from the fraction between normal–hearing dynamic range (assumed to be 100 dB) and the individual listener's dynamic range. This is approximated by the difference between the individual hearing–impaired listener's uncomfortable loudness level (*UCL*, assumed to be 100 dBHL) and the mean audiogram ($HL(f)$):

$$\beta_{est}(f) = \frac{100\,dB}{100\,dB - HL(f)} \tag{6.6}$$

From Table 6.5 it can be seen that the difference between the fitted β–values (β_{fit}) and this simple prediction is very small. This shows that for a very common type of hearing loss as it is represented by the mean audiogram of our subjects, β is easily predictable from the audiogram.

To investigate whether this simple transformation is able to predict the value of β when individual data is considered, Figure 6.7 shows the distribution of the ratios between β_{est} and β_{fit} calculated for each subject and each frequency measured in the audiogram.

Freq. [Hz]	125	250	500	1000	1500	2000	3000	4000	6000	8000
Freq. [ERB]	4.0	6.8	10.7	15.6	18.8	21.1	24.6	27.1	30.7	33.2
β_{fit}	1.522	1.455	1.620	1.838	2.107	2.172	2.298	2.437	2.707	2.972
β_{est}	1.333	1.389	1.587	1.786	1.923	2.128	2.174	2.381	2.632	2.500

Table 6.5: *β–factors for several frequencies estimated from the mean hearing thresholds taken across the group of hearing–impaired subjects (row β_{fit}) and the respective fitted β–factors for MODEL–3 (β_{fit}, fit of a quadratic polynomial, same as squares in lower right panel of Figure 6.6).*

Figure 6.7 shows that the center of the distribution is close to one, supporting the finding that β, and thus the slope of the loudness function, can be on average estimated by Equation 6.6. On the other hand, Figure 6.7 shows that the distribution is very broad, so that Equation 6.6 for some subjects may predict β–values that are one half or twice of the β–values fitted to the data. The predicted dynamic range of the subject would be approximately twice or half of the dynamic range observed in the experiment in this case. Therefore, Equation 6.6 is not generally applicable to predict slopes of the loudness functions on an individual basis from the audiogram. However, Equation 6.6 works quite well on average and it might be used to calculate appropriate initial estimates for β or when no loudness scaling data for narrowband stimuli is available.

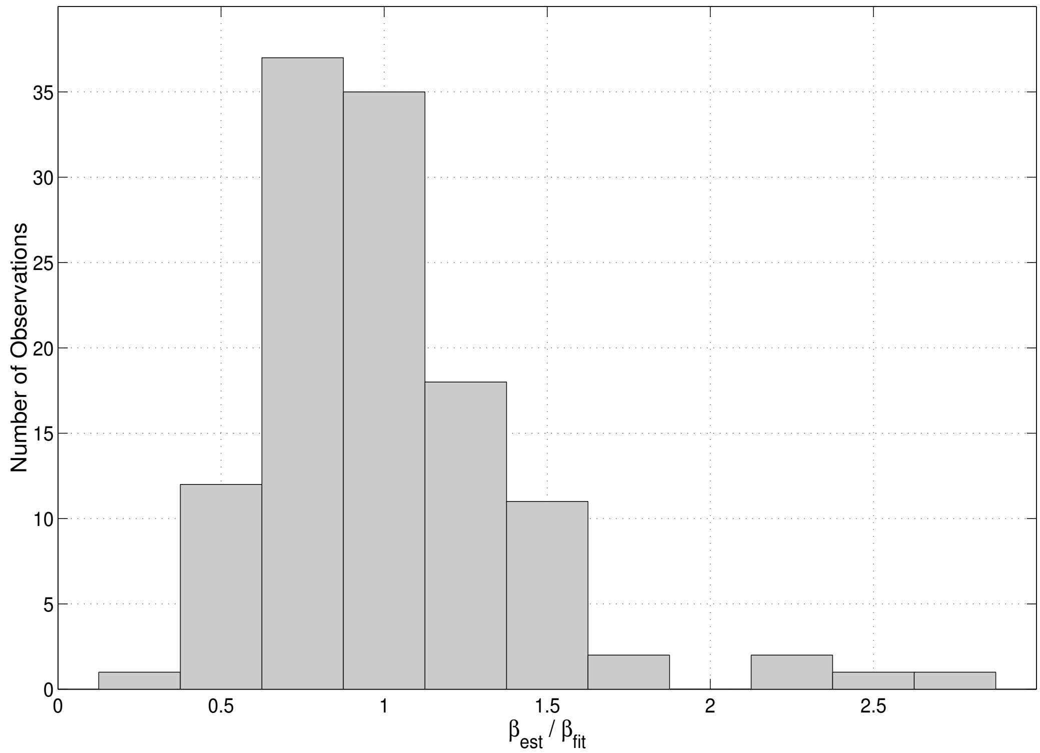

Figure 6.7: *Histogram of the ratios between estimated and fitted β–factors calculated for each subjects and each frequency measured in the audiogram.*

6.4.2 Normal–hearing Subjects

Figure 6.8 shows predictions of normal–hearing loudness functions for the narrowband stimuli in categorical units (CU) as they are predicted by the three models (dotted, dash–dotted and dashed lines for *MODEL–1*, *MODEL–2* and *MODEL–3*, respectively) together with the average experimental data already shown in Figure 6.1 (straight lines, shaded area). The predictions of *MODEL–2* and *MODEL–3* were calculated using the models standard settings for normal–hearing subjects, i.e., absolute threshold was set to zero dBHL and the exponent factor β was set to one. Here and in the following the sone–CU transformation proposed for *MODEL–3* (chapter 5, Equation 5.6) was used to transform the model predictions of *MODEL–1* from the sone–scale into categorical units. This is appropriate because in the case of modeling normal–hearing loudness perception, *MODEL–1* differs from *MODEL–3* only in that it uses the ELC–correction instead of the MAF–correction, which only differ for frequencies below 1 kHz.

In fact it can be concluded from Figure 6.8 that *MODEL–1* and *MODEL–3* give approximately the same predictions for frequencies at 1 kHz and above. For frequencies below 1 kHz and at higher levels the predictions of *MODEL–1* are closer to the predictions made by *MODEL–2*, which is due to the ELC–correction used in both models. At threshold and at low levels *MODEL–1* again predicts similar loudness functions as *MODEL–3*. This is because both models account for normal–hearing absolute threshold in the same way, i.e., the subtractive term in Equations 6.4 and 6.2 provides quantities that are taken relative to the correction applied in the first stage of the model. Compared to the experimental data (straight lines in Figure 6.8), Figure 6.8 shows that *MODEL–3* gives best overall predictions of the narrowband data measured in the group of normal–hearing subjects. As expected from the discussion on the fitting of the β–polynomial in the previous section, *MODEL–3* only slightly underestimates the dynamic range of the normal–hearing subjects at low frequencies and slightly overestimates it at higher frequencies, whereas this effect is more pronounced in the predictions of *MODEL–2*. In general *MODEL–2* also shows stronger deviations from the experimental data at low levels, which might be due to the inaccurate prediction of normal–hearing threshold and due to the transformation between the sone and the CU–scale that was proposed by Blum [1999] for this model (see section 5.2.3).

Figure 6.9 shows the predictions of the three models together with the experimental data for the broadband stimuli. Note that the parameters of the models were not fitted to these data, but are based on narrowband data. In general, the models are capable to predict that the upwardly concave loudness functions for narrowband stimuli become more linear with increasing stimulus bandwidth. In addition spectral loudness summation is predicted by the three models, which can be quantified by taking the level difference between a narrowband and a broadband signal at the same loudness. However, some deviations from the experimental data are found. None of the models is able to accurately predict the experimental data at high levels and for bandwidths above 5 Bark. However, as mentioned above, there is some uncertainty in the measured data because 'negative loudness summation' is observed in the data for the highest stimulus levels (see Section 6.3.5) and this effect is in principle not included in the models.

Interestingly, all models show comparatively good agreement with the subjects responses near threshold (shaded area in Figure 6.9), i.e., the model predictions do not show the lim-

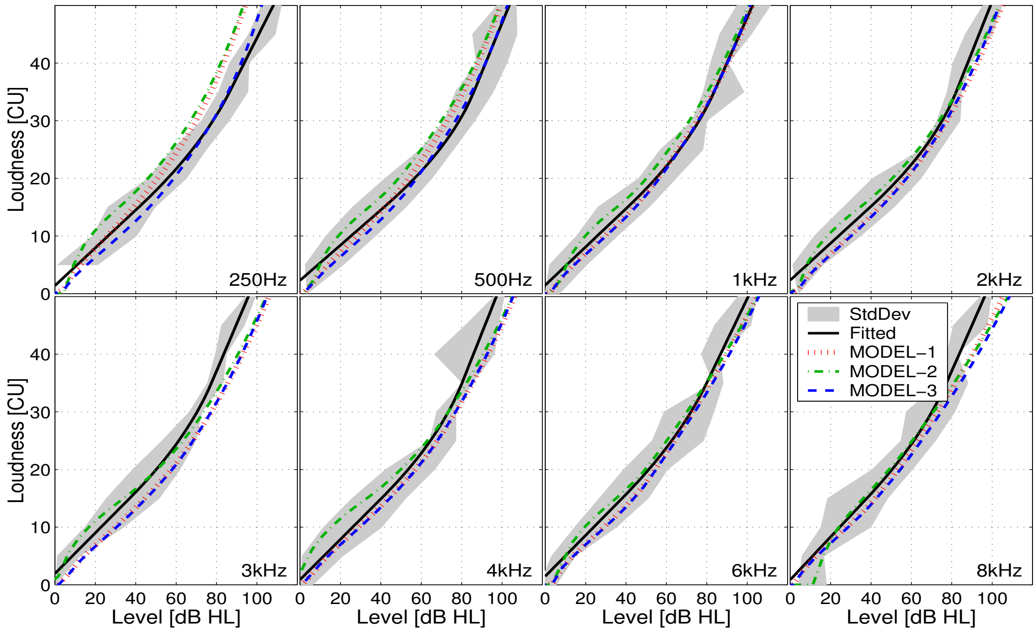

Figure 6.8: *Loudness functions predicted by MODEL–1 (dotted lines), MODEL–2 (dash–dotted lines) and MODEL–3 (dashed lines) for the group of normal–impaired subjects and third–octave narrowband stimuli. Predictions of MODEL–2 and MODEL–3 are calculated for the standard setting (without fitting of the β–factor). As in Figure 6.1, the shaded areas show plus minus one standard deviation of the mean across the subjects responses at fixed categorical loudness and the solid lines are fitted to mean data.*

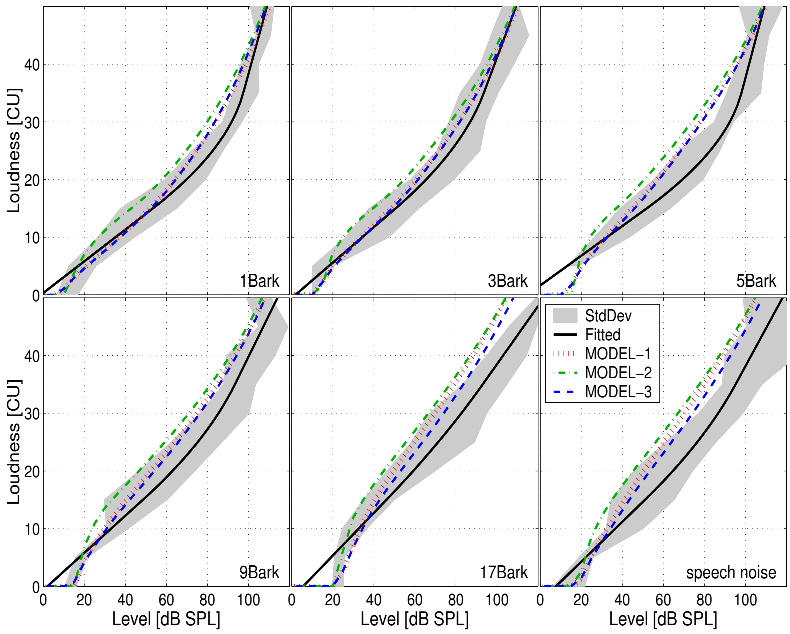

Figure 6.9: *Same as Figure 6.8 but for broadband stimuli, i.e., 1, 3, 5, 9 and 17 Bark wide uniform exciting noise and speech simulating noise (left to right and top to bottom, respectively).*

itations found for the model function fitted to the subject's responses (straight line in Figure 6.9). This is expected for *MODEL–1* and *MODEL–3*, because these models carefully model loudness near threshold. However, this finding is not obvious for *MODEL–2* that did not accurately predict (refer to Figure 5.2, chapter 5). These deviations are not revealed here because the broadband stimuli are spectrally concentrated in a frequency range where the model predicts absolute threshold quite correctly.

In general it can be concluded from Figure 6.9 that the experimental data is better predicted by *MODEL–1* and *MODEL–3* compared to *MODEL–2*. As expected for the prediction of loudness perception in normal–hearing listeners, the difference in the predictions of *MODEL–1* and *MODEL–3* are very small with a small advantage for *MODEL–3*.

6.4.3 Hearing–impaired Subjects

Figure 6.10 shows the model predictions for the group of hearing–impaired subjects and for the narrowband stimuli. In contrast to the predictions for the normal–hearing subjects of *MODEL–2* and *MODEL–3* shown in the previous sections, the predictions made for the hearing–impaired subjects used individual fitted β–factors (quadratic polynomial). As can be seen from Figure 6.10, all models are able to predict increased hearing threshold and increased slope of the loudness functions quite reasonably. The predictions of *MODEL–2* and *MODEL–3* are very similar and do represent the experimental data with good accuracy. Only at high frequencies, i.e., for high hearing thresholds and at high levels, *MODEL–1* tends to predict a negative curvature of the loudness functions and thereby predicts too shallow loudness functions as compared to the experimental data. This might occur for two reasons. First, in *MODEL–1* the bandwidth of the auditory filters depends on hearing loss in a way that the bandwidth is increased at low input levels, but changes less with level compared to the level dependence of the auditory filter bandwidth assumed for normal–hearing subjects. If the assumed increase of the filter bandwidth with level is too small, this leads to a too small increase of the spread of excitation for narrowband stimuli and therefore loudness would increase less with increasing stimulus level. Second, steeper loudness functions are modeled by increasing the subtractive term in Equation 6.4 in *MODEL–1*. As can be seen from Equation 6.4 and as it is already mentioned in Launer [1995, Figure 7.2, chapter 7], this affects mainly the slope of the loudness function at low levels, whereas the subtractive term in Equation 6.4 is more or less negligible at high levels. Note, however, that the adjustment of *MODEL–1* only concerns the audiogram and that standard settings were used for the other free parameters of the model. Better predictions could presumably be achieved by fitting these parameters as well.

Figure 6.11 shows the predictions of the three models together with the experimental data for the broadband stimuli. *MODEL–3* gives slightly better predictions than *MODEL–2*, whereas *MODEL–1* again predicts too shallow slopes, so that the residual dynamic range is overestimated. In addition *MODEL–1* seems to overestimate the loudness for the stimuli with the larger bandwidth and therefore predicts a higher amount of loudness summation than it is observed in the experimental data. However, the differences in the predictions of the three models are again not very large.

Figures 6.12 and 6.13 show individual model predictions for subjects hhr (subject hh, right ear) and bul (subject bu, left ear). These subjects were selected because they have

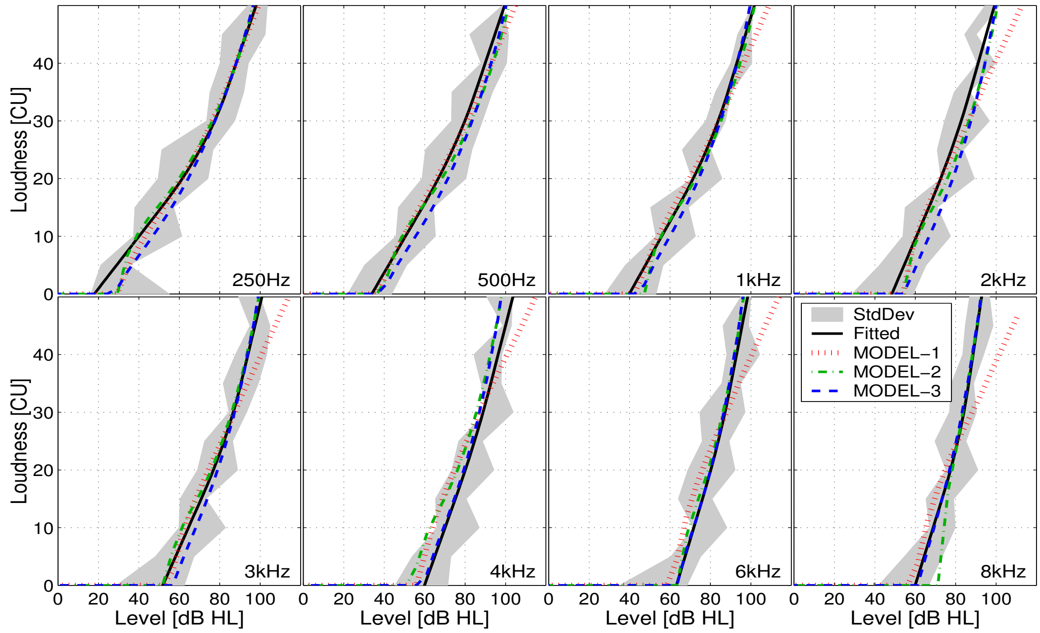

Figure 6.10: *Same as Figure 6.8 for the group of hearing–impaired subjects and with a fitting of the β–factor for MODEL–2 and MODEL–3.*

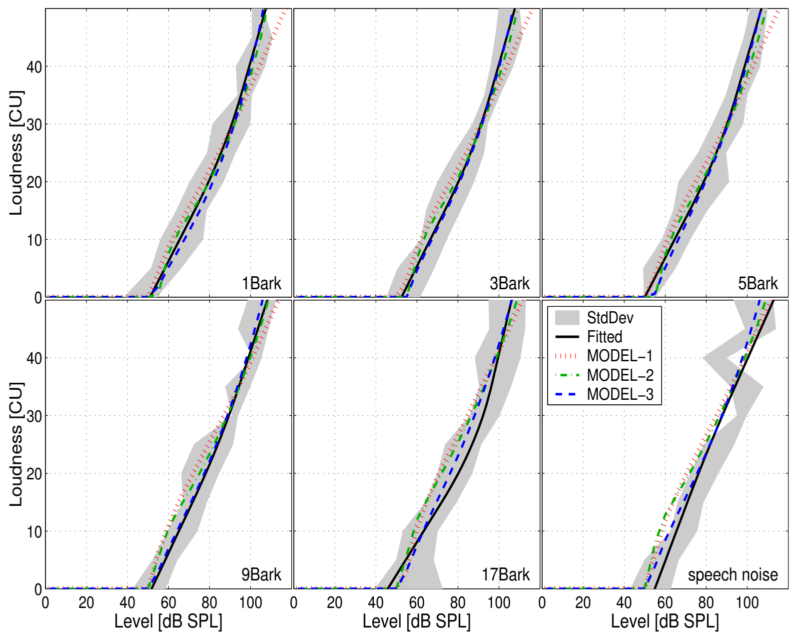

Figure 6.11: *Same as Figure 6.9 for the group of hearing–impaired subjects and with fitting of the β–factor for MODEL–2 and MODEL–3.*

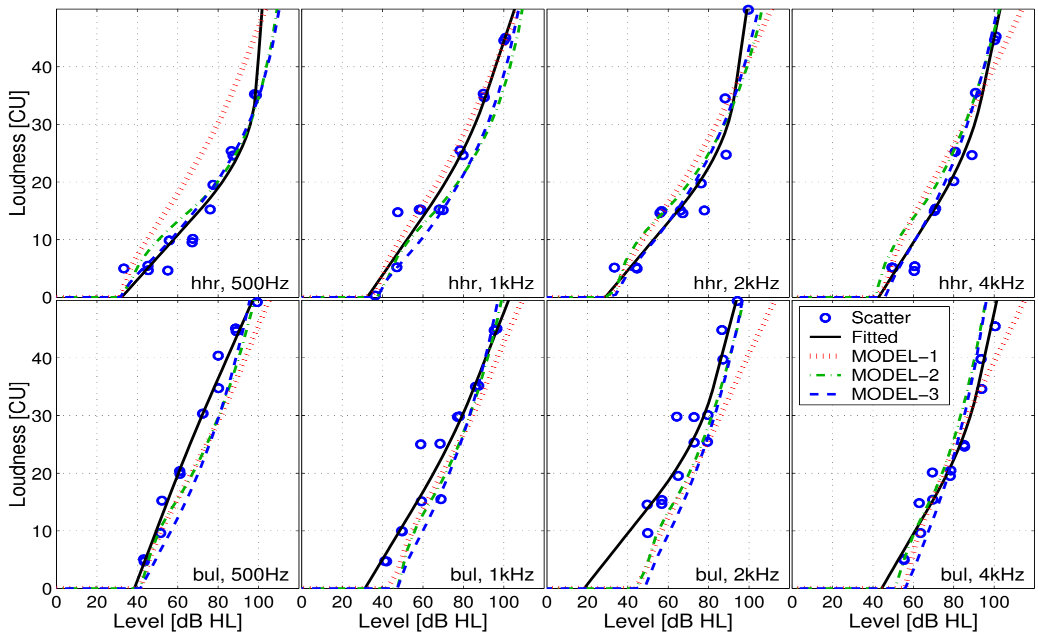

Figure 6.12*: Loudness functions for narrowband noises centered at 0.5, 1, 2, and 4 kHz (left to right panels, respectively) predicted by MODEL–1 (dotted lines), MODEL–2 (dash–dotted lines) and MODEL–3 (dashed lines) for the two hearing–impaired subjects hhr and bul (upper and lower panels, respectively). Predictions of MODEL–2 and MODEL–3 are calculated with fitting of the β–factor. Solid lines show the model functions fitted to the loudness scaling data (circles).*

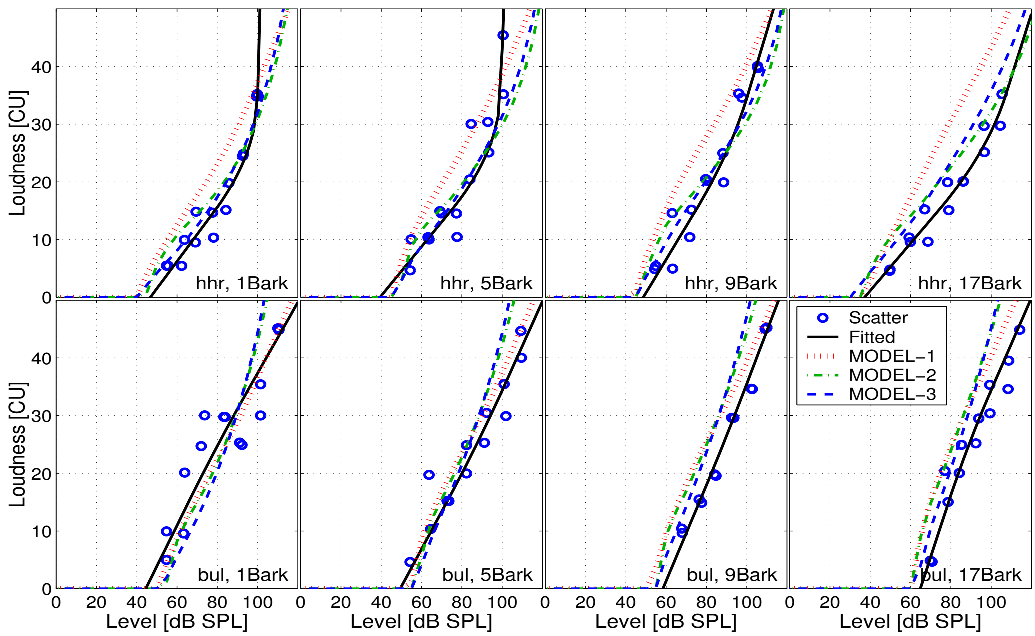

Figure 6.13*: Same as Figure 6.12 but for broadband stimuli, i.e., 1, 5, 9 and 17 Bark wide uniform exciting noise (left to right panels, respectively).*

similar audiograms for the frequencies shown in Figure 6.12 but showed the largest difference with respect to the recruitment phenomenon. This can be seen from the fitted β–values when β is fitted by a constant value in *MODEL–2* and *MODEL–3* ($\beta(\textit{MODEL–2}, hhr) = 1.22$, $\beta(\textit{MODEL–3}, hhr) = 1.39$, $\beta(\textit{MODEL–2}, bul) = 2.29$, $\beta(\textit{MODEL–3}, bul) = 2.52$, see Table D.2). As expected from the larger β–values fitted for subject bul, a smaller dynamic range is predicted by *MODEL–2* and *MODEL–3* than for subject hhr both for the narrowband stimuli (Figure 6.12) and the broadband stimuli (Figure 6.13). On the other hand *MODEL–1* predicts only slightly steeper loudness functions for subject bul. This is because only the subjects audiogram was considered in the individual adjustment of *MODEL–1*.

6.5 Discussion and Conclusions

The predictions of three recent loudness models for stationary sounds accounting for normal and impaired loudness perception were compared to loudness data obtained in loudness scaling experiments of narrowband and broadband noise stimuli with normal–hearing and hearing–impaired subjects. The models are all based on the four processing stages proposed by [Zwicker, 1958; Zwicker and Fastl, 1990] and account for hearing impairment by a component that relates to a raised hearing threshold and a component that relates to a loss of compression in the cochlea. However, they differ in several relevant aspects. The model proposed by Marzinzik *et al.* [1996b] (*MODEL–2*) and the model introduced in chapter 5 (*MODEL–3*) account for raised hearing threshold by linearly attenuating the excitation evoked by the input stimulus. They account for the loss of compression by increasing the compressive exponent in the formula relating excitation with specific loudness (Equations 6.1 and 6.2). In the model proposed by Moore *et al.* [1999b] (*MODEL–1*), on the other hand, hearing impairment is accounted for by partitioning the overall hearing loss into a component due to OHC damage and a component due to IHC damage. Whereas both, IHC and OHC loss, contribute to raised hearing threshold, the amount of hearing loss that attributes to OHC loss is used to increasing the subtractive term in the equation relating excitation with specific loudness OHC loss (Equation 6.4), which steepens the predicted loudness function mainly at low levels.

MODEL–3 differs from *MODEL–2* with respect to a better prediction of normal–hearing threshold and adapts the model to equal–loudness level contours (ELLC's) recently suggested by the Committee Draft ISO/CD 226 [2000]. In addition, *MODEL–3* uses the empirically derived formula for the transformation between the sone–scale and the categorical loudness-scale (CU–scale) introduced in chapter 5 (cubic polynomial fit to narrowband loudness scaling data), whereas *MODEL–2* is tested with the transformation proposed by Blum [1999]. The extensions made in *MODEL–3* have been discussed in detail in chapter 5.

The predictions of *MODEL–2* and *MODEL–3* for the group of normal–hearing subjects show that *MODEL–3* better predicts normal–hearing threshold, loudness perception at moderate input levels and at high input levels and low frequencies than *MODEL–2*. In comparison to *MODEL–1*, the extensions introduced in *MODEL–3* have the consequence that both models predict normal–hearing loudness perception in the same way, except for stimuli having significant low frequency components, i.e., in the frequency range where the modifications made to *MODEL–3* with respect to the new ELLC's affect the predictions. Thus, the predictions of the two models for the narrowband stimuli above 1 kHz are very similar and are found to be of good accuracy, indicating that the proposed transformation between sone and CU is applicable to both models. At lower frequencies only *MODEL–3* is capable to predict the loudness functions accurately, because it is the only model that accounts for new ELLC's. However, the analysis of the fitted β–factor for the normal–hearing subjects showed, that the predictions of *MODEL–3* would further improve when still higher ELLC's at low frequencies (larger dynamic range) and slightly lower ELLC's at high frequencies (smaller dynamic range) would be assumed. This can be explained by the fact that the modification made to *MODEL–3* with respect to new ELLC's (use of the MAF–correction instead of the ELLC-correction) is only a moderate approximation towards new ELLC's as they are suggested by our data (see chapter 5) as well as by several other studies [Betke, 1991; Fastl *et al.*, 1990; Gabriel *et al.*, 1994; Suzuki *et al.*, 1989; Watanabe and Møller, 1990].

The data obtained from the loudness scaling experiments of the broadband stimuli for the group of normal–hearing subjects showed the effect of loudness summation at medium levels (up to about 90 dBSPL), which was in principle modeled by all models. However, at higher levels our data show that loudness summation rapidly decreases and is even found to be negative at the highest presentation levels. It might be suggested that the subject's loudness ratings are influenced by other criteria such as annoyance when high levels are presented and that stimuli with a larger bandwidth might be better tolerated when judged in terms of annoyance than narrowband stimuli. However, due to the principles of the loudness models, negative loudness summation can not be predicted by these models. Although it is difficult to obtain reliable experimental data for these high loudness levels, further experimental evidence on this effect is therefore desirable.

While the predictions of *MODEL–2* and *MODEL–3* show good agreement with the narrowband loudness scaling data measured in the hearing–impaired subjects, *MODEL–1* only yields a good agreement for subjects with a hearing threshold below about 40 dBHL. The predictions of *MODEL–1* and *MODEL–2* for the broadband loudness scaling data measured in the hearing–impaired subjects tend to overestimate the effect of loudness summation, whereas *MODEL–3* predicts the data reasonably well. In general *MODEL–1* predicts too shallow slopes of the loudness functions at high levels combined with a hearing threshold above about 40 dBHL. This might be explained by the fact that the subtractive term in Equation 6.4 becomes negligible when the excitation of the stimulus raises or by the way the broadening of auditory filters is restricted due to OHC loss. Although the predictions of *MODEL–1* where calculated using the standard settings of the model (i.e., partitioning of the total hearing loss into 80% and 20% relating to OHC and IHC loss, respectively, no definition of dead regions) it can be assumed that steeper loudness functions at high levels can not be achieved by a more judicious choice of the fitted parameters without deteriorating the predictions at low levels.

One potential drawback of all three loudness models discussed in this study is that the fitting of the models free parameters require more data on the individual hearing loss than provided by the audiogram. This is due to the fact that several studies have shown that the elevated hearing threshold and the reduced dynamic range are more or less independent from one another [Hohmann, 1993; Kießling, 1995; Kießling *et al.*, 1994; Launer, 1995; Launer *et al.*, 1996; Moore *et al.*, 1999d]. This was also found in this study by investigating the distribution of the ratios between the estimated and fitted β–values. In *MODEL–2* and *MODEL–3* this is accounted for by a rigorous fit of the parameters influencing β (which accounts for reduced dynamic range) to the loudness scaling data obtained from loudness scaling of narrowband stimuli. However, in cases where only the data from the audiogram is available for the adjustment of *MODEL–3*, we suggest that Equation 6.6 gives an appropriate estimate of the parameter accounting for reduced dynamic range, i.e., the value of β. Note, however, that substantial individual deviation from this average value may occur, especially for cases with a hearing loss quite dissimilar to those considered here.

In general it was found for the loudness scaling data employed in this study that *MODEL–1* better predicted loudness perception for the group of normal–hearing subjects when compared to *MODEL–2*, whereas *MODEL–2* better predicted loudness perception for the group of hearing–impaired subjects. Overall, *MODEL–3* yields better predictions than *MODEL–1* and *MODEL–2*. Hence, this model should be employed in future studies, especially when low–frequency data, categorical loudness scaling data or data from hearing–impaired listeners are considered.

Chapter 7

Summary and Outlook

The general aim of this thesis was to develop measurement methods and models of loudness perception appropriate for rehabilitative audiology, i.e., the application of these methods to the restoration of normal loudness perception in hearing–impaired listeners. This goal has been partially achieved by

- studying the influence of loudness perception and its measurement in a field–test with hearing–impaired listeners and different compression schemes (chapter 2). Although several recommendations for the setting of the dynamic compression parameters could be derived from the experiments, some deviations from the fitting rational of establishing normal loudness perception in hearing–impaired (especially for the broadband stimuli) called for a further methodological and theoretical investigation.
- investigating the different methods for assessing loudness in hearing–impaired listeners with respect to its usability in practical applications (chapter 3). Although both loudness matching and categorical loudness scaling methods are shown to have their respective advantages and problems, loudness scaling seems to be more appropriate in practical applications with hearing–impaired listeners (i.e., rehabilitative audiology) since it provides information on loudness perception for the listeners' complete dynamic range in a time efficient way.
- the development and evaluation of a new loudness model (Chapter 5 and 6) that is capable of predicting the perceived loudness for hearing–impaired listeners not only on a ratio scale (sone–scale) but also in terms of the categorical loudness rating (CU–scale, according to the "Oldenburger Hörflächenskalierung"). This "Oldenburg loudness model" is a modification of Zwicker's loudness model and has the advantage of accounting for recent data on equal–loudness level contours and of modeling loudness perception in hearing–impaired listeners with a two–component approach. The first component accounts for a loss of sensitivity and the second component ("compression loss") is characterized by a decrease of the compressive exponent in the formula relating the excitation pattern with specific loudness. The adaptation of the model to account for recent data on equal–loudness level contours also showed that the model yields a better prediction of this data obtained from loudness scaling experiments (chapter 5). However, further optimizations of the model may have to be considered when the recent data on equal–loudness level contours are standardized by the International Organization for Standardization.

Although this work is a solid basis for applying loudness measurements and prediction of loudness sensation in a variety of practical applications, it also may serve as the basis of future work. In particular, the "Oldenburg loudness model" presented here has only been evaluated with a limited number of hearing–impaired subjects showing only a limited variation across different hearing losses. Hence, a larger variability of experimental parameters (such as, e.g., type of hearing loss, type of stimuli, type of experiments) should be employed to test the proposed loudness model for its general applicability. Specifically, the temporal properties of loudness and loudness models have not been considered here, which should be a major issue in future work.

Another line of research following from the current thesis will be to test the performance of hearing–aid algorithms with respect to loudness perception and its restoration, both in an objective way (with the model proposed here) and a subjective way (loudness perception in the aided and unaided condition). From the comparison between both measures, one could gain more insights into the applicability of the current model to different situations with and without hearing aids and into the actual performance of different hearing–aid algorithms. Thus, the loudness model proposed here might be transformed into a valuable tool for the developers of hearing aids to asses the relative benefit and shortcomings of their respective algorithms.

Yet another possible application would be to incorporate the loudness model described here into the actual signal processing in a hearing aid. It is conceivable that an "intelligent" hearing aid compares the loudness impression that a normal listener would receive in the respective acoustical situation with that of the individual hearing–impaired hearing–aid user. From this, it would derive an appropriate (frequency–dependent) amplification based on the assumption that both loudness impressions should match as closely as possible. Such an algorithm would either has to 'invert' the loudness model for hearing–impaired listeners or would require empirical rules and approximations to such an inversion.

Taken together, the current thesis provides a step into the general aim of rehabilitative audiology, i.e, to provide the individual hearing–impaired subject with devices and methods to overcome her or his hearing handicap.

Appendix A

Field test of 3–Channel Dynamic Compression schemes

A.1 The Wearable Digital Hearing Device

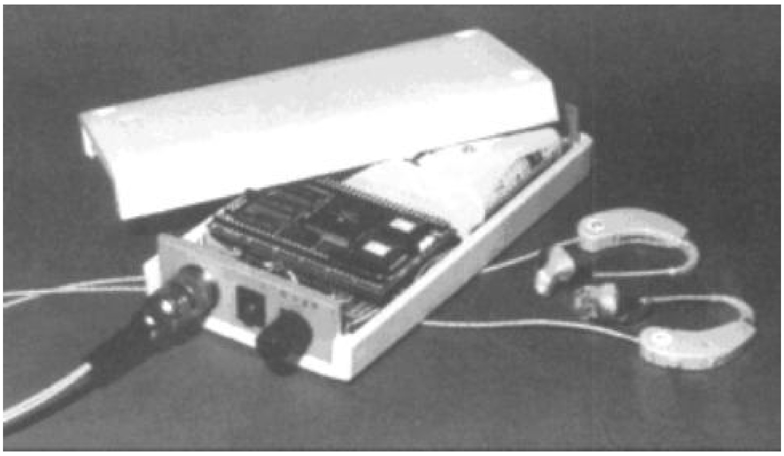

Figure A.1: *The wearable hearing–aid device (DASi–2).*

The wearable device (called DASi–2) was developed by Raß [1996]. It is based on a Motorola DSP56L002 signal processor (40 MHz clock frequency, 20 MIPS computational speed) embedded in a complete stand–alone system, the miniKit56002 (see Figure A.2). This credit–card sized module includes the required data and program memory and a serial interface for connecting the device to a PC. A single–chip Codec (AD and DA converter) acts as an interface to the analogue In–The–Ear (*ITE*) hearing–aids (Siemens Cosmea M). The Codec provides 16 bit stereo samples at 18.9 kHz. The digital hearing–aid prototype is powered from a 6 Volt NiCd accumulator package providing a capacity of 3.4 Ah. This enables an operation time of about 10 hours without charging the accumulators. About 50% of the unit's volume and weight are due to the accumulator package (total volume 190 mm, 100 mm, 40 mm, total weight approx. 1 kg, see Figure A.1).

As a comfortable user interface for fitting purposes, Rass has written the PC–program

'*HGTools*' (platform: Microsoft Windows). With this software it is possible to interactively adjust the parameters which are instantaneously transmitted to the signal processing unit. Thus, the effect of new settings can immediately be heard. Having set all parameters, these can be stored permanently in a non–volatile flash memory, which can hold up to four fitted algorithms for different acoustic situations. The user can select the program by a switch mounted on the digital device. Another switch allows to adjust the volume of the device in 16 steps with step–size 1.5 or 3 dB.

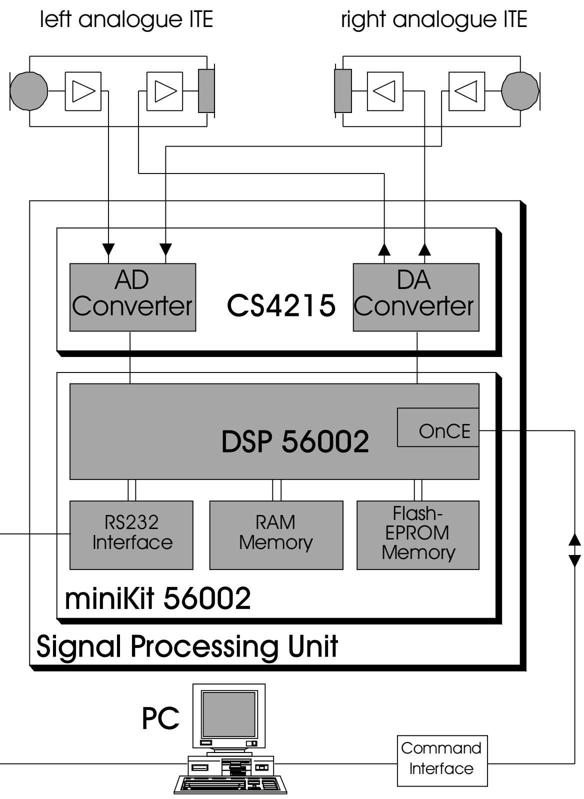

Figure A.2: *Block diagram of the DASi–2. Algorithms and parameters are downloaded by a PC. Two ITE hearing–aids are used for the in- and output of the acoustical signal.*

A.2 FFT Signal Processing Framework

Signal processing in the frequency domain can be efficiently realised using the *Overlap–Add FFT* processing scheme [Allen, 1977]. A block of input samples is multiplied with a window function (e.g., Hanning window). The sequence of input blocks overlap each other (e.g., by half of the block length). After adding zeros, the data block is transformed to the frequency domain with a *FFT* (fast fourier transform). The input spectra are multiplied with the transfer function, which may vary from block to block. To reconstruct the corresponding time signal, an inverse FFT transform is applied. The resulting sample blocks are again multiplied with a window function to reduce time domain aliasing effects. Finally the data blocks are added with an appropriate overlap to receive the processed output signals.

Such an algorithm is realised in the Overlap–Add module of the DASi–2 Kernel (Figure A.3). The implementation transforms the stereo input signal blocks with a single complex FFT into the frequency domain. The spectra of left and right channel are written to specific memory locations which can be accessed by the user's code. The programmer only has to write a subroutine performing the spectral modification. The processed spectra of left and right channel are transfered back to the Overlap–Add module which performs the inverse FFT, applies the optional output window and adds up the processed blocks of samples with the correct overlap. Parts of the algorithm which do not work frequency specific can also be realised in the time domain. This is useful, for example, to implement a peak clipping before the output samples are transferred to the *DAC* (digital to analog converter).

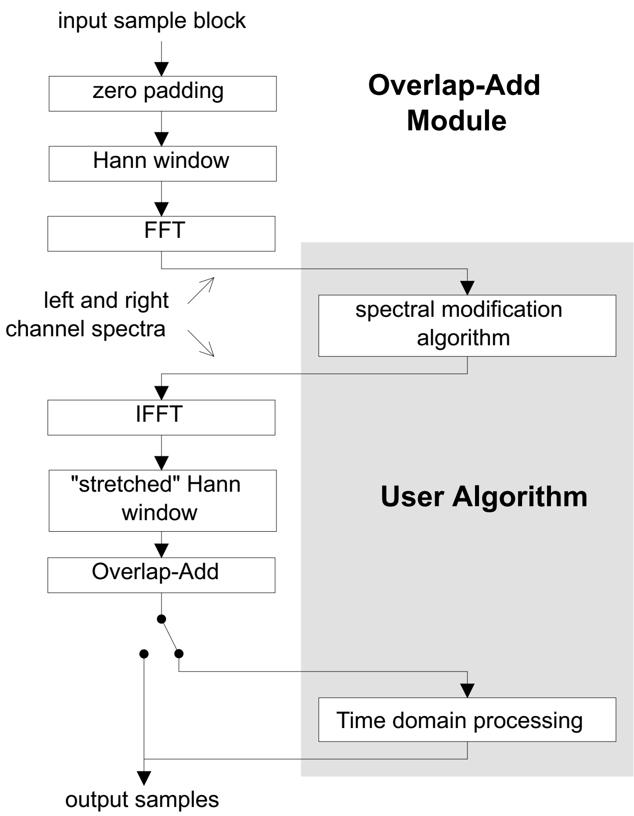

Figure A.3: *Block diagram of the Overlap–Add framework.*

However, the algorithms tested in chapter 2 were implemented in the frequency domain, so this feature was not used. The input signal was sampled with 18.9 kHz. Blocks of length 180 samples (duration 9.5 msec) where windowed by a Hanning window, padded with zeros to length 256 samples and a 256–point FFT was applied. The overlap of successive frames was set to 50% (90 samples). Three channel dynamic compression was applied to the input spectra. The output spectra was transformed into the time domain (*IFFT*, inverse FFT) and reconstructed by Overlap–Add. No windowing on the output was applied.

A.3 Instructions and Questionnaires

Versuchsablauf

Während dieses Experimentes werden Sie über den Lautsprecher folgende Signale hören:

- Musik
- Sprache in einer Cafeteria
- Sprache

Die Signale werden Ihnen bei drei verschiedenen Lautstärken dargeboten, die von Normalhörenden als:

- leise
- mittellaut
- laut

empfunden werden.

Jedes Signal möchten Sie sich bitte in zwei Versionen anhören. Sie können an dem Hörgerät zwischen zwei Programmen wählen. Wir bitten Sie, die zwei Programme bezüglich des Kriteriums "Genereller Eindruck" zu vergleichen. Der generelle Eindruck umfaßt den Klang, die Ausgewogenheit der Töne, bei Darbietung von Sprache auch die Sprachverständlichkeit und bei Musik z.B. die Natürlichkeit, die Klarheit usw.

Bitte beantworten Sie die folgende Frage:

Welche Version hat den deutlich oder etwas besseren "generellen Eindruck" oder sind beide Programme gleich ?

Vergleichen Sie bitte für jedes Signal die beiden Programme so lange, bis Sie sich ein Urteil gebildet haben. Teilen Sie Ihre Antwort bitte der Versuchsleitung mit.

Wenn Sie Ihr Urteil abgegeben haben, wird Ihnen ein weiterer Vergleich dargeboten, bei dem Sie bitte genauso verfahren wie oben beschrieben.

Wir bedanken uns für Ihre Mitarbeit bei dieser Studie !

Figure A.4: *Instructions given to the subjects before the paired comparison experiment.*

Bitte beurteilen Sie die verschiedenen Programme des Hörgerätes

Wie empfinden Sie die Verständlichkeit von Sprache in ruhiger Umgebung ?

a:	☐ ungenügend	☐ mangelhaft	☐ befriedigend	☐ gut	☐ sehr gut
b:	☐ ungenügend	☐ mangelhaft	☐ befriedigend	☐ gut	☐ sehr gut
c:	☐ ungenügend	☐ mangelhaft	☐ befriedigend	☐ gut	☐ sehr gut
d:	☐ ungenügend	☐ mangelhaft	☐ befriedigend	☐ gut	☐ sehr gut

Bestes Programm: ☐ a ☐ b ☐ c ☐ d

Wie finden Sie die Lautstärke ?

a:	☐ viel zu leise	☐ zu leise	☐ richtig	☐ zu laut	☐ viel zu laut
b:	☐ viel zu leise	☐ zu leise	☐ richtig	☐ zu laut	☐ viel zu laut
c:	☐ viel zu leise	☐ zu leise	☐ richtig	☐ zu laut	☐ viel zu laut
d:	☐ viel zu leise	☐ zu leise	☐ richtig	☐ zu laut	☐ viel zu laut

Bestes Programm: ☐ a ☐ b ☐ c ☐ d

Wie natürlich empfinden Sie den Klang ?

a:	☐ sehr natürlich	☐ natürlich	☐ etwas unnatürlich	☐ unnatürlich	☐ sehr unnatürlich
b:	☐ sehr natürlich	☐ natürlich	☐ etwas unnatürlich	☐ unnatürlich	☐ sehr unnatürlich
c:	☐ sehr natürlich	☐ natürlich	☐ etwas unnatürlich	☐ unnatürlich	☐ sehr unnatürlich
d:	☐ sehr natürlich	☐ natürlich	☐ etwas unnatürlich	☐ unnatürlich	☐ sehr unnatürlich

Bestes Programm: ☐ a ☐ b ☐ c ☐ d

Wie klar empfinden Sie den Klang ?

a:	☐ sehr unklar	☐ unklar	☐ etwas unklar	☐ klar	☐ sehr klar
b:	☐ sehr unklar	☐ unklar	☐ etwas unklar	☐ klar	☐ sehr klar
c:	☐ sehr unklar	☐ unklar	☐ etwas unklar	☐ klar	☐ sehr klar
d:	☐ sehr unklar	☐ unklar	☐ etwas unklar	☐ klar	☐ sehr klar

Bestes Programm: ☐ a ☐ b ☐ c ☐ d

Wie ist die Sprachverständlichkeit bei einem Theaterstück oder Vortrag ?

a:	☐ ungenügend	☐ mangelhaft	☐ befriedigend	☐ gut	☐ sehr gut
b:	☐ ungenügend	☐ mangelhaft	☐ befriedigend	☐ gut	☐ sehr gut
c:	☐ ungenügend	☐ mangelhaft	☐ befriedigend	☐ gut	☐ sehr gut
d:	☐ ungenügend	☐ mangelhaft	☐ befriedigend	☐ gut	☐ sehr gut

Bestes Programm: ☐ a ☐ b ☐ c ☐ d

Verständlichkeit von Sprache in Störgeräuschen
(lautes Cafe, Unterhaltung an einer Bushaltestelle)

a:	☐ ungenügend	☐ mangelhaft	☐ befriedigend	☐ gut	☐ sehr gut
b:	☐ ungenügend	☐ mangelhaft	☐ befriedigend	☐ gut	☐ sehr gut
c:	☐ ungenügend	☐ mangelhaft	☐ befriedigend	☐ gut	☐ sehr gut
d:	☐ ungenügend	☐ mangelhaft	☐ befriedigend	☐ gut	☐ sehr gut

Bestes Programm: ☐ a ☐ b ☐ c ☐ d

Bitte beurteilen Sie die allgemeine Klangqualität

a:	☐ sehr gut	☐ gut	☐ befriedigend	☐ mangelhaft	☐ ungenügend
b:	☐ sehr gut	☐ gut	☐ befriedigend	☐ mangelhaft	☐ ungenügend
c:	☐ sehr gut	☐ gut	☐ befriedigend	☐ mangelhaft	☐ ungenügend
d:	☐ sehr gut	☐ gut	☐ befriedigend	☐ mangelhaft	☐ ungenügend

Bestes Programm: ☐ a ☐ b ☐ c ☐ d

Bitte beurteilen Sie die allgemeine Klangqualität von Musik

a:	☐ sehr gut	☐ gut	☐ befriedigend	☐ mangelhaft	☐ ungenügend
b:	☐ sehr gut	☐ gut	☐ befriedigend	☐ mangelhaft	☐ ungenügend
c:	☐ sehr gut	☐ gut	☐ befriedigend	☐ mangelhaft	☐ ungenügend
d:	☐ sehr gut	☐ gut	☐ befriedigend	☐ mangelhaft	☐ ungenügend

Bestes Programm: ☐ a ☐ b ☐ c ☐ d

Figure A.5: *Questionaire filled out by the subjects during the field test.*

A.4 Sentence Test Data

	unaided	**LIN**	**AV**	**SC**
BD	-2.2 dB	-0.2 dB	4.8 dB	4.3 dB
EJ	-4.1 dB	-1.7 dB	-1.0 dB	-1.8 dB
GH	-2.1 dB	-1.3 dB	-0.9 dB	0.0 dB
HM	-1.7 dB	-3.3 dB	-2.1 dB	-2.7 dB
MW	2.8 dB	5.9 dB	3.2 dB	5.0 dB

Table A.1: *Results of the speech intelligibility measurements in noise with algorithms LIN, AV and SC, as well as unaided. Given are the SRT's for each subject with each algorithm.*

Algo1	**Algo2**	**Significance**
unaided	LIN	0,1794
unaided	AV	0,1620
unaided	SC	0,1126
LIN	AV	0,4953
LIN	SC	0,3100
AV	SC	0,7688

Table A.2: *Results of a paired T–Test*[2] *calculated for the data samples shown in Table A.1 with the null hypothesis 'samples are equal'. Here the null hypothesis can not be rejected at a Significance level of* 0.05 *for all combinations (i.e. Significance is greater* 0.05*) and therefore, there is no significant difference in the SRT's measured using algorithms LIN, AV and SC, as well as unaided.*

A.5 Additional Loudness Scaling Data

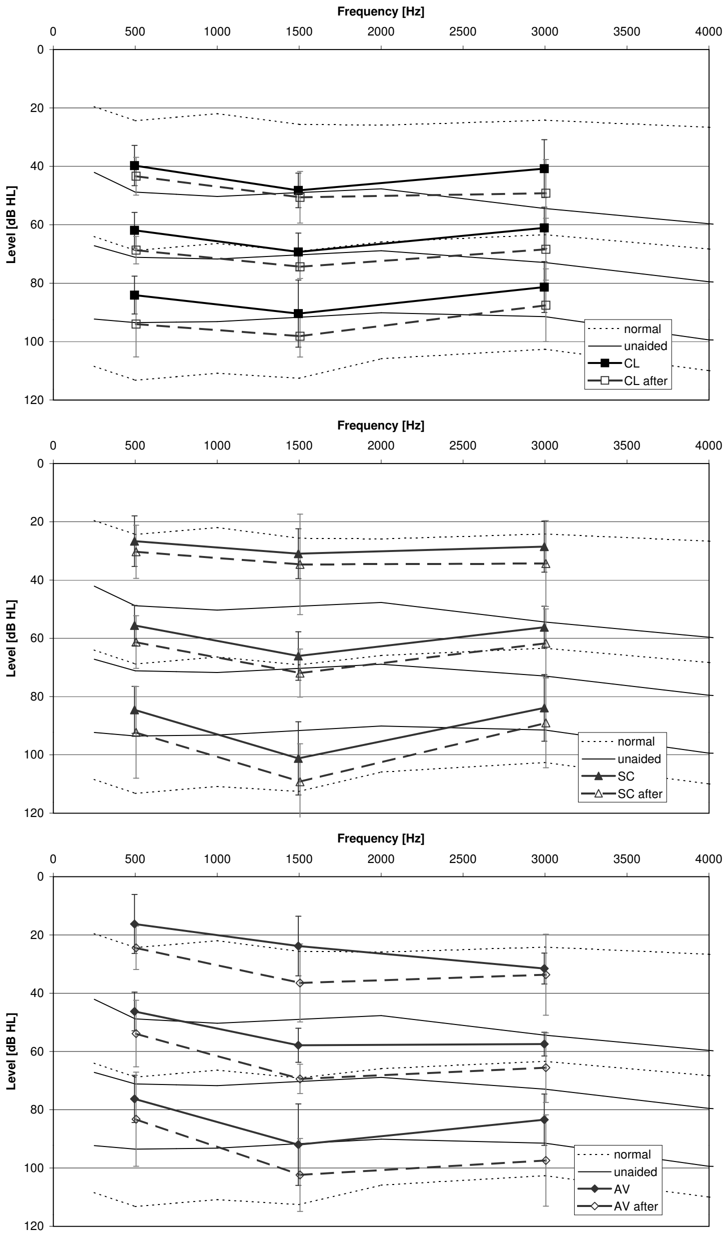

Figure A.6: *Results of the narrowband loudness scaling experiments with algorithms CL (top), SC (middle), and AV (bottom), as well as unaided. Each plot shows curves of equal–loudness for the loudness impressions 'very soft', 'medium', and 'very loud', respectively (mean values over the four hearing–impaired subjects that participated in measurements after the field test). Solid lines show the data observed before the field test. Dashed lines show the data observed after the field test. The dotted lines show normal–hearing data.*

Appendix B

Comparing Loudness Matching with Loudness Scaling

B.1 Subject Groups

Subject	Group A	Group B	Group C	Group D
BG	•			•
CR	•			
JA	•		•	
KS		•		•
JV		•	•	
OW		•		
GG			•	
JO			•	
AA				•
BF				•
HB				•
VG				•

Table B.1: *A filled circle in the Table indicates that the subject on the same row in the left column participated in the experiments performed with the group denoted at the top of the column.*

B.2 Model Function to Parameterize the Loudness Function

To parameterize the loudness functions observed in the loudness scaling experiments, a model function proposed by Brand *et al.* [1998] was used. It consists of two linear parts which have independent positive slope values m_{lo} and m_{hi} and which are connected at the L_{cut}. For smoothing the sudden change of the slope at the L_{cut} a Bezier fit is applied between the categories L_{15} and L_{35} (levels corresponding to loudness categories 15 and 35, respectively):

$$F(L) = \begin{cases} 25 + m_{\text{lo}}(L - L_{\text{cut}}) & , \quad L \leq L_{15} \\ BEZIER(L, L_{\text{cut}}, L_{15}, L_{35}) & , \quad L_{15} < L < L_{35} \\ 25 + m_{\text{hi}}(L - L_{\text{cut}}) & , \quad L \geq L_{35} \end{cases} \tag{B.1}$$

Because of the smoothing the L_{cut} parameter in Eq. (B.1) does not represent the medium loudness level L_{25} but the level where the two linear parts would meet when they were not smoothed.

B.3 Individual Data

Figures B.1 to B.4 show the individual data (mean and standard deviation) corresponding to the averaged data presented in chapter 3. For monaural measurements the data is averaged across both ears.

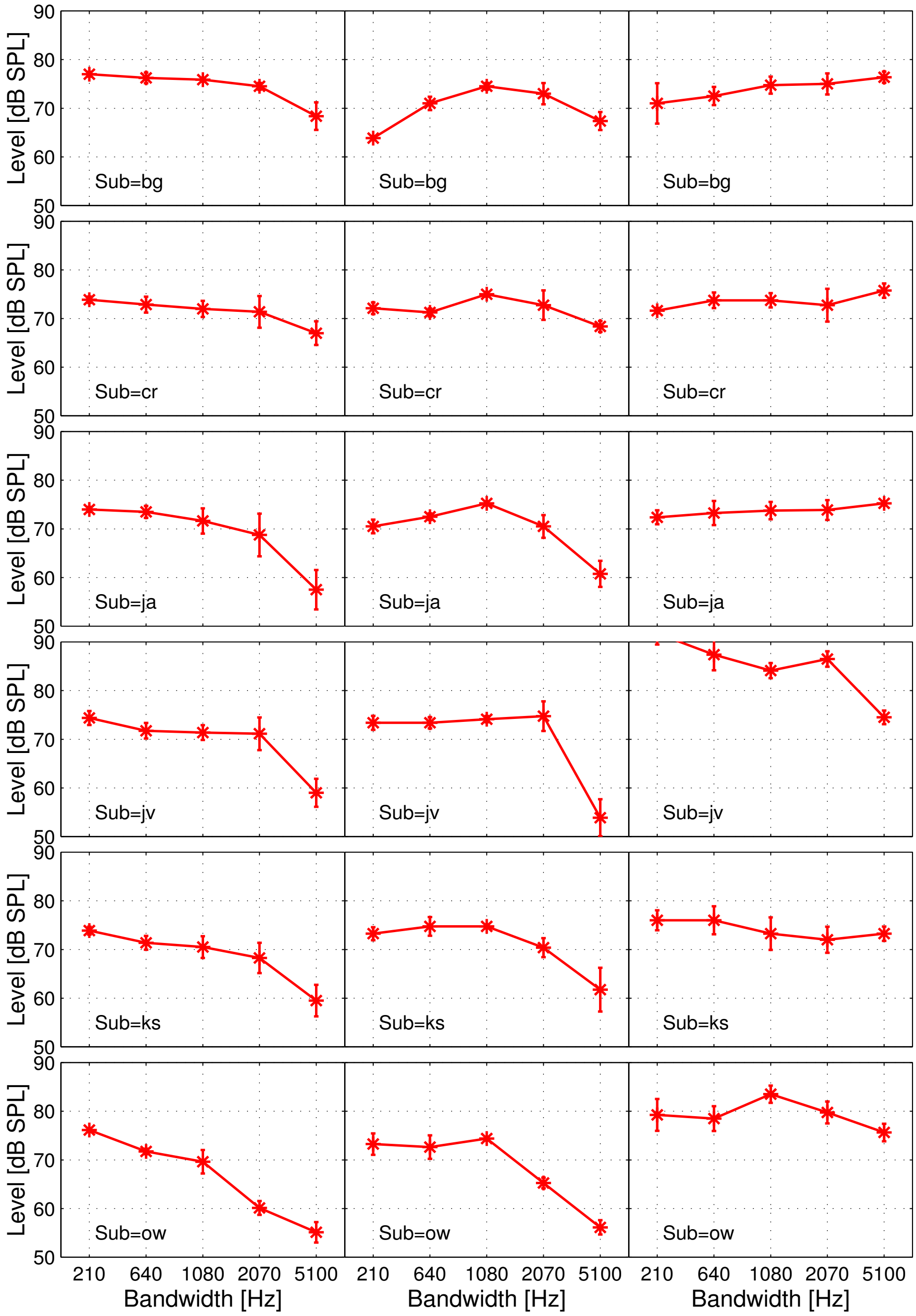

Figure B.1: *Individual equal–loudness levels as a function of the test–signal bandwidth measured with non–interleaved tracks (paradigm LM–METHOD–1) for subjects that contributed data to Figure 3.2. Three reference bandwidths were tested: 210 Hz (1 Bark, left panel), 1080 Hz (5 Bark, mid panel) and 5100 Hz (17 Bark, right panel). The level of the reference signals was fixed at 75 dB SPL. Each panel shows the averaged data across all trials performed by the subject. The vertical bars show plus minus one standard deviation of the intra–individual mean.*

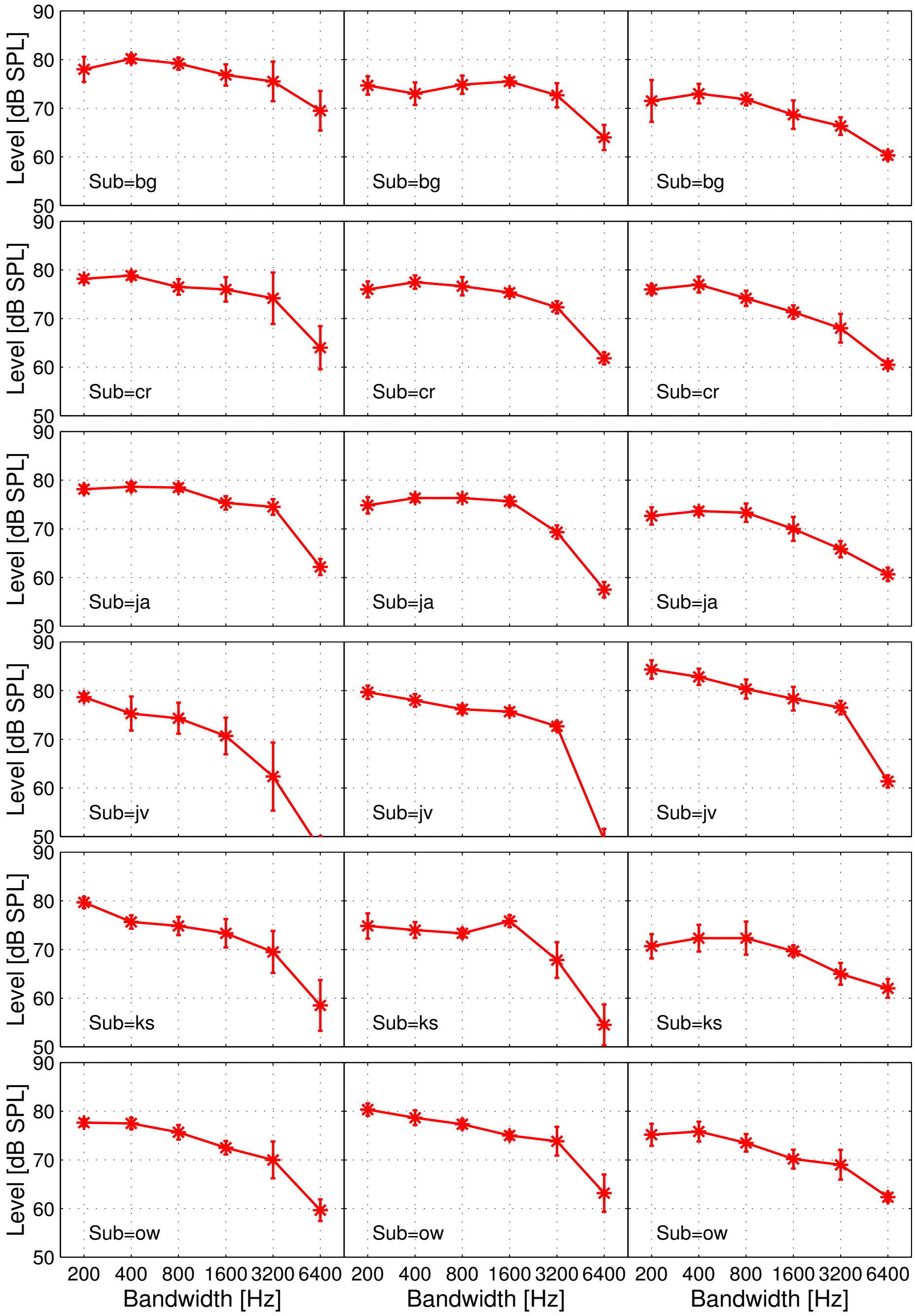

Figure B.2: *Individual equal–loudness levels as a function of the test–signal bandwidth measured with interleaved tracks (paradigm LM–METHOD–4) for subjects that contributed data to Figure 3.3. The presentation level of the reference signals was a priory corrected according to the expected amount of loudness summation: 200 Hz at 78 dB SPL (left panel), 1600 Hz at 75 dB SPL (mid panel) and 6400 Hz at 60 dB SPL (right panel). Each panel shows the averaged data across all trials performed by the subject. The vertical bars show plus minus one standard deviation of the intra–individual mean.*

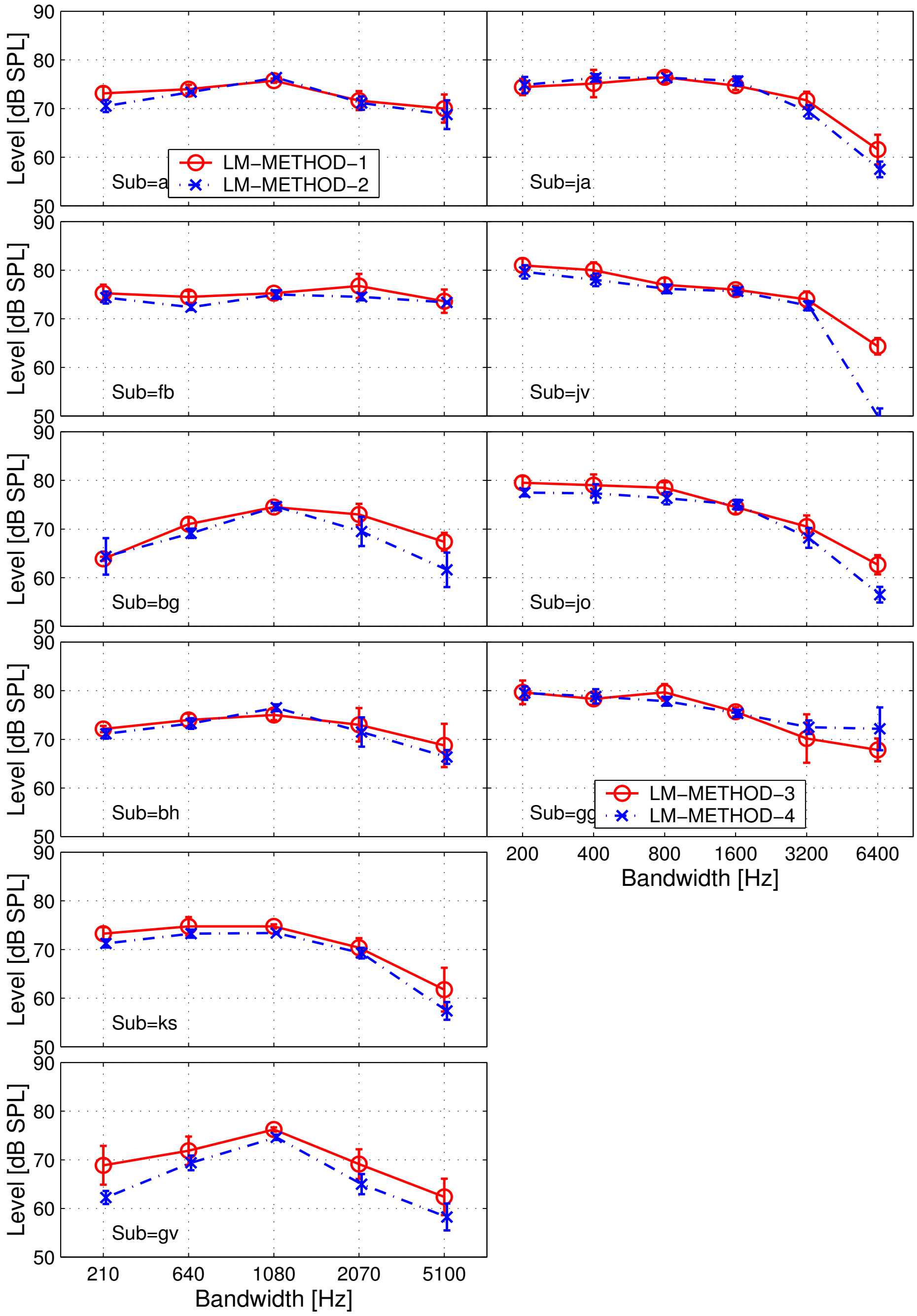

Figure B.3: *Individual equal–loudness levels as a function of the test–signal bandwidth measured with subsequent (×, dash–dotted lines) and interleaved (○, solid lines) tracks for subjects that contributed data to Figure 3.5. Left panel shows the results for subjects in group D and methodological paradigms LM–METHOD–1 and LM–METHOD–2, right panel shows the results for subjects group C and the methodological paradigms LM–METHOD–3 and LM–METHOD–4. The bandwidths of the reference signals in the left and right panel was 1600 Hz and 1080 Hz, respectively. The level of the reference signals was fixed at 75 dB SPL. Each panel shows the averaged data across all trials performed by the subject. The vertical bars show plus minus one standard deviation of the intra–individual mean.*

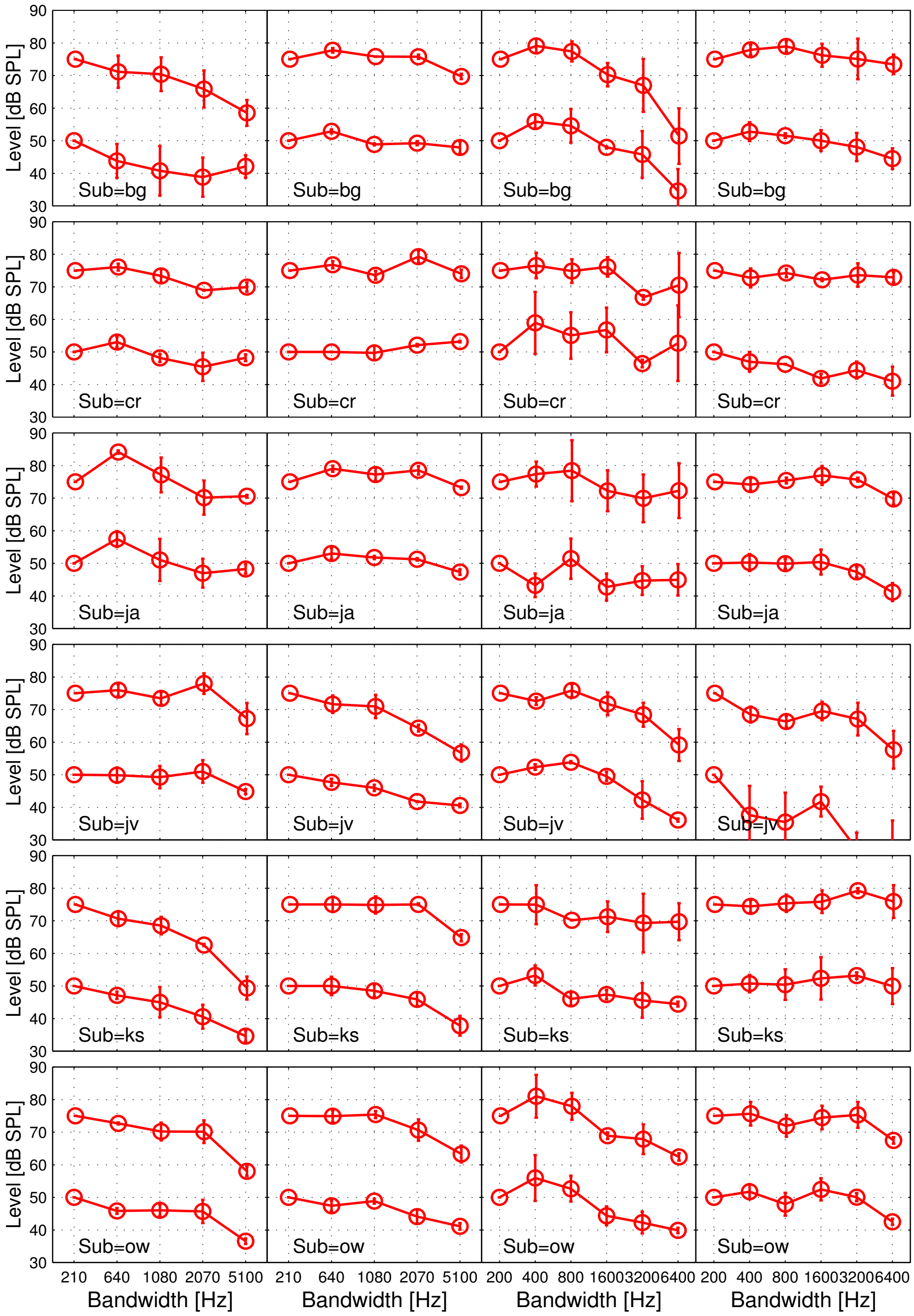

Figure B.4: *Individual equal–loudness levels across bandwidth calculated from the loudness scaling data for subjects that contributed data to Figures 3.6 and 3.7. From the left panels to the right the results for measurement paradigms LS–METHOD–1, LS–METHOD–2, LS–METHOD–3 and LS–METHOD–4 are shown, respectively. Each panel shows the averaged data across the data calculated for that paradigm/subject. The vertical bars show plus minus one standard deviation of the intra–individual mean.*

B.4 Rank Order for Experiment 3

Figure B.5 shows the rank order of the amount of loudness summation for measurement paradigms *LM–METHOD–1* and *LM–METHOD–2* tested in experiment 3 for subjects in group D. Figure B.6 shows the corresponding data for group C and measurement paradigms *LM–METHOD–3* and *LM–METHOD–4*. The amount of loudness summation was calculated by the difference in level at equal–loudness between the signal with the smallest and the signal with the largest bandwidth under test.

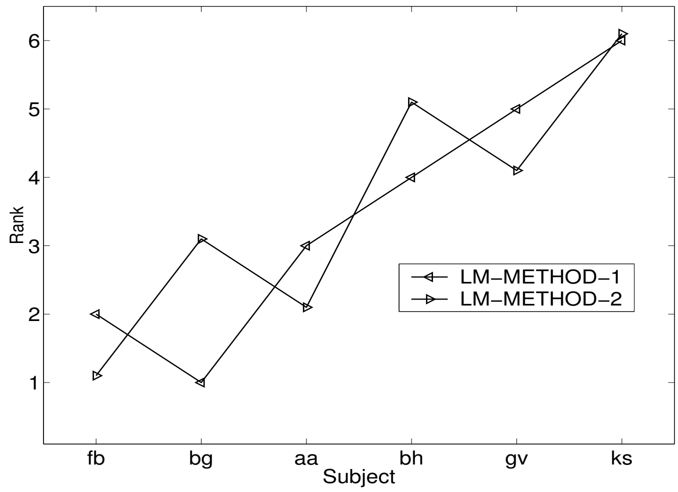

Figure B.5: *The amount of loudness summation between the stimulus with the smallest and the largest bandwidth is plotted in terms of the rank order for measurement paradigms LM–METHOD–1 and LM–METHOD–2 for subjects in group D.*

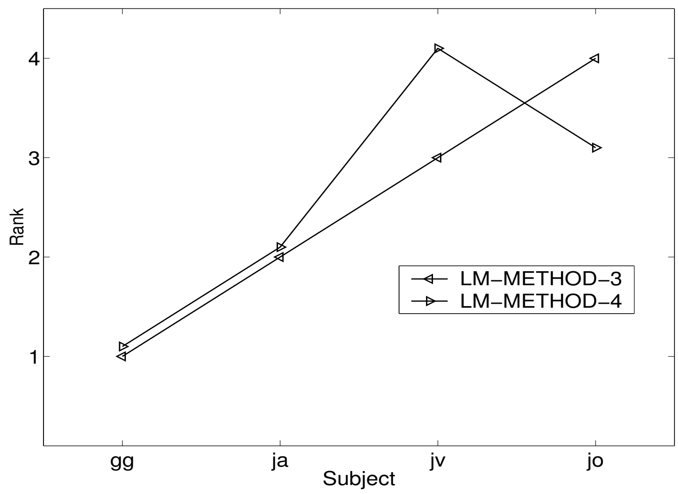

Figure B.6: *The amount of loudness summation between the stimulus with the smallest and the largest bandwidth is plotted in terms of the rank order for measurement paradigms LM–METHOD–3 and LM–METHOD–4 for subjects in group C.*

B.5 Correlation between the Measurement Paradigms

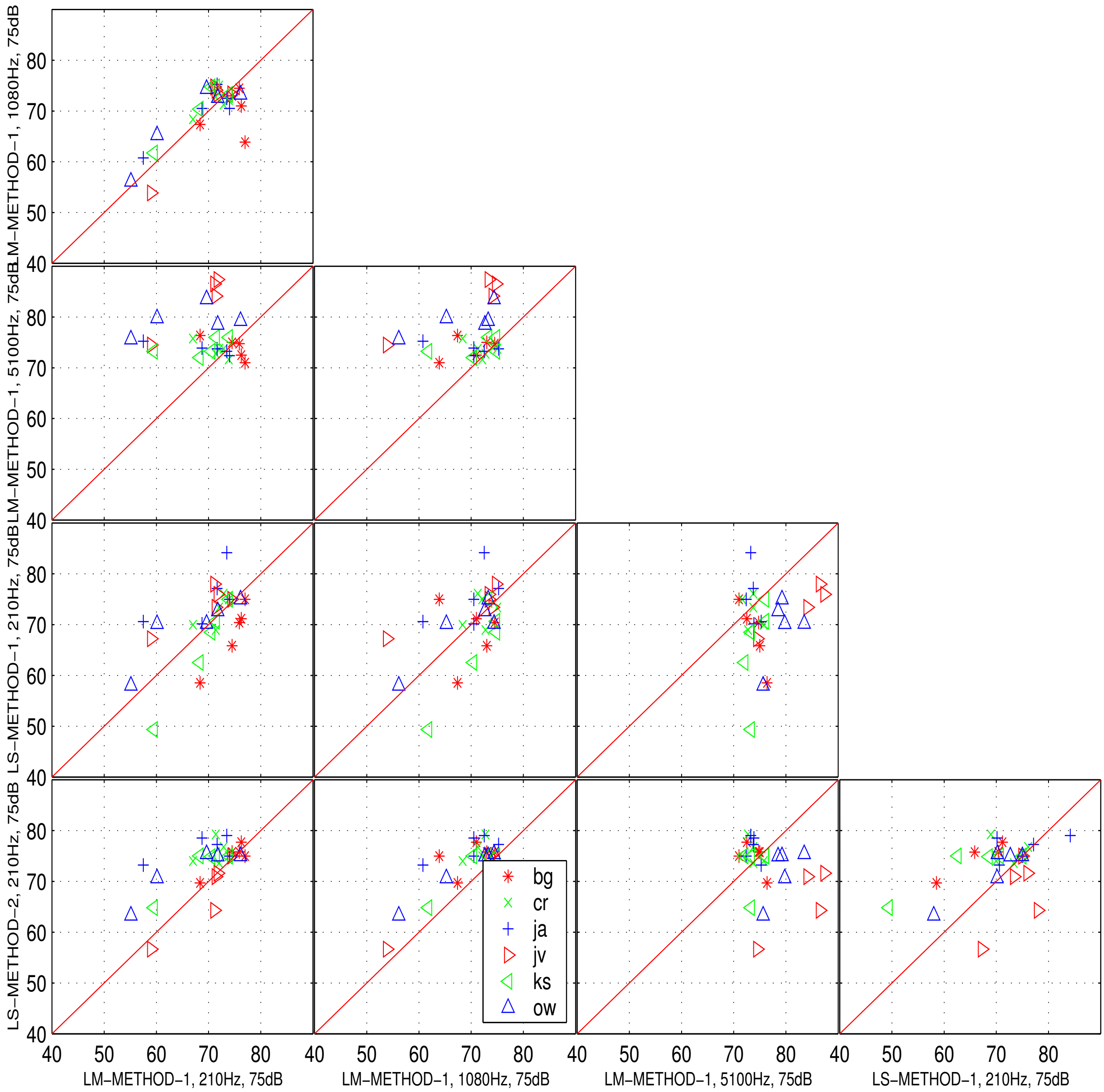

Figure B.7: *Correlation between measurement paradigms employing UEN–stimuli. Each panel shows the intra–individual mean levels adjusted for the test stimuli to produce the same loudness as the reference stimulus. The data for two measurement paradigms (including reference bandwidth and level) are plotted against another. The most bottom abscissa labels and most left ordinate labels denote the measurement paradigm, reference bandwidth and reference level, respectively. Each symbol denotes the data for one subject.*

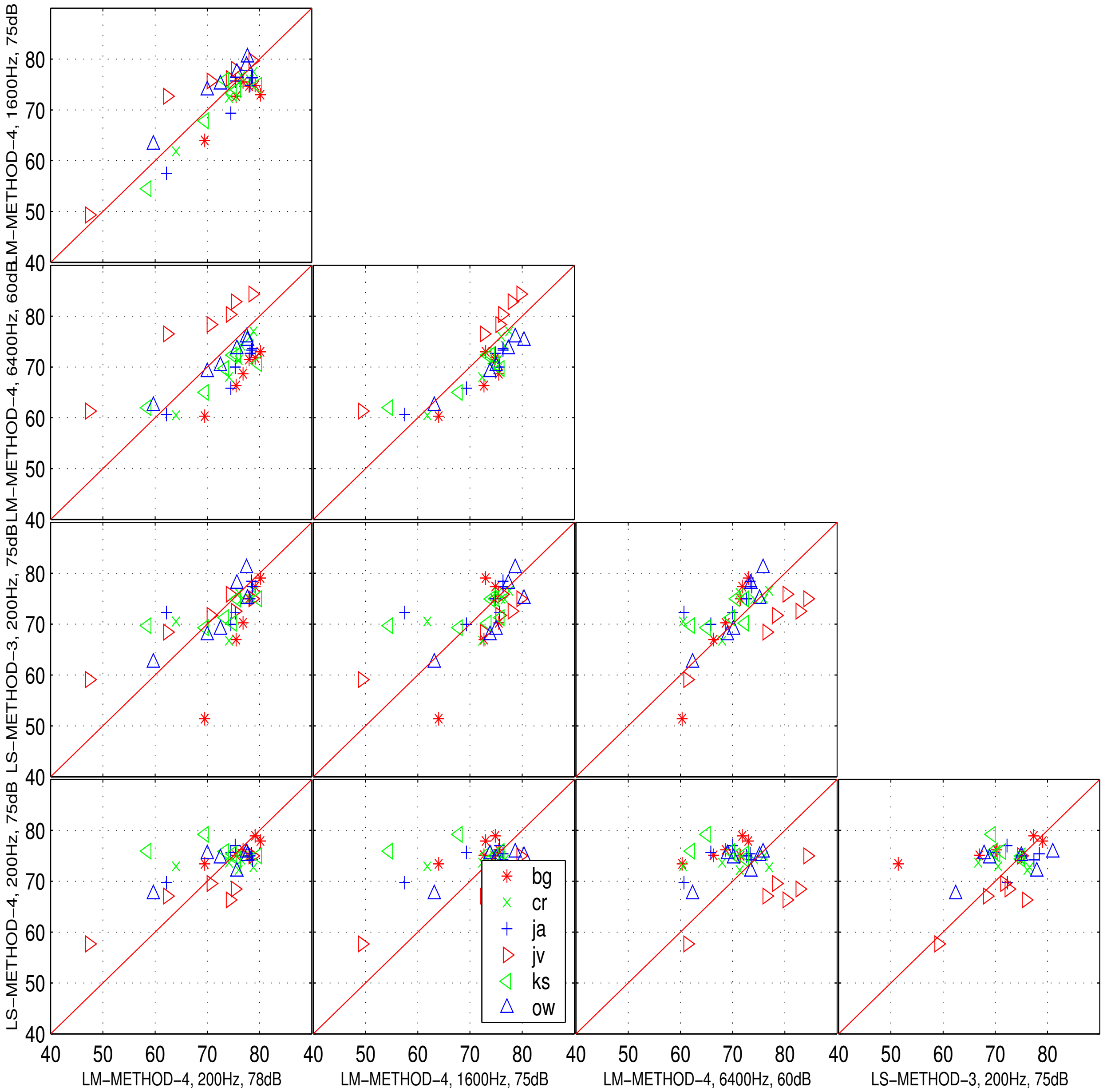

Figure B.8: *Same as Figure B.7 for measurement paradigms employing LNN–stimuli.*

Appendix C

The Oldenburg Loudness Model

C.1 Comparison between recent ISO 226 and its preliminary Refinement

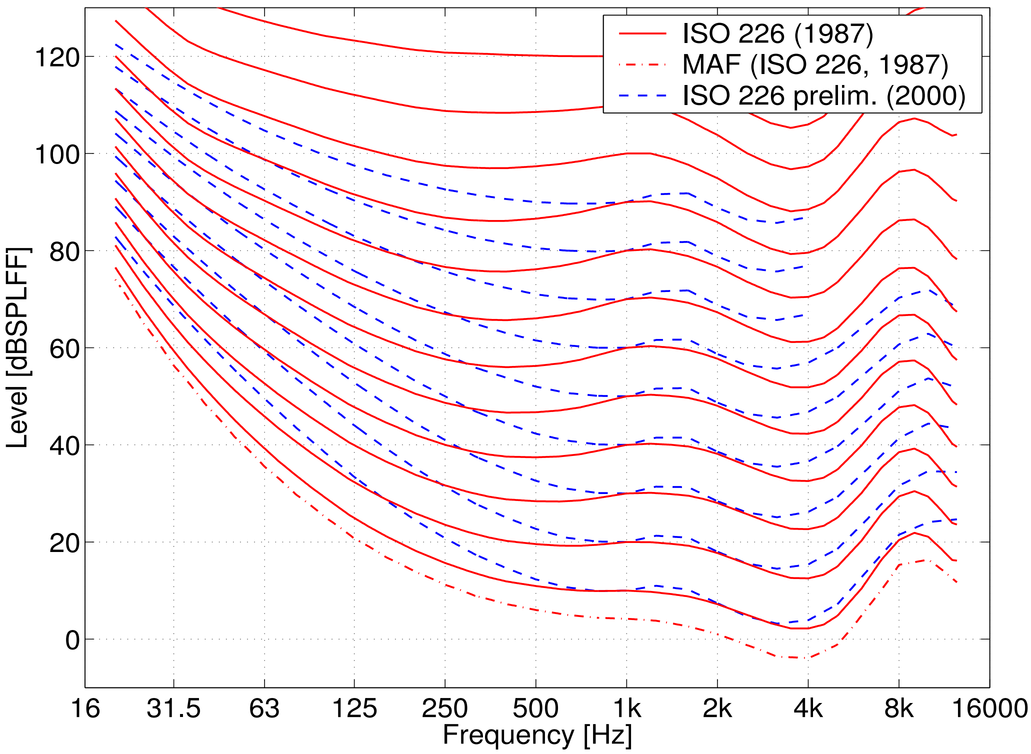

Figure C.1: *Equal–loudness level contours in phon (ELLC-P). Solid lines show ELLC-P's according to ISO 226(E) [1987]. Dashed lines show data from it's recent preliminary draft revision [Committee Draft ISO/CD 226, 2000]. The dash–dotted line shows MAF according to ISO 226(E) [1987].*

Appendix D

Evaluation of the extended Oldenburg Loudness Model

D.1 Audiogram Data of the Hearing–impaired Subjects

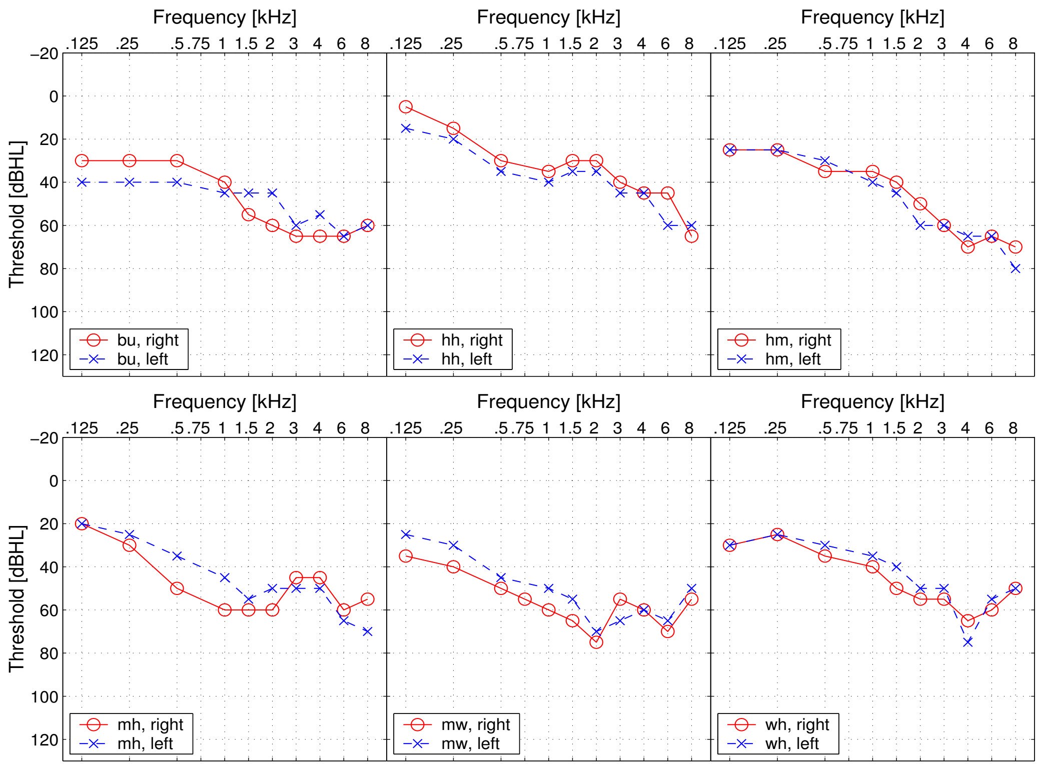

Figure D.1: *Pure tone audiogram (air conduction thresholds) in dB hearing level (dB HL) for the hearing–impaired subjects.*

D.2 Fitted Polynomial Coefficients

For *MODEL–2* and *MODEL–3* the pre–factor β was individually fitted to the narrow–band loudness scaling data of each subject by fitting the coefficients of the polynomial

$$\beta = \beta(f_{ERB}) = a \cdot (f_{ERB})^2 + b \cdot f_{ERB} + c. \tag{D.1}$$

in three ways: a *constant* polynomial fit (i.e., $a \equiv 0$ and $b \equiv 0$, c fitted), a *linear* polynomial fit (i.e., $a \equiv 0$, b and c fitted) and a *cubic* polynomial fit (i.e., a, b and c fitted). The respective coefficients for the group of normal–hearing and impaired–hearing subjects are summarized in tables D.1 and D.2.

MODEL–2, normal–hearing subjects							
		constant	linear		quadratic		
Subject	Ear	c	b	c	a	b	c
aa	left	0.864	0.014	0.565	-0.000	0.019	0.518
aa	right	0.881	0.013	0.576	-0.002	0.121	-0.480
bf	left	0.957	0.009	0.759	-0.000	0.019	0.671
bf	right	0.883	0.023	0.347	-0.000	0.033	0.256
gb	left	0.896	0.009	0.691	-0.000	0.014	0.642
gb	right	0.971	0.012	0.697	0.000	0.008	0.737
hb	left	0.912	0.022	0.419	-0.001	0.084	-0.179
hb	right	0.949	0.012	0.689	-0.001	0.071	0.145
sk	left	1.081	0.014	0.781	-0.000	0.018	0.747
sk	right	1.087	0.021	0.661	-0.001	0.084	0.074
vg	left	0.848	0.013	0.534	-0.000	0.033	0.348
vg	right	0.836	0.013	0.518	-0.001	0.039	0.264

MODEL–3, normal–hearing subjects							
		constant	linear		quadratic		
Subject	Ear	c	b	c	a	b	c
aa	left	0.956	0.021	0.527	-0.000	0.033	0.404
aa	right	0.981	0.006	0.841	-0.002	0.088	0.074
bf	left	1.047	0.003	0.975	0.000	-0.001	1.015
bf	right	0.982	0.016	0.623	0.000	-0.002	0.783
gb	left	0.976	0.006	0.836	0.000	-0.003	0.911
gb	right	1.050	0.004	0.969	0.001	-0.022	1.190
hb	left	1.016	0.013	0.742	-0.001	0.040	0.483
hb	right	1.065	0.009	0.876	-0.001	0.054	0.480
sk	left	1.166	0.005	1.063	0.001	-0.017	1.235
sk	right	1.196	0.012	0.950	-0.000	0.031	0.787
vg	left	0.920	0.009	0.718	0.000	0.008	0.724
vg	right	0.910	0.009	0.706	-0.000	0.018	0.614

Table D.1: *Fitted coefficients for MODEL–2 (upper table) and MODEL–3 (lower table) for the group of normal–hearing subjects.*

MODEL–2, impaired–hearing subjects							
		constant	linear		quadratic		
Subject	Ear	c	b	c	a	b	c
bu	left	2.290	0.072	1.099	0.004	-0.066	2.183
bu	right	1.721	0.093	0.403	-0.001	0.135	0.035
hh	left	1.449	0.067	0.293	0.004	-0.086	1.500
hh	right	1.216	0.070	-0.043	0.002	-0.019	0.711
hm	left	1.655	0.085	0.560	0.002	0.008	1.199
hm	right	1.480	0.097	0.259	0.000	0.093	0.299
mh	left	1.528	0.051	0.697	0.002	-0.033	1.290
mh	right	1.584	0.033	0.975	0.001	-0.003	1.209
mw	left	1.677	0.038	1.087	-0.003	0.169	0.035
mw	right	1.585	-0.006	1.715	0.000	-0.012	1.758
wh	left	2.199	0.073	1.077	0.001	0.017	1.535
wh	right	1.323	0.012	1.140	-0.002	0.087	0.641

MODEL–3, impaired–hearing subjects							
		constant	linear		quadratic		
Subject	Ear	c	b	c	a	b	c
bu	left	2.522	0.053	1.600	0.005	-0.130	3.055
bu	right	2.019	0.087	0.698	0.001	0.065	0.883
hh	left	1.628	0.048	0.825	0.005	-0.133	2.077
hh	right	1.393	0.049	0.580	0.003	-0.052	1.307
hm	left	2.133	0.247	-1.594	0.013	-0.281	3.107
hm	right	1.986	0.228	-1.592	0.003	0.124	-0.977
mh	left	1.725	0.037	1.110	0.002	-0.046	1.661
mh	right	1.760	0.022	1.350	0.001	-0.012	1.562
mw	left	2.163	0.042	1.366	-0.003	0.161	0.478
mw	right	1.679	-0.022	2.182	0.001	-0.067	2.528
wh	left	2.415	0.059	1.440	0.003	-0.049	2.354
wh	right	1.628	-0.004	1.716	-0.003	0.119	0.786

Table D.2: *Fitted coefficients for MODEL–2 (upper table) and MODEL–3 (lower table) for the group of impaired–hearing subjects.*

References

Aibara R., Welsh J.T., Puria S., and Goode R.L. (**2001**)
Human middle–ear sound transfer function and cochlear input impedance. Hear. Res. **152**, 100–109. 80

Allen J.B. (**1977**)
Short term spectral analysis, synthesis and modification by discrete Fourier transform. IEEE Trans. Acoust., Speech, Signal Processing **25**, 235–238, ISSN 1053–587X. 110

Allen J.B., Hall J.L., and Jeng P.S. (**1990**)
Loudness growth in 1/2-octave bands (LGOB) – A procedure for the assessment of loudness. J. Acoust. Soc. Am. **88**, 745–753. 35, 62

Appell J.E. and Hohmann V. (**1998**)
Messung der Lautheitssummation bei Normal- und Schwerhörenden. In *Fortschritte der Akustik – DAGA '98*, DPG-Verlag, Bad Honnef. 47

Appell J.E., Hohmann V., and Kollmeier B. (**1995**)
Vergleich verschiedener digital realisierter Signalverarbeitungs–Strategien für Dreikanal–Hörgeräte mit Dynamikkompression.. Audiologische Akustik **34(3)**, 134–143. 7, 9, 15, 30, 35

Appell J.E., Gabriel B., Kollmeier B., Marzinzik M., Raß U., Steeger G.H., and Wittkop T. (**1999**)
Implementation and Field-Test Evaluation of Simplified Noise Suppression and Dynamic Compression Algorithms on a DSP Prototype Hearing Aid. Technical report, Carl von Ossietzky Universität Oldenburg, AG medizinische Physik. Deliverable SP-3a of the SPACE Project (DE 3012). 16

Benson D., Clark T.M., and Johnson J.S. (**1992**)
Patient experience with multiband full dynamic range compression. Ear and Hear. **13(5)**, 320–330. 30

Betke K. (**1987**)
Kurven gleicher Pegellautstärke im Frequenzbereich 50 Hz bis 1 kHz. In *Fortschritte der Akustik – DAGA '87*, DPG Kongreßgesellschaft mbH, Bad Honnef. 71

Betke K. (**1991**)
Hörschwellen und Kurven gleicher Pegellautstärke im ebenen Schallfeld. Ph.D. thesis, Universität Oldenburg. 67, 71, 80, 97, 105

Blum R. (**1999**)
Experimente und Modellvorstellungen zur Lautheitswahrnehmung. Master's thesis, Universität Oldenburg. 51, 62, 63, 66, 69, 73, 78, 79, 81, 95, 99, 104

Blum R., Hohmann V., Dau T., and Kollmeier B. (**1998**)
Vergleich verschiedener Methoden Meßmethoden zur Lautheit. In *Fortschritte der Akustik – DAGA '98*, DPG-Verlag, Bad Honnef. 62

Brand T. (**2000**)
Analysis and optimization of psychphysical procedures in audiology. Ph.D. thesis, Universität Oldenburg, BIS-Verlag Oldenburg. 37, 38, 72, 88, 93

Brand T. and Hohmann V. (**2001a**)
Design and Evaluation of an Adaptive Procedure for Categorical Loudness Scaling. J. Acoust. Soc. Am. *(in preparation)*. 36, 38, 62

Brand T. and Hohmann V. (**2001b**)
Effect of Hearing Loss, Center Frequency, and Bandwidth on the Shape of Loudness Functions in Categorical Loudness Scaling. Audiology *(accepted)*. 93

Brand T. and Kollmeier B. (**1996**)
Adaptive Testverfahren in der Audiologie. In *Fortschritte der Akustik – DAGA '96*, DEGA e.V., Oldenburg, pp. 60–63. 17

Brand T., Hohmann V., and Kollmeier B. (**1997a**)
Adaptive categorical loudness scaling. In *Contributions to Psychological Acoustics – 7th Oldenburg Symposium on Psychological Acoustics*, edited by A. Schick and M. Klatte, University of Oldenburg, Oldenburg, ISBN 3-8142-0572-3. 38

Brand T., Hohmann V., and Kollmeier B. (**1997b**)
Die adaptive Hörflächenskalierung. In *Hörflächenskalierung, Buchreihe Audiologische Akustik*, edited by B. Kollmeier, median–verlag, Heidelberg, pp. 146–166. ISBN 3–922766–26–9. 38

Brand T., Hohmann V., and Kollmeier B. (**1997c**)
Meßgenauigkeit der (adaptiven) kategorialen Hörflächenskalierung. In *Fortschritte der Akustik – DAGA '97*, DEGA e.V., Oldenburg. 38

Brand T., Hohmann V., and Kollmeier B. (**1998**)
Krümmung der Lautheitsfunktion in Abhängigkeit von der Bandbreite des Signals bei Normal- und Schwerhörenden. In *Fortschritte der Akustik – DAGA '98*, DPG-Verlag, Bad Honnef. 39, 62, 72, 77, 88, 117

Bray V.H., Chabries D.M., Davis K.L., and Johnson J.S. (**1998**)
Digital Signal Processing (DSP) derived from a Nonlinear Auditory Model. Poster at AAA – Los Angelos – April 4, 1998. 84

Harten-de Bruijn v.H.E., Kreveld v.S.G.M., Dreschler W.A., and Verschuure J. (**1996**)
Design of two syllabic non-linear multichannel signal processors and the results of speechtests in noise. Technical report, Heardip Report. 30

Brunt M.A. (**1994**)
Test of cochlear function. In *Handbook of clinical Audiology*, edited by J. Katz, Williams and Wilkins, Baltimore. 5, 34

Bustamente P.K. and Braida L.D. (**1987**)
Principal-Component amplitude compression for hearing impaired. J. Acoust. Soc. Am. **82(4)**. 8

Buus S., Müsch H., and Florentine M. (**1998**)
On loudness at threshold. J. Acoust. Soc. Am. **104(1)**, 399–410. 53

Buus S., Florentine M., and Poulson T. (**1999**)
Temporal integration of loudness in listeners with hearing losses of primarily cochlear origin. J. Acoust. Soc. Am. **105(6)**, 3464–3480. 35

Byrne D. (**1986**)
The National Acoustics Laboraties' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. Ear and Hear. **7**, 257–265. 8

Chalupper J. (**2000**)
Modellierung der Lautstärkeschwankung für Normal– und Schwerhörende. In *Fortschritte der Akustik – DAGA 2000*, DPG-Verlag, Bad Honnef, pp. 254–255. 6, 52

Committee Draft ISO/CD 226 (**2000**)
First ISO/CD 226 "Acoustics — Normal equal–loudness level contours". International Organization for Standardization, Geneva. Reference number ISO/TC 43 N 952. 71, 72, 73, 75, 76, 80, 83, 86, 104, 127

de Genaro S., Braida L.D., and Durlach N.I. (**1986**)
Multichannel compression for several listeners. J. Reh. Res. Dev. **23/1**. 8

Elberling C. (**1999**)
Loudness Scaling Revisited. J. Am. Acad. Audiol. **10**, 428–260. 34

Fastl H., Jaroszewski A., Schorer E., and Zwicker E. (**1990**)
Equal Loudness Contours between 100 and 1000 Hz for 30, 40 and 70 phon. Acustica **70**, 197–201. 67, 71, 73, 80, 97, 105

Fechner G.T. (**1860**)
Elemente der Psychophysik. Breitkopf und Härtel, Leipzig. Mehrbändiges Werk. 5

Festen J.M. (**1999**)
Spectro-temporal modulation transfer and speech intelligibility for multi–band amplitude compression. Handout. 8, 30

Fletcher H. (**1995**)
Speech and hearing in communication. Acoustical Society of America, ISBN 1-56396-393-0 (hbk.). Edited by J. B. Allen. 52

Fletcher H. and Munson W.A. (**1933**)
Loudness, its definition, measurement and calculation. J. Acoust. Soc. Am. **5**, 82–108. 5, 6

Fletcher H. and Steinberg J. (**1924**)
The dependence of the loudness of a complex sound upon the energy in the various frequency regions of the sound. Phys. Rev. **24**, 306–317. 6

Florentine M. and Zwicker E. (**1979**)
A model of loudness summation applied to noise–induced hearing loss. Hear. Res. **1**, 121–132. 6, 46, 52, 59, 60

Florentine M., Buus S., and Poulsen T. (**1996**)
Temporal Integration of Loudness as a Function of Level. J. Acoust. Soc. Am. **99(3)**, 1633–1644. 35, 42, 47

Fowler E.P. (**1936**)
A method for the early detection of otosclerosis. Arch. Otolaryngol. **24**, 731–741. 5, 34

Fröhlich T. (**1993**)
Digitale Signalverarbeitung für Hörgeräte: Mehrkanalige Lautheitskorrektur im Frequenzbereich. Ph.D. thesis, ETH Zürich. 8, 84

Gabriel B. (**1996**)
Equal–Loudness Level Contours: Procedures, Factors and Models.. Ph.D. thesis, Universität Oldenburg. Shaker Verlag, Aachen. 34, 71, 86, 97

Gabriel B., Kollmeier B., and Mellert V. (**1994**)
Einfluß verschiednener Meßmethoden auf Kurven gleicher Pegellautstärke. In *Fortschritte der Akustik – DAGA '94*, DPG-Verlag, Bad Honnef, pp. 1085–1088. 67, 71, 97, 105

Gatehouse S. (**1992**)
The time course and magnitude of perceptual acclimatization to frequency response: Evidence from monaural fitting of hearing aids. J. Acoust. Soc. Am. **92**, 1258–1268. 30

Gescheider G.A. and Hughson B.A. (**1991**)
Stimulus context and absolute magnitude estimation: A study of individual differences. Percept. Psychophys. **50(1)**, 45–57. 62

Glasberg B.R. and Moore B.C.J. (**1990**)
Derivation of auditory filter shapes from notched-noise data. Hear. Res. **47**, 103–138. 57, 70, 85

Goedegebure A., M. H., Maas A.J.J., and J. V. (**1996a**)
Fitting of hearing aids. Technical report, Heardip Report. 8

Goedegebure A., M. H., Maas A.J.J., and J. V. (**1996b**)
Evaluation of compression in field. Technical report, Heardip Report. 30

Hansen M. (**2000**)
Einfluß der Kompressionszeitkonstanten auf subjektive Sprachverständlichkeit und Klangqualität von Högeräten. In *Fortschritte der Akustik – DAGA 2000*, DEGA e.V., Oldenburg, pp. 260–261. 8

Hartmann W.M. and Pumplin J. (**1988**)
Noise power fluctuations and the masking sine signals. J. Acoust. Soc. Am. **83**, 2277–2289. 34, 36

Hellbrück J. (**1991**)
Loudness scaling: Ratio scales, intersubject variability and some other problems. In *Contributions to Psychological Acoustics*, edited by A. Schick, BIS, Oldenburg, volume 5. 62

Heller O. (**1985**)
Hörfeldaudiometrie mit dem Verfahren der Kategorienunterteilung (KU). Psycholog. Beiträge **27**, 478–493. 35

Hellman R.P. and Meiselman C.H. (**1988**)
Prediction of individual loudness exponents from cross–modality matching. J. Speech Hear. Res. **31**, 605–615. 62

Hellman R.P. and Meiselman C.H. (**1990**)
Loudness relation for individuals and groups in normal and impaired hearing. J. Acoust. Soc. Am. **88**, 2596–2606. 54, 59, 60, 61, 62

Hohmann V. (**1993**)
Dynamikkompression für Hörgeräte — Psychoakustische Grundlagen und Algorithmen, volume 93 of *Biotechnik*. VDI-Verlag, Düsseldorf. 35, 52, 60, 62, 73, 84, 105

Hohmann V. and Kollmeier B. (**1995a**)
The effect of multichannel dynamic compression on speech intelligibility. J. Acoust. Soc. Am. **97(2)**, 1191–1195. 8, 12, 30

Hohmann V. and Kollmeier B. (**1995b**)
Weiterentwicklung und klinischer Einsatz der Hörfeldskalierung.. Audiologische Akustik **34(2)**, 48–64. 16, 19, 35, 37, 69, 72, 88

Hohmann V. and Kollmeier B. (**1996**)
Perceptual models for hearing aid algorithms.. In *Psychoacoustics, Speech and Hearing Aids*, edited by B. Kollmeier, World Scientific, Singapore, pp. 193–202. 84

Hohmann V., Appell J.E., Marzinzik M., and Kollmeier B. (**1997**)
Speech processing for hearing aids I: Dynamic compression motivated by a loudness model. Acustica · acta acustica **83(4)**. 84

Hübner R. and Ellermeier W. (**1993**)
Additivity of loudness across critical bands: A critical test. Percept. Psychophys. **54**, 185–189. 35

Humes L.E. and Jesteadt W. (**1991**)
Models of the effects of threshold on loudness growth and summation. J. Acoust. Soc. Am. **90(4)**, 1933–1943. 53

ISO 226(E) (**1987**)
Acoustics — Normal equal–loudness level contours. International Organization for Standardization, Geneva. Reference number ISO 226:1987(E). 54, 56, 65, 66, 67, 69, 71, 72, 74, 75, 80, 86, 94, 97, 127

Jesteadt W.A. (**1980**)
An adaptive procedure for subjective judgments. Percept. Psychophys. **28(1)**, 85–88. 35

Jesteadt W.A., ed. (**1997**)
Modeling sensorineural Hearing Loss, Lawrence Erlbaum Assoc., Mahwah, NJ. 84

Kießling J. (**1995**)
Zum überschwelligen Lautheitsanstieg bei Schallempfindungsschwerhörigen — Konsequenzen für die Hörgeräte–Entwicklung und –Anpassung. Audiologische Akustik **34(2)**, 82–89. 52, 60, 84, 105

Kießling J. (**1996**)
Scaling methods for the selection, fitting and evaluation of hearing aids. In *Psychoacoustics, Speech and Hearing Aids*, edited by B. Kollmeier, World Scientific, Singapore, pp. 297–306. ISBN 981022561X. 35, 62

Kießling J. and Steffens T. (**1991**)
Clinical evaluation of a programmable three channel automatic gain control amplification system. Audiology **30**, 70–81. 15

Kießling J., Schubert M., and Wagner I. (**1994**)
Lautheitsskalierung – Ein Verfahren zum quantitativen Recruitmentnachweis. HNO **42**, 350–357. 52, 60, 62, 74, 84, 105

Kohlrausch A., Fassel R., van der Heijden M., Kortekaas R., van de Par S., Oxenham A.J., and Püschel D. (**1997**)
Detection of Tones in Low-noise Noise: Further Evidence for the Role of Envelope Fluctuations. Acustica · acta acustica **83**, 659–669. 36

Kollmeier B. (**1996**)
Computer-controlled speech audiometric techniques for the assessment of hearing loss and the evaluation of hearing aids. In *Psychoacoustics, Speech and Hearing Aids*, edited by B. Kollmeier, World Scientific, Singapore, pp. 57–68. 40

Kollmeier B. (**1997a**)
Aktuelle und zukünftige Entwicklungen. In *Versorgung und Rehabilitation mit Hörgeräten*, edited by J. Kießling, B. Kollmeier, and G. Diller, Georg Thieme Verlag, Stuttgart, pp. 111–131, ISBN 3-13-106821-3. 8

Kollmeier B. (**1997b**)
Hörflächenskalierung — Grundlagen und Anwendung der kategorialen Lautheitsskalierung für Hördiagnostik und Hörgeräteversorgung, volume 2. Median-Verlag, Heidelberg. 16, 35, 62

Kollmeier B. and Wesselkamp M. (**1997**)
Development and Evaluation of a German Sentence Test for objective and subjective Speech Intelligibility Assessment. J. Acoust. Soc. Am. **102(4)**, 2412–2421. 17

Kollmeier B., Müller C., Wesselkamp M., and Kliem K. (**1992**)
Weiterentwicklung des Reimtests nach Sotscheck. In *Moderne Verfahren der Sprachaudiometrie, Buchreihe Audiologische Akustik*, edited by B. Kollmeier, median–verlag, Heidelberg, pp. 216–237, ISBN 3–922766–15–3. 90

Launer S. (**1995**)
Loudness Perception in Listeners with Sensorineural Hearing Impairment. Ph.D. thesis, Universität Oldenburg. 6, 42, 46, 51, 52, 56, 59, 60, 61, 62, 63, 66, 69, 70, 73, 74, 78, 84, 85, 86, 97, 101, 105

Launer S., Holube I., Hohmann V., and Kollmeier B. (**1996**)
Categorical loudness scaling in hearing-impaired listeners - Can loudness growth be predicted from the audiogram?. Audiologische Akustik **35(4)**, 156–163, ISSN 0172–8261. 6, 52, 60, 74, 84, 105

Launer S., Hohmann V., and Kollmeier B. (**1997**)
Modeling loudness growth and loudness summation in hearing-impaired listeners. In *Modeling sensorineural hearing loss.*, edited by W. Jesteadt, Lawrence Erlbaum Assoc., Mahwah, NJ, pp. 175–185. 6, 84, 93

Levitt H. (**1968**)
Testing for sequential Dependencies. J. Acoust. Soc. Am. **43(1)**, 65–69. 35, 47

Levitt H. (**1971**)
Transformed up–down methods in psychoacoustics. J. Acoust. Soc. Am. **49(2)**, 467–477. 35, 37

Lippmann R.P., Braida L.D., and Durlach N.I. (**1981**)
Study of multi-channel amplitude compression and linear amplification for persons with sensorineural hearing loss. J. Acoust. Soc. Am. **69**, 524–534. 8

Lunner T., Hellgren J., Arlinger S., and Elberling C. (**1998**)
Non-linear signal processing in digital hearing aids. Scan. Audiol. **27**, 40–49. Suppl. 49. 30

Maré M.J., Dreschler W.A., and Verschuure H. (**1992**)
The effects of input–output configuration in syllabic compression on speech perception. J. Speech Hear. Res. **35**, 675–685. 32

Marks L.E. (**1994**)
"Recalibrating" the auditory system: the perception of loudness. J. Exp. Psychol. Hum. Percept. Perform. **20(2)**, 382–396. 34

Marzinzik M. (**1996**)
Dynamikkompression für Hörgeräte basierend auf einem psychoakustischen Modell der Lautheitswahrnehmung. Master's thesis, Universität Oldenburg. 51

Marzinzik M., Appell J.E., Hohmann V., and Kollmeier B. (**1996a**)
Evaluation of dynamic compression algorithms using a loudness model for hearing impaired listeners. In *Psychoacoustics, Speech and Hearing Aids*, edited by B. Kollmeier, World Scientific, Singapore, pp. 203–206. 84

Marzinzik M., Hohmann V., Appell J.E., and Kollmeier B. (**1996b**)
Zur Modellierung der Lautheitswahrnehmung bei Normalhörenden und Innenohrschwerhörigen. Audiologische Akustik **35(4)**, 136–144. 6, 51, 52, 61, 63, 66, 69, 71, 73, 74, 75, 80, 83, 84, 85, 104

Marzinzik M., Hohmann V., Appell J.E., and Kollmeier B. (**1997**)
Evaluation of different multi-channel dynamic compression algorithms with regard to recruitment compensation, quality and speech intelligibility. In *Seventh Oldenburg Symposium on Psychological Acoustics*, edited by A. Schick and M. Klatte, BIS, Universität Oldenburg, volume 7, pp. 619–630. 8

Moore B.C.J. (**1993**)
Frequency analysis and pitch perception. In *Human Psychophysics*, edited by W.A. Yost, A.N. Popper, and R.R. Fay, Springer-Verlag, New York, pp. 56–115. 57, 59

Moore B.C.J. (**1995**)
Perceptual Consequences of Cochlear Damage. Oxford University Press, Oxford. 5, 34, 52, 59, 60, 61, 74, 84, 85, 93

Moore B.C.J. (**1999**)
Use of a loudness model for hearing aid fitting: III. A general method for deriving initial fittings for hearing aids with multi-channel compression. Brit. J. Audiol. **33(4)**, 241–258, ISSN 0300-5364. 84

Moore B.C.J. (**2000**)
Use of a loudness model for hearing aid fitting. IV. Fitting hearing aids with multi-channel compression so as to restore 'normal' loudness for speech at different levels. Brit. J. Audiol. **34(3)**, 165–177, ISSN 0300-5364. 84

Moore B.C.J. and Glasberg B.R. (**1987**)
Formulae describing frequency selectivity as a function of frequency and level and their use in calculating excitation patterns. Hear. Res. **28**, 209–225. 57, 70, 85

Moore B.C.J. and Glasberg B.R. (**1996**)
A revision of Zwicker's loudness model. Acustica **82(2)**, 335–345, ISSN 0001-7884. 5, 6, 51, 52, 56, 57, 59, 61, 66, 70, 71, 74, 76, 80

Moore B.C.J. and Glasberg B.R. (**1997**)
A model of loudness perception applied to cochlear hearing loss. Auditory Neuroscience **3**, 289–311, ISSN 1023-618X. 6, 52, 74, 87

Moore B.C.J. and Glasberg B.R. (**1998**)
Use of a loudness model for hearing-aid fitting. I. Linear hearing aids. Brit. J. Audiol. **32(5)**, 317–335, ISSN 0300-5364. 84

Moore B.C.J., Johnson J.S., Clark T.S., and Pluvinage V. (**1992**)
Evaluation of a dual–channel full dynamic range compression system for people with sensorineural hearing loss. Ear and Hear. **13(5)**, 349–370. 35

Moore B.C.J., Glasberg B.R., and Vickers D.A. (**1996**)
Factors influencing loudness perception in people with cochlear hearing loss. In *Psychoacoustics, Speech and Hearing Aids*, edited by B. Kollmeier, World Scientific, Singapore, pp. 7–18. ISBN 981022561X. 6, 52, 74, 85, 87

Moore B.C.J., Huss M., and Baer T. (**1997**)
A model for the prediction of thresholds, loudness, and partial loudness. J. Audio Eng. Soc. **45(4)**, 224–240, ISSN 0004-7554. 6, 52, 80

Moore B.C.J., Alcantara J.I., Stone M.A., and Glasberg B.R. (**1999a**)
Use of a loudness model for hearing aid fitting: II. Hearing aids with multi-channel compression. Brit. J. Audiol. **33(3)**, 157–170, ISSN 0300-5364. 84

Moore B.C.J., Glasberg B.R., and Vickers D.A. (**1999b**)
Further evaluation of a model of loudness perception applied to cochlear hearing. J. Acoust. Soc. Am. **106(2)**, 898–907. 6, 52, 74, 83, 84, 85, 87, 94, 104

Moore B.C.J., Peters R.W., and Stone M.A. (**1999c**)
Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. J. Acoust. Soc. Am. **105(1)**, 400–411. 8, 30

Moore B.C.J., Vickers D.A., Plack C.J., and Oxenham A.J. (**1999d**)
Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism. J. Acoust. Soc. Am. **106(5)**, 2761–2778, ISSN 0300-5364. 84, 105

Moore B.C.J., Huss M., Vickers D.A., Glasberg B.R., and Alcantara J.I. (**2000**)
A test for the diagnosis of dead regions in the cochlea. Brit. J. Audiol. **34(4)**, 205–224, ISSN 0300-5364. 6, 95

Moore B.C.J., Huss M., Vickers D.A., and Baer T. (**2001**)
Psychophysics of dead regions. In *Physiological and Psychophysical Bases of Auditory Function - Proceedings of the 12th International Symposium of Hearing*, edited by D.J. Breebaart, A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs, and R. Schoonhoven, Shaker, Maastricht, ISBN 90-423-0115-5. 6, 85

Nábělek J.V. (**1983**)
Performance of hearing–impaired listeners under various types of amplitude compression. J. Acoust. Soc. Am. **74(3)**, 776. 8

Neuman A.C., Bakke M.H., Mackersie C., Hellman S., and Levitt H. (**1995**)
Effect of releasetime in compression hearing aids: Paired comparison judgement of quality. J. Acoust. Soc. Am. **98(6)**, 3182–3187. 8

Pascoe D.P. (**1978**)
An approach to hearing aid selection. Hear. Instr. **29**, 12–16. 35, 62

Patterson R.D. and Moore B.C.J. (**1986**)
Auditory filters and excitation patterns as representations of frequency resolution. In *Frequency Selectivity in Hearing*, edited by B.C.J. Moore, Academic, London, pp. 123–177. 57, 70, 85

Patterson R.D., Nimmo-Smith J., Holdsworth J., and Rice P. (**1987**)
An efficient auditory filterbank based on the gammatone function. Paper presented at a meeting of the IOC Speech Group on Auditory Modelling at RSRE. 57, 70, 85

Paulus E. and Zwicker E. (**1972**)
Programme zur automatischen Bestimmung der Lautheit aus Terzpegeln oder Frequenzgruppen-Pegeln. Acustica **27**, 253–266. 6, 52

Plomp R. (**1988**)
The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation transfer function. J. Acoust. Soc. Am. **83**, 2322–2327. 8, 12

Plomp R. and Mimpen A.M. (**1979**)
Speech-reception threshold for sentences as a function of age and noise level. J. Acoust. Soc. Am. **66**, 1333–1342. 84

Port E. (**1963**)
Über die Lautstärke einzelner kurzer Schallimpulse. Acustica **13**, 212–223. 41, 42, 47

Poulton E.C. (**1989**)
Bias in Quantifying Judgements. Lawrence Erlbaum, Hillsdale, New Jersey. 35, 47

Press W.H., Flannery B.P., Teukolsky S.A., and Vetterling W.T. (**1992**)
Numerical Recipes in C. Cambridge University Press, Cambridge, 2nd edition. 60, 70, 73, 74, 95

Pumplin J. (**1985**)
Low–noise noise. J. Acoust. Soc. Am. **78(1)**, 100–104. 34, 36

Puria S., Peake W.T., and Rosowski J.J. (**1997**)
Sound–pressure measurements in the cochlear vestibule of human–cadaver ears. J. Acoust. Soc. Am. **101(5)**, 2754–2770, ISSN 0004-7554. 80

Raß U. (**1996**)
A wearable signal processor system for the evaluation of digital hearing aid algorithms. In *Psychoacoustics, Speech and Hearing Aids*, edited by B. Kollmeier, World Scientific, Singapore, pp. 273–276. ISBN 981022561X. 109

Raß U. and Steeger G. (**2000**)
A high performance pocket–size system for evaluations in acoustic signal processing. Acustica **86(2)**, 374–375, ISSN 0001-7884. 9, 14

Reckhardt C. (**2000**)
Factors influencing equal–loudness level contours. Ph.D. thesis, Universität Oldenburg. 67, 71, 80, 86, 97

Rosowski J.J. (**1991**)
The effects of external- and middle-ear filtering on auditory threshold and noise-induced hearing loss. J. Acoust. Soc. Am. **90(1)**, 124–135. 57

Sachs L. (**1991**)
Angewandte Statistik: Anwendung statistischer Methoden. Springer Verlag, Berlin u. a. ISBN 3–540–52085–6.

Scharf B. (**1961**)
Loudness summation under masking. J. Acoust. Soc. Am. **64**, 1036–1040. 42, 47

Serpanos Y.C., O'Malley H., and Gravel J.S. (**1997**)
The relationship between loudness intensity functions and the click-ABR wave V latency. Ear and Hear. **18(5)**, 409–419. 62

Steinberg J.C. and Gardner M.B. (**1937**)
The dependency of hearing impairment on sound intensity. J. Acoust. Soc. Am. **9**, 11–23. 5, 8, 34

Stevens S.S. (**1953**)
On the brightness of light and the loudness of sounds. Science **118**, 576. 6

Stevens S.S. (**1955**)
The measurement of loudness. J. Acoust. Soc. Am. **27**, 815–827. 42, 47

Stevens S.S. (**1956**)
Calculation of loudness of Complex Sounds. J. Acoust. Soc. Am. **28**, 807–832. 5, 6

Stevens S.S. (**1957**)
On the psychophysical law. Psychol. Rev. **64**, 153–181. 5, 52, 53

Stevens S.S. (**1966**)
Power–group transformations under glare, masking and recruitment. J. Acoust. Soc. Am. **9**, 11–23. 53

Stone M.A., Moore B.C.J., Wojtczak M., and Gudgin E. (**1997**)
Effects of fast–acting high–frequency compression on the intelligibility of speech in steady and fluctuating background sounds. Brit. J. Audiol. **31(4)**, 257–273, ISSN 0300-5364. 30

Stone M.A., Moore B.C.J., Alcantara J.I., and Glasberg B.R. (**1999**)
Comparison of different forms of compression using wearable digital hearing aids. J. Acoust. Soc. Am. **106(6)**, 3603–3619, ISSN 0300-5364. 8

Suzuki S., Suzuki Y., Kono S., Sone T., Kumagai M., Miura H., and Kado H. (**1989**)
Equal-loudness level contours for pure tones under free field listening condition (I) - Some data and considerations of experimental conditions. J. Acoust. Soc. Jap. **-**, 329–338. 67, 71, 73, 80, 97, 105

Tejero–Calado J.C., Rutledge J.C., and Nelson P.B. (**1998**)
Combination Compression and Linear Gain Processing for Digital Hearing Aids, IEEE, Hong Kong. 8

Tyler R.S. (**1986**)
Frequency resolution in impaired listeners. In *Frequency Selectivity in Hearing*, edited by B.C.J. Moore, Academic Press, London, pp. 309–371. 85

Verhey J.L. (**1998**)
Psychoacoustics of spectro-temporal effects in masking and loudness patterns. Ph.D. thesis, Universität Oldenburg, BIS-Verlag Oldenburg. 34, 35, 42, 47

Verschuure J. and Dreschler W.A. (**1996**)
Dynamic compression hearing aids. In *Psychoacoustics, Speech and Hearing Aids*, edited by B. Kollmeier, World Scientific, Singapore, pp. 153–164, ISBN 981022561X. 8, 12

Verschuure J., Benning F.J., van Cappellen M., Dreschler W.A., and Boeremans P.P. (**1998**)
Speech Intelligibility in Noise with Fast Compression Hearing Aids. Audiology **29(3)**, 127–150. 30

Villchur E. (**1989**)
Comments on "The negative effect of amplitude Compression in Multichannel hearing aids in the light of the modulation transfer function". J. Acoust. Soc. Am. **86**, 425–427. 12

Walker G. and Byrne D. (**1984**)
The effect of multiband compression/expansion amplification on the intelligibility of nonsense syllables in noise. J. Acoust. Soc. Am. **76(3)**, 746. 8, 30

Watanabe T. and Møller H. (**1990**)
Hearing Thresholds and Equal Loudness Contours in Free Field at Frequencies below 1 kHz. J. of Low Frequency & Vibration **9(4)**, 135–148. 67, 71, 80, 97, 105

Weber E.H. (**1905**)
Tastsinn und Gemeingefühl, volume 149 of *Ostwalds Klassiker der exakten Wissenschaften*. Engelmann, Leipzig. 5

Wesselkamp M., Kliem K., and Kollmeier B. (**1992**)
Erstellung eines optimierten Satztests in deutscher Sprache. In *Moderne Verfahren der Sprachaudiometrie, Buchreihe Audiologische Akustik*, edited by B. Kollmeier, median–verlag, Heidelberg, pp. 330–343, ISBN 3–922766–15–3. 17

White M.W. (**1986**)
Compression Systems for Hearing Aids and Cochlear Prosthesis. J. Reh. Res. Dev. **23/1**, 25–39. 8

Working Group on Communication Aids for the Hearing–Impaired (**1991**)
Speech–perception aids for hearing–impaired people: Current status and needed research. J. Acoust. Soc. Am. **90**, 637–685. 8

Zwicker E. (**1958**)
Über psychologische und methodische Grundlagen der Lautheit. Acustica **8**, 237–258. 5, 6, 51, 52, 55, 56, 57, 58, 59, 62, 66, 69, 70, 83, 84, 104

Zwicker E. (**1960**)
Ein Verfahren zur Berechnung der Lautstärke. Acustica **10**, 304–308. 52

Zwicker E. (**1966**)
Ein Beitrag zur Unterscheidung von Lautstärke und Lästigkeit. Acustica **17**, 22–25. 47

Zwicker E. and Fastl H. (**1990**)
Psychoacoustics - Facts and Models. Springer-Verlag Berlin Heidelberg New York, ISBN 3-540-52600-5. 6, 34, 36, 46, 51, 52, 55, 59, 69, 70, 83, 84, 90, 104

Zwicker E. and Fastl H. (**1999**)
Psychoacoustics - Facts and Models. Springer-Verlag Berlin Heidelberg New York, 2nd edition. 52

Zwicker E. and Scharf B. (**1965**)
A model of loudness summation. Psychol. Rev. **72**, 3–26. 52

Zwicker E. and Terhardt E. (**1980**)
Analytical expressions for critical–band rate and critical bandwidth as a function of frequency. J. Acoust. Soc. Am. **68(5)**, 1523–1525. 57

Zwicker E., Flottorp G., and Stevens S.S. (**1957**)
Critical bandwidth in loudness summation. J. Acoust. Soc. Am. **29**, 548–557. 41

Zwislocki J.J. (**1965**)
Analysis of some auditory characteristics. In *Handbook of Mathematical Psychology*, edited by R.D. Luce, R. Bush, and E. Galanter, John Wiley, New York, volume III. 5, 53

Zwislocki J.J. (**1975**)
The role of the external and middle ear in sound transmission. In *The Nervous System, Vol. 3. Human Communication and its Disorders*, edited by D.B. Tower and E.L. Eagles, Raven Press, New York, pp. 45–55. 57

Index

Erklärung

Hiermit erkläre ich, daß ich die vorliegende Arbeit selbstständig verfaßt und keine anderen als die angegebenen Hilfsmittel benutzt habe.

Oldenburg, den 10. August 2001

Jens-Ekkehart Appell

Danksagung

Mein herzlicher Dank gilt Prof. Dr. Dr. Birger Kollmeier für die Ermöglichung dieser Arbeit, die ausgezeichneten Arbeitsbedingungen, die vielen konstruktiven und motivierenden Diskussionen, besonders in der Endphase der Arbeit, und für seine freundschaftliche und offene Art.

Meinem “Mitstreiter in Sachen Dynamikkompression”, Dr. Volker Hohmann möchte ich ganz besonders für die vielen Anregungen, ungezählten Diskussionen und immer neuen Ideen für ein Histogramm danken. Außerdem danke ich ihm, sowie Dr. Torsten Dau, Dr. Birgitta Gabriel, Dr. Thomas Brand und Dr. Jesko Verhey für das kritische Lesen dieser Arbeit.

Für die gute Zusammenarbeit in der Frage “Was ist gleich laut ?” danke ich Giso Grimm und Dr. Jesko Verhey.

Dr. Stefan Launer danke ich für den regen Austausch insbesondere zu Beginn der Arbeit.

Ebenso möchte ich Uwe Raß für seine Arbeiten an der Hard– und Software des tragbaren Echtzeithörgerätes danken.

Den Mitgliedern der Arbeitsgruppe Medizinische Physik danke ich für die gute Arbeitsatmosphäre, Diskussionsbereitschaft und die vielen “helping hands”, zum Beispiel an und in den Servern.

Den Mitarbeiterinnen und Mitarbeitern des Hörzentrum Oldenburg und dabei insbesondere Dr. Birgitta Gabriel und Kerstin Sommer danke ich für die Unterstützung bei vielen Messungen.

Anita Gorges und Bajo Meenen danke ich für Rat und Tat bei technischen Problemen.

Dr. Joachim Neumann, Dr. Torsten Dau, Dr. Ralf Peter Derleth und Helmut Riedel danke ich für das freundschaftliche “universitäre” Zusammenwohnen.

Danken möchte ich auch Stephan Albani, Dr. Daniel Berg und Dr. Thomas Wittkop für ihre vielfältige Unterstützung in allen Lebenslagen.

Ganz besonders danke ich Kira.

Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft, dem Bundesministerium für Bildung und Forschung, sowie der Europäischen Union gefördert.

Lebenslauf

Am 25. September 1966 wurde ich, Jens–Ekkehart Appell, als drittes Kind von Barbara Helene Appell, geb. Sostmann und Dr. Franz August Wilhelm Ehrhart Appell in Rotenburg a. d. Fulda geboren. Von 1972 bis 1976 besuchte ich die Schlotschule Melsungen (Grundschule), von 1976 bis 1982 die Gesamtschule Melsungen (Orientierungs– und Mittelstufe) und von 1982 bis 1985 das Geschwister-Scholl-Gymnasium in Melsungen.

Von September 1985 bis April 1987 leistete ich meinen Zivildienst beim Jugendwerk der Arbeiterwohlfahrt Bezirksverband Hessen–Nord in Kassel.

Im April 1987 nahm ich das Philosophiestudium an der Georg–August–Universität in Göttingen auf und wechselte zum Wintersemester in den Diplomstudiengang Physik. Mein Vordiplom in Physik legte ich am 26. April 1990 ab. Von Juli 1992 bis Juli 1994 fertigte ich am Dritten Physikalischen Institut in Göttingen, sowie in der Arbeitsgruppe Medizinische Physik von Prof. Dr. Dr. Kollmeier an der Carl von Ossietzky Universität meine Diplomarbeit mit dem Thema "Simulation, Anpassung und Test von dreikanaligen Signalverarbeitungsstrategien für Hörgeräte" an und legte am 1. Juli 1994 die Diplomprüfung in Physik an der Georg–August–Universität in Göttingen ab.

Von Juli 1994 bis July 2001 arbeitete ich als wissenschaftlicher Mitarbeiter in der Arbeitsgruppe Medizinische Physik unter anderem in Projekten zur Entwicklung moderner Hörgerätealgorithmen (Drittmittelprojekte des BMBF und der EU), sowie in Projekten zu Experimenten und Modellen für die Lautheitswahrnehmung mit. Dort fertigte ich unter Anleitung von Prof. Dr. Dr. Kollmeier die vorliegende Dissertation an.

Neben dieser Tätigkeit habe ich in mehreren Industrieprojekten des Hörzentrum Oldenburg als Softwareentwickler, wissenschaftlicher Berater und Projektleiter gearbeitet.