The role of loudness perception in hearing aid fitting

Temporal, spectral and binaural effects

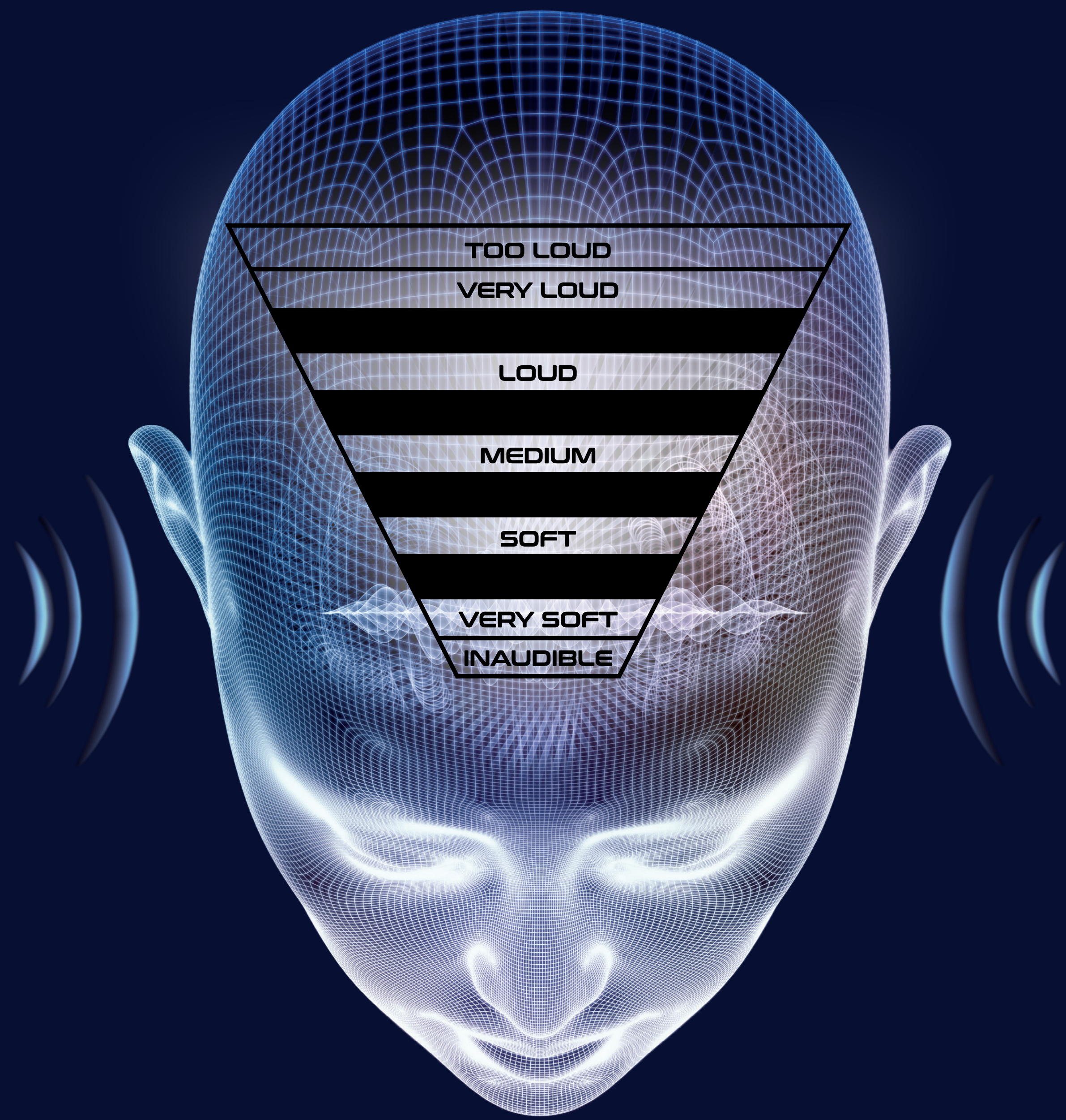
van Beurden, M.F.B.

Publication date
2023

Document Version
Final published version

The role of loudness perception in hearing aid fitting:

temporal, spectral and binaural effects



Maarten van Beurden

The role of loudness perception in hearing aid fitting: temporal, spectral and binaural effects
Maarten van Beurden

The role of loudness perception in hearing aid fitting:
temporal, spectral and binaural effects

Maarten van Beurden

Colofon

The role of loudness perception in hearing aid fitting: temporal, spectral and binaural effects – Maarten van Beurden
ISBN: 978-94-6419-769-3

Cover design by BLOCC creatieve communicatie
Printed by Gildeprint, Enschede

The role of loudness perception in hearing aid fitting:

temporal, spectral and binaural effects

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. P.P.C.C. Verbeek

ten overstaan van een door het College voor Promoties ingestelde commissie,

in het openbaar te verdedigen in de Agnietenkapel

op vrijdag 14 juli 2023, te 13.00 uur

door Maarten Franciscus Bernardus van Beurden

geboren te Hilvarenbeek

Promotiecommissie

Promotor	prof. dr. ir. W.A. Dreschler	AMC-UvA
Copromotor	dr. M. Boymans	Libra Revalidatie & Audiologie
Overige leden	prof. dr. W.J. Fokkens	AMC-UvA
	prof. dr. ir. J.C.M. Smits	AMC-UvA
	dr. ir. P. Brienesse	AMC-UvA
	prof. dr. ir. J.M. Festen	Vrije Universiteit Amsterdam
	prof. dr. P. van Dijk	Rijksuniversiteit Groningen
	dr. ir. J.A.P.M. de Laat	LUMC

Faculteit der Geneeskunde

Contents

Chapter 1

General introduction and outline of the thesis

1.1 Introduction

Our hearing is a fantastic instrument. If the ears function normally uncountable different sounds can be detected, discriminated, and identified at diverging sound levels, ranging from the faint whisper of the wind to the overwhelming sound of a departing airplane. Unfortunately, not everyone is blessed with normal hearing. Hearing impairment has a high prevalence with possibly severe consequences in daily life. The World Health Organization (WHO) defines disabling hearing loss as a hearing loss greater than 40 decibels (dB) in the better ear for adults and greater than 30 dB in the better ear for children. According to this definition over 5% of the world's population has disabling hearing loss (WHO, April 2021). This adds up to 430 million people. The WHO estimates that in 2050 over 700 million people will have disabling hearing loss.

Hearing refers to the perception of sounds or vibrations. In physics, sound is a vibration that propagates as an acoustic wave through a transmission medium, such as gas, liquid or solid (Wikipedia). The basic form of a sound is a sinusoidal wave that is defined by its frequency, amplitude and the speed of sound of the medium in which it travels. In principle every sound can be described as a combination of sinusoidal waves with different amplitudes, and frequencies. As such, sound is well-defined by physical quantities.

Our hearing, however, is not able to perform a perfect analysis of sound physics. The ability of our hearing to perceive sound, is influenced by the absolute threshold of hearing (e.g. the pure-tone audiogram), supra-threshold perception (e.g. frequency resolution, temporal resolution, compression), and binaural co-operation (e.g. directional hearing).

Furthermore, perception is by definition a subjective experience. Sound perception is thus not only determined by the physical quantities of a sound, but also by psychological characteristics of a listener, such as mood, expectation, experience, and attention. Therefore, psychoacoustics, the scientific study of sound perception, does not only deal with the physical characteristics of sound such as amplitude or frequency, but also with the perceptual responses to these physical quantities, such as loudness and pitch.

In this thesis, the focus is on loudness perception. Loudness is the subjective perception of sound pressure level. It is the attribute of sound on which sounds are ordered perceptually from quiet to loud. Loudness spans the entire dynamic range of hearing, from hearing threshold to the threshold of pain. Although loudness is mainly related to the sound pressure level of a sound, duration, bandwidth, masking and unilateral or bilateral presentation also influence loudness perception.

1.2 The physiology of loudness

To understand the properties of loudness, a basic understanding of the physiology of our hearing is needed. The human auditory system consists out of three main components, the outer ear, the middle ear, and the inner ear. A schematic overview of the human ear is shown in figure 1.1. The outer ear consists of the pinna and the ear canal, together with the middle ear that consists of the eardrum and the ossicles, the outer ear is responsible for transmission of sound to the inner ear.

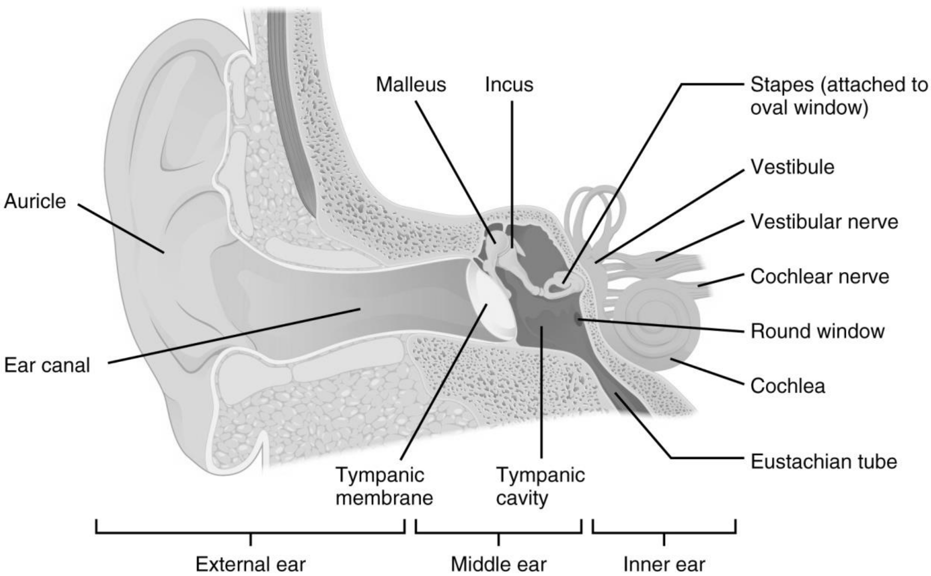


Figure 1.1. *Schematic representation of the human ear (http://commons.wikimedia.org with modifications).*

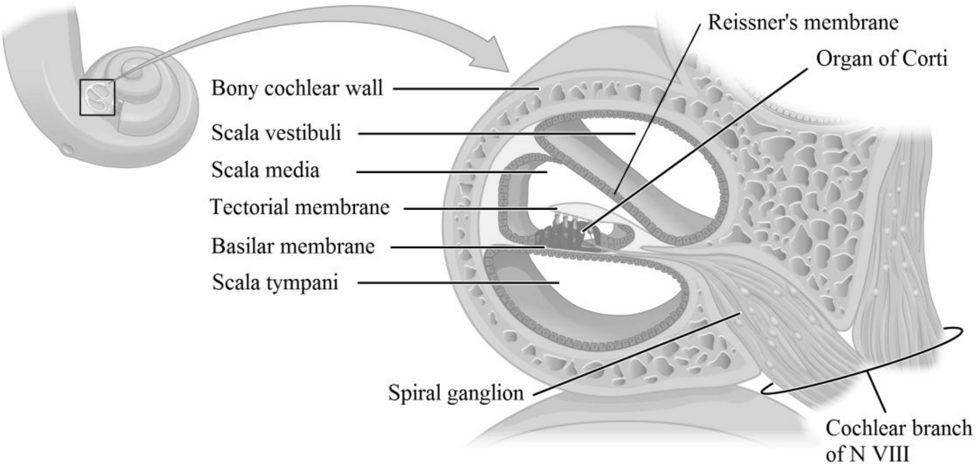


Figure 1.2. *Schematic representation of the cochlea (http://commons.wikimedia.org with modifications).*

The inner ear or cochlea is the part of the ear where vibration is transferred into neural activity and is believed to play an important role in loudness perception. The cochlea consists of a fluid-filled spiral-shaped hollow tube making 2.5 turns around its axis. The start of the cochlea is called the basal end, the tip of the spiral is called the apical end (or apex). In a cross-section of the cochlear tube, see figure 1.2, three chambers can be seen: the scala vestibuli, the scala media and the scala tympani. The chambers are separated from each other by two membranes. Reissner's membrane separates the scala vestibuli and the scala media and the basilar membrane separates the scala tympani and the scala media. Because the base of the basilar membrane is 5 times narrower and about 100 times stiffer than the apex, high frequency sounds cause greatest vibration near the base of the membrane, and low frequency sounds cause greatest vibration near the apex. This way, the basilar membrane performs a frequency-to-place conversion. On top of the basilar membrane lies the organ of Corti. The organ of Corti houses the primary sensory receptor cells known as hair cells.

The hair cells derive their name from the hair-like stereocilia extending from their apical end. Two types of hair cells can be distinguished: outer hair cells and inner hair cells. The inner hair cells are believed to be the main sensory cells transforming vibration into neural activity. The outer hair cells amplify the movement of the basilar membrane. This is called the cochlear amplifier. As a result of the cochlear amplifier the healthy cochlea is more sensitive to low-level sounds than a damaged cochlea. This leads to a highly compressive response. At mid-level input levels (30 to 80 dB SPL) a change in input sound pressure of 50 dB leads to a change of slightly less than 10 dB in the velocity of the basilar membrane, while in a damaged cochlea this relationship is essentially linear (e.g., Robles *et al.*, 1986; Yates, 1990). This compression implies that the gain of the cochlear amplifier at low levels is greater than its gain at high levels. The cochlear amplifier is also responsible for the sharpening of the frequency response of the cochlea (Oghalai, 2004; Oxenham and Bacon, 2003). Again, the response is nonlinear with sharper tuning at low levels. Cochlear damage reduces the level dependency of frequency tuning at the basilar membrane, making the response more linear.

The loudness function shows features similar to the mechanical input/output function measured at the basilar membrane (Florentine *et al.*, 1996). As is the case in the input/output curve of the basilar membrane, the loudness function grows more slowly at moderate levels than at low and high levels. Loudness therefore seems to be related to the mechanical excitation of the basilar membrane. The critical band, an important concept involved in spectral loudness summation, is also thought to correspond with a specific area on the basilar membrane. With one critical band assumed to cover a length of about 0.9 mm (Moore, 2003).

Although many properties of hearing such as intensity and frequency sensitivity can be largely explained by processes in the cochlea, the central auditory nervous system plays an important role in the processing of complex auditory signals. The central auditory nervous system is complex and does not only transmit information from the cochlea to the auditory cortex (the afferent system), but also transmits information from the auditory cortex back to the cochlea (the efferent system).

A schematic overview of the auditory pathway is shown in figure 1.3. The ascending pathway starts with the cochlear nerve that terminates at the cochlear nucleus.

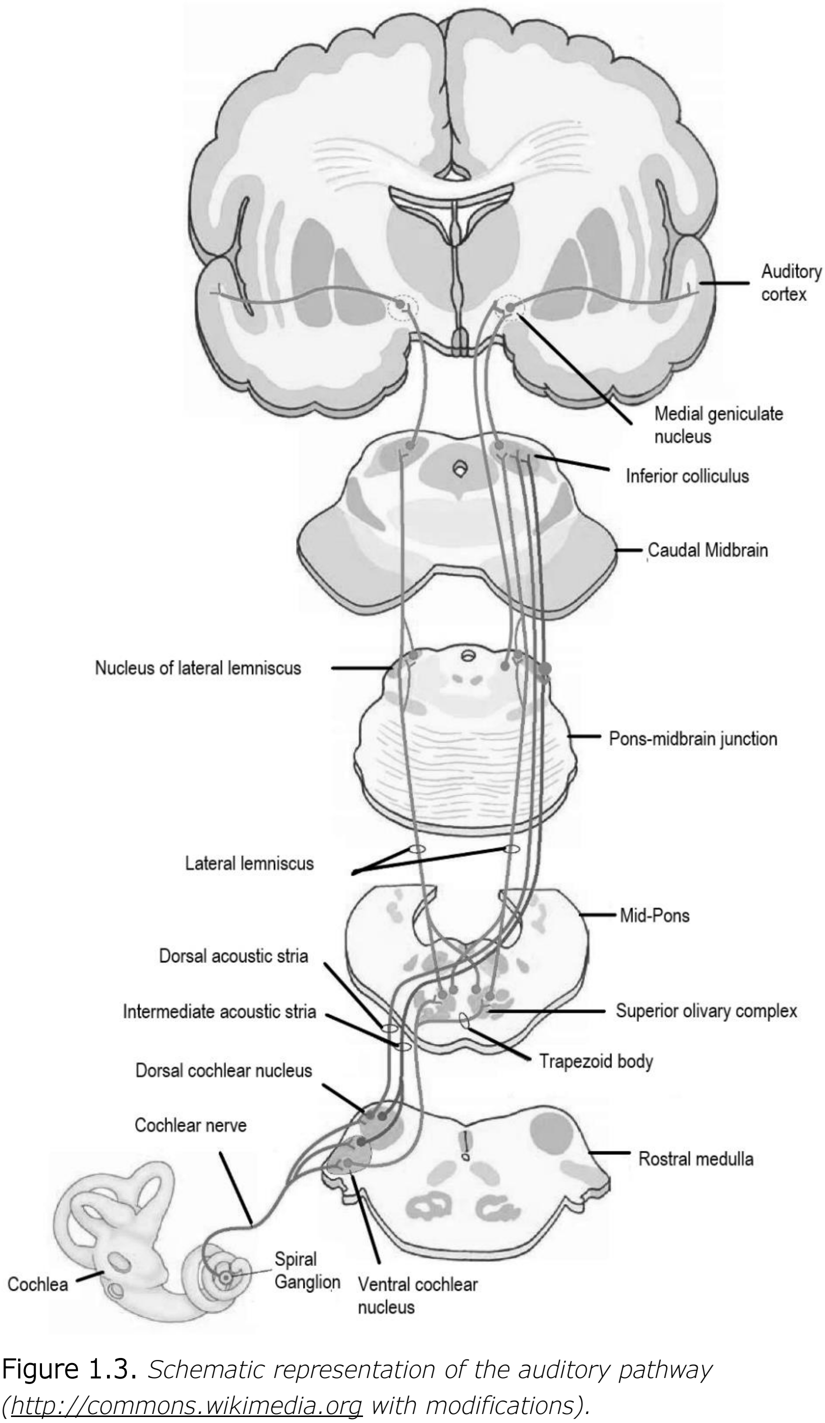


Figure 1.3. *Schematic representation of the auditory pathway (http://commons.wikimedia.org with modifications).*

At the cochlear nucleus each cochlear nerve fiber branches, one branch goes to the dorsal cochlear nucleus (DCN) and the other branch to the ventral cochlear nucleus (VCN). From the ventral cochlear nucleus some fibers pass across the midline to the cells of the superior olivary complex (SOC), whereas others connect with olivary cells at the same side. The superior olivary complex is therefore the first stage in the auditory nervous system that receives binaural information. Fibers from the dorsal cochlear nucleus cross the midline of the brain, where they are joined by the fibers from the ventral cochlear nuclei of both sides and from the olivary complex in the nuclei of the lateral lemniscus (nLL). Most of the fibers of the lemniscus end in the inferior colliculus (IC), but some fibers bypass the colliculus and end, together with the fibers from the colliculus at the medial geniculate nuclei (MGN). From this point the fibers project to the temporal lobe in the primary auditory cortex. About half of the fibers of the auditory pathways cross the midline while others ascend on the same side of the brain. Information from each ear is therefore represented in both the right and the left primary auditory cortex.

The representation of loudness in the central auditory nervous system has been investigated with functional magnetic resonance imaging (fMRI) (e.g.; Röhl *et al.* 2010; Röhl and Uppenkamp 2012; Schreiner and Malone, 2015). Röhl and Uppenkamp (2012) showed with fMRI that for normal hearing subjects, perceived loudness is reflected by corresponding neural activity in the auditory cortex, but not in the auditory brainstem. This illustrates that loudness perception is not complete before the level of the auditory cortex. Furthermore, Röhl *et al.* (2010) showed that while neural activation in the auditory brainstem reflects bandwidth in a linear fashion, a link between perceived loudness and neural activity could be solely observed in the primary auditory cortex.

In general, the efferent system may be regarded as providing a sort of feedback loop. The olivocochlear (OC) bundle, which arises from the olivary complex, is believed to be involved in modifying the analysis made in the cochlea. There are two groups of olivocochlear efferents: medial OC (MOC) and lateral OC (LOC) efferents. LOC efferents innervate primary auditory-nerve (AN) fibers, their function may be to reduce damage to AN fibers from excessive activation by traumatic sounds (Darrow *et al.*, 2007; Fuente, 2015). MOC efferents innervate the outer hair cells and act to turn down the gain of the cochlear amplifier (Guinan Jr., 2018). Of course, there is also the well-known feedback loop to the stapedius muscle, in which loud sounds lead to a contraction of the stapedius muscle in the middle ear, protecting the ear for loud sounds. In this feedback loop it is again at the level of the superior olivary complex, where the ascending information about the intensity of the sounds is fed back in a descending path towards the stapedius muscle.

1.3 Basic properties of loudness

Although loudness is mainly related to sound pressure level, other sound properties also influence loudness perception. In this paragraph the main influences of, frequency, duration, bandwidth, masking, bilateral presentation are briefly reviewed. First, however, a short introduction in loudness scales. Several loudness scales have been developed to relate sound pressure level and loudness. The two most well-known loudness scales are the phon scale (ISO 226, 2003) and the sone scale (Stevens, 1936, 1955, 1956). Both scales define the loudness of a reference signal to which the loudness of other sounds can be compared. For the phon scale, the reference is the intensity of a 1000 Hz pure tone in dB SPL. The intensity of the 1000 Hz tone at which the sound is perceived equally loud as the test signal is the loudness in phons. Thus, if a given sound is perceived equally loud as a 1000 Hz pure tone of 60 dB SPL, the loudness of that sound is 60 phon. On the phon scale a 10 phon change in loudness approximately corresponds roughly with a doubling of loudness. For some measurements a loudness scale with a more linear relationship between intensity and loudness is more appropriate.

The sone scale (Stevens, 1936, 1955, 1956) is developed to provide such a linear scale. In the sone scale 1 sone is equal to 40 phons. Contrary to the phon scale, the sone scale has only one reference point. The loudness of other sounds can be derived by comparing the sound to the reference signal of 1 sone. A sound that is perceived as twice as loud as the reference signal has a loudness of 2 sones, a sound half as loud as the reference sound has a loudness of 0.5 sone, and so on. The sone scale has especially been used in experiments on the relationship between the loudness of sound and sound pressure level. In a first approximation the loudness in sones is a power law function of the signal intensity, with an exponential of 0.3.

As noted before loudness is the subjective evaluation of sound pressure level and spans the entire dynamic range from auditory threshold to the threshold of pain. As our hearing is not equally sensitive at all frequencies, the relationship between loudness and sound pressure level is frequency dependent. This can be readily seen in a plot of the equal-loudness contours (figure 1.4). Fletcher and Munson (1933) were the first to measure equal-loudness contours with headphones. Test subjects compared pure tones at various frequencies to a reference tone at 1000 Hz. The test tones were provided with 10 dB increments in stimulus intensity and the intensity of the reference tone was adjusted until the test subjects perceived the same loudness as the test tone. The lowest equal-loudness contour represents the absolute threshold of hearing, the highest contour is the threshold of pain.

In practice it is not acceptable to measure the threshold of pain. Therefore, in clinical practice normally the threshold of uncomfortable loudness (UCL) is measured instead. The threshold of uncomfortable loudness can be defined in several different ways and the outcome values strongly depend on the instruction given to a subject. Mostly an ascending staircase is used and the subject is asked to indicate when the sound becomes uncomfortably loud. As the term uncomfortably loud is rather vague, often more elaborate instructions are added to specify this term. For instance, the British Society of Audiology recommends the following instructions: “I will gradually make the sound louder in your ear, and you must press the button (or raise your hand) as soon as the sound becomes uncomfortable (uncomfortably loud). This is not a test to find the loudest sound you can tolerate; it is a test to find what level of sound you find uncomfortable. You should press the button (or raise your hand) only when the sound becomes uncomfortable; but make sure you press (raise) it as soon as the sound reaches that level.”

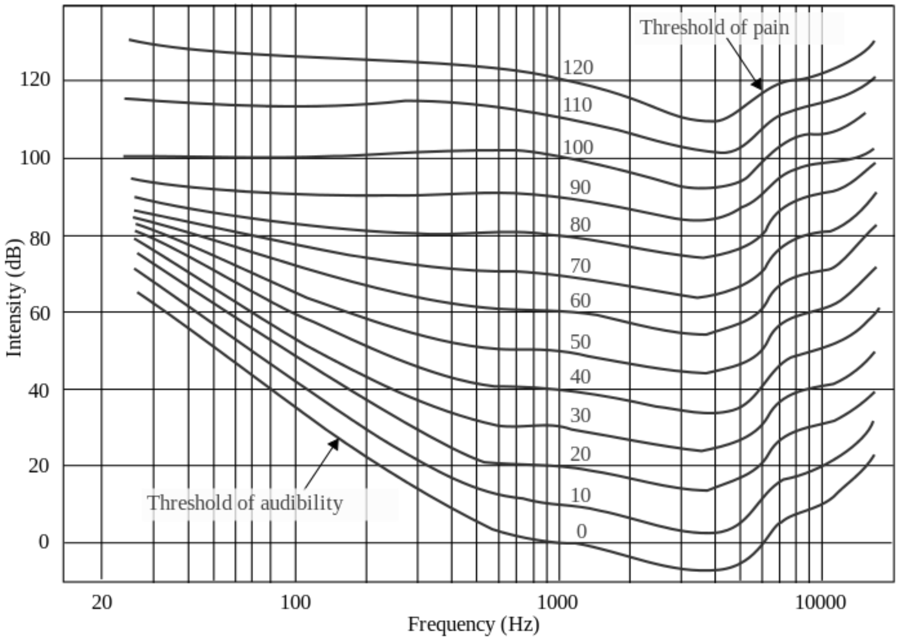


Figure 1.4. *Equal-loudness contours (http://commons.wikimedia.org with modifications).*

In this thesis we determined the UCL based on categorical loudness scaling measurements. In categorical loudness scaling a subject is instructed to indicate how loud a sound is based on several named categories ranging from “inaudible” to “too loud”, see paragraph 1.4. In this case the threshold of uncomfortable loudness is defined as the level at which the sound is judged as too loud. The threshold of uncomfortable loudness normally lies close to the threshold of pain.

Since the first measurements of equal-loudness contours by Fletcher and Munson, the equal-loudness contours have been revised several times. The most recent version of the equal-loudness contours is described in ISO 226 (2003). The equal-loudness contours clearly show that loudness varies with frequency as the equal-loudness contours are not flat lines.

Loudness has been shown to increase with signal duration for durations up to a time constant of 100 – 200 ms. This is called temporal integration (Munson, 1947; Stephens, 1973; Buus *et al.*, 1997). The amount of temporal integration depends on level (Buus *et al.*, 1997) with larger effects at moderate levels than at low and high levels. The equal-loudness-ratio-hypothesis (ELRH) proposed by Florentine *et al.* (1996) states further that the loudness ratio between equal-level long and short signals with the same spectrum is independent of level and spectrum.

The effect of bandwidth depends on spectral filtering in the ear. The loudness of a narrowband sound is determined by the excitation within an auditory channel (critical band) centered at the center frequency of the sound. A broadband sound excites several auditory channels. The loudness of the broadband signal is calculated by summing the loudness estimates in each auditory channel. Since the auditory system is normally compressive, the sum yields a larger overall loudness than the overall loudness of a sound which has the same intensity but excites only one auditory channel. This effect that at equal sound pressure level a broadband sound is perceived louder than a narrowband sound is called spectral loudness summation (e.g., Zwicker *et al.*, 1957; Scharf 1959; Cacace and Margolis, 1985).

In daily life most sounds are not perceived in silence but are accompanied by other sounds. In that case the perception of that sound can be influenced by the other sound, or sounds. This is called auditory masking. Masking can occur in the frequency domain and in the temporal domain. In the frequency domain masking occurs when the excitation patterns of two sounds overlap. An excitation pattern shows how much energy comes through each auditory filter and is thought to resemble the pattern of vibration on the basilar membrane. (Moore and Glasberg, 1983). A normal excitation pattern has a steep slope at the low frequency side and a shallower slope at the high frequency side. As a result, low frequency sounds mask high frequency sounds better than the other way around, this is called upward spread of masking. The amount of spectral masking depends on several factors, such as the frequency resolution of the subject’s hearing, the frequency difference between the sounds and the intensity of the sounds. Hearing-impaired listeners normally show broader than normal excitation patterns, due to broadening of the auditory filters (Glasberg and Moore, 1986).

Temporal masking occurs when a sound obscures a sound immediately preceding the sound (backward masking), or immediately following the sounds (forward masking). The amount of masking attenuates exponentially from the onset or offset of a sound, with backward masking lasting approximately 20 ms and forward masking lasting approximately 100 ms. Forward masking has been extensively investigated, as it is believed that forward masking is closely related to cochlear processing. A special case of temporal masking is the temporal effect. The temporal effect refers to the finding that a higher signal-to-masker ratio may be needed to detect a brief signal at the onset of a masker than if it is presented with an onset delay. The loudness of a (partially) masked signal is lower than the loudness of an unmasked signal of the same intensity.

A sound presented simultaneously to both ears (bilateral presentation) is normally perceived to be louder than the same sound presented to one ear (unilateral presentation). This effect is called binaural summation (for a review see: Sivonen and Ellermeier, 2011). It is currently believed that binaural loudness summation is not complete, that is, the ratio between binaural loudness perception (two ears) and monaural loudness perception (one ear) is less than 2 (Zwicker and Zwicker, 1991; Whilby *et al.* 2006; Marozeau *et al.* 2006).

1.4 Measurements procedures for loudness

Measuring loudness is not straightforward. Many different measurement procedures have been developed to measure the loudness of sound. In principal these measurement procedures can be grouped into two categories; loudness rating, and loudness matching.

<u>Loudness rating procedures</u> lead to loudness growth functions (Marks and Florentine, 2011). Examples of loudness rating procedures are magnitude estimation, magnitude production, and loudness scaling. In magnitude estimation a sound is presented to a listener, and the listener is asked to assign a number to the perceived loudness of that sound. Magnitude estimation is often conducted by cross-modality scaling in which the loudness of a sound is related to the length of a string, or the brightness of a light. In magnitude production the listener is asked to adjust the level of a sound until

it stands in some prescribed relation to a reference stimulus. For instance, a listener is asked to adjust the intensity of a sound in order to make it twice as loud, or half as loud as the reference stimulus.

In categorical loudness scaling (Brand and Hohmann, 2002; Elberling, 1999) the loudness of a sound is not related to a specific number or reference signal. Instead the listener is given a fixed set of terms (response categories) ranging from "inaudible" to "too loud". The response categories are usually assigned meaningful names, such as "very soft", ''medium", and "very loud", thereby relating to a subject's listening experience. As a result, loudness scaling requires minimal training. Another advantage is that it efficiently can cope with a large range of different intensities. The drawback of loudness scaling is that it is less suitable for measuring small differences as the response alternatives are limited and the results are often analyzed based on a fit through the data.

<u>Loudness matching</u> can be regarded as a special case of magnitude production in which the aim is to obtain equal loudness between the variable stimulus and the reference stimulus. Loudness matching procedures are often used to investigate loudness differences between sounds with different signal characteristics. Loudness matching is often performed with a two-interval, two-alternative forced-choice paradigm. This means that a listener is presented with two stimuli in a sequence with a pause between them and is forced to choose between the two stimuli based on a certain criterium. For instance, the listener is asked to indicate which of the two stimuli is the loudest. A forced choice procedure needs a decision strategy to determine the presentation level of the next trial. For an overview of modern procedures, see Leek (2001). The most applied method is the staircase procedure. In this procedure the level of the test stimulus depends on one or more of the previous responses. In a simple up-down staircase the stimulus level is increased when a response is negative (the test stimulus is softer) and decreased when a response is positive (the test stimulus is louder). The simple one-up one-down procedure targets the 50% performance level. Levitt (1992) also described transformed up-down procedures that target other points on the psychometric function. For instance, a two-up, one-down procedure targets the 70,7% level on the psychometric function. The estimate of the target point on the psychometric function is commonly based

on the average of the levels of a specified number of level-reversals. A forced choice procedure is an adaptive procedure as the level of the test stimulus is adapted based on the previous trials. Normally also the step size (the amount of difference between stimulus values) is decreased during a track to improve the accuracy.

Both in loudness matching as in loudness rating the ordering of the presentation levels can have a substantial impact on the results (Silva and Florentine, 2006; Brand and Hohmann, 2002; Jenstad *et al.* 1997). Bias may occur due to the tendency of listeners to involve experience in earlier trials in their loudness judgment. To force listeners to base their judgment on the stimulus on hand randomization and interleaving procedures have been developed. With randomization stimulus levels are presented in random order across blocks of trials. Randomization is used in experiments where a wide range of levels are measured to prevent order effects that can occur when levels are presented in increasing or decreasing order (i.e.: Mapes-Riordan and Yost, 1999). Interleaving means that trials of several stimulus pairs are measured in the same block. As randomization, interleaving prevents order effects.

1.5 Loudness models

Because loudness is one of the main attributes of the perception of sound, the prediction of loudness of specific sounds is important for hearing aid fitting, and for regulation and legislation regarding sound exposure in environmental, occupational and recreational settings. For this purpose, models have been developed to predict loudness in both normal hearing and hearing-impaired listeners. Fletcher and Munson (1933) were the first to calculate the loudness of steady sounds from the intensities of its frequency components. Zwicker and co-workers (Zwicker, 1958; Zwicker and Scharf, 1965; Zwicker *et al.*, 1984; Zwicker and Fastl, 1990) developed a more elaborate model that can account for a variety of data on loudness perception. Glasberg and Moore (Moore and Glasberg, 1996; Moore *et al.*, 1997; Glasberg and Moore, 2002) revised and extended Zwicker's model. The models of Zwicker and Glasberg and Moore consist of several stages that follow more or less the path of sounds traveling through the auditory system. The first stage describes the transfer through the outer and middle ear. The second stage describes the calculation of the excitation pattern from the effective spectrum reaching the cochlea. This stage is related to the basilar membrane mechanics in the cochlea. The basilar membrane reacts for specific frequencies particularly strong at specific points along the membrane. This tonotopic organization of the sensitivity to frequency along the basilar membrane provides a high frequency selectivity of the cochlea and can be modeled by a filter bank. Zwicker based the filter bank on measurements of critical bands, Glasberg and Moore on measurements of auditory filters. The excitation patterns show how much energy is transmitted through each filter in the filter bank. In the next stage the excitation patterns are transformed from excitation level to specific loudness, which is the loudness per critical band, or per auditory filter. Finally, the specific loudnesses in all bands/filters are summed to provide the instantaneous loudness.

Most of the models described above were developed for steady sounds and can correctly predict many loudness characteristics such as the loudness growth function, equal-loudness contours, and spectral loudness summation. In daily life most sounds are not steady sounds but are dynamic sounds. Therefore, also dynamic loudness models have been developed that can account for temporal integration data. For instance, Glasberg and Moore (2002) added an extra stage to their model to make the model applicable to time-varying sounds. In this stage they calculated short-term and long-term loudness from the instantaneous loudness using averaging mechanisms similar to an automatic gain control system, with attack and release times. Another model designed for dynamic signals is the Dynamic Loudness Model by Chalupper and Fastl (2002). This model roughly follows the same steps as the model by Glasberg and Moore, except that it encompasses an extra dynamic stage after the calculation of the excitation patterns that models the effects of forward masking. In subsequent versions of this Dynamic Loudness Model temporal components have been introduced at several stages of the model (Rennies *et al.,* 2009).

Except for the Dynamic Loudness Model by Chalupper and Fastl (2002) most loudness models assume little interaction between spectral and temporal effects. However, interactions between temporal and spectral characteristics of sounds have been shown. Verhey and Kollmeier (2002) measured spectral loudness summation in normal hearing listeners for short (10 ms) and long (1000 ms) signals. At the same reference level larger loudness summation was found for short than for long signals. These results are supported by Fruhmann *et al.* (2003) and Anweiler and Verhey (2006). A similar interaction between the spectral and temporal characteristics of a signal has been observed in experiments with modulated signals (i.e., Moore *et al.*, 1999; Grimm *et al.*, 2002; Gockel *et al.*, 2003).

1.6 Loudness and hearing loss

Loudness is an important perceptual characteristic of sound. At both sides of the scale (either sounds that are too soft or sounds that are too loud), sounds can easily give rise to complaints in daily life. This holds for normal hearing listeners, but especially for hearing-impaired listeners.

In most hearing-impaired listeners the dynamic range of hearing is reduced as hearing threshold is elevated by the hearing loss, while the threshold of pain is normally not or only slightly elevated (Kamm *et al.*, 1978; Shapiro, 1979). In combination with tinnitus the threshold of pain can even be reduced (Sanchez *et al.*, 2016). As a result, most hearing-impaired listeners show loudness recruitment, loudness growth increases at a higher rate than normal. This change in loudness growth is normally strongest at levels near hearing threshold. Near the threshold of pain loudness growth is generally comparable with that in normal hearing listeners. However, the exact form of the loudness function near threshold in case of hearing loss is still under debate and appears to differ significantly between individual listeners.

Marozeau and Florentine (2007) reanalyzed data from the literature and found that individual loudness-growth functions encompass a wide range of shapes. Some loudness growth functions correspond to the rapid growth hypothesis. In this hypothesis the loudness at threshold is the same for hearing-impaired listeners and normal hearing listeners, and from threshold to midlevels loudness grows more rapidly for hearing-impaired listeners, than for normal hearing listeners. At high levels loudness is the same or approaches that of normal hearing listeners. Other loudness growth functions correspond to the softness imperception hypothesis (Florentine, 2003). In this hypothesis loudness at threshold is higher for hearing-impaired listeners than for normal hearing listeners, but loudness growth at and near threshold is similar for hearing-impaired and normal hearing listeners. At midlevels loudness growth is faster-than-normal for hearing-impaired listeners and at higher levels the loudness growth function approaches that of a normal hearing listener again.

1.7 Consequences for hearing aid fitting

Altered loudness perception can have a severe impact on a subject's quality of life. If soft sounds become inaudible, detection of warning signals and speech understanding can be reduced. If loud sounds become uncomfortable too easily, one may tend to avoid such sounds. Both problems can limit participation in daily life. Therefore, one of the major aims in hearing aid fitting is loudness normalization. Loudness normalization is the restoration of the loudness perception of a hearing-impaired listener to the loudness perceived by a typical normal hearing listener. Several gain prescription rules have been developed to fulfill this aim. The two most well-known are NAL-NL2 (Dillon, 2012) and DSL I/O (Bagatto *et al.,* 2005; Scollie *et al.,* 2005). These two prescription rules do not deal with loudness normalization in the same way. DSL I/O performs true loudness normalization (Cornelisse *et. al,.*1995; Scollie *et al.*, 2005). That is, it aims to normalize loudness per frequency band. NAL-NL2 does not normalize loudness per frequency band but equalizes loudness for the mid-frequencies (called loudness equalization). The difference between the two prescription rules is caused by a different rational behind both procedures. The basic rationale behind NAL-NL2 is not to normalize loudness, but to optimize the speech intelligibility index with a constrained on overall loudness. This approach has the secondary effect that it normalizes loudness between 500-2000 Hz. Because normally loudness is dominated by the low frequencies in normal hearing listeners, loudness equalization generally provides less gain for the low frequencies than loudness normalization.

1.8 Outline of this thesis

To perform loudness normalization or loudness equalization, it is necessary to know how loudness is perceived by normal hearing listeners and how loudness perception is changed in hearing-impaired listeners. Therefore, all hearing aid prescription rules include information derived from loudness measurements in normal hearing, and hearing-impaired listeners. For instance, in the early version of DSL (Seewald *et al.*, 1985) the desired sensation levels (DSLs) are based on data on comfortable listening levels across a range of hearing levels and maximum speech recognition (Kamm *et al.*, 1978; Pascoe, 1978; Gengel *et al.*, 1971; Erber and Witt, 1977; Macrae, 1986; Smith and Boothroyd, 1989). The NAL-NL prescription rules (NAL-NL1, NAL-NL2) use a loudness model (Moore and Glasberg, 1997, 2004) to predict loudness growth for a given hearing loss.

The current approach of the hearing aid prescription rules that derive loudness normalization from historic loudness data or loudness models has a few limitations. Three of these limitations are addressed in this thesis. The first limitation is that the loudness data used in the hearing aid prescription rules have all been collected with static stimuli. However, in daily life most sound environments consist of dynamic stimuli. Hearing aids are becoming gradually better in analyzing sound environments and the improved processing speed and processing power of hearing aids make it possible to make real-time gain adjustments. Therefore, it is becoming more and more important to predict the impact of dynamic changes on loudness perception. Dynamic loudness models have been developed (i.e.: Glasberg and Moore, 2002; Chalupper and Fastl, 2002) to include the influence of duration on loudness perception. These dynamic models, however, are not yet able to correctly predict the interactions between the spectral and dynamic properties of sound. To improve the dynamic loudness models more data on dynamic interactions in loudness perception is needed.

In the first part of this thesis two research questions are addressed concerning the loudness of dynamic sounds and the role of the cochlear amplifier in normal hearing listeners:

A) How does bandwidth influence loudness of short stimuli and series of noise bursts?

Verhey and Kollmeier (2002) showed that in normal hearing listeners spectral and temporal loudness effects interact. To model this interaction effect a time or bandwidth dependent cochlear gain is needed. In Chapter 2 two experiments are presented that aim to contribute to the literature on time-dependent spectral loudness summation. The first experiment on time dependent spectral loudness summation replicates part of the study by Verhey and Kollmeier (2002). The second experiment investigates the influence of bandwidth on the loudness of a series of noise bursts. In these studies, an alternative procedure for the classical loudness matching procedure has been introduced. The new procedure shows the same qualitative results as the classical loudness matching procedure, but the absolute effects are smaller.

B) Is the temporal effect also apparent at supra-threshold levels?

The temporal effect refers to the finding that a higher signal-to-masker ratio may be needed to detect a brief signal at the onset of a masker than if it is presented with an onset delay. One of the explanations for this phenomenon may be a change in amplification in the cochlea with time, as is the case in the hypotheses to explain time dependent spectral loudness summation. In Chapter 3 temporal effects will be presented that were measured at threshold and at supra-threshold levels.

The second limitation of current models on loudness perception is that the hearing aid prescription rules of NAL and DSL are in principle threshold based. DSL allows to enter frequency specific Loudness Discomfort Levels (LDLs) for the determination of the output-limiting levels. But in case frequency specific LDLs are not available, DSL predicts the maximum output based on the LDL data reported by Pascoe (1978). Martin *et al.* (1998) and Mueller (2003) found in surveys that only 60% of the audiologists perform LDL measurements on hearing aid candidates. Furthermore, Mueller (2003) found that only 27% of the audiologist perform frequency specific LDL measurements. The majority only use speech stimuli. NAL-NL1 and NAL-NL2 do not even allow to enter individual Loudness Discomfort Levels (LDLs).

Thus, in most cases it is assumed that the form of the entire loudness growth curve can be based on the audiometric threshold alone. However, among others Formby *et al.* (2017) have shown that subjects with equal audiometric thresholds can have considerably different LDLs. This may explain that discomfort with loud sounds, is one of the major complaints by hearing aid users (Jenstad *et al.,* 2003; Kochkin, 2000).

As an intermezzo between part 1 and 2 in which loudness experiments are described in well defined groups of normal hearing (part 1) and hearing-impaired listeners (part 2), Chapter 4 describes loudness measurements in several clinical groups. The main research question in this intermezzo is:

C) What can we learn from measuring the entire loudness function in a clinical population?

The third limitation of current models on loudness perception is that the historic loudness data for hearing-impaired listeners were collected for unaided conditions. In unaided conditions the audibility has a major impact on loudness perception. It is therefore not surprising that many studies found less spectral loudness summation for hearing-impaired listeners than for normal hearing listeners (Appell and Hohmann, 1998; Brand and Hohmann, 2001; Bonding and Elberling, 1980; Garnier *et al.*, 1999).

Implicitly the hearing aid prescription rules assume that loudness normalization does not have an impact on loudness perception for hearing-impaired listeners. Another assumption is that spectral loudness summation for hearing-impaired listeners is equal to or smaller than for normal hearing listeners. A similar assumption is made for binaural loudness summation. Binaural loudness summation for hearing-impaired listeners is assumed to be less than or equal to binaural loudness summation for normal hearing listeners and corrections for bilateral hearing aid prescription are therefore only small.

Recently however, Oetting *et al.* (2016) showed that narrowband loudness compensation does not guarantee normal loudness perception for hearing-impaired listeners for broadband signals, especially when a sound is presented bilaterally. The second part of this thesis describes two follow-up studies that investigate the effects of hearing impairment on spectral loudness summation and binaural loudness summation more thoroughly. The research questions in the second part of the thesis are:

D) How does hearing loss configuration influence spectral and binaural loudness perception?

In Chapter 5 the study by Oetting *et al.* (2016) is replicated for a larger group of hearing-impaired listeners. The group is subdivided into four groups with different hearing loss to investigate the influence of the shape of the audiogram on the perceived loudness. As in Oetting *et al.* (2016) loudness equalization is applied.

E) How does frequency content influence spectral and binaural loudness perception?

In Chapter 6 loudness perception for a group of hearing-impaired listeners is again investigated for unilateral and bilateral presentation. However, in this study no narrow-band loudness equalization is applied and loudness perception of low-pass filtered, high-pass filtered, and broadband pink noise is compared.

F) An important final research question in the second part of this thesis is: do we need individual loudness measurements for loudness normalization?

In Chapter 5 and 6 considerable individual variability is found in loudness perception, especially for higher intensities.

The results in Chapter 4, 5, and 6 show that the assumptions made in prescription rules do not always match with the needs in the individual case. Therefore, individual loudness measurements may be regarded as a good and safe way to ensure appropriate aided loudness perception at high levels.

Chapter 2

Bandwidth and duration dependency of loudness

This chapter is a compilation of two papers:

Van Beurden, M.F.B., and Dreschler, W. A. (2005). Bandwidth dependency of loudness in series of short noise bursts. Acta Acustica united with Acustica, 2005, 91(1), 1020-1024.

Van Beurden, M.F.B., and Dreschler, W. A. (2007). Duration dependency of spectral loudness summation, measured with three different experimental procedures. In Hearing - From Sensory Processing to Perception, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, J.L. Verhey, (Springer, Berlin), pp. 237-245.

Abstract

This chapter describes two experiments that examine how loudness perception is influenced by the interaction between signal duration and signal bandwidth. In experiment 1 spectral loudness summation at different signal durations is measured with three loudness procedures. In experiment 2 the influence of different burst-and inter-burst-durations on the perceived loudness of a series of noise bursts was measured as a function of bandwidth.

Experiment 1 shows with three different measuring procedures that spectral loudness summation is larger at short signal durations. This is in line with previous experiments. The amount of loudness summation depends on the measuring procedure. Experiment 2 also shows an interaction between bandwidth and the temporal structure of the signal. Our results suggest that strong modulations (pulse trains) lead to a larger loudness for larger bandwidths. This is in line with the results of experiment 1, as stronger modulations correspond with shorter signal durations. However, the exact relationship between temporal structure and loudness remains to be established.

2.1 Introduction

Loudness perception depends on many physical parameters such as intensity, frequency, bandwidth, duration, and temporal structure. Many studies have investigated the influence of one or more of these parameters on loudness perception (i.e.: Zwicker *et al.*, 1957; Florentine *et al*, 1996; Moore *et al.* 1999a; Verhey and Kollmeier, 2002). It was found that loudness increases when the bandwidth exceeds a critical bandwidth, this effect is called *spectral* loudness summation (Zwicker *et al.*, 1957; Zwicker and Scharf, 1965; Verhey and Kollmeier, 2002; Cacace and Margolis, 1985). Another property of loudness is that its magnitude increases as the duration increases up to a "critical duration" of about 200 ms, which is known as *temporal* loudness summation or temporal integration (Zwicker, 1965; Florentine *et al.*, 1996; Buus *et al.*, 1997, 1998).

Often spectral loudness summation and temporal loudness integration are treated separately (Zwicker *et al.*, 1957; Zwicker and Scharf, 1965; Buus *et al.*, 1997; Port, 1963), and no interaction on loudness perception is presumed. However, there are several studies that indicate that the influences of spectral and temporal loudness integration cannot simply be added in order to achieve the overall loudness, but that an interaction does exist. Verhey and Kollmeier (2002) showed that spectral loudness summation depends on the duration of a sound; spectral loudness summation for a noise band with a center frequency of 2000 Hz was found to be 6 to 8 dB larger for a duration of 10 ms than for a duration of 1000 ms. This finding was confirmed by Chalupper (2002) but these results contrast with earlier measurements in which no dependency on signal duration was found (Zwicker and Scharf, 1965; Port, 1963).

An interrelation between the spectral and the temporal domain is also evident in the loudness perception of modulated signals. The results of studies on the influence of modulations on loudness are not entirely consistent, but they suggest that for narrowband signals at medium modulation rates (20-100 Hz) an amplitude modulation leads to a decrease in loudness (Moore *et al.*, 1999a; Grimm *et al.*, 2002; Zhang and Zeng, 1997; Gockel *et al.*, 2002; Moore *et al.*, 1998), whereas for broadband signals an amplitude modulation results in a small increase in loudness (Moore *et al.,* 1999b; Verhey and Kollmeier, 2002; Grimm *et al.*, 2002; Zhang and Zeng, 1997; Moore *et al.,* 1998). The temporal structure thus leads to a different loudness perception depending on the bandwidth of the modulated sound.

Several explanations for the change in loudness perception as a result of modulation have been proposed. Moore *et al.* (1998) explained the altered loudness perception of modulated sounds by the compression that occurs on the basilar membrane. Their explanation could explain the slight decrease in loudness at small bandwidths, but not the slight increase at large bandwidths. An interesting explanation for the increased loudness of broadband modulated sounds is given by Grimm et al. (2002). Based on their data on loudness of modulated sounds as a function of bandwidth and data from Verhey and Kollmeier (2002) on duration dependent spectral loudness summation, they stated that if a modulated sound is interpreted as a sequence of short signals, the difference in spectral loudness summation between short and long signals can explain the loudness difference between modulated and unmodulated signals.

Garner (1948) measured the loudness of repeated short tones. He showed that a series of repeated short tones can be louder than a steady tone of the same peak intensity. He also found that this loudness difference depended on frequency, intensity, repetition rate, and signal duration, which all interacted. For instance, the rate at which loudness increased with repetition rate was found to depend on the duration of the repeated tones. The shorter the duration was, the larger the change in loudness with a change in repetition rate. Garner (1948) explained his results for repeated short tones in terms of a spread of energy over a larger range of frequencies, which leads to spectral loudness summation. In later studies with modulated signals it was ensured that short noise bursts had the same frequency contents as the continuous reference signal. A simple spread of energy over frequencies can therefore not explain these results.

Pollack (1951) measured auditory threshold and loudness of repeated bursts of noise and found that an interrupted noise was perceived louder than a continuous noise with the same energy. The maximal difference was found for repetition rates of 2-10 Hz. Pollack (1951) explained his data with an increased neural response for interrupted stimuli, especially with a sharp transient “on” time. This is in line with experimental results from Carter (1972). He showed that decreasing the rise time of a series of short noise bursts increased the perceived loudness. On the basis of a similar hypothesis Grimm *et al.* (2002) evaluated the benefits of a contrasting model, which emphasized rapid temporal level fluctuations. They concluded that such a model gave encouraging results. The main limitation of the contrasting model was its failure to predict the bandwidth dependency of modulated sounds. An explanation by a model that emphasizes rapid temporal fluctuations seems in line with the findings of Gockel *et al.* (2002) and Carlyon and Datta (1997) that a high peak factor increases the loudness of a signal, however see (Preece and Wilson, 1988).

Similar to Pollack (1951), Reichardt and Niese (1965), Reichardt (1970), Niese (1959), and Port (1963a, b) measured the loudness of repeated noise bursts. They found that the loudness was determined by the root mean square (rms) for fast repetitions of the noise bursts. In contrast, for long bursts and long inter-burst durations loudness of the sequence of bursts was determined by the loudness of a single burst. Reichardt and Niese (1965) simulated those results by assuming a time constant of 25 ms in their loudness meter. Port showed that the level difference between the series of noise bursts and the continuous reference signal increased as the bandwidth of the signals increased. This result is consistent with the explanations for the difference in loudness of amplitude modulated narrow band and broadband signals in Grimm *et al.* (2002).

Many different measurement procedures have been used to study loudness perception. In loudness matching a subject has to compare the loudness of a target signal to the loudness of a reference signal at a certain level. In loudness scaling a subject has to judge the loudness of a single signal on a particular scale for a set of signal levels. Specific advantages of the measuring procedures are that loudness matching is the more accurate procedure, while loudness scaling is more appropriate when loudness perception is assessed over a large range of levels.

In this study we describe the results of two experiments.

1. In the first experiment the interaction between temporal and spectral signal characteristics is investigated by measuring spectral loudness summation at different signal durations. The aim of this experiment is twofold, the first aim is to replicate the results by Verhey and Kollmeier (2002), the second aim is to compare the results of three experimental procedures: one based on loudness scaling and two procedures based on loudness matching.

2. In the second experiment the interaction between temporal and spectral signal characteristics is investigated by measuring the influence of different burst-and inter-burst-durations on the perceived loudness as a function of bandwidth. For this purpose we use series of noise bursts with equal spectral content but differences in burst- and inter-burst-durations as a first approximation of strongly modulated signals. The experimental procedure applied in this experiment is one of the two loudness matching procedures in experiment 1.

2.2 Methods

2.2.1 Measurement procedures

In experiment 1 three different types of loudness measurement procedures are applied, one loudness scaling procedure and two loudness matching procedures:

a) For loudness scaling the Oldenburg-Adaptive CAtegorical LOudness Scaling (ACALOS) procedure, designed by Brand and Hohmann (2002), is applied. This is a loudness scaling procedure with 11 response categories, 5 named categories, 4 un-named intermediate categories and 2 limiting categories, which correspond to categorical loudness levels from 0 to 50. The level assigned to a given loudness category x is termed the "categorical loudness level" Lx. The procedure consists out of two phases. In the first phase the limits of the auditory range are estimated by an interleaved ascending and descending stimulus sequence. In the second phase the four named intermediate categorical loudness levels are estimated. This last phase consists out of two blocks. In the first block the four named intermediate categorical loudness levels are estimated by linear interpolation between the two limits of the auditory range, which are the values at L5 (very soft) and L50 (too loud). In the second block the named intermediate categorical loudness levels are estimated by a modified least-squares fit of a linear model function. In this study 3 iterations of the final block were applied. In the analysis each ACALOS measurement was fitted with a model function consisting of two linear parts with independent slopes and a free cut-point. This model function is a slightly different function than the model function applied by Brand and Hohmann (2002), because their model function had a fixed cut-point at 25 CU. In the loudness scaling procedure each condition was measured three times.

b) The first loudness matching procedure (called matching 1) is an adaptive two-interval, two-alternative forced choice procedure similar to the procedure used by Verhey and Kollmeier (2002). In each trial the subject hear two sounds, a reference signal and a test signal, separated by a 400 ms silent interval. Test and reference signals are presented in random order and with equal a priori probability. The listeners indicate which signal is louder by pressing a button on a two-button console. A simple one-up one-down procedure is used, which converges at the 50% point of the psychometric function. The initial step size of 4 dB is decreased to 2 dB after the second reversal in the adaptive tracking procedure and is held constant for the next eight reversals. To reduce biases several interleaved tracks are used.

c) The second matching procedure (called matching 2) has been designed to reduce the uncertainty region, especially around the equal loudness point, where matching is a very difficult task. Conventional procedures as used in other studies (Florentine *et al.*, 1996, 1998; Verhey, and Kollmeier, 2002; Grimm *et al.*, 2002) have the drawback that these procedures assume that there is only one distinct point at which the loudness difference between the test and reference signals is zero. In the case of loudness matching however there is a certain region in which it is impossible to discriminate between the loudness of the test signal and the loudness of the reference signal. Matching 2 is designed to make matching close to the equal loudness point easier by changing the task from differentiating the louder of two signals to discriminating if there is a loudness difference in a signal pair or not. Instead of comparing the loudness of two signals, the task is to compare the loudness differences of two stimuli pairs. The procedure is essentially equal to the procedure for matching 1 as described above. The only difference is that each interval consists of a pair of signals instead of a single signal and the subject’s task is to indicate in which interval the loudness difference was larger.

In each trial two pairs of sounds are presented, each pair consisting of a reference signal and a test signal, separated by 400 ms. The two sound pairs are separated by 800 ms. In both pairs the reference signal have the same intensity, but in random order with an equal a priori probability one of the two test signals has an intensity increase of 2 dB (this value is just above the intensity of a just noticeable difference (Ozimek and Zwislocki, 1996)). Test and reference signal are presented in random order with equal a priori probability, but the order is the same in both pairs of a trial. The listeners indicate in which sound pair the loudness difference is larger by pressing a button on a two-button console. A simple one-up one-down procedure is used, which converges at the 50% point of the psychometric function. If the listener indicates that the interval containing the intensity increase has the greater loudness difference, the level of the test signals is decreased, otherwise it is increased.

The procedure is based on the assumption that there is a small loudness uncertainty region of about 2 dB, where it is very hard to indicate the difference between the loudness of two signals. The procedure converges to the point where the subject can hear a loudness difference in only one of the sound pairs. This means that the two signals of the other pair are so close together that they fall inside the loudness uncertainty region. The level of the variable sound within this loudness uncertainty region is used to estimate the equal loudness level. All starting levels are chosen randomly from a set of levels ranging from –6 to 6 dB. At the beginning, the step size is 4 dB. It is decreased to 2 dB after the second reversal in the adaptive tracking procedure and held constant for the next eight reversals. A reversal is defined as a change in preference for the interval having the greater loudness difference, the interval with the intensity lift or the interval without the intensity lift. The level difference between test and reference signal needed to obtain the same loudness is determined by calculating the mean of the levels of the variable that fell in the loudness uncertainty region at the last six reversals see figure 2.1. The increase of the test signal can occur in either the first or the second pair. This position of the increased test signal is randomized between the first and second interval in order to ensure that subjects are not able to follow the one-up one-down procedure.

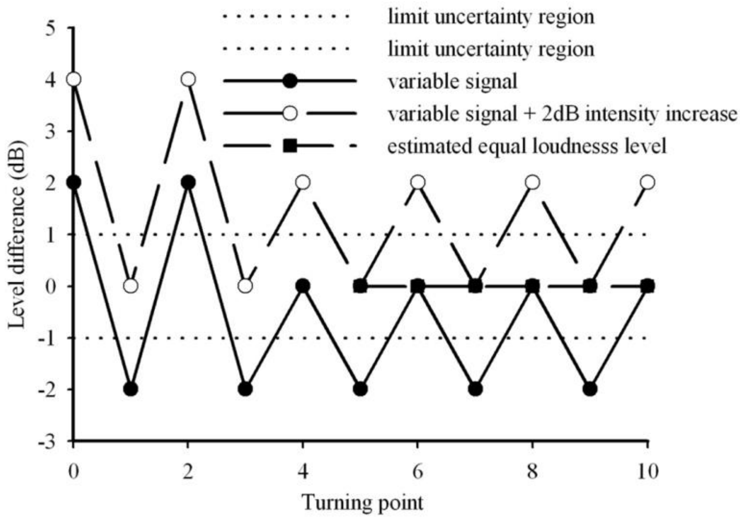


Figure 2.1. *An example of an outcome of the loudness difference procedure. At each turning point the level differences between the variable and the reference and between the variable + 2 dB increase and the reference are shown. The estimated equal loudness level is constructed from the level differences at the upper turning points of the variable and at the lower turning points of the variable + 2 dB increase. The dotted lines present the assumed loudness uncertainty region.*

Around the equal loudness point the task is effectively to determine if there is a loudness difference between a sound pair or not. In loudness matching 1 interleaved tracks were applied to ensure that the subjects would not know which stimulus was the variable in a particular trial. Because if a subject knows which signal is the variable and which is the reference, they can tend to ignore the reference signal and shift their attention towards comfortable loudness. The task in loudness matching 2 is expected to be less sensitive to a shift to the comfortable loudness level as a result of ignoring the fixed reference signal, than the task in the conventional procedure (Florentine *et al*, 1996, Buus *et al.*, 1997). Therefore, no interleaved tracks are used in the adapted procedure. By presentation of one track at a time, the subjects can better focus their attention on the small loudness differences of the signals under consideration. In experiment 1 all three measurement procedures have been applied, while in experiment 2 only loudness matching type 2 was applied.

2.2.2 Stimuli and apparatus

The stimuli applied in both experiments are based on low-noise noise (LNN). In both cases the low-noise noise (LNN) have a peak factor, defined as $W = \overline{x^4} / \left(\overline{x^2}\right)^2$ of approximately 1.7 (± 5 iterations) for each bandwidth applied in the experiments. The noises are generated from pink noise with a method similar to the method described by Kohlrausch et al. (1997). Besides restricting the bandwidth by zeroing the components in the power spectrum outside the original bandwidth, pink noise is created by performing an appropriate amplitude transformation. The entire procedure providsd a pink noise with a well defined bandwidth.

The stimuli are generated in Matlab with a sampling rate of 20 kHz. A computer controls the stimulus generation, registers the subjects' responses and executes the adaptive procedure. The stimuli are played by the computer, converted from digital to analogue by a D/A converter (TDT DA 3-2) and low pass filtered at 8 kHz (TDT FT6). The output of the low pass filter is attenuated by a programmable attenuator (TDT PA4), led to a headphone buffer (TDT HB6), and presented monaurally via headphones (TDH 39). The noises are gated with a raised-cosine rise and fall of 6.67 ms. The equivalent rectangular duration of such a rise and fall is 1.67 ms shorter than the duration between the half-amplitude points and amounts thus to 5.0 ms. A 12.5 ms signal consists then out of the rise and fall and a 7.5 ms steady state portion. The calibration of all signals is based on the long-term rms level of each signal measured in dB SPL. Sound pressure levels are measured using the artificial ear B&K 4153 and the sound level meter B&K 2260 Investigator.

2.2.2.1 Experiment 1

In experiment 1, the test and reference signals are band-limited low-noise noise signals geometrically centered around 2000 Hz. In the loudness matching procedures the reference signal has a bandwidth of 800 Hz and the test signals have bandwidths of 1600, 3200, and 6400 Hz. Measurements are performed at two signal duration: 25 and 1000 ms. The level of the reference signal in the loudness matching procedures is roved between 54 dB SPL and 66 dB SPL. The level of the test signals follows the roving of the reference signal in order to avoid a disturbance of the adaptive procedure.

In the loudness scaling procedure no reference bandwidth is needed and test signals have bandwidths of 100, 400, 1600, 3200 and 6400 Hz. Measurements are performed at two durations: 25 and 400 ms.

2.2.2.2 Experiment 2

In experiment 2 the test signals are trains of low-noise noise bursts. The noise bursts are made from the same noise as the reference noise and because of the smooth rise and fall the test signals have equal spectral properties as the reference noise. Test and reference signals are matched for two bandwidths, 200 Hz and 3200 Hz wide, geometrically centered around 2000 Hz.

Table 2.1. *Overall intensity of the series of noise bursts (middle column) and intensity of the single noise burst (right column) for different conditions (left column).*

Burst duration / inter-burst duration (ms)	RMS level dB SPL total signal train	RMS level dB SPL noise bursts
12.5/12.5	60	63
25/25	60	63
50/50	60	63
12.5/87.5	60	69
25/75	60	66
Continuous noise	60	60

Five different trains of noise bursts with different burst and inter-burst durations are used (see column 1 in Table 2.1). The durations of both test and reference signal are 1000 ms. The level of the train of noise bursts is defined in dB rms of a long-term signal. This means that the rms of each separate noise burst depends on the specific series of noise bursts. Table 2.1 gives an overview of the rms levels of the total signal and the corresponding rms levels of the noise bursts for each combination used in experiment 2. The experimental procedure in experiment 2 is loudness matching 2.

2.2.3 Subjects

In experiment 1 nine normal hearing subjects (4 male, 5 female) participated in the loudness matching experiments. The age of the subjects ranged from 18 to 34 years. Two of the subjects were members of the Audiology department; the other subjects were paid volunteers without previous experience with loudness matching experiments. In the loudness scaling experiment twelve other normal hearing subjects (5 male, 7 female) in the age from 18 to 36 participated. The subjects were paid volunteers without previous experience with loudness scaling experiments.

In experiment 2 eight subjects (1 male, 7 female) participated. The age of the subjects ranged from 18 to 25 years. Seven of these subjects were paid and had no previous experience with loudness matching experiments.
All subjects had normal auditory thresholds (i.e. absolute threshold in quiet < 15 dB HL) and no previous history of any hearing problems.

2.3 Results

2.3.1 Experiment 1

Figure 2.2 shows the results of loudness procedures matching 1 and matching 2 at a center frequency of 2000 Hz and with a reference bandwidth of 800 Hz. The figure shows the differences between the level of the test signal and the reference signal (ΔL) at equal loudness as a function of the bandwidth of the test signal. A negative level difference means that the test signal needs a lower level to be judged as equally loud as the reference signal. Signal durations were 25 ms (circles) and 1000 ms (squares). The error bars indicate plus and minus one standard error of the mean.

The figure shows that:

a) Loudness summation increases with increasing bandwidth.
b) Loudness summation is larger for the 25ms signals than for the 1000 ms signals.
c) Loudness summation is larger for matching 1 than for matching 2.

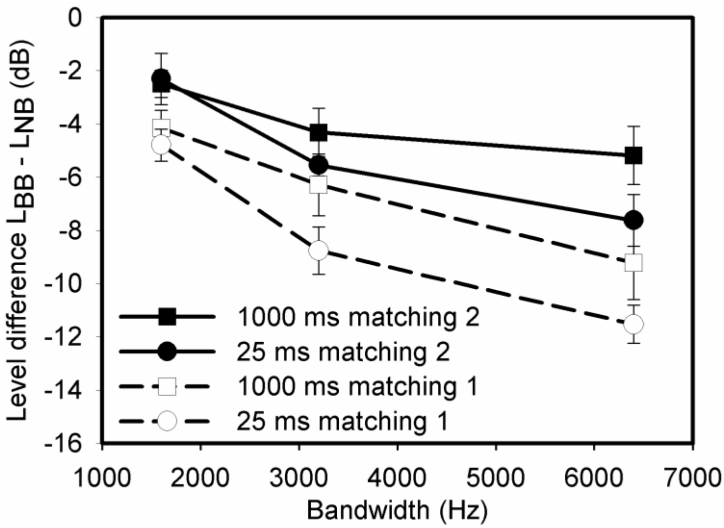


Figure 2.2. *Results of the matching 1 (solid lines) and matching 2 (dashed lines) at 25 ms (circles) and 1000 ms (squares). Error bars indicate plus and minus one standard error of the mean.*

The results of the loudness scaling procedure for five bandwidths and two signal durations are presented in figure 2.3. Each ACALOS measurement was fitted with a model function consisting of two linear parts with independent slopes and a free cut-point. Therefore, each fit was characterized by four parameters. The fits shown are based on the average of the parameters across subjects. The loudness curve of a subject for a certain condition was calculated on the basis of all loudness points obtained in the three runs of the loudness scaling procedure for that condition.

The figure shows that:

a) The slopes of the higher-intensity part are usually steeper than for the low-intensity part. These concave forms are found for all signal bandwidths and both signal durations.
b) The low-intensity slope is less steep for the 25 ms signals.
c) Furthermore, the loudness curves are ordered according to bandwidth, with a larger bandwidth leading to a higher loudness at the same level. This is what is expected from spectral loudness summation.
d) Finally, a comparison of both figures shows that corresponding levels yield a higher loudness level for the 400 ms signals than for the 25 ms signals, as would be expected from temporal integration.

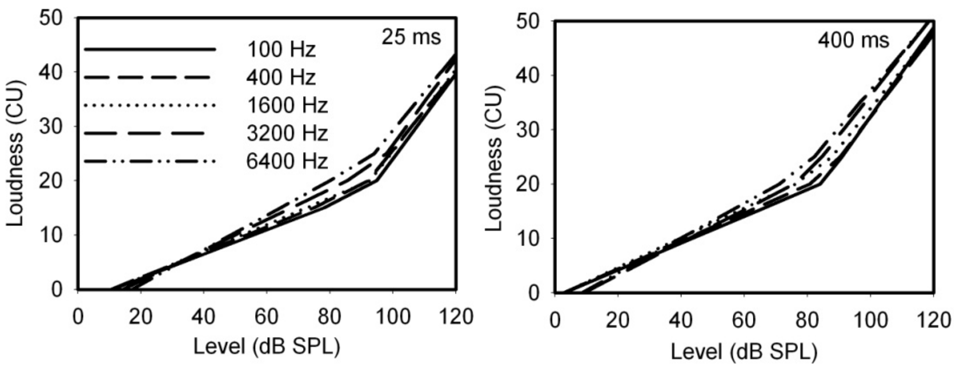


Figure 2.3. *Results loudness scaling for 25 ms and 400 ms signals for five different bandwidths.*

Loudness matching procedures only allow to compare loudness differences at one reference level at a time. The amount of information provided by loudness matching is therefore limited, unless the measurement is conducted at multiple reference levels, which takes a lot of time. Alternatively, loudness scaling provides a full loudness growth curve of the test signal under consideration. However, this loudness growth curve is based on a model fit and that fit might not be completely accurate at all reference levels.

One of the objectives of this study is to compare the amount of spectral loudness summation obtained with the three different procedures. However, it should be stressed that it is only possible to make a qualitative comparison between the loudness matching procedures and the loudness scaling procedure. There are three reasons why a quantitative comparison is not possible.

1. Loudness scaling is performed in a different group of subjects than loudness matching. With the current small group sizes this may lead to small differences in the average effect of spectral loudness summation.

2. The reference bandwidth of 400 Hz applied in the loudness matching procedures is not measured in the loudness scaling experiment. For calculation of the spectral loudness summation in the loudness scaling experiment, 800 Hz is taken as the reference bandwidth. A broader reference bandwidth leads to slightly less spectral loudness summation.

3. The duration of the long duration signal is 400 ms in the loudness scaling procedure, while it is 1000 ms in the matching procedures. Although the impact of temporal integration above 200 ms is supposed to be small (Buus *et al.*, 1997), it cannot be ruled out that the different signal duration has an impact on spectral loudness summation.

To make it possible to compare the results of the loudness scaling procedure to the results of the loudness matching procedure, we calculated the loudness differences between the broadband signals and the narrowband signal at a specified level of the narrowband signal for each of the loudness scaling procedures. The level of the narrowband reference signal in the loudness matching procedures was 60 dB SPL. Therefore, the level differences at equal loudness were calculated relative to the loudness of the 400 Hz narrowband signal at 60 dB SPL. Table 2.2 shows the calculated summation data. At 25 ms the loudness of the narrowband reference signal at 60 dB SPL is 11.6 CU while at 400 ms the loudness is

14.5 CU. So, there are slight loudness differences between the reference signals at different durations. It is important to realize that these differences in reference loudness are also present in the loudness matching procedures.

Table 2.2. *Spectral loudness summation difference between short and long duration signals in dB SPL. For loudness matching 1 and 2 the reference bandwidth is 400 Hz and the duration of the long signal is 1000 ms. For loudness scaling the reference bandwidth is 800 Hz and the duration of the long signal is 400 ms. In all procedures the duration of the short duration signal is 25 ms.*

	Summation difference Matching 1	Summation difference Matching 2	Summation difference Scaling
1600 Hz	0.64	-0.14	-1.58
3200 Hz	2.47	1.23	2.75
6400 Hz	2.31	2.44	1.14

2.3.2 Experiment 2

Experiment 2 measured the loudness difference between a continuous signal and a series of noise bursts. Figure 2.4 shows the average results of the eight subjects for noise bursts with bandwidths of 3200 Hz and 200 Hz, plotted as a function of condition, defined by the burst and the inter-burst duration, indicated by t1/t2. The y-axis represents the level difference between the series of noise bursts and the continuous reference signal at equal loudness. The level of each signal is defined as the peak level of the noise burst. The error bars indicate the standard error of the mean.

The large standard errors indicate that the differences between subjects are considerable. Despite these large differences a few interesting observations can be made in the average results.

1. The main effect of bandwidth is that the effect of the temporal structure on loudness are about 1 dB stronger at 3200 Hz.

2. A General Linear Model Repeated Measures ANOVA shows that both bandwidth and condition have a significant effect. A Tukey posthoc test shows that the 12.5/87.5 condition is significantly different from

all other conditions and that the 25/75 condition is significantly different to all conditions except the 25/25 condition. It is interesting to see that with a bandwidth of 3200 Hz, the presentation of half of the signal in the three conditions with a duty cycle of 50% (12,5/12,5, 25/25, and 50/50) hardly alters the loudness perception, as the level difference between the series of noise bursts and the continuous signal is almost zero. The decrease in duty cycle is mainly compensated by an increase in burst level. This increase is similar in size to the increase expected from the assumption that the overall intensity determines the loudness (dashed line in figure 2.4).

3. Finally, a comparison with the broken line indicating equal rms of the series of tone bursts and continuous reference signal indicates that the series of noise bursts need less intensity to sound equally loud as the continuous noise. In other words, for a given intensity the series of noise bursts are perceived louder than a continuous noise signal with the same spectral content and overall intensity.

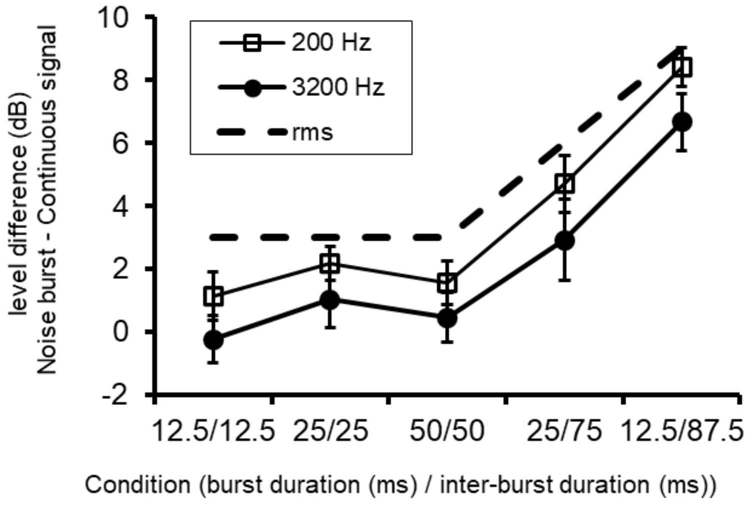


Figure 2.4. *Group results of temporal integration of series of noise bursts with a bandwidth of 200 Hz and 3200 Hz, the error bars indicate the standard error of the mean. Given levels are the rms levels of the noise bursts. The dashed line indicates the level of equal rms of the total 1000 ms series of noise bursts.*

2.4 Discussion

2.4.1 Experiment 1

Although the scaling procedure and both matching procedures have been conducted with slightly different stimuli the same trends can be observed. First, in all three procedures spectral loudness summation is larger for short signals than for long signals. This corresponds well with the findings of Verhey and Kollmeier (2002) and Chalupper (2002). The fact that duration-dependent spectral loudness summation has been found in each of the three measuring procedures provides extra support for the existence of a temporal effect and excludes possible artifacts due to the measurement procedure.

The amount of spectral loudness summation difference depends on the amount of summation, which is in agreement with the results of Brand and Hohmann (2002). The duration-dependency of spectral loudness summation is small, when the loudness summation is small. As loudness summation increases, the loudness summation difference also increases. However, the maximum amount of loudness summation difference between short and long signals seems limited. In all three procedures the amount of loudness summation at a bandwidth of 3200 Hz and 6400 Hz is approximately the same. A further investigation with even broader bandwidths is needed to confirm our observation that a ceiling effect may be present. It would be interesting to determine the bandwidth at which the summation difference between long and short signals reaches the maximum value.

There are also differences between the procedures, especially with respect to the amount of summation found. This is probably a consequence of procedural differences. In the second matching procedure we assumed that interleaving of the different conditions was not necessary. Verhey (1999) found that an adaptive procedure with interleaved tracks leads to larger loudness summation. The differences we found between matching 1 and matching 2 correspond to the differences found between an interleaved and a non-interleaved procedure. An experiment with matching 2 with interleaved tracks could clarify if the difference between the two matching procedures is truly caused by the choice to not apply interleaving in matching 2. The difference between the two matching procedures may also be influenced by the choice for a 2 dB uncertainty range. It is equally interesting to investigate the impact of smaller or larger size of the uncertainty range.

The long duration condition of the scaling procedure is conducted with a 400 ms signal instead of a 1000 ms signal. The influence of this difference in signal duration may be expected to be negligible, as the effect of temporal loudness integration is thought to be limited to approximately 200 ms. The scaling procedure is much less sensitive than the two matching procedures at one specific loudness. The results depend heavily on the definition of the fitting curve. Nevertheless, the results correspond reasonably well with the matching results and give also a hint towards the level dependency of the effect. At low levels there is almost no spectral loudness summation and therefore the summation difference is also very small. Around the cut-point of the fitting curve, which seems to lie at the lower side of the most comfortable loudness region, both the spectral summation and the summation difference are largest. At higher levels they tend to decrease again.

2.4.2 Experiment 2

The results of experiment 2 show that series of noise bursts are perceived to be louder than the continuous signal for a given overall rms presentation level in dB SPL. This is in agreement with the literature on repeated bursts of noise. Garner (1948), Pollack (1951) and Port (1963a, b) showed that repeated bursts of sound were perceived louder than a continuous sound with the same total energy for tones and white noises. Garner (1948), who only used tones, found a maximum level difference between equally loud series of sound bursts and continuous signals of 2-3 dB. Pollack (1951) who used broadband white noise signals, found level differences between series of noise bursts and continuous signals up to about 5 dB. If we assume that we may combine the results of these two studies, this leads to a difference of 2-3 dB between the level of a series of short tones and the level of a series of short noises at equal loudness. Port (1963a, b) found differences of the same magnitude between a series of 2 ms noise bursts with a 50 Hz burst-frequency and a continuous noise of the same bandwidth. The level differences between the loudness of series of narrow-band signals (200 Hz) and broadband signals (3200 Hz) found in this study were on average 1 dB, which is smaller. This probably results from the smaller bandwidth difference in this study. As Port (1963a, b) has shown the level difference increases as the bandwidth difference increases.

The finding, that at a bandwidth of 3200 Hz, the presentation of only half of the signal hardly alters loudness perception at the 50/50 condition, seems in line with the hypothesis from Reichardt (1965) and Niese (1959) that loudness perception has a decay time of about 25 ms. In the 50/50 condition the interburst-time is larger than this decay time and the loudness hardly decreases. In the 25/25 and 12.5/12.5 condition the interburst-time is equal or smaller than this decay time and especially in the 12.5/12.5 condition a loudness decrease is expected, this was not confirmed in the results of this study.

Slightly unexpected are the insignificant differences between the different conditions when the results are given in rms of the total signal. If increased spectral loudness summation for short signals causes the increased loudness perception for series of noise bursts (as hypothesized by Grimm et al. (2002)), the spectral loudness summation differences between the two bandwidths would be expected to be larger in the conditions with short noise bursts than in the conditions with a longer burst duration. Such an effect is not apparent in our results. This may however be a consequence of the large spread in the data.

2.5 Conclusions

Experiment 1 shows with three different measuring procedures that spectral loudness summation is larger at short signal durations. This is in line with the results of Verhey and Kollmeier (2002). Although the amount of summation differs in the different procedures the summation differences are approximately the same. This indicates that duration dependent spectral loudness summation does not depend on the measuring procedure. Our data do show a possible ceiling effect in the amount of spectral loudness summation differences between short and long signals.

Experiment 2 also shows an interaction between bandwidth and the temporal structure of the signal. Strong modulations (pulse trains) lead to a larger loudness for larger bandwidths. This experiment confirms that spectral and temporal characteristics interact, and that broad bandwidths in combination with short duration increase loudness perception. However, the exact relationship between temporal structure and loudness remains to be established.

Acknowledgments

The authors thank László Körössy for his help in developing and programming the used measuring procedure. We also thank Koen Rhebergen for his comments on earlier versions of this paper. This research was supported by a grant from the Heinsius-Houbolt Fund.

Chapter 3

Partial loudness at masker onset indicates temporal effects at supra-threshold levels

Van Beurden, M.F.B., Dreschler, W.A. (2018). Partial loudness at masker onset indicates temporal effects at supra-threshold levels. Hear Res. 370, 168-180.

Abstract

This study examines temporal effects both at threshold and at supra-threshold levels. The level needed to detect a short-duration 4.0-kHz signal was measured for signals presented with different onset delays relative to a 300-ms broadband noise masker: 100 ms and 5 ms before the onset of the masker and 5 ms and 100 ms after the onset of the masker. Loudness matches between the signal in quiet and the signal at the same four onset delays were obtained for five presentation levels of the short-duration signal and for three masker levels. The temporal effect was defined as the level difference between the signals near masker onset and the signals well before or well after masker onset, needed to reach threshold and/or achieve equal loudness. Both at threshold and at supra-threshold levels temporal effects were observed consistent with a decrease in gain at the masker frequency during the course of the masker. The temporal effect was not restricted to simultaneous masking, but was also found for backward masking. In both cases the temporal effects were stronger at supra-threshold levels than at threshold. This may be caused by a transient effect at masker onset. The almost simultaneous onset of the signal and the masker makes it difficult for subjects to separate signal from the masker, especially when the signal level is close to masked threshold.

3.1 Introduction

"Overshoot", also called "the temporal effect" (Zwicker, 1965; Bacon and Smith, 1991; Strickland, 2001, 2004, 2008; Strickland and Krishnan 2005), refers to the finding that a higher signal-to-masker ratio may be needed to detect a brief signal at the onset of a masker than if it is presented with an onset delay, or if it is preceded by another sound (precursor condition).

There is a growing body of evidence showing that overshoot may be explained to a large part by changes in amplification in the cochlea. Schmidt and Zwicker (1991) were the first to propose a reduction of the amplification of the active process in the course of a long noise masker as an explanation for overshoot and von Klitzling and Kohlrausch (1994) developed this idea in more detail. More recently Strickland (2001, 2004, 2008) and Strickland and Krishnan (2005) have shown that input-output functions derived from overshoot data are consistent with a frequency-specific decrease in gain at the masker frequency during the course of masker stimulation. A decrease in gain during the course of the masker can explain overshoot because less gain leads to less compressive behavior of the input-output function of the basilar membrane. The input-output function grows with a slope near 1 for low input levels, grows at a reduced rate for medium input levels, and then may grow more rapidly again for high input levels (e.g. Ruggero and Rich, 1991). The input level of the brief signal is usually much higher than the input level of the masker. As a result of these differences in input level the signal and masker are on a different portion of the input-output function. Generally it is assumed that a signal can be detected at a constant signal-plus-masker to masker ratio. If the signal is on the higher compressive part of the input-output function and the masker is on the lower linear part, the amount of compression determines the input-level difference needed to obtain this constant signal-plus-masker to masker ratio criterion at the output. For an input-output function with more compression, the input of the signal has to be stronger to satisfy the output criterion, than for an input-output function with less compression. A lower gain during the course of the masker may therefore explain overshoot.

Factors that impair the active process in the cochlea also tend to decrease overshoot. Cochlear hearing loss has been shown to lead to reduced overshoot (Bacon *et al.*, 1988; Kimberley and Nelson, 1989; Bacon and Takahashi, 1992; Strickland and Krishnan, 2005). The same holds true for temporary hearing loss induced by noise exposure (Champlin and McFadden, 1989) and aspirin use (McFadden and Champlin, 1990). The reduction of overshoot found in these experiments was due to an improvement of the threshold for the signal at masker onset. As a result thresholds became the same at masker onset as with a longer masker onset delay.

A possible explanation for overshoot could be that a change occurs in cochlear amplification during the course of the masker, for instance a decrease in the gain of the cochlear active process during the course of the masker. As Bacon and Savel (2004) pointed out it is important to realize that basilar-membrane mechanics do not -by themselves- change over time. Thus some other factor must be responsible for the decrease in threshold over time. The decrease in gain of the active process could be mediated by the medial olivocochlear bundle that feeds back from the level of the olivary complex to the outer hair cells in the cochlea (Guinan, 2006; Russell and Murugasu, 1997; Cooper and Guinan, 2006), as argued extensively in literature on overshoot (e.g., Schmidt and Zwicker, 1991; Strickland, 2001; Roverud and Strickland 2010).

Reduced basilar-membrane compression by mediation of the medial olivocochlear reflex (MOCR) has also been proposed as a possible explanation for reduction of the mid-level hump in intensity discrimination (Roverud and Strickland, 2015a, 2015b) and improved detection of low-frequency amplitude modulation (Almishaal *et al.* 2017). Roverud and Strickland (2015a, 2015b) studied intensity discrimination for short (e.g. 30 ms), high-frequency (~6 kHz) tone stimuli in background noise in which the noise was either short (50 ms), or long (150 ms). In quiet such stimuli show poorer discrimination limens at mid-levels (around 50 dB SPL), which has been termed the “severe departure from Weber’s law” (Carlyon and Moore, 1986), or the “mid-level hump” (Nizami, 2006). One explanation for the mid-level hump is that it reflects the decreasing slope of the input/output (I/O) function and the onset of compression. A MOCR induced change in cochlear compression could therefore lead to a change in the mid-level hump. Roverud and Strickland (2015a, 2015b) found a reduction in the mid-level hump for the long duration noises consistent with this theory. A mid-level hump can also be observed in

AM detection thresholds. Almishaal *et al.* (2017) found that a notched-noise precursor decreased the mid-level hump in AM-detection similar to the results for intensity discrimination.

A decrease in gain of the active process during the course of a signal has also been posited as an explanation for the time dependence of spectral loudness summation. Loudness summation is the increase in loudness with an increase in bandwidth of a fixed-level stimulus above a certain critical bandwidth. Several authors (Anweiler and Verhey, 2006; Chalupper, 2002; Grimm *et al.*, 2002; Verhey and Kollmeier, 2002; Verhey *et al.*, 2006) have shown that loudness summation is higher for short-duration than for long-duration stimuli. The difference in loudness summation for short and long durations can be explained by a reduction in gain over time, as follows. The cochlea can be characterized as containing a bank of bandpass filters whose center frequencies span the range from 50 to 15000 Hz (Moore and Glasberg, 1997). The bandwidths of the filters increase with increasing center frequency and can be expressed in equivalent rectangular bandwidths (ERB_N, Glasberg and Moore, 1990). For a given overall level, the Level/ERB_N is lower for broadband signals than for narrowband signals. As a consequence, broadband signals are processed on a lower region of the input-output function of the basilar membrane than the narrowband signals and the broadband signals are subjected to higher gain values from the active process than narrowband signals. If the gain of the active process reduces over time, this may reduce the loudness of broadband signals more than the loudness of narrowband signals, leading to a decrease in loudness summation. In addition Verhey *et al.* (2006) have shown that the time dependence of loudness summation is decreased for subjects with sensorineural hearing loss, which is consistent with the observation that overshoot decreases in case of sensorineural hearing loss.

The results from Roverud and Strickland (2015a, 2015b), Almishaal *et al.* (2017) and the literature on time-dependent loudness summation suggest that a time-dependent change in cochlear compression may not be restricted to signal threshold. In the present study a direct comparison between overshoot at threshold and overshoot-like temporal effects at supra-threshold levels was made. For supra-threshold levels a loudness matching paradigm

was utilized. The masker and target in the two experiments were the same, as were the temporal positions of the target relative to the masker.
In principle, measuring an overshoot-like temporal effect in the loudness domain means measuring partial loudness. In the past, partial loudness has mainly been investigated with loudness-matching procedures where the loudness of the target in quiet is matched to that of the target in a background noise. If this is repeated for various fixed levels of the target, a masked loudness-matching function (MLMF) can be constructed, showing the level of the target alone as a function of the level of the target in the mixture at the point of equal loudness. Several studies of partial loudness have investigated the MLMF for a pure tone as a function of the frequency of the tone and the bandwidth of the masking noise (Hellman and Zwislocki, 1964; Hellman, 1970; Pavel and Iverson 1981, Florentine *et al.*, 1998). However, all these studies used only one relatively long onset delay of the target in the masking noise.

Gockel *et al.* (2003) investigated the partial loudness of periodic complex tones in background noise and the partial loudness of noise in a background of periodic complex tones. The complex tones had a fundamental frequency (F0) of 62.5 or 250 Hz. The components were added in either random phase (RPH) or cosine phase (CPH). As a part of their experiments Gockel *et al.* (2003) investigated the influence of onset effects. They used two conditions. “Synchronous onset” (in which the target and masker were gated synchronously) and “asynchronous onset” (in which the masker was gated on 400 ms before the target). For both the tonal target in the noise masker and the noise target in the tonal masker Gockel *et al.* (2003) found that near threshold the levels of the target in the masker were often matched to significantly lower levels of the target alone for the asynchronous condition than for the synchronous condition. At low target-to-background ratios the asynchrony of the onset appeared to help the subjects in hearing out the target as a separate sound from the background. These results therefore show that also at supra-threshold levels an onset effect can be found.

In this study two backward masking conditions were included because time windows play an important role in loudness perception, as was shown by Heeren *et al.* (2011). Heeren *et al.* (2011) investigated to what extent spectral loudness summation occurred for nonsimultaneously presented frequency components. Spectral loudness summation was measured for sequences of short tone pulses with varying frequencies, randomly chosen from a set of five frequencies. In addition, spectral loudness summation was measured for the simultaneous presentation of all five frequencies. The pulse duration was 10, 20, 50, or 100 ms and the inter-pulse interval ranged from 0 to 390 ms. Considerable nonsimultaneous spectral loudness summation was found for short pulse durations and short inter-pulse intervals, and a small effect was observed even for the largest inter-pulse interval. Heeren *et al.* (2011) thus showed that non-overlapping stimulus components can influence loudness perception of the whole signal complex. Temporal windows up to 400 ms were also observed in intensity discrimination in forward masking (Zeng and Turner, 1992; Oberfeld, 2007). More importantly Zeng *et al.* (1991) found evidence suggesting that under some conditions suprathreshold intensity discrimination has a much longer time course of recovery from forward masking than was found for detection. The backward masking condition was aimed to investigate whether interaction between non-overlapping stimulus components also occurs in partial loudness perception.

3.2 Methods

3.2.1 Stimuli and apparatus

The stimuli were generated using Matlab with a sampling rate of 44.1 kHz. A computer controlled the stimulus generation, registered the subjects' responses and controlled the adaptive procedure. The stimuli were generated by the computer, played by an Echo Audio Gina sound card, sent to a headphone buffer (TDT HB6), and presented monaurally via headphones (Sennheiser HDA 200). The signal was a 4-kHz sinusoid. The signal frequency was chosen because previous studies have shown temporal effects with noise maskers to be larger at higher frequencies (Carlyon, 1987; Bacon and Takahashi, 1992; Strickland 2001, 2004).

The signal duration was 10 ms, including 5-ms cosine on- and offset ramps (no steady state). The masker was a broadband filtered white noise with a frequency range of 400-6000 Hz ($0.1f_s$ and $1.5f_s$, where f_s is the signal frequency). The masker duration was 300 ms, including 5-ms cosine on- and offset ramps.

3.2.2 Test procedures

Subjects were tested in a double-walled sound-attenuating booth. All stimuli were presented monaurally to the better ear, based on the average hearing thresholds at 500, 1000, 2000, and 4000 Hz. Experiment 1 used a detection task to measure overshoot. Thresholds were measured with a short signal onset delay of 5 ms (Δ+5) and a long onset delay of 100 ms (Δ+100). Thresholds were also measured for negative onset delays of -15 and -110 ms. This means that the silent interval between signal offset and masker onset was 5 (Δ-5) and 100 ms (Δ-100). Four masker levels were used, 30, 40, 50, and 60 dB SPL, but not all subjects were tested using all masker levels. Three subjects were tested using masker levels of 30, 40, and 50 dB SPL and one subject was tested using masker levels of 40, 50, and 60 dB SPL.

Experiment 1 was based on the procedure described by Strickland (2004, 2008). Thresholds were measured using a three-interval forced-choice adaptive tracking procedure with a two-up, one-down stepping rule. This estimates the 71% correct point on the psychometric function (Levitt, 1971). Temporal intervals were marked visually on a computer monitor and subjects responded via a computer keyboard. Visual feedback was provided. The step size was initially 5 dB, and it was decreased to 2 dB after the second reversal. Threshold was taken as the average of the signal levels at the last even number of reversals at the smaller step size in a block of 50 trials. Blocks for which the standard deviation of the reversal points was 5 dB or greater were discarded. The final threshold estimate was an average of three thresholds.

Experiment 2 was a loudness-matching experiment. The subjects had to match the loudness of a signal in a noise with that of a reference signal in quiet. The masker levels were 30, 40, and 50 dB SPL. The reference signal in quiet was held constant and the level of the signal in the noise was varied. The reference levels of the signal were 35, 40, 45, 50, and 55 dB SPL for the 30 dB masker, 40, 45, 50, 55, and 60 dB SPL for the 40-dB-masker, and 45, 50, 55, 60, and 65 dB SPL for the 50-dB-masker. The same four onset delays were used as in the first experiment. To ensure equal interval lengths in the matching procedure, the signal in quiet was constructed as a signal in a virtual masker with a masker level of -∞.

Loudness matching was conducted using a two-interval two-alternative forced-choice adaptive tracking procedure with a one-up, one-down stepping rule. This estimates the 50% point on the psychometric function (Levitt, 1971). In each trial the subject heard two sounds, the reference signal in quiet and the test signal. The silent interval between the signals was 500 ms. The test and reference signals were presented in random order and with equal a priori probability. The subject indicated which signal was louder by pressing the corresponding key on a keyboard. If the subject indicated that the test signal was louder, its level was reduced in the next trial. Otherwise it was increased. The initial step size was 8 dB. This was divided by two after each reversal until it reached 2 dB. Seven reversals were obtained using the 2-dB step size. The level difference between the test and reference signal at equal loudness for one track was determined by calculating the mean of the levels at the last four reversals. Three matches were obtained for each subject and pair of stimuli. Matches for which the standard deviation was 5 dB or greater were discarded and repeated in the last session. This occurred in less than 3% of the matches and was not restricted to a particular condition. In 6 occasions, only two matches with a standard deviations less than 5 dB were obtained due to limitations in the availability of the subject, or the inability of the subject to obtain a stable match.

To reduce biases that occur when stimuli from only one stimulus pair are matched in loudness in a series of trials, interleaved adaptive tracks were used (Florentine *et al.*, 1996, Verhey and Kollmeier, 2002). Concurrent loudness matches were obtained for all five reference signals at a specific masker level in the same run. On each trial, the track was chosen randomly from all possible tracks, i.e. from all tracks that had not yet been terminated. To ensure that the interleaved tracks converged at roughly the same time, the random choice of tracks was restricted by requiring that each track be selected once in random order before any track could be selected again. Only tracks with the same signal onset time were interleaved. The sequence of interleaved tracks was chosen randomly from all possible combinations of onset time and masker noise level.

3.2.3 Subjects

Eight subjects participated in both experiments. Subject one was the first author, and the others were paid volunteers. All subjects had normal audiograms for the test ear with hearing thresholds better than or equal to 15 dB HL at octave frequencies between 0.125 and 8 kHz.

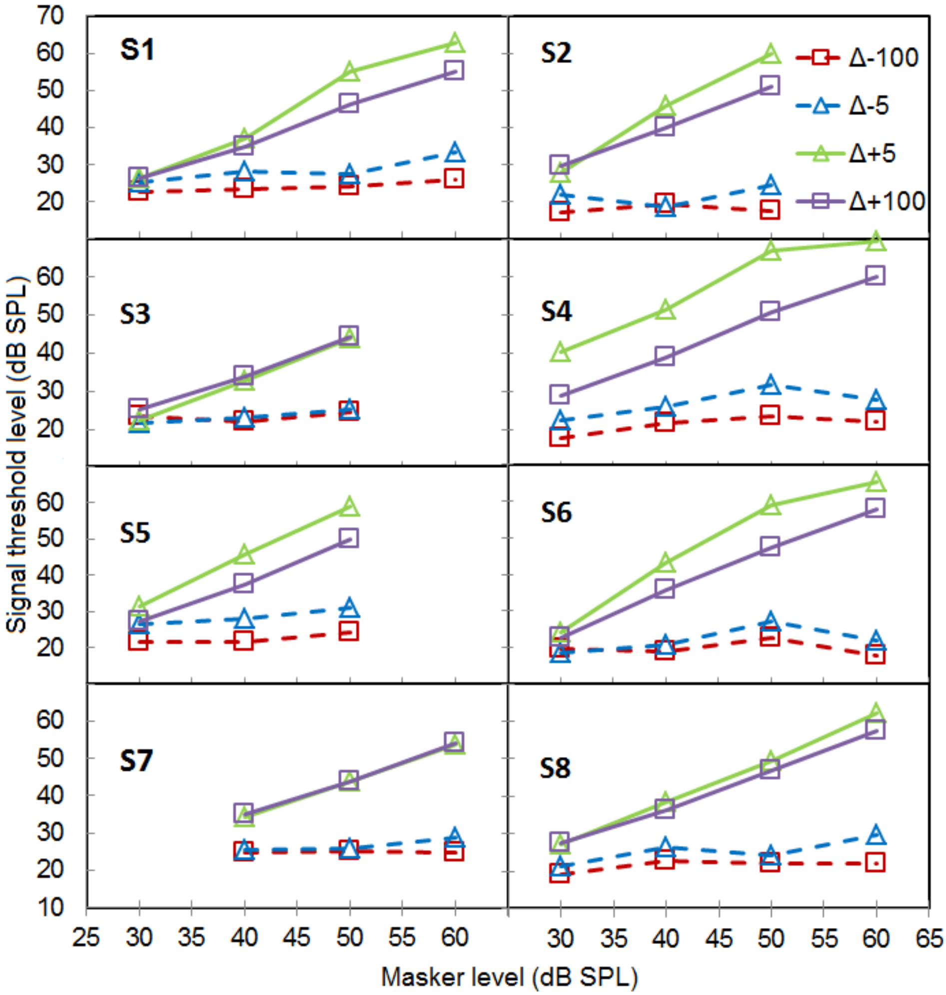


Figure 3.1. *Individual levels at which the signal was just detectable with the masker present plotted as a function of masker level. Squares represent a large difference in onset delay (signal presented 100 ms before or 100 ms after masker onset) and triangles represent a short onset delay (signal presented 5 ms before or 5 ms after masker onset). Backward masking conditions are indicated by dashed lines (Δ-) and simultaneous masking conditions are indicated by solid lines (Δ+).*

3.3 Results

3.3.1 Thresholds

Figure 3.1 shows the individual growth-of-masking functions for the conditions in which the signal was presented prior to the masker (dashed lines) or simultaneous with the masker (solid lines). Triangles and squares show the results for small and large differences in onset delay, respectively. As observed by others (Bacon, 1990; Bacon and Savel, 2004; Wright, 1995, 1997) there were considerable individual differences in the temporal masking effect, as indicated by comparing thresholds for the conditions Δ+5 (triangles, connected with solid lines) and Δ+100 (plotted as squares with solid lines). Three subjects (S3, S7, and S8) showed no or very little temporal effect, while the other five subjects showed a clear temporal effect.

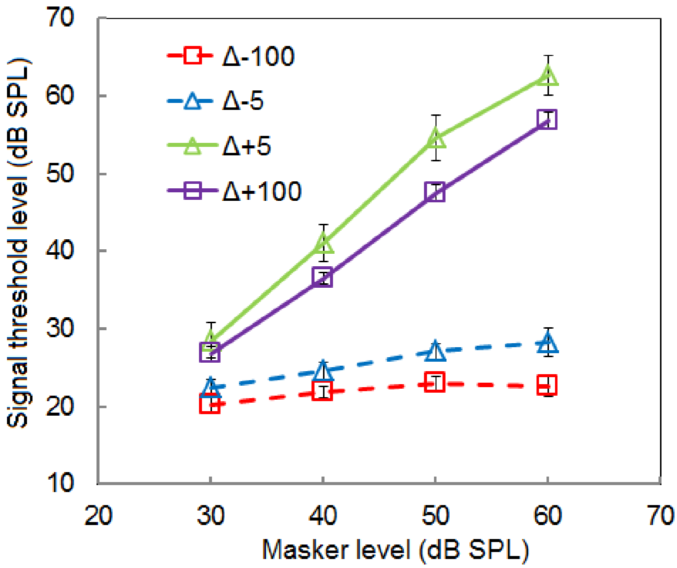


Figure 3.2. *Averaged levels at which the signal was just detectable in the masker plotted as a function of masker level. Vertical bars show plus and minus one standard error across the number of subjects that completed the condition. Symbols are as for figure 3.1.*

Figure 3.2 shows average data. When the signal was presented prior to the onset of the masker (Δ-100 and Δ-5), threshold was hardly influenced by the masker level, even when the time between signal onset and masker onset was as short as 5 ms. However, when the signal was simultaneous with the masker, threshold increased almost linearly with masker level. For the 100-ms onset delay (Δ+100) the slope of the masking function was approximately 1.0, but for the short delay (Δ+5) the slope was higher than 1.0. For three subjects the slope tended to decrease between 50 and 60 dB SPL (see Fiure 3.1). Others (e.g. Bacon, 1990; Bacon and Savel, 2004) have observed similar, multi-segment growth-of-masking functions for a signal presented at the beginning of a broadband masker. The standard error of the mean was largest for the Δ+5 condition.

Figure 3.1 also shows a small temporal masking effect for the signal presented prior to the masker for six of the subjects. Figure 3.2 shows that the average threshold was 2-6 dB higher for the signal presented 5 ms prior to the masker (Δ-5) than for the signal presented 100 ms prior to the masker (Δ-100). This indicates a small amount of backward masking. This is consistent with the literature on backward masking (Dolan and Small, 1984; Penner, 1974). The backward masking appears to be more or less independent of masker level.

To assess the significance of differences between conditions two analyses were performed on the average values of the three tests with repeated measures ANOVA in JASP (JASP Team, 2018). Temporal position of probe and masker (‘condition’, 4 categories) and masker level were the within-subjects factors. In analysis 1 we included all four masking levels. Here the number of subjects was reduced to four because of missing values for four of the subjects at low or high masker levels (30 dB and 60 dB, respectively). In analysis 2 we analyzed the two conditions that were measured in all eight subjects (masker levels 40 dB and 50 dB). In analysis 2 the results for ‘condition’ were corrected according to Huyhn-Feldt, because sphericity could not be assumed. In analysis 1 the main effects of ‘condition’ ($F(3,9)=56.25$ and ‘masker level’ ($F(3,9)=57.3$ were highly significant ($p<0.001$). Pair wise comparisons with Bonferroni correction in analysis 1 revealed that the signal levels at threshold were significantly different from each other ($p<0.001$) for all conditions, and condition S100-S5 was slightly less significant ($p=0.002$). Pair wise comparisons in analysis 1 with Bonferroni correction for masker level revealed that the signal levels at threshold were highly significantly different between masker level 30 dB and all other levels ($p<0.001$). Masker level 40 dB was significantly different from masker level 50 dB and 60 dB

($p=0.041$, $p=0.004$ resp.) and masker levels 50 and 60 dB were also significantly different ($p=0.043$). In analysis 2 the main effects of 'condition' $F(1.3,9.3) = 80.20$ and 'masker level' $F(1,7)=135.48$ were highly significant ($p<0.001$). Pair wise comparisons with Bonferroni correction revealed that for all conditions and for both masker levels the signal levels at threshold were highly significantly different from each other ($p<0.001$), while condition S100-S5 was slightly less significant ($p=0.003$).

3.3.2 Loudness

Figure 3.3 shows individual MLMFs for the masker level of 50 dB SPL. The individual data for masker levels of 30, and 40 dB SPL are presented in the appendix. The level of the reference signal is plotted as a function of the matching level of the test signal. Again, the conditions in which the signal was prior to and simultaneous with the masker are represented by dashed and solid lines, respectively. The conditions with small and large onset delays are represented by triangles and squares, respectively. As for experiment 1, there were considerable individual differences. However, generally the trends were similar across subjects. For each subject, the levels of the test signals were highest for condition Δ+5 and these levels were almost independent of the reference level. For condition Δ-100, the levels of the test signal were lowest and were strongly dependent on the reference level. The results for condition Δ+100 at reference levels of 55 dB and above were mainly close to the results for condition Δ-100, except for subjects 1 and 7. This indicates that at those reference levels the loudness of the test signal was hardly influenced by the masker and resembled the loudness in quiet. At low reference levels the matching level of the test signal for the Δ+100 condition was higher than for the Δ-100 condition. This was caused by partial masking of the signal at these levels. This partial masking at low reference levels occurred for all masker levels.

The highest variability among subjects occurred for the condition Δ-5. For subjects 1, 4, 5 and 7, the levels of the masked signals were close to those for condition Δ+5. For subjects 6 and 8 the results fell between those for condition Δ+5 and those for conditions Δ+100 and Δ-100. For subjects 2 and 3, the results fell close to those for conditions Δ+100 and Δ-100. The variability between subjects in the Δ-5 condition was smaller for the other masker levels (see appendix A).

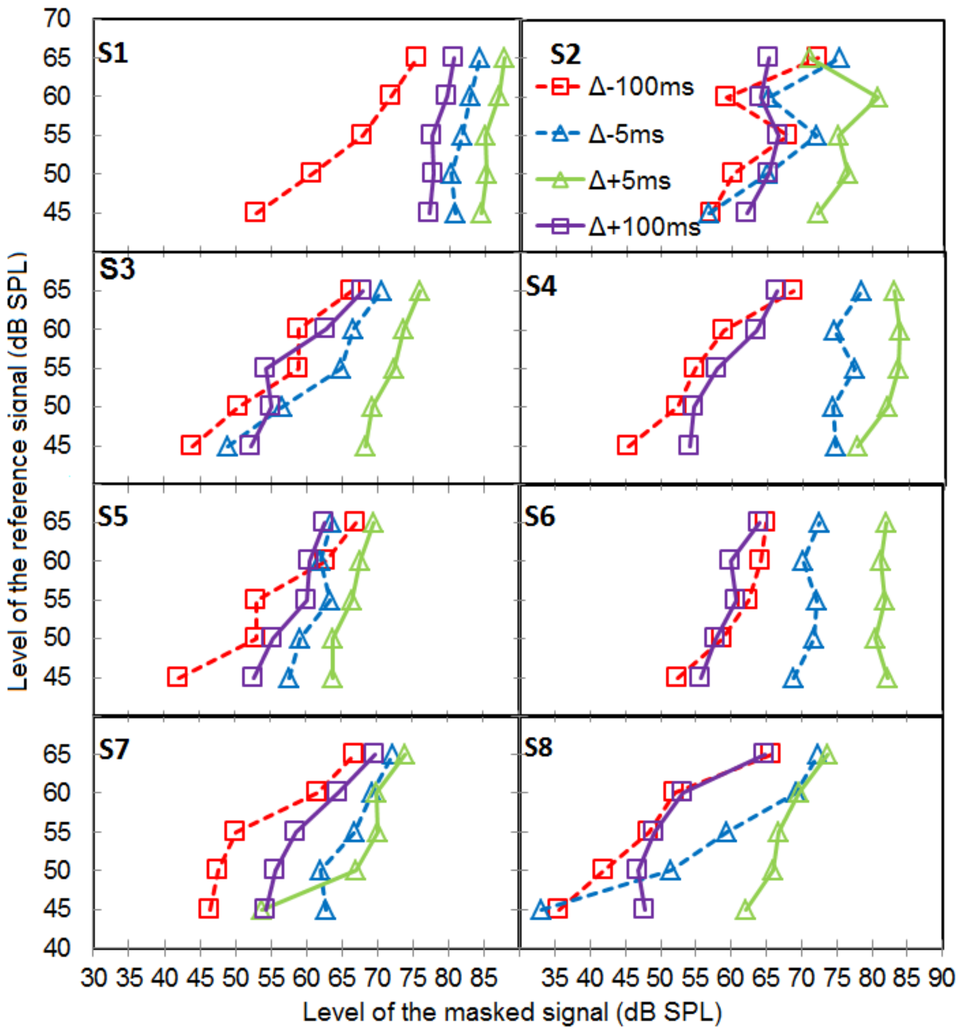


Figure 3.3. *Individual MLMFs for a masker level of 50 dB SPL.*

A comparison with the data for masker levels of 30 and 40 dB SPL also reveals that for condition Δ-5 the results show a different dependency on masker level than for conditionΔ+5. Where the MLMFs for conditionΔ+5 shift to the left with decreasing masker level, the MLMFs for condition Δ-5 appear to more or less independent from masker level.

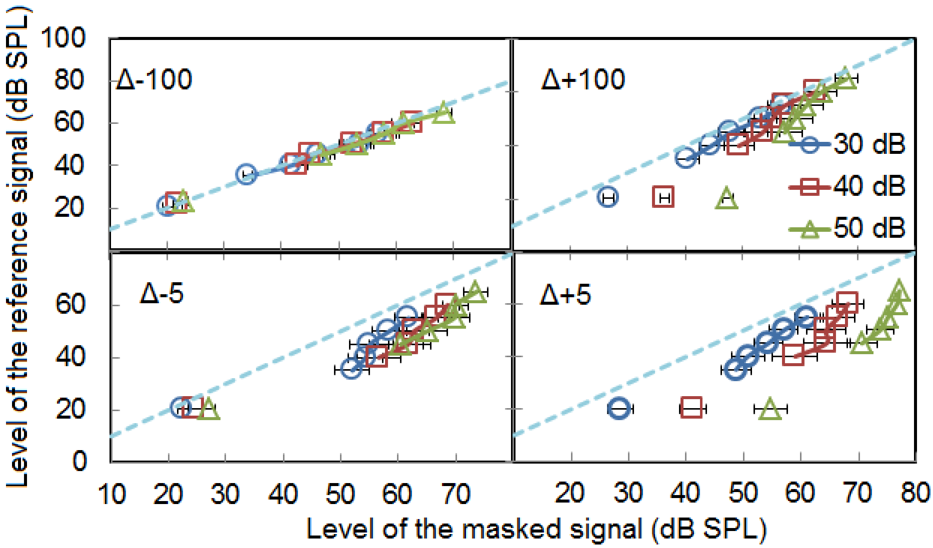


Figure 3.4. *Averaged masked loudness-matching functions (MLMF) for each condition for masker levels of 30 dB SPL (circles), 40 dB SPL (squares) and 50 dB SPL (triangles). The individual symbols at the left side of each plot show the threshold values for the signal in the noise masker as measured in the threshold experiment. The dashed line represents equal levels of the signal in noise and the signal in quiet.*

Figure 3.4 shows the mean data across all subjects (solid lines) separately for each condition. The dashed diagonal represents combinations of masked signals and reference signals that are predicted to have equal loudness at the same level. The main interest is the deviation from equal loudness at the same level. Therefore figure 3.5 represents the level differences between masked signal and reference signal as a function of the level difference between the reference signal and the masker. For condition Δ-100, the level of the masker did not influence the loudness of the test signal, and the curves for the different masker levels overlap. The loudness matches for condition Δ+100 were clearly influenced by the masker, as expected. For the 30 dB SPL masker, the loudness of the signal was only slightly influenced by the masker and the matches fell close to 0. For the 40 dB SPL and 50 dB SPL maskers the level of the test signal needed to be increased to match the loudness of the reference signal at low levels. This suggests that the loudness of the test signal was decreased by the noise, which is in agreement with earlier findings (Hellman and Zwislocki, 1964; Hellman, 1970; Pavel and Iverson 1981).

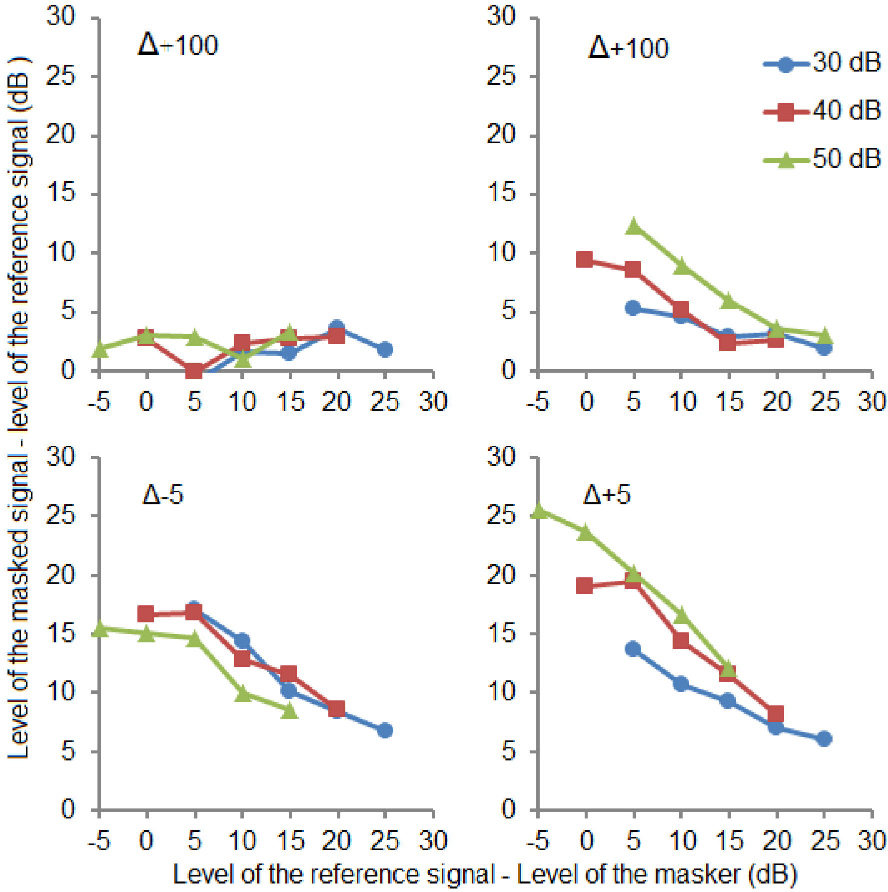


Figure 3.5. *Level differences between the masked signal and the reference signal as a function of the level difference between the reference signal and the masker for masker levels of 30 dB SPL (circles), 40 dB SPL (squares) and 50 dB SPL (triangles).*

The loudness growth functions for condition Δ+5 resemble the loudness growth functions for condition Δ+100, since the levels of the test signals needed to obtain equal loudness with the reference signals were increased at low levels. However, the effect was much larger, as though the masker had a much higher level. These findings support the hypothesis that an overshoot-like temporal effect occurs in the loudness domain, as the masker has a larger effect on the loudness of a masked signal at masker onset than later on in the masker. Remarkably, an overshoot-like temporal effect was also present for the signals presented just before the masker (Δ-5). The results for condition Δ-5 resemble those for condition Δ+5. However the dependency on masker level is different. Where for condition Δ+5 the temporal effect increases with increasing masker level, for condition Δ-5 the amount of temporal effect is roughly equal for all masker levels, and even appears to decrease slightly for the masker level of 50 dB SPL.

To assess the significance of differences between conditions a three-way repeated measures ANOVA was conducted in JASP (JASP Team, 2018) on the average values across all three retests with position of probe and masker ('condition', 4 categories), masker level ('masker', 3 levels), and reference signal level ('reference', 3 levels) as within-subjects repeated measures. In the analysis we only included the reference signal levels 45, 50, and 55 dB SPL, as these were the only levels measured at all masker levels. All main effects were highly significant ($p<0.001$), as were the interactions 'condition X masker' ($F(6,42)=13.46$, $p<0.001$), 'condition X reference($F(6,42)=6,12$, $p<0.001$), and the three-way interaction 'condition X masker X reference' ($F(12,84)=1.59$, $p<0.001$). Pair wise Bonferroni corrected comparisons revealed that for all masker levels and reference signal levels the test signal levels at equal loudness were significantly different from each other. Pair wise Bonferroni corrected comparisons further revealed significant differences between the test signal level at equal loudness for all conditions.

3.3.3 Comparison of the temporal effects for threshold and loudness

In the threshold experiment, subjects S3 and S7 showed no overshoot and subject S8 showed only a very small amount of overshoot. In the matching experiments this lack of an overshoot-like effect was not replicated. All three subjects showed results comparable to those for the other subjects with a comparable amount of a temporal effect in the loudness domain. For these subjects the levels at threshold may still have been on the linear part of the input-output function, where a temporal effect is not apparent, while the higher levels in the matching experiment may reach the compressive part of the input-output function, leading to a clear temporal effect.

Figure 3.6 compares temporal effects for threshold and loudness for conditions Δ+5 and Δ+100 (left panel) and for conditions Δ-5 and Δ-100 (right panel). Temporal effects were calculated as the test-signal level for the condition with the signal near masker onset minus the test-signal level for the condition with the signal well inside or well before the masker onset for each masker level. As the level range of the reference signals in quiet varied with masker level, some curves are composed of one or two points only.

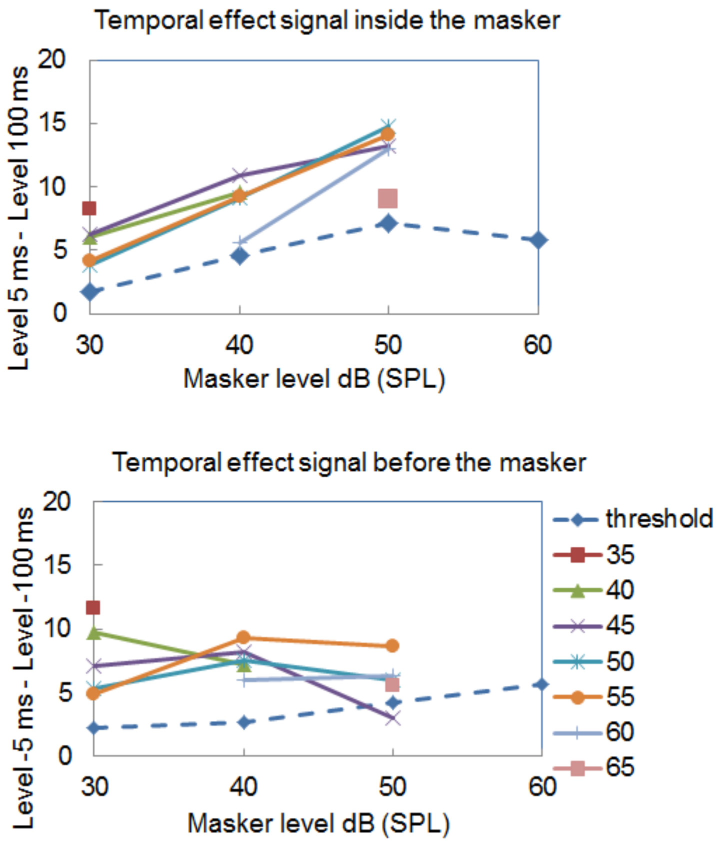


Figure 3.6. *Temporal effect calculated by subtracting the level of the signal presented far from masker onset from the level of the signal presented near masker onset. The left panel shows the results for simultaneous masking and the right panel the results for backward masking. The temporal effect is shown for the threshold experiment (dashed line) and the loudness experiment (solid lines for different masker levels).*

It is clear, however, that overshoot-like temporal effects were larger in the loudness domain than at threshold, and there is a trend for overshoot-like effects to increase with increasing masker level both at threshold and at supra-threshold levels. On the other hand, overshoot-like effects were reduced with increasing level of the reference signal, given a fixed masker level. This reduction in overshoot-like effects for higher signal levels is consistent with the hypothesis that overshoot originates from a different cochlear gain at masker onset with respect to masker center. The higher cochlear gain at onset is assumed to influence the input-output function mainly at lower input levels. At higher input levels the input-output functions with different cochlear gains are assumed to converge. Overshoot-like effects should therefore be weaker for high than for low input levels.

Overshoot-like effects for the backward masking conditions showed the same trends as for the simultaneous masking conditions. Overshoot-like effects were stronger for supra-threshold conditions than at threshold and increased with increasing masker level except for the signal reference level of 45 dB SPL. However, the level dependence was much stronger for backward masking than for simultaneous masking. Especially for the masker level of 30 dB SPL, overshoot-like effects were much larger at low than at high signal reference levels. Overshoot-like effects for condition Δ-5 can also be calculated relative to condition Δ+100. In this case the strength of the overshoot-like effects were comparable to the degree of overshoot in simultaneous masking for masker levels of 30 and 40 dB SPL, but considerably less for the 50 dB SPL masker level. This finding is in line with the observation that the MLMFs in condition Δ-5 were comparable to those in condition Δ+5 for masker levels of 30 and 40 dB SPL but not 50 dB SPL.

3.4 Discussion

3.4.1. Overshoot at threshold

Five of the eight subjects in this study showed clear overshoot at threshold. Overshoot was small at low levels and increased with increasing masker level. This corresponds to findings in the literature (Bacon, 1990; Bacon and Savel, 2004; Wright, 1995, 1997; von Klitzing and Kohlrausch, 1994; Strickland, 2004). From the literature it is known that the amount of overshoot varies considerably between subjects. It is somewhat remarkable that three out of eight subjects did not show clear overshoot, but in some other studies similar effects were found for individual subjects (Subject S4 in Bacon and Savel (2004); Subject S5 in Bacon (1990), Subject L5 in Strickland (2004)). As we tested more subjects than in most other studies on overshoot the chance of including subjects with no overshoot was higher. Inter-individual differences in overshoot may reflect a difference in efferent feedback by the medial olivocochlear bundle. A lack of efferent feedback might lead to an absence of overshoot.

3.4.2. Overshoot-like effects above threshold

Partial loudness has been investigated in several studies (Zwicker, 1963; Hellman and Zwislocki, 1964; Scharf, 1964; Stevens and Guirao, 1967; Hellman, 1970; Gockel *et al.*, 2003). In most studies the onset of the signal was later than the onset of the masker, to make it easier for subjects to "hear out" the signal. Our condition Δ+100 resembles these studies, except for the short duration of the signal. Except for Gockel *et al.* (2003) we are not aware of any studies that investigated partial loudness for a signal presented at the onset of the masker. Gockel *et al.* (2003) showed a small effect of onset delay (5.4 dB), but the duration of their signal was 700 ms, which is much longer than the 10-ms duration of our signal, which makes it hard to compare their data to our data, as temporal integration may influence the responses to the target signal differently in the two paradigms.

In our loudness matching data, the MLMFs for condition Δ+5 were higher than for condition Δ+100, which is consistent with our threshold data and with the literature on overshoot. The increased level of a brief tone presented shortly after the onset of a masker relative to the level of an equally loud brief tone presented later in the masker suggests a higher instantaneous loudness of the masker at masker onset with respect to the instantaneous loudness at the steady state of the masker. As indicated in the introduction, a higher instantaneous loudness of the masker at masker onset is compatible with a decrease in cochlear gain during the course of the masker, similar to the hypothesized origin of overshoot. The results for condition Δ-5 also show a temporal effect. However, the dependency on masker level for the Δ-5 condition is less pronounced than for the Δ+5 condition. This suggests that for the Δ-5 condition energetic masking does not play a major role.

At first glance, the results of the loudness matching experiment resemble those of experiments on loudness enhancement and loudness decrement (Elmasian and Galambos, 1975; 1980). Loudness enhancement is the effect that the loudness of a target signal is increased when the target signal is preceded or followed by a conditioner that is higher in level than the target. Loudness decrement is the effect that the loudness of a target is reduced if the conditioner level is lower in level than the target level. However, in those previous experiments the authors concluded that the influence of the conditioner (in our case the masker) on the matches to the target could be summarized by a single rule: judgments of the target were shifted in the direction of the conditioner. When the conditioner was more intense than the target, target loudness was increased, and when the conditioner was less intense, target loudness was decreased. The effect of a forward masker on intensity resolution has been shown to depend on the perceptual similarity between masker and standard (Schlauch *et al.*, 1997; Oberfeld, 2007; Oberfeld, 2008). The similarity hypothesis predicts that a reduction in the masker-standard similarity reduces the effect of the masker. As the masker in this study was a broadband signal and the target was a narrowband signal, loudness enhancement is not expected to have a large effect in our experiments. At least in the Δ-100 condition it is clear that loudness enhancement or decrement did not occur, as the loudness of test and reference signal were equal in this condition. For the Δ-5 condition it is unlikely that loudness enhancement or decrement played an important role. Assuming that the masking noise acted as the conditioner, higher masker levels (corresponding with a more intense conditioner) should lead to more loudness enhancement or less loudness decrement. However, with higher masker levels the loudness of the target tone was decreased. In this respect the relationship between target signal and masker for the Δ-5 condition resembles more closely the relationship found in simultaneous masking, where the loudness of the masker reduces the loudness of the signal.

3.4.3. Threshold and loudness results combined

The results presented in this paper clearly demonstrate that overshoot-like effects can also be observed at supra-threshold levels. Existing hypotheses to explain overshoot in detection may therefore also be relevant for loudness perception. However, an explanation in terms of a time-dependent gain factor as in the literature on overshoot cannot account for: 1) the fact that overshoot-like effects in the loudness domain were larger than overshoot at threshold and 2) the fact that the MLMFs for the Δ-5 condition were more or less independent from masker level.

Moore *et al.* (1997) assumed that sounds at absolute threshold have a small but finite loudness (Buus and Müsch, 2008). This loudness was assumed to be constant regardless of the frequency and spectral content of the signal. They also assumed that a sound at masked threshold has the same loudness as a sound at absolute threshold. In our study we did not measure loudness growth functions down to threshold. But if the above-mentioned assumptions hold true, with decreasing signal levels the loudness matches should converge on the masked threshold.

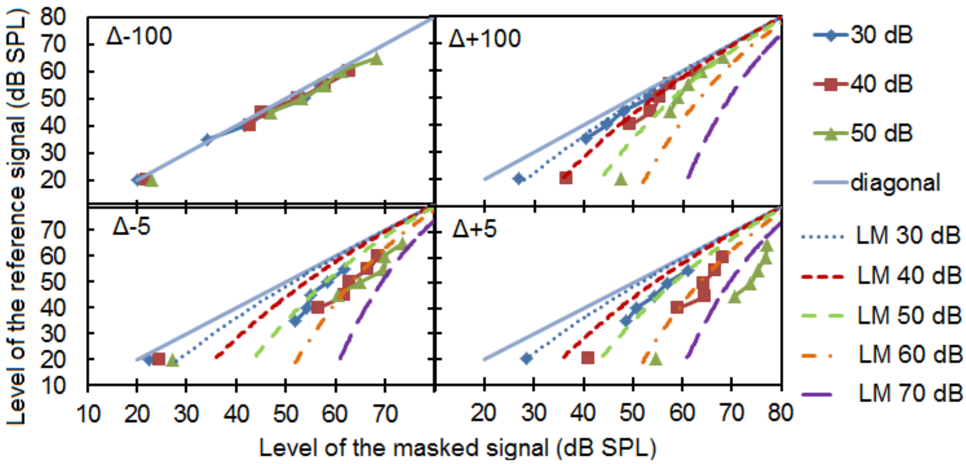


Figure 3.7. *Averaged MLMFs for each condition for masker levels of 30 dB SPL (circles), 40 dB SPL (squares) and 50 dB SPL (triangles). Dotted and dashed lines show predicted MLMFs for simultaneous masking for several masker levels.*

Figure 3.7 shows the results of this extrapolation. The threshold of the signal in quiet was assumed to be equal to the threshold in condition Δ-100. The dotted lines indicate short-term MLMFs calculated with the model for predicting the audibility of time-varying sounds in the presence of background sounds described by Glasberg and Moore (2005). The inputs to the model were a 4-kHz pure tone with a 10-ms duration as the signal and a 400-6000 Hz broadband noise as the masker, corresponding to the stimuli used in the experiments. As the onset delay of the signal with respect to masker onset did not influence the outcome of the model, arbitrarily the calculations were done with an onset delay of 5 ms. The model calculations serve to give an indication what a dynamic loudness model for partial loudness would predict for the shapes of the MLMFs for simultaneous masking at several masker levels. For condition Δ+100 the predicted MLMFs from the model are reasonably well in agreement with the measured partial loudness. For condition Δ+5 the measurements are more in line with MLMFs calculated with a higher masker level. Partial loudness at masker onset therefore resembles partial loudness of signals with a longer onset delay at higher masker levels. However, this does not necessarily mean that at masker onset the instantaneous loudness of the masker is indeed elevated. More central processing also seems to play a role, as will be discussed later. The model calculations for simultaneous masking are also shown in the panel for condition Δ-5, as for masker levels of 30 and 40 dB the slopes of the measured MLMFs correspond reasonably well with the modeled MLMFs. As for condition Δ+5 they follow MLMFs calculated for higher masker levels. Thus, although the model is not designed for this nonsimultaneous masking condition, it reasonably predicts the slopes of the MLMFs.

Threshold values for condition Δ-100 were comparable for all masker levels and were in line with the loudness matches. As expected, at Δ+100 the threshold values were influenced by masker level and the masked thresholds correspond closely to the modeled MLMFs, however, for the conditions with the signal near the onset of the masker the thresholds are not in line with the modeled MLMFs. The largest deviations between the measured thresholds and the thresholds predicted by the model occurred at Δ-5. Here the measured MLMFs were close to the modeled MLMFs. However, the measured thresholds were not close to the modeled thresholds and that yielded large level differences between measured thresholds and expected thresholds based on the modeled MLMFs. At Δ+5 the thresholds also lay at lower masked levels than expected from the modeled MLMFs. It appears that near masker onset the modeled MLMFs lead to threshold estimates that do not match the measured thresholds.

An explanation for the differences between the expected and measured masked signal thresholds near masker onset, might be that the subjects experience perceptual difficulties in separating the loudness of the signal and the masker. A similar effect was observed by Gockel *et al.* (2003). In their discussion of the onset effect they state “It was if part or all of the energy in the background was assigned to the target”. It appears that the near coincidence of signal and masker onset made it harder to differentiate the signal loudness from the masker loudness for the Δ-5 and the Δ+5 condition than for the Δ+100 condition. This hypothesis has also been postulated as an explanation for the gating effects in CMR. Hatch *et al.* (1995) have suggested that the smaller CMRs for gated maskers than for continuous maskers may be associated with competing cues related to auditory grouping. The synchronous onset may promote perceptual grouping of the signal and the masker, making it more difficult to separate the signal from the masker.
In the threshold task perceptual separation of the signal and the masker is not required. For threshold detection it is in principle enough if a change in the combined stimulus is detected. However, the loudness measurement task is much more demanding. In addition to detecting the signal, the subject has to separate the signal from the noise, has to identify the loudness of the signal and put a value on the loudness of the signal. In this higher-order task, subjects may be stronger influenced by the context (the loudness of the masker) and the degree of influence may be higher for a signal presented at masker onset than for a signal presented later.

The influence of noise conditions has also been widely investigated for intensity discrimination. Intensity discrimination is closely related to loudness, e.g. Oberfeld (2008) found a significant correlation between loudness changes and intensity-difference limens (DLs) in forward masking. In this study and several related studies on the effects of forward masking on intensity discrimination (Oberfeld and Stahn, 2012; Oberfeld *et al.* 2014) the hypothesis was postulated that the effects on intensity discrimination are determined by an inclusion of the masker intensities in the decision variable. The present study suggests that this hypothesis can be extended to the loudness domain. Especially at masker onset the loudness of the masker seems to be partly included in the loudness judgments of the signal. However, where in literature on loudness enhancement a conditioner enhances, or decreases the loudness of the target signal depending on whether it is stronger than the signal or not, in our results the masker always decreases the loudness of the target signal. As is the case in partial masking. As a consequence the intensity of the masked signal needs to be higher at equal loudness than the intensity of the reference signal.

As in loudness enhancement, the similarity effect is expected to be stronger when the perceptual distance between the signal and the masker is smaller (Schlauch *et al.*, 1997; Oberfeld, 2007; Oberfeld, 2008). Besides that the similarity effect may explain the onset effect (larger similarity as a result of almost simultaneous onset), it may also be the reason of a different dependency on masker level between the Δ-5 condition and the Δ+5 condition. In the Δ-5 condition the perceptual differences between the target signal and the masker increase with increasing masker level, making it easier for the subjects to separate the signal from the masker at a masker level of 50 dB SPL than at a masker level of 30 and 40 dB SPL. In the Δ+5 condition this effect is reduced by the simultaneous masking effect that greatly depends on masker level.

Also in overshoot experiments interactions between masker and signal beyond energetic masking have been shown. Bacon and Moore (1987) measured overshoot in a tone-on-tone simultaneous masking experiment and investigated the influence of transient responses on overshoot. They measured the level of a sinusoidal 1250-Hz masker at which the masker just masked a 20-ms 1000-Hz sinusoid presented at 10 dB SL. They found that the masker level needed to mask the signal was lower when the signal was presented at the onset of the masker than when it was presented at the temporal center of the masker. Although the experimental method deviates from the method used in most other studies on overshoot (Bacon *et al.*, 1988; Kimberley and Nelson, 1989; Bacon and Takahashi, 1992; Strickland and Krishnan, 2005), because the level of the masker was varied in this study and not the level of the signal, the results are consistent with these studies. Bacon and Moore (1987) also investigated the change in masker level produced by gating a 20-ms, 500-Hz transient masker with the signal, where the transient masker level was 30 dB below the level required to just mask the 10-dB SL signal. They showed that at least part of the elevation in threshold in the presence of a short-duration masker at the beginning of a longer duration masker may have been due to the transient responses to the masker affecting detection of the signal. This would mean that part of the effect did not originate from excitation in the signal "channel" due to the masker.

Further, Scharf *et al.* (2008) investigated the role of attention in an overshoot experiment in which they either randomized the signal frequency (frequency uncertainty) or kept it steady (frequency certainty), using a broadband and a narrow-band noise. They found that frequency uncertainty led to less overshoot for broadband noise, and to more overshoot for narrow-band noise. They explained this difference by the hypothesis that the listener can prepare for the signal frequency before any stimulation if the frequency is known beforehand. The onset of a broadband noise temporarily diverts attention to a wide range of frequencies, thereby raising the masked threshold for a signal close to noise onset. However, after 200 ms, focusing on the (known) signal frequency is reestablished and there is no extra threshold elevation. This difference in thresholds is overshoot. In contrast, the onset of a narrow-band noise that is centered on the same critical band as the signal does not displace focusing to frequencies away from the target. Consequently, detection is as good right after the onset of a narrow-band noise as a few hundred milliseconds later. No overshoot is present with frequency certainty for narrow-band noise. The experiments from Bacon and Moore (1987) and Sharf *et al.* (2008) show that a transient effect may play a role in overshoot experiments. Difficulties in the perceptual separation of a target from a masker at supra-threshold levels may be a consequence of a transient effect.

It would be interesting to measure partial loudness over a range of signal levels closer to masked threshold to see how the MLMFs behave near masked threshold. Would we see a sharp transition at lower levels towards threshold as subject S8 shows for the condition Δ-5 condition, where the MLMF was close to that for condition Δ+5 at high levels, but shifted to the MLFM for condition Δ-100 at low levels? Or would we still see a gap between the extrapolated threshold based on the matching data and the detection data, as seen for condition Δ+5? More research is needed to clarify this point and to investigate if the observed temporal effect at supra-threshold levels is also found in hearing-impaired subjects.

3.5 Conclusions

The results show that overshoot-like effects occur at supra-threshold levels. Near masker onset the equal loudness level of a short test signal relative to a reference signal in quiet is up to 15 dB higher than that of a short signal in the temporal center of the masker. The partial loudness of short signals at masker onset is reasonably well in agreement with the predictions of an existing loudness model with the assumption that the effective level of the masker is larger at onset than at steady state. Therefore, the results of this study are consistent with the hypothesis that the gain of the cochlear active process decreases during acoustic stimulation. The partial masking at supra-threshold levels appears to be more than energetic masking only. Especially at masker onset, subjects appear to have difficulty separating the signal from the masker, which leads to a stronger temporal effect in the loudness judgment task than in the threshold detection task. More research is needed to further investigate the effect of maskers on the loudness of signals near onset.

Appendix 3A: Individual masking data for 30 dB and 40 dB masker levels

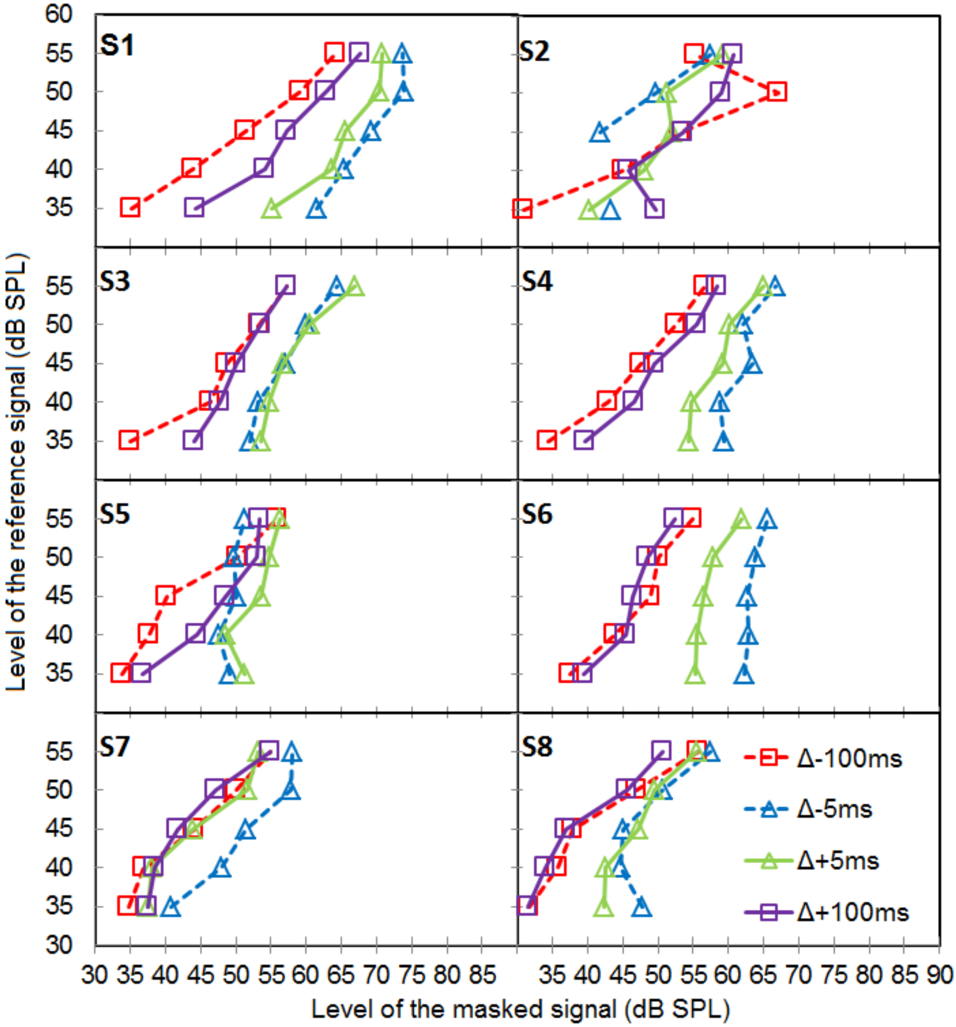


Figure 3A.1. *Individual MLMFs for the masker level of 30 dB SPL. Symbols are as in figure 3.1.*

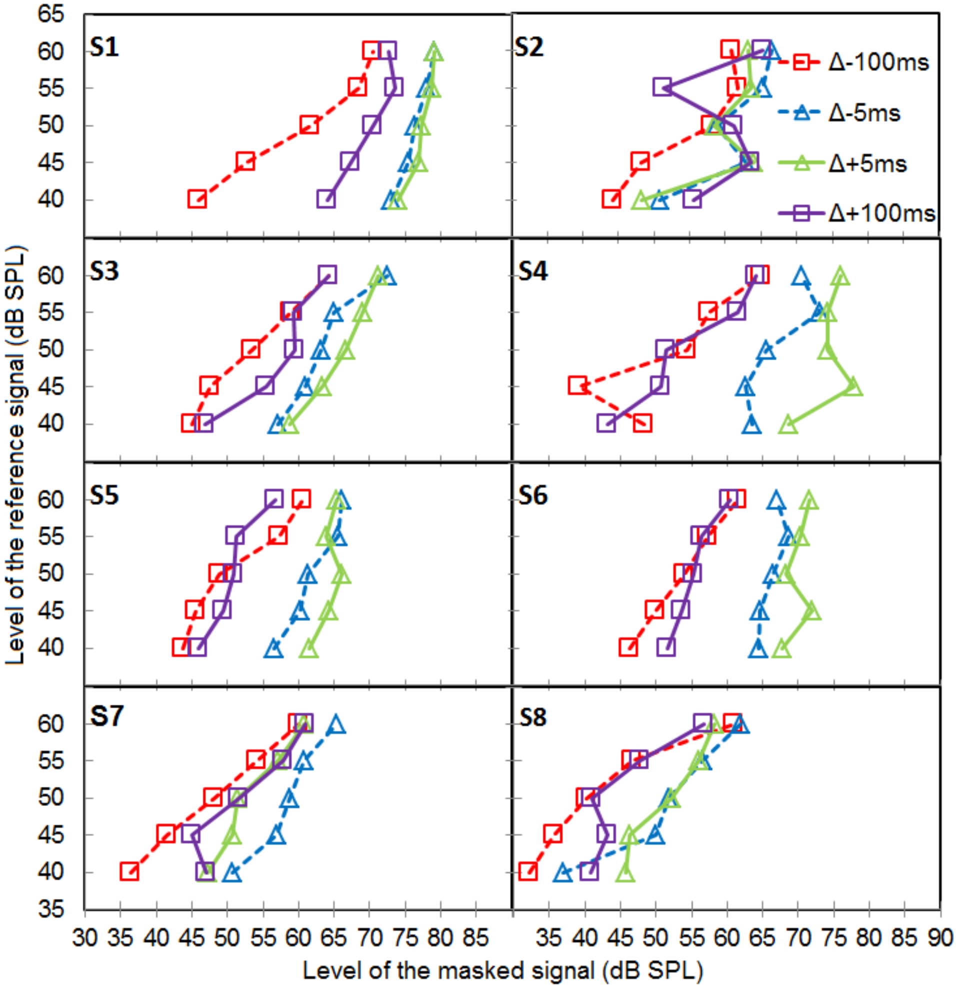


Figure. 3A.2. *Individual MLMFs for the masker level of 40 dB SPL. Symbols are as in figure 3.1.*

Chapter 4

Clinical applications of loudness scaling

Parts of this work has been published as:

Van Beurden, M.F.B., Boymans, M. Jansen, N., Dreschler, W.A. (2007). In Auditory signal processing in hearing-impaired listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR). Edited by T. Dau, J. M. Buchholz, J. M. Harte, T. U. Christiansen. ISBN: 87-990013-1-4.

Abstract

Fitting rules used in auditory rehabilitation usually have their main focus on detection thresholds. In state-of-the-art nonlinear hearing aids supra-threshold measures of hearing are also important and some of this information can be derived from loudness scaling.

In three studies we examined the added value of loudness scaling for clinical applications. In experiment 1 we performed loudness scaling in a group of musicians with primarily normal hearing. We measured loudness scaling with two narrowband signals (750 Hz and 3 kHz) and a broadband signal and investigated the relation between the form of the loudness curve and audiometric thresholds. In experiment 2 we examined the difference between monaural and binaural loudness perception in a subgroup of the musicians measured in experiment 1. Finally, in experiment 3 we examined the relationships between self-reported problems and measures obtained from loudness scaling in a different group of hearing impaired employees.
Our findings show that the form of the loudness curve cannot be well predicted from threshold measures alone. Both the slope of the upper part of the loudness curve and the level of uncomfortable loudness are not correlated to threshold. Obviously supra-threshold data on loudness perception contain information that cannot be derived directly from the pure-tone audiogram. This illustrates the added value of loudness scaling in the individual case.

Furthermore, our findings suggest that hearing loss mainly influences the slope of the lower part of the loudness curve and not the upper part. Binaural loudness summation on the other hand seems to mainly influence the upper part of the loudness curve and not the lower part. Finally, it is shown that very concave loudness functions may be associated with poorer speech understanding in noise.

4.1 Introduction

Fitting rules used in auditory rehabilitation are mainly based on measurements of the auditory thresholds and the uncomfortable loudness levels (UCL), i.e. DSL i/o (Cornelisse *et al.*, 1995), or even by measurements of auditory threshold only, i.e. NAL-NL1 (Byrne *et al.*, 2001). Based on these measurements the amount of gain and compression is selected. It would seem to be more appropriate to base the amount of gain and compression on a measurement of the complete shape of the loudness function rather than on a measurement of the extremes of the scale. However, measuring individual loudness functions is only interesting, when two conditions are fulfilled. First the reliability of the loudness function must be good enough to obtain reliable and reproducible individual differences. Second the loudness function must contain information that is not available from threshold measurements alone.

The present study is concerned with the second condition. The first condition has already been extensively investigated previously for several categorial loudness scaling methods (i.e.: Beattie *et al.*, 1997; Palmer and Lindley, 1998; Al-Salim *et al.*, 2010; Oetting *et al.*, 2014). These studies show that subjects can indeed make reliable loudness judgements with loudness scaling procedures. For instance, Al Salim *et al.* (2010) showed that correlations describing the reliability of mean stimulus-level within category exceeded 0.92 at all frequencies.

In this study we collected the results on loudness scaling obtained with the Adaptive CAtegorical LOudness Scaling (ACALOS) procedure (Brand and Hohmann, 2002) in three large studies on auditory performance. The parameters in each study varied according to the specific needs of the study.

The loudness scaling results of these studies were analyzed to investigate:

1. The shape of the loudness function and its relation to auditory threshold (experiment 1).
2. The difference between monaural and binaural measurements (experiment 2).
3. The relation between the loudness function and listening effort (experiment 3).

4.2 Methods

In the three experiments described both normal hearing and hearing impaired subjects participated. Subjects were defined as hearing impaired if one or more thresholds at the frequencies 500 Hz, 1000 Hz, 2000 Hz, or 4000 Hz exceeded 20 dB (HL). The major part of the subjects classified as hearing impaired had a mild high frequency hearing loss.
All measurements were conducted in a sound treated booth.

Experiment 1:
In this experiment 223 musicians from three orchestras participated. The group could be divided into 178 normal hearing and 45 hearing impaired subjects.
All subjects in experiment 1 performed ACALOS tests for three signals. Two signals were narrowband noises (1/3 octave-band) at 750 Hz and 3 kHz and the third signal was a broadband white noise. The signals were delivered by a loudspeaker at a distance of approximately 1 m. The hearing impaired subjects performed this experiment without hearing aids.

Experiment 2:
In this experiment a subgroup of 52 musicians from experiment 1 participated. The subgroup consisted out of 48 normal hearing subjects and 4 hearing impaired subjects. In experiment 2 the subjects performed ACALOS at 1 kHz and 4 kHz (1/3 octave-band noises) for monaural and binaural presentations. For the binaural measurements the same signal was presented at both ears (diotic presentation). The signals were delivered by TDH 39-headphones.

Experiment 3:
In this experiment 14 hearing impaired employees of a printing office performed the same tests as in experiment 1. The subjects performed ACALOS tests for three signals, 1/3 octave-band noises at 750 Hz and 3 kHz and a broadband white noise. The signals were delivered by a loudspeaker at a distance of approximately 1 m. The 14 subjects in this experiment also completed a questionnaire on listening effort in daily life (SSQ, part 3; Gatehouse and Noble, 2004). The questionnaire contained questions about

listening effort in silence and in noise. The questionnaire consisted out of four possible answers; 'no effort', 'little effort', 'moderate effort' and 'high effort'. Six subjects performed the tests without hearing aids, the other eight subjects performed the tests with hearing aids, corresponding to their daily practice.

4.3 Procedures

The loudness scaling procedure used was the Adaptive CAtegorical LOudness Scaling (ACALOS) procedure designed in Oldenburg by Brand and Hohmann (2002). This is a loudness scaling procedure with 11 response categories, 5 named categories, 4 un-named intermediate categories, and 2 limiting categories, which correspond to categorical loudness levels from 0 to 50. The level assigned to a given loudness category x is termed the "categorical loudness level" Lx. An example of the response scale is given in figure 4.1.

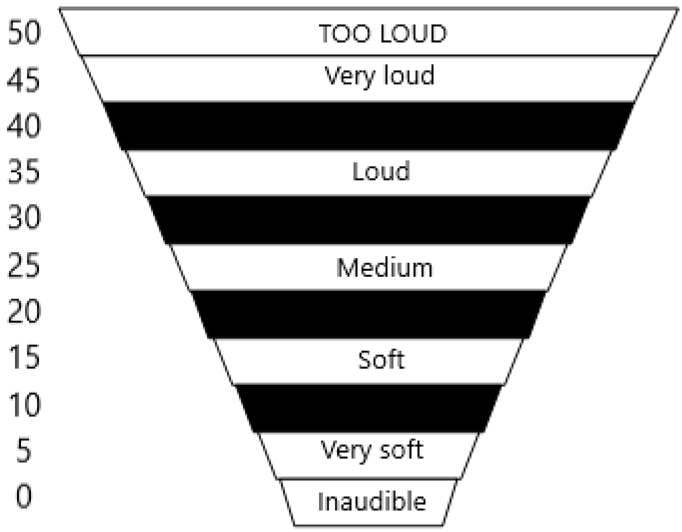


Figure 4.1. *Response scale, consisting of 11 response alternatives. The numbers on the left side indicate the categorical (units). They were not visible to the subjects during the tests.*

The procedure consists out of two phases. In the first phase the limits of the auditory range are estimated by an interleaved ascending and descending stimulus sequence. In the second phase the four named intermediate categorical loudness levels are estimated. This last phase consists out of two blocks. In the first block the four named intermediate categorical loudness levels are estimated by linear interpolation between the two limits of the auditory range, which are the values at L5 (very soft) and L50 (too loud). In the second block the named intermediate categorical loudness levels are estimated by a modified least-squares fit of a linear model function. In this study three iterations of the final block have been applied. The data is fitted with a model function consisting of two linear parts with independent slopes m_{low} and m_{high}. The two parts are connected at 25 CU. The transition area between the loudness categories L_{15} and L_{35} is smoothed with a Bezier fit Brand and Hohmann, 2002).

4.4 Results

4.4.1 Experiment 1: Loudness scaling in freefield

The data of the normal hearing and hearing-impaired subjects were analysed separately. In figure 4.2 the mean fit and 5th and 95th percentile are presented for the normal hearing listeners. The spread in the data is large for all three signals. In the loudness scaling data the level at 5 CU was defined as the threshold value of that signal.
Correlations were calculated between the thresholds estimated with loudness scaling (CU5) and several pure tone averages ($PTA_{0.5,\ 1\ kHz}$, $PTA_{2,3,4,kHz}$, $PTA_{0.5,\ 1,2,3,4\ kHz}$) obtained from pure tone audiometry. Pure tone averages were chosen for correlation, because the frequencies of the stimuli used in the loudness scaling experiment did not correspond to frequencies measured in pure tone audiometry and because the stimuli in the loudness scaling experiment were 1/3-octave narrowband noises and a wideband noise and not pure tones.

For the normal hearing subjects no significant correlations were found. Correlations between the levels at CU5 and CU50 were also not significant. This indicates that for normal hearing subjects no strong relationship exists between the audiometric thresholds and the thresholds obtained from loudness scaling.

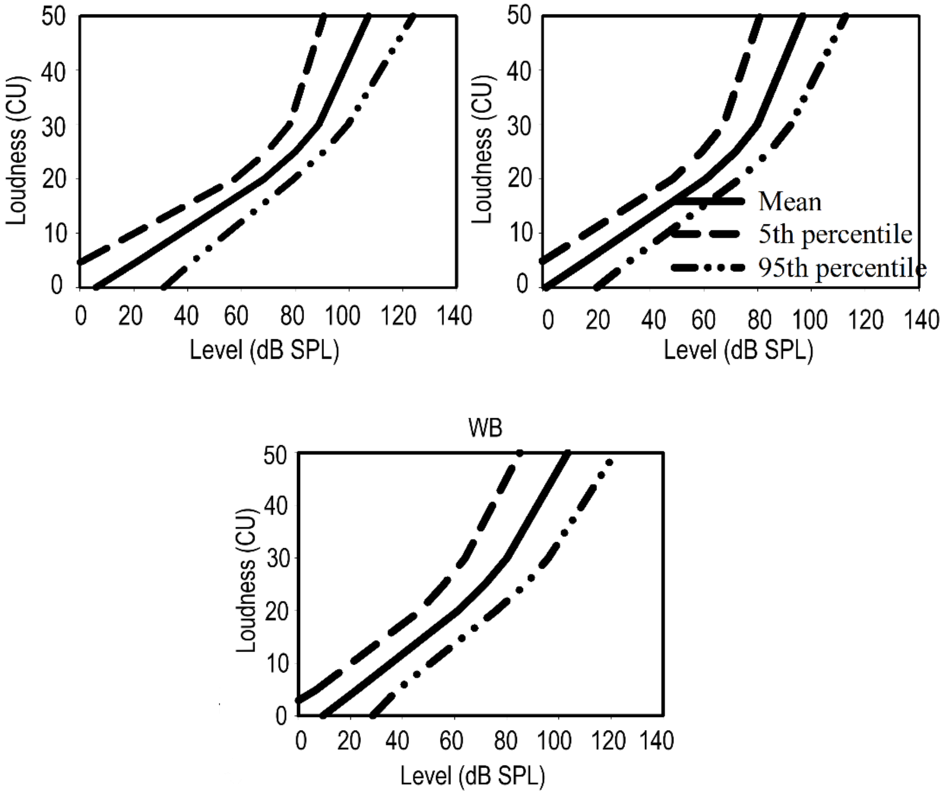


Figure 4.2. *Mean loudness functions for normal hearing subjects and the 5th and 95th percentile ranges for 1/3 octave bands around 750 and 3000 Hz and for a wideband noise.*

However, for the group of hearing-impaired subjects significant correlations were found ($p<0.01$) between CU5 values and pure tone averages. The correlations are shown in table 4.1. Again, no significant correlation between CU5 and CU50 values was found.

Table 4.1. *Correlations between audiometric thresholds and thresholds obtained from loudness scaling (CU5) for hearing impaired subjects.*

Audiometric threshold	Threshold ACALOS	Correlation
$PTA_{0.5,1\ kHz}$	$CU5_{750Hz}$	0.55 (0.000)
$PTA_{2,3,4\ kHz}$	$CU5_{3kHz}$	0.65 (0.000)
$PTA_{0.5,1,2,3,4\ kHz}$	$CU5_{WB}$	0.39 (0.008)

For the total group of normal hearing and hearing-impaired subjects correlations were calculated between the dynamic range (DR) as defined by the level difference between CU50 and CU5 and the values characterizing the steepness of the loudness function: m_{low} and m_{high}. Correlations are shown in table 4.2. All correlations were significant ($p<0.05$). The dynamic range is stronger correlated to m_{low} than to m_{high}, as would be expected as generally the slope described by m_{low} is shallower than the slope described by m_{high}. There is a significant correlation between m_{low} and CU5 (0.61, $p<0.001$). The lower slope is therefore for a large part determined by CU5. There was no significant correlation between m_{high} and audiometric thresholds or CU5. This suggests that the upper part of the loudness curve is more or less independent of threshold.

Table 4.2. *Correlations between lower and upper slopes of the loudness function and the dynamic range DR (CU50-CU5).*

m_{low}	DR_{750Hz}	-0.66 (0.000)
	DR_{3kHz}	-0.50 (0.000)
	DR_{WB}	-0.65 (0.000)
m_{high}	DR_{750Hz}	-0.45 (0.000)
	DR_{3kHz}	-0.31 (0.000)
	DR_{WB}	-0.36 (0.000)

4.4.2 Experiment 2: Loudness scaling with headphones

In Fig. 4.3 average loudness curves are shown for monaural and binaural measurements at 1 and 4 kHz. The results show that binaural signals presented at equal levels are perceived louder than monaural signals for higher levels. This effect is somewhat stronger at 1 kHz than at 4 kHz. A paired T-test shows significant differences in level between monaural and binaural measurements for CU20, CU25, CU30 and CU50, but not for CU5. This holds for both the 1 kHz and the 4 kHz stimuli. Level differences between measurements at the right and left ear were not significant for 4 kHz and 1 kHz except for CU25 and CU30 at 1 kHz.

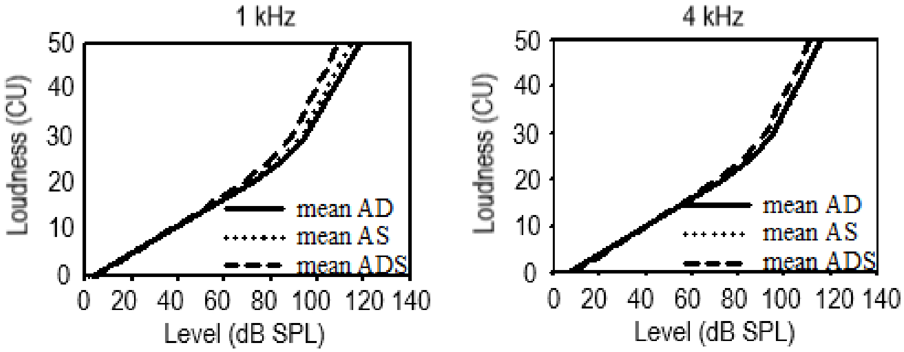


Figure 4.3. *Average loudness functions for the right ear (AD), the left ear (AS) and binaural (ADS) for the test frequencies 1 kHz and 4 kHz.*

4.4.3 Experiment 3: Loudness scaling aided and unaided

In this experiment the outcome of a questionnaire on listening effort in daily life was correlated with measures obtained from loudness scaling. The highest correlations were obtained with the ratio of the slope values m_{high} and m_{low}. This ratio is 1 if the loudness curve is linear, >1 if the loudness curve is concave and <1 if the curve is convex. Figure 4.4 shows the effort of listening in noise versus the ratio m_{high}/m_{low}.

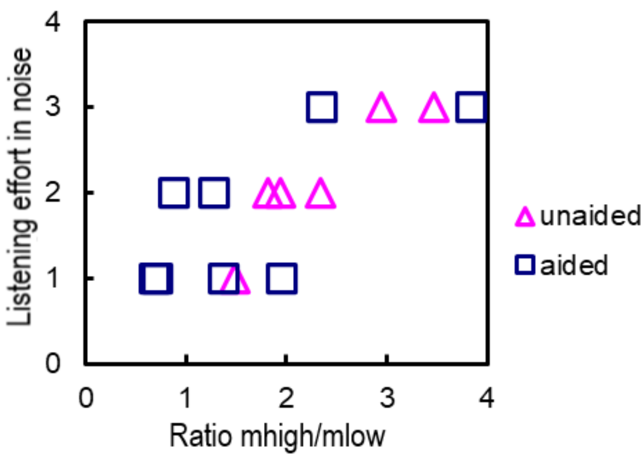


Figure 4.4. *Listening effort in noise as a function of the ratio m_{high}/m_{low}. The numbers correspond to: 0 "no effort", 1 "little effort", 2 "moderate effort" and 3 "high effort".*

The ratios between mhigh and mlow show that most loudness functions are concave. The relationships with listening effort data suggest that very concave loudness functions increase listening effort in noise, both in unaided and aided conditions.

4.5 Discussion

In the experiments described above we measured loudness scaling for normal hearing subjects and hearing-impaired subjects, with headphones and in free-field, aided and unaided. In all studies we evaluated the extra information contributed by loudness scaling.

4.5.1 Experiment 1

The results of experiment 1 show that it is hard to define one single normal loudness function. The spread in loudness functions within the group of normal hearing subjects is large. This result is in line with the data presented in ISO 16832 (2006) where in a normal hearing population the intensity levels at a particular loudness level differ as much as 20-30 dB.

In our group of mostly normal hearing listeners the correlations between audiometric thresholds and the thresholds from loudness scaling are weak. This is a consequence of the shallow slopes of the loudness function at low levels in the normal hearing subjects. The shallower the slope of the loudness curve becomes, the larger impact a small change in slope has on the threshold estimate. As a result of the shallow slopes near threshold, the threshold estimates based on loudness scaling vary considerably more than the threshold estimates based on audiometric measurements.

Brand and Hohmann (2002) already noticed that some normal hearing listeners reported that they were forced to respond less accurate than they could, especially at low levels. This is a consequence of the choice of a limited response scale. Normal hearing listeners have a large range of intensities in which they estimate a signal to be soft (very soft to medium). As can be seen in figures 4.2 and 4.3 this range can encompass 60-80 dB. For this entire range the normal hearing listeners only have five response alternatives. This means that every response alternative encompasses an intensity range of over 10 dB, while the just noticeable difference is approximately 1-2 dB. This leads to the problem that a listener can hear a loudness difference, but cannot report the difference as the response alternatives are too limited. In this and other studies performed in our lab, normal hearing subjects reported the same problem. The lack of correlations between the audiometric thresholds and the loudness scaling based thresholds is therefore a reminder, that the fitted loudness function is only an approximation of the real loudness curve.

In the group of hearing-impaired listeners the correlations between audiometric threshold and threshold extrapolated from the loudness function increase. As the slope of the lower part of the loudness function steepens with hearing loss, the sensitivity to small differences in the slope of the loudness curve decreases. This decreases the variability in the threshold estimates from loudness scaling. Together with the fact that the hearing-impaired subjects form a more heterogeneous group this leads to higher correlations between audiometric thresholds and thresholds from loudness scaling in the group of hearing-impaired subjects.

The correlations are not very strong, but nevertheless they imply that in hearing-impaired listeners thresholds based on loudness scaling are related to the thresholds from standard audiometric testing. Note that the correlation between audiometric thresholds and estimated thresholds from loudness scaling is highest for 3 kHz. This is the frequency were the largest variation in hearing losses is found.

This is in line with data from Al-Salim *et al.* (2010). In a group of primarily hearing-impaired listeners with a larger range of hearing losses than in our study, Al-Salim *et al.* (2010) found that the slope of the low-level portion of the categorical loudness scaling function varied in a predictable manner with audiometric threshold, with slope increasing as audiometric threshold increased.

The group results lead to another interesting finding. Hearing loss in this population mainly influences the lower part of the loudness function. The higher part of the loudness function seems to be more or less independent from both audiometric thresholds and thresholds obtained from loudness scaling. The results in this study strongly support the notion that recruitment is mainly focused on low and medium levels. The upper parts of the loudness curve show no consistent steepening with increasing threshold and therefore show no signs of recruitment. This suggests that for normalization of loudness, compression should be applied mainly for the lower levels and linear amplification at high levels, as otherwise the shape of the normal loudness function will be distorted.

4.5.2 Experiment 2

The results of the second study show a clear difference between monaural and binaural loudness measurements. Binaurally presented signals are clearly perceived louder than monaurally presented signals. It is interesting to see that this effect mainly occurs at the higher levels. At lower levels no binaural summation was found. This is in line with a few other studies that show more binaural loudness summation at high levels than at low levels (Reynolds and Stevens, 1960; Scharf and Fishken, 1970; Whilby et al., 2006; Zwicker and Zwicker, 1991). However, data from for instance Marks (1978) do not show any effect of level on binaural loudness summation. The binaural loudness data may have implications for hearing aid fitting. In hearing aid fitting loudness normalization is often one of the main targets. However hearing aid fitting targets are commonly per ear and not bilateral. If binaural loudness summation is indeed level dependent this should be taken in consideration in the hearing aid fitting.

4.5.3 Experiment 3

The third and final study was done in a small and very heterogeneous group of subjects. Therefore no strong conclusions may be drawn from this study. The fact that the relationship between linearity of the loudness function and listening effort in noise also appears in the unaided measurements, shows at least that the effect is not created by inadequate hearing aid fitting. On the other hand if the relationship between high listening effort and very concave loudness functions can be confirmed in follow-up studies, this knowledge may have important consequences for hearing aid fitting. It would be very interesting to investigate if reducing the concaveness of a loudness function would decrease listening effort. In that case compression schemes should be adapted to avoid high levels of compression. In some cases one may even consider to apply expansion.

4.6 Conclusions

In conclusion, these studies show that loudness scaling can give us more insight in loudness perception:

- The spread in loudness functions in normal hearing subjects is large. It is therefore questionable if it is appropriate to take the average curve for normal hearing listeners as the ultimate reference for loudness perception. It could be important to pay more attention to individual loudness curves in order to obtain normal loudness perception for the individual.
- Threshold estimates from loudness scaling do not correlate well with audiometric thresholds. At least for normal hearing listeners threshold estimates based on loudness scaling are much less accurate than standard audiometry.
- In hearing-impaired listeners the threshold estimates based on loudness scaling correlate better with thresholds measured obtained by standard audiometry than in normal hearing subjects. The correlations increase at higher hearing losses.
- Hearing impairment hardly affects the upper part of the loudness curve (mhigh), The most dominant changes due to hearing impairment are found in the lower part of the loudness curve.
- Binaural loudness summation seems to occur mainly at the upper part of the loudness curve. The loudness function is steeper for binaurally presented signals.
- We found a weak tendency that very concave loudness functions may be associated with higher listening effort.

Further studies should investigate in more detail how the loudness function is changed by hearing impairment. Points of interest are:

- The influence of more severe hearing impairment on the shape of the loudness function. In experiment 1 only mild hearing losses were included. For these hearing losses m_{high} did not change with hearing loss. For hearing aid fitting it is important to know if this also holds for higher hearing losses.
- The influence of hearing impairment on binaural loudness summation. In experiment 2 binaural loudness was only present at higher levels. Is this also the case for more severe hearing losses?

- Is there indeed a relation between the concaveness of the loudness function and the amount of listening effort? And does this relationship also holds for more severe hearing losses?

Acknowledgements

E.J.M. Jansen contributed to the conference paper that formed the basis of this chapter.

Chapter 5

Potential consequences of spectral and binaural loudness summation for bilateral hearing aid fitting

Van Beurden, M.F.B., Boymans, M., van Geleuken, M., Oetting, D., Kollmeier, B., Dreschler, W.A. (2018). Potential consequences of spectral and binaural summation for bilateral hearing aid fitting. Trends Hear. 22, 1-15.

Abstract

Aversiveness of loud sounds is a frequent complaint by hearing-aid users, especially when fitted bilaterally. This study investigates whether loudness summation can be held responsible for this finding. Two aspects of loudness summation should be taken into account: spectral loudness summation for broadband signals and binaural loudness summation for signals that are presented binaurally. In this study, the effect of different symmetrical hearing losses was studied. Measurements were obtained with the widely used technique of Adaptive Categorical Loudness Scaling. For large bandwidths, spectral loudness summation for hearing-impaired listeners was found to be greater than that for normal hearing listeners, both for monaurally and binaurally presented signals. For binaural loudness summation, the effect of hearing loss was not significant. In all cases, individual differences were substantial.

5.1 Introduction

Nowadays, the majority of listeners with hearing loss (HL) are fitted bilaterally. The use of two hearing aids has increased over the last decades and reached values of about 75% in the US (Kochkin, 2009) and about 70% in Europe (see EuroTrak Germany 2018). Bilaterally fitted hearing aids have been shown to improve speech intelligibility both in quiet and in noise and to improve localization (Boymans *et al.* 2008; 2009; Köbler and Rosenhall, 2002; Noble and Gatehouse 2006). However, aversiveness of loud sounds remains a problem. In several studies on the benefit of hearing aids, aversiveness of sounds has been found to be negatively influenced by hearing aid fitting (Abrams *et al*., 2012; Cox *et al*., 2011; Löhler *et al*., 2016), an effect that in bilaterally fitted subjects might be stronger than in unilaterally fitted subjects (Boymans *et al.,* 2009). Loudness complaints remain a major reason for revisiting the hearing aid dispenser (Jenstad *et al.,* 2003), and aversiveness of loud sounds is one of the main reasons to be dissatisfied with a hearing aid fitting (Hickson *et al.,* 2010). Discomfort of loud sound and its importance for hearing aid fitting has been extensively investigated (e.g.: Formby *et al*., 2017; Hawley *et al.*, 2017; Mueller and Bentler, 2005). The relationship between measured loudness discomfort levels and ratings for satisfaction, however, is weak (Zaugg *et al.,* 2016), and the loudness judgments within the same loudness category varied across listeners within a group by as much as 50-60 dB (Formby *et al.,* 2017).

It is generally accepted that hearing aid rehabilitation involves successive steps, starting with a first fit based on a prescriptive formula, followed by individual fine tuning based on subjective responses or technical measurements using in situ responses. Over the years, a number of prescriptive formulas have been developed. The linear prescriptive formulas (e.g., NAL-R, FIG6, POGO) have been replaced by non-linear prescriptions, such as NAL-NL2 (Dillon, 2012) and DSL I/O (Cornelisse et. Al. 1995, Bagatto *et al.* 2005, Scollie *et al.* 2005), taking into account that the amount of gain required is not only frequency dependent but also level dependent.

Nonlinear fitting formulas show some relationship with the loudness growth at different frequencies. The level of detail of knowledge about loudness perception required for an effective first-fit setting is still in debate. But the dynamic range as the frequency-dependent range between the individual hearing thresholds and the levels of uncomfortable loudness is generally accepted and applied in different forms in nonlinear prescriptive formulas.

Due to the fact that hearing loss is often strongly frequency dependent, loudness growth is usually measured with narrowband signals. Loudness curves measured in individual hearing-impaired (HI) listeners can be compared with loudness curves of normal hearing (NH) listeners and thus transferred into level-dependent gain prescriptions for hearing aid amplification settings to normalize loudness (Herzke and Hohmann, 2005).

However, in this approach, two aspects of loudness perception are not taken into account: spectral loudness summation (in case of the presentation of broadband signals instead of narrowband signals) and binaural loudness summation (in case of bilateral presentation instead of unilateral). This includes also the binaural loudness perception of broadband signals. This combined effect has to be considered because often two hearing aids are worn and they will typically process broadband signals as speech or environmental sounds.

These types of loudness summation may require individual corrections. Recent data of hearing-impaired listeners (Oetting *et al.* 2016) showed large individual differences in spectral loudness summation and binaural loudness summation after careful narrowband loudness normalization. Some of the listeners showed loudness perception for binaural broadband signals that was fully in agreement with normal hearing reference data, whereas others showed a higher-than-normal loudness summation of up to 30 dB SPL for the binaurally presented broad-band signals. Given the magnitude of the inter-individual differences found, it can be assumed that these findings are relevant for loudness adjustments during bilateral hearing aid fittings.

In this study we measured spectral and binaural loudness summation, separately as well as the combination for binaurally presented sounds using categorical loudness scaling (Brand and Hohmann, 2002). In a study by Oetting *et al.* (2016), mild to moderate hearing losses were tested corresponding to audiometric configurations of N1-N3 and S1 (Bisgaard *et al.* 2010). It is not clear whether the effect of individual variation decreases, remains constant, or increases with increasing hearing loss. Therefore, in this study, a broader range of hearing losses (audiometric configurations: N2-N4 and S2-S3; Bisgaard *et al.*, 2010) were included. The focus was on a larger variety of hearing losses and a potential effect of the Bisgaard *et al.*'s (2010) classification on the individual variation. The main questions of the current study are (a) whether the shape of the audiogram can explain individual differences and (b) if several characteristics of the hearing loss and hearing loss compensation strategy are possible predictors for the amount of spectral and binaural loudness summation.

5.2 Methods

5.2.1 Subjects

The inclusion criteria were age above 18 years and native Dutch speakers with mild to moderate symmetrical hearing losses (differences between both ears at 0.5, 1, 2 and 4 kHz < 10dB) selected from clinical files. Their pure-tone audiograms were classified according to the 10 standard audiograms N1 to N7 and S1 to S3, as suggested by Bisgaard *et al.* (2010). Bigaard *et al.* (2010) defined typical audiograms that cover the entire range of audiograms met in clinical practice. Their classification consists of seven audiograms for flat and moderately sloping hearing loss (N1-N7) and three audiograms for steep hearing loss (S1-S3), with higher numbers corresponding to greater hearing loss. The individual audiogram was taken as the average audiogram of the right and left ear. The classification was based on the lowest root-mean-square error of the individual audiogram and the standard audiograms.

Thirteen women and sixteen men participated with an average age of 69 and a standard deviation of 4 years. Twenty-two listeners had a flat or moderately sloping audiogram classified N2 (9), N3 (10), or N4 (3). Seven listeners had steep sloping audiograms classified as S2 (4) or S3 (3). For reasons of comparison, reference data from 9 normal hearing (NH) listeners measured by Oetting *et al.* (2016) were used. Because of the small numbers in the S2 and S3 groups, these 7 listeners were taken together in one group with steep sloping losses: group S. The distributions of the hearing losses for each standard audiogram are given in figure 5.1, whiskers mark minimum and maximum values.

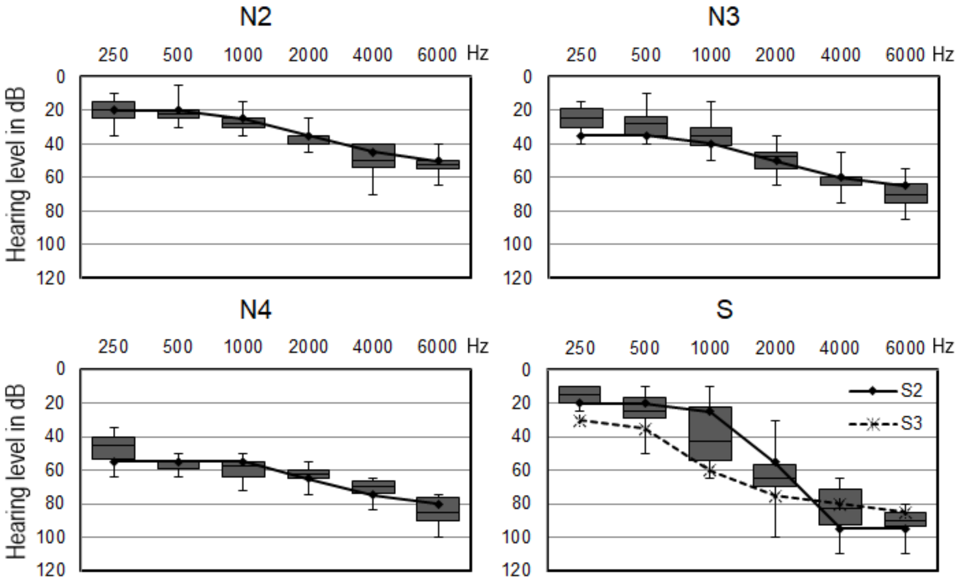


Figure 5.1. *The distribution of the audiograms for each standard according to Bisgaard et al. (2010). Whiskers mark minimum and maximum values.*

5.2.2 Equipment

All measurements were conducted in a sound-insulated booth in two sessions of about 2 hours each. Pure-tone audiograms (air and bone conduction) were measured with DECOS audiometers, using TDH39 headphones. Sennheiser HAD 200 headphones were used for the loudness categorical loudness scaling procedure using the framework for psychoacoustic experiments (Ewert, 2013). Signals were presented using a RME Fireface UC at 44.1 kHz. Headphones were calibrated with a Brüel & Kjær artificial ear type 4153, a 0.5-inch microphone type 4134, a microphone preamplifier type 2669, and a measuring amplifier type 2610. Headphones were free-field equalized according to ISO 389-8 (2004) and levels are expressed as the equivalent free-field level in dB SPL(FF).

5.2.3 Procedure

5.2.3.1 Loudness Scaling

Categorical loudness scaling was performed to measure the individual loudness perception. During the loudness scaling procedure, listeners had to rate the perceived loudness on an 11-point scale from "not heard" to "too loud", which were transformed into numerical values in "Categorical Units" (CUs) from 0 to 50. Stimuli were presented in a pseudorandom order with levels between -10 and 105 dB HL. A monotonically increasing loudness function was fitted to the responses for each of the Adaptive Categorical Loudness Scaling measurements using the BTUX fitting method (Oetting *et al.*, 2014). The model function consists of two linear parts with independent slopes m_{low} and m_{high} with a smooth transition range (see Brand and Hohmann, 2002).

5.2.3.2 Stimuli

Two types of signals were used:
In part I, loudness functions in different frequency regions were assessed with narrowband stimuli. For this purpose, one-third octave low-noise noises (Kohlrausch *et al.*, 1997) were used. These narrowband stimuli had center frequencies of 250, 500, 1000, 2000, 4000, and 6000 Hz.
In part II, loudness summation effects were assessed. For these experiments, stimuli that consisted of uniformly exciting noise (UEN; Fastl and Zwicker, 2007) with bandwidths of 1, 5, and 17 Barks were used, referred to as UEN1 (bandwidth: 210 Hz), UEN5 (1080 Hz) and UEN17 (5100 Hz), respectively. The UEN noises were centered on the Barkscale at 10.5 Bark (1370 Hz) and were designed so that each Bark band had equal signal energy.

In addition to the UEN a speech-shaped noise, referred to as IFnoise (International Female noise, Holube, 2011), was included in the test battery. The IFnoise was generated to match the spectral shape of the long-term average speech spectrum for females (Byrne *et al.*, 1994).
All stimuli were 1-s noises with 50-ms rise and fall ramps and identical to the stimuli used by Oetting *et al.* (2016).

5.2.3.3 Part I: Narrowband loudness functions

In part I of the measurements, the narrowband LNN-signals were presented monaurally to the right and left ears with randomized order of the test frequencies. Hearing-impaired listeners with an even number started the monaural conditions with the right ear, whereas hearing-impaired listeners with an uneven number started the monaural conditions with the left ear.

5.2.3.4 Narrowband loudness normalization.

Before loudness summation was determined for the broadband signals (see part II) the UEN1, UEN5, UEN17, and IFnoise, the noises were corrected for each hearing-impaired listener individually aiming to present signal levels that produce the same loudness levels within each narrowband as for the average normal hearing listener (narrowband loudness normalization). For this purpose the broadband signals were filtered in six nonoverlapping frequency bands having the same center frequencies as the narrowband signals. The required gain for each frequency band was defined as the difference in level for each loudness category between the individual loudness functions of the narrowband signals and the average normal hearing loudness function. An example is given in figure 5.2(a). The narrowband normalization method and the normal hearing reference data (dashed in figure 5.2(a)) were identical as applied in Oetting *et al.* (2016).

Gain limitations for the narrowband compensation were based on the maximum applicable level covered by the ethics approval. The loudness function for narrowband gain compensation was artificially limited to 105 dB HL at 50 CU. This leads to a gain reduction for listeners with loudness functions exceeding 105 dB HL for 50 CU, and narrowband loudness compensation will not be achieved. If the level of an amplified signal would have exceeded 105 dB HL, it was attenuated after narrowband loudness compensation to 105 dB HL by a broadband attenuation factor. The required amplification to restore normal loudness was calculated for the left and right ear separately. Theoretically, this may have caused a slight deviation from the principle to present the stimuli after loudness compensation in the binaural conditions diotically.

To quantify the remaining dynamic range of the impaired ears, for each narrowband signal the compression ratio (CR) was calculated which is defined as the ratio between input and output level at 40 and 80 dB HL input level according to the following equation:

$$CR = \frac{\Delta in}{\Delta out} = \frac{\Delta in}{\Delta in - \Delta gain} = \frac{80 - 40}{80 - 40 - (G40 - G80)} = \frac{40}{40 - \Delta gain} \qquad Eq. (1)$$

An example is given in figure 5.2(b) with a compression ratio of 1.9 indicating high gain values for low input levels and gains below 10 dB for high input levels.

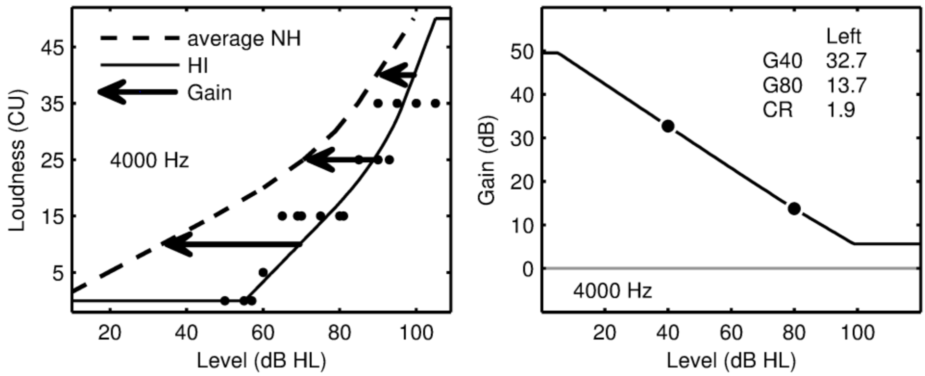


Figure 5.2. *(a) Gain to restore the narrowband loudness perception at 4000 Hz. Gains are defined as the horizontal difference between the individual loudness function (solid line) and the normal hearing reference (dashed line). (b) Level-dependent gain for narrowband loudness compensation from the example in (a). The gains result in a compression ratio of 1.9 for a gain difference around 19 dB (Equation (1)). CU=categorical unit; HI=hearing-impaired; NH=normal hearing; HL=hearing level.*

5.2.3.5 Part II: Spectral and binaural loudness summation.

In part II, the loudness-compensated UENs and the IFnoise were presented to the hearing-impaired listeners, first monaurally and thereafter binaurally. As in the first part of the measurements, hearing-impaired listeners with an even number started the monaural conditions with the right ear and hearing-impaired listeners with an uneven number started the monaural conditions with the left ear. To assess spectral loudness summation, levels for equal loudness of the narrowband UEN1 were compared with the more broadband signals UEN5, UEN17, and IFnoise.

Binaural loudness summation was assessed in two ways. First, spectral loudness summation for the binaural conditions was calculated parallel to the monaural conditions, that is, levels for equal loudness of the narrowband UEN1 were compared with the signals with an increasing bandwidth: UEN5, UEN17, and IFnoise. Second, binaural loudness summation was calculated as the level difference at equal loudness between the monaural loudness function (average of the right and left ear) and the binaural loudness function for all signals: UEN1, UEN5, UEN17, and IFnoise.

5.3 Results

5.3.1 Part I: Monaural loudness

The narrowband loudness normalization fitting method typically showed decreasing gains with increasing presentation level (figure 5.2(b)). By exception, gain increased slightly for frequencies without hearing loss.

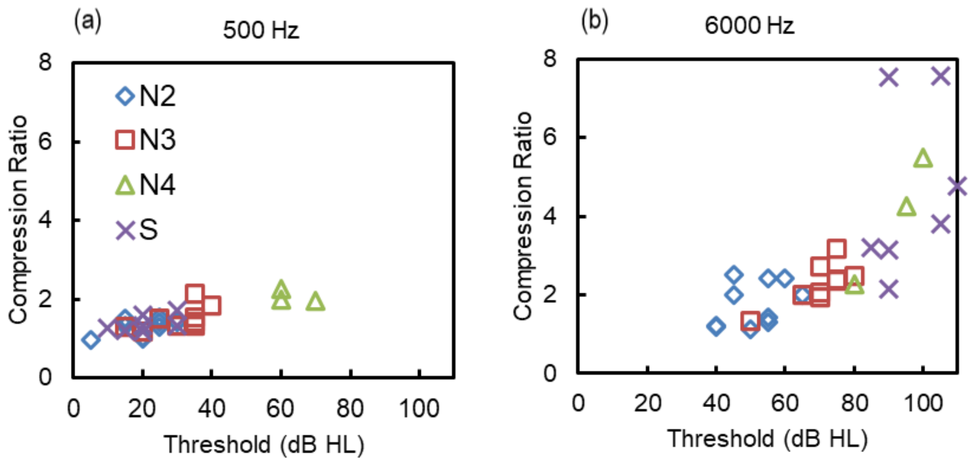


Figure 5.3. *(a) Compression ratios at 500 Hz for the left ear. (b) Compression ratios at 6000 Hz for the left ear. HL=hearing level.*

Figure 5.3 shows the compression ratios for the narrowband signals at 500 Hz and 6000 Hz as a function of the hearing threshold for the left ear. The symbols indicate the audiometric classification of the ear. The compression ratios associated with the narrowband signals show increasing values with increasing hearing threshold, especially for hearing losses above 60 dB HL.

The compression ratios are closely related to hearing threshold (500 Hz: r=0.788, p=<0.001; 6000 Hz: r=0.752, p=<0.001).

Figure 5.4 shows the monaural results of the signals with increasing bandwidth, for the right ear (upper row) and left ear (lower row). The levels on the x-axis represent the unaided input signal levels before amplification. Every solid line is a result of a single hearing-impaired listener. The dotted line represents the mean level of the hearing-impaired listeners measured in this study, and the striped line is the mean of nine normal hearing listeners (measured by Oetting *et al.* 2016).

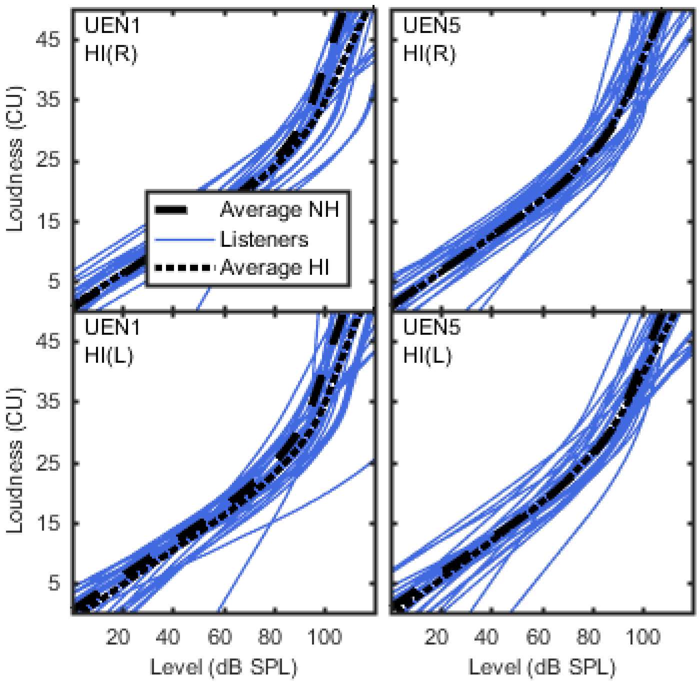


Figure 5.4a. *Loudness functions for UEN1 and UEN5 including individual and average monaural data. Upper row shows the results for the right ear, and lower row shows the results for the left ear. CU=categorical unit; HI=hearing-impaired; UEN=uniformly exciting noise; Ifnoise=International Female noise.*

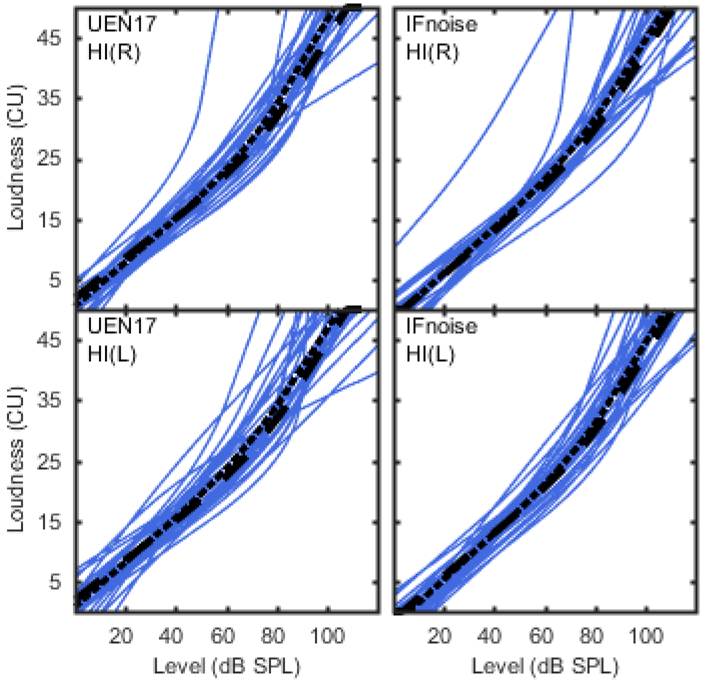


Figure 5.4b. *Loudness functions for UEN17 and IFnoise including individual and average monaural data. Upper row shows the results for the right ear, and lower row shows the results for the left ear. CU=categorical unit; HI=hearing-impaired; UEN=uniformly exciting noise; Ifnoise=International Female noise.*

The mean loudness curves for the hearing-impaired listeners are close to the mean loudness curves for the normal hearing listeners, but at high input levels of UEN1 loudness for hearing-impaired listeners was found to be slightly smaller than for normal hearing listeners, that is, loudness appears to be undercompensated for UEN1.

The average loudness functions for hearing-impaired listeners with UEN5 are almost the same as for the normal hearing listeners. At UEN17 and IFnoise, the average loudness curve for the hearing-impaired listeners is shifted slightly to lower input levels, relative to the loudness function for the normal hearing listeners, suggesting a slight overcompensation. In this case, the shift is mainly caused by listeners with N3 and N4 audiograms.

5.3.2 Part IIa: Spectral Loudness summation

Spectral loudness summation is defined as the level difference between the signals (UEN5, UEN17 and IFnoise) and the narrowband signal (UEN1). To asses spectral loudness summation in more detail, the level differences for equal (categorical) loudness (LDEL) with respect to UEN1 (center frequency 1370 Hz) were calculated for UEN5, UEN17, and IFnoise, as shown in figure 5.5. LDELs calculated with respect to UEN1 will be referred to as spectral LDELs (SLDELs).

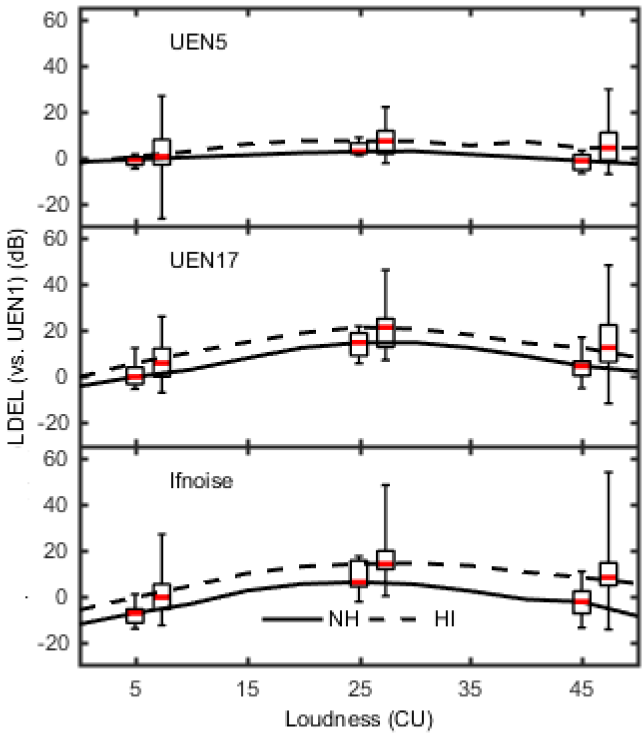


Figure 5.5. *Spectral loudness summation for the normal hearing listeners and hearing-impaired listeners, expressed as the level difference for equal loudness (SLDEL) with the narrowband UEN1 as the reference signal. The lines show median values across listeners. To assess interindividual variability, the boxplots show the results at 5, 25, and 45 CU. Whiskers indicate the observed range for the listeners, and the boxplots were horizontally shifted to increase readability. CU=categorical unit; HI=hearing-impaired; NH=normal hearing; UEN=uniformly exciting noise; Ifnoise= International Female noise; LDEL=level difference for equal loudness.*

Positive values indicate that a higher UEN1 level is required to match the loudness level of the test signal. The median SLDEL values for normal hearing listeners for each loudness category are shown as lines in the upper panel of figure 5.5 (replicated from Oetting *et al.,* 2016). Three boxplots per signal were included to show the interquartile range of the individual results for the loudness categories "very soft" (5 CU), "medium" (25 CU), and "very loud" (45 CU). Whiskers mark the entire range, that is, minimum and maximum values of the SLDELs. The differences in SLDELs for the normal hearing and hearing-impaired listeners are not strongly dependent on the categorical loudness levels. The average difference is about 8 dB for UEN17 and IFnoise. Hearing-impaired listeners show a larger spectral loudness summation effect than normal-nearing listeners. Detailed analysis reveals that this is due to the combined effect of a less than normal loudness perception for UEN1 and a higher than normal loudness perception for UEN17 and IFnoise.

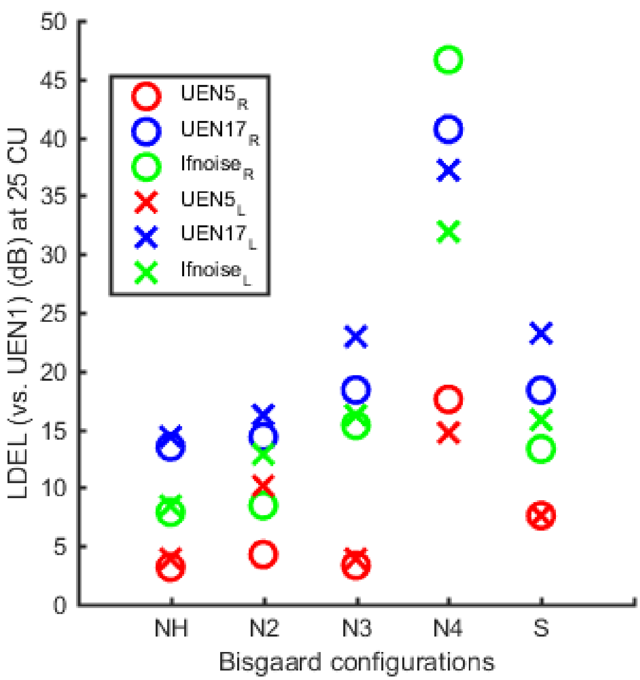


Figure 5.6. *Average SLDELs at 25 CU with respect to UEN1 for the different audiograms configurations for the UEN5, UEN17, and IFnoise. CU=categorical unit; UEN=uniformly exciting noise; Ifnoise=international Female noise.*

Figure 5.6 shows SLDEL's at CU 25 for the different audiogram configurations according to Bisgaard. At the left-hand site the SLDELs for normal hearing listeners are shown. The SLDELs for N2 do not deviate much from those for normal hearing listeners. The SLDELs for N4 listeners are clearly higher than for normal hearing listeners even with the UEN5 signal. For the N3 and S audiograms, the SLDELs for UEN17 and IFnoise are somewhat higher than for normal hearing listeners. Although the listeners were selected for symmetrical hearing losses, SLDELs for both ears may show some variation.

Table 5.1: *A Four-Way Mixed-Design ANOVA on the LDEL values with one Between-Subjects Factor (audiogram classification: NH, N2, N3, N4, S) and three Within-Subjects Repeated Measures (three signals: UEN5, UEN17, and IFnoise; three loudness categories: 5, 25, and 45 CU; two ears: left and right). The significance level was set at 0.05. The Greenhouse-Geisser correction was used because sphericity of the data could not be assumed.*

	df	F	p
Ear	1.000	0.95	0.943
Ear x Audiogram	4.000	612.49	0.021
Loudness	1.647	12914.98	< .001
Loudness x Audiogram	6.590	997.64	0.003
Signal	1.529	6046.06	< .001
Signal x Audiogram	6.114	568.10	< .001
Ear x Loudness	1.599	908.93	0.014
Ear x Loudness x Audiogram	6.396	190.48	0.390
Ear x Signal	1.884	75.15	0.095
Ear x Signal x Audiogram	7.535	89.17	0.008
Loudness x Signal	2.746	772.83	< .001
Loudness x Signal x Audiogram	10.984	46.65	0.488
Ear x Loudness x Signal	2.915	21.78	0.645
Ear x Loudness x Signal x Audiogram	11.659	33.36	0.607

Note. The significance level was set at 0.05. The Greenhouse-Geisser correction was used because sphericity of the data could not be assumed.

A four-way mixed-design ANOVA was conducted on the SLDEL values with one between-subjects factor (audiogram-classification: NH, N2, N3, N4, S) and three within-subjects repeated measures (three signals: UEN5, UEN17, and IFnoise; three loudness categories: 5, 25, and 45 CU; two ears: left and right), see Table 5.1. The significance level was set at 0.05. The

Greenhouse-Geisser correction was used whenever sphericity of the data could not be assumed.

As should be the case in symmetrical hearing losses, there was no significant effect of the test ear. A significant effect was shown for loudness category $F(1.6, 54.4)=46.6$, $p=<0.001$, and signal $F(1.5, 73.2)=82.5$, $p=<0.001$. Bonferroni-corrected comparisons of loudness category and signal showed significant differences between all loudness categories and bandwidths. With regard to audiometric configuration, N3, N4 and S differed significantly from normal hearing with p-values of 0.044, <0.001 and 0.014 respectively. Hearing loss category N2 was not significantly different from normal hearing.

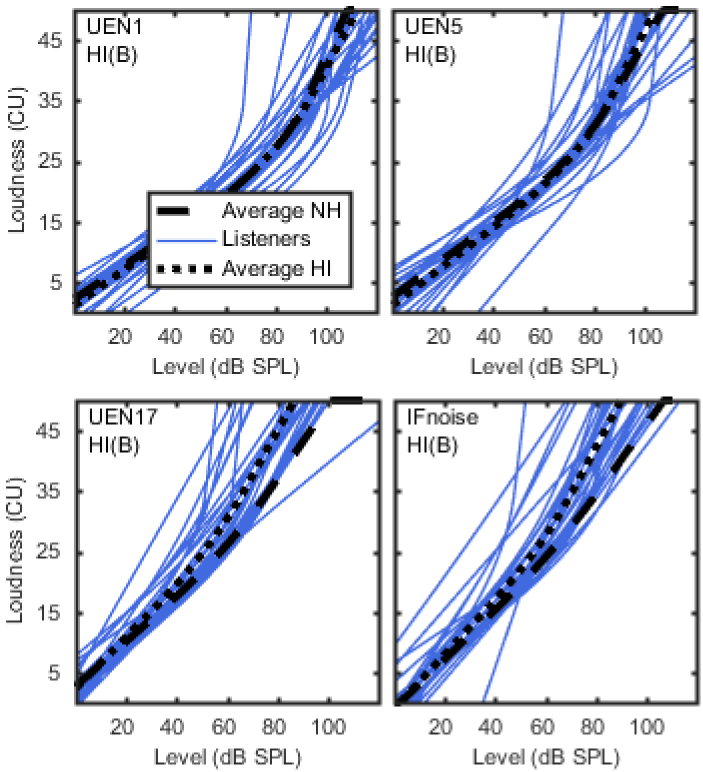


Figure 5.7. *Individual and average data of the binaural conditions for all different audiograms. CU=categorical unit; UEN=uniformly exciting noise; Ifnoise=international Female noise; HI=hearing-impaired.*

5.3.3 *Part IIb: Binaural loudness summation*

Figure 5.7 shows the individual and average results for all binaural conditions in the same way as the monaural data was shown in figure 5.4. Again the mean loudness functions for the hearing-impaired listeners are close to the average normal hearing loudness functions for the signals UEN1 and UEN5.

That is, in the binaural condition, the loudness normalization procedure seems to restore loudness to normal for the narrowband signals. For the broadband signals UEN17 and IFnoise, however, the loudness functions are clearly shifted to the higher-than-normal loudness. This indicates that the binaural broadband signals - despite the loudness normalization based on monaural NB signals - are perceived by the hearing-impaired listeners as louder than by the normal hearing listeners. As in the monaural condition SLDELs were calculated with respect to UEN1. The results of this calculation can be seen in figure 5.8. As in the monaural conditions the variability is much larger for the hearing-impaired listeners than for the normal hearing-listeners.

A three-way mixed-design ANOVA was conducted on the SLDEL values with one between-subjects factor (audiogram classification) and two within-subjects repeated measures (three signals: UEN5, UEN17, and IFnoise; three loudness categories: 5, 25, and 45 CU). The significance level was set at 0.05. The three-way interaction was not significant $F(13.2, 108.6)=0.74$, $p=0.719$. The two-way interactions between signal and loudness category $F(3.3, 108.6) =30.08$, $p < .001$, bandwidth and hearing loss category $F(6.9, 56.9)=8.4$, $p<0.001$, and loudness category and hearing loss category $F(7.5, 61.9)=3.5$, $p=0.003$ were significant. There was a significant effect of loudness category $F(1.9, 61.9)=75,5$ and bandwidth $F(1.7, 56.9)=149.8$ both with a $p<0.001$.

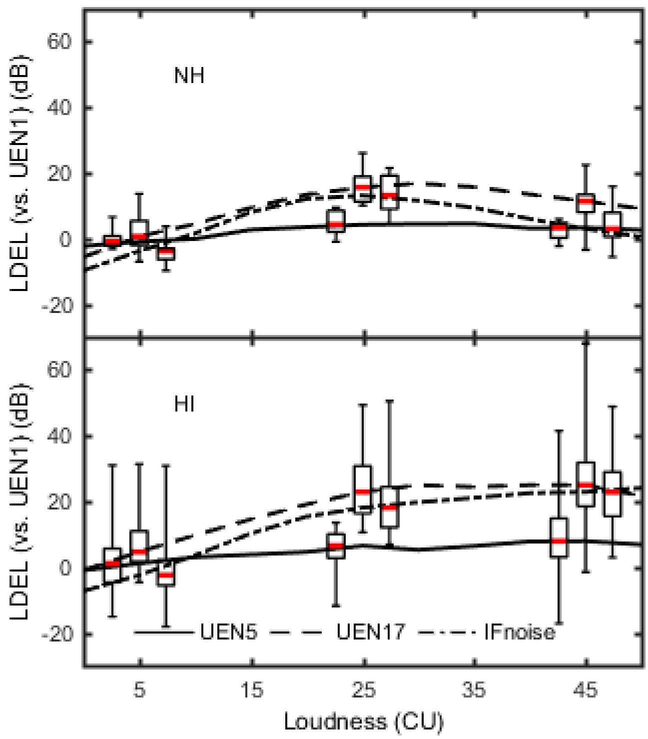


Figure 5.8. *Spectral loudness summation of binaural sounds for the normal hearing (upper panel) and hearing-impaired (lower panel) listeners, expressed as the LDEL with the narrowband UEN1 (center frequency 1370 Hz) as reference. The lines show median values across listeners. To assess inter-individual variability, the boxplots show the individual results at 5, 25, and 45 CU. Whiskers indicate the observed range for the listeners and the boxplots were horizontally shifted to increase readability. CU=categorical unit; UEN=uniformly exciting noise; Ifnoise=international Female noise. NH=normal hearing; HI=hearing-impaired; LDEL=level difference for equal loudness.*

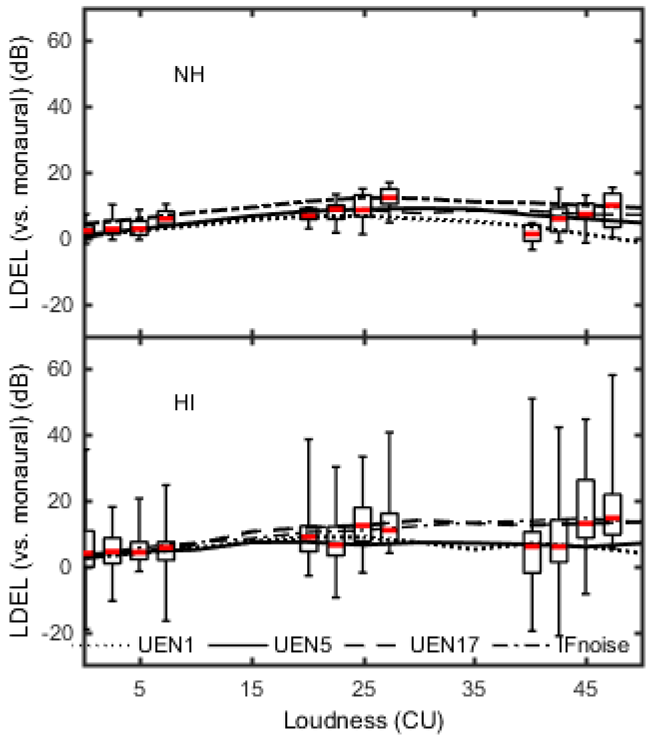


Figure 5.9. *Binaural loudness summation of broadband sounds for the normal hearing (upper panel) and hearing-impaired (lower panel) listeners, expressed as the LDEL) with the mean results of the right and left ears as reference. Solid lines show median values across listeners. To assess inter-individual variability the boxplots show the inter-individual results at 5, 25, and 45 CU. Whiskers indicate the observed range for the listeners and the boxplots were horizontally shifted to increase readability. CU=categorical unit; UEN=uniformly exciting noise; Ifnoise=international Female noise. NH=normal hearing; HI=hearing-impaired; LDEL=level difference for equal loudness.*

Bonferroni-corrected comparisons of loudness category showed significant differences between the SLDELs at 5 CU and at 25 CU and 45 CU, but not between the LDELs at 25 CU and 45 CU. For the different bandwidths all differences were significant. With regard to hearing loss, category N3, N4 and S differed significantly from normal hearing with p-values of 0.024, <0.001 and 0.044 respectively. Hearing loss category N2 was not significantly different from normal hearing.

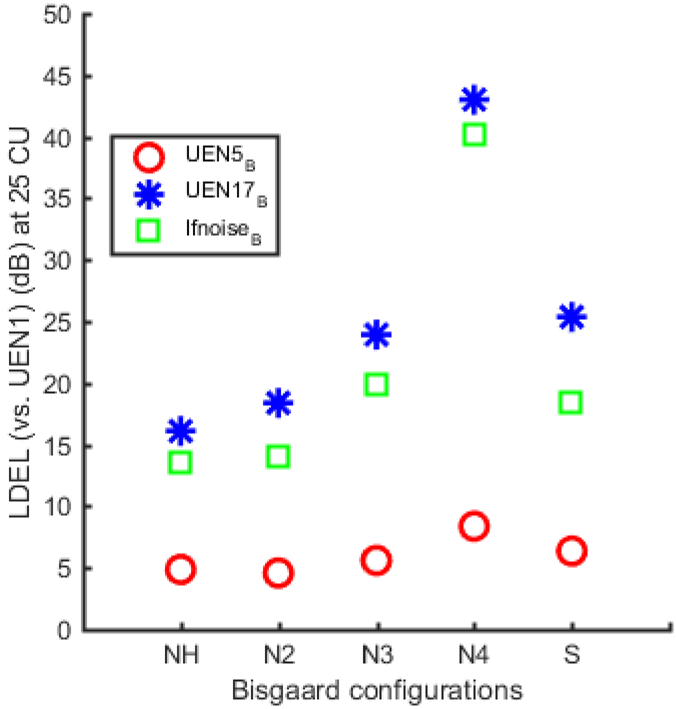


Figure 5.10. *SLDEL (re UEN1) for the different audiometric classifications tested for binaural signals presented at 25 CU. CU=categorical unit; UEN=uniformly exciting noise; Ifnoise=international Female noise. NH=normal hearing; HI=hearing-impaired; LDEL=level difference for equal loudness.*

Figure 5.10 shows the mean SLDELs at 25 CU for the binaural conditions as a function of the audiogram classification. As in the monaural conditions the results for N2 are close to normal for all signals. For N3, N4 and S the mean SLDELs for the broadband signals are larger than for the normal hearing listeners. Especially for N4 the SLDELs are higher than normal, as was also observed in the monaural conditions.

Figure 5.8 showed spectral loudness summation of binaural sounds. It is also possible to calculate the binaural loudness summation of broadband sounds, thatis, the binaural summation with respect to the monaural signals. Figure 5.9 shows the mean binaural level differences at equal loudness (BLDELs) between the mean of the right and left ear and the binaural level for all stimuli. For UEN1 and UEN5, the binaural summation for the hearing-impaired listeners is similar to the binaural summation of the normal hearing listeners. For UEN17 and IFnoise (the broadband conditions), binaural summation is increased for the hearing-impaired listeners, especially at higher CUs. This trend was observed for all hearing loss configurations (not shown). Again, the variability in the results for the hearing-impaired listeners is much larger than for the normal hearing listeners. A three-way mixed-design ANOVA was conducted on the BLDEL values with one between-subjects factor (audiogram-classification) and two within-subjects repeated measures (four signals: UEN1, UEN5, UEN17, and IFnoise; three loudness categories: 5, 25, and 45 CU). The significance level was set at 0.05. The three-way interaction was not significant $F(4.6,152.2)=1.59$, $p=0.068$. The two-way interaction between signal and loudness category $F(4.6, 152.2)=4.829$, $p < .001$ was significant. The other two-way interactions were not. There was a significant effect of loudness category $F(1.3,43.9)=13.20$ and bandwidth $F(2.7, 88.7)=16.07$ both with a $p<0.001$. There was no main effect of audiogram-classification. Post hoc analysis showed significant differences between CU25 and CU45 with respect to CU5, but not between each other. For the different signals no differences were found between UEN1 and UEN5 and between UEN17 and IFnoise. All other differences were significant with $p<0.001$.

5.3.4 Predictability of the binaural loudness for broadband signals

As binaural loudness is not routinely measured in clinical practice it is important to check if the amount of binaural loudness for broadband signals can be predicted on the basis of monaural measurements. Therefore correlations (Pearson's r) were calculated between the levels of the binaural IFnoise at 45 CU and test results derived from monaural measurements. To characterize the audiometric data, pure tone averages (PTAs) were calculated: $PTA_{.5,1,2,4}$ (PTA), $PTA_{.25,.5,1}$ (PTA_{low}), $PTA_{1,2,4}$ (PTA_{high}). In addition, the loudness levels at 45 CU from the monaural measurements for the four signals were selected. Besides, the compression ratios of the unaided loudness curves for the 6 low-noise noises were taken. Results are given in table 5.2. Significant values ($p = 0.05$) are indicated with asterisks.

Table 5.2. *Correlation values (Pearson's r) between the levels at 45 CU of the binaural IFnoise and several test values for both the right and the left ear.*

Right ear				Left ear			
	Pearson's *r*		p		Pearson's *r*		p
PTA	-0.409	*	0.028	PTA	-0.407	*	0.028
PTA low	-0.327		0.084	PTA low	-0.433	*	0.019
PTA high	-0.380	*	0.042	PTA high	-0.340		0.071
UEN1	-0.169		0.380	UEN1	-0.071		0.714
UEN5	0.304		0.109	UEN5	0.305		0.107
UEN17	0.466	*	0.011	UEN17	0.428	*	0.020
IFnoise	0.460	*	0.012	IFnoise	0.452	*	0.014
CR 250	-0.305		0.108	CR 250	0.010		0.958
CR 500	-0.219		0.253	CR 500	-0.225		0.241
CR 1000	-0.122		0.528	CR 1000	-0.205		0.287
CR 2000	-0.170		0.377	CR 2000	-0.043		0.824
CR 4000	-0.048		0.803	CR 4000	-0.180		0.350
CR 6000	-0.200		0.299	CR 6000	-0.299		0.116

Note. Significant values at a level of p = 0.05 are marked with an asterisk.
CR = compression ratio; PTA = pure-tone average; UEN = uniformly exciting noise; IFnoise = International Female noise.
*p<0.05. **p<0.01. ***p<0.001.

In most cases, the correlation coefficients were weak, but some significant correlations were found. There is a trend of negative correlations (not all of them reach the level of significance) between the different PTA-parameters and binaural loudness for the IFnoise, indicating that less binaural loudness is found for higher hearing losses, despite the correction based on monaural loudness and the high level.

Positive significant correlations were found for the broadband monaural signals (UEN17 and IFnoise), indicating that binaural loudness is increased when spectral loudness summation is higher. The strong correlation between monaural and binaural results appears to be determined by spectral summation effects for the N4 audiograms. All correlations with the compression ratios of the low-noise noises used to perform the loudness equalization were nonsignificant.

All hearing-impaired listeners completed the Abbreviated Profile of Hearing Aid Benefit questionnaire during the study. No clear correlations between the answers on the aided or unaided Abbreviated Profile of Hearing Aid Benefit questionnaire and the amount of binaural loudness summation was found.

5.4 Discussion and conclusions

5.4.1 Monaural spectral loudness summation

The results in this study are an extension of the work by Oetting *et al.* (2016), as Oetting *et al.* (2016) measured only a small sample of hearing losses with a relative small hearing loss (N1:1, N2: 4, N3: 1, S1: 4). In this study, group sizes were increased and moderate to severe hearing losses were included. In both studies, spectral loudness summation for hearing-impaired listeners was found to be higher than in normal hearing listeners. Oetting *et al.* (2016) compared monaural spectral loudness summation as measured with the current procedure to several studies from the literature (Appell and Hohmann 1998; Brand and Hohmann 2001; Bonding and Elberling 1980; Garnier *et al.* 1999; Strelcyk *et al.* 2012; Verhey *et al.* 2006). The ʌ-shape and 10-15 dB spectral loudness summation found by Oetting *et al.* (2016) for normal hearing listeners were in agreement with data from the literature with some minor exceptions. For hearing-impaired listeners Oetting *et al.* (2016) noted that literature data showed a decrease in spectral loudness summation with increasing hearing loss, which was in contrast with their own results. However, all studies agreed about the fact that the variability in the results of hearing-impaired listeners was large.

In this study, the effects on spectral loudness summation were largest for the largest bandwidths UEN17 and IFnoise (in correspondence with i.e. Zwicker, 1958) and increased with the degree of hearing loss in agreement with the results by Oetting *et al.* (2016). Spectral loudness summation was negligible for UEN5. Compared to the data by Oetting *et al.* (2016), this study shows slightly more undercompensated loudness for UEN1. This effect is seen for subjects in all Bisgaard classifications and seems to be larger for subjects with higher losses (N4 and S3). This could be due to the fact that loudness functions for narrowband gain compensation were limited to 105 dB HL and obviously the calculated gain values were not sufficient to achieve a complete narrowband loudness compensation.

Only a few other studies investigated spectral loudness summation for different degrees of hearing loss. Bonding and Elberling (1980) measured spectral loudness summation for different degrees of flat hearing loss (PTAs of 0.5, 1 and kHz of 25 dB, 40 dB and 50 dB HL). The flat audiograms in this study correspond to PTAs of 27 dB (N2), 42 dB (N3) and 58 dB (N4). Bonding and Elberling (1980) found that SLDELs for hearing-impaired listeners were smaller than for normal hearing listeners at the same reference level of the narrowband signal, with no clear effect of the degree of hearing loss on the maximum SLDELs reached. The reference level as defined by Bonding and Elberling (1980) does not ensure equal loudness for normal hearing listeners and hearing-impaired listeners at the same reference level. Therefore, it is not clear whether the difference in SLDELs for normal hearing listeners and hearing-impaired listeners would still be found, if their results were analyzed according to a loudness scale that ensured equal loudness for both groups, as in this study. The bandwidth of 1600 Hz used by Bonding and Elberling may contribute further to the absence of an effect of degree of hearing loss on the maximum SLDEL in their study, as in this study, the effect of the degree of hearing loss was only apparent for the broadband signals UEN17 and IFnoise but was small for UEN5 (1080-Hz bandwidth).

Strelcyk *et al.* (2012) showed that the method of hearing loss compensation can influence the amount of absolute spectral loudness summation. They used three different multichannel compression systems and measured their effects on loudness summation. As in the current study, they compensated the broadband loudness signals for the degree of hearing loss. In contrast to our study, their compensation strategy was not loudness based but threshold based. Strelcyk *et al.* (2012) found no difference in the maximum SLDELs between a 230-Hz wide reference signal and a 1600-Hz wide test signal for the hearing-impaired listeners included (with a flat hearing loss and a PTA across 0.5, 1, 2 and 4 kHz of 55 dB, which is in between our N3 and N4 listeners) relative to normal hearing listeners. This is not in contrast with our results as the larger SLDELs for hearing-impaired listeners in this study were found for UEN17 and IFnoise and not for UEN5 and UEN1. As in Bonding and Elberling (1980), for hearing-impaired listeners, the level of the maximum was shifted to higher levels for the signals centered around 1 kHz. This is in line with the current data, as maximum SLDELs were found around 25 CU, and the level at which 25 CU is reached is shifted to higher levels for subjects with increased hearing loss.

In this study, flat audiograms (N2, N3 and N4) were compared with (a few) sloping audiograms (S), as loudness models predict more spectral loudness summation in the high-frequency region than in the low-frequency region (DIN, 199l; ANSI, 2007). Nevertheless Schlittenlacher *et al.* (2015) found only minor differences in loudness summation between lower (125-1000 Hz), middle (500-2000 Hz), and higher (1.25-5 kHz) noises when these signals were compared with a 1 kHz-tone. More hearing loss in the high frequency could therefore give rise to less spectral loudness summation. However, in this study, no clear differences were found between flat audiograms and sloping audiograms. The largest deviations were found for audiograms classified as N4. N4 audiograms mainly deviate from the other audiogram configurations in the low-frequency region.

It is not clear why monaural spectral loudness summation is increased for hearing-impaired listeners. In theory, the widening of the critical bands for hearing-impaired listeners should lead to a decrease in spectral loudness summation (e.g., Moore and Glasberg, 2004). In the S audiograms, the hearing loss in the high frequencies (2, 4 and 6 kHz) is close to or greater than the hearing loss in N4, but the spectral loudness summation is clearly smaller. This suggests that spectral loudness summation in the N4 listeners is mainly influenced by the low frequencies. The relatively large gains in the low frequencies for the N4 hearing-impaired listeners may have caused upward spread of masking. If this was the case, the narrowband normalization strategy used in this study may have induced higher loudness values, that could have been interpreted as spectral loudness summation. This requires further research.

5.4.2 Binaural loudness summation

Oetting *et al.* (2016) summarized the findings on binaural loudness summation in other studies with hearing-impaired listeners (Dermody and Byrne, 1975; Hawkins *et al.* 1987; Moore *et al.* 2014; Whilby *et al.* 2006). They concluded that the BLDELs between monaural and binaural stimuli fell between 5 and 8 dB. This corresponds reasonably well with the median results found for UEN1 and UEN5, where BLDELs were found ranging from 2.8 dB at low loudness categories up to 8.0 dB at high loudness categories. Similar binaural loudness summation values were also found for normal hearing listeners (for an overview, see Whilby *et al.* 2006).

In this study and in the study by Oetting *et al.* (2016), signals with larger bandwidths (UEN17 and IFnoise) were used than in other studies. In the current studies, average BLDELs were found of 11.5 and 12.3 dB for UEN17 and IFnoise, respectively. Thus, binaural loudness summation increased with increasing bandwidth. On average, spectral and binaural loudness summation seem to add, causing an extra strong bandwidth dependency for the combined effects of binaural and spectral loudness summation.

Because the monaural results sorted by audiogram class (figure 5.6) show great similarities with the binaural results (figure 5.10), it is tempting to assume a common origin. However, the analysis of the effect of presentation (monaural to the right, monaural to the left and binaural) shows that the binaural results are significantly different from both monaural results, while the monaural data do not significantly differ from each other. This is reflected in figure 5.11. In this figure, the hearing-impaired listeners classified as N4 show binaural summation close to normal, which seems to imply that the large combined spectral and binaural loudness summation is mainly caused by the large spectral monaural loudness summation. For hearing-impaired listeners classified as N3 and S, binaural summation for the broader bandwidths is larger than normal, suggesting a separate binaural effect next to the spectral loudness effect.

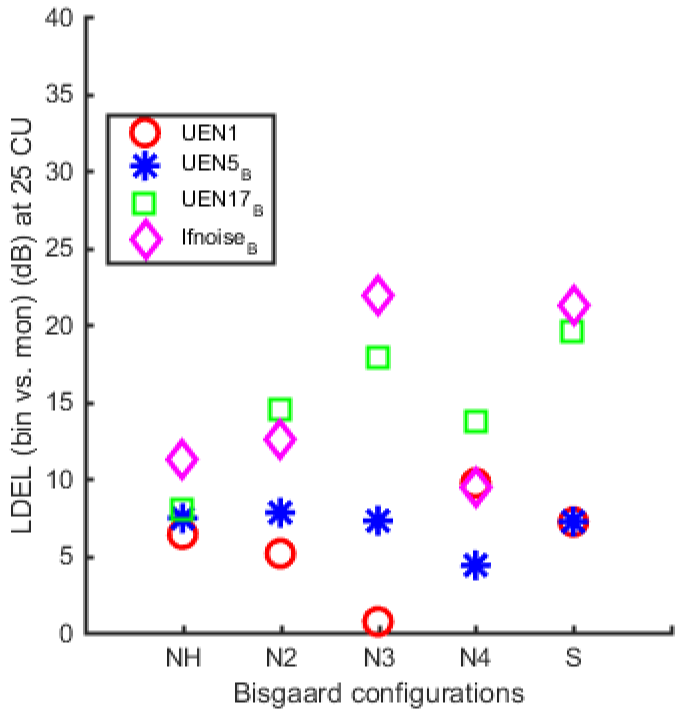


Figure 5.11. *BLDEL (re monaural signals) for the different audiometric classifications tested for binaural signals at 25 CU. CU=categorical unit; UEN=uniformly exciting noise; Ifnoise=international Female noise. NH=normal hearing; HI=hearing-impaired; LDEL=level difference for equal loudness.*

5.4.3 Interactions between spectral and binaural summation

This study shows a hearing loss dependency for spectral loudness summation of binaurally presented signals. For small hearing losses (N2), spectral loudness summation of binaural sounds is the same as for normal hearing listeners. For the larger hearing losses (N3, N4, and S), spectral loudness summation of binaural sounds tends to be higher than normal, with extremely high values for N4.

With respect to binaural summation of broadband sounds itself (binaural conditions versus monaural conditions), no clear hearing loss dependency was found. Binaural summation of broadband sounds appears to be larger for N3 and S than for normal hearing listeners, but statistically, this difference was not significant. However, in binaural loudness summation of broadband sounds, the effect of bandwidth did lead to a statistically significant difference: UEN17 and IFnoise showed more binaural loudness summation of broadband sounds than UEN1 and UEN5. The finding that listeners with an N4 audiogram showed large spectral loudness summation for binaural signals, but normal binaural loudness summation, indicates that spectral loudness summation of binaurally presented sounds may be a complex combination of the effects of bandwidth and hearing loss.

5.4.4 The effects of hearing loss

Hearing loss may influence binaural loudness perception in two ways. First of all, central gain may be increased. Eggermont (2017) reviewed the influence of acquired hearing loss on the central auditory system and found increased spontaneous firing rates and increased neural synchrony at the level of the auditory cortex. Salvi *et al.* (2017) reviewed a comprehensive series of experiments aimed to determine how loss of the inner hair cells type I system affects hearing in chinchillas. They concluded that the results suggest that when the neural output of the cochlea is reduced, the central auditory system compensates by turning up its gain so that weak signals once again become comfortably loud. Chen *et al.* (2014) found a correlation between salicylate-induced hyperactivity in the central auditory systems of rats with behavioral evidence of loudness hyperacusis. Excessive increases of the central gain may thus convert recruitment into loudness hyperacusis.

Increased binaural loudness summation could also be explained by a decrease of contralateral suppression in hearing-impaired listeners activated by the medial olivo-cochlear (MOC) system. MOC feedback to the cochlea is believed to control cochlear gain and to enable modulation of auditory nerve activity (Guinan 2006; Guinan and Gifford 1988; Warr 1975). However, in a study by Wilson *et al.* (2017) in children with autism spectrum disorder MOC inhibition of transient otoacoustic emissions was on average larger at all frequencies for a group with severe hyperacusis compared with a group without severe hyperacusis. The stronger activity of the MOC in the groups related to hyperacusis is not compatible with the hypothesis of decreased contralateral suppression. Wilson *et al.* (2017) propose an increased gain in the central auditory pathways as an explanation for the increased MOC effect.

5.4.5 Limitations of the current approach

The results presented in this study have to be considered in relation to the choices made for the experimental setup. In categorical loudness scaling, the choice of the number of response alternatives is a factor that influences the slope of the loudness curve. With an increasing number of response alternatives, the knee point of the fitted loudness function tends to shift to lower intensities (Brand, 2007). The LDELs presented are therefore influenced by the choice for the procedure for categorical loudness scaling. In several studies, the reproducibility of categorical loudness scaling has been investigated (Al-Salim *et al.* 2010; Cox *et al.* 1997; Rasetshwane *et al.* 2015; Robinson and Gatehouse, 1996) and has been found to be good on group level. Rasetshwane *et al.* (2015) found that categorical loudness scaling was reliable even at an individual level and leads to comparable results with other loudness measurements when the categorical units are transformed to phons. They calculated the standard deviation of the signed differences between test and retest for 22 subjects and found a mean of 4.22 dB.

The LDEL values are also influenced by the choice for the current compression system with six nonoverlapping channels. As Strelcyk *et al.* (2012) showed, the compressor influences the loudness summation after loudness equalization. More channels or other choices for the frequency limits could lead to different LDELs while the underlying real physiological loudness summation processes have not been changed.

The LDEL values further depend on the selected broadband test signals. In some subjects, we found high spectral loudness summation for the IFnoise and lower spectral loudness summation for the UEN17. It might be that the narrowband loudness compensation applied to the signals lead to a lower perceived bandwidth of the UEN17 signal compared to the IFnoise.

Finally, the measurement setup will influence the measured LDELs. The choice for a specific headphone (HAD 200) and equalization method (free-field equalization) defines the signal at the eardrum. Another setup would inherently have resulted in a different signal at the eardrum. For instance, Thiele *et al.* (2014) found that the 50% speech reception threshold measured with the HDA200 headphones with free-field correction was on average 5.1 dB lower than for loudspeakers. Thus, even widely used equalization methods do not guarantee equal sound characteristics at the eardrum.

The narrowband loudness normalization method used in this study is not directly suitable for use in hearing aid fitting, as normalizing narrowband loudness does not guarantee normal binaural broadband loudness. Furthermore, normal binaural broadband loudness does not guarantee optimal speech understanding and optimal comfort. The large interindividual differences in binaural loudness perception are an important finding, but implications on hearing aid fitting require further research.

5.4.6 Clinical implications

The results of this study confirm the findings by Oetting *et al.* (2016) that spectral loudness summation of binaurally presented sounds can be extremely large in HI listeners. Although a significant effect of the audiometric configuration on the amount of spectral loudness summation of binaural sounds was found, the variability in each group was that large that the spectral loudness summation of binaurally presented sounds could not be predicted from the audiometric classification alone. The correlation matrix shows that other predictors based on audiogram or monaural loudness measurements also fail to give a good prediction of the amount of the combined spectral and binaural loudness summation. As we encounter in daily practice very often broadband sounds presented in a binaural situation, spectral and binaural loudness summation are highly relevant features. The current hearing aid fitting rules based on monaural threshold measurements utilize average gain corrections for bilateral fittings that are identical for all hearing-impaired listeners. NAL-NL2 propose bilateral compensation factors (reductions in gain) relative to an unilateral fitting ranging from 2 dB for input levels below 40 dB to 6 dB for input levels at 90 dB SPL and above regardless of signal bandwidth (Keidser et al. 2012).

Our results show a clear bandwidth-dependency of binaural loudness summation with individual binaural summation effects higher than 30 dB for broadband input signals presented at mediate to high levels (see figure 5.9). In our approach, input levels are processed according to the six-channel compressor with independent compression ratios to compensate the narrowband loudness perception. Loudness summation expressed as output levels will therefore give a smaller effect size. However, the effect is still sizable, as the mean compression ratio is 2.1:1 averaged over all frequencies for the Bisgaard classes N3, N4 and S. Taken compression into account, the average amount of binaural loudness summation in output terms is still in excess of 14 dB for the more severe hearing losses. With individual differences ranging from about 30 dB at 25 CU to over 60 dB at 45 CU (cf. figure 5.9), taking the effect of compression into account still leaves output level differences in individual binaural loudness summation between 14 to 29 dB. These values are in accordance with the large inter-individual differences in LDLs found by Formby *et al.* (2017) for monaural warble tones.

Therefore, there is need to adjust fitting rules for bilaterally fitted hearing aids to take the large individual differences in loudness summation into account. Regarding the high variability in the individual data, it seems to be imperative to determine individual amounts of gain correction based on separate tests of loudness perception, including spectral and binaural loudness summation.

Acknowledgements

We thank Marije Wolvers with her help with the statistical analysis. We also thank Addy Mols for her participation in the data collection. The authors would also like to thank Josef Schlittenlacher and two anonymous reviewers for their constructive comments.

Chapter 6

Uni-and bilateral spectral loudness summation and binaural loudness summation with loudness matching and categorial loudness scaling

Van Beurden, M.F.B., Boymans, M., van Geleuken, M., Oetting, D., Kollmeier, B., Dreschler, W.A. (2021). Uni- and bilateral spectral loudness summation and binaural loudness summation with loudness matching and categorical loudness scaling. Int J Audiol. 60(5), 350-358. Doi: 10.1080/14992027.2020.1832263.

Abstract

Objective: Current hearing-aid prescription rules assume that spectral loudness summation decreases with hearing impairment and that binaural loudness summation is independent of hearing loss and signal bandwidth. Previous studies have shown that these assumptions might be incorrect. Spectral loudness summation was measured and compared for loudness scaling and loudness matching.

Design: In this study the effect of bandwidth on binaural summation was investigated by comparing loudness perception of low-pass filtered, high-pass filtered, and broadband pink noise at 35 Categorical Units for both unilateral and bilateral presentation.

Study Sample: Sixteen hearing-impaired listeners. Results: The results show that loudness differences between the three signals are different for bilateral presentation than for unilateral presentation. In specific, binaural loudness summation is larger for the low-pass filtered pink noise than for the high-pass filtered pink noise. Finally, individual variability in loudness perception near loudness discomfort level was found to be very large.

Conclusions: Loudness matching is offered as a fast and reliable method to measure individual loudness perception. As discomfort with loud sounds is one of the major problems encountered by hearing aid users, measurement of individual loudness perception could improve hearing aid fitting substantially.

6.1 Introduction

In studies on the benefit of hearing aids aversiveness of loud sounds remains an important reason for dissatisfaction with hearing aids (Kochkin, 2000; Jenstad *et al.,* 2003; Boymans *et al.*, 2008; Hickson, *et al.*, 2010; Franks and Beckmann, 1985; EuroTrak Germany, 2018). However, in most clinical settings individual frequency specific loudness discomfort levels (LDLs) are not routinely measured (Mueller, 2003), as loudness measurements are deemed time-consuming and tedious for patients to perform (Formby *et al.*, 2017). In many hearing aid fittings, only hearing thresholds are included and the uncomfortable loudness level is estimated, for instance NAL-NL1 and NAL-NL2 do not allow to enter patient-specific LDLs (Keidser *et al.* 2011). And although most hearing-aid users wear binaural hearing aids (Kochkin, 2009; EuroTrak Germany 2018) and speech and environmental sounds are broadband signals, most of the current prescription rules such as NAL-NL2 (Dillon, 2012) and DSL I/O (Cornelisse *et. al,.*1995, Bagatto *et al.,* 2005, Scollie *et al.,* 2005) are based on monaural threshold measurements with narrow-band signals, e.g. pure tones.

To prescribe the right amount of gain for broadband binaural sounds several assumptions about loudness perception are made in the prescription rules. The first assumption is that spectral loudness summation for hearing-impaired listeners is equal to or lower than spectral loudness summation for normal hearing listeners. Several studies have shown lower-than-normal spectral loudness summation for hearing-impaired listeners (e.g., Bonding and Elberling, 1980; Brand and Hohmann, 2001; Florentine and Zwicker, 1979; Garnier et al., 1999; Verhey et al., 2006). The most recent loudness model used in prescription rules is the model by Moore and Glasberg (2004). This model assumes a lower compression for cochlear hearing loss leading to decreased spectral loudness summation. Prescription rules based on this model therefore assume decreased spectral loudness summation. However, some recent studies have shown that spectral loudness summation for individual hearing-impaired listeners can be clearly higher than spectral loudness summation for normal hearing listeners, after an appropriate compensation of the hearing loss (Oetting *et al.,* 2016; van Beurden *et al.*, 2018; Rasetshwane *et al.*, 2018).

Rasetshwane *et al.* (2018) measured the effect of monaural hearing aid amplification on spectral loudness summation. They compared unaided loudness summation to two aided conditions. In both aided conditions gain prescriptions were based on loudness growth curves measured for narrow-band signals at specific frequencies with categorical loudness scaling. In one of the aided conditions, they also accounted for the effects of suppression, the reduction in the cochlear response to a sound due to the simultaneous presence of other sounds. Rasetshwane *et al.* (2018) found that spectral loudness summation was higher-than-normal for the aided conditions, although the differences between moderate hearing-impaired listeners and normal hearing listeners decreased when suppression was taken into account.

Oetting *et al.* (2016) investigated spectral and binaural loudness summation for nine normal hearing listeners and ten hearing-impaired listeners. Loudness perception was quantified by categorical loudness scaling for six narrowband signals and four broadband signals. The hearing-impaired listeners had slight-to-moderate sensorineural hearing losses with pure-tone averages across 500, 1000, 2000, and 4000 Hz between 20 and 44 dB HL. For these listeners, frequency- and level-dependent amplification was used to match the narrowband monaural loudness functions of the normal hearing listeners. The required gain levels were defined as the differences for each loudness category between the level of the average normal hearing loudness function and the level of the individual narrowband loudness functions. For the broadband signals, the gain prescriptions derived from narrowband sounds were used. The results indicated that spectral loudness summation was slightly higher for hearing-impaired listeners than for normal hearing listeners for monaurally presented sounds and substantially larger for binaurally presented sounds. Van Beurden *et al.* (2018) extended the experiments of Oetting *et al.* (2016) with a larger group of hearing-impaired listeners. They measured spectral loudness summation for monaural and binaural broadband signals in twenty-nine hearing-impaired listeners using the same narrowband loudness compensation as used by Oetting *et al.* (2016). The hearing losses were grouped according to the Bisgaard classification (Bisgaard *et al.*, 2010) N2, N3, N4, and a group with ski-sloping audiograms (i.e. Bisgaard category S2 or S3). Van Beurden *et al.* (2018) showed that the degree of spectral loudness summation significantly increased for large bandwidths for categories N3, N4, and S.

The second assumption made in hearing aid prescription rules is that binaural loudness summation has the same effect for all hearing losses and for all signals. For bilateral fittings, the prescription rules apply fixed gain corrections that are the same for all hearing-impaired listeners. NAL-NL2 proposes bilateral compensation factors (reductions in gain) relative to a unilateral fitting ranging from 2 dB for input levels below 40 dB to 6 dB for input levels at 90 dB SPL and above, regardless of the signal bandwidth (Keidser *et al.,* 2012). DSL m[i/o] prescribes a reduction of 3 dB for bilateral fitting re. unilateral fittings (Scollie *et al.* 2005) for speech. However, Van Beurden *et al.* (2018) have shown that binaural loudness summation depends on the amount of hearing loss and on the level of presentation. Furthermore, Oetting *et al.* (2016) and van Beurden *et al.* (2018) showed that spectral loudness summation was a few decibels larger in bilateral presentation than in unilateral presentation.

Ewert and Oetting (2018) investigated if the differences in loudness perception between normal hearing and hearing-impaired listeners in the studies by Oetting *et al.* (2016) and van Beurden *et al.* (2018) could be attributed to narrowband loudness compensation or to differences in loudness summation in normal hearing and hearing-impaired listeners. They used an equal categorical loudness noise (ECLN) that was composed of the six narrowband noises used in the narrowband loudness compensation applied in the study by Oetting *et al.* (2016). The levels of the six narrowband noises were adjusted to produce an equal categorical loudness. Hearing-impaired listeners showed lower narrowband loudness values compared to normal hearing listeners, indicating an increased spectral loudness summation. Also, in the case of bilateral presentation seven out of ten hearing-impaired listeners showed higher spectral loudness summation compared to normal hearing listeners.

The analysis of the effects of the different audiometric configurations on spectral loudness summation by van Beurden *et al.* (2018) suggested that for hearing-impaired listeners the low frequency components were more important in spectral loudness summation than the high-frequency components. Such an unequal distribution of perceptual weights has been observed earlier in normal hearing subjects, where the highest weight was given to the lowest noise band for a signal composed of three noise bands (Oberfeld *et al.,* 2012) and to the highest and lowest frequencies for a ten-tone complex (Joshi *et al.* 2016).

Finally, current prescription rules implicitly assume that loudness growth is the same for subjects with the same hearing loss. However, Oetting *et al.* (2016), van Beurden *et al.* (2018), and Rasetshwane *et al.* (2018) all showed that individual differences in spectral loudness summation were substantial, even for subjects with similar degrees of hearing loss. Large inter-individual differences in the loudness domain have been found earlier in literature on the measurement of LDL. In several large-scale studies (e.g. Kamm *et al.* 1978, Formby *et al.* 2017) LDLs were found to vary considerably for subjects with different hearing losses but were also found to vary significantly for subjects with the same degree of hearing loss. Formby *et al.* (2017) reported loudness judgments within the same loudness category to vary across a group of 30 normal hearing listeners by as much as 50 to 60 dB.

Nelson *et al.* (2018) allowed a group of mild to moderate hearing-impaired listeners to self-adjust hearing aid gain for speech understanding in a variety of quiet and noisy listening conditions. They also found a large between-subject variability with the range of selected gains spanning about 40 dB. Perry *et al.* (2019) analyzed the individual variability of gain values chosen in several studies on self-adjustment and concluded that the range of selected gains depended on the range of gain values made available to the subjects. In audiologist driven fittings, the range of available gain values may be unconsciously limited by the tendency of an audiologist not to deviate too far from the gain proposed by the prescription rule. This may explain why we usually do not observe this large ranges in routine hearing aid fitting.

The present study was designed to further investigate the interaction between spectral content and binaural loudness summation with a focus on the role of low and high frequencies in spectral and binaural loudness perception. For this purpose, the loudness of a broadband noise was compared to the loudness of its lower and higher frequency part. The loudness of these three signals was measured with loudness scaling for both unilateral and bilateral presentation for a group of hearing-impaired listeners. The results were analyzed in terms of Level Difference at Equal Loudness

(LDEL) for spectral effects in unilateral and bilateral presentation, binaural effects, and inter-individual differences.

Loudness scaling provides information on the entire loudness range but is a fairly time-consuming procedure. While information on individual loudness growth appears to be indispensable for a good hearing aid fitting, it may not be necessary to measure loudness growth for each separate presentation level, when the loudness differences between signals are known for a representative presentation level. As measurement time is an important limitation in the clinic, a second objective was to investigate if loudness matching could be suitable as a more time efficient alternative to measure these loudness differences.

6.2 Methods

6.2.1 Subjects

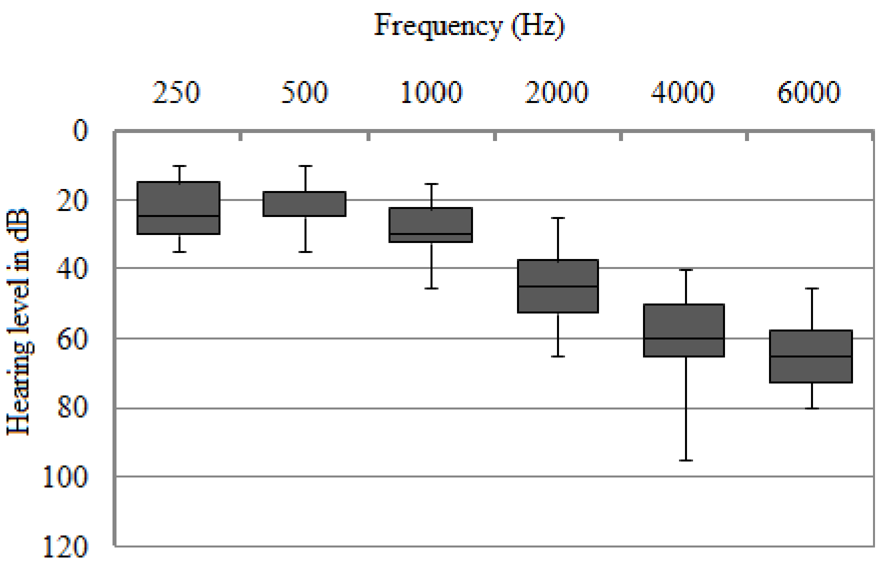


Figure 6.1. *The distribution of the audiograms for the sixteen subjects. Whiskers mark minimum and maximum values.*

Sixteen adult hearing-impaired listeners participated in the study. Inclusion criteria for the hearing-impaired listeners were native Dutch speakers with mild to moderate symmetrical hearing losses (differences between both ears at 0.5, 1, 2, and 4 kHz <10 dB) selected from clinical files. Seven men and nine women participated with an average age of 68 years. The distributions of the hearing thresholds for all listeners are given in figure 6.1. Whiskers mark minimum and maximum values.

6.2.2 Equipment

All measurements were conducted in a sound-insulated booth in a session of about 2 hours. Pure-tone audiograms with air and bone conduction were measured less than 4 weeks earlier with DECOS audiometers using TDH39 headphones. Sennheiser HDA 200 headphones were used for both the categorical loudness scaling procedure and for the loudness matching procedure using the experimental approach described by Ewert (Ewert, 2013). Signals were presented using a RME Fireface UC DA convertor at 44.1 kHz sampling frequency. Headphones were calibrated with a Brüel & Kjær artificial ear type 4153, a 0.5-inch microphone type 4134, a microphone preamplifier type 2669, and a measuring amplifier type 2610. Headphones were free-field equalized according to ISO 389-8 (2004) and levels are expressed as the equivalent free-field levels in dB SPL(FF).

6.2.3 Stimuli

The stimuli were a broadband pink noise (BB) with a bandwidth between 100 Hz and 16000 Hz and low-pass and high pass filtered versions of this same pink noise (LP and HP, respectively), see figure 6.2. The unilaterally presented signals were always presented to the left ear. The LP noise and HP noise were obtained by filtering the original BB noise with an eight order Butterworth filter with a cut-off frequency of 1400 Hz. The perceptual center of the BB noise would be approximately 1700 Hz. The cut-off frequency was chosen a few 100 Hz lower to put it more in line with the point where the hearing loss in the audiogram starts to drop. All stimuli were 1-s noises with 50-ms rise and fall ramps.

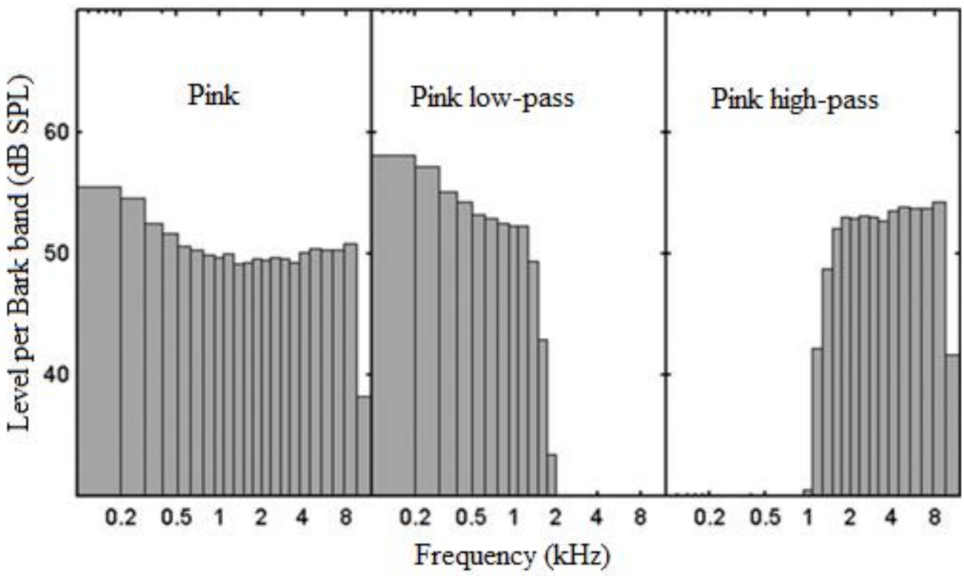


Figure 6.2. *Bark spectrum of the three test signals at 65 (dB SPL).*

6.2.4 Procedures

6.2.4.1 Loudness scaling

Categorical loudness scaling was performed using the ACALOS procedure (Brand and Hohmann, 2002) to measure the loudness perception over the whole dynamic range. During the loudness-scaling procedure listeners had to rate the perceived loudness on an 11-point scale from “not heard” to “too loud”. The scales were transformed into numerical values in “Categorical Units” (CU) from 0 to 50. Stimuli were presented in a pseudo-random order with levels between -10 to 105 dB SPL. A monotonically increasing loudness function was fitted to the responses for each of the ACALOS measurements using the BTUX fitting method (Oetting *et al.*, 2014). The model function consists of two linear parts with independent slopes m_{low} and m_{high} with a smooth transition range (see Brand and Hohmann, 2002).

Table 6.1. *Overview of the fifteen pairs of low-pass filtered (LP), high-pass filtered (HP) and broadband (BB) pink noise signals that were matched in this study. The columns show the reference signals and the rows the test signals.*

Reference → / Test Signal ↓	HP Unilateral	BB Unilateral	LP Bilateral	HP Bilateral	BB Bilateral
LP Unilateral	1	2	3	4	5
HP Unilateral	--	6	7	8	9
BB Unilateral	--	--	10	11	12
LP Bilateral	--	--	--	13	14
HP Bilateral	--	--	--	--	15

6.2.4.2 Loudness matching

Loudness matching was performed to measure the loudness of 15 pairs of signals, see table 6.1. In the loudness scaling procedure the subjects always judged the loudness of a unilateral, or bilateral presented signal. In the loudness matching procedure subjects also compared unilateral signals with bilateral signals. Loudness matching was conducted using a two-alternative forced-choice procedure with a one-up one-down adaptive rule, converging to the 50% point on the psychometric function (Levitt, 1971). In each trial the subject heard two sounds, the reference signal and the test signal. The silent interval between the signals was 500 ms. The test and reference signals were presented in random order and with equal *a priori* probability. The subject indicated which signal was louder by selecting the first or the second signal on a touchscreen. If the subject indicated that the test signal was louder, its level was reduced in the next trial and vice versa. The initial step size was 10 dB. This was decreased to 5 dB after the first upper reversal and to 3 dB after the second upper reversal. The maximal presentation level was set to 105 dB SPL. All comparisons were interleaved to reduce biases that occur when stimuli from only one stimulus pair are matched in loudness in a series of trials (Florentine *et al.*, 1996, Verhey and Kollmeier, 2002).

The reference level in the loudness matching procedure was chosen as the level where the loudness function of the loudness scaling procedure corresponded to 35 CU (“loud”). Setting the reference level in the matching procedure equal to the level at CU 35 in the scaling procedure ensured that the results of both procedures could be analyzed at equal loudness for all subjects. Previous studies showed large individual differences in loudness curves. The choice of a common reference level (i.e. 90 dB SPL) would have led to a different point on the loudness curve for each subject, making comparisons between subjects and procedures very complex. In this study, the main interest was on the loudness at the high end of the loudness curve, where complaints of discomfort about loud sounds can occur. The 35 CU level was chosen for being the highest point on the loudness curve that provided enough room for the matching procedure to move around the point of equal loudness without reaching levels that could be perceived as too loud. As the loudness scaling results show that loudness measurements at 35 CU are closely related to loudness measurements at 50 CU, we assume that the measurements at 35 CU provide a first approximation for the loudness perception of discomfortable loud sounds. Ideally the start level of the test signal lies equally often above and beneath the estimated level of equal loudness to avoid range effects. For safety reasons the test level always started beneath the estimated level of equal loudness at a level corresponding to 25 CU of the loudness function. This might cause a bias towards lower loudness levels. No level-roving was applied. The safety limit of 105 dB SPL also caused a few missing data points, as in some cases the level at 35 CU was above our safety limit of 105 dB SPL. In these cases, no starting point could be determined for the matching procedure. In other cases, the matching procedure led to matches with levels above this same limit. In that case, the result was also undetermined.

6.2.4.3 Measurement protocol

There was a fixed order starting with loudness scaling followed by loudness matching, because the results from the loudness scaling procedure were used as input for the reference levels in the loudness matching procedure. The whole measurement set was repeated once in the same session to obtain a retest. In the retest the reference levels for the loudness matching procedure were based on the retest results of the loudness scaling procedure rather than on the test results. The reference levels in the retest could therefore deviate from the reference levels in the test.

To complete the loudness scaling procedure for one stimulus approximately two minutes were needed. Measurement time for all six conditions (three signals, both unilateral, and bilateral) mounted up to twelve minutes. The loudness matching procedure was faster: fifteen conditions could be matched in twelve minutes.

6.3 Results

6.3.1 Effect of frequency spectrum

The main interest of this study was on loudness differences because of different frequency spectra (low-pass, high-pass, broadband) and different presentation modes (unilateral vs. bilateral) in hearing-impaired listeners. These differences have been defined as LDEL at 35 CU.

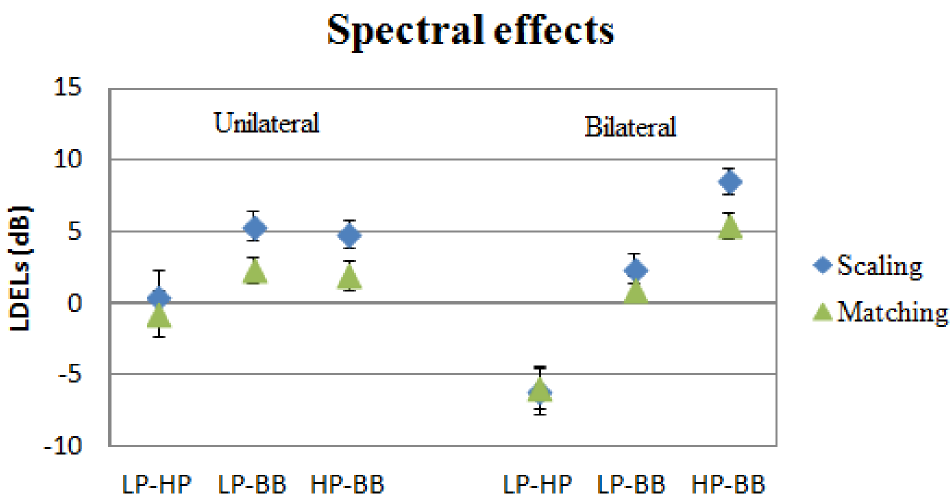


Figure 6.3. *LDEL at a loudness level of 35 CU for unilateral and bilateral presentation. Diamonds represent loudness scaling data, and triangles represent loudness matching data. The first signal is the test signal, and the second signal the reference signal. Positive values indicate that the test signal reaches equal loudness at a higher level than the reference signal. The error bars indicate the standard error of the average spectral loudness effect across all sixteen hearing-impaired subjects.*

Figure 6.3 shows the loudness differences between the three stimuli for unilateral and bilateral presentation. The loudness differences were first calculated for test and retest separately and then averaged. Positive values were expected when the reference signal was the broadband signal and the test signal was a filtered signal (LP-BB, HP-BB).

The error bars indicate the standard error of the average spectral loudness effect across all sixteen hearing-impaired subjects. Both procedures showed the same trends. For unilateral presentation, the LDEL for LP-HP was approximately zero. As expected, the LDELs for LP-BB and HP-BB were both positive, indicating that the loudness of 35 CU was reached at a lower level for BB noise than for HP or LP filtered pink noise.

For bilateral presentation the results were different. The LP-HP showed LDEL levels of around -6 dB, signifying that the LP noise levels were 6 dB lower compared to the HP noise levels at equal loudness. The LDEL for LP-BB was around 2 dB whereas the LDEL for HP-BB was around 7 dB. The difference of 5 dB is in the same range of the LDEL for the direct comparison of LP-HP. As will be shown below the differences in spectral effects between unilateral and bilateral presentation appear to be associated with a higher binaural loudness summation for LP noise than for HP noise.

A three-way repeated measures ANOVA with spectral loudness differences (LP-HP, LP-BB, and HP-BB), procedure (loudness scaling, loudness matching), and presentation mode (unilateral, bilateral) as within-subjects variables was conducted. In case sphericity could not be assumed, Huynh-Feldt correction was applied. All main effects were significant ($F(1,15.4)=14.9$, $p=0.001$, $F(1,15)=10.6$, $p=0.005$, $F(1.15)=6.7$, $p=0.021$ for spectral loudness differences, procedure and presentation mode, resp.). The interaction effect between presentation mode and spectral loudness differences was also significant ($F(1.4,20.8)=26.0$, $p=<0.001$). Bonferroni corrected post hoc analysis showed a significant difference of 1.9 dB between loudness scaling and loudness matching ($p<0.001$), a significant effect between unilateral and bilateral presentation ($p=0.018$), and highly significant differences between LP-HP and LP-BB ($p<0.001$), LP-HP and HP-BB ($p<0.001$), and LP-BB and HP-BB ($p=0.031$). Because of the significant interaction effect between presentation mode and spectral loudness differences, separate two-way repeated measures ANOVAs were calculated for both unilateral and bilateral presentation to investigate the spectral differences in more detail. Bonferroni corrected post hoc analysis showed for unilateral presentation only a highly significant difference of 4.0 dB between LP-HP and LP-BB. For bilateral presentation highly significant differences were found between all conditions with a mean difference between LP-HP and LP-BB of 7.7 dB, a mean difference between LP-HP and HP-BB of 13.0 dB, and a mean difference between LP-BB and HP-BB of 5.2 dB. For bilateral presentation the mean differences have to be interpreted with caution as there was a highly significant interaction effect between procedure and spectral loudness differences ($F(2.30)=8.2$, $p<0.001$).

Calculating mean differences with standard errors, as in figure 6.3, obscures the view on the individual variability. In figure 6.4, the levels of the test signal in the matching procedure are plotted as a function of the levels of the categorical loudness scaling procedure at 35 CU. As a result of the cross-over design in the loudness matching procedure the loudness of a signal in the loudness matching procedure was measured with respect to several different reference signals. For instance, the monaurally presented low-pass filtered noise was the test signal in five different comparisons. This means that these five conditions should lead to the same loudness estimate. The points shown for the loudness matching procedure in figure 6.4 are averages of all the conditions in which the signal was the test signal. Note that the number of conditions is different for the five conditions, decreasing from five to two (see table 6.1). As binaural BB noise was never the test signal in the loudness matching procedure no data points are available for this stimulus. Figure 6.4 shows that the individual variability is large. The lowest and highest level of the monaural LP noise at which subjects rated the loudness as “loud” (35 CU) varied from below 70 dB to above 100 dB in both procedures. Comparable ranges of 30 dB between lowest and highest level for a loudness of 35 CU are found for the other signals. Be aware that the signals were presented unaided.

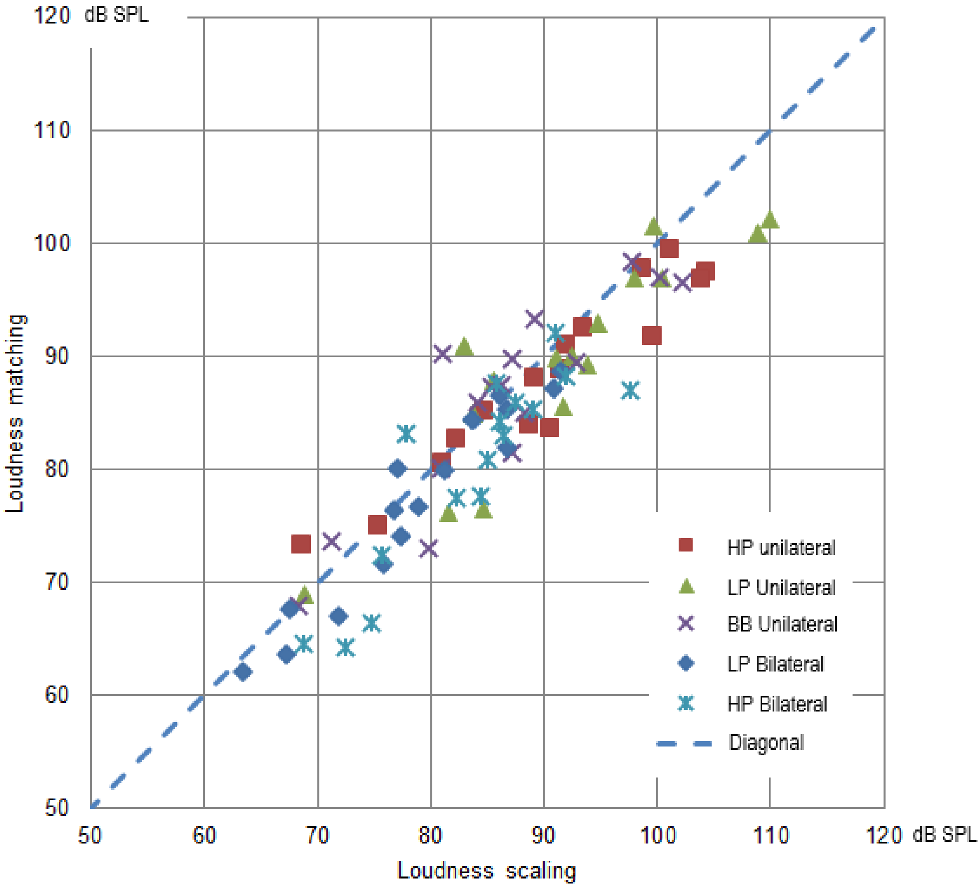


Figure 6.4. *Absolute values for the level at which the signal is perceived to have a loudness equal to 35 CU. The levels for loudness matching are plotted as a function of the levels for loudness scaling with each marker representing one subject.*

6.3.2 Binaural effects

Figure 6.5 shows binaural summation at a loudness level of 35 CU for all three stimuli. Binaural summation expressed as LDEL was first calculated for test and retest separately and then averaged. Error bars indicate the standard error. Figure 6.5 clearly shows differences in LDEL. Binaural summation was largest for the LP noise and lowest for the HP noise. A two-way repeated measures ANOVA with signal (LP, HP, and BB) and procedure (loudness scaling, loudness matching) as within-subjects variables was conducted. All main effects were significant ($F(1,15)=8.28$, $p=0.012$, $F(2,30)=25.34$, $p<0.001$), but the interaction effect was not ($F(2,30)=2.4$, $p=0.108$). Bonferroni corrected post-hoc analysis showed a highly significant difference of 2.0 dB between loudness scaling and loudness matching ($p=0.001$), and highly significant differences between LP noise and HP noise and between HP noise and BB noise ($p<0.001$). The difference between LP noise and BB noise was at the significance level ($p=0.050$).

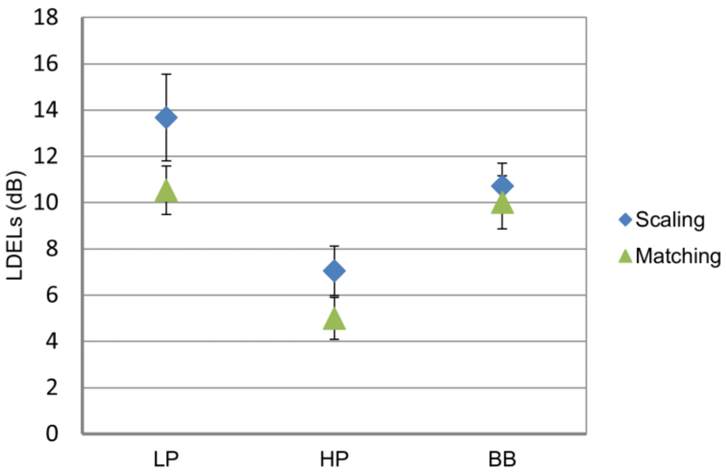


Figure 6.5. *LDEL between unilateral and bilateral presentation for the three signals.*

6.3.3 Intra-test reproducibility

To assess the intra-test reproducibility of both procedures correlations between test and retest were calculated for both procedures. For the loudness scaling procedure, the levels at 35 CU were used for these correlations, as this was the loudness used as the reference level in the loudness matching procedure. For all six conditions (three signals, unilateral and bilateral presentation) test-retest correlations for loudness scaling were high with correlation coefficients above 0.80.

The correlation coefficients for the loudness matching procedure were slightly lower than for the loudness scaling procedure. The average correlation coefficient across all fifteen conditions was 0.81, with a range from 0.70 for HP(B)-LP(B) to 0.89 for HP(B)-Pink(M). The lower correlation coefficients in the matching procedure partly originate from the choice to base the reference levels in the retest conditions of the matching procedure on the outcome of the retest conditions of the scaling procedure. Because of this choice the reference levels differ for test and retest, introducing an extra source of variance.

The pooled standard deviation across subjects ($\sqrt{\frac{1}{16}\sum \sigma_i{}^2}$,with σ_i the standard deviation of each subject) calculated for each condition showed a range of 2.9-4.3 dB for loudness scaling and a range of 3.2-6.3 dB for loudness matching. The highest pooled standard deviation was for the conditions with the low-pass filtered pink noise. Calculations of pooled standard variations across conditions for each subject showed a range of 1.7-5.4 dB for loudness scaling and 1.8-8.0 dB for loudness matching. Pearson's R correlation coefficient between the pooled standard deviations of the loudness scaling procedure and the loudness matching per subject was 0.7, showing that on average subjects with larger standard variations in loudness scaling also showed larger standard variations in loudness matching. The pooled standard deviation across all subjects and conditions amounted to 3.7 dB for loudness scaling and 4.5 dB for loudness matching. Note however that in the loudness matching procedure almost three times as many conditions were measured as in the loudness scaling procedure.

6.3.4 Correspondence between Loudness scaling and Loudness matching.

One of the main goals of this study was to compare loudness matching with loudness scaling. The previous sections showed that both procedures produced significantly different results in spectral and binaural effects. However, this appears to be mainly caused by a static offset between the two procedures. The offset is probably a result of the choice to set the starting level of the test signal always below the level of the reference signal. Verhey

(1999) showed that the starting level of a matching procedure influenced the amount of spectral loudness summation. Figure 6.3 shows that both procedures are well correlated. Correlations were calculated for the data shown in figure 6.3. The results show that correlations for all signals were excellent (0.89 and higher).

6.4 Discussion

6.4.1 Spectral effects

Spectral loudness summation is investigated generally by varying the bandwidth of a noise or tone complex geometrically centered around a certain frequency, where the center frequency is usually in the range between 1000 and 3200 Hz (e.g., Zwicker *et al.,* 1957; Verhey and Kollmeier, 2002; Anweiler and Verhey, 2006; Bonding and Elberling, 1980; Rasetshwane *et al.* 2018). As a consequence of varying the center frequency of the signals in our experiments, the results of the current study cannot be compared directly with other studies on spectral loudness summation. However, it is at least expected that in accordance with the literature the broadband pink noise will show the largest spectral loudness summation effect. Zwicker *et al.* (1957) found that loudness was less strong for a four-tone complex with equal sound pressure level for each tone centered around 500 Hz than for a four-tone complex centered around 2000 Hz when the spacing in Hz between the tones was identical. However, in this study in unilateral presentation no difference in loudness was found for the two signals. This may be caused by the impact of hearing loss. Hearing loss has been shown to decrease spectral loudness summation (e.g. Scharf and Hellman, 1966; Florentine and Zwicker, 1979; Garnier *et al.* 1999, Rasetshwane *et al.*, 2018). As the subjects had more hearing loss in the higher frequencies, this may have decreased loudness for the high-pass filtered pink noise. Following this argumentation, the results for unilaterally presented signals are reasonably in line with existing knowledge.

However, the most striking result of this study is the large difference in spectral effects between unilateral and bilateral presentation. While the broadband pink noise still has the lowest level at equal loudness in bilateral presentation (in accordance with the current knowledge on spectral loudness summation), the balance between the low-pass filtered pink noise and the high-pass filtered pink noise is shifted dramatically when unilateral presentation is changed to bilateral presentation. This has not been described in other studies so far, as to our knowledge spectral loudness summation has only been studied monaurally. The main reason only monaural summation has been studied is probably the assumption that binaural loudness summation has no interaction with spectral loudness summation (i.e., Moore *et al.,* 2014). However, there are studies that point to higher binaural loudness summation for broadband signals compared to narrow-band signals (e.g. Zwicker and Zwicker, 1991; Oetting *et al.* 2016, 2018; Algom *et al.*, 1989; Scharf, 1968). Oetting (2016) presented an effective extension of Zwicker's loudness model to consider the increased binaural summation for broadband signals. In our data the large differences in spectral loudness effects between the low-pass and the high-pass filtered pink noises indicate large differences in binaural loudness summation for the two signals.

Hawkins *et al.* (1987) measured binaural loudness summation with three different paradigms for both normal hearing and hearing-impaired listeners. The amount of binaural loudness summation strongly depended on measurement paradigm. With a loudness balance paradigm similar to the loudness matching procedure in this study Hawkins *et al.* (1987) found binaural loudness summation for hearing-impaired listeners to be slightly lower for the 4000 Hz pure tone, than for the 500 Hz pure tone (8.1 dB vs. 10.3 dB, resp.), which is in qualitative agreement with our data.

The unexpected large differences in LDEL between LP noise and HP noise when comparing the unilateral and bilateral condition (binaural level difference for equal loudness; BLDEL) was further investigated by plotting the individual BLDEL measured with the loudness scaling procedure for the filtered signals against the BLDEL of the BB, see figure 6.6. Individual values were between 1 dB and 25 dB. On average a larger BLDEL for BB noise is associated with a larger binaural loudness summation for LP noise and HP noise. This trend is stronger for LP noise than for HP noise. The slope of the linear regression line for LP noise is 0.78, the slope for HP noise is 0.54.

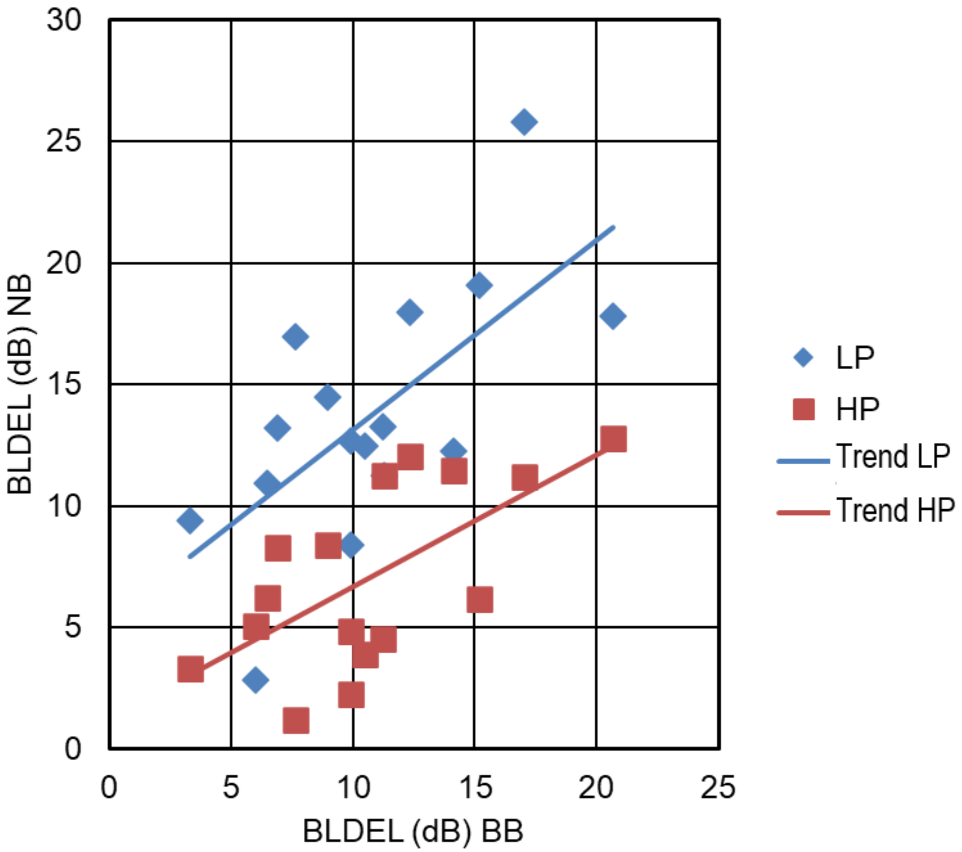


Figure 6.6. *BLDEL are shown for LP noise and HP noise as a function of the BLDEL for BB noise. Trend lines are added for clarity. Each marker represents a single subject.*

The individual results substantiate that the mechanism responsible for the increased binaural summation has its major effect at the frequencies below 1400 Hz. A speculative hypothesis could be that hearing loss does not only lead to tonotopic reorganization in the cochlea, but also leads to perceptual reorganization at more central levels of the hearing system. Indications of tonotopic map changes in human auditory cortex following hearing loss have been shown (for a review see Eggermont, 2017), and increases in central gain have also been postulated as a possible explanation for tinnitus and hyperacusis (i.e. Auerbach *et al.*, 2014).

6.4.2 Reproducibility

Loudness matching yielded lower correlation coefficients for test-retest than loudness scaling. However, part of the differences can be explained by a shift in reference level in the loudness matching procedure between test and retest, as in the loudness matching procedure the reference levels were always based on the outcome of the corresponding loudness scaling measurement. The reference levels in the retest of the loudness matching procedure were therefore based on the levels of retest of the loudness scaling procedure. As the correlation between test and retest was not perfect in loudness scaling this introduced extra variability in the loudness matching procedure. Analysis of the correlation coefficients for conditions of loudness matching within the same test showed higher correlations that are well in agreement with the values found in loudness scaling. The intra-test correlation coefficients for the loudness matching procedure were better for the retest results than for the test results. This may indicate a learning effect for some subjects. The inter-test correlations between loudness scaling and loudness matching were excellent with an average correlation coefficient over the five conditions of 0.92 in both the test as well as the retest results. Taken together, the loudness matching procedure seems to provide reliable and reproducible results, that are well in agreement with the results of the loudness scaling procedure.

6.4.3 Clinical implications

This study clearly shows that hearing-impaired listeners can already perceive bilaterally presented broadband signals as loud at relatively low sound pressure levels (see figure 6.4). If we want to ensure that hearing-impaired listeners aided with two hearing aids do not experience discomfort from loud sounds, we need to pay special attention to the effects of spectral and binaural loudness summation. Oetting *et al.* (2018) presented a dynamic compressor considering the bandwidth of the input signal for the gain calculation. After narrow-band loudness compensation, they measured aided loudness functions for signals with different bandwidths. The deviations of these loudness functions from average normal hearing loudness functions were used for gain corrections in the bandwidth-adaptive dynamic compressor. Normal loudness for natural signals with different bandwidths could be restored if these gain corrections made. The calculated gain corrections after narrowband loudness compensation showed large interindividual differences for binaural broadband signals.

The current prescription rules ignore the variability in loudness perception between individual subjects. In the current data at a loudness level of 35 CU the level difference between the subject with the highest and lowest sensitivity was more than 25 dB. And this range increased further for loudness levels above 35 CU. This is in agreement with the findings from Formby *et al.* (2017) who found loudness judgments within the same loudness category to vary across listeners by as much as 50 to 60 dB. Van Beurden *et al.* (2018) found comparable ranges between lowest and highest levels judged to have equal loudness for international female noise and uniformly exciting noise of 17 Bark bandwidth, presented unilaterally and bilaterally after narrow-band loudness compensation. These data therefore show that loudness compensation based on narrow-band measurements, does not guarantee that loudness for broadband signals will be well compensated as well.

As individual loudness functions for broadband signals cannot be predicted from narrowband measurements and differ for unilateral and bilateral presentation additional measurements are needed to assess spectral loudness summation and binaural loudness summation. Our results suggest that only adding a unilateral and bilateral measurement of a broadband signal, will not suffice to predict summation effects for all signals. Measurements of band pass filtered signals appear to be needed to provide detailed information. To determine an appropriate minimal set additional research is needed. As loudness scaling is fairly time consuming, a faster measurement procedure may be needed to make hearing aid fitting based on individual loudness perception clinically feasible. Loudness matching is shown to provide reliable results in approximately one-third of the time needed for the ACALOS-procedure.

The huge differences in loudness perception between hearing-impaired subjects for especially bilaterally presented broadband stimuli make it highly unlikely that one generic prescription rule will be able to predict an acceptable amplification for loud sounds for all hearing-impaired listeners. Hearing aid fitting should be based on individual loudness growth measures.

Furthermore, the individual differences in spectral and binaural loudness summation and the interactions between these two make it highly unlikely that the current options to control the Maximum Power Output (MPO) of a hearing aid will provide enough flexibility to properly reduce the amplification at the upper part of the loudness scale. An adaptive compressor as proposed by Oetting *et al.* (2018) will be required to cope with the large individual differences.

6.5 Conclusions

The current study challenges some of the common assumptions on how loudness perception is altered by hearing loss. Bilateral presentation of a signal is usually assumed to increase loudness perception with a fixed amount regardless of the bandwidth of the signal. However, loudness perception measurements in this experiment clearly showed a different bandwidth dependency in bilateral presentation than in unilateral presentation. The results from a group of hearing-impaired listeners included in this study suggest that the low frequency part of the signal has a stronger contribution to the total loudness sensation than the high frequency part in bilateral presentation compared to unilateral presentation. This appears to be a consequence of more binaural loudness summation for the low frequency part of the spectrum than for the high frequency part.

Secondly, individual differences in loudness perception usually are not taken into account in hearing-aid prescriptions. The current study however confirms that individual variability in loudness perception near loudness discomfort levels is too large to ignore in hearing aid amplification. Measuring individual loudness perception for broadband and binaural signals should be considered in hearing aid fitting procedures. Which set of signals would be the best choice, needs to be investigated further. This study shows that loudness matching appears to be a reliable and time-efficient procedure to quickly measure loudness differences between signals with larger bandwidths and presentation modes (unilateral vs. bilateral).

Chapter 7

General discussion

7.1 Introduction

When someone's hearing decreases as a result of sensorineural hearing loss, the threshold of hearing increases, but in most cases the threshold of uncomfortable loudness does not increase by the same amount. The threshold of uncomfortable loudness generally remains the same and can even decrease. The dynamic range, the area between hearing threshold and the threshold of uncomfortable loudness, is therefore smaller for most hearing-impaired listeners than for normal hearing listeners. Due to the smaller dynamic range hearing-impaired listeners experience altered loudness perception. Often recruitment occurs, i.e. a higher than normal growth in perceived loudness as the signal level increases.

When the degree of hearing loss exceeds a certain limit, generally hearing aids are recommended to compensate for the hearing loss. Hearing aids are basically amplifiers. They provide frequency-dependent gain to compensate for the frequency-dependent hearing loss. To compensate for the altered loudness growth, a more sophisticated type of signal processing is applied: compression. Compression serves two purposes. The first purpose of compression is to prevent high input levels becoming too loud without peak-clipping. The second purpose is to restore loudness growth in hearing-impaired listeners as good as possible.

To prescribe the right amount of amplification in hearing aids, prescription rules have been available since 1940 (Watson and Knudsen, 1940). In the 1980s these prescriptive formulas were based on linear (level-independent) amplification. For example, National Acoustic Laboratories-Revised (NAL-R; Byrne and Dillon, 1986), prescription of gain and output (POGO; McCandless and Lyregaard, 1983), Libby one-third gain (Libby, 1986), and Berger (Berger *et al.*, 1980). Some 10 years later the first nonlinear prescription rules started to emerge such as DSL i/o (Cornelisse *et al.*, 1995), IHAFF (Cox, 1995), FIG6 (Killion and Fikret-Pasa, 1993), and NAL-NL1(Byrne *et al.*, 2001). DSL i/o, NAL NL1 and his successor NAL-NL2 (Keidser *et al.*, 2011) are currently the most used prescription rules worldwide. And although in both prescription rules, loudness normalization is an important aspect, both rules follow different rationales. DSL i/o aims to normalize loudness perception based on comfortable listening levels across hearing levels in order to maximize speech intelligibility (Kamm *et al.*, 1978; Pascoe, 1978 ; *Gengel et al.*, 1971; Erber and Witt, 1977; Macrae, 1986; Smith and Boothroyd, 1989).

The NAL-NL 1 and 2 prescription rules aim to optimize speech intelligibility with a constraint on overall loudness. This approach equalizes loudness across the mid-frequencies (Byrne *et al.*, 2001; Johnson and Dillon, 2011). In the NAL-NL prescription rules (NAL-NL1, NAL-NL2) the overall loudness is calculated with the loudness model of Moore and Glasberg (1997, 2004). Despite all efforts to normalize loudness, complaints about aversiveness of loud sounds remain an important reason for dissatisfaction with hearing aids (Kochkin, 2000; Jenstad *et al.,* 2003; Boymans *et al.*, 2008; Hickson, *et al.*, 2010; Franks and Beckmann, 1985; EuroTrak, 2018).

In this thesis three limitations of the current approach to normalize loudness perception with hearing aids are addressed. *The first limitation concerns the loudness perception of dynamic sounds.* In the last decades the computing power of hearing aids has increased dramatically, making it possible to realize gain adjustments on small time scales. However, loudness models as the Dynamic Loudness Model by Chalupper and Fastl (2002) and the model of loudness applicable to time varying sounds by Glasberg and Moore (2002), do not correctly predict loudness perception of dynamic sounds (Rennies *et al.*, 2009; 2010). This shows that the exact mechanism in which loudness is processed by our hearing is still not completely understood. To take optimal advantage of the increased computing power, a better understanding about loudness perception of dynamic sounds in normal hearing listeners and about the changes in loudness perception of dynamic sounds in hearing-impaired listeners is needed.

The second limitation is that the loudness function in the hearing aid prescription rules is only based on measurements of the auditory threshold. However, Marozeau and Florentine (2007) have shown that the loudness growth function can have several forms above threshold. And for instance, Formby *et al.* (2017) have shown that in a group of normal hearing listeners loudness judgments within the same loudness category can differ in intensity by as much as 50-60 dB.

Finally*, the third limitation is that it is assumed that spectral and binaural loudness summation are not influenced by frequency specific gain changes.* However, Oetting *et al.* (2016) showed that restoring narrowband loudness perception to normal does not ensure normal loudness perception for broadband signals in hearing-impaired listeners due to changes in spectral loudness summation. If the sounds are presented binaurally (which is usually the case) also the effects of binaural loudness summation should be taken into account.

The first part of this thesis (chapters 2 and 3) focuses on the first limitation and deals with dynamic aspects of loudness perception in normal hearing listeners. It gives more insight into the loudness effects at short time scales. The loudness models applied in hearing aid prescription rules have been developed for stationary sounds. They are not suited to predict interactions between signal properties such as signal duration and signal bandwidth (chapter 2) and the interactions between signals with different onset times (chapter 3).

The second and third limitations are addressed in chapters 4, 5 and 6. In the second part of this thesis (chapters 5 and 6) the interaction between spectral loudness summation and presentation mode (unilateral or bilateral) in hearing-impaired listeners is further investigated. The individual variability often seen in loudness measurements is another important issue in part two of this thesis, and is also one of the topics in the intermezzo in chapter 4. The fact that loudness is not a physical quantity that can be measured objectively, but always contains a subjective element, makes loudness measurements susceptible to biases. Loudness measurements depend on the judgment of the individual listener, and this judgment can be influenced by prior experiences with loud sounds, mood, interpretation of the instruction, etc...Therefore full account must be taken of the measurement procedure itself, because this can influence the results, as especially presentation order and stimulus range can give rise to biases. The results presented in this thesis question the possibility of calculating an appropriate individual gain from a threshold-based loudness model. If the latter is not possible, individual loudness measurements for a collection of narrowband and broadband stimuli appear to be necessary to fulfill the aim of loudness normalization for an individual hearing-impaired listener.

7.2 Spectral loudness summation and temporal integration

Loudness perception has been studied extensively from the end of the 19th century, when Weber and Fechner started to study the human response to a physical stimulus in a quantitative fashion. Since those early days much has been learned about loudness perception and several loudness models have been proposed, initially for steady signals (i.e. Fletcher and Munson, 1933; Zwicker and Scharf, 1965; Moore and Glasberg, 1996, 2004; Moore *et al.*, 1997) and later also for dynamic signals (Glasberg and Moore, 2002; Moore *et al.,* 2016; Chalupper and Fastl 2002). The loudness models for steady signals are mainly developed to calculate loudness growth (increase of loudness with increasing intensity) and spectral loudness summation (increase of loudness with increasing bandwidth). The dynamic models aim to calculate temporal integration, the increase in loudness with increasing signal duration, as well.

Spectral loudness summation is assumed to be a consequence of the combination of two properties of the auditory system: spectral filtering and cochlear compression by a compressive non-linearity. Spectral filtering refers to the finding that the auditory system processes intensity separately in critical bands and that the overall loudness is determined by summing the specific loudnesses in each critical band. Since the calculation of the specific loudness values normally encompasses a compressive non-linearity, the sum of specific loudnesses yields a larger overall loudness than the overall loudness of a signal which has the same intensity but excites only one band. This phenomenon is called spectral loudness summation.

In most studies, spectral loudness summation, and temporal loudness integration are investigated separately (for a review see, Verhey and Kollmeier, 2002; Anweiler and Verhey, 2006). Only a few studies investigated both aspects of loudness perception. When comparing spectral loudness summation between short and long signals it is important to define the common reference. Loudness summation can be compared at equal loudness level of the reference signal, or at equal intensity level. In the first case, the intensity level of the short signals will be higher than for the long signals, as temporal integration increases the loudness of the long signals. Earlier studies (Zwicker, 1965; Port, 1963a) showed that spectral loudness summation was the same for short and long signals given an identical reference loudness. At equal reference level, this would mean that spectral loudness summation is smaller for short signals than for long signals.

However, studies on temporal loudness integration (Florentine *et al.*, 1996; Buus *et al.*, 1997) suggested that spectral loudness summation should be the same for short and long sounds at equal reference level. As the loudness growth curve is compressed at medium intensities, equal spectral loudness summation at equal reference level implies that at equal reference loudness spectral loudness summation should be larger for short than for long sounds, which is qualitatively in agreement with the results by Boone (1973). In contradiction with these results other studies found spectral loudness summation not to be the same, but to be even larger for short signals than for long signals at equal level (Verhey and Kollmeier, 2002; Anweiler and Verhey, 2006; Verhey *et al.*, 2006).

In chapter 2 we investigated the research question how bandwidth influences loudness perception of short stimuli and series of noise bursts. The results of the first experiment in chapter 2 correspond with the studies on time-dependent spectral loudness summation (Verhey and Kollmeier, 2002; Anweiler and Verhey, 2006), showing larger spectral loudness summation for shorter signals at equal reference level. In the second experiment, a series of short stimuli combined in a pulse train also led to larger loudness for larger bandwidths. This is in agreement with the results from Verhey and Uhlemann (2008). They found that up to repetition rates of 50 Hz, the magnitude of spectral loudness summation for the sequences of noise bursts was the same as for the single short noise burst. The highest repetition rate applied in experiment 2 was 40 Hz. To be in full agreement with the results of Verhey and Uhlemann (2008) the magnitude of spectral loudness summation in experiment 2 should be determined by the magnitude of spectral loudness summation in a single noise burst. In that case the magnitude of spectral loudness summation was expected to be larger for the 12.5 ms noise bursts than for the 25 ms, and 50 ms noise bursts. This was not observed in our results.

Loudness models that treat spectral loudness summation and temporal integration sequentially (i.e.: Glasberg and Moore, 2002; Chalupper and

Fastl, 2002) cannot account for time dependent spectral loudness summation, as these models predict the same loudness differences between narrow-band and broadband signals irrespective of the signal duration. The effect of bandwidth found in the series of noise bursts can also not be accounted for by such a model. For a correct prediction of time dependent spectral loudness summation either the compressive non-linearity, or the critical bandwidth has to vary with signal duration. Rennies *et al.* (2009; 2015) investigated the 'model of loudness applicable to time-varying sounds' by Glasberg and Moore (2002) and the 'Dynamic Loudness Model (DLM)' by Chalupper and Fastl (2002) and several modified versions of the DLM with respect to their ability to predict the time dependence of spectral loudness summation and other spectro-temporal effects. Both original models consist of an auditory filterbank and a compressive non-linearity that does not depend on the temporal properties of the signal. As expected, the original models predicted the same spectral loudness summation for signals with short and long duration at equal level. In order to be able to predict time dependent loudness summation Rennies *et al.* (2009) investigated three modifications to DLM in which the auditory filterbank and/or the compressive non-linearity were made time dependent.

One of the modifications to DLM suggested and evaluated by Rennies *et al.* (2009) was a model with adaptive compression, where larger compression was assumed at signal onset as a possible explanation for time dependent spectral loudness summation, as suggested earlier by Verhey and Kollmeier (2002). This hypothesis was also used to explain overshoot (von Klitzling and Kohlrausch, 1994; Strickland, 2001, 2004, 2008). Rennies *et al.* (2009) showed that a model with adaptive compression can indeed predict time-dependent spectral loudness summation. However, to account for the differences in spectral loudness summation between brief and long signals as measured by Verhey and Uhlemann (2008), unrealistically high levels of compression at signal onset are needed. A second problem with the adaptive compression model is that it violates the modified equal-loudness-ratio-hypothesis (ELRH) proposed by Anweiler *et al.* (2006). The original ELRH by Florentine *et al.* (1996) states that the loudness ratio between long and short signals with the same spectrum presented at an equal level is independent of level and spectrum. This has been shown to be inconsistent with experimental data on duration effects in spectral loudness summation

(Verhey *et al.*, 2002; Anweiler *et al.*, 2006; Verhey *et al.*, 2006), as the ELRH predicts an amount of loudness-summation which is independent of duration. The modified ELRH assumes a smaller loudness ratio for broadband signals than for narrow-band signals (see Anweiler *et al.,* 2006) and is consistent with their data measured for moderate levels. However, even the modified ELRH does not account for a loudness ratio that increases with increasing level, as predicted by a model with adaptive compression.

Rennies *et al.* (2009) also investigated a modified DLM with adaptive auditory filters (for short signals the widths of the auditory filters are decreased and thus the number of auditory filters is increased), and a bandwidth-dependent integration model. In this last modification the temporal integration window depends on the bandwidth of the stimulus, such that the loudness of a narrow-band signal is derived from integration over a longer duration than for a broadband signal. At short durations this assumption leads to incomplete temporal integration for narrow-band signals relative to broadband signals, which results in a larger loudness difference between the narrow-band and broadband signals and thus in more loudness summation for short signals than for long signals. A modified version of DLM, in which such a bandwidth-dependent temporal window is introduced, yields good predictions of the data on loudness summation from several studies (Rennies *et al.* 2009; Heeren *et al.* 2011).

Our results presented in chapter 2 confirm that spectral loudness summation depends on signal duration both in stationary sounds and in series of noise bursts. We agree with Rennies et al. (2009) that the bandwidth-dependent integration model seems to be the most promising modification of the current dynamic loudness model. However, more research is needed to determine the time constants involved in this model.

7.3 Partial loudness at signal onset

At threshold a higher signal-to-masker ratio is needed to detect a brief signal at the onset of a masker than when it is presented with an onset delay, or when it is preceded by another sound (precursor condition). This effect is called “overshoot”, or the temporal effect (Zwicker, 1965; Bacon and Smith, 1991; Strickland, 2001, 2004, 2008; Strickland and Krishnan 2005). The temporal effect has been investigated extensively and is believed to result from a decrease in cochlear gain during the course of a masker (e.g., Von Klitzing and Kohlrausch, 1994; Strickland, 2001, 2004, 2008; Strickland and Krishnan 2005; Bacon and Savel, 2004). In chapter 3 we investigated the temporal effect at threshold and at supra-threshold levels for simultaneous masking conditions and backward masking conditions. Our results show that a temporal effect is not restricted to threshold, but also occurs at supra-threshold levels. In chapter 3 we also show that near masker onset the partial loudness curves can be well fitted by model predictions for partial loudness (Glasberg and Moore, 2002) if a higher masker level is assumed to be present than is presented actually. This shows a strong resemblance with the bandwidth-dependent integration model proposed by Rennies *et al.* (2009).

The bandwidth-dependent integration version of DLM is implemented by Rennies *et al.* (2009) as an amplification at onset. The amount of amplification increases with increasing bandwidth and it is only applied for a short time after stimulus onset. The assumed amplification at masker onset results in an instantaneous loudness at masker onset which is higher than the steady state response. The authors state that such an accentuation of the stimulus onset is not uncommon in the auditory system. If detection of a short signal in a masker is assumed to depend on a constant signal-plus-masker to masker ratio, a higher instantaneous loudness at masker onset may lead to a higher signal level at masker onset to fulfill the constant "signal-plus-masker to masker" ratio relative to the signal level needed at steady state. This is consistent with the data on the overshoot effect. Grimm *et al.* (2002) also suggest an approach in which the onset of the signal receives more emphasis. These authors show that the contrasting model gives the best predictions for their data on loudness of fluctuating sounds.

Emphasis on the onset of a signal is also found in research on temporal weights. In experiments using sounds with fluctuating levels during the course of the signal, it is shown that not all temporal portions of a sound receive equal weights (Fischenich *et al.,* 2019, 2020; Oberfeld and Plank, 2011; Pedersen and Ellermeier, 2008). The beginning of the signal receives a larger weight than latter portions of the signal. This is called the primacy effect. The primacy effect and the temporal effect share several characteristics. For instance, the primacy effect has approximately the same time course as the temporal effect with a decay time of approximately 200-300 ms (Oberfeld *et al.*, 2018). Fischenich *et al.* (2019) further show that the primacy effect also occurs in background noise and is almost independent of loudness level.

Common explanations for the primacy and the temporal effect, suggest that the effects could be related to the firing of the auditory nerve, as the auditory nerve shows a peak at the onset of a sound and adapts to a steady state firing rate after a few milliseconds. The time courses of the primacy effect and the temporal effect are however much longer than a few milliseconds, suggesting that more central auditory processing plays a role also. For instance, for the temporal effect the medial olivocochlear reflex (MOCR) is assumed to turn down the gain in the cochlea in response to sounds (i.e. Roverud and Strickland, 2010).

A second explanation for the primacy effect is called attention orientation. Due to the abrupt onset of the noise, the attention is captured and directed to the beginning of the stimulus (Oberfeld and Plank, 2011). Attention orientation to the onset of the masker has been suggested by Scharf *et al.* (2008) and Bacon and Moore (1987), as an explanation for overshoot. However, if attention orientation plays a role, a reduction of the primacy effect is expected in background noise, as the onset is less abrupt in a background noise than in silence, but this reduction is not found. Therefore, the hyphothesis is that attention orientation to signal onset does not play a major role in the primacy effect. This does not automatically mean that attention orientation doesn’t play a role in the temporal effect either. On a higher level the attention orientation effect can cause confusion between the signal and the masker, making it harder to differentiate the signal from the masker. Higher levels of processing are also considered to play a role in another possible explanation for the primacy effect, where the hypothesis is that the primacy effect may stem from an “evidence integration” process (Fischenich *et al.*, 2019). This hypothesis suggests that evidence is collected sequentially during each trial and that the decision is made as soon as sufficient information has been accumulated ignoring further evidence during the trial. Following this reasoning a longer duration may have to be processed to collect enough evidence for the presence of the signal at masker onset than at the center of the masker, leading to a higher loudness perception of the masker. This hypothesis can explain the effect that a pre-masker signal is influenced by the masker. Near masker onset the uncertainty increases to detect another signal and the integration window of attention is increased around masker onset, encompassing a certain time window before masker onset.

The temporal effect has been investigated extensively for signals presented at threshold levels (i.e.: Zwicker, 1965; Bacon and Smith, 1991; Strickland, 2001, 2004, 2008; Strickland and Krishnan 2005), but not for supra-threshold levels. As shown in chapter 3 the temporal effect is stronger at supra-threshold levels than it is at threshold. The current results, however, are not accurate enough to determine the exact transition of the temporal effect from threshold to supra-threshold levels. More data on the time scales involved with the supra-threshold effects are needed to determine which of the hypotheses mentioned above may provide the best explanation of the supra-threshold temporal effect.

7.4 Spectral summation in hearing-impaired listeners

Spectral loudness summation is a major topic in the second part of this thesis. Until recently spectral loudness summation in hearing-impaired listeners was assumed to be equal to or less than in normal hearing listeners (Bonding and Elberling, 1980; Brand and Hohmann, 2001; Verhey *et al.*, 2006; Moore and Glasberg, 2004). In chapter 5 we showed that for broadband sounds spectral loudness summation can be higher in hearing-impaired listeners than in normal hearing listeners. Especially for larger hearing losses, corresponding to audiogram-classifications N3, N4, and S according to Bisgaard *et al.* (2010), the amount of spectral loudness summation can be substantial. This is found both for unilateral and bilateral presentations. The cause of higher spectral loudness summation in hearing-impaired listeners is still unclear. As stated earlier, summation is a result from the interaction between the cochlear non-linearity and the auditory filterbank. To achieve an increase in spectral loudness summation there are theoretically two paths: a narrowing of the auditory filters, which will cause a summation across more filters, and/or a change of the cochlear non-linearity which produces a higher loudness per auditory filter. Auditory filters broaden with hearing impairment (e.g., Bernstein and Oxenham, 2006; Carney and Nelson, 1983; Glasberg and Moore, 1986; Lutman, Gatehouse and Worthington, 1991; Oxenham and Bacon, 2003). Therefore, it is unlikely that the increased spectral loudness summation in hearing-impaired listeners is caused by summing loudnesses over an increased number of auditory filters.

It is more likely that the loudness per auditory filter is increased, as is the case in time-dependent-loudness summation. It is well known that hearing impairment changes loudness growth (Moore, 1998; Buus and Florentine, 2002). The loudness in the auditory filters can thus be different for hearing-impaired listeners and normal hearing listeners. In current loudness models hearing impairment is assumed to have a local impact on loudness growth. For instance, the model by Moore and Glasberg (2004) calculates an elevation in absolute threshold, a reduction in compressive non-linearity and a loss of frequency selectivity for each auditory filter separately. Thus, loudness growth in each auditory filter is independent of the loudness growth in the other auditory filters. With this design the loudness model by Moore and Glasberg (2004) predicts decreased spectral loudness summation for hearing-impaired listeners in line with the results by Bonding and Elberling (1980).

However, the results in chapter 5 and 6 are not in line with this model. These results suggest that for broadband signals loudness growth may be influenced by a central change in gain, encompassing several auditory filters simultaneously, as hearing loss in the high frequencies also appears to influence loudness perception at low frequencies. This is in line with the hypothesis for time-dependent spectral loudness summation. A trade-off between spectral and temporal properties such as suggested in the bandwidth-dependent integration model can be explained by a central processing stage.

Pieper *et al.* (2018) investigated several loudness models that simulated individual hearing thresholds by a cochlear gain reduction and linear attenuation prior to an internal threshold. They showed that such an approach was insufficient to account for individual loudness perception, in particular at high stimulus levels near uncomfortable loudness. To improve their loudness model based on a transmission-line model, they added a frequency dependent post gain. This post gain could be interpreted as a central gain occurring at higher stages as a result of peripheral deafferentation. The post gain improved the predictions of the individual variations in the steepness of the loudness function and the variation in the uncomfortable loudness level independently of the hearing loss. In a follow-up study Pieper *et al.* (2021) showed that it was crucial to include bandwidth-dependent weightings of the signals in the monaural path to improve the prediction of monaural spectral loudness summation. The modeling efforts of Pieper *et al.* (2018, 2021) thus confirm that central bandwidth-dependent processing stages have to be included in loudness models to allow the prediction of individual loudness perception.

In the experiments, described in chapters 5 and 6, most hearing losses were flat or mildly sloping. To investigate the influence of a central gain component, spectral loudness summation of more steeply sloping audiograms with different cut-off frequencies for the start of the slope could be of interest. Another approach could be to measure spectral loudness summation in subjects with asymmetric hearing losses and compare the results between ears.

7.5 Binaural loudness summation

Central processing is certainly involved in binaural hearing. When we hear a sound with two ears instead of one, that sound is normally perceived louder. This phenomenon is called binaural loudness summation. The first loudness models (Zwicker, 1958; Zwicker and Scharf, 1965; Moore and Glasberg, 1996) were developed to predict the loudness of sounds presented bilaterally. However, psychophysical experiments on loudness perception are often conducted unilaterally. Therefore, it is important to model the differences between loudness of unilaterally and bilaterally presented sounds. For diotic presentation (the same sound at each ear), the overall loudness is normally assumed to be double that for each ear separately (Hellman and Zwislocki, 1963) corresponding to a LDEL of approximately 10 dB SPL, or slightly less than double as a result of binaural inhibition (Moore *et al.*, 2014), corresponding to a LDEL of 5-6 dB SPL.

The results in chapter 5 and 6 clearly show that the relationship between unilaterally and bilaterally presented signals is not that straightforward for hearing-impaired listeners. Binaural loudness summation depends on bandwidth, amount of hearing loss, and frequency content. For the relatively narrow-band signals UEN1 and UEN5 the amount of binaural loudness summation is comparable to the amount of binaural loudness summation found in literature for all classes of hearing loss. However, chapter 5 shows that binaural loudness summation is increased for broadband signals (UEN17 and IFNoise) and this increase is substantial. For audiometric configurations "N3" and "S" the LDEL can be more than twice as large (in decibels) as for the narrow-band signals. This implies severe consequences for the applicability of prescription rules for hearing aid amplification that will be discussed in paragraph 7.7.

Chapter 6 shows further that binaural loudness summation is larger for LP filtered pink noises, than for HP filtered pink noises. As if in spectral loudness summation a central processing stage seems to be active that determines the gain in all auditory filters and supplies more gain in case hearing loss is present in one or several auditory filters. This is again in line with modeling efforts by Pieper *et al.* (2021). They showed that binaural loudness summation could be accounted for by a single bandwidth-independent parameter. As in the case of spectral loudness summation (see paragraph 7.4), in case of binaural loudness summation we expect that testing subjects with asymmetric audiograms could also provide further insights into a possible role of a central processing stage.

7.6 Individual variability

Loudness models generally assume that it is possible to define an average normal loudness function. However, even in a population of normal hearing listeners loudness growth curves can differ quite substantially. This is clearly shown in ISO 16832 (2006) where in a normal hearing population the intensity levels at a particular loudness level may differ as much as 20-30 dB. The results for a group of normal hearing musicians presented in chapter 4 confirm this range of individual differences. Formby *et al.* (2017) found an even larger range of levels at equal loudness in a group of normal hearing listeners. They found loudness judgments within the same loudness category to differ by as much as 50-60 dB.

Individual variability in hearing-impaired listeners is at least as large as in normal hearing listeners (i.e., see Bentler and Cooley, 2001). This thesis shows that individual variability in hearing-impaired listeners is even larger than normal, see chapters 5 and 6. Furthermore, chapter 5 shows that normalizing loudness based on narrow-band loudness measurements at six frequencies, does not fully decrease the variability in loudness growth curves for broadband signals. For bilaterally presented broadband signals the variability even increases, especially at the upper part of the loudness scale. The level at which a sound is judged too loud (corresponding to a loudness level of 50 CU) differed by as much as 40 dB between subjects.

As shown in chapter 4 the correlation between the intensity level near absolute threshold (CU5) and the intensity level of uncomfortable loudness (CU50) was not significant. This means that the threshold of uncomfortable loudness is not directly related to absolute threshold. The dynamic range and therefore also the slopes of the loudness curves differ for each normal hearing listener. This is also found in hearing-impaired listeners. Marozeau and Florentine (2007) summarize research on loudness growth in hearing-impaired listeners and they show large individual differences in loudness growths with increasing sound pressure level for hearing-impaired listeners.

For sensorineural hearing losses Marozeau and Florentine (2007) describe two types of loudness growth functions: rapid-growth types and softness imperceptions types (Moore, 2004). For the rapid-growth type, loudness at threshold is the same for normal hearing and hearing-impaired listeners but loudness grows more rapidly from threshold to mid-levels for hearing-impaired listeners than for normal hearing listeners. For the softness imperception type, the loudness grows similar to normal near threshold and faster-than-normal at mid-levels. However, the absolute loudness at threshold is in these cases larger than normal. Individual variability in loudness perception is also common in studies investigating loudness differences between signals with different signal properties, such as duration (i.e.: Florentine *et al.*, 1996), bandwidth (i.e.: Verhey and Kollmeier, 2002), and different listening conditions, such as forward masking (i.e.: Oberfeld, 2007). This individual variability is also seen in chapters 2 and 3 of this thesis. Not all subjects showed an equal amount of spectral loudness summation or temporal effect.

There are several causes for individual variability in loudness measurements. First, the variability may arise from a genuine difference in loudness perception caused by differences in peripheral or central hearing ability. Temporal bone studies of individuals with hearing loss have identified several abnormalities in the cochlea associated with hearing loss, including the amount of inner and outer hair cell loss, spiral ganglion cell loss, stria vascularis atrophy, and stiffening of the basilar membrane (Gacek and Shuknecht, 1969; Suga, and Lindsay, 1976; Keithley, 2020). Several studies have attempted to relate these abnormalities to specific audiometric patterns of hearing loss (for review see: Nelson and Hinojosa, 2006) and these studies suggest that different audiometric patterns are associated with different underlying pathologies. If different cochlear abnormalities cause different audiometric threshold patterns, it is likely that they also can cause different patterns of loudness perception. Although the hearing losses in our studies were sorted according to audiometric threshold patterns, we cannot guarantee that all subjects in the same hearing loss category had the same cochlear abnormalities. Further on in the auditory pathway, fMRI activation in several regions within the auditory cortex as well as in certain stages of the ascending auditory pathway appears to be a direct linear reflection of loudness (i.e.: Behler and Uppenkamp, 2016). As hearing loss has been shown to alter central auditory processes (Eggermont, 2017), it is likely that differences in central auditory processing can also lead to differences in loudness perception.

In general, loudness perception can be influenced by non-auditory factors. For instance, Parker *et al.* (2012) showed that the expectation of a listener about the loudness of a sound influences the perceived loudness. Epstein and Florentine (2012) showed that the amount of binaural loudness summation was significantly less for speech presented via a loudspeaker with visual cues than for speech presented via earphones with visual cues, or speech presented via loudspeakers or earphones without visual cues. Loudness perception is also influenced by emotion. On a short time scale, Asutay and Västfjäll (2012) showed that the same auditory stimulus was reported as being louder, when it was conditioned with an aversive experience, compared to when it was used as a control stimulus. On a larger time scale, Wallén *et al.* (2012) showed that subjects with higher scores on emotional exhaustion scored higher on the hyperacusis questionnaire and showed lower uncomfortable loudness levels. Loudness perception can also be altered by appropriate training. In a population of persons with decreased dynamic range Formby *et al.* (2015, 2017) showed that a sound-therapy based intervention could expand the auditory dynamic range for listeners with sensorineural hearing loss.

Finally, it is important to realize that loudness perception is influenced by the measurement procedure. This effect is well recognized in the psychophysical test procedures, where randomization and interleaving are used to minimize order effects. Still caution must be taken when comparing loudness results obtained with different measurement procedures. As seen on several occasions in this thesis, results from different measurement procedures correspond qualitatively, but not quantitatively. This is observed in chapter 2 where the amount of spectral loudness summation clearly depends on the matching procedure, but also in chapter 6, where the amount of spectral and binaural summation is different for the matching procedure and the scaling procedure. The measurement procedure may therefore influence the amount of gain needed to normalize loudness perception.

7.7 Research questions and clinical applications

In the previous paragraphs several aspects of loudness perception have been discussed in which the experimental findings do not always match with currently established hypotheses and models. In this paragraph the answers to our research questions are summarized and the implications of the deviations between the theoretical models and the experimental results for the clinical application of loudness models are discussed.
The first part of this thesis addresses two research questions:

A) How does bandwidth influence loudness of short stimuli and series of noise bursts?

and

B) Is the temporal effect also apparent at supra-threshold levels?

The results in chapter 2 are in agreement with the results of Verhey *et al.* (2006): spectral loudness summation is larger for short stimuli than for long stimuli. Furthermore, loudness summation is larger in series of noise bursts than in a continuous signal. In chapter 3 is shown that the temporal effect is also apparent at supra-threshold levels.

Which consequences can these results have for clinical application?
Time dependent spectral loudness summation has been shown to be absent in most hearing-impaired listeners (Verhey *et al.*, 2006). The same holds true for overshoot at threshold (Bacon *et al.*, 1988; Kimberley and Nelson, 1989; Bacon and Takahashi, 1992; Strickland and Krishnan, 2005). And although the temporal effect above threshold has not been investigated for hearing-impaired listeners, it is likely that the temporal effect above threshold is reduced in hearing-impaired listeners. Both time dependent loudness summation and the temporal effect emphasize the onset of a sound. Obviously, this effect is decreased in hearing-impaired listeners. To restore the onset effect to normal, fast compression schemes should provide a stronger compression for transients than for stationary sounds (Verhey *et al.*, 2006). This approach is not in agreement with the commonly applied compression schemes in hearing aid fitting. The lack of a temporal effect may also contribute to some of the problems with hearing in noisy environments, as these problems seem to be related to a loss of temporal resolution (i.e. Feng *et al.*, 2010, Vermeire *et al.* 2016). As the temporal effect is generally not taken into account in hearing aid fitting, one would expect that the loudness for short and dynamic sounds is underestimated for hearing-impaired listeners, as the absence of the temporal effect decreases the loudness for these sounds relative to normal hearing listeners. This effect is not found in clinical practice thus far, as shown by Smeds *et al.* (2004). They showed that hearing-impaired listeners generally prefer less gain than prescribed from loudness normalization or loudness equalization fitting procedures when listening to everyday sounds. A possible reason for this discrepancy may be that time dependent spectral loudness summation and the temporal effect are commonly investigated with monaural stimulus

presentation. As shown in the second part of this thesis, bilateral presentation of sounds can have a large impact (up to 20 dB) on loudness perception for individual hearing-impaired listeners. The larger binaural summation effect may therefore mask the more subtle temporal effect.

In Chapter 4 the main research question was:

C) What can we learn from measuring the entire loudness function in a clinical population?

The results in chapter 4 show that even in a population of normal hearing subjects loudness functions can vary considerably and that auditory threshold is not always a good predictor for the loudness function at high levels. More importantly, the results suggest that binaural loudness summation mainly influences the upper part of the loudness function, making it quite unlikely that loudness perception for bilaterally presented sounds can be correctly predicted from threshold measurements.

D) Do we need individual loudness measurements for loudness normalization?

The results found in chapter 4 are closely related to the results presented in chapters 5 and 6. As the results in chapter 4 have shown, it is not straightforward to define normal loudness. The goal to restore loudness to normal as done in chapter 5 and in many fitting rules for hearing aids should therefore be applied with some caution and should be revisited. Normal loudness can vary considerably between subjects. The results of Smeds *et al.* (2006a, b) raise the question if loudness normalization is a good criterion for hearing aid fitting. The same observation can be made in chapter 5. After loudness normalization based on individual loudness curves measured at six frequencies, the loudness curves are nicely distributed around the average normal loudness curve for unilaterally presented noises (both narrow-band and broadband) and bilaterally presented narrow-band noises. However, for the hearing-impaired listeners almost all loudness curves for bilaterally presented broadband noises are shifted towards lower intensity levels. This means that the hearing-impaired listeners judge broadband sounds of the same intensity louder than the average normal hearing listener when the signals are presented bilaterally. The hearing-impaired listeners would therefore need less gain for these bilateral stimuli than calculated on basis of the unilaterally presented narrow-band loudness measurements. An interesting point is that the studies by Smeds *et al.* (2006a, b) and by van Beurden *et al.* (2018), both deal with broadband signals presented bilaterally to the hearing-impaired listeners. In most studies on binaural loudness perception, that find normal or smaller-than-normal binaural loudness summation, only narrow-band, or limited broadband signals have been investigated (Whilby *et al.,* 2006; Moore *et al.*, 2014). Hawkins *et al.* (1987) did not measure a difference in binaural loudness summation between normal hearing and hearing-impaired subjects for a speech spectrum noise, but they didn't define the exact bandwidth of the speech spectrum noise.

E) How does hearing loss configuration influence spectral and binaural loudness perception?

Chapter 5 shows that for larger hearing losses, corresponding to audiogram-classifications N3, N4, and S according to Bisgaard *et al.* (2010), the amount of spectral loudness summation can be considerably larger than for normal hearing. Binaural loudness summation is larger for audiogram classifications N3, and S.

F) How does frequency content influence spectral and binaural loudness perception?

As the results in chapter 6 show, the low frequency content of the broadband signal delivers the main contribution to the loudness of the broadband signal, especially for bilateral presentation. At this point a good understanding about the mechanisms behind binaural loudness summation in hearing-impaired listeners is still lacking. As discussed above central processes leading to a central gain compensation may account for part of the effects, but more research is needed to determine the exact mechanism. It is questionable if it will be possible to predict spectral and binaural loudness summation based on unilateral threshold measurements as is now commonly done in hearing aid fitting. It is more likely that individual broadband and binaural loudness measurements will be needed to obtain a more accurate estimation of an appropriate loudness compensation in the individual case.

These conclusions are in line with the conclusions drawn by Pieper *et al.* (2021). In a series of papers Pieper *et al.* (2016, 2018, 2021) developed an individual binaural loudness model for hearing aid fitting. For listeners with a slight-to-moderate sensorineural hearing loss this model was able to successfully account for spectral and binaural loudness summation. The model was based on loudness scaling measurements with narrowband and broadband stimuli similar to the stimuli used in the experiments by Oetting *et al.* (2016) and van Beurden *et al.* (2018). To achieve an accurate prediction of individual loudness perception for broadband and binaural signals Pieper *et al.* (2018, 2021) needed a bandwidth-dependent monaural loudness parameter to account for individual spectral loudness summation and they needed a bandwidth-independent parameter to account for binaural loudness summation.

Oetting *et al.* (2018) calculated individual signal-dependent gain corrections for four signals with different bandwidths and for both unilateral and bilateral presentation based on loudness scaling measurements after narrowband loudness compensation. They showed that these gain corrections resulted in normal loudness ratings for real-world test signals. They also showed large inter-individual differences in gain corrections. This finding again underlines that threshold based fitting strategies and even narrowband loudness compensation procedures do not ensure normal loudness perception for binaural broadband stimuli, when based on average relations. Individual binaural broadband corrections are needed to obtain the individual parameters characterizing appropriate loudness perception for hearing-impaired listeners.

For individual loudness corrections as suggested by Oetting *et al.* (2018) and Pieper *et al.* (2021) many loudness measurements are needed. Ideally, frequency specific narrowband loudness assessment for both ears is combined with measurements of spectral loudness summation and binaural loudness summation. As loudness scaling takes a few minutes per stimulus, loudness scaling is not well suited for the measurement of a large set of stimuli in a clinical setting. For clinical use a more time efficient measurement procedure is needed.

With loudness matching twice as many conditions can be measured as with loudness scaling in the same amount of time. Therefore, loudness matching might be applied to increase the amount of individual loudness measurements without exceeding an acceptable measurement time. Loudness matching of course does not provide a full loudness growth curve. However, it may not be necessary to measure loudness growth curves for each stimulus. For hearing-impaired listeners the narrowband loudness compensation applied in chapter 5 generally leads to normal loudness growth functions at low to medium intensity levels. The deviations with regard to the loudness growth functions for normal hearing listeners occur at medium to high levels. It therefore seems to be possible to estimate the loudness growth curve as soon as the deviation from the normal loudness growth function at high levels is known. Further research should provide more insight in an optimal set of stimuli to be used in clinical practice.

At the end, two important considerations need attention. First, most research on loudness perception has been performed with stimuli in quiet. Normal listening environments, however, often include some kind of background noise. Studies on partial loudness are sparse, and are mainly conducted on normal hearing subjects and with tonal or narrowband signals (Zwicker, 1963; Stevens and Guirao, 1967; Houtgast, 1974; Langhans and Kohlrausch, 1992; Spiegel *et al.*, 1981). Studies on loudness perception in hearing-impaired listeners, and with broadband stimuli are still lacking. Therefore, it is still to be seen, if loudness normalization in quiet ensures normal loudness perception in background noise in hearing-impaired listeners.

A second point to keep in mind is that the considerations given above apply uniquely to loudness perception. For a good hearing aid fitting (close to) normal loudness perception is not the only objective. Another, and perhaps even more important, objective is to optimize speech perception in quiet and in noise. Although it seems logical that normal loudness perception will lead to optimal speech perception, this is not guaranteed. It is therefore also important to investigate what the impact of normalizing loudness perception as described above is on speech perception both in quiet and in noise.

7.8 Final Conclusions

In this thesis we have shown that there are still several aspects of loudness perception that are not well understood. In the first part of the thesis we show that there are two temporal effects for short signals that are not yet covered in the current loudness models. Leading to the following results:

1) Spectral loudness summation depends on the duration of the signal. This duration dependency is shown to have a small effect in series of noise bursts.
2) The temporal effect has been shown to be also present at supra-threshold levels.
3) The temporal effect is even larger at supra-threshold levels than at threshold and is not confined to simultaneous masking, but can also be found shortly before onset of the masker.

In the second part of this thesis we have shown that spectral and binaural loudness summation can be much larger in hearing-impaired listeners than previously expected. Leading to the following results:

4) Larger hearing losses and more sloping hearing losses classified according to Bisgaard et al. (2010) lead to larger than normal spectral loudness summation both in unilateral as in bilateral presentation.
5) In hearing-impaired listeners binaural loudness summation is larger for the low frequency part of a broadband sound than for the high frequency part.
6) The variability in loudness perception in both normal hearing and hearing-impaired listeners is too large to ignore in loudness normalization. Individual loudness measurements are necessary since they may be expected to improve hearing aid fitting.

References & List of Abbreviations

References

Abrams, H.B., Chisolm, T.H., McManus, M., McArdle, R. (2012). Initial-fit approach versus verified prescription: comparing self-perceived hearing aid benefit. J. Am. Ac. Audiol. 23, 768-778. Doi: 10.3766/jaaa.23.10.3.

Algom, D., Rubin, A., and Cohen-Raz, L. (1989). Binaural and temporal integration of the loudness of tones and noises. Percept. Psychophys. 46, 155–166. Doi: 10.3758/bf03204975.

Almishaal, A., Bidelman, G.M., Jennings, S.G. (2017). Notched-noise precursors improve detection of low-frequency amplitude modulation. J. Acoust. Soc. Am. 141(1), 324-333. Doi: 10.1121/1.4973912.

Al-Salim, S.C., Kopun, J.G., Neely, S.T., Jesteadt, W., Stiegemann, B., Gorga, M.P. (2010). Reliability of categorical loudness scaling ad its relation to threshold. Ear. Hear. 31(4), 567-578. Doi: 10.1097/AUD.0.b013e3181da4d15.

ANSI S3.4. (2007). Procedure for the computation of loudness of steady sounds. Washington, DC: American National Standards Institute.

Anweiler, A. K. and Verhey, J. L. (2006). Spectral loudness summation for short and long signals as a function of level. J. Acoust. Soc. Am. 119, 2919-2928. Doi: 10.1121/1.2184224.

Appell, J.-E., Hohmann, V. (1998). Messung der Lautheitssummation bei Normal-und Schwerhörenden. In: DAGA - Fortschritte Der Akustik. Zürich, Switzerland.

Asutay, E., Västfjäll, D. (2012). Perception of loudness is influenced by emotion. PLoS One. 7(6): e38660. Doi: 10.1371/journal.pone.0038660.

Auerbach, B.D., Rodrigues, P.V., and Salvi, R.J. (2014). Central gain control in tinnitus and hyperacusis. Front. Neurol. 24(5), 1-21. Doi: 10.3389/fneur.2014.00206.

Bacon, S. P., and Smith, M. A. (1991). Spectral, intensive, and temporal factors influencing overshoot. Q. J. Exp. Psychol. A 43, 373–399. Doi: 10.1080/14640749108400

Bacon, S.P., and Savel, S. (2004). Temporal effects in simultaneous masking with on- and off-frequency noise maskers: Effects of signal frequency and masker level. J. Acoust. Soc. Am. 111(4), 1674-1682. Doi: 10.1121/1.1689344.

Bacon, S.P., Takahashi, G.A. (1992). Overshoot in normal-hearing and hearing-impaired subjects. J. Acoust. Soc. Am. 91, 2865-2871. Doi: 10.1121/1.402967.

Bacon, S.P., Moore, B.C.J., (1987). Transient masking and the temporal course of simultaneous tone-on-tone masking. J. Acoust. Soc. Am. 81, 1073-1077. Doi: 10.1121/1.395125.

Bacon S.P., Hedrick, M.S., and Grantham, D.W. (1988). Temporal effects in simultaneous pure-tone masking in subjects with high-frequency sensorineural hearing loss. Audiology 27, 313-323. Doi: 10.3109/00206098809081602.

Bacon, S.P. (1990). Effect of masker level on overshoot. J. Acoust. Soc. Am. 88, 698-702. Doi: 10.1121/1.399773.

Bagatto, M., Moodie, S., Scollie, S., Seewald, R., Moodie, S., Pumford, J., liu, K.P. (2005). Clinical protocols for hearig instrument fitting in the desired sensation level method. Trends Amplif. 9(4), 199-226. Doi: 10.1177/108471380500900404.

Beattie, R.C., Huynh, R.C., Ngo, V.N., Jones, R.L. (1997). IHAFF loudness contour test: reliability and effects of approach mode in normal-hearing subjects. J. Am. Acad. Audiol. 8(4), 243-56.
Behler, O., Uppenkamp, S. (2016). The representation of level and loudness in the central auditory system for unilateral stimulation. Neuroimage. 139, 176-188. Doi: 10.1016/j.neuroimage.2016.06.025.
Bentler, R.A., Cooley, L.J. (2001). An examination of several characteristics that affect the prediction of OSPL90 in hearing aids. Ear Hear. 22(1), 58-64. Doi: 10.1097/00003446-200102000-00006.
Berger, K.W., Hagberg, E.N., Rane, R.L. (1980) A reexamination of the one-half gain rule. Ear Hear. 1(4), 223-5. Doi: 10.1097/00003446-198007000-00009.
Bernstein J. G. W., Oxenham A. J. (2006). The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss. J. Acoust. Soc. Am. 120, 3929-3945. Doi: 10.1121/1.2372452.
Bisgaard, N., Vlaming, M.S., Dahlquist, M. (2010). Standard audiograms for the IEC 60118-15 measurement procedure. Trends Amplif. 14(2), 113-120. Doi: 10.1177/1084713810379609.
Bonding, P., Elberling, C. (1980). Loudness summation across frequency under masking and in sensorineural hearing loss. Audiology 19, 57-74. Doi: 10.3109/00206098009072649.
Boone, M.M. (1973). Loudness measurements on pure tone and broad band impulsive sounds. Acust. Acta Acust. 29(4), 198-204.
Boymans, M., Goverts, S.T., Kramer, S.E., Festen, J.M., Dreschler, W.A. (2008). A prospective Multi-centre study of the benefits of bilateral hearing aids. Ear Hear. 29(6), 930-941. Doi: 10.1097/AUD.0b013e31818713a8.
Boymans, M., Goverts, S.T., Kramer, S.E., Festen, J.M., Dreschler, W.A. (2009). Candidacy for bilateral hearing aids: a retrospective multicenter study. J. Speech Lang. Hear. Res. 52(1), 130-140. Doi: 10.1044/1092-4388(2008/07-0120).
Brand, T., Hohmann, V. (2001). Effect of hearing loss, centre frequency, and bandwidth on the shape of loudness functions in categorical loudness scaling. Int. J. Audiol. 40, 92-103. Doi: 10.3109/00206090109073104.
Brand, T., Hohmann, V. (2002). An adaptive procedure for categorical loudness scaling. J.Acoust.Soc.Am. 112, 1597-1604. Doi: 10.1121/1.1502902.
Brand, T. (2007). Loudness Scaling. 8th EFAS congress/10th Congress of the German Society of Audiology, 1-7, Heidelberg, Germany, 6-9 June.
Buus, S., Florentine, M., and Poulsen, T. (1997). Temporal integration of loudness, loudness discrimination and the form of the loudness function. J. Acoust. Soc. Am. 101(2), 669-680. Doi: 10.1121/1.417959.
Buus, S., and Müsch, H. (1998). On loudness at threshold. J. Acoust. Soc. Am. 104, 399-410. Doi: 10.1121/1.423295.
Buus, S., Musch, H., and Florentine, M. (1998). On loudness at threshold. J. Acoust. Soc. Am. 104(1), 399-410. Doi: 10.1121/1.423295.
Buus. S., and Florentine, M. (2002). Growth of loudness in listeners with cochlear hearing losses: Recruitment reconsidered. J. Assoc. Res. Otolaryngol. 3, 120-139. Doi: 10.1007/s101620010084.
Byrne, D., Dillon, H. (1986). The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. Ear Hear. 7(4), 257-65. Doi: 10.1097/00003446-198608000-00007.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., et al (1994). An international comparison of long-term average speech spectra. J.Acoust.Soc.Am. 96, 2108-2120. Doi: 10.1121/1.410152.

Byrne, D., Dillon, H., Ching, T., Katsch, R., Keidser, G. (2001). NAL-NL1 procedure for fitting nonlinear hearing aids: characteristics and comparisons with other procedures. J. Am. Acad. Audiol. 12, 37-51. Doi:10.1055/s-0041-1741117.

Cacace, A.T., Margolis, R.H. (1985). On the loudness of complex stimuli and its relationship to cochlear excitation. J Acoust Soc Am. 78(5), 1568-73. Doi: 10.1121/1.392793.

Carlyon, R. P. and Datta, A. J. (1997). Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system. J. Acoust. Soc. Am. 101, 3636-3647. Doi: 10.1121/1.418324.

Carlyon, R.P. (1987). A release from masking by continuous, random, notched noise. J. Acoust. Soc. Am. 81, 418-426. Doi: 10.1121/1.395117.

Carlyon, R.P., and Moore, B.C. (1986). Detection of tones in noise and the "severe departure" from Weber's law. J. Acoust, Soc. Am. 79(2), 461-464. Doi: 10.1121/1.393533.

Carney A. E., Nelson D. A. (1983). An analysis of psychophysical tuning curves in normal and pathological ears. J. Acoust, Soc. Am. 73, 268-278. Doi: 10.1121/1.388860

Carter, N. L. (1972). Effects of rise time and repetition rate on the loudness of acoustic transients. J.Sound Vib. 2, 227-239. Doi: 10.1016/0022-460X(72)90909-1.

Chalupper, J. (2002). Perzeptive Folgen von Innenohrschwerhörigkeit: Modellierung, Simulation und Rehabilitation (Perceptive consequences of inner ear hearing impairment: modelling, simulation and rehabilitation), Shaker Verlag, Aachen.

Chalupper, J. and Fastl, H. (2002). Dynamic loudness model (DLM) for normal and hearing-impaired listeners. Acust. Acta Acust. 88, 378-386.

Champlin, C.A., and McFadden, D. (1989). Reductions in overshoot following intense sound exposure. J. Acoust. Soc. Am. 85, 2005-2011. Doi: 10.1121/1.397853.

Chen, G.D., Radziwon, K.E., Kashanian, N., Manohar, S., Salvi, R. (2014) Salicylate-induced auditory perceptual disorders and plastic changes in nonclassical auditory centerd in rats. Neural. Plast. 2014;658741.Doi: 10.1016/j.heares.2011.08.001.

Cooper, N.P., and Guinan, J.J.Jr. (2006). Efferent-mediated control of basilar membrane motion. J. Physiol. 576, 49-54. Doi: 10.1113/jphysiol.2006.114991.

Cornelisse, L.E., Seewald, R.C., Jamieson, D.G. (1995). The input/output formula: A theoretical approach to the fitting of personal amplification devices. J. Acoust. Soc. Am. 97(3), 1854-1864. Doi: 10.1121/1.412980.

Cox, R.M. (1995). Using Loudness Data for Hearing Aid Selection: The IHAFF Approach. Hear.J. 48, 10, 39-44. Doi: 10.1097/00025572-199502000-00001.

Cox, R.M., Alexander, G.C., Taylor, I.M., Gray, G.A. (1997). The contour test of loudness perception. Ear. Hear. 18(5), 388-400. Doi: 10.1097/00003446-199710000-00004.

Cox, R.M., Schwartz, K.S., Noe, C.M., Alexander, G.C. (2011). Preference for one or two hearing aids among adult patients. Ear. Hear. 32(2), 181-197. Doi: 10.1097/AUD.0b013e3181f8bf6c.

Darrow, K.N., Maison, S.F., Liberman, M.C. (2007). Selective removal of lateral olivocochlear efferents increases vulnerability to acute acoustic injury. J Neurophysiol. 97(2), 1775-85. Doi: 10.1152/jn.00955.2006.
Dermody, P., Byrne, D. (1975). Loudness summation with binaural hearing aids. Scand. Audiol. 4, 23-28. Doi: 10.3109/01050397509075010.
Dillon, H. (2012). Hearing Aids Second Edition. Sydney, Australia: Boomerang Press. 226.
DIN 45631 (1991). Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum-Verfahren nach E. Zwicker (Procedure for calculating loudness levels and loudness). Deutches Institut für Normung, Berlin.
Dolan, T.G., and Small, Jr., A.M. (1984). Frequency effects in backward masking. J. Acoust. Soc. Am. 75, 932-936. Doi: 10.1121/1.390540 75.
Eggermont, J.J. (2017). Acquired hearing loss and brain plasticity. Hear. Res. 343, 176-190. Doi: 10.1016/j.heares.2016.05.008.
Elberling, C. (1999). Loudness scaling revisited. J. Am. Acad. Audiol. 10(5), 248-60. Doi:10.1055/s-0042-1748496.
Elmasian, R. and Galambos, R. (1975). Loudness enhancement: monaural, binaural, and dichotic. J. Acoust. Soc. Am. 58, 229-234.Doi: 10.1121/1.380650.
Elmasian, R. and Galambos, R. (1980). Loudness enhancement and decrement in four paradigms. J. Acoust. Soc. Am. 67(2), 601-607. Doi:10.1121/1.383937.
Epstein, M., Florentine, M. (2012). Binaural loudness summation for speech presented via earphones and loudspeaker with and without visual cues. J. Acoust. Soc. Am. 131(5), 3981-8. Doi: 10.1121/1.3701984.
Erber, N.P., Witt, L.H. (1977). Effects of stimulus intensity on speech perception by deaf children. J. Speech. Hear. Dis. 42, 271–278. Doi: 10.1044/jshd.4202.271.
EuroTrak Germany 2018. Trends EuroTrak 2009-2015 & MT9 (USA). Accessed 9 September 2019. www.ehima.com.
Ewert, S.D. (2013). AFC - a modular framework for running psychoacoustic experiments and computational perception models. 39 Tagung der Deutschen Arbeitsgemeinschaft für Akustik (DAGA).
Ewert S.D., Oetting, D. (2018). Loudness summation of equal loud narrowband signals in normal-hearing and hearing-impaired listeners. Int. J. Audiol. 57(sup3), S71-S80. Doi: 10.1080/14992027.1380848.
Fastl, H., Zwicker, E. (2007). In: Psychoacoustics: Facts and Models, third ed. Springer, Berlin.
Feng, Y., Yin, S., Kiefte, M., Wang, J. (2010). Temporal resolution in regions of normal hearing and speech perception in noise for adults with sloping high-frequency hearing loss. Ear Hear. 31(1), 115-25. Doi: 10.1097/AUD.0b013e3181bb69be.
Fischenich, A., Hots, J., Verhey, J., Oberfeld, D. (2019). Temporal weights in loudness: Investigation of the effects of background noise and sound level. PLoS One. 14(11), e0223075. Doi: 10.1371/journal.pone.0223075.
Fischenich, A., Hots, J., Verhey, J.L., Oberfeld, D. (2020). The effect of silent gaps on temporal weights in loudness judgments. Hear Res. 395, 108028. Doi: 10.1016/j.heares.2020.108028.
Fletcher, H., Munson, W.A. (1933) Loudness, its definition, measurement and calculation. J. Acoust. Soc. Am. 5, 82-108. Doi.org/10.1121/1.1915637

Florentine, M., Zwicker, E. (1979). A model of loudness summation applied to noise-induced hearing loss. Hear. Res. 1(2), 121-132. Doi: 10.1016/0378-5955(79)90023-6.

Florentine, M., Buus, S., Poulsen, T. (1996). Temporal integration of loudness as a function of level. J.Acoust. Soc. Am. 99(3), 1633-1644. Doi: 10.1121/1.415236.

Florentine, M., Buus, S., and Robinson, M. (1998). Temporal integration of loudness under partial masking. J. Acoust. Soc. Am. 104(2 Pt 1), 999-1007. Doi: 10.1121/1.423314.

Florentine, M. (2003). It's not recruitment—gasp! It's softness imperceptions. Hear. J. 56(3), 10-15 Doi: 10.1097/01.HJ.0000293012.17887.b4.

Formby, C., Hawley, M.L., Sherlock, L.P., Gold, S., Payne, J., Brooks, R., Parton, J.M., Juneau, R., Desporte, E.J., Siegle, G.R. (2015). A Sound Therapy-Based Intervention to Expand the Auditory Dynamic Range for Loudness among Persons with Sensorineural Hearing Losses: A Randomized Placebo-Controlled Clinical Trial. Semin. Hear. 36(2), 77-110. Doi: 10.1055/s-0035-1546958.

Formby, C., Payne, J., Yang, W., Wu, D., Parton, J.M. (2017). Repeated measurement of absolute and relative judgments of loudness: clinical relevance for prescriptive fitting od aided target gains for "soft", "comfortable", and "loud, but OK" sound levels. Semin. Hear. 38(1), 26-52. Doi: 10.1055/s-0037-1598064.

Franks, J.R., Beckmann, N.J. (1985). Rejection of hearing aids: attitudes of a geriatric sample. Ear. Hear. 6(3), 161-166. Doi: 10.1097/00003446-198505000-00007.

Fruhmann, M., Chalupper, J., and Fastl, H. (2003). Zum einfluss von innenohrschwerhörigkeit auf die lautheitssummation (Influence of cochlear hearing impairment on spectral loudness summation). In DAGA 2003—Fortschritte der Akustik, Proceedings of the 29th Annual Meeting of the Deutsche Gesellschaft für Akustik e.V., pp. 253–254.

Fuente, A. (2015). The olivocochlear system and protection from acoustic trauma: a mini literature review. Front. Syst. Neurosci. 22, 9:94. Doi: 10.3389/fnsys.2015.00094.

Gacek, R.R., Schuknecht, H.F. (1969). Pathology of presbycusis. Int. Aud. 8, 199-209. Doi: 10.3109/05384916909079061.

Garner, W. R. (1948). The loudness of repeated short tones. J. Acoust. Soc. Am. 20, 513-527. Doi: 10.1121/1.1906405.

Garnier, S., Micheyl, C., Arthaud, P., Berger-Vachon, C., Collet, L. (1999). Temporal loudness integration and spectral loudness summation in normal-hearing and hearing-impaired listeners. Acta Otolaryngol. 119, 154-157. Doi: 10.1080/00016489950181567.

Gatehouse, S, and Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). Int J Audiol. 43(2), 85-99. Doi:10.1080/14992020400050014.

Gengel, R.W., Pascoe, D., Shore, I. (1971). A frequency-response procedure for evaluating and selecting hearing aids for severely hearing-impaired children. J. Speech. Hear. Dis. 36, 341–353. Doi: 10.1044/jshd.3603.341.

Glasberg, B.R., Moore, B.C. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. J Acoust Soc Am. 79(4), 1020-33. Doi: 10.1121/1.393374.

Glasberg, B.R., Moore, B.C.J. (1990). Derivation of auditory filter shapes from notched-noise data. Hear. Res. 47, 103-138. Doi: 10.1016/0378-5955(90)90170-t.

Glasberg, B.R., Moore, B.C.J. (2002) A Model of Loudness Applicable to Time-Varying Sounds. J. Audio Eng. Soc. 50(5), 331-342.
Glasberg, B.R., Moore, B.C.J. (2005). Development and evaluation of a model for predicting the audibility of time-varying sounds in the presence of background sounds. J. Audio. Eng. Soc. 53(10), 906, 918.
Gockel, H., Moore, B. C. J., and Patterson, R. D. (2002). Influence of component phase on the loudness of complex tones. Acust.Acta Acust. 88, 369-377.
Gockel, H., Moore, B.C.J., and Patterson, R.D. (2003). Asymmetry of masking between complex tones and noise: partial loudness. J. Acoust. Soc. Am. 114(1), 349-360. Doi: 10.1121/1.1582447.
Grimm, G., Hohmann, V., and Verhey, J. L. (2002). Loudness of fluctuating sounds. Acust. Acta Acust. 88, 359-368.
Guinan J.J. Jr. (2006). Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans. Ear. Hear. 27, 589-607. Doi: 10.1097/01.aud.0000240507.83072.e7.
Guinan, J.J. Jr., Gifford, M.L. (1988). Effects of electrical stimulation of efferent olivocochlear neurons on cat auditory-nerve fibers. I. Rate-level functions. Hear. Res. 33, 97-113. Doi: 10.1016/0378-5955(88)90023-8.
Guinan, J.J. Jr. (2018). Olivocochlear efferents: Their action, effects, measurement and uses, and the impact of the new conception of cochlear mechanical responses. Hear Res. 362, 38-47. Doi: 10.1016/j.heares.2017.12.012.
Hatch, D.R., Arné, B.C., Hall, J.W. 3rd. (1995). Comodulation masking release (CMR): effects of gating as a function of umber of flanking bands and masker bandwidth. J. Acoust. Soc. Am. 97(6), 3768-3774. Doi: 10.1121/1.412392.
Hawkins, D.B., Prosek, R.A.,Walden, B.E., Montgomery, A.A. (1987). Binaural loudness summation in the hearing impaired. J. Speech Lang. Hear. Res. 30, 37-43. Doi: 10.1044/jshr.3001.37.
Hawley, M.L., Sherlock, L.P., Formby, C. (2017). Intra- and intersubject variability in audiometric measures and loudness judgments in older listeners with normal hearing. Semin. Hear. 38(1), 3-35. Doi: 10.1055/s-0037-1598063.
Heeren, W., Rennies, J., Verhey, J.L. (2011). Spectral loudness summation of nonsimultaneous tone pulses. J. Acoust. Soc. Am. 130(6), 3905-3915. Doi: 10.1121/1.3652866.
Hellman, R.P. (1970). Effect of noise bandwidth on the loudness of a 1000-Hz tone. J. Acoust. Soc. Am. 48, 500-504. Doi: 10.1121/1.1912164.
Hellman, R.P., Zwislocki, J.J. (1963). Monaural Loudness Function at 1000 cps and Interaural Summation. J. Acoust. Soc. Am. 35, 856-865. Doi: 10.1121/1.1918619.
Hellman, R.P., and Zwislocki, J. (1964). Loudness function of a 1000-cps tone in the presence of a masking noise. J. Acoust. Soc. Am. 36, 1618-1627. Doi: 10.1121/1.1919255.
Herzke, T., Hohmann, V. (2005). Effects of instantaneous multiband dynamic compression on speech intelligibility. EURASIP J. App. Sig. P., 18, 3034-43. Doi: 10.1155/ASP.2005.3034.
Hickson, L., Clutterbuck, S., Khan, A. (2010). Factors associated with hearing aid fitting outcomes on the IOI-HA. Int. J. Audiol. 49(8), 586-595. Doi: 10.3109/14992021003777259.
Holube, I. (2011). Speech intelligibility in fluctuating maskers." Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium

on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. The Danavox Jubilee Foundation, 2012.
Houtgast, T. (1974). Lateral suppression and loudness reduction of a tone in noise. Acustica 30, 214-221.
ISO 389-8, 2004. Acoustics - Reference zero for the calibration of audiometric equipment.
ISO 226, 2003. Acoustics – Normal equal-loudness-level contours.
ISO 16832, 2006. Acoustics — Loudness scaling by means of categories
JASP Team, 2018, JASP (Version 0.9) [Computer software].
Jenstad, L.M., Cornelisse, L.E., Seewald, R.C. (1997). Effects of test procedure on individual loudness functions. Ear Hear. 18(5), 401-8. Doi: 10.1097/00003446-199710000-00005.
Jenstad, L.M., Van Tasell, D.J., Ewert, C. (2003). Hearing Aid Troubleshooting Based on Patients' Descriptions. J. Am. Acad. Audiol. 14(7), 347-360. Doi: 10.1055/s-0040-1715754
Johnson, E.E., Dillon, H. (2011). A comparison of gain for adults from generic hearing aid prescriptive methods: impacts on predicted loudness, frequency bandwidth, and speech intelligibility. J. Am. Acad. Audiol. 22(7), 441-59. Doi: 10.3766/jaaa.22.7.5.
Joshi, S.N., Wróblewski, M., Schmid, K.K., Jesteadt, W. (2016). Effects of relative and absolute frequency in the spectral weighting of loudness. J. Acoust. Soc. Am. 139(1), 373-383. Doi: 10.1121/1.4939893.
Kamm, C., Dirks, D.D., Mickey, M.R. (1978). Effects of sensorineural hearing loss on loudness discomfort level and most comfortable loudness judgments. J. Speech Hear. Res. 21(4), 668-681. Doi: 10.1044/jshr.2104.668.
Keidser G., Dillon H., Flax M., Ching T., Brewer S. (2011) The NAL-NL2 prescription procedure. Hearing Research 1(1): 88–90. Doi: 10.4081/audiores.2011.e24.
Keidser, G., Dillon, H., Carter, L., O'Brien, A. (2012). NAL-NL2 empirical adjustments. Trends Amplif. 2012, 16(4), 211-223. Doi: 10.1177/1084713812468511.
Keithley, E.M. (2020). Pathology and mechanisms of cochlear aging. J. Neurosci. Res. 98(9), 1674-1684. Doi: 10.1002/jnr.24439.
Killion, M.C., Fikret-Pasa, S. (1993). The 3 types of sensorineural hearing loss: loudness and intelligibility considerations. Hear. J. 46(11), 31-36.
Kimberley, B.P., and Nelson, D.A. (1989). Temporal overshoot in simultaneous-masked psychophysical tuning curves from normal and hearing-impaired listeners. J. Acoust. Soc. Am. 85(4), 1660-1665. Doi: 10.1121/1.397954.
Köbler, S., Rosenhall, U. (2002). Horizontal localization and speech intelligibility with bilateral and unilateral hearing aid amplification. Int. J. Audiol. 41(7), 395-400. Doi: 10.3109/14992020209090416.
Kochkin, S. (2009). MarkeTrak VIII: 25-year trends in the hearing health market. Hearing Review 16, 12-31.
Kochkin, S. (2000). MarkeTrak V: "Why my hearing aids are in the drawer": the consumers' perspective. The Hear. J. 53, 34-41.
Kohlrausch, A., Fassel, R., Heijden, M. Van Der, Kortekaas, R., Par, S. Van De, Oxenham, A.J., Püschel, D. (1997). Detection of tones in low-noise noise: further evidence for the role of envelope fluctuations. Acta Acust. 83, 659-669.

Langhans, A., Kohlrausch, A. (1992). Differences in auditory performance between monaural and dichotic conditions. I: masking thresholds in frozen noise. J. Acoust. Soc. Am. 91(6), 3456-70. Doi: 10.1121/1.402834.
Levitt, H. (1971). Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49, 467-477. Doi: 10.1121/1.1912375.
Libby, E. (1986). The 1/3–2/3 insertion gain hearing aid selection guide. Hear. Instrum. 37, 27–28.
Löhler, J., Akcicek, B., Kappe, T., Schlattmann, P., Schönweiler, R. (2016). The influence of frequency-dependent hearing loss to unaided APHAB scores. Eur. Arch. Otorhinolaryngol. 273(11), 3587-3593. Doi: 10.1007/s00405-016-3966-9.
Lutman M. E., Gatehouse S., Worthington A. G. (1991). Frequency resolution as a function of hearing threshold level and age. J. Acoust. Soc. Am. 89, 320-328. Doi: 10.1121/1.400513.
Macrae, J. (1986). Relationships between the hearing threshold levels and aided speech discrimination of severely and profoundly deaf children. In NAL Report No.107. Canberra: Australian Government Publishing Service.
Marks, L.E. (1978). Binaural summation of the loudness of pure tones. J. Acoust. Soc. Am. 64, 107-113. Doi: 10.1121/1.381976.
Mapes-Riordan, D., Yost, W.A. (1999). Loudness recalibration as a function of level. J. Acoust. Soc. Am. 106(6), 3506-11. Doi: 10.1121/1.428203.
Marks, L. E., and Florentine, M. (2011). Measurement of loudness, Part I: Methods, problems, and pitfalls. in Loudness, edited by M. Florentine, A. N. Popper and R. R. Fay (Springer, New York).
Marozeau, J., Epstein, M., Florentine, M., Daley, B. (2006). A test of the binaural equal-loudness-ratio hypothesis for tones. J Acoust Soc Am. 120(6), 3870-7. Doi: 10.1121/1.2363935.
Marozeau, J., Florentine, M. (2007). Loudness growth in individual listeners with hearing losses: A review. J. Acoust. Am. 122(3), EL82-EL87. Doi: 10.1121/1.2761924.
Martin, F.N., Champlin, C.A., Chambers, J.A. (1998) Seventh survey of audiometric practices in the United States. J. Am. Acad. Audiol. 9, 95-104.
McCandless, G., Lyregaard, P. (1983). Prescription of gain/ output (POGO) for hearing aids. Hear. Instrum. 35, 16-21.
McFadden, D., and Champlin, C.A. (1990). Reductions in overshoot during aspirin use. J. Acoust. Soc. Am. 87, 2634-2642. Doi: 10.1121/1.399056.
Moore, B.C., Glasberg, B.R. (1983) Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J Acoust Soc Am. 74(3), 750-753. Doi: 10.1121/1.389861.
Moore, B., Glasberg, B. (1996). A revision of Zwicker's loudness model. Acustica-Acta Acustica. 82, 335-345.
Moore, B.C.J., Glasberg, B.R. (1997). A model of loudness perception applied to cochlear hearing loss. Auditory Neuroscience 3(3), 289-311.
Moore, B.C.J., Glasberg, B.R., and Baer, T. (1997). A model for the prediction of thresholds, loudness and partial loudness. J. Audio Eng. Soc. 45(4), 224-240.
Moore, B.C.J. (1998). Cochlear Hearing Loss. Whurr London.
Moore, B. C. J., Launer, S., Vickers, D.A., Baer, T. (1998). Loudness of modulated sounds as a function of modulation rate, modulation depth, modulation waveform and overall level. - In: Psychophysical and physiological advances in

hearing. A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (eds.). Whurr, London.

Moore, B.C., Vickers, D.A., Baer, T., Launer, S. (1999a). Factors affecting the loudness of modulated sounds. J. Acoust. Soc. Am. 105(5), 2757-72. Doi: 10.1121/1.426893.

Moore, B. C. J., Vickers, D. A., Plack, C. J., Oxenham, A. J. (1999b). Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism. J. Acoust. Soc. Am. 106, 2761-2778. Doi: 10.1121/1.428133.

Moore, B.C.J., (2003). An Introduction to the Psychology of Hearing. London: Academic Press.

Moore, B.C.J., Glasberg, B.R. (2004). A revised model of loudness perception applied to cochlear hearing loss. Hear. Res. 188(1-2), 70-88. Doi: 10.1016/S0378-5955(03)00347-2.

Moore, B.C.J. (2004). Testing the concept of softness imperception: Loudness near threshold for hearing-impaired ears. J. Acoust. Soc. Am. 115(6), 3103-3111. Doi: 10.1121/1.1738839.

Moore, B.C.J., Gibbs, A., Onions, G., Glasberg, B.R. (2014). Measurement and modeling of binaural loudness summation for hearing-impaired listeners. J. Acoust. Soc. Am. 136, 736-747. Doi: 10.1121/1.4889868.

Moore, B.C., Glasberg, B.R., Varathanathan, A., Schlittenlacher, J. (2016). A Loudness Model for Time-Varying Sounds Incorporating Binaural Inhibition. Trends Hear. 20. Doi: 10.1177/2331216516682698

Mueller, H.G. (2003). Fitting test protocols are more honored in the breach than in the observance. The Hearing Journal, 56(10), 19-26. Doi: 10.1097/01.HJ.0000292915.14697.f3

Mueller, G.H., Bentler, R.A. (2005). Fitting hearing aids using clinical measures of loudness discomfort levels: an evidence-based review of effectiveness. J. Am. Acad. Audiol. 16(7), 461-472. Doi: 10.3766/jaaa.16.7.6.

Munson, W.A., (1947). The growth of auditory sensation. J. Acoust. Soc. Am. 19, 584. Doi: 10.1121/1.1916525.

Nelson, E.G., Hinojosa, R. (2006). Presbycusis: a human temporal bone study of individuals with downward sloping audiometric patterns of hearing loss and review of the literature. Laryngoscope 116(9 Pt. 3 Suppl. 112), 1-12. Doi: 10.1097/01.mlg.0000236089.44566.62.

Nelson, P.B., Perry, T.T., Gregan, M., Van Tasell, D.J. (2018). Self-adjusted amplification parameters produce large between-subject variability and preserve speech intelligibility. Trends Hear. 22, 1-13. Doi: 10.1177/2331216518798264.

Niese, H. (1959). Die Trägheit der Lautstärkebildung in Abhängigkeit von Schallpegel. Hochfrequenztechn.u.Elektroak. 68, 143-152.

Nizami, L. (2006). The intensity-difference limen for 6.5 kHz: an even more severe departure from Weber’s law. Percept. Psychophys. 68(7), 1107-1112. Doi: 10.3758/bf03193713.

Noble, W., Gatehouse, S. (2006). Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the Speech, and Qualities of Hearing scale (SSQ). Int. J. Audiol. 45(3), 172-181. Doi: 10.1080/14992020500376933.

Oberfeld, D. (2007). Loudness changes induced by a proximal sound: Loudness enhancement, loudness recalibration, or both? J. Acoust. Soc. Am. 121(4), 2137, 2148. Doi: 10.1121/1.2710433.
Oberfeld, D. (2008). The mid-difference hump in forward-masked intensity discrimination. J. Acoust. Soc. Am. 123(3), 1571-1581. Doi: 10.1121/1.2837284.
Oberfeld, D., Plank, T. (2011). The temporal weighting of loudness: effects of the level profile. Atten. Percept. Psychophys. 73(1), 189-208. Doi: 10.3758/s13414-010-0011-8.
Oberfeld, D., Stahn, P. (2012). Sequential grouping modulates the effect of non-simultaneous masking on auditory intensity resolution. PLoS One, 7(10), 1-11. Doi: 10.1371/journal.pone.0048054.
Oberfeld, D., Heeren, W., Rennies, J., Verhey, J. (2012). Spectro-temporal weighting of loudness. PloS One 7(11), e50184. Doi: 10.1371/journal.pone.0050184.
Oberfeld, D., Stahn, P., Kuta, M. (2014). Why do forward maskers affect auditory intensity discrimination? Evidence from "molecular psychophysics". PLoS One, 9(6), 1-14. Doi: 10.1371/journal.pone.0099745.
Oberfeld, D., Jung, L., Verhey, J.L., Hots, J. (2018). Evaluation of a model of temporal weights in loudness judgments. J. Acoust. Soc. Am. 144(2), EL119. Doi: 10.1121/1.5049895.
Oetting, D., Brand, T., Ewert, S.D. (2014). Optimized loudness-function estimation for categorical loudness scaling data. Hearing research 316, 16-27. Doi: 10.1016/j.heares.2014.07.003.
Oetting, D., Hohmann, V., Appell, J.-E., Kollmeier, B., Ewert, S.D. (2016). Spectral and binaural loudness summation for hearing-impaired listeners. Hearing research. 335, 179-192. Doi: 10.1016/j.heares.2016.03.010.
Oetting, D., Hohmann, V., Appell, J.E., Kollmeier, B., Ewert, S.D. (2018) Restoring perceived loudness for listeners with hearing loss. Ear Hear. 39(4), 664-678. Doi: 10.1097/AUD.0000000000000521.
Oghalai, J.S. (2004). The cochlear amplifier: augmentation of the traveling wave within the inner ear. Curr. Opin. Otolaryngol. Head Neck Surg. 12(5), 431-8. Doi: 10.1097/01.moo.0000134449.05454.82.
Oxenham, A.J., Bacon, S.P. (2003). Cochlear compression: perceptual measures and implications for normal and impaired hearing. Ear Hear. 24(5), 352-66. Doi: 10.1097/01.AUD.0000090470.73934.78.
Ozimek, E. and Zwislocki, J. J. (1996). Relationships of intensity discrimination to sensation and loudness levels: dependence on sound frequency. J.Acoust.Soc.Am. 100(5), 3304-3320. Doi: 10.1121/1.416993.
Palmer, C.V., Lindley, G.A. 4th. (1998). Reliability of the Contour Test in a population of adults with hearing loss. J. Am. Acad. Audiol. 9(3), 209-15.
Parker, S., Moore, J.M., Bahraini, S., Gunthert, K., Zellner, D.A. (2012). Effects of expectations on loudness and loudness difference. Atten Percept Psychophys. 74(6), 1334-42. Doi: 10.3758/s13414-012-0326-8.
Pascoe, D. (1978). An approach to hearing aid selection. Hear. Instr. 29, 12-16, 36.
Pavel, M., and Iverson, G.J. (1981). Invariant characteristics of partial masking: implications for mathematical models. J. Acoust. Soc. Am. 69(4), 1126-1131. Doi: 10.1121/1.385692.

Pedersen, B., Ellermeier, W. (2008). Temporal weights in the level discrimination of time-varying sounds. J. Acoust. Soc. Am. 123(2), 963-72. Doi: 10.1121/1.2822883.

Penner, M.J. (1974). Effect of masker duration and masker level on forward and backward masking. J. Acoust. Soc. Am. 56(1), 179-182. Doi: 10.1121/1.1903250.

Perry, T.T., Nelson, P.B., Van Tasell, D.J. (2019). Listener factors explain little variability in self-adjusted hearing aid gain. Trends Hear. 23, 1-17. Doi: 10.1177/2331216519837124.

Pieper, I., Mauermann, M., Oetting, D., Kollmeier, B., Ewert, S.D. (2016). Physiological motivated transmission-lines as front end for loudness models. J. Acoust. Soc. Am. 139(5), 2896. Doi: 10.1121/1.4949540.

Pieper, I., Mauermann, M., Oetting, D., Kollmeier, B., Ewert, S.D. (2018). Physiologically motivated individual loudness model for normal hearing and hearing impaired listeners. J Acoust. Soc. Am. 144(2), 917-930. Doi: 10.1121/1.5050518.

Pieper, I., Mauermann, M., Kollmeier, B., Ewert, S.D. (2021). Toward an Individual Binaural Loudness Model for Hearing Aid Fitting and Development. Front Psychol. 12, 1-23. Doi: 10.3389/fpsyg.2021.634943.

Pollack, I. (1951). On the Threshold and Loudness of repeated Bursts of Noise. J. Acoust. Soc. Am. 23, 646-650. Doi: 10/1121/1.1906814.

Port, E. von (1963a). Über die Lautstärke einzelner kurzer Schallimpulse. Acustica. 13, 212-223.

Port E. von (1963b). Zur Lautstärkeempfindung und Lautstärkemessung von pulsierenden Geräuschen (On the loudness perception and the loudness measurement of pulsating noises). Acustica 13, 224–233.

Preece, J. P. and Wilson, R. H. (1988). Detection, loudness, and discrimination of five-component tonal complexes differing in crest factor. J. Acoust. Soc. Am. 84(1), 166-171. Doi: 10.1121/1.396983.

Rasetshwane, D.M., Trevino, A.C., Gombert, J.N., Liebig-Trehearn, L., Kopun, J.G., Jesteadt, W., Neely, S.T., Gorga, M.P. (2015). Categorical loudness scaling and equal-loudness contours in listeners with normal hearing and hearing loss. J. Acoust. Soc. Am. 137(4), 1899-1913. Doi: 10.1121/1.4916605.

Rasetshwane, D.M., High, R.R., Kopun, J.G., Neely, S.T., Gorga, M.P., Jesteadt, W. (2018). Influence of suppression on restoration of spectral loudness summation in listeners with hearing loss. J.Acoust. Am. 143(5), 2994-3008. Doi: 10.1121/1.5038274.

Reichardt, W. and Niese, H. (1965). Die Addition der Schallerregungen in den einzelnen Frequenzgruppen bei impulsiven Geräuschen. Acustica. 16, 295-304.

Reichardt, W. (1970). Subjective and objective measurement of the loudness level of single and repeated impulses. J. Acoust. Soc. Am. 47, 1557-1562. Doi: 10.1121/1.1912087.

Reichardt, W. (1965). Zur Trägheit der Lautstärkebildung. Acustica. 15, 345-354.

Rennies, J., Verhey, J.L., Chalupper, J., Fastl, H. (2009). Modeling temporal effects of spectral loudness summation. Acust. Acta Acust. 95, 1112-1122.

Rennies, J., Verhey, J.L., Fastl, H. (2010). Comparison of loudness models for time-varying sounds. Acust. Acta Acust. 96, 383-396. Doi: 10.3813/AAA.918287.

Rennies, J., Wächtler, M., Hots, J., Verhey, J.L. (2015). Spectro-temporal characteristics affecting the loudness of technical sounds: data and model predictions. Acust. Acta Acust. 101, 1145-1156.
Reynolds, G.S., and Stevens, S.S. (1960). Binaural summation of loudness. J. Acoust. Soc. Am. 32, 1337-1344. Doi: 10.1121/1.1907903.
Robinson, K., Gatehouse, S. (1996). Test-retest reliability of loudness scaling. Ear Hear. 17(2), 120-123. Doi: 10.1097/00003446-199604000-00005.
Robles, L., Ruggero, M.A., Rich, N.C. (1986). Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases. J Acoust Soc Am. 80(5), 1364-74. Doi: 10.1121/1.394389.
Röhl, M., Uppenkamp, S. (2010). An auditory fMRI correlate of impulsivity. Psychiatry Res. 28, 181(2), 145-50. Doi: 10.1016/j.pscychresns.2009.09.002.
Röhl, M., Uppenkamp, S. (2012). Neural coding of sound intensity and loudness in the human auditory system. J Assoc Res Otolaryngol. 13(3), 369-79. Doi: 10.1007/s10162-012-0315-6.
Roverud, E., Strickland, E.A. (2010). The time course of cochleair gain reduction measured using a more efficient psychophysical technique. J. Acoust. Soc. Am. 128(3), 1203-1214. Doi: 10.1121/1.3473695.
Roverud, E., Strickland, E.A. (2015a). Exploring the source of the mid-level hump for intensity discrimination in quiet and the effects of noise. J. Acoust. Soc. Am. 137(3), 1318-1335. Doi: 10.1121/1.4908243.
Roverud, E., Strickland, E.A. (2015b). The effect of ipsilateral, contralateral, and bilateral broadband nooise on the mid-level hump in intensity discrimination. J. Acoust. Soc. Am. 138(5), 3245-3261. Doi: 10.1121/1.4935515.
Ruggero, M.A., and Rich, N.C. (1991). Furomeside alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane. J. Neurosci. 11, 1057-1067. Doi: 10.1523/JNEUROSCI.11-04-01057.1991.
Russell, I.J., and Murugasu, E. (1997). Medial efferent inhibition suppresses basilar membrane responses to near characteristic frequency tones of moderate to high intensities. J.Acoust. Soc.Am. 102(3), 1734-1738. Doi: 10.1121/1.420083.
Salvi, R., Sun, W., Ding, D., Chen, G.D., Lobarinas, E., Wang, J., Radziwon, K., Auerbach, B.D. (2017). Inner hair cell loss disrupts hearing and cochlear function leading to sensory deprivation and enhanced central gain. Front. Neurosci. 10 (621), 1-18. Doi: 10.3389/fnins.2016.00621.
Scharf, B. (1959). Critical bands and the loudness of complex sounds near threshold. J. Acoust. Soc. Am. 31, 365-370. Doi.org/10.1121/1.1907725.
Scharf, B. (1968). Binaural loudness summation as a function of bandwidth. In Y. Kohasi (Ed.), Reports of the 6th International Congress on Acoustics (Vol. I, pp. 25-28). Tokyo: Marazen A-3-5.
Scharf, B., Hellman, R.P. (1966). Model of loudness summation applied to impaired ears. J. Acoust. Soc. Am. 40(1), 71-78. Doi: 10.1121/1.1910066.
Scharf, B. (1964). Partial masking. Acustica 14, 16-23.
Scharf, B. and Fishken, D. (1970). Binaural summation of loudness reconsidered. J. Exp. Psychol. 86. 374-379. Doi: 10.1037/h0030159.
Scharf, B., Reeves, A., and Giovanetti, H. (2008). Role of attention in overshoot: frequency certainty versus uncertainty. J. Acoust. Soc. Am. 123(3), 1555-1561. Doi: 10.1121/1.2835436.

Schlittenlacher, J., Ellermeier, W., Hashimoto, T. (2015). Spectral loudness summation: shortcomings of current standards. J. Acoust. Soc. Am. 137 (1), EL25-EL31. Doi: 10.1121/1.4902425.
Schlauch, R.S., Lanthier, N., Neve, J. (1997). Forward-masked intensity discrimination: Duration effects and spectral effects. J. Acoust. Soc. Am. 102(1), 461-467. Doi: 10.1121/1.419610.
Schmidt, S., and Zwicker, E. (1991). The effect of masker spectral asymmetry on overshoot in simultaneous masking. J. Acoust. Soc. Am. 89(3), 1324-1330. Doi: 10.1121/1.400656.
Schreiner, C.E., Malone, B.J. (2015). Representation of loudness in the auditory cortex. Handb Clin Neurol. 129, 73-84. Doi: 10.1016/B978-0-444-62630-1.00004-4.
Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Laurnagaray, D., Beaulac, S., Pumford, J. (2005). The desired sensation level multistage input/output algorithm. Trends Amplif. 9(4), 159-197. Doi: 10.1177/1380500900403.
Seewald, R.C., Ross, M., Spiro, M.K. (1985). Selecting amplification characteristics for young hearing-impaired children. Ear Hear. 6(1), 48-53. Doi: 10.1097/00003446-198501000-00013.
Shapiro, I. (1979). Evaluation of relationship between hearing threshold and loudness discomfort level in sensorineural hearing loss. J. Speech. Hear. Disord. 44(1), 31-6. Doi: 10.1044/jshd.4401.31.
Silva, I., Florentine, M. (2006). Effect of adaptive psychophysical procedure on loudness matches. J. Acoust. Soc. Am. 120(4), 2124-31. Doi: 10.1121/1.2336747.
Sivonen, V., Ellermeier, W. (2011). Binaural loudness. in Loudness, edited by M. Florentine, A. N. Popper, and R. R. Fay (Springer, New York).
Smeds, K. (2004). Is normal or less than normal overall loudness preferred by first-time hearing aid users? Ear Hear. 25(2), 159-72. Doi: 10.1097/01.aud.0000120364.13671.e2.
Smeds, K., Keidser, G., Zakis, J., Dillon, H., Leijon, A., Grant, F., Convery, E., Brew, C. (2006a). Preferred overall loudness. I: Sound field presentation in the laboratory. Int. J. Audiol. 45(1), 2-11. Doi: 10.1080/14992020500190144.
Smeds, K., Keidser, G., Zakis, J., Dillon, H., Leijon, A., Grant, F., Convery, E., Brew, C. (2006b) Preferred overall loudness. II: Listening through hearing aids in field and laboratory tests. Int. J. Audiol. 45(1), 12-25. Doi: 10.1080/14992020500190177.
Smith, L.Z., Boothroyd, A. (1989). Performance intensity function and speech perception in hearing impaired children. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, St. Louis, MO.
Spiegel, M.F., Picardi, M.C., Green, D.M. (1981) Signal and masker uncertainty in intensity discrimination. J. Acoust. Soc. Am. 70(4), 1015-9. Doi: 10.1121/1.386951.
Stephens, S.D.G. (1973). Auditory temporal integration as a function of intensity. J. Sound Vib. 30(1), 109-126. Doi: 10.1016/S0022-460X(73)80054-9.
Stevens, S. S. (1936). A scale for the measurement of a psychological magnitude: loudness. Psychol. Rev. 43(5), 405–416. Doi: 10.1037/h0058773.

Stevens, S. S. (1955). The measurement of loudness. J. Acoust. Soc. Am. 27, 815–829. Doi: 10.1121/1.1908048.
Stevens, S.S. (1956). The direct estimation of sensory magnitudes-loudness. Am. J. Psychol. 69(1), 1-25.
Stevens, S.S., and Guirao, M. (1967). Loudness functions under inhibition. Percept. Psychophysics. 2, 459-465.
Strelcyk, O., Nooraei, N., Kalluri, S., Edwards, B. (2012). Restoration of loudness summation and differential loudness growth in hearing-impaired listeners. J. Acoust. Soc. Am. 132, 2557-2568. Doi: 10.1121/1.4747018.
Strickland, E. A. (2001). The relationship between frequency selectivity and overshoot. J. Acoust. Soc. Am. 109, 2062-2073. Doi: 10.1121/1.1357811.
Strickland, E. A. (2004). The temporal effect with notched-noise maskers: analysis in terms of input-output functions. J. Acoust. Soc. Am. 115, 2234-2245. Doi: 10.1121/1.1691036.
Strickland, E. A. and Krishnan, L. A. (2005). The temporal effect in listeners with mild to moderate cochlear hearing impairment. J. Acoust. Soc. Am. 118, 3211-3217. Doi: 10.1121/1.2074787.
Strickland, E. A. (2008). The relationship between precursor level and the temporal effect. J. Acoust. Soc. Am. 123(2), 946-954. Doi: 10.1121/1.2821977.
Suga, F., Lindsay, J.R. (1976). Histopathological observations of presbycusis. Ann. Otol. Rhinol. Laryngol. 85(2 pt.1), 169-84. Doi: 10.1177/000348947608500201.
Thiele, C., Wardenga, N., Lenarz, T., Büchner, A. (2014). Testing the comparability of free-field and HDA200 headphone measurements for the Freiburg speech test. HN) 62(2), 115-120. Doi: 10.1007/s00106-013-2789-x.
Van Beurden, M.F.B., Boymans, M. Jansen, N., Dreschler, W.A. (2007). In Auditory signal processing in hearing-impaired listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR). Edited by T. Dau, J. M. Buchholz, J. M. Harte, T. U. Christiansen. ISBN: 87-990013-1-4.
Van Beurden, M.F.B., and Dreschler, W. A. (2005). Bandwidth dependency of loudness in series of short noise bursts. Acta Acustica united with Acustica, 2005, 91(1), 1020-1024.
Van Beurden, M.F.B., and Dreschler, W. A. (2007). Duration dependency of spectral loudness summation, measured with three different experimental procedures. In Hearing - From Sensory Processing to Perception. Edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann U, Mauermann M, Uppenkamp, S, Verhey JL, (Springer, Berlin), pp. 237-245.
Van Beurden, M.F.B., Dreschler, W.A. (2018). Partial loudness at masker onset indicates temporal effects at supra-threshold levels. Hear Res. 370, 168-180. Doi: 10.1016/j.heares.2018.10.012.
Van Beurden, M.F.B., Boymans, M., van Geleuken, M., Oetting, D., Kollmeier, B., Dreschler, W.A. (2018). Potential consequences of spectral and binaural summation for bilateral hearing aid fitting. Trends Hear. 22, 1-15. Doi: 10.1177/2331216518805690.
Van Beurden, M.F.B., Boymans, M., van Geleuken, M., Oetting, D., Kollmeier, B., Dreschler, W.A. (2021). Uni- and bilateral spectral loudness summation and binaural loudness summation with loudness matching and categorical loudness scaling. Int J Audiol. 60(5), 350-358. Doi: 10.1080/14992027.2020.1832263.

Verhey, J.L., (1999). Psychoacoustics of spectro-temporal effects in masking and loudness perception. BIS-Verlag, Oldenburg, ISBN 3-8141-0662-2, http://docserver.bis.uni-oldenburg.de/publikationen/bisverlag/verpsy99/verspy99.html.
Verhey, J.L., and Kollemier, B. (2002). Spectral loudness summation as a function of duration. J.Acoust. Soc. Am. 111(3), 1349-1358. Doi: 10.1121/1.1451065.
Verhey, J.L., Anweiler, A.-K., Hohmann, V. (2006). Spectral loudness summation as a function of duration for hearing-impaired listeners. Int. J. Audiol. 45, 287-294. Doi: 10.1080/14992020500485692.
Verhey, J.L., Uhlemann, M. (2008). Spectral loudness summation for sequences of short noise bursts. J. Acoust. Soc. Am. 123(2), 925-34. Doi: 10.1121/1.2822318.
Vermeire, K., Knoop, A., Boel, C., Auwers, S., Schenus, L., Talaveron-Rodriguez, M., De Boom, C., De Sloovere, M. (2016). Speech Recognition in Noise by Younger and Older Adults: Effects of Age, Hearing Loss, and Temporal Resolution. Ann. Otol. Rhinol. Laryngol. 125(4), 297-302. Doi: 10.1177/0003489415611424.
von Klitzing, R. and Kohlrausch, A. (1994). Effect of masker level on overshoot in running- and frozen-noise maskers. J. Acoust. Soc. Am. 95, 2192-2201. Doi: 10.1121/1.408679.
Wallén, M.B., Hasson, D., Theorell, T., Canlon, B. (2012). The correlation between the hyperacusis questionnaire and uncomfortable loudness levels is dependent on emotional exhaustion. Int J Audiol. 51(10), 722-9. Doi: 10.3109/14992027.2012.695874.
Warr W.B. (1975). Olivocochlear and vestibular efferent neurons of the feline brain stem: their location, morphology and number determined by retrograde axonal transport and acetylcholinesterase histochemistry. J. Comp. Neurol. 161(2), 159-181. Doi: 10.1002/cne.901610203.
Watson, N.A., Knudsen, V.O. (1940). Selective amplification in hearing aids. J.Acoust. Soc.Am. 11, 406. Doi: 10.1121/1.196053.
Whilby, S., Florentine, M., Wagner, E., Marozeau, J. (2006). Monaural and binaural loudness of 5- and 200-ms tones in normal and impaired hearing. J. Acoust. Soc. Am. 119, 3931-3939. Doi: 10.1121/1.2193813.
Wilson, U.S., Sadler, K.M., Hancock, K.E., Guinan, J.J.Jr., Lichtenhan, J.T. (2017). Efferent inhibition strength is a physiological correlate of hyperacusis in children with autism spectrum disorder. J. Neurophysiol. 118, 1164-1172. Doi: 10.1152/jn.00142.2017.
World Health Organization (2021). World report on hearing. Geneva.
Wright, B.A. (1995). Detectability of simultaneously masked signals as a function of signal bandwidth for different signal delays. J. Acoust. Soc. Am. 98, 2493-2503. Doi: 10.1121/1.413280.
Wright, B.A. (1997). Detectability of simultaneously masked signals as a function of masker bandwidth and configuration for different signal delays. J. Acoust. Soc. Am. 101, 420-429. Doi: 10.1121/1.417987.
Yates, G.K. (1990). Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions. Hear Res. 50(1-2), 145-62. Doi: 10.1016/0378-5955(90)90041-m.

Zaugg, T.L., Thielman, E.J., Griest, S., Henry, J.A. (2016). Subjective reports of trouble tolerating sound in daily life versus loudness discomfort levels. Am. J. Audiol. 25, 359-363. Doi: 10.1044/2016_AJA-15-0034.

Zeng, F-g., Turner, C.W., Relkin, E.M. (1991). Recovery from prior stimulation II: Effects upon intensity discrimination. Hear. Res. 55, 223-230. Doi: 10.1016/0378-5955(91)90107-k.

Zeng, F.-g., and Turner, C.W. (1992). Intensity discrimination in forward masking. J. Acoust. Soc. Am. 92(2), 782-787. Doi: 10.1121/1.403947.

Zhang, C. Y. and Zeng, F. G. (1997). Loudness of dynamic stimuli in acoustic and electric hearing. J. Acoust. Soc. Am. 102, 2925-2934. Doi: 10.1121/1.420347.

Zwicker, E., Flottorp, G., Stevens, S.S. (1957). Critical bandwidth in loudness summation. J. Acoust. Soc. Am. 29, 548-557. Doi: 10.1121/1.1908963.

Zwicker, E. (1958). Über psychologische und methodische Grundlagen der Lautheit. Acustica 8, 237-258.

Zwicker, E. (1963). Über die Lautheit von ungedrosselten und gedrosselten Schallen. Acustica 13, 194-211.

Zwicker, E. (1965). Temporal effects in simultaneous masking by white-noise bursts. Acoust. Soc. Am. 37, 653. Doi: 10.1121/1.1909389.

Zwicker, E., Scharf, B. (1965). A model of loudness summation. Psychol. Rev. 72, 3-26. Doi: 10.1037/h0021703.

Zwicker E., Fastl H., Dallmayr, C. (1984). BASIC-Program for calculating the loudness of sounds from their 1/3-oct band spectra according to ISO 532 B. Acustica, 55, 63-67.

Zwicker, E. and Fastl, H. (1990). Psychoacoustics - facts and models. Springer.

Zwicker, E. and, Zwicker, U.T. (1991). Dependence of binaural loudness summation on interaural level differences, spectral distribution, and temporal distribution. J. Acoust. Soc. Am. 89(2), 756-764. Doi: 10.1121/1.1894635.

List of abbreviations

ACALOS	Adaptive Categorical Loudness Scaling
AD	Auris Dexter (right ear)
ADS	Auris dexter et sinister (right and left ear)
AM	Amplitude Modulation
AN	Auditory Nerve
AS	Auris Sinister (left ear)
BB	Broadband
BLDEL	Binaural Level Differences for Equal Loudness
CMR	Comodulation Masking Release
CPH	Cosine Phase
CR	Compression Ratio
CU	Categorical Units
dB	Decibel
DCN	Dorsal Cochlear Nucleus
DL	Difference Limen
DLM	Dynamic Loudness Model
DSL	Desired Sensation Level
ELRH	Equal-Loudness-Ratio-Hypothesis
ERB	Equivalent Rectangular Bandwidth
FF	Free-Field
fMRI	functional Magnetic Resonance Imaging
F_s	Signal frequency
HI	Hearing Impaired
HL	Hearing Level
HP	High-Pass
(k)Hz	(kilo)Herz
I/O	Input/Output
IC	Inferior Colliculus
IFnoise	International Female Noise
ISO	International Organization for Standardization
LDEL	Level Differences for Equal Loudness
LDL	Loudness Discomfort Level
LNN	Low-Noise Noise
LOC	Lateral Olivocochlear
LP	Low-Pass
MGN	Medial Geniculate Nuclei
MLMF	Masked Loudness-Matching Function
MOC	Medial Olivocochlear
MOCR	Medial Olivocochlear Reflex
NAL	National Acoustic Laboratories
NH	Normal Hearing
nLL	nuclei of the Lateral Lemniscus
OC	Olivocochlear
PTA	Pure Tone Average
Rms	Root mean square
RPH	Random Phase
SLDEL	Spectral Level Differences for Equal Loudness
SOC	Superior Olivary Complex
SPL	Sound Pressure Level
UCL	Uncomfortable Loudness
UEN	Uniformly Exciting Noise
VCN	Ventral Cochlear Nucleus
WHO	World Health Organization

References & List of Abbreviations

Summary

&

Samenvatting

Summary

One of the main aims in individual hearing aid fitting is to restore loudness perception with hearing aids close to normal. This aim assumes knowledge about normal loudness perception for a variety of sounds and intensities, and knowledge about how loudness perception changes with hearing loss.

Chapter 1 gives a brief introduction into the basic knowledge on the relationship between loudness perception and the physiology of hearing, how to measure loudness and how to model loudness perception. Loudness measurements and loudness models have led to hearing aid prescription rules that can be used to determine the appropriate gain to compensate for a certain hearing loss. There are at least three limitations in the collection of the loudness data that form the basis for these prescription rules, and these are addressed in this thesis.

The first limitation is that the loudness data are based on measurements with static stimuli, while daily listening conditions most of the time consist of dynamic stimuli. Part 1 of this thesis consists of chapter 2 and 3, and deals with the first limitation. It contains experiments with normal hearing listeners with a focus on the role of the cochlear amplifier in loudness perception of short stimuli.

The second limitation is that the hearing aid prescription rules of NAL and DSL are in principle threshold based, while the loudness growth curve cannot be predicted based on auditory threshold only. In an intermezzo, consisting of chapter 4, the second limitation is treated.
The third limitation is that loudness data for hearing-impaired listeners have been collected for unaided conditions. The influence of loudness normalization in the aided situation is not verified. Part two, consisting of chapter 5 and 6, deals with the third limitation. It contains experiments with hearing-impaired listeners and focuses on the interaction between spectral and binaural loudness summation.

In chapter 2 two experiments on the time dependency of loudness summation were described. The first experiment was an experiment on duration dependent loudness summation, similar to that by Verhey and

Kollmeier (2002). In this experiment two loudness matching procedures were compared, one with a more traditional and one with a more experimental response task. Loudness scaling was added to investigate loudness summation for a large range of levels. Nine normal hearing subjects participated in this experiment.

In the second experiment, the loudness of series of noise bursts was investigated, with well-defined noise bursts as a first approximation of dynamic sounds found in daily life. The series of noise bursts had equal spectral content but differed in burst- and inter-burst-durations. The experiment was performed with two bandwidths, 3200 Hz and 200 Hz. The measuring procedure was the experimental loudness matching procedure of the first experiment. Eight normal hearing subjects participated in this experiment.

Adaptive CAtegorical LOudness Scaling (ACALOS) by Brand and Hohmann (2002) was used as the loudness scaling procedure. The traditional loudness matching procedure was a standard adaptive two-interval, two-alternative forced choice procedure. The experimental loudness matching procedure was a variation on this traditional procedure. The procedure was designed to make matching close to the equal loudness point easier by changing the task from differentiating the loudness of two signals to discriminating if there was a difference in a signal pair or not. We expected that with this task listeners would shift their judgment less to the comfortable loudness level and would ignore the fixed reference sound more easily than in the original task.

In experiment 1 all three procedures showed duration-dependent loudness summation, as was also found by Brand and Hohmann (2002). The amount of loudness summation was smaller for the experimental loudness matching procedure than for the traditional loudness matching procedure. The loudness scaling procedure showed that the duration-dependency of the loudness summation was strongest at mid-levels and seemed to be small at low and high levels.

The results of the second experiment indicated that series of noise bursts needed less intensity to sound equally loud as the continuous noise. The main effect of bandwidth was that the effects of the temporal structure on loudness was about 1 dB stronger for the 3200 Hz bandwidth than for the 200 Hz bandwidth. Unexpectedly, the differences between the different conditions lost their significance when the results were presented in the rms of the total signal. If increased spectral loudness summation for short signals causes an increased loudness perception for series of noise bursts, the spectral loudness summation differences between the two bandwidths would be expected to be larger in the conditions with shorter noise bursts than in conditions with a longer noise burst duration. However, this effect was not apparent in the results.

In Chapter 3, we examined temporal effects both at threshold and at supra-threshold levels. In the context of our focus on loudness of dynamic sounds, it is important to know how changes in signal intensity affect loudness judgments. The ultimate change in signal intensity is signal onset. “Overshoot”, also called “the temporal effect”, refers to the finding that a higher signal-to-masker ratio may be needed to detect a brief signal at the onset of a masker than if it is presented with an onset delay, or if it is preceded by another sound (precursor condition). There is a growing body of evidence showing that overshoot may be explained to a large part by changes in cochlear gain. Loudness models, however, assume a steady gain throughout the course of a signal.

The level needed to detect a short-duration 4 kHz signal was measured for signals presented with different onset delays relative to a 300 ms broadband noise masker: 100 ms and 5 ms before the onset of the masker and 5 ms and 100 ms after the onset of the masker. Loudness matches between the signal in quiet and the signal at the same four onset delays were obtained for five presentation levels of the short-duration signal and for three masker levels. The temporal effect was defined as the level difference between the signals near masker onset and the signals well before or well after masker onset, needed to reach threshold and/or to achieve equal loudness. Both at threshold and at supra-threshold levels temporal effects were observed consistent with a decrease in gain at the masker frequency during the course of the masker. The temporal effect was not restricted to simultaneous masking but was also found in backward masking conditions. In both cases the temporal effects were stronger at supra-threshold levels than at threshold levels. This may be caused by a transient effect at masker onset. The almost simultaneous onsets of signal and masker made it difficult for subjects to separate the signal from the masker, especially when the signal level was close to masked threshold.

Fitting rules used in auditory rehabilitation usually have their main focus on detection thresholds. However, in state-of-the-art hearing aids the prescribed gain is normally nonlinear. To prescribe an appropriate nonlinear gain, information about loudness growth is needed. A measurement procedure that can provide information about loudness growth is loudness scaling.

In Chapter 4, the added value of loudness scaling for clinical applications was examined in three studies. In the first study loudness scaling was performed by a group of musicians with normal hearing and mild hearing losses. Loudness scaling was measured with two narrowband signals (1/3-octave around 750 Hz and 3 kHz) and a broadband signal and the relation with audiometric threshold was investigated. The results of the first study showed that it is hard to define one single normal loudness function. The inter-individual spread in loudness functions within the normal hearing subjects was large. The correlations between audiometric thresholds and the thresholds from loudness scaling were also weak. The correlation coefficients between audiometric thresholds and thresholds from loudness scaling increased if a hearing loss was present. This was a logical consequence from the steepening of the loudness function at low levels in subjects with cochlear hearing loss, which led to a more accurate threshold estimation in the loudness scaling data.

The group results led to another interesting finding. Hearing loss in this population mainly influenced the lower part of the loudness function. The higher part of the loudness function was relatively independent from both audiometric thresholds and thresholds obtained from loudness scaling. The results in this study strongly suggest that recruitment is limited to low and medium levels. The upper part of the loudness curve showed no consistent steepening with increasing threshold and therefore no recruitment. This suggests that for normalization of loudness, compression should be applied mainly for the lower levels with linear amplification at high levels, as otherwise the shape of the normal loudness function will be distorted.

In a second study the difference between monaural and binaural loudness perception in a subgroup of musicians was examined. The results of the second study showed a clear difference between monaural and binaural loudness measurements. Binaurally presented signals were clearly perceived louder than monaurally presented signals. This effect mainly occured at the higher levels. At low levels hardly any binaural summation was found. The binaural loudness data may have implications for hearing aid fitting. In hearing aid fitting loudness normalization is often one of the main targets. However, in Real Ear Measurements (REM) hearing aid fittings are normally evaluated monaurally and not binaurally. If binaural loudness summation is indeed level dependent this should be taken in consideration during hearing aid fitting.

Finally, in the third study described in chapter 4 the correlations between self-reported problems and measures obtained from loudness scaling in a different group of hearing-impaired employees was examined. This study was done in a very heterogeneous group of subjects. Therefore, no strong conclusions may be drawn from this study. In this study loudness measurements in freefield were compared to results of a questionnaire on listening effort. The ratio between m_{low} and m_{high} (which determines the concaveness of the loudness function) showed a weak correlation with listening effort in noise. With a higher ratio corresponding to higher listening effort. The relationship between linearity of the loudness function and listening effort in noise also appeared in the unaided measurements, suggesting that the effect was not created by inadequate hearing aid fitting. If the relationship between high listening effort and very concave loudness functions can be confirmed in follow-up studies, this knowledge may have important consequences for hearing aid fitting. In that case extremely concave loudness functions should be avoided. This means that the use of strong compression needs some precautions.

In chapter 5, categorical loudness scaling (Brand & Hohmann, 2002) was applied in order to measure spectral and binaural loudness summation after narrowband loudness compensation. In a study by Oetting *et al.* (2016) mild to moderate hearing losses were tested for groups of hearing-impaired listeners with pure-tone audiograms corresponding to audiometric configurations of N1–N3 and S1 (Bisgaard *et al.,* 2010). It is not clear whether the effect of individual variation decreases, remains constant, or even increases with increasing hearing loss. Therefore, in this study, a broader range of hearing losses (audiometric configurations: N2–N4 and S2–S3; Bisgaard *et al.,* 2010) was included. The focus was on the effect of the individual variability of loudness summation in a larger variety of hearing losses according to the classification by Bisgaard *et al.* (2010). The main questions of this study were (a) whether the shape of the audiogram could explain individual differences and (b) if several characteristics of the hearing loss and hearing loss compensation strategy were possible predictors for the amount of spectral and binaural loudness summation.

Before loudness summation was determined for the broadband signals (UEN1, UEN5, UEN17, and Ifnoise), the noises were corrected for each hearing-impaired listener individually aiming to present signal levels that produce the same loudness level within each narrowband as for the average normal hearing listener (narrowband loudness normalization). For this purpose, the broadband signals were filtered in six non-overlapping frequency bands having the same center frequencies as the narrowband signals. The required gain for each frequency band was defined as the difference in level for each loudness category between the individual loudness functions of the narrowband signals and the average normal hearing loudness function.

In this study, the effects on spectral loudness summation were largest for the signals with largest bandwidths UEN17 and Ifnoise (in correspondence with, i.e., Zwicker, 1958). In agreement with the results by Oetting *et al.* (2016) spectral loudness summation increased with the degree of hearing loss. This study also showed a hearing loss dependency for spectral loudness summation of binaurally presented signals. For small hearing losses (N2), spectral loudness summation of binaural sounds was the same as for normal hearing listeners. For the larger hearing losses (N3, N4, and S), spectral loudness summation of binaurally presented sounds tended to be higher than normal. With respect to binaural summation of broadband sounds itself (binaural conditions versus monaural conditions), no clear hearing loss dependency was found.

In chapter 6, loudness perception of low-pass filtered, high-pass filtered, and broadband pink noise was compared in sixteen hearing-impaired listeners for both unilateral and bilateral presentation. Current hearing-aid prescription rules assume that spectral loudness summation decreases for hearing-impaired listeners and that binaural loudness summation is independent of hearing loss and signal bandwidth. Previous studies have shown that these assumptions might be incorrect.

The experiment in chapter 6 was designed to investigate the validity of the assumption that binaural loudness summation has the same effect on all signals. With a special focus on the role of low and high frequencies in spectral and binaural loudness summation. For this purpose, the loudness of a broadband noise was compared to the loudness of its lower and higher frequency part. The loudness of these three signals was measured in a group of hearing-impaired listeners using loudness scaling and loudness matching for signals that were presented unilaterally and bilaterally. The results were analyzed for spectral effects in unilateral and bilateral presentation, for binaural effects, and for inter-individual differences.

Changes in loudness perception found in unilateral presentation were found to be enlarged in bilateral presentation. In the group of hearing-impaired listeners investigated in this study, this led to a larger contribution of the low frequency part of the signal to the total loudness sensation than the high frequency part in bilateral presentation compared to unilateral presentation. This appeared to be a consequence of more binaural loudness summation for the low frequency part of the spectrum than for the high frequency part.

Secondly, individual differences in loudness perception usually are not taken into account in hearing-aid prescriptions. An important outcome of the current study was that individual variability in loudness perception near loudness discomfort levels was so large that it must be taken into consideration in individual hearing-aid fitting. Information on individual loudness growth appears to be indispensable for a good hearing aid fitting. Loudness scaling provides information on the entire loudness range, but loudness scaling is a fairly time-consuming procedure. It may not be necessary to measure loudness growth for each separate signal, when the loudness differences between signals are known. As measurement time is an important issue in the clinic, a second objective was to investigate if loudness matching could be applied as a reliable and more time efficient alternative for categorical loudness scaling. This appears to be the case as the individual variability does not seem to be influenced by the measurement procedure (loudness scaling, or loudness matching), or narrow-band loudness compensation.

Measuring individual loudness perception for broadband and binaural signals should be considered in the hearing aid verification toolbox. Which set of signals would be the best choice, could be a topic for further research, but loudness matching appears to be a reliable and time efficient procedure to quickly measure loudness differences between signals with different bandwidths and presentation modes (unilateral vs. bilateral).

The results of the studies described in this thesis lead to six general conclusions:

1) Spectral loudness summation depends on the duration of the signal. This duration dependency is shown to have a small effect in series of noise bursts.
2) The temporal effect has been shown to be also present at supra-threshold levels.
3) The temporal effect is even larger at supra-threshold levels than at threshold and is not confined to simultaneous masking, but can also be found shortly before onset of the masker.
4) Larger hearing losses and more sloping hearing losses classified according to Bisgaard et al. (2010) lead to larger than normal spectral loudness summation both in unilateral as in bilateral presentation.
5) In hearing-impaired listeners binaural loudness summation is larger for the low frequency part of a broadband signal than for the high frequency part.
6) The variability in loudness perception in both normal hearing and hearing-impaired listeners cannot be ignored in loudness normalization. Therefore it is expected that individual loudness measurements will provide a positive influence on hearing aid fittings.

Samenvatting

Eén van de belangrijkste doelen bij het individueel aanpassen van hoortoestellen is het herstellen van de luidheidsperceptie met de hoortoestellen naar zo normaal mogelijk. Dit doel veronderstelt kennis omtrent luidheidsperceptie voor een verscheidenheid aan geluiden en intensiteiten en kennis over hoe luidheidsperceptie verandert ten gevolge van gehoorverlies.

Hoofdstuk 1 geeft een korte introductie in de basiskennis over de relatie tussen luidheidsperceptie en de fysiologie van het gehoor, hoe luidheid te meten en hoe luidheidsperceptie te modelleren. Luidheidsmetingen en luidheidsmodellen hebben geleid tot hoortoestelvoorschrijfregels, die gebruikt kunnen worden om de juiste versterking te bepalen om een bepaald gehoorverlies te compenseren. Er zijn ten minste drie beperkingen bij het verzamelen van de luidheidsgegevens, die de basis vormen voor deze voorschrijfregels en deze worden in dit proefschrift behandeld.

De eerste beperking is dat de luidheidsgegevens gebaseerd zijn op metingen met statische stimuli, terwijl dagelijkse luisteromstandigheden meestal bestaan uit dynamische stimuli.

Deel 1 van dit proefschrift bestaat uit hoofdstuk 2 en 3 en behandelt de eerste beperking. Het bevat experimenten met normaalhorende luisteraars met een focus op de rol van de cochleaire versterking in luidheidsperceptie van korte stimuli.

De tweede beperking is dat de hoortoestelvoorschrijfregels van NAL en DSL in principe gebaseerd zijn op drempelwaarden, terwijl de opbouw van de luidheidscurve niet kan worden voorspeld op basis van alleen de gehoordrempel. In een intermezzo, bestaande uit hoofdstuk 4, wordt de tweede beperking behandeld.
De derde beperking is dat luidheidsgegevens voor slechthorende luisteraars zijn verzameld voor omstandigheden zonder hoortoestellen. De invloed van luidheidnormalisatie zelf op luidheidsperceptie is niet gecontroleerd.

Deel 2, bestaande uit hoofdstuk 5 en 6, behandelt de derde beperking. Het bevat experimenten met slechthorende luisteraars en richt zich op de interactie tussen spectrale en binaurale luidheidssommatie.

In hoofdstuk 2 werden twee experimenten aangaande de tijdsafhankelijkheid van luidheidssommatie beschreven. Het eerste experiment was een experiment aangaande tijdsafhankelijke luidheidssommatie, vergelijkbaar met dat van Verhey en Kollmeier (2002). In dit experiment werden twee luidheidsmatching procedures vergeleken, één met een meer traditionele en één met een meer experimentele antwoordtaak. Luidheidschaling was toegevoegd om luidheidssommatie over een groter intensiteitsbereik te onderzoeken. Negen normaalhorende proefpersonen werkten mee aan het experiment.

In het tweede experiment werd de luidheid van een reeks korte ruisjes onderzocht met de goed gedefinieerde noise bursts als eerste benadering van dynamische geluiden die in het dagelijks leven voorkomen. De serie van korte ruisjes hadden gelijke spectra, maar verschilden in ruisduur en pauzetijd tussen de ruisjes. Het experiment werd uitgevoerd met twee bandbreedtes, 3200 Hz en 200 Hz. De meetmethode was de experimentele luidheidsmatching procedure van het eerste experiment. Acht normaal horende proefpersonen namen deel aan het experiment.

Adaptive CAtegorical LOudness Scaling (ACALOS) van Brand en Hohmann (2002) werd gebruikt als luidheidsschalingprocedure. De traditionele procedure voor luidheidsmatching was een standaard adaptieve gedwongen keuze procedure met twee intervallen en twee alternatieven. De experimentele luidheidsmatching procedure was een variatie op deze traditionele procedure. De procedure was ontworpen om het matchen dicht bij het punt van gelijke luidheid makkelijker te maken door de taak te veranderen van het differentiëren van de luidheid van twee signalen naar het onderscheiden of er een verschil was in een signaalpaar of niet. We verwachtten dat met deze taak luisteraars minder hun beoordeling zouden verschuiven naar de intensiteit van comfortabele luidheid en het vaste referentie geluid makkelijker zouden negeren dan bij de originele taak.

In experiment 1 lieten alle drie de procedures duurafhankelijke luidheidssommatie zien, zoals ook werd gevonden door Brand en Hohmann (2002). De hoeveelheid luidheidssommatie was kleiner voor de experimentele luidheidsmatching procedure dan voor de traditionele luidheidsmatching procedure. De luidheidsschaling procedure liet zien dat de duurafhankelijkheid van de luidheidssommatie het sterkste was in de midden intensiteiten en klein leek te zijn bij lage en hoge intensiteiten.

De resultaten van het tweede experiment duidden erop dat de reeks korte ruisjes minder intensiteit nodig hadden om even luid te klinken als de continue ruis. Het hoofdeffect van bandbreedte was dat de effecten van de temporele structuur op luidheid ongeveer 1 dB sterker was voor de 3200 Hz bandbreedte dan voor de 200 Hz bandbreedte. Onverwacht verloren de verschillen tussen de verschillende condities hun significantie wanneer de resultaten werden weergegeven in de effectieve waarde van het totale signaal. Als sterkere spectrale luidheidssommatie voor korte signalen een toegenomen luidheidsperceptie voor de reeks korte ruisjes veroorzaakt dan valt te verwachten dat de verschillen in spectrale luidheidssommatie tussen de twee bandbreedtes groter zou zijn in de condities met korte ruisduur dan in condities met een langere ruisduur. Echter dit effect kwam niet naar voren in de resultaten.

In hoofdstuk 3 onderzochten we temporele effecten bij zowel drempelniveaus als bovendrempelige niveaus. In de context van onze focus op de luidheid van dynamische signalen is het belangrijk om te weten hoe veranderingen in signaalintensiteit luidheidsbepalingen beïnvloeden. De ultieme verandering in signaalintensiteit is het begin van een signaal. “Overshoot” ook wel “het temporele effect” genoemd, refereert naar de bevinding dat een hogere signaal-maskeerder ratio nodig lijkt te zijn om een kort signaal te detecteren tijdens het begin van een maskeerder dan als het is aangeboden met een vertraging, of als het wordt voorafgegaan door een ander geluid (precursor conditie). Er is toenemend bewijs, dat toont dat overshoot voor een groot deel verklaard kan worden door veranderingen in de cochleaire versterking. Luidheidsmodellen gaan echter uit van een stabiele versterking gedurende de duur van een signaal.

Het niveau benodigd om een kort 4 kHz signaal te detecteren werd gemeten voor signalen met verschillende starttijden relatief tot het begin van een 300 ms breedbandige maskeerruis: 100 ms en 5 ms voor het begin van de maskeerder en 5 ms en 100 ms na het begin van de maskeerder.

Luidheidsvergelijkingen tussen het signaal in stilte en het signaal op dezelfde vier starttijden werden verkregen voor vijf intensiteiten van het korte signaal en voor drie maskeerniveaus. Het temporele effect werd gedefinieerd als het intensiteitsverschil tussen signalen dichtbij het begin van de maskeerder en signalen ruim voor of na het begin van de maskeerder, benodigd om de drempel te bereiken en/of het bereiken van gelijke luidheid. Zowel bij de drempel als op bovendrempelige intensiteiten werden temporele effecten gevonden, die in overeenstemming waren met een afname van de versterking op de maskeerfrequentie gedurende de duur van de maskeerder. Het temporele effect was niet beperkt tot simultane maskering, maar werd ook gevonden bij condities met backward masking. In beide gevallen waren de temporele effecten sterker op bovendrempelige niveaus dan op drempelwaarden. Dit zou veroorzaakt kunnen worden door een overgangseffect bij het begin van de maskeerder. De bijna gelijktijdige aanvang van het signaal en de maskeerder maakte het moeilijk voor proefpersonen om het signaal van de maskeerder te scheiden, met name wanneer het signaalniveau dichtbij de gemaskeerde drempel lag.

Hoortoestelaanpasregels in hoorrevalidatie richten zich normaalgesproken vooral op detectiedrempels. Echter bij geavanceerde hoortoestellen is de voorgeschreven versterking normaalgesproken niet-lineair. Om een geschikte niet-lineaire versterking voor te schrijven is informatie over de luidheidsopbouw nodig. Een meetprocedure die informatie kan geven over luidheidsopbouw is luidheidsschaling.

In hoofdstuk 4 werd de toegevoegde waarde van luidheidschaling voor klinische toepassingen onderzocht in drie studies. In de eerste studie werd luidheidschaling uitgevoerd door een groep musici met normaal gehoor en licht gehoorverlies. Luidheidsschaling werd gemeten met twee smalbandige signalen (1/3-octaaf rond 750 Hz en 3 kHz) en een breedbandig signaal en de relatie met de audiometrische drempel werd onderzocht. De resultaten van de eerste studie toonden dat het moeilijk is om een eenduidige normale luidheidsfunctie te definiëren. De individuele spreiding in luidheidsfuncties binnen de normaalhorende proefpersonen was groot. De correlaties tussen de audiometrische drempels en de drempels op basis van luidheidschaling waren ook zwak. De correlatiecoëfficiënten tussen audiometrische drempels en drempels gebaseerd op luidheidschaling namen toe als er gehoorverlies aanwezig was. Dit was een logisch gevolg van het steiler worden van de luidheidsfunctie bij lage intensiteiten in proefpersonen met een cochleair gehoorverlies, wat leidde tot een nauwkeurigere schatting van de drempel in de luidheidschalingsdata.

De groepsresultaten leidden tot een andere interessante constatering. Gehoorverlies in deze populatie beïnvloedde vooral het onderste deel van de luidheidscurve. Het bovenste deel van de luidheidsfunctie was relatief onafhankelijk van zowel audiometrische drempels als van drempels verkregen uit luidheidschaling. De resultaten in deze studie suggereren sterk dat het verschijnsel recruitment beperkt is tot lage en medium intensiteiten. Het bovenste deel van de luidheidscurve toont geen consistente toename van de steilheid van de luidheidscurve met een toenemende drempel en dus geen recruitment. Dit suggereert dat voor normalisatie van luidheid compressie met name moet worden toegepast voor de lagere intensiteiten met lineaire versterking voor hoge intensiteiten, omdat anders de vorm van de luidheidsfunctie afwijkt van normaal.

In een tweede studie werd het verschil tussen monaurale en binaurale luidheidsperceptie onderzocht in een subgroep van musici. De resultaten van de tweede studie lieten een duidelijk verschil zien tussen monaurale en binaurale metingen. Binauraal aangeboden signalen werden duidelijk luider waargenomen dan monauraal aangeboden signalen. Dit effect trad vooral op bij de hogere intensiteiten. Op lage niveaus werd nauwelijks binaurale sommatie gevonden. De binaurale luidheidsdata kunnen gevolgen hebben voor hoortoestelaanpassingen. Bij hoortoestelaanpassingen is luidheidsnormalisatie vaak één van de hoofddoelen. Echter, hoortoestelaanpassingen worden bij Real Ear Measurements (REM) normaalgesproken monauraal geëvalueerd en niet binauraal. Als binaurale luidheidssommatie inderdaad niveauafhankelijk is dan zou dit meegenomen moeten in de wije waarop hoortoestellen worden aangepast.

Ten slotte, in de derde studie beschreven in **hoofdstuk 4** werden de correlaties tussen zelf gerapporteerde problemen en resultaten verkregen uit luidheidsschaling onderzocht in een andere groep van slechthorende werknemers. Deze studie werd uitgevoerd in een zeer heterogene groep proefpersonen. Daarom mogen er geen sterke conclusies getrokken worden uit deze studie. In deze studie werden luidheidsmetingen in het vrije veld vergeleken met resultaten van een vragenlijst over luisterinspanning. De ratio tussen m_{low} en m_{high} (Welke de concaafheid van de luidheidsfunctie bepaalt) vertoonde een zwakke correlatie met luisterinspanning in achtergrondruis: een hogere ratio correspondeerde met meer luisterinspanning. De relatie tussen lineariteit van de luidheidsfunctie en luisterinspanning in achtergrondruis kwam ook naar voren in de metingen zonder hoortoestel, suggererend dat het effect niet veroorzaakt werd door een inadequate hoortoestelaanpassing. Als de relatie tussen hoge luisterinspanning en erg concave luidheidsfuncties kan worden bevestigd in vervolgstudies dan kan deze kennis belangrijke gevolgen hebben voor hoortoestelaanpassingen. In dat geval zouden extreem concave luidheidsfuncties vermeden moeten worden. Dit betekent dat voorzichtigheid geboden is bij het gebruik van sterke compressie.

In **hoofdstuk 5** werd gebruik gemaakt van categorale luidheidsschaling (Brand en Hohman, 2002) om na smalbandige luidheidcompensatie spectrale en binaurale luidheidssommatie te meten. In een studie van Oetting *et al.* (2016) werden groepen slechthorenden met lichte en matig ernstige gehoorverliezen getest met toonaudiogrammen corresponderend met de audiometrische configuraties N1-N3 en S1 (Bisgaard *et al.*, 2010). Het is niet duidelijk of het effect van individuele spreiding afneemt, gelijk blijft, of zelfs toeneemt met toenemend gehoorverlies. Daarom werd in deze studie een grotere variatie van gehoorverliezen geïncludeerd (audiometrische configuraties: N2-N4 en S2-S3; Bisgaard *et al.*, 2010). De nadruk lag op het effect van de individuele spreiding in luidheidssommatie bij een grotere verscheidenheid aan gehoorverliezen volgens de classificatie van Bisgaard *et al.* (2010). De hoofdvragen in deze studie waren (a) of de vorm van het audiogram individuele verschillen zou kunnen verklaren en (b) of verschillende kenmerken van het gehoorverlies en van de compensatiestrategie voor het gehoorverlies mogelijke voorspellers waren voor de hoeveelheid spectrale en binaurale sommatie.

Voordat luidheidssommatie werd bepaald voor de breedbandige signalen (UEN1, UEN5, UEN17, and IGnoise) werden de ruizen voor elke slechthorende luisteraar afzonderlijk gecorrigeerd met als doel om geluidsniveaus te presenteren, die hetzelfde luidheidsniveau opleverden binnen elke band als voor de gemiddelde normaalhorende luisteraar (smalbandige luidheidscompensatie). Voor dit doel werden de breedbandige signalen gefilterd in zes niet-overlappende frequentiebanden met dezelfde centerfrequentie als de smalbandige signalen. De benodigde versterking voor elke frequentieband werd gedefinieerd als het verschil in intensiteit voor elke luidheidscategorie tussen de individuele luidheidsfuncties van de smalbandige signalen en de gemiddelde normale luidheidsfunctie.

In deze studie waren de effecten op de spectrale luidheidssommatie het grootst voor de signalen met de grootste bandbreedte, UEN17 en IFnoise (in overeenstemming met bijvoorbeeld Zwicker, 1958). In overeenstemming met de resultaten van Oetting *et al.* (2016) nam spectrale luidheidssommatie toe met de mate van gehoorverlies. Deze studie liet ook duidelijk een afhankelijkheid van gehoorverlies zien voor de spectrale sommatie van binauraal aangeboden signalen. Voor lichte gehoorverliezen (N2) was de spectrale luidheidssommatie voor binaurale geluiden hetzelfde als voor normaalhorende luisteraars. Voor de grotere gehoorverliezen (N3, N4 en S) neeg spectrale luidheidssommatie voor binauraal aangeboden geluiden groter te zijn dan normaal. Wat betreft de binaurale sommatie van breedbandige signalen zelf (binaurale condities versus monaurale condities) werd er geen duidelijke afhankelijkheid van gehoorverlies gevonden.

In **hoofdstuk 6** werd de luidheidsperceptie van laagdoorlaat gefilterde, hoogdoorlaat gefilterde en breedbandige roze ruis vergeleken in zestien slechthorende luisteraars voor zowel unilaterale als bilaterale aanbieding. De huidige hoortoestel voorschrijfregels gaan ervan uit dat spectrale luidheidssommatie afneemt in slechthorende luisteraars en dat binaurale luidheidssommatie onafhankelijk is van gehoorverlies en bandbreedte van het signaal. Voorgaande studies hebben aangetoond dat deze aannames incorrect zouden kunnen zijn.

Het experiment in **hoofdstuk 6** werd ontworpen om de validiteit te onderzoeken van de aanname dat binaurale luidheidsommatie hetzelfde effect heeft voor alle signalen. Met speciale aandacht voor de rol van de lage en de hoge frequenties in spectrale en binaurale luidheidsommatie. Voor dit doel werd de luidheid van een breedbandig signaal vergeleken met de luidheid van twee signalen, die ofwel de lagere frequentie componenten van het breedbandige signaal bevatten, ofwel de hogere frequentie componenten bevatten. De luidheid van deze drie signalen werd gemeten in een groep slechthorenden met behulp van luidheidschaling en luidheidmatching voor signalen, die unilateraal en bilateraal zijn aangeboden. De resultaten werden geanalyseerd voor spectrale effecten in unilaterale en bilaterale aanbieding, binaurale effecten en interindividuele verschillen.

Veranderingen in luidheidsperceptie bleken groter te zijn bij bilaterale aanbieding dan bij unilaterale aanbieding. In de groep slechthorenden, die in deze studie onderzocht zijn, leidde dit tot een grotere bijdrage van het laagfrequente deel van het signaal aan de totale luidheidsbeleving dan het hoogfrequente deel in bilaterale aanbieding vergeleken met unilaterale aanbieding. Dit leek een consequentie te zijn van het feit dat het laagfrequente deel van het spectrum meer binaurale luidheidsommatie gaf dan het hoogfrequente deel.

Ten tweede wordt bij het voorschrijven van hoortoestellen meestal geen rekening gehouden met individuele verschillen in luidheidsperceptie. Een belangrijke uitkomst van het huidige onderzoek was dat de individuele variabiliteit in luidheidsperceptie in de buurt van het niveau van onaangename luidheid zo groot was dat er rekening mee moet worden gehouden bij het aanpassen van individuele hoortoestellen. Informatie over de individuele luidheidsopbouw lijkt onmisbaar voor een goede hoortoestelaanpassing. Luidheidschaling geeft informatie over het gehele luidheidsbereik, maar luidheidsschaling is een vrij tijdrovende procedure. Het is misschien niet nodig om de luidheidsopbouw voor elk afzonderlijk signaal te meten, wanneer de luidheidsverschillen tussen signalen bekend zijn. Aangezien meettijd een belangrijke factor is in de kliniek, was een tweede doelstelling om te onderzoeken of luidheidsmatching zou kunnen worden toegepast als een betrouwbaar en tijdbesparend alternatief voor categorische luidheidsschaling. Dit lijkt het geval te zijn, omdat de individuele variabiliteit niet lijkt te worden beïnvloed door de meetprocedure (luidheidschaling, of luidheidmatching), of smalbandige luidheidscompensatie.

Overwogen zou moeten worden of het meten van individuele luidheidsperceptie voor breedbandige en binaurale signalen niet een vaste plaats verdient in de gereedschapskist voor hoortoestelverificatie. Welke set signalen de beste keuze zou zijn, is onderwerp voor verder onderzoek, maar luidheidsmatching lijkt een betrouwbare en tijdbesparende procedure te zijn om snel luidheidsverschillen te meten tussen signalen met verschillende bandbreedtes en aanbiedingsvormen (unilateraal vs. bilateraal).

De resultaten van de studies beschreven in dit proefschrift leiden tot zes algemene conclusies:

1) Spectrale luidheidssommatie is afhankelijk van de duur van het signaal. Deze duurafhankelijkheid blijkt een klein effect te hebben in reeksen van ruizen.
2) Er is aangetoond dat het temporele effect ook aanwezig is op bovendrempelige niveaus.
3) Het temporele effect is zelfs groter bij bovendrempelige intensiteiten dan op drempelniveau en is niet beperkt tot gelijktijdige maskering, maar kan ook kort voor het begin van de maskeerder worden gevonden.
4) Grotere gehoorverliezen en sterker aflopende gehoorverliezen geclassificeerd volgens Bisgaard *et al.* (2010) leiden tot een groter dan normale spectrale luidheidssommatie, zowel bij unilaterale als in bilaterale aanbieding.
5) Bij slechthorende luisteraars is de binaurale luidheidssommatie groter voor het laagfrequente deel van het spectrum van een breedbandsignaal dan voor het hoogfrequente deel.
6) De variabiliteit in luidheidsperceptie bij zowel normaal horenden als slechthorenden kan niet genegeerd worden bij luidheidnormalisatie. Daarom is de verwachting dat individuele luidheidsmetingen een positieve invloed hebben bij hoortoestelaanpassingen.

Dankwoord

Dankwoord

De basis van dit proefschrift is reeds gelegd in 2002, toen ik aangenomen werd op een combinatieplaats voor klinisch fysicus-audioloog in opleiding en promovendus op het AMC te Amsterdam. Toen ik in 2008 mijn opleiding tot klinisch fysicus-audioloog afrondde en ik alleen de artikelen behorend bij hoofdstuk 2 en 4 van dit proefschrift gepubliceerd had gekregen, had ik niet verwacht dat dit proefschrift er ooit nog zou komen. In de jaren daarna heeft mijn promotieonderzoek ook een aantal jaren volledig stilgelegen. Het kantelpunt kwam eind 2013. Bij de borrel na de promotieplechtigheid van mijn oud-collega Thamar van Esch benaderde mijn promotor Wouter mij met de vraag of ik interesse had om mee te helpen aan het Fit2Ears project van het AMC en de universiteit van Oldenburg. Dit project draaide om luidheidsschaling en daar had ik eerder in mijn promotieonderzoek ook al mee gewerkt. “Misschien,” zei Wouter, “kan je hiermee zelfs je proefschrift voltooien.” En zie, 10 jaar later is dat inderdaad gelukt!

Bedankt Wouter dat je mij indertijd gevraagd hebt om als onderzoeker deel te nemen aan het Fit2Ears project. Daarnaast bedankt dat je, in al die jaren die het heeft gekost om dit proefschrift te voltooien, in mij bent blijven geloven en mij bent blijven steunen. Zo ben ik ook dank verschuldigd aan Monique Boymans, die in het Fit2Ears project een belangrijke bijdrage heeft geleverd aan de ordening en analyse van alle meetgegevens. Mede door haar inzet kon ik op een rijdende trein springen. Bovendien heeft zij ondanks haar drukke baan bij Libra altijd tijd gevonden om de proefversies van mijn artikelen en later de hoofdstukken van mijn proefschrift door te nemen en te voorzien van commentaar. Dat is het geheel ten goede gekomen.

Die rijdende trein werd mede mogelijk gemaakt door Dirk Oetting en zijn collega’s in Oldenburg. Het tweede deel van mijn proefschrift is een aanvulling op hun fantastische werk, waarop ik mocht voortborduren. De software voor onze experimenten werd door hen ter beschikking gesteld en ook zij hebben mijn artikelen van bruikbaar commentaar voorzien. Ontzettend leerzaam waren ook de discussies tijdens gezamenlijke overleggen met de mensen van het AMC en Oldenburg in Assen.

Maar ook tijdens het werk aan het eerste deel van mijn proefschrift heb ik veel steun mogen ontvangen. Tegelijkertijd met Koen Rhebergen ben ik indertijd aan mijn promotieonderzoek begonnen. Jarenlang hebben wij een kamer op het AMC gedeeld en dat is mij altijd prima bevallen. Ook Bas Franck en Jan Koopman hebben in die eerste jaren veel voor mij betekend. Zij hebben mij veel geleerd over wat het is om onderzoek te doen en hoe je experimenten moet opzetten. En bij dat laatste was László Körössy ook onontbeerlijk. László heeft mij feitelijk leren programmeren. Met een engelengeduld heeft hij mij steeds weer verder geholpen als ik vastliep en ik kon altijd bij hem terecht met vragen.

Het soort onderzoek dat wij hebben gedaan, kan je niet met onderzoekers alleen. Voor psychofysisch onderzoek heb je proefpersonen nodig. In de loop van de tijd heb ik heel wat proefpersonen gehad: studenten van het AMC, muzikanten van orkesten in de omgeving en slechthorende vrijwilligers uit de klinische populaties van AC Tilburg en het AMC. Zonder hen had ik dit proefschrift niet kunnen schrijven. Dus ook aan al die anonieme proefpersonen, die tijd wilden steken in mijn onderzoek en hun best hebben gedaan om zich door de bij vlagen toch behoorlijk saaie experimenten heen te worstelen, dank je wel.

Al die proefpersonen heb ik overigens niet allemaal zelf gezien. Ik heb het geluk gehad dat bij verschillende experimenten collega's bereid waren om een deel van de metingen voor hun rekening te nemen. In dat kader wil ik mijn dank uitspreken aan Miranda Neerings, Mirjam van Geleuken en Addy Mols.

Onderzoek is alleen mogelijk als er financiering voor beschikbaar is. Daarom wil ik hierbij ook mijn dank aanspreken aan het Heinsius-Houboltfonds, dat een deel van mijn promotieonderzoek financieel heeft ondersteund.

Wetenschappelijk onderzoek kan bovendien niet zonder kritische beoordeling door andere onderzoekers. Daarom wil ik de reviewers van mijn verschillende artikelen bedanken. Hoewel het bij vlagen frustrerend was om weer een beoordeling met op- en aanmerkingen te ontvangen, is het de kwaliteit van mijn artikelen zonder meer ten goede gekomen. In dit kader wil ik ook prof. dr. W.J. Fokkens, prof. dr. J.C.M. Smits, dr.ir. P. Brienesse, prof. dr. ir. J.M.

Festen, prof. dr. P. van Dijk en dr. ir. J.A.P.M. de Laat hartelijk bedanken voor het beoordelen van dit proefschrift en hun bereidheid om deel te nemen aan de oppositie.

Ten slotte wil ik mijn gezin bedanken. In de loop der tijd heb ik een beetje leren relativeren, maar zeker in de beginjaren leverde een deadline rond een artikel of revisie regelmatig stress op en dat maakt me niet altijd gezelliger. Bedankt Ine, dat je me bent blijven steunen. En jongens, wat fijn dat jullie zo zelfstandig en flexibel zijn. Regelmatig ben ik op mijn papadag naar het AMC gegaan en was ik te laat terug om jullie van school te halen. Fijn dat jullie het prima vonden om even naar opa en oma te gaan of even alleen thuis te zijn. En papa, hoe vaak heb jij wel niet gevraagd hoe het met mijn proefschrift ging. Nou, hier is het dan eindelijk!

Curriculum Vitae & PhD Portfolio

Curriculum Vitae

Maarten van Beurden was born on March 23th 1978 in Hilvarenbeek and grew up in Baarle-Nassau. After graduation in 1996 from the Odulphus Lyceum (gymnasium) in Tilburg, he studied Applied Physics at the University of Twente in Enschede. His Master thesis was done in the group of Biofysical Techniques. In august 2002 he started working in the Academic Medical Centre in Amsterdam (now Amsterdam UMC). This position was a combination of a position as a Medical Physicist in Audiology in training at the ENT department and a position as a PhD student at the department for Clinical and Experimental Audiology. As a PhD student he conducted several psychophysical experiments on loudness perception under supervision of Prof. dr. ir. W.A. Dreschler.

In 2008 he completed his training as a Medical Physicist in Audiology and started as a Medical Pysicist in Audiology at Libra Revalidatie & Audiologie in Tilburg, where he still works. His work as a PhD student died out between 2008 and 2011 and was revived in 2015. Leading eventually to this thesis.

Publications

Journal Articles

- Van Beurden, M.F.B., Dreschler, W.A. (2018). Partial loudness at masker onset indicates temporal effects at supra-threshold levels. Hear Res. 370, 168-180.
- Van Beurden, M.F.B., Boymans, M., van Geleuken, M., Oetting, D., Kollmeier, B., Dreschler, W.A. (2018). Potential consequences of spectral and binaural summation for bilateral hearing aid fitting. Trends Hear. 22, 1-15.
- Van Beurden, M.F.B., Boymans, M., van Geleuken, M., Oetting, D., Kollmeier, B., Dreschler, W.A. (2021). Uni- and bilateral spectral loudness summation and binaural loudness summation with loudness matching and categorical loudness scaling. Int J Audiol. 60(5), 350-358.

Other publications

- Van Beurden, M.F.B., and Dreschler, W.A. (2005). Bandwidth dependency of loudness in series of short noise bursts. Acta Acustica united with Acustica, 2005, 91(1), 1020-1024.
- Van Beurden, M.F.B., and Dreschler, W.A. (2007). Duration dependency of spectral loudness summation, measured with three different experimental procedures. In Hearing - From Sensory Processing to Perception, edited by Kollmeier, B., Klump, G., Hohmann, V., Langemann, U., Mauermann, M., Uppenkamp, S., Verhey, J.L., (Springer, Berlin), pp. 237-245.
- Van Beurden, M.F.B., Boymans, M., Jansen, N., Dreschler, W.A. (2007). In Auditory signal processing in hearing-impaired listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR). Edited by T. Dau, J. M. Buchholz, J. M. Harte, T. U. Christiansen. ISBN: 87-990013-1-4.

PHD PORTFOLIO

Name PhD student: M.F.B. van Beurden
PhD period: 2002-2009 and 2015-2022
Name PhD supervisor: Prof. dr. ir. W.A. Dreschler

Courses	Year	Workload (ECTS)
Helmholtz Core Program for Audiology Seminars, workshops and master classes	2002-2005	3
Second Helmholtz Retreat 2003, Bergen	2003	1
Third Helmholtz Retreat 2005, Bergen	2005	1
Practical Biostatistics, Amsterdam	2003	1
Sensory Systems, Utrecht	2004	1.5
Beeldvormende technieken, medische diagnostiek voor fysici, Amsterdam	2004	1
Medische Ethiek, Rotterdam	2006	1
Good Clinical Practice, Amsterdam	2007	1
Objectieve audiometrie, Leiden	2008	1
“Management en Organisatie”, Rotterdam	2008	1.5

Presentations	Year	Workload (ECTS)
Loudness perception; methods, models, and new concepts Second Helmholtz retreat 2003	2003	0.5
Loudness summation as a function of signal duration Scientific meeting Werkgemeenschap Auditief Systeem (WAS)	2004	0.5

Bandwidth dependency of loudness of series of noise bursts. 8th Jahrestagung der Deutsche Gesellschaft für Audiologie	2005	0.5
Loudness perception; measuring procedures Third Helmholtz Retreat 2005	2005	0.5
Modeling of time dependent loudness summation XIVth International Symposium on Hearing	2006	0.5
Modeling of time dependent loudness summation Scientific meeting Werkgemeenschap Auditief Systeem (WAS)	2006	0.5
Partial loudness at signal onset, the temporal effect above threshold 20th Jahrestagung der Deutsche Gesellschaft für Audiologie	2017	0.5
Uni- en bilaterale spectrale luidheidsommatie en binaurale luidheidsommatie Nederlandse Vereniging voor Audiologie Wintervergadering 2020	2020	0.5

(inter)national conferences	Year	Workload
NVKF-conferences, Woudschoten, The Netherlands	2002-2021	5
LHCA-symposia, Amsterdam, The Netherlands	2016-2021	2
8th Jahrestagung der Deutsche Gesellschaft für Audiologie, Göttingen, Germany	2005	1
XIVth International Symposium on Hearing, Cloppenburg, Germany	2006	1

International Symposium on Auditory and Audiological Research, Nyborg, Denmark	2007	1
18th Jahrestagung der Deutsche Gesellschaft für Audiologie, Bochum, Germany	2015	1
Phonak European Pediatrics Conference, Berlin, Germany	2016	1
20th Jahrestagung der Deutsche Gesellschaft für Audiologie, Aalen, Germany	2017	1
International Conference for Adults, Frankfurt, Germany	2019	1
Hearing4All Symposium, Oldenburg, Germany	2020	0.2
8th International Pediatric Audiology Conference	2021	0.5
Virtual Conference on Computational Audiology (VVCA)	2021	0.2
Future of Audiology, Amsterdam, The Netherlands	2021	0.2

Other	Year	Workload (ECTS)
Scientific Meetings Werkgemeenschap Auditief Systeem (WAS)	2002-2005	2
Scientific Meeting Nieuwe Audiologen Nederland (NAN)	2002-2008	1
Nederlandse Vereniging voor Audiologie (twice a year)	2002-2021	3
ENT department Scientific Research Days	2002-2009; 2019	1