CARL VON OSSIETZKY universität OLDENBURG

Loudness models examined in the light of findings from loudness judgments and neural loudness correlates

Von der Fakultät für Medizin und Gesundheitswissenschaften der Carl von Ossietzky Universität Oldenburg zur Erlangung des Grades und Titels eines

Doktors der Naturwissenschaften (Dr. rer. Nat.)

angenommene Dissertation

von Herrn Florian Schmidt

geboren am 7. Dezember 1985

In Buxtehude

Gutachter: Prof. Dr. Dr. Birger Kollmeier
Weitere Gutachter: Prof. Dr. Jesko Verhey
Disputationsdatum: 23.09.2019

Abstract

Current loudness models for natural stimuli such as speech or music still provide predictions that differ from subjective loudness perception. This is also evident in practice, where simple level measures are often preferred for loudness evaluation.

Loudness models are based on the processing of the auditory system. While models can already reproduce the processing in the ear relatively well, neuronal processing is not well understood yet. However, cognitive effects seem to have a major influence on the perception of music. This thesis focuses on the evaluating current loudness models and finding approaches for modifications to improve their performance.

In the first part of the thesis, overall loudness judgements of music excerpts from different genres were collected by paired comparison. Level measures initially showed better predictions than loudness models. However, the loudness models could be significantly improved by specific pre-processing by applying a low-pass filter on the instantaneous loudness and including the psychoacoustic sharpness of music excerpts into the prediction procedure. Furthermore, a loudness transformation into categorical units showed a better representation of the loudness judgements.

In the second part of the thesis, neuronal loudness processing was investigated by electroencephalography. Two studies were carried out for this purpose. In the first study, a correlation between the amplitude of cortical potentials and modelled loudness of a music excerpt was found. In particular, Sone-loudness correlated better with early potentials, whereas categorical loudness correlated better with later potentials. In the second study a paradigm was implemented to study correlations between neurophysiological responses and contextual loudness effects that are associated with central processing rather than peripheral processing. The cortical correlates found in the previous study showed a significant correlation with loudness.

This thesis has shown that loudness judgements and their neuronal representation reflects a complex hierarchical process with neurosensory processing steps as well as more central adaptation and recalibration processes that will have to be incorporated into more sophisticated loudness models of the future.

Zusammenfassung

Aktuelle Lautheitsmodelle liefern für natürliche Stimuli wie Sprache oder Musik noch immer abweichende Vorhersagen von der subjektiven Lautheitswahrnehmung. Dies zeigt sich auch in der Praxis, wo für die Lautheitsbewertung einfachen Pegelmaßen oft der Vorzug gegeben wird.

Lautheitsmodelle orientieren sich an der Verarbeitung des auditorischen Systems. Während die Prozesse der Verarbeitung im Ohr bereits relativ gut modelliert sind, ist das Wissen über Prozesse der neuronalen Verarbeitung eher schlicht. Allerdings scheinen gerade kognitive Effekte einen großen Einfluss auf die Wahrnehmung von Musik zu haben. Im Kern dieser Arbeit wird die Qualität aktueller Lautheitsmodelle auf den Prüfstand gebracht, um danach nach Möglichkeiten zu schauen, auf welchen Stufen diese Modelle für bessere Vorhersagen modifiziert werden können.

Im ersten Teil der Arbeit wurden Gesamtlautheitsurteile von Musikausschnitten unterschiedlicher Genres durch einen Paarvergleich erhoben. Pegelmaße zeigten zunächst bessere Vorhersagen als Lautheitsmodelle. Jedoch durch gezielte Vorverarbeitung durch Tiefpassfilterung der instantanen Lautheit und Berücksichtigung der psychoakustischen Schärfe der einzelnen Musikausschnitte, konnten die Lautheitsmodelle deutlich verbessert werden. Außerdem konnte gezeigt werden, dass eine Lautheitstransformation in kategoriale Einheiten die Lautheitsurteile besser abgebildet hat.

Im zweiten Teil der Arbeit wurde die neuronale Lautheitsverarbeitung durch Elektroenzephalographie untersucht. Hierfür wurden zwei Studien durchgeführt. In der ersten Studie wurde eine Korrelation zwischen der Amplitude kortikaler Potentiale mit modellierter Lautheit eines Musikausschnittes gefunden. Insbesondere zeigte sich, dass Sone-Lautheit besser mit frühen Potentialen korrelierte, wohingegen kategoriale Lautheit besser mit späteren. Zur weiteren Untersuchung der neuronalen Lautheitskorrelate wurde in der zweiten Studie ein Paradigma entworfen, deren Ursache eher in der zentralen Verarbeitung anstatt der peripheren verortet wird. Die in der vorherigen Studie gefundenen kortikalen Korrelate zeigten auch hier einen signifikanten Zusammenhang mit der Lautheit.

Diese Arbeit hat gezeigt, dass Lautheitsurteile und ihre neuronale Repräsentation einen komplexen, hierarchischen Prozess darstellen mit neurosensorischen Verarbeitungsschritten sowie zentraleren Anpassungs- und Rekalibrierungsprozessen, die in Lautheitsmodellen der Zukunft integriert werden müssen.

Contents

Abbreviations

ABR	Auditory brainstem response
AFC	Alternative forced choice
AEP	Auditory evoked potential
ANOVA	Analysis of variance
ASSR	Auditory steady state response
BTL	Bradley-Terry-Luce method
dBA	A-weighted decibel
dBB	B-weighted decibel
DINA1	DIN 45631 / A1 standards of Zwicker Loudness Model for instationary sounds
DLM	Dynamic loudness model
DLMext	Extension of the DLM
EBU R	European Broadcast Union - recommendation
EEG	Electroencephalography
EFR	Envelope following response
ERB	Equivalent rectangular bandwidth
ERP	Event related potential
EPSP	Excitatory postsynaptic potentials
FFR	Frequency following response
FFT	Fast Fourier transform
fMRI	Functional magnetic resonance imaging
iFFT	Inverse fast Fourier transform
ITU	International Telecommunication Union

MEG	Magnetoencephalography
MLR	Middle latency response
MMN	Mismatched negativity
RLB	Revised Low-frequency B-weighting
RMS	Root mean square
SPL	Sound pressure level
TVL	Time varying loudness model

1 Introduction

The energy density of an audible pressure wave in the direction of its propagation is called sound intensity. *Loudness* is its perceptual correlate. While for physical quantities there are more or less precise ways to measure and to model data, perceptual quantities like loudness are much harder to deal with. The reason is that in most cases measurements are only feasible by asking subjects about their inferred sensory impression. Nevertheless, models of loudness perception can be realized by imitating the essential stages of physiological processing of sound along the auditory pathway.

First attempts were made in Weber-Fechner's logarithmic law - which is one of the most important laws of psychophysics - by considering the relationship between stimulus and sensation. Stevens (1957) developed a similar transfer function (Steven's power-law) which, however, reflects the transformation from intensity to loudness. Related to this is a loudness growth function with the unit 'Sone' whose parameters are obtained by psychoacoustic measurements using the ratio-scaling method for magnitude estimation. An alternative loudness growth function can be derived from categorical loudness scaling, as proposed by Heller (1985). Categorical loudness scaling determines the loudness with the unit 'CU' in the full auditory dynamic range in terms of categories like 'soft' and 'loud' as a function of the sound level. There are conceptual differences between the two scaling methods that historically led to the formation of opposing groups (Hellbrück, 1993). Nevertheless, both loudness functions provide reasonable loudness scales that differ only in the transformation of low and high intensities and in the degree of compression (Launer, 1995; Heeren *et al*., 2013).

The spectral properties of the sound have also a major impact on loudness perception. The frequency-dependent contribution to loudness was clearly illustrated in the equal loudness contours by Fletcher and Munson (1933). By referring pure tones of different frequencies to a 1 kHz pure tone in dB SPL the unit 'phon' is derived (DIN ISO 226; 2006). Based on the equal loudness contours and the concept of loudness summation across frequencies, Zwicker (ISO 532B, 1975) developed a multiband model (DIN 45631) for stationary sounds that predicts the loudness of narrowband as well as broadband signals. Recently, loudness models were developed with improvements in dealing with non-stationary stimuli (Chalupper and Fastl, 2002; Glasberg and Moore, 2002; Rennies *et al.*, 2009). Furthermore, physiologically motivated models suggest modeling the inner ear mechanics in more detail, involving a serial processing as realized in a transmission line model for the cochlea (Epp *et al.*, 2010; Pieper *et al.*, 2016).

Despite these considerable efforts and progress in improving loudness models, in practice, they have not yet replaced simpler level measures. This is clearly illustrated by the fact that in almost every sound device the loudness adjustment is based on dB increments. Beyond that, there are some areas where the loudness is mainly projected from levels. For example, loudness of traffic noise is treated by using the equivalent continuous sound level in dB(A). Another example occurs in the treatment of loudness prediction of music where newly developed level measures were commercially used by broadcast stations (e.g. EBU R-128, 2014)

supported by recent studies that have shown that there are deficiencies of loudness models predicting the loudness of music (Skovenborg and Nielsen, 2004; Vickers, 2010a). Fastl *et al*. (2003) showed that the dB(A) level cannot discriminate loudness differences between railway noise and road traffic noise (also known as 'railway bonus', i.e. the preference of railway noise to road traffic noise at the same A-weighted energy-equivalent level). With their study, they also raise the question whether such preference phenomena are due to loudness differences in music. Therefore, the research of loudness perception of music is of particular interest.

However, it is a rather complex task to define the term music in a few words. There are many kinds of sounds that are collectively referred to as music, like various interactions of running speech respectively singing voices, harmonic, percussive and synthetic instruments, everyday noises and various sound samples. Furthermore, the broad spectrum of sound scenarios requires the consideration of various parameters that determine the loudness. These include spectral and temporal loudness integration as well as effects caused by amplitude modulation. Moreover, it is suggested that for music there are also effects at higher stages of auditory processing that affect the loudness judgment, e.g. preferences (Cullari and Semanchick, 1989), pitch (Neuhoff *et al*., 1999), increased musical experience with age (Fucci, 1999), hearing expectation for different music genres (Barrett & Hodges, 1995), or context effects (Arieh and Marks, 2003a). The diversity of the sounds and the amount of loudness effects to be considered demonstrate the difficulty of modeling the loudness of music.

To analyze loudness effects at higher stages of auditory pathway, insights into the processing between the inner ear and the perceptual judgment would be beneficial. Therefore, a deep understanding of the neural processing of loudness would be necessary. Unfortunately, neural sound processing is not completely understood yet. However, in the past many studies suggested correlations between neural activity and loudness (Pratt and Sohmer, 1977; Hegerl *et al.*, 1994; Serpanos *et al.*, 1997; Langers *et al.*, 2007; Cai *et al.*, 2008; Soeta and Nakagawa, 2008; Röhl and Uppenkamp 2012; Behler and Uppenkamp, 2016). Neural activity can be studied by measuring quantities that correlate with it, such as the change of the electric field by electroencephalography (EEG). Studies about investigating correlations between loudness and such neural derived sizes are essentially done by comparing conditions in which only the loudness changes. A popular approach that is based on this idea is to search for a neural correlate of the loudness growth function. Therefore, neural representations are examined searching for a similar compression as the loudness transformation using stimuli with different acoustic levels (e.g. Ménard *et al.*, 2008; Florentine *et al.*, 2011; Silva and Epstein, 2010; Behler and Uppenkamp, 2016; Eeckhoutte *et al*., 2016). Evidence was found that loudness is reflected in the brainstem (Serpanos *et al.* 1997; Silva and Epstein, 2010), thalamus (Madell and Goldstein, 1972) and in parts of the auditory cortex (Hegerl *et al.*; 1994; Röhl and Uppenkamp, 2012; Behler and Uppenkamp; 2016). However, there is still some dispute whether these correlates reflect the sensorially processed intensity or already the loudness perception (Näätänen and Picton, 1987; Darling and Price, 1990; Hart *et al.*, 2002; Röhl and Uppenkamp, 2012).

There are numerous neuroscience studies on the perception of music, but hardly any of these studies deals with the perception of loudness. This is due to the fact that in most cases evoked neural responses to acoustic stimuli can only be sufficiently analyzed in subsequent silence. This complicates the analysis of continuous stimuli. However, it is well studied that music often elicits several features that are measurable by EEG or MEG, e.g. event related potentials (Besson and Macar, 1987; Patel *et al.*, 1998; Miranda and Ullman, 2007) or brain oscillations related to the spectral properties of the stimulus (Doelling and Poeppel, 2015).

The central question of this thesis is: Can current loudness models provide more reliable predictions for the loudness of music in comparison to level measures, and if so, at what stages should the models be modified in order to improve their predictions? In order to find indications about the stage in which the respective model should be modified it would be useful to identify neural representations of the loudness of music or loudness in general along the auditory pathway. Consequently, the question arises whether a respective neural representation of loudness is corresponding more closely to the sensorially processed intensity of the sound or rather to more final processing stages representing the perceived loudness.

In order to address this question, three aspects have to be clarified which determine the methodological procedure: (i) Which method should be used to measure loudness judgments? (ii) Which loudness models should be examined? (iii) Which measurement method should be used to investigate the neural processing of loudness? (iv) How should the relationship between the neural correlate and loudness be demonstrated?

(i) In assessing the loudness of music, the estimation of one value representing the overall loudness of the stimulus is of particular interest, e.g. in the broadcasting where different pieces of music should be set equally loud (Vickers, 2010b) or in medical research where the hearing loss is related to one equivalent continuous sound level (Gunderson *et al*. 1997). Scaling methods such as matching, magnitude estimation or categorical scaling provide their own scale. It is to be expected that models adapted to one scale will work less well for the other scale, e.g. sone-models will predict the loudness of categorically scaled stimuli worse than CU-models. Therefore, a paired comparison method was used which provides an interval scale of the overall loudness of respective music pieces using the Bradley-Terry-Luce model. Thus, level measures, sone-models as well as CU-transformed models can be assessed objectively. Moreover, due to the difficulty of the task of reducing the loudness of such a multi-faceted stimulus as music to one overall value, the paired comparison, which requires little effort, seems to be an excellent method. This method is used in the first study (Chapter 3) to compare loudness models and level measures.

(ii) In this thesis, the current standard loudness models are examined. Therefore, the recommendations of the International Organization for Standardization (ISO) are considered that propose two models for loudness evaluation of time-varying sounds: those based on measurements from Stevens, ISO 532A, and the Zwicker-based ISO 532B. These two models are also proposed in respective national standards: ANSI-S3.4 and DIN 45631 / A1. Apart from the DIN 45631 / A1, there are other Zwicker-based models: the Dynamic Loudness Model (DLM) by Fastl and Chalupper (2002) and its extension by Rennies *et al.* (2009). These models are compared with the standard level measures. The unweighted and A- and B-weighted level measures are used

as well as the recommendation of the european broadcasting union, the EBU R-128 standard, for level-based loudness evaluation of music. In Chapter 3, all these measures are used. The second study (Chapter 4) uses only the DLM, the A-weighted and unweighted sound pressure level and the EBU R-128 recommendation.

(iii) In studies on the neural activity of humans non-invasive monitoring methods are required. The most prominent representatives are functional magnetic resonance imaging (fMRI), electroencephalography (EEG) and magnetoencephalography (MEG). Each of these methods enhances different aspects in terms of temporal and spatial resolution. For the neural investigations of this thesis only EEG measurements were considered. EEG, as opposed to fMRI, provides a higher temporal resolution taking care of particularly complex time-varying sounds such as music. Of course, MEG also provides high temporal resolution and, in addition, high spatial resolution for the cortex. However, EEG requires a much simpler measurement setup, which initially favors the EEG over the MEG method. Finally, neural loudness research for EEG is most advanced comparing these three methods. Therefore, the choice of the EEG method appears to be the most plausible. The EEG method was used in Chapter 4 and Chapter 5 (third study).

(iv) In this thesis, two approaches are used to explain the relationship between loudness and its neural correlate. In the first approach, the (cortical) EEG response to an excerpt of a piece of music is investigated at different sound levels to find a representation of the loudness growth function. This approach is described in Chapter 4. In the second approach, the neural representation of loudness recalibration is examined which is assumed to be associated with central processing of loudness (Arieh and Marks, 2003a). To illustrate this relatively small loudness effect, a matching paradigm is used for a few conditions. A change in parts of the EEG response as the condition changes is an indication of a correlate of neural loudness processing. This approach is reflected in Chapter 5.

The thesis is divided into two major sections. The first section deals with research concepts and employed methods with the focus on the basics of loudness and EEG processing (Chapter 2). These topics are treated with special consideration to their relevance for music. The second section is divided into three chapters (Chapter 3-5). In this section, several experimental studies are described and discussed that have been designed to provide contributions to the two research problems stated above.

In Chapter 3 the prediction performance of level measures and loudness models is investigated by comparing these measures with psychoacoustic results from a scaling procedure. The aim is to improve the performance of loudness predictions and to understand where the deficiencies in sound processing of the models are. The purpose of the study in Chapter 4 is to expose the neural processing of the loudness of music. Therefore, the relationship between the loudness of a music stimulus and spectro-temporal features of the EEG response with regard to amplitude and latency is investigated, while also differences of the correlations with loudness models and level measures are shown. The study in Chapter 5 focuses on the neural representation of central loudness effects, more exactly: loudness recalibration, to address the question whether these neural correlates are related to sensorially processed intensity of the sound or rather to the loudness perception. Hence, the cortical EEG response is investigated during a psychoacoustical loudness recalibration experiment

with different inter-stimulus intervals. Finally, the last Chapter 6 gives a summary, some concluding remarks and an outlook to the intended future research in the field of loudness modelling of music.

2 Research concepts and methods employed

2.1 Loudness

Loudness is an auditory measure of perception and can be essentially defined as the perceived intensity of a sound. However, loudness is not just constituted by the transformation of sound intensity. The spectral content of the sound and its fluctuations over time strongly affect the processing in the auditory system. At the peripheral level, i.e. the outer ear, middle ear and inner ear, various transformation processes take place containing spectral shaping, forming a filter bank and amplitude compression. This processing of the sound intensity at early stages of the auditory system results in the specific loudness of respective frequency channels. Spectral and temporal integration arise at later stages.

There are three temporal types of loudness (Fig. 2.1) that are important for classifying the perception of time-varying sounds: (1) short-term loudness, (2) long-term loudness and (3) overall loudness. (1) The short-term loudness represents the momentary impression of the loudness of a short segment of sound. Hence it includes the fine structure of the loudness. (2) Long segments of sound are reflected by the long-term loudness. The fine structure of loudness becomes secondary while the average loudness of the segment is highlighted. (3) The overall loudness corresponds to a summary assessment of the entire experience profile converted from a stream of long-term loudness judgements. Several transformation processes in the outer, middle and inner ear result in a quantity called ‘instantaneous loudness’ that cannot be consciously perceived. At later stages in the auditory system temporal integration occurs, transforming instantaneous loudness into short-term loudness and long-term loudness.

The development of loudness models is intended to predict the loudness of sounds. Essentially, these models combine some contributing factors for loudness and apply them based on the simulation of series of processing stages in the auditory system. These contributing factors will now be considered in more detail. The following subchapters were mostly based on Zwicker and Fastl (1999), Florentine *et al.* (2011), Moore (2013) and Kießling *et al.* (2018). They give a good review of research in this field.

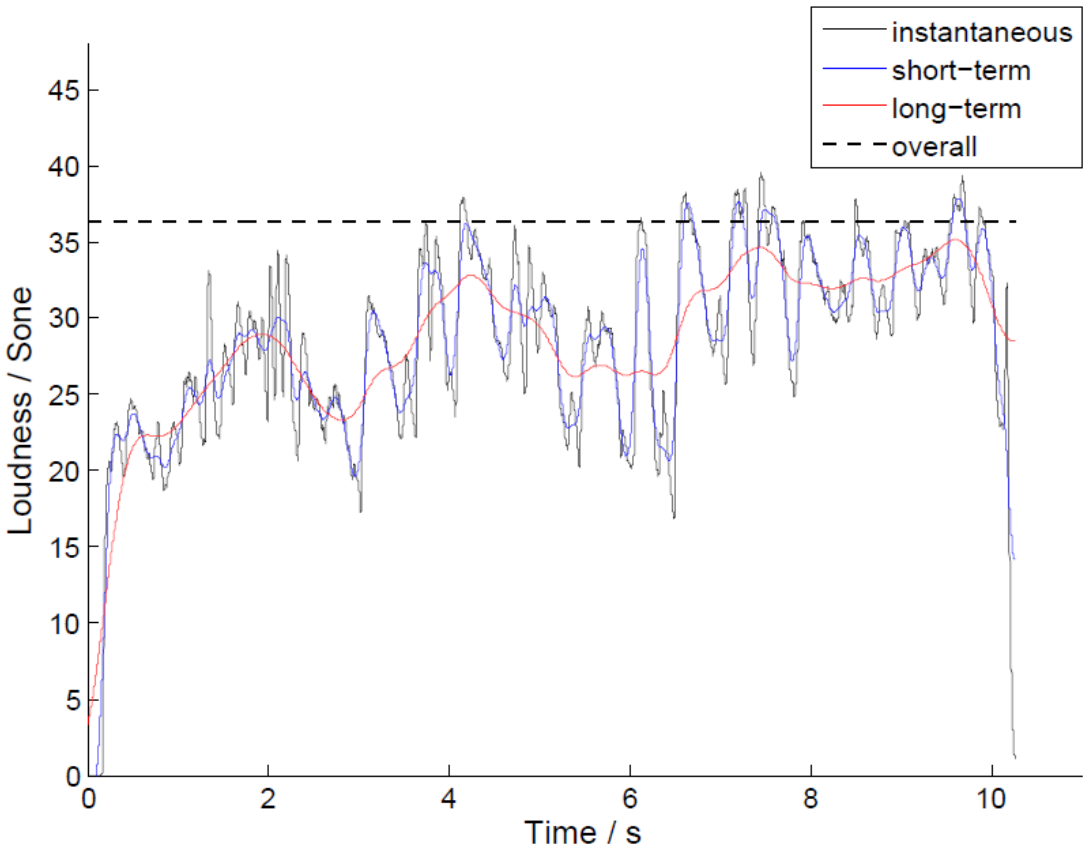

Fig. 2.1: Example of the temporal types of loudness and overall loudness. Instantaneous loudness is presumed to be the last form of processing before conscious perception. The short-term loudness represents the momentary impression. The long-term loudness rather reflects the average loudness of a sound segment. The overall loudness corresponds to a summary assessment of the entire experience.

2.1.1 Compression

The transformation of the physical magnitude of a sound stimulus into the perceived magnitude corresponds to a non-linear transfer function. The Weber-Fechner-law (2.1) is a well-known approximation to this transfer function and serves as a motivation of the sound level measures. The physical magnitude, the intensity *I*, is logarithmically transformed to perceived magnitude, the sensation *S*, with some scaling factor a, b.

$$S = \log(a \cdot I^b) \tag{2.1}$$

The corresponding setting of the parameters results in the representation of the sound intensity level L_I in dB (1.2):

$$L_I = 10 \cdot \log_{10}\left(\frac{I}{I_0}\right) \tag{2.2}$$

However, the sound intensity level is not a ratio scale. This means that, for example, a doubling of the level does not cause a doubling of the perceived loudness. In fact, the empirical 10-dB rule (2.4) tends to apply

where doubling of the loudness is caused by a level increase of 10 dB. In this rule, a ratio is transformed into a distance. This is owed to the logarithmic transformation of Eq. (2.1). Therefore, S. Stevens (1957) proposed an experiment in which the loudness of a sinusoidal tone was determined by absolute magnitude estimation (see section 2.3). The resulting functional relationship between estimated magnitudes and sound levels corresponds to a power law as a transfer function Eq. (2.3) with a scaling factor *k* and a compressive factor α.

$$S = k \cdot I^{\alpha} \tag{2.3}$$

If the exponent $\alpha \approx 0.3$ this transfer function provides a similar compressive characteristic as in Eq. (2.2). In this case, *S* represents the perceived loudness in Sone. Generally, *k* shall be such that, *S* = 1 Sone for a sinusoidal tone at 1 kHz and 40 dB SPL (see Fig. 2.2). The compression satisfies the 10-dB rule which can be seen by in Eq.'s (2.4) and (2.5).

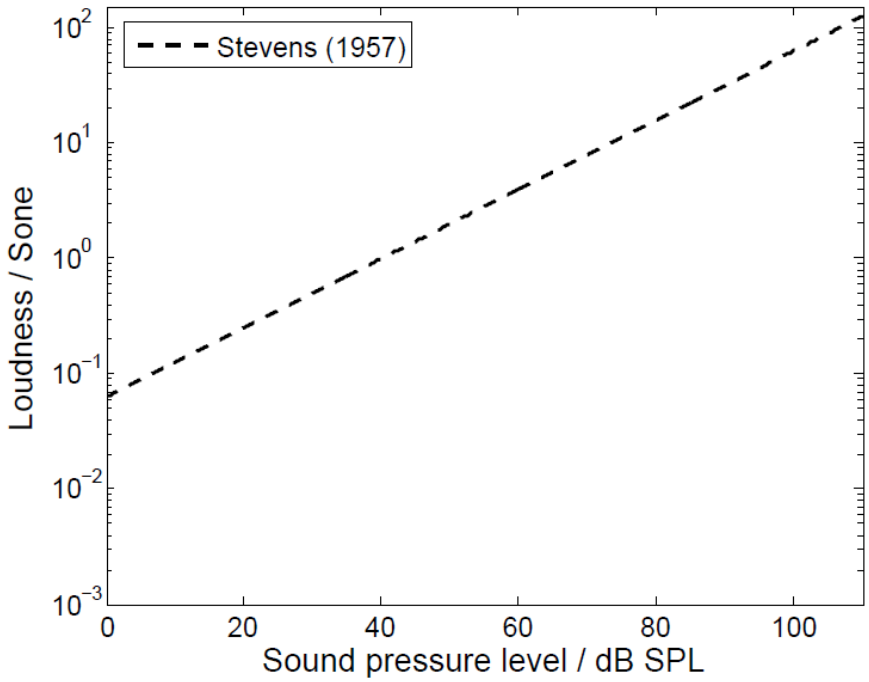

Fig. 2.2: The loudness function of a 1 kHz sinusoidal tone according to Stevens power-law, α =0.3.

$$2 \cdot S \triangleq 10 \cdot \log\left(\frac{I}{I_0}\right) + 10 = 10 \cdot \log\left(\frac{10 \cdot I}{I_0}\right) \tag{2.4}$$

$$k \cdot (10 \cdot I)^{0.3} = 10^{0.3} \cdot k \cdot I^{0.3} \approx 2 \cdot k \cdot I^{0.3} = 2 \cdot S \tag{2.5}$$

Doubling the loudness *S* can be achieved by a tenfold increase of the sound intensity *I*, i.e. the 10-dB rule Eq. (2.4). Substituting this intensity into Equation (2.3) the loudness is only twice as high, Eq. (2.5). A *compression* takes place. However, at absolute hearing threshold, the loudness should be zero, which is not met by Eq. (2.3). Moreover, it is questionable whether humans are really able to judge the ratios of loudness.

Another way to determine the loudness perception is to use categorical scaling (Heller, 1985; Brand and Hohmann, 2002), in which the loudness is scaled by categories (e.g. "very quiet", "quiet", "medium", "loud"

and "very loud"). In a second step, these rough categories are divided into finer categorical units (CU). Categorical scaling seems a bit more sensible than magnitude estimation, since it is not clear whether a person is able to transfer the perceived loudness to a ratio scale at all (Heller, 1985). Nevertheless, the transfer function from sound pressure level to CU-loudness is similar to the Sone scale. Deviations from the loudness in sone occur especially at very low and at very high levels (Launer, 1995). In Heeren *et al*. (2013), the difference in scaling is summarized by a transformation from sone to CU Eq. (2.6).

$$\mathrm{CU} = 2{,}6253 \cdot \lg(\mathrm{sone} + 0{,}0887)^3 + 0{,}7799 \cdot \lg(\mathrm{sone} + 0{,}0887)^2$$
$$+8{,}0856 \cdot \lg(\mathrm{sone} + 0{,}0887) + 13{,}4493 \tag{2.6}$$

2.1.2 Spectral effects

The loudness is highly dependent on the frequency. This is illustrated very clearly by the equal loudness contours (Fig. 2.3). They are an array of curves which represent equal loudness for sinusoidal tones of different frequencies and sound pressure levels. Each contour corresponds to a loudness value whose unit is represented in phon. An alternative approach to model the frequency dependence of loudness is possible by considering k and α in Eq. (2.3) as function of a frequency.

There are several causes of this frequency dependence. At early stages, the transmission of the signal from free-field through outer and middle ear has to be considered by a correction factor.

Further, it has been suggested that in the inner ear sound evokes a spread of excitation along the basilar membrane (Zwicker, 1958). The spectral content of the sound is represented by the resulting excitation pattern. From this consideration, physiologically motivated frequency scales were derived such as the Bark-scale or the equivalent rectangular bandwidth (ERB) scale. Traunmüller (1990) proposed a decent transformation of the frequency from Hz to Bark in Eq. (2.7). A transformation to ERB is proposed by Moore and Glasberg (1996) in Eq. (2.8). However, recent experiments showed that loudness models based on ERB scale (e.g. ANSI S3.4, 2007) overestimate the loudness of broadband sounds (Schlittenlacher *et al*., 2012).

$$f_{Bark} = \left(\frac{26.81 \cdot f_{Hz}}{1960 + f_{Hz}}\right) - 0.53 \tag{2.7}$$
$$f_{ERB} = 24.7 \cdot (4.37 \cdot f_{Hz} + 1) \tag{2.8}$$

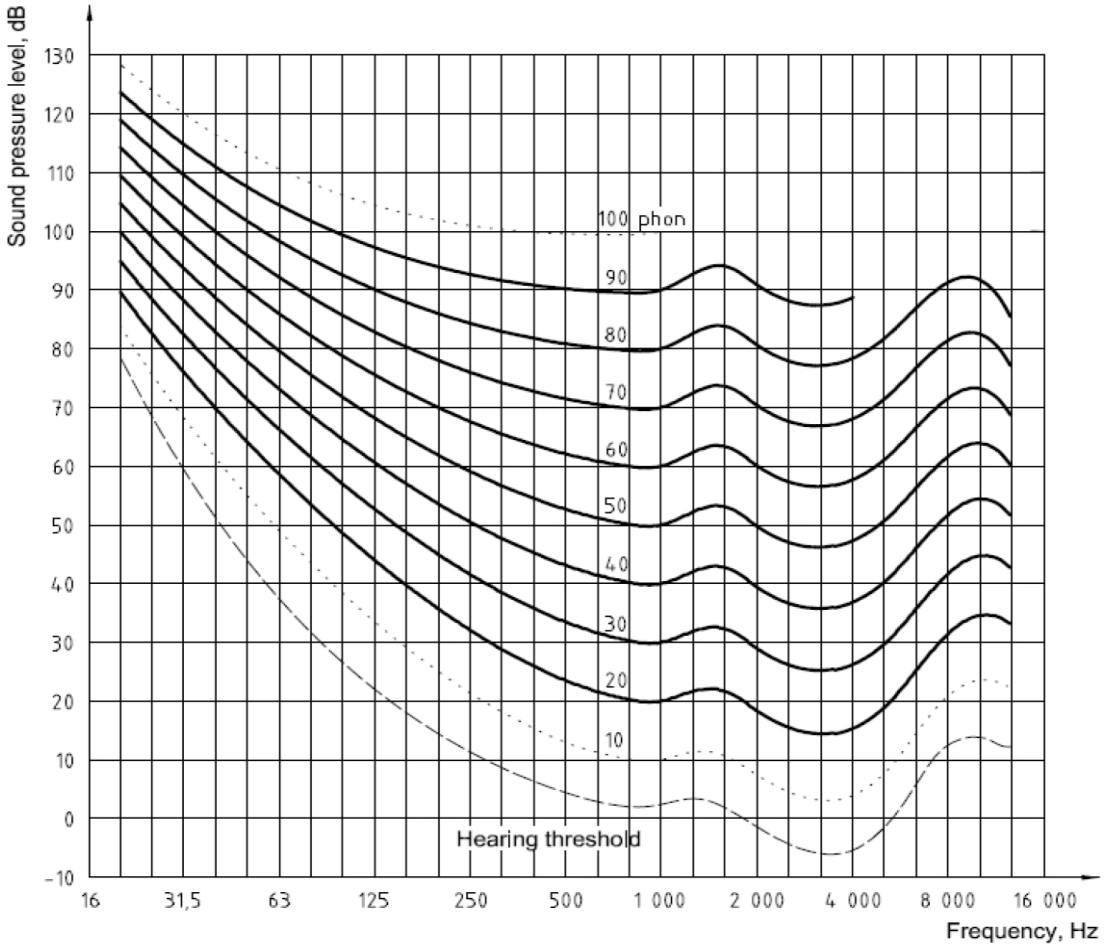

Fig. 2.3: The Equal loudness contours for loudness levels from 10 to 100 phons (ISO 226, 2003). The curves < 20 phon and at 100 phon (dashed) are based on interpolation and extrapolation.

The specification of the frequency scale is the concept of critical bandwidth. The incoming sound is separated in the inner ear in a bank of overlapping critical band filters that are adjusted corresponding to those physiological frequency scales. A nonlinear compression in each filter transforms the excitation into specific loudness. By integrating the specific loudness across frequencies total loudness can be derived. This relationship between bandwidth and loudness is called spectral loudness summation. As the bandwidth of a signal widens, the loudness increases for a constant total signal power. This can be demonstrated by considering the difference of the integrated loudness between a narrow band S_N and broadband S_B tone complex in Eq. (2.9). Due to the fact that a compression $\alpha < 1$ constitutes a concave function, the Jensen inequality applies (Rennies *et al.*, 2009).

$$S_B = \sum_{n=1}^{m} I^{\alpha} > S_N = \left(\sum_{n=1}^{m} I\right)^{\alpha}, \quad \alpha < 1. \tag{2.9}$$

Spectral loudness summation can also occur with amplitude modulation. If the spectral sidebands of the modulation signal lie in adjacent auditory filter, this effect occurs.

There is also a complex masking interaction of signals with different frequencies. It is known from everyday life that communication almost always takes place with background noises. When the spectral content of this noise coincides with the actual signal, masking takes place and the loudness of the signal decreases. This effect is called spectrally partial masked loudness. This masking even affects adjacent frequency ranges, i.e. when the spectral range of the masker is adjacent to that of the signal.

2.1.3 Temporal effects

Loudness takes some time to build up. If the exposure time of a stimulus is less than a certain critical value – around 100 ms – its loudness increases with duration (Scharf, 1978; Zwicker and Fastl, 1999). This effect is commonly known as temporal integration. The duration of a stimulus also affect the spectral loudness summation (Verhey and Kollmeier, 2001). They showed that the level difference between narrowband noise and equally loud broadband noise is larger for 10-ms bursts than for 1000-ms bursts.

Spectral partial masking also occurs temporally and is referred to as pre- and post-masking. A masker can influence the loudness perception of the previously presented content up to about 50 ms and the subsequent content up to about 150 ms.

2.1.4 Contextual effects

Equal loudness judgments may be evaluated as being very unequal under a different set of contextual conditions. Therefore these contexts seem to have biasing effects on the loudness judgment. If a preceded context determines the condition under which judgements are given this context has a preceding effect. An extreme case of such a preceding context effect is the temporary loudness shift. Continued high level sound exposure causes a decreased perception in the range of 20-40 dB with varying recovery times over at least several minutes (Hirsh and Ward, 1952). In this extreme case, more peripheral causes seem to be responsible for the altered loudness perception. Multiple mechanisms can be the cause of this loss of perception or at least can be interactively involved: a change in hair cell activity, a reduction of stereocilia rootlet length (Liberman and Dodds, 1987), a reduction in neural activity and temporary sensory-cell degradation due to the rapid production of metabolic waste products during increased activity. However, preceding context effects might also have a central origin caused by processes of judgment. In a study of Marks (1988) magnitude estimation was performed for the loudness of 500 Hz and 2.5 kHz tones. In one contextual condition, the 500 Hz tone was presented at low and the 2.5 kHz tone at high SPLs. In the other condition, these level settings were reversed (Fig. 2.4). It was demonstrated that the assessments of the low SPLs were overrated compared to the high SPLs. This effect induced by the context is called “loudness recalibration” by Arieh and Marks (2003a). It has remarkable consequences for the measurement of equal loudness contours with the usual matching methods, since the interval of the level range produces context effects that generate biases (Gabriel, 1996). Loudness recalibration does not happen instantaneously but needs some time for

temporal adaptation. Arieh and Marks (2003a) showed that this adaptation process can last over 2 seconds and make up 10 dB of differences in loudness perception. Contextual biasing also happens when the preceding information exceeds the subsequent one**,** also known as primacy effect (Fiebig and Sottek, 2015)**.** For example, this effect was studied by Ponsot *et al.* (2013) by examining the loudness of a series of 1 kHz pure tone segments with similar length of 125 ms but varying in sound level. They found a primacy effect for flat sound level profiles.

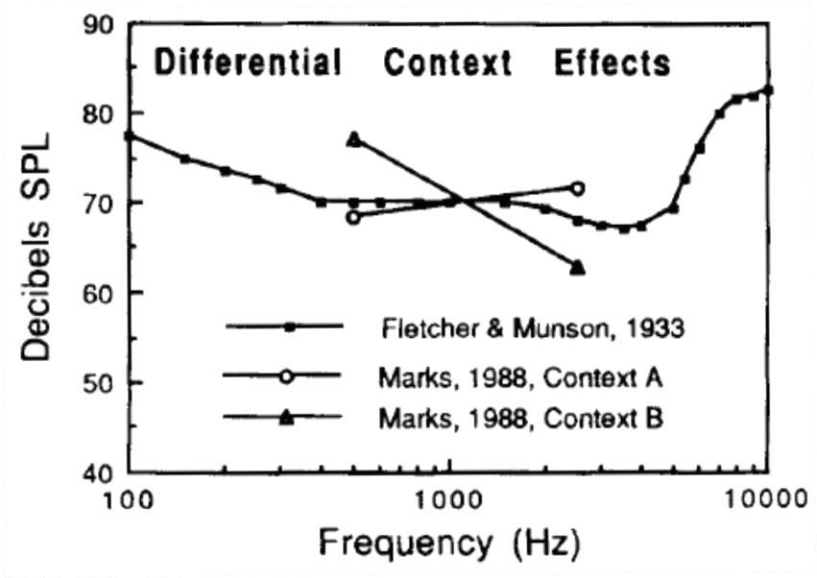

Fig. 2.4: Context effect compared to the corresponding equal loudness contour (see Fig. 2.3); the perceived loudness of 500 Hz and 1500 Hz tones with different contextual sets of stimuli (Marks, 1993).

Similar observations have been made by other studies (Susini *et al.*, 2002; Oberfeld and Plank, 2005; Pedersen and Ellermeier, 2008). However, an opposite effect is also well known, where the subsequent information exceeds the preceding one**.** This subsequently contextual biasing is called recency effect and it is often allocated to a cognitive distortion of memory. Höger *et al.*, 1988 found that the recency effect could significantly bias loudness assessments.

2.1.5 Cognitive processing

At the highest stages of auditory processing, sounds are consciously perceived and reflected. Therefore, it is associated with cognitive processing. Due to the complexity, this processing stage has only an effect on the overall judgments or at least long-term judgements. On the other hand, short-term judgments happen too spontaneously as impressions on this level of consciousness can be included quickly enough. At this level, different perceptual parameters are reflected in parallel, which can lead to confusion and mixing of them. It is suggested that for music cognitive processing may considerably affect the loudness judgment. There are several perceptual or psychological parameters that have been shown to interfere with loudness. Some of them, which are related to music, are to be mentioned now.

Cullari and Semanchick (1989) investigated whether or not the loudness of music is affected by how much music is subjectively liked, i.e. preference, and found a negative correlation between loudness and preference. Similar findings were made by Kuwano *et al.* (1992) when they investigated the relationship between loudness and annoyance, which is nearly the counterpart to preference. They found a positive correlation between loudness and annoyance for a number different stimuli (music, speech, traffic noise,…). Toepken and Weber (2013) demonstrated in an elaborated measurement method the discrimination between influence of loudness and preference in the case of multi-tone stimuli. They were able to reduce the shared variance of both parameters from 35% to 8%. Barrett and Hodges (1995) even went one step further. They identified differences in the preferred sound levels for several music genres. For example, subjects preferred to listen to heavy metal at 92.9 dB while country music was preferred at much lower sound levels of 73.8 dB. This shows that there are specific hearing expectations in loudness of different music genres, which finally could influence the loudness judgments of music sounds.

Music is also received in different loudness depending on age (Barrett and Hodges, 1995) but furthermore, this age dependency applies also directly to the loudness judgement (Fucci, 1999). Fucci (1999) investigated how different age groups estimated the loudness of rock music which was presented at different sound levels. Older subjects perceived rock music louder at each level than younger subjects. However, at the highest presented levels (80-90 dB) the children surpass the older subjects and showed the highest loudness estimation.

Changes in auditory pitch are not always easy to distinct perceptually from changes in loudness. For example, the perceived pitch of an approaching train can rise due to the appearing rising dynamic intensity change while the observed frequency actually falls (Neuhoff and McBeath, 1996). The rising dynamic make it difficult to accurately track the falling frequency. Hence, pitch and loudness interact under dynamic conditions. Neuhoff *et al.*, 1999 confirmed this result and presumed that this interaction is centrally processed in the auditory system. It may be an analytic process and has derived from the benefit of recognizing naturally occurring covariation of frequency and intensity. Regardless of this interpretation, the interaction between pitch and loudness could obviously complicate the loudness judgment while listening to music.

2.2 Loudness measures

All loudness measures are designed to estimate the loudness perception (often) by imitating the physiology of hearing. However, the mathematical complexity of these measures can be very different.. In the following the functioning of different level measures and loudness models is presented. Here we use the term 'level measure' in contrast to ‘loudness model’ rather than the term 'single-band model' versus 'multi-band model' as it is frequently found in the literature (Skovenborg and Nielsen, 2004; Vickers, 2010a). The common single-band models have no spectral signal separation and are unlike loudness models "less" physiologically motivated, but show a successful input-output characteristic. Furthermore, they are all based on logarithmic level representation, which is why we prefer this choice of conceptual terms.

2.2.1 Level measures

Basically, all the level measures share a logarithmic compression as described in section 2.1.1. Another common feature is the use of frequency weightings. In general, these are frequency weightings that attenuate the low-frequency part of the spectrum (Fig. 2.5). This is the region where the hearing is least sensitive, particularly concerning sounds below 100 Hz. For some frequency weightings, high frequency regions are also attenuated so that only the region is amplified, where the hearing is most sensitive (around 1-4 kHz). The frequency dependent sensitivity of hearing is also affected by the absolute levels of the sound. Therefore, the frequency weighting is generally dependent on the range of the sound pressure level. The most popular frequency weightings are modeled on the (historic!) equal loudness contours from the old Fletcher & Munson curves (cf. section 2.1.2). The A-weighting corresponds to the 30-phon contour, that means to soft sounds. The B-weighting corresponds to the 70-phon contour and at high levels where all frequencies contribute more or less equally to the loudness sensation the C-weighting is usually considered. However, there are several other frequency weightings that were developed to serve other purposes, e.g. several perceptual measures (loudness, annoyance, etc.) of a certain category of sounds. For the loudness of music the Revised Low-frequency B-weighting (RLB) (Soulodre and Norcross, 2003) is frequently recommended (Skovenborg and Nielsen, 2004; Vickers, 2010a; Ponsot *et al.*, 2016) and is used in the ITU (ITU-R BS.1770-2, 2011) and EBU standards (EBU R-128, 2014).

Level meters imply a common approach for short-/long-term loudness prediction which is based on long-term energy integration, the equivalent sound level *Leq* in dB (Eq. 2.10). It allows an envelope extraction of the signal. Basically, Equation (1.9) is just another expression for the Root-Mean-Square level (RMS).

$$Leq(t) = 10 \cdot \log_{10}\left(\frac{1}{T}\int_0^T \left(\frac{p(t)}{p_{ref}}\right)^2 dt\right) \tag{2.10}$$

The time constant *T* represents the time interval of interest and it also corresponds to low-pass filtering. The commonly used time constants are the ’fast’ weighting, *T* = 125 ms, and the ‘slow’ weighting, *T* = 1 s. The gradual integration of the sound signal generates a low-pass filtered signal *Leq*(*t*) .

Beyond these similarities in the processing of the level measures, there usually follow further processing steps, serving to estimate the overall loudness. *Leq*(*t*) can be described by the distribution of dB values across time. The peak (RMS-peak) of such a distribution is often used as the representative of the overall level.

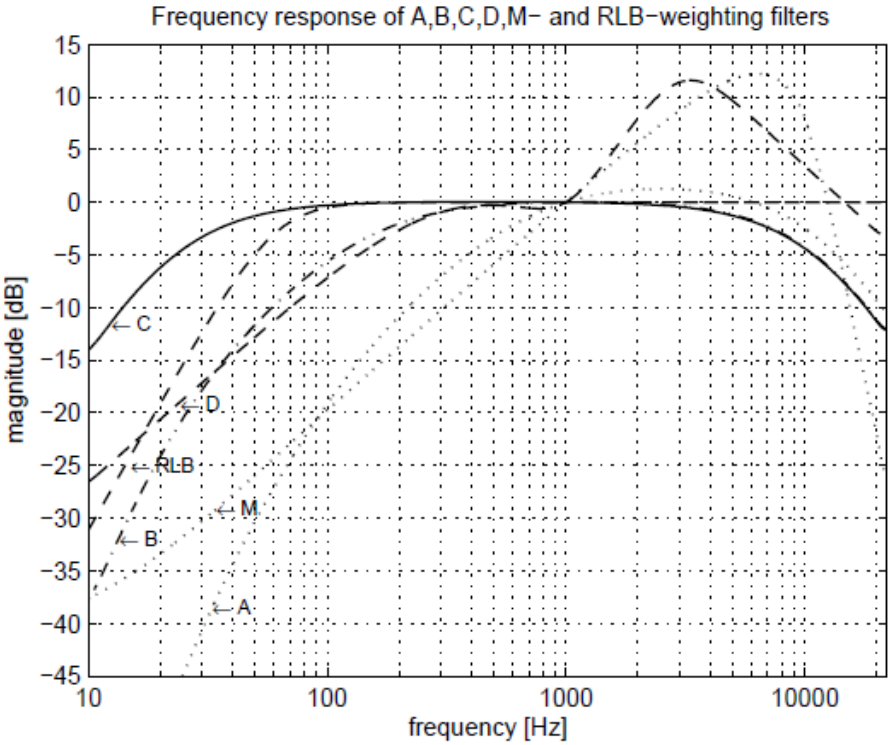

Fig. 2.5: Frequency weighting filters of different level measures (A, B, C, D, M, RLB). (Skovenborg and Nielsen, 2004)

2.2.2 Loudness models

Loudness models are designed to reproduce the complex process of auditory processing with respect to loudness (which may be quite distinct from other auditory effects, such as, e.g. masking, scene analysis or spatial hearing). There are different degrees of imitation of physiological processes. The simplest models show a focus on the frequency-place transformation in the inner ear, while more complex models also include the mechanics in the middle and inner ear (Epp *et al.*, 2010; Pieper *et al.*, 2016). Furthermore, auditory models exist which mimic other parts of the auditory pathway, e.g. mimicking the chemical processes in the hair cells (Meddis, 1988). In this section, only the 'simple' (dynamic) loudness models are briefly explained, since only those were used in the studies.

In recent years, three loudness models have received particular interest: (1) the time varying loudness model (TVL) of Glasberg and Moore (2002), (2) the Zwicker model for instationary sounds (DINA1) recommended in the DIN 45631 / A1 standard (2010), and (3) the dynamic loudness model (DLM) of Chalupper and Fastl (2002). Furthermore, Rennies *et al.* (2009) performed a major extension of the DLM (DLMext). These models are able to estimate the loudness of time varying sounds and basically the structure of their processing steps are almost similar. One of the main differences resides in the choice of critical bandwidth, which is discussed in section 2.1.2. The ERB scale is applied by the TVL whereas the DINA1 and the DLM rely on the Bark scale. Besides these essential differences, there are other minor aspects such as some filter processes that are handled differently. A detailed comparison of the model structures is provided by Appell *et al.* (2001) or

Rennies *et al.* (2010). A brief summary of the processing steps of these models is given below. A schematic diagram of two models (DLM and TVL) is shown in Fig. 2.6.

(1) In the first stage, a high-pass filter is used for the outer and middle transformation of the sound. (2) Subsequently, a critical-band filter bank is applied to separate the sound into different filtered time signals and envelopes are calculated. (3) In the next stage, the excitation is determined by a correction factor. Temporal (forward) and spectral (upward spread) masking effects can be accounted. Furthermore, the specific loudness can be calculated (loudness transformation). (4) Spectral loudness summation is applied by integrating the specific loudness-time pattern along the frequency dimension resulting in the instantaneous loudness. (5) Finally, short-term loudness can be derived by the temporal integration using a low-pass filter. Therefore, the cut-off frequency is approximately between about 8 and 14 Hz. A further stage is recommended by Glasberg and Moore (2002) for transforming short-term loudness into long-term loudness by applying another low-pass filter with a cut-off frequency at 0.5 Hz.

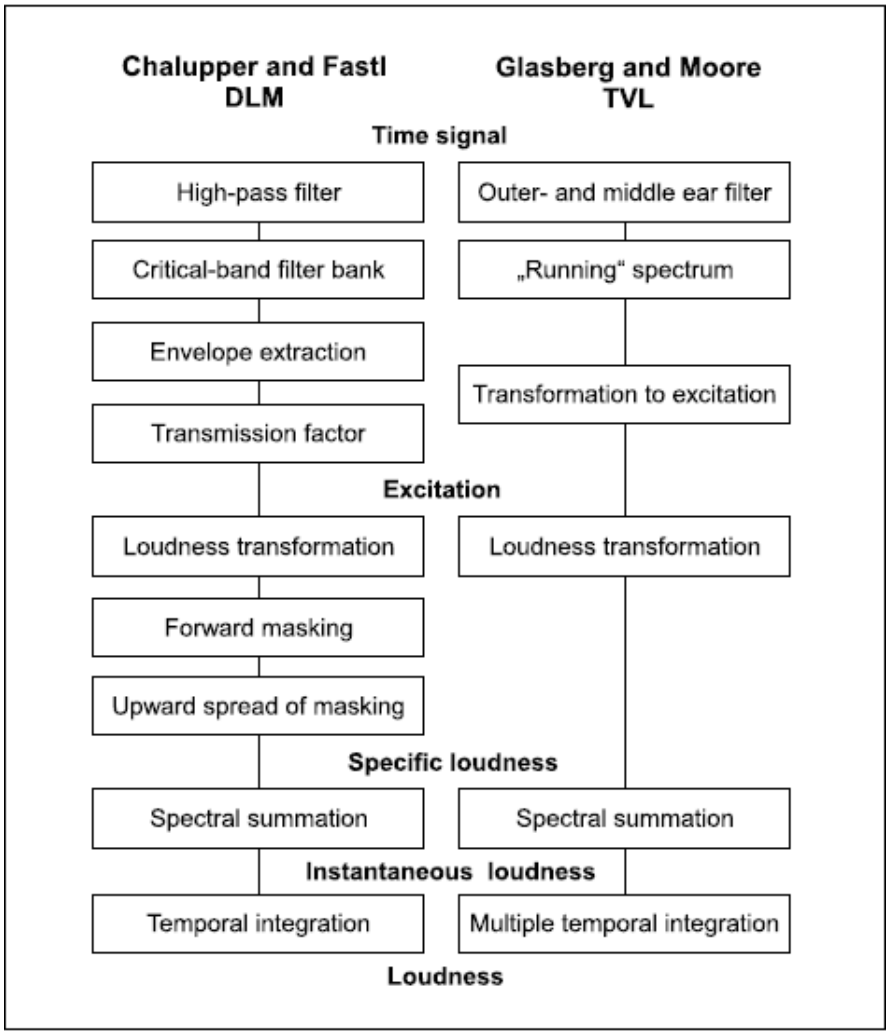

Fig. 2.6: Schematic structure of two loudness models: DLM (left) and TVL (right) (Rennies *et al.*, 2010).

There are different propositions to scale the short-term loudness down to overall loudness. A reasonable approach is to use distribution parameters like the peak, the mean (Glasberg and Moore, 2002) or certain percentiles, e.g. the 95-percentile N5 (Chalupper and Fastl, 2002), representing the distribution of the loudness. It is assumed that mainly loud frames affect the overall loudness. However, there is still disagreement on the question of whether peak or high percentiles are generally best suited for this purpose (Fiebig and Sottek, 2015). On the other hand, there are weighted averaging methods in the process of which, for example, periods with louder frames are weighted more heavily. This method seems particularly useful in order to reduce the influence of periods of silence on the estimation of the overall loudness (silence gating). This approach is implemented in the EBU R-128 standard as an extension to the ITU-R-BS. 1770. Both standards use a modified RLB weighting ("K-weighting") and silent gating calculates for 400 ms blocks each by a threshold criterion. In contrast to the ITU recommendation, the EBU R-128 offers an estimator for the overall loudness as well as an estimator for the true loudness peak.

2.3 Psychophysical methods

Loudness is a perceptual quantity and is usually measured by psychophysical methods. There are several approaches to classify these methods. A simple one is to differentiate the generated scaling level and the scaling method (Rajamanickam, 2002).

The scaling level can be either: A) nominal, B) ordinal, C) interval or D) ratio.

A) The nominal type differentiates between items or subjects based only on their categories. The categories have neither measurable distance to each other nor a preferred order. The nominal type represents the lowest level of scale information. B) The ordinal type provides a rank order by which data can be sorted. Yet, there is a lack of information about the distance between the ranks as well as any proposition about the relationship to values beyond the data. C) The interval type informs about the distance between items. D) The ratio type provides information about the ratio between items. It is also a metric scale.

The scaling method can be either: A) production, B) estimation or C) comparison. A) Production is an active scaling method in which the subject is required to adjust a certain category or magnitude of a perceptual size. B) Estimation is a passive scaling method in which the subject is presented a configuration of a particular category or magnitude of a perceptual size and then asked to make an assessment. C) Comparison is also a passive method in which the subject has to compare the perceptual size of several items to each other.

The most popular methods for loudness measurement are the following: (1) magnitude estimation which is assumed to provide a ratio scale level. This method was preferred by S. Stevens. (2) Categorical loudness scaling which allows interval scale level. (3) Matching, a comparison method which in general provides interval or in some cases ratio scale level. (4) Paired comparison which allows ordinal and, under certain

conditions, interval and ratio scale level. In this thesis only method 3 and 4 were used. Therefore, these methods are explained in more detail in the next two sections.

2.3.1 Matching

Probably the most frequently used matching-method is the 'Two Alternative Forced Choice' method (2-AFC) perhaps beside from cross modality matching. However, at this point we will only further look at the 2-AFC procedure, since it is used in this thesis (cf. Chapter 5). The 2-AFC procedure is a method to find the point of subjective equality of two stimuli. It was already used in the 19th century in the beginning of psychophysics by G. Fechner and has been modified many times in the course of time (e.g. the introduction of the staircase procedure by G. Békésy in 1960 or implementations of different transformed up-down methods by Levitt in 1971).

Both, a test signal and a reference signal are presented at intervals to a subject in successive trials. The subject must then decide which of these two intervals has been perceived as louder. The subject is forced to choose one of the two alternatives ("forced choice") excluding the option of two equally loud intervals. This method aims at approaching the point of equal loudness. An adaptive procedure with a staircase procedure is mostly used. The sound level of the test signal is changed by a predetermined level increment (staircase) after each trial, depending on the response of the subject while the reference signal is continuously retained on a fixed level. The level change depending on the response is described by the up-down rule. The level of the test signal should be increased if it is perceived to be softer than the reference signal and vice versa. However, the accuracy of the point of equal loudness can be increased if level adaptation is performed only in the case of multiple constant responses of the subject. This can be time consuming, though. Hence, the up-down rule should be chosen based on the level of difficulty of the loudness judgment which depends on the use of the specific stimulus and paradigm. After a certain number of trials, the level increment is reduced to enhance accuracy. This usually depends on a predetermined critical number of reversals. Reversals are those trials where the sequence of stimulus presentations changes from an ascending to a descending sequence of stimulus intensities, or the reverse pattern. These turnaround points are recorded and their average defines the stimulus threshold value. Furthermore, having achieved a number of trials or a specified number of reversals the procedure will be terminated. To increase the accuracy it is useful to repeat the measurement or at least to select the starting value as a high-level and once as a low-level. Moreover, it is recommended to interleave the AFC-sequences of several conditions in order to disguise the up-down rules for the test subjects. The result of this method provides sound levels of multiple test signals relative to a common reference at the point of equal loudness.

2.3.2 Paired comparison

Rankings are ubiquitous these days: the current table in the football league, the top ten lists of most popular movies, the music charts, or the list of top manager salaries. In almost all areas of life comparable objects or their properties are evaluated and put into rankings. Paired comparison is the easiest way to bring non-directly measurable variables into an unambiguous ranking. It is important to mention that other methods (magnitude estimation, ratio production, categorical scaling) have a higher degree of difficulty in accomplishing the task. These methods presume that subjects can fall back on a widespread inner scale that are used to measure items. If the task is too demanding the subjects, it can easily lead to extreme inaccuracies.

The advantage of paired comparison is that it breaks down the task of creating a ranking to the lowest level of difficulty. On the other hand, there are some well-known disadvantages: (1) Paired comparisons require a great deal of time in implementation. *m* items to be compared result in $\binom{m}{2}$ pairs for comparisons, in which the subjects must examine the items for the feature to be scaled. It should be noted that with regard to the effort, the 2-AFC method mentioned in the previous section exceeds this simple paired comparison by far. (2) Contradictory statements manifested as cyclic triads (A > B, B > C, C > A) indicate some uncertainty in the rating of the feature. However, cyclic triads could also be a sign of multi-dimensionality of a feature. (3) Paired comparisons usually only provide an ordinal scale level. This may mean that in some cases scale information unnecessarily remains unused although it could have been provided by the subjects when using another psychophysical method. Therefore, the Bradley-Terry-Luce method (BTL-method) is often recommended to increase the scale level (e.g. Ellermeier *et al.*, 2004; Wickelmaier and Schmid, 2004; Tsukida and Gupta, 2011).

The BTL method starts with the result from the pairwise comparisons. This is a count matrix *M* of the number of times that each option was preferred over every other option (Eq. 2.11).

$$M_{ij} = \begin{cases} \#\text{ of times option } i \text{ preferred over option } j, & i \neq j \\ 0, & i = j \end{cases} \tag{2.11}$$

Sorting the sums of each row M_i by size the ranking order can easily be obtained. Obviously, transforming *M* into ordinal scaled data leads to an information loss. Hence, we apply the BTL-method that is used to establish data on a ratio scale level by postulating a relationship between preference probabilities and scale values (Ellermeier *et al.*, 2004):

$$p_{ij} = \frac{\pi(i)}{\pi(i)+\pi(j)} \tag{2.12}$$

in which p_{ij} is the probability that a subject prefers the option *i* over *j* and where π is the scale value. The scale values can be estimated, for example, by using an iterative maximum likelihood method. A common approach is to use the distance μ_{ij} between $\pi(i)$ and $\pi(j)$ to build the BTL-scale (Elo, 1965). This can be done by applying the logit-transformation (Tsukida and Gupta, 2011) which is realized by the logarithm of the scale values,

$$\mu_{ij} = s \cdot \left(log(\pi(\mathrm{i})) - log(\pi(j))\right) \tag{2.13}$$

in which *s* is a scale parameter. In literature $log(\pi)$ is sometimes called the log-BTL scale value (e.g. Dittrich *et al.*, 2000). Furthermore, it is important to notice that equations (2.12) and (2.13) implicate that to gain decent distances it has to be applied:

$$p_{ij} \neq 0 \wedge p_{ji} \neq 0 \tag{2.14}$$

Therefore, it is sometimes necessary to divide the matrix *M* into subgroups, i.e. submatrices $M_{ij}^{(k)}$ in which every member *i, j* satisfies Eq. (2.14). The quality distance of the group $M^{(k)}$ to $M^{(k\pm1)}$ is calculated using the quality distance of the common members of these groups.

2.4 Electroencephalography

Neural activation in the brain generates an electric field on the scalp. This field can be measured by electroencephalography (EEG). This non-invasive imaging technique makes it possible to measure neural activity at a high temporal resolution (Teplan, 2002). However, the local resolution is rather low. This is due to the fact that temporal superposition of many neural processes and the electric field shielded by the skull and skin restrain the separation of sources.

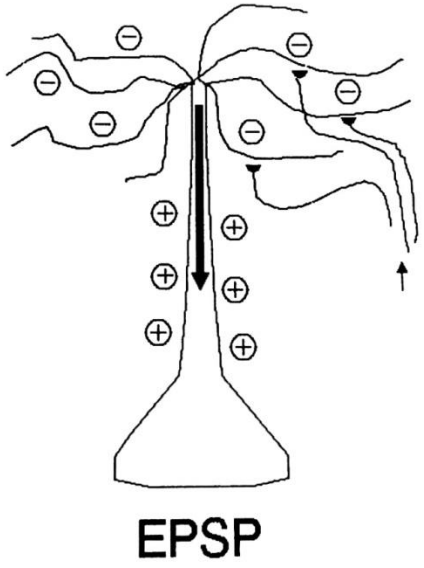

Fig. 2.7: Schematic representation of the formation of dipole fields on pyramidal cells. Exciting postsynaptic potentials (EPSP) indicate an influx (-) in the area of the apical dendritic branches (Scherg, 1991).

2.4.1 Anatomical basics

Changes in the electrical potential of a neuron are caused by synaptic excitation. In the process, transmitters are released inducing a local membrane current. This results in the depolarization of the nerve cell. The electric charge transfer can be described as a dipole. EEG mainly measures excitatory postsynaptic potentials (EPSP), which are caused by an influx of ions in the region of the apical dendritic branches (Fig. 2.7). The activity of pyramidal neurons in the celebral cortex mostly contributes to the macroscopic field due to their architecture and rather symmetric orientation. A similar orientation of the neurons towards each other is essential, since in the far field on the scalp individual dipoles generated by single cells are hardly recordable – contrary to a collective dipole of large populations of active neurons. However, only nerve endings and bends contribute to the far field, since an asymmetric intraaxial current flow is necessary for dipole formation (Scherg, 1991). The dipole clusters are not activated synchronously resulting in dispersion effects. This slight desynchronization produces low-pass filter effects, especially for brain regions where late neuronal processing operations take place (Scherg, 1991). The electric field of these dipole clusters on the scalp is essentially affected by the intervening layers: scalp, skull and brain. These layers have different electrical conductivities. Especially the low conductivity of the skull has a shielding effect on the electric field. As a result, the field appears on the scalp spatially widened with attenuated intensity. Finally, these weak electrical signals can be detected on the scalp as potential difference by using at least two electrodes.

2.4.2 Auditory evoked potentials

Electrical activity on the scalp evoked by acoustic stimuli originates from different dipole sources which are characterized by corresponding delay times and their measurability by respective groups of electrodes. This relationship is well studied for the EEG response to stimuli of clicks and tone bursts (Picton *et al.*, 1974; Radeloff *et al.*, 2014). These stimuli evoke significant voltage fluctuations known as auditory evoked potentials (AEP), illustrated in Fig. 2.8. These potentials are assigned to three areas of the brain. However, these assignments have not been conclusively clarified yet. The auditory brainstem response (ABR) includes the wave I-V, which are components with an early latency (0-10 ms). The middle latency response (MLR) consisting of P0, Na, Pa, Nb and P1 arises at 10-80 ms and is associated with the Thalamus (P0, Na) and Cortex (Pa, Nb, P1). Late latency (> 80 ms) AEPs are cortical in origin and are composed of N1, P2 and N2. All of these components range from 0.1 µV to 2.5 µV in amplitude, provided that voltage is measured by electrodes between vertex and mastoids.

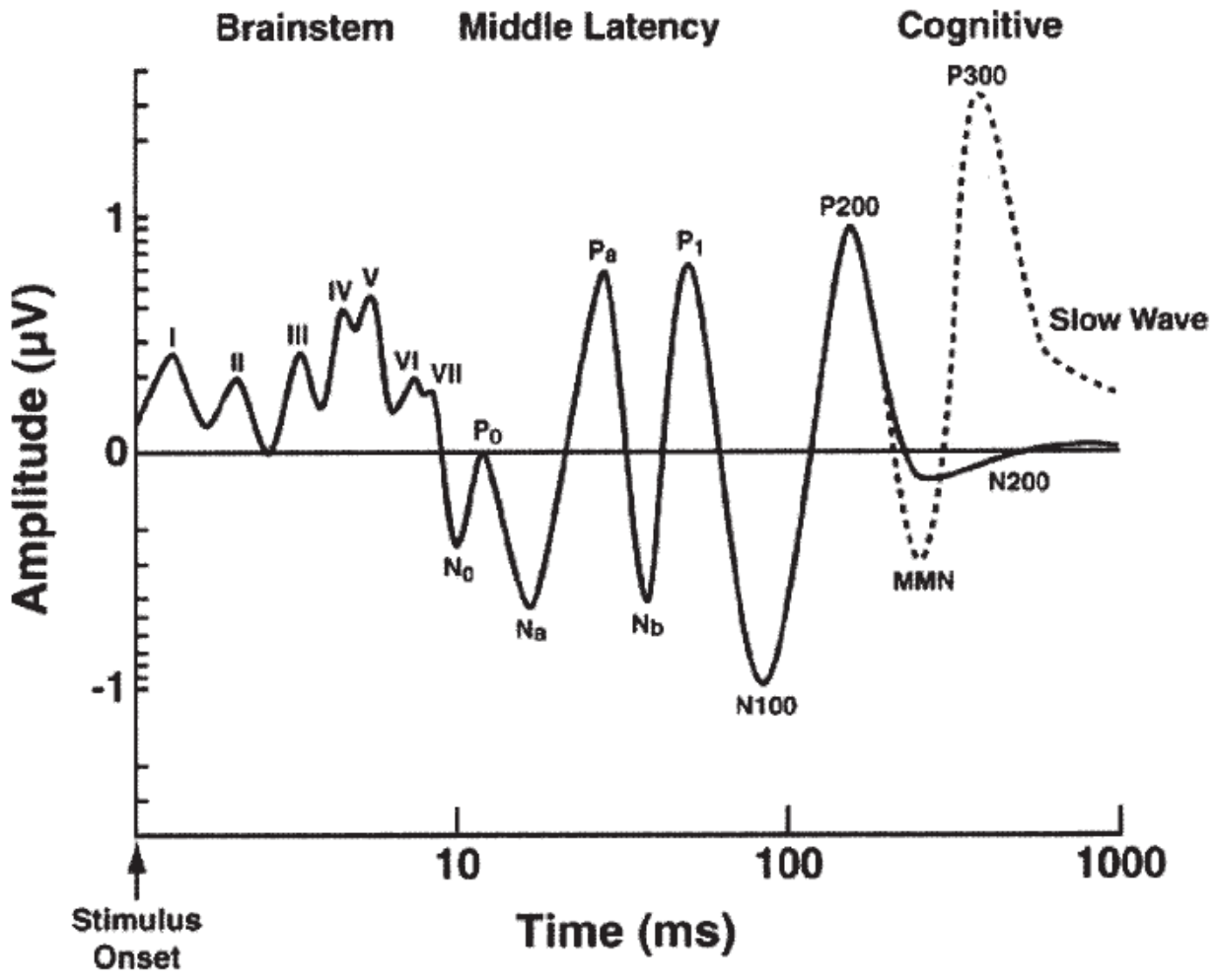

Fig. 2.8: Schematic illustration of the auditory evoked potential (AEP) and event related potentials (ERP: MMN, P300). (Cahn and Polich, 2006).

2.4.3 Event related neural activity

Beyond the AEPs, perceptible events can elicit neural activity. A distinction is made between neural oscillations and event related potentials (ERP).

Neural oscillation is repetitive neural activity that is commonly sinusoidal (Teplan, 2002). It has been categorized into five basic groups: delta (0.5-4 Hz), theta (4-8 Hz), alpha (7.5-13 Hz), beta (13-30 Hz) and gamma (> 30 Hz). Usually, the amplitude is measured from peak to peak and may reach a magnitude of up to 100 µV. Event related neural oscillation is mainly found in alpha activity. Often, an abruptly increasing alpha activity is related to closing the eyes and a decreasing alpha activity correlates with eye opening or is induced by cognitive mechanisms like thinking and calculating. The latter plays an essential role in EEG paradigms that involve a task. However, mental states like relaxation, stress, alertness, resting, hypnosis and

sleep can also induce neural oscillations. In fact, they are even more associated with them. It should be mentioned that the transition of the meaning of events to the mental states is smooth.

Event related potentials can be induced by specific sensory, cognitive, or motor events. They often appear as additional components next to the AEP. The most well-known ERP, the P300 (Fig. 2.8), is a central-parietal component with a latency of 300 ms (Polich, 2007). It is triggered by events that attract attention. Therefore, they need stimulation paradigms that repeatedly involve attention tasks (usually oddball-paradigms).

Furthermore, there are ERPs that are elicited auditorily. The mismatch negativity (MMN) belongs to this type. The MMN is a change-specific component of the auditory ERP which can be elicited either by the brain's automatic response to any change in auditory stimulation or by different kinds of abstract changes in auditory stimulation, e.g. grammar violations (Näätänen *et al.*, 2007). The MMN is a frontocentral component seen as a negative displacement with a latency of 150 - 250 ms (Fig. 2.8).

2.4.4 Neural entrainment

Neural entrainment is the capacity of the brain to synchronize the neural activity with the frequency and the modulation of a stimulus. It is assumed that different coupling mechanisms exist between stimulus and EEG response (Schnitzler and Gross, 2005). However, the discourse of the question how these coupling mechanisms work is still in process. Most studies deal with two entrainment phenomena: frequency following response (FFR) and envelope following response (EFR).

The FFR encodes spectral features of the stimulus. It reflects particularly well the pitch of the stimulus which is associated with the fundamental frequency (Krishnan *et al.*, 2004). It is suggested that phase-locked activity in a population of neural elements within rostral brainstem underlies the FFR generation (Worden and Marsh, 1968; Glaser *et al.*, 1976). The modulation of the fundamental frequency in the stimulus provides appropriately modulated FFRs and, surprisingly, a low-frequency response corresponding to the envelope of the frequency modulation (Rance and Rickards, 2002). This additional coded response to the frequency modulation is related to the EFR (John *et al.*, 2001). Krishnan *et al.* (2004) has impressively shown that FFRs can be found for the pitch of different words.

The EFR encodes the envelope of the stimulus (Picton *et al.*, 2003; Aiken and Picton, 2008). In part, this also includes the envelope of the frequency modulation, but mainly this term refers to the temporal envelope of the stimulus (Ding and Simon, 2014). The term auditory steady state response (ASSR) is also frequently used, but Purcell *et al.*, 2004 rightly argued that this choice of terminology is not appropriate due to the instationary nature of the response. A popular hypothesis for the origin of EFR is the "onset tracking hypothesis". The temporal envelope of a natural sound usually has many acoustic "edges", e.g. onsets and offsets. These edges can elicit AEPs (see section 2.4.2.). Hence, it has been proposed that EFRs are superpositions of edge-related AEPs (Howard and Poeppel, 2010). However, this hypothesis is currently controversial (Simon and Ding,

2014). The findings of Thwraites *et al*. (2016) undermine this hypothesis by showing that loudness even better represents the EFR than the envelope does.

2.4.5 Correlates to loudness in the encephalogram

For a considerable time there have been attempts to find a physiological and objective correlate to loudness that can be determined passively without a report of the subject. There are several approaches that differ mainly in searching for sources at different stages of the auditory pathway. For example, otoacoustic emissions provide a correlate of the compressed intensity related to loudness of the outer hair cells (Neely *et al.*, 2003; Epstein and Florentine, 2005). However, it is suggested that loudness perception is primarily generated cortically in the region at the posterior medial Heschl's gyrus (e.g. Röhl and Uppenkamp, 2012; Behler and Uppenkamp, 2016; Thwaites *et al.*, 2016). Therefore, correlates from earlier sources of the auditory pathway are rather intermediate results of loudness processing and may be associated with instantaneous and/or partial loudness.

Already many decades ago, the encephalogram was examined for correlates of loudness. Bauer *et al.*, 1974 tried to determine a perceptual correlate of loudness within the AEP. Further studies on the topic provided results with contradictory evidence (Bauer, 1974; Babkoff *et al.*, 1984; Darling and Price, 1990; Serpanos, 1997; Fobel and Dau, 2004; Junius and Dau, 2005; Dau et al., 2005; Silva and Epstein, 2010; Silva and Epstein, 2012). Most authors agreed on the fact that sound intensity is reflected in the strength and the latency of the ABR (Pratt and Sohmer, 1977; Serpanos *et al.*, 1997). However, there is evidence that further components of the AEP are related to loudness. Madell and Goldstein (1972) discovered correlations between the peak-to-peak strength of some MLR components and loudness estimates. They emphasized the early components P0 and Na. Furthermore, the peak-to-peak strength of the cortical components N1 and P2 also correlates with sound intensity (Pratt and Sohmer, 1977; Hegerl *et al.*, 1994).

The EEG response to continuous stimuli often does not allow the examination of AEPs, as the elicited potentials tend to overlap or cancel each other out. However, this EEG response often shows a relationship to the envelope of the stimulus and is therefore associated with envelope following response (EFR). In a simple scenario in which amplitude modulated sinusoids evoke the EFR, the modulation frequency and their harmonics are present in the long-term spectrum of the EEG. Several studies showed an interdependence between the amplitude of the fundamental frequency of the EFR and categorically scaled loudness (Ménard *et al.*, 2008; Castro *et al.*, 2008; Emara and Kolkaila, 2010; Eeckhoutte *et al.*, 2016). Moreover, in an MEG study, Thwaites *et al.* (2016) found several components of the EFR corresponding to the loudness of speech. They used cross correlation to find the best correlating latencies and found four components at 45 ms, 100 ms, 165 ms and 275 ms. Furthermore, they were able to prove that the last component shows higher correlation to the short-term loudness than to the instantaneous loudness.

2.4.6 Signal distortions in the encephalogram

During the EEG recording of a specific brain response to a stimulus, interfering signals from different sources superimpose on other to produce the measured signal. Neural processes always take place in the brain independently of any stimuli and produce spontaneous brain activity that is not correlated with the stimulus. This is called background activity. The voltage measured in the EEG of this background activity is in the range of 40-50 µV (peak-to-peak). This is ten times higher compared to the strength of the components in the AEP. By assuming that the brain responds equally to the same stimulus, the uncorrelated background activity can be attenuated by the factor $1/\sqrt{N}$ by averaging over the repetitions *N*. The mean power of the background activity may change over the course of repetitions. Therefore, it is recommended to use weighted averaging according to the inverse mean power of each repetition. Riedel *et al.* (2001) proposed a method that uses this kind of averaging. Furthermore, they modified this method by estimating the mean power iteratively.

There is also recorded activity that is not of cerebral origin. This activity is termed artifact. Artifacts can be of physiological or extraphysiological origin. Physiological artifacts are generated by the body except the brain. Extraphysiological artifacts arise from sources of the environment or the equipment. However, if background activity is evoked by clearly attributable processes identifiable by conspicuous patterns, cerebral artifacts are the appropriate term.

The three most common extraphysiological artifacts are: (1) change of electrical conductivity between skin and electrode, (2) AC power and traction supply and (3) electrical devices. (1) The contact between electrodes and scalp is usually ensured by using contact gel with a high electrical conductivity. The change in skin resistance due to sweat or the change in electrode positioning slowly causes changes in the measured voltage over several hundred volts. This artifact can be reduced by using a high-pass filter with a low-cut frequency between 0.3 and 1 Hz. Electrode artifacts can be reduced by averaging across a cluster of adjacent electrodes. The electrical field of the environment is decisively influenced (2) by the AC power supply in the walls of houses and often in cities by traction supply. These artifacts superimpose the EEG recordings with an oscillation corresponding to the respective AC power supply. In Germany, for example, AC power supply generates an oscillation of 50 Hz and the traction supply an oscillation of 16 Hz. An electrically screened booth can reduce these two environmental artifacts. Alternatively, notch filters or bandpass filters can be used. (3) Electronic devices to stimulate and record the EEG should also be placed outside the booth. They also contain the AC power supply. In order to prevent measuring the electrical field of a headset that may be similar to the stimulus, tube headphones should be used.

Physiological artifacts are dominated by muscle movement. The heartbeat generated by the heart muscle can be found in the EEG recording. Furthermore, the movements of the jaw (e.g. chewing), neck and facial muscles seriously interfere the EEG recording due to their short distance to the electrodes. These artifacts occur spectrally broadband with a peak-to-peak strength of 100-200 µV. It is very difficult to reduce muscle artifacts in the EEG. Therefore, it is recommended to instruct participants to avoid muscle movement as far

as possible during the recording. Otherwise affected repetitions disturb the averaged result due to the high strength of these artifacts. To prevent this, repetitions that are affected by such artifacts are typically removed. Reducing the number of repetitions, however, would also reduce the signal to noise ratio. Alternatively, a spectral averaging method introduced by Vardi and Zhang (2000) that applies the multivariate L1-median might be considered to reduce artifacts. This averaging method is extremely robust with regard to artifacts because it allows keeping repetitions which are affected by artifacts (more details in chapter 4).

Another important class of physiological artifacts is associated with eye movement. These ocular artifacts include: Eye blinks, roving eye movement and eye flutter (can also be counted to the facial muscles), just to mention the most prominent ones. Eye blinks are considered to be a major problem for EEG recording: firstly because of the high strength of the elicited fluctuation (100 µV) and secondly because of the need to moisten the eyes regularly. The same methods (removing and reducing) as in the other muscle artifacts can be used to reduce the influence of these artifacts. However, the use of artifact rejection is recommended due to the clearly local definition of these artifacts. The most common approaches are: (1) to regress the interfering signals using an electrode near the eye that records the regressor or (2) to separate and remove the eye components by using independent component analysis (Makeig *et al.*, 1996). This method can also be used to reject the heart beat artifact.

Cerebral artifacts are evoked potentials or brain waves that are not primarily related to the perceptual processing of the stimulus. Sometimes experimental paradigms can cause these artifacts to correlate with the stimulus. Paradigms with a task that involves the use of a button box elicit motor evoked potentials. Using screens that change in sync with acoustic stimulation trigger correlated with visual evoked potentials. These potentials show a very small fluctuation (< 10 µV). Hence, they are difficult to remove afterwards. Instead, such artifacts should be anticipated and avoided when designing an experimental paradigm. Finally, increased alpha activity is often an undesirable phenomenon in EEG recording. Alpha waves (8-13 Hz, 10-20 µV, sinusoidal activity) often occur during tasks, mental states of relaxed wakefulness or when participants close their eyes (Hughes and Crunelli, 2005). This artifact is usually treated by removing repetitions affected by alpha waves using a threshold criterion. Parietal and frontal electrodes are mainly affected by alpha waves. However, the central electrodes show contours which are less affected by the alpha waves. The choice of the measuring electrodes can therefore reduce the effect of the alpha waves.

3 Deficiencies of models for overall loudness estimation of music

Abstract

Despite considerable progress in improving models for loudness perception, sound level measures are in common use for many practical applications, like, e.g., estimating the loudness of musical stimuli, due to their simple implementation and decent performance (Skovenborg and Nielsen, 2004; Vickers, 2010a). This is in part owed to the familiarity with the characterization of sounds along a decibel scale, which, as a technical measure, sometimes may suggest a certain grade of objectivity, as it can be read from a sound level meter. In fact, improved level measures have recently been developed and commercially used to predict the loudness perception of music [e.g. EBU R 128-2014]. However, the origin of the performance deficiencies of loudness models when used for real-life applications is still unclear. In this study we compare the prediction performance of level measures and loudness models with psychoacoustic results from a scaling procedure, with the aim to try and improve the performance of loudness predictions. In a paired comparison paradigm 29 listeners were doing overall loudness judgments of 14 music sequences with 10 s duration each. The Bradley-Terry-Luce model was used to construct a scale of loudness judgments. To improve the validity, leave-one-out cross validation with multiple regressions was employed. Our findings confirm the superior performance of level measures, but also approaches to modify existing loudness models can be derived. The prediction performances of loudness models can be improved by using a low-pass filter around 4 Hz for instantaneous loudness in the preprocessing. Furthermore, overall loudness judgements seem to be strongly affected by the perceived sharpness of the sound.

3.1 Introduction

The prediction of loudness, i.e. the perceptual correlate of sound intensity, for complex sounds like music is still a challenge (e.g. Skovenborg, 2004; Vickers 2010a). This is particularly evident in the fact that, for commercial applications, simple level measures are preferred to sophisticated loudness models (EBU R 128-2014). Indeed, there are many kinds of sounds that are collectively referred to as music, like various interactions of running speech respectively singing voices, harmonic, percussive and synthetic instruments, everyday noises and various sound samples. It is a rather complex task to define the term music in a few words. Furthermore the broad spectrum of sound scenarios requires the consideration of various parameters that determine the loudness. These include spectral and temporal loudness integration as well as effects caused by amplitude modulation. Moreover it is suggested that for music there are also effects at a higher stages of auditory processing that affect the loudness judgment, e.g. preferences (Cullari and Semanchick,

1989), pitch (Neuhoff *et al*., 1999), increased musical experience with age (Fucci, 1999), hearing expectation for different music genres (Barrett and Hodges, 1995), or context effects (Arieh and Marks 2003).

In recent years loudness models have undergone a great improvement in dealing with non-stationary stimuli (Fastl and Chalupper, 2002; Glasberg and Moore, 2002; Rennies *et al*., 2009). In the case of the loudness prediction of music the main objectives are to determine the momentary stream of loudness, i.e. the instantaneous loudness, as well as the resulting overall loudness which is assumed to be the instantaneous loudness combined across certain time windows to a summary assessment. It is unclear how the overall loudness transformation proceeds because of the complexity of the task to integrate changes in loudness to one representative value. Many suggestions have already been made to formulate a general, robust approach. Recently, it has become very popular to consider the instantaneous loudness stream as a time-independent distribution and to use parameters of the distribution, e.g. percentiles, as an estimator of the overall loudness (e.g. Chalupper and Fastl, 2002; Rennies *et al*., 2015, Fiebig and Sottek, 2015). The 95% percentile is recommended as a robust measure (Fastl and Chalupper, 2002). An alternative approach is the use of a weighted average of the instantaneous loudness. The simplest realization for example is to ignore the silent sections of the stimulus for assessment, often called as “silence gating” (EBU R 128-2014). Glasberg and Moore (2002) introduced two more stages while processing the instantaneous loudness which they called ‘short-term loudness’ and long-term loudness’. For the realization they recommended different adapted low-pass filters. Rennies *et al*. (2015) showed that for the overall loudness prediction of fluctuating technical sounds the performance of some loudness models can be improved by using longer time constants, an approach equivalent to low-pass filtering. Finally, it should be noted that the overall loudness may also be influenced by a composite of several other perceptual parameters, or at least they can be influenced by other parameters than just the loudness.

In our study we test the performance of overall loudness predictions of different loudness models and level measures. We hypothesized that the loudness models perform the task better due to the consideration of loudness effects relevant to music. However, deficiencies of these models have already been reported (Skovenborg, 2004; Vickers 2010a). Hence we also want to optimize the adjustments of the models to improve their performance. Therefore, we follow the recommendations of Glasberg and Moore (2002) and Rennies *et al*. (2015) and implement a short-term loudness stage from which the overall loudness is estimated. Finally, we investigate the influence of one other perceptual measure on the overall loudness, that is, psychoacoustic sharpness, because this bears some resemblance to timbre, which is one central feature of musical sounds. Using multiple regression methods it is tested to what extent the overall loudness judgement of music is influenced by sharpness perception.

3.2 Methods

3.2.1 Participants

29 listeners, fourteen females and fifteen males, aged 18 to 32, participated in the study. All of them had normal audiograms (i.e. absolute thresholds in quiet ≤ 15 dB HL for frequencies between 125 Hz and 10 kHz), with no previous history of any hearing problems. Written informed consent was obtained and all participants were paid volunteers. The study was carried out in accordance with the Declaration of Helsinki 2008.

3.2.2 Stimuli and apparatus

All participants listened to 14 music examples of different music genres (classical music, jazz, rock, hip hop, pop) which each had a duration of 10 s. The RMS level of all music examples was between 70 and 85 dB SPL (RMS). The music examples were chosen according to their dynamic range, which should be comparatively low to make the task easy to perform. All stimuli were stored digitally with a sampling rate of 44.1 kHz and a resolution of 16 bit. During the experiments, the participants sat in a double-walled hearing booth. The sounds were D/A-converted by an external audio interface (RME-Fireface UC), digitally attenuated with Matlab R2015b and presented diotically via headphones (Sennheiser HD 650).

3.2.3 Procedure

The stimuli were compared in loudness with each other as pairs. In each trial two 10-s music examples were successively presented, separated by a 500 ms silent interval. After the presentation the listeners were forced to indicate which of the two signals was louder by clicking with the mouse on the corresponding dialog button on a computer screen. During the measurement each trial was repeated once in reversed order to prevent primacy- and recency-effects. All trials were presented in pseudo-randomized order. After finishing the loudness comparisons, the subjects were asked to evaluate the preference of the played music sequences according to their individual taste.

The full experiment with 14*13 comparisons of 10-s stimuli took approximately two hours including the response times and one or two short breaks. Therefore, to reduce the amount of time needed for this procedure, only 10 participants did the full experiment. After that, the number of comparisons was reduced for the rest of the participants by omitting those judgements that had a univocal outcome for the initial group of 10 listeners (e.g., the softest vs. the loudest music samples), as these were assumed to be the same for all participants.

3.2.4 Data evaluation

A: BTL-method

The result of the pairwise comparisons is a count matrix *M* of the number of times that each option was preferred over every other option,

$$M_{ij} = \begin{cases} \#\text{ of times option } i \text{ preferred over option } j, & i \neq j \\ 0, & i = j \end{cases} \tag{3.1}$$

Sorting the sums of each row M_i by size we obtain easily the rank order in loudness averaged across all subjects. Obviously, transforming *M* into ordinal scaled data leads to an information loss. Hence we applied the Bradley-Terry-Luce method (BTL) that establish data on a ratio scale level by postulating a relationship between preference probabilities and scale values (Ellermeier, 2004):

$$p_{ij} = \frac{\pi(i)}{\pi(i)+\pi(j)} \tag{3.2}$$

where p_{ij} is the probability that a subject prefers the option *i* over *j* and where π is the scale value. In our study we used a maximum likelihood method of Wickelmaier and Schmid (2004) to estimate the BTL scale values. The distance μ_{ij}between $\pi(i)$ and $\pi(j)$ can be determined by applying the logit-transformation (Tsukida and Gupta, 2011) which is simply realized by the logarithm of the scale values,

$$\mu_{ij} = s \cdot \big(log(\pi(i)) - log(\pi(j))\big) \tag{3.3}$$

where *s* is a scale parameter. In literature $log(\pi)$ is sometimes called the log-BTL scale value (e.g. Dittrich *et al.*, 2000). In our study this logarithmised value is the scale value we mainly deal with (henceforth referred to as BTL-value). Equations 3.2 and 3.3 implicate that to gain decent distances it has to apply:

$$p_{ij} \neq 0 \wedge p_{ji} \neq 0 \tag{3.4}$$

Therefore the matrix *M* has to be divided into groups i.e. submatrices $M_{ij}^{(k)}$ where every member *i, j* satisfies condition (4). The quality distance of the group $M^{(k)}$ to $M^{(k+1)}$ is calculated using the quality distance of the common members of these groups.

B: Loudness models, level measures and sharpness

Different models and level measures were used to predict overall loudness. On the one hand the level measures were: sound pressure level (dB SPL), A-weighted and B-weighted sound pressure level (dB(A), dB(B)) and the EBU R-128 standard (EBU R-128, 2014) that is applied by the European Broadcasting Union and is based on the ITU-R BS 1770-2 recommendation with the addition of a silent gating function. On the other hand the loudness models were: the dynamic loudness model (DLM) of and Chalupper and Fastl (2002), an extension of this model that includes the consideration of temporal effects of spectral loudness summation (DLMext, Rennies *et al.*, 2009), the time varying loudness model (TVL) of Moore and Glasberg, 2002 and the newest revision of the Zwicker model, the DIN 45631 / A1 standard (DIN 45631 / A1, 2010).

Loudness models predict the temporal change of loudness, i.e. instantaneous loudness. On the other hand the common method to estimate the change of level over time is to use the root mean square (RMS) value of different integration times: 'fast' corresponds to a 125 ms time constant and 'slow' corresponds to a time constant of one second. In our study different time constants from 10 ms to 1 s were used. There are different propositions to scale the instantaneous loudness resp. level down to overall loudness resp. level. A reasonable approach is to use statistical parameters like the percentiles that represent the distribution of the loudness resp. level over time. In our study percentiles from 70% to 100% (maximum peak) were considered for optimization. The EBU R-128 standard provides three loudness parameters: Maximum true peak level, loudness range and the so-called program loudness. In our case program loudness is recommended because it represents the integrated loudness over the duration of a radio program.

For the consideration of sharpness as one additional parameter to improve the prediction of loudness judgements, we used the sharpness model DIN 45692 (DIN 45692, 2009) that is recommended by the DIN standard.

All models were taken from the Fraunhofer IDMT Sound Quality and Speech Intelligibility Prediction Toolbox (SIP-Toolbox).

C: Leave-one-out cross validation

The prediction performance of the different models and level measures was evaluated by doing regression analysis between BTL-values and predictions. We used leave-one-out cross validation which is a robust method to test model ability (Todeschini, 2010). By comparing the Predictive Error Sum of Squares PRESS with the Total Sum of Squares TSS (i.e. the variance) the cross validated coefficient of determination R_{cv}^2 can be derived.

$$R_{cv}^2 = 1 - \frac{PRESS}{TSS} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3.5}$$

where y_i represents the ith- left-out sample of the experimental response (BTL-value) whether $\hat{y}_i$ is the predicted response of the regression model, which in our case is a linear model.

$$\hat{y} = a_1 \cdot x + a_2 \tag{3.6}$$

where the vector $\vec{a} = (a_1, a_2)$ represents the model parameters. Furthermore, repetitions of this procedure are necessary i to estimate the bias of R_{cv}^2. Therefore, we select the permutations of n-2 samples as a subset of our total set (n = 14 music sequences) and perform the leave-one-out cross validation with $\binom{14}{2} = 91$ repititions. Including other variables into the prediction only the regression model has to be extended to a multiple regression model:

$$\mathrm{y} = \sum(a_{1i} \cdot x) + a_2 \tag{3.7}$$

where now a matrix a_{1i} represents the i[th] model parameter. Including too many independent variables would of course result in overfitting. But overfitting is easily recognized as it is reflected in a decrease of R_{cv}^2 with an increasing variance. In contrast to that a significant increase of R_{cv}^2 will be interpreted as an indicator for more explained variance by the variables. However, it became apparent that the number of variables may not exceed *i* = 2 owed to the low number of samples. In our study we carry out a 1-sided T-test to compare the loudness models with their variable extensions (in our case: + sharpness).

3.3 Results

3.3.1 Short-term loudness

The adjustment of the modeled short-term loudness is done by adjusting the optimal cut-off frequency of the low-pass filter. The overall loudness is then gained by selecting the distinct percentile from the provided distribution of short-term loudness values. As mentioned, the maximum or the 95%-percentile had been recommended before to be the best choice (see above, Chalupper and Fastl, 2002). We varied both, the cut-off frequency and the employed percentile and compared the achieved prediction performance by calculating the averaged cross-validated R^2 for the loudness models and level meters. The results of this optimization procedure are shown for the loudness models in Fig. 1.

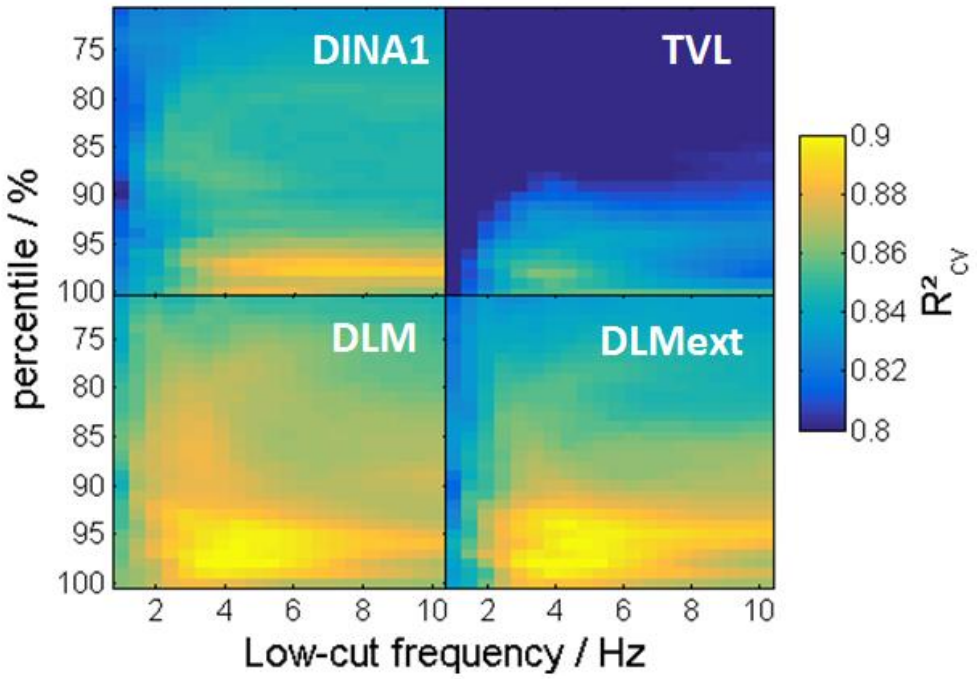

Fig. 3.1: The prediction performance represented by leave-one-out cross-validated coefficient of determination R^2_{cv} of the loudness models: the Zwicker model DIN 45631 / A1 - DINA1 (top left), the time-varying loudness model - TVL (top right) the dynamic loudness model - DLM (bottom left) and the extended dynamic loudness model – DLMext (bottom right). The low-cut filtering frequency was varied from 1 Hz to 10.5 Hz and also the percentile of the loudness distribution over time from 70% to 100% (i.e. the maximum).

It becomes clear that low-pass filtering as well as high percentiles improve the processing performance of the models. The DLM, the extended DLM and the TVL show optimal performance around 4 Hz cut-off frequency whereas the Zwicker model DINA1 performs well at higher cut-off frequencies around 8 Hz. In Fig. 4 the improved prediction power using a low-pass filtering stage in contrast to the use of no filter is well demonstrated. All models had in common that the 97%-percentile seems to predict the overall loudness best.

For the following analysis the optimal adjustments of low-cut frequency and percentile were calculated for each model and level measure.

3.3.2 Sound level measures vs loudness models

Our leave-one-out cross validation procedure provides averaged R_{cv}^2 and corresponding standard deviations which can be used for statistical comparison. By using the same approach as above we were searching for the optimal cut-off frequency respectively time constant and percentile for the overall loudness prediction of the sound level measures (Fig. 2).

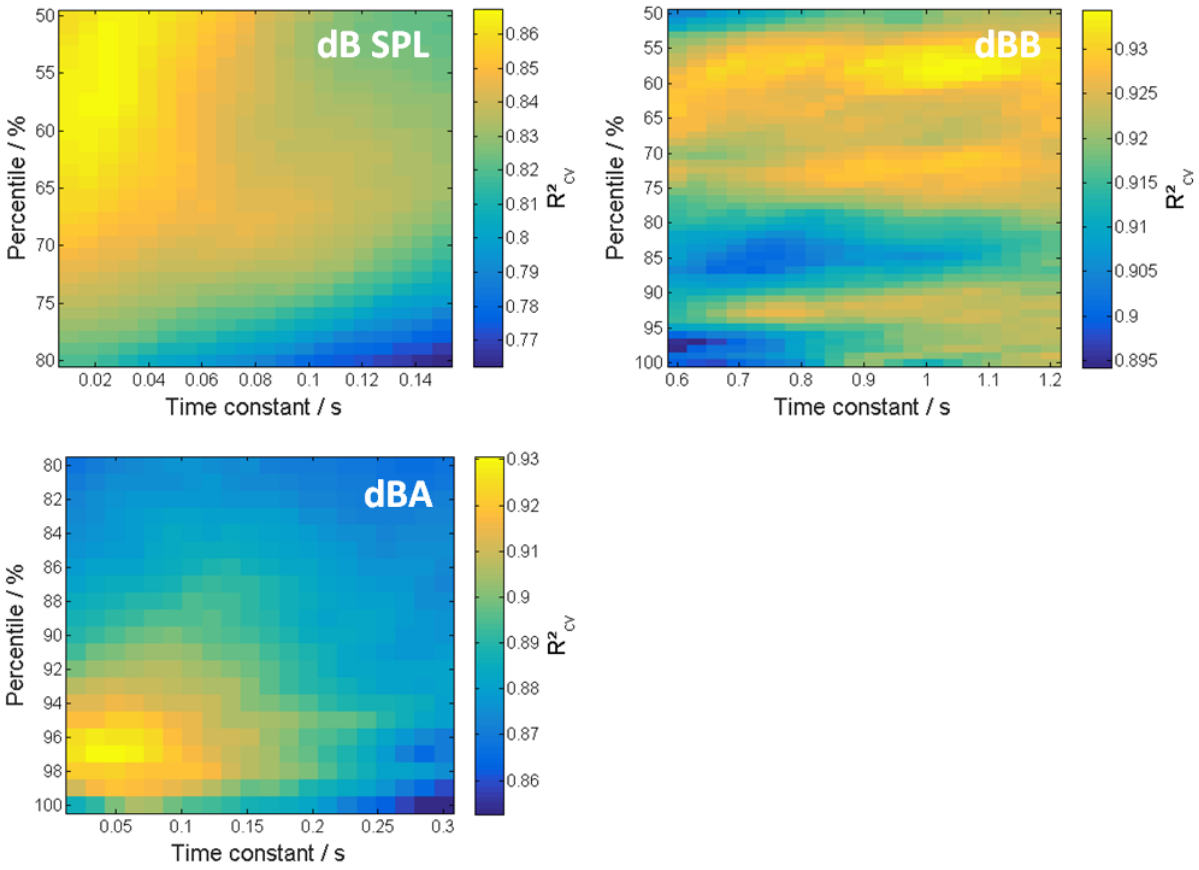

Fig. 3.2: Prediction performance of the A-weighted sound pressure level: leave-one-out cross-validated coefficient of determination R_{cv}^2 for different time constants and different percentiles. The time constant is reciprocal to the low-pass filter cutoff frequency. The percentile represents a location parameter of the intensity distribution over time.

Unlike the models, the optimal percentiles differ considerably. The unweighted sound pressure level and the B-weighted level show best performance at a percentile around 55 %, which is quite similar to the mean. The A-weighted level, on the other hand, is similar to the models and shows best performance at a percentile at 97 %. The optimal time constant for the unweighted sound pressure level as well as the A-weighted level is smaller than 100 ms, which is often referred to as the “fast” measuring mode for sound level meters. The B-weighted level is optimally adjusted with a time constant of around 1 s which is often referred as the “slow”

measuring mode for sound level meters. Figure 3 provides an overview of the prediction performance of each model and level measure after low-pass filtering and percentile extraction. The plot demonstrates that the frequency weighted sound level measures (dB(A), dB(B) and EBU R-128) outperform all the loudness models (T-test: $p < 0.01$) in terms of their prediction performance. The unweighted sound pressure level and the time varying loudness model show the worst performance to predict the results from the BTL scaling experiment.

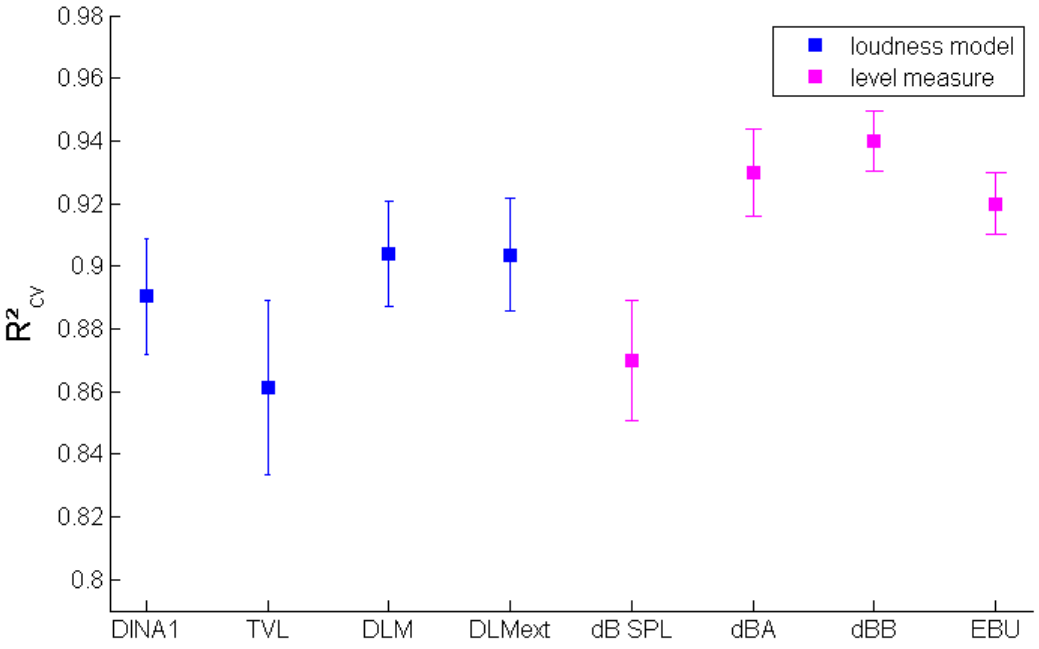

Fig. 3.3: An overview of the prediction performance represented by leave-one-out cross-validated coefficient of determination R_{cv}^2 of the loudness models: the Zwicker model DIN 45631 / A1 (DINA1), the time-varying loudness model (TVL), the dynamic loudness model (DLM) and the extended dynamic loudness model (DLMext); and of the level measures: sound pressure level (dB SPL), A-weighted sound pressure level (dBA), B-weighted sound pressure level (dBB) and the EBU R-128 standard (EBU).The error bars represent the standard deviation of R_{cv}^2.

3.3.3 The influence of sharpness to the overall loudness

In order to determine the influence of sharpness on the overall loudness judgement we have extended the input parameters of the regression model by adding the modeled sharpness as one additional regressor. The results in Fig. 4 show that the prediction performance of the loudness models strongly increases (T-test: $p < 0.01$). However, level measures show no degradation in their performance. An overview of this is shown in Table 1. Furthermore the results indicate that there is a stronger improvement of the model performance by including sharpness to the prediction process than by adjusting only the low-pass filter.

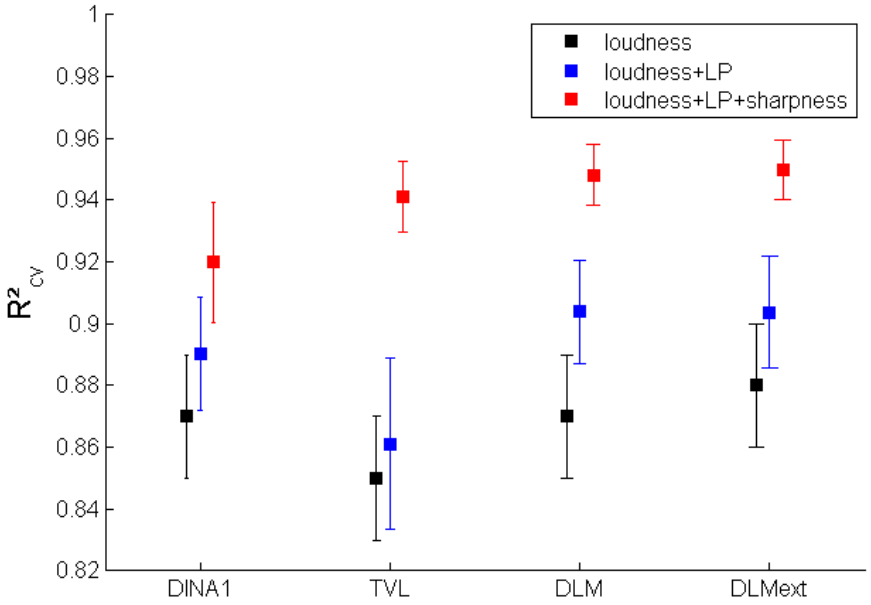

Fig. 3.4: An overview of the prediction performance represented by leave-one-out cross-validated coefficient of determination R_{cv}^2 of the loudness models: the Zwicker model DIN 45631 / A1 (DINA1), the time-varying loudness model (TVL), the dynamic loudness model (DLM) and the extended dynamic loudness model (DLMext). Modifications to the "standard" procedure of overall loudness prediction (black): a short-term loudness stage done by low-pass filtering (LP, blue) and furthermore including modeled sharpness to the prediction process (red). The error bars represent the standard deviation of R_{cv}^2.

3.4 Discussion

In this study we compared the prediction performance of different level measures and loudness models for the overall loudness of music of various genres. We considered the effect of low-pass filtering - as one implementation of short-term loudness transformation at an intermediate stage of preprocessing - on the prediction performance of overall loudness. Furthermore we included sharpness into the prediction procedure as one example for a psychoacoustic parameter that might influence the overall loudness judgement.

It has been shown that low-pass filtered instantaneous loudness resp. level is an improved basis for overall loudness estimation. This finding indicates that the fine structure of the instantaneous loudness has no big effect on the overall judgement, as determined in the BTL scaling experimnet. For the loudness models considerd here it seems to be sensible to transform instantaneous loudness into short-term loudness before using them to estimate the overall loudness. This is in line with the results of Rennies *et al.* (2015). They found that for predicting the loudness of technical sounds the time constants of loudness models should be much longer than commonly used. We confirm that this observation is also valid for music sounds.

Model, Sound Level meter	Prediction quality / R_{cv}^2		
	Loudness	+ modifications	
		+ Low-pass filtering	+ Low-pass filtering + Sharpness
DIN 45631 / A1	0,87 ± 0,02	0,89** ± 0,02	0,92** ± 0,02
TVL	0,85 ± 0,02	0,86* ± 0,03	0,94** ± 0,01
DLM	0,87 ± 0,02	0,90** ± 0,02	0,95** ± 0,01
DLMext	0,88 ± 0,02	0,90** ± 0,02	0,95** ± 0,01
dB SPL	0,83 ± 0,03	0,87** ± 0,01	0,83 ± 0,03
dBA	0,89 ± 0,03	0,93** ± 0,01	0,91** ± 0,02
dBB	0,93 ± 0,01	0,94* ± 0,01	0,93 ± 0,01
EBU R-128	0,92 ± 0,02	-	-

* $p < 0.05$ (significant) ** $p < 0.001$ (highly significant)

Table 3.1: Model comparison. Prediction quality of different loudness models and sound level measures including modified low-pass filtering and the combination with sharpness for the BTL scaled loudness judgements represented by the R_{cv}^2 as a function of statistical location parameters.

Although low-pass filtering improves the prediction of loudness models our investigations showed that the level measures considered in this study outperform them. We found that that the A- and B-weighted level and the EBU R-128 standard surpass all our provided loudness models in prediction performance. Comparing them to each other the performance was equally well. The lower performance of the loudness models against level measures is in line with the results of Skovenberg and Nielsen (2004) who observe similar poor findings for models based on the loudness model ISO 532B by Zwicker (ISO 532B, 1975). These consistent findings are rather surprising given that loudness models provide a much more accurate frequency weighting than the implementations in level measures. One explanation for this might be that, in contrast to the level measures, there is an loss of information about the temporal structure in the loudness models that has a stronger effect on the overall loudness than the inaccuracies in evaluating the spectral structure. And indeed, deficiencies in modelling the temporal loudness change by overrating the spectral structure were seen in further studies (Verhey and Kollmeier, 2002; Heeren *et al*., 2011; Rennies *et al*., 2015). But at this point it is hard to see why level measures should be closer to the real temporal structure of the loudness.

All loudness models provided improved performance by considering the estimated sharpness of the music samples as one additional parameter. Sharpness is strongly related to the spectral structure of the stimulus. Hence, this indicates that this parameter supplies the loudness models additional information about the

spectrum. On the other hand the predicition performance of level measures is not improved by this additional information.

It may also be possible that sharpness plays an important role in estimating the overall loudness of music. As mentioned before, other perceptual measures can influence the loudness judgements (Barrett, 1995; Cullari, 1989; Kantor-Martynuska, 2009). Skovenborg and Nielsen (2004) did a similar study comparing the quality of many level measures and loudness models with each other. They used speech and music stimuli with durations of 10 to 15 s and with a dynamic range similar to those used in our study. However, they used a different method with the task to adjust equal loudness between pairs of stimuli. Finally the level differences in dB were compared with the predictions. Their findings emphasised that standard frequency weighted level measures (e.g. dB(A)) show the worst performance. However, Zwicker-based loudness models were also achieving only mediocre results. A level meter similar to the EBU R-128 standard was rated as considerably better, just behind some self-developed level measures. The slight discrepancy between their finding and our results may be attributed to the selection of the stimuli. In their case the audio material widely varied from speech stimuli to music which led to a diminished performance that apparently affected the level measures.

Vickers (2010a) emphasised that level measures had done surprisingly well against the more sophisticated loudness models in many listening tests. Still, he also stated that he was limited because compromises had to be made in the choice of frequency weighting. This can be critical when the audio material is too different as argued beforeThere might be different optimal addapted frequency weightings for different sound scenarios. Furthermore, Vickers (2010a) highlighted the difficulty of estimating the loudness of music because of the wide variety of effects that can affect the loudness perception (dynamic compression, clipping, listening fatigue…). In his study he presented the most popular level measures and loudness models and showed different strategies for estimating the loudness for dynamic signals. His recommanded selection of models coincides with our choice. Besides the statistical approach of estimating overall loudness which was also used here he suggested alternatively a frame weighted averaging method where the 'long-term loudness matching level' is calculated. In some situations the statistical approach reaches its limits where the frame weighting approach is more robust, e.g. by evaluating stimuli including periods of silence. Silence gating is already implemented in the EBU R-128 standard. However, in our study stimuli were carefully selected to avoid this problem.

We already mentioned that Rennies *et al*. (2015) recently presented a similar study in which loudness models were tested for a variety of technical sounds. These sounds are non-stationary and they can be distinguished to be either dominated by temporal properties or rather by spectral properties in a much better way than music stimuli. The investigation of sounds within the transition of being dominated by spectral and temporal properties is important as for these it is especially challenging to predict their loudness. In general, temporal properties are crucial for the examination of music, but in the course of loudness compression ("loudness war", Vickers, 2010b) also spectral characteristics become a more important issue.

3.5 Summary

The main findings of the current study are:

- Low-pass filtering improves the prediction performance of the loudness models and level measures.
- Using loudness models a low-cut frequency of around 4 Hz and a percentile at 97% provide appropriate predictions.
- The A- and B-weighted sound level and the EBU R-128 standard outperform the loudness models in prediction performance.
- Including sharpness into the prediction the performance of the loudness models could be considerably improved

3.6 Appendix

3.6.1 A general numerical approach to validate the BTL-method

There are a few limitations where the BTL model cannot estimate a reasonable interval scale, for example if the set of stimuli has some natural structure (Debreu, 1960; Luce and Suppes, 1965). In most critical cases, the information used to estimate the binomial likelihood function is too small or inaccurate. A simulation of the paired comparison with random numbers can test at what point the BTL model becomes inaccurate. Therefore, real scale values N_s are randomly generated from a distribution *D* (e.g., uniform distribution). When these scale values are compared with each other, normal distributed random numbers $X(\mu, \sigma)$ are generated for each comparison. The scale values correspond to μ while σ is freely selectable parameter. In this way, a matching matrix can be generated with any desired number of repetitions N_r of the pair comparison. Hereafter, a few limiting cases of the BTL model are considered by applying parameter variation (σ, *D*, N_r). The prediction power of the BTL-model is determined by evaluation the linear fit between estimated and real scale values. The mean squared error (MSE) of the fit represents the quality of the model prediction. Furthermore, the BTL-model is compared with two other (simple) scaling models: the ordinal ranking and the average of the binomial probability $\frac{1}{n}\sum_j^n p_{ij}$ (pbin).

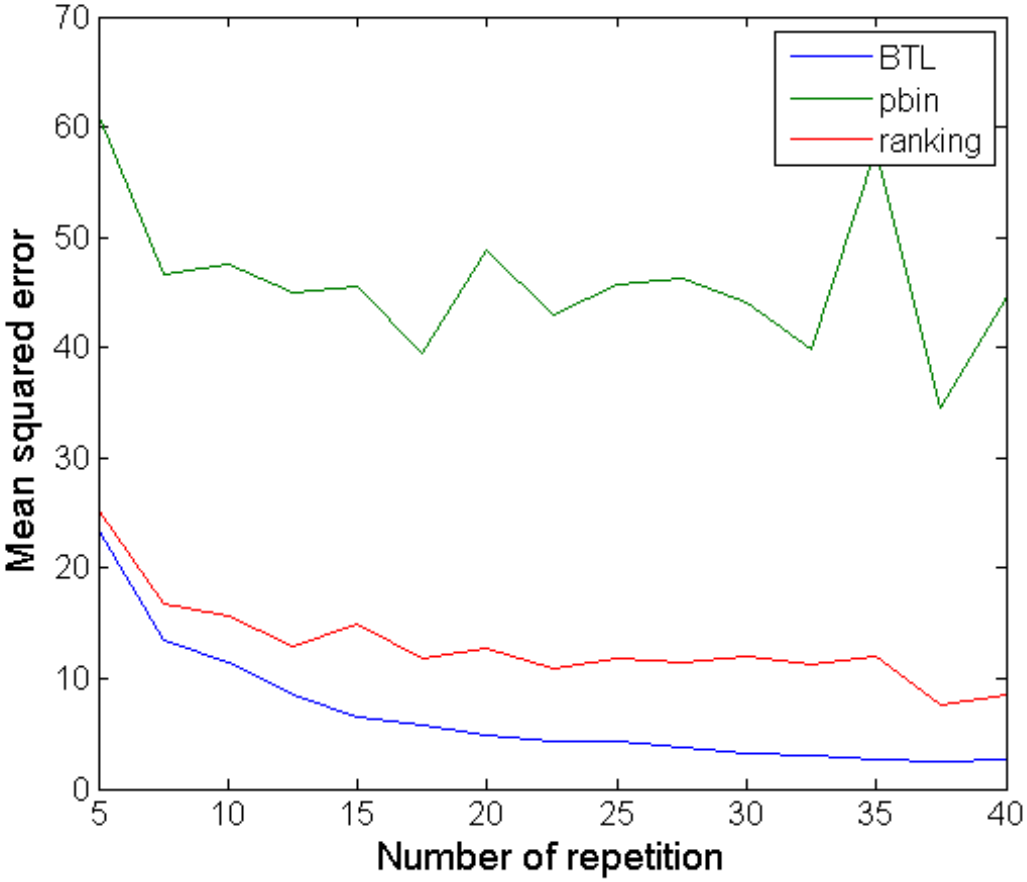

Fig. 3.5: Variation of repetitions (subjects). Comparison of the quality of different scale estimates in a pair comparison: Ranking of stimuli (green) mean binomial probability (red) and Bradley-Terry-Luce scale (blue).

Figure 3.5 illustrates the prediction power of the BTL-model by varying the numbers of repetitions. The real scale values are generated from a uniform distribution of the range of 0 to 100. Furthermore, the other parameter were set to $N_s = 20$ and $\sigma = 12$ (the unit of the standard deviation corresponds to the units of the real scale values). The MSE are the averaging of repetitions by a Monte Carlo procedure with n = 50. The result demonstrates that in all conditions the BTL-model outperforms the two simple scaling models. However, with decreasing numbers of repetitions (< 15) the accuracy of BTL-model decreases rapidly. Therefore, it can be deduced that for this procedure the number of subjects should be high enough.

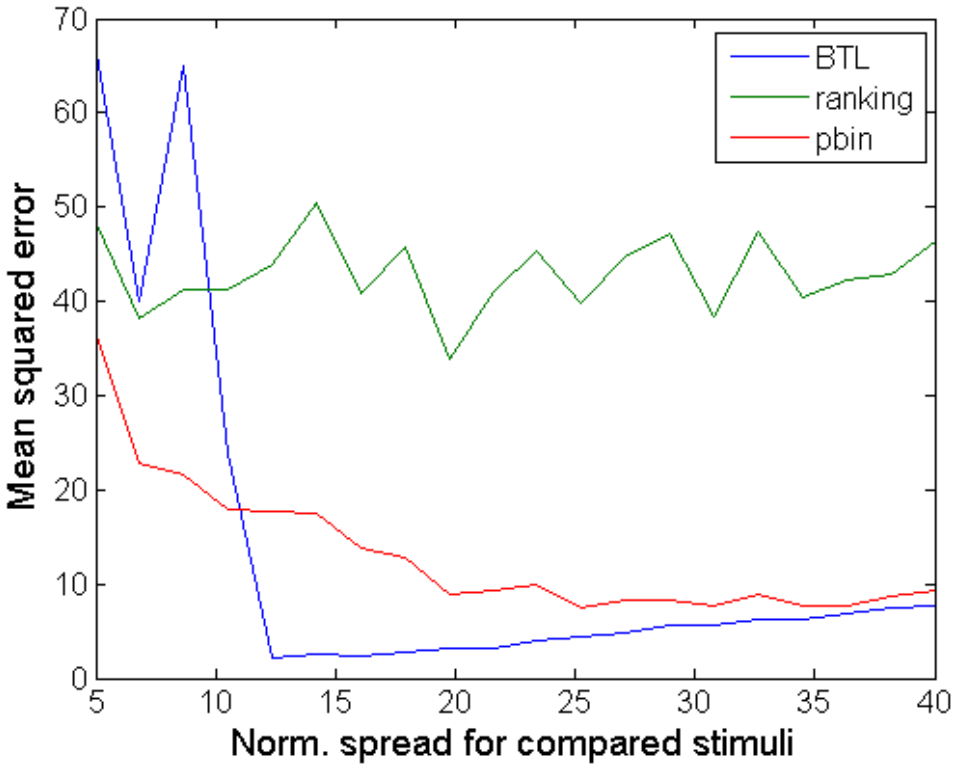

Fig. 3.6: Variation of the variance of the (normalized) real scale values. Comparison of the quality of different scale estimates in a pair comparison: Ranking of stimuli (green, ranking) mean binomial probability (red, pbin) and Bradley-Terry-Luce scale (BTL, blue).

The next simulation shows the prediction power of the BTL-model for varying variance of the scale values during the simulated pair comparison (Fig. 3.6). As before, the real scale values are generated from a uniform distribution of the range of 0 to 100 and the parameter were similarly set: $N_s = 20$ and $N_r = 30$. The MSE are averaging of repetitions by a Monte Carlo procedure with n = 50. With a low spread (< 5), the accuracy of the estimated BTL-values highly decreases. The performance of the BTL-model is even below the accuracy of the other two models. This case will occur in a paired comparison if the set of stimuli has some natural structure (s. a.). With increasing spread beyond the optimum of ~ 12, the accuracy decreases slightly due to the difficulty of discriminating the estimated scale values. This effect can be compensated by increasing the number of repetitions.

In the third simulation two different distributions for the real scale values are considered: a uniform distribution and an exponential distribution with two accumulation points. While for uniformly distributed stimuli the equidistance of the ranks in an ordinal scale does not significantly affect the quality of the prediction (Fig. 3.7A) large deviations are produced for non-uniform distributions (Fig. 3.7B). On the other hand, the BTL-method again outperforms both other methods. The MSE-statistic shows that in both cases the BTL method provides robust predictions. The binomial-based method also provides satisfactory results. However, it shows deviations at the edges of the scale for non-uniform distributed stimuli.

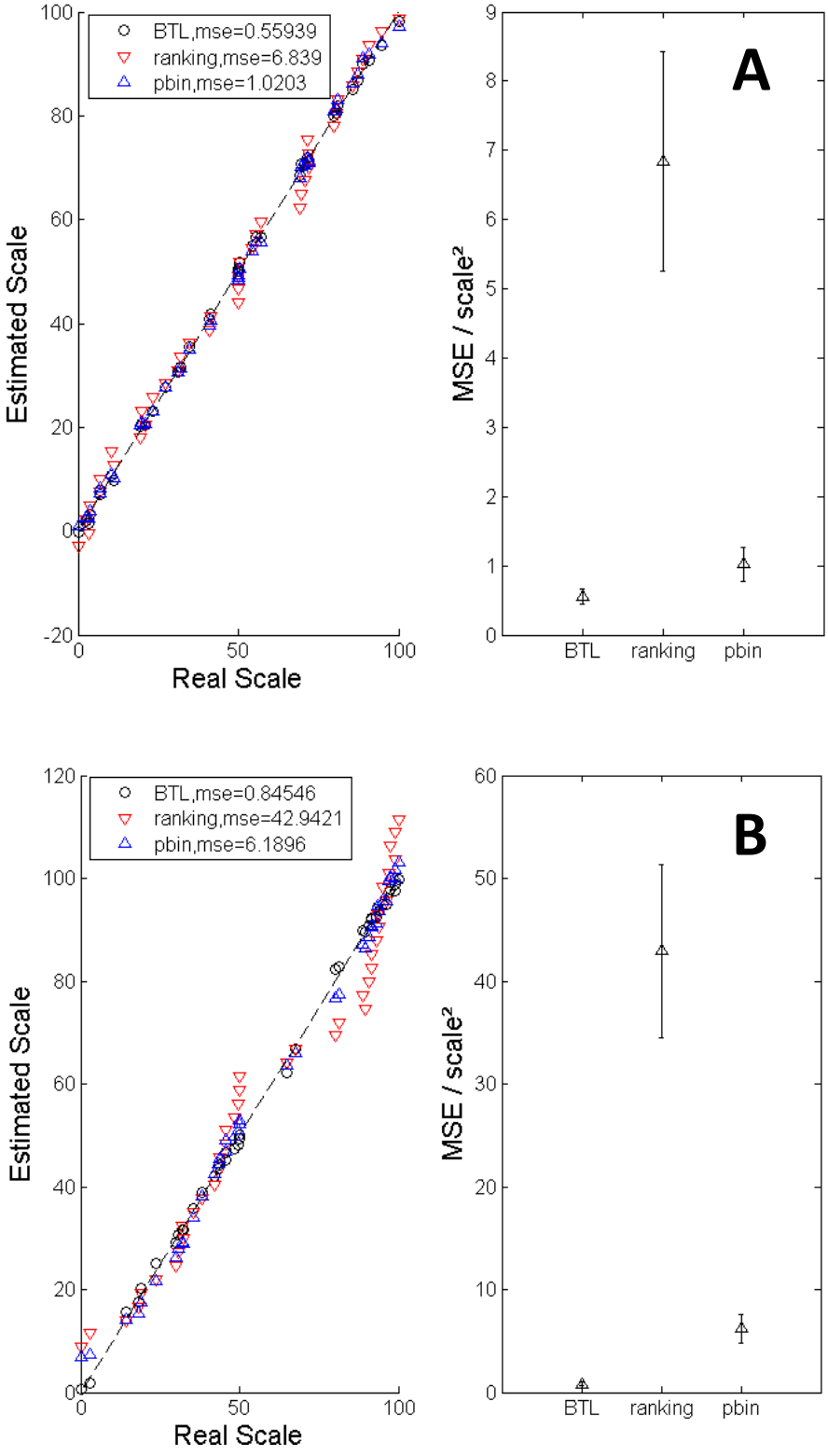

Fig. 3.7: Comparison of the distributions of the scale values of the stimuli. A. Uniformly distributed stimuli are robustly determined by all three estimators. B. Distributions with accumulation points (here two combined exponential distributions) are much less well estimated by rank order (Rang, red) or mean binomial probability (pbin, blue) than by the BTL model (BTL, black).

3.6.2 Results of the loudness scaling using pair comparison and the BTL method from Chapter 3 (amendments)

In the third chapter of this work, the overall loudness of different music sequences was scaled by a pairwise comparison. Due to the framework of the publication, some aspects could not be dealt with, which should now be made up. The experiment originally used 20 stimuli (cf. Chapter 3.2.2). However, to compare the models in a cross-validated (multiple) linear regression analysis, each modeled loudness scale must be approximately linear to the experimental loudness scale. This is only possible for a small loudness range. Hence, the number of stimuli at the edges of the scale has been reduced.

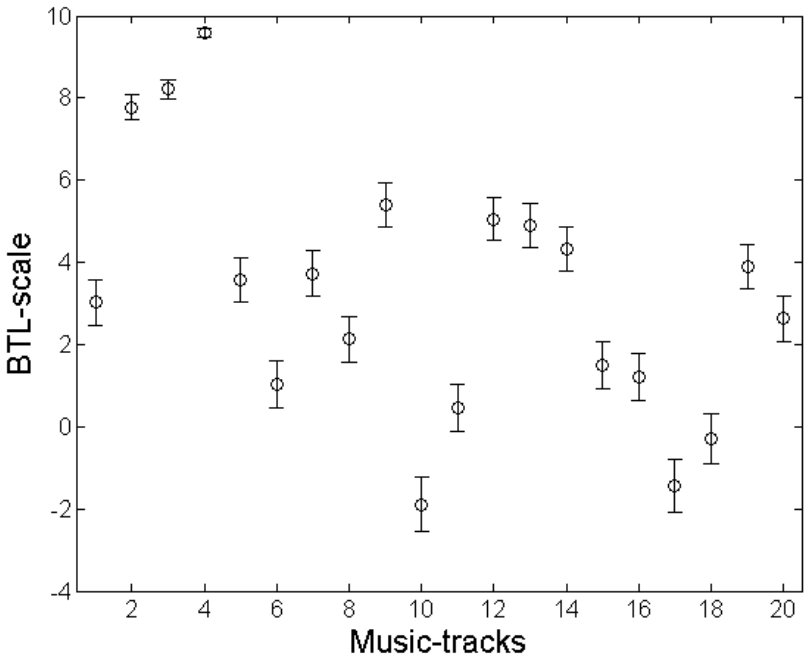

Fig. 3.8: Log-BTL scale of a pair comparison from chapter 3. The loudness of 20 different pieces of music is determined by a pair comparison. The error bars represent the standard error of the estimate.

Considering all 20 stimuli, the loudness scale shown in Fig. 3.8 is obtained using the method in Section Appendix A. The error bars can be derived by Eq. (A.5). The goodness of the fit was tested by using Eq. (A.6) with $\chi^2 = 170{,}3$ and f = 171 (freedom of degree) which results in not refusing the calculated scale. The unit of the log BTL scale is not fixed (size and zero position). Only the linearity of the scale can be assumed. Therefore, for model comparisons, the scale can be transformed arbitrarily linearly.

In Fig. 3.9, levels, the Sone-models and the CU-transformed models are compared with the BTL scale. The level measure shows a slight convex nonlinearity, while the sone-models show concave nonlinearities. In contrast, the CU-transformed models show almost a linear course. In conclusion, the results indicate that in terms of linearity categorical loudness best matches BTL loudness. Furthermore, it is well demonstrated in

Fig. A5 that for the Sone-models and the level measures linearity can only apply on the 14 stimuli in the middle loudness range.

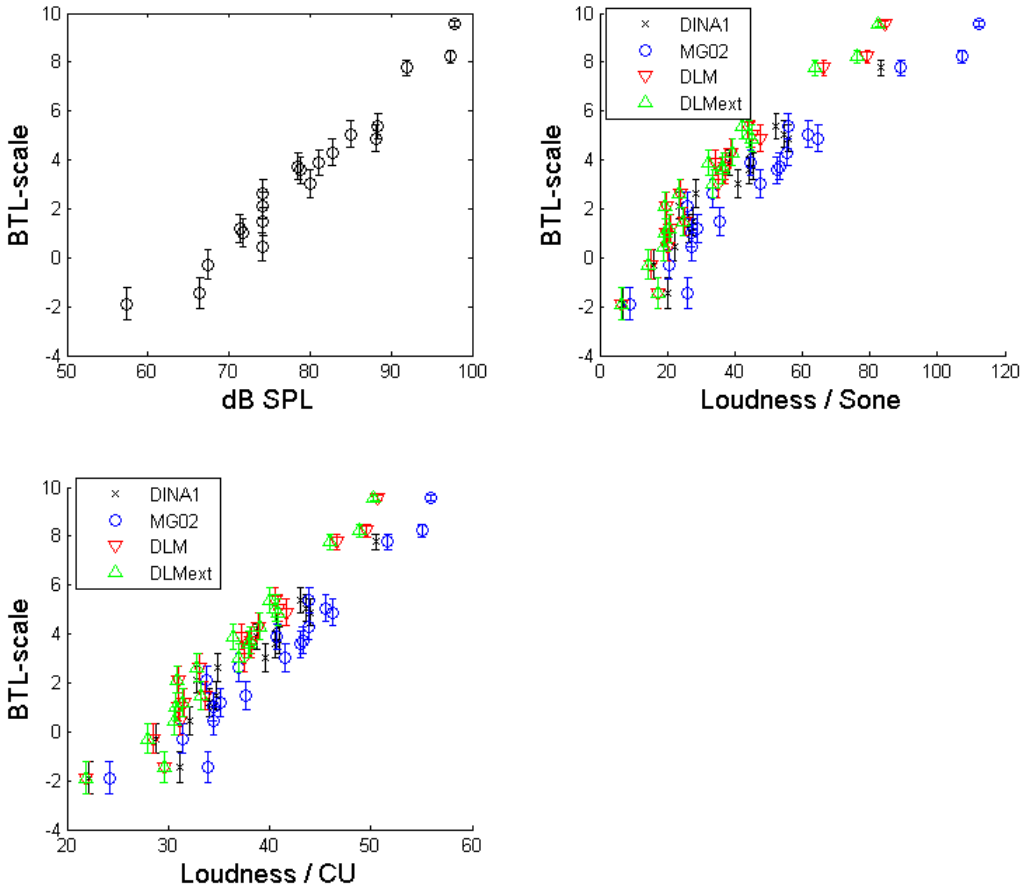

Fig. 3.9: Comparison of the continuous sound level with Sone and CU models. The level measure shows a slight convex nonlinearity, while the Sone-models show concave nonlinearities. In contrast, the CU-transformed models show almost a linear course.

4 Cortical entrainment to the loudness of music in the amplitude and latency of the envelope following response

Abstract

Recent studies have indicated that loudness is represented in the human auditory cortex by the envelope following response (EFR). However, it is not entirely clear whether there is a specific feature of the EFR which allows estimating the loudness of an arbitrary stimulus. In this study we investigate the relationship between the loudness of a music stimulus and spectro-temporal features of the EFR with regard to amplitude and latency. EEG-data of nine normal hearing subjects listening to an excerpt of a Tchaikovsky piano concert were recorded and evaluated. Using cross correlation techniques we found cortical correlates to the overall loudness around 50 ms at 11 Hz and around 150 ms at 4 Hz as well as to the instantaneous loudness of the stimulus around 145 ms at 10 Hz and around 240 ms at 2 Hz. Furthermore the amplitude long-term spectrum of the EEG response shows proportional relationship between the amplitude at the 1.25 Hz modulation frequency of the stimulus and the overall loudness. The results suggest that amplitudes and latencies of certain frequency bands may be used to estimate loudness growth for music stimuli.

4.1 Introduction

Loudness is the perceived intensity of a sound. It is assumed that the momentary stream of loudness, i.e. the instantaneous loudness, is combined to a summary assessment, determining the overall loudness of a non-stationary stimulus. In Bauer *et al.* (1974) tried to determine a perceptual correlate of loudness within the auditory evoked potential (AEP). Further studies on the topic provided results with contradictory evidence (Pratt and Sohmer, 1977; Babkoff *et al.*, 1984; Darling and Price, 1990; Serpanos *et al.*, 1997; Silva and Epstein, 2010, 2012). Most authors agreed however that sound intensity is reflected in the strength and the latency of the auditory brainstem response (ABR) (Pratt and Sohmer, 1977; Serpanos *et al.*, 1997). It was further noticed that for late evoked (cortical) potentials with latencies of 100 ms or more, thought to be associated with cognitive processes, the strength also correlates with sound intensity (Pratt and Sohmer, 1977; Hegerl *et al.*, 1994).

Nowadays, new paradigms allow evoking and analyzing the continuous EEG response which is strongly affected by the envelope of the stimulus (Picton *et al.*, 2003; Aiken and Picton 2008). In a simple scenario where amplitude modulated sinusoids evoke the envelope following responses (EFR), the modulation frequency and their harmonics are present in the long-term spectrum of the EEG. This response is also called auditory steady state response. Several studies showed an interdependence between EFR and categorically scaled loudness (Ménard *et al.*, 2008; Castro *et al.*, 2008; Emara and Kolkaila, 2010; Eeckhoutte *et al.*, 2016).

Beyond that, EFR are also found for speech (Thwaites *et al.*, 2016, Aiken and Picton, 2008; Ding and Simon, 2014; Liberto *et al.*, 2015) and music stimuli (Doelling and Poeppel, 2015). For example, Doelling and Poeppel (2015) found in an MEG study cortical entrainment while their participants were listening to music. They demonstrated that the modulation frequency of the stimulus needs to be at least 1 Hz or above to allow for an MEG correlate of cortical entrainment.

Although there are a substantial number of publications dealing with the EFR, the analysis of those responses has to be handled with care because their shape varies widely across subjects and for different stimuli. On the other hand there is no "gold standard" in measuring loudness (magnitude estimation, comparison, categorical scaling) and also, there is up to now no general model of loudness perception available that is valid for all types of complex stimuli. The latter is well reflected by misjudgments of currently used loudness models (e.g. Rennies *et al.* 2015; Skovenborg and Nielsen, 2004). Central processing beyond the cochlea like context effects (Arieh and Marks, 2003a) may also contribute to the general discrepancy between empirical data and modelled loudness. In summary there is no complete understanding of the loudness of complex sounds as well as of their evoked EEG response.

Common features of the auditory evoked potentials are the latencies and amplitudes of the various peaks of the evoked potential generated along the ascending auditory pathway and by processing in auditory cortex. In this study we aim to find features of the envelope-following response that correlate with the envelope or the overall intensity of a music stimulus. We investigate the energy of the EEG-response at several latencies and in several frequency bands. For this, a short music sequence with a certain periodicity in acoustic features is used as a stimulus. It has a high dynamic range within the sequence with steep attack times on each beat and hence with the expectation to record distinct stimulus related features in the EEG-response. One question is also whether these EEG features are actually correlates of loudness perception or possibly simply reflect the changes in sound intensity. Different measures in terms of sound level as well as in terms of loudness will be applied which reflect on the one hand sound intensity and on the other hand loudness. Correlation coefficients will be compared to assess which of the applied measures is best represented in the EEG response. This approach somehow follows the study by Thwaites *et al.*, 2016 who were comparing the correlation coefficients between MEG-responses to running speech and short-term and long-term loudness determined by a loudness model to identify the areas of loudness processing in the brain.

4.2 Materials and methods

4.2.1 Stimulus

The auditory stimulus used in the experiments was a music excerpt of the piano concert in B flat minor No 1 of P. Tchaikovsky (Fig. 4.1a), taken from an audio CD recording. The duration of the stimulus was 20 s. At the

begin and the end Hanning ramps of 30 ms length were used to reduce sharp on- and offsets. The tempo of the orchestral per¬formance was 75 beats per minute which corresponds to 1.25 Hz (Fig. 4.1b).

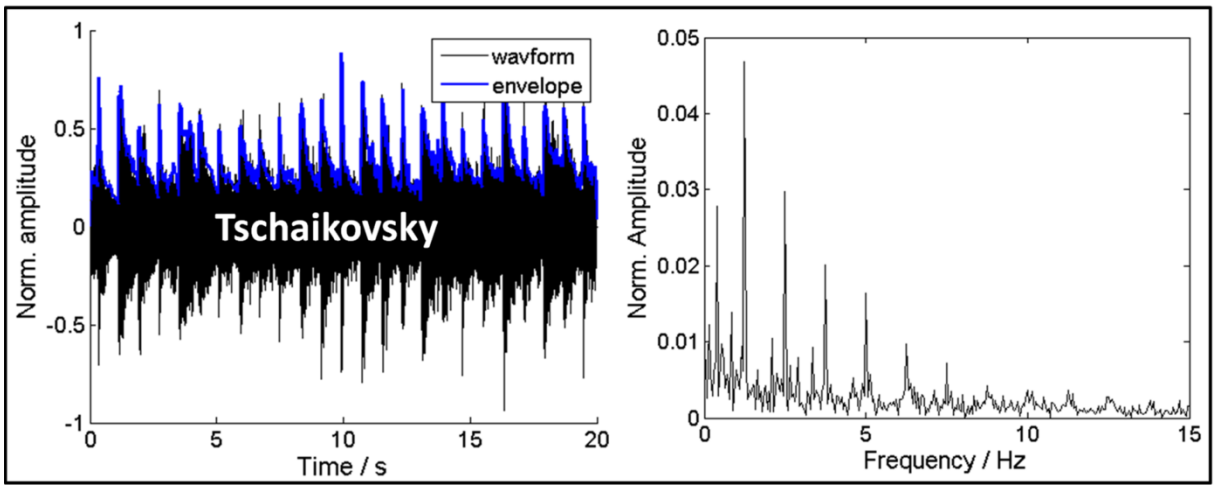

Fig. 4.1a-b: a: Wavform and modelled loudness (employing DLM) of the stimulus: a music excerpt (measures 13-21) of the first movement of the piano concert in B flat minor No. 1 of P. Tchaikovsky. b: Amplitude modulation spectrum of the stimulus exhibiting a peak modulation frequency at 1,25 Hz and its harmonics.

The stimulus was converted via a digital-analog converter(RME ADI-8 DS, sampling frequency: 44.1 kHz) and presented diotically using an RME-Digi 96-8 PAD soundcard in six different level conditions (40, 50, 60, 70, 80 and 90 dB SPL) and digitally attenuated with MATLAB (7.3.0 R2006b). The stimulus presentation was performed via a Trucker-Davis HB7 headphone amplifier and insert tube- headphones (Etymotic Research ER-2). For comparability with previous ASSR-loudness studies (Ménard *et al.*, 2008; Eeckhoutte *et al.*, 2016), a block paradigm was chosen. Each block represents 50 repetitions of one condition. A pause between the repetitions was set at 500 ms with a maximal time-jitter of 50 ms. The order of the six blocks was pseudo randomized for each subject.

4.2.2 EEG data acquisition

The EEG-response was recorded with a Biosemi Active Two system using 68 channels with the electrodes placed according to the international 10-20 system, using an electrode cap. Contact gel (Parker typ no. 15-25) was used to ensure good contact between electrodes and scalp. The recordings were collected and digitalized with ActiView (6.03) at a sampling frequency of 1024 Hz without using any filters.

4.2.3 Subjects

Nine volunteers, five male (S2, S3, S4, S5, S7) and four female (S1, S6, S8, S9), with normal hearing participated in the experiments. All were right-handed, between 20-30 years old, and they were paid for their

participation. During the EEG recordings, the participants watched a silent movie of their choice. They did not have a specific task to perform.

All experimental procedures were approved by the ethics committee of the University of Oldenburg.

4.2.4 Data processing

All data were processed offline using MATLAB Version R2011b and the corresponding signal processing toolbox. The averaged responses of the A1- and A2-electrode were used as virtual reference between the mastoids. The response at the Cz-electrode was considered for further data evaluation. After high pass filtering at 0.3 Hz to reduce the electrode drift, the data were downsampled to 256 Hz sampling frequency. The data were then separated into epochs of 20 s length and a baseline correction was performed by subtracting the averaged response of the 100 ms before each epoch.

Usually epochs affected by artefacts have to be removed because the arithmetic average is vulnerable to high amplitudes. Reducing the number of epochs, however, would also reduce the signal to noise ratio. In our study we used a spectral averaging method introduced by Vardi and Zhang (2000) applying the multivariate L1-median to reduce artefacts:

$$F_n(\omega) = FFT\big(x_n(t)\big) \tag{4.1}$$

$$L_1(\omega) \rightarrow \frac{1}{N}\sum_n(\|F_n(\omega) - L_1(\omega)\|) \tag{4.2}$$

$$\langle x(t)\rangle_{L_1} = iFFT\big(L_1(\omega)\big) \tag{4.3}$$

In Eq. 4.1 $F_n(\omega)$ is the discrete complex spectrum of the n-th epoch of the time signal $x_n(t)$ provided by the Fast Fourier transform algorithm (*FFT*). The L1-median $L_1(\omega)$ is the minimized averaged distance between $F_n(\omega)$ in the complex plane (Eq. 4.2). Finally, the averaged time signal (Eq. 4.3) can be calculated by transforming $L_1(\omega)$ back into the time domain using the inverse FFT (*iFFT*). We used the L1-median algorithm implemented in the the LIBRA Matlab-toolbox (Verboven and Hubert, 2005). Essentially, this averaging method is extremely robust with regard to artefacts, and it allows keeping epochs which are affected by artefacts.

4.2.5 Statistical analysis

Four different measures for sound intensity or loudness, respectively, were examined in this study. In addition to the sound pressure level (dB SPL) we applied two further quantities based on a certain frequency weighting: the A-weighted sound pressure level (dBA) and the EBU R-128 standard (European Broadcasting Union) that has been recommended to predict the loudness of music (EBUR-128, 2014). For the different

measures based on the level we used a 125 ms time constant ('fast') to derive an envelope. For the overall level the root mean square (RMS) over the full range was applied. A more physiological approach to predict the loudness would be to use models that take effects of spectral and temporal loudness integration into account. Here the dynamic loudness model (DLM) developed by Chalupper and Fastl (2002) was used. This loudness model provides an estimation of the time course of loudness (instantaneous loudness).

A common approach to analyze EEG-data is to measure a stimulus related response (e.g. auditory evoked potentials, event related potentials or in our case: the EFR), identify specific components and then investigate the changes of the amplitude and the latency of these components by changing the stimulus condition.

In this study the data were analyzed and evaluated with respect to their relationship to loudness in three ways: A) EFR-amplitude and overall loudness B) EFR-amplitude and instantaneous loudness C) EFR-latency and overall loudness.

A) EFR-amplitude and overall loudness

The EFR for music stimuli is dominated by certain frequencies (Doelling and Poeppel, 2015). Mainly the fundamental frequency of the stimulus envelope and its harmonics can be found in the EFR. It is therefore appropriate to investigate the strength of these frequencies in the EEG long-term spectrum to find the relationship between EFR and overall loudness. This idea is encouraged by several studies showing a relationship between loudness and the amplitude spectrum by use of sinusoidally amplitude-modulated stimuli (Ménard *et al.*, 2008; Castro *et al.*, 2008; Emara and Kolkaila, 2010; Eeckhoutte *et al.*, 2016). Finally the significance of the relationship can be tested by performing an ANOVA across sound level conditions and amplitudes over the subjects.

B) EFR-amplitude and instantaneous loudness

The relationship between the EFR amplitude and the instantaneous loudness can be investigated by analyzing the temporal envelope of the loudness of the stimulus and its corresponding EEG response. The EFR is expected to follow largely the beats of the music excerpt because of their sharp onset shape. Now measuring the amplitude peaks in different time frames provides the requested characteristic quantities for the evoked peaks in the EFR. However, according to the complexity of the EEG response to continuous stimuli and due the low number of available epochs caused by the comparatively long stimulus duration, a specific procedure for data analysis is required. In this study a new approach is introduced to perform this task.

Time-frequency analysis was applied to the current EEG-signal (z) to create a spectrogram $x^{(\Delta f)}(t, f)$ by calculating the envelope of frequency bands between f = 3 Hz and f = 28 with Δf = 5 Hz bandwidth (bandpass filtering (BP$*$…) by doing forward-backward-filtering with an elliptical filter) using the absolute value of the Hilbert Transformation (|H{…}|) (Eq. 4.4).

$$x^{(\Delta f)}(t, f) = |\mathrm{H}\{\mathrm{BP}(f, \Delta f) * z\}| \tag{4.4}$$

This procedure is also illustrated in Fig. 4.2. To create the cross spectrogram$\rho(\tau, f)$, $x^{(\Delta f)}(t, f)$ is correlated with the stimulus loudness $y(t)$ (Eq. 4.7). On the stimulus side for the correlation we only consider the maximum loudness values y_i (Eq. 4.5) which are the onset peaks at $t_i = t_{i-1} + 800\ \text{ms}, t_0 = 330\ \text{ms}$ when the beat occurs (see Fig. 4.2 at the top left).

$$y_i = \max\big(y(n)\big), \{n\epsilon[t_i - 15\ ms, t_i + 15\ ms]\}, i = 1,\ldots,25. \tag{4.5}$$

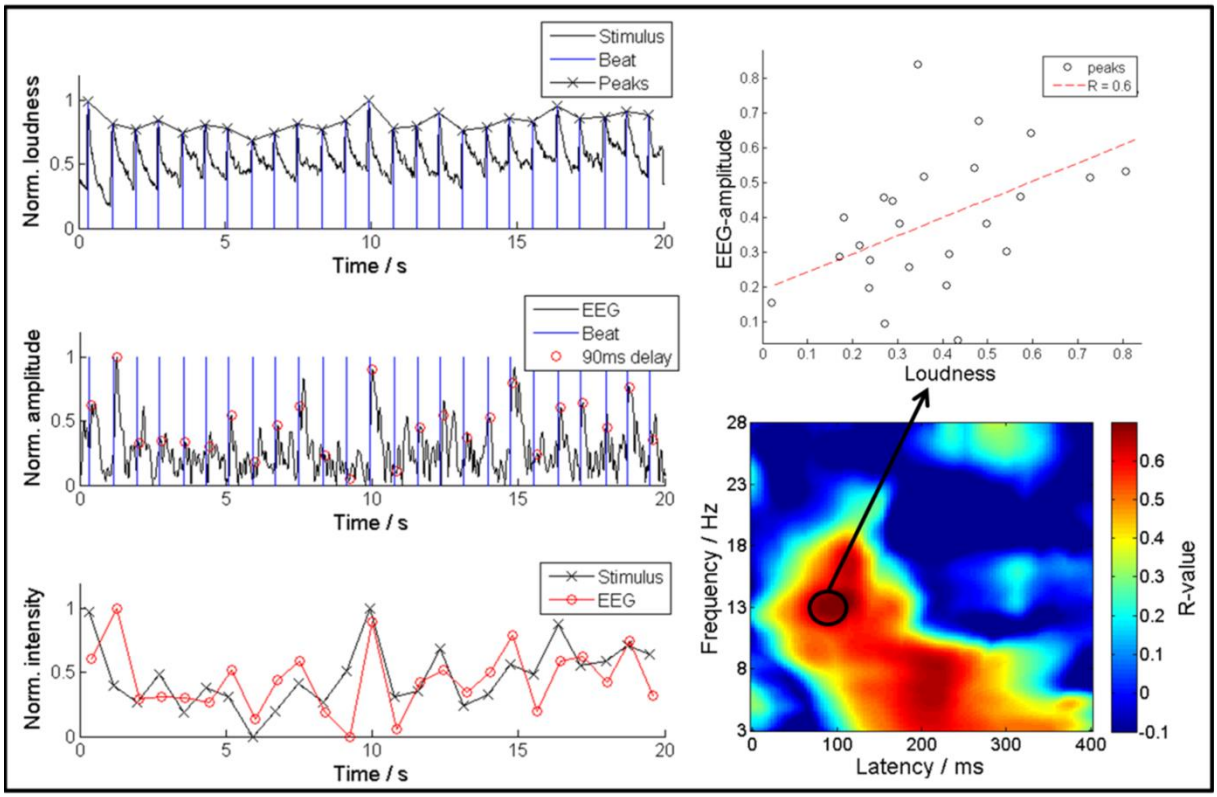

Fig. 4.2: Procedure to create cross spectrograms. Loudness peaks at the beat (top left) were exemplary correlated with their corresponding EEG peaks extracted from a 30 ms time frame 90 ms delayed to the beat and bandpass filtered between 11,5 Hz and 15,5 Hz (middle left). These loudness peaks and EEG amplitude peaks are plotted over time (bottom left). The calculated Pearson correlation coefficient R (top right) is used to create a cross spectrogram for different delays (latency) and frequency bands (bottom right).

On the EEG side we extract the same number of amplitude values but only those which are temporally attributed to the beat by a physiologically reasonable delay τ of not more than 400 ms. Now, for each frequency band amplitude peaks $x_i(\tau, f)$ (Eq. 4.6) in a short time frame of 30 ms were collected.

$$x_i(\tau, f) = \max\left(x^{(\Delta f)}(n, f)\right), \{n\epsilon[t_i + \tau - 15\ ms, t_i + \tau + 15\ ms]\}, i = 1,\ldots,25. \tag{4.6}$$

In this way *N* = 25 EEG amplitude peaks and stimulus onset peaks were correlated with each other giving a correlation coefficient $\rho(\tau, f)$ (Eq. 4.7). But as mentioned above this procedure was repeated for different frequency bands *f* and time frames *τ*.

$$\rho(\tau, f) = \sum_{i=1}^{N}(x_i(\tau, f) - \mu_x) \cdot \big(y_i - \mu_y\big) \Big/ \sigma_x(\tau, f) \cdot \sigma_y, \text{ with} \tag{4.7}$$

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i(\tau, f) \text{ and in a similar way to } \mu_y, \sigma_x(\tau, f) \text{ and } \sigma_y. \quad (4.8)$$

C) EFR latency and overall loudness.

In a similar way as in section B) EFR latency was examined by considering cross spectrograms. This time, however, we considered the entire temporal information $y(t)$ and $x^{(\Delta f)}(t, f)$ for correlation and not only the maximum loudness peaks. In this way *N* = 256*20 samples of the EEG signal and the stimulus signal were correlated with each other, respectively with repetitions for different frequency bands *f* and time frames *τ*. Apart from this the signal processing to provide the cross spectrograms was done just the same as above (Eq. 4.9).

$$\rho(\tau, f) = \sum_{i=1}^{N}\left(x^{(\Delta f)}(i+\tau, f) - \mu_x\right) \cdot \left(y_i - \mu_y\right) \Big/ \sigma_x(\tau, f) \cdot \sigma_y \quad (4.9)$$

By varying the sound level condition we expect to derive cross spectrograms with almost similar correlation patterns. Comparing these patterns for temporal shifts provides a relationship between EFR latency and overall loudness represented by the sound level conditions. Finally by performing an ANOVA between sound level conditions and the latency shifts over the subjects the significance of the relationship can be tested.

4.3 Results

4.3.1 EEG-response to the stimulus

Figure 4.3a shows the individual long-term amplitude spectrum of the EEG response to the stimulus at 70 dB SPL. For all subjects, the spectrum is dominated by a fundamental frequency at 1.25 Hz and their corresponding harmonics. Note that the fundamental frequency is identical to the tempo of the music (75 beats/min). This provides evidence that the EFR has a periodic structure in this case. Similar observations were made for the EEG responses at the other sound levels employed.

In Fig. 4.3b a short time segment of 450 ms duration of the EEG response for each beat is shown (-50 to 400 ms) for subject 5. The 25 beats of the music excerpt are plotted over time where 0 ms represents position of the beat. Within the subjects the EFR to every beat shows resemblances to each other. It is striking that the EFR latency tends to decrease with ongoing stimulation.

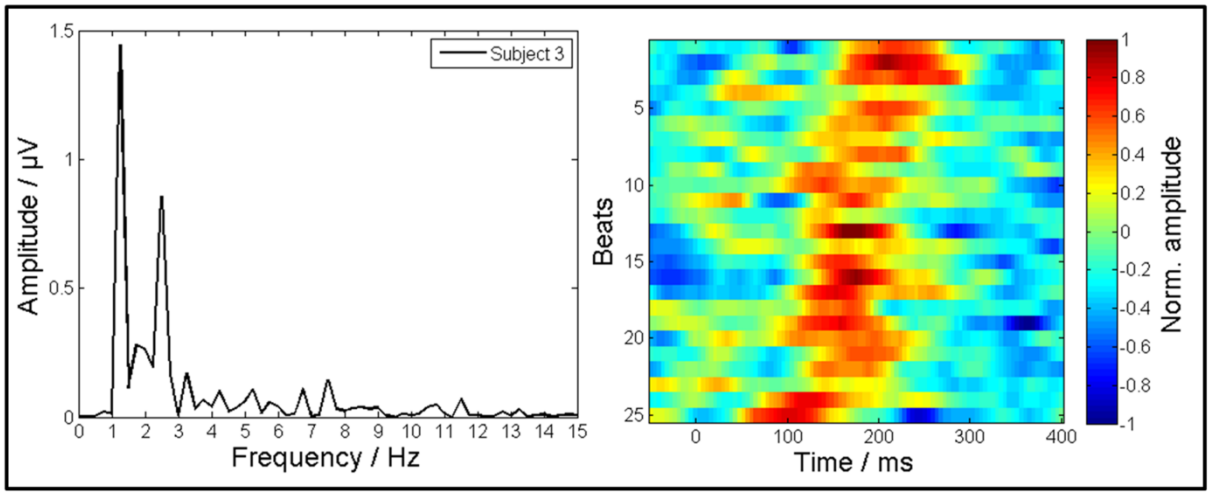

Fig. 4.3a-b: a: Amplitude long-term spectrum of the EEG-response at 70 dB SPL RMS (top left) of subject 3 b: EEG-response to the beats of the stimulus averaged over 50 epochs of subject 5. The color represents the normalized amplitude of the EEG response (top right).

4.3.2 Instantaneous loudness dependency of the EEG amplitude

Figure 4.4 shows the cross spectrograms for the four different measures for sound intensity resp. loudness described in the methods section: Sound pressure level, A-weighted sound pressure level, Dynamic loudness model and the EBU R-128 standard. Each cross spectrogram depicts the correlation between the respective measure and the EEG response for different latencies and frequency bands. The results indicate that there are two components in the EFR which show significant correlation: (DLM) $R_{140\,ms,9\,Hz}$ = 0.43, p = 0.000016, $R_{236\,ms,5,6\,Hz}$ = 0.41, p = 0.00039 with the modelled loudness of the stimulus (top right)

whereas there was no significant correlation with the sound pressure level (top left). Frequency-weighted levels like dBA (bottom left) and the EBU R-128 standard (bottom right) achieved similarly good results dBA: $R_{144\ ms,9\ Hz}$ = 0.47, p = 0.0000071, $R_{240\ ms,5,7\ Hz}$ = 0.42, p = 0.00039; EBU R-128: $R_{132\ ms,9\ Hz}$ = 0.44, p = 0.00012, $R_{240\ ms,5,6\ Hz}$ = 0.44, p = 0.00019. Generally speaking the early component that correlates with loudness occurs in the alpha-band of the EEG signal around 140 ms and 9 Hz. The late component occurs in the theta-band of the EEG signal around 240 ms and 5.6 Hz.

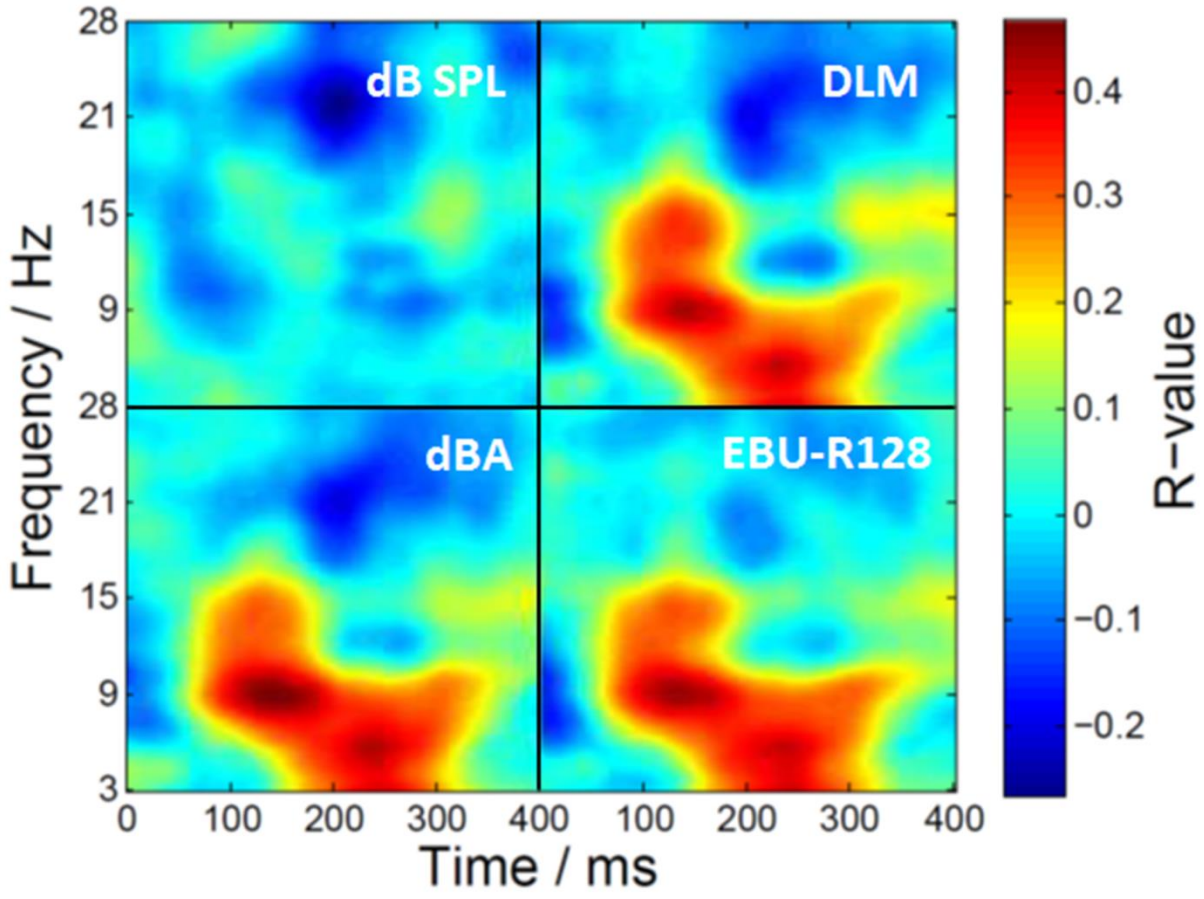

Fig. 4.4: Cross spectrogram between EEG-response and stimulus loudness: a) dB SPL b) DLM c) dBA and d) EBU R-128.

4.3.3 Dependency of the long-term amplitude spectrum on overall level

Figure 4.5 shows the relationship between the sound pressure levels of the stimulus (40 dB, 50 dB, 60 dB, 70 dB, 80 dB, 90 dB) and the overall amplitude of the EFR represented by their fundamental frequency at 1.25 Hz averaged over the nine participants and 50 repetitions. With increasing level the EFR-amplitude increases too, but decreases at higher levels (ANOVA: F = 5.919, p = 0.0002). Other frequencies of the long-term spectrum did not show significant differences between their amplitude and sound pressure level.

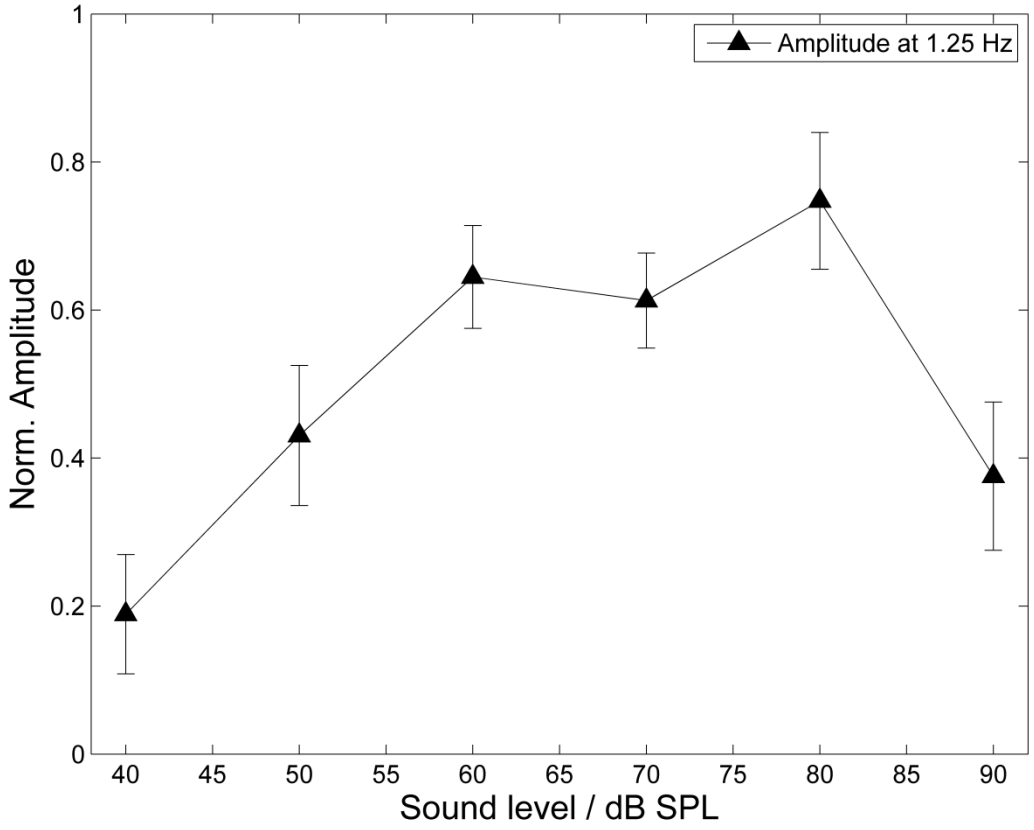

Fig. 4.5: Amplitude at 1.25 Hz of the long-term spectrum of the EEG-response averaged over 9 subjects and 50 epochs as a function of sound pressure level of the stimulus. Error bars indicate standard errors. The amplitude was normalized for every subject.

4.3.4 Dependency of the latency of the EFR on overall level

Figure 4.6a shows cross spectrograms between stimulus envelope and the respective EEG response of an exemplary subject at two different levels (50 dB and 80 dB). The red and the black circles mark a significant correlation at two positions in the time-frequency plane between stimulus envelope and EEG response. These correlating components can be extracted for all subjects and conditions and can be put into context with the overall level of the stimulus (Fig. 4.6b). The red circles are representing the component around 50 ms and 11 Hz, the black circles the component at 150 ms and 4 Hz. With increasing level the latency of these two components decreases significantly (ANOVA: $F_{50\ ms,11\ Hz}$ = 3.01, p = 0.0027, $F_{150\ ms,4\ Hz}$ = 2.55, p = 0.0477).

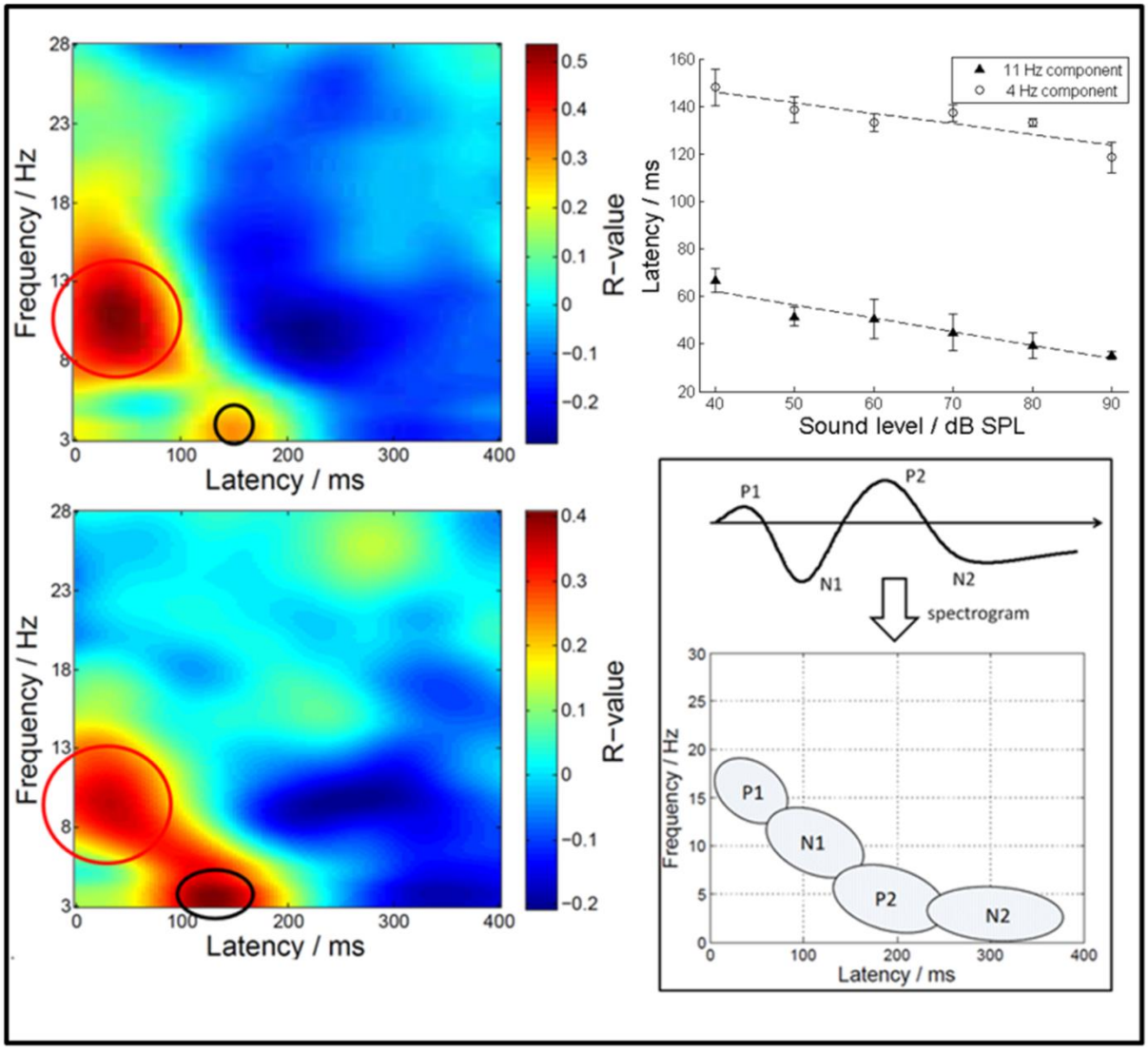

Fig 4.6a-c: a) Cross spectrograms between stimulus envelope (exemplary at 50 dB and 80 dB) and their EEG responses (left). Two correlating components are marked (red and black circles) b) The latencies of the correlating components over sound intensity of the stimulus. Error bars indicate standard errors. (top right) c) schematic representation of late auditory evoked potentials temporal and spectro-temporal (bottom right).

4.4 Discussion

We analysed the amplitudes of the EFR as a function of loundess change within the stimulus and as a function of six different sound level conditions, by evaluating the EFR long-term spectrum. The results indicate that the EFR amplitude increases with increasing sound level from 40 dB to 80 dB at the modulation frequency of 1.25 Hz in the EEG long-term spectrum, as expected. At higher levels, however, the intensity of the amplitude seems to saturate. Up to a certain point saturation is expected because the overall EEG amplitude is limited by saturation of firing neural activity.

Further the correlation between EFR amplitude and stimulus intensity could be shown by evaluating their cross spectrograms as demonstrated in Figure 4.4. Two correlating components were identified, one around 145 ms at 10 Hz, and a second one around 240 ms at 2 Hz. These components can be associated with cortical neural activity due to their comparatively long latencies. This result is in line with some earlier studies that show that the amplitude of cortical evoked potentials is sensitive to changes in sound intensity (Hegerl *et al.*, 1994). Comparing the cross spectrograms (Fig. 4b-d) with an idealized spectrogram of an auditory evoked potential as sketched in Fig. 4.6c, the components identified with our procedure are in line with the spectro-temporal appearance of AEP. The first component can be associated with the N1-P2 complex of the AEP whereas the second component is linked with the P2-N2 complex.

By evaluating the cross spectrograms between stimulus envelope and the EFR spectrogram as demonstrated in Figure 4.6 for all conditions a correlation between overall level and the latencies of two EFR components was demonstrated. These two components were found around 40-70 ms at 11 Hz and around 120-150 ms at 4 Hz (Fig. 6a). Again, the components may be associated with the AEP complexes: the first component with the P1-N1 complex in AEP whereas the second component with the N1--P2 complex (Fig. 6c). We cannot rule out that in this analysis using the overall level we are dealing with the same components as in the analysis employing the instantaneous level resp. loudness, irrespective of the difference in latencies as the deviations may simply be caused by the different approaches of the two methods. Nevertheless, it appears that there are three different components that correlate with the level resp. loudness. Both methods were able to detect one common component (N1-P2) and each of them one unique component (P2-N2, P1-N1). However, the cross correlation of the time signal of the whole stimulus with its corresponding EEG response holds the risk to detect only correlating activation of certain distinct events in the stimulus rather than a relationship to the stimulus loudness. Therefore we prefer the method making use of the processed and thereby reduced data to correlate only values that represent the loudness of the stimulus in the corresponding activation.

Our results confirm the results of previous studies that the amplitude of the EFR correlates with level resp. loudness (Ménard *et al.*, 2008; Castro *et al.*, 2008; Emara and Kolkaila, 2010; Eeckhoutte *et al.*, 2016). However, most of these studies dealt with simple stimuli like amplitude modulated sinusoids (e.g. Ménard *et al.*, 2008; Eeckhoutte *et al.*, 2016). To our knowledge, our data represent the first demonstration of loudness correlation with a rather complex music stimulus for a parametric change of stimulation level. But our findings match closely with previous studies. The dominance of cortical components in the EFR is in line with the observations of Doelling and Poeppel (2015). The modulation frequency of our music sequence was 1.25 Hz which is above this limit suggested by Doelling and Poeppel (2015). Beyond that we were able to show that this cortical response correlates with the stimulus loudness and can be separated in two components.

Other studies also suggested that loudness perception is primarily done cortically (e.g. Röhl and Uppenkamp, 2012; Behler and Uppenkamp, 2016). They showed in an fMRI study that the correlation between the BOLD signal and perceived loudness increases at later stages of the ascending auditory pathway and reaches its

peak at the posterior medial Heschl's gyrus. Similar observations were done by Thwaites *et al.*, 2016. They found in an MEG study several cortical components corresponding to the loudness of speech. They used cross correlation to find the best correlating latencies and found four components at 45 ms, 100 ms, 165 ms and 275 ms. Furthermore they were able to prove that the latest component shows higher correlation to the short-term loudness (i.e. low pass filtered) than to the instantaneous loudness. Hence they concluded that the latest component represent a percept that already reflects processing at a higher stage in the hierarchy of loudness processing. The loudness related latencies in our study were in the same range as the ones reported by Thwaites *et al.* (2016) although we could identify only two components instead of four. This might be owed to the use of EEG instead of MEG data in our study, the difference in length and type of the stimulus, or the difference in the processing strategy of the data.

4.5 Conclusion

The main findings of the current study are:

- There is a relationship between the overall level of the stimulus and the amplitude at 1.25 Hz of the EFR long-term spectrum. The EFR amplitude increases with increasing sound level from 40 dB to 80 dB and seems to saturate from 80 dB to 90 dB.
- Two components of the cortical response are correlated to the instantaneous stimulus loudness, one around 145 ms at 10 Hz, and a second one around 240 ms at 2 Hz.
- Without any frequency weighting the envelope of stimulus is not correlated to the strength of the cortical response.
- The overall level is also correlated to the latency of two components of the cortical response, one around 50 ms at 11 Hz and a second one around 150 ms at 4 Hz.

In conclusion, the current data suggest that the main factor on which level resp. loudness correlates in EEG data are based on is the type of frequency weighting employed. The finding that even a comparatively simple measure like e.g. EBU R-128 results in a reasonable correlation with recorded EEG responses might at some stage allow for the implementation of an EEG-based brain-computer interface, which may be of use in an adaptive hearing aid with automatic adjustment of amplification, based on the perceived loudness of fluctuating stimuli.

4.6 Appendix: Neural loudness processing of perceived music by ERP

Previously, the EEG response to music was investigated to find neural loudness correlates. A sequence from a classical piece of music lasting 20 s was presented at 6 different sound levels (cf. Chapter 4.2.1). Furthermore, in Chapter 4.2.5B the amplitude of the envelope following response (EFR) was correlated to the instantaneous loudness of the stimulus at different frequency bands and latency shifts. However, only one level condition was considered in this study. In the concluding discussion in Chapter 4.4, it was pointed out that EFR is dominated by auditory evoked potentials (AEP). This is not surprising. It is well studied that music often elicits several features that are measurable by event related potentials that correspond to the cortical deflections of the AEPs. In this section, it is examined whether the peak amplitude of the P1, N1, P2 and N2 of the ERP of all 6 conditions correlates with the loudness of the eliciting events.

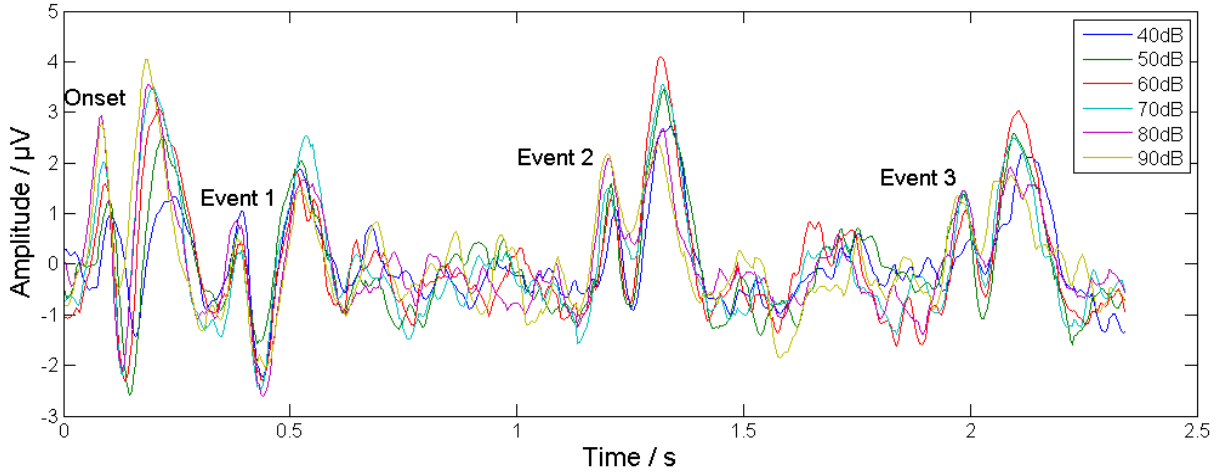

Fig. 4.7: EEG response (0 - 2.5 s) to the stimulus for all levels. Stimulus onset and three events on the beat characterized by high loudness change elicit an ERP. The ERP of the first event is superimposed by the N2 of the onset ERP.

4.6.1 Level versus loudness

The first 4 ERPs of the 6 different conditions elicited by the stimulus are illustrated in Figure 4.7. The first ERP is the onset response, i.e. the response of the beginning of the stimulation. After that, mainly the beat generates the subsequent ERPs. It can be seen from the onset response that both latencies and the amplitude peaks of the individual potentials depend on the stimulus level. For further analysis, the peaks of the potentials for each event are extracted. The onset response is not used for the analysis. Furthermore, the ERP of the first event is not used because the ERP is adapted by the onset response. After peak extraction, the amplitude values of each potential can be correlated with the calculated levels and the modeled loudness (Dynamic loudness model) of these events. In Figure 4.8 a comparison between level and loudness is shown by illustrating their correlations with the individual potentials. Significant correlations were mainly found for

correlations between the loudness model and the amplitude of N1 and P2 potentials. The R-values for each condition and the corresponding p-values can be found in Table B1.

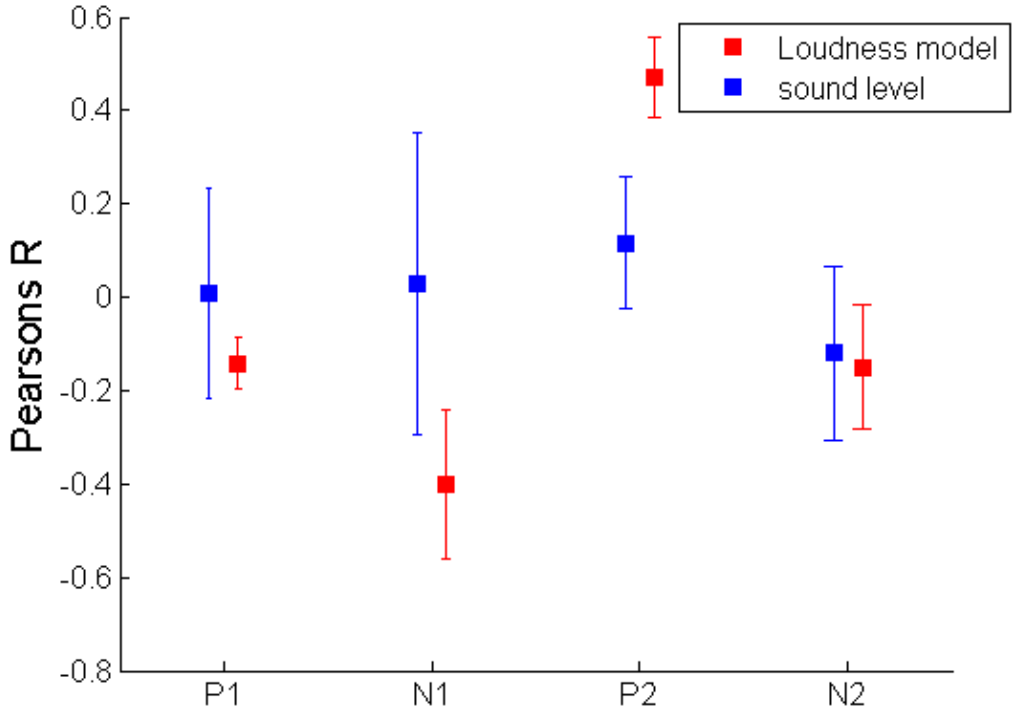

Fig. 4.8: Comparison of the correlation of level and loudness model with the peak amplitude of the cortical potentials of the ERP (P1, N1, P2 and N2). The error bars represent the standard deviation across the conditions.

	Condition / dB SPL (RMS)	Cortical potentials P1	N1	P2	N2
	40	-0,13	-0,69**	0,60**	0,02
DLM (Sone)	50	-0,18	-0,42*	0,51*	-0,09
	60	-0,03	-0,43*	0,35	-0,34
	70	-0,18	-0,36	0,48*	-0,15
	80	-0,15	-0,24	0,40	-0,07
	90	-0,18	-0,27	0,47*	-0,27
Sound level (dB SPL)	40	-0,28	-0,26	0,06	0,15
	50	0,04	-0,19	0,25	-0,22
	60	0,24	-0,27	0,30	-0,39
	70	-0,19	0,05	0,14	-0,11
	80	0,29	0,47*	0,02	-0,16
	90	-0,04	0,37	-0,07	0,01

Tab. 4.2: Correlation of level and loudness model with the peak amplitude of the cortical potentials (P1, N1, P2 and N2) for different six different level conditions. * $p < 0.05$ (significant), ** $p < 0.01$ (highly significant).

4.6.2 Loudness change

The research on N1 and P2 suggests that they reflect a matching process, that is, whenever a stimulus is presented, it is matched with previously experienced stimuli (Sur *et al.*, 2009). Hence, it makes sense to consider the loudness change instead of the loudness at the time of the event. The loudness change is the loudness at a certain time minus the average loudness over a recent time range. The recent time range is defined by a period of remembrance. Therefore, we use the term memory-time to describe this period. By varying the memory time, it can be examined which time range has to be considered for the matching process of the N1 and P2. Figure 4.9 illustrates this analysis by plotting the R values of the Pearson correlation for both potentials over the memory-time. The black vertical lines in this figure show the moment of the previous event. If the memory-time exceeds this value, the loudness change value contains the loudness of the previous event. The correlation between the loudness change and the N1 increases until the period of memory-time reaches the moment of the previous event. Thereafter, the correlation decreases again. This indicates that the loudness change should contain the recent loudness up to the previous event. This conclusion is even more marked if the correlation function between loudness change and the P2 is considered. A large increase in correlation is shown for memory-times containing the previous event.

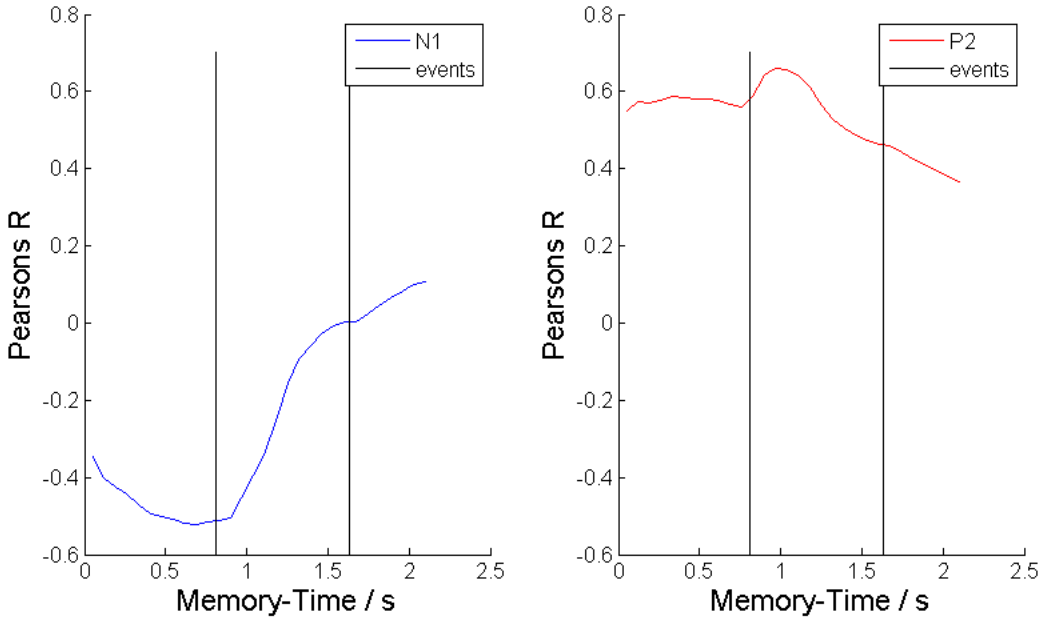

Fig. 4.9: Correlation between the peak amplitude of cortical potentials N1 and P2 and the loudness change with variation of memory-time. The memory-time is the time range whose mean loudness serves as a reference for the calculation of the loudness change. The events (black) are the time frames when strong loudness changes in the stimulus elicit ERPs.

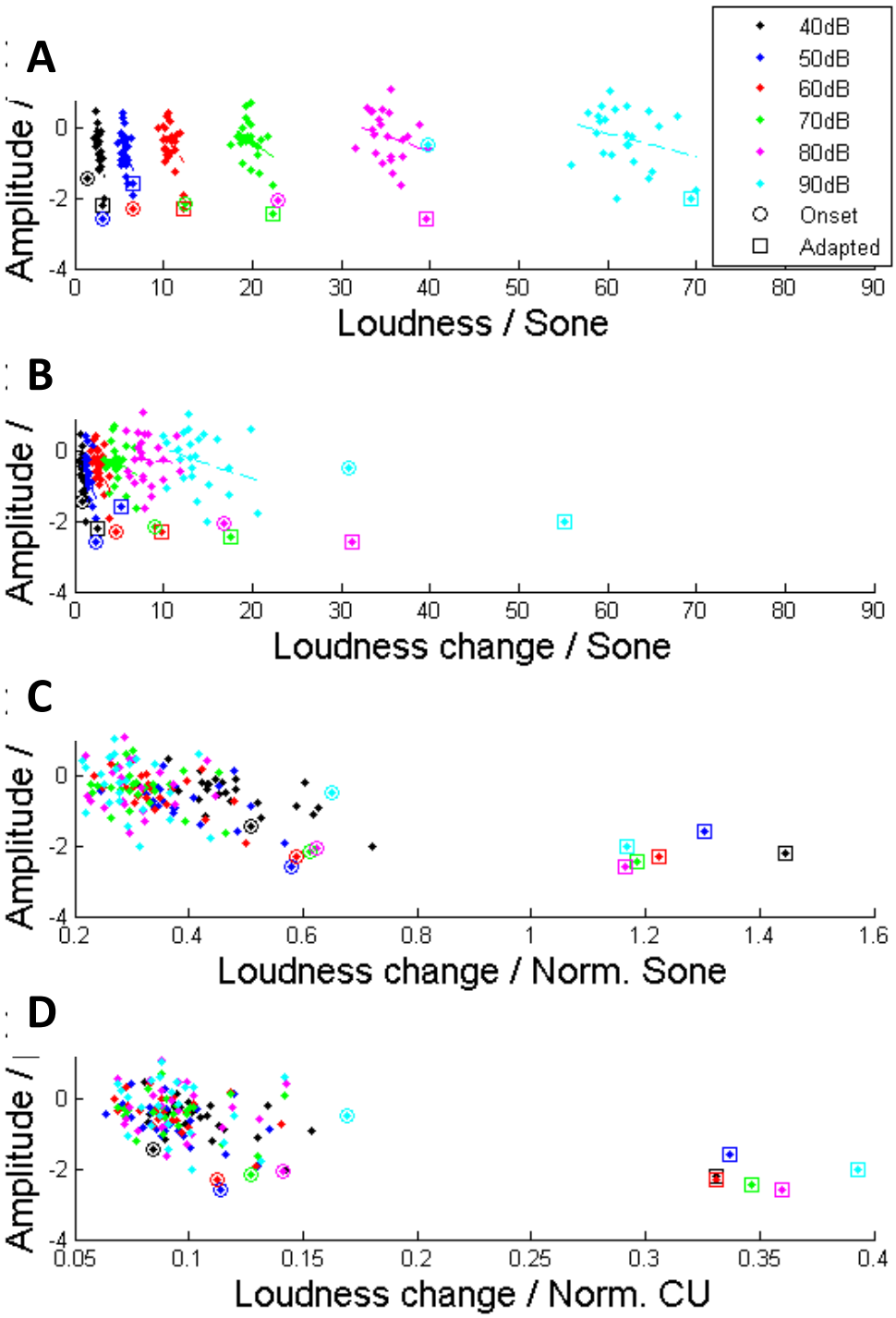

Fig. 4.10: The peak amplitude of N1 of all event-triggered ERPs of the stimulus for different level conditions plotted against different loudness transformations. A. Calculated Sone-loudness of the DLM; Circles are the N1 of the ERPs of the stimulus onset; Squares are the N1 of the ERPs of the first event, which is still superimposed by the N2 of the onset response. B. Loudness change instead of loudness using a memory-time of 0.95 s. C. Recalibrated (or normalized) loudness change in Sone calculated by considering the ratio of loudness and overall loudness (95 percentile) of the respective condition. D. The Sone-loudness is transformed into CU-loudness by using the transformation function from Heeren *et al.* (2013).

4.6.3 Normalized loudness

The Previous analyses have shown that loudness, and especially loudness change, are best related to maximum deflections of N1 and P2 potentials. However, so far the conditions were examined individually and not interrelated. Using the example of N1, the first (A) and second (B) plots in Figure 4.10 show that loudness or loudness change as variables are not sufficient to explain all conditions together. The amplitudes of the EEG response seem to be at about the same level, regardless of condition. This suggests that a recalibration process is taking place, i.e. a charge to the averaged received loudness. This can simply be modeled by normalizing the loudness change to the overall loudness of each condition:

$$\mathrm{EEG} \propto \frac{L(t) - \langle L \rangle_{\tau,t}}{\langle L \rangle} \tag{4.10}$$

where τ represents the memory-time, L the loudness and *EEG* the amplitude of the N1 and P2 peaks. In Fig. 4.10C it can be shown that in this way all N1 peaks can be explained by one model. Furthermore, even the onset response (circles) can be integrated into the model. An integration of the adapted ERP (squares, the ERP influenced by the onset response) can possibly succeed if the influence of the onset response is regressed.

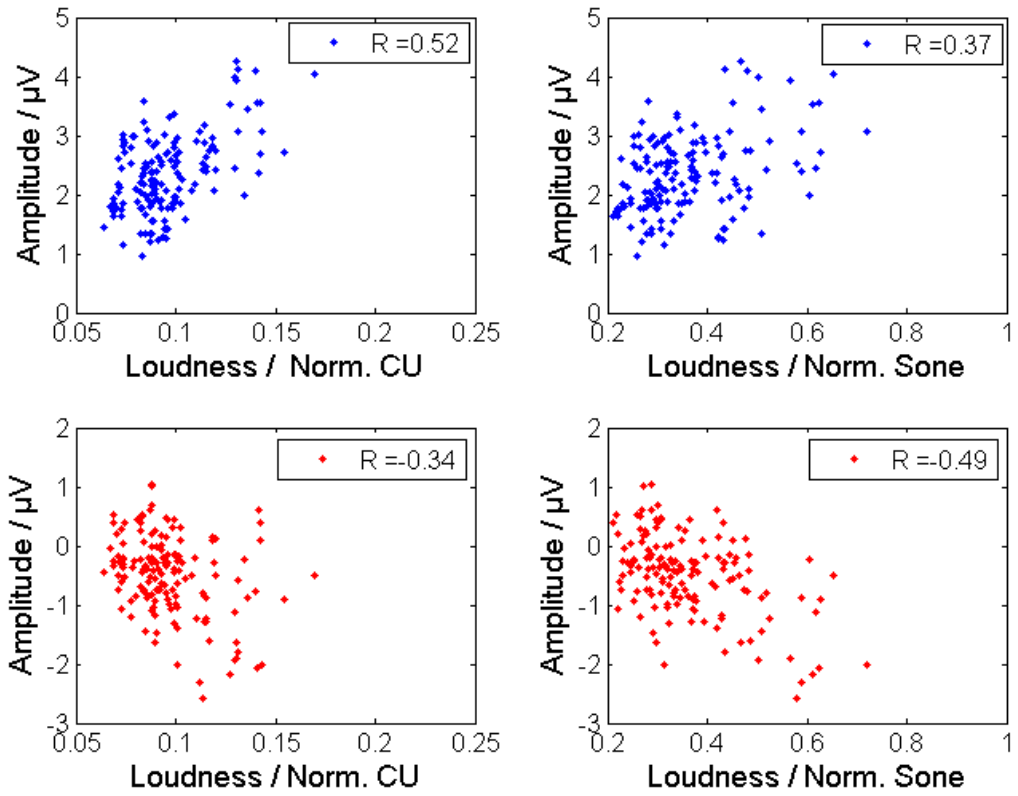

Fig. 4.11: Correlation of the peak amplitude of N1 (red, bottom) and P2 (blue, top) with the modeled Sone (right) and CU loudness (left). Normalized loudness means recalibrated loudness change as in Fig. 4 C-D.

Finally, the question arises whether the peaks of N1 and P2 can be better explained by Sone or CU-loudness growth function. Therefore, the normalized loudness change in Sone and CU is correlated with the amplitude peaks of the two potentials. The CU-loudness was derived from the Sone-loudness using the transformation function from Heeren *et al.* (2013).

$$\mathrm{CU} = 2{,}6253 \cdot \lg(\mathrm{sone} + 0{,}0887)^3 + 0{,}7799 \cdot \lg(\mathrm{sone} + 0{,}0887)^2$$

$$+8{,}0856 \cdot \lg(\mathrm{sone} + 0{,}0887) + 13{,}4493 \qquad (4.11)$$

Figure 4.11 illustrates that the N1 is more related to the Sone-loudness (R = -0,49), while the P2 is more related to the CU-loudness (R = 0,52).

4.6.4 Summary

The results have shown that the peak of the excursion of N1 and P2 is not related to the level, but to the loudness. As already mentioned in Chapters 4 and 5, this claim is still controversial, as studies with pure tone pulses miss the expected compression that occurs with loudness (Näätänen and Picton, 1987; Hegerl *et al.* 1994). The problem of these studies is the use of narrowband sounds, which induce neural adaptation (Chapter 5), and the disregard of the matching function of N1 and P2 (Sur *et al*., 2009), which has also been shown in this experiment. Here, it was found that the normalized loudness change can explain the N1 and P2 together across all conditions. Surprisingly, music proves to be an extremely suitable stimulus for investigating the relationship between loudness and cortical potentials because it is a broadband sound and is evidently quite robust to neural adaptation. Bearing in mind the matching function of N1 and P2, a fundamental problem arises for almost all EEG paradigms: the analysis of the auditory evoked potentials is only possible by repeating the presentation of the conditions. A paradigm with repetitions has the undesirable effect of affecting the beginning of the epoch of an EEG response from the end of the previous epoch. This is especially a problem for potentials like the N1 and P2 that compare current events with previous ones. Paradigms involving repetitions also recalibrate the cortical potentials to a averaged received loudness. This is ignored in almost every EEG loudness study. However, there were differences in the strength of the correlation in the choice of loudness growth function.

In this experiment, it turned out that the N1 rather represents the Sone-loudness whereas the P2 rather represents the CU loudness. This implies that categorical loudness is a processing step at later stages in which sensory processed loudness is categorized into categories like ‘soft’, ‘loud’ or ‘too loud’. It is also indicates that the transition from N1 to P2 also represents the transition from sensory to perceptual loudness.

5 Neural representation of loudness: Cortical evoked potentials in a loudness recalibration experiment

Abstract

Loudness recalibration comprises differences in loudness judgments of a target stimulus according to the presence of a preceding recalibrating tone. Increasing the inter-stimulus intervals (ISI) between recalibrating tone and target typically enhances the effect of the recalibrating tone which then reduces the loudness of the target stimulus. On the one hand the recalibration effect is interesting for loudness research in neuroscience since identical stimuli show different loudness, thus providing evidence whether neural EEG responses indicate a coding of physical intensity or if they provide a loudness correlate. On the other hand, finding neural stages that provide loudness correlates may help to segregate whether loudness recalibration is a change in the decisional process - represented on rather late processing stages - or if the underlying reason is a specific sensory adaptation of the neural stimulus representation. To find out if any cortical response in the latency range of 75 to 510 ms behave like a loudness correlate, we investigated the EEG response during a psychoacoustical loudness recalibration experiment with different ISI. With increasing ISI the strength of the N1-P2 deflection of the respective electroencephalography response decreases in a similar way as the loudness perception of the target tone pulse. This indicates a representation based on loudness rather than on intensity at the corresponding processing stage. Assuming that the N1-P2 deflection does not represent a decision-processing stage this indicates furthermore that context effects cause an adaptive change of the neural stimulus representation rather than changing only the decisional processes.

5.1 Introduction

Loudness is an auditory measure of perception and can essentially be defined as the perceived intensity of a sound. However, it is well known for decades that besides intensity other physical parameters such as spectral and temporal properties contribute to loudness perception (Stevens, 1957; Zwicker, 1958; Zwicker and Fastl, 1999; Moore, 2013). It is assumed that crucial steps of loudness processing are located in the peripheral parts of the auditory system (i.e. outer and inner ear, brainstem and thalamus). While many loudness effects are linked to the physical properties of the stimulus there are experiments indicating loudness differences for identical stimuli caused by changeable contexts (for a review see, e.g., Arieh and Marks, 2011). In psychophysics the common idea for a long time was that these contextual effects reflect relatively late processes of judgment, i.e., a bias in response, rather than changes in the internal perceptual representation of sound intensity (Anderson 1975; Stevens 1958; Treisman, 1984). However, more recent psychoacoustical research on contextual effects of loudness show that besides decisional processes there is evidence for explicit changes in the underlying loudness representation (e.g., Schneider and Parker, 1990;

Algom and Marks, 1990). Neurophysiological indications, confirming those changes of loudness representation would provide important information for the understanding of loudness processing. Neurophysiological correlates of contextual loudness effects that are related to processing stages which are not explicitly involved in decision processing or are even known to be not consciously accessible would provide clear evidence for a context related adaptation of neural loudness representation.

From the neurophysiological side many studies have shown that the change of sound intensity is represented by respective changes of neural activity, in the brainstem (Bauer *et al.*, 1974; Serpanos *et al.*, 1997; Fobel and Dau, 2004; Junius and Dau, 2005; Dau *et al.*, 2005) as well as in the auditory cortex (Näätänen and Picton, 1987; Hegerl *et al.*, 1994; Hoppe *et al.*, 2001; Mulert *et al.*, 2002; Mariam *et al.*, 2012; Potter *et al.*, 2017). Both can be measured indirectly by electroencephalography (EEG). Generally, those studies treated the analysis of auditory evoked potentials (AEP) that can be simply generated by clicks or tone pulses. While most authors agree on the representation of sound intensity in AEPs it remains rather unclear whether the observations are indicating only a correlation to intensity or a link to the perceived loudness. Studies investigating this topic provide results with contradictory evidence (Pratt and Sohmer, 1977, Babkoff *et al.*, 1984, Näätänen and Picton, 1987; Darling and Price, 1990, Serpanos *et al.*, 1997, Hoppe *et al.*, 2001; Silva and Epstein, 2010, 2012). Since sound intensity is the factor with the greatest influence on loudness, intensity and loudness show a close covariation making it rather difficult to distinguish whether a neural response is better correlated with one or the other. Previous studies typically tried to detect compression effects in the neural response to distinguish loudness from sound intensity (Menard *et al.*, 2008; Castro *et al.*, 2008; Emara and Kolkaila, 2010; Eeckhoutte *et al.*, 2016; Behler and Uppenkamp, 2016). It can be assumed that a major part of the compression of the auditory dynamic range takes place in the cochlear. Consequently, these studies are not clear on whether this peripherally compressed intensity is coded in the respective AEPs (which may be expanded at other stages). On the contrary, correlations between neurophysiological responses and contextual loudness effects might provide evidence for the representation of loudness rather than sound intensity on the corresponding processing stages.

Studies of the context effects of loudness typically show that the presentation of a relatively strong inducing tone reduces the loudness of a succeeding weaker tone of the same frequency (Marks, 1988; Mapes-Riordan and Yost, 1999; Marks, 1994; Scharf, 2002; Nieder *et al.*, 2003; Wagner and Scharf, 2006). The amount of reduction depends on the duration between successive tones which was shown by Arieh and Marks (2003). They used a paradigm that investigated loudness perception of a 2500 Hz tone pulse at 60 dB SPL – referred to as target tone - while varying the loudness context. The respective loudness context is realized by a prior presented tone pulse at 80 dB SPL of the same frequency – referred to as recalibrating tone. The loudness in a given context is then measured by adjusting a third tone pulse (comparison tone) at 500 Hz - presented 1 s after the target tone – to be equally loud (cf. Fig. 1).

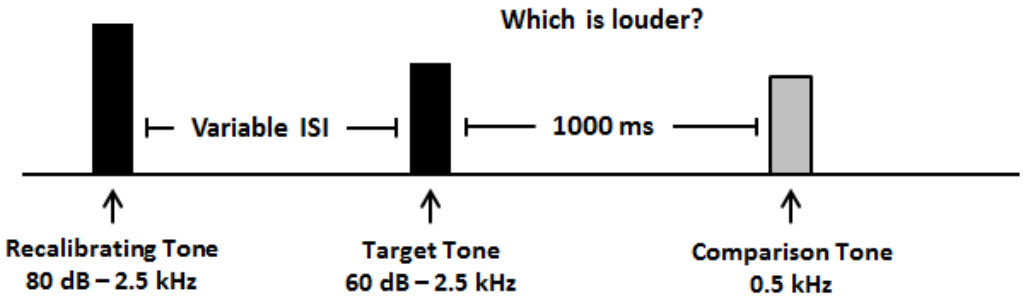

Fig. 5.1: The stimulus sequences used to measure the adaptation of loudness recalibration. The recalibrating tone creates the context of the target tone. By varying the inter-stimulus interval (ISI) between the recalibrating tone and the target tone the adaptation process of the loudness recalibration is measured. The Comparison tone level was adjusted in an adaptive procedure to determine the loudness of the target tone.

They showed that high level contexts result in a reduced loudness of the target tone. They termed this loudness adjustment to the previous context as recalibration effect[1]. Furthermore, they investigated the time course of the loudness recalibration by varying the length of the inter-stimulus intervals (ISI), i.e. the intervals between the two tone pulses. The result of this study was that loudness recalibration decreases with increasing length of the interval until it converges after about 2 s to a fixed loudness value. They suggested that the prior tone pulse determines the context of the loudness while the human auditory system needs some time to complete this recalibrating process.

Previous EEG studies investigated the change of the cortical AEP for a series of tone pulses using the same frequency with varying stimulus onset asynchrony and ISI (Davies *et al.*, 1966; Nelson and Lassman, 1968; Lanting *et al.*, 2013). These studies found a strong decrease of the neural response strength to the second and later presented tone pulses with respect to the first tone. This decrease of strength is referred to as "repetition suppression". It can be assumed to be a consequence of an overloading-related reduction of synchronous firing neurons or specific neural circuits. Furthermore, they found that by increasing the length of the intervals between tone pulses, the repetition suppression decreases, i.e. the related cortical AEPs/responses increase again. Assuming that the strength of the cortical components correlates positively with loudness as has been proposed in the EEG literature (Näätänen and Picton, 1987; Hegerl *et al.*, 1994; Hoppe *et al.*, 2001; Mulert *et al.*, 2002; Mariam *et al.*, 2012; Potter *et al.*, 2017) these findings appear to contradict the psychoacoustical results of Arieh and Marks (2003a). However, Lanting *et al.*, (2013) showed that the different cortical components have different adaptation properties. They suggested that mainly cortical components at later stages, particularly the vertex-positive deflection around 200 ms (P2), are involved in the decrease of the repetition suppression. Therefore, it might be possible to observe loudness recalibration to less affected components, for example the N1. A magnetoencephalographic study performed

[1] More recent literature often refers to this effect as induced loudness reduction (ILR). However, we keep the nomenclature used in the study by Arieh and Marks which is partially reproduced here

by Lu *et al.* (1992) supports this idea by suggesting a relationship between loudness recalibration and the vertex-negative deflection around 100 ms, the N1m. However, the relationship found by Lu *et al.* (1992) is based on the observation of a similar time constant for the increase of N1m and the observed changes in loudness, independent of observing psychoacoustically an increase or decrease in loudness. Another study by Oberfeld (2010) measured the cortical response to a target tone and its loudness for two different level conditions (30 and 60 dB) with a recalibrating tone at 90 dB and an ISI of 120 ms. Their results showed that, in addition to repetition suppression, there was a similar change in N1 amplitude and, furthermore, P2 amplitude with loudness.

The goal of the current study is to find features in the cortical AEP responses that possibly correlate directly with the loudness differences caused by context effects. Therefore, we included the psychoacoustic paradigm of Arieh and Marks (2003a) directly into an EEG experiment. The synchronous measurement of both provides two possible advantages: (a) The recorded EEG data is directly linked to the psychoacoustic outcome – no effects have to be considered due to different attention or physiologic status of the subject, which may occur when performing EEG and psychoacoustic measurements at different times. (b) A possible enhancement of the neural activation, since recent studies recommended active listening tasks to enhance neural activation related to the investigated features (Öhman and Lader, 1972; Bennington and Polich, 1999; O'Sullivan *et al.*, 2015).

The features we investigate are the condition-related changes of strength and latency of the cortical components in the AEP. The cortical components that we consider are the vertex-negative deflection around 100 ms (N1), the vertex-positive deflection around 200 ms (P2) and the vertex-negative deflection around 250 ms (N2). These selected components are popular candidates to represent cortical activation at different stages. We tested if (1) the strength of cortical components increase with increasing loudness; and (2) the latency of cortical components is changing with loudness.

The eliciting target stimulus is identical in the different ISI-conditions and due to the sufficiently long ISIs, effects caused by peripheral interactions of recalibrating tone and target tone (e.g. forward masking) appear unlikely. Therefore, a relationship between cortical AEPs and context-related loudness changes would provide evidence for the neural representation of loudness rather than a representation of intensity input on this processing stage. The N1-P2 deflection is assumed to be associated with sensory evoked potentials that are most probably not representing conscious processes such as attention or decision making (Polich, 2007) while later AEPs, like N2 may already reflect as well "cognitive control" mechanisms (Folstein and Petten, 2008). Therefore, a correlation between earlier cortical AEPs and the contextual loudness would provide some evidence that the neural representation of the stimulus loudness is adapted rather than observing a bias in response only, whereas a correlation only with N2 or later AEPs would indicate the opposite.

5.2 Methods

5.2.1 Subjects

Twelve subjects, six male (S1, S3, S6, S7, S10, S11) and six female (S2, S4, S5, S8, S9, S12), with clinically normal hearing participated in the experiments. All had hearing thresholds ≤ 15 dB HL at standard audiometric frequencies between 125 and 8000 Hz. The subjects were right-handed, between 20-30 years old and were paid volunteers.

All experimental procedures were approved by the ethics committee of the University of Oldenburg.

5.2.2 Stimulation and recording

The stimuli used in the experiment are different sequences of sinusoidal tone pulses with overall duration of 50 ms including 5 ms cosine rise and decay. A sequence generally consisted of one 2500 Hz recalibrating tone at 80 dB SPL, one 2500 Hz target tone at 60 dB SPL and one 500 Hz comparison tone with adjustable sound level. The four tested conditions were presented in pseudo-randomized order and differ in the length of the ISI between recalibrating and target tone: (i) without a recalibrating tone, and with a recalibration tone at (ii) 150, (iii) 525, and (iv) 1650 ms. For all conditions the ISI between target and comparison tone was 1000 ms. Signal generation and conditioning including attenuation was performed digitally on a PC by a MATLAB R2006b (the Mathworks) based custom made software. The stimuli were digital-to-analog converted at a sampling frequency of 44,1 kHz using a Fireface UCX (RME) as external sound device and were presented diotically via ER 2 insert earphones (Etymotic Research) driven by a HB7 headphone buffer (Tucker Davies Technologies).

The subjects response to the psychoacoustic task via a button-response box, which sends a response specific trigger signal to the EEG recording system to be stored in the EEG data and gives back the response to the stimulation PC to control the psychoacoustic measurement procedure.

Similar to the study of Arieh & Marks (2003), a randomized and adaptive two-alternative forced choice two-up, two-down (2-AFC-2up-2down) procedure was used for every track, based on the framework of the AFC software package, a tool designed to run psychoacoustic measurements in Matlab (Ewert, 2013). This adaptive rule converges to the target probability of 50 % on the psychometric function. The loudness of the 2500-Hz target tone at 60 dB SPL was estimated from the results for the measurement of two interleaved tracks, referred to as ascending and descending track respectively. The step size of the sound level of the comparison tone changed stepwise during the adaptive process. After three reversals of direction, i.e. the transition points of the staircase procedure, the step size of 4 dB decreased to 2 dB. The whole procedure ended after nine reversals in each track. The ascending track started with a comparison tone sound level at 40 dB SPL, the descending track at 80 dB SPL.

Due to the adaptive 2-AFC procedure, the numbers of trials in each run varied. The average number of runs was 156, but varied between 77 and 210.

Parallel to the loudness estimation procedure EEG was recorded with a Biosemi Active Two system using 64 channels with the electrodes placed according to the international 10-20 system. Contact gel (Signa gel Electrode Gel, Parker) was used to ensure good contact between electrodes and scalp. The electrode offset was not higher than 10mV. The recordings were collected and digitalized on a second PC using the ActiView software (6.03, Biosemi) at 1024 Hz sampling frequency without using any filters. Each session lasted 10–15 min. A 10 min break separated consecutive sessions.

5.2.3 Data processing and analysis

The psychoacoustic data from the adaptive loudness matches provide a series of loudness judgments that gradually converge to the value of equal loudness. The point of equal loudness between the target and reference was calculated by averaging the last six reversal points. Subsequently, the mean value was determined from the ascending and descending track. The listeners were instructed to ignore the recalibrating tone and judge which tone was louder, the target or the comparison.

All EEG data was processed offline using MATLAB. The average of all electrodes was used as virtual reference electrode. For data evaluation a cluster of nine electrodes around the Fcz-electrode (Fcz, F1, F2, Fz, C1, C2, C3, C4, Cz) was considered. Generally, diotical stimulation favors the recording at central electrodes. Furthermore, the findings of Lanting at al. (2013) regarding the topographic activation of the responses N1 and P2 to the target tone showed that some of the frontal electrodes were most appropriate to measure the AEP to the consecutive tone pulse. After applying a high-pass filtering at 0.5 Hz to reduce the electrode drift and a low-pass filtering at 8 Hz to reduce alpha-wave artefacts, the data were down-sampled to 64 Hz sampling frequency. Further artefact reduction caused by eye movement was done by using independent component analysis. (MATLAB toolbox eeglab 4.4b). For the two short ISI-conditions (150 ms, 525 ms) there was a temporal overlap between the auditory evoked potentials of the recalibrating and the target tone. Therefore, the response to the recalibrating tone from the longest ISI-condition (1650 ms) was subtracted from these conditions. Afterwards the data were separated to epochs. The epochs were averaged using an iterated weighted averaging procedure (Riedel and Kollmeier, 2003) with two iterations. This procedure provides weights for each epoch according to the estimated amount of noise contamination in it.

In the focus of our investigation were the amplitudes and latencies of the cortical components N1, P2 and N2. We considered using the peak-to-peak amplitude between the P2 and N1 deflections and between the N2 and P2 deflections as well as the mean amplitude over different time windows as representations of these components. Peak-to-peak amplitudes of the components were extracted by selecting the minimum amplitude for the negative components (N1, N2) or the maximum for the positive component (P2) within physiologically reasonable time windows, respectively. The respective time windows were chosen for N1

from 75 to 170 ms, for P2 from 170 to 310 ms and for N2 from 200 to 510 ms. The width of these windows was chosen based on the measured temporal occurrence of the components in the mean EEG response to the target tone. The position of each peak was used as representation of the latency of the respective cortical component.

5.2.4 Statistical analysis

We aim to test whether the trend of the evoked responses of the target tone is driven by a decrease of the repetition suppression, or driven by the adaptation of the loudness recalibration. Therefore, we compared the order of the averaged AFC judgements for the different ISI-conditions with the order of the strength of the cortical components. An analysis of variance for repeated measures (rANOVA) including the Mauchly-test for sphericity and a corresponding Greenhouse-Geisser correction was done to find out whether the means of the prospective neural correlates over subjects differ about conditions. This was also done for the sound levels derived from the AFC judgements. Furthermore, we compared the changes in the loudness matches to the changes in the EEG amplitudes within subjects. We used the approach of Bland and Altman (1995) for the calculation of correlation coefficients with repeated observations. This method allows to investigate the direct relationship between individual loudness matches and individual EEG responses.

5.3 Results

Figure 5.2 illustrates the results of the AFC judgements for the three ISI-conditions. By increasing the ISI between the recalibrating tone and the target tone the level of the adjusted comparison tone decreases. This implies a decrease in loudness of the target tone. The level difference between the ISI-conditions 150 ms and 525 ms is about 4 dB, whereas it amounts to only 2.5 dB between 525 ms and 1650 ms. The decreasing effect of the recalibrating tone on the loudness perception of the target tone shows high significance (rANOVA: F = 8.789, p < 0.004, epsilon = 0.862, $\mathrm{df}_{\mathrm{between}} = 1.724$, $\mathrm{df}_{\mathrm{error}} = 18.973$).

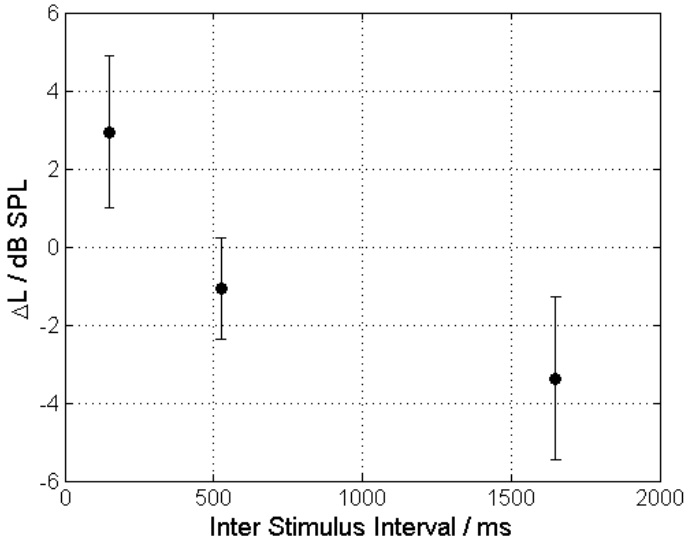

Fig. 5.2: The level of the comparison tones was adjusted in equal loudness to the 60 dB target tone for the baseline condition and several inter-stimulus interval conditions. ΔL represents the level difference between the adjusted comparison tone level under baseline condition and the comparison tone levels under ISI condition. The error bars indicate the standard errors across subjects. If ΔL > 0 the target tone of this ISI condition is perceived louder than in the baseline condition and vice versa.

The recalibrating tone has also a considerable effect on the EEG-response of the target tone. The strength of cortical components is clearly reduced for all conditions compared with the baseline-response (Fig. 5.3). But across the ISI-conditions the mean amplitudes of the different cortical components show disparate behavior (Fig. 5.4). The mean amplitude difference of N1-P2 decreases with increasing ISI significantly (rANOVA: F =4.9, p = 0.023, epsilon = 0.862, $\mathrm{df}_{\mathrm{between}} = 1.724$, $\mathrm{df}_{\mathrm{error}} = 18.963$). On the other hand, the mean amplitude difference of P2-N2 increases proportionally with the ISI (rANOVA: F = 4.68, p = 0.02 epsilon = 1, $\mathrm{df}_{\mathrm{between}} = 2$, $\mathrm{df}_{\mathrm{error}} = 22$). However, these significant effects could only be shown for the mean amplitude not for the peak-to-peak amplitude. The recalibrating tone seems to have no significant effect on the latencies of the cortical components (Fig. 4). The results are summarized in more detail in Table 1 including the absolute amplitudes of the cortical components for peak-to-peak and mean amplitude methods as well as their corresponding latencies.

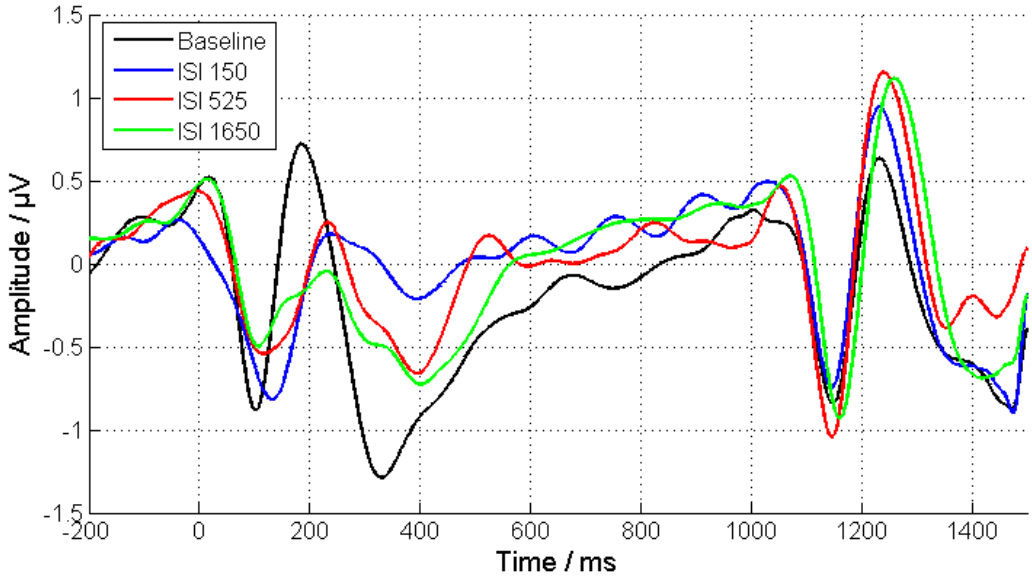

Fig. 5.3: The averaged EEG-response to the target tone (starting at 0ms) and to the loudness-adjusted comparison tone (starting at 1050 ms) for the different ISI-conditions. The averaged EEG response to a 60 dB tone pulse and to a corresponding adjusted comparison tone with the absence of a recalibrating tone is represented by the black line (Baseline condition). Both tone pulses have a similar cortical response, characterized by a pronounced N1 with a peak at 100 ms, a pronounced P2 with a peak 190 ms, and an N2 with a peak at 320 ms with an extended reload. Comparing the conditions, for all inter-stimulus intervals a reduced strength of the target tone response is shown. Furthermore, the strength of their N1-P2 deflections shows a decrease with increasing inter-stimulus interval.

Correlation analysis between individual loudness matches and individual EEG responses showed a significant correlation between N1-P2 amplitude and loudness matches ($R_{\mathrm{peak-to-peak}} = 0.31$, p = 0.03; $R_{\mathrm{mean}} = 0.38$, p = 0.01). There was no significant correlation between the P2-N2 amplitude individual loudness. The latencies of the cortical components also do not correlate with the individual loudness matches.

Figure 5.5 provides the topographic activation (topoplot) of the EEG responses to the target tone (Fig. 5a), the comparison tone (Fig 5b) and to the target tone of the baseline condition. This was done by applying a principal component analysis (PCA) across the 64 electrodes and the time range of 350 ms after each tone pulse. This method is a common approach to find out which electrodes have the highest portion of the evoked potentials (Tremblay *et al.*, 2014). For the purpose of visualization, the weightings of the first PCA-component are displayed as topoplots. The topoplots represent the activation of the electrodes that are involved in the first component. When comparing the activation resulting from the target tone with the one from the comparison tone or from the target tone of the baseline condition, respectively, a shift of the relevance of electrodes is visible from the frontal to the central electrodes.

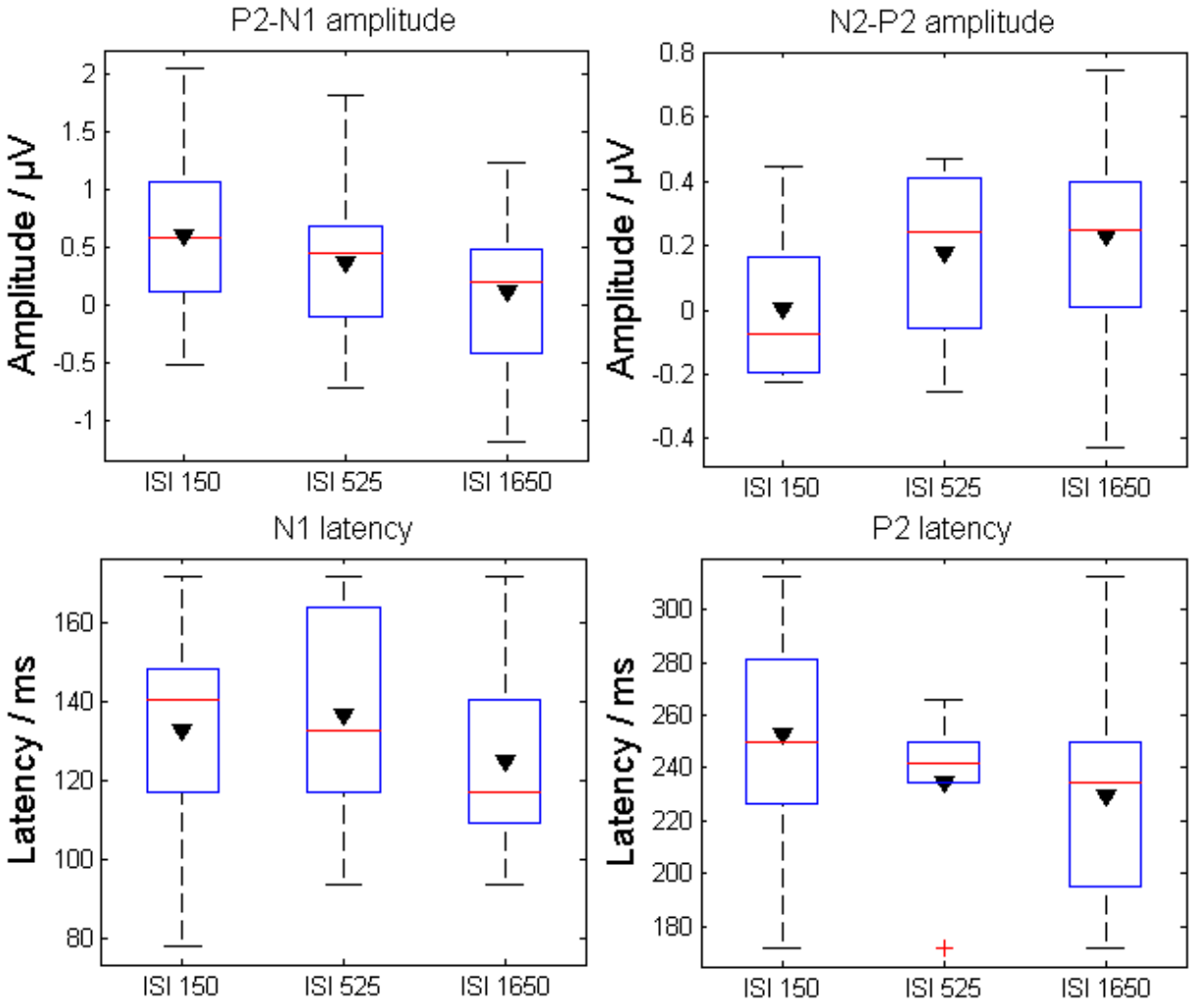

Fig. 5.4: Boxplots of the investigated features of the EEG response to the target tone for the three different ISI-conditions. At the top amplitude differences between the N1-P2 (top left) and P2-N2 (top right) are depicted. The amplitudes were extracted by using the mean amplitude method. At the bottom the latencies of the N1 (left) and of the P2 (right) are displayed. The mean values across the subjects are marked by black triangles.

Figure 5.5 provides the topographic activation (topoplot) of the EEG responses to the target tone (Fig. 5.5a), the comparison tone (Fig 4.5b) and to the target tone of the baseline condition. This was done by applying a principal component analysis (PCA) across the 64 electrodes and the time range of 350 ms after each tone pulse. This method is a common approach to find out which electrodes have the highest portion of the evoked potentials (Tremblay *et al.*, 2014). For the purpose of visualization, the weightings of the first PCA-component are displayed as topoplots. The topoplots represent the activation of the electrodes that are involved in the first component. When comparing the activation resulting from the target tone with the one from the comparison tone or from the target tone of the baseline condition, respectively, a shift of the relevance of electrodes is visible from the frontal to the central electrodes.

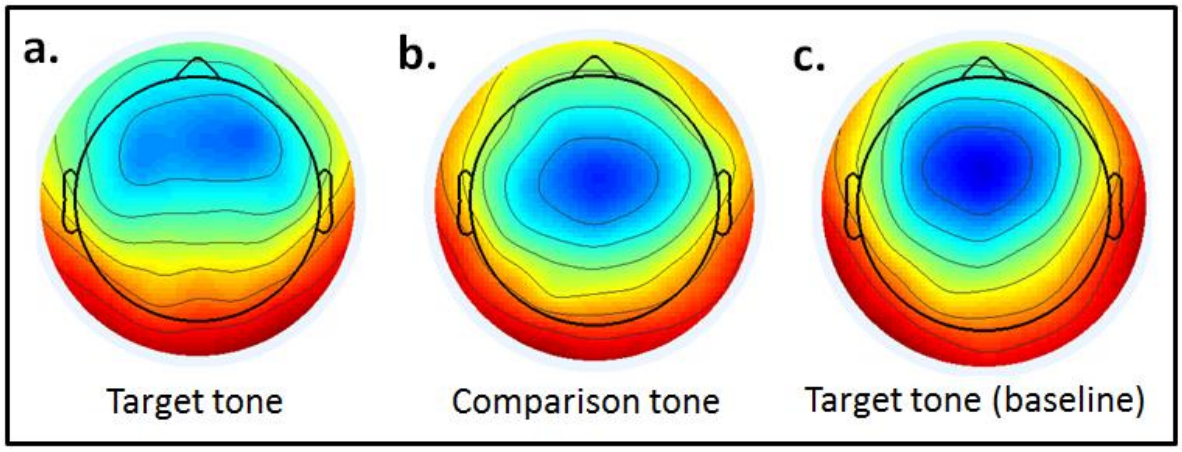

Fig. 5.5a-c: Topographic map of the PCA weightings / eigenvector of the first component to represent the neural activity of each tone response over the scalp; a. target tone response; b. comparison tone response c. target tone response at the baseline condition. Comparing the weightings of the responses to target tone (a) and the comparison tone (b) in the case of the occurrence of a recalibrating tone a shift of the most relevant electrodes towards the front of the scalp can be seen. In comparison, the response to target tone of the baseline condition (c) shows a similar activation pattern as the response to the comparison tone (b). The change from a frontocentral to a more central topographic representation may indicate a change in the underlying neural sources involved in the processing of the first and the second of two successive tones at the same frequency.

5.4 Discussion

The recalibration effect for loudness found by Arieh and Marks (2003a) was successfully reproduced for the three tested conditions in the psychoacoustical experiment. However, level differences between the conditions are slightly lower than found in Arieh and Marks (2003a). This result can be related to the different experimental context, given, e.g., by the additional task to keep eye and muscle movement as low as possible in the EEG recording setup while performing the loudness judgment. In addition, the task performed here did not include visual feedback which differs from the task provided by Arieh and Marks (2003a). This difference in the experimental design is necessary, since EEG responses evoked by visual stimuli could interfere with the recordings. It is noticeable that the loudness of the target tone at an ISI of 150 ms to the recalibration tone is perceived higher than without the recalibration tone (baseline condition). This special case is in line with the observations of other studies. Within ISI shorter than about 200 ms, recalibrating and target tone can interact in a way that leads to loudness enhancement of the target tone. This is particularly the case when the level of the recalibrating tone is higher than the target tone level (Elmasian *et al.*, 1980; Oberfeld, 2007).

The amplitude of the P2-N2 deflection - as part of the cortical response to the target tone – increase with increasing duration of the ISI. This decrease of repetition suppression is in line with the expectations of previous studies (Davis *et al.*, 1966; Nelson and Lassman, 1968, Lanting *et al.*, 2013). However, comparing

the loudness matches with the amplitudes of the P2-N2 deflection within subjects no correlation was found. On the contrary, the amplitude of the N1-P2 deflection decreases with increasing duration of the ISI which is in line with the psychoacoustically measured decrease in loudness. This effect was most evident when examining the mean amplitude, since the peak-to-peak amplitude only showed significant correlation when compared with the individual loudness matches. In addition, the correlation coefficient for the peak-to-peak amplitude was slightly lower than for the mean amplitude. This is probably due to the fact that peak extraction cannot take temporally wider deflections into account. The reduction of the N1-P2 deflection with decreasing loudness is in agreement with the findings by Hoppe *et al.* (2001) in CI users. For the different ISI the target stimuli are identical in the current study. That means that the changes in loudness as well as in the coinciding N1-P2 deflection, as observed here, are not related to the physical stimulus intensity. Since the ISI used here should have a sufficient long duration to rule out cochlear interactions between recalibrating tone and target tone this holds as well for the magnitude of the neural input to the auditory system.

cortical components		ISI-conditions			ANOVA results	
		150 ms	525 ms	1650 ms	F-value	p-value
Amplitude – Peak-to-peak	P2-N1 / µV	1.35 ± 0.25	1.10 ± 0.21	0.94 ± 0.20	3.12	.06
	N2-P2 / µV	0.94 ± 0.10	1.21 ± 0.18	1.25 ± 0.14	2.67	.09
Amplitude – Mean amplitude	P2-N1 / µV	0.59 ± 0.21	0.37 ± 0.20	1.11 ± 0.20	6.1	.006
	N2-P2 / µV	0.001 ± 0.061	0.18 ± 0.074	0.229 ± 0.095	7.65	.002
Latency	N1 / ms	132.8 ± 30	136.7 ± 29	125 ± 27	0.5	.61
	P2 / ms	252.6 ± 39	234.4 ± 31	229.2 ± 41	1.29	.29

Table 5.1: Cortical features of the EEG-response to the target tone for different ISI to the recalibrating tone and their corresponding ANOVA results. The N1-P2 and N2-P2 deflections were extracted by using two different approaches: peak-to-peak and mean amplitude. Latencies were extracted at the position of the peaks of N1 and P2 components. One-way ANOVA was performed across ISI-conditions and 12 subjects.

Compared to the expected strength of a normal cortical response (1-2 µV for the deflections of individual components, e.g. the response to the comparison tone) a strong reduction in the strength of the cortical response was observed for both N1-P2 as well as P2-N2 deflections. This generally coincides with the results of previous studies where a decrease of repetition suppression for the cortical components was observed (Davis *et al.*, 1966; Nelson and Lassman, 1968, Lanting *et al.*, 2013). The N1-P2 deflection shows this overall

reduction as well[2] but providing no decrease of repetition suppression with increasing ISI. Therefore, it might be possible that the change in the underlying AEP component (related to the time course of loudness recalibration) is even larger than observed here but that it is partially masked by the general trend of a decrease of repetition suppression. Nevertheless, the representation of the recalibration effect in the N1-P2 deflection suggests that loudness is represented in the auditory cortex. This finding is in agreement with the results of some previous studies (Röhl and Uppenkamp 2012; Behler and Uppenkamp 2016, Thwaites *et al.*, 2016). Behler and Uppenkamp (2016) showed in a functional magnetic resonance imaging study that the correlation between the blood oxygenation level dependent (BOLD) signal and loudness increases at later stages in the auditory pathway. Maximum correlation was reached at the posterior medial Heschl's gyrus. Similar observations have been done by Thwaites *et al.* (2016). They found in an MEG study several cortical components corresponding to the loudness of speech estimated by a loudness model. Using a cross correlation analysis, they found four components with significant correlation at latencies of 45 ms, 100 ms, 165 ms and 275 ms. The latest of these components provides a higher correlation with short-term loudness (i.e. low-pass filtered) than with the instantaneous loudness. Hence they concluded that the latest component represents short-term loudness, which is seen as a more final loudness representation on higher auditory stages. At latencies between 75 ms and 310 ms we even found that the N1-P2 deflection correlated to contextual effects in loudness, suggesting as well that in this range of latencies the neural representation of loudness is fairly complete.

Loudness recalibration reflects central rather than peripheral processing of sound intensity. However, it is not clear whether this effect still belongs to the sensory processing or is rather an issue of a shift in response criteria (Schneider and Parker, 1990; Algom and Marks, 1990, Arieh and Marks, 2003b). Investigating the time neural of processing can be useful to clarify this issue. For example, in Arieh and Marks (2003b) listeners had to detect weak tones in a choice decision task. They found that in conditions that produce loudness recalibration, the listeners show increased response times and higher error rates compared to control conditions. According to Luce (1986), a positive relation between response time and error rate is a strong indicator for a sensory rather than for a decisional change. That means, from the psychoacoustical point of view, the effect of loudness recalibration is - at least partially - a change in the sensory representation of the target-tone rather than a response bias. From the neurophysiological side the N1-P2 deflection is not

[2] With respect to the effect strength of the repetition suppression for tone pulses of either different or same frequencies the current results are in line with the considerations of Lanting *et al.* (2013). They argued that for consecutive tones with different features, e.g. frequency, the subsequent tone will recruit unadapted neurons and thus show almost no repetition suppression while this shall not apply for similar tones. This is line with psychoacoustic findings indicating the largest loudness reduction when the recalibration tone and the target fall within the same critical band (Marks and Warner, 1991). However, Lanting *et al.* (2013) contradicts other AEP studies on prepulse inhibition (e.g. Schall *et al.*, 1996) who measured frequency-independent repetition suppression. In our results this is reflected by comparing the amplitude of the responses of the target and the comparison tone. It can be clearly seen that the strengths of the cortical responses to the comparison tone are higher than to the target tone. The target tone response is a subsequent tone pulse to the recalibrating tone and both have the same frequency. The comparison tone, however, is the subsequent tone pulse to the target tone with a different frequency than its predecessor. This observation is best illustrated by comparing the cortical responses for the baseline condition where the strength is of the same order of magnitude.

expected to reflect processing stages that are already involved into decisional processes but more in providing specific feature traces of the stimuli and maybe to some extent complete stimulus representations (for a review see Näätänen and Winkler, 1999). Under this premise the found relation between N1-P2 deflection and loudness recalibration time course provide some neurophysiological evidence, that loudness recalibration causes a context-related adaption of the sensory stimulus representation.

It should be noted, that in the current EEG-analysis a low-pass filter is used with a relatively low adjusted cutoff frequency of 8 Hz which reduces the amplitudes of the measured cortical components. Highly affected by this filtering is the P1 component. Similar to Hoppe *et al.* (2001) who used a cut off frequency at 10 Hz this makes a quantitative analysis of the P1 amplitude almost impossible. Therefore, only the N1-P2 and P2-N2 deflections are investigated here. This has two major drawbacks: (1) it significantly reduces the temporal resolution of the components; (2) it impedes a separate consideration of the deflections. In particular, it would be interesting to investigate the N1 and P2 separately, since they represent different dipole sources (Crowley and Colrain, 2004). The reason why this filtering adjustment is necessary are frequently observed alpha waves during the experiment presumable induced by the effort of the AFC-task processing. This is in agreement with the findings by Klimesch (1999), who showed that cognitive and memory performance was reflected by an increase of alpha activity. This has to be seen as a clear drawback of the synchronous measurement of AEP and psychoacoustic tasks. Indeed, other factors could influence the current AEP results as well. The recalibrating tone pulse as the prior to the target tone has a higher sound level than the sound level of the target tone as a prior to the comparison tone which was adjusted to equal loudness. Therefore, the prior tone with the higher sound level generates repetition suppression with more strength. Furthermore, the cognitive load differs during the presentation of the target and comparison tone. Both disparities could alter the results in some way. We also compared the total activity across the scalp of the responses to the target tone with the comparison tone by determining their activation patterns. These patterns were obtained from the weightings of the principal components. Figure 5 shows the weightings on a topographic map across the scalp of the first component to the target tone and the comparison tone response of the conditions using a recalibrating tone and to the target tone response of the baseline condition. In all cases the central sources were accountable for the evoked potentials. However, if a recalibrating tone has been presented, a shift of the most relevant electrodes towards the front of the scalp can be seen. This change from a frontocentral to a more central topographic representation on the scalp suggest a change of the underlying neural sources involved in the processing of the first and the second of two successive tones at the same frequency. More details on these aspects as well as on an identification of possibly counteractive AEP components reflecting a release of neural repetition suppression and loudness recalibration as discussed above are needed to provide a deeper understanding of the underlying mechanisms. This might be possible by a detailed source reconstruction obtained from less noisy data in future studies, e.g., by MEG measurements using a similar paradigm.

5.5 Summary and Conclusion

Neural representation of the loudness recalibration effect in the EEG response was demonstrated by using an AFC paradigm during the EEG recordings. The strength of the N1-P2 deflection shows the same trend with the time course of loudness recalibration, i.e., a decrease with increasing ISI; whereas the P2-N2 deflection shows a recovery from repetition suppression indicated by a relative increase with increasing ISI. In the presence of a conditioner tone at 80 dB SPL both deflections for all ISI are clearly reduced - compared to the baseline condition. The following conclusions are drawn:

a) Since the target stimuli for all ISI are identical the processing stages reflected by N1-P2 deflection provide a representation of loudness- rather than only physical sound intensity.
b) The finding that loudness recalibration is already reflected as early as in the N1-P2 deflection provides neurophysiological evidence that context effects cause an adaption of the neural loudness representation of the stimulus and not (only) a shift in the decisional processes.

6 Summary and outlook

One objective of this thesis was to examine how well current models can predict loudness for music, in comparison to more basic level measures, and what kind of modifications would need to improve the predictions. The second objective was to supplement the psychoacoustical judgements and model predictions with potential neural correlates of loudness in order to obtain an insight how sound intensity is transformed into a loudness impression by the different stages of auditory processing.

For this purpose, firstly, in Chapter 3 a paired comparison task was implemented and analyzed using the Bradley-Terry-Luce method (BTL). This allows to transform the rank-scaled comparison data obtained for the overall loudness judgements of various sequences of music pieces to an interval scale. Subsequently, the resulting BTL-scale was compared with the predictions of models and level measures. This was done by considering the linearity of the function representing the relationship between the derived loudness scale and the models resp. the level measures. The results showed that for all investigated Sone-models this functional relationship is not linear. Transforming the Sone-scales of the models into categorical units, this relationship is largely linear. For the sound pressure level, the shape of the function shows also a high degree of linearity, however, with a slight curvature at low levels. The discrepancy between the experimentally determined BTL scale and the Sone-models can only be modelled if one of two alternative explanations is valid: Either the respective loudness models or the Sone scale are not fully suitable for estimating the overall loudness of music. However, the high degree of linearity between the BTL scale and the model results when transformed into CU indicates that the Sone-transformation would need to be corrected.

Based on this, only parts of the BTL scale were considered for further analysis. This was done to ensure linearity for all measures and models. The leave-one-out cross-validated R^2 served as a measure of the goodness of the prediction. It was shown that the standard frequency-weighted sound level (e. g., dB(A)) outperforms the considered loudness models. This was even the case when optimizing those parameters that are partially adjustable for an overall loudness assessment: the cut-off frequency of the low-pass filter to simulate the short-term loudness transformation, and the percentile as a measure of the distribution of the overall loudness. However, the parameter variation showed that the predictive power of loudness models still increases at a cut-off frequency around 4 Hz and a percentile at 97%. Incorporating a modeled overall sharpness of the music stimuli into the predictions of the loudness models the accuracy of the prediction increased further.

These findings hint towards necessary modifications of loudness models that have not been considered in previous studies on model comparisons (Skovenborg *et al.*, 2004, Vickers, 2010). After all, such an application of a low-pass filter to mimic a short-term loudness stage was proposed in some previous studies (Chalupper and Fastl, 2002; Glasberg and Moore, 2002), although not at such a low filter cut-off frequency as used here, i.e., 4 Hz in contrast to 8 Hz (Chalupper and Fastl, 2002) and 15 Hz (Glasberg and Moore, 2002). Similar findings were reported by Rennies *et al.* (2015). They showed that the estimation of the overall loudness of

some fluctuating technical sounds could be improved by using a cut-off frequency of around 2 Hz. The improvement of the performance by processing the information of overall sharpness is in line with the findings of Fastl (2007) who assumed that sharpness affects the loudness perception of music. However, due to the fact that the investigated frequency-weighted level measures performed so well in the current study with no use of information about the sharpness of the music pieces, one could also assume that rather spectral cues are the reason for improvement of the loudness models since sharpness implies information about the spectral envelope of the respective sound.

In Chapter 4, the investigation of loudness processing of music is continued by supplementing the subjective loudness data with objective measurements. This was attempted by analyzing neural activity recorded by cortical EEG responses. Therefore, a sequence from a classical piece of music lasting 20 s was presented at six different sound levels. The peak amplitude of the deflections P1, N1, P2 and N2 in the event related potentials were correlated to the predicted loudness of these eliciting events. The results demonstrate that the peak of the deflections N1 and P2 show a higher correlation to the loudness of the stimuli - as predicted by the dynamic loudness model of Chalupper and Fastl (2002) - than to the sound pressure level. This finding supports the hypothesis from previous fMRT- and MEG-studies that loudness perception is mainly processed in auditory cortex (Röhl and Uppenkamp, 2012; Behler and Uppenkamp, 2016; Thwaites *et al.*, 2016).

It should be noted that for the loudness prediction of single conditions, even simple frequency weightings like the A-weighted sound pressure level or the EBU-R128 standard provide a correlation similar to that of the loudness model, although being merely a very rough approximation of auditory processing. For correlation across all conditions, it is loudness that is best related to the N1 and P2. This relation was less distinct for the level measures. However, a correlation across all conditions could only be carried out in a meaningful way by considering the ratio between the loudness change and overall loudness as derived from the loudness model instead of considering the absolute loudness. The overall loudness was determined by the average across the course of the instantaneous loudness for the complete stimulus. The loudness change was determined by the difference between the instantaneous loudness and the mean value of a time-window of 1 s prior to the current instant in time. A better representation of the neural activity of N1 and P2 by loudness change than by absolute loudness is in line with previous research on event related potentials that suggests that N1 and P2 reflect a matching process. Whenever a stimulus is presented it is matched with previously experienced stimuli (Sur *et al.*, 2009). Furthermore, it was found that the N1 characteristics show the best correlation to Sone-loudness whereas the P2 response shows a closer link to the CU-loudness. This finding allows the interpretation that the transition from N1 to P2 reflects a logarithmic transformation of purely sensory processing into the perception of loudness.

Most previous studies that investigated whether sound level or loudness is represented by specific neural responses compared the different growth characteristics of sound level and perceived loudness with the growth characteristics of respective neural responses (e. g. Ménard *et al.*, 2008; Silva and Epstein, 2010; Behler and Uppenkamp, 2016). These approaches can provide some evidence at which specific processing

stage the neural representation matches better the loudness rather than the physical level of an acoustic stimulus. It remains still unclear whether a closer correlation between perceived loudness and neural representation is just reflecting modifications by the ongoing sensory (pre-) processing of the neural stimulus representation or actually constitutes a (completed) loudness perception process including a kind of judgment or classification. However, when dealing with stimuli like music for which loudness judgments are strongly affected by individual preferences and experiences, a clear discrimination line is desirable between neural responses representing already a kind of judged loudness in contrast to neural responses reflecting just a stimulus representation being optimized for loudness perception - and therefore showing properties closely linked to loudness for most types of less complex stimuli.

If we assume that physical identical stimuli result in almost the same neural (pre)processing before a kind of loudness-classification processing is done, the recording of neural responses within a paradigm during which identical stimuli are judged as different loud by the same subject would provide useful information about the question if a specific neural response is representing indeed 'perceived' loudness or not. In Chapter 5, therefore, a paradigm was implemented to study correlations between neurophysiological responses and contextual loudness effects, i.e. loudness recalibration. Arieh and Marks (2003a) showed that loudness recalibration is subject to an adaptation process. The adaptation of the recalibration can be demonstrated by the variation of the inter-stimulus interval (ISI) between two tone pulses. A subsequent tone pulse of different frequency can be used for loudness matching. This procedure was performed during the EEG recording. While the strength of a large part of the cortical response to the target tone was attenuated according to the neural repetition suppression effect, it was shown that the strength of the N1-P2 component was sensitive to the recalibration effect. Loudness as well as the strength of the N1-P2 component decreased with increasing magnitude of the ISI. In conclusion, loudness changes due to recalibration are represented on the stage of the auditory cortices at the respective latencies in the range of 75 to 310 ms. This finding coincides with the results in Chapter 4, where the amplitude of N1 and P2 were identified as loudness correlates. The results from Chapter 5 provide stronger evidence that the interpretation of N1 and P2 as indicators for perceived loudness is correct. The N1 and P2 components not only reflect level effects and spectral effects, but also central loudness effects. Hence, level measures cannot adequately represent these cortical potentials with generators beyond primary auditory cortex.

Although the effect of loudness recalibration and its adaptation process has hardly been studied enough for series of tone pulses or tone pulse complexes to allow detailed predictions for music (or at least periodic repetitions of tone complexes as stylized music) it may be assumed that some of the deviations between predictions from loudness models and neural representation found in Chapter 4 (as reflected in low correlation) can be explained by the loudness recalibration process. Therefore, the loudness recalibration effect might be a good and relevant starting point to include aspects of neural processing into loudness models for complex stimuli in order to improve the accuracy and applicability of loudness models for complex stimuli such as music. Several psychoacoustical findings on loudness recalibration already provide insight into basic properties of loudness generation (Marks and Warner, 1991; Mapes-Riordan and Yost, 1999; Nieder *et*

al., 2003; Arieh and Marks, 2003b; Arieh *et al.*, 2005). Based on these findings a new effective processing stage at the end of sensory processing for loudness modeling appears conceivable since loudness judgement seems to change according to the history of the stimulus perceived so far. One feature of such a processing stage would probably be that it essentially operates on specific loudness because recalibration affects mainly stimuli with the same comparable spectral content as stimuli presented before. For further improvements of such a stage, additional psychoacoustic investigations would be needed, especially to examine how level differences and the interval between masker and masked sound affect loudness specifically in different frequency bands.

Outlook

In summary, this thesis has shown that current loudness models are deficient in the treatment of music. This was reflected by deviations in the prediction of the overall loudness of various pieces of music (Chapter 3) as well as in the prediction of instantaneous loudness on the level of neural representation (Chapter 4). The observation that frequency-weighted level measures (A-weighted, B-weighted, ITU-R BS.1770-2) not only kept up in loudness prediction, but even partially outperformed the psychoacoustical loudness models (DIN 45631/A1, TVL, DLM, DLMext) indicates that a more detailed modeling of processes in the auditory periphery provides no benefit for loudness predictions of complex stimuli like music. These models have been adjusted for pure tones and for broadband noise. They are apparently not capable for loudness predictions of stimuli strongly affected by central processing due to preferences and experiences of the listener. In this thesis various suggestions were made to modify or extend these loudness models. The findings in Chapter 3 suggest that the CU scale provide a better matching scale for the loudness of complex stimuli than the Sone scale. This is in line with the results in Chapter 4. The P2 component of the event related response shows a closer link to the CU loudness than to the Sone loudness. This suggests that the loudness transformation of current models should be based on categorical units which can be done for example by adding a subsequent CU transformation (Appell *et al.*, 2001; Heeren *et al.*, 2013) to Sone-models.

Operating on overall loudness predictions of music sequences, a modified short-term loudness stage simulated by a low-pass filter around 4 Hz was recommended in Chapter 3. However, it is unclear if this can be seen as a relevant part reflecting the integration process needed to derive the overall loudness estimate from an internal instantaneous loudness process, or if such a low-pass filter is already an integral part of continuous loudness processing, or if it is a combination of both. An experiment is needed in which this aspect can be tested for the continuous loudness judgment of music. This experiment should contain both: continuous loudness judgements and overall loudness judgements similar to those described by Schlittenlacher *et al.* (2014). The variation of differently adjusted low-pass filters as in this thesis could then shed light on this issue. Furthermore, the overall loudness of music seems to be affected by the perceived sharpness. It could be shown that a linear combination of the estimated overall loudness and estimated sharpness could increase the prediction power for the real overall loudness (see Chapter 3). It is unclear whether the influence of sharpness on loudness is a judgment bias or indeed a necessary contribution to the

sensory processed intensity in the calculation of loudness. The overall sharpness and some measure of 'instantaneous' sharpness should be tested by multiple regressions if they may help to improve the estimated short-term loudness of the models of music sequences.

Finally, it has been found that there are neural correlates to loudness context effects, that is, the adaptation process of recalibration of loudness appears to be coded in the EEG responses. Further experiments should examine the brainstem response to see if the loudness recalibration effect is already related to early stages of processing. Moreover, the effect should be demonstrated for more inter-stimulus intervals to strengthen the validity of the findings presented here. It should be further examined whether the inclusion of the psychoacoustical paradigm into the EEG experiment was crucial for the elicitation of the founded neural correlate to loudness. This can be tested by using a paradigm in which the tone pulse series were passively presented. The last experiment gives some evidence that complex processes like the contextual loudness effects, which are beyond neural signal preprocessing, are represented in the cortical deflections N1 and P2. However, it cannot be excluded that the final neural representation of a loudness percept for stimuli as complex as music (which are affected by very specific individual preferences) is firstly finalized on even higher processing stages. Therefore, a transfer of the tone-pulse paradigms from Marks (1993) or Arieh and Marks (2003a) to paradigms with more complex and periodic sounds and, after all, to music would be desirable.

Taken together, this thesis contributes to reducing the deficiencies of current loudness models in dealing with music. Focus of this work was not the modeling of the periphery of loudness processing but rather central aspects. Psychoacoustical and neural findings emphasize the benefit of categorical loudness scaling. Estimating the overall loudness of music requires correct preprocessing of short-term loudness and, furthermore, sharpness must be integrated into the model. In the EEG, the loudness change is reflected in the N1 (by Sone) and in the P2 (by CU). At this cortical level there is also a neural representation of loudness recalibration. Hence, loudness judgements and its neural representation has been shown in this thesis to be a complex, hierarchical process that assumedly requires neurosensory processing steps (including a logarithmic transform and an appropriate spectral weighting) as well as more central adaptation and recalibration processes that will have to be incorporated into more sophisticated loudness models of the future.

7 Appendix: Bradley-Terry-Luce method

The result of a pairwise comparison experiment is a count matrix *M* of the number of times that each option was preferred over every other option,

$$(1)\quad M_{i,j} = \begin{cases} \#\text{ of times option } i \text{ preferred over option } j, & i \neq j \\ 0, & i = j \end{cases}$$

Since pair comparisons correspond to a binomial process, the probability $p_{i,j}$ that a subject prefers the option *i* over *j* can be estimated from $M_{i,j}$ and the number of comparisons. The Bradley-Terry-Luce method (BTL) that establishes data on an interval scale level by postulating a relationship between preference probabilities and scale values (Ellermeier, 2004):

$$p_{ij} = \frac{\pi_i}{\pi_i + \pi_j} \tag{A.1}$$

where $\pi_i$is a value associated to each option. However, the real scale values corresponding to the stimuli can only be obtained by logarithmisation (Tsukida and Gupta, 2011)

$$\mu_i = s \cdot log(\pi_i) \tag{A.2}$$

where *s* is a scale parameter. In order to estimate π_i the likelihood function of the model has to be set up and maximized. For binomially distributed random variables, the likelihood function takes the shape:

$$L = \prod_{i<j} p_{ij}^{M_{i,j}} \cdot \left(1 - p_{ij}\right)^{M_{j,i}} \tag{A.3}$$

For this, we recommend the numerical method proposed by Wickelmaier and Schmid (2004). The confidence intervals *e* for the maximum likelihood estimated scale values can be estimated by determining the covariance matrix (A5) of the inverse negative Hessian matrix H (A.4). The Hessian matrix of the log-likelihood function is defined as the square matrix of second partial derivatives with respect to the model parameters:

$$\mathrm{H} = \begin{pmatrix} \frac{\partial^2 \log L}{\partial {\pi_1}^2} & \cdots & \frac{\partial^2 \log L}{\partial \pi_1 \partial \pi_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \log L}{\partial \pi_k \partial \pi_1} & \cdots & \frac{\partial^2 \log L}{\partial {\pi_k}^2} \end{pmatrix} \tag{A.4}$$

$$\mathrm{C} = \begin{bmatrix} -H & 1 \\ 1 & 0 \end{bmatrix}^{-1} \qquad e = 1{,}96 \cdot \sqrt{\mathrm{Diag}\left(\mathrm{cov}(C)\right)} \tag{A.5}$$

Furthermore, the goodness of the fit can be derived by comparing the likelihood of the model with the saturated model that fits the data perfectly (Wickelmaier and Schmid, 2004). This can be done by applying the likelihood ratio test (Eq. A6).

$$\chi^2 = -2 \log\left(\frac{L}{L_{sat}}\right) \tag{A.6}$$

This value can be compared with a critical value of the χ^2-distribution that corresponds to a defined level of significance with the degree of freedom $f = \binom{n}{2} - n + 1$.

References

Aiken, S. J., Picton, T. W., 2008. Human Cortical Responses to the Speech Envelope. Ear & Hearing Vol. 29, No. 2, 139-157.

Algom D., and Marks L. E., 1990. Range and regression, loudness scales, and loudness processing: Toward a context-bound psychophysics. Journal of Experimental Psychology: Human Perception and Performance 16, 706–727.

Anderson N. H., 1975. On the role of context effects in psychophysical judgment. Psychological Review 82, 462–482.

ANSI S3.4, 2007. Procedure for the computation of loudness of steady sounds. New York, NY: Author.

Appell, J.-E., Hohmann, V., Kollmeier, B., 2001. Review of Loudness models for normal and hearing-impaired listeners based on the Model proposed by Zwicker. Z. Audiol. Vol.40, No.4, 140-154.

Arieh, Y., Marks, L. E., 2003a. Time course of loudness recalibration: Implications for loudness enhancement. J. Acoust. Soc. Am. Vol. 114, No. 3, 1550-1556.

Arieh, Y., Marks, L. E., 2003b. Recalibrating the Auditory System: A Speed-Accuracy Analysis of Intensity Perception. Journal of Experimental Psychology: Human Perception and Performance Vol. 29, No. 3, 523-536.

Arieh, Y., Kelly, K., Marks, L. E., 2005. Tracking the time to recovery after induced loudness reduction. J. Acoust. Soc. Am. Vol. 117, No. 6, 3381-3384.

Babkoff, H., Pratt, H., Kempinski, D., 1984. Auditory brainstem evoked potential latency-intensity functions: A corrective algorithm. Hearing Research Vol. 16, 243-249.

Barrett, D. E., Hodges, D. A., 1995. Music Loudness Preferences of Middle School and College Students. Texas Education Research.

Bauer, J. W., Elmasian, R. O., Galambos, R., 1974. Loudness enhancement in man. I. Brainstem-evoked response correlates. J. Acoust. Soc. Am. Vol. 57, No. 1, 165-171.

Behler, O., Uppenkamp, S., 2016. The representation of level and loudness in the central auditory system for unilateral stimulation. NeuroImage Vol. 139, 176-188.

Bennington, J. Y., Polich, J., 1999. Comparison of P300 from passive and active tasks for auditory and visual stimuli. International Journal of Psychophysiology Vol. 34, 171-177.

Besson, M., Macar, F., 1987. An event-related-potential analysis of incongruity in music and other non-linguistic contexts. Psychophysiology Vol. 24, 14–25.

Bland, J. M., Altman, D. G., 1995. Calculating correlation coefficients with repeated observations: Part 1—Correlation within subjects. British Medical Journal 310, 446.

Brand, T., Hohmann, V., 2002. An adaptive procedure for categorical loudness scaling. J. Acoust. Soc. Am. Vol. 112, 1597-1604.

Cahn, B. R., Polich, J., 2006. Meditation States and Traits: EEG, ERP, and Neuroimaging Studies. Psychological Bulletin Vol 132, No. 2, 180-211.

Cai, S., Ma, W. D., Young, E. D., 2008. Encoding Intensity in Ventral Cochlear Nucleus Following Acoustic Trauma: Implications for Loudness Recruitment. JARO 10, 5-22.

Castro, F. Z., de Prat, J. J. B., Zabala, E. L., 2008. Loudness and auditory steady-state responses in normal-hearing subjects. International Journal of Audiology Vol. 47, 269-275.

Chalupper, J., Fastl, H., 2002. Dynamic Loudness Model (DLM) for Normal and Hearing-Impaired Listeners. Acta Acustica United with Acustica Vol. 88, 378-386.

Crowley, K., Colrain, I., 2004. A review of the evidence for P2 being an independent component process: Age, sleep, and modality. Clinical Neurophysiology, 115, 732–744.

Cullari, S., Semanchick, O., 1989. Music Preference and Perception of Loudness. Perceptual and Motor Skills Vol. 68, No. 1, 186-186.

Darling, R. M., Price, L. L., 1990. Loudness and Auditory Brain Stem Evoked Response. Ear & Hearing Vol. 11, No. 4, 289-295.

Dau T., Wegner O., Mellert V., Kollmeier B., 2000. Auditory brainstem responses with optimized chirp signals compensating basilar membrane dispersion. J. Acoust. Soc. Am. Vol. 3, 1530–1540.

Davis, H., Mast, T., Yoshie, N., Zerlin, S., 1966. The Slow Response of the Human Cortex to Auditory Stimuli: Recovery process. EEG Clin. Neurophysiol. Vol. 21, 105-113.

Debreu, G., 1960. Review of R. D. Luce's Individual choice behavior: A theoretical analysis. American Economic Review, 50, 186-188.

DIN 45631 / ISO 532B: 1991-03. Calculation of loudness level and loudness from the sound spectrum - Zwicker method. Deutsches Institut für Normung.

DIN 45631 / A1, 2010. Calculation of loudness level and loudness from the sound spectrum - Zwicker method - Amendment 1: Calculation of the loudness of time-variant sound. Deutsches Institut für Normung.

DIN 45692, 2009. Measurement technique for the simulation of the auditory sensation of sharpness. Deutsches Institut für Normung.

DIN ISO 226, 2006. Akustik – Normalkurven gleicher Lautstärkepegel, Dt. Norm, Beuth, Berlin.

Ding, N., Simon, J. Z., 2014. Cortical entrainment to continuous speech: functional roles and interpretations. Frontiers in Neuroscience Vol. 8, Article 311.

Dittrich, R., Katzenbeisser, W., Reisinger, H., 2000. The analysis of rank ordered preference data based on Bradley-Terry Type Models. OR Spektrum Vol. 22, 117-134.

Doelling, K. B., Poeppel, D., 2015. Cortical entrainment to music and its modulation by expertise. PNAS Vol. 112, No. 45, E6233-E6242.

EBU R-128, 2014. Loudness Normalisation and Permitted Maximum Level of Audio Signals. European Broadcasting Union.

Eeckhoutte, M. Van, Wouters, J., Francart, T., 2016. Auditory steady-state responses as neural correlates of loudness growth. Hearing Research Vol. 342, 58-68.

Ellermeier, W., Mader, M., Daniel, P., 2004. Scaling the Unpleasantness of Sounds According to the BTL Model: Ratio-Scale Representation and Psychoacoustical Analysis. Acta Acustica united with Acustica Vol. 90, 101-107.

Elmasian, R., Galambos, R., Bernheim, A., 1980. Loudness enhancement and decrement in four paradigms, J. Acoust. Soc. Am. Vol. 67, 601–607.

Elo, A. E., 1965. Age changes in master chess performances. Journal of Gerontology, 20, 289-299.

Emara, A. A. Y., Kolkaila, E. A., 2010. Prediction of Loudness Growth in Subjects with Sensorineural Hearing Loss Using Auditory Steady State Response. Int. Adv. Otol. Vol. 6, No. 3, 371-379.

Epp, B., Verhey, J. L., Mauermann, M., 2010. Modeling cochlear dynamics: Interrelation between cochlea mechanics and psychoacoustics. J. Acoust. Soc. Am. Vol. 128, No. 4, 1870–1883.

Epstein, M., Florentine, M., 2005. Inferring basilar-membrane motion from tone-burst optoacoustic emissions and psychoacoustic measurements. J. Acoust. Soc. Am. Vol. 117, 263–274.

Ewert S. D., 2013. AFC – A modular framework for running psychoacoustic experiments and computational perception models. Proc. AIA-DAGA 2013, Merano, Italy, 1326–1329.

Fastl, H., Fruhmann, M., Ache, S., 2003. Railway Bonus for Sounds without Meaning? Proc. WESPAC8, Melbourne, Australia.

Fastl, H., 2007. Psychoacoustics, sound quality and music. Proc. Inter-Noise 2007, Istanbul, Turkey.

Fiebig, A., Sottek, R., 2015. Contribution of Peak Events to Overall Loudness. Acta Acustica united with Acustica Vol. 101, 1116-1129.

Fletcher, H., Munson, W.A., 1933. Loudness, its definition, measurement and calculation. J. Acoust. Soc. Am. Vol. 5, 82-108.

Florentine, M., Popper, A., Fay, R.R., 2011. Loudness. Springer Handbook of Auditory Research Vol. 37, No. 14.

Fobel O., Dau T., 2004. Searching for the optimal stimulus eliciting auditory brainstem responses in humans. J. Acoust. Soc. Am. Vol. 116, No. 4, 2213-2222.

Folstein J. R., Petten C. Van, 2008. Influence of cognitive control and mismatch on the N2 component of the ERP: A review. Psychophysiology 45, 152-170.

Fucci, D., 1999. Children Scaling Rock Music. Acoust. Soc. Am. 138th Meeting Lay Language Papers.

Gabriel, B., 1996. Equal-loudness Level Contours: Procedures, Factors and Models. Ph.D Thesis. Oldenburg University, Oldenburg, Germany.

Glasberg, B. R., Moore, B. C. J., 2002. A Model of Loudness Applicable to Time-Varying Sounds. J. Audio Eng. Soc. Vol. 50, No. 5, 331-342.

Glaser, E. M., Suter, C. M., Dasheiff, R., Goldberg, A., 1976. The human frequency following response: its behavior during continuous stimulation. Electroencephalogr. Clin. Neurophysiol. 40, 25-32.

Gunderson E., Moline J., Catalano P., 1997. Risks of developing noise-induced hearing loss in employees of urban music clubs. Am. J. Ind. Med. Vol. 31, 75-79.

Hart, H.C., Palmer, A.R., Hall, D. A., 2002. Heschl's gyrus is more sensitive to tone level than non-primary auditory cortex. Hearing Research Vol. 171, 177–190.

Heeren, W., Rennies, J., Verhey, J. L., 2011. Spectral loudness summation of nonsimultaneous tone pulses. J. Acoust. Soc. Am. Vol. 130, No. 6, 3905-3915.

Heeren, W., Hohmann, V., Appell, J. E., Verhey, J. L., 2013. Relation between loudness in categorical units and in phons and sones. J. Acoust. Soc. Am. Vol. 133, No. 4, 314-319.

Hegerl, U., Gallinat, J., Mrowinski, D., 1994. Intensity dependence of auditory evoked dipole source activity. International Journal of Psychophysiology Vol. 17, 1-13.

Hellbrück J., 1993. Hören. Physiologie, Psychologie und Pathologie. Hogrefe-Verlag, Göttingen.

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

Hirsh, I. J., Ward, W. D., 1952. Recovery of the auditory threshold after strong acoustic stimulation. J. Acoust. Soc. Am. Vol. 24, No. 131, 131-141.

Höger, R., E. Matthies, E. Letzing, 1988. Physikalische versus psychologische Reizintegration: Der Mittelungspegel aus wahrnehmungspsychologischer Sicht. Zeitschrift für Lärmbekämpfung Vol. 35, 163-167.

Hoppe U., Rosanowski F., Iro H., Eysholdt U., 2001. Loudness perception and late auditory evoked potentials in adult cochlear implant users. Scandinavian Audiology 30, 119–125.

Howard, M. F., Poeppel, D., 2010. Discrimination of Speech Stimuli Based on Neuronal Response Phase Patterns Depends on Acoustics But Not Comprehension. J. Neurophysiol. Vol. 104, 2500–2511.

Hughes, S. W., Crunelli, V., 2005. Thalamic Mechanisms of EEG Alpha Rhythms and Their Pathological Implications. Neuroscientist Vol. 11, 357–372.

ISO 226, 2003. Acoustics – Normal Equal-Loudness Contours. Geneva, Switzerland.

ISO 532B, 1975. Acoustics-method for calculating loudness level.

ITU-R BS.1770-2, 2011. Algorithms to measure audio programme loudness and true-peak audio level. International Telecommunication Union.

John, M. S., Dimitrijevic, A., Roon, P. v., Picton, T., 2001. Multiple Auditory Steady-State Reponses to AM and FM Stimuli. Audiol. Neurootol. Vol. 6, 12-27.

Junius D., Dau T., 2005. Influence of cochlear traveling wave and neural adaptation on auditory brainstem responses. Hearing Research 205, 53-67.

Kantor-Martynuska, J., 2009. The Listener's Temperament and Perceived Tempo and Loudness of Music. European Journal of Personality Vol. 23, 655-673.

Kellaris, J. J., Powell Mantel, S., Altsech, M. B., 1996. Decibels, Disposition, and Duration: The Impact of Musical Loudness and Internal States on Time Perceptions. Advances in Consumer Research Vol. 23, 498-503.

Kießling, J., Kollmeier, B., Baumann, U., 2018. Versorgung mit Hörgeräten und Hörimplantaten. 3. Auflage, Georg Thieme Verlag KG, Stuttgart, New York.

Klimesch, W., 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Research Reviews Vol. 29, 169-195.

Krishnan, A., Xu, Y., Gandour, J. T., Cariani, P. A., 2004. Human frequency-following response: representation of pitch contours in Chinese tones. Hearing Research Vol. 189, 1-12.

Kuwano S., Namba S., Florentine M., Zheng D. R., Hashimoto T., 1992. Factor analysis of the timbre of noise – comparison of the data obtained in three different laboratories. Proc. Acoust. Soc. Jpn. N92–4–3, 559–560.

Langers, D. R. M., Dijk, P. v., Schoenmaker, E. S., Backes, W. H., 2007. fMRI activation in relation to sound intensity and loudness. NeuroImage 35, 709-718.

Lanting, C. P., Briley, P. M., Sumner, C. J., Krumbholz, K., 2013. Mechanisms of adaptation in human auditory cortex. J. Neurophysiol. Vol. 110, 973-983.

Launer, S., 1995. Loudness Perception in Listeners with Sensorineural Hearing Impairment. Ph.D Thesis. Oldenburg University, Oldenburg, Germany.

Liberman, M. C., Dodds, L. W., 1987. Acute ultrastructural changes in acoustic trauma: Serial-section reconstruction of stereocilia and cuticular plates. Hearing Research Vol. 26, 45-64.

Liberto, G. M. Di, O'Sullivan, J. A., Lalor, E. C., 2015. Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Current Biology Vol. 25, 1-9.

Lu, Z. L., Williamson, S. J. & Kaufman, L., 1992. Behavioral lifetime of human auditory sensory memory predicted by physiological measures. Science 258, 1668–1670.

Luce, R. D., Suppes, P., 1965. Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), Handbook of mathematical psychology (Vol. 3, 249-410). New York: Wiley.

Luce, R.D., 1986. Response Times: Their Role in Inferring Elementary Mental Organization. New York: Oxford University Press.

Lütkenhöner, B., Steinsträter, O., 1998. High-Precision Neuromagnetic Study of the Functional Organization of the Human Auditory Cortex. Audiol. Neurootol. Vol. 3, 191-213.

Madell, J. R., Goldstein, R., 1972. Relation between loudness and the amplitude of the early components of the averaged electroencephalic response. Journal of Speech and Hearing Research Vol. 15, 134-141.

Makeig, S., Bell, A. J., Jung, T. P., Sejnowski, T. J., 1996. Independent component analysis of electroencephalographic data. In Touretzky, D., Mozer, M., Hasselmo, M., editors. Adv. Neural Inf. Process. Syst.; 8: 145-151.

Mapes-Riordan, D., Yost, W. A., 1999. Loudness recalibration as a function of level. J. Acoust. Soc. Am. Vol. 106, No. 6, 3506-3511.

Mariam, M., Mustaffa, I., Muzzammil, K., 2012. Feasibility of N1-P2 Habituation of Differentiate Loudness Levels. Proc. ICoBE, Penang.

Marks, L. E., 1988. Magnitude estimation and sensory matching. Perception & psychophysics Vol. 43, No. 6, 511-525.

Marks, L. E., Warner, E., 1991. Slippery Context Effect and Critical Bands. Journal of Experimental Psychology: Human Perception and Performance Vol. 17, No. 4, 986-996.

Marks, L. E., 1993. Contextual Processing of Multidimensional and Unidimensional Auditory Stimuli. Journal of Experimental Psychology: Human Perception and Performance Vol. 19, No. 2, 227-249.

Marks, L. E., 1994. Recalibrating the auditory system: The perception of loudness. Journal of Experimental Psychology: Human Perception and Performance, 20, 382-396.

McKinnon, C. A., 2009. Louder than hell: Power, volume and the brain. In R. Hill and K. Spracklen (eds), Heavy Fundamentalisms: Music, Metal, and Politics, Oxford: Inter-Disciplinary Press, 113-126.

Meddis, R., 1988. Simulation of auditory-neural transduction: Further studies. J. Acoust. Soc. Am. Vol. 83, No.3, 1056-1063.

Ménard, M., Gallégo, S., Berger-Vachon, C., Collet, L., Thai-Van, H., 2008. Relationship between loudness growth function and auditory steady-state response in normal-hearing subjects. Hearing Research Vol. 235, 105-113.

Miranda, R. A., Ullman, M. T., 2007. Double dissociation between rules and memory in music: an event-related potential study. NeuroImage Vol. 38, 331-345.

Moore, B. C. J., Glasberg, B. R., 1996. A Revision of Zwicker's Loudness Model. Acta Acustica Vol. 82, 335-345.

Moore, B. C. J., 2013. An Introduction to the Psychology of Hearing. London, England: Academic Press, 6rd edn.

Mulert, C., Juckel, G., Augustin, H., Hegerl, U., 2002. Comparison between the analysis of the loudness dependency of the auditory N1/P2 component with LORETA and dipole source analysis in the prediction of treatment response to the selective serotonin reuptake inhibitor citalopram in major depression. Clinical Neurophysiology Vol. 113, 1566-1572.

Näätänen R., Picton T., 1987. The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. Psychophysiology 24, No. 4, 375-425.

Näätänen R., Winkler I., 1999. The concept of auditory stimulus representation in neuroscience. Psychological Bulletin 125, 826-59.

Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: A review. Clinical Neurophysiology Vol. 118, 2544-2590.

Neely S. T., Gorga, M. P., Dorn, P. A., 2003. Cochlear compression estimates from measurements of distortion-product otoacoustic emissions. J. Acoust. Soc. Am. Vol. 114, 1499-1507.

Nelson, D. A., Lassman, F. M., 1968. Effects of Intersignal Interval on the Human Auditory Evoked Response. J. Acoust. Soc. Am. Vol. 44, No. 6, 1529-1968.

Neuhoff, J., G., McBeath, M. K., 1996. The Doppler Illusion: The Influence of Dynamic Intensity Change on Perceived Pitch. Journal of Experimental Psychology: Human Perception and Performance Vol. 22, No. 4, 970-985.

Neuhoff, J. G., McBeath, M. K., Wanzie, W.C., 1999. Dynamic Frequency Change Influences Loudness Perception: A Central, Analytic Process. Journal of Experimental Psychology Vol. 25, No. 4, 1050-1059.

Nieder, B., Buus, S., Florentine, M., Scharf, B., 2002. Interactions between test- and inducer-tone durations in induced loudness reduction. J. Acoust. Soc. Am. Vol. 114, No. 5, 2846-2855.

Oberfeld, D., Plank, T., Temporal Weighting of Loudness: Effects of a Fade In. In Deutsche Gesellschaft für Akustik (Ed.), Fortschritte der Akustik—DAGA 2005, 227–228. Berlin: DEGA.

Oberfeld, D., 2007. Loudness changes induced by a proximal sound: Loudness enhancement, loudness recalibration, or both? J. Acoust. Soc. Am. Vol. 121, 2137-2148.

Oberfeld, D., 2010. Electrophysiological correlates of intensity resolution under forward masking. In: Lopez-Poveda, E.A., Palmer, A.R., Meddis, R. (Eds.), Advances in Auditory Research: Physiology, Psychophysics and Models. Springer, New York, 99-110.

Öhman A., Lader M., 1972. Selective attention and 'habituation' of the auditory averaged evoked response in humans. Physiology & Behavior 8, 79-85.

O'Sullivan, J. A., Shamma, S. A., Lalor, E. C., 2015. Evidence for Neural Computations of Temporal Coherence in an Auditory Scene and Their Enhancement during Active Listening. The Journal of Neuroscience Vol. 35, No. 18, 7256-7263.

Patel, A.D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P.J., 1998. Processing syntactic relations in language and music: An event-related potential study. J. Cogn. Neurosci. Vol. 10, 717–733.

Pedersen, B., Ellermeier, W., 2008. Temporal weights in the level discrimination of time-varying sounds. . J. Acoust. Soc. Am. Vol. 123, No. 2, 963-972.

Picton, T. W., Hillyard, S. A., Krausz, H. I., Galambos, R., 1974. Human auditory evoked potentials. I. Evaluation of components. Electroencephalography and Clinical Neurophysiology Vol. 36, 179-190.

Picton, T. W., Skinner, C. R., Champagne, S. C., Kellett, A. C. J., Maiste, A. C., 1987. Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. J. Acoust. Soc. Am. Vol. 82, No. 1, 165-178.

Picton, T. W., John, M. S., Dimitrijevic, A., Purcell, D., 2003. Human auditory steady-state responses. International Journal of Audiology Vol. 42, No. 4, 177-219.

Pieper, I., Mauermann, M., Kollmeier, B., Ewert, S. D., 2016. Physiological motivated transmission-lines as front end for loudness models. J. Acoust. Soc. Am. Vol. 139, No.5, 2896-2910.

Polich, J., 2007. Updating P300: An integrative theory of P3a and P3b. Clinical Neurophysiology Vol. 118, 2128-2148.

Ponsot, E., Susini, P., Saint Pierre, G., Meunier, S., 2013. Temporal loudness weights for sounds with increasing and decreasing intensity profiles. J. Acoust. Soc. Am. Vol. 134, No. 4, EL321-EL326.

Ponsot, E., Susini, P., Meunier, S., 2016. Loudness Processing of Time-Varying Sounds: Recent advances in psychophysics and challenges for future research. INTER-NOISE and NOISE-CON Congress and Conference Proceedings Vol. 253, No. 2, 6437–6442.

Potter, T., Li, S., Nguyen, T., Nguyen, T., Ince, N., Zhang, Y., 2017. Characterization of Volume-Based Changes in Cortical Auditory Evoked Potentials and Prepulse Inhibition. Scientific Reports Vol. 7, 1-9.

Pratt, H., Sohmer, H., 1977. Correlations between psychophysical magnitude estimates and simultaneously obtained auditory nerve, brain stem and cortical responses to click stimuli in man. EEG Clin. Neurophysiol. Vol. 43, 802-812.

Purcell, D. W., John, S. M., Schneider, B. A., Picton, T. W., 2004. Human temporal auditory acuity as assessed by envelope following responses. J. Acoust. Soc. Am. Vol. 116, No. 6, 3581-3593.

Radeloff, A., Cebulla, M., Shehata-Dieler, W., 2014. Akustisch evozierte Potentiale: Grundlagen und klinische Anwendung. Laryngo. Rhino. Otol. Vol. 93, 625–637.

Rajamanickam, M., 2002. Modern Psychophysical and Scaling Methods and Experimentation. Concept Publishing Company.

Rance, G., Rickards, F., 2002. Prediction of Hearing Threshold in Infants Using Auditory Steady State Evoked Potentials. J. Am. Acad. Audiol. Vol. 13, 236-245.

Rennies, J., Verhey, J. L., Chalupper, J., Fastl, H., 2009. Modeling Temporal Effects of Spectral Loudness Summation. Acta Acustica united with Acustica Vol. 95, 1112-1122.

Rennies, J., Verhey, J. L., Fastl, H., 2010. Comparison of loudness models for time-varying sounds. Acta Acustica united with Acustica Vol. 96, 383-396.

Rennies, J., Wächtler, M., Hots, J., Verhey, J., 2015. Spectro-Temporal Characteristics Affecting the Loudness of Technical Sounds: Data and Model Predictions. Acta Acustica united with Acustica Vol. 101, 1145-1156.

Riedel, H., Granzow, M., Kollmeier, B., 2001. Single-sweep-based methods to improve the quality of auditory brain stem responses Part II: Averaging methods. Z. Audiol. Vol. 40, No. 2, 62-85.

Röhl, M., Uppenkamp, S., 2012. Neural Coding of Sound Intensity and Loudness in the Human Auditory System. Journal of Association for Research in Otolaryngology Vol. 13, 369-379.

Schall, U., Schön, A., Zerbin, D., Eggers, C., Oades, R.D., 1996. Event-related potentials during an auditory discrimination with prepulse inhibition in patients with schizophrenia, obsessive-compulsive disorder and healthy subjects. Intern. J. Neuroscience Vol. 84, 15-33.

Scharf, B., 1978. Loudness. In E. C. Carterette & M . D. Friedman (Eds.), Handbook of perception Vol. 4, New York: Academic Press.

Scharf, B., Buus, S., Nieder, B., 2002. Loudness enhancement: induced reduction in disguise? J. Acoust. Soc. Am. Vol. 112, No. 3, 807-810.

Scherg, M., 1991. Akustisch evozierte Potentiale – Grundlagen, Entstehungsmechanismen, Quellenmodell. Stuttgart, Kohlhammer.

Schlittenlacher, J., Ellermeier, W., Hashimoto, T., 2012. Loudness model extension improving predictions for broadband sounds. Proc. Inter-Noise, 5495-5505.

Schlittenlacher, J., Hashimoto, T., Kuwano, S., Namba, S., 2014. Overall Loudness of Short Time-varying Sounds. Proc. Inter-Noise 2014, Melbourne, Australia.

Schneider B., Parker S.,1990. Does stimulus context affect loudness or only loudness judgments? Perception & Psychophysics 48, 409–418.

Schnitzler, A., Gross, J., 2005. Normal and pathological oscillatory communication in the brain. Nature reviews Neuroscience Vol. 6, 285-296.

Serpanos, Y. C., O'Malley, H., Gravel, J. S., 1997. The Relationship between Loudness Intensity Functions and the Click-ABR Wave V Latency. Ear & Hearing Vol. 18, No. 5, 409-419.

Silva, I., Epstein, M., 2010. Estimating loudness growth from tone-burst evoked responses. J. Acoust. Soc. Am. Vol. 127, No. 6, 3629-2642.

Silva, I., Epstein, M., 2012. Objective estimation of loudness growth in hearing-impaired. J. Acoust. Soc. Am. Vol. 131, No. 1, 353-362.

Skovenborg, E., Nielsen, S. H., 2004. Evaluation of Different Loudness Models with Music and Speech Material. Proc. Audio Engineering Society 117th Convention, 1-34.

Soeta, Y., Nakagawa, S., 2008. The effects of pitch and pitch strength on an auditory-evoked N1m. NeuroReport 19, 783–787.

Soulodre, G. A., Norcross, S. G., 2003. Objective Measures of Loudness, Proc. AES 115th Convention.

Stevens, S. S., 1957. On the Psychophysical Law. Psychol. Rev. Vol. 64, 153-181.

Stevens S. S., 1958. Adaptation-Level vs the Relativity of Judgment. The American Journal of Psychology 71, 633-646.

Sur, S., Sinha, V., 2009. Event-related potential: An overview. Ind. Psychiatry J., vol. 18, no. 1, 70-73.

Susini, P., McAdams, S., Smith, B. K., 2002. Global and Continuous Loudness Estimation of Time-Varying Levels. Acta Acustica united with Acustica Vol. 88, 536-548.

Teplan, M., 2002. Fundamentals of EEG measurement. Measurement science review Vol. 2, Sec. 2.

Thwaites, A., Glasberg, B. R., Nimmo-Smith, I., Marslen-Wilson, W. D., Moore, B. C. J., 2016. Representation of Instantaneous and Short-Term Loudness in the Human Cortex. Frontiers in Neuroscience Vol. 10, Article 183.

Todeshini, R., 2010. Useful and unuseful summaries of regression models. moleculardescriptors.eu Tutorial 5.

Toepken, S., Weber, R., 2013. Differentiating between Loudness and Preference in the Case of Multi-tone Stimuli. Proc. Mtgs. Acoust. Vol. 19, 050190.

Traunmüller, H., 1990. Analytical expressions for the tonotopic sensory scale. J. Acoust. Soc. Am. Vol. 88, No. 1, 97-100.

Treisman M., 1984. A theory of criterion setting: An alternative to the attention band and response ratio hypotheses in magnitude estimation and cross-modality matching. Journal of Experimental Psychology: General 113, 443-463.

Tremblay K.L., Ross B., Inoue K., McClannahan K., Collet G., 2014. Is the auditory P2 response a biomarker of learning? Frontiers in Systems Neuroscience 8, 1-13.

Tsukida, K., Gupta, M. R., 2011. How to Analyze Paired Comparison Data. UW Electrical Engineering No. UWEETR-2011-0004.

Vardi, Y., Zhang, C., 2000. The multivariate L1-median and associated data depth. PNAS Vol. 97, No. 4, 1423-1426.

Verboven, S., Hubert, M., 2005. LIBRA: a MATLAB Library for Robust Analysis. Chemometrics and Intelligent Laboratory Systems Vol. 75, 127-136.

Verhey, J. L., Kollmeier, B., 2001. Spectral loudness summation as a function of duration. J. Acoust. Soc. Am. Vol. 111, No. 3, 1349-1358.

Vickers, E., 2010a. Metrics for Quantifying Loudness and Dynamics. Proc. AES 129th Convention.

Vickers, E., 2010b. The Loudness War: Background, Speculation and Recommendations. Proc. AES 129th Convention.

Wagner E., Scharf B., 2006. Induced loudness reduction as a function of exposure time and signal frequency. J. Acoust. Soc. Am. Vol. 119, No. 2, 1012-1020.

Wickelmaier, F., Schmid, C., 2004. A Matlab function to estimate choice model parameters from paired-comparison data. Behavior Research Methods Vol. 36, No. 1, 29-40.

Worden, F. G., Marsh, J. T., 1968. Frequency following (microphonic-like) neural responses evoked by sound. Electroencephalogr. Clin. Neurophysiol. 25, 45-52.

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

Zwicker, E. und Fastl, H., 1999. Psychoacoustics: Facts and models, Bd. 3. Springer Berlin.

Danksagung

Gott sei Dank, dass ich so viele Unterstützer für diese Arbeit hatte: meine Eltern Astrid und Axel und meine Familie, meine Freunde und Kollegen und Birger, von dem ich viel lernen durfte und der mir diese Arbeit ermöglicht hat.

Abschließende Erklärung

Ich versichere hiermit, dass ich die vorliegende Dissertation selbständig verfasst und nur die angegebenen Hilfsmittel verwendet habe.

_____ Oldenburg, den _____