Comparison of loudness models for time-varying sounds[3]

Jan Rennies[1)*], Jesko L. Verhey[1)], Hugo Fastl[2)]

[1)] AG Neuroakustik, Institut für Physik, Carl von Ossietzky Universität Oldenburg, D-26111 Oldenburg, Germany. jan.rennies@uni-oldenburg.de

[2)] AG Technische Akustik, Lehrstuhl für Mensch-Maschine-Kommunikation, Technische Universität München, D-80333 München, Germany

Summary

The loudness of a sound depends, among other parameters, on its temporal shape. Different loudness models were proposed to account for temporal aspects in loudness perception. To investigate different dynamic concepts for modeling loudness, predictions were made with the two current loudness models of Glasberg and Moore [J. Acoust. Soc. Am. **50**, 331–341 (2002)] and Chalupper and Fastl [Acta Acustica united with Acustica **88**, 378–386 (2002)] for a set of time-varying sounds. The predicted effects of duration, repetition rate, amplitude-modulation, temporal asymmetry, frequency modulation and the systematic variation of spectro-temporal structure on loudness were compared to data from the literature. Both models predicted the general trends of the data for single, repeated and asymmetric sound bursts and amplitude-modulated sounds. The model of Chalupper and Fastl seems to agree slightly better with loudness data for sounds with strong spectral variations over time, since it includes a dynamic stage which allows spectral loudness summation also for non-synchronous frequency components.

PACS no. 43.66.Cb, 43.66.Ba, 43.66.Mk

1. Introduction

Models for the prediction of loudness are valuable tools since they can at least partly replace time consuming subjective test. Accordingly, they are applied in a number of fields, e.g. in the assessment of noise emissions or the development and optimization of algorithms in hearing aids. Due to the practical relevance, different standards have been developed describing procedures to compute loudness (e.g. [2, 3]). However, all current loudness models are limited in their applicability to some extend. For example, the standardized procedures to calculate loudness mentioned above only provide valid loudness values for signals that are stationary. Since it is desirable to have a loudness model applicable to a wider range of sounds, it is first necessary to know the capabilities and limitations of current loudness models. This study compares the predictions of two elaborate current loudness models representing different concepts for a set of time-varying sounds.

In general, loudness models can be subdivided into models for stationary signals and those for time-varying sounds. Models for stationary signals disregard temporal properties of the sound and are based on the long-term spectrum of the signal. Apart from a weighting of the frequencies they also account for the effect of bandwidth on the overall loudness. If the bandwidth of a sound is varied while keeping the overall intensity fixed, loudness remains constant as long as the bandwidth is smaller than a critical bandwidth, for larger bandwidths, loudness increases (e.g. [4, 5, 6, 7, 8, 9]). This effect called spectral loudness summation is believed to result from an analysis of the incoming sound by a bank of overlapping critical-band filters followed by a compressive nonlinearity in each filter that transforms the intensity to specific loudness, and a final loudness summation across channels. The bandwidth of the auditory filters and the amount of compression affect spectral loudness summation, i.e. the narrower the auditory filters and the higher the compression, the larger the amount of spectral loudness summation (see [10]). This concept of spectral loudness summation has been implemented in a number of loudness models, which successfully predict the loudness of stationary sounds as perceived by normal-hearing (e.g. [11, 12, 13, 14, 15, 16, 17]) and hearing-impaired listeners (e.g. [18, 19, 20, 21]).

Most natural sounds, however, are non-stationary and have time-varying properties which also affect their loudness. For example, several studies found that loudness of sounds with the same intensity increases with duration (e.g. [22, 23, 24, 25, 26, 27, 28, 29, 30]). This effect is commonly referred to as temporal integration of loudness. It is

* now at the Fraunhofer Institute for Digital Media Technology, Oldenburg, Germany

[3] This study was partly done at the Technical University of Munich and partly at the University of Oldenburg. Part of the results were presented at the joint DAGA/NAG conference in Rotterdam in March 2009 [1].
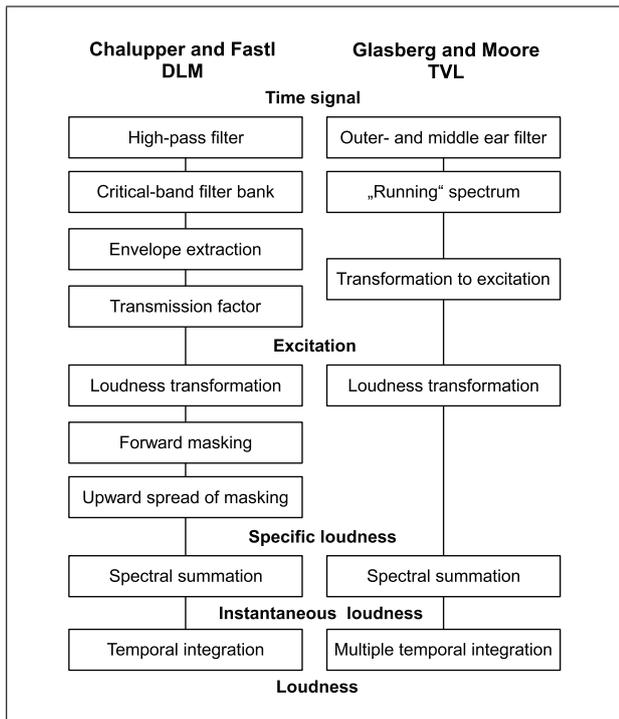
Figure 1. Schematic structure of the loudness models [20] (DLM, left) and [38] (TVL, right).

usually modeled by assuming that the intensity or some other transformation of the signal is analyzed by a leaky integrator (see e.g. [26]). There is considerable variability with respect to the time constants of the leaky integrator, ranging from 25 ms [31, 32, 33] and 100 ms [25] to 200 ms [22, 34, 35]. Some studies indicate that temporal integration may involve more than one time constant (e.g. [26]). In addition, the decay time of the leaky integrator may be longer than the rise time as suggested by e.g. Port [23], Kumagai *et al.* [36] and Ogura *et al.* [37] to account for the loudness of repeated sound bursts.

The are only a few models that were proposed to account for both spectral and temporal aspects of loudness, among them are the models of Chalupper and Fastl [20] and Glasberg and Moore [38]. Both models are based on the model originally developed by Zwicker [12, 14] and thus have a similar general structure. However, there exist some fundamental differences as far as their dynamic properties are concerned. For example, Chalupper and Fastl [20] only used a single time constant, but included temporal effects of post-masking to describe the dynamic behavior of their model. Glasberg and Moore [38] implemented a more elaborate temporal integration stage using several time constants in order to correctly predict loudness of amplitude-modulated sounds.

In this study, predictions of the two models are compared to data. For temporal integration of loudness, data shown by Poulsen [26] and Pedersen *et al.* [39] were used. The large number of participants (up to approximately 300 listeners) in the latter study ensured that the data resemble loudness perception of the average normal-hearing listener. The comparison of predictions to the data pro-

vide insights into the accuracy of the attack time constant of the temporal integration stage in the models. Unfortunately, a large data set as the one of temporal integration does not exist for other aspects of temporal loudness perception. Thus, for loudness of sequences of noise bursts, data from a study of Port [23] were used. The loudness of sequences of noise bursts provides insights into the release time constants of the models. For the loudness of amplitude-modulated sounds, model predictions were compared to the results of several studies [40, 41, 42, 43]. The comparison was included in the present study because the models show conceptual differences in the procedure to calculate loudness for this type of stimuli. For the loudness of temporally asymmetric stimuli, data of Stecker and Hafter [44] were used. Stecker and Hafter [44] showed that the auditory image model (AIM, [45]) failed to predict this temporal aspect of loudness. The present study investigates if this is also true for the two elaborate loudness models. Finally, the interplay of the temporal and spectral characteristics of the loudness models is studied using loudness data for sounds with time-varying spectra: on the one hand, predictions are compared to data of Zwicker [46] on the loudness of stimuli with a distinct spectro-temporal pattern. On the other hand, the ability of the models to predict the data of Zwicker [47, 48] for loudness of frequency-modulated sinusoids is investigated.

2. Model structures

2.1. Loudness model by Chalupper and Fastl

The Dynamic Loudness Model (DLM, [20]) accounts for several aspects of dynamic loudness perception. The basic structure of the DLM is illustrated on the left side of Figure 1. The input time signal is high-pass filtered using a Butterworth filter with a cut-off frequency of 50 Hz to account for the lower limit of the audible frequency range. In the following stage, a bank of overlapping auditory filters is applied. The frequency basis of the filter bank is the critical-band-rate scale or z-scale [49, 50], which approximates the frequency representation in the inner ear.

Relative to this scale, the DLM uses 24 equidistant filters with center frequencies from 50 Hz (0.5 Bark) to 13500 Hz (23.5 Bark), i.e. with center frequencies equivalent to the critical-band center frequencies as described by Zwicker [49]. Accordingly, the width Δf of the filters is the critical bandwidth (CB, in Hz), which is related to center frequency f_c (in kHz) by

$$\Delta f = 25 + 75[1 + 1.4 f_c^2]^{0.69}. \tag{1}$$

At the output of the filter-bank stage, 24 band-pass filtered time signals are available. For each channel, a temporal window with an equivalent rectangular duration (ERD) of 4 ms is temporally shifted along the signal in steps of 2 ms to compute the short-term root-mean-square (rms) value. The form of the temporal window was chosen according to masking experiments by Moore *et al.* [51] and Plack and Moore [52], who suggested to describe each side of

the window as the sum of two rounded-exponential functions. The transmission of sound from free-field through outer and middle ear is accounted for by a correction factor in the next stage, resulting in the quantity excitation E. The excitation is then transformed to specific loudness in several steps. Firstly, the quantity main loudness is calculated applying the compressive relation between excitation and loudness and accounting for loudness near threshold in a way very similar to the original model [12, 14]. The exponent describing the compression has a value of 0.23. Then, effects of forward masking are included (the influence of backward masking is neglected). This is achieved in a non-linear stage by appending temporal tails to peaks of the specific loudness. The time constants are chosen according to forward masking experiments by Zwicker [53], and accordingly depend on level and duration (see [54] and [55] for details).

Subsequently, spectral masking is accounted for according to DIN 45631 [2]. The resulting specific loudness-time pattern $N'(z, t)$ is then integrated along the frequency dimension, which gives the so-called instantaneous loudness as a function of time. This instantaneous loudness can be interpreted as "an intervening variable which is not available for conscious perception" [38]. In the final stage of the model, the instantaneous loudness is integrated using a first-order low-pass filter with a cut-off frequency of 8 Hz. The resulting quantity is called short-term loudness and can be described as "the loudness perceived at any instant" [38]. When the loudness of different sounds is compared, an assessment of the overall loudness is required. Zwicker [46, 48] found that the peak value of the short-term loudness is the dominant aspect when globally judging the loudness of short sounds. Accordingly, for the simulations in the present study, the peak value of the short-term loudness was taken as an estimate of the global loudness for the DLM for all simulations.

2.2. Loudness model by Glasberg and Moore

Glasberg and Moore [38] developed a loudness model applicable to time-varying sounds (time-varying loudness model, referred to as TVL in this study) on the basis of their earlier models for stationary sounds [15, 17], which were in turn based on the loudness model by Zwicker [12, 14]. The general structure is schematically shown on the right side of Figure 1. As in the DLM, the time signal of the stimulus is used as input to the model. A fixed filter represents the combined effect of the transfer function from free-field to ear drum and of the transmission through the middle ear. As an intermediate variable, the excitation pattern is calculated from the effective spectrum reaching the cochlea (i.e. after accounting for the transfer through outer and middle ear). To obtain a spectrum which approximates the spectral and temporal resolution of the hearing system for the different frequency regions, the filtered time signal is analyzed using six parallel Fast Fourier Transforms (FFTs), each assigned to a different frequency range. Hamming windows with lengths of 64, 32, 16, 8, 4, and 2 ms are used to compute components

at the frequency regions 20 to 80 Hz, 80 to 500 Hz, 500 to 1250 Hz, 1250 to 2450 Hz, 2450 to 4050 Hz, and 4050 to 15000 Hz, respectively. The short-term spectra are calculated by shifting the temporal analysis windows - all aligned at their temporal centers - along the time signal in steps of 1 ms. Each millisecond, the excitation pattern is calculated from the resulting spectra in the same way as in the stationary model [17], accounting for the width of the auditory filters, their level dependence and their variation with center frequency. For the same reasons as Chalupper and Fastl [20], Glasberg and Moore [38] use a transformed frequency scale which more closely relates to the representation of sound in the auditory system than a liner frequency scale in Hertz. However, instead of the critical-band-rate scale, they use a scale based on the equivalent rectangular bandwidth (ERB) of the auditory filters, as estimated in notched-noise experiments [56, 57, 58].

The ERB (in Hz) as a function of center frequency f_c (in kHz) can be described by

$$\text{ERB} = 24.7(4.37 f_c + 1). \qquad (2)$$

The excitation pattern is transformed into specific loudness as in the stationary loudness model [17]. As in the DLM, compression and the influence of hearing threshold are included in the transformation. The compressive exponent has a value of 0.20. The specific loudness is then summed across frequency. After this stage of the model, instantaneous loudness is available at a sampling rate of 1 ms, i.e. the same rate at which the spectra and excitation patterns are computed. The instantaneous loudness, which closely follows the temporal envelope of the input signal, is integrated using an attack time constant of about 22 ms and a release time constant of about 50 ms, resulting in the short-term loudness. The short-term loudness is subsequently integrated with a temporal window similar to the one used for the derivation of the short-term loudness but now with longer time constants: 99 ms for the attack time constant and 2000 ms for the release time constant. The resulting long-term loudness is meant to describe loudness sensations that are built rather slowly, e.g. for sounds modulated at a very slow rate. As for the DLM, the maximum of the short-term loudness was used to assess the overall loudness of short stimuli. For longer signals, the complex temporal integration stage in the TVL offers more flexibility to account for the loudness of slowly time-varying stimuli. Glasberg and Moore [38] suggested to use the mean value of the long-term loudness when describing the loudness of amplitude-modulated sounds. This suggestion was followed in the present study for both amplitude and frequency-modulated stimuli.

2.3. Principal differences between the models

While the two models share a common general structure, there are some fundamental differences, which influence the predictions of loudness for stationary and time-varying sounds. One difference is the way the models account for the transmission characteristics of outer and middle ear.

While the DLM uses a transmission factor, the TVL uses a fixed filter prior to auditory filtering. The structure of the DLM was motivated by a standard (DIN 45631), whereas the stage of the TVL seems to be more reasonable from the physiological point of view. Furthermore, the frequency scales and widths of the auditory filters in the two models differ, since they are based on the CB (DLM), or the ERB (TVL). The differences between CB and ERB are largest at frequencies below 500 Hz, where the CB is constant at about 100 Hz, while the ERB decreases monotonically with frequency down to about 25 Hz. Thus, especially for sounds with low-frequency components, different predictions of the two models may occur. In the original implementation of the TVL, Glasberg and Moore [38] allowed for a binaural sound input by calculating the loudness at each ear separately and finally summing the loudnesses across ears to compute the overall loudness. While this principle was refined in the mean time based on more recent experimental evidence [16], the fact that, in principle, binaural loudness can be computed may be regarded as an advantage of the TVL over the DLM, which only computes loudness of diotic stimuli. Since the main topic of the current study is the prediction of loudness of time-varying sounds, this potential advantage will not be considered in the following.

The functions relating specific loudness to level are different in the two models (cf. [59]). On the one hand, both models account for the influence of absolute threshold, which results in a steeper increase of loudness with level below about 40 dB SPL for narrowband signals. On the other hand, the DLM assumes a simple exponential increase for levels larger than 40 dB SPL, while the TVL predicts a steeper, less compressive loudness growth at very high levels (≥ 100 dB). For modeling temporal aspects of loudness, another fundamental difference is that the TVL computes both short-term and long-term loudness, while only short-term loudness is derived from the instantaneous loudness in the DLM using a simple low-pass filter. This means that attack and release time constants for temporal integration are the same in the DLM, while the release time constants are longer than the attack time constants in the TVL for both short-term loudness and long-term loudness. In principle, this more complex integration stage makes the TVL more flexible since more parameters are used to predict loudness of time-varying sounds. On the other hand it implies that, for a single paradigm, there are several possibilities to compute loudness and the user of the model has to decide whether to use long-term or short-term loudness. Finally, the models differ in the exact numerical values of the internal parameters, in particular in the compressive exponent and the time constants used to describe temporal integration.

3. Predictions

3.1. Temporal integration of loudness for single tone bursts

Figure 2 illustrates how loudness integrates over time by showing the response of the loudness models to a tone
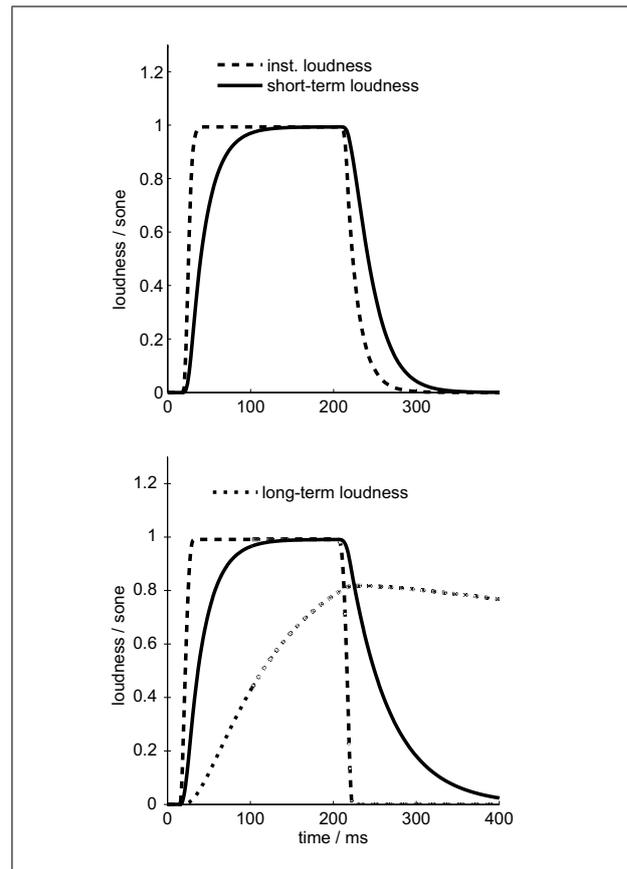

Figure 2. Predicted loudness in response to a 1-kHz tone burst with a level 40 dB SPL and a duration of 200 ms obtained with the DLM (top) and the TVL (bottom). Both panels show the instantaneous loudness (dashed line) and short-term loudness (solid line). For the TVL, the dotted line additionally indicates the predicted long-term loudness.

pulse. The figure shows the instantaneous (dashed line) and short-term loudness (solid line) calculated by both models in response to a 1-kHz tone burst at 40 dB SPL and a duration of 200 ms including 10-ms raised-cosine ramps at on- and offset. Additionally, for the TVL, the long-term loudness is shown (dotted line, right panel). The instantaneous loudness closely follows the excitation in the TVL, while a slower decay is calculated by the DLM. This is due to the post-masking included before the final temporal integration. The short-term loudness shows a very similar built-up in both models, while the decay is faster in the DLM. The short-term loudness reaches a value of 1 sone in both models, i.e. the loudness reaches the value of a continuous 1-kHz tone. This is expected, since a duration of 200 ms is in the range of time constants typically used to describe temporal integration of loudness (see e.g. [60]). Thus, for this type of stimuli, the maximum of the short-term loudness is a good estimate of the overall loudness. The long-term loudness does not reach a stationary value of 1 sone within 200 ms. It should be noted that the long-term loudness was not meant to describe stimuli like short tone burst, but rather to assess the loudness of longer stimuli (see below).
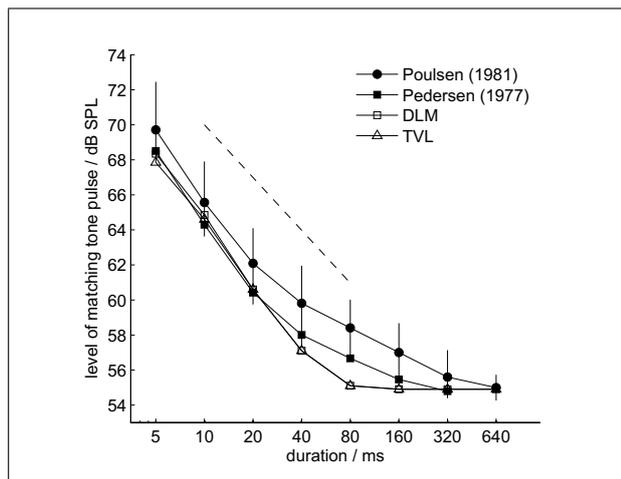
Figure 3. Sound pressure level at equal loudness of 1-kHz tone bursts as a function of burst duration measured by Poulsen [26, filled circles] and in an international round Robin test [39, filled squares]. Error bars represent 95%-confidence intervals of the data of Poulsen [26]. Predictions by the DLM (squares) and the TVL (triangles) are shown with open symbols. The dashed line indicates −3 dB per doubling of duration.

A subset of the data of Poulsen [26] on temporal integration of loudness is represented by filled circles in Figure 3. The sound pressure level at equal loudness is shown as a function of duration for 1-kHz tone bursts. Error bars indicate 95% confidence limits. The level of the longest duration (640 ms) was 55 dB SPL and all stimuli were filtered using a third-octave filter to avoid clicks. After filtering, the levels were corrected such that filtered and unfiltered signals had the same level. For comparison, data of Pedersen *et al.* [39] collected in an international Round Robin Test are shown. This set of data represents results of approximately 300 subjects measured in 21 laboratories. The two studies measured a comparable dependence of loudness on duration, although the effect was slightly smaller in the Round Robin Test. The data suggest a monotonic decrease of sound pressure level with increasing duration for durations smaller than 320 ms (i.e., 640 ms for the data of Poulsen [26]). The slope is slightly less than −3 dB per doubling for durations larger than 40 ms (see dashed line in Figure 3). It is slightly steeper than −3 dB per doubling of duration for short durations smaller than 20 ms. The corresponding predictions of the DLM and the TVL are very similar as indicated by open symbols. For durations below about 40 ms, the predicted level differences are in good agreement with the data of Pedersen *et al.* [39]. The predicted slopes are slightly steeper than −3 dB per doubling. When the duration increases beyond 80 ms, no increase in loudness is predicted by the models.

3.2. Temporal integration of loudness for repeated noise bursts

The loudness of tone pulses was also investigated for repeated pulses as a function of repetition rate. Figure 4 shows the data of Port [23, filled circles], which were mea-

sured in a matching experiment determining the level at equal loudness of a sequence of 2-ms noise bursts and a continuous reference noise for repetition rates between 2 and 500 Hz. The overall duration of sequence and reference was 1.2 s. The noise was 550 Hz wide and centered at 2.5 kHz. Prior to presentation to the subjects, signals were filtered using an one-third octave filter to avoid clicks. The data show a monotonic decrease of the level difference with increasing repetition rate. For the shortest repetition rate of 2 Hz, the sequence needed an about 19-dB greater level to be perceived as equally loud as the continuous noise. At the largest repetition rate of 500 Hz, i.e. when the single bursts were directly concatenated, the levels of sequence and equally loud reference noise were the same. The interquartile ranges in Port's study varied between about 3 and 10 dB for the individual data points. Open symbols in Figure 4 represent the predicted level differences of the DLM (squares) and the TVL (triangles), which were obtained as the mean of ten simulation runs. Lines indicate standard deviations. Additionally, the dashed line shows predictions obtained using the average of the long-term loudness. In agreement with the data, the simulated level difference decreases with increasing repetition rate, i.e. both models correctly predict an increasing loudness of a sequence of short bursts as the repetition rate is increased. At the largest rate of 500 Hz, the level difference is 0 dB. At lower repetition rates, the predicted level differences based on the short-term loudness are slightly larger than measured by Port [23]. On average, the overestimation amounts to about 3 dB for the DLM and about 6 dB for the TVL. Thus, especially at low repetition rates, the TVL underestimates the loudness of sequence and predicts a greater level difference at equal loudness. The predicted decrease of the level difference with increasing repetition rate is steeper in the TVL than in the DLM. Predictions based on the long-term loudness differ considerably from the data. For the slowest repetition rates, the predicted level difference is more than 20 dB greater than observed by Port [23].

3.3. Loudness of temporally asymmetric signals

Stecker and Hafter [44] measured the loudness of bursts of tones with the same duration, but with either quickly rising and slowly falling level (damped stimuli) or vice versa (ramped stimuli). For tone frequencies between 330 and 6000 Hz, they found that loudness was larger for ramped than for damped stimuli, although spectrum, duration and intensity were the same. This recency effect could only be partly modeled with the auditory image model (AIM, [45]). Since AIM was not explicitly designed as a loudness model, loudness of the same stimuli as used by Stecker and Hafter [44] was predicted with the two loudness models under consideration in the present study to investigate if the models can account for such a temporal asymmetry in loudness perception. The lower panels of Figure 5 show the predicted instantaneous (gray lines) and short-term loudness (black lines) for both models for a carrier frequency of 1.5 kHz and a peak level of 80 dB SPL for
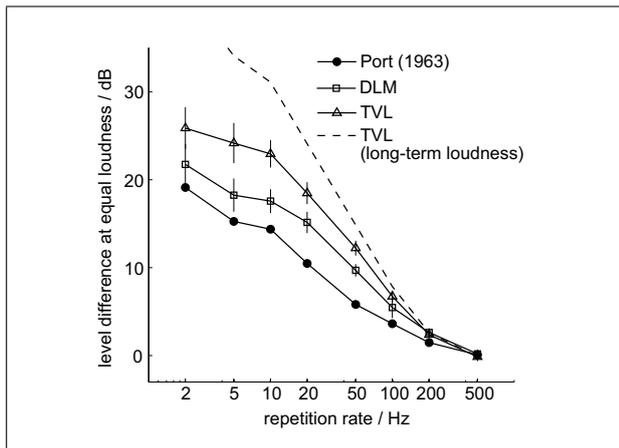
Figure 4. Level difference between repeated 2-ms bursts of narrowband noise and an equally loud continuous noise as a function of repetition rate. Filled circles show the data measured by Port [23], open squares and triangles represent the predictions of the DLM and the TVL, respectively. Error bars for the simulations represent standard deviations over 10 simulations.

Table I. Loudness in sones as predicted by the models for ramped and damped envelopes for different carrier frequencies as used by Stecker and Hafter [44]. Additionally, level differences between ramped and damped signals at equal loudness as derived from the loudness ratios are indicated in italics.

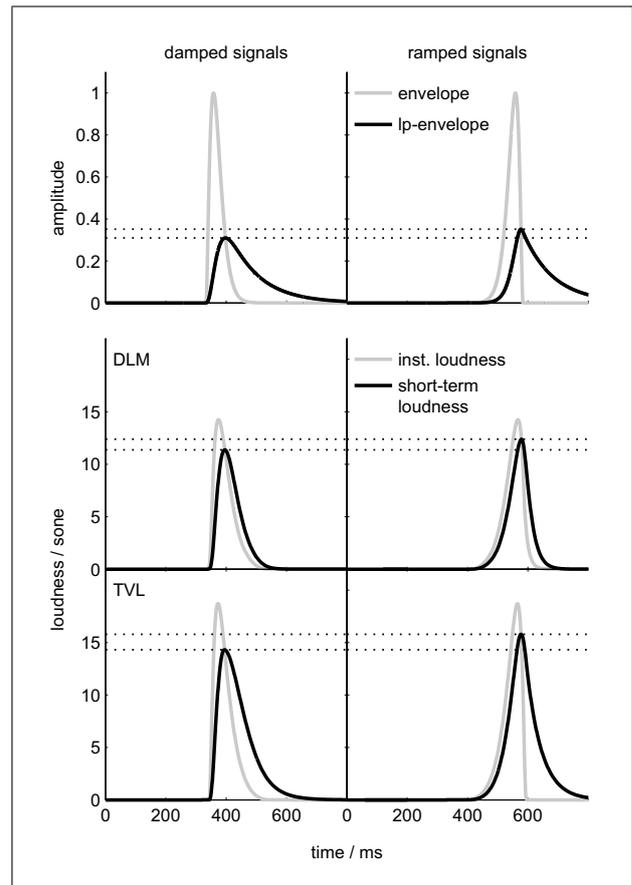|	330 Hz	700 Hz	1500 Hz	3000 Hz	6000 Hz
DLM					
ramped	11.93	12.29	12.38	14.04	11.39
damped	11.02	11.33	11.38	12.89	10.47
ΔL / dB	*1.14*	*1.17*	*1.22*	*1.23*	*1.22*
TVL					
ramped	9.40	12.11	15.79	20.82	6.39
damped	8.76	11.07	14.32	18.83	5.78
ΔL / dB	*1.02*	*1.30*	*1.41*	*1.45*	*1.45*

Figure 5. Top panels: envelopes as used by Stecker and Hafter [44, gray] and the corresponding envelopes after low-pass filtering with $\tau=100$ ms (black) for damped (left) and ramped (right) envelopes. Dotted lines indicate the maxima of the low-pass filtered envelopes. Lower panels: Corresponding instantaneous (gray) and short-term loudness (black) as predicted by the DLM (mid panels) and the TVL (bottom panels) for a carrier frequency of 1.5 kHz and a peak level of 80 dB SPL. Dotted lines represent maxima of the short-term loudness of damped and ramped signals.

both damped (left) and ramped envelopes (right). The parameter describing the widths of the envelopes was *pt*=-3 for the damped stimulus and *pt*=+3 for the ramped stimulus (see Figure 1 in [44]). Dashed lines indicate the maximum of the short-term loudness. Both models predict a greater loudness for the ramped stimulus, when this maximum is taken as a measure of overall loudness as proposed in the two loudness models for this type of stimuli. Table I summarizes the predicted loudnesses for different carrier frequencies modulated by ramped or damped envelopes. The predictions show that loudness is larger for the ramped than for the damped envelope for all carrier frequencies, which agrees with the results of Stecker and Hafter [44]. The level differences needed to predict equal loudness for both envelopes were calculated from the sone ratio and are indicated in italics in Table I. Although the absolute predicted sone values differ between the models, the level differences are similar and lie between 1 dB and 1.5 dB. This quantitatively agrees with the experimental data shown in Figure 2 of [44].

3.4. Loudness of amplitude-modulated sinusoids

A further example of time-varying sounds is an amplitude modulated tone. Bauch [40] measured the level difference between an unmodulated and a sinusoidally modulated carrier tone as a function of modulation frequency. His data for a carrier frequency of 1 kHz, a modulation depth of $m = 0.5$ and a reference level of 45 dB SPL are indicated by filled diamonds in the top panel of Figure 6. He found a negative level difference for modulation rates below about 10 Hz and for high modulation rates above 200 Hz. For intermediate modulation rates, the level difference was close to zero. Results were similar for a carrier frequency of 4 kHz, but the decrease in level difference occurred at higher modulation frequencies (bottom panel of Figure 6). Bauch [40] found no significant difference between the results of two subjects. Zhang and Zeng [41] measured similar results (filled circles) as observed by Bauch [40] for the 1-kHz carrier at the same level and modulation depth for six listeners. Zhang and Zeng [41]

reported a range of two standard deviations of about 3 dB (not shown). Moore *et al.* [42, 43] measured the level difference between equally loud unmodulated and modulated 4-kHz carriers at a modulation depth of $m = 0.5$. Their data for a comparable level (40 dB SL) are indicated by filled triangles in the bottom panel of Figure 6. In contrast to previous studies, they found a slightly positive level difference for intermediate modulation rates and, for very low modulation rates, that the perceived loudness corresponded to a level between the rms level and the peak level. As a measure of the inter-subject variability, the studies report standard errors between about 1 and 5 dB (not shown).

In general, the two loudness models (open symbols) predict similar level differences when unmodulated and sinusoidally modulated tones are matched in loudness. In particular, a negative level difference is only predicted for low and high modulation rates. For modulation rates below about 200 Hz, the difference between the predictions is less than one dB. For modulation rates above 200 Hz, the DLM predicts slightly smaller level differences than the TVL. For this range of modulation frequencies, the TVL provides a better fit to the data of [40] and [41].

3.5. Loudness of frequency-modulated sinusoids

Zwicker [47] measured the loudness of strongly frequency-modulated (FM) sounds using a carrier frequency of 1.5 kHz and a frequency deviation of 700 Hz, i.e. the instantaneous frequency varied between 800 and 2200 Hz. Figure 7 shows the level difference between a frequency-modulated tone of 50 dB SPL and an equally loud unmodulated tone at the carrier frequency as a function of modulation rate (filled circles)[1]. Triangles and squares indicate the corresponding predictions of the two loudness models.

The data of Zwicker [47] are characterized by two quasi-steady-state conditions and a transition region. An almost constant level difference was found for modulation frequencies up to about 16 Hz. For intermediate modulation frequencies, the level differences increased with modulation frequency up to a maximum level difference, which was reached at about 64 Hz. For modulation frequencies larger than 64 Hz, the level difference was independent of modulation frequency. The interquartile ranges at medium levels varied between 4 and 12 dB. The simulations show that both models in principle reproduce the results, i.e. both predict a constant level difference for low modulation frequencies and another, higher steady state for large modulation frequencies. In general, predicted level differences are larger than in the experimental data. At low modulation frequencies between 1 and 16 Hz, the average level difference is 0 dB in the data, 2 dB predicted by the DLM and almost 3 dB predicted by the TVL. For large modulation
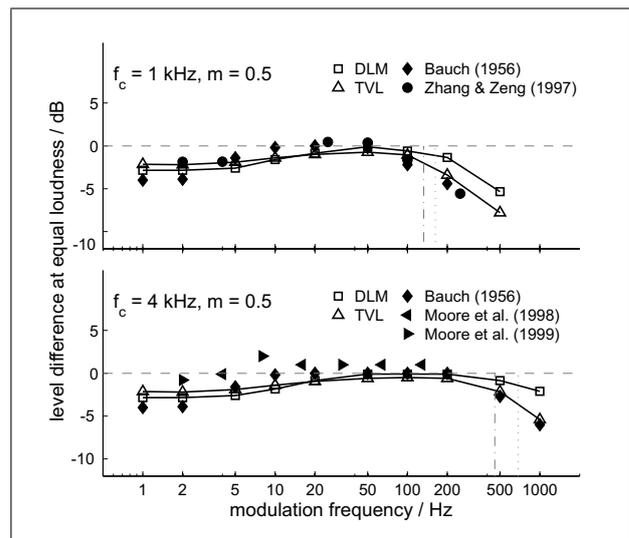

Figure 6. Data on the influence of amplitude modulation on loudness (filled symbols) and the corresponding predictions of the DLM (open squares) and the TVL (open triangles). The level difference between the unmodulated and the modulated carrier at equal loudness is shown for carrier frequencies of 1 kHz (top) and 4 kHz (bottom) as a function of modulation frequency. The modulation depth was $m = 0.5$, the level of the modulated sounds was 45 dB SPL for both carrier frequencies. Critical bandwidth (dotted) and equivalent rectangular bandwidth (dash-dotted) at the carrier frequencies are indicated by vertical lines.
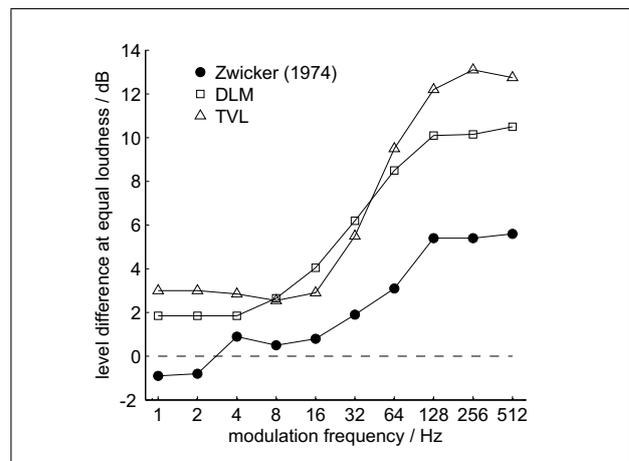

Figure 7. Data on the influence of frequency modulation on loudness (filled circles) taken from [47] and corresponding predictions of the DLM (squares) and the TVL (triangles). The level difference between unmodulated and equally loud modulated carrier at 1.5 kHz is shown as a function of the modulation frequency. The frequency deviation was 700 Hz.

frequencies, the deviations between data and predictions amount to about 4 dB for the DLM and 7 dB for the TVL.

3.6. Loudness of pulses forming different spectro-temporal patterns

Zwicker [46] used trains of tone pulses whose temporal and spectral structures were varied systematically in order to investigate the interaction of loudness integration

---

[1] Zwicker did not adjust the level of an unmodulated tone to match the loudness of the modulated tones, but used a critical-band wide noise as comparison signal [47]. In the present study, results are presented as level difference between unmodulated and modulated tone as derived from Zwicker's data.
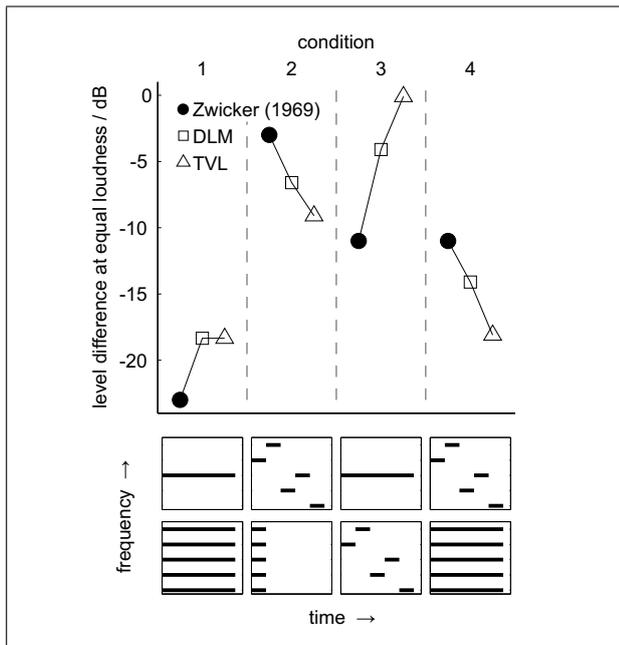
Figure 8. Level differences between test and reference signals at equal loudness for four experimental conditions. Filled circles indicate data of Zwicker [46], open squares and triangles represent predictions of the DLM and the TVL, respectively. The pictograms below show the spectro-temporal pattern of the test (lower row) and reference (upper row) stimuli for each condition.

over time and frequency. Figure 8 shows a subset of his data (filled circles) and predictions of the DLM and the TVL for four different test and reference signals. For each condition, the level difference between test and reference stimulus is given. The pictograms in the lower part of Figure 8 indicate the spectro-temporal structure of reference and test signal (upper and lower row, respectively). In general, all level differences measured by Zwicker [46] were negative indicating that the level of the test stimulus was always lower than that of the reference in order to obtain equal loudness. The interquartile ranges reported by Zwicker [46] varied between the different conditions and ranged from about 3 to 11 dB.

In the first condition, a reference tone pulse of 100 ms duration, a level of 70 dB SPL and a frequency of 1.85 kHz was matched in loudness to a stimulus, which consisted of the sum of five 100-ms tone pulses of frequencies 1000, 1370, 1850, 2500, and 3400 Hz[2]. Each pulse of the latter had a level of 70 dB SPL. The given level difference indicated in Figure 8 is calculated as the difference between the reference level and the level of each of the five pulses. The results indicate that the reference tone had to be about 23 dB higher in level to be perceived as equally loud as the test stimulus. Both models slightly underestimate the level difference in this condition and predict only about 18 dB.

---

[2] Zwicker did not discuss how he avoided clicks at stimulus on- and off-sets [46]. In the simulations of the present study, $\cos^2$-ramps of 2.5 ms were used to reduce the influence of spectral broadening.

In the second condition, the reference stimulus was a pulse train of five 20-ms pulses without pauses between the pulses. Each pulse had a level of 70 dB SPL. The frequencies of the pulses had the fixed temporal order 1370, 2500, 1000, 3400 and 1850 Hz. The test signal was a 20-ms burst, which consisted of the sum of five pulses with the same frequencies. The experimental data indicate only a small difference of about 3 dB between the equally loud reference and test stimulus. Both models predict larger differences of 6.5 dB (DLM) and 9 dB (TVL).

In the third condition, the same train of 20-ms tone pulses as above was matched in loudness to a 100-ms test tone of 1.85 kHz at 70 dB SPL. As in the first condition, the level difference was based on the reference level and the level of each pulse of the test stimulus at equal loudness. While the measured data show a level difference of about 11 dB, the TVL predicts the same level at equal loudness for these two stimuli; an effect of about 4 dB is indicated by the DLM.

In the fourth condition, the loudness of the pulse train was compared to that of the sum of five 100-ms tone bursts. Both models overestimate the experimentally found level difference of 11 dB by 3 dB (DLM) and 7 dB (TVL).

In summary, for the given stimuli both models predict the same level difference in a classical paradigm where the loudness of a narrowband stimulus is compared to that of a broadband stimulus (condition 1) and slightly underestimate the measured effect. In the remaining conditions, the predictions of the DLM are always closer to the experimental data than those of the TVL. The experimental data as well as the model predictions are self-consistent. Comparing a tone to a sum of five tones (condition 1) yields the same level difference as the combined effect found in conditions 3 and 4, where the same tone is matched in loudness to the pulse train (condition 3) and the pulse train is matched to the sum of five tones (condition 4).

4. Discussion

4.1. Temporal integration of loudness

For short sound bursts as considered the present study, the peak of the short-term loudness is a reasonable measure for the overall loudness (see e.g. [48, 38]). Thus, the attack time constant of the temporal integration stage determines overall loudness, since the peak value is not affected by the shape of the loudness decay. The simulations of the present study showed that both loudness models could accurately predict loudness of short tone bursts up to about 40 ms. However, when the duration was increased beyond about 80 ms, no increase of loudness was predicted by either model. In contrast, the data of Poulsen [26] suggest that temporal integration continuous at least up to durations of 320 ms, i.e. the implemented integration stages saturate earlier than suggested by the data. Poulsen [26] also found this result for different levels and frequencies (not shown here), which was in close agreement with results of the international Round Robin Test on impulsive

noise [39] involving a large number of subjects in different laboratories. Poulsen [26] argued that, if a model with only a single time constant was used, best agreement between data and predictions for intermediate levels was obtained with a time constant of about 100 ms, which was in line with data of Zwislocki [35]. Additionally, Poulsen [26] proposed a model with a longer ($\tau \approx 100$ ms) and a shorter time constant ($\tau \approx 5$ ms) to account for the steeper increase of loudness for very short durations at intermediate and high levels. Other studies varied in their results on the time constants underlying temporal integration of loudness. Takeshima *et al.* [27] found that the loudness of 1-kHz tone bursts increased even for durations up to 10 s, which would require a time constant much longer than 100 ms (such as used e.g. in the TVL). The predicted saturation of temporal integration at about 80 ms indicates that the effective attack time constants used to compute the short-term loudness in both models are slightly too short to account for the data of Poulsen [26] and Pedersen *et al.* [39].

While attack time constants can be estimated with single bursts, the investigation of repeated noise bursts can give an insight into the release time constants underlying the perception of loudness. When, effectively, a fast decay time constant determines loudness perception, slowly repeated bursts are processed quasi-independently, while a slow decay time constant results in a combined processing of repeated bursts already at relatively low repetition rates.

A comparison of data and predictions in the paradigm of Port [23] shows that both models predict a larger level difference for slowly repeated 2-ms bursts. Since the predicted and measured level differences decay to 0 dB at the largest repetition rate, the predicted decay of the level difference is slightly steeper than observed by Port [23], especially in the TVL. Part of this effect may be due to differences in the calibration of the signals. Unlike Poulsen [26], Port [23] did not mention an adjustment of the levels after band-pass filtering the signals to ensure the same level before and after the filtering. Accordingly, no such level adjustment was made in the simulations. Especially for short bursts, band-pass filtering reduces the overall level. To obtain equal loudness, this results in a corresponding larger level difference for sequences of slowly repeated noise bursts.

Another possible reason for the differences between model predictions and experimental data is the value of the release time constants. For example, the measured level difference decays by about 9 dB between repetition rates of 2 to 20 Hz, whereas the models predict a decay of 5.5 to 6 dB. This indicates that the release time constant in the models may be too fast.

One may argue that the auditory processing of a sequence of short bursts is similar to that of amplitude-modulated sounds (e.g. [61, 62]). For such sounds, Glasberg and Moore [38] suggested to use a mean value of the long-term loudness as a measure of the overall loudness. However, the predictions in Figure 4 show that the loudness of repeated 2-ms bursts cannot be accurately de-

scribed using this measure. In particular for slow repetition rates, the predicted level difference is too large. This results from the slow built-up time constant used to derive the long-term loudness from the short-term loudness. Thus, a very long decay time constant is not sufficient to describe the dependence of loudness of repeated noise bursts on repetition rate. In summary, the paradigm of Port [23] could be well described using an attack time constant similar to those used in the TVL and the DLM for the short-term loudness in combination with a slightly longer release time constant.

4.2. Temporal asymmetries in loudness perception

Measuring the loudness of ramped and damped envelopes, Stecker and Hafter [44] found a recency effect, i.e. they found that the stimulus whose peak energy was closer to the end was perceived louder than if the peak was close to stimulus onset. They argued that the effect results from "decay suppression", a mechanism that may reduce the effect of reverberation on perception of sound in reverberant rooms. Predictions of the present study showed that both loudness models can quantitatively account for this effect. Thus, "decay suppression" seems to be unnecessary to account for the difference in loudness between ramped and damped sounds. The current predictions seem to be at odds with Stecker and Hafter [44], who concluded that the auditory-image model (AIM, [45]) was unable to account for the effect, since AIM predicted a strong dependence of the effect on the signal frequency which was not found in the data. The present study showed that the two loudness models show a similar loudness difference between ramped and damped sounds for low and high-frequency tones. This discrepancy between the model predictions in [44] and the present study is presumably due to the fact that Stecker and Hafter [44] used a different method to derive overall loudness from the excitation. They [44] used the temporal average of the excitation calculated across the whole stimulus duration as a measure of overall loudness, whereas in the present study, the peak value of the short-term loudness determined loudness which is the common way to determine loudness within the two loudness models for short signals. The top panels of Figure 5 illustrate the influence of using the peak or mean value to derive global judgments. The gray lines represent the normalized temporal envelopes used by Stecker and Hafter [44] for damped (left) and ramped (right) stimuli. The black lines indicate filtered envelopes using a first-order low-pass filter with a time constant of 100 ms, i.e. the simplest approximation of a temporal integration stage. While the mean energy of the low-pass filtered envelopes is the same, the dashed lines show that a larger peak value is reached for the ramped envelope. Thus, already a very simple model of temporal integration can account for this temporal asymmetry in loudness perception, provided the maximum is used to assess overall loudness. A similar asymmetry in the height of the maximum excitation can also be observed for all frequencies in Figure 6 of [44].

4.3. Dynamic processes for loudness of amplitude-modulated sounds

The loudness of a sinusoidally amplitude-modulated carrier tone depends on modulation frequency. The data of different studies suggest that, in principle, three regions can be distinguished: For very low modulation frequencies, the modulated signal is louder than the unmodulated signal at the same level. Bauch [40] argued that the ear was able to follow the slow envelope modulations and the peak amplitudes determined loudness perception. Increasing the modulation frequency impedes the hearing system's ability to closely follow the envelope fluctuations. Accordingly, the magnitude of the level difference between modulated and unmodulated signal at equal loudness decreases. This is true as long as the two side components in the spectrum of the modulated signal are within the critical band of the carrier frequency. For large modulation frequencies, the side components can be resolved by different auditory filters. Loudness is then dominated by spectral loudness summation, which increases the loudness of the modulated signal. Accordingly, the level difference increases. The comparison between simulations and data shown in Figure 6 suggests that the models can predict these main experimental results. For low and intermediate modulation frequencies, the predictions of both models are similar. This is in agreement with the finidngs in [61]. They showed for their data similar predictions of the previous versions of these two models.

For higher modulation rates, the DLM predicts smaller absolute level differences than the TVL and slightly underestimates the level differences measured by Bauch [40]. At least part of this difference can be understood by the different auditory filters used in the models. The critical bandwidth (CB) used in the DLM is larger than the equivalent rectangular bandwidth (ERB), which determines the frequency resolution in the TVL. In Figure 6, CB and ERB are indicated by dotted and dash-dotted vertical lines, respectively, for both carrier frequencies. Thus, the TVL can resolve spectral components at a lower modulation rate and the increase of the level difference between modulated and unmodulated signal occurs at a lower modulation rate. The difference between CB and ERB is larger at a center frequency of 4 kHz than at 1 kHz. Accordingly, the difference between the model predictions is larger at the higher carrier frequency as shown in Figure 6. Another factor, which might add to the different model predictions at large modulation frequencies is the amount of spectral loudness summation. As described in the introduction, the amount of spectral loudness summation depends on the auditory filtering and the compression in each filter. As mentioned above, the TVL uses the slightly narrower ERB instead of the CB. Additionally, the compression is slightly larger than in the DLM (see Sections 2 2.1 and 2.2). Thus, a slightly larger spectral loudness summation is expected in the TVL compared to the DLM. Since the increased loudness of modulated signals at large modulation frequencies is due to spectral loudness summation, the larger level difference predicted by the TVL is expected.

Figure 6 shows that the predictions of both models agree with the experimental data at low and medium modulation frequencies. The TVL gives a better fit at large modulation rates and also predicts that, at very low modulation rates, a level between peak level (corresponds to -3 dB) and rms level (corresponds to 0 dB) determines the loudness of modulated tones. This was not measured by Bauch [40], but agrees with more recent studies (see Figure 6, [42, 43]). These predictions of the TVL result from the more sophisticated temporal-integration stage, which offers several options for a measure of overall loudness. In the present study, the mean of the long-term loudness was taken to assess loudness of amplitude-modulated sounds, as suggested by Glasberg and Moore [38]. However, the predictions of the DLM are generally comparable using a simple low-pass filter to describe temporal integration and the maximum of the short-term loudness as a measure of overall loudness. This indicates that other parameters such as spectral resolution and compression are more important than the choice of time constants for this kind of loudness comparison.

4.4. Dynamic processes for loudness of frequency-modulated sounds

The predictions of the two loudness models for frequency-modulated tones are similar. In agreement with data of Zwicker [47], both models predict a constant level difference for the low modulation rates and no variation of the level difference for high modulation rates. In analogy to the amplitude-modulation paradigm, the ear is able to follow the modulation at low modulation frequencies. In this case, the instantaneously perceived frequency eliciting the largest loudness determines overall loudness. It is likely that the 1-dB larger level difference for the TVL at low modulation rates is due to different frequency-dependent attenuations applied in the models (e.g. middle-ear correction). In the region from 800 to 2200 Hz, i.e. the frequency range covered during one period of the frequency modulation, a slightly different attenuation of the frequency components in the two models may lead to different loudnesses of tones at these frequencies. These different loudnesses then determine the loudness of the frequency-modulated tone for low modulation frequencies.

At large modulation rates, the ear no longer follows the modulation, but integrates over several periods such that, effectively, a broadband signal is perceived. The predicted level difference is about 2 dB larger for the TVL than for the DLM, which is in line with the assumption of an increased spectral loudness summation as discussed above. For intermediate modulation frequencies, the predictions of the models differ slightly. A shallower increase of the level difference with modulation frequency is predicted by the DLM, while the TVL predicts a rather steep transition between the two steady states. The DLM predicts a larger level difference for modulation rates between 8 and 32 Hz, while it is smaller for low and high modulation frequencies. One possible reason for this is that these modulations are too fast for the ear to follow closely, but

that a specific loudness decaying slowly in each channel allows a spectral loudness summation of successive frequencies such that the overall loudness is increased. Since this mechanism is not implemented in the TVL, spectral summation can only take place for components falling into the same temporal analysis window. At large modulation frequencies, this increased loudness in the DLM relative to the TVL is disrupted, since all frequencies are processed quasi-simultaneously. Accordingly, the larger spectral loudness summation of the TVL causes a larger level difference. The measured overall change in level difference from low to high modulation rates is about 7 dB. The DLM predicts about the same dynamic range (difference less than 1 dB), whereas the TVL overestimates this effect by about 3 dB. Thus, the DLM predicts both a slope of the transition and an overall dynamic range, which agree slightly better with Zwicker's data than the TVL. As discussed above, the slope of the transition may be the result of the forward-masking stage, which is not included in the TVL.

4.5. Dynamic processes for loudness of sounds with systematic spectro-temporal patterns

The four conditions taken from Zwicker [46] shed some light on the combinations of spectral and temporal effects in loudness perception. The first condition determines spectral loudness summation for the complex tone. The data indicate that the single tone had a level about 23 dB larger than each tone in the equally loud five-tone complex. About 7 dB ($= 10 \log_{10}(5)$) of the observed effect result from the fact that the given level difference is based on the level of each pulse rather than on the overall level of the test stimulus, as chosen by Zwicker [46]. The remaining 16 dB can be attributed to spectral loudness summation. Both models predict the same level difference between test and reference signal (see Figure 8). This seems to contradict the conclusion of the previous sections that a larger spectral loudness summation is expected in the TVL. However, spectral loudness summation depends on spectral content, center frequency, and level, which were different in the paradigm of Zwicker [46] and the studies discussed above. The different parameters are likely to affect the predicted amount of spectral loudness summation in the two models.

The third condition addresses the question if spectral loudness summation also takes place for non-synchronous frequencies. The TVL does not predict such an effect: it predicts about the same loudness at the same level for the sequence of tones and a tone with the same total duration. Provided that the loudness of the different frequency components of the sequence is similar, this can be expected since the TVL nearly instantaneously integrates across-frequency prior to temporal integration.

In contrast, the data indicate that the sequence required an about 11 dB lower level to be as loud as the tone. Zwicker [46] argued qualitatively that specific loudness should increase rapidly in an excited critical band, but

should decay only slowly such that spectral loudness summation could take place even for non-synchronous tone pulses (see his Figure 8). The DLM includes such a stage: depending on level and duration, this stage appends loudness tails to peaks of the specific loudness. Effectively, this increases the temporal overlap of energy from different auditory filters and allows the effect of spectral loudness summation. As a consequence, the DLM predicts spectral loudness summation for the tone-pulse sequence. This results in smaller differences between predictions and data for the DLM than for the TVL.

The same is true in the fourth condition. The TVL predicts the same level difference for the last and the first condition. This is expected from the predicted level difference (0 dB) for the third condition. The experimental data show a smaller difference which supports the hypothesis that spectral loudness summation still contributes to the loudness of the sequence of tones reducing the difference between the two equally-loud stimuli of condition 4. Since the DLM predicts a residual spectral loudness summation for sequences of different frequencies in the third condition, it also agrees better with Zwicker's data in the fourth condition.

The second condition investigates a trade-off of two effects. On the one hand, only a residual effect of spectral loudness summation is expected for the 100-ms long sequence of tone pulses compared to the 20-ms tone complex. On the other hand, the sequence is five times as long as the tone complex which increases the loudness due to temporal integration. In this condition, Zwicker [46] found only a small level difference of about 3 dB indicating that the effects nearly cancel out one another. In contrast, the TVL underestimates the loudness of the pulse sequence and predicts a level difference of 9 dB. The DLM predicts about 7 dB.

In conclusion, the data indicate that spectral loudness summation also takes place for non-synchronous frequency components. This hypothesis is supported by the predictions of the models. In conditions where the spectral content varies over time, the predictions of the DLM agree better with the data than those of the TVL. However, there are still quantitative differences between 3 dB (condition 4) and 7 dB (condition 3). This means that the influence of slowly decaying specific loudness might not be sufficiently accounted for. If the general amount of predicted spectral loudness summation was matched to the data of the classical spectral-loudness-summation paradigm in condition 1 (e.g. by slightly increasing the compression), the prediction of the TVL would still be 0 dB for condition 3, but that of the DLM would fit better, resulting in an even larger discrepancy between the model predictions.

4.6. Limitations of the models

Although a number of temporal effects in loudness perception can be accounted for by the DLM and the TVL, there remain several limitations of the models. Temporal asymmetries such as found by Stecker and Hafter

[44] were also observed for markedly longer stimuli (e.g. [63, 64, 65, 66, 67, 68]). For example, Susini *et al.* [68] found that sounds with increasing level lead to greater global loudness judgments than sounds with decreasing level for durations between 2 and 20 s. They proposed "that global impressions resulted from a memory process dominated by the last parts of the sound sequence" [68]. All of these studies used stimuli with durations of several seconds or even minutes. It is unlikely that the maximum of the short-term loudness (as used in most of the present study) is an adequate measure of overall loudness for this type of stimuli. The predicted difference between ramped and damped envelopes of the study of Stecker and Hafter [44] was due to the fact that the temporal integration stages of the models did not reach a stationary state in the short overall duration of the stimuli. For slowly varying sounds with durations of several seconds or minutes, the maximum of the short-term loudness would be the same irrespective of the temporal position of the loud parts of the stimulus. It is possible, however, that e.g. the maximum of the long-term loudness as computed in the TVL could predict part of these recency effects, which would constitute an advantage of the TVL over the DLM. In a similar way, longer time constants than used in the DLM are necessary to account for temporal integration data for signals of several seconds [27]. The long-term loudness of the TVL might be a first approximation, while the DLM is not able to explain this effect.

More recent studies have also found primacy effects in loudness perception. Pedersen and Ellermeier [69] and Oberfeld [70] have shown that the beginning of broadband noise contributed significantly more to the overall loudness perception than portions in the temporal center or end. So far, the underlying mechanisms are not completely understood. Given the uncertainty concerning the reason for this effect, neither of the two loudness models contains a mechanism giving special weight to stimulus onsets and, thus, the models cannot predict such a primacy effect.

A temporal effect in loudness perception, which may be related to the above-mentioned primacy effect, is the duration dependence of spectral loudness summation. Several studies have shown that the level difference between broadband and equally loud narrowband noise bursts is considerably larger for short bursts (typically 10 ms) than for long bursts of typically 1000 ms [71, 21, 72, 73]. Rennies *et al.* [74] have shown that both the DLM and the TVL predict the same level difference for short and long signals and fail to predict this effect[3].

Glasberg and Moore [38] argued that their model could not predict the influence of the relative phase of components in a tone complex, since the model is based on the short-term power spectrum. Depending on the relative phases, stimuli with the same power spectrum can evoke different loudness perceptions (e.g. [75]). An auditory filter bank with realistic phase responses would be needed to account for this effect. Neither TVL nor DLM contain such a stage and, thus, cannot predict the effect.

5. Summary and conclusions

The dynamic properties of the dynamic loudness model (DLM, [20]) and the model for time-varying loudness (TVL, [38]) were tested by comparing their predictions with data from the literature on the loudness of time-varying sounds.

In general, the models predict the main trends observed in the data. Both models predict an effect of duration on loudness similar to the data. However, the time constant for temporal integration seems to be slightly too small. The comparison with data on the loudness of amplitude-modulated sounds indicates that, for low to moderate modulation rates, both models predict almost the same effects (difference smaller than 1 dB) of modulation rate on the level difference at equal loudness. This suggests that, at least for amplitude-modulated sounds, it may not be necessary to introduce a second loudness (long-term loudness) as done in the TVL to describe loudness of these sounds. Long-term loudness was also unable to account for the effect of repetition rate on loudness of sequences of noise bursts. One major difference between the TVL and the DLM is that only the latter includes a dynamic stage simulating a slow decay of specific loudness in each auditory channel. This was especially relevant for the paradigm of a study of Zwicker [46]. Using stimuli with a systematically varied spectro-temporal structure, Zwicker [46] found that spectral loudness summation took place even when tone pulses at different frequencies were presented non-synchronously. Predictions of the DLM agreed better with Zwicker's data, although quantitative differences still existed. A slightly better match between the data and predictions of the DLM than those of the TVL was also found for the loudness of frequency-modulated sounds. Thus, the DLM may provide a better prediction for loudness of signals with strong spectral variations over time than the TVL.

The TVL may be advantageous when considering binaural effects since it is able to compute binaural loudness, which may be important when the loudness in non-artificial acoustic situations is investigated. Additionally, the computation of long-term loudness may be beneficial when the loudness of very long signals is assessed. In conclusion, a combination of the crucial stages of the two models might lead to a more accurate and generally applicable loudness model for time-varying sounds.

---

[3] Rennies *et al.* [74] have presented an extension of the DLM that accounts for duration effect in spectral loudness summation. Since the goal of the present study was to compare only well-established models it was decided to use the original version of the DLM rather than the extended version.

from http://web.arch.usyd.edu.au/~sfer9710/PsySound3/index.php on October 2, 2007.

References

[1] J. Rennies, J. L. Verhey: Comparison of loudness models for artificial and environmental sounds. DAGA 2009 - Fortschritte der Akustik, Proceedings of the 34th Annual Meeting of the Deutsche Gesellschaft für Akustik e.V., 2009, 388–391.

[2] DIN 45631: Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum, Verfahren nach E. Zwicker (Procedure for calculating loudness level and loudness). 1991.

[3] ANSI/ASA S3.4-2007: American National Standard Procedure for the Computation of Loudness of Steady Sounds. 2007.

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

[5] E. Zwicker, R. Feldtkeller: Über die Lautstärke von gleichförmigen Geräuschen (Loudness of uniform sounds). Acustica **5** (1955) 303–316.

[6] E. Zwicker, G. Flottorp, S. S. Stevens: Critical band width in loudness summation. J. Acoust. Soc. Am. **29** (1957) 548–557.

[7] B. Scharf: Loudness of complex sounds as a function of the number of components. J. Acoust. Soc. Am. **31** (1959) 783–785.

[8] A. T. Cacace, R. H. Margolis: On the loudness of complex stimuli and its relation to cochlear excitation. J. Acoust. Soc. Am. **78** (1985) 1568–1573.

[9] B. Schneider: The additivity of loudness across critical bands: A conjoint measurement approach. Percept. Psychophys. **43** (1988) 211–222.

[10] J. L. Verhey, M. Uhlemann: Spectral loudness summation for sequences of short noise bursts. J. Acoust. Soc. Am. **123** (2008) 925–934.

[11] H. Fletcher, W. A. Munson: Relation between loudness and masking. J. Acoust. Soc. Am. **9** (1937) 1–10.

[12] E. Zwicker: Über psychologische und methodische Grundlagen der Lautheit (On psychological and methodical principles of loudness). Acustica **8** (1958) 237–258.

[13] E. Zwicker, B. Scharf: A model of loudness summation. Psychological review **72** (1965) 3–26.

[14] E. Zwicker, H. Fastl: Psychoacoustics. 2 ed. Springer, Berlin, 1999.

[15] B. C. J. Moore, B. R. Glasberg: A revision of Zwicker's loudness model. Acta Acustica united with Acustica **82** (1996) 335–345.

[16] B. C. J. Moore, B. R. Glasberg: Modeling binaural loudness. J. Acoust. Soc. Am. **121** (2007) 1604–1612.

[17] B. C. J. Moore, B. R. Glasberg, T. Baer: A model for the prediction of thresholds, loudness and partial loudness. J. Audio Eng. Soc. **45** (1997) 224–239.

[18] M. Florentine, E. Zwicker: A model of loudness summation applied to noise-induced hearing loss. Hear. Res. **1** (1979) 121–132.

[19] B. C. J. Moore, B. R. Glasberg: A model of loudness perception applied to cochlear hearing loss. Auditory Neurosci. **3** (1997) 289–311.

[20] J. Chalupper, H. Fastl: Dynamic loudness model (DLM) for normal and hearing-impaired listeners. Acta Acustica united with Acustica **88** (2002) 378–386.

[21] M. Fruhmann, J. Chalupper, H. Fastl: Zum Einfluss von Innenohrschwerhörigkeit auf die Lautheitssummation (Influence of cochlear hearing impairment on spectral loudness summation). DAGA 2003 - Fortschritte der Akustik, Proceedings of the 29th Annual Meeting of the Deutsche Gesellschaft für Akustik e.V., 2003, 253–254.

[22] W. A. Munson: The growth of auditory sensation. J. Acoust. Soc. Am. **19** (1947) 584–591.

[23] E. Port: Zur Lautstärke und Lautstärkemesseung von pulsierenden Geräuschen (Loudness and loudness measurement of pulsating sounds). Acustica **13** (1963) 224–233.

[24] G. Ekman, B. Berglund, U. Berglund: Loudness as a function of the duration of auditory stimulation. Scand. J. Psychol. **7** (1966) 201–208.

[25] E. Zwicker: Ein Beitrag zur Lautstärkemessung impulshaltiger Schalle (A contribution to the loudness of impulsive sounds). Acustica **17** (1966) 11–22.

[26] T. Poulsen: Loudness of tone pulses in a free field. J. Acoust. Soc. Am. **69** (1981) 1786–1790.

[27] H. Takeshima, Y. Suzuki, S. Kono, T. Sone: Growth of the loudness of a tone burst with a duration up to 10 s. J. Acoust. Soc. Jpn. (E) **9** (1988) 295–300.

[28] M. Florentine, S. Buus, T. Poulsen: Temporal integration of loudness as a function of level. J. Acoust. Soc. Am. **99** (1996) 1633–1644.

[29] S. Buus, M. Florentine, T. Poulsen: Temporal integration of loudnss, loudness discrimination, and the form of the loudness function. J. Acoust. Soc. Am. **101** (1997) 669–680.

[30] M. Epstein, M. Florentine: A test of the Equal-Loudness-Ratio hypothesis using cross-modality matching functions. J. Acoust. Soc. Am. **118** (2005) 907–913.

[31] H. Niese: Die Trägheit der Lautstärke in Abhängigkeit vom Schallpegel (Sluggishness of loudness as a function of level). Hochfrequenz- und Elektroakustik **68** (1959) 143–152.

[32] W. Reichardt: Zur Trägheit der Lautstärkebildung (Sluggishness of loudness perception). Acustica **15** (1965) 345–354.

[33] W. Reichardt: Subjective and objective measurement of loudness level of single and repeated impulses. J. Acoust. Soc. Am. **47** (1970) 1557–1562.

[34] S. S. Stevens: Procedure for calculating loudness: Mark VI. J. Acoust. Soc. Am. **33** (1961) 1577–1585.

[35] J. J. Zwislocki: Temporal summation of loudness: An analysis. J. Acoust. Soc. Am. **46** (1969) 431–441.

[36] M. Kumagai, Y. Suzuki, T. Sone: A study on the time constants for an impulsive sound level meter (A study on loudness of impact sounds). J. Acoust. Soc. Jpn. **5** (1984) 31–36.

[37] Y. Ogura, Y. Suzuki, T. Sone: A temporal integration model for loudness perception of repeated impulsive sounds. J. Acoust. Soc. Jpn. **12** (1991) 1–11.

[38] B. R. Glasberg, B. C. J. Moore: A model of loudness applicable to time-varying sounds. J. Audio Eng. Soc. **50** (2002) 331–341.

[39] O. Pedersen, P. Lyregaard, T. Poulsen: The round robin test on impulsive noise. Rep. No 22, The Acoustics Laboratory, Technical University of Denmark, 1977. 1-180.

[40] H. Bauch: Die Bedeutung der Frequenzgruppe für die Lautheit von Klängen (The role of critical bands for the loudness of sounds). Acustica **6** (1956) 41–45.

[41] C. Zhang, F. Zeng: Loudness of dynamic stimuli in acoustic and electric hearing. J. Acoust. Soc. Am. **102** (1997) 2925–2945.

[42] B. C. J. Moore, S. Launer, D. A. Vickers, T. Baer: Loudness of modulated sounds as a function of modulation rate, modulation depth, modulation waveform and overall level. – In: Psychophysical and Physiological Advances in Hearing. A. S. A. R. Palmer, A. Rees, R. Meddis (eds.). Whurr, London, 1998, 465–472.

[43] B. C. J. Moore, D. A. Vickers, T. Baer: Factors affecting the loudness of modulated sounds. J. Acoust. Soc. Am. **105** (1999) 2757–2772.

[44] G. C. Stecker, E. R. Hafter: An effect of temporal asymmetry on loudness. J. Acoust. Soc. Am. **107** (2000) 3358–3368.

[45] R. Patterson, M. H. Allerhand, C. Giguere: Time-domain modelling of peripheral auditory processing: A modular architecture and software platform. J. Acoust. Soc. Am. **98** (1995) 1890–1894.

[46] E. Zwicker: Der Einfluss der zeitlichen Struktur von Tönen auf die Addition von Teillautheiten (Influence of the temporal structure of tones on the summation of partial loudnesses). Acustica **21** (1969) 16–25.

[47] E. Zwicker: Loudness and excitation patterns of strongly frequency modulated tones. – In: Sensation and measurement, papers in honor of S. S. Stevens. B. S. H. R. Moskowitz, J. Stevens (eds.). D. Reidel, Dordrecht/Boston, 1974, 325–335.

[48] E. Zwicker: Procedure of calculating loudness of temporally variable sounds. J. Acoust. Soc. Am. **62** (1977) 675–682.

[49] E. Zwicker: Subdivision of the audible frequency range into critical bands (Frequenzgruppen). J. Acoust. Soc. Am. **33** (1961) 248.

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

[51] B. C. J. Moore, B. R. Glasberg, C. J. Plack, A. K. Biswas: The shape of the ears's temporal window. J. Acoust. Soc. Am. **83** (1988) 1102–1116.

[52] C. J. Plack, B. C. J. Moore: Temporal window shape as a function of frequency and level. J. Acoust. Soc. Am. **87** (1990) 2178–2187.

[53] E. Zwicker: Dependence of post-masking on masker duration and its relation to temporal effects in loudness. J. Acoust. Soc. Am. **57** (1984) 219–223.

[54] R. Widmann, R. Lippold, H. Fastl: Ein Computerprogramm zur Simulation der Nachverdeckung für Anwendung in akustischen Messsystemen (A computer program for simulations of forward masking for application in acoustic measurement systems). DAGA 1998 - Fortschritte der Akustik, Proceedings of the 24th Annual Meeting of the Deutsche Gesellschaft für Akustik e.V., 1998, 96–97.

[55] J. Chalupper: Simulation von Mitörschwellen Normal- und Schwerhörender mit einem dynamischen Lautheitsmodell (Simulation of masked thresholds for normal and hearing-impaired listeners using a dynamic loudness model). DAGA 2002 - Fortschritte der Akustik, Proceedings of the 28th Annual Meeting of the Deutsche Gesellschaft für Akustik e.V., 2002, 490–491.

[56] B. C. J. Moore, B. R. Glasberg: Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J. Acoust. Soc. Am. **74** (1983) 750–753.

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

[58] B. R. Glasberg, B. C. J. Moore: Derivation of the auditory filter shapes from notched-noise data. Hear. Res. **47** (1990) 103–138.

[59] J. E. Appell: Loudness models for rehabilitative audiology. Ph.D. thesis, Universität Oldenburg, 2002.

[60] B. Scharf: Loudness. – In: Handbook of Perception, Volume IV. Hearing. E. C. Carterette, M. P. Friedman (eds.). Academic Press, New York, 1978, 187–242.

[61] G. Grimm, V. Hohmann, J. L. Verhey: Loudness of fluctuating sounds. Acta Acustica united with Acustica **88** (2002) 359–368.

[62] M. F. B. van Beurden, W. A. Dreschler: Bandwidth dependency of loudness in series of short noise bursts. Acta Acustica united with Acustica **91** (2005) 1020–1024.

[63] H. Fastl: How loud is a passing vehicle? Proc. Inter-Noise '87, 1987, vol. II, 993–996.

[64] R. Höger, E. Matthies, E. Letzing: Physikalische versus psychologische Reizintegration: Der Mittelungspegel aus wahrnehmungspsychologischer Sicht (Physical versus psychological integration of stimuli: average level from the point of view of psychology of perception). Zeitschrift für Lärmbekämpfung **35** (1988) 163–167.

[65] G. Canévet, B. Scharf: The loudness of sounds that increase and decrease continuously in level. J. Acoust. Soc. Am. **88** (1990) 2136–2142.

[66] R. S. Schlauch: A cognitive influence on the loudness of tones that change continuously in level. J. Acoust. Soc. Am. **92** (1992) 758–765.

[67] P. Susini, S. McAdams, B. K. Bright: Global and continuous loudness estimation of time-varying levels. Acta Acustica united with Acustica **88** (2002) 536–548.

[68] P. Susini, S. McAdams, B. K. Bright: Loudness asymmetries for tones with increasing and decreasing levels using continuous and global ratings. Acta Acustica united with Acustica **93** (2007) 623–631.

[69] B. Pedersen, W. Ellermeier: Temporal weights in level discrimination of time-varying sounds. J. Acoust. Soc. Am. **123** (2008) 963–972.

[70] D. Oberfeld: Does a rhythmic context have an effect on perceptual weights in auditory intensity processing? Can. J. Exp. Psychol. **62** (2008) 24–32.

[71] J. L. Verhey, B. Kollmeier: Spectral loudness summation as a function of duration. J. Acoust. Soc. Am. **111** (2002) 1349–1358.

[72] A. K. Anweiler, J. L. Verhey: Spectral loudness summation for short and long signals as a function of level. J. Acoust. Soc. Am. **119** (2006) 2919–2928.

[73] J. L. Verhey, M. Uhlemann: Spektrale Lautheitssummation von pulsierenden Geräuschen (Spectral loudness summation of pulsating sounds). DAGA 2007 - Fortschritte der Akustik, Proceedings of the 33rd Annual Meeting of the Deutsche Gesellschaft für Akustik e.V., 2070, 847–848.

[74] J. Rennies, J. L. Verhey, H. Fastl: Modeling temporal effects of spectral loudness summation. Acta Acustica united with Acustica **95** (2009) 1112–1122.

[75] H. Gockel, B. C. J. Moore, R. Patterson: Influence of component phase on the loudness of complex tones. Acta Acustica united with Acustica **88** (2002) 369–377.