ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 95 (2009) 1112 – 1122

DOI 10.3813/AAA.918243

Modeling Temporal Effects of Spectral Loudness Summation

Jan Rennies*, Jesko L. Verhey
AG Neurosensorik, Institut für Physik, Carl von Ossietzky Universität Oldenburg, 26111 Oldenburg, Germany.
jesko.verhey@uni-oldenburg.de

Josef Chalupper**, Hugo Fastl
AG Technische Akustik, Lehrstuhl für Mensch-Maschine-Kommunikation, Technische Universität München, 80333 München, Germany

Summary

Recent studies have shown that spectral loudness summation depends on duration. Modifications of a current loudness model were investigated with respect to their ability to predict this effect. The original version of the model could not simulate the duration dependence of spectral loudness summation. To reconcile the model with the loudness data, three different mechanisms accentuating temporal onsets of sounds were tested: (i) adaptive compression, (ii) adaptive auditory filters and (iii) bandwidth-dependent integration. A comparison between simulations and experimental data indicated that, in principle, all mechanisms lead to increased spectral loudness summation for short noise bursts, but bandwidth-dependent integration may be the most realistic approach. Such a modified model also predicts the spectral loudness summation of repeated noise bursts as a function of repetition rate.

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

1. Introduction

We are constantly exposed to various kinds of sound differing in level, duration and spectral content. Especially high-level, impulsive sounds contribute to auditory sensations such as annoyance or loudness, even though their effect on equivalent continuous sound pressure level is small (e.g. [2]). In order to correctly assess these effects on auditory perception, it is therefore desirable to be able to predict the loudness of these sounds. The present study addresses the problem of current models to predict the higher spectral loudness summation for short single and repeated noise bursts compared to stationary sounds. Several mechanisms were hypothesized to account for this duration effect in spectral loudness summation. The validity of these hypotheses was tested by modifying an existing model incorporating these mechanisms and comparing the predictions of the modified model with experimental data on temporal effects in spectral loudness summation.

Several studies indicate that the loudness of sounds with constant overall intensity increases if the bandwidth is widened beyond a certain critical bandwidth (e.g. [3, 4, 5]). This effect known as spectral loudness summation has been successfully implemented in loudness models for stationary sounds (e.g. [6, 7, 8, 9, 10]). The common assumption of these models is that spectral loudness summation is the result of an analysis of the incoming sound with a bank of overlapping auditory filters, followed by a compressive loudness transformation of the output of each filter, and a subsequent summation across filters. For simplicity, consider a sound consisting of m spectral components of intensity I. If the entire intensity is confined to a single auditory filter, the resulting loudness, N_{narrow}, is smaller than if the components are distributed over different auditory filters due to the compression with an exponent $\alpha < 1$:

$$N_{\text{broad}} = \sum_{n=1}^{m} I^{\alpha} > \left(\sum_{n=1}^{m} I \right)^{\alpha} = N_{\text{narrow}}, \quad \alpha < 1. \tag{1}$$

Both the value of α and the width of the auditory filters influence the amount of spectral loudness summation.

Several studies have shown that spectral loudness summation depends on duration [11, 12, 13, 14]. Even though the amount of spectral loudness summation varied between the individual studies, all showed that the level difference between narrowband noise and equally loud broadband noise was larger for 10-ms bursts than for 1000-ms bursts by amounts from 4 dB [13] to 8 dB [14].

To describe temporal aspects of loudness perception, different loudness models have been proposed (e.g. [15,

Received 8 June 2009,
accepted 31 July 2009.

* now at the Fraunhofer Institute for Digital Media Technology IDMT, 26129 Oldenburg, Germany

** now at the Siemens Audiological Engineering Group, 91058 Erlangen, Germany

16, 17, 18, 19]). Zwicker [15] showed that temporal integration of loudness for single tone bursts could be predicted when a low-pass circuit with a time constant of 100 ms was used. Ogura *et al.* [17] extended this approach using the same time constant to describe the attack of loudness perception, and another, much longer time constant of 5 s to describe the decay. Using two time constants they could explain an increase in loudness for a series of repeated noise bursts when the repetition rate was increased. Glasberg and Moore [18] published a loudness model with a multiple temporal-integration stage. By computing what they called short-term and long-term loudness, they were able to predict temporal integration of tone bursts and the loudness of amplitude-modulated signals. The dynamic properties of these models are expressed in a more or less sophisticated temporal integration stage. Commonly, this is the last stage of the models, i.e. effects such as spectral masking or spectral loudness summation are accounted for prior to temporal integration. Verhey and Kollmeier [11] argued that such a model structure is a realization of the equal-loudness ratio hypothesis (ELRH, e.g. [20, 21, 22]). The ELRH assumes that the loudness ratio between long and short signals with the same spectrum is only determined by the two durations, and is independent of level and spectrum. Therefore, according to the ELRH, the loudness versus intensity curves of short and long sounds are vertical transposes of one another (provided they have the same spectrum and loudness is expressed in sones on a logarithmic scale). The ELRH predicts the same magnitudes of spectral loudness summation for short and long signals when compared at the same reference level (see right panel of Figure 2 in [13])[1]. This prediction is not consistent with the experimental data on duration effects in spectral loudness summation described in the literature [11, 12, 13, 23, 14].

In contrast to the loudness models mentioned above, the model of Zwicker [16] and the dynamic loudness model (DLM) of Chalupper and Fastl [19] contain an additional dynamic stage before loudness is summed across frequency and integrated over time by a low-pass filter. This additional stage takes temporal masking into account by appending post-masking tails to loudness peaks to describe temporal persistence of loudness in each auditory channel. Thus, loudness is not switched off instantaneously at stimulus offsets. However, Verhey and Kollmeier [11] argued that these models also fail to predict duration-dependent spectral loudness summation, since only the decay part is modeled, which does not affect spectral loudness summation of short noise bursts.

[1] The prediction may be different for a comparison at the same reference loudness, especially in regions where the slope of the loudness functions changes with level (see also [11]). It should be noted that the ELRH does not predict that, for very long signals, salient events may receive higher weights in overall loudness, nor does the ELRH predict recency effects. This is also true for all models used in the present study since they use the maximum of a low-pass filtered version of the instantaneous loudness as a measure of the overall loudness. However, these effects are usually observed on a time scale of several seconds or minutes, i.e. well beyond the signal durations considered in the present study.

The aim of the present study was to investigate the simulation of the dynamics of spectral loudness summation. From the loudness models mentioned above, the DLM may be preferable, since it is the only one containing a dynamic stage prior to spectral summation. This entanglement of temporal and spectral processing may be advantageous when describing the dynamic properties of spectral loudness summation of single and repeated noise bursts.

In the first step, therefore, the original version of the DLM was used and its ability to predict spectral loudness summation for short and long signals was investigated. Then, modifications to the model were made by implementing the mechanisms of adaptive compression, adaptive auditory filters and bandwidth-dependent integration. The ability of these approaches to account for the duration dependence of spectral loudness summation was compared to the data of Verhey and Uhlemann [14] for single noise bursts. Subsequently, the newly introduced dynamics of the modified model were used to predict the loudness ratio between short and long signals for narrowband and broadband noise, as presented in [13]. It was concluded that the most promising approach is the one assuming a bandwidth-dependent integration. This approach was then used to predict the level dependence of spectral loudness summation described by Anweiler and Verhey [13], the spectral loudness summation of repeated noise bursts described by Verhey and Uhlemann [14] and spectral loudness summation for an intermediate duration described by Verhey and Kollmeier [11].

2. General model structure

The structure of the DLM is schematically shown in Figure 1. For a detailed description of the model, the reader is referred to [19]. Briefly, in the first stage, the lower limit of the audible frequency range is accounted for by a high-pass filter. In the following stage, a bank of overlapping critical-band filters is applied[2]. The temporal envelopes of the filtered signals are extracted at a rate of 500 Hz using a temporal window. A subsequent correction for the transmission through the middle ear results in the time-dependent quantity excitation $E(z,t)$. In order to simplify implementation and to be compliant to the DIN 45631 [24] standard, this transmission is modeled after the filter bank. In each filter corresponding to the critical-band rate z in Bark, $E(z,t)$ is transformed into specific loudness $N'(z,t)$ by

$$N'(z,t) = N_0 \cdot \left(\frac{E_{THQN}(z)}{s(z)E_0}\right)^{\alpha} \cdot \left[\left(1 - s(z) + s(z)\frac{E(z,t)}{E_{THQN}(z)}\right)^{\alpha} - 1\right], \quad (2)$$

where $E_{THQN}(z)$ is the excitation at threshold in quiet and E_0 is the excitation corresponding to the reference sound

[2] In the original implementation of the model, the critical-band filters do not overlap perfectly at their -3-dB points. The same implementation was used throughout the present study.

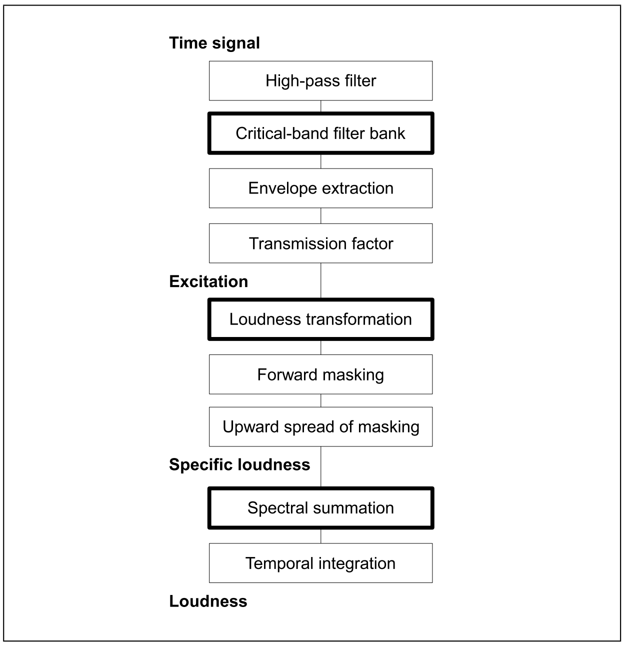

Figure 1. Schematic structure of the dynamic loudness model (DLM) of Chalupper and Fastl [19]. Bold edges mark the three stages which were modified in the present study.

intensity of 10^{-12} W/m^2. N_0 is a constant used to calibrate the model. The quantity $s(z)$ is the frequency-dependent threshold factor and accounts for differences in the absolute threshold across frequency caused by internal noise [19, 25]. The compressive exponent α has a value of 0.23. This value is in the range of compressive exponents found in the literature (see section 4.1) and, among other things, leads to the well-established doubling of loudness for a 10-dB increase for narrowband signals [25].

In the following dynamic stage, the influence of forward masking is included by appending temporal tails to loudness peaks. Effects of backward masking are neglected. Subsequently, nonlinear spectral masking effects are accounted for by modeling the upward spread of excitation. A summation of loudness across frequency is realized by computing the area under the specific-loudness-critical-band-rate curve for each time sample. In the last stage, the resulting time-dependent quantity (instantaneous loudness) is low-pass filtered in order to account for temporal integration of loudness. At the output of the model, loudness in sones is available as a function of time. In this form, the loudness model can easily be used to predict loudness for hearing-impaired listeners by adapting $E(z)$ and $E_{THQN}(z)$ to individual hearing losses (see [19]). However, in the present study, this ability is not exploited since the focus is on the loudness perception of normal-hearing listeners.

3. Modifications of the model

In the literature, three different mechanisms were proposed, which are described in detail below. They are based on modifications at different stages of the model, which are marked by bold edges in Figure 1. The compressive behavior is changed in the transformation from excitation to loudness (equation 2), adaptive filters are introduced in the auditory filter stage, and in the third approach, a bandwidth-dependent integration is applied at the level of spectral summation. Since the modifications are aimed at predicting the increased spectral loudness summation for short signals compared to long signals, each mechanism is implemented to give special weight to onsets of the input sound. For long signals, this effect is outweighed by the final temporal integration stage since larger loudness values are reached for later time samples (see section 5.5.4). For very short signals, however, temporal integration does not reach a stationary value and an increased loudness at stimulus onset determines overall loudness. Thus, for stationary sounds, the modified versions of the DLM predict essentially the same loudness as the original version, while the predictions differ for very short or quickly fluctuating sounds.

In the present study, the maximum of the loudness as a function of time was taken as an estimate of the loudness of a stimulus for all simulations, following a suggestion by Zwicker [16]. A similar approach was chosen by Glasberg and Moore [18] for stimuli with durations comparable to the ones considered in the present study. The sampling rate for the calculation of loudness was increased to 1000 Hz to achieve sufficiently accurate sampling for very short stimuli.

3.1. Adaptive compression

Verhey and Kollmeier [11] suggested that adaptive compression might be responsible for the duration dependence of spectral loudness summation. If larger compression (i.e. a smaller value of α) is assumed at stimulus onset, then more spectral loudness summation is expected for short than for long stimuli. In order to estimate the increase in spectral loudness summation produced by modifying the compression, simulations were carried out with the original exponent for 1000-ms long signals and with more compression for 10-ms long noise bursts. A transition between the two parameters could be used to implement this approach in a single model, but for reasons outlined below, this was not made in the present study.

3.2. Adaptive auditory filters

Verhey and Uhlemann proposed adaptive auditory filters as an alternative underlying mechanism [26, 14]. They suggested that sharper auditory filters at stimulus beginning, which broaden over time, might explain the larger spectral loudness summation for short signals, since the sound energy is distributed over more auditory filters prior to compression and summation across frequency.

The loudness model of Chalupper and Fastl [19] uses a critical-band filter bank of 24 overlapping auditory filters to analyze incoming sound signals, i.e. the center frequencies and bandwidths of the filters are implemented as suggested by Zwicker [27]. For the model version with adaptive auditory filters, the width of the auditory filters was decreased for short signals while keeping the point of overlap constant, i.e. a larger number of filters was used. In this way, the same frequency range was covered with more filters. For a given bandwidth of the input sound, the energy was thus distributed over more filters, which increased the amount of spectral loudness summation. Verhey and Uhlemann [26] used a simplified loudness model with a gammatone filter bank [28] where the width and steepness of the filters were fitted to their data on spectral loudness summation using 1000-ms long signals. Another fitting process with the same stimuli was performed to account for the data of short noise bursts, for which narrower filters were necessary. The present study tested if this approach could also be used in the DLM. In order to estimate the spectral loudness summation produced by modifying the auditory filter width, simulations were carried out with the original filter bank used by Chalupper and Fastl [19] for 1000-ms long signals and with a larger number of narrower filters for 10-ms long noise bursts. Again, a transition between the two parameter sets could be used to implement this approach in a single loudness model, but this was not made in the present study for reasons given below.

There is some controversy in the literature as to the temporal development of the auditory filters. To the knowledge of the authors, the study by Strickland [29] is the only one that found narrower auditory filters at stimulus onset, which broadened over time, i.e. that found a temporal development of the filters, which might explain an increased spectral loudness summation for short signals. Strickland [29] measured filters that were about half as wide at stimulus onsets than at later times. Therefore, in the present study, the filter width was reduced by a factor of two for the 10-ms long stimuli to investigate the possible increase in predicted spectral loudness summation.

3.3. Bandwidth-dependent integration

A third alternative, proposed by Fruhmann *et al.* [12] and Verhey and Uhlemann [14], was motivated by a signal-detection experiment performed by van den Brink and Houtgast [30][3]. Van den Brink and Houtgast [30] measured masked thresholds of noise stimuli with a maximum bandwidth of three octaves with a center frequency of 1600 Hz, and a maximum duration of 100 ms. Systematically varying the bandwidth at a fixed duration of 100 ms, they found increasing thresholds when the bandwidth was increased beyond about 1/3 oct. Analogously, for a fixed bandwidth of 3 octaves and systematically varied durations of the stimulus, they found increasing thresholds when the duration exceeded a value of about 30 ms. They argued that efficient integration of information was only possible for a restricted time-frequency region. Fruhmann *et al.* [12] and Verhey and Uhlemann [14] suggested that this principle of efficient integration might explain the duration dependence of spectral loudness summation: if summation across frequency is enhanced at stimulus onset, then more spectral loudness summation is expected for short stimuli. To test this hypothesis, a modified DLM with an amplification at stimulus onset was used. The amount of amplification was determined by the bandwidth and the time course of the stimulus. On the one hand, the amplification increased with bandwidth, i.e. the larger the bandwidth of the stimulus, the higher the amplification. On the other hand, the amplification was triggered only at stimulus onsets, i.e. the amplification was only applied for a short time. The amplification at stimulus onset results in an instantaneous loudness at onset which is higher than the steady state response. Such an accentuation of the stimulus onset is a common property of the auditory system since most neurons along the auditory pathway respond briskly to transients [31]. An accentuation of stimulus onset is also found in recent studies on temporal weights in loudness perception. Pedersen and Ellermeier [32] found a strong primacy effect, i.e. a significantly higher weight to the first few milliseconds, and Verhey and Rennies [33] showed that this effect is bandwidth dependent. In this way, for stationary broadband sounds, the amplification only caused a faster build-up of loudness, but the steady-state loudness was the same as without amplification, i.e. the same as for the original model. For short broadband stimuli not reaching the steady state of temporal integration, however, the amplification caused an increased overall loudness.

This principle required a detection of both bandwidth and temporal onsets of the exciting stimulus. Since the model was based on different auditory filters, the number of "active" filters, n, could serve as an indicator for the bandwidth of the stimulus. Whether or not a filter was active was determined by means of a relative threshold: when the specific loudness in a channel exceeded one third of the maximum specific loudness in any channel, it was considered active, otherwise it was inactive. This approach was based on the assumption that the frequency region excited the most dominates loudness perception. The bandwidth-dependent part of the amplification was thus $C \cdot n$, where the value of the relative threshold and $C = 0.29$ were set to obtain an increased predicted spectral loudness summation for short noise bursts as observed in the literature.

The detection of an onset at a given time sample was based on a causal approach: the excitation at the sample was compared to the mean excitation over the preceding 5 ms (including the current sample). If the actual excitation was larger than the average excitation over the preceding 5 ms by a certain ratio (threshold), then an onset was detected. This onset detector was applied to each auditory channel. As an effective implementation in the model, a binary vector was defined containing ones for time samples at which an onset was detected in any channel, and zeros otherwise. This quantity was subsequently smoothed by a low-pass filter, resulting in a vector with values between zero and one. This low-pass filtered binary variable represented a time factor. The overall ampli-

[3] Data from detection experiments are commonly used to incorporate spectral and temporal properties of the auditory system into loudness models. For example, excitation patterns (as an interim stage of loudness models) are usually derived from psychoacoustical masking patterns. Forward masking data were introduced by Zwicker [16] to account for the temporal decay of loudness.

fication was the product between this time factor and the bandwidth-dependent amplification. The parameters introduced by this implementation were the threshold for onset detection (6 dB), and the attack (10 ms) and release (5 ms) time constants of the low-pass filter. They were adjusted to fit recent data on spectral loudness summation of single and repeated noise bursts [14].

4. Results of the simulations

In the following, model predictions are always presented as an average of ten simulation runs, using a different realization of noise for each run.

4.1. Spectral loudness summation for single noise bursts

Verhey and Uhlemann [14] measured the level difference at equal loudness between a test noise band and a 400-Hz wide reference noise band with the same center frequency of 2000 Hz as a function of the test noise bandwidth. The reference level was 70 dB SPL. The same stimulus parameters were used in the simulations of the present study. Figure 2 compares their data (symbols) to the predictions of the original model and of the three versions of the modified model described above (curves). In agreement with the data, the predicted level difference ΔL between a test noise band and the equally loud 400-Hz wide reference noise decreases as the test bandwidth increases, i.e. spectral loudness summation is predicted by the DLM in its original form. However, two major discrepancies between data and simulations can be observed. Firstly, the DLM generally underestimates the observed spectral loudness summation. The maximum predicted level difference between reference and test signals is about 13 dB, while it amounts to 19 dB in the experimental data for long stimuli. This difference is larger than the interindividual standard errors which were always below 3 dB (see Figure 3 in [14]). A better fit of the model predictions to the experimental data on spectral loudness summation for long signals is achieved when the exponent α is set to 0.12 instead of 0.23 (not shown). Inter-individual variations in compression are often observed in studies on loudness growth functions for tones [34, 22, 35, 13]. Canévet *et al.* [36] observed such variability also at group level when measuring loudness functions in different groups of listeners. Thus, it is not unreasonable to assume that for the limited number of listeners (12) participating in the experiments of Verhey and Uhlemann [14] a higher than average compression is observed. However, since the general goal of loudness model for normal hearing listeners is to fit a wide range of data from different studies, it was decided not to adjust the compression for stationary signals to reflect the compression of the group of listeners of Verhey and Uhlemann [14]. Apart form the underestimation of the spectral loudness summation for long signals, the DLM predicts a slightly smaller level difference for short signals and thus, it fails at simulating the 6-dB larger spectral loudness summation observed experimentally for short signals by Verhey and Uhlemann [14].

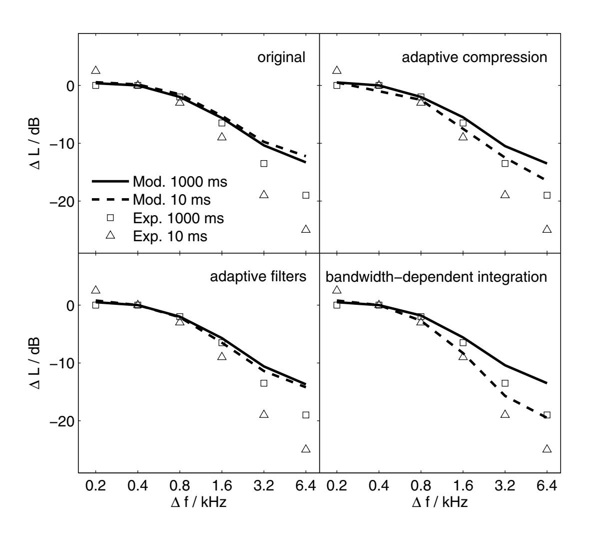

Figure 2. Level difference between a test noise band and the equally loud 400-Hz wide reference noise as a function of test bandwidth for 10- and 1000-ms long bursts, as measured by Verhey and Uhlemann [14]. Squares and triangles show data for long and short signals, respectively. Solid and dashed lines show the corresponding predictions, which are shown for the original model and the modified versions including adaptive compression, adaptive filters and bandwidth-dependent integration.

When the compressive exponent is changed to 0.1 for short signals, while it is kept the same for long stimuli (i.e. $\alpha = 0.23$), more spectral loudness summation for short signals is predicted (Figure 2, top right panel). The level difference simulated by the modified DLM is about 3 dB larger for short stimuli, which is half the difference found experimentally. This change in compression is already larger than by a factor of two, which is the change in bandwidth for the second modification of the model. Therefore, higher compressions were not considered.

The bottom left panel of Figure 2 compares data and simulations obtained when narrower auditory filters are used for short stimuli. As for the approach with an increased compression, the modified model predicts slightly more spectral loudness summation for short than for long signals. However, the difference is very small and cannot explain the measured effect. The predicted level differences were larger when 1000-ms long signals were used in combination with the narrower filters (not shown). This indicates that spectral broadening has a detrimental effect on the gain in spectral loudness summation due to narrower filters. Since a further reduction of the filter width did not substantially increase spectral loudness summation for short signals and, in addition, a higher reduction of filter width seems unrealistic (cf. [29]), no further adjustments for the parameters of the model version with time-varying filters were made.

The bottom right panel of Figure 2 shows that when bandwidth-dependent integration is introduced in the DLM, simulated level differences are in qualitative agree-

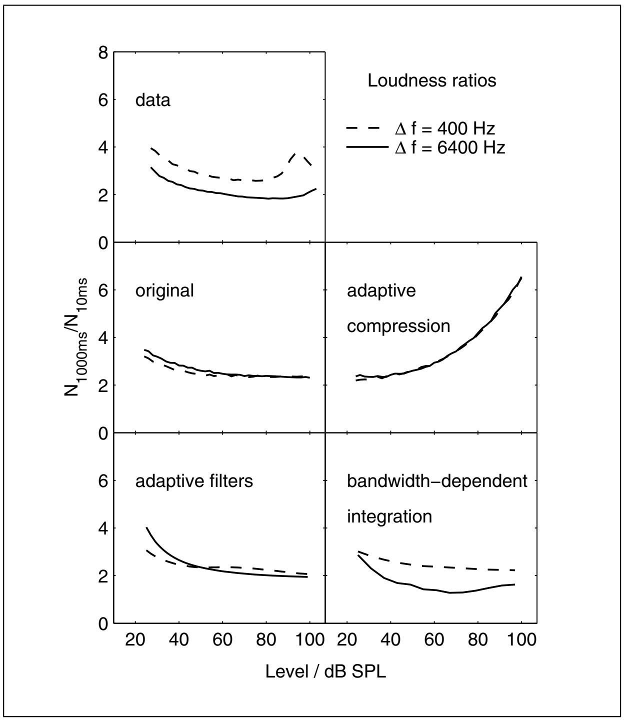

Figure 3. Loudness ratios between noise bursts with durations of 1000 and 10 ms as a function of level for bandpass noises with bandwidths of 6400 (solid) and 400 Hz (dashed). The upper panel shows data from Anweiler and Verhey [13]. The remaining panels show predictions of the original model and its modified versions.

ment with the data for single noise bursts presented by Verhey and Uhlemann [14]. For the long signals, the simulated level differences are up to 6 dB smaller than for short signals, as observed experimentally. In conclusion, the DLM predicts less spectral loudness summation for long signals than measured by Verhey and Uhlemann [14]. All of the implemented mechanisms can in principle predict a duration dependence of spectral loudness summation. However, only the bandwidth-dependent integration seems to predict significantly more spectral loudness summation for short than for long signals.

4.2. Loudness ratio between long and short signals

To further investigate the effects of the modifications, loudness ratios between long and short signals were computed as a function of level. Bandpass noise bursts with durations of 10 and 1000-ms, and bandwidths of 400 and 6400 Hz were used for the simulations (the center frequency was 2000 Hz). The ratios were computed by dividing the loudness functions for long signals by those for short signals. The middle and lower panels of Figure 3 show the predicted loudness ratios for the original version of the DLM and the three modified versions. For comparison, data from Anweiler and Verhey [13], obtained by categorical loudness scaling, are shown in the top panel.

The original version of the model predicts that loudness ratios slightly decrease with level, approaching a constant value for high levels. This is a good first approximation to the experimental data. However, the data indicate a different ratio for the two bandwidths: the ratio found by Anweiler and Verhey [13] is larger for the smaller bandwidth. For both bandwidths, a slight decrease with level can be observed. The maximum in the ratio function of the narrowband stimuli at about 95 dB SPL is probably an artifact of their fitting procedure.

When more compression is used for short signals, the slope of the loudness function is shallower than for long signals. Thus, the loudness ratio increases with level (mid-right panel of Figure 3). Again, the loudness ratio is independent of bandwidth. Neither of theses trends is observed in the experimental data. The approach of adaptive auditory filters yields a slightly larger loudness ratio for the narrow bandwidth than for the broad bandwidth at medium to high levels. Between 50 and 80 dB SPL, the loudness ratios are about 2.3 and 2.1 for the 400 and 6400 Hz bandwidths, respectively. For lower levels, the ratio for the broad bandwidth becomes larger than for the narrow bandwidth. The loudness ratios obtained using bandwidth-dependent integration agree reasonably well with the data of Anweiler and Verhey [13]. That is, the ratio between long and short signals slightly decreases with level and depends on bandwidth, being larger for the narrower bandwidth.

4.3. Spectral loudness summation as a function of level

The simulations of spectral loudness summation for single noise bursts show that adaptive compression results in a loudness ratio between long and short signals which increases with level. This is at odds with experimental data (see top panel of Figure 3). Additionally, the effect of dynamic auditory filters was very small and possibly counteracted to a large extent by spectral broadening for short signals. Therefore, these mechanisms were discarded in the rest of this paper and other effects of spectral loudness summation were investigated only using the approach of a bandwidth-dependent integration.

Several studies have investigated the dependence of spectral loudness summation on level. Wagner *et al.* [37] and Anweiler and Verhey [13] showed that the level difference $\Delta L = L_{400Hz} - L_{6400Hz}$ between equally loud narrowband and broadband signals was largest at medium reference levels and decreased toward lower or higher levels, where the level difference approached zero. Similar results were reported by Zwicker *et al.* [5] for tone complexes with different spacing between the individual components. The mean level differences between a tone complex with widely spaced components and a tonal reference were largest at medium levels, smaller at a high reference level, and close to zero at low levels.

Figure 4 shows the data of Anweiler and Verhey [13] (top panel) together with the predictions of the original (bottom left panel) and the modified version of the DLM using a bandwidth-dependent gain (bottom right panel). For long signals, both model versions show a maximum spectral loudness summation at medium levels which is

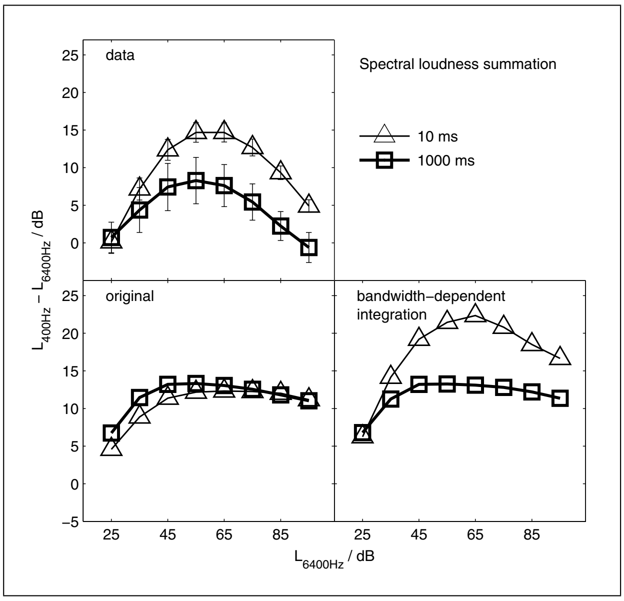

Figure 4. Level difference between equally loud 400- and 6400-Hz wide noise bands as a function of the level of the broadband reference stimulus. Data from Anweiler and Verhey [13] are shown in the top panel for bursts with durations of 10 (triangles) and 1000 ms (squares). Error bars indicate standard errors across 12 listeners. The corresponding predictions are shown for the original version of the DLM (bottom left panel) and for the modified version, using bandwidth-dependent integration (bottom right panel).

also found in the data. The predicted maximum of spectral loudness summation is 13 dB. This is slightly larger than observed by Anweiler and Verhey [13] (8 dB) but is in reasonable agreement with a maximum of 10 to 13 dB found in previous studies on the level dependence of spectral loudness summation [21, 37].

The DLM in its original form predicts no increased spectral loudness summation for short signals, i.e. the predicted curve for short signals never exceeds that for long signals (bottom left panel of Figure 4). For low levels, the original version of the model predicts a small level difference, which increases with level and reaches a maximum at about 45 dB SPL. A moderate decrease can be observed for higher levels. The difference between the maximum ΔL and the ΔL at a reference level of 95 dB SPL is about 2 dB. The shapes of the curves for short and long signals are similar, even though the maximum for 10-ms long signals is at 55 dB SPL. The modified loudness model also shows an increasing level difference from low to moderate levels. For reference levels larger than about 35 dB SPL, the predicted level difference is larger for short than for long signals. This is similar to the experimental data of Anweiler and Verhey [13] (see top panel). As for the original version of the DLM, a maximum is reached at a reference level of 45 dB SPL for 1000-ms long signals. For 10-ms long signals, the maximum occurs at a larger reference level, which agrees with the experimental data. For higher levels, the modified DLM predicts a decreasing spectral loudness summation. At a reference level of 95 dB

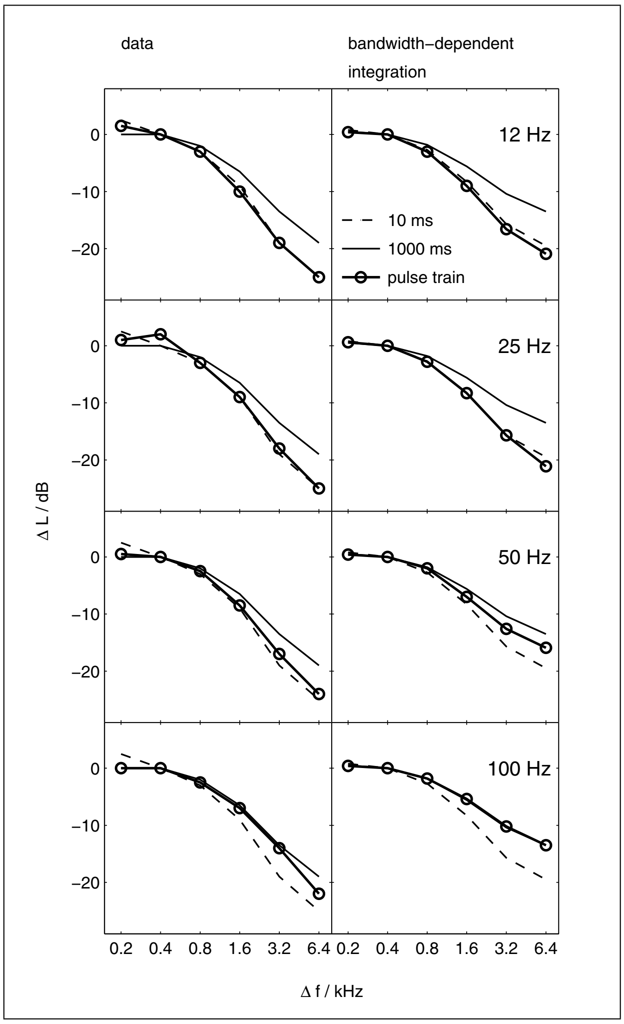

Figure 5. Level difference between a test stimulus and the equally loud 400-Hz wide reference stimulus at a center frequency of 2000 Hz as a function of test bandwidth for repeated noise bursts (circles). The left column shows data from Verhey and Uhlemann [14], the right column shows the corresponding simulations obtained using bandwidth-dependent integration. Repetition rates are indicated in the right panels. For comparison, solid and dashed lines show data (left) and simulations (right) for long and short single bursts, respectively.

SPL, the level difference is still 11 dB for long signals and about 17 dB for short signals. This is not observed in the experimental data, in which the level difference is about 5 and 0 dB for short and long signals, respectively, at this reference level. Thus, at high levels spectral loudness summation is overestimated by the original and the modified versions of the DLM.

4.4. Spectral loudness summation for repeated noise bursts

The left column of Figure 5 shows data of Verhey and Uhlemann [14] for repeated noise bursts (circles). The level differences between equally loud test stimuli and the 400-Hz wide reference are shown as a function of test bandwidth. The center frequency of the noise bursts was 2000 Hz. The stimuli consisted of repeated 10-ms long

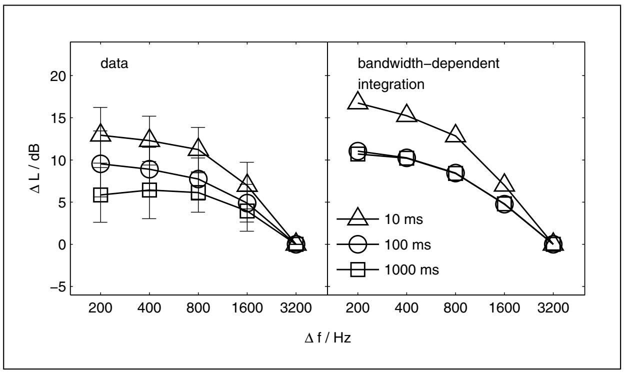

Figure 6. Level difference between equally loud noise bands and the 3200-Hz wide reference as a function of bandwidth. Data from Verhey and Kollmeier [11] for burst durations of 10 (triangles), 100 (circles) and 1000 ms (squares) are shown in the left panel. The corresponding predictions obtained with the modified DLM using bandwidth-dependent integration are shown in the right panel.

noise bursts (including 2.5 ms $\cos^2$ ramps) with an inter-burst interval equal to the inverse of the repetition rate reduced by the duration of a single burst. The number of bursts per stimulus was equal to the repetition rate. The repetition rate is indicated in the right panels of Figure 5. The right column shows the predictions of the modified DLM using bandwidth-dependent integration. For comparison, thin solid and dashed lines represent data and simulations for single noise bursts of 1000 and 10 ms duration, respectively. The experimental data indicate that up to a repetition rate of 25 Hz, spectral loudness summation of pulse trains is the same as for single bursts of 10-ms duration. For a repetition rate of 100 Hz, spectral loudness summation is similar to that for a single, 1000-ms long burst. A transition begins to occur at a repetition rate of about 50 Hz. This dependence of spectral loudness summation on repetition rate is also predicted by the modified model.

4.5. Spectral loudness summation for an intermediate duration

Verhey and Kollmeier [11] measured the level difference at equal loudness between narrowband and broadband sounds with a center frequency of 2000 Hz for burst durations of 10, 100 and 1000 ms. The left panel of Figure 6 shows their data for reference noise bursts with a bandwidth of 3200 Hz and a level of 65 dB SPL. In agreement with Verhey and Uhlemann [14], they found more spectral loudness summation for 10-ms bursts than for 1000-ms bursts. Furthermore, Verhey and Kollmeier [11] found up to about 3 dB more spectral loudness summation for 100-ms bursts than for 1000-ms bursts.

The corresponding predictions of the modified DLM using bandwidth-dependent integration are shown in the right panel. For the long signals the amount of spectral loudness summation in Verhey and Kollmeier [11] is smaller than the predicted effect, indicating a lower than average compression for the listeners in their study. A good fit between the measured and predicted amount of spectral loudness summation for the 1000-ms signals can be achieved by setting the exponent to 0.3 (not shown). As before, the modified model predicts duration-dependent spectral loudness summation: the maximum level difference is about 6 dB larger for 10-ms long bursts than for 1000-ms long bursts. This agrees quantitatively with the data by Verhey and Uhlemann [14], but is about 1 dB less than measured by Verhey and Kollmeier [11], as shown in the left panel of Figure 6. Moreover, the predicted level difference between noise bursts with bandwidths of 3200 and 200 Hz and with a duration of 100 ms is only about 0.4 dB larger than that for 1000-ms long signals.

5. Discussion

5.1. Comparison with the Equal-Loudness-Ratio Hypothesis

The simulations show that the original version of the DLM cannot simulate the duration dependence of spectral loudness summation. In contrast to the experimental data, the model predicts a slightly smaller level difference between equally loud narrowband and broadband signals for short than for long durations. This occurs due to spectral broadening of the short stimulus, which effectively decreases the difference in bandwidth between the test and reference stimuli. The mid-left panel of Figure 3 shows that the predicted ratio between the loudness of long and short stimuli is approximately constant, except for a slight increase at low levels. This is in agreement with data of Epstein and Florentine [22, 35], who measured loudness functions for short and long 1000-Hz tones. Thus, the predictions of the DLM are consistent with the ELRH, which assumes that the loudness ratio between long and short stimuli is independent of level [20]. Additionally, Buus *et al.* [21] suggested that the loudness ratio is independent of spectrum, i.e. the same ratio was assumed for broadband and narrowband signals. The results of the present study show that this is predicted by the DLM (mid-left panel of Figure 3). Thus, the DLM in its original form predicts a loudness ratio independent of both level and spectrum and satisfies the ELRH. A corollary of the ELRH, however, is that spectral loudness summation should be essentially independent of duration. Figure 2 shows that this is also predicted by the DLM. In contrast, experimental data suggest that spectral loudness summation is larger for short signals [11, 12, 13, 14].

Decreasing the compressive exponent for short signals in the framework of the DLM yields slightly more spectral loudness summation for short signals. However, larger compression implies a shallower loudness function. Accordingly, adaptive compression leads to a loudness ratio between long and short signals which increases with level. This is inconsistent with experimental data, which indicate a constant or slightly decreasing loudness ratio (see [22, 35, 13] and top panel of Figure 3).

5.2. Comparison with a modified ELRH

Anweiler and Verhey [13] showed that most of the increase in spectral loudness summation for short signals was consistent with a modified ELRH, i.e. the hypothesis that loudness ratios are independent of level but not of spectrum. The ratios derived from their data are about 2.7 for 400 Hz and 2.0 for 6400-Hz wide stimuli in the range from 40 to 80 dB SPL (Figure 3). Thus, a larger ratio was found between short and long narrowband signals than for broadband signals. In the present study, two mechanisms were investigated which effectively introduce spectrum-dependent loudness ratios. When narrower filters are used at stimulus onset, the loudness of short, broadband stimuli is slightly enhanced. This leads to a decreased loudness ratio for broadband stimuli (cf. bottom left panel of Figure 3). The resulting loudness ratios for 400 and 6400-Hz wide noise bursts decrease with level. They differ by about a factor of 1.1 in the level range from 50 to 80 dB SPL. Simulations shown in the bottom left panel of Figure 2 indicate that a larger difference between the loudness ratios is necessary to account for the entire duration effect of spectral loudness summation. The simulations showed that the gain in spectral loudness summation is largely counteracted by the effect of spectral broadening, which may cast doubt on the hypothesis that adaptive auditory filters are the mechanism underlying increased spectral loudness summation.

Bandwidth-dependent integration was implemented as an alternative mechanism to introduce a duration dependence of spectral loudness summation. Effectively, this approach also causes bandwidth-dependent loudness ratios by increasing the loudness of short, broadband stimuli. In the current implementation, the increased integration efficiency is modeled as an effective amplification factor at stimulus onsets. The parameters were adapted to yield good agreement between data and model predictions of spectral loudness summation, which resulted in loudness ratios of about 2.4 and 1.5 in the range from 40 to 80 dB SPL.

It is worth noting that the modified model correctly predicts that the maximum amount of spectral loudness summation occurs at a higher level for short than for long signals. This finding could not be reproduced by the modified ELRH by Anweiler and Verhey [13] (see their Figure 4), which predicted the same reference level for maximum spectral loudness summation for short and long signals.

5.3. Dynamic stages in current loudness models

Figure 5 shows that the dependence of spectral loudness summation on repetition rate is reasonably well predicted by the modified version of the DLM using an integration process that depends on bandwidth and time. A transition from short burst to continuous signals begins to occur at a repetition rate of about 50 Hz, in agreement with data of Verhey and Uhlemann [14]. In the original version of the DLM, the dynamic properties are implemented in several stages: (i) a temporal window is used to extract

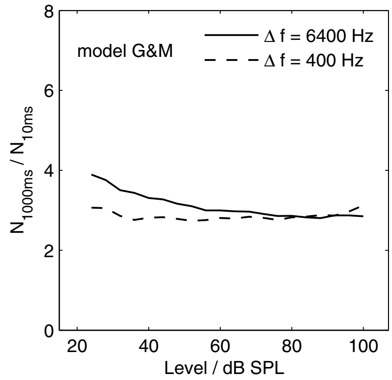

Figure 7. Loudness ratios predicted by the model of Glasberg and Moore [18] between bandpass noise of 1000-ms and 10-ms duration for bandwidths of 6400 Hz (solid line) and 400 Hz (dashed line).

the short-term rms-value, the form and width of which determine how sequences of short sound bursts are processed, (ii) forward-masking curves describe nonlinear effects for points in time succeeding peaks of the stimulus and (iii) temporal integration is modeled by a simple low-pass filter with a cut-off frequency of 8 Hz. As shown in Figure 2, the combination of these stages cannot predict the duration dependence of spectral loudness summation.

A more sophisticated temporal integration stage involving several time constants is used in the loudness model for time-varying sounds of Glasberg and Moore [18]. In order to compute short-term loudness and long-term loudness, time constants of about 22 and 50 ms, and 100 and 2000 ms are used, respectively. Figure 7 shows loudness ratios predicted by their model. The maximum of the short-term loudness was taken as a measure of the overall loudness of the noise bursts, following the suggestion of Glasberg and Moore [18]. The predicted curves indicate that this model also fails to predict the duration dependence of spectral loudness summation. As for the DLM, loudness ratios between long and short signals slightly decrease with level and approach a constant value, i.e. the ERLH is approximately fulfilled. At low levels, the loudness ratio for broadband stimuli is larger than that for narrowband noise. As discussed above, the opposite would be needed to account for greater spectral loudness summation of short stimuli, which means that the loudness model of Glasberg and Moore [18] cannot correctly predict the duration-dependence of spectral loudness summation.

In conclusion, loudness models need an additional dynamic stage to correctly describe the dynamics of spectral loudness summation. This stage has to be introduced prior to the integration of loudness across frequency. In order to keep the model applicable to all kinds of sounds without any a priori knowledge of spectrum or temporal shape, a detection of temporal onsets as used in the present study is needed. Of the three mechanisms proposed in the literature, a signal-adaptive spectro-temporal window as a bandwidth-dependent integration seems the most promising.

Within the framework of the DLM, this approach yielded the best correspondence between model predictions and experimental data on spectral loudness summation for short and long signals at moderate levels (see e.g. Figure 2). An inspection of equation 2 shows that the transformation from excitation to loudness used in the DLM may, however, be unrealistic at high levels. The limiting case of equation 2 for large input levels is a simple power law, i.e. $N'(z,t) \sim E^{\alpha}$. In contrast, several studies found a steeper, less compressive relation at high levels (e.g., [21, 22, 20, 13]). This steeper increase of loudness with level decreases the horizontal distance between the curves for broadband and narrowband signals at high levels, i.e. the magnitude of spectral loudness summation. Thus, in order for the DLM to correctly predict the dependence of spectral loudness summation on level over the whole level range shown in Figure 4, a modified loudness transformation is needed which accounts for the less compressive relation at high levels, such as used e.g. by Glasberg and Moore [18].

5.4. Interaction of spectral and temporal integration

The modifications of the DLM in the present study enhanced the loudness of short, broadband sounds using different mechanisms, which affected loudness predictions at stimulus onsets. As a consequence, for long, broadband sounds the modified models apply the same bandwidth-dependent amplification at the initial onset as for short stimuli. When the maximum of the loudness as a function of time is used to describe the overall loudness of a sound (as in the present study), however, then the initially increased loudness is outweighed by the temporal integration, which results in a stationary value at a later, larger loudness. For signals much shorter than the time constant of the temporal integrator, stationarity is never reached and the increased loudness at the onset results in a larger maximum loudness. This principle implies that, if a very large amplification is given at the initial onset, then a loudness "overshoot" could occur, and the temporal integration stage may not reach its stationary value at a larger loudness. In this case, the peak loudness in the beginning would determine overall loudness and the difference in amplification between long and short signals would disappear. In this case, the duration-dependence of spectral loudness summation could no longer be predicted. Therefore, when the modeling of loudness is based on mechanisms acting at stimulus onsets, the influence of the temporal integration stage has to be considered and the two stages need to be adjusted correspondingly.

In the present study, the temporal-integration stage of the DLM was not changed. The modifications were implemented such that no loudness overshoot was observed at the onsets of long, broadband stimuli and the parameters were optimized using data for 10- and 1000-ms long stimuli, as used by Verhey and Uhlemann [14].

The discrepancies between model prediction and data on spectral loudness summation for intermediate durations (Figure 6) suggest that further adjustments of amplification and the temporal integration stage may be needed to quantitatively predict spectral loudness summation as a function of duration.

6. Summary

Three mechanisms proposed in the literature to account for the duration dependence of spectral loudness summation were implemented as modifications of the DLM [19]. The results are summarized as follows:

- Adaptive compression: a higher compression at stimulus onset causes increased spectral loudness summation for short signals. The effect is too small, however, to account for the experimental results. Furthermore, different compression for short and long signals implies different slopes of the loudness functions, which has not been observed in experimental studies.
- Adaptive auditory filters: spectral loudness summation is increased when the width of auditory filters is narrowed. In the framework of the DLM, however, the effect is significantly reduced for short durations due to effects of spectral broadening.
- Bandwidth-dependent integration: a bandwidth-dependent temporal window was proposed in signal-detection experiments. In the DLM such a window, realized as amplification at stimulus onsets, can predict increased spectral loudness summation for short single and repeated noise bursts.
- The modified DLM underestimates spectral loudness summation for signals with a duration of 100 ms and overestimates spectral loudness summation for very high levels. A more sophisticated temporal-integration stage and a modified loudness transformation may be needed to account for these effects.

Acknowledgements

This work was partly supported by the Deutsche Forschungsgesellschaft (SFB tr31). We thank Densil Cabrera for providing part of the Matlab code of the loudness model by Glasberg and Moore [18] in the framework of the PsySound3 project. It was downloaded from http://web.arch.usyd.edu.au/~sfer9710/PsySound3/index on October 2, 2007.

References

[1] J. Rennies, J. L. Verhey, H. Fastl: Modellierung der Dynamik der spektralen Lautheitssummation (Modeling the dynamics of spectral loudness summation). DAGA 2008 – Fortschritte der Akustik, Proceedings of the 34th Annual Meeting of the Deutsche Gesellschaft für Akustik e.V., 2008, 449–450.

[2] J. Vos, G. F. Smoorenburg: Penalty for impulse noise, derived from annoyance ratings for impulse and road-traffic sounds. J. Acoust. Soc. Am. **77** (1985) 193–201.

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

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

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

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

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

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

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

[10] 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.

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

[12] 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.

[13] 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.

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

[15] 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.

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

[17] 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.

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

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

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

[21] 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.

[22] 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.

[23] M. F. B. van Beurden, W. A. Dreschler: Duration dependency of spectral loudness summation, measured with three different experimental procedures. – In: Hearing – From sensory processing to perception. B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, J. Verhey (eds.). Springer, Berlin, 2007, 237–245.

[24] DIN 45631: Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum, Verfahren nach E. Zwicker (Calculation of loudness level and loudness from the sound spectrum - Zwicker method). 1991.

[25] H. Fastl, E. Zwicker: Psychoacoustics - Facts and models. 3rd edition. Springer, Berlin, 2007.

[26] 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., 2007, 847–848.

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

[28] V. Hohmann: Frequency analysis and synthesis using a gammatone filterbank. Acta Acustica united with Acustica **88** (2002) 433–442.

[29] E. A. Strickland: The relationship between frequency selectivity and overshoot. J. Acoust. Soc. Am. **109** (2001) 2062–2073.

[30] W. A. C. van den Brink, T. Houtgast: Spectro-temporal integration in signal detection. J. Acoust. Soc. Am. **88** (1990) 1703–1711.

[31] P. Heil: Auditory cortical onset responses revisited. I. First-spike timing. J. Neurophysiol. **77** (1997) 2616–264.

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

[33] J. L. Verhey, J. Rennies: Temporal weighting in loudness perception: Effect of bandwidth. DAGA/NAG 2009 – International conference on acoustics, 2009.

[34] N. F. Viemeister, S. P. Bacon: Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones. J. Acoust. Soc. Am. **84** (1988) 172–178.

[35] M. Epstein, M. Florentine: Loudness of brief tones measured by magnitude estimation and loudness matching (L). J. Acoust. Soc. Am. **119** (2006) 1943–1944.

[36] G. Canévet, R. Hellman, B. Scharf: Group estimation of loudness in sound fields. Acustica **60** (1986) 277–282.

[37] E. Wagner, M. Florentine, S. Buus, J. McCormack: Spectral loudness summation and simple reaction time. J. Acoust. Soc. Am. **116** (2004) 1681–1686.