Influence of time constants and compression on the prediction of temporal integration of loudness

Jan Hots[1,3], Jan Rennies[2], Jesko L. Verhey[1]
[1] *Dept. of Experimental Audiology, Otto von Guericke University, D-39120 Magdeburg*
[2] *Fraunhofer IDMT / Hearing, Speech and Audio Technology, D-26129 Oldenburg*
[3] *Email: jan.hots@med.ovgu.de*

Introduction

Several studies have shown that the loudness of a sound depends on its duration (e.g., [1]). For equal-level sounds, an increase in the duration yields an increase in the loudness, at least for short durations. Or in other words, for short durations the level at equal loudness decreases as the duration increases. For durations of several hundred milliseconds, the loudness of sounds with different durations is almost the same at the same level.

The effect of duration on loudness is commonly referred to as temporal integration of loudness. In simple models using the instantaneous intensity as an input it is described in terms of a leaky integrator with a time constant of about 100 to 200 ms [2]. Some studies argue that more than one time constant is required to describe temporal integration over a wide range of durations [3, 4]. For example, in [3] it was shown that a better fit to temporal integration data is obtained when processing the instantaneous intensity with two serially aligned lowpass filters.

Recently, Rennies and coauthors [5] investigated to what extent current models of loudness perception of non-stationary sounds such as the dynamic loudness model (DLM, [6]) and the model of time-varying loudness (TVL, [7]) predict temporal integration of loudness using data of Poulsen [3]. Both models predicted almost the same levels at equal loduness as a function of duration. For both models a good correspondnace between data and model predictions was obtained at very short and at long durations. However, the duration beyond which the predicted loudness did hardly change with duration was shorter than found in the data and the temporal integration function was too steep for smaller durations. The aim of this study is to shed some light on the reasons of these deviations from the data and how the temporal integration stage has to be changed within the model to improve the prediction. The study starts with simple models on the basis of the instantaneous intensity. Then the effect of peripheral compression is investigated. In the final step, loudness was predicted with modifications of the extended version of the DLM (eDLM) proposed in [8]. The extended version was used since it includes a bandwidth dependent onset enhancement which may affect loudness at very short durations where spectral splatter is no longer negligible. The data for a reference level of 55 dB SPL of the longest signal of Poulsen [3] were used, since the data were obtained for a wide range of durations and, at this level, peripheral compression has a large effect on temporal integration [9].

Models

Leaky integrator

Four different simple leaky integrator models were tested. In the first stage of each of these models the instantaneous intensity of the input signal was calculated by squaring the amplitude. In two models this was directly followed by the integration stage which either consisted of a single leaky integrator ($\tau = 80$ ms; model one LI) or two consecutive leaky integrators ($\tau_1 = 10$ ms and $\tau_2 = 50$ ms; model two LI). These linear models and time constants were proposed in [3] to predict the data considered here. In the other model versions a compressive stage preceded the integration stage. The instantaneous intensity values were raised to the power of $\alpha = 0.3$. This value of α was chosen since loudness approximately doubles when intensity is increased by 10 dB [1]. Two nonlinear leaky integrator model versions including compression were tested. The integration stage was realized as a combination of two integrators, one with a short and one with a long time constant. In one of these models (two LI c ser), the integrators were serially arranged. In the other model version (two LI c par), they were arranged in parallel and the output of the two integrators was set to their mean value, as proposed in [4]. In these two model versions, the time constants were fitted to the data by minimizing the root of the mean quadratic deviation between predictions and the data (rms_{diff}). In all model versions, the overall loudness was determined by the maximum of the output of the integration stage.

DLM

The basic structure of the eDLM [8] is the same as the one of the Dynamic Loudness Model (DLM, [6]). In addition, the eDLM uses a bandwidth-dependent amplification at stimulus onset to account for the larger spectral loudness summation for short signals than for long signals. The temporal integration stage of the model is realized as a first-order low-pass filter with a cut-off frequency of 8 Hz (which corresponds to an effective time constant of 86.5 ms, when the compressive relation between intensity and specific loudness ($\alpha = 0.23$) is accounted for). The maximum of its output is taken as an estimate of the overall loudness.

To investigate the influence of the temporal integration stage on the prediction of temporal integration data in this model framework, the original version of the eDLM was compared to a modified version (eDLM mod) with two parallel leaky integrators, one with a short and one with a long time constant, as proposed in [4]. The output of this temporal integration stage, i.e., the short-term loudness, was set to the mean of the two integrators. Time constants were optimized in the same way as for the leaky integrator models.

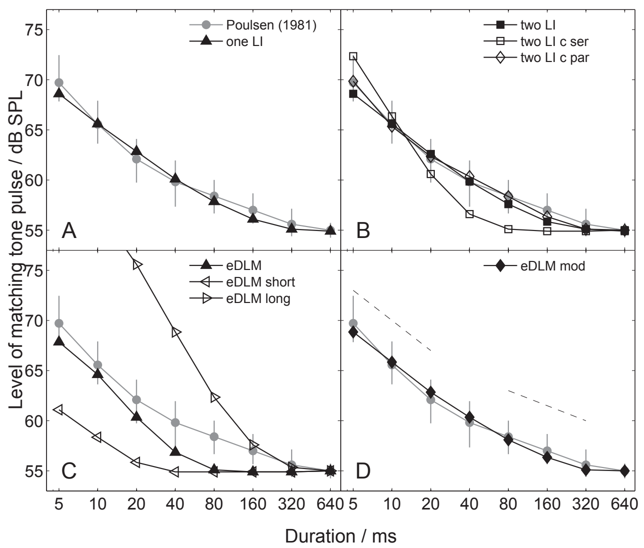

Figure 1: Temporal integration data and predictions. Experimental data for a reference level of 55 dB [3] are shown in all panels (gray dots, errorbars indicate 95% confidence limits). **A** Predictions of a linear leaky integrator model with one time constant (one LI, filled triangles). **B** Predictions of linear leaky integrator models with two time constants: linear consecutive (two LI, filled squares), with compression consecutive (two LI c ser, squares) and with compression arranged in parallel (two LI c par, diamonds), the latter two with optimized time constants. **C** Predictions of the eDLM with default parameter set (filled triangles), with a short time constant (eDLM short, left-pointing triangles) and with a long time constant (eDLM long, right-pointing triangles). **D** Predictions of the eDLM with modified integration stage and time constants fitted to the data (diamonds).

Results and Discussion

Figure 1 shows data and model predictions of temporal integration of loudness for various models. Each panel shows the levels at equal loudness of 1-kHz tone bursts with durations of 5 to 320 ms compared to a 1-kHz tone with a duration of 640 ms and a level of 55 dB SPL as a function of the signal duration. The experimental data of Poulsen [3] for a reference level of 55 dB SPL are shown in all panels of Figure 1 with gray dots. Errorbars indicate the 95% confidence interval. Black symbols in Figure 1 indicate model predictions.

Panel A shows model predictions of the single leaky integrator (one LI) proposed by Poulsen [3] with black triangles. The good correspondence between data and predictions of this model (root of the mean quadratic deviation between predictions and data $\mathrm{rms}_{diff} = 0.6$ dB; maximum difference $\max_{diff} = 1.1$ dB) indicates that a single time constant already accounts for most aspects of the data.

Panel B shows predictions of the two LI model proposed by Poulsen [3] with filled squares. There is hardly any difference in the predictions of this model containing a second time constant compared to the predictions of the one LI model ($\mathrm{rms}_{diff} = 0.7$ dB and $\max_{diff} = 1.2$ dB for the two LI model) shown in panel A. This is presumably due to the choice of the short time constant which was close to the shortest signal duration of 5 ms.

In general, both models, one LI and two LI, predict the data well. This is also true of the slope of the predicted temporal integration function which was close to -3 dB per doubling for short durations and decreased towards longer durations.

A limitation of the above models are that they do not include compression as a fundamental property of the auditory system. Predictions of the two LI model with a compression stage (two LI c ser) are shown by open squares in Panel B. The optimized effective time constants of this model are $\tau_{short} = 1$ ms and $\tau_{long} = 59$ ms (for effective time constants see, e.g., [2]). The resulting temporal integration function was considerably steeper than the experimental data ($\mathrm{rms}_{diff} = 2.1$ dB, $\max_{diff} = 3.3$ dB). Thus, the deviations between data and predictions reach values up to 3.3 dB at intermediate durations. Note that τ_{short} is so short that the model does not benefit from the second integrator, i.e., it is essentially a single leaky integrator model.

Empty diamonds in panel B indicate predictions of the model with two integrators arranged in parallel preceded by a compression stage (two LI c par). The optimized effective time constants were $\tau_{short} = 17$ ms and $\tau_{long} = 300$ ms. This model version provides a good match between predictions and data for all durations ($\mathrm{rms}_{diff} = 0.4$ dB, $\max_{diff} = 0.7$ dB).

Panel C shows predictions of the unmodified eDLM (filled triangles). The predictions are very similar to those shown for the DLM and the TVL in [5]. For the two shortest durations (5 ms and 10 ms) and the longest duration (640 ms) there is a good agreement between predictions and data but the model showed deviations at intermediate durations ($\mathrm{rms}_{diff} = 2.0$ dB, $\max_{diff} = 3.3$ dB). The similarity between predictions of the original DLM (see [5]) and the eDLM indicates that the additional stage for the duration effect in spectral loudness summation did not alter the predictions for temporal integration of loudness for tones.

The predicted slope of more than -3 dB per doubling results from the compression included in the model (cf. panel B). However, the temporal intergation function was not as steep as in the simulation of the two LI c ser model (top right panel). This is presumably mainly due to effects of spectral splatter at short durations, i.e., it results from an increased loudness of short signals due to spectral loudness summation. A simple change of the time constant in this model does not lead to a substantial improvement of the predictions. The effect of the time constant is shown in panel C with model predictions using a shorter effective time constant of $\tau_{short} = 35\,\mathrm{ms}$ (left-pointing open triangles) and a longer effective time constant of $\tau_{long} = 398\,\mathrm{ms}$ (right-pointing open triangles) than used in the original model.

In panel D, the predictions of the eDLM with a modified temporal integration stage (eDLM mod) are shown (filled diamonds). The optimized time constants of the two parallel leaky integrators had effective values of $\tau_{short} = 35\,\mathrm{ms}$ and $\tau_{long} = 398\,\mathrm{ms}$. For the whole range of durations this model provided a good prediction of the data ($\mathrm{rms}_{diff} = 0.6\,\mathrm{dB}$, $\max_{diff} = 0.9\,\mathrm{dB}$). The slope of the predicted temporal integration function was about -1.5 dB per doubling for medium to long durations and approached -3 dB per doubling for short durations (slopes are visualized by dashed lines in the bottom right panel). The good prediction is primarily due to the parallel arrangement of the leaky integrators. Note that, with only a single leaky integrator, the slope was too large, as shown in panel C of Figure 1 and it would be even larger for two serially aligned integrators.

In summary, the present study shows that a parallel organization of leaky integrators with different time constants overcomes the problem of too steep predicted temporal integration functions in models with cochlear compression. The improvement of temporal integration by using the modified temporal integration stage was demonstrated within the model framework of the eDLM. Given the previously shown similarities of the TVL and the DLM with respect to temporal integration it is reasonable to assume that a similar change of the stage calculating the short-term loudness would lead to a similar improvement in prediction accuracy of the TVL for the temporal integration data considered here.

References

[1] Epstein, M., Marozeau, J.: Loudness and intenstity coding. *The Oxford Handbook of Auditory Science – Hearing* **3**, Edited by C.J. Plack, (Oxford University Press, ISBN 978–0–19–923355–7) (2010) 45–69

[2] Verhey, J.L.: Temporal resolution and temporal integration. *The Oxford Handbook of Auditory Science – Hearing* **3**, Edited by C.J. Plack, (Oxford University Press, ISBN 978–0–19–923355–7) (2010) 105–121

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

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

[5] Rennies, J., Verhey, J.L., Fastl, H.: Comparison of loudness models for time-varying sounds. Acta Acust. united Ac. **96** (2010) 383–396

[6] Chalupper, S., Fastl, H.: Dynamic loudness model (DLM) for normal and hearing-impaired listeners. Acta Acust. united Ac. **88** (2002), 378–386

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

[8] Rennies, J., Verhey, J.L., Chalupper, S., Fastl, H.: Modeling temporal effects of spectral loudness summation. Acta Acust. united Ac. **95** (2009) 1112–1122

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