Modelling adaptation and amplitude-modulation processing

Carolin Iben, Stephan D. Ewert
Medizinische Physik, Carl-von-Ossietzky Universität Oldenburg, 26111 Oldenburg
E-Mail:carolin.iben@uni-oldenburg.de, stephan.ewert@uni-oldenburg.de

Introduction

The computational auditory signal processing and perception model (CASP, [1]) accounts for various aspects of simultaneous and non-simultaneous masking in human listeners. The model introduced improvements over the auditory pre-processing stages of the perception model (PEMO, [2]). These pre-processing stages were used in a number of monaural and binaural applications and modelling studies. The current study focuses on two main aspects of the pre-processing stages of CASP, adaptation and modulation processing. Both stages interact and are responsible for main properties of the model, like the ability to predict Weber's law in intensity discrimination tasks, forward masking, and modulation masking. The ability to predict Weber's law is coupled to the properties of the adaptation stage. Slow temporal fluctuations (below about 2 Hz) and the steady state component of the stimuli are quasi-logarithmically transformed. This transform is followed by a linear modulation lowpass filter and a fixed-variance internal noise, yielding the models ability to predict Weber's law. In contrast, amplitude fluctuations or modulations (AM) are linearly processed with regard to their modulation depth. As a consequence, the consecutive processing by linear bandpass modulation filters and the addition of a fixed-variance internal noise, does not account for Weber's law in the AM-depth domain as reported in [3]. Another key property of the adaptation stage is the transformation of amplitude fluctuations to a nearly level-independent internal representation. This property enables the model to correctly account for the relatively small effect of overall level on modulation perception (at least for medium range levels). In the past, the auditory filterbank in PEMO and CASP was typically restricted to a range of filters determined by the bandwidth of the stimuli, partly for computational optimization purposes. The usage of such *a-priori* knowledge is, however, not practical if the model is applied as front-end for quality assessment or in hearing aid algorithms.

Here, interactions between the non-linear auditory filters in the CASP model and the adaptation stage were investigated. The modulation processing stage was modified to allow for the correct prediction of AM-depth discrimination based on [3]. An energy weighting of auditory channels was introduced to counteract interactions between the non-linear auditory filters and the adaptation stage, and to allow for an automated, signal-driven reduction of computational load by ignoring channels with small contribution to the overall model prediction. The goal of the study was to provide a more generalized auditory pre-processing model for the prediction of psychoacoustic data and as a front-end for speech and audio-quality assessment (e.g., PEMO-Q [4]).

Extensions of the CASP model (eCASP)

Power-weighting of auditory channels

In the CASP model, the non-linear auditory filterbank compresses and thus reduces the AM-depth of fluctuating stimuli at medium levels. The included non-linear filters account for frequency-specific compression, according to physiology. This causes the maximum amount of compression to be applied to amplitude fluctuations of the spectral components near the center frequency of the filter. In contrast, dynamic fluctuations of off-frequency components (particularly low-frequency components) are linearly processed and only attenuated by the filter. One disadvantage of this frequency-specific compression in combination with the adaptation stage of [2] in the CASP model is that the AM-depth of the fluctuations is best represented in auditory filters far (about one octave) above the original carrier frequency after the adaptation stage. In such filters tuned two about twice the carrier frequency, the attenuation is compensated by the level-invariant mapping of AM-depth in the adaptation stage, while the AM-depth was not reduced due to the absence of compression (the carrier fell in the linear low-frequency tail of the filter). This interaction of the non-linear auditory filterbank and the adaptation stage can cause implausible internal representations with a maximum excitation in auditory channels far off the dominant frequencies in the stimuli. In contrast, the internal representations of the stimuli in PEMO always show the best representation of AM in or close to the auditory filters that are maximally excited by the stimuli. Such a behaviour appears more realistic and is also suited as auditory front-end in, e.g., quality assessment (PEMO-Q) without pre-selection of "appropriate" auditory filters based on *a-priori* knowledge.

Here, an energy weighting of the auditory channels was introduced. The root-mean-square (RMS) of the output of the auditory filters was calculated and averaged during the template generation process (for details see [1,2]). The resulting RMS values were normalized to the maximum channel and applied as channel-dependent weights to the template. This procedure introduced an automatic selection of channels with high spectral excitation. Additionally, filter channels with weighting coefficients 40 dB below the maximum were not considered to automatically reduce the computational load.

Modification of modulation processing

The modulation filters in the CASP model were replaced against broader modulation filters as suggested in [5]. These filters were directly fitted to modulation masking data in the context of the much simpler envelope power-spectrum model (EPSM) and were so far not tested in the context of the PEMO and CASP model. A second extension was the introduction of a non-linear transformation of the (bandpass) modulation filter output based on [3]. In contrast to a logarithmic transform suggested for PEMO in [3], the transformation $s_t = as/(b+s)$ was here applied to the output of each modulation filter, s, with $a = 60$ and $b = 12$.

Model evaluation and discussion

Model predictions of the extended CASP (eCASP) were compared to psychoacoustic data and predictions of the original CASP for three critical experiments. The psychoacoustic data were taken from the literature. The model predictions were obtained with the same measurement framework (AFC for Matlab) and procedure (3-interval, 3-alternative forced-choice with 1-up, 2-down rule) as used to obtain the psychoacoustic data.

Figure 1 shows detection threshold data (black circles) and predictions of the CASP (red squares) and eCASP (blue triangles) as a function the offset-onset time in a forward masking task (for details see [1]).

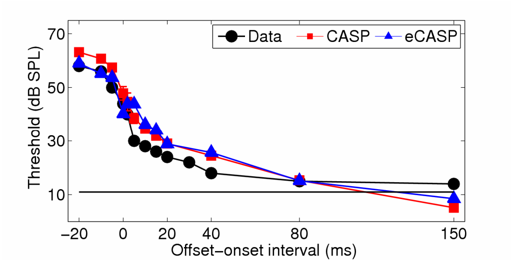

Figure 1: Forward-masked thresholds as a function of the offset-onset interval between the envelope of a 60-dB, 200-ms broadband masker and a 10-ms probe at 4 kHz.

In the region of simultaneous masking (-20 to 0 ms) and for large separations of the masker and probe, eCASP shows a slightly better agreement with the data.

Figure 2 shows model predictions and data in a modulation masked-threshold pattern (MTP) paradigm used to estimate the modulation filter shape in [6] (see also for details).

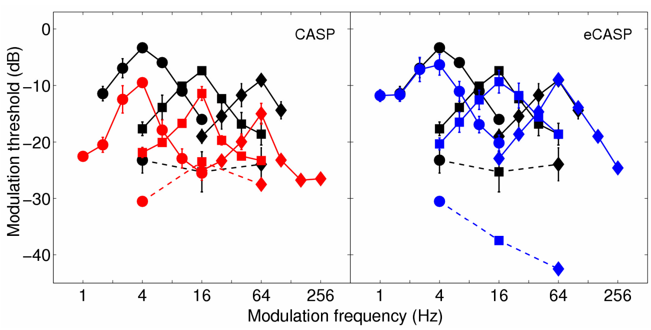

Figure 2: Masked-threshold patterns and modulation detection thresholds (dash-dotted at the bottom) for modulation frequencies of 4, 16, and 64 Hz (2.8 kHz carrier). The modulation masker was a narrowband noise (bandwidth 1.4, 5.6, and 22.3 Hz, respectively) shifted -2 to 2 octaves (2/3-oct steps) relative to the probe modulation.

The eCASP model shows a much better agreement with the peaked MTPs, however, modulation detection thresholds (bottom) were considerably overestimated. This mismatch is related to the so far not optimized non-linear transform at the output of the modulation filters.

Figure 3 shows the predictions and data for an AM-depth discrimination experiment (for details see [3]) in the same style as in Fig. 1. The eCASP model accounts for Weber's law in the AM-domain (1-dB criterion indicated by the horizontal line) as observed in the data for the sine carrier. For the noise carrier, eCASP shows too high discrimination thresholds for small reference modulation depths.

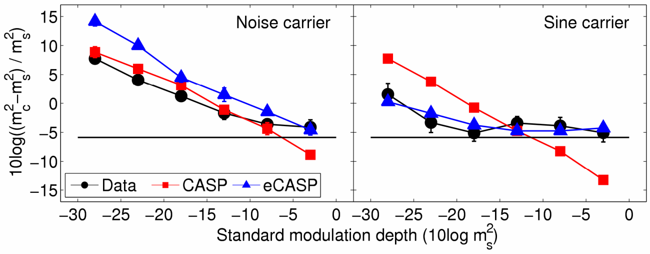

Figure 3: AM-depth discrimination thresholds expressed as Weber fraction as a function of the reference modulation depth. The left and right panel are for a broadband noise carrier (2-4 kHz) and a sine carrier (4 kHz), respectively.

Conclusions and outlook

Overall, the extended CASP model (eCASP) showed better agreement with the data than the original model without the requirement to pre-select the appropriate auditory filter ranges. There were, however, still considerable deviations to the data that require further modification of the non-linear transformation in the AM domain. A further future step is a separation of the adaptation stage in a fast stage with a time constant about 5 ms and a slower stage, to allow for binaural processing after fast adaptation. [This work was supported by the BMBF ("Modellbasierte Hörsysteme").]

References

[1] Jepsen, M. L., Ewert, S. D., und Dau, T.: Modeling spectral and temporal masking in the human auditory system. J. Acoust. Soc. Am. 124 (2008), 422-438

[2] Dau, T., Kollmeier, B., and Kohlrausch, A.: Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. J. Acoust. Soc. Am. 102 (1997), 2892–2905

[3] Ewert, S. D., und Dau, T.: Internal and external limitations in amplitude-modulation processing. J. Acoust. Soc. Am. 116 (2004), 478-490

[4] Huber, R., Kollmeier, B.: PEMO-Q - A new Method for Objective Audio Quality Assessment using a Model of Auditory Perception. IEEE Transactions on Audio, Speech and Language processing, Vol. 14, no. 6 (2006), 1902 - 1911.

[5] Ewert, S. D., and Dau, T.: Characterizing frequency selectivity for envelope fluctuations. J. Acoust. Soc. Am. 108 (2000), 1181-1196

[6] Ewert, S. D., Verhey, J. L., and Dau, T.: Spectro-temporal processing in the envelope-frequency domain. J. Acoust. Soc. Am. 112 (2002), 2921-2931