Evaluation of a physiological ear model for the simulation of nonlinear masking effects

Frank Baumgarte (baumgart@tnt.uni-hannover.de)

Universität Hannover, Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Appelstrasse 9A, D-30167 Hannover, Germany

Summary A physiological model of the human auditory system with the aim at generating masked thresholds for arbitrary complex sound signals is evaluated with respect to nonlinear mechanisms in masking. In [1, 2] it was shown that the ear model is able to rebuild the level-dependency of spectral and temporal masking patterns, the asymmetry of masking between noise and tone, and the different detection thresholds for amplitude and frequency modulation. The simulation results in this paper indicate that nonlinear mechanisms such as the additivity of masking, suppression, and distortion product detection are reproduced as well by the physiological ear model. Since the model represents a unified approach comprising the most important masking effects of complex sounds, it is suitable for applications in audio signal processing.

1 INTRODUCTION

Audio coding as well as audio quality assessment represent application examples for perceptual models which require the processing of complex audio signals. In contrast, most psychoacoustical measurements are carried out with simple signal configurations like combinations of narrow-band noises or pure tones. Therefore, the task of such perceptual models consists not only of rebuilding the required psychoacoustical results, but also of performing an extrapolation for more complex signals. A common goal in these applications is the generation of masked thresholds for the evaluation of the audibility of signal distortions. Currently used perceptual models for masked threshold generation widely ignore the highly nonlinear properties of sound processing in the auditory system. Therefore, the influence of the nonlinearity on masking can not be considered sufficiently.

In this paper, the modelling of three nonlinear mechanisms affecting masking will be investigated: (1) The perception of distortion products created in the cochlea. (2) The suppression of a test signal by a suppressor signal. (3) The additivity of masking in case of two superimposed maskers. All three mechanisms can influence the amount of masking but are not considered in currently applied perceptual models. A physiological ear model published in [1, 2] is used for the simulation of these nonlinear mechanisms. The main part of this model is a nonlinear cochlear model, which rebuilds the behavior of the outer hair cells (OHCs) as “cochlear amplifiers” with saturation. It was already shown, that this physiological model rebuilds masking properties like the classic simultaneous masking patterns, post-masking, and the detection of amplitude and frequency modulation. Due to the physiological approach the modelling results are valid independently from signal complexity.

Chapter 2 contains a brief description of the physiological model. A more complete one can be found in [1, 2]. The simulation results of the three nonlinear mechanisms are outlined in Chapter 3. Concluding remarks are summarized in Chapter 4.

2 PHYSIOLOGICAL EAR MODEL

The physiological ear model consists of sub-models for the outer, middle, and inner ear, complemented by a sub-model for the neural processing in the central auditory system (see Figure 1). The sub-models up to the neural processing stage are extensively based on physiology. A psychoacoustical modelling approach was chosen for the central auditory system. The audio signal is input to the outer and middle ear model (OME) realized as a linear filter. The cochlear model consists of sections connected in series which represent the mechanical properties of small longitudinal cochlear segments including the fluids, the basilar membrane (BM) and the outer hair cells (OHCs). In each section the BM velocity is derived for a given oval window input signal from the OME model. The OHC models act as cochlear amplifiers, which mechanically amplify the travelling wave at low input levels and saturate at high levels. The BM velocity signal combined with the amplified signal from the OHC results in the inner hair cell (IHC) stimulation signal. The IHC model transforms the stimulation by a square function and low-pass filtering into the neural activity signal. The following sub-model for the neural processing (NP) determines the specific loudness [3] as internal representation of the input audio signal. Specific loudness changes are processed by a short-time integrator and evaluated by a threshold detector in the NP model.

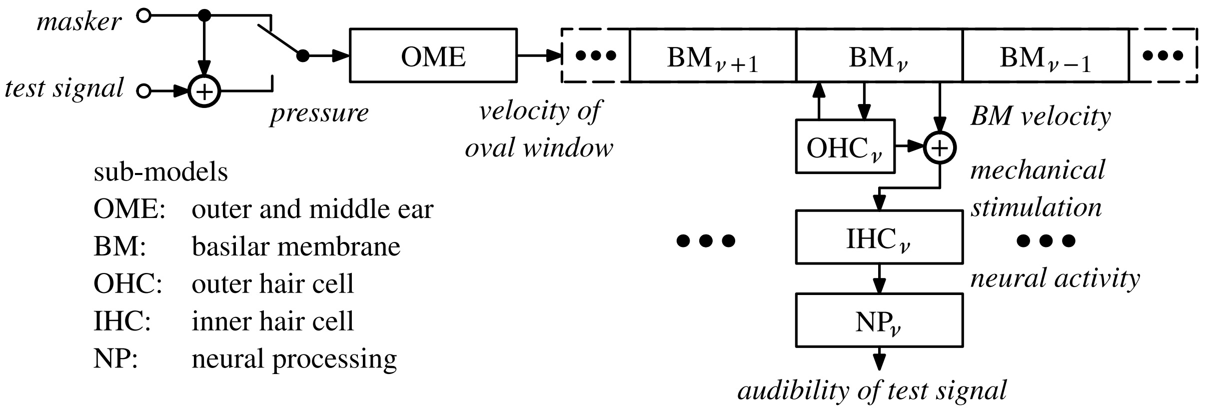

FIGURE 1 Physiological ear model. All sections have the same structure as the outlined section ν.

With a given masker the masked threshold for a test signal is determined by the model corresponding to a psychoacoustical measurement procedure. In the first step the specific loudness of the masker is derived as internal representation. This masker representation is stored in a memory inside the NP model. The masker with superimposed test signal is processed in the next steps. The change of the internal signal representation is derived by calculating the ratio of specific loudnesses from the masker with and without superimposed test signal. After a short-time integration the audibility of the test signal is detected independently in each model section by a comparison of the ratio with an internal threshold value. By repeating this step, the test signal level is iteratively adjusted so that it converges on the masked threshold level.

3 MODELLING RESULTS

The three nonlinear mechanisms examined below were selected for a further evaluation of the physiological ear model, because these mechanisms are assumed to be important for masking by complex signals.

Suppression arises in the cochlea due to the interaction of a masker and test signal. The saturation of the cochlear amplifiers results in a reduced test signal gain, if the amplifiers are partially saturated by the masker signal compared to the case of a test signal presented without masker. Masking is affected by suppression because the test signal will be less detectable if suppression is present.

The amount of suppression can be observed in physiological measurements of firing rates on the auditory nerve. This is performed by first presenting the test tone alone and monitoring the mean firing rate of a nerve fiber connected to an IHC near to the characteristic-frequency (CF) position. The CF position is the longitudinal BM location where the maximum velocity response to the test tone is observed. A suppression tuning curve is determined by the superposition of a suppressor tone. The suppressor level is adjusted at different frequencies so that the monitored mean firing rate is reduced by a given factor. The simulation results from the physiological ear model in Figure 2 reproduce the shape of the suppression areas on both sides of the test tone frequency in qualitative agreement with [4].

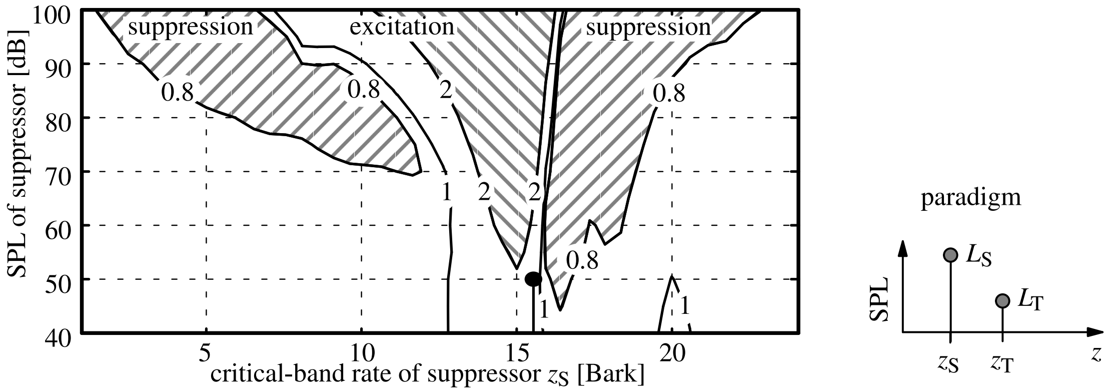

FIGURE 2 Suppression tuning curves for a test tone at 3 kHz (15.7 Bark) and L_T = 50 dB SPL. A factor of 0.8 is equal to a suppression of 20% (S: suppressor, T: test signal).

Distortion products are created in the cochlea due to the nonlinearity of the cochlear amplifiers [3]. Odd order distortions dominate because of the approximately symmetrical saturating characteristic. In case of psychoacoustical tone-on-tone masking measurements the cubic difference tone is often audible which has a frequency of $f_{d3} = 2f_M - f_T$ for $f_T > f_M$. The cubic difference tone can be detected at a considerably lower test tone level than the test tone itself so that the masked threshold will be reduced accordingly.

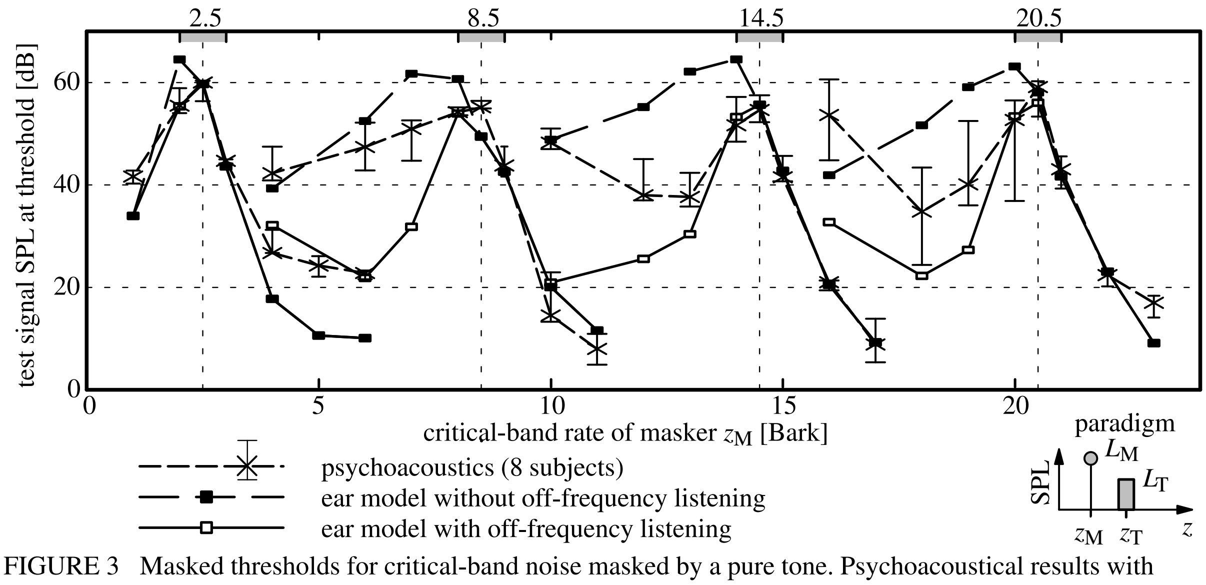

FIGURE 3 Masked thresholds for critical-band noise masked by a pure tone. Psychoacoustical results with mean and interquartile ranges and modelling results. Four different critical-band wide noises were used as test signal at z_T = 2.5, 8.5, 14.5, and 20.5 Bark fixed center frequency. The masker SPL was constant at L_M = 80 dB SPL (M: masker, T: test signal).

Figure 3 shows psychoacoustical masked thresholds for single pure tone maskers and a critical-band wide noise as test signal. The measure was performed for a masker at different frequencies and a test signal at a fixed center frequency. The psychoacoustical masked thresholds for the 2.5 and 7 kHz test signal are significantly reduced for critical-band rates up to 3 Barks below the test signal frequency due to the distortion product detection. As expected, the physiological ear model does not rebuild the reduced masked threshold, if the change detection is restricted to a small surrounding of the test signal CF location. If the detection is unrestricted, e.g. off-frequency listening is enabled, the audible distortion products are detected and result in a decreased masking. The ear model shows a higher sensitivity to distortion products than observed in the measurements. It detects distortion products for the 1-kHz test signal and shows lower masked thresholds below the test signal frequency for the 2.5 and 7 kHz test signals.

The combination of two or more maskers usually results in an increased masked threshold compared to each single masker and is called additivity of masking [3]. Standard models predict the masked threshold of the combined maskers by calculating the intensity sum of the masked threshold of each individual masker. This results in a maximum masking increment of 3 dB for the case of two maskers. In contrast, psychoacoustical measurements show that the increment can amount 13 to 20 dB referred to the maximum threshold of the two single maskers. This masking increase is near to the higher value (20 dB) for maskers which have a spectral distance of more than the critical bandwidth [5, 6].

In the simulation two pairs of narrow-band Gaussian noises with 50 Hz bandwidth were used. The center frequencies of the first pair were 1 kHz and 2.1 kHz. The maximum masking increment amounts to 22.6 dB in excess of the maximum masked threshold of each single masker. The center frequencies of the second pair were 1.8 kHz and 2.1 kHz, a frequency distance of one critical band. In this case the maximum masking increment is 13.3 dB in comparison to the masked thresholds of the individual maskers. This result shows that the considerable masking excess compared to an intensity summation is reproduced by the model and that higher masking increments are obtained for maskers with a larger frequency distance consistent with psychoacoustical results [5].

4 CONCLUSIONS

A physiological ear model was evaluated with respect to nonlinear mechanisms in masking. The examined mechanisms were the detection of distortion products, suppression, and the additivity of masking. Special signal configurations known from psychoacoustics were used in order to emphasize each mechanism in the simulation. The comparison of measurements and modelling results shows that these mechanisms are rebuild by the model. The detection of distortion products is overestimated by the model for the investigated signals while the results for suppression and additivity of masking correspond to physiological and psychoacoustical observations. However, the degree of quantitative agreement cannot be finally assessed on the basis of the small number of different signals in this study.

The nonlinear mechanisms are not limited to the signal configurations used here but are also important for more complex signals. For example, the additivity of masking results in a significantly increased masked threshold of superimposed maskers which will be underestimated by models which neglect these nonlinear mechanisms. This is a reason why the physiological ear model is expected to be better suited for complex audio signals.

REFERENCES

[1] Baumgarte, F. ‘‘A Physiological Ear Model for Auditory Masking Applicable to Perceptual Coding,’’ *presented at the 103rd AES Convention*, Preprint 4511, New York, NY, (1997).

[2] Baumgarte, F. ‘‘Evaluation of a Physiological Ear Model Considering Masking Effects Relevant to Audio Coding,’’ *presented at the 105th AES Convention*, Preprint 4789, San Francisco, CA, (1998).

[3] Zwicker, E.; Fastl, H; *Psychoacoustics: Facts and Models*, Springer, New York, 1990.

[4] Arthur, R. M.; Pfeiffer, R. R.; Suga, N. ‘‘Properties of ‘two-tone inhibition’ in primary auditory neurones,’’ *J. Physiol.* **212**, pp. 593–609 (1971).

[5] Lutfi, R. A.; ‘‘Additivity of simultaneous masking,’’ *J. Acoust. Soc. Am.* **73** (1), pp. 262–267, (1983).

[6] Moore, B. C. J.; ‘‘Additivity of simultaneous masking, revisited,’’ *J. Acoust. Soc. Am.*, **78** (2), pp. 488–494, (1985).