Prediction of masking thresholds for Schroeder phase maskers: masker level effect

Václav Vencovský[1]
[1] *Musical Acoustics Research Center, Academy of Performing Arts in Prague, 118 00 Prague, Czech Republic*
Email: vaclav.vencovsky@gmail.com

Abstract

Schroeder phase maskers are harmonic complex tones with starting phases of individual harmonics given by an equation proposed by Schroeder [1]. Relative phase between the harmonics may affect masking thresholds. This masker phase effect depends on the masker level: difference between masking thresholds increases with increasing level. This study used four auditory models to predict masking thresholds for Schroeder phase maskers of various levels. The models contained different algorithms simulating response of the basilar membrane (BM): dual resonance nonlinear (DRNL) filterbank of Lopez-Poveda and Meddis [2]; a transmission line model of Verhulst *et al.* [3]; a transmission line model of Baumgarte [4]; and a hydrodynamic model of Nobili *et al.* [5]. The algorithms were extended by an inner hair cell model and a decision device. The DRNL filterbank model predicted the same masker phase effect for the lowest and highest masker level. Transmission line models showed the opposite dependence: the masker phase effect decreased with increasing level. The hydrodynamic model for some of the maskers predicted thresholds with qualitative agreement with behavioral data.

Introduction

Schroeder phase maskers used in this study were harmonic complex tones whose spectral components have equal amplitudes and starting phases are given by

$$\theta_n = C\pi n(n+1)/N, \tag{1}$$

where n is the nth spectral component and N is the overall number of spectral components. The complexes with starting phases calculated for $C = +1$ are called "positive Schroeder phase" and for $C = -1$ "negative Schroeder phase" complexes. Behavioral studies (e.g. [6]) showed that although the positive and negative Schroeder phase complexes have approximately same, flat temporal envelope, they may produce different masking thresholds (differences up to 20 dB). These masker phase effects depend also on the masker level [7, 8]. All these effects can be accounted for on a peripheral level and since many of cochlear models cannot predict the masker phase effects, the stimuli put a strong constraint on their function [9].

This study uses four different cochlear models to predict the effects of level on masking thresholds. Since also the parameters of the cochlear models and the used method of prediction affect the predicted thresholds, the study is not aimed to prove any of the used modeling approaches as wrong. Instead, the aim is to compare the predicted masking thresholds for different types of cochlear models with given parameters.

Cochlear models

Nobili et al. model

The cochlear model proposed by Nobili *et al.* [5] approximates the BM as an array of damped oscillators (with mass and stiffness). The oscillators are coupled through the incompressible fluids in the cochlea. The coupling is modeled by the method of Green's functions. The active function of the cochlea is modeled by a feedback force which undamps the oscillators. This study uses the Nobili et al. model proposed with parameters and dimensions of the human cochlea [5]. I have changed the damping of the oscillators in order to increase the frequency selectivity of the model. Table 1 shows equivalent rectangular bandwidths (ERB) of the simulated cochlear filters measured in six discrete outputs of the model (with characteristic frequency (CF) of 0.125, 0.25, 0.5, 1, 2 and 4 kHz). The values of ERB_{GM} are ERB estimated from behavioral data [10]. Responses of the model are nonlinear and the model can simulate otoacoustic emissions [5].

Table 1: Nobili et al. model: equivalent rectangular bandwidths (ERB) of the simulated cochlear filters. ERB_{GM}: behavioral data [10]

level	characteristic frequency (kHz)					
(dB SPL)	0.125	0.25	0.5	1	2	4
20	43	62	89	141	225	390
40	43	62	90	148	245	521
60	43	70	122	201	337	818
80	54	98	168	307	528	1107
ERB_{GM}	38	52	79	133	241	456

Table 2: DRNL model: ERB of the simulated cochlear filters. ERB_{GM}: behavioral data [10].

level	characteristic frequency (kHz)					
(dB SPL)	0.125	0.25	0.5	1	2	4
20	37	48	73	117	204	303
40	37	48	73	156	272	403
60	37	49	99	172	287	418
80	47	77	130	245	391	541
ERB_{GM}	38	52	79	133	241	456

Table 3: Verhulst et al. model: ERB of the simulated cochlear filters. $\mathrm{ERB_{GM}}$: behavioral data [10].

level	characteristic frequency (kHz)					
(dB SPL)	0.125	0.25	0.5	1	2	4
20	30	48	104	141	221	492
40	43	64	96	159	246	493
60	74	129	151	369	576	801
80	76	138	204	486	1272	2349
$\mathrm{ERB_{GM}}$	38	52	79	133	241	456

Table 4: Baumgarte model: ERB of the simulated cochlear filters. $\mathrm{ERB_{GM}}$: behavioral data [10].

level	characteristic frequency (kHz)					
(dB SPL)	0.125	0.25	0.5	1	2	4
20	48	68	103	177	351	620
40	51	74	120	214	387	716
60	52	83	139	246	438	822
80	59	94	162	298	538	1010
$\mathrm{ERB_{GM}}$	38	52	79	133	241	456

DRNL model

Lopez-Poveda and Meddis [2] proposed a filterbank model composed of a dual-resonance nonlinear (DRNL) type of filters. The DRNL filter has two parallel bandpass processing paths, one with a linear and the other with a compressive nonlinear gain. Each path consists of gammatone filters. I have set the spacing of the DRNL filters in the filterbank to follow the spacing of the Nobili et al. cochlear model – 300 filters between 20 Hz and 17 kHz. Table 2 shows ERB of the DRNL filters in the filterbank (measured in the filters with CF of 0.125, 0.25, 0.5, 1, 2 and 4 kHz).

Verhulst et al. model

Verhulst *et al.* [3] designed a tranmission line model (long-wave approximation) of the cochlea. The BM is modeled as an array of damped oscillators (with mass and stiffness) coupled via incompressible fluids in the cochlea. The model describes the traveling wave on the BM as a sum of backward- and forward-traveling waves. Shera *et al.* [11] showed that this wave-equation formulation is a different mathematical representation of the same cochlear mechanics as the hydrodynamic formulation used by Nobili *et al.*. The model simulates the active function of the cochlea by means of the approach described by Shera [12]. This approach allows to reach the level near-invariant impulse responses as was measured in the live mammalian cochlea. I have set the spacing of the oscillators to follow the spacing used in the Nobili et al. cochlear model – 300 oscillators between 20 Hz and 17 kHz. Table 3 shows ERB measured in discrete outputs (CF of 0.125, 0.25, 0.5, 1, 2 and 4 kHz) of the Verhulst et al. model.

Baumgarte model

Baumgarte [4] proposed a transmission line model of the cochlea. The model, as well as the Nobili et al. and Verhulst et al. models, simulates the BM as an array of damped oscillators (with mass and stiffness). The model is very similar to the Verhulst et al. model. It differs mainly in the approach used to simulate the active amplification of the outer hair cells (OHCs). It has two amplifiers. The first amplifier affects the BM vibrations. The second amplifier then further amplifies the BM vibrations without coupling back to the BM. However, there is no direct physiological evidence for the second amplifier. I have set spacing of the oscillators to follow the spacing used in the Nobili et al. cochlear model – 300 oscillators ranging between 20 Hz and 17 kHz. Table 4 shows ERB of the simulated cochlear filters (measured in the model outputs with CF of 0.125, 0.25, 0.5, 1, 2 a 4 kHz).

Experiments

The experiments employ behavioral data from three different studies [7, 8, 13]. The behavioral masking thresholds measured in the studies are compared with the thresholds predicted by an auditory model employing four different cochlear models: the Nobili et al., the DRNL cochlear model, the Verhulst et al., and the Baumgarte model.

Stimuli

Stimuli used in the experiments were constructed according to the description given in the perceptual studies [7, 8, 13]. The studies used harmonic complex tones as maskers and pure tones with frequency f_s as test tones. The maskers with fundamental frequency f_0 contained spectral components with a frequency ranging between $0.4f_s$ and $1.6f_s$. Four different configurations of the maskers and test tones were used.

$f_s = 1\,\mathrm{kHz}$, $f_0 = 100\,\mathrm{Hz}$: The configuration was adapted from [8, 13]. The fundamental frequency, f_0, of the masker was 100 Hz, the duration was 320 ms and it was ramped on and off with 10-ms raised-cosine ramps. The test tone with a frequency, f_s, of 1 kHz was temporally center within the masker, its duration was 260 ms and it was ramped on and off with 30-ms raised-cosine ramps. The overall level of the masker was 75 and 90 dB SPL.

$f_s = 1\,\mathrm{kHz}$, $f_0 = 50\,\mathrm{Hz}$: The configuration was adapted from [8]. The masker had $f_0 = 50\,\mathrm{Hz}$, duration of 320 ms and was ramped on and off with 30-ms raised-cosine ramps. The test tone with frequency $f_s = 1\,\mathrm{kHz}$ was temporally center within the masker, its duration was 260 ms and it was ramped on and off with 30-ms raised-cosine ramps. The overall level of the masker was 40, 60 and 85 dB SPL.

$f_s = 2\,\mathrm{kHz}$, $f_0 = 100\,\mathrm{Hz}$: The configuration was adapted from [7]. The masker had $f_0 = 100\,\mathrm{Hz}$ and the test tone had $f_s = 2\,\mathrm{kHz}$. The duration of the masker and test tone was 300 ms and it was ramped on and off with 30-ms raised-cosine ramps. The overall level of the masker was 50, 70 and 90 dB SPL.

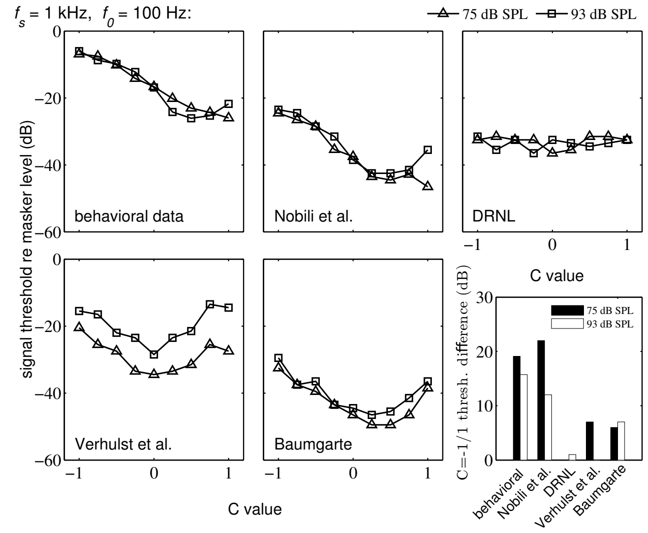

Figure 1: Top panel: Signal threshold re masker level as a function of the masker phase curvature (parameter C). Each panel shows the thresholds obtained by different means: listening tests, Nobili et al. model, DRNL model, Verhulst et al. model and Baumgarte model. The bar graph shows the difference between the thresholds for negative (C=-1) and positive (C=1) Schroeder phase maskers.

$\boldsymbol{f_s = 4\,\mathrm{kHz},\ f_0 = 100\,\mathrm{Hz}}$**:** The configuration was adapted from [8]. The masker had $f_0 = 100\,\mathrm{Hz}$, duration of 320 ms and was ramped on and off with 30-ms raised-cosine ramps. The test tone with frequency $f_s = 4\,\mathrm{kHz}$ was temporally center within the masker, its duration was 260 ms and it was ramped on and off with 30-ms raised-cosine ramps. The overall level of the masker was 40, 60 and 85 dB SPL.

Method of predictions

In order to predict the masking thresholds, I have extended the cochlear models by an inner hair cell (IHC) model, a model of auditory nerve (AN) synapse and a modulation filterbank. The output (between corresponding channels) from the modulation filterbank is compared using an optimal detector. The same models of the IHC and AN synapse, the modulation filterbank and the optimal detector were used in [9]. The optimal detector compares the model outputs in response to a masker plus a test tone and a masker only with a template. The template is calculated from the model output in response to a masker plus a suprathreshold (about 10 dB above threshold) test tone and a masker only. The optimal detector than decides if the test tone was detected. The masking thresholds were measured using a tracking algorithm. The smallest step size was 1 dB.

Results

$\boldsymbol{f_s = 1\,\mathrm{kHz},\ f_0 = 100\,\mathrm{Hz}}$**:** The behavioral and predicted masking thresholds are shown in Fig. 1. Each of the five largest panels show the masking thresholds relative to the masker level. Abscissa of the graphs represents the masker phase curvature (parameter C). The bar graph shows a difference between the thresholds for positive (C=-1) and negative (C=1) Schroeder phase maskers. Although these two maskers have approximately flat temporal envelopes, they produce different masking thresholds. Moreover the difference between the masking thresholds depends on the masker level. This is demonstrated in the bar graph in Fig. 1. The behavioral data were reproduced from [8] (75-dB masker) and from [13] (93-dB masker). The legends in Fig. 1 show the masker level.

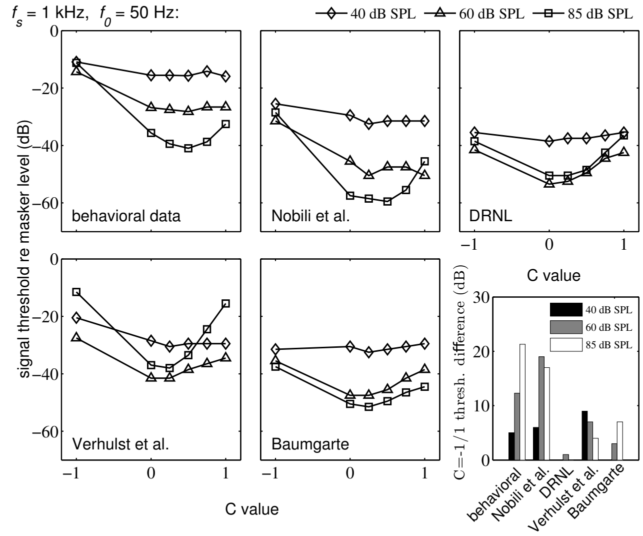

Figure 2: Same as in Fig. 1.

$\boldsymbol{f_s = 1\,\mathrm{kHz},\ f_0 = 50\,\mathrm{Hz}}$**:** Fig. 2 shows the behavioral and predicted masking thresholds. The behavioral data were reproduced from [8]. The panels show the same type of data as the panels in Fig. 1 (results of the previous experiment ($\boldsymbol{f_s = 1\,\mathrm{kHz},\ f_0 = 50\,\mathrm{Hz}}$**:**). The data were obtained using maskers with a level of 40, 60 and 85 dB SPL.

$\boldsymbol{f_s = 2\,\mathrm{kHz},\ f_0 = 100\,\mathrm{Hz}}$**:** The results are shown in Fig. 3. The behavioral data were reproduced from [7].

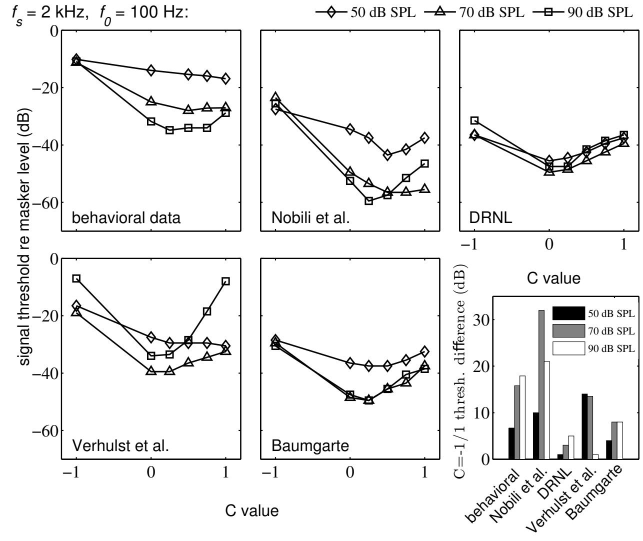

Figure 3: Same as in Fig 1.

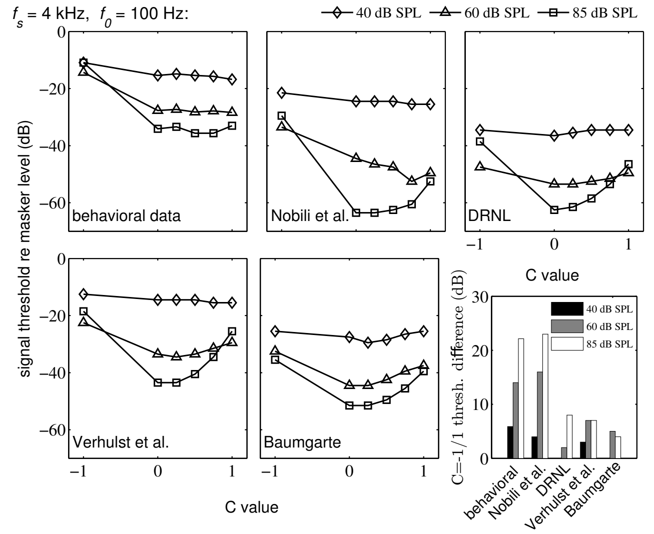

Figure 4: Same as in Fig. 1.

The panels show the same type of data as the panels in Fig. 1 and 2. The data were obtained using maskers with a level of 50, 70 and 90 dB SPL.

$f_s = 4\,\mathrm{kHz},\ f_0 = 100\,\mathrm{Hz}$:

The results are shown in Fig. 3. The behavioral data were reproduced from [8]. The panels show the same type of data as the panels in Fig. 1, 2 and 3. The data were obtained using maskers with a level of 40, 60 and 85 dB SPL.

Conclusion

The predicted thresholds largely differed from the behavioral data – none of the models showed a quantitative agreement with the behavioral data. The best qualitative agreement – for some of the maskers – was reached using the Nobili et al. cochlear model. The Nobili et al. model was the only model which predicted higher differences than about 13 dB between the thresholds for negative (C=-1) and positive (C=1) Schroeder phase maskers. Although, the predicted differences were for some maskers higher than for the behavioral data (Fig. 2 and Fig. 3). Both transmission line cochlear models (Verhulst et al. and Baumgarte) often showed smaller differences between thresholds for positive and negative Schroeder phase maskers which contrasts with behavioral data (see the bar graphs in the figures).

This study cannot distinguish if the disagreement with the behavioral data is caused by the used modeling approaches, by the model parameters or the used method (and parameters) of prediction.

Acknowledgements

Supported by the Institutional Endowment for the Long Term Conceptual Development of Research Institutes, as provided by the Ministry of Education of the Czech Republic in the year 2015: the "Sound quality" project at the Academy of Performing Arts in Prague

References

[1] Schroeder, M.R.: Synthesis of low peak-factor signals and binary sequences with low autocorrelation. IEEE Trans. Inf. Theory., 16 (1970) 85–89.

[2] Lopez-Poveda, E.A., Meddis R.: A human nonlinear cochlear filterbank. J. Acoust. Soc. Am. 110 (2001), 3107-3118

[3] Verhulst, S., Dau, T., Shera C.A.: Nonlinear time-domain cochlear model for transient stimulation and human otoacoustic emission. J. Acoust. Soc. Am. 132 (2012), 3842-3848

[4] Baumgarte, F. A Physiological Ear Model for Auditory Masking Applicable to Perceptual Coding. Proc. 103rd AES Convention, New York, (1997)

[5] Nobili, R., Vetešník, A., Turicchia, L., Mammano F.: Otoacoustic emissions from residual oscillations of the cochlear basilar membrane in a human ear model. J. Assoc. Res. Otolaryngol.4 (2003), 478-494

[6] Smith, B.K., Sieben U.K., Kohlrausch, A., Schroeder M.R.: Phase effects in masking related to dispersion in the inner ear. J. Acoust. Soc. Am. 80 (1987), 1631-1637

[7] Shen, Y., Lentz, J.J.: Level dependence in behavioral measurements of auditory-filter phase characteristics. J. Acoust. Soc. Am. 126 (2009), 2501–2510

[8] Oxenham, A., Dau, T.: Towards a measure of auditory-filter phase response. J. Acoust. Soc. Am., 110 (2001), 3169-3178

[9] Oxenham, A., Dau, T.: Reconciling frequency selectivity and phase effects in masking. J. Acoust. Soc. Am., 110 (2001), 1525-1538

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

[11] Shera, C.A., Tubis, A., Talmadge, C.L. Do Forward- and Backward-Traveling Waves Occur Within the Cochlea? Countering the Critique of Nobili et al. J. Assoc. Res. Otolaryngol., 5 (2004), 349-359

[12] Shera, C.A.: Intensity-invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics. J. Acoust. Soc. Am. 110 (2001), 332-348

[13] Oxenham, A.J., Dau, T.: Masker phase effects in normal-hearing and hearing-impaired listeners: Evidence for peripheral compression at low signal frequencies. J. Acoust. Soc. Am. 110 (2004), 2248-2257