Modeling the masking of tones by Schroeder-phase harmonic tone complexes

Eugen Rasumow[1], Martin Hansen[1] and Stephan Ewert[2]
[1] Institut für Hörtechnik und Audiologie, Fachhochschule OOW, D-26121 Oldenburg
[2] Medizinische Physik, Carl von Ossietzky Universität Oldenburg, D-26111 Oldenburg

24th March, 2009

I. Introduction

The amount of masking on a sinusoidal signal does not solely depend on the power but can also depend on the phase relationship of adjacent frequency components of the masker and the corresponding phase characteristic of the auditory filter. This effect is apparent in experiments where a sinusoidal signal is masked by a harmonic tone complex (*htc*) with variable Schroeder phase. Initial research shows a masking difference of more than 35 dB when altering the phase of the *htc*. This dynamic range of the measured data cannot be simulated when using a linear filter with optimized phase characteristics at the stage of the basilar membrane (*BM*) in a framework like desribed in [7] (see figure 1 and [8]). The consideration

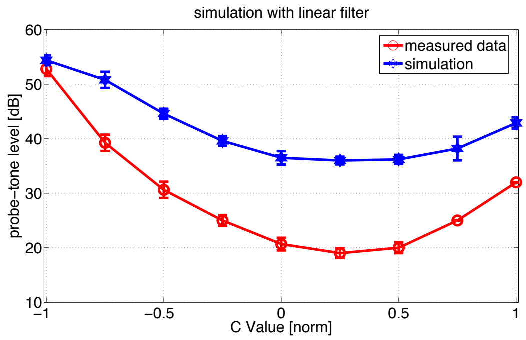

Figure 1: Measured masking (red line) and simulation (blue line) of a *htc* with $f_0 = 50$ Hz on a sine wave with $f_s = 2$ Hz. The curvature of the phase of the *htc*-masker is varied on the abscissa according to eq. (2). The simulation was accomplished using the perception model according to [7] and a linear *BM*-filter with optimized phase characteristics.

of compression might improve the simulations. However, current studies using the **D**ual **R**esonance **N**on-**L**inear (DRNL)-filters according to [2] have shown only little improvement of the simulations. This study deals with the refinement of the phase response of the *DRNL*-filters and its application in the framework of the computational model of human auditory signal processing and perception (*CASP*) according to [3]. The estimation of the phase response of the *BM* at the characteristic frequency (*CF*) of 2 kHz was accomplished according to Oxenham and Dau in [4]. The masker in these experiments were *htc* with modified Schroeder phase according to eq. (2).

$$htc = \frac{1}{N} \sum_{n=N_1}^{N_2} sin(2\pi n f_0 t + \theta(n)), \qquad (1)$$

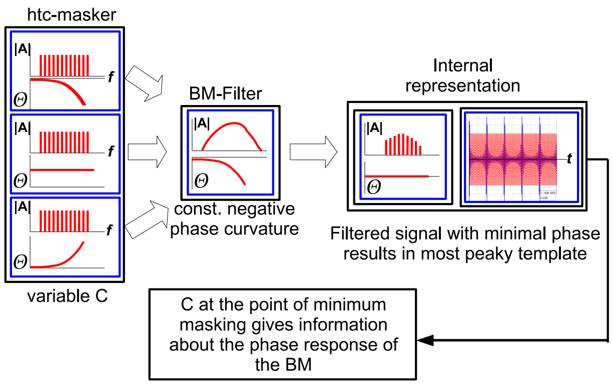

Figure 2: Paradigm for estimating the constant curvature of the phase response: Which curvature of the masker causes the most-peaky internal representation, i.e. the minimum in masking?

where f_0 is the fundamental frequency and n is the counter of the harmonic. The sum is bounded from the N_1th to the N_2th harmonic and N is the total number of frequency components. Each frequency component has its own starting phase $\theta(n)$.

$$\theta(n) = C \cdot \frac{\pi n(n-1)}{N}. \qquad (2)$$

The constant phase-curvature of the *htc* is given by

$$\frac{d^2\theta}{df^2} = C \cdot \frac{2\pi}{N f_0^2}. \qquad (3)$$

Modifying the variable C enables to vary sign and curvature of the phase, resulting in a variation of the temporal envelope of the masker. When using these *htc* as maskers, the variation of C between 1 and -1 can change the amount of masking over 35 dB as can be seen in figure 1. The current explanation is that the phase of the *htc* interacts with the phase response of the *BM* and produces an internal representation with a characteristic phase response. The fact that a *htc* with equal phase for all components results in the most-peaky envelope of the *htc* enables an estimation of the phase response of the *BM*. The curvature of the phase response of the *BM* is assumed to be constant inside the passband of a filter (see fig. 4 in [6]). Hence the constant phase-curvature of a certain *htc* can interact with the phase-curvature of the *BM* and cause a very peaky internal representation that corresponds to a minimum in masking in the measured data. This paradigm is illustrated in figure 2. The C_{min} of minimum masking can be converted into the negative phase-curvature of the *BM* via eq. (3). The curvature of the *htc*-masker at minimal masking has the opposite sign as the *BM* phase curvature it interacted with. Using this paradigm we can estimate the second derivative of the *BM* phase response with respect to frequency at a certain *CF*, i.e. the curvature of the phase response of an auditory filter. The knowledge of the basic function of the phase response as well as of the group delay (1st derivative of the phase response with respect to the frequency) is not necessary for simulating the dispersion due to the phase response of the *BM*. The group delay solely contains information about the delay of the system and thus is only essential for simulations with a fixed time reference. The absolute location of the phase response can be ignored since only the envelope of the signals is processed inside the framework. A variation of the absolute location of a phase response does not alter the envelope of a filtered signal (see [9]). The subsequent procedure is to estimate the phase-curvature of the *BM* via the explained experiment and to implement this phase response to an existing *DRNL*-filter. Furthemore, we investigate whether there is a distinctive correlation between the phase-curvature and the level of the masker.

II. Methods

II. A. Stimuli

The threshold of a sinusoidal signal with $f_s = 2$ kHz was measured in the presence of a *htc*[1] with an overall level of $L_{htc} = 60$, 75 and 80 dB SPL and fundamental frequencies of $f_0 = 25$, 50 and 100 Hz. The number of sinusoidal components ranged from 97 to 49 and 25, respectively. The procedure was accomplished according to [4]. The bandwidth of the *htc*-masker was 800 Hz $< f <$ 3200 Hz throughout the whole experminent. The duration of the maskers was set to 320 ms including 10 ms raised cosine on- and offset ramps. The total duration of the signal was 260 ms, including 30 ms raised cosine on- and offset ramps. The signal was temporally centered within the *htc*-masker. The starting phase of the probe-tone was randomized from trial to trial to avoid a constant superposition of the signal and a component of the *htc*-masker. The starting phases of the *htc*-masker was chosen according to eq. (2). A modification of the initial experimental procedure is the choice of the C (according to eq. (1)) of the *htc*-masker. Since a related research was accomplished before, the C of minimal masking (C_{min}) could be predicted approximately. Thus the C was chosen C = (-1 -0.75 -0.5 0 0.25 0.5 0.6 0.7 0.75 1) for the htc_{100}-masker, C = (-1 -0.75 -0.5 0 0.2 0.25 0.3 0.5 0.75 1) for the htc_{50}-masker and C = (-1 -0.75 -0.5 0 0.1 0.2 0.25 0.5 0.75 1) for the htc_{25}-masker. The stimuli were generated digitally in *MATLAB* and played out via a RME ADI 2 interface. All signals were converted with a 24-bit resolution and a sampling rate of 48 kHz. The playback in *MATLAB* at 24-bit resolution was accomplished with msound which is a kind of interface between *MATLAB* and the soundcard. It is available at `http://www.hoertechnik-audiologie.de/web/file/Forschung/Software.php#msound`. All signals were presented to the better ear of each subject through a Sennheiser HDA 200 headphone in a sound-attenuating booth.

[1] The fundamental frequency f_0 of a *htc* will be stated as index subsequently.

II. B. Procedure

The threshold of the sine wave probe-tone in presence of the *htc*-masker was measured using a 3 *AFC* method in conjunction with the 1-up 2-down tracking rule. This led to the point of 70,7 % correct on the psychometric function. Subsequent intervals were separated by 500 ms quietness. Each interval contained the *htc*-masker and only one randomly chosen interval contained the probe-tone. The task of the subject was to tell the interval including the probe-tone. Each experiment was accomplished using *Psylab*, version 2.1. *Psylab* is a collection of scripts in *MATLAB* that enables the procedure of any n-AFC experiments. It is available at `http://www.hoertechnik-audiologie.de/psylab/`. The initial step size was set to 8 dB and was bisected in value after each reversal. The minimal step size was 1 dB. The thresholds were measured at least three times in each condition. Thresholds around the minimum in masking were measured more often to ensure a more precise estimation of the phase curvature. Each subject accomplished the experiments in a number of sessions lasting approximately two hours.

II. C. Listeners

Three listeners (one female and two male) participated in the experiments. The listeners were 25 and 28 years old and had absolute hearing thresholds better than 15 dB HL in the frequency range between 125 Hz and 8 kHz. Only subject *nf* had no prior experience with psychoacoustic experiments. The subjects *nf* and *sm* were payed for their participance.

III. Results and discussion

III. A. Analysis of the measured data

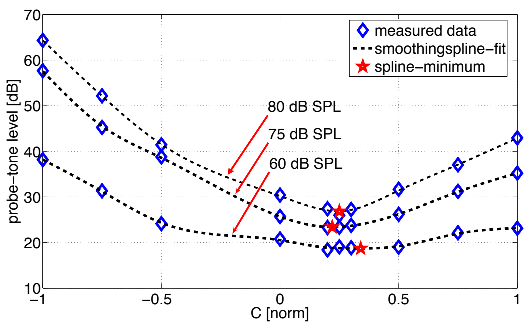

Figure 3: Data for masking with the htc_{50}-masker. The blue markers show the pooled data of all subjects and the black dashed line is the interpolating smoothing spline. The red star indicates the minimum of the interpolation for further interpretation.

htc-level [dB SPL]	C for minimal masking			
	htc_{100}	htc_{50}	htc_{25}	
60	0.65	0.34	0.04	$1.264\ 10^{-5}$
75	0.63	0.22	0.09	$1.215\ 10^{-5}$
80	0.55	0.25	0.17	$1.264\ 10^{-5}$
	$1.53\ 10^{-5}$	$1.38\ 10^{-5}$	$1.04\ 10^{-5}$	mean phase-curvature [rad/Hz^2]

Table 1: The minima C_{min} of the interpolated data and the corresponding phase curvature of the auditory filter for each condition.

The thresholds have a similar shape for all subjects in each condition. They only differ in their dynamic range and an absolute offset. Subsequently only the representative results and simulations for masking with htc_{50} will be illustrated. The experiments and simulations with other fundamental frequencies produce analog outcomes. Since it is assumed that the absolute location of the measured data is not affected by the phase dispersion on the basilar membrane the thresholds of all subjects were pooled. It is of great interest for the current research to estimate the C_{min} at minimal masking. Thus the pooled data were interpolated with a smoothing spline[2]. These curves for masking with a htc_{50}-masker are shown in figure 3. The pooled data for each condition or rather their interpolation were analysed concerning their minima. These minima can be converted into the phase curvature by interverting the sign of eq. (3). The pooled data indicate no characteristic relationship between phase-curvature and the level of the masker.

The minima and the corresponding curvature for each condition are listed in table 1. The corresponding mean phase curvature for all levels and all measured conditions is $\frac{d^2\theta}{df^2} \approx 1.32 \cdot 10^{-5}$ rad/Hz^2. This value correlates well with initial research results (see [8]).

III. B. Simulation of the measured data

The aim of the current research was to simulate the measured data in an appropriate way and to verify whether the consideration of compression allows for simulation of the measured dynamic range. A *DRNL*-filter according to [2], for instance, consists of 1st order Gammatone (GT) filters and 2nd order Butterworth-lowpass filters that have an improper phase response. Hence the phase characteristisc of the discrete filter stages was varied. The own variation of the *DRNL*-filter is illustrated in figure 4.

The filter-coefficients of the lowpass-filters were used unchanged, though the filtering was accomplished using the *filtfilt*-function in *MATLAB*. This function performs quasi zero-phase filtering by processing the input data in the forward and reverse directions. It should be considered that this function acts like a cascading of two filters. Thus the initial cascading should be halved when using the *filtfilt*-function. The desired phase characteristics were achieved by replacing the GT-filters with *IIR*-filters with similar magnitude response and optimized phase characteristics. The number of cascading of the GT-filters is two in the linear and four in the non-linear path. Therfore the phase characteristics of each single *IIR*-filter contained one halve respectively one quarter of the desired curvature.

[2]The choice of the interpolation is important for the estimation of the minimum C_{min}. The interpolation with a smoothing spline, as implemented in *MATLAB*, was chosen since this interpolation does not alter the characteristic shape of the measured data and results in consistent C_{min} for each condition.

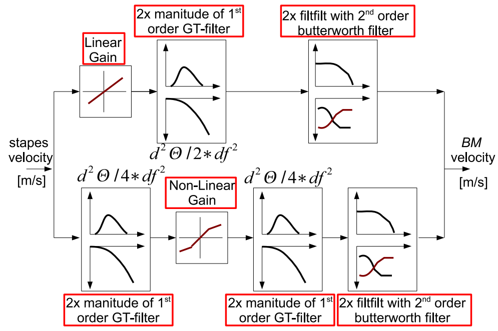

Figure 4: Own variation of the *DRNL*-filter. The refinement of the phase characteristics consist in the phase characteristic of the GT-filters. To avoid further influence of the phase characteristic of the lowpass filters, this stage is filtered backward and then forward using *filtfilt* in *MATLAB*.

This filter was included into the *CASP*-model according to [3]. All stages were kept as in [3] excluding from the *BM*-filter and the modulation frequency analysis-stage. Instead of using the whole modulation filterbank, in the following simulations only a lowpass-filter with a cut-off frequency of f_g = 8 Hz was used as in the perception model (see [7]). This was done because this constellation of the *CASP* allowed for better simulations of the measured data. The results of the simulations with a htc_{50}-masker using the *DRNL*-filter according to Jepsen et al. are depicted in figures 5. The results of the simulation for masking with htc_{50} using the own refined *DRNL*-filter are illustrated in figure 6.

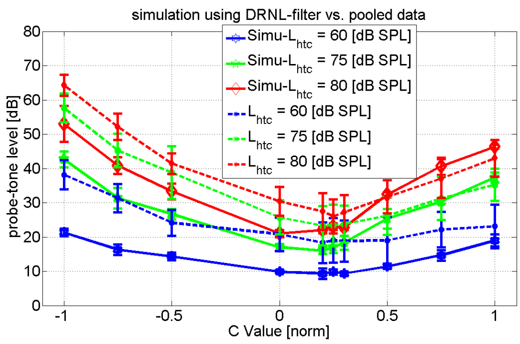

Figure 5: Comparison of the simulation results using the *DRNL*-filter according to Jepsen et al. with the pooled measured data. The dashed lines indicate the simulated data and the solid lines indicate the pooled data. The different colors indicate the level of the masker. Red lines correspond to 80 dB, green lines to 75 dB and blue line to 60 dB overal masker level.

III. C. Discussion of the simulated data

The use of the settings of the *CASP*-framework as described before enables for good simulations of the

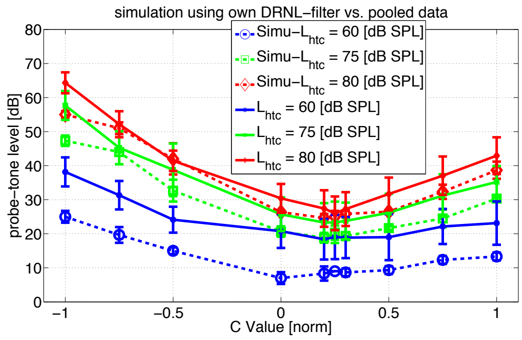

Figure 6: Comparison of the simulated data using the own refinement of the *DRNL*-filter with the pooled data. Dashed lines indicate simulated data and solid lines indicate the pooled data. Different colors represent different levels of the masker.

dynamic range of the measured data when using the own *DRNL*-filter as well as the *DRNL*-filter according to Jepsen et al.. Comparing the simulations using linear filters and *DRNL*-filters at the stage of the *BM* shows the necessity of consideration of compression in an auditory framework like the *CASP*. Although the dynamic range of the simulated data resembles the measured data well, the minimum in masking cannot be simulated using *DRNL*-filters that are composed of *GT*-filters and lowpass filters with inappropriate phase characteristics. This effect can be seen in the shifted minima as well as an improper bias of the simulated data. This effect becomes even more visible when the *DRNL*-filter according to [2] is applied (not shown in this article). The refinement of the phase characteristics of the own *DRNL*-filter yields a more precise simulation of the minima and the bias of the simulated data compared to the *DRNL*-filter according to Jepsen et al. (see figure 6).

In general the simulations approximate the measured data quite well when using higher levels but get less precise when maskers with lower levels are applied. This is evident for simulations with all applied fundamental frequencies whereas the dependence on the phase of the masker is still simulated quite well with the refined phase response of the own *DRNL*-filter. Possible causes for this absolute deviation could be an inacurate magnitude response of the used *DRNL*-filter or an inappropriate use of the modulation filterbank at the stage of modulation frequency analysis. A different cut-off frequency at this stage can alter the envelope of the internal representation and thus control the performance in the simulated experiments. Another reason might be, that the subjects did not reach their optimal-performance during the measurement period. If one compares the simulation to the data of the most experienced subject *elr*, for instance, this absolute offset disappears for higher levels. Thus a longer measurement period might bear a solution for this problem.

IV. Summary

The subjects in the accomplished experiments show a common dependence on the phase of the masking *htc*. A variation of C according to eq. (2) effects a variation of the threshold of the probe-tone of more than 35 dB. The shapes of the measured thresholds and the minima in masking are similar for all subjects. The measured data show no consistent relationship between phase curvature and level of the masker throughout all subjects. The interpolation of the pooled data lead to a mean phase curvature of $\frac{d^2\ \theta}{df^2} \approx 1.32 \cdot 10^{-5}$ rad/Hz^2. The consideration of compression at the *BM*-stage of the *CASP* enabled good simulations of the dynamic range of the measured data. The own modification of the *DRNL*-fliter with optimized phase charateristics enabled the simulation of the dynamic range of the measured data *as well as* the point of minimal masking.

The simulation of masking with a *htc* with lower levels produces data with an absolute deviation. Hence, further investigations concerning the modulation frequency analysis-stage in conjunction with a refinement of the phase characteristics of the *DRNL* are necessary.

References

[1] Andrew J. Oxenham and Stephan D. Ewert, Estimates of auditory filter phase response at and below characteristic frequency (L), J. Acoust. Soc. Am. 117 (4):1713-1716, 2005

[2] Enrique A. Lopez-Poveda and Ray Meddis. A human nonlinear cochlear filterbank. J. Acoust. Soc. Am., 110(6):3107-3118, 2001.

[3] Morten L. Jepsen, Stephan D. Ewert and Torsten Dau, A computational model of human auditory signal processing and perception J. Acoust. Soc. Am., 124(1):422-438, 2008

[4] Andrew J. Oxenham and Torsten Dau, Towards a measure of auditory-filter phase response. J. Acoust. Soc. Am., 110(6):3169-3177, 2001

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

[6] Christopher A. Shera, Frequency glides in click responses of the basilar membrane and auditory nerve: Their scaling behavior and origin in traveling-wave dispersion, J. Acoust. Soc. Am. 109 (5), 2001

[7] Dau, T., Püschel, D. und Kohlrausch, A. (1996). A quantitative model of the effective signal processing in the auditory system. I.Model structure. J. Acoust. Soc. Am., 99(6):3615-3622.

[8] Eugen Rasumow and Martin Hansen, Übertragungsfunktion der Basilarmembran, DAGA 2008, Dresden; 673-674

[9] Fettweis, A. (1977). On the Significance of Group Delay in Communication Engineering. International Journal of Electronics and Communication, 31(9):342-348.