An auditory model that explains masking by Schroeder-phase complexes

Akira Nishimura

Department of Media and Cultural Studies, Faculty of Informatics, Tokyo University of Information Sciences,
1200-2, Yato-cho, Wakaba-ku, Chiba, 265-8501 Japan

(Received 25 June 2001, Accepted for publication 9 July 2001)

Keywords: Temporal window, Gammachirp filter, Masking period pattern, Simultaneous masking
PACS number: 43.66.Ba, 43.66.Dc, 43.66.Mk, 43.66.Nm

1. Introduction

Phase effects in masking experiments using harmonic complex maskers are usually associated with strong variations in the masker envelope. Representative examples are positive and negative Schroeder-phase complexes. A Schroeder-phase complex with a fundamental frequency f_0 in Hz is given by:

$$S(t) = \sum_{m=n_1}^{n_h} \sin(2\pi m f_0 t + \text{sign} \times \text{phase}(m)) \quad (1)$$

$$\text{phase}(m) = \pi m(m-1)/(n_h - n_1 + 1) \quad (2)$$

where n_1 is the lowest harmonic number and n_h is the highest harmonic number. When the *sign* is positive, $S(t)$ represents a positive Schroeder phase complex (POS), and, when the *sign* is negative, it represents a negative Schroeder phase complex (NEG). POS and NEG are time-reversed and inverted versions of each other. Both phase relationships produce very flat envelopes. However, the masked thresholds of a pure tone of frequency and phase identical to a component of the Schroeder-phase masker vary by about 20 dB between POS and NEG maskers.

In the present study we propose an auditory model that quantitatively explains the dependence of the masked threshold on the phase spectra of the harmonic complex maskers. This model is demonstrated to be capable of accurately predicting the masked thresholds reported by other researchers.

2. Masking experiments

Smith *et al.* [1] measured the masked thresholds of a pure tone of frequency and phase identical to a Schroeder-phase masker component under conditions of controlled fundamental frequencies of masker and maskee frequencies. Kohlrausch and Sander [2] reported results for a sine-phase masker in addition to replicated experiments of Smith *et al.* Both Smith and Kohlrausch discussed that the phase response of the auditory filter transforms the flat envelope of the input into an output containing marked peaks and dips, and that this leads to a substantial threshold variation between POS and NEG maskers. These previous reports, however, did not include clear quantitative analysis of these threshold variations.

Carlyon and Datta [3] measured masking period patterns with respect to the sound pressure level, number of components, and phase of flanking components of Schroeder-phase maskers. They showed that several features in the data could be simulated using a nonlinear model of the

auditory periphery as described by Giguère and Woodland [4].

Nishimura [5] focused on the characteristics of the phase function of Schroeder-phase complexes, deriving the phase function

$$\text{phase}(m) = \sum_{k=1}^m (k\theta\pi/180) \quad (3)$$

where θ in degrees is regarded as a parameter that controls the delay time of each component. Nishimura showed that the masked thresholds of 50-Hz band noise from positive and negative phase masker are dependent on θ .

3. Masking model

The proposed auditory model consists of a series of modules that each simulate specific functions of the auditory periphery. A detector is connected to the output of the last module. The modules are as follows: a fixed filter for outer/middle ear response, a Gammachirp filter bank for auditory filters [6], a half-wave rectifier, a direct current (DC) component adder, and a symmetric rounded-exponential temporal integrator with equivalent rectangular duration of 2 ms. The half-wave rectifier represents a simple approximation of the sum of responses of inner hair cells. The DC component adder is intended to simulate the absolute threshold of a pure tone corresponding to the center frequency of the auditory filter. The main feature of this model is the phase response of the Gammachirp filters, emphasizing peaks in the POS masker waveform and flattening the envelope of the NEG masker waveform.

At the detection stage, the absolute differences between the running output levels of the temporal integrator for each channel of the masker+signal waveforms and the masker alone (SMR: Signal+Masker to Masker Ratio) are calculated. The maximum SMR (MAXSMR) is regarded as an index of detection of a short signal in the measurement of masking period patterns. The mean SMR (MEANSMR) is regarded as an index of detection of a continuous signal in simultaneous masking conditions. In both cases, the signal is considered to be detected in the frequency channel exhibiting the largest MAXSMR or MEANSMR.

An initial MAXSMR or MEANSMR value is required for practical masking simulation, and is calculated using a signal of masked threshold level as model input under arbitrary experimental conditions. The smallest masked threshold condition is usually selected for obtaining the initial

value; in the present simulation, the initial values of MAXSMR are in the range 6 to 8 dB, and MEANSMMR is 0.4 to 1.1 dB. The simulated masked thresholds under other conditions are then determined by searching signal levels that give the same SMMR value as the initial result.

The most remarkable feature of this model compared to Carlyon and Datta's model is that this model requires only a small number of parameters that can be estimated through psychoacoustic experiments, such as Gammachirp filter parameters that can be obtained from notched noise masking data [6].

4. Comparison of masking data and predictions by model

A rough search of Gammachirp filter parameters b and c was conducted in order to fit the results of the model simulations to the masking data obtained from the experiments reviewed in section 2. Figures 1 to 4 show the masking data in the left column and the results of the model simulations in the right column, with parameter values b and c . All data points are expressed relative to the level of a single masker component. The signal frequency is 1,100 Hz, except for the experiment in Fig. 1, and the fundamental frequency of the masker is 100 Hz, except for the experiment in Fig. 2.

The model with $b \approx 1$ is a close simulation of the

Fig. 1 Smith *et al.* (1986) data and results of model.

Fig. 2 Kohlrausch and Sander (1995) data and results of model.

Fig. 3 Carlyon and Datta (1997) data and results of model.

Fig. 4 Nishimura (1999) data and results of model.

tendencies of the masking data, and of other experimental data not displayed here. The model parameter c must be twice the value obtained from the fitting to the notched noise data [6] in order to provide a similar approximation. A large negative c is required in order to obtain the appropriate negative curvature of the phase function for the Gammachirp filter, which produces pronounced peaks and dips in the POS output and a flattened-envelope in the NEG output. However a large negative c also causes the amplitude spectrum of the filter to have a steeper slope at high frequencies region and a shallower slope at low frequencies.

There is less agreement between the data and the model in the large θ region ($\theta \geq 64$) in Fig. 4. As the masker waveforms in this region differ considerably from the Schroeder-phase complexes, cues other than the amplitude envelope will need to be employed for signal detection in this region.

References

- [1] B. K. Smith, U. K. Sieben, A. Kohlrausch and M. R. Schroeder, "Phase effects in masking related to dispersion in the inner ear", *J. Acoust. Soc. Am.*, **80**, 1631 (1986).
- [2] A. Kohlrausch and A. Sander, "Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets", *J. Acoust. Soc. Am.*, **97**, 1817 (1995).
- [3] R. P. Carlyon and A. J. Datta, "Masking period patterns of Schroeder-phase complexes: Effects of level, number of components, and phase of flanking components", *J. Acoust.*

- Soc. Am.*, **101**, 3648 (1997).
- [4] C. Giguère and P. C. Woodland, “A computational model of the auditory periphery for speech and hearing research. I. Ascending path”, *J. Acoust. Soc. Am.*, **95**, 331 (1994).
- [5] A. Nishimura, “Changes in masked thresholds of a narrowband noise by initial phase of the harmonic complex masker”, *Proc. Spring Meet. Acoust. Soc. Jpn.*, 3-2-1, 399 (1999).
- [6] T. Irino and R. D. Patterson, “A time-domain, level-dependent auditory filter: The gammachirp”, *J. Acoust. Soc. Am.*, **101**, 412 (1997).