A Computational Model of Co-modulation Masking Release

Masashi UNOKI and Masato AKAGI

6 Feb. 1998
IS-RR-98-0006P

School of Information Science
Japan Advanced Institute of Science and Technology, Hokuriku
Asahidai 1–1, Tatsunokuchi
Nomi, Ishikawa, 923-12, JAPAN
unoki@jaist.ac.jp, akagi@jaist.ac.jp

ISSN 0918-7553

A Computational Model of Co-modulation Masking Release

Masashi UNOKI and Masato AKAGI
School of Information Science,
Japan Advanced Institute of Science and Technology
1-1 Asahidai, Tatsunokuchi, Ishikawa-ken, 923-1292 Japan

Abstract

This paper proposes a computational model of co-modulation masking release (CMR). It consists of two models, our auditory segregation model (model A) and the power spectrum model of masking (model B), and a selection process that selects one of their results. Model A extracts a sinusoidal signal using the outputs of multiple auditory filters and model B extracts a sinusoidal signal using the output of a single auditory filter. For both models, simulations similar to Hall *et al.*'s demonstrations were carried out. Simulation stimuli consisted of two types of noise masker, bandpassed random noise and AM bandpassed random noise. It was found that in model A, the signal threshold decreases as the masker bandwidth increases for AM bandpassed noise. In contrast, in model B, the signal threshold increases as the masker bandwidth increases up to 1 ERB and then it remains constant for both noises. The selection process selects the sinusoidal signal with the lowest signal threshold from the two extracted signals. As a result, the signal threshold of the pure tone extracted using the proposed model shows the similar properties to Hall *et al.*'s demonstrations. The maximum amount of CMR in the proposed model is about 8 dB.

Key words: auditory scene analysis, two acoustic source segregation, gammatone filter, Co-modulation masking release (CMR)

I. Introduction

In investigations for frequency selectivity of the auditory system, the power spectrum model of masking [Patterson *et al.*, 1986] is widely accepted to explain the phenomenon of masking. In this model, it is assumed that when a listener tries to detect a sinusoidal signal with a particular center frequency amid background noise, he makes use of the output of a single auditory filter having a center frequency close to the signal frequency and having the highest signal-to-masker ratio. In addition, it is assumed that the stimuli are represented by long-term power spectra, and that the masking threshold for the sinusoidal signal is determined by the amount of noise passing through the auditory filter. With these assumptions, the power spectrum model can explain masking phenomena such as simultaneous masking. However, this model cannot explain all masking phenomena be-

cause the relative phases of the components and the short-term fluctuations in the masker are ignored.

In 1984, Hall *et al.* have demonstrated that across-filter comparisons could enhance the detection of a sinusoidal signal in a fluctuating noise masker [Hall *et al.*, 1984]. The crucial feature for achieving this enhancement was that the fluctuations should be coherent or correlated across different frequency bands. They called this across-frequency coherence in their demonstrations "co-modulation." Therefore, the enhancement in signal detection obtained using coherent fluctuation, i.e., this phenomenon of reduced masking threshold, was called "Co-modulation Masking Release (CMR)". Many psychoacoustical experiments were carried out [Moore , 1997,Moore , 1992,Willen *et al.*, 1997] and the same phenomenon was demonstrated repeatedly. The condition when CMR can occur was revealed, but a less computational model using this condition was proposed.

On the other hand, we have been tackling the problem of segregating the desired signal from noisy signal based on auditory scene analysis (ASA) [Bregman , 1990, Bregman , 1993]. We stress the need to consider not only the amplitude spectrum but also the phase spectrum when attempting to completely extract the signal from a noise-added signal which both exist in the same frequency region [Unoki *et al.*, 1997]; based on this stance, we seek to solve the problem of segregating two acoustic sources — the basic problem of acoustic source segregation using regularities (ii) and (iv) of the following regularities: (i) common onset and offset; (ii) gradualness of change; (iii) harmonicity; and (iv) changes occurring in an acoustic event [Bregman , 1993].

This paper proposes a computational model of CMR that consists of two models, our auditory segregation model and the power spectrum model of masking proposed by Patterson *et al.*, and a selection process.

II. Computational model of CMR

Our computational model of CMR is shown in Fig. 1. This model consists of two models (A and B) and a selection process. In this model, it is assumed that $f_1(t)$ is a sinusoidal signal (pure tone) and $f_2(t)$ is two types of noise masker (bandpassed random noise and AM bandpassed random noise) whose center frequency is the same as the signal frequency. It is also assumed that the localized $f_1(t)$ is added to $f_2(t)$. Since the proposed model can observe only mixed signal $f(t)$, it can extract a sinusoidal signal $f_1(t)$ using two models (A and B). Model A is the auditory segregation model we proposed [Unoki *et al.*, 1997]. Model B is the power spectrum model of masking [Patterson *et al.*, 1986]. We consider that in the computational model of CMR these two models work in parallel and extract a sinusoidal signal from the masked signal. Here, let $\hat{f}_{1,\mathrm{A}}(t)$ and $\hat{f}_{1,\mathrm{B}}(t)$ be the sinusoidal signals extracted using models A and B, respectively. The fundamental idea arises from the fact that the masking threshold increases as the masker bandwidth increases up to the bandwidth of the signal auditory filter (1 ERB) and then it either remains constant or decreases depending on the coherency of fluctuations. In other words, model B can explain part of CMR by using the output of a single auditory filter for the case that the masker bandwidth increases up to 1 ERB, and Model A can explain part of CMR by using the outputs of multiple auditory filter for the case that the masker bandwidth exceeds over 1 ERB.

III. Model A: Auditory segregation model

The auditory segregation model, shown in Fig. 2, consists of three parts: (a) an auditory filterbank, (b) separation block, and (c) grouping block. The auditory filterbank is constructed using a gammatone filter as an "analyzing wavelet." The separation block uses physical constraints related to heuristic regularities (ii) and (iv) proposed by Bregman [Bregman , 1993]. The grouping block synthesizes each separated parameter and then reconstructs the extracted signal using the inverse wavelet transform. These blocks are described in the next section.

A. Auditory filterbank

In model A, an auditory filterbank is constructed using the wavelet transform, where the basic function is the gammatone filter. The impulse response of the gammatone filter [Patterson *et al.*, 1991] is given by

$$gt(t) = At^{N-1}e^{-2\pi b_f t}\cos(2\pi f_0 t), \qquad t \geq 0, \tag{1}$$

where $\mathrm{ERB}(f_0) = 24.7(4.37f_0/1000 + 1)$ and $b_f = 1.019\mathrm{ERB}(f_0)$. To determine phase information, we extend the impulse response of the gammatone filter, which is a basic wavelet. This basic wavelet is represented by

$$\psi(t) = At^{N-1}e^{j2\pi f_0 t - 2\pi b_f t}, \tag{2}$$

using the Hilbert transform. Then, an auditory filterbank is constructed using the wavelet transform.

$$\tilde{f}(a,b) = \frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty} f(t)\overline{\psi\left(\frac{t-b}{a}\right)}dt, \tag{3}$$

$$f(t) = \frac{1}{D_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{f}(a,b)\psi\left(\frac{t-b}{a}\right)\frac{dadb}{a^2}, \tag{4}$$

where a is the "scale parameter," b is the "shift parameter," and $a, b \in \mathbf{R}$ with $a \neq 0$. In addition, function $\bar{\psi}$ is the conjugate of ψ. Here, since $\phi(t)$ is a complex basic wavelet, the integral wavelet transform can be represented by

$$\tilde{f}(a,b) = |\tilde{f}(a,b)|e^{j\arg(\tilde{f}(a,b))}, \tag{5}$$

where $|\tilde{f}(a,b)|$ is the amplitude spectrum and $\arg(\tilde{f}(a,b))$ is the phase spectrum.

Finally, an auditory filterbank is designed with a center frequency f_0 of 1 kHz, a band-passed region from 100 Hz to 10 kHz, and number of filters K of 128. This auditory filterbank is implemented, using the discrete wavelet transform with the following conditions: sampling frequency $f_s = 20$ kHz, the scale parameter $a = \alpha^p, -\frac{K}{2} \leq p \leq \frac{K}{2}, \alpha = 10^{2/K}$, and the shift parameter $b = q/f_s$, where $p, q \in \mathbf{Z}$. This is a constant Q filterbank whose center frequency is 1 kHz; the bandwidth of the auditory filter is 1 ERB, as shown in Fig. 3. In addition, we compensate for the group delay by adjusting the peak in the envelopes of Eq. (2) for all scale parameters, which is called "alinement processing," because the group delay occurs for each scale, as predicted from the impulse response of Eq. (2).

B. Separation and Grouping

Model A treats the problem of segregating two acoustic sources as follows:

First, we can observe only the signal $f(t)$:

$$f(t) = f_1(t) + f_2(t), \tag{6}$$

where $f_1(t)$ is the desired signal and $f_2(t)$ is a noise masker. The observed signal $f(t)$ is decomposed into its frequency components by an auditory filterbank. Second, outputs of the k-th channel, which correspond to $f_1(t)$ and $f_2(t)$, are assumed to be

$$f_1(t) : A_k(t)\sin(\omega_k t + \theta_{1k}(t)) \tag{7}$$

and

$$f_2(t) : B_k(t)\sin(\omega_k t + \theta_{2k}(t)). \tag{8}$$

Here, ω_k is the center frequency of the auditory filter and $\theta_{1k}(t)$ and $\theta_{2k}(t)$ are the input phases of $f_1(t)$ and $f_2(t)$, respectively. Since the output of the k-th channel $X_k(t)$ is the sum of Eqs. (7) and (8), it is represented by

$$X_k(t) = S_k(t)\sin(\omega_k t + \phi_k(t)). \tag{9}$$

Therefore, the amplitude envelopes of the two signals $A_k(t)$ and $B_k(t)$ can be determined by

$$A_k(t) = \frac{S_k(t)\sin(\theta_{2k}(t) - \phi_k(t))}{\sin\theta_k(t)} \tag{10}$$

and

$$B_k(t) = \frac{S_k(t)\sin(\phi_k(t) - \theta_{1k}(t))}{\sin\theta_k(t)}, \tag{11}$$

where $\theta_k(t) = \theta_{2k}(t) - \theta_{1k}(t)$ and $\theta_k(t) \neq n\pi, n \in \mathbf{Z}$. Since the amplitude envelope $S_k(t)$ and the output phase $\phi_k(t)$ are observable (See the next section.), then if the input phases $\theta_{1k}(t)$ and $\theta_{2k}(t)$ are determined, $A_k(t)$ and $B_k(t)$ can be determined by the above equations. Finally, all the components are synthesized from Eqs. (7) and (8) in the grouping block. Then $f_1(t)$ and $f_2(t)$ can be reconstructed by the grouping block using the inverse wavelet transform. Here, $\hat{f}_{1,\mathrm{A}}(t)$ and $\hat{f}_{2,\mathrm{B}}(t)$ are the reconstructed $f_1(t)$ and $f_2(t)$, respectively.

In this paper, we assume that $f_1(t)$ is a sinusoidal signal, $f_2(t)$ is a noise masker, and the center frequency of the auditory filter corresponds to the signal frequency. Therefore, we consider the problem of segregating $f_1(t)$ from $f(t)$ when $\theta_{1k}(t) = 0$ and $\theta_k(t) = \theta_{2k}(t)$.

C. Calculation of the four physical parameters

In this section we calculate the four physical parameters amplitude envelope $S_k(t)$, output phase $\phi_k(t)$, and input phases $\theta_{1k}(t)$ and $\theta_{2k}(t)$.

The amplitude envelope $S_k(t)$ is calculated by

$$S_k(t) = |\tilde{f}(\alpha^{k-\frac{K}{2}}, t)|, \tag{12}$$

where $|\tilde{f}(a,b)|$ is the amplitude spectrum [Unoki *et al.*, 1997]. The output phase $\phi_k(t)$ is calculated by

$$\phi_k(t) = \int \left(\frac{d}{dt} \arg \left(\tilde{f}(\alpha^{k-\frac{K}{2}}, t) \right) - \omega_k \right) dt, \tag{13}$$

where $\arg(\tilde{f}(a,b))$ is the phase spectrum [Unoki *et al.*, 1997].

In this paper, we assume $\theta_{1k}(t) = 0$. Since $\theta_k(t) = \theta_{2k}(t) - \theta_{1k}(t)$, we must find the input phase $\theta_{2k}(t)$. It can be determined by applying three physical constraints, derived from regularities (ii) and (iv), as follows.

Firstly, we use regularity (ii), which is the gradualness of change. This regularity means that "a single sound tends to change its properties smoothly and slowly (gradualness of change)" [Bregman , 1993]. We consider the term "slowly" in this regularity as the following physical constraint, in order to apply it to the amplitude envelope $A_k(t)$.

Physical constraint 1 *Temporal differentiation of the amplitude envelope $A_k(t)$ must be represented by an R-th-order differentiable polynomial $C_{k,R}(t)$ as follows:*

$$\frac{dA_k(t)}{dt} = C_{k,R}(t) \tag{14}$$

□

Applying **Physical constraint 1** to Eq. (10), we get a linear differential equation, which we solve to get a general solution of the input phase $\theta_{2k}(t)$:

$$\theta_{2k}(t) = \arctan \left(\frac{S_k(t) \sin \phi_k(t)}{S_k(t) \cos \phi_k(t) + C_k(t)} \right), \tag{15}$$

where $C_k(t) = -\int C_{k,R}(t)dt + C_{k,0}$. The $C_k(t)$ is called the "unknown function."

Therefore, if $C_k(t)$ is determined, then $\theta_{2k}(t)$ is uniquely determined by Eq. (15). Although it is possible to estimate the coefficients $C_{k,r}, r = 0, 1, \cdots, R$ by considering this problem as an optimization problem, we assume that, in small segment Δt, $C_{k,R}(t) = C_{k,0}$. Therefore this means that Eq. (14) is equivalent to $dA_k(t)/dt = 0$ and that the amplitude envelope $A_k(t)$ does not fluctuate in small segment Δt.

Next, we use regularity (ii) to segregate each small segment Δt. Regularity (ii) means that "each physical parameter must retain temporal proximity in the bound ($t = T_r$) between pre-segment ($T_r - \Delta t \leq t < T_r$) and post-segment ($T_r \leq t < T_r + \Delta t$)" for this regularity to apply to physical parameters. This is considered in the following physical constraint.

Physical constraint 2 *In the bound ($t = T_r$) between pre-segment and post-segment, each of the physical parameters $A_k(t)$, $B_k(t)$, and $\theta_{2k}(t)$ must be connected within ΔA, ΔB, and $\Delta\theta$, respectively. That is,*

$$\begin{aligned} |A_k(T_r + 0) - A_k(T_r - 0)| &\leq \Delta A \\ |B_k(T_r + 0) - B_k(T_r - 0)| &\leq \Delta B \\ |\theta_{2k}(T_r + 0) - \theta_{2k}(T_r - 0)| &\leq \Delta\theta. \end{aligned} \tag{16}$$

□

From Eqs. (10), (11), and (15), the amplitude envelopes, $A_k(t)$ and $B_k(t)$, and the input phase $\theta_{2k}(t)$ are functions of the unknown coefficient. Therefore, by considering the above relationships, we can interpret **Physical constraint 2** in order to determine $C_{k,0}$, which is restricted within

$$C_{k,\alpha} \leq C_{k,0} \leq C_{k,\beta}, \tag{17}$$

where $C_{k,\alpha}$ and $C_{k,\beta}$ are the upper-limited and lower-limited $C_{k,0}$ in the bound between the two segments.

Finally, we apply regularity (iv), which means that "many changes take place in an acoustic event that affect all the components of the resulting sound in the same way and at the same time" [Bregman , 1993]. This regularity is considered as the following physical constraint.

Physical constraint 3 *The normalized amplitude envelope of the output of the kth channel must be approximately equal to that of the ℓth channel as follows:*

$$\frac{B_k(t)}{\|B_k(t)\|} \approx \frac{B_{k\pm\ell}(t)}{\|B_{k\pm\ell}(t)\|}, \qquad \ell = 1, 2, \cdots, L. \tag{18}$$

□

Here, a masker envelope $B_k(t)$ is a function of $C_{k,0}$ from Eqs. (11) and (15), and let $\hat{B}_k(t)$ be a masker envelope $B_k(t)$ determined by any $C_{k,0}$. We consider the **physical constraint 3** to select an optimal coefficient $C_{k,0}$ when the correlation between $B_k(t)$ and $B_{k\pm\ell}(t)$ becomes maximum at any $C_{k,0}$ within the region of Eq. (17), as follows:

$$\max_{C_{k,\alpha} \leq C_{k,0} \leq C_{k,\beta}} \frac{\langle \hat{B}_k, \hat{\hat{B}}_k \rangle}{||\hat{B}_k|| ||\hat{\hat{B}}_k||}, \tag{19}$$

where $\hat{B}_k(t)$ is the masker envelope given by any $C_{k,0}$, and $\hat{\hat{B}}_k(t)$ is the masker envelope given by the outputs of the $k \pm \ell$th auditory filters as follows:

$$\hat{\hat{B}}_k(t) = \frac{1}{2L} \sum_{\ell=-L, \ell\neq 0}^{L} \frac{\hat{B}_{k+\ell}(t)}{\|\hat{B}_{k+\ell}(t)\|}. \tag{20}$$

Hence, the above computational process can be summarized as follows: (a) a general solution of the input phase $\theta_{2k}(t)$ is determined using physical constraint 1; (b) a candidate of $C_{k,0}$ that can uniquely determine $\theta_{2k}(t)$, is determined using physical constraint 2; (c) an optimal $C_{k,0}$ is determined using physical constraint 3; and (d) $\theta_{2k}(t)$ can be uniquely determined by the optimal $C_{k,0}$.

The algorithm of model A is shown in Fig. 4 and the relationship between amplitude envelopes in auditory filters is shown in Fig. 5. $\hat{\hat{B}}_{k+\ell}(t)$ is determined using amplitude characteristics as shown in Fig. 5, and using procedures (1)–(5) in Fig. 4.

In this paper, we consider the problem of segregating a sinusoidal signal in the masked signal in which the localized $f_1(t)$ is added to $f_2(t)$. Therefore, when we solve the above problem using the proposed method, we must know the duration, which is two acoustic signals exist in the same time region. The duration can be determined by detecting the onset and offset of $f_1(t)$. In Fig. 4, by focusing on the temporal deviation of $S_k(t)$ and $\phi_k(t)$, we can determine onset $T_{k,\text{on}}$ and offset $T_{k,\text{off}}$ of $f_1(t)$ as follows:

1. Onset $T_{k,\text{on}}$ is determined by the nearest maximum point of $|\frac{d\phi_k(t)}{dt}|$ (within 25 ms) to the maximum point of $|\frac{dS_k(t)}{dt}|$.

2. Onset $T_{k,\text{off}}$ is determined by the nearest maximum point of $|\frac{d\phi_k(t)}{dt}|$ (within 25 ms) to the minimum point of $|\frac{dS_k(t)}{dt}|$.

The segregated duration is $T_{k,\text{off}} - T_{k,\text{on}}$.

IV. Model B: the power spectrum model of masking

In the power spectrum model [Patterson *et al.*, 1986], it is assumed that when a listener is trying to detect a sinusoidal signal with a particular center frequency in a background noise, he uses of the output of a single auditory filter whose center frequency is close to the signal frequency, and which has the highest signal-to-masker ratio. Therefore, it can be considered that the component passed through the single auditory filter only affects masking. In particular the masking threshold for a sinusoidal signal is determined by the amount of noise passing through the auditory filter.

The power spectrum model consists of Model B as shown in Fig. 6. The output of the auditory filter $X_k(t)$ is one of the outputs of the auditory filterbank. This filter consists of the gammatone filter whose center frequency is 1 kHz and bandwidth is 1 ERB. In this model, the sinusoidal signal $\hat{f}_{1,\text{B}}(t)$ extracted from the masked signal $f(t)$ is the output of the single auditory filter $X_k(t)$.

V. Simulations

A. Co-Modulation Masking Release

Hall *et al.* measured the masking threshold for a sinusoidal signal in one of their experiments, in which the center frequency was 1 kHz and the duration was 400 ms, as a function of the bandwidth of a continuous noise masker, keeping the spectrum level constant [Hall *et al.*, 1984]. They used two types of masker, which were both centered at 1 kHz, as follows:

- A random noise masker: This had irregular fluctuations in amplitude, and the fluctuations in different frequency regions were independent.

- An amplitude modulated random noise masker: This was a random noise that was modulated in amplitude at an irregular, show rate; a noise lowpass filtered at 50 Hz was used as a modulator. Therefore, fluctuations in the amplitude of the noise in different spectral regions were the same.

This across-frequency coherence was called "co-modulation" by them.

Fig. 7 shows the results of that experiment. For the random noise (denoted by R), the signal threshold increased as the masker bandwidth increased up to about 100-200 Hz, and then remained constant. This is exactly as expected from the traditional model of masking. The auditory filter at this center frequency had a bandwidth of about 130 Hz. Hence, for noise bandwidths up to about 130 Hz, increasing the bandwidth the

filter increased the noise passing through the filter, so the signal threshold increased. In contrast, increasing the bandwidth beyond 130 Hz did not increase the noise passing through the filter, so the threshold did not increase. The pattern for the modulated noise (denoted by M) was quite different. For noise bandwidths greater than 100 Hz, the signal threshold decreased as the bandwidth increased. This indicates that subjects could compare the outputs of different auditory filters to enhance signal detection. The fact that the decrease in threshold with increasing bandwidth only occurred with modulated noise indicates that fluctuations in the masker are critical and that the fluctuations need to be correlated across frequency bands. Hence, this phenomenon has been called "co-modulation masking release (CMR)." The amount of CMR in that experiment, defined as the difference in thresholds for random noise and modulated noise, was at most about 10 dB [Moore , 1997].

In this paper, our simulation conditions were the same as the conditions in the above experiment. Simulations of the proposed model were done to examine whether the model can simulate the property of CMR.

B. Simulations for Model A

1. Stimuli

To consider conditions equivalent to the experimental ones used by Hall *et al.*, in this simulation we assume that $f_1(t)$ was a sinusoidal signal , where a center frequency was 1 kHz, duration was 400 ms and the amplitude envelope was constant, and that $f_2(t)$ was two types of bandpassed noise masker having center frequency close to the signal frequency. In addition, we adjust the bandwidths of the auditory filters, which is equivalent to the masker bandwidth, in stead of the two maskers were made by fixing the masker bandwidth to 1 kHz. One was a bandpassed random noise $f_{21}(t)$ and other was an AM bandpassed random noise $f_{22}(t)$. This masker was amplitude modulated $f_{21}(t)$, where the modulation frequency was 50 Hz and the modulation rate was 100%. Here, the power of the noise masker $f_2(t)$ was adjusted so that $\sqrt{f_{21}(t)^2/f_{22}(t)^2} = 1$. Moreover the power ratio between $f_1(t)$ and $f_2(t)$, i.e., the SNR (signal-to-noise ratio), was -6.6 dB.

In this simulation, the mixed signals were $f_R(t) = f_1(t) + f_{21}(t)$ and $f_M(t) = f_1(t) + f_{22}(t)$, corresponding to the stimuli labeled R and M, respectively. Simulation stimuli consisting of 10 sinusoidal signals were formed by varying the onset and 30 maskers of two types were formed by varying random seeds. Thus, the total number of stimuli was 300. As an example, one of the two types of mixed signals is shown in Fig. 8. Here, a sinusoidal signal $f_1(t)$ is masked visually in the all-mixed signal, but we can hear the sinusoidal signal from $f_M(t)$ because of the CMR; however, we cannot hear the sinusoidal signal from $f_R(t)$ because of the masking.

2. Conditions and procedure

In this paper, we set the parameters for $\Delta t = 3/(f_0 \cdot \alpha^{k-\frac{K}{2}})$, $\Delta A = |A_k(T_r - \Delta t) - A_k(T_r - 2\Delta t)|$, $\Delta B = 0.01 S_{\max}$, and $\Delta\theta = \pi/20$, where $S_{\max}$ is the maximum of $S_k(t)$.

In their demonstration of CMR, Hall *et al.* measured the masking threshold as a function of the masker bandwidth. Our simulation conditions can be considered to be the same as the experimental ones used by Hall *et al.* since we measured the SNR of the extracted sinusoidal signal $\hat{f}_{1,A}(t)$ as a function of the number of adjacent auditory filters

L, which is equivalent to the masker bandwidth, where the masker bandwidth is fixed. Therefore, $\theta_{2k}(t)$ is uniquely determined by the amplitude envelope $\hat{\hat{B}}_k(t)$ as a function of L from Eqs. (15), (19), and (20). The bandwidths related to $L = 1$, 3, 5, 7, 9, 11 are 207, 352, 499, 648, 801, 958 Hz, respectively.

C. Results and discussion

Simulations were carried out according the conditions mentioned above. The results are shown in Fig. 10, where the vertical and horizontal axes show the improved SNR of the extracted sinusoidal signal $\hat{f}_{1,A}(t)$ and the bandwidth related to L, respectively. Moreover, the real line and the error bar show the mean and standard deviation of the SNR of the signal $\hat{f}_{1,A}(t)$ extracted from 300 mixed signals, respectively. It was found that for the mixed signal $f_M(t)$, a sinusoidal signal $\hat{f}_{1,A}(t)$ became detectable as the number of the adjacent auditory filters L increased, but for the mixed signal $f_R(t)$, $\hat{f}_{1,A}(t)$ was not detectable as L increased. Therefore, the results show that a sinusoidal signal is more detectable when the components of the masker have the same amplitude modulation pattern in different frequency regions or when the fluctuations in the masker envelopes are coherent. Hence, model A simulates the phenomenon of reduction from masking using the outputs of multiple auditory filters.

D. Simulations for Model B

1. Stimuli

These simulations assumed that $f_1(t)$ was the same 10 sinusoidal signals as those used as the stimuli in model A and that $f_2(t)$ was 45 bandpassed random noise maskers of two types formed by varying random seeds (five types) and by varying the bandwidth (nine types). Thus, the total number of stimuli was 450. The masker bandwidths were 33, 67, 133, 207, 352, 499, 648, 801, and 958 Hz. Three of these bandwidths were related to 1/4, 1/2, and 1 ERB, respectively. The remainder were bandwidths related to L.

2. Condition and procedure

In model B, in order to measure the masking threshold as a function of the masker bandwidth, we measure the SNR of the sinusoidal signal $\hat{f}_{1,B}(t)$ extracted for the masking threshold as a function of the masker bandwidth.

3. Results and discussion

Simulations were carried out according to the above mentioned conditions. The results are shown in Fig. 11, where the vertical and horizontal axes show the improved SNR of the extracted sinusoidal signal $\hat{f}_{1,B}(t)$ and the masker bandwidth, respectively. Moreover, the real line and the error bar show mean and standard deviation of the SNR, respectively. It was found that the SNR for the extracted sinusoidal signal $\hat{f}_{1,B}(t)$ increased as the masker bandwidth increased, independent on the type of masker. In particular, as the masker bandwidth increased up to 1 ERB the masking threshold (SNR) increased as a function and then remained constant. Hence, model B simulates the phenomenon of simultaneous masking, using the output of a single auditory filter.

E. Considerations for Computational model of CMR

The results of simulations for the two models show that model A simulates the phenomenon of CMR/simultaneous masking by coherence/incoherence between the fluctuations of amplitude envelope of a masker when the masker bandwidth increases above 1 ERB. Moreover, model B simulates the phenomenon of simultaneous masking in which the threshold increases as a function of the masker bandwidth as the masker bandwidth increases up to 1 ERB and then the threshold remains constant. The selection process therefore selects the lowest of these masking thresholds. In other words, it selects the highest SNR of the signal extracted from $\hat{f}_{1,A}(t)$ and $\hat{f}_{1,B}(t)$, and let $\hat{f}_1(t)$ be the extracted signal with the highest SNR. Thus, from Figs. 10 and 11 the proposed model has the characteristics of the masking threshold shown in Fig. 12. In the selection process, the extracted signal with the lowest threshold is selected from the signals extracted using the two models. These characteristics show that the phenomenon of CMR is similar to Hall *et. al.*'s results. Hence, it can be interpreted that the proposed model is a computational model of CMR. The maximum amount of CMR in Hall *et al.*'s demonstrations was about 10 dB. In contrast, the maximum amount of CMR in our model was about 8 dB.

VI. Conclusions

In this paper, we proposed a computational model of co-modulation masking release (CMR). This model consists of two models, our auditory segregation model (model A) and the power spectrum model of masking (model B), and a selection process that selects one of their results. The mechanisms for extracting a sinusoidal signal from a masked signal work as follows: model A uses the outputs of multiple auditory filters and model B uses the output of a single auditory filter. Simulations of the two models were carried out using two types of noise masker, the same as Hall *et al.*'s demonstration conditions, bandpassed random noise and AM bandpassed random noise. In model A, the signal threshold decreased depending on the type of masker and the masker bandwidth. In the case of bandpassed random noise, the signal threshold did not vary as the masker bandwidth increased. In contrast, for AM bandpassed noise, the signal threshold decreased as the masker bandwidth increased. In model B, the signal threshold increased as the masker bandwidth increased up to 1 ERB and then remained constant for both noise maskers. The selection process then selected the highest SNR from the sinusoidal signals extracted from the results of the two models. As a result, the characteristics of the proposed model show that the phenomenon of CMR is similar to Hall *et al.*'s results. The maximum amount of CMR in the proposed model was about 8 dB.

Hence, the proposed model can be interpreted as a computational model of CMR. It was also shown that regularity (iv) is one clue to CMR.

References

A. S. Bregman. (1990). Auditory Scene Analysis: The Perceptual Organization of Sound, MIT Press, Cambridge, Mass.

A. S. Bregman. (1993). "Auditory Scene Analysis: hearing in complex environments," in Thinking in Sounds, (Eds. S. McAdams and E.Bigand), pp. 10–36, Oxford University Press, New York.

Brain C.J. Moore. (1997). An Introduction to the Psychology of Hearing, 4th ed., Academic Press, San Diego.

Brain C.J. Moore. (1992). "Comodulation Masking release and Modulation Discrimination Interface," in The Auditory Processing of Speech, from Sound to Words, (Edited by M. E. H. Schouten), pp. 167–183, Mouton de Gruyter, NewYork.

C. K. Chui. (1992). An Introduction to Wavelets, Academic Press, Boston, MA.

Hall, J. W. and Fernandes, M. A. (1984). "The role of monaural frequency selectivity in binaural analysis," J. Acoust. Soc. Am. 76, 435–439.

Hall, J. W. and Grose, J. H. (1988). "Comodulation masking release: Evidence for multiple cues," J. Acoust. Soc. Am. 84, pp. 1669–1675.

Patterson, R. D. and Moore, B. C. J. (1986). Auditory filters and excitation patterns as representations of frequency resolution. In Frequency Selectivity in Hearing (ed. B. C. J. Moore), Academic Press, London and New York.

Patterson, R. D. and John Holdsworth. (1991). A Functional Model of Neural Activity Patterns and Auditory Images, Advances in speech, Hearing and Language Processing, vol. 3, JAI Press, London.

Masashi Unoki and Masato Akagi. (1997). "A Method of Signal Extraction from Noise-Added Signal," IEICE, Vol. J80-A, No. 3, pp. 444–453, March (in Japanese).

Willen A. C. van den Brink, Tammo Houtgast, and Guido F. Smoorenburg. (1992). "Effectiveness of Comodulation Masking Release," in The Auditory Processing of Speech, from Sound to Words, (Eds. M. E. H. Schouten), pp. 167–183, Mouton de Gruyter, NewYork.

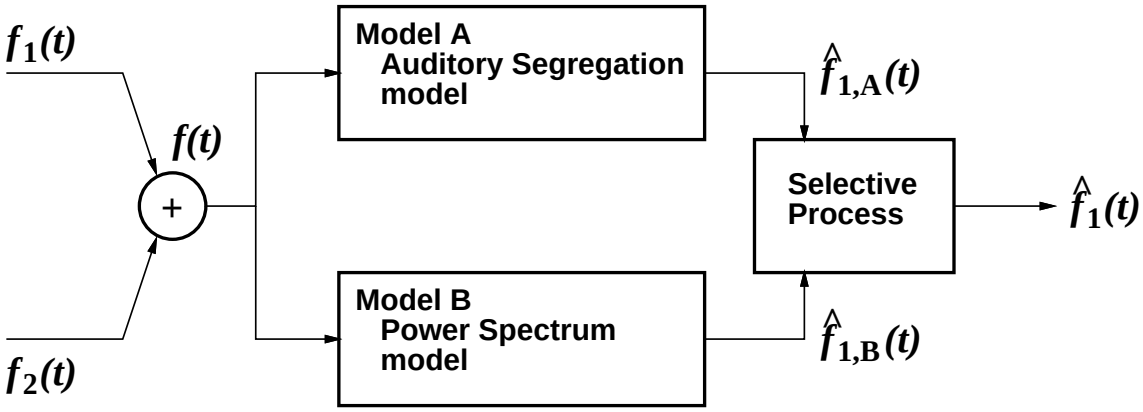

FIG. 1. Computational model of CMR. This model consists of two models, our auditory segregation model (model A) and the power spectrum model of masking (model B), and a selection process that selects one of their results.

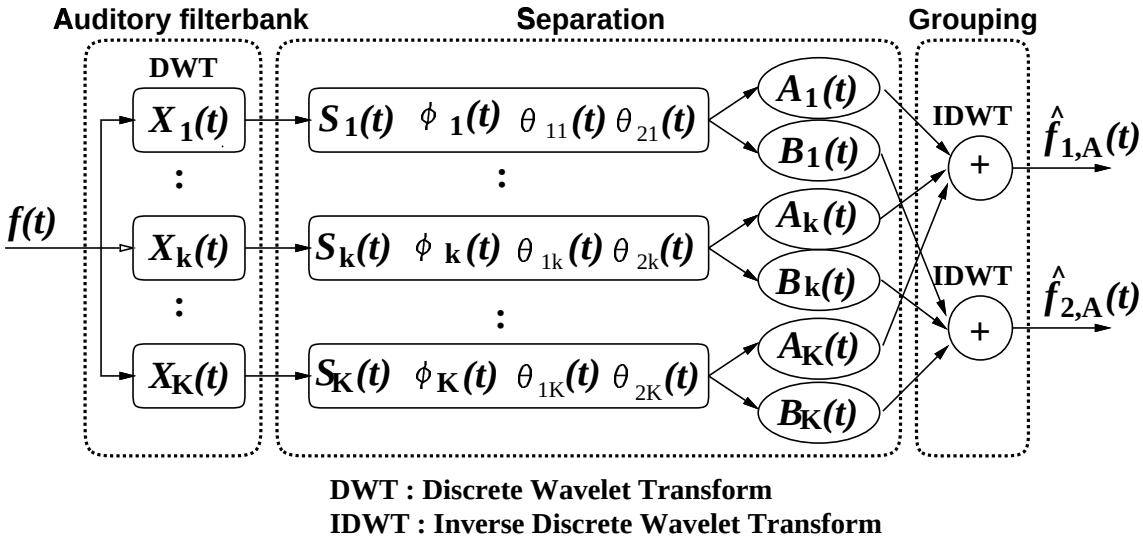

FIG. 2. Model A: an auditory segregation model. This model consists of three parts: (a) an auditory filterbank, (b) separation block, and (c) grouping block.

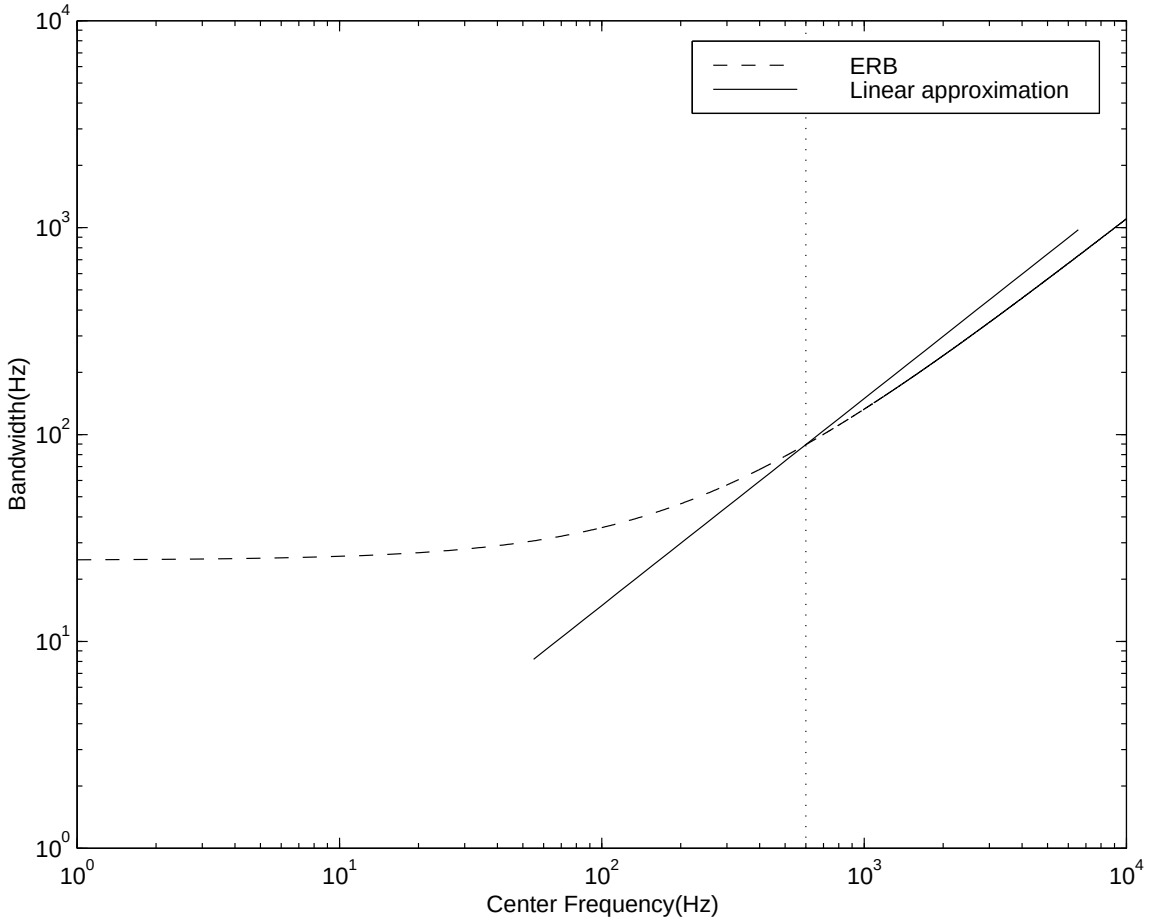

FIG. 3. Relationship between center frequency and ERB. Dashed-line shows ERB corresponding to the center frequency and solid-line shows linear approximation of ERB at 600 Hz.

decompose $f(t)$ into its frequency components using the auditory filterbank from Eq. (9);
for $k := 1$ **to** K **do**
determine $S_k(t)$ and $\phi_k(t)$ from Eqs. (12) and (13);
detect onset T_{on} and offset T_{off} from $dS_k(t)/dt$ and $d\phi_k(t)/dt$;
let the segregated duration be $T_{k,\text{on}} \leq t \leq T_{k,\text{off}}$;
split the segregated duration into N segments $\Delta t = M/f_0$;
for $n := 1$ **to** N **do**
determine $C_{k,\alpha} \leq C_{k,0} \leq C_{k,\beta}$;
for $C_{k,0} := C_{k,\alpha}$ **to** $C_{k,\beta}$ **do**
determine $\hat{\theta}_k(t)$ for $C_{k,0}$;
determine $\hat{A}_k(t)$ and $\hat{B}_k(t)$;
In adjacent auditory filters(Fig.5), $(\ell = 1, 2, \cdots, L)$;
(1) determine $\hat{A}_{k\pm\ell}(t)$ from amplitude characteristics of adjacent filters;
(2) determine $S_{k\pm\ell}(t)$ and $\phi_{k\pm\ell}(t)$ from Eqs. (12) and (13);
(3) determine $\hat{\theta}_{k\pm\ell}(t)$ from Eq. (15), using $\hat{A}_{k\pm\ell}(t)$, $S_{k\pm\ell}(t)$, and $\phi_{k\pm\ell}(t)$;
(4) determine $\hat{B}_{k\pm\ell}(t)$ from Eq. (11);
(5) determine $\hat{\hat{B}}_k(t)$ form Eq. (20);
(6) determine a correlation $\text{Corr}(\hat{B}_k(t), \hat{\hat{B}}_k(t))$ from Eq. (19);
end
determine unknown parameter $C_{k,0}$ from Eq. (19) within $C_{k,\alpha} \leq C_{k,0} \leq C_{k,\beta}$
determine $\theta_k(t)$ from Eq. (15);
determine $A_k(t)$ from Eq. (10);
determine $B_k(t)$ from Eq. (11);
end
determine components from Eqs. (7) and (8);
end
reconstruct $\hat{f}_1(t)$ and $\hat{f}_2(t)$ using the wavelet filterbank (inverse wavelet transform) from Eqs. (10) and (11);

FIG. 4. Segregation algorithm.

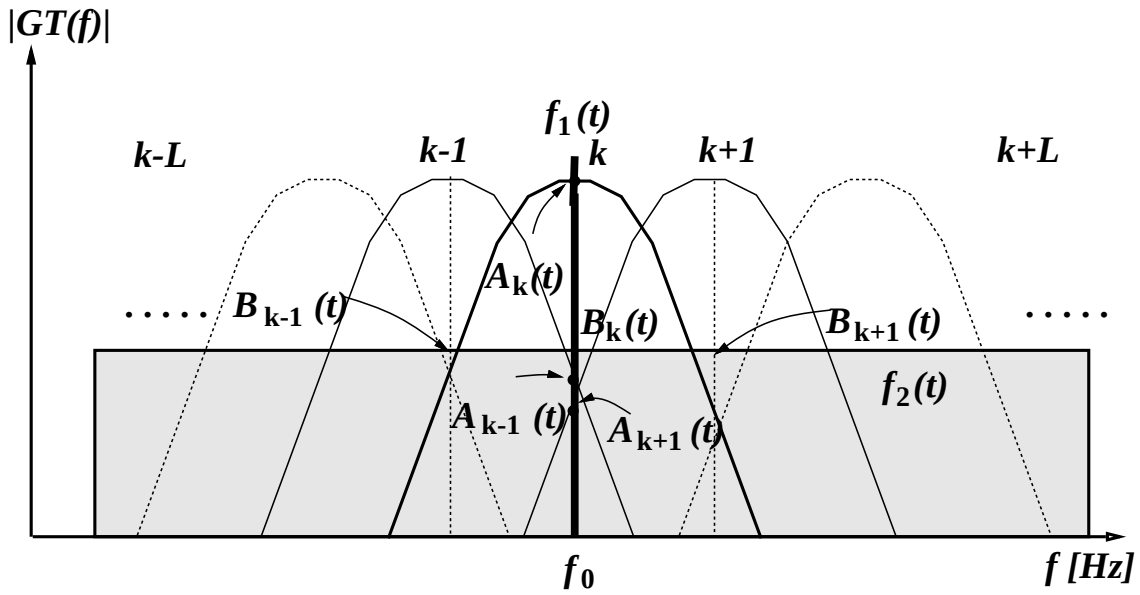

FIG. 5. Bandpassed characteristics of a sinusoidal signal $f_1(t)$ and bandpassed noise $f_2(t)$ in the adjacent channel.

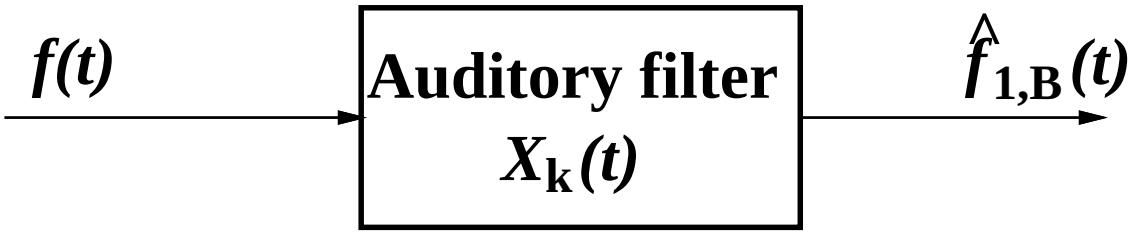

FIG. 6. Model B: a power spectrum model of masking.

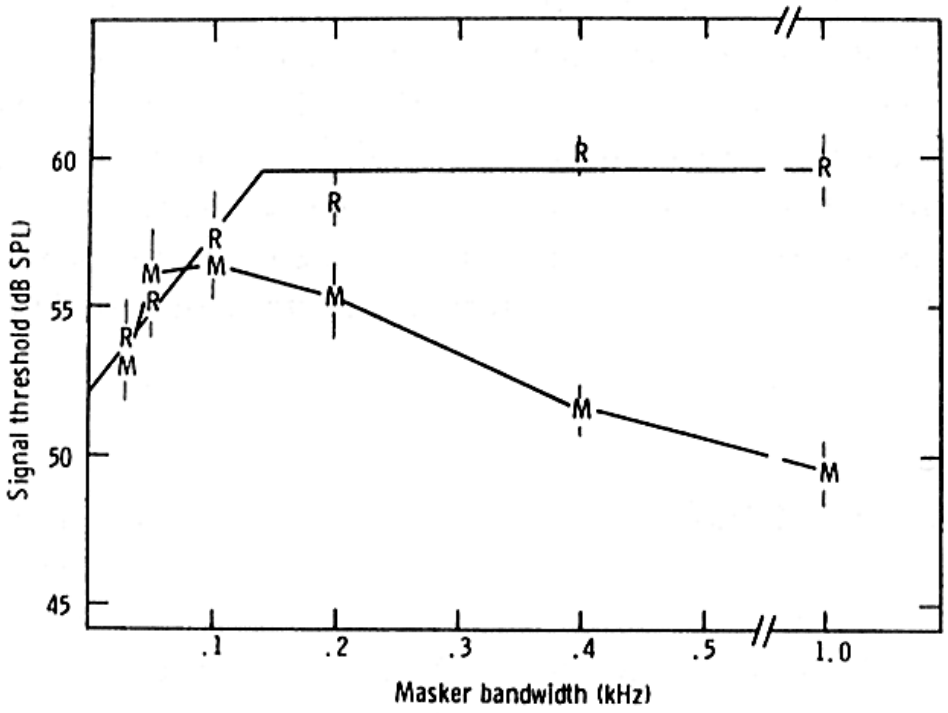

FIG. 7. Results of CMR (Hall *et al.*, 1984). The points labeled 'R' are thresholds for 1 kHz signal centered in a band of random noise, plotted as a function of the bandwidth of the noise. The points labeled 'M' are the thresholds obtained when the noise was amplitude modulated at an irregular, low rate.

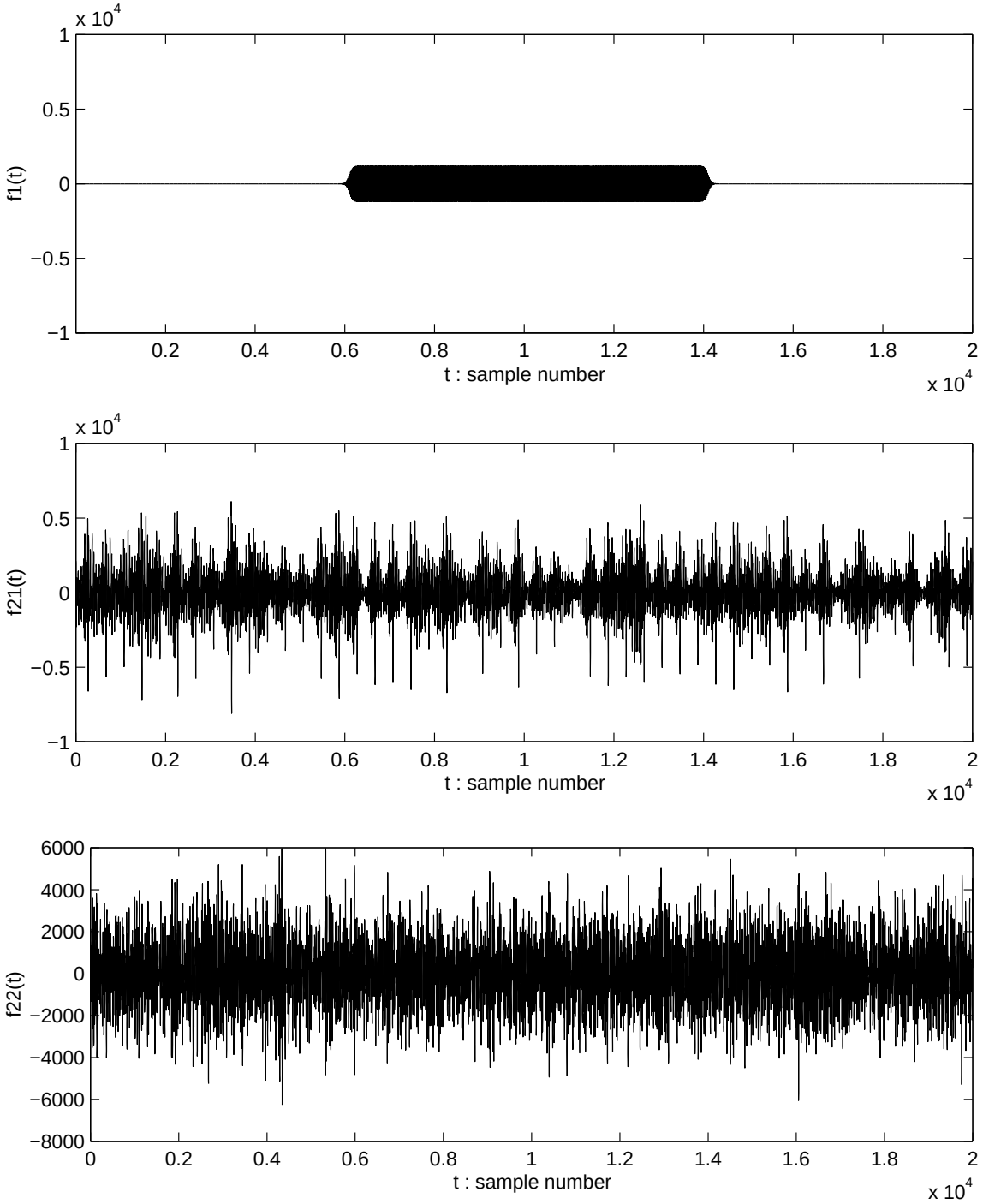

FIG. 8. Stimuli: a sinusoidal signal $f_1(t)$ (top), a bandpassed random noise $f_{21}(t)$ (middle), and an AM bandpassed noise $f_{22}(t)$ (bottom).

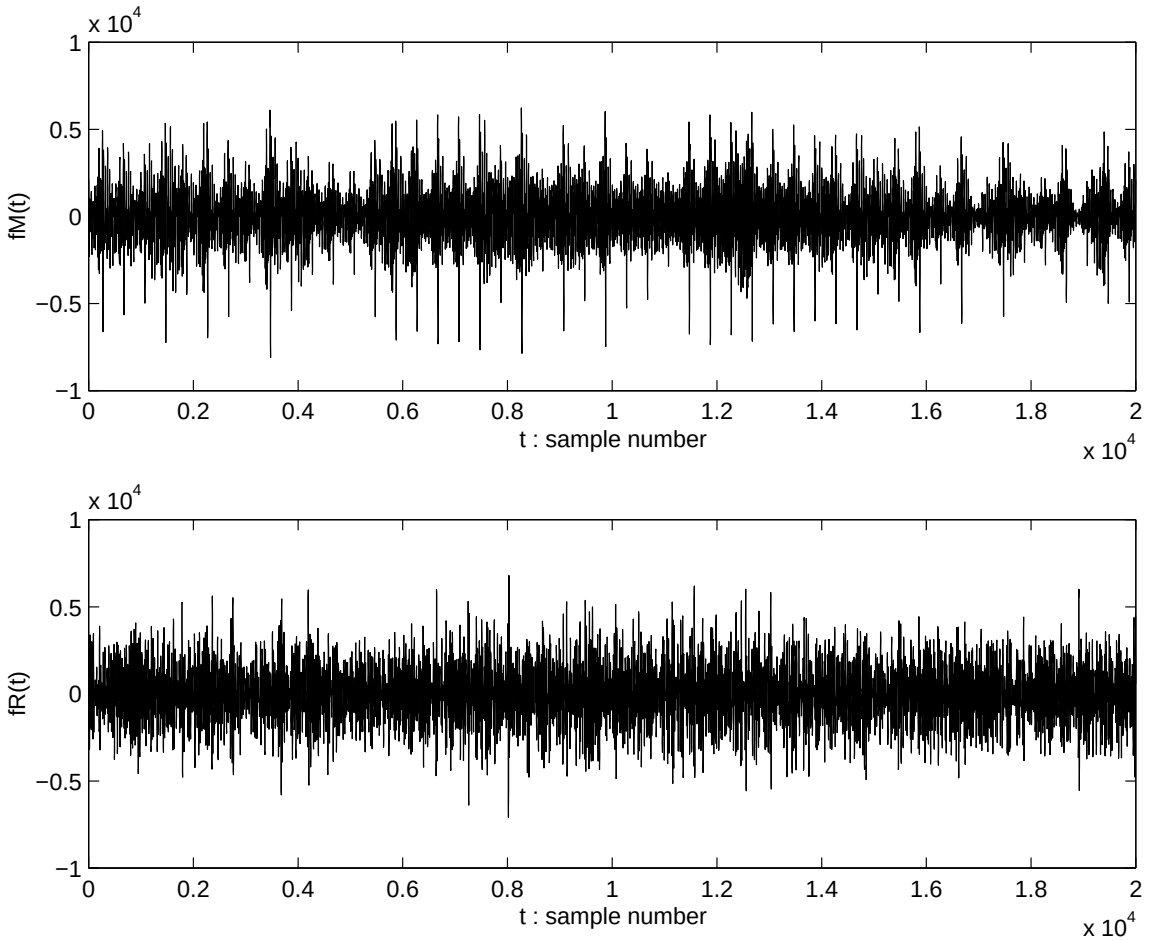

FIG. 9. Mixed signals $f_M(t)$ (top) and $f_R(t)$ (bottom).

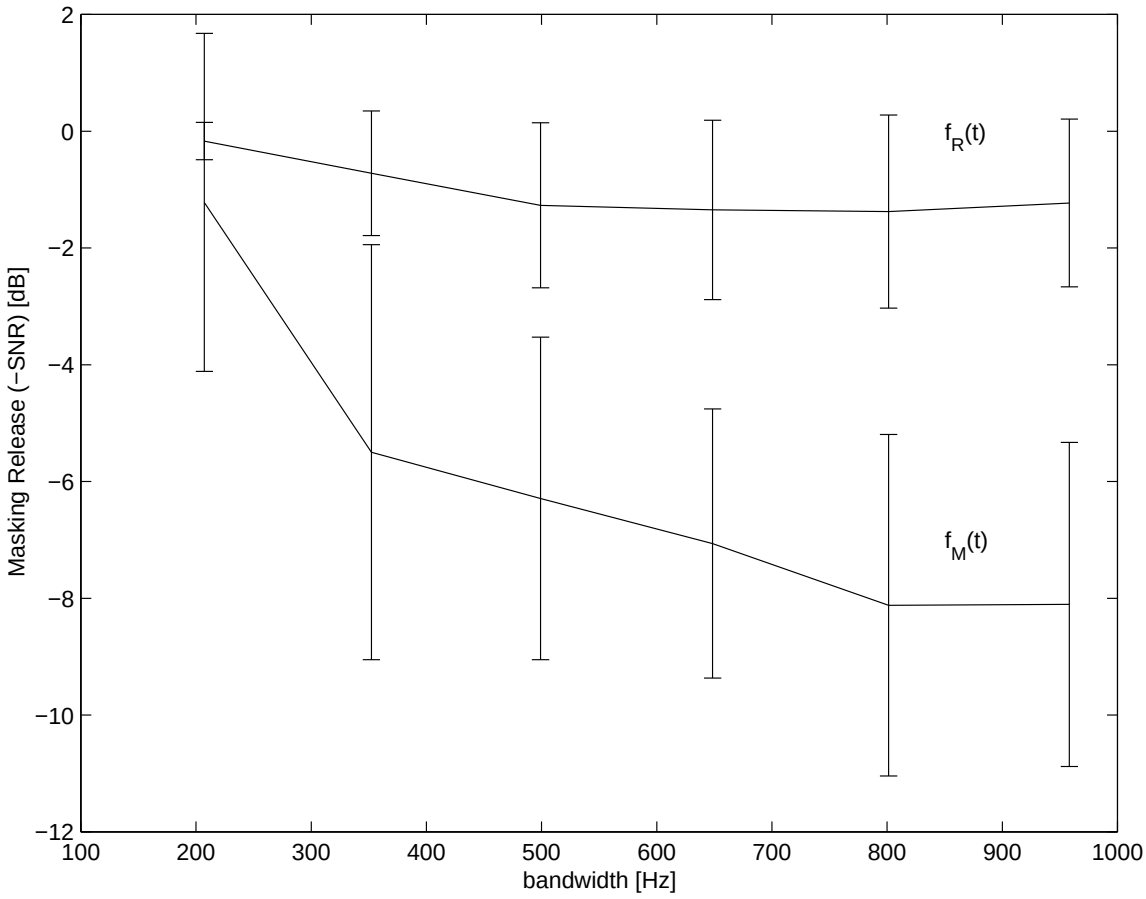

FIG. 10. Relationship between the bandwidth related to the number of adjacent auditory filters and the SNR for the extracted signal $\hat{f}_{1,A}(t)$. The vertical and horizontal axes show the improved SNR of the extracted sinusoidal signal $\hat{f}_{1,A}(t)$ and the bandwidth related to L, respectively. The real line and the error bar show the mean and standard deviation of the SNR of the signal $\hat{f}_{1,A}(t)$ extracted from 300 mixed signals, respectively.

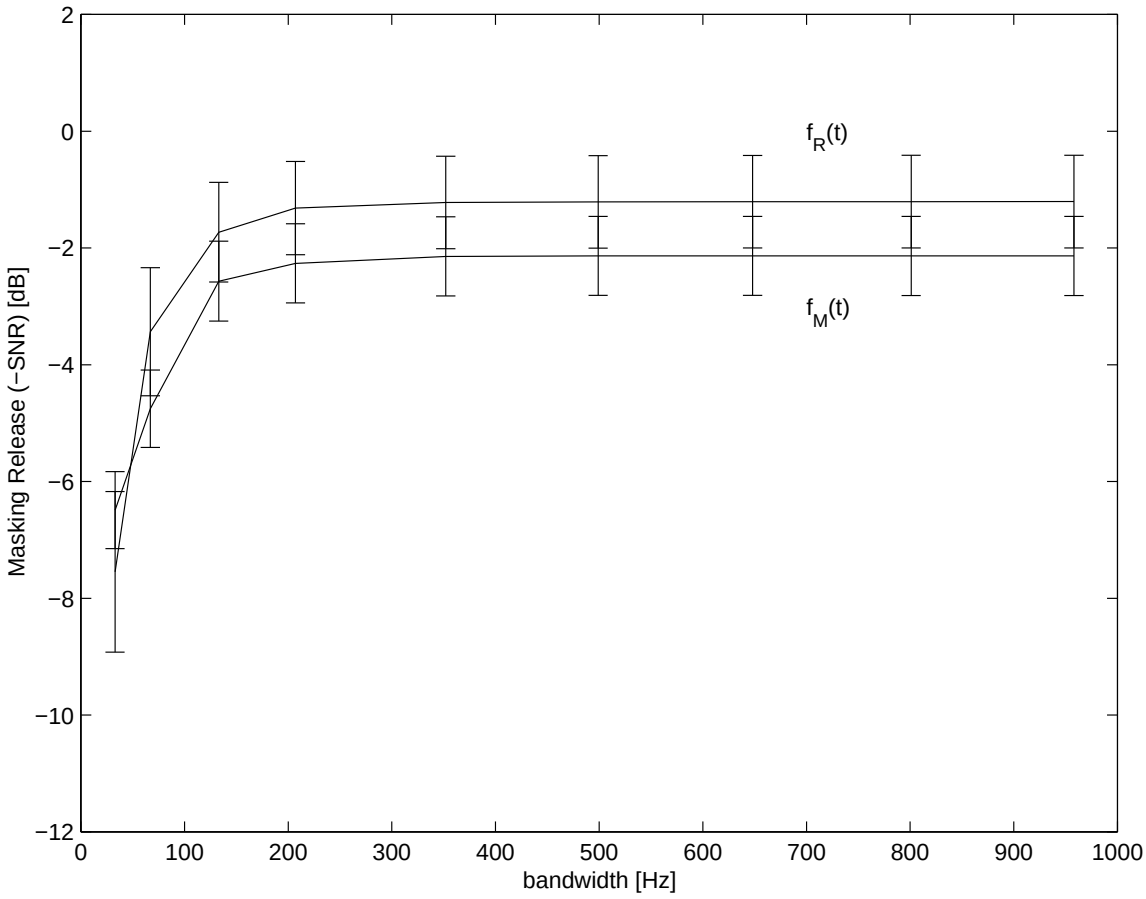

FIG. 11. Relationship between the masker bandwidth and the SNR for the extracted signal $\hat{f}_{1,B}(t)$. The vertical and horizontal axes show the improved SNR of the extracted sinusoidal signal $\hat{f}_{1,B}(t)$ and the bandwidth related to L, respectively. The real line and the error bar show the mean and standard deviation of the SNR of the signal $\hat{f}_{1,B}(t)$ extracted from 300 mixed signals, respectively.

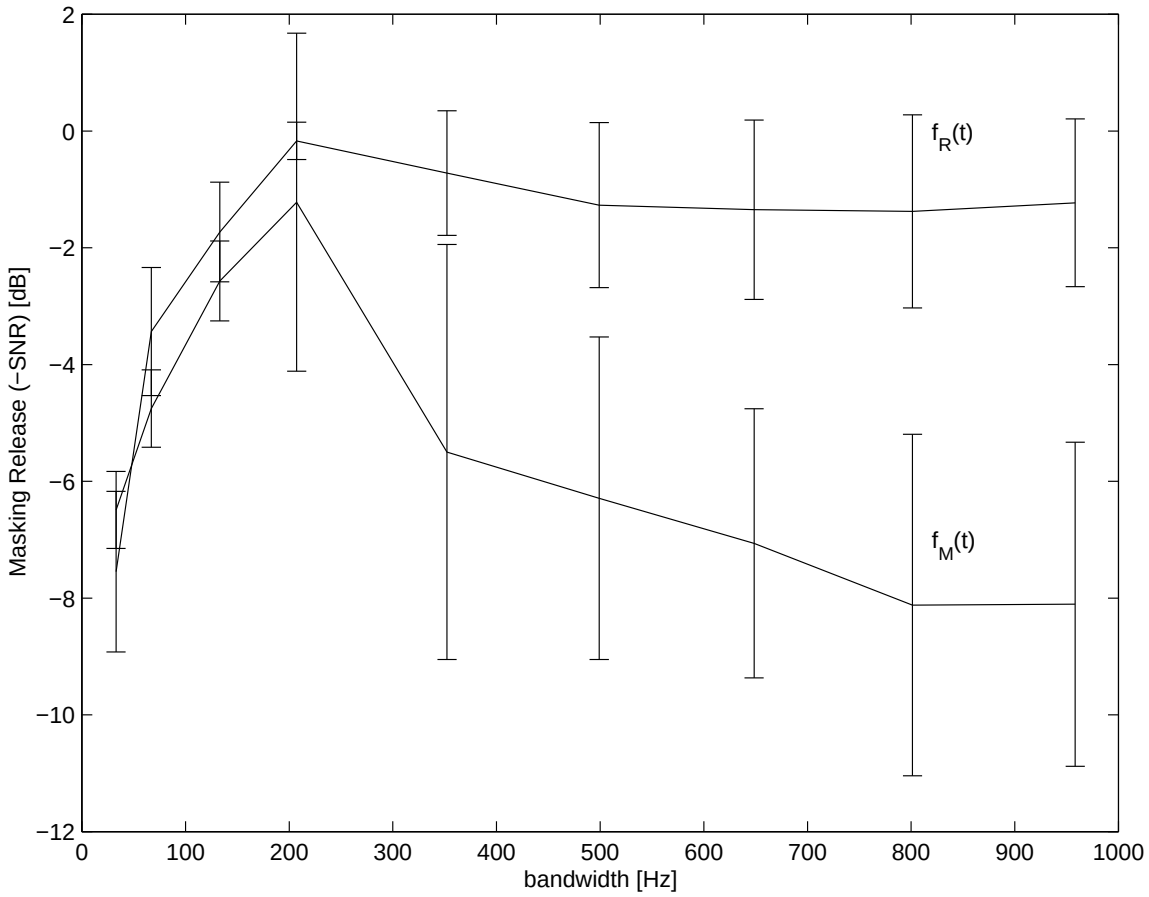

FIG. 12. Relationship between the masker bandwidth and the SNR for the extracted signal. This characteristic was obtained by the result of the selection process from Figs. 10 and 11.