Computational Models of Auditory Function
S. Greenberg and M. Slaney (eds.)
IOS Press, 2001

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 Nomi Ishikawa, 923-1292, Japan

1. Introduction

In investigations of the frequency selectivity of the auditory system, the power-spectrum model of masking [6] is widely accepted as an explanation of the phenomenon of masking. This model assumes that when a listener tries to detect a sinusoidal signal amid background noise he makes use of the output of a single auditory filter having its center frequency close to the signal frequency and having the highest signal-to-masker ratio. In addition, it assumes 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 explains many masking phenomena such as simultaneous masking. However, this model cannot explain all masking phenomena because it ignores the relative phases of the components and the short-term fluctuations in the masker.

In 1984, Hall *et al.* demonstrated that across-filter comparisons can enhance the detection of a sinusoidal signal in a fluctuating noise masker [3]. The crucial feature for achieving this enhancement is that the fluctuations are 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 reduction in masking threshold, was called "Co-modulation Masking Release" (CMR). Many psychoacoustical experiments were carried out [7][4][9] and the same phenomenon was repeatedly demonstrated. The experiments revealed the condition when CMR can occur. But so far, no computational model has been proposed that takes advantage of across-frequency coherence.

On the other hand, the human auditory system can easily segregate the desired signal in a noisy environment that simultaneously contains speech, noise, and reflections. Recently, this ability of the auditory system has been regarded as a function of an active scene analysis system. Called "Auditory Scene Analysis" (ASA), it has become widely known as a result of Bregman's book [1]. Bregman reported that the human auditory system uses four heuristic regularities related to acoustic events to solve the problem of Auditory Scene Analysis. These regularities are (i) common onset and offset, (ii) gradualness of change, (iii) harmonicity, and (iv) changes occurring in the acoustic event [2].

In this work we tackle the problem of segregating the desired signal from a noisy signal [8] using Bregman's regularities [2]. We stress the need to consider not only the amplitude spectrum but also the phase spectrum when attempting to completely extract the signal from noise, both of which are present in the same frequency region [8]. Based on this approach,

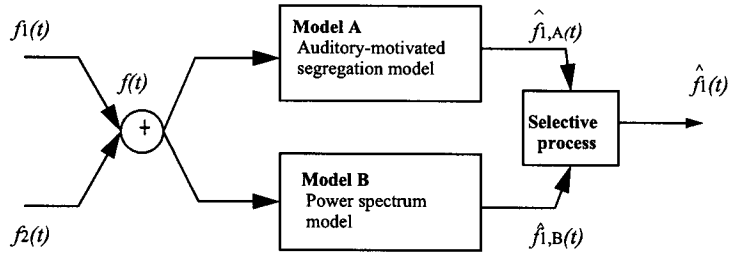

Figure 1 Computational model of CMR. This model consists of two models — our auditory-motivated segregation model (model A) and the power spectrum model of masking (model B) — followed by a selection process that selects one of their results.

we seek to solve the problem of segregating two acoustic sources — the basic problem of acoustic source segregation using regularities (ii) and (iv) of Bregman's principles [2].

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

2. Computational Model of CMR

Our computational model of CMR is shown in Figure 1. It consists of two models (A and B) and a selection process. In this model, we assume that $f_1(t)$ is a sinusoidal signal and $f_2(t)$ is one of two types of noise masker (bandpassed random noise and AM bandpassed random noise) whose center frequency is the same as the signal frequency. We also assume that the sinusoidal signal $f_1(t)$ is added to $f_2(t)$. Since the proposed model can observe only the mixed signal $f(t)$, it extracts the sinusoidal signal $f_1(t)$ using the two models (A and B). Model A is the auditory-motivated segregation model we proposed earlier [8]. Model B is the power spectrum model of masking [6].

We propose a computational framework for CMR, where 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 coherence of the fluctuations. Thus, model B can explain part of CMR by using the output of a single auditory filter when the masker bandwidth increases up to 1 ERB. Model A can explain part of CMR by using the outputs of multiple auditory filters when the masker bandwidth exceeds 1 ERB.

3. Model A: Auditory-Motivated Segregation Model

The auditory-motivated segregation model shown in Figure 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 [2]. The grouping block synthesizes each separated parameter and then reconstructs the extracted signal using the inverse wavelet transform.

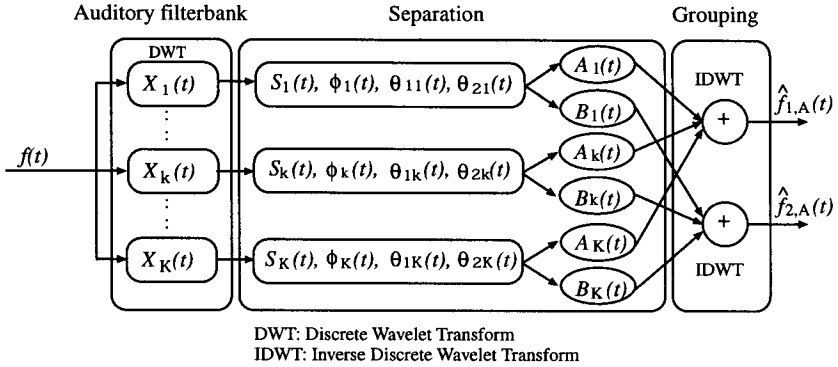

Figure 2 Model A: an auditory-motivated segregation model. This model consists of three parts: (a) an auditory filterbank, (b) separation block, and (c) grouping block.

3.1 Auditory Filterbank

An auditory filterbank is constructed using the wavelet transform, where the basic function $\psi(t)$ is the impulse response of the gammatone filter [5] which is represented using the Hilbert transform.

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

where $\mathrm{ERB}(f_0) = 24.7(4.37(f_0/1000)+1)$ and $f_b = 1.1019\ \mathrm{ERB}(f_0)$. This is a constant Q filterbank having a center frequency f_0 of 1 kHz, a bandpass region from 100 Hz to 10 kHz, and 128 channels. The bandwidth of each auditory filter is 1 ERB. In addition, we compensate for the group delay by adjusting the peak in the envelopes of Equation (1) for all scale parameters, which is called "alignment processing," because a different group delay occurs at each scale.

3.2 Separation and Grouping

First, we can observe only the signal $f(t)$, where $f(t) = f_1(t) + f_2(t)$, $f_1(t)$ 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, the output of the k-th channel, corresponding to $f_1(t)$ and $f_2(t)$, are assumed to be narrow-band sinusoids

$$f_1(t):\ A_k(t)\sin(\omega_k t + \theta_{1k}(t)), \tag{2}$$

and

$$f_2(t):\ B_k(t)\sin(\omega_k t + \theta_{2k}(t)). \tag{3}$$

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 Equations (2) and (3),

$$f(t) = S_k(t)\sin(\omega_k t + \phi_k(t)). \tag{4}$$

Therefore, the amplitude envelopes of the two signals $A_k(t)$ and $B_k(t)$ is equal to

$$A_k(t) = \frac{S_k(t)\sin(\theta_{2k}(t) - \phi_k(t))}{\sin\theta_k(t)}, \tag{5}$$

and

$$B_k(t) = \frac{S_k(t)\sin(\phi_k(t) - \theta_{1k}(t))}{\sin\theta_k(t)}, \tag{6}$$

where $\theta_k(t) = \theta_{2k}(t) - \theta_{1k}(t)$ and $\theta_k(t) \neq n \times \pi, n \in \mathbf{Z}$. Since the amplitude envelope $S_k(t)$ and the output phase $\phi_k(t)$ are observable, then if $\theta_{1k}(t)$ and $\theta_{2k}(t)$ are determined, $A_k(t)$ and $B_k(t)$ can be determined using the equations above. Finally, all the components are synthesized from Equations (2) and (3) 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, $\hat{f}_{1,A}(t)$ and $\hat{f}_{2,A}(t)$ are the reconstructed versions of $f_1(t)$ and $f_2(t)$, respectively.

In this paper, we assume that 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 $\theta_k(t) = \theta_{2k}(t)$.

3.3 Calculating the Four Physical Parameters

The amplitude envelope $S_k(t)$ and phase $\phi_k(t)$ of $X_k(t)$ are determined using the amplitude and phase spectra. Since $\theta_{1k}(t) = 0$, 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 shown below [8].

Constraint 1. Gradualness of change (slowness)

Regularity (ii) means that "a single sound tends to change its properties smoothly and slowly (gradualness of change)" [2]. The first constraint we describe as "slowness," is $dA_k(t)/dt = C_{k,R}(t)$, where $C_{k,R}(t)$ is an R-th-order differentiable polynomial. By applying this constraint to Equation (5), and solving the resulting linear differential equation, we obtain

$$\theta_{2k}(t) = \arctan\left(\frac{S_k(t)\sin\phi_k(t)}{S_k(t)\cos(\phi_k(t) - C_k(t))}\right), \tag{7}$$

where $C_k(t) = \int C_{k,R}(t)dt + C'_{k,0}$. Here, we assume that in a small segment, Δt, $C_{k,R}(t) = C_{k,0}$.

Constraint 2. Gradualness of change (smoothness)

The second constraint we describe as "smoothness." At the boundary ($t = T_r$) between the earlier segment ($T_r - \Delta t \leq t < T_r$) and succeeding segment ($T_r \leq t < T_r + \Delta t$),

$$\left|A_k(T_r + 0) - A_k(T_r - 0)\right| \leq \Delta A \tag{8}$$

$$|B_k(T_r+0)-B_k(T_r-0)| \le \Delta B \tag{9}$$

$$|\theta_{2k}(T_r+0)-\theta_{2k}(T_r-0)| \le \Delta\theta \tag{10}$$

From the above relationships, we can use this constraint to determine $C_{k,0}$, which must satisfy $C_{k,\alpha} \le C_{k,0} \le C_{k,\beta}$. The variables $C_{k,\alpha}$ and $C_{k,\beta}$ are the upper and lower values of $C_{k,0}$ when $\theta_{2k}(t)$ is determined by substituting any value of $C_{k,0}$ for $C_{k,R}(t)$ and then Equations (8)–(10) are satisfied.

Constraint 3. Changes occurring in an acoustic event (regularity)

Regularity (iv) 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" [2]. The third constraint, which we describe as regularity, is

$$\frac{B_k(t)}{\|B_k(t)\|} \sim \frac{B_{k\pm l}(t)}{\|B_{k\pm l}(t)\|}, \qquad l = 1, 2, \ldots, L\ , \tag{11}$$

where L is the number of adjacent auditory filters.

Here, a masker envelope $B_k(t)$ is a function of $C_{k,0}$ from Equations (6) and (7). We consider this constraint to select an optimal coefficient $C_{k,0}$ using

$$\max_{C_{k,\alpha} \le C_{k,0} \le C_{k,\beta}} \frac{\langle \hat{B}, \tilde{B} \rangle}{\|\hat{B}\| \cdot \|\tilde{B}\|}\ , \tag{12}$$

where $\hat{B}_k(t)$ is the masker envelope given by any $C_{k,0}$, and

$$\tilde{B}_k(t) = \frac{1}{2L} \sum_{l=-L,\, l\neq 0}^{L} \frac{\hat{B}_{k+l}(t)}{\|\hat{B}_{k+l}(t)\|}. \tag{13}$$

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

In this chapter, we consider the problem of segregating a masked sinusoidal signal in which the localized signal $f_1(t)$ is added to the noise $f_2(t)$. Therefore, when we solve the above problem using the proposed method, we must know the duration for which two acoustic signals overlap. This can be determined by detecting the onset and offset of $f_1(t)$. By focusing on the temporal deviation of $S_k(t)$ and $\phi_k(t)$, we can determine onset $T_{k,\mathrm{on}}$ and offset $T_{k,\mathrm{off}}$ of $f_1(t)$ as follows:

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

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

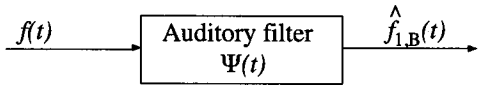

Figure 3 Model B: a power spectrum model of masking.

4. Model B: The Power Spectrum Model of Masking

In the power spectrum model [6], we assume that when a listener is trying to detect a sinusoidal signal with a particular center frequency amid background noise, he uses 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, we assume that only the component passed through a single auditory filter 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 Figure 3. This filter consists of a gammatone filter whose center frequency is 1 kHz and bandwidth is 1 ERB. In this model, the sinusoidal signal $\hat{f}_{1,B}(t)$ extracted from the masked signal $f(t)$ is the output of the single auditory filter $X_k(t)$.

5. Simulations

5.1 Co-modulation Masking Release

Hall *et al.* measured the masking threshold for a sinusoidal signal in one of their experiments as a function of the bandwidth of a continuous noise masker. They used a center frequency of 1 kHz, a duration of 400 ms and kept the spectrum level constant [3]. They used two types of masker — a random noise masker and an amplitude modulated random noise masker — which were both centered at 1 kHz. The random noise masker had irregular fluctuations in amplitude, and the fluctuations in different frequency regions were independent. The amplitude-modulated masker was a random noise that was amplitude modulated at an irregular, slow rate; a noise that was 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.

Figure 4 shows the results of that experiment. For the random noise (denoted by R), the signal threshold increased as the masker bandwidth increased up to ca. 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 ca. 130 Hz. Hence, for noise bandwidths up to about 130 Hz, increasing the bandwidth 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 threshold decreased with increasing bandwidth only 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 10 dB [3].

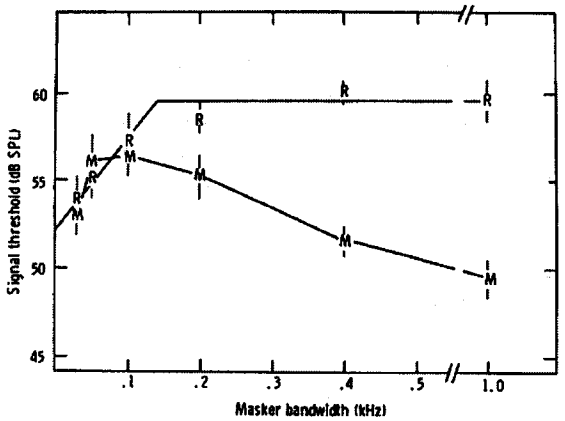

Figure 4 Results for CMR (Hall *et al*, 1984). The points labeled 'R' are thresholds for a 1-kHz signal centered in a band of random noise, plotted as a function of the bandwidth of the noise. The points labeled "M" are the thresholds obtained when the noise was amplitude modulated at an irregular,

5.2 Simulations for Model A

5.2.1 Stimuli and Procedure

We considered conditions equivalent to the experimental ones used by Hall *et al*. In this simulation we assumed that $f_1(t)$ was a sinusoidal signal, where the center frequency was 1 kHz, the duration was 400 ms, and the amplitude envelope was constant, and the masker $f_2(t)$ was two types of bandpassed noise having its center frequency close to the signal frequency. One was a bandpassed random noise $f_{21}(t)$ and other was an AM bandpassed random noise $f_{22}(t)$. The AM masker was calculated by amplitude modulating $f_{21}(t)$, 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, we must determine the number of adjacent auditory filters, L, to use in Equation (11). However we don't know this number when CMR occurs. We don't know which channels actually contribute to the CMR effect observed in psychoacoustics. We assume that the number of relevant auditory filters required in this model is simply determined by the total masker bandwidth. To realize the different experimental conditions, the initial bandwidth of the masking noise($f_{21}(t)$ and $f_{22}(t)$) was kept constant at 1 kHz, and only the number of auditory filters to be processed by the model was adjusted. 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 in Figure 4 labeled R and M, respectively. Simulation stimuli, consisting of 10 sinusoidal signals, were formed by varying the onset. 30 maskers of the two types were generated by varying the random seeds. Thus, the total number of stimuli was 300. For example, one of the two types of mixed signals is shown in Figure 5. 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.

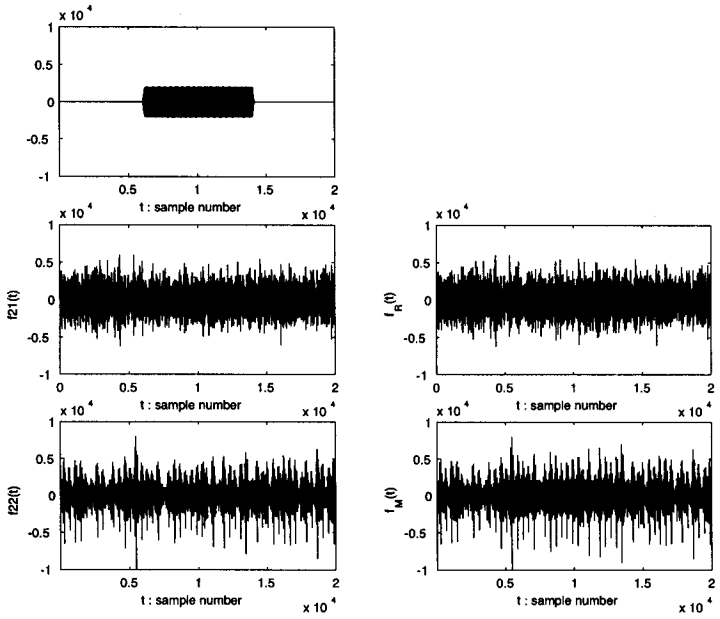

Figure 5 Stimuli: a sinusoidal signal $f_1(t)$ (left-top), a bandpassed random noise $f_{21}(t)$ (left-middle), and an AM bandpassed noise $f_{22}(t)$ (left-bottom). Mixed signals $f_R(t)$ (right-top) and $f_M(t)$ (right-bottom).

In this paper, we set the parameter as $\Delta t = 3/(f_0\alpha^{k-K/2})$, $\Delta A = |A_k(T_r-\Delta t)-A_k(T_r-2t\Delta)|$, $\Delta B = 0.01S_{\max}$, $\Delta\theta = \pi/20$, and $S_{\max}$ 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.

In their demonstration of CMR, Hall *et al* measured the masking threshold as a function of the masker bandwidth. Our simulation conditions are equivalent 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 $\tilde{B}_k(t)$ as a function of L from Equations (7), (12), and (13). The bandwidths related to L=1, 3, 5, 7, 9, 11 are 207, 352, 499, 648, 801, 958 Hz, respectively.

5.2.2 Results and Discussion

Simulations were carried out according to the conditions described above. The results are shown in Figure 6, 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 line and the error bars 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

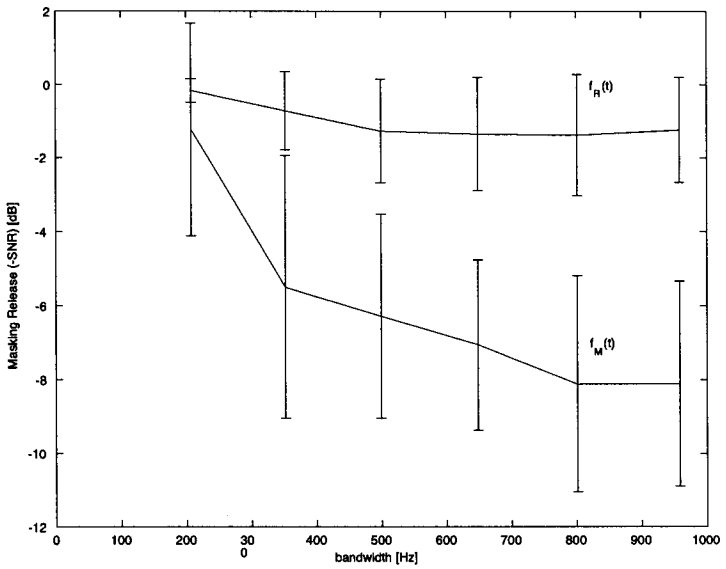

Figure 6 Relationship between the bandwidth related to the number of adjacent auditory filters and the SNR for the extracted signal $\hat{f}_{1,A}(t)$. The vertical and horizontal axes show the improved SNR of the extracted sinusoidal signal $\hat{f}_{1,A}(t)$ and the bandwidth related to L, respectively. The real line and the error bars show the mean and standard deviation of the SNR of the signal $\hat{f}_{1,A}(t)$ extracted from 300 mixed signals, respectively.

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 reduction of masking using the outputs of multiple auditory filters.

5.3 Simulations for Model B

5.3.1 Stimuli and Procedure

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 because we don't need to determine the number of adjacent filters, L, and we can control the masker bandwidth directly. Three of these bandwidths were related to 1/4, 1/2, and 1 ERB, respectively. The remainder were the same bandwidths used in the simulations for model A.

In model B, in order to measure the masking threshold as a function of the masker bandwidth, we measure the SNR of the extracted sinusoidal signal $\hat{f}_{1,B}(t)$ from noise-added signal as a function of the masker bandwidth, using the same evaluation measure of masking threshold in model A.

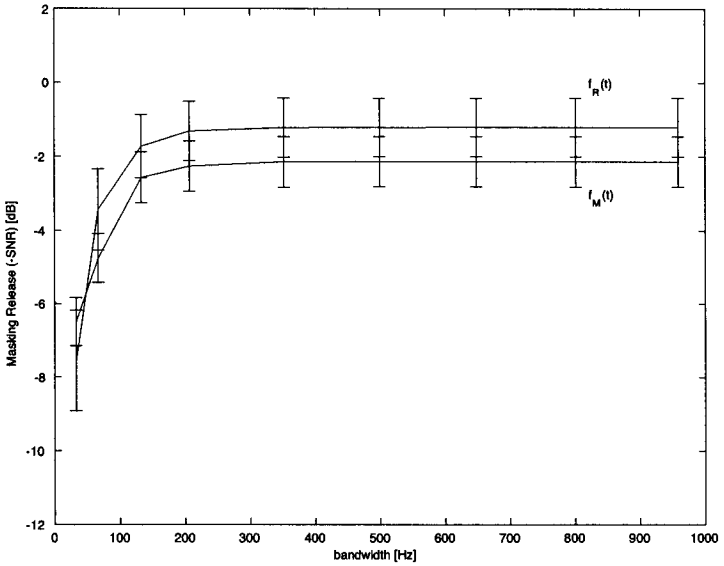

Figure 7 Relationship between the masker bandwidth and the SNR for the extracted signal $\hat{f}_{1,B}(t)$. The vertical and horizontal axes show the improved SNR of the extracted sinusoidal signal $\hat{f}_{1,B}(t)$ and the bandwidth related to L, respectively. The real line and the error bars show the mean and standard deviation of the SNR of the signal $\hat{f}_{1,B}(t)$ extracted from 300 mixed signals, respectively.

5.3.2 Results and Discussion

Simulations were carried out according to the descriptions above. The results are shown in Figure 7, 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 line and the error bars show the mean and standard deviation of the SNR, respectively. Figure 7 shows 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 and then remained constant. Hence, model B simulates the phenomenon of simultaneous masking using the output of a single auditory filter.

5.4 Considerations for Computational Model of CMR

The results of simulations for the two models show two types of CMR behavior. Model A simulates the phenomenon of CMR/simultaneous masking by using the coherence of the fluctuations in the amplitude envelope of a masker as the masker bandwidth increases above 1 ERB. By contrast, model B simulates 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 then $\hat{f}_1(t)$ is the extracted signal with the highest SNR. Thus, based on the results in Figures 6 and 7, the proposed model has the masking

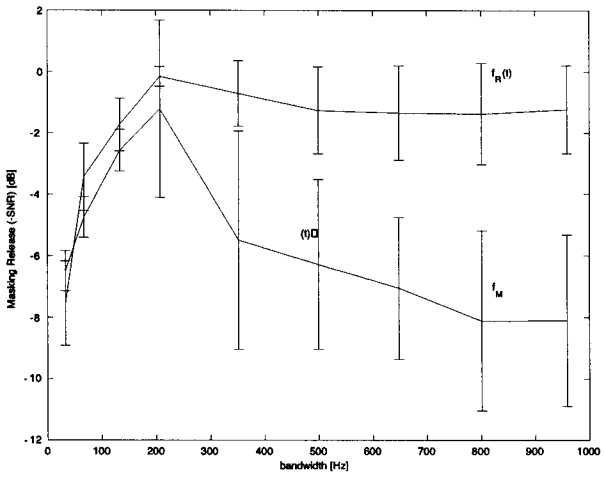

Figure 8 Relationship between the masker bandwidth and the SNR for the extracted signal. This characteristic was obtained from the result of the selection process from Figures 6 and 7.

threshold shown in Figure 8. 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, the proposed model is a computational model of CMR. The maximum amount of CMR in Hall *et al.*'s demonstrations was about 10 dB, whereas in our model it is about 8 dB.

6. Conclusions

In this paper, we have proposed a computational framework for CMR. This framework consists of two models, our auditory-motivated segregation model (model A) and the power spectrum model of masking (model B), as well as 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 two models. As a result, the characteristics of the proposed model show that the phenomenon of CMR closely corresponds to Hall *et al.*'s results. The maximum amount of CMR in the proposed model was about 8 dB.

Hence, the proposed model is a computational model of CMR. We also showed that signal slowness and smoothness — related to regularity (ii) — and the same fluctuation pattern

in different frequency regions — related to regularity (iv) — are all important cues to explain CMR.

Acknowledgments

This work was supported by Grant-in-Aid for science research from the Ministry of Education (Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists and 10680374) and by CREST (Core Research for Evolutional Science and Technology) of the Japan Science and Technology Corporation (JST).

References

[1] Bregman, A. S. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press, 1990.

[2] Bregman, A. S. "Auditory Scene Analysis: hearing in complex environments." In *Thinking in Sound: The Cognitive Psychology of Human Audition*, S. McAdams and E. Bigand (eds.), New York: Oxford University Press, pp. 10–36, 1993.

[3] Hall, J. W. and Fernandes, M. A. "The role of monaural frequency selectivity in binaural analysis." *J. Acoust. Soc. Am.* 76: 435–439, 1984.

[4] Hall, J. W. and Grose, J. H. "Comodulation masking release: Evidence for multiple cues." *J. Acoust. Soc. Am.* 84: 1669–1675, 1988.

[5] Patterson, R. D. and Holdsworth, J. "A functional model of neural activity patterns and auditory images." In *Advances in Speech, Hearing and Language Processing*, Volume 3, London: JAI Press, 1991.

[6] Patterson, R. D. and Moore, B. C. J. "Auditory filters and excitation patterns as representations of frequency resolution." In *Frequency Selectivity in Hearing*, B. C. J. Moore (ed.), London: Academic Press, pp. 123–178, 1986.

[7] Moore, B. C. J. "Comodulation masking release and modulation discrimination interface." In *The Auditory Processing of Speech, from Sound to Words*, M. E. H. Schouten (ed.), New York: Mouton de Gruyter, pp. 167–183, 1992.

[8] Unoki, M and Akagi, M. *Method of Signal Extraction from Noise-Added Signal, Electronics and Communications in Japan, Part 3, Vol. 80, No. 11, 1997*. Translated from *IEICE, Vol. J80-A*, No. 3, pp. 444–453, March and http://www.jaist.ac.jp/~unoki/index-e.html.

[9] van den Brink, W. A. C., Houtgast, T., and Smoorenburg, G. F. "Effectiveness of comodulation masking release." In *The Auditory Processing of Speech, from Sound to Words*, M. E. H. Schouten (ed.), New York: Mouton de Gruyter, 1992.