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, 923-1211, 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. The selection process selects the sinusoidal signal with the lowest signal threshold from the two extracted signals. 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. 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.

1. INTRODUCTION

In investigations for frequency selectivity of the auditory system, the power spectrum model of masking [1] 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 because 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 [2]. 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 [3–5] 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) [6]. 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 [7]; 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 [6].

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.

2. 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 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 [7]. Model B is the power spectrum model of masking [1]. 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

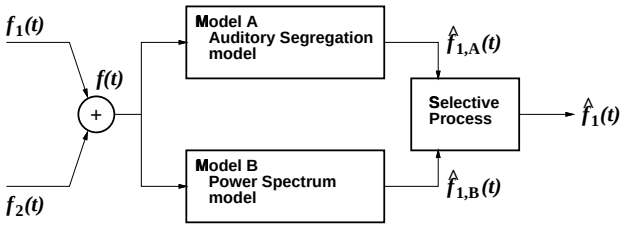

Figure 1: Computational model of CMR.

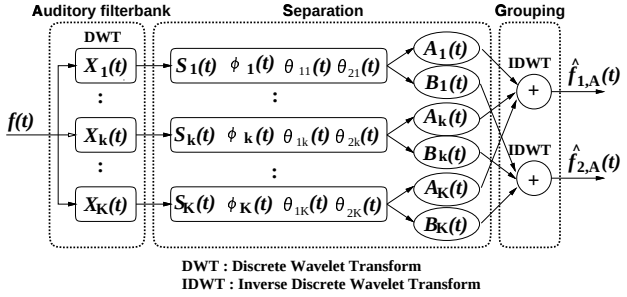

Figure 2: Model A: an auditory segregation model.

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.

3. 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 [6]. The grouping block synthesizes each separated parameter and then reconstructs the extracted signal using the inverse wavelet transform.

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 [8] which is represented using the Hilbert transform.

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

where $\mathrm{ERB}(f_0) = 24.7(4.37f_0/1000 + 1)$ and $b_f = 1.019\mathrm{ERB}(f_0)$. This is a constant Q filterbank whose a center frequency f_0 is 1 kHz, a bandpassed region from 100 Hz to 10 kHz, and number of channel of 128; the bandwidth of the auditory filter is 1 ERB. In addition, we compensate for the group delay by adjusting the peak in the envelopes of Eq. (1) for all scale parameters, which is called "alinement processing," because the group delay occurs for 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, outputs of the k-th channel, which correspond to $f_1(t)$ and $f_2(t)$, are assumed to be

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

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 Eqs. (2) and (3),

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

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

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

and

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

where $\theta_k(t) = \theta_{2k}(t) - \theta_{1k}(t)$ and $\theta_k(t) \neq n\pi, n \in \mathbf{Z}$. 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 by the above equations. Finally, all the components are synthesized from Eqs. (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, $\hat{f}_{1,\mathrm{A}}(t)$ and $\hat{f}_{2,\mathrm{B}}(t)$ are the reconstructed $f_1(t)$ and $f_2(t)$, respectively.

In this paper, we assume that 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. Calculation of 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 [7].

1. Gradualness of change (slowness)

   Regularity (ii) means that "a single sound tends to change its properties smoothly and slowly (gradualness of change)" [6]. First constraint, considered 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 it to Eq. (5), and solving the resulting linear differential equation, we obtain

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

   where $C_k(t) = -\int C_{k,R}(t)dt + C_{k,0}$. Here, we assume that, in small segment Δt, $C_{k,R}(t) = C_{k,0}$.

2. Gradualness of change (smoothness)

   Second constraint, considered as "smoothness", is that, in the bound ($t = T_r$) between pre-segment ($T_r - \Delta t \le t < T_r$) and post-segment ($T_r \le t < T_r + \Delta t$),

$$|A_k(T_r+0) - A_k(T_r-0)| \le \Delta A \quad (8)$$
$$|B_k(T_r+0) - B_k(T_r-0)| \le \Delta B \quad (9)$$
$$|\theta_{2k}(T_r+0) - \theta_{2k}(T_r-0)| \le \Delta\theta. \quad (10)$$

   From the above relationships, we can interpret this constraint in order to determine $C_{k,0}$, which is restricted within $C_{k,\alpha} \le C_{k,0} \le C_{k,\beta}$, 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.

3. Changes taken in an acoustic event

   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" [6]. Third constraint, considered as this regularity, is

$$\frac{B_k(t)}{\|B_k(t)\|} \approx \frac{B_{k\pm\ell}(t)}{\|B_{k\pm\ell}(t)\|}, \qquad \ell = 1, 2, \cdots, L. \quad (11)$$

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

$$\max_{C_{k,\alpha} \le C_{k,0} \le C_{k,\beta}} \frac{\langle \hat{B}_k, \hat{\hat{B}}_k \rangle}{||\hat{B}_k|| ||\hat{\hat{B}}_k||}, \quad (12)$$

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

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

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 $\theta_{2k}(t)$, 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 $C_{k,0}$.

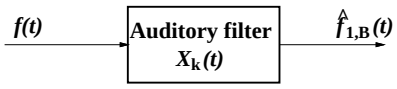

Figure 3: Model B: a power spectrum model of masking.

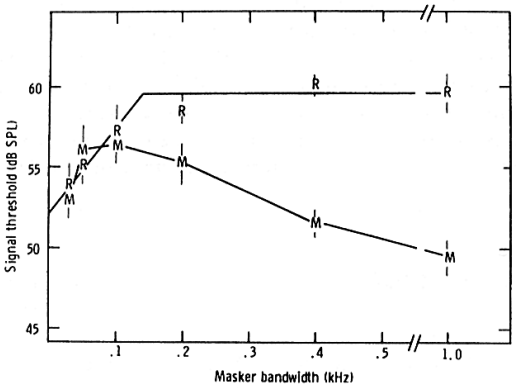

Figure 4: Results of CMR (Hall *et al.*, 1984). The points labeled 'R' are thresholds for 1 kHz signal centered in a band of random noise, plotted as a function of the bandwidth of the noise. The points labeled 'M' are the thresholds obtained when the noise was amplitude modulated at an irregular, low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

4. MODEL B: THE POWER SPECTRUM MODEL OF MASKING

In the power spectrum model [1], 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. 3. 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 $\hat{f}_{1,\mathrm{B}}(t)$ 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, 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 [2]. They used two types of masker, a random noise masker and an amplitude modulated random noise masker, which were both centered at 1 kHz, as follows: The former had irregular fluctuations in amplitude, and the fluctuations in different frequency regions were independent. The later 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.

Fig. 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 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 [2].

5.2. Simulations for Model A

5.2.1. Stimuli and procedure

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 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

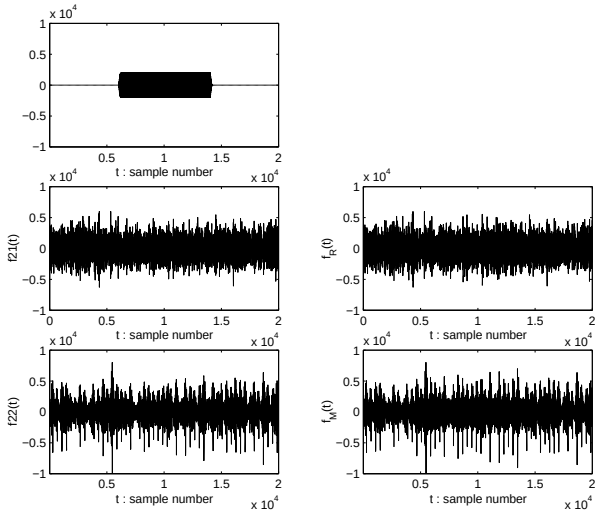

Figure 5: Stimuli: a sinusoidal signal $f_1(t)$ (left-top), a bandpassed random noise $f_{21}(t)$ (left-middle), and an AM bandpassed noise $f_{22}(t)$ (left-bottom). Mixed signals $f_R(t)$ (right-top) and $f_M(t)$ (right-bottom).

$\sqrt{f_{21}(t)^2/f_{22}(t)^2} = 1$. 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. 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.

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 $\hat{B}_k(t)$ as a function of L from Eqs. (7), (12), and (13). The bandwidths related to $L = 1, 3, 5, 7, 9, 11$ are 207, 352, 499, 648, 801, 958 Hz, respectively.

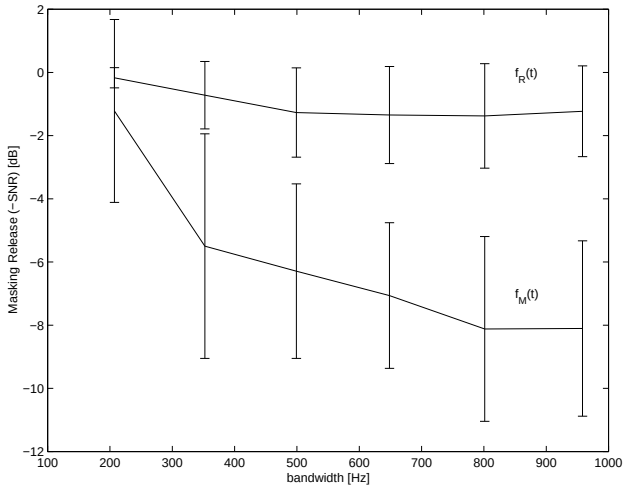

Figure 6: Relationship between the bandwidth related to the number of adjacent auditory filters and the SNR for the extracted signal $\hat{f}_{1,A}(t)$.

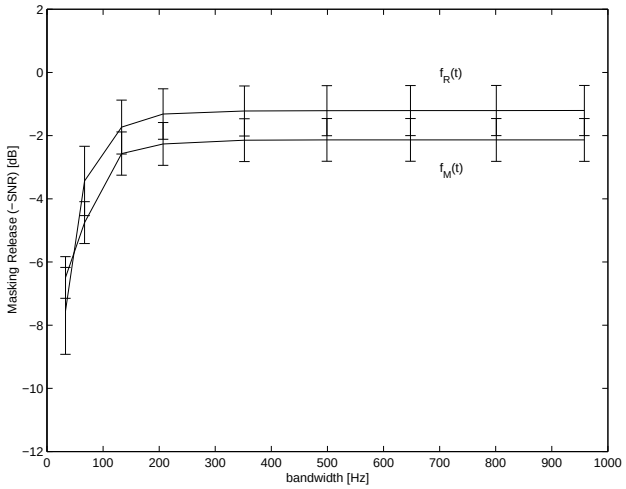

Figure 7: Relationship between the masker bandwidth and the SNR for the extracted signal $\hat{f}_{1,B}(t)$.

5.2.2. Results and discussion

Simulations were carried out according the conditions mentioned above. The results are shown in Fig. 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 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.

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. Three of these bandwidths were related to 1/4, 1/2, and 1 ERB, respectively. The remainder were the same bandwidths related to 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 sinusoidal signal $\hat{f}_{1,B}(t)$ extracted for the masking threshold as a function of the masker bandwidth.

5.3.2. Results and discussion

Simulations were carried out according to the above mentioned conditions. The results are shown in Fig. 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 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.

5.4. 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

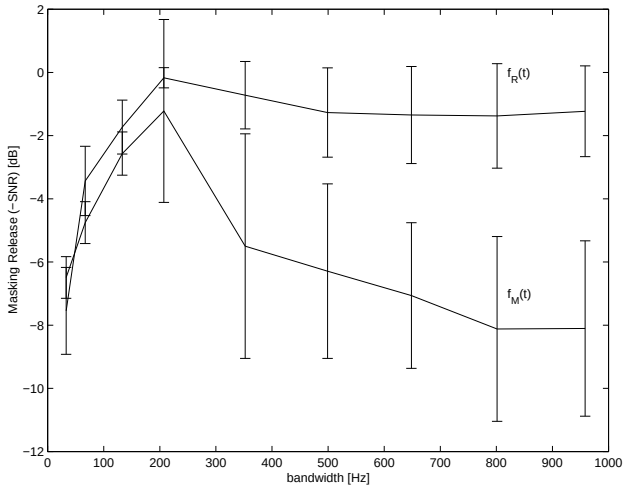

Figure 8: Relationship between the masker bandwidth and the SNR for the extracted signal. This characteristic was obtained by the result of the selection process from Figs. 10 and 11.

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. 6 and 7 the proposed model has the characteristics of the masking threshold shown in Fig. 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, 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.

6. CONCLUSIONS

In this paper, we proposed a 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. 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.

7. ACKNOWLEDGEMENTS

This work was supported by CREST.

8. REFERENCES

[1] 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.

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

[3] 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.

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

[5] 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.

[6] 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.

[7] Masashi Unoki and Masato Akagi. (1997). "A 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 (in Japanese) and http://www.jaist.ac.jp/~unoki/index-e.html.

[8] 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.