Estimates of auditory filter shape using simultaneous and forward notched-noise masking

Masashi Unoki, Chin-Tuan Tan
School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, 923-1292, JAPAN, email: {unoki, chintuan}@jaist.ac.jp

In this paper, we attempted to estimate the auditory filter shape using both simultaneous and forward masking experiments with notched-noise masker. In simultaneous masking experiment, masked threshold of the probe signal in the presence of a notched-noise masker were measured for five normal hearing subjects. In each trial, the level of the probe signal was varied while masker level was kept at a constant level. The frequencies of the probe signals were 0.5, 1.0, 2.0, and 4.0 kHz. The width and position of the notched-noise maskers were the same as in Glasberg and Moore (2000). In forward masking experiment, masked threshold were measured for six normal hearing subjects in similar notch conditions at the same frequencies, with the masker varied in the presence of a constant level of probe signal. The levels of probe signal were presented at 10, 20, and 30 dB above the absolute threshold of the subjects. A double roex auditory filter was then individually fitted to the data collected in these two experiments by the PolyFit procedure. ERB (Equivalent rectangular bandwidth) and the ratio of ERBs were calculated from these filters as measures to compare the tuning of the derived filters. The results suggest that the tuning of the auditory filters derived from the forward masking data was considerably sharper than those derived from simultaneous masking data. The tuning of the auditory filters becomes much sharper as the center frequency increases. However, the degree of sharpening of the tuning of the auditory filter by the forward masking data was not consistent at different center frequencies.

1 Introduction

Conceptually, the nature of frequency selectivity in human auditory system can be characterized by the analysis nature of a series of overlapping bandpass filters, which are often referred to as the auditory filters. Forward and simultaneous masking experiments [1-4, 9-11] using notched-noise masker were commonly used in deriving psychophysical data for the auditory filter estimation. In these masking experiments, the masked threshold of a sinusoidal signal in the presence of a noise with a spectral notch in which the width and position of the notch varies relatively to the signal frequency is measured. The advantage of using notched-noise as masker, is that it can avoid the off-frequency listening. However, the psychophysical data collected using notched-noise varies from the way the masking experiment is conducted.

Apparently, the tunings of auditory filters derived from the data collected in forward masking experiment was sharper than those in the simultaneous masking experiment, especially when the signal level is low. It is commonly believe that the tuning of the filter would be affected by the cochlear nonlinearity such as the effect of suppression. In the past studies, the tuning of the auditory filter derived from the simultaneous masking data is wider than that of the filter derived from the non-simultaneous (forward) masking data [7]. Glasberg and Moore (1982) showed that a 3-dB bandwidth of the estimated filter at low level from the forward masking is sharper than that of the estimated filter in the simultaneous masking [2]. Heinz et al. (2002) showed that tuning is generally sharpest when the stimuli are at low level and suppression may affect tuning estimates more at high characteristic frequencies (CFs) than at low CFs [5].

One of the popular measures in estimating the tuning of the auditory filters would be Equivalent Rectangular Bandwidth (ERB) [3]. If the suggestion of Heinz et al. (2002) holds, that suppression affects change with frequency, the comparison of the ERBs derived from simultaneous and forward masking experiments would be a good indication of this observation.

In this paper, we attempt to estimate auditory filter shapes using both simultaneous and forward masking experiments with notched-noise masker and, study the difference in the tuning of the derived filter from these two sets of masking data in term of center frequencies and signal level. ERBs and the ratios of ERBs were calculated from the derived auditory filters. A comparison of ERBs was made across these two sets of masking data.

2 Simultaneous and forward masking with notched-noise masker

2.1 Stimuli

Figure 1 shows the shape of the stimulus used in these notched-noise masking experiments. Subject was
required to detect a brief sinusoidal signal, which was referred to as a “probe” signal, in the presence of a noise with a spectral notch designed to be placed about the frequency region of the probe signal, which was referred to as “notched-noise masker”. The level of the probe signal at which it is just audible for the subjects, in the presence of the notched-noise masker was referred to as “masked threshold”. \( f_c \), \( \Delta f_n \), \( P_s \), and \( N_n \) denote signal frequency (Hz), notch width (Hz), probe signal level (dB SPL), and masker level (dB SPL/Hz), respectively.

Masked threshold were measured in simultaneous and forward masking with notched-noise masker as shown in Fig. 1. The signal frequencies in quest are 0.5, 1.0, 2.0, and 4.0 kHz. The notched-noise masker consists of two bands of Gaussian white noise centered below and above signal frequency, \( f_c \), each with a bandwidth of 0.4 \( f_c \). The noise bands were generated in the spectral domain by setting all spectral components beyond the desired pass band to zero.

In the simultaneous masking experiment, the relative notch width for each notched-noise masker centred at \( f_c \) was defined as the normalized deviation of each edge of the normalized notch from \( f_c \), denoted as \( \Delta f_n / f_c \). There were seven conditions in which the notch was symmetrically placed about the signal frequency. The value of \( \Delta f_n / f_c \) in these conditions were 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, and 0.6. There were another twelve conditions in which the notch was asymmetrically placed about the signal. The combination of lower and upper normalized deviations \( \Delta f_n / f_c \) in asymmetric conditions were (0.1, 0.3), (0.3, 0.1), (0.2, 0.4), (0.4, 0.2), (0.3, 0.5), (0.5, 0.3), (0.4, 0.6), (0.6, 0.4), (0.5, 0.7), (0.7, 0.5), (0.6, 0.8), and (0.8, 0.6). In total, there are 19 notch-conditions.

Thresholds were measured for a 200-ms signal (10-ms raised-cosine ramps and 180 ms steady state) in the presence of 200-ms masker (10-ms raised cosine ramps). The inter-stimulus interval was kept at 500 ms, as shown in Fig. 2(a). At a fixed masker level \( N_n \), masked thresholds were measured at signal frequencies of 0.5, 1, 2, and 4 kHz with the notch width of the masker and the signal level \( P_s \) varying. \( N_n \) was chosen as (1) at \( f_c = 0.5 \), 1.0, and 2.0 kHz, \( N_n = 27.3, 37.3, \) and 47.3 dB SPL/Hz; and (b) at \( f_c = 4.0 \) kHz, \( N_n = 17.3, 27.3, \) and 37.3 dB SPL/Hz. The minimum value of the low-frequency edge of the lower noise band \( f_{min} \) was kept at 40 Hz. With the masker presented at 3 different levels, a total of 57 conditions (\( 19 \times 3 \)) was tested at each signal frequency.

Similarly, there were seven conditions in which the notch was placed symmetrically about the signal frequency in the forward masking experiment. The value of \( \Delta f_n / f_c \) in these conditions were 0.0, 0.05, 0.1, 0.15, 0.2, 0.3, and 0.4. Together with the others six asymmetric conditions in which the combination of the lower and upper normalized deviations \( \Delta f_n / f_c \) were (0.25, 0.05), (0.05, 0.25), (0.3, 0.1), (0.1, 0.3), (0.2, 0.4), and (0.4, 0.2), a total of 13 notch-conditions were set for the forward masking experiment.

Thresholds were measured for a brief 10-ms signal (5-ms raised-cosine ramps: no steady state) in the presence of 300-ms masker gated with 15-ms raised-cosine ramps as shown in Fig. 2(b). The silent interval or gap between the masker and the signal was 5 ms and the inter-stimulus interval was kept at 500 ms. At a fixed probe signal level \( P_s \), masked thresholds were measured at the same signal frequencies as in the simultaneous masking experiment with the masker level \( N_n \) varying. Probe signal level was chosen as 10,
20, and 30 dB above the absolute threshold (10, 20, and 30 dB SL) of the subjects. With the signal presented at 3 different levels, a total of 39 conditions (13 x 3) was tested at each signal frequency.

All stimuli were re-generated digitally at a sampling frequency of 48 kHz and presented via Tucker-Davis Technologies (TDT) system III real-time processor (TDT RP2). The masker and signal were separately attenuated by two different programmable attenuators (TDT PA5) before they were mixed (TDT SM5) and passed through a headphone buffer (TDT HB7) for presentation. The stimuli were presented monaurally to the subjects in a double-walled sound attenuating booth via Etymotic Research ER2 insert earphones. This earphone has a flat frequency response at the eardrum up to about 14 kHz. The level of stimuli were verified using B&K 4152 Artificial Ear Simulator with the eardrum up to about 14 kHz. The level of stimuli were verified using B&K 4152 Artificial Ear Simulator with a 2 cm³ coupler (B&K DB 0138) and B&K 2231 Modular Precision Sound Level Meter.

2.2 Subjects

Eight normal-hearing listeners (AH, MT, YI, YY, MU, JN, YT, and TT), aged from 21 to 35, participated in these masking experiments. The fifth subject was the first author. Five subjects (AH, MT, YI, YY, and MU) participated in simultaneous masking experiment while six subjects (AH, MT, YI, JN, YT, and TT) participated in forward masking experiments. First three subjects participated in both experiments. All subjects had absolute thresholds, measured by a standard audiometric tone test using RION AA-72B audiometer, at octave frequency between 125 and 8000 Hz of 15 dB HL or less for both ears. All subjects were given, at least, two hours of practice.

2.3 Procedure

Masked thresholds were measured using a three alternative forced-choice (3AFC) three-down one-up procedure that tracks the 79.4% point on the psychometric function [6]. Three intervals of stimuli were presented sequentially in each trial. Subjects were required to identify the interval which carries the probe signal using the numbered push-buttons on the response box. A feedback was provided by lighting up the LEDs corresponding to the correct interval on the response box. A feedback was provided by lighting up the corresponding led after each trial. A run was terminated after 12 reversals. The step size was 5 dB for the first 4 reversals, and 2 dB thereafter. Threshold was defined as the mean signal level at the last 8 reversals.

2.4 Results

The mean masked thresholds of the five subjects in simultaneous masking experiment were shown in Fig. 3. The abscissa of the plots in Fig. 3, shows the smaller of the two values of \( \Delta f_s/f_s \), and the ordinate shows masked threshold. Circles denote the mean masked thresholds in the symmetric notched-noise conditions. Left-pointing triangles denote the mean masked thresholds in the asymmetric notched-noise conditions where \( \Delta f_s/f_s \) for the lower noise band is 0.2 greater than that for upper noise band. Right-pointing triangles denote the mean masked thresholds in the asymmetric notched-noise conditions where \( \Delta f_s/f_s \) for the upper noise band is 0.2 greater than that for lower noise band. The symbol “+” in each plot denote the mean absolute threshold of the subjects.

In general, the masked threshold decreases as the width of the notch increases, and approaches the level of the mean absolute threshold as the notch widens further. For all signal frequencies, masked thresholds measured in asymmetric notch conditions (right- and left-pointing triangles) were different from their corresponding masked threshold measured in symmetric notch condition (circle). This indicates that the auditory filter shapes are asymmetrical.

Similarly, the mean masked thresholds of the six subjects in forward masking experiment were shown in Fig. 4. Fig. 4 uses the same symbol notation as in Fig. 3. The ordinate in Fig. 4 shows the masker level at the masked threshold. Mean absolute thresholds were 21.8, 10.8, 11.5, and 10.3 dB SPL at 0.5, 1.0, 2.0, and 4.0 kHz, respectively. The plots show that the masker level required to mask the signal increases as notch width increases. Furthermore, the left-pointing triangles were consistently higher than the right-pointing triangles suggests that the auditory filters are asymmetric, with a steeper high frequency slope.

3 Filter shape estimation

3.1 The power spectrum model of masking

If the roll-off of the noise band is as steep as in Fig. 1, it is possible to write a function that relates the probe signal level at the masked threshold to the integral of the auditory filter. In this paper, this relationship was used to estimate the auditory filter shape [3]. If the auditory filter shape was represented as the weighting function, \( W(f) \), then masked threshold predicted by the power spectrum model of masking was given by

\[
P_e = K + N_s + 10 \log_{10} \left\{ \int_{f_{\min}}^{f_{\max}} W(f) df + \int_{f_{\min}}^{f_{\max}} W(f) df \right\}
\]

(1)

where \( P_e \) is the “probe” signal level at threshold, \( N_s \) is
Table 1: Parameters derived from the mean simultaneous masking data.

<table>
<thead>
<tr>
<th>Frequency $f_c$ (Hz)</th>
<th>$p_l$</th>
<th>$t$</th>
<th>$p_u$</th>
<th>$G_{max}$</th>
<th>$K$ (dB)</th>
<th>$P_0$ (dB)</th>
<th>RMS error (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>22.0</td>
<td>5.01</td>
<td>20.9</td>
<td>41.5</td>
<td>-3.02</td>
<td>12.0</td>
<td>1.34</td>
</tr>
<tr>
<td>1000</td>
<td>29.1</td>
<td>5.41</td>
<td>26.0</td>
<td>49.4</td>
<td>-1.62</td>
<td>7.79</td>
<td>1.55</td>
</tr>
<tr>
<td>2000</td>
<td>41.3</td>
<td>9.69</td>
<td>25.3</td>
<td>45.1</td>
<td>-1.62</td>
<td>4.72</td>
<td>1.24</td>
</tr>
<tr>
<td>4000</td>
<td>24.2</td>
<td>3.83</td>
<td>23.4</td>
<td>57.7</td>
<td>0.698</td>
<td>10.5</td>
<td>1.79</td>
</tr>
</tbody>
</table>

Table 2: Parameters derived from the mean forward masking data.

<table>
<thead>
<tr>
<th>Frequency $f_c$ (Hz)</th>
<th>$p_l$</th>
<th>$t$</th>
<th>$p_u$</th>
<th>$G_{max}$</th>
<th>$K$ (dB)</th>
<th>$P_0$ (dB)</th>
<th>RMS error (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>20.5</td>
<td>3.37</td>
<td>48.3</td>
<td>39.3</td>
<td>-42.0</td>
<td>28.2</td>
<td>2.05</td>
</tr>
<tr>
<td>1000</td>
<td>26.0</td>
<td>7.72</td>
<td>44.0</td>
<td>33.6</td>
<td>-49.9</td>
<td>18.5</td>
<td>2.36</td>
</tr>
<tr>
<td>2000</td>
<td>32.7</td>
<td>5.04</td>
<td>75.1</td>
<td>41.3</td>
<td>-54.7</td>
<td>13.0</td>
<td>2.08</td>
</tr>
<tr>
<td>4000</td>
<td>30.8</td>
<td>9.10</td>
<td>108</td>
<td>30.8</td>
<td>-55.4</td>
<td>13.6</td>
<td>1.61</td>
</tr>
</tbody>
</table>

the spectrum level of the masker, and $K$ is a constant which is related to the efficiency of the detection mechanism following the auditory filter. The limits on the filter integrals were from $f_{i,\text{min}}$ to $f_{i,\text{max}}$ for the lower noise band and from $f_{u,\text{min}}$ to $f_{u,\text{max}}$ for the upper noise band. This model is often referred to as a “power spectrum model of masking”, as it simply assumes that the fluctuations within the noise bands can be ignored.

### 3.2 roex auditory filter

The standard double roex (rounded-exponential) filter, roex($p, w, t$), proposed by Glasberg and Moore (2000) [4], is used to estimate the shape of the auditory filter. $W(f)$ in Eq. (1) is represented as

$$W(g) = \begin{cases} (1-w)(1+p,g)\exp(-p,g) \\ + w(1+tg)\exp(-tg), & f \leq f_i \\ (1+p,g)\exp(-p,g), & f > f_i \end{cases}$$

where the normalized frequency is defined as $g = |f - f_i|/f_i$, and $w$ is relative gain term defined as $w = 1/(G_{\text{max}} + 1)$. The gain of the tip filter is relative to a schematic I/O function, which models the I/O function of the basilar membrane. As the gain of tip filter varies with input level, the gain of the I/O function, $G_{\text{max}}$, is defined as

$$G_{\text{max}} = 0.9L + A + B\left(1 - \frac{1}{1 + \exp(-0.05(L - 50))}\right) - L$$

where $L$ is the input signal level, $A = -0.0894G_{\text{max}} + 10.894$, $B = 1.1789G_{\text{max}} - 11.789$, and $G_{\text{max}}$ is parameter that can determine filter characteristics of the I/O function. Subscripts of “lin” and “dB” denote linear- and log-scale, and $G_{\text{max}} = 10^{G_{\text{max}}/10}$.

### 3.3 Fitting procedure

A total six parameters, inclusive of four parameter ($p_l$, $t$, $G_{\text{max}}$, and $p_u$) of the double roex auditory filter and two non-filter parameters ($K$ and $P_0$), were determined by using the PolyFit procedure [1, 4]. The parameter $P_0$ was used in the form of $10\log_{10}(10^{P_0/10} + 10^{5.6/10})$ to estimate the masked thresholds by the auditory filter. In this procedure, precochlear processing (MidEar correct) was used in the same way as in [4], to incorporate the effects of outer and middle ear.

The influence of off-frequency listening effect was included by locating the auditory filter that produces the best signal-to-noise ratio when estimating the thresholds. There parameters were optimized by minimizing the root mean square (rms) error between the masked thresholds and the predicted thresholds of the auditory filter.

### 3.4 Result

The mean data for both simultaneous and forward masking experiments were used to estimate auditory filter shapes. In this paper, the shape of the filter was estimated in the same manner as in [1, 2, 4, 9] using the fitting procedure, described in the above.
Figure 3: Mean masked thresholds in the simultaneous masking with notched-noise masker for (a) 500 Hz, (b) 1000 Hz, (c) 2000 Hz, and (d) 4000 Hz.

Figure 4: Mean masked thresholds in the forward masking with notched-noise masker for (a) 500 Hz, (b) 1000 Hz, (c) 2000 Hz, and (d) 4000 Hz.

The optimized values for the six parameters of the double roex filter and the rms error at each signal frequency fitted to the mean data of simultaneous masking and forward masking were tabulated in Tables 1 and 2 respectively. Similarly, the thresholds estimated by these fitted auditory filters are shown in Figs. 3 and 4 correspond to the parameters in Tables 1 and 2. The solid line shows the estimated thresholds in symmetric notch condition. Dotted and dashed lines show the estimated thresholds in the asymmetric notch conditions, which is accompanied by the right- and left-pointing triangles to denote the different $\Delta f'$, $f$ of upper and lower band as shown in section 2.4.

Figure 5: Auditory filter shapes with center frequencies between 500 and 4000 Hz, derived from the mean data in the simultaneous masking conditions (solid curves) and the mean data in the forward masking (dashed curves).

The shapes of the fitted auditory filters centered at the signal frequencies of 0.5, 1.0, 2.0, and 4.0 kHz were plotted in Fig. 5, as a function of the signal level (10, 20, and 30 dB SL). Signal was set to the levels of 10, 20, and 30 dB above the mean absolute thresholds correspondingly. In this figure, solid and dashed lines showed the resulting derived auditory filter shapes from simultaneous and forward masking data. Each auditory filter function provided good fits to the respective data. As expected, the shapes of the filters derived from forward masking data has a sharper tuning than those derived from simultaneous masking data.

ERBs were derived for each individual auditory filter at different dB SL’s. When signal level is at 10 dB SL, the ERBs at centre frequencies 500, 1000, 2000, and 4000 Hz with forward masking data were 71, 124, 176, and 340 Hz, while the corresponding values with simultaneous masking data were 94, 145, 252, and 670 Hz. When signal level increases to 30 dB SL, ERBs with forward masking data were 81, 130, 180, and 361 Hz, at the same center frequencies, while the values with simultaneous masking data were 100, 146, 253, and 671.
On the whole, ERBs derived from the forward masking data are smaller than those derived from the simultaneous masking data, indicating the tuning of the filter derived from the forward masking data is sharper. When signal level increases in dBSL, ERBs from both sets of data increase, but the values derived from the forward masking data are still smaller in values. At higher signal frequencies, the ERBs from the forward masking data are relatively much smaller. The tuning of the filter derived from the forward masking data seems to become sharper as the signal frequency increases.

To further analyze the sharpness in the tuning for these two sets of auditory filters, the ratios of the ERBs from the simultaneous masking data to those corresponding ERBs from the forward masking data were calculated. The ratios at 10 dBSL were not consistent at different center frequencies, indicating that the degree of sharpening the tuning of the auditory filter by the forward masking data is not equally done at different center frequency. When the ratios were recalculated at 30 dBSL, a similar observation is found. However, the range of the ratios (1.12 to 1.86) at 30 dBSL is smaller than the range of ratios (1.17 to 1.97) at 10dBSL. The sharpening of the auditory filter seems to becoming slightly more consistent at all signal frequencies in quest by the forward masking data.

4 Summary

This paper estimated auditory filter shapes using both simultaneous and forward masking with notched noise masker. The auditory filter derived from forward masking data demonstrated a sharper tuning than those derived from simultaneous masking data. This study also showed that the tuning of the filter becomes sharper as the signal frequency increases. However, the ratios of the ERBs seem to suggest that there is no consistent degree of sharpening the auditory filter at different signal frequencies by forward masking data.

The differences between the auditory filters derived from the forward masking data and the simultaneous masking data might have demonstrated the need of incorporating more realistic suppression mechanism into the auditory filter function. Furthermore, beyond the context of this paper, a trend of decreasing $K$ values was observed in forward masking data, which could be related to the decay of forward masking effect. Incorporating temporal integration window might be necessary in a further study with forward masking experiments.

Acknowledgements

This work was supported by special coordination funds for promoting science and technology (supporting young researchers with fixed-term appointments).

References


