Improving auditory filter estimation with level-dependent cochlear noise floor

Toshio Irino[1,*], Kenji Yokota[1], and Roy D. Patterson[2]

[1]Faculty of Systems Engineering, Wakayama University, 930 Sakaedani Wakayama, 640-8510, Japan
[2]Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street Cambridge, CB2 3EG, UK
[*]Corresponding author: irino@wakayama-u.ac.jp

Abstract

When developing models of human hearing for communication devices, it is important to have an accurate representation of auditory filter (AF) shape. AF shape has been traditionally estimated by the combination of notched-noise (NN) masking experiment and power spectrum model (PSM) of masking. AFs of hearing impaired (HI) listeners were sometimes estimated extremely broader than ones expected from physiological observation when NN thresholds rapidly converged onto the absolute threshold (AT) as notch width increases. The overestimation happened probably because the conventional PSM does not adequately include the effect of the cochlear noise floor associated with AT. This paper tried to clarify and solve the problem through NN measurements and a new formula of the PSM.

We measured a detailed set of NN threshold values for normal-hearing (NH) listeners, including low-level noises at four center frequencies (500, 1000, 2000, and 4000 Hz) to show how threshold converges onto the AT as notch width increases at low noise levels. We incorporated AT into the PSM for the AF shape estimation by introducing the level-dependent cochlear noise floor, $N_c^{(LD)}$, which is a function of the NN masker level from the base level directly calculated from the hearing level (HL) of 0 dB. We estimated the AF shapes for the four center frequencies simultaneously and compared the $N_c^{(LD)}$ model with a fixed noise floor model, $N_c^{(Fx)}$ and the conventional P_0 model in which an arbitrarily constant, P_0, had been introduced to represent the low-level threshold limit. The $N_c^{(LD)}$ model provided an excellent fit and a major reduction in the rms error of the AF shape estimation when comparing the P_0 and $N_c^{(Fx)}$ models. We also examined the frequency distribution of the cochlear noise floor in quiet, which provides the basis of the AT and AF shape estimation. It was found the frequency distribution associated with the HL of 0 dB was optimal regardless of the frequency dependency for the detector SNR, K, in the PSM. It implies that the AT can be explained by this noise floor in quiet.

1 Introduction

When developing models of human hearing for communication devices, it is important to have an accurate representation of auditory filter (AF) shape. Traditionally, the shape is estimated using a notched-noise (NN) experiment in which threshold for a sinusoidal signal is measured in the presence of a broad band of noise (Moore, 2012; Patterson, 1976; Patterson, Unoki, &

Irino, 2003). A notch is created in the noise around the signal frequency, f_s, and tone threshold is repeatedly measured as the width of the notch is increased. The resultant NN threshold function is assumed to provide an estimate of the shape of the integral of the auditory filter at that signal frequency. The shape of the filter itself is derived from the threshold function using a relatively simple power spectrum model (PSM) (Fletcher, 1940) of tone-in-noise masking.

As notch width continues to increase, the descent of NN threshold is eventually limited by absolute threshold (AT). Thus, the curve that describes AT as a function of signal frequency, AT(f_s) , also describes the lower boundary for NN threshold as a function of f_s. For young normal listeners, the form of AT(f_s) is well known; it is codified in ANSI hearing-level standard (ANSI_S3.6-2010, 2010) as HL-0dB (Hearing Level for 0 dB). This suggests that the PSM of masking used to derive an AF shape from a set of NN thresholds should include the constraints imposed by AT.

Moore (2012) pointed out that AFs of hearing impaired (HI) listeners differed considerably with the individual. Some of them had extremely broad filters; others had filters with the opposite asymmetry to those typically observed (Glasberg & Moore, 1986). But there was no clear relationship between AF shape and the broadly tuned component of the cochlear traveling wave observed physiologically (Pickles, 2013; von Békésy & Peake, 1990). Often with these HI listeners, the sound pressure level (SPL) of the NN masker was not far above their elevated AT value and NN threshold converges onto AT at a relatively narrow notch width. The conventional PSM does not include the effect of AT, and in such cases the bandwidth of the AF is likely to be overestimated. The incorporation of AT into the PSM would increase the stability of the AF fit and reduce the number of free coefficients required for a good fit.

This paper shows that AT(f_s) can be incorporated into the PSM by assuming that there is a broad-band noise floor in the cochlea which combines with the NN masker as it appears in the cochlea, and together they determine the threshold value observed in any given condition of the experiment. It is also demonstrated that the level-dependency of the noise floor, which may vary with NN masker levels, plays an important role in the AF shape estimation. Moreover, it is necessary to know the wide-range distribution of the noise floor when we estimate the filter shapes for various center frequencies simultaneously(Patterson, Unoki, & Irino, 2003). Buss et al. (2016) reported that the frequency distribution of the internal "self-generating noise" of NH listeners is similar to the 0-dB HL function(ANSI_S3.6-2010, 2010) on a dB scale. Although it gives the first-order approximation, it was not proved to be optimum in the AF shape estimation. It is another question to be answered in this paper.

This paper is organized as follows. Section 2 presents a large NN experiment performed with NH listeners and a high proportion of wide notch conditions and low NN levels, to produce a detailed record of how AT interacts with NN threshold. Section 3 shows how the PSM of masking was extended to include the noise floor dictated by AT. Section 4 presents a quantitative comparison of the conventional and extended PSMs for AF shape estimation. Finally, in Section 5, we confirm that the frequency distribution of the noise floor is reasonable in the extended PSM.

2 Experiment

The NN experiment in this study is similar to those in the more recent AF studies (e.g., Baker and Rosen (2002), Glasberg and Moore (2000); see Moore (2012), for details). NN threshold for a sinusoidal signal (0.5, 1.0, 2.0 or 4.0 kHz) was repeatedly measured in the presence of a NN

masker using an adaptive, two-alternative, forced-choice procedure (Levitt, 1971). The main difference in the design of the current experiment was the inclusion of masker conditions with low spectrum levels, where AT is observed to limit the descent of NN threshold in wide notches.

2.1 Notched-noise conditions

The signal frequencies (f_s) were 0.5, 1.0, 2.0 and 4.0 kHz. The normalized frequency distances from the signal to the nearer edges of the lower and upper noise bands $\{\Delta f_l/f_s, \Delta f_u/f_s\}$ were $\{0, 0; 0.1, 0.1; 0.2, 0.2; 0.4, 0.4; 0.3, 0.5; 0.5, 0.3\}$. The same notches were used at each spectrum level. The bandwidth of the noise was 0.4 of the normalized signal frequency. The spectrum levels (N_0) were $\{38, 28, 18, 8, -2, -12\}$ dB when $f_s = 2.0$ kHz and $\{40, 30, 20, 10, 0, -10\}$ dB when f_s was 0.5, 1.0, or 4.0 kHz. At each signal frequency, threshold was measured for six noise spectrum levels in a random order[1].

2.2 Listeners

The experiment was performed with NH listeners rather than HI listeners because their audiograms exhibit less variability and they were willing to participate in the long sessions required to collect NN thresholds for such a wide range of SPLs and notch widths.

In total, 26 NH listeners participated in the experiment; they ranged from 19 - 28 years old; there were 14 males and 12 females. They all had hearing levels (HLs) less than 20 dB between 125 and 8000 Hz. To make up the four groups of 8 observers required by the design, one man and one woman participated at two of the four signal frequencies, and two different men participated at three signal frequencies. The experiment was approved by the local ethics committee of Wakayama University and all of the listeners provided informed consent before participating in the experiment.

2.3 Signal generation and measurement procedure

The sinusoidal signals and the NN maskers were generated digitally at a sampling rate of 48 kHz with 24-bit resolution using MATLAB 2017a on a Mac mini with MacOS 10.12. The signal and the masking noise had the same, 200-ms, duration. The onset and offset were rounded with the rise and fall of a 10-ms hanning window. The stimuli were presented over headphones (PM-1, OPPO) via a USB interface (HA-1, OPPO) at a 48-kHz sampling rate and 24-bit resolution. The listeners were seated in a sound attenuated room (RION AT62W). The headphone levels were calibrated with a sound level meter (Type 2250-L, Bruël & Kjær) and an artificial ear (Type 4153, Bruël & Kjær).

Signal threshold was measured using a two-interval, two-alternative, forced-choice procedure and the transformed up-down method of Levitt (1971). In one interval, the masker was presented on its own; in the other, the signal and masker were presented simultaneously. Listeners were asked to select the interval containing the signal using a graphical user interface. Feedback regarding the correct answer was indicated visually after the listener's response. There was a brief training session lasting about 20 minutes to familiarize the listener with the threshold procedure.

[1]The 2.0-kHz f_s condition was performed first. The conditions with the lowest masker levels (-10 or -12 dB) were measured separately, after those for the other five noise levels when it became clear that the interaction of NN threshold with AT continued below 0 dB

2.4 Results

For the four signal frequencies, Figure 1 shows average NN threshold for the eight listeners at the six masker levels (solid lines), along with their average AT (dashed line). The thresholds associated with the two highest noise levels, 30 and 40 dB[2], remain well above AT out to the widest notches. At lower noise levels (20, 10 and 0), however, threshold is limited by the proximity of AT, and NN threshold at the -10 dB noise level converges onto AT at the wider notch widths. The set of curves shows that NN threshold does eventually converge onto AT at all signal frequencies. This in turn suggests that NN threshold should be assumed to converge onto AT in the PSM of masking. In earlier studies, although AT was routinely measured, it was not included in the data set used to derive the shape and gain of the auditory filter, nor was AT directly represented in the power-spectrum model used to derive filter shape and filter gain.

3 Extension of the power spectrum model of masking

In the notched noise (NN) experiments, the PSM was used for estimating signal threshold, $\hat{P}_s$ (on a dB scale), with the following equations:

$$\hat{P}'_s = K' + 10\log_{10}\hat{P}_{ext}, \tag{1}$$

$$\hat{P}_{ext} = \int_{f_{a_{min}}}^{f_{a_{max}}} N_0(f)\cdot W(f)df, \tag{2}$$

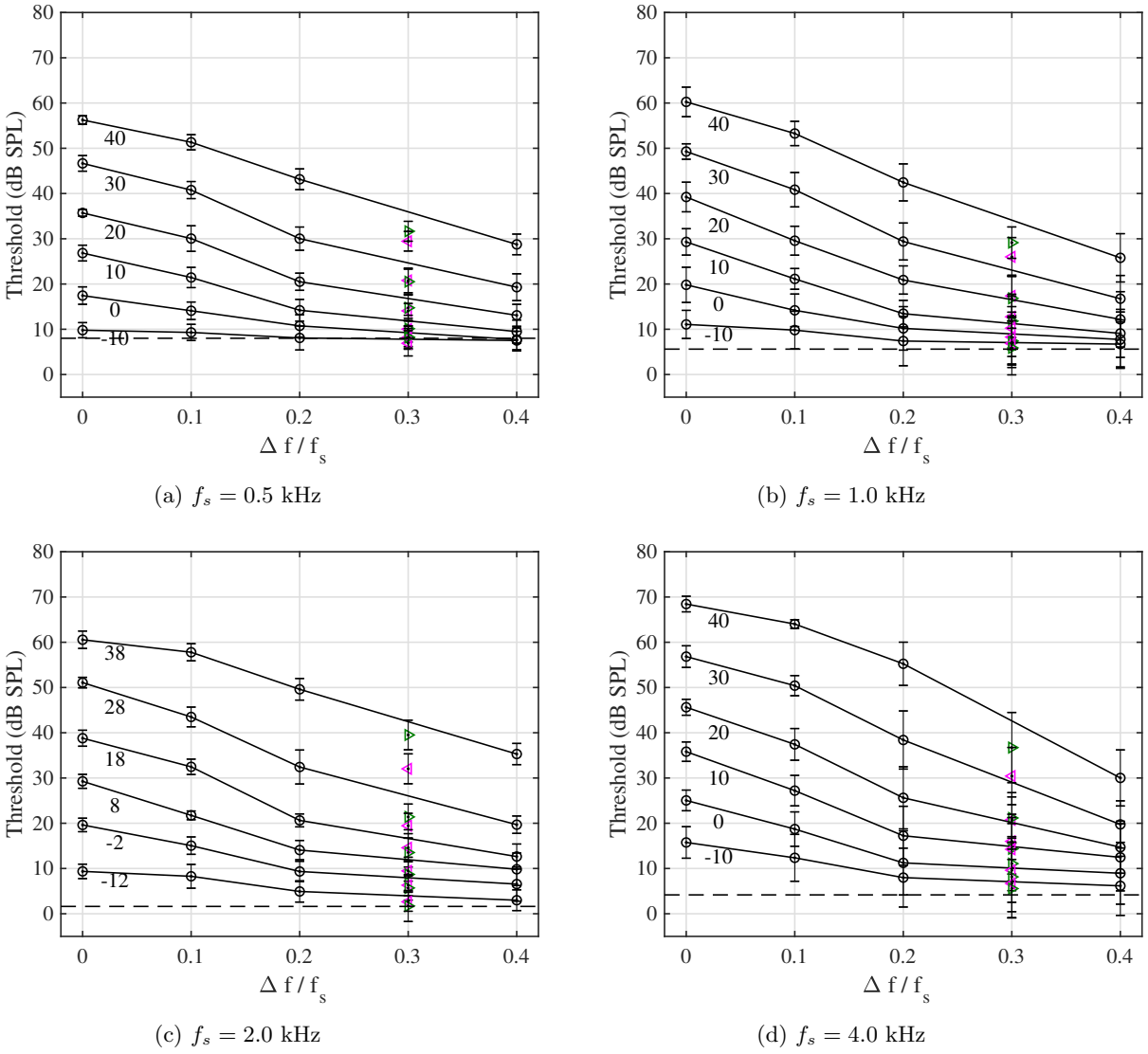
$$N_0(f) = \begin{cases} N_0\cdot|T(f)|^2 & \text{when } \{f|f_{l_{min}}\le f\le f_{l_{max}}, f_{u_{min}}\le f\le f_{u_{max}}\} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where K' is the signal-to-noise ratio (SNR) at the output of the auditory filter and $\hat{P}_{ext}$ (on a linear scale) is an estimate of the external noise that passes through the auditory filter. Note that, to avoid confusion, the parameters with a prime (e.g.,K', $\hat{P}'_s$) represent level on a dB scale hereafter. $N_0(f)$ is the spectrum level of the noise, i.e., power density function, and $W(f)$ is the power weighting function of the auditory filter. $f_{a_{min}}$ and $f_{a_{max}}$ are the cochlear frequency range [3]. $f_{l_{min}}$, $f_{l_{max}}$, $f_{u_{min}}$, and $f_{u_{max}}$ specify lower and upper noise bands of NN. In Eq. 3, $T(f)$ is the transfer function of sound from the audio device to the input of the cochlea. It is dependent on how sounds are delivered (e.g., free field or headphone) and where the noise level is defined, which in this study was the ear drum. $T(f)$ was the transfer function of the middle ear, $T_{mid}(f)$ (Aibara et al., 2001; Glasberg & Moore, 2006; Puria, Peake, & Rosowski, 1997) as shown in Fig. 2(b).

When the auditory filter is modeled with the compressive gammachirp (GC), $G_C(f)$, as described in Appendix A (see also Irino and Patterson (2001) and Patterson, Unoki, and Irino (2003)), the filter weighting function, $W(f)$ becomes $|G_C(f)|^2$. The GC filter was used as the filter function in the current paper because it provides a better representation of the level-dependence and compression of the auditory filter than the conventional roex filter (Patterson & Nimmo-Smith, 1980); moreover, the GC filter requires fewer parameters (Unoki et al., 2006).

[2] At 2 kHz, they are 28 and 38 dB which are -2 dB less than those appear in this subsection.

[3] In the current simulation, $f_{a_{min}}$ = 100Hz and $f_{a_{max}}$ = 12kHz because the gains of the GC filters between 0.5 kHz and 4 kHz are sufficiently small beyond this range. Another reason is to avoid ill-defined noise level at low and high frequencies affecting estimation of the noise floor at the reference frequency, 1 kHz.



(a) $f_s = 0.5$ kHz

(b) $f_s = 1.0$ kHz

(c) $f_s = 2.0$ kHz

(d) $f_s = 4.0$ kHz

Figure 1 — Average NN threshold (solid lines) for eight listeners, and their average AT (dashed line). The signal frequencies (f_s) were 0.5 kHz (a), 1.0 kHz (b), 2.0 kHz (c), and 4.0 kHz (d). The abscissa is normalized notch width ($\Delta f/f_s$). The circles (○) show symmetric notch conditions; the right-pointing triangles (▷), at $\Delta f/f_s = 0.3$, show conditions with additional shifting of the upper noise band by 0.2; the left-pointing triangles (◁), at $\Delta f/f_s = 0.3$, show conditions with additional shifting of the lower noise band by 0.2. The parameter beneath each threshold curve is noise spectrum level which was the same for the lower and upper bands throughout the experiment. The noise levels for the triangles are the same as for the threshold curves just above them.

3.1 Incorporating absolute threshold into the estimation of NN threshold

In conventional NN experiments, the notched noise level is well above absolute threshold (AT), so AT can be ignored in the derivation of AF shape. The PSM was extended to include AT by assuming that there is an internal noise floor that limits NN threshold and the power at the

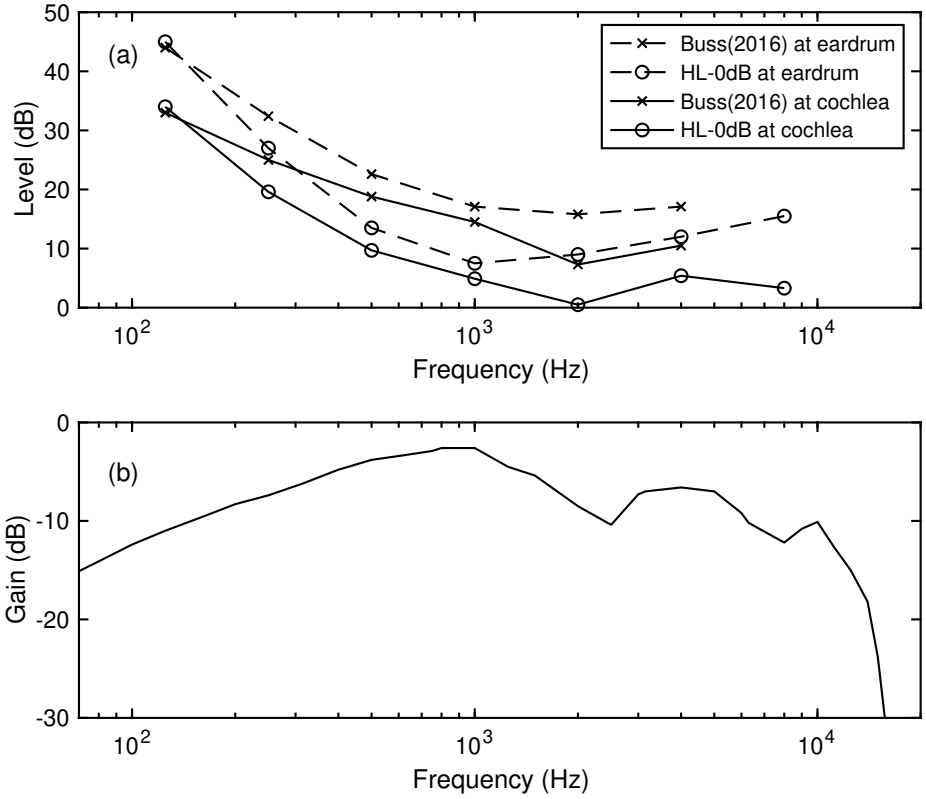


Figure 2 — (a) Relationship between Self-generating noise (Buss et al., 2016) and HL-0dB function at the ear drum (dashed lines) and at the cochlear input (solide lines). (b) The middle ear transfer function, $T_{mid}(f)$ (Glasberg & Moore, 2006) for compensating between them.

output of the corresponding AF is

$$\hat{P}_{int} = \int_{f_{a_{min}}}^{f_{a_{max}}} N_c(f) \cdot W(f) df, \tag{4}$$

where $N_c(f)$ is the spectrum level of the internal noise floor. NN threshold depends on both the internal noise $\hat{P}_{int}$ in Eq.4 and the external noise $\hat{P}_{ext}$ in Eq.2. so that NN threshold can be predicted as

$$\hat{P}'_s = K' + 10\log_{10}(\hat{P}_{int} + \hat{P}_{ext}). \tag{5}$$

Absolute threshold, $\hat{P}_{abs}$, can be estimated when the noise floor level $N_c(f)$ is equal to $N_c^{(Q)}(f)$ which is the level in quiet, i.e., $\hat{P}_{ext} = 0$. Then AT can be estimated with the PSM as

$$\hat{P}'_{abs} = K' + 10\log_{10}\left\{\int_{f_{a_{min}}}^{f_{a_{max}}} N_c^{(Q)}(f) \cdot W(f) df\right\}. \tag{6}$$

3.2 Frequency distribution of the cochlear noise floor

We would like to know whether AT is completely determined by the internal noise floor in quiet, $N_c^{(Q)}(f)$, or whether some other factor is also involved. Buss et al. (2016) reported that

the distribution of internal "self-generated noise" of NH listeners is similar to the 0-dB HL function(ANSI_S3.6-2010, 2010) on a dB scale as shown in Fig.2(a). If we assume that the distribution of the cochlear noise floor, $N_c^{(Q)}(f)$, is indeed the 0-dB HL function, $L_{HL0}(f)$, it can be represented as

$$N_c^{(Q)}(f) = N_c^{(Q)}(f_{ref}) \cdot |T_{mid}(f)|^2 \cdot \frac{L_{HL0}(f)}{L_{HL0}(f_{ref})} \cdot \frac{ERB_N(f_{ref})}{ERB_N(f)}. \quad (7)$$

The noise floor, $N_c^{(Q)}(f)$, is a spectral power density function and should be multiplied by the AF bandwidth, $ERB_N(f)$, to convert a linear power function, $L_{HL0}(f)$. The function is normalized at a reference frequency, f_{ref}, which is 1 kHz in this case. $T_{mid}(f)$ is the middle ear transfer function shown in Fig. 2(b). The noise floor, $N_c^{(Q)}(f)$, is uniquely determined by a constant $N_c^{(Q)}(f_{ref})$ which is involved in the filter estimation process. The rationale for this definition will be presented in Section 5.

3.3 Level dependence of the noise floor

Irino et al. (2018) assumed that the cochlear noise floor, $N_c(f)$, in Eq. 4 would be dependent on the level of the external NN, in which case, $N_c(f)$ should be greater than $N_c^{(Q)}(f)$ in Eq. 7. This is because distortion products would be generated by cochlear nonlinearity (e.g. Gaskill and Brown (1990) and Hall (1972)) and its distribution could spread widely even beyond the frequency regions of the NN. They found that the estimation error was significantly reduced when the noise floor was made level dependent $N_c^{(LD)}(f)$. In this paper, on a dB scale, it ${N_c}^{(LD)'}(f)$ was defined as

$$\begin{aligned} {N_c}^{(LD)'}(f) &= {N_c}^{(Q)'}(f) + n_{LD} \cdot (N_0'(f) - {N_c}^{(Q)'}(f)) & (8) \\ &= (1 - n_{LD}) \cdot {N_c}^{(Q)'}(f) + n_{LD} \cdot N_0'(f). & (9) \end{aligned}$$

where ${N_c}^{(Q)'}(f)$ is the noise floor in quiet, i.e., when there is no external sound. Equation 9 means that the noise floor increases from its quiet level as the external noise level, $N_0'(f)$, increases [4]. The proportionality coefficient for the level dependence is n_{LD} in dB/dB. This equation is a revised version of Eq. 9 in Irino et al. (2018) with one less parameter. It increases the stability of filter estimation. The noise level becomes a value which is linearly interpolated with the ratio of $n_{LD} : (1 - n_{LD})$ between ${N_c}^{(Q)'}(f)$ and $N_0'(f)$. This model will be referred to as the "$N_c^{(LD)}$ model" in what follows. When $n_{LD} = 0$, it becomes a level-independent, fixed function, ${N_c}^{(Q)'}(f)$, which will be referred to as the "$N_c^{(Fx)}$ model."

3.4 Estimation of the filter shape

The coefficients of the auditory filter were estimated using a least-squares method (Moré, 1978) to minimize the error between measured and predicted NN thresholds (P_s' and $\hat{P}_s'$ in Eq. 5)

[4] We assume this simple formula with frequency-independent n_{LD} is sufficient to support a first order approximation of the level dependence. There are several aspects of the noise floor which need to be considered for accurate simulation. Firstly, the external noise (NN) is bandpass noise with width $2 \times 0.4\Delta f_s/f_s$. Although it is roughly constant on a logarithmic frequency axis, the location of the noise band in the NN condition may affect the noise floor. Secondly, the transfer function $T(f)$ from the headphones to the cochlea may also affect the noise floor, although it is invariant across noise level. We assumed these factors are relatively small with respect to the effect of $\tilde{N}_0$ level change. In any case, we do not know the exact characteristics of the noise currently and it would be difficult to take all of the factors into account. This led us to use this simple formula in this study.

and the error between the measured and predicted ATs (P'_{abs}, and $\hat{P}'_{abs}$ in Eq. 6) [5]. Namely,

$$\mathbf{c}_{\mathbf{gc}}^{(\mathbf{N_c})} = \underset{\mathbf{c_{gc}}}{\arg\min} \left\{ \frac{1}{N} \sum_{i=1}^{N} (P'_{s_i} - \hat{P}'_{s_i})^2 + (P'_{abs} - \hat{P}'_{abs})^2 \right\}, \tag{10}$$

where $\mathbf{c}_{\mathbf{gc}}^{(\mathbf{N_c})}$ is a vector of the GC coefficients, $\{b_1, c_1, f_{rat}^{(0)}, f_{rat}^{(1)}, b_2, c_2\}$ (see Appendix A), plus the constants $\{K', N_c^{(Q)'}(f_{ref}), n_{LD}\}$.

3.5 Conventional $\mathbf{P_0}$ threshold limit

In the NN experiment, threshold asymptotes to a low level somewhat above absolute threshold even when the NN level is relatively high (Patterson & Nimmo-Smith, 1980) (see Fig. 1). Glasberg and Moore (2000) introduced a term, P_0, to represent the lower limit of NN threshold and prevent it from distorting the representation of the tails of the auditory filter. In this case,

$$\hat{P}_s'^{(P_0)} = 10 \log_{10} \left\{ 10^{\hat{P}'_s/10} + 10^{P'_0/10} \right\} \tag{11}$$

$$= 10 \log_{10} \left\{ 10^{(K' + \hat{P}'_{ext})/10} + 10^{P'_0/10} \right\} \tag{12}$$

The coefficients of the auditory filter were estimated using the least-squares method to minimize the error between the measured thresholds, P_s, and the thresholds predicted by the model, $\hat{P}_s$; that is

$$\mathbf{c}_{gc}^{(P_0)} = \underset{\mathbf{c}_{gc}}{\arg\min} \left\{ \frac{1}{N} \sum_{i=1}^{N} (P'_{s_i} - \hat{P}_{s_i}^{(P_0)'})^2 \right\} \tag{13}$$

where $\mathbf{c}_{gc}^{(P_0)}$ is a vector of the GC coefficients, $\{b_1, c_1, f_{rat}^{(0)}, f_{rat}^{(1)}, b_2, c_2\}$, plus the constants $\{K \& P_0\}$. Glasberg and Moore (2000) showed that the use of P_0 is effective in reducing estimation error. They suggested that P_0 is related to AT but they did not explain the relationship in detail. This model will be referred to as the "P_0 model" in what follows and is used as a conventional model to compare the performance with the $N_c^{(LD)}$ model.

4 The effect of a level-dependent noise floor

The $N_c^{(LD)}$, $N_c^{(Fx)}$, and P_0 models were compared to evaluate the effect of level dependence on the goodness of the filter estimation.

4.1 Procedure

The auditory filters of 500, 1000, 2000, and 4000 Hz were simultaneously estimated by using all of the 144 thresholds (= 36 × 4 probe frequencies) shown in Fig. 1. It is similar to the global fitting with P_0 in Patterson, Unoki, and Irino (2003). They reported that filter shape can be accurately determined using a GC filter with six, frequency independent coefficients and two non-filter parameters P_0 and K which were quadratic functions of frequency. The number of

[5] We also introduced several constraints to improve the stability of the fitting process and to restrict the filter shape coefficients to a reasonable range. There were limits on the GC coefficients, the bandwidth, and the slope of the IO function. The constraints were introduced as error terms with small weighting values.

Table 1 — Number of coefficients in selected fits of the $N_c^{(LD)}$, $N_c^{(Fx)}$, and P_0 (Patterson, Unoki, & Irino, 2003) models with the rms errors of NN threshold and AT that each model produced.

	Number of coefficients						NN error	AT error
model	total	GC	K	$N_c^{(Q)}$	n_{LD}	P_0	(dB)	(dB)
$N_c^{(LD)}$	11	6	3	1	1	-	1.64	2.23
	10	6	2	1	1	-	1.68	2.10
	9	6	1	1	1	-	1.65	2.58
$N_c^{(Fx)}$	10	6	3	1	-	-	2.66	3.16
	9	6	2	1	-	-	2.64	2.45
	8	6	1	1	-	-	2.62	2.61
P_0	12	6	3	-	-	3	2.40	5.05

coefficients for each parameter is listed in the bottom part of Table 1 and for the P_0 model was 12 in total. Based on this knowledge, we set the number of coefficients for the $N_c^{(LD)}$ and $N_c^{(Fx)}$ models as shown in Table 1. The number of GC filter coefficients was the same (6). $N_c^{(Q)'}(f_{ref})$ in Eq. 7 and n_{LD} in Eq. 9 were set to constants (1). The SNR at the output of the auditory filter, K, was set to a constant, a linear function or a quadratic function of the normalized frequency, E_f, defined as

$$E_f = \frac{ERB_{N_{number}}(f)}{ERB_{N_{number}}(f_{ref})} - 1, \tag{14}$$

$ERB_{N_{number}}(f)$ is the number of equal rectangular bands in Cam (Moore, 2012) and $f_{ref} =$ 1 kHz.

In the estimation, each model was fitted to the 144 thresholds 10 times, using different initial values for the GC coefficients, chosen randomly within a range ±20% of the summary coefficient values reported in (Patterson, Unoki, & Irino, 2003). The best of the 10 filter set was selected as the one that minimized the rms error of the NN threshold.

4.2 Results

4.2.1 Estimation error

The right two columns of Table 1 show the rms estimation errors of NN threshold and AT. The table shows that the NN threshold errors of the $N_c^{(LD)}$ model were between 1.64 dB and 1.68 dB, 0.9 dB and 0.7 dB, smaller than those of the $N_c^{(Fx)}$ and P_0 models, respectively. The $N_c^{(LD)}$ model also requires fewer coefficients than the P_0 model. The AT errors of the $N_c^{(LD)}$ model were also much smaller than those of the P_0 model.

Thus, the introduction of the level dependence coefficient, n_{LD}, in Eq. 9 reduces the estimation error of the NN threshold, which implies that the cochlear noise floor is dependent on the external noise (NN) level, and the estimation of AF shape should take this into account.

With the $N_c^{(LD)}$ model, NN threshold error is effectively independent of the form of K; the version with 9 coefficients is as effective as the one with 11 coefficients. In other words a $N_c^{(LD)}$ model with a fixed K is sufficient to explain the NN threshold data.

This result also suggests that the distribution of the HL-0dB threshold largely explains the frequency dependence which necessitated the frequency dependent terms associated with the

Table 2 — Values from the global fit for the nine-coefficient $N_c^{(LD)}$ model and the twelve-coefficient P_0 model. E_f is normalized frequency as defined in Eq. 14. The NN threshold and AT errors are listed in the last two columns. The bottom row, $P_0^{(PUI)}$ shows the values for the P_0 model reported by Patterson, Unoki, and Irino (2003). The NN error with the asterisk cannot be fairly compared with the two above it because the NN threshold data were from a different experiment.

model	No.	Filter coefficients						Non-filter coefficients			NN err	AT err
	coeff	b_1	c_1	$f_{rat}^{(0)}$	$f_{rat}^{(1)}$	b_2	c_2	K	$N_c^{(Q)}/P_0$	n_{LD}	(dB)	(dB)
$N_c^{(LD)}$	9	2.17	-2.76	0.676	0.0085	1.79	2.77	-4.49	-20.30	0.20	1.65	2.58
P_0	12	2.26	-2.07	0.536	0.0099	2.23	2.77	-5.75 $0.076E_f$ $+1.48E_f^2$	2.64 $-8.01E_f$ $+9.50E_f^2$	-	2.40	5.05
$P_0^{(PUI)}$	12	1.81	-2.96	0.467	0.0109	2.17	2.20	-3.73 $-4.89E_f$ $+8.30E_f^2$	16.80 $-1.27E_f$ $+5.74E_f^2$	-	3.71*	-

non-filter parameters K and P_0. Indeed, it suggests that the arbitrary coefficient P_0 is not required for accurate AF shape estimation.

4.2.2 Filter coefficients

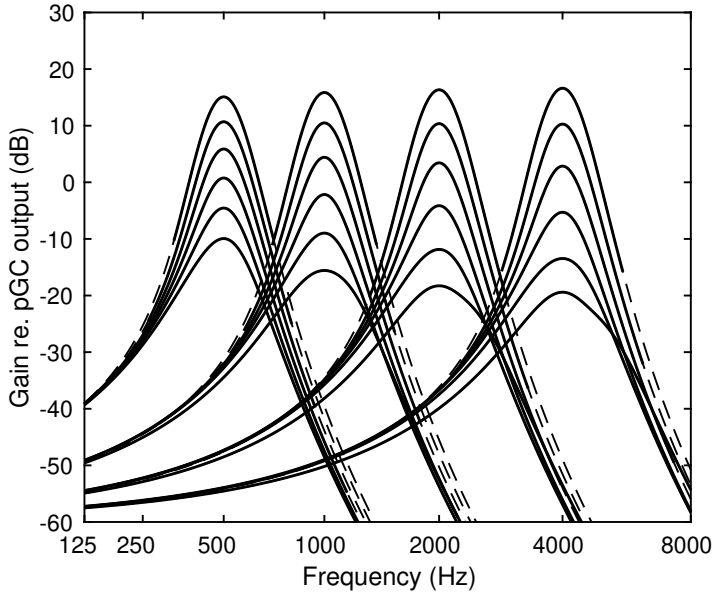
The filter coefficients of the nine-coefficient $N_c^{(LD)}$ model and the twelve-coefficient P_0 model are listed in Table 2. For comparison the GC filter coefficients of the P_0 model reported by Patterson, Unoki, and Irino (2003) are also listed in the bottom row. The main difference is the numbers of the non-filter coefficients. The level-dependency coefficient of the cochlear noise floor n_{LD} in Eq. 9 was 0.20 dB/dB, which is approximately the same as the minimum slope of the compression function shown in Fig. 4b and described in, e.g., Moore (2012) and Pickles (2013). This may indicate that the distortion products associated with the NN masker produce a cochlear noise floor with an increasingly compressive growth rate. The $N_c^{(LD)}$ model includes this effect which may well account for the reduction in NN error to less than 2 dB. The corresponding value for the P_0 model with the current NN data is 2.40 dB and the value reported in Patterson, Unoki, and Irino (2003) is 3.71 dB for the previous NN data set.

4.2.3 Filter shape

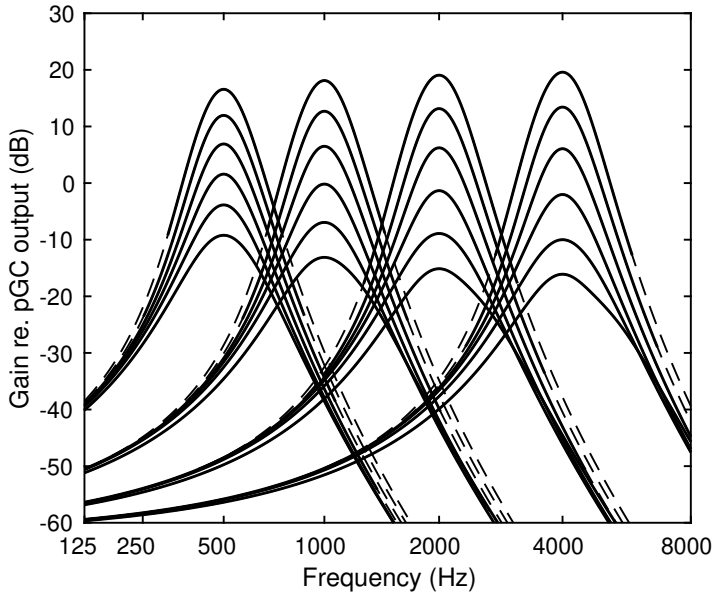
The filter shapes associated with the coefficients listed in Table 2 are shown in Fig. 3a for the nine-coefficient $N_c^{(LD)}$ model and Fig. 3b for the twelve-coefficient P_0 model. The filter shapes of the $N_c^{(LD)}$ model are sharper than those of the P_0 model. This means the $N_c^{(LD)}$ model can repress the effect of NN threshold convergence onto AT at low-levels as shown in Fig. 1.

4.2.4 Bandwidth and IO function

Figure 4a shows the bandwidth of the nine-coefficient $N_c^{(LD)}$ model shown in Fig. 3a. When the level of the cochlear input is between 30 dB and 50 dB, the bandwidth was effectively fixed at approximately 1.5 times of the standard ERB of NH listeners, ERB_N. Above 60 dB, the bandwidth increased rather rapidly as the level increased. The rate of increase was slightly slower at 500 Hz than at the higher signal frequencies, perhaps because the dynamic range of the 500 Hz filter is relatively smaller. This is slightly different from the result presented



(a) Nine-coefficient $N_c^{(LD)}$ model

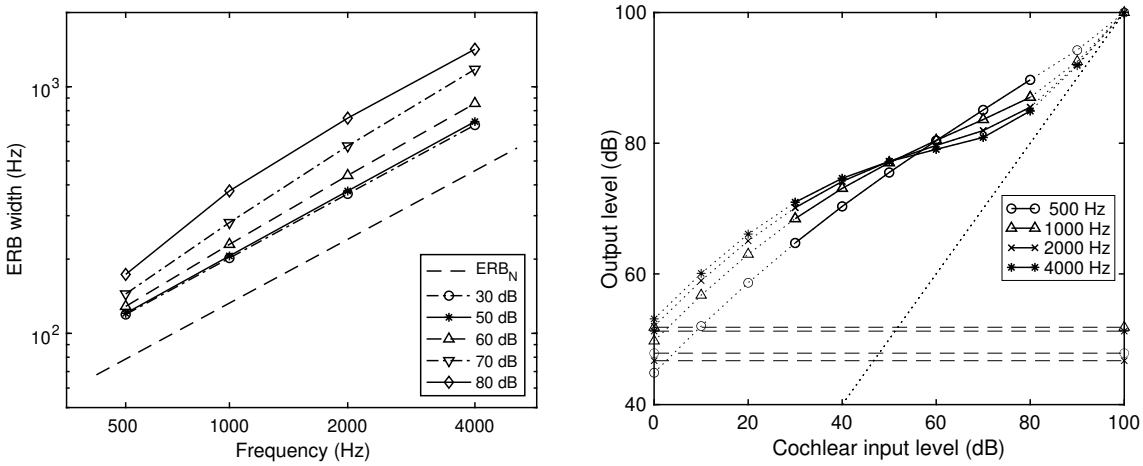


(b) Twelve-coefficient P_0 model

Figure 3 — Filter shapes using the coefficients listed in Table 2. The center frequencies are 500, 1000, 2000, and 4000 Hz and the five lines at each frequency correspond to cochlear input levels every 10 dB between 30 and 80 dB.

in Patterson, Unoki, and Irino (2003) where the bandwidth increased almost linearly between 30 dB and 70 dB.

Figure 4b shows the IO function for the nine-coefficient $N_c^{(LD)}$ model shown in Fig. 3a. The slope of the IO function decreases as the center frequency increases. The minimum slope was 0.46 dB/dB at 500 Hz, 0.32 dB/dB at 1000 Hz , 0.23 dB/dB at 2000 Hz, and 0.18 dB/dB at 4000 Hz. The IO slopes are roughly consistent with those in Patterson, Unoki, and Irino (2003).



(a) Bandwidth (b) IO function

Figure 4 — Bandwidth and IO function of the nine-coefficient $N_c^{(LD)}$ model shown in Fig. 3a. The lines represent the values when the center frequencies are 500, 1000, 2000, and 4000 Hz. The IO function is drawn so that the output level is 100 dB when the cochlear input level is 100 dB.

5 Estimation of the cochlear noise floor in quiet

In Section 4, it was demonstrated that AF estimation was successful when the cochlear noise floor in quiet was defined as $N_c^{(Q)}(f)$ in Eq. 7. It was assumed that $N_c^{(Q)}(f)$ could be directly from the HL-0dB function. The purpose of this section is to confirm this assumption. It is equivalent to assuming that the frequency dependence of AT is entirely determined by the internal cochlear noise floor.

AT can be estimated with the PSM as shown in Eqs. 6. In the PSM, the SNR detector, K, is after the cochlear filter, and so may be involved in AF estimation. It is possible to make K a frequency-dependent function as listed in Table 1, and there could be a trade-off in the frequency dependence of the cochlear noise floor and that of K. If so, $N_c^{(Q)}(f)$ could be different from Eq. 7, being at least partly determined by the frequency distribution of K.

In this section, the distributions of $N_c^{(Q)}(f)$ and K are rewritten to check for a the trade-off between them, and to find a plausible estimate of $N_c^{(Q)}(f)$.

5.1 Procedure

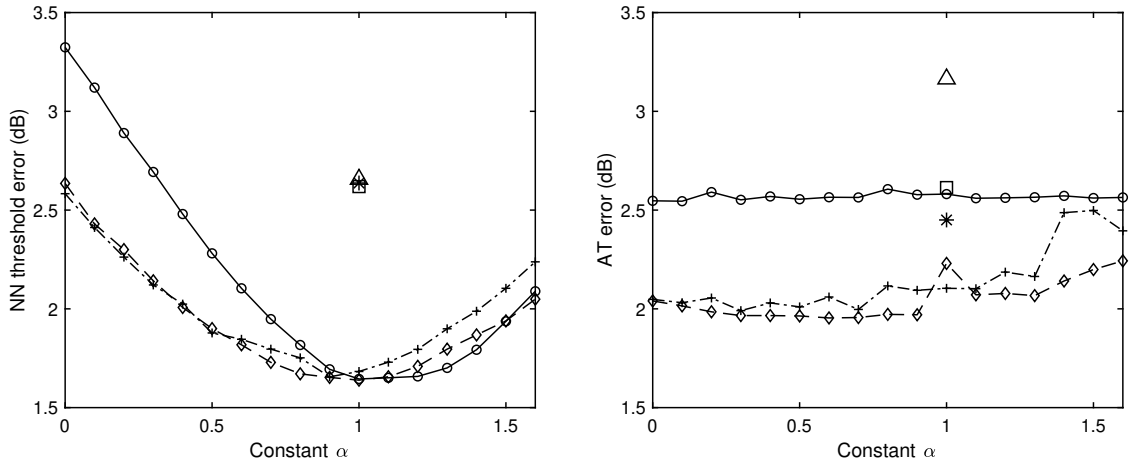
The distribution of the cochlear noise floor in Section 4 was defined as in Eq. 7 from the HL0-dB function in Fig. 2(a). Although there would be number of potential variants, we introduced a constant α into Eq. 7 to reduce or enhance the spectrum distribution of the HL-0dB function as

$$N_c^{(Q)}(f,\alpha) \;=\; \left[\, |T_{mid}(f)|^2 \cdot \frac{L_{HL0}(f)}{L_{HL0}(f_{ref})} \,\right]^{\alpha} \cdot N_c^{(Q)}(f_{ref}) \cdot \frac{ERB_N(f_{ref})}{ERB_N(f)}. \qquad (15)$$

It is equivalent to Eq. 7 when $\alpha = 1$ and is equivalent to a uniformly exciting noise (Glasberg & Moore, 2000) on the $\mathrm{ERB_{N\ number}}$ axis when $\alpha = 0$. The constant α determines the noise floor

function in the proportion from the HL-0dB function. The dynamic range of the distribution is reduced when $0 < \alpha < 1$ and is emphasized when $\alpha > 1$.

The fitting procedure was similar to that in Section 4. The auditory filter was estimated after the proportion constant α was set to every 0.1 step between 0 and 1.6. The fit was performed 10 times with different initial coefficients and the best of 10 filter was selected as the one that minimized the rms error of the NN threshold. We compared the $N_c^{(LD)}$ models when K was constant, linear, or quadratic functions of frequency.



(a) NN threshold error (b) AT error

Figure 5 — Estimation errors (dB) of the NN threshold (a) and the AT (b) as a function of the proportionality constant α in Eq. 15. Lines with circles (o), pluses (+), and diamonds show the results from the $N_c^{(LD)}$ model with constant, linear, and quadratic functions of K , respectively. Square, asterisk, and triangle show the errors listed in Table 1 when using the $N_c^{(Fx)}$ model with constant, linear, and quadratic functions of K and $\alpha = 1$.

5.2 Result

Figure 5a shows the estimation error of NN threshold as a function of the proportionality constant α of $N_c^{(Q)}(f, \alpha)$ in Eq. 15. The lines with circles (o), pluses (+), and diamonds show the results from the $N_c^{(LD)}$ model with constant, linear, and quadratic functions of K. These lines were approximately parabolic and had minimum values of 1.64 dB, 1.68 dB, and 1.64 dB when α was 1.0, 0.9, and 1.0, respectively. The estimation was best when $\alpha \approx 1$, independent of the function order of K. As definition of Eq. 15 with $\alpha = 1$ is equivalent to Eq. 7. This implies that AT can be properly formulated by the cochlear noise floor function shown in Eq. 7 without the need to make the SNR detector, K, frequency dependent.

Figure 5b shows AT error as a function of the proportionality constant α. The error of the $N_c^{(LD)}$ model when K is a constant (line with circle) is approximately 2.6 dB. This value is greater than that of the $N_c^{(LD)}$ model when K is either a linear or a quadratic function α. It appears that a frequency dependent K can reduce AT error but this is limited to 0.6 dB improvement at maximum. The effect of the AT error on AF shape was much smaller than that of NN threshold because the number of the NN thresholds was 144 while the number of the ATs was 4.

In summary, AT can be well modeled by the cochlear noise floor function shown in Eq.7,

independent of the detector SNR, K.

6 Conclusions

This paper provided a detailed set of the NN threshold values, including low-level noises at four center frequencies (500, 1000, 2000, and 4000 Hz), to show how threshold converges onto AT as notch width increases at low noise levels. We assumed that the cochlear noise floor limits threshold at wide notch widths. Then, we extended the power spectrum model of masking to explain both the AT and the NN thresholds simultaneously by introducing a level-dependent cochlear noise floor, $N_c^{(LD)}$. The distribution of the cochlear noise floor in quiet, $N_c^{(Q)}$, was assumed to be directly defined by the 0-dB HL function, which was defined by AT of NH listeners. The level dependence was set to be proportional to the external noise level. The GC auditory filter was estimated by minimizing the errors between the experimental data and the predicted data simultaneously for all the four center frequencies. The estimation error for the $N_c^{(LD)}$ model were much less than those for both the conventional P_0 model and the fixed noise floor, $N_c^{(Fx)}$, model. The $N_c^{(LD)}$ model with nine coefficients produces much smaller estimation errors than the P_0 model with twelve coefficients. The resultant filter shapes were sharper than those estimated by the P_0 model. This implies that the $N_c^{(LD)}$ model can successfully repress the effect of threshold convergence onto AT at low-levels as shown in Fig. 1, and thus, the $N_c^{(LD)}$ model is the better representation for auditory filter estimation.

Finally, we showed that the NN estimation error was minimum when the $N_c^{(Q)}$ function was directly set to the HL-0dB function, regardless of the frequency dependence of the detector SNR, K. This suggests that AT is solely determined by the cochlear noise floor in quiet, and is independent of the detector SNR following the cochlear filter.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP16H01734 and JP21H03468. The authors wish to thank Toshie Matsui, Hiroki Matsuura and Anzu Nakama for assisting in the data collection.

References

Aibara, R., Welsh, J. T., Puria, S., & Goode, R. L. (2001). Human middle-ear sound transfer function and cochlear input impedance. *Hearing research*, *152*(1-2), 100–109.

ANSI_S3.6-2010. (2010). Specification for audiometers [(American National Standards Institute, New York, USA, 2010)].

Baker, R. J., & Rosen, S. (2002). Auditory filter nonlinearity in mild/moderate hearing impairment. *The Journal of the Acoustical Society of America*, *111*(3), 1330–1339.

Buss, E., Porter, H. L., Leibold, L. J., Grose, J. H., & Hall III, J. W. (2016). Effects of self-generated noise on estimates of detection threshold in quiet for school-age children and adults. *Ear and hearing*, *37*(6), 650.

Fletcher, H. (1940). Auditory patterns. *Reviews of modern physics*, *12*(1), 47.

Gaskill, S. A., & Brown, A. M. (1990). The behavior of the acoustic distortion product, 2 f 1- f 2, from the human ear and its relation to auditory sensitivity. *The Journal of the Acoustical Society of America*, *88*(2), 821–839.

Glasberg, B. R., & Moore, B. C. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *The Journal of the Acoustical Society of America*, *79*(4), 1020–1033.

Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing research*, *47*(1-2), 103–138.

Glasberg, B. R., & Moore, B. C. (2000). Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. *The Journal of the Acoustical Society of America*, *108*(5), 2318–2328.

Glasberg, B. R., & Moore, B. C. (2006). Prediction of absolute thresholds and equal-loudness contours using a modified loudness model. *The Journal of the Acoustical Society of America*, *120*(2), 585–588.

Hall, J. (1972). Auditory distortion products f 2- f 1 and 2f 1- f 2. *The Journal of the Acoustical Society of America*, *51*(6B), 1863–1871.

Irino, T., & Patterson, R. D. (1997). A time-domain, level-dependent auditory filter: The gammachirp. *The Journal of the Acoustical Society of America*, *101*(1), 412–419.

Irino, T., Yokota, K., Matsui, T., & Patterson, R. D. (2018). Auditory filter derivation at low levels where masked threshold interacts with absolute threshold. *Acta Acustica united with Acustica*, *104*(5), 887–890.

Irino, T., & Patterson, R. D. (2001). A compressive gammachirp auditory filter for both physiological and psychophysical data. *The Journal of the Acoustical Society of America*, *109*(5), 2008–2022.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America*, *49*(2B), 467–477.

Moore, B. C. (2012). *An introduction to the psychology of hearing*. Brill.

Moré, J. J. (1978). The levenberg-marquardt algorithm: Implementation and theory. In *Numerical analysis* (pp. 105–116). Springer.

Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *The Journal of the Acoustical Society of America*, *59*(3), 640–654.

Patterson, R. D., Allerhand, M. H., & Giguere, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *The Journal of the Acoustical Society of America*, *98*(4), 1890–1894.

Patterson, R. D., & Nimmo-Smith, I. (1980). Off-frequency listening and auditory-filter asymmetry. *The Journal of the Acoustical Society of America*, *67*(1), 229–245.

Patterson, R. D., Unoki, M., & Irino, T. (2003). Extending the domain of center frequencies for the compressive gammachirp auditory filter. *The Journal of the Acoustical Society of America*, *114*(3), 1529–1542.

Pickles, J. (2013). *An introduction to the physiology of hearing*. Brill.

Puria, S., Peake, W. T., & Rosowski, J. J. (1997). Sound-pressure measurements in the cochlear vestibule of human-cadaver ears. *The Journal of the Acoustical Society of America*, *101*(5), 2754–2770.

Unoki, M., Irino, T., Glasberg, B., Moore, B. C., & Patterson, R. D. (2006). Comparison of the roex and gammachirp filters as representations of the auditory filter. *The Journal of the Acoustical Society of America*, *120*(3), 1474–1492.

von Békésy, G., & Peake, W. T. (1990). *Experiments in hearing*. Acoustical Society of America.

Appendix

A Compressive Gammachirp Filter

We describe the formula and parameters of the compressive gammachirp filter here. A brief summary of the development of the gammatone and gammachirp filterbanks over the past is provided in Appendix A of (Patterson, Unoki, & Irino, 2003).

The complex form of the gammachirp auditory filter (Irino & Patterson, 1997) is

$$\begin{aligned} g_c(t) &= at^{(n_1-1)}\exp(-2\pi b_1 ERB_N(f_{r1})t) \\ &\quad \exp(2j\pi f_{r1}t + jc_1 \ln t + j\phi_1) \end{aligned} \tag{16}$$

where time $t > 0$; a is amplitude; n_1 and b_1 are parameters defining the envelope of the gamma distribution; c_1 is the chirp factor; f_{r1} is the asymptotic frequency; $ERB_N(f_{r1})$ is the equivalent rectangular bandwidth (Glasberg & Moore, 1990); $j\phi_1$ is the initial phase; and is the natural logarithm of time. When $c_1 = 0$, Eq. 16 reduces to the complex impulse response of the gammatone filter (Patterson, Allerhand, & Giguere, 1995). The Fourier magnitude spectrum of the analytic gammachirp filter in Eq.16 is

$$|G_{CA}(f)| = a_\Gamma \cdot |G_T(f)| \cdot \exp(c_1\theta_1(f)), \tag{17}$$

$$\theta_1(f) = \arctan\left(\frac{f - f_{r1}}{b_1 ERB_N(f_{r1})}\right). \tag{18}$$

$|G_T(f)|$ is the Fourier magnitude spectrum of the gammatone filter, and $\exp(c_1\theta_1(f))$ is an asymmetric function since $\theta_1(f)$ is an anti-symmetric function centered at the asymptotic frequency, f_{r1} (Eq. 17). a_Γ is the relative amplitude of the magnitude spectrum of the gammatone filter.

Irino and Patterson (2001) decomposed the asymmetric function, $\exp(c_1\theta_1(f))$, into separate low-pass and high-pass asymmetric functions in order to represent the passive basilar membrane and the subsequent level-dependent component separately in the filter function. The resulting 'compressive' gammachirp filter, $|G_C(f)|$, is

$$\begin{aligned} |G_C(f)| &= \{a_\Gamma \cdot |G_T(f)| \cdot \exp(c_1\theta_1(f))\} \cdot \exp(c_2\theta_2(f)) \\ &= |G_{CP}(f)| \cdot \exp(c_2\theta_2(f)) \end{aligned} \tag{19}$$

$$\theta_1(f) = \arctan\left(\frac{f - f_{r1}}{b_1 ERB_N(f_{r1})}\right), \tag{20}$$

$$\theta_2(f) = \arctan\left(\frac{f - f_{r2}}{b_3 ERB_N(f_{r2})}\right). \tag{21}$$

Conceptually, this compressive gammachirp is composed of a level-independent, 'passive' gammachirp filter, $|G_{CP}(f)|$, that represents the passive basilar membrane, and a level-dependent, high-pass asymmetric function (HP-AF) , $\exp(c_2\theta_2(f))$, that represents the active mechanism in the cochlea. The peak frequency of the passive gammachirp, f_{p1}, is

$$f_{p1} = f_{r1} + c_1 b_1 ERB_N(f_{r1})/n_1 \tag{22}$$

The center frequency of the high-pass asymmetric function, f_{r2}, is determined by the following equation to introduce the level dependence.

$$f_{r2} = (f_{rat}^{(0)} + f_{rat}^{(1)} \cdot P'_{gcp}) \cdot f_{p1} \tag{23}$$

where P'_{gcp} is the output level of the passive gammachirp in a dB scale. When the slope factor, $f_{rat}^{(1)}$, is positive, f_{rat} and f_{r2} increase as the output level increases. This means that the HP-AF shifts upward relative to the passive gammachirp. As the result, the gain of the composite, compressive gammachirp reduces as observed physiologically and psychoacoustically. In summary, there are six parameters of the compressive gammachirp as $\{b_1, c_1, f_{rat}^{(0)}, f_{rat}^{(1)}, b_2, c_2\}$.