

Comparison of across-frequency integration strategies in a binaural detection model

Jeroen Breebaart

*Dolby Laboratories, Level 3, 35-51 Mitchell Street, McMahons Point, NSW 2060
Australia
jeroen.breebaart@dolby.com*

Abstract: Breebaart *et al.* [J. Acoust. Soc. Am. **110**, 1089–1104 (2001)] reported that the masker bandwidth dependence of detection thresholds for an out-of-phase signal and an in-phase noise masker (N_0S_π) can be explained by principles of integration of information across critical bands. In this paper, different methods for such across-frequency integration process are evaluated as a function of the bandwidth and notch width of the masker. The results indicate that an “optimal detector” model assuming independent internal noise in each critical band provides a better fit to experimental data than a best filter or a simple across-frequency integrator model. Furthermore, the exponent used to model peripheral compression influences the accuracy of predictions in notched conditions.

© 2013 Acoustical Society of America

PACS numbers: 43.66.Ba, 43.66.Pn, 43.66.Rq [QJF]

Date Received: July 3, 2013 Date Accepted: August 21, 2013

1. Introduction

Since the introduction of the concept of critical bands by Fletcher¹ in the 1940s, the frequency resolution of the human auditory system has been modeled by a set of nonlinearly spaced band-pass filters.² Subsequent to splitting signals into critical bands, most auditory models apply a peripheral nonlinear compressive function on the critical-band signal levels (cf. Refs. 3–7). In the case of binaural signal detection models, the peripheral compression is followed by a binaural process that uses inputs from the two ears to derive a representation of the input signals that accounts for the lower detection thresholds observed in a dichotic condition compared to a diotic presentation. Various mechanisms to account for this phenomenon have been proposed in the past, including (but not limited to) an equalization-cancellation⁸ or a cross-correlation⁹ stage, which in essence aim at providing an increased sensitivity in terms of the signal-to-masker ratio in each critical band. The next stage in most models constructs a decision statistic from the information available in one or more critical bands that can be used to predict a detection threshold or psychometric curve. We can identify three groups of models based on their way of processing information in critical bands.

- (i) Best filter model: In this approach, only the auditory filter that gives the best or most reliable result for the task at hand is considered;⁷ this model, thus, includes any off-frequency listening if that would be beneficial for the task at hand.
- (ii) Integrator model: This type of model is based on linear summation or integration of information across filters. The per-band information may be a specific loudness,⁶ a specific detectability based on energetic principles,¹⁰ or more advanced stimulus attributes.¹¹ To limit the accuracy or sensitivity of the model, the integrated decision variable is required to change by a minimum amount to ensure detectability.
- (iii) Optimal detector model: These models assume that information in each critical band is corrupted by internal noise that is independent across critical bands.

Integration of information across bands often involves weights to improve the detectability or accuracy of the task of the model by minimizing the effect of the noise on the integrated decision variable.^{4,5,11,12,18}

In this paper, the differences in model predictions for the three groups outlined above will be investigated for a binaural signal detection task. Analytical expressions for stationary signals will be used to predict the masker bandwidth and notch width dependence of signal detection thresholds in an N_0S_π detection task (e.g., the detection of an out-of-phase signal S_π in the presence of an in-phase noise masker N_0). Furthermore, the effect of the amount of peripheral compression is investigated by varying the peripheral compressive exponent. In particular, signal detection models based on peripheral adaptation loops^{4,5} have an effective compression exponent of $(0.5)^5 \approx 0.031$ and use an optimal detection strategy for combining information across critical bands, while loudness models⁶ are based on an exponent of 0.2 with a subsequent integrator model. Bernstein *et al.*³ and Goupell¹³ have shown benefits of peripheral compression to predict thresholds in low-noise noise conditions. Variation of the compressive exponent allows us to determine whether this parameter is critical in terms of model predictions and to determine whether potential interactions with across-frequency processing methods exist.

2. Method

2.1 Transformation of signal spectrum to internal representation

The power spectral density of a signal at frequency f is denoted by $X^2(f)$. The signal power, or excitation level, $\sigma_X^2(b)$, in auditory filter with index b is then given by

$$\sigma_X^2(b) = \int_0^\infty \Gamma^2(b,f)H^2(f)X^2(f) df, \tag{1}$$

with $H(f)$ a band-pass filter with skirts of 6 dB/Oct and -3 dB cut-off frequencies of 1000 and 4000 Hz to model the combined outer and middle ear transfer function, and $\Gamma(b,f)$ the magnitude transfer function of auditory filter b . In the simulations, a third-order^{4,14} gammatone filter bank was used with filter bandwidths equal to the equivalent rectangular bandwidth (ERB)^{6,15} and a spacing of ten filters per ERB. The transformation of excitation levels $\sigma_X^2(b)$ to an “internal representation,” $y(b)$, involves a compressive nonlinearity with exponent $\alpha < 1$ and an additive term A to account for a threshold in quiet

$$y_X(b) = (\sigma_X^2(b) + A)^\alpha. \tag{2}$$

Values of $\alpha = 0.031$, $\alpha = 0.1$, and $\alpha = 0.2$ will be used to investigate the effect of this parameter. In a signal detection task, the (expected) value for $y_{N+S}(b)$ as a result of a noise masker, N , and target signal, S , under the assumption of a stationary and independent signals is given by

$$y_{N+S}(b) = (\sigma_S^2(b) + \sigma_N^2(b) + A)^\alpha, \tag{3}$$

which for $\sigma_S^2 \ll \sigma_N^2$ leads to the following approximation:

$$y_{N+S}(b) - y_N(b) \approx \alpha(\sigma_N^2(b) + A)^\alpha \left(\frac{\sigma_S^2(b)}{\sigma_N^2(b) + A} \right). \tag{4}$$

2.2 Detectability index

For the best filter model, the detectability index, d_1 , in a detection task equals the maximum change in $y(b)$ across filters, b , due to the addition of the signal adjusted with an internal noise parameter β_1

$$d_1 = \frac{1}{\beta_1} \max_b (y_{N+S}(b) - y_N(b)). \tag{5}$$

The integrator model, on the other hand, has a detectability index, d_2 , that is obtained by integrating the changes in $y(b)$ across critical bands, b ,

$$d_2 = \frac{1}{\beta_2} \int_0^\infty (y_{N+S}(b) - y_N(b)) db. \tag{6}$$

The optimal detector model is based on the assumption that each observation of $y_{n+s}(b) - y_n(b)$ is subjected to an additive internal noise, and that detection is based on a weighted sum of these noisy observations. For a zero-mean, constant variance, independent internal noise, the detectability index, d_3 , for the optimal detector model is given by⁴

$$d_3^2 = \frac{1}{\beta_3^2} \int_0^\infty (y_{N+S}(b) - y_N(b))^2 db. \tag{7}$$

2.3 Calibration of β and A

The binaural advantage for detecting an out-of-phase signal compared to monaural conditions is not modeled explicitly. Instead, any binaural processor responsible for that phenomenon is modeled as a black box providing an increased ratio of σ_S^2 over σ_N^2 in each critical band, which, in essence, is the basis of the equalization-cancellation theory.⁸ Given the formulation of Eq. (4), this increased ratio can be incorporated by a reduction in the internal noise parameter, β . Specifically, the value of β was set to result in an N_0S_π threshold at 500 Hz equal to 41 dB sound pressure level (SPL) based on a wide-band noise masker with a power spectral density of 40 dB/Hz. Subsequently, the value of A was determined to result in a threshold-in-quiet of 4 dB SPL at 500 Hz. This calibration procedure was carried out independently for each of the three models given in Sec. 2.2.

3. Results

N_0S_π detection thresholds as a function of the masker bandwidth are shown in Fig. 1. The masker level was set to 70 dB SPL and arithmetically centered at 500 Hz irrespective of the bandwidth; the frequency of the signal amounted to 500 Hz. The different line styles represent the thresholds according to the best filter model, the integrator model, and the optimal detector model. The triangles are experimental data replotted from van de Par and Kohlrausch.¹⁶ The three panels from left to right correspond to $\alpha = 0.031$, $\alpha = 0.1$, and $\alpha = 0.2$, respectively. The number between brackets in the legends represents the root mean square error of the model predictions with respect to the experimental data. The three models demonstrate equal thresholds for a masker bandwidth of 1 kHz because all models were calibrated for such a wide-band noise condition. For smaller bandwidths, the best-filter model shows the highest thresholds, and the integrator model predicts the best performance. The optimal detector model generally shows the best match with the experimental data for all three values of α . When comparing model predictions across the three panels representing different values of α , it can be observed that an increase in α brings the three model approaches closer together, indicating that the efficiency or benefit from across-frequency integration is higher for smaller values of α . Furthermore, for $\alpha = 0.031$ (left panel), the fit between experimental data and optimal detector predictions is exceptionally good for

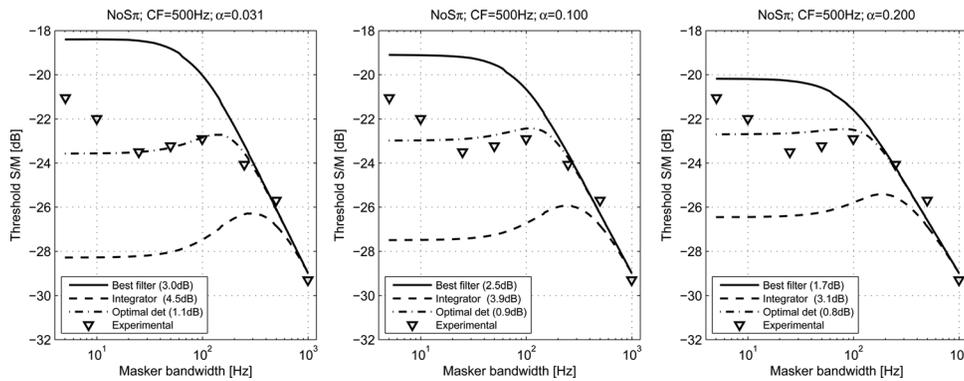


Fig. 1. N_0S_π detection thresholds expressed as signal-to-masker level in dB as a function of the masker bandwidth at 500 Hz center frequency. Different line styles indicate different models (see legend). Experimental data shown by the triangles are average thresholds across subjects replotted from Van de Par and Kohlrausch (Ref. 16). Left panel: $\alpha = 0.031$; middle panel: $\alpha = 0.1$; right panel: $\alpha = 0.2$.

bandwidths of 25 Hz and beyond, but errors amount to more than 1 dB for smaller bandwidths. When the root mean square error is considered, the optimal detector model with $\alpha = 0.2$ (dashed-dotted line in the right panel of Fig. 1) provides the best fit with the experimental data.

The detection thresholds as a function of the notch width of a noise masker are visualized in Fig. 2. The masker power spectral density was set to 30 dB/Hz, and the frequency of the signal was 500 Hz. Similar to the format of Fig. 1, different line styles represent the various modeling approaches and the triangles represent experimental data replotted from Nitschmann and Verhey.¹⁴ In contrast to the experiments with bandwidth as independent variable, varying the notch width does not reveal large differences between models. Thresholds decrease with increasing masker notch width with a slope that depends on the value of α . The best fit with experimental data is obtained for $\alpha = 0.2$ (right panel of Fig. 2), while smaller values of α result in a too steep decrease with masker notch width.

4. Discussion and conclusions

Despite the relatively simple binaural auditory model that was employed in this study using a linear filter bank and only two calibration parameters (β and A), the N_0S_π thresholds as a function of masker bandwidth could be predicted accurately using an

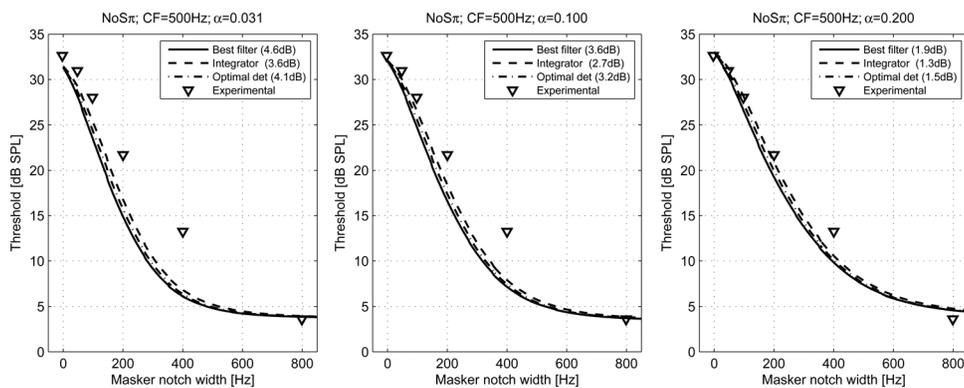


Fig. 2. N_0S_π detection thresholds expressed in dB SPL as a function of the masker notch width at 500 Hz center frequency. Different line styles indicate different models (see legend). Experimental data shown by the triangles are average thresholds replotted from Nitschmann and Verhey (Ref. 14). Left panel: $\alpha = 0.031$; middle panel: $\alpha = 0.1$; right panel: $\alpha = 0.2$.

optimal detector model. The other two modeling approaches (e.g., the best filter and integrator model) show substantially larger discrepancies between experimental data and model predictions. More specifically, the integrator model overestimates the benefit of across-frequency integration, resulting in too low thresholds for narrow-band maskers. The best-filter model, on the other hand, has no means to combine information across filters and, hence, the predicted thresholds for narrow-band maskers are too high. Modifying the compressive nonlinearity by means of changing the exponent, α , changes the predicted thresholds of the models, but does not influence the conclusion on what across-frequency integration process is most capable of predicting thresholds as a function of bandwidth. For the three values tested, the optimal detector model provides the best fit with experimental data, while none of the models is able to predict the increase in thresholds for masker bandwidths of 10 Hz and below.

In notch-widening experiments, the nature of the stimulus limits detection to a very limited number of critical band outputs and therefore across-frequency integration is not providing a significant detection advantage over a best-filter model. Interestingly, the larger value of $\alpha = 0.2$ gives a fairly good fit with experimental data without requiring any across-channel processes as proposed by Nitschmann and Verhey.¹⁴ Such less aggressive compression could therefore serve as alternative explanation for the relatively shallow decrease of thresholds with notch width.

The reasonably good prediction for both bandwidth and notch-width dependencies may serve as an argument to use $\alpha = 0.2$ in binaural signal detection models instead of the more common value of $\alpha = 0.031$, which effectively provides logarithmic compression.⁵ Inspection of Eq. (4), however, reveals that changing α has an influence on the masker-level dependency of thresholds. In particular, a larger value of α will result in a shallower slope of thresholds as a function of masker level. This effect is visualized in Fig. 3. Narrow-band (50-Hz wide) N_0S_π thresholds are shown as a

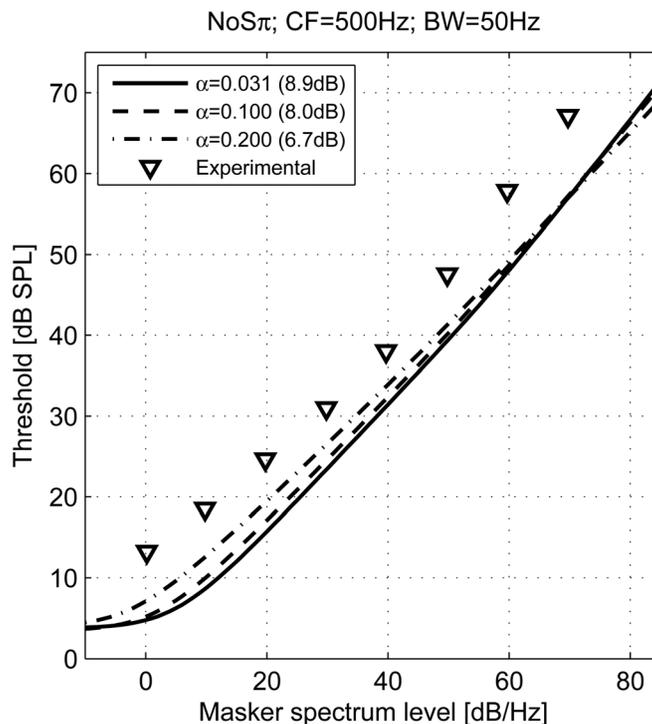


Fig. 3. N_0S_π detection thresholds expressed in dB SPL as a function of the masker level. Different line styles indicate different values for α ; experimental data shown by the triangles are replotted from Hall and Harvey (Ref. 17). The bandwidth of the masker amounted to 50 Hz.

function of the level of the masker. The center frequency of the stimulus was 500 Hz. The lines represent different values of α ; the triangles are experimental data replotted from Hall and Harvey.¹⁷ All model predictions were performed using an optimal detector strategy. As expected, the slope of the model predictions decreases with increasing α . Although the model predicts a higher overall sensitivity than the experimental data, one could argue that for masker levels up to 50 dB/Hz, $\alpha = 0.2$ results in a slope that is quite similar to the slope observed in experimental data while for higher masker levels, $\alpha = 0.031$ seems to provide the best predictions. Overall, $\alpha = 0.2$ results in the smallest root mean square error (see legend of Fig. 3). The data nevertheless suggest that a linear filter bank model followed by a single compressive exponent may be too simplistic to accurately cover such masker level dependencies.

References and links

- ¹H. Fletcher, "Auditory patterns," *Rev. Mod. Phys.* **12**, 47–65 (1940).
- ²B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138 (1990).
- ³L. R. Bernstein, S. van de Par, and C. Trahiotis, "The normalized interaural correlation: Accounting for $\text{NoS}\pi$ thresholds obtained with Gaussian and 'low-noise' masking noise," *J. Acoust. Soc. Am.* **106**(2), 870–876 (1999).
- ⁴J. Breebaart, S. van de Par, and A. Kohlrausch, "Binaural processing model based on contralateral inhibition. I. Model structure," *J. Acoust. Soc. Am.* **110**(2), 1074–1088 (2001).
- ⁵T. Dau, D. Püschel, and A. Kohlrausch, "A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure," *J. Acoust. Soc. Am.* **99**(6), 3615–3622 (1996).
- ⁶B. C. J. Moore, B. R. Glasberg, and T. Baer, "A model for the prediction of thresholds, loudness, and partial loudness," *J. Audio Eng. Soc.* **45**(4), 224–240 (1997).
- ⁷M. van der Heijden and A. Kohlrausch, "Using an excitation-pattern model to predict auditory masking," *Hear. Res.* **80**, 38–52 (1994).
- ⁸N. I. Durlach "Equalization and cancellation theory of binaural masking-level differences," *J. Acoust. Soc. Am.* **35**, 1206–1218 (1963).
- ⁹H. S. Colburn and N. I. Durlach, "Models of binaural interaction," in *Handbook of Perception*, edited by E. Varterette and M. Friedman (Academic, New York, 1978), Vol. IV, pp. 467–518.
- ¹⁰S. van de Par, A. Kohlrausch, G. Charestan, and R. Heusdens, "A new psychoacoustical masking model for audio coding applications," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (2002), Vol. 2, pp. II-1805–II-1808.
- ¹¹S. A. Davidson, R. H. Gilkey, H. S. Colburn, and L. H. Carney, "Binaural detection with narrowband and wideband reproducible noise maskers. III. Monaural and diotic detection and model results," *J. Acoust. Soc. Am.* **119**(4), 2258–2275 (2006).
- ¹²M. Florentine and S. Buus, "An excitation-pattern model for intensity discrimination," *J. Acoust. Soc. Am.* **70**(6), 1646–1654 (1981).
- ¹³M. J. Goupell, "The role of envelope statistics in detecting changes in interaural correlation," *J. Acoust. Soc. Am.* **132**(3), 1561–1572 (2012).
- ¹⁴M. Nitschmann and J. L. Verhey, "Binaural notched-noise masking and auditory filter shape," *J. Acoust. Soc. Am.* **133**(4), 2262–2271 (2013).
- ¹⁵B. C. J. Moore and B. R. Glasberg, "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**(3), 750–753 (1983).
- ¹⁶S. van de Par and A. Kohlrausch, "Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters," *J. Acoust. Soc. Am.* **106**(4), 1940–1947 (1999).
- ¹⁷J. W. Hall and A. D. G. Harvey, "Noso and nos π thresholds as a function of masker level for narrow-band and wideband masking noise," *J. Acoust. Soc. Am.* **76**, 1699 (1984).
- ¹⁸J. Breebaart, S. van de Par, and A. Kohlrausch, "Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters," *J. Acoust. Soc. Am.* **110**(2), 1089–1104 (2001).