Modeling off-frequency binaural masking for short- and long-duration signals

Marc Nitschmann, Ifat Yasin, G. Bruce Henning, and Jesko L. Verhey

Citation: The Journal of the Acoustical Society of America **142**, EL205 (2017);
View online: https://doi.org/10.1121/1.4996438
View Table of Contents: http://asa.scitation.org/toc/jas/142/2
Published by the Acoustical Society of America

Modeling off-frequency binaural masking for short- and long-duration signals

Marc Nitschmann,[1] Ifat Yasin,[2] G. Bruce Henning,[3]
and Jesko L. Verhey[1,a)]
[1]*Department of Experimental Audiology, Otto von Guericke University, Leipziger Str. 44,*
39120 Magdeburg, Germany
[2]*Department of Computer Science, University College London, Malet Place,*
London WC1E 6BT, United Kingdom
[3]*Institute of Ophthalmology, University College London, 11-43 Bath Street,*
London EC1V 9EL, United Kingdom
marc.nitschmann@uni-oldenburg.de, i.yasin@ucl.ac.uk, g.henning@ucl.ac.uk,
jesko.verhey@med.ovgu.de

Abstract: Experimental binaural masking-pattern data are presented together with model simulations for 12- and 600-ms signals. The masker was a diotic 11-Hz wide noise centered on 500 Hz. The tonal signal was presented either diotically or dichotically (180° interaural phase difference) with frequencies ranging from 400 to 600 Hz. The results and the modeling agree with previous data and hypotheses; simulations with a binaural model sensitive to monaural modulation cues show that the effect of duration on off-frequency binaural masking-level differences is mainly a result of modulation cues which are only available in the monaural detection of long signals.

[QJF]
Date Received: May 15, 2017 **Date Accepted:** July 14, 2017

1. Introduction

When the interaural characteristics of signal and masker differ (dichotic conditions) the masker is often much less effective in reducing the detectability of the signal than when the signal and the masker are identical in the ears (the diotic condition). The difference in detectability is referred to as the "binaural masking-level difference" (BMLD). BMLDs are frequently measured in the presence of a diotic noise, N_0, and defined as the difference in decibels (dB) between the threshold for detecting a signal that is in phase at the ears, S_0, and a signal that is 180° out of phase at the ears, S_π. On-frequency BMLDs are usually measured where the masker spectra contain the frequency of the signal (Jeffress, 1948; Durlach, 1963; Colburn, 1973). Very often in the real world, however, the spectra of masker and signal do not overlap completely, so studying the effect on BMLDs of increasing frequency separation between signal and masker (off-frequency BMLDs) may be as important as the traditional on-frequency BMLD. The present study focuses on binaural masking patterns as a typical paradigm to study off-frequency BMLDs. We shall refer to the difference in frequency between the center frequency of the noise and the frequency of the tonal signal as "Δf" (Hz). We investigate off-frequency BMLDs using different signal durations and compare our results with model predictions. Somewhat surprisingly, monaural cues seem to be crucial in understanding binaural masking patterns.

Binaural masking patterns describe BMLDs measured using a band of masking noise and a range of tonal signal frequencies both below and above the center frequency of the noise. Most measured binaural masking patterns have used signals (and maskers) of relatively long duration, such as the study by Zwicker and Henning (1984) who measured binaural masking patterns for a 600-ms signal, a masker (N_0) centered on 250 Hz, and a range of signal frequencies. For their narrowband (10-Hz wide) noise condition they reported a steep decrease in the BMLD with increasing frequency difference between the signal and masker (Δf, Hz). This steep decrease in the BMLD for noise bands centered on 250 Hz has been also found in later studies using signal durations from 300 to 600 ms (Henning *et al.*, 2007; Buss and Hall, 2010; Nitschmann and Verhey, 2012).

For similarly long-duration stimuli, a steep decrease in BMLD was also found when $|\Delta f|$ was increased for noise centered at frequencies greater than 250 Hz (Nitschmann and Verhey, 2012; Buss and Hall, 2010). Buss and Hall (2010) argued

---

[a)] Author to whom correspondence should be addressed.

that the steep decline in off-frequency BMLD with increasing $|\Delta f|$ is dominated by additional monaural cues that are available in off-frequency conditions. One such cue is the beating between signal and masker. Nitschmann and Verhey (2012) showed that a model sensitive to beating cues predicted their binaural masking patterns. Their model assumed an analysis of the envelope fluctuations of the stimulus with a modulation filterbank (Dau *et al.*, 1997) in the *monaural* pathways of the model whereas the *binaural* pathway of the model was not sensitive to beating cues.

Beating cues are prominent for long signals where the period of the beat frequency is considerably shorter than the signal duration. This implies that for very short signals, beating cues should be reduced or even absent. Thus, the decline in off-frequency BMLD with increasing $|\Delta f|$ should be reduced for the short signals when compared with long signals. So far, only Henning *et al.* (2007) have studied the effects of short stimulus durations on the binaural masking pattern, comparing off-frequency BMLDs with a noise centered at 250 Hz and with masker and signals gated together for total durations of either 12 or 600 ms. Results of the two subjects who participated in that experiment showed a steeper decline in BMLD for the longer-duration stimuli than for shorter signals, in agreement with the above hypothesis about the importance of beating cues.

The present study investigates to what extent the effect of duration on the off-frequency BMLD is predicted by the model of Nitschmann and Verhey (2012) originally developed for the prediction of binaural masking patterns with long signals. Model predictions are compared to binaural masking patterns with short- and long-duration stimuli presented in noise centered on 500 Hz. New threshold measurements were conducted (i) to investigate if the effect of duration is similar at 500 Hz (a frequency widely used in binaural studies) to that found in the literature for 250 Hz, (ii) to base the comparison between model predictions and behavioral measurements on a larger data set than that of Henning *et al.* (2007), and (iii) to ensure that the same stimuli and procedure were used for the behavioral measurements and for the threshold simulations.

2. Methods

2.1 Measurements

Masking patterns were obtained with 12- or 600-ms tone bursts in the presence of an 11-Hz wide Gaussian noise masker with a constant spectrum level within the band and an overall level of 67 dB sound pressure level (SPL). The noise band was arithmetically centered on 500 Hz and was generated in the frequency domain by transforming a Gaussian noise into the frequency domain via a fast Fourier transform and setting all Fourier components outside the desired passband to zero. A subsequent inverse Fourier transform on the complex buffer pair yielded the desired noise waveform. The signals and maskers were gated on and off together with the same $\cos^2$ ramps. The ramp duration was 6 ms for the 12-ms stimulus and 50 ms for the 600-ms stimulus. The masker was always presented in-phase at the ears (N_0), and the signal was either presented in-phase at the ears (S_0), or 180° out-of-phase at the ears (S_π); the two interaural phase conditions are conventionally specified as N_0S_0 and N_0S_π.

Frequencies of the tonal target spanned a range ±100 Hz around the center frequency of the noise in steps of 50 Hz. A schematic plot of a spectrum for a signal presented frequency below the masker is shown in the inset of the top right panel of Fig. 1. Thresholds were measured with a 3-interval 3-alternative forced-choice procedure. Each of the three intervals of a trial contained a different realization of the noise masker. One randomly chosen interval contained the signal. The task of the listener was to indicate the interval containing the signal by pressing the corresponding button on a keyboard. For a given frequency of the target tone, signal amplitude was adaptively varied using a two down, one up rule to estimate the 71% correct level for signal detectability (Levitt, 1971)—two correct responses produced a reduction in signal level, one incorrect response produced an increase in signal level. Each block of adaptive runs started with a clearly audible signal. The initial step size for level changes was 6 dB; the step size was reduced to 3 dB after the first change from a sequence of increasing signal levels to one with decreasing levels (called an upper reversal) and to 1 dB at the second upper reversal. The adaptive run continued for another six reversals with the 1 dB step size. Threshold was estimated as the mean of the levels obtained at the six final reversals. The block of runs was repeated 3 times. The mean of the thresholds across three blocks was taken as the threshold for the listener.

A total of eight normal-hearing listeners participated and was tested individually in sound-attenuating booths. Signals were presented via HD650 headphones

(Sennheiser HD650, Wedemark, Germany) that were calibrated using artificial ear (Brüel & Kjaer type 4153, Nærum, Denmark) and driven in phase.

2.2 Simulations

The model used to predict the data is essentially the same as the one used in Nitschmann and Verhey (2012). It is based on the effective binaural processing model by Zerbs (2000) extended by a modulation filterbank (Dau *et al.*, 1997) in the monaural processing stage. The present study used modulation filters with center frequencies equal to, or less than, half the center frequency of the auditory filter. To simulate the frequency decomposition at the level of the cochlea, a filterbank of fourth-order gammatone bandpass filters that are one equivalent rectangular bandwidth ($\mathrm{ERB_N}$) wide was used. One of these filters was centered on the center frequency of the masker and the other filters were positioned in 1-$\mathrm{ERB_N}$ steps below and above this filter. Only filters within a two-octave range around the center frequency of the masker were used for the simulations. The binaural processing part is an equalization–cancellation model (Durlach, 1963). Thresholds were estimated with the same procedure as in the experiments using the model as an artificial observer where the model does a pattern matching of the internal representation of the current stimulus minus the average internal representation of the masker alone with an internal representation of the target signal. This latter representation is calculated prior to the experiment as the difference of an average representation of the masker plus suprathreshold signal and the average internal representation of the masker alone. The model chooses that interval of a trial that has the highest correlation with the internal representation of the target signal (for details, see Dau *et al.*, 1996). The final threshold estimate was calculated as the mean across at least 36 threshold estimates from simulated experimental tracks.

3. Results and discussion

Psychometric functions for detecting tones in noise are parallel when plotted as percentage correct vs signal level (dB) for similar N_0S_0 and N_0S_π conditions at different frequencies and at different durations (Yasin and Henning, 2012). Consequently our data can be summarized by a "threshold," i.e., a cut at a single performance level—71% correct in our case. Figure 1 shows the mean binaural masking patterns (upper two panels) for long (left panel) and short (right panel) signals. Signal levels are shown in dB SPL at threshold (where the signal level has been measured before applying the gating window) as a function of the difference between the frequency of the target tone and the center frequency of the noise (500 Hz), Δf. Circles and inverted triangles

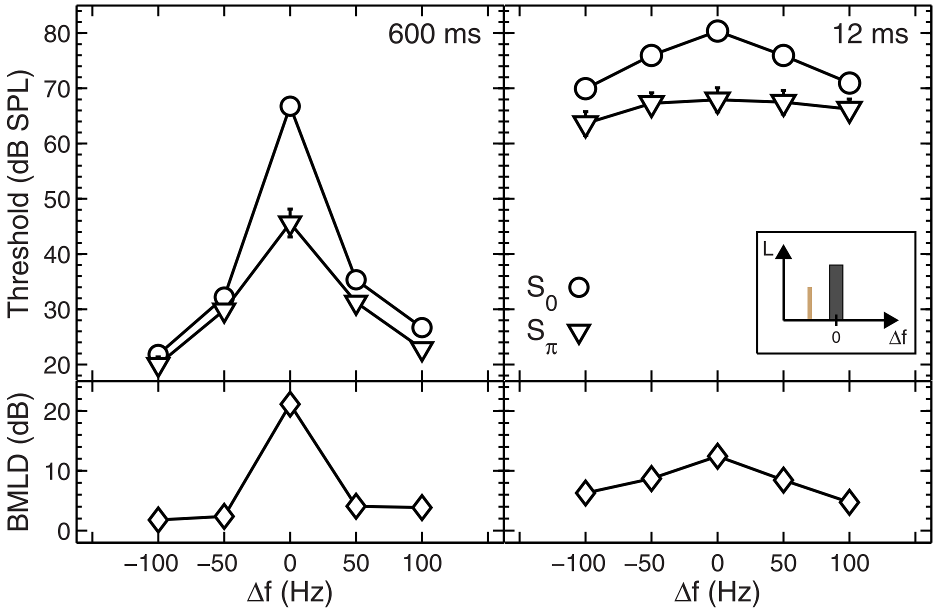

Fig. 1. (Color online) Mean binaural masking patterns and corresponding BMLDs obtained in the presence of in-phase noise (N_0) arithmetically centered on 500 Hz. Circle and inverted triangle symbols in the upper panels represent mean data for N_0S_0 and N_0S_π, respectively. Left and right columns present binaural masking patterns (upper panels) and the BMLD (bottom panels) for 600- and 12-ms duration stimuli, respectively. Standard error bars are also shown (when they were larger than the symbol size). Signal frequency is represented as Δf—the difference in hertz from the center frequency (500 Hz) of the noise. The inset in the upper right panel shows a schematic plot of a spectrum for a signal presented frequency below the masker.

represent the averages across listeners for the diotic (N_0S_0) and dichotic (N_0S_π) conditions, respectively. Error bars indicate plus and minus one standard error of the mean calculated from the individual thresholds.

The corresponding BMLDs are shown in the lower two panels. Error bars are not shown since they are smaller than the symbol size. The masking patterns for the long signals are slightly steeper than the corresponding masking patterns for a different set of listeners and a slightly lower masker level (60 dB SPL) shown in Nitschmann and Verhey (2012) but the effect of Δf on the BMLD is essentially the same. An on-frequency BMLD of slightly more than 20 dB is measured in the present study and in Nitschmann and Verhey (2012). For a target-tone frequency of 100 Hz below the center frequency of the masker, the BMLD is small (approximately 2 dB) in both studies. For a signal frequency 100 Hz above center frequency of the masker, the BMLD is slightly larger than for the same spectral distance below the signal frequency [4 dB in the present study and 3 dB in Nitschmann and Verhey (2012)].

Masking patterns for the short-duration signal are shallower than those for the long-duration signal and the difference in shape between the diotic and dichotic signals is less marked for short-duration signals than for the long-duration signals. This is consistent with the 250-Hz data shown in Henning *et al.* (2007). In both studies, the on-frequency BMLD for the short signal is smaller than that for the long signal and the decrease in BMLD as the frequency difference between masker and signal increases is smaller for the short than for the long signals. In contrast to the present data, thresholds in Henning *et al.* (2007) tend to increase toward lower signal frequencies with the longer duration stimuli, presumably as a consequence of the low signal levels which are already close to the threshold in quiet which increases with decreasing frequency in the frequency region around 250 Hz. For the frequency region of the present study around 500 Hz, thresholds in quiet are lower and thus there is hardly any influence of the threshold in quiet on the masking patterns.

In the following data analyses, Mauchly's Test of Sphericity was initially conducted and shown not to be significant, so sphericity could be assumed for the subsequent analysis of variance (ANOVA) analyses. *Post hoc* pairwise comparisons were conducted with a Bonferroni correction to keep type I error at 5%.

A within-subject ANOVA was conducted on the data for the S_0 signal with factors signal duration (12 or 600 ms) and signal frequency (400, 450, 500, 550, or 600 Hz). There was a significant effect of signal duration [$F(1,7) = 1918.7$, $p < 0.001$ (two-tailed) with effect size, $\eta^2, = 0.99$)], signal frequency [$F(4,28) = 419.9$, $p < 0.001$ (two-tailed) with effect size, $\eta^2 = 0.98$] and a significant interaction between signal duration and frequency [$F(4,28) = 231.89$, $p < 0.001$ (two-tailed) with effect size, $\eta^2 = 0.97$]. *Post hoc* pairwise comparisons showed a significant ($p < 0.001$) difference between thresholds for the 12 and 600 ms S_0 signal for all signal frequencies.

A within-subject ANOVA was conducted on the data for the S_π signal with factors signal duration (12 or 600 ms) and signal frequency (400, 450, 500, 550, or 600 Hz). There was a significant effect of signal duration [$F(1,7) = 422.3$, $p < 0.001$ (two-tailed) with effect size, $\eta^2 = 0.99$], signal frequency [$F(4,28) = 43.29$, $p < 0.001$ (two-tailed) with effect size, $\eta^2 = 0.98$], and a significant interaction between signal duration and frequency [$F(4,28) = 38.253$, $p < 0.001$ (two-tailed), with effect size, $\eta^2 = 0.85$]. *Post hoc* pairwise comparisons showed a significant ($p < 0.001$) difference between thresholds for the 12 and 600 ms S_π signal for all signal frequencies.

A repeated 2-way ANOVA was conducted on the values of BMLD with factors signal duration (12 or 600 ms) and signal frequency (400, 450, 500, 550, or 600 Hz). There was a significant effect of signal frequency [$F(4,28) = 42.03$, $p < 0.001$ (two-tailed) with effect size, $\eta^2 = 0.86$] and significant interaction between signal duration and signal frequency [$F(4,28) = 20.54$, $p < 0.001$ (two-tailed), with effect size, $\eta^2 = 0.75$].

Post hoc pairwise comparisons showed that the largest BMLD occurred for a 500-Hz signal at the longer duration of 600-ms; this BMLD was significantly greater than the BMLD generated by a 12-ms signal at any frequency ($p < 0.01$). The BMLD generated by the 600-ms signal reduces steeply for signal frequencies below or above 500 Hz, such that the BMLD for the 600-ms duration signal was significantly smaller for signal frequencies of 400, 450, 550, and 600 Hz than the "peak" BMLD for a 500-Hz signal at the shorter duration of 12 ms ($p < 0.01$).

Figure 2 presents the BMLDs that were predicted by the model of Nitschmann and Verhey (2012) shown by gray filled symbols connected with a dashed line. In addition, the measured BMLDs are redrawn from the lower panels of Fig. 1 (open symbols). As expected, and as shown previously (Nitschmann and Verhey, 2012), the model predicts the BMLDs for the long signals. The large decrease in

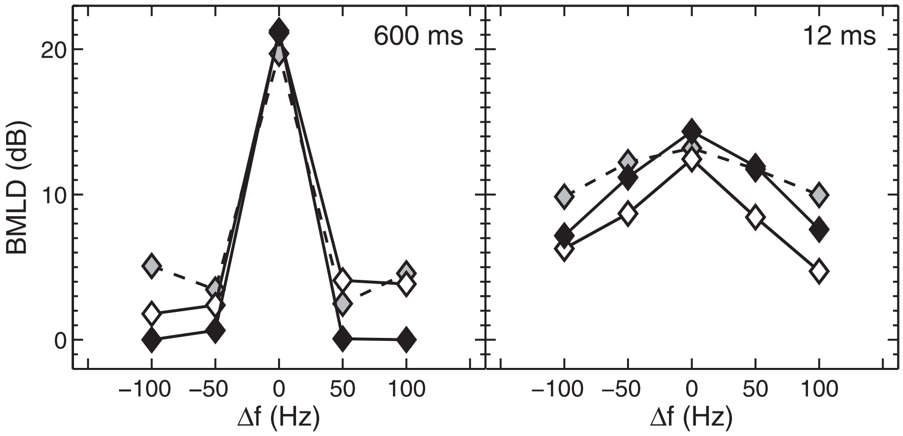

Fig. 2. In the left panel, the BMLD is shown as a function of the signal frequency for the 600-ms signal; the right panel shows the BMLD for the 12-ms signal. In each panel, mean measured BMLDs are shown with open symbols connected by a black solid line; predictions from the model of Nitschmann and Verhey (2012) are shown with gray filled symbols connected by a dashed line; and predictions from a modified model are shown with filled black symbols connected by a solid black line. In this modified model, the binaural filter bandwidth was 1.9 times wider than the monaural filter bandwidth.

thresholds with increasing frequency difference between signal and masker predicted by the model is due to modulation cues that, with these long signal and masker durations, contribute to signal detection in the diotic case when the signal is not centered in the masking noise band (but not in the dichotic condition). Note that the original model predicts, in contrast to the data, an increase in BMLD when $|\Delta f|$ is increased from 50 to 100 Hz. This was already observed in Nitschmann and Verhey (2012) and is presumably a consequence of the center frequency of the highest modulation filter that was used for the simulations.

The model predicts a smaller on-frequency BMLD for short than for long signals. The model also predicts that the decrease in the BMLD as the absolute frequency difference between masker and signal increases should be smaller for the short than for the long signals. Both predictions are consistent with the experimental data. The predicted reduction in the BMLD with increasing absolute frequency difference between masker and signal is due to residual beating cues: the beat frequency for a spectral distance of 100 Hz below or above the masker center frequency is so rapid that even for 12-ms signals, there is already more than one beat. Thus, beating cues can still play a role in diotic signal detection even with 12-ms signals. Note that the predicted decrease is slightly smaller than observed in the data. This may point toward effectively wider binaural filters than monaural filters, as suggested by notched-noise data (e.g., Nitschmann and Verhey, 2013). In order to test this hypothesis, the data were also predicted using a modified model where the bandpass filter for binaural processing was 1.9 times larger than that for monaural processing. [A factor of 1.9 was derived for 500 Hz from notched-noise data (Nitschmann and Verhey, 2013).] Predictions with such a modified model are also shown in Fig. 2 with black filled symbols connected with a solid line. As expected, the modified model predicts a steeper decrease than the original version. The predicted change of the BMLD with signal frequency for the short tonal signal is now close to the measured data. For long signals, the modified model predicts, in contrast to the original model but in agreement with the data, a decrease in BMLD as $|\Delta f|$ increases from 50 to 100 Hz. The model prediction of the decrease in BMLD with spectral separation between signal and masker is slightly larger than in the measured results. This may indicate that the assumed binaural filter widths of 1.9 $\mathrm{ERB_N}$ derived from the previous notched-noise study using a different set of listeners (Nitschmann and Verhey, 2013) may produce a fit that is slightly too broad for the current data set.

In summary, the BMLD data with a short duration signal may reflect two mechanisms: A residual modulation cue and an effectively wider binaural filter. Further experiments are required to test this hypothesis.

Acknowledgments

We thank the Royal National Institute for the Deaf (now Action on Hearing Loss) for a Knowledge Transfer Grant and the Deutsche Forschungsgemeinschaft (DFG, SFB trr 31) for supporting this project.

References and links

Buss, E., and Hall, J. W. III (**2010**). "The role of off-frequency masking in binaural hearing," J. Acoust. Soc. Am. **127**, 3666–3677.

Colburn, H. S. (**1973**). "Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination," J. Acoust. Soc. Am. **54**, 1458–1470.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). "A quantitative model of the 'effective' signal processing in the auditory system: I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Henning, G. B., Yasin, I., and Witton, C. (**2007**). "Remote masking and binaural masking level differences," in *Hearing—From Basic Research to Applications*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer Verlag, Heidelberg), pp. 457–466.

Jeffress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–39.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Nitschmann, M., and Verhey, J. L. (**2012**). "Modulation cues influence binaural masking-level difference in masking-pattern experiments," J. Acoust. Soc. Am. **131**, EL223–EL228.

Nitschmann, M., and Verhey, J. L. (**2013**). "Binaural notched-noise masking and auditory-filter shape," J. Acoust, Soc. Am. **133**, 2262–2271.

Yasin, I., and Henning, G. B. (**2012**). "The effects of noise-bandwidth, noise-fringe duration, and temporal signal location on the binaural masking-level difference," J. Acoust. Soc. Am. **132**, 327–338.

Zerbs, C. (**2000**). "Modelling the effective binaural signal processing in the auditory system," Ph.D. thesis, Carl-von-Ossietzky-Universität, Oldenburg, Germany.

Zwicker, E., and Henning, G. B. (**1984**). "Binaural masking-level differences with tones masked by noises of various bandwidths and levels," Hear. Res. **14**, 179–183.