Across-frequency integration in spatial release from masking

Norbert Kopčo
Department of Cybernetics and AI, Technical University, Košice, Slovakia and
Hearing Research Center, Boston University, **kopco@bu.edu**

Spatial separation of a target (T) stimulus from a masker (M) often improves detectability of the target, a phenomenon known as the spatial release from masking (SRM). When the masker is a noise, two main factors contribute to SRM: changes in the target-to-masker ratio dominate the performance at high frequencies, while binaural processing dominates at low frequencies. Previous neurophysiological studies (e.g., Lane et al., ISH 2003) suggested that, at the level of inferior colliculus, the SRM of broadband stimuli is determined by a single unit – the one that is most sensitive in the given T/M spatial configuration. Based on this observation, Lane et al. proposed a simple model that used the assumption that the channel with the most favorable signal-to-noise ratio also determines behavioral performance. The current study evaluated this model psychophysically. First, several T/M spatial configurations were selected based on the criterion that they must have a narrowband spectral region with very favorable SNR (re. other spectral regions). The stimuli were then filtered so that they would activate mainly the peripheral channel with the most favorable SNR. Detection thresholds were then measured for the filtered and the unfiltered stimuli, both binaurally and monaurally. Large differences (up to 10 dB) in performance were observed, with binaural thresholds generally better than the corresponding monaural thresholds, which, in turn, were better than the single-channel thresholds. These results support the single-channel model only partially. However, they do not prove that across-channel integration plays a role in spatial release from masking.

1 Introduction

Detectability of a target sound (T) presented concurrently with another sound, a masker (M), is influenced, among other things, by the relative spatial position of the target and the masker. In most cases, spatial separation of T from M improves the target detectability, i.e., it leads to a spatial release from masking (SRM). Previous studies of SRM [1,2,3] suggested that, for non-speech targets masked by noise, two factors influence SRM: 1) the relative ratio of the target and the masker energy (TMR) in the peripheral filters of both ears, and 2) binaural processing (mostly for T stimuli with low-frequency content.

Lane et al. [1] performed a study of SRM in which they measured human performance when detecting a chirp-train stimulus masked by noise. They measured performance with broadband and lowpass-filtered stimuli, and tried to predict the data using a simple model (called the single-best-filter model, SBF) that only considered processing in a single peripheral channel - the one with the most favorable TMR, and ignored binaural, across-frequency or amplitude modulation processing. The SBF model accurately predicted broadband performance. However, the model was unable to predict the lowpass and the broadband data at the same time because the lowpass thresholds were worse than the broadband thresholds, while the model predicted identical performance. Lane et al. proposed that this discrepancy was due to across-frequency integration of the peripheral auditory information, which the model did not consider.

The goal of the present study was to more directly evaluate the hypothesis that across-channel integration is important in SRM, and that it was the missing integration part of the model that lead to the failure in the predictions of the Lane et al. data. We first replicated the results of the previous study, to make sure that possible differences in the results do not come from different experimental procedures. Then, we analyzed the outputs of the model peripheral filters for various spatial configurations of the T and M, and chose several prototypical spatial configurations. The selected configurations ranged from a configuration where one peripheral filter had clearly the most favorable TMR (and thus small effect of integration would be expected even if the integration was important) to a configuration where multiple channels had approximately equally favorable TMR (and thus there was plenty of opportunity for the across-channel integration to influence results). For the chosen spatial configurations the threshold TMRs were measured binaurally, monaurally, and with the target stimulus pre-filtered by the most-favorable model peripheral filter so that the across-channel integration is minimized. If across-channel integration improves performance then the pre-filtered thresholds were expected to be worse than the broadband thresholds. If not, then the thresholds were expected to be similar.

The present study was performed mostly with broadband stimuli for which the binaural contribution to SRM was expected to be small. This was important because otherwise, it might have been hard to distinguish the contribution of binaural processing from the contribution of across-frequency integration, since both these factors were expected to improve performance.

2 Methods

2.1 Experimental procedure

The study consisted of two experiments. Five subjects with normal hearing participated in each experiment. Both experiments were performed in a virtual auditory environment, generated using non-individualized human head-related transfer functions (HRTFs). The target stimulus was a 200-ms long 40-Hz train of exponentially growing chirps with white spectrum in the range of 300-12,000 Hz (for lowpass conditions, the target was lowpass-filtered at 1,500 Hz). The masker was a white noise with frequency range of 200-14,000 Hz (for lowpass conditions, lowpass-filtered at 2,000 Hz). To determine the 79.8%-correct threshold TMR, 3-down-1-up 3-interval adaptive procedure was used, varying the T level. 3-interval, 2-alternative forced-choice procedure was used to collect responses. Stimuli were delivered via insert headphones in a quiet room.

In experiment 1 (described also in [5]), various azimuthal configurations of T and M were tested as indicated in Figure 1a. Most stimuli were broadband (except for three lowpass stimuli) and all were presented binaurally.

In experiment 2 (described also in [4]), only five azimuthal configurations were used (see panels “a” of Figures 2 to 6), chosen to examine the character of the across-frequency integration. For each spatial configuration, broadband binaural, broadband monaural, and several pre-filtered thresholds were measured (listed in panels “b” of Figs. 2 to 6). Gammatone filter [6] was used to pre-filter the signal so that the best-TMR peripheral auditory filter is activated most by the pre-filtered target.

2.2 Model

The “single-best-filter” model implemented to predict the data was identical to that used in the Lane et al. study [1]. The model computes the TMR in peripheral auditory channels a function of frequency, but does not allow for any across-frequency integration of information or any binaural processing. The model consists of a bank of 60 log-spaced gammatone filters [6] for each ear. For each spatial configuration, the root-mean-squared energy at the output of every filter is separately computed for the target and the masker. The model assumes that the filter with the largest TMR (over the set of 120) determines threshold. The only free parameter in the model, the TMR yielding 79.4% correct performance, was fit to match the measured threshold when broadband target and masker were at the same location.

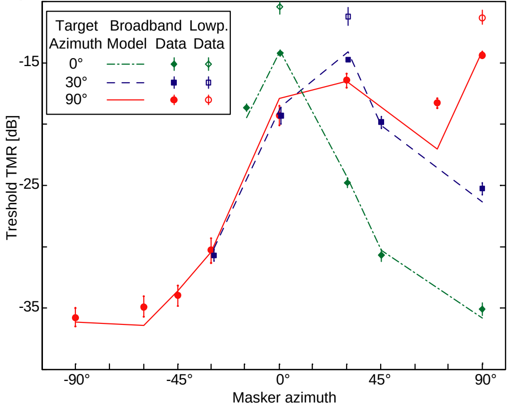
3 Results

3.1 Experiment 1

The results of Experiment 1 are summarized in Figure 1. The main goal of this experiment was to compare the results obtained with the current experimental procedures to those of Lane et al.

Figure 1a shows the measured (symbols) and predicted (lines) thresholds as a function of the masker azimuth, for T azimuth fixed at 0°, 30°, or 90°. There is a very

a) Mesured and predicted threshold TMR



b) CF and ear of filter used for prediction

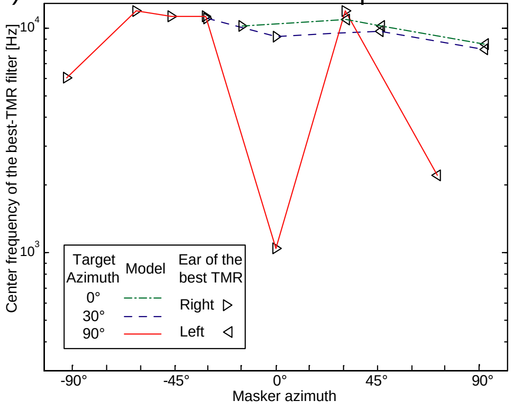
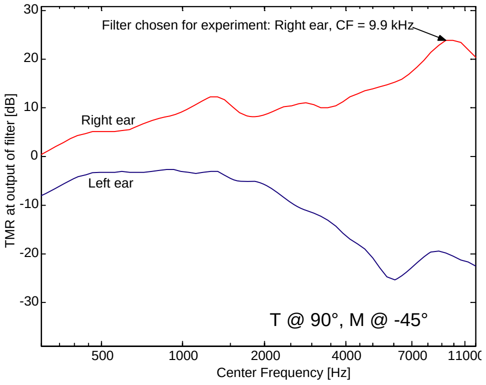


Figure 1: a) Measured and predicted threshold TMR for spatial configurations tested in Exp 1, plotted as a function of the Masker azimuth for a fixed target azimuth. b) Center frequency and ear (left vs. right) of the best-TMR filter based on which the corresponding prediction in panel a) was generated.

good match between the predicted and measured broadband thresholds. However, as in the previous study, the lowpass thresholds (open symbols) are consistently worse than the corresponding broadband thresholds. In addition, there is one broadband threshold (T @ 90°, M @ 70°, diamond symbol vs. the full line) mispredicted by the model. In this case, as well as in the lowpass cases, the model predicts that performance should be better than actually observed.

Figure 1b shows, for each broadband prediction from panel 1a, the ear (left or right) and the center frequency of the best-TMR peripheral filter on which the prediction was based. Most predictions were based on filters with high CF. However, the incorrect broadband prediction, as well as all the (incorrect) lowpass predictions, were based on filters with low CF. These results are very similar to results of Lane et al., suggesting that there might be a difference in the accuracy of predictions of high-CF vs. low-CF filters.


a) TMR at output of model peripheral filters

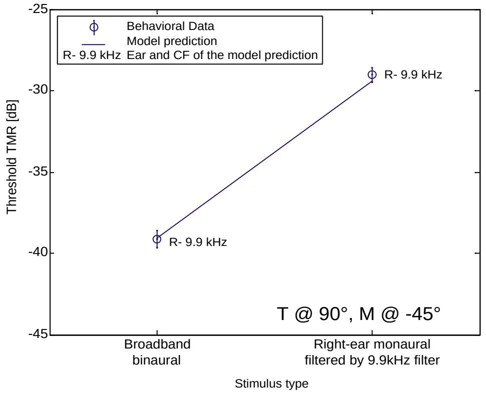
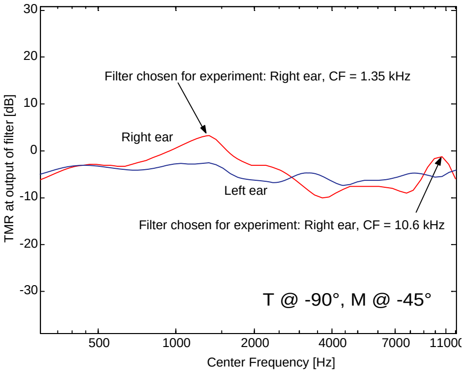

b) Mesured and predicted threshold TMR

Figure 2: a) TMR in each model peripheral filter in both ears when T is at 90° and M at -45°. Arrow points to the filter that was chosen to for pre-filtering.
b) Measured (x-subject mean and SE) and predicted threshold TMRs for the stimulus conditions for which threshold was measured with T at 90° and M at -45°

3.2 Experiment 2

In experiment 2, thresholds were measured in 5 spatial configurations. Figure 2 describes the results obtained with T at 90° and M at -45°. This spatial configuration was chosen because there is a single high-CF peripheral filter for which the TMR is much better than for the other filters (see Figure 2a). To test whether this filter in deed determines performance, threshold was measured with broadband binaural target and with target and masker presented monaurally to the right ear, with the target pre-filtered by the chosen model filter.

Symbols in Figure 2b show the two measured thresholds. The broadband threshold is approximately 10 dB better than the threshold obtained with the monaural pre-filtered target. However, this difference is well described by the model (line), suggesting that the difference is not due to across-channel integration, but simply due to double filtering of the stimulus (first, the pre-filtering to generate the narrowband stimulus, and second, the actual peripheral auditory filtering).


a) TMR at output of model peripheral filters

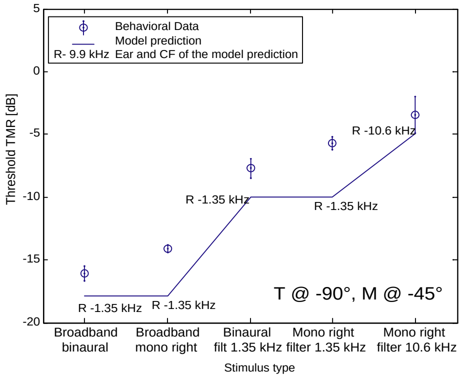

b) Mesured and predicted threshold TMR

Figure 3: Description of figure identical to Figure 2. Spatial configuration with T at -90° and M at -45°.

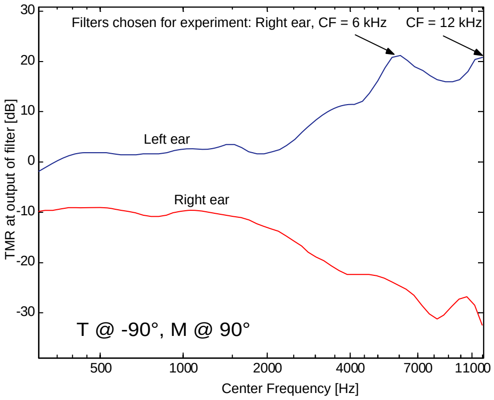
The second chosen spatial configuration was T at -90° and M at -45°. As shown in Figure 3a, this configuration is interesting because there is a single low-CF dominant channel (there is also a relatively good high-CF channel that was included in the measurement to distinguish its potential contribution). Thus, across-channel integration, as well as binaural processing, might contribute to the broadband threshold.

Five different thresholds were measured in this spatial configuration (see Fig. 3b). The best threshold was obtained with broadband binaural presentation, followed by broadband monaural presentation. The narrowband pre-filtered thresholds were several dB worse than the broadband ones. The binaural thresholds are always a little bit better than the corresponding binaural thresholds, suggesting that there is some binaural contribution.

The model always predicted better performance than observed (lines vs. symbols in Fig. 3b). Since the broadband prediction is based on a filter with low CF, this model error is consistent with the errors observed in experiment 1. Moreover, if the model was fitted to the broadband thresholds (i.e., to the leftmost two data points), the narrowband thresholds would be predicted accurately (imagine shifting the whole line up by 3 dB). Thus, no across-channel integration is necessary to explain these data.

The third spatial configuration in Experiment 2 was T at -90° and M at 90°. As shown in Fig. 4a, in the left ear there are two high-CF channels with very favorable TMR. Three thresholds were measured (Fig. 4b), one broadband binaural, and one narrowband monaural for each of the two candidate channels. The results show that both the binaural and the 6-kHz monaural threshold are well predicted by the model, which means that considering the 6-kHz channel alone is sufficient to predict broadband performance. There is a large difference between the model prediction and the data for the 12-kHz threshold. This difference is probably a combination of inconsistent listener performance (note the large standard error bar) and some border effect, since the best filter is the filter with the highest CF considered.

a) TMR at output of model peripheral filters



b) Mesured and predicted threshold TMR

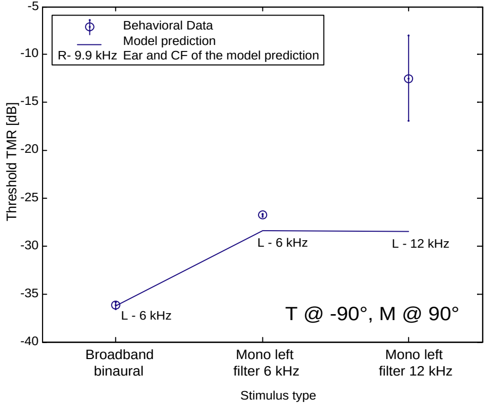
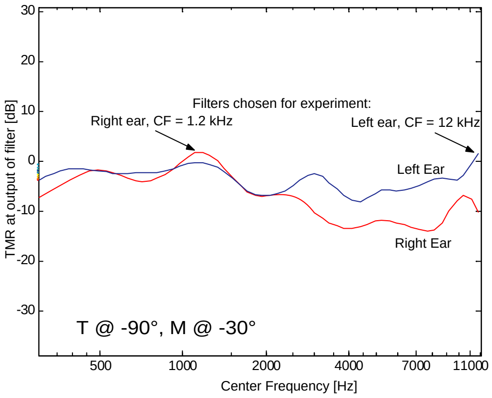


Figure 4: Description of figure identical to Figure 2. Spatial configuration with T at -90° and M at 90°.

a) TMR at output of model peripheral filters



b) Mesured and predicted threshold TMR

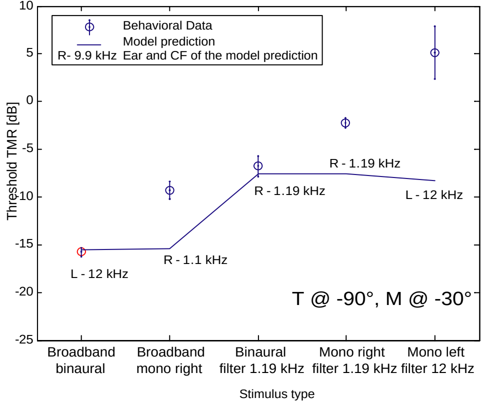
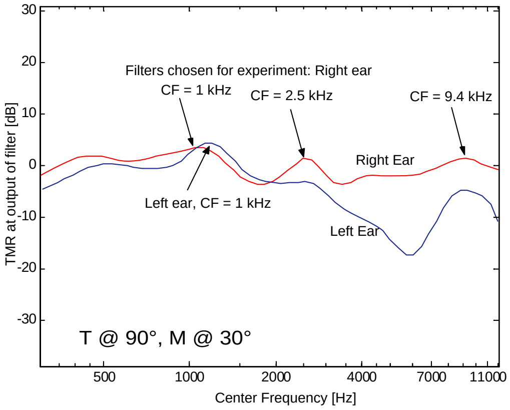


Figure 5: Description of figure identical to Figure 2. Spatial configuration with T at -90° and M at -30°.

The results obtained with T at -90° and M at -30° are shown in Figure 5. Fig. 5a shows that there are at least two channels with very good TMR in this configuration. The results show that binaural processing influenced the broadband binaural threshold (in Fig. 5b this threshold is much better than the others). However, there is still a good match between the two binaural thresholds and their predictions, as well as between the right-ear monaural thresholds and their predictions, suggesting that no across-channel integration needs to be evoked. Note that here again the left-ear monaural threshold is incorrectly predicted, probably for reasons similar to those discussed above for Figure 4.

The most challenging spatial configuration was that with T at 90° and M at 30° (Figure 6) where there are multiple low- and high-CF channels in both ears with approximately equal TMR (Fig. 6a). Figure 6b shows eight different measured and predicted thresholds, considered in this configuration. First, comparison of the broadband binaural threshold (left-most circle) to the right-ear (second from left) and the left-ear (third from right) threshold shows that binaural processing contributed to the detection of broadband target. However, when binaural processing is accounted for by fitting the model to the monaural broadband thresholds (dashed lines), both left- and right-ear thresholds can be predicted by the best-TMR channels. Particularly interesting is the comparison of the predictions and data in the right ear (second through fifth point from the left in Fig 6b). Here, the broadband prediction is based on the low-CF channel, however, it is the high-CF channel that gives the lowest narrowband threshold. Moreover, the narrowband thresholds improve with increasing CF while the model predictions worsen with increasing CF, resulting in the low-CF threshold being much worse than predicted. This discrepancy is again consistent with the errors discussed above, in which the model had the tendency to predict better performance if the prediction was based on a low-CF channel.

a) TMR at output of model peripheral filters



b) Mesured and predicted threshold TMR

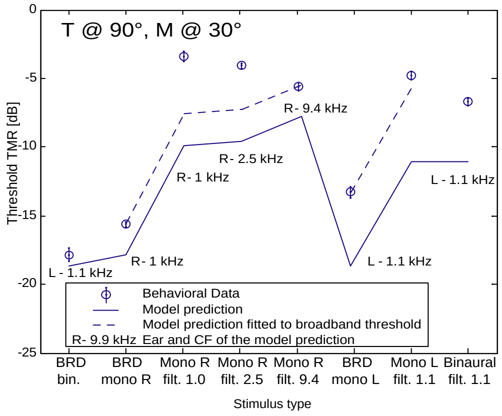


Figure 6: Description of figure identical to Figure 2. Spatial configuration with T at 90° and M at 30°.

4 Discussion and conclusions

The results of this study do not support the hypothesis that across-channel integration is necessary when considering spatial unmasking of stimuli with varying bandwidth. However, there were several occasions when contribution of binaural processing was observed, so considering binaural processing is important even for these broadband stimuli.

The only re-occurring error of the model was that the model tended to predict better performance than measured when the prediction was based on a low-CF channel. First, this error is probably not due to the model's lack of binaural processing or across-frequency integration, because both these mechanisms would make predictions even better, i.e., the error would be larger. Instead, the errors might be due to several other assumptions that the model makes. First, the model uses a gammatone filter bank to model peripheral processing. The observed errors in predictions might result from the gammatone filter being a more accurate model of auditory periphery at high frequencies than at low frequencies. Second, the model assumes that the threshold TMR is constant and independent of the filter CF. Again, assuming that the threshold TMR is higher at lower frequencies could correct the errors in predictions. And last, the stimuli used in this study produce 40-Hz amplitude modulation at the output of the peripheral filters. It might be that this modulation is used as a detection cue and that this cue is more efficient at higher filter CFs than at low CFs.

Further studies are needed to determine the actual source of this error, as well as to fully understand the importance of modulation, binaural, and across-channel processing for spatial release from masking of non-speech and speech stimuli.

Acknowledgement

I would like to thank Branislav Beníkovský and Anton Baša whose diploma theses were important sources of information for this study. The research reported here was partially supported by grants from the Slovak Scientific Grant Agency (grant VEGA 1/1059/04) and the U.S. National Academy of Sciences.

References

[1] C.C. Lane, N. Kopco, B. Delgutte, B. G. Shinn-Cunningham, and H. S. Colburn. ‘A cat's cocktail party: Psychophysical, neurophysiological, and computational studies of spatial release from masking’ In: *Auditory signal processing: Physiology, psychoacoustics, and models.* (Pressnitzer, D., de Cheveigné, A, McAdams, S., and Collet, L., eds), pp 327-333, Springer, New York. (Proc. International Symposium on Hearing, Dourdan, France, Aug. 24-29, 2003)

[2] Saberi, K., Dostal, L., Sadralodabai,T., Bull, V., and Perrott, D.R. ‘Free-field release from masking.’ *J. Acoust. Soc. Am.* 90, 1355-1370. (1991)

[3] Good, M.D., Gilkey, R.H., and Ball, J.M. ‘The relation between detection in noise and localization in noise in the free field.’ In R.H. Gilkey and T.R. Anderson (Eds), *Binaural and Spatial Hearing in Real and Virtual Environments.* Lawrence Erlbaum Associates, Mahwah, N.J, pp 349–376. (1997)

[4] B. Beníkovský, ‘*Spatial unmasking of broadband stimuli in a virtual auditory environment*’, Unpublished diploma thesis, TU Košice (2004)

[5] A. Baša, ‘*Význam priestorového vnímania a spracovania modulovaných podnetov pri počúvaní v komplexnom prostredí (Importance of spatial hearing and processing of modulated stimuli for listening in complex environments.)*’, Unpublished diploma thesis, TU Košice (2005)

[6] Johannesma, P.I.M. ‘The pre-response stimulus ensemble of neurons in the cochlear nucleus.’ In: B.L. Cardozo, E. de Boer, and R. Plomp (Eds.), *IPO Symposium on Hearing Theory*. IPO, Eindhoven, The Netherlands, pp. 58-69. (1972).