Spatial release from masking as a function of the spectral overlap of competing talkers (L)[a)]

Virginia Best,[b)] Eric R. Thompson,[c)] Christine R. Mason, and Gerald Kidd, Jr.
Department of Speech, Language and Hearing Sciences and Hearing Research Center, Boston University, Boston, Massachusetts 02215

(Received 3 February 2013; revised 28 March 2013; accepted 8 April 2013)

This study tested the hypothesis that the reduced spatial release from speech-on-speech masking typically observed in listeners with sensorineural hearing loss results from increased energetic masking. Target sentences were presented simultaneously with a speech masker, and the spectral overlap between the pair (and hence the energetic masking) was systematically varied. The results are consistent with increased energetic masking in listeners with hearing loss that limits performance when listening in speech mixtures. However, listeners with hearing loss did not exhibit reduced spatial release from masking when stimuli were filtered into narrow bands.
 [http://dx.doi.org/10.1121/1.4803517]

PACS number(s): 43.71.Ky, 43.66.Pn, 43.66.Dc [JFC] Pages: 3677–3680

I. INTRODUCTION

Listening in speech mixtures can be difficult as a result of at least two kinds of auditory masking. "Energetic masking" (EM) occurs when a competing sound interferes with the peripheral representation of a target sound as a result of acoustic overlap. "Informational masking" (IM) refers to interference that occurs despite a target that is well-represented peripherally (for review, see Kidd *et al.*, 2008b). In the context of speech perception, the interference caused by noise provides an example of predominantly EM and that caused by other distracting talkers an example of predominantly IM. However, in many listening situations, both of these kinds of masking likely affect speech understanding (Brungart, 2005; Stone *et al.*, 2012).

Spatially separating competing sounds can reduce both EM and IM, although spatial release from masking (SRM) can be much larger for situations dominated by IM than those dominated by EM (e.g., Freyman *et al.*, 1999; Freyman *et al.*, 2001; Arbogast *et al.*, 2002; Marrone *et al.*, 2008a). Ultimately, it is the balance of IM and EM that determines how much benefit spatial separation will provide when listening in speech mixtures (Kidd *et al.*, 2010).

Attempts to understand why listeners with sensorineural hearing impairment (HI) are more susceptible to interference than those with normal hearing (NH) have generally concluded that increased EM is more of a problem than increased IM (e.g., Helfer and Freyman, 2008; Agus *et al.*, 2009). Indeed one of the hallmarks of sensorineural hearing loss is increased susceptibility to noise. Furthermore, it has been suggested that the increased EM in HI listeners may be part of the reason they experience reduced spatial benefits in speech mixtures (Arbogast *et al.*, 2005; Marrone *et al.*, 2008b; Best *et al.*, 2011), but the evidence is far from conclusive.

Arbogast and colleagues (2005) measured SRM under conditions in which EM was reduced by reducing the spectral overlap of target and masker talkers. To achieve this, they processed each talker into a set of narrow spectral bands using sine-vocoding and presented mutually exclusive subsets of these bands for the target and the masker. They found that HI listeners showed less SRM than NH listeners under these conditions; this suggested that the reduced spatial benefit was not related to EM. However, a control in which the masker talker was replaced by noise indicated that significant amounts of EM remained in the HI listeners, presumably as a result of reduced frequency selectivity. Thus the authors were unable to rule out the idea that increased EM was a factor limiting SRM.

The aim of the current experiment was to revisit the basic approach of Arbogast and colleagues but with several modifications designed to give us better control over the amount of EM present. First, we attempted to achieve better spectral isolation between the target and masking talkers by using fewer, more widely spaced, bands. Our hope here was that in the case of least overlap, we might eliminate EM completely even in the HI group. Second, we systematically varied the spectral overlap of the speech stimuli from non-overlapping to completely overlapping so we could examine performance across a continuum. We also included a noise control as a direct measure of the amount of EM present in each spectral overlap condition. We predicted that as spectral overlap (and EM) was reduced, both intelligibility and spatial benefit would be more comparable between NH and HI listeners.

II. METHODS

A. Participants

Seven NH listeners (18–28 yr of age; mean, 20) and seven HI listeners (18–42 yr of age; mean, 24) participated. The NH listeners were screened to ensure that their pure-tone thresholds were in the normal range for octave frequencies from 250 to 8000 Hz. The HI listeners had mild to moderately severe, bilateral, symmetric, sloping, sensorineural hearing

[a)]Portions of this work were presented at the Association for Research in Otolaryngology MidWinter Meeting, San Diego, February 2012.
[b)]Present address: National Acoustic Laboratories, Macquarie University, NSW, Australia.
[c)]Present address: Ball Aerospace and Technology Corp., Fairborn, OH 45324.

losses, with pure-tone averages in the range 24–62 dB (mean, 37 dB). Five of the seven HI listeners were regular bilateral hearing aid users, but their aids were removed for testing. All listeners were paid for their participation.

B. Stimuli

The experiment made use of a corpus of monosyllabic words recorded at Boston University's Hearing Research Center (for details see Kidd *et al.*, 2008a). A target sentence was created by concatenating five words from the same male talker to create a syntactically correct but unpredictable sentence that always started with a person's name (e.g., Bob found four red toys). Speech maskers were created by concatenating random words spoken by a different male talker than the target talker. To ensure that the target was fully masked, each speech masker was created using enough words to exceed the length of the target sentence (five to seven words). Two such masker strings were generated on each trial and were added and treated as one masker for the purposes of setting the target-to-masker ratio (TMR). In conditions containing a noise masker, a speech masker was generated as just described and then its magnitude spectrum was extracted and used to shape a randomly generated broadband noise.

Stimuli were either presented with their natural spectrum ("unfiltered") or they were filtered into narrow spectral bands. We chose simple filtering instead of vocoding (cf. Arbogast *et al.*, 2005) to maintain the natural structure of speech as much as possible while manipulating the spectral overlap. In filtered conditions, the target was filtered into four bands with logarithmically spaced center frequencies (310, 676, 1475, and 3218 Hz) and a bandwidth of 20% of center frequency. These bands were selected on the basis of pilot studies to be spaced well apart and to give good intelligibility in quiet when combined. Filtering was done using the "fir1" command in MATLAB (MathWorks Inc.). Maskers were filtered into four bands in the same way with center frequencies that either matched those of the target (100% overlap) or were shifted up or down by various amounts (see Fig. 1). In the condition of least overlap (0% overlap), the masker bands were shifted to be centered in the spectral gaps between target bands, giving rise to center frequencies of 210, 458, 999, 2179 Hz (downward shift) or 458, 999, 2179, 4752 Hz (upward shift). Three intermediate shifts were also tested that were equal in log-frequency and spanned the range between 0% and 100% overlap (referred to as 25%, 50%, and 75% overlap).

Target stimuli were presented at a fixed level of 45 dB sound pressure level (SPL) (calculated post-filtering in the filtered conditions) and masker levels were varied according to the TMR. Five TMRs were chosen per listener group, masker type, and filtering condition, to cover a suitable range of the psychometric function. To improve the audibility of stimuli in the HI group, individualized linear gain was provided to the stimulus according to the NAL-RP prescription (Byrne *et al.*, 1991) before delivery via the headphones.

While the target sentence was always presented diotically, the masker sentence was presented with one of two spatial configurations. In the "co-located" configuration, the masker sentence was also diotic. In the "separated" configuration, the masker sentence was lateralized to the right side by introducing an interaural time difference of 0.6 ms.

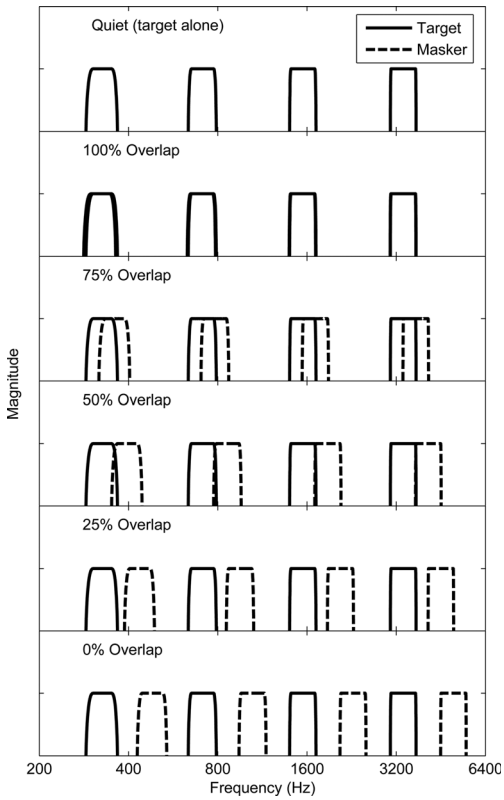

FIG. 1. Schematic of the different spectral overlap conditions. The target (top panel) was always composed of four narrow bands at fixed frequencies. The masker was also composed of four narrow bands, but they were shifted in frequency such that they completely overlapped with the target (100% overlap), or were centered in between the target bands (0% overlap) with three intermediate levels of overlap. Both up- and downward shifts in frequency were tested, but only upward shifts are illustrated here.

C. Procedures

Digital stimuli were generated on a PC using MATLAB software, D/A converted, and attenuated using Tucker-Davis Technologies hardware (System II), and presented over Sennheiser headphones (HD 280 Pro). Listeners were seated in a sound-treated booth fitted with a monitor and mouse and indicated their responses by clicking with the mouse on a graphical user interface. Each word in the response was scored separately and used to generate scores in percent correct for each TMR under each condition. Logistic fits to psychometric functions were generated for each condition, and threshold was defined as the TMR corresponding to the 50% correct point on these fits.

Listeners first participated in a short experiment using unfiltered stimuli to provide a baseline of performance and to familiarize them with the stimuli and task. This experiment began with one or two 20-trial blocks in which the unfiltered target was presented diotically in quiet. Following this, four masked conditions (speech/noise masker, co-located/separated configuration) were completed again using unfiltered stimuli. Each masked block consisted of 25 trials (five repeats at each of five TMRs).

Following the initial experiment using unfiltered stimuli, listeners then moved on to the main experiment in which the

stimuli were filtered as described in Sec. II B. The main experiment began with a number of 45-trial blocks in which listeners identified the filtered target sentences in quiet (five trials for each of the nine masker band combinations). NH listeners completed two to four blocks until performance was above 90% for two consecutive blocks; the mean score across these two blocks is reported in the following text. HI listeners completed three to six blocks and were given small amounts (3–8 dB) of additional gain if initial scores were very low. The mean score across the best two blocks is reported below. Listeners then completed seven blocks each of the four masked conditions (speech/noise masker, co-located/separated configuration) for the filtered stimuli. Each of the four conditions was completed once in a random order before moving on to the next set of four. Each block consisted of 45 trials (one trial per combination of five TMRs and nine spectral overlaps). For the purposes of analysis, the upward and downward shifts giving the same proportion of overlap were combined.

At the completion of the main experiment, a short test was conducted to estimate the sensation level of the filtered target. For NH listeners, five repetitions of target sentences at each of ten target levels (from 5 to 50 dB SPL in 5 dB steps) were presented and a psychometric function generated. Sensation level was calculated by subtracting the level giving rise to 50% correct identification from the presentation level. For HI listeners, the same procedure was followed, except target levels from 10 to 70 dB SPL (in 10 dB steps) were used and the NAL-RP gain (plus any additional gain they received) was provided.

III. RESULTS

A. Quiet performance

For unfiltered stimuli, NH and HI groups had mean quiet identification scores of 99% and 96%, respectively. For filtered stimuli, mean quiet identification scores were 93% and 89%, with little variation in scores as a function of band choice, confirming that both targets and maskers were equally intelligible. Sensation levels were 29 and 17 dB on average in the NH and HI groups, respectively, and individual sensation levels were correlated with quiet performance scores for filtered stimuli [$r = 0.73$, $p < 0.005$]. Thus it appears that the NAL-RP gain did not completely restore audibility for the HI group.

B. Noise maskers

The left column of Fig. 2 shows results for each listener group when the masker was noise. The top and middle rows show thresholds for co-located and separated configurations, respectively, and the bottom row shows the difference between these (i.e., the SRM). For unfiltered stimuli (leftmost points), mean co-located and separated thresholds were very similar for the NH and HI groups. SRM was similar in the two groups (4.1 and 5.1 dB) and not significantly different according to a t-test [$t(12) = 1.1$, $p = 0.28$].

For filtered stimuli, thresholds improved as spectral overlap was reduced in both co-located and separated configurations. Note that thresholds for the lower amounts of spectral overlap were *better* than in the unfiltered condition despite

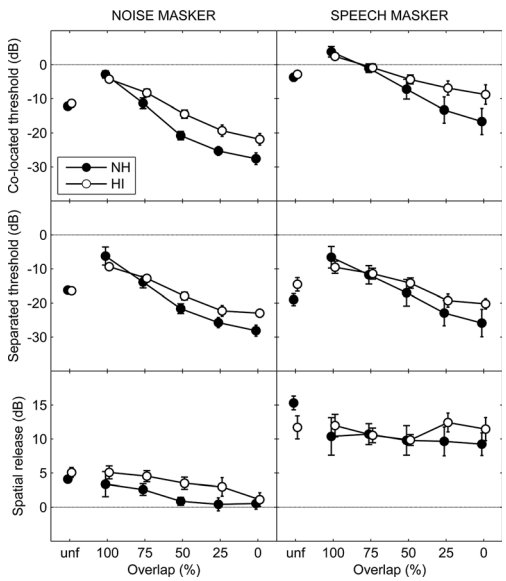

FIG. 2. Results for each listener group as a function of spectral overlap (upward and downward shifts have been averaged). The left and right columns show data for noise and speech maskers, respectively. The top and middle rows show target-to-masker ratios at threshold for co-located and separated configurations, respectively, and the bottom row shows the subtraction of these (SRM). The isolated leftmost points in each panel show results for the condition in which no filtering was applied to the stimuli.

the impoverished speech signals. The HI group showed poorer thresholds than the NH group for all levels of spectral overlap except 100%, suggesting that noise in non-target bands caused more EM in this group. Poorer thresholds for the HI group even in the condition of least overlap (0%) suggests that we did not achieve our original goal of creating a condition in which EM was reduced so much as to equalize performance in the two groups. Nonetheless, the difference between groups in this condition (around 6 dB) is smaller than the difference between groups reported by Arbogast and colleagues for a denser arrangement of more target and masker bands (around 13 dB; Arbogast *et al.*, 2005).

Both groups showed a modest SRM for noise maskers that ranged from 3.4/5.1 dB for 100% overlap to 0.5/1.1 dB for 0% overlap. The drop in SRM with decreasing spectral overlap likely reflects the fact that there was less masking to release. Moreover, SRM was slightly lower in the NH group overall, suggesting that they experienced less EM to begin with. An analysis of variance (ANOVA) revealed that the SRM was significantly affected by overlap condition [$F(4,48) = 3.3$, $p < 0.05$] and listener group [$F(1,12) = 9.7$, $p < 0.01$], with no significant interaction [$F(4,48) = 0.3$, $p = 0.88$].

C. Speech maskers

The right column of Fig. 2 shows results for each listener group when the masker was speech. For unfiltered stimuli (leftmost points), NH and HI listeners performed similarly in the co-located configuration, but the NH listeners had lower thresholds in the separated configuration. Accordingly, the NH group had a larger SRM on average (15.3 vs 11.7 dB) but this difference did not reach significance [$t(12) = 1.8$, $p = 0.09$].

For filtered stimuli, the overall pattern of results was very similar to that observed for noise maskers. Again,

thresholds improved with reduced spectral overlap, surpassing those seen for unfiltered stimuli. One expected difference was that thresholds in the co-located condition were poorer for speech maskers than noise maskers in both groups, and this can be attributed to additional IM caused by target-masker similarity. Thresholds for the separated configuration, however, were approximately similar for speech and noise maskers, suggesting that most of the IM was alleviated by the imposed spatial differences.

The HI group showed poorer thresholds than the NH group for all levels of spectral overlap except the most highly overlapping in both spatial configurations (see Fig. 2). A relatively large SRM was observed for both groups in the speech masker condition. Surprisingly, no systematic difference in SRM was seen in any of the unfiltered conditions. An ANOVA revealed no significant effect of overlap condition [$F(4,48) = 0.3$, $p = 0.87$] or listener group [$F(1,12) = 0.5$, $p = 0.49$], and no significant interaction [$F(4,48) = 0.5$, $p = 0.77$].

IV. DISCUSSION

For speech masked by noise, both NH and HI listeners showed improved thresholds when the spectral overlap of the target and masker was reduced. This is consistent with the results of other studies (e.g., Arbogast *et al.*, 2002, 2005; Apoux and Healy, 2010) and primarily reflects a reduction in EM. However, this improvement was more marked in NH than HI listeners, suggesting that the HI group had some residual EM even when the acoustic overlap of the sounds was minimized by our processing. This may reflect the poor spectral resolution in these listeners; this would result in more overlap in the neural representation of the competing sounds for a given acoustic overlap. Spectral release from masking was also apparent for speech maskers, although again was stronger in the NH group.

For all listeners, SRM was larger for speech maskers than for noise maskers consistent with previous studies. However, a surprising finding in this study was that the HI group did not show reduced SRM for speech-on-speech masking like that observed in previous experiments (e.g. Arbogast *et al.*, 2005; Marrone *et al.*, 2008b). Although there was a tendency in this direction for unfiltered speech, it disappeared when targets and maskers were filtered into narrow bands. The NH group showed less SRM for filtered than unfiltered stimuli, closing the gap between groups (and even reversing the order of the groups for some spectral overlaps). For some spectral overlaps, the small SRM in the NH group may be a headroom issue, related to the large *spectral* release from masking already obtained. However, this cannot explain why the NH and HI groups showed equivalent SRM in the most highly overlapping conditions. Perhaps here the spectral "gaps" introduced to our stimuli alleviated the smearing of target and masker energy that reduced frequency resolution would normally cause, thus alleviating to some extent the increased EM shown by HI listeners in unfiltered mixtures.

Finally, thresholds in the separated speech masker configuration differed consistently between listener groups and were closely related to the corresponding noise thresholds. As the noise thresholds represent the amount of EM in the different spectral overlap conditions, this correspondence suggests that "best performance" in this task was determined by EM. HI listeners experienced more EM, and thus their performance was ultimately more limited. It is possible that this limit restricts the benefits achievable in these listeners under many circumstances not tested here, e.g., given attentional cues, amplification by hearing aids, etc.

V. CONCLUSION

Listeners with hearing loss did not exhibit reduced spatial release from masking when stimuli were filtered into narrow bands. However, the results are consistent with increased energetic masking in this population that limits performance when listening in speech mixtures.

ACKNOWLEDGMENTS

This work was supported by grants from NIH/NIDCD and AFOSR.

Agus, T. R., Akeroyd, M. A., Gatehouse, S., and Warden, D. (**2009**). "Informational masking in young and elderly listeners for speech masked by simultaneous speech and noise," J. Acoust. Soc. Am. **126**, 1926–1940.

Apoux, F., and Healy, E. W. (**2010**). "Relative contribution of off- and on-frequency spectral components of background noise to the masking of unprocessed and vocoded speech," J. Acoust. Soc. Am. **128**, 2075–2084.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2005**). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **117**, 2169–2180.

Best, V., Mason, C. R., and Kidd, G., Jr. (**2011**). "Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the temporal overlap of competing talkers," J. Acoust. Soc. Am. **129**, 1616–1625.

Brungart, D. (**2005**). "Informational and energetic masking effects in multitalker speech perception," in *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluwer Academic, Dordrecht), pp. 261–267.

Byrne, D. J., Parkinson, A., and Newall, P. (**1991**). "Modified hearing aid selection procedures for severe-profound hearing losses," in *The Vanderbilt Hearing Aid Report II*, edited by G. A. Studebaker, F. H. Bess, and L. B. Beck (York, Parkton, MD), pp. 295–300.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Helfer, K. S., and Freyman, R. L. (**2008**). "Aging and speech-on-speech masking," Ear Hear. **29**, 87–98.

Kidd, G., Jr., Best, V., and Mason, C. R. (**2008a**). "Listening to every other word: Examining the strength of linkage variables in forming streams of speech," J. Acoust. Soc. Am. **124**, 3793–3802.

Kidd, G., Jr., Mason, C. R., Best, V., and Marrone, N. (**2010**). "Stimulus factors influencing spatial release from speech-on-speech masking," J. Acoust. Soc. Am. **128**, 1965–1978.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (**2008b**). "Informational masking," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer Handbook of Auditory Research, New York), pp. 143–190.

Marrone, N., Mason, C. R., and Kidd, G., Jr. (**2008a**). "Tuning in the spatial dimension: Evidence from a masked speech identification task," J. Acoust. Soc. Am. **124**, 1146–1158.

Marrone, N., Mason, C. R., and Kidd, G., Jr. (**2008b**). "The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms," J. Acoust. Soc. Am. **124**, 3064–3075.

Stone, M. A., Fullgrabe, C., and Moore, B. C. J. (**2012**). "Notionally steady background noise acts primarily as a modulation masker of speech," J. Acoust. Soc. Am. **132**, 317–326.