Research Note

Listeners Experience Linguistic Masking Release in Noise-Vocoded Speech-in-Speech Recognition

Navin Viswanathan,^a Kostas Kokkinakis,^a and Brittany T. Williams^a

Purpose: The purpose of this study was to evaluate whether listeners with normal hearing perceiving noise-vocoded speech-in-speech demonstrate better intelligibility of target speech when the background speech was mismatched in language (linguistic release from masking [LRM]) and/or location (spatial release from masking [SRM]) relative to the target. We also assessed whether the spectral resolution of the noise-vocoded stimuli affected the presence of LRM and SRM under these conditions.

Method: In Experiment 1, a mixed factorial design was used to simultaneously manipulate the masker language (within-subject, English vs. Dutch), the simulated masker location (within-subject, right, center, left), and the spectral resolution (between-subjects, 6 vs. 12 channels) of noise-vocoded target-masker combinations presented at +25 dB signal-to-noise ratio (SNR). In Experiment 2, the study was

repeated using a spectral resolution of 12 channels at +15 dB SNR.

Results: In both experiments, listeners' intelligibility of noise-vocoded targets was better when the background masker was Dutch, demonstrating reliable LRM in all conditions. The pattern of results in Experiment 1 was not reliably different across the 6- and 12-channel noise-vocoded speech. Finally, a reliable spatial benefit (SRM) was detected only in the more challenging SNR condition (Experiment 2).

Conclusion: The current study is the first to report a clear LRM benefit in noise-vocoded speech-in-speech recognition. Our results indicate that this benefit is available even under spectrally degraded conditions and that it may augment the benefit due to spatial separation of target speech and competing backgrounds.

In typical listening conditions, listeners are often faced with the task of perceiving speech in the presence of competing background speech. This situation, often described as the cocktail party effect (Cherry, 1953), has been well studied to determine what factors affect performance in such settings (for a detailed review, see Bronkhorst, 2000). The interference exerted by the background signal is broadly classified into two kinds. *Energetic masking* is used to refer to the interference effect of distracting background signals on target speech that is due to the spectro-temporal overlap of the signals. Any aspect of interference that goes beyond energetic masking is subsumed under the term *informational masking*. In speech-in-speech perception, the relative contributions of energetic and informational masking

have been shown to vary depending on listening conditions and the similarity of the signals to be segregated (Bronkhorst, 2000, 2015). An often-observed finding is that listeners typically demonstrate improved performance when the target and background speech originate from different spatial locations in the acoustic field—a phenomenon termed as spatial release from masking (SRM; Plomp & Mimpen, 1981). A typical explanation for SRM is that this reduction in masking occurs when listeners utilize spatial acoustic cues to perceptually segregate and selectively attend to the target source (e.g., see Brungart, 2001; Freyman, Balakrishnan, & Helfer, 2001; Freyman, Helfer, McCall, & Clifton, 1999; Hawley, Litovsky, & Culling, 2004). SRM is a well-documented effect and, in listeners with normal hearing, often produces a substantial benefit, especially when multiple maskers are present and the maskers and target speech are composed of talkers that can be confused with one another (Hawley et al., 2004). In this study, we will focus on a different source of improvement in perceptual performance in speech-in-speech recognition: one that arises when the background speech is linguistically distinct from the target. We will also examine how this form of masking release interacts with SRM.

^aSpeech-Language-Hearing Sciences and Disorders, University of Kansas, Lawrence

Correspondence to Navin Viswanathan: navin@ku.edu

Editor-in-Chief: Frederick (Erick) Gallun

Editor: Matthew Fitzgerald

Received June 2, 2017

Revision received August 18, 2017

Accepted October 21, 2017

https://doi.org/10.1044/2017_JSLHR-H-17-0215

Disclosure: The authors have declared that no competing interests existed at the time of publication.

A set of recent studies has investigated the effect of manipulating the linguistic properties of background speech relative to the target. Broadly, listeners' performance in speech-in-speech recognition is improved when the similarity of the target and competing maskers are reduced (e.g., Miller & Licklider, 1950). Applied to a situation that involves competing speech in a foreign language, the target-masker linguistic similarity hypothesis suggests that listeners can more efficiently segregate two competing streams when the background speech differs along specific linguistic dimensions from the target speech. For instance, Van Engen and Bradlow (2007) demonstrated that native English speakers' report of English target sentences was improved when the background consisted of non-native Mandarin babble rather than native English babble. To describe the performance improvement resulting from the target-masker linguistic mismatch, terms such as (same) language interference effect (Cooke, Garcia Lecumberri, & Barker, 2008; Mattys, Brooks, & Cooke, 2009), target-masker language mismatch effect (e.g., Brouwer, Van Engen, Calandruccio, & Bradlow, 2012), and linguistic release from masking (LRM; e.g., see Calandruccio, Leibold, & Buss, 2016; Viswanathan, Kokkinakis, & Williams, 2016) have been widely used. To parallel our previous discussion of SRM, we will use LRM in the remainder of this article to describe this benefit.

Several follow-up studies have demonstrated that LRM effects extend to other masking languages (e.g., non-native Dutch maskers with English target speech; Brouwer et al., 2012) as well as to accented versions of a language (e.g., Mandarin-accented English on the native dialect of American English; Calandruccio, Dhar, & Bradlow, 2010). Broadly, the magnitude of LRM effects is directly related to the overall typological similarity of the languages being tested (Calandruccio, Brouwer, Van Engen, Dhar, & Bradlow, 2013). Furthermore, the effect of LRM persists even when non-native speakers are tested on the report of their L2 with interfering L1 backgrounds. However, in these cases, the LRM benefit is diminished (Mandarin-English bilinguals, Van Engen & Bradlow, 2007; Dutch-English bilinguals, Brouwer et al., 2012).

In a recent study, Viswanathan et al. (2016) assessed the effects of simultaneously varying the spatial location and the language of background speech relative to the target. The authors noted that typical studies of LRM use spatially colocated target-masker combinations, making it unclear whether LRM persists when the target and masker are spatially separated. To investigate this, the authors presented competing speech maskers that were spoken in either Dutch or English from a simulated location to the left, the right, or the same (frontal) location as the English target speech. Across different signal-to-noise ratios (SNRs), they found that the LRM benefit persisted even when the masker and target were spatially separated, indicating that this benefit may be available to listeners under conditions of typical listening. Second, they also detected a reliable SRM, suggesting that these benefits may co-occur in these conditions.

The studies reviewed thus far indicate that while speech-in-speech recognition is challenging due to the presence of

competing signals, listeners with normal hearing are able to cope with this challenge by attuning to specific spectrotemporal variations in the signal to segregate the target from the background. However, for listeners with impaired access to the acoustic signal, this process of perceptual segregation in everyday listening conditions proves to be especially difficult. For instance, listeners with hearing impairment show considerably poorer speech intelligibility during typical multitalker listening tasks. In particular, for individuals fitted with cochlear implant (CI) devices, who only have limited access to the target signal (Fu & Nogaki, 2005), speech-in-speech recognition remains extremely challenging. When addressing the various deficits in speech intelligibility experienced by CI users in the presence of background noise or competing speech, one prime example is the lack of a reliable SRM. With target speech in front and competing speech emanating from various other locations, the ability of CI listeners to make use of spatial cues in order to perceptually separate and identify the target speech from one or more interfering sources has been shown to only marginally improve regardless of the number, type, and spatial location of maskers (Loizou et al., 2009).

A typical approach to simulate the challenges that result from the reduced spectral resolution in CI devices is through the use of noise-vocoded speech (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). In quiet, even with as few as four active frequency channels, vocoded speech can be understood remarkably well (e.g., Dorman, Loizou, & Rainey, 1997; Shannon et al., 1995). However, in the presence of background noise, more spectral bands are necessary (8–12) in order to achieve the same level of performance (Dorman, Loizou, Fitzke, & Tu, 1998). In fact, when listening to speech processed through a noise vocoder with a relatively small number of channels, the loss of spectral information is compounded by the lack of reliable spatial cues, making it more difficult for the listeners to determine the spatial location of the masker. Hence, these listeners often experience more difficulty when attempting to perceptually segregate the two streams effectively (Qin & Oxenham, 2003). Freyman, Balakrishnan, and Helfer (2008) investigated what factors determine whether typical hearing listeners benefit from the spatial separation of target and masker when presented with noise-vocoded speech. Interestingly, in their Experiment 1, the authors found that when listeners were presented with nonsense sentences in a two-talker babble under six-channel noise vocoding, SRM was not detected. They suggested that the lack of SRM in these impoverished signals could result from the lack of clear spatial cues to the locations of competing streams or from an overall reduction of informational masking in the colocated condition or be due to the nature of listening challenges in these conditions. Specifically, they argued that because listeners required highly positive SNRs to even achieve modest intelligibility performance (around 30% correct) in these conditions, the target was already clearly distinct from the competing masking speech, precluding any additional benefits from SRM. In support of this hypothesis, in their Experiment 2, they found robust SRM when

listening performance could be tested at lower SNRs using a four-alternative choice task. This finding again indicates that listeners with impaired access to the competing signals do not always benefit from a mismatch in the spatial location between the target and competing speech. This finding, along with others, indicates that when listeners with normal hearing are presented with noise-vocoded target–masker stimuli, SRM benefits either decrease significantly or become entirely absent (Best, Marrone, Mason, & Kidd, 2012; Freyman et al., 2001; Stickney, Zeng, Litovsky, & Assmann, 2004). In contrast, other studies have reported a contradictory pattern of findings showing a consistent SRM under noise-vocoded conditions (e.g., Garadat, Litovsky, Yu, & Zeng, 2009, with 8- and 16-channel vocoding; Swaminathan, Mason, Streeter, Best, & Kidd, 2013, with 8-channel vocoding). Broadly, these findings suggest that listeners with impaired access to the competing signals can, but do not always, benefit from a mismatch in the spatial location between the target and competing speech. Listening difficulty in these situations is accentuated when listeners only have access to a small number of spectral channels providing independent envelope cues either in listeners with CI through electrical stimulation or in typical hearing listeners presented with vocoded speech. However, it is currently unknown whether these perceivers experience LRM under these listening conditions. The experiments reported in the current study aim to address this gap by investigating whether the masking release due to a linguistic mismatch (LRM benefit) is available to listeners perceiving noise-vocoded speech. We suggest that this investigation is warranted because, as noted, LRM has not been examined in noise-vocoded speech or with CI listeners, making it unclear whether this benefit is indeed available under such impoverished listening situations. By examining LRM and SRM simultaneously, we will be able to clarify the nature of challenges in these situations by further dissociating the relative contributions of auditory spatial information, diminished informational masking in colocated conditions, as well as the need for highly positive SNRs. To do so, in Experiment 1, we examined the effect of simultaneously varying the masker language and simulated location relative to the target in reduced spectrum speech-in-speech recognition situations at +25 dB. In addition, we tested whether altering the spectral resolution of noise-vocoded speech by using either 6-channel (similar to Freyman et al., 2001) or 12-channel vocoding affects the observed LRM and SRM effects. In Experiment 2, we used a more challenging SNR (+15 dB) to test the same factors only in the higher spectral resolution 12-channel condition. These pairs of experiments investigated whether LRM and SRM systematically covary under these listening conditions.

Experiment 1

Method

Participants

Twenty-six undergraduate students from the University of Kansas ranging between 18 and 21 years old participated

in two speech perception experiments (median age was 20 years). All listeners were native speakers of American English with no knowledge of Dutch. All participants reported no history of speech, hearing, or language disorders and received course credit for their participation.

Materials

All stimuli were recorded by a speaker reading visually presented sentences in a sound-attenuating booth at a 44.1-kHz sampling rate and a 16-bit resolution. Unprocessed versions of these stimuli were used by Viswanathan et al. (2016). The target stimuli were chosen from the Bamford-Kowal-Bench Standard Sentence Test (Bench, Kowal, & Bamford, 1979) and were recorded by one female native English speaker. Each sentence contained three or four key words for a total of 169 key words (e.g., *The CLOWN had a FUNNY FACE*). Background sentences were produced by two female native English speakers and consisted of syntactically well-formed, semantically simple sentences (e.g., *The wrong shot led the farm*) from the Syntactically Normal Sentence Test (Nye & Gaitenby, 1974). For the Dutch background conditions, translated versions of these sentences (e.g., *Het verkeerde schot leidde de Boerderij*) recorded by two female native Dutch speakers were used. A two-talker babble masker was created by mixing waveforms of each pair of speakers for each language. Following Brouwer et al. (2012), unequal amounts of energetic masking were reduced by minimizing the differences in the long-term average speech spectrum of each language masker. Random portions were excised from this babble to serve as background stimuli for each trial. The background stimuli always preceded the target onset by 0.5 s and persisted 0.5 s after the target offset. Root-mean-square levels of all target sentences were equalized to 65 dB SPL, and the background maskers were equalized to 40 dB SPL to produce an SNR of +25 dB. Through pilot testing of three listeners and consistent with past studies (e.g., Freyman et al., 2008), it was established that this SNR level allows the listener to have access to the content of the target speech while avoiding floor and ceiling effects.

Signal Processing

For each target and masker configuration, three different spatial listening conditions were created by processing the stimuli with premeasured head-related transfer functions (HRTFs) in order to simulate the virtual spatial locations of the target and interfering speech (see Kayser et al., 2009; Viswanathan et al., 2016). These HRTFs were recorded inside an anechoic chamber at the University of Oldenburg (Kayser et al., 2009) using behind-the-ear hearing aid dummy shells placed on the pinnae of an artificial head-and-torso simulator.

The two-talker babble masker originated from either the front of the listener, the right of the listener, or the left of the listener, creating the following three spatial location combinations: (a) target and masker both at 0° azimuth (colocated), (b) target at 0° azimuth and masker at +90° azimuth, and (c) target at 0° azimuth and masker at −90°

azimuth. The left and right ear (binaural) stimuli containing the combined auditory target and masking streams were generated for each spatial location in MATLAB by convolving each target and masking stimulus with each pair of the direction-dependent HRTFs (-90° , 0° , $+90^\circ$).

To simulate CI signal processing, two separate signal processors with 6- and 12-channel noise-band vocoders were used as described in Shannon et al. (1995). The target and masker speech stimuli were first processed through a pre-emphasis high-pass filter (2000-Hz cutoff frequency) with a 3-dB per-octave roll-off, and then band-pass filtered into logarithmically spaced frequency bands between 50 and 8000 Hz using sixth-order Butterworth filters. The equivalent rectangular bandwidth scale was used to allocate the channels (6 or 12) with the specific bandwidth (Glasberg & Moore, 1990). The cutoff and center frequencies for the 6- and 12-channel vocoder processors are provided in Table 1. For each filtered frequency band signal, the temporal envelope was extracted by using full-wave rectification followed by low-pass filtering using a second-order Butterworth filter with a cutoff frequency of 150 Hz. The extracted temporal envelope from each frequency band was then used to modulate a white noise signal delivered by a pseudo-random number generator, and the resulting signal was filtered through the same sixth-order Butterworth filters used for the frequency band selection. Finally, the signals obtained at each frequency band were summed up, and the overall acoustic level was readjusted to match the original sound level. The noise vocoder has been widely utilized as a simulation tool for CI studies, and vocoder simulations based on noise-band vocoders have been previously shown to predict the pattern or trend in performance observed in CI listeners fairly well (e.g., see Dorman et al., 1998; Qin & Oxenham, 2003; Shannon et al., 1995).

Procedure

The listeners were tested individually, seated in front of a computer in a sound-isolated room. Stimuli were presented binaurally through Sennheiser HD-558 headphones (Sennheiser electronic GmbH & Co. KG). Prior to testing,

Table 1. The cutoff and center frequencies (in hertz) for the 6- and 12-channel vocoders, respectively.

Channel	6-Channel			12-Channel		
	Lower	Center	Upper	Lower	Center	Upper
1	50	140	229	50	88	126
2	229	394	558	126	178	230
3	558	860	1161	230	300	369
4	1161	1713	2265	369	464	558
5	2265	3278	4290	558	686	814
6	4290	6145	8000	814	988	1161
7				1161	1395	1630
8				1630	1948	2265
9				2265	2695	3125
10				3125	3708	4290
11				4290	5078	5866
12				5866	6933	8000

listeners transcribed 20 noise-vocoded sentences presented in quiet. Each participant received either 6- or 12-channel noise-vocoded sentences that corresponded with their experimental trials. The listener was first presented with an example of the noise-vocoded sentence, then with the unprocessed version of the same sentence, and, finally, the noise-vocoded version was repeated. All listeners were able to recognize the noise-vocoded sentences in silence after training with at least 75% accuracy. During each experimental trial, one target sentence was presented with the masker. Participants listened to sentences spoken by a female native English speaker in the presence of two-talker background speech. They were informed that the background speech would be perceived from various azimuth locations. Their task was to type what they heard from the target speaker and report individual words if they were unable to identify the whole sentence. Each participant listened to 20 sentences from each of the three spatial locations in two background languages. The presentation order of these sentences (i.e., spatial location and background language) was completely randomized for each subject. The channels of the target-background combinations, which determine the difficulty of listening conditions, were manipulated between subjects. All others were within-subject factors. All responses were scored by an experimenter, blind to condition, based on three to four key words per sentence according to the key words specified in the Bamford-Kowal-Bench sentence lists (Bench et al., 1979). Responses received partial scores for correctly reported key words, including homophones.

Results and Discussion

Data from all subjects ($N = 26$) were included in the final analyses. Figure 1 depicts intelligibility scores (percent key words correctly reported) by location, masker language, and channel. The left and right panels depict performance in the 6- and 12-channel conditions, respectively.

To statistically evaluate these patterns, the intelligibility scores were transformed into logit values (see Warton & Hui, 2011 for preferring this over arcsine transforms) to normalize error variance before being submitted to a 2 (channel: 6 and 12) $\times$ 2 (masker language) $\times$ 3 (location) mixed analysis of variance (ANOVA). There was a main effect of channel, $F(1, 23) = 235.36$, $p < 0.001$, $\eta_p^2 = .911$, confirming that listeners' intelligibility improved when they were presented with sentences processed with a higher number of spectral channels. In addition, the main effect of masker language, $F(1, 23) = 46.61$, $p < 0.001$, $\eta_p^2 = .67$, was significant, but that of location, $F(2, 46) = 1.61$, $p = 0.21$, $\eta_p^2 = .07$, was not, indicating that listeners were able to benefit from LRM but not from SRM. The Channel $\times$ Language, $F(1, 23) = 0.53$, $p = 0.821$, $\eta_p^2 = .002$, and Channel $\times$ Location, $F(2, 46) = 2.42$, $p = 0.100$, $\eta_p^2 = .10$, interactions were not significant. However, there was a significant Masker Language $\times$ Location, $F(2, 46) = 26.14$, $p < 0.001$, $\eta_p^2 = .53$, interaction observed. Finally, there was a marginally significant Channel $\times$ Location $\times$ Language, $F(1, 23) = 3.07$, $p = 0.056$, $\eta_p^2 = .12$, interaction indicating

Figure 1. Sentence recognition performance (percent correct) for 26 listeners tested in signal-to-noise ratio (SNR) = 25 dB in the presence of two-talker English (white) and two-talker Dutch (gray) maskers in the 6-channel (left) and 12-channel (right) noise-vocoded conditions. Each open circle represents the score of a single participant. Boxplots in the two masking conditions (English and Dutch) are plotted for each spatial location condition (colocated, separated).

that the dependence of the effect of language on location was especially pronounced in the 12-channel condition.

To explore the Language $\times$ Location interaction further, we performed an additional post hoc analysis. First, we calculated an average intelligibility score in the spatially separated conditions by averaging the intelligibility scores observed in the -90° and $+90^\circ$ spatial configurations for each individual subject and for each of the two masker languages. Following this, we calculated the LRM scores by subtracting the intelligibility in English from the Dutch conditions in both spatial conditions for each individual subject (see Figure 2). We submitted these LRM difference scores to a 2 (spatial location: colocated and separated) $\times$ 2 (channel: 6 and 12) mixed ANOVA. Figure 2 depicts the average LRM scores observed for the colocated and the spatially separated locations. This analysis confirmed that there was a more pronounced LRM benefit in the spatially separated (average = 7.96%) than in the colocated condition (average = 3.12%), $p = .009$. We did not detect differences in the overall amounts of LRM in the two different channel conditions ($p = .134$).

Experiment 2

Experiment 1 demonstrated that listeners benefit from LRM but do not exhibit SRM when the noise-vocoded stimuli were presented at a 25-dB SNR. The release from

masking attributed to spatial separation between the competing streams did not depend on the spectral resolution of the noise-vocoded speech. In Experiment 2, in order to test whether SRM effects can be obtained at a lower SNR as per Freyman et al. (2008) and to evaluate whether LRM effects are replicated in more difficult conditions, we repeated Experiment 1 using an SNR of 15 dB instead. To prevent the possibility of floor effects in performance, we only used 12-channel noise-vocoded stimuli.

Method

Participants

Twelve undergraduate students from the University of Kansas ranging between 18 and 21 years old participated in these experiments (median age was 20 years). All listeners were native speakers of American English with no knowledge of Dutch, reported no history of speech or hearing disorders, and received course credit. None had participated in Experiment 1.

Materials

The materials from Experiment 1 were reused, except that the target and the background were combined to produce an SNR of 15 dB by normalizing the root-mean-square level of the original background masker signals to 50 dB SPL.

Figure 2. The individual benefit due to linguistic release from masking plotted in percentage points for all the listeners tested in each spatial location (unfilled bars indicating colocated and hashed indicating separated) in the (a) signal-to-noise ratio (SNR) = 25 and 12-channel, (b) SNR = 25 and 6-channel, and (c) SNR = 15 and 12-channel conditions. Each open circle represents the benefit estimated for every individual. RMS = root-mean-square.

Procedure

A procedure identical to the experimental procedure in Experiment 1 was used.

Results and Discussion

Data from all subjects ($N = 12$) were included in the final analyses. Figure 3 depicts intelligibility scores (percent key words correctly reported) by location and masker language. To statistically evaluate these observed patterns, logit-transformed intelligibility scores were submitted to a 2 (masker language) $\times$ 3 (location) within-subject ANOVA. The main effect of masker language, $F(1, 11) = 16.57$, $p = 0.002$, $\eta_p^2 = .60$, was significant, indicating that listeners again benefited from LRM. Unlike Experiment 1, the main effect of location, $F(2, 11) = 17.86$, $p < 0.001$, $\eta_p^2 = .62$, was also significant, indicating that listeners also benefited from SRM. Finally, we did not find evidence that SRM and LRM modulated each other because the Location $\times$ Language interaction, $F(2, 22) = 2.47$, $p = 0.11$, $\eta_p^2 = .18$, was not significant.

Experiment 2 revealed two key findings. First, the LRM effect is replicated in the lower SNR condition, suggesting that this benefit may be available in a range of listening situations with noise-vocoded speech. Second, unlike Experiment 1, we detected a pronounced SRM effect. This finding adds further support to the finding by Freyman et al. (2008) that SRM is likely to be detected when noise-vocoded listening is tested under appropriately difficult

SNR conditions. However, further testing is required to determine whether the SRM results from the perceived differences in spatial location or is due to having access to other binaural spatial benefits such as head-shadow (better-ear listening) and binaural squelch.

General Discussion

To summarize, we presented the first report of LRM in listeners tested with noise-vocoded target-masker combinations. Our results demonstrate that even under spectrally degraded conditions, there are considerable differences in the masking produced by the native and foreign language maskers. A benefit due to LRM was obtained under both channel conditions that were tested in Experiment 1 and a lower SNR tested in Experiment 2. Furthermore, similar to the findings reported in Freyman et al. (2008), while perceiving noise-vocoded speech, there was no overall benefit under experimental conditions in which the target was spatially separated from the masker location in Experiment 1 for either channel conditions likely due to the relatively high (positive) SNR that was chosen. Interestingly, our follow-up analyses indicated that there was a benefit when both language and the location of the background were mismatched with those of the target (Viswanathan et al., 2016). In Experiment 2 at a lower but positive SNR, we found evidence of both SRM and LRM. This would seem to suggest that in more difficult settings, listeners with normal hearing can reliably benefit from both spectro-temporal mismatches

Figure 3. Sentence recognition performance (percent correct) for 12 listeners tested in signal-to-noise ratio (SNR) = 15 dB in the presence of two-talker English (white) and two-talker Dutch (gray) maskers in the 12-channel noise-vocoded condition. Each open circle shows the score of a single participant. Boxplots in the two masking conditions (English and Dutch) are plotted for each spatial location condition (colocated, separated).

and the spatial dissimilarity between the target and the masker.

Further research is required to ascertain under what experimental conditions listeners experience LRM and SRM benefits. In particular, it would be pertinent to test whether LRM is observed in listeners with CIs and whether spatially separating the target from the masker can modulate the overall benefit similar to our findings reported in Experiment 1. Because CI users are only stimulated through a relatively small number of spectral channels, it remains to be seen whether they could benefit from LRM when perceiving speech in the presence of foreign competing background speech. For patients with a severe-to-profound hearing loss who rely on CIs to understand speech, our findings could have important clinical implications for the development of new aural rehabilitation strategies. For instance, it may be possible to improve speech-in-speech recognition performance in CI listeners through the development of tailored training materials that can be adjusted to modulate the difficulty of listening by altering the linguistic similarity of the competing background.

A notable limitation of the current study is the reliance on intelligibility measures, which limits the range of SNRs that can be used to evaluate these benefits while also avoiding floor or ceiling effects. To remedy this issue, we suggest that future investigation could utilize stimulus response thresholds in order to determine whether and how SRM and LRM modulate each other. Nevertheless, the current findings set the stage for a systematic exploration of LRM, its relationship with SRM, as well as the relative contributions of energetic and informational masking in spectrally impoverished listening conditions.

Acknowledgments

This research was supported by National Science Foundation Grant BCS-1431105, awarded to Navin Viswanathan, and by National Institute on Deafness and Other Communication Disorders Grant DC000052 T32, awarded to Brittany T. Williams under a predoctoral traineeship.

References

- Bench, J., Kowal, Å., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, 13, 108–112.
- Best, V., Marrone, N., Mason, C. R., & Kidd, G., Jr. (2012). The influence of non-spatial factors on measures of spatial release from masking. *The Journal of the Acoustical Society of America*, 131(4), 3103–3110.
- Bronkhorst, A. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acoustica*, 86, 117–128.
- Bronkhorst, A. W. (2015). The cocktail-party problem revisited: Early processing and selection of multi-talker speech. *Attention, Perception, & Psychophysics*, 77(5), 1465–1487.
- Brouwer, S., Van Engen, K. J., Calandruccio, L., & Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *The Journal of the Acoustical Society of America*, 131, 1449–1464.
- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, 109, 1101–1109.
- Calandruccio, L., Brouwer, S., Van Engen, K. J., Dhar, S., & Bradlow, A. R. (2013). Masking release due to linguistic and phonetic dissimilarity between the target and masker speech. *American Journal of Audiology*, 22, 157–164.
- Calandruccio, L., Dhar, S., & Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *The Journal of the Acoustical Society of America*, 128, 860–869.
- Calandruccio, L., Leibold, L. J., & Buss, E. (2016). Linguistic masking release in school-age children and adults. *American Journal of Audiology*, 25(1), 34–40.
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *The Journal of the Acoustical Society of America*, 25, 975–979.
- Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, 123, 414–427.

- Dorman, M. F., Loizou, P. C., Fitzke, J., & Tu, Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels. *The Journal of the Acoustical Society of America*, 104, 3583–3585.
- Dorman, M. F., Loizou, P. C., & Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *The Journal of the Acoustical Society of America*, 102, 2403–2411.
- Freyman, R., Helfer, K., McCall, D., & Clifton, R. (1999). The role of perceived spatial separation in the unmasking of speech. *The Journal of the Acoustical Society of America*, 106, 3578–3587.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 109, 2112–2122.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2008). Spatial release from masking with noise-vocoded speech. *The Journal of the Acoustical Society of America*, 124, 1627–1637.
- Fu, Q. J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology*, 6, 19–27.
- Garadat, S. N., Litovsky, R. Y., Yu, G., & Zeng, F. G. (2009). Role of binaural hearing in speech intelligibility and spatial release from masking using vocoded speech. *The Journal of the Acoustical Society of America*, 126, 2522–2535.
- Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47, 103–138.
- Hawley, M. L., Litovsky, R. Y., & Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *The Journal of the Acoustical Society of America*, 115, 833–843.
- Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., & Kollmeier, B. (2009). Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. *EURASIP Journal on Advances in Signal Processing*, 2009, 1–10.
- Loizou, P. C., Hu, Y., Litovsky, P., Yu, G., Peters, R., Lake, J., & Roland, P. (2009). Speech recognition by bilateral cochlear implant users in a cocktail-party setting. *The Journal of the Acoustical Society of America*, 125, 372–383.
- Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, 59, 203–243.
- Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America*, 22, 167–173.
- Nye, P. W., & Gaitenby, J. H. (1974). The intelligibility of synthetic monosyllabic words in short, syntactically normal sentences. *Haskins Laboratories status report on speech research*, 37(38), 169–190.
- Plomp, R., & Mimpen, A. M. (1981). Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech reception threshold for sentences. *Acustica*, 48, 325–328.
- Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *The Journal of the Acoustical Society of America*, 114, 446–454.
- Shannon, R., Zeng, F., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303–304.
- Stickney, G., Zeng, F.-G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *The Journal of the Acoustical Society of America*, 116, 1081–1091.
- Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., & Kidd, G. (2013). Spatial release from masking for noise-vocoded speech. *Proceedings of Meetings on Acoustics*, 19, 1–8.
- Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, 121(1), 519–526.
- Viswanathan, N., Kokkinakis, K., & Williams, B. T. (2016). Spatially separating foreign language masker from target results in spatial and linguistic masking release. *The Journal of the Acoustical Society of America*, 140, 465–470.
- Warton, D. I., & Hui, F. K. (2011). The arcsine is asinine: The analysis of proportions in ecology. *Ecology*, 92(1), 3–10.