

## Binaural sensitivity and release from speech-on-speech masking in listeners with and without hearing loss

Lucas S. Baltzell,<sup>1,a)</sup> Jayaganesh Swaminathan,<sup>1,b)</sup> Adrian Y. Cho,<sup>1,c)</sup> Mathieu Lavandier,<sup>2</sup> and Virginia Best<sup>1,d)</sup>

<sup>1</sup>*Department of Speech, Language, and Hearing Sciences, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215, USA*

<sup>2</sup>*University of Lyon, ENTPE, Laboratoire Génie Civil et Bâtiment, Rue Maurice Audin, F-69518 Vaulx-en-Velin Cedex, France*

### ABSTRACT:

Listeners with sensorineural hearing loss routinely experience less spatial release from masking (SRM) in speech mixtures than listeners with normal hearing. Hearing-impaired listeners have also been shown to have degraded temporal fine structure (TFS) sensitivity, a consequence of which is degraded access to interaural time differences (ITDs) contained in the TFS. Since these “binaural TFS” cues are critical for spatial hearing, it has been hypothesized that degraded binaural TFS sensitivity accounts for the limited SRM experienced by hearing-impaired listeners. In this study, speech stimuli were noise-vocoded using carriers that were systematically decorrelated across the left and right ears, thus simulating degraded binaural TFS sensitivity. Both (1) ITD sensitivity in quiet and (2) SRM in speech mixtures spatialized using ITDs (or binaural release from masking; BRM) were measured as a function of TFS interaural decorrelation in young normal-hearing and hearing-impaired listeners. This allowed for the examination of the relationship between ITD sensitivity and BRM over a wide range of ITD thresholds. This paper found that, for a given ITD sensitivity, hearing-impaired listeners experienced less BRM than normal-hearing listeners, suggesting that binaural TFS sensitivity can account for only a modest portion of the BRM deficit in hearing-impaired listeners. However, substantial individual variability was observed. © 2020 Acoustical Society of America.

<https://doi.org/10.1121/10.0000812>

(Received 30 July 2019; revised 7 February 2020; accepted 11 February 2020; published online 6 March 2020)

[Editor: Mathias Dietz]

Pages: 1546–1561

### I. INTRODUCTION

Listeners are often tasked with attending to a particular target sound source while ignoring competing sound sources. When these sound sources consist of speech, this task is commonly referred to as the “cocktail party” problem (Cherry, 1953). In the context of the cocktail party problem, spatial release from masking (SRM) refers to an improvement in intelligibility when competing talkers are spatially separated relative to when they are co-located with the target talker. While normal hearing (NH) listeners typically experience significant SRM (ranging from a few dB to ~25 dB; e.g., Bronkhorst and Plomp, 1992; Swaminathan *et al.*, 2016), hearing impaired (HI) listeners often experience less SRM under the same listening conditions, even when the effects of age are taken into account (e.g., Marrone *et al.*, 2008; Gallun *et al.*, 2013; Best *et al.*, 2012).

SRM is thought to arise from both monaural and binaural advantages conferred by spatially separating the target talker from competing talkers (Zurek, 1993; Hawley *et al.*, 2004). Monaural advantages arise in configurations where the head casts an acoustic “shadow” that attenuates the

competing talkers and thus improves the signal-to-noise ratio (SNR) in one ear (the “better ear”). Binaural advantages come from the processing of interaural time differences (ITDs), which improve the detectability of the target through binaural unmasking when target and masker energy are present simultaneously. In addition, ITDs (and interaural level differences; ILDs) provide the perception of spatial position to support source segregation and selective attention (Gallun *et al.*, 2005; Kidd and Colburn, 2017). For speech maskers, this can provide a release from masking greater than can be accounted for by improvements in effective SNR provided by better-ear listening and binaural unmasking (Freyman *et al.*, 1999).

There has been a significant amount of debate about the relative contribution of these different advantages to SRM (e.g., Hawley *et al.*, 1999; Brungart and Iyer, 2012; Schoenmaker *et al.*, 2016; Culling *et al.*, 2004; Kidd *et al.*, 2010; Ihlefeld and Litovsky, 2012; Glyde *et al.*, 2013) and the answer appears to depend strongly on the nature of the task. For example, for stimulus configurations where a speech target is masked by competing speech and the primary challenge is to segregate the talkers (i.e., when “informational masking” is high), binaural unmasking seems to contribute relatively little to SRM (Schoenmaker *et al.*, 2016). Indeed, while a number of binaural speech intelligibility models are available that can predict SRM for speech in noise, these models generally fail to predict the large SRM that is commonly observed when a speech target

<sup>a)</sup>Electronic mail: lbaltzel@bu.edu

<sup>b)</sup>Also at: University of the Pacific, 155 5th Street, San Francisco, CA 94103, USA.

<sup>c)</sup>Also at: Speech and Hearing Biosciences and Technology Program, Harvard University, Cambridge, MA 02138, USA.

<sup>d)</sup>ORCID: 0000-0002-5535-5736.

is masked by competing speech (see [Appendix B](#)). In these cases, it appears that the differences in perceived spatial position afforded by binaural cues are what dominate SRM (see [Sec. IVC](#)). The fact that HI listeners show reduced SRM in multi-talker listening environments leads naturally to the hypothesis that degraded binaural sensitivity may be responsible.

Across a range of monaural and binaural tasks, HI listeners have been shown to have degraded temporal sensitivity (e.g., [Strelyck and Dau, 2009](#); [Hopkins and Moore, 2011](#); [Gallun et al., 2014](#)). Variations in the acoustic waveform over time are encoded by auditory nerve (AN) fibers that phase-lock to these variations and temporal sensitivity depends on the fidelity of this phase-locking. However, a distinction is typically made between the temporal fine structure (TFS) and temporal envelope (ENV) components of the neural code. The TFS corresponds to the rapid variations in the acoustic waveform, while the ENV corresponds to the slow changes in amplitude over time ([Swaminathan and Heinz, 2012](#)). In the context of speech perception, HI listeners seem to be as proficient as NH listeners at making use of ENV information but less proficient at making use of TFS information, consistent with a loss of sensitivity to fine timing ([Lorenzi et al., 2006](#); [Moore, 2008](#); though see [Swaminathan et al., 2014](#)). In the context of binaural hearing, the most salient ITD cues are carried by the TFS, and degraded temporal sensitivity should impair the ability to utilize ITD cues. Indeed, many HI listeners are relatively insensitive to fine-structure ITDs as measured using pure tones ([Füllgrabe and Moore, 2018](#); [Best and Swaminathan, 2019](#)). It stands to reason that degraded TFS sensitivity may limit the ability of HI listeners to experience SRM by degrading access to ITD cues, though direct evidence for this hypothesis is lacking.

A number of findings from NH listeners suggest that TFS sensitivity supports SRM in speech mixtures. [Ruggles et al. \(2011\)](#) reported large individual differences in SRM for NH listeners and found that these differences were correlated with monaural TFS sensitivity. [Swaminathan et al. \(2016\)](#) tested the hypothesis that binaural TFS supports SRM by decorrelating the TFS across the two ears and measuring the effect on SRM. Speech stimuli were noise-vocoded, and the noise carriers across the two ears were correlated (same noise token) or uncorrelated (independent noise tokens). They found that SRM was significantly reduced (but not eliminated) when the TFS was decorrelated across ears.

[Drennan et al. \(2007\)](#) found that the ability of NH listeners to lateralize vocoded word tokens depended on the degree of TFS interaural correlation, as did their ability to experience binaural release from masking (BRM). As a brief aside, we use the term BRM here rather than SRM since stimuli were spatialized with ITDs rather than a full set of spatial cues (as would be experienced in free-field listening, see [Sec. IVC](#) for further discussion). The word tokens were filtered into six bands, and independent noise was systematically introduced to the TFS in the left and right ears in each

band in order to interaurally decorrelate the TFS. Lateralization accuracy monotonically decreased with interaural decorrelation, as did BRM, suggesting not only that binaural TFS is critical for lateralization, but that it supports BRM. In that study, though, monaural TFS cues were systematically degraded along with binaural TFS cues as noise was added to the TFS, making it difficult to tease apart the effects of noise vocoding and the effect of interaural decorrelation on BRM. This is because as the target and maskers become less distinct, the benefit of perceived location may be attenuated.

There is rather limited evidence showing that reduced SRM in HI listeners is a direct result of poor TFS sensitivity. In a group of HI listeners, [Strelyck and Dau \(2009\)](#) found that some (but not all) measures of binaural TFS sensitivity were significantly correlated with speech reception thresholds (SRTs) for spatially separated speech and noise maskers. [Neher et al. \(2012\)](#) found a relationship between a measure of TFS and SRTs in a spatially separated speech-on-speech task, although the relationship was not significant when age was accounted for. [Lőcsei et al. \(2016\)](#) measured BRM for speech in noise or babble in NH and HI listeners and found no relationship to binaural TFS sensitivity as measured via ITD discrimination for 250 Hz tones. Similarly, [King et al. \(2017\)](#) found no relationship between SRM for a speech-on-speech task and ITD thresholds at 500 Hz. On the other hand, [Papesh et al. \(2017\)](#) argued that a physiological measure of binaural sensitivity (based on auditory evoked potentials) was a better predictor of SRM with competing talkers than age and/or hearing loss. Thus, the relationship between hearing loss, binaural TFS sensitivity, and SRM warrants further investigation. Furthermore, if it is the case that poor binaural TFS sensitivity disrupts SRM in HI listeners, it is important to know whether this disruption accounts for some or all of the observed SRM deficits. Indeed, a number of studies have suggested that *non-spatial* factors (e.g., audibility, spectro-temporal resolution, cognitive capacity) significantly contribute to the reduced SRM observed in HI listeners ([Neher et al., 2009](#); [Best et al., 2011](#); [Best et al., 2012](#); [Best et al., 2013](#); [Best et al., 2017](#); [Rana and Buchholz, 2018](#); [Kidd et al., 2019](#)).

The goal of the present study was to re-examine the hypothesis that poor sensitivity to binaural TFS in HI listeners drives (or contributes to) their reduced BRM in a speech-on-speech task. To do this, we took the approach of simulating a loss of binaural TFS sensitivity in young NH listeners and comparing performance to that of a group of young HI listeners. Following [Swaminathan et al. \(2016\)](#), we decorrelated the TFS across the two ears to simulate loss of binaural TFS sensitivity. In order to simulate a range of binaural TFS sensitivity, the degree of TFS interaural correlation was systematically varied from fully correlated to fully uncorrelated. Using speech processed in this way, we measured both ITD sensitivity and BRM as a function of TFS interaural correlation. The results in NH listeners enabled us to establish a relationship between performance on these two tasks, as access to binaural TFS was

systematically varied in an otherwise healthy auditory system. Then, by observing how closely the results for individual HI listeners follow this relationship, we were able to estimate to what extent declines in binaural TFS sensitivity can explain their declines in BRM.

We found that while both ITD thresholds and BRM depended on TFS interaural correlation, HI listeners experienced less BRM than NH listeners for a given ITD sensitivity, suggesting that binaural TFS sensitivity accounts for only a modest portion of their BRM deficit. However, we observed large individual variability to the extent in which binaural TFS sensitivity accounts for BRM in our sample of HI listeners.

## II. METHODS

### A. Participants

Eleven NH listeners (five female) between the ages of 19 to 30 years (mean age = 23 years) participated in this study (pure-tone audiometric thresholds  $\leq 20$  dB from 250 Hz to 8 kHz). All NH listeners were fluent in English and were native English speakers. Nine HI listeners (four female) between the ages of 21 to 44 years (mean age = 28 years) also participated. Audiograms of these listeners averaged over left and right ears, are shown in Fig. 1 [see Supplemental Fig. 1(B) for left-ear and right-ear audiograms].<sup>1</sup> Listeners with asymmetric hearing loss were not recruited for this study, where asymmetry referred to a difference in pure tone average (PTA) across ears that exceeded 10 dB for low frequencies (PTA<sub>low</sub>: average of thresholds from 250 to 1000 Hz) and/or exceeded 15 dB for high frequencies (PTA<sub>high</sub>: average of thresholds from 2000 to 8000 Hz). All HI listeners were fluent in English, although one learned English as a second language. Experiments were conducted at Boston University, and all procedures were reviewed and approved by the Institutional Review Board. All participants provided informed consent prior to testing. All listeners were recruited from the Boston area and were, for the most

part, college students or recent graduates. Many of these listeners (both NH and HI) had previously participated in psychoacoustic experiments in the lab.

### B. Stimuli

Speech stimuli were drawn from a corpus of monosyllabic words recorded by Sensimetrics Corporation (Malden, MA) at a sampling rate of 50 kHz (see Kidd *et al.*, 2008). The corpus contained recordings from multiple male and female talkers and was designed so that individual words could be combined to form sentences according to the structure <name> <verb> <number> <adjective> <object>. There were eight possible words in each category. In the BRM task, sentences were constructed by randomly drawing a word from each category for a given talker. Only female talkers were used, and talkers were drawn randomly on each trial.

### C. Stimulus processing

#### 1. Vocoder

Following Swaminathan *et al.* (2016), the TFS interaural correlation was manipulated by vocoding the speech stimulus using different noise carriers for the left and right channels. To create the vocoded speech stimulus, a pair of pink noise signals (Noise<sub>1</sub> and Noise<sub>2</sub>) were independently generated, and were combined according to the symmetric-generator method described by Hartmann and Cho (2011) for a desired carrier interaural correlation  $rc$ ,

$$\text{Carrier}_{\text{LEFT}} = \alpha \text{Noise}_1 + \beta \text{Noise}_2, \quad (1)$$

$$\text{Carrier}_{\text{RIGHT}} = \alpha \text{Noise}_1 - \beta \text{Noise}_2, \quad (2)$$

$$\text{where } \alpha = \sqrt{\frac{rc + 1}{2}} \text{ and } \beta = \sqrt{1 - \alpha^2}.$$

Hartmann and Cho (2011) demonstrated that this method of defining two noise signals with a desired correlation results in relatively minimal variability in correlation across pairs of noise signals (in this case, Carrier<sub>LEFT</sub> and Carrier<sub>RIGHT</sub>).

The speech stimulus was passed through a bank of 32 band-pass analysis filters with equal bandwidth on a logarithmic frequency scale spanning 80–8000 Hz. The band-pass filters were created using the auditory chimera package described in Smith *et al.* (2002). The analytic signal was first generated for each band using the Hilbert transform. For each band, the ENV was extracted as the magnitude of the Hilbert analytic signal, followed by low-pass filtering below 150 Hz with a fourth-order Butterworth filter. The broadband noise carriers were passed through the same bank of analysis filters. The TFS for each band was extracted as the cosine of the phase of the Hilbert analytic signal. The TFS was multiplied by the ENV for each band. These products were once again passed through the bank of analysis filters in order to remove any spectral splatter and summed across bands to create the vocoded speech stimulus.

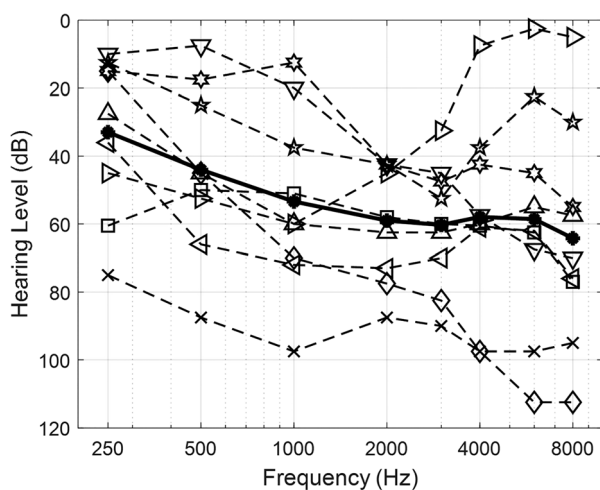


FIG. 1. Audiometric thresholds for HI listeners. Threshold curves of individual listeners (averaged across ears) are indicated by dotted lines, and the mean is indicated by a solid line. Unique symbols are used for each listener, and these symbols refer to the same listener in all subsequent figures.



2. TFS interaural correlation at output of vocoder

Five different carrier interaural correlation ( $rc$ ) values were chosen: 1, 0.75, 0.5, 0.25, and 0. However, to account for the effects of vocoding, and for the fact that the carrier interaural correlation reflected both ENV and TFS components, we wanted to quantify the TFS interaural correlation at the *output* of the vocoder. Shown in Fig. 2(A), we applied a Hilbert transform to the output of the vocoder. As above, the TFS was extracted as the angle of the Hilbert analytic signal, and the ENV was extracted as the magnitude of the Hilbert analytic signal, low-pass filtered below 150 Hz (fourth-order Butterworth filter). To identify samples where there was little to no energy in the stimulus, a threshold  $\xi$  was defined as 5% of the root-mean-square (rms) of the ENV (Goupell and Hartmann, 2007). Since the TFS is not well defined when the ENV is at or near zero, all samples where the ENV did not equal or exceed this threshold were removed from the TFS. This operation mimics the neural coding of TFS, since phase-locking only occurs when sufficient energy is present in the stimulus. The correlation between the left and right channels of the resulting TFS was computed to yield the TFS interaural correlation at the output of the vocoder. This procedure was repeated for all experimental stimuli at each carrier interaural correlation  $rc$ . The five carrier interaural correlation values are shown in the left column of Fig. 2(B), and the corresponding TFS interaural correlations at the output of the vocoder ( $r$ ) are shown in the right column. The interaural correlations in the right column were assumed to be the true TFS interaural correlation of the acoustic stimuli presented to the listeners.

3. TFS interaural correlation at output of AN model

Since the relationship between “acoustic” TFS (extracted from the acoustic stimulus) and “neural” TFS

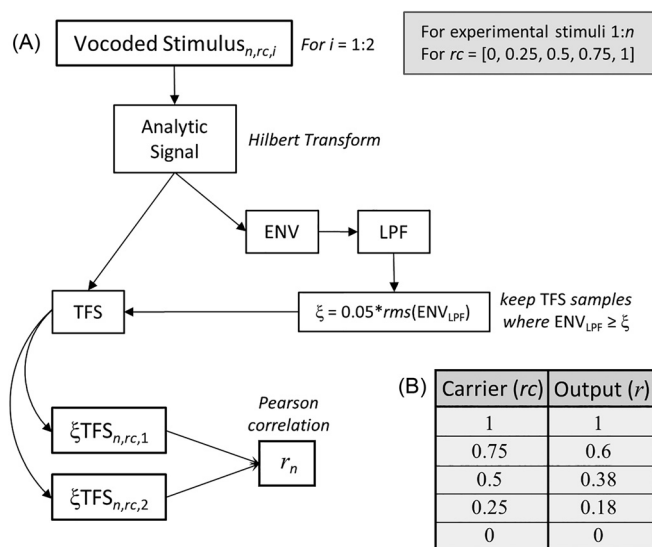


FIG. 2. (A) Schematic illustrating the calculation of TFS interaural correlation at the output of the vocoder. (B) The table in the bottom right shows the relationship between the interaural correlation of the noise carrier and the TFS interaural correlation of the vocoded stimulus.

(phase-locked activity in the AN) is not always obvious (Ghitza, 2001; Zeng et al., 2004; Shamma and Lorenzi, 2013), we quantified the interaural correlation of TFS and ENV components at the output of a phenomenological AN model. Specifically, we quantified the TFS interaural correlation preserved by the AN, as well as the ENV interaural distortion introduced by using different carriers in the two ears. We used the computational AN model described by Bruce et al. (2018) to quantify the similarity of neural TFS and ENV coding across the left and right ears (Heinz and Swaminathan, 2009).

The details of this method and our specific implementation are described in Appendix A. To summarize, the AN model was used to generate spike-train responses for individual AN fibers, and neural cross-correlation coefficients  $\rho_{TFS}$  and  $\rho_{ENV}$  were computed across the two ears for each fiber. The coefficients were normalized by the strength of coding for each fiber and combined across fibers to yield an integrated estimate of TFS and ENV interaural correlation.

The integrated neural coefficients for each carrier interaural correlation value  $rc$ , averaged over vocoded stimuli, are shown in Fig. 3. Also shown are the average acoustic interaural correlations over the same stimuli. We see that the neural TFS interaural correlations are similar to the acoustic TFS interaural correlations [Fig. 2(B), second column] across  $rc$  values. We also see that the neural interaural ENV correlations are similar to the acoustic interaural ENV correlations.

D. Procedure

After vocoding, stimuli were spatialized by taking an FFT of the signal, shifting the phase of each frequency

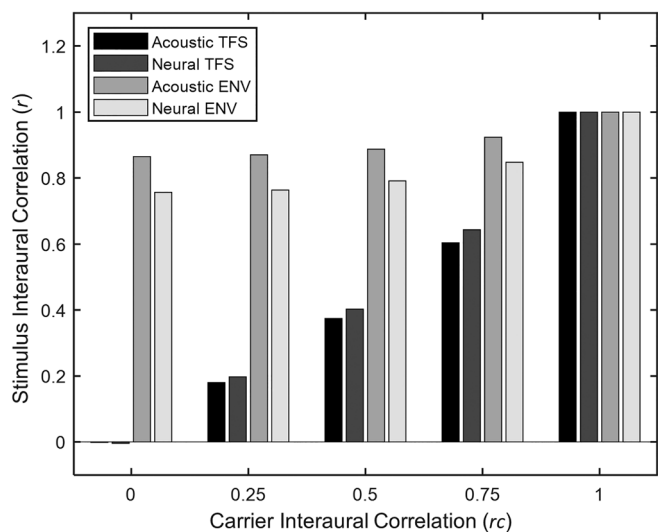


FIG. 3. Comparison of interaural correlations for TFS and ENV components of vocoded stimuli derived from the acoustic signal and from the output of an AN model (Bruce et al., 2018). The computation of the acoustic TFS interaural correlations are described in Fig. 2, and the acoustic ENV interaural correlations were computed from the low-pass filtered magnitude of the Hilbert analytic signal of the vocoded stimuli. The carrier interaural correlation refers to correlation of the noise carriers across ears at the input to the vocoder.

component corresponding to a desired ITD, and taking the IFFT of the resulting spectrum. Besides ITD cues, no other binaural cues were available to the listener, and stimuli were presumably perceived as inside the head. For HI listeners, a gain filter was computed based on the audiogram following the NAL-RP fitting algorithm (Dillon, 2012), and was applied to all stimuli subsequent to spatialization. Gain was based on the average audiogram across ears, and the same gain was applied to each ear. For one HI listener (×) the gain filter resulted in peak clipping at the maximum presentation level for the BRM task, and so the maximum presentation level of the target was adjusted (see Sec. IID 2).

Stimuli were presented via Sennheiser HD 280 headphones (Wedemark, Germany) to listeners seated in a double-walled sound-attenuating chamber (IAC Acoustics, North Aurora, IL). The digital signals were generated on a PC outside of the booth and then routed through an RME HDSP 9632 24-bit soundcard (Haimhausen, Germany). Stimulus presentation level was normalized based on headphone calibration to a broadband noise, and inverse filtering based on the headphone frequency response was not applied.

### 1. Experiment 1: ITD

A single word from the speech corpus (“two”) spoken by a single female talker was used to measure ITD thresholds in a 2-alternative forced choice (2AFC) lateralization task. This word token was chosen because it had been used in previous studies of spatial sensitivity in HI listeners (Best *et al.*, 2011; Best and Swaminathan, 2019). This word token had a duration of 459 ms and was fixed in level at 70 dB sound pressure level (SPL). On each trial, the word token was presented in two intervals, separated by a 500-ms inter-stimulus interval (ISI). In the first interval, the ITD was always 0  $\mu$ s, and served as a reference for the second interval. In the second interval, the ITD was either left-leading or right-leading, and the listener was instructed to indicate whether the word token in the second interval was presented from the left or right of midline. The ITD was adaptively varied according to a 2-down/1-up tracking procedure in  $\log_{10}$ (ITD) steps, with thresholds corresponding to approximately 71% correct on the psychometric function. ITDs were initially varied in step sizes of 0.2  $\log_{10}$  units and then in step sizes of 0.1  $\log_{10}$  units after the fourth reversal. Each track consisted of at least 20 trials and at least 12 reversals and began with an ITD of 500  $\mu$ s. An upper limit of 2 ms was defined such that if the value of the track equaled or exceeded this upper limit at least four times, the track was considered unreliable and no threshold was obtained. An ITD of 2 ms is well outside of the ecologically valid range for human listeners but is well below the echo suppression threshold at which the signals across the two ears are heard as separate sources (Yang and Grantham, 1997). Correct answer feedback was provided during testing.

ITD thresholds were obtained for each of the five vocoding conditions [Fig. 2(B)]. ITD thresholds were also

obtained for the “natural” recording of the word without vocoding. Each listener completed five blocks, where each block contained one track for each condition. The first block was considered practice, and conditions were presented in descending order of difficulty (natural, followed by  $r=1$ , followed by  $r=0.6$ , etc.). The first block was not included in the analysis. In each of the four experimental blocks, the order of conditions was randomized. A total of four tracks were run for each condition (one track per block).

Since unreliable tracks were obtained for a number of both HI and NH listeners in conditions with lower interaural correlation values [Supplemental Fig. 1(A)],<sup>1</sup> data were pooled across tracks and psychometric functions were fit for each listener in each condition (using the “psignifit” package). This was done so that threshold estimates were not biased towards estimates obtained for reliable tracks. Thresholds were obtained by extracting ITD values corresponding to 71% correct from the psychometric function, and if the psychometric function did not reach 71% correct, no threshold was obtained (unmeasured). Furthermore, in order to exclude threshold estimates that were too close to the upper limit of 2 ms, all thresholds exceeding 1.25 ms were treated as unmeasured.

### 2. Experiment 2: BRM

Performance in speech mixtures was assessed as follows, following Swaminathan *et al.* (2016). On each trial, the listener heard three sentences spoken by three different randomly chosen female talkers from eight available female talkers. Sentences were constructed by concatenating (without pause) the monosyllabic words recorded in isolation. All sentences were gated on simultaneously. One sentence was designated as the target and always contained the <name> “Sue” with the other words being randomly selected from the available choices (e.g., “Sue lost two red socks”). The two masker sentences contained randomly selected words from each category that differed from the target sentence and from each other. Listeners were instructed to identify all the words in the sentence spoken by the target talker (first word always “Sue”). The set of possible words was displayed graphically on the computer monitor, and subjects clicked on the words they perceived as constituting the sentence spoken by the target talker.

The target sentence was always presented with an ITD of 0  $\mu$ s, corresponding to a source at 0° azimuth. In “separated” conditions, the two masker sentences were presented with an ITD of  $\pm 690 \mu$ s (positive left-leading, negative right-leading), corresponding to sources at approximately  $\pm 90^\circ$  (Moore, 2004; Chap. 7). In the “co-located” condition, the two masker sentences were presented with an ITD of 0  $\mu$ s. The level of each masker sentence was fixed at 55 dB SPL, and the level of the target sentence was adaptively varied according to a 1-down/1-up tracking procedure, with thresholds corresponding to approximately 50% correct on the psychometric function. The level of the target was varied in 6 dB steps initially and then in 3 dB

steps after the third reversal. The maximum level of the target was set to 70 dB for all listeners except for one HI listener (×), for whom the maximum target level was set to 62 dB to avoid peak clipping. Correct answer feedback was provided during testing. Responses were counted as correct if the listener successfully identified at least three of the four words (excluding <name>).

Each listener completed four blocks, where each block contained one track for each condition. In each of these blocks, the order of conditions was randomized. Separated ( $\pm 690 \mu\text{s}$ ) thresholds were obtained for each of the five vocoded conditions. Co-located ( $\pm 0 \mu\text{s}$ ) thresholds were also obtained for the  $r = 1$  vocoded condition, though in the interest of time, co-located thresholds were not obtained for the other values of  $r$ . Prior to the four experimental blocks, listeners completed a practice track for the separated  $r = 1$  and  $r = 0$  conditions. Target to masker ratio (TMR) refers to the difference between the intensity of the target and the intensity of each masker on a dB scale. Across conditions, BRM was defined as the difference between the separated TMR at threshold and the co-located TMR at threshold for  $r = 1$ . Individual thresholds for each condition are based on the average of four tracks. For HI listeners, separated thresholds were also obtained for “natural” stimuli without vocoding.

III. RESULTS

A. ITD thresholds

Shown in Fig. 4, ITD thresholds systematically increased as a function of TFS interaural decorrelation for both NH (Fig. 4A) and HI (Fig. 4B) listeners. Unmeasured thresholds are indicated above the hash marks. Only listener × had unmeasured thresholds in all conditions and was excluded from further analysis.<sup>2</sup> For other listeners, unmeasured thresholds were treated as missing values in the statistical analysis. A linear mixed-effects model (“lmer” package, R) was fit to the data, treating condition ( $r = [0, 0.2, 0.4, 0.6, 1]$ ) and hearing status (NH vs HI) as fixed effects and listener as a random effect. Condition was treated as a continuous variable, and hearing status was treated as a categorical variable. The natural condition was not included in the model. ITD thresholds were  $\log_{10}$  transformed for statistical analysis, and effective degrees of freedom for the pooled sample variance were approximated using the Welch-Satterthwaite method. An analysis of variance (ANOVA) (“lmerTest” package, R) revealed a significant main effect of condition [ $F(1,56.85) = 514.01, p < 0.001$ ] on ITD threshold, but did not reveal a significant main effect of hearing status [ $F(1,17.97) = 1.99, p = 0.18$ ] or a significant interaction [ $F(1, 56.85) = 3.32, p = 0.07$ ]. While our decision to not include unmeasured thresholds may have contributed to our failure to find a significant effect of hearing status, arbitrarily assigning a value to these unmeasured thresholds would have complicated our interpretation of the statistical results. A separate paired t-test did not reveal a significant difference in ITD thresholds

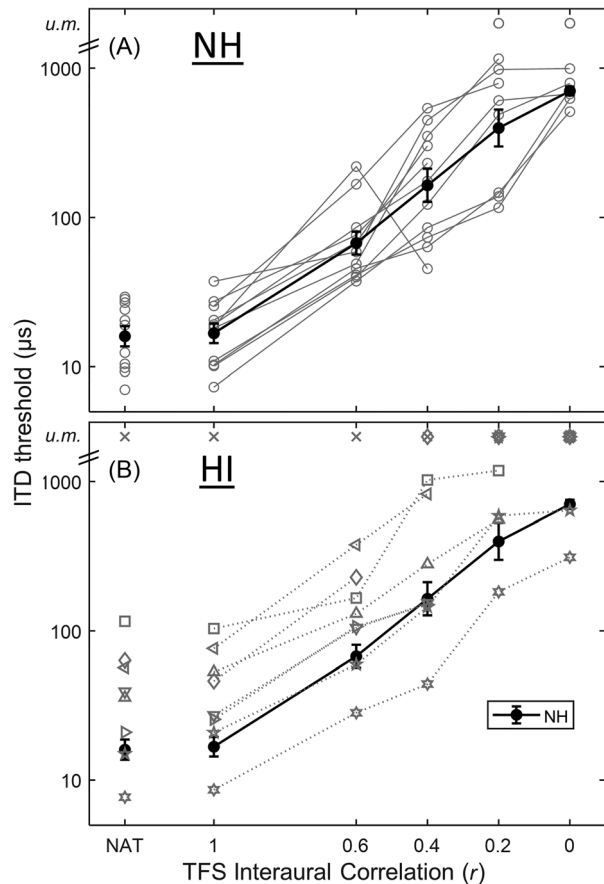


FIG. 4. ITD thresholds as a function of TFS interaural correlation for (A) NH listeners and (B) HI listeners. (A) Individual NH psychometric functions are displayed as solid gray lines, with the geometric mean (excluding unmeasured thresholds) displayed as a solid black line (standard error bars). (B) Individual HI psychometric functions are displayed as dotted gray lines with unique marker symbols, and *u.m.* refers to unmeasured thresholds (points above the hash mark). The NH mean is replotted for comparison. NAT refers to the natural condition (without vocoding).

between the natural and  $r = 1$  conditions for NH listeners ( $p = 0.72$ ) or for HI listeners ( $p = 0.87$ ).

B. Threshold TMRs

Shown in Fig. 5, TMRs at threshold for speech mixtures systematically increased as a function of TFS interaural decorrelation for both NH [Fig. 5(A)] and HI [Fig. 5(B)] listeners. A linear mixed-effects model treating condition and hearing status as fixed effects and listener as a random effect was fit to the data. The co-located condition was not included in the model. An ANOVA revealed a significant effect of condition [ $F(1,78) = 277.02, p < 0.001$ ] and hearing status [ $F(1,18) = 22.11, p < 0.001$ ] on threshold TMRs, as well as a significant interaction [ $F(1,78) = 52.83, p < 0.001$ ]. Because the interaction between condition and hearing status was significant, a separate mixed-effects model was fit using only the HI data (not including the natural condition), and a separate ANOVA was performed revealing a significant main effect of condition on threshold TMRs [ $F(1,35) = 35.97, p < 0.001$ ].

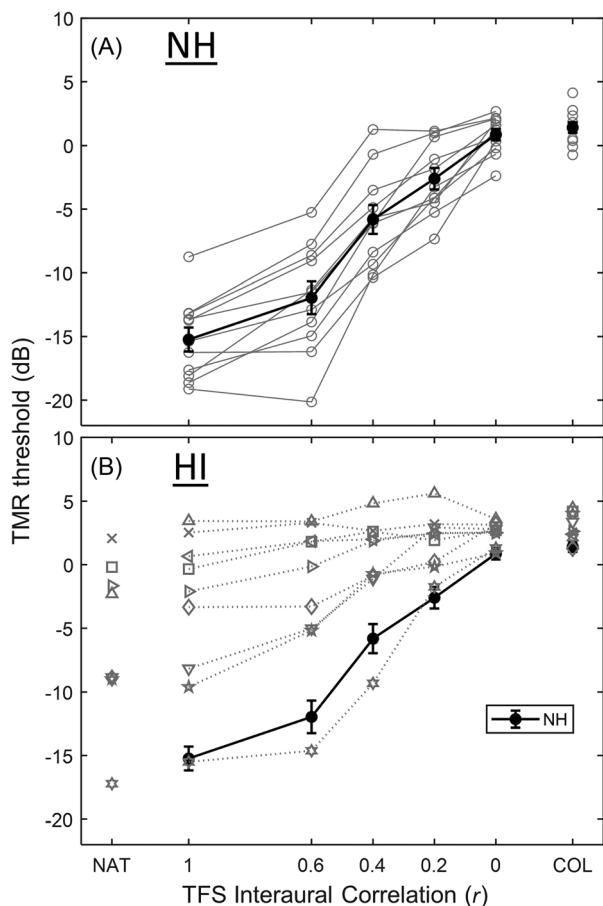


FIG. 5. TMRs at threshold as a function of TFS interaural correlation for (A) NH listeners and (B) HI listeners. (A) Individual NH psychometric functions are displayed as solid gray lines, with the mean displayed as a solid black line (standard error bars). (B) Individual HI psychometric functions are displayed as dotted gray lines with unique marker symbols (same as Fig. 4). The NH mean is replotted for comparison. COL refers to the co-located condition, and NAT refers to the natural condition (without vocoding).

A separate paired t-test did not reveal a significant difference in threshold TMRs between the  $r = 0$  and co-located conditions for NH listeners ( $p = 0.38$ ) or for HI listeners ( $p = 0.34$ ). A separate paired t-test also did not reveal a significant difference between the natural and the  $r = 1$  condition for HI listeners ( $p = 0.49$ ).

**C. Relationship between ITD sensitivity and BRM**

BRM was calculated for each correlation condition (by subtracting individual threshold TMRs from the co-located threshold in the  $r = 1$  condition for that listener). Figure 6(A) shows mean BRM for the NH group as a function of mean ITD thresholds for different values of interaural correlation (open symbols with error bars). We see that ITD thresholds and BRM jointly depend on  $r$  such that both ITD sensitivity and BRM get worse as  $r$  decreases. We summarize this relationship as a least squares linear fit through the NH group means (dashed line).

To the extent that TFS interaural decorrelation simulates loss of binaural TFS sensitivity, we can think of the

least squares fit through the NH data as a prediction for how performance on these tasks may be related in HI listeners. Specifically, if loss of binaural TFS sensitivity accounts for BRM deficits in HI listeners, we would expect the HI data to follow the NH fit, even if the range of performance was compressed. Put differently, if binaural TFS is responsible for a listener’s poor BRM, then the NH fit indicates how bad their ITD sensitivity should be. If loss of binaural TFS sensitivity does not account for the BRM deficits in HI listeners, we would not expect a relationship between ITD sensitivity and BRM in these listeners. BRM and ITD thresholds for HI listeners are shown in Fig. 6(A) as individual symbols. These points in general lie far from the NH fit, and thus do not follow the predicted relationship.

This point can be further illustrated by quantifying the prediction error for each HI listener in the  $r = 1$  condition, where the prediction error refers to the difference in BRM between the least squares fit through the NH data and the observed BRM of the HI listener, for a given ITD sensitivity. Consistent with previous reports, we see a range of ITD thresholds for HI listeners in the  $r = 1$  condition, from 8 to 107  $\mu\text{s}$ , suggesting that our HI listeners had a large range of binaural TFS sensitivity. Over this range however, all HI listeners experienced less BRM than NH listeners with similar ITD thresholds [Fig. 6(B)]. That being said, the large individual differences in prediction error (and the fact that these differences were non-monotonic over ITD sensitivity) suggest substantial individual differences in the amount of BRM deficit accounted for by binaural TFS sensitivity.

**D. Predicting BRM from ITD thresholds and audiogram**

The severity of each HI listener’s hearing loss (with the exception of listener  $\times$  who was excluded from this analysis) was quantified using  $\text{PTA}_{\text{low}}$  and  $\text{PTA}_{\text{high}}$  as defined in Sec. II A. We determined the degree to which BRM can be predicted by  $\text{PTA}_{\text{low,high}}$  and ITD sensitivity, specifically in the  $r = 1$  condition, in two ways (Fig. 7). First, the correlation coefficient was separately computed between BRM and ITD sensitivity, between BRM and  $\text{PTA}_{\text{low}}$ , and between BRM and  $\text{PTA}_{\text{high}}$ . These coefficients were squared yielding the percent of BRM variance explained, separately for ITD sensitivity ( $r^2 = 0.73$ ;  $p = 0.007$ ),  $\text{PTA}_{\text{low}}$  ( $r^2 = 0.78$ ;  $p = 0.004$ ), and  $\text{PTA}_{\text{high}}$  ( $r^2 = 0.12$ ;  $p = 0.4$ ). Second, a partial least squares (PLS) regression was fit with BRM as the dependent variable, and with ITD thresholds and  $\text{PTA}_{\text{low,high}}$  as independent variables. The PLS regression performs a singular value decomposition on the combined matrix of dependent and independent variables, identifying transformations that maximize the covariance between the variables. Because we had three independent variables, three PLS components were computed, but since the second and third PLS components explained only a negligible amount of the variance in the transformed dependent variable ( $<3\%$  and  $<1\%$ , respectively), only the first PLS component was considered. A correlation coefficient was computed between the transformed dependent variable and transformed combination of independent variables corresponding to



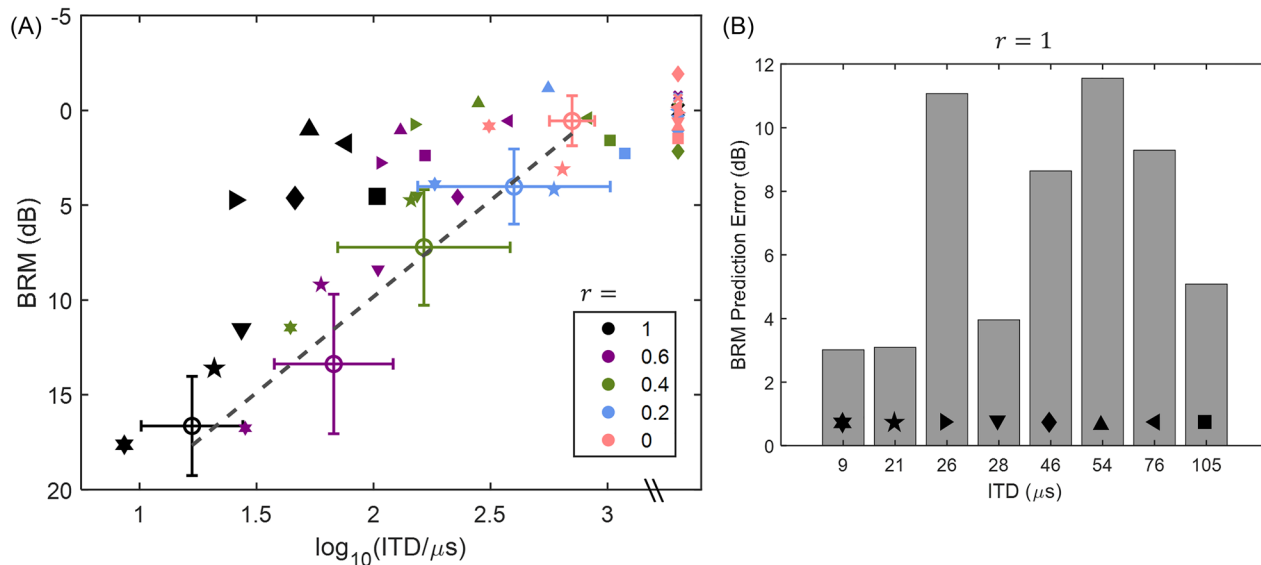


FIG. 6. (Color online) (A) BRM as a function of ITD threshold for different interaural TFS correlation values (legend). For NH listeners, the mean for each condition is indicated by an unfilled circle, with the standard deviation in each dimension indicated by an error bar, and the dashed line shows a linear fit through the means. Individual HI listeners are indicated by unique filled symbols (same as Figs. 3 and 4). Unmeasured ITD thresholds are shown to the right of the hash mark. (B) Difference in BRM between each HI listener in the  $r = 1$  condition (black symbols) and the linear fit to the NH data at the corresponding ITD threshold. Listener  $\times$ , who had an unmeasured threshold in the  $r = 1$  condition, was excluded from this display.

the first PLS component, and this correlation coefficient was squared to yield the combined variance explained ( $r^2 = 0.83$ ,  $p = 0.002$ ). By optimally combining independent variables (ITD,  $\text{PTA}_{\text{low}}$ , and  $\text{PTA}_{\text{high}}$ ), we can account for a greater variance in BRM than can be accounted for by any independent variable alone. However, this improvement was modest, suggesting that the ITD thresholds and  $\text{PTA}_{\text{low,high}}$  are similarly correlated with BRM.

For NH listeners, we computed the correlation coefficient between BRM and ITD thresholds in the  $r = 1$

condition, and this coefficient was squared to yield the variance explained ( $r^2 = 0.07$ ;  $p = 0.43$ ). The lack of significant correlation suggests that ITD thresholds were not predictive of BRM in our sample of NH listeners. Furthermore, neither  $\text{PTA}_{\text{low}}$  ( $r^2 = 0.12$ ;  $p = 0.29$ ) nor  $\text{PTA}_{\text{high}}$  ( $r^2 = 0.1$ ;  $p = 0.35$ ) were predictive of BRM in NH listeners.

#### IV. DISCUSSION

##### A. Summary and interpretation of results

The results of this study can be summarized by three main points. First, both ITD sensitivity as measured with speech and BRM systematically increase with TFS interaural correlation, for both NH and HI listeners. Second, for a given ITD sensitivity, HI listeners experience less BRM than NH listeners, suggesting that binaural TFS sensitivity cannot fully account for the difference in BRM between NH and HI listeners. Third, individual variability in the HI group suggests that the role of binaural TFS sensitivity in supporting BRM depends on the listener, and that BRM depends in part on the severity of that listener's hearing loss.

That TFS interaural correlation jointly supports ITD sensitivity and BRM in NH listeners is not surprising, and is consistent with the hypothesis that ITD cues in the TFS support BRM (Swaminathan *et al.*, 2016). That TFS interaural correlation supports ITD sensitivity in HI listeners is also not surprising, though the fact that it supports BRM in HI listeners is worthy of note. While studies have consistently found that HI listeners experience less SRM than NH listeners (e.g., Marrone *et al.*, 2008; Neher *et al.*, 2009; Glyde *et al.*, 2013c; Gallun *et al.*, 2013), HI listeners in these studies tended to experience *some* SRM. In previous studies, stimuli were either presented over loudspeakers or

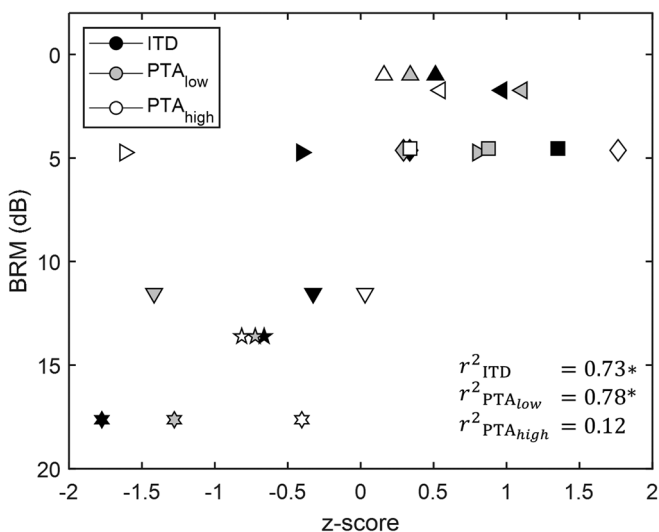


FIG. 7. BRM as a function of ITD threshold (black) and PTA (gray/white) for HI listeners in the  $r = 1$  condition. ITD thresholds and PTAs have been converted to z-scores. Individual listeners are indicated by unique symbols (same as previous figures). Listener  $\times$  was excluded from this analysis and is not shown. Variance explained ( $r^2$ ) for each dependent variable is shown and an asterisk indicates significance.



spatialized with head-related transfer function (HRTFs), and in the current study, they were spatialized by applying ITDs. Therefore, the fact that HI listeners experienced BRM in our study suggests that ITD cues alone are sufficient, for most HI listeners, to provide some spatial benefit.

When comparing the relationship between ITD sensitivity and BRM, we find that HI listeners experience less BRM than NH listeners for a given ITD threshold (Fig. 6). This result suggests that poor TFS sensitivity, reflected in poor ITD sensitivity, is not the primary factor limiting BRM in these listeners. This conclusion is in agreement with other studies demonstrating an influence of non-spatial factors on SRM, such as reduced audibility (Best *et al.*, 2017; Rana and Buchholz, 2018) and reduced cognitive capacity in the case of older HI listeners (e.g., Neher *et al.*, 2009). In the current study, cognitive effects were minimized by recruiting only younger listeners with hearing loss, who were well matched in age to our young NH group.

Several studies have shown that simulating reduced audibility of speech in NH listeners reduces SRM (e.g., Glyde *et al.*, 2015; Best *et al.*, 2017). In addition, increases in overall level can improve SRM in listeners with HI (Jakien *et al.*, 2017; Rana and Buchholz, 2018). Furthermore, it appears that SRM can be further increased in HI listeners by the provision of additional linear gain (above that prescribed by the NAL-RP formula also used in our study) or by the provision of non-linear gain (Glyde *et al.*, 2015; Rana and Buchholz, 2018). Interestingly, however, while  $PTA_{low}$  was highly correlated with BRM in the present study,  $PTA_{high}$  was not (see also Neher *et al.*, 2009; King *et al.*, 2017). Given that audibility was more likely to be a problem at higher frequencies where speech energy is low and the NAL-RP gain we provided does not provide full compensation, this result suggests that for our BRM task, audibility per se is not the issue, but rather access to low-frequency information.

There are a number of important differences between the ITD task and the BRM task though that are worth some consideration. First, there is not necessarily an equivalence between the ability to detect ITDs at threshold and the ability to use suprathreshold ITDs to segregate competing talkers. Our experimental design assumes that since both of these abilities rely on the binaural system, they will be equally susceptible to a loss of binaural TFS sensitivity. It is possible though that HI listeners are particularly impaired in their ability to make use of suprathreshold ITDs.

Second, there is not necessarily an equivalence between making use of spatial cues in quiet and in the presence of competing sounds. For instance, Best *et al.* (2011) found that while localization accuracy of speech in quiet was similar for NH and HI listeners, localization accuracy of speech in speech mixtures was much worse for HI listeners. Certainly, any loss of frequency resolution in HI listeners (e.g., Hopkins and Moore, 2011) that makes separating target energy from masker energy more difficult could also have binaural consequences. For example, ITD cues from both the target and the masker are more likely to be present

in the same channel, limiting the ability to extract valid ITD information from that channel. Indeed, one explanation of the current result is that reduced spectral resolution has little to no effect on ITD sensitivity in quiet, but a large effect when competing talkers are present. Third, while the ITD task presents listeners with static ITDs, the BRM task requires listeners to make use of binaural cues that fluctuate over time. With symmetric speech maskers (and a diotic target), the dominant ITD will fluctuate, and optimal performance may depend on following these sometimes rapid fluctuations. It is possible that individual differences in binaural integration windows, which determine how well these fluctuations can be followed, may contribute to individual differences in BRM. More specifically, if HI listeners have sub-optimal binaural integration windows (Hauth and Brand, 2018), this may help account for the deficit in BRM not accounted for by ITD sensitivity. We are unaware of any studies comparing binaural integration windows for NH and HI listeners, though Hu *et al.* (2017) found longer integration windows for cochlear implant users compared to NH listeners.

What can we conclude then about the source of the BRM deficit in HI listeners? For the majority of HI listeners in the present study, ITD sensitivity appeared to account for only a modest portion of their BRM deficit relative to NH listeners. Indeed, low frequency hearing loss was more predictive of BRM than ITD sensitivity. It is possible that reduced audibility of speech, and/or poorer spectro-temporal resolution, contributed to the BRM deficits observed. It is worth emphasizing that these conclusions are based on our relatively small and heterogeneous population of young HI listeners. It could be expected that sensitivity to ITDs would play a larger role for older listeners, given the well-established effects of aging on TFS processing (e.g., Füllgrabe and Moore, 2018). However, the lack of a clear relationship between ITD thresholds (for pure tones) and SRM in older HI listeners suggests this may not be the case (Neher *et al.*, 2012; Lócsei *et al.*, 2016; King *et al.*, 2017).

## B. Individual variability

We found clear individual differences in the patterns of performance for our sample of HI listeners. For listeners ★ and ★, SRM was well predicted by their ITD sensitivity, which is to say that these HI listeners resembled NH listeners with some amount of TFS interaural decorrelation. Thus, binaural TFS sensitivity seems to account for much of the observed BRM in these listeners. However, for the other HI listeners, BRM was consistently worse than predicted by their ITD sensitivity, suggesting that binaural TFS sensitivity does not fully account for their BRM deficit.

Furthermore, individual BRM prediction errors did not straightforwardly (monotonically) depend on ITD sensitivity [Fig. 6(B)]. For instance, while listener ► had a lower ITD threshold (26  $\mu$ s) than listener ▼ (28  $\mu$ s), the prediction error for listener ► was 11 dB, while for listener ▼, it was only 4 dB. On the other hand, while listener ▼ had a lower ITD

threshold than listener  $\blacklozenge$  ( $46 \mu\text{s}$ ), the prediction error for listener  $\blacklozenge$  was larger (9 dB) than for listener  $\blacktriangledown$ . Again, this is consistent with the hypothesis that SRM in HI listeners does not reliably depend on binaural TFS sensitivity.

Both ITD sensitivity and low frequency hearing loss accounted for some of the variance in BRM (in the  $r = 1$  condition) for HI listeners. However, low frequency hearing loss accounted for a greater portion (Fig. 7). While we cannot make strong conclusions on this point given our relatively small sample size, this result is consistent with other studies that have reported significant relationships between hearing loss and SRM in larger groups of listeners (e.g., Gallun *et al.*, 2013; Glyde *et al.*, 2013c). Furthermore, in view of the fact that a number of other studies have also found that SRM was predicted by low frequency hearing loss (Neher *et al.*, 2009; King *et al.*, 2017), it is perhaps worth noting that sensitivity to ITDs carried by the TFS is highest at lower frequencies.

Interestingly, for NH listeners, we found that ITD sensitivity accounted for a small portion of the variance in BRM for the  $r = 1$  condition and that the correlation was not significant. While we did not necessarily expect a wide range of ITD sensitivity for NH listeners, this result is somewhat inconsistent with Ruggles *et al.* (2011), who reported that SRM for NH was significantly correlated with monaural TFS sensitivity.

### C. Relationship between SRM, BRM, and the binaural masking level difference (BMLD)

As stated in the Introduction, SRM refers to improvements in thresholds due to target and masker being presented from different locations (either over loudspeakers or via HRTFs), while BRM refers to improvements in thresholds due to target and masker being presented with different ITDs. For speech presented in noise, where informational masking is minimal, SRM captures the combined effect of binaural unmasking and better-ear listening, while BRM only captures binaural unmasking (Appendix B). For speech targets presented with speech maskers, however, where informational masking can be substantial, both SRM and BRM can be dominated by the effect of perceived position. In these cases, SRM can be defined as the sum of better-ear listening, binaural unmasking, and perceived position, while BRM can be defined as the sum of binaural unmasking and perceived position. However, in cases where the total masking release from masking experienced by the listener is dominated by perceived position, the distinction between SRM and BRM is minimized.

We note that the binaural unmasking component of BRM/SRM is related to the BMLD, which describes the improved detection of a signal in a masker given differences in interaural phase. While large BMLDs have been previously reported for both speech and tones in noise (e.g., Levitt and Rabiner, 1967; Hirsh and Burgeat, 1958), the relationship between the BMLD and speech intelligibility is not one-to-one, and even large BMLDs can result in a

relatively small improvement in intelligibility for the same stimuli (e.g., Levitt and Rabiner, 1967). This was reflected in our model results (Appendix B), where we show that the predicted contribution of binaural unmasking was modest for our stimuli ( $< 3$  dB across all conditions), at least compared to the BRM experienced by the majority of listeners. Nonetheless, since binaural unmasking likely did contribute modestly to the BRM we report, it is perhaps relevant to consider studies that have examined the relationship between BMLD and ITD thresholds in NH and HI listeners. While some of these studies reported a significant correlation in performance across the two tasks (e.g., Hall *et al.*, 1984; Koehnke *et al.*, 1986; Strelcyk and Dau, 2009), other studies reported a less straightforward relationship (Koehnke *et al.*, 1995; Bernstein *et al.*, 1998). To the extent that individual differences in BRM in our task were driven by individual differences in binaural unmasking, our results are broadly consistent with those studies finding a weak relationship between BMLDs and ITD thresholds. Overall then, it appears that the ability to benefit from supra-threshold ITDs is not always limited by ITD sensitivity at threshold.

Our choice of stimuli that emphasized informational masking and minimized non-spatial segregation cues (by using highly synchronized same-gender talkers; following Swaminathan *et al.*, 2016) has several advantages and disadvantages. One consequence of this choice is that the task is exceptionally challenging in the co-located condition. As seen in Fig. 5, co-located TMRs at threshold are close to 0 dB for both NH and HI listeners, with relatively little variability. This suggests that all listeners require a positive TMR before they can reliably understand the target. A related consequence is that there is much room for improvement and listeners exhibit very large amounts of BRM. An advantage of this experimental design is that it allows us to (1) extend the range of performance and thus measure differences across individuals and across interaural correlation conditions, and (2) minimize the influence of individual differences in non-spatial abilities. One disadvantage, however, is that it is difficult to generalize these results to realistic listening environments where informational masking is generally less prominent (e.g., see Westermann and Buchholz, 2015). It would be interesting in future studies to determine whether the relationship between binaural TFS sensitivity and SRM/BRM in NH and HI listeners is different under more realistic conditions.

### D. On the utility of ITDs and ILDs for SRM

The speech-on-speech task in the current study largely followed the design of Swaminathan *et al.* (2016). However, while stimuli were spatialized using KEMAR HRTFs in Swaminathan *et al.* (2016), stimuli were spatialized by applying ITDs in the current study. This means that listeners in the current study did not have access to ILDs and were entirely dependent on ITDs. For interaurally correlated stimuli, ITDs were available in both the TFS and in the speech

envelope, while for interaurally uncorrelated stimuli, ITDs were only available in the envelope. The fact that ITD thresholds were for the most part measurable in the  $r = 0$  condition suggests that listeners can use envelope ITD cues to make lateral discriminations for broadband speech stimuli even when carried by uncorrelated fine structure. This is consistent with studies suggesting that listeners can use envelope ITD cues when ITDs in the TFS are unusable (e.g., Moore *et al.*, 2018). However, ITD thresholds in this condition were quite large, and sometimes fell outside the ecological range of ITDs ( $\sim 700 \mu\text{s}$ ), especially for HI listeners.

Importantly, the ability to make use of envelope ITD cues to lateralize did not seem to translate to experiencing BRM in speech mixtures. The difference in TMR between the  $r = 0$  condition and co-located condition was not significant (Fig. 5), suggesting that envelope ITDs were not sufficient to support BRM in this task. This is somewhat surprising since Swaminathan *et al.* (2016) found significant SRM ( $\sim 15$  dB) in the  $r = 0$  condition and suggested that this release was likely due to envelope ITDs. It is possible that the SRM they observed was instead supported by the ILDs in their stimuli. The release afforded by ILDs, in this case, is likely not due to improvements in SNR but rather to improved segregation of the talkers due to perceived position (or release from informational masking). This conjecture is supported by the results of a binaural speech intelligibility model, which predicts less than 2 dB of SRM in the  $r = 0$  condition for stimuli spatialized with HRTFs (Swaminathan *et al.*, 2016) or with ITDs only (present study; see Appendix B). The literature is somewhat mixed on whether ILDs alone can support SRM in speech mixtures, with some studies suggesting that this support is minimal (e.g., Culling *et al.*, 2004; Ihlefeld and Litovsky, 2012), and others suggesting this support is as robust as that observed for ITDs (Gallun *et al.*, 2005; Glyde *et al.*, 2013b). It seems that the significant SRM reported by Swaminathan *et al.* (2016) using HRTFs was likely driven by ILD cues that provided a strong segregation cue. It is also possible that access to ILD cues *in addition* to envelope-ITD cues supported a larger release than would be observed with ILD cues alone.

### E. Conclusions

The goal of the present study was to re-examine the hypothesis that poor sensitivity to binaural TFS in HI listeners contributes to their reduced BRM in speech mixtures. We measured both ITD sensitivity for speech in quiet and BRM as a function of TFS interaural correlation in young NH and HI listeners. By characterizing this relationship in NH listeners, we determined the extent to which access to binaural TFS cues jointly supports ITD sensitivity and BRM in an otherwise healthy auditory system. We found that both ITD thresholds and BRM depended on TFS interaural correlation in both groups of listeners. However, HI listeners tended to experience less BRM than NH listeners for a given

ITD sensitivity. We conclude that while binaural TFS sensitivity may account for BRM in some HI listeners, it fails to do so for the majority of cases.

### ACKNOWLEDGMENTS

This work was supported by NIH-NIDCD Award No. DC015760. The international mobility of ML at Boston University (BU) was funded by BU, ENTPE, and the Fondation pour l'audition (Speech2Ears grant). We would like to thank Gerald Kidd and Chris Mason for use of the facilities, assistance with recruitment, and helpful discussions.

### APPENDIX A: AN MODELLING OF INTERAURAL CORRELATIONS

Stimuli were constructed using noise carriers for the left and right ears with systematically decreasing correlations in order to systematically remove access to binaural TFS cues. While our goal was not to remove access to binaural ENV cues, narrowband filtering introduces intrinsic envelope fluctuations, and a disadvantage of using noise vocoders is that within each frequency band, the intrinsic fluctuations from the ENV component of the noise carrier can disrupt the modulating speech ENV (Whitmal *et al.*, 2007). Interaural decorrelation of the noise carriers then, can result in interaural decorrelation of the ENV component in addition to the TFS component of the vocoded stimulus. Furthermore, the relationship between “acoustic” TFS (extracted from the acoustic stimulus) and “neural” TFS (phase-locked activity in the AN) is not obvious and should be approached with caution (Ghitza, 2001; Zeng *et al.*, 2004; Shamma and Lorenzi, 2013). For example, while acoustic TFS is well defined even at high frequencies, neural TFS is much more robust at lower frequencies, within the limits of phase-locking.

Following Swaminathan *et al.* (2016), we quantified the interaural correlation of TFS and ENV components at the output of a phenomenological AN model. We used the computational AN model described by Bruce *et al.* (2018) to generate spike-train responses for individual AN fibers, and calculated neural cross-correlation coefficients  $\rho_{\text{TFS}}$  and  $\rho_{\text{ENV}}$  to quantify the similarity of neural TFS and ENV coding across the left and right ears (Heinz and Swaminathan, 2009).

The definition of the neural cross-correlation coefficient  $\rho$  between two stimuli A (left ear) and B (right ear) follows the form of a normalized cross-correlation, the covariance of A and B divided by the square root of the product of the individual variances of A and B, where values  $\rho$  range from 0 (uncorrelated) to 1 (fully correlated)

$$\rho_{\text{TFS}} = \frac{\text{difcor}_{\text{AB}}}{\sqrt{\text{difcor}_{\text{AA}} \text{difcor}_{\text{BB}}}}, \quad (\text{A1})$$

$$\rho_{\text{ENV}} = \frac{(\text{sumcor}_{\text{AB}} - 1)}{\sqrt{(\text{sumcor}_{\text{AA}} - 1)(\text{sumcor}_{\text{BB}} - 1)}}. \quad (\text{A2})$$



As we will see, the difcor and sumcor signals are defined over a range of delays, but assuming no delay between the stimuli (or correcting for a characteristic delay), both the difcor and sumcor signals have a straightforward interpretation at a delay of zero: the difcor signal reflects the strength of TFS coding, and the sumcor signal reflects the strength of ENV coding. Therefore, when the neural cross-correlation coefficient  $\rho$  is referred to as a single value, it is referring to the zero-delay coefficient.

Figure 8 illustrates signals needed to compute  $\rho_{TFS}$  and  $\rho_{ENV}$  for an example AN fiber with a center frequency (CF) of 1 kHz in response to an example vocoded speech stimulus with an interaural correlation of zero. The difcor and sumcor signals are both derived from shuffled cross-correlograms (SCC), which are robust all-order inter-spike interval histograms for pairs of stimuli [A,B] computed over multiple stimulus repetitions (Joris, 2003; Louage *et al.*, 2004). The shuffled auto-correlogram (SAC) is a special case of the SCC for a stimulus pair [A,A]. Shown in Figs. 8(A)–8(C), the SCC describes the number of coincident spikes in each delay bin (for details of normalization, see Louage *et al.*, 2004). To derive the difcor and sumcor signals, we compute an SCC for a stimulus pair [A<sup>+</sup>,B<sup>+</sup>] and an SCC for the cross-polarity stimulus pair [A<sup>+</sup>,B<sup>-</sup>]. The difcor is defined as the difference between SCC(A<sup>+</sup>,B<sup>+</sup>) and SCC(A<sup>+</sup>,B<sup>-</sup>), while the sumcor is defined as the sum of SCC(A<sup>+</sup>,B<sup>+</sup>) and SCC(A<sup>+</sup>,B<sup>-</sup>).

The logic is that inverting the polarity of the stimulus inverts the acoustic TFS but not the acoustic ENV. Neural

activity that is phase-locked to the acoustic TFS will, therefore, invert along with the stimulus, while neural activity that is not phase-locked to the acoustic TFS (including neural ENV coding) will be unaffected by the stimulus inversion. For example, let us consider the left column of Fig. 8, which shows the SCC, difcor, and sumcor signals for a stimulus pair [A,A], where A is a vocoded speech stimulus presented to the left ear. The thick line in Fig. 8(A) shows SCC(A<sup>+</sup>,A<sup>+</sup>), which has three important features: an oscillation at the CF of the AN fiber, a DC shift above a normalized coincidence number of one, and a symmetric damping with increasing delay. The oscillation reflects neural TFS coding, the DC shift reflects neural ENV coding, and the damping reflects the characteristic falloff of an all-order inter-spike interval histogram. Notice that the difcor signal recovers the oscillation and the sumcor recovers the DC shift already present in the SCC(A<sup>+</sup>,A<sup>+</sup>). The thin line in Fig. 8(A) shows SCC(A<sup>+</sup>,A<sup>-</sup>), and we see that the oscillation is inverted relative to SCC(A<sup>+</sup>,A<sup>+</sup>), reflecting neural TFS coding. We also see that the DC shift is not inverted, reflecting neural ENV coding. Taking the difference between SCC(A<sup>+</sup>,A<sup>+</sup>) and SCC(A<sup>+</sup>,A<sup>-</sup>) effectively reverses the inversion such that the oscillations will sum and the DC shift will cancel, boosting the neural TFS while suppressing the neural ENV. Conversely, taking the sum cancels the oscillations while summing the DC shift, boosting the neural ENV while suppressing the neural TFS. The raw sumcor, however, can contain high-frequency artifacts and

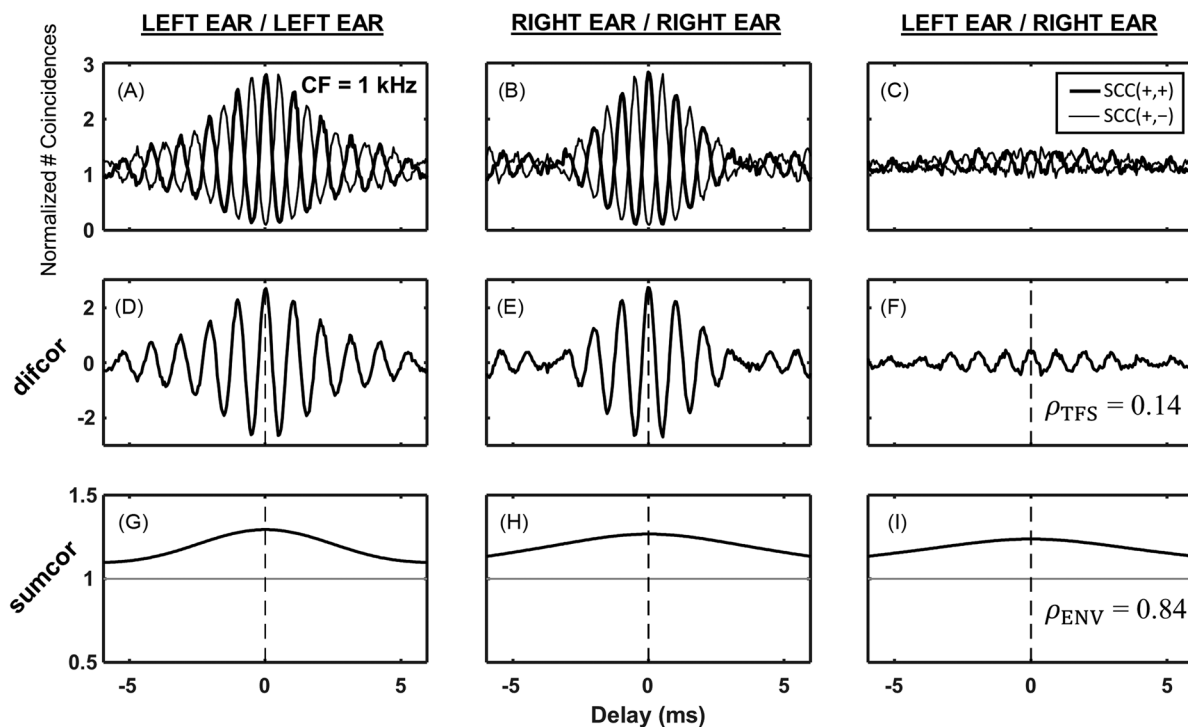


FIG. 8. Example of the correlogram analysis used to compute the neural cross-correlation coefficients that quantified the interaural correlation of neural ENV and neural TFS in response to noise-vocoded speech. In this example, the vocoded stimulus has a carrier interaural correlation of zero ( $\alpha = 0$  and  $\beta = 1$  from equations 1-2). Column 1 shows (A) SCC, (D) difcor, and (G) sumcor signals for the left ear [left, left]. Column 2 shows (B) SCC, (E) difcor, and (H) sumcor signals for the right ear [right, right]. Column 3 shows (C) SCC, (F) difcor, and (I) sumcor signals for the left and right ears [left, right]. The neural cross-correlation coefficient  $\rho_{TFS}$  is computed from the difcor signals (D, E, F) at a delay of zero (dotted vertical line) following Eq. (A1). The neural cross-correlation coefficient  $\rho_{ENV}$  is computed from the sumcor signals (G, H, I) at a delay of zero (dotted vertical line) following Eq. (A2).



can reflect neural activity not associated with the coding of the acoustic ENV (Swaminathan and Heinz, 2012). The sumcor signal was low-pass filtered below 150 Hz in order to remove this activity.

To continue our example, let us now consider the right column of Fig. 8, which shows the SCC, difcor, and sumcor signals for a stimulus pair [A,B]. In this case, the acoustic TFS is uncorrelated across A and B, while the acoustic ENV is highly correlated. In Fig. 8(C), we see that the oscillation component of the SCCs is dramatically reduced but that DC shift component is relatively unchanged compared to the SCCs in Figs. 8(A) and 8(B). Indeed, when we extract the difcor and sumcor signals, we see that the difcor signal in Fig. 8(F) is dramatically smaller than the difcor signals in Figs. 8(D) and 8(E) and that the sumcor signal in Fig. 8(I) is similar to the sumcor signals in Figs. 8(G) and 8(H). Using Eqs. (A1) and (A2), we can compute  $\rho_{TFS}$  from the difcor signals [Figs. 8(D)–8(F)] and  $\rho_{ENV}$  from the sumcor signals [Figs. 8(G)–8(I)], and thereby quantify the relationship between interaural correlations in the acoustic stimulus and in the neural coding of that stimulus, separately for TFS and ENV components.

For the stimulus in this example, the neural TFS interaural correlation (for an AN fiber with a CF of 1 kHz) is close to zero ( $\rho_{TFS} = 0.14$ ), consistent with the fact that the carrier interaural correlation was set to zero. Conversely, the neural interaural ENV correlation is close to one ( $\rho_{ENV} = 0.84$ ), consistent with the fact that the same speech envelope was used to modulate the carrier in the left and right ears.

To quantify the neural interaural correlation for a particular stimulus, we need to integrate neural coding across a range of AN fibers. We defined a set of AN fibers with CFs corresponding to the CFs of the vocoder filterbank between 200 and 8000 Hz. Because the energy in the stimulus was not equal across bands, and because low-frequency fibers have more robust TFS coding while high-frequency fibers have more robust ENV coding, a simple summation of  $\rho_{TFS}$  and  $\rho_{ENV}$  across channels was not appropriate. The rms of the difcor signal *within* each ear is a magnitude measure that depends on both the energy in the stimulus and the strength of neural TFS coding within a fiber, while the rms of the sumcor depends on both the energy in the stimulus and the strength of neural ENV coding. To correct for differences in strength of neural coding across fibers then,  $\rho_{TFS}$  was normalized by the rms of the difcor signal within each ear, and  $\rho_{ENV}$  was normalized by the rms of the sumcor signal within each ear, irrespective of interaural correlation

$$\hat{\rho}_{TFS} = \rho_{TFS} * \frac{\text{rms}(\text{difcor}_{AA}) + \text{rms}(\text{difcor}_{BB})}{2}, \quad (\text{A3})$$

$$\hat{\rho}_{ENV} = \rho_{ENV} * \frac{\text{rms}(\text{sumcor}_{AA}) + \text{rms}(\text{sumcor}_{BB})}{2}, \quad (\text{A4})$$

where A = right ear and B = left ear.

After normalization,  $\hat{\rho}_{TFS}$  and  $\hat{\rho}_{ENV}$  were summed across channels to yield integrated TFS and ENV interaural

correlations *across* the AN. Integrated  $\hat{\rho}_{TFS}$  and  $\hat{\rho}_{ENV}$  coefficients were calculated for all experimental stimuli at all five carrier interaural correlation values. In order to rescale the integrated coefficients to have a maximum of one, all  $\hat{\rho}_{TFS}$  coefficients were divided by the maximum  $\hat{\rho}_{TFS}$  coefficient across interaural correlation values, and all  $\hat{\rho}_{ENV}$  coefficients were divided by the maximum  $\rho_{ENV}$  coefficient.

## APPENDIX B: PREDICTION OF THE ENERGETIC COMPONENTS OF SRM

A model has been used here to evaluate the energetic component of SRM (or more precisely, BRM) for the stimuli in the present study as well as a subset of stimuli from Swaminathan *et al.* (2016). This energetic component is calculated as the sum of two contributions: the better-ear advantage and binaural unmasking. The model is a slight revision of the original model of Collin and Lavandier (2013), which predicts binaural speech intelligibility in the presence of multiple non-stationary noises. From the signals generated at the ears by the target and those generated by the sum of all maskers, it computes a better-ear SNR and a binaural unmasking advantage. Here, the predicted SRM was obtained by computing the difference in better-ear SNR between the co-located and separated conditions, and adding the corresponding binaural unmasking advantage.

The predictions are based on short-term predictions averaged across time. The model uses 24-ms half-overlapping Hann windows to compute the better-ear ratio and 300-ms windows to compute the binaural unmasking advantage (longer windows are used in order to take into account binaural sluggishness). A ceiling of 20 dB is also applied as the maximum better-ear ratio allowed by frequency band and time frame (Cubick *et al.*, 2018). For the present study, 100 trials were simulated, where each trial consisted of a single target sentence and a pair of masker sentences. The target sentences were averaged after truncation to the duration of the shortest sentence. The model computes the long-term statistics (spectra at the ears and interaural parameters) of the target only once. These long-term target statistics are combined with the short-term masker statistics to compute the better-ear and binaural unmasking components within each time frame (before averaging). After the two masker sentences were summed to yield the masker token, masker tokens were concatenated into a single signal *across* trials, and the averaged target and concatenated masker signals were each equalized in root-mean-square power so that the absolute level and interaural level difference was equal to their average across the original 100 tokens. To compare predictions across conditions, the same original sentences (target and masker) were used to generate the signals in each condition.

Figure 9 presents the predicted SRM obtained in the HRTF (Swaminathan *et al.*, 2016) condition for  $r=0$ , and in the ITD conditions of the current study ( $r=0, 0.2, 0.4, 0.6, 1$ ). In these latter conditions, the variation in SRM is mainly due to the increase of binaural unmasking advantage

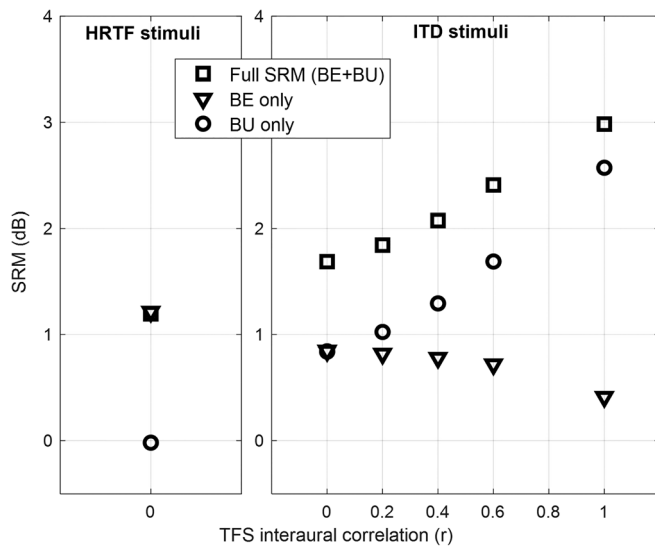


FIG. 9. Predicted SRM computed as the difference in better-ear ratio (BE) and binaural unmasking advantage (BU) between the co-located condition ( $r=0$  for the HRTF-spatialized stimuli,  $r=1$  for the ITD-spatialized stimuli) and separated condition ( $r$  on the  $x$  axis).

with increasing interaural coherence. The binaural unmasking computation relies on the determination of the target and masker ITDs. For the  $r=0$  condition, the ITD is not defined and its determination is random, leading to an advantage of 0.8 dB in all conditions (co-located or separated). For the HRTF stimuli, SRM was defined as the difference in TMR between co-located and separated configurations in the  $r=0$  condition and thus the model found 0 dB of binaural unmasking release. However, for the ITD stimuli, SRM was always defined in reference to the co-located  $r=1$  condition. Therefore, for the ITD stimuli in the  $r=0$ , a binaural unmasking release of 0.8 dB was predicted.

The better-ear contribution of 1.3 dB with the HRTF stimuli is very limited compared to the one that has been observed in other symmetric configurations with other speech material (e.g., the 5-dB SRM for symmetric modulated maskers at  $\pm 60$  degrees reported by Ewert *et al.*, 2017). This could be due to the use of highly synchronized speech tokens, which provide little opportunity for head-shadow advantages. Also, there was a left/right asymmetry in the HRTFs used by Swaminathan *et al.* (2016), which led to a slightly negative SRM ( $-0.7$  dB) when computed with the stationary version of the binaural model (Jelfs *et al.*, 2011) applied directly to the HRTFs, suggesting that the SRM of 1.3 dB we observed might slightly underestimate the SRM that would be experienced if HRTFs were not normalized. The very small better-ear contribution to SRM with the ITD stimuli could be due in part to the slight misalignment of words due to the ITD, and in part to distortions in the envelope as a consequence of using interaurally decorrelated carriers (Fig. 3).

This analysis suggests that when carriers were interaurally uncorrelated ( $r=0$ ), the energetic component of SRM was 1.3 dB for HRTF-spatialized stimuli, and 1.7 dB for ITD-spatialized stimuli. We also see that the energetic

component of SRM systematically increases from 1.7 to 3 dB as a function of interaural correlation for ITD-spatialized stimuli.

<sup>1</sup>See supplementary material at <https://doi.org/10.1121/10.0000812> for Supplemental Fig. 1.

<sup>2</sup>Listener  $\times$  was also tested without applying the NAL-RP gain filter. Instead of generating stimuli at a nominal level of 70 dB and then applying a gain filter, stimuli were presented at 100 dB with no gain filter. ITD thresholds remained unmeasurable.

Bernstein, L. R., Trahiotis, C., and Hyde, E. L. (1998). "Inter-individual differences in binaural detection of low-frequency or high-frequency tonal signals masked by narrow-band or broadband noise," *J. Acoust. Soc. Am.* **103**(4), 2069–2078.

Best, V., Carlile, S., Kopčo, N., and van Schaik, A. (2011). "Localization in speech mixtures by listeners with hearing loss," *J. Acoust. Soc. Am.* **129**(5), EL210–EL215.

Best, V., Marrone, N., Mason, C. R., and Kidd, G. (2012). "The influence of non-spatial factors on measures of spatial release from masking," *J. Acoust. Soc. Am.* **131**(4), 3103–3110.

Best, V., Mason, C. R., Swaminathan, J., Roverud, E., and Kidd, G. (2017). "Use of a glimpsing model to understand the performance of listeners with and without hearing loss in spatialized speech mixtures," *J. Acoust. Soc. Am.* **141**(1), 81–91.

Best, V., and Swaminathan, J. (2019). "Revisiting the detection of interaural time differences in listeners with hearing loss," *J. Acoust. Soc. Am.* **145**(6), EL508–EL513.

Best, V., Thompson, E. R., Mason, C. R., and Kidd, G. (2013). "An energetic limit on spatial release from masking," *J. Assoc. Res. Otolaryngol.* **14**(4), 603–610.

Bronkhorst, A. W., and Plomp, R. (1992). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," *J. Acoust. Soc. Am.* **92**(6), 3132–3139.

Bruce, I. C., Erfani, Y., and Zilany, M. S. A. (2018). "A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites," *Hear. Res.* **360**, 40–54.

Brungart, D. S., and Iyer, N. (2012). "Better-ear glimpsing efficiency with symmetrically-placed interfering talkers," *J. Acoust. Soc. Am.* **132**(4), 2545–2556.

Cherry, E. C. (1953). "Some experiments on the recognition of speech, with one and with two ears," *J. Acoust. Soc. Am.* **25**(5), 975–979.

Collin, B., and Lavandier, M. (2013). "Binaural speech intelligibility in rooms with variations in spatial location of sources and modulation depth of noise interferers," *J. Acoust. Soc. Am.* **134**(2), 1146–1159.

Cubick, J., Buchholz, J. M., Best, V., Lavandier, M., and Dau, T. (2018). "Listening through hearing aids affects spatial perception and speech intelligibility in normal-hearing listeners," *J. Acoust. Soc. Am.* **144**(5), 2896–2905.

Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (2004). "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," *J. Acoust. Soc. Am.* **116**(2), 1057–1065.

Dillon, H. (2012). *Hearing Aids* (Boomerang Press, Sydney).

Drennan, W. R., Won, J. H., Dasika, V. K., and Rubinstein, J. T. (2007). "Effects of temporal fine structure on the lateralization of speech and on speech understanding in noise," *J. Assoc. Res. Otolaryngol.* **8**(3), 373–383.

Ewert, S. D., Schubotz, W., Brand, T., and Kollmeier, B. (2017). "Binaural masking release in symmetric listening conditions with spectro-temporally modulated maskers," *J. Acoust. Soc. Am.* **142**(1), 12–28.

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**(6), 3578–3588.

Füllgrabe, C., and Moore, B. C. J. (2018). "The association between the processing of binaural temporal-fine-structure information and audiometric threshold and age: A meta-analysis," *Trends Hear.* **22**, 233121651879725.

- Gallun, F. J., Diedesch, A. C., Kampel, S. D., and Jakien, K. M. (2013). "Independent impacts of age and hearing loss on spatial release in a complex auditory environment," *Front. Neurosci.* **7**(252), 1–11.
- Gallun, F. J., Mason, C. R., and Kidd, G. (2005). "Binaural release from informational masking in a speech identification task," *J. Acoust. Soc. Am.* **118**(3), 1614.
- Gallun, F. J., McMillan, G. P., Molis, M. R., Kampel, S. D., Dann, S. M., and Konrad-Martin, D. L. (2014). "Relating age and hearing loss to monaural, bilateral, and binaural temporal sensitivity," *Front. Neurosci.* **8**, 172.
- Ghitza, O. (2001). "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," *J. Acoust. Soc. Am.* **110**(3), 1628.
- Glyde, H., Buchholz, J. M., Dillon, H., Cameron, S., and Hickson, L. (2013b). "The importance of interaural time differences and level differences in spatial release from masking," *J. Acoust. Soc. Am.* **134**(2), EL147–EL152.
- Glyde, H., Buchholz, J. M., Nielsen, L., Best, V., Dillon, H., Cameron, S., and Hickson, L. (2015). "Effect of audibility on spatial release from speech-on-speech masking," *J. Acoust. Soc. Am.* **138**(5), 3311–3319.
- Glyde, H., Cameron, S., Dillon, H., Hickson, L., and Seeto, M. (2013c). "The effects of hearing impairment and aging on spatial processing," *Ear Hear.* **34**(1), 15–28.
- Goupell, M. J., and Hartmann, W. M. (2007). "Interaural fluctuations and the detection of interaural incoherence. III. Narrowband experiments and binaural models," *J. Acoust. Soc. Am.* **122**(2), 1029–1045.
- Hall, J. W., Tyler, R. S., and Fernandes, M. A. (1984). "Factors influencing the masking level difference in cochlear hearing-impaired and normal hearing listeners," *J. Speech Hear. Res.* **27**, 145–154.
- Hartmann, W. M., and Cho, Y. J. (2011). "Generating partially correlated noise—A comparison of methods," *J. Acoust. Soc. Am.* **130**(1), 292–301.
- Hauth, C. F., and Brand, T. (2018). "Modeling sluggishness in binaural unmasking of speech for maskers with time-varying interaural phase differences," *Trends Hear.* **22**, 233121651775354.
- Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (1999). "Speech intelligibility and localization in a multi-source environment," *J. Acoust. Soc. Am.* **105**(6), 3436–3448.
- Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (2004). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," *J. Acoust. Soc. Am.* **115**(2), 833–843.
- Heinz, M. G., and Swaminathan, J. (2009). "Quantifying envelope and fine-structure coding in auditory nerve responses to chimaeric speech," *J. Assoc. Res. Otolaryngol.* **10**(3), 407–423.
- Hirsh, I. J., and Burgeat, M. (1958). "Binaural effects in remote masking," *J. Acoust. Soc. Am.* **30**, 827–832.
- Hopkins, K., and Moore, B. C. J. (2011). "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise," *J. Acoust. Soc. Am.* **130**(1), 334–349.
- Hu, H., Ewert, S. D., McAlpine, D., and Dietz, M. (2017). "Differences in the temporal course of interaural time difference sensitivity between acoustic and electric hearing in amplitude modulated stimuli," *J. Acoust. Soc. Am.* **141**(3), 1862–1873.
- Ihlefeld, A., and Litovsky, R. Y. (2012). "Interaural level differences do not suffice for restoring spatial release from masking in simulated cochlear implant listening," *PLoS One* **7**(9), e45296.
- Jakien, K. M., Kampel, S. D., Gordon, S. Y., and Gallun, F. J. (2017). "The benefits of increased sensation level and bandwidth for spatial release from masking," *Ear Hear.* **38**(1), e13–e21.
- Jelfs, S., Culling, J. F., and Lavandier, M. (2011). "Revision and validation of a binaural model for speech intelligibility in noise," *Hear. Res.* **275**, 96–104.
- Joris, P. X. (2003). "Interaural time sensitivity dominated by cochlea-induced envelope patterns," *J. Neurosci.* **23**(15), 6345–6350.
- Kidd, G., Best, V., and Mason, C. R. (2008). "Listening to every other word: Examining the strength of linkage variables in forming streams of speech," *J. Acoust. Soc. Am.* **124**(6), 3793–3802.
- Kidd, G., and Colburn, H. S. (2017). "Informational masking in speech recognition," in *The Auditory System at the Cocktail Party*, edited by J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay (Springer, New York), pp. 75–109.
- Kidd, G., Mason, C. R., Best, V., and Marrone, N. (2010). "Stimulus factors influencing spatial release from speech-on-speech masking," *J. Acoust. Soc. Am.* **128**(4), 1965–1978.
- Kidd, G., Mason, C. R., Best, V., Roverud, E., Swaminathan, J., Jennings, T., Clayton, K., and Steven Colburn, H. (2019). "Determining the energetic and informational components of speech-on-speech masking in listeners with sensorineural hearing loss," *J. Acoust. Soc. Am.* **145**(1), 440–457.
- King, A., Hopkins, K., Plack, C. J., Pontoppidan, N. H., Bramsløw, L., Hietkamp, R. K., Vatti, M., and Hafez, A. (2017). "The effect of tone-coding on spatial release from masking for old, hearing-impaired listeners," *J. Acoust. Soc. Am.* **141**(4), 2591–2603.
- Koehnke, J., Colburn, H. S., and Durlach, N. I. (1986). "Performance in several binaural-interaction experiments," *J. Acoust. Soc. Am.* **79**(5), 1558.
- Koehnke, J., Culotta, C. P., Hawley, M. L., and Colburn, H. S. (1995). "Effects of reference interaural time and intensity differences on binaural performance in listeners with normal and impaired hearing," *Ear Hear.* **16**(4), 331–353.
- Levitt, H., and Rabiner, L. R. (1967). "Predicting binaural gain in intelligibility and release from masking for speech," *J. Acoust. Soc. Am.* **42**, 820–829.
- Lócssei, G., Pedersen, J. H., Laugesen, S., Santurette, S., Dau, T., and MacDonald, E. N. (2016). "Temporal fine-structure coding and lateralized speech perception in normal-hearing and hearing-impaired listeners," *Trends Hear.* **20**, 233121651666096.
- Lorenzi, C., Gilbert, G., Cam, H., Garnier, S., and Moore, B. C. J. (2006). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," *Proc. Natl. Acad. Sci.* **103**(49), 18866–18869.
- Louage, D. H. G., van der Heijden, M., and Joris, P. X. (2004). "Temporal properties of responses to broadband noise in the auditory nerve," *J. Neurophys.* **91**(5), 2051–2065.
- Marrone, N., Mason, C. R., and Kidd, G. (2008). "The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms," *J. Acoust. Soc. Am.* **124**(5), 3064–3075.
- Moore, B. C. J. (2004). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, London).
- Moore, B. C. J. (2008). "The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people," *J. Assoc. Res. Otolaryngol.* **9**(4), 399–406.
- Moore, B. C. J., Heinz, M. G., Braida, L. D., and Léger, A. C. (2018). "Effects of age on sensitivity to interaural time differences in envelope and fine structure, individually and in combination," *J. Acoust. Soc. Am.* **143**(3), 1287–1296.
- Neher, T., Behrens, T., Carlile, S., Jin, C., Kragelund, L., Petersen, A. S., and Schaik, A. V. (2009). "Benefit from spatial separation of multiple talkers in bilateral hearing-aid users: Effects of hearing loss, age, and cognition," *Int. J. Audiol.* **48**(11), 758–774.
- Neher, T., Lunner, T., Hopkins, K., and Moore, B. C. J. (2012). "Binaural temporal fine structure sensitivity, cognitive function, and spatial speech recognition of hearing-impaired listeners (L)," *J. Acoust. Soc. Am.* **131**(4), 2561–2564.
- Papesh, M. A., Folmer, R. L., and Gallun, F. J. (2017). "Cortical measures of binaural processing predict spatial release from masking performance," *Front. Hum. Neurosci.* **11**, 124.
- Rana, B., and Buchholz, J. M. (2018). "Effect of audibility on better-ear glimpsing as a function of frequency in normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **143**(4), 2195–2206.
- Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. G. (2011). "Normal hearing is not enough to guarantee robust encoding of supra-threshold features important in everyday communication," *Proc. Natl. Acad. Sci.* **108**(37), 15516–15521.
- Schoenmaker, E., Brand, T., and van de Par, S. (2016). "The multiple contributions of interaural differences to improved speech intelligibility in multitalker scenarios," *J. Acoust. Soc. Am.* **139**(5), 2589–2603.
- Shamma, S., and Lorenzi, C. (2013). "On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system," *J. Acoust. Soc. Am.* **133**(5), 2818–2833.
- Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). "Chimaeric sounds reveal dichotomies in auditory perception," *Nature* **416**(6876), 87–90.
- Strelcyk, O., and Dau, T. (2009). "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing," *J. Acoust. Soc. Am.* **125**(5), 3328.
- Swaminathan, J., and Heinz, M. G. (2012). "Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise," *J. Neurosci.* **32**(5), 1747–1756.

- Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Roverud, E., and Kidd, G. (2016). "Role of binaural temporal fine structure and envelope cues in cocktail-party listening," *J. Neurosci.* **36**(31), 8250–8257.
- Swaminathan, J., Reed, C. M., Desloge, J. G., Braida, L. D., and Delhorne, L. A. (2014). "Consonant identification using temporal fine structure and recovered envelope cues," *J. Acoust. Soc. Am.* **135**(4), 2078–2090.
- Westermann, A., and Buchholz, J. M. (2015). "The influence of informational masking in reverberant, multi-talker environments," *J. Acoust. Soc. Am.* **138**(2), 584–593.
- Whitmal, N. A., Poissant, S. F., Freyman, R. L., and Helfer, K. S. (2007). "Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience," *J. Acoust. Soc. Am.* **122**(4), 2376–2388.
- Yang, X., and Grantham, D. W. (1997). "Echo suppression and discrimination suppression aspects of the precedence effect," *Percept. Psychophys.* **59**(7), 1108–1117.
- Zeng, F.-G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y.-Y., and Chen, H. (2004). "On the dichotomy in auditory perception between temporal envelope and fine structure cues (L)," *J. Acoust. Soc. Am.* **116**(3), 1351–1354.
- Zurek, P. M. (1993). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. Studebaker and I. Hochberg (Allyn and Bacon, Needham Heights, MA), pp. 255–276.