Determining the energetic and informational components of speech-on-speech masking in listeners with sensorineural hearing loss

Gerald Kidd, Jr.,^{a)} Christine R. Mason, Virginia Best, Elin Roverud, Jayaganesh Swaminathan,^{b)} Todd Jennings, and Kameron Clayton^{c)}

Department of Speech, Language and Hearing Sciences, Boston University, Boston, Massachusetts 02215, USA

H. Steven Colburn

Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, USA

(Received 20 July 2018; revised 19 November 2018; accepted 18 December 2018; published online 25 January 2019)

The ability to identify the words spoken by one talker masked by two or four competing talkers was tested in young-adult listeners with sensorineural hearing loss (SNHL). In a reference/baseline condition, masking speech was colocated with target speech, target and masker talkers were female, and the masker was intelligible. Three comparison conditions included replacing female masker talkers with males, time-reversal of masker speech, and spatial separation of sources. All three variables produced significant release from masking. To emulate energetic masking (EM), stimuli were subjected to ideal time-frequency segregation retaining only the time-frequency units where target energy exceeded masker energy. Subjects were then tested with these resynthesized “glimpsed stimuli.” For either two or four maskers, thresholds only varied about 3 dB across conditions suggesting that EM was roughly equal. Compared to normal-hearing listeners from an earlier study [Kidd, Mason, Swaminathan, Roverud, Clayton, and Best, *J. Acoust. Soc. Am.* **140**, 132–144 (2016)], SNHL listeners demonstrated both greater energetic and informational masking as well as higher glimpsed thresholds. Individual differences were correlated across masking release conditions suggesting that listeners could be categorized according to their general ability to solve the task. Overall, both peripheral and central factors appear to contribute to the higher thresholds for SNHL listeners. © 2019 Acoustical Society of America. <https://doi.org/10.1121/1.5087555>

[JLJ]

Pages: 440–457

I. INTRODUCTION

There is considerable evidence that sensorineural hearing loss (SNHL) adversely affects a listener's ability to hear and understand the speech of one specific talker in the presence of competing talkers. This evidence is varied and consists of subjective reports of the experience of communication difficulty (e.g., Gatehouse and Noble, 2004; Noble and Gatehouse, 2006; Agus *et al.*, 2009) as well as the findings from laboratory-based studies of speech-on-speech (SOS) masking (e.g., Festen and Plomp, 1990; Arbogast *et al.*, 2005; Marrone *et al.*, 2008a,b; Best *et al.*, 2012, 2013a; Gallun *et al.*, 2013; Dawes *et al.*, 2013; Srinivasan *et al.*, 2017; Jakien *et al.*, 2017). Despite this accumulation of evidence, the causes for the difficulties experienced by listeners with SNHL in multiple-talker sound fields remain poorly understood. The limitations of our understanding of this problem are due in part to the complexity of solving the SOS masking task in general, which places demands on processing at multiple sites in the system spanning the auditory pathway from sensory transduction in the periphery to

linguistic processing in the brain (e.g., Freyman *et al.*, 2004; Mattys *et al.*, 2012; Calandruccio *et al.*, 2013, 2017; Kidd *et al.*, 2014; Roberts and Summers, 2018). This raises the possibility that adverse effects of SNHL on performance could occur at multiple stages of processing of the stimulus or at different stages for different individuals with otherwise similar audiometric profiles (e.g., Shinn-Cunningham, 2017). Furthermore, the outcome of the SOS masking experiment itself is quite sensitive to the degree of uncertainty that is present in the listener's task. Listener uncertainty can be a difficult variable to control or to quantify and the particulars of the experimental design—which determine the uncertainty inherent to the task—may exert a profound influence on the outcome of the experiment (e.g., discussion in Kidd and Colburn, 2017). Differences in experimental design across studies comparing SOS masking in listeners with normal hearing (NH) to listeners with SNHL—the speech materials used, the availability of low-level segregation cues, the means by which the target source is designated, etc.—contribute to the difficulty in drawing firm conclusions about the differences in performance that have been observed.

In this article we examine these issues by testing a group of listeners with SNHL under conditions that exactly match (except for the frequency-specific gain used to compensate for hearing loss, described later) those tested in an earlier

^{a)}Electronic mail: gakidd@bu.edu

^{b)}Current address: Starkey Hearing Research, Berkeley, CA 94704, USA.

^{c)}Current address: Harvard Medical School, Boston, MA 02115, USA.

study of listeners with NH (Kidd *et al.*, 2016). The focus of this study is on one aspect of this complicated problem: attempting to better understand the interaction between SNHL and the *type of masking* that is present in the listening environment. By “type of masking” we are referring to the distinction that often is made between masking that originates primarily in the auditory periphery, termed energetic masking (EM), and masking that originates primarily at higher levels of the auditory system, termed informational masking (IM) (e.g., Kidd *et al.*, 2008b). Energetic masking occurs due to the overlap of excitation in peripheral neural structures and limits the information available for processing at subsequent levels in the system (e.g., Durlach *et al.*, 2003a) while IM occurs despite a neural representation of the target stimulus from the periphery that is sufficient for the observer to solve the task (e.g., Kidd *et al.*, 2002b; Leibold *et al.*, 2018). Because the primary cause of IM is listener uncertainty, and therefore reflects the internal state of the observer, it can be difficult to isolate and control as a variable in experimental design and performance may depend on factors such as stimulus predictability and the *a priori* knowledge available to the observer. Furthermore, because the processing that occurs subsequent to peripheral transduction depends on the fidelity of the representations that are received at each successive stage, factors that affect EM may manifest as differences in performance on tasks intended to measure aspects of auditory perception and cognitive processing (e.g., Kidd *et al.*, 2002a; Shinn-Cunningham and Best, 2008). For listeners with SNHL, making effective use of incomplete or distorted peripheral representations may depend on rapidly “filling in” the missing information at some stage of processing and presumably draws heavily on the predictability provided by context (see Başkent *et al.*, 2016, for a review). When attempting to determine the origins of the difficulties encountered by SNHL listeners in complex listening situations, such as in multiple-talker sound fields created in the laboratory, controlling the type of masking—EM vs IM—may be crucial to understanding the factors governing performance. However, very few studies comparing SOS masking in NH and SNHL listeners have attempted to explicitly take into account or vary in a controlled manner the degree of EM and IM in the experimental design (cf. Arbogast *et al.*, 2005; Agus *et al.*, 2009; Best *et al.*, 2013a).

Among the more effective tools for separating EM from IM in SOS masking is “ideal time-frequency segregation” (ITFS), a technique originally developed as a means for studying computational auditory scene analysis (e.g., Wang, 2005) but which was applied to the SOS masking problem by Brungart and colleagues (2006, 2009). ITFS processing usually assumes *a priori* knowledge of the target and masker signals so that the energy relations in each time-frequency (T-F) unit may be specified exactly and, in that sense, is fundamentally different than normal human perception. The use of ITFS to separate EM from IM is based on the premise that listeners extracting speech in a mixture of sounds likely rely only (or predominantly) on the subset of T-F units in which the energy of the target source relative to the energy of the masking source(s) (e.g., Cooke, 2006; Li and Loizou, 2007; Healey and Vasko, 2018) exceeds a specified value (termed

the level criterion, or LC). The T-F units in which masker energy dominates target energy logically fall under the definition of EM because of the assumption that the neural response in a small T-F unit would be driven by the properties (i.e., amplitude and timing) of the higher-energy source. Thus, performing ITFS on a SOS mixture, and using the processed stimulus as the target speech, emulates EM because it eliminates the T-F units dominated by masker energy which presumably contribute little to target speech intelligibility. However, two relevant points about ITFS processing—both of which were made by Brungart *et al.* (2006) in their original article—are that, first, the extent in time and frequency of the T-F units used in the analysis may not be matched to the limits of the internal representations in the human auditory system. The choice of the T-F window has some bearing on the interpretation of the differences in performance observed between NH and SNHL listeners because the internal time-frequency representations between the two groups may differ (e.g., Kidd *et al.*, 1984) and factors may affect the neural signal that are unaccounted for by the configuration of peripheral loss (e.g., Henry and Heinz, 2012; Henry *et al.*, 2016; Chambers *et al.*, 2016; review in Carney, 2018). Second, although the assumption that ITFS processing removes the energetically masked T-F units is reasonable and straight-forward, ITFS processing *also* eliminates the IM that is present in those masker-dominated T-F units. Thus, ITFS processing may achieve the analog of perceptual segregation almost perfectly (“almost” here means that some unwanted masker energy is still present within target-dominated T-F units and some useful target energy below the LC value may be discarded). What this means is that the difference in intelligibility observed between the “glimpsed” target stimulus (the speech that remains after ITFS processing) and the intelligibility of the masked “natural” target stimulus (prior to ITFS processing)—which is termed “additional masking”—is an indicator of the IM caused by the masker (e.g., Brungart *et al.*, 2006; Kidd *et al.*, 2016; Buss *et al.*, 2017). Both of these factors must be considered in the interpretation of experiments using ITFS processed stimuli and its application to the study of SOS masking in SNHL.¹

In the present study, performance in SOS masking conditions was measured in a group of young adult listeners with SNHL for conditions in which the stimulus was presented naturally and for the same conditions following ITFS processing. Furthermore, these two processing approaches (termed “natural” and “glimpsed,” respectively) were applied to several types of maskers and in a variety of presentation conditions including a high IM baseline, three conditions in which explicit source segregation cues were provided, and a noise masker condition that was intended as a low IM control. These stimuli and presentation conditions allowed us to address the following sets of questions: First, do listeners with SNHL exhibit greater than normal EM and/or IM in SOS masking conditions? And, conversely, how does the benefit obtained from stimulus manipulations intended to perceptually segregate target speech from masker speech in listeners with SNHL compare to the corresponding benefit found for listeners with NH? Second, is the release from SOS masking due to various source segregation cues correlated among individuals with

SNHL? Recent work by Kidd *et al.* (2016) indicates that NH listeners tend to be categorized according to their general ability to use source segregation cues to overcome IM. Would this same pattern be observed for listeners with SNHL? A related question is whether the release from masking observed for individual listeners with SNHL depends on the degree of hearing loss. These questions will be addressed using a correlational approach that parallels that reported by Kidd *et al.* (2016). Third, does a constant level of speech identification performance (i.e., at the 50% correct “threshold” point) correspond to the availability of a constant proportion of target energy in the glimpsed stimulus? Because ITFS processing essentially removes the effect of the masker—any masker—it is plausible that all that matters in the glimpsed stimulus is how much of the target remains. This would imply that a constant level of target identification performance would be directly related to a constant proportion of the target information retained after processing. If that is not the case across listeners and/or conditions, it would suggest that other factors influence the results. Some of the factors that could lead to this result include insufficient audibility of the stimulus due to inadequate amplification (e.g., Glyde *et al.*, 2015; Best *et al.*, 2017; Rana and Buchholz, 2018) and/or a reduced ability of the listener with SNHL to constitute a coherent stream of speech from the sparse “glimpses” of the target (e.g., Bernstein *et al.*, 2013) that are available in the ITFS processed/reconstructed speech. Comparison of glimpsed target speech intelligibility across listeners and maskers may help to answer this question.

II. METHODS

A. Subjects

Twelve adult listeners ranging in age from 18 to 40 years with bilateral SNHL participated. Ten listeners participated in the main portion of the study while in a subsequent condition using a noise masker, eight SNHL listeners participated including six of the original listeners plus two additional listeners. The correlation between age and hearing loss (as defined by the 4-frequency pure-tone average of 500, 1000, 2000, and 4000 Hz) was not significant, with $r=0.13$ and $p(1\text{-tailed})=0.34$. Figure 1 shows the group mean audiogram and standard errors of the means. As a group the hearing losses were bilaterally symmetric and sloped about 10 dB/octave over the range of frequencies tested. All subjects received compensation for their participation in the experiments. For the SNHL listeners the stimuli were amplified using individualized linear gain according to the NAL-RP prescriptive formula (Byrne *et al.*, 1991) determined separately for each ear. Levels are specified prior to the application of gain. Eight NH listeners matched in mean age to the group of SNHL listeners (approximately 24 years) participated in the noise masker conditions.

B. Stimuli

The speech materials consisted of a set of 40 monosyllabic words divided into five syntactic categories: <name> <verb> <number> <adjective> <object> with eight

FIG. 1. Group mean audiometric thresholds and standard errors of the means for the subjects with sensorineural hearing loss.

exemplars in each category (Kidd *et al.*, 2008a). The speech targets and maskers were the same as those that were used in Kidd *et al.* (2016). The words were spoken by 16 young-adult talkers with an equal number of males and females. The average fundamental frequency of the female talkers was 201 Hz (s.d. 20 Hz) and of the male talkers was 109 Hz (s.d. 12 Hz). The recordings were produced by Sensimetrics Corporation (Malden, MA). All talkers recorded all words. The words were spoken individually so that they could be concatenated in any order without causing acoustic differences due to across-word coarticulation. The sentences were used for both the “target” that was to be identified and the speech “maskers” that were to be ignored. The target sentence was designated by the name “Sue” which was used for the first word in that sentence and was not scored. The remaining four words in the target sentence were chosen at random from the eight exemplars in each of the other four syntactic categories (after names) and were scored. An example is “Sue found three red shoes.” An 8×5 grid of the corpus of test words was displayed on the subject’s monitor. The subjects were instructed to mouse-click the five words in order, one from each column left to right, comprising the target sentence. The words used for the sentences within a trial were always mutually exclusive as were the talkers which were selected randomly, with replacement, from eight possible talkers on each trial.

The noise maskers were speech-spectrum-shaped (based on the measured long-term average spectrum of our group of female talkers) and speech-envelope-modulated (using the envelopes of the recorded words as spoken in the speech masker conditions). Thus, the noise maskers were essentially single-channel vocoded stimuli with the same long term average spectrum and envelopes as the speech maskers but were unintelligible.

C. Procedures

The procedures were identical to those used by Kidd *et al.* (2016) with exceptions as noted below. The target sentence was always spoken by a female talker and presented from 0° azimuth (directly in front of the listener). There were either two or four masker sentences presented concurrently with the target sentence with the onset of the first word in each sentence time aligned and natural variation in alignment as word durations varied for the subsequent words. In the reference condition (designated as the “baseline” condition), these masker sentences also were spoken by female talkers from 0° azimuth (i.e., same sex talkers, intelligible speech, and colocated presentation). There were three other comparison conditions tested which were expected to produce a release from masking. These “masking release” conditions were (a) the use of male voices for the masker sentences (“male”), (b) the time-reversal of the words forming the masker sentences (“reversed”) on a word-by-word basis, and (c) the spatial separation of the masker sentences from the target sentence (“spatial”). Each of these four conditions also was subjected to ITFS signal processing (e.g., Wang, 2005; Brungart *et al.*, 2006) that was used to exclude T-F units in which masker energy was greater than the target energy (i.e., the LC was 0 dB). The ITFS processing was intended to isolate the available target “glimpses” contributing to intelligibility after the masker-dominated T-F units were removed. After processing, the remaining T-F units were reassembled and presented under the same four conditions as above to measure speech intelligibility. A more detailed description of this procedure is given in Sec. II D. This full set of conditions produced a total of 16 combinations [i.e., 2 signal processing conditions (natural or glimpsed) $\times$ 2 number-of-masker conditions (2 or 4) $\times$ 4 masker type conditions (baseline, male, reversed, spatial)].

During testing subjects were seated in individual double-walled sound-treated IAC (Industrial Acoustics Corporation, North Aurora, IL) booths and listened through Sennheiser HD280 Pro headphones (Sennheiser, Inc., Wedemark, Germany). For each condition a psychometric function was measured in which the target level was fixed at 55 dB sound pressure level and the masker level was varied to test 6 target-to-masker ratios (T/Ms) per condition. The sound levels were specified prior to the ITFS processing. The T/Ms were selected separately for each condition based on preliminary listening. Note that a T/M of 0 dB indicates that all talkers (target and each masker) were the same rms level individually. Thus, the expected “signal-to-noise ratio” (S/N) would be specified as -3 dB for the 2-talker masker and -6 dB for the 4-talker masker.

The experimental session began with a 30-trial block of the target sentences alone. This served the purpose of familiarizing the listeners with the sentence structure and response mode and as a check on the intelligibility of the target in quiet at the planned presentation level. All subjects scored 100% correct on this first block of trials. The next step in the familiarization stage consisted of a written description of the experimental conditions with three example trials presented at descending T/Ms for each condition. No criterion was set for performance in these practice/example trials.

The experimental test blocks began immediately following the quiet testing and familiarization. Each 30-trial block that followed consisted of one of the 16 experimental conditions with a mixed ordering of 5 trials at each of the 6 T/Ms. The order of conditions tested was randomized in each 16-block set (one block of each condition) and 4 sets of blocks were completed by each subject. This resulted in 80 scored words (4 scored words per sentence $\times$ 5 trials per block $\times$ 4 sets of blocks) at each T/M for each condition. Correct-answer feedback was given on all quiet, familiarization, and experimental trials. The entire experiment required 3–4 sessions of approximately two hours each including rest breaks.

D. Signal processing/stimulus generation

The target and masker sentences for all conditions were pregenerated and stored for playback during the experiment at a sampling rate of 44.1 kHz. All stimuli were convolved with impulse responses that had been recorded in our laboratory via the KEMAR manikin for the source positions (at a distance of 5 ft from the listener) used in the experiment. Therefore, a version of the target and maskers was available separately for each ear. The digital waveforms were D/A converted through an RME HDSP 9632 (ASIO) 24-bit sound card (Audio AG, Haimhausen, Germany). All signal processing and experimental control was implemented via MATLAB (MathWorks, Inc., Natick, MA).

All of the targets and the maskers, except for the maskers in the spatial condition, were processed for the source position of 0° azimuth. In the spatially separated condition, the masker placement was symmetric around the target. When there were two masker talkers the source positions were $\pm 90^\circ$ azimuth (to the left and right of the subject facing straight ahead) and when there were four masker talkers they were positioned at $\pm 45^\circ$ and $\pm 90^\circ$. Next, the masker sentences were scaled to produce the desired T/M and summed for each ear. The stimuli for each ear were then subjected to the ITFS analysis which resulted in a matrix of T-F values representing the energy in each of those units. According to that algorithm, a 1 represents each unit containing target energy equal to or greater than the masker energy while a 0 is used for the remaining masker-dominated T-F units. For the natural speech conditions, all T-F units were retained in the resynthesis for both target and summed masker (an all 1's binary mask; e.g., Wang, 2005). For the glimpsed conditions, an LC value of 0 dB was applied such that the target energy was compared to the summed masker energy for each T-F unit and each ear separately. Of the summed target and maskers for each ear, only the units with a value of 1 were retained while the units with a 0 were discarded. The synthesis was completed following application of the binary mask and the left ear and right ear signals were stored for playback. It should be noted that in the colocated conditions the targets and maskers all originated from 0° azimuth but still could have slightly different resynthesized waveforms for the two ears due to small asymmetries in the KEMAR recordings.

In order to keep the processing the same in all conditions, the analysis and resynthesis required to produce

the glimpsed conditions also were applied to the “natural” conditions. The ITFS analysis was the same as that used by Brungart *et al.* (2006, 2009) in which 128 frequency channels were analyzed with 20-ms windows (sequential windows overlapping by 10 ms). In our adaptation of that method, the frequency channels spanned 80 to 8000 Hz.

Figure 2 illustrates the result of ITFS processing for a single target sentence in two speech maskers (left column) and the same sentence in two noise maskers (right column) at three T/Ms (rows). The black areas indicate the T-F units *retained* in the ITFS processing. Thus, this is a plot showing the binary masks and is not a plot of the spectrograms. The words comprising the target and maskers were the same in all panels (the envelopes of the masker words modulated the speech-shaped noise maskers). The noise maskers that were used as a high-EM control for the 2-talker masker conditions were affected somewhat differently than the speech maskers with respect to the consequences of ITFS processing. Most notably, each successive reduction in the T/M excludes a greater number of target-dominated T-F units for noise than does the corresponding reductions in T/M for speech (compare left and right columns).

FIG. 2. Plots of the time-frequency (T-F) units retained (i.e., the binary masks; denoted by black regions) after applying ideal time-frequency segregation (ITFS) to the masked target stimuli. The left column shows representative masks for the target in speech maskers while the right column shows representative masks for the target in noise maskers. The same target words were used for all cases and the same pair of masker word sequences was used in the left column and the envelopes of those same words were used to modulate the noise maskers in the right column (see text). Each row contains stimuli processed at a specified target-to-masker ratio decreasing as indicated to the right of the figure.

III. RESULTS

In the following description of the results of the present study, the corresponding findings from Kidd *et al.* (2016) for NH listeners also are shown for comparison. Furthermore, the results from the two studies were combined in some of the statistical analyses. With the exception of the noise condition in which 8 new NH listeners (plus 2 who had participated in the earlier study) were recruited, the label “NH” refers to the subjects in the Kidd *et al.* (2016) study.

Table I contains the individual and group mean T/Ms at threshold for the SNHL listeners for both natural and glimpsed conditions for the 2- and 4-talker maskers. The thresholds were obtained from the fitted psychometric functions (least-squares fits to logistic functions) for 50% correct identification performance. The corresponding group mean values for the NH listeners from Kidd *et al.* (2016, p. 136) also are shown in Table I.

The results (excluding the noise masker) were evaluated with a repeated-measures analysis of variance (ANOVA) consisting of the within-subjects factors of *number* of maskers (2 vs 4), *processing* condition (natural vs glimpsed) and *masker* type (baseline, male, reversed, and spatial) and the between-subjects factor of *group* (NH vs SNHL). Shapiro-Wilk tests of normality and visual inspection supported the use of ANOVA. All of the within-subjects factors *number* $F(1,14) = 416.2$, *processing* $F(1,14) = 668.3$, and *masker* $F(1.9,26.3) = 83.8$ were significant at the $p < 0.001$ level (the Greenhouse-Geisser adjustment of degrees of freedom was used to correct for the only violation of sphericity, based on Mauchly’s test, for the factor of *masker type*). Likewise, the between-subjects factor of *group* also was significant $F(1,14) = 11.7$, $p = 0.004$. All three of the 2-way interactions and the 3-way interaction among the within-subjects factors were significant at the $p < 0.05$ level. These within-subjects trends are considered later. Furthermore, all of the 2-way, 3-way, and the 4-way interactions with *group* were significant with the exception of the 3-way interaction of *group* $\times$ *number* $\times$ *masker*.

A. Comparisons across groups: Natural speech and noise

The group mean T/Ms at threshold and the associated standard errors of the means for the 2- and 4-talker masker conditions are plotted in the upper row of Fig. 3. The noise masker condition also is plotted in the upper left panel and the thresholds in noise for the individual NH and SNHL listeners are given in Table II. The abscissa in Fig. 3 is masker condition while the ordinate is threshold T/M in dB. The thresholds for the SNHL group are plotted as filled/black bars while the thresholds for the NH subjects are plotted as open/white bars. Except as noted, this convention is used in subsequent figures as well. The differences in thresholds between groups are plotted on the ordinate in the lower row with the 2-talker and noise results in the lower left panel and the 4-talker results in the lower right panel.

First, for the natural 2-talker masker (upper left panel), thresholds were near 0 dB T/M in the baseline condition for both groups. The thresholds for the three speech masking

TABLE I. Individual target-to-masker ratios in dB at threshold for the 2- and 4-talker maskers and group means with standard errors of the means for the ten listeners with SNHL. The thresholds for both glimpsed and natural stimuli are given for all four masker conditions. Also, the group means and standard errors of the means from the six listeners with NH from [Kidd et al. \(2016\)](#) are given below the SNHL data.

	Glimpsed				Natural			
	Baseline	Male	Reversed	Spatial	Baseline	Male	Reversed	Spatial
2-talker masker								
1	-24.4	-28.5	-24.6	-21.8	-1.3	-11.6	-5.8	-7.7
2	-23.3	-26.5	-24.0	-20.9	4.4	-10.0	4.3	-7.5
3	-28.4	-28.3	-22.9	-24.5	1.7	-10.6	-1.7	-3.2
4	-28.8	-26.9	-25.2	-23.7	-2.1	-16.5	-9.1	-12.0
5	-26.0	-28.1	-24.0	-23.5	-0.6	-17.5	-13.0	-19.1
6	-27.9	-32.7	-31.8	-32.1	-3.1	-19.6	-16.8	-24.4
7	-23.9	-21.2	-22.0	-19.3	2.7	-5.8	-0.2	1.3
8	-27.3	-24.4	-23.8	-27.0	1.3	-9.3	-7.6	-5.2
9	-21.0	-23.7	-20.5	-20.5	2.4	-0.8	2.8	3.9
10	-27.5	-32.4	-26.1	-26.7	1.4	-17.4	-2.2	-11.2
Mean	-25.8	-27.3	-24.5	-24.0	0.7	-11.9	-4.9	-8.5
se	0.8	1.1	1.0	1.2	0.8	1.9	2.1	2.7
NH Mean	-30.6	-28.7	-29.6	-29.4	-0.4	-21.8	-17.1	-19.6
NH se	0.8	0.9	1.2	0.8	1.1	1.9	1.9	1.1
4-talker masker								
1	-16.3	-20.8	-14.7	-15.8	2.4	-4.7	-1.2	-6.0
2	-16.0	-17.4	-14.7	-12.6	6.0	-0.2	3.7	-2.6
3	-16.8	-16.5	-18.2	-14.4	3.7	-3.9	1.2	0.4
4	-17.3	-16.6	-16.2	-17.2	2.9	-13.5	0.4	-7.0
5	-15.3	-15.1	-16.0	-14.5	3.1	-7.1	-2.0	-5.4
6	-20.7	-22.3	-19.7	-22.0	2.8	-9.1	-3.1	-12.2
7	-13.7	-15.3	-12.3	-13.2	4.7	0.0	2.2	1.5
8	-15.6	-15.2	-15.6	-16.6	4.6	-6.6	-1.9	-0.2
9	-12.6	-11.6	-14.3	-12.7	5.9	2.1	4.4	6.9
10	-19.6	-19.8	-17.6	-18.6	2.8	-5.7	-1.3	-3.6
Mean	-16.4	-17.1	-15.9	-15.8	3.9	-4.9	0.2	-2.8
se	0.8	1.0	0.7	0.9	0.4	1.5	0.8	1.7
NH Mean	-18.7	-20.4	-19.8	-19.9	3.3	-10.7	-4.5	-9.1
NH se	0.8	0.8	1.0	0.7	0.6	0.7	1.2	1.1

FIG. 3. The upper two panels show group mean T/Ms at threshold in dB with standard errors of the means for NH (from [Kidd et al., 2016](#)) and SNHL listeners for the natural speech stimuli. The left panel contains the results from the 2-talker masker while the right panel shows the results from the 4-talker masker. The lower row of panels shows the difference in thresholds between NH and SNHL groups for the 2- and 4-talker masker conditions. The results from the noise masker condition are shown in the left column only.

TABLE II. Target-to-masker ratios at threshold for the natural and glimpsed noise masker for individual SNHL and NH listeners. The group means and standard errors of the means also are given.

Subject Number	SNHL		NH	
	Glimpsed	Natural	Glimpsed	Natural
1	-18.8	-10.9	-17.4	-16.9
2			-18.2	-16.1
3	-12.0	-10.0	-17.4	-12.8
4			-19.1	-16.0
5	-16.0	-10.7	-18.4	-10.0
6	-17.4	-16.5	-20.6	-12.9
7			-18.0	-14.8
8	-18.2	-9.3	-13.5	-8.3
9	-12.3	-3.5		
10				
11	-14.5	-5.4		
12	-16.1	-6.3		
Mean	-15.6	-9.1	-17.8	-13.5
se	0.9	1.4	0.7	1.1

release conditions (male, reversed, and spatial) were lower than in the baseline condition in all cases and, for both groups, the order of thresholds was the same with the lowest thresholds obtained for the male condition followed by the spatial condition and then the reversed condition. For the masking release conditions, the thresholds for the SNHL group were much higher than those for the NH listeners. In terms of the differences in thresholds between the two groups (lower left panel), the NH thresholds were all lower than the corresponding thresholds for the SNHL group with the average difference between groups for the three masking release conditions being about 11.1 dB (−8.4 vs −19.5 dB).

For the noise maskers (2 maskers “baseline” only; upper left panel of Fig. 3), the group mean thresholds were lower than baseline for both groups but were higher than the average thresholds for the three speech masking release conditions. However, it should be noted that these were different

listener groups for speech and noise conditions as described above. Here too the SNHL thresholds were higher than the NH thresholds with the difference between groups being about 4.4 dB (lower left panel of Fig. 3; Table II).

The group mean T/Ms at threshold and the associated standard errors for the 4-talker masker conditions are plotted in the upper right panel of Fig. 3. Group mean thresholds for SNHL and NH groups in the baseline condition were 3.9 and 3.3 dB, respectively. For the three speech masking release conditions (noise was not tested), the group-mean thresholds were lower than baseline in all cases. For both groups, the rank order of thresholds was the same as was noted above for the 2-talker masker with the lowest thresholds obtained for the male condition followed by the spatial condition and then the reversed condition. The average of the thresholds for the masking release conditions was about −2.5 for the SNHL listeners and −8.1 dB for the NH listeners. The differences in thresholds between NH and SNHL groups for these conditions are plotted in the lower right panel of Fig. 3. As was the case for the 2-talker masker, the difference between groups for baseline was small—less than 1 dB—with larger differences between groups found for the three masking release conditions. However, the group mean thresholds differed less for the 4-talker masker than for the 2-talker masker with the average difference for the masking release conditions only being about 5.6 dB for the 4-talker masker compared with just over 11 dB for the 2-talker masker.

B. Comparisons across groups: Glimpsed speech and noise

The group mean thresholds for the 2-talker masker in the glimpsed conditions are plotted in the upper left panel of Fig. 4. As expected, the thresholds were much lower than those shown for the corresponding conditions for the natural speech stimuli in Fig. 3. Here, the differences across masker type were much less than for natural speech with the group mean threshold for baseline and the average threshold of the three masking release conditions roughly being the same.

FIG. 4. Same as Fig. 3 but for the glimpsed stimulus conditions. The upper two panels show group mean T/Ms at threshold in dB with standard errors of the means for NH (from Kidd *et al.*, 2016) and SNHL listeners for the 2-talker (left) and 4-talker (right) maskers. The lower row of panels shows the difference in thresholds between NH and SNHL groups for the 2- and 4-talker masker glimpsed conditions. The results from the noise masker condition are shown in the left column only.

All of the thresholds for the SNHL group were higher than the corresponding thresholds for the NH group. For the glimpsed noise masker, the group mean thresholds were higher for both groups than all of the speech masker cases. Furthermore, the difference between noise thresholds for the SNHL and NH groups was relatively small at about 2 dB. The implications of the differences between glimpsed noise and glimpsed speech thresholds are considered later.

The thresholds for the glimpsed 4-talker masker conditions are shown in the upper right panel of Fig. 4. In general, the trends noted above for the glimpsed 2-talker masker were also found for the 4-talker masker. In all cases the glimpsed thresholds were much lower than the corresponding thresholds for the natural stimuli. And, as was the case for the 2-talker masker, the thresholds varied little across masker types with baseline and the three masking release conditions being roughly equal. However, all of the thresholds for the glimpsed 4-talker masker conditions were elevated relative to the corresponding thresholds for the 2-talker masker (compare upper left and right panels of Fig. 4).

The differences between groups - the extent to which the NH thresholds are lower than the SNHL thresholds—are shown for the glimpsed 2-talker condition in the lower left panel of Fig. 4. With the exception of baseline, all of the group differences shown in the lower left panel of Fig. 4 were smaller than the corresponding group differences found for the natural speech stimuli in the lower left panel of Fig. 3. In the lower right panel, the differences between groups for the glimpsed 4-talker masker stimuli are shown. These group differences again were smaller than were found for the natural stimuli—with the exception of baseline—and were roughly the same as the group differences for the glimpsed 2-talker masker shown in the lower left panel.

C. Derived measures: Masking release

Two comparisons derived from the thresholds shown in Figs. 3 and 4—release from masking (the reduction in

threshold relative to the threshold in baseline condition for the natural masker) and additional masking (the increase in threshold for the natural stimuli relative to the corresponding threshold for the glimpsed stimuli)—are plotted in Fig. 5 for each masker condition. The upper row is for release from masking and the lower row is for additional masking. The values for the 2-talker masker are plotted in the left panels and the values for the 4-talker masker are plotted in the right panels. As in Figs. 3 and 4, the values for the SNHL listeners are indicated by filled bars while the values for the NH listeners are indicated by the open bars. First, in all cases, the release from masking was less for the SNHL listeners than for the NH listeners and the release from masking found for the 2-talker masker generally was greater than that found for the 4-talker masker. The largest release from masking for either number of maskers occurred for the male masker condition where the NH listeners exhibited more than 20 dB of release from masking for the 2-talker masker while the SNHL listeners exhibited about 12 dB of release. Those numbers declined to about 14 and 8 dB, respectively, for the 4-talker masker. The smallest release from masking was found for the reversed condition for both groups and both numbers of maskers. The release from masking computed for noise used the same baseline 2-talker speech masker as a reference. This comparison was included because it provides an estimate of the release from IM due to substitution of equal-level noise maskers for the speech maskers—and was 9.8 dB for SNHL and 13.1 dB for NH.

D. Derived measures: Additional masking

The computation of additional masking used here (specified in dB change in thresholds) is conceptually the same as that reported by Brungart *et al.* (2006). Additional masking was interpreted by them as an estimate of the amount of IM that is present in a speech masking task. For baseline, the additional masking was more than 25 dB for both groups for the 2-talker masker (Fig. 5, lower left panel)

FIG. 5. This figure contains two sets of values derived from the group mean T/Ms at threshold shown in Figs. 3 and 4. In the upper row of panels, the release from masking (computed relative to the threshold in the baseline condition) is plotted for both NH (open/white bars) and SNHL (filled/black bars) groups for the 2-talker masker (upper left panel; and noise) and for the 4-talker masker (upper right panel). In the lower row, the additional masking (increase in thresholds produced by the maskers in the natural conditions relative to the corresponding thresholds in the glimpsed conditions) is plotted for both groups. The values for the 2-talker masker and noise are shown in the lower left panel and the values for the 4-talker masker are shown in the lower right panel. The error bars indicate ± 1 S.E. of the differences computed across individual listeners.

and more than 20 dB for the 4-talker masker (Fig. 5, lower right panel). By way of comparison, the SNHL listeners exhibited about 2–4 dB less additional masking than did the NH listeners for those conditions. For the three masking release conditions, the greatest additional masking was found for the reversed masker and the least for the male masker. Importantly, the additional masking was *greater* for the SNHL listeners than for the NH listeners for all maskers other than baseline. This was the case even though the group mean glimpsed thresholds were higher for the SNHL listeners than for the NH listeners (see Fig. 4). For the SNHL group, the average additional masking computed across the three masking release conditions was 16.8 and 13.8 dB for the 2- and 4-talker maskers, respectively. For the NH that pattern was reversed with less additional masking found for the 2-talker masker (9.6 dB) than for the 4-talker masker (11.9 dB). It also is of interest to note that the additional masking was lower for the noise masker condition than for the speech masker conditions at 6.6 dB for SNHL and 4.4 dB for NH. This result is consistent with the view that Gaussian noise causes predominantly EM and is close to the range of values for NH listeners (2–4 dB) reported by Brungart *et al.* (2006).

E. Comparison of 2- and 4-talker maskers

As noted above, the thresholds for the 4-talker masker were higher than the corresponding thresholds for the 2-talker masker in all cases (cf. Brungart *et al.*, 2009; Yost, 2017). This trend is apparent from the comparisons of thresholds plotted in Fig. 6 and is supported by the statistical analysis described previously. The abscissa is masker type

FIG. 6. The increase in group mean thresholds (T/M in dB) for the 4-talker masker conditions relative to the 2-talker masker conditions for natural (upper panel) and glimpsed (lower panel) stimuli. The abscissa is the masker condition and the open and filled bars show values for NH and SNHL listeners, respectively.

while the ordinate is the 2-talker masker threshold subtracted from the 4-talker masker threshold. The upper panel is for the natural speech stimulus while the lower panel is for the glimpsed speech stimulus. Note that the T/Ms at threshold are based on the levels per masker and thus the expected overall S/N is 3 dB higher for the 4-talker masker at any given T/M. For the natural speech conditions (upper panel), the increases in thresholds for the 4-talker maskers relative to the 2-talker maskers were all smaller for the SNHL listeners than for the NH listeners and were less for the baseline condition than for the three masking release conditions. In the extreme case, the threshold for the 4-talker reversed speech masker was only about 5 dB higher than for the 2-talker masker for SNHL listeners but was almost 13 dB higher for NH listeners. As discussed later, these differences are related to the greater EM inherent to the 4-talker masker than to the 2-talker masker as evidenced by the higher proportions of T-F units dominated by the 4-talker maskers at equal S/Ns (see Fig. 6, p. 141, in Kidd *et al.*, 2016).

For the glimpsed speech conditions (lower panel), the differences between NH and SNHL groups were smaller than the corresponding differences between groups for the natural speech conditions except for the baseline condition. Also, there was less variation across masker types with all of the group mean differences for the glimpsed stimuli falling in the range of about 8–12 dB. No consistent pattern was noted for these group differences either across listener groups or, within groups, across masker types.

F. Individual differences

As with the group mean results, the individual data obtained from the current study are combined with those from Kidd *et al.* (2016) so that comparisons between SNHL and NH listeners may be made more directly. The relationship between thresholds observed under the three masking release conditions for the natural speech stimuli are shown for individual subjects for both the 2- and 4-talker maskers in Fig. 7. The left column shows the results from the 2-talker masker while the right column shows the results from the 4-talker masker. Each row is a plot of the T/Ms at threshold from a different pair of the three masking release variables. The open symbols are for the six NH listeners from Kidd *et al.* (2016) while the filled symbols are thresholds from the ten SNHL listeners tested here. The slopes of the straight-line fits and the associated correlation coefficients for each group of listeners are given in Table III for both the 2-talker and 4-talker maskers. In general, the positive correlations among the thresholds for the different masking release variables indicated that the SNHL listeners, as was also the case for the NH listeners from Kidd *et al.*, tended to perform similarly, relative to the members of their group of listeners, with respect to how well they could use *any* of the three source segregation cues provided. That is, within a group, listeners who were successful at using one masking release variable to overcome SOS masking tended to also be successful using another masking release variable and the converse was true for the poorer performing listeners. In terms of the correlations reported in Table III, all of the values were statistically

FIG. 7. Individual thresholds for both SNHL and NH listeners are plotted for the three masking release conditions for natural speech. The left column shows the results from the 2-talker masker while the right column shows the results from the 4-talker masker. Each row shows the thresholds for a different pair of maskers: reversed by male (upper), spatial by male (center), and spatial by reversed (lower). Linear least-squares fits to the data for NH and SNHL subjects are shown separately in each panel while the associated correlation coefficients and slopes are given in Table III.

significant at the $p < 0.05$ level with the exception of the 4-talker reversed by spatial case for NH. Although we do not fully understand the origins of the consistent individual differences across NH listeners, the similar correlations among

TABLE III. Correlation coefficients and slopes of the linear least-squares fits to the data shown in Fig. 7. The values were computed for all three pairs of the “masking release variables” for the natural speech stimuli: male, reversed and spatial.

	Male vs reversed		Male vs spatial		Reversed vs spatial	
	Pearson r	Slope	Pearson r	Slope	Pearson r	Slope
2 maskers						
NH	0.83 ^a	0.84	0.91 ^a	0.53	0.95 ^a	0.55
SNHL	0.75 ^a	0.86	0.93 ^a	1.36	0.83 ^a	1.06
4 maskers						
NH	0.74 ^a	1.24	0.96 ^a	1.41	0.62	0.54
SNHL	0.72 ^a	0.39	0.76 ^a	0.86	0.72 ^a	1.50

^aStatistically significant values of the Pearson r ($p < 0.05$).

masking release variables for the NH and SNHL listeners warrant consideration of the extent to which the underlying mechanisms are the same. The factor that uniquely applies to the SNHL group, of course, is the presence and magnitude of their hearing loss—as reflected in audiometric thresholds—which could contribute to the high positive correlations.

Figure 8 shows the thresholds for the 2-talker masker conditions on the ordinate plotted as a function of the 4-frequency (500, 1000, 2000, and 4000 Hz) audiometric average along the abscissa. The thresholds for the SNHL listeners are plotted as triangles while the thresholds for the NH listeners are plotted as circles. Note that, in this figure, the open symbols are for the natural speech conditions while the filled symbols are for the glimpsed speech conditions. Each panel contains the results for both natural and glimpsed stimuli from a particular masker condition (clockwise from upper left): baseline, male, reversed, and spatial. Also shown in the baseline panel are the data from the noise masker condition for both natural and glimpsed conditions. Note again that a different group of NH listeners was tested for the noise condition and that 2 of the SNHL listeners (out of 8 total) also were new and in this figure are shown only for the noise condition. The thresholds for the SNHL listeners were fit with lines (linear least-squares fits) separately for glimpsed and natural stimulus conditions in each panel. The NH thresholds are to the left of the SNHL data in each panel. Because the pattern of results for the 4-talker masker was very similar to that

FIG. 8. Thresholds for individual listeners in each of the 2-talker masker conditions (clockwise from upper left): baseline, male, spatial, and reversed for the 2-talker maskers plotted as a function of hearing loss (average of audiometric thresholds measured at 500, 1000, 2000, and 4000 Hz). The triangles are for the SNHL listeners while the circles are for NH listeners from Kidd *et al.* (2016). Open symbols indicate natural speech conditions while black symbols indicate glimpsed speech conditions. The lighter shaded/gray symbols plotted in the upper left panel are for the noise maskers. Linear least-squares fits to the SNHL thresholds are shown in each panel and the corresponding correlation coefficients and slopes are given in Table IV.

for the 2-talker masker, those data were not displayed. However, the slopes and correlation coefficients of the straight line fits to the SNHL results—both for the 2-talker and noise maskers shown in Fig. 8 and for the 4-talker masker (not shown)—are given in Table IV.

There are several features of the results contained in Fig. 8 and Table IV that are noteworthy. First, the positive slopes of the straight-line fits mean that thresholds increased with increasing hearing loss. This was true for every condition tested for both 2- and 4-talker maskers and for noise. The asterisks beside the correlations shown in Table IV indicate that all of the correlations except for two—one of which was baseline—were statistically significant. Second, there were systematic differences in slopes across conditions. The slopes were steepest for the natural speech masking release conditions where they ranged from 0.29 (male) to 0.43 (spatial). For baseline, the slopes were much shallower at 0.1 and 0.05 for 2- and 4-talker maskers, respectively. In contrast, the corresponding slopes for the glimpsed functions also were shallow and fell within a narrow range from 0.14 (male) to 0.17 (spatial). A similar trend—steeper slopes for the natural speech masking release conditions—was found for the 4-talker masker but was less pronounced. In that case, the slopes for the natural speech masking release conditions ranged from 0.12 (reversed) to 0.29 (spatial) while the corresponding ranges for glimpsed conditions were from 0.06 (reversed) to 0.15 (spatial). Third, the slopes in the noise conditions were intermediate to the baseline and masking release conditions for the unprocessed stimuli (0.2) and were similar to the other conditions for glimpsed (0.15).

As an aside, the same correlational analyses for the results from the SNHL group were obtained with age replacing pure-tone average as a variable. In all cases, except the 4-talker masker reversed speech glimpsed condition ($r = -0.66$, $p = 0.04$), none of the correlations reached statistical significance (even without applying a correction for multiple tests). In fact, for all conditions other than the glimpsed noise condition, the computed correlations were negative.

TABLE IV. Correlation coefficients and slopes of the linear least-squares fits to the data shown in Fig. 8. The lines fit the individual SNHL listener thresholds plotted as a function of the 4-frequency average thresholds (i.e., degree of hearing loss).

	Glimpsed		Natural	
	Pearson r	Slope	Pearson r	Slope
2 maskers				
Baseline	0.55	0.09	0.67 ^a	0.10
Male	0.59 ^a	0.14	0.77 ^a	0.29
Reversed	0.83 ^a	0.16	0.69 ^a	0.30
Spatial	0.69 ^a	0.17	0.78 ^a	0.43
Noise	0.89 ^a	0.15	0.75 ^a	0.20
4 maskers				
Baseline	0.74 ^a	0.12	0.59 ^a	0.05
Male	0.70 ^a	0.14	0.77 ^a	0.23
Reversed	0.47	0.06	0.73 ^a	0.12
Spatial	0.81 ^a	0.15	0.85 ^a	0.29

^aStatistically significant values of the Pearson r ($p < 0.05$).

IV. DISCUSSION

A. Summary of the differences in performance between SNHL and NH listeners

In all cases, both for the natural speech stimuli and the glimpsed speech stimuli, the group mean thresholds, as computed by extracting equal performance points on the psychometric functions, were higher (poorer) for the SNHL listeners than for the NH listeners (Figs. 3 and 4). The largest differences between groups were found for the three masking release conditions (male, reversed, and spatial) for the natural speech 2-talker maskers. Thus, our findings are consistent with the conclusion that these large group differences indicate a reduced ability of the SNHL listeners to exploit the three source segregation cues that were available to reduce IM under those conditions. It should be pointed out, though, that the trends found in the results were very similar for the NH and SNHL listeners; for example, the order of group mean thresholds for the different speech masker conditions was the same and, as discussed further below, the correlations of thresholds across cue variables for individual subjects within groups were similar. The same pattern of results for the group means was found for the 4-talker maskers as was found for the 2-talker maskers (Table I) but the differences—both between groups and across masker conditions within groups—generally were less pronounced, likely because performance was influenced more by EM for the 4-talker masker (discussed further below). The individual differences were prominent for the higher IM speech masker conditions but were reduced for the higher EM noise masker control.

1. Release from masking

A common metric of the strength of a segregation cue in multiple source masking experiments is release from masking. In these experiments release from masking was computed by taking the difference between the thresholds for the baseline condition and the three speech masking release conditions and noise. By this metric, release from masking also was reduced for the SNHL group relative to the NH group in every case for both the 2- and 4-talker maskers (upper row, Fig. 5). Again, the group mean differences were greater for the 2-talker masker than for the 4-talker masker. This is consistent with the interpretation that the SNHL listeners were less able to use source segregation cues in overcoming IM than were NH listeners. Because, in our view, there is more IM for the 2-talker masker than for the 4-talker masker, there is also more IM that may be released when a segregation cue is provided and the masking that is present is more likely to be dominated by IM when IM is high overall. Thus, the group differences in release from masking for the 2-talker masker are particularly revealing of the difficulties the SNHL listeners experienced under the high IM conditions. One complication in the interpretation of these estimates of release from masking, however, is that the glimpsed thresholds (in all conditions including baseline) were higher for the SNHL listeners as a group and, on an individual basis, increased as hearing loss increased (Fig. 8). The glimpsed

thresholds reflect performance after accounting for the T-F units dominated by EM and after eliminating IM. Thus, the T/Ms at threshold for the natural speech stimuli do not directly reflect how much masking is present in a given condition overall if the useable target energy as reflected by the glimpsed thresholds (i.e., the T-F units dominated by the target) differs across groups. The factors influencing glimpsed thresholds are considered further later.

A second complication with using the baseline condition as a reference for determining masking effectiveness/release is that there is an additional segregation cue that is available in the form of differences in level between the target and masker talkers. The influence of this cue, which has been discussed previously in several other studies (e.g., Brungart, 2001; Marrone *et al.*, 2008c; Iyer *et al.*, 2010; Best *et al.*, 2013b), depends on the relative level of the sounds and the number of sound sources in a complex way but clearly is most effective when the target talker is the highest level source. This means that once the target is higher in level than the competitors it increases in prominence/salience to the point where it can be easily selected from the mixture of talkers, creating a “ceiling” for the baseline thresholds that could effectively compress the range of masking release. Because the thresholds in the masking release conditions for the 4-talker masker were higher than for the 2-talker masker, especially for the SNHL listeners, the potential for the level cue to influence the estimate of masking release in that case is greater than for the 2-talker masker. This means that conditions in which thresholds exceed 0 dB T/M and that rely on such thresholds for computing differences must be interpreted with caution.

2. Additional/informational masking

Because the glimpsed stimulus comprises all of the information (within the constraints of the process discussed above) available to the listener to solve the speech identification task in the natural listening conditions, the limitations on target speech information due to a failure to fully compensate for the loss of sensitivity in SNHL presumably are taken into account when comparisons of performance are referenced to the glimpsed thresholds. This is relevant for any reduction in target information in the acoustic signal due to insufficient amplification in any of the conditions tested here because that reduction would be the same across all conditions. Thus, analysis of the results in terms of additional masking, which uses the glimpsed thresholds as a reference, is particularly useful for interpreting the present results. Also, the issue of range compression due to a relative level cue is greatly reduced because the thresholds in the masking release conditions—even for the SNHL listeners—tend to be at negative T/Ms where segregation by level is difficult. Based on the comparisons between SNHL and NH groups (Fig. 5), additional masking was greater in all cases for the SNHL listeners than for the NH listeners for the conditions in which explicit segregation cues were provided; i.e., the masking release conditions. Again, this trend was more pronounced for the 2-talker masker than for the 4-talker masker. These findings lead to the conclusion that IM was

greater for the SNHL listeners than for the NH listeners in the subset of the conditions tested here where source segregation cues were explicitly provided, largely because the SNHL listeners were less able to take advantage of these source segregation cues. This was much less apparent in the baseline condition where the source segregation cues were more limited and the ceiling effect due to relative level was more likely to influence the results. These trends were apparent for additional masking even though the glimpsed thresholds were higher for the SNHL group and increased as the amount of hearing loss increased for individual subjects (Fig. 8).

3. Energetic masking

With respect to the evidence obtained in the present study that EM contributed to the higher masked thresholds for the SNHL group than for the NH group, the most relevant data, perhaps, are those obtained using the noise masker. This is because the noise masker condition produces the highest proportion of EM of all of the conditions tested (e.g., review in Kidd *et al.*, 2008b; cf. related in Schubotz *et al.*, 2016; Culling and Stone, 2017). Although the thresholds for the SNHL listener group were higher than those for the NH listener group for the noise masker, consistent with the prediction of greater EM (e.g., Arbogast *et al.*, 2005; Best *et al.*, 2013a), the difference (4.4 dB) was considerably less than the difference between groups found for the speech masking conditions, especially the masking release conditions for the 2-talker masker (e.g., 12.2 dB for the reversed speech masker). Thus, to the extent that the thresholds obtained in the noise masker condition provide an accurate estimate of the relative EM in the two groups, the difference between groups is insufficient to account for the differences found in the speech masker conditions.

However, beyond the simple comparison of threshold T/Ms in the unprocessed noise there are other findings that are at least as relevant; in particular, the magnitude of the additional masking—reflecting the benefit of ITFS processing—for the noise masker. There are two reasons for the importance of additional masking in understanding the influence of EM: first, the glimpsed thresholds for the SNHL group were higher in noise than they were for the NH group. This group mean difference in glimpsed thresholds presumably is not a reflection of differences in EM but rather is due to other factors as discussed above.² Second, the possibility of spread of masking from masker-dominated frequency regions into target-dominated frequency regions in the unprocessed conditions is eliminated by ITFS processing. Therefore, if the SNHL group experienced greater spread of masking from masker dominated T-F units into adjacent target dominated T-F units than the NH group, it should be reflected by greater additional masking in noise. For the group mean data, that was indeed the case. However, the amount of additional masking for noise was only about 2 dB greater for the SNHL group than for the NH group (Fig. 5). These results were analyzed with a repeated measures ANOVA having the within-subjects factor of processing condition (natural vs glimpsed) and the between-subjects factor of group (SNHL vs NH). This analysis indicated that

both main factors, processing condition [$F(1,14)=49.9$, $p<0.001$] and group [$F(1,14)=56.4$, $p=0.024$], were significant while the interaction of processing condition by group was not [$F(1,14)=2.04$, $p=0.175$]. This relatively small effect (2 dB) together with the lack of a significant interaction in the results of the ANOVA, again are consistent with the view that the differences in EM between groups played only a minor role in the trends that were observed in the speech masking conditions. So, to the extent that the contribution of EM to the overall amount of masking may be estimated by the additional masking found for the noise masker conditions, differences in EM appear to be insufficient to account for the large group differences found in the natural speech masking release conditions.

B. Individual differences among the SNHL listeners

The high positive correlations among the thresholds for the three masking release variables found for the SNHL listeners—similar to the trends for the NH listeners from Kidd *et al.* (2016; replotted in Fig. 7)—are especially important because the benefits derived from the cues presumably depend on mechanisms that may be, to a large extent, independent. For example, the benefit from a talker sex difference likely is due primarily to discriminating differences in fundamental frequency and vocal tract length whereas exploiting the spatial separation of sources depends on utilizing interaural time and intensity differences arising from differences in source azimuth. Utilizing masker time reversal potentially is even more complex in that it is less directly related to changes in any simple stimulus property than it is to the negation of intelligibility, meaning that the explicit confusions sometimes associated with high IM SOS masking conditions (e.g., as revealed by an analysis of error patterns; e.g., Brungart *et al.*, 2001; Kidd *et al.*, 2005, 2016) largely are mitigated. The significance of the finding that thresholds were highly correlated among the masking release variables for the SNHL listeners is that, despite the fact that thresholds increase with degree of hearing loss in all cases (Fig. 8), it appears to be a reflection of a more general problem affecting multiple abilities rather than being strongly related to reduced resolution of any specific cue per se (see related discussion in Shinn-Cunningham and Best, 2008, and Best *et al.*, 2012, 2017). The similarity of the patterns of individual differences between NH and SNHL listeners, as indicated in Fig. 7, suggests that some portion of these differences in overcoming masking—likely overcoming IM—is *not* a consequence of hearing loss but rather is due to other aspects of auditory perception common to both NH and SNHL listeners that vary in a consistent way across individuals. It also cannot be ruled out that this common factor is not directly related to perceptual segregation per se but instead reflects another higher-level capability such as working memory or executive function (e.g., McCoy *et al.*, 2005; Clayton *et al.*, 2016; Roverud *et al.*, 2017).

With respect to individual differences within the SNHL group, it is of interest to ascertain whether the correlations among the thresholds found in the masking release conditions for natural speech (Fig. 7) also would be present after

taking into account the differences that exist between subjects for the glimpsed thresholds. There are two factors in this correlational analysis that are of interest: first, it is clear from Fig. 8 that thresholds for all of the natural speech masking release conditions increase as a function of hearing loss. It also is apparent that the glimpsed thresholds increase with hearing loss too, although the slopes of the functions are considerably steeper for the natural speech masking release conditions than the corresponding slopes for the same conditions for the glimpsed stimuli (e.g., Table IV). The steeper slopes imply that factors other than those responsible for the increase in glimpsed thresholds with increasing hearing loss (e.g., insufficient audibility of the high frequency glimpses) contribute to the poor performance on the natural speech stimuli in the masking release conditions, especially for the 2-talker masker. However, the question remains whether the high correlations observed for the results plotted in Fig. 7 also would be found after the variation in glimpsed thresholds was taken into account. The second point of interest is whether the additional masking observed for the three natural speech masking release conditions is predicted by the thresholds in noise on an individual subject basis. Although it was concluded above that differences in EM could not account for the large group mean differences in the additional masking observed in the masking release conditions, it still is possible that, within the SNHL group, the variation across individuals is driven by underlying differences in EM. If that were true we might expect that the thresholds obtained in the masking release conditions would be predicted by the thresholds found in noise.

Of the group of eight subjects with SNHL who took part in the noise masker conditions, six also had participated in the same conditions (i.e., masking release conditions for natural speech) shown in Fig. 7. For this subset of SNHL listeners it is possible to compute correlations of additional masking with the other conditions tested in the study. These data are plotted in Fig. 9. In this figure, the values that are plotted are additional masking (increase in thresholds for the natural speech condition above the thresholds for the glimpsed speech condition) for all three pairs of masker types among the male, spatial and reversed conditions. The left two columns are for the 2-talker masker and the 4-talker masker, respectively. The right column is for noise and the three masking release variables for the 2-talker masker.

The only statistically significant correlations in this set of comparisons (correlations given in each panel marked with asterisk where significant) were among pairs of the three masking release variables for the 2-talker masker. In those cases, the listeners roughly followed the pattern seen in the threshold T/Ms plotted in Fig. 7 with high correlations found among masking release variables. This result makes sense based on the conclusion that the subjects differed the most with respect to IM, as discussed above. Because additional masking is thought to be proportional to the amount of IM that is present (e.g., Brungart *et al.*, 2006; Kidd *et al.*, 2016) correlations based on the conditions highest in IM should reflect the differences between subjects more clearly than conditions where EM is stronger. No significant correlations were observed for comparisons that involved maskers

FIG. 9. Additional masking in dB for the natural speech conditions for individual SNHL listeners for each of the 2-talker masking release conditions (male, reversed and spatial) plotted one against the other (left column); additional masking for each of the 4-talker masking release conditions plotted one against the other (center column); and additional masking for the 2-talker masking release conditions plotted as a function of the additional masking in noise (right column). Correlation coefficients are indicated in the lower right of each panel with significant correlations ($p < 0.05$) denoted by asterisks.

that had a higher proportion of EM - noise and the 4-talker masker. It should be noted that these correlations were based only on the six available SNHL listeners and the generality of the trends should be confirmed on a larger group of subjects. Overall, it appears that the large, systematic differences between subjects apparent in the left column of Fig. 9 were due primarily to differences in IM rather than to differences in EM.

With respect to the consistency of individual differences in performance across conditions for NH listeners, which was based on the study by Kidd *et al.* (2016) (see also Durlach *et al.*, 2003b; Durlach *et al.*, 2005) with only six observers, there are some relevant findings from Swaminathan *et al.* (2015) who reached similar conclusions. Swaminathan *et al.* studied a larger group of subjects ($N = 24$) using an experimental paradigm that overlapped some of the conditions tested here including the same speech corpus and similar procedures. In their study, a high positive

correlation was found between the masking release obtained by masker time reversal and spatial separation of target and masker talkers. They were focused on the role of musical training in achieving release from SOS masking and in general the listener group with musical training exhibited lower thresholds and greater masking release than the non-musician group for both of these variables (see replot of Swaminathan *et al.* individual data in Kidd and Colburn, 2017, Figure 4.4, p. 90). Swaminathan *et al.* (2015) concluded that the group differences in performance were not due to differences in the peripheral mechanisms of their subjects but were related instead to musicianship, either in terms of differences in innate ability to segregate and attend to one talker/source in a mixture of talkers/sources and/or to musical training that fostered that ability. The findings of Swaminathan *et al.* (2015) are consistent with the conclusion by Kidd *et al.* (2016) that NH listeners may be categorized according to whether they generally are able to use various

types of cues to release IM rather than demonstrating facility with one type of cue but not another. The fact that Swaminathan *et al.* also found a difference due to musical training reinforces the view that these individual differences were related to central factors more than to peripheral factors (see also Clayton *et al.*, 2016). This same conclusion may be extended to the current findings with the caveat that the degree of SNHL also affected thresholds in all of the relevant cases confirming that peripheral hearing loss also contributes to the performance deficits found here.

C. Glimpsed thresholds

With respect to the findings that the glimpsed thresholds were higher for the SNHL listeners than for the NH listeners, one possibility, which is supported by recent studies of spatial release from masking in listeners with SNHL (e.g., Glyde *et al.*, 2015; Best *et al.*, 2017; Rana and Buchholz, 2018), is that the higher glimpsed thresholds were due to a lack of audible high-frequency information for the listeners with hearing loss. According to this interpretation, the NAL-RP gain algorithm applied here did not fully restore audibility for the SNHL listeners at the higher frequencies where hearing loss is greater (note that different amplification schemes, e.g., multiband compression, could exert a different effect on audibility than was observed here; cf. Dillon, 2012). This would cause some of the high-frequency glimpses to be inaudible and others to be less useful due to a low sensation level, a particular problem if those T-F units had high T/Ms. It is reasonable to expect that a reduction in the audibility of a portion of the information-bearing T-F units would require an increase in level to reach speech reception threshold. In the study by Best *et al.* (2017), larger group differences in glimpsed speech thresholds between NH and SNHL listeners were reported than those found here (they found differences of roughly 5 and 7 dB in their study for 4- and 2-talker maskers, respectively). One methodological difference between studies is that, in the present experiments, the target level was fixed and psychometric functions were generated by varying the level of the maskers whereas the converse was true for the Best *et al.* study. This may have increased the influence of reduced audibility in the Best *et al.* study, particularly at the lower T/Ms. However, the finding that the glimpsed thresholds in this study were higher for the SNHL group than for the NH group in all cases, and the general increase in glimpsed thresholds as hearing loss increased, were consistent with the reduced audibility hypothesis. It also is possible that listeners with SNHL are less able to synthesize the sparse glimpsed stimulus into a coherent stream of speech. Removing portions of the stimulus (i.e., the masker-dominated T-F units) alters the within-channel representation of the envelope by replacing segments of the target-masker mixture with gaps, even though it is assumed that the remaining glimpses contain all of the useful target information. It is possible that the higher glimpsed thresholds for the SNHL listeners could reflect difficulty processing the degraded envelopes comprising this type of stimulus.

Because ITFS processing is intended to emulate the effects of the EM that is present in any given masked condition while also concomitantly mitigating IM, a plausible hypothesis is that a given level of identification performance for a glimpsed target stimulus (e.g., threshold) would correspond to a constant proportion of glimpsed target energy retained regardless of the masker type. Some support for this constancy hypothesis was reported by Kidd *et al.* (2016) who found that the proportion of energy retained in the glimpses following ITFS processing was about the same for the 2-talker masker and the 4-talker masker at group mean thresholds. However, it is not clear whether this proposed constancy would hold for different types of maskers (e.g., speech vs noise) or for the different subject groups. If the SNHL listeners experienced difficulty with audibility of the target stimulus even after the application of the frequency-specific gain, we might expect that the proportion of energy needed to achieve threshold for the glimpsed stimulus would be greater for those listeners than for the NH listeners—to compensate for the inaudible high-frequency glimpses—and that this effect would increase as hearing loss increased. Figure 10 shows the proportion of target energy retained at threshold in the glimpsed stimulus at each individual listener's threshold plotted as a function of hearing levels (i.e., 4-frequency audiometric average) for speech and noise maskers. These values are for the baseline 2-talker speech masker condition and the noise masker conditions. Straight lines were fit to the SNHL values (linear least-squares fits) and both speech and noise functions were significantly correlated with the 4-frequency threshold averages. The slope of the line for the noise masker was more than twice the slope computed for the speech masker, implying that the factors underlying performance for the two glimpsed maskers may in part differ. A paired-comparisons t-test (speech vs noise) conducted for the data

FIG. 10. The proportion of energy retained for the glimpsed stimulus conditions at threshold for individual subjects plotted as a function of the 4-frequency audiometric average thresholds. The values for both NH and SNHL listeners are shown for speech (baseline) and noise maskers. The circles are values for NH subjects while the triangles fit with straight lines are for the SNHL subjects.

from the six subjects who were tested on both maskers was significant [$t = -4.57$, $df = 5$, $p = 0.006$] confirming that the group means were different for the two types of maskers. It is of interest too that the energy retained at glimpsed threshold for this young-adult SNHL population for either speech or noise maskers was not significantly correlated with age over the range of ages tested here.

The main points to be made from plotting the glimpsed data in these coordinates are that, first of all, a higher proportion of energy retained is needed to achieve threshold target identification performance for the noise masker condition than for the 2-talker speech masker condition. These findings violate the proposition that a fixed level of speech identification performance corresponds to a constant proportion of target energy retained and implies that the way that the masker energy is distributed (distribution of energy across frequency over time) affects performance even after performing ITFS processing. This observation holds even though the noise masker was shaped to match the long term average spectrum of the speech maskers used in this study and had broadband envelopes that matched the envelopes of the words used as speech maskers. This suggests that even speech-shaped speech-envelope modulated noise has some limitations as an energetic masking control for actual speech. The second point is that, within masker type, there was some overlap between the values for the NH and SNHL listeners (Fig. 10). The difference in glimpsed thresholds for noise and speech maskers was greater in general than the difference within masker type across subjects. Further studies of how the parameters of the ITFS processing (e.g., number of bands, temporal window duration) may interact with masker type (e.g., spectral density and distribution of masker levels within T-F units) and presence and degree of hearing loss appears to be an important next step in this line of research.

V. SUMMARY AND CONCLUSIONS

The findings from this SOS masking study indicate that listeners with sensorineural hearing loss—with linear amplification applied to account for audibility—exhibited more masking under all stimulus conditions than their normal-hearing counterparts. There was evidence that both EM and IM were greater for the SNHL listeners for these stimuli and conditions. The greatest difference, though, occurred for the high IM conditions where explicit source segregation cues were provided. In those cases, the SNHL listeners were less able to take advantage of the cues to overcome IM. So, to answer the first set of questions posed in the Introduction, SNHL listeners exhibited more IM than NH listeners under conditions where source segregation cues normally are effective. Furthermore, the SNHL listeners also exhibited greater EM than normal—for example they had higher thresholds in noise—but the magnitude of the effect was small relative to the results under the high IM conditions.

With regard to the second set of questions posed in the Introduction, the correlations among thresholds for the SNHL listeners on the three masking release variables were very similar to those found by Kidd *et al.* (2016) for NH listeners. Although this study was not intended to determine the mechanisms underlying the masking release caused by each

variable, the similarly high correlations among individuals in the two groups suggests that at least some of these differences are due to common underlying mechanisms that do not depend on the presence or degree of hearing loss. It is important to emphasize that the different segregation cues: different sex talkers, time reversal and spatial separation of sources, likely vary in the extent to which the combination of low-level and high-level processes foster perceptual segregation. Male vs female speech, for example, is distinguished primarily based on differences in fundamental frequency, which is often thought of as a relatively low-level stimulus variable. Separating time-reversed speech from natural speech may be facilitated by reliable spectrotemporal differences between sounds but also appears to involve linguistic processing (e.g., Kidd and Colburn, 2017). Although similarly strong correlations were found for SNHL listeners among these variables, one important difference relative to NH listeners is that performance varied with the degree of loss. This was apparent for the glimpsed stimuli, but was more pronounced for the natural speech masking release conditions—a finding indicating greater additional masking attributed to IM.

Finally, there is strong evidence that a given level of speech identification performance is not closely tied to a fixed/constant proportion of target energy available in the glimpsed stimulus. We found differences in the glimpsed thresholds across subject groups and noted that the glimpsed thresholds increased as hearing loss increased, particularly in noise. This effect likely is due in part to insufficient audibility. In that case, the SNHL listeners could miss the glimpses having low overall energy, but which retained sufficient target information (above the LC value) to contribute to intelligibility. This would be consistent with the findings of Glyde *et al.* (2015), Best *et al.* (2017), and Rana and Buchholz (2018), but was not directly examined here. Furthermore, the proportion of target energy at threshold varied considerably depending on the type of masker with higher proportions needed to achieve threshold for noise maskers than for speech. Undoubtedly, these results are related to the distribution of T-F units retained following ITFS analysis, and not simply the energy in the remaining glimpses, but this issue requires further study.

Overall, at least three factors appear to be present and to adversely affect the performance of the SNHL listeners in this task: reduced audibility and/or impaired/inefficient use of the available glimpses (as indicated by higher thresholds following ITFS processing); greater than normal EM, likely due to increased spread of excitation from masker-dominated spectral regions into adjacent regions; and increased IM due to reduced source segregation ability. Although there is evidence supporting the role of each of these factors taken from the various experimental conditions tested, all three—and others not explicitly tested or identified here—likely interact in natural listening situations to limit speech recognition performance of listeners with SNHL attempting to communicate in multiple-talker sound fields.

ACKNOWLEDGMENTS

The work is supported by NIH-NIDCD Grant Nos. DC004545, DC013286, DC00100, and AFOSR FA9550-16-

10372. The authors are grateful to Jan Rennie-Hochmuth and Christopher Conroy for comments on an earlier version of the manuscript. We also would like to thank Lorraine Delhorne for help in recruiting and testing some of the subjects.

¹As opposed to estimating the EM that is present in individual T-F units, the estimation of the contribution of IM from individual T-F units is not independent of the presence/properties of other T-F units; that is, the IM value of any specific T-F unit depends on the ensemble of T-F units (e.g., the masker could be intelligible speech or unintelligible time-reversed speech).

²The presence of noise within target dominated T-F units does not affect intelligibility as far as we have been able to ascertain for the conditions tested here. This conclusion is based on our own informal listening tests in which the exact same glimpsed stimulus was presented with the noise present in the target-dominated T-F units (after ITFS processing) versus when the noise was removed from those T-F units and only the “clean” glimpses of the target waveform were presented.

- Agus, T. L., Akeroyd, M. A., Noble, W., and Bhullar, N. (2009). “An analysis of the masking of speech by competing speech using self-report data (L),” *J. Acoust. Soc. Am.* **125**, 23–26.
- Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (2005). “The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners,” *J. Acoust. Soc. Am.* **117**, 2169–2180.
- Başkent, D., Clarke, J., Pals, C., Benard, M. R., Bhargava, P., Saija, J., Sarampalis, A., Wagner, A., and Gaudrain, E. (2016). “Cognitive compensation of speech perception with hearing impairment, cochlear implants, and aging: How and to what degree can it be achieved?,” *Trends Hear.* **20**, 1–16.
- Bernstein, J. G. W., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M., and Leek, M. R. (2013). “Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners,” *J. Am. Acad. Audiol.* **24**, 293–306.
- Best, V., Marrone, N., Mason, C. R., and Kidd, G., Jr. (2012). “The influence of non-spatial factors on measures of spatial release from masking,” *J. Acoust. Soc. Am.* **131**, 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**, 81–91.
- Best, V., Thompson, E., Mason, C. R., and Kidd, G., Jr. (2013a). “Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the spectral overlap of competing talkers,” *J. Acoust. Soc. Am.* **133**, 3677–3680.
- Best, V., Thompson, E. R., Mason, C. R., and Kidd, G., Jr. (2013b). “An energetic limit on spatial release from masking,” *J. Assoc. Res. Otolaryn.* **14**, 603–610.
- Brungart, D. S. (2001). “Informational and energetic masking effects in the perception of two simultaneous talkers,” *J. Acoust. Soc. Am.* **109**, 1101–1109.
- Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (2006). “Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation,” *J. Acoust. Soc. Am.* **120**, 4007–4018.
- Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (2009). “Multitalker speech perception with ideal time-frequency segregation: Effects of voice characteristics and number of talkers,” *J. Acoust. Soc. Am.* **125**, 4006–4022.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (2001). “Informational and energetic masking effects in the perception of multiple simultaneous talkers,” *J. Acoust. Soc. Am.* **110**, 2527–2538.
- Buss, E., Leibold, L. J., Porter, S. L., and Grose, J. H. (2017). “Speech recognition in one- and two-talker maskers in school-age children and adults: Development of perceptual masking and glimpsing,” *J. Acoust. Soc. Am.* **141**, 2650–2660.
- Byrne, D. J., Parkinson, A., and Newall, P. (1991). “Modified hearing aid selection procedures for severe-profound hearing losses,” in *The Vanderbilt Hearing Aid Report II*, edited by G. A. Studebaker, F. H. Bess, and L. B. Beck (York Press, Parkton, MD), pp. 295–300.
- Calandruccio, L., Brouwer, S., Van Engen, K., Dhar, S., and Bradlow, A. (2013). “Masking release due to linguistic and phonetic dissimilarity between the target and masker speech,” *Am. J. Audiol.* **22**, 157–164.
- Calandruccio, L., Buss, E., and Bowdrie, K. (2017). “Effectiveness of two-talker maskers that differ in talker congruity and perceptual similarity to the target speech,” *Trends Hear.* **21**, 1–14.
- Carney, L. H. (2018). “Supra-threshold hearing and fluctuation profiles: Implications for sensorineural and hidden hearing loss,” *J. Assoc. Res. Otolaryn.* **19**, 331–352.
- Chambers, A. R., Resnik, J., Yuan, Y., Whitton, J. P., Edge, A. S., Liberman, M. C., and Polley, D. B. (2016). “Central gain restores auditory processing following near-complete cochlear denervation,” *Neuron* **89**, 867–879.
- Clayton, K. K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A. D., and Kidd, G., Jr. (2016). “Executive function, visual attention and the cocktail party problem in musicians and non-musicians,” *PLoS One* **11**(7), e0157638.
- Cooke, M. (2006). “A glimpsing model of speech perception in noise,” *J. Acoust. Soc. Am.* **119**, 1562–1573.
- Culling, J. F., and Stone, M. A. (2017). “Energetic masking and masking release,” in *The Auditory System at the Cocktail Party*, edited by J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay (Springer Nature, New York), pp. 41–73.
- Dawes, P., Munro, K. J., Kalluri, S., and Edwards, B. (2013). “Unilateral and bilateral hearing aids, spatial release from masking and auditory acclimatization,” *J. Acoust. Soc. Am.* **134**, 596–606.
- Dillon, H. (2012). *Hearing Aids* (Boomerang Press, Turrumura).
- Durlach, N. I., Mason, C. R., Gallun, F. J., Shinn-Cunningham, B., Colburn, H. S., and Kidd, G., Jr. (2005). “Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers,” *J. Acoust. Soc. Am.* **118**, 2482–2497.
- Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003a). “Note on informational masking,” *J. Acoust. Soc. Am.* **113**, 2984–2987.
- Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. C., Arbogast, T. L., Colburn, H. S., and Kidd, G., Jr. (2003b). “Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity,” *J. Acoust. Soc. Am.* **114**, 368–379.
- Festen, J. M., and Plomp, R. (1990). “Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing,” *J. Acoust. Soc. Am.* **88**, 1725–1736.
- Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2004). “Effect of number of masker talkers and auditory priming on informational masking in speech recognition,” *J. Acoust. Soc. Am.* **115**, 2246–2256.
- Gallun, F. J., Kappel, S. D., Diedesch, A. C., and Jakien, K. M. (2013). “Independent impacts of age and hearing loss on spatial release in a complex auditory environment,” *Front. Neurosci.* **7**, 252.
- Gatehouse, S., and Noble, W. (2004). “The speech, spatial and qualities hearing scale (SSQ),” *Int. J. Audiol.* **43**, 85–99.
- 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**, 3311–3319.
- Healey, E. W., and Vasko, J. L. (2018). “An ideal quantized mask to increase intelligibility and quality of speech in noise,” *J. Acoust. Soc. Am.* **144**, 1392–1405.
- Henry, K. S., and Heinz, M. G. (2012). “Diminished temporal coding with sensorineural hearing loss emerges in background noise,” *Nat. Neurosci.* **15**, 1362–1364.
- Henry, K. S., Sushrut, K., and Heinz, M. G. (2016). “Distorted tonotopic coding of temporal envelope and fine structure with noise-induced hearing loss,” *J. Neurosci.* **36**, 2227–2237.
- Iyer, N., Brungart, D. S., and Simpson, B. D. (2010). “Effects of target-masker contextual similarity on the multimasker penalty in a three-talker diotic listening task,” *J. Acoust. Soc. Am.* **128**, 2998–3010.
- Jakien, K. M., S. D. Kappel, S. Y. Gordon, and Gallun, F. J. (2017). “The benefits of increased sensation level and bandwidth for spatial release from masking,” *Ear Hear.* **38**(1), e13–e21.
- Kidd, G., Jr., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (2005). “The advantage of knowing where to listen,” *J. Acoust. Soc. Am.* **118**, 3804–3815.
- Kidd, G., Jr., Arbogast, T. L., Mason, C. R., and Walsh, M. (2002a). “Informational masking in listeners with sensorineural hearing loss,” *J. Assoc. Res. Otolaryn.* **3**, 107–119.
- Kidd, G., Jr., Best, V., and Mason, C. R. (2008a). “Listening to every other word: Examining the strength of linkage variables in forming streams of speech,” *J. Acoust. Soc. Am.* **124**, 3793–3802.

- Kidd, G., Jr., 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 Nature, New York), pp. 75–109.
- Kidd, G., Jr., Mason, C. R., and Arbogast, T. L. (2002b). "Similarity, uncertainty and masking in the identification of nonspeech auditory patterns," *J. Acoust. Soc. Am.* **111**, 1367–1376.
- Kidd, G., Jr., Mason, C. R., and Best, V. (2014). "The role of syntax in maintaining the integrity of streams of speech," *J. Acoust. Soc. Am.* **135**, 766–777.
- Kidd, G., Jr., Mason, C. R., and Feth, L. L. (1984). "Temporal integration of forward masking in listeners having sensorineural hearing loss," *J. Acoust. Soc. Am.* **75**, 937–944.
- Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2008b). "Informational masking," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer Science+Business Media, New York), pp. 143–190.
- Kidd, G., Jr., Mason, C. R., Swaminathan, J., Roverud, E., Clayton, K. K., and Best, V. (2016). "Determining the energetic and informational components of speech-on-speech masking," *J. Acoust. Soc. Am.* **140**, 132–144.
- Leibold, L. J., Buss, E., and Calandruccio, L. (2018). "Developmental effects in masking release for speech-in-speech perception due to a target/masker sex mismatch," *Ear Hear.* **39**, 935–945.
- Li, N., and Loizou, P. (2007). "Factors influencing glimpsing of speech in noise," *J. Acoust. Soc. Am.* **122**, 1165–1172.
- Marrone, N. L., Mason, C. R., and Kidd, G., Jr. (2008a). "Effect of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms," *J. Acoust. Soc. Am.* **124**, 3064–3075.
- Marrone, N. L., Mason, C. R., and Kidd, G., Jr. (2008b). "Evaluating the benefit of hearing aids in solving the cocktail party problem," *Trends Amplif.* **12**, 300–315.
- Marrone, N. L., Mason, C. R., and Kidd, G., Jr. (2008c). "Tuning in the spatial dimension: Evidence from a masked speech identification task," *J. Acoust. Soc. Am.* **124**, 1146–1158.
- Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). "Speech recognition in adverse conditions: A review," *Lang. Cogn. Process.* **27**, 953–978.
- McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., and Wingfield, A. (2005). "Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech," *Quart. J. Exp. Psychol.* **58**, 22–33.
- Noble, W., and Gatehouse, S. (2006). "Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the Speech, Spatial, and Qualities of Hearing Scale (SSQ)," *Int. J. Audiol.* **45**, 172–181.
- Rana, B., and Bucholtz, J. (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**, 2195–2206.
- Roberts, B., and Summers, R. J. (2018). "Informational masking of speech by time-varying competitors: Effects of frequency region and number of interfering formants," *J. Acoust. Soc. Am.* **143**, 891–900.
- Roverud, E., Best, V., Mason, C. R., and Kidd, G., Jr. (2017). "The influence of word sequence length and source segregation on serial recall of masked speech," in *Proceedings of the Annual Midwinter Research Meeting of the Association for Research in Otolaryngology*, Baltimore, MD
- Shinn-Cunningham, B. G. (2017). "Cortical and sensory causes of individual differences in selective auditory attention ability among listeners with normal hearing thresholds," *J. Speech Lang. Hear. Res.* **60**, 2976–2988.
- Shinn-Cunningham, B. G., and Best, V. (2008). "Selective attention in normal and impaired hearing," *Trends Amplif.* **12**(4), 283–299.
- Schubotz, W., Brand, T., Kollmeier, B., and Ewert, S. D. (2016). "Monaural speech intelligibility and detection in maskers with varying amounts of spectro-temporal speech features," *J. Acoust. Soc. Am.* **140**, 524–540.
- Srinivasan, N. K., Stansell, M., and Gallun, F. J. (2017). "The role of early and late reflections on spatial release from masking: Effects of age and hearing loss," *J. Acoust. Soc. Am.* **141**, EL185–EL191.
- Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V. A., Kidd, G., Jr., and Patel, A. D. (2015). "Musical training and the cocktail party problem," *Sci. Rep.* **5**, 11628.
- Wang, D. L. (2005). "On ideal binary mask as the computational goal of auditory scene analysis," in *Speech Separation by Humans and Machines*, P. Divenyi (Kluwer Academic, Norwell, MA), pp. 181–197.
- Yost, W. A. (2017). "Spatial release from masking based on binaural processing for up to six maskers," *J. Acoust. Soc. Am.* **141**, 2093–2106.