The importance of a broad bandwidth for understanding "glimpsed" speech[a]

Virginia Best,[b] Elin Roverud, Lucas Baltzell, Jan Rennies,[c] and Mathieu Lavandier[d]
Department of Speech, Language and Hearing Sciences, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215, USA

(Received 13 May 2019; revised 25 September 2019; accepted 11 October 2019; published online 11 November 2019)

When a target talker speaks in the presence of competing talkers, the listener must not only segregate the voices but also understand the target message based on a limited set of spectrotemporal regions ("glimpses") in which the target voice dominates the acoustic mixture. Here, the hypothesis that a broad audible bandwidth is more critical for these sparse representations of speech than it is for intact speech is tested. Listeners with normal hearing were presented with sentences that were either intact, or progressively "glimpsed" according to a competing two-talker masker presented at various levels. This was achieved by using an ideal binary mask to exclude time-frequency units in the target that would be dominated by the masker in the natural mixture. In each glimpsed condition, speech intelligibility was measured for a range of low-pass conditions (cutoff frequencies from 500 to 8000 Hz). Intelligibility was poorer for sparser speech, and the bandwidth required for optimal intelligibility increased with the sparseness of the speech. The combined effects of glimpsing and bandwidth reduction were well captured by a simple metric based on the proportion of audible target glimpses retained. The findings may be relevant for understanding the impact of high-frequency hearing loss on everyday speech communication.

© *2019 Acoustical Society of America*. https://doi.org/10.1121/1.5131651

[MIM] Pages: 3215–3221

I. INTRODUCTION

It is well accepted that the reduced audibility experienced by listeners with hearing loss limits the intelligibility of speech. As such, the primary goal of hearing-aid amplification is to restore audibility towards normal sensation levels, and this is generally done by providing frequency-dependent gain according to the severity of the loss at each frequency (Dillon, 2012). However, in addition to restoring audibility, prescriptive amplification must also deal with the somewhat conflicting goal of maintaining a comfortable loudness level despite abnormal growth of loudness with intensity (recruitment) in listeners with sensorineural hearing loss. Thus, most prescriptions represent a compromise and do not lead to sensation levels equivalent to those of normally hearing listeners. This is especially true for higher frequencies, where hearing losses tend to be more severe. This compromise often does not affect speech in quiet, where a well-fitted hearing aid may support near-perfect speech intelligibility. Part of the reason for this is that speech is a highly redundant signal, and robust information in one frequency region can counteract the loss of information in another region. Moreover, band importance functions suggest that the higher frequencies in speech (>4 kHz) contribute less to intelligibility than mid frequencies (1–3 kHz; ANSI S3.5-1997).

Understanding the impact of audibility on speech intelligibility in the presence of competing sounds is more complicated. It is often reported that listeners with hearing loss—even if it is a relatively mild loss—experience substantial difficulties in the noisy situations they encounter in their everyday lives (e.g., Gatehouse and Noble, 2004). Situations involving multiple people talking at once (parties, restaurants, etc.) seem to be a particular problem. In these situations, where the interfering sounds contain fluctuations in energy, there will be moments when the speech of interest is masked to different degrees by the interference and information is lost. However, it is thought that listeners make use of other moments (so-called "glimpses") where the speech of interest dominates the acoustic mixture. Much of our knowledge about this process comes from studies that have used "interrupted" speech. In these studies, intelligibility is measured for speech that is periodically interrupted by silence or by noise (e.g., Miller and Licklider, 1950; Howard-Jones and Rosen, 1993). It seems that listeners use contextual information to fill in the missing information, and that strong linguistic skills offer an advantage (e.g., Benard *et al.*, 2014). Listeners with hearing loss perform more poorly than normally hearing listeners for interrupted speech (e.g., Baskent *et al.*, 2010), as do older listeners with relatively good hearing (e.g., Bologna *et al.*, 2018).

In competing-talker situations, listeners are faced with two challenges. First, they must segregate the acoustic mixture in order to isolate the speech spoken by the target talker. This can be much more difficult when the interference is speech than when it is noise, because the interference is

[a] Portions of this work were presented in "The importance of high-frequency information for understanding 'glimpsed' speech," Proceedings of the Midwinter Meeting of the Association for Research in Otolaryngology, Baltimore, MD, February 2019.

[b] Electronic mail: ginbest@bu.edu

[c] Also at: Fraunhofer IDMT, Hearing Speech and Audio Technology and Cluster of Excellence Hearing4all, Oldenburg, Germany.

[d] Also at: Univ Lyon, ENTPE, Laboratoire Génie Civil et Bâtiment, Rue M. Audin, F-69518 Vaulx-en-Velin Cedex, France.

more similar to the signal of interest, and often highly distracting. Historically, many studies have focused on this aspect of the "cocktail party problem" (see review in Kidd and Colburn, 2017). However, even if the different talkers are successfully segregated, listeners are faced with a second challenge, which is to understand the target message based on an interrupted representation of the target speech. In this case, unlike for many studies of interrupted speech, the interruptions are not periodic and may occur at different moments in different frequency regions depending on the characteristics of the competing talkers. Two recent studies focused on this aspect of the problem, and tried to understand how the use of target glimpses in speech mixtures affects listener performance (Kidd *et al.*, 2016; Best *et al.*, 2017). In those studies, performance was compared for speech-on-speech mixtures before and after the application of a "glimpsing model" in which only the time-frequency regions dominated by the target were retained. This procedure eliminates the requirement for listeners to segregate the mixture, but captures their ability to make use of target glimpses. These studies found that for a population of young adult listeners (with and without hearing impairment), those who had trouble in speech-on-speech masking conditions tended to also have trouble with glimpsed speech. This suggests that understanding glimpsed speech, rather than the segregation of competing talkers, might be the critical problem in some cases.

Here, we extended that basic approach to examine the importance of a broad audible bandwidth in speech mixtures. Specifically, we tested the hypothesis that loss of audibility in the high frequencies affects performance for sparse representations of speech more than for intact speech. Because sparse speech has less redundancy than intact speech, the idea is that any loss of information will more dramatically affect speech intelligibility. Alternatively, it might be that glimpsed speech is as robust as intact speech to reductions in bandwidth. Such an outcome would suggest that negative effects of reduced audibility on performance in competing-talker situations are related to the ability to segregate competing talkers rather than to the ability to piece together sparse glimpses of speech. Either way, the results were expected to further our understanding of why hearing loss, even if it is restricted to the high frequencies, can have such a dramatic impact on performance in multitalker listening situations.

II. METHODS

A. Participants

Ten adults (ages 19–39 years, mean age 25) participated in experiment 1, and ten adults (ages 19–40 years, mean age 24) participated in experiment 2. Five participants were common to both experiments, and completed experiment 1 first. The first author was a participant in both experiments, and the second author was a participant in experiment 1. All participants had normal hearing at audiometric frequencies from 250 Hz to 8 kHz. External participants were paid for their participation and gave informed consent. The procedures were approved by the Boston University Institutional Review Board.

B. Stimuli

Two experiments were conducted that differed only in the speech materials used to measure intelligibility (closed-set matrix sentences in experiment 1 and naturally spoken open-set sentences in experiment 2). Experiment 2 was primarily intended to provide a confirmation of the general pattern of results of experiment 1. In addition, the use of two different kinds of materials provided the potential for insights into effects of bandwidth that apply quite broadly to the perception of sentence-level speech versus those that might be sensitive to the particular speech characteristics.

Experiment 1 used a corpus of monosyllabic words that has been described previously (Kidd *et al.*, 2008). This corpus contains 40 words (eight in each of five word categories), and is typically used to create matrix-style sentences by concatenating one word from each of the five word categories (e.g., "Sue bought two red toys"). Stimuli for this experiment were created by combining three different sentences (one target and two maskers). On each trial, these three sentences were spoken by three different female talkers (selected at random from a set of eight).

Experiment 2 used meaningful open-set sentences from the Harvard/IEEE corpus (Rothauser *et al.*, 1969). This corpus contains 720 unique sentences, each containing five keywords (that are scored) and a number of connecting words (that are not scored). An example is: "A large size in stockings is hard to sell." Stimuli for this experiment were created by combining three different sentences (one target and two maskers) from a single female talker (different from the talkers used in experiment 1).

In both experiments, the target had a nominal level of 65 dB sound pressure level (SPL) (before any processing was applied; see below), and the root-mean-square level of each masker sentence was scaled relative to the target sentence to achieve target-to-masker ratios (TMRs) of 0, −10, and −20 dB. A condition was also included in which there were no maskers present. All stimuli were presented diotically.

For each stimulus, a simple glimpsing model was applied to estimate the sparse version of the target that is potentially available given the TMR and the random draw of the competing sentences. The model was based on the approach of Wang (2005) and Brungart *et al.* (2006). In brief, the signals were analyzed using 128 frequency channels logarithmically spaced between 80 Hz and 8 kHz, and 20-ms time windows with 50% overlap. Time-frequency tiles in which the target energy exceeded the total masker energy were assigned a mask value of 1, and the remaining tiles were assigned a value of 0. The mask was then applied to the clean target signal (the target signal with no maskers), such that only the time-frequency regions containing a mask value of 1 were retained before the signal was resynthesized. Using this approach, the sparseness of the mask varies with TMR (see Fig. 1) and in the resulting speech there are fewer glimpses at poorer TMRs. It is worth noting that application of the mask to the *clean target* differs from the more common approach of applying the mask to the *mixture* of target and maskers. The difference is that the retained stimulus contained no low-level masker components, allowing us to

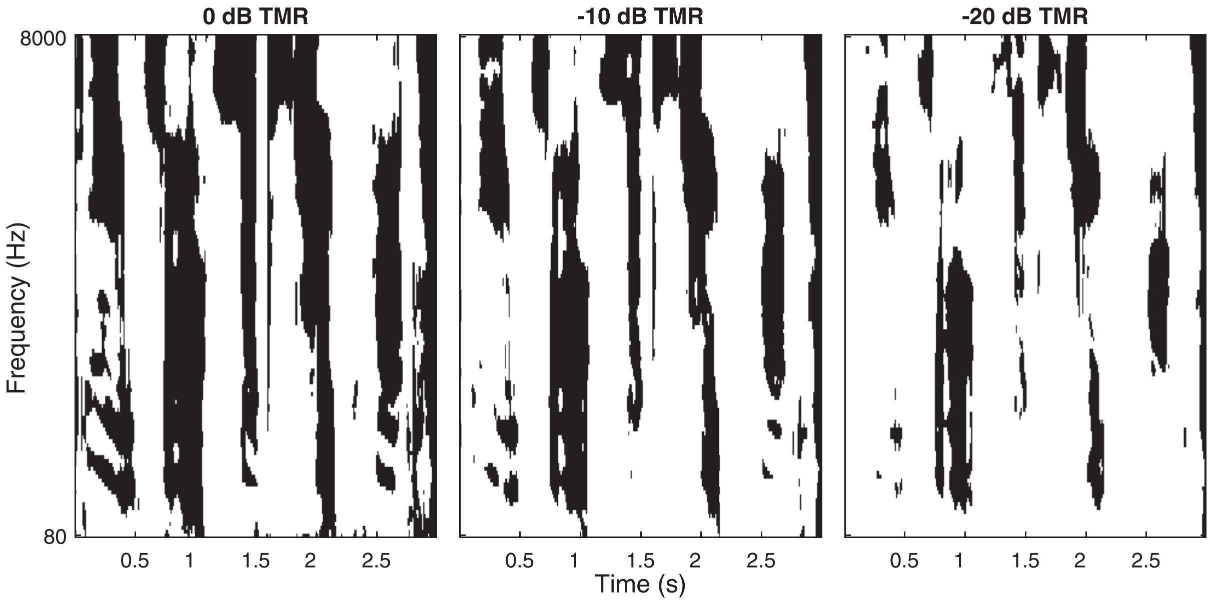

FIG. 1. Binary masks generated for an example target sentence in the presence of two masker sentences at a TMR of 0 dB (left), −10 dB (middle), and −20 dB (right). Black regions indicate target-dominated time-frequency regions. This example uses stimuli from experiment 1 but would look qualitatively similar for the stimuli of experiment 2.

focus exclusively on the issue of sparseness in the target. Also note that for the condition with no maskers, the glimpsing model was still applied but with an all-ones mask, resulting in "intact" speech.

After applying the glimpsing model, we adopted an approach from Silberer and colleagues (2015) to characterize the effect of bandwidth on performance in each of the glimpsed conditions. The availability of high-frequency information was systematically varied by low-pass filtering the glimpsed stimuli at 13 different frequency cutoffs from 500 to 8000 Hz in third-octave steps. The low-pass filter was implemented as a brick-wall filter in the frequency domain.

C. Procedures

Stimuli were controlled in MATLAB (MathWorks Inc., Natick, MA) and presented via a 24-bit soundcard (RME HDSP 9632, Haimhausen, Germany) through a pair of headphones (Sennheiser HD280 Pro, Wedemark, Germany). Participants sat in a double-walled sound-treated booth fitted with a computer monitor, keyboard and mouse.

In experiment 1, participants were instructed to listen to the sentence and to select one word from each of the five word categories. Responses were given by selecting from a grid of 40 (five categories × eight options) presented on the monitor. Correct-answer feedback was provided. Experiment 2 made use of a self-scoring method originally introduced by Culling and Colburn (2000). Participants heard the target sentence and then were required to type their response into a text box on the screen. After pressing enter, the correct transcript of the sentence was displayed with keywords capitalized, and the participant scored their response by choosing 0–5 from a drop-down menu. Participants were told that the precision of their typed responses was less important than their selected score. They were encouraged to type their responses rapidly (to limit memory effects) and not to worry about spelling errors and typing errors. In determining their score, they were told that homonyms should be counted as correct (e.g., "bear" instead of "bare"), but that if they heard the wrong form of a word (e.g., "dog" instead of "dogs") that should be counted as an error. Transcripts of the responses given on each trial were stored along with the scores, so that the experimenters could check for any systematic problems with the accuracy of self-scoring after each participant had completed testing.[1]

Each experiment was completed in a single session of approximately two hours. The session began with one training block, followed by 52 blocks of testing. The training block consisted of 10 trials of intact speech with the widest bandwidth (8000 Hz) and was intended to familiarize participants with the stimuli and the response method. Each testing block consisted of 10 trials in one of the 52 experimental conditions (four glimpsed conditions × 13 cutoff frequencies). The order of these blocks was randomized across participants, as was the particular set of sentences used to test a given condition.

D. Calculation of minimum bandwidth

The concept of the "minimum bandwidth" as described by Silberer and colleagues (2015) was adopted to quantify the point at which a loss of high-frequency information had an impact on intelligibility. For each participant, in each of the glimpsed conditions, the percentage of words correct was calculated as a function of cutoff frequency. After log-transforming the cutoff frequency axis, logistic functions were fitted to these functions using the psignifit toolbox version 2.5.6 for MATLAB which implements the maximum-likelihood method described by Wichmann and Hill (2001). The lower asymptote was set to chance performance (12.5% for experiment 1; 0% for experiment 2) while the upper asymptote was left as a free parameter representing optimal performance (which fell at or near the 8 kHz performance level). The minimum bandwidth was then defined as the lowest cutoff frequency that produced performance equivalent to optimal performance, operationalized as the cutoff frequency corresponding to 0.9 of the dynamic range (from chance to optimal performance) on the logistic function. Minimum bandwidth estimates that fell above 8 kHz (meaning that the function had not reached a plateau by 8 kHz) were capped at 8 kHz.

E. Analysis of available speech information

A stimulus analysis was conducted to explore whether the combined effects of glimpsing and low-pass filtering on performance could be understood in terms of the available speech information. Since both manipulations result in the loss of portions of the target across the spectrotemporal plane, we used a simple version of the "glimpse proportion" metric proposed by Cooke and colleagues (Cooke, 2006; Tang *et al.*, 2016). Briefly, spectrotemporal excitation patterns were generated for a given stimulus using 34 gammatone filters (equally spaced on the ERB scale with center frequencies from 100 to 7500 Hz) and 10-ms time windows (as suggested by Tang *et al.*, 2016). A threshold was applied to the excitation pattern to exclude time-frequency units below a certain level of audibility for an average normally hearing listener [defined as 25 dB hearing level (HL) as per Tang *et al.*, 2016] and generate a binary representation of the suprathreshold time-frequency units. This binary representation was created for the intact/unfiltered stimulus as well as for the same stimulus after glimpsing and/or low-pass filtering. Then, within each frequency channel, the proportion of time bins retained in the glimpsed/filtered target was calculated. Finally, these proportions were weighted according to the band importance function for "average speech" as defined in the speech intelligibility index (ANSI S3.5-1997, Table I) and summed. The result was a glimpse proportion value that ranged from 0 (no glimpses retained) to 1 (all glimpses retained). A stable estimate of the glimpse proportion for each condition was obtained by applying it to 50 randomly generated stimuli (separately for experiments 1 and 2).

III. RESULTS

Group means (and standard deviations) of the raw percent correct data are plotted in Fig. 2 for experiment 1 (left panel) and experiment 2 (right panel). The data are plotted as a function of the low-pass cutoff frequency, and the four lines represent the four glimpsed conditions. In experiment 1, focusing first on performance with the full bandwidth (8 kHz cutoff), the expected drop in performance with decreasing TMR can be seen. Whereas scores for intact speech were at ceiling (100%), there was a slight drop for TMRs of 0 and −10 dB (99% and 91%) and a dramatic drop for the −20 dB TMR condition (66%). More importantly, there were clear differences across glimpsed conditions in both the slopes of the functions and in the point at which performance began to drop as the cutoff frequency was lowered. For intact speech, performance stayed at ceiling for much of the range and then dropped steeply when the cutoff frequency fell below about 1 kHz. As the TMR was reduced, resulting in sparser representations of the sentences, best performance decreased as expected, but also performance started to drop at a higher cutoff frequency and with a shallower slope. Slope values in experiment 1 ranged from 24% per third-octave step (intact) to 7% per third-octave step (−20 dB TMR). In experiment 2, a broadly similar pattern of performance was observed, although scores were lower overall and the effects of glimpsing and filtering were more severe. For full-bandwidth speech (8 kHz cutoff), performance dropped slightly from intact speech (98%) to a TMR of 0 dB (94%) and then fell dramatically for TMRs of −10 dB (57%) and −20 dB (14%). Again, there were differences across glimpsed conditions in

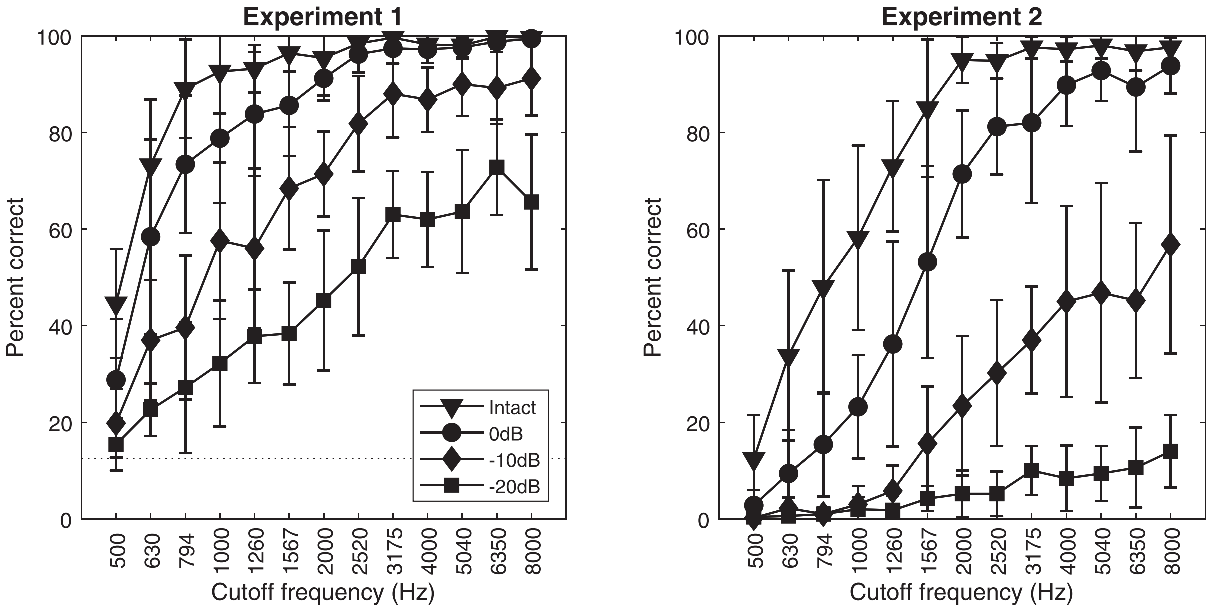

FIG. 2. Group-mean performance in percent correct as a function of low-pass cutoff frequency for the four different glimpsed conditions (intact speech, and glimpsed speech according to TMRs of 0, −10, and −20 dB). The dashed line for experiment 1 indicates chance performance. Error bars in both panels show across-subject standard deviations.

the pattern of performance as the cutoff frequency was lowered. For intact speech, performance dropped only for cutoffs below about 2 kHz, whereas for the sparser conditions, the decline began at higher cutoff frequencies and exhibited a shallower slope. Slope values in experiment 2 ranged from 18% per third-octave step (intact) to 7% per third-octave step (−20 dB TMR).

Figure 3 shows group-mean minimum bandwidths (corresponding to 0.9 of the dynamic range for that glimpsed condition) that were extracted from the fits to individual data. This representation confirms the observations made above, that a wider bandwidth was required to maintain optimal performance in the glimpsed conditions compared to the intact condition. In Experiment 1, a bandwidth of 939 Hz was sufficient for optimal performance with intact speech, whereas for −20 dB glimpsed speech, the full bandwidth of 8 kHz was required. A repeated-measures analysis of variance (ANOVA) conducted on the minimum bandwidths confirmed that the effect of condition was significant [F(3,27) = 159.435, p < 0.001)]. *Post hoc* comparisons (paired t tests with Bonferroni correction) indicated that minimum bandwidths for each condition were significantly different from the adjacent conditions. In experiment 2, a bandwidth of 1618 Hz was sufficient for optimal performance with intact speech, whereas the full bandwidth of 8 kHz was required for glimpsed speech at −10 and −20 dB TMR. A repeated-measures ANOVA conducted on the minimum bandwidths confirmed that the effect of condition was significant [F(3,27) = 80.892, p < 0.001)]. *Post hoc* comparisons indicated that minimum bandwidths were significantly different for intact vs 0 dB, and for 0 dB vs −10 dB (−10 dB and −20 dB were both capped at 8 kHz for all listeners so that comparison is not meaningful).

Figure 4 plots behavioral performance as a function of glimpse proportion, separately for each experiment. Each data point represents the combination of one glimpsed condition and one cutoff frequency (52 in total per experiment). The percent correct value represents the across-subject mean (from Fig. 2) and the glimpse proportion value represents the mean value obtained from the simulation. The solid lines in each panel show logistic fits to the data points (generated using the psignifit toolbox). The metric appears to capture the behavioral data reasonably well, with a mean absolute error around the fit of 3.9 percentage points (experiment 1) and 5.0 percentage points (experiment 2). Note that the mapping functions differ across the two experiments, with a higher glimpse proportion needed for the same level of performance in experiment 2. For example, a glimpse proportion of 0.78 was needed for 50% correct performance in experiment 2, whereas the same performance was reached in experiment 1 with a glimpse proportion of only 0.62.[2]

IV. DISCUSSION

Previous studies have shown that low-pass filtering has an impact on speech intelligibility in noise (e.g., Studebaker *et al.*, 1987) and in the presence of competing talkers (e.g., Kidd *et al.*, 2010). Moreover, Silberer and colleagues (2015) demonstrated that the minimum bandwidth required for optimal intelligibility is wider for speech in noise than speech in quiet. Our results corroborate these previous findings. However, the novelty in our approach is that we implemented the glimpsing model to capture the sparseness of the target speech while completely eliminating the interference. In this way, we are able to attribute the effects we see directly to the availability and use of target glimpses, rather than, for example, the ability to segregate the target from the interference. We conclude that a broad bandwidth of speech information becomes increasingly important when the speech is sparsely represented as it is in mixtures of competing talkers.

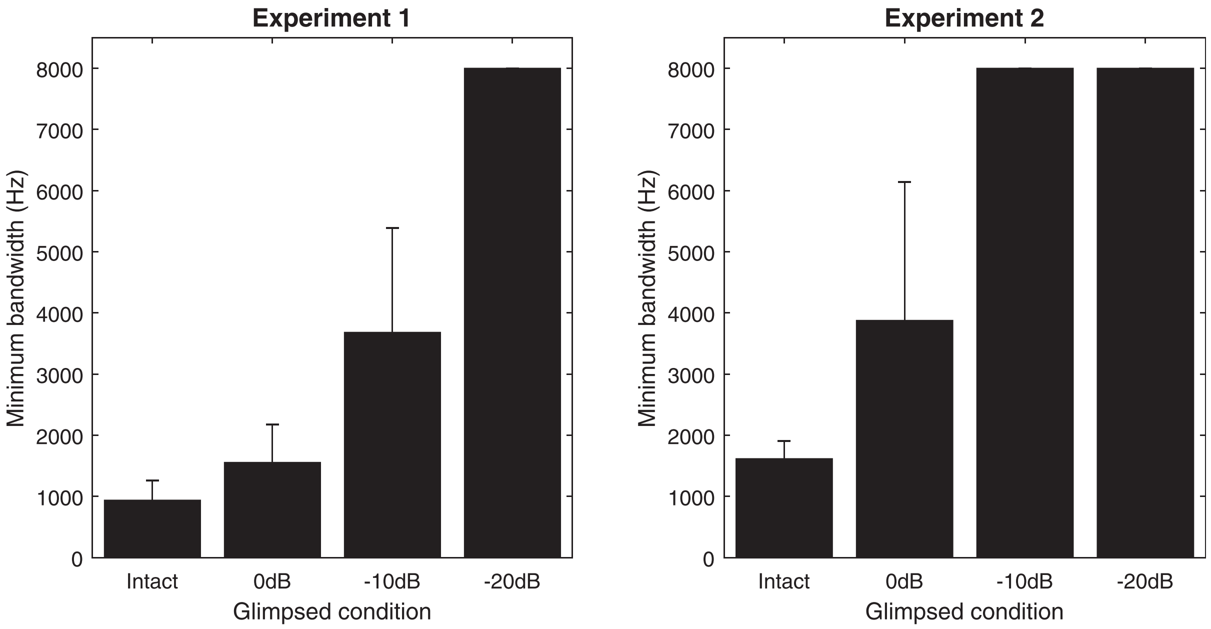

FIG. 3. Group-mean minimum bandwidth required for optimal performance in each of the glimpsed conditions. Error bars in both panels show across-subject standard deviations.

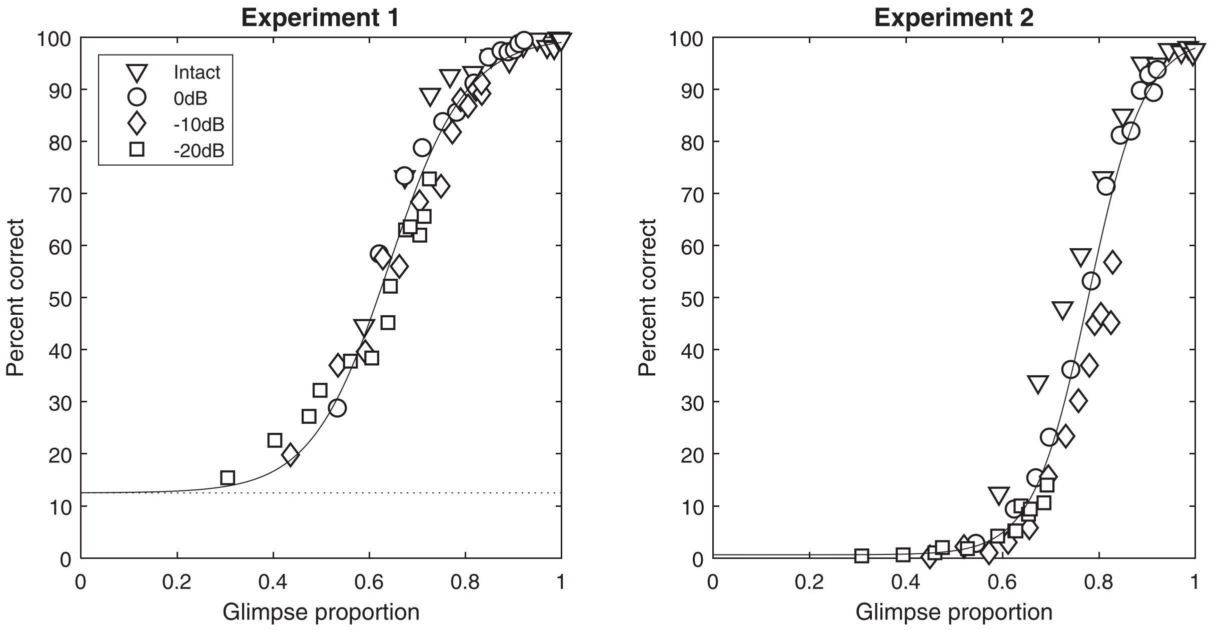

FIG. 4. Performance as a function of glimpse proportion. Each data point represents mean performance for one combination of glimpsed condition and cutoff frequency (52 in total) and the solid line shows a logistic fit to the data points.

Our stimulus analysis showed that the combined effect of glimpsing and low-pass filtering on performance could be quite well described by a relatively simple measure of the available speech information (the glimpse proportion). This metric provides a useful framework for understanding why performance starts to decline at a higher cutoff frequency for glimpsed speech as compared to intact speech: when competing sounds obscure enough of a target sound to reduce the glimpse proportion below some critical level, then further reductions in the glimpse proportion (such as those caused by filtering) become relevant. The metric can also explain the differences in slope observed in Fig. 2, i.e., why there is a more gradual change in performance with changes in cutoff frequency for glimpsed speech. Put differently, larger changes in bandwidth are required to obtain equivalent changes in performance, which makes sense if one considers the fact that less information is added/subtracted for a given change in bandwidth with glimpsed speech. Note that although we focused here on high-frequency information, because of its relevance to the most common kind of hearing loss, we would expect similar patterns of performance for other kinds of bandwidth reduction. For example, the basic sparseness-bandwidth tradeoff should apply to high-pass filtered stimuli just as it does to low-pass filtered stimuli. This is a prediction that could be explicitly tested in future studies. It would also be interesting to explore the influence of the glimpse proportion parameters (e.g., spectral and temporal resolution) which were simply fixed here using published values from Tang *et al*. (2016).

One compelling aspect of these results was that the glimpse proportion metric was able to account quite well for performance across two experiments that used very different kinds of speech materials. This is possible simply by assuming different mapping functions from glimpse proportion to speech intelligibility. While comparisons across the two experiments should be made with caution due to the fact that the participant groups were not identical, Fig. 4 suggests that a larger glimpse proportion was required for the same level of intelligibility with open-set as compared to closed-set materials. This parallels the well-known speech intelligibility index [ANSI S3.5-1997; and see Kryter (1962)], where the function relating the value of the index to performance depends on the nature of the message that is being conveyed (including context, predictability, set-size, etc.). The general idea that restraining the response set reduces the speech information required for equivalent levels of performance is entirely consistent with earlier studies showing effects of set size on speech reception thresholds in noise (e.g., Miller *et al.*, 1951). For the purposes of this study, the important point is that a loss of bandwidth, in combination with a sparse representation of speech, may have a particularly dramatic effect for realistic, open-set speech materials. Two previous studies are relevant here, which reported large effects of low-pass filtering on the intelligibility of periodically interrupted open-set sentences (Lacroix *et al.*, 1979; Bhargava and Baskent, 2012). On the other hand, it is worth keeping in mind that real-world speech stimuli are often accompanied by lip-reading cues, which provide redundant information and may lower the acoustic bandwidth required for good intelligibility (Silberer *et al.*, 2015). Future experiments using naturalistic speech materials that also incorporate visual cues, competing talkers, and other kinds of noise, may help to better estimate the bandwidth requirements of real-world communication situations.

While low-pass filtering is an extremely crude analog of the loss of audibility experienced by listeners with hearing loss, our results suggest that reductions in high frequency audibility have a dramatic impact on the intelligibility of sparse speech like that which is available in multitalker listening situations. The implication of this is that reduced audibility in this region, even if it stems from relatively mild hearing losses or from the tradeoffs associated with prescribing hearing-aid gain, may be functionally important. Consistent with this are the results of several studies that have shown beneficial effects of extending the bandwidth of amplification for listeners with hearing loss (Moore *et al.*, 2010; Levy *et al.*, 2015) or systematically increasing the amount of amplification in the high-frequency region (Glyde *et al.*, 2015) in mixtures containing competing talkers at different spatial locations. Taken together, these results suggest that a complete picture of the difficulties experienced by listeners with hearing loss in their everyday lives may require a closer consideration of the audibility of high-frequency information (see related discussions in Monson *et al.*, 2014; Moore, 2016).

ACKNOWLEDGMENTS

This work was supported by NIH-NIDCD Awards Nos. DC015760 and DC04545. The international mobility of M.L. at Boston University was partly funded by ENTPE and the Fondation pour l'Audition (Speech2Ears grant). The authors would like to thank Gerald Kidd, Chris Mason, and Ganesh Swaminathan for many helpful discussions. The authors are also grateful to Tang and colleagues for making their MATLAB code available online.

[1] While in several cases the self-given score did not correspond to the score the experimenter would assign based on the typed transcripts, these mismatches were occasional and non-systematic (with roughly as many overestimates as underestimates). Indeed, for a handful of participants it was confirmed that rescoring the data based on the transcripts would have a negligible effect on their final result. Thus, the original scores were retained for all participants, as these provide the most direct indication of how many words were heard correctly (i.e., according to the participant and without any interpretation by the experimenters).

[2] It is worth pointing out that the general pattern of results shown in Fig. 4 does not change dramatically if one uses a different band importance function to calculate the glimpse proportion metric. For example, if one omits the band importance function and gives equal weight to all frequencies, the glimpse proportion values corresponding to 50% correct shift only slightly to 0.66 and 0.80 for experiments 1 and 2, respectively. Interestingly, the mean absolute error around the fits in this case is slightly higher for experiment 1 (4.3 percentage points) but slightly *lower* for experiment 2 (3.6 percentage points), raising the question of how appropriate the standard band importance function is under these conditions (see related discussions in Healy *et al.*, 2013; Shen and Kern, 2018).

ANSI (**1997**). ANSI S3.5-1997, *Methods for Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).

Başkent, D., Eiler, C. L., and Edwards, B. (**2010**). "Phonemic restoration by hearing-impaired listeners with mild to moderate sensorineural hearing loss," Hear. Res. **260**, 54–62.

Benard, M. R., Mensink, J. S., and Başkent, D. (**2014**). "Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities," J. Acoust. Soc. Am. **135**, EL88–EL94.

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.

Bhargava, P., and Başkent, D. (**2012**). "Effects of low-pass filtering on intelligibility of periodically interrupted speech," J. Acoust. Soc. Am. **131**, EL87–EL92.

Bologna, W. J., Vaden, K. I., Ahlstrom, J. B., and Dubno, J. R. (**2018**). "Age effects on perceptual organization of speech: Contributions of glimpsing, phonemic restoration, and speech segregation," J. Acoust. Soc. Am. **144**, 267–281.

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.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Culling, J. F., and Colburn, H. S. (**2000**). "Binaural sluggishness in the perception of tone sequences and speech in noise," J. Acoust. Soc. Am. **107**, 517–527.

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

Gatehouse, S., and Noble, W. (**2004**). "The speech, spatial and qualities of 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.

Healy, E. W., Yoho, S. E., and Apoux, F. (**2013**). "Band importance for sentences and words reexamined," J. Acoust. Soc. Am. **133**, 463–473.

Howard-Jones, P. A., and Rosen, S. (**1993**). "Uncomodulated glimpsing in 'checkerboard' noise," J. Acoust. Soc. Am. **93**, 2915–2922.

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

Kidd, G., and Colburn, H. S. (**2017**). "Informational masking in speech recognition," in *The Auditory System at the Cocktail Party*, edited by J. C. Middlebrooks and J. Z. Simon (Springer Nature, New York), pp. 75–109.

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

Kidd, G., 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.

Kryter, K. D. (**1962**). "Methods for the calculation and use of the Articulation Index," J. Acoust. Soc. Am. **34**, 1689–1697.

Lacroix, P. G., Harris, J. D., and Randolph, K. J. (**1979**). "Multiplicative effects on sentence comprehension for combined acoustic distortions," J. Speech Hear. Res. **22**, 259–269.

Levy, S. C., Freed, D. J., Nilsson, M., Moore, B. C. J., and Puria, S. (**2015**). "Extended high-frequency bandwidth improves speech reception in the presence of spatially separated masking speech," Ear Hear. **36**, e214–e224.

Miller, G. A., Heise, G. A., and Lichten, W. (**1951**). "The intelligibility of speech as a function of the context of the test materials," J. Exp. Psychol. **41**, 329–335.

Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.

Monson, B. B., Hunter, E. J., Lotto, A. J., and Story, B. H. (**2014**). "The perceptual significance of high-frequency energy in the human voice," Front. Psychol. **5**, 1–11.

Moore, B. C. J. (**2016**). "A review of the perceptual effects of hearing loss for frequencies above 3 kHz," Int. J. Audiol. **55**, 707–714.

Moore, B. C. J., Füllgrabe, C., and Stone, M. A. (**2010**). "Effect of spatial separation, extended bandwidth, and compression speed on intelligibility in a competing-speech task," J. Acoust. Soc. Am. **128**, 360–371.

Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.

Shen, Y., and Kern, A. B. (**2018**). "An analysis of individual differences in recognizing monosyllabic words under the speech intelligibility index framework," Trends Hear. **22**, 1–14.

Silberer, A. B., Bentler, R., and Wu, Y. H. (**2015**). "The importance of high-frequency audibility with and without visual cues on speech recognition for listeners with normal hearing," Int. J. Audiol. **54**, 865–872.

Studebaker, G. A., Pavlovic, C. V., and Sherbecoe, R. L. (**1987**). "A frequency importance function for continuous discourse," J. Acoust. Soc. Am. **81**, 1130–1138.

Tang, Y., Cooke, M., Fazenda, B. M., and Cox, T. J. (**2016**). "A metric for predicting binaural speech intelligibility in stationary noise and competing speech maskers," J. Acoust. Soc. Am. **140**, 1858–1870.

Wang, D. (**2005**). "On ideal binary mask as the computational goal of auditory scene analysis," in *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluwer Academic, Norwell, MA), pp. 181–197.

Wichmann, F. A., and Hill, N. J. (**2001**). "The psychometric function: I. Fitting, sampling and goodness-of-fit.," Perc. Psych. **63**, 1293–1313.