Speech recognition for multiple bands: Implications for the Speech Intelligibility Index

Larry E. Humes and Gary R. KiddMSS

Citation: The Journal of the Acoustical Society of America **140**, 2019 (2016); doi: 10.1121/1.4962539
View online: http://dx.doi.org/10.1121/1.4962539
View Table of Contents: http://asa.scitation.org/toc/jas/140/3
Published by the Acoustical Society of America

Speech recognition for multiple bands: Implications for the Speech Intelligibility Index

Larry E. Humes[a] and Gary R. Kidd
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405-7002, USA

(Received 2 March 2016; revised 28 July 2016; accepted 15 August 2016; published online 26 September 2016)

The Speech Intelligibility Index (SII) assumes additivity of the importance of acoustically independent bands of speech. To further evaluate this assumption, open-set speech recognition was measured for words and sentences, in quiet and in noise, when the speech stimuli were presented to the listener in selected frequency bands. The filter passbands were constructed from various combinations of 20 bands having equivalent (0.05) importance in the SII framework. This permitted the construction of a variety of equal-SII band patterns that were then evaluated by nine different groups of young adults with normal hearing. For monosyllabic words, a similar dependence on band pattern was observed for SII values of 0.4, 0.5, and 0.6 in both quiet and noise conditions. Specifically, band patterns concentrated toward the lower and upper frequency range tended to yield significantly lower scores than those more evenly sampling a broader frequency range. For all stimuli and test conditions, equal SII values did not yield equal performance. Because the spectral distortions of speech evaluated here may not commonly occur in everyday listening conditions, this finding does not necessarily represent a serious deficit for the application of the SII. These findings, however, challenge the band-independence assumption of the theory underlying the SII.
© *2016 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4962539]

[MSS] Pages: 2019–2026

I. INTRODUCTION

In this report, we explore the recognition of words and sentences presented as various combinations of acoustically independent (non-overlapping) frequency bands. Gathering such data may seem unnecessary because the Articulation Index (AI) framework, more recently described as the Speech Intelligibility Index [SII; American National Standards Institute (ANSI), 1997], was derived originally from considerable systematic research on the perception of filtered speech (e.g., French and Steinberg, 1947; Fletcher and Galt, 1950), primarily nonsense syllables and words in quiet. However, it had been known for some time that the AI framework did not provide an accurate account of the perception of speech when filtered into multiple acoustically independent or spectrally non-overlapping bands (e.g., Kryter, 1960, 1962b). This early finding had been confirmed several times since for young normal-hearing adults, typically by demonstrating that two widely spaced narrow bands of speech, each yielding low speech-recognition scores (e.g., <20% correct), demonstrated super-additivity via high levels of performance (e.g., >80% correct) when combined. This finding was demonstrated and systematically investigated in a series of studies on the open-set recognition of meaningful sentences by Warren and colleagues spanning a decade or more (e.g., Warren *et al.*, 1995; Warren *et al.*, 1997; Warren and Bashford, 1999; Warren *et al.*, 2005; Bashford *et al.*, 2000). The focus in these studies was on the open-set recognition of meaningful sentences, primarily custom-made recordings of the Central Institute for the Deaf (CID) Everyday Sentences described by Davis and Silverman (1970). However, Warren *et al.* (1997) demonstrated that such super-additivity also could be observed for the recognition of monosyllables, although the magnitude of the effect was generally smaller compared to that observed with sentences. This sentence advantage for two-band super-additivity was explained in terms of additional top-down resources facilitating the reconstruction of sentences presented as sparse, widely separated bands of speech. Warren *et al.* (1997) also demonstrated that noise inserted between the spectral fragments facilitated recognition; once again, more so for meaningful sentences than for isolated monosyllables. It should be noted that it has seldom been the case that both words and sentences had been used as test materials and the focus was on the simple demonstration of super-additivity for a small number of bands and band combinations rather than a systematic evaluation of the AI/SII framework.

Although questioning the application of the AI/SII framework was a practical byproduct of this work, this was not viewed to be a critical problem. As noted by Kryter (1962b) in this same context, such extreme filtering conditions are not common everyday occurrences. As a result, it was argued that this finding *per se* does not pose serious limitations on the application of the AI/SII framework as an engineering tool that can be applied to most everyday listening situations. To the extent that many everyday listening situations may involve the integration of spectro-temporal glimpses of the speech stimulus (e.g., Howard-Jones and Rosen, 1993; Buss *et al.*, 2004; Cooke, 2006; Hall *et al.*, 2008a,b), however, it could be argued that integration of isolated speech fragments across frequency (and time) is, in fact, a very common everyday occurrence and one for which the AI/SII framework would be expected to apply.

[a]Electronic mail: humes@indiana.edu

The observed super-additivity of two or more acoustically independent bands of speech, in both quiet and noise, has had considerable importance for the theory underlying the AI/SII framework, as well as for our general understanding of speech perception. This is most apparent in the recent work of Healy and Warren (2003), Healy and Bacon (2007), Apoux and Healy (2009, 2010, 2012), and Healy *et al.* (2013). This work of Healy and colleagues posed serious questions about the assumed independence of bands comprising the speech-recognition performance estimates derived via the AI/SII framework. From this work, it followed that it would not be possible to derive the relative importance of a given frequency band to speech recognition unless this could be determined in a wide array of multi-band contexts. That is, the importance of a given band could only be established in the presence of a number of other contributing bands given that speech perception is typically a multi-band process. Apoux and Healy (2012) developed a novel approach to determining the relative importance of various frequency bands, referred to as the "compound method." Briefly, this method compares the speech-recognition performance obtained for n randomly selected bands to that measured for the same bands less one (n–1 bands), with the difference in performance attributed to the additional band present for the n-band condition. Most data were obtained for $n = 5$ to 10. Over many trials, the importance of band X is established over a wide range of conditions employing n bands where the other specific non-X bands included in the set are varied systematically.

Most recently, Healy *et al.* (2013) used the compound method to establish importance functions for the standard recordings of the Revised Speech Perception in Noise (R-SPIN) sentence materials (Bilger *et al.*, 1984) and the CID W-22 monosyllables (Hirsh *et al.*, 1952). In the current study, as well as our earlier work on *temporal* glimpses of broad-band speech (Kidd and Humes, 2012), we made use of the same R-SPIN materials, but used these materials to examine *both* the perception of monosyllables and sentences. That is, we used identical speech stimuli to evaluate the effects of various spectral-band patterns on speech recognition for *both* monosyllabic words and sentences. As in our prior work in the temporal domain, we excised the final word from the R-SPIN sentences, which is the sole response keyword for these sentence materials, and used these acoustically identical stimuli to explore performance differences between meaningful sentences and monosyllabic words. To examine the influence of semantic context on performance with meaningful sentences, performance was also evaluated with both the R-SPIN low-predictability (PL) and high-predictability (PH) sentence contexts.

Our approach to evaluating the open-set recognition of speech filtered into multiple bands provided a more direct assessment of the validity of the AI/SII framework for such conditions than either the two-band super-additivity approach or the compound-method. We constructed several multi-band speech stimuli, each band designed to contribute an equivalent amount (0.05) of importance for that condition according to the AI/SII framework. For example, several combinations of 12 bands, each contributing 0.05 to overall importance for a total importance value of 0.60, were evaluated. In this example, as a result of their equivalent importance of 0.60, open-set recognition performance would not be expected to differ significantly across these various 12-band combinations. This prediction was evaluated here for overall importance values of 0.40, 0.50, and 0.60 for monosyllabic words in quiet and in speech-shaped noise, and for sentences using keywords acoustically identical to the monosyllables in noise at an overall importance value of 0.60. Additional details regarding the series of experiments in this study follow in Sec. II.

II. METHOD

A. Subjects

Nine groups of listeners participated in this study. All but two of the groups consisted of 10 subjects. Five groups of subjects listened to words in quiet, two groups listened to words in noise, and two groups listened to sentences in noise. For one of the initial test conditions (isolated words, with SII = 0.4), overall performance for the words in quiet was lower than expected and testing was terminated after four subjects completed the task. One of the isolated-word groups tested with SII = 0.6 in quiet included only eight subjects. All subjects were young normal-hearing (YNH) listeners (ages 19–26 years, mean = 22.2 years), and all were paid for their participation in this experiment. All subjects were native speakers of English. All listeners had pure-tone thresholds ≤25 dB hearing level (ANSI, 2004) for all octave frequencies between 250 and 8000 Hz, normal tympanograms, and normal otoscopic examinations.

B. Stimuli

The stimuli were the standard (original) R-SPIN sentences spoken by a male talker (Kalikow *et al.*, 1977; Bilger *et al.*, 1984). The 400 R-SPIN sentences without background noise were digitized for use as PH (N = 200) and PL (N = 200) sentence stimuli. For monosyllabic word testing, the target word (always the last word in a sentence) was copied from each sentence. The boundary between the target word and the preceding word was selected to preserve the intelligibility of the target word while minimizing any audible trace of the preceding word. All edits of the word stimulus file were made at zero crossings of the waveform to minimize transients. The onset and offset of each digitally copied word was carefully examined to eliminate leading and trailing information (silence or other words) without removing any of the target word. Amplitudes were adjusted to achieve the same root-mean-square (RMS) level for all words. The intelligibility of the full set of 400 stimuli (200 words, each spoken in a predictability-high, PH, and a predictability-low, PL, context) was assessed in an open-set word-identification test in which each stimulus was presented once to a separate group of 24 YNH subjects. Based on this testing, a subset of 250 stimuli (125 words spoken in both a PH and a PL context) was selected for this study by including only words, with durations between 300 ms and 600 ms, which were correctly identified by at least 21 of the

24 pilot subjects. For testing with full sentences, the corresponding 250 sentences were used (i.e., including both PH and PL context). For monosyllabic testing, only 125 stimuli (the unique words) were used. The 125 single-word stimuli were randomly selected from PH and PL contexts in approximately equal (62/63) numbers.

We derived 20 equally important bands, each contributing 0.05 importance to the SII, from the critical-band importance function for the SPIN materials in Table B.1 of the ANSI (1997) SII standard. The values in the ANSI standard were derived from the data of Bell *et al*. (1992) and represent an average for the PL and PH monosyllables. The cumulative importance from the ANSI standard critical-band values was plotted as a function of the upper cut-off frequency of successively higher critical bands (using the critical band passband values in the SII standard). These values were then fit with the following best-fitting ($r^2 = 0.999$) three-parameter exponential:

$$I = -0.0779 + 1.1382(1 - e^{(-0.0005f)}), \tag{1}$$

where I is the cumulative importance and f is frequency in Hz. From this cumulative importance function, the 20 bands yielding equal (0.05) importance values were generated. These bands appear in Table I. Filters were constrained least-squares finite impulse response (FIR) multiband filters, 4000-order with extremely steep rejection rates (Warren *et al.*, 2004) as illustrated in Fig. 1 for a sample long-term spectrum of multiband speech in a complementary multiband notched noise.

A presentation level of 85 dB sound pressure level was used for the broad-band stimuli and the overall levels of the filtered bands were not adjusted following filtering. This relatively high level was used for comparisons with hearing-impaired listeners in other studies who require higher presentation levels for audibility.

Various patterns of band combinations were explored in this study. For materials presented in quiet and at the presentation level used (and assuming the validity of the SII importance function) one can simply add up the number of 0.05-bands included to compute the resulting SII value. Thus, conditions employing 8, 10, or 12 of the 20 bands in quiet would yield SII values of 0.4, 0.5, and 0.6, respectively, regardless of the distribution of those bands. Rather than describe the myriad of band combinations used in this study, we will present graphic illustrations of each of the band patterns alongside the speech recognition results obtained for that distribution of bands. The band patterns constructed were designed to evaluate the importance of the number of bands and their relative distribution across the spectrum for a given SII value. In addition, in some test conditions, broad-band noise was used as masker. This noise was shaped spectrally to match the long-term spectrum of the speech and matched in RMS level to the speech [for the nominal signal-to-noise ratio (SNR) of 0 dB]. In one set of test conditions, the noise was notched to be complementary to the pattern of speech bands (see example in Fig. 1), filling only the portions of the spectrum without any speech energy. Noise conditions were included for several reasons, including their common occurrence in everyday listening, their utility in increasing the difficulty of the test conditions for which high-context sentences comprised the test stimuli, and to evaluate the assumed asymptotic SNR (+15 dB) within the AI/SII framework.

TABLE I. Lower (Fc-low) and upper (Fc-high) cut-off frequencies, in Hz, for the 20 bands used in this study.

Band	1	2	3	4	5	6	7	8	9	10
Fc-low (Hz)	102	227	336	449	568	692	823	962	1111	1270
Fc-high (Hz)	226	335	448	567	691	822	961	1110	1269	1439
	11	12	13	14	15	16	17	18	19	20
Fc-low (Hz)	1441	1626	1829	2052	2303	2591	2930	3344	3884	4684
Fc-high (Hz)	1625	1828	2051	2302	2590	2929	3343	3883	4683	6307

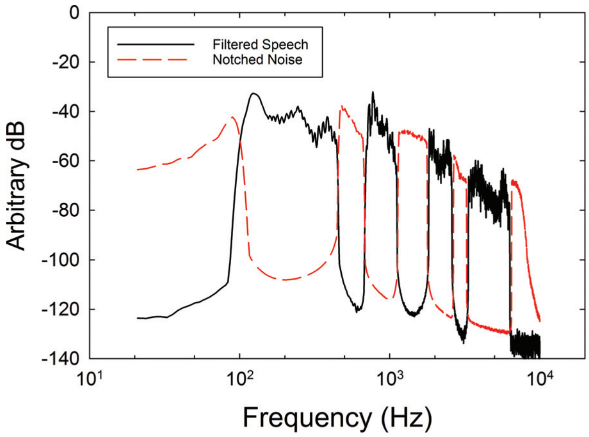

FIG. 1. An illustration of the amplitude spectrum for one of the multi-band speech stimuli (black solid) to show the steep slopes of the filters used. The dashed red line shows one example of the complementary notched noise spectrum used in some conditions.

C. Procedure

All testing was done in a single-walled sound-treated booth that met or exceeded ANSI guidelines for permissible ambient noise for earphone testing (ANSI, 1999). Stimuli were presented to the right ear, using an Etymotic Research (Elk Grove Village, IL) ER-3A insert earphone. A disconnected earphone was inserted in the left ear to block extraneous sounds. Stimuli were presented by computer using Tucker Davis Technologies (Alachua, FL) System 3 hardware (RP2 16-bit D/A converter, HB6 headphone buffer). Each listener was seated in front of a touchscreen monitor, keyboard, and mouse. On each trial, the word "LISTEN" was presented on the monitor, followed by the presentation of a word or sentence 500 ms later. The subject's task was to type the word they just heard or, for sentences, the last word heard, using the computer keyboard. Subjects were instructed to make their best guess if they were unsure. The next trial was initiated by either clicking on (with the mouse) or touching a box on the monitor labeled "NEXT." In addition to responses that were spelled correctly, homophones and phonetic spellings were scored as correct responses.

For each group of subjects, 125 isolated words or 250 sentences were presented randomly with different spectral

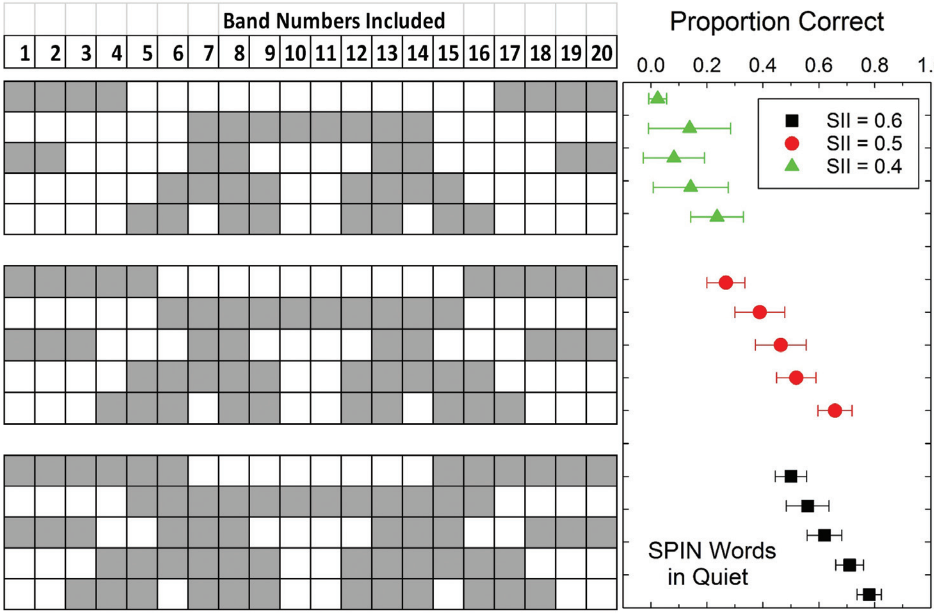

FIG. 2. Mean proportion-correct scores, plus 95% confidence interval around those means, for several multiple-band conditions for monosyllabic words presented in quiet. The grid in the left-hand portion of the figure illustrates the specific pattern of speech bands present in the stimulus. The bands are numbered 1–20 and represent the passbands provided in Table I. Each band present is shaded in grey and each represents an overall importance value of 0.05 in the SII. Results are shown for various patterns yielding SII values of 0.4 (green triangles), 0.5 (red circles), and 0.6 (black squares).

filters based on five patterns of equal-importance bands (shown in Figs. 2–5). Each unique combination of word or sentence with a spectral filter was presented only one time. There was a total of 625 stimuli for the groups of subjects presented with isolated words, and 1250 stimuli (625 PH and 625 PL sentences) for the sentence groups.

Testing consisted of five trial blocks of 125 trials for isolated words, and blocks of 250 trials (with short breaks after every 125 trials) for sentences. Testing was completed in 1 to 3 sessions of no longer than 90 min each. Each of the 625 words or 1250 sentences (125 or 250 speech tokens × 5 passband conditions) was presented once. A different word or sentence was selected randomly (from the full set of 125 or 250 tokens, respectively) on each trial and the sequence of trials cycled through a random permutation of the five passband conditions in each consecutive set of five trials. For sentences, PL and PH tokens were selected randomly, with the constraint that no target word be repeated within each half (125 trials) of a block. Prior to data collection, all subjects were presented with 20 practice trials, which included four examples of each of the five passband conditions using words that were not used in the main experiment.

III. RESULTS AND DISCUSSION

A. R-SPIN words in quiet

Figure 2 illustrates the group results for five similar band patterns at each of three SII values: 0.4 (green symbols), 0.5 (red symbols), and 0.6 (black symbols). The means and 95% confidence intervals about those means for percent-correct open-set monosyllabic-word recognition performance are plotted in the right-hand portion of Fig. 2 for each SII value. In the grid to the left in Fig. 2, the corresponding band pattern is illustrated schematically. Each of the 20 bands filled in grey in this grid illustrates the bands that included speech energy for each condition. Recall that only four subjects were tested for the SII = 0.4 condition (top grid, green symbols) due to low overall performance at this SII value. As a result, the 95% confidence intervals are noticeably larger for these data. Nonetheless, the pattern observed in the data for the SII = 0.4 condition is similar to that observed for the other two SII values (0.5 and 0.6) in Fig. 2. In general, across all three SII values, the top condition in each equal-SII group, the condition with speech energy only at the lower and upper extremes, yields the

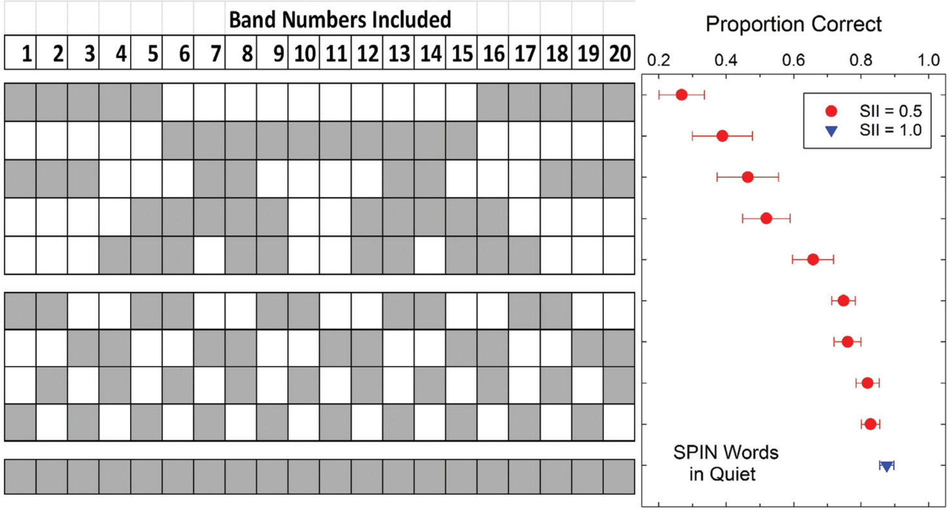

FIG. 3. Mean proportion-correct scores, plus 95% confidence interval around those means, for several multiple-band conditions for monosyllabic words presented in quiet. Figure layout as in Fig. 2. Results are shown for various patterns yielding an SII value 0.5 (red circles) and one value for SII = 1.0 (blue inverted triangle).

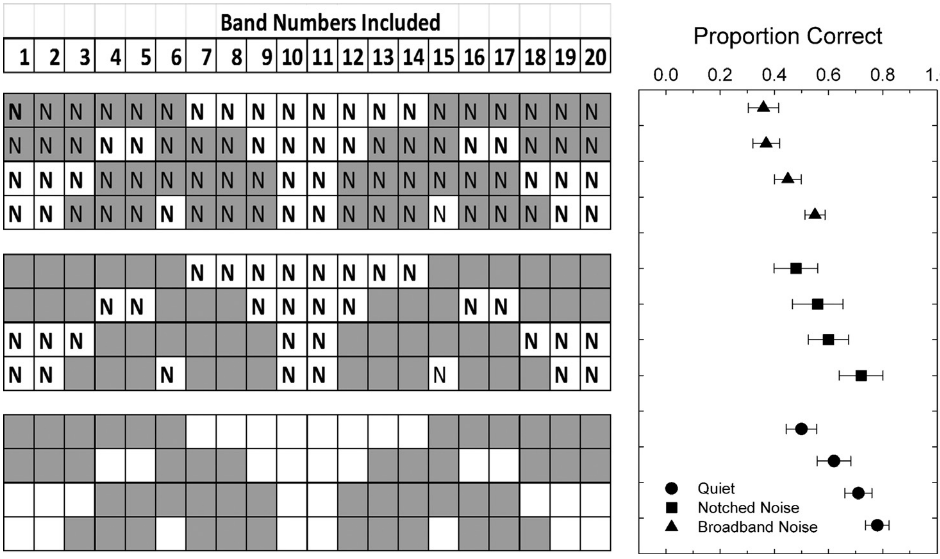

FIG. 4. Mean proportion-correct scores, plus 95% confidence interval around those means, for several multiple-band conditions for monosyllabic words presented in quiet and in noise. Figure layout as in Fig. 2. Results are shown for various patterns yielding an SII value 0.6 in quiet (bottom, circles), in spectrally notched noise (middle, squares), and in broad-band spectrally matched noise (top, triangles). "N" in the schematic illustration of the band patterns on the left is used to represent the bands filled with noise. The overall rms SNR was +15 dB for the broad-band noise.

lowest scores whereas the bottom condition in each equal-SII group with a mix of low-, mid-, and high-frequency bands, yields the highest performance. In general, moreover, performance increases as the SII increases from 0.4 to 0.5 to 0.6. Note, however, that for each of the three SII values, several significant differences in performance are observed across the conditions within that set of band patterns. Clearly, equal SII values do not yield equal performance for these R-SPIN words in quiet.

The data in Fig. 3 make this point even more strongly. Here, a wider range of conditions were explored, all with SII = 0.5. The top portion of the grid represents band patterns and data for the same SII = 0.5 condition from Fig. 2. The lower portion of the figure shows new data: means and 95% confidence intervals for four additional SII = 0.5 patterns and one SII = 1.0 condition. For SII = 0.5, across the nine band patterns examined for this SII value, R-SPIN word-recognition scores range from about 26% correct to 83% correct. Clearly, by examining the 95% confidence intervals, performance across several of the band patterns differs significantly from performance with other band patterns despite an equivalent SII value (0.5). Further, it appears that band patterns with a fairly low SII value can result in fairly good speech recognition performance, as long as they include a sampling of low-, mid-, and high-frequency bands. Note that the two band patterns for SII = 0.5 at the bottom of Fig. 3 (inclusion of either all odd- or even-numbered bands), yields scores almost equivalent to that obtained for all 20 bands (SII = 1.0).

B. R-SPIN words in noise

In the next set of conditions, we included background noise. In one case, broad-band speech-shaped noise matching the spectrum of the speech materials was presented at a SNR of 15 dB. This is the noise floor assumed in the use of a range of +15 to −15 dB SNR values for each band in the calculation of the SII (ANSI, 1997). That is, there should be no difference in SII value, and the measured performance, for the SII = 0.5 band patterns in quiet or at an SNR of 15 dB. We also examined performance when this same noise was excluded from the speech passbands and only added to the regions devoid of speech energy (as depicted previously in Fig. 1). Figure 4 provides the means and 95% confidence

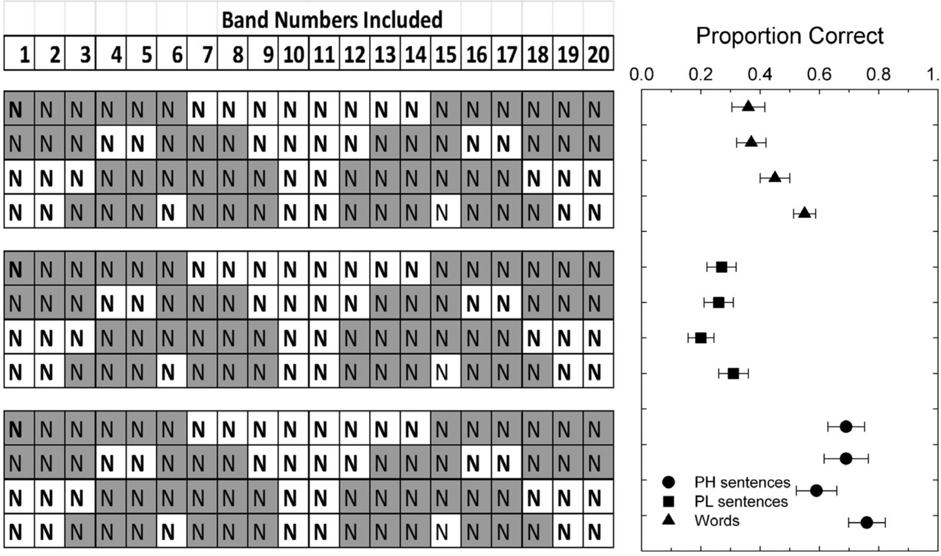

FIG. 5. Mean proportion-correct scores, plus 95% confidence interval around those means, for several multiple-band conditions for monosyllabic words (triangles) and sentences (squares, circles) presented in quiet and in noise. Figure layout as in Fig. 2. Results are shown for both predictability-low (squares) and predictability-high (circles) SPIN sentences. Broad-band spectrally matched noise was used throughout, but the SNR was +15 dB for the monosyllables and +5 dB for the sentences. "N" in the schematic illustration of the band patterns on the left is used to represent the bands filled with noise.

intervals for this set of conditions. The group of band patterns and data at the bottom of Fig. 4 represent the same data obtained in quiet for SII = 0.6, shown previously in Fig. 2. The middle group of band patterns and data illustrates the condition for the complementary notched background noise. As expected, there are no significant differences between the quiet and notched-noise pairs for corresponding band patterns. Finally, the top group of band patterns and data in Fig. 4 are for broad-band noise at an SNR of 15 dB. These results show a similar pattern of performance across band patterns as was observed in quiet and in the notched complementary-band noise, but the overall performance for each band pattern is significantly lower for the broad-band masker in several cases. This may suggest that the range of SNRs included in the SII for the R-SPIN materials may not be −15 to +15 dB, but perhaps −18 to +12 dB, as suggested for other speech materials previously by others (e.g., Kryter, 1962a; Pavlovic, 1987).

C. R-SPIN sentences in noise

Data were also obtained for the PL and PH R-SPIN sentences in broad-band speech-shaped noise at an SNR of 5 dB. The band patterns used for the words in quiet at SII = 0.6 were applied to these sentences. The results for the open-set recognition of final (key) words in PL (squares) and PH (circles) sentence contexts are shown in the right-hand portion of Fig. 5, once again as means and 95% confidence intervals. For comparison, the word-only scores for the SII = 0.6 band pattern at an SNR of 15 dB are provided in the upper portion of Fig. 5. Because the SNR values differ between the sentences (5 dB) and the words-only (15 dB), the absolute performance levels should not be compared. Rather, it is the relative pattern of results across the changes in band pattern for a given test material that is of interest. In contrast to the findings with isolated words (shown here and in Figs. 2–4), which show a tendency for performance to improve across the set of band patterns as ordered in the figures, performance was more uniform across conditions for the sentences. It is only the lower pair of each set of four band patterns (which include more mid-frequency information at the expense of the upper and lower bands) that differ significantly from one another for sentences.

D. Additional data for SII = 0.4

One of the interesting observations from the data presented at the bottom on Fig. 3 for R-SPIN words in quiet was that alternating odd-numbered or even-numbered bands over a broad frequency range, yielding an SII of 0.5, resulted in open-set recognition scores comparable to that for the full bandwidth condition or SII = 1.0. Recall that performance for the SII = 0.4 condition was so low that collection of those data was terminated after the completion of four subjects. We wondered whether a similar "packaging" of the bands for SII = 0.4 might also result in considerably higher scores. In particular, stimuli were filtered into bands 3, 5, 7, 9, 12, 14, 16 and 18 to produce a band pattern similar to all-odd or all-even bands over a broad frequency range for an SII value of 0.4. This band pattern was then applied for all speech and noise stimuli used in this study: words in quiet; words in notched noise, 15-dB SNR; words in broad-band speech-shaped noise, 15-dB SNR; and PL and PH sentences in broad-band speech-shaped noise, 5-dB SNR.

TABLE II. Comparison of means and standard deviations for speech-recognition scores (proportion correct) for groups of ten listeners for conditions yielding an SII or 0.4 or 0.6. For both SII values the overall frequency range encompassed is similar, but for SII = 0.4, the bands (8 in number) are less densely packed in that range compared to SII = 0.6 (12 bands).

	SII = 0.4		SII = 0.6	
Condition	Mean	Standard Deviation	Mean	Standard Deviation
Words, quiet	0.73	0.09	0.71	0.08
Words, notched noise	0.60	0.12	0.60	0.12
Words, broadband noise	0.41	0.05	0.45	0.08
PL sentences, broadband noise	0.19	0.08	0.20	0.07
PH sentences, broadband noise	0.49	0.13	0.59	0.11

Results are presented in Table II, along with data from conditions with a similar band pattern with an SII of 0.6 (N = 10 for each condition). Clearly, such a band pattern greatly facilitated speech-recognition performance. For word stimuli in quiet, for example, whereas performance was very low with the other band patterns examined previously, ranging from 2% to 24% (Fig. 2), performance with this band pattern was 73%. Recall that the SII was 0.4, including 8 bands, each having an importance value of 0.05, for all of these cases. Yet, performance with SII = 0.4 varied from 2% to 73% for words in quiet. This is comparable to the wide variation in performance from 27% to 83% illustrated previously in Fig. 3 for SII = 0.5. When we examined the scores for SII = 0.4 in Table II across all five test conditions, it was apparent that performance was very similar across these same five test conditions to that observed for SII = 0.6, consisting of bands 4–9 and 12–17. The means and standard deviations for this SII = 0.6 band pattern are in the far right columns of Table II. Note that the frequency region spanned by the two band patterns examined in Table II is very similar, bands 3–18 for SII = 0.4 and bands 4–17 for SII = 0.6, but that the band pattern for SII = 0.4 has fewer contiguous bands and fewer total bands (8) than the pattern for SII = 0.6 (total number of bands = 12). Yet, as shown in Table II, across a wide range of test conditions, performance is virtually identical. This is very similar to the observation made previously regarding the data in Fig. 3 for SII = 0.5 and the odd-numbered or even-numbered band patterns. Performance in these two conditions for SII = 0.5 was 82% and 83% correct whereas the full-bandwidth SII = 1.0 condition yielded only a slightly higher score of 88% correct. Clearly, those data, together with these additional data in Table II across a wider range of test conditions, indicate that it is not necessary to include the entire frequency region to achieve comparably high levels of performance. Rather, several separate spectral glimpses across that same broad frequency range can suffice for both word and sentence recognition and in quiet as well as in noise.

Comparing performance for quiet and notched-noise conditions in Table II and in Fig. 3 reveals that filling the notches between bands with noise (+15 dB SNR) had only a slight impact on performance with observed decreases ranging from 2 to 13 percentage points, depending on the pattern and the SII value. Thus, the high levels of performance achieved with relatively sparse coverage of a broad range of frequencies does not appear to be due to the use of redundant information by the auditory system in the places corresponding to the frequency regions of the notches.

IV. GENERAL DISCUSSION

The results of this series of experiments generally agree with the findings of Warren and colleagues (Warren *et al.*, 1995, 1997, 2005) and, more recently, Healy and colleagues (Healy and Warren, 2003; Apoux and Healy, 2009, 2012; Healy *et al.*, 2013). We did not, however, measure the speech-recognition performance associated with each of the individual bands comprising our band patterns to test for super-additivity, as would be required to directly compare our findings to those of Warren and colleagues (Warren *et al.*, 1995, 1997, 2005). Nor did we make use of the compound method to establish the importance of each band to permit direct comparisons to the findings of Healy and colleagues (Healy and Warren, 2003; Apoux and Healy, 2009, 2012; Healy *et al.*, 2013). However, like both sets of previous studies, as well as the earlier work of Kryter (1960, 1962b), we find that open-set speech-recognition performance for multiple bands poses serious questions about the assumptions of the AI/SII framework. In particular, speech-recognition performance over a range of iso-SII band patterns can vary by 60–70 percentage points depending upon how the band pattern is configured. This problem, moreover, manifests itself for words in quiet, words in noise, and sentences (both low and high context) in noise.

Perhaps the problem lies in the importance function assumed here to generate the 20 equally important bands. The validity of this importance function, as noted, has already been challenged by Healy *et al.* (2013), which led to the multi-band compound method of derivation. Our bandwidths for the 20 equally important bands were derived, as described above, from the cumulative importance function taken from the ANSI (1997) SII standard for the R-SPIN materials. The importance values underlying the cumulative function were obtained in the traditional way, speech passed through a series of high-pass and low-pass filters, rather than a way, such as the compound method, which makes use of speech stimuli covering a broad frequency range throughout. To determine whether this was the likely source of wide performance variation for the iso-SII conditions in this study, we compared the cumulative importance function used here to that derived by Healy *et al.* (2013). This comparison is provided in Fig. 6, where it can clearly be seen that the two cumulative importance functions are quite similar. The values derived with the compound method are slightly, but consistently, lower than those used in this study. This, however, is mainly due to the extension of the compound-method importance function (and the measurement of performance) to higher frequencies by Healy *et al.* (2013). Differences in the assumed underlying importance functions do not seem to offer a viable explanation for the wide range of performance differences observed here under iso-SII conditions. This observation also does not appear to be unique to a particular test material or condition, having been observed for words and sentences, and in quiet as well as with different patterns of background noise at different SNRs. It is also not unique to a particular SII value, having been observed at least for SII = 0.4, 0.5, and 0.6.

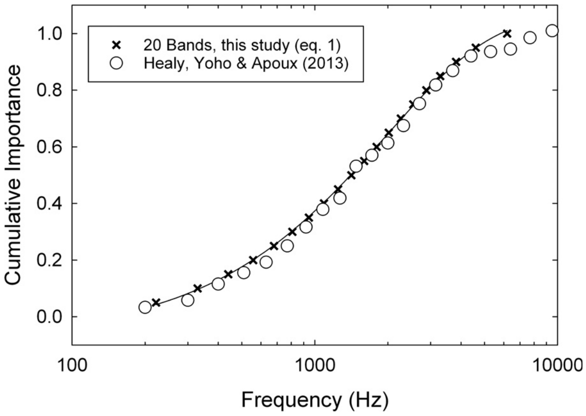

FIG. 6. The cumulative importance function from ANSI (2004) derived from the 20-band importance values for the SPIN stimuli [Eq. (1), × symbols] compared to that derived by Healy *et al.* (2013) for the same materials using the compound method (circles).

For all stimuli and test conditions, equal SII values did not yield equal performance. These findings clearly challenge the band-independence assumption of the theory underlying the SII.

Aside from implications for the validity of assumptions underlying the SII, the results in this study suggest that all spectral information across a broad frequency range is not critical for high levels of open-set speech-recognition performance. First, it is apparent that band patterns that include only the lowest and highest bands or only center bands yield the poorest performance (see Figs. 2 and 3). On the other hand, in several cases, it was demonstrated that presentation of every other band over a broad frequency range, roughly 400–4000 Hz, yielded performance comparable to that achieved with full coverage of that frequency region. Clearly, although an equal number of equally important bands may yield a constant SII value, performance for that number of bands and SII value is *not* constant. The relative redundancy or independence of the speech information in the specific bands combined is critical. Deleting half of the speech bands in an every-other-band fashion has little effect on overall speech perception because of the acoustical redundancy between the retained and deleted bands which are in close proximity to one another. Deleting an equal number of bands, all from either the low or high frequencies, on the other hand, has a strong negative impact on performance because of the independence of the deleted and retained bands drawn from opposite regions of the spectrum.

The observation that every other band can be deleted without much impact on speech-recognition performance may have implications for listeners with inner ear pathology, such as the recently observed phenomenon of "hidden hearing loss" (e.g., Liberman, 2015) following noise exposure or with advancing age. Corrupted physiological input from the periphery to the brain may be likened to missing bands in the various band patterns studied here. If so, then scattered regions of corrupted neural input from the periphery may not be detrimental for open-set speech recognition in quiet or in noise. More concentrated contiguous regions of pathology, however, could prove to be much more devastating for the listener.

ACKNOWLEDGMENTS

This work was supported, in part, by a research grant from the National Institute on Aging (R01 AG008293). We thank Kristen Baisley and Kristin Quinones for their assistance with data collection.

ANSI (**1997**). S3.5 (R2007), *American National Standard Methods for the Calculation of the Speech Intelligibility Index* (Acoustical Society of America, New York).

ANSI (**1999**). S3.1, *American National Standard Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms* (Acoustical Society of America, New York).

ANSI (**2004**). S3.6, *American National Standard Specification for Audiometers* (Acoustical Society of America, New York).

Apoux, F., and Healy, E. W. (**2009**). "On the number of auditory filter outputs needed to understand speech: Further evidence for auditory channel independence," Hear. Res. **255**, 99–108.

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

Apoux, F., and Healy, E. W. (**2012**). "Use of a compound approach to derive auditory-filter-wide frequency-importance functions for vowels and consonants," J. Acoust. Soc. Am. **132**, 1078–1087.

Bashford, J. A., Jr., Warren, R. M., and Lenz, P. W. (**2000**). "Relative contributions of passband and filter skirts to the intelligibility of bandpass speech: Some effects of context and amplitude," Acoust. Res. Letters Online **1**(2), 31–36.

Bell, T. S., Dirks, D. D., and Trine, T. D. (**1992**). "Frequency-importance functions for words in high- and low-context sentences," J. Speech Hear. Res. **35**, 950–959.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (**1984**). "Standardization of a test of speech perception in noise," J. Speech Hear. Res. **27**, 32–48.

Buss, E., Hall, J. W., and Grose, J. W. (**2004**). "Spectral integration of synchronous and asynchronous cues to consonant identification," J. Acoust. Soc. Am. **115**, 2278–2285.

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

Davis, H., and Silverman, S. R. (**1970**). *Hearing and Deafness*, 3rd ed. (Holt, Rinehart, and Winston, New York), pp. 492–495.

Fletcher, H., and Galt, R. H. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**, 89–151.

French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.

Hall, J. W., Buss, E., and Grose, J. H. (**2008a**). "The effect of hearing impairment on the identification of speech that is modulated synchronously or asynchronously across frequency," J. Acoust. Soc. Am. **123**, 955–962.

Hall, J. W., Buss, E., and Grose, J. H. (**2008b**). "Spectral integration of speech bands in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **124**, 1105–1115.

Healy, E. W., and Bacon, S. P. (**2007**). "Effect of spectral frequency range and separation on the perception of asynchronous speech," J. Acoust. Soc. Am. **121**, 1691–1700.

Healy, E. W., and Warren, R. M. (**2003**). "The role of contrasting temporal amplitude patterns in the perception of speech," J. Acoust. Soc. Am. **113**, 1676–1688.

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

Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, E. G., Eldert, E., and Benson, R. W. (**1952**). "Development of materials for speech audiometry," J. Speech Hear. Disord. **17**, 321–337.

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

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.

Kidd, G., and Humes, L. E. (**2012**). "Effects of age and hearing loss on the recognition of interrupted words in isolation and in sentences," J. Acoust. Soc. Am. **131**, 1434–1448.

Kryter, K. D. (**1960**). "Speech bandwidth compression through spectrum selection," J. Acoust. Soc. Am. **32**, 547–556.

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

Kryter, K. D. (**1962b**). "Validation of the articulation index," J. Acoust. Soc. Am. **34**, 1698–1702.

Liberman, M. C. (**2015**). "Hidden hearing loss," Sci. Am. **313**, 48–53.

Pavlovic, C. V. (**1987**). "Derivation of primary parameters and procedures for use in speech intelligibility predictions," J. Acoust. Soc. Am. **82**, 413–422.

Warren, R. M., and Bashford, J. A., Jr. (**1999**). "Intelligibility of 1/3-octavespeech: Greater contribution of frequencies outside than inside the nominal passband," J. Acoust. Soc. Am. **106**, L47–L52.

Warren, R. M., Bashford, J. A., Jr., and Lenz, P. W. (**2004**). "Intelligibility of bandpass filtered speech: Steepness of slopes required to eliminate transition band contributions," J. Acoust. Soc. Am. **115**, 1292–1295.

Warren, R. M., Bashford, J. A., Jr., and Lenz, P. W. (**2005**). "Intelligibilities of 1-octave rectangular bands spanning the speech spectrum when heard separately and paired," J. Acoust. Soc. Am. **118**, 3261–3266.

Warren, R. M., Hainsworth, K. R., Brubaker, B. S., Bashford, J. A., Jr., and Healy, E. W. (**1997**). "Spectral restoration of speech: Intelligibility is increased by inserting noise in spectral gaps," Percept. Psychophys. **59**, 275–283.

Warren, R. M., Riener, K. R., Bashford, J. A., Jr., and Brubaker, B. S. (**1995**). "Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits," Percept. Psychophys. **57**, 175–182.