Published in final edited form as:

J Exp Psychol Hum Percept Perform. 2011 February ; 37(1): 296–302. doi:10.1037/a0020411.

An alternative to the computational Speech Intelligibility Index estimates: Direct measurement of rectangular passband intelligibilities

Richard M. Warren, James A. Bashford Jr., and Peter W. Lenz

Department of Psychology, University of Wisconsin-Milwaukee, Post Office Box 413, Milwaukee, Wisconsin 53201-0413

Richard M. Warren: rmwarren@uwm.edu

Abstract

The need for determining the relative intelligibility of passbands spanning the speech spectrum has been addressed by publications of the American National Standards Institute (ANSI). When the Articulation Index (AI) standard [ANSI, S3.5, 1969] was developed, available filters confounded passband and slope contributions. The AI procedure and its updated successor, the Speech Intelligibility Index (SII) standard [ANSI, S3.5, 1997, R2007], cancel slope contributions by using intelligibility scores for partially masked highpass and lowpass speech to calculate passband importance values; these values can be converted to passband intelligibility predictions using transfer functions. However, by using very high-order digital filtering, it is now possible to eliminate contributions from filter skirts and produce rectangular passbands. Employing the same commercial recording and the same one-octave passbands published in the SII standard (Table B. 3), the present study compares Rectangular Passband Intelligibility (RPI) with SII estimates of intelligibility. The directly measured RPI differs from the computational SII predictions. Advantages resulting from direct measurement are discussed.

Keywords

Speech perception; Filtered speech; Speech intelligibility; Speech Intelligibility Index; Rectangular Passband Intelligibility

Investigators at Bell Telephone Laboratories during the first half of the twentieth century realized the importance of determining the intelligibility of speech passbands centered at different regions of the speech spectrum, and how the bands interact when heard together. Yet at that time, it was not possible to obtain direct measurement of intelligibility for a passband restricted to a particular range of frequencies. To generate a speech band it is necessary to employ filtering, and it was recognized that the transition band slopes accompanying passbands could contribute appreciably to intelligibility. Kryter (1960) estimated that contributions could be made by information as far as 30 dB down the slopes.

Correspondence concerning this article should be addressed to R.M. Warren, University of Wisconsin-Milwaukee, Department of Psychology, PO Box 413, Milwaukee, WI 53201 (rmwarren@uwm.edu).

The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/xhnp

The extent of information contained within transition band slopes is illustrated by a study in which “everyday” sentences were limited to a band consisting solely of a pair of 115 dB/octave slopes (considered very steep by analog filtering standards). When the highpass and lowpass slopes met at the common cutoff frequency of 1.5 kHz, the resulting band had an intelligibility of 77% (Warren, Riener, Bashford, & Brubaker, 1995).

When pushed to high orders, the analog filter banks available to the early investigators produced unacceptable distortions; these include frequency-dependent time delays resulting in asynchronies that increase with spectral separation, as well as a tendency to oscillate with rapid changes in amplitude, resulting in ringing and spectral smearing. Therefore, it was not possible for these investigators to eliminate contributions from the slopes accompanying passbands, or to determine the extent of their relative contributions. However, their studies (e.g., by Fletcher, 1929; Fletcher & Galt, 1950; and French & Steinberg, 1947) led to a rather complex computational procedure for canceling contributions from filter slopes. This procedure was used to estimate passband intelligibilities by the Articulation Index (AI) published by the American National Standards Institute (ANSI S3.5, 1969, R1986), and the Speech Intelligibility Index (SII) published as its successor (ANSI S3.5, 1997, R2007). The filter slopes required for deriving these standards can be quite shallow: the instructions for deriving the AI state, “For the purpose of this standard the slopes of the filter skirts should be equal to or greater than 18 dB/octave” (ANSI S3.5, 1969, R1986). The indices provide relative “importance” values for individual passbands covering the speech spectrum. These importance values are considered to be “highly correlated” with intelligibility (see page 1 of the SII). They are derived from data obtained by combining the intelligibility scores of systematically graded cutoff frequencies for highpass and for lowpass speech. Ceiling scores that would otherwise occur for those bands having high lowpass cutoffs and low highpass cutoffs are avoided by adding noise that partially masks the speech. The importance values of individual bands, as well as the summed importance values of contiguous bands, can be converted to intelligibility estimates by application of appropriately designed transfer functions. However, as will be discussed, it is possible to obtain direct measurements of intelligibility for passbands heard singly and together. The present study compares these measured intelligibilities with intelligibility predictions based upon the SII.

Finite impulse response (FIR) digital filtering can eliminate contributions from flanking transition bands while leaving passbands intact and maintaining synchrony of passbands across the spectrum. The degrading of intelligibility by ringing and spectral smearing can also be avoided.

In order to determine the steepness of slopes required to eliminate their contributions, FIR filtering was used by Warren, Bashford, and Lenz (2004) to produce a series of increasingly steep slopes for a 1/3-octave band of “everyday” sentences centered on 1.5 kHz. It was found that reaching asymptotic intelligibility required FIR lengths greater than 7,000 taps (44.1 kHz sampling) that produced highpass and lowpass slopes steeper than 2.8 dB/Hz (averaging approximately 5,000 dB/octave). Even with transition bands averaging 1,200 dB/octave, over 30% of the intelligibility of a 1/3-octave passband was contributed by its slopes.

The ability to separate contributions from passbands and transition bands indicated that it might be possible to construct an alternative to the computational AI/SII procedure. An initial study was undertaken to demonstrate the feasibility of directly measuring the intelligibility of passbands spanning the speech spectrum.

Warren, Bashford, and Lenz (2005), used FIR lengths of 8,000 taps to generate slopes of 3.2 dB/Hz. These slopes were used to produce six 1-octave rectangular bands of sentences that

spanned the speech spectrum (center frequencies ranging from 0.25 to 8 kHz). The highest intelligibility score (55%) for these bands was obtained for the 2-kHz band, with intelligibility decreasing monotonically with increasing distance from 2 kHz. In order to ensure that 8,000-tap filtering was sufficient to remove transition band contributions, as well as to ensure that high-order filtering did not introduce artifacts interfering with intelligibility, the filter taps were increased to 20,000, producing slopes of 8 dB/Hz; no significant effect upon the intelligibility of any of the passbands was produced by the additional filtering. The Conclusion section of this paper stated that a comparison of intelligibility values obtained by using rectangular passbands with those obtained using the standard procedure could be accomplished "...by obtaining direct intelligibility measurements for the same speech stimuli and passbands that had been employed for estimates based upon the SII." The present study implemented this suggestion. Rather than employing sentences (as used in the Warren et al. 2005 study), the same word lists employed for data published in the SII standard (ANSI S3.5, 1997, R2007) were used.

The SII Table B.3 entitled, "Octave band importance functions for various speech tests," contains data for six different speech tests (composed of nonsense syllables, monosyllabic words, and sentences), and provides importance values for their one-octave passbands covering the effective speech spectrum (center frequencies 0.25, 0.5, 1, 2, 4, and 8 kHz). One of the importance functions listed in the table consists of the data reported in a study by Studebaker and Sherbecoe (1991) that was based upon a commercially available recording of a monosyllabic word list. This study also provided a transfer function for converting their importance values to intelligibility estimates.

The present study employed a copy of the same recording used by Studebaker and Sherbecoe. The speech was digitally filtered to produce six 1-octave passbands with effectively vertical slopes having the same center frequencies listed in SII Table B.3. Listeners presented with these rectangular passbands were able to provide intelligibility scores directly, without the intermediate stages of importance values and transfer functions required by the AI/SII procedure.

This Rectangular Passband Intelligibility (RPI) procedure was employed in Experiment 1 to obtain intelligibility scores for five separate groups of 30 participants. Each group was presented with one of five presentation levels ranging from 45 to 70 dBA. Experiments 2, 3a, and 3b taken together were designed to determine the intelligibility scores for all 15 possible pairings of the six 1-octave passbands.

Method

Participants

The total of 242 participants employed in Experiments 1 through 3b were native English speakers, who were undergraduate participants enrolled in classes at the University of Wisconsin-Milwaukee. They received payment in exchange for their participation. All participants were native monolingual English speakers who reported having no hearing problems. They had normal bilateral hearing, as measured by pure tone thresholds of 20 dB HL or better at octave frequencies from 250 to 8,000 Hz.

Stimulus Processing

The stimuli consisted of the CID (Central Institute for the Deaf) W-22 monosyllabic word lists (Hirsh et al., 1952), along with a carrier phrase ("You will say...") that had been recorded by a male speaker on phonograph disks by Technisonic Studios. As stated in the introduction, a copy of this commercially available recording had been used previously in

constructing SII-based frequency importance functions and intelligibility estimates for the same 1-octave bands.

The phonograph recording was transferred into digital format (16-bit, 44.1 kHz sampling frequency) using a Fisher MT-640 turntable, with the output passed through a Panasonic Ramsa WR-133 mixing board and into a DigiDesign Audiomedia 3 PCI soundcard running on a Macintosh G3 computer equipped with DigiDesign Sound Designer II software. Occasional transient artifacts (“pops” and “clicks”) occurring with the phonograph playback were removed using visual and auditory cues with the help of Sound Designer II waveform editing capabilities. There were 200 stimuli (test word plus carrier phrase) in the original recording: 180 were selected for use as experimental stimuli, and of the remainder, ten were used as practice stimuli. The slow-peak levels of the stimuli were matched to correct for level differences in the original recording, and to ensure maximum signal-to-noise ratio when filtered. Measuring signal level was accomplished using a flat-plate coupler in conjunction with a Brüel and Kjaer Model 2230 sound-level meter set at A-scale weighting (as were all amplitude measurements in this study).

Filtering to produce the six 1-octave bands centered at 0.25, 0.5, 1, 2, 4, and 8 kHz was accomplished using the *fir1* function in MATLAB 5.2.1 running on a Macintosh G4 computer. Slopes of 3.2 dB/Hz were produced by two successive stages of 4,000-tap filter lengths (44.1 kHz sampling). The two stages were used to ensure that all stopband frequencies were well below threshold. Finally, the slow-peak levels of the individual rectangular-band stimuli were matched to ± 0.2 dB.¹ Mixing of the bandpass signals was done on a Macintosh G3 using Sound Designer II software. Figures 1 and 2 compare the speech spectra for passband and broadband stimuli.

General Procedure

Before receiving each stimulus condition in an experiment, participants became acquainted with the effects of filtering by listening to the set of 10 practice stimuli. These were first presented broadband, and then filtered in the same manner as the test stimuli (carrier phrase plus keyword) that followed. The test stimuli in each experiment were grouped into sets, which were presented in a fixed order to all listeners, and the different filtering conditions were applied to the sets pseudorandomly, with the single restriction that each filtering condition was applied an equal number of times to each set across listeners. Each participant served in only one of the eight experimental groups, and each participant heard the keywords only once during the experiment. None of the participants had served in any previous experiment with filtered speech.

Participants were tested individually while seated with the experimenter in an audiometric chamber. The stimuli were delivered diotically through Sennheiser HD 250 Linear II headphones. Participants were told to repeat the word following the stimulus carrier phrase “You will say...” as accurately as possible, and to guess if unsure. The number of target words reported correctly was recorded by the experimenter. Testing occurred in a single session lasting approximately 30 minutes.

¹In the unfiltered broadband stimulus, the six individual 1-octave bands did not have matched levels, but had average amplitudes that differed over a range of 23 dB. It might be thought that since adjusting all bands to the same average amplitude would produce deviations from the normal spectral profile, the “unnatural” relative amplitudes would reduce the intelligibility of paired bands. However, it has been reported that there were no changes in the dual-band intelligibility scores when the relative amplitudes of two rectangular bands were varied over a range of 30 dB (Warren, Bashford, & Lenz, 2003).

Experiments

Experiment 1: Intelligibilities of the six 1-octave rectangular bands (center frequencies 0.25 to 8 kHz) presented at five different levels—Bashford, Warren, and Lenz (2005) had reported that unlike broadband speech, the intelligibility of a rectangular speech band declines at levels of 65 dB and higher. For that reason it was decided to determine the effect of amplitude on the intelligibility of this recording and to conduct the other experiments at what was determined to be an optimal level. A total of 150 participants (five separate groups of 30 participants) each received 180 keywords (six sets of 30 keywords) at peak-rms levels of either 45, 50, 55, 60, or 70 dB. The six sets of stimuli presented at each level were reduced to 1-octave bands having center frequencies of 0.25, 0.5, 1, 2, 4, and 8 kHz (the same frequencies that had been used by the SII for reporting importance values). Based on the results obtained in Experiment 1, the subsequent experiments were conducted with rectangular bands at 50 dB.

Experiment 2: Intelligibilities of pairs consisting of adjacent 1-octave bands—Thirty participants each received 180 keywords (5 sets of 36 keywords). Each set consisted of one of the five contiguous pairings of the 1-octave bands. These pairs were: 0.25 + 0.5 kHz, 0.5 + 1 kHz, 1 + 2 kHz, 2 + 4 kHz, and 4 + 8 kHz.

Experiment 3a: Pairs separated by spectral gaps of one octave—Thirty-two participants each received 180 keywords (4 sets of 45 keywords). Each set consisted of one of the four pairings of the 1-octave bands in Experiment 1. These pairs separated by a 1-octave gap were: 0.25 + 1 kHz, 0.5 + 2 kHz, 1 + 4 kHz, and 2 + 8 kHz.

Experiment 3b: Pairs separated by spectral gaps of two, three, and four octaves—Thirty participants each received 180 keywords (6 sets of 30 keywords). Each set consisted of pairings of the 1-octave bands in Experiment 1. These sets consisted of the three pairs of 1-octave bands separated by gaps of two octaves (0.25 + 2 kHz, 0.5 + 4 kHz, 1 + 8 kHz), the two pairs separated by three octaves (0.25 + 4 kHz and 0.5 + 8 kHz), and the single pair separated by four octaves (0.25 + 8 kHz).

Results

Experiment 1

Figure 3 shows the average intelligibility scores (defined as the percent correct repetition of keywords) and standard errors for the six 1-octave bands at each presentation level. It can be seen that intelligibility was highest for the 2-kHz band for all five presentation levels, and that the scores for the higher and lower bands decreased with the extent of their separation from the 2-kHz band. For the 2-kHz band (and for the two bands with the next highest intelligibilities) there was a decrease in scores for bands presented at levels above 60 dB. This decline may seem surprising, since broadband speech can remain fully intelligible at levels above 90 dB. The drop in intelligibility for narrowband speech at relatively low signal levels appears to be due to a lack of lateral suppression normally produced in broadband speech by flanking spectral components: Bashford, Warren, and Lenz (2005) reported that intelligibility of rectangular speech bands at levels of 65 and 75 dB was increased substantially by the addition of flanking stochastic noise bands.

The data were subjected to a mixed two-factor (speech level X center frequency) analysis of variance, which yielded significant main effects of presentation level [$F(4,145)=5.27$, $p<0.001$, $\eta_p^2=0.127$] and speech band center frequency [$F(5,725)=955.2$, $p<0.0001$, $\eta_p^2=0.863$], and a significant interaction of those factors [$F(20,725)=1.65$, $p<0.05$, $\eta_p^2=0.006$]. Tukey HSD tests for the main effect of presentation level revealed ($p<0.05$)

that repetition accuracy was lower for the 70 dB level than for the 45, 50, and 55 dB levels, which did not differ significantly (the 60 dB level did not differ from any of the other levels). Tukey tests for the main effect of speech band center frequency revealed ($p < 0.05$) that performance was highest for the 2-kHz band and declined increasingly for the 1-kHz, 0.5-kHz, and 4-kHz bands. Intelligibility was lowest, and equivalent, for the 0.25 kHz and 8-kHz bands. Simple-effects tests for the effect of speech-band center frequency were significant at all presentation levels [$F(5,145) \geq 144.9$, $p < 0.0001$, $\eta_p^2 \geq 0.861$], and subsequent Tukey tests indicated that the interaction effect was probably due to a nonsignificant difference between the 0.5-kHz and 4-kHz bands, that was obtained only at the 70 dB presentation level.

Experiment 2

The average paired intelligibility scores and standard errors for the five adjacent 1-octave bands are shown in Table 1 (along with their summed individual scores). The data for the paired intelligibilities were subjected to a repeated measures analysis of variance, which yielded a significant effect of band composition [$F(4,116) = 345.1$, $p < 0.0001$, $\eta_p^2 = 0.923$]. Subsequent Tukey HSD tests indicated ($p < 0.05$) that intelligibility was highest for the 1 + 2 kHz pair, next highest and equivalent for the 0.5 + 1 kHz pair and 2 + 4 kHz pair, next highest for the 0.25 + 0.5 kHz pair, and lowest for the 4 + 8 kHz pair.

Experiment 3a

The average intelligibility scores and standard errors for the four pairs of 1-octave bands separated by one octave are shown in Table 1 (along with their summed individual scores). The data for the paired intelligibilities were subjected to a repeated measures analysis of variance, which yielded a significant effect of band composition [$F(3,93) = 91.7$, $p < 0.0001$, $\eta_p^2 = 0.747$]. Subsequent Tukey HSD tests indicated ($p < 0.05$) that intelligibility was highest for the 0.5 + 2 kHz pair, next highest, and equivalent, for the 1 + 4 kHz and 2 + 8 kHz pairs, and lowest for the 0.25 + 1 kHz pair.

Experiment 3b

The average intelligibility scores and standard errors for the six pairs of 1-octave bands separated by more than one octave are shown in Table 1 (along with their summed individual scores). The data for the paired intelligibilities were subjected to a repeated measures analysis of variance which yielded a significant effect of band composition [$F(5,145) = 214.1$, $p < 0.0001$, $\eta_p^2 = 0.881$]. Subsequent Tukey HSD tests indicated ($p < 0.05$) that intelligibility was highest, and equivalent, for all band pairs separated by two octaves, was next highest for the 0.5 + 8 kHz pair, and lowest for the 0.25 + 4 kHz and 0.25 + 8 kHz pairs, which did not differ significantly.

The black diagonal boxes in Table 1 show the intelligibilities and standard errors of the six 1-octave rectangular bands presented singly at 50 dB, and the boxes to the left of the black diagonal show the intelligibilities and standard errors for each of their pairwise combinations determined in Experiments 2, 3a, and 3b. The square boxes to the right of the black diagonal show that each of the dual band pairings exhibits a hyperadditivity.

Summary and Discussion of Experiments Employing Direct Intelligibility Measurement

Experiment 1 demonstrated that there were no significant changes in intelligibility scores for the rectangular 1-octave bands of the CID word lists over the range from 45 to 60 dB. However, at 70 dB a decrease in intelligibility occurred at mid-range frequencies in keeping

with other studies employing rectangular bands (Bashford, Warren, & Lenz, 2005; Warren, Bashford, & Lenz, 2005). Figure 3 and Table 1 show that the octave band centered on 2-kHz had the highest intelligibility. Among other cues, this band had the advantage of lying in the frequency range of two formants, F2 and F3 (see Petersen & Barney, 1952). Intelligibility decreased monotonically for the word lists as the distance between the higher and lower bands increased from the position of the 2-kHz band, as was also reported for sentences in the Warren et al., 2005 study. The similarities observed for direct intelligibility measures of sentence and word lists resemble those appearing for importance values in SII Table B.3.

Comparisons of Direct Intelligibility Measurements with SII Intelligibility Predictions

Although the present study used the same recording of word lists and the same passbands employed for the SII, the procedures employed for determining the intelligibilities of speech bands differed greatly.

As discussed briefly in the introduction, when the program to determine the intelligibility of passbands was being developed, there was no means to measure their intelligibilities directly: Passband intelligibilities would be confounded with contributions from transition bands produced by the analog filters available at that time. In order to determine passband intelligibilities, it was necessary to find some way to eliminate out-of-band contributions. The procedure used by both the Articulation Index (AI) [ANSI S3.5, 1969, R1986] and the Speech Intelligibility Index (SII) [ANSI S3.5, 1997, R2007] cancelled transition-band contributions of analog filters by combining intelligibility scores of highpass and lowpass speech partially masked by noise. This cancellation permitted the procedure to be used with transition-band slopes as shallow as 18 dB/octave (see ANSI S3.5, 1969, R1986, p. 7). Of course, slopes steeper than this were available at that time -- for example, the paper entitled "Validation of the articulation index," Kryter (1962) employed slopes of 60 dB/octave.

The application of the AI/SII procedure is exemplified by the study of Studebaker and Sherbecoe (1991). Prior to gathering their experimental data used for calculating the band importance values incorporated in Table B.3 of the SII, Studebaker and Sherbecoe familiarized individuals in their single group of eight graduate students with the test words by presenting them with the entire set of stimuli heard broadband without added noise. During the series of subsequent experimental sessions, each lasting approximately 2 hours, intelligibility scores were obtained from individuals presented with the same keywords under approximately 300 conditions in which the highpass and lowpass speech bands were presented along with noise.

Without the partial masking of speech by the addition of high levels of noise, ceiling effects would prevent calculation of passband importance values. The noise employed by Studebaker and Sherbecoe was not highpass or lowpass filtered to match the filtered speech, but was broadband Gaussian noise shaped to match the average long-term spectrum of the unfiltered speech. The signal/noise ratio employed with the SII procedure varied in 2 dB steps from -10 to +8 dB depending upon the level required to partially mask individual highpass and lowpass speech stimuli. Some of these ratios would be expected to produce a quite considerable reduction in intelligibility. As pointed out by Steeneken and Houtgast (2002), the presence of these high noise levels can produce appreciable changes in the relative contributions made by different classes of phonemes. Thus, the presence of noise may require weighting functions that vary with the phoneme content of the particular speech sample. By eliminating the presence of added noise in the present study, such phoneme-selective effects are avoided.

As detailed by Studebaker and Sherbecoe on pages 430-434, the intelligibility scores for the partially masked highpass and lowpass bands were combined and processed using several successive stages to produce a “relative” transfer function, which was then used to generate the passband importance values. Unlike passband intelligibilities, adjacent passband importance values were considered to be additive. The final processing stage involved the use of several empirical constants to derive the published “absolute” transfer function. This second transfer function can convert importance values to SII intelligibility estimates for single or conjoined adjacent passbands.

As can be seen in Figure 4, both the SII estimates and direct measurements yielded passband intelligibility scores that were highest at 2 kHz, and exhibited monotonic decreases with distance from 2 kHz, with the exception of the anomalous SII 0.25 kHz band.

Figure 5 shows that SII estimates and the measured intelligibilities of paired adjacent bands differ, but are in somewhat closer agreement than are the scores for single passbands shown in Figure 4.

Figures 4 and 5 provide a means of comparing the results obtained by the two methods. It can be seen that the intelligibilities predicted by the standard SII procedure are higher than those obtained by the rectangular passband procedure used in the present study. This can be attributed to the SII participants’ greater familiarity with the keywords, resulting from their repeated presentation under a large number of highpass and lowpass conditions, as well as having the participants hear all the keywords broadband without added noise before the delivery of the experimental stimuli. In contrast, participants in the present study heard each target word only once.

An Alternative to the SII: The Rectangular Passband Intelligibility (RPI) Procedure

The use of rectangular passbands can make it possible to dispense with the intelligibility measurements of partially masked highpass and lowpass speech and eliminates the need for calculation of passband importance values and the construction of transfer functions required by the SII.

The RPI procedure permits the direct measurement of the intrinsic intelligibilities of passbands spanning the speech spectrum when heard singly, and when heard together, whether the bands are adjacent or separated. It can also be used to measure the effects produced by extraneous sounds and other factors degrading intelligibility by employing a two-stage process. In the first of these stages, the intelligibility of single or multiple rectangular passbands of desired bandwidth(s) can be measured under optimal listening conditions (i.e., eliminating factors known to reduce intelligibility). In the second stage, any factor or factors that modify intelligibility (such as noise, reverberation, and transmission distortion) can be introduced, and the resulting intelligibility measured: The extent of changes produced by these factors can be obtained by subtraction.

To summarize: The RPI procedure has several advantages over the SII procedure. These include:

1. The RPI procedure is much simpler than that of the SII. It eliminates the need for determining the intermediate stage of passband importance values, and the calculation of two transfer functions (one function to go from the intelligibility-based data for highpass and lowpass speech to passband importance values, the other to go from the importance values to intelligibility estimates).

2. The RPI procedure can measure the intelligibility of passbands heard in quiet without being degraded by extraneous sounds or distortions. The values obtained are of interest both for theory and for practical applications. Although the SII procedure can estimate these values, it is based upon and is designed to estimate the intelligibility of speech passbands heard under conditions that reduce intelligibility. When desired, the RPI procedure also can be employed under conditions that reduce intelligibility.
3. The RPI procedure can provide intelligibility measurements for single as well as any combination of passbands, whether adjacent or separated.
4. The RPI procedure has the intrinsic advantage of direct measurement over estimation or prediction.

It is hoped that the RPI procedure will be of use in conducting future basic research as well as contributing to the development of practical applications that include auditory prostheses, telephony, and architectural acoustics.

Acknowledgments

The project described was supported by Grant Number R01DC000208 from the National Institute On Deafness And Other Communication Disorders. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute On Deafness And Other Communication Disorders or the National Institutes of Health.

References

- ANSI S3.5, 1969. Methods for the Calculation of the Articulation Index. American National Standards Institute; New York: Reaffirmed, 1986.
- ANSI S3.5, 1997. Methods for the Calculation of the Speech Intelligibility Index. American National Standards Institute; New York: Reaffirmed, 2007.
- Bashford JA Jr, Warren RM, Lenz PW. Enhancing intelligibility of narrowband speech with out-of-band noise: Evidence for lateral suppression at high-normal intensity. *Journal of the Acoustical Society of America* 2005;117:365–369. [PubMed: 15704428]
- Fletcher, H. *Speech and Hearing*. London: McMillan; 1929.
- Fletcher H, Galt RH. The perception of speech and its relation to telephony. *Journal of the Acoustical Society of America* 1950;22:89–151.
- French NR, Steinberg JC. Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America* 1947;19:90–119.
- Hirsh II, Davis H, Silverman SR, Reynolds EG, Eldert E, Benson RW. Development of materials for speech audiometry. *Journal of Speech Hearing Disorders* 1952;17:321–337.
- Kryter KD. Speech bandwidth compression through spectrum selection. *Journal of the Acoustical Society of America* 1960;32:547–556.
- Kryter KD. Validation of the articulation index. *Journal of the Acoustical Society of America* 1962;34:1698–1702.
- Peterson GE, Barney HL. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 1952;24:175–184.
- Steeneken HJM, Houtgast T. Phoneme-group specific octave-band weights in predicting speech intelligibility. *Speech Communication* 2002;38:399–411.
- Studebaker GA, Sherbecoe RL. Frequency-importance and transfer functions for recorded CID W-22 word lists. *Journal of Speech Language and Hearing Research* 1991;34:427–438.
- Warren RM, Bashford JA Jr, Lenz PW. Intelligibility of dual rectangular speech bands: Implications of observations concerning amplitude mismatch and asynchrony. *Speech Communication* 2003;40:551–558.

- Warren RM, Bashford JA Jr, Lenz PW. Intelligibility of bandpass filtered speech: Steepness of slopes required to eliminate transition band contributions. *Journal of the Acoustical Society of America* 2004;115:1292–1295. [PubMed: 15058351]
- Warren RM, Bashford JA Jr, Lenz PW. Intelligibilities of 1-octave rectangular bands spanning the speech spectrum when heard separately and paired. *Journal of the Acoustical Society of America* 2005;118:3261–3266. [PubMed: 16334905]
- Warren RM, Riener KR, Bashford JA Jr, Brubaker BS. Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits. *Perception & Psychophysics* 1995;57:175–182. [PubMed: 7885815]

Figure 1.

Long-term average spectrum of 180 CID W-22 monosyllables, with carrier phrase, subjected to two successive stages of 4,000-order FIR filtering that produced a 1-octave rectangular band centered at 2-kHz. This plot for the filtered band is superimposed upon the corresponding portion of the spectrum of the parent broadband stimulus, but with a +6 dB offset of the passband to facilitate comparison of the two spectra. The passband's stopband is also shown, and exhibits no spectral ripples. The plots were generated offline via discrete Fourier transform (DFT) with linear averaging of spectra using a 64,000-point (0.7 Hz resolution) Blackman window across the entire set of digitally recorded stimuli. Metric Halo Laboratories, Inc. produced the DFT software.

Figure 2.

Averaged spectra of the 2-kHz passband of one of the target words, “jam”. The word was contained in the final 600 ms excised from its carrier phrase. In panel A, the bandpass spectrum of the word is superimposed upon the corresponding time-aligned broadband spectrum, but with a +18 dB offset to facilitate comparison of the spectra. In panel B, the bandpass spectrum is superimposed upon the broadband spectrum with no offset in level. Panel C permits a more stringent evaluation of the spectral match. Spectra were matched in level as in panel B, but the bandpass stimulus was phase inverted and added to the broadband stimulus. The resulting cancellation was generally greater than 60 dB, demonstrating that there was no appreciable distortion produced by the 8,000 filter taps.

Figure 3.

Experiment 1: Mean percent intelligibility scores and standard errors as measured for the six 1-octave passbands spanning the speech spectrum, presented at five different levels from 45 to 70 dB (N=30 at each level). For further details, see text.

Figure 4.

Experiment 2: Mean percent intelligibility scores as measured for the six 50 dB 1-octave passbands spanning the speech spectrum (N=30), compared with the predicted intelligibility scores for the same passbands based on the Speech Intelligibility Index. For further details, see text.

Figure 5.

Experiments 3a and 3b: Mean percent intelligibility scores as measured for the five adjacent pairings of the six 1-octave passbands spanning the speech spectrum (N=30), compared with the predicted intelligibility scores for the same pairs of passbands based on the Speech Intelligibility Index (SII). For further details, see text.

Intelligibilities of individual rectangular 1-octave passbands spanning the speech spectrum, and their deviations from additivity when paired. The matrix summarizes the results obtained in Experiments 1, 2, 3a, and 3b for bands presented at 50 dB. The dark diagonal boxes show the percent intelligibility scores and standard errors (in parentheses) for the six 1-octave rectangular passbands when heard alone. The boxes to the left of the diagonal show the dual-band intelligibility scores and standard errors for all 15 possible pairings of the six bands. The boxes to the right of the major dark diagonal show that each of the 15 pairs of passbands exhibits a more-than-additive intelligibility: Within each of the square boxes, the upper left shows the intelligibility of the 15 pairings when both bands were heard together, and the lower right shows the sum of the intelligibilities of these bands when heard separately. To take as an example, the 2 kHz and the 4 kHz passbands heard together had an intelligibility of 64.3%; when these bands were heard separately, their intelligibilities were 47.0% and 9.3% respectively, which yields a sum of 56.3%. It can be seen that the intelligibility of the bands when paired was always greater than the summed intelligibilities of the constituent individual bands.

Table 1

	.25 kHz	.5 kHz	1 kHz	2 kHz	4 kHz	8 kHz
8 kHz	8.4% (1.4)	40.4% (1.6)	50.9% (2.1)	56.4% (1.6)	14.1% (2.0)	2.1% (0.5)
4 kHz	12.0% (1.3)	53.7% (1.8)	60.0% (2.1)	64.3% (2.0)	9.3% (1.4)	14.1% 11.4%
2 kHz	56.4% (1.4)	86.2% (1.9)	83.7% (1.4)	47.0% (1.7)	64.3% 56.3%	56.4% 49.1%
1 kHz	46.5% (1.8)	67.0% (1.6)	26.9% (1.3)	83.7% 73.9%	60.0% 36.2%	50.9% 29.0%
.5 kHz	26.0% (1.3)	18.8% (1.2)	67.0% 45.7%	86.2% 65.8%	53.7% 28.1%	40.4% 20.9%
.25 kHz	0.4% (0.4)	26.0% 19.2%	46.5% 27.3%	56.4% 47.4%	12.0% 9.7%	8.4% 2.5%
DUAL-BAND INTELLIGIBILITY & (SE)						
DUAL-BAND SYNERGY						