

Measuring the critical band for speech^{a)}

Eric W. Healy^{b)}

*Department of Communication Sciences and Disorders, Arnold School of Public Health,
University of South Carolina, Columbia, South Carolina 29208 and Psychoacoustics Laboratory,
Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102*

Sid P. Bacon

*Psychoacoustics Laboratory, Department of Speech and Hearing Science, Arizona State University, Tempe,
Arizona 85287-0102*

(Received 25 June 2003; revised 15 November 2005; accepted 26 November 2005)

The current experiments were designed to measure the frequency resolution employed by listeners during the perception of everyday sentences. Speech bands having nearly vertical filter slopes and narrow bandwidths were sharply partitioned into various numbers of equal log- or ERB_N-width subbands. The temporal envelope from each partition was used to amplitude modulate a corresponding band of low-noise noise, and the modulated carriers were combined and presented to normal-hearing listeners. Intelligibility increased and reached asymptote as the number of partitions increased. In the mid- and high-frequency regions of the speech spectrum, the partition bandwidth corresponding to asymptotic performance matched current estimates of psychophysical tuning across a number of conditions. These results indicate that, in these regions, the critical band for speech matches the critical band measured using traditional psychoacoustic methods and nonspeech stimuli. However, in the low-frequency region, partition bandwidths at asymptote were somewhat narrower than would be predicted based upon psychophysical tuning. It is concluded that, overall, current estimates of psychophysical tuning represent reasonably well the ability of listeners to extract spectral detail from running speech.

© 2006 Acoustical Society of America. [DOI: 10.1121/1.2162176]

PACS number(s): 43.71.Es, 43.71.An, 43.66.Fe [PFA]

Pages: 1083–1091

I. INTRODUCTION

The division of the auditory spectrum into a series of critical bands by auditory filters is a primary stage of auditory processing. The importance of the critical band (CB) is underscored by its relevance to a wide range of auditory phenomena, including masking and loudness. The psychophysical techniques currently used to measure the critical band (e.g., Patterson, 1976; Patterson and Moore, 1986; Glasberg and Moore, 1990) employ simple tone and noise stimuli, which allow precise control over the input stimulus. However, accurate estimates of frequency tuning become more difficult to obtain as stimulus conditions are made more complex.

An important question is whether the CB measured using relatively simple stimuli represents the frequency resolution underlying the perception of complex sounds such as speech. One technique that has been employed to examine the normal frequency resolution employed when processing speech is to present listeners with stimuli having reduced frequency specificity and to examine resulting reductions in performance. However, performance on broadband speech in quiet can be quite good despite extensive spectral smearing (Baer and Moore, 1993; but also see Boothroyd *et al.*, 1997).

Only when speech was presented in background noise did performance suffer with spectral smearing beyond the width of a psychophysical CB (e.g., Celmer and Bienvenue, 1987; ter Keurs *et al.*, 1992, 1993; Baer and Moore, 1993).

Indeed, the minimum frequency resolution required for normal-hearing listeners to understand broadband speech is extremely low: Shannon *et al.* (1995) demonstrated high levels of sentence recognition under conditions in which spectral information was reduced to that conveyed by only three or four broad amplitude-modulated (AM) noise bands. In accord with results involving spectral smearing, broadband speech presented in background noise required additional spectral detail and therefore additional AM frequency channels to reach ceiling intelligibility values (Dorman *et al.*, 1998).

Due to the robust nature of speech and the multiplicity of cues available to the listener, the amount of spectral smearing required to reduce broadband speech intelligibility from ceiling values cannot tell us about the normal frequency resolution employed when processing speech, or the maximum spectral resolution listeners can employ to gain useful information from the signal. This is also true when examining the smallest number of AM carrier bands capable of providing intelligibility of broadband speech. Instead, these measures provide useful information concerning robustness or the resistance of speech to spectral degradation. As smearing or spectral reduction becomes more and more severe,

^{a)}Portions of this work were presented at the 139th Meeting of the Acoustical Society of America, Atlanta, GA, 30 May–3 June 2000 [J. Acoust. Soc. Am. **107**, 2914 (2000)].

^{b)}Electronic mail: ewh@sc.edu

these manipulations provide information concerning the minimum spectral resolution that can provide intelligibility.

In the current experiments, a technique for measuring the frequency resolution employed by listeners when processing everyday sentences is employed. This technique allows the measurement of frequency resolution within a restricted range of frequencies and the establishment of the speech critical band (*S*-CB).

This measurement is based upon several recent and related findings. The first involves the mechanism allowing for the high intelligibility of narrow-band sentences: When spectral information was removed from narrow-band filtered speech, while maintaining temporal information, intelligibility fell from values near 100% to values near 0%. However, when a minimal spectral contrast was reintroduced by partitioning the band into a pair of juxtaposed temporal patterns, some intelligibility returned (Healy and Warren, 2003). Thus, the near-perfect intelligibility observed when sentences are filtered to a narrow "spectral slit" (Warren *et al.*, 1995) is attributable to contrasting temporal patterns of amplitude fluctuation within the narrow band.

The second finding involves the role of the skirts in the intelligibility of filtered speech: When the CID sentences (Davis and Silverman, 1978) were filtered using a fixed pass-band of 1/3 octave centered at 1500 Hz, but different slopes, mean intelligibility scores fell from values near 100% for slopes of approximately 100 dB/octave, to below 20% for bands created using high-order FIR filters that produced slopes over 1000 dB/octave (Healy, 1998; Warren and Bashford, 1999; Warren *et al.*, 2004). Thus, much of the contrasting pattern information providing for narrow-band sentence intelligibility can reside within the filter skirts, and a detailed examination of spectro-temporal speech information requires extremely precise filtering.

The third finding involves the influence of spectral overlap of contrasting temporal patterns on intelligibility: When a pair of juxtaposed narrow patterns (AM carrier bands) had shallower filter slopes so that they overlapped spectrally, intelligibility was found to be reduced relative to conditions in which steep filtering was employed to eliminate acoustic overlap of the adjacent patterns (Healy and Warren, 2003).

In the current study, a narrow band of speech having extremely steep filter slopes was sharply partitioned into increasing numbers of component bands, and the temporal envelope of each was used to modulate a corresponding carrier band. The AM carriers were summed and presented to listeners. Spectral detail within the band was manipulated by changing the number of component bands, while holding the overall bandwidth constant. Performance was expected to increase as the number of component bands increased and eventually reach asymptote. The amount of spectral detail in the signal at the point of asymptote provides an estimate of frequency resolution employed when processing the speech band, as additional spectral detail is ineffective.

Physiological limitations of the auditory system apply to the processing of all signals. However, the frequency resolution employed during the processing of speech may potentially be governed by a number of factors including those attributable to the auditory system and, in the case of speech,

those of the signal. According to the band importance functions of the articulation and speech intelligibility indexes (ANSI, 1986; 1997), speech possesses maximum density of information in the region surrounding approximately 1500 Hz. This spectral density is reflected in higher relative contributions to intelligibility of fixed width (e.g., 1/3 octave) bands. One possible outcome of the current study is that the speech signal will lack sufficient spectral density of information to take full advantage of the resolving power of the auditory system, especially in the low or high regions of the speech spectrum, and resolution will instead be governed by attributes of the signal. Alternatively, resolution may be quite fine and match or even exceed that predicted by psychophysical tuning.

The goal of the current study was to assess the component bandwidth at asymptote, a measure of the *S*-CB, in each frequency region of the speech spectrum. Frequency resolution was measured in the middle speech frequencies in Experiment 1, and in the high and low frequency regions in Experiments 2 and 3. In Experiment 4, the influence of different carrier types was assessed.

II. EXPERIMENT 1: MEASURING RESOLUTION IN THE MIDDLE FREQUENCIES

A. Experiment 1a

A preliminary experiment was performed to determine the relationship between bandwidth and intelligibility for the sentence materials employed. The restriction of speech to a narrow band was required to reveal the maximum resolution of contrasting temporal patterns within the band. This information was then used to guide the selection of overall bandwidths used for measuring resolution in Experiment 1b.

1. Method

a. Subjects. A group of 12 young adult listeners participated and received either course credit or money in compensation. All were native speakers of English between the ages of 18 and 40 years (mean age of all 124 subjects tested=22 years) and had pure-tone audiometric thresholds of 20 dB HL or better at octave frequencies from 250 to 8000 Hz (ANSI, 1996). Care was taken to ensure that none of the listeners had any prior exposure to the sentence materials. These characteristics and compensation procedures were the same for all listeners tested in this study.

b. Stimuli. The stimuli were based upon the standard recordings (20161-Hz sampling, 16-bit resolution) of the Hearing In Noise Test (HINT, Nilsson *et al.*, 1994). The sentences were filtered to a single narrow band having a width of 2/3, 5/6, or 1 octave centered at 1500 Hz. Filtering was performed using a single pass through a 2000-order digital FIR filter implemented in MATLAB. These parameters produced extremely steep filter slopes, which measured well over 1000 dB/octave. A secondary filtering pass having a slightly wider bandwidth corresponding to the points where the near-vertical slopes intersected the noise floor served to further attenuate the noise floor and resulted in increased signal-to-noise ratio (S/N). The level of each sentence in the 1-octave band was scaled to play back at a slow rms peak level of 70 dBA, and the narrower bandwidths were created by filtering this equated band. The processed digital signals

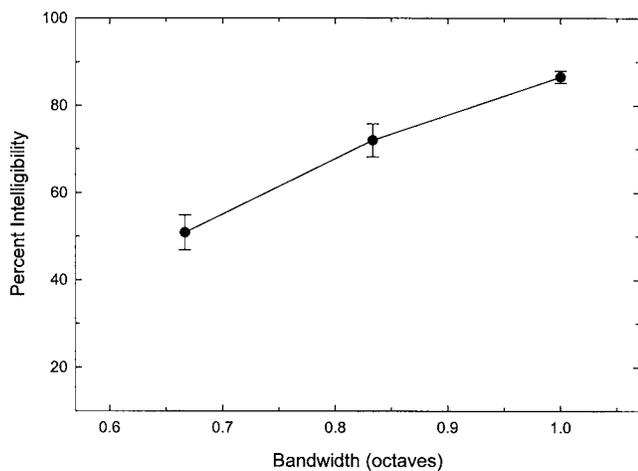


FIG. 1. Group mean intelligibility scores and standard errors for speech bands centered at 1500 Hz and having the bandwidths indicated.

were converted to analog form, amplified (Crown D75), and delivered diotically through TDH-49P headphones mounted in MX/51 cushions.

c. Procedure. Subjects were tested individually, seated with the experimenter in a single-walled audiometric booth located within an acoustically-treated room. Each listener heard 50 test sentences at each of the three bandwidths. The order in which conditions were heard was balanced across subjects so that each appeared in each serial position twice. In addition, to control for potential differences in the difficulty of the 50-sentence sets, the sentence list-to-condition correspondence was balanced so that each list was heard an equal number of times in each condition. Prior to the first condition, subjects heard a single list of 10 HINT practice sentences first broadband, then filtered in a manner corresponding to the first-heard condition. The practice list was repeated in corresponding filtered form before each of the two subsequent test conditions. Subjects were instructed to repeat each sentence aloud after hearing it. They heard each sentence only once, received no feedback, and were encouraged to guess if unsure of the content. The experimenter controlled the presentation of sentences and scored the proportion of component words reported correctly.

2. Results

Figure 1 shows the group mean intelligibility scores and standard errors for the three speech bandwidths presented in this experiment. Intelligibility increased with increasing bandwidth, reaching 87% at 1 octave. It is important in the subsequent experiment to employ a bandwidth at or near the value sufficient for intelligibility, but below ceiling values. Therefore, the data of Experiment 1a provide a guide for the selection of overall bandwidths for further examination in Experiment 1b.

B. Experiment 1b

In this experiment, spectral information within the narrow speech band was quantized by partitioning the band, and removing spectral information from each partition by replacing it with a carrier band that was amplitude modulated by the envelope of the corresponding speech partition.

1. Method

a. Subjects. A total of 30 listeners participated using selection and compensation procedures employed previously.

b. Stimuli. Because it provides appreciable intelligibility, but yields scores below ceiling values, the 1-octave speech band was first selected for the measurement of frequency resolution. In addition to that band, four additional conditions were prepared by partitioning the band into 2, 4, 6, and 10 equal log-width subbands. These subbands were contiguous, meeting at the -6 dB cutoffs. The lowest and highest partitions were created using a low pass or high pass, and the inner bands were created using a bandpass. The FIR filter order for this processing was increased to 6000, to further enhance the acoustic isolation of the juxtaposed bands. Because the FIR filter is linear in phase, all component bands were exactly aligned in time. The partitioning of the speech band is shown in Fig. 2.

Low-noise noise (LNN), which is noise engineered to have extremely small fluctuations in amplitude (Pumplin, 1985; Hartmann and Pumplin, 1988), was selected for the carrier signal. Kohlrausch *et al.* (1997) have described convenient methods for generating LNN, one of which ("Method 1") was used in the current study. This method involves the division of the waveform by its envelope in a series of iterations (100 in this case). Low-noise noise carriers were selected over Gaussian noise carriers because the random amplitude fluctuations of the narrow-band noises could potentially dilute the temporal details of the speech. They were selected over tonal carriers to allow spectral density to remain constant as the number of partition bands changed. Carrier bands having the same frequency composition as the speech partitions were created by summing sinusoidal components having appropriate amplitude and phase, and 0.5-Hz spacing. This component spacing produced a repeated noise having a duration that was sufficiently long (2 s) to not substantially interfere with the perception of the sentences. The LNN carrier bands were separated by 0.5 Hz so that when combined, the entire array would have equal spacing of components.

The amplitude envelope was extracted from each speech partition by full-wave rectification and low-pass filtering (2000-order FIR, 100-Hz cutoff) and applied to a corresponding LNN carrier band by multiplying on a sample point-by-point basis. The AM LNN carriers were then post-filtered to strictly restrict them to the frequency region of origin using the same filters employed to create the speech partitions. Because different filter orders were employed for the inner (6000 order) and outer (2000 order) cutoffs of the lowest- and highest-frequency bands, corrections for the different group delays were performed to ensure exact temporal alignment of the modulated carriers. These manipulations were all implemented in MATLAB.

The AM carriers comprising each condition (1, 2, 4, 6, and 10 bands) were assembled for presentation to listeners. Because this processing preserved the relative overall level of each component band, this resulting array maintained the spectral profile of the original speech band (see Fig. 2 lowest two panels). Each sentence in each condition was presented at a slow rms peak level of 70 dBA using the apparatus employed in the previous experiment. In addition to these conditions based on the 1-octave speech band, additional confirmatory conditions were prepared using bands that were narrower ($2/3$ octave) and wider ($3/2$ octave). These overall bandwidths were chosen so that when parti-

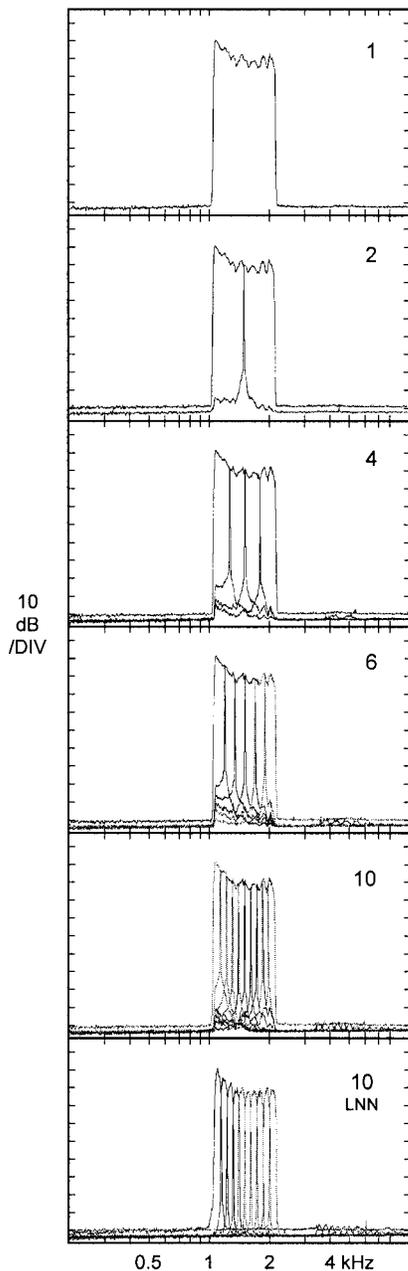


FIG. 2. Overlaid average amplitude spectra for the 1-octave speech band partitioned into 1, 2, 4, 6, and 10 subbands. The bottom panel shows the average amplitude spectra for a corresponding array of 10 amplitude-modulated low-noise noise carrier bands.

tioned into the same numbers of component bands, correspondences in bandwidth would occur. For example, a 1/6-octave component bandwidth is obtained by both dividing the 1-octave band into six and the 2/3-octave band into four. These bands were also centered at 1500 Hz and were presented as arrays of 1, 2, 4, 6, and 10 AM carriers using the same procedures employed to create the 1-octave stimuli.

c. Procedure. Separate groups of 10 listeners each were employed for the three overall bandwidths. They heard 30 test sentences in each of the five spectral resolution conditions. A practice list of 10 sentences was presented in broadband form at the beginning of the session, and was presented first as the overall speech band, then again in a form matching the particular experimental condition prior to each condition. The sentence list-to-condition correspon-

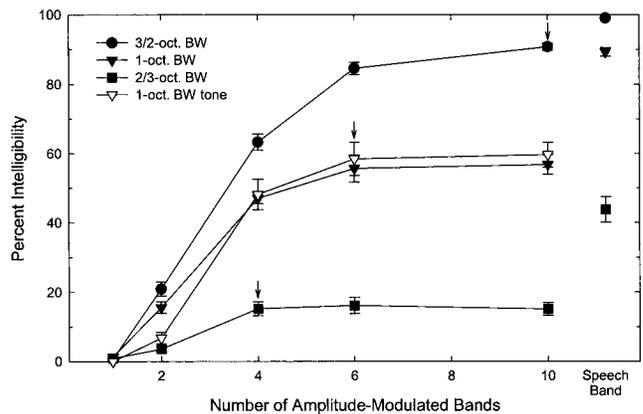


FIG. 3. Group mean intelligibility scores and standard errors for arrays of amplitude-modulated low-noise noise (LNN) carriers having increasing frequency resolution. Shown are scores for overall bandwidths of 3/2, 1, and 2/3 octave all centered at 1500 Hz. Asymptotic performance (indicated by arrows) occurred at 10, 6, and 4 bands respectively, which corresponds to a partition bandwidth of approximately 1/6 octave in each case. At the far right are scores for the corresponding intact speech bands. Also shown (as open symbols) are scores for additional conditions in which pure-tone carriers replaced the LNN carriers.

dence was balanced, and the conditions were presented in ascending order (1, 2, 4, 6, and 10 carriers), so that listeners were familiarized with the general procedure during the lower intelligibility conditions (1 and 2 carriers). Following these five AM carrier conditions, listeners heard 50 additional sentences in the overall speech band condition. This allowed performance across the two types of stimuli (AM carrier array versus speech band) to be compared within the same group of listeners. As before, each sentence was played only once and the experimenter scored the proportion of component words correctly recalled.

2. Results

The group mean intelligibilities for the 1-octave overall bandwidth conditions are presented in Fig. 3 as closed inverted triangles. Asymptote was defined throughout by scores differing by less than 2%. As can be seen, performance increased with increasing number of partitions/carriers, with no further increases after six. Also presented is the performance of these listeners on the 1-octave speech band (far right). It was found that performance at asymptote for the AM carriers was below that of the corresponding speech band.

The scores for the confirmatory conditions are also presented in Fig. 3. Performance was found to increase and reach asymptote at four partitions in the 2/3-octave condition (squares). In the 3/2-octave condition (circles), performance continued to increase to the largest number of partitions employed (10), reaching an apparent asymptote at a high level of intelligibility. In each case, the partition bandwidth at asymptote, a measure of the *S-CB*, matches or approximates 1/6 octave.

To further confirm the point of asymptote, the 1-octave conditions were recreated using pure-tone carriers in place of LNN carriers. The tones had a frequency corresponding to the (log) center of each partition. Five additional listeners were recruited and experienced procedures identical to those

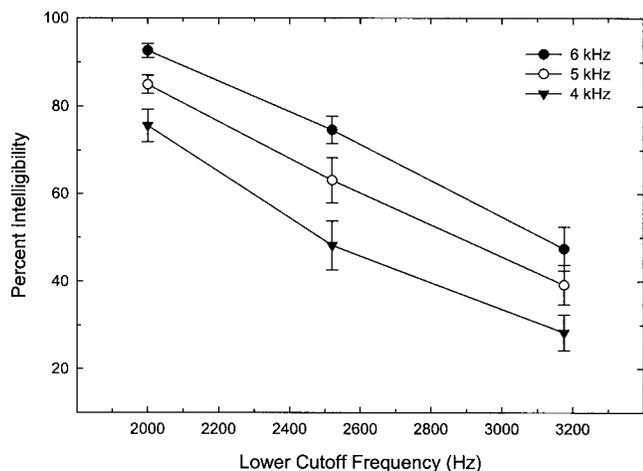


FIG. 4. Group mean intelligibility scores and standard errors for speech bands having lower cutoff frequencies indicated on the abscissa and upper cutoff frequencies as the parameter.

employed earlier, except that the intact speech band was not presented after the modulated carrier arrays. As with the LNN carriers, performance increased and reached asymptote at six bands (see Fig. 3, open symbols). This correspondence between tonal and LNN carriers was observed despite large differences in the spectral density of the tonal carrier array as the number of carriers increased, versus the relative constancy of the LNN arrays.

III. EXPERIMENT 2: MEASURING RESOLUTION AT HIGH FREQUENCIES

A. Experiment 2a

As in Experiment 1, results from a preliminary experiment in which spectral region was related to intelligibility were used to select bands for resolution measurement.

1. Method

A group of nine listeners served. The HINT sentences were filtered using the same primary- and secondary-filtering techniques employed in Experiment 1a. A set of three lower cutoff frequencies (2000, 2520, and 3175 Hz) and three upper cutoff frequencies (4000, 5000, and 6000 Hz) yielded nine speech-band conditions. The sentence list-to-condition correspondence was balanced, and the conditions were presented in a different random order for each listener. Listeners heard 20 sentences in each condition, and each condition was preceded by practice as in Experiment 1a. Each sentence was set to 70 dBA and delivered diotically over Sennheiser HD 250II headphones, which were selected to provide a frequency response that was wider than the audiometric headphones employed in Experiment 1. As before, the experimenter was seated with the subject within an audiometric booth, controlled the presentation of sentences, and recorded the responses.

2. Results

Group mean intelligibility scores are shown in Fig. 4. These results show how intelligibility varies as a function of

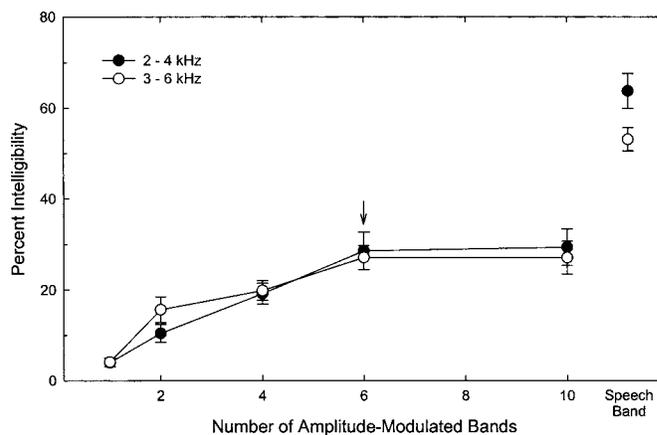


FIG. 5. Group mean intelligibility scores and standard errors for arrays of amplitude-modulated LNN carriers having increasing frequency resolution. Shown are scores for a pair of bands in the high speech frequency region. Asymptotic performance (indicated by the arrow) occurred at 6 bands for both 1-octave bandwidth conditions, corresponding to a partition bandwidth of 1/6 octave. At the far right are scores for the corresponding intact speech bands.

lower cutoff, and indicate that information up to 6 kHz contributes to intelligibility for these materials under these processing conditions.

B. Experiment 2b

1. Method

Twenty listeners were randomly divided into two groups of ten each. The standard 1-octave bandwidth selected in Experiment 1b was again employed, but the band was transposed upward in frequency to 2–4 kHz in one condition and to 3–6 kHz in a second. The preparation of amplitude-modulated LNN stimuli was the same as that employed in Experiment 1b, and again yielded bands represented by 1, 2, 4, 6, and 10 carriers. The testing procedures were also the same as those of Experiment 1b, with the exception that subjects heard 30, rather than 50 sentences in the overall speech band condition at the end of the session, and that the Sennheiser headphones were employed.

2. Results

The group mean intelligibility scores are shown in Fig. 5. In accord with the results from the mid-frequency region, performance increased and reached asymptote at six bands for both of the 1-octave conditions, corresponding to a component bandwidth of 1/6 octave in both cases.

IV. EXPERIMENT 3: MEASURING RESOLUTION AT LOW FREQUENCIES

A. Experiment 3a

1. Method

The sentences were again filtered using the procedures of Experiment 1a. Four upper cutoff frequencies (476, 566, 673, and 800 Hz) were combined with two lower cutoffs (100 and 200 Hz) to produce eight speech-band conditions.

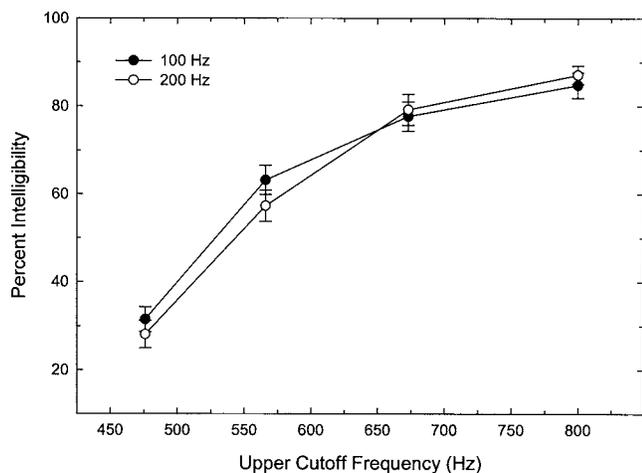


FIG. 6. Group mean intelligibility scores and standard errors for speech bands having the upper cutoff frequencies indicated on the abscissa and the lower cutoff frequencies as the parameter.

Eight listeners heard 30 sentences in each condition, and all procedures were otherwise identical to those of Experiment 2a.

2. Results

Group mean intelligibility scores are shown in Fig. 6. Little difference was observed between the 100 and 200 Hz lower cutoffs, indicating that spectral information below 200 Hz contributes little to intelligibility of these materials under these conditions.

B. Experiment 3b

The partitioning of the speech bands in Experiments 1 and 2 was performed in logarithmic units, because psycho-physical tuning, as measured by the equivalent rectangular bandwidth (ERB_N , Glasberg and Moore, 1990; Moore, 2003) follows a simple logarithmic function in the mid- and high-frequency regions, and is approximately constant in width at 1/6 octave. However, in the current experiment, partitioning was performed directly in ERB_N units, because the two functions diverge sharply below approximately 1 kHz. The ERB_N increases to approximately 1/2 octave at 100 Hz, and 1/6-octave bands from 100 to 1000 Hz range from 0.3 to 0.9 ERB_N .

1. Method

A total of 30 listeners were randomly divided into three groups of 10 each. Three overall bandwidths were employed. A four- ERB_N condition spanned the region from 312 to 603 Hz (ERB_N numbers 8–12), a six- ERB_N condition spanned the region from 257 to 698 Hz (ERB_N numbers 7–13), and a 10- ERB_N condition spanned the region from 163 to 921 Hz (ERB_N numbers 5–15). The four- and six- ERB_N bandwidths were divided as before (1, 2, 4, 6, and 10 bands), and the 10- ERB_N bandwidth was divided into 2, 5, 10, 15, and 20 bands. This partitioning was performed in equal ERB_N -width units. Other aspects of the creation of LNN carrier arrays, as well as the apparatus and procedures, were the same as those of Experiment 2b.

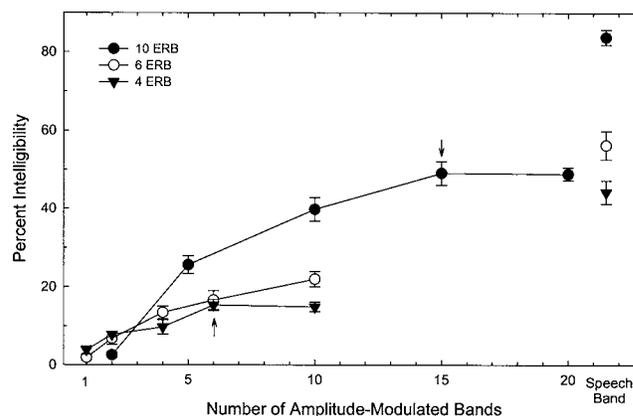


FIG. 7. Group mean intelligibility scores and standard errors for arrays of amplitude-modulated LNN carriers having increasing frequency resolution. Shown are scores for three bands in the low speech frequency region. Asymptotic performance (indicated by arrows) occurred at 6 bands for the 4 ERB_N overall bandwidth condition and at 15 bands for the 10 ERB_N condition, corresponding to a partition bandwidth of 0.67 ERB_N in each case.

2. Results

Group mean intelligibilities are displayed in Fig. 7. The data for the six- ERB_N condition indicates that performance continues to increase beyond six partitions, which suggests that spectral resolution in this region exceeds 1 ERB_N . The data for the four- ERB_N condition also indicate resolution exceeding 1 ERB_N . The asymptote at six bands corresponds to a component bandwidth of 0.67 ERB_N . Finally, the data for the 10- ERB_N condition show an asymptote at 15 bands, which also corresponds to a component bandwidth of 0.67 ERB_N .

V. EXPERIMENT 4: THE INFLUENCE OF DIFFERENT CARRIER TYPES

In the current experiment, differences in performance between the arrays of AM carriers at asymptote and the intact speech bands were examined using different carrier types within a single group of subjects.

A. Method

Ten listeners participated. Performance in one condition (the standard 1-octave overall bandwidth represented by six carriers from Experiment 1b) was compared across four carrier types: pure tone, frequency-modulated (FM) tone, LNN, and Gaussian noise. The pure tones again had frequencies corresponding to the center of each partition. The FM tones had a nominal frequency matching that of the tonal carriers, and had sinusoidal frequency modulation at an average rate of 28 Hz and an average range of 64 Hz. The rate and range values for each FM carrier matched the mean of the dominant frequency modulations of the corresponding LNN carrier (LNN is essentially a flat amplitude, random FM signal; see Kohlrausch *et al.*, 1997). The Gaussian noise band carriers were created using the same filtering used to partition the speech band. These four conditions, plus the 1-octave speech band, were heard in a different random order for each subject. Thirty sentences were heard in each condition, and all

other stimulus preparation methods, apparatus, balancing, and testing procedures were the same as those employed in Experiment 1b.

B. Results

The group mean intelligibility scores (and standard errors) for the five conditions are as follows: speech band: 87.8% (1.7), pure tone: 60.8% (2.8), FM tone: 55.9% (3.1), LNN: 46.5% (2.4), and Gaussian noise: 36.3% (2.9). Thus, none of the carrier arrays reached the intelligibility of the 1-octave speech band.

Performance was poorest with the Gaussian noise carriers, where random amplitude fluctuations of the narrow-band carrier may potentially dilute the temporal details of the speech. Performance was best in the tonal carrier condition, despite the impoverished amplitude spectrum of the 6-tone array and the tonal timbre of this signal. Scores were only slightly reduced when sinusoidal FM was applied to the tonal carriers. Finally, performance in the LNN-carrier condition was below that of the tones, but above that of the Gaussian noise.

VI. DISCUSSION

Measures of auditory frequency resolution have typically employed simple stimuli as well as procedures not compatible with spectro-temporally complex speech signals. Using techniques to quantize spectral information within the speech signal, the current experiments were designed to measure frequency resolution of everyday speech materials in each spectral region separately. This region-specific information is not available when spectral smearing or reduction is applied to all frequency regions simultaneously, or from experiments examining the relation between auditory tuning and broadband speech.

The *S*-CB has been defined here as the width of spectral bands comprising the acoustic speech signal at performance asymptote. This value gives the spectral resolution of the signal beyond which information content does not increase, and reflects the ability to extract spectral detail from the acoustic speech signal. It was important that narrow bandwidths were employed in the current study: If the bandwidths far exceeded required for full intelligibility, then the opportunity to observe the processing of contrasting patterns within the band at maximum resolution would have been lost. For example, if Experiment 1b were repeated with an overall speech bandwidth of two or three octaves, scores would reach ceiling before reaching asymptote. These wider bandwidth conditions would eventually approach those of Shannon *et al.* (1995). Similarly, in the low- and high-frequency regions, it was important to only examine regions of the spectrum that contributed to intelligibility.

When spectral information within the speech band was quantized, scores increased with increasing spectral resolution and eventually reached asymptote, presumably when the spectral resolution of the stimulus matched that of the speech processing system. In the middle speech frequencies (Experiment 1), the component bandwidth at asymptote was found to be 1/6 octave across four conditions employing four dif-

ferent groups of subjects. This correspondence occurred despite differences in overall bandwidth, differences in the number of AM carriers in the array, differences in the level of performance at asymptote, and differences in carrier signal type. It therefore appears that the measurement of the *S*-CB is somewhat robust. In Experiment 2, the *S*-CB was also found to be 1/6 octave in width for a pair of conditions covering the higher regions of the speech spectrum. Because 1/6-octave bands in the region from 1000 to 6000 Hz range from 0.9 to 1.0 ERB_N , it is concluded that, in this frequency region, the size of the *S*-CB approximately matches that of the psychophysical CB. In Experiment 3, resolution was examined in the lower speech frequency region. In contrast to Experiments 1 and 2, it was found that the *S*-CB was somewhat narrower than the psychophysical CB. Again, despite differences in bandwidth, mean intelligibility level, and number of component bands at asymptote, the conditions converged to yield a measurement of the *S*-CB at 0.67 ERB_N . It is unclear why the two measures of frequency resolution diverge somewhat at low frequencies. However, the *S*-CB is within approximately 25 Hz of the CB at a frequency of 500 Hz and so it may be concluded that, overall, current estimates of psychophysical tuning also describe reasonably well the ability of normal-hearing listeners to extract spectral detail from the spectro-temporally complex speech signal.

Because high-context sentences were employed, the current results are restricted to these materials and it is possible that different results will be obtained for individual words, syllables, or even phonemes. However, the goal of the current study was to assess resolution of everyday speech, and the results indicate that the functional density of information in this material is quite high. It is interesting to note that the observed match between speech resolution and psychophysical tuning in the mid and high frequencies, but the modest divergence in the low frequencies would not be predicted from the shape of the band-importance functions of the Articulation or Speech Intelligibility Indexes.

The intelligibility of the array of AM carriers at asymptote does not reach that of the intact speech band. This difference in performance was observed both across and within subjects. It appears that a speech band having spectral information quantized in the fashion employed here does not completely retain characteristics of the signal required for full intelligibility. This is especially evident from conditions in which spectral resolution far exceeds that required for asymptotic performance (e.g., 2/3-octave band represented by 10 carriers in Fig. 3).

A control condition was employed to ensure that this difference in performance between the AM carriers and the intact speech band was not due to some disruption caused by the severe filtering used to partition the speech bands. Two of the subjects from Experiment 1b participated in additional testing at the end of the session. They heard six lists of ten HINT sentences each, alternating between the 1-octave speech band and the 1-octave speech band divided into six partitions and recombined. The average intelligibility score for the two conditions differed by less than 1% (89.9% intact

speech band, 90.5% recombined speech partitions), suggesting that artifacts associated with the extreme filtering did not hinder performance.

A second control condition was designed to investigate possible interference produced by beating of the juxtaposed carrier bands. This test was based on modulation detection interference (Yost and Sheft, 1989). If beating of the carrier array was interfering with the temporal details of the speech, then the introduction of an array of carriers to a separate frequency region may have some perceptible influence on a speech band. An array of six unmodulated LNN bands mimicking those employed as carriers in the 1-octave overall bandwidth condition of Experiment 1 was transposed up in frequency by 2000 Hz and presented along with the 1-octave speech band at an equal average spectrum level. An additional condition employed an array of tones having frequencies corresponding to the centers of the transposed LNNs. Both informal listening and formal intelligibility testing indicated no interference.

Previous work has shown that filter skirts as steep as 100–200 dB/octave can contribute considerably to narrow-band speech intelligibility. However, it is difficult to know precisely how far down the filter skirts the fluctuating information is available, and it is difficult to know to what extent the severe spectral tilt of information lying along the slope can affect recognition. In addition, it has been found that intelligibility of narrow juxtaposed temporal speech patterns can suffer when they possess filter skirts that cause them to overlap spectrally. The use of extremely steep filter slopes allows the elimination of these complicating effects.

The current results have implications not only for the processing of speech by normal-hearing listeners, but also for speech perception by individuals having a hearing impairment (HI). Sensorineural HI is often characterized by broadened auditory tuning (for a review, see Moore, 1998). Although it may be assumed that tuning that exceeds the normal psychophysical CB will negatively impact speech perception, the empirical evidence has been mixed (e.g., Celmer and Bienvenue, 1987; Moore and Glasberg, 1987; Thibodeau and van Tasell, 1987; Dubno and Dirks, 1989; Turner and Henn, 1989; Dubno and Schaefer, 1992; ter Keurs *et al.*, 1993; Ching *et al.*, 1997). Because the ability to extract spectral information was found to be quite fine in the current investigation, and to match the psychophysical tuning of normal-hearing listeners in the high frequency region where most hearing loss occurs, these data indicate that any broadening beyond normal may be expected to hinder the ability to extract normally usable information from the speech signal.

The frequency resolution of speech is also a question of importance for the prosthetic treatment of hearing loss. Modern cochlear implants employ an array of electrodes that stimulate different portions of the cochlea with temporal information derived from corresponding regions of the speech spectrum. Although a current goal of cochlear implant development is enhance the poor frequency representation of these devices and provide as much spectral specificity as possible,

mimicking the frequency resolution employed normally when processing speech can be considered an ultimate goal of cochlear prosthetics.

VII. SUMMARY AND CONCLUSIONS

When spectral information was quantized within speech bands at or near the minimum bandwidth required for intelligibility, performance was found to asymptote consistently across a variety of conditions at a partition bandwidth matching or approximating 1/6 octave for frequency regions from approximately 1000 to 6000 Hz. Thus, in these regions, the size of the *S*-CB approximately matches that of the psychophysical CB. However, in the region below approximately 1000 Hz, the ability to process spectral contrasts within the running speech signal is somewhat finer than would be suggested by psychophysical tuning.

ACKNOWLEDGMENTS

This research was supported by NIDCD Grant Nos. DC01376 and DC05795. The authors thank Dave Eddins for providing the code used to create the LNN and Stuart Rosen for helpful comments on an earlier version of this manuscript.

- American National Standards Institute. (1996). *ANSI-S3.6 (1996), Specifications for Audiometers*. (New York).
- American National Standards Institute. (1986). *ANSI-S3.5, 1969 (R 1986), American National Standard Methods for the Calculation of the Articulation Index* (New York).
- American National Standards Institute. (1997). *ANSI-S3.5. (1997), American National Standard Methods for the Calculation of the Speech Intelligibility Index*. (New York).
- Baer, T., and Moore, B. C. J. (1993). "Effect of spectral smearing on the intelligibility of sentences in noise," *J. Acoust. Soc. Am.* **94**, 1229–1241.
- Boothroyd, A., Mulhearn, B., Gong, J., and Ostroff, J. (1997). "Simulation of sensorineural hearing loss: Reducing frequency resolution by uniform spectral smearing," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Mahwah, NJ).
- Celmer, R. D., and Bienvenue, G. R. (1987). "Critical bands in the perception of speech signals by normal and sensorineural hearing loss listeners," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Nijhoff, Dordrecht).
- Ching, T., Dillon, H., and Byrne, D. (1997). "Prediction of speech recognition from audibility and psychoacoustic abilities of hearing-impaired listeners," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Mahwah, NJ).
- Davis, H., and Silverman, S. R. (1978). *Hearing and Deafness*, 4th ed. (Holt, Rinehart, and Winston, New York).
- Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (1998). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels," *J. Acoust. Soc. Am.* **104**, 3583–3585.
- Dubno, J. R., and Dirks, D. D. (1989). "Auditory filter characteristics and consonant recognition for hearing-impaired listeners," *J. Acoust. Soc. Am.* **85**, 1666–1675.
- Dubno, J. R., and Schaefer, A. B. (1992). "Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners," *J. Acoust. Soc. Am.* **91**, 2110–2121.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- Hartmann, W. M., and Pumplin, J. (1988). "Noise power fluctuations and the masking of sine signals," *J. Acoust. Soc. Am.* **83**, 2277–2289.
- Healy, E. W. (1998). "A minimum spectral contrast rule for speech recognition: Intelligibility based upon contrasting pairs of narrow-band amplitude patterns," Doctoral dissertation, University of Wisconsin-Milwaukee.
- 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.
- Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Püschel, D. (1997). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," *Acta Acoustica united with Acoustica* **83**, 659–669.
- Moore, B. C. J. (1998). *Cochlear Hearing Loss* (Whurr, London).
- Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, London).
- Moore, B. C. J., and Glasberg, B. R. (1987). "Relationship between psychophysical abilities and speech perception for subjects with unilateral and bilateral cochlear hearing impairments," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Nijhoff, Dordrecht).
- Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Am.* **95**, 1085–1099.
- Patterson, R. D. (1976). "Auditory filter shapes derived with noise stimuli," *J. Acoust. Soc. Am.* **59**, 640–654.
- Patterson, R. D., and Moore, B. C. J. (1986). "Auditory filter shapes and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London).
- Pumplin, J. (1985). "Low-noise noise," *J. Acoust. Soc. Am.* **78**, 100–104.
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- ter Keurs, M., Festen, J. M., and Plomp, R. (1992). "Effect of spectral envelope smearing on speech reception I.," *J. Acoust. Soc. Am.* **91**, 2872–2880.
- ter Keurs, M., Festen, J. M., and Plomp, R. (1993). "Effect of spectral envelope smearing on speech reception. II," *J. Acoust. Soc. Am.* **93**, 1547–1552.
- Thibodeau, L. M., and Van Tasell, D. J. (1987). "Tone detection and synthetic speech discrimination in band-reject noise by hearing-impaired listeners," *J. Acoust. Soc. Am.* **82**, 864–873.
- Turner, C. W., and Henn, C. C. (1989). "The relation between vowel recognition and measures of frequency resolution," *J. Speech Hear. Res.* **32**, 49–58.
- Warren, R. M., and Bashford, Jr., J. A. (1999). "Intelligibility of 1/3-octave speech: Greater contribution of frequencies outside than inside the nominal passband," *J. Acoust. Soc. Am.* **106**, L47–L52.
- Warren, R. M., Bashford, Jr., J. A., 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., Riener, K. R., Bashford, Jr., J. A., and Brubaker, B. S. (1995). "Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits," *Percept. Psychophys.* **57**, 175–182.
- Yost, W. A., and Sheft, S. (1989). "Across-critical-band processing of amplitude-modulated tones," *J. Acoust. Soc. Am.* **85**, 848–857.