Published in final edited form as:

J Acoust Soc Am. 2000 September ; 108(3 Pt 1): 1264–1268.

Intelligibility of bandpass speech: Effects of truncation or removal of transition bands

Richard M. Warren^{a)}, James A. Bashford Jr.^{b)}, and Peter W. Lenz^{c)}

Department of Psychology, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201-0413

Abstract

An intelligibility of over 90% was reported for keywords in “everyday” 1/3-octave sentences centered on 1500 Hz and having steep transition band slopes of 100 dB/octave [Warren *et al.*, *Percept. Psychophys.* **57**, 175–182 (1995)]. A subsequent study by Warren and Bashford [*J. Acoust. Soc. Am.* **106**, L47–L52 (1999)] found that it was not the 1/3-octave passband, but the transition bands that were chiefly responsible for this high intelligibility: When the passband and transition bands were segregated using filter slopes of 1000 dB/octave, the isolated passband had an intelligibility score of only 24%, while the pair of transition bands had a score of over 80%. In the present study, experiment 1 examined the distribution of information along the transition bands' slopes by truncation at graded downpoints: Truncation at downpoints of 40 dB or more produced no significant change in intelligibility. Experiment 2 closed the gap separating the transition bands so that their slopes intersected at 1500 Hz. This triangular band had a negligible passband (as defined conventionally by 3-dB downpoints) and an intelligibility score of 60%; truncation at downpoints of 50 dB or more produced no significant change in intelligibility. Experiment 3 determined the intelligibilities of rectangular bands (1000-dB/octave slopes) centered on 1500 Hz. Their bandwidths ranged from 3 to 12 semitones in 1-semitone steps, resulting in intelligibility scores increasing monotonically from 14% to 94%. Calculations based upon experiments 2 and 3 showed that the triangular band truncated at 30-dB downpoints had half the intelligibility of a rectangular band having the same frequency range.

Introduction

As vocal productions become more complex and approximate normal discourse, identification becomes more accurate and resistant to distortions, masking, and deletions. At the level of sentences, a narrow spectral band can be sufficient to achieve high intelligibility. Warren *et al.* (1995) investigated the extent to which keywords in narrow-band sentences could be identified when they lacked spectral components required for the accurate identification of isolated phonemes and words. The stimuli chosen for this study had been designed to represent “everyday American speech” by Silverman and Hirsh (1955). When listeners heard 1/3-octave bands centered on 1500 Hz with steep slopes of 100 dB/octave, it was found that intelligibility was well over 90%. Stickney and Assmann (1997) verified the high intelligibility of narrow-band sentences, and also demonstrated that intelligibility dropped when the keywords in the sentences used for scoring had a lower contextual probability. When the narrow-band speech consisted of isolated words, a further large drop in performance occurred.

^{a)}Electronic mail: rmwarren@csd.uwm.edu

^{b)}Electronic mail: bashford@csd.uwm.edu

^{c)}Electronic mail: plenz@csd.uwm.edu

In order to examine the basis for the high intelligibility of words in 1/3-octave sentences, Warren and Bashford (1999) used the same digitally recorded stimuli employed in the earlier Warren *et al.* study. It was found that the average score of over 90% correct identification for keywords was due not to the nominal 1/3-octave passband, but rather to the flanking 100-dB/octave transition bands. This was determined by separating the passband (as defined conventionally by the 3-dB downpoints) and the pair of transition bands using 2000-order finite impulse response (FIR) filtering that produced nearly vertical slopes of approximately 1000 dB/octave. The isolated rectangular 1/3-octave passband had an intelligibility of only 24%, while the isolated transition bands (two right triangles separated by a 1/3-octave gap) had a much higher intelligibility score of 83% despite their severe spectral tilt.

The contrast between the low intelligibility of the passband and the remarkably high intelligibility of the steep transition bands raised several new questions that were addressed by the present study, using the same digitally recorded 1/3-octave sentences as the earlier experiments.

I. Overall Plan and Rationale

Experiment 1 examined the distribution of functional information along the transition bands by truncating the slopes at graded downpoints using 2000-order FIR filtering. In order to avoid ceiling effects that would obscure the loss of information, the passband was excised, leaving the isolated transition bands with a below-ceiling intelligibility score of approximately 80% prior to truncation.

Since the pair of transition bands employed in experiment 1 were separated by a 1/3-octave gap, the speech spectrum being sampled was discontinuous. It has been reported that two widely separated narrow bands of the everyday sentences, when presented together, have intelligibility scores well above the scores for each of the bands heard separately (Warren *et al.*, 1995). To avoid such interactive enhancement associated with dual bands, experiment 2 closed the gap separating the transition bands, creating a single band consisting of slopes of 100 dB/octave intersecting at 1500 Hz. This triangular band had a negligible bandwidth, as defined conventionally by the 3-dB downpoints along the slopes. Preliminary observations had indicated that closing the 1/3-octave gap reduced intelligibility, but that the pair of intersecting slopes nevertheless had a considerably higher intelligibility than the 1/3-octave rectangular passband having the same center frequency. In experiment 2, the slopes of the triangular band were truncated at graded downpoints using the 2000-order FIR filtering as in experiment 1, and the intelligibilities were determined for various bandwidths (defined in this case as the range between the truncation downpoints) for this band consisting solely of 100-dB/octave slopes.

In experiment 3, the intelligibility was determined for ten bands having bandwidths ranging from 3 to 12 semitones in 1-semitone steps. These bands were each centered logarithmically on 1500 Hz, and were truncated at the passband borders with slopes of 1000 dB/octave. While these rectangular bands lack appreciable transition bands, the triangular bands lack appreciable passbands. By comparing the intelligibilities of truncated triangular bands (as determined in experiment 2) with the intelligibilities of rectangular bands extending over the same frequency range (as determined in experiment 3), we can determine the extent to which the listeners were able to compensate for the 100-dB/octave distortion of the normal spectral profile. In addition, the rectangular bands in experiment 3 provide a measure of the passband bandwidth needed to achieve an intelligibility comparable to the near ceiling value obtained with the intact 1/3-octave band (passband plus 100-dB/octave transition bands) reported previously for the same sentences.

II. Subjects

The 60 listeners (20 in each of the three experiments) were native English speakers with no known hearing problems, who were recruited from the introductory Psychology courses at the University of Wisconsin-Milwaukee. They received course credit for their participation.

III. Stimulus Processing CID (Central Institute for the Deaf)

The filtered sentences employed in this study consisted of the 100 (10 lists of 10) “everyday” sentences (Silverman and Hirsh, 1955). They contain 500 keywords that are used for scoring of intelligibility. Another 25 sentences served as practice stimuli and were taken from the high-predictability sublist of the revised SPIN (speech perception in noise) test (Kalikow *et al.*, 1977; Bilger *et al.*, 1984). All sentences had been recorded in digital form (44.1-kHz sampling, 16-bit quantization), and were used previously to study the intelligibility of 1/3-octave sentences (Warren *et al.*, 1995; Warren and Bashford, 1999). They were produced by a male speaker with no evident regional accent, who had an average voicing frequency of about 100 Hz.

Prior to preparation of the filtered stimuli, all sentences were transduced by a Sennheiser HD 250 Linear II headphone. The slow-rms peak level of each sentence was matched to within ± 0.2 dBA using an artificial ear in conjunction with a Brüel & Kjær Model 2230 digital sound-level meter.

The two types of filtering described by Warren and Bashford (1999) were used in this study. Slopes of approximately 1000 dB/octave were produced using MATLAB function *fir1* with order $n = 2000$, and when slopes of 100-dB/octave were required, they were produced by low-pass followed by high-pass filtering using MATLAB function *ellip* with order $n = 7$.

IV. Procedure

A. General procedure

A repeated measures design was used, with the 10 CID lists presented pseudorandomly in a balanced order. Before receiving each stimulus condition, listeners were familiarized with the effects of filtering by listening to a set of SPIN sentences (five for each of the five conditions in experiments 1 and 2, and two for each of the ten conditions in experiment 3). These practice sentences were first presented broadband and then filtered in the same manner as the test sentences that followed. Listeners were tested individually while seated with the experimenter in an audiometric chamber. The stimuli were delivered diotically through Sennheiser HD 250 Linear II headphones. Subjects were told to repeat each sentence as accurately as possible, and to guess if unsure. The number of keywords reported correctly for each condition was recorded by the experimenter. Testing occurred in a single session lasting approximately 30 min.

B. Experiment 1: Truncation of isolated transition bands

The levels of the transition bands were the same as those employed by Warren and Bashford (1999), and were based upon a slow-rms peak level of 75 dB for the intact 1/3-octave sentences. This resulted in an average peak level of 69 dB for the pair of transition bands as measured after excision of the 1/3-octave passband. Listeners received 20 of the 100 sentences in each of the five conditions. These consisted of the intact transition bands with slopes of 100 dB/octave plus four truncated derivatives having 1000-dB/octave cutoffs at downpoints along both slopes of -30 , -40 , -50 , and -60 dB, respectively (see Fig. 1). All truncations had a negligible effect on overall levels (less than 0.1 dB). The spectra of the stimuli are shown in Fig. 1.

C. Experiment 2: Truncation of triangular bands

The triangular band sentences were produced by setting the cutoff frequencies of both the low-pass and high-pass 100-dB/octave filters at 1500 Hz. As in experiment 1, in addition to the intact transition bands, four additional stimuli were employed having nearly vertical truncations of the slopes occurring at downpoints of -30, -40, -50, and -60 dB, respectively. The level of all five stimulus sets was 75 dB (as in experiment 1, truncation had no appreciable effect upon the overall level). The spectra of the stimuli are shown in Fig. 2.

D. Experiment 3: Graded widths of rectangular bands

Listeners heard ten sets of ten rectangular-band sentences having slopes of 1000 dB/octave. All 100 sentences were centered logarithmically on 1500 Hz, and their bandwidths varied in 1-semitone steps from 3 semitones (1/4 octave) to 12 semitones (one octave). The one-octave band had a level of 75 dBA. The narrower bands, which maintained the same spectrum level for corresponding frequencies, were consequently reduced in overall peak level (the lowest average peak level was 68 dB for the three-semitone band). The spectra of the stimuli are shown in Fig. 3.

V. Results and Discussion

The intelligibility scores obtained in experiments 1 and 2 are shown in Fig. 4, and scores for experiment 3 are shown in Fig. 5. These scores were subjected to separate analyses of variance (ANOVAs), which yielded significant main effects for truncation of the transition bands in experiment 1 [$F(4,76) = 14.94, p < 0.0001$], for truncation of the triangular band in experiment 2 [$F(4,76) = 14.90, p < 0.0001$], and for variation of the width of the rectangular bands in experiment 3 [$F(9,261) = 135.13, p < 0.0001$]. Subsequent Tukey HSD tests indicated that truncation of the transition bands in experiment 1 lowered intelligibility only when the frequencies beyond the 30-dB downpoints were removed ($p < 0.05$). In contrast, intelligibility of the triangular band in experiment 2 was reduced by truncation not only at minus 30 dB, but also when frequencies beyond the 40-dB downpoints were removed ($p < 0.05$ by Tukey HSD test). Finally, Helmholtz single-df contrast analysis of the results from experiment 3 indicated that intelligibility improved consistently with increasing width of the rectangular band ($F \geq 14.4, p < 0.001$ or better) until reaching asymptote in the ten-semitone bandwidth condition.

As shown in Fig. 4, there was a small monotonic decrease in average scores in experiment 1 as the truncation downpoint moved from 60 to 40 in 10-dB steps, but no significant change from the score for the intact transition bands occurred until the 30-dB cutoff was reached. If we accept the simplifying assumption that only frequencies between the two 30-dB downpoints on the transition bands (1043 and 2067 Hz, respectively) contributed to intelligibility, then frequencies spanning approximately one octave were responsible for the near ceiling intelligibility of the nominal 1/3-octave band (passband plus intact transition bands) reported by Warren and Bashford (1999).

As shown in Fig. 4, intelligibility scores decreased when the gap separating the two transition bands in experiment 1 was closed, resulting in the high-pass and low-pass slopes intersecting at 1500 Hz. This decrease in scores (from 84% to 60%) presumably occurred because of the greater redundancy of information in a restricted spectral range (Pollack, 1968; Steeneken and Houtgast, 1999). The effects of truncation obtained in experiment 2 were also somewhat different from those obtained in experiment 1: As shown in Fig. 4, the mean intelligibility score for the intact triangular band was unchanged by truncation at the 60-dB downpoints, but the scores decreased in a linear fashion as truncations moved up the

slopes. The reduction in intelligibility relative to the intact triangular band became significant at the 40-dB downpoints. Using the conservative assumption that only information limited to the frequencies lying between the two 40-dB downpoints contributed to intelligibility, then the effective bandwidth of the triangular band (in terms of intelligibility) extended from 1083 to 1947 Hz (10.2 semitones).

Figure 5 shows the monotonic increase in intelligibility in experiment 3 occurring when the bandwidths of rectangular bands centered upon 1500 Hz were increased from 3 semitones (intelligibility of 14%) to 12 semitones (intelligibility of 94%). Warren and Bashford (1999) reported that the 1/3-octave (four semitone) rectangular passband of the everyday sentences had an intelligibility of 24%, a value close to the 28% obtained for this bandwidth in the present study.¹ It also can be seen that intelligibility scores increase monotonically up to the 12-semitone limit, but scores from 10 semitones (88%) to 12 semitones (94%) increase only slightly.

In experiment 2, when the two transition bands were brought together and intersected at 1500 Hz, the intelligibility was 60%. Despite the absence of an appreciable passband, this triangular band had an intelligibility score corresponding to a rectangular bandwidth between six and seven semitones as measured in experiment 3. When the triangular band was truncated at 30 dB, the cutoff frequencies of 1178 and 1820 Hz spanned a range of 7.5 semitones, and the truncated band had an intelligibility score of 34%. A rectangular band at the same center frequency and spanning the same 7.5-semitone range had an intelligibility of about 70% as calculated from the data in experiment 3. Hence, the spectral tilt of 100-dB/octave reduced the scores of the triangular band by about 50% over this frequency range.

The ability of listeners to recognize words in spectrally limited and distorted sentences demonstrates the redundancy of acoustic information in speech. It seems probable, as suggested by Greenberg (1996) and by Shannon *et al.* (1995), that the patterns of temporal modulation occurring at local frequencies can allow listeners to comprehend degraded speech. Listeners may have the ability to utilize this temporal information despite the distortion of relative spectral amplitudes by (1) Using familiarity with the normal spectral profile of speech to compensate for filter slopes by applying an inverse transfer function of the type suggested by Watkins and Makin (1996); or (2) by applying a direct cross-frequency integration of information concerning local patterns of amplitude fluctuation without the need to restore the normal spectral profile (this latter possibility is currently being tested experimentally).

It appears that listeners having mastery of a language can achieve comprehension in a variety of ways. Although narrow-band sentences can be understood despite the absence and/or amplitude distortion of formants normally present in speech, Remez *et al.* (1994) have shown that sentences can be understood when the information consists solely of sinusoids that track the frequency and amplitude fluctuation of formants.

It is suggested that appropriate processing strategies can be selected and their results compared and combined as required by the specific task demands, integrating not only top-down and bottom-up, but also processing across similar hierarchical levels reached by different routes. This flexible nature of speech processing is emphasized by the LAME (lateral access from multilevel engrams) model elaborated by Warren (1999) that

¹In another study in our laboratory, Eric Healy used the same recorded sentences in his 1998 doctoral dissertation on the intelligibility of tones and noise modulated by narrow-band speech. Using a between-groups design, he also obtained data for three different rectangular bandwidths of unmodulated speech. The intelligibility scores were similar to those obtained for equivalent bands using a within-group design as shown in Fig. 5.

encompasses this mélange of strategies allowing listeners to understand speech both when intact and when acoustically degraded.

Acknowledgments

This work was supported in part by NIH, through Grant No. DC 00208.

References

- Bilger RC, Nuetzel JM, Rabinowitz WM, Rzeczkowski C. Standardization of a test of speech perception in noise. *J Speech Hear Res* 1984;27:32–48. [PubMed: 6717005]
- Greenberg, S. Understanding speech understanding: Towards a unified theory of speech perception. In: Ainsworth, WA.; Greenberg, S., editors. *Proceedings of the Workshop on the Auditory Basis of Speech Perception*. Keele University; UK: 1996. p. 1-8.
- Kalikow DN, Stevens KN, Elliott LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am* 1977;61:1337–1351. [PubMed: 881487]
- Pollack I. Effect of high pass and low pass filtering on the intelligibility of speech in noise. *J Acoust Soc Am* 1968;20:259–266.
- Remez RE, Rubin PE, Berns SM, Pardo JS, Lang JM. On the perceptual organization of speech. *Psychol Rev* 1994;101:129–156. [PubMed: 8121955]
- Shannon RZ, Zeng FG, Wygonski J, Kamath V, Ekelid M. Speech recognition with primarily temporal cues. *Science* 1995;270:303–304. [PubMed: 7569981]
- Silverman SR, Hirsh IJ. Problems related to the use of speech in clinical audiometry. *Ann Otol Rhinol Laryngol* 1955;64:1234–1245. [PubMed: 13283495]
- Steeneken HJM, Houtgast T. Mutual dependence of the octave-band weights in predicting speech intelligibility. *Speech Commun* 1999;20:109–123.
- Stickney GS, Assmann PF. Intelligibility of bandpass-filtered speech. *J Acoust Soc Am* 1997;102:3134. Abstract.
- Warren, RM. *Auditory Perception: A New Analysis and Synthesis*. Cambridge University Press; Cambridge: 1999.
- Warren RM, Bashford JA Jr. Intelligibility of 1/3-octave speech: Greater contribution of frequencies outside than inside the nominal passband. *J Acoust Soc Am* 1999;106:L47–L52. [PubMed: 10573914]
- Warren RM, Riener KR, Bashford JA Jr, Brubaker BS. Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits. *Percept Psychophys* 1995;57:175–182. [PubMed: 7885815]
- Watkins AJ, Makin SJ. Effects of spectral contrast on perceptual compensation for spectral-envelope distortion. *J Acoust Soc Am* 1996;99:3749–3757. [PubMed: 8655806]

FIG. 1.

Long-term averaged spectra based upon the 100 1/3-octave CID sentences having the nominal passband removed through steep FIR filtering. The isolated transition bands, having slopes of about 100-dB/octave, were presented either intact or truncated as shown with cutoffs at downpoints (DPs) along slopes of -30, -40, -50, or -60 dB, relative to the peak of the respective transition band. The plots were generated offline via discrete Fourier transform (DFT) with linear averaging of spectra obtained using a 8192-point (186 ms, 5-Hz resolution) Blackman window across the entire set of digitally recorded sentences. Metric Halo Laboratories, Inc. produced the DFT software.

FIG. 2.

Long-term averaged spectra of the 100 “triangular band” CID sentences produced with a common low-pass and high-pass cutoff of 1500 Hz for the 100-dB/octave filters. These contiguous transition bands were presented either intact or truncated with cutoffs at downpoints along slopes of -30 , -40 , -50 , or -60 dB, relative to the peak of the triangular band. The plots were generated offline as described for Fig. 1.

FIG. 3.

Long-term averaged spectra of the 100 CID sentences subjected to steep FIR filtering to produce rectangular bands with nearly vertical slopes that were each geometrically centered at 1500 Hz. Their bandwidths ranged from 3 to 12 semitones in steps of 1 semitone. The plots were generated offline as described for Fig. 1.

FIG. 4.

Mean intelligibility scores and standard error confidence limits ($N = 20$) for the transition band stimuli in experiment 1 and the triangular band stimuli in experiment 2, which were presented both intact and truncated at various downpoints.

FIG. 5. Mean intelligibility scores and standard error confidence limits ($N = 20$) for rectangular bands ranging in width from 3 to 12 semitones (1/4 octave to 1 octave) as presented in experiment 3.