Published in final edited form as:

Acoust Res Lett Online. 2000 October ; 1(2): 31–36.

Relative contributions of passband and filter skirts to the intelligibility of bandpass speech: Some effects of context and amplitude

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

Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin
53201-0413

James A. Bashford: bashford@uwm.edu; Richard M. Warren: rmwarren@uwm.edu; Peter W. Lenz: plenz@uwm.edu

Abstract

Warren *et al.* (1995) reported over 90% intelligibility for everyday sentences reduced to a 1/3-octave band (center frequency 1,500 Hz, slopes 100 dB/octave, slow-rms peak levels 75 dB). To investigate the basis of this high intelligibility, Warren and Bashford (1999) partitioned the sentences. Surprisingly, the rectangular 1/3-octave passband had only 24% intelligibility, whereas the filter skirts separated by a 1/3-octave notch had an intelligibility of 83%, despite their severe spectral tilts. Experiment 1 of the present study substituted monosyllabic words for sentences. Wholeband intelligibility was 26%, the passband 4%, and the filter skirts 16%. Experiment 2 measured intelligibility for 1/3-octave sentences having peak levels ranging from 85 down to 35 dB. Whole band intelligibility ranged from 90% to 68%, and the filter skirt pairs had from two to four times the passband's intelligibility (which did not vary significantly with level). Hence, steep (100 dB/octave) filter skirts make the dominant contribution to intelligibility of nominally 1/3-octave speech across a wide range of presentation levels.

1. Introduction

Warren, Riener, Bashford, and Brubaker (1995) reported that "everyday" sentences (Silverman and Hirsh, 1955) are more than 90% intelligible when limited to a nominal 1/3-octave band centered at 1.5 kHz and having conventionally steep filter skirt slopes of 100 dB/octave. Warren and Bashford (1999) subsequently partitioned the bandpass filtered sentences with 2000-order FIR filtering and found that the high intelligibility was due primarily not to the nominal 1/3-octave passband (only 24% intelligible when presented alone) but rather to the filter skirts, which were about 83% intelligible when presented without the passband. These results indicate a remarkable ability to compensate for severe spectral tilt and suggest that the conventional specification of bandwidth in terms of 3-dB downpoints along the filter skirts may be inappropriate and potentially misleading. The present study examines the generality of these findings using stimuli differing in the amount of contextual information and presentation level.

© 2000 Acoustical Society of America

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

PACS numbers: 43.71Es, 43.71

The high intelligibility of narrowband speech reported by Warren *et al.* (1995) was verified by Stickney and Assmann (1997) using high-predictable SPIN (Speech Perception in Noise) sentences. However, they also observed substantially lower intelligibility for keywords in low-predictable SPIN sentences and observed a further large decrease for keywords excised from the sentences. Experiment 1 of the present study was designed to determine whether the relative contributions of passband and filter skirts, as reported by Warren and Bashford (1999), are maintained in the absence of sentence context. For this purpose, lists of monosyllabic words were produced in citation form and subjected to the same 1/3-octave filtering and subsequent FIR partitioning procedures employed in the earlier study.

Warren and Bashford (1999) presented their sentence stimuli in the high-normal range, with slow-rms peak levels set at 75 dBA for the intact 1/3-octave band condition. Because level changes may substantially alter the range of audible frequencies in the filter skirts, it would be expected that their contribution to intelligibility would be more dependent upon level than that of the passband. Experiment 2 of the present study examined the effect of amplitude upon sentence intelligibility for the intact band and the isolated component conditions by varying peak levels over a 50-dB range.

2. General Method

2.1 Subjects

The 20 listeners in Experiment 1 and the 120 in Experiment 2 were undergraduate student volunteers from the University of Wisconsin. They either received experimental credit or were paid for their participation. They were native speakers of English who had no known hearing deficits.

2.2 Stimuli

The test stimuli presented in Experiment 1 were ten 50-word lists of phonetically balanced monosyllables (American National Standards Institute S3.2-1960 [R1971]). An additional 25 monosyllables served as practice stimuli for familiarizing listeners with the different filtering conditions. The test stimuli presented in Experiment 2 were based on the same recordings used by Warren and Bashford (1999). They were the 100 (10 lists of 10) "everyday" sentences, which contain 500 phonetically balanced keywords used for scoring (Silverman and Hirsh, 1955). An additional 25 practice sentences, used to familiarize listeners with the different filtering conditions, were drawn from the high-predictability subset of the revised SPIN test (Kalikow *et al.*, 1977; Bilger *et al.*, 1984). All stimuli were digitally recorded (44.1 kHz sampling, 16-bit quantization) by a male speaker having an average voicing frequency of approximately 100 Hz and no obvious regional accent.

Fig. 1 shows the long-term averaged spectra of the five filtering conditions presented in Experiments 1 and 2. A standard 1/3-octave band, geometrically centered at 1.5 kHz with filter skirt slopes of about 100 dB/octave, was produced via low-pass followed by high-pass filtering using the function *ellip* in MATLAB with order $n = 7$. These filtered words and sentences were digitally rescaled in level so that their slow-rms peaks were matched to within ± 0.2 dBA on playback (see stimulus A in Fig. 1). Then four reduced versions of the standard 1/3-octave band were produced using the function *fir1* in MATLAB with order $n = 2000$, resulting in slopes greater than 1,000 dB/octave. The reduced versions of the bandpass speech consisted of: 1. the standard band subjected to 1/3-octave bandpass FIR filtering from 1,336 to 1,684 Hz that removed both filter skirts (Fig. 1B); 2. the standard band subjected to FIR notch filtering at the same cutoffs, removing the passband (Fig. 1C); 3. the standard band FIR lowpass filtered, removing the upper filter skirt (Fig. 1D); and 4. the standard band FIR highpass filtered, removing the lower filter skirt (Fig. 1E).

2.3 Procedure

In both experiments, the five filtering conditions were presented within-subjects and assigned pseudorandomly to successive pairs of stimulus lists (word lists in Experiment 1, sentence lists in Experiment 2) so that each filtering condition was applied an equal number of times (4) to each list across subjects. Prior to receiving the test stimuli in a given filtering condition, listeners were presented with 5 practice stimuli. Each word or SPIN sentence was first heard broadband and then, following a 1-s delay, filtered in the same manner as the test stimuli that followed. The stimuli were delivered diotically through Sennheiser HD 250 Linear II headphones. For the monosyllable word lists in Experiment 1, slow-rms peak levels were 75 dBA for the intact band (Fig. 1A) and averaged 73 dB for the passband (Fig. 1B), 69 dB for the pair of filter skirts (Fig. 1C), 74.5 dB for the passband plus lower skirt (Fig. 1D), and 73.5 dB for the passband plus upper skirt (Fig. 1E). These relative level differences across filtering conditions also apply to the sentences presented in Experiment 2. Five separate groups of listeners ($N = 20$) were each presented the sentences at a different level. Those levels, measured as slow-rms peaks for the intact, 1/3-octave band condition were 85, 65, 55, 45, and 35 dBA, respectively. The data obtained earlier by Warren and Bashford (1999) for sentences presented at 75 dB were also included in the present analysis.

3. Experiment 1 results

Fig. 2 presents the intelligibility scores for the five filtering conditions for the monosyllabic words presented in Experiment 1 and the sentences presented under equivalent conditions in the earlier study by Warren and Bashford (1999). Although intelligibility overall was much lower for isolated monosyllables, the relative performance across the various partitioning conditions is similar to that observed for the sentences. Analysis of variance [$F(4,76) = 34.5$, $p < .0001$] and subsequent Tukey HSD tests ($p < .01$) indicated that monosyllable intelligibility was highest (about 26%) for the intact 1/3-octave band and lowest (only 4%) when the 100-dB/octave filter skirts were removed, leaving only the nominal passband. In contrast, intelligibility of the isolated pair of filter skirts was 16% or about 4 times that of the passband. Thus, the filter skirts contributed proportionately at least as much to intelligibility as they did when sentential context was present.

4. Experiment 2 results

Fig. 3 presents sentence intelligibility scores for the five filtering conditions across presentation levels in Experiment 2.

As expected based on the findings of Warren and Bashford (1999), intelligibility was lowest overall for the isolated passband, averaging 22% across levels, and was highest for the intact band, which averaged 85% intelligibility across levels [$F(4,456) = 463.1$, $p < .0001$; $p < .01$ by Tukey HSD test]. Also in keeping with the earlier findings, the isolated pair of filter skirts produced substantially higher intelligibility than the isolated passband at all presentation levels [$F(4,76) \geq 38.2$, $p < .0001$; $p < .05$ or better by Tukey HSD test] and ranged from twice the passband's intelligibility at 35 dB to four times its intelligibility at 65 dB.

Averaged across filtering conditions, intelligibility was highest at the 75-dB level used by Warren and Bashford (1999) and was significantly lower at 85 dB and lowest in the 35-dB condition [$F(5,114) = 21.8$, $p < .0001$; $p < .05$ by Tukey HSD tests]. Simple effects tests revealed that the isolated passband was unaffected by presentation level [$F(5,114) = 1.2$, $p > .3$], whereas intelligibility of the intact band, although equal to about 90% at intermediate levels from 75 to 55 dB, did decrease significantly at the lower presentation levels of 45 and 35 dB, [$F(5,114) = 23.0$, $p < .0001$; $p < .05$ by Tukey HSD test]. Given the nonsignificant

effect of level on performance with the isolated passband, it then appears that the reduction of intelligibility for the intact band at the lowest presentation levels is due to effects of level upon information in the filter skirts.

The overall intelligibility decrement observed at the 85-dB presentation level appears to be due chiefly to the spread of masking, which would be expected to degrade information in the steeply sloping filter skirts. The loss of intelligibility at the lower presentation levels for all conditions except the passband is an expected consequence of the reduced audibility and dynamic range of frequency components within the filter skirts of the intact band A and the reduced conditions C, D, and E. However, even at the lowest presentation level of 35 dB, intelligibility was nearly four times greater with the filter skirts presented along with the passband in condition A (68% intelligible) than when the filter skirts were absent in condition B (18% intelligible).¹

5. Summary

For stimuli as diverse as isolated monosyllabic words and everyday sentences, 1/3-octave passbands (centered at 1.5 kHz) make less than one-third the contribution to intelligibility made by flanking filter skirts having 100-dB/octave slopes. Filter skirt information dominates over a wide range of presentation levels (at least 50 dB) with levels from 55 to 75 dB yielding 90% intelligibility or better for 1/3-octave everyday sentences. The effect of level on intelligibility involves a tradeoff between audibility and masking of information in the filter skirts. Performance increased with level from 35 dB up to 75 dB, but a further increase in level from 75 to 85 dB significantly lowered intelligibility for all stimuli except the isolated passband, presumably due to the greater vulnerability of the sloping filter skirts to the spread of masking.

Acknowledgments

This work was supported in part by NIH through Grant DC 00208.

References

- ANSI. USA standard: Method for the measurement of monosyllabic word intelligibility. New York: American National Standards Institute; 1960, R1971. ANSI S3.2-1960.
- ANSI. American national standard methods for the calculation of the articulation index. New York: American National Standards Institute; 1969, R1986. ANSI S3.5-1969.
- ANSI. American national standard methods for the calculation of the speech intelligibility index. New York: American National Standards Institute; 1997. ANSI S3.5-1997.
- Bilger RC, Nuetzel JM, Rabinowitz WM, Rzeczkowski C. Standardization of a test of speech perception in noise. *J. of Speech and Hearing Res.* 1984; 27:32–48. [PubMed: 6717005]
- 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]
- Pavlovic CV. Band importance functions for audiological applications. *Ear and Hearing.* 1994; 15:100–104. [PubMed: 8194673]

¹Pavlovic (1994) summarized reports on the relative importance function of the articulation index for 1/3-octave bands of sentences. For bands centered near 1,500 Hz, only about 10% of the total information is present. Accepting this value, our results indicate that under the conditions we employed, 10% of the information available in the broadband sentences permits about 18–25% of the keywords in the 1/3-octave passband to be identified. However, the articulation index does not permit any estimate of the relative contributions of filter slopes upon intelligibility, because any slopes greater than 18 dB/octave may be used without influencing the values of the frequency importance functions (see ANSI S3.5-1969, R1986, 1997).

- Silverman SR, Hirsh IJ. Problems related to the use of speech in clinical audiometry. *Annals of Otology, Rhinology, and Laryngology*. 1955; 64:1234–1245.
- Stickney GS, Assmann PF. Intelligibility of bandpass-filtered speech. *J. Acoust. Soc. Am.* 1997; 102:3134A.
- 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]
- Warren RM, Bashford JA Jr. Intelligibility of 1/3-octave speech: Greater contribution from frequencies outside than inside the nominal passband. *J. Acoust. Soc. Am.* 1999; 106:L47–L52. [PubMed: 10573914]

Fig. 1.

Long-term averaged spectra based upon the 100, 1/3-octave "everyday" sentences: (A) passband with flanking filter skirts; (B) passband alone; (C) filter skirts alone; (D) passband and lower filter skirt; (E) passband and higher filter skirt. Plots were generated offline via discrete Fourier transform with linear averaging of spectra obtained using an 8192-point (186 ms, 5 Hz resolution) Blackman window.

Fig. 2. Mean intelligibility scores with standard error confidence limits ($N = 20$) for the monosyllabic words presented in Experiment 1 and the sentences presented by Warren and Bashford (1999) under five filtering conditions: (A) an intact 1/3-octave band with 100 dB/octave filter skirts; (B) the passband alone; (C) the filter skirts alone; (D) the passband and lower filter skirt; and (E) the passband and higher filter skirt.

Fig. 3. Mean intelligibility scores with standard error confidence limits ($N = 20$) for everyday sentences presented at 6 different slow-rms peak levels under five filtering conditions, including: (A) an intact 1/3-octave band with 100 dB/octave filter skirts; (B) the passband alone; (C) the filter skirts alone; (D) the passband and lower filter skirt; and (E) the passband and higher filter skirt.