Veterans Administration

Journal of Rehabilitation Research and Development, Vol. 23 No. 1 1986
pages 17-24

Multichannel Syllabic Compression[a] for Severely Impaired Listeners

STEVEN DE GENNARO, Sc. D.
LOUIS D. BRAIDA, Ph. D.
NATHANIEL I. DURLACH, M.A.

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Abstract — Two listeners with congenital hearing losses characterized by flat audiograms and dynamic ranges of 18-33 dB were tested with three compression systems and one (reference) linear amplification system. The compression systems placed progressively larger amounts of speech energy within the listener's residual dynamic range, by raising to audibility and compressing 25, 50, and 90 percent of the short-term input amplitude distribution in each of 16 frequency bands. The comparison linear system was defined by adjusting six octave-wide bands of speech to comfortable levels. System performance was evaluated with nonsence CVC syllables presented at a constant input level and spoken by two talkers. Extensive training was provided to ensure stable performance. The results were notably speaker-dependent, with compression consistently providing better performance for one speaker, linear amplification for the other. Averaged over speakers, however, there was no net advantage for any of the compression systems for any listener. The use of high compression ratios and large input ranges tended to degrade perception of initial consonants and vowels. Under some conditions, however, final consonant scores were higher with compression than with linear amplification. Compression generally enhanced the distinction between stops and fricatives, but degraded spectral-concentration and relative-intensity cues required to identify place of articulation.

INTRODUCTION

Amplitude compression has often been suggested as a means of overcoming the reduced dynamic range and the recruitment characteristic of many types of sensorineural hearing loss: e.g., Braida et al., 1979 (1). However, syllabic compression has not generally improved speech reception for moderately impaired listeners when compared with carefully chosen linear amplification coupled with some form of overall level normalization: e.g. (2, 3, 4, 5). This study, further details of which are available in De Gennaro, 1982 (6), was concerned with assessing the utility of multiband syllabic compression for listeners with severe sensorineural hearing loss, since these listeners represent the population believed most likely to benefit from some form of amplitude processing: see Vilchur, 1973 (7), and Peterson, 1980 (8). There was, however, no clear method of specifying an "optimum" compression strategy for a particular listener, due to both the lack of prior analytic studies and the large number of free parameters required to specify a multiband compression system.

[a]This research was supported by a grant from the National Institute of Neurological and Communicative Disorders and Stroke, Grant No. R01-NS12846.

[b]Dr. De Gennaro is currently with the IBM Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598

NOTE: Please address correspondence to: L. D. Braida, 36-747 M.I.T., Cambridge, Massachusetts 02139.

The purpose of this research was to conduct experiments to define appropriate compression characteristics (given properties of the hearing loss and the speech materials) and also to evaluate the usefulness of compression for severely impaired listeners relative to carefully chosen linear amplification.

METHODS

Compression System Design

Intuitively, it seems that there is a trade-off involved in the use of compression to increase the speech energy presented to an impaired listener. While intelligibility should be enhanced by raising a wider range of speech sounds to audibility, the corresponding alterations of normal intensity cues can result in perceptual degradations. Therefore, compression of the entire speech range into the auditory area of a listener with a severe hearing loss is not likely to be optimal, since the high compression ratios required degrade resolution for both overall intensity and details of spectral shape. On the other hand, if only peak speech levels were made audible, intelligibility would also be low because many speech elements would remain inaudible.

To determine the "optimal" range of speech levels to present to an impaired listener, three compression systems which differed parametrically in the amount of speech energy presented above the listener's elevated detection threshold were investigated. These systems placed 25, 50, or 90 percent of the short-term amplitude distributions in each of 16 frequency bands within the listener's residual auditory area.

The static characteristics for these systems were linear below the compression thresholds in each band, and employed a single compression ratio within the compression region. In each band, the compression threshold was set at the desired cumulative level (25, 50 or 90 percent), and the gain reflected the amplification required to raise the cumulative level to the listener's detection threshold. The degree of compression applied to each band was determined through an empirical fitting process in which listeners chose among systems with identical compression thresholds and band gains, but different static compression ratios.

To provide a reasonable match between the speech spectrum and the listener's narrow residual dynamic range, the fitting procedure was conducted in six separate frequency bands. Four of the frequency bands, centered at 500, 1000, 2000, and 4000 Hz, were each one octave wide. The lowest frequency band, centered at 200 Hz., was one and two-thirds octaves wide, and the highest frequency band, centered at 7100 Hz., was two-thirds of an octave wide.

A 16-band compression system, as described by Coln in 1979 (9), with bandwidths roughly equal to critical bandwidths, true rms detectors, and independent compression applied to each band, was used to process speech signals. An independent rms detector was used to determine the short-term signal level for each band. The time constant of the detector was inversely proportional to the width of the corresponding input filter, so that the faster level-variations in the wider bands could be followed. Detector time constants were determined by single-pole lowpass filters and varied from 18 ms in the lowest frequency band to 0.2 ms in the highest frequency band: see de Gennaro et al., 1981 (10). In the final system implementations, the compression ratio selected for each of the 16 bands was derived by linearly smoothing the compression ratios chosen by the listeners across frequency.

Linear System Design

The comparison linear amplification system was a variant of the OMCL (Octaves at Most Comfortable Level) system defined by Lippmann et al. in 1981 (5). It was chosen because it had generally provided high levels of performance when compared to other linear and compression systems in several previous studies. The frequency gain characteristic for the linear amplification system was determined empirically by having the listeners adjust isolated bands of speech to levels that were consistent with maximum intelligibility and long-term comfort.

The same six frequency bands used to determine the static compression ratios for the compression systems were used in this fitting process.

Subjects

Two listeners, both of them 29-year-old women with severe bilateral sensorineural hearing loss of congenital origin, participated in this study. Each had relatively constant dynamic range over the frequency range of 125 to 8000 Hz. Both ears were tested for one listener, who is identified as "Subject TTR" for the right ear results and as "Subject TTL" for the left ear results. This listener used a hearing aid for her right ear only.

One ear was tested for the second listener, "Subject PBR". This listener also used a hearing aid in her right ear only.

Pure tone detection thresholds were determined

through an adaptive forced-choice paradigm while discomfort thresholds were measured with pulsed tones using an adaptive tracking algorithm. Detection and discomfort thresholds are presented in Figure 1.

Subject PBR had the most severely restricted residual dynamic range, with a mean dynamic range of only 18 dB between 125 Hz and 8000 Hz. Subject TTL had an average range of 21 dB, while subject TTR had the largest residual range with an average of 33 dB.

Speech Materials

The speech materials used in the identification experiments were nonsense CVC syllables spoken in an /a/-CVC context by two talkers, one male (Speaker BH), and one female (Speaker CL). The CVC syllables were formed randomly from a set of 16 initial consonants, 6 vowels, and 16 final consonants. The consonant set consisted of /p, t, k, b, d, g, f, θ, s, ʃ, v, γ, z, ʒ, tʃ, dʒ, h/; the vowel set /ɛ, a, i, I, u, U/.

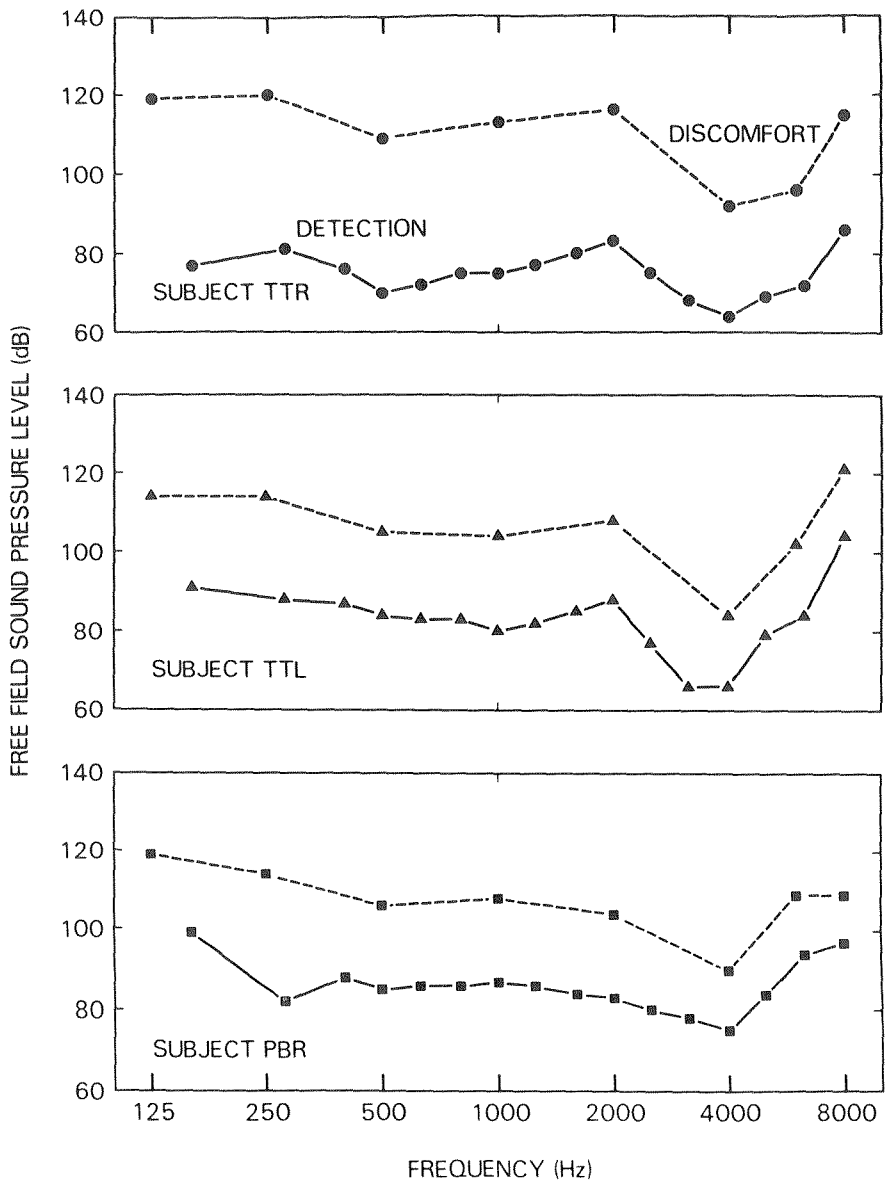


FIGURE 1
Tone detection and discomfort thresholds of the subjects tested.

TABLE 1
Average Static compression ratios*

Speaker BH			
	System		
Subject	25%-C	50%-C	90%-C
TTR	1.1	1.8	1.8
TTL	2.1	3.7	4.2
PBR	1.8	3.6	3.9

Speaker CL			
	System		
Subject	25%-C	50%-C	90%-C
TTR	1.3	1.6	1.9
TTL	1.8	3.1	4.0
PBR	2.1	2.9	3.3

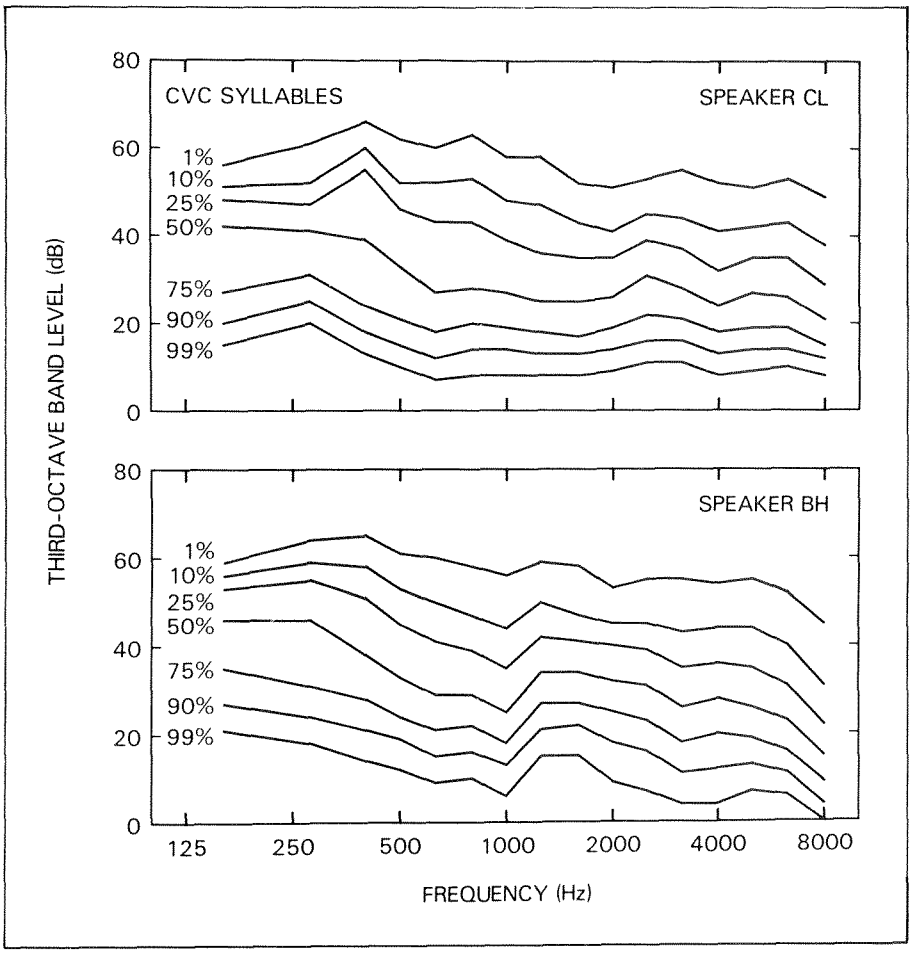
*Results have been averaged over the 16 frequency-dependent compression ratios used for each subject and each system.

The 800 CVC syllables for each speaker were stored digitally (with 12-bit resolution and 10-kHz audio bandwith) to allow random presentation sequences, in order to reduce order effects and learning of stimulus artifacts. In addition, the overall rms level of each utterance was normalized. The cumulative amplitude level distributions derived (10) from measurements made on 800 CVC syllables spoken by each of the talkers are presented in Figure 2. These distributions, along with the detection thresholds for each subject, were used to define the compression thresholds and band gains for the 25, 50, and 90 percent compression systems.

Amplitude Properties of Processed Speech Materials

In general, the listeners were quite consistent in the selection of static compression ratios for the 25, 50, and 90 percent systems, and in the selection of the frequency gain characteristic for the linear OMCL system. The compression ratios chosen are presented in Table 1. The subjects generally responded to increases in the range of audible speech, from 25 to 50 to 90 percent, with higher effective

FIGURE 2
Cumulative amplitude level distributions for the speech materials used in the intelligibility tests.



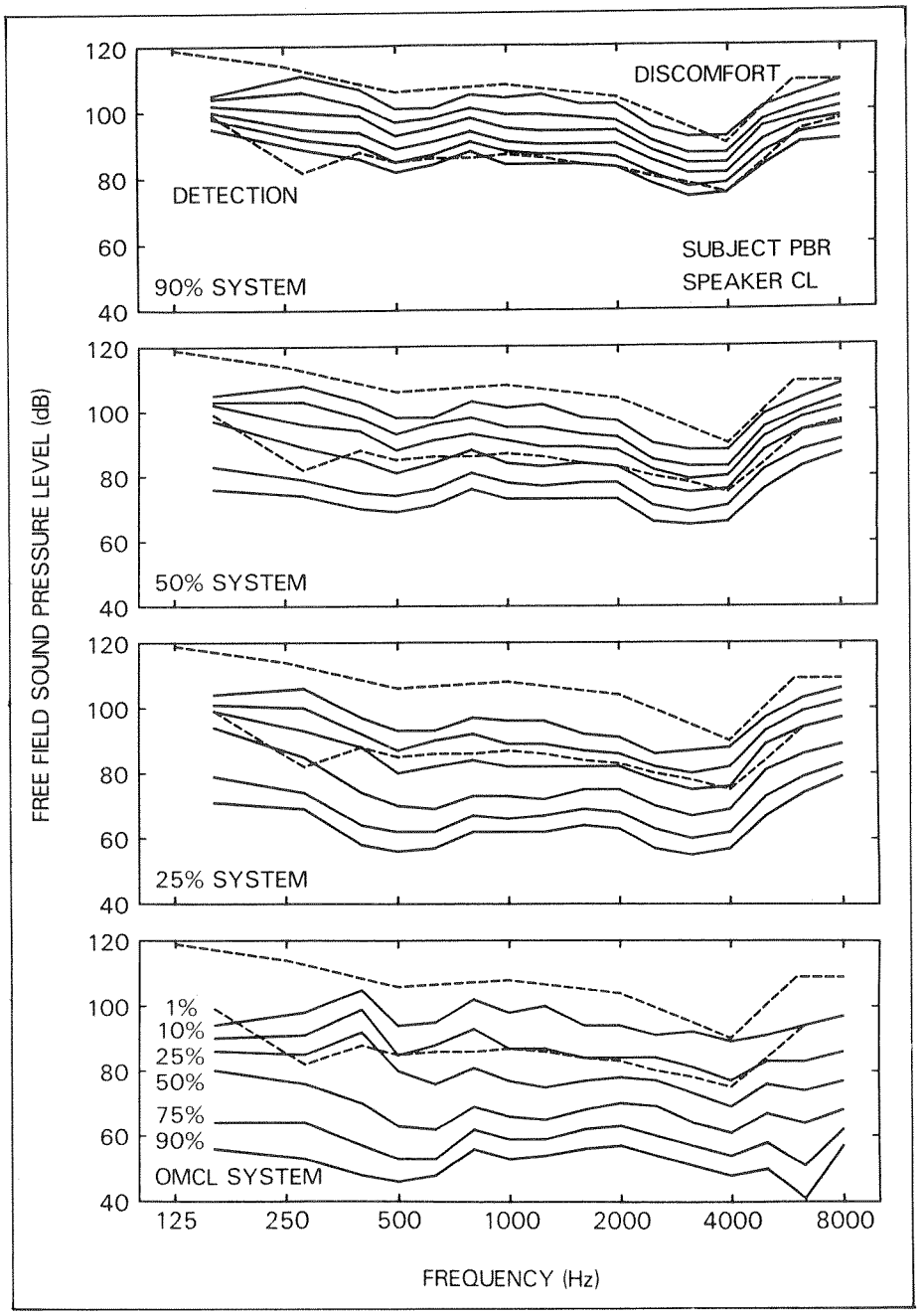


FIGURE 3
Relation of cumulative level distributions for processed speech materials (speaker CL) to the detection and discomfort levels for subject PBR.

compression ratios. Also, smaller compression ratios were chosen by the subject with the largest dynamic range.

After fitting, amplitude distributions of the short-term output levels for each of the systems investigated were measured. For example, the cumulative level distributions for the three compression systems and the comparison linear system for subject PBR and speaker CL are shown in Figure 3. Measurements such as these suggest that the subjects selected compression characteristics that presented the peak 1 percent speech levels significantly below the discomfort thresholds measured with pulsed tones.

The frequency gain characteristics chosen by the subjects also tended to flatten the cumulative level distributions across frequency to provide a better match to the detection thresholds. For the less severely impaired subject, TTR, approximately 50 percent of the speech range was presented above threshold with the OMCL linear system. For subjects TTL and PBR, roughly 10-to-25 percent of the speech range was audible with linear amplification. Again, the subjects selected presentation

levels that held the 1 percent peak levels significantly below the measured discomfort levels in most frequency bands.

Identification Experiments

Given the severity of the impairments and the number of experimental conditions, extensive training was provided to ensure a stable basis for system comparisons. In a study of compression published by Peterson in 1980 (8), large variations in performance were observed during initial presentations of both compressed and linearly processed materials. Consistent performance was achieved only after 8,000 to 10,000 trials for a severely impaired subject; even then, additional learning appeared to take place, but at a fairly slow rate.

The experiments in this study were structured to separate training effects from the comparative analysis of system performance. Identification experiments were conducted over 16 sessions for each subject. In each session, all four systems were presented with a practice list of 50 CVC items, followed by a test list of 150 to 200 CVC items. The results from the first 8000 trials were considered "training" and were not included in the final system comparisons. For the smallest unit of analysis (1 system, 1 speaker, after training), each subject completed 1000 identification trials. Discarding the 200 practice trials, this provided roughly 50 presentations of each initial consonant, 133 presentations of each vowel, and 50 presentations of each final consonant.

The primary measures of system performance in this study were the average percent correct scores for the initial consonant, vowel, and final consonant components of the CVC test materials. Overall system performance was characterized by the phoneme and item percent correct scores.

RESULTS

Phoneme and item scores obtained by each subject with each system are presented in Table 2. In general, the three compression systems tested in this study were, at best, equivalent to the comparison linear system in terms of overall performance levels. There was no net improvement with compression, whether measured by phoneme or item scores, for any of the subjects. Compression was better than linear amplification only in the final consonant position for the two more severely impaired subjects, and these improvements with compression were small — between 4 and 7 percentage points.

For the least severely impaired subject, TTR, the 50 percent and 90 percent compression systems and the OMCL linear system were statistically equivalent, and were better than the 25 percent compression system by roughly 5 percentage points. In general, however, subject TTR achieved quite high levels of performance as demonstrated by phoneme scores of 75-to-81 percent across the linear and compression systems tested.

For the two more severely impaired subjects, TTL and PBR, the 25 percent and 50 percent compression systems and the OMCL linear system were generally equivalent in terms of overall performance, and were better than the 90 percent system by 6-to-11 percentage points. However, the performance scores for these subjects were rather low, ranging between 52 percent and 69 percent of phonemes correct.

The results of the identification experiments, however, were highly speaker-dependent for the two more severely impaired subjects. Compression consistently improved performance with one speaker (CL) while degrading performance with the other (BH). In many instances, the difference in performance across speakers was comparable to or larger than the corresponding differences across systems. For example, for the 90 percent system, scores for speaker CL were 14 percentage points higher than for speaker BH — while for the OMCL system, they were 8 points lower. Although the causes of this speaker dependence are not well understood, it may be due to differences in articulation and differences in masking effects associated with greater concentration of low-frequency energy with speaker BH. There may also have been more significant in-

TABLE 2
Identification scores*
Phoneme and item scores

Phoneme identification scores				
	System			
Subject	OMCL-Lin	25%-C	50%-C	90%-C
TTR	80.1	75.4	81.3	80.4
TTL	67.9	69.0	67.5	57.4
PBR	57.5	59.1	56.0	51.6

Item identification scores				
	System			
Subject	OMCL-Lin	25%-C	50%-C	90%-C
TTR	50.5	40.9	52.3	50.6
TTL	29.4	32.8	30.0	21.1
PBR	14.9	17.1	14.8	12.9

*Scores have been averaged over the two speakers.

teractions between the dynamic properties of the compressor and the lower fundamental frequency of speaker BH.

The systems produced different error patterns — Even when comparable levels of performance were found across the systems tested, analysis of the response error patterns indicated that phonetic feature reception was significantly different under compression and linear amplification. While consonant feature reception was usually influenced by speaker and by syllable position, several broad trends were noted, particularly for the two more severely impaired subjects. Compression generally improved the subject's ability to distinguish between stops and fricatives. On the other hand, compression degraded the spectral-concentration and relative-intensity cues required to identify the particular stop or fricative that had been presented. And under compression, place of articulation and duration were significantly degraded. The observed confusions under compression are consistent with the spectral flattening introduced by the independent action of the 16 bands in the speech processing system.

With compression, /θ, ʃ, s/ were often labeled as /f/, and /γ, ʒ, z/ were often confused with /v/. Confusions were also common between /ʒ/ and /z/, and between /ʃ/ and /s/ with compression. With linear amplification, confusions between stops and fricatives (for example, between /θ/ and /t/, /γ/ and /d/, and between /b, d, g/ and /v/) were more common.

With the 90 percent compression system, vowel perception was also degraded for the two more severely impaired subjects, by as much as 8 to 20 percentage points. These subjects often incorrectly responded with /ɛ/ to presentations of the other five vowels in the test set. This result is also consistent with the spectral flattening caused by independent compression in the different bands of the compression system, since /ɛ/ was the most neutral of the vowels tested, with the most uniformly spaced formants and the flattest spectrum.

CONCLUSIONS

The compression systems tested in this study were at best equivalent to the comparison linear system when performance was measured by either phoneme or item scores for isolated CVC syllables. Compression did improve final-consonant scores for the two more severely impaired subjects, but it often significantly degraded initial-consonant and vowel scores. The results of the identification experiments were highly speaker-dependent. Compression consistently improved performance with one speaker while degrading performance with the other. Averaged over speakers, however, there was no net advantage for compression.

This study was also concerned with determining appropriate compression system characteristics given properties of the subject's hearing loss. Usually, the best performance was achieved with systems employing moderate compression ratios and compression regions spanning the top 25-to-50 percent of the short-term input speech range. Compression of 90 percent of the input range into the residual auditory area often resulted in substantially degraded speech reception.

High levels of performance were generally achieved with linear amplification for the less severely impaired subject, for speech materials that were presented at a constant input level. It has been suggested that the use of speech materials with fixed overall presentation levels provides a false advantage for comparison linear systems, since compression is capable of equalizing presentation level differences. While it is clear that the introduction of significant item-to-item level variation will degrade performance with linear amplification, the results of these experiments indicate that performance will also be significantly degraded with compression. If syllabic compression is used to compensate for overall presentation level variations, then compression ratios and input ranges much larger than those used in the current study would be required. Given the degradations observed with the 90 percent system, it is likely that compressing wider input ranges into the subject's residual auditory range, with high compression ratios and a large number of independently processed bands, would result in severely degraded performance. Similar degradations have been observed by Bustamante for multiband compression-limiters (11).

The results of this study suggest that well-chosen linear amplification, coupled with some form of automatic volume control to reduce long-term level variations, may be more beneficial than syllabic compression for subjects with flat losses and dynamic ranges of 30 dB or more. While normal hearing was not restored by the systems tested in this study, the less severely impaired subject achieved phoneme scores of roughly 80 percent on nonsense CVC syllables under linear amplification with materials presented at a fixed overall level.

The performance scores for the subjects with more severely restricted dynamic ranges were generally

quite low even under the best presentation conditions. For those subjects, it is clear that a simple mapping of the input speech range into the residual auditory area is inadequate, particularly if the mapping is carried out independently in a large number of frequency bands. While it is certainly necessary to overcome the lack of audibility, the mapping must also preserve important acoustic differences between speech sounds. A more appropriate compression strategy would reduce overall level variations while maintaining, or perhaps emphasizing, relevant acoustic cues. For example, the relatively good performance achieved with 25 percent of the speech range above threshold suggests that preservation of spectral peaks is more important to intelligibility than the audibility of lower amplitude speech levels. Future work should concentrate on systems that employ interband control algorithms to preserve relevant short-term features while reducing overall level variations ■

REFERENCES

1. Braida LD, Durlach NI, Lippmann RP, Hicks BL, Rabinowitz WM, Reed CM: Hearing Aids: A Review of Past Research on Linear Amplification, Amplitude Compression, and Frequency Lowering. ASHA Monograph Number 19, 1979.
2. Barfod J: Multichannel Compression Hearing Aids: Effects of Recruitment on Speech Intelligibility. Report No. 11, The Acoustics Laboratory, Technical University of Denmark, 1976.
3. Abramovitz R: The Effects of Multichannel Compression, Amplification, and Frequency Shaping on Speech Intelligibility for Hearing Impaired Subjects. Unpublished Ph. D. Thesis, The City University of New York, 1979.
4. Braida LD, Durlach NI, De Gennaro SV, Peterson PM, Bustamante DK: Review of Recent Research on Muliband Amplitude Compression for the Hearing Impaired. **In** The Vanderbilt Hearing-Aid Report, State of the Art Research Needs, G. A. Studebaker and F. H. Bess, Eds. Upper Darby, Pennsylvania: Monographs in Contemporary Audiology, 1982.
5. Lippmann RP, Braida LD, Durlach NI: Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss. J Acoust Soc Am 69(2): 524-534, 1981.
6. De Gennaro SV: An Analytic Study of Syllabic Compression for Severly Impaired Listeners. Unpublished Ph. D. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., Cambridge, Mass., 1982.
8. Peterson PM: Further Studies of the Perception of Amplitude Compressed Speech by Impaired Listeners. Unpublished S.M. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., Cambridge, Mass., 1980.
7. Villchur E: Signal processing to improve speech intelligibility in perceptive deafness. J Acoust Soc Am 53:1646-1657, 1973.
9. Coln MCW: A Computer Controlled Multiband Amplitude Compressor. Unpublished S.M. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., Cambridge, Mass. 1979.
10. De Gennaro SV, Krieg KR, Braida LD, Durlach NI: Third-Octave Analysis of Multichannel Amplitude-Compressed Speech. Proceedings of IEEE ICASSP, 1981.
11. Bustamante DK: The Effect of Compression Limiting on Speech Intelligibility for the Sensorineural Hearing Impaired. Unpublished S.M. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., Cambridge, Mass., 1981.