Multichannel compression processing for profound deafness

EDGAR VILLCHUR
Foundation for Hearing Aid Research, Woodstock, New York 12498

Abstract—Three-channel amplitude compression followed by frequency shaping was used to process test sentences for five profoundly deaf subjects, and the recognition scores were compared to scores achieved with frequency shaping only. At preferred levels, the scores of three of the five subjects showed a statistically significant but not dramatic advantage for compression; the averages of the scores for these three subjects were 29.2 percent for the uncompressed speech and 39.5 percent for the compressed speech. The preferred-level scores of the other two subjects did not show a statistically significant advantage for compression; averages were 33.7 percent for uncompressed speech and 36.1 percent for compressed speech. Tests at input levels 10 dB and 15 dB below preferred levels were also given to four of the subjects (the fifth had to leave the experiment early). In the reduced-level tests all four subjects showed a statistically significant advantage for compression. The averages of reduced-level uncompressed scores for the four subjects were 10.7 percent and 15.2 percent at -15 dB and -10 dB levels, compared to compressed scores at these levels of 31.7 percent and 32.5 percent. When visual cues were added to the auditory presentation in an exploratory experiment with one subject, the benefit of compression carried over into the higher scores.

INTRODUCTION AND HISTORY

Villchur (17) reported a significant improvement in speech recognition by hearing-impaired subjects, both in quiet and with speech interference, when the signal was processed by two-channel amplitude compression followed by frequency shaping. The subjects had moderately severe to severe sensorineural hearing impairment. Both high-frequency and low-frequency shaping were used in the processing, but the improvement was measured relative to an uncompressed reference signal whose only frequency shaping was in the region below 750 Hz. The rationale for not using high-frequency emphasis in the linear reference was that a hearing aid with high-frequency emphasis (but without compression) might amplify real-life sounds like the ringing of a telephone or the clatter of silverware to unacceptable levels, and that the relative high-frequency level of the reference signal used in the experiment was at least equal to that provided by the real-ear response of hearing aids at the time, which is to say equal to the relative high-frequency level that was known to be acceptable in the real world.

Of nine studies of multichannel compression published since then, five have provided at least qualified confirmation of Villchur’s results, and four have contradicted them. Yanick (20) reported results similar to Villchur’s, showing even greater improvement; Gregory and Drysdale (6) briefly confirmed Villchur’s results and then described an alternate method of compression involving high-frequency carrier clipping; Mangold and Leijon (11) reported an advantage to compression processing for some subjects, particularly in noise, over a linear reference that included high-frequency shaping, but they differed from Villchur in that a high compression threshold and high compression ratios were used;
Lippmann et al. (10) found an advantage to compressed speech at optimum level over a flat reference, an advantage that disappeared when the linear reference had proper frequency shaping but reappeared at lower input levels to the compressor; and Laurence et al. (9) showed laboratory results similar to those of Lippmann et al., and a marked advantage to multichannel compression processing when it was used in a wearable hearing aid.

On the negative side, Barfod (2) found no advantage to compression over a linear reference with frequency shaping; O’Loughlin (13) concluded that compensation for recruitment (accelerated loudness growth) was not an important element in improving speech intelligibility for the hearing impaired (although the scores of two of his six subjects were increased by compression over a linear reference with frequency shaping, one from 30 percent to 49 percent); Nábělek (12) found no advantage to multichannel compression in quiet and a disadvantage in noise and/or reverberation; and Walker et al. (19) reported no general benefit from compression, although they considered the subject still open. Walker et al. used high compression thresholds and an expansion mode below compression threshold, so that a large part of the speech elements of their test material was subjected to expansion rather than compression.

In the present compression experiments, the uncompressed reference differs from that used in Villchur’s 1973 study in that it is processed by both low- and high-frequency shaping. The change was made for the following reasons: 1) High-frequency emphasis in terms of true insertion gain is now used successfully in non-compression hearing aids; 2) Lippmann et al., who confirmed Villchur’s results against a flat reference, showed the importance of high-frequency emphasis in the reference signal; and 3) Skinner (15) showed that high-frequency emphasis without compression can have significant benefit for moderately impaired subjects. These findings are in contrast to the earlier report of Davis et al. (4) that 8 out of 12 subjects with accentuated high-frequency loss did not benefit from a 6 dB/oct high-frequency emphasis.

The linear-reference scores of the 1973 Villchur experiment were reduced by the lack of high-frequency emphasis in the uncompressed presentations. I suggest that the linear-reference scores in many of the compression studies that followed were increased by the artificially small dynamic range of the test presentations. When the dynamic range in any one frequency band of the test speech is small, simple frequency shaping is more likely to be capable of placing the test material within the residual dynamic range of hearing of a subject, even though real-life speech might not fit.

Speech tests, including those used in compression studies, are routinely recorded with the talker monitoring a vu meter to keep his or her voice level constant. This procedure puts a human compressor into the circuit, providing partly precompressed speech. The intersyllabic and word-to-word amplitude changes of speech have been almost eliminated in a test that is intended to evaluate the effect of compression on amplitude changes. *Intrasyllabic* amplitude differences remain, but these make up only part of the dynamic range of speech encountered in normal social communication. The dynamic range of conversational speech is increased by the stress given to some words and syllables and the drops in level assigned to others, by differences in characteristic talker levels, and by talker distances that vary from several feet to the very short distance for the listener’s own voice. To test the effects of compression with single-syllable material at a single level is comparable to testing the effects of high-frequency emphasis with speech material from which most words with high-frequency consonants have been eliminated. In each case, the choice of test material prevents the operative element of the processing from being effective.

The purpose of compressing speech for the hearing-impaired listener is to fit the large dynamic range of speech encountered by a hearing aid user into the restricted residual dynamic range of hearing resulting from recruitment. Amplitude compression is thus likely to be useful only if the listener’s residual dynamic range of hearing is restricted enough to reject, even after optimum frequency shaping, weak acoustic cues significant to speech recognition. For this condition to be represented in an experiment, the recruitment of the subject must be great enough, and the speech must have a dynamic range large enough, so that at comfortable overall levels significant cues fall below the subject’s dynamic range of hearing. In this experiment the dynamic range of the peaks within a typical test list was about 7 dB, and the test lists were presented to each subject at three different levels, as described in
“Procedure” under Test Materials and Presentation.

A compressor is commonly thought of as a device that reduces gain as the signal level increases. In the present application it is more useful to think of a compressor as a device that increases gain as the signal level decreases, so that low-intensity elements of speech can be amplified into the reduced dynamic range of a subject’s hearing while high-intensity elements remain at optimum level.

EXPERIMENT 1: Auditory presentation without visual cues

Subjects

Five profoundly deaf adults aged 32 to 70, 4 male and 1 female, served as subjects. Table 1 shows the hearing threshold levels and other data for each subject. (The hearing threshold level, formerly called the hearing loss, is the difference in dB between normal threshold SPL and impaired threshold SPL.)

All subjects had had at least acceptable hearing at one time, all had good speech, and all had been given very low discrimination scores in their standard audiological workups (4 to 14 percent on W-22 lists). The use of subjects whose onset of profound deafness had occurred prior to the acquisition of language would have added an unnecessary problem to the experiment: the presentation to subjects of language sounds that were unfamiliar.

The subjects’ residual dynamic ranges of hearing for speech, as reflected in their choices of preferred speech levels relative to their hearing thresholds, was greater than had been anticipated, possibly because of a conductive component in their deafness. The short-term maximum of the unprocessed speech level chosen by each subject is listed in Table 1.

Equipment

A block diagram of the processing circuit is shown in Figure 1. The thresholds of compression for low, middle, and high channels were set at 30 dB, 35 dB, and 40 dB below the maximum short-term, all-pass speech level. The latter level was adjusted for 0 dB on the input vu meter, ignoring occasional overshoots—following the vu-measurement procedure described in ASA Standard C16.5-1954 (1)—and was called the reference input level. When a reference-level signal was fed to any channel, the gain of that channel remained the same for all settings of the compression-ratio control. (The compression ratio is the numerical ratio of input level change in dB to output level change in dB.) When the input signal exceeded 0 dB on an occasional peak, the compressor reduced the channel gain; when the input signal level was lower than reference level the compressor increased channel gain. Thus, if an input signal were alternately 2 dB above and 8 dB below reference level in a channel whose compression ratio was 2, the corresponding output levels would be 1 dB above and 4 dB below the reference output. With this design the subject was able to adjust compression ratios for what he or she thought was maximum intelligibility without making any substantial change in the all-pass output level.

The reasons for using more than one compression channel are: 1) The subject’s residual dynamic range of hearing is restricted in different amounts in different frequency regions, requiring different amounts of compression in different regions; and 2)

Table 1.

Hearing threshold levels (subject threshold SPL minus normal threshold SPL, in dB) of the subjects of this experiment. NR means no response at the 120-dB HL limit of the audiometer. Where discomfort levels were within the capability of the audiometer they appear in parentheses, in HL’s. The short-term maximum of the unprocessed speech level chosen by the subject, in SPL, is also listed.

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>No. of yrs. of profound deafness</th>
<th>All-pass speech level</th>
<th>250 Hz</th>
<th>500 Hz</th>
<th>700 Hz</th>
<th>1 kHz</th>
<th>1.5 kHz</th>
<th>2 kHz</th>
<th>3 kHz</th>
<th>4 kHz</th>
<th>6 kHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject AB</td>
<td>32</td>
<td>23</td>
<td>119</td>
<td>70</td>
<td>85</td>
<td>80(115)</td>
<td>85(115)</td>
<td>85(120)</td>
<td>90</td>
<td>90</td>
<td>115</td>
<td>NR</td>
</tr>
<tr>
<td>Subject NJ</td>
<td>48</td>
<td>26</td>
<td>123</td>
<td>70</td>
<td>95</td>
<td>100</td>
<td></td>
<td></td>
<td>110</td>
<td>115</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Subject JS</td>
<td>69</td>
<td>4</td>
<td>114</td>
<td>55</td>
<td>70</td>
<td>105</td>
<td>NR</td>
<td>NR</td>
<td></td>
<td></td>
<td></td>
<td>NR</td>
</tr>
<tr>
<td>Subject BG</td>
<td>70</td>
<td>10</td>
<td>130</td>
<td>65</td>
<td>95</td>
<td>105</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NR</td>
</tr>
<tr>
<td>Subject JT</td>
<td>66</td>
<td>21</td>
<td>113</td>
<td>60</td>
<td>70(115)</td>
<td>(110)</td>
<td>80</td>
<td>105</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>
a multichannel compressor, unlike a single-band compressor, can act independently on simultaneous signals when each of the signals occupies a different frequency band.

The transition frequencies between channels, and the attack and release times for each channel were adjustable by the experimenter; the values used for these characteristics are discussed in the “Procedure” section. Either the subject or the experimenter could adjust the compression ratio for each channel, the low-frequency equalization slope in both linear and compression modes, the high-frequency equalization slope in both linear and compression modes, and the output level into the earphone. The subject made his adjustments on a small, lap-held control box. The experimenter could offer the subject a choice between two processing conditions by switching between two values of a particular characteristic. A printout provided an automatic record of the final processing characteristics in both linear and compression modes.

The compression ratios were continuously variable from 1.0 (no compression) to 20. The range of low- and high-frequency equalization, also continuously variable, is shown in Figure 2. The experimenter could choose between 400 Hz and 800 Hz as the transition frequency for low-frequency attenuation in either mode, and he could apply a 3-kHz, 18 dB/oct high-frequency rolloff to the signal in either mode.

The processing equipment was assembled into a single unit housed in a suitcase. The basic compressors were dbx 303S cards. Overall design and manufacture was by Etymotic Research, with initial assistance from Daniel Queen Associates.

Speech tests were recorded on cassettes and played through a TEAC A-170S cassette deck. The tests were presented to subjects through a TDH-39 earphone housed in a circumaural mounting described by Villchur (16) and made as a prototype by Telephonics. A real-ear, free-field calibration of the earphone had been performed, so that measurements of electrical input to the earphone were readily converted to SPL’s delivered to the subject.

Procedure

Processing calculations and adjustments—Preferred listening levels for unprocessed speech were measured, presenting each subject with practice lists of sentences similar to those to be used in the tests. The subject was asked to adjust a level control for maximum understanding consistent with all-day comfort, as though he or she were adjusting the volume control of a hearing aid. The subject’s residual range of hearing for speech was then calculated at half-octave intervals, as the number of dB between the subject’s pure-tone threshold SPL at a given frequency and the SPL at the same frequency of a hypothetical equal-loudness contour pegged to the subject’s preferred listening level for speech.

The absolute 500-Hz level of this equal-loudness contour was set at the short-term maximum of the subject’s preferred all-pass listening level. The con-
Figure 2.
Range of frequency equalization provided by the processor for either compressed or uncompressed signals. The two low-frequency rolloff curves are for 400-Hz and 800-Hz transition frequencies: the sharply rolled-off high-frequency curves represent the 3-kHz rolloff switched in.

tour was drawn with the same shape as the subject’s threshold curve but with half the instantaneous slope. This half-slope formula was based on an examination of the measured thresholds and equal-loudness contours of hearing-impaired subjects published by Villchur (17), by Barfod (2), and by Lippmann et al. (10), which showed that the formula produced a reasonable first approximation of the measured contour values. In Figure 3 a comparison is made between the 6-subject average of measured equal-loudness contours at preferred speech levels reported by Villchur, and the average contour for these subjects calculated from their thresholds.4

The normal dynamic range of hearing that was taken as analogous to the impaired-subject dynamic range described above was the distance at a given frequency between normal threshold and the 84-phon ISO (1961) equal-loudness contour. The 500-Hz level of the 84-phon contour, 80 dB SPL, is equal to the short-term, maximum all-pass rms level of male conversational speech5 reported by Dunn and White (1940), and therefore presumed to be comparable in loudness to the 500-Hz level used for the subject contour. Figure 4 shows the approximate range of sound-pressure levels for conversational speech measured in half-octave bandwidths by Dunn and White, and the proportionate position of these speech levels in the range between normal threshold and the 84-phon contour.

The compression ratios used for the initial presentation of processed speech to the subject were calculated for each channel, as the normal average dynamic range of hearing over the frequency range of the channel divided by the subject’s average dynamic range of hearing in that channel.

In Figure 5 the Dunn-White speech band, amplified to subject NJ’s preferred all-pass level, is plotted against her calculated residual dynamic range of hearing. A large part of the speech remains at levels below her useful hearing. The processing of this
experiment was designed to make more speech elements audible to the subject by reducing the amplitude range of the speech band (by compression) and then bending the band to a position within the residual dynamic range of the subject’s hearing (by frequency equalization). The ideal position for any speech element was assumed to be one between the subject’s threshold and equal-loudness contour proportionate to the corresponding position for normals shown in Figure 4. A curve of the maximum levels of speech placed in such a proportionate position is plotted in Figure 5, directly under the subject’s calculated equal-loudness contour. The effect of three-channel compression on maximum levels of the speech band in Figure 5, using compression ratios calculated for subject NJ, is also shown. The equalization at a given frequency that is needed to complete the processing is the difference at that frequency between the maximum level of the compressed, unequalized speech and the ideal level, a difference indicated by the arrows in Figure 5. Lower intensity speech elements will assume their proper levels as a result of the compression, except when they are produced at the same time as higher level elements in the same channel.

Equalization for the linear mode can be plotted in a similar way, calculating from maximum uncompressed speech levels at given frequencies. It is apparent from Figure 5 that while frequency equalization without compression can bring some of the inaudible speech elements into this subject’s dynamic range of hearing (such equalization amplifies only speech maxima to levels equal to those with compression), conversational speech has too large a dynamic range for the lower-intensity elements to fit.

The compression and equalization processing for each subject were calculated and placed in the
The choice of transition frequencies was made (shown in Table 2) based on the frequencies at which the subject’s residual dynamic range of hearing changed most abruptly. The subject was then asked to readjust, in the compression mode and in the linear mode where applicable, the overall level, low- and high-frequency equalization, and the compression ratios for each channel. Adjustments were made on the lap-held control box while the subject listened to a 20-second sentence repeated on an endless tape. The sentence was similar in character to the short sentences used in the actual tests, except that the dynamic range of peak vu deflections was about 10 dB. The experimenter could change the function of either of two knobs on the control box; one knob controlled independent compression ratios for each of the three channels and another knob controlled low- and high-frequency equalization in either the linear or the compression mode. When the subject had completed the adjustment of a particular characteristic, the value chosen was substituted for the previous setting of the processor and became operative for all further adjustments.

The repeatability of adjustment was sometimes poor, and so a second method of allowing the subjects to make choices was provided. Two values of a particular processing characteristic were placed in the processor memory. The endless tape and/or practice lists were then presented to the subject with each of these characteristics in turn, and the subject was asked to choose the one that provided better understanding. When the characteristic being tested was low-frequency equalization, different
values of equalization were accompanied by different level settings that had been chosen previously by the subject as appropriate to the particular equalization. The results of the choice procedure proved to be more repeatable than the results of the adjustment procedure. After the values of compression and equalization were set, the subject was given an opportunity to readjust the level in both linear and compression modes.

Table 2 shows the values of compression ratios and frequency equalization calculated for each subject, compared to the corresponding values that were chosen by the subject and used in the experiment. The 3-kHz cutoff was used in both compressed and uncompressed modes for all subjects except BG.

The compressor attack and release times used at first—approximately 1-msec attack and 20-msec release—were the same as those used in the earlier Villchur (17) two-channel experiment, in which the speech tests had consisted of single-syllable test items at different levels. In the present experiment, where whole sentences or phrases were used as test material, some of the subjects made unsolicited comments about the compressed speech to the effect that the words sounded jumbled together. With
longer release times there is less of a tendency for the quiet intervals between words to be filled in with reverberant and noise elements, and longer time constants were tried and approved by the subjects. The approximate attack and release times used in the final tests were: 17 ms attack and 200 ms release for the low channel; 13 ms attack and 160 ms release for the middle channel; and 6 ms attack and 75 ms release for the high channel. The effect of these long release times on closely spaced amplitude differences within a syllable are examined in the Discussion section.

**Test materials and presentation**—Short sentences or phrases composed of words with a high proportion of high-frequency consonants were used as speech tests. Typical examples are: “Lock the suitcase,” “Stay in the sun,” and “Noise in the street.” Twelve of these lists are shown in the Appendix. The word-to-word dynamic range of the peak vu deflections of scored items in a typical list of 10 sentences was of the order of 7 dB. Articles were not scored.

The IEEE (7) modified Harvard sentences (e.g., “The birch canoe slid on the smooth planks”), which were recorded with about the same dynamic range as the other lists, were used for two subjects whose speech recognition was judged good enough for more difficult test material. Subject AB was tested with both types of sentences, but subject JS was available only long enough to take the IEEE tests at preferred level.

Each IEEE sentence was scored on the basis of five key words. Partial credit was given in scoring either type of sentence by calculating the percent of phonemes correct in a word. All responses were written. Subjects were encouraged to guess if they weren’t sure of the answer, and to write down any parts of words that they heard. All tests were presented monaurally. Linear and compressed presentations of lists were alternated.

Lists were presented at preferred level, at 10 dB below preferred level (reduced at the input to the processor), at 15 dB below preferred input level, and with speech-spectrum noise added at a signal-to-noise ratio before processing of either 10 dB or 15 dB. The noise tests were given to three subjects, two of whom found the noise at −10 dB in the compressed mode too unpleasant. Finally, lists were presented to four subjects live, read by the same male talker as in the recording, while the subject wore his or her hearing aid. Subject JT did not use a hearing aid.

It was initially planned to achieve list equivalence by presenting each list in both linear and compressed modes, alternating the order of linear and compressed presentations. However, due to time constraints, not all subjects were able to complete the entire experiment.

<table>
<thead>
<tr>
<th>Subject AB</th>
<th>Compression Ratio</th>
<th>Post-compression Equalization</th>
<th>Linear-Mode Equalization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Mid</td>
<td>High</td>
</tr>
<tr>
<td>Calculated</td>
<td>(450)</td>
<td>(1470)</td>
<td>2.5</td>
</tr>
<tr>
<td>Chosen</td>
<td>2.0</td>
<td>2.5</td>
<td>3.0</td>
</tr>
<tr>
<td>Subject NJ</td>
<td>(450)</td>
<td>(1470)</td>
<td>2.9</td>
</tr>
<tr>
<td>Calculated</td>
<td>1.8</td>
<td>2.6</td>
<td>5.0</td>
</tr>
<tr>
<td>Chosen</td>
<td>1.9</td>
<td>3.1</td>
<td>4.1</td>
</tr>
<tr>
<td>Subject JS</td>
<td>(550)</td>
<td>(770)</td>
<td>1.9</td>
</tr>
<tr>
<td>Calculated</td>
<td>1.8</td>
<td>1.9</td>
<td>3</td>
</tr>
<tr>
<td>Chosen</td>
<td>2.2</td>
<td>3.1</td>
<td>2.8</td>
</tr>
<tr>
<td>Subject BG</td>
<td>(450)</td>
<td>(1470)</td>
<td>1.6</td>
</tr>
<tr>
<td>Calculated</td>
<td>2.2</td>
<td>3.1</td>
<td>2.8</td>
</tr>
<tr>
<td>Chosen</td>
<td>1.6</td>
<td>2.2</td>
<td>2.1</td>
</tr>
<tr>
<td>Subject JT</td>
<td>(450)</td>
<td>(1000)</td>
<td>1.9</td>
</tr>
<tr>
<td>Calculated</td>
<td>2.0</td>
<td>2.4</td>
<td>3.4</td>
</tr>
</tbody>
</table>
pressed presentation and waiting several weeks before presenting a list a second time. It became apparent, however, that this procedure was capable of giving a significant advantage to the mode that provided lower scores, since the subject had more correct answers to remember when the first presentation was in the higher-score mode. In a series of tests with 100 sentences (280 words) subject NJ scored 23.4 percent in the linear mode and 32.6 percent in the compression mode on lists in which the linear mode was presented first, while she scored 31.1 percent in the linear mode and 36.1 percent in the compression mode when the compression mode was presented first. The initial plan of presentation was therefore abandoned, and different lists were used for each mode, with the assignment of odd- and even-numbered lists to linear or compressed presentation alternated among subjects. In the final tests 130 IEEE sentences (650 words) were used for subject JS who left the experiment early; 320 of the shorter sentences (about 925 words), divided among the different conditions, were used for the latecomer BG; and almost double that number of words were used for each of the other three subjects.

Results
A comparison between recognition scores for the linear and compression modes appears in Table 3. Except for subject JS, scores are given for three input levels. Scores are also given for tests with the subject's own monaural aid, and at preferred level with speech-spectrum noise.

At preferred levels, three of the five subjects showed a statistically significant benefit from compression. The differences between linear scores and compression scores at preferred levels were significant at the 99 percent confidence level (single-tailed Student's t test, treated as single comparisons) for subjects JS and NJ, and at just under the 95 percent confidence level for subject AB. The averages of scores for these three subjects were 29.2 percent for uncompressed speech and 39.4 percent for compressed speech.

The differences between linear scores and compression scores at preferred level for the other two subjects were not significant at the 90 percent confidence level. The averages of preferred-level scores for these two subjects were 33.9 percent for the linear mode and 36.1 percent for the compression mode.

All four subjects to whom the reduced-level tests were given showed a statistically significant benefit from compression at these input levels, at the 95 percent confidence level or better. The averages of linear-presentation scores for these four subjects were 15.2 percent at 10 dB below preferred level and 10.7 percent at 15 dB below preferred level, compared to 31.7 percent and 32.5 percent for compressed presentations at -10 dB and -15 dB.

In the noise tests the scores of two subjects were higher in the compression mode than in the linear

Table 3.
Individual scores for the different conditions of presentation of test material. Standard error of the mean (σ/√N) appears next to scores, in parentheses. Subject JT did not use a hearing aid.

<table>
<thead>
<tr>
<th>Subject AB</th>
<th>Uncompressed</th>
<th>Compressed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-15 dB</td>
<td>-10 dB</td>
</tr>
<tr>
<td>IEEE</td>
<td>22.6 (1.5)</td>
<td>38 (3.7)</td>
</tr>
<tr>
<td></td>
<td>74 (1.4)</td>
<td>69.8 (6.3)</td>
</tr>
<tr>
<td></td>
<td>11 (4.4)</td>
<td>10.3 (3.7)</td>
</tr>
<tr>
<td></td>
<td>34.1 (6.0)</td>
<td>30.7 (8.6)</td>
</tr>
<tr>
<td>Subject NJ</td>
<td>Uncompressed</td>
<td>Compressed</td>
</tr>
<tr>
<td></td>
<td>9.3 (1.0)</td>
<td>11.5 (1.6)</td>
</tr>
<tr>
<td></td>
<td>20.9 (3.3)</td>
<td>22.8 (2.8)</td>
</tr>
<tr>
<td>Subject JS</td>
<td>Uncompressed</td>
<td>Compressed</td>
</tr>
<tr>
<td>(IEEE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject BG</td>
<td>Uncompressed</td>
<td>Compressed</td>
</tr>
<tr>
<td></td>
<td>7.1 (5.7)</td>
<td>9.2 (4.7)</td>
</tr>
<tr>
<td></td>
<td>32.4 (3.1)</td>
<td>31.8 (1.2)</td>
</tr>
<tr>
<td>Subject JT</td>
<td>Uncompressed</td>
<td>Compressed</td>
</tr>
<tr>
<td></td>
<td>9.5 (2.5)</td>
<td>16.1 (1.9)</td>
</tr>
<tr>
<td></td>
<td>22.8 (4.1)</td>
<td>21.8 (1.8)</td>
</tr>
</tbody>
</table>
mode, but only one of these compression scores showed an advantage that was significant at even the 80 percent confidence level. The noise-test score of a third subject was reduced significantly in the compression mode.

The live-voice hearing aid scores were all lower than scores for linear tests presented through the processor. The subject who scored zero with his hearing aid was having trouble with his earmold, but the trouble wasn’t bad enough to prevent him from making daily use of his aid.

Discussion

While compression processing produced statistically significant improvements in recognition scores, the improved scores were still so low that it is necessary to consider whether the improvement would be significant to the subject’s communication abilities. Experiment II sheds some light on the question by combining auditory and visual presentation, but the only authoritative answer to the question can come from real-life field tests.

Two types of perceptive distortion that would not have responded to compression/equalization processing—and that may have contributed to the low scores—are the loss of frequency selectivity (the ability to discriminate among simultaneous speech elements of different frequency) and the loss of temporal resolution (the ability to distinguish temporal patterns without blurring). Such distortions are likely to be worst in the profoundly deaf. Restoring acoustic speech cues to the residual dynamic range of hearing of the present subjects through amplitude- and frequency-dependent gain may have been a necessary but insufficient condition for good speech recognition.

Three-channel rather than two-channel compression was used because experiments with an electronic model of profoundly impaired hearing as described by Villchur (18) had suggested the superiority of three-channel over two-channel compression. There is no evidence that two-channel compression is not adequate for less than profound impairment, and only indirect evidence provided by the model that it is not adequate for the profoundly deaf.

Relatively long release times were preferred by the subjects, and although no formal study was made, the longer releases appeared to produce better results. The increased delay in releasing the compressor from its low-gain state following a high-intensity signal prevents silent intervals between speech elements from being filled in as quickly. In a multichannel system this delay will interfere with compressor action only when a sudden drop in signal amplitude occurs within the same channel. It should be noted that an effect on silent intervals similar to that produced by longer release times can be produced by the use of higher thresholds of compression (such as those used in the 1973 Villchur compression experiment, which were about 10 dB higher than in this experiment) and/or by the use of an expansion mode below compression threshold. In either case, noise and reverberant speech elements fall away more rapidly with decreasing input level.

Most of the subjects of the 1973 Villchur experiment showed a significant benefit from compression in the presence of two-voice speech interference. The steady-state noise used as interference in the present experiment had a higher level in the compression mode than much of the varying-level test speech, something that would not have been true with speech interference. Speech-spectrum noise imitates only the frequency distribution of speech, not its amplitude and time distribution. It is much less representative of real-life interference than are voices, and after reconsideration I am of the opinion that voices, such as were used in the 1973 Villchur experiment, would have provided a more valid interference element than steady-state noise.

The hearing aids used by subjects AB and NJ, which performed about as well as the processor in its linear mode, had frequency equalization similar to the linear equalization used for these subjects in the processor.

EXPERIMENT 2: Auditory presentation combined with visual cues

Subjects and equipment

Experiment 2 was exploratory in nature and was conducted with only one of the subjects, NJ. She was chosen because of the marked disparity between her low scores on the test material and the ease of communicating with her. The equipment was the same as that used in Experiment 1, except that an electret microphone (Realistic 33-1050) was substituted for the cassette player, and an insert earphone (Etymotic Research ER-3) was substituted for the
circumaural earphone. The microphone was in the same small room as the subject, and the large amount of acoustic gain combined with compression and high-frequency emphasis created unacceptable acoustic feedback from the circumaural earphone in spite of its noise-exclusion design. The ER-3 earphone was coupled to the subject's own earmold, and provided the necessary acoustic isolation.

Procedure
Test sentences were spoken into the microphone with the talker's face in clear view of the subject. It became apparent that the short-sentence lists used for subject NJ in Experiment I were not difficult enough to provide adequate resolving power between linear and compression modes, and so the more difficult IEEE sentences were used. Four 50-word lists were presented in each mode, with the mode alternated after each list. The processor level control was set to equal, at close microphone distance, the subject's previously chosen preferred level, but the microphone distance was changed for two of the lists. The dynamic range of the test was thereby increased and the reverberation element varied. One additional list was presented with the sound turned off. No list was used more than once.

Results
The recognition scores for auditory-plus-visual presentation of the IEEE lists in each mode, and at each of the different microphone distances, appears in Table 4. The average recognition score in the linear mode was 69.8 percent, compared to an average score of 80.4 percent for the compression mode. Since these averages include scores at different microphone distances, the variation of scores within each mode is too great for a meaningful calculation of the statistical significance of the difference between averages. The score for visual-only presentation was 39.2 percent.

Discussion
The increase in lipreading scores by the addition of auditory signals in either the linear or the compression mode was greater than might be predicted from the subject's low scores for auditory presentation alone of much easier test material. This result is consistent with the findings of Breeuwer and Plomp (3), who reported that when the visual-only presentation of speech was supplemented by certain frequency-selective sound-pressure information there was a large increase of intelligibility.

The advantage for compression carries over to the higher scores of the combined auditory/visual presentation, but no firm conclusions can be drawn until Experiment 2 is repeated with more subjects and, even more important, is carried out with wearable hearing aids in a real-life environment.

ACKNOWLEDGMENTS
I am indebted to Barbara Kruger, Director of Audiology and Speech-Language Pathology at the Albert Einstein College of Medicine, for arranging the supply of subjects and of working space at the College and for her invaluable cooperation in the project; to Neal A. Sloane, Supervisor of Audiology at the Hospital of the Albert Einstein College of Medicine, for selecting subjects and doing their audiological workups; to Mead C. Killion, President of Etymotic Research, Inc., for his major contributions in equipment design and in discussing the project; and finally to the five subjects of this experiment who showed such unfailing patience and good will.

Table 4.
Scores for subject NJ in the live visual/auditory tests. Each score is for a different list of 10 IEEE sentences (50 scored words).

<table>
<thead>
<tr>
<th>Microphone distance to talker’s lips</th>
<th>3 in</th>
<th>6 in</th>
<th>12 in</th>
<th>3 in, slower speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncompressed</td>
<td>74.8</td>
<td>69</td>
<td>52</td>
<td>83.5</td>
</tr>
<tr>
<td>Compressed</td>
<td>88.3</td>
<td>80.1</td>
<td>61.3</td>
<td>92</td>
</tr>
<tr>
<td>Visual only</td>
<td></td>
<td></td>
<td>39.2</td>
<td></td>
</tr>
</tbody>
</table>

FOOTNOTES
O'Loughlin suggested that the improvement was the result of the limiting action of the compressors, but he did not check his hypothesis by retesting at the same input level with compression thresholds at high speech levels. In such a retest the limiting action of his compressors—providing the subject with protection against high-intensity sounds—would have been maintained while compressor action that increased the relative gain of weak speech sounds would have been eliminated.
High-frequency emphasis reduced the dynamic range of speech usable by Skinner's subjects, and the higher the presentation level the less the amount of high-frequency emphasis that was most effective or that the subject would tolerate. Skinner suggested that these results indicated the desirability of using level-dependent frequency response: Such response implies the variation of high-frequency emphasis that was reduced by multichannel compressors whose compression ratios vary with frequency.

Of the studies showing no advantage to compression, Barfod (2) and Nabelek (12) used single-syllable test items at one level, while O'Loughlin (13) varied the level of his single-syllable test items only after compression (which does not engage the effect of the compressor on level changes).

The half-slope formula may change somewhat at the high threshold and speech levels of this experiment, but the subjects were able to compensate for any such change when they readjusted the compression ratios.

Dunn and White (5) measured male conversational speech at 30 cm as having an all-pass rms level, exceeded 10 percent of the time, of 80 dB SPL. (Use of the 10 percent level is consistent with the vu speech measurements of this experiment, in which occasional peaks are ignored.) This 80 dB level needs to be corrected for the present experiment by two factors, but one counterbalances the other. Most conversational speech occurs at a distance of 1 m or more rather than 30 cm, which in a typical indoor environment would reduce the level reported by Dunn and White by 6 to 8 dB. On the other hand the speech tests in this experiment were presented monaurally; as pointed out by Scharf (14), monaural listening at speech levels requires about 8 dB more amplitude than binaural listening for the same loudness.

There were some exceptions to this system—for example, when a score for the first presentation was very low—but the exceptions were distributed randomly. The scores showed no evidence of a systematic difference in difficulty between odd and even lists; one of the two subjects whose preferred-level scores were not improved significantly by compression had been assigned odd lists and the other had been assigned even.

In calculating average scores, subject AB was assigned a single score for each presentation mode, equal to the average of scores for the two types of sentences.

REFERENCES


APPENDIX: Examples of the test lists

Stop the bus
Is she smiling?
Take the subway
It's going to rain
Do you like the movies?
It's time for lunch
A pair of shoes
Get a haircut
He just had breakfast
Put on the dress

Dogs and kittens
Yellow leaves
Is it too loud?
Don't stand there
Rinse it in water
Are you thirsty?
The blue ocean
The train is fast
Light up the screen
He moves too slowly

Did you speak?
The stock market
They are silly
Wear a scarf
Ten spoons
Trucks are noisy
Can't you stay?
What does it cost?
A tall building
Read these letters

The sixth avenue bus
The cream is sour
Lost in the forest
She hasn't even started
The coffee spilled
Please have some breakfast
Listen to the noise
A dark dress
Steak and potatoes
Sit here quietly

The stars are bright
Did you go yesterday?
A baseball bat
He teaches in High School
There are six trees
A box of candy
Open the suitcase
The wind is cold
Turn around slowly
Sit at the desk

Dress quickly
What is the answer?
Listen to the wind
Serve the tea
My jacket is wet
Don't sleep late
A simple task
Winter snow
A sandy beach
Does she wear ties?

Don't run so fast
State your case
The boy has a dirty face
What city are you from?
The second act
Railroad tracks
She is sad
Look at the horse!
He drives too slowly
A pile of stones

Which is best?
Rainbow colors
A dozen sheets
She lost the bracelet
A graceful dance
Laughter and shouting
The sting of a bee
The clash of steel
Cats like fish
His car stalled

The sun is hot
Do you hear the music?
The show has started
Stay in the sun
Buy a ticket
A smooth surface
Traffic is heavy
The road is rough
Hang up your suit
Start the car

The dogs are barking
Thunder and lightning
Fudge is sweet
Write it down
A salt pickle
Seven large books
It costs too much
The river is quiet
This window doesn't open
Noise in the street

The worst one
A railway station
Throw the switch
Walk on the path
Spread the butter
He uses a cane
A nice piece of cake
How pretty she is!
An ice cream soda
Stay on the sidewalk

A fruit stand
Strip off the paint
She loves parties
Brass polish
An old castle
It won't last
Pass the sugar
The mast of a ship
Have some supper
The first day