Department of Veterans Affairs

Journal of Rehabilitation Research
and Development Vol. 30 No. 1, 1993
Pages 49–72

Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality, and response times

Thomas Baer, PhD; Brian C.J. Moore, PhD; Stuart Gatehouse, PhD
Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, England; MRC Institute of Hearing Research, Scottish Section, Glasgow Royal Infirmary, Glasgow G31 2ER, Scotland

Abstract—This paper describes a series of experiments evaluating the effects of digital processing of speech in noise so as to enhance spectral contrast, using subjects with cochlear hearing loss. The enhancement was carried out on a frequency scale related to the equivalent rectangular bandwidths (ERBs) of auditory filters in normally hearing subjects. The aim was to enhance major spectral prominences without enhancing fine-grain spectral features that would not be resolved by a normal ear. In experiment 1, the amount of enhancement and the bandwidth (in ERBs) of the enhancement processing were systematically varied. Large amounts of enhancement produced decreases in the intelligibility of speech in noise. Performance for moderate degrees of enhancement was generally similar to that for the control conditions, possibly because subjects did not have sufficient experience with the processed speech. In experiment 2, subjects judged the relative quality and intelligibility of speech in noise processed using a subset of the conditions of experiment 1. Generally, processing with a moderate degree of enhancement was preferred over the control condition, for both quality and intelligibility. Subjects varied in their preferences for high degrees of enhancement. Experiment 3 used a modified processing algorithm, with a moderate degree of spectral enhancement, and examined the effects of combining the enhancement with dynamic range compression. The intelligibility of speech in noise improved with practice, and, after a small amount of practice, scores for the condition combining enhancement with a moderate degree of compression were found to be significantly higher than for the control condition. Experiment 4 used a subset of conditions from experiment 3, but performance was assessed using a sentence verification test that measured both intelligibility and response times. Scores on both measures were improved by spectral enhancement, and improved still more by enhancement combined with compression. The effects were statistically more robust for the response times. When expressed as equivalent changes in speech-to-noise ratio, the improvements were about twice as large for the response times as for the intelligibility scores. The overall effect of spectral enhancement combined with compression was equivalent to an improvement of speech-to-noise ratio by 4.2 dB.

Key words: *compression, hearing impairment, response times, spectral enhancement, speech intelligibility.*

INTRODUCTION

People with moderate sensorineural hearing impairment often complain of difficulty in understanding speech in noise. They can understand speech reasonably well in one-to-one conversation in a quiet room, but they have great difficulty when there is background noise or reverberation, or when more than one person is talking. This difficulty appears to be related to a variety of abnormalities in the perception of sound (1) and it persists even when

Address all correspondence and requests for reprints to: Dr. T. Baer, Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England.

the speech is amplified sufficiently (by a hearing aid) to be well above the threshold for detection (1,2).

Reduced frequency selectivity is a well-documented abnormality that is associated with sensorineural hearing loss and which can affect speech perception in noise. Frequency selectivity refers to the ability of the ear to resolve a complex sound into its frequency components. This ability is often characterized by describing the ear as containing a bank of overlapping bandpass filters, known as the auditory filters (3). The characteristics of these filters for normally hearing people have been reasonably well established (4,5,6,7). Sensorineural hearing loss, and particularly cochlear hearing loss, is associated with broader-than-normal auditory filters, that is, reduced frequency selectivity (8,9). Several studies have shown that the ability to understand speech in noise is correlated with measures of auditory filter bandwidth, although the effects of filter bandwidth are difficult to separate from the effects of a simple loss of sensitivity to weak sounds, since the two are highly correlated (10,11,12). It seems likely that impaired frequency selectivity is at least partly responsible for reduced ability to hear speech in noise, although this causal link has not been universally accepted (13).

One mechanism by which impaired frequency selectivity could affect speech perception in noise involves the perception of spectral shape. The recognition of speech sounds requires a determination of their spectral shapes, especially the locations of spectral prominences (usually formants). One representation of spectral shape in the auditory system is called the excitation pattern. The excitation pattern of a given sound may be defined as the magnitude of the outputs of the auditory filters in response to that sound as a function of filter center frequency (4,6). The excitation pattern resembles a smoothed version of the spectrum. Broader auditory filters produce a more highly smoothed representation of the spectrum. If spectral features are not sufficiently prominent, they may be smoothed to such an extent that they become imperceptible. In one study where degree of spectral contrast was varied, the contrast (decibel [dB] difference between peaks and valleys in the spectrum) required for vowels to be identified was shown to be greater for impaired than for normal listeners (14). Adding a noise background to speech fills in the valleys between the spectral peaks and thus reduces their prominence, exacerbating the problem of perceiving them for people with broadened auditory filters.

A second possible effect of reduced frequency selectivity on speech perception in noise is connected with the temporal patterns at the outputs of individual auditory filters. The perceived frequency of a given formant and/or the fundamental frequency of voicing may be partly determined by the time pattern at the outputs of the auditory filters tuned close to the formant frequency (15,16). Background noise disturbs this time pattern, which may lead to reduced accuracy in determining these frequencies. This effect would be greater in a person with reduced frequency selectivity, since broader filters generally pass more background noise.

If reduced frequency selectivity impairs speech perception, then enhancement of spectral contrasts might improve it for the hearing-impaired person. Either of the two mechanisms outlined above, one based on degradation of spectral shape and the other on degradation of temporal patterns, provides a rationale for performing spectral enhancement. If spectral features are smoothed by an impaired auditory system, then preprocessing the signal to enhance spectral contrasts can produce an excitation pattern that more nearly resembles the excitation pattern evoked by an unprocessed signal in a normal auditory system. The impaired auditory system can be thought of as convolving the spectrum with a smoothing function, and spectral contrast enhancement can be thought of as a partial deconvolution process. If temporal patterns are disturbed by the noise passing through a broadened auditory filter, then enhancing those portions of the spectrum where the signal-to-noise ratio is highest (the peaks) and suppressing those portions where it is lowest (the valleys) should minimize this effect.

Several authors have described attempts to improve speech intelligibility for the hearing impaired by enhancement of spectral features. Boers (17) processed a set of sentences so as to increase the level differences between peaks and valleys in the spectrum. Noise was added *after* the processing, and the effects of the processing were assessed by measuring the speech-to-noise ratio required for 50 percent of the words to be understood. Overall, the processing reduced intelligibility, although two impaired listeners did show a slight improvement with the processed signals. Even if it had systematically improved intelligibility, this kind of processing

would not be feasible with naturally occurring signals; with these the speech would already be contaminated with noise, and the processing would have to operate on the speech-plus-noise.

Summerfield, et al. (18), synthesized "whispered" speech sounds, and investigated the effect of narrowing the bandwidths of the formants (spectral resonances) used in synthesis. Narrowing these bandwidths led to both sharper spectral peaks and greater peak-to-valley ratios. However, it had only small effects on speech intelligibility; identification of consonants at the end of syllables tended to be slightly better for both normal and impaired listeners when the formant bandwidths were half their nominal normal values. Speech intelligibility in noise was not tested.

Simpson, et al. (19), described a method of digital signal processing of speech in noise so as to increase differences in level between peaks and valleys in the spectrum. Before spectral enhancement, the spectra were smoothed to eliminate minor peaks and ripples, using smoothing filters based on the properties of the auditory filters in normal ears. The enhancement was also done on a frequency scale related to the frequency resolution of normal ears (4). The enhancement procedure involved convolving the spectrum with a Difference-of-Gaussians (DoG) filter. This operation is similar to taking a smoothed second derivative of the spectrum. The spectral pattern obtained in this way was used to construct a gain function to enhance the original spectrum. The intelligibility of the speech in speech-shaped noise was measured using subjects with moderate sensorineural hearing loss. The results showed small but reasonably consistent improvements in speech intelligibility for the processed speech. The processing used by Simpson, et al. ran at about 200 times real time on a reasonably fast laboratory computer (Masscomp 5400 with floating-point accelerator).

Stone and Moore (20) described a speech-processing system similar to that used by Simpson, et al., but one that was simpler, and based on analog electronics running in real time, using a 16-channel band-pass filter bank. Each channel generated an "activity function" that was proportional to the magnitude of the signal envelope in that channel, averaged over a short period of time. A positively weighted activity function from the nth channel was combined with negatively weighted functions from channels $n - 2$, $n - 1$, $n + 1$, and $n + 2$, giving a correction signal used to control the gain of the band-pass signal in the nth channel. Recombining the band-pass signals resulted in a signal with enhanced spectral contrast. Two different experiments were described, the first using the activity function as described, and the second using a nonlinear transform of the activity function. In both experiments, several different weighting patterns were used in calculating the correction signal. The intelligibility of speech in speech-shaped noise processed by the system was measured for subjects with moderate sensorineural hearing loss. In both experiments, no improvement in intelligibility was found. However, subjective ratings of the stimuli used in the second experiment indicated that some subjects judged the processed stimuli to have both higher quality and higher intelligibility than unprocessed stimuli.

Bunnell (21) described a method of digital signal processing to enhance spectral contrasts. Contrasts were enhanced mainly at middle frequencies, leaving high and low frequencies relatively unaffected. Unlike the processing used by Simpson, et al. (19), and by Stone and Moore (20), the enhancement was performed on a spectral envelope that was calculated with a linear frequency scale (using a cepstral smoothing technique) rather than a scale reflecting auditory frequency selectivity. Small improvements were found in the identification of stop consonants presented in quiet to subjects with sloping hearing losses. No measurements of the intelligibility of speech in noise were reported.

Several other authors have described methods of processing speech in noise aimed mainly at enhancing speech quality and/or intelligibility for normal listeners or as preprocessors for speech recognition devices. Lim (22) reviews work done prior to 1983. Many of the techniques that have been developed result in improvements of signal-to-noise ratio (SNR) without any improvement in intelligibility, and many have been plagued by artifacts such as the introduction of spurious sounds as a result of enhancing random spectral peaks. Cheng and O'Shaughnessy (23) described a method similar to that used by Simpson, et al. (19), but differing in several details. They reported an improvement in subjective quality for speech in white noise, based on informal tests with normal listeners. They used two alternative algorithms—one for

Journal of Rehabilitation Research and Development Vol. 30 No. 1 1993

low-noise conditions where the improvement in SNR was modest but speech quality (naturalness) was retained or enhanced, and the other for high-noise conditions, where there was a large improvement in SNR but speech quality was degraded. No formal measurements of speech intelligibility were made.

Clarkson and Bahgat (24) filtered signals into several contiguous frequency bands and expanded the envelope in each band, so as to enhance spectral contrast. A measure of spectral variance was used to control the amount of expansion. Listening trials with a simplified real time system showed small, but reasonably consistent, improvements at 0-dB speech-to-noise ratio in a modified rhyme test.

In this paper, we describe a series of experiments aimed at further developing the technique of Simpson, et al. (19). Experiment 1 was a parametric study using processing similar to that described by Simpson, et al. The objective was to find optimum values of two of the parameters used in the processing. The intelligibility of speech in speech-shaped noise was measured for several different conditions involving spectral enhancement. Experiment 2 was carried out using a subset of the conditions from experiment 1, to determine whether the spectral enhancement produced improvements in subjective judgments of speech quality and intelligibility. Experiment 3 investigated the effect of combining spectral enhancement with amplitude compression, with a modified enhancement algorithm, again using measures of the intelligibility of speech in speech-shaped noise. Finally, experiment 4 used a subset of the conditions from experiment 3, but performance was evaluated in a test measuring both speech intelligibility and response time. Although the experiments were primarily concerned with the intelligibility and quality of speech in noise, informal listening tests were carried out using speech in quiet. In all cases, the quality of the processed speech was judged to be good, by both normal and hearing-impaired listeners.

EXPERIMENT 1

Method of Speech Enhancement

The technique used for spectral enhancement was similar to that described by Simpson, et al. (19), and involved manipulation of the short-term spectrum of the speech in noise. Sampled segments of the signal were windowed, smoothed, spectrally enhanced, and then resynthesized using the overlap-add technique (25). Each step is described below. The steps are also illustrated in **Figure 1**.

The speech in noise was low-pass filtered at 4 kHz (Fem EF16, 100 dB/oct slope) and sampled at a 10-kHz rate with 12-bit resolution using a Masscomp 5400 computer with EF12M analog-to-digital converter. A 12.8-ms segment of the signal was weighted with a 12.8-ms Hamming window; the segment was padded with 64 zeros at the start and 64 zeros at the end. A 256-point fast Fourier transform (FFT) of the windowed segment was calculated, giving 128 magnitude values and 128 phase values. The phase values were stored and subsequent operations were carried out only on the magnitude spectrum.

To avoid enhancing spectral details that would be undetectable even for a normal ear, the magnitude spectrum was transformed to an auditory excitation pattern, using the convolution procedure described by Moore and Glasberg (4). This involved calculating the output of an array of simulated auditory filters in response to the magnitude spectrum. Each side of each auditory filter is modeled as an intensity-weighting function, assumed to have the form of the rounded-exponential filter described by Patterson, et al. (26):

$$W(g) = (1 + pg)\exp(-pg), \qquad [1]$$

where g is the normalized distance from the center of the filter (distance from center frequency divided by center frequency, $\Delta f_c/f_c$) and p is a parameter determining the slope of the filter skirts. The value of p was assumed to be the same for the two sides of the filter. The equivalent rectangular bandwidth (ERB) of this filter is $4f_c/p$.

The ERBs of the auditory filters were assumed to increase with increasing center frequency, as described by Moore and Glasberg (4). As a result of this calculation, the original 128 magnitude values were replaced with 128 new values, representing a smoothed version of the original spectrum. The smoothing tended to remove minor irregularities in the spectrum, but to preserve peaks corresponding to major spectral prominences in the speech.

An enhancement function was derived from the excitation pattern by a process of convolution with a DoG function (on an ERB scale). This function is the sum of a positive Gaussian and a negative

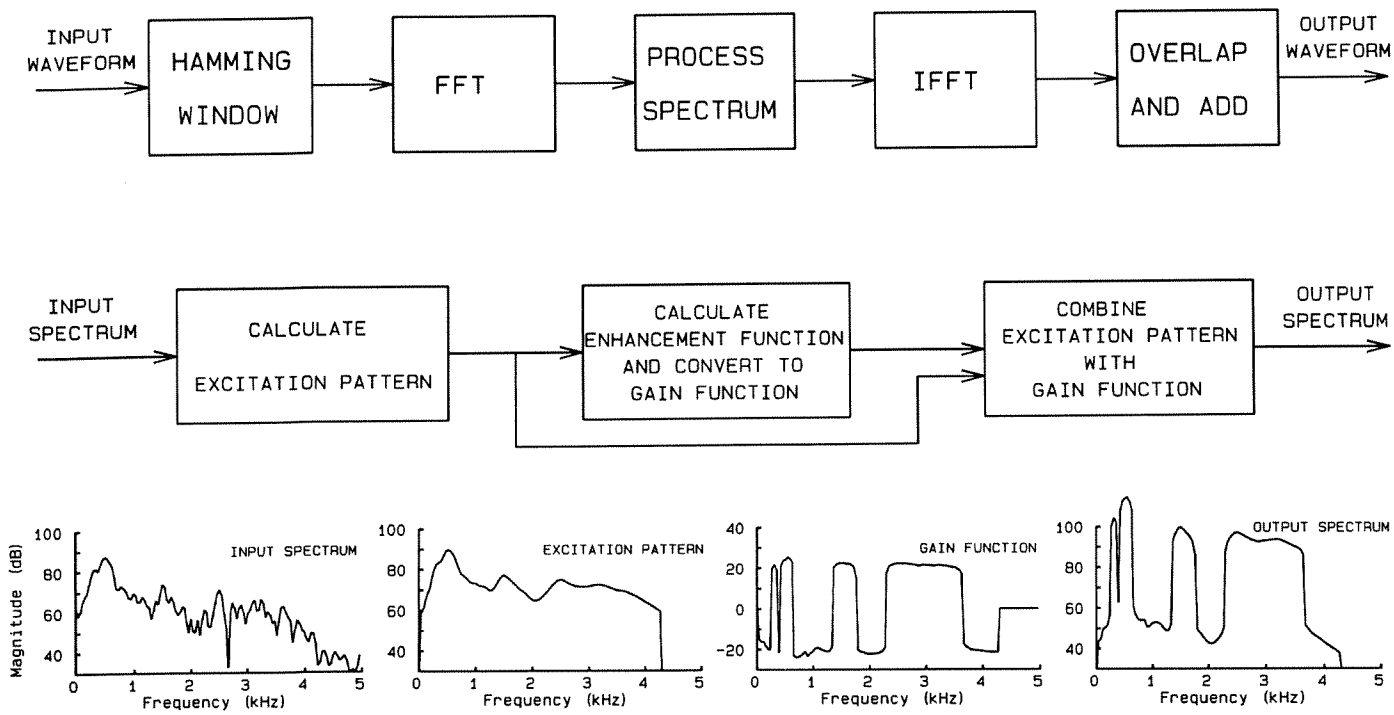

Figure 1.
Schematic diagram of the sequence of stages involved in the enhancement processing of Experiment 1. The top row shows all stages of the processing. The middle row shows the "process spectrum" stage in more detail. The bottom row shows an example of the spectral processing for a particular frame, for condition E3B2.

Gaussian that has twice the bandwidth of the positive Gaussian, as described by the following equation:

$$\mathrm{DoG}(\Delta f) = (1/2\pi)^{1/2}[\exp\{-(\Delta f/b)^2/2\} - (1/2)\exp\{-(\Delta f/2b)^2/2\}], \qquad [2]$$

where Δf is the deviation from the center frequency, and b is a parameter determining the bandwidth of the DoG function. Note that the total area of this function, summed over positive and negative parts, is zero. In these experiments three values of b were used, chosen so that the width of the positive lobe (between the zero-crossing points) was either 0.5, 1.0, or 2.0 times the ERB of the auditory filter with the same center frequency (4). Thus, the width of the DoG function increased with increasing center frequency. The three bandwidths used will be referred to as B.5, B1, and B2.

The DoG function was centered on the frequency of each of the 128 magnitude values of the excitation pattern in turn. For a given center frequency of the DoG function, the value of the excitation pattern at each frequency (in linear power units) was multiplied by the value of the DoG function at that same frequency, and the products obtained in this way were summed. The magnitude value of the excitation pattern at that center frequency was then replaced by that sum.

The enhancement function derived in this way was then used to modify the excitation pattern. At center frequencies where the enhancement function was positive, the excitation pattern was increased in magnitude; at center frequencies where the enhancement function was negative, the excitation pattern was decreased in magnitude. This was achieved in the following way. Let the absolute value of the enhancement function at a particular center frequency be denoted by abs(ENF) and the corresponding sign (positive or negative) of the enhancement function be denoted sign(ENF). The value of the enhancement function was converted to a decibel-like quantity by calculating

$$G = \log\{\mathrm{abs}(\mathrm{ENF}) + 1\} \times \mathrm{sign}(\mathrm{ENF}). \qquad [3]$$

The value of abs(ENF) was generally large (in the thousands), but 1 was added to it to avoid the possibility of taking the logarithm of zero. The value of G was then scaled by a certain factor, E,

and added to the magnitude of the excitation pattern at that center frequency—the excitation level being expressed in decibels. The degree of enhancement of the spectrum was determined by the size of the factor E; values used were 0.3, 0.6, and 0.9, corresponding to small, medium, and large amounts of enhancement. These degrees of enhancement will be referred to as E3, E6, and E9, respectively.

The magnitude values from the enhanced excitation pattern, expressed in linear amplitude units, were then combined with the original phase values, and an inverse FFT was used to produce a 25.6-ms segment of spectrally enhanced speech in noise. This process was repeated every 6.4 ms, and the resultant overlapping segments were summed to give a complete processed waveform.

In summary, the processing had the effect of enhancing spectral contrast in the magnitude spectrum while preserving the phase spectrum. The processing was performed with three degrees of enhancement (E3, E6, and E9) and three values for the width of the DoG function (B.5, B1, and B2), giving nine experimental conditions in total. The condition E3B1 is similar to that used by Simpson, et al. (19). In addition, two control conditions were used. In one, the speech in noise was processed through all stages except those involving enhancement. Thus, the spectrum was smoothed in the conversion to the excitation pattern, but was otherwise unaltered; this corresponds to processing with the value of E set to 0. We refer to this condition as E0. In the second control condition, referred to as NULL, the speech in noise was passed through all stages except the conversion to the excitation pattern and the enhancement. The conversion to an excitation pattern has the effect of putting a high frequency emphasis on the spectrum; this happens because the ERB of the auditory filter increases with center frequency. Since the NULL condition did not involve conversion to an excitation pattern, the high frequency emphasis was obtained in this condition by increasing the power spectrum at a given frequency by an amount proportional to the ERB of the auditory filter at that center frequency. The overall level of the speech-plus-noise was equalized for all conditions.

Figure 2 shows an example of the spectra of stimuli processed using conditions NULL (top panel), E0 (middle panel), and E3B2 (bottom panel). The signal was a synthesized neutral vowel presented in speech-shaped noise at a signal-to-noise ratio of 0 dB. The figure shows the long-term-average spectra of the processed stimuli, not the spectra of individual frames; the effects of the enhancement processing were generally more pronounced in the latter. Note how the spectral level between the formants, especially the second and third formants, is decreased by the processing.

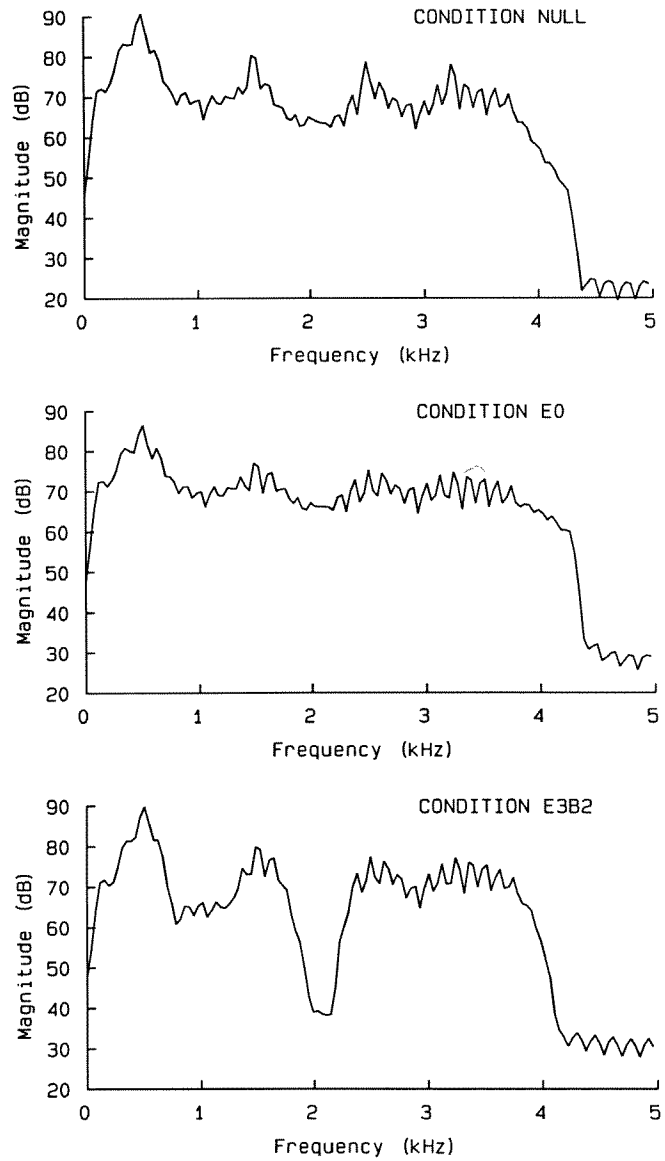

Figure 2.
Example of the effects of the processing used in Experiment 1, showing long-term-average spectra of a neutral vowel in noise, processed using the control condition (NULL–top panel), condition E0 (middle panel) and condition E3B2 (bottom panel).

Stimuli

The stimuli were the first 11 lists from the Adaptive Sentence Lists (ASL) (27) presented in a continuous background of noise with the same long-term-average spectrum as the sentences. Sentences were presented at 12-sec intervals, leaving ample time for subject responses. Most subjects were tested at a speech-to-noise ratio of 0 dB, both speech and noise levels being specified in terms of root-mean-square pressures. Subjects 4 and 6, who scored poorly at this speech-to-noise ratio, were tested using a ratio of +3 dB. The score was the number of key words identified (out of the 45 in each list). Stimuli were recorded on digital audio tape (DAT) and presented via a Quad amplifier and Monitor Audio MA4 loudspeaker.

Subjects

Eleven subjects were tested. All were diagnosed as having bilateral sensorineural hearing loss, probably of cochlear origin. Their audiograms and other relevant information are presented in **Table 1**. Most were experienced hearing aid users.

Experimental Design

A Latin Square design was used. All subjects were tested with the 11 ASL lists presented in the same, ascending order. Each subject was tested once in each of the 11 conditions, with the order of conditions counterbalanced across subjects. Thus, for each subject, a different list was used for each of the 11 conditions, and for each condition a different list was used for each of the 11 subjects.

Procedure

The subject sat in a sound-attenuating room facing the loudspeaker at a distance of 1.3 m. Seven of the subjects, those who normally wore hearing aids without any compression circuity or other "signal processing," listened using their own hearing aids. Initially, they were asked to adjust the volume controls on their aids to the setting that they would use for normal conversation. Then, they were presented with ASL list 12 (i.e., not one of the 11 test lists) processed using condition NULL and the level was varied until they indicated that it was at their preferred listening level. Subject 3, who normally wore hearing aids incorporating compression, and subjects 2, 5, and 8, who did not normally use their aids, were tested unaided; the level of the stimuli was adjusted to their preferred listening level. The adjustments were usually completed well before the list was completed. The remainder of list 12 was used as practice. In a few cases, list 13, also processed in condition NULL, was used for further practice.

Testing proper then started. Subjects were presented with 11 test lists, with a brief rest between each list. Subjects were told to repeat back as many words as they could, and to make a guess when they were not sure. They were told that the task would be quite difficult, and they were not expected to hear every word.

Results

The scores for each subject for each condition and the mean scores across subjects are shown in **Table 2**. The mean scores do not differ greatly across conditions, but tend to be lower for the conditions involving a high degree of enhancement (E9). To assess the significance of these effects, the data were subjected to an analysis of variance (ANOVA) with factor condition, with the data blocked across list number and subject (28). In this analysis, the proportions correct were transformed using the expression arcsine ($\sqrt{\text{proportion correct}}$). This transform makes the scores follow a normal distribution more closely. The effect of condition was significant: $F(10,90) = 5.06$, $p < 0.001$. The GENSTAT package used gave estimates of the standard errors of the differences between the mean scores for the different conditions. These standard errors were used to assess the significance of the differences between means (28, p. 81). The mean score for condition E9B.5 was significantly lower ($p < 0.01$) than the mean scores for all other conditions. The score for condition E9B1 was significantly lower than the scores for conditions NULL ($p < 0.01$), E0 ($p < 0.01$), E3B2 ($p < 0.02$), and E6B1 ($p < 0.05$). Scores for the other conditions did not differ significantly.

Overall, these results are disappointing. In contrast to the results of Simpson, et al. (19), the processing did not improve speech intelligibility relative to the control conditions; and a high degree of processing led to a significant worsening of intelligibility. The processing condition giving the highest scores was one involving a moderate degree of enhancement, E3B2. If scores for this condition are compared with the mean scores for the two

Journal of Rehabilitation Research and Development Vol. 30 No. 1 1993

Table 1.
Characteristics of the hearing impaired subjects used in the experiments.

				Frequency in kHz					
Subject	**Age**	**Sex**	**Ear**	**0.25**	**0.5**	**1.0**	**2.0**	**4.0**	**8.0**
1	74	M	L	35	30	35	40	55	55
			R	30	35	35	50	65	60
2	75	F	L	25	10	5	30	60	75
			R	15	10	5	35	65	80
3	67	M	L	60	55	55	60	65	75
			R	50	50	55	60	70	70
4	68	M	L	40	65	75	85	85	>100
			R	55	60	70	85	95	>100
5	72	M	L	30	20	35	55	95	80
			R	35	30	40	60	85	95
6	82	M	L	85	80	80	90	95	>100
			R	60	65	80	85	80	85
7	70	M	L	25	30	60	60	75	80
			R	10	10	60	55	65	80
8	78	M	L	40	50	55	60	80	>100
			R	25	25	40	45	70	80
9	69	F	L	35	35	45	45	65	85
			R	65	50	55	70	90	100
10	68	M	L	70	55	45	35	50	60
			R	30	35	45	40	40	50
11	73	F	L	45	50	50	50	55	55
			R	50	50	55	50	55	55
12	71	M	L	30	40	55	30	50	95
			R	30	40	45	30	30	60
13	62	M	L	45	50	65	60	40	70
			R	35	40	55	60	60	55
14	68	M	L	30	40	45	60	70	85
			R	20	35	45	50	55	75
15	72	F	L	35	45	50	55	60	70
			R	30	25	40	55	65	65
16	63	M	L	15	30	40	60	65	55
			R	20	25	40	70	70	60
17	64	M	L	25	15	20	40	65	70
			R	30	20	20	30	60	75
18	69	M	L	20	40	30	35	60	80
			R	15	30	40	40	55	75

Subjects 1-11 took part in Experiment 1. Subjects 1, 4, 7, 9, 10, and 11 took part in Experiment 2. Subjects 8-13 took part in Experiment 3. Subjects 14-18 took part in Experiment 4. Absolute thresholds are given in dB HL.

control conditions, NULL and E0, we find that seven subjects performed better with the processing and four performed more poorly.

There may be several reasons why Simpson, et al., found significant improvements in speech intelligibility with processed stimuli, whereas we did not. The first possibility is connected with the fact that Simpson, et al., used only one processing condition

Table 2.
Results of Experiment 1, showing the score for each subject in each condition (number of words correct out of 45) and the mean score for each condition.

						Condition					
Subject	Null	E0	E3B.5	E3B1	E3B2	E6B.5	E6B1	E6B2	E9B.5	E9B1	E9B2
1	41	41	40	40	33	37	43	34	30	38	33
2	38	42	38	40	42	37	42	44	30	36	38
3	36	30	35	32	36	30	31	31	16	22	38
4	34	36	31	23	34	33	34	28	19	28	31
5	37	42	41	39	31	30	37	37	19	38	31
6	34	34	33	29	35	33	27	23	22	26	24
7	36	32	36	38	35	28	34	38	35	36	41
8	45	40	34	40	43	35	41	37	23	26	24
9	30	36	31	40	32	38	35	29	31	37	36
10	28	34	29	28	32	32	35	35	36	34	33
11	40	34	40	35	44	36	33	39	35	33	38
Mean	36.3	36.5	35.3	34.9	36.1	33.6	35.6	34.1	26.9	32.2	33.4

and one control condition. They gave subjects two practice lists (one control and one enhanced) and then tested subjects using six sentence lists for each condition. This gave subjects a reasonably large amount of experience with the processed stimuli. In contrast, each subject in our experiment listened to each condition only once, using a single sentence list. It may be that subjects require a more extended practice period to get a benefit from the processing. It should also be noted that the use of six lists per condition greatly reduces the inherent variability in the data compared with our use of a single list.

A second possible factor only became apparent after the main part of the experiment was completed. We discovered that the enhancement process had an undesired side effect; it tended to produce a high-frequency deemphasis. The spectral level of frequencies above 600 Hz tended to be 3–6 dB lower in the experimental conditions than in the control conditions. This may have offset any potential improvements in intelligibility produced by the enhancement process.

EXPERIMENT 2

Ratings of Speech Quality and Intelligibility

There have been several reports in the past of processing that improves the subjective quality of speech in noise without improving intelligibility (22). Previous work involving judgments of speech quality has mainly used normally hearing subjects, although Stone and Moore (20) reported such measurements for hearing-impaired subjects. Processing that improves speech quality without changing intelligibility may be useful as a means of making listening more pleasant and less effortful. Hence, we decided to investigate whether our processing led to any improvements in subjective speech quality.

Two tests were performed where six hearing-impaired subjects (subjects 1, 4, 7, 9, 10, and 11 from experiment 1) made pair-wise subjective comparisons between sentences in noise processed using conditions E0, E3B1, E3B2, E6B2, and E9B2 of experiment 1, rating them for sound quality in one set of tests and intelligibility in the other set. In addition, we used a condition resembling E3B1, but with the unprocessed signal added back to the processed signal. This had the effect of slightly reducing the amount of enhancement, but also of somewhat reducing the audibility of undesired side-effects of the processing, specifically a slight "gurgling" quality. This condition resembles the processing used by Simpson, et al. (19), and will be denoted by E3B1 + U.

For all 15 possible pairs of processing conditions, 10 pairs of sentences were compared. On a

Journal of Rehabilitation Research and Development Vol. 30 No. 1 1993

given trial the same sentence was presented twice, the sentences differing only in the way they were processed. Five different sentences were used, taken from ASL list 1, chosen because performance in experiment 1 was especially poor for these sentences. They were edited so that each sentence was approximately centered in 3 sec of its masking noise. There was an interval of 500 ms between the noises for the two sentences in a pair. Following the end of the noise for the second sentence, there were 5.5 sec of silence during which the subject indicated which sentence had the higher quality or intelligibility. Five of the sentence pairs were presented as condition A followed by condition B, while the other five were presented as condition B followed by condition A. The order of presentation of the sentence pairs was randomized for both comparison of processing condition and order of presentation within each individual test. All editing was done digitally using a Masscomp 5400 computer system. Final stimuli were recorded on digital audio tape (Sony DTC 1000ES).

In the first test, subjects were asked to indicate which sentence in each pair had the higher sound quality in terms of pleasantness. In the second test, they were asked to indicate which sentence in each pair they felt was more intelligible. In addition to making a forced-choice decision, each subject was asked to make a confidence rating on each trial, by giving a number indicating how large the difference appeared to be.

For each processing pair, a distance metric was calculated by adding together the 10 (signed) confidence ratings. For each pair of conditions, AB, the sign was positive if B was selected and negative if A was selected. An analysis of the results showed that, for each subject, there was a high correlation between the number of selections and the distance metric; correlations ranged from 0.69 to 0.99, and were typically over 0.8. This indicates that the measures have a fairly high degree of reliability and internal consistency.

Results

The results are summarized in **Table 3** (quality judgments) and **Table 4** (intelligibility judgments). Each cell in each table shows the number of B selections (out of 10) with the distance score in parentheses. For example, in the quality judgments of subject 7, condition E3B1 was preferred over condition E0 nine times out of ten, with a distance metric of 5.0. It should be noted that the results showed a bias for the second sentence in a pair to be preferred over the first. This bias was controlled for by our procedure of balancing the order of processing conditions across pairs, but it probably had the effect of somewhat reducing the overall differences between pairs of conditions.

Two overall measures of the scores for each condition were also calculated. For the first, the preferences were summed across all five pairs of conditions involving a given condition. For example, the summed preference score for condition E0 was equal to the total number of times that condition was preferred over the other five conditions; the maximum value for this score is 50. For the second measure, the signed distance measures for a given condition were summed for all comparisons involving that condition. The signs were chosen so that a positive score would indicate an overall preference for that condition. These scores are also shown in **Table 3** and **Table 4**.

Consider first the quality judgments (**Table 3**). For subjects 7, 9, 10, and 11, conditions E3B1 and E3B1 + U were preferred over the control condition, E0. This is apparent both from the numbers of selections and from the distance scores. Preferences for the other processing conditions varied more across subjects. Subject 11 preferred enhanced speech over the control condition for all conditions involving enhancement. Her overall scores were lowest for the control condition, and highest for the conditions involving the greatest degree of enhancement (E6B2 and E9B2). For subjects 7 and 9, both the overall preference scores and the overall distance scores were highest for conditions E3B1 and E3B1 + U, and lowest for condition E9B2. For subjects 1 and 4, preferences were clearly lowest for the condition involving the greatest degree of enhancement (E9B2). For subject 10, preferences were less clear cut, but there was a consistent trend for the conditions involving enhancement to be preferred over the control condition. Overall, the results indicate that the quality of the enhanced speech in noise was generally preferred over that in

Table 3.
Results of Experiment 2 for the judgments of the quality of speech in noise.

Subject	A condition	B condition E3B1	ENH	E3B2	E6B2	E9B2	Global
1	E0	1 (−4)	5 (1)	3 (−2)	2 (−6)	0 (−11)	39 (22)
	E3B1		8 (1)	3 (−1)	1 (−5)	2 (−7)	27 (8)
	ENH			4 (−1)	0 (−5)	1 (−10)	38 (18)
	E3B2				3 (−4)	2 (−8)	25 (8)
	E6B2					0 (−8)	16 (−12)
	E9B2						5 (−44)
4	E0	4 (1)	7 (9)	6 (5)	3 (−7)	1 (−19)	29 (11)
	E3B1		3 (0)	5 (−2)	4 (2)	1 (−16)	31 (17)
	ENH			6 (2)	7 (5)	1 (−16)	26 (18)
	E3B2				5 (−4)	2 (−8)	30 (17)
	E6B2					3 (−5)	26 (50)
	E9B2						8 (−64)
7	E0	9 (5.0)	6 (2.5)	4 (−0.5)	2 (−4.0)	1 (−6.0)	28 (3)
	E3B1		5 (−0.5)	4 (−2.5)	2 (−3.5)	4 (−1.5)	34 (13)
	ENH			4 (−1.5)	4 (−2.5)	3 (−4.5)	30 (10.5)
	E3B2				2 (−5.)	3 (−3.0)	27 (3.5)
	E6B2					4 (−2.5)	15 (−12.5)
	E9B2						15 (−17.5)
9	E0	9 (17)	8 (13)	3 (−8)	3 (−8)	0 (−26)	27 (12)
	E3B1		4 (−2)	6 (4)	4 (−7)	0 (−28)	35 (50)
	ENH			2 (−11)	3 (−10)	0 (−25)	37 (57)
	E3B2				4 (−6)	1 (−18)	26 (9)
	E6B2					0 (−21)	24 (−10)
	E9B2						1 (−118)
10	E0	7 (4)	8 (4)	7 (3)	9 (7)	6 (3)	13 (−21)
	E3B1		4 (−1)	6 (2)	4 (2)	5 (−2)	28 (3)
	ENH			4 (−2)	5 (0)	5 (−1)	28 (6)
	E3B2				6 (−1)	3 (−1)	28(5)
	E6B2					5 (0)	29 (8)
	E9B2						24 (−1)
11	E0	9 (13)	10 (18)	10 (21)	10 (18)	9 (16)	2 (−86)
	E3B1		6 (2)	5 (−3)	8 (11)	8 (12)	22 (−9)
	ENH			5 (2)	7 (7)	10 (19)	24 (−8)
	E3B2				9 (11)	9 (14)	22 (−5)
	E6B2					6 (2)	38 (45)
	E9B2						42 (63)
Mean	E0	65	73	55	48	28	46
	E3B1		50	48	38	33	59
	ENH			42	43	33	61
	E3B2				48	33	53
	E6B2					30	50
	E9B2						32

Preference scores (ranging from 0 to 10) indicate the number of times the B condition was preferred over the A condition. Distance scores (in parentheses) are positive if the B condition was preferred and negative if the A condition was preferred. The global scores indicate the total number of times (out of 50) that a given condition was preferred, and the total distance (in parentheses). The more positive the distance, the more that condition was preferred overall. For the means, only preferences are shown (as percentages).

Journal of Rehabilitation Research and Development Vol. 30 No. 1 1993

Table 4.
Results of Experiment 2 for the judgments of intelligibility.

Subject	A condition	B condition E3B1	ENH	E3B2	E6B2	E9B2	Global
1	E0	6 (1)	4 (0)	7 (2)	5 (0)	4 (0)	24 (−3)
	E3B1		3 (−3)	6 (0)	3 (−1)	5 (1)	29 (4)
	ENH			7 (0)	2 (0)	6 (1)	22 (−4)
	E3B2				4 (−1)	5 (0)	31 (3)
	E6B2					7 (2)	17 (−4)
	E9B2						27 (4)
4	E0	9 (8)	6 (0)	8 (11)	8 (11)	9 (16)	10 (−46)
	E3B1		5 (−2)	5 (0)	5 (7)	4 (−3)	30 (6)
	ENH			4 (0)	5 (5)	5 (1)	27 (−8)
	E3B2				6 (1)	6 (3)	25 (7)
	E6B2					4 (1)	30 (23)
	E9B2						28 (18)
7	E0	7 (2.5)	7 (5.0)	9 (6.5)	9 (5.5)	7 (4.5)	11 (−24)
	E3B1		2 (−1.5)	4 (0.5)	6 (1.0)	7 (3.5)	28 (−1)
	ENH			7 (2.0)	7 (1.0)	6 (0.5)	19 (0)
	E3B2				7 (3.5)	6 (2.5)	27 (3)
	E6B2					8 (2.0)	31 (9)
	E9B2						34 (13)
9	E0	4 (−9)	5 (2)	4 (−6)	3 (−11)	1 (−18)	33 (42)
	E3B1		6 (3)	6 (0)	4 (−7)	1 (−19)	27 (14)
	ENH			3 (−11)	4 (−6)	1 (−24)	33 (46)
	E3B2				4 (−4)	2 (−21)	27 (8)
	E6B2					4 (−8)	21 (−20)
	E9B2						9 (−90)
10	E0	6 (1)	4 (0)	7 (2)	5 (0)	4 (0)	24 (−3)
	E3B1		3 (−3)	6 (0)	3 (−1)	5 (1)	29 (4)
	ENH			7 (0)	2 (0)	6 (1)	22 (−4)
	E3B2				4 (−1)	5 (0)	31 (3)
	E6B2					7 (2)	17 (−4)
	E9B2						27 (4)
11	E0	10 (13)	10 (11)	10 (15)	10 (17)	10 (20)	0 (−76)
	E3B1		6 (0)	5 (0)	8 (7)	9 (13)	22 (−7)
	ENH			7(2)	9 (8)	8 (9)	22 (−8)
	E3B2				10 (9)	9 (14)	23 (−6)
	E6B2					7 (2)	40 (39)
	E9B2						43 (58)
Mean	E0	70	60	75	67	58	34
	E3B1		42	53	48	52	55
	ENH			58	48	53	48
	E3B2				58	55	55
	E6B2					62	52
	E9B2						56

the control condition for moderate degrees of enhancement. As the degree of enhancement was increased, subject 10 showed little change in preference, subject 11 showed an increase in preference, and subjects 1, 4, 7, and 9 showed a decrease.

Consider now the intelligibility judgments (**Table 4**). For subjects 4, 7, and 11, all conditions involving enhanced speech in noise were judged to give higher intelligibility than the control condition, as indicated both by the numbers of selections and

by the distance scores. The overall preference and distance scores were lowest for the control condition. Subjects 1 and 10 did not show clear preferences for any condition, and their distance scores were all rather low. Subject 9 did not show clear preferences for conditions E3B1 and E3B1 + U relative to the control condition, but tended to prefer conditions E0, E3B1, and E3B1 + U over conditions involving large degrees of enhancement (E6B2 and E9B2).

In summary, the results of experiment 2 showed that, for judgments of both quality and intelligibility, speech in noise processed using a moderate degree of enhancement was generally preferred over the control condition. The results for higher degrees of enhancement varied across subjects. Subject 11 preferred the highest degree of enhancement both for quality and for intelligibility. For several other subjects, quality decreased for the highest degree of enhancement.

EXPERIMENT 3

Experiment 3 was similar to experiment 1, in that it involved measures of the intelligibility of speech in noise for several processing conditions. However, the processing differed from that used in experiment 1 in several ways. The first difference was in the way that the enhancement was performed. Instead of the DoG function, a function based on the difference between two rounded-exponential functions was used. This is equivalent to calculating two excitation patterns and taking the difference between them.

A second difference between experiments 1 and 3 was in the way that the enhancement signal was transformed into a gain function to modify the spectral shape of the signal. The transformation in experiment 3 was tailored to limit the maximum gain at any frequency to 20 dB (to avoid excessive increases in sound level) and was scaled so that, most of the time, the gain value was within reasonable limits. In addition, the enhancement function was applied to the original magnitude spectrum, rather than to the (normal) excitation pattern. This meant that only major spectral features were enhanced, but fine-grain spectral features were not smoothed in the conversion to an excitation pattern.

Finally, experiment 3 differed from experiment 1 by including conditions using fast-acting compression. This was done because the enhancement processing had the effect of expanding the dynamic range of the speech in noise. Potentially, this expansion could create problems for hearing-impaired subjects, who often have loudness recruitment and an associated reduction in usable dynamic range. The expansion of dynamic range produced by the enhancement processing might have offset the potential advantages to be gained from enhancement of spectral contrast. The compression used in experiment 3 was intended to compensate for the dynamic range expansion.

Method of Processing

Many of the stages in the processing were the same as used in experiment 1. Therefore, only the stages that were different will be described. The magnitude spectrum of a windowed sample of speech in noise was determined as before. An enhancement function was calculated by convolution of the power spectrum with the sum of a positive rounded-exponential function and a negative rounded-exponential function. For both the positive and negative functions, the ERB varied with center frequency according to equations described by Moore and Glasberg (4). The positive function had an ERB that was 0.5 times the "normal" value suggested by Moore and Glasberg, while the negative function had an ERB that was 2.0 times the normal value. The factors of 0.5 and 2.0 were chosen on the basis of informal listening tests. The sum of the two functions had a positive lobe whose width was approximately 0.67 ERB, which is intermediate between the B.5 and B1 values for the DoG function used in experiment 1. Each of the rounded exponentials was scaled, by dividing by its own ERB, so that the area under it was unity; thus, the area under the sum of the positive and negative rounded-exponentials was always zero.

The enhancement function will be designated D(f). It was converted to a gain function according to the following rules:

$$\text{Gain}(f) = 10^{K0.3D(f)} \quad \text{for } D(f) \leq 0 \quad [4]$$

$$\text{Gain}(f) = 10-(10-1)10^{-K0.3D(f)} \quad \text{for } D(f) \geq 0 \quad [5]$$

The resulting gain function was used to modify the original magnitude spectrum of the sample of speech

Journal of Rehabilitation Research and Development Vol. 30 No. 1 1993

in noise by multiplying the magnitude value at each frequency by the value of the gain function at that frequency. The form of equation [5] was chosen so as to limit the maximum gain at any frequency to a factor of 10 (20 dB). The value of the constant K was chosen to give a degree of enhancement comparable to the E3 enhancement of experiment 1. We refer to this processing condition as ENH. Stimuli for the control condition, E0, were obtained by processing stimuli in the same way but with the constant K set to zero. Subsequent to the enhancement processing, the stimuli in condition ENH were digitally filtered (by adjusting the magnitude spectrum prior to calculating the inverse FFT) so that the long-term-average spectrum of the processed noise matched the long-term-average spectrum of the noise in the control condition.

Examples of spectra for stimuli processed using conditions E0 and ENH are given in **Figure 3**. The stimulus was the same neutral vowel in noise as used for **Figure 1**. Note that compared with the processing condition E3B2 of experiment 1 (lower panel in **Figure 1**), condition ENH gave rise to sharper spectral peaks associated with the formants, and a greater spectral valley between the third and fourth formants. The difference can be attributed to the fact that the enhancement function used in experiment 1 was applied to the (normal) excitation pattern, whereas the enhancement function in experiment 3 was applied to the original spectrum.

Four conditions using compression were also run. The compression was implemented using an algorithm described by Robinson and Huntington (29). It was based on the use of a 20-ms sliding rectangular window. The rms value of the waveform within the window was calculated for each position of the window, and that value was used to calculate a gain function applied to the waveform sample at the center of the window. The compression took two forms. The first gave a moderate amount of compression, used a compression ratio of 2 and a compression threshold 10 dB below the peak value of the speech plus noise. We refer to this condition as C10/2. The second used a greater amount of compression, with a compression ratio of 3 and a compression threshold 15 dB below the peak value of the speech plus noise. We refer to this condition as C15/3. The compression was applied both alone and following the enhancement processing. This gave two additional conditions involving both enhancement and compression, ENHC10/2 and ENHC15/3.

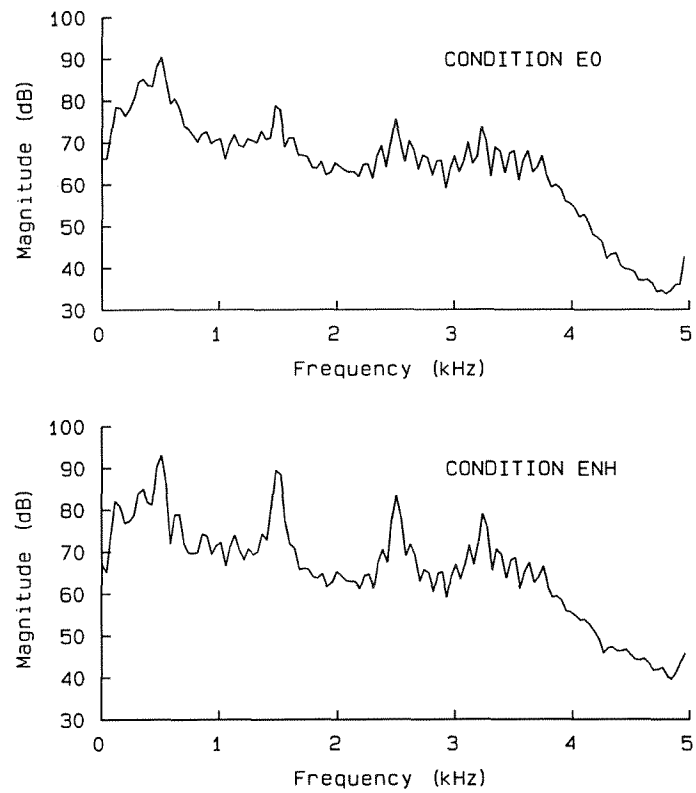

Figure 3.
Example of the effects of the processing used in Experiment 3 showing long-term-average spectra of a neutral vowel in noise, processed using the control condition (E0–top panel) and condition ENH (bottom panel).

In summary, six conditions were tested: the control condition, E0; a condition involving enhancement alone, ENH; two conditions involving compression alone, C10/2 and C15/3; and two conditions involving both enhancement and compression, ENHC10/2 and ENHC15/3. The overall level of the speech-plus-noise was equalized for all conditions.

Stimuli

The stimuli were lists 13–18 from the Adaptive Sentence Lists presented in a background of noise with the same long-term-average spectrum as the sentences. All subjects were tested at a speech-to-noise ratio of 0 dB, both speech and noise levels being specified in terms of root-mean-square pressures. Subjects were tested without using their hearing aids. In order to compensate for the lack of

Table 5.
Results of Experiment 3, showing the score for each subject in each condition (number of words correct out of 45) for the first and second tests, and the mean score for each condition.

		Condition					
Subject	Test	E0	C10/2	C15/3	ENH	ENHC10/2	ENHC15/3
8	1	40	42	37	37	36	34
	2	44	45	42	43	44	38
9	1	41	43	32	41	35	32
	2	36	43	36	37	42	33
10	1	32	30	39	42	30	23
	2	35	42	43	41	41	27
11	1	41	40	30	38	41	31
	2	37	43	38	41	44	38
12	1	42	43	40	41	33	41
	2	42	36	36	38	44	39
13	1	19	23	21	30	26	23
	2	29	24	15	38	34	24
Mean	1	35.8	36.8	33.2	38.2	33.5	30.7
	2	37.2	38.8	35.0	39.7	41.5	33.2

aids, which usually give a high-frequency emphasis, the off-tape signals were passed through a spectrum shaping network that rolled off at 12 dB/octave below 200 Hz, was "flat" from 200 to 400 Hz, and rose smoothly to +2 dB at 600 Hz and +15 dB at 4 kHz. This form of spectral shaping is similar to that commonly used in commercial hearing aids.

The level of the replayed speech in noise was adjusted for each subject to the value that they found comfortable for everyday conversation in a domestic environment. Other aspects of the stimuli were the same as for experiment 1.

Subjects

Six subjects were tested, four of whom had been used in experiment 1. All were diagnosed as having bilateral sensorineural hearing loss, probably of cochlear origin. They are subjects 8–13 in **Table 1**. Most were experienced hearing aid users.

Experimental Design

A double Latin Square design was used. Each subject was tested once in each of the six conditions, with the order of testing of conditions counterbalanced across subjects. This was then repeated but with the order of testing "rotated" so that the order of conditions for a given subject was different for the two Latin Squares. In each Latin Square, one ASL list was used for each subject and each condition.

Procedure

The procedure was essentially the same as for experiment 1. Subjects were given one practice list which was also used for adjusting the noise level.

Results

The raw scores are given in **Table 5**, which also shows the mean score across subjects for each condition for each of the two Latin Squares. Inspection of the data revealed a trend for performance to be better for the second Latin Square than for the first (i.e., there was a practice effect). Therefore, an ANOVA was conducted with factors condition and order of testing (first or second Latin Square) with the data blocked across subjects and lists (28). As for the data of experiment 1, the proportions correct were transformed using the expression arcsine($\sqrt{\text{proportion correct}}$). The analysis revealed a significant effect of condition, $F(5,50) = 5.89$, $p < 0.001$, and order of testing $F(1,50) = 12.04$, $p = 0.001$. The interaction of condition and order of testing approached, but did not reach, significance, $F(5,50) = 1.79$, $p = 0.13$.

Considering the mean scores for both Latin Squares, the highest scores were obtained for condi-

Journal of Rehabilitation Research and Development Vol. 30 No. 1 1993

tions ENH and C10/2 and the lowest for the conditions involving the greatest amounts of compression, C15/3 and ENHC15/3. *Post-hoc* tests, conducted as described earlier, showed that the mean scores for conditions ENH and C10/2 were significantly higher than the mean scores for conditions C15/3 and ENHC15/3 ($p < 0.01$ in all cases). The mean score for condition ENHC15/3 was also significantly lower than the mean score for the control condition, E0 ($p < 0.01$). Thus, a large amount of compression has deleterious effects. However, the scores for conditions E0, C10/2, ENH, and ENHC10/2 did not differ significantly from one another.

It seems reasonable to consider separately the scores for the second Latin Square, since there was evidence for improvements with practice. For the second Latin Square, the highest score overall was obtained for condition ENHC10/2, the condition involving both enhancement and a moderate degree of compression. The mean score for this condition (92 percent) was significantly greater than that for the control condition (82.6 percent) ($p < 0.05$). However, it was not significantly greater than the mean scores for conditions ENH (88 percent) and C10/2 (86.3 percent). The results of the second Latin Square for conditions E0, ENH, and ENHC10/2 are shown separately for each subject in **Figure 4**. For subject 8, the differences between conditions were limited by a ceiling effect; scores were close to perfect for all conditions. All of the other subjects scored better in condition ENHC10/2 than in condition E0.

In summary, the results of experiment 3 indicate that a large amount of compression, either used alone or in combination with spectral contrast enhancement, has deleterious effects on the intelligibility of speech in noise. The results showed clear effects of practice, suggesting that subjects may require time to get used to novel types of processing. The results for the second Latin Square (i.e., those obtained after a small amount of practice) indicated that the condition involving the combination of spectral contrast enhancement and a moderate amount of compression, ENHC10/2, gave a significantly higher mean score than the control condition.

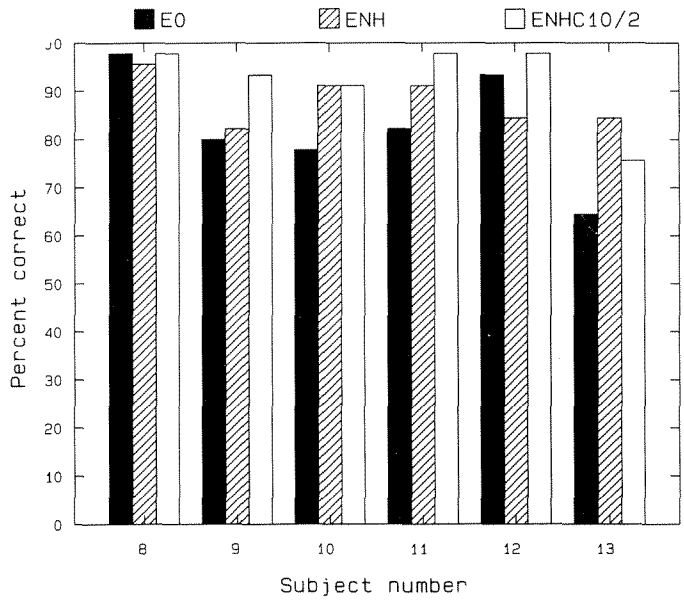

Figure 4.
Results of Experiment 3 for the second Latin Square, for the control condition (E0), the condition involving enhancement (ENH), and the condition involving enhancement combined with a moderate degree of compression (ENHC10/2).

Discussion

Although the results of experiment 3 suggest that the intelligibility of speech in noise may be improved by the enhancement of spectral contrasts, especially when combined with a moderate amount of compression, the effects were small. The small size of the effects probably arose partly from the lack of experience of the subjects with the processed stimuli; the results showed clear evidence of practice effects. This raises a dilemma. We wished to compare performance on several conditions, but we also wanted to avoid the possibility of subjects learning the sentence lists through repeated presentations. This latter requirement meant that it was not possible to give the subjects extensive practice on each condition.

A second factor that may have limited the size of the effects is related to the trade-off between accuracy and time/effort. Our subjects were effectively given as much time as they wanted to respond after each sentence had been presented. In difficult listening conditions, subjects may have devoted more effort and/or more time to the task of identifying each sentence. This would have resulted in reduced differences between conditions in comparison to the hypothetical situation where equal effort and/or time were devoted to all conditions.

It has previously been suggested that traditional speech intelligibility scores access only one component of disability and benefit, and that further information may be obtained by investigation of response times to speech stimuli (30,31,32,33). The response time aspects have previously been interpreted in terms of ease of listening. It may be argued that some or perhaps all of the benefits of spectral enhancement may accrue not from improvement of intelligibility, but rather from advantages to the listener in terms of the decreased difficulty (i.e., decreased effort required) in identifying the speech signal due to the sharper distinction of spectral cues. The availability of the sentence verification test, which yields measures of both speech intelligibility and response times, enabled this idea to be tested directly. A further advantage of this test is that, after an initial practice period, there is little evidence for improvements over time, and the materials themselves cannot be memorized. This made it possible to gather much more data for each subject and condition than in the earlier experiments.

EXPERIMENT 4

The Sentence Verification Test

The sentence verification test uses a closed vocabulary to construct four-word sentences from an overall vocabulary of 32 words. There are four alternatives for the first word in the sentence (LIZ, LYNNE, LEN, BEN), 12 alternatives for the second word (SOLD, SHOWED, STOLE, STORED, WORE, STITCHED, DROVE, CRASHED, CRACKED, CORKED, READ, TORE), 12 alternatives for the third word (FOUR, MORE, TWO, FEW, TWEED, CLOTH, FAST, SPORTS, GLASS, JAM, ROAD, STREET), and four alternatives for the fourth word (CAPS, CARS, JARS, MAPS). Of the 144 combinations of the second and third words, there are 82 for which there is at least one fourth word which makes the sentence unequivocally silly (nonsense) and at least one fourth word which makes the sentence unequivocally sensible (e.g., BEN SOLD STREET MAPS is sensible, while BEN SOLD STREET JARS is silly). Any combination of a fourth word with a second word–third word pair that may be considered equivocal with regard to sense/nonsense is not employed in the test. The eventual sentences require identification of the second, third, and fourth words in the sentence before a decision regarding the sense/nonsense of the sentence may be made.

The 32 words were stored as digitized waveform files which were isolated from sentences spoken by a single male talker and were concatenated to produce the desired sentences. During the construction of the test, care was taken to ensure that the intonation contours of the items, and other aspects, such as duration of voicing, were similar across items. This was done to remove extraneous cues not directly associated with the intelligibility of the individual word.

Following presentation of the sentence to the listener, the subject was asked to indicate whether the sentence was "silly" or "sensible" via a touch sensitive computer screen, and the response time for that decision (verification time) was recorded. This verification was followed by the identification component, for which four potential alternatives for the first word in the sentence, four for the second, four for the third, and four for the fourth were displayed on the touch sensitive computer screen. The subject was required to identify the components of the sentence. The test may be run either adaptively (yielding a signal-to-noise ratio for criterion performance) or at a fixed signal-to-noise ratio (yielding a percent correct score for the intelligibility component). The verification component of the test yields a median response time for all or a subset of the items for the cognitive decision concerning the sense/nonsense of the sentences. Evaluations of the within-session and between-session stability of the test for both normally hearing and hearing-impaired subjects has shown that there are no significant long-term learning effects associated with repeated administration of the closed vocabulary.

Processing of the Sentence Verification Test Items

Due to hardware constraints, the sentence verification test was available only in Glasgow. Hence, the stimuli to be processed were recorded in Glasgow, sent to Cambridge for processing, and then returned to Glasgow, using digital audio tape (DAT) as the recording medium. The 32 individual words constituting the vocabulary for the sentence verification test were each recorded at signal-to-noise ratios of 0, +3, +6, +9, and +12 dB, where the signal level was defined as the mean level of the speech peaks, and the noise level (shaped noise with

Journal of Rehabilitation Research and Development Vol. 30 No. 1 1993

the same long-term spectrum as the single male speaker) was defined as the rms level. A 1,000 Hz sine wave was included at the beginning of the recording to provide a reference level. These recordings were then sent to Cambridge and processed as described earlier, using three of the processing conditions from experiment 3: the control condition, E0; the condition involving enhancement alone, ENH; and the condition involving both enhancement and a moderate degree of compression, ENHC10/2. The processed stimuli were subjected to the high-frequency emphasis described for experiment 3, before being recorded on DAT tape. Each condition was recorded on a separate tape. The tapes were then returned to Glasgow.

The 15 sets of the 32 words (three conditions by five signal-to-noise ratios) were each redigitized using a CED 1401 laboratory interface and stored as individual waveform files. These waveform files were concatenated during testing to produce the required sentences.

Test Conditions

Because the processing was done with the noise added to the speech, the test had to be administered at fixed signal-to-noise ratios. Each subject was tested both unaided and with the level of the speech-plus-noise adjusted to a comfortable value. According to the experimental design described below, the required condition and signal-to-noise ratio was identified and a total of 55 sentences were delivered to the subject via a Grason-Stadler GSI 16 Audiometer and a Goodmans B41 loudspeaker in a sound-treated room with the subject seated 2 m from the loudspeaker at 0° azimuth. The first five of these sentences were not scored, but were regarded as practice within each individual run. The remaining 50 sentences were used, giving a score out of 200 for the identification component of the test. For the verification component of the test, only those sentences that were correctly identified (each of the four constituent words in the sentence identified correctly) *and* verified (correctly labeled as being either silly or sensible) were used. The median of the response times for the verification process using this subset of sentences was then derived. Thus, each run of the sentence verification test yielded an identification score out of 200 (here expressed as percent correct) and a response time (verification time) for the decision regarding the sense/nonsense of the sentence.

Subjects

The five subjects were all established users (at least 12 months) of a single post-aural BE10 series National Health Service hearing aid. The characteristics of the subjects are shown in **Table 1** (subjects 14–18). They all had broadly symmetric bilateral sensorineural losses of moderate degree, with greater losses at high frequencies than at low. All subjects had taken part in earlier experiments using the sentence verification test and were familiar with its form and configuration.

Experimental Design

The experiment consisted of five sessions for each subject, usually conducted at weekly intervals. Each session used seven complete runs of the sentence verification test as configured above. During each session, data were gathered for a pair of signal-to-noise ratios for each of the three processing conditions (E0, ENH, and ENHC10/2). An initial complete run for one of the signal-to-noise ratio/processing conditions was employed as practice, as previous experience with the sentence verification test suggested that optimal stability is achieved if this is done. The signal-to- noise ratios for each session were selected from a blocked design across subjects. Within each signal-to-noise ratio, the order of the three conditions was selected randomly. During the course of the five sessions, each subject was tested twice for each of the signal-to-noise ratios and each of the processing conditions.

Results

To show the overall form of the results, the mean of the two repetitions for each subject/signal-to-noise ratio/condition combination was taken and then the scores for the five subjects were averaged. The results are summarized in **Figure 5**. Error bars show 95 percent confidence limits. The figure shows the expected trend of increasing intelligibility and decreasing response times as the signal-to-noise ratio increases. For the identification component, there appear to be modest but consistent advantages at most signal-to-noise ratios for both of the processed conditions over the control condition (E0). For the

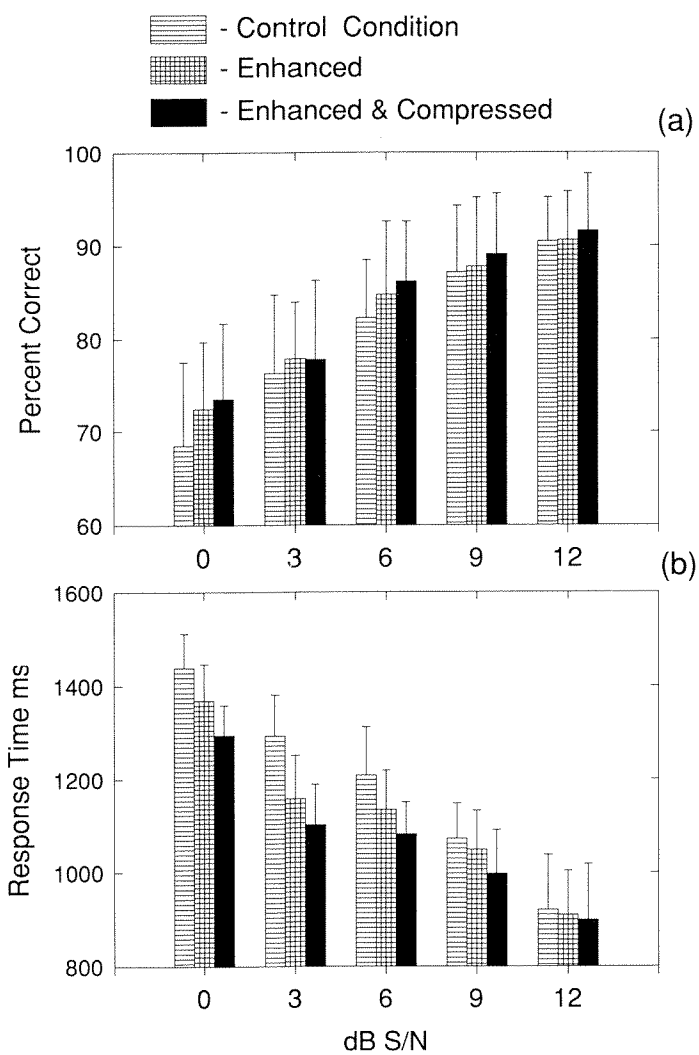

Figure 5.
Results of Experiment 4, showing the mean (and 95% confidence intervals for the mean) as a function of original signal-to-noise ratio and processing condition. Panel *(a)* shows scores for the identification component of the sentence verification test in terms of percent correct. Panel *(b)* shows the verification component in terms of response time.

response time component, the advantages of the processing conditions are larger, relative to the confidence limits, and there is a clear tendency for the processing condition involving both enhancement and compression (ENHC10/2) to give shorter response times than the condition involving enhancement alone (ENH).

The results of the five subjects were subjected to a repeated-measures ANOVA, using the GENSTAT package, with the following dependent variables: (i) percent correct score; (ii) arcsine ($\sqrt{\text{proportion correct}}$)—this measure makes the scores follow a normal distribution more closely; (iii) response time; and, (iv) square root of response time—again, this measure makes the scores follow a normal distribution more closely.

The results for the transformed variables (ii) and (iv) were similar to those for the untransformed variables (i) and (iii), so the latter will be presented to facilitate interpretation. In the ANOVA, there were three within-subject factors. The independent variables were: (i) the signal-to-noise ratio (0, 3, 6, 9, and 12 dB); (ii) the condition (linear, enhanced, enhanced and compressed); and, (iii) replication (first and second replicate).

For the percent correct scores, there was a highly significant effect of signal-to-noise ratio [$F(4,16) = 97.1$, $p < 0.001$], as expected from **Figure 5**, and a significant effect of condition [$F(2,8) = 6.65$, $p < 0.02$]. The main effect of replicate was not significant, and none of the interactions was significant. The mean score for condition ENH was 1.76 percent greater than that for condition E0 (standard error = 0.54), and this difference was statistically significant ($p < 0.02$). The mean score for condition ENHC10/2 was 2.73 percent greater than that for condition E0, and again this difference was significant ($p < 0.001$). The mean difference between conditions ENH and ENHC10/2, 0.97 percent, was not significant.

The ANOVA for the response time component of the sentence verification test showed highly significant effects of signal-to-noise ratio [$F(4,16) = 333.6$, $p < 0.001$] and of condition [$F(2,8) = 31.4$, $p < 0.001$]. The main effect of replicate was not significant, but there was a significant interaction between signal-to-noise ratio and processing condition [$F(8,32) = 4.07$, $p < 0.002$], consistent with the greater effect of condition at low signal-to-noise ratios apparent in **Figure 5**. The mean response time for condition ENH was 62.8 ms less than that for condition E0 (standard error = 10.1 ms), and this difference was statistically significant ($p < 0.001$). The mean response time for condition ENHC10/2 was 113 ms less than that for condition E0, and again this difference was significant ($p < 0.001$). The mean difference between conditions ENH and ENHC10/2, 52.8 ms, was also significant ($p < 0.001$). Thus, the results show that there are significant advantages for the processed conditions

compared with the control condition, with combined enhancement and compression giving bigger advantages than enhancement alone. The advantages are statistically more robust for the response-time component of the test than for the identification component.

The magnitudes of the effects described above, especially the response times, are difficult to interpret because of the somewhat complex nature of the sentence verification test. One way of relating the effects to other, more familiar measures, is to convert the differences in percent correct scores or response times to equivalent changes in signal-to-masker ratio. The data in **Figure 5** indicate that both the percent correct scores and the response times are approximately linearly related to the signal-to-noise ratio, for signal-to-noise ratios between 0 and +6 dB. For the control condition, each 1-dB increment in signal-to-noise ratio produces a 2.3 percent change in the percent correct score and a 38.3-ms change in the response time. These relationships were used to transform the magnitudes of the differences between conditions into equivalent changes in signal-to-noise ratio in dB.

For the percent correct scores, the difference between conditions E0 and ENH was equivalent to a 0.8-dB change in signal-to-noise ratio, while the difference between conditions E0 and ENHC10/2 was equivalent to 1.2 dB. For the response times, the difference between conditions E0 and ENH was equivalent to a 1.6-dB change in signal-to-noise ratio, while the difference between conditions E0 and ENHC10/2 was equivalent to 3.0 dB. Thus, the benefits of processing are approximately twice as large for the response-time component as for the identification component. If percent correct and response time can be regarded as subcomponents of an overall benefit from processing, then condition ENH gave an overall benefit of 2.4 dB compared with the control condition, and condition ENHC10/2 gave an overall advantage of 4.2 dB compared with the control condition.

The fully factorial, repeated-measures nature of the experimental design enabled individual differences to be investigated. For the identification scores, a general linear model (GLIM) analysis was conducted based on a logistic model, assuming that errors were distributed according to a binomial distribution. Here the proportion correct (PrC) is the dependent variable in an equation of the form:

$$PrC = 1/(1+\exp(-(B0+B1*X1+B2*X2+\ldots))) \quad [6]$$

where X1, X2, etc. are indices for specific values of the independent variables (signal-to-noise ratios, and processing conditions) and their interactions. The procedure produced estimates of the values of the parameters, B0, B1, B2, etc., referenced to a specific baseline, namely the mean for the control condition at 0 dB signal-to-noise ratio. B0 is the parameter estimate for the baseline itself. The effect of replicate was not significant and is not included in the analysis. The results are summarized in **Table 6**.

Table 6.
Summary of the results of the logistic regression analysis (GLIM) of the identification scores.

		S/N ratio				Condition		Interaction
Subject	Baseline	3 dB	6 dB	9 dB	12 dB	ENH	ENHC10/2	
14	0.315	0.282	0.691	0.836	1.234	0.169	0.182	N.S.
	(0.078)	(0.085)	(0.089)	(0.091)	(0.098)	(0.072)	(0.072)	
15	0.674	0.270	0.855	1.081	1.209	0.186	0.064	N.S.
	(0.083)	(0.089)	(0.098)	(0.103)	(0.106)	(0.080)	(0.078)	
16	0.816	0.442	0.627	1.157	1.549	−0.073	0.229	N.S.
	(0.087)	(0.096)	(0.099)	(0.112)	(0.125)	(0.084)	(0.088)	
17	1.050	0.327	1.001	1.260	2.041	0.116	0.277	$p < .01$
	(0.095)	(0.101)	(0.117)	(0.125)	(0.164)	(0.094)	(0.097)	
18	1.370	0.254	0.906	1.632	1.385	0.258	0.359	N.S.
	(0.104)	(0.112)	(0.130)	(0.164)	(0.151)	(0.105)	(0.108)	

Parameter estimates (and associated standard errors) are referenced to a baseline, namely the control condition with 0-dB signal-to-noise ratio. A parameter estimate for the baseline itself is given and an indication of the significance of the interaction of signal-to-noise ratio with condition.

To return to a percent correct score from a parameter estimate, the equation

$$\text{Percent correct} = 100/(1+e^{-(\text{sum of estimates})}) \quad [7]$$

is used. Thus, for subject 14 the figure of 0.315 for the baseline (control condition at 0 dB signal-to-noise) is equivalent to a score of 57.8 percent. Using the properties of the logistic regression, the effect of a combination of factors may be assessed by simple addition of the parameter estimates. Thus the estimate associated with a signal-to-noise ratio of 3 dB in the control condition is 0.315 + 0.282 = 0.597 (equivalent to a percent correct score of 64.5 percent) while the estimate associated with a signal-to-noise ratio of 3 dB in the enhanced condition for subject 14 is 0.315 + 0.282 + 0.169 = 0.766 (equivalent to a percent correct score of 68.3 percent).

The data in **Table 6** suggest that the processing conditions give different effects for different subjects. For example, subject 16 showed no benefit for enhancement alone, but showed a clear benefit for enhancement with compression. In contrast, subject 15 showed a clear benefit for enhancement alone, and showed less benefit for enhancement with compression. Subjects 14, 17, and 18 showed some benefit from enhancement alone, and showed larger benefits from enhancement with compression, although the differences between conditions ENH and ENHC10/2 were not significant. The interaction of signal-to-noise ratio with processing condition was significant only for subject 17.

The results for the response time estimates were analyzed using an identical linear model but assuming that errors were normally distributed. The results for the five subjects are shown in **Table 7**. The pattern was similar to that for **Table 6**, though now there was a significant interaction between signal-to-noise ratio and condition for subjects 16 and 18. Overall, the effects of condition were more robust (as can be seen by comparing the parameter estimates with their associated standard errors). For subjects 14 and 15, the benefit of condition ENH was not significant, but the benefit of condition ENHC10/2 was significant. For subjects 16, 17, and 18, there were significant benefits in both conditions. The benefits tended to be larger in condition ENHC10/2 than in condition ENH, but the differences were not significant. It is noteworthy that the identification scores of subject 16 did not show a benefit for condition ENH, whereas the response-time scores did.

Although the experiment contained relatively small numbers of subjects, the nonhomogeneous pattern of results does suggest that, in future experiments, it would be worthwhile to investigate further the characteristics of individual subjects to try to find the predictors of benefit from enhancement.

GENERAL SUMMARY, DISCUSSION, AND CONCLUSIONS

The results of experiment 1 were disappointing, in that they failed to show any significant benefits of

Table 7.
Summary of the results of the logistic regression analysis (GLIM) for the response times.

		S/N ratio				Condition		Interaction
Subject	**Baseline**	**3 dB**	**6 dB**	**9 dB**	**12 dB**	**ENH**	**ENHC10/2**	
14	1357 (29.8)	−212 (33.3)	−248 (33.3)	−350 (33.3)	−513 (33.3)	−37 (25.8)	−111 (25.8)	N.S.
15	1469 (25.8)	−160 (28.8)	−213 (28.8)	−293 (28.8)	−433 (28.8)	−20 (22.3)	−85 (22.3)	N.S.
16	1383 (29.8)	−160 (33.3)	−220 (33.3)	−310 (33.3)	−403 (33.3)	−71 (25.8)	−83 (25.8)	$p < .01$
17	1535 (29.3)	−165 (32.7)	−198 (32.7)	−343 (32.7)	−432 (32.7)	−117 (25.4)	−165 (25.4)	N.S.
18	1378 (29.3)	−217 (32.8)	−248 (32.8)	−355 (32.8)	−512 (32.8)	−69 (25.4)	−121 (25.4)	$p < .01$

the enhancement processing, although the results did indicate that a large degree of spectral enhancement can have deleterious effects. In hindsight, the failure to find positive effects of the processing in experiment 1 probably can be attributed in part to the experimental design, which did not give subjects the opportunity to practice in the different conditions.

The results of experiment 2 showed that subjective ratings of both quality and intelligibility were affected by the processing, but the effects varied across subjects. For judgments of both quality and intelligibility, speech in noise processed using a moderate degree of enhancement was generally preferred over the control condition. The results for higher degrees of enhancement varied across subjects. Subject 11 preferred the highest degree of enhancement both for quality and intelligibility. For several other subjects, quality decreased for the highest degree of enhancement.

The results of experiment 3 indicated that a large amount of compression, either used alone or in combination with spectral contrast enhancement, had deleterious effects on the intelligibility of speech in noise. The results showed clear effects of practice, suggesting that subjects may require time to get used to novel types of processing. The results for the second Latin Square (i.e., those obtained after a small amount of practice) indicated that the condition involving the combination of spectral contrast enhancement and a moderate amount of compression, ENHC10/2, gave a significantly higher mean score than the control condition.

Taken together, the results of experiments 1, 2, and 3 indicate that high degrees of enhancement, or high degrees of compression, generally have deleterious effects. In other words, too much processing is a bad thing! However, the results of experiment 2 indicate that a moderate amount of spectral enhancement can lead to improved subjective ratings of quality and intelligibility, and the results of experiment 3 indicate that, after some practice, a moderate degree of spectral enhancement, combined with a moderate degree of compression, can give better results than those obtained with unprocessed speech.

Experiments 1 and 3 suffered from the problem that subjects were given rather little practice in each condition. This was forced upon us because, with the limited number of sentence lists available to us, it would not have been possible to give extensive practice without subjects memorizing the lists. The limited number of sentence lists created a second problem; it was impossible to gather a large amount of data for each condition. This meant that some of the effects observed were of marginal statistical significance. A third problem was that the measure used, the percent correct of words identified in short sentences, may not have been suitable for revealing all of the effects of the processing. Specifically, the measure probably did not tap the dimension of "ease of listening," which can be especially important in everyday situations involving decision making and selective attention.

The Sentence Verification Test used in experiment 4 was intended to overcome these problems. The test can be administered repeatedly without substantial learning effects, and it includes a measure of response time which is probably related to ease of listening. The results showed highly significant benefits of the processing, with spectral enhancement alone being superior to the control condition, and enhancement combined with compression being superior to enhancement alone. When expressed in terms of equivalent changes in signal-to-masker ratio, the benefits were about twice as great for the response time measures as for the identification scores, and they were also statistically more robust for the response time measures. This suggests that the major benefits of the processing may be in terms of increased ease of listening rather than in intelligibility.

The results of experiment 4 indicate that the improvement in the intelligibility score produced by processing alone was equivalent to a change in signal-to-noise ratio of about 0.8 dB, a relatively modest amount. The results of Simpson, et al. (19) for spectral processing with a similar degree of enhancement (although implemented using a somewhat different algorithm), showed typical improvements in intelligibility, relative to the control condition of about 7 percent. For the speech materials used by Simpson, et al., each 1-dB change in speech-to-noise ratio produces about an 11 percent change in intelligibility (34). Thus, the 7 percent change in intelligibility is equivalent to about a 0.6-dB change in speech-to-noise ratio. This is comparable to the 0.8-dB change found in experiment 4.

It should be emphasized that the overall effect of the processing found in experiment 4 was larger

than this. If the changes in intelligibility and in response times were both expressed in terms of equivalent change in speech-to-noise ratio, the net effect was an improvement (relative to the control condition) of 2.4 dB for enhancement alone, and 4.2 dB for enhancement combined with compression.

Experiments 1 and 3 used a Latin Square design, which makes it difficult to analyze the effects of individual differences. However, the results of experiment 2 showed clear evidence of individual differences in the judged pleasantness and intelligibility of the processed stimuli. Similarly, both the intelligibility measures and the response time measures of experiment 4 revealed clear individual differences. Further research is needed to clarify why these differences occur, and to establish whether they can be related to individual differences in psychoacoustic factors such as frequency selectivity.

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council (UK) and the Hearing Research Trust (UK). We thank Brian Glasberg for assistance during many stages of this work and Adrian Davis for assistance with statistical analysis. Joseph Alcántara gave helpful comments on an earlier version of this paper.

REFERENCES

1. Plomp T. Auditory handicap of hearing impairment and the limited benefit of hearing aids. J Acoust Soc Am 1978;63:533-49.
2. Plomp R. A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. J Speech Hear Res 1986;29:146-54.
3. Fletcher H. Auditory patterns. Rev Mod Phys 1940;12:47-65.
4. Moore BCJ, Glasberg BR. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J Acoust Soc Am 1983;74:750-3.
5. Patterson RD, Moore BCJ. Auditory filters and excitation patterns as representations of frequency resolution. In: Moore BCJ, editor. Frequency selectivity in hearing. London: Academic, 1986: 123-77.
6. Moore BCJ, Glasberg BR. Formulae describing frequency selectivity as a function of frequency and level and their use in calculating excitation patterns. Hear Res 1987;28:209-25.
7. Glasberg BR, Moore BCJ. Derivation of auditory filter shapes from notched-noise data. Hear Res 1990;47:103-38.
8. Glasberg BR, Moore BCJ. Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. J Acoust Soc Am 1986;79:1020-33.
9. Tyler RS. Frequency resolution in hearing-impaired listeners. In: Moore BCJ, editor. Frequency selectivity in hearing. London: Academic Press, 1986: 309-71.
10. Dreschler WA, Plomp R. Relations between psychophysical data and speech perception for hearing-impaired subjects. I. J Acoust Soc Am 1980;68:1608-15.
11. Dreschler WA, Plomp R. Relations between psychophysical data and speech perception for hearing-impaired subjects. II. J Acoust Soc Am 1985;78:1261-70.
12. Glasberg BR, Moore BCJ. Psychoacoustic abilities of subjects with unilateral and bilateral cochlear impairments and their relationship to the ability to understand speech. Scand Audiol Suppl 1989;32:1-25.
13. Havens JF, Geisler CD. Speech recognition and frequency selectivity for hearing impaired listeners. Proceedings of the International Conference on Acoustics Speech and Signal Processing; 1991 May 14-17; Toronto (ON). New York: IEEE, 1991: 3629-32.
14. Leek MR, Dorman MF, Summerfield Q. Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners. J Acoust Soc Am 1987;81:148-54.
15. Rosen S, Fourcin A. Frequency selectivity and the perception of speech. In: Moore BCJ, editor. Frequency selectivity in hearing. London: Academic, 1986: 373-487.
16. Young ED, Sachs MB. Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibres. J Acoust Soc Am 1979;66:1381-403.
17. Boers PM. Formant enhancement of speech for listeners with sensorineural hearing loss. IPO Ann Prog Rep 1980;15:21-8.
18. Summerfield AQ, Foster J, Tyler R, Bailey PJ. Influences of formant narrowing and auditory frequency selectivity on identification of place of articulation in stop consonants. Speech Commun 1985;4:213-29.
19. Simpson AM, Moore BCJ, Glasberg BR. Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners. Acta Otolaryngol Suppl 1990;469:101-7.
20. Stone MA, Moore BCJ. Spectral feature enhancement for people with sensorineural hearing impairment: effects on speech intelligibility and quality. J Rehabil Res Dev 1992;29(2):39-56.
21. Bunnell HT. On enhancement of spectral contrast in speech for hearing-impaired listeners. J Acoust Soc Am 1990;88:2546-56.
22. Lim JS. Speech enhancement. New Jersey: Prentice Hall, 1983.
23. Cheng YM, O'Shaughnessy D. Speech enhancement based conceptually on auditory evidence. IEEE Trans Sig Proc 1991;39:1943-54.
24. Clarkson PM, Bahgat SF. Envelope expansion methods for speech enhancement. J Acoust Soc Am 1991;89:1378-82.
25. Allen JB. Short term spectral analysis, synthesis and

modification by discrete Fourier transform. IEEE Trans Acoust Speech Sig Proc 1977;25:235-8.

26. Patterson RD, Nimmo-Smith I, Weber DL, Milroy R. The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold. J Acoust Soc Am 1982;72:1788-803.
27. MacLeod A, Summerfield Q. A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use. Brit J Audiol 1990;24:29-43.
28. Alvey N, Galwey N, Lane P. An Introduction to GENSTAT. London: Academic Press, 1982.
29. Robinson CE, Huntington DA. The intelligibility of speech processed by delayed long-term-averaged compression amplification. J Acoust Soc Am 1973;54:314.
30. Hecker MH, Stevens KN, Williams CE. Measurements of reaction time in intelligibility tests. J Acoust Soc Am 1966;39:1188-9.
31. Pratt RL. On the use of reaction time as a measure of intelligibility. Brit J Audiol 1981;15:253-5.
32. Cox RM, Alexander GC, Gilmore C. Development of the Connected Speech Test (CST). Ear Hear 1987; 8(Suppl):119S-26S.
33. Gatehouse S, Gordon J. Response times to speech stimuli as measures of benefit from amplification. Brit J Audiol 1990;24:63-8.
34. Laurence RF, Moore BCJ, Glasberg BR. A comparison of behind-the-ear high fidelity linear aids and two-channel compression hearing aids in the laboratory and in everyday life. Brit J Audiol 1983;17:31-48.