Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task

Michael A. Stone[a] and Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

(Received 24 September 2003; revised 26 April 2004; accepted 30 June 2004)

Using a cochlear implant simulator, Stone and Moore [J. Acoust. Soc. Am. **114**, 1023–1034 (2003)] reported that wideband fast-acting compression led to poorer intelligibility than slow-acting compression in a competing speech task. Compression speed was varied by using different pairs of attack and release times. In the first experiment reported here, it is shown that attack times less than about 2 ms in a wideband compressor are deleterious to intelligibility. In experiment 2, fast wideband compression was applied to the target and background either before or after mixing. The former reduced the modulation depth of each signal but maintained the independence between the two signals, while the latter introduced "comodulation." Using simulations with 6 and 11 channels, intelligibility was higher when compression was applied before mixing. In experiment 3, wideband compression was compared with multichannel compression; the latter led to reduced comodulation effects. For 6 channels, the position of the compressor, either wideband or within each channel, had no effect on intelligibility. For 11 channels, channel compression severely degraded intelligibility compared to wideband compression, presumably because of the greater reduction of across-channel contrasts. Overall, caution appears necessary in the use of fast-acting compression in cochlear implants, so as to preserve intelligibility.
[DOI: 10.1121/1.1784447]

PACS numbers: 43.66.Ts, 43.71.Gv, 43.66.Mk [DDO]   Pages: 2311–2323

I. INTRODUCTION

Cochlear implants all contain some form of compression system to map the wide range of sound levels encountered in everyday life into the limited dynamic range (6–20 dB) available at each electrode (Clark *et al.*, 1990; Moore, 2003a; Zeng, 2004). There are many parameters that may be varied in the design of compression systems, among which are the attack and release times, which determine whether the compression should be described as "fast" (also called syllabic compression) or "slow" (also called automatic volume control). For reviews, see Dillon (1996) and Moore (1998).

We (Stone and Moore, 2003a) have previously described the use of a "noise-vocoder" simulation of a cochlear implant (Shannon *et al.*, 1995) to assess the effect of the speed of response of a wideband (single-channel) dynamic range compressor on speech intelligibility in a competing speech task. The compressors assessed were representative of those used in commercial cochlear implant systems. We found that fast-acting compression degraded speech intelligibility when compared to slow-acting compression: the slow-acting compression itself produced no significant degradation compared to no compression at all. We suggested that the reduction could be explained by two factors: (1) Fast compression introduces correlated fluctuations in amplitude in different frequency bands, which may promote perceptual fusion of the target and background sounds (Bregman *et al.*, 1985; Bregman, 1990; Hall and Grose, 1990; Carrell and Opie, 1992; Moore *et al.*, 1993; Darwin and Carlyon, 1995); (2) Fast compression reduces amplitude modulation depth and intensity contrasts; the importance of this effect has been the subject of much debate (Villchur, 1973; Plomp, 1988; 1994; Moore, 1990; 2003b; Dillon, 1996). For brevity, we refer to factor (1) as "comodulation" and factor (2) as "modulation reduction."

The comodulation introduced by fast-acting single-channel compression is actually of a somewhat unusual type. When two fluctuating signals (e.g., two talkers) with a similar overall level are mixed, the gain at any instant is determined mainly by the level of the signal that has the higher level. Peaks in one signal result in a reduction in gain (although the peak remains a peak following compression) that decreases the output level of the other signal. Hence, the comodulation of the wideband stimuli is out of phase; increases in level of one signal result in decreases in level of the other signal, and vice versa. While the perceptual grouping of two signals appears to be affected by the similarity in the envelopes of the signals, there is little evidence to support the idea that the relative phase of modulation of the signals is important, and there is some evidence to the contrary. For example, the perceptual segregation of pairs of vowels is not influenced by the relative phase of amplitude modulation of the vowels (Summerfield and Culling, 1992). Also, modulation detection interference, which is believed to result at least partly from perceptual grouping processes (Hall and Grose, 1991; Moore and Shailer, 1992), is not greatly affected by the phase of modulation of the interferer relative to that of the target (Moore and Shailer, 1992; Moore, 1992; Hall *et al.*, 1995). Thus, the type of comodulation introduced by a single-channel fast-acting compressor

[a]Electronic mail: mas19@cam.ac.uk

could decrease perceptual segregation even though it is out of phase.

When considering the effects of comodulation, another factor needs to be taken into account. The gain of a wideband compressor is controlled by the components of the input that are most intense, and these tend to be the lower-frequency components of speech signals. When the signal is split into frequency channels (by the auditory system or by an implant processor), the temporal fluctuations in channels passing the most intense components may be almost independent of those in remote (typically high-frequency) channels. Under these conditions, the target and background become comodulated *in phase* in the remote channels, because the same gain changes are imposed on the target and background. Thus, the pattern of comodulation is actually rather complex and varies depending on the nature, number, and center frequency of the channels involved, as well as on the statistical properties of the signal.

In our earlier experiments, we compared two systems with markedly different attack and release times, but with the same (high) compression ratio. Since both attack and release times were varied, we were not able to assess whether there were any specific effects associated with a fast attack time. However, there are theoretical reasons, discussed below, for thinking that such effects might occur. The first experiment reported here investigated the effect on intelligibility of varying the attack time of a wideband (single-channel) compressor, over the range 0.125 to 8 ms, while keeping other aspects of the compressor performance as similar as possible across conditions.

The data reported in our earlier experiments did not allow us to assess the relative importance of comodulation and modulation reduction. Experiments 2 and 3 were intended to clarify this issue. Experiment 2 compared the effect on intelligibility of wideband compression applied before or after mixing the target signal with the background interference. Compression before mixing produces modulation reduction without comodulation, while compression applied after mixing produces both modulation reduction and comodulation. The third experiment compared the intelligibility produced by use of a single, wideband compressor to that produced by a system employing multichannel independent compression, for a simulated implant with either 6 or 11 channels. The single wideband compressor produced both modulation reduction and comodulation, while the multichannel independent compression produced only limited comodulation; the comodulation occurred within but not across channels. If comodulation is an important effect, we would expect better performance with multichannel than with single-channel compression. However, other factors may influence the outcome, as discussed later.

II. EXPERIMENT 1: EFFECT OF ATTACK TIME ON INTELLIGIBILITY

A. Introduction and rationale

For an acoustic hearing-aid application, ANSI (1996) defines the attack time as the time taken for the output of the compressor to settle to within 2 dB of its steady state value after a 25-dB step increase in level at its input. In some systems, it is important to avoid high signal levels so as to avoid distortion (as in broadcasting or in hearing aids) and/or discomfort (as in hearing aids or cochlear implants). In such systems, it has been common to use fast attack times in the compressors. This ensures that the gain reacts rapidly to sudden increases in the input level. However, there are at least two reasons why a fast attack time might have detrimental effects; these are discussed below.

A fast change in gain is equivalent to a fast amplitude modulation of the signal being controlled. The fast modulation generates spectral sidebands around the frequency components in the original signal, which, in an acoustic aid, may momentarily mask other components present in the signal. In a cochlear-implant system, a fast-acting wideband compressor acting before the separation into channels may lead to spectral components within one or more channels that were not present in the original signal. The spread in frequency of the spectral sidebands is an inverse function of the attack time. The level of the sidebands increases with increasing compression ratio; high compression ratios produce large changes in gain. For input modulation rates slower than the reciprocal of the sum of the attack and release times, the gain changes can be almost as large as the signal modulation depth (Braida *et al.*, 1982; Stone and Moore, 1992). If the attack time is made longer, this reduces the level and spread of the sidebands, but the sidebands last for a longer time. However, with sufficiently long attack times, the spread becomes negligible. Hence, the deleterious effects of spectral spreading are likely to be greater for short attack times.

A second possible detrimental effect of short attack times in a wideband compressor is connected with the comodulation effect described in the Introduction. At the onset of a sound, a fast attack time results in a rapid decrease in gain, so any co-occurring sound has an abrupt decrease in level imposed on it. In other words, an abrupt increase in level of one sound becomes associated with an abrupt decrease in level of the other. The faster the attack time of the compressor, the more likely it will be to ‘‘track’’ rapid onsets in level of the input signal. There is evidence that the auditory system contains a mechanism for detecting abrupt changes in level that is not sensitive to the polarity of the change (Macmillan, 1973; Hafter *et al.*, 1996). Also, synchrony of onsets has been shown to be a powerful cue for perceptual grouping (Rasch, 1978; Bregman, 1990; Darwin and Carlyon, 1995). The synchronous abrupt changes in the two sounds produced by gain changes in the compressor may thus promote perceptual fusion of the two sounds even though the changes are in opposite directions. Also, as described in the Introduction for the general case of comodulation, when the signal is split into channels, the abrupt changes may actually be in phase for channels remote from those passing the most intense components.

In this experiment, we varied the attack time of the wideband compressor over a wide range and investigated its effect on intelligibility using a noise-vocoder simulation of a cochlear implant. To isolate the effects described above from effects related to modulation reduction, small adjustments to the release time were also applied, in such a way that modulation reduction was held constant as the attack time was varied. Because of the number of conditions assessed, the number of channels in the implant simulation during the main test was fixed at a value (8) similar to the number of independent channels typically achieved in cochlear implants (Clark *et al.*, 1990; Friesen *et al.*, 2001; Moore, 2003a).

B. Signal processing for simulation of a cochlear implant

The signal-processing software, written in MATLAB,™ was the same as used by Stone and Moore (2003a). Briefly, the signal to be processed was "pre-emphasized" by applying a gain rising at 3.3 dB/octave between 500 and 4000 Hz, giving a total of 10 dB. Below 500 Hz the gain was 0 dB, and above 4000 Hz it was 10 dB. The signal was then filtered into one of a specified number of channels. All filters were implemented using a finite-impulse-response (FIR) method, which introduced a frequency-independent time delay. They were designed such that the low-pass edge of one filter had the complementary response to the high-pass edge of the next higher filter. The outputs of all filters were time aligned. Addition of the channel outputs gave virtually perfect recombination: within the frequency range 100 to 7800 Hz, the ripple was less than ± 0.2 dB.

After filtering, the channel envelopes were extracted by full-wave rectification and low-pass filtering. The channel envelopes were each used to modulate a broadband white noise, after which the modulated noises were bandlimited to the same widths as those of the corresponding analysis filters. Time delays were added to the channel signals to compensate for the time delays introduced by the FIR filters (these delays varied with center frequency), such that the relative timing of envelope signals across frequency was the same for the input and the output. The filtered noise bands were then added back together and a linear-phase 6-pole infinite impulse response high-pass filter with a corner frequency of 100 Hz was applied to remove spurious low-frequency noise.

C. Subjects and equipment

Twenty volunteer subjects (6 males, 14 females, aged 18–41 years), all university undergraduates or graduates, were selected on the basis of their having audiometric thresholds $\leqslant 15$ dB HL at octave frequencies between 125 and 8000 Hz and at 3000 and 6000 Hz. None had prior experience of the processing. All were native speakers of British English. Subjects attended one session, which lasted nearly 2 h after initial audiometric screening. Subjects were paid for their attendance. The equipment was the same as used by Stone and Moore (2003a). Signals were generated via a high-quality sound card in a PC, passed through a mixing desk, and presented diotically through Sennheiser HD580 headphones. Each subject was seated in a sound-isolated, double-walled chamber.

The speech material for both target and background was the same as used by Stone and Moore (2003a). The target sentences were from the ASL corpus (MacLeod and Summerfield, 1990), and were spoken by a male speaker of British English. Additionally, four lists of 17 sentences each were prepared from recordings by a female speaker of British English of lists 43 to 57 of the IEEE sentence lists (IEEE, 1968). These four lists were used for training only. All speech (both target and background) was filtered so as to have the same long-term average spectrum as specified in ANSI (1997) for normal conversational levels. The sentence material used for the test proper comprised lists of 15 sentences with 3 keywords in each. The background was continuous running speech produced by a male speaker of British English. Pauses and hesitations were removed by hand editing. The input level of the target speech to the processing was 67 dB SPL (unweighted). The presentation level was 68 dB SPL (unweighted).

D. Procedure

Since the processed signals were novel to the subjects, a training period was necessary before testing proper began. Stone and Moore (2003a) reported that a training period of at least one-half hour was necessary to achieve stable results, but even after this, some learning effects were apparent in the counterbalanced data. Compared to Stone and Moore (2003a), we lengthened the initial training period and made it more interactive.

The initial training comprised two phases. In the first phase, four IEEE sentence lists were presented using progressively decreasing target-to-background ratios (99 to 10 dB) and decreasing numbers of channels (16 to 8). If, after the second presentation of a sentence, the keywords were still not repeated correctly, the experimenter spoke the correct answer and then made a further presentation to allow the subject to verify the answer. The background speaker was not the same as used in the test proper.

In the second phase of initial training, the subject was familiarized with the target and background speakers for the test proper. Two lists were presented, one processed as 8 channels with a +10-dB target-to-background, and the second as 8 channels with a +5-dB target-to-background. These 30 training sentences were not used again either for training or testing. The processing conditions used in the initial training were varied in a counterbalanced order across subjects. This completed the initial training, but before each condition was tested, there was an additional training phase. Eighteen further sentences were presented up to three times, using 8 channels and a +2-dB target-to-background ratio.

After all the training, the intelligibility test was conducted at a signal-to-background ratio of +2 dB. A total of 33 target sentences (which had not been used in the training) was used for each condition, with 3 keywords in each. For each presentation of a target sentence, the background was ramped up over 0.25 s, the ramp starting 1 s before the target sentence. After the target sentence was completed, the background continued for about a further 0.5 s before being ramped down over 0.25 s. Subjects were encouraged to respond after each presentation, even if the sentence appeared to be nonsense. Five conditions differing in the compression system used were tested in a counterbalanced order across subjects.

E. Equating compression conditions

The primary purpose of this experiment was to compare four fast-acting compression systems similar to the fast-acting system used by Stone and Moore (2003a), but having attack times of 0.125, 0.5, 2, and 8 ms. A fifth system was also tested which was intended to be more effective than the other systems in reducing envelope modulations over the range of modulation rates important for speech perception (Plomp, 1983; Drullman *et al.*, 1994a; 1994b). This system is described in more detail later. The fast-acting system used by Stone and Moore (2003a) had an attack time of 2 ms, a release time of 240 ms, a compression ratio of 7, and a compression threshold of 55 dB SPL. We used the same compression ratio and compression threshold as before. The compression threshold is lower than usually employed in cochlear implant processors, but is not unrealistically low. The relatively low threshold was chosen to ensure that the compressor was active during presentation of the target speech.

The minimum attack time of 0.125 ms is close to the instantaneous attack time used to prevent overload in some electrical circuits, for example as a limiter in broadcasting. The maximum attack time is similar to the maximum value found in fast-acting compressors in commercial hearing aids. When a relatively long attack time is used with such a high compression ratio, it is desirable to incorporate a small delay to the audio path, typically about half the attack time in duration, so as to prevent large ''overshoots'' at signal onsets. Such a delay has been included in experimental acoustic hearing aids (Robinson and Huntington, 1973; Bustamante and Braida, 1987; Baer *et al.*, 1993; Verschuure *et al.*, 1996; Stone *et al.*, 1999). Here, a delay equal to half the attack time was inserted into the audio path for all the configurations of ''fast'' compression. The maximum delay of 4 ms might be subjectively noticeable in some situations (Agnew and Thornton, 2000), but it would not be subjectively disturbing or objectively degrading (Stone and Moore, 2002; 2003b).

In this experiment, we wished to isolate the effects of fast attack times on spectral spreading and comodulation of onsets from the effect of modulation reduction. To do this, we equated modulation reduction across conditions by making small adjustments to the release time. Stone and Moore (2003a) described a metric, called the fractional reduction in modulation, f_r, which allowed comparison of the effect of different compression systems on envelope modulations. Typically, f_r is a function that varies with envelope frequency, and it depends on both the attack and the release time of a compression system. In order to match f_r between the systems used here, the release time, t_r (ms), was set to

$$t_r = 244 - 2t_a, \tag{1}$$

where t_a is the attack time (ms). For the range of attack times tested here, the release time varied only over a small range (228–243.75 ms). The value of f_r for the four fast compression systems is plotted as a function of envelope modulation frequency in Fig. 1 (curve marked ''Fast''); the curve was almost identical for the four attack times used.

The fifth system tested, mentioned earlier, was similar to that of Robinson and Huntington (1973) and of one of the

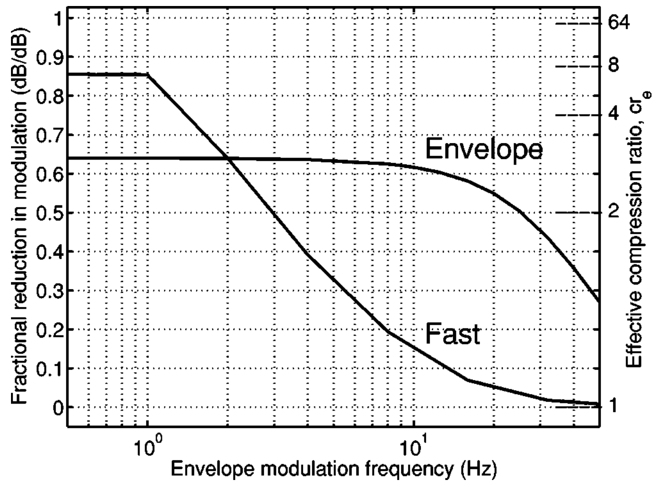

FIG. 1. Effect on envelope modulations of the two compressor types under test, as a function of the modulation rate of the input. The left ordinate shows the fractional reduction in modulation and the right ordinate shows the effective compression ratio. The curve marked ''Fast'' shows results for all four attack times used with the fast compressor. The curve marked ''Envelope'' shows results for the envelope compressor.

compressors used in the ''FULL-4'' system of Stone *et al.* (1999). The envelope of the signal was extracted and passed through a 2-pole Bessel-derived low-pass filter, resulting in nearly symmetric attack and release times of 10 and 13 ms, respectively. Note that the attack time is similar to that for the fast system with the longest attack time (8 ms). In order to achieve the near-symmetry of attack and release times for the fifth system, the audio signal was delayed by 6.3 ms before the gain signal was applied. Again, this delay has no significant perceptual effects. The low-pass filter design had negligible overshoot in its step response, and the two-pole design, with a corner frequency of 24 Hz, ensured that the filter output was very low for envelope modulation rates corresponding to voice fundamental frequencies (f_r was nearly zero at an envelope rate of 100 Hz). This envelope compressor achieved the ''target'' f_r for envelope rates up to about 10 Hz. The resulting compressed signal had intermodulation distortion less than about 3% (Moore *et al.*, 1999). The compression ratio, 2.78, was chosen to be equal to the effective compression ratio of the fast-acting systems, for an envelope modulation rate of 2 Hz. This rate corresponds to the lower limit of ''useful'' envelope modulation rates in speech, as found by Drullman *et al.* (1994a; 1994b). The value of f_r for the envelope compression system is plotted as a function of envelope modulation frequency in Fig. 1 (curve marked ''Envelope''). The envelope compressor achieves far more effective compression than the ''fast'' compressors over the range of envelope modulation rates significant for speech perception (Drullman *et al.*, 1994a; 1994b). We therefore expected the envelope compressor to produce lower intelligibility than any of the fast compressors. The compression threshold of the envelope compressor was the same as for the fast compressors, namely 55 dB SPL.

Stone and Moore (2003a) used a target-to-background ratio of +5 dB throughout their three experiments, and, for an 8-channel system, the average score was around 70%–75%. Typical slopes of psychometric functions for the intelligibility of speech against an interfering voice have been

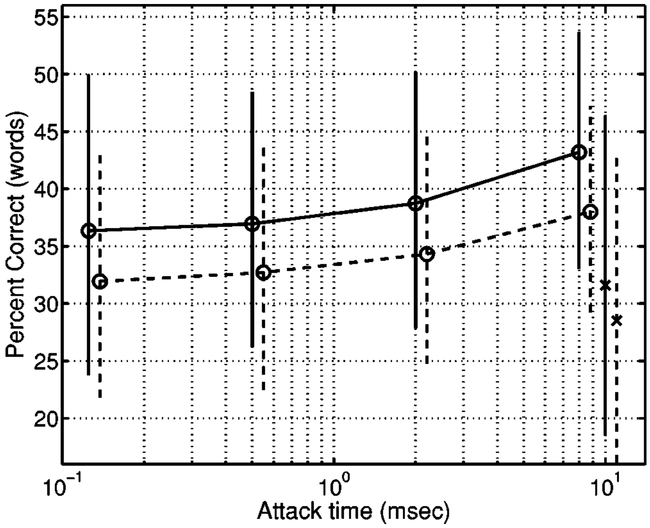

FIG. 2. Results of experiment 1. The solid and dashed lines show data corrected and uncorrected for learning effects, respectively. The connected lines show data for the fast compressors. The separate crosses show results for the envelope compressor. Error bars show ± 1 SD. The target-to-background ratio was 2 dB, and eight channels were used in the simulation.

reported as between 7% and 13%/dB around the 50%-point on the psychometric function (Baer and Moore, 1994; Festen and Plomp, 1990). To avoid ceiling effects and to bring the scores nearer to 50%, the target-to-background ratio was reduced to $+2$ dB for these experiments.

F. Results

The scores were transformed into rationalized arcsine units (RAU; Studebaker, 1985). This was done to make the variance more uniform across conditions. Despite the training given, there was evidence that performance increased across successive conditions, i.e., a further learning effect occurred. To test for this, the scores for each subject were divided by the average score for that subject across the five processing conditions. Then, the score for each condition for a given subject was divided by the mean score for that condition across subjects. The resulting ‘‘normalized’’ data were then time ordered and the mean and standard deviation of the scores were calculated across all subjects, for each time slot. Using a *t*-test, we compared the means for the first and last conditions presented. There was a highly significant difference between the two ($t=3.52$, 38 df, $p<0.005$, one-tailed), implying that a learning effect was present.

The time-ordered averaged scores were expressed relative to the average score obtained in the last condition. The resulting relative scoring rates for the first four conditions tested (in time order) were 0.72, 0.91, 0.92, 0.90. The learning effects, if uncorrected, would make the data more ‘‘noisy,’’ and might obscure the effects of the different conditions. Hence, to correct for the learning effect, the scores for each of the first four conditions tested for each subject were divided by the relative score for that condition, e.g., the score for the second condition tested was divided by 0.91.

The solid line in Fig. 2 shows the mean learning-corrected word intelligibility scores, transformed back from RAUs into percentages, as a function of the attack time of the fast compressor. The mean corrected score for the envelope compressor is shown as a single cross with solid error bars at the right of the figure. Corresponding uncorrected scores are shown by dashed lines. The error bars show standard deviations (SDs) across subjects. The mean scores tend to improve with increasing attack time for the fast systems. The mean score is somewhat lower for the envelope compressor than for any of the fast systems.

The learning-corrected scores (in RAUs) were subjected to a one-way analysis of variance (ANOVA), with factor compression system. The ANOVA showed a significant effect of compression system; $F(4,76)=5.03$, $p=0.001$. Based on Fisher's least significant differences test, the difference between scores for the fast compressors for attack times of 8 and 2 ms was not significant (one-tailed *t*-test, $t=1.66$, 19 df, $p\cong 0.067$). The difference between the fast compressor with an attack time of 8 ms and all the remaining compression systems was significant at $p<0.025$ ($t>2.33$, 19 df, one-tailed). As expected, the score for the ‘‘envelope’’ compressor was significantly worse than for any of the fast compressors, $t\geqslant 1.83$, $p<0.05$ (one-tailed).

Although we were aiming to get overall scores around 50%, the scores were generally lower than this. The target-to-background ratio was lower than used in our earlier experiment, which had given scores around 70%–75% (Stone and Moore, 2003a). Evidently, we reduced the target-to-background ratio too much, possibly because we assumed a slope for the psychometric function that was based on experiments using unprocessed speech in noise. Also, psychometric functions for speech in fluctuating backgrounds, processed using a noise-vocoder simulation, can vary markedly in slope across subjects (Qin and Oxenham, 2003). The large error bars in Fig. 2 reflect a near-bimodal distribution in subject mean scores. Splitting the learning-corrected mean scores of each subject into bins of 5-RAU width, seven subjects scored 20–25, one scored 25–30, seven scored 35–40, and the remaining five scored 40 or more. We have not previously observed such a large spread in individual means in a nominally homogeneous group.

Despite the ‘‘noise’’ in the data, the results do show that decreasing the attack time from 8 to 0.5 or 0.125 ms, while holding modulation reduction constant, resulted in a significant reduction in intelligibility. Thus, it would seem desirable to avoid very fast attack times in compression circuits for cochlear implants. On the other hand, the envelope compressor, which had a longer attack time than the other compressors (and a shorter release time), but was more effective in reducing the modulation depth of the signal, produced the lowest intelligibility of all of the compressors. Thus, modulation reduction appears to be an important effect, separate from the effects of spectral spreading and comodulation of onsets resulting from the use of fast attack times.

III. EXPERIMENT 2: ASSESSING THE ROLES OF COMODULATION AND MODULATION REDUCTION

A. Introduction

The aim of this experiment was to assess the relative importance of comodulation and modulation reduction by applying wideband compression to the target and background, either before or after mixing. The fast compressors used in experiment 1 and by Stone and Moore (2003a) had attack times that were much smaller than the release times. Compressors with asymmetric attack and release times produce large distortion of the shape of the temporal envelope (Stone and Moore, 1992). This distortion on its own may have contributed to the reduction of intelligibility found by Stone and Moore (2003a) for the fast compressor relative to the slow compressor. Since we wished to investigate the role of modulation reduction rather than distortion of envelope shape, we decided to use a form of compression that produced minimal distortion of envelope shape. To do this, we used the envelope compressor of experiment 1, which had near-symmetric attack and release times.

In one condition, the compression was applied before the target and background signals were mixed. This would not be possible in a "real-world" environment, but it allowed us to apply modulation reduction to both signals while preserving their independence: comodulation should not then occur. We call this condition "INDEP." In a second condition, the compression was applied after mixing the target and background. This resulted in both modulation reduction and comodulation. We call this condition "COMOD." Each condition was tested using both a 6-channel and an 11-channel simulation; these channel numbers bracket the 8 channels used in experiment 1. On the basis of a pilot trial, target-to-background ratios of +8 and +3 dB were used for the 6- and 11-channel simulations, respectively, in order to achieve similar intelligibility. These ratios led to higher mean scores than for experiment 1.

B. Subjects and equipment

Eight volunteer subjects (4 males, 4 females, aged 19–31 years), all university undergraduates or graduates, were selected on the same basis as for experiment 1, except that one subject was a native speaker of American-, rather than British English. All subjects had previously been exposed to the processing, at times ranging from a few days to 18 months earlier. Subjects attended one session, which lasted just over 1 h. They were paid for their attendance. Since all the subjects had prior exposure to the ASL material, the BKB speech corpus (Bench and Bamford, 1979) was used. The sentences were spoken by a female speaker of British English. The background was based on a male speaker, but a different one from that used for experiment 1. The equipment and signal-processing method were the same as for experiment 1.

The input level of the target signal to the processing was 67 dB SPL (unweighted), and the presentation level of the mixed signal was 68 dB SPL for both conditions, as measured when the target and background were present at the same time. The envelope compressor had a compression ratio of 2.78 and the compression threshold was 55 dB SPL, the same as used in experiment 1.

C. Procedure

The training method for this experiment was almost the same as for experiment 1. The replay of the four IEEE sentence lists was again at increasing levels of difficulty. The second stage of training used the 32 sentences from BKB lists 1 and 2. Consecutive triads of lists drawn from BKB list numbers 3, 5, 6, 8, 9, 10, 11, 13, 14, 18, 19, and 21 were used for assessing each condition. Scores for each condition were the number of keywords correct from 33 sentences. The four conditions (INDEP versus COMOD and two channel numbers) were tested in a counterbalanced design.

D. Equating target-to-background ratios before and after compression

To allow a fair comparison of conditions INDEP and COMOD, it was necessary to ensure that the target-to-background ratios were comparable. In condition INDEP, the target and background were independently compressed, and then had to be mixed at the same target-to-background ratio as used for condition COMOD. This required measurement of the rms value of the uncompressed target and background for condition COMOD, and of the compressed target and background for condition INDEP. However, measurement of the rms value of compressed speech is not straightforward.

If the energy of *uncompressed* speech is estimated for short successive segments (frames), a histogram of the frame energies typically shows two peaks, a main peak defined by the speech energy and a peak at a lower level that represents the background noise in the recording. To measure the "real" speech energy, a measurement threshold needs to be selected (note that this threshold has nothing to do with the compression threshold used in the compressor). Only frames with energy above this measurement threshold are used in the calculation of the rms value. Choosing the measurement threshold to lie near the minimum of the valley, Stone and Moore (2003a) found that about 80% of 10-ms frames in continuous running speech were included in their measure of rms value. Compression alters the shape of the histogram of signal levels, which makes it more difficult to estimate an appropriate measurement threshold. Here, we chose the measurement threshold so that the percentage of 10-ms frames included in the measure of rms value was the same as for Stone and Moore, i.e., 80%. The measurement threshold was expressed relative to the level at the output of the compressor for a sinusoidal input with the same rms value as the input speech; this measure is denoted $\mathrm{rms_{sin}}$. For condition INDEP, the measurement threshold for the target was set to 5 dB below $\mathrm{rms_{sin}}$. For the background, which was at a lower input level than the target, relatively more of the signal fell below the compression threshold. To allow for this, the measurement threshold was reduced so as to meet the 80% criterion. The measurement threshold for the background was 7 dB below $\mathrm{rms_{sin}}$ for the +3-dB target-to-background ratio, and about 10 dB for the +8-dB target-to-background ratio.

E. Results

The scores were transformed into RAUs. Each data point was then normalized as described for experiment 1. The data were then time ordered and the mean and standard deviation of the score were calculated for each time slot. Mean scores across time-ordered conditions, expressed rela-

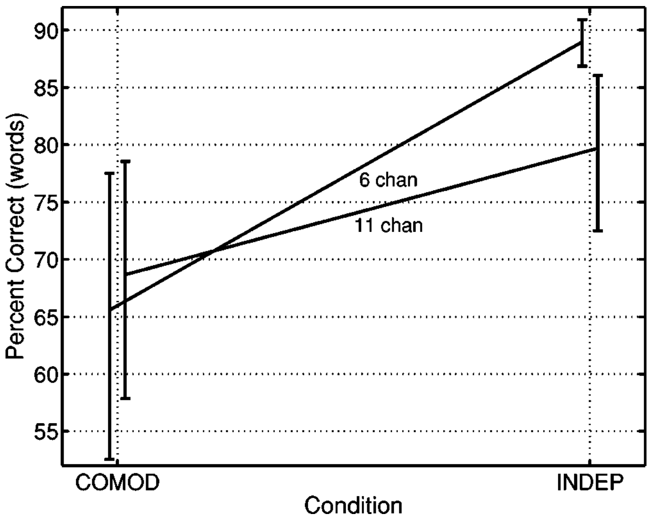

FIG. 3. Results of experiment 2 not corrected for learning effects (which were not significant for this experiment). Error bars show ±1 SD. The target-to-background ratio was 8 dB for the 6-channel system and 3 dB for the 11-channel system.

tive to the mean score for the last condition tested, were 1.03, 0.94, and 0.99. Using a *t*-test, we compared the means for the first and last conditions presented. The difference was not significant ($t=1.71$, 14 df, $p>0.1$, two-tailed). This was not surprising since the subjects were experienced with the processing. Consequently, no compensation was necessary for learning effects.

The mean scores, transformed from RAUs back into percentages, are plotted in Fig. 3. The error bars show SDs across subjects. For both channel numbers, intelligibility was higher for the INDEP condition, where the target and background were mixed after compression.

To assess the significance of the measured effects, an ANOVA was conducted with factors condition (INDEP or COMOD) and number of channels (6 or 11). One potential problem in comparing scores across channel numbers is that the signal-to-background ratio was different for the 6- and 11-channel systems. However, Friesen *et al.* (2001) have shown that psychometric functions for identifying sentences in steady background noise are roughly parallel for systems with different numbers of channels, for channel numbers from 4 to 20. We believe that this makes it reasonable to compare results across channel numbers. There was a significant effect of condition, $F(1,7)=20.67$, $p=0.003$, and a significant interaction between condition and number of channels, $F(1,7)=10.23$, $p=0.015$. The effect of condition was significant for both channel numbers ($t>2.41$, 7 df, $p<0.05$, two-tailed). There was no significant difference between the scores for the 6- and 11-channel systems in condition COMOD ($t=0.89$, 7 df, $p>0.4$), but there was a significant difference between the two channel numbers for the INDEP condition ($t=3.42$, 7 df, $p<0.02$).

The finding that performance was worse in condition COMOD than in condition INDEP supports the idea that at least part of the deleterious effect of fast-acting compression occurs because the compression introduces comodulation between the target and background. The finding that the difference between conditions COMOD and INDEP was larger for 6 channels than for 11 channels can be explained in the following way. In both conditions, the compressors acted on the wideband signal. When a compressed wideband signal is split into a number of frequency channels, the degree of modulation reduction of each channel signal (relative to what would be observed for an uncompressed wideband input signal) is less than that applied to the wideband signal; the channel signals are compressed less than the wideband signal (see experiment 3 and Table I, which is described later). The difference between the wideband modulation reduction and the channel modulation reduction increases with increasing number of channels. Less modulation reduction implies smaller changes in gain over time, which in turn implies less comodulation. This could explain why the difference between conditions INDEP and COMOD was smaller for the 11- than for the 6-channel system.

We conclude that comodulation does contribute significantly to the deleterious effect on intelligibility produced by fast compression, but the influence of comodulation decreases as the number of channels increases.

IV. EXPERIMENT 3: COMPRESSION OF THE WIDEBAND SIGNAL VERSUS COMPRESSION OF THE CHANNEL SIGNALS

A. Introduction

In some cochlear implant systems, the front-end (wideband) compression is followed by a second stage of fast-acting or instantaneous compression applied to individual channels, to map the channel signals into the current range between threshold and the most comfortable level (MCL) at each electrode (Clark *et al.*, 1990; Moore, 2003a; Zeng, 2004). This prevents uncomfortably loud levels (ULL) being reached, even allowing for loudness summation across electrodes. In other systems, all of the compression is performed by a front-end compressor (Eddington, 1983). An alternative configuration of the processing in a cochlear implant would be to have no front-end compressor, and to perform all of the necessary compression using fast-acting compressors within each channel. This approach has two desirable effects. First, comodulation effects would be more restricted in bandwidth. For example, a momentary peak in low-frequency energy of a background sound would lead to a reduction in gain, and a comodulation effect, only for low-frequency channels. Second, the spectral sidebands produced by rapid changes in gain in a given channel would have no influence on other channels. Drawbacks of this approach are that it would reduce modulation depth and intensity contrasts within each channel, *and* would reduce across-channel contrasts, which convey spectral information (Plomp, 1988; Drullman *et al.*, 1996). When a noise-vocoder implant simulator is used, which removes spectral fine structure, subjects may be especially dependent on across-channel contrasts. The same may be true of real cochlear implants.

The purpose of experiment 3 was to compare intelligibility obtained using a single-channel wideband compressor with that obtained using independent compressors within each channel of the implant simulation. For brevity, we refer to these as ‘‘wideband’’ compression and ‘‘channel’’ com-

TABLE I. Comparison of compression ratios for the wideband and channel compressors of experiment 3. Part A is for the 6-channel system, which used a target-to-background ratio of 5 dB. Part B is for the 11-channel system, which used a target-to-background ratio of 2 dB. The first line in each section shows the channel number. The second line shows the slope of the input–output function for each channel, measured as described in the text, for the system with compression applied to the wideband signal. The third line shows the corresponding effective compression ratio (the reciprocal of the slope). The fourth line shows the nominal compression ratio that was set for each channel of the system with independent channel compression, in order to match the effective compression ratio of the wideband compressor.

Condition	Channel number										
(A) 6-channel system			1	2	3	4	5	6			
Wideband slope			0.49	0.59	0.72	0.61	0.89	1.02			
Wideband effective CR			2.04	1.69	1.39	1.64	1.12	0.98			
Channel nominal CR			4.29	2.34	2.20	2.66	1.57	1.22			
(B) 11-channel system	1	2	3	4	5	6	7	8	9	10	11
Wideband slope	0.55	0.59	0.64	0.74	0.75	0.69	0.66	0.66	0.87	0.90	1.02
Wideband effective CR	1.82	1.69	1.56	1.35	1.33	1.45	1.52	1.52	1.15	1.11	0.98
Channel nominal CR	3.51	2.75	2.20	2.34	2.41	2.57	2.75	2.97	1.83	1.61	1.33

pression, respectively. The simulation used both 6 channels and 11 channels, with target-to-background ratios of +5 and +2 dB, respectively. In order to provide a fair comparison between the two compression arrangements, it was necessary to equate the effective amount of compression for the two. Our procedure for doing this is described next.

B. Equating the effect of a multichannel compression system to that of a single-channel compression system

To make the two conditions comparable, it was necessary to adjust the action of the independent channel compressors so that they produced a mapping of short-term signal levels between input and output similar to that achieved by the single-channel compression system. The steps involved in this are described below.

Two 90-sec-long speech files were generated comprising two male speakers mixed at the same target-to-background ratios as used in the experiment. The speakers were from our corpus of manually edited continuous speech (Stone and Moore, 2003a). First, the implant simulation was performed with no compression active. The level in successive 125-ms rectangular time windows (frames) was measured for each channel signal, prior to the stage of recombination. This provided a statistics file with which the effects of the compression systems were compared. Second, the implant simulation was performed with the single-channel wideband compressor active, and the same analysis was performed. The compressor was the same as one of the fast compressors of experiment 1, having attack and release times of 2 and 240 ms, respectively, a compression ratio of 7, and a compression threshold of 55 dB SPL. For each channel, a scatter plot was produced of the short-term level with compression versus the short-term level without compression (equivalent to a plot of output versus input). As is typical for compression systems (Elberling and Hansen, 1999), the points fell on a straight line (on a dB versus dB scale) only over a small range of levels. A linear regression over this range of levels was performed. Initially, frames were selected which lay in the 30-dB range below the maximum input level for a given channel. The mean and standard deviation of the levels in these frames were calculated. The linear regression was then performed over those frames lying between maximum channel level and the lesser of either 3 dB or 1 standard deviation below the channel mean level. The regression slopes provided a first approximation to the reciprocal of the compression ratio required for the independent channel compressors, so as to provide the same mapping of input to output as for the wideband compressor. These regression slopes are given by the second line of each section in Table I. The corresponding ‘‘effective’’ compression ratios are shown in line 3 of each section. The decrease of compression ratio with increasing frequency occurred because the gain of the wideband compressor was largely determined by the more intense low-frequency components in the input speech. This occurred even though, as described earlier, ‘‘pre-emphasis’’ was applied before wideband compression, so as to reduce the dominance of low frequencies in setting the gain.

Next, the implant simulation was performed with compression only in the individual channels. The channel compressors had the same attack and release times as the single-channel wideband compressor, but the compression thresholds for each compressor were set 13 dB below the rms value of each channel signal (for the wideband compressor, the compression threshold was 13–14 dB below the rms value of the target plus background at the input to the compressor, the exact value depending on target-to-background ratio). The channel rms value was calculated from the channel envelope signal, which was determined by full-wave rectification of the output of a given channel filter and low-pass filtering using a linear-phase filter with a response that was −3 dB at 50 Hz and −40 dB at 100 Hz (Stone and Moore, 2003a). The channel rms value was calculated in two stages: first, the rms value was calculated for all samples (at the sample rate of 16 kHz) of the channel envelope. Then, the rms value was recalculated using only those samples whose level exceeded the initial rms value−10 dB. Using the same 125-ms frame-selection and measurement criteria as before, a linear regression was performed on the scatter plot of output frame levels versus input frame levels. The channel compression ratios were adjusted iteratively until the regression slopes within each channel were the same for the multichannel compression system and the single-channel compression system. The required compression ratios were larger than the reciprocal of the regression slopes of each channel estimated for the wideband compression system (shown in line 3 of Table I), since the effective compression ratio achieved by the fast compressor was less than the nominal compression ratio (see Fig. 1). The nominal compression ratios needed to achieve the target values are shown in line 4 of each section of Table I.

C. Subjects, equipment, and stimuli

Eight volunteer subjects (3 males, 5 females, aged 20–35 years) were selected on the same basis as for experiment 1, except that three subjects were not native English speakers. These subjects were, however, fluent in English, and the pattern of their results did not differ markedly from that of the other subjects. No subject had participated in experiment 1 or 2, and none was familiar with the speech material or processing method. Subjects were paid for their attendance. The equipment and signal-processing methods were the same as for experiments 1 and 2. The ASL corpus was used. The input level of the target signal to the processing was 67 dB SPL (unweighted), and the presentation level of the mixed signal was 68 dB SPL.

D. Procedure

The four conditions (two arrangements of compression by two channel numbers) were tested in a counterbalanced design. The training procedure was almost identical to that for experiment 1.

E. Reduction of modulation spread in multichannel compression

In the introduction to experiment 1, we discussed how rapid changes in gain produce spectral sidebands around each signal component. In a multichannel acoustic hearing aid, such sidebands could fall outside the passband of the channel containing the signal component that gave rise to the sidebands, resulting in cross-channel effects. In an implant, the output of each channel is fed to a single electrode, so such cross-channel effects do not occur. This was implicitly simulated in our software since, after modulation of the noise carrier by the compressed channel signal, bandpass filtering was used to create the channel signals and to reinstate the long-term average spectral shape of the original speech spectrum plus the pre-emphasis (Stone and Moore, 2003a). The out-of-original channel sidebands were therefore prevented from interfering with adjacent channel signals in our simulation.

F. Results

The scores were transformed into RAUs. We assessed whether there was a learning effect using the same method as for experiments 1 and 2. There was a significant difference between average scores for the first and last conditions ($t=6.16$, 14 df, $p<0.0005$, one-tailed). The normalized scores for the successive conditions were 0.69, 0.88, and 0.95 times the score in the final condition. The change in scores over time is similar to that found for experiment 1. Scores for successive conditions were corrected as described for experiment 1.

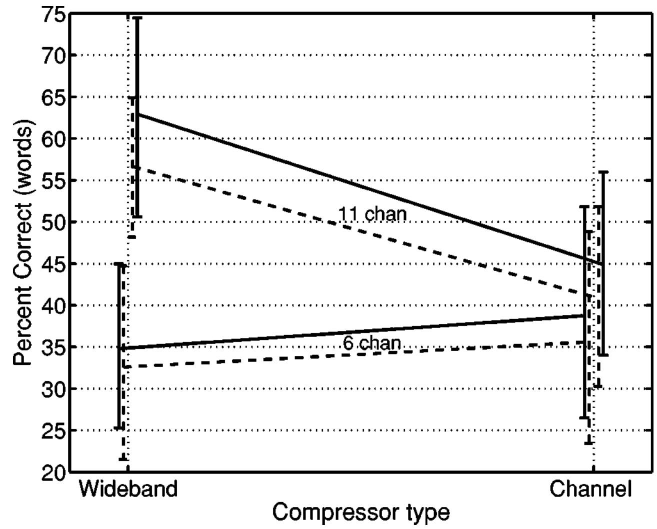

FIG. 4. Results of experiment 3. The solid and dashed lines show data corrected and uncorrected for learning effects, respectively. Error bars show ± 1 SD. The target-to-background ratio was 5 dB for the 6-channel system and 2 dB for the 11-channel system.

The corrected scores, transformed from RAUs back into percentages, are plotted for each compressor type in Fig. 4, as solid lines. Corresponding uncorrected scores are shown as dashed lines. The error bars show SDs across subjects. For the 11-channel system, channel compression led to markedly lower intelligibility than wideband compression. For the 6-channel system, intelligibility was similar for channel and wideband compression.

The corrected scores were subjected to an ANOVA with factors number of channels (6 or 11) and type of compression (wideband or channel). There was a significant effect of number of channels, $F(1,7)=96.0$, $p<0.001$, and a significant effect of compressor type, $F(1,7)=10.18$, $p=0.015$. There was also a significant interaction between number of channels and compressor type, $F(1,7)=38.72$, $p<0.001$. This interaction arose because, for the 6-channel system, there was no significant difference in intelligibility between compressor types ($t=1.44$, 7 df, $p>0.1$, two-tailed), while, for the 11-channel system, performance for wideband compression was significantly better than for channel compression ($t=6.40$, 7 df, $p<0.001$, two-tailed).

The lower intelligibility with channel than with wideband compression for the 11-channel system presumably occurred because of the reduction in across-channel spectral contrasts produced by the channel compression. It seems that this effect was more than enough to offset the potentially beneficial effect of channel compression in restricting the bandwidth of comodulation effects. However, for the 6-channel system, intelligibility was similar for channel and wideband compression. Perhaps this reflects a more even balance between the two opposing factors. With the 6-channel system, spectral information was represented in a highly quantized manner, and subjects may have relied more on within-channel patterns of amplitude modulation than on across-channel comparisons.

The score for the 6-channel wideband compressor is lower (35%) than the 55% found previously under similar test conditions (Stone and Moore, 2003a). This may be related to the use of a different background talker in the two studies. The background talker in the earlier study spoke with a high degree of amplitude modulation, possibly allowing more ''listening in the dips.''

V. DISCUSSION

A. Recommended attack times

The data presented here suggest that several factors can interact in a complex way to influence the efficacy of compression systems, when speech is processed so as to present primarily envelope cues in different frequency bands. We have established previously that fast-acting compression leads to lower intelligibility than slow-acting compression (Stone and Moore, 2003a). We have shown here that changes in the attack time of the fast compressor, while holding the fractional reduction in modulation constant, can also affect intelligibility. Attack times less than about 2 ms, typically associated with the need to avoid overload in electro-acoustic systems with limited dynamic range, are more deleterious than longer attack times. The deleterious effect of very short attack times can probably be attributed to the two factors discussed earlier, namely the spectral spreading and the onset comodulation produced by rapid changes in gain.

The results of experiment 1 did not reveal a ''threshold'' attack time below which intelligibility rapidly deteriorated. Rather, there was a small progressive deterioration as the attack time was decreased below 8 ms, with the biggest change occurring between 8 and 2 ms. We tentatively conclude that attack times should exceed ''several'' ms in order to avoid the worst effects of fast compression. Since the best performance was achieved for the longest attack time of 8 ms, the ''optimal'' attack time may well be longer than 8 ms.

B. Importance of preserving envelope shape

In experiment 1, we also assessed an envelope compressor which had an attack time of 10 ms, longer than that of the fast system with the longest attack time. However, the envelope compressor had a markedly shorter release time than any of the fast compressors. The near-symmetric attack and release times of this compressor, plus the fact that gain changes were applied to the delayed audio signal, meant that the compressor produced much less distortion of the envelope *shape* than the fast compressors. This might be expected to improve intelligibility. However, the envelope compressor was much more effective than the fast compressors at reducing modulation depth over the range of modulation rates important for speech perception. This might be expected to impair intelligibility. The results showed that the envelope compressor produced significantly lower intelligibility than any of the fast compressors, which suggests that, for this particular compressor, the deleterious effect of reduced envelope modulation outweighed any potential advantage from better preservation of envelope shape. However, it would be desirable to conduct further experiments manipulating the asymmetry of attack and release times while holding the fractional reduction in modulation as constant as possible. This would help in deciding whether preservation of envelope shape is important.

It should be noted that modulation reduction, as measured by the fractional reduction in modulation, depends on both the attack and release time (as well as compression ratio and threshold). It seems to us likely that modulation reduction is a more relevant and important factor than attack or release times alone.

C. Relative importance of modulation reduction and comodulation

Experiment 2 compared the effect of compressing the target and background before mixing (condition INDEP) and after mixing (condition COMOD). The two conditions introduced similar amounts of modulation reduction, but only condition COMOD introduced comodulation of the target and background. The results clearly showed poorer performance for condition COMOD, indicating that intelligibility can be reduced by comodulation of the target and background. The difference in intelligibility between conditions INDEP and COMOD was smaller for the 11-channel system than for the 6-channel system. This smaller change, despite the more adverse target-to-background ratio used, which would have led to greater comodulation, may have happened because increasing the number of channels in condition COMOD led to a decrease in the overall degree of comodulation of the target and background.

Experiment 3 compared the effect of wideband compression and channel compression. For the 11-channel system, performance was significantly worse for the latter, suggesting that reduction of across-channel contrasts, which convey spectral information, can produce a reduction in intelligibility, as argued by Plomp (1988). However, the effect was not apparent when 6 channels were used.

It seems clear that both modulation reduction and comodulation can lead to reduced intelligibility in a competing speech task. Comodulation appears to play a smaller role as the number of channels increases, while modulation reduction plays a greater role.

D. Hybrid fast and slow compression

A possible way of avoiding the deleterious effects of fast-acting compression described above is to use a two-stage compression process, similar to the DUAL-4 system described by Stone *et al.* (1999) for acoustic hearing aids. In that system, a slow-acting compressor, with a low-to-moderate compression ratio, was followed by fast-acting channel compression, again with low-to-moderate compression ratios. The slow-acting compressor reduced gross variations in signal level between environments so as to present a narrower range of levels to the fast-acting compressors. These compressors then mapped the remaining dynamic range of the signal into the reduced dynamic range of the impaired cochlea. For an implant, where the dynamic range that can be applied to a single electrode is very small, more compression is necessary. This ''extra'' compression should probably be incorporated in the slow-acting compressor, especially in implants with a large number of channels (greater than 6); otherwise, the deleterious effects associated with reduced within- and across-channel intensity contrasts, as revealed in our experiments, will play a significant role. There is at least one commercially available cochlear implant system that incorporates a slow-acting wideband compressor followed by fast-acting channel compressors.

E. Applicability to acoustic hearing aids for profound hearing loss

We have focused here on the effects of fast-acting compression on performance with a cochlear-implant simulator. However, the results may also have implications for the design of acoustic hearing aids for people with sensorineural hearing loss, especially severe to profound loss. Such people have reduced frequency selectivity (Pick *et al.*, 1977; Glasberg and Moore, 1986; Moore, 1998), and some are relatively insensitive to temporal fine structure cues (Rosen and Fourcin, 1986; Moore, 1998; Moore and Moore, 2003), but have a good ability to use temporal envelope cues (Bacon and Gleitman, 1992; Moore *et al.*, 1992; Turner *et al.*, 1995). Any form of signal processing that adversely affects the use of temporal envelope cues might therefore be expected to have a deleterious effect on performance.

F. Limitations of the simulation

The noise-vocoder stimulation of a cochlear implant used here has also been employed by many others. However, it should be noted that it does not adequately simulate all aspects of a cochlear implant system. With a real cochlear implant, the available dynamic range on each electrode is very small, while the simulation is used with normally hearing listeners who have a wide dynamic range. In an implant, an additional stage of fast-acting or instantaneous compression is nearly always applied to the signal for each electrode. Such a stage was not used in our simulation (or by others using similar simulations), because this would result in only a small portion of the dynamic range of the (normally hearing) listeners being used. Also, the use of instantaneous compression would lead to severe spectral spreading, which would mean that the ''channel'' signals were no longer confined to specific spectral regions.

A second limitation of the simulation is that, because the cutoff frequency of the envelope filters used in the simulation was relatively low (50 Hz), information about the fundamental frequency of the target and background talkers was largely removed. In a real cochlear implant, at least some information about fundamental frequency may be conveyed. Thus, it would be desirable to evaluate different forms of amplitude compression with real cochlear implant users, and not just with a simulation.

VI. CONCLUSIONS

Using a simulation of a cochlear implant assessed with normal-hearing listeners, we have explored the factors underlying the reduction in speech intelligibility produced by fast-acting compression, as demonstrated by Stone and Moore (2003a). The following are our main findings and conclusions.

(1) When the fractional reduction in modulation depth is held constant over a wide range of modulation rates, intelligibility decreases as the attack time of a fast compression system is decreased from 8 to 0.125 ms, keeping release times in the range 228–244 ms. This can probably be accounted for by the rapid changes in gain associated with fast attack times, which lead to two main effects: (a) correlated abrupt changes in amplitude of the target and background, which may inhibit their perceptual segregation; and (b) spectral artifacts associated with the modulation of components in the original signal.

(2) A fast-acting envelope compressor with an attack time of 10 ms and a release time of 13 ms led to lower intelligibility than any of the compressors described in (1) above, even though the envelope compressor produced less distortion of envelope shape and avoided most of effects (a) and (b) above. The poorer intelligibility produced by the envelope compressor can be attributed to the fact that it was much more effective than the compressors described in (1) in reducing the modulation depth of the input signal.

(3) Applying compression independently to the target and background, prior to mixing, led to higher intelligibility than when compression was applied after mixing. This indicates that comodulation of the target and background contributes to the reduced intelligibility produced by fast-acting compression. The effect was greater for a 6-channel simulation than for an 11-channel simulation, suggesting that the importance of comodulation decreases with increasing number of channels.

(4) For an 11-channel simulation, compression applied independently to the channel signals resulted in lower intelligibility than compression applied to the wideband signal, even though comodulation effects would have been smaller for the independent channel compression. For this number of channels, reduction of across-channel contrasts, which convey spectral information, was probably the dominant factor leading to poor performance for the channel compression. For the 6-channel simulation, performance was similar for wideband and channel compression, suggesting that the deleterious effects of reduced across-channel contrasts for the latter were offset by the reduction in comodulation effects.

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council (UK). We thank Tom Baer for comments on an earlier version of the manuscript. We also thank Andrew Faulkner and two anonymous reviewers for helpful comments.

Agnew, J., and Thornton, J. M. (**2000**). ''Just noticeable and objectionable group delays in digital hearing aids,'' J. Am. Acad. Audiol. **11**, 330–336.

ANSI (**1996**). ANSI S3.22-1996. ''Specification of hearing aid characteristics'' (American National Standards Institute, New York).

ANSI (**1997**). ANSI S3.5-1997. ''Methods for the calculation of the speech intelligibility index'' (American National Standards Institute, New York).

Bacon, S. P., and Gleitman, R. M. (**1992**). ''Modulation detection in subjects with relatively flat hearing losses,'' J. Speech Hear. Res. **35**, 642–653.

Baer, T., and Moore, B. C. J. (**1994**). ''Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech,'' J. Acoust. Soc. Am. **95**, 2277–2280.

Baer, T., Moore, B. C. J., and Gatehouse, S. (**1993**). ''Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality, and response times,'' J. Rehabil. Res. Dev. **30**, 49–72.

Bench, J., and Bamford, J. (**1979**). *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic, London).

Braida, L. D., Durlach, N. I., De Gennaro, S. V., Peterson, P. M., and Bustamante, D. K. (**1982**). ''Review of recent research on multiband amplitude compression for the hearing impaired,'' in *The Vanderbilt Hearing-Aid Report*, edited by G. A. Studebaker and F. H. Bess (Monographs in Contemporary Audiology, Upper Darby, PA).

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (Bradford, MIT Press, Cambridge, MA.).

Bregman, A. S., Abramson, J., Doehring, P., and Darwin, C. J. (**1985**). ''Spectral integration based on common amplitude modulation,'' Percept. Psychophys. **37**, 483–493.

Bustamante, D. K., and Braida, L. D. (**1987**). ''Multiband compression limiting for hearing-impaired listeners,'' J. Rehabil. Res. Dev. **24**, 149–160.

Carrell, T. D., and Opie, J. M. (**1992**). ''The effect of amplitude comodulation on auditory object formation in sentence perception,'' Percept. Psychophys. **52**, 437–445.

Clark, G. M., Tong, Y. C., and Patrick, J. F. (**1990**). *Cochlear Prostheses* (Churchill Livingston, Edinburgh).

Darwin, C. J., and Carlyon, R. P. (**1995**). ''Auditory grouping,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).

Dillon, H. (**1996**). ''Compression? Yes, but for low or high frequencies, for low or high intensities, and with what response times?'' Ear Hear. **17**, 287–307.

Drullman, R., Festen, J. M., and Houtgast, T. (**1996**). ''Effect of temporal modulation reduction on spectral contrasts in speech,'' J. Acoust. Soc. Am. **99**, 2358–2364.

Drullman, R., Festen, J. M., and Plomp, R. (**1994a**). ''Effect of reducing slow temporal modulations on speech reception,'' J. Acoust. Soc. Am. **95**, 2670–2680.

Drullman, R., Festen, J. M., and Plomp, R. (**1994b**). ''Effect of temporal envelope smearing on speech reception,'' J. Acoust. Soc. Am. **95**, 1053–1064.

Eddington, D. K. (**1983**). ''Speech recognition in deaf subjects with multichannel intracochlear electrodes,'' Ann. N.Y. Acad. Sci. **405**, 241–258.

Elberling, C., and Hansen, K. V. (**1999**). ''Hearing instruments—interaction with user preference,'' in *Auditory Models and Non-linear Hearing Instruments*, edited by A. N. Rasmussen, P. A. Osterhammel, T. Andersen, and T. Poulsen (Holmens Trykkeri, Copenhagen, Denmark).

Festen, J. M., and Plomp, R. (**1990**). ''Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing,'' J. Acoust. Soc. Am. **88**, 1725–1736.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). ''Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants,'' J. Acoust. Soc. Am. **110**, 1150–1163.

Glasberg, B. R., and Moore, B. C. J. (**1986**). ''Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments,'' J. Acoust. Soc. Am. **79**, 1020–1033.

Hafter, E. R., Bonnel, A. M., and Gallun, E. J. (**1996**). ''Detection of change without regard to its valence,'' J. Acoust. Soc. Am. **99**, 2541.

Hall, J. W., and Grose, J. H. (**1990**). ''Comodulation masking release and auditory grouping,'' J. Acoust. Soc. Am. **88**, 119–125.

Hall, J. W., and Grose, J. H. (**1991**). ''Some effects of auditory grouping factors on modulation detection interference (MDI),'' J. Acoust. Soc. Am. **90**, 3028–3035.

Hall, J. W., Grose, J. H., and Mendoza, L. (**1995**). ''Across-channel processes in masking,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).

IEEE (**1968**). *Recommended Practice for Speech Quality Measurements* (Institute of Electrical and Electronics Engineers, New York).

MacLeod, A., and Summerfield, Q. (**1990**). ''A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use,'' Br. J. Audiol. **24**, 29–43.

Macmillan, N. A. (**1973**). ''Detection and recognition of intensity changes in tone and noise: The detection-recognition disparity,'' Percept. Psychophys. **13**, 65–75.

Moore, B. C. J. (**1990**). ''How much do we gain by gain control in hearing aids?'' Acta Oto-Laryngol., Suppl. **469**, 250–256.

Moore, B. C. J. (**1992**). ''Across-channel processes in auditory masking,'' J. Acoust. Soc. Jpn. (E) **13**, 25–37.

Moore, B. C. J. (**1998**). *Cochlear Hearing Loss* (Whurr, London).

Moore, B. C. J. (**2003a**). ''Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants,'' Otol. Neurotol. **24**, 243–254.

Moore, B. C. J. (**2003b**). ''Speech processing for the hearing-impaired: Successes, failures, and implications for speech mechanisms,'' Speech Commun. **41**, 81–91.

Moore, B. C. J., and Moore, G. A. (**2003**). ''Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects,'' Hear. Res. **182**, 153–163.

Moore, B. C. J., Peters, R. W., and Stone, M. A. (**1999**). ''Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips,'' J. Acoust. Soc. Am. **105**, 400–411.

Moore, B. C. J., and Shailer, M. J. (**1992**). ''Modulation discrimination interference and auditory grouping,'' Philos. Trans. R. Soc. London, Ser. B **336**, 339–346.

Moore, B. C. J., Shailer, M. J., and Black, M. J. (**1993**). ''Dichotic interference effects in gap detection,'' J. Acoust. Soc. Am. **93**, 2130–2133.

Moore, B. C. J., Shailer, M. J., and Schooneveldt, G. P. (**1992**). ''Temporal modulation transfer functions for bandlimited noise in subjects with cochlear hearing loss,'' Br. J. Audiol. **26**, 229–237.

Pick, G., Evans, E. F., and Wilson, J. P. (**1977**). ''Frequency resolution in patients with hearing loss of cochlear origin,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).

Plomp, R. (**1983**). ''The role of modulation in hearing,'' in *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer, Berlin).

Plomp, R. (**1988**). ''The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function,'' J. Acoust. Soc. Am. **83**, 2322–2327.

Plomp, R. (**1994**). ''Noise, amplification, and compression: Considerations of three main issues in hearing aid design,'' Ear Hear. **15**, 2–12.

Qin, M. K., and Oxenham, A. J. (**2003**). ''Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers,'' J. Acoust. Soc. Am. **114**, 446–454.

Rasch, R. A. (**1978**). ''The perception of simultaneous notes such as in polyphonic music,'' Acustica **40**, 21–33.

Robinson, C. E., and Huntington, D. A. (**1973**). ''The intelligibility of speech processed by delayed long-term averaged compression amplification,'' J. Acoust. Soc. Am. **54**, 314.

Rosen, S., and Fourcin, A. (**1986**). ''Frequency selectivity and the perception of speech,'' in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London).

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.

Stone, M. A., and Moore, B. C. J. (**1992**). ''Syllabic compression: Effective compression ratios for signals modulated at different rates,'' Br. J. Audiol. **26**, 351–361.

Stone, M. A., and Moore, B. C. J. (**2002**). ''Tolerable hearing-aid delays. II. Estimation of limits imposed during speech production,'' Ear Hear. **23**, 325–338.

Stone, M. A., and Moore, B. C. J. (**2003a**). ''Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task,'' J. Acoust. Soc. Am. **114**, 1023–1034.

Stone, M. A., and Moore, B. C. J. (**2003b**). ''Tolerable hearing-aid delays. III. Effects on speech production and perception of across-frequency variation in delay,'' Ear Hear. **24**, 175–183.

Stone, M. A., Moore, B. C. J., Alcántara, J. I., and Glasberg, B. R. (**1999**).

''Comparison of different forms of compression using wearable digital hearing aids,'' J. Acoust. Soc. Am. **106**, 3603–3619.
Studebaker, G. (**1985**). ''A rationalized arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.
Summerfield, Q., and Culling, J. F. (**1992**). ''Auditory segregation of competing voices: Absence of effects of FM or AM coherence,'' Philos. Trans. R. Soc. London, Ser. B **336**, 357–366.
Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). ''Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **97**, 2568–2576.
Verschuure, J., Maas, A. J. J., Stikvoort, E., de Jong, R. M., Goedegebure, A., and Dreschler, W. A. (**1996**). ''Compression and its effect on the speech signal,'' Ear Hear. **17**, 162–175.
Villchur, E. (**1973**). ''Signal processing to improve speech intelligibility in perceptive deafness,'' J. Acoust. Soc. Am. **53**, 1646–1657.
Zeng, F.-G. (**2004**). ''Compression and cochlear implants,'' in *Compression: From Cochlea to Cochlear Implants*, edited by S. P. Bacon, A. N. Popper, and R. R. Fay (Springer, New York).