Veterans Administration

Journal of Rehabilitation Research and Development Vol. 24 No. 4
Pages 193–206

Two-state compression of spectral tilt: Individual differences and psychoacoustical limitations to the benefit from compression

MARK P. HAGGARD,* JULIAN R. TRINDER, JOHN R. FOSTER, and ANN-CATHRINE LINDBLAD**

MRC Institute of Hearing Research, University of Nottingham, Nottingham, United Kingdom, and Department of Technical Audiology, Karolinska Institutet, Stockholm, Sweden

Abstract—A psychoacoustic rationale was developed for a hearing aid design in which compression of spectral tilt was incorporated without any instantaneous nonlinear distortion. This involved switching between a 'flat' and a 'rising' frequency response; the switching was done slowly to avoid audible transients and was controlled by feedback derived from comparison of output levels in low- and high-frequency channels, approximating voiced/unvoiced detection. The effect of this switching process was to narrow the distribution of spectral tilt values compared with the input. Asynchrony between the switching and the triggering speech structures was avoided by also delaying the signal path. Unfortunately, hearing-impaired listeners performed more poorly on the switching system than on either of the control 'flat' or 'rising' frequency-responses. An explanation is offered (on the basis of growing evidence from perceptual experiments) of the perceptual importance of temporal envelope contours within individual frequency bands. It was possible, in part, to predict individuals' results in the switching condition from age and audiometric or psychoacoustic characteristics. The results suggest a modification to the switching design, and they point to an intrinsic limit to the ability of all hearing aids of the compression type to enhance intelligibility.

*Address correspondence to: Mark P. Haggard, MRC Institute of Hearing Research, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom.

**Ann-Cathrine Lindblad's work was supported by the Swedish National Board for Technical Development.

INTRODUCTION

Hearing-impaired persons whose disorder is of cochlear origin suffer impairment of frequency resolution and temporal resolution as well as of absolute sensitivity. Amplification is therefore, at best, a partial solution, and those concerned with rehabilitation have turned their attention to further aspects of signal-processing. While not able directly to reverse externally the internally degrading transformation (4), the signal-processing may, in theory at least, optimally match important properties of the (speech) signal to residual capacity (10). Various bases for such matching have been suggested.

In most cochlear hearing losses the raised threshold is not accompanied by a correspondingly raised loudness discomfort level; thus there is a restriction of the sensory dynamic range. For enhancing the effective dynamic range, the most investigated signal-processing principle to date is that of amplitude compression. Compression should, in comparison with linear amplification, place more of the time-varying speech spectrum sufficiently above threshold for it to be discriminated, but without generating loudness discomfort (we refer to this approach as the first therapeutic hypothesis). This applies both in general and to a particular broad frequency region. Circumstantial but widespread clinical experience suggests that compression of various types can

facilitate acceptance of an aid in individual cases with limited sensory dynamic range, while not in general leading to improved performance in word identification tasks (20), even when the sophistication of multi-band compression is introduced (2,3,17). The overall degree of compression required, totalled across the spectrum, may need to be moderated because of associated problems such as annoying amplification of background noise when speech is absent. Partly for this reason, several investigators have employed multiband compression. This is based on the reasonable assumption, stated for example by Villchur (25), that the local degree of compression should be tailored to the degree of hearing loss not only across patients, but also across frequency bands within a single patient. Evidence for any generally higher performance with such sophisticated systems is, however, marginal, especially when tested on clinical test materials that vary little in level (17), as is customary for diagnostic purposes. Another possible reason is the tendency to diminish short-term spectral differences (10,26). The specific difficulty in showing advantages for compression on performance measures is additional to the general difficulty of showing differences between aid designs with intelligibility-type tests (22). The benefit obtained from compression appears to depend upon the degree of amplitude fluctuation in the test material and on other particular circumstances in the material and the individual (20,3).

One reason for the variability of results could be a delicate tradeoff underlying the possibility of benefits from compression. All forms of compression tend to reduce the acoustical differences between classes of sound and as a result tend to make them less discriminable and identifiable. The first therapeutic hypothesis therefore relies on the assumption that gross differences between the more discriminable classes can sustain a reduction in discriminability in terms of intensity or spectral tilt; that the loss due to errors in this respect will be smaller or less important than the improvement in within-class discrimination achieved by making the general conditions more favorable for discrimination of fine acoustical differences. In circumstances of such a delicate tradeoff, one would expect large individual differences in response to any one set of parameter settings for a given compression scheme and hence no overall significant advantage. Where there are reliable individual differences but no overall mean advantage, one would further expect that a subset of hearing-impaired listeners could be delimited as appropriate candidates for the scheme and that procedures for custom-fitting the electroacoustic parameters could be developed.

Compression could assist tolerance of rising frequency-responses appropriate to optimise performance in cases of markedly sloping audiograms (the second therapeutic hypothesis). Haggard and Trinder (13) developed a version of this hypothesis, reconciling demands of frequency response with dynamic range on the basis of the following formulation of the dilemma for optimising the frequency response of a linear amplifier. In hearing impairment, effective upward spread of masking results from the predominance of low frequencies in speech, coupled with degraded frequency resolution. For flat and gradually-sloping losses, the optimum frequency response for word identification performance will therefore be one that pre-emphasizes high frequencies. However, a steeply rising response will occasionally cause sounds with intense high-frequency components, such as /s/, to exceed the top of the dynamic range and cause discomfort, leading to poor acceptance of such pre-emphasising aids (24).

Avoidance of this problem by using overall lower gain settings or flattening the frequency response will, at the other extreme, again depress performance. There must be a plateau of optimally sloping frequency-responses between the extremes mentioned. One can escape from this dilemma by postulating that a rising response will be appropriate for a natural subclass of sounds with a falling long-term spectrum, and that a flat response will be more appropriate for a subclass having a flat spectrum. We may consider the source characteristics of voiced and voiceless sounds, falling and flat respectively, as satisfying the criteria for those subclasses (5). If we further regard the spectral effects of articulatory vocal tract modifications as averaging out over the long term, and if we cumulate over the continuous changes in vocal tract configuration for voiced and voiceless portions separately, then two separate long-term spectral slopes will be obtained, similar to those of the corresponding energy sources.

Figure 1 summarizes the result of an illustrative analysis of a short sample of one speaker's speech that is consistent with this subdivision of short-term amplitude spectra according to their source characteristics. The whispered and fricated sounds jointly

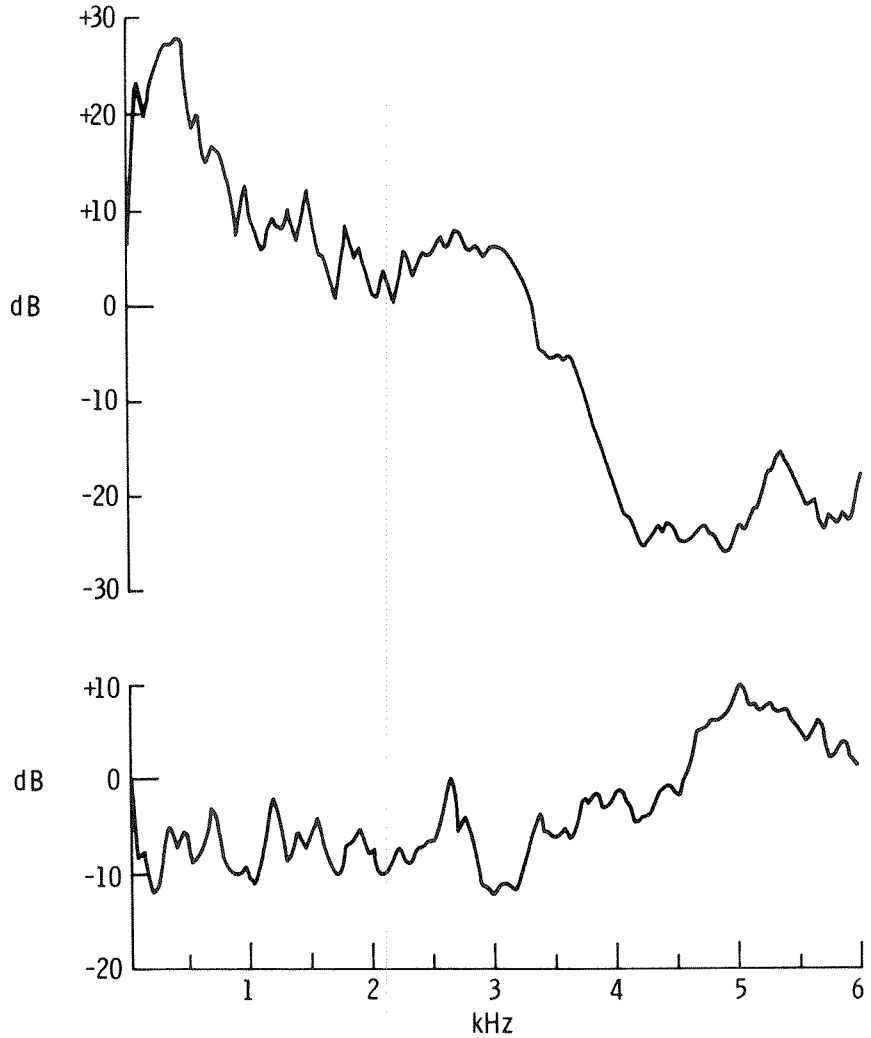

Figure 1.
Illustrative analysis of the averaged spectra of sentences from a single speaker: (a) "Where were you a year ago? We were away in Wallawalla". (voiced); (b) "Sixteen psychiatrists visited the Institute yesterday" (whispered). The first contains only phonemically voiced sounds and was produced with normal voice; the second contains many voiceless plosives and fricatives and was furthermore produced with whisper in place of voice.

have a long-term spectrum much flatter than that of the voiced sounds.

To deal differently with the two subclasses, Haggard and Trinder (12,13) proposed an aid embodying a speech-driven alternation between a rising and a flat characteristic; if the two natural classes of speech event were gross enough, for example approximating the voiced and voiceless sounds, then a favorable tradeoff of improvement in within-class discrimination (for example, the place-of-articulation feature) should occur with only small deterioration in between-class (voicing) discrimination. This gives a way of implementing a long-term statistical nonlinearity of response, while avoiding having signal paths that are nonlinear in the short term, with their many associated problems of nonlinear distortion: adjustment, specification, calibration, and optimization of time constants. A main problem in the controlled comparisons of hearing aid designs is that the characteristics of nonlinear systems, even those such as AGC, involving nonlinearity with 'memory' rather than instantaneous distortion, are intrinsically difficult to specify. There is therefore a further incentive to seek the main benefits hypothesized for compression by slightly different and more specifiable means.

Simplicity is also desirable. A main obstacle for compression aids is their behavior in noise when dynamics that may be inappropriate for the speech are imposed in response to the noise. However, a well-controlled study on an aid of basically high quality (15) has shown favorable results for a two-

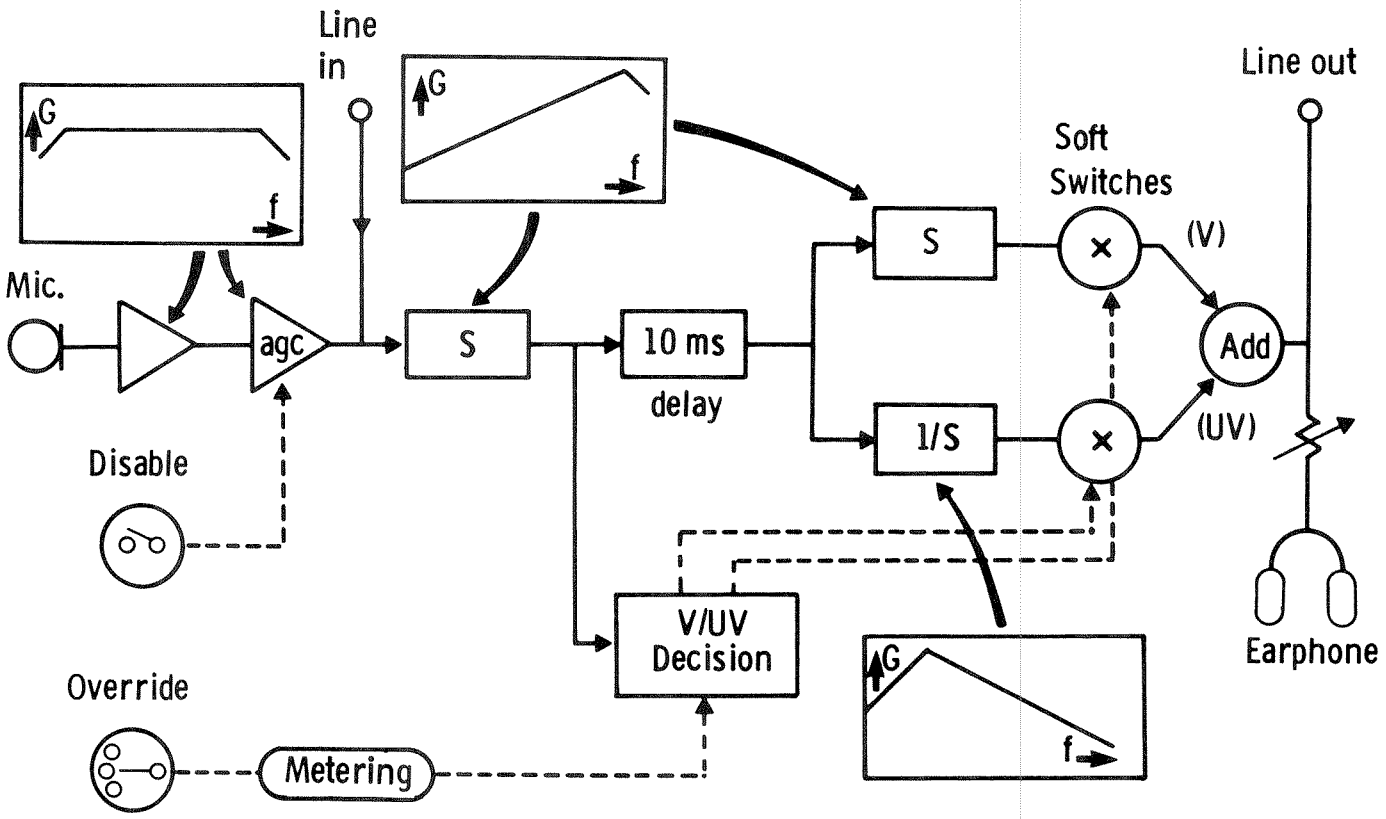

Figure 2.
Design of the Switching Hearing Aid. For details see text, also Haggard and Trinder (12,13).

channel compression device, even in mildly noisy conditions. This encourages further experimentation with the parameters of compression schemes. The complexity of such signal-processing aids and the resulting bulk of their circuitry involved remains sufficient (even with prospective implementation by purpose-built integrated circuits) to constitute a slight cosmetic barrier. The claimed extra benefits[1] would have to outweigh this cosmetic cost as well as financial cost, in relation to what a conventional aid offers. Thus, for the present, simplicity of circuitry is also a design constraint in achieving the hypothesized benefits from compression, for reasons both of cost and specifiability.

This paper reports some results with a prototype of the Haggard-Trinder aid (12,13) which serve to spotlight a possible psychoacoustic limitation to the benefit available from all compression schemes, whether based upon instantaneous waveform nonlinearity negative gain feedback (AGC), or upon more sophisticated digital schemes that would minimize the disadvantages of the former. The conditions tested employ an input signal-to-noise (S/N) ratio of +5 dB which is somewhat adverse for hearing-impaired listeners and avoids ceiling effects in performance measures. Such a noise level also represents a stringent test of a main problem with compression schemes.

METHOD

Signal-processing and parameter adjustment

The aid design, shown schematically in **Figure 2**, incorporates soft switching (i.e., switching gradually) between two channels, one with a nominally 'flat' and one with a 'rising' frequency response, +9 dB/octave in the implementation used. [Further details are given in (12) and (13).] The switching is determined by the ratio of energy in the band up to 1 kHz to that in a 3 kHz to 6 kHz band. This is

relatively simple and crude by modern standards for design of detectors for the phonetic feature voiced/unvoiced, but it may also be seen more generally as a spectral tilt measure for supplying a binary tilt correction (by AGC with negative feedback) so as to flatten the short-term spectrum. The device is thus in some respects a two-state approximation to a frequency-dependent compressor, or two-band compressor, except that in compressing spectral tilt it exerts only minor compression on the total amplitude envelope. A separate overall AGC stage can, of course, be incorporated.

Hysteresis and nonlinear smoothing are incorporated into the switching decision to allow a more valid decision; this causes approximately 10 ms delay in the switching control signal, which is then resynchronized with the waveform by also delaying the waveform by 10 ms with a charge-coupled delay device. In this way instantaneous waveform distortion and the undesirable side-effects of feedback time constants are both avoided. Switching transients are avoided by using a 'slow' switching with a time constant of approximately 5 ms. Such a design is particularly relevant to digital implementation, not merely because the compression is achieved in binary fashion, but because the control functions lend themselves to simple programming. The two audio filter functions postulated (or any other small set) could even be implemented digitally via stored filter impulse responses, while keeping the processing demand within bounds that can be foreseen as feasible within the near future. (The prototype used here was, however, implemented in analog.)

Suitable tilt thresholds for the switching action were judged on listening with continuous speech. In the presence of only the speech-shaped masking noise used in the tests, the 'switching' condition selected the 'rising' frequency response, as required in the design for speech approximating the long-term average spectrum. The relative gains of the 'flat' channel and the 'rising' channel were adjusted for clarity and comfort in pilot listening to continuous speech above white noise at 85 dB SPL. These conditions only partly simulate those obtaining in the severely impaired: however, given the assumption of similar uncomfortable loudness levels for the normally-hearing judges and impaired listeners, the adjustment will tend to conservatism concerning the tolerability of the high levels at high frequencies generated by the +9 dB/oct condition. In application the relative level could be an individually adjustable parameter, but control was desired here to obtain an overall view of the value of the scheme without introducing too many covariates. The balance of the two frequency responses so obtained entailed an equal gain (and hence a spectral pivot point) at 1.7 kHz. This is approximately the frequency region below and above which equal statistical contributions to monosyllable intelligibility are made according to French and Steinberg (9). The two linear characteristics were also tested separately as control conditions.

Test materials

Forced-choice word identification tests based upon consonant features can offer measures somewhat specific to spectro-temporal processing, at the expense of general representativeness. More important, they offer tightly-controlled ways of measuring small differences in the efficacy of this processing. For the problem in hand, it is particularly required that test materials be sensitive to the vulnerable features in middle and high frequencies, which are materially affected in moderate hearing impairments. (Low-frequency prosodic cues of intonation and amplitude contour are in any case transmitted by most hearing aids and by all auditory systems except the most profoundly impaired.) The FAAF test (Four Alternative Auditory Feature Test; see Foster and Haggard (6,7,8)) has documented linguistic and psychoacoustic properties. In the FAAF test the 80 stimulus words comprise each word drawn from 20 four-word vocabularies; they are embedded in a carrier phrase and were used here with a 10-second intertrial interval. The vocabulary is composed on the rhyme test principle by sampling two values of each of two articulatory/acoustic features of consonants that distinguish minimally-paired real words in terms of the phoneme at a single syllabic position (e.g. mail, bail, nail, dale). Reference experiments on normally hearing subjects with highpass, lowpass, bandpass and bandstop conditions have further permitted the definition of indices of the use of low-, mid-, and high-frequency information (CVL, CVM, CVH) from subscores of detailed directional error patterns (7). The frequency bands used in the reference experiments, and defining what we here denote as 'low', 'mid' and 'high'

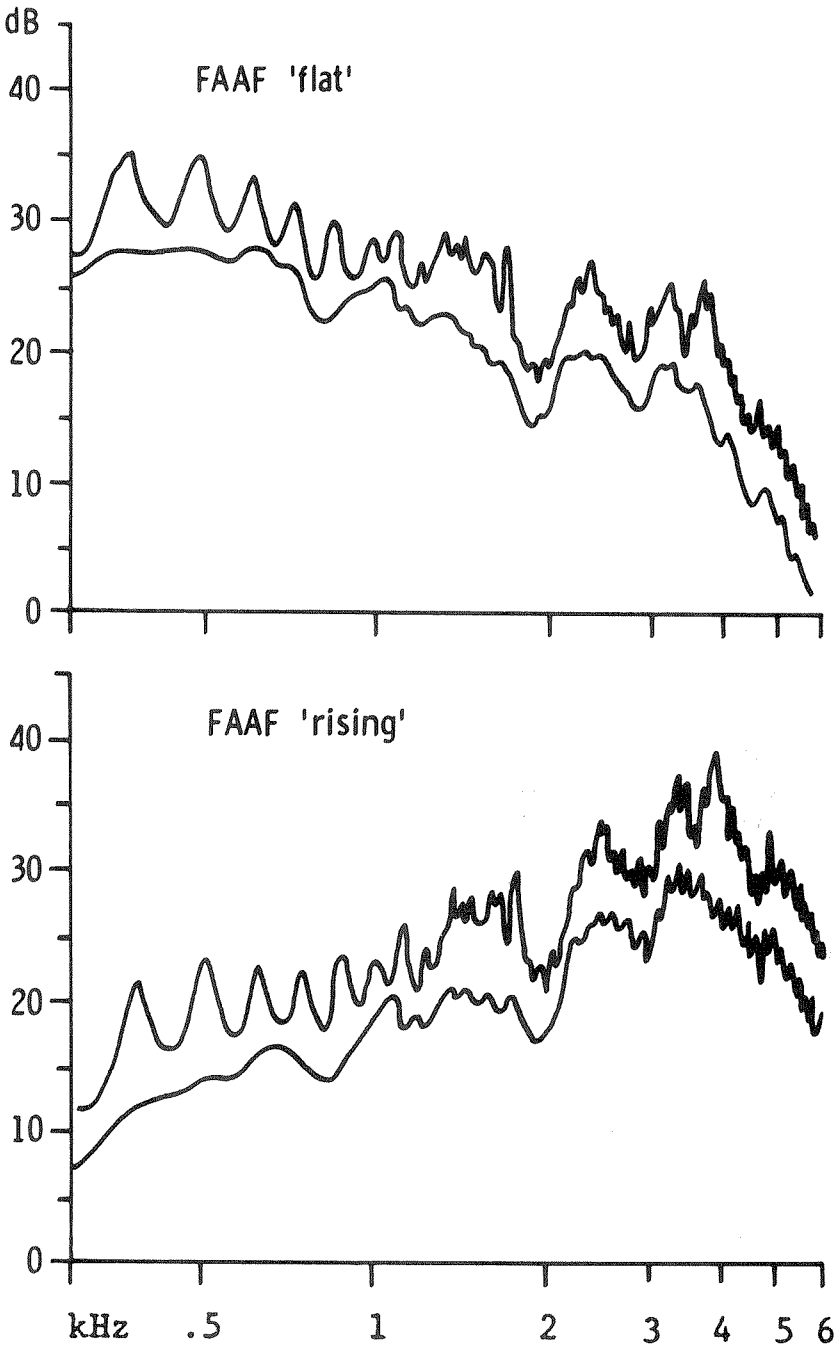

Figure 3.
Long Term Spectra of Test Materials. Each graph shows the 'long-term' spectrum levels for samples of 10 test items plus noise and for noise alone. Samples were of approximately 5 s duration. The FAAF test was masked by speech-spectrum-shaped noise modelled on continous speech of the same talker. The nominal signal-to-noise ratio before experimental filtering was +5 dB, and this holds also for the levels in the spectral bands.

frequencies, are 0.1 to 1.3 kHz, 1.6 to 2.4 kHz and 3.6 to 6.0 kHz, respectively. The scale values of the subscores are not in percentage units but in standard deviation units of the distributions on the discriminant function (selected and weighted from the raw feature errors) that best differentiates conditions in the reference data containing and not containing the band in question.

Electroacoustic conditions

The calibration level of each test is defined as the mean of the peak rms levels (fast meter-setting) of the test words in A-weighted SPL, henceforth denoted dB(A). White noise filtered to give the long-term spectrum of the same talker was added to the FAAF materials. Tapes were recorded with a signal-to-noise ratio of +5 dB (all measurements quoted

are in dB(A) prior to processing). This was done by monitoring levels from TDH 39P 100-ohm headphones with MX-41/AR cushions in a Bruel and Kjaer wideband artificial ear (type 4153) with a microphone (type 4134). Test tapes were then prepared by recording the speech plus noise after processing through the prototype aid, either fixed in its 'flat' or 'rising' mode or in its 'switching' mode. The spectra resulting are shown in **Figure 3**.

Experimental stimuli for this test and for an audiovisual test (not reported here) were recorded on videotape to minimize calibration problems with presentation conditions. At playback a Sony Videorecorder VO-1810 fed a Marantz 1072 amplifier and Jessop attenuator driving TDH 39P 100-ohm headphones with MX-41/AR cushions for monaural listening. Testing was conducted in a quiet room in a hospital annex.

Subjects

The test subjects were obtained from records and follow-up visits at a National Health Service clinic in a middle-sized English country town. All but one had been fitted with their first hearing aid within the last year (median 3 months). Twenty-eight subjects were tested, 12 males and 16 females, with sensorineural hearing loss unselected as to severity, asymmetry, aetiology or audiogram shape: The mean air-conduction hearing loss (0.5–4.0 kHz) in the usually aided ear was 49 dB and ranged from 35 to 73 dB. The mean air-bone gap (0.5–2.0 kHz) was in 21 cases less than 10 dB. In 7 cases, however, the mean air-bone gap exceeded 10 dB, ranging from 11 to 17 dB. The conductive component in these losses was considered marginal for the purposes of this study, and these subjects were included as being mainly sensorineural. The ages ranged from 42 to 69 years with a median age of 64 years.

Design and procedure

The three speech test conditions (two frequency responses plus one 'switching' condition) were embedded in a design counterbalanced for order across subjects, but which included further conditions not discussed here (14). A different ordering of the test items was used for each condition, although response-sequence learning is in fact minimal with these tests because of their forced-choice format. A freefield test with the subject's own aid, involving only 40 trials, also was included mostly for motivational reasons as the comparison with a real hearing aid cannot be adequately controlled electroacoustically to be informative. There the speech level was set to 70 dB SPL and the aid was adjusted by the subject for a comfortable listening level as in real life.

The FAAF test words were heard in each of three conditions, counterbalanced for order: 'flat'; 'rising' and 'switching' (between the other two characteristics). The other tests and the questionnaire occurred in the fixed order shown in **Table 1**, not being part of a counterbalanced experimental design. They were interleaved with the components of the design to avoid boredom in the subjects. The single test (80 words) in each condition allows no scope for learning and hence gives an unfavorable evaluation of any new form of signal-processing. However, relative performances with little prior experience are of some importance because they determine whether an aid will be accepted and used subsequently.

Simplified clinical measures of frequency resolution were taken at 2 kHz and audiograms were available. The measure of frequency resolution approximated the psychoacoustical tuning curve (PTC)

Table 1.
Test complement for the study.

Uncomfortable loudness level	for running speech for each frequency response: 'flat' and 'rising' (and 'switching').
Earphone audio tests	(FAAF) 80 trials × 2 frequency responses.
Earphone audio test	(FAAF) 80 trials 'switching' condition.
Pure tone audiometry	bone conduction at 0.5, 1.0, and 2.0 kHz; air conduction at 7 frequencies.
"Freefield" own-aid test	(FAAF) 40 trials.
Hearing aid questionnaire	questions on impairment, handicap, disability, benefit and use of the aid.
Electroacoustic check	hearing aid function, gain, and distortion.
Psychoacoustic tuning curves	levels of 1/15 octave noises at 1.25, 2.0 and 2.5 kHz which just mask a 2.0 kHz pulsed probe tone at 10 dB SL.

as described by Zwicker and Schorn (27) and simplified for clinical application by Lutman and Wood (19). This involves finding the levels of 1/15 octave narrow-band maskers at 1.25, 2.0, and 2.5 kHz that just mask a 10 dB SL tone at 2 kHz. The procedure furnished three measures, denoted PTCL, PTCH and PTCE. The PTCL and PTCH measures reflect frequency resolution and are respectively the differences between the intensities required of the narrow band maskers at 1.25 kHz or at 2.50 kHz and the intensity required of a narrow band noise at 2.0 kHz to just mask the 2.0 kHz tone at 10 dB SL. PTCE is the difference between the absolute and the masked threshold for the 10 dB SL tone, i.e., a measure of the amount of simultaneous masking from a noise all of whose energy would contribute to masking in an ear with normal frequency resolution. As there is no extra energy on the skirts of the auditory filter to be differentially incorporated according to the filter's width, this measure is thus an efficiency measure; it may be equated with the non-resolution component that contributes to the Critical Ratio, and presumably has a partly central basis.

The usually aided ear of each subject was used. Uncomfortable loudness level (ULL) judgments were made before the first test with each frequency response on a passage of running speech set to the same calibration level as the test stimuli. Instructions emphasized that a level just tolerable in the long-term should be set; it was not a short-term endurance task. The presentation levels were set initially according to the rule: ULL − 10 dB. In the 'rising' condition this gave a level that was reported in six cases as not consistent with comfort and ease of listening. Achieving the highest comfortable listening level (HCL) involved an increase by 5 dB in four of these cases and a decrease by 5 dB in two. For two of these subjects a parallel modification of the rule was required in the 'flat' condition also. Responses were made on printed sheets by circling the response word.

RESULTS

Differences between conditions

The top line of data in **Table 2** shows the mean and standard deviation of the percentage identification scores. A one-way analysis of variance on the three main conditions confirmed ($F_{(2,54)} = 63.72$; $p<0.001$), that the conditions differ significantly. Incorporating listening level as a covariate in the corresponding analysis of covariance did not alter this result, so differences are probably not due to condition-linked gain adjustments. All paired differences, bar that between 'flat' and 'rising' are highly significant. The 'switching' condition is on average worse than the 'flat' or 'rising' conditions of which it is composed.

In analyzing and optimizing the effects of signal-processing, the particular classes of feature distinctions affected are of as much interest as the global benefit in percent correct, as they can indicate the confirmation or otherwise of the therapeutic hypotheses (3) and the possible direction of optimization of parameters of the processing. The FAAF test offers two ways of doing this, via traditional phonetic features of place, manner, and voicing, and via frequency-specific subscores. As a corollary of the first therapeutic hypothesis, the difference between predominantly high- and low-frequency sounds is reduced by the tilt compression and hence the voicing distinction should be degraded, while place distinctions within classes of voicing or manner should improve or at least deteriorate insignificantly. Manner distinctions are too diverse to predict for, as some could display either tendency.

The middle field of results in **Table 2** gives the percentages of errors in the three features classes for the three processing conditions. Contrary to the

Table 2.
FAAF test speech identification scores by conditions (N = 28).

Condition	'Flat'		'Rising'		'Switching'	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Overall % correct	79.1	8.3	77.6	7.0	66.4	7.6
PLACE errors	12.0	5.7	11.4	4.1	16.8	6.7
MANNER errors	7.4	3.6	7.8	4.4	12.5	5.5
VOICING errors	7.2	5.1	10.9	6.2	11.8	7.8
CVL: use of low frequency	2.78	1.71	0.35	2.77	-0.58	2.75
CVM: use of mid frequency	3.43	1.65	3.68	1.67	0.60	2.79
CVH: use of high frequency	3.90	2.30	5.42	1.24	3.14	3.00

hypothesis, the general degradation in the 'switching' condition with respect to the 'flat' and 'rising' conditions was reflected in the significantly greater numbers of errors in distinctions of place ($t=5.82$, $p<0.001$; $t=5.40$, $p<0.001$ respectively) and also of manner ($t = 4.63$, $p<0.001$; $t=5.09$, $p<0.001$ respectively). A further subdivision into voiced and unvoiced place distinctions showed that the effect was greatest for unvoiced place. However, significantly fewer errors of voicing occurred in the 'flat' condition than in the 'switching' ($t=3.67$, $p<0.01$) or in the 'rising' ($t=3.44$, $p<0.01$) conditions, indicating that the expected degradation in voicing due to the effects of the tilt compression was also parallelled by a degradation in the 'rising' condition.

The results of the frequency-specific FAAF subscores in the bottom field of **Table 2** are of more use in documenting the particular effects of the switching scheme.[2] The decrement for 'switching' is characteristic of the features associated with all frequencies of speech energy. For low-frequency speech features, 'switching' was not significantly worse than the linear characteristic lacking low-frequency gain (i.e., the 'rising' characteristic), although it was slightly worse on mean scores ($t=1.67$, $p>0.1$). For high-frequency features, 'switching' was likewise not significantly worse than the characteristic lacking high-frequency gain, i.e., the 'flat' ($t=1.61$; $p>0.1$). Thus the scores representing use of information in both low- and high-frequency bands suggest that performance with 'switching' was no better in each frequency region than that associated with the characteristic giving lesser gain there; it was not similar to the performance on the control condition having the locally higher gain in the way desired and expected by the rationale for the aiding scheme.

For the mid-frequency features, 'switching' was considerably worse than either of the two linear characteristics ($t=4.76$, $p>0.001$; $t=5.47$, $p>0.001$) even though the long term mid-frequency spectral differences among all these conditions were very small. This is at first sight an unexpected finding, because it entails that the changes of about ± 2 dB in the octave around 1.7 kHz are paradoxically more deleterious than the steps of 10 dB or more in those octaves separated from 1.7 kHz by an octave or more. However, the more distant a frequency region from this region, the more rarely will it happen that appreciable energy of both periodic and aperiodic origins are contiguous within that region, and the less is the need for sequential reference, and the less will be the effect of the switching action. Individual differences throw further light[3] upon this explanation.

Despite the possible contribution of some hearing on the unaided ear, the own-aid freefield condition (mean score 72.9 percent; s.d. 8.8 percent) was, as expected, significantly worse than each of the two linear amplification conditions (against 'flat' $t=2.94$, $p<0.01$; against 'rising' $t=2.21$, $p<0.05$). It was still better than the 'switching' condition ($t=3.16$; $p<0.01$).

Individual differences

The present data have been used elsewhere (14) in a more general discussion of ways of predicting individual differences in disability. Only aspects relevant to the interpretation of the switching decrement and to delimitation of signal-processing aids are reported and discussed here.

For differential prescription of hearing aids it is potentially valuable to know to what extent the decrement for switching (and potentially other types of compression) was predictable from the configuration of each person's hearing impairment. To this end a difference score $[(F+R)/2-S]$ was computed between the mean of the 'flat' and 'rising' conditions and the 'switching' condition. Multiple regressions were run to predict values of this difference score and to predict the values of the absolute score in the 'switching' condition. Independent variables were age, the mean low-mid HL (0.5 and 1.0 kHz), high-frequency slope of HL (defined as the mean low-mid HL minus the mean HL at 2 kHz and above), and the variables from the PTC technique. Almost half of the variance in the difference score is explicable in this regression; results are summarized in **Table 3**. PTCE and mean HL did not contribute significantly to these regressions. The corresponding regressions were also run for the difference scores $F-S$ and $R-S$ separately, with results similar to those for the combined analysis. Inspection of the quartile (seven subjects) having least decrement for 'switching' (-7.5 percent or less) illustrated the import of the **Table 3** regression more concretely. Of these seven, six had better or equal performance in the 'rising' compared with the 'flat' condition. The remaining one lacking this

Table 3.
Summary of a stepwise multiple regression of individual characteristics with [(F+R)/2−S]. This expression is the difference in score between two linear conditions (combined mean) and the condition involving switching between those two conditions.

Predictor Variable	Simple r	Cumulative Adjusted Variance Explained	Coefficient for Prediction Formula
High Slope	0.54	0.27	0.55
PTCL	-0.14	0.35	-0.16
Age	0.41	0.41	0.47
PTCH	-0.32	0.45	-0.37
(Constant)			(20.00)
$F_{(4,23)} = 6.45$; $p<0.002$			

property but had extreme high-frequency loss, as implied by the sign of entry for slope in the regression, but he had no extreme values on any other obviously relevant variable.

The relation of the results in the 'switching' condition to those in the other two conditions, therefore, confirms the second therapeutic hypothesis, concerning frequency-dependent compression, to the extent of giving relative benefit from high-frequency information to those who need extra amplification to receive this information at all, even though this extra amplification is part-time. In other words, the comparison does realize some of the benefits of a rising frequency-response, but this is not an absolute benefit. It introduces, at least in the short-term and at the noise-level used here, greater deleterious side-effects that vary in a way that is only partly accountable for by the listeners' ages. The contributions of PTCL and PTCH were marginal after taking the tied variable of age into account, and the reverse was also true. Thus, greater age and poor resolution, not separable here, generally accompanied a particularly low performance in the 'switching' condition. The simple correlations between the PTC variables and the absolute score in the 'switching' condition were positive and significant ($r=0.55$, $p<0.001$; $r=0.71$, $p<0.001$). Together with the regression, this pattern suggests that individuals with poor frequency resolution, far from benefitting most (as the general appeal for signal-processing intends) have their performance most disrupted by the switching. The first therapeutic hypothesis, therefore, does not hold for this form of compression.

Linear conditions

Although mean overall scores for the two linear conditions were close, the error-types differed radically (7) in a way that may be made use of in determining appropriate frequency responses. The difference score (F−R) was partly predictable from a mixture of audiometric and PTC information (R^2 [corrected] = 0.36; $F_{3,24} = 4.81$; $p<0.01$). A rising frequency response was favored by an audiogram steeply sloping to poor high-frequency thresholds, by good PTCL (not excessive upward masking) and by poor PTCH. The difference in overall scores (F−R) was also predictable from the combination of the two FAAF subscores in their respectively least favorable conditions: the high-frequency subscore in the 'flat' condition and the low-frequency subscore in the 'rising' (unfavorable) condition ($R^2 = 0.41$; $F_{2,25} = 10.25$; $p<0.001$), and highly predictable from the conjunction of these variables ($R^2 = 0.59$; $F_{6,21} = 7.64$; $p<0.001$). The direction of the contributions by the subscores in these regressions supports the idea that the continuous variable (F−R) reflects the underlying information on which an appropriate choice of frequency response can be made.[4]

DISCUSSION

While some learning over time may well occur, the target of any complex signal-processing scheme for hearing aids must, for motivational reasons, be an initial performance that is no worse than that for the unprocessed control condition. From this point of view, the mean 12 percent decrement for switching looks unfavorable, but should not be viewed as being entirely negative. It need not rule out some improved version of the present class of scheme for some listeners at more favorable S/N ratios. There remains potential benefit from individually optimizing the relative gains of the two linear characteristics and some parameters of the switching action, if this could be done quickly and reliably. However, other more promising ideas should have priority.

The aid design copes, not very differently from the way a two-channel compression aid would, with the allegedly important effects of loudness discomfort at high frequencies. Given this, it is desirable to understand the basis of its deleterious side-effects, which may also be present in some degree in conventional compression systems. It is worth explaining this negative result because a similar negative result has been obtained by Bustamante and Braida (1). They transformed the spectro-temporal variations of speech into statistical principal components, the first comprising the overall amplitude, the second (PC2) the spectral tilt and so forth. Compression of PC2 (a continuous analogue of the dichotomous state-switching used here) gave degradation rather than enhancement of intelligibility. Likewise, results of a sophisticated digital implementation of a similar scheme by Levitt et al. (16) are in absolute terms more favorable, but attribute the majority of the benefit to compression of the overall amplitude, not the spectral tilt. While there remains a debate over the value of overall amplitude compression, to which the comparisons here do not contribute, there are now three pieces of negative evidence on spectral tilt compression.

How can we explain the relatively small amplitude shifts being so deleterious as attested by the vulnerability of the mid-frequency scores? We have distinguished three non-exclusive hypotheses which the data obtained do not serve to evaluate but which raise important issues for signal processing hearing aids.

Temporal referencing

Few of the group tested have usable hearing at high frequencies. While most have usable hearing at low frequencies, there is much variance at the mid-frequencies, which are going to be crucial for the analysis of place-of-articulation information. This on its own is not sufficient explanation, and perceptual use of temporal intensity contrasts also needs to be postulated. The idea that the switching design (and indeed any form of tilt compression) undermines important temporal intensity relationships is favored by recent experiments on the perception of speech sounds. It has been shown that cue values of certain acoustical features are most appropriately considered as temporally relative. For example, Ohde and Stevens (21) showed that the phonetic cue values towards place-of-articulation of phonologically voiceless phonemes are not just a matter of the absolute shape of the spectrum in the acoustically voiceless portions. They depend also on the time course of amplitude in relationship to the neighboring vowel formants, particularly in the high-frequency third formant region.

In another type of perceptual experiment, Summerfield et al. (23) have shown that a percept of a vowel can be created in a sound with a uniform spectrum totally lacking the usual formant structure of a vowel, provided that "valleys" in the spectrum precede the plateau so as to create increases in intensity at specific frequencies, as would occur in vowel onsets. The effect appears to be based at least partly upon the physiological phenomenon of adaptation, which acts as a gain control, compressing the dynamic range of response to steady-states while enhancing response to transients. This temporal relativity does not deny that absolute shapes of spectra play an important role in natural speech perception, but it commends specifications of perceptual cues in terms of intensity change with time. A fuller justification of such an approach is given in Haggard (11).

Summerfield et al. (23) showed that only 2 dB of contrast were required to sustain perception of a vowel via their contrast effect. Therefore, gain shifts of only a few decibels could disrupt the analysis of temporal patterns. In the present experiment a band of about 1/2 octave around 1.7 kHz undergoes a gain shift less than 2 dB on switching between the two linear characteristics; the switching action in the remainder of the mid-frequency region suffers changes that, while still quite small, are large enough to undermine sequential intensity relationships upon which listeners depend. The present result is consistent with Ohde and Stevens (21) in suggesting that sequential intensity referencing occurs across changes between the voiced and the unvoiced excitation mode; also, that the typical duration in one mode appears to be insufficient for a self-sufficient referencing framework to be established.

The relatively poor predictability in the switching condition with mid-frequency sounds is consistent with the temporal referencing hypothesis. The reliable individual variation in mid-frequency scores in the 'switching' condition is unrelated[3] to that variation determining overall score or low- or high-frequency subscores, and accounted for in **Table 3.**

It was also separate from the variation underlying the scores in the other conditions. Furthermore, multiple regressions from auditory characteristics, analogous to those of **Table 3** but using mid-frequency subscores in the switching condition as dependent variable, were not significant. In contrast, for the low- and high-frequency scores, the regressions were significant, involving PTCH and age (for low-frequency subscores) and PTCH and audiogram slope (for high-frequency subscores). This contrast further supports the existence of special factors governing individual differences in the extraction of information from the mid-frequency region in the switching condition. Here the intensity differences caused by the switching were smallest, but apparently most mis-interpretable as arising from the speech itself. The above analysis suggests that retaining temporal contrasts within broad bands may partly account for the more favorable results obtained with a two-channel compressor by Lawrence et al. (15) than in other studies using a larger number of differentially compressed bands.

Noise driving

Only one S/N ratio was tested here, a relatively adverse one for the hearing-impaired. It is possible (although unlikely from the Bustamante and Braida (1) and Levitt (16) findings) that the poor result is specifically related to the noise. Even when the switching action is controlled by the speech (as it is here) there is noise present whose spectral intensity in a particular frequency band will quite frequently exceed that of the speech. Thus, switching to the 'flat' characteristic on the advent of an /s/ will lead to an increase in low-frequency noise more typical of an /f/. All speech-processing devices that are analytic enough to take feature decisions pay for their great simplification and data-reduction with serious errors at adverse S/N ratios which the intelligent ear cannot retrieve. The switching aid is no exception.

Sensitivity to broad-spectrum cues

Some hearing-impaired listeners, lacking the ability to resolve local spectral maxima, will have to give relatively more reliance to overall spectral tilt; compressing tilt must then collapse much of their perceptual vowel space and blur the consonant feature place-of-articulation. The greater decrement for those with poor frequency resolution is consistent with this. Such listeners might do best not with overall compression but with expansion of spectral tilt, although this may aggravate the problems of achieving comfortable levels unless careful attention is paid to overall compression.

SUMMARY AND CONCLUSIONS

1. A design of signal-processing hearing aid was constructed to give some of the hypothesized benefits of frequency-selective amplitude compression in a simple fashion by compressing spectral tilt. This involved switching between two frequency responses, one flat and one of +9 dB/octave (which individually give about equal intelligibility); the switching was triggered by the overall slope of the short-term speech spectrum so as to reduce the dynamic range of the speech signal throughout the spectrum. This scheme gave worse performance on word identification than either of the constituent linear frequency responses alone.
2. The extent of the disadvantage for the switching condition was moderately predictable by multiple regression using listeners' ages, their frequency resolution at 2 kHz, and their high-frequency audiogram slope. With such individual prognosis available, there is a justification for proceeding with signal-processing schemes that implement some form of compression as a supplement to tailoring of frequency responses in an individual's total amplification requirement, but the prediction or adjustment of parameter values may be laborious.
3. The obtained disadvantage for the switching throws light upon some of the limitations with compression systems generally. One likely factor, with independent perceptual evidence, is the loss of the normal temporal reference framework for intensity changes within the frequency bands where the listener's information processing is most valuable.

ACKNOWLEDGMENTS

We thank our patient subjects for participating; Mr. A. G. Beeden, FRCS, and Mrs. C. Copperwheat of the E.N.T. Department, Bedford General Hospital, for providing access to patients; and Mr. J. Beaumont of the hospital administration for providing space in which to test them. Dr. A. Q. Summerfield produced the analysis yielding **Figure 1**, and also contributed constructive criticism of an earlier draft of this work.

FOOTNOTES

[1]The clinically observed benefit from compression in some seriously impaired patients (and from more complicated schemes) might be measured not so much as specifically enhanced transmission of basic dimensions of the information of auditory stimuli, but as greater ease and comfort of listening, or as lesser need to adjust volume controls. Some advantages may be economic, such as facilitating the fine specification, or adjustment of aids implementing other principles such as the tailoring of the frequency response.

[2]The validity of the subscores is supported (7) as expected, by the fact that on the low-frequency subscore, 'flat' was better than 'rising' ($t=4.89$; $p<0.001$), and on the high-frequency subscore 'rising' was better than 'flat' ($t=4.25$; $p<0.001$). The mid-frequency subscore did not give a significant difference between 'flat' and 'rising' ($t=0.87$; $p>0.3$). Thus, near-equality of overall scores does not mean perceptual indifference to the two conditions.

[3]The three correlations for the three pairings of absolute overall scores for the three main conditions were all greater than 0.6 ($p<0.001$), attesting to reasonable reliability. When the low- or high-frequency subscores were used, consistent rankings of subjects likewise emerged, for the corresponding pairings of conditions, despite the much smaller number of responses on which each score was based; correlations were all greater than 0.4 ($p<0.05$), so in these respects there is a fairly stable ranking of performance across subjects. The two correlations pairing the mid-frequency score in the 'switching' condition with the other conditions provided the only exceptions to this general pattern of reliability ($r=0.07$, NS; $r=0.18$, NS, between 'switching' and 'flat' and 'rising' respectively). This idiosyncrasy of the mid-frequency subscores is not to be explained by any generally poorer reliability of mid-frequency subscores as such; the scores' values were, for example, positively correlated between the flat and rising conditions ($r=0.58$; $p<0.001$). The idiosyncrasy problem is therefore specific to the 'switching' condition. This is possibly due to the fact that particular feature errors composing the mid-frequency subscore depended particularly upon formant frequency changes, which would have correlates in changes of spectral tilt in the mid-frequencies.

[4]For further tests of this type of psychoacoustical prediction see Lutman and Clark (18). The optimum frequency response may be construed as, and sensitively measured as, a midway point of balance between two other points on the function relating score to slope of frequency response: one point lies at a rising slope value where further advantage in high-frequency information is more than offset by the disadvantage in low-frequency information, the other point lies at a shallow or falling slope value where advantage in low becomes accompanied by greater disadvantages in high frequencies. The FAAF subscore technique would enable these two points to be determined without steep filtering, i.e., without the unrepresentative lack of spectral relationships and the unrepresentative removal of between-frequency masking which that would entail. It is a reasonable assumption that the score/slope function inscribes a roughly symmetrical inverted U-shape, with an optimum at about +4 dB/oct for the average listener here. The two disadvantageous combinations of subscore with condition (high/flat and low/rising) should then provide a rather direct reflection of the displacement of an individual's inverted U-shape from the zero-slope origin. The prediction of over 40 percent of the variance in the (F − R) difference from subscores based on some 30 percent of the 160 trials favors further investigation of error patterns in defining optimum frequency responses.

REFERENCES

1. **Bustamante DK and Braida L:** Principal-component amplitude compression of speech for the sensorineural hearing-impaired. *J Rehabil Res Dev* (this issue).
2. **Byrne D and Walker G:** The effects of mulitchannel compression and expression amplification on perceived quality of speech. *Austral J Audiol* 4:1-8, 1982.
3. **de Gennaro S, Braida LD, and Durlach NL:** Multichannel syllabic compression for severely impaired listeners. *J Rehabil Res Dev* 23(1):17-24, 1986.
4. **Evans EF:** Peripheral auditory processing in normal and abnormal ears. *Scand Audiol Suppl* 6:9-48, 1978.
5. **Fant CGM:** *Acoustical theory of speech production*. The Hague: Mouton, 1960.
6. **Foster JR and Haggard MP:** FAAF - An efficient analytical test of speech perception. *Proc Inst Acoust* 9-12, 1979.
7. **Foster JR and Haggard MP:** The Four-Alternative Auditory Feature test—A general-purpose test with multiple special-purpose applications and subscores. *Proc Inst Acoust* 8(7):167-174, 1986.
8. **Foster JR and Haggard MP:** The Four-Alternative Auditory Feature test (FAAF)—Linguistic and psychometric properties of the material with normative data in noise. *Brit J Audiol*, In Press, 1987.
9. **French NR and Steinberg JC:** Factors governing the intelligibility of speech sounds. *J Acoust Soc Am* 19:90-119, 1947.

10. **Haggard MP:** New and old conceptions of hearing aids. In *Hearing Science and Hearing Disorders*, Lutman ME and Haggard MP, Eds. London: Academic Press, 1983.
11. **Haggard MP:** Temporal patterning in speech. In *Time Resolution in Auditory Systems*, Michelsen A, Ed. Heidelberg: Springer, 1985.
12. **Haggard MP and Trinder JR:** Linear Hearing Aid. British Patent 8200317, 1982.
13. **Haggard MP and Trinder JR:** Problems in realizing the potential of integrated circuits in future aids for the hearing-impaired. In *High Technology Aids for the Disabled*, Perkins WJ, Ed. London: Butterworth, 1983.
14. **Haggard MP, Lindblad A-C, and Foster JR:** Psychoacoustical and audiometric prediction of auditory disability for different frequency responses at listener-adjusted presentation levels. *Audiol* 25:277-298, 1986.
15. **Lawrence R, Moore BC, and Glasberg B:** A comparison of behind-the-ear high fidelity linear aids in the laboratory and everyday life. *Brit J Audiol* 17:31-48, 1983.
16. **Levitt H, Neuman A, and Torasker J:** Orthogonal polynomial compression amplification for the hearing impaired. *Proceedings of the 10th Annual Conference on Rehabilitation Technology*, San Jose CA, 1987. Washington, DC: Association for the Advancement of Rehabilitation Technology, 1987.
17. **Lippman RP, Braida L, and Durlach NI:** Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss. *J Acoust Soc Am* 69:524-534, 1981.
18. **Lutman ME and Clark J:** Speech identification under simulated hearing-aid frequency response characteristics in relation to sensitivity, frequency resolution and temporal resolution. *J Acoust Soc Am* 80:1030-1040, 1986.
19. **Lutman ME and Wood EJ:** A simple clinical measure of frequency resolution. *Brit J Audiol* 19:1-8, 1985.
20. **Nabelek I:** Performance of hearing-impaired listeners under various types of amplitude compression. *J Acoust Soc Am* 74:776-791, 1983.
21. **Ohde R and Stevens KN:** Effect of burst amplitude on the perception of stop-consonant place of articulation. *J Acoust Soc Am* 74:706-714, 1983.
22. **Shore I, Bilger R, and Hirsh IJ:** Hearing aid evaluation: reliability of repeated measurements. *J Speech Hear Dis* 25:152-170, 1960.
23. **Summerfield AQ, Haggard MP, Foster JR, and Gray S:** Perceiving vowels from uniform spectra. *Percept Psychophys* 35:203-213, 1984.
24. **Thompson G and Lassman F:** Relationships of auditory distortion test results to speech discrimination through flat vs selective amplifying systems. *J Speech Hear Res* 12:594-606, 1969.
25. **Villchur E:** Signal-processing to improve speech intelligibility in perceptive deafness. *J Acoust Soc Am* 53:1646-1657, 1973.
26. **White MW:** Compression systems for hearing aids and cochlear prostheses. *J Rehabil Res Dev* 23:25-40, 1986.
27. **Zwicker E and Schorn K:** Psychoacoustical tuning curves in audiology. *Audiol* 17:120-140, 1978.