Taylor & Francis
Taylor & Francis Group

Original Article

International Journal of Audiology 2006; 45:130–152

*Stuart Gatehouse**
Graham Naylor†
Claus Elberling†

*MRC Institute of Hearing Research, Glasgow Royal Infirmary, Glasgow, UK
†Oticon A/S Research Centre, Snekkersten, Denmark

Linear and nonlinear hearing aid fittings – 1. Patterns of benefit

Adaptación de auxiliares auditivos lineales y no lineales – 1. Patrones de beneficio

Key Words
Compression
Benefit
Speech intelligibility
Listening comfort
Satisfaction

Abbreviations
AVC: automatic volume control
HTL: hearing threshold level
IG: insertion gain
SNHL: sensorineural hearing loss
ULL: uncomfortable listening level
WDRC: wide dynamic range compression

Abstract
We evaluated the benefits of fast-acting WDRC, slow-acting AVC, and linear reference fittings for speech intelligibility and reported disability, in a within-subject within-device masked crossover design on 50 listeners with SNHL. Five hearing aid fittings were implemented having two compression channels and seven frequency bands. Each listener sequentially experienced each fitting for a 10-week period. Outcome measures included speech intelligibility under diverse conditions and self-reported disability. At a group level, each nonlinear fitting was superior to the linear references for benefits in listening comfort, listener satisfaction, reported intelligibility and speech intelligibility. Slow-acting AVC outperformed the fast-acting WDRC fittings for listening comfort, while for reported and measured speech intelligibility the converse was true. For listener satisfaction there were no group differences between the nonlinear fittings. Analysis in terms of fittings for individual listeners revealed subsets with definite divergences from the group data and hence a need for candidature criteria. There are systematic differences between the benefits of nonlinear and linear fittings, and also within nonlinear fittings with fast versus slow time constants. The patterns of benefit and individual optima depend on the domain of outcome being assessed.

Sumario
Evaluamos en cincuenta sujetos con SNHL los beneficios sobre la inteligibilidad del lenguaje de adaptaciones de auxiliares auditivos con WDRC de acción rápida, con AVC de acción lenta y con referencia lineal, así como la discapacidad reportada, con un diseño de enmascaramiento cruzado que permite juzgar las diferencias intra-sujeto y en función del propio dispositivo. Se implementaron cinco adaptaciones con dos canales de compresión y siete bandas de frecuencia. Cada sujeto utilizó secuencialmente las diferentes adaptaciones durante un período de diez semanas. Las medidas de resultado incluyeron inteligibilidad del lenguaje bajo diversas condiciones y la discapacidad auto-reportada. La adaptación con AVC de acción lenta superó la de WDRC de acción rápida en relación con la comodidad para escuchar, aunque para la inteligibilidad del lenguaje, tanto medida como reportada, ocurrió lo contrario. En relación con la satisfacción del sujeto, no existieron diferencias de grupo entre las adaptaciones no lineales. El análisis, en términos en adaptaciones para sujetos individuales, reveló sub-grupos con divergencias definitivas en relación a los datos grupales, y por tanto, la necesidad de establecer criterios para la selección de candidatos. Existen diferencias sistemáticas entre los beneficios de las adaptaciones lineales y no lineales, así como entre las adaptaciones no lineales y las constantes temporales rápidas versus las lentas. El patrón de beneficio y de optimización individual depende del ámbito de resultado bajo escrutinio.

As technological advances have expanded the scope of options for hearing aid processing and fitting features, then the task of evaluating which features provide additional benefit in what circumstances and in what combinations has grown accordingly. Features such as frequency responses, maximum power output, compression ratios, number of channels, etc., have been investigated to an extent whereby there has emerged a broad appreciation of what is likely to be beneficial, or at least not unreasonable (Byrne, 1982; Cox, 1995; Dillon, 2001; Stone et al, 1999). Advocates of different individual rationales argue the case for a particular approach and its incremental benefits, but few would attempt to argue the case that competing approaches are fundamentally misplaced. As such, there is now a range of rationales which find acceptance in commercial products, for which some at least have received a degree of experimental underpinning. The generic, as opposed to manufacturer-specific, rationales offer little guidance about time constants in nonlinear amplitude compression systems (see for example Cornelisse et al, 1995; Valente & van Vliet, 1997; Byrne et al, 2001). Given the specific attention that has been paid to other parameters such as compression thresholds, compression ratios, and frequency-gain characteristics, this is perhaps surprising, especially given the wide range of time constants that are available, from milliseconds to seconds. There is a literature evaluating time constants in hearing aids with indeterminate or conflicting results. There are several reasons why a consensus has not developed.

Firstly, details of implementation of signal processing tend to differ from study to study, so for example results obtained with one system having a particular arrangement of channels and input/output curves might not be obtained with another configuration. 'Compression', even 'wide dynamic range compression' (WDRC), can be configured in many ways, with different goals in mind (Dillon, 1996; Moore 1998). Examples include the objectives of normalising loudness across frequency, of equalising loudness across frequency, or of maximising moment-to-moment audibility. Thus we might expect to find a diversity of

ISSN 1499-2027 print/ISSN 1708-8186 online
DOI: 10.1080/14992020500429518

Received: January 6, 2005
Accepted: September 20, 2005

Stuart Gatehouse
MRC Institute of Hearing Research, Scottish Section,
Glasgow Royal Infirmary, Queen Elizabeth Building,
16 Alexandra Parade, Glasgow G31 2ER, UK. E-mail: stuart@ihr.gla.ac.uk

conclusions in different studies. In this respect, the study presented here is no more generalisable than others have been, though as will be seen, the primary conclusions concern precisely diversity rather than consensus.

Secondly, compression provides a nonlinear signal transformation between input and output, in which the transformation applied to any given portion of the input signal depends on the spectral and temporal properties of preceding signal portions. Thus the results of studies of compression are bound to depend on the characteristics of the test signals used. This is especially true if the object of study is time constants, since varying these can fundamentally alter the nature of the nonlinearity. Therefore, if one wishes to have some confidence in the general validity of the results, it is important to ensure that test conditions reflect a relevant diversity of signal characteristics (Gatehouse et al, 2003). This has not always been the case.

Thirdly, the statistical power of many studies in the literature is insufficient to provide robust conclusions. It has been noted by at least one government evaluative body (The National Institute for Clinical Excellence in the United Kingdom National Health Service) that in comparison to other medical and clinical fields, the audiological literature is far from robust (National Institute for Clinical Excellence, 2000). Any survey of the hearing aid literature will identify many experiments, often yielding contradictory results, which are often small in size with limited statistical leverage and restricted in scope when placed in the context of the many factors that are known to affect outcome.

Lastly, any effective nonlinear amplification by definition introduces distortions into the spectral and temporal structure of the signals it operates on, some of which may not be immediately 'comprehensible' to the auditory system of an unaccustomed listener. Thus acclimatisation periods may be needed even for studies where the only outcomes are objective measurements made in the laboratory. The fact that some studies without acclimatisation do find effects of experimental variables does not mean that all effects will show up immediately.

Only a very small subset of the literature on compression amplification deals with studies in which the design includes a direct contrast between alternative time-constant settings. In the following review, we divide these into four groups according to their conclusions:

1. no effect of time-constants;
2. fast compression superior to slow compression;
3. slow superior to fast; and
4. mixed results with indications of different optima for different listeners.

(For a summary of the findings, see Table 1.)

In the first group, Bentler and Nelson (1997) studied subjective preference, rated intelligibility, objective intelligibility of nonsense syllables in cafeteria noise, and use time for each of four settings available over a single two-week period, for 14 subjects with a two-channel device. Attack time was held constant at 1 ms, and release times varied between (LF/HF =) 20/150, 20/35, 100/35, and 500/7 ms. Group mean data showed no significant trends. Note that no condition covered long time-constants in both LF and HF channels, and the compression kneepoints were relatively high (71 and 68 dB SPL in LF and HF channels respectively), so the degree to which compression was active is limited. Novick et al (2001) looked for effects of release time on speech reception in spatially distributed noise in combination with directional-microphone hearing aids with two compression channels. Ten subjects participated. Clear group mean effects occurred for the contrasts anechoic versus classroom environment and unilateral versus bilateral fitting. However, no overall effect of release time was observed. It should be noted that no acclimatisation period was included, and that the noise signals used were not significantly modulated. Moore et al (2004) looked for differences in speech recognition performance between linear amplification and four compression systems, of which three implemented fairly standard one-stage WDRC compression amplification (though in 20 channels). Time-constant variations were:

- very fast in all channels;
- fast at LF increasing to slow at HF; and
- slow at LF decreasing to fast at HF.

Vowel-consonant-vowel (VCV) tokens were presented against background noises of one talker, babble, and cafeteria noise, and consonant scores were recorded. No 'consistent or significant' differences were found between the four compression schemes. Note that only five subjects were used, and no acclimatisation period was provided before testing. Muller et al (2004) looked for differences in the gain setting preferred by 18 subjects as release time was varied (40, 160, or 640 ms) in a two-channel device. No effects were observable. It is open to doubt whether the subjects were able to determine their 'preferred gain' with sufficient precision for any effect of time-constants to be measurable, should it have existed.

In the second group, Nabelek and Robinette (1977) compared seven different combinations of time-constants ranging from (attack/release =) 6/30 to 130/580 ms. They found faster compression to yield higher percent correct on a word recognition task, using ten subjects. The effect was more pronounced as the stimulus SPL decreased. Since there was no background noise, this probably indicates the effect of increased audibility for less intense sounds with fast compression. Note that seven different commercial devices were used, providing a potentially strong confound. Jerlvall and Lindblad (1978) studied the effect of varying time constants in a single-channel setup with attack/release varying from 0.05/10 ms to 5/1000 ms. Recognition scores for CVC target words in carrier sentences were measured against steady-state white background noise at +5 and +60 dB SNR, with seven listeners. In the noisy condition, scores deteriorated for time constants longer than 1.5/300 ms, mostly due to increasing confusions amongst voiceless consonants. The compression ratio was 6:1, which would be very unusual in a modern WDRC system. Moore et al (1993), in one of the studies using their dual front-end AGC system, considered the effect of varying the release time in the HF channel of the second stage. Release times varied between 5 and 320 ms. SRTs were measured against steady-state noise and 12-talker babble (which is almost steady-state). For all five listeners, SRTs improved with shorter release times, being on average about 4 dB lower with the 5 ms release time than with the 320 ms release time. This rather clear result carries the caveat of applying to a compression system structure which is not immediately compatible with the majority seen elsewhere.

Table 1. Literature summary on the effects of hearing aid time constants

Study	*N*	*Conclusions*	*Caveats*
Group 1: No effect			
Bentler and Nelson (1997)	14	No effect of release time	No very long release time, high kneepoint
Novick et al (2001)	10	No effect of release time	No acclimatisation, no modulated noise
Moore et al (2004)	5	No effect of speed of compression	Small N, no acclimatisation
Muller et al (2004)	18	No effect of release time	. . . on preferred gain
Group 2: Fast superior			
Nabelek and Robinette (1977)	10	Faster compression gives better word recognition	No background noise, different devices for different time constants
Jerlvall and Lindblad (1978)	7	Word recognition in noise degrades with attack/release greater than 1.5/300 ms	Only steady-state noise, high compression ratio (6:1)
Moore, Glasberg and Stone (1993)	5	For all listeners, SRT improved with shorter release time in HF	. . . of second stage in two-stage compression system
Group 3: Slow superior			
King and Martin (1984)	10	Weak trend towards preference for release time = 1000 ms over 50 ms	High compression ratio (5:1)
Neuman et al (1998)	20	Longer release time gives better subjective quality	No acclimatisation
Hansen (2002)	6	All listeners preferred very long release times	15 channels, no acclimatisation
Group 4: Mixed			
Schweitzer and Causey (1977)	19	Word recognition best with release time of 90 ms, shorter or longer is worse	No background noise, I/O curve resembles output limiter
Neuman et al (1995)	20	No main effect of release time, but interaction (longer release times preferred for higher-level noise backgrounds); significant individual deviations from group trends	No acclimatisation
van Toor and Verschuure (2002)	38	No overall preferred compression speed (However, those subjects who obtained most benefit from fast obtained more benefit with the fast setting than those subjects who obtained most benefit with slow did with the slow setting. Those subjects who obtained most benefit from fast showed greater variation in benefit with the fast setting than those subjects who obtained most benefit with slow showed with the slow setting. Perhaps each individual hearing aid user needs different time constants for optimal performance.)	Design not fully balanced

In the third group, King and Martin (1984) contrasted release times of 50 and 1000 ms, measuring 'speech tracking' performance and preference with speech in babble at +7 dB SNR. With 10 listeners, a weak trend towards preference for the longer release time was observed, being described by subjects as 'the speaker's voice stood out better from the background'. The compression ratio (5:1) was very high by today's standards. Neuman et al (1998) recorded categorical judgements of sound quality for a single-channel WDRC system. The attack time was 5 ms, and release times were 60, 200, and 1000 ms. In a laboratory setup, 20 listeners heard speech against backgrounds of ventilation, apartment, and cafeteria noise and rated clarity, pleasantness, background noise, loudness, and overall impression. Longer release time was found to be associated with greater pleasantness, less background noise, and less loudness. The effect of release time increased with increasing compression ratio, as one might expect. Hansen (2002) conducted a parametric study of the effect of attack and release times in a 15-channel WDRC system on sound quality and speech intelligibility, as subjectively rated via paired comparisons. Recordings from real everyday situations were reproduced with spatial information. Attack times (equal in all channels) varied from 1 ms to 100 ms, and release times (also equal in all channels) from 40 ms to 4000 ms. All six hearing-impaired listeners showed significant preference for the longest release time, particularly for signals containing speech. With the longest release time, the effect of attack time was small. In this study the compression ratio was 2.1:1 in all channels, and the compression threshold was very low (ca. 20 dB SPL).

In the fourth group, Schweitzer and Causey (1977) studied the effect of varying release time in a single-channel device. Attack time was always 10 ms and release times ranged between 15 and 320 ms. Nineteen listeners showed on average best word recognition scores with a release time of 90 ms, being about 12% better than with the shortest or longest release times. The I/O curve of the device resembled output limiting more than WDRC, and no background noise was present, so the general validity of the result is open to doubt. Neuman et al (1995) obtained paired-comparison judgements of quality from 20 subjects listening to speech against a variety of noise backgrounds through a single-channel system with release times of 60, 200, and 1000 ms. No main effect of release time was found, but a significant interaction existed, such that the longer release times were preferred for the higher-level noise backgrounds. In addition, significant individual deviations from group trends were found, suggesting that different listeners had different release-time preferences. Van Toor and Verschuure (2002) carried out a study with a rather complex design, in which direct contrasts of 'Slow', 'Intermediate', and 'Fast' compression in a four-channel device occurred. Attack and release times covaried, but in the same direction. 'Slow' had (attack LF-HF)/(release LF-HF) =(64–32)/(2048–1024) ms, 'Intermediate' had (16–2)/(512–64) ms, and 'Fast' had 2/(64–16) ms. Substantial periods of acclimatisation were included, and 38 subjects took part. SRT measurements were made with three types of noise (steady-state, one-talker-modulated noise, and car noise), and statistical analysis was based on the benefit scores (aided minus unaided performance). No overall preference for one of the settings was found. However it was found that subjects who obtained most benefit from 'Fast' obtained more benefit with the 'Fast' setting than subjects who obtained most benefit with 'Slow' did with the 'Slow' setting. It may in addition be construed from their Table 14 that subjects who obtained most benefit from 'Fast' showed greater variation in benefit with the 'Fast' setting than subjects who obtained most benefit with 'Slow' showed with the 'Slow' setting. The authors speculated that perhaps each individual hearing aid user needs different time constants for optimal performance.

Taken together, the overall picture is indeed one characterised by indeterminate and conflicting results. Almost all studies carry caveats that leave their results (whether null or not) open to question. The only conclusion which might be resistant to these caveats is that hinted at by Neuman et al (1995) and van Toor and Verschuure (2002), namely that individual listeners show divergent preferences or optima.

A traditional approach to hearing disability and hearing aid benefit has assumed a natural focus on speech intelligibility, though hopefully with a realisation that speech intelligibility in the laboratory is an incomplete picture. However, as the understanding of auditory disability (generally considered to be reduced ability to perform useful auditory tasks, such as understanding speech, localising sounds, parsing the auditory scene, etc.) has expanded, the range of relevant hearing functions, abilities, and disabilities has increased, and undermined a simple concentration on the segmental intelligibility of a constrained speech stimulus (Noble, 1998; Gatehouse and Noble, 2004). There is a growing appreciation of the complexity of goals and objectives in hearing rehabilitation and the ways in which hearing aid fittings can either promote or compromise their realisation (Gatehouse, 1994). It is therefore imperative that any evaluation of hearing aid fitting and processing features embraces as wide a range as possible of the goals and objectives that are likely to be relevant to hearing-impaired listeners.

Evaluation of hearing aid processing and fitting features is subject to many confounds, particularly when different physical devices are compared or when either or both of the test subject or experimenter has prior knowledge of the feature under investigation. Robust and appropriately masked (blinded) comparisons are the exception rather than the rule (National Institute for Clinical Excellence, 2000). This manuscript reports an experiment designed to investigate the impact of different time constants in hearing aids on multiple dimensions of hearing aid benefit whilst minimising the confounds which might compromise interpretation of the results. Because of the multidimensional nature of hearing disability and hearing aid benefit, the experimental design samples a wide range of conditions for both laboratory testing and self-reports of everyday listening so that the different benefits delivered in different domains might be appropriately captured.

Although it has been stated above that there is little guidance on hearing aid time constants, a more substantive justification for concentration on their effects is required. We contend that the different conceptual bases for either fast or slow time constants in hearing aids provide that justification. This contention is presented in more detail in the following.

It is well established that sensorineural hearing loss entails abnormal function beyond simple attenuation of sound, and includes to varying degrees reduced dynamic range between threshold of hearing and threshold of uncomfortable listening, widened auditory filters, and abnormalities in the temporal processing of auditory signals. Input signals become distorted in the temporal, spectral, and intensity domains, and are required to be processed through a perceptual channel with reduced dynamic range. Speech consists of an acoustic signal with rapid changes in both frequency and time which is required to be mapped into the abnormal periphery. Fast-acting wide dynamic range compression systems attempt to overcome some of the distortions and nonlinearities in sensorineural hearing loss by maximising the effective moment-to-moment audibility of a speech signal. As such, they must act rapidly over periods of time comparable to the rapid fluctuations in the speech signal. In this manuscript we reserve the term wide dynamic range compression (WDRC) for nonlinear implementations with (at least some element of) fast-acting compression. In contrast, slow-acting automatic volume control systems (AVC) are designed to provide the listener with access to the external auditory world with minimal processing and distortion artefacts whilst adapting to the longer-term changes as the listener's auditory environment alters, or as listeners move from environment to environment.

This design goal implies time constants commensurate with the environmental changes themselves.

The contrasting goals of fast-acting wide dynamic range compression systems and slow-acting automatic volume control systems provide the prima facie case for investigating the benefits that they deliver and the domains within which those benefits reside for hearing-impaired listeners.

Methods

Listeners

Listeners were identified from the clinic lists at the audiology department at Glasgow Royal Infirmary according to the following criteria. All listeners were users of a single postaural hearing aid as supplied by the United Kingdom National Health Service (NHS), with a use score of at least 50 on the Glasgow Hearing Aid Benefit Profile (Gatehouse, 1999). At the time of the study, the NHS in Scotland employed linear devices with output limiting via peak clipping. Listeners were invited to take part in the study, and gave informed consent as part of hospital ethical approval. No age or other selection criteria were used, though the process of obtaining informed consent did detail the experimental processes and thus deselected participants who felt they would not be able to comply. Air and bone conduction thresholds were assessed (British Society of Audiology, 1981) and listeners were selected who had bilateral symmetric sensorineural hearing loss, defined as average (over 500, 1000, and 2000 Hz) air-bone gap <10 dB with no single frequency air-bone gap >15 dB, and average interaural asymmetry (over 500, 1000, 2000, and 4000 Hz) <15 dB with no single asymmetry >20 dB. Listeners with reverse slopes (>10 dB between 500 and 1000 Hz against 2000 and 4000 Hz) were excluded. Listeners' hearing thresholds were selected to fall within the fitting range (see Table 2) of the device used throughout the experiment. (The experiment used a single device termed 'JUMP-1' – see subsequent sections for a full description.)

Each listener was tested on the set of outcome measures described below while wearing their NHS hearing aid prior to formal enrolment in the study to familiarise them with the requirements of the experiment and to check their willingness to complete the study protocols.

Table 2. Statistics of the hearing thresholds for the 50 listeners in the experiment, and the upper bound of the JUMP-1 fitting range

	Frequency				
	250 Hz	*500 Hz*	*1000 Hz*	*2000 Hz*	*4000 Hz*
Hearing thresholds dB HL					
Minimum	10	10	10	15	40
Lower quartile	20	25	25	35	50
Median	30	30	35	47.5	60
Upper quartile	40	40	45	55	70
Maximum	70	60	60	65	75
JUMP-1 fitting range dB HL					
Upper bound	75	80	80	80	75

Of the 61 listeners who entered the preliminary evaluation using their NHS aids, seven withdrew when the extent of the test protocol, and hence the commitment required, became clear. Of the 54 listeners who formally entered the experiment, two withdrew because of ill health, and a further two listeners failed to complete the experimental design. Statistics of the thresholds for the 50 listeners who form the experimental pool, and details of the JUMP-1 fitting range, are shown in Table 2. There were 28 females and 22 males, while the mean age of the participants was 67.1 years and ranged from 54 to 82 years.

Stratification

There is an established literature which links hearing losses that slope from low to high frequencies and dynamic ranges (the difference between threshold of uncomfortable listening and threshold of hearing) that are more restricted at high compared to low frequencies with increased benefit from amplitude compression systems (Moore et al, 1992; Lunner et al, 1997a; Lunner et al, 1997b). Although this article focuses on the relative benefits and domains of benefit for linear and nonlinear fittings in a reasonably representative group of hearing aid wearers, the focus of a companion article is on the predictors of benefit (Gatehouse et al, 2006). We have therefore adopted a stratification system which aligns with our major hypotheses. These hypotheses are developed in more detail in the companion article but briefly consist of an expectation that candidature for different fittings is related to the established audiometric slope/dynamic range and in addition the parameter of auditory ecology (Gatehouse et al, 1999). Prior to entry into the experiment no information was available in the ecological domain, although as described in the companion article we have developed both measures and reports of listeners' auditory environments and demands. As part of the routine clinical processes at the host clinic in Glasgow Royal Infirmary, pre-fitting data on the Glasgow Hearing Aid Benefit Profile (Gatehouse, 1999) was obtained for each listener. The Glasgow Hearing Aid Benefit Profile consists of reports on four prespecified listening circumstances and additional reports on up to four listening circumstances that are volunteered by listeners. We have used the number of these additionally specified listening circumstances as a preliminary index of the range of auditory environments encountered by individual listeners.

When identifying potential participants from the clinical records, the difference in dynamic range at 2000 Hz and 4000 Hz compared to that at 500 Hz and 1000 Hz, along with the surrogate index of auditory ecology described above, were used to select candidates for each of four cells. Table 3 cross-tabulates

Table 3. Crosstabulation of the stratification variables of mean difference in dynamic range (ULL – threshold) at 2000 and 4000 Hz compared to 500 and 1000 Hz, and the number of volunteered listening circumstances in the free section of the Glasgow Hearing Aid Benefit Profile

Number of volunteered situations on the GHABP		*Dynamic range at 2000 and 4000 Hz minus Dynamic range at 500 and 1000 Hz*	
		≤ 10 dB	>10 dB
	≤ 1	10	15
	>1	9	16

the difference in dynamic range between high and low frequencies against the number of additional circumstances specified on the Glasgow Hearing Aid Benefit Profile, with a cut-off value of 1 for the number of circumstances and 10dB for the difference in dynamic range.

Prior to the current experiment, data were collected on several hundred hearing aid fittings at the host clinic at Glasgow Royal Infirmary and, after implementation of the audiometric criteria for the current experiment, the median difference in dynamic range between high and low frequencies was close to 10 dB for this large control group. In the current experiment we therefore have a slight overrepresentation of listeners with greater differences in dynamic range between low and high frequencies, as shown in Table 3, which could compromise the statistical leverage of the experiment to a minor degree.

The primary purpose of the stratification was to allow sufficient statistical leverage separately for the major hypothesised dimensions of audiometric characteristics and auditory ecology, and this theme is developed further in the companion article. Although the sample under test is not truly representative of the host clinic population, this will always be the case in a research context. The statistical objective has been achieved, albeit at the expense of a slight departure from a truly representative clinical sample.

Statistical Power

During the planning phase of the study, the experiment was powered to detect a within-subject between-fitting difference of 4% on the mean scores across conditions on the speech intelligibility tests described subsequently for $p < .05$ at 80% power. This required at least 40 complete data sets for each fitting if Bonferroni corrections for within-subject multiple contrasts were applied. Our initial assumptions about drop-out rates and non-usability of data proved to be pessimistic, and this power level was therefore exceeded.

Hearing aid fittings

All subjects were fitted unilaterally, on the previously aided ear (23 left and 27 right ears). All the fittings were implemented in the same behind-the-ear (BTE) device, Oticon's JUMP-1 (Naylor, 1997). Confounding variables due to differing electroacoustics, cosmetics, or expectations thereby were removed from the study, and crucial aspects of the fittings could be masked from test subjects and experimenters. The JUMP-1 platform has access to two channels of amplitude compression with flexible time constants, and employs seven bands of frequency shaping. The JUMP-1 devices were programmed via custom software to convert the desired audiological and acoustical characteristics into internal JUMP-1 register values. Initial testing of the devices and software in standard couplers confirmed their integrity.

All the fittings were specified and implemented in a consistent manner, so as to simplify the experimental design and avoid unwanted or uncontrolled additional variables. At the same time it was desired that each fitting should be representative of the philosophy it embodied without undue compromises. Five fittings were designed. All five were nonlinear in so far as they all contained a compression hard-limiter against excessive output levels; but apart from that, two were linear (constant gain across input levels) and three contained nonlinear amplitude compression. As a reference which would be recognisable by the broad hearing aid research community, a one-channel implementation of the *NAL-RP* prescription including a volume control was chosen. Since one of the main points of interest of the study was the influence of 'fast' versus 'slow' compression in multichannel nonlinear prescriptions, two nonlinear prescriptions were designed to epitomise 'fast' and 'slow' two-channel compression (hereafter called '*FAST-FAST*' and '*SLOW-SLOW*'). A third nonlinear prescription (*FAST-SLOW*) was a hybrid between the 'fast' and 'slow' prescriptions, and was included due to the good performance it showed in Lunner (1997b), and also the proprietary interest it held for the second and third authors. For the compression prescriptions, the volume control (VC) was disabled and had no effect. This is typical in commercial designs and avoids a complicating factor in the analysis of results, though it does introduce a potentially important difference with the *NAL-RP* fitting. As a bridge between the NAL fitting and the two-channel WDRC designs having no VC, a two-channel linear prescription without VC (*LINEAR*) was added. The inclusion of this second linear reference allows more direct interpretation of the origins of any differential performance between linear and nonlinear fittings.

For each of the two compression channels, the input-output function was described by three straight-line segments: linear gain for low input levels, constant compression ratio for medium input levels, and constant output for high input levels. The parameters included in the specifications were as follows:

1. insertion gain (IG) for speech-spectrum input at 65 dB SPL free field, at the centre of each of the seven bands in the JUMP-1's filterbank;
2. compression ratio (CR) in each of the two compressor channels;
3. compression kneepoint in each of the two compressor channels;
4. maximum power output (MPO) for each of the two compressor channels;
5. attack and release time constants for the level detectors in the two compression channels;
6. crossover frequency between the two channels; and
7. volume control locked/unlocked.

Each fitting consisted of a set of rules for generating values of these parameters. These rules are described below, and the distinguishing features of the five prescriptions are summarised in Table 4. Since all hearing losses were sensorineural in nature, there was no need to adjust the prescriptions of gain or compression for conductive components of the loss.

The channel crossover frequency could have been chosen as a function of the audiogram, on the basis of some property of the supposed input signal, or by adjustment for wearer preference. However, a constant value across all subjects and all multichannel prescriptions was desired for a homogeneous experimental design. Of three available options (1000, 1500, and 2500 Hz) in the JUMP-1, 1500 Hz was chosen. This is a reasonable choice for most audiograms, and is a good choice as a boundary between low-frequency and high-frequency energy in speech signals.

The *NAL-RP* prescription used the frequency-gain characteristic from (Byrne and Dillon, 1986). The other four fittings shared a gain characteristic based on (Lunner et al, 1997a). For these, the IG for a static speech-shaped input at 65 dB SPL

Table 4. Summary of the parameters of the JUMP-1 fittings

	NAL-RP	*LINEAR*	*SLOW-SLOW*	*FAST-FAST*	*FAST-SLOW*
Channel split	n/a		1500 Hz		
Limiting	Macrae and Dillon 1986		Worst case pure-tone UCL-12 dB		
Frequency response	NAL-RP		POGO II − 4 dB (LF) and POGO II − 8 dB (HF) for speech at 65 dB		
Compression kneepoint LF Compression kneepoint HF	n/a	n/a	50 dB SPL free field in channel bandwidth (speech signal assumed)		
Attack/release times LF (ms) Attack/release times HF (ms)	10/20	10/20	10/640	10/40	10/40 10/640
Compression ratio LF Compression ratio HF	1.0	1.0	Composite loudness model (Elberling 1999)		
Volume control	Yes		No		

free field, was equal to POGO-II (Schwartz et al, 1988) minus 4 dB in the LF channel and POGO-II minus 8 dB in the HF channel, with smoothing across the LF-HF channel boundary. The inclusion of the *LINEAR* fitting in the design allows a contrast with the *NAL-RP* fitting to determine whether any material effects of static frequency-gain characteristic were present, though the prediction was that this would not occur. The four two-channel fittings implement the same static frequency-gain characteristic. Table 5 shows the prescribed and realised insertion gains for the two sets of static frequency-gain characteristics.

In the *NAL-RP* prescription, the frequency-gain characteristic was combined with the MPO formula of Macrae and Dillon (1986). The four two-channel prescriptions used an MPO prescription which ensured that the MPO was at least 12 dB below puretone UCL at all UCL measurement frequencies. This is similar to the approach of (Lunner et al, 1997a), and again the *LINEAR* versus *NAL-RP* contrast allowed appreciation of any MPO effects. The mean MPO for the *NAL-RP* prescriptions exceeded that for the other four prescriptions by 7 dB in the low-frequency channel, and by 4 dB in the high-frequency channel.

If the CR is to vary with hearing loss, it has to be chosen by some rule, and any reasonable basis for such a rule must aim to adjust the input dynamic range towards that of the hearing aid user. Thus a CR increasing with hearing loss is reasonable. Pure-tone loudness considerations are a reasonable approach to obtain a suitable rule (whether or not loudness compensation is the stated aim of a prescription). The formula used to prescribe CR for the nonlinear prescriptions is based on statistical data from loudness measurements on normal and impaired listeners (Elbering, 1999), and is shown below:

$$CR = \begin{cases} 1/(1-0.0081^{*}\overline{HTL}), \text{for } \overline{HTL} \leq 70 \\ 1/(0.615-0.0026^{*}\overline{HTL}), \text{for } \overline{HTL} \geq 70 \end{cases}$$

where $\overline{HTL}$ is the average hearing threshold within the compression channel bandwidth in dB HL.

This is a combination of four published sets of statistics (Elberling & Nielsen, 1993; Launer et al, 1996; Ricketts & Bentler, 1996; Keissling, 1995) where each set is weighted according to the sample size (Elberling, 1999). The formula is applied to the HTL averaged across each compression channel's frequency bandwidth to obtain the CR for that channel (equal to the estimated inverse slope of the loudness function for the average HTL). The CR for the low-frequency channel was 1.4, and for the high-frequency channel was 1.8 averaged across the fittings for the 50 listeners.

For homogeneity of the experimental design, the attack times of the level detectors in all the prescriptions were equal, at a nominal 10 ms (internal circuit value). This is rather longer than is normal for the output limiter in a linear prescription, but not unusual for compression prescriptions. Actual resulting attack times measured according to IEC 60118-2 (1983) varied from approximately 8 to 15 ms, dependent on other compression parameters. Release times of nominal 20 ms were chosen for the output limiter in the linear fittings. For the nonlinear fittings, release times were chosen so as to be indisputably 'fast' (nominal 40 ms) or 'slow' (nominal 640 ms). Thus in fact the only difference between the three nonlinear settings was the release time, which varied by a factor of 16:1. Measured IEC release times varied between approximately 80% and 100% of the nominal values, again dependent on other compression parameters. No procedures were implemented to deal with overshoots in the compression fittings.

All three nonlinear fittings shared the same compression kneepoint. Assuming a free-field spectrum shape approximating the long-term spectrum of speech (Cox & Moore 1988), compression in a given channel began when the free-field level within that channel's frequency bandwidth reached 50 dB SPL.

Table 5. Prescribed and realised insertion gains (dB) for a 65dB static wideband input signal averaged across the fifty listeners

Frequency (Hz)	*Single-channel (NAL-RP)*		*Two-channel fittings*	
	Prescribed	*Realised*	*Prescribed*	*Realised*
250	−2	0	1	0
700	11	11	11	12
1200	18	19	13	12
1900	19	18	16	16
2900	20	21	19	20
3900	22	21	22	23
4900	23	22	24	22

The *NAL-RP* fitting presupposes a user volume control, and we wished this to be a reference condition which could be related to other studies if desired. Therefore the volume control was enabled for the *NAL-RP* fitting. It was disabled in all other cases, which has the advantage that a source of uncontrolled variance is removed from the design. On the other hand, it may be (where benefit ratings are concerned) an 'unfair' restriction on a linear prescription like *LINEAR*. Nonlinear fittings are commonly made without a VC, so major problems of subject acceptance were not anticipated with *SLOW-SLOW*, *FAST-FAST* or *FAST-SLOW*. A need for subject counselling was, however, anticipated with *LINEAR* fittings. Such counselling might be considered a form of experimental bias.

The gain and filtering blocks of the JUMP-1 were adjusted prescriptively on the basis of proprietary values of full-on insertion gain for the BTE configuration in question and the frequency responses of the bank of filters, so as to approach the target values as closely as possible. At the first subject visit with the hearing aid in place, real-ear measurements were made to obtain subject-specific corrections to the prescription. These corrections were then used in all prescriptions for the given subject.

Proprietary transformations were available between free field sound levels and electrical voltage at the input to the level detectors, for each of the seven filter bands. These were used along with an assumed free field signal spectrum (the speech spectrum of Cox & Moore, 1988) and the filter and gain settings after insertion gain calibration, to obtain for each compression channel the level detector input voltage corresponding to the desired compression kneepoint. Power summation across filter bands was assumed.

For a few subjects with very mild losses at low frequencies, the specified compression kneepoint could not be set at a sufficiently small value in the LF compression channel. For these subjects the kneepoint was up to 6 dB higher than prescribed, and while insertion gain above the achieved kneepoint was correct, gain below the kneepoint was up to 2 dB lower than prescribed.

In all the fittings, maximum power output was set by implementing a hard compression limiter (constant output level for varying input level) as the final straight-line segment of the gain versus input characteristic. For a reference setting of the device (full-on gain and maximum output), measurement data was available for the output achieved in an IEC711 coupler as a function of signal frequency. The manner in which the hard limiter was implemented corresponds to a simple reduction of the MPO versus frequency curve by a constant amount for all frequencies within a given compression channel. The amount of reduction required was derived in different ways for the *NAL-RP* and the two-channel prescriptions. The MPO prescription for the two-channel case was that the difference between the MPO-versus-frequency curve and the UCL-versus-frequency curve should be 12 dB at the frequency where they were closest together. Hence the MPO was at least 12 dB below UCL at all frequencies within a given channel. By transforming the full-on MPO and UCL curves to equivalent units (e.g. dB SPL in a standard coupler), the amount of MPO reduction required in each compression channel is easily determined. The *NAL-RP* prescription generated a desired value of three-frequency averaged (3FA) MPO (500, 1000, and 2000 Hz). The difference between this value and the 3FA MPO for the full-on condition yields the amount in dB by which the MPO should be reduced from full-on.

All the prescriptions were pilot-tested on a separate group of five subjects. They were found to be acceptable and without glaring shortcomings, which might have rendered the main experiment invalid. In particular, the *LINEAR* prescription was accepted without a volume control.

Experimental Design

The five fittings described above are referred to hereafter as:

- *NAL-RP*
- *LINEAR*
- *SLOW-SLOW*
- *FAST-FAST*
- *FAST-SLOW.*

Each of the 50 listeners wore each of the five fittings for a 10-week period. This length of time was chosen to ensure that any effects of prolonged acclimatisation to the fittings (Cox & Alexander, 1992; Gatehouse, 1989; Gatehouse, 1993; Gatehouse and Killion, 1993) were incorporated, should any have occurred. More importantly, such a period was thought to be necessary to allow listeners the opportunity to gain sufficient experience with each of the fittings across the range of acoustical environments and perceptual demands encountered in everyday life, to allow meaningful reports of the performance and characteristics of each fitting.

Each listener wore the *NAL-RP* fitting first on entry into the formal study. This fitting alone had an active volume control. As described above, the presence of two linear fittings in the design aimed to avoid any incremental benefits or deficits of a nonlinear over linear fitting being simply a surrogate of either frequency response or the presence/absence of a volume control. Following the *NAL-RP* fitting, each of the other four fittings was worn for a 10-week period in a randomised order. At the end of each 10-week block, listeners attended for two sessions of three hours duration during which the outcome assessments were performed, and the programming, fitting, and fine-tuning of the subsequent fitting was performed.

At each stage of the experiment, the fitting under investigation was masked as far as was practicable from both the test subject and the experimenter. The fitting software contained a hidden key which mapped a nominal fitting code onto each of the five actual fittings. That key was not available to the audiologist. At the outset, the fittings and design of the study were described to the volunteer subjects, but at each stage the fitting under assessment was not disclosed. Clearly it is not possible to mask from a listener whether a volume control is operative, and while theoretically possible, this is also difficult to mask from an experimenter when assessing outcome. The overall experiment cannot therefore be described as a true double-blind randomised control trial because the *NAL-RP* always occurred first, and because the *NAL-RP* fitting had an active volume control which the wearer could not fail to be aware of. The design is though both double blind and randomised across the *LINEAR*, *SLOW-SLOW*, *FAST-FAST* and *FAST-SLOW* fittings, and comparison of the *NAL-RP* and *LINEAR* fittings allows an appreciation of order effects, effects of incomplete blinding for both subjects and experimenters, as well the effects of frequency response and access to a volume control.

Verification and fine-tuning

The fitting rationales summarised in Table 4 yield one frequency/gain characteristic for a 65 dB steady-state speech-shaped noise input for the *NAL-RP* fitting, and one single nominal target for the other fittings (this single target is nominal because in reality the differing time constants in the fittings will result in differing actual characteristics for a complex time-varying signal such as speech). The programming software, which converted the desired fitting characteristics into register values in the JUMP-1, assumed a constant average value for the real ear to coupler difference (RECD) for adult ears. For each listener and each fitting, the real ear insertion gain (REIG) for a 65 dB steady-state speech shaped noise was derived. Software trimmers were used to manipulate the gain for each of the seven frequency bands in the JUMP-1 to minimise difference between the measured and target REIG. For all listeners this was achieved to within 2.5 dB for each frequency band. The process represents the acoustical fine-tuning and verification of the fittings. Note that all of the other parameters in Table 4 will be either unaltered or affected equally by this difference between the assumed and actual RECD.

Any clinical hearing aid fitting takes as its starting point a number of assumptions which may not necessarily apply to any given hearing-impaired individual, and hence a process of fine-tuning is needed to ensure an optimal fit. For this study, it was important to keep the fittings as homogeneous as possible across the subjects, and therefore fine-tuning for an *acceptable* rather than optimal fit was performed. At the first fitting visit, listeners were presented with continuous speech at levels of 65 and 80 dB SPL. The gain and MPO setting for the low and high-frequency compressor channels were adjusted for acceptability and to alleviate feedback. No adjustment >5 dB for any setting was required to achieve acceptability. Thus no attempt has been made to identify an optimal frequency response, MPO, compression ratio or compression kneepoint for an individual listener. The experiment represents a set of fittings and conditions which are maximally homogeneous, and thus maximises the extent to which any differences that might emerge can be attributed to the conceptual rationale underpinning a fitting, rather than the idiosyncratic implementation and adjustment for individual listeners.

Outcome Measures

Speech tests

Speech identification ability was assessed using the Four Alternative Auditory Feature test (FAAF, Foster & Haggard, 1987). The FAAF test consists of an 80-item vocabulary grouped into 20 sets, each with four items. Each of the four items differs only in either the initial or final consonant. The four response alternatives were offered to the listeners on a touch-sensitive computer screen, prior to the onset of the stimulus. The stimulus consisted of a target word embedded in a carrier phrase ('Can you hear *target* clearly?'), spoken by a single male talker. Listeners responded by selecting one of the four alternatives on the touch screen. No feedback was given. The levels of the speech tokens were calibrated by isolating the target words, concatenating into a single waveform file, and generating a steady-state noise file with the same spectrum and root-mean-square (RMS) amplitude as the targets. This speech calibration file was used to set the various speech presentation levels. All waveform files were Windows wav files with a sampling rate of 44,100 Hz. A 60-second duration waveform file with the same spectrum and RMS level as the FAAF target words served as one of the noise files. The initial sampling point of the noise file was varied randomly, and the noise was gated on and off in synchrony with the FAAF carrier phrase. Both the speech and the noise were presented in a sound-reduced enclosure from a single loudspeaker located at 0 degrees azimuth and 1 metre from the centre of the subject's head. Levels were calibrated in dB SPL at the position of the subject's head, with the listener removed from the sound field.

In addition to the steady-state speech-shaped noise, the ICRA noises (Dreschler et al, 2001) with two-talker and six-talker envelopes were used. The waveform files for each of these were adjusted to have the same RMS amplitude as the steady-state noise. A total of 18 test conditions in a $3\times2\times3$ matrix were used, consisting of:

1. speech presentation levels of 55, 65 and 75 dB SPL;
2. signal-to-noise ratios of $+5$ and $+10$dB; and
3. steady-state, ICRA-2 and ICRA-6 noise.

Given the potential advantages of nonlinear over linear processing, this range of test conditions was chosen to sample the range of acoustical conditions that listeners might experience in everyday life.

Presentations were grouped in blocks of six conditions with the noise type held constant. Within each block the presentation level and signal-to-noise ratio were interleaved randomly, thereby removing any opportunity to adjust the volume control in the *NAL-RP* condition. The order of testing by noise type was chosen randomly. The 18 conditions were presented both unaided and aided, and the whole procedure was repeated twice, once at each of the two assessment visits following the 10-week trial/acclimatisation period for each fitting. Each test session began with a single practice run in a condition chosen at random. All results in this paper are constructed as benefit (aided minus unaided) scores. In total, there were 36 benefit scores for each hearing aid fitting (three presentation levels, two signal-to-noise ratios, three noise types, and two replicates). In the subsequent data-reduction phase, each of the 36 values was treated as an independent variable. For the *NAL-RP* fitting, the volume control was set to comfortable listening for the FAAF material presented at 65 dB SPL in quiet. Detailed analyses of the interactions between the hearing aid fittings, the speech-test conditions, severity of impairment and cognitive capacity are available in Gatehouse et al (2003).

Self-report measures

The speech tests described above sample a relatively rich set of conditions. A similar approach was adopted in the self-report domain by employing a range of questionnaire instruments to evaluate different aspects of hearing aid outcome (Gatehouse, 1994; Noble, 1998). Each self-report instrument was completed in a paper and pencil interview format at the end of the 10-week trial with each hearing aid fitting.

The Abbreviated Profile of Hearing Aid Benefit (APHAB) is an established instrument for the assessment of auditory disability (Cox, 1997; Cox & Alexander, 1995). It consists of

24 statements about communication difficulty, each of which is rated on a seven-point response scale, separately for unaided and aided listening. Benefit scores are constructed by taking the difference between the aided and unaided responses. The APHAB yields four subscales corresponding to:

1. benefit in ease of communication;
2. benefits in background noise;
3. benefit in environments with reverberation; and
4. reduction in the aversiveness of sounds.

Table 6 shows the APHAB benefit scores aggregated over the five hearing aid fittings. The results are comparable to published data (Cox & Alexander, 1995). Note that the aversiveness scores take negative values, indicating that hearing aids have disadvantages in this respect.

Aspects of satisfaction with hearing aid fittings were assessed using the Satisfaction with Amplification in Daily Life questionnaire (SADL, Cox & Alexander, 1999). The standard questionnaire consists of 15 statements, each of which is rated on a seven-point response scale, and is completed with reference to aided listening. The SADL yields subscales for:

1. positive effect;
2. negative features;
3. service and cost; and
4. personal image.

In the context of this experiment where the hearing aids and services were free of charge, and the repeat fittings were conducted by the same audiologists, the items regarding service and cost were omitted, and only three subscales constructed.

Table 6. Mean (and standard deviation) of the scores in the self-report domain aggregated over the five fittings for the benefit score on the APHAB, SADL scores, GHABP scores, HAPQ scores, and scores from the visual analogue ratings

	Mean	*Standard deviation*
Abbreviated profile of hearing aid benefit		
Ease of communication	32.7	18.7
Background noise	37.3	16.9
Reverberation	33.4	18.3
Aversiveness	−19.3	15.7
Satisfaction with amplification in daily life		
Positive effect	5.1	1.0
Negative features	3.7	1.2
Personal image	5.3	0.9
Glasgow Hearing Aid Benefit Profile		
Use	82.4	16.9
Benefit	65.1	16.3
Satisfaction	65.0	17.0
Hearing Aid Performance Questionnaire		
Speech variations	6.0	2.2
Environmental sounds	6.0	2.3
Intense sounds	5.1	2.4
Direct visual analogue rating scales		
Perceived intelligibility	5.4	2.6
Listening comfort	6.2	1.9
Overall rating	6.7	1.4

Table 6 shows the SADL scores aggregated over the five fittings. The results are comparable to published values obtained in other contexts (Cox & Alexander, 1999).

The Glasgow Hearing Aid Benefit Profile (GHABP) can be used to assess (among other dimensions) hearing aid use, hearing aid benefit, and satisfaction with hearing aids, over a range of prespecified and subject-specified listening circumstances leading to material disability (Gatehouse, 1999). The existence of the prespecified listening circumstances, and the nomination of the subject-specified circumstances was established prior to the start of the formal experiment, and held constant thereafter. Table 6 shows the GHABP scores aggregated over the five fittings, which are comparable to published UK data (Gatehouse, 1999).

The APHAB, SADL, and GHABP are established self-report instruments for the assessment of benefits of amplification. For each of these outcome measures, the present mean data are comparable to published values and support the adequacy of the hearing aid fittings employed in the experiment. In addition, a new questionnaire, termed the Hearing Aid Performance Questionnaire (HAPQ), was piloted and developed with the specific aim of differentiating linear and nonlinear fittings. Table 7 contains an abbreviated description for each of the 26 items which make up the HAPQ. The full text for item 1 is '*How good or bad is your hearing aid at dealing with the following circumstances and environments? – When two people with very different voice levels are talking*'. Each item was rated on a visual analogue scale from 0–10 with text labels of very poor, rather poor, acceptable, rather good, and very good at scale points 1, 3, 5, 7, and 9 respectively (Figure 1). Factor analysis of the data from the HAPQ yields 3 principal components accounting for 76% of the variance. Following varimax rotation, Table 7 identifies the HAPQ items with factor loadings >0.7 for each of the rotated factors. Inspection of Table 7 suggests that factor 1 can be interpreted as 'variations in speech', factor 2 as 'variations in nonspeech environmental sounds', and factor 3 as 'environments with intense sounds'. These three subscales were then constructed consisting of the unweighted average of the HAPQ items with factor loadings >0.7. Table 6 shows the HAPQ subscale scores aggregated over the 5 fittings. Note that only those HAPQ items with a factor loading >0.7 contribute to a subscale. This differs from a factor-score approach where each variable contributes to the factor score irrespective of its factor loading.

Finally, three direct visual analogue rating scales were administered using the same graphical layout as for the HAPQ for:

1. rating of speech clarity;
2. rating of listening comfort; and
3. an overall rating of the performance of each hearing aid fitting (Gabrielsson & Sjogren, 1979).

Together, the APHAB, SADL, GHABP, HAPQ, and direct ratings yield the 16 outcome scales in the self-report domain shown in Table 6.

Data reduction

The rich set of outcome measures in the self-report and laboratory-performance domains furnishes this experiment

Table 7. Shortform description for each item on the Hearing Aid Performance Questionnaire (HAPQ), and its allocation to the Speech variations, Environmental sounds and Intense sounds subscales

	HAPQ subscale		
Item from the Hearing Aid Performance Questionnaire	*Speech variations*	*Environmental Sounds*	*Intense Sounds*
Two people talking with different voice levels	•		
One person talking at comfortable level	•		
People talking quickly	•		
Listening to loud voices	•		
Listening to music		•	
Listening to music at high volume			•
Making speech as understandable as possible	•		
Making speech as natural as possible	•		
Short loud sounds not uncomfortable			•
Continuous loud sounds not distressing			•
Detect sounds that change in level		•	
Hear tone of speaker's voice	•		
Hear changes in speaker's voice	•		
Keep background noise to a minimum		•	
Allow own voice to sound natural	•		
Able to catch beginning of new sentence	•		
Make aware of kind of room		•	
Comfortable listening in car or bus			•
Understand speech in car or bus	•		
Hear traffic noise without being invasive			•
Hear alerting sounds		•	
Quiet sounds loud enough		•	
Hear what people say when at other side of room	•		
Judge mood of speaker	•		
Aware of accents or dialects	•		
Recognise different voices	•		

with opportunities for differences between the hearing aid fittings to emerge, though at the potential expense of increased data complexity. It was therefore desirable to reduce the number of outcome measures to a more manageable set, via a factor analytical approach. All data reduction and analysis was performed using the Statistical Package for the Social Sciences (SPSS) version 11. The 16 self-report variables for the amalgamation of each of the five fittings and 50 subjects were subjected to a principal components analysis, operating on the covariance rather than the correlation matrix. For the default procedure whereby Eigenvalues greater than unity result in factor extraction, four principal components were identified, though the Eigenvalue of the fourth factor was marginal at 1.06. Furthermore, when 10% of the data was randomly deleted, the fourth factor oscillated around the criterion value, depending on which 10% was deleted. The fourth factor was not stable. The procedure was repeated with a manual override that only three factors be extracted, provided the criterion value was satisfied. With this restriction in place, random deletion of 10% of the data did not disturb the factor extraction or its structure, and

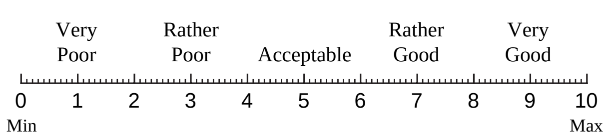

Figure 1. Response format for the visual analogue scales.

hence yielded a stable result. The three-factor solution accounts for 74% of the variance in the 16 input variables.

Although an orthogonality constraint is often applied in factor analyses, there is no good reason why dimensions of hearing aid outcome should be orthogonal (that is uncorrelated). Indeed, one might in general expect that different aspects of outcome would be correlated, either positively or negatively. The three extracted factors were therefore subjected to an oblique rotation, which does not insist on orthogonal solutions, but allows varying degrees of inter-relationships. The loading of each of the 16 input variables on the three resulting factors is shown in Table 8, with factor loadings >0.7 identified in bold, underlined font. Factor 1 is loaded heavily by APHAB ease of communication, APHAB aversiveness, SADL negative features, HAPQ environmental sounds, and direct rating of listening comfort. We interpret this factor in terms of *Listening Comfort* and use that label hereafter. Factor 2 is loaded most heavily by APHAB aversiveness, SADL positive effect, SADL personal image, GHABP satisfaction, HAPQ speech variations, HAPQ environmental sounds, and direct rating overall. We interpret this factor as *Satisfaction* with the hearing aid fitting. Factor 3 is loaded by APHAB ease of communication, APHAB background noise, APHAB reverberation, GHABP benefit, HAPQ speech variations, HAPQ environmental sounds, and direct rating of speech clarity. We interpret this factor as *Reported Intelligibility*. It should be remembered that the labels attached to each of these factors are an external construction rather than

Table 8. Factor loadings for each of the sixteen outcome scales and subscales on the three factors in the self-report domain with factor loadings >0.7 identified in bold, underlined font

	Self-report factor		
	Listening Comfort	*Satisfaction*	*Reported Intelligibility*
Abbreviated profile of hearing aid benefit			
Ease of communication	**<u>.81</u>**	.61	**<u>.84</u>**
Background noise	.60	.62	**<u>.78</u>**
Reverberation	.57	.61	**<u>.75</u>**
Aversiveness	**<u>.72</u>**	**<u>.79</u>**	.52
Satisfaction with amplification in daily life			
Positive effect	.57	**<u>.81</u>**	.57
Negative features	**−.75**	−.68	−.58
Personal image	.58	**<u>.79</u>**	.46
Glasgow Hearing Aid Benefit Profile			
Use	.24	.44	.22
Benefit	.57	.62	**<u>.82</u>**
Satisfaction	.61	**<u>.85</u>**	.68
Hearing Aid Performance Questionnaire			
Speech variations	.58	**<u>.82</u>**	**<u>.85</u>**
Environmental sounds	**<u>.86</u>**	**<u>.81</u>**	**<u>.70</u>**
Intense sounds	**<u>.77</u>**	.59	.62
Direct visual analogue rating scales			
Speech clarity	.62	.52	**.72**
Listening comfort	.87	.66	.65
Overall rating	.69	**<u>.81</u>**	.68

an inevitable consequence of the analysis. They are, however, consistent with the constructs which underpin each of the 16 variables entering the data reduction procedure. Each of the three benefit factor scores was derived using the SPSS regression option which yields each factor score represented as a standardised z-score with a mean of zero and unity standard deviation (see endnote for a brief discussion of the nomenclature).

The benefit scores for speech intelligibility via the FAAF test were reduced to a single factor, which accounts for 79% of the variance in the 32 input benefit scores. This single factor approach will obscure the ways in which different fittings yield differential benefits in different test conditions, but is parsimonious in that only differences between fittings that are pervasive across test conditions will emerge. Consideration of the ways in which differences between the fittings cluster in particular test conditions is available in Gatehouse et al (2003).

In general the reduction of the questionnaire data to three self-report factors, and the FAAF benefit scores into a single factor was designed to increase statistical reliability and stability, as well as facilitating identification of overall differences. Table 9 shows the correlations between the four benefit factors, aggregated over the 50 subjects and five fittings (i.e. 250 data points for each correlation coefficient). Both *Listening Comfort* and *Reported Intelligibility* correlate with *Satisfaction* with coefficients 0.75–0.79, but are less correlated with each other (0.58). The *Speech Test* factor correlates at 0.66 with the factor in the self-report domain interpreted as *Reported Intelligibility*, at 0.40 with *Satisfaction*, and at 0.28 with *Listening Comfort*. This pattern of correlation lends support to the interpretations and labels attached to the various factors.

Table 9. Correlation coefficients between the four eventual factors used for the substantive analysis

	Satisfaction	*Reported Intelligibility*	*Speech Test Performance*
Listening comfort	.79	.58	.28
Satisfaction		.75	.40
Reported intelligibility			.66

The reduction of the outcome data to four factors should not be construed as implying that all the dimensions of hearing aid fitting outcome are included in this experiment, and no others exist. Rather, with the outcome measures used it was possible to distinguish these factors. Other outcome measures might have revealed other dimensions by probing other aspects of hearing function, for example behaviour in complex and dynamically varying listening environments.

Results

Group Differences

Each of the 50 listeners experienced each of the five fittings for a 10-week period in a randomised block design. Outcome was assessed on the four derived benefit factors of:

1. *Listening Comfort*;
2. *Satisfaction*;
3. *Reported Intelligibility*; and
4. *Speech Test Benefit*.

Figure 2 contains the mean and 95% confidence interval for each of the five fittings, separately for each of the four benefit factors.

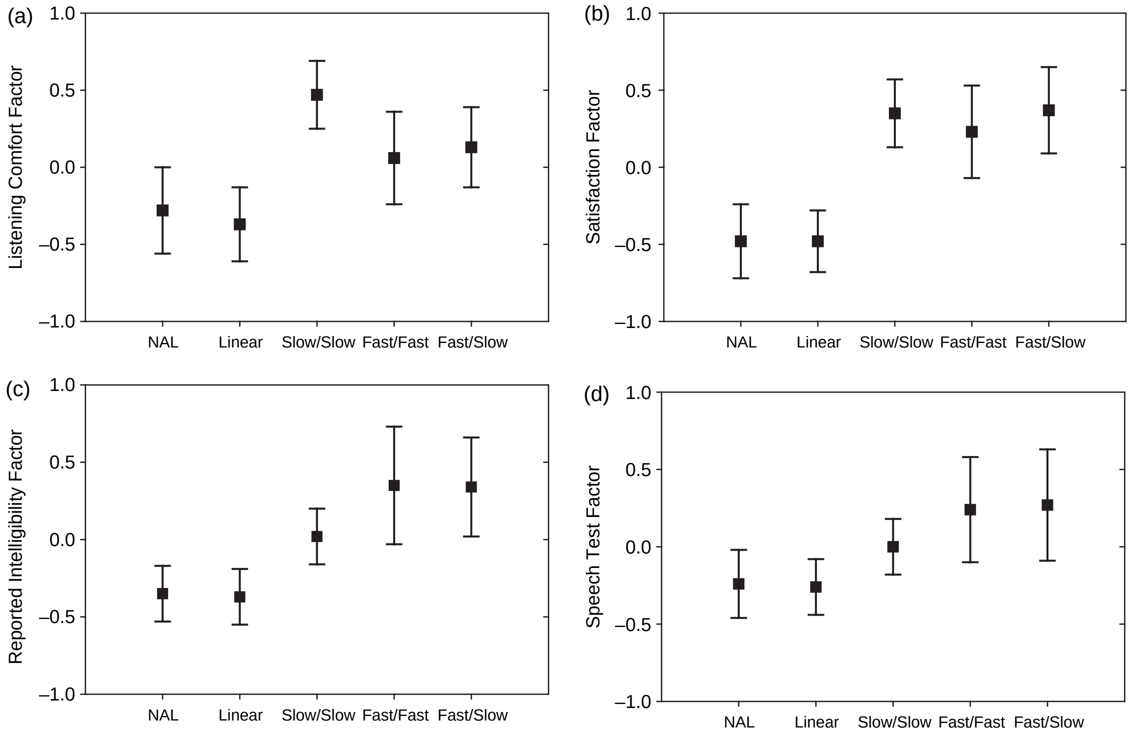

Figure 2. Mean (and 95% confidence interval) of (a) *Listening Comfort* factor, (b) *Satisfaction* factor, (c) *Reported Intelligibility* factor and (d) *Speech Test* factor for the five hearing aid fittings. The Y-axis is a standardised z-score.

Prior to any formal statistical consideration, a number of trends are evident. Across the four benefit factors, there would not appear to be any group differences between the *NAL-RP* and *LINEAR* fittings, even though they differ in frequency response, MPO setting, and availability of a volume control. Thus any differences that emerge between any of the nonlinear and linear fittings cannot be ascribed to these factors. The presence of the two linear control conditions allows this interpretation, which would not have been possible otherwise. The presence of the *NAL-RP* condition allows a reference which has wide application and high face validity. Figure 2 suggests that for *Listening Comfort*, each of the nonlinear fittings provides superior outcome to the two linear fittings, with an advantage for the *SLOW-SLOW* fitting over both *FAST-FAST* and *FAST-SLOW*. For the *Satisfaction* benefit factor, each of the nonlinear fittings appears to outperform the two linear fittings, with no differentiation amongst them. For both *Reported Intelligibility* and for *Speech Test Performance* the three nonlinear fittings provide greater benefit than the two linear fittings, with both *FAST-FAST* and *FAST-SLOW* delivering more benefit than *SLOW-SLOW*.

The magnitudes of the differences in Figure 2 require interpretation. The benefit factors are dimensionless standardised z-scores, with means of zero and standard deviations of unity. Unlike % correct on a speech test, or even scale points on an APHAB or SADL score, they have no familiarity or anchoring in audiology. Hence some method to appreciate the likely clinical and scientific significance (as opposed to statistical significance) of any differences is required. In the clinical trials literature (Pocock, 1990) population standard deviations (PSD) are used to represent effect sizes, with 0.33 PSD considered a modest but worthwhile effect, 0.67 PSD a substantial effect, and 1.00 PSD a large effect. Thus any effect size which achieves a magnitude of >0.33 PSD can be regarded as material and of interest. In the context of this experiment, the standardised z-scores that form the four benefit factors represent the population of listeners and hearing aid fittings in the design. Consequently, the effect sizes reported in this paper are expressed directly in PSD units. The above representation in terms of population standard deviation is used here to calibrate the potential meaning of any differences which might achieve statistical significance in a formal analysis.

As each of the 50 listeners experienced each of the five fittings in a within-subject crossover design, the 95% confidence intervals in Figure 2 do not necessarily reveal differences that achieve statistical significance. For each of the four benefit factors, a repeated measures analysis of variance (ANOVA) was conducted, with hearing aid fitting specified as a within-subject factor. As the fitting had five data levels, simple pairwise comparisons via t-tests might capitalise upon chance and assign significance erroneously. Post hoc comparisons using Bonferroni corrections were employed to avoid this. Tables 10, 11, 12, and 13 summarise the ANOVAs for the *Listening Comfort*, *Satisfaction*, *Reported Intelligibility*, and *Speech Test Performance* benefit factors respectively. Each table contains, for each of the five fittings, the mean and standard error of the benefit factor, the mean difference from each of the other four fittings, the statistical significance of that difference after Bonferroni correction, and the 95% confidence interval of the difference between the hearing aid fittings. Since the Bonferroni correction is conservative, any statistically significant findings that result are highly unlikely to be erroneous.

Table 10. Descriptive statistics (mean and standard error) of the *Listening Comfort* factor for each of the five fittings, and results of post hoc multiple comparisons (after Bonferroni correction) between each fitting and each of the four remaining alternatives

Listening Comfort factor					
				95% Confidence interval	
Mean (Standard error)		*Mean difference*	*Sig.*	*Lower bound*	*Upper bound*
NAL-RP	*LINEAR*	.13	.66	−.13	.39
−0.25	*SLOW-SLOW*	−.73**	.01	−.99	−.46
(0.14)	*FAST-FAST*	−.29*	.02	−.55	−.03
	FAST-SLOW	−.35**	.01	−.61	−.09
LINEAR	*NAL-RP*	−.13	.66	−.39	.13
−0.38	*SLOW-SLOW*	−.85**	.01	−1.11	−.59
(0.12)	*FAST-FAST*	−.42**	.01	−.68	−.16
	FAST-SLOW	−.48**	.01	−.74	−.22
SLOW-SLOW	*NAL-RP*	.73**	.01	.46	.99
0.48	*LINEAR*	.85**	.01	.59	1.11
(0.11)	*FAST-FAST*	.44**	.01	.18	.70
	FAST-SLOW	.38**	.01	.12	.64
FAST-FAST	*NAL-RP*	.29*	.02	.03	.55
0.04	*LINEAR*	.42**	.01	.16	.68
(0.16)	*SLOW-SLOW*	−.44**	.01	−.70	−.18
	FAST-SLOW	−.06	.97	−.32	.20
FAST-SLOW	*NAL-RP*	.35**	.01	.09	.61
0.10	*LINEAR*	.48**	.01	.22	.74
(0.14)	*SLOW-SLOW*	−.38**	.01	−.64	−.12
	FAST-FAST	.06	.97	−.20	.32

*The mean difference is significant at the .05 level.
**The mean difference is significant at the .01 level.

Table 11. Descriptive statistics (mean and standard error) of the *Satisfaction* benefit factor for each of the five fittings, and statistical results of post hoc multiple comparisons (after Bonferroni correction) between each fitting and each of the four remaining alternatives

Satisfaction benefit factor					
				95% Confidence interval	
Mean (Standard error)		*Mean difference*	*Sig.*	*Lower bound*	*Upper bound*
NAL-RP	*LINEAR*	.06	.97	−.19	.31
−0.44	*SLOW-SLOW*	−.83**	.01	−1.08	−.58
(0.11)	*FAST-FAST*	−.62**	.01	−.87	−.37
	FAST-SLOW	−.81**	.01	−1.06	−.56
LINEAR	*NAL-RP*	−.06	.97	−.31	.19
−0.50	*SLOW-SLOW*	−.89**	.01	−1.14	−.64
(0.10)	*FAST-FAST*	−.68**	.01	−.93	−.43
	FAST-SLOW	−.87**	.01	−1.12	−.62
SLOW-SLOW	*NAL-RP*	.83**	.01	.58	1.08
0.39	*LINEAR*	.89**	.01	.64	1.14
(0.12)	*FAST-FAST*	.21	.14	−.04	.46
	FAST-SLOW	.02	.10	−.23	.27
FAST-FAST	*NAL-RP*	.62**	.01	.37	.87
0.18	*LINEAR*	.68**	.01	.43	.93
(0.16)	*SLOW-SLOW*	−.21	.14	−.46	.04
	FAST-SLOW	−.19	.23	−.44	.06
FAST-SLOW	*NAL-RP*	.81**	.01	.56	1.06
0.37	*LINEAR*	.87**	.01	.62	1.12
(0.16)	*SLOW-SLOW*	−.02	.10	−.27	.23
	FAST-FAST	.19	.23	−.06	.44

*The mean difference is significant at the .05 level.
**The mean difference is significant at the .01 level.

Table 12. Descriptive statistics (mean and standard error) of the *Reported Intelligibility* factor for each of the five fittings, and statistical results of post hoc multiple comparisons (after Bonferroni correction) between each fitting and each of the four remaining alternatives

Reported Intelligibility factor					
				95% Confidence interval	
Mean (Standard error)		*Mean difference*	*Sig.*	*Lower bound*	*Upper bound*
NAL-RP	*LINEAR*	.06	.98	−.20	.31
−0.35	*SLOW-SLOW*	−.37**	.01	−.62	−.11
(0.08)	*FAST-FAST*	−.65**	.01	−.90	−.39
	FAST-SLOW	−.78**	.01	−1.03	−.52
LINEAR	*NAL-RP*	−.06	.98	−.31	.20
−0.40	*SLOW-SLOW*	−.42**	.01	−.68	−.16
(0.10)	*FAST-FAST*	−.70**	.01	−.96	−.44
	FAST-SLOW	−.83**	.01	−1.09	−.58
SLOW-SLOW	*NAL-RP*	.37**	.01	.11	.62
0.02	*LINEAR*	.42**	.01	.16	.68
(0.10)	*FAST-FAST*	−.28*	.02	−.54	−.03
	FAST-SLOW	.41**	.01	−.67	−.16
FAST-FAST	*NAL-RP*	.65**	.01	.39	.90
0.30	*LINEAR*	.70**	.01	.44	.96
(0.09)	*SLOW-SLOW*	.28*	.02	.03	.54
	FAST-SLOW	−.13	.64	−.39	.13
FAST-SLOW	*NAL-RP*	.78**	.01	.52	1.03
0.43	*LINEAR*	.83**	.01	.58	1.09
(0.17)	*SLOW-SLOW*	.41**	.01	.16	.67
	FAST-FAST	.13	.64	−.13	.39

*The mean difference is significant at the .05 level.
**The mean difference is significant at the .01 level.

Inspection of Tables 10, 11, 12, and 13 confirms the absence of any significant differences between the *NAL-RP* and *LINEAR* fittings for any of the four benefit factors. Table 10 shows that all three nonlinear fittings are superior to the two linear fittings for *Listening Comfort*. The advantage of *SLOW-SLOW* over the two linear fittings at 0.73–0.85 is substantial, and over *FAST-FAST* and *FAST-SLOW* at 0.38–0.44 is material. Although inferior to *SLOW-SLOW*, both *FAST-FAST* and *FAST-SLOW* show material advantages of 0.29–0.35 over the linear fittings for *Listening Comfort*.

Table 11 shows that all three of the nonlinear fittings offer substantially greater *Satisfaction* over linear fittings of magnitude 0.62–0.89, but do not differ among themselves. In contrast to the two different patterns in Tables 10 and 11, Tables 12 and 13 show advantages in both *Reported Intelligibility* and *Speech Test Performance* for the *FAST-FAST* and *FAST-SLOW* fittings. The advantages over the two linear rationales are substantial (0.55–0.83), and still material when compared to *SLOW-SLOW* (0.24–0.41).

The results clearly demonstrate the group advantages of nonlinear over linear fittings for all four benefit factors, but also show that the magnitude and patterns of the advantages depend on the choice of outcome domain. Furthermore, the relative advantages for the different nonlinear fittings emerge differently, with *SLOW-SLOW* optimal for *Listening Comfort*, and with *FAST-FAST* and *FAST-SLOW* optimal for *Reported Intelligibility* and *Speech Test Performance*.

Individual optima

The foregoing analysis has considered group differences. Of equal interest is the pattern of fittings that provide the most beneficial amplification for individual listeners. Given the results of the group analysis, these patterns might also be expected to vary across outcome domain. For each listener, there were four benefit factor values for each of the five fittings. For each of the benefit factors separately, the given listener's scores for each of the fittings were inspected in order to elucidate patterns of individual optimal fittings. The fitting which achieved the highest score was designated the 'winner' on that benefit factor, the fitting which achieved the second highest score was designated '2nd', and so on until the fitting with the lowest score was designated the 'loser'. The process was repeated for each benefit factor on each listener, and the results are shown in Tables 14, 15, 16, and 17. Each of the tables contains:

1. the number of occasions (out of the 50 listeners) that each fitting achieved a particular rank;
2. the number of occasions above expressed as a percentage (each row then sums to 100%);
3. the mean of the benefit factor scores for each intersection of fitting and rank; and
4. the standard error of that factor score.

In viewing Tables 14 through 17 it should be borne in mind that differences in row percentages of 7% between any two columns achieve statistical significance at $p < 0.01$ (Wilcoxon signed rank

Table 13. Descriptive statistics (mean and standard error) of the *Speech Test* factor for each of the five fittings, and results of post hoc multiple comparisons (after Bonferroni correction) between each fitting and each of the four remaining alternatives

Speech Test factor					
				95% Confidence interval	
Mean (Standard error)		*Mean difference*	*Sig*	*Lower bound*	*Upper bound*
NAL-RP	*LINEAR*	−.01	1.00	−.27	.25
−0.30	*SLOW-SLOW*	−.32**	.01	−.58	−.05
(0.11)	*FAST-FAST*	−.56**	.01	−.82	−.30
	FAST-SLOW	−.61**	.01	−.87	−.34
LINEAR	*NAL-RP*	.01	1.00	−.25	.27
−0.29	*SLOW-SLOW*	−.31**	.01	−.57	−.04
(0.10)	*FAST-FAST*	−.55**	.01	−.81	−.29
	FAST-SLOW	−.60**	.01	−.86	−.33
SLOW-SLOW	*NAL-RP*	.32**	.01	.05	.58
0.02	*LINEAR*	.31**	.01	.04	.57
(0.10)	*FAST-FAST*	−.24*	.05	−.51	−.02
	FAST-SLOW	−.29*	.02	−.55	−.08
FAST-FAST	*NAL-RP*	.56**	.01	.30	.82
0.26	*LINEAR*	.55**	.01	.29	.81
(0.17)	*SLOW-SLOW*	.24*	.05	.02	.51
	FAST-SLOW	−.05	.99	−.31	.22
FAST-SLOW	*NAL-RP*	.61**	.01	.34	.87
0.31	*LINEAR*	.60**	.01	.33	.86
(0.19)	*SLOW-SLOW*	.29*	.02	.08	.55
	FAST-FAST	.05	.99	−.22	.31

*The mean difference is significant at the .05 level.
**The mean difference is significant at the .01 level.

Table 14. Breakdown of the within-listener rankings of the five fittings for the *Listening Comfort* factor (Each cell also contains the mean and standard error [SE] of the *Listening Comfort* scores which achieve the given rank.)

Listening Comfort factor		*NAL-RP*	*LINEAR*	*SLOW-SLOW*	*FAST-FAST*	*FAST-SLOW*
Winner	Count	12	4	21	9	4
	%	24%	8%	42%	18%	8%
	Mean	.82	.78	.85	1.81	.88
	SE of mean	.23	.34	.13	.27	.26
2nd	Count	5	14	6	9	16
	%	10%	28%	12%	18%	32%
	Mean	.06	.35	.47	.48	.78
	SE of mean	.22	.17	.37	.25	.20
3rd	Count	2	5	15	12	16
	%	4%	10%	30%	24%	32%
	Mean	−.41	−.07	.46	−.40	.22
	SE of mean	.72	.17	.20	.16	.14
4th	Count	19	13	7	5	6
	%	38%	26%	14%	10%	12%
	Mean	−.39	−.85	−.40	−.67	−.63
	SE of mean	.14	.13	.18	.17	.35
Loser	Count	12	14	1	15	8
	%	24%	28%	2%	30%	16%
	Mean	−1.19	−1.10	−.98	−.69	−1.33
	SE of mean	.21	.16	n/a	.13	.23

Table 15. Breakdown of the within-listener rankings of the five fittings for the *Satisfaction* benefit factor (Each cell also contains the mean and standard error [SE] of the *Satisfaction* scores which achieve the given rank.)

Satisfaction factor		*NAL-RP*	*LINEAR*	*SLOW-SLOW*	*FAST-FAST*	*FAST-SLOW*
Winner	Count	6	3	23	9	9
	%	12%	6%	46%	18%	18%
	Mean	.80	.40	.61	1.23	1.54
	SE of mean	.29	.30	.14	.41	.29
2nd	Count	6	6	8	11	19
	%	12%	12%	16%	22%	38%
	Mean	−.09	.27	.50	.44	.88
	SE of mean	.26	.32	.28	.32	.16
3rd	Count	6	9	10	15	10
	%	12%	18%	20%	30%	20%
	Mean	−.29	−.20	.65	.27	−.10
	SE of mean	.20	.17	.27	.20	.12
4th	Count	15	18	5	6	6
	%	30%	36%	10%	12%	12%
	Mean	−.59	−.70	−.45	−.55	−.45
	SE of mean	.13	.11	.24	.28	.16
Loser	Count	17	14	4	9	6
	%	34%	28%	8%	18%	12%
	Mean	−.92	−.97	−.68	−.86	−1.37
	SE of mean	.17	.19	.25	.23	.19

test) and hence are unlikely to be chance findings (even in the absence of corrections for multiple comparisons).

Table 14 shows that for *Listening Comfort*, the fitting *SLOW-SLOW* was an optimum option (winner) for 21 of the 50 listeners (42%), and for only one listener was it the worst fitting that could have been chosen. For 24% and 8% *NAL-RP* and *LINEAR* respectively were optimum, but for 24% and 28% respectively were the poorest option. In fittings containing some

Table 16. Breakdown of the within-listener rankings of the five fittings for the *Reported Intelligibility* factor (Each cell also contains the mean and standard error [SE] of the *Reported Intelligibility* scores which achieve the given rank.)

Reporter intelligibility factor		*NAL-RP*	*LINEAR*	*SLOW-SLOW*	*FAST-FAST*	*FAST-SLOW*
Winner	Count	**7**	6	**7**	**12**	18
	%	14%	12%	14%	24%	36%
	Mean	.05	.85	.46	2.07	.98
	SE of mean	.15	.18	.27	.36	.15
2nd	Count	9	7	10	10	4
	%	18%	14%	20%	20%	28%
	Mean	.06	−.10	.15	.50	1.25
	SE of mean	.17	.17	.15	.16	.27
3rd	Count	6	14	18	7	5
	%	12%	28%	36%	14%	10%
	Mean	.05	−.24	.21	.04	−.05
	SE of mean	.22	.08	.17	.21	.26
4th	Count	18	10	10	8	4
	%	36%	20%	20%	16%	8%
	Mean	−.51	−.71	−.51	−.38	−.70
	SE of mean	.12	.09	.11	.20	.11
Loser	Count	10	13	5	13	9
	%	20%	26%	10%	26%	18%
	Mean	−.94	−1.10	−.48	−.92	−1.19
	SE of mean	.07	.12	.17	.18	.20

Table 17. Breakdown of the within-listener rankings of the five fittings for the *Speech Test* factor (Each cell also contains the mean and standard error [SE] of the *Speech Test* scores which achieve the given rank

Satisfaction factor		*NAL-RP*	*LINEAR*	*SLOW-SLOW*	*FAST-FAST*	*FAST-SLOW*
Winner	Count	2	8	3	22	15
	%	4%	16%	6%	44%	30%
	Mean	.16	.48	.34	.99	1.56
	SE of mean	.80	.22	.24	.19	.27
2nd	Count	5	5	16	8	16
	%	10%	10%	32%	16%	32%
	Mean	.70	.00	.31	.72	.42
	SE of mean	.39	.27	.16	.40	.14
3rd	Count	13	6	20	5	6
	%	26%	12%	40%	10%	12%
	Mean	.11	−.40	−.02	−.22	.37
	SE of mean	.16	.20	.15	.29	.26
4th	Count	21	14	9	4	2
	%	42%	28%	18%	8%	4%
	Mean	−.39	−.28	−.40	−.94	−.97
	SE of mean	.14	.12	.24	.12	.57
Loser	Count	9	17	2	11	11
	%	18%	34%	4%	22%	22%
	Mean	−1.02	−.70	−.56	−.88	−1.36
	SE of mean	.16	.18	.17	.20	.25

aspect of fast-acting compression (*FAST-FAST* and *FAST-SLOW*), for 18% and 8% respectively of listeners this gave the best result for *Listening Comfort*, but for 30% and 16% respectively the poorest outcome. Although *SLOW-SLOW* was most often the 'winner' for *Listening Comfort*, the mean score (0.85) with which *SLOW-SLOW* achieved that rank was not conspicuously greater than for the other fittings. The nine listeners who found *FAST-FAST* to provide greatest *Listening Comfort* found *FAST-FAST* greatly more comfortable on average (1.01) than the 21 listeners whose optimum was *SLOW-SLOW*.

Table 15 shows the results for the *Satisfaction* benefit factor. Although the group analyses showed similar mean scores for the three nonlinear fittings (see Figure 2), the individual analyses identify *SLOW-SLOW* as the optimal fitting for 46% of listeners, against 18% for both *FAST-FAST* and *FAST-SLOW*. Note though that the mean *Satisfaction* scores with which the rank is achieved was 0.61 for *SLOW-SLOW* and 1.23–1.54 for fittings containing fast-acting compression. Again though, *SLOW-SLOW* was the poorest fitting for only 8% of listeners, against 34% and 28% respectively for the linear rationales (*NAL-RP* and *LINEAR*), and 18% and 12% respectively for fast compression (*FAST-FAST* and *FAST-SLOW*).

The patterns in Table 16 for *Reported Intelligibility*, and Table 17 for *Speech Test Performance* are similar. *FAST-FAST* was optimal for 24% and 44% respectively, while *FAST-SLOW* was optimal for 36% and 30% of listeners. Note though that *FAST-FAST* was the poorest choice for 26% and 22%, and *FAST-SLOW* for 18% and 22%. *SLOW-SLOW* was an infrequent optimum (14% and 6%), but was also an infrequent poorest choice (10% and 4%). When *SLOW-SLOW* was the optimum, the score (0.24–0.27) with which it achieves that rank is low compared with those for *FAST-FAST* and *FAST-SLOW* (0.98–2.07). Conversely, the losing score (−0.56 to −0.48) for *SLOW-SLOW* is superior to the corresponding scores (−0.88 to −1.36) for *FAST-FAST* and *FAST-SLOW*. The relative concentration of 'winners' with high scores and 'losers' with low scores is reflected in the wider 95% confidence intervals for *FAST-FAST* and *FAST-SLOW* for *Reported Intelligibility* and *Speech Test Performance* in Figure 2. This pattern of findings corresponds closely to those reported by van Toor and Verschuure (2002). Note also that there are still a substantial minority of listeners in Tables 16 and 17, for whom a linear fitting provides an optimum for speech intelligibility. Thus linear fittings would appear to be appropriate for material proportions of the population under study, irrespective of the domain of outcome.

Thus the AVC (*SLOW-SLOW*) fitting can be regarded as a 'safe' option as it is often good for *Listening Comfort* while rarely disastrous for *Reported Intelligibility* or *Speech Test Performance*. In contrast the WDRC fittings (*FAST-FAST* and *FAST-SLOW*) often deliver either optimal or poorest intelligibility to different individual listeners.

The differing distributions in the patterns of individual optima can be seen in Figure 3, with *SLOW-SLOW* prominent for *Listening Comfort* and *Satisfaction*, and concentration on *FAST-FAST* and *FAST-SLOW* for the differing aspects of speech intelligibility. The analyses to generate Tables 14–17 and Figure 3 take no account of the margin by which a fitting achieves its status as the 'winner', and the differing patterns could be simple reflections of immaterial marginal differences between rationales. Using the assumption that one-third of a population standard deviation yields an effect size which is clinically and scientifically material in magnitude, the analysis above was therefore repeated, but now only designating a

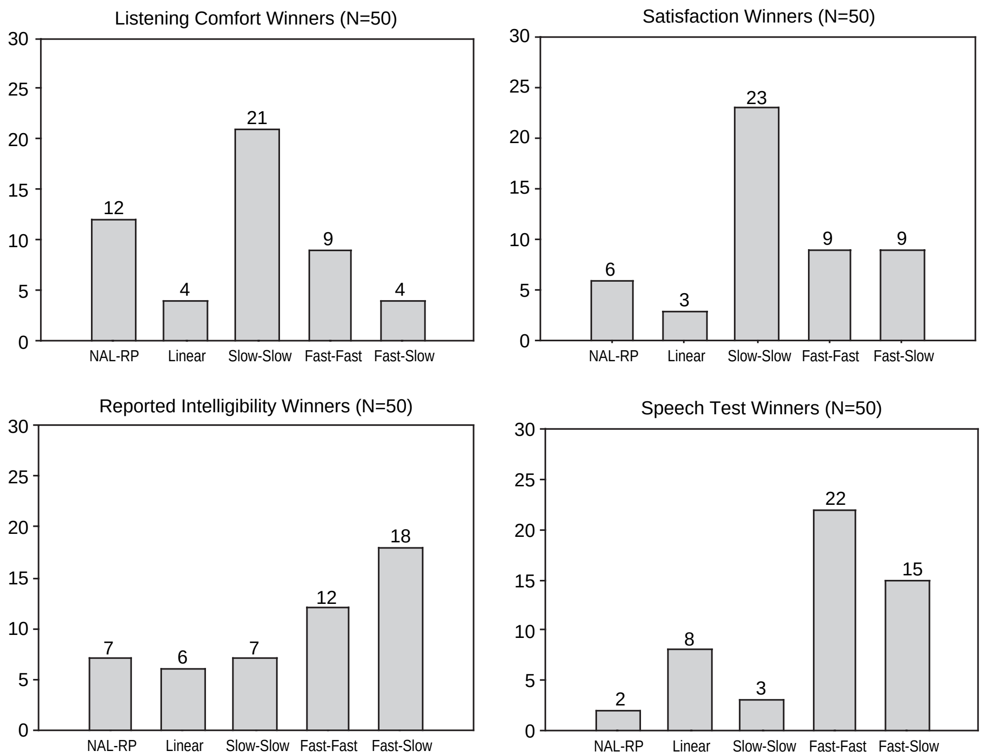

Figure 3. Distribution of the hearing aid fittings, which achieve the highest within-listener score for (a) *Listening Comfort,* (b) *Satisfaction,* (c) *Reported Intelligibility* and (d) *Speech Test* performance

particular fitting as a 'winner' if it achieved that status with a score on a benefit factor which exceeds that of the '2nd' fitting by more than 0.33. The results are shown in Figure 4. Of the 50 listeners, 31 achieved an optimum *Listening Comfort* fitting which was a 'clear winner', with corresponding values of 24 for *Satisfaction*, 32 for *Reported Intelligibility*, and 28 for *Speech Test Performance*. Comparison of Figures 3 and 4 shows that for three of the four benefit factors, the patterns of prominence were maintained. The exception is the *Satisfaction* benefit factor, where the prominence of *SLOW-SLOW* in Figure 3 was attenuated in Figure 4. This demonstrates a compression in scores across the rationales for cases where *SLOW-SLOW* is an apparent optimum for *Satisfaction*.

The analyses so far have demonstrated a clear dissociation between performance in the different outcome domains, both at the group and individual levels. It is of interest to compare how the individual optima vary for different outcomes. Table 18 cross-tabulates the individual optima for *Reported Intelligibility* against *Speech Test Performance*. Overall the correspondence was good, and lends further support to the interpretation of the third factor in the self-report domain in terms of speech intelligibility. In contrast, Table 19, which cross-tabulates the *Reported Intelligibility* factor and the *Listening Comfort* optima, shows greater dissociation, with two particular points to note. The most populated cell represents eleven listeners for whom the *SLOW-SLOW* fitting was best for *Listening Comfort*, and the *FAST-SLOW* fitting was best for *Reported Intelligibility*. Such a pattern lends support to the contention (at least for some listeners) that hearing aids with multiple programs offer advantages when listeners encounter differing environments with differing perceptual demands, though differences between such programs are usually formulated in terms of frequency-gain characteristics rather than time constants (Keidser et al, 1995; Keidser et al, 1996; Keidser, 1996). The five listeners in Table 19 for whom the *NAL-RP* fitting was best for *Listening Comfort* but for whom a nonlinear fitting was optimal for *Reported Intelligibility* are of interest. At the end of the experiment, all 50 listeners were given a choice of how the JUMP-1 should be programmed for continued use. These same five listeners were willing to accept (and continued to use) the device with a nonlinear fitting, but only if the volume control was activated. Thus there would appear to be a number of subjects for whom access to an active volume control is a requirement rather than an option. The figure of 5/50 = 10% requiring access to a volume control is if anything rather low when compared to the sparse and somewhat incompatible data available in the literature (Surr et al, 2001; Kochkin, 2003).

Discussion

The results reported from this experiment showed clear differences between the fittings at a group level. The existence of the differences, their magnitude and statistical and clinical implications are direct results of a robust design with adequate statistical leverage. A within-subject, within-device crossover design with appropriate masking of the features under comparison from both experimental test subject and investigator allowed effects to emerge which might otherwise be concealed

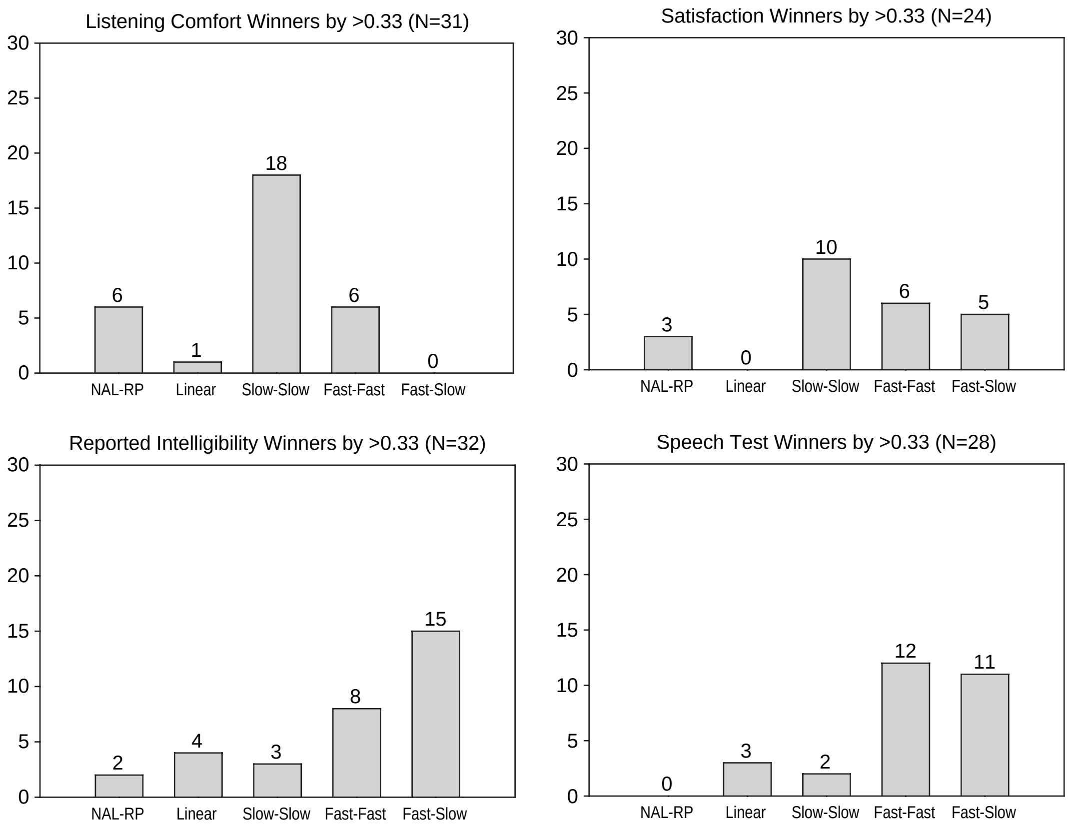

Figure 4. Distribution of the hearing aid fittings, which achieve the highest within-listener score for (a) *Listening Comfort*, (b) *Satisfaction*, (c) *Reported Intelligibility* and (d) *Speech Test* performance, but which also exceed the score of the runner-up by one third of a standard deviation in that benefit factor

by a variety of confounds. The incorporation of a wide range of self-report outcomes and conditions for laboratory evaluation of speech intelligibility allowed these differences to emerge. Not only is statistical power increased by incorporating a wide outcome set, in addition the idiosyncrasies and focus of one particular outcome or set of outcomes are avoided. A comprehensive experiment with appropriate design has allowed a set of findings to emerge. It should be noted that the absence of any significant contrasts between the two control fittings, *NAL-RP* and *LINEAR*, means that none of the findings can be attributed to the choice of values for the static frequency-gain characteristic or the MPO used in the experiment.

At the group level, the absolute and relative benefits of different fittings manifested themselves differently across the range of outcome domains. In general nonlinear fittings were superior to linear fittings for both reported and measured speech intelligibility, listening comfort and listener satisfaction at the group level. When considering listener satisfaction, the three nonlinear rationales were not differentiated, while for listening comfort, the *SLOW-SLOW* fitting scores more highly than either of the rationales (*FAST-FAST* or *FAST-SLOW*) containing fast time constants. In direct contrast to this, reports or laboratory measures of speech intelligibility were maximal for the rationales containing a fast time constant in at least one channel.

The findings of diversity at group level were replicated and indeed extended when the data are analysed for individual listeners. In addition to different fittings exhibiting differing

Table 18. Crosstabulation of the fittings which, for each listener, provide the maximum *Reported Intelligibility* factor score and the maximum *Speech Test* factor score

		Speech Test winner				
		NAL-RP	*LINEAR*	*SLOW-SLOW*	*FAST-FAST*	*FAST-SLOW*
Reported Intelligibility winner	*NAL-RP*	2	5	–	–	–
	LINEAR	–	3	3	–	–
	SLOW-SLOW	–	–	–	7	–
	FAST-FAST	–	–	–	12	–
	FAST-SLOW	–	–	–	3	15

Table 19. Crosstabulation of the fittings which, for each listener, provide the maximum *Reported Intelligibility* factor score and the maximum listening comfort factor score

		Comfort winner				
		NAL-RP	*LINEAR*	*SLOW-SLOW*	*FAST-FAST*	*FAST-SLOW*
Reported Intelligibility winner	*NAL-RP*	4	1	2	–	–
	LINEAR	3	3	–	–	–
	SLOW-SLOW	–	–	7	–	–
	FAST-FAST	2	–	1	8	1
	FAST-SLOW	3	–	11	1	3

performance dependent on the domain of benefit, within any given domain individual listeners showed distinct optima which diverged from the mean trend. Thus not only was there no fitting which on average is superior for all of the domains of outcome, but different listeners exhibited different optima even within a given domain of outcome. This finding is consistent with those reported by Neuman et al (1995), and van Toor and Verschuure (2002). Although fittings with fast time constants are on average optimum for speech intelligibility, there are significant numbers of listeners who are best served with either a slow-acting automatic volume control fitting or even a linear fitting. It should be noted that it is not only the case that such listeners did not gain from the fitting which on average is optimum, they would have been actively penalised by that fitting relative to their individual optimum fitting. Even though fittings with fast time constants were on average superior for speech intelligibility and there were many listeners for whom this was the individual optimum, there were significant numbers for whom such a fitting delivered significantly inferior performance.

Although the JUMP-1 platform used in this experiment is flexible and highly suited to the masked crossover design, it would be appropriate to consider what limitations its architecture might impose upon the generalisability of these results and the options that have been selected within the various implementations. Each of the choices that have been made for the parameters of the fittings (for example, the frequency-gain characteristic, the maximum power output, the compression ratios, the compression thresholds etc.) are all within the range of those which would be recognised as reasonable, even though any one individual or group might argue that a different specific choice might have been more appropriate. We contend that a reasonable default position should be that until a specific hypothesis can be mounted as to how and why a different choice of parameter might have led to a different outcome, then the results should be regarded as potentially generalisable – for example to more complex forms of compression characteristic. Indeed it is possible to argue for example, that the results might have become even more contrasted between the slow and fast implementations had a larger number of compression channels or more widely spaced time constants been employed. There are arguments in the literature that time constants longer than 600 ms or shorter than 20 ms might be appropriate (see for example Hansen 2002; Bray & Nilsson, 2002). Adoption of such an approach is likely to increase rather than decrease the differences between the rationales that have been identified in the current experiment.

Although the contrasts within the current data are restricted to time constants, we would contend that the general lessons regarding the ways in which significant and material results can emerge as a consequence of a robust and sensitive design are ones that should find application in other contexts. The current results are illuminating in that they provide valuable insight into the ways in which different fitting rationales can deliver benefits and disadvantages in various outcome domains, but there remains the substantial challenge of understanding the ways in which different hearing-impaired listeners with apparently similar characteristics exhibit different degrees of benefit and different individual optima. This challenge is taken up in the companion article (Gatehouse et al, 2006).

Acknowledgements

We would like to thank Patrick Howell and Jacqueline MacDonald who performed the hearing aid fittings and the testing of the patients. We are grateful to the hearing-impaired volunteers who generously gave of their time. We would also like to thank the three reviewers for their constructive comments on our original draft manuscript.

The hearing aids were supplied by Oticon A/S. All resources for the collection of data were supplied by core funding to the MRC Institute of Hearing Research by the United Kingdom Medical Research Council and the Chief Scientist Office of the Scottish Executive Health Department.

Note

Throughout this manuscript we use the term ‘benefit factor’ to describe the advantages that might accrue following successful hearing aid fitting. Within that general construct, advantages might accrue in terms of overall listener satisfaction, listening comfort and perceived intelligibility. We do not use the terms ‘benefit’ and ‘satisfaction’ as referring to fundamentally dissociated domains.

References

British Society of Audiology 1981. Recommended procedure for pure-tone audiometry using a manually operated instrument. *Br J Audiol*, 15, 213–216.

Bentler, R. & Nelson, J. 1997. Assessing release-time options in a two-channel AGC hearing aid. *J Am Acad Audiol*, 6, 43–51.

Bray, V. & Nilsson, M. 2002. What digital hearing aids can do: Another perspective. *Hear J*, 55:4 (60–61, 64).

Byrne, D. 1982. Theoretical approaches for hearing aid selection. In GA Studebaker & FH Bess (eds.) *The Vanderbilt hearing aid report: state of the art – research needs* 175–179. Upper Darby, PA: Monographs in Comtemporary Audiology.
Byrne, D. & Dillon, H. 1986. The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid (175–179). *Ear Hear*, 7(4), 257–265.
Byrne, D., Dillon, H., Ching, T., Katsch, R. & Keidser, G. 2001. NAL-NL1 procedure for fitting nonlinear hearing aids: characteristics and comparisons with other procedures. *J Am Acad Audiol*, 12(1), 37–51.
Cornelisse, L.E., Seewald, R.C. & Jamieson, D.G. 1995. The input/output formula: a theoretical approach to the fitting of personal amplification devices. *J Acoust Soc Am*, 97(3), 1854–1864.
Cox, R. 1995. Using loudness data for hearing aid selection: The IHAFF approach. *Hear J,* 48(2) 10, 39–44.
Cox, R.M. 1997. Administration and application of the APHAB. *Hear J*, 50(4), 32–48.
Cox, R.M. & Alexander, G.C. 1992. Maturation of hearing aid benefit subjective and objective measurements. *Ear Hear*, 13(3), 131–141.
Cox, R. & Alexander, G.C. 1995. The abbreviated profile of hearing aid benefit. *Ear & Hear*, 16(2), 176–186.
Cox, R.M. & Alexander, G.C. 1999. Measuring Satisfaction with Amplification in Daily Life: the SADL scale. *Ear Hear*, 20(4), 306–20.
Cox, R.M. & Moore, J.N. 1988. Composite speech spectrum for hearing aid gain prescriptions. *J Speech Hear Res*, 31, 102–107.
Dillon, H. 1996. Compression? Yes, but for low or high frequencies, for how or high intensities, and with what response times. *Ear Hear*, 17(4), 287–307.
Dillon, H. 2001. *Hearing aids*. Sydney: Boomerang Press.
Dreschler, W.A., Verschuure, H., Ludvigsen, C. & Westermann, S. 2001. ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing aid assessment. *Audiology*, 40, 148–157.
Elberling, C. 1999. Loudness scaling revisited. *J Am Acad Audiol*, 10(5), 248–260.
Elberling, C. & Nielsen, C. 1993. The dynamics of speech and the auditory dynamic range in sensorineural hearing impairment. In J. Beilin & G.R. Jensen (eds.) *Recent developments in hearing instrument technology. 15th Danavox Symposium*., pp. 99–133.
Foster, J.R. & Haggard, M.P. 1987. The Four Alternative Auditory Feature test (FAAF) – linguistic and psychometric properties of the material with normative data in noise. *Br J Audiol*, 21, 165–174.
Gabrielsson, A. & Sjogren, H. 1979. Perceived sound quality of hearing aids. *Scand Audiol*, 8, 159–169.
Gatehouse, S. 1989. Apparent auditory deprivation effects of late onset: the role of presentation level. *J Acoust Soc Am*, 86(6), 2103–2106.
Gatehouse, S. 1993. Role of perceptual acclimatisation in the selection of frequency responses for hearing aids. *J Am Acad Audiol*, 4(5), 296–306.
Gatehouse, S. 1994. Components and determinants of hearing aid benefit. *Ear Hear*, 15(1), 30–49.
Gatehouse, S. 1999. Glasgow Hearing Aid Benefit Profile: derivation and validation of a client-centered outcome measure for hearing aid services. *J Am Acad Audiol*, 10(2), 80–103.
Gatehouse, S., Elberling, C., & Naylor, G. 1999. Aspects of auditory ecology and psychoacoustic function as determinants of benefits from and candidature for nonlinear processing in hearing aids. In Auditory models and nonlinear hearing instruments. *18th Danavox Symposium*, 221–233.
Gatehouse, S. & Killion, M. 1993. HABRAT: Hearing aid brain rewiring accommodation time. *Hear Instrum*, 44(10), 29–32.
Gatehouse, S., Naylor, G. & Elberling, C. 2003. Benefits from hearing aids in relation to the interaction between the user and the environment. *Int J Audiol*, 42, S77–S85.
Gatehouse, S., Naylor, G., & Elberling, C. (2006). Linear and nonlinear hearing aids – 2. patterns of candidature. *Int J Audiol*, 45, 153–171.
Gatehouse, S. & Noble, W. 2004. The Speech, Spatial and Qualities of Hearing Scale (SSQ). *Int J Audiol*, 43, 85–99.
Hansen, M. 2002. Effects of multichannel compression time constants on subjectively perceived sound quality and speech intelligibility. *Ear Hear*, 23(4), 369–380.
International Electrotechnical Commission, 1983. Hearing aids. Part 2: Hearing aids with automatic gain control circuits. IEC 60118–2–1983, Geneva:IEC.
Jerlvall, L.B. & Lindblad, A.C. 1978. The influence of attack time and release time on speech intelligibility. A study of the effects of AGC on normal hearing and hearing impaired subjects. *Scand Audiol Suppl*, 6, 341–353.
Keidser, G. 1996. Selecting different amplification for different listening conditions. *J Am Acad Audiol*, 7(2), 92–104.
Keidser, G., Dillon, H. & Byrne, D. 1995. Candidates for multiple frequency response characteristics. *Ear Hear*, 16(6), 562–74.
Keidser, G., Dillon, H. & Byrne, D. 1996. Guidelines for fitting multiple memory hearing aids. *J Am Acad Audiol*, 7, 406–418.
Keissling, J. 1995. Loudness growth in sensorineural hearing loss – Consequences for hearing aid design and fitting. Audiol. *Acoust*, 34(2), 82–89.
King, A.B. & Martin, M.C. 1984. Is AGC beneficial in hearing aids? *Br J Audiol*, 18(1), 31–38.
Kochkin, S. 2003. Isolating the impact of the volume control on customer satisfaction. *Hearing Review*, 10(1), 26–35.
Launer, S., Holube, I., Hohmann, V. & Kollmeier, B. 1996. Categorical loudness scaling in hearing-impaired listeners – Can loudness growth be predicted from the audiogram? *Audiol Acoust*, 35(4), 156–163.
Lunner, T., Hellgren, J., Arlinger, S. & Elberling, C. 1997a. A digital filterbank hearing aid: Predicting user preference and performance for two signal progressing algorithms. *Ear Hear*, 18(1), 12–25.
Lunner, T., Hellgren, J., Arlinger, S. & Elberling, C. 1997b. A digital filterbank hearing aid: Three DSP algorithms – User preference and performance. *Ear Hear*, 18(5), 373–387.
Macrae, J. & Dillon, H. 1986. Gain, frequency response, and maximum output requirements for hearing aids. *J Rehab Res Dev*, 33(4), 363–376.
Moore, B.C.J. 1998. A comparison of four methods of implementing automatic gain control (AGC) in hearing aids. *Br J Audiol*, 22, 93–104.
Moore, B.C.J, Glasberg, B.R. & Stone, M.A. 1993. Effect on the speech reception threshold in noise of the recovery time of the compressor in the high-frequency channel of a two-channel aid. *Scand Audiol Suppl*, 38, 82–91.
Moore, B.C.J., Johnson, J.S., Clark, T.M. & Pluvinage, V. 1992. *Evaluation of a dual-channel full dynamic range compression system for people with sensorineural hearing loss Ear Hear*, 13, 349–370.
Moore, B.C.J., Stainsby, T.H., Alcantara, J.I. & Kuhnel, V. 2004. The effect on speech intelligibility of varying compression time constants in a digital hearing aid. *Int J Audiol*, 43, 300–409.
Muller, T.F., Harris, F.P. & Ellison, J.C. 2004. Effect of Release Time on Preferred Gain and Speech Acoustics. *J Am AcadAudiol*, 15, 605–615.
Nabelek, I.V. & Robinette, L.N. 1977. A comparison of hearing aids with amplitude compression. *Audiology*, 16(1), 73–85.
National Institute for Clinical Excellence (NICE), July 2000. Guidance on Hearing Aid Technology. ISBN 1-84257–023–4.
Naylor, G. 1997. Technical and audiological factors in the implementation and use of digital signal processing hearing aids. *Scand Audiol*, 26(4), 223–229.
Neuman, A.C., Bakke, M.H., Mackersie, C., Hellman, S. & Levitt, H. 1995. Effect of release time in compression hearing aids: Paired-comparison judgements of quality. *J Acoust Soc Am*, 98(6), 3182–3187.
Neuman, A.C., Bakke, M.H., Mackersie, C., Hellman, S. & Levitt, H. 1998. The effect of compression ratio and release time on the categorical rating of sound quality. *J Acoust Soc Am*, 103(5), 2273–2281.
Noble, W. 1998. *Self-assessment of Hearing and Related Functions*. Whurr Publishers Ltd.
Novick, M.L., Bentler, R.A., Dittberner, A. & Flamme, G.A. 2001. Effects of release time and directionality on unilateral and bilateral hearing aid fittings in complex sound fields. *J Am Acad Audiol*, 12, 534–544.
Pocock, S J. 1990. *Clinical Trials: A Practical Approach*. London: John Wiley & Sons Ltd.

Ricketts, T.A. & Bentler, R.A. 1996. The effect of test signal type and bandwidth on the categorical scaling of loudness. *J Acoust Soc Am*, 99, 2281–2287.

Schwartz, D., Lyregaard, P. & Lundh, P. 1988. Hearing aid selection for severe-to-profound hearing loss. *Hear J*, 41(2), 13–17.

Schweitzer, H.C. & Causey, G.D. 1977. The relative importance of recovery time in compression hearing aids. *Audiology*, 16(1), 61–72.

Stone, M.A., Moore, B.C.J., Alcantara, J.I. & Glasberg, B.R. 1999. Comparison of different forms of compression using wearable digital hearing aids. *J Acoust Soc Am*, 106, 3603–3619.

Surr, R.K., Cord, M.T. & Walden, B.E. 2001. Response of hearing aid wearers to the absence of a user-operated volume control. *Hear J*, 54(4), 32–36.

Valente, M. & van Vliet, D. 1997. The independent hearing aid fitting forum (IHAFF) protocol. *Trends in Amplification*, 2, 6–35.

Van Toor, T. & Verschuure, H. 2002. Effects of high-frequency emphasis and compression time constants on speech intelligibility in noise. *Int J Audiol*, 41, 379–394.