A Review of Past Research on Changes in Hearing Aid Benefit over Time

Christopher W. Turner, Larry E. Humes, Ruth A. Bentler, and Robyn M. Cox

Hearing aid benefit refers to a relative change in performance on a particular measure between aided and unaided listening conditions. A number of studies in recent years have investigated the hypothesis that hearing aid benefit increases over time after the initial fitting of the aid. Both objective (speech recognition) and subjective (questionnaire) measures have been used to measure hearing aid benefit. Some studies have reported a positive increase over time in group mean benefit, and some have reported no change in benefit, whereas none have reported a group mean negative change in benefit. However, individual subjects in these studies can show changes in benefit in either a positive or negative direction. The variability across subjects in each study has been large in comparison with the observed amount of benefit increase. In this review of the literature, it is argued that the studies present essentially similar results and the range of values across subjects in the various studies shows considerable overlap. Although there does appear to be a tendency for hearing aid benefit to increase over time, there are other, much stronger, factors influencing changes in hearing aid benefit that make it impossible at present to predict which patients will show a reliable increase (or decrease) in hearing aid benefit over time.

(Ear & Hearing 1996;17;14S–28S)

The general question "Does hearing aid benefit change over time?" is of considerable importance to those involved in clinical practice as well as hearing aid research, design, and manufacturing. Clinicians who dispense hearing aids need to know how much emphasis to place on initial evaluations of hearing aids. If benefit changes significantly over time, longer-term assessment and follow-up procedures may be more appropriate. For the manufacturer, the existence of hearing aid benefit changes over time means that the results of listener-based comparisons among various hearing aid models and/or circuit designs may depend on the timing of the evaluation. Researchers have assigned various names to changes in hearing aid benefit over time. For example, Gatehouse (1992) used the term "acclimatization," whereas Cox & Alexander (1992) used "maturation" to describe the effect. Both terms refer to an improvement in speech recognition over time as the patient learns to use the newly available speech cues made audible by the hearing aid amplification.

At least 18 studies providing information on the nature of acclimatization are available in the literature. Slightly more than half of these used actual hearing aids; the remainder used headphones. Nearly all used as subjects the most common type of first-time hearing aid users (usually middle-age to elderly, high-frequency hearing loss patients). Each study was generally designed to improve on the previous ones, with better control conditions and more accurate measures. Yet, the pattern of results observed does not change much across these studies. The results of the statistical tests in the various studies differ and the authors tend to make correspondingly different conclusions in their articles. This review, which overviews and integrates the results of the many studies on the topic, allows us to see that actually a rather consistent picture of the expected effect for the typical clinical hearing aid user has emerged.

This literature review is organized around several fundamental questions:

- Does hearing aid benefit change over time? If so, what abilities are changing and what are the magnitudes of the changes? What does this mean for individual patients and/or larger group studies (both previous and future) of hearing aids?
- Is there any clear or known relation between etiology or other individual subject factors and the amount of acclimatization?
- What is the time course of changes in benefit?
- Does hearing aid benefit change differentially between two hearing aids? That is, would initial versus subsequent listener comparisons between hearing aids yield different rankings?

Does Hearing Aid Benefit Change over Time?

This question lies at the heart of much of the research conducted to date. The answer(s) to this question are critically linked to the definition of "benefit" as well as to the identification and control

Departments of Communication Sciences and Disorders, Syracuse University, Syracuse, New York (C.W.T.); Speech and Hearing Sciences, Indiana University, Bloomington, Indiana (L.E.H.); and Speech Pathology and Audiology, University of Iowa, Iowa City, Iowa (R.A.B.); and the Speech and Hearing Center, University of Memphis, Memphis, Tennessee (R.M.C)

0196/0202/96/1703-0014S$3.00/0 • Ear & Hearing •

of the many external factors that can influence hearing aid benefit.

The simplest definition of "hearing aid benefit" is the improvement in "hearing ability" due to the hearing aid. "Hearing ability" can be further subdivided into objective and subjective measures, with objective measures including tests of speech recognition yielding numerical scores of "percent correct," or something similar, and subjective measures including questionnaire or interview responses. By their nature, objective measures of hearing ability usually involve a formal speech-recognition task, whereas the subjective measures can address not only speech recognition but also other aspects of everyday life. There is no consensus as to what measures provide a definitive picture of hearing ability or hearing aid benefit (e.g., Bess, 1982; National Institutes of Health, 1994). Certainly, objective measures have a built-in appeal to many researchers, in that the data are "hard" and less influenced by biases and variability. However, for the purpose of fitting and dispensing hearing aids, subjective opinions from the user are certainly an important, if not the most important, outcome measure. Often, objective and subjective measures of hearing ability from the same individuals do not even agree with one another. It seems unlikely that this complex issue will be resolved to everyone's complete satisfaction in the near future. In the meantime, we must attempt to examine a broad range of the various measures of hearing ability.

Hearing aid benefit is frequently defined as the difference between a patient's performance with a hearing aid and their performance without a hearing aid. However, there is a disadvantage in using this single-number metric to describe hearing aid benefit. It may obscure the source of any changes in benefit over time. For example, if a patient's unaided scores decline over time and the aided scores remain stable over the same period, benefit will show an increase. This is clearly a different sort of "acclimatization" than would be observed if increases in benefit are due to increases in the aided scores over time, even though both cases show an increase in benefit. For this reason, it is helpful to examine not only the "benefit" (or difference) scores, but also the raw scores that are used to derive the benefit measures.

The amount of benefit typically shown by today's commercially available hearing aids is related to the problems in defining hearing aid benefit. Often, despite the fact that a hearing aid may provide a large amount of insertion gain for a patient, the accompanying increases in word recognition on a monosyllabic test are disappointingly small. There are several factors to consider here that may influence the amount of benefit measured with a hearing aid. First, when benefit is measured for speech presented at average preferred listening levels (approximately 70 dB SPL) for patients with mild to moderate hearing losses, the additional gain provided by the hearing aid may not markedly increase speech recognition performance. Hearing aids are most effective in restoring the audibility of soft speech sounds; therefore, they may not exhibit much of an effect for speech presented at higher levels. Second, if hearing aid benefit is measured by using speech in noise as the test material, the amount of benefit provided by the hearing aid may again be limited, because the hearing aid will amplify both the speech and the noise. Third, changes in monosyllabic word-recognition scores are not related to changes in the ability to understand everyday speech in a simple manner. Boothroyd & Nittrouer (1988) described the complex relations between recognition of individual phonemes, words, and whole sentences. For example, a rather moderate percentage of correctly identified monosyllables may correspond to nearly perfect sentence recognition. The relation between phoneme- or word-recognition scores obtained in the clinic and the corresponding performance in real-world situations where some type of connected discourse is the signal of interest is presumably even less direct. It is also known that sentence and connected-discourse speech materials designed to simulate "real world" speech signals do not depend as strongly on high-frequency information as do nonsense-syllable or monosyllabic-word lists typically used for diagnostic or research purposes. It is clear that the measured amount of hearing aid benefit can be made to vary considerably for the same subject and hearing aid merely by changing the type of test materials or the listening conditions.

Another important issue is to determine if observed changes in hearing aid benefit over time are the result of "other" factors. A "pure" acclimatization effect is presumably due to increased ability to recognize speech due to "perceptual learning" of the newly audible speech cues via the new hearing aid. Other extraneous variables such as changes in the listener's volume control or other hearing aid controls could confound results. Increasing familiarity with the test materials or procedure (procedural learning) must also be factored out or eliminated. One way to do this is to include unaided listening conditions, as well as control groups composed of longstanding hearing aid users. The same cautions hold for subjective measures. Subjective acclimatization should not include changes over time in such factors as familiarity of test materials, ease of operation of the hearing aid, earmold comfort, etc.

An additional caution concerning ceiling and floor effects is warranted in examining the data from various studies. Certainly, if speech-recognition materials are too easy or too hard, yielding initial-session scores near 0 or 100% for the listeners, it may be impossible to measure any changes in benefit. Thus, only test scores lying away from the extremes of the scale should be considered informative. On the other hand, if the initial-session scores from a hearing-impaired patient are not at or near 100%, yet are equal to the scores observed for experienced normal-hearing listeners under the same listening conditions, it suggests that the hearing-impaired patient's scores could not go any higher than this. Because it is very unlikely that any hearing aid would be capable of making speech recognition better than normal for a hearing-impaired patient, we could conclude that the patient has learned to use the new speech cues to full advantage immediately upon being fit with the aid. This means that acclimatization does not occur for that subject for the test materials used.

With this background in mind, this review attempts briefly to review the existing literature on acclimatization. The emphasis is placed first on the analysis of group data. Later, some attention is devoted to the magnitude and variability of hearing aid benefit measures and the implications for individual clinic patients is addressed.

More than 50 yr ago, Watson & Knudson (1940) presented data relevant to the time course of hearing aid benefit. The main purpose of their study was to compare patients' word recognition scores under conditions of "selective" versus "uniform" amplification (provided via laboratory equipment and headphones). At least 17 subjects participated in the study; however, results as a function of time are presented for only one selected subject. The authors state "Some observers, particularly those with a large perceptive [sensorineural] loss, do not hear well with uniform amplification when it is first tried, because they have never heard some of the speech sounds before, or have not heard them for a long time. Such an observer was No. 12, with a severe perceptive loss at all frequencies" (page 412). The authors claim that this subject was thoroughly acquainted with the word lists before any testing. Technically speaking, "benefit" was not measured here, because unaided speech recognition scores were not obtained. The results for this single subject show that initially the speech-recognition scores for both amplification schemes were poor (about 10%) and that, over the course of 10 to 12 wk, recognition scores increased for both conditions to over 50%. These general results held for speech presented either at a most comfortable level or for speech presented at a fixed number of decibels above the speech reception threshold. This is surely a large change in speech recognition over time, compared with the effects reported in subsequent studies. However, this selected case study is perhaps only interesting from a historical and anecdotal perspective, inasmuch as the results of the other subjects are not presented. In addition, the 10 to 12 wk "acclimatization period" is misleading, in that their subjects did not wear these amplification devices on a daily basis, but only during the actual test sessions. Therefore, the acclimatization shown by this subject may actually represent only the most initial portion of the learning curve, perhaps comparable only to what a typical patient seen in a clinic today may experience within the first few hours of wearing his or her new aid.

Interest in this topic was revived beginning in 1992 with the publication of several studies. Cox & Alexander (1992) reported both objective (150-item Connected Speech Test [CST]) and subjective (Profile of Hearing Aid Benefit [PHAB]) data for 10 subjects obtained shortly after the subject received a new hearing aid and then again 10 wk later. A significant increase in "benefit" (aided minus unaided performance) was reported for one of the four listening environments (low noise and reverberation with full visual cues) of the CST. The group improvement was approximately 6 rau (roughly equivalent to 6% correct). For the other three environments, two showed nonsignificant decreases and the third showed a marginally significant increase in benefit. For the PHAB, a significant increase in benefit was reported for all five speech communication subscales. No control group was included, so the potential effects of procedural learning in the CST and potential effects of subject biases on the PHAB cannot be discounted. Subjects were allowed to adjust their own volume controls at each session, and "minor fitting modifications" were performed on some aids during the study. The hearing aid gain was monitored, however, and no significant changes over time were observed. In addition, raw aided and unaided scores are not reported, so it is unknown if any changes in benefit were due to decreases in unaided scores. Nonetheless, their study presented modern evidence suggesting the existence of acclimatization, and served to heighten interest in this topic.

Gatehouse (1992) measured objective hearing aid benefit using the 80-item Four Alternative Auditory Feature (FAAF) test in four new hearing aid users over the course of 12 wk. The speech materials were presented in a noise background. Recognition of the FAAF speech materials is dependent on high-frequency speech information. Several conditions were tested. In some, the subjects wore actual hearing aids, whereas in others speech testing was performed under high-pass amplification presented via headphones to simulate the hearing aids. The subjects' other (unaided) ear served as a control. For the hearing aid conditions, benefit increased from 7 to 15% over the 12 wk period (after correcting the data by subtracting any changes occurring in the control ear). For headphone conditions, the results followed a similar pattern. When the raw scores for unaided and aided conditions are examined, it is seen that in some conditions, a decrease in the unaided scores occurred over time. Subjects were allowed to adjust the volume controls of their aids. However, gain remained essentially stable after 5 wk, whereas improvements in benefit continued until the end of the 12 wk period. As with the Cox & Alexander (1992) study, Gatehouse's report served to revive interest in this topic, and demonstrated the value of including control conditions and reporting raw data to assist in interpreting the results.

Taylor (1993) reported amplification benefit in 58 elderly new hearing aid users over the course of 1 yr. Both aided objective (speech reception thresholds, 50-item Northwestern University No. 6 monosyllables in noise and quiet) and subjective (Hearing Handicap Inventory for the Elderly [HHIE]) measures were used. The Northwestern University No. 6 speech test is individual monosyllabic word recognition emphasizing mid- to high-frequency speech information. No significant improvements in objective measures for either quiet or noise backgrounds were noted, and the HHIE showed changes only for the 12 wk measure, and this was in the negative direction. Volume control settings were not fixed in this study, and the "initial" test session did not occur until 3 wk after the hearing aid was dispensed. No results from unaided control conditions were reported.

Bentler, Niebuhr, Getta, and Anderson (1993a, b) published the results from 39 "new" and 26 longstanding hearing aid users over the course of 1 yr. Objective measures included the Speech Perception in Noise 25-item tests and a 62-item Nonsense Syllable Test (NST). The Speech Perception in Noise speech test is a sentence task, with individual "target" words being scored. The NST monosyllables emphasize high-frequency speech information. Subjective measures included the Understanding Speech subsection of the Hearing Performance Inventory and a Qualitative Judgment test. Significant improvements over time were not noted for most tests, the exception was a subjective measure relating to speech in quiet. Volume controls were set at the initial test session by the investigator. Subsequently, the subject was allowed to set the volume control and some modifications of the hearing aid fitting controls were performed throughout the experiment. No unaided control conditions were included. Another potentially important factor to consider is that the "new" hearing aid users included some patients who had trial-period experience with hearing aids before the initial test session.

Horwitz (1996) followed 13 listeners with newly fitted hearing aids and also a control group of 13 longstanding hearing aid users. Both objective (192-item NST) and subjective (PHAB) scores were obtained over an 18 wk period. For the NST testing, two volume-control conditions were run, the first with volume controls fixed in the same position as the initial test session, the second allowing the subjects to adjust the volume control themselves for each session. Group mean NST scores significantly increased for the new hearing aid users in both the fixed and adjusted volume control settings. In contrast, the NST scores for the longstanding user group only increased for the adjusted volume control condition. Unaided scores remained stable for both groups. The increase in objectively measured benefit observed in the new user group was approximately 6%. The subjective measures of benefit did not show a significant improvement in benefit for the new users. These results suggest that the acclimatization observed for the objective measures was not dependent on increasing the volume-control settings. Nor was the increase due to procedural learning effects, inasmuch as a corresponding increase in word recognition was not observed in the fixed-volume, longstanding (control) group. It also suggests that significant increases in objective benefit may not necessarily be accompanied by a corresponding significant subjective improvement.

Humes, Halling, Schmitt, Coughlin, Wilson, and Kinden (1995) measured speech recognition for 102-item NST syllable lists both in quiet and in noise, and the 100-item Hearing in Noise Test speech test over a 24 wk period in 10 new and 10 experienced hearing aid users. All subjects were fit with new binaural hearing aids at the initialization of the study. The NST syllables are particularly sensitive to high-frequency speech information. The Hearing in Noise Text is a sentence test designed to more closely represent real-world communication abilities, and therefore is primarily sensitive to low- and midfrequency speech information. In addition, the subjective Hearing Aid Performance Inventory and HHIE scales were also administered. No significant increase over time was noted in any of the measures or groups in this study.

Cox, Alexander, Taylor, and Gray (1995) measured speech recognition on a 300-item CST and a 24-item Speech Pattern Contrast (SPAC) test in 22 elderly, first-time hearing aid users. The SPAC test is specifically designed to score performance in terms of speech features. Presentation levels of the speech and a background of multitalker babble were held constant in the subjects' ear canal over the testing period. In addition, a small control group of experienced hearing aid users participated in the experiment. Over a period of 12 wk, a 4% improvement in speech recognition on the CST was noted in the group data; this result was statistically significant. No such increase was observed in the control group or in the unaided scores for the test ear. Cox et al. hypothesized that acclimatization effects would be related to the amount of high-frequency gain newly provided to the subjects. This was not supported by the data; however, the difference in audibility across subjects was relatively small. These authors also hypothesized that acclimatization effects would be exhibited mainly in the high-frequency speech sounds in the SPAC. Again the data did not support this hypothesis.

In addition, another report is available in the literature that may relate to acclimatization. In a retrospective look at patient records, Arkis & Burkey (1994) reported clinical (CNC) word-recognition scores for 105 patients, the first measure taken before a hearing aid fitting, the second taken a few months afterward. A 5% increase in word recognition was noted for the aided ear. However, because all testing was performed at 30 dB SL under headphones, this study did not specifically test the situation for acclimatization occurring under more realistic conditions of listening to newly amplified sound via the patients' hearing aid.

Demorest & Walden (1984), Malinoff & Weinstein (1989), and Seyfried (1990), using only questionnaire measures, showed that subjectively measured hearing aid benefit increased over the first few weeks of use, and then diminished during subsequent months. On the other hand, a number of similar studies suggest that subjective hearing aid benefit remains stable over long periods of time (e.g., Brooks, 1989; Henrichson, Noring, Lindemann, Christensen, & Parving, 1991; Mulrow, Tuley, & Aguilar, 1992; Schum, 1992). These studies, despite the lack of control conditions designed to factor out nonacclimatization factors, imply that subjective evidence of acclimatization is not robust.

How can one summarize this literature with respect to the first question, "Does hearing aid benefit increase over time?" First, it appears that a significant effect of increasing benefit over time may exist as a real phenomenon. Although the increase in benefit observed in some studies may have resulted from increasing volume controls or from decreasing unaided scores in some studies, even when these factors are ruled out, as was done in several studies, an acclimatization effect often remains. No studies reported to date have shown a significant group mean decrease in benefit over time, as might be expected if the true acclimatization value was zero, with individual studies yielding results randomly distributed about this mean of zero. In Table 1, the various studies are summarized. In Figure 1, the group mean scores for initial and final word-recognition testing are plotted, both for aided and unaided conditions for the studies where these raw quantities are published. Figure 1 shows the reader the types of raw scores obtained in the various conditions of these studies.

There have also been a number of studies on "auditory deprivation," or the apparent decrease over time in speech-recognition ability in the unaided ear when a patient with a bilaterally symmetric hearing loss is fit with only one hearing aid. These studies are reviewed in detail in a companion article by Neuman in this issue. However, some of these studies may provide data related to the acclimatization effect, in that the speech-recognition scores of the aided ear are also provided from clinic visits soon (within 6 to 8 wk) after the initial fitting and then extending over time. Most of these studies cover long periods of time, typically 4 to 5 yr. All used headphones instead of hearing aids for testing. The speech measures used were sensitive enough to exhibit the effects of auditory deprivation. These studies are summarized in Table 2. None of these studies, except for Arkis & Burkey (1994), which is listed in Table 1, support the existence of increasing benefit over time in the aided ear. Interestingly, the Arkis & Burkey study does not support the existence of auditory deprivation when the scores for the unaided ear are examined, although their measures only extended to 33 wk postfitting.

Why then is there an apparent disagreement between various studies in terms of whether benefit changes over time? Table 1 and Figure 1 show that neither choice of speech materials nor presence or absence of background noise during speech testing can explain the pattern of results. High-frequency word lists were used in several studies that did show an effect, as well as in several studies that did not show an effect. Although a few of these studies may have suffered from ceiling effects in some conditions, one of those showed an acclimatization effect (Cox, Alexander, Taylor, & Gray, 1995) and one did not (Bentler, Niebuhr, Getta, & Anderson, 1993a). Instead it is helpful to view this research from a perspective in which each experiment is taken to be a sampling of the clinical population as a whole. The few studies that have shown large objectively measured effects for small numbers of subjects and the few showing no effects probably represent samples

TABLE 1. Summary of recent studies of acclimitization effect

Study	*N*	New/Exp	Mon/Bin	Stim Deliv	Dep Meas	Timetable	Adaptation?
Malinoff & Weinstein, 1989	25	New	Mon	HA	HHIE	3, 12, 52 wk	N, large improvement after 3 wk, decreased after
Gatehouse, 1992	4	New?	Mon	HA & headphones	%CNC (FAAF), +10 SNR, 65 dB	12 weeks, 0–6 wkly, 8–12 biwk	Y, after 4 wk, 10–15% incr
Cox & Alexander, 1992	10–17(12)	8 New, 9 Exp	9 Mon, 3 Bin	HA	%CST (a/v), +7, +2 SNR, 55, 64 dB, T; PHAB	10 wk	Y?, 4–5% & dependent on listening conditions (CST)
Mulrow et al., 1992	192	?	Mon	HA	HHIE, Denver Quick Scale	16, 32, 52 wk	N
Taylor, 1993	58	New	37 Mon, 21 Bin	HA	%CNC, Q, +10 SNR, 62 dB; HHIE	3, 12, 24, 52 wk	N, HHIE large improvement after 3 wk, decreased after
Bentler et al., 1993a	65	39 New, 26 Exp	55 Mon, 10 Bin	HA	%SPIN, +8 SBR; %NST, Q, +5 SNR; 65–75 dB	4, 12, 24, 52 wk	N
Bentler et al., 1993b	65	39 New, 26 Exp	55 Mon, 10 Bin	HA	HPI-38, Speech Communication subscales	4, 12, 24, 52 wk	N?, "Speech in Quiet" only
Gatehouse, 1993	36	Exp	Mon	HA	%CNC (FAAF), +7 SNR; SVT in Noise	0, 8, 16 wk; use of improved HA (NAL vs. NHS)	Y, NAL 3–5% or 1 dB SNR better than NHS by 8 wk
Arkis & Burkey, 1994	105	New?	70 Mon, 35 Bin	Headphones	%CNC, Q, 30 dB SL	mean test-retest interval of 33 wk	Y, 5% increase for Mon and Bin aided
Horwitz, 1995	26	13 New, 13 Exp-Control	Mon	HA	%NST, +20 SNR, 70 dB; PHAB	0, 3, 6, 10, 14, 18 wk	Y, new users, 7–8% increase, 0–18 wk, NST & PHAB
Humes et al., 1995	20	10 New, 10 Exp	Bin	HA	%NST, Q, 70 dB; %HINT, +7 SNR, 70 dB; HAPI, HHIE	0, 1, 2, 4, 8, 12, 24 wk (HINT through 8 wk only)	N
Cox et al., 1995	27	22 New, 5 Exp-Control	Mon	HA	%CST, +1 or +3 SNR, 63 dB; %SPAC, Q, 63 dB	0, 12 weeks (CST); 0, 3, 6, 9, 12 wk (SPAC)	Y, 4–5% increase after 12 wk for group, but only 3 of 22 Ss

N = sample size; New/Exp = new or experienced hearing-aid users fit with new hearing aid or, for some experienced users, their own hearing aids (control group); Mon/Bin = monaural or binaural fitting of hearing aids; Dep Meas = dependent measure of acclimitization used and listening conditions (levels for speech signals are all in dB SPL unless noted otherwise); SNR = signal-to-noise ratio in dB; Q = quiet; timetable = test intervals used in study; Acclim? = presence or absence of a significant acclimitization effect (Y = yes, there was significant effect; N = no, there was not a significant effect; Y? = multiple measures with most, but not all, suggesting significant effect; N? = multiple measures with most, but not all, indicating no significant effect).

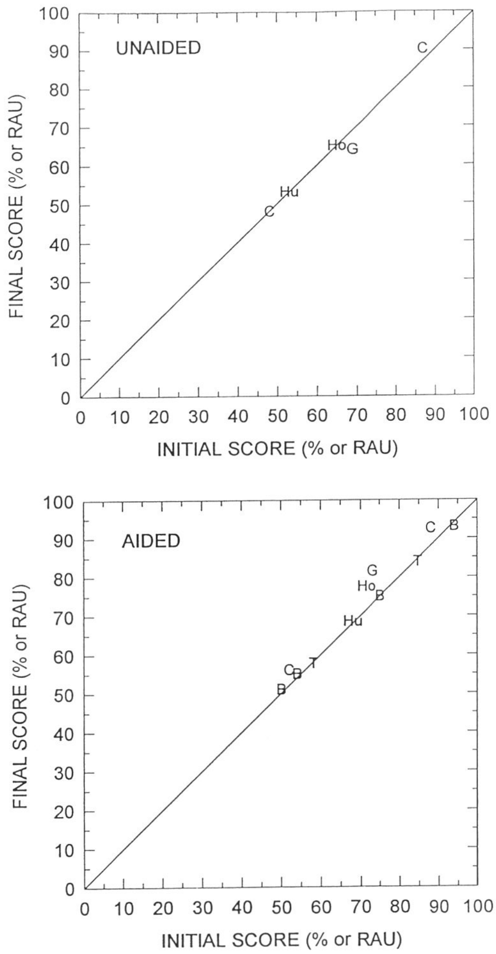

Figure 1. The group mean initial versus final speech-recognition scores are plotted for various studies that delivered speech through subjects' (first-time) hearing aids. Data for both unaided and aided conditions from the fitted ear(s) (where the data were available from the published studies) are shown in the upper and lower panels, respectively. The data from Gatehouse (1992) are for headphone conditions D and G in that study, which were designed to simulate the actual aided conditions. Aided and unaided raw scores in percentages correct for the fitted ear with the actual hearing aid were not reported in that study. B = Bentler, Niebuhr, Getta, and Anderson, 1993; G = Gatehouse, 1992; T = Taylor, 1992; Ho = Horwitz, 1995; C = Cox et al., 1995; Hu = Humes, 1995.

from the edges of the population distribution as a whole.

In Figure 2, a schematic representation of various studies' results is shown; these studies were chosen because they represent the kind of data that might be obtained in typical clinical situations in which a patient is fit with his or her first hearing aid. This figure displays the distribution of subjects' "change in benefit" over the duration of each study for the seven studies that tested speech recognition of words or syllables (measured in percent correct) using subjects' own newly issued hearing aids (for "new" hearing aid users). For each study, a normal-shaped distribution for the quantity "change in benefit" over time is assumed and is represented by a different line type. The height (number of subjects), mean, and SD across subjects for each distribution is taken from the published reports or from the authors themselves.[1] The results of all seven published studies are shown. Although the word lists differ between studies, we have combined across studies, because it is not clear what speech materials are the "best" for representing the "real world" benefits of a hearing aid and, although converting to a common metric would be the ideal approach, it is also not clear what metric is appropriate across the various test materials and/or listening conditions. The subjects are those typically seen in most hearing aid clinics—adults with mild to severe high-frequency hearing losses. It is immediately evident that the variability across subjects in nearly every study is at least as large as, if not larger than, the size of the "benefit change," and the distribution of subjects' values in the various studies show considerable overlap with one another.

Figure 2 therefore demonstrates that the various studies do not really differ substantially from one another; the differences between them are certainly consistent with expected differences in sampling this population. Because each experiment is actually a sampling of the population distribution conducted to obtain an estimate of the entire population, the summed results (upper solid-line curve in Fig. 2) of these seven studies can perhaps be taken as a better estimate of the distribution of the entire population than any single experiment. The all studies' group "Sum" distribution has a mean of approximately 3% for the quantity "change in benefit," with a standard deviation of 9%. Thus, although there would appear to be a real effect of hearing aid acclimatization, this effect is not large compared to the variability observed across patients. Therefore, it is not surpris-

[1] For two studies in this figure (Bentler, Niebuhr, Getta, & Anderson, 1993a; Taylor, 1993), the quantities "benefit" or "change in benefit" cannot be exactly determined; instead, the researchers obtained only the aided speech-recognition scores. In these cases, it was assumed that the unaided scores did not change over time, and the quantity "change in aided scores" was used as an estimate of "change in benefit." In studies in which several listening environments were included, we averaged across the conditions to yield a single result for the study as a whole. For the Taylor (1993) study, the standard deviation across subjects for the quantity "change in benefit" was not available from the author. Because "change in benefit" is derived from the difference between final and initial scores, the variance of this value depends on the correlation between these two values across individual subjects. We therefore estimated this quantity from the standard deviation of the raw scores and by assuming a mild correlation between the initial and final test results, in line with what was seen in the other studies.

TABLE 2. Summary of acclimitization effects as measured in the aided ear in studies of auditory deprivation

Study	N	Age (yr)	Mon/Bin	Stim Deliv	Dep Meas	Timetable (yr)	Acclim?
Silman, Gelfand, & Silverman, 1984	67	M = 58	Mon	Headphones	%CNC (W-22s), Q, 40 dB SL (re: SRT)	0 to 4–5 y	N
Silverman, 1989	2	54, 82	Mon	Headphones	%CNC (W-22s), Q, 40 dB SL (re: SRT)	0 to 8 or 12	N
Silverman & Silman, 1990	2	12, 34	Mon	Headphones	%CNC (W-22s), Q, 40 dB SL (re: SRT)	0 to 6 or 11	N
Silverman & Emmer, 1993	8	26–57	Mon	Headphones	%CNC (W-22s), Q, 40 dB SL (re: SRT)	0 to 2–10	N
Hurley, 1993	9	56–63	Mon	Headphones	%CNC (W-22s), Q, 40 dB SL (re: SRT)	0 to 4–8	N
Hattori, 1993	35	4–5	17 Mon, 18 Bin or Alt Mon	Headphones	"NST" max	mean test-retest interval = 15	N?, 20% increase in Mon-aided only

N = sample size; Age = age of hearing-aid users at time of fit; Mon/Bin = monaural or binaural fitting of hearing aids; Dep Meas = dependent measure of acclimitization used and listening conditions (Q = quiet); timetable = test intervals used in study; Acclim? = presence or absence of a significant acclimitization effect (Y = yes, there was significant effect; N = no, there was not a significant effect; Y? = multiple measures with most, but not all, suggesting significant effect; N? = multiple measures with most, but not all, indicating no significant effect). In every study shown here, the initial measurement of performance was at the initial fitting of the hearing aids (as opposed to several other such studies in which the first measurements were made "within 6–8 weeks" of the initial fitting). Also, it has been assumed that all subjects are new users, although this has not typically been stated in the studies.

ing that some studies have shown a significant effect for acclimatization and others have not. In other words, a patient's speech-recognition scores will often change between the initial and final evaluations. One factor influencing this change is a positive-direction acclimatization effect (several percentage points). There are also much stronger, as yet unidentified, factors influencing the amount of change observed (either in the positive or negative direction). The effects of these unidentified factors appear as across-subject variability. The large variability across subjects would suggest that further studies, using current-generation hearing aids, seeking to test for the existence of acclimatization will yield results similar to those published already.

It should also be pointed out that the average magnitude of "change in benefit," although in the neighborhood of only a few percentage points in most studies, is not much smaller than the magnitude of the initial benefit provided by the hearing aid in many of these studies. For example, Horwitz (1995) found approximately 7% initial benefit from the hearing aid (SD 9%). The change in this benefit over time was of nearly the same magnitude and variance. Cox & Alexander (1992) found a mean initial benefit (across conditions) of about 3%, and

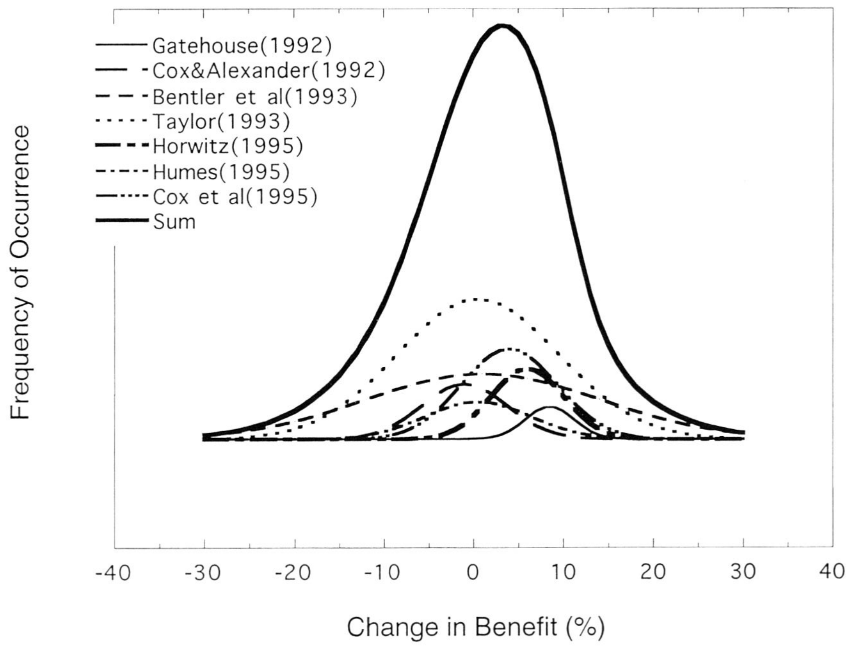

Figure 2. The distribution of values for "change in benefit" for the subjects in seven studies using "first-time" hearing aids. The seven individual studies are indicated by the lighter lines; the topmost heavy line represents the all studies' "Sum." See text and footnote 1 for further details.

the change in benefit across conditions was near 0. Some of the other studies in which unaided results were obtained show a similar pattern. This points out a general problem with all studies in the area of hearing aid evaluations: that many speech-recognition measures of hearing aid performance do not show large effects of hearing aid benefit, especially when conversational level speech, or speech in noise, is used as the test material. The magnitude of improvement due to the aid is small, and the variance of the data across subjects is quite large. Certainly, attempts to study changes in hearing aid benefit can only be as good as our ability to measure hearing aid benefit in general. Whether "improved" speech tests or testing conditions will change this situation in the future remains unanswered.

Statistical tests on group data, such as analyses of variance, are extremely sensitive to detecting significant differences between groups. However, they do not directly inform us as to the relevance of these significant group differences when applied to individual patients. The data show that the degree of improvement in benefit over time is quite variable across subjects, with some subjects showing little or no improvement over time, and some even showing decreases in benefit. The more items used in a speech-recognition test, the more reliable it will be (Thornton & Raffin, 1978), and confidence intervals for single administrations of a speech test can be determined by knowing the number of items and the patient's score. Similarly, critical differences between scores (such as the difference between initial and final test scores in an acclimatization experiment) can also be calculated to examine how many patients show an individually significant change in hearing aid benefit. The magnitude of change tends to be so small in the studies described above that reliable estimates of acclimatization would be difficult to obtain for individual patients under normal clinical settings, in which word lists of 100 items or less are typically used. For example, although Cox, Alexander, Taylor, and Gray (1995) did find a significant group effect for changes in hearing aid benefit over time, only 3 of 22 individual subjects showed a significant effect based on confidence intervals for the 300-item word lists used in the study. The data of Horwitz (1995), which was based on 192-item word lists, indicate that only 4 of 13 new hearing aid users showed reliable differences between initial and final word recognition scores. These findings suggest that even when the variability associated with speech-recognition testing is minimized by using longer (and therefore more reliable) word lists, a large variation still remains among subjects in terms of changes in hearing aid benefit over time. Subjective measures, although presumably possessing a good deal of practical face validity, may be even less reliable.

It may be more profitable in the future for researchers to study acclimatization in terms of selected groups of subjects or amplification strategies, looking to answer the questions surrounding acclimatization for those special circumstances in which it is expected to exist. This should not be confused with studying acclimatization only in subjects who show acclimatization: such a strategy will not allow us to progress beyond the anecdotal or case study type of research. Instead, we should try to identify specific circumstances that reliably lead to a change in benefit over time. Some of these issues are addressed in the following sections. It is certainly possible that newer designs in hearing aids may also exhibit acclimatization characteristics. For example, Yund & Buckles (1995), in a laboratory study of compression algorithms, reported a change in hearing aid benefit over the course of several listening sessions.

Is There Any Relation between Individual Subject Factors and the Amount of Acclimatization?

Individual subject factors that may be related to the degree of acclimatization include such things as etiology, magnitude of hearing loss, age of onset of hearing loss and/or age at time of hearing aid fitting, the amount of initial benefit derived from the hearing aid, and the amount of hearing aid gain prescribed. If acclimatization is related to the patient learning to use "new" (or previously inaudible) speech cues, one might expect that any of these factors could influence the amount of benefit change observed over time. Factors such as these might be responsible for the relatively large variability across subjects observed in all the studies described above. The research, however, has not identified any consistent factors that could be used to predict a priori which subjects will show a change in benefit over time.

Most of the studies have examined the relation between degree of hearing loss (and/or the closely related factor, amount of hearing aid gain). Gatehouse (1992) reported an acclimatization effect in his four subjects. His subjects had high-frequency hearing losses of approximately 50 to 60 dB, and the insertion gain for their new (first) hearing aid was usually more than one-half of that amount. Gatehouse (1993), using a larger group of previous hearing aid users who were fit with a "more appropriate" hearing aid, that had increased high-frequency gain, also showed acclimatization, but the magnitude of the effect in the 1993 study was considerably smaller than that in the 1992 study. These two studies might suggest that the greater the amount of newly audible information, the greater the amount of acclimatization. However, in Cox & Alexander (1992), a study that did not show a robust acclimatization effect across a majority of listening conditions, subjects also had high-frequency hearing losses of 50 to 70 dB. They plotted the amount of initial benefit against the amount of long-term benefit for their subjects on the CST speech test. There was no clear predictive relation between the amount of initial benefit and the change in benefit for their subjects, with some subjects showing low initial benefit coupled with a large change over time, and others showing an opposite relation. For the subjective benefit measures used by these investigators, again, no clear predictive relation was observed between the amount of initial benefit and the amount of acclimatization. A large group of subjects with a wide range of hearing losses was used by Bentler, Niebuhr, Getta, and Anderson (1993a, b). There was no relation between the amount of hearing loss and the change in benefit over time, nor was the amount of insertion gain correlated with the amount of acclimatization. Horwitz (1995) did find a mild positive correlation between the degree of hearing loss and the amount of acclimatization, but no such relation was found by Humes et al. (1995). Cox, Alexander, Taylor, and Gray (1995) and Horwitz (1995) found no relation between the amount of acclimatization and the amount of high-frequency gain provided to the new users of hearing aids.

In summary, no convincing evidence exists linking the amount of acclimatization to any particular subject factors. Typically, a factor identified in one study is not replicated by others. This topic remains unresolved.

What Is the Time Course of Acclimatization?

One may attempt to estimate the time course of acclimatization by examining the studies discussed above. Of particular interest is the amount of time required for performance to reach an asymptote. If such a figure could be determined, then clinical evaluations of hearing aids would not be considered "valid" unless they were obtained after this time period. Unfortunately, because of the magnitude of the acclimatization effect in relation to the much larger sources of variability surrounding "change in benefit" measures, there is no clear answer to this question.

At face value, the studies that did not show a group acclimatization effect (Bentler, Niebuhr, Getta, & Anderson, 1993a, b; Humes, Halling, Schmitt, Coughlin, Wilson, & Kinden, 1995; Taylor, 1993) suggest that acclimatization does not occur; therefore, performance asymptotes soon after or immediately after the receipt of the new hearing aid. In the Bentler, Niebuhr, Getta, and Anderson studies, "new" hearing aid users included some who had worn hearing aids during the past year or more recently. This would imply that any acclimatization effects that may have occurred were complete within the period before the initial evaluation. The Taylor (1993) study did not obtain initial measures until 3 wk after the hearing aid fitting. This would imply that any acclimatization effects that have occurred were complete within a much shorter time period (3 wk).

On the other hand, the studies showing some acclimatization effect (Cox & Alexander, 1992; Cox, Alexander, Taylor, & Gray, 1995; Gatehouse, 1992; Gatehouse, 1993; Horwitz, 1995) seem to indicate that the acclimatization effect is not completed after 10 to 18 wk, as none show an asymptote in benefit at the time of the final test session. Gatehouse (1992) and Horwitz (1995) both show that some of the largest changes in benefit occur in the time period between 3 to 18 wk postfitting. Although Cox et al. (1995) did show increasing benefit continuing beyond 12 wk in some subjects, these increases were due to declining unaided scores. In summary, the literature again cannot provide a clear picture of the time course of acclimatization because of the large variability across subjects and studies.

Does Hearing Aid Benefit Change Differentially between Two Hearing Aids?

We now move beyond the original basic question of the existence of acclimatization toward more direct clinical implications. If every patient showed a similar and reliable acclimatization effect under all circumstances, there would be few consequences in terms of evaluating and choosing hearing aids. The clinician could merely assume that the hearing aid benefit measured at the initial fitting would increase by a given amount over time, regardless of the hearing aid. Conversely, if acclimatization could be reliably assumed to not occur, initial fitting data would also be valid. However, if the amount of acclimatization depended on the hearing aid and/or patient, then a comparison between two aids on the basis of data gathered at the initial fitting would be invalid at a later date. The single subject's data presented by Watson & Knudson (1940) is an example of such a comparison made over time. At the end of the testing period, both the uniform and the selective amplification strategies yielded similar scores, although the amount of time for the two effects to reach asymptote was different.

In a more comprehensive study, Walden, Schwartz, Williams, Holum-Hardegen, and Crowley (1983) compared speech recognition results from 35 patients who were "issued" one of three dissimilar hearing aids on an alternating basis (the subjects were in a resident-based, military rehabilitation program). Examining only the 20 subjects who showed significant differences between aids at the initial evaluation, Walden et al. looked to see how stable the scores remained over a 1 wk period. After a week of alternating use, overall group mean aided speech recognition scores did not improve for any of the hearing aids (this study would therefore not support the existence of acclimatization over the first week). Because the group scores did not change over time, this suggests that rank orderings of hearing aids obtained at an initial session would remain valid over time. However, the individual results did not strongly support such an interpretation. Of the 20 original subjects who showed significant speech recognition differences between these very different hearing aids on the initial testing, only six subjects retained the same ordering of hearing aids after the 1 wk period. For 7 of the 20 subjects, significant differences between aids disappeared. Based on this pattern of results, it is also likely that some of the other 15 subjects (of the 35 originals), who did not show significant differences between aids on the initial testing, would also have shown some differences on the retests. In other words, even when group data might support a particular conclusion concerning hearing aids, these results cannot reliably be applied to individuals.

Gatehouse (1993) looked at the progress of 36 patients who were fit with a new hearing aid but not new users of hearing aids. The subjects were fit with a new frequency response that provided more high-frequency amplification than the old hearing aid. Scores on the 80-item FAAF and a Sentence Verification Test were measured up to 16 wk after the fitting of the new hearing aid. A statistically significant 4% improvement in FAAF speech recognition was noted in the group data. Scores were also obtained with the old hearing aid, and they remained stable over the 16 wk period, indicating that the acclimatization effect observed in this study was due solely to improvements in scores obtained using the new hearing aids. Thus, in group data, acclimatization to a different hearing aid was observed in this study. However, for individuals, changes of this magnitude would be considerably more difficult to observe.

These studies suggest, therefore, that either or both of the following occur: 1) acclimatization does not exist at a magnitude large enough to overcome the individual variability in speech testing with different hearing aids, or 2) acclimatization proceeds at unequal rates across subjects and hearing aids. Although these studies are not sufficient to answer the larger question posed at the beginning of this section, they do imply that, although initial-session rankings of hearing aids in individuals may not be valid, waiting to re-evaluate hearing aids after an "acclimatization" period may also not add much new information. This dilemma seems to be strongly related to the limitations of our test instruments as used for evaluating hearing aids. Indeed, these limitations have led the majority of clinicians today to eliminate speech-recognition testing from their hearing aid fitting protocols.

SUMMARY

The hypothesis that patients will learn and improve in their ability to use new speech cues over time after being fit with a new hearing aid has been studied by a number of investigators. Some studies have reported a statistically significant group increase in hearing aid benefit, as measured by increases in aided speech recognition scores over time. A number of studies have found no change in benefit over time. No studies have shown a significant group decrease in benefit. However, individual subjects in these studies can show changes in benefit in either a positive or negative direction. This leads to the conclusion that, although hearing aid benefit tends to increase over time, there are other larger sources of variability that can influence change in benefit scores in either a positive or negative direction. Some of this variability is the result of speech tests that use a small number of items. However, a substantial variability still exists in the data from studies using longer word lists as well. This variability across patients is probably exacerbated by the possibility that current speech-recognition measures do not accurately reflect hearing aid benefit.

In view of the size and variability of the acclimatization effect, it may only have some immediate consequences for large-scale hearing aid studies, where a difference of a few percentage points in group speech-recognition results may be of significance. One also cannot rule out the possibility that acclimatization effects will be larger or more consistent for future generations of hearing aids, or for laboratory studies of new speech-processing techniques.

For current clinical practice, however, the variability of the effect across patients makes it impossible at this time for a clinician to predict which individual patients will show a reliable increase in benefit over time. For individual subjects, no factors have yet been identified that can predict ahead of time which patients will show an increase (or decrease) in benefit. In addition, the magnitude of the acclimatization effect is not large enough in relation to typical test-retest accuracy of individual tests to allow the clinician even to observe the effect with confidence in the majority of patients. Because the clinician at present can neither predict nor even reliably measure the effects of acclimatization in individual patients in today's typical clinical settings, there would appear to be minimal consequences of acclimatization for current clinical practice.

Acknowledgments:

The other participants at the Erikshom Workshop in 1995 were responsible for 3 days of sharing of data and ideas with the authors. This conference was sponsored by the Oticon Foundation. Their inspiration and help was invaluable to this project. Amy Horwitz deserves tremendous thanks for providing the first author with a thorough literature review of the topic in her dissertation. This work was supported in part by grants from NIDCD, NIA, and the VA.

Address for correspondence: Christopher W. Turner, Department of Communication Sciences and Disorders, Syracuse University, 805 South Crouse Avenue, Syracuse, New York 13244-2280.

Received January 23, 1996; accepted March 20, 1996

Implications of Acclimatization for Hearing Aid Fitting Practices and Research

Denis Byrne

Auditory deprivation has been advanced as a logical argument for bilateral fitting in severe hearing losses; unilateral fitting may lead to a reduction of function in the unaided ear that may or may not be restored by subsequent bilateral fitting. However, for people with mild hearing losses, unilateral fitting may still permit binaural functioning, because many sounds will be audible in the unaided ear. There may conceivably be some advantage in that ear being acclimatized to low sound levels. If one ear becomes acclimatized to low-intensity sounds and the other ear becomes acclimatized to high-intensity sounds, this may extend the intensity range over which auditory functioning is effective. It is not intended here to advocate unilateral fitting but simply to point out that there is evidence from the acclimatization, as well as other, data to suggest that bilateral fitting may not necessarily be the better choice for all people whose hearing losses are only mild or moderate.

The existence of acclimatization implies that evaluations performed immediately after hearing aid fitting may underestimate the potential benefit. Comparative evaluations could also lead to the wrong conclusions because there could be a bias favoring the pattern of sound that most closely resembles what the client is used to hearing. It is arguable whether the deferring of clinical evaluations is justified considering that acclimatization effects, for current types of hearing aids, seem relatively small (Turner et al., this issue). If acclimatization is to be allowed for, then the data of Gatehouse suggest that it may be desirable to defer evaluation until 3 or 4 mo postfitting. This raises the question: fitting with what? Whatever fitting is chosen may bias any future comparative evaluation. The possibility of acclimatization emphasizes the importance of striving to achieve the best possible fitting from the outset so that the hearing aid wearer will acclimatize to the type of sound that is capable of providing most benefit. The client should be encouraged to get used to the aid that should be most beneficial rather than the one that sounds best initially. Although a compromise may be necessary, it seems short sighted to compromise too much before acclimatization has had a chance to occur. This suggested way of proceeding presupposes that the audiologist knows what is likely best for the client, which gets back to the importance of starting with the best possible prescriptive procedure.

Many experiments have compared two or more types of amplification, including some that the subject has had little experience in using. If no difference is found, or if a new type of amplification is found to be inferior, the question arises: would performance with the new type of amplification have been better if the subject had been experienced in using it? Thus, the absence of benefit for a novel amplification scheme is not conclusive unless there has been a reasonable opportunity for acclimatization to occur. Ideally, experiments should be designed so that the subject is given sufficient experience to acclimatize. This is likely to incur difficulties as the inclusion of acclimatization periods will increase the length and complexity of studies. Furthermore, it requires that all amplification schemes be available in a wearable form, a requirement that may be difficult and costly to meet for complex types of signal processing.

In designing experiments, we need to consider the probable size of the effects we are seeking to measure relative to the size of acclimatization effects, and we need to consider how long it is likely to take for acclimatization to become complete. It seems reasonable to assume that, with current types of hearing aids and mild or moderately postlingually hearing-impaired clients, acclimatization effects will be small (Turner et al., this issue), although not necessarily unimportant, as the effects of interest may also be small. With new types of processing, which may present a highly novel perceptual experience, it is not known what to expect about the size of acclimatization effects or how long it will take for such effects to become evident. Experience with cochlear implants suggests that some people may take a long time to show the full benefits of some types of processing and that acclimatization effects may be large for some individuals (Tyler & Summerfield, this issue). It appears that acclimatization depends on characteristics of the processing and characteristics of the client, such as hearing loss and experience, i.e., on "modulators" of learning (Robinson & Summerfield, this issue).

There is insufficient basis for specific recommendations about how, if at all, research designs should allow for acclimatization. Certainly, researchers should be aware of the possibility of acclimatization and that this could be especially important when evaluating new types of amplification. It seems desirable to encourage current trends to produce new processing schemes in wearable forms so that evaluations can incorporate field trials as well as laboratory tests. Recognition of possible acclimatization may affect details of research procedures. For example, some experiments involve field comparisons of different amplification provided in two memories of a multiple memory hearing aid. It may be necessary to discuss the possibility of acclimatization to persuade the subject to give both memories a fair trial rather than mostly using the one that initially sounds better.

In summary, acclimatization needs to be considered in designing and conducting hearing aid research and in interpreting results. Acclimatization may be especially important for evaluating some new types of signal processing schemes but may also be significant if large variations in frequency response or compression characteristics are being considered. Research designs that account for acclimatization may be complicated, and the benefits need to be weighed against the costs of pursuing this course.

Outcome Measures for Hearing Aid Evaluation—Properties and Requirements in Acclimatization Research

Stuart Gatehouse

Any evaluation of the acclimatization literature should take account of the statistical limitations of the procedures and measures used to assess hearing aid benefit. Turner and colleagues (this issue) point out that tests of speech identification ability have limitations, both in their relationship to "real world" benefit, and also in their ability to detect the benefits of amplification and the marginal (i.e., differential) benefits of alternative amplification schemes. Their review contains a synthesis of the literature and concludes that acclimatization effects are difficult to demonstrate using currently available speech tests and are of modest magnitude. The review raises a number of points that merit amplification.

The authors point out that acclimatization effects (revealed by tests of speech-perception in noise) appear to be of similar magnitude to the immediate benefit of amplification (i.e., the difference between aided and unaided scores shortly after fitting) for group mean data, although for a particular condition they are smaller than the between-subject standard deviation in test scores. It would appear to be the limitations inherent in currently available speech identification procedures in their ability to demonstrate the benefits of amplification that restrict the experimenter's ability to demonstrate changes in that benefit across time. If an experimenter has reason to infer that, for a group of individuals, amplification does indeed confer benefit (as evidenced, for example, by the listeners making continual use of a hearing aid for substantial portions of their daily lives), then a difficulty in demonstrating benefit on a speech test configured in a particular way should not be interpreted as questioning the existence of that benefit. The difficulties in demonstrating the benefits of a hearing aid on speech tests do not necessarily lead us to logically doubt their existence. More particularly, the interpretation of the importance of an acclimatization effect (of comparable magnitude to the initial benefit of amplification) must be considered in the same way as a consideration of the importance of the benefit itself.

The demands on the properties of evaluation instruments become even more stringent when the requirement is to distinguish the differential performance of two amplification systems. Unfortunately, for hearing-impaired listeners, researchers have not been successful in identifying signal-processing strategies or ways of adjusting hearing aids that confer on the hearing-impaired listener differential advantages that improve speech identification abilities by large amounts (say, 10 to 20% on a conventional speech test). Thus, it is clear that the likely marginal (i.e., incremental) benefits of improvements in hearing aid fitting and signal processing are also themselves likely to be comparable to the size of acclimatization effects across groups of subjects. This conclusion has direct implications for the interpretation of the importance of such effects.

The particular limitations of speech identification procedures have restricted (if not negated) their use in clinical practice in the selection and adjustment of parameters of amplification for individual listeners.

We see their replacement by measures based on subjective judgments by the hearing-impaired listener (e.g., the techniques of paired comparisons) or by evaluation using self-report scales of hearing disability and handicap. Such latter instruments have high face validity and are an essential component of any rehabilitation and evaluation scheme, but their individual psychometric and discriminatory abilities in assessing amplification per se and alternatives in amplification are, as yet, undocumented. Certainly, the methodologic investigations that accompany speech identification measures have not yet been conducted and convincingly demonstrated in the self-report domain. A general set of properties that should accompany outcome measures that are used in acclimatization research include evidence that the outcome measure of interest is indeed sensitive to the changes induced in the acoustical environment by the hearing aid (i.e., the evaluation instrument can demonstrate a difference between aided and unaided ability). Furthermore, the evaluation instrument should be sensitive to systematic alterations in the degree of the change in the acoustic environment induced by the amplification system. The attempts to use the fine structure of speech tests to relate, e.g., changes in amplification characteristics (audibility as a function of frequency) to one of speech features are unlikely to be productive. Such fine detail (e.g., in the frequency domain) is likely only to be available from psychoacoustic abilities that can be assessed by procedures constrained by frequency and intensity in addition to ear of stimulation (e.g., Byrne & Dirks, this issue).

The review pays particular attention to the generality and detectability of acclimatization effects in individuals as opposed to group mean data, as this is of particular importance to the management of individual hearing-impaired listeners. Data sets are analyzed with regard to the critical difference in speech identification scores from particular configurations of speech tests, which can be regarded as exceeding a 95% confidence interval (Thornton & Raffin, 1978). Such an approach addresses the question of how large an individual difference in performance has to be, such that it is unlikely (usually at the 5% level) to be the result of uncontrolled variability in the speech test itself (i.e., the number of occasions where a difference would be attributed as being "present" when it is actually absent, is fixed at the 5% level). Such an approach, however, is incomplete, as it does not address the issue of the power of the assessment procedure to detect a difference when it is actually "present." This assessment requires a power analysis based not only on the alpha parameter (set at 5% in the previous example) but, in addition, on the beta parameter. The latter describes the probability that a procedure can detect the presence of an effect when it, in fact, does exist given the variability of the assessment method. One way of viewing the application of a speech test (with a given list, length, and set number of items on an individual subject) is to conduct a power analysis for the difference between two proportions. If a range of examples is taken in which the size of the effect that it is desired to detect ranges between 3% (the group mean figure postulated in the review) and 7% (a not unreasonable estimate of the upper bound of the mean acclimatization effect across a group of subjects) for a baseline score of 70% using a set of speech tokens containing between 200 and 300 items (the examples drawn in the review from Horwitz & Cox), then application of such a power analysis (Lipsey, 1990), at a fixed alpha level of 0.05 for two-tailed analyses, yields a power between 0.1 and 0.35. That is, assuming an acclimatization effect existed for all subjects, the speech assessment procedures have the statistical power to detect them between 10 and 35% of individuals, given the inherent variability of the procedures. It is intriguing to note that these values correspond very closely to the 3 of 22 subjects from Cox and to the 4 of 13 subjects from Horwitz (values taken from the review of Turner and colleagues—although Figure 9 of the Cox study appears to show 4 of 22 subjects exceeding the 95% critical difference) for whom a material acclimatization effect was observed using the conventional assessment of critical differences. Thus, another interpretation of these individual subject results (as opposed to the interpretation that suggests the limited generality of acclimatization effects) is that the assessment procedures (even where they consist of speech tests with 200 to 300 items and, thus, far exceeding the lengths used in clinical practice) have the statistical power to detect effects only in that limited proportion of cases. Such a consideration, of course, does not prove that acclimatization effects are, indeed, general, but does demonstrate that the published data are not incompatible with their generality given the outcomes used.

The review and this amplification show clearly that the acclimatization literature is bedeviled by the inherent limitations in our procedures for the assessment of hearing aid benefit. It appears likely that future elucidation of the important questions surrounding acclimatization are likely only to be resolved by significant advances in our understanding and assessment of the benefits delivered by amplification schemes (or conceivably by intensively laborious experiments where hearing-impaired listeners undergo speech tests containing many hundreds of test items). Thus, the importance for the clinical practice of audiology of acclimatization phenomena will remain limited, given the limited utility of speech assessment procedures in themselves for such purposes because of the time available. However, the application of group mean data (e.g., in laboratory research and field trials of amplification systems) demonstrate that for acclimatization effects displaying comparable magnitude to both the differential performance of amplification systems and the absolute benefits delivered by those systems, it is unwise to ignore their potential influences. The incorporation of acclimatization effects into research and evaluation designs depends, of course, on the potential consequences for future action of both type 1 and type 2 errors within the experiment.