Hearing Aid Evaluation: Limitations of Present Procedures and Future Requirements

L'évaluation prothétique: limites des procédures actuelles et besoins futurs

Stuart Gatehouse
MRC Institute of Hearing Research
Glasgow Royal Infirmary

Abstract

This paper discusses the underlying conceptual problems of techniques employed in hearing aid selection and evaluation processes in four classes: target gain formulae, consumers' judgement, speech intelligibility, and self-report. Although technology for delivering and assessing acoustic properties is now largely in place, all the prescriptive formulae suffer from the absence of systematic large scale external validation studies to determine their overall success, their relative effectiveness, and the effect of deviations from the standard. As the complexity of hearing aids increases in the future, the demands on the tools and procedures used for selection and evaluation will increase. At present all the procedures have received insufficient systematic investigation on large enough representative populations for standardized instruments to emerge. Despite the legion of conceptual and practical problems in developing and validating such instruments, they remain a priority for both research and clinical applications if future technology is to be applied to the maximum benefit of the hearing impaired.
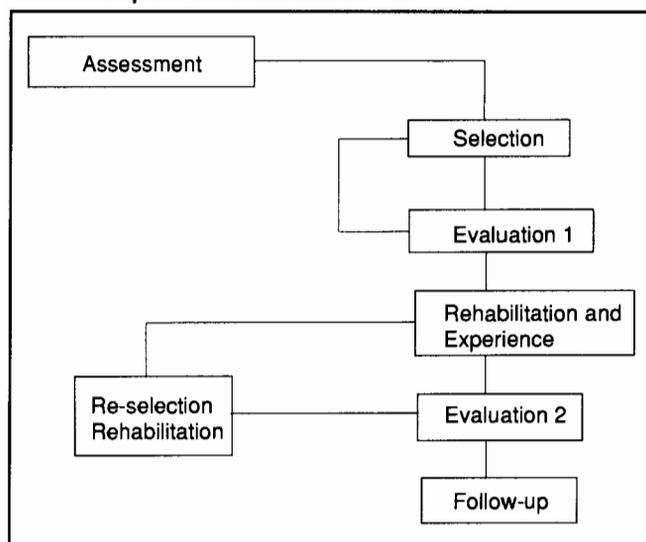
Résumé

L'auteur analyse les problèmes théoriques en rapport aux techniques utilisées pour l'évaluation et la sélection des prothèses auditives dans quatre catégories : formules des gains-cibles, jugement des utilisateurs, intelligibilité de la parole et auto-rapport. Même si la technologie fournit et évalue les propriétés acoustiques, toutes les formules prescriptives sont dénuées d'études de validation externes et systématiques effectuées sur une grande échelle pour déterminer leur réussite globale, leur efficacité relative et l'effet des écarts par rapport à la norme. Plus les aides auditives deviendront complexes et plus la demande d'outils et de procédures pour la sélection et l'évaluation sera plus forte. À l'heure actuelle, les procédures n'ont pas fait l'objet de suffisamment de recherches systématiques sur des populations assez représentatives pour développer des instruments normalisés. Malgré les nombreux problèmes théoriques et pratiques pour mettre au point et valider de tels instruments, ils demeurent une priorité pour la recherche et les applications en milieu clinique afin que les malentendants profitent au maximum des avantages de la technologie future.

Introduction

The term "hearing aid evaluation" encompasses an enormous range of audiological activities ranging from the verification of the electroacoustical characteristics of a hearing aid in a standard coupler to the application of hearing disability and handicap inventories to address issues in the public health domain. The very use of the term evaluation immediately raises questions such as "evaluation of what?" and "for what purpose?". Relevant responses can include the performance of the hearing aid itself, the relative performance of a given hearing aid to a possible competitor, and the degree to which a hearing aid resolves the hearing difficulties experienced by an individual or a specific group of hearing impaired persons. Although the term hearing aid evaluation is widely used (indeed it forms the title for one of the pre-eminent text books in the field: Skinner, 1988), the term actually encompasses the two logically separate processes of hearing aid selection and hearing aid evaluation. In hearing aid selection, emphasis is placed on the ability of a particular instrument to meet a certain target (as in probe microphone measurements) or perhaps to "win" a tournament on the basis of paired comparisons or speech intelligibility scores. In contrast, the evaluation of a hearing aid fitting places emphasis on the ability of the rehabilitation package to alleviate the auditory impairments, disabilities, and handicaps suffered by an individual or investigates population groups as part of the evaluation of health care delivery systems. Although these two processes are logically separate, many techniques used within each are common to both. While there are numerous techniques in the audiological literature recommended for both selection and/or evaluation of a hearing aid fitting, this article aims to consider the broad applications and limitations associated with each class of instrument.

Prior to proceeding to such an aim, a conceptual framework has to be established within which the selection and evaluation issues can be addressed, and the schematic shown in Figure 1 is put forward for this purpose. It must stressed that this very simplified scheme is in no way intended to

50

JSLPA Monogr. Suppl. 1, Jan. 1993 / ROA Suppl. de monogr. n° 1, janv. 1993

Figure 1. Simplified representation of the selection and evaluation processes.

Table 1. Classification of selection and evaluation procedures.

Target Formulae
Consumer Judgement
Speech Intelligibility Measures
Hearing Disability and Handicap Inventories

supplant or replace detailed models of the auditory rehabilitation process (e.g., Goldstein & Stephens, 1981; Stephens, 1987), but is merely used as a framework to highlight the issues in hearing aid selection and evaluation discussed in the present article. The scheme in Figure 1 envisages an assessment phase during which certain auditory abilities of the subject are measured (including, for example, pure tone thresholds, perhaps comfortable/uncomfortable listening levels, speech recognition abilities, etc). Usually some assessment of the hearing disability and handicap suffered by the individual is included, and the decision is made that provision of a hearing aid is an appropriate part of the rehabilitation process for the individual. The scheme then envisages a process of hearing aid selection and initial evaluation (labelled Evaluation 1 in Figure 1) that iterates until some predefined criteria are achieved. Then follows a period of rehabilitation and experience with the hearing aid, followed by a second evaluation phase labelled Evaluation 2 in Figure 1. This second evaluation phase may initiate a process of re-selection and further rehabilitation depending upon the achievement of a second (and perhaps different) set of predetermined criteria. Then, depending upon the health care delivery system within whose context the service takes place, there may be further periods of follow-up. Throughout this article the scheme in Figure 1 will be used as the basis of discussion for a variety of approaches to selection and evaluation.

Classification of Selection and Evaluation Instruments

Although the individual tests and questionnaires either employed or advocated for hearing aid selection and evaluation may be numerous, they can be classified usefully into the

four areas shown in Table 1. In putting forward this classification some explanation is required. The class, target formulae, is intended to encompass all of those procedures for which a set of measurements taken from the subject are used to produce a target for some aspect of the electroacoustic characteristics of the hearing aid. The measurements that are used for this basis may be as simple as pure tone thresholds (Berger, Hagberg, & Raine, 1984; Byrne & Dillon, 1986; McCandless & Lyregaard, 1983) or may rely on measures of comfortable and/or uncomfortable listening levels (Cox, 1985; Skinner, Pascoe, Miller, & Popelka, 1982). The overall classification of target gain formally also includes methods based on the articulation index (Berger, 1990; Pavlovic, 1988; Rankovic, 1991), which may be considered as a generalization of these approaches.

Consumer judgements are processes whereby the listener is asked to make assessments on one or more dimensions (e.g., speech intelligibility, speech quality) about a particular hearing aid. These judgements may be on the basis of paired comparisons (Punch & Parker, 1981) or direct magnitude estimation (for a description of the techniques, see Pavlovic, Sorin, Roymiguiere, & Lucas, 1990), which can then be used to select and/or evaluate a hearing aid fitting via some form of tournament strategy (Studebaker, Bisset, & Van Orf, 1982) or by some form of adaptive tracking procedure such as the simplex algorithm (Neuman, Levitt, Mills, & Schwander, 1987).

The use of performance tests of speech intelligibility in hearing aid selection and evaluation has an attractively high degree of face validity because improving speech perception is one of the underlying goals in managing hearing impairment and disability. Indeed for many years it came to be regarded as the "gold standard." Although long established as an approach to the problem (Carhart, 1946), its limitations in differentiating between hearing aids have since been exposed, although encouraging recent developments are noted later in this article.

The use of self-assessment instruments of hearing disability and handicap has a controversial and checkered history in audiology (for review, see Schow & Gatehouse, 1990). Their application to hearing aid evaluation has received renewed interest, given that one of the primary aims in hearing aid provision is to improve communication abilities via a reduction in disability and handicap.

JSLPA Monogr. Suppl. 1, Jan. 1993 / *ROA Suppl. de monogr. n° 1, janv. 1993*

51

Having identified these four classification areas for the procedures used in hearing aid selection and evaluation, their application to the simplified scheme in Figure 1 is addressed to assess their effectiveness and limitations with regard to some of the concerns raised in the introduction to this article.

Applications to Selection and Evaluation

From inspection of Figure 1 it may not be immediately apparent why the separation between the initial selection and evaluation processes has been advocated. Indeed practices for many years did not recognize such a separation and used a purely comparative procedure whereby hearing aids were ranked using measures of speech intelligibility in quiet and noise (Carhart, 1946), thus intimately linking the selection and evaluation processes. However, with the increasing popularity of prescriptive regimes, the separation, although perhaps not recognized, has been inherent in practice. Prior to the advent of instrumentation measuring real ear insertion gain or real ear aided response of a hearing aid in a clinical setting, a common approach would have been to use one of the target gain formulae for the initial selection, which would then be evaluated or validated, perhaps by clinical measures of functional gain. If the measured insertion gain proved to be inadequate, then the procedure would loop back to the selection process, and the hearing aid would be adjusted or an alternative hearing aid chosen. Using this approach, the differentiation between selection and evaluation is apparent in the particular test employed, but still logically remains within the domain of target gain (that is, has the desired/prescribed characteristic been achieved). With the advent of probe tube systems to measure in a clinical setting the output characteristics of hearing aids in the wearer's own ear rather than in standard couplers, the approach whereby hearing aids are prescribed and fitted purely on the basis of a desired target characteristic has become even more popular. Here the target characteristic is specified, and sound pressure level measurements in the wearer's ear canal are conducted to assess the degree to which the target has been achieved. Note that the selection and evaluation processes are very poorly differentiated.

However, such a blurring of distinction is not inevitable, nor is it necessarily desirable. It is entirely possible, and indeed is widely practised, that the selection process might be based on a target formula, while the initial evaluation is based on some aspect of consumer preference. Thus, for example, the initial selection might be derived from the revised NAL-R prescription formula (Byrne & Dillon, 1986). Patient report scales of sound quality and speech intelligibility then might suggest that an unacceptable degree of high frequency emphasis has been prescribed and achieved. This might in turn lead to a revision of the target by a reduction in the amount of high frequency energy prescribed. It is almost certain that in this instance the target gain would also be part of the evaluation process by measuring the extent to which the target had been achieved. We see here the beginnings of a separation between the selection and evaluation functions.

This separation becomes more apparent when the second evaluation phase (labelled Evaluation 2 in Figure 1) takes place after a period of rehabilitation and experience with the hearing aid. The particular timing and content of this evaluation depends largely on the context of the health care delivery system within which the hearing aid is prescribed. It could lie within the domain of consumer preference (listeners might conduct formal reports of speech intelligibility and sound quality listening through the hearing aid, perhaps but not necessarily in relation to a competitor) or it might be based upon speech intelligibility tests (again with the fitted hearing aid considered in isolation or in comparison to potential alternatives). Alternatively, self-report questionnaires of hearing disability and handicap may be used to assess the degree to which the listener's needs and difficulties stemming from their hearing problem have been resolved by the fitting of the hearing aid and the rehabilitation process. Any deficiencies identified at this stage may be used to trigger a process of re-selection and further rehabilitation that attempts to overcome the deficiencies identified.

It is at this later stage that the logical differences between selection and evaluation become apparent. When considering the strengths and limitations of the four types of processes identified earlier (Table 1), it is important to specify whether the procedure under scrutiny is intended to select an appropriate hearing aid or to evaluate that selection. The evaluation may be either in relation to the possible alternatives or to the degree hearing disability and handicap suffered by the listener has been overcome.

This article does not attempt a detailed critique of each and every formula, instrument, test, or procedure used in selection and evaluation, but attempts to identify the more general conceptual advantages and limitations within each of the four domains, and outline areas for research.

Target Formulae

Although the most common use of target formulae is the prescription of the desired gain of a hearing aid as a function of frequency, this is not the only application. For example, a similar approach may be employed for prescribing the maximum acoustic output of the hearing aid (e.g., SSPL90). Indeed similar prescriptive approaches may be used in specifying and adjusting the particular parameters of non-linear hearing aids (e.g., those with full dynamic range compression). This article is not concerned with the details regarding the particu-

52

JSLPA Monogr. Suppl. 1, Jan. 1993 / *ROA Suppl. de monogr. n° 1, janv. 1993*

lar measures from the listener upon which the prescription is based (e.g., threshold, comfortable listening levels, or uncomfortable listening levels) or the process by which the prescription is arrived at (e.g., coupler measurements using correction factors, real ear measures, or some combination of the two). The range of procedures advocated in the literature does of course have a variety of underlying rationales, but whether they are as unsophisticated as the simple half-gain rule or as potentially complex in their derivation as procedures based on the articulation index, they share certain characteristics. The advances in hearing aid technology and clinical instrumentation now mean that despite the continuing need for standardization of terminology and procedures, issues concerning the degree of accuracy to which measurements can be achieved and the ability of hearing aids to physically deliver the desired target are no longer a major concern. In previous years when the inaccuracies in measurement (e.g., functional gain in a clinical setting) and the physical capabilities of hearing aids (e.g., the ability to deliver smooth high frequency responses) were a concern, the evaluation or validation of a particular prescriptive regime tended to take a back seat. However, we are now faced with the situation whereby if we knew what we wanted to achieve in terms of the characteristics of a hearing aid, with no consideration to cost, the technology is in place both to deliver the target and to assess whether the target was in fact being delivered. This issue of evaluation and validation might be considered academic if the competing prescriptive regimes all produced targets that were similar or within the range of measurement errors. However, this is not the case. Skinner (1988) has shown that among six of the commonly advocated prescriptive strategies for listeners with not untypical threshold and loudness level characteristics, the differences with frequency of prescribed gain (e.g., the slope of the frequency response) for the six procedures can cover a range of some 20 to 25 dB. Thus there clearly are material differences between the available prescriptive regimes.

What characterizes all of the target formulae is their lack of validation in large-scale clinical populations. Although some attempts have been made to validate the formulae, particularly those advocated by the National Acoustics Laboratory (Byrne, 1984; Byrne, 1986a, 1986b), there is a conspicuous lack of substantive validation studies. The evaluation of target formulae may take place at two levels: (1) comparative and (2) absolute. In the former, the issue at stake is whether one prescriptive approach (e.g., the NAL-R formula) leads to a better result than a competing approach (e.g., one based on the articulation index), while the latter aims to evaluate to what extent a hearing aid fitted according to a particular formula alleviates the hearing impairment, disability, and handicap (WHO, 1980) suffered by the listener. Although studies on small numbers of subjects (who are quite likely to be atypical of the overall target population) may be valuable

in narrowing the field of competing prescriptions, realistic evaluation can only take place in the context of large-scale clinical populations with the range of hearing levels and age encountered in actual practice. Because of the heterogeneity of hearing impaired listeners, such studies, of necessity, will have to contain large numbers of subjects to ensure the applicability of the findings. These studies are expensive and difficult to conduct, and although the measurement and physical issues are no longer a bar to their progression, they do require external validation instruments that are sufficiently stable and sensitive to differentiate between outcomes, which although successful in an overall sense, might have comparative differences. Despite the lack of and deficiencies of evaluation instruments that are generally available, later sections of this article suggest some progress towards these goals. It is useful to identify the objectives of the required validation studies of target formulae. They should aim to achieve the following:

1. An assessment of the relative efficiency of competing target prescriptions (e.g., gain and SSPL90, eventually compression characteristics also) for different populations in terms of relevant characteristics such as auditory thresholds, configuration, and age.

2. The degree to which a hearing aid fitting according to a target formulae can achieve the alleviation of hearing impairment, disability, and handicap.

3. The limits of deviation from the target that are acceptable for aims 1 and 2 still to be met.

These are ambitious requirements and not necessarily attractive or glamorous to prospective researchers, but if some order is to be brought to the multiplicity of competing claims amongst prescriptive formulae, they are going to be ultimately necessary.

Consumer Judgement

Allowing or encouraging the consumer to choose what sounds best on the basis of intelligibility or quality has a high degree of face validity and has been developed into a formidable research methodology using techniques of paired comparisons and magnitude estimation in a variety of tournament and adaptive (e.g., simplex) strategies. Though no comprehensive documentary evidence is available, it is apparent that most clinical practice contains an element whereby the listener is allowed to fine tune a particular hearing aid characteristic on the basis of the above. For both this process and the more comprehensive research methodologies, questions concerning the starting value and range of alternative characteristics can arise. There are however some more fundamental issues to be addressed.

JSLPA Monogr. Suppl. 1, Jan. 1993 / *ROA Suppl. de monogr. n⁰ 1, janv. 1993*

53

It has been known for at least two decades (Thomson & Lassman, 1969) that listeners with sensorineural hearing impairment (which is almost inevitably more severe at the high frequencies) express an initial preference for a frequency response that contains less selective high frequency amplification than would be deemed desirable. It is this initial preference for what is familiar and resistance to change that contributes at least in part to the organization in Figure 1 of a rehabilitation and experience phase followed by a second evaluation stage. From a service rather than research point of view this certainly cannot be achieved in the clinic or office within a single patient visit. However, from a research point of view this raises quite difficult problems. If the objective of the research encompasses adaptive selection strategies (e.g., Neuman et al., 1987), then the type of amplification (if any) prior to evaluation could influence the outcome, and the stability of the results in the longer term remains open to question. There is a need for research on the overall stability and effects of previous experience on the results from both tournament and adaptive selection strategies. This is not to argue against the genuine merits of such approaches in terms of potential sensitivity and relatively low demand on time, but rather to caution against their acceptance and application in settings in which the performance has not been properly documented.

Any procedure that attempts to conduct tests in the clinic or laboratory to identify an optimum or maximally acceptable characteristic for everyday use faces a series of requirements on the speech materials that are used and the acoustic conditions (such as presentation level, signal-to-noise ratio, degree of reverberation, type of competing signal) employed in the investigation. There is evidence that such parameters can influence the outcome (e.g., Sullivan, Levitt, Whang, & Hennessey, 1988). At a group level, the experimenter or clinician requires that the conditions under test bear a sufficiently direct relationship to the overall needs and conditions encountered by hearing impaired listeners for selections and evaluations conducted in the laboratory or clinic to be broadly applicable. At an individual level the problem becomes more pointed because not all listeners necessarily will experience the same range of acoustic conditions. It is argued here that one of the requirements of future research is to identify and standardize a restricted set of speech material and presentation conditions for viable cross-comparisons to be made, for both selection and evaluation purposes. Of course, this requirement lies not only in the domain of consumer judgement when speech and other materials are employed, but also applies equally to performance tests of speech intelligibility.

Speech Intelligibility Measures

Comparative evaluation of hearing aids using speech intelligibility measures became enshrined in audiological practice

(e.g., Carhart, 1946), and many tests have been developed and standardized for assessment purposes including monosyllabic word tests, sentence tests, nonsense syllables, and connected discourse tests with the aim of differentiating between hearing aid fittings. Subsequently, the usefulness of comparative evaluations was brought into question following a number of studies that showed that differences between hearing aids were simply not detectable (Shore, Bilger, & Hirsh, 1960; Resnick & Becker, 1963; Walden, Schwartz, Williams, Holum-Hardegen, & Crowley, 1983) unless the parameters of the hearing aid were so manipulated as to be grossly inappropriate (Jerger, Speaks, & Malquist, 1966). It, of course, remains unclear whether this failure to differentiate between hearing aids is due to the fact that the aids really were broadly equivalent or simply was because the speech tests used were just not sensitive enough to reveal any differences. The difficulties encountered when using traditional speech intelligibility measures can be easily appreciated when, for an individual subject, the critical difference for significance are assessed as by Thornton and Raffin (1978).

However, the interest in the issue of speech recognition assessment has not died perhaps because of the need to develop meaningful ways of evaluating the performance of new technology and because material advances have been made in the methodology of speech identification by the adoption of, for example, more robust psychometric procedures based upon adaptive tests. Examples of these successes in a research context are the use of an adaptive sentence procedure for the evaluation of a two-channel compression hearing aid (Moore, 1987), and application of an adaptive procedure in noise to study the effectiveness of noise suppression circuits (Van Tasell, Larsen, & Fabry, 1988). However, it must be appreciated that these speech measures are time-consuming and although viable on relatively small subject sets in research environments, their application to widespread clinical practice is not likely to be feasible unless the time devoted to the selection and evaluation process is greatly increased. It must also be appreciated that the tests have been made more sensitive by the application of conditions which might not necessarily find an analogue in listening conditions in everyday life. For example, a particular type of processing and hearing aid configuration that shows benefit in one sensitive test condition (e.g., the 50% point on the psychometric function of an adaptive speech-identification-in-noise procedure as in Moore, 1987) may not show benefit in more relevant listening conditions.

Another recent development is the advocacy of performance measures of speech identification ability that include indices not only of simple intelligibility, but also a component of perceptual effort required in the speech perception process. This is accessed by measures of response times (Gatehouse & Gordon, 1990; Gatehouse, 1992a). Recent evidence

54

JSLPA Monogr. Suppl. 1, Jan. 1993 / *ROA Suppl. de monogr. n° 1, janv. 1993*

suggests that such approaches can differentiate between hearing aid processing conditions that are extremely difficult to identify using traditional procedures (Baer, Moore, & Gatehouse, 1992). However, evidence has also been presented that calls into question the ability of speech identification measures, however they may be configured, to differentiate between conditions because of the potential existence of relatively long-term perceptual acclimatization effects (Gatehouse 1989, 1992b). This work suggests that differences between conditions only become apparent after the listener has substantial experience of listening through the appropriate environment. Note that this process is differentiated from the acceptability argument applicable to consumer judgements and involves an underlying perceptual process rather than problems associated with familiarity.

Of course, in considering appropriate speech intelligibility measures, the same issues of appropriateness of laboratory speech materials and acoustic conditions to everyday life arise as in the area of the consumer judgements discussed in the previous section. As such, the same requirements apply, although research effort is now being directed to the identification and evaluation in standard forms of such environments (e.g., Cox & Alexander, 1991).

Much of the foregoing discussion has been directed at hearing aid selection rather than hearing aid evaluation, and the area of evaluation remains a relatively untapped application for speech intelligibility measures, not with the aim of comparing the relative efficacy of two hearing aid fittings, but rather with the aim of assessing how well a particular hearing aid fitting overcomes the speech intelligibility deficits of the impaired listener. Given the construction of appropriate listening environments and speech materials, a future development could be the adoption of a small number of recognized targets for hearing aid provision that had been validated on large clinical samples and against which the large-scale deployment of new technology could be assessed. The existence of, say, between four and eight standardized sets of test material and acoustic environments could be used to ascertain on group or individual data to what extent hearing aid provision and the rehabilitation process has alleviated the speech recognition difficulties with respect to the performance of appropriate control populations.

Hearing Disability and Handicap Inventories

Whatever the success or failure of hearing aid fitting based on target formulae, consumer judgements, and speech intelligibility measures, it has to be recognized that one of the prime goals of managing hearing disability and handicap is to decrease the self-perceived problems of the hearing impaired

individual, and therefore, ultimately only those technologies that materially contribute to that process are likely to be successful. Numerous self-report instruments of hearing disability and handicap are available (for a review, see Schow & Gatehouse, 1990), and particular instruments have been advocated as potentially useful in the evaluation of hearing aid benefit (Newman, Weinstein, Jacobson, & Hugg, 1991; Demorest, 1985). There have been relatively few examples of differences between hearing aids or hearing aid provision strategies that have demonstrated such self-report instruments. (For an exception, see Birk-Nielsen [1976] who assessed the benefits of monaural vs. binaural hearing aids using such a methodology). This is perhaps due to the rather non-specific nature of such instruments, although recently effort has been devoted to developing instruments specifically aimed at assessing hearing aid benefit (Cox & Gilmore, 1990). A more serious limitation perhaps is the relatively time-consuming nature of such studies, which, by their very rationale, have to be conducted serially over time rather than allowing simultaneous comparisons between hearing aids in the clinic or laboratory. This is of course their very strength, given that the assessments made by the subject of the two hearing aids or provision strategies actually take place in the subject's own listening environment. However, this does place considerable resource constraints on the size of such studies, and it is likely that comprehensive experiments are only likely to take place when considerable advantages have been demonstrated in the laboratory.

Although the development of self-report instruments targeted at hearing aid benefit has taken place, these instruments are somewhat simple in nature and merely assess the degrees of difficulty before and after hearing aid provision (or the change in difficulty induced by hearing aid provision) on a set of situation specific questions. The questions may be worded in such a way to tap primarily into the speech communication difficulties associated with the situation, or the psychological, social, or emotional responses to that situation. However, there is good reason to argue from a conceptual point of view that other dimensions of initial disability and handicap can occur and that potentially influential variables lie not only in the domains of difficulty, but also in the domains of frequency of exposure to the difficulty, the importance of the difficulty to the individual, and the effect that the difficulty has on an individuals ability to perform adequately the tasks they require in their everyday life. The underlying dimensions of hearing aid benefit have received somewhat more attention (Brooks, 1979; Schow & Nerbonne, 1989), and individual components of hearing aid use, benefit, and satisfaction have been identified.

Work is currently in progress that attempts to identify and relate the underlying dimensions of hearing disability and handicap to those of hearing aid benefit (Gatehouse,

1992c). Preliminary findings suggest that the recognition of the multidimensional nature of these domains can improve insight into the ways in which particular types of hearing aid provision may or may not benefit hearing impaired individuals. In the same way as advocated for speech intelligibility measures, the development and standardization of instruments aimed at hearing aid benefit could be used to monitor the effectiveness of rehabilitation programmes with particular characteristics and types of hearing aids with respect to the needs, disabilities, and handicaps of the hearing impaired listener.

Agenda for the Future

Throughout this article I have attempted to identify those areas in which future research might be employed beneficially to overcome some of the underlying conceptual limitations of the procedures and processes involved in hearing aid selection and evaluation. The future demands on target formulae are likely to increase given the almost inevitable increases in the complexity of technology and, therefore, the number of ways in which a particular hearing aid or sets of hearing aids may be configured. Thus, the validation on appropriately large and appropriately stratified clinical samples of the competing target schemes remains a research priority, with a specific aim of providing reliable data on the deviation from the target standards that are acceptable. However, this research presupposes the existence of stable and sensitive evaluation procedures with which to validate the target formulae themselves.

Although there has been much progress in the methodologies employed in consumer judgements and speech intelligibility, there still remains much to establish and standardize. A set of appropriate listening environments in terms of speech material and acoustic presentation conditions awaits concerted agreement amongst researchers and health care professionals so that realistic comparisons can be made between laboratories and clinical populations. This article has identified deficits in the methodology of consumer judgement instruments in terms of the overall stability and effects of previous experience. Although new and more stable methods of assessing speech intelligibility are becoming available, the challenge of providing instruments appropriate for clinical rather than research settings remains formidable.

At a more general level, this author would advocate a switch of emphasis from the selection process to the evaluation process with the specific suggestion that a set of self-report and performance measures of speech intelligibility be developed and standardized over an appropriately representative (but relatively sensitive) set of speech materials and acoustic conditions likely to be encountered by the hearing

impaired listener. These standards may then be used as evaluation methods to judge: (1) the success or failure of the overall rehabilitation process in an individual; (2) the degree to which health care delivery systems are meeting the needs of the hearing impaired population; and when new technology is employed in widespread clinical practice; (3) the benefits with respect to the overheads in terms of a laboratory experiment (e.g., the 50% point on a performance intensity function in noise) and in relation to the disability and handicaps that the technology is attempting to overcome.

Address all correspondence to: Stuart Gatehouse, Scientist-in-Charge, MRC Institute of Hearing Research, Scottish Section, Queen Elizabeth Building, Glasgow Royal Infirmary, 16 Alexandra Parade, Glasgow G31 2ER

References

Baer, T., Moore, B.C.J., & Gatehouse, S. (1992). Spectral contrast enhancement of speech in noise: Effects on intelligibility, quality and response times for listeners with sensorineural hearing impairment. *Journal of Rehabilitation Research and Development.* In Press.

Berger, K.W., Hagberg, E.N., & Raine, R.L. (1984). Prescription of hearing aids: Rationale procedures and results. *4th Ed. Kent, OH: Herald.*

Berger, K.W. (1990). The use of an articulation index to compare three hearing aid prescriptive methods. *Audecibel, 39*(3), 16-19.

Birk-Nielsen, H. (1976). Effect of monaural vs. binaural hearing aid treatment. *Hearing Journal, 8,* 9-29.

Brooks, D.N. (1979). Counselling and its effect on hearing aid use. *Scandinavian Audiology, 8,* 101-107.

Byrne, D. (1984). *Evaluation of the National Acoustic Laboratories procedure for selecting the gain and frequency response of a hearing aid.* Unpublished PhD Thesis, Macquarie University, Sydney, Australia.

Byrne, D. (1986a). Effects of bandwidth and stimulus time on most comfortable loudness levels of hearing impaired listeners. *Journal of the Acoustical Society of America, 80,* 484-493.

Byrne, D. (1986b). Effects of frequency response characteristics on speech discrimination and perceived intelligibility and pleasantness of speech for hearing impaired listeners. *Journal of the Acoustical Society of America, 80,* 494-504.

Byrne, D., & Dillon, H. (1986). The National Acoustics Laboratories (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear & Hearing, 7,* 257-265.

Carhart, R. (1946). Tests for selection of hearing aids. *Laryngoscope, 56,* 780-794.

Cox, R.M. (1985). Hearing aids and aural rehabilitation: A structured approach to hearing aid selection. *Ear & Hearing, 6,* 226-239.

Cox, R.M., & Gilmore, C. (1990). Development of the profile by hearing aid performance (PHAP). *Journal of Speech and Hearing Research, 33,* 343-357.

56

JSLPA Monogr. Suppl. 1, Jan. 1993 / *ROA Suppl. de monogr. n° 1, janv. 1993*

Cox, R.M., & Alexander, G. (1991). Hearing aid benefit in everyday environments. *Ear & Hearing, 12*, 127-139.

Demorest, M. (1985). Techniques for measuring hearing aid benefit through self-report. In M. Pickett (Ed.) *Symposium on Hearing Aid Technology: Its Present and Future*. Washington, DC: Gallaudet Press.

Gatehouse, S. (1989). Apparent auditory deprivation effects of late onset: The role of presentation level. *Journal of the Acoustical Society of America, 86*(6), 2103-2106.

Gatehouse, S., & Gordon, J. (1990). Response times to speech stimuli as measures of benefit from amplification. *British Journal of Audiology, 24*, 63-68.

Gatehouse, S. (1992a). Development and evaluation of a sentence verification test for the assessment of hearing aid benefit. *Ear & Hearing*. Manuscript Submitted For Publication.

Gatehouse, S. (1992b). The time-course and magnitude of perceptual acclimatization to frequency responses: Evidence from monaural fitting of hearing aids. *Journal of the Acoustical Society of America*. In Press.

Gatehouse, S. (1992c). [Determinants and components of auditory disability and hearing aid benefit]. Unpublished data.

Goldstein, D.P., & Stephens, S.D.G. (1981). Audiological rehabilitation: Management model 1. *Audiology, 20*, 432-452.

Jerger, J., Speaks, C., & Malquist, C. (1966). Hearing aid performance and hearing aid selection. *Journal of Speech and Hearing Research, 9*, 136-149.

McCandless, G.A., & Lyregaard, P.E. (1983). Prescription of gain/output (POGO) for hearing aids. *Hearing Instruments, 34*(1), 16-21.

Moore, B.C.J. (1987). Design and evaluation of a two-channel compression hearing-aid. *Journal of Rehabilitation Research and Development, 24*, 181-192.

Neumans, A.C., Levitt, H., Mills, R., & Schwander, T. (1987). An evaluation of three adaptive hearing aid selection strategies. *Journal of the Acoustical Society of America, 82*, 1967-1976.

Newman, C.W., Weinstein, B.E., Jacobson, G.P., & Hugg, A. (1991). Test-retest reliability of the hearing handicap inventory for adults. *Ear & Hearing, 12*, 355-357.

Pavlovic, C.V. (1988). Articulation index predictions of speech intelligibility in hearing aid selection. *ASHA, 30*(6), 63-65.

Pavlovic, C.V., Sorin, C., Roymiguiere, J-P., & Lucas, J-P. (1990). Cross validation between a direct magnitude estimation technique and a paired comparison technique for assessing the quality of text-to-speech synthesis systems. *Journal Acoustique, 3*, 75-83.

Punch, J.L., & Parker, C.A. (1981). Pairwise listener preferences in hearing aid evaluation. *Journal of Speech and Hearing Research, 24*, 366-374.

Rankovic, C.M. (1991). An application of the articulation index to hearing aid fitting. *Journal of Speech and Hearing Research, 34*, 391-402.

Resnick, D.M., & Becker, M. (1963). Hearing aid evaluation - a new approach. *ASHA, 5*, 659-699.

Shore, I., Bilger, I.C., & Hirsh, I.J. (1960). Hearing aid evaluations: Reliability of repeated measures. *Journal of Speech and Hearing Disorders, 25*, 152-170.

Schow, R.L., & Gatehouse, S. (1990). Fundamental issues in self-assessment of hearing. *Ear & Hearing, 11*(5), Suppl. 6-16.

Schow, R.L., & Nerbonne, M.A. (1989). *Introduction to aural rehabilitation*. Second Edition. Austin, Texas: Pro-Ed.

Skinner, M.W., Pascoe, D.P., Miller, J.D., & Popelka, G.R. (1982). Measurements of the region of the optimum placement of speech energy within the listener's auditory area: A basis for selecting amplification characteristics. In G.A. Studebaker & F.H. Bess (Ed.) *Vanderbilt Hearing Aid Report* (161-169). Upper Darby, PA: Monographs in Contemporary Audiology.

Skinner, M.W. (1988). *Hearing aid evaluation*. Englewood Cliffs, NJ: Prentice Hall.

Stephens, S.D.G. (1987). Auditory rehabilitation. *British Medical Bulletin, 43*, 999-1026.

Studebaker, G.A., Bisset, J.D., & Van Ort, D.M. (1982). Paired comparison judgements of relative intelligibility in noise. *Journal of the Acoustical Society of America, 72*, 80-92.

Sullivan, J.A., Levitt, H., Whang, J.Y., & Hennessey, A.M. (1988). An experimental comparison of four hearing aid prescription methods. *Ear & Hearing, 9*, 22-32.

Thomson, G., & Lassman, F. (1969). Relationship of auditory distortion test results to speech discrimination through flat versus selected amplification systems. *Journal of Speech and Hearing Research, 12*, 594-686.

Thornton, A., & Raffin, M. (1978) Speech discrimination scores modelled as a binominal variable. *Journal of Speech and Hearing Research, 21*, 507-518.

Van Tasell, D., Larsen, S., & Fabry, D. (1988). Effects of an adaptive filter hearing-aid on speech recognition in noise by hearing-impaired subjects. *Ear & Hearing, 9*, 15-21.

Walden, B.E., Schwartz, D.M., Williams, D.L., Holum-Hardegen, L.L., & Crowley, J.M. (1983). Test of the assumptions underlying comparative hearing aid evaluations. *Journal of Speech and Hearing Disorders, 48*, 264-273.

World Health Organization (WHO). (1980). *International classification of impairments, disabilities and handicaps: A manual of classification relating to the consequence of disease*, 25-43.

JSLPA Monogr. Suppl. 1, Jan. 1993 / *ROA Suppl. de monogr. n⁰ 1, janv. 1993*

57