Paired Comparisons as a Fine-Tuning Tool in Hearing Aid Fittings

FRANCIS K. KUK

The Need to Go Beyond Prescriptive Formulas

The use of prescriptive formulas to specify frequency-gain characteristics of hearing aids simplifies hearing aid selection. Satisfaction is assumed when the measured gain of the hearing aids matches the gain prescribed by the formulas. Knowledge of individual preference and psychophysical skills are not required or needed when using this selection process. An implicit requirement for the use of prescriptive formulas is that audiologic indices [e.g., thresholds, most comfortable listening levels (MCLs), and loudness discomfort levels (LDLs)] can be measured reliably and accurately. Furthermore, when predicting 2-cc coupler frequency-gain characteristics, the formulas require that the individual wearer have ear canal characteristics that resemble those of the average ear.

Unfortunately, neither of these requirements can be met all the time. For example, several investigators have shown that auditory thresholds could vary from 4 to 9 dB hearing level (HL) upon retest (Byrne and Dillon, 1981; Skinner and Miller, 1983). Suprathreshold loudness judgments (i.e., MCLs, LDLs) can be even more variable depending on instructions, stimulus type, and psychophysical method employed to measure such indices (Cox and Bisset, 1982; Skinner and Miller, 1983). Fluctuations in the listener’s internal criteria could also affect the reliability of such indices. Additionally, the presence of a fluctuating loss (e.g., Meniere’s disease) or of an ear with atypical middle ear resonance and impedance characteristics may result in the prescribed frequency-gain response as being inappropriate for the hearing aid wearer (Gilman et al, 1981). These observations suggest that, unless one can control the intra- and intersubject variability seen in audiologic measurements and account for the individual differences when specifying frequency-gain response, there is the potential that the prescribed frequency-gain response may be inappropriate for the wearer.

Assuming that audiologic indices can be defined accurately and that target gain may be achieved, individual frequency-gain responses specified by group data (as in the prescriptive approach) may not be satisfactory to all hearing aid wearers. For example, Neuman et al (1987) showed that seven of eight hearing aid wearers preferred frequency-gain responses that were different from those prescribed when using an MCL approach (Pascoe, 1978). Kuk and Pape (1992) reported that 18 of 20 subjects selected frequency-gain responses that devi-
ated from the National Acoustic Laboratories' (NAL) prescription (Byrne and Dillon, 1986). On the other hand, Byrne and Cotton (1988) reported that only 10 to 20% of their hearing aid wearers selected a frequency-gain response that deviated significantly from the NAL prescription. These observations suggest that additional measures are needed to verify the appropriateness of the prescribed frequency-gain response to ensure maximum satisfaction in all wearers. Alternative methods to select frequency-gain response on hearing aids may be necessary. There is increasing evidence that hearing aid wearers prefer a unique frequency-gain response from their hearing aids when they listen in various acoustic environments and/or when they use different criteria to select preferred frequency-gain responses. For example, Byrne (1986) found that preferred frequency-gain response that was selected with a criterion of “pleasantness” was different from that selected based on the judgment of “intelligibility.” Kuk (1990) demonstrated that the preferred insertion gain for listening to one’s own vocalization is different from listening to externally presented stimuli. Tecca and Goldstein (1984) showed that hearing-impaired listeners preferred less low-frequency gain from a hearing aid as the stimulus level was increased. Kuk et al (1994) showed that the preferred frequency responses for speech in noise ratio (SNR) of the input signal. Presently, no guidelines are available to specify how these frequency-gain responses should be prescribed. The concerned clinician needs to ensure that the measured frequency-gain response provides optimal performance under these special listening situations or that the frequency-gain response can be modified to accommodate the listener’s needs in these situations.

Advances in technology improve the sophistication of hearing aids. Today, 70% of hearing aids sold in the United States are either nonlinear analog hearing aids (21%) or programmable and digital hearing aids (49%) that use nonlinear signal processing (Skafte, 2000). Although these newer compression hearing aids and digital signal processing (DSP) hearing aids promise new and better ways to serve individuals with hearing impairment, prescriptive formulas that are available were based on single-channel, linear signal processing hearing aids. Thus, these formulas did not consider the effect of the number of processing channels and release times on the actual real-world output of the hearing aids (Kuk and Ludvigsen, 1999). Consequently, although the output of multichannel nonlinear hearing aids may match a prescriptive target, the wearer may be dissatisfied with the real-world output of the hearing aids. Unlike linear hearing aids with which there are at least several generally accepted approaches for their fitting, there are only a few manufacturers of nonlinear hearing aids that provide guidelines to clinicians on fitting these devices. Unfortunately, methods to evaluate the adequacy of the manufacturers’ guidelines or of the performance of these devices are lacking. As in traditional linear hearing aids, verification of optimal settings on these devices is necessary to ensure success with the special circuits. One approach that can help to verify the appropriateness of the selected hearing aid settings and can potentially be helpful in hearing aid selection and fine-tuning is the paired comparison technique.

What Is Paired Comparison?

The method of paired comparison was attributed to Fechner as a data collection technique for the study of sensory perception (Thurstone, 1927). In this method, an experimental subject makes binary decisions on a number of stimuli that are presented in pairs. The experimenter usually sets the criteria of judgment. The subject’s task is to indicate which one of the two comparison stimulus intervals meets the experimenter’s criterion. In hearing aid research, the results of the comparison may show the relative “perceptual” distances among a list of items being compared (Punch et al, 1980), or the comparison may show the relative ranking
of the comparison hearing aids (Punch, 1978). Moreover, the comparison may result in a recommendation of the best hearing aids (Studebaker et al, 1978). Finally, a combination of settings on programmable or digital hearing aids may be recommended based on the results of this process (Levitt et al, 1978; Neuman et al, 1987; Kuk and Pape, 1992).

Figure 4–1 is a block diagram representation of the sequence of events involved in a paired comparison trial. Two settings (e.g., different low-frequency gains) on a programmable hearing aid are compared. There are three important components in a single paired comparison trial: the instructions, the presentation, and the decision. In the instructional stage, the clinician explains the task and specifies the criterion for judgment to the listener. For example, one may use the following instructions for selection of a hearing aid setting that maximizes speech intelligibility:

You will listen to a short passage played back twice, each time with a different hearing aid setting. As the passage is played back, I will indicate to you which is hearing aid A and which is hearing aid B. After you have listened through both hearing aid settings, you must indicate which hearing aid (A or B) yields more intelligible speech. By that I mean the hearing aid setting with which you can understand more of the spoken passage. You must indicate one preference even though you may find these two hearing aid settings to sound very similar. I will gladly repeat the presentation if you so desire. Do you have any questions?

During the presentation stage, the clinician adjusts the hearing aids to the appropriate settings for comparison. These two combinations of settings are stored in the temporary memory of each programmable/digital hearing aid. The clinician activates the first setting of the hearing aid, alerts the listener of hearing aid setting A, and plays back the stimulus. Typically, short (between 10 and 20 seconds) discourse passages can be used as stimuli for comparison. Afterward, the clinician activates the second setting of the hearing aid, announces hearing aid setting B, and plays back the same passage for comparison. This process is repeated until the listener is ready to make a decision on the preferred hearing aid setting. Because of the way the hearing aid settings are labeled, paired comparison is also commonly known as AB comparison.

The listener must make a preference judgment on the two hearing aid settings in the decision stage. He or she must decide if hearing aid setting A or hearing aid setting B meets the criterion of better speech intelligibility regardless of the similarity or dissimilarity between the two comparison hearing aids. Responses like "no difference" and "they both sound good/bad" are not acceptable. If such is the response, the listener will be re instructed, and the same stimulus will be presented again.

Although the performance of hearing aids selected using paired comparison technique has been favorable, absolute performance is not measured, and thus its performance is not guaranteed. The hearing aid settings that are selected using paired comparisons reflect the listener's relative judgment for the settings available for comparison. Maximum satisfaction is guaranteed only if at least one of the available combinations results in maximum satisfaction. Direct measurement of listener satisfaction with the

![Figure 4–1. Sequence of events in a single paired comparison trial.](image-url)
selected hearing aid settings is needed to determine absolute performance.

**History of the Use of Paired Comparison in Hearing Aid Selection**

A requirement for using paired comparison to evaluate hearing aids in the clinic is the ability to switch rapidly among various electroacoustic settings (or hearing aids) for comparison. This was impossible using conventional hearing aids without undue delay during the switching process. Zerlin (1962) is credited with first proposing a manageable way of performing paired comparison.

In Zerlin’s method, speech was recorded through two hearing aids that were coupled to separate couplers. Output from the couplers was recorded on two separate tracks of a magnetic tape. Different pairs of hearing aids were connected to the couplers, and processed speech was recorded in the sequence in which it would be presented during the evaluation. The output from the tape recordings was presented to the listeners through earphones. Listeners switched between the two tracks of the tape and indicated preferences for one of the two taped-speech segments. This was used to indicate their preference for the hearing aid that was used to process the speech signal. Figure 4–2 shows a schematic diagram of how stimuli were recorded and played back in the earlier paired comparison trials. This approach of recording and then playing back the stimuli was improved by subsequent investigators. For example, a Knowles Electronic Manikin for Acoustic Research (KEMAR) with a Zwislocki coupler was used to record in the sound field to approximate the acoustic effects associated with the head and body baffle and ear canal response (Punch, 1978). Hearing aid receivers and/or earmolds were used as output transducers in later studies (Studebaker et al, 1980).

Although this approach allows rapid comparison between speech processed by different hearing aid settings, this is impractical for clinical use because of the labor involved in the recording process, the limitations in the number and variations of hearing aid settings that can be compared, the time involved in clinical comparison, and the difficulty of transferring the results of the paired comparison to commercial hearing aid use. Paired comparison was primarily used as a research tool.

Clinical pairwise quality judgments have been made with a master hearing aid (Watson and Knudsen, 1940; Pascoe, 1975). Using this approach, the clinician adjusts the settings on a master hearing aid (the size of a portable audiometer), and speech (taped or live) is delivered through headphones to the

![Figure 4-2: Instrumentation involved in original paired comparison trials.](image)
listeners, who then indicate their preference as the settings are adjusted and presented in pairs. The setting that is preferred will be recommended. Although the labor involved in recording speech stimuli is eliminated, accurate simulation of true hearing aid performance in commercial hearing aids is still a major difficulty with the master hearing aid approach. Guidelines to vary the settings systematically were also lacking.

Levitt et al (1978) were the first to vary the frequency-gain responses systematically on a wearable master hearing aid using an adaptive procedure (the simplex procedure). In this procedure, listeners identified nonsense syllables while wearing a master hearing aid that was adjusted to various settings. Through selective testing of choice electroacoustic settings, the simplex procedure converges at a combination of settings on the master hearing aid that yields a maximum speech understanding score. Although the method had significant appeal, it was not adopted for clinical use because of its complexity, inefficiency, and dependence on a computer. Several hours were needed to select a frequency-gain response using the simplex procedure.

Neuman et al (1987) modified the simplex procedure to incorporate the use of pairwise comparison. Rather than comparing speech recognition scores among the selected settings, the modified simplex procedure used subjective judgment of relative intelligibility as a criterion and allowed systematic pairwise comparison of settings. These modifications substantially reduced the time to select a frequency-gain response.

The advent of digital and programmable hearing aids facilitates the use of paired comparisons in a clinical environment. Four features of such hearing aids contributed significantly in this regard. The reader is warned that not all digital and all programmable hearing aids share all four features. First, the wide range of electroacoustic adjustments on some digital and programmable hearing aids allows them to be viewed as stand-alone wearable master hearing aids and to be used for a wide range of listening conditions and for a wide range of hearing loss configurations.

Second, in many digital and programmable hearing aids, switching of settings can be performed rapidly through the use of a programming cord or via a remote control device using frequency-modulated (radio-frequency and ultrasonic) signals. This allows rapid switching between comparison settings.

Third, the availability of multiple memories in some digital and programmable hearing aids also facilitates comparison of frequency-gain responses in the clinic and in everyday listening situations.

Fourth, several manufacturers of digital and programmable hearing aids can interface their units directly to external computers so that paired comparison can be performed in an automatic manner. The computer controls for stimulus delivery, tracks responses, and adjusts settings on the hearing aid according to defined rules and algorithms. This could significantly reduce the time involved in the comparison and improve the reliability in which comparisons are made. Indeed, Kuk (1992) demonstrated the feasibility of adapting the use of the modified simplex procedure to select frequency-gain responses for a commercially available programmable multimemory hearing aid. Some recent digital hearing aids also have automated algorithms that allow for paired comparisons.

Paired Comparison Techniques

Although paired comparison involves only binary decisions on pairs of stimuli, stimulus pairing (i.e., the manner in which the different settings are paired and compared) affects the information available from the comparison. The different strategies in which this technique has been used in hearing aid research include the round-robin tournament, single- and double-elimination tournaments, simple up-down procedure, and the modified simplex procedure. Although these strategies can be performed either manually or through the assistance of a PC-based software, the
use of a computer with custom software could greatly facilitate comparison.

**Round-Robin Tournament**

The object of a round-robin tournament is to rank order the available hearing aids or different settings within hearing aids based on some defined criteria. In this approach, every hearing aid (or combinations of settings) is paired and compared with every other hearing aid (or settings). For \( N \) hearing aids or combinations of settings, a round-robin tournament involves \( N(N - 1)/2 \) pairs of comparisons. As a result, hearing aids are ranked according to the frequency in which they are chosen.

Figure 4–3 illustrates the manner in which hearing aids containing the necessary settings appropriate to achieve one of four prescriptive formulas (i.e., A, B, C, and D) are compared in a round-robin tournament (i.e., \( N = 4 \)). For four hearing aid settings, the number of comparisons would be \( 4(4 - 1)/2 \) or 6. This includes the comparisons between A and B, A and C, A and D, B and C, B and D, and C and D (bottom row). Subjects may be asked to judge which one of two hearing aid settings provides clearer speech or better intelligibility (or any other criteria) as they listen to discourse passages presented in a noise background. Assuming that the winners of these comparisons are B, A, A, B, B, and C, respectively (second row), the relative ranks of these four hearing aid settings will be B, A, C, and D because B has the most wins and D has no wins at all. As indicated earlier, the result of a round-robin tournament is a rank order of hearing aid prescriptions based on subject preference.

The advantage of the round-robin tournament is its ability to rank order hearing aids. In addition to selecting the best hearing aid (or setting) among the comparisons, the round-robin tournament also allows the study of relationships among electroacoustic parameters and their relative contribution to the perceptual process. For example, Punch et al. (1980) used this approach to study factors governing subjective preference for hearing aid processed speech.

From a clinical standpoint, the round-robin tournament may be practical only if (1) specific ranking information is needed among the comparison hearing aids; and (2) the number of comparison hearing aids (or settings) is small (i.e., under four). Other methods may be more efficient to verify the

---

**Round Robin Tournament**

4 Hearing aids - A, B, C, D

# comparisons - \( N(N-1)/2 \), or \( 4(4-1)/2 = 6 \)

![Round Robin Tournament Diagram](image)

**Figure 4–3.** Example of a round-robin tournament. Letters A to D represent four different hearing aids each fitted using specific prescriptive formulas.
appropriateness of a selected frequency-gain response. For example, Neuman et al (1987) reported that it required an average of 83.8 minutes to complete a round-robin tournament, but only 8.0 minutes to complete a modified simplex procedure when the same combinations of settings (a total of 25) are compared.

**Single-Elimination Tournament**

The object of an elimination tournament is to determine which one of several hearing aids (or combinations of settings within a hearing aid) is more preferable than the others. In this approach, each hearing aid is first compared with one other hearing aid. Winners of each pair of comparison are further compared several times to result in an overall winner. For \( N \) hearing aids where \( N \) is an integer power of 2, there will be \( \frac{N}{2}-1 \) pairs of comparison hearing aids.

An example of a single-elimination tournament is shown in Figure 4–4 where eight hearing aids are compared. Letters identify the hearing aids in the comparison (A–H). In the first round of comparison, A is compared to B, C to D, E to F, and G to H. The winners of the comparison, in this case A, C, E, and G, are compared in a second round where A is compared to C, and E is compared to G. The winners of the second round of comparison, A and E, are further compared in the third round to determine a winner (A). A hearing aid is eliminated if it loses any comparison. However, a hearing aid can only become the winner when it has won \( M \) rounds of comparison where \( 2^M \) equals \( N \), the total number of hearing aids. For example, in Figure 4–4, three rounds of comparisons are needed to determine a winner when eight \((2^3)\) hearing aids are compared. No ranking information is available from the results of an elimination tournament.

The manner in which hearing aids are paired in a single-elimination tournament is important. Comparison hearing aids can be paired either by random assignment or according to some seeding rules. A good seeding rule is to pair hearing aids with greater difference in their frequency-gain characteristics during the first round. Variability in the selection will be increased if hearing aids with similar electroacoustic characteristics are paired (White and Studebaker, 1978; Montgomery et al, 1982).

**Single Elimination**

\[ \#\text{comparison} = N-1, \text{ or } 8-1 = 7 \]

![Diagram](image)

*Figure 4–4.* Example of a single-elimination tournament. Letters A to H represent eight different hearing aids for comparison.
Double-Elimination Tournament

When two hearing aids with relatively similar characteristics are compared, it is possible that the "better" hearing aid is eliminated because of random error. To safeguard a "good" hearing aid from being eliminated in the early rounds of comparison, Studebaker and his colleagues (1979a, 1980) proposed the use of a double-elimination tournament as a means to compare hearing aids (or combinations of settings within a single hearing aid). Instead of being eliminated after only one loss, each hearing aid has to lose twice before it is eliminated from the tournament. Higher test-retest reliability was reported with the double-elimination tournament than with the single-elimination tournament (Studebaker et al, 1979a).

Figure 4–5 illustrates a double-elimination tournament. Assume again that the same eight hearing aids (A–H) are compared. During the first round of comparison, hearing aid A is compared to hearing aid B, C to D, E to F, and G to H. Winners of this round of comparison (in this case A, C, E, and G) are further compared within the winners' bracket. On the other hand, losers in this round are compared among themselves (B with D and F with H) to yield the winners in the losers' bracket. During the second round of comparison between A and C and E and G, the losers of these comparisons in the winners' bracket (C and G) are compared with the winners (B and F) in the losers' bracket after the first round of comparison. In essence, the losers in the winners' bracket are given one more chance to compare with the winners in the losers' bracket. The winner in each bracket (A in winners' bracket and F in losers' bracket) is compared one last time to determine the overall winner (i.e., hearing aid A). There are $2(N - 1)$ pairs of comparisons for $N$ hearing aids, where $N$ is an integer power of 2. In this example, there are a total of $2(8 - 1)$ or 14 pairs of comparisons.

Simple Up-Down Procedure

The "goodness" of the verified settings using a round-robin or an elimination tournament depends on the "goodness" of all settings selected for comparison. One may need to compare with a large number of settings to

**Double Elimination**

# comparison = $2(N-1)$ or $2(8-1) = 14$

![Diagram showing a double-elimination tournament](image)

Figure 4-5. Example of a double-elimination tournament. Letters A to H represent eight different hearing aids for comparison.
ensure that a recommended setting is indeed the best setting available on the hearing aids. Because of potentially large numbers of settings that need to be compared and the time constraints in a clinical setting, the use of round-robin and elimination tournaments may not be practical.

An adaptive procedure is an estimation procedure in which a listener’s response on a test trial dictates the direction of stimulus change in the next trial. This helps to focus comparisons on only those settings contributing to the estimation and allows one to fine-tune a selected setting without testing a large number of settings. As in all estimation procedures, one assumes the existence of only one combination of hearing aid settings that will optimize listener preference. The goal of the adaptive procedure is to estimate that setting in the shortest amount of time while satisfying the clinician’s and/or the hearing aid wearer’s criteria.

The simple up-down procedure is an adaptive procedure that allows verification (and fine-tuning) of settings in one dimension. In using this procedure, the clinician first estimates the initial hearing aid setting that the listener may find most satisfactory. This initial setting is the initial estimate. This initial setting can be based on any prescriptive formula, on manufacturer recommendations, or on clinician intuition. When the method of paired comparison is used, this initial estimate is compared with another setting that differs from the initial estimate on the same adjustable parameter of the hearing aid. A selected criterion (e.g., relative intelligibility) is used for comparison between two settings. Any acoustic stimuli can be used for judgment (e.g., discourse presented in noise). Comparison continues with different settings that are varied in a systematic manner until the criterion to terminate comparison is met.

Figure 4–6 is an illustration of a simple up-down procedure to determine optimal low-frequency setting on a hearing aid. Assume that the initial estimate corresponds to a low-frequency setting at interval 3 (open rectangle). The listener compares this setting with interval 4 (black rectangle). The direction of comparison during the first trial may be arbitrary or purposeful. The change in adjustment between these settings is termed

**Simple Up-Down**

![Figure 4–6](image-url)
a step, and the magnitude of such change is termed the step size. Assume that the listener prefers the setting with more low-frequency gain (i.e., interval 4); this indicates that the direction of the next comparison should be in the direction of more low-frequency gain (i.e., intervals 4 and 5 will be compared). A comparison in the same direction is termed a run (trials 1 and 2). If during the second trial the listener prefers less low-frequency gain (i.e., interval 4 to interval 5), a reversal (in the direction of preference) has occurred. In Figure 4–6, reversals occurred after trials 2, 3, and 4. Although one can continue the comparison indefinitely, comparison is terminated after the third reversal, with interval 4 being the listener’s preferred low-frequency setting on the hearing aid. This preferred setting is termed the final estimate. In this case, the initial estimation of optimal low-frequency gain differs from the listener’s preference by one interval.

The simple up-down procedure differs from the round-robin and elimination tournaments in the manner in which settings are selected for comparison and the number of comparisons. Assume that eight intervals on the low-frequency dimension are compared. The round-robin tournament requires all eight intervals to be compared with each other to result in 28 (or 8(8 – 1)/2) pairs of comparisons. Seven (or 8 – 1) pairs of comparison are needed in the single-elimination tournament and 14 (or 2(8 – 1)) pairs are needed in the double-elimination tournament. In the simple up-down procedure illustrated previously, only four comparisons are needed to estimate the preferred low-frequency gain.

It is important to note that whereas all intervals are compared in the round-robin and elimination tournaments, only the intervals near the final estimate are compared several times in the simple up-down procedure. Intervals 1, 2, 6, 7, and 8 are not compared. This observation, however, is only true when the initial estimate is close to the final estimate and when the listener is consistent in his or her judgments. More comparisons will be needed if these assumptions are not met.

Because not all settings are evaluated in the simple up-down procedure, an occasional listener not showing strong preference for a particular setting may reveal random preference during the evaluation and lead to errors in the estimation procedure. The round-robin and elimination tournaments, because of their sampling of all comparison settings, will not be as affected by homogeneous preference.

There are several issues to consider when using a simple up-down procedure. These issues are fully discussed by Levitt (1970, 1978, 1992) but will be briefly reviewed for completeness.

**Choice of Initial Estimate**

In theory, the initial estimate should be as similar to the final estimate as possible to ensure maximum efficiency. The disadvantage of a dissimilar initial estimate is that it will require more comparisons to reach the final estimate. Kuk and Lau (1995a) compared the convergence time (time to reach the final estimate) of four initial estimates [NAL revised (NAL-R), low- and high-frequency variations of NAL-R, and a flat response). They found that all four initial estimates resulted in the same final estimate, but the convergence time differed. The NAL-R setting showed the least convergence time and was the closest to the final estimate. The advantage of an initial estimate that is very different from the final estimate is that these estimates may enable listeners to discriminate better the differences between stimulus intervals and to be better familiarized with the test routine. Levitt (1970, 1978, 1992) recommended that data obtained from the first run should not be analyzed so as to minimize bias resulting from an inappropriate initial estimate.

Different approaches have been attempted to determine the best initial estimate. For example, Neuman et al (1987) used a frequency-gain response that placed the aver-
age speech spectrum at the user’s MCL as the initial estimate. Kuk and Pape (1992) used
the frequency-gain response recommended by NAL-R as the initial estimate. Both stud-
ies, however, showed that the final estimates deviated from the initial estimates for a ma-
jority of subjects. Furthermore, the listening conditions, types of stimuli, and criteria used
to make the comparison can alter the final estimate. In other words, there may not be one
fixed optimal initial estimate for all test conditions. Clinicians who use this technique
for the first time may consider following the approach of Leijon et al (1991) and Kuk
and Pape (1992) by using frequency-gain response recommended by NAL-R as the ini-
tial estimate. After trying this on a few listeners (perhaps 10 to 20), it should become
possible for clinicians to choose initial estimates based on their knowledge of the final
estimates obtained under the same listening condition. Furthermore, the default settings
recommended by most manufacturers of digital and programmable hearing aids are
also good initial estimates to use when performing paired comparisons with a commer-
cial hearing aid.

An alternative approach is to compare several distinct settings (less than four) on
the hearing aid using a tournament strategy. The winner of the comparisons may be used
as the initial estimate for selecting the best setting available on the hearing aids.

STEP SIZE
The ideal step size for statistical efficiency de-

pends on the consistency of the listener’s re-
sponse and the accuracy of the preceding esti-
mates. Theoretically, a small step size should
be used if the listener is consistent in his or
her response and if the estimate is close to the
final estimate. When using small step sizes,
the precision of the estimation is increased,
but at the expense of increasing evaluation
time. A large step size allows one to converge
at the final estimate quickly, but at the ex-

pense of losing estimation precision. In prac-
tice, one would not know how consistent a

listener is until after the comparison has
started. As a general rule, the step size used
must be larger than the just noticeable differ-
ence (JND) of the wearer. For the average
hearing-impaired wearer, this is roughly 5 dB
gain change in the low- or high-frequency
settings (Kuk, 1994). Because hearing-
impaired wearers differ in their psychophysi-
cal abilities, the use of a fixed step size may
not be the most efficient.

Robbins and Monro (1951) recommended
starting with a large step size (arriving at the
vicinity of the final estimate sooner) and
gradually reducing the step size as compari-
on continues. Mathematically, the step size
on run $N$ is $d/N$, where $d$ is the step size
used on the first run. For example, the step
size for the third run may be reduced to one
interval if the step size for the first run was
set at three intervals. In practice, this rule
may be difficult to implement on commer-
cial analog hearing aids because most can
vary only in limited intervals on each pa-
rameter. This may not be a problem with
digital and programmable hearing aids. A
compromise is to start the first run with a
step size of two intervals and reduce it to
one interval after the first reversal.

TERMINATION RULE
Rules to terminate an adaptive procedure
are necessary after the clinician is reasonably
certain that the final estimate reflects the lis-
tener’s preference. In the example provided
earlier, three reversals were required to ter-
minate the comparison. In theory, one can
terminate a comparison after the first rever-
sal. The accuracy, however, of the estimation
may not be acceptable. Precision and reliabil-
ity in estimation improve as the number
of reversals is increased. However, the time
involved in the estimation will be increased
also. In practice, three reversals can be used
with fair reliability (Neuman et al, 1987; Kuk
and Pape, 1992).

The simple up-down procedure, in its var-
ious forms, has been used to estimate thresh-
olds, to estimate MCL (Wall and Gans, 1984),
and to select and verify low-frequency gain on a noise reduction hearing aid (Kuk et al, 1992).

**Modified Simplex Procedure**

The simplex procedure was originally proposed by Box (1957) as a means to optimize productivity. It was adapted by Levitt et al (1978) and later modified by Neuman et al (1987) as an alternative to select the optimal settings on more than one parameter on a wearable master hearing aid. Like the simple up-down procedure, the modified simplex procedure assumes the existence of one and only one combination of settings on a hearing aid that optimizes listening under a specific condition.

Like the simple up-down procedure, the process starts with an estimation of the listener’s preferred settings on the hearing aid (prescriptive phase to determine initial estimate). This combination is compared with other combinations of settings in a systematic manner (adaptive phase to determine final estimate) until the termination rules are met. The goal of the comparison is to select a combination of settings that reflects the listener’s preference for the listening condition.

Figure 4–7 shows the matrix representation of the frequency-gain response of a programmable hearing aid in the high- and low-frequency dimensions. Each cell on the matrix represents a unique combination of high- and low-frequency gain. For example, in Figure 4–7, cell (2L, 2H) represents the frequency response with a low-frequency setting of 2 and a high-frequency setting of 2. A total of $5 \times 5 = 25$ combinations of electro-acoustic settings are available for comparison in this example.

Assume that cell (2L, 2H) represents the initial estimate (I). Comparisons will be performed in the low- and high-frequency dimensions with the initial estimate as the vertex of the comparison. The direction of the first comparison can be arbitrary (i.e., cells with more or less gain can be chosen for comparison). In this example, we choose cells with less gain. In the low-frequency region, one compares (2L, 2H) with (3L, 2H), and in the high-frequency region one compares (2L, 2H) with (2L, 3H). Assume that the listener prefers less low-frequency gain and more...
high-frequency gain. That is, cell (3L, 2H) is preferred over cell (2L, 2H), and cell (2L, 2H) is preferred over cell (2L, 3H). The winning cells are indicated by a number sign (#).

The vertex of the next comparison will be formed by coordinates of the winning cells. Because “3L” and “2H” are the winning intervals in the low- and high-frequency regions, respectively, cell (3L, 2H) becomes the new vertex (II). The direction of the new comparison is with cells with even less low-frequency gain (i.e., 4L) and greater high-frequency gain (i.e., 1H). The comparison cells will be (4L, 2H) and (3L, 1H). Assuming that the listener selects cell (3L, 2H) over cell (4L, 2H) and cell (3L, 2H) over cell (3L, 1H), the new vertex is again formed at cell (3L, 2H). However, the direction of the next comparison is reversed to cells with more low-frequency gain and less high-frequency gain. This is recorded as the first reversal.

As illustrated in Figure 4–7, the second reversal is encountered after the third comparison when cell (3L, 2H) is compared with cell (2L, 2H) in the low-frequency region and cell (3L, 3H) in the high-frequency region. The third reversal is encountered after the fourth comparison when cell (3L, 2H) is compared with cell (4L, 2H) in the low-frequency region and cell (3L, 1H) in the high-frequency region. If the termination rule is set at three reversals, the new vertex formed at cell (3L, 2H) will become the final estimate or preferred frequency-gain response for the listener.

This example shows a case in which all the comparison settings are within the range of values available on the specific instrument. For listeners who may prefer a parametric value that is beyond the range available on the instrument, Levitt (1978) recommended that testing proceed with a value within the available range as if a reversal had occurred.

Issues that are important to consider in the simple up-down procedure are also important to consider in the modified simplex procedure. These issues include initial estimate, step size, and termination rules. An additional consideration in the modified simplex procedure is the optimal number of parameters to include in the comparison. Increasing the number of parameters to compare will increase the number of unique combinations of settings dramatically. In general, the number of combinations is given by \( N^m \), where \( N \) is the number of stimulus intervals within a dimension and \( m \) is the number of dimensions. For example, if three dimensions each with five intervals are compared, a total of \( 5^3 \) (125) settings will be available for comparison. Although not all combinations are compared, it is inevitable that more comparisons are necessary as the number of dimensions is increased. To keep the number of comparisons manageable, a rule of thumb is to compare only those parameters whose optimal settings one cannot easily predict.

Similar to the simple up-down procedure, listeners who do not show a strong preference for only one combination of settings (i.e., multiple preferences or same preference for all settings) may exhibit random preference judgments during paired comparisons and lead to errors in the estimation process. A sampling of the listener’s preference prior to using the modified simplex procedure may be helpful. Despite this potential limitation, Neuman et al (1987) and Kuk and Pape (1992) failed to find any of their subjects \((N = 8\) and \(N = 20\), respectively) who showed multiple preferences in their studies. In a subsequent study, Kuk and Lau (1996a) reported that six of their seven subjects showed a strong preference during their paired comparison judgments when the stimulus was presented at a positive SNR. However, only four subjects showed a strong preference when a negative SNR was used. This suggests the possibility that the stimulus conditions (and possibly criteria also) may affect the strength of the preference.

The modified simplex procedure yields a final estimate that agrees remarkably well with those selected with round-robin and double-elimination tournaments. Neuman et al (1987) showed that four of eight subjects participating in their study chose the same setting as the final estimate using the three procedures. The remaining subjects chose an
adjacent cell as the final setting. Furthermore, an average of 8.0 minutes was required to complete a modified simplex procedure, but 36.3 minutes was necessary to complete a double-elimination tournament, and 83.8 minutes was necessary to complete a round-robin tournament. Based on the results of these studies, the modified simplex procedure may be the most efficient of the three procedures.

**Advantages of the Paired Comparison Technique**

The use of paired comparison as a clinical tool to select hearing aid settings has been suggested since the late 1970s (Punch, 1978; Punch and Howard, 1978; Punch et al, 1980; Studebaker et al, 1978, 1979a,b, 1980; Punch et al, 1991). It is generally agreed that this method provides a viable alternative with which to verify and select frequency-gain characteristics of hearing aids. Some of the advantages of the paired comparison technique in relation to conventional methods are discussed in the following sections.

**Greater Sensitivity**

The most frequently cited advantage of paired comparison is its sensitivity over speech recognition tests in differentiating the improvements provided by settings of different amplification systems even when such differences are not apparent when using speech recognition tests (Zerlin, 1962; Witter and Goldstein, 1971; Punch, 1978; Punch and Howard, 1978; Studebaker et al, 1978, 1982; Tecca and Goldstein, 1984; Studebaker and Sherbecoe, 1988).

Studebaker et al (1980) compared the number of times a hearing aid was correctly selected in a paired comparison task with the absolute difference in word recognition scores between two comparison hearing aids. They reported that correct identification increased exponentially as the difference in mean word recognition scores between the two comparison hearing aids increased. For example, at a 0-dB SNR, subjects with a hearing loss required an 8% difference in word recognition scores between two comparison hearing aids to correctly select the one with better intelligibility in 75% of the comparisons. Normal-hearing subjects require only a 3% difference in word recognition scores to achieve the same level of selection.

Byrne and colleagues (Byrne, 1986; Murray and Byrne, 1986; Byrne et al, 1990) evaluated the sensitivity of paired comparison to differentiate frequency-gain responses that are more homogeneous in electroacoustic characteristics than those reported in earlier studies. Despite the homogeneity, these authors reported that 70 to 80% of the comparisons showed a significant preference for one frequency-gain response, whereas few significant differences were observed when using speech recognition testing.

Purdy and Pavlovic (1992) compared the frequency responses selected using paired comparison, magnitude estimation, and category rating tasks in elderly listeners. Their results showed that all three procedures are equally sensitive to differentiate small changes in frequency-response characteristics.

Preminger et al (2000) compared speech recognition scores and speech intelligibility ratings with frequency responses selected with a modified simplex approach and that selected by NAL-R. Two of the seven subjects reported significantly better speech understanding in the real world, whereas the rest of the subjects reported similar speech understanding with both methods of frequency response selection.

**Equal or Greater Reliability**

The reliability of the method of paired comparison, or the consistency with which subjects select a preferred hearing aid, is reportedly high. Zerlin (1962) showed that 7 of 11 subjects ranked the same hearing aid first in both test and retest trials when using a paired comparison procedure. Studebaker et al (1978) revealed correlations ($R$) in excess of 0.70 for paired comparison data obtained across different subject populations and in different lab-
oratory settings. For the same subjects, significantly lower correlations were obtained (between −0.40 and 0.70) on word recognition tests. This illustrates the relative consistency with which reliability data on paired comparison were obtained in different laboratories. Punch and Beck (1980) and Punch and Parker (1981) also showed higher reliability when using pairwise quality judgments in comparison with word recognition testing.

The reliability of paired comparison judgment may be affected by test conditions and choice of judgment criteria. For example, Punch (1978) found that subjects were most reliable when male speech was used as the stimulus, whereas music yielded the least reliable selection. Punch and Howard (1978) showed higher reliability when using clarity judgments of connected discourse presented in quiet than with intelligibility judgments of discourse presented in noise. Studebaker et al (1979b) demonstrated higher reliability in intelligibility rankings at a 0-dB SNR than at an SNR of +7 dB. Kuk and Pape (1992) reported similar findings.

Individuals with hearing loss may not be as reliable during paired comparison testing as normal hearing listeners. Studebaker et al (1980) compared the reliability of pairwise judgment of relative intelligibility and word recognition score [Northwestern University (NU)-6] between normal hearing and subjects with hearing loss. When paired comparison data were analyzed, 83% of normal hearing subjects and 54% of subjects with a hearing loss ranked the same hearing aid first in both test and retest sessions. When data from word recognition tests were analyzed, 42% of normal hearing subjects and 49% of subjects with a hearing loss ranked the same hearing aid first in both test and retest sessions. Schwartz et al (1979) also obtained similar findings. These data suggest that the reliability of paired comparison is as good as, if not better than, speech recognition.

Kuk and Pape (1992) evaluated the within-session and between-session reliability in which elderly hearing aid wearers (N = 20) selected their preferred frequency-gain response using pairwise clarity judgment of discourse passages as the criterion. An average of 80% of all subjects showed less than a 5-dB variation in their frequency-gain response selection upon retest. A similar consistency was seen when between-session and within-session data were examined. Stelmachowicz et al (1994), Eisenberg and Dirks (1995), and Eisenberg et al (1997) also showed similar reliability. This reflects minimal learning effects and suggests that this method can be expected to yield reliable results when clear instructions are provided to elicit judgments. On the other hand, Purdy and Pavlovic (1992) reported poorer reliability with paired comparison technique than the use of category rating.

Valid Predictor of Hearing Aid Performance

Two approaches to validate the results of paired comparison include correlation with speech recognition scores and real-world evaluation of the selected settings. Results of correlation between paired comparison judgments and speech recognition scores varied according to the criteria and test conditions used in the paired judgment. Punch and Howard (1978) reported low correlation (R = −0.46, to 0.34) between results of paired comparison of relative intelligibility and sentence scores on the Central Institute for the Deaf (CID) Sentence Test. On the other hand, Studebaker et al (1978) using the same criterion, found excellent correlation (R = 0.98) between results of paired comparison and scores on the speech perception in noise (SPIN) test.

Studebaker et al (1980) demonstrated the validity of paired comparison judgment of intelligibility by reporting that almost 73% of hearing aids chosen as the best during paired comparison judgment also received the highest mean speech recognition score. Neuman et al (1987) also showed that six of eight subjects obtained equal or higher individual speech recognition scores with hearing aids selected using paired comparison than those selected using an MCL approach (Pascoe, 1978).
On the other hand, pairwise judgment of speech quality may not yield a high correlation when evaluating intelligibility. For example, Punch and Parker (1981) reported negligible correlations between subjective judgments of speech quality and speech recognition scores despite significant correlation ($R = 0.70$) between relative intelligibility judgment and speech recognition scores. Despite the low correlation between speech quality judgment and measured speech intelligibility, Punch and Parker did not observe poorer measured speech intelligibility for hearing aids selected on the basis of quality. A later study (Punch and Beck, 1986) confirmed the low correlation between speech quality judgments and speech intelligibility scores. On the other hand, Studebaker and Sherbecoe (1988) found that hearing aids selected on the basis of pairwise quality judgment provided better measured speech intelligibility than hearing aids selected with magnitude estimation of speech intelligibility. This points to the potential difference between hearing aids selected with paired comparison and those with magnitude estimation.

Kuk and Lau (1996a) compared the preferred frequency gain response obtained through paired comparison and category rating. Although the preferred frequency response selected with paired comparison was always unique (i.e., one setting only) and was always highly rated, several subjects also showed similarly high ratings for adjacent frequency responses. This suggests that the results of paired comparison are similar to, and possibly more sensitive than, category rating. Eisenberg and her colleagues (1991, 1995, 1997) also showed high validity of the paired comparison technique compared to category rating.

Hearing aid settings that are selected with paired comparison may result in increased user satisfaction than those selected using a prescriptive method. Kuk and Pape (1993) evaluated listeners’ everyday satisfaction with hearing aids selected with the NAL-R formula and those selected with pairwise clarity judgment of discourse passages read by a male speaker and mixed in a babble noise (SNR = +5 dB). Listeners completed a questionnaire to indicate their satisfaction with the hearing aid in 22 listening situations. Subjects with a sloping hearing loss ($N = 10$) showed similar preference for hearing aids fit with either approach. Of the nine subjects who have a relatively flat hearing loss, seven showed significantly higher satisfaction for hearing aids selected with the pairwise approach, and only one showed higher satisfaction for a hearing aid selected with the NAL-R formula. Four of the seven subjects selected more low-frequency gain than NAL-R recommendation, whereas the remaining three subjects selected less low-frequency gain than NAL-R specification.

In a subsequent study, Kuk (1994) proposed a screening procedure for the use of the modified simplex. Subjects who were screened to prefer an alternate frequency response from the NAL-R were further engaged in the modified simplex to select their preferred frequency response in different noise backgrounds. When asked to complete a questionnaire on their real-world satisfaction of the selected frequency response, the majority of subjects rated the frequency response selected with the simplex procedure higher than that prescribed by the NAL-R. Preminger et al (2000) had a similar finding in some of their subjects.

The results of these studies show that hearing aids selected with the paired comparison technique are just as effective as, if not more effective than, those selected with a prescriptive method. This suggests that the use of paired comparison to verify hearing aid fitting may further enhance the fitting of hearing aids for some wearers.

**Ability to Judge Several Subjective Attributes**

The method of paired comparison has been used with different criteria. This includes “overall quality” (Jeffers, 1960; Witter and Goldstein, 1971; Punch, 1978; Harris and Goldstein, 1979; Punch and Beck, 1980; Sullivan et al, 1988; Leijon et al, 1991; Kuk et al, 1992), “intelligibility” (Zerlin, 1962; Stude-
baker et al., 1982; Sullivan et al., 1988; Byrne et al., 1990), and “pleasantness” or “naturalness” (Byrne, 1986; Murray and Byrne, 1986; Byrne and Cotton, 1988; Leijon et al., 1991). Other scales, for example “hollowness” (Kuk et al., 1992), “noise interference” (Kuk et al., 1990), and “amount of distortion,” although not used with paired comparison, could potentially be useful as criteria for paired comparison judgment. Kuk and Tyler (1990) demonstrated that listeners with hearing impairment could differentiate among various subjective criteria. These criteria could be useful in evaluating nonlinear and newer types of signal processing hearing aids.

Ability to be Performed Under More Listening Conditions

Byrne (1991) indicated that speech testing performed in quiet or in conditions of poor SNR (e.g., <−10 dB) does not help to differentiate amplification systems and restricts the test conditions for which speech recognition tests can be performed. He further indicated that paired comparison judgments are less susceptible to the ceiling effect than are speech recognition tests. Consequently, paired comparison may be used in more test conditions (e.g., different SNR, different types of noise backgrounds) than speech recognition tests to verify the appropriateness of settings selected for a programmable or digital hearing aid. Kuk et al. (1994) used paired comparison to measure the preferred frequency response at different SNRs as well as different overall levels to estimate the preferred long-term frequency response characteristics of hearing aids that allow automatic gain regulation.

Reduced Testing Time

The reliability of clinical speech recognition test is related to the number of items on the test. For example, Studebaker (1982) calculated that a 10% difference between two speech scores is considered significant at the 5% probability level if 135 test items are presented. At least 3381 test items must be presented if the same level of confidence is desired for only a 2% difference between test scores. Consequently, the time required to obtain a reliable result may be substantial.

Paired comparison procedures can be completed in substantially reduced time. Yet the results can be equally satisfactory, if not providing greater satisfaction than the settings that are selected using speech recognition tests. For example, an average of only 8 minutes is required to select the low- and high-frequency gain settings on a master hearing aid (Neuman et al., 1987). This makes paired comparison technique an ideal clinical tool with which to examine a large number of hearing aid settings under different listening conditions.

Minimal Involvement of Auditory Memory

In the method of paired comparison, acoustic stimuli processed by two hearing aid settings are presented sequentially with minimal time delay between presentations. This reduces any memory factor that may affect the sensitivity of the judgment. Studebaker (1982) suggested that judgment of small differences between stimuli is easier when performed in a comparative mode of minimal delay than in an isolated mode.

Simple Instructions and Easy Task

Listeners are instructed to choose from one of two stimulus intervals that meets the set criterion (e.g., better sound quality). A verbal or manual response (i.e., press one of two response buttons) is usually accepted. The task can be performed by listeners of all ages. For example, Eisenberg and Levitt (1991), and Eisenberg and Dirks (1995) reported that almost all hearing-impaired children are capable of performing paired comparison tasks by 6.5 years of age. The oldest subject (88 years old) in a study by Kuk and Pape (1992) reported no difficulty completing the paired comparison task.

Wide Applications

The method of paired comparison has been used to study sound quality perception of

**Individualized Fitting**

The method of paired comparison not only verifies if the selected settings are appropriate for the individual listener but also specifies new settings if alternative settings are preferred. The result of paired comparison procedures can lead to a more appropriate combination of settings tailored to the individual’s preference and psychophysical limitations. Furthermore, there is the psychological advantage to the listeners that they are actually involved in the hearing aid evaluation process. Unlike information provided by prescriptive formulas, final settings recommended by paired comparison procedures are not restricted to frequency-gain response only. Other electroacoustic parameters, such as compression settings (e.g., release time, compression ratios) or different types of signal-processing techniques may also be examined using this technique.

**Integrating Paired Comparison in Clinical Hearing Aid Fitting**

Although previous research has demonstrated that the technique of paired comparison is a powerful research tool, its clinical potential may not be realized unless it is implemented properly. This section discusses the considerations in implementing this technique in the clinical setting.

**Timing for Paired Comparison**

An important issue to consider is when and for whom paired comparison should be used. Although age of the subjects is not a major factor (Eisenberg and colleagues, 1991, 1995, 1997; Kuk and Pape, 1992), it is unclear if the amount of hearing aid experience of the subjects would affect the reliability of the judgments, especially if the judgments are made during the initial visit. Kuk et al (2002) showed that experienced linear hearing aid wearers with a severe-to-profound hearing loss did not experience the full benefit of a digital nonlinear hearing aid especially for low-level input sounds until after 1 month of use of the device. This raises the question of possible acclimatization effect as a confounding variable in the observed results. One may question if the results of paired comparison reflect the wearer’s experience with his or her previous hearing aid settings (for experienced wearers; unaided hearing for new wearers) and not a quest for the optimal setting?

To examine this issue, Kuk and Lau (1996b) correlated the preferred insertion gain on a programmable hearing aid obtained under six listening conditions (speech/noise levels of 55/50, 65/60, 75/70, 50/55, 60/65, 70/75 dBA) and one vocalization condition with the wearers’ used gain. The results showed that the amount of preferred insertion gain for listening was not correlated with the wearers’ used gain at any frequencies in any condition. On the other hand, the preferred insertion gain obtained during vocalization correlated significantly ($p < 0.05$) with the wearers’ used gain. The authors suggested that wearer insertion gain could affect preferred insertion gain only when the test conditions are identical to those that the hearing aid wearers experience in everyday life (i.e., own voice). In typical clinical situations, the preferred insertion gain for listening is likely determined by the stimulus characteristics and subjective
preference. The observation of acclimatization in Kuk et al (2002) originated from the extra gain for low-input sounds [50-dB sound pressure level (SPL)] available on the digital nonlinear hearing aids that were used during the study. Many of the soft sounds that were inaudible to the wearers became audible, and over time, became meaningful. The preferred frequency responses were determined at a suprathreshold level with a linear hearing aid (Kuk and Lau, 1996b). Stimulus audibility was never an issue, and thus hearing aid experience should not confound the results of paired comparisons. Berger and Hagberg (1982) and Leijon et al (1990) reported similar impressions.

Although wearer experience may not affect the result of paired comparison greatly (at least for frequency response adjustment), there are practical concerns to suggest delaying its use until later times. First, its high sensitivity to the stimulus condition suggests that the choice of the stimulus conditions would limit the outcome of the comparison. It will be impossible to select the optimal stimulus condition for every hearing aid wearer during the initial visit without knowing what listening conditions are encountered by the wearers, and which conditions are difficult for them. Second, although the use of prescriptive formulas has its limitations, if used properly such formulas can yield adequate fitting for a good portion of potential hearing aid wearers. Furthermore, even though the results of paired comparison may yield a different set of frequency response characteristics from the prescriptive targets, the difference in responses may not lead to functional difference in real life because of the wearers’ psychophysical abilities or listening environments. Thus, the time spent in paired comparison may not improve the fit of the hearing aid. Third, the time needed to carry out the procedure will necessarily limit the time that can be available for other activities during the initial visit, for example counseling, verification, and so on. For these reasons, it may be strategically desirable to delay the use of paired comparison to 2 to 3 weeks after the wearers have been fitted with their hearing aids. Frequency-gain responses that are selected with a prescriptive formula or recommended by the manufacturers should be used during the interim. This 2- to 3-week delay gives the wearers an opportunity to become acquainted with the amplified sound. In addition, it gives them an opportunity to identify problem areas with the use of the hearing aids, as well as the listening situations in which use of the selected frequency-gain response are less than satisfactory. Such information will form the basis for selecting stimulus materials to use during subsequent paired comparisons.

Although paired comparison can and should be delayed for subsequent visits, the recommended frequency-gain response should be verified as optimal for the wearer at least objectively during the initial fitting. This means that one must ensure that the chosen settings provide adequate gain for audibility of the softest speech sounds (around 20-dB HL across audiogram) and comfort for loud sounds (above 100-dB SPL). If real-ear target gain match is desired, make sure that the target gain formula has considered the effects of nonlinear signal processing, effect of multiple channels, release time, etc. (Kuk and Ludvigsen, 1999) in the gain formulation and that a composite signal be used as the stimulus. Furthermore, because the aided threshold represents the softest sound that a wearer can hear with his or her nonlinear hearing aid, such an index should be measured to determine if the goal of audibility for meaningful soft sounds is met (Kuk, 2001).

Thus the paired comparison technique should be used in the clinical setting during the follow-up visit as a fine-tuning (or troubleshooting) tool when the wearer has been given the opportunity to experience the recommended hearing aid settings for 2 to 3 weeks. Clinicians should start with the manufacturer’s recommended initial settings (because many are adjusted for device-specific modification to the prescriptive target) and only fine-tune (via paired comparison) the recommended settings when the wearer re-
turns after the trial period with specific complaints on the sound quality/performance of the hearing aid. To provide the best solution to the wearer, the following procedural steps are important to consider.

UNDERSTAND THE WEARERS’ COMPLAINTS

Many clinicians take the wearers’ complaints at face value when trying to solve them. For example, a wearer may say, “I can hear far but I cannot hear close.” Some clinicians will increase the gain parameter of the hearing aid that controls gain for low input (below 60-dB SPL), assuming sounds that are presented at close proximity will be at a low-input level. On the other hand, many wearers use the same phrase to describe their experience in a noisy restaurant where they can hear people from tables away, but not understand the person sitting across the table from them. In essence, this is a speech-in-noise problem and not an audibility problem. Adjusting the gain parameter for low input (below 60-dB SPL) would probably not solve the wearer’s complaint.

Thus, it is critically important to understand the wearer’s complaint thoroughly. Ask questions that may help clarify the wearer’s complaint. One needs to know “what” is wrong, “when” the problem occurs, “who” is involved, “where” the problem occurs, and “how” the problem occurs. Knowing the answers to these questions allows the clinician to determine if electroacoustic adjustment of the hearing aid settings is even necessary. Many times, the wearers’ complaints may have to do with restrictive situations, for example, “buzz” interference with security alarms, talking to an old patient in a noisy hospital, etc., where counseling and setting realistic expectations for the hearing aid would be more appropriate instead. If there is indeed a problem, knowing the answers to these questions will help the clinician understand the environments so that the appropriate test conditions for paired comparison can be set up and the appropriate electroacoustic parameters adjusted during fine-tuning.

CHOOSE THE RIGHT STIMULUS/TEST CONDITIONS

An advantage of paired comparison is the potentially infinite number of test conditions in which one can evaluate the selected hearing aid settings. In theory, one should compare hearing aid settings in listening conditions that are identified as difficult by the hearing aid wearers. The task for the clinician is to have replications (i.e., on cassette tape or compact disc) of these listening conditions and to evaluate the wearers using appropriate stimulus materials. This, however, may be clinically impractical because no two wearers’ listening environments are identical. Furthermore, there is the difficulty of physically re-creating the various listening environments in a clinical setting. A compromise may be to use standardized listening conditions for general purposes, but to have available a few representative sound effects that would represent the listening situations frequently encountered by the population of hearing aid wearers who are served in the community.

It is important to realize that the level of the stimulus used and the SNRs employed in the comparison could affect the appropriateness of the selected frequency-gain response. Specifically, our experience indicates that frequency-gain responses selected with female discourse passages presented at 62-dB SPL [root mean square (RMS), A-scale] in the presence of babble noise and at a favorable SNR (>+10 dB) are appropriate for typical listening situations. The same speech stimulus presented in quiet at a low input level of 50-dB SPL is appropriate for selecting optimal settings for soft sounds, whereas the same stimulus presented at 72-dB SPL and at a negative SNR may simulate loud, noisy situations. These test conditions can be marked for general use with paired comparison. Cox and Alexander (1991) and Pearsons et al (1977) also offered some speech and noise levels that can be used in different listening conditions. Sound effects that may be useful for selection of frequency-gain responses in specific listening situations in-
clude wind noise, music (orchestral and individual instruments), office noise (typewriter noise, computer fan noise), restaurant noise, and traffic noise. Additionally, one may perform paired comparison in an empty office or a large hall to verify or select frequency-gain response for listening in reverberant environments. The levels at which these sound effects may be presented need further investigation.

The criterion that one uses to verify or select a frequency-gain response is dependent on the needs of the hearing aid wearer for the frequency-gain response. If the intent is to maximize speech understanding ability, such criterion must be specified. If the intent is to search for a “natural”-sounding hearing aid, such criterion must be indicated. An earlier section summarizes some of the criteria that were used (and may be used) during paired comparison. It is important to remember that the criterion used will affect the outcome of the comparison (Byrne, 1986; Kuk, 1990). A criterion of “clarity” is recommended for general use because it is the criterion by which most hearing aid wearers judge their hearing aids (Hagerman and Gabrielsson, 1985).

**Determine the appropriate parameters to adjust**

The reason for asking wearers specific questions on their complaints is that one may gain insights on which electroacoustic parameters to adjust. Unlike linear hearing aids, which use the same gain for all input levels, nonlinear hearing aids employ different electroacoustic parameters to control gain at different input levels. For example, Kuk (2000) indicated that the compression threshold of wide dynamic range compression (WDRC) hearing aids influences gain for low-input sounds (below 60-dB SPL). Medium input sounds (around 60- to 75-dB SPL) are controlled by gain parameter called Gain_{medium}, or hearing threshold level (HTL). High-level inputs (above 85-dB SPL) are controlled by a parameter that may be labeled as Gain_{90}, Gain_{loud}, uncomfortable loudness (UCL), or saturation SPL (SSPL) depending on the manufacturer. Consequently, if one knows that the complaint is primarily occurring at a medium input level, one may focus on the Gain_{medium} parameter in the correct frequency band during the fine-tuning. It is not uncommon to have wearer complaints that may involve more than one electroacoustic parameter. The clinician must be thorough to include all the involved parameters.

**Determine the right form of paired comparison**

The type of information that one is interested to know influences the form of paired comparison task, that is, whether one should perform a single comparison, round-robin, elimination tournaments, or adaptive strategies. As indicated previously, if one is interested to know if one of several alternatives would provide better satisfaction to the wearer, one may perform a single comparison, round-robin, or elimination tournament. It is not necessary that these frequency-gain responses be systematically related to each other. Indeed, it may be worthwhile to choose alternative settings that are sufficiently different from the prescribed frequency-gain response to ensure maximum notable differences between comparison frequency-gain responses. Although these tasks could indicate if any of the alternatives is more preferred than the original setting, none of these could ensure that the chosen settings are the most optimal for the wearer, which would be dependent on the appropriate choice of comparison alternatives. On the other hand, if one were interested in the optimal settings for the wearer, an adaptive strategy, either a simple up-down procedure for unidimensional comparison or a modified simplex for multidimensional comparison, would be preferable.

**Implementation on commercial hearing aids**

Although the majority of current digital and programmable hearing aids have the capa-
bility to allow implementation of paired comparison tasks, many do not have structured algorithms that allow automatic stimulus presentation and response tracking. Instead, a second memory or multiple memories are usually available that the clinician can utilize to store the alternative settings. The clinician, however, has to manually track the responses and update the comparison settings. This limits the type of paired comparison procedures that can be performed in the clinic. It is foreseeable that more manufacturers of digital hearing aids will implement some of the mentioned procedures in their software in the near future.

Meanwhile, we illustrate how a hearing aid with two memories (permanent or temporary) can be used for paired comparison purposes. For example, the wearer complains that the original setting of the hearing aid results in other people’s voices being too “boomy.” This would suggest excessive gain for conversational input (60- to 70-dB SPL) in the low frequency. Only the Gain medium parameter in the low frequency would need to be adjusted. Male speech produced at a conversational input level (around 62-dB SPL) should be chosen as the stimulus. Because only one parameter is involved, a simple up-down procedure (Fig. 4–6) would be a good procedure to zero in on the optimal amount of low-frequency gain for clear and “non-boomy” perception. To proceed with the comparison, the clinician sets one memory (A) to the original setting of the hearing aid, and the other memory (B) to the alternate setting of the hearing aid. In this case, memory B should have 5 dB less gain in the low frequency (5 dB is the default step size) than memory A. The clinician instructs the wearer, turns on the test stimulus, and lets the wearer listen through settings A and B. After the wearer has decided on the preferred setting, the content of A and B will be updated so that one contains the preferred setting and the other contains the setting for the next comparisons. Comparison continues until the termination rule is met. The specific instructions and the example of how a potential run sequence may be done was described previously.

**NUMBER OF COMPARISONS FOR STATISTICAL SIGNIFICANCE**

I have been describing paired comparison as a single comparison event; that is, A and B are compared once to reach a decision. The assumption is that the decision is true and error-free. In reality, human responses are seldom error-free. Variability in response behaviors becomes more noticeable when the two comparison settings are very similar. Indeed, when the two settings are indiscriminable to the wearer, one would expect that both settings would have the same chance of being selected (i.e., 50%) given the nature of paired comparison tasks. This means that the selected setting would have a 50% chance of being the wrong choice!

Increasing the number of comparisons can improve the accuracy of the selection. By modeling the paired comparison task with the binomial theorem, Kuk and Lau (1995b) were able to predict the minimum number of comparisons and the number of times a particular setting must be chosen to reach a given level of statistical significance. Figure 4–8 summarizes this information. The y-axis reports on the level of significance and the x-axis reports on the number of chosen responses. The number of comparisons is the displayed parameter. If one draws a horizontal line at a particular probability (or level of significance), one can read off the number of times a particular setting should be selected for the chosen response to have that probability. The probability for up to 10 comparisons is provided. For example, if the line is drawn at $p = 0.3$, one can see that a subject has to choose the same response two out of two times, or three out of three times to ensure that the choice has less than 30% chance of being in error. However, if four comparisons are made, the same individual has to choose the same setting three out of four times to reach that probability level. If 10 comparisons are made, the same response
has to be selected seven times before one can say with 70% certainty that the chosen response is truly preferable to the wearer. As one can imagine, the more stringent the criterion (lower probability), the more times a particular response has to be chosen. Although most statistical analysis uses a 0.05 or 0.1 level of significance, we have found that a more relaxed criterion of $p = 0.3$ can produce relatively accurate and reliable selection for clinical purposes.

**COMPLAINTS INVOLVING MORE THAN ONE PARAMETER**

There may be two alternatives to solving complaints that require adjustment in more than one parameter. For example, a complaint of “unclear speech” may require lowering the gain in the low-frequency channel and increasing the gain in the mid-/high-frequency channels. With a commercial system that has only two memories, one approach is simply to compare the original setting with another setting that has decreased low-frequency gain and increased mid-frequency gain stored in the second memory. The result of the single comparison (performed the necessary number of times) specifies the final setting.

If one desires to perform adaptive comparison manually to optimize the final settings, one would need a commercial hearing aid that satisfies the following two requirements. The first is that the electroacoustic dimension(s) under comparison must have fixed intervals so that comparisons can proceed in discrete steps. Almost all commercial programmable and digital hearing aids satisfy this requirement. The second requirement is that the unit must have at least $(N + 1)$ temporary or permanent memories to allow paired comparison in $N$ dimensions, that is, two memories for one dimension, three memories for two dimensions, and so on.

One can illustrate how a modified simplex procedure can be manually implemented on a commercial digital hearing aid for selection of optimal high- and low-frequency settings. Because settings on two electroacoustic dimensions are compared (i.e., $N = 2$), the hearing aid must have at least three memories (i.e., $N + 1$) to facilitate comparison. A matrix representation of the combinations of settings similar to that illustrated in Figure 4–7 must be prepared prior to the comparison so that
one can manually track listener responses and update memory content accurately.

Assume that one follows the same sequence of comparisons shown in Figure 4–7. In the first round of comparison, one may store the frequency-gain setting represented by cell (2L, 2H) into memory no. 1. Frequency-gain responses represented by cells (3L, 2H) and (2L, 3H) can be stored in memories 2 and 3, respectively. Memory assignment is arbitrary. The same sequence of comparison illustrated in Figure 4–1 can be followed. It is extremely important to indicate on the matrix the winner of each pair of comparison so that settings for the next round of comparison can be easily determined. In the example illustrated in Figure 4–7, settings represented by cells (3L, 2H), (3L, 1H), and (4L, 2H) are stored in each memory of the hearing aid for the second round of comparison. In this way, memory update is only necessary after every round of comparison. Obviously, manual tracking of wearer response and the frequent update can be tedious and labor intensive. Thus, the use of a computer is extremely helpful and necessary when paired comparisons are made in more than one electroacoustic dimension.

Integration with Other Measures

The paired comparison technique is a relative measure. The result of the comparison does not guarantee that the particular programmable or digital hearing aid includes the best frequency-gain response or processing algorithm for the specific hearing aid wearer or if the hearing aid is appropriate for the wearer. Although this is a potential limitation of this technique, one can easily overcome it with the use of additional measures to ensure wearer satisfaction for the selected frequency-gain response. For example, the use of real-ear measures could ensure that the wearer is actually using some amplification from the hearing aid and not responding to “phantom attributes” during paired comparison. Frequently, direct questioning of wearer satisfaction with the selected or verified frequency-gain response will yield valuable information on the appropriateness of the selected device. Regular follow-up examinations and readjustment of settings (if necessary) may also enhance the wearer’s satisfaction with the hearing aid.

The technique of paired comparison is an important tool that can enhance the success of a hearing aid fitting. It should be regarded as part of a fitting battery. Integration of results on more than one measure is necessary to ensure maximum wearer satisfaction with the selected hearing aid.

References


CHAPTER 4  ■  PAIRED COMPARISONS AS A FINE-TUNING TOOL IN HEARING AID FITTINGS  149


Kuk F. Adaptation to enhanced dynamic range compression (EDRC)—examples from the SENSO P38 hearing aid. Semin Hear 2001;22(2):161–171.


