Loudness and Satisfaction Ratings for Hearing Aid Users
DOI: 10.3766/jaaa.20.4.7

Peter J. Blamey*
Lois F.A. Martin*

Abstract

Background: Hearing aids amplify low-intensity sounds to make them audible while keeping high-intensity sounds at an acceptable loudness for listeners with impaired hearing.

Purpose: The purpose of this analysis was to assess loudness and satisfaction at the same time using a combined loudness and satisfaction questionnaire to rate 18 everyday environmental sounds.

Research Design: Ten sets of data from four studies, covering three conditions, were analyzed. The three conditions were unaided, wide dynamic range compression (WDRC), and adaptive dynamic range optimization (ADRO). In total, there were 61 subjects giving over 3,000 pairs of ratings for loudness and satisfaction.

Results: The analysis found a strong relationship between loudness and satisfaction ratings for this set of listeners and conditions. The maximum satisfaction ratings corresponded to sounds with "comfortable" loudness ratings. Satisfaction was lowest for sounds that were "uncomfortably loud." Sounds that were very soft or inaudible also received low satisfaction ratings unless the sounds were expected to be soft, such as the sound of one's own breathing.

Conclusions: Hearing aid fittings that place most sounds at a comfortable level are likely to be more satisfactory than hearing aid fittings that produce more sounds close to hearing thresholds or discomfort levels. Aided conditions gave higher loudness and satisfaction ratings than the unaided condition, and the ADRO hearing aids gave significantly higher satisfaction ratings than the WDRC hearing aids.

Key Words: Comfort, hearing aid, loudness, satisfaction

Abbreviations: ADRO = adaptive dynamic range optimization; ESQ = Environmental Sounds Questionnaire; HTL = hearing threshold level; LDL = loudness discomfort level; NAL-NL1 = National Acoustic Laboratories—Nonlinear 1; PAL = Profile of Aided Loudness; WDRC = wide dynamic range compression

It is curious that most hearing aid fittings are based on measurements of what the listener can only just detect (hearing threshold levels [HTLs]) and/or what the listener finds uncomfortably loud (loudness discomfort levels [LDL]) instead of what the listener wants to hear. It is safe to assume that people want to listen to something in between HTL and LDL, but relatively little research has been directed at determining which part of this hearing range produces the most satisfactory hearing sensations, and most of the research that has been done is restricted to speech sounds in quiet and in noise. The main purpose of the analysis presented here is to investigate the loudness of nonspeech sounds heard through the hearing aid at loudness levels intermediate between threshold and discomfort, and the level of listener satisfaction with the perceived loudness of those sounds.

Examples of studies relevant to this topic include the work of the following researchers: Cox (1989), who developed a method of predicting comfortable loudness

*Dynamic Hearing Pty Ltd, Richmond, Victoria, Australia

Peter Blamey, Ph.D., Dynamic Hearing Pty Ltd, 2 Chapel Street, Richmond, Victoria, 3121, Australia; E-mail: pblamey@dynamichearing.com.au

Disclosure: The authors are inventors of the ADRO® amplification scheme and minority shareholders in Dynamic Hearing Pty Ltd, and both are former employees of Dynamic Hearing Pty Ltd. Dr. Blamey is a paid consultant to Dynamic Hearing Pty Ltd. Dr. Blamey is owner of Australia Hears Pty Ltd, a hearing aid company distributing ADRO and WDRC hearing aids in Australia.
levels across the frequency range from HTL measures and comfortable loudness data at 500 and 4000 Hz; Ricketts (1996), whose “data do not appear to support the use of additional clinical time to obtain individual loudness growth measures”; Munro and Patel (1998), who reported that “subjects did not express real-world auditory discomfort when the MPO (maximum power output) value matched the ULL [uncomfortable loudness level] value”; Elberling (1999), who concluded that “categorical loudness scaling cannot, in general, provide significant information for the fitting process”; Bentler and Cooley (2001), who found that the threshold of discomfort was highly variable and less than 22% of the variance could be explained by HTL; and Smeds (2004), who found that first-time hearing aid users preferred speech to be presented with less than normal loudness. After reviewing nearly 200 articles on the clinical effectiveness of using LDLs in the fitting of hearing aids, Mueller and Bentler (2005) found only three that met their criteria for evidence-based studies, and concluded that they were unable to make a strong recommendation concerning the clinical use of LDL measures. This same recent review also reports that many hearing aid users are dissatisfied with the loudness of their hearing aids. Kochkin (2002) found that 83% of hearing aid users feel they need to hear more soft sounds and 81% want increased comfort for loud sounds. There is thus a need for further research to shed light on how to achieve a hearing aid fitting with satisfactory loudness for the majority of hearing aid users.

The most comprehensive body of research on loudness perception and its application to hearing aids is the work of Moore and colleagues at Cambridge University. They developed a loudness model (Moore and Glasberg, 2004) and applied it to the fitting of hearing aids that use linear and nonlinear amplification (Moore and Glasberg, 1998; Moore et al 1999a, 1999b; Moore, 2000). Their theoretical loudness predictions can be combined with a variety of rationales to produce an initial hearing aid fitting. One rationale, called CAMEQ, aims to amplify speech to give equal loudness per critical band over the frequency range from 500 to 5000 Hz, and another, called CAMREST, aims to restore normal specific loudness patterns over a wide range of speech levels (Moore et al, 2001; Alcantara et al, 2004). Other hearing aid fitting protocols are also designed to produce near-normal loudness growth (Cornelisse et al, 1995; Cox, 1995; Gitles and Niquette, 1995) or to maximize the intelligibility of speech without exceeding normal loudness (Byrne et al, 2001). All of the fitting procedures cited in this paragraph predict suitable gains and output levels for individual hearing aid users from their hearing thresholds. These initial fittings do not take into account potential individual differences in loudness growth or individual preferences. Usually, the initial fitting will be “fine-tuned” to the individual’s preferences by the audiologist using the manufacturer’s fitting software, or by the hearing aid wearer using the volume control on the hearing aid. One of the reasons for fine-tuning is to increase the individual’s level of satisfaction with the loudness of sounds heard through the hearing aid beyond the level of satisfaction achieved by the initial fitting.

In order to evaluate the loudness of sounds in everyday situations, Palmer et al (1999) developed the Profile of Aided Loudness (PAL). Respondents were asked to rate the loudness of 12 everyday environmental sounds and give an overall satisfaction rating for the loudness of that sound. Normative data were obtained from 41 normally hearing subjects who rated the 12 sounds, which were subsequently grouped into three loudness levels: loud, average, and soft. By comparing aided responses for respondents with impaired hearing against the norms, it is possible to determine whether amplification provides a normalized loudness response where soft sounds are perceived as soft, and loud sounds are perceived as loud. In this way, the PAL can provide verification for a nonlinear hearing aid fitting using a loudness normalization rationale (Palmer et al, 1999; Mueller, 1999). The loudness and satisfaction rating scales and the procedure from the PAL provide operational definitions of loudness and loudness satisfaction in the present study (see Appendix 1).

The PAL has been extended to a set of 18 sounds using the same rating scales (see Appendix 1) to form the Environmental Sounds Questionnaire (ESQ). This extended questionnaire was used in four studies to investigate loudness and satisfaction with a new hearing aid amplification scheme, ADRO® (adaptive dynamic range optimization; Blamey, 2005). The speech intelligibility in quiet and in noise, and listener preferences for ADRO were also compared with other amplification schemes in these studies. The ADRO amplification scheme and fitting method are based on comfortable levels in contrast to most WDRC (wide dynamic range compression) fitting methods that are based entirely on thresholds and loudness discomfort levels. The comparison of loudness and satisfaction ratings for ADRO and WDRC is thus of particular theoretical and practical interest. The data for the ESQ collected in these four studies was collated and used for the present analysis. In all, there were five different hearing aid processing schemes (three using compression and two using ADRO) in addition to the unaided condition. The goals of the analysis were:

- to compare the loudness and satisfaction ratings for aided and unaided listening conditions;
to establish the relationship (if any) between loudness and satisfaction; and

to compare the loudness and satisfaction ratings of soft, average, and loud level sounds.

The specific hypotheses addressed were these:

- Environmental sounds will be rated as louder and more satisfactory in aided conditions relative to unaided conditions.
- Sounds perceived to be at a comfortable level will be more satisfactory than sounds that are either too soft or too loud.
- ADRO will produce greater loudness satisfaction than WDRC because its rationale is to make sounds comfortable.
- Ratings of loudness and satisfaction will be different for sounds with soft, average, and loud intensity levels.

### METHOD

The data for the present studies came from four investigations that are summarized in Table 1. The main goal of each study was to compare different hearing aid processing schemes in a group of subjects who had used each processing scheme for at least four weeks of acclimatization. The data for the ESQ were collected toward the end of the acclimatization period by listening to the sounds listed in Appendix 1 in real-life situations and filling in the loudness and satisfaction ratings. The ESQ forms were returned at the end of the acclimatization period. The important differences to note between the studies are that they were carried out with different hearing aid conditions, with different subjects, and by different investigators. One study (Mispagel and Valente, 2006) was carried out in the United States and the other three in Australia.

The first study (Study 1) was a clinical trial of the 64-channel ADRO amplification scheme in comparison with three-channel WDRC in a behind-the-ear hearing aid (Blamey et al, 2005). There were 19 participants with hearing losses ranging from mild to profound. Both amplification schemes were implemented in the same hearing aids, and the participants could switch between them at will during the six-week acclimatization period. The ESQ was filled out in the two aided conditions. The second study (Study 2) has been described in detail by Blamey et al (2004) and included 22 participants with mild-to-moderate hearing loss. Half the participants were fitted with an ADRO 64-channel in-the-ear hearing aid for four weeks experience followed by a nine-channel WDRC hearing aid for four weeks while for the remaining half of the subjects the conditions were reversed. During each take-home phase the subjects filled in the ESQ for both aided and unaided conditions.

The third study (Study 3) compared ADRO with four-channel WDRC for ten listeners with moderate-to-severe hearing loss. The study was conducted by Dynamic Hearing at the request of one of the large hearing aid manufacturers, using their top-end WDRC product. In this study the WDRC output levels were matched to the fitted ADRO output levels for speech at 55 dB SPL and 75 dB SPL (as far as the compression ratios available would allow). The volume control on the hearing aid was disabled. The participants could switch between ADRO and WDRC at will during the four-week acclimatization period, and the ESQ was used in the two aided conditions. The data for this study are unpublished. The fourth study (Study 4) investigated the effect of the number of channels on sound quality, speech intelligibility in quiet and in noise, and on acceptability (Mispagel and Valente, 2006). Ten subjects were fitted with a pair of ADRO 32-channel behind-the-ear hearing aids and a pair of ADRO 64-channel behind-the-ear hearing aids (in random order). During each six-week take-home phase, subjects filled out the ESQ for aided and
Hearing Instruments and Fitting Methods

The ADRO strategy is a multichannel nonlinear adaptive dynamic range optimization strategy (Blamey, 2005). It uses four statistical rules (fuzzy logic rules) to optimize the output level of the hearing aid independently in 32 or 64 narrow-frequency channels. The comfort rule ensures that the output level does not exceed the comfort target more than 10% of the time, and the audibility rule ensures that the output level does not fall below the audibility target more than 30% of the time. An important feature of the ADRO rationale is to align the comfort target to the listener’s perceptual level of comfortable loudness. The interactive fitting process used with ADRO allowed the listener to adjust this comfortable level according to their subjective preferences. This was done using loudness balancing of narrow-band noise at nine frequencies followed by an overall adjustment of comfort targets while listening to live speech through the hearing aid. This fitting process is described in detail in a conference paper (Blamey, 2006). The audibility target is typically set 20 dB below the comfort target for frequency regions with a wide dynamic range and less than 20 dB for frequency regions with a very narrow dynamic range of hearing. Sudden loud sounds are limited to a maximum output level in each channel by the hearing protection rule. Low-level environmental noise is kept at an acceptable soft output level by the background noise rule that imposes a maximum gain at each frequency. In such cases, the output may be below the audibility target or even below the threshold of the listener. Generally, the maximum gain is shaped to maximize sound quality and intelligibility for soft speech.

The WDRC hearing aid used in Study 1 was a three-channel device with 32 bands of frequency shaping adjustable at 11 audiological frequencies and fast time constants (10 msec attack and 80 msec release). The WDRC device used in Study 2 was a commercial nine-channel hearing aid with 15 bands of frequency shaping and used a WDRC strategy with low compression thresholds and compression ratios between 1:1 and 6:1. The hearing aid featured dual time constants (fast: 10 msec attack and 200 msec release times, slow: 200 msec attack and 800 msec release). Compression thresholds, compression ratios, maximum gains, and output limits for both the nine-channel and three-channel hearing aids were set using the NAL-NL1 (National Acoustic Laboratories—Nonlinear 1) nonlin-

ear prescription (Byrne et al, 2001). The WDRC hearing aid for Study 3 was also a commercial device with four compression channels and 16 frequency shaping bands. The WDRC time constants were quite slow, requiring 5 to 15 sec to achieve a 20 dB change in gain. The WDRC parameters for Study 3 were set to match as closely as possible the output levels for speech at 55 and 75 dB SPL from the ADRO fitting for each subject. The WDRC fitting software would not allow compression ratios higher than 3.3, and it was not possible to achieve a perfect match at both levels. WDRC was not used in Study 4.

In each study, the alternative amplification schemes were implemented on the same hardware platform so that differences in results could only be attributed to differences in the digital signal processing algorithms or the fittings. Studies 1 and 2 used no additional DSP algorithms such as noise reduction, feedback cancellation, or directional microphone. The WDRC device in Study 3 had an additional noise reduction algorithm that may have reduced the loudness of some sounds. Both ADRO devices in Study 4 used an adaptive directional microphone that may have affected the loudness and satisfaction ratings of some sounds depending on their input level and direction.

Procedure

The ESQ uses the loudness and satisfaction rating scales from the PAL (Palmer et al, 1999). Subjects were asked to rate the loudness of 18 different sounds ranging from inaudible to uncomfortably loud according to an eight-point scale (see Appendix 1 for details including instructions to participants). In addition, subjects were asked to rate the satisfaction with the loudness of the sound according a five-point scale ranging from “just right” to “not good at all.” A sound could be omitted from rating if the subject did not encounter that sound during the acclimatization period. Sounds 1, 3, 8, 10, 12, and 17 were classified as “loud”; sounds 2, 5, 6, 13, 15, and 16 were classified as “average”; and sounds 4, 7, 9, 11, 14, and 18 were classified as “soft” by people with normal hearing. These groups of sounds were treated separately in the analysis. These three groups are called “loudness levels” to avoid confusion with the eight “loudness rating” categories.

A copy of the ESQ was filled in for each condition by each participant in each study. The ESQ was provided toward the end of the acclimatization period of each study after a stable and acceptable fitting had been established for each participant in each condition. The participant was asked to fill in the ESQ based on their experience in the field (i.e., not with recorded sounds), and the ESQ was collected at the end of the acclimatization period.
Data Reduction and Analysis

Each condition of each study generated a matrix of responses (eight loudness rating categories × five satisfaction rating categories). Ideally, these matrices should be compared using nonparametric statistical tests such as the multifactor Chi-Squared test (Winer et al., 1991), but the results of this analysis are not intuitive and are difficult to display graphically. In total, the four studies yielded 3384 pairs of responses in 188 blocks: 72 blocks for ADRO, 52 for WDRC, and 64 in the unaided condition (see Table 1). The data were represented graphically by calculating mean ratings to illustrate the recurring patterns in the data. For analyses with more than one independent factor, analysis of variance (ANOVA) was used. The reader is reminded that the data are actually categorical and therefore do not meet the usual assumptions for the application of ANOVA. The results of the ANOVA analyses are thus only approximate. For very large numbers of observations and small numbers of factors, the approximation is a reasonable one. The mean rating values in the figures also need to be interpreted with caution. A higher mean rating indicates that the distribution of responses was skewed more toward the high end of the response scale.

To reduce the data and simplify the analysis, the data for the four studies have been combined for the statistical analyses. As a result of the different numbers of subjects and the different conditions tested in each study, it was not possible to perform balanced ANOVA, and a Generalized Linear Model (GLM) was used. Each study is presented separately at first, so the reader can judge the similarities and differences between the studies, and then combined to provide a “big-picture” view of the main features of the data.

RESULTS

Figure 1 shows the mean loudness rating for each loudness level (on the horizontal axis) for each condition (see legend in each box) for each study. The scale on the vertical axis is the same for each pane of the figure. Each point is the average of about 100 ratings, so the standard error of the mean (shown by the error bars) is quite small. The data for each condition show an increase in mean loudness rating for increasing loudness level. The unaided conditions for Study 2 and 4 produce a fairly straight line on the graphs, although the line is lower on the graph for Study 4 than Study 2. Most of the lines for the aided conditions show a slight curvature. This probably indicates that perceived loudness is not “normal” for these conditions. The loudness ratings for the aided conditions are higher than the ratings for the unaided conditions, showing that the hearing aids were indeed amplifying the sounds for the listeners. The differences between the data for the different studies are probably due to the different hearing losses of the participants, as summarized in Table 1. It was unexpected that the largest differences in perceived loudness between WDRC and ADRO occurred for Study 3, which is the one where an attempt was made to match the output levels of ADRO and WDRC for running speech at 55 and 75 dBA SPL. This resulted in a good match for sounds at the soft level, but WDRC was louder for the average and loud levels. It is likely that this outcome was caused by a restriction in the WDRC hearing aid manufacturer’s fitting software that did not allow compression ratios higher than three.

Figure 2 shows the data combined for all four studies. ANOVA showed significant main effects for loudness levels (F = 608.3; df = 2, 3120; p < 0.001) and conditions (F = 421.9; df = 2, 3120; p < 0.001). There was no significant interaction effect between loudness level and condition (F = 0.84; df = 4, 3120; p = 0.501). Bonferroni pairwise comparisons of all data points in Figure 2 showed there was no significant difference between the loudness ratings for the ADRO and WDRC conditions at any loudness level. The differences between pairs of data points at different loudness levels and/or aided/unaided pairs were all significantly different with p < 0.001.

The ESQ was derived from the PAL by adding new sounds to each of the sound level categories. The effect of adding these new sounds was assessed using ANOVA to compare the mean loudness and satisfaction ratings for new and existing sounds at each loudness level in the aided listening conditions combined. No significant difference between new and existing sounds was found for either loudness or satisfaction ratings, but there was a significant interaction between loudness level and new/existing sounds for loudness ratings (p < 0.001). The new sounds were rated 0.35 higher in loudness at the soft level (t = 3.12, p < 0.05) and 0.35 lower in loudness at the loud level compared with the original sounds in the PAL. Thus the inclusion of the new sounds in the ESQ may have had a small effect on the average loudness ratings at soft and loud levels in Figures 1 and 2 but are unlikely to have affected the satisfaction ratings in Figures 3 and 4.

Figure 3 shows the mean satisfaction rating for each loudness level for each condition for each study. The data for each aided condition show a decrease in mean satisfaction rating for increasing loudness level. The unaired condition for Study 2 shows a flatter curve, and the unaired condition for Study 4 shows a slight increase. The mean satisfaction ratings for the aided conditions tend to be higher than the ratings for the unaired condition, except for loud sounds, where they are about equal. The mean satisfaction ratings for
ADRO in Studies 1 to 3 tended to be a little higher than those for WDRC. The differences between the data for the different studies are probably due to the different hearing losses of the participants, as summarized in Table 1.

Figure 4 shows the satisfaction rating data combined for all four studies. Analysis of variance showed significant main effects for loudness levels ($F = 58.1; df = 2, 3118; p < 0.001$) and conditions ($F = 48.3; df = 2, 3118; p < 0.001$). Bonferroni pairwise comparisons of the factors showed that there were significant differences between the satisfaction ratings for the ADRO, WDRC, and the unaided condition for the full data set ($p < 0.001$). For individual points on the curves, the Bonferroni procedure showed no significant differences between the mean satisfaction ratings for the three loudness levels in the unaided condition, or between the three conditions at the loud level. ADRO at the soft level had a significantly higher mean satisfaction rating than any other combination of condition and loudness level ($p < 0.001$) except for WDRC at the soft level. The three points on the ADRO curve were significantly different from one another, as were the three points on the WDRC curve.

The final stage of analysis separated the data into groups according to the loudness level of the input sound as rated by listeners with normal hearing (“soft,” “average,” and “loud”) to see whether the relationship between loudness ratings and satisfaction ratings was the same for each level of sounds (Fig. 5). It is clear from Figure 5 that the relationship is somewhat different for “soft” sounds that receive significantly higher satisfaction ratings than the “average” and “loud” sounds when their loudness rating is low. There is little difference between the data for “average” and “loud” sounds. All three groups

![Figure 1](image1.png)

**Figure 1.** Mean loudness ratings for each condition of each study. The sounds are grouped by loudness level. The error bars show ± one standard error of the mean.

![Figure 2](image2.png)

**Figure 2.** Mean loudness ratings for each loudness level for the three conditions, ADRO, WDRC, and unaided. The data from all four studies have been combined in this figure. The error bars show the standard error of each mean value.
of sounds achieved maximum satisfaction ratings for the “comfortable” loudness rating category. Analysis of variance indicated significant main effects of loudness rating category ($F = 68.7; \text{df} = 7, 3103; p < 0.001$) and loudness level ($F = 13.8; \text{df} = 2, 3103; p < 0.001$). The interaction between these factors was also significant ($F = 3.33; \text{df} = 14, 3103; p < 0.001$). Bonferroni pairwise comparisons of the loudness rating categories indicated that the satisfaction ratings for the “comfortable” category were significantly higher than for all other loudness categories ($p < 0.001$). Bonferroni pairwise comparisons of the loudness levels indicated that “soft” level sounds received significantly higher satisfaction ratings than “average” and “loud” level sounds ($p < 0.001$), and there was no significant difference in the mean satisfaction ratings for “average” and “loud” level sounds.

**DISCUSSION**

The data in Figures 1 and 2 show that the hearing aids were performing their primary function of amplifying sounds, that is, making them louder than in the unaided condition. One of the goals of the PAL was to verify that hearing aid fittings were successful in achieving loudness normalization if this was part of their amplification rationale (Mueller, 1999; Palmer et al., 1999). From this point of view, one could consider Figures 1 and 2 to be a sort of perceptual input/output function using loudness instead of sound pressure level for the units on the input (horizontal axis) and output (vertical axis). Palmer et al. indicate that normal-hearing listeners rated “soft” sounds on the input loudness level as “very soft” or “soft”; “average” level sounds were rated as “comfortable but slightly soft,” “comfortable,” or “comfortable but slightly loud”; and “loud” level sounds were rated as “loud but okay” or “uncomfortably loud.” On this basis, they suggested that normal loudness perception would correspond to

![Figure 3](image-url) Mean satisfaction ratings for each condition of each study. The sounds are grouped by loudness level. The error bars show ± one standard error of the mean.

![Figure 4](image-url) Mean satisfaction rating for each loudness level for the three conditions. The data from all four studies have been combined in this figure. The error bars show the standard error of each mean value.
mean loudness ratings of 2 for the “soft” sounds, 4 for “average” sounds, and 6 for “loud” sounds. These values would give a much steeper perceptual input/output function than any of those shown in Figures 1 and 2. This is probably not very surprising for the aided conditions because neither ADRO nor WDRC fitted with NAL-NL1 is designed to give normal loudness perception. The shallower than normal slope in the unaided condition is quite surprising because the literature suggests that listeners with impaired hearing have steeper loudness growth than listeners with normal hearing. Another way of looking at this is that in the unaided condition, the participants rated “loud” level sounds as “comfortable” on average, while “soft” level sounds were still rated as “soft.” After amplification, these “soft” level sounds were rated within the “comfortable” range, that is, the hearing impaired listeners perceived them to be louder than normal. After amplification, “loud” sounds were perceived to be slightly softer than normal, with a mean loudness rating below 6.

Part of the shallow slope is accounted for by the inclusion of the new sounds in the ESQ (see paragraph three of the results section); however, this is not sufficient to account for the full effect. We speculate that “soft” level sounds are still rated as “soft” by people with impaired hearing because the data were collected in the course of the ESQ respondents’ normal experience where sound sources can occur at a wide range of distances from far away to near the listener. In the field, listeners will hear each of the sounds in the ESQ at input levels ranging from their hearing threshold (at relatively far distances) up to the actual level of the sound when they are close to the source. This results in a sampling bias that is different for normal-hearing people and people with impaired hearing in the unaided condition: the normal-hearing people can hear (and rate) sounds from further away. These distant sound sources will never be rated by hearing-impaired listeners because they do not hear them, and so the range of input levels rated by the ESQ is more limited and biased to higher input levels. This bias is more important for “soft” sounds than for “average” or “loud” level sounds, resulting in a shallower-than-normal loudness growth function for environmental sounds heard in real environments.

To find out whether the study participants thought these differences from normal loudness were a good thing or a bad thing, it is necessary to look at the satisfaction ratings. For “soft” and “average” level sounds, the amplification resulted in an increase in mean satisfaction ratings, but not for sounds at the “loud” level (see Figs. 3 and 4). This implies that it is a good thing for “soft” and “average” sounds, but not necessarily for “loud” sounds, to be louder than normal. It is worth noting that the loudness rating “loud but okay” does correspond to the satisfaction rating “okay” as would be expected by the use of the word “okay” in both ratings. Figure 4 also shows higher mean satisfaction ratings for ADRO than for WDRC at each loudness level, although the mean loudness ratings were very close for these two conditions in Figure 2.

Having explored the data a little, it is now time to address the specific goal in this analysis: to determine whether there is a relationship between loudness and satisfaction ratings. Figure 5 demonstrates the relationship between loudness and satisfaction ratings on a single graph by showing the mean satisfaction rating for all sounds that received a particular loudness rating for each of the three loudness levels. Because the ESQ was used in the field, the participants would have been exposed to the sounds with a wide range of input levels at the microphone, depending not only on the sound pressure level of the source but also the distance between the sound source and the listener. Thus the loudness ratings for each level of sound were spread over the entire range of responses from “did not hear” to “uncomfortably loud.” Obviously there would have been very few sounds at the “soft” level that were rated as “uncomfortably loud,” and few at the “loud” level rated as “did not hear,” but there were enough to produce reasonably reliable mean satisfaction ratings. Figure 5 shows that if a sound is rated as “comfortable,” it is likely to receive a higher satisfaction rating than if it is rated as either louder or softer. This relationship suggests that a hearing aid that presents sounds with “comfortable” loudness will provide maximum satisfaction for the listener. There are some obvious exceptions to this observation: Anecdotal reports and studies of noise reduction indicate that listeners prefer some sounds (such as traffic noise for example) to be softer and some low-level environmental sounds (such as computer fans) to be inaudible.
The idea that making all sounds “comfortable” will provide maximum satisfaction is counterintuitive but plausible if comfort is a major requirement for a hearing aid to be considered satisfactory. In a recent review of 45 studies on hearing aid satisfaction, Wong et al (2003) found that comfort was one of many factors, but this appeared to refer to physical comfort of the hearing aid in the ear of the wearer rather than comfortable loudness. Loudness was not referred to as a major factor affecting satisfaction with the hearing aid, although Kochkin (2002) has consistently found that audibility for soft sounds and comfort for loud sounds are in the top five requirements for hearing aid users. The high satisfaction ratings achieved with FM microphones and other assistive listening devices that are designed to produce a comfortable output signal for speech at a distance in noisy environments is also an indication of the important relationship between comfortable loudness and satisfaction.

The lines in Figure 5 confirm that maximum loudness satisfaction is provided by presentation of average and loud level sounds at a “comfortable” level regardless of the level of the input sound. There is a significant difference between “soft” level sounds and the rest. These sounds maintained a reasonably high satisfaction rating even when the loudness rating was “very soft” or even “do not hear,” indicating that the participants liked to hear these sounds but were not particularly dissatisfied if they did not hear them or if the sounds were not very loud. By contrast, the participants were not satisfied if the “average” and “loud” level sounds were perceived as “soft” or “very soft.” This is evidence that satisfaction depends on the type of sound as well as its perceived loudness.

It is important to note that loudness satisfaction for environmental sounds may not equate to overall satisfaction with hearing aids or even with overall satisfaction with the sound produced by hearing aids. Overall satisfaction is likely to depend on other factors such as sound quality, intelligibility of speech, and naturalness of the hearing aid user’s own voice.

The relationships shown in Figure 5 help to explain the strong preferences that have been found for ADRO hearing aids (Blamey et al, 2005) and cochlear implant processors (James et al, 2002; Dawson et al, 2004) when compared to alternative processing schemes such as WDRC. The ADRO processor is designed to place average and loud sounds relatively close to the “comfortable” loudness, where satisfaction is highest, whereas other schemes tend to produce a wider range of output levels than ADRO. Figures 2 and 4 together show that loudness is not the only factor contributing to the strong preferences for ADRO in hearing aids, because ADRO produced consistently higher satisfaction ratings (Fig. 4) without consistent loudness differences from WDRC (Fig. 2).

CONCLUSIONS AND IMPLICATIONS FOR HEARING AID DESIGN AND FITTING

This analysis found a strong relationship between satisfaction ratings and perceived loudness of environmental sounds. It was found that satisfaction ratings also depended on the input level of the sound and the amplification condition.

The data imply that “average” and “loud” environmental sounds should be presented close to the “comfortable” loudness level to maximize loudness satisfaction for environmental sounds. They should not be too loud or too soft. “Soft” sounds should not be amplified above “comfortable” loudness, but it is acceptable for them to be perceived as “soft” or “very soft.” This recommendation is clearly different from other hearing aid fitting rationales such as loudness normalization.

Maximizing loudness satisfaction may be different from maximizing overall satisfaction with the hearing aid. Overall satisfaction with hearing aids depends on many factors, and the relative importance of the loudness of environmental sounds has not been assessed. Most hearing aids are optimized for listening to speech, and it is not known whether the relationship between loudness and satisfaction is the same for speech sounds as it is for environmental sounds. It is hoped that the approach used in this analysis will be helpful in addressing these important questions.

Acknowledgments. The authors would like to thank Karen Mispagel and Michael Valente for providing access to raw data from their study of ADRO and commenting on an early version of this manuscript. We would also like to thank the participants in the studies and the research audiologists who collected the data in the other studies: Karyn Galvin, Kamilla Krauze, Hayley Fiket, and Jutta Hau. The data collection protocols were approved by the Human Research and Ethics Committee of the Royal Victorian Eye and Ear Hospital under Projects 02/458H and 03/592 and the Human Studies Committee of Washington University Medical Center. ADRO® is the registered trademark of Dynamic Hearing.

REFERENCES


Appendix 1. Instructions and Items for the Environmental Sounds Questionnaire

During this week, please listen to each of the following sounds with each of your hearing aid programs. Please enter your responses to indicate the loudness of the sound and your satisfaction with that loudness level for each program. For rating the loudness of the sound use the following loudness scale:

7 = uncomfortably loud
6 = loud but okay
5 = comfortable but slightly loud
4 = comfortable
3 = comfortable but slightly soft
2 = soft
1 = very soft
0 = do not hear
x = don’t know, e.g., did not encounter that sound

For rating your satisfaction with the loudness level use the following satisfaction scale:

5 = just right
4 = pretty good
3 = okay
2 = not too good
1 = not good at all

For example, you might rate a particular sound as “very soft.” If “very soft” is your preferred level for this sound, then you would rate your loudness satisfaction as “just right.” If, on the other hand, you think the sound should be louder than “very soft,” then your loudness satisfaction rating might be “not too good” or “not good at all.” The loudness satisfaction rating is not related to how pleasing or easy to hear the sound is but, rather, how satisfied you are with the loudness level perceived.

1. Dog barking close by.
2. Travelling in a car with the windows closed.
3. Traffic noise when standing on the curb of a busy road.
4. Your own breathing.
5. Washing machine.
6. Running water, such as a toilet or shower.
7. Car indicator signal.
8. A motorbike passing by.
9. Chewing soft food.
10. Vacuum cleaner.
12. Door slamming.
13. Telephone ringing close by.
15. Microwave oven beeping.
16. Hair dryer or electric shaver.
17. Lawn mower.