Real-Ear Measurement of Hearing Threshold and Loudness

¹Larry E. Humes, Ph.D., ²Chas Pavlovic, Ph.D., ²Victor Bray, Ph.D.
and ²Michele Barr, M.A.

¹Department of Speech and Hearing Sciences, Indiana University, Bloomington, IN 47405, and

²ReSound Corporation, Redwood City, CA 94063

For at least the past sixty years, the predominant type of electronic circuit present in hearing aids has been a single-channel amplifier designed to be linear over most of its operating range with its maximum output limited via peak clipping (Berger, 1984). Although this is presently still the case (Hawkins and Naidoo, 1993), it is apparent that hearing aids designed to be *nonlinear* over most of their operating range are increasingly encountered and are now offered by a wide range of manufacturers (Berkey et al, 1992; HIA, 1995). Moreover, many of these contemporary devices are two- or three-channel instruments with programmable control of the nonlinear characteristics available for each channel. Research into the physiological and perceptual nature of the sensorineural hearing loss experienced by the vast majority of hearing-aid wearers (see review by Van Tasell, 1993) and advancements in engineering (see Killion, 1993) have largely led to this trend toward nonlinear hearing aids. Briefly, it has been established that the normal inner ear provides a compression of the input sound intensity that appears to be dependent on the presence of normally functioning outer hair cells. Listeners with sensorineural hearing loss attributed to underlying cochlear pathology typically lack normally functioning outer hair cells and the corresponding compression of input sound intensity. Nonlinear hearing aids have been designed to restore this normal compression of sound intensity prior to delivering the amplified sound to the inner ear of the listener with sensorineural hearing loss.

The ways in which hearing aids are fit and evaluated have also undergone considerable evolution over the past sixty years. This evolution, primarily concerned with the fitting of linear hearing aids, has been reviewed recently elsewhere (Humes,

1996). The most commonly accepted approach to fitting linear hearing aids involves the use of some prescriptive formula, based either on thresholds or suprathreshold loudness measurements [see Humes (1991), McCandless (1994), and Humes and Halling (1994) for reviews], to generate target real-ear gain. Once targets are generated, the hearing aid is fit to the patient and adjusted to achieve a match to the targeted gain values *for a moderate level input signal*. For contemporary nonlinear hearing aids, unlike their linear predecessors, the gain at low and high input levels is not designed to be the same as that at the moderate input levels used to generate and confirm the targeted values. For nonlinear devices, the gain is generally less as the input increases over the range from 50 to 100 dB SPL, although the manner in which it decreases with increases in input level varies with the particular circuit incorporated into the hearing aid. Moreover, the desired level-dependent gain characteristics may vary across channels for multi-channel hearing aids.

Rationale for Prescriptive Formulae

The increasing popularity of nonlinear hearing aids has led many to reconsider the ways in which hearing aids are fit and evaluated. As a result, interest has been rekindled in the use of real-ear measurement of threshold and loudness as one of the available methods. The authors, and others, have argued previously, however, that such approaches are probably desirable for *all* hearing aids, linear and nonlinear alike (Skinner et al, 1982; Kiesling, 1987; Skinner, 1988; Cox and Alexander, 1991; Humes and Houghton, 1992; Kruger and Kruger, 1994). To see why such an approach might be preferable even for linear hear-

ing aids, it may be informative to review the rationale and assumptions of one of the more popular prescriptive approaches, POGO (Prescription of Gain and Output, see McCandless and Lyregaard, 1983; Schwartz et al, 1988). As far as the prescribed gain characteristics are concerned, POGO attempts to amplify a moderate level speech input signal (65 dB SPL) so as to be comfortably loud at all frequencies (or slightly below that at 250 and 500 Hz). The simple prescriptive formula used in this approach was derived by examining data from large numbers of subjects on the relation between threshold and most comfortable loudness at several frequencies.

Under the assumption that average speech is at most comfortable loudness in listeners with normal hearing, then knowledge of the dependence of most comfortable loudness on hearing threshold in listeners with impaired hearing can be used to predict the desired amount of real-ear gain needed for this same average speech level. This predictive relation between threshold and most comfortable loudness, however, will only work well for "the average" hearing-impaired listener with a specified hearing threshold. Thus, two hearing-impaired listeners, each having a hearing threshold of 50 dB HL at 1000 Hz, may have most comfortable loudness levels that are elevated 15 and 35 dB (on average, 25 dB) above the normal value and, according to POGO, would both require 25 dB of real-ear gain at this frequency. The actual real-ear gain required in these two cases, however, would be 15 and 35 dB so that achieving a good match to the POGO target would underamplify by 10 dB for one hearing-aid wearer and overamplify by the same amount for the other (assuming a goal of restoring amplified speech to a most comfortable loudness at each frequency). This example illustrates a problem common to all threshold-based prescriptive procedures that have as their goal the restoration of some loudness-based criterion. Whenever this is the case, POGO and NAL (Byrne and Tonisson, 1976; Byrne and Dillon, 1986) probably being two of the most commonly used examples (Martin and Morris, 1989), an average relation between threshold and some loudness criterion is assumed and few listeners with hearing impairment may prove to be truly "average." By way of analogy, it could be established that most 6'-tall men have a shoe size of 10, but it would clearly be a mistake to attempt to fit *every* 6' male customer at the shoe store with a size-10 shoe.

Several studies have documented the considerable variability that exists in loudness judgments,

such as MCL and LDL, for listeners having the same degree of hearing loss (e.g., Kamm et al, 1978; Hawkins et al, 1987; Valente et al, 1994). The results of these studies imply that it would be difficult to accurately predict loudness judgments from thresholds. This difficulty in prediction, however, is not due to poor reliability of LDLs or other loudness-scaling judgments (e.g., Robinson and Gatehouse, 1996), but to true individual differences in loudness perception among listeners with similar hearing loss.

Direct Measures of the Dynamic Range of Hearing

Rather than attempt to estimate the desired shoe size from the customers height, one could measure the size of the foot directly and select the appropriate shoe size from such measurements. This, of course, is the way most shoes are sold in this country. Analogously, one could measure the most comfortable loudness (if that is the desired goal of the fitting strategy) directly, rather than attempting to estimate it from threshold. Indeed, several such loudness-based approaches have been developed (Skinner et al 1982, 1988; Cox, 1983, 1985, 1988; Humes and Halling, 1994). Regardless of the specific loudness-based criterion pursued by a particular method, a fundamental premise underlying these approaches is that the desired criterion can not be reliably predicted indirectly and must be measured directly. If multiple loudness criteria, such as "soft," "comfortably loud" and "uncomfortably loud," are to be used to establish gain for low, moderate and high input levels with nonlinear hearing aids, then *each* of these must also be measured directly. In general, as the number of loudness criteria incorporated into the fitting strategy increases, the time required for direct measurement of these criteria also increases. The additional time required to perform the loudness measurements has probably been one of the biggest obstacles to the widespread use of direct, loudness-based fitting approaches.

The direct measurement of the desired loudness criteria eliminates one of the problems with indirect threshold-based methods: the poor predictability of loudness criteria from the hearing thresholds. Direct measurements of loudness, however, do not eliminate *all* of the problems. Of course, one must have a reliable, valid and efficient way of measuring the loudness criteria to be incorporated into the desired fitting approach.

Although this is not a trivial issue, it is one that has been addressed adequately (e.g., Allen et al, 1990).

Perhaps a more evasive issue involves the errors introduced by the use of a variety of standardized couplers in various aspects of the hearing-aid fitting and evaluation process. Under the assumption that the most appropriate place to confirm the functional benefits of the hearing aid is in the wearer's ear while the hearing aid is being worn, then we are again confronted with the need to consult average relational data, this time on the relation between sound pressure levels measured in the average ear and those measured in a wide assortment of couplers, including the NBS-9A 6-cm³ coupler, the HA-1 or HA-2 2-cm³ coupler, and the occluded-ear simulator (Zwislocki coupler).

This is not to say that the issue is one simply of having so many different coupler-to-real ear relations to consider. Rather, even one such relation is plagued with the same problems regarding variability of individual subjects around the "average" values built into the relation. If, for example, one chooses to use insert earphones calibrated in an occluded-ear simulator (Zwislocki coupler) to make threshold and loudness measurements and measures the gain characteristics of the hearing aid in a testbox using the same occluded ear-simulator, problems still exist in translating the testbox measurements to the actual real-ear sound pressure level produced by the hearing aid *in the wearer's earcanal* when worn in the sound field. Such translation problems exist because of individual differences in occluded-ear volumes between the insert earphone and the hearing aid, in variations in microphone location within and across hearing-aid types, in the geometry and resulting acoustics of the earcanal, concha and other outer-ear cavities, in the diffraction effects associated with the presence of the head and body in the sound field during the real-ear measurements, and in variations in coupling (vents, horns, leaks, etc.) between the testbox measurements and the sound-field measures. Although extensive catalogs of average correction factors could be developed, and have been already to some extent (e.g., Bentler and Pavlovic, 1989; Dillon, 1991), these corrections are again only applicable to the "average" hearing-aid wearer with physical characteristics identical to those for whom the relations were derived.

From the foregoing it should be clear that it is problematic at best to attempt to combine the assortment of measurements required to fit and evaluate a hearing aid on the patient, even using

direct and multiple measures of loudness, because of the dependence on average relations between the real ear and a particular standardized coupling system and the restricted set of conditions under which the standardized coupler measurements are performed. This, of course, assumes that one or more of the various behavioral and acoustic measurements performed with the patient or on the hearing aid are defined outside of the patient's earcanal. That is, this situation becomes problematic because we wish to evaluate hearing-aid performance with the actual hearing aid on the patient's ear in the sound field under a representative set of acoustic conditions, yet have a variety of behavioral measurements needed in the evaluation, including threshold and loudness criteria, defined as sound pressure levels generated in one or more standardized couplers.

Real-Ear Measures

With the advent of real-ear probe-microphone measurements in the early 1980s, clinicians have been able to reliably measure the real-ear unaided and aided responses (REUR and REAR) for specified acoustic inputs in the sound field. The primary focus with these measurements, however, has been in the derivation of real-ear insertion responses (REIR, $REIR = REAR - REUR$) to evaluate the match between observed and prescribed insertion gain. Since the primary objective for which probe-microphone systems had been developed was the measurement of insertion gain for linear hearing aids at moderate input levels, it was adequate to design systems that permitted such measurements for input levels as low as 60-70 dB SPL. This: (1) permitted use of input levels representative of those encountered in everyday, conversational speech; (2) approximated those used in standard coupler evaluations of the hearing aid; and (3) was well above the noise floor existing in the front end of the measurement systems.

It is clear, however, that probe-microphone systems are capable of measuring much lower levels in the ear canal as evidenced by their more recent application to the clinical measurement of otoacoustic emissions (Probst et al, 1991). It is now possible, therefore, to measure the entire dynamic range of hearing, from threshold to uncomfortable loudness, with the corresponding levels referenced to sound pressure levels *in the ear canal of the hearing-aid wearer*. Consequently, it is also now possible to reference the sound levels corresponding to these various behavioral mea-

tures, *together with the output of the patient's hearing aid*, to the same point of measurement: the earcanal of the hearing-aid wearer. The reliability and validity of these acoustic measurements, moreover, has been established in the contexts of both hearing-aid real-ear-measurement and otoacoustic-emission measurement (e.g., Humes et al, 1988; Dirks and Kincaid, 1987; Probst et al, 1991; Dirks et al, 1996).

It has been emphasized previously (Humes, 1988; Humes and Houghton, 1992) that such an approach involving direct, real-ear confirmation of the theoretical objective of the fitting approach adopted for use by the audiologist should be preferred even for linear instruments. Returning to our previous example using POGO and its objective of amplifying speech to MCL, there is now no reason that this objective can not be confirmed directly on the patient's ear while wearing the hearing aid. That is, rather than attempting to predict most comfortable loudness from threshold and then translating this objective to an insertion-gain target, one can now measure the most comfortable loudness level in terms of earcanal SPL and then present a speech-shaped test signal to the patient, while the hearing aid is being worn, to verify that the aided speech spectrum follows the most-comfortable loudness contour.

The theoretical objective of POGO (amplifying conversational speech to most comfortable loudness) had previously been translated into an insertion-gain target for the patient because that's what could be conveniently measured by existing equipment in the patient's earcanal. The advent of improved probe-microphone technology, together with the microphone's interface, permits a wide range of sound pressure levels to be measured in the patient's earcanal. In addition, low-cost, high-quality "sound cards," developed for widely-used multi-media applications with personal computers, are generally available. These trends, together with the continued development of sophisticated software for such applications and the programmability and electroacoustic flexibility of many contemporary hearing aids, all make it possible to bypass this previously necessary insertion-gain "middle man" and move straight ahead to direct real-ear confirmation of the theoretical objectives of a particular approach.

It should be noted further that, although the focus here has been placed on loudness-based theoretical objectives and measurements, the same applies to fitting rationales not directly based on loudness measurements. The Desired Sensation

Level method (DSL; Seewald, 1992, 1995), for example, has as its objective the amplification of the full 30-dB dynamic range of speech so as to place this range at or above threshold from 250 through 6000 Hz while making sure that the peaks of speech (top of the 30-dB dynamic range) are not amplified beyond an uncomfortable level. The DSL procedure, designed for application to pediatric populations, is simply based on threshold measurements and the application of a number of correction factors to convert all behavioral and hypothetical speech spectra to the same earcanal reference. The problems with such corrections have already been noted in the discussion of loudness-based rationales and also apply here, as has been acknowledged by Seewald (1992). In principle, there is no reason this approach cannot be implemented using real-ear measurement of the sound pressure levels corresponding to threshold and the dynamic range of speech signals, ranging from soft to loud speech, to directly confirm the approach's theoretical objectives in the ear of the hearing-aid wearer. More recent versions of the DSL approach (Seewald, 1995), have moved much closer to such an implementation. The required behavioral measurements, however, are often difficult or impossible to obtain from the pediatric populations for which the DSL approach was designed.

An assumption throughout much of this discussion has been that the measurements required of the theoretical objective can, in fact, be performed by the patient. Aside from the increased testing time needed for additional loudness measurements required in many methods noted previously, limited patient capability is another reason audiologists frequently default to threshold-based rationales. If the audiologist is unable to obtain loudness measures from the patient, due either to time limitations or other constraints, then direct confirmation of a loudness-based theoretical objective is not feasible. In some cases, the only information available on which to base a fitting is hearing threshold and this may itself be only an estimate derived from auditory brainstem responses, otoacoustic emissions, or binaural sound-field responsiveness to warble tones.

For most adult patients, and many children and adolescents, however, this is not the case. Valid and reliable measurements of various loudness measures, and certainly of threshold, are rarely impossible to obtain. This being the case, it is usually possible to obtain direct, real-ear, confirmation of most hearing-aid fitting rationales on most potential hearing-aid wearers.

In the remainder of this paper, the authors would like to review the development and evaluation of a system designed to accomplish this objective. Although other similar systems either have been or are being developed by a variety of manufacturers, our focus is placed on the Real-Ear Loudness Mapping (RELM) system being developed by ReSound Corporation. The loudness-based objective incorporated into this method, as well as the techniques for the loudness measurements, have undergone considerable research and development with which we are familiar.

Real-Ear Loudness Mapping (RELM)

Introduction

The RELM system has been under development and evaluation at ReSound Corporation over the past two years. In reality, though, its development has been evolving over a much longer period in that it borrows heavily from two previous hearing-aid fitting and evaluation systems: (1) the Loudness Growth of Octave Bands (LGOB) protocol (Allen et al, 1990) incorporated as a means of fitting of ReSound hearing aids since 1988; and (2) the CHASE system created in 1988 (Humes and Houghton, 1992). Essentially, the CHASE system was a real-ear based system that was designed to perform a comprehensive set of tasks involved in the fitting and evaluation of linear and nonlinear hearing aids, including insertion-gain measurement, speech-recognition testing, automated administration and scoring of subjective hearing-aid performance surveys and questionnaires, and real-ear loudness measurements. The real-ear loudness measurements, however, were not obtained using a rigorous procedure of established reliability and had not undergone sufficient clinical evaluation to ensure their widespread use clinically. The CHASE system remained a tool for clinical research used at a handful of sites prior to being acquired by ReSound Corporation in 1994. In the intervening two years, the focus has been placed on merging and enhancing the real-ear aspect of the CHASE loudness-measurement system with the now, long-established LGOB loudness-measurement procedure.

Loudness Growth in Octave Bands (LGOB)

The LGOB method (Allen et al, 1990) has proven to be a reliable and valid means of measuring the growth of loudness in listeners with sensorineural hearing loss efficiently (Pluinage,

1989). As noted previously, the LGOB protocol has been incorporated as an option for the clinical fitting and evaluation of ReSound hearing aids since their appearance on the market in 1988. In ReSound's clinical version of the LGOB system, the stimuli have been presented via insert earphones calibrated in reference to either a 2-cm³ or a Zwislocki coupler. The stimuli presented via these insert earphones are 1/2 octave bands of noise centered at octave intervals from 500 through 4000 Hz. Following presentation of three successive bursts of a particular noise stimulus, the patient responds by rating the loudness as either TOO LOUD, VERY LOUD, LOUD, COMFORTABLE, SOFT, VERY SOFT or DID NOT HEAR using a custom button pad referred to as the Personal Selector (Pluinage, 1989). Task instructions incorporate definitions for each of the loudness categories (Allen et al, 1990). For example, for "very soft" it is indicated that "you would ask someone talking this loud to speak up" while for the "very loud" category "you would ask someone speaking this loud not to shout." Several practice trials are presented prior to actual data collection. These practice trials serve a two-fold purpose of familiarizing the listener with the stimuli and task and establishing the lower and upper limits of the listener's dynamic range at each frequency. The latter purpose serves to reduce the time spent during the testing phase in the presentation of stimuli that are either inaudible or beyond uncomfortable loudness. During the training phase of the LGOB procedure, loudness growth is measured sequentially in each frequency region such that all intensities are presented randomly at a particular frequency prior to advancing to the next frequency. In the test phase of the LGOB procedure, stimulus frequency and intensity are both presented in random order.

RELM Equipment

To date, the RELM system has been a prototype consisting of a desktop 80486-50 MHz, IBM-compatible personal computer with a 16-bit audio board installed, a probe-microphone with preamplifier assembly, and an amplified loudspeaker. The arrangement of this equipment for the testing conducted to date with the prototype is illustrated in Figure 1. Note that the loudspeaker (Acoustic Research, AR570) is at an azimuth of 45 degrees from the test ear and 18 inches from the center of the subject's head. The probe-microphone system used in this prototype is the Etymotic Research

Figure 1. Schematic illustration of the arrangement of the equipment used in the RELM prototype system.

ER-7C. It is important to note that this system is just a prototype and that the final hardware configuration is likely to differ from that depicted in Figure 1.

After the system is calibrated at the location of the listener's head without the listener present (method of substitution), the RELM system proceeds to the LGOB measurements with the probe-tube microphone positioned in the earcanal of the test ear. The nontest ear is occluded with a deeply seated foam earplug. At present, the LGOB procedure requires approximately 15 minutes per ear for completion, including both the practice and testing phases at all frequencies.

Reliability of RELM Measures

One of the first questions to be addressed with the RELM system was the reliability and validity of the loudness measurements. To address these issues, loudness measurements were obtained in a group of 35 normal-hearing adults (14 males and 21 females) ranging in age from 21 to 50 yrs (Mean = 29.8 yrs.). Normal hearing was defined as air-conduction pure-tone thresholds no greater than 10 dB HL (ANSI, 1989) from 250 through 6000 Hz and a normal tympanogram. Thresholds were measured for twenty of the subjects and screened at 10 dB HL for the remainder. Subjects were tested with the RELM system and then retested by the same examiner after a period of at least two, but not more than eight, days.

Figure 2 provides the mean loudness contours obtained for both the test (filled symbols) and re-

test (unfilled symbols) sessions from 35 normal-hearing subjects. This figure plots the earcanal sound pressure levels corresponding to the loudness ratings indicated to the left of each contour. All sound pressure levels plotted are 1/3-octave-band levels calculated from Fast Fourier Transforms of the probe-mic measurements. At the bottom of this figure, the mean thresholds for the test and retest sessions have been entered for the 20 subjects for whom threshold measurements were obtained (unconnected, unfilled squares and triangles). These values are not measurements performed by the RELM system, but represent air-conduction thresholds measured using standard clinical procedures and a clinical audiometer prior to the RELM measurements in each session. Within the RELM system, the examiner enters air-conduction thresholds in dB HL re: ANSI (1989) from a previously obtained audiogram. These values are then converted to equivalent free-field values (Bentler and Pavlovic, 1989). Next, the patient's own sound-field-to-earcanal transfer function, measured in RELM, is used to estimate the patient's earcanal SPL at threshold. This process has the advantage of not requiring additional time to remeasure thresholds already established during the hearing evaluation while simultaneously incorporating direct measurement of the most idiosyncratic portion of the coupler-to-earcanal transfer function. As can be seen in Figure 2, the estimated thresholds for these twenty subjects (unconnected open squares and triangles) are generally 5-10 dB above the DID NOT HEAR ratings which were directly measured by the RELM system. (Although the means for the DID NOT HEAR contour were calculated from the data for all 35 subjects, corresponding means calculated from just the 20 subjects for whom thresholds were measured were within 1-2 dB of the means for the entire group at all frequencies.)

The reliability of these real-ear loudness measurements was evaluated in several ways. First, it is apparent from visual inspection of Figure 2 that there is reasonable agreement between the test and retest measurements at all frequencies and for all loudness categories. Second, this was confirmed statistically by computing the test-retest differences and performing several repeated-measures analyses of variance (ANOVAs). A series of seven ANOVAs were performed, one for each loudness category, to determine if the test-retest differences were significantly different from zero and if the test-retest differences varied significantly with frequency. The results of these ANOVAs are

Figure 2. Mean test (filled symbols) and retest (unfilled symbols) data for the RELM loudness ratings from 35 normal-hearing adults. Mean threshold values for test (triangles) and retest (squares), measured with clinical procedures and equipment for 20 of the 35 subjects, are shown as unfilled, unconnected symbols near the bottom of the figure.

summarized in Table 1. Note that only two significant ($p < .01$) effects are observed: test-retest differences were significantly different from zero for the SOFT loudness rating and the test-retest differences varied significantly with frequency for the VERY SOFT category. It should be noted, however, that the test-retest differences for the DID NOT HEAR and COMFORTABLE loudness categories would also have differed significantly from zero had a significance criterion of $p < .05$ been used.

Another approach to assessing the reliability of these measurements is to compute test-retest correlation coefficients. Table 2 provides the Pearson- r correlations between the test and retest results from RELM for the eight test frequencies and the seven loudness categories. All but six of the 56 test-retest correlations are at least of moderate strength and statistically significant at either the $p < .01$ or $p < .05$ level. Most of the exceptions occur for the test frequency of 500 Hz and the loudness rating of VERY LOUD. It is unclear as to why 500 Hz would be less reliable, although the dB SPL values for this frequency tended to vary over a narrower range for most loudness ratings in comparison to the other frequencies. (In general, correlations decrease in size as the range of values being correlated decreases.) The VERY LOUD category may be slightly more variable, on the other hand, because the listeners have difficulty making consistent distinctions between LOUD, VERY LOUD and TOO LOUD. Given that the range of dB SPL values across subjects was narrow (approximately 25 dB) for these 35 normal-hearing subjects for each loudness category and frequency, it is impressive that about 90% of the test-retest correlations were of at least moderate strength and statistically significant.

Finally, another way of examining the variability of these measurements across the test and retest sessions is to examine the standard deviations and distribution of the signed test-retest differences. These values are depicted for each of the eight frequencies of the RELM system in Figure 3. The values depicted are averages across the seven loudness ratings at each frequency. The unfilled bars provide the standard deviations of the test-retest differences which provide an indication

Table 1. Summary of results of repeated-measures ANOVAs on the signed test-retest differences for the RELM loudness measurements

Loudness Rating	Differences $<$ or $>$ 0?	Effect of Frequency?
DID NOT HEAR	NO, $F(1, 34) = 5.53$	NO, $F(7,238) = 1.45$
VERY SOFT	NO, $F(1, 34) = 0.03$	YES, $F(7,238) = 3.23^*$
SOFT	YES, $F(1, 34) = 7.79^*$	NO, $F(7,238) = 0.65$
COMFORTABLE	NO, $F(1, 34) = 4.70$	NO, $F(7,238) = 0.36$
LOUD	NO, $F(1, 34) = 3.03$	NO, $F(7,238) = 1.05$
VERY LOUD	NO, $F(1, 34) = 0.00$	NO, $F(7,238) = 0.44$
TOO LOUD	NO, $F(1, 34) = 0.33$	NO, $F(7,238) = 1.00$

* $p < .01$

Table 2. Test-retest Pearson-r correlations for the RELM loudness rating (N = 35)

Loudness Rating	Frequency (kHz)							
	0.25	0.5	1.0	1.5	2.0	3.0	4.0	6.0
DID NOT HEAR	0.55	0.25	<i>0.39</i>	0.49	0.55	<i>0.37</i>	0.32	0.69
VERY SOFT	0.85	0.19	0.58	0.66	0.64	0.77	0.54	0.57
SOFT	0.77	0.72	0.56	0.78	0.72	0.75	0.56	0.72
COMFORTABLE	0.65	0.52	0.57	0.77	0.75	0.70	0.45	0.64
LOUD	0.61	0.62	0.61	0.80	0.83	0.74	0.51	0.73
VERY LOUD	<i>0.42</i>	0.17	0.23	<i>0.35</i>	0.23	0.47	0.47	0.44
TOO LOUD	0.47	<i>0.37</i>	0.44	0.67	0.68	0.63	<i>0.39</i>	0.69

bold = $p < .01$; *italics* = $p < .05$

of the average *magnitude* of the test-retest differences regardless of the direction of the difference. That is, these standard deviations indicate that the average size of the test-test difference was approximately 4-5 dB at and below 3000 Hz and about 6-7 dB at and above 4000 Hz. The mean differences between test and retest are, in fact, much smaller than this (<1-2 dB) as indicated previously in Figure 2. For clinical applications with individual listeners, however, the group mean differences are not as important as the average size of the difference, regardless of direction. For example, if one makes ten measurements at test, the first five having a value of 70 dB and the last five having a value of 80 dB and then, on retest, observes the opposite pattern in which the first five measurements are 80 dB and the last five are 70 dB, then the test-retest differences in this case will be five values of -10 dB and five values of +10 dB. In computing the arithmetic mean of these differences, the positive and negative values cancel each other resulting in a mean test-retest difference of 0 dB. This obviously does not reflect the fact that the magnitude of the test-retest difference was always 10 dB in this example, half the time test being greater than retest and half the time retest being greater than test. The standard deviation of these test-retest differences, however, is 10 dB which accurately reflects the magnitude of the test-retest differences, regardless of their direction.

Figure 3 also indicates the dB difference between the 5th and 95th percentiles for each test frequency, averaged across the seven loudness categories (striped bars). These values indicate that 90% of the subjects had test-retest differ-

ences less than or equal to about + or - 9 dB at and below 3000 Hz and less than or equal to about + or - 12 dB at 4000 and 6000 Hz.

In summary, considering the results of the ANOVAs on the test-retest differences, the test-retest correlations, and an average magnitude for the test-retest differences of about 5 dB across loudness categories and frequencies, the reliability of the loudness ratings in the RELM system is very acceptable for clinical use. This variability is considered acceptable in that it is not unlike that observed for other measurements used in audiology that are either acoustic measurements (such as real-ear gain) or behavioral measurements (such as MCL or LDL). The variability reported here for the RELM measurements, moreover, represents the combined variability inherent in

Figure 3. Mean standard deviations (unfilled bars) and 5th-95th percentile ranges (striped bars) for the signed test-retest differences for the RELM loudness ratings.

the acoustic measurements of real-ear sound pressure level and in the behavioral measurement of loudness.

Validity of RELM Measures

Measurements of any kind, however, can be reliable or repeatable without necessarily being valid. Assessing the validity of these measurements is difficult in that generally accepted or standardized loudness contours are not available for loudness categorization. Most classic loudness contours, for example, are based on the phon or sone loudness scales and are not expressed in dB SPL values referenced to the ear canal of the listener [see review by Scharf (1978)]. To assess the validity of the RELM loudness contours, the RELM contours were compared to the equal-loudness contours of Robinson and Dadson (1956) that form the basis of an ISO standard on loudness contours (ISO/R 131-1959E).

To do so, however, required several adjustments or corrections to both the RELM contours and the standard Robinson and Dadson contours. First, because the Robinson and Dadson contours were obtained in a free field at 0-degree azimuth, the RELM contours underwent two corrections: (1) the mean ear canal response for the 35 subjects was subtracted from the RELM loudness contours to correct them to sound-field SPLs (vs. ear canal SPLs) at 45-degrees azimuth; (2) a 3-dB adjustment in level was applied to every data point, except that at 2000 Hz, to correct for the difference in azimuth (0 vs. 45 degrees) between the two sets of data (Blauert, 1983). Once the RELM contours were corrected by these two adjustments, the dB SPL value for each RELM loudness contour at 1000 Hz was noted and the corresponding phon equal-loudness contour from Robinson and Dadson was consulted. For example, the adjusted RELM COMFORTABLE loudness contour had a sound-field SPL of 73 dB at 1000 Hz and the 73-phon equal-loudness contour from Robinson and Dadson became the reference contour for these data. The Robinson and Dadson equal-loudness contours, however, were obtained for binaural pure tones, rather than monaural 1/2-octave bands of noise as in the RELM procedure. As a result, two adjustments were applied to the corresponding Robinson and Dadson equal-loudness contours, one to account for the increased loudness of the broader bandwidth stimuli used in RELM and the other for the decreased loudness associated with monaural versus binaural stimulus presentation. From data summarized in Scharf

(1978) and Humes and Jesteadt (1991), the following two corrections were applied to the Robinson and Dadson reference contours: (1) increase of 10 dB to account for binaural summation of loudness for moderately loud to loud stimuli (> 40 phons); and (2) decrease of 4 dB for the use of pure tones instead of 1/2-octave bands to measure the loudness. This correction, a net correction of 6 dB, was applied to all of the Robinson and Dadson reference contours, except the 4-phon contour (threshold), and at all frequencies. Corrections were not made for bandwidth or binaural summation at the lowest (threshold) contour because these effects are negligible at these low levels.

Figure 4 provides a comparison of RELM contours to Robinson and Dadson (1956) loudness contours, both adjusted as described in the preceding paragraph. The symbols and dotted lines depict the RELM loudness contours and the solid lines represent the corresponding equal-loudness contours from Robinson and Dadson (1956) in

Figure 4. Comparison of adjusted RELM loudness contours (symbols connected by dotted lines) to adjusted equal-loudness contours in phons from Robinson and Dadson (1956). See text for a description of the various adjustments required to enable this comparison.

phons. Despite the number of corrections required to make this comparison, the agreement between the two sets of data is remarkable. In general, the differences between the RELM contours and the corresponding equal-loudness contour in phons is less than 3 dB, except at 1000 Hz and 6000 Hz where it is somewhat higher (about 5 dB at 1000 Hz and 8 dB at 6000 Hz). In summary, the RELM system appears to provide both reliable and valid measures of loudness across frequency and throughout the listeners dynamic range.

The reader may find it odd that a number of correction factors were applied to the data to assess its validity when the authors had argued previously in this paper against the use of such correction factors. The correction factors, however, are appropriate to use for mean data from groups of subjects as has been done in the preceding section on the validation of RELM measurements. For validation, one is only interested in establishing that one set of average data measured in one fashion agrees with another set of average data measured in some other fashion. We are not proposing that these same average corrections could be applied to the individual data of each subject to derive valid loudness contours from the RELM measurements (or vice versa).

Aided and Unaided Speech-Noise Measurements in the RELM System

Once the RELM system has been used to measure the loudness mapping across frequency and intensity with reference to the listener's earcanal, the next step is to make acoustic measurements of a speech-spectrum noise in the earcanal of the listener with the probe-tube microphone still in place and using the same equipment used to perform the loudness measurements. This is a purely acoustic measurement and requires no participation on the part of the listener. Speech-spectrum shaped noise is digitally generated with reference to levels measured in the sound field during the initial calibration of the RELM system. The 1/3-octave levels of the speech-shaped noise are derived from those in the pending ANSI standard, S3.79, on the Speech Intelligibility Index (formerly Articulation Index). The RELM system presents three speech-noise signals in sequence with the level of each noise user-selectable from an available range of 50-85 dB (in 5-dB steps). In the evaluation of this prototype RELM system, levels of 50, 65 and 80 dB SPL were selected. The levels in each 1/3-octave band from 200 through 6300 Hz are measured using the same FFT-based

spectral analysis methods used in the calibration and loudness-mapping procedures.

The results of the speech-noise measurements from the same 35 normal-hearing adults are provided in Figure 5 for both the test (filled symbols) and retest (unfilled symbols) sessions. It is immediately apparent that there is excellent agreement between the test and retest measurements at 65 and 80 dB SPL, but not at 50 dB SPL. The discrepancy at 50 dB between test and retest, however, is artifactual in nature. During the initial test session the sequence of speech-noise levels was 50, 65 and 80 dB, in that order. These measurements are completed in a matter of seconds and it was clear that the first measurement often "caught the listener by surprise," frequently while they were moving or talking. The 50-dB measurement is obviously the most likely one to be affected by extraneous low-level noise. For the retest measurements, it was decided to use the same three speech-noise levels, but in opposite order (80, 65, then 50 dB). This served the purpose of alerting the subject to the presentation of the noises with the highest intensity noise in the stimulus set, the one least likely to be affected by low-level extraneous noise. Immediate test-retest measurements were performed with this new sequence (80-65-50 dB) during the second test session with 15 of the 35 subjects and were found to be in excellent agreement at all three presentation levels (equivalent to the negligible test-retest differences seen

Figure 5. Mean test (filled symbols) and retest (unfilled symbols) 1/3-octave levels for the speech noise presented at levels of 50, 65 and 80 dB SPL.

in Figure 5 at 65 and 80 dB). Because the retest values at 50 dB in Figure 5 are felt to be more valid than the test values, given the change in presentation sequence, the retest values at each level are considered “the norms” for this portion of the RELM system. This also illustrates the “vulnerability” of the acoustic measurements at low input levels (50 dB) to background noise in the prototype system and may dictate changes in the design of the final system to permit reliable and valid use of low-level inputs.

Repeated-measures ANOVAs were performed separately for each of the three speech-noise levels. As might be expected from visual inspection of Figure 5, the signed test-retest differences did not differ significantly from zero at 65 dB [$F(1,34) = 3.8$, $p > .05$] or 80 dB [$F(1,34) = 1.5$, $p > .05$], but did at 50-dB [$F(1,34) = 30.5$, $p < .01$]. Similarly, there was no effect of frequency on the test-retest differences at 65 dB [$F(15,510) = 1.63$, $p > .05$] or 80 dB [$F(15,510) = 0.97$, $p > .05$], but there was at the 50-dB speech-noise level [$F(15,510) = 11.81$, $p < .01$]. Given the change in the sequence of presentation levels between test and retest, as described above, the significant effects observed at 50 dB are expected. A more accurate reflection of the test-retest variability can be obtained by focusing on the data obtained for the 65 and 80 dB speech-noise levels.

Figure 6 provides the mean standard deviations (unfilled bars) and the mean dB range (striped bars) encompassing the 5th through 95th percentiles for the signed test-retest differences for the 65- and 80-dB speech-noise levels (50-dB level omitted). As noted previously in discussion of comparable data for the loudness measurements from RELM (Figure 3), the mean standard deviation provides an indication of the average magnitude of the test-retest differences, regardless of sign. This appears to be approximately 2-3 dB, although it approaches 4 dB in the higher frequencies. The striped bars indicate that 90% of the subjects had test-retest difference less than + or - 5 dB through 3000 Hz and less than + or - 8.5 dB at 4000 and 6000 Hz for the 65-dB and 80-dB speech-noise levels. Comparison of Figures 3 and 6 indicates that, not surprisingly, these acoustic measurements are even more reliable than the results of the loudness-mapping measurements. For the real-ear loudness-mapping procedure, one has the same variability associated with the acoustic real-ear measurements *plus* the additional variability associated with the behavioral task of judging and categorizing loudness.

Figure 6. Mean standard deviations (unfilled bars) and 5th–95th percentile ranges (striped bars) for the signed test-retest differences for the RELM speech-noise measurements averaged across the 65 and 80 dB presentation levels.

Regarding the validity of these measurements, Figure 7 provides a comparison of the 1/3-octave band levels measured with the RELM system (unfilled symbols) and the targeted values from the pending ANSI SII standard for the 50-, 65-, and 80-dB speech-noise levels. The RELM values are simply the measured earcanal SPLs shown previously in Figure 5 minus the mean earcanal re-

Figure 7. Comparison of RELM speech-noise 1/3-octave levels (unfilled symbols) to the targeted 1/3-octave levels (solid lines) from the pending ANSI standard on the Speech Intelligibility Index (SII). All dB values are in sound-field SPL.

sponse of the 35 subjects. In general, targeted and measured speech-noise levels are within 2-3 dB at all frequencies and for all three speech-noise levels. The exceptions to this are the 1/3-octave levels at 4000, 5000 and 6300 Hz for the 50-dB speech-noise. The SPLs measured for these three data points are not valid measures of the speech noise, but represent the measurement of the noise floor in this prototype system. The RELM system performs noise-floor measurements during the acoustic measurements to allow the clinician to discern valid from invalid speech-noise measurements. The mean noise floor values at 4000, 5000 and 6300 Hz are within a couple dB of the speech-noise levels for the 50-dB speech noise at these same frequencies. If the RELM system is to be used to measure a 50-dB speech-noise signal from 200-6300 Hz, then the internal noise must be reduced in the high frequencies below those values encountered in the prototype system.

Finally, it is important to note that the shape of the speech-noise spectrum in Figure 7 is changing as the level increases from 50 to 80 dB. That is, the speech-noise is designed to represent the change in spectrum that accompanies real-life increases in vocal effort and corresponding speech intensity. As vocal effort increases from soft to normal to raised to shout, for example, the spectrum of the speech signal is not just elevated vertically by the same amount at all frequencies, but shows an upward shift in the peak of the spectrum. This is represented in the speech spectra incorporated into the ANSI SII standard and within the RELM system.

Using RELM to Fit Hearing Aids

Having established the reliability and validity of the loudness and speech-noise measurements made with this prototype of the RELM system, how can these measures be used to fit and evaluate hearing aids? One approach is to use these normative data as a guide to restoring normal loudness for soft, conversational and loud speech. Figure 8, for example, combines the three sets of mean speech-noise levels from Figure 5 (filled symbols) with the mean loudness contours from Figure 2 (dashed lines). (For the lowest speech-noise level of 50 dB, the unfilled circles represent an adjustment to the high-frequency 1/3-octave levels based on the discrepancy noted in Figure 7 at these frequencies that is due to the noise floor of the RELM prototype.) This provides an indication of where speech, from soft to loud, falls

within the loudness map of the normal-hearing listener and can be used to provide target amounts of gain to restore normal loudness across a wide range of frequencies and for a range of input intensities. For example, from Figure 8, it is apparent that normal loudness for the 65-dB speech-noise signal means that it should follow the SOFT loudness contour from 400 through 1000 Hz and should be 5-10 dB below that loudness contour at lower and higher frequencies. One can now make the RELM loudness measurements in an ear with hearing impairment to obtain their loudness map, observe the location of the SOFT loudness contour, and measure the unaided speech-noise levels for the 65-dB level (The speech-noise levels measured in the unaided ear of the hearing impaired should be about the same as the normative values, *on average*, since these are acoustic measurements made in the earcanal).

From these measurements, we can derive target gain values so as to amplify the 65-dB speech-noise to match the levels of the SOFT loudness contour from 400 through 1000 Hz and to fall between the VERY SOFT and SOFT loudness contours at lower and higher frequencies; that is, to

Figure 8. Overlay of the three speech-noise contours at 50, 65 and 80 dB from Figure 5 (retest values) on top of the RELM loudness contours from Figure 2 (test values). The unfilled circles at the higher frequencies for the 50-dB speech noise represent estimates of the true speech spectrum unaffected by the noise floor. These values were estimated from the differences between measured and targeted 1/3-octave levels for the 50-dB speech noise at the frequencies (observed previously in Figure 7).

restore normal loudness for this broadband, 65-dB speech noise. In similar fashion, targets can also be derived for the hearing-impaired listener at the 50- and 80-dB speech-noise levels in which the objective is to have the aided speech-noise levels positioned at the same locations within the impaired ear's auditory map as has been observed in the normal ear. In this way, normal loudness will be restored across a wide range of frequencies and intensities and can be confirmed using direct loudness measurements and acoustic speech-noise measurements obtained from the earcanal of the listener.

The RELM system, following the completion of the loudness mapping and the unaided speech-noise measurements, generates targets for the speech-noise measurements and instructs the clinician to insert the hearing aid while leaving the probe tube in place, then repeats the speech-noise measurements in this aided condition. The aided measurements of speech-noise 1/3-octave levels are then compared to those targeted to restore normal loudness for the 50-dB and 80-dB speech-noise levels and, if necessary, the hearing aids parameters are adjusted accordingly to improve the match to target. Aided measurements are repeated (5-10 seconds required) until a satisfactory match is obtained to the targets.

Effects of Loudness Summation on the Interpretation of RELM Results

Regarding the results plotted in Figure 8, at first glance, it might seem odd that "loud" speech would fall so close to the COMFORTABLE loudness contour and "normal" speech so near the SOFT loudness contour. One should keep in mind, however, that the loudness contours have been obtained with half-octave bands of noise whereas the loud or shouted speech is a broadband stimulus. This difference in bandwidth produces this apparent vertical discrepancy between the loudness contours and the speech-noise levels. As the bandwidth of the stimulus is increased, its loudness increases, especially for stimuli of at least moderate intensity (Scharf, 1978). Thus, if the loudness contours were established several different times, each time increasing the bandwidth of the stimuli used for the loudness judgments, the SPLs of the various loudness contours would progressively decrease with a maximum drop of about 8-10 dB. That is, for a constant loudness criterion (i.e., "LOUD"), as bandwidth is increased, the sound intensity required to produce the desired loudness sensation decreases. In

this way, loud or shouted speech (80 dB) would fall more nearly along the LOUD loudness contour and normal speech (65 dB) would fall closer to the COMFORTABLE loudness contour when the loudness contours were established for stimuli of broader bandwidth than used to produce the results in Figure 8. A similar mismatch between the spectrum of high-intensity (LOUD) speech and loudness ratings for narrow-band stimuli has been noted by Cox (1995) in the development of the Contour Test and a similar bandwidth-based argument has been used to explain the mismatch.

Of course, the RELM system could be used to target directly and confirm in the patient's earcanal a number of various theoretical objectives. For example, if one's objective is to establish that soft speech (50 dB) is above threshold at all frequencies and that shouted speech (80 or 85 dB) is never uncomfortable (TOO LOUD), an appropriate target can be specified and directly confirmed in the ear of the patient. The objective of restoring normal loudness across a wide range of frequencies and intensities, however, accomplishes this goal of assuring audibility for soft speech and precluding discomfort for loud sounds while also restoring normal loudness which will likely lead to good or pleasant sound quality. At present, this is the preferred objective for which the authors have designed the RELM system, although it can be used to evaluate a number of theoretical objectives. Given the lack of data supporting one theoretical rationale over another, however, this decision is based only on the assumption that it is desirable to restore normal loudness, rather than simply assuring that soft speech is audible and loud speech is not uncomfortable. With tools like RELM now available to directly confirm the accomplishment of various theoretical objectives for hearing-aid fitting and evaluation in the patient's earcanal, determination of "the ideal" theoretical objective can now be pursued reliably and validly. The authors look forward to the resolution of this important issue in the future.

REFERENCES

- Allen JB, Hall JL, Peng PS. (1990). Loudness growth in 1/2-octave bands (LGOB)—A procedure for the assessment of loudness. *J Acoust Soc Am* 88:745-753.
- American National Standards Institute. (1989). *Specifications for audiometers, ANSI S3.6-1989*. New York: American National Standards Institute.
- American National Standards Institute (pending). *Methods for the calculation of the speech intelligibil-*

- ity index, ANSI S3.79. Draft version 3.1, 7/3/93. New York: American National Standards Institute.
- Bentler RA, Pavlovic CV. (1989). Transfer functions and correction factors used in hearing aid evaluation and research. *Ear Hear* 10:58–63.
- Berger KW. (1984). *The Hearing Aid; Its Operation and Development*, 3rd ed. National Hearing Aid Society.
- Berkey DA, Marion MW, Robinson ME, Van Vliet DD (eds). (1992). New Technology: Programmable Hearing Aids. *Sem Hearing* 13(2):105–192.
- Blauert J. (1983). *Spatial Hearing*. Cambridge, MA: MIT Press.
- Byrne D, Tonisson W. (1976). Selecting the gain of hearing aids for persons with sensorineural hearing impairments. *Scand Audiol* 5:51–59.
- Byrne D, Dillon H. (1986). The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear Hear* 7:257–265.
- Cox RM. (1983). Using ULCL measures to find frequency/gain and SSPL90. *Hear Instrum* 34(7):17–21, 39.
- Cox RM. (1985). Hearing aids and aural rehabilitation: A structured approach to hearing aid selection. *Ear Hear* 6:226–239.
- Cox RM. (1988). The MSUv3 hearing instrument prescription procedure. *Hear Instrum* 39(1):6–10.
- Cox RM. (1995). Evaluation of the Contour Test for the measurement of loudness. Presented at the American Academy of Audiology convention, 1995.
- Cox RM, Alexander GC. (1991). Evaluation of an in situ output probe-microphone method of hearing aid fitting verification. *Ear Hear* 11:31–39.
- Dillon H. (1991). Allowing for real-ear venting effects when selecting the coupler gain of hearing aids. *Ear Hear* 12:406–416.
- Dirks DD, Kincaid GE. (1987). Basic acoustic considerations of ear canal probe measurements. *Ear Hear* 8:60S–67S.
- Dirks DD, Ahlstrom JB, Eisenberg LS. (1996). Comparison of probe insertion methods on estimates of ear canal SPL. *J Amer Acad Audiol* 7:31–38.
- Hawkins DB, Walden B, Montgomery A, Prosek R. (1987). Description and validation of an LDL procedure designed to select SSPL90. *Ear Hear* 8:162–169.
- Hawkins DB, Naidoo SV. (1993). Comparison of sound quality and clarity with asymmetrical peak clipping and output limiting compression. *J Amer Acad Audiol* 4:221–228.
- HIA (1995). *Hearing Industries Association Statistical Report*, October 23, 1995.
- Humes LE. (1988). And the winner is . . . *Hear Instrum* 39(7):24–26.
- Humes LE. (1991). Prescribing gain characteristics of linear hearing aids. In Studebaker GA, Bess FH, Beck LB (eds): *The Vanderbilt Hearing Aid Report II*. Parkton, MD: York Press, 13–22.
- Humes LE. (1996). Evolution of prescriptive fitting approaches. *Amer J Audiol* 5:19–23.
- Humes LE, Halling DC (1994). Overview, rationale and comparison of suprathreshold-based gain prescription methods. In: Valente M (ed): *Strategies for Selecting and Verifying Hearing Aid Fittings*. New York: Thieme Medical Publishers, 19–37.
- Humes LE, Houghton R. (1992). Beyond insertion gain. *Hear Instrum* 43(3):32–35.
- Humes LE, Jesteadt W. (1991). Models of the effects of threshold on loudness growth and summation. *J Acoust Soc Am* 90:1933–1943.
- Humes LE, Hipskind NM, Block M. (1988). Insertion gain measured with three probe tube system. *Ear Hear* 9:108–112.
- International Organization for Standardization. (1959). *Expression of the Physical and Subjective Magnitudes of Sound, ISO/R 131–1959*. Geneva: ISO.
- Kamm C, Kirks DD, Mickey MR. (1978). Effects of sensorineural hearing loss on loudness discomfort level and most comfortable loudness. *J Speech Hear Res* 21:668–681.
- Kiessling J. (1987). In situ audiometry (ISA). *Hear Instrum* 38(1):28–29.
- Killion MC. (1993). An attempt to present high-fidelity for the hearing impaired. In: Beilin J, Jensen G (eds.): *Recent Developments in Hearing Instrument Technology*. Copenhagen: Stougaard Jensen, 167–229.
- Kruger B, Kruger FM. (1994). Future trends in hearing aid fitting strategies: With a view towards 2020. In: Valente M (ed.): *Strategies for Selecting and Verifying Hearing Aid Fittings*. New York: Thieme Medical Publishers, 300–341.
- Martin FN, Morris LJ. (1989). Current audiologic practices in the United States. *Hear J* 43(4):25–44.
- McCandless GA, Lyregaard PE. (1983). Prescription of gain and output (POGO) for hearing aids. *Hear Instrum* 34(1):16–21.
- McCandless G. (1994). Overview and rationale of threshold-based hearing aid selection procedures. In: Valente M (ed.): *Strategies for Selecting and Verifying Hearing Aid Fittings*. New York: Thieme Medical Publishers, 1–18.
- Pluinage V. (1989). Clinical measurement of loudness growth. *Hear Instrum* 39(4):19–22.
- Probst R, Lonsbury-Martin BL, Martin GK. (1991). A review of otoacoustic emissions. *J Acoust Soc Amer* 89:2027–2067.
- Robomsp DW, Dadspms RS. (1956). A re-determination of the equal-loudness relation for pure tones. *Brit J Appl Phys* 7:166–181.
- Robinson K, Gatehouse S. (1996). Test-retest reliability of loudness scaling. *Ear Hear* 17:120–123.
- Scharf B. (1978). Loudness. In: Carterette EC, Friedman MP (eds.): *Handbook of Perception*, Vol. IV. New York: Academic, 187–242.
- Schwartz DM, Lyregaard PE, Lundh P. (1988). Hearing aid selection for severe-to-profound hearing loss. *Hear J* 41(2):13–17.
- Seewald RC. (1992). The desired sensation level

- method for fitting children: Version 3.0. *Hear J* 45(4):36,38–41.
- Seewald RC. (1995). The desired sensation level (DSL) method for hearing aid fitting in infants and children. *Phonak Focus*, 20:1–20.
- Skinner MW. (1988). *Hearing Aid Evaluation*. Englewood Cliffs, NJ: Prentice-Hall.
- Skinner MW, Pascoe DP, Miller JA, Popelka GR. (1982). Measurements to determine the optimal placement of speech energy within the listener's auditory area: A basis for selecting amplification characteristics. In: Studebaker GA, Bess FH (eds.): *The Vanderbilt Hearing Aid Report*. Upper Darby, PA: Monographs in Contemporary Audiology, 161–169.
- Valente M, Potts LG, Valente M, Vass W, Goebel J. (1994). Intersubject variability of real-ear sound pressure level: conventional and insert earphones. *J Amer Acad Audiol* 5:390–398.
- Van Tasell DJ. (1993). Hearing loss, speech, and hearing aids. *J Speech Hear Res* 36:228–244.