Loudness growth in 1/2-octave bands (LGOB)—A procedure for the assessment of loudness

Jont B. Allen, J. L. Hall, and P. S. Jeng[a)]
AT&T Bell Laboratories, 600 Mountain Ave, Murray Hill, New Jersey 07974-2070

(Received 5 November 1989; accepted for publication 16 March 1990)

In this paper, a method that has been developed for the assessment and quantification of loudness perception in normal-hearing and hearing-impaired persons is described. The method has been named LGOB, which stands for loudness growth in 1/2-octave bands. The method uses 1/2-octave bands of noise, centered at 0.25, 0.5, 1.0, 2.0, and 4.0 kHz, with subjective levels between a subject's threshold of hearing and the "too loud" level. The noise bands are presented to the subject, randomized over frequency and level, and the subject is asked to respond with a loudness rating (one of: VERY SOFT, SOFT, OK, LOUD, VERY LOUD, TOO LOUD). Subject responses (normal and hearing-impaired) are then compared to the average responses of a group of normal-hearing subjects. This procedure allows one to estimate the subject's loudness growth relative to normals, as a function of frequency and level. The results may be displayed either as isoloudness contours or as recruitment curves. In its present form, the measurements take less than 30 min. The signal presentation and analysis is done using a PC and a PC plug-in board having a digital to analog converter.

PACS numbers: 43.66.Fe, 43.66.Cb, 43.66.Yw, 43.66.Ts [NFV]

INTRODUCTION

This paper describes a procedure we have been using for assessing loudness growth in normal and hearing-impaired subjects. We have named the method LGOB, which stands for Loudness Growth in 1/2-Octave Bands.

LGOB is a method for quickly and accurately assessing loudness over frequency and level. This is useful for the evaluation of hearing impairment and was developed to fit multiband compression and automatic signal processing (ASP) hearing aids. The procedure has been designed to measure loudness growth in normal and hearing-impaired subjects, to be easy for elderly subjects, and to run in a reasonable amount of time.

The measurement method is based on absolute-loudness judgments of presentations of 1/2-octave bands of periodic noise centered at 0.25, 0.5, 1.0, 2.0, and 4.0 kHz that have been randomized over level and frequency. Each of 15 equispaced levels in each band was presented three times per band during the course of the experiment.

We have experienced no difficulty with subjects being able to perform the task, and the test–retest data for both normal and hearing-impaired subjects have shown a reasonable degree of consistency.

LGOB is an extension of procedures developed by Garner (1953), Galanter and Messick (1961), Pascoe (1978), and Geller and Margolis (1984). Our results differ in that we have, first, normalized the data against normal subjects to show recruitment, and, second, displayed it as isoloudness contours. Third, we have trained the subjects during an initial phase, and at the same time estimated their dynamic range of hearing on a band-by-band basis. Fourth, we have automated the procedure to greatly reduce the measurement time. And finally, we have extended the dynamic range of the measurement system and simultaneously increased the measurement speed by using computer generated signals with 16-bit digital converters and programmable attenuators. The increased dynamic range was necessary to measure over the entire dynamic range of hearing for normals.

We believe that the method will be useful in a clinical environment for fitting hearing aids and, perhaps, might be applied to the diagnosis of hearing problems that previously have been difficult to categorize or measure.

I. METHODS

A. Procedure

Each subject to be tested was seated in a sound-proof booth and given a set of instructions that indicated that they would be hearing different test signals, and that they should rate the loudness of these signals using one of the six responses: TOO LOUD—6, VERY LOUD—5, LOUD—4, OK—3, SOFT—2, VERY SOFT—1. Besides these categories, the tester also entered NO RESPONSE—0, if the subject failed to respond to the stimulus presentation. The numerical value shown after each rating is strictly for accounting purposes, and was not known to the subject. The instructions defined the meaning of the categories for the subjects. A copy of the instructions has been included as Appendix A. The subject was asked if he or she was ready, and then the test began.

False alarms were not a problem, since there were no trials in which a stimulus was not presented. Signaling the subjects upon a stimulus presentation was unnecessary because of the 5- to 8-s pace between presentations. The subjects quickly learned that they would receive the next stimulus immediately after they had responded to the previous one. If the subjects responded with a TOO LOUD response, they were given a few extra seconds to recover. The subjects

[a)] Also at the Center for Research in Speech and Hearing Sciences, City University of New York, New York, New York 10036-8099.

were given a box with a safety button they could push if they felt that they needed to stop the stimulus for any reason. The safety button was incorporated to make the subjects feel in control and at ease. Subjects pushed the button only infrequently.

After each signal presentation, the subject responded with one of the descriptive words, and the tester entered the response on the PC keyboard, using a one-letter abbreviation. During the testing phase, if the subject changed his or her mind, the tester had the ability to enter the correction. Such errors occurred about once or twice per subject run (e.g., one time per 15–30 min).

The procedure has two phases, a "limits and practice" phase, and a "data collection" phase. During the first phase, the subject estimated the upper and lower bounds of the sound pressure used for testing, and simultaneously obtained practice on the task. For the results reported here, the first phase began with monotonically descending presentations, starting at 60 dB SPL, with 5-dB decrements. These continued until the subject gave NO RESPONSE, followed by ascending presentations with 5-dB increments, starting from 65 dB SPL, and terminating on the first TOO LOUD.

The second phase is data collection with presentations randomized over the five bands and 15 levels, as described below.

In a subsequent application of this method, we made several modifications which will be described in Sec. III of this paper. These changes sped up the program and allowed the upper and lower bounds to be dynamic.

B. Stimuli

The five 1/2-octave bands of periodic noise were generated via fast Fourier transform (FFT) as the sums of sinusoids with random phases. All of the amplitudes of the sine components were identical. The stimuli had a base period of 32 ms (512 points at a 16-kHz sampling rate). By using an FFT, each tone component was continuous at the 32-ms boundaries, giving clean, artifact-free, periodic noise samples. The resulting frequency spacing of the stimulus components was 31.25 Hz. As previously stated, the center frequencies were 0.25, 0.5, 1.0, 2.0, and 4.0 kHz. (Note that the 250-Hz band consists of six lines separated by 31.5 Hz, starting with the seventh harmonic of 31.25 and ending with the 12th.) The phase of each sine component was different each time the program was run. The signal was gated on and off with a 15-ms linear ramp and was on for 0.5 s and off for 0.5 s. Three such noise bursts were used for each presentation.

In each frequency band, 15 different levels were used, equispaced on a dB scale, with each amplitude-frequency condition presented three times. Thus the total number of randomized trials was $225 = 15$ (levels) $\times 5$ (frequency bands) $\times 3$ (presentations). If the subjects took 8 s per response, then the procedure would take 34 min, which was greater than the typical measurement time. Because of transducer power limitations, we frequently could not reach the TOO LOUD level in some bands. This was especially true for the hearing-impaired subjects. It was also true for the normals in the 250-Hz band.

C. Equipment

The LGOB procedure was programmed in Lahey FORTRAN (Lahey Computer Systems, Inc., P.O. Box 6091, Incline Village, NV) on an AT&T PC-6300 (an IBM XT compatible). The signals were presented through a signal processing board that plugs into the PC-XT bus and has a 16-bit digital to analog converter (Ariel DSP-16, Ariel Corp., 433 River Road, Highland Park, NJ 08904). The levels of the presented stimuli were varied using a combination of a programmable attenuator (Wavetek 617) that had an 80-dB range in 1-dB steps (the attenuator was connected to the PC via a National GPIB interface) and scaling via integer arithmetic on the signal processing board. The total dynamic range of the system was about 120 dB. However, this large dynamic range was seldom needed for these experiments. The minimum usable signal level was limited by a noise floor that was more than 90 dB below the maximum level. The harmonic distortion floor was determined by the transducer.

D. Calibration

Before running each subject, the system was calibrated using a 1/2-octave noise band centered at 1 kHz with a Brüel & Kjaer 4157 coupler fitted with a B&K 4155 microphone and a B&K 2230 sound-level meter. The transducer package was a behind-the-ear transducer (Knowles ED 1932 series) coupled to the ear canal via an E-A-R® (Division of Cabot Corp.) foam eartip. Correction factors were introduced to compensate for the frequency response of the in-the-ear transducer as measured with the 4157 coupler using each of the five 1/2-octave bands of noise. Pure tone thresholds were made on the subjects using an Etymōtic ER-2 insert earphone which was calibrated with a DB-100 (Industrial Research Products, Inc.) artificial ear. The ED receiver was required to reach the TOO LOUD sound levels. The ER-2 was chosen so that we could measure thresholds at 8 kHz.

E. Subjects

Two sets of subjects were run, a normal group of 15 ears, one per subject, and a hearing-impaired group of 16 ears, from 12 subjects. The normal group was selected from young adults, with no history of hearing loss, and having a normal audiogram. Only one ear was used in these subjects in order to gain the maximum amount of information about intersubject variability with the minimum amount of testing. Pure-tone thresholds were taken for each of the normal subjects, and then the subjects were run on the LGOB test. All testing was done in a sound booth. The subjects' responses were monitored using an intercom. Eleven of the subjects had similar loudness-response relationships, and the results from these 11 subjects were averaged to get the "average–normal" curves used in the remainder of this paper.

Two ears were used for those hearing-impaired subjects that wore binaural hearing aids. If a subject wore one hearing aid, only the aided ear was used. For the hearing-impaired ears, we ran pure-tone thresholds, bone-conduction thresholds, and word lists at three sound-pressure levels. The only selection that was done on the hearing-impaired group was an attempt to ensure that they had a cochlear loss rather than middle ear or retrocochlear loss, based on their audiological history and air–bone gap.

II. RESULTS

In Fig. 1, responses from the 1-kHz band are shown for one of the normal-hearing subjects. The subject's responses are shown as the symbols in the two panels. A loudness rating of 6 corresponds to the TOO LOUD level, while a 0 corresponds to NO RESPONSE, as defined in Sec. I A. Many subjects suggested that the number of categories was insufficient, and they frequently volunteered responses that were between categories. In these cases, we forced the subject to respond in one of the seven categories. The effect of this quantization to the nearest response category is easily seen in the loudness growth plots of this figure. Because the quantization was so strong, we found that we could model it. This model allowed us to estimate the *transition levels* between the response category boundaries. We also estimated the *average-response levels*. The two estimates, when used together, gave us different information about the underlying subject response. The transition-level estimate was somewhat less robust than the average-level estimate (e.g., a σ of 3.9 dB for the average levels and 4.9 dB for the transition levels, as described in Fig. 3). The average levels, on the other hand, are meaningless at the end points (TOO LOUD and NO RESPONSE).

The two panels show the different methods used for fitting the data points. In modeling ordinal data, one must take special care. It is incorrect to average OK responses with LOUD responses, because they represent different things to the subject. We have approached this problem in two ways. First, we have formed averages over each response category. The second approach was to assume an underlying continuous curve, and use an assumption of forced quantization. We call this model the nearest integer (NINT) model.

In the upper panel, we have drawn a quantized curve through the data points. Consider the following model: Assume that a continuous underlying monotonic response curve exists which is $R(\wp)$, where $\wp$ is the sound pressure. We choose R so that the residual error between the measured

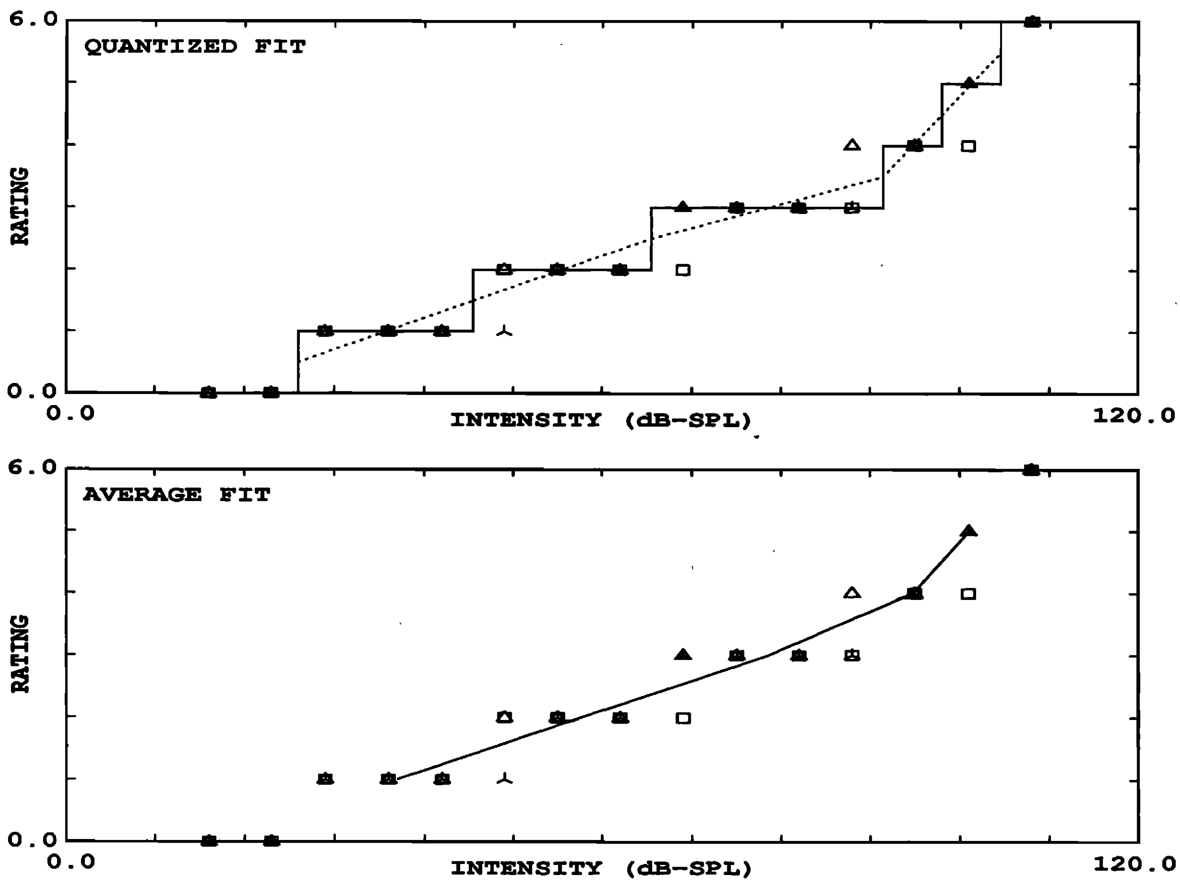

FIG. 1. This figure shows the two methods used for forming the two estimates of the loudness growth curve. The dotted line in the upper panel shows the results of estimating the *response-transition* levels. The solid line going through most of the data points results from quantizing the response-transition curve (the dotted line) to the nearest integer response (the ordinate). We view this quantization as being similar to the subject's forced choice of an adjectival category. The solid line gives the smallest estimation error, in that it minimizes the error [see discussion of Eq. (1)]. The lower panel shows the results of estimating the *response-average levels* (solid line), which is obtained by averaging the sound levels, in dB, for each rating. In this way, we avoid meaningless averages across loudness ratings. The raw subject data are shown by the symbols. The three different symbols at each pressure level code temporal order of the presentations.

data $M(\wp)$ and the quantized response NINT[$R(\wp)$] [where NINT ( ) is the nearest-integer function] is minimized. In the upper panel of Fig. 1, the dashed line is $R(\wp)$, the solid line is NINT(R), and the symbols define M.

Find $R(\wp)$ so that

$$\min_{R(\wp)} \sum_{\wp} |\mathrm{NINT}(R(\wp)) - M(\wp)|. \quad (1)$$

The specification of $R(\wp)$ is not complete until its parametrization is specified. We parametrized R as straight lines between coordinates $(\wp_l, R_l)$, where $\wp_l$ is the lth abscissa coordinate, and $R_l = R(\wp_l)$. In general, one may vary $\wp_l$ as well as R_l in forming the minimization. However, in the procedure that has been used in this paper, the underlying curve $R(\wp)$ has been further constrained to have transitions that are always halfway between two sound–pressure measurement levels. This has probably resulted in slightly inferior estimates of R relative to what might be obtained by the more general procedure. These joined line segments are shown as a dashed line in the upper panel of Fig. 1 and form one estimate of the loudness response curve in terms of its *transition levels* between category boundaries.

The solid line in the lower panel connects the averages (in dB) of all the sound–pressure levels corresponding to a given loudness category. This curve forms our second estimate of the loudness level curve, which we call the *average levels*. Knowing the transition levels as well as the average levels estimates $R(\wp)$ for two different sets of $\wp$.

In Fig. 2, we show typical results for one of our hearing-impaired subjects. The frequency band is indicated in the upper right corner of each panel of the plot. For each band, a solid line connects the average data, while a broken line connects the transition levels. The different symbols correspond to the three ratings at each level and are not significant in the present context.

All of the normal-hearing subjects, and most of the hearing-impaired subjects, regularly displayed more rapid increase in rating at the higher end of the scale between LOUD (4) and TOO LOUD (6). This effect is seen in the 0.25- and 0.5-kHz bands of Fig. 2, starting at a level of about 95 dB SPL.

In the lower right-hand panel, the data have been plotted as a function of frequency to form a family of isoresponse contours. In this figure, each solid line connects the average levels as a function of frequency, while the transition levels are connected by dotted lines. The region delimiting OK level is marked by □'s. The Δ's mark the average OK level. We marked the OK range because of its general importance to our application of fitting hearing aids. Marking the OK range in this manner made the plots easier to read, since there are so many lines on the plots.

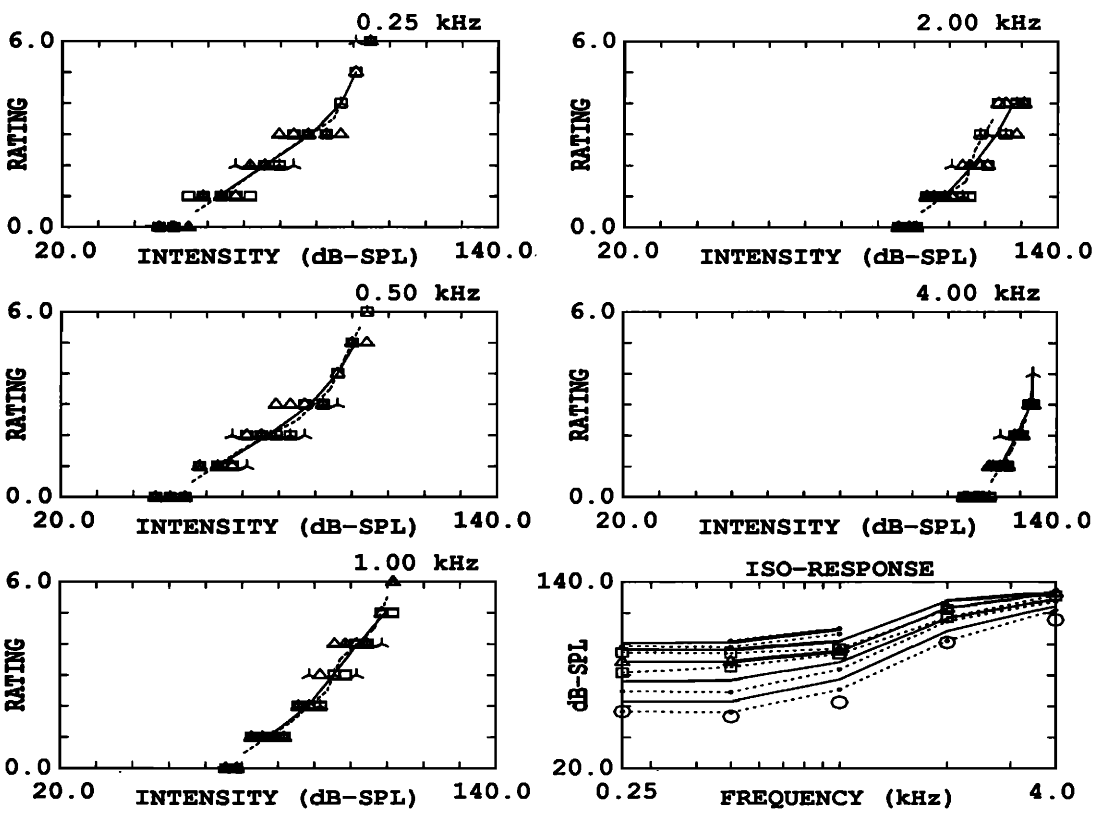

FIG. 2. We show here loudness growth data for one of our subjects (105) as a function of intensity for each band, and in the lower right panel, the resulting isoresponse contours. The center frequency of the noise is given in the upper right-hand corner of each panel. The raw subject data are shown by the symbols. The three different symbols at each intensity code temporal order of the presentations. The data have been smoothed by two procedures, as described in the text [see Eq. (1)]. Response-transition levels are shown by the dashed curves, and the response-average levels are given by the solid lines. In the lower right panel, Δ's mark the OK response average levels and □'s mark the OK-response transition intensities for the subject. Pure-tone thresholds are shown as O's.

In Fig. 3, we show a similar set of curves but for averages over the database of 11 normal-hearing subjects. The two methods used for smoothing the data almost superimpose in this case. Four subjects out of the 15 original normal subjects were excluded from the average–normal database. Two of these subjects had loudness curves that were above the average–normals, and two were below. The standard deviation (s.d.) over the population of 11 subjects is 3.9 dB (computed in dB units), and the s.d. for the original 15 normal-hearing subjects is 7.8 dB, averaged over bands, levels, and subjects. No intrasubject errors are presented here, other than the test–retest data discussed in Fig. 6.

At the bottom right panel of Fig. 3, we have indicated average–normal pure-tone thresholds with the O's. Figure 4 is this same panel expanded to show more detail. The difference between the thresholds for pure tones and for 1/2-octave bands of noise is believed to be due to the difference in transducers and corresponding couplers used to calibrate the transducers.

To further explore the difference between these two threshold measures, we measured one subject's (JBA) thresholds to a 1/2-octave band of noise and a pure tone at 4 kHz using a single transducer, and without moving the transducer between measurements. In this case, we found the thresholds to be the same. We interpret this experiment as indicating that the difference we observe at high frequencies is due to the two different transducers *and* their calibrations using the two different couplers.

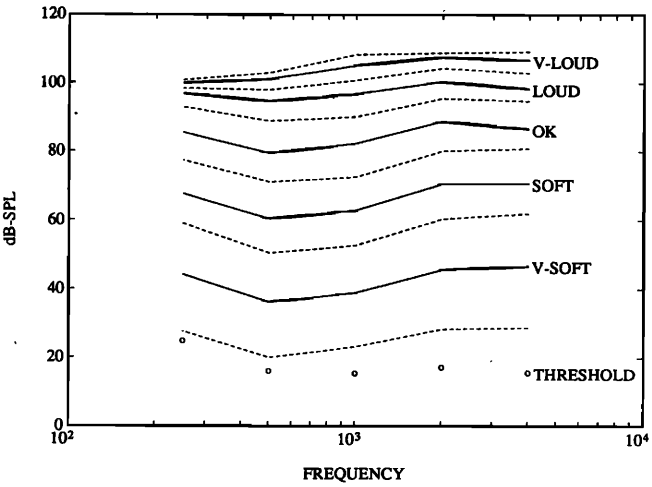

FIG. 4. The isoresponse contours from Fig. 3 expanded for clarity.

When comparing the normal subject's data from Fig. 3 to the hearing-impaired subject's data within each frequency band, we find that each hearing-impaired subject shows elevated thresholds, while in some frequency bands, the subject has near normal TOO LOUD levels. This difference is the result of a reduced dynamic range, or equivalently, recruitment. Recruitment is frequently described as an increase in the slope of the loudness-growth curves. The recruitment for subject 105 (Fig. 2) is most obvious at the lower sound-

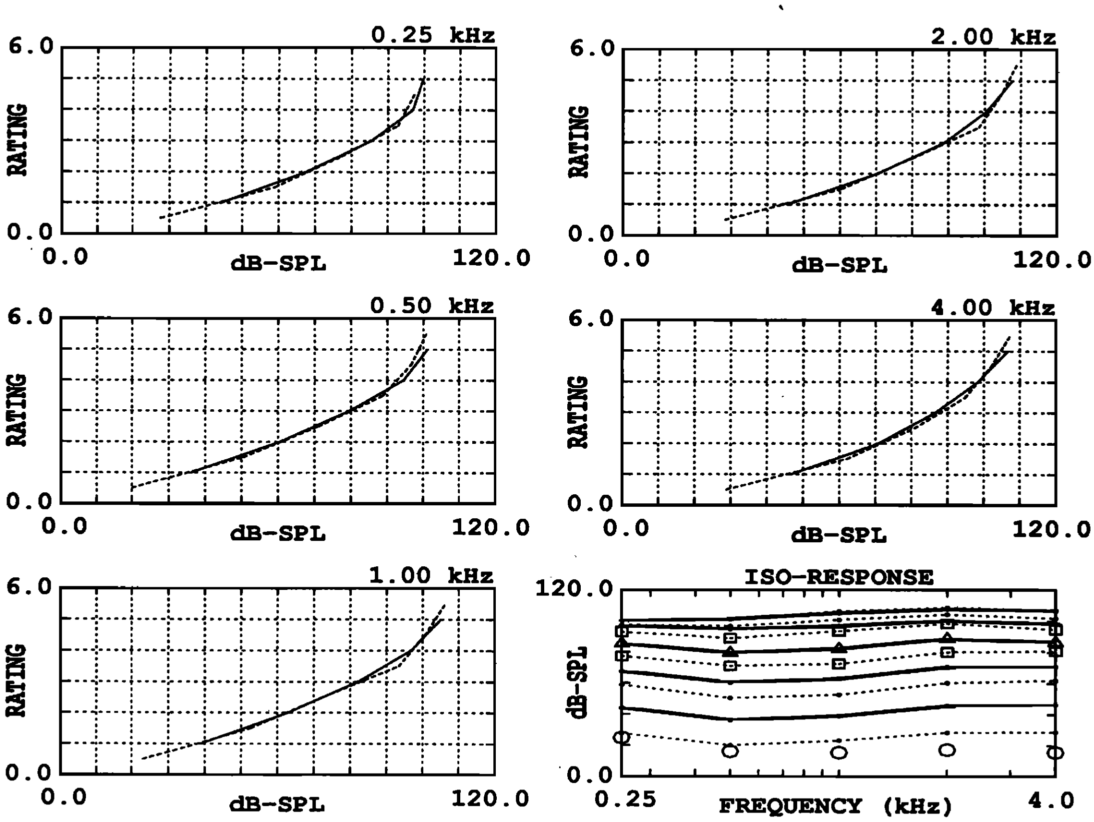

FIG. 3. This plot is the same as that of Fig. 2, except that it is for an average of 11 normal subjects. The individual data points are not shown. The s.d., averaged over subjects, bands, and levels, for the average-response levels (solid line) is 3.9 dB, and for the transition-response levels (dashed line) is 4.9 dB. The lower right panel gives the isoresponse contours for the average–normal subjects. The symbols are the same as in Fig. 2.

pressure levels and for the higher bands, as may be seen in Fig. 5.

In Fig. 5, subject 105's sound-pressure level for a given category is used as the abscissa, while the average–normal subject's sound-pressure level for the same category is used on the ordinate, on a band-by-band basis. Plotted in this way, the data show the classical recruitment of a person having a unilateral hearing loss, when plotting the sound-pressure level in the bad ear against that in the good ear for an equal loudness response. A conductive loss would be represented by a parallel shift, while a cochlear loss is seen as a change in dynamic range (a nonparallel curve). In this plot, as before, the Δ's represents the sound pressure corresponding to the average OK level, while the □'s have been placed on the boundaries of OK.

In Fig. 6, we show test–retest reliability for a normal subject (JBA, subject 131) made on different days. It is likely that the 5- to 7-dB differences observed are caused by slight changes in the ear canal pressure due to the transducer coupling (e.g., the depth of the transducer placement in the ear canal), rather than being due to subjective measurement variations, or other variations over time, because the largest differences are only in the lower bands, and are largely independent of level, from 20 to 90 dB SPL. Ear canal probe-tube measurements are needed to quantify this type of variation further. We have found our results to be reproducible over days and weeks. In a separate experiment, we measured loudness of three frequency bands replicated four times with each of six normal-hearing subjects. A statistical analysis of these data gave a within subject s.d. of 2.9 dB. This estimate of the s.d. was averaged over subjects, bands, and level. The NO REPONSE and TOO LOUD categories were not included in this analysis.

In Fig. 7, we compare adjectival loudness to loudness in sones for each band for the normal data of Fig. 3. Starting from Stevens' data for 1/2-octave bands in a diffuse sound field (Stevens, 1956, p. 838, Fig. 22), we applied two correction factors, one for diffuse to free-field, and one for free-field to eardrum. The total correction was (0.75, 0.10, − 0.33, 13.3, 15.15) dB at (0.25, 0.5, 1.0, 2.0, 4.0) kHz. This correction represents the effects of head diffraction and ear canal resonance. We plot here the loudness rating of Fig. 3 as a function of the measured sone value for each frequency band. From this figure, we see that the OK rating is equiva-

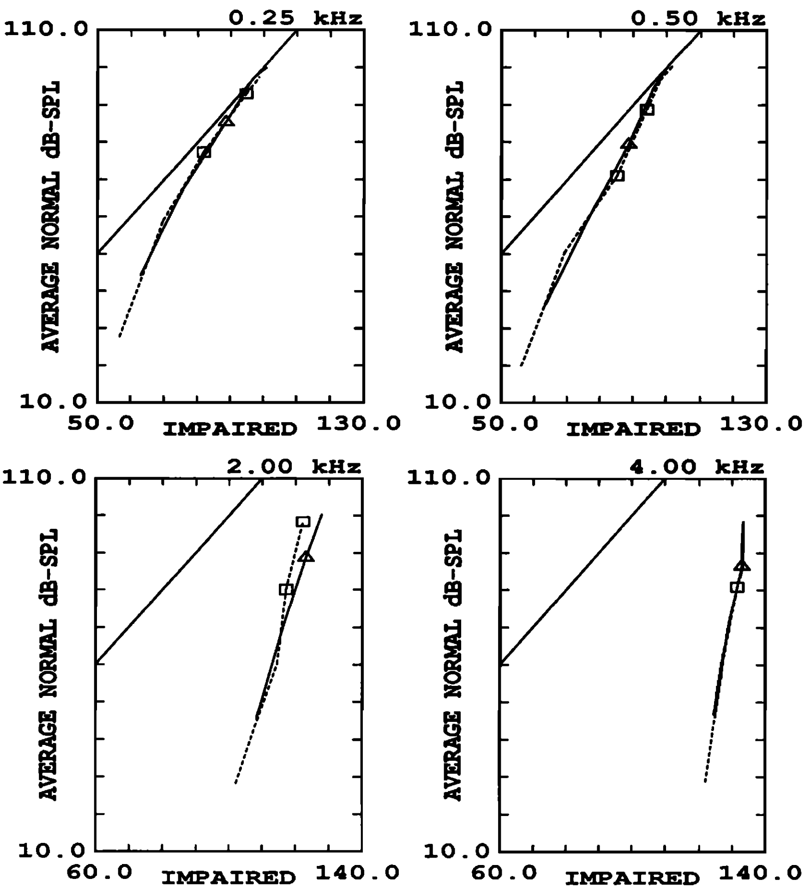

FIG. 5. When we plot the hearing-impaired subject against the average normal-hearing subjects, we see recruitment, or abnormal growth of loudness. As before, the symbols indicate the range of OK-responses. Both axes are in dB SPL. The solid straight line corresponds to normal hearing. When the loudness-growth curve drops away from the normal curve, the subject requires an intensity for a given response that is larger than the average–normal subject's intensity by the horizontal difference between the two curves. In this figure, we show the recruitment plots for subject 105. Note how the recruitment increases for the higher frequency bands, and that the subject's hearing is almost "normal" in the 250-Hz band at the OK level (e.g., 80–95 dB SPL).

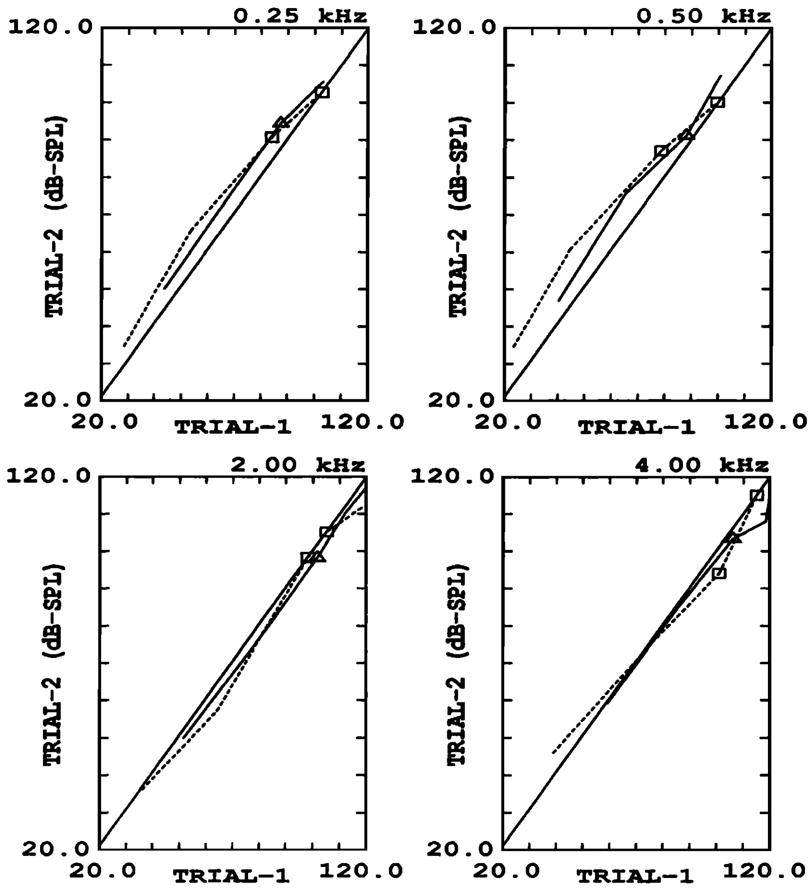

FIG. 6. Test–retest data show small errors that are typically less than 10 dB. Systematic (level independent) errors may be accounted for by earphone placement variations.

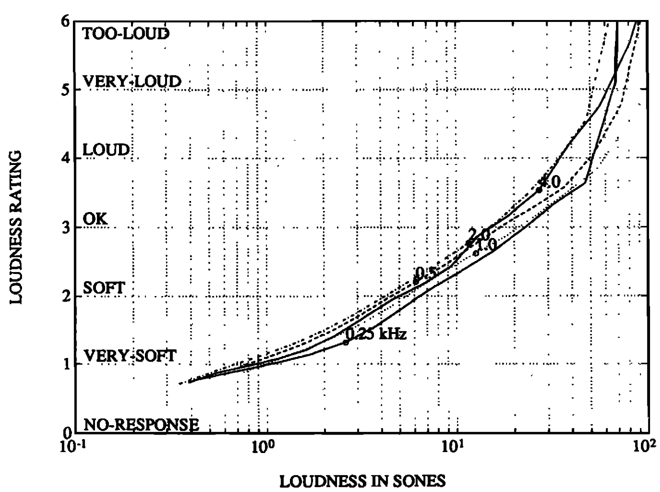

FIG. 7. We have used Stevens' (1956) loudness data for 1/2-octave bands of noise as a function of pressure level for comparison to our loudness ratings data of Fig. 3. The Stevens' data was measured relative to a diffuse field. After transforming it to eardrum pressure (see text), we plotted the rating data of Fig. 3 against the Stevens' sone values for each frequency band. The clustering of the curves across frequency bands is consistent with the view that the loudness ratings and the sone scale are both measures of the same perceptual quantity in normal-hearing subjects.

lent to about 18 sones. It would seem from this curve that the loudness ratings for the average–normal population are highly correlated with the sone scale.

III. DISCUSSION

The LGOB procedure described here has allowed us to measure absolute loudness, loudness growth and recruitment as a function of frequency, and isoresponse contours for both normal-hearing and hearing-impaired subjects, in under 30 min. The data of Fig. 7 indicate that for normal-hearing subjects, this procedure appears to be functionally equivalent to time-consuming *sone* measurements. While 30 min is long for a clinical procedure, given the amount of information obtained from the test, the procedure could be useful in the clinic for fitting nonlinear hearing aids, such as multiband compression hearing aids, which attempt to compensate for recruitment.

Weaknesses in the procedure were discovered, as is discussed in Sec. IV, but, overall, the strategy worked well. Our original application of fitting compression hearing aids, which required knowledge of the loudness growth curves for the subjects as a function of frequency, was easily met using LGOB. The use of a seven-category scale was good, although we felt that perhaps one more category would be an improvement (see below). The OK rating could have been misinterpreted by our subjects. However, we believe that the subjects understood the simple concept behind the rating scale, and were not confused because of the subject instructions (Appendix A), which limited any potential misunderstanding. The subjects understood that OK is an easily spoken alternative to COMFORTABLE. The subjects did not question its usage.

This suprathreshold measurement of hearing loss evaluation could turn out to be useful for: (1) fitting hearing aids, because many new hearing-aid fitting procedures now frequently require such information, usually in the form of the most comfortable level (MCL) and loudness discomfort level (LDL); (2) quantifying hearing loss, since this measurement allows the simple diagnosis of both cochlear loss, in the form of recruitment curves, and may allow for diagnosis of conductive loss; and in (3) collecting large amounts of loudness data to test various loudness models against experiment. For example, the loudness of simultaneous bands of noise could be estimated for loudness summation experiments. It would also be interesting to rerun these tests at various bandwidths.

We have not been able to determine if retrocochlear loss may be diagnosed by LGOB since we did not have a retrocochlear subject base. However, we have seen one anomaly that has not been easy to explain. With two "normal" subjects (based on pure-tone thresholds) we saw normal thresholds levels with abnormal LOUD and TOO LOUD levels. That is, these subjects showed a form of "recruitment," but had normal thresholds. No explanation for these unusual responses has yet been found.

Many new hearing-aid fitting procedures require an estimate of the subject's MCL or LDL. Procedures that attempt to measure these points in isolation are subject to a large variance across subjects, probably due to the difficulty in defining these levels. Dirks and Kamm (1982) and Kamm *et al.* (1978) investigated the measurement of the LDL and MCL in isolation using adaptive methods and showed significant dependence on the exact instructions used in specifying those levels. It has also been observed that when the MCL or LDL is measured by a sequential method (e.g., monotonically increasing or decreasing levels), large subject-dependent biases are frequently present, again leading to a measurement method with a large degree of uncertainty (Neuman and Levitt, personal communcation). We believe that we have minimized these problems using the LGOB procedure. First, *by randomizing the presentations over the entire hearing range we seem to be able avoid both the definitional problems*, probably because the end points (e.g., threshold and TOO LOUD) are more firmly anchored, and the sequential bias, because the presentations are random. The fact that most of our normal subjects (11 our of 15) give similar results is consistent with the possibility that the loudness scale (VERY SOFT, SOFT, OK, LOUD, VERY LOUD, and TOO LOUD) may be interpreted on an absolute scale. Namely, unlike the Dirks and Kamm experiment, the subjects are less sensitive to the instructions because of their strong previous experience with the descriptive categories. This prevous experience might be functional *only* in the context of randomized presentations that span the entire dynamic range of their hearing. Second, we feel that it is important to *not* require the subject to project into the future as some instructions have attempted to do. For example, an instruction of the sort: *Turn down the level when you feel that you could not listen at this level for more than five minutes*, means different things to different subjects.

In our definition of LGOB, we chose 1/2-octave bands for the test stimulus because we wished to average over subject and transducer variations. At frequencies near the frequencies of cochlear emission, 10-dB threshold variations have been seen, both with normal and hearing-impaired ears. In principle, transducer variations could be calibrated out, but impedance variations between the coupler and the subject's ear will affect the results. Our choice of 1/2-octave bands was a compromise between covering most of the frequency spectrum (e.g., 50%) and having sufficient frequency resolution with a small number of test noise bands. Finally, we also felt that periodic noise was perceptibly closer to speech than were tones, and we were interested in estimates of loudness growth of speech for fitting compression hearing aids. (Speech does not work at all in a test of this type because a 500-ms segment of speech has large loudness variability.)

IV. IMPROVEMENTS

We have been looking for ways to further improve the procedure. In a recent version, we tried several modifications. These modifications seemed to be significant improvements in terms of speed, accuracy and convenience. First, in the training phase of the program, we used random presentations over level and frequency, rather than monotonic changes. This has the advantage that the subject practices on a procedure that, from the subject's point of view, seems identical to the testing phase. In fact, it differs in that we used only five levels spread over the range of testing levels during the practice phase, versus 15 levels during the testing phase. We believe that the subjects find this training less confusing because of its similarity to the testing phase.

At the same time, we introduced a dynamic boundary algorithm that allowed us to estimate the upper and lower boundaries continuously, more accurately, and faster. For example, during the training phase, we monitor the subject's response to the *highest* presentation level in each band. If the response is OK, we increase the maximum level by 12 dB. If the response is LOUD, we increase the level by 6 dB. If the response is VERY LOUD, we increase the maximum by 2 dB. If the level is TOO LOUD, we *decrease* the maximum by 3 dB. When we start the testing phase, we use a maximum presentation level that gave the TOO LOUD level during the training phase.

After the final limits have been found in the practice phase using five levels per band, 15 levels are established between these limits per band (uniformly on a log scale), for the testing phase. This adaptive strategy in the testing phase reduces the number of TOO LOUD and NO RESPONSE responses that we elicit from the subject. We found that we frequently gave only two TOO LOUD presentations (once during the training phase, and once during the testing

TABLE I. Increments that were used to dynamically vary the limits.

Minimum limit		Maximum limit	
Response	Change	Response	Change
OK or more	−18	TOO LOUD	−3
SOFT	−9	VERY LOUD	+2
VERY SOFT	−5	LOUD	+6
NO RESPONSE	+3	OK or less	+12

phase) in each band using this procedure. Since the procedure is dynamic, the subjects are somewhat protected from receiving too many TOO LOUD presentations. Table I shows the maximum and minimum limit increments, in dB, that were used for the adaptive control as a function of the subject's response.

Other changes that we feel would be an improvement, but which we have not tried, are to add a VERY VERY SOFT level, to provide user feedback via a subject response light that would tell the subject that a stimulus has been delivered, and to include a set of subject buttons that, when hit, would display the selected response on the tester's console. If the subject were capable of running the test alone, then the subject's button could be set to control the responses directly without the tester's intervention. Finally, it might be reasonable to increase the density of pressure levels in the LOUD to TOO-LOUD range due to the increased rate of growth. Alternatively, it might be possible to bias the subject's responses using different instructions about the VERY-LOUD range using different wording, to decrease the large slope in that region.

ACKNOWLEDGMENTS

We would like to thank Arlene Neuman for suggesting the Pascoe method as a method for estimating loudness growth. We would also like to thank Professor Harry Levitt for many helpful discussions and comments on the manuscript.

APPENDIX: INSTRUCTIONS TO THE SUBJECT

3/3/87

Loudness Estimation

The purpose of this experiment is to determine how loud some noise bursts sound to you. After each stimulus presentation (a series of three short noise bursts), please tell the experimenter how loud the noise bursts sounded, using one of the following categories:

Very soft
You would ask someone talking this loud to speak up.

Soft
Soft conversation level.

OK
Most comfortable conversational level.

Loud
Loud conversational level.

Very loud
You would ask someone speaking this loud not to shout.

Too loud
Uncomfortably loud.

There are no "right" or "wrong" answers. All that matters is how loud the noise bursts sound to you. You may very well find yourself using some response categories more often than others. This is perfectly all right.

In the first part of the experiment, the noise bursts will first decrease, then increase, in loudness. In the second part of the experiment, the various kinds of noise bursts will be presented in random order. Please stay in the soundproof room, without removing the earpiece, until the experiment is completed. The experiment takes about twenty minutes.

Do you have any questions?

Dirks, D., and Kamm, C. (**1976**). "Psychometric functions for loudness discomfort and most comfortable loudness level," J. Speech Hear. Res. **19**, 613–627.

Galanter, E., and Messick, S. (**1961**). "The relation between category and magnitude scales of loudness," Psychol. Rev. **68** (6), 363–372.

Geller, D., and Margolis, R. H. (**1984**). "Magnitude estimation of loudness I: Application to hearing aid selection," J. Speech Hear. Res. **27**, 20–27.

Garner, W. R. (**1953**). "An informational analysis of absolute judgments of loudness," J. Exp. Pscyhol. **46** (5), 373–380.

Kamm, C., Dirks, D. D., and Mickey, M. R. (**1978**). "Effect of sensorineural hearing loss on loudness discomfort level and most comforable loudness judgments," J. Speech Hear. Res. **21**, 668–681.

Neuman, A., and Levitt, H. (**1985**). Personal communcation.

Pascoe, D. P. (**1978**). "An approach to hearing aid selection," Hear. Instrum. **29**, 12–16, 36.

Stevens, S. S. (**1956**). "Calculation of the loudness of complex noise," J. Acoust. Soc. Am. **28**, 807–832.