Journal of Speech and Hearing Research, Volume 32, 816–828, December 1989

COMFORTABLE LOUDNESS LEVEL: STIMULUS EFFECTS, LONG-TERM RELIABILITY, AND PREDICTABILITY

ROBYN M. COX
Memphis State University and Veteran's Administration Medical Center, Memphis

This paper reports the results of a series of investigations of comfortable loudness levels with particular reference to their application to hearing aid gain prescriptions. Experiment 1 studied the effects of several stimulus waveforms, bandwidths, and durations on comfortable loudness levels for normal and hearing impaired listeners. Speech band comfort levels were found to be significantly higher than equal-duration noise band or warble tone comfort levels. Comfortable loudness levels were found to be independent of warble tone modulation parameters and of stimulus bandwidth (stimuli did not exceed critical bandwidths). In Experiment 2, reliability of comfortable loudness levels was evaluated in hearing-impaired subjects over two consecutive 1-year periods. Results indicated that comfortable loudness levels were slightly less reliable than thresholds. In addition, the results were consistent with a hypothesis that exposure to amplified sound produces a small increase in comfortable loudness levels. In Experiment 3, data from 67 hearing-impaired subjects were used to develop regression equations for prediction of comfortable loudness levels. Thresholds at the test frequencies were combined with comfortable loudness data at 500 Hz and 4000 Hz. The prediction method was then evaluated using a new group of 25 subjects. Accuracy of predictions of comfort levels was substantially better with the new method than with an older method that relied exclusively on threshold data. Relevance of the outcomes to hearing aid fitting procedures is discussed.

KEY WORDS: Comfortable loudness, hearing aid gain, stimulus duration, speech bands

Several studies support the contribution of comfortable loudness levels in determining appropriate hearing aid gain (e.g., Leijon, Eriksson-Mangold, & Bech-Karlsen 1984; Lippmann, Braida, & Durlach, 1981; Sullivan, Levitt, Hwang, & Hennessey, 1988). Overall, this work suggests that comfortable loudness levels are closely related to desired frequency/gain function in a hearing aid fitting. A number of investigators have proposed that comfortable loudness for speech signals should be used as the basis for hearing aid gain prescriptions. Recent work by Byrne (1986a) has lent support to this approach. His study demonstrated that a frequency response based on comfortable loudness for narrow bands of speech babble produced more intelligible and pleasant speech than any of three other approaches tried.

Potential measures of comfortable loudness levels include the lower or upper limits of the comfortable loudness range, the range itself, the middle of the range, or the point that is labelled as the "most" comfortable loudness level (MCL). The purpose of this article is to report the results of several investigations of comfortable loudness as characterized by the upper limit of the comfortable loudness range. In previous work, this measure has been called the ULCL. However, because this mneumonic for upper limit of comfortable loudness is often confused with UCL (*uncomfortable* loudness), this terminology has been changed. Thus, the mneumonic used to designate the upper limit of the comfortable loudness range is HCL (highest comfortable loudness).[1] These studies were undertaken to explore the application of comfortable loudness measurements in hearing aid prescription procedures.

Despite research evidence supporting the use of comfortable loudness levels in hearing aid prescription strategies, several factors operate against the direct measurement of comfortable loudness, especially for speech stimuli, in hearing aid fitting protocols. They include the following: (a) appropriate speech stimuli are not readily available; (b) reliability of comfortable loudness measurement is reputed to be poor; and (c) comfortable loudness measurement is time-consuming. The studies reported here addressed these three issues. The results should be applicable to any hearing aid prescription procedure that utilizes the concept of loudness comfort for speech signals as a factor in determining appropriate hearing aid amplification. In addition, these data have implications for other types of investigations in which comfortable loudness is measured.

EXPERIMENT 1: EFFECT OF STIMULUS PARAMETERS ON COMFORTABLE LOUDNESS LEVELS

The theoretical framework for hearing aid fitting often

---

[1] The upper limit of the comfortable loudness range was selected for measurement on the basis of several investigations that suggested it was more robust and repeatable than other points in the comfortable loudness area, especially when measured using a descending approach (e.g., Berger, Varavvas, & Vottero, 1982; Lucker, Grzybmacher, & Ventry, 1978; Ventry & Johnson, 1978). However, psychophysical procedures that utilize a set of response labels to explore the comfortable loudness range may produce equally reliable results for other points on the comfortable loudness continuum (e.g., Hawkins, et al., 1987; Pluvinage, 1989; Sammeth, Birman, & Hecox 1989; Skinner, Pascoe, Miller, & Popelka, 1982).

0022-4685/89/3204-0816$01.00/0

invokes the concept of comfortable loudness for normal conversational speech. This is operationally defined in terms of comfortable loudness for ⅓-octave bands of speech babble as a function of frequency. Because narrow bands of speech babble are not readily available, different narrow-band stimuli may be used for actual comfortable loudness measurements. However, the relationship between comfortable loudness for speech bands and comfortable loudness for other narrow-band stimuli has received relatively little attention. Cox and Bisset (1982) compared noise band and speech-band HCLs. The speech babble stimuli were presented continuously to maximize their resemblance to real speech. The noise bands were pulsed 2.5 times per second with an on-time of 200 ms. Results were interpreted as indicating that speech-band and noise-band comfort levels were essentially equal. Byrne (1986b) reported that speech-band MCLs were measured at significantly higher levels than pure tone MCLs in one study but that he was unable to replicate this finding in a second study. These two reports with somewhat contradictory findings appear to exhaust the literature on this topic.

In addition to the potential effects of stimulus waveform, comfortable loudness levels may be affected by stimulus bandwidth in hearing impaired persons with sloping audiograms, even for sub-critical narrow band stimuli. This possibility is suggested by the observation that stimulus bandwidth affects thresholds in these individuals (e.g., Orchik & Mosher, 1975). Some workers have suggested using a highly frequency-specific stimulus, such as a 5%–10% warble (frequency modulated) tone, for loudness measurements (Cox, 1985; Hawkins, Walden, Montgomery, & Prosek, 1987). However, a brief listening comparison between warble tones having, for example, rectangular versus sinusoidal modulation waveforms, reveals that different types of warble tone stimuli sound rather different qualitatively. Furthermore, certain modulation parameters of warble tones affect their thresholds (Barry & Resnick, 1978). These kinds of considerations suggest that the comfortable loudness for warble tones may vary with their bandwidth and/or their modulation parameters. Thus a warble-tone comfort level may not be an accurate estimate of the comfort level for a ⅓-octave speech band with the same nominal frequency.

To further explore the relationship between comfortable loudness for speech bands and narrow band stimuli having different waveforms and bandwidths, a series of studies was undertaken in which comfortable loudness levels were measured for warble tones, speech bands, and noise bands.

EXPERIMENT 1A: EFFECT OF SPECTRAL FACTORS

In this investigation, normal-hearing subjects provided HCL and threshold data for six stimuli. The goals were to explore the relationship between speech-band and warble-tone HCLs, and to replicate the finding of Cox and Bisset (1982) regarding equivalence of speech-band and noise-band HCLs. Thresholds were measured in addition to HCLs because it was of interest to determine whether any effects of stimulus on comfort levels would be paralleled by corresponding effects on thresholds.

METHOD

Subjects. One ear was tested for each of 15 normal hearing young adults. Their pure tone thresholds were <15 dB HL (re ANSI, 1969) in the range 250 through 4000 Hz. Because of potential ambient noise masking, subjects with hearing at or better than 0 dB SPL were excluded from the study.

Stimuli. Nominal test frequencies were 250, 1000, and 4000 Hz. The six stimuli at each frequency were (a) ⅓-octave noise band; (b) ⅓-octave band of 6-talker babble; (c) ⅓-octave, sinusoidally modulated warble tone; (d) ⅓-octave, rectangularly modulated warble tone; (e) ±5% sinusoidally modulated warble tone; and (f) ±5% rectangularly modulated warble tone. All warble tones were modulated at the rate of 5 Hz. Figure 1 illustrates the long-term RMS spectra of the ⅓-octave stimuli at 1000 Hz. The ±5% warble tones were simply narrower versions of the warble tones seen in this figure. It should be noted that even though some of the stimuli seen in Figure 1 had similar long-term spectra, their short-term spectra were rather different.

Procedure. Testing was conducted in a double-walled audiometric room. Stimuli were digitized with a 5-kHz audio bandwidth and routed through an audiometer (Fonix, model 3100) that was controlled by a microcomputer.

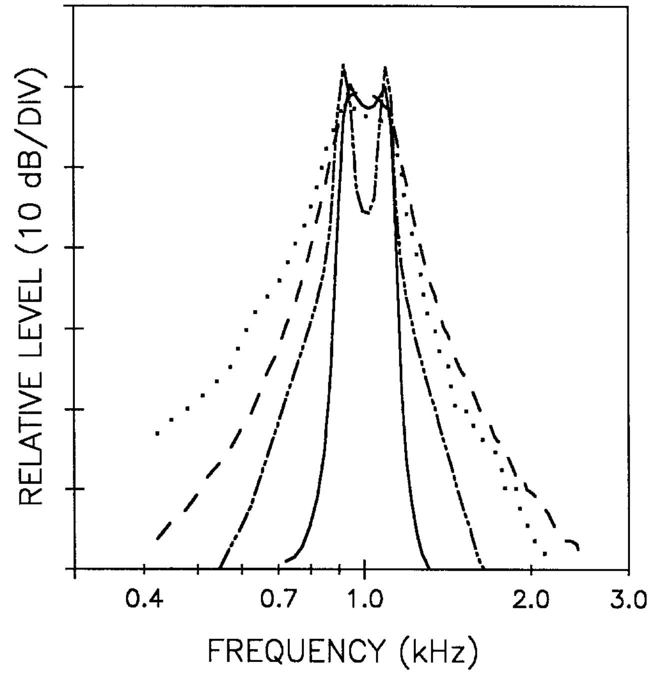

FIGURE 1. Long-term RMS spectra of ⅓-octave bandwidth stimuli at 1000 Hz used in Experiment 1a. Amplitude placement is arbitrary. Dotted line = speech band, dashed line = noise band, dash-dot line = rectangularly-modulated warble tone, solid line = sinusoidally-modulated warble tone.

An insert earphone (Etymotic ER3A) was used to present the stimuli monaurally. The nontest ear was occluded with a compressible foam earplug.

Thresholds were measured using standard audiometric procedures including an ascending approach mode and a 5-dB increment. HCLs were measured using the procedure described in Cox (1985 Appendix A). The procedure consisted of two phases for each new stimulus: (a) a search phase during which the approximate location of the HCL was determined, and (b) a test phase during which the final value was fixed. The test phase incorporated a descending approach mode and a 5-dB decrement. Subjects were instructed to respond whenever the stimulus was presented at a level that would be comfortable for long-term listening. After a response, the level was increased and another descending run initiated. The final HCL was the highest level at which the subject responded in two out of three runs during the test phase.

Reliability was checked within each test session using standard audiometric procedure. This involved retesting the first stimulus after several stimuli had been tested. Agreement within 5 dB between initial and repeated measurements was required. If this criterion was not met, testing continued until levels stabilized.

HCLs were tested first for all subjects. All other variables were counterbalanced or randomized to minimize order effects. Stimuli were calibrated in terms of the long-term RMS level produced by the ER3A earphone in a DB-0138 $2cm^3$ coupler, measured using an integrating sound level meter (Larson Davis, model 800B). Each stimulus was 1200 ms in duration with a rise/fall time of 25 ms. A 1.7-s response window was provided after each stimulus presentation. During this time, the subject responded by depressing a button if the stimulus was judged to be comfortable for long-term listening.

RESULTS AND DISCUSSION

To explore the effects of stimuli on HCLs and thresholds, the data were subjected to two separate analyses of variance (ANOVA), one for HCL data and one for threshold data (both expressed in long-term RMS levels). Variables were stimulus (six), and frequency (three). Post hoc testing utilized the Student-Newman-Keuls procedure. Results indicated that, overall, the six different stimuli did not produce significantly different thresholds ($p > .05$). Mean thresholds were 20.4, 6.8, and 10.5 dB SPL for 250, 1000, and 4000 Hz respectively. In contrast to the threshold results, there were significant differences among mean HCLs for different stimuli [$F(5,70) = 4.13$, $p < .01$]. Specifically, the mean speech-band HCL was higher than the HCLs for the other five stimuli ($p < .05$). The noise band and warble tone comfort levels were not significantly different from each other. The solid symbols in Figure 2 illustrate the mean long-term RMS HCL levels; the pattern was similar for all three test frequencies.

Because speech bands were found to be comfortable at higher levels than noise bands or warble tones, it is tempting to postulate that this outcome was related to the meaningful nature of speech compared to the other stimuli. Previous research has shown that meaningful sounds are acceptable at higher levels than nonmeaningful ones (Kerrick, Nagel, & Bennett, 1968). However, it is important to note that ⅓-octave bands of speech babble do not resemble speech very closely—they sound more like rumbles and jingles, depending on frequency. In addition, subjects were not informed that they were listening to speech-band stimuli. Thus, the meaningfulness of the stimuli may not be a satisfactory explanation of the interstimuli differences in HCL levels shown in Figure 2.

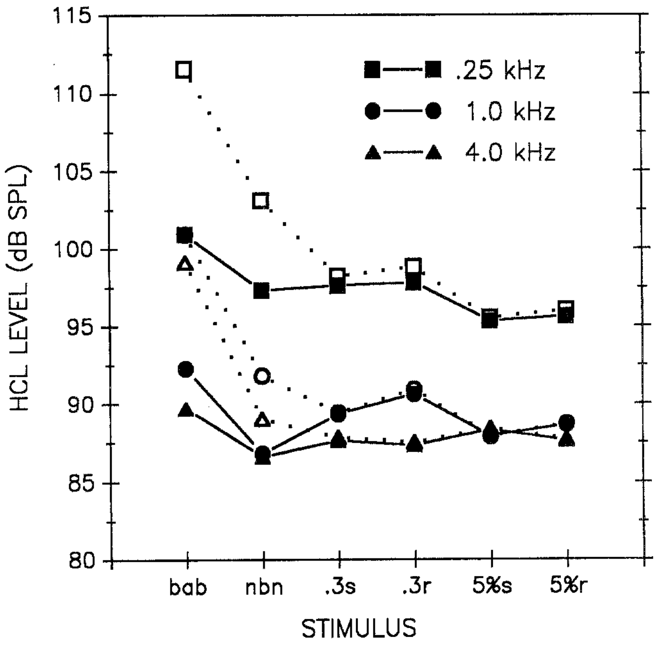

FIGURE 2. HCL levels for each test stimulus in Experiment 1a. Solid symbols = average long-term RMS levels, open symbols = 99th percentile levels. Bab = speech band, nbn = noise band, .3s = ⅓-octave sinusoidally-modulated warble tone, .3r = ⅓-octave rectangularly-modulated warble tone, 5%s = ±5% sinusoidally-modulated warble tone, 5%r = ±5% rectangularly-modulated warble tone.

The finding that different stimuli were not equally comfortable at equal long-term RMS levels indicates that factors in addition to overall SPL are important in the judgment of comfortable loudness. One stimulus component that could be implicated in comfortable loudness judgments is the peak distribution. For example, we may hypothesize that a stimulus with frequent, high, short-term levels would be judged less comfortable than one with the same long-term SPL but fewer high short-term levels. To explore this possibility, 20-ms level distributions were determined for each stimulus using the method described by Cox, Matesich, and Moore (1988). The open symbols in Figure 2 show the mean HCL levels plotted in terms of the 99th percentile of each stimulus; only 1% of RMS levels in 20-ms samples would exceed these levels. Thus, this measure is akin to a measurement of the peaks of each stimulus. As Figure 2 implies, only the speech-band and noise-band stimuli showed substantial variation in short-term levels. For the warble tone stimuli, RMS and peak levels were essentially the same. The figure illustrates that expressing HCLs in terms of the peak levels exaggerates rather than reduces the differences among speech-band, noise-band and warble tone comfort levels.

Further examination of the 20-ms distribution measurements suggested that the short-term percentile level at which the six stimuli were most closely equated was different for the different frequencies—approximately the 50th percentile for 250-Hz stimuli, the 60th percentile for 1000-Hz stimuli, and the 75th percentile for 4000-Hz stimuli. This observation implies that as frequency increases, there is a systematic increase in the short-term level of the stimulus that listeners find most salient when judging whether the stimulus is comfortably loud.

To explore the relationship between speech-band HCLs and HCLs for the other stimuli, linear correlation coefficients were computed and are shown in Table 1. The mean correlation coefficient between speech-band and noise-band HCLs was .80, a value somewhat lower than the .85 obtained by Cox and Bisset (1982) with hearing-impaired subjects. This outcome may be the result of greater homogeneity among the normal-hearing subjects in the present study. The correlations between speech-band and warble tone HCLs were quite similar to the correlations between speech-band and noise-band HCLs at 250 and 1000 Hz. At 4000 Hz, the warble tone/speech-band correlations were somewhat lower than the noise-band/speech-band correlation. However, this outcome was not statistically significant at the .05 level (Ferguson, 1966, p. 189, equation 12.13) and was not replicated in a related study with hearing-impaired listeners (see Experiment 1c). In addition, there is no clear evidence in Table 1 that either bandwidth or modulation waveform affected the relationship between warble tone and speech-band HCLs. On the whole, the results given in the table indicate that, at least for normal hearers, warble tone HCLs are about equal to noise band HCLs as predictors of speech-band HCLs regardless of the modulation waveform or bandwidth of the warble tones.

TABLE 1. Correlation coefficients between speech-band HCLs and HCLs for five other stimuli. Standard errors of estimate (dB) for prediction of speech-band HCL from each other stimulus are shown in parentheses: nbn = ⅓-octave noise band; .3s = ⅓-octave sinusoidally modulated warble tone; .3r = ⅓-octave rectangularly modulated warble tone; 5%s = ±5% sinusoidally modulated warble tone; 5%r = ±5% rectangularly modulated warble tone.

Freq (Hz)	*nbn*	*.3s*	*.3r*	*5%s*	*5%r*
250	.93	.93	.96	.85	.82
	(3.9)	(4.0)	(3.1)	(5.6)	(6.1)
1000	.67	.78	.62	.66	.69
	(6.9)	(5.8)	(7.3)	(6.9)	(6.3)
4000	.81	.66	.70	.48	.67
	(4.4)	(5.5)	(5.3)	(6.5)	(5.5)
Mean	.80	.79	.76	.66	.73

Because the speech-band HCLs were significantly higher than the noise-band HCLs, these data did not replicate the results of Cox and Bisset (1982) in which speech-band and noise-band HCLs were found at essentially equal levels. Instead, the present data resemble the observation by Byrne (1986b) in which speech-band MCLs were measured at significantly higher levels than pure tone MCLs. The most obvious difference between the present study and that of Cox and Bisset is in the temporal characteristics of the stimuli. In the former study, the speech band was presented continuously, whereas the noise band was pulsed 2.5 times per second. In the present study, both speech band and noise band were delivered in 1200-ms pulses. Another study was performed to explore the effects of this difference.

EXPERIMENT 1B: EFFECT OF DURATION

In this investigation, an attempt was made to resolve the discrepancy in outcome between Experiment 1a and the study by Cox and Bisset (1982). Normal hearing listeners provided comfortable loudness data for ⅓-octave speech-band and noise-band stimuli at 1000 Hz. Stimulus duration was varied to approximate the temporal conditions of Experiment 1a and those used by Cox and Bisset.

METHOD

Subjects. One ear was tested for each of 10 normal hearers.

Stimuli. Four stimuli were used: (a) ⅓-octave speech band with 5-s duration; (b) ⅓-octave speech band with 1200-ms duration; (c) ⅓-octave noise band with 1200-ms duration; and (d) ⅓-octave noise band presented in 4 pulses with 400-ms on-time and 350-ms off-time. Stimuli (a) and (d) were intended to approximate the continuous speech babble and pulsed noise band, respectively, used by Cox and Bisset (1982). Stimuli (b) and (c) were identical to the conditions used in Experiment 1a.

Procedure. All testing conditions were the same as those used in Experiment 1a except that the stimuli were presented using a TDH-49 earphone and calibrated in a NBS-9A 6 cm^3 coupler. Stimuli were presented in random order.

RESULTS AND DISCUSSION

The mean long-term RMS HCL level for each stimulus is shown in Table 2. A one-way analysis of variance revealed a significant effect due to stimulus [$F(3,27) = 6.78$, $p < .002$]. Post hoc testing with the Student-Newman-Keuls procedure revealed, as shown in the table, that the 1200-ms speech-band stimulus resulted in an HCL level that was significantly higher ($p < .05$) than that seen for any other stimulus. As found in Experiment 1a, the 1200-ms speech-band comfort level was significantly

TABLE 2. Mean long-term RMS HCL levels (dB SPL) for the four stimuli used in Experiment 1b. Underlining connects stimuli for which the HCLs were not significantly different at the .05 level.

speech band 1200-ms	*speech band 5-s*	*noise band 400-ms*	*noise band 1200-ms*
95.5	92.0	91.0	88.5

higher than the 1200-ms noise band comfort level. In addition, the outcome of Cox and Bisset (1982) was replicated in that the pulsed noise band comfort level was at essentially the same level as the 5-s speech-band comfort level.

These results underscore the importance of the temporal characteristics of stimuli chosen for comfortable loudness measurement. Clearly, as the on-time of a stimulus decreased, it was judged by these listeners to be comfortably loud at higher levels. Cox and Bisset (1982) apparently chose temporal parameters for their speech-band and noise-band stimuli that resulted in equivalent comfort levels for the two stimuli. However, Experiments 1a and 1b were consistent in indicating that, when speech-band and noise-band stimuli are presented in constant duration pulses, the speech-band comfort level is found at a higher long-term RMS level than the noise band comfort level.

Previous investigators have noted an effect of stimulus duration on comfortable loudness level tracking (Melnick, 1967; Rintlemann & Carhart, 1964). In each case, a pulsed stimulus was tracked at a higher MCL than the same stimulus presented continuously. Although attempts have been made to relate this phenomenon to temporal integration of loudness and loudness memory effects, these factors have not been found to fully explain the observation. The results of Experiments 1a and 1b indicate that the effect of stimulus duration on comfortable loudness is observed at the upper limit of the comfortable loudness range as well as at the MCL and that this effect is seen with another psychometric procedure as well as tracking. Because all of the stimuli were longer than the 200-300-ms needed for full loudness development in normal ears, the result cannot be attributed to temporal integration effects.

Experiments 1a and 1b examined loudness comfort judgments made by normal hearers. Although there is no reason to postulate that hearing-impaired persons use different guidelines for judging loudness from those used by normal hearers, certain aspects of the hearing loss itself could result in different outcomes. For example, for persons with sloping audiometric configuration, variations in stimulus bandwidth similar to those used in Experiment 1a might affect loudness judgments because low frequency spectral components are more audible in the wider bandwidth stimuli. Thus, ⅓-octave and ±5% warble tones might have different comfort levels for a hearing-impaired listener even though this is not seen in normal hearers. Similarly, because temporal processing is often defective in hearing-impaired ears, the effect of stimulus duration on comfortable loudness may be different from that observed in normal hearers. To explore these possibilities, an additional study was performed in which HCLs were measured for several stimuli using hearing-impaired listeners.

EXPERIMENT 1C: HEARING-IMPAIRED LISTENERS

In this study, several stimulus variables investigated in normal hearers were evaluated in hearing-impaired listeners to assess the acceptability of generalizing the findings of Experiments 1a and 1b to hearing-impaired persons.

METHOD

Subjects. One ear was tested for each of 18 persons with bilateral sensorineural hearing loss. Mean pure tone hearing level at 4000 Hz was 63 dB HL re (ANSI), 1969 (*SD* = 10.8 dB). Mean pure tone audiogram slope between 2000 Hz and 4000 Hz was 16.4 dB (*SD* = 14.7 dB).

Stimuli. Eight stimuli, all centered at 4000 Hz, were used: (a) ⅓-octave multitalker babble, 5-s duration; (b) ⅓-octave noise band, 400-ms duration; (c) ⅓-octave sinusoidally modulated warble tone, 400-ms duration; (d) ⅓-octave rectangularly modulated warble tone, 400-ms duration; (e) ±5% sinusoidally modulated warble tone, 400-ms duration; (f) ±5% rectangularly modulated warble tone, 400-ms duration; (g) ±5% sinusoidally modulated warble tone, 1200-ms duration; and (h) ±5% rectangularly modulated warble tone, 1200-ms duration. The 400-ms stimuli were presented in four pulses with 350-ms off-time.

These stimuli were chosen to allow examination of the effects of stimulus type (noise band and warble tone), bandwidth (⅓-octave and ±5%), duration (1200-ms vs. 400-ms), and modulation waveform (sinusoidal and rectangular) on comfort levels for hearing impaired listeners. In addition, the comfortable loudness relationships were determined between "continuous" (5 s) speech bands and the other stimuli.

Procedure. HCL test procedure and instrumentation were the same as used in Experiment 1b. In addition, threshold was tested for each stimulus. HCLs were tested first. Stimuli were presented in random order.

RESULTS AND DISCUSSION

Mean long-term RMS HCL and threshold data are given in Table 3. Separate analyses of variance were performed on each data set with post hoc testing using the Student-Newman-Keuls procedure. The outcomes are indicated in the table. In addition, this table gives the correlation coefficient computed between speech-band HCLs and HCLs for each other stimulus. Figure 3 illustrates the distribution of differences between HCL levels for speech bands and corresponding HCL levels for each

Table 3. Mean long-term RMS thresholds and HCLs (dB SPL) measured for eight 4000-Hz stimuli using hearing-impaired subjects. Underlining connects stimuli that were not significantly different at the .01 level. Values in parentheses are correlation coefficients (r) between HCLs for speech-band stimuli and corresponding HCLs for each other stimulus: bab = ⅓-octave speech band; nbn = ⅓-octave noise band; .3s = ⅓-octave sinusoidally modulated warble tone; .3r = ⅓-octave rectangularly modulated warble tone; 5%s(a) = ±5% sinusoidally modulated warble tone, 1200 ms; 5%s(b) = ±5% sinusoidally modulated warble tone, 400 ms; 5%r(a) = ±5% rectangularly modulated warble tone, 1200 ms; 5%r(b) = ±5% rectangularly modulated warble tone, 400 ms.

Test	*bab*	*.3r*	*.3s*	*nbn*	*5%r(b)*	*5%r(a)*	*5%s(b)*	*5%s(a)*
Thr.	68.2	71.9	72.5	73.6	74.2	74.7	75.6	75.8
HCL	102.4	101.4	101.7	103.3	101.9	101.9	103.1	101.4
r		(.89)	(.86)	(.97)	(.94)	(.90)	(.92)	(.87)

other stimulus (data for all stimuli are combined together in this Figure, $N = 126$). These results supported the following conclusions:

1. There were no significant differences among mean HCLs for the eight stimuli. Furthermore, the difference between the HCL for the continuous speech band and the corresponding HCL for pulsed noise band or warble-tone stimuli was 5 dB or less in 84% of comparisons. Finally, correlations between speech-band HCLs and HCLs for other stimuli were high.
2. There were no significant differences between HCLs for 1200-ms and 400-ms pulsed warble tones.
3. Several significant threshold differences were seen between ⅓-octave and ±5% bandwidth stimuli, as would be expected with persons having sloping audiometric contours [$F(7,119) = 19.86$, $p < .001$]. However, there were no differences between HCLs for stimuli with these two bandwidths.

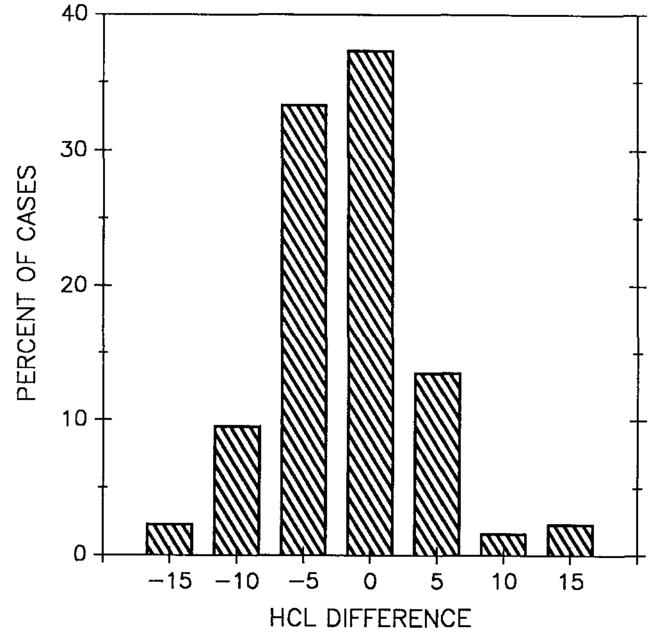

Figure 3. Distribution of differences between HCLs for speech bands and HCLs for other stimuli used in Experiment 1c. $N = 126$.

Overall, the comfortable loudness data obtained from hearing-impaired listeners were very similar to the corresponding data reported for normal hearers.

EXPERIMENT 2: LONG-TERM RELIABILITY OF COMFORTABLE LOUDNESS LEVELS

The repeatability of the measured comfort levels is one of the major issues affecting the usefulness of these measures in hearing aid prescription strategies. Clearly, comfortable loudness measures must be reasonably stable over time if they are to serve as the basis for a frequency/gain function intended for long-term use. Numerous studies of the day-to-day reliability of comfortable loudness measures have been reported (see Skinner, 1988, chapter 5, for a review). However, only the study of Christen and Byrne (1980) appears to have assessed the reliability of comfortable loudness measures over an extended period of time. These investigators measured MCL in test and retest sessions separated by 5–19 months for 6 hearing-impaired subjects. The results indicated that although MCLs measured on successive days were reasonably similar (usually within 10 dB), one of the 6 subjects (17%) showed a difference of about 22 dB between test and retest that were separated by 18 months. The investigators concluded that comfortable loudness measurements were not sufficiently reliable to serve as the basis for hearing aid fittings.

To further evaluate the long-term stability of comfortable loudness measurements, an investigation was undertaken in which comfortable loudness levels were measured on several occasions over a period of 2 years. In addition to assessing the absolute repeatability of the comfortable loudness level, it was of interest to examine the possible effects of daily use of amplification on comfortable loudness levels. If exposure to amplified sound results in an increase in comfortable loudness levels, this could have implications for loudness-based gain prescriptions, particularly for persons who are obtaining their first hearing aid.

Because threshold measurements have been suggested as predictors of loudness comfort levels, long-term reliability was evaluated for threshold measures as well as for HCL measures.

METHOD

Subjects. One ear was tested for each of 10 subjects. Ages were 23 to 88 with a mean of 67 years. All had essentially sensorineural hearing loss of mild to severe extent. Table 4 gives the means and standard deviations (*SD*s) of pure tone thresholds in the test ears. Seven subjects obtained their first hearing aid after the first HCL test. All subjects wore amplification throughout the time period of the study. Reported hearing aid use ranged from 1 to 16 hrs/day with a mean of 8.5 hrs/day.

Stimuli. Nominal test frequencies were 500, 800, 1000, 1600, 2500, and 4000 Hz. Stimuli were either narrow bands of noise with ⅓-octave bandwidth and 24 dB/octave slope or −5% warble tones with rectangular modulation and modulation rate of 3.5/sec. For each subject, the stimulus type was held constant throughout the study.

Procedure. HCLs and thresholds for the test stimuli were determined three times for each subject. Elapsed time between the first and second tests ranged from 7 to 24 months with a mean of 15. Between the second and third tests, times ranged from 10 to 17 months with a mean of 11. Stimuli were delivered using either a supra-aural earphone (Telephonics TDH-39 in MX-41/AR cushion), calibrated in an NBS-9A 6 cm^3 coupler; or using an insert earphone (Danavox SMW) attached to an earmold and calibrated in an HA-2 2 cm^3 coupler. For each subject, the transducer type was held constant throughout the study.

HCL and threshold test procedures were the same as those used in Experiment 1 with the exception that the test administration was controlled manually rather than by microcomputer. Testing was done in a double-walled sound-treated room. Thresholds were measured first, followed by HCLs. Testing began at the lowest frequency and proceeded to each higher frequency. A different tester, who was not aware of the results of previous tests, collected the data for each test session.

TABLE 4. Means and standard deviations *(SD)* of pure tone thresholds in the test ears of subjects (N = 10) in Experiment 2. Data are in dB HL re ANSI, 1969.

Freq (Hz)	*Mean*	*SD*
250	31.5	16.3
500	35.5	13.4
1000	45.0	12.0
2000	55.0	15.8
4000	65.5	14.6

RESULTS AND DISCUSSION

To evaluate any systematic changes in thresholds or HCLs over time, these data were subjected to a three-way repeated measures analysis of variance. Variables were frequency (six), test (threshold/HCL), and session (three). Results indicated that there was a significant main effect for session [$F(2,18) = 3.98$, $p < .04$] but no significant session-test interaction. Further inspection revealed that this outcome indicated an increase in both thresholds and HCLs over the two years of the study. Collapsing across frequencies, mean thresholds for the three sessions were 57.8, 58.9, and 60.2 dB SPL. Analogous mean values for HCLs were 97.2, 100.7, and 100.8 dB SPL. Thus, there was a small progression in hearing loss and HCL levels over the time of the investigation.

Test-retest differences for both thresholds and HCLs were determined for session 1 versus session 2 and for session 2 versus session 3. Another three-way ANOVA was run on these data [frequency(six) × test(two) × session differences(two)]. None of the main effects or interactions was significant ($p > .05$). As a result, test-retest data were combined across frequencies and sessions. Figure 4 illustrates the distributions of test-retest differences for both thresholds and HCLs (N = 120 for each distribution). The mean test-retest differences were 1.2 dB and 1.8 dB for thresholds and HCLs, respectively (this is consistent with the slight upward trend in the data for both measures). The standard deviations of the test-retest differences were 6.7 dB and 8.0 dB for thresholds and HCLs, respectively.

In a study of the day-to-day reliability of HCL measures, Cox and Bisset (1982) found that the standard deviation of test-retest differences was 5–6 dB. The pre-

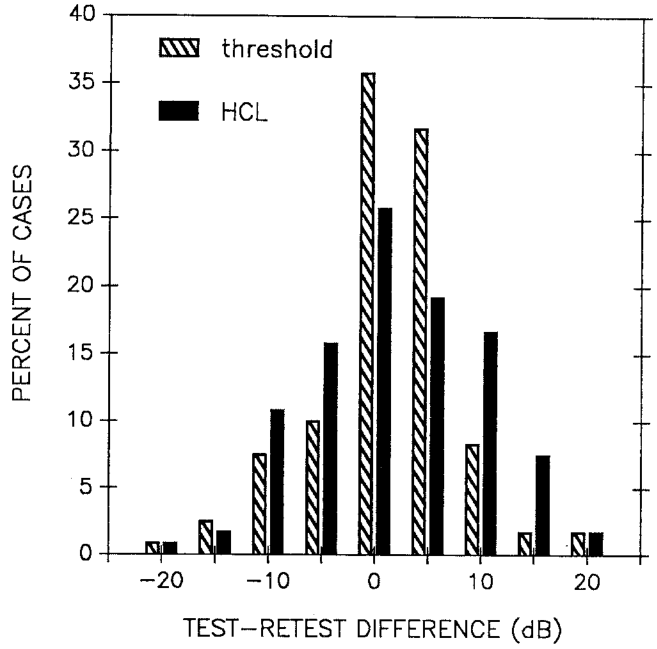

FIGURE 4. Distributions of test-retest differences for thresholds and HCLs measured in Experiment 2. N = 120.

sent data indicate that the long-term reliability of comfortable loudness levels is somewhat less than the day-to-day reliability.

Perhaps the most surprising aspect of these data is the relatively poor long-term reliability of the threshold measures. The test-retest difference *SD* of 6.7 dB found in this study is larger than that usually reported for threshold measures. For example, Arlinger and Jervall (1987) recently reported test-retest standard deviation values of 4.5–5.5 dB for pure tone thresholds under earphones at the frequencies tested in this study. Other investigators have reported similar values. Most previous evaluations of threshold reliability have differed from this study in employing pure tones, a smaller increment, supra-aural earphones, and a test-retest interval of one day or less. There is little reason to postulate that the use of warble tones, narrow band noises, a 5-dB step size, or an insert receiver would increase threshold variability because there are numerous reported studies that indicate the contrary (e.g., Byrne & Dillion, 1981; Jervall & Arlinger, 1986; Larsen, et al, 1988). Hence, the most likely reason for the larger test-retest threshold differences would appear to be related to the long time between tests, possibly including different tester criteria and small variations in equipment performance. These same factors would be expected to affect the variability of HCL measures.

These results indicate that, over the long term, comfortable loudness levels were less repeatable than thresholds. However, the difference between the two measures was not very large: 95% of threshold test-retest differences were within ±13 dB, whereas a range of ±16 dB was required to encompass 95% of the HCL test-retest differences. The HCL results are consistent with the MCL data reported by Christen and Byrne (1980) in suggesting that a long-term difference of 20 dB or more may be seen in a small proportion (2.5% in our data) of comfortable loudness test-retest measures. However, it is worth noting that test-retest differences of 20 dB were also seen in a small proportion of threshold tests.

Because thresholds and HCLs both varied from test to retest, it seemed possible that the variations in the two measures were related. This was explored by computing least-squares correlation coefficients between test-retest differences for thresholds and HCLs at each frequency. Of the six test frequencies, only the correlation at 4000 Hz was significant ($r(18) = .42$, $p < .05$, one-tailed test). This outcome suggests that at 4000 Hz, about 18% of the variation in HCLs could be attributed to variation in thresholds. At the other test frequencies, thresholds and HCLs appeared to vary quite independently.

As noted earlier, we were interested in evaluating the possible effects of daily use of amplification on comfortable loudness. Seven of the subjects in the present study obtained their first hearing aid shortly after the first test session and then wore amplification on a daily basis. The other 3 subjects were established hearing aid wearers, having worn their instruments more than 12 hr/day for 2–5 years. It was reasoned that if exposure to amplified sound resulted in an increased comfortable loudness level, the range from threshold to HCL would increase after a period of hearing aid use. Data were collapsed across frequency to explore this issue. Results indicated that, for the subgroup of new users, the mean HCL/threshold range in the first test session was 38 dB, whereas the analogous range over the second and third sessions averaged 42 dB. In the subgroup of experienced hearing aid users, the mean HCL/threshold range in the first test session was 40 dB and remained at 40 dB across the next two test sessions. Although the sample is too small to justify definite conclusions, this outcome is consistent with a hypothesis that use of amplification increases comfortable loudness levels. However, the amount of increase seen here, 4 dB on average, would not have a major impact on gain prescriptions.

EXPERIMENT 3: PREDICTABILITY OF COMFORTABLE LOUDNESS LEVELS

In spite of research evidence that supports the use of comfortable loudness data in hearing aid prescriptions, loudness-based prescription strategies are sometimes not used because the additional time required to measure comfortable loudness levels is not available. One possible tactic in this event is to predict the loudness measure using threshold data. Numerous investigators have been interested in the relationship between hearing thresholds and comfortable loudness levels (e.g., Byrne & Murray, 1985; Cox & Bissett, 1982; Kamm, Dirks, & Mickey, 1978; Shapiro, 1975). Some of these authors have concluded that comfortable loudness levels cannot be predicted with acceptable accuracy from thresholds. However, most authors have not reported regression analyses for the two measures.

Cox (1988) investigated the relationship between thresholds and HCLs for ⅓-octave noise bands for 45 listeners with sensorineural hearing impairment. Regression equations to predict HCLs from thresholds were generated. It was noted that the standard deviation of the differences between predicted and actual HCLs was 7–8 dB. Thus, about 20% of the predicted HCLs would be more than 10 dB different from the HCL that would have been obtained by direct measurement.

To some extent, the apparent inaccuracy of predictions of comfortable loudness levels from thresholds is due to the lack of perfect reliability of the threshold and loudness measurements on which the regression equations are calculated. It is possible to estimate the underlying relationship between threshold and loudness measures by applying the correction for attenuation (Nunnally, 1978, p. 237). For this correction, the correlation coefficient empirically derived between threshold and loudness variables is divided by the square root of the product of the reliability coefficients of each variable. The result is the correlation coefficient that would be expected if threshold and loudness measures were each perfectly repeatable.

Table 5 gives the corrected correlation coefficients between threshold and HCL measures at several frequencies. The necessary data were drawn from several sources. The reliability coefficients for threshold measures were estimated from the data for 27 subjects reported by Jerger (1962). Reliability coefficients for HCL measures were derived from data for 16 subjects collected in the study reported by Cox and Bisset (1982). The correlation coefficient between threshold and HCL measures was computed using the data from 45 subjects reported by Cox (1988). All these investigations used subject groups with sensorineural, mild to moderately severe hearing loss. The squared, corrected correlation coefficient reported for each frequency in Table 5 indicates the proportion of the variance in HCL measures that can be attributed to variation in threshold measures, excluding measurement error. It is noteworthy that this value is less than 50% even after the lack of reliability of threshold and HCL data has been accounted for. This outcome indicates that comfortable loudness levels include a large component that cannot be predicted from hearing loss alone.

As a result of these considerations, an investigation was undertaken to develop a more accurate method of predicting comfortable loudness levels. Analysis of HCL residuals (differences between directly measured HCLs and HCLs predicted from thresholds) for hearing-impaired subjects revealed that the correlations between residuals at different frequencies were rather high. This indicates that if an individual's comfortable loudness levels are higher than average at one frequency, they are likely to be higher than average at other frequencies also. It was hypothesized that this fact could be exploited to permit more accurate predictions of comfort levels at a particular frequency by combining threshold at that frequency with one or more comfort level residuals from other frequencies. This procedure would still require measurement of comfort levels at some frequencies but would obviate the need to measure them at every test frequency.

METHOD

Subjects. Group 1 comprised 67 adult subjects yielding 70 test ears with sensorineural hearing loss. This group included the 45 subjects whose data were analyzed and reported by Cox (1988). Table 6 gives the means and standard deviations of thresholds in the test ears. Group 2 was composed of 25 different adults with sensorineural hearing loss. Table 7 gives the means and standard deviations of thresholds in the test ears for this group.

TABLE 5. Corrected correlation coefficients between threshold and HCL measures at several frequencies: r_{tt} = threshold reliability coefficient; r_{uu} = HCL reliability coefficient; r_{tu} = measured correlation between the threshold and HCL; r_{tu}' = corrected correlation between the threshold and HCL.

Freq (Hz)	r_{tt}	r_{uu}	r_{tu}	r_{tu}'	$(r_{tu}')^2$
500	.93	.71	.36	.44	.20
1000	.93	.78	.59	.69	.48
2500	.95	.85	.61	.68	.46
4000	.95	.88	.64	.70	.49

TABLE 6. Means and standard deviations *(SD)* of thresholds in the test ears of Group 1 subjects (N = 70) in Experiment 3. Data are in dB SPL.

Freq (Hz)	*Mean*	*SD*
250	56	13.4
500	50	15.3
800	51	15.5
1000	51	15.1
1600	55	17.7
2500	62	16.7
4000	62	16.3

Stimuli. Nominal test frequencies were 250, 500, 800, 1000, 1600, 2500, and 4000 Hz. Stimuli were either ⅓-octave noise bands, −5% warble tones with rectangular modulation, or ±5% warble tones with sinusoidal modulation. Stimuli were delivered using either a supra-aural earphone (Telephonics TDH-39 or TDH-49 in MX-41/AR cushion) calibrated in an NBS-9A 6 cm^3 coupler; or using an insert earphone (Danavox SMW) attached to an earmold and calibrated in an HA-2 2 cm^3 coupler. For each subject, both factors were held constant for all data collection.

Procedure. HCLs and thresholds were recorded retrospectively from clinical records for each subject in Group 1 at each test frequency. These data were subjected to simple regression analysis to produce equations for prediction of HCL from threshold at each frequency. Next, the regression equations were used to generate HCL residuals (difference between predicted and measured HCL) for each Group 1 subject at each frequency.

Multiple regression analyses were then performed for each test frequency to examine the relationship between measured HCL and combinations of threshold at the test frequency with residuals at one or more other frequencies. The overall aim was to develop a method of predicting HCL for which the confidence interval on the predicted value was as small as the confidence interval on a single direct measurement of HCL.

Finally, thresholds and HCLs were determined at each test frequency for the Group 2 subjects. These data were

TABLE 7. Means and standard deviations *(SD)* of thresholds in the test ears of Group 2 subjects (N = 25) in Experiment 3. Data are in dB SPL.

Freq (Hz)	*Mean*	*SD*
250	55	17.8
500	47	17.1
800	48	15.1
1000	50	14.7
1600	55	15.0
2500	66	13.8
4000	75	13.8

used to test the prediction equations developed from the Group 1 data. Measured HCL was compared with predicted HCL at each frequency for this group.

RESULTS AND DISCUSSION

Table 8 gives the simple regression equations derived to predict HCLs from thresholds for the subjects in Group 1. The table also shows the standard error of estimate associated with each equation. It is noteworthy that the standard errors of estimate for this group were in the range 8.5–11.0 dB and were, therefore, larger than the 7–8 dB found for the subgroup of 45 subjects reported by Cox (1988). One possible explanation for this increase might be the inclusion of warble tone and ⅓-octave noise band data in the same analysis (the earlier analysis focused on noise band stimuli exclusively). In any event, results from this larger group present an even worse picture of the accuracy of predictions of comfortable loudness levels from thresholds.

The multiple regression analyses indicated that predictions of HCL were not substantially improved when thresholds were combined with an HCL residual from any other single frequency. However, predictions of HCL were greatly improved when HCL residuals at 500 and 4000 Hz were both included in the equation. Table 9 gives the equations for predicting the HCLs at other test frequencies, using HCL residuals from 500 and 4000 Hz. Determination of HCLs using this procedure would require (a) measurement of thresholds at each test frequency; (b) direct measurement of HCLs at 500 and 4000 Hz; (c) derivation of the residuals at these frequencies by comparing the measured HCL with the predicted HCL from the appropriate equations in Table 8; and (d) computation of predicted HCLs at the other five test frequencies using the equations in Table 9.

The standard errors of estimate for this method of predicting HCLs are also shown in Table 9. At most frequencies, they are half (or less) of the standard errors associated with the equations in Table 8. In addition, this table gives the squared correlation coefficient at each frequency, indicating the proportion of the variance in measured HCLs that can be attributed to variance in thresholds at the test frequency plus variance in HCL residuals at 500 and 4000 Hz for that subject. Comparison with the analogous statistic in Table 5 confirms that the accuracy of predicted HCLs is dramatically improved when information about the residuals at 500 and 4000 Hz is included in the procedure.

TABLE 8. Regression equation and standard error of estimate (S_e) for prediction of HCL (dB SPL) from threshold (dB SPL) at each test frequency for subjects in Group 1, Experiment 3. Thr. = threshold.

Freq (Hz)	*Equation*	S_e *(dB)*
250	HCL=.26(Thr.)+92	10.0
500	HCL=.19(Thr.)+87.5	9.6
800	HCL=.26(Thr.)+83	8.6
1000	HCL=.31(Thr.)+78	9.9
1600	HCL=.43(Thr.)+69	10.8
2500	HCL=.46(Thr.)+66	10.2
4000	HCL=.51(Thr.)+62	9.9

TABLE 9. Regression equation and standard error of estimate (S_e) for prediction of HCL (dB SPL) from threshold at the test frequency and residuals for HCLs at 500 Hz and 4000 Hz. Thr. = threshold at test frequency (dB SPL); $R_{0.5}$ = HCL residual at 500 Hz (dB); $R_{4.0}$ = HCL residual at 4000 Hz (dB).

Freq (Hz)	*Equation*	S_e *(dB)*	r^2
250	HCL = .26(Thr.) + .9($R_{0.5}$) − .08($R_{4.0}$) + 92	5.7	.74
800	HCL = .28(Thr.) + .56($R_{0.5}$) + .25($R_{4.0}$) + 81	4.4	.80
1000	HCL = .26(Thr.) + .56($R_{0.5}$) + .39($R_{4.0}$) + 80	5.0	.80
1600	HCL = .41(Thr.) + .19($R_{0.5}$) + .84($R_{4.0}$) + 70	4.6	.88
2500	HCL = .43(Thr.) + .1($R_{0.5}$) + .86($R_{4.0}$) + 68	4.1	.89

As noted above, the overall aim in devising this new procedure for predicting HCLs was to develop a predictive scheme for which the confidence interval on the predicted value was as small as the confidence interval on a single direct measurement of HCL. Confidence intervals on HCL measures were estimated using the data reported by Cox and Bisset (1982). In this investigation of hearing-impaired subjects, five separate measures of HCL were obtained, separated in time by intervals ranging from 1 day to 1 month. The standard deviation of repeated HCLs was determined for each subject. To compute the confidence intervals, the within-subject standard deviations of repeated HCLs were determined for the group as a whole (square root of mean variance) at 500, 1000, and 2000 Hz. Ninety-five percent confidence intervals (1.96 × *SD*) for a single measure of HCL on a typical subject, were determined to be ±10.8, ±9.0, and ±8.0 dB at 500, 1000, and 2000 Hz, respectively. Analogous confidence intervals may be constructed for predictions of HCLs using the standard errors of estimate reported in Table 9. These 95% confidence intervals (1.96 $\times S_e$) were ±11.2, ±8.6, ±9.8, ±9.0, and ±8.0 dB at 250, 800, 1000, 1600, and 2500 Hz, respectively. Thus, the accuracy of a predicted HCL using one of the equations in Table 9 was essentially equal to the accuracy of a single measured HCL in the same frequency region.

Testing the equations. The regression equations given in Table 9 for predicting HCLs were optimized for the 70 test ears in Group 1. A second group of subjects was used to test the application of these equations to other hearing-impaired persons. Figure 5 illustrates the comparison of measured HCLs and HCLs predicted using the equations in Table 9 for the 25 subjects in Group 2. At each frequency, the diagonal lines indicate the 95% confidence interval for the predicted HCL. Confidence intervals were determined on the assumption that the actual HCL should be equal to the predicted HCL ±1.96(S_e) from Table 9. In a group of 25, 1–2 subjects can be expected to fall outside this confidence interval. Inspection of the five panels in Figure 5 reveals that at four frequencies 2–3 observations actually fell outside the confidence interval. This result suggests that, at these

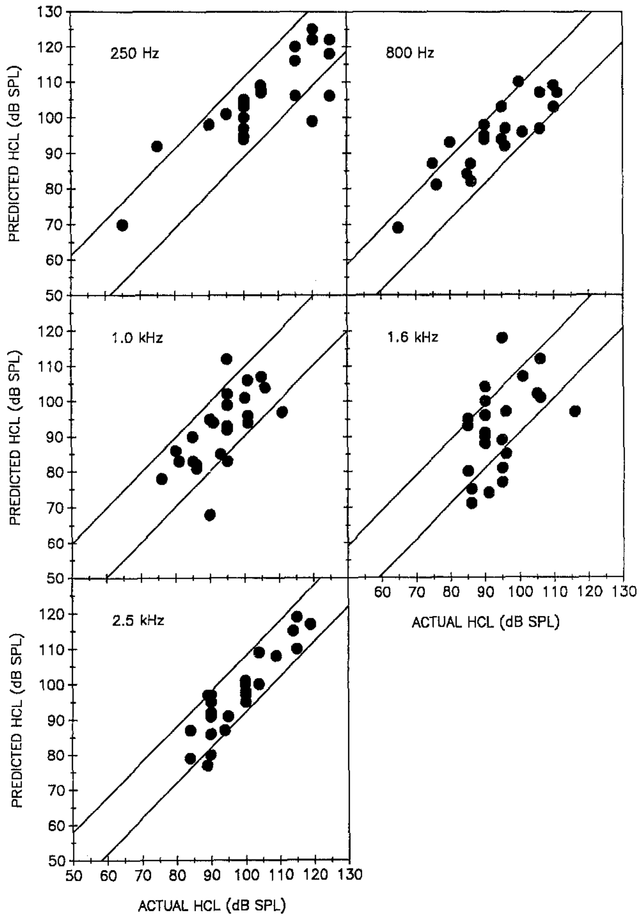

FIGURE 5. Actual (measured) and predicted HCLs at five frequencies, determined in Experiment 3. Diagonal lines depict the 95% confidence interval for the predicted HCL levels.

frequencies, the equations in Table 9 gave rather accurate predictions of HCLs.

Further examination of Figure 5 reveals that the accuracy of predicted HCLs at 1600 Hz was substantially poorer than at other frequencies. It is tempting to suggest that this outcome was observed because 1600 Hz is relatively distant on the frequency scale from both 500 and 4000 Hz and, thus, HCL residuals at those frequencies were not very germane to HCL prediction at 1600 Hz. However, this explanation is not consistent with the results obtained for the subjects in Group 1. Alternatively, this outcome may have been the result of sampling error. Further study with additional subjects will be necessary to resolve this issue.

GENERAL DISCUSSION: IMPLICATIONS FOR HEARING AID PRESCRIPTION PROCEDURES

The results of the studies reported here permit some recommendations regarding application of comfortable loudness measurement in hearing aid prescription procedures.

Because speech-based stimuli are not readily available, it is desirable to identify an alternate (more available) stimulus for which the comfortable loudness level is the same as for a corresponding continuous speech band. However, as the results of Experiment 1 suggest, a simple substitution of another stimulus with similar bandwidth, duration, etc., is not likely to produce this result. Experiment 1a indicated that when pulse duration was equated, stimuli with similar long-term spectra did not necessarily yield equal comfort levels; speech bands and noise bands had very similar long-term spectra but had different HCL long-term RMS levels. Also, rectangularly and sinusoidally modulated warble tones had rather different long-term spectra but the same HCL long-term RMS levels. Furthermore, the different stimuli were not equated for loudness comfort when adjusted to equal 20-ms peak levels. In fact, the 20-ms distribution level at which comfort levels for different stimuli were approximately equal seemed to change with frequency, approaching the peak level as frequency increased. This is consistent with a hypothesis that stimulus peak levels have a greater impact on loudness comfort at high frequencies than at low frequencies. Overall, these results indicated that comfortable loudness levels for noise bands and warble tones underestimate comfortable loudness for equal-duration speech bands regardless of whether calibration is performed in terms of overall levels or 20-ms peak levels.

When stimulus duration was varied in Experiment 1b, stimuli with shorter duration yielded higher mean comfort levels than did the same stimuli with longer duration.[2] This outcome suggested that the effect of duration could be used to offset the difference between comfort levels for speech bands and those for noise bands and warble tones. It was hypothesized that comfort levels for continuous speech bands could be rather accurately estimated using comfort levels for interrupted noise bands or warble tones. This hypothesis was substantiated with hearing-impaired listeners in Experiment 1c. In addition, results of this study combined with the data of Cox and Bisset (1982) suggest that, at least for hearing-impaired listeners, pulse duration of the interrupted stimulus can vary in the range from 0.2 to 1.0 s (assuming a duty cycle of about 50%) without significantly affecting the measured comfort levels. Furthermore, the outcome of Experiment 1 indicated that comfortable loudness levels are independent of warble tone bandwidth or modulation waveform even though these factors have been found to affect hearing thresholds.[3]

Overall, these results indicate that the comfortable loudness level for either a pulsed ⅓-octave noise band or

---

[2] However, the difference in comfort levels between 1200-ms and 400-ms stimuli was not statistically significant.

[3] It should be kept in mind that bandwidths of all stimuli in this study were equal to or less than normal critical bandwidths. If stimuli with supercritical bandwidths were compared, differences in comfortable loudness levels due to bandwidth could be expected to occur.

a pulsed warble tone, (sinusoidal or rectangular) could be expected to yield a fairly accurate estimate of the corresponding comfortable loudness level for a continuous speech band. The finding that noise bands and warble tones were equally good substitutes for speech bands in comfortable loudness measures suggests that perhaps other types of stimuli, such as pure tones or damped wave trains, could also be used for this purpose. However, Byrne (1986b) reported data suggesting that MCLs for pure tones were poor predictors of MCLs for speech bands. Thus, it would appear that these results cannot be assumed to apply to other types of stimuli.

The results of Experiment 2 indicate that when comfortable loudness levels are measured using standardized procedures and instructions, the long-term reliability of these levels is only minimally poorer than that of thresholds. Comparison with previous studies suggested that both measures are less reliable over the long term than over day-to-day assessments. This outcome is not consistent with the position sometimes advanced that comfortable loudness levels are too variable to be a useful basis for hearing aid prescriptions. Although comfortable loudness measurements do have certain limitations (not all listeners can do the task and the measurements are time consuming), data from the present study indicate that lack of reliability should not be a major concern.

Hearing aid prescription methods may elect to predict comfort levels from thresholds in order to conserve the time required for direct comfort level measurements. However, there is evidence that comfortable loudness levels are not determined solely, or even mostly, by threshold sensitivity (see Table 5). The specific factors (in addition to hearing thresholds) that determine comfortable loudness levels were not investigated in the studies reported here. Nevertheless, it was found in Experiment 3 that these unknown factors have quite orderly effects across frequency for a given individual. Although future work may refine the new procedure described in Experiment 3, the results of this study strongly imply that comfortable loudness levels may be predicted rather accurately if hearing thresholds at the frequencies of interest are combined with direct measurements of loudness comfort at 500 and 4000 Hz. This would result in a considerable reduction in time required to generate comfortable loudness data and would, therefore, address a major impediment to clinical application of comfort level measurements.

Finally, even though the studies reported here suggest that comfortable loudness levels for continuous speech bands can be estimated rather accurately, quickly, and with good reliability, significant questions remain regarding the merit of basing hearing aid gain prescriptions on these comfortable loudness levels. Definitive investigations are needed to determine the relationship between comfortable loudness levels and optimal frequency response for hearing-impaired persons. Unless a strong relationship can be demonstrated in several different laboratories, the value of comfortable loudness levels in hearing aid prescription will remain controversial.

ACKNOWLEDGMENTS

Several colleagues and students helped with data collection and/or analysis in these studies: Genevieve Alexander, Pamela Clasgens, Tania Houk, Jeffrey Moore, and Ronald Peck. Software was written by Robert Joyce, Joseph Matesich, and Joseph Knack.

This work was supported in part by VA Rehabilitation Research and Development funds. Also supported in part by the Center for Research Initiatives and Strategies for the Communicatively Impaired, Memphis State University.

REFERENCES

American National Standards Institute (1970). *American National Standard specification for audiometers* (ANSI S3.6–1969). New York: Author.

Arlinger, S. D., & Jervall, L. B. (1987). Reliability in warble-tone sound field audiometry. *Scandinavian Audiology, 16,* 21–27.

Barry, S. J., & Resnick, S. B. (1978). Absolute thresholds for frequency-modulated signals: Effects of rate, pattern, and percentage of modulation. *Journal of Speech and Hearing Disorders, 43,* 192–199.

Berger, K. W., Varavvas, D. M., & Vottero, D. M. (1982). Test-retest reliability of MCLs, UCLs and ARTs. *Hearing Instruments, 33*(5) 16–19.

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

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

Byrne, D., & Dillon, H. (1981). Comparative reliability of warble tone thresholds under earphones and in sound field. *Australian Journal of Audiology, 3,* 12–14.

Byrne, D., & Murray, N. (1985). Relationship of HTLs, MCLs, LDLs and psychoacoustic tuning curves to the optimal frequency response characteristics of hearing aids. *Australian Journal of Audiology, 7,* 7–16.

Christen, R., & Byrne, D. (1980). Variability of MCL measurements: Significance for hearing aid selection. *Australian Journal of Audiology, 2,* 10–18.

Cox, R. M. (1985). A structured approach to hearing aid selection. *Ear and Hearing, 6,* 226–239.

Cox, R. M. (1988). The-MSU hearing instrument prescription procedure. *Hearing Instruments, 39(1),* 6–10.

Cox, R. M., & Bisset, J. D. (1982). Prediction of aided preferred listening levels for hearing aid gain prescription. *Ear and Hearing, 3,* 66–71.

Cox, R. M., Matesich, J. S., & Moore, J. N. (1988). Distribution of short-term RMS levels in conversational speech. *Journal of the Acoustical Society of America, 84,* 1100–1104.

Ferguson, G. A. (1966) *Statistical analysis in psychology and education.* New York: McGraw-Hill.

Hawkins, D. B., Walden, B. E., Montgomery, A., & Prosek, R. A. (1987). Description and validation of an LDL procedure designed to select SSPL90. *Ear and Hearing, 8,* 162–169.

Jerger, J. (1962). Comparative evaluation of some auditory measures. *Journal of Speech and Hearing Research, 5,* 3–17.

Jervall, L., & Arlinger, S. (1986). A comparison of 2-dB and 5-dB step size in pure-tone audiometry. *Scandinavian Audiology, 15,* 51–56.

Kamm, C., Dirks, D. D., & Mickey, M. R. (1978). Effect of sensorineural hearing loss on loudness discomfort level and most comfortable loudness judgements. *Journal of Speech and Hearing Research, 21,* 668–681.

Kerrick, J. S., Nagel, D. C., & Bennett, R. L. (1968). Multiple ratings of sound stimuli. *Journal of the Acoustical Society of America, 45,* 1014–1017.

Larsen, V. D., Cooper, W. A., Talbott, R. E., Schwartz, D. M., Ahlstrom, C., & De Chiccis, A. R. (1988). Reference threshold sound pressure levels for the TDH-50 and ER-3A earphones. *Journal of the Acoustical Society of America, 84,* 46–51.

Leijon, A., Eriksson-Mangold, M., & Bech-Karlsen, A. (1984). Preferred hearing aid gain and bass-cut in relation to hearing aid fitting. *Scandinavian Audiology, 13,* 157–161.

Lippmann, R. P., Braida, L. D., & Durlach, N. I. (1981). Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss. *Journal of the Acoustical Society of America, 69,* 524–534.

Lucker, J. R., Grzybmacher, N., & Ventry, I. V. (1978). Effect of instructional set on the measurement of the comfortable loudness range. *Journal of Speech and Hearing Disorders, 43,* 168–172.

Melnick, W. (1967). Comfort level and loudness matching for continuous and interrupted signals. *Journal of Speech and Hearing Research, 10,* 99–109.

Nunnally, J. C. (1978). *Psychometric theory.* New York: McGraw-Hill.

Orchik, D. J., & Mosher, N. L. (1975). Narrow band noise audiometry: The effect of filter slope. *Journal of the American Audiology Society, 1,* 50–53.

Pluvinage, V. (1989). Clinical measurement of loudness growth. *Hearing Instruments, 40(3),* 28–34.

Rintlemann, W. R., & Carhart, R. (1964). Loudness tracking by normal hearers via Bekesy audiometer. *Journal of Speech and Hearing Research, 7,* 79–93.

Sammeth, C. A., Birman, M., & Hecox, K. E. (1989). Variability of most comfortable and uncomfortable loudness levels to speech stimuli in the hearing impaired. *Ear and Hearing, 10,* 94–100.

Shapiro, I. (1975). Prediction of most comfortable loudness levels in hearing aid evaluation. *Journal of Speech and Hearing Disorders, 40,* 434–438.

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

Skinner, M. W., Pascoe, D. P., Miller, J. D., & Popelka, G. R. (1982). Measurements to determine the optimal placement of speech energy within the listener's auditory area. In G. A. Studebaker & F. Bess (Eds.), *The Vanderbilt hearing aid report* (pp. 161–169). Upper Darby, PA: Monographs in Contemporary Audiology.

Sullivan, J. A., Levitt, H., Hwang, J., & Hennessey, A. (1988). An experimental comparison of four hearing aid prescription methods. *Ear and Hearing,* 9, 22–32.

Ventry, I. M., & Johnson, J. I. (1978). Evaluation of a clinical method for measuring comfortable loudness for speech. *Journal of Speech and Hearing Disorders, 49,* 149–159.

Received February 14, 1989
Accepted May 5, 1989

Requests for reprints should be sent to Robyn M. Cox, Memphis State University, Memphis Speech & Hearing Center, 807 Jefferson Avenue, Memphis, TN 38105.