J Am Acad Audiol 11: 570–574 (2000)

Effect of Interstimulus Interval on Subjective Categorical Loudness Judgments

Melissa N. Ruscetta*
Catherine V. Palmer*†

Abstract

The Independent Hearing Aid Fitting Forum uses the visual input/output locator algorithm and the Contour Test of Loudness Perception to achieve the goal of restoring normal loudness perception with amplification. This method presupposes that subjective categorical loudness judgments are valid and reliable when using the procedure outlined by the test developers. There is no indication in the instructions of the Contour Test of a specific duration for interstimulus interval (ISI). The effect of ISI on loudness perception is important to establish because of potential time-error effects. Time-error refers to the extent to which the judged magnitude of a second stimulus varies with the time interval by which it follows the first stimulus. Past research has indicated that judgments of a second stimulus are shifted in the direction of the intensity of the preceding stimulus and that this effect intensifies with shorter ISIs. The current experiment was designed to examine whether a change in ISI produced a change in the loudness rating of the subsequent stimulus for subjective categorical loudness judgment testing. A trend toward lower, median dB values for ratings 2 to 5 at 500 Hz in a group of subjects with normal hearing was noted when 1-sec intervals were used in comparison with longer ISIs. No trends for the effect of ISI were noted at 3000 Hz. The findings provide ISI recommendations for loudness judgment test administration.

Key Words: Contour Test of Loudness Perception, interstimulus interval, subjective categorical loudness judgments

Abbreviations: DAT = digital audiotape, IHAFF = Independent Hearing Aid Fitting Forum, ISI = interstimulus interval, KEMAR = Knowles Electronics' Mannikin for Auditory Research, VIOLA = visual input/output locator algorithm, WDRC = wide dynamic range compression

The reduced dynamic range and loss of nonlinear function of the outer hair cells of individuals with sensorineural and/or mixed hearing impairment has been one motivation for the use of compression circuitry in hearing aids. One goal of this type of circuitry may be to restore normal loudness perception. Hearing aids with wide dynamic range compression (WDRC) achieve this goal by amplifying the input sounds to intensity levels within the individual's reduced dynamic range while systematically reducing gain as the input level increases.

With the dawn of this type of circuitry came the need for fitting strategies that determine the compression kneepoint(s) and compression ratio setting(s) needed for the hearing aids to restore normal loudness perception. The Independent Hearing Aid Fitting Forum (IHAFF) began meeting in 1993 to establish a protocol for fitting compression hearing aids so that they normalize loudness perception. In order to implement the goal, a standard test for loudness perception and a fitting rule that would prescribe appropriate gain and compression characteristics using the loudness data were needed. The Contour Test of Loudness Perception (Cox et al, 1997) is recommended by IHAFF for obtaining loudness perception data. It is designed to determine loudness categorizations for pulsed warble tones of varying frequency. Individuals with hearing impairment will typically demonstrate different dynamic ranges at different frequencies. Often, the dynamic range at higher frequencies will be smaller than at lower frequencies due to a higher threshold of audibility with unchanged level of discomfort. Many

*Department of Communication Science and Disorders, University of Pittsburgh, †Department of Otolaryngology, University of Pittsburgh Medical Center, Pittsburg, Pennsylvania

Reprint requests: Melissa N. Ruscetta, University of Pittsburgh, 4033 Forbes Tower, Pittsburgh, PA 15260

WDRC hearing aids are equipped with at least two frequency channels that can be programmed with different gain and compression characteristics. For these reasons, obtaining loudness data from at least one low and one high frequency is recommended with the Contour Test. The test uses seven categories of loudness, adapted from Hawkins et al (1987). Beginning at threshold for the test frequency, the individual is asked to categorize the loudness of each tone until he or she responds that the tone is uncomfortably loud (rating 7). This procedure is repeated two more times for the test frequency, and the median intensity values are determined for each loudness category. This information then can be applied to the visual input/output locator algorithm (VIOLA) (Cox, 1995), which uses an established fitting rule (see Cox, 1995) and allows the clinician to determine the hearing aid gain, output, and compression characteristics that should most closely allow for normal categorical loudness perception.

This procedure presupposes that subjective loudness judgments are valid and reliable. Unfortunately, because the judgments associated with the Contour Test are subjective, the validity of the procedure cannot be unequivocally determined. However, it is important to adhere to consistent experimental procedures and to be aware of changes in outcome that might result from variation in parameters that are not under specific control.

Good test–retest reliability has been shown for various loudness scaling procedures, including the Contour Test (Ricketts and Bentler, 1996; Robinson and Gatehouse, 1996; Cox et al, 1997; Palmer and Lindley, 1998). Factors that may influence the reliability of the test have been examined and are controlled for through explicit procedures to be used in manual administration of the Contour Test (e.g., patient instructions, sequencing of the stimuli in terms of intensity, optimal test signal, and duration of the stimuli).

Instructions are a critical factor for reliability. They should be specific and read verbatim to each subject every time testing occurs. The instructions for the Contour Test are as follows:

> The purpose of this test is to find your judgments of the loudness of different sounds. You will hear sounds that increase and decrease in volume. You must make a judgment about how loud these sounds are. Pretend you are listening to the radio at that volume. How loud would it be? After each sound, tell me which of these categories best describes the loudness. Keep in mind that an uncomfortably loud sound is louder than you would ever choose on your radio no matter what mood you are in (Cox et al, 1997, p. 391).

Another consideration is the sequencing of the stimuli in terms of intensity level. Both Woods et al (1973) and Ventry and Johnson (1978) conducted studies that illustrate the differences between using an ascending versus descending approach. These data, in conjunction with data from Allen et al (1990) and Hellbruck et al (1995), who demonstrate the effects of random level sequencing, illustrate that a specific sequence of presentation will alter subjective loudness judgments. Note that in such studies, the category assigned to a particular stimulus tended to be similar to the category assigned to the preceding stimulus (Ward and Lockhead, 1970; Jesteadt et al, 1977). In general, descending and random approaches produce mean decibel ratings that are higher for each category than those obtained using an ascending approach. Ascending sequencing was chosen for the Contour Test because it is easily implemented manually and via computer.

Also, one must consider the influence of test signal and bandwidth on loudness perception. Data obtained by Ricketts and Bentler (1996) reveal that altering the stimulus bandwidth for both low- and high-frequency weighted stimuli causes a significant difference in categorical scaling of loudness using a nine-category rating scale (Allen et al, 1990). Wideband signals were judged as significantly louder than narrowband signals for both noise and speech stimuli. The Contour Test and resulting VIOLA fitting program require warbled pure-tone stimuli.

Duration of stimuli may have some effect on the categorization of varying intensity tones. Several investigators (Henning, 1970; Wier, 1980; Florentine, 1986) have shown that the level difference (in dB) between two tones that is just discriminable decreases as duration of the stimuli increases. Therefore, it may be easier to tell the difference between two sequential tones if those tones are longer in duration. If one is able to detect a level difference more easily, one may tend to place the tones into different loudness categories simply because they perceive them as different in intensity, without regard for the actual definition of the particular category. For the Contour Test, the warble tones are presented in groups of four 200-msec pulses with a 50 percent duty cycle. This duration was chosen because it is an average length and is similar to the pulsed stimuli that are presentable through an audiometer. Additionally, this duration is what was used to develop the corresponding fitting algorithm and must be implemented to ensure appropriate selection of hearing aid parameters.

There is no indication in the instructions of the Contour Test of what specific interstimulus interval (ISI) should be employed. Cox (1999) recommends manual administration of the Contour Test, citing that "computerized loudness testing. . .is too error-prone" (p. 18). ISI duration is not defined in this procedure. The impact that differing ISIs might have on the loudness categorization results is unclear.

The effect of ISI on loudness perception is important to establish because of potential time-error effects. Time-error refers to when the judged magnitude of the second stimulus varies with the time interval by which it follows the first. The second stimulus is compared with the first (Woodworth and Schlosberg, 1954). Needham (1934) and Postman (1946) found time-errors on the order of 2 or 3 dB at ISIs ranging from 1 to 6 seconds. In two experiments done by Elmasian and Galambos (1975) and Elmasian et al (1980), results revealed enhancement and decrement of the perceived loudness of the second stimulus depending on whether the first stimulus was louder or softer than the second stimulus. When the first stimulus was louder than the second stimulus, enhancement of the second was evidenced. Decrement of the second stimuli was noted when the first stimulus was softer than the second stimulus. These experiments allow for the following generalization: judgments of the second stimulus are shifted in the direction of the intensity of the preceding stimulus, and this effect intensifies with shorter ISIs.

The goal of this study was to determine if presentation gap length produces a change in the loudness rating of the second stimulus, if the change is frequency dependent, and if one of four ISIs presented in this study should be recommended for clinical use based on the results.

METHODS

Subjects

Eighteen adult subjects, naive with respect to loudness perception testing, presented with normal hearing bilaterally (defined as no threshold greater than 20 dB HL re ANSI, 1969) across frequencies (250–8000 Hz). Testing was performed using a Virtual model 320 audiometer (ANSI, 1989) with the subject sitting in a single-walled Industrial Acoustics Company sound booth.

Stimuli

Three steps were employed to generate the final experimental stimuli. In step 1, a Panasonic professional digital audiotape (DAT) deck (Model SV 3700) was used to record 500 and 3000 Hz warbled, pulsed tones of 200-msec length and with a 50 percent duty cycle. The stimuli were generated by a Virtual Model 320 audiometer and routed through Etymotic ER-3A insert earphones. The earphones were placed in the ear canal of a Knowles Electronics' Mannikin for Auditory Research (KEMAR) seated in a single-walled Industrial Acoustics Company sound booth. Recordings were made through a Zwislocki coupler that was attached to an ER-11 ½-inch microphone preamplifier (Etymotic Research, Inc.) and placed at the end of KEMAR's ear canal. From the preamplifier, the signal was routed to a QSC professional stereo amplifier (Model 1200) and then to the DAT recorder. The signals were recorded at each frequency beginning at a low level and at ascending 5-dB steps to the output limit of the audiometer. In step 2, a portable Panasonic DAT recorder (Model SV 250) was connected to the Model SV 3700 DAT recorder for dubbing the stimuli with the specified ISI. Eight experimental digital tapes were created. Each frequency (500 and 3000 Hz) was recorded with 1-, 2-, 3-, and 6-sec gaps placed between each 5-dB step. For step 3, the decibel levels of the signals were measured. The Model SV 3700 DAT recorder served as auxiliary input to the Virtual audiometer (Model 320). The signals were again routed through the Etymotic ER-3A insert earphones seated in KEMAR's ear. Also seated in KEMAR's ear was a probe microphone connected to a Virtual model 340 real-ear system. The real-ear system allows digital readings of the sound pressure level present in the ear canal. The limitations of the playback equipment dictated a finite range of intensities that were available for presentation. We felt that it was important to ensure that the lowest (and first) signal presented elicit a rating of category one. As noted earlier, presenting signals in a random sequence as opposed to in an ascending fashion beginning with a very soft sound alters the loudness judgments. The audiometer amplifier was manipulated until sound pressure levels for the 500- and 3000-Hz signals were from 20 to 75 dB SPL.

Procedures

Each subject was read the instructions for the Contour Test (Cox et al, 1997). Additional instructions included that the subject should answer as quickly as possible after the stimulus had ended and that if he or she did not answer before the next presentation to skip that presentation and respond to the subsequent one. Each subject also received a copy of the seven categories of loudness to look at throughout testing.

There are 24 possibilities for the order in which four ISIs can be presented. One order was assigned randomly to each of the 18 subjects from the 24 possibilities. Order of frequency was counterbalanced across the subjects. To familiarize subjects with the task prior to the start of data collection, subjects rated a 1000-Hz signal with no specified ISIs.

All contour testing was administered according to the procedure outlined by Cox et al (1997) in the right ear only. The opposite ear was left unoccluded. Stimuli were presented through an Etymotic ER-3A insert earphone to each subject. The subjects were seated in a single-walled Industrial Acoustics Company sound booth. Their loudness ratings were recorded for each presentation level. For each subject, the SPL decibel level corresponding to his or her median response for each category, after three runs at each ISI, was calculated. The range of intensities used did not produce ratings from one through seven for every individual. All individuals classified the most intense sound presented as at least a "four (comfortable)."

RESULTS

The means of the median dB values obtained for each category across all 18 subjects were calculated at each ISI. Four means (one from each ISI) for each of the seven categories were compared using a one-way repeated-measure ANOVA. This procedure was applied to the low- and high-frequency results separately. Table 1 provides these mean data and standard deviations.

Results of the ANOVA were significant ($p < .05$) for loudness categories two through five at 500 Hz and for loudness categories one and three at 3000 Hz. A Tukey, critical difference, post hoc analysis showed that mean values for categories two and three are significantly lower ($p = .005$ and $.0001$, respectively) at the 1-sec ISI than at any of the other ISIs for 500 Hz. For categories four and five of the 500-Hz signal, mean values at the 1-sec ISI were significantly lower than at both 2- and 3-sec ISIs ($p = .001$ and $.027$, respectively). At 3000 Hz, mean values for 2- and 6-sec ISIs were significantly different from one another for category three ($p = .042$), and mean values for 1- and 2-sec ISIs were significantly different from mean values for 3- and 6-sec ISIs for category one ($p = .040$). However, no trends were indicated at this higher frequency.

Table 1 Average, Median dB SPL Values and (SDs) for 500 and 3000 Hz, at the Four ISIs

Category	*1 sec*	*2 sec*	*3 sec*	*6 sec*
500 Hz				
1	25.6 (3.8)	26.4 (4.1)	26.7 (4.2)	26.9 (3.9)
2	36.1 (5.6)	39.2 (6.5)	38.9 (7.2)	38.6 (4.5)
3	42.4 (4.1)	46.1 (6.3)	46.3 (5.4)	45.1 (4.7)
4	49.3 (6.4)	53.3 (7.1)	51.9 (6.7)	51.4 (5.4)
5	54.1 (8.0)	57.5 (6.6)	57.5 (8.0)	57.0 (6.9)
6	60.9 (5.8)	62.0 (6.0)	61.1 (6.6)	63.0 (6.0)
7	61.7 (8.2)	65.8 (6.6)	65.0 (5.8)	65.7 (6.7)
3000 Hz				
1	27.8 (3.5)	27.5 (3.5)	26.7 (3.4)	26.7 (3.8)
2	41.1 (6.3)	41.4 (5.9)	40.6 (6.8)	39.4 (6.6)
3	46.7 (5.9)	47.5 (6.0)	46.9 (7.3)	45.0 (6.6)
4	53.6 (7.6)	54.4 (8.9)	54.7 (8.1)	53.3 (8.9)
5	60.0 (9.8)	60.6 (9.5)	60.6 (9.5)	60.0 (8.8)
6	63.3 (6.2)	62.8 (7.8)	61.7 (7.8)	62.2 (7.0)
7	66.7 (7.5)	62.5 (8.7)	63.0 (5.7)	66.3 (6.4)

DISCUSSION

The dB SPL ratings for soft sounds in this group of subjects are similar to those of Cox et al (1997). Mean dB SPL ratings for category 7 ratings are lower for this group of individuals than for the Cox et al (1997) group. This difference in actual dB SPL levels did not impact the current intent of this study and may be explained by the fact that no reinstruction was offered (encouraging individuals to wait until sound was truly uncomfortable before using the #7 rating) during the procedure. Figure 1 illustrates the difference in value of the mean decibel values across the four ISIs at 500 Hz for categories that revealed significant differences in post hoc analysis. A trend is noted toward a significantly lower decibel value for the category rating when stimuli were presented 1 sec apart as opposed to when presentations were further apart for the 500-Hz data. These data suggest that, for a group of normal-hearing individuals, presenting signals at 1-sec or smaller intervals may lower the individual's decibel rating for categories 2 to 5 by as much as 4 dB SPL for

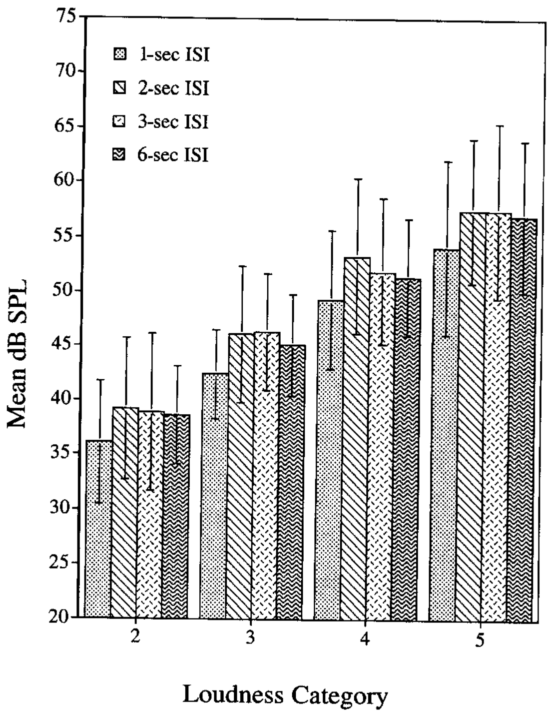

Figure 1 Mean dB SPL values and SDs across ISIs for categories 2 to 5 at 500 Hz.

low-frequency stimuli. Although it might be interesting to know what happens at even shorter ISIs than those tested here, it is not plausible to think that one might manually administer the Contour Test at rates faster than 1 sec or that a patient could keep up with the task. Also, the data do not support the notion that ISIs longer than 6 sec will produce a significant change, nor is it in the interest of clinical time management to allow for ISIs that are longer than 6 seconds. To summarize, when administering the Contour Test manually, 2-sec ISIs may allow the clinician to avoid time-error effects in the patient's responses, particularly when testing low frequencies. Future research must include individuals with varying degrees of hearing loss in order to determine if the trends noted for normal-hearing individuals are maintained.

Acknowledgment. This article was presented at the 12th Annual Convention of the American Academy of Audiology, Chicago, March 2000.

REFERENCES

Allen JB, Hall JL, Jeng PS. (1990). Loudness growth in ½ octave bands (LGOB): a procedure for the assessment of loudness. *J Acoust Soc Am* 88:745–753.

American National Standards Institute. (1969). *American National Standard for Audiometry.* (ANSI S3.6—1969). New York: ANSI.

American National Standards Institute. (1989). *American National Standard Specifications for Audiometers.* (ANSI S3.6—1989). New York: ANSI.

Cox RM. (1995). Using loudness data for hearing aid selection: the IHAFF approach. *Hear J* 48:10, 44.

Cox RM. (1999). Five years later: an update on the IHAFF fitting protocol. *Hear J* 52:10–18.

Cox RM, Alexander GC, Taylor IM, Gray GA. (1997). The Contour Test of Loudness Perception. *Ear Hear* 18:388–400.

Elmasian R, Galambos R. (1975). Loudness enhancement: monaural, binaural, and dichotic. *J Acoust Soc Am* 58:229–234.

Elmasian R, Galambos R, Bernheim A. (1980). Loudness enhancement and decrement in four paradigms. *J Acoust Soc Am* 67:601–607.

Florentine M. (1986). Level discrimination of tones as a function of duration. *J Acoust Soc Am* 79:792–797.

Hawkins DB, Walden BE, Montgomery AA, Prosek RA. (1987). Description and validation of an LDL procedure designed to select SSPL90. *Ear Hear* 8:162–169.

Hellbruck J, Thomamuller D, Zeitler A. (1995). Hearing field audiometry with groups of normal hearing persons in natural sound fields. *Audiol Akustik* 34:60–81.

Henning GB. (1970). A comparison of the effects of signal duration on frequency and amplitude discrimination. In: Plomp R, Smoorenburg GF, eds. *Frequency Analysis and Periodicity Detection in Hearing.* Sitjthoff, Leiden, The Netherlands.

Jesteadt W, Luce RD, Green DM. (1977). Sequential effects in judgment of loudness. *J Exp Psychol Hum Percept Perform* 3:92–104.

Needham JG. (1934). The time-error as a function of continued experimentation. *Am J Psychol* 46:558–567.

Palmer CV, Lindley GA. (1998). Reliability of the Contour Test in a population of adults with hearing loss. *J Am Acad Audiol* 9:209–215.

Postman L. (1946). The time-error in auditory perception. *Am J Psychol* 56:193–219.

Ricketts TA, Bentler RA. (1996). The effect of test signal type and bandwidth on the categorical scaling of loudness. *J Acoust Soc Am* 99:2281–2287.

Robinson K, Gatehouse S. (1996). Test–retest reliability of loudness scaling. *Ear Hear* 17:120–123.

Ventry IM, Johnson JI. (1978). Evaluation of a clinical method for measuring comfortable loudness for speech. *J Speech Hear Disord* 43:149–159.

Ward LM, Lockhead GR. (1970). Sequential effects and memory in category judgment. *J Exp Psychol* 84:27–34.

Wier C. (1980). Pure-tone intensity discrimination as a function of frequency, level and duration. *J Acoust Soc Am* 67(Suppl 1):S22.

Woods RW, Ventry IM, Gatling LW. (1973). Effect of ascending and descending measurement methods on comfortable loudness levels for pure tones. *J Acoust Soc Am* 54:205–206.

Woodworth RS, Schlosberg H. (1954). *Experimental Psychology.* Rev Ed. New York: Holt, Rinehart, and Winston.