Perceptual weights for loudness judgments of six-tone complexes[a]

Walt Jesteadt,[b] Daniel L. Valente,[c] and Suyash N. Joshi[d]
Center for Hearing Research, Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

Kendra K. Schmid
Department of Biostatistics, College of Public Health, University of Nebraska Medical Center, Omaha, Nebraska 68198

(Received 18 October 2013; revised 22 June 2014; accepted 26 June 2014)

Subjects with normal hearing (NH) and with sensorineural hearing loss (SNHL) judged the overall loudness of six-tone complexes comprised of octave frequencies from 0.25 to 8 kHz. The level of each tone was selected from a normal distribution with a standard deviation of 5 dB, and subjects judged which of two complexes was louder. Overall level varied across conditions. In the "loudness" task, there was no difference in mean level across the two stimuli. In the "sample discrimination" task, the two complexes differed by an average of 5 dB. For both tasks, perceptual weights were derived by correlating the differences in level between matched-frequency tones in the complexes and the loudness decision on each trial. Weights obtained in the two tasks showed similar shifts from low to high frequency components with increasing overall level. Simulation of these experiments using a model of loudness perception [Moore and Glasberg (2004), Hear Res. **188**, 70–88] yielded predicted weights for these stimuli that were highly correlated with predicted specific loudness, but not with the observed weights. © *2014 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4887478]

PACS number(s): 43.66.Cb, 43.66.Fe, 43.66.Ba [FJG] Pages: 728–735

I. INTRODUCTION

Leibold *et al.* (2007) asked listeners to compare two narrowband sounds consisting of five logarithmically spaced sinusoidal components and judge which was louder. In one sound, the level of each component varied randomly from trial to trial while in the others, all components were equal in level. The correlation between the level of each variable component and a listener's decisions provided a measure of the relative contribution or perceptual weight of that component in the loudness judgment (Lutfi, 1995; Richards and Zu, 1994). A similar approach has been used in a wide range of studies (e.g., Doherty and Lutfi, 1996; Kortekaas *et al.*, 2003; Leibold *et al.*, 2009; Lentz, 2007; Willihnganz *et al.*, 1997), but all of these studies used a sample discrimination paradigm where the levels of components in both stimuli were selected at random from one of two distributions that differed in overall level. Subjects in these studies were asked to judge which interval contained components drawn from the distribution that was higher in level rather than which interval was louder.

Doherty and Lutfi (1996) obtained perceptual weights for sample discrimination using six-tone complexes with octave frequencies (250–8000 Hz) that had a mean difference in level of 5 dB. They reported that subjects with sensorineural hearing loss (SNHL) placed greater weight on higher-frequency components that were in the region of greater hearing loss and speculated that listeners with SNHL pay more attention to the information within the region of their hearing loss to compensate for the degraded sensory information in those regions. The levels used to test SNHL subjects, however, were higher than the overall levels used for subjects with normal hearing (NH). Leibold *et al.* (2009) used high- and low-pass noises to elevate thresholds for NH subjects and tested them at multiple levels in the task used by Doherty and Lutfi (1996). Leibold *et al.* found greater weight associated with higher frequency stimuli as overall level increased. Results reported by Kortekaas *et al.* (2003) and Lentz (2007) also pointed to a level-by-frequency interaction. Together these results suggest that differences in level rather than degree of hearing loss may have accounted for the differences in perceptual weight obtained by Doherty and Lutfi (1996) at higher frequencies.

Oberfeld *et al.* (2012) obtained perceptual weights for both time and frequency using stimuli made up of three noise bands divided into ten temporal segments. Their listeners were tested in a single-interval sample discrimination task in which they indicated their confidence that a loud or soft sound had been presented. The listeners assigned greater weight to the lowest frequency band and to the earlier segments.

Leibold *et al.* (2007) demonstrated that measures of perceptual weight could be used to explore the properties of loudness summation and that the model of loudness proposed by Moore *et al.* (1997) could be used to predict

---

[a] A portion of this work was presented at the 2011 Annual Midwinter Meeting of the Association of Research in Otolaryngology.

[b] Author to whom correspondence should be addressed. Electronic mail: walt.jesteadt@boystown.org

[c] Present address: JMP Division, SAS Institute, Inc., 100 SAS Campus Drive, Cary, NC 27513.

[d] Present address: Oticon Center of Excellence for Hearing and Speech Sciences, Technical University of Denmark, Ørsteds Plads Building 352, 2800 Kongens Lyngby, Denmark.

perceptual weights in loudness-summation conditions (Leibold and Jesteadt, 2007). Like other models of loudness (Chalupper and Fastl, 2002; Zwicker and Scharf, 1965; Stevens, 1961), Moore *et al.* (1997) assumed a multi-stage process that includes linear filtering by the outer and middle ear, transformation of the spectrum to an excitation pattern reflecting potential masking of some spectral components by others, transformation of excitation to specific loudness reflecting growth of loudness as a function of level at each frequency, and summation of the specific loudness across frequency bands and across ears. There were, however, a number of unresolved issues. The existing loudness models account for most effects of level, frequency, and bandwidth in the literature, but Leibold *et al.* (2007) found that the Moore *et al.* (1997) loudness model failed to predict the pattern of perceptual weights in their widest bandwidth condition, where five sinusoidal components were spread over a 2119-Hz bandwidth. The model also failed to predict the level effects observed by Kortekaas *et al.* (2003), Lentz (2007) and Leibold *et al.* (2009) using wider band stimuli. All of those studies used a sample discrimination task rather than a loudness task. The differences between the two tasks are minor, but given the literature on the complex relation between the slope of the loudness function and intensity resolution (e.g., Schlauch and Wier, 1987; Zwislocki and Jordan, 1986), there is some doubt as to whether the two tasks yield equivalent information.

Results obtained with multi-tone complexes suggested that measures of perceptual weight obtained with these stimuli might be used to determine the relative contributions of different frequency regions to the overall loudness of broadband sounds. This would be far more efficient than using loudness matching to assess the increase in loudness associated with the addition of individual bands of noise to an existing noise stimulus (Pollack, 1951, 1952) and might yield different and more clinically relevant results than measuring loudness growth at individual frequencies in isolation (e.g., Cox *et al.*, 1997).

Our goal, therefore, was to assess the perceptual weight associated with individual components of a broadband stimulus over a range of levels in a task where listeners were instructed to judge loudness rather than to discriminate between samples differing in average level and to compare the results to those predicted by the model of loudness proposed by Moore *et al.* (1997). Because we were using stimuli similar to those used by Doherty and Lutfi (1996), we extended the study to include a small group of listeners with hearing loss, allowing us to compare data for listeners with normal hearing and hearing loss at the same levels.

II. EXPERIMENT 1: EFFECTS OF LEVEL ON PERCEPTUAL WEIGHTS FOR LOUDNESS

A. Listeners

Eight NH listeners and four SNHL listeners were paid to participate. The NH listeners (ages 21–31) had hearing threshold $\leq$20 dB hearing level (HL) at the audiometric frequencies. The SNHL listeners (ages 38, 20, 57, and 67 yr with respect to the order in which results are reported in the following text) had been seen previously in our clinic at the Boys Town National Research Hospital (BTNRH). The losses were diagnosed as sensorineural based on the absence of differences between air and bone conduction and the lack of retrocochlear signs.

B. Procedure

Stimuli were generated digitally in MATLAB at a sampling rate of 44.1 kHz and were presented to subjects via 24-bit digital-to-analog convertors (DAC) (Digital Audio Labs, CardDeluxe). A remote passive attenuator coupled the outputs of the DAC with Sennheiser HD250 Linear II headphones. Stimuli were delivered to the left ear of the subject in a double walled sound-attenuated booth. The same equipment was used in all experiments.

Quiet thresholds were obtained at octave frequencies from 250 to 8000 Hz using a standard two-interval forced-choice adaptive procedure with a decision rule that estimated 71% correct (Levitt, 1971). Tones were presented for 300 ms with 10-ms cos2 ramps. The initial step size of 8 dB was reduced to 2 dB after four reversals. Thresholds were averaged across two 50-trial blocks. Mean thresholds for NH listeners and individual thresholds for SNHL listeners are shown in Fig. 1.

In the main conditions, NH listeners were presented with two six-tone complexes made up of octave frequencies from 250 to 8000 Hz. The tone complexes had a total duration of 300 ms with 10-ms cos^2 ramps. An individual trial consisted of a 300-ms warning interval, two 300- ms observation intervals separated by 500 ms, and an answer interval. Listeners were given visual markers for the warning, observation and answer intervals in a message window on a keypad that they used to indicate which interval contained the louder sound. For a given presentation, the level of each of the six tones was selected from a Gaussian distribution with a mean of X dB and a standard deviation of 5 dB. X was fixed for a block of 100 trials at 45, 55, 65, or 75 dB sound pressure level (SPL). The levels were not corrected for

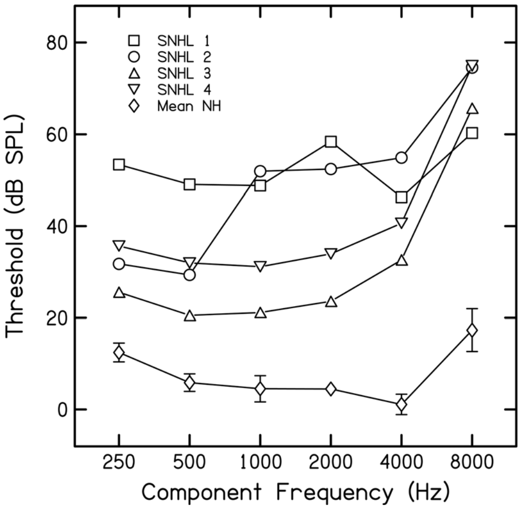

FIG. 1. Mean absolute threshold for six listeners with normal hearing (NH) and individual thresholds for those with hearing loss in experiment 1. The error bars for NH listeners here and in all later figures indicate ±1 standard error.

deviations from a flat headphone response in the equal-SPL conditions. As a result, the actual levels of the 250 Hz tones were 6.8 dB below the nominal level, while those at 8000 Hz were 2.9 dB high.

In this loudness task, all stimuli in a given block of trials were generated from the same distribution, and there was no difference on average across components or between the two complexes presented on a given trial. Listeners made 1000 decisions per condition in blocks of 100 trials. Five blocks were run at each level in ascending order, then five more blocks were run in descending order. In a second set of conditions, the levels at each frequency were adjusted based on the quiet threshold at that frequency, and the mean level of the tones for NH listeners was set to 30, 40, 50, or 60 dB sensation level (SL). These levels were run in ascending order with 1000 trials per level.

The levels for the SNHL listeners were adjusted to a range of 55–85 dB SPL and 20–50 dB SL to allow more levels to be tested without use of components that were below threshold in the SPL conditions or were at levels greater than 100 dB SPL in the equal-SL conditions. Each SNHL listener was tested in a subset of the SPL and SL conditions selected on the basis of her or his audiogram. Only three of the four SNHL listeners were available for testing in the equal-SL conditions. All four SNHL listeners were tested at 75 dB SPL, and three were tested at 20 dB SL. Only those data were included in the statistical analyses.

Perceptual weights were obtained for all conditions using the loudness model described by Moore and Glasberg (2004), a revision of the 1997 model that generates loudness estimates for SNHL as well as NH listeners. Loudness estimates were obtained for pairs of stimuli with the statistical properties described in the preceding text, and a vote was cast for the louder of the two stimuli. Component levels and votes for 1000 such trials were then analyzed using the procedures used in the analysis of data for NH and SNHL listeners. NH listeners were simulated in the model by entering the mean audiogram shown in Fig. 1, corrected to dB HL. SNHL listeners were simulated by entering their individual audiograms and making the default assumption that 10% of the hearing loss was due to loss of inner hair cells (Moore and Glasberg, 2004). The levels for individual components in the SPL and SL conditions used in the model were adjusted to correct for the response of the Sennheiser HD250 headphones as measured in a flat-plate coupler. The levels in the model were adjusted to agree with the physical levels in all cases.

C. Results and discussion

Following Lutfi (1995), the difference between the levels of each of the six tones across the two intervals and the decision regarding which tone was louder were used to estimate the perceptual weight associated with each tone using the following linear regression model:

$$D=\sum w_i x_i+C,$$

where D is the listener's response, x_i is the difference in the level of the ith component across the two intervals, w_i is the weight applied to the ith component, and C is a constant. The weights were normalized so that the sum of the six weights was equal to 1. Weights are generally referred to as having been "assigned" to individual components, but this is not meant to imply an intentional process.

TABLE I. Split halves reliability for perceptual weights. Each of the N weights is based on 1000 trials. Listeners NH 7 and NH 8 were judged to be unreliable.

Subject	r	N
NH-1	0.91	48
NH-2	0.87	48
NH-3	0.97	24
NH-4	0.85	48
NH-5	0.95	24
NH-6	0.95	48
NH-7	0.46	48
NH-8	0.09	48

Because the loudness task does not yield a measure of performance such as percent correct that could be used to verify attention to the task, we began by assessing the reliability of the perceptual weights using a split-halves analysis in which weights computed for odd-numbered trials were compared to those computed for even-numbered trials across the six frequencies and four mean levels for the SPL conditions. The correlation coefficients in Table I suggested that one of the eight NH listeners was making random responses and that another was highly inconsistent. We therefore restricted further analyses to the six NH listeners with reliability coefficients of 0.85 and above. The high degree of reliability for these six listeners suggests that valid estimates of perceptual weight can be obtained with far fewer than 1000 trials.

Mean normalized weights for the six NH listeners for the SPL conditions and corresponding weights from the model are shown in Fig. 2. Results for the SL conditions are shown in Fig. 3. In both sets of conditions, increased weight was given to higher frequency components at higher overall levels, as observed by Leibold *et al.* (2009), but the more obvious effect is that increased weight was given to low-

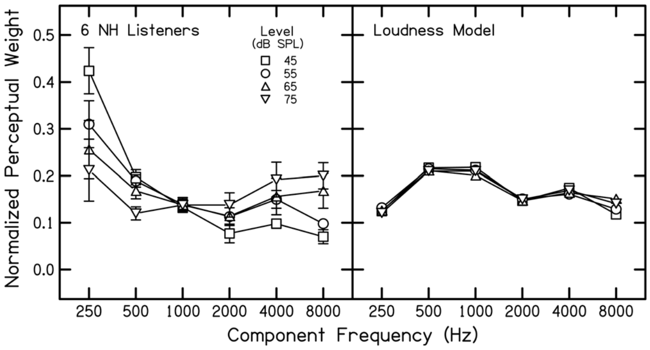

FIG. 2. The left panel shows the mean perceptual weight assigned to the component frequencies by six listeners in experiment 1 when all components were equal in dB SPL. The right panel shows results obtained when predictions from the Moore and Glasberg (2004) loudness model were used to generate perceptual weights.

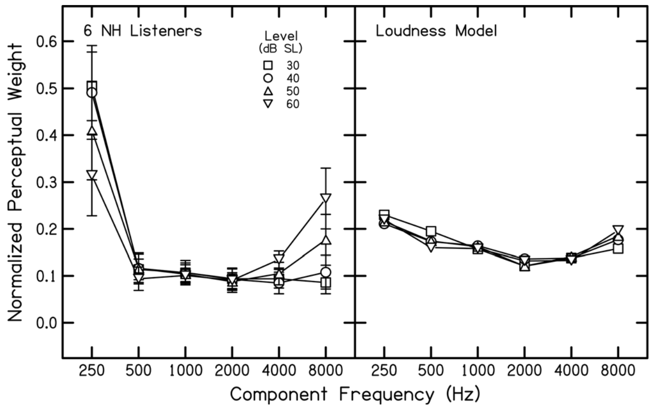

FIG. 3. The left panel shows the mean perceptual weight assigned to the component frequencies by six listeners in experiment 1 when all components were equal in dB SL. The right panel shows results obtained when predictions from the Moore and Glasberg (2004) loudness model were used to generate perceptual weights.

frequency components at lower levels. Leibold *et al*. (2009) did not observe the low-frequency effect.

The data for SPL and SL conditions were included in a single repeated-measures analysis of variance (RM ANOVA) with the SPL vs SL distinction as a nested variable to obtain a sample size sufficient for convergence of the unstructured covariance matrix. The degrees of freedom were adjusted using the Kenward–Rogers method. Data for 1000 Hz were omitted from the analysis so that the value of the normalized weights would not be constrained to sum to 1.0 at all levels. The data show an effect of frequency $[F(4,29.7) = 6.40;\ p < 0.001]$ and a level × frequency interaction $[F(28,64.4) = 4.58;\ p < 0.0001]$, but no significant effect of level. After adjusting for multiple comparisons, 250 and 500 Hz differ from 2000, 4000, and 8000 Hz at 45 dB SPL. At 20 and 30 dB SL, 250 Hz differs from all other frequencies. Weights obtained from the model show little shift in frequency across levels with greater weight at 500 and 1000 Hz. Individual normalized weights for the four SNHL listeners and estimates from the model are shown for SPL conditions in Fig. 4 and SL conditions in Fig. 5. In the equal-SPL conditions, SNHL-1 and -2 show a greater contribution by the low frequencies in agreement with data from the NH listeners. SNHL-2, -3, and -4 show a peak at 4 kHz in agreement with the mean data for NH listeners and with results reported by Doherty and Lutfi (1996). Weights obtained from the model show little contribution by the lowest frequency component. In the equal-SL conditions, all three SNHL listeners assigned weight to the lowest frequency component, but SNHL-3 assigned greater weight to the highest frequencies. Results from the model show the highest predicted weight at 8000 Hz in all cases because the level required to achieve equal SL was 20-30 dB higher at 8000 Hz than at the lower frequencies. Greater weight is assigned at higher levels because all components vary over the same fixed range defined in dB and that range spans a

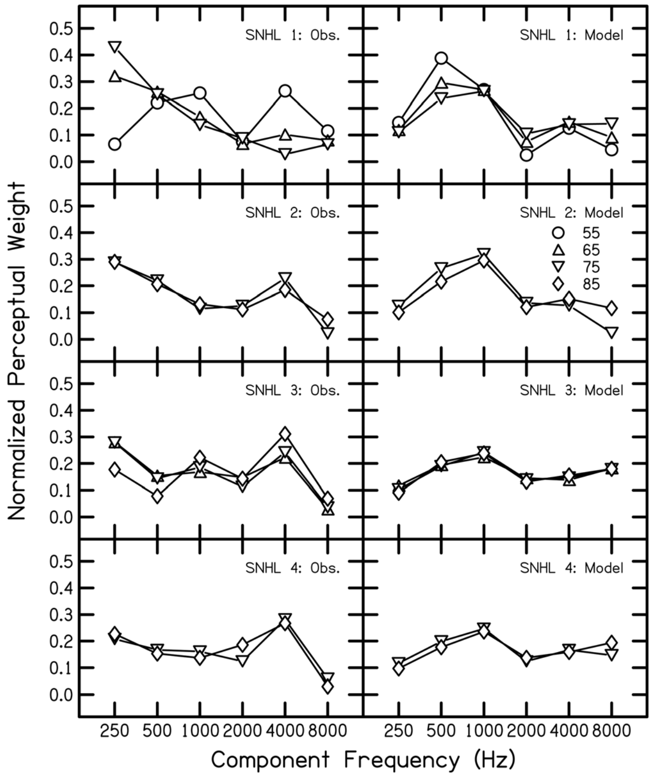

FIG. 4. The left panels show the perceptual weight assigned to the component frequencies by individual listeners with sensorineural hearing loss (SNHL) in experiment 1 when all components were equal in dB SPL. The right panels show results obtained when predictions from the Moore and Glasberg (2004) loudness model for a listener with the same degree of hearing loss were used to generate perceptual weights.

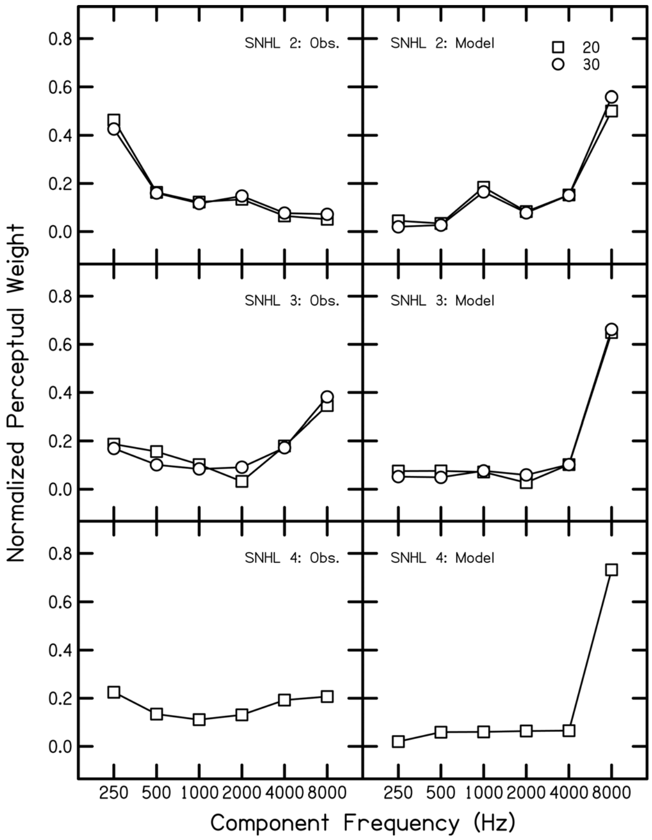

FIG. 5. The left panels show the perceptual weight assigned to the component frequencies by individual listeners with sensorineural hearing loss (SNHL) in experiment 1 when all components were equal in dB SL. The right panels show results obtained when predictions from the Moore and Glasberg (2004) loudness model for a listener with the same degree of hearing loss were used to generate perceptual weights.

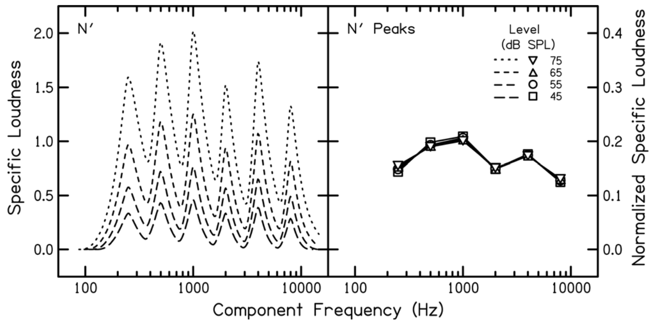

FIG. 6. The left panel shows specific-loudness patterns generated by the Moore and Glasberg (2004) loudness model for six components that were equal in dB SPL. The right panel shows peaks of the loudness patterns, normalized so that the six peaks sum to 1.0, as is the case with normalized perceptual weights. The overlapping symbols indicate that the patterns are parallel, with a form similar to that of perceptual weights generated by the model, as shown in Fig. 2.

wider range of loudness in sones at higher levels than at lower levels.

The normalized weights for SPL and SL conditions for the SNHL listeners were included in a single RM ANOVA with the SPL vs SL distinction as a nested variable as in the analysis of the data for NH listeners. The effect of frequency was significant only in the equal-SL condition [$F(4, 8.68) = 7.89$; $p = 0.006$], where 250 Hz received greater weight than the other frequencies.

Loudness models, including the Moore and Glasberg (2004) model used here, arrive at a loudness estimate by integrating the area under a specific loudness pattern. Examples of loudness patterns predicted by the model in the equal-SPL conditions are shown in the left panel of Fig. 6. To facilitate a comparison between loudness patterns and weights predicted by the model, the values of specific loudness at each of the six frequencies were determined by linear interpolation, then the six values for each pattern were normalized to sum to 1.0 to arrive at a measure of the relative loudness of the components. A plot of the normalized values in the right panel of Fig. 6 shows that relative loudness in the model does not change as a function of level. The functions are similar in shape to those for the predictions of perceptual weight obtained for the model, shown in Fig. 2. For these widely spaced frequency components, the correlation between the perceptual weights and normalized specific loudness pattern is high ($r = 0.92$). The assumptions regarding specific loudness in the model determine the contributions of individual frequencies to total loudness and therefore determine the predicted perceptual weights. This suggests that the operations used to determine perceptual weights in actual listeners, where the underlying specific loudness is unknown, can be used to assess specific loudness. The problem is that the weights obtained from listeners with normal hearing do not agree with the weights predicted by the model ($r = -0.25$) or with the normalized specific loudness pattern ($r = -0.11$). Nonetheless this analysis supports the use of weights as a measure of specific loudness.

Both the loudness task used here and the sample discrimination task used by Doherty and Lutfi (1996) and Leibold *et al.* (2009) show increased weight assigned to high-frequency components at high levels, but results obtained with the two tasks appear to differ at low frequencies. Leibold *et al.* (2009) found no effect of level at low frequencies. Leibold *et al.* (2007) showed better agreement between sample discrimination data and the prediction of the loudness model at the lowest frequency (397 Hz) than at the highest frequency (2519 Hz) in their widest bandwidth condition. These differences in results indicated a need to repeat a subset of the conditions used in experiment 1 in a sample discrimination task.

III. EXPERIMENT 2: EFFECTS OF LEVEL ON PERCEPTUAL WEIGHTS FOR SAMPLE DISCRIMINATION

A. Listeners

Five of the eight NH listeners tested in experiment 1 participated in this experiment. NH-7 and NH-8 were not included as a result of the reliability analysis summarized in Table I and NH-4 was not available for further testing.

B. Procedure

The stimuli were identical to those used in experiment 1 except that all components for one interval of each trial were drawn from a distribution with a mean value of 45, 55, 65, or 75 dB SPL, while all components of the other interval, selected at random, were drawn from a distribution with a mean value of 50, 60, 70, or 80 dB SPL, respectively. The six-tone complex generated from distributions with the higher mean value would be expected to be louder and was designated as the correct response. The data collection program reported the percentage of correct responses at the end of every block of 100 trials, but no trial-by-trial feedback was given and the instructions emphasized judgments of total loudness. Listeners made 1000 decisions at each of the four levels. The levels were presented in ascending order to half of the subjects and in descending order to the other half.

C. Results and discussion

Perceptual weights were derived for the sample-discrimination task by sorting the trials into two groups based on the interval with the higher average level, computing multiple regression weights for each group of 500 trials as described for experiment 1, then averaging the two weights for each of the six components before normalizing them to sum to 1. Figure 7 shows mean weights for the same five listeners in the two tasks. Data in the left panel differ from those in the left panel of Fig. 2 due to the absence of NH-4. The pattern is comparable for the two procedures. An ANOVA of the normalized weights with 1000 Hz excluded from the analysis, as in experiment 1, showed a level × frequency interaction [$F(12,15.7) = 6.99$; $p = 0.0003$], but no significant effect of level, frequency, or task. There was no significant task × level or task × frequency interaction. As expected, the un-normalized weights for the loudness task were larger by a factor of 2, on average, than those for sample discrimination. The fact that there was a net difference, on average, between the levels of the stimuli in the two

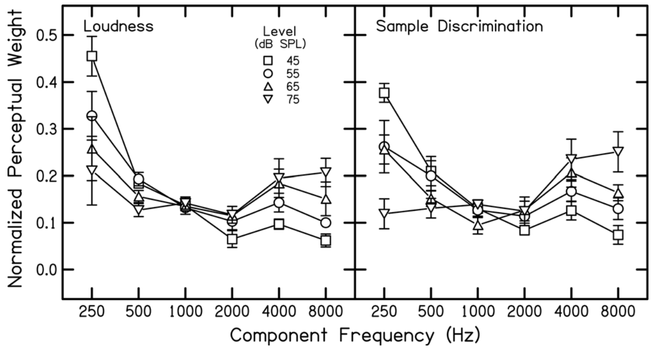

FIG. 7. A comparison of perceptual weights obtained in a loudness task and a sample discrimination task by the same group of five NH listeners.

observation intervals reduced the size of the weights in the sample discrimination task despite the fact that weights were computed separately for trials where samples from the higher distribution were presented in the first or second interval. This occurs because the average difference in levels between the two intervals accounts for a substantial portion of the decision variance (Richards and Zu, 1994).

While the raw weights were lower in the sample discrimination task, sample discrimination has the advantage of providing information on the performance of the subjects. The five subjects had a mean percent correct across all levels of 86.5 and a corresponding d' value of 1.60. The observed d' values increased significantly with level [F (3,12) = 8.68; $p = 0.0025$]. This might at first glance be attributed to the "near-miss" to Weber's law, the well-known result that performance in intensity resolution tasks improves with level (McGill and Goldberg, 1968). A more detailed analysis suggests that this is not the case. Performance in any given sample discrimination condition is determined by internal noise and by the degree to which listeners use optimum weights (Berg, 1990). Ideal performance could be achieved in this task by weighting the six frequencies equally. The normalized weights in Fig. 7 show that weights are more equal at high levels (up and down triangles) than at low levels (squares and circles). The pattern can be reduced to a single measure of observed weighting use, d'_{wgt} (Berg, 1990). Values of d'_{obs} and d'_{wgt} are shown in Fig. 8. Values of d'_{obs} increased significantly with level [F (3,12) = 8.68; $p = 0.0025$]. Values of d'_{wgt} appear to parallel those for d'_{obs}, but the effect of level did not reach significance in that case. The pattern of results suggests that in this case improved performance at higher levels is due to the more equal use of weights across frequencies rather than reduced internal noise.

The reliability analysis summarized in Table I provides a means of assessing the level of attention of individual listeners to the task and a check on the quality of the data, but it cannot be done until data collection has been completed. The sample discrimination task provides information after every block of trials. Listeners in the loudness task reported that many of their decisions were arbitrary because the two stimuli presented on some trials did not differ noticeably in loudness. The sample discrimination task reduces the number of such trials by creating an average difference in intensity across the two observation intervals. It is similar in this regard to using transformed up-down rules to target points above and below 50% in loudness matching tasks (Jesteadt, 1980).

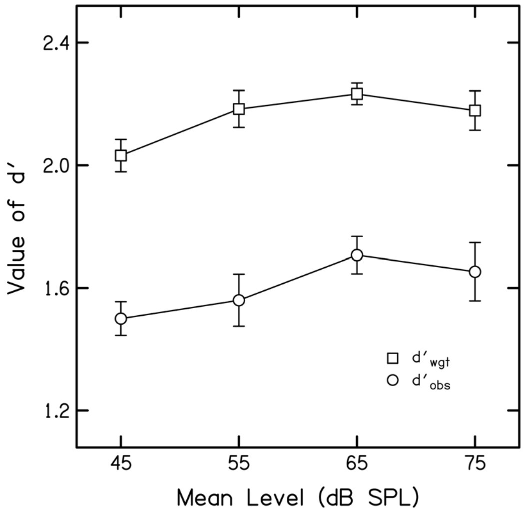

FIG. 8. Mean values of d'_{obs} and d'_{wgt} for five listeners in the sample discrimination task.

IV. GENERAL DISCUSSION

Data have been presented concerning the contribution of individual components of a multi-tone complex to the overall loudness of the complex. The results can be considered as a series of contrasts: (1) Component levels equated in dB SPL vs dB SL, (2) listeners with normal hearing vs listeners with sensorineural hearing loss, (3) loudness judgments vs sample discrimination, and (4) observed data vs predictions of the loudness model.

A. Component levels equated in dB SPL vs dB SL

A significant frequency-by-level interaction was observed for NH listeners in both the equal-SPL (Fig. 2) and equal-SL (Fig. 3) conditions. In the equal-SL conditions, this effect was confined to the lowest and highest frequencies with little or no effect of level over the range from 500 to 4000 Hz. The main difference in the results for equal-SPL and equal-SL conditions occurs in the predictions of the model, which are discussed in Sec. IV D.

B. Listeners with normal hearing vs hearing loss

The data for SNHL listeners are quite limited. In the equal-SPL conditions, the SNHL listeners did not show the shift in weight to higher frequencies at higher levels observed in NH listeners. Three of the four SNHL listeners assigned maximum weight to the 4000-Hz component, a result consistent with the pattern observed by Doherty and Lutfi (1996), but the effect of frequency was not significant in the RM ANOVA. There was greater agreement between NH and SNHL listeners in the equal-SL conditions, but the data set may be too small to draw valid conclusions. Inclusion of a component at 8000 Hz limited the range of levels that could be included in both the equal-SPL and equal-SL conditions when testing SNHL listeners. Inclusion of a larger number of frequencies spread over a narrower frequency range in both the audiogram and the test stimuli would facilitate efforts to relate perceptual weights to regions of hearing loss.

C. Loudness judgments vs sample discrimination

The data support the use of a sample-discrimination task as a measure of perceptual weights for loudness in that results observed in the current study with instructions to respond on the basis of loudness were similar to those in earlier sample-discrimination studies. Most of the previous uses of sample discrimination have given listeners correct-answer feedback (e.g., Doherty and Lutfi, 1996; Lutfi and Jesteadt, 2006; Oberfeld *et al.*, 2012), and all but Oberfeld *et al.* described it to listeners as an intensity resolution task rather than one involving judgments of overall loudness. Results reported here may have been different if the listeners had been given a more complete description of the properties of the stimuli and feedback regarding their answers, but the intent was to make the loudness and sample-discrimination tasks comparable, not to emphasize potential differences.

The use of sample discrimination to measure perceptual weights for loudness was of some concern because measures of intensity resolution and measures of loudness are not always in close agreement (Hellman *et al.*, 1987; Schlauch and Wier, 1987; Zwislocki and Jordan, 1986). Experimental manipulations of the slope of the loudness function do not, in general, result in corresponding changes in intensity-discrimination thresholds. Hellman *et al.* (1987), for example, measured growth of loudness and intensity discrimination for 1000-Hz tones presented in narrow-band or wideband noise. Although the loudness functions were steeper for tones in narrow-band noise, there was no difference in intensity discrimination. The relation between the perceptual weight assigned to an individual component of a multi-tone complex and slope of the loudness function for that component is unclear. Leibold *et al.* (2007) found good agreement between perceptual weights and intensity resolution for individual components in five-tone complexes when the tones were closely spaced in frequency, but less agreement when the tones were more widely spaced. The bandwidth in their widest spacing condition was 2119 Hz, much less than the bandwidth in the current study. The intensity resolution thresholds reported by Leibold *et al.* (2007) suggested that listeners were basing their decisions on the change in intensity of the total complex not the change at a single frequency. If their listeners had been able to focus on the frequency where the increment was added, the threshold for detection of the increment should have been lower than the observed value.

One goal of experiment 2 was to obtain loudness judgments in a sample-discrimination framework that would provide a measure of percent correct that could be used to monitor attention to the task and that would enable analyses of weighting efficiency and internal noise. The results suggest that it is feasible to do that without changing the nature of the loudness judgments.

D. Observed data vs predictions of the loudness model

The loudness model (Moore *et al.*, 1997; Moore and Glasberg, 2004) was not specifically designed to predict results of studies of perceptual weights. The application of it in this context assumes that decisions in these tasks are based on the total loudness of the multi-tone complex. Given that assumption, it is simple to obtain a loudness estimate for two complexes and to have the model vote for the louder of the two. The resulting predicted perceptual weights show virtually no effect of overall level on the spectral loudness profile in any of the conditions tested, for either NH or SNHL listeners. The pattern of weights as a function of frequency also differs markedly for NH listeners vs the model in Fig. 2, where all components were presented at equal SPL. The model predicts that greater weight will be assigned to the more audible mid-frequency components because the loudness difference in sones for a given difference in decibels is greater at higher levels. The opposite effect is observed in the data. The pattern of weights as a function of frequency predicted by the model was in greater agreement with the data in the condition where all components were presented at equal SL as shown in Fig. 3. Equating the components in dB SL resulted in presentation of the highest and lowest frequency components at higher physical levels. The model was more sensitive to this shift than the listeners were, bringing the predictions of the model more in line with the data.

It is not clear whether the low-frequency emphasis observed in the data is associated with the lower edge of the tone complex or is associated with low frequencies *per se*. Because the model takes known peripheral processes into account and provides an accurate description of the loudness of narrow-band sounds, it is reasonable to assume that the low-frequency emphasis and shift in perceptual weight with increasing level may be due to more central processes. An alternative explanation of the results is that the weights reflect sensitivity to change in the level of a given component rather than the contribution of that component to total loudness. The increased weight assigned to low frequencies might then be attributed to the presence of steeper loudness functions at low frequencies. Oberfeld *et al.* (2012) considered this explanation for the low-frequency effect observed in their data and found that the predicted effect on perceptual weights was smaller than the observed effect. For the present conditions, any effect of steeper loudness functions at low frequencies should have been observed in the predictions generated by the model because it assumes steeper loudness functions at low frequencies, although the difference at 250 Hz is small. Listeners may be more sensitive to changes in components at the edge of a tone complex in a way not captured by the loudness model. Listeners are better able to hear out partials in inharmonic complex tones, for example, when the partials are at the lower or upper edge of the tone complex (Moore *et al.*, 2006; Moore and Ogushi, 1993). In profile analysis experiments, however, listeners show degraded performance when the signal is added to a component at the edge of the stimulus spectrum (Green and Berg, 1991; Green *et al.*, 1987; Green and Mason, 1985). The assumption that increased weight reflects a greater contribution to total loudness will be tested in a future study using a loudness matching task.

V. CONCLUSIONS

(1) Listeners assigned greater weight to the lowest and highest frequency in the six-tone complexes used in the current study than would be predicted by the loudness models proposed by Moore *et al.* (1997) and Moore and Glasberg (2004).

(2) The weight shifted from low to high frequencies when the complexes increased in level.

(3) The pattern of perceptual weights as a function of frequency was in better agreement with the model when the six tones were presented at equal SL than at equal SPL.

(4) Introduction of an average intensity difference between the two intervals provides the advantages of a sample discrimination task without altering the nature of the loudness judgments.

(5) Limited data for listeners with sensorineural hearing loss show effects at the lowest and highest frequency that are comparable to those for listeners with normal hearing, but no change in perceptual weight with level, over the limited range of levels available.

(6) It is possible that the increased perceptual weight associated with the highest and lowest frequency components reflects greater salience of those components rather than a greater contribution of those components to the overall loudness of the complex.

ACKNOWLEDGMENT

This research was supported by NIH Grant Nos. R01 DC 011806, T32 DC000013, and P30 DC 004662. The authors thank Dr. Moore and Dr. Glasberg for providing source code for the loudness model, Dr. Moore and Dr. Oberfeld for helpful comments on earlier versions of the manuscript, and Robin High from the Department of Biostatistics at the College of Public Health, University of Nebraska Medical Center for his help with statistical analysis.

Berg, B. G. (**1990**). "Observer efficiency and weights in a multiple observation task," J. Acoust. Soc. Am. **88**, 149–158.

Chalupper, J., and Fastl, H. (**2002**). "Dynamic loudness model (DLM) for normal and hearing-impaired listeners," Acust. Acta Acust. **88**, 378–386.

Cox, R. M., Alexander, G. C., Taylor, I. M., and Gray, G. A. (**1997**). "The contour test of loudness perception," Ear Hear. **18**, 388–400.

Doherty, K. A., and Lutfi, R. A. (**1996**). "Spectral weights for overall level discrimination in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **99**, 1053–1058.

Green, D. M., and Berg, B. G. (**1991**). "Spectral weights and the profile bowl," Q. J. Exp. Psychol. A **43**, 449–458.

Green, D. M., and Mason, C. R. (**1985**). "Auditory profile analysis: Frequency, phase, and Weber's law," J. Acoust. Soc. Am. **77**, 1155–1161.

Green, D. M., Onsan, Z. A., and Forrest, T. G. (**1987**). "Frequency effects in profile analysis and detecting complex spectra changes," J. Acoust. Soc. Am. **81**, 692–699.

Hellman, R., Scharf, B., Teghtsoonian, M., and Teghtsoonian, R. (**1987**). "On the relation between the growth of loudness and the discrimination of intensity for pure tones," J. Acoust. Soc. Am. **82**, 448–453.

Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments," Percept. Psychophys. **28**, 85–88.

Kortekaas, R., Buus, S., and Florentine, M. (**2003**). "Perceptual weights in auditory level discrimination," J. Acoust. Soc. Am. **113**, 3306–3322.

Leibold, L. J., and Jesteadt, W. (**2007**). "Use of perceptual weights to test a model of loudness summation," J. Acoust. Soc. Am. **122**, EL69–EL73.

Leibold, L. J., Tan, H., and Jesteadt, W. (**2009**). "Spectral weights for sample discrimination as a function of overall level," J. Acoust. Soc. Am. **125**, 339–346.

Leibold, L. J., Tan, H., Khaddam, S., and Jesteadt, W. (**2007**). "Contributions of individual components to the overall loudness of a multitone complex," J. Acoust. Soc. Am. **121**, 2822–2831.

Lentz, J. J. (**2007**). "Variation in spectral-shape discrimination weighting functions at different stimulus levels and signal strengths," J. Acoust. Soc. Am. **122**, 1702–1712.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lutfi, R. A. (**1995**). "Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks," J. Acoust. Soc. Am. **97**, 1333–1334.

Lutfi, R. A., and Jesteadt, W. (**2006**). "Molecular analysis of the effect of relative tone level on multitone pattern discrimination," J. Acoust. Soc. Am. **120**, 3853–3860.

McGill, W. J., and Goldberg, J. P. (**1968**). "A study of the near-miss involving Weber's law and pure-tone intensity discrimination," Percept. Psychophys. **4**, 105–109.

Moore, B. C. J., and Glasberg, B. R. (**2004**). "A revised model of loudness perception applied to cochlear hearing loss," Hear. Res. **188**, 70–88.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness, and partial loudness," J. Audio Eng. Soc. **45**, 224–240.

Moore, B. C. J., Glasberg, B. R., Low, K.-E., Cope, T., and Cope, W. (**2006**). "Effects of level and frequency on the audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **120**, 934–944.

Moore, B. C. J., and Ohgushi, K. (**1993**). "Audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **93**, 452–461.

Oberfeld, D., Heeren, W., Rennies, J., and Verhey, J. (**2012**). "Spectro-temporal weighting of loudness," PLOS One **7**, 1–14.

Pollack, I. (**1951**). "On the measurement of the loudness of white noise," J. Acoust. Soc. Am. **23**, 654–657.

Pollack, I. (**1952**). "The loudness of bands of noise," J. Acoust. Soc. Am. **24**, 533–538.

Richards, V. M., and Zhu, S. (**1994**). "Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients," J. Acoust. Soc. Am. **95**, 423–434.

Schlauch, R. S., and Wier, C. C. (**1987**). "A method for relating loudness-matching and intensity-discrimination data," J. Speech Hear. Res. **30**, 13–20.

Stevens, S. S. (**1961**). "Procedure for calculating loudness: Mark VI," J. Acoust. Soc. Am. **33**, 1577–1585.

Willihnganz, M. S., Stellmack, M. A., Lutfi, R. A., and Wightman, F. L. (**1997**). "Spectral weights in level discrimination by preschool children: Synthetic listening conditions," J. Acoust. Soc. Am. **101**, 2803–2810.

Zwicker, E., and Scharf, B. (**1965**). "A model of loudness summation," Psychol. Rev. **72**, 3–26.

Zwislocki, J. J., and Jordan, H. N. (**1986**). "On the relations of intensity jnd's to loudness and neural noise," J. Acoust. Soc. Am. **79**, 772–780.