Effects of relative and absolute frequency in the spectral weighting of loudness^{a)}

Suyash Narendra Joshi^{b)} and Marcin Wróblewski

Center for Hearing Research, Boys Town National Research Hospital, Omaha, Nebraska 68131, USA

Kendra K. Schmid

Department of Biostatistics, College of Public Health, University of Nebraska Medical Center, Omaha, Nebraska 68198, USA

Walt Jesteadt

Center for Hearing Research, Boys Town National Research Hospital, Omaha, Nebraska 68131, USA

(Received 10 April 2015; revised 18 December 2015; accepted 30 December 2015; published online 19 January 2016)

The loudness of broadband sound is often modeled as a linear sum of specific loudness across frequency bands. In contrast, recent studies using molecular psychophysical methods suggest that low and high frequency components contribute more to the overall loudness than mid frequencies. In a series of experiments, the contribution of individual components to the overall loudness of a tone complex was assessed using the molecular psychophysical method as well as a loudness matching task. The stimuli were two spectrally overlapping ten-tone complexes with two equivalent rectangular bandwidth spacing between the tones, making it possible to separate effects of relative and absolute frequency. The lowest frequency components of the “low-frequency” and the “high-frequency” complexes were 208 and 808 Hz, respectively. Perceptual-weights data showed emphasis on lowest and highest frequencies of both the complexes, suggesting spectral-edge related effects. Loudness matching data in the same listeners confirmed the greater contribution of low and high frequency components to the overall loudness of the ten-tone complexes. Masked detection thresholds of the individual components within the tone complex were not correlated with perceptual weights. The results show that perceptual weights provide reliable behavioral correlates of relative contributions of the individual frequency components to overall loudness of broadband sounds. © 2016 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4939893>]

[VMR]

Pages: 373–383

I. INTRODUCTION

Loudness is generally defined as the perceived magnitude of a sound that primarily changes with intensity (Florentine, 2011). Here, “primarily” is a key adjective, since loudness is also affected by other physical dimensions such as duration, frequency, and bandwidth (for review, Jesteadt and Leibold, 2011). Almost all the sounds in our acoustic environment, including speech, are broadband in nature. Nevertheless, more research has been focused on the loudness of sinusoidal stimuli than broadband sounds. This discrepancy may be the reason why loudness discomfort remains a major source of dissatisfaction with hearing-aid prescription (Kochkin, 2010), despite use of loudness data in hearing aid fitting. Understanding the laws that govern the loudness of broadband sounds is, therefore, clinically important as well as ecologically relevant. This is one of a series of studies exploring the use of molecular psychophysics to

assess the contributions of individual frequency bands to the total loudness of broadband sounds.

Zwicker *et al.* (1957) investigated the effect of the frequency spacing of tones within a multi-tone complex on the loudness of the complex. They showed that the loudness did not change when all the components remained within a single critical band. The loudness increased, however, when spacing of the tones occupied a bandwidth larger than a critical band. This trend of spectral summation of loudness was observed for noise stimuli as well. Zwicker (1958) and Zwicker and Scharf (1965) proposed a model of loudness perception for broadband sounds, which had a significant impact on all later models. In their model, stimuli were transformed to produce geometric excitation patterns, which were then converted to specific loudness patterns using a function that described a power-law type relationship between the degree of excitation and loudness. In this and later models, specific loudness (N') was defined as the loudness per critical band. The overall loudness in sones was obtained by calculating the area under the curve for N' plotted against a frequency scale with units of critical bands, Barks (in the model of Zwicker, 1958) or equivalent rectangular bandwidths (ERBs) (Glasberg and Moore, 1990) in the model of Moore and colleagues (Moore *et al.*, 1997). In

^{a)} A portion of this work was presented at the Annual Meeting of the American Auditory Society in Scottsdale, AZ in March, 2012.

^{b)} Present address: Hearing Systems, Technical University of Denmark, Ørstedts Plads Building 352, 2800 Kongens Lyngby, Denmark. Electronic mail: suyash@suyash-joshi.com

summary, the loudness of a broadband sound is generally assumed to be a linear summation of specific loudness across auditory filters. This modeling framework has served as the basis for all modern psychoacoustical models of loudness perception (e.g., Chalupper and Fastl, 2002; Moore and Glasberg, 2004; Moore *et al.*, 1997). The loudness model described by Moore and colleagues serves as an ANSI standard for calculation of loudness (ANSI, 2007) as well as a basis for hearing aid prescriptions and for their validation (e.g., NAL, CAMEQ) (Moore and Sk, 2013).

Leibold *et al.* (2007) used molecular psychophysical methods to determine the relative contribution of each component of a five-tone complex to the loudness of the complex. In their study, the components were logarithmically spaced, centered at 1000 Hz; the lowest bandwidth tested was 46 Hz and the highest was 2119 Hz. Their analysis showed that for the conditions where bandwidth did not exceed a critical band, the perceptual weights for the components of a complex were relatively uniform. For conditions in which the bandwidth exceeded a single critical band, weights for components differed significantly from one another, suggesting differential contributions to the overall loudness. Leibold *et al.* (2007) used the loudness model of Moore and Glasberg (2004) to predict the weighting patterns they observed. The model was successful in predicting the observed weights for all but the widest bandwidth condition, where the components were spread across several critical bands and did not mask one another. In that condition, the model predicted relatively uniform weights, but the observed weights continued to show the same emphasis on the lowest and highest components observed at intermediate bandwidths. The results for the widest bandwidth condition appeared to contradict the assumption common to standard loudness models that the loudness of a broadband sound is a simple linear summation of loudness across the auditory filters.

To further investigate summation of loudness across the auditory filters, Jesteadt *et al.* (2014) used a six-tone complex, components of which were widely separated over a five-octave range (octave spaced frequencies, from 250 to 8000 Hz). Such a stimulus avoids the possible confounding effect of one component masking another and reducing its contribution to the overall loudness. The effect of overall level was investigated in two conditions, one in which the mean level of all the tones was equal in sound pressure level (SPL) and one where levels were corrected for threshold, resulting in components equal in sensation level (SL). In both cases, at low overall levels, low frequencies dominated the loudness. With increasing level, however, the low frequency emphasis was reduced and high frequencies received more weight. Although the loudness model (Moore and Glasberg, 2004) failed to predict the observed weighting function or any change in relative weights with increasing level, Jesteadt *et al.* (2014) showed that weights from the model correlated with the specific loudness predictions of the model, suggesting that the underlying specific loudness of the components of a broadband sound could be assessed behaviorally using the perceptual weights. An increase in high frequency weight with increasing level has also been

reported for sample discrimination of intensity (Kortekaas *et al.*, 2003; Leibold *et al.*, 2009) and discrimination of spectral shapes (Lentz, 2006).

Central to our investigation in this article is the question posed by the studies reviewed above, regarding the calculation of the specific loudness of individual frequency bands or auditory filter outputs in current loudness models. One can consider three possible explanations of the pattern of low- and high-frequency emphasis in the existing perceptual-weights data. The first is based on differences in growth of loudness with increasing level at different frequencies. For example, loudness grows more rapidly at low frequencies than at middle or high frequencies (Hellman and Zwisllocki, 1968). The model would fail to reproduce the observed frequency emphasis in perceptual weights if it underestimates how fast loudness grows at a particular frequency. The second is to assume that the model makes correct assumptions regarding growth of loudness, but that the loudness is indeed summed unequally across the auditory filters, and that the weights reflect those differences. The third is to assume that the weights do not reflect contributions to overall loudness but rather reflect sensitivity to the changes in the levels of individual components used to derive the weights.

Differences in the growth of loudness as a function of frequency are often estimated from equal loudness contours (e.g., Jesteadt and Leibold, 2011; Scharf, 1978). Although Glasberg and Moore (2006) showed that their loudness model can predict the standard equal-loudness contours reasonably well, a direct comparison of their predicted contours and those in the current ISO (2003) standard [Jesteadt and Leibold (2011), Fig. 5.5] suggests that the model may underestimate the steepness of the loudness function at low frequencies. Jesteadt *et al.* (2014) showed low frequency emphasis at low overall levels in both equal-SL and equal-SPL conditions. The lowest frequency component was presented at levels closer to quiet threshold than other components in the equal-SPL condition and at a higher SPL in the equal-SL condition. Therefore, it is unlikely that audibility of a tone or absolute level could account for the increased weight given to the lowest-frequency component in both conditions. Recently, Oberfeld *et al.* (2012) obtained perceptual weights for loudness of noise made up of multiple bands in a condition where levels of the noise bands were equated to produce equal loudness. In line with the results of Jesteadt *et al.* (2014), their listeners put highest weight on the lowest component of the complex. Oberfeld *et al.* (2012) demonstrated that the weights could not be accounted for by assuming different slopes of the loudness function at different frequencies. These results provide indirect support for the second and third explanations, in which the model fails to predict the weights because summation across the auditory filters carries unequal weighting or the weights reflect sensitivity to change rather than contribution to total loudness. The series of experiments reported below was designed to investigate the unequal weighting of loudness across the filters and its dependence on stimulus parameters.

II. EXPERIMENT 1: EFFECT OF A COMPONENT'S RELATIVE POSITION ON ITS SPECIFIC LOUDNESS

The components of the six-tone complex used by Jesteadt *et al.* (2014) were spread over a wide frequency range to avoid masking of one component by another. It is therefore unlikely that the emphasis on the lowest and highest components observed by Jesteadt *et al.* was due to reduced masking of those components. It might have been due to their edge position in the complex for some reason other than masking or to the specific frequencies of those components. For example, Moore and Ohgushi (1993) tested an effect of spacing between components of an inharmonic-tone complex on their audibility. They showed that when components were separated by equal ERB units, the lowest and the highest edge partials were most audible in all the conditions. Audibility of the inner partials, however, also increased with increasing spacing. While the method used in their study is not directly comparable to ours, their thesis regarding a “salience” of the partials is certainly applicable. This experiment was designed to investigate the effects of relative and absolute position of the components of a complex tone on the perceptual weight in a loudness comparison task.

A. Listeners

Six listeners with normal hearing participated in the study. All listeners had audiometric thresholds below 15 dB

hearing level (HL) at the octave audiometric frequencies. All listeners were compensated for their time of participation in the study. The experimental procedures were approved by the institutional review board at the Boys Town National Research Hospital.

B. Stimuli

Two ten-tone complexes, components of which were two-ERB apart, were used to study the frequency specificity of the contributions of individual components to total loudness. The low-frequency complex included components at 208, 313, 444, 605, 806, 1055, 1364, 1747, 2223, and 2812 Hz. The high-frequency complex included components at 806, 1055, 1364, 1747, 2223, 2812, 3544, 4452, 5578, and 6975 Hz. The fifth component of the low complex (806 Hz) was the first (lowest) component of the high complex. The frequency overlap of the two complexes is shown in Fig. 1(A). This stimulus configuration made it possible to distinguish between absolute frequency, represented by the low and high complexes, and relative frequency, represented by position within the complex. The stimuli were 500-ms long with 10-ms onset and offset cosine-squared ramps. Stimuli were created as individual sinusoids in Matlab with a sampling rate of 44 100 Hz with 24-bit resolution. For each interval of every trial, the level of each sinusoid was selected from a rectangular distribution with a range of 16 dB in 1-dB

FIG. 1. (Color online) (A) Spectral configuration of the two ten-tone complexes. The lowest frequency of the High complex was the fifth component in the Low complex. This stimulus configuration was designed to test the effect of absolute frequency and relative position of the component in a tone complex on the perceptual weight it received. (B) Illustration of the task used to obtain perceptual weights. Level of each component in the “signal” and “non-signal” interval was selected from a distribution with a mean of 48 and 45 dB, respectively. Listeners were asked to pick the interval that was louder.

steps and the ten sinusoids were then added in random phase. The low-frequency complex and the high-frequency complex were presented in separate blocks. The order of presentation of the blocks was randomized.

C. Methods

Listeners were seated in a double-walled sound attenuating booth. Stimuli were generated using 24-bit digital-to-analog convertors (Digital Audio Labs, CardDeluxe) and presented monaurally to the left ear through Sennheiser 250-HD headphones.

In a two-interval forced choice (2IFC) task, listeners were asked to pick the interval that was louder. For the “signal” interval, the mean of the level distribution for the individual components was 48 dB SPL (Overall mean level of 58 dB); for the “non-signal” interval it was 45 dB SPL (Overall mean level of 55 dB). A visual description of a sample trial is shown in Fig. 1(B). The two intervals, each 500-ms long, were separated by a 500-ms inter-stimulus-interval. Each trial began with a 500-ms light emitting diode (LED) light on a response box indicating the beginning of the trial. Separate lights to cue the listener highlighted each interval. There was no gap between the initial warning light offset and the onset of the light indicating the beginning of the first interval. Stimulus onset was simultaneous with the onset of the interval lights. Although each trial could be marked with a “correct” and “incorrect” interval, listeners received no feedback regarding their response. Data were collected in blocks of 100 trials and appropriate breaks were given to the listeners. Each listener completed 1000 trials for each of the two ten-tone complexes. Blocks for the two ten-tone complexes were run in a random order.

The Moore and Glasberg (2004) model was tested for its ability to predict listener’s responses. It was used because it represents the ANSI standard, has been used for development and evaluation of hearing aid fitting strategies and was used by Leibold *et al.* (2007) and Jesteadt *et al.* (2014). The model was provided with a headphone response of Sennheiser HD-250 headphones, measured on a flat plate coupler and was run with the same configuration of the stimuli on every trial that each listener received. The model

estimated the loudness of the two stimuli on every trial as an artificial observer and voted for the louder interval. This method avoided any confounds such as differences in weights due to a rove distribution or changes in actual levels due to headphone calibration issues. Responses by the model were then analyzed in the same way as responses by the individual listeners.

D. Results

For both conditions tested here, listeners performed at a minimum of 70% correct in the sample discrimination task, with mean of 77% correct for the low complex and 78.9% correct for the high complex. Perceptual weights for the components were obtained using the difference in the level of each component between the two intervals (interval 2 – interval 1) to predict the listener’s responses in a linear model, as shown in Eq. (1) (Lutfi, 1995; Richards and Zhu, 1994), where x_i is the difference in level between the two intervals for component x , w_i represents the weight given to that component, and C is additive internal noise. Weights were calculated separately using multiple linear regression for trials when the correct interval was first or second, and then averaged. Because the signal was in the same interval for all trials included in a given regression analysis, only the differences in the levels of the components could contribute to systematic differences in the responses (Berg, 1990). The same method was used to obtain the weights predicted by the loudness model

$$D = \sum_{i=1}^{i=10} w_i x_i + C. \quad (1)$$

Weights were normalized to sum to unity, since our goal was to investigate the “relative” contributions of the components of the complex. Mean normalized observed weights as well as the model predictions are shown in Fig. 2.

This experiment was designed to determine whether frequencies below 500 Hz receive higher weight because lower frequencies contribute more to total loudness than current models predict or because their position at the lower edge of

FIG. 2. (Color online) Mean of the normalized perceptual weights and their standard deviations from experiment 1. The weights were obtained using a multiple regression model to predict a listener’s responses on every trial using the difference in level between interval 2 and interval 1 for each individual component. Weights predicted by the loudness model are shown with a dashed line.

the complex makes them more salient. For the low complex, the lowest and highest frequency components both received greater weight than was predicted by the model. For the high complex, the lowest frequency component received greater weight than all but one other component, but the model actually predicted this effect. The similarity of the patterns of weights in the two panels suggests that relative position is important. The effect of relative position is clearer in Fig. 3, where the normalized weights are plotted for individual subjects. Subjects 1 and 3 assign lower relative weight to the lowest component when it is part of the low complex than when it is part of the high complex and subject 6 shows no difference in weights for the two complexes. The importance of the sixth component in the mean data for the high complex is due to subject 4.

A repeated measures analysis of variance (ANOVA) comparing the weights assigned to the ten components of each complex based on their relative frequencies showed a significant effect of relative frequency within the complex [$F(9,54.97) = 2.12$; $p = 0.043$], but no significant interaction of relative frequency and absolute frequency [$F(7,55.22) = 0.56$; $p = 0.78$].¹ For the six overlapping frequencies, statistical analysis indicated no significant difference in the overall weight assigned to those frequencies in the low and high complexes [$F(1,9.43) = 1.56$; $p = 0.242$], no significant differences among the weights assigned to the six components [$F(5,23.24) = 2.07$; $p = 0.105$] and no significant interaction of relative frequency and absolute frequency [$F(5,23.24) = 0.63$; $p = 0.68$]. The six frequencies were not weighted differently when they were the lowest six components than when they were the highest six components.

III. EXPERIMENT 2: RELATIONSHIP BETWEEN AUDIBILITY AND SPECIFIC LOUDNESS

The slope of the loudness function is frequency dependent (Scharf, 1978). The information provided by the level

rove may be greater for frequencies where the slope of the loudness function is steeper than for frequencies where it is shallower. Although this is a plausible assumption, data in the literature suggest that neither frequency (Jesteadt *et al.*, 1977) nor the slope of the loudness function (Hellman *et al.*, 1987) affects performance in intensity discrimination. Berg (1990) and Lutfi and Jesteadt (2006) found that more intense tones receive more weight when tones are presented sequentially in a sample discrimination task, even when the more intense tones carry less information. Turner and Berg (2007) and Oberfeld *et al.* (2013) ruled out forward masking as a contributor to this level dominance, but greater audibility could still be a factor.

Leibold *et al.* (2007) investigated the relationship between the audibility of a component and the perceptual weight it received in a loudness comparison task. They found that edge frequencies of closely spaced tone complexes had lower thresholds than frequencies in the middle of the complex. As the spacing between the components increased, thresholds became approximately equal among the components. The correlation between the audibility thresholds and the perceptual weight weakened with increasing spacing between the tones. The widest bandwidth tested in their study, however, was only 2119 Hz. Jesteadt *et al.* (2014) obtained weights for components that were presented at equal SL, avoiding the effect of threshold differences across components. Oberfeld *et al.* (2012) obtained weights for noise bands that were matched in loudness, avoiding any “loudness” dominance. Both studies, however, found that the listeners put higher weight on the low frequency component of their stimuli. These results could not be reconciled with level dominance, unless the response to each component is affected when presented simultaneously with other components. For example, low-frequency maskers more readily mask high-frequency signals, a phenomenon known as the upward spread of masking. Mutual masking might alter or enhance the audibility of certain components of the

FIG. 3. (Color online) Normalized perceptual weights for six individual subjects in experiment 1, plotted as a function of component number within the low and high complexes.

complex tones, producing level dominance that is reflected in the perceptual weights.

This experiment tested the relationship between listener’s masked thresholds for each component in the complex and the weights obtained in experiment 1. If masked thresholds show a systematic trend, where components that receive highest weight have lowest masked threshold and vice versa, it would suggest that the weights represent level dominance created by mutual masking.

A. Methods

The six listeners from experiment 1 participated in this experiment, but only four of the six completed all of the conditions. The stimuli used in the experiment were identical to stimuli used in experiment 1. Masked thresholds for each component in the presence of the other nine components of the complex were obtained using a 2IFC adaptive procedure to estimate the 70.7% point on a psychometric function (Levitt, 1971). In the two-interval task, the “incorrect” interval contained the nine non-target components of a complex, each at 45 dB SPL before correcting for the headphone response, while the “correct” interval contained the target tone, along with the other nine components. The level of the non-signal complex was roved randomly across trials with a

range of ± 3 dB with 1 dB steps to avoid use of overall loudness as a cue for detection of the addition of a signal frequency. The two intervals, each 500 ms long, were separated by 500 ms of silence. Listening intervals were marked by lights and each trial was preceded by a 500 ms light indicating the beginning of a trial, as in experiment 1. Listeners were instructed to choose the interval that contained a signal tone. Correct-answer feedback was provided after the listener’s response.

Each block of trials began with the target tone presented at 60 dB SPL, so that the listener could clearly hear out the signal tone. The step size was 5 dB until after the second reversal and 2 dB for the remaining trials. Each block consisted of 50 trials and four threshold estimates were obtained for each component. The task is described in Fig. 4(A). Thresholds were obtained for all the components in a random order.

B. Results

Average thresholds for each component of both complexes did not differ much across the listeners. Therefore the arithmetic means across listeners are shown in Fig. 5. Moore *et al.* (1997) described a version of the loudness model that can be used to predict the loudness of one component of a

FIG. 4. (Color online) (A) An illustration of the task used to obtain the masked threshold for individual components. The signal interval included the signal tone at a level which was changed adaptively. The non-signal interval contained only 9 tones presented at equal level with mean of 45 dB/component and their levels were roved across trials in 3 dB range. (B) Illustration of the loudness matching task used in experiment 3. The standard complex included one component that was either incremented or decremented by 8 dB. The level of the comparison tone complex was changed adaptively to match the loudness of the standard complex.

FIG. 5. (Color online) Mean masked thresholds and their standard deviation for individual components of the ten-tone complexes in experiment 2. Predictions of the partial loudness model with a criterion of 5 phons are shown with dashed lines.

complex stimulus in the presence of other components and noted that partial loudness can be used to estimate threshold, if one assumes that the signal at threshold has some uniform low loudness. They proposed 2 phons as an appropriate value. The partial-loudness model was used to predict the masked thresholds for all the components for individual subjects, for assumed threshold criteria of 1 to 10 phons. A criterion value that minimized the least square error between model predictions and observed thresholds was then determined for individual listeners. The optimum values for loudness at threshold for the four listeners ranged from 4 to 6 phons. Mean thresholds predicted by the partial loudness model using a loudness criterion of 5 phons are shown in Fig. 5. This criterion provides the best fit to the mean data of listeners for all the frequencies in both complexes. This value is higher than the 2-phon criterion proposed by Moore *et al.*, possibly due to the level rove in the non-signal interval.

The pattern of thresholds for individual components clearly differed for the low and high complexes. Thresholds for the lowest components in the low complex were elevated because absolute thresholds were higher for frequencies below 500 Hz. Thresholds near the high-frequency edge of both complexes were lower than those closer to the center of the complex. The thresholds of the six components common to both complexes did not differ significantly as a group across the two complexes [$F(1,7.102) = 0.79$; $p = 0.403$], but did differ across the six components [$F(5,13.03) = 4.89$; $p = 0.0098$]. The interaction of relative frequency and absolute frequency was not significant [$F(5,13.03) = 2.12$; $p = 0.127$]. A repeated-measures ANOVA of data for all ten components comparing thresholds components in the low- and high-frequency complexes in terms of their relative position showed significant differences between the low and high complexes [$F(1,18.2) = 6.01$, $p = 0.024$], a significant effect of component position [$F(9,40.56) = 2.84$, $p = 0.011$] and a significant complex by position interaction [$F(9,40.56) = 4.11$, $p = 0.0008$]. As seen in Fig. 5, the partial loudness model correctly predicts that thresholds of the lowest components of the low complex are elevated relative to the lowest components of the high complex. This is consistent with the loudness model prediction

(Fig. 2) that the lowest components of the low complex will make a smaller contribution to overall loudness than higher frequency components in the same complex. The perceptual-weights data do not agree with this prediction. An analysis of the perceptual weights from experiment 1 for the four listeners with complete data in experiment 2, using threshold as an additional predictor of the perceptual weights, showed no significant difference in overall weight between the complexes or relative position within the complexes and a non-significant interaction of relative and absolute frequency, as in the earlier analysis. Not surprisingly, this analysis showed no effect of threshold on perceptual weight [$F(1,18.62) = 0.30$; $p = 0.59$].

IV. EXPERIMENT 3: CONTRIBUTIONS TO TOTAL LOUDNESS ESTIMATED FROM LOUDNESS MATCHING

Results of experiment 1 suggest that perceptual weights obtained at lower-edge-frequencies in the loudness task are determined by their relative position within a broadband signal. Results of the second experiment show that the perceptual weights observed in the first experiment are not due to the audibility or masking of individual components by one another. The threshold configurations differ significantly from one complex to the other while the weights do not. In line with suggestions from Jesteadt *et al.* (2014) and Oberfeld *et al.* (2012), results of the first two experiments suggest that the global loudness of broadband sounds is based on an unequally weighted sum of specific loudness across the auditory filters.

If all the frequencies do not contribute equally to the overall loudness, this should also be seen in loudness matching. For example, an X dB increment or decrement of a component that receives higher weight should produce more change in overall loudness than the same increment or decrement of a component that receives lower weight. In this experiment, we determined the change in loudness level of ten-tone complexes resulting from 8-dB increments or decrements to individual components.

A. Methods

The same six listeners from experiment 1 participated in this experiment. The stimuli used in the experiment were identical to stimuli used in experiment 1. The effects of an 8-dB increment or decrement in the level of a single component (component number 1, 2, 5, 6, 9, and 10) on the overall loudness of the complex were tested in two separate conditions. An 8-dB increment on one component produced a 1.85-dB increase in the overall level of the complex while an 8-dB decrement produced a 0.58-dB decrease.

Loudness matching was carried out in a two-IFC task using interleaved adaptive tracks (Jesteadt, 1980). The test complex, depending on the condition, contained either an increment or a decrement of 8 dB on one component of the ten-tone complex. The remaining nine components were presented at 45 dB SPL. The comparison complex contained all of the components at equal level and that level was adapted to determine the level required for equal loudness. The listener's task was to indicate the interval that was louder [Fig. 4(B)]. Data were collected in 100-trial blocks, 50 trials for each of the two interleaved tracks. On one track, the beginning level of the comparison complex was 35 dB/component, and was increased adaptively, while on the other track, the beginning level was 55 dB/component and was decreased adaptively. The initial step size of 8 dB was reduced to 2 dB after two reversals. The 70.7% and 29.1% points were estimated from the means of the reversals after the decrease in step size, and the average of the two was used to determine the level of the complex with ten equal-level components required to match the loudness of the complex where the level of one component had been altered. Each listener completed two loudness matches for each of the six components, in random order.

Each trial consisted of a 500 ms warning interval and two observation intervals, each 500 ms long, which were separated by 500 ms of silence. An LED light on a response box marked each interval. Listeners received no feedback regarding their responses.

B. Results

The mean loudness matches across listeners for an increment or decrement of 8 dB on various components of a

multi-tone complex are shown in Fig. 6, along with predictions from the loudness model. Upward pointing triangles show conditions where 8 dB was added to the component and downward pointing triangles show conditions where 8 dB was subtracted from the component. It is not surprising that complexes with one component incremented by 8 dB resulted in higher loudness matches than those with one component decremented by 8 dB. The data were analyzed, therefore, in terms of the absolute magnitude of the deviations from the 55 dB midpoint. A repeated measures ANOVA was used to assess the effects of absolute frequency (low vs high complex), relative frequency (the six positions within each complex), the direction of the change in level (−8 dB vs +8 dB) and the pairwise interactions of those three main effects. The magnitude of the deviations from the 55 dB midpoint did not differ for the low vs high complexes [$F(1,16.41) = 1.03$; $p = 0.326$], but the effect of an 8 dB decrement was greater than that of an increment [$F(1,16.41) = 6.05$; $p = 0.025$] and the effect of relative position within each complex was significant as well [$F(5,54.24) = 8.42$; $p < 0.0001$]. The difference between the effect of an increment or decrement did not differ for low vs high complexes [$F(1,17.18) = 0.88$; $p = 0.360$] or as a function of relative position within the complexes [$F(5,54.24) = 1.09$; $p = 0.376$], but the effect of relative position was marginally greater for low complexes compared to high complexes [$F(5,54.24) = 2.17$; $p = 0.071$].

The shape of the loudness matching pattern is consistent with the shape of the pattern of perceptual weights observed in experiment 1 and the earlier studies (Jesteadt et al., 2014; Leibold et al., 2007). The edge components that receive higher weight also produced a larger change in the overall loudness, compared to the components in the middle of the same complex. Statistical analyses incorporating both perceptual weights and matching data, however, failed to show any significant relation between the two measures.

V. DISCUSSION

Perceptual weights for loudness were obtained for two different sets of stimuli in experiment 1. Even though the task used to obtain the perceptual weights was subjective in

FIG. 6. (Color online) Mean levels (and their standard deviations) of comparison complexes required to match the loudness of standard complexes in experiment 3. Upward triangles show conditions where the standard complex included one component with an increment of 8 dB and downward triangles show conditions with a decrement of 8 dB. Predictions of the loudness model are shown with dashed lines.

the perspective of the listener, the presence of an average difference in level between the two observation intervals made it possible to designate a “correct” and an “incorrect” interval. In the correct interval, the level of each component was drawn from a distribution whose mean was $\mu + \Delta$ dB [Fig. 1(B)], with $\Delta = 3$. The only difference between a pure “sample discrimination” task and the task used in experiment 1 is that listeners did not receive feedback regarding the “correct” answer. [Jesteadt et al. \(2014\)](#) showed that the reliability of listeners weighting patterns could be assessed with percent correct performance in the “loudness comparison” task and that moving from loudness judgments in a task with no net difference in level between the two observation intervals to loudness comparison did not alter normalized perceptual weights. Using percent-correct responses, we calculated d' values for individual listeners for experiment 1. The mean value was 1.03 for the low complex and 1.20 for the high. For further comparison, we calculated performance of the ideal observer using Eq. (2). In this equation, N is the number of components, Δ is the increment in dB, and σ is the standard deviation of the rove²

$$d'_{\text{ideal}} = \sqrt{\sum_{i=1}^N \left(\frac{\Delta_i}{\sigma_i} \right)^2}. \quad (2)$$

For a given a rectangular distribution with a range of 16 dB, $\sigma = 4.62$ and Eq. (2) predicts $d' = 2.05$ for both of the tone-complexes. The difference between the predicted and the observed performance could be attributed to either the larger amount of internal noise in listeners or lack of feedback. A cube root rule ([Lutfi and Gilbertson, 2011](#)) predicts $d' = 1.40$, still higher than the observed values.

Performance may be lower in this task than in other sample discrimination tasks because the components are too close to one another in frequency to be processed independently by the auditory system. This would also contribute noise to the assessment of perceptual weights because changes in the effective level of any given component would be altered by the changes in level of adjacent components. To test this, a second set of weights predicted by the model was obtained by using a version of the model developed by [Chen et al. \(2011\)](#) to estimate the “excitation” in dB for the auditory filter centered on each component, for every trial each listener received. The model of [Chen et al. \(2011\)](#) provides direct access to the estimated excitation pattern but is similar to the model of [Moore and Glasberg \(2004\)](#) in other respects. Weights were then obtained by using the difference between the excitation value for each component between interval 2 and interval 1 to predict the difference in overall loudness estimated by the model for the two intervals. This allowed us to use values that represented an “internal representation” of the stimuli as predicted by the model rather than differences in input level. The weights calculated using this method did not differ from the weights obtained using the level difference between the two intervals, and therefore are not shown. A similar strategy to calculate the perceptual weights using excitation patterns has been used previously by [Lentz \(2006\)](#). Her study also reported no difference

between the weights calculated using excitation patterns and the weights calculated using physical levels of the individual components.

A. Specific loudness and perceptual weights

Although a calculation of specific loudness is an important stage in modelling the overall loudness of broadband sounds, direct measurements of specific loudness have not been possible. Previously, contributions of frequency components have been measured by matching the loudness of band-limited signals to the loudness of various reference signals. For example, [Pollack \(1952\)](#) determined the loudness-level of noise bands of various bandwidths and frequency regions and concluded that the mid-frequencies make larger contributions to the overall loudness of broadband sounds than high or low frequencies. Equal loudness contours are derived using a similar method and show that higher levels of low frequency signals are required to match the loudness of reference mid frequency signals. These results are generally interpreted as indicating that low frequencies carry lower specific loudness than the mid-frequencies, when presented at equal levels. This notion contrasts with the results of recent studies that have obtained relative weights using molecular psychophysical methods in loudness comparison tasks ([Jesteadt et al., 2014](#); [Oberfeld et al., 2012](#)). [Oberfeld et al. \(2012\)](#) considered effects of differences in slopes of loudness functions used to predict the spectral weights for loudness of broadband sounds and found that the steeper loudness functions cannot account for the “edge” emphasis observed. [Jesteadt et al. \(2014\)](#), based on their model analysis, suggested that spectral weights obtained in a loudness comparison task may represent the relative specific loudness of the spectral components. To test this, we obtained loudness matches between equal-level multi-tone complexes and complexes where the level of a test component was either increased or decreased by 8 dB. As seen in Fig. 6, the results of this experiment are qualitatively consistent with the perceptual weights obtained in experiment 1. A change in the level of the components that receive higher weight also results in greater change in the loudness of the overall complex. These results provide a foundation for interpretation of the perceptual weights obtained in this and previous studies and confirm the hypothesis of [Jesteadt et al. \(2014\)](#) that perceptual weights reflect relative specific loudness of spectral components of broadband sounds.

The edge frequencies in multi-tone stimuli have been shown to be either more salient or less salient depending on the psychophysical task. [Moore and Ohgushi \(1993\)](#) found that detecting a presence of mistuned frequencies near the edge of an inharmonic tone complex was easier than in the middle of the tone complex. In a profile analysis task, however, listeners performed more poorly when a signal was added to the edge frequencies than the central frequencies of the complex tones ([Green and Berg, 1991](#)). Results presented here demonstrate edge effects in the contribution of individual components to the loudness of broadband sounds. The masked thresholds obtained in experiment 2 rule out the effect of audibility of the edge components as a reason for

such effects. The edge emphasis observed in perceptual weights from experiment 1 and loudness matching results from experiment 3 suggest some mechanism that enhances the contribution of spectral edges to the loudness of broadband sounds.

B. Peripheral vs central auditory processing

Loudness is commonly described as a product of the overall neural excitation produced by a sound. Importantly, this definition does not specify a stage in the auditory system at which the loudness is being calculated from the excitation. Many investigators have attempted to relate psychophysical loudness functions to input-output characteristics of the cochlea. It is reasonable to assume that the perceived loudness reflects the effective output of cochlear processing, and therefore these studies have achieved a certain amount of success in relating loudness functions for “narrowband signals” to specific physiological measures such as otoacoustic emissions and auditory brainstem responses (Epstein and Silva, 2009; Neely *et al.*, 2003; Rasetshwane *et al.*, 2013; Schlauch *et al.*, 1998; Thorson *et al.*, 2012). These schemes, however, have never been tested for signals that are broadband in nature. Earlier studies have shown that auditory nerve responses cannot account for the trends observed spectral summation of loudness (Pickles, 1983; Relkin and Doucet, 1997). Recent neuroimaging studies using fMRI to investigate the neural correlates of loudness perception found that the auditory brainstem activation showed a linear relationship with the physical levels of the broadband signals while the nonlinearities observed in spectral summation were only seen in the auditory cortex activation (Langers *et al.*, 2007; Röhl *et al.*, 2011; Röhl and Uppenkamp, 2012). They concluded that the spectral summation of loudness takes place in the auditory cortex. Ernst *et al.* (2008) presented a tone in background noise at several signal-to-noise ratios and measured cortical activation at various regions of interest using fMRI. Their data showed that separate regions were activated when the tone was at a level close to its threshold (lateral Heschl’s gyrus) compared to above threshold (Planum temporale). They concluded that the activation in the Planum temporale was likely related to the increase in specific loudness of the tone presented in noise.

A motivation for this study was to explore the use of perceptual weights as a more objective measure of loudness for broadband sounds that might be clinically useful for the fitting of hearing aids. The results suggest that measures of the loudness of narrowband signals presented in isolation may not be applicable to real world sounds, which are broadband in nature. Our results support the hypothesis of Leibold *et al.* (2007) that the perceptual weights for widely spaced stimuli reflect the influence of more “central” processing stages that are involved in spectral summation of loudness. It is unclear, however, at what stage in the auditory pathway this dichotomous coding of loudness may arise. Further investigation in this matter would require physiological recordings at more central stages of auditory processing that might be responsible for the weighting of loudness across frequencies.

C. Implications for models of loudness and intensity perception

The matching data suggest that loudness changes produced by a component as part of a larger broadband complex are significantly different than when measured for that same component in isolation. Further experiments that can explicitly test the effect of such spectral context on loudness of the individual components may allow us to understand the rules that govern loudness of broadband sounds. In light of these results, it becomes clear why loudness models can reproduce the equal loudness contours and slopes of the loudness functions with accuracy, yet fail to reproduce the perceptual weights observed in the data. The models have been calibrated with data that predominantly reflect the loudness of narrowband sounds in isolation. Moore’s model has been revised to reflect more realistic assumptions regarding nonlinearities in cochlear processing (Chen *et al.*, 2011), yet these modifications do not improve the predictions of perceptual weights. This supports our hypothesis that the contribution of the individual components to overall loudness is influenced by additional processes central to the auditory nerve, aspects that are missing in the loudness models. Inclusion of a more central stage that will account for changes in contributions to loudness of frequencies presented in the context the other sounds may improve the predictions of the loudness model over a wide range of conditions.

VI. CONCLUSIONS

An overall goal of the experiments reported here was to further investigate the feasibility of using perceptual weights to assess the relative contribution of specific frequency bands to the overall loudness of broadband sounds. Psychophysical data and model simulations are presented in three experiments. Based on the data reported here and the review of the literature, we conclude as follows:

- (1) Specific loudness is weighted across frequency channels before spectral integration of loudness.
- (2) This weighting is dependent on the relative position in the complex signal and not the absolute frequency of the components.
- (3) Such weighting is likely due to more central stages of auditory processing that provide gain to the spectral edges.
- (4) Perceptual weights provide a reliable and fast estimate of the quantity that is referred to as specific loudness.
- (5) Loudness models fail to predict these weights, possibly due to a lack of central processing stages.

ACKNOWLEDGMENTS

The authors would like to thank Tom Creutz and Hongyang Tan for their help with the data collection and data analysis software, Robin High for his help with statistical analyses, Brian Moore and Brian Glasberg for sharing the code for their loudness model, and Dr. Robert Lutfi and an anonymous reviewer for helpful comments on the manuscript. This research was supported by NIH Grants Nos. RO1 DC 011806 and P30 DC 004662.

- ¹The ANOVAs reported here were run with PROC GLIMMIX using SAS/STAT software, v. 9.4. Adjustments for multiplicity were made with the simulation technique (Westfall *et al.*, 2011). The Kenward-Roger adjustment was used, as is standard in repeated measures, resulting in fractional degrees of freedom.
- ²Equation (2) is not strictly correct because the stimulus distributions were uniform rather than Gaussian. Because the calculation involves 20 such distributions, however, 10 in each observation interval, use of a Gaussian approximation is appropriate.
- ANSI (2007). S3.4-2007, *Procedure for the Computation of Loudness of Steady Sounds* (American National Standards Institute, New York).
- 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," *Acta Acust. Acust.* **88**, 378–386.
- Chen, Z., Hu, G., Glasberg, B. R., and Moore, B. C. (2011). "A new method of calculating auditory excitation patterns and loudness for steady sounds," *Hear. Res.* **282**, 204–215.
- Epstein, M., and Silva, I. (2009). "Analysis of parameters for the estimation of loudness from tone burst otoacoustic emissions," *J. Acoust. Soc. Am.* **125**, 3855–3864.
- Ernst, S. M., Verhey, J. L., and Uppenkamp, S. (2008). "Spatial dissociation of changes of level and signal-to-noise ratio in auditory cortex for tones in noise," *Neuroimage* **43**, 321–328.
- Florentine, M. (2011). "Loudness," in *Loudness*, edited by M. Florentine (Springer, New York), pp. 1–15.
- Glasberg, B. R., and Moore, B. C. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- Glasberg, B. R., and Moore, B. C. (2006). "Prediction of absolute thresholds and equal-loudness contours using a modified loudness model," *J. Acoust. Soc. Am.* **120**, 585–588.
- Green, D. M., and Berg, B. G. (1991). "Spectral weights and the profile bowl," *Q. J. Exp. Psychol. A* **43**, 449–458.
- Hellman, R. P., Scharf, B., Teghtsoonian, M., and Teghtsoonian, M. (1987). "On the relation between the growth of loudness and the discrimination of intensity for pure tones," *J. Acoust. Soc. Am.* **82**, 448–453.
- Hellman, R. P., and Zwischlocki, J. J. (1968). "Loudness determination at low sound frequencies," *J. Acoust. Soc. Am.* **43**, 60–64.
- ISO (2003). 226, *Acoustics—Normal Equal-loudness Contours* (International Organization for Standardization, Geneva).
- Jesteadt, W. (1980). "An adaptive procedure for subjective judgments," *Percept. Psychophys.* **28**, 85–88.
- Jesteadt, W., and Leibold, L. J. (2011). "Loudness in the laboratory, Part I: Steady-state sounds," in *Loudness*, edited by M. Florentine (Springer, New York), pp. 109–144.
- Jesteadt, W., Valente, D. L., Joshi, S. N., and Schmid, K. K. (2014). "Perceptual weights for loudness judgments of six-tone complexes," *J. Acoust. Soc. Am.* **136**, 728–735.
- Jesteadt, W., Wier, C. C., and Green, D. M. (1977). "Intensity discrimination as a function of frequency and sensation level," *J. Acoust. Soc. Am.* **61**, 169–177.
- Kochkin, S. (2010). "MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing," *Hear. J.* **63**, 19–32.
- Kortekaas, R., Buus, S., and Florentine, M. (2003). "Perceptual weights in auditory level discrimination," *J. Acoust. Soc. Am.* **113**, 3306–3322.
- Langers, D. R., van Dijk, P., Schoenmaker, E. S., and Backes, W. H. (2007). "fMRI activation in relation to sound intensity and loudness," *Neuroimage* **35**, 709–718.
- 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. (2006). "Spectral-peak selection in spectral-shape discrimination by normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **120**, 945–956.
- 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 Gilbertson, L. (2011). "Effect of decision weights and internal noise on the growth of d' with N ," *J. Acoust. Soc. Am.* **130**, EL329–EL333.
- 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.
- Moore, B. C., and Glasberg, B. R. (2004). "A revised model of loudness perception applied to cochlear hearing loss," *Hear. Res.* **188**, 70–88.
- Moore, B. C., 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., and Sèk, A. (2013). "Comparison of the CAM2 and NAL-NL2 hearing aid fitting methods," *Ear Hear.* **34**, 83–95.
- Moore, B. C. J., and Ohgushi, K. (1993). "Audibility of partials in inharmonic complex tones," *J. Acoust. Soc. Am.* **93**, 452–461.
- Neely, S. T., Gorga, M. P., and Dorn, P. A. (2003). "Cochlear compression estimates from measurements of distortion-product otoacoustic emissions," *J. Acoust. Soc. Am.* **114**, 1499–1507.
- Oberfeld, D., Heeren, W., Rennie, J., and Verhey, J. (2012). "Spectro-temporal weighting of loudness," *PLoS One* **7**, 1–14.
- Oberfeld, D., Kuta, M., and Jesteadt, W. (2013). "Factors limiting performance in a multitoned intensity-discrimination task: Disentangling non-optimal decision weights and increased internal noise," *PLoS One* **8**, 1–14.
- Pickles, J. O. (1983). "Auditory-nerve correlates of loudness summation with stimulus bandwidth, in normal and pathological cochleae," *Hearing Res.* **12**, 239–250.
- Pollack, I. (1952). "The loudness of bands of noise," *J. Acoust. Soc. Am.* **24**, 533–538.
- Rasetshwane, D. M., Neely, S. T., Kopun, J. G., and Gorga, M. P. (2013). "Relation of distortion-product otoacoustic emission input-output functions to loudness," *J. Acoust. Soc. Am.* **134**, 369–383.
- Relkin, E. M., and Doucet, J. R. (1997). "Is loudness simply proportional to the auditory nerve spike count?," *J. Acoust. Soc. Am.* **101**, 2735–2740.
- 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.
- Röhl, M., Kollmeier, B., and Uppenkamp, S. (2011). "Spectral loudness summation takes place in the primary auditory cortex," *Hum. Brain Mapping* **32**, 1483–1496.
- Röhl, M., and Uppenkamp, S. (2012). "Neural coding of sound intensity and loudness in the human auditory system," *J. Assoc. Res. Otolaryngol.* **13**, 369–379.
- Scharf, B. (1978). "Loudness," in *Handbook of Perception: Vol. 4. Hearing*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), pp. 187–242.
- Schlauch, R. S., DiGiovanni, J. J., and Ries, D. T. (1998). "Basilar membrane nonlinearity and loudness," *J. Acoust. Soc. Am.* **103**, 2010–2020.
- Thorson, M. J., Kopun, J. G., Neely, S. T., Tan, H., and Gorga, M. P. (2012). "Reliability of distortion-product otoacoustic emissions and their relation to loudness," *J. Acoust. Soc. Am.* **131**, 1282–1295.
- Turner, M. D., and Berg, B. G. (2007). "Temporal limits of level dominance in a sample discrimination task," *J. Acoust. Soc. Am.* **121**, 1848–1851.
- Westfall, P., Tobias, R. D., and Wolfinger, R. D. (2011). *Multiple Comparisons and Multiple Tests Using SAS@R*, 2nd ed. (SAS Institute, Cary, NC).
- Zwicker, E. (1958). "Über psychologische und methodische Grundlagen der Lautheit" ("On the psychological and methodological bases of loudness"), *Acustica* **8**, 237–258.
- Zwicker, E., Flottorp, G., and Stevens, S. S. (1957). "Critical bandwidth in loudness summation," *J. Acoust. Soc. Am.* **29**, 548–557.
- Zwicker, E., and Scharf, B. (1965). "A model of loudness summation," *Psychol. Rev.* **72**, 3–26.