Contribution of frequency bands to the loudness of broadband sounds: Tonal and noise stimuli^{a)}

Walt Jesteadt,^{1,b)} Marcin Wróblewski,^{1,c)} and Robin High²

¹Center for Hearing Research, Boys Town National Research Hospital, Omaha, Nebraska 68131, USA

²Department of Biostatistics, College of Public Health, University of Nebraska Medical Center, Omaha, Nebraska 68198, USA

(Received 21 January 2019; revised 6 May 2019; accepted 28 May 2019; published online 19 June 2019)

Contributions of individual frequency bands to judgments of total loudness can be assessed by varying the level of each band independently from one presentation to the next and determining the relation between the change in level of each band and the loudness judgment. In a previous study, measures of perceptual weight obtained in this way for noise stimuli consisting of 15 bands showed greater weight associated with the highest and lowest bands than loudness models would predict. This was true even for noise with the long-term average speech spectrum, where the highest band contained little energy. One explanation is that listeners were basing decisions on some attribute other than loudness. The current study replicated earlier results for noise stimuli and included conditions using 15 tones located at the center frequencies of the noise bands. Although the two types of stimuli sound very different, the patterns of perceptual weight were nearly identical, suggesting that both sets of results are based on loudness judgments and that the edge bands play an important role in those judgments. The importance of the highest band was confirmed in a loudness-matching task involving all combinations of noise and tonal stimuli. © 2019 Acoustical Society of America.

<https://doi.org/10.1121/1.5111751>

[JLJ]

Pages: 3586–3594

I. INTRODUCTION

A recent paper describes use of a perceptual weights paradigm to determine the relative contributions of different frequency bands to the overall loudness of broadband stimuli made up of 15 adjacent bands of noise, with either equal levels per band or band levels adjusted to approximate the long-term average speech spectrum (LTASS) (Jesteadt *et al.*, 2017). Listeners were presented with a pair of sounds on each trial and the level of each of the 15 two-equivalent-rectangular-bandwidths (ERB) wide (Glasberg and Moore, 1990) bands was varied randomly over a range of ± 6 dB from one presentation to the next. The contribution of each band to total loudness, or its perceptual weight, can be estimated from the correlation between the difference in the level of that band from interval to interval and the listener's judgment of which interval contained the louder sound (Berg, 1990; Lutfi, 1995; Richards and Zhu, 1994).

A major question in the interpretation of the data obtained by Jesteadt *et al.* (2017) is whether listeners were basing their decisions on loudness or on some other attribute of the sounds, such as timbre or pitch. The task requires decisions concerning sounds that are similar in loudness. In an effort to be consistent, listeners may rely on some other cue.

The same concern applies to data obtained in loudness matching tasks. In the current study, data were obtained with both perceptual-weight and matching procedures using the 15-band noise stimuli used by Jesteadt *et al.* (2017) and tone complexes made up of 15 tones at the center frequencies of the noise bands. The two types of stimuli sound very different and might be expected to have different timbre and pitch cues, but they produce similar results when evaluated with current loudness models (Moore, 2014; Moore *et al.*, 1997). Observation of similar results for the two types of stimuli in measures of perceptual weight and in loudness matches would provide support for the assumption that listeners' loudness judgments are in fact based on loudness.

II. EXPERIMENT 1: PERCEPTUAL WEIGHTS FOR TONAL AND NOISE STIMULI

A. Listeners

Nine listeners, four males and five females with a mean age of 39, participated in the study. All had audiometric thresholds at or below 15 dB Hearing Level (HL) at the octave audiometric frequencies. Five of the listeners had participated in an earlier study in which perceptual weights were obtained for noise stimuli (Jesteadt *et al.*, 2017). The experimental procedures were approved by the institutional review board at the Boys Town National Research Hospital.

B. Stimuli

The stimuli were generated digitally in MATLAB at a sampling rate of 44.1 kHz and were presented using 24-bit digital-to-analog convertors (Digital Audio Labs, CardDeluxe),

^{a)}Portions of this work were presented at the Annual Meeting of the American Auditory Society in Scottsdale, AZ, USA March 2016, and at the 171st Meeting of the Acoustical Society of America in Salt Lake City, UT, USA, May 2016.

^{b)}Electronic mail: walt.jesteadt@boystown.org

^{c)}Also at: Pacific University, College of Health Professions, School of Audiology, Hillsboro, OR 97123, USA.

through a PreSonus HP4 headphone distribution amplifier (Baton Rouge, LA), to one ear using Sennheiser HD-25-1 II headphones (Wedemark, Germany). Each noise or tone complex stimulus had a total duration of 1 s, with 100-ms cosine-squared ramps. The noise stimuli were made up of 15 frequency bands, each 2-ERB wide, as described in Table I. The bands were created separately by adding sinusoids with 1-Hz spacing in random phase. In one spectral condition, the mean level for each band was chosen to approximate the average diffuse-field LTASS spectrum for male and female talkers given by Moore *et al.* (2008). In the remaining three conditions with equal levels for all 15 bands, the mean levels per band were 43, 51, or 59 dB Sound Pressure Level (SPL). These levels were chosen to span the range of levels represented in the LTASS stimuli. Additional details regarding calibration can be found in Jesteadt *et al.* (2017). The tone complex stimuli were made up of 15 sinusoids at the center frequencies of the noise bands, with mean levels equal to the mean levels of the corresponding noise bands.

C. Methods

All testing was completed in a double-walled sound attenuating booth. An individual trial consisted of a 500-ms warning interval, two 1-s observation intervals separated by 500-ms, and an answer interval in which the listener indicated which interval contained the louder noise. The intervals were marked on a simulated response box displayed on a computer screen and the listener responded using a mouse. The levels of the 15 noise bands or tones were varied independently over a range of ± 6 dB in 2-dB steps in each interval of a given trial. In an effort to create uniform stimuli that could be presented in a way that would allow assessment of the consistency of listeners' loudness judgments, 98 stimulus

pairs were created in which each of the seven deviations from mean level (+6, +4, +2, 0, -2, -4, -6 dB) occurred 14 times in each of the 15 bands. The stimulus lists were randomized in five different trial orders and each order was presented twice, so that each stimulus pair was presented ten times in the 980 trials per spectral condition.

The noise or tone complex in one of the two observation intervals, selected at random, was incremented by 1 dB to increase the likelihood that one interval would be perceived as louder than the other. The stimulus with the 1-dB increment was not necessarily the more intense or the louder stimulus because the level of each of the 15 bands varied over a range of ± 6 dB around the mean. No feedback was given following a listener's response. Each listener completed 98-trial blocks for each of the eight conditions (four spectral conditions for noise or tonal stimuli) in random order and testing continued with new random orders until the ten lists per condition had been presented.

D. Results

For every trial, the difference in level between the first and second interval for each of the 15 bands was used to predict the listener's response in a linear model. Weights were calculated separately using multiple linear regression for trials when the interval with a 1-dB increment was first or second. Weights were then normalized such that the mean of the absolute values of each set of 15 weights summed to 1.0 (Kortekaas *et al.*, 2003; Oberfeld *et al.*, 2012). Following a repeated-measures analysis of variance (ANOVA) described below that showed no effect of presentation interval, the normalized weights were averaged across presentation interval, then means and standard errors were computed across listeners. Figure 1 shows mean normalized weights across nine listeners in eight conditions. Data for individual listeners in two conditions (51 dB per band and LTASS) are shown in Fig. 2.

Perceptual weights obtained using a version the Moore *et al.* (1997) loudness model are also shown in Fig. 1. The MATLAB code provided by Dr. Josef Schlittenlacher was written to implement ISO 532-2 (2017). The model was run as described by Jesteadt *et al.* (2017), using 15 tonal components in place of the 15 noise bands. An analysis of results based on four replications of 500 trials each showed no significant differences between weights obtained using the noise-band and tone-complex stimuli. The values in Fig. 1 are weights predicted by the model averaged across the two types of stimuli.

A repeated measures ANOVA using a Kenward-Roger adjustment for the degrees of freedom showed no significant difference between normalized weights obtained from trials where the higher level components (on average) were in the first interval or the second interval and no interaction of Interval with any other variable. A comparison of the means of the four noise conditions with the means of the four tone conditions showed a significant effect of center frequency [$F(14\ 250) = 17.53$; $p < 0.0001$], and a significant interaction of center frequency with stimulus type (noise vs tone complex) [$F(14\ 218) = 4.38$; $p < 0.0001$], but no effect of

TABLE I. End points and center frequencies of the fifteen 2-ERB bands. The lowest and highest frequencies are equivalent to ERB_N values of 3, 5, ..., 31, 33 Cams. The final column gives the diffuse-field level per band of the LTASS noise in dB SPL. These levels result in 1/3rd octave levels that approximate the diffuse-field values given by Moore *et al.* (2008) in Table II, Column 5.

Band Number	Lowest Frequency	Highest Frequency	Center Frequency	LTASS Level/Band
1	87	163	119	54.1
2	164	258	206	57.1
3	259	375	312	55.9
4	376	520	442	59.0
5	521	700	604	58.2
6	701	924	805	54.2
7	925	1201	1054	50.3
8	1202	1545	1363	49.4
9	1546	1972	1746	47.3
10	1973	2502	2222	45.3
11	2503	3158	2811	44.4
12	3159	3973	3543	42.7
13	3974	4984	4450	41.5
14	4985	6239	5577	41.3
15	6240	7795	6974	40.8

FIG. 1. (Color online) Mean normalized perceptual weights for nine listeners for noise and tone-complex stimuli in four spectral conditions. The dashed line in each panel shows results obtained for noise stimuli by Jesteadt *et al.* (2017). The heavy solid line shows weights predicted by the Moore *et al.* (2007) model.

stimulus type [$F(1,218)=0.18$; $p=0.671$]. The lack of a main effect of stimulus type was expected because the normalization process creates approximately equivalent weights across frequency in all conditions. Follow-up comparisons of noise and tone-complex data showed that the only significant difference was at the highest frequency where the weight assigned in the tone-complex conditions was greater than that in the noise conditions [$t(218)=7.16$; $p<0.001$]. An analysis of the effect of presentation level for the noise and tone-complex conditions, excluding the LTASS conditions, showed a marginally significant difference only in the next-to-highest band (centered on 5577 Hz), where the perceptual weight decreased with presentation level in the noise conditions, but increased with level in the tone conditions [$t(190.5)=-1.959$; $p=0.051$]. Once again, the weights for tone complexes were more similar to the earlier data obtained for noise bands (Jesteadt *et al.*, 2017) than to the weights obtained here for noise bands.

E. Discussion

The pattern of results for perceptual weights obtained using 15 bands of noise is similar to that obtained for 15 tones located at the center frequencies of the noise bands. Perceptual weights obtained with the noise stimuli show a significantly larger high-frequency edge effect than the weights obtained with the tone complexes, but the importance of this result is reduced by the fact that the high-frequency edge effect for the noise stimuli is greater than that observed in the earlier study (Jesteadt *et al.*, 2017). The lowest band also receives more weight than loudness models would predict in all but the LTASS condition and that effect

FIG. 2. (Color online) Data for individual listeners in two of the four spectral conditions, 51 dB/band and LTASS. Many listeners show similar patterns of weights in noise and tone-complex conditions.

is similar for noise stimuli and tone complexes. These results do not entirely rule out contributions by factors such as pitch and salience to the loudness judgments, but any model that accounted for the results based on such factors would have to make similar predictions for noise stimuli and tone complexes. The Moore *et al.* (1997) loudness model and later variations (Moore, 2014) predict that the loudness of a stimulus made up of 15 noise bands will be approximately equal to the loudness of a stimulus made up of tones at the center frequencies of the noise bands and therefore predict equivalent perceptual weights for the two types of stimuli, the result that has been obtained. The problem, of course, is that the model does not predict the large observed edge effects for either type of stimulus. As discussed in Jesteadt *et al.* (2017), the

model does predict a high-frequency edge effect in the spectral conditions that have equal power per band as a result of spread of excitation above the highest band. That effect is observable in three panels of Fig. 1, but is too small to account for the data.

In comparing weights obtained with the LTASS spectrum to those obtained in the three spectra with equal levels per band, it is important to consider the fact that higher weights are generally assigned to stimulus components that are higher in level (Berg, 1990; Lutfi and Jesteadt, 2006; Oberfeld, 2015; Oberfeld and Plank, 2011). In a condition where all components were individually matched for loudness, Oberfeld *et al.* (2012) observed a higher weight on the lowest component, but not on the highest component. In the LTASS conditions, where the lowest frequency and highest frequency bands were lower in overall level than those in the mid frequencies, the low-frequency edge effect was eliminated, but the high-frequency edge effect remained, despite the greatly reduced level and loudness of the highest frequency band.

Jesteadt *et al.* (2017, Fig. 9) showed that the loudness model predicts that perceptual weights will be determined by specific loudness or by the loudness of the individual components, those two quantities being highly correlated in these stimuli. The loudness-levels of the 15 noise bands and 15 tones in each of the four spectral conditions are shown in Table II. The patterns as a function of frequency are similar to the perceptual weights generated by the model shown in Fig. 1. The conditions with equal levels per band show greater loudness in the higher frequencies (but not the highest frequency), while the LTASS condition shows greater emphasis in the mid to lower frequencies. It seems likely that adjustments to the levels of individual components in either spectral condition to approximate equal loudness per component would increase the observed low- and high-frequency edge effects.

An effort has been made in earlier studies to confirm the perceptual weights results using a loudness matching procedure (Jesteadt *et al.*, 2017; Joshi *et al.*, 2016). If the highest band receives the most weight, then all other things being

equal, there should be a larger difference in levels required for equal loudness when comparisons involve a stimulus with the highest band removed than when a low or middle band is removed. The matching data supported the importance of the highest band. The agreement between perceptual weights and matching data was far from perfect but was generally greater than that observed in studies that have evaluated correlations between alternative measures of frequency or temporal analysis (e.g., Festen and Plomp, 1981, 1983). The availability of noise and tone-complex stimuli with similar perceptual weights provides an opportunity for a matching study with comparisons involving stimuli of different types as well as those of the same type.

III. EXPERIMENT 2: LOUDNESS MATCHES FOR TONAL AND NOISE STIMULI

A. Listeners

Six listeners, three males and three females with a mean age of 35, participated in the study. Five of the six (561, 562, 567, 585, and 604) participated in experiment 1.

B. Stimuli

The stimuli were noise and tone complexes from the 51 dB SPL/band and LTASS conditions from experiment 1. The stimuli either had all 15 components or were missing bands 1, 8, or 15. For the 15-band stimuli, all bands were at the mean levels used in experiment 1. Levels per band of the 14-band stimuli were increased to equate them with the 15-band stimuli in overall level. For the 51 dB SPL/band condition, the increase was 0.3 dB. For the LTASS condition, the increase depended on the band that had been removed, 0.36 dB for stimuli missing band 1, 0.12 dB for stimuli missing band 8, and 0.02 dB for stimuli missing band 15.

C. Methods

For each spectral condition, 51 dB/band or LTASS, loudness matches were obtained for eight stimuli that were fixed in

TABLE II. Loudness levels in phons of individual noise bands and tones predicted by a version of the Moore *et al.* (1997) loudness model for the conditions in experiment 1. LTASS levels are given in Table I.

Component Number	43 dB SPL per band	43 dB SPL per tone	51 dB SPL per band	51 dB SPL per tone	59 dB SPL per band	59 dB SPL per tone	LTASS noise	LTASS tones
1	22.8	19.9	31.1	27.8	39.4	36.2	34.4	31.0
2	29.7	27.4	37.3	35.3	45.3	43.1	43.3	41.2
3	34.4	31.6	41.8	39.2	49.5	46.9	46.5	43.9
4	37.1	34.9	44.7	42.2	52.0	50.1	52.0	50.1
5	39.0	36.4	46.3	43.9	53.8	51.6	53.0	50.8
6	40.4	37.9	47.7	45.6	55.2	53.2	50.7	48.5
7	41.1	38.7	48.5	46.2	55.9	54.0	47.8	45.6
8	41.1	38.6	48.5	46.2	56.0	53.9	47.0	44.8
9	41.4	38.9	48.8	46.4	56.2	54.2	47.0	42.9
10	43.0	40.5	50.5	48.0	57.9	55.8	45.3	42.5
11	46.0	43.4	53.4	51.1	60.9	59.0	47.2	44.9
12	46.3	44.1	53.8	51.7	61.2	59.6	46.1	43.8
13	44.9	42.3	52.1	50.2	59.7	57.8	43.3	41.0
14	41.4	39.0	48.8	46.5	56.2	54.3	40.0	37.3
15	37.9	35.5	45.4	42.9	52.6	50.6	35.9	33.4

level and served as standards. The eight stimuli were noises or tone complexes with all 15 bands present or bands 1, 8, or 15 missing. The comparison stimuli, adapted in level, were either noises or tone complexes with all 15 bands present. Loudness matches were obtained for noises or tone complexes serving as standards paired with noises or tone complexes serving as comparison stimuli. The order of presentation of the standard and comparison stimuli was randomized from trial to trial. The 16 conditions for each spectrum included four control conditions where both the standard and comparison conditions consisted of 15-band stimuli of the same type or the opposite type.

Matches were obtained using a two-track adaptive procedure to estimate the 29- and 71-percent points on the psychometric function (Jesteadt, 1980). The initial level of the comparison stimulus in one track was 10 dB above the level of the standard stimulus. In the other track, the initial level was 10 dB below the level of the standard stimulus. The initial step size of 4 dB was reduced to 2 dB after the fourth reversal in each track. Each 100-trial block included 50 trials in each track, selected in random order. The means of reversal points obtained with the 2-dB step size were used to estimate the 29- and 71-percent points which were then averaged to estimate the 50-percent point. Two 100-trial blocks were completed for each condition. The conditions were run in random order. The level of the fixed-level standard was subtracted for the 50-percent point for each listener to convert the data to differences in dB required for equal loudness.

D. Results

Figure 3 shows means and standard errors for the six subjects for loudness matches in the four control conditions for each spectrum where all 15 bands were present. Matches of stimuli to themselves are unbiased, as expected. In loudness matches of noise and tone-complex stimuli, the 50-percent point for the tone complexes was below the fixed level of the noise stimuli, while the 50-percent point for the noise stimuli was above the fixed level of the tone complexes. When the two stimulus types were presented at the same level, the tone-complex stimuli were louder on average and a change in level of 2.4 dB was required to equate them in loudness.

FIG. 3. (Color online) Means and standard errors for the six subjects for loudness matches in the four control conditions where all 15 bands were present. For both the 51 dB/band and LTASS spectra, the stimuli consisting of 15 tones are louder than the stimuli consisting of 15 noise bands.

Although data for five of the six listeners indicated that the tonal stimuli were louder than the noise stimuli, data for one listener indicated that the noise stimuli were consistently louder, as shown in Fig. 4.

A repeated measures ANOVA of data for the four conditions where stimuli were matched to themselves showed no significant effects. A repeated measures ANOVA of the data for all six listeners for the four conditions where stimuli consisting of 15 tones were matched to stimuli consisting of 15 noise bands, using a Kenward-Roger adjustment for the degrees of freedom, showed no overall difference between the two spectral conditions [$F(1,7.7) = 0.09$; $p = 0.774$], but a significant effect of fixed-level stimulus [$F(1,8.76) = 14.54$; $p = 0.0043$] and a nearly significant interaction of fixed-level stimulus and spectrum [$F(1,5.66) = 6.04$; $p = 0.052$]. The effect of fixed-level stimulus in these conditions where the comparisons were always between noise and tone stimuli reflects the fact that tonal stimuli were louder than noise stimuli when both stimuli were presented at the same overall level. The interaction of fixed-level stimulus and spectrum reflects a greater effect of the fixed-level stimulus in the LTASS condition than in the 51 dB/band condition.

In the remaining 12 conditions for each spectrum, stimuli with all bands present were matched in loudness to stimuli with one band missing. To separate the effect of stimulus type (tone or noise) from the effect of removing one band, the data for individual listeners for these conditions were corrected for the differences shown in Fig. 4 whenever the comparisons involved a fixed stimulus of one type and an adapted stimulus of the other. Figure 5 shows means and standard errors for the six subjects for loudness matches between stimuli with all 15 bands present and those where band 1, 8, or 15 was missing.

Separate repeated measures ANOVAs were completed for the 51 dB/band and LTASS spectral conditions. For the 51 dB/band conditions, there were significant effects of the band missing [$F(2,8.66) = 16.02$; $p = 0.0012$], the stimulus type that was adapted in level [$F(1,18.8) = 6.90$; $p = 0.0167$], and the stimulus type that was fixed in level [$F(1,18.8) = 16.18$; $p = 0.0007$]. There were no significant interactions. For the LTASS conditions, there were significant effects of the band missing [$F(2,6.53) = 9.68$; $p = 0.0111$], the stimulus type that was

FIG. 4. (Color online) Means and standard errors for two replications for individual listeners for control conditions where the adapted and fixed-level stimuli differed. One listener, 567, found the noise stimuli to be louder than the tone complexes in both spectral conditions.

FIG. 5. (Color online) Means and standard errors of the levels of 15-component stimuli required to match stimuli where one of the 15 components (a noise band or a tone) was removed. Comparisons of noise and tonal stimuli have been adjusted for the observed overall difference in loudness shown in Fig. 3. As in Figs. 3 and 4, conditions where the noise was adapted have darker shading (blue) than conditions where the tones were adapted (red).

adapted in level [$F(1,8.24) = 6.97$; $p = 0.0290$], the stimulus type that was fixed in level [$F(1,8.64) = 4.79$; $p = 0.0576$], and significant interactions of the band missing with both the adapted [$F(2,5.75) = 15.27$; $p = 0.0050$] and fixed-level stimulus [$F(2,6.77) = 9.97$; $p = 0.0096$]. Both interactions reflect the large effect of removing band 8 in the noise-fixed/tones-adapted condition.

The data in Figs. 3–5 are based on dB differences between the 50-percent points on the psychometric functions, estimated by averaging values obtained for the two tracks used to estimate the 29- and 71-percent points. The difference in dB between the two tracks provides a measure of the slope of the psychometric function and an indication of the variability or difficulty associated with the matches. Figure 6 shows the means and standard errors for the six subjects for this slope measure in the four control conditions shown in Fig. 3. As would be expected, subjects showed more variability in decisions when judging the relative loudness of noise and tonal stimuli than when judging the loudness of stimuli of the same type.

Figure 7 shows the data for the slope measure for the conditions shown in Fig. 5. Note that the y axis range has been expanded to accommodate larger values for the LTASS spectrum. The variability is again somewhat greater when the matches involve stimuli of different types, but there is no consistent difference as a function of the band that is missing.

The same version of the Moore *et al.* (1997) loudness model used to predict perceptual weights was used to predict the loudness levels of the 14- and 15-band stimuli. When the overall level of the 15-band stimuli is changed over the range used in this study, the model predicts that loudness level increases by 0.84–0.86 dB for every 1-dB increase in stimulus level. These slope values were used in conjunction

FIG. 6. (Color online) Means and standard errors for the six subjects showing the variability in loudness matches in the four control conditions where all 15 bands were present, as represented by the difference in dB between the 29- and 71-percent points on the psychometric function. A greater difference indicates a shallower psychometric-function slope.

with the differences in loudness level for the 14- and 15-band stimuli to predict the results obtained in the matching conditions. The data in Fig. 3 indicated that when the 15-band tonal and noise stimuli were presented at the same level, the tone-complex stimuli were louder on average and a change in level of 2.4 dB was required to equate them in loudness. The model predicts that the noise stimuli will be louder and that a change of 0.6 dB will be required to equate them in loudness. For the conditions where one band was removed, the model agrees with the majority of the data in predicting the greatest effect for removal of band 15 (0.65 dB), a smaller effect for band 8 (0.45 dB), and no effect for band 1, but the differences are smaller than those observed in the data. The lack of effect for removal of band 1 occurs because the 14- and 15-band

FIG. 7. (Color online) Means and standard errors for the six subjects showing the variability in loudness matches in the conditions where one of the 15 bands was missing, as represented by the difference in dB between the 29- and 71-percent points on the psychometric function. A greater difference indicates a shallower psychometric-function slope. As in earlier figures, conditions where the noise was adapted have darker shading (blue) than conditions where the tones were adapted (red).

stimuli were presented at the same overall level. This adjustment in level was sufficient to balance the change in bandwidth when the lowest band was removed, but not when the highest band was removed.

E. Discussion

Five of the six listeners found broadband stimuli consisting of 15 tones to be louder than stimuli consisting of 15 noise bands. The average difference in level required for a loudness match for all six listeners was 2.4 dB. The loudness model predicts a difference of 0.6 dB in the other direction. The time-domain waveforms of the two types of stimuli appear similar, although the noise stimuli have a higher peak-to-rms ratio. The stimuli consisting of tones clearly sound tonal and it is well established that tonality contributes to both loudness and annoyance (Kryter and Pearsons, 1963; Lee and Wang, 2018; Ryherd and Wang, 2008). Current spectral loudness models do not account for this effect.

Loudness matches in conditions where the stimuli consisted of 15 noise bands or tones presented at 51 dB per band showed a larger effect of removing the highest component and little or no effect of removing the lowest one, despite the weight assigned to the lowest one in that condition in experiment 1. The effect of removing the highest component was greater for the noise stimuli than for the tonal stimuli, perhaps because removal of the highest tonal component may not have contributed to a significant reduction in tonality. The pattern was similar in conditions where stimulus levels followed an LTASS spectrum, with the exception of the large effect observed when the eighth noise band was removed and the adapted stimulus in the loudness match consisted of 15 tones. We cannot account for that result. With the exception of that condition, the pattern of results in conditions where one noise band or tonal component was removed was similar whether the adapted stimulus consisted of noise bands or tones and similar, but attenuated, when the component removed was a tone rather than a noise band.

Loudness matches were compared to perceptual weights in earlier studies reported by Joshi *et al.* (2016) and Jesteadt *et al.* (2017). Joshi *et al.* (2016) obtained perceptual weights for stimuli consisting of ten tones, equal in level, with a frequency range of either 208–2812 Hz or 806–6975 Hz and found that tones at the low- and high-frequency edges received greater weight regardless of their absolute frequency. In the loudness matching portion of the study, they used a two-track adaptive procedure to determine the level of a stimulus with ten tones at 45 dB SPL per tone required to match a stimulus where one of the tones (1, 2, 5, 6, 9, or 10) had been incremented or decremented by 8 dB. They found the greatest effects at the low- and high-frequency edges, but no significant correlation between the matching data and the perceptual weights data for individual listeners. Jesteadt *et al.* (2017) obtained perceptual weights and loudness matches for stimuli consisting of 15 bands of noise, the same stimuli used in the noise conditions of the present study. In the matching portion of the study, they found a large effect of removing the highest band, an intermediate effect of removing band 8 and a small effect of removing the

lowest band. In a follow-up study with stimuli fixed in level, removal of the highest band produced a large effect, but removal of the band 8 or the lowest band produced only a small one. Those results closely parallel the results obtained here in the noise conditions.

The most obvious way to assess the contribution of a given frequency band to total loudness is to remove the band as in this study and in Jesteadt *et al.* (2017). Removing edge bands alters the bandwidth of the stimuli, however, and Jesteadt *et al.* (2017) found that removal of the two highest frequency bands increased the weight assigned to the next highest frequency, complicating the interpretation of the loudness-matching data. If the goal is to explore the effect of level variation in the perceptual weights task by use of loudness matching, it might be preferable to determine the effect of an increase in level rather than a large decrease in level on loudness matches. The only available data where reported by Joshi *et al.* (2016), who found that an 8-dB reduction in an edge component resulted in a greater effect on overall loudness than an 8-dB increment and that both effects were greater than the model would predict.

In general, studies using loudness matching suggest that the highest band makes a greater contribution to overall loudness than loudness models would predict, in agreement with data obtained in studies of perceptual weight. With the exception of Joshi *et al.* (2016), however, the loudness-matching data do not agree with the perceptual-weights data on the importance of the lowest band.

IV. GENERAL DISCUSSION

Perceptual-weights data obtained with tonal and noise stimuli suggest that the highest- and lowest-frequency components make greater contributions to overall loudness than current models of loudness would predict. The agreement between data obtained with 15 tones and 15 noise bands appears to rule out contributions of pitch and timbre to this result. Loudness-matching data confirm the importance of the highest-frequency component, but generally fail to confirm the importance of the lowest-frequency component. An important difference in the stimuli used in the perceptual-weights and matching paradigms is that all bands are varying in level from one presentation to the next in the perceptual-weights paradigm, whereas the bands are fixed in level in the matching paradigm. It is hard to see, however, how variability in the levels of bands above the lowest frequency band could result in that band receiving additional weight in the perceptual weights paradigm. It would be possible to obtain a measure of perceptual weight using stimuli where all components were equal in level with the exception of one component, selected at random, that was completely absent. Those stimuli would be in closer agreement with the stimuli used in the matching task. As an alternative, listeners could be asked to match stimuli in loudness when one of the stimuli had a missing component, but the other components varied in level from one presentation to the next, as in the perceptual weights paradigm. Either approach might result in greater agreement between the two sets of data.

The similarity of perceptual-weights data obtained with noise and tonal stimuli equated in level per ERB_n provides support for some basic assumptions underlying the Moore *et al.* (1997) loudness model. The model assumes that it is the total excitation within an auditory filter that matters, not the distribution of energy across frequencies within the filter. It makes similar predictions for noise and tonal stimuli and the perceptual weights data are in fact similar. The model predicts that the highest frequency component will receive some additional weight as a result of spread of excitation to higher auditory filters that contribute to loudness summation, but fails to account for the size of the effect that is routinely observed.

The interpretation of loudness matches involving comparisons of noise and tonal stimuli was complicated by the fact that tonal stimuli were generally found to be louder than noise stimuli with the same levels per ERB_n , a result not predicted by the loudness model. It is well established, however, that tonal stimuli are found to be louder and more annoying than non-tonal stimuli of the same overall level (see Lee and Wang, 2018 for a recent review). An additional level of cognitive processing might be required to account for the fact that noise and tonal stimuli have the same perceptual weights, as predicted by the loudness model, but not the same overall loudness. The additional level of processing might also account for the exaggerated edge effects observed in the perceptual weights paradigm, consistent individual differences in perceptual weights and other loudness phenomena not addressed by peripheral loudness models.

Cognitive and memory effects have been more widely explored in studies of temporal weighting of loudness, where the initial and final segments of a sound are generally found to receive greater weight (Dittrich and Oberfeld, 2009; Oberfeld, 2009, 2015; Oberfeld and Plank, 2011; Oberfeld *et al.*, 2018; Pedersen and Ellermeier, 2008; Rennies and Verhey, 2009). Pedersen and Ellermeier (2008) found that correct-answer feedback led to a reduction in the weight assigned to the initial and final segments, a shift to more optimum temporal weights. A change in spectrum in the middle of a temporal pattern resulted in increased weight at the onset of the new sound. They noted that these results, combined with individual differences, argued against the “automatic” temporal integration assumed in most loudness models and in favor of individual processing of the temporal segments, as in a multiple-looks model (Viemeister and Wakefield, 1991). Factors such as salience and memory might then play a role at the cognitive level. Oberfeld (2009, 2015) failed to find an effect of feedback on the pattern of temporal weights, but concluded that “higher-level effects like memory and attention” played an important role in the primacy effect, although top-down control was limited.

The primacy effect observed in studies of the temporal weighting of loudness and the high-frequency spectral edge effect for noise and tonal stimuli described here are both resistant to reductions in level. Oberfeld and Plank (2011) found that reduction in the level of early components, which might be expected to reduce the primacy effect because lower-level components generally receive less weight, simply shifted the primacy effect to the initial remaining

components that were equal in level. Likewise, the highest frequency component in the LTASS conditions received additional weight despite its reduced level. Jesteadt *et al.* (2017) found that removing the two highest frequency components caused the additional weight to shift to the next highest component, similar to the temporal shift observed by Oberfeld and Plank (2011). The loudness levels predicted by the model, shown in Table II, suggest that the LTASS stimuli and the stimuli with equal levels per band all decrease in loudness over the three bands below 500 Hz. The decrease in loudness is greater in the LTASS stimuli, which also decrease by 5 dB in level over that low-frequency range. The differences in level or loudness at the low frequencies or the differences across all frequencies are enough to eliminate the low-frequency edge effect in the perceptual weights assigned to the LTASS stimuli but not stimuli with equal levels per band. Despite larger differences in level and loudness at higher frequencies, both types of stimuli show the high-frequency edge effect.

Current loudness models cannot account for all aspects of the temporal or spectral weighting of loudness observed in perceptual-weights data. The issue is whether factors such as salience and tonality that appear to play a role in temporal and spectral weighting are unique to the perceptual-weights paradigm or play a more general role in loudness judgments that should be incorporated in the models. Data obtained with loudness matching provide some support for the view that additional factors contribute to judgments of the loudness of broadband sounds, but are not in total agreement with the perceptual-weights data.

ACKNOWLEDGMENTS

The authors would like to thank Tom Creutz for his help with the data-collection and data-analysis software, Brian Moore, Brian Glasberg, and Josef Schlittenlacher for sharing the code for their loudness models, Katyarina Brunette, Oluwaseye Ogun, Sara Walker, and Katie Thrailkill for help with data collection, and Daniel Oberfeld for helpful comments on a previous version of the paper. This research was supported by NIH Grant Nos. R01 DC 011806, P30 DC 004662, and 1 U54 GM115458, which funds the Great Plains IDEa-CTR Network.

- Berg, B. G. (1990). “Observer efficiency and weights in a multiple observation task,” *J. Acoust. Soc. Am.* **88**, 149–158.
- Dittrich, K., and Oberfeld, D. (2009). “A comparison of the temporal weighting of annoyance and loudness,” *J. Acoust. Soc. Am.* **126**, 3168–3178.
- Festen, J. M., and Plomp, R. (1981). “Relations between auditory functions in normal hearing,” *J. Acoust. Soc. Am.* **70**, 356–369.
- Festen, J. M., and Plomp, R. (1983). “Relations between auditory functions in impaired hearing,” *J. Acoust. Soc. Am.* **73**, 652–662.
- Glasberg, B. R., and Moore, B. C. J. (1990). “Derivation of auditory filter shapes from notched-noise data,” *Hear. Res.* **47**, 103–138.
- ISO (2017). 532-2, *Methods for Calculating Loudness—Part 2: Moore-Glasberg Method* (International Organization for Standardization, Geneva, Switzerland).
- Jesteadt, W. (1980). “An adaptive procedure for subjective judgments,” *Percept. Psychophys.* **28**, 85–88.
- Jesteadt, W., Walker, S. M., Ogun, O. A., Ohlrich, B., Brunette, K. E., Wroblewski, M., and Schmid, K. (2017). “Relative contributions of

- specific frequency bands to the loudness of broadband sounds," *J. Acoust. Soc. Am.* **142**, 1597–1610.
- Joshi, S. N., Wróblewski, M., Schmid, K. K., and Jesteadt, W. (2016). "Effects of relative and absolute frequency in the spectral weighting of loudness," *J. Acoust. Soc. Am.* **139**, 373–383.
- Kortekaas, R., Buus, S., and Florentine, M. (2003). "Perceptual weights in auditory level discrimination," *J. Acoust. Soc. Am.* **113**, 3306–3322.
- Kryter, K. D., and Pearsons, K. S. (1963). "Some effects of spectral content and duration on perceived noise level," *J. Acoust. Soc. Am.* **35**, 866–883.
- Lee, J., and Wang, L. M. (2018). "Development of a model to predict the likelihood of complaints due to assorted tone-in-noise combinations," *J. Acoust. Soc. Am.* **143**, 2697–2707.
- 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.
- Moore, B. C. J. (2014). "Development and current status of the 'Cambridge' loudness models," *Trends Hear.* **18**, 1–29.
- 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., Stone, M. A., Fullgrabe, C., Glasberg, B. R., and Puria, S. (2008). "Spectrotemporal characteristics of speech at high frequencies, and the potential for restoration of audibility to people with mild-to-moderate hearing loss," *Ear Hear.* **29**, 907–922.
- Oberfeld, D. (2009). "The temporal weighting of the loudness of time-varying sounds reflects both sensory and cognitive processes," in *Fechner Day 2009. Proceedings of the 25th Annual Meeting of the International Society for Psychophysics*, edited by M. A. Elliott, S. Antoniević, S. M. Berthaud, B. Bargary, and C. Martyn (Galway, Ireland), pp. 261–266.
- Oberfeld, D. (2015). "Are temporal loudness weights under top-down control? Effects of trial-by-trial feedback," *Acta Acust. united Ac.* **101**, 1105–1115.
- Oberfeld, D., Heeren, W., Rennies, J., and Verhey, J. (2012). "Spectro-temporal weighting of loudness," *PLoS One* **7**, 1–14.
- Oberfeld, D., Hots, J., and Verhey, J. (2018). "Temporal weights in the perception of sound intensity: Effects of sound duration and number of temporal segments," *J. Acoust. Soc. Am.* **143**, 943–953.
- Oberfeld, D., and Plank, T. (2011). "The temporal weighting of loudness: Effects of the level profile," *Atten. Percept. Psychophys.* **73**, 189–208.
- Pedersen, B., and Ellermeier, W. (2008). "Temporal weights in the level discrimination of time-varying sounds," *J. Acoust. Soc. Am.* **123**, 963–972.
- Rennies, J., and Verhey, J. L. (2009). "Temporal weighting in loudness of broadband and narrowband signals," *J. Acoust. Soc. Am.* **126**, 951–954.
- 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.
- Ryherd, E. E., and Wang, L. M. (2008). "Implications of human performance and perception under tonal noise conditions on indoor noise criteria," *J. Acoust. Soc. Am.* **124**, 218–226.
- Viemeister, N. F., and Wakefield, G. H. (1991). "Temporal integration and multiple looks," *J. Acoust. Soc. Am.* **90**, 858–865.