Relative contributions of specific frequency bands to the loudness of broadband sounds^{a)}

Walt Jesteadt,^{1,b)} Sara M. Walker,¹ Oluwaseye A. Ogun,¹ Brenda Ohlrich,¹ Katyarina E. Brunette,¹ Marcin Wróblewski,¹ and Kendra K. Schmid²

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 26 October 2016; revised 31 August 2017; accepted 5 September 2017; published online 25 September 2017)

Listeners with normal hearing (NH) and sensorineural hearing loss (SNHL) were asked to compare pairs of noise stimuli and choose the louder noise in each pair. Each noise was made up of 15, two-ERB_N (equivalent rectangular bandwidth) wide frequency bands that varied independently over a 12-dB range from one presentation to the next. Mean levels of the bands followed the long-term average speech spectrum (LTASS) or were set to 43, 51, or 59 dB sound pressure level (SPL). The relative contribution of each band to the total loudness of the noise was determined by computing the correlation between the difference in levels for a given band on every trial and the listener's decision on that trial. Weights for SNHL listeners were governed by audibility and the spectrum of the noise stimuli, with bands near the spectral peak of the LTASS noise receiving greatest weight. NH listeners assigned greater weight to the lowest and highest bands, an effect that increased with overall level, but did not assign greater weight to bands near the LTASS peak. Additional loudness-matching and paired-comparison studies using stimuli missing one of the 15 bands showed a significant contribution by the highest band, but properties other than loudness may have contributed to the decisions. © 2017 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.5003778>]

[MAS]

Pages: 1597–1610

I. INTRODUCTION

Current loudness models (e.g., Moore *et al.*, 1997; Moore and Glasberg, 2004, 2007; Chalupper and Fastl, 2002), along with earlier models (Stevens, 1956, 1972; Zwicker and Feldtkeller, 1967) and related standards (ANSI S3.4, 2007; ISO 532-2, 2017), attempt to account for the loudness of a large variety of sounds using a fixed set of parameter values. Most models make predictions regarding the loudness of broadband stimuli by assuming that loudness is summed across individual auditory filters or critical bands after correction for spread of excitation. The contribution of individual auditory filter outputs or critical bands to the total loudness is assumed to be governed by the same factors that govern the level and spacing of equal-loudness contours (ISO 226, 2003). The equal-loudness contours themselves were generated using narrowband sounds, typically sinusoids. Zwicker (1958) incorporated data obtained in studies with broadband noise into his initial model, but the majority of the data used to test models of loudness have been obtained with narrowband stimuli.

The ability of the models and standards to correctly predict the loudness of broadband sounds has been verified in studies where listeners rate the loudness of those sounds or adjust the level of a narrowband sound to match the broadband sound in loudness (Kryter and Pearsons, 1963;

Meunier *et al.*, 2000; Schlittenlacher *et al.*, 2011; Schlittenlacher *et al.*, 2015; Stevens, 1972). This approach does not provide a sensitive measure of the contribution of individual bands to total loudness because changes in the level of individual bands have only a small impact on the predicted loudness. Only changes in overall level or large changes in spectral shape result in measurable differences in the model estimates. Recent studies demonstrate that current models overpredict the loudness of broadband sounds and that the effect is greater at higher levels for broadband sounds centered at higher frequencies (Meunier *et al.*, 2000; Schlittenlacher *et al.*, 2015; Schlittenlacher *et al.*, 2017), although one study found that current models were accurate in predicting the loudness of many technical sounds (Rennies *et al.*, 2015).

Two alternative approaches may provide better data regarding the contributions of individual bands. One is to assess the effect of adding or subtracting individual bands using loudness ratings or loudness matching, using a constant-bandwidth stimulus as the standard (Keidser *et al.*, 2002; Holte and Margolis, 1987; Pollack, 1952). The other is to vary the level of individual bands independently and use a reverse correlation approach to determine the contribution of each band to loudness judgments (Jesteadt *et al.*, 2014; Joshi *et al.*, 2016; Leibold *et al.*, 2007; Oberfeld *et al.*, 2012). Results from both approaches will be presented here.

Pollack (1952) appears to have been the first to assess the contribution of individual frequency bands to the loudness of broadband noise. He divided white noise (60–5800 Hz) into ten bands, each 250 mels wide (Stevens *et al.*, 1947), and removed successive bands by filtering. Listeners adjusted the

^{a)}A portion of this work was presented at the Annual Meeting of the American Auditory Society in Scottsdale, AZ, in March, 2015.

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

level of the filtered noise to equate it in loudness to the unfiltered broadband noise. The conditions included a series where low- and high-frequency bands were removed in pairs or successive low- or high-frequency bands were removed individually. The overall level of the broadband noise was varied as well. The results where the bands were removed in pairs show all of the expected effects of loudness summation, predating Zwicker *et al.* (1957). The low- and high-pass noise conditions also showed effects of loudness summation. Pollack found that the mid frequencies (500–3000 Hz) contributed more to total loudness than the lower or higher frequencies, as one would predict from equal-loudness contours, but did not take into account the fact that the stimuli with mid frequencies present were always greater in bandwidth than those without mid frequencies. He concluded that “the loudness of a white noise is increased as its bandwidth is increased *even* when noise frequencies to which the ear is relatively less sensitive are added” (Pollack, 1952, p. 537, emphasis in the original).

The broadband sound of greatest interest is speech. Holte and Margolis (1987) and Keidser *et al.* (2002) have compared the loudness of different frequency bands to one another, but neither assessed the contribution of individual bands to total loudness. Holte and Margolis (1987) recorded speech samples and determined the level of the total speech waveform that was equal in loudness to that of each of seven third-octave bands filtered from the speech waveform. The bands were centered on 250, 400, 800, 1000, 2000, 3150, and 4000 Hz. Keidser *et al.* (2002) filtered different frequency bands from a broadband noise with the long-term average speech spectrum (LTASS) as described by Byrne *et al.* (1994), and had listeners adjust the level of one of six bands to match the loudness of a standard band centered on 1500 Hz. The LTASS noise had an overall level of 65 or 85 dB sound pressure level (SPL) and the bands were either 1 octave or 1 equivalent rectangular bandwidth (ERB_N) wide (Glasberg and Moore, 1990). Keidser *et al.* (2002) corrected the earlier data for differences between the spectrum of the Holte and Margolis (1987) speech sample and the LTASS spectrum and concluded that both sets of data showed that low-frequency bands were louder than those at 1000 Hz and above. They also concluded that the low-frequency bands of the LTASS noise, therefore, made a greater contribution to total loudness and noted that Moore (2000) had reached a similar conclusion.

The methodology used in studies of perceptual weight provides a means of assessing the relative contributions of specific frequency bands to overall loudness when all bands are present, as they would be when listening to speech under normal conditions. Listeners are presented with two broadband sounds on every trial and the levels of frequency bands within the sounds are varied independently from one presentation to the next. The contribution of each band to total loudness is estimated from the correlation between the difference in level of that band on every trial and the listener’s judgment of which interval contained the louder sound (Lutfi, 1995; Richards and Zhu, 1994). This paradigm has been used to explore spectral weighting (Doherty and Lutfi, 1996; Jesteadt *et al.*, 2014; Joshi *et al.*, 2016; Kortekaas *et al.*, 2003;

Leibold *et al.*, 2007; Leibold *et al.*, 2009; Lentz, 2006), temporal weighting (Dittrich and Oberfeld, 2009; Lutfi and Jesteadt, 2006; Oberfeld and Plank, 2011; Oberfeld *et al.*, 2013), and the combination of the two (Oberfeld *et al.*, 2012). In general, the lowest and highest frequency bands and the initial bands in temporal sequences receive greater weight. Higher level bands receive greater weight than lower level bands (Lutfi and Jesteadt, 2006) and spectral and temporal effects are independent of one another (Oberfeld *et al.*, 2012). Many of these studies have used stimuli made up of multiple tones rather than noise bands and none have obtained loudness judgments for speech stimuli or LTASS noise.

Only Doherty and Lutfi (1996) and Jesteadt *et al.* (2014) have included listeners with sensorineural hearing loss (SNHL) in studies of spectral weighting of loudness or intensity perception. In one of the earliest studies in this literature, Doherty and Lutfi (1996) noted that 13 out of 14 SNHL listeners placed greater weight on components of a 6-tone complex that were either the highest frequency in the complex (8000 Hz) or a frequency within a region of steeply sloping hearing loss (4000 Hz). The high-frequency edge effect was not as strong in listeners with normal hearing (NH), although those listeners placed high weight on the lowest component. Leibold *et al.* (2009) noted that the SNHL listeners had been tested at a higher overall level than the NH listeners and that the difference in level may have contributed to the difference in frequency weighting. Jesteadt *et al.* (2014) obtained data from four SNHL listeners using the same six-tone complex as used by Doherty and Lutfi, with tones presented at equal SPL and retested three of the four listeners using tones presented at equal sensation level (SL). In agreement with Doherty and Lutfi, they found greatest weight assigned to the 4000-Hz component in the equal-SPL condition, but large individual differences in the equal-SL condition.

The goal of the present study was to obtain spectral weights for broadband noise stimuli, including LTASS noise, and following Doherty and Lutfi (1996), to compare results obtained for listeners with NH to those for listeners with SNHL.

II. EXPERIMENT 1. PERCEPTUAL WEIGHTS FOR NOISE STIMULI

A. Listeners

There were 16 NH and 16 SNHL listeners that participated in the study. NH listeners, 5 males and 11 females ranging in age from 22 to 55 years old with a mean age of 40 yr, had audiometric thresholds below 15 dB hearing level (HL) at the octave audiometric frequencies from 250 to 8000 Hz. SNHL listeners, nine males and seven females ranging in age from 19 to 70 years old with a mean age of 50 yr, had mild to moderate hearing loss. All listeners had normal middle ear admittance (>0.2 mmho, a unit of static admittance) as measured with tympanometry. All listeners had participated in an earlier study in which they made loudness judgments (Rasetshwane *et al.*, 2015). Audiograms for the 16 SNHL listeners, grouped by the slope of the audiogram, are shown in Fig. 1. All listeners were compensated for their time of participation in the study. The

FIG. 1. Inverted audiograms for 16 listeners with hearing loss, grouped by slope of the hearing loss as a function of frequency. Slopes were obtained from a least-squares fit to the dB HL values as a function of log frequency over the entire range of available frequencies.

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 converters (Digital Audio Labs, CardDeluxe, Chanhassen, MN), through a PreSonus HP4 headphone distribution amplifier (Baton Rouge, LA), to one ear using Sennheiser HD-25-1 II headphones (Wedemark, Germany). Each noise had a total duration of 1 s, with 100-ms cosine-squared ramps, and was made up of 15 frequency bands, each 2-ERB_N wide, as described in Table I. The bands were created separately by adding sinusoids with 1-Hz spacing in random phase. Thirteen independent samples of each band were generated with levels ranging from +6 dB to -6 dB around the mean level, in 1-dB steps. The data collection program then generated 15-band stimuli by selecting 1 of the 13 samples of each band at random and summing those waveforms. In one spectral condition, the mean level for each band was chosen to approximate the average diffuse-field LTASS 1/3-octave levels for male and female talkers given by Moore *et al.* (2008), in their Table II, column 3. The calibration was verified by using KEMAR to determine the levels produced by the headphones at the eardrum. These levels were equivalent to eardrum levels given by Moore *et al.* (2008) in their Table II, column 5. The diffuse-field levels of the current stimuli are shown in Fig. 2. In the remaining three spectral conditions, the mean diffuse-field levels per band were 43, 51, or 59 dB SPL as measured using a Larson Davis AEC101 artificial ear (Depew, NY). These levels were chosen to span the range of levels represented in the LTASS stimuli.

TABLE I. Frequency ranges of the 15 2-ERB_N 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 1/3 octave levels that approximate the diffuse-field values given by Moore *et al.* (2008) in their Table 2, Column 3.

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

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 noise in one of the two observation intervals, selected at random, was incremented by 1 dB to create a nominal correct answer and reduce the uncertainty regarding loudness judgments. The noise with the 1-dB increment was not necessarily the more intense or the louder noise and no feedback was given following a listener's response. Each listener completed five 100-trial blocks for each of the four conditions in random order.

SNHL listeners were given an initial block of 6 trials in the 43-dB/band condition and 5 of the 16 listeners were not tested at that level because they could not hear the stimuli. All 16 listeners were able to hear the stimuli in the 51-dB/band condition, but some bands were inaudible for the SNHL listeners in all of the conditions because the stimuli were fixed in level and there was no compensation for hearing loss.

D. Results

Perceptual weights for the 15 noise bands were obtained in each of 2 ways. In the first, we followed procedures used in prior studies in the auditory research literature, as described recently by Joshi *et al.* (2016). 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 "correct" interval was first or second. Because the signal was

FIG. 2. Acoustic calibration data for the LTASS spectrum condition compared to LTASS diffuse-field levels estimated by Moore *et al.* (2008). Levels per band in the other three spectral conditions are shown in schematic form. Because the 2-ERB_N wide bands increase in width with frequency, the equal-level conditions are similar to pink noise.

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). 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).

Perceptual weights were also obtained following procedures used in the vision research literature, as described by Ahumada (2002), Murray (2011), and others. In this classification-image approach, all stimuli judged to be louder are considered as one group and those judged to be softer are considered as another. The mean level is then computed for each noise band for the stimuli in each group and the difference in mean levels across groups is treated as an estimate of perceptual weight. Ahumada (2002) has shown that these two methods yield the same expected values if the distribution of levels within each band is Gaussian. The distributions in the current data set were rectangular, but mean weights computed with the two approaches were nearly identical and the correlation of one set of weights with the other, across all listeners, conditions and frequency bands ($N=1920$) was 0.96. Weights obtained using multiple linear regression were used in the remainder of the analyses.

Perceptual weights were also obtained using a version of the Moore *et al.* (1997) loudness model. For a given trial, loudness estimates were obtained from the model for 2 15-band stimuli whose levels varied from band to band with the same statistics as the actual stimuli. The model then voted for the louder of the two stimuli and this was repeated for 500 trials. A multiple regression analysis was used to determine the relation between the differences in level for the 15 bands and the votes. Four replications were run for NH listeners. Input to the model was assumed to be monaural with a diffuse-field correction, consistent with the calibration of the stimuli, which were specified to the model as diffuse-

field levels. All estimates of perceptual weights were obtained using MATLAB code that was written by Dr. Josef Schlittenlacher to implement the recent revision of ISO 532-2, based on Moore and Glasberg (2007), but other versions of the model (see Moore, 2014) would have yielded the same results, given a diffuse-field correction.

A repeated-measures analysis of variance (RM-ANOVA) that included data for both NH and SNHL listeners showed no significant difference between normalized weights obtained from trials where the “signal” was in the first interval or the second interval and no interaction of interval with any other variable. There was a significant effect of center frequency [$F(14,453) = 3.49$; $p < 0.001$], a significant interaction of center frequency with spectral condition [$F(42,1018) = 3.38$; $p < 0.001$], a significant interaction of the frequency effect with hearing loss [$F(14,453) = 9.97$; $p < 0.001$], and a significant interaction of the frequency effect with spectral condition and hearing loss [$F(42,1017) = 1.84$; $p = 0.001$]. Figure 3 shows mean perceptual weights for the 16 NH listeners in the 4 conditions. Means and standard errors were computed after averaging across intervals. Dashed lines show weights predicted by the Moore *et al.* (1997) model.

The perceptual weights for specific center frequencies in the LTASS condition were compared to those for the three conditions where all bands were equal in level, treated as a group. Then differences were considered within the group of equal-level conditions. The p -values for individual follow-up comparisons have been adjusted for multiple comparisons using the “simulate” method (Westfall *et al.*, 2011). The weight associated with the lowest center frequency in the LTASS condition was significantly lower ($p = 0.008$) than the average weight at that frequency in the equal-level conditions. None of the other LTASS comparisons reached significance. For the three equal-level conditions, the weight associated with the lowest center frequency in the 43-dB/band condition was significantly lower than the weight for that frequency in the 59-dB/band condition ($p = 0.02$) and marginally lower than the weight for that frequency in the 51-dB/band condition ($p = 0.07$). At 442 Hz (band 4), the weight in the 51-dB/band condition was significantly lower than the weight for that frequency in the 43-dB/band condition ($p = 0.04$). At 2222 Hz (band 10), the weight in the 59-dB/band condition was significantly lower than the weight in the 51-dB/band condition ($p = 0.02$). At the highest center frequency, 6974 Hz, the weight in the 59-dB/band condition was significantly higher than the weight in the 51-dB/band condition ($p < 0.01$) and in the 43-dB/band condition ($p < 0.01$).

In the three conditions where all bands were equal in level, greater weight was assigned to the lowest and highest frequency bands and this effect increased with increasing level per band. In the LTASS condition, greater weight was assigned to the two highest frequency bands than to mid- or low-frequency bands despite the fact that these bands were below the peak of the LTASS spectrum. The peak (at ~ 500 Hz) did not receive additional weight and the lowest frequency band received less weight than in the conditions with equal levels per band. The decreased weight for band 13 (4450 Hz) and increased weight for band 12 in the LTASS condition may have been due to the low level of

FIG. 3. (Color online) Means and standard errors for normalized perceptual weight for 16 listeners with NH in the 4 spectral conditions. The dashed lines show weights obtained using the Moore *et al.* (1997) model.

band 13 that can be seen in Fig. 2. This may have reduced the weight of band 13 and may have increased the weight associated with band 12 as a result of spread of excitation. The effects observed in the mean data reach statistical significance, but the perceptual weights for individual listeners were highly variable as shown for the 51-dB per band and LTASS conditions in Fig. 4. Weights for the remaining two conditions were omitted for clarity.

Figure 5 shows means and standard errors of perceptual weights for the 16 SNHL listeners in the 4 conditions. The grey symbols without error bars in each panel show the mean values for the 16 NH listeners from Fig. 3 for comparison. Despite the variability in the audiograms shown in Fig. 1, the variability in perceptual weights for the SNHL listeners in Fig. 5 is comparable to that for the NH listeners in Fig. 3, particularly in the 51- and 59-dB per band conditions. The significant interaction of the frequency effect with spectral condition and hearing loss noted above is evident in the differences between the two sets of symbols in each panel. The highest frequency bands were given less weight by the SNHL listeners, presumably because they were less audible, but note that the highest band received more weight than the next two highest bands in all but the 59-dB per band condition. The lowest band was assigned less weight in all but the 59-dB per band condition and, unlike the data for the NH listeners, weights in the LTASS condition appeared to follow the LTASS spectrum. Perceptual weights for individual SNHL listeners for the 51-dB per band and LTASS conditions are shown in Fig. 6. As in Fig. 4, differences across listeners were clearly greater than differences across spectral conditions despite the significant effect of spectral condition.

E. Discussion

The data for NH listeners indicate that the lowest frequency band received more weight than the loudness model

predicts for the conditions with equal levels per band, and this effect increased with increasing overall level of the stimuli, whereas the loudness model shows no corresponding level effect. The conditions with equal levels per band show a similar increase in weight with increasing level at the highest frequencies and the LTASS condition shows a high-frequency effect despite reduced energy at high frequencies.

The loudness model predicts smaller contributions from the lowest frequencies, consistent with the extensive data available regarding equal-loudness contours. For broadband stimuli, the model is based on the assumption that the specific loudness associated with individual frequency bands is summed. To a first approximation, total loudness can be predicted by a combination of frequency weighting (consistent with equal-loudness contours) followed by loudness summation. The data suggest that either the assumptions regarding specific loudness or loudness summation are incorrect or that the two interact in a way not captured by the model (Schlittenlacher *et al.*, 2015; Schlittenlacher *et al.*, 2017). Another interpretation, reviewed in Sec. V, is that the use of perceptual weights as a measure of contributions to total loudness is incorrect. The procedures used to estimate perceptual weight also assume loudness summation in that the decision is assumed to be based on the sum of the contributions of the individual bands. Thus, the deviation between model and data is presumably due to the assumptions regarding the specific loudness of individual bands in the presence of other bands or to the validity of the perceptual weights. Using tone complexes, Joshi *et al.* (2016) have shown that the lowest frequency receives additional weight whether it is at 208 or 808 Hz, suggesting that the contribution of specific frequency bands to total loudness depends on their position relative to other tones in the complex. Jesteadt *et al.* (2014) also found greater weight associated with the lowest frequency of a six-tone complex than predicted by the loudness model, but in that case, the weight decreased with increasing level, while the weight associated with the highest frequency increased.

For the conditions with equal levels per band, the model predicts greater weight around 3000 Hz because that is the point at which the diffuse-field correction results in the greatest level at the tympanic membrane. The diffuse-field correction should be playing an equal role, of course, in both the model and the data. It is noteworthy that the model also assigns greater weight to the highest frequency band than to the bands just below it, although not as much weight as actual listeners do. The additional weight associated with the highest frequency band has two potential sources. One source in both the model and real listeners is that the effective level in each 2-ERB_N band is influenced by the levels of the adjacent bands that are fluctuating from presentation to presentation, and these effects are not included in the calculation of perceptual weights. The noise resulting from fluctuations in the level of adjacent bands reduces the estimated perceptual weights. The highest and lowest bands have only one adjacent band and are thus less influenced by this effect. The second source contributing to higher weights in both the model and real listeners is that changes in the level of edge bands influence the level of excitation in bands outside of

FIG. 4. (Color online) Normalized perceptual weights for individual listeners with NH in the 51 dB/band and LTASS conditions. Although the lowest and highest bands are given more weight on average, there are large individual differences in the pattern of weights that are moderately consistent across the two conditions.

the range of the physical stimuli. The additional bands contribute to total loudness, giving more weight to the edge bands. Given upward spread of excitation, this effect should make a greater contribution to the weight associated with the highest frequency band than to the weight associated with the lowest frequency band. If additional bands above 7795 Hz are specified as part of the input to the model, the weight assigned by the model to the band from 6240 to 7795 Hz is reduced and that occurs whether or not the additional bands are assumed to vary in level from interval to interval. This suggests that the second source, upward spread of excitation, is a more important contributor to the high-frequency edge effect in the model than the reduced influence from variability of neighboring bands. The effect of removing bands was explored in

experiment 2, but not the effect of adding bands. It is possible that additional stationary noise above 7795 Hz would reduce the edge effect observed in listeners, but that has not been tested.

Results for the LTASS condition differed from those for the conditions with equal levels per band in that the lowest frequency band received less weight, perhaps because it was presented at a lower level than the adjacent band. The two highest frequency bands in the LTASS condition received more weight than adjacent bands, just as in the conditions with equal levels per band, despite the difference in spectral shape. The loudness model assigns more weight to frequency bands at the peak of the LTASS spectrum and does not assign greater weight to the highest frequency band than to adjacent bands.

FIG. 5. (Color online) Means and standard errors for normalized perceptual weight for 16 listeners with SNHL in the 4 spectral conditions are shown by the black symbols. The grey symbols without error bars in each panel show the mean values for NH listeners from Fig. 3 for comparison.

The perceptual weights for individual listeners in Fig. 4 show greater differences across listeners than across conditions. L4, for example, placed the greatest weight on higher frequency bands above 1000 Hz, with the exception of the band from 3974 to 4984 Hz, with a similar pattern for the 51-dB/band and LTASS conditions. L8 showed the opposite pattern, while weights for L12 were relatively flat with the exception of the highest bands. The individual patterns were highly reproducible.

The mean data for SNHL listeners in Fig. 5 show the effects of hearing loss in reducing the weight assigned to the highest frequency bands, but the weight assigned to the lowest band was sometimes reduced as well. Even though the highest bands were less audible as a result of sloping hearing losses, the data showed increased weight on the highest band in three of the four conditions. The mean data for the LTASS condition showed greatest weight assigned to the spectral peak, a pattern we expected to see in the data for NH listeners, but in this case it was no doubt due to the reduced audibility of the higher frequency bands. The data for individual SNHL listeners in Fig. 6 do not show a consistent relation between audibility and perceptual weight except for cases where the highest bands were inaudible. If the data are pooled across frequency bands and listeners, however, the correlation between perceptual weight and hearing loss at the middle frequency of each band (estimated by interpolation between the nearest two points on the audiogram) was significant for all four conditions. The r values of -0.39 , -0.44 , -0.36 , and -0.39 for the conditions with 43, 51, or 59-dB/band or an LTASS spectrum, respectively, indicate that perceptual weight is reduced for frequencies with greater hearing loss, irrespective of the specific frequencies involved. This result is not surprising, but runs counter to the conclusions drawn by Doherty and Lutfi (1996).

III. EXPERIMENT 2. EFFECT OF REDUCED BANDWIDTH

The SNHL listeners placed less weight on the highest frequency bands when those bands were greatly reduced in audibility. To determine how perceptual weight would be redistributed in NH listeners if the highest bands were inaudible, we tested a subset of six NH listeners using stimuli where the two highest bands were not included.

A. Methods

The six listeners included in the second experiment were L1, L3, L6, L7, L10, and L16 from the first experiment, whose results are shown in Fig. 4. The procedures were identical to those in the original study except that only the 51-dB/band and LTASS conditions were included and those stimuli had 13 bands with center frequencies ranging from 119 Hz to 4450 Hz. The new data were compared to data collected six months earlier in experiment 1.

B. Results

Mean perceptual weights for the six listeners in the 13- and 15-band conditions are shown in Fig. 7. Data for the 15-band condition differ somewhat from those in Fig. 3 because the means reflect results for only 6 of the original 16 listeners. The weights assigned to the highest two bands in the 15-band stimuli were clearly shifted to the highest remaining bands when bands 14 and 15 were removed. The dotted and dashed lines show results for 13- and 15-band stimuli obtained with the Moore *et al.* (1997) model. The dashed lines, which are reproduced from Fig. 3, show increased weight on the highest band in the 51-dB/band condition, but not the LTASS condition. When the two highest bands were not included in the model, as shown by the dotted lines, the weight assigned to band 13 increased in both conditions, but not as much as the effect observed in the listeners' data.

C. Discussion

As noted in the Introduction, several studies have found that the highest frequencies contribute more to total loudness than loudness models would predict. As described in the discussion of experiment 1, the Moore *et al.* (1997) model also assigns greater weight to the highest frequency band than to the bands just below it for the conditions with equal levels per band and further tests with the model suggest that this is due to spread of excitation. The effect of spread of excitation can also account for the increased weight associated with band 13 as the edge band, compared with band 15. Changes in the level of the edge band result in changes in excitation at frequencies above the edge band that contribute to greater changes in loudness. This occurs in both the 13- and 15-band conditions, but the available range of audible frequencies is greater in the 13-band case. The effect can be seen in both the model predictions and the data for the 51-dB/band condition and for model predictions in the LTASS condition.

FIG. 6. Normalized perceptual weights for 16 individual listeners with hearing loss in the 51 dB/band and LTASS conditions. Each row corresponds to a panel in Fig. 1. Within each row, the leftmost panel contains data for the listener represented by squares in Fig. 1, followed by circles, upward-pointing triangles and downward pointing triangles. Again, there are large individual differences in the pattern of weights that are consistent across the two conditions. The weight assigned to higher frequencies declines across panels from upper left to lower right as hearing loss in the high frequencies increases.

IV. EXPERIMENT 3. LOUDNESS MATCHES WITH ONE BAND MISSING

Joshi *et al.* (2016) altered the level of a single tone in a ten-tone complex and assessed the effect on overall loudness in a matching paradigm. They found that changes in the level of components that received more weight also produced a larger change in overall loudness level. To verify that noise bands differed in their contributions to overall loudness, seven NH listeners from experiment 1 were asked to adjust the level of a 15-band stimulus to match the loudness of a 14-band stimulus where either band 1, 8, or 15 was missing. Levels of the other bands were set to the midpoints of the ranges used in experiment 1.

A. Methods

The seven listeners included in the third experiment were L1, L5, L6, L7, L9, L10, and L16 whose results are shown in Fig. 4. Stimuli with 14 and 15 bands were presented in random order in a two-interval forced choice (2IFC) adaptive task, and listeners were asked to indicate which interval they perceived as louder. Durations were the same as those used in experiment 1. The 14-band stimuli remained constant and served as standards. The 15-band stimuli were adapted based on decisions made by the listener regarding whether they were perceived as softer or louder than the 14-band stimuli. The two-track adaptive paradigm was based on Jesteadt (1980). For one track, the 15-band

FIG. 7. (Color online) Normalized perceptual weights for six listeners with NH in the 51-dB/band and LTASS spectral conditions. Dark grey circles and black squares show data for stimuli with 15 bands from experiment 1, for these 6 listeners. Light grey circles and squares show data for stimuli with the two highest bands missing. Dotted and dashed lines show model predictions for the stimuli with 13 and 15 bands, respectively.

stimulus started at a level 10 dB above that of the 14-band stimulus and it was reduced in level after two judgments that it was louder. It was increased in level after one judgment that the 14-band stimulus was louder. The step size was reduced from 4 dB to 2 dB after four reversals in the track. For the other track, the 15-band stimulus began at a level 10 dB lower than for the 14-band stimulus and its level was increased after two judgments that the 14-band stimulus was louder and decreased after one judgment that the 15-band stimulus was louder. The procedure determined the level necessary for the 15-band stimuli in one interval to be perceived as louder than the comparison 14-band stimuli in the other interval 71% and 29% of the time. These two values were then averaged to estimate the 50% point. Twelve conditions, composed of stimuli missing band 1, 8, or 15 at each level (43, 51, or 59-dB/band or LTASS), were randomly repeated four times in 100-trial blocks with one condition per block.

To verify results obtained with the adaptive matching task, the same listeners were tested in a constant-stimulus procedure in which stimuli with 14 and 15 bands were equated in level and presented in random order in a 2IFC

procedure in which listeners were asked to vote for the louder stimulus. Both the 14- and 15-band stimuli were set to the same overall level as the 15-band stimuli in experiment 1 (see Table I). The four spectral conditions were tested separately, but trials where the 14-band stimulus was missing band 1, 8, or 15 were interleaved at random in the same block of trials. Listeners completed three 100-trial blocks for each spectral condition, resulting in 100 trials for each missing band in each condition.

B. Results

Figure 8 shows means (and standard errors) for the seven listeners for normalized perceptual weights for the three bands, mean loudness matches when each of the three bands was missing, and mean values of d' for the constant stimulus procedure. Greater weight was assigned to the lowest and highest bands than the middle band for these seven listeners, but the pattern is different in the other two panels, where removal of only the highest band shows a large effect. Note that the highest band for the LTASS condition shows a large effect for all three measures despite its greatly reduced physical level.

RM-ANOVAs were carried out for the data in each of the three panels. This was followed by correlational analyses to assess the degree of agreement across panels. For the subset of perceptual-weights data from experiment 1 shown in the left panel, there was a significant effect of noise band [$F(2,17.9)=3.83$; $p=0.041$] and a marginally significant interaction of noise band and spectral condition [$F(6,15.1)=2.73$; $p=0.053$]. *Post hoc* tests indicated significantly greater weight assigned to band 15 than to bands 1 or 8, but only for the LTASS condition. For the matching data in the middle panel, there was a marginally significant effect of the missing band [$F(2,7.65)=4.26$; $p=0.057$], but no effect related to spectral condition. *Post hoc* tests indicated a significantly greater reduction in loudness when band 15 was missing than when band 1 was missing. For the constant-stimulus data in the right panel, there was a significant effect of the missing band [$F(2,10.3)=9.70$; $p=0.0043$] and a

FIG. 8. (Color online) Means and standard errors for three measures of the impact of individual bands on overall loudness. The left panel shows mean perceptual weights from experiment 1 for these seven listeners for bands 1, 8, and 15. The middle panel shows the reduction in level (dB) for 15-band stimuli required to match the loudness of a 14-band stimulus, missing band 1, 8, or 15. This provides a direct measure of the contribution of each band to total loudness. The right panel shows a d' conversion of the percentage of times that the 15-band stimulus was judged to be louder than a 14-band stimulus presented at the same overall level.

TABLE II. Pearson r values showing the degree of correlation between measures of the contribution of individual bands to overall loudness for seven subjects. Values significantly different than zero in a one-tailed test ($p = 0.05$) are shown in bold.

Noise band	Condition	Weight vs level decrease	Weight vs d'	Level decrease vs d'
1	43	0.674	0.687	0.813
	51	0.858	0.715	0.717
	59	0.964	0.850	0.822
	LTASS	0.489	0.632	-0.063
8	43	0.336	-0.305	0.577
	51	-0.851	-0.106	-0.105
	59	0.624	0.484	0.048
	LTASS	-0.015	-0.801	0.418
15	43	-0.466	-0.075	0.165
	51	0.017	0.822	0.073
	59	-0.171	0.738	0.077
	LTASS	-0.652	0.622	-0.312

significant effect of spectral condition [$F(3,7.24) = 4.56$; $p = 0.043$], but no significant interaction. *Post hoc* tests indicated a significantly higher value of d' when band 15 was missing than when bands 1 or 8 were missing, as well as a significantly higher value of d' for the LTASS spectral condition than for 43-dB/band.

Agreement among the different measures of the contribution of individual bands can be assessed by correlational analyses to compare results for the seven listeners on one measure to the results on the other two measures, but the power is limited by the low number of listeners. Correlations, shown in Table II, were nonetheless significant for all but the LTASS stimulus when comparing the three different measures of the contribution of band 1. There was no systematic pattern of correlation across the measures for the two higher bands. A correlation measure of the agreement across panels in Fig. 8 shows high agreement between the matching and d' measures ($r = 0.84$), but less agreement of either measure with perceptual weight ($r = 0.42$ for weight vs matching; $r = 0.70$ for weight vs d').

C. Discussion

In the conditions with fixed levels per band, removal of 1 of the 15 bands would result in a 0.3 dB decrease in overall level. Removal of band 15 in the LTASS stimulus reduced its overall level by 0.02 dB. The 14- and 15-band stimuli were presented at equal levels in the constant-stimulus task, correcting for the difference in level, but even if the difference in overall level had been present, it would have been barely detectable. Discrimination in the fixed-level task may have been based on differences in pitch or timbre. Trials with different bands missing were included in the same block to create variability in the perception of the stimuli in an effort to reduce reliance on factors other than loudness, but even with variability from trial to trial, performance was surprisingly good on trials where band 15 was missing in one of the stimuli.

Both low- and high-frequency edge effects are apparent in the perceptual weights data in the left panel of Fig. 8, but

only the high-frequency edge effect is present in the other two panels. This suggests that different factors may contribute to the additional weight associated with the lowest and highest bands, only some of which are present in the other tasks. The data in Fig. 7 present the greatest problem in relating the results across panels in Fig. 8. The data in Fig. 7 indicate that removal of bands 14 and 15 results in increased weight on band 13, with a greater contribution of band 13 to overall loudness. If that is the case, removal of one edge band in the matching or fixed-level task would not be expected to have a significant effect on overall loudness because there would be a new edge band taking its place.

V. GENERAL DISCUSSION

The noise used in three of the four conditions in these studies, with equal levels for each 2-ERB_N band, is similar to pink noise, which has equal power per octave. If white noise, broadband noise with equal spectrum level at all frequencies, had been used in these conditions, the increased width of the higher frequency bands would have caused them to have more power and the increased weight associated with high-frequency bands could have been attributed to the use of higher levels per critical band. The pattern of weights in the three conditions varying in level is different than that observed for six tones with octave spacing (Jesteadt *et al.*, 2014). Both cases show an increase in the weight associated with the highest frequency as overall level increases. For tonal stimuli, the weight associated with the lowest frequency is reduced as the level increases, but for the noise stimuli, the weight increases. This could be due to differences in the perception of the two types of stimuli or to the fact that the relative levels of the tonal stimuli differed from the relative levels of the noise stimuli in the current study. A similar shift in emphasis from low to high frequencies with increasing stimulus level was observed for the tonal stimuli, however, whether the six tones were presented at equal SPL or equal sensation level (Jesteadt *et al.*, 2014).

The weights for listeners with NH are not in good agreement with predictions of the Moore *et al.* (1997) model in that the data show greater edge effects and less contribution by frequencies near the middle of the range. The edge effects observed in the data may be related to the sharp cutoffs in the noise stimuli. The noises were generated by the addition of sinusoids, so the cutoffs were sharper than they would have been if the stimuli had been generated by filtering broadband noise, and this may have resulted in pitch effects. The transitions from band to band were also sharp and differences in the level of adjacent bands could be as much as 12 dB. Results for the LTASS condition, however, argue against sharp cutoffs as an important factor. The LTASS stimuli had a gradual reduction in level across the higher frequency bands, but the highest band still received additional weight. Other factors may contribute to the salience of the edge components (e.g., Moore and Ohgushi, 1993; Moore *et al.*, 2006) causing them to receive more weight in listeners' decisions. An effort was made to test this hypothesis in an earlier study by estimating thresholds for intensity discrimination of each component and there was no relation to

perceptual weight (Leibold and Jesteadt, 2007). It could still be the case that the listeners' decisions are based on small differences between the stimuli unrelated to loudness. All of the measures used here, including the loudness matching task, are based on two-interval paradigms and require listeners to choose the louder of two sounds that are similar in overall loudness. In an effort to make consistent decisions from trial to trial, listeners could be relying on cues other than overall loudness.

The perceptual weights sample discrimination paradigm is robust and not sensitive to the range or distribution of the changes in band intensity around the mean level. Jesteadt *et al.* (2003) found no difference between results obtained with distributions that were Gaussian in dB or Gaussian in power. The current data were reanalyzed to determine whether weights obtained using only level decrements of a given band would differ from those obtained using only level increments. Subsets of trials where band 1, 8, or 15 had a decrement were analyzed separately from trials where those bands had increments, and weights were obtained for all bands. Bands 13 and 14 received more weight when band 15 had a decrement rather than an increment, but there were no other differences. The result for band 15 can be explained in terms of spread of excitation because the levels of bands 13 and 14 would have more impact on total loudness on trials where band 15 was low.

The models of loudness proposed by Moore and colleagues (reviewed by Moore, 2014) are the basis of the current ANSI (2007) standard and have many elements in common with models proposed by Zwicker and colleagues (e.g., Zwicker, 1958; Zwicker and Feldtkeller, 1967), which are the basis of the corresponding German loudness standard (DIN 45631, 1991). These models differ in their assumptions regarding the width of the spectral units underlying loudness summation (ERB_N vs critical bands) and other details, and

they make different predictions regarding the loudness of broadband sounds (e.g., Schlittenlacher *et al.*, 2015; Schlittenlacher *et al.*, 2017). To our knowledge, estimates of perceptual weight have not been obtained for models related to DIN 45631 and it is possible that the two categories of models would predict different edge effects.

The interpretation of perceptual weights as a measure of the contribution of specific frequency bands to the loudness of broadband sounds is based on a number of assumptions. The most important one, discussed above, is that the listeners' decisions are truly based on loudness. The loudness matching experiments were intended to address that issue, but those decisions as well might have been based on some other aspect of the sounds. Another assumption is that an analysis of the *differences* in the levels of bands, from one presentation to the next, can provide information about the contribution of the band rather than the contribution of the change in level. A third assumption is that we can safely ignore differences in the spread of excitation across bands as band levels vary. Some insight into the latter two assumptions can be provided by the model itself.

The left-hand panels of Fig. 9 show comparisons of specific loudness (N') predicted by the model and estimates of perceptual weight obtained with the model for two spectral conditions, 51-dB/band and LTASS. In the 51-dB/band condition, N' increases monotonically from the band centered on 119 Hz to the band centered on 3543 Hz. This reflects a combination of upward spread of excitation and the contribution of ear-canal resonance. N' then decreases before increasing for the final band. The increase there reflects spread of excitation to frequencies above the highest band in the physical stimulus. The total loudness predicted by the model includes the contribution of excitation at frequencies above the highest band. An additional four ERB_N (from 7795 to 12120 Hz) were therefore included in the N' estimate for the highest

FIG. 9. Comparison of three measures of the contribution of individual bands of noise to total loudness obtained using the Moore *et al.* (1997) model. The left-hand panels show comparisons of perceptual weight predicted by the model and estimates of specific loudness obtained with the model for two spectral conditions, 51-dB/band and LTASS. The right-hand panels show comparisons of estimates of specific loudness and estimates from the model of the loudness level of each band in isolation. The right-hand axes have been scaled in both cases to minimize the squared deviation of the two sets of symbols.

band to reflect the contribution of that frequency region. The weights predicted by the model show a similar pattern, with less increase from band to band in the low frequencies because the variability in band levels from interval to interval reduces the impact of the spread of excitation. The weights show an increase for the final band due to the spread of excitation in the absence of noise in the frequency region above that band. The agreement between N' and perceptual weight predicted by the model is greater for the LTASS stimulus shown in the lower panel. Spread of excitation to frequencies above the noise was again included in the estimate of N' , but did not result in an increase of N' or perceptual weight for the highest frequency band due to the lower level of that band. Within the confines of the model, perceptual weight agrees well with specific loudness, suggesting that perceptual weight should provide a measure of the contribution of the band, not just the difference in band level. Similar results were obtained in earlier work (Jesteadt *et al.*, 2014; Joshi *et al.*, 2016). Spread of excitation plays a role as well, however, and could be contributing to the observed differences between perceptual weight and N' .

The right-hand panels of Fig. 9 provide insight into the impact of spread of excitation and masking effects across bands. Here, specific loudness is compared to estimates from the model of the loudness level of each band in isolation. Deviations between the two measures are small even in the LTASS condition, suggesting that the relative loudness level of a band in isolation, as predicted by the model, gives a good indication of its contribution to total loudness, as predicted by the model. In this case, the contributions of frequencies above the upper edge of the highest band have not been included in the estimate of N' because the goal was a band-to-band comparison. Note that the patterns of perceptual weight and other measures generated by the model show large differences across these two conditions, while the perceptual weights for individual listeners shown in Fig. 4 looked similar for the two conditions in many cases.

SNHL listeners tended to put less weight on high-frequency bands than NH listeners and showed a greater difference in the mean data between the LTASS condition and those with equal levels per band. The mean data for SNHL listeners still showed a high-frequency edge effect in three of the four conditions, and many of the individual listeners showed no difference between the LTASS condition and the condition with 51-dB/band. The weights were highly reliable, but differed markedly from listener to listener. There was a significant negative correlation between degree of hearing loss at a given frequency and perceptual weight associated with that frequency in all four conditions, and the relation between hearing loss and perceptual weight did not differ across the four arbitrary groups shown in Fig. 6. If the highest frequency band receives greater weight because it is at the edge of the stimulus, the highest audible band might be expected to be given greater weight by SNHL listeners. Doherty and Lutfi (1996) observed such a pattern in their study using six-tone complexes where SNHL listeners placed maximum weight on the tone at 4000 Hz and less weight on 8000 Hz. They attributed it to an effort by the listeners to compensate for their hearing loss because they

observed no such effect in listeners with NH. Because the two groups of listeners were not tested at the same level, however, it is difficult to separate effects of level from effects of hearing loss. The use of six tones with octave spacing by Doherty and Lutfi (1996) and Jesteadt *et al.* (2014) makes it difficult to relate differences in spectral weight to features of the audiogram because neither measure has sufficient grain. In the present study, the inter-listener variability apparent in weights for both NH and SNHL listeners may have reduced the correlation between audiograms and perceptual weights.

An alternative interpretation of the perceptual weights data is that listeners are basing their decisions on pitch, timbre, or some other perceptual property of the stimuli rather than loudness and that alternative property is subject to an edge effect. The third experiment was designed to address that issue by asking listeners to make loudness matches using stimuli where band levels were not varying from interval to interval. Although the data show a significant difference in loudness associated with the highest band, results from the second experiment suggest that the 14th band should make a bigger contribution to loudness once the 15th band is removed, largely eliminating any difference in loudness between 14- and 15-band noises. It might be possible to account for the results of both studies by assuming that the 15th band contributes to a pitch difference between the stimuli and that there is an interaction between pitch and loudness (Jesteadt and Neff, 1982; Stevens, 1935).

One of our goals in exploring the use of perceptual weights to determine the contribution of specific frequency bands to the loudness of broadband sounds was the development of better procedures to be used in fitting hearing aids. Current efforts to restore normal loudness when fitting hearing aids rely on tests using narrowband sounds (e.g., Cox, 1995) or models based primarily on data obtained with narrowband sounds (e.g., Moore, 2000). Oetting *et al.* (2016) recently reported results for loudness judgments by NH and SNHL listeners obtained using narrow- and broadband sounds in monaural and binaural conditions. Use of results obtained with narrowband sounds to restore normal loudness in SNHL listeners resulted in louder than normal broadband sounds, and this effect was greatest in binaural conditions. They note that this abnormal binaural loudness summation might be one reason that new hearing-aid users frequently complain that sounds are too loud (Jenstad *et al.*, 2003). The results of the current study suggest that spectral weights can be determined with high reliability and could be used to identify frequency regions that were making the strongest contributions to loudness complaints, but that would require a greater understanding of edge effects and individual differences.

VI. CONCLUSIONS

Use of a perceptual-weights paradigm to determine the contribution of specific frequency bands to the loudness of broadband sounds suggests that the lowest and especially the highest frequency bands carry more weight than would be predicted from equal-loudness contours or the audibility of

individual bands. This is true for SNHL as well as NH listeners, with the caveat that noise bands that are inaudible carry no weight.

An experiment using a matching paradigm with NH listeners suggests that the data could deviate from estimates of loudness level based on loudness models by 3 dB or more. An experiment using a perceptual-weights paradigm with stimuli reduced in bandwidth, however, suggests that the weight assigned to the highest bands, the source of the 3-dB difference, transfers to the highest remaining bands when a band is removed. This should greatly reduce any overall difference in loudness in the matching paradigm, but does not have that effect. One way to reconcile these results is to assume that listeners' judgments are influenced by properties other than loudness.

The source of the edge effects observed in the perceptual weights paradigm is unknown. It is noteworthy that the Moore *et al.* (1997) loudness model generates edge effects when used to predict perceptual weights, but they are generally smaller than those observed in the data.

Perceptual weights are subject to large individual differences, but are clearly reproducible for individual listeners.

ACKNOWLEDGMENTS

The authors would like to thank Tom Creutz and Hongyang Tan for their help with the data collection and data analysis software; Brian Moore, Brian Glasberg, and Josef Schlittenlacher for sharing the code for the loudness model; and Katie Thraillkill for help with manuscript preparation and editing. We thank Marc Brennan, Suyash Joshi, and Lori Leibold for comments on earlier versions. This paper benefitted greatly from reviews by Michael Stone, Brian Moore, and Joseph Schlittenlacher. The work described in experiment 3 was submitted by S.M.W. to the University of Nebraska-Lincoln as an Doctor of Audiology (AuD) capstone project. This research was supported by National Institutes of Health (NIH) Grants Nos. R01 DC 011806 and P30 DC 004662.

Ahumada, A. J. (2002). "Classification image weights and internal noise level estimation," *J. Vision* **2**, 121–131.

ANSI (2007). ANSI 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.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Cox, R., Hagerman, B., and Ludvigsen, C. (1994). "An international comparison of long-term average speech spectra," *J. Acoust. Soc. Am.* **96**, 2108–2120.

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

Cox, R. M. (1995). "Using loudness data for hearing aid selection: The IHAF approach," *Hear. J.* **48**, 10–44.

DIN (1991). 45631, "Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum, Verfahren nach E. Zwicker" ("Procedure for calculating loudness levels and loudness method according to E. Zwicker") (Deutsches Institut für Normung, Berlin).

Dittrich, K., and Oberfeld, D. (2009). "A comparison of the temporal weighting of annoyance and loudness," *J. Acoust. Soc. Am.* **126**, 3168–3178.

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

Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.

Holte, L., and Margolis, R. H. (1987). "The relative loudness of third-octave bands of speech," *J. Acoust. Soc. Am.* **81**, 186–190.

ISO 226 (2003). *Acoustics—Normal Equal-Loudness Contours* (International Organization for Standardization, Geneva, Switzerland).

ISO 532-2 (2017). *Methods for Calculating Loudness—Part 2: Moore-Glasberg Method* (International Organization for Standardization, Geneva, Switzerland).

Jenstad, L. M., Van Tasell, D. J., and Ewert, C. (2003). "Hearing aid troubleshooting based on patients' descriptions," *J. Am. Acad. Audiol.* **14**, 347–360.

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

Jesteadt, W., and Neff, D. L. (1982). "A signal-detection-theory measure of pitch shifts in sinusoids as a function of intensity," *J. Acoust. Soc. Am.* **72**, 1812–1820.

Jesteadt, W., Nizami, L., and Schairer, K. S. (2003). "A measure of internal noise based on sample discrimination," *J. Acoust. Soc. Am.* **114**, 2147–2157.

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.

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.

Keidser, G., Katsch, R., Dillon, H., and Grant, F. (2002). "Relative loudness of low- and high-frequency bands of speech-shaped babble, including the influence of bandwidth and input level," *J. Acoust. Soc. Am.* **111**, 669–671.

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.

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

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

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

Lentz, J. J. (2006). "Spectral-peak selection in spectral-shape discrimination by normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **120**, 945–956.

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.

Meunier, S., Marchioni, A., and Rabau, G. (2000). "Subjective evaluation of loudness models using synthesized and environmental sounds," in *Proceedings of INTERNOISE 2000*, Nice, France.

Moore, B. C. J. (2000). "Use of a loudness model for hearing aid fitting. IV. Fitting hearing aids with multi-channel compression so as to restore 'normal' loudness for speech at different levels," *Br. J. Audiol.* **34**, 165–177.

Moore, B. C. J. (2014). "Development and current status of the 'Cambridge' loudness models," *Trends Hear.* **18**, 1–29.

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

Moore, B. C. J., and Glasberg, B. R. (2007). "Modeling binaural loudness," *J. Acoust. Soc. Am.* **121**, 1604–1612.

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

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

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

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.
- Murray, R. F. (2011). "Classification images: A review," *J. Vision* **11**, 1–25.
- Oberfeld, D., Heeren, W., Rennie, J., and Verhey, J. (2012). "Spectro-temporal weighting of loudness," *PLoS One* **7**, e50184.
- Oberfeld, D., Kuta, M., and Jesteadt, W. (2013). "Factors limiting performance in a multitone intensity-discrimination task: Disentangling non-optimal decision weights and increased internal noise," *PLoS One* **8**, e79830.
- Oberfeld, D., and Plank, T. (2011). "The temporal weighting of loudness: Effects of the level profile," *Atten. Percept. Psychophys.* **73**, 189–208.
- Oetting, D., Hohmann, V., Appell, J. E., Kollmeier, B., and Ewert, S. D. (2016). "Spectral and binaural loudness summation for hearing-impaired listeners," *Hear. Res.* **335**, 179–192.
- Pollack, I. (1952). "The loudness of bands of noise," *J. Acoust. Soc. Am.* **24**, 533–538.
- Rasetshwane, D. M., Trevino, A. C., Gombert, J. N., Liebig-Trehearn, L., Kopun, J. G., Jesteadt, W., Neely, S. T., and Gorga, M. P. (2015). "Categorical loudness scaling and equal-loudness contours in listeners with normal hearing and hearing loss," *J. Acoust. Soc. Am.* **137**, 1899–1913.
- Rennie, J., Wächter, M., Hots, J., and Verhey, J. (2015). "Spectro-temporal characteristics affecting the loudness of technical sounds: Data and model predictions," *Acta Acust. Acust.* **101**, 1145–1156.
- 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.
- Schlittenlacher, J., Ellermeier, W., and Avci, G. (2017). "Simple reaction time for broadband sounds compared to pure tones," *Atten. Percept. Psychophys.* **79**, 628–636.
- Schlittenlacher, J., Ellermeier, W., and Hashimoto, T. (2015). "Spectral loudness summation: Shortcomings of current standards," *J. Acoust. Soc. Am.* **137**, EL26–EL31.
- Schlittenlacher, J., Hashimoto, T., Fastl, H., Namba, S., Kuwano, S., and Hatano, S. (2011). "Loudness of pink noise and stationary technical sounds," in *Proceedings of INTERNOISE 2011*, Osaka, Japan.
- Stevens, S. S. (1935). "The relation of pitch to intensity," *J. Acoust. Soc. Am.* **6**, 150–154.
- Stevens, S. S. (1956). "Calculating the loudness of complex noise," *J. Acoust. Soc. Am.* **28**, 807–832.
- Stevens, S. S. (1972). "Perceived level of noise by Mark VII and Decibels (E)," *J. Acoust. Soc. Am.* **51**, 575–601.
- Stevens, S. S., Egan, J. P., and Miller, G. A. (1947). "Methods of measuring speech spectra," *J. Acoust. Soc. Am.* **19**, 771–780.
- Westfall, P., Tobias, R. D., and Wolfinger, R. D. (2011). *Multiple Comparisons and Multiple Tests Using SAS*, 2nd ed. (SAS Institute, Cary, NC).
- Zwicker, E. (1958). "Über psychologische und methodische Grundlagen der Lautheit" ("Psychological and methodical basis of loudness"), *Acustica* **8**, 237–258.
- Zwicker, E., and Feldtkeller, R. (1967). *Das Ohr als Nachrichtenempfänger (The Ear as a Communication Receiver)* (Hirzel, Stuttgart).
- Zwicker, E., Flottorp, G., and Stevens, S. S. (1957). "Critical bandwidth in loudness summation," *J. Acoust. Soc. Am.* **29**, 548–557.