

A Model of Loudness Applicable to Time-Varying Sounds*

BRIAN R. GLASBERG AND BRIAN C. J. MOORE, *AES Member*

Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK

Previously we described a model for calculating the loudness of steady sounds from their spectrum. Here a new version of the model is presented, which uses a waveform as its input. The stages of the model are as follows. (a) A finite impulse response filter representing transfer through the outer and middle ear. (b) Calculation of the short-term spectrum using the fast Fourier transform (FFT). To give adequate spectral resolution at low frequencies, combined with adequate temporal resolution at high frequencies, six FFTs are calculated in parallel, using longer signal segments for low frequencies and shorter segments for higher frequencies. (c) Calculation of an excitation pattern from the physical spectrum. (d) Transformation of the excitation pattern to a specific loudness pattern. (e) Determination of the area under the specific loudness pattern. This gives a value for the “instantaneous” loudness. The short-term perceived loudness is calculated from the instantaneous loudness using an averaging mechanism similar to an automatic gain control system, with attack and release times. Finally the overall loudness impression is calculated from the short-term loudness using a similar averaging mechanism, but with longer attack and release times. The new model gives very similar predictions to our earlier model for steady sounds. In addition, it can predict the loudness of brief sounds as a function of duration and the overall loudness of sounds that are amplitude modulated at various rates.

0 INTRODUCTION

In an earlier paper [1] we described a model for the calculation of the loudness of steady sounds from the spectra of those sounds. Many of the concepts used in developing the model were based on the work of earlier researchers [2]–[6]. However, for certain types of sounds our model gave more accurate predictions than earlier models.

Two aspects of our model limited its application. First the model required as input a specification of the spectrum of the sound. For most applications, such as estimating the loudness of environmental sounds arising from heating, ventilation, and air-conditioning (HVAC), this meant that spectra had to be calculated (usually in one-third-octave bands) from the time waveforms of the sounds. While one-third-octave spectra give adequate resolution for sounds with smooth broad-band spectra, they do not always give adequate resolution for sounds with discrete spectral components. A second limitation was that the model was applicable only to steady sounds. Many everyday sounds, such as speech and music, are time-varying, and it would be useful to have a way of calculating the loudness of such sounds. For sounds like speech, there are two aspects of

the loudness impression: the listener can judge the short-term loudness, for example, the loudness of a specific syllable; or the listener can judge the overall loudness of a relatively long segment, such as a sentence. We will refer to the latter as the long-term loudness. The long-term loudness probably reflects relatively “high-level” cortical processes and involves memory. The long-term loudness impression can persist for several seconds after a sound has ceased.

The model presented in this paper is intended to overcome these two limitations: it uses as input the time waveform of a sound, as might be picked up by a microphone in a specific location; and it produces running estimates of the short-term and long-term loudness. A loudness model capable of operating on the time waveform of a sound was described some years ago by Zwicker and coworkers [7]–[9]. Although the model was implemented using analog electronics, this was both costly and complex, and the model did not find widespread application. A somewhat simpler model, implemented using digital signal processing, was described by Stone et al. [10], but this model includes only a very simple form of temporal integration and does not account adequately for the loudness of amplitude-modulated sounds, a topic that will be discussed in more detail. Loudness is also calculated from the

*Manuscript received 2001 July 11; revised 2002 March 15.

time waveforms of sounds in some commercially available devices. However, it appears that these devices base their calculations on the method described in ISO 532 [11], which is intended for use with steady sounds. Hence the outputs of these devices may not correspond accurately to the loudness impression evoked by time-varying sounds.

One aspect of loudness that we wished to account for with our new model was the long-term loudness of amplitude-modulated sounds. When listening to a sound that is amplitude modulated at a moderate rate, say 10 Hz, it is easy to hear that the sound is fluctuating, but at the same time one gets an overall impression of loudness—the long-term loudness. However there is some controversy as to the physical measure that correlates best with the long-term loudness. Data on the long-term loudness of amplitude-modulated sinusoidal carriers were presented by Bauch [12]; see also Zwicker and Fastl [6]. Subjects were required to adjust the level of a steady tone to match the loudness of a modulated tone. For very low modulation rates, below about 10 Hz, the loudness of the modulated tone corresponded to its peak level. In other words, the steady tone was judged equal in loudness to the modulated tone when its level was equal to the peak level of the modulated tone. For higher rates, up to about one-half of the critical bandwidth, the loudness decreased and corresponded to the rms level. For rates larger than about one-half of the bandwidth of the auditory filter [13], the loudness increased, as the spectral components were resolved by the auditory system. Results were similar for carrier frequencies from 400 to 4000 Hz and for levels ranging from 30 to 70 dB SPL.

Hellman [14] used a magnitude estimation method to estimate the loudness of two-tone complexes as a function of the frequency separation of the tones. Her conclusions were similar to those of Bauch; for small frequency separations (corresponding to low beat rates) the loudness corresponded to the peak level of the two tones, whereas for larger separations it corresponded to the rms level.

Fastl [15] obtained loudness matches between a sinusoid and a narrow-band noise, both centered at 8.5 kHz. A noise band has inherent amplitude fluctuations, which increase in rate with increasing bandwidth. He found that, for bandwidths from 10 Hz up to about 300 Hz, the sinusoid and the noise were judged as equally loud when their rms values were roughly equal. For greater bandwidths, the sinusoid had to have a higher level than the noise for the loudness to match. This effect amounted to about 4 dB for a bandwidth of 700 Hz. Fastl [15] did not obtain loudness matches for noise bandwidths less than 10 Hz, so it is not known whether the loudness would have increased for a noise with very slow fluctuations.

Moore et al. [16] presented data on the loudness of amplitude-modulated tones. Loudness matches were made between a steady 4000-Hz sinusoid and a 4000-Hz sinusoidal carrier that was amplitude modulated at various rates and depths. The modulated and steady sounds were presented in regular alternation, and matches were made with both the steady tone varied and the modulated tone varied. The mean results differed from those of Bauch

[12]. For modulation rates above 4 Hz, at the point of equal loudness the modulated tone had a slightly higher rms level than the steady tone and a markedly higher peak level. For modulation rates of 4 Hz and below, the modulated and steady tones had almost equal rms levels at the point of equal loudness. There was no significant effect of overall level or of modulation depth. Similar results were found by Moore et al. [17] for both sinusoidal carriers and carriers consisting of white noise and noise with the same long-term average spectrum as speech.

Zhang and Zeng [18] attempted to replicate some of the results of Bauch [12], but using an adaptive procedure with two interleaved staircases to determine the point of equal loudness [19]. They used both a two-tone complex (that is, a pair of beating tones) and a three-tone complex with components added in two different starting phases, producing either amplitude modulation or quasi-frequency modulation. All sounds were centered at 1000 Hz. For intermediate modulation rates, at the point of equal loudness the steady sound had a slightly lower rms level than the modulated sound, as was also reported by Moore et al. [16], [17]. However, for very low modulation rates (below 10 Hz) the steady sound had a slightly higher rms level than the modulated sound at the point of equal loudness. The results for very low rates fell between those reported by Bauch [12] and those reported by Moore et al. [16], [17].

Grimm et al. [20] obtained loudness matches between an unmodulated broad-band (3800-Hz-wide) noise centered at 2000 Hz and a noise that was multiplied by a sinusoid. For frequencies of the multiplying sinusoid between 4 and 32 Hz, the steady noise had to be about 1 dB higher in level than the multiplied noise to obtain equal loudness. This effect is in the same direction as reported by Zhang and Zeng [18]. In another experiment the frequency of the multiplying sinusoid was fixed at 8 Hz while the noise bandwidth was varied. For the narrow-band carriers (200-Hz-wide noise or sinusoid) the results were as reported by Moore et al. [16], [17]; the modulated sound had a slightly (about 1-dB) higher level than the unmodulated sound at the point of equal loudness. For wide-band carriers the effect reversed.

Based on this overview of the literature, we conclude that for carriers that are amplitude modulated at very low rates, the perceived loudness corresponds to a level between the rms level and the peak level. For sounds that are modulated at intermediate rates, the loudness corresponds to a level close to but slightly below the rms level. For sounds that are modulated at high rates, the spectral sidebands may be resolved (at least for sinusoidal carriers), which usually leads to an increase in loudness; the modulation rate at which this first occurs increases with increasing center frequency [12]. We sought to develop a model that would give predictions of this form.

A second aspect of loudness perception that we wished to account for was the way that loudness changes with duration, for a fixed intensity; this is often described as temporal integration for loudness. Data on this topic show considerable variability across studies; for reviews and data see Scharf [21], Florentine et al. [22], and Buus et al. [23]. The variability reflects the great difficulty that sub-

jects seem to have in comparing the loudness of sounds of different durations. However, it is generally agreed that, for a fixed intensity, loudness increases with increasing duration for durations up to 100–200 ms, and then remains roughly constant. For durations less than 100 ms, the loudness increases by roughly 10 phons for each ten-fold increase in duration, although this rule may fail at very short durations as the spectrum of the sound broadens considerably for very short durations. We sought to develop a model that would predict these basic features of temporal integration of loudness.

1 DETAILS OF MODEL

In what follows we assume a sampling rate of 32 kHz and 16-bit resolution, although the model can readily be adapted for use with other sampling rates and resolutions, such as 48 kHz and 24 bit.

1.1 Transfer through the Outer and Middle Ear

The transfer of sound through the outer and middle ear can be modeled using fixed filters, although the filtering produced by the outer ear depends on the direction of incidence of the sound relative to the head [24]. The transfer function of the outer ear for frontal incidence used in our model is given in Moore et al. [1, fig. 2]. The assumed transfer function of the middle ear is given in Moore et al. [1, fig. 3]. In the version of the model described here, the combined effect of the outer and middle ear is modeled by a single finite impulse response (FIR) filter with 4097 coefficients. We decided to simulate the effects of the outer and middle ear in this way, rather than by modifying magnitude values in the fast Fourier transform (FFT), because the spectral smearing produced

by time-window analysis did not allow sufficient accuracy at low frequencies.

The filter was designed using the FIR2 function in MATLAB. The transfer characteristic of this filter, for a sound with frontal incidence, is shown in Fig. 1. The filter is constructed so as to have a gain of 0 dB at 1000 Hz. Other filters can be used for different directions of sound incidence. The output of the filter can be considered as representing the sound reaching the cochlea (the inner ear).

1.2 Calculation of Running Spectrum and Excitation Pattern

The cochlea can be characterized as containing a bank of bandpass filters whose center frequencies span the range from about 50 to 15 000 Hz [1], [25]. The bandwidths of the filters increase with increasing center frequency. For a filter centered around 100 Hz the equivalent rectangular bandwidth (ERB) is about 35 Hz, whereas at 10 000 Hz it is about 1100 Hz [13]. The filters are level dependent, the low-frequency slopes becoming less steep with increasing level [26]–[29]. The magnitude of the output of each filter in response to a given sound, plotted as a function of filter center frequency, is called the excitation pattern of that sound [25], [30], and the calculation of the excitation pattern is an important stage in most loudness models.

Some researchers have proposed time-domain models of the level-dependent auditory filter bank [31]–[34]. However, these models are computationally intensive, and are unsuitable for applications where results are required quickly, or in real time. Also, some of the models have parameters chosen to fit data from animals, and they do not provide an accurate characterization of human audi-

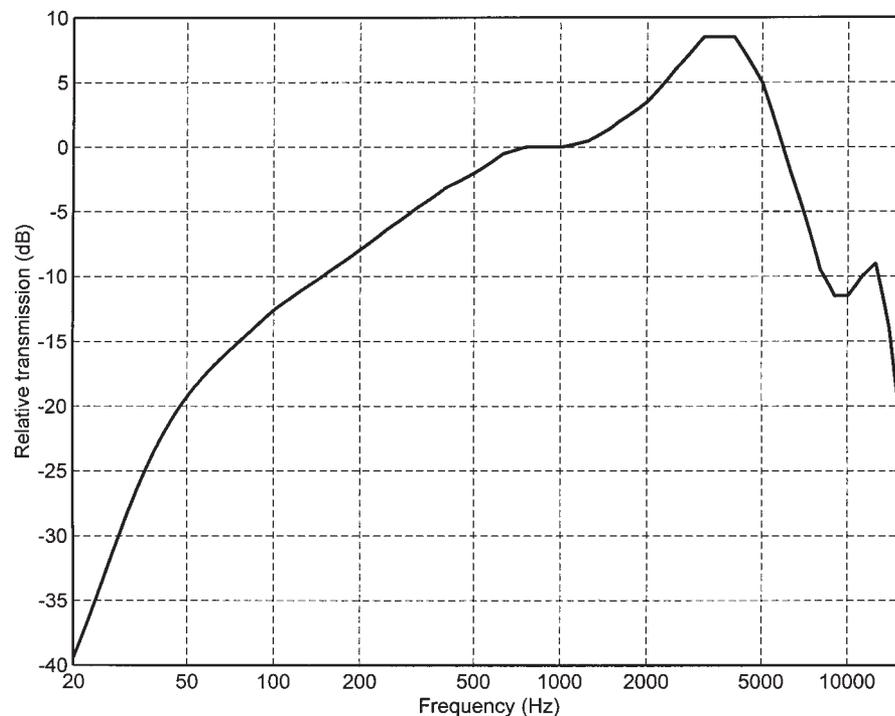


Fig. 1. Transfer function of finite impulse response digital filter used to simulate effects of outer and middle ear. Gain scaled to be 0 dB at 1000 Hz.

tory filters. We have adopted a computationally efficient method of calculating the excitation pattern, based on an initial spectral analysis using the FFT. To obtain spectral resolution at low frequencies comparable to that of the auditory system, the analysis of relatively long (about 64-ms) segments of the input signal is required. However, for high center frequencies, the use of such long segments would have the effect of limiting temporal resolution in a way that does not occur in the auditory system. For example, when a high carrier frequency is used, amplitude modulation at rates up to several hundred hertz can be detected without the resolution of spectral sidebands [35], [36]. To give adequate temporal resolution at high frequencies, analysis using much shorter signal segments (about 2 ms) is required.

To deal with these problems, our model calculates six FFTs in parallel, using signal segment durations that decrease with increasing center frequency. The six FFTs are based on Hanning-windowed segments with durations of 2, 4, 8, 16, 32, and 64 ms, all aligned at their temporal centers. The windowed segments are zero padded, and all FFTs are based on 2048 sample points. All FFTs are updated at 1-ms intervals.

Each FFT is used to calculate spectral magnitudes over a specific frequency range; values outside that range are discarded. These ranges are 20 to 80 Hz, 80 to 500 Hz, 500 to 1250 Hz, 1250 to 2540 Hz, 2540 to 4050 Hz, and 4050 to 15 000 Hz for segment durations of 64, 32, 16, 8, 4, and 2 ms, respectively. An excitation pattern is calculated from the short-term spectrum at 1-ms intervals, using exactly the same method as described in Moore et al. [1]. Briefly, the outputs of the auditory filters are calculated for center frequencies spaced at 0.25-ERB intervals, taking into account the known variation of the auditory filter shape with center frequency and level [13]. The excitation pattern is then defined as the output of the auditory filters as a function of the center frequency.

1.3 Calculation of Instantaneous Loudness

The next stage in the model is the calculation of what we call the “instantaneous” loudness. We assume that the instantaneous loudness is an intervening variable which is not available for conscious perception. It might correspond, for example, to the total activity in the auditory nerve, measured over a very short time interval, such as 1 ms. The perception of loudness depends on the summation or integration of neural activity over times longer than 1 ms. This summation process is modeled subsequently.

The calculation of instantaneous loudness from the excitation pattern is done in the same way as in our earlier model [1]; the reader is referred to that paper for details. Briefly, the excitation pattern is transformed to a specific loudness pattern, and the area under the loudness pattern is summed to give the instantaneous loudness for a given ear. The transformation from excitation to specific loudness involves a compressive nonlinearity, meant to resemble the compression that occurs in the cochlea [37], [38]. If two ears are being used, the instantaneous loudness is summed across ears to give the overall instantaneous loudness.

1.4 Calculation of Short-Term Loudness

The short-term loudness, the loudness perceived at any instant, is calculated using a form of temporal integration or averaging of the instantaneous loudness which resembles the way that a control signal is generated in an automatic gain control (AGC) circuit. Such a control signal has an attack time T_a and a release time T_r . This was implemented in the following way. We define S'_n as the running (averaged) short-term estimate of loudness at the time corresponding to the n th time frame (updated every 1 ms), S_n as the calculated instantaneous loudness at the n th time frame, and S'_{n-1} as the running loudness at the time corresponding to frame $n - 1$.

If $S_n > S'_{n-1}$ (corresponding to an attack, as the instantaneous loudness at frame n is greater than the short-term loudness at the previous frame), then

$$S'_n = a_a S_n + (1 - a_a) S'_{n-1} \quad (1)$$

where a_a is a constant related to T_a ,

$$a_a = 1 - e^{-T_i/T_a}. \quad (2)$$

Here T_i is the time interval between successive values of the instantaneous loudness (1 ms in this case).

If $S_n \leq S'_{n-1}$ (corresponding to a release, as the instantaneous loudness is less than the short-term loudness), then

$$S'_n = a_r S_n + (1 - a_r) S'_{n-1} \quad (3)$$

where a_r is a constant related to T_r ,

$$a_r = 1 - e^{-T_i/T_r}. \quad (4)$$

The values of a_a and a_r were set to 0.045 and 0.02, respectively. The value of a_a was chosen to give reasonable predictions for the variation of loudness with duration; see later for details. The value of a_r was chosen to give reasonable predictions of the overall loudness of amplitude-modulated sounds. The fact that a_a is greater than a_r means that the short-term loudness can increase relatively quickly when a sound is turned on, but it takes somewhat longer to decay when the sound is turned off. The decay may correspond to the persistence of neural activity at some level in the auditory system and it may be related to forward masking [6], [39].

1.5 Calculation of Long-Term Loudness

The long-term loudness is calculated from the short-term loudness, again using a form of temporal integration resembling the operation of an AGC circuit. Denote the long-term loudness at the time corresponding to frame n by S''_n . If $S'_n > S''_{n-1}$ (corresponding to an attack, as the short-term loudness at frame n is greater than the long-term loudness at the previous frame), then

$$S''_n = a_{al} S'_n + (1 - a_{al}) S''_{n-1} \quad (5)$$

where a_{al} is a constant related to the attack time of the averager [as described in Eq. (2)].

If $S_n' \leq S_{n-1}''$ (corresponding to a release, as the short-term loudness is less than the long-term loudness), then

$$S_n'' = a_{ri} S_n' + (1 - a_{ri}) S_{n-1}'' \quad (6)$$

where a_{ri} is a constant related to the release time of the averager.

The values of a_{al} and a_{ri} were set to 0.01 and 0.0005, respectively. The fact that a_{al} is greater than a_{ri} means that the overall loudness impression can increase relatively quickly when a sound is turned on, but it takes a long time to decay when the sound is turned off.

2 PREDICTIONS OF THE MODEL FOR STEADY SOUNDS

Is it important that the new model make predictions of loudness and absolute threshold for steady sounds comparable in accuracy to those of our earlier model [1]. Hence we start by presenting predictions for steady sounds. We then go on to consider the loudness of time-varying sounds. In deriving the predictions for steady sounds, we used signals of relatively long duration, typically greater than 2000 ms, and determined the long-term loudness between 600 and 2000 ms after the start of the signal. For sounds with very slow amplitude fluctuations, such as narrow bands of noise, the long-term loudness was averaged over this time period.

2.1 Absolute Thresholds

As in our earlier model [1], the absolute threshold of a sound corresponds to the level at which the loudness is 0.003 sones or 2 phons. Hence the absolute threshold for any sound can be predicted simply by determining the level of the sound required to give a predicted loudness of 0.003 sones. The lowest solid curve in Fig. 2 shows thresholds predicted by the new model for sinusoids as a

function of frequency. These are thresholds for binaural listening in free field (frontal incidence), otherwise known as the minimum audible field (MAF). For comparison, the dashed line shows the corresponding predictions of our earlier model [1]. There is a very good correspondence between the predictions of the two models (discrepancies are nearly all less than 1 dB), and both sets of predictions correspond well to the empirical values published in ISO 389-7 [40].

2.2 Loudness as a Function of Loudness Level

Fig. 3 shows the relationship between the level of a 1-kHz sinusoid (loudness level in phons) and calculated loudness in sones, assuming binaural presentation in the free field. The solid and dashed curves show predictions of the new model and our earlier model, respectively. The correspondence is very good. The relationship shown in Fig. 3 is used within the model to transform the calculated loudness in sones to the loudness level in phons.

2.3 Equal-Loudness Contours

Fig. 2 shows equal-loudness contours predicted by the new model for binaural listening in the free field (frontal incidence). The lowest curve is the absolute threshold (MAF) curve, corresponding to a loudness level of 2 phons. The contours are very similar to those predicted by our earlier model [1]. Note that the contours are not intended to correspond to the contours published as ISO 226 [41]. The ISO 226 contours are now believed to contain systematic errors [42], [43]. A new standard for equal-loudness contours is currently being developed, and we hope that the contours predicted by our model will correspond reasonably well to those in the new standard.

2.4 Loudness as a Function of Bandwidth

If the bandwidth of a sound is varied keeping the overall intensity fixed, the loudness remains constant as long

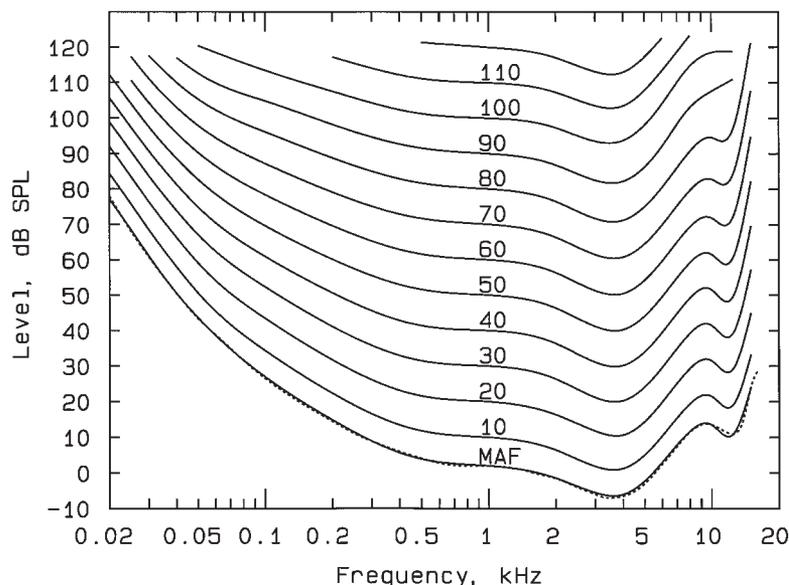


Fig. 2. Predicted equal-loudness contours for binaural listening in free field with frontal incidence. Parameter is loudness level. Lowest curve is MAF for binaural listening. . . . MAF predicted by earlier model [1].

as the bandwidth is less than a certain value (called the CB for loudness). If the bandwidth is increased beyond the CB, the loudness increases, except when the overall sound level is very low [44]–[47]. The model can predict this effect.

Fig. 4 shows the predictions of the model for a noise band centered at 1000 Hz, with an overall level of 10, 20, 40, 60, or 80 dB SPL. It was assumed that the sound was presented binaurally in a free field with frontal incidence. The predicted curves resemble the empirically obtained

results in general form, showing the following features:

- 1) The loudness is invariant with bandwidth for bandwidths less than about 160 Hz.
- 2) For the three higher levels, the loudness increases for bandwidths greater than about 160 Hz, and the rate of increase is slightly greater for medium than for high levels.
- 3) At low levels the loudness increases very little with increasing bandwidth and even decreases at the largest bandwidths.

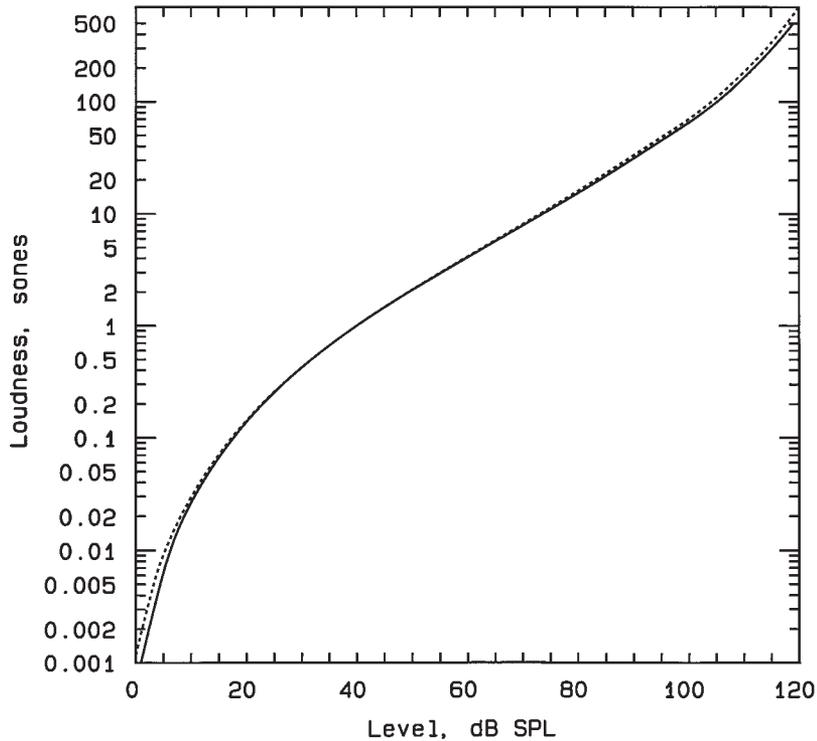


Fig. 3. Predicted loudness in sones, plotted as a function of level of a 1-kHz sinusoid presented binaurally in free field (frontal incidence).

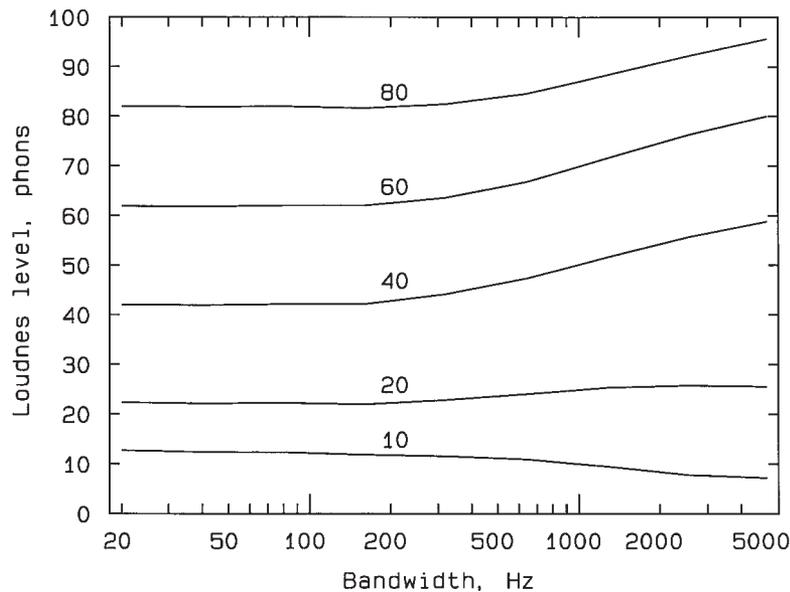


Fig. 4. Loudness level as a function of bandwidth for five different fixed overall noise levels. Noise was geometrically centered at 1000 Hz.

3 LOUDNESS OF TIME-VARYING SOUNDS

The model provides estimates of the instantaneous loudness, the short-term loudness, and the long-term loudness, each updated every 1 ms. Recall that the instantaneous loudness is assumed to be an intervening variable that is not available to conscious perception. We assume that the loudness of a brief sound is determined by the maximum value of the short-term loudness.

3.1 Response to a Tone Burst

An example of the output of the model is given in Fig. 5, which shows the instantaneous loudness, the short-term loudness, and the long-term loudness plotted as a function of time for a 200-ms 4000-Hz tone burst, gated on and off abruptly. Several features can be noted in this figure. First the instantaneous loudness shows local maxima at times corresponding to the onset and offset of the sound. These are caused by spectral spreading related to the abrupt gating; this effect is discussed in more detail later. Related to this, the short-term loudness also shows a slight rise at the end of the signal. Second both the short-term and the long-term loudness take some time to build up, the long-term loudness taking longer. In fact, the long-term loudness does not quite reach its asymptotic value. Third both the short-term and the long-term loudness take some time to decay when the input ceases. We assume that the long-term loudness may correspond to a memory for the loudness of an event (a tone burst in this case), which can last for several seconds, although it can probably be “reset” by a new sound event.

3.2 Temporal Integration of Loudness

As noted in the Introduction, for a sinusoid of fixed peak level, and for durations below about 100 ms, the loudness level increases by roughly 10 phons for each tenfold increase in duration; this is equivalent to a 3-phon increase per doubling of duration. For very short dura-

tions, this relationship can break down because the spectrum of the signal becomes sufficiently broad to cause an increased spread of the excitation pattern. In that case a loudness summation effect occurs, similar to that described in Section 2.4, and the loudness level is slightly greater than would be expected from the 3-phon-per-doubling relationship. Fig. 6 shows the predicted loudness level as a function of duration (time on a logarithmic scale) for four cases. The signal was a gated sinusoid with a level of 60 dB SPL and a frequency of either 1000 or 4000 Hz. The signal either had abrupt onsets and offsets (gated at zero crossings) or had 5-ms raised-cosine ramps at the onset and offset.

For the 4000-Hz signal without ramps, the loudness level increases by roughly 3 phons per doubling of duration over the range of durations from 5 to 80 ms. For longer durations the increase in loudness level becomes more gradual, and for durations beyond 160 ms the loudness level reaches an asymptote. All of these features accord well with empirical data [21]. When the duration is decreased from 4 to 2 ms, the loudness level decreases by much less than 3 phons. This effect can be attributed to the broadening of the signal’s spectrum for very short durations. When 5-ms ramps are added to the signal, the loudness level decreases more progressively with decreasing duration. Again, these effects are consistent with empirical data. It is noteworthy that the signal with ramps is very slightly less loud than the signal without ramps for the longer durations. This indicates that the spectral spread of an abruptly gated signal produces a small increase in its loudness even when the duration is as great as 200 ms.

The pattern of results for the 1000-Hz signal with 5-ms ramps is similar to that for the 4000-Hz signal with ramps, except that the overall loudness is lower. However, at 1000 Hz the spectral spread of the signal plays a role for slightly longer durations. This happens because the ERB of the auditory filter is smaller at 1000 than at 4000 Hz. For the 1000-Hz signal without ramps, the spectral splatter plays

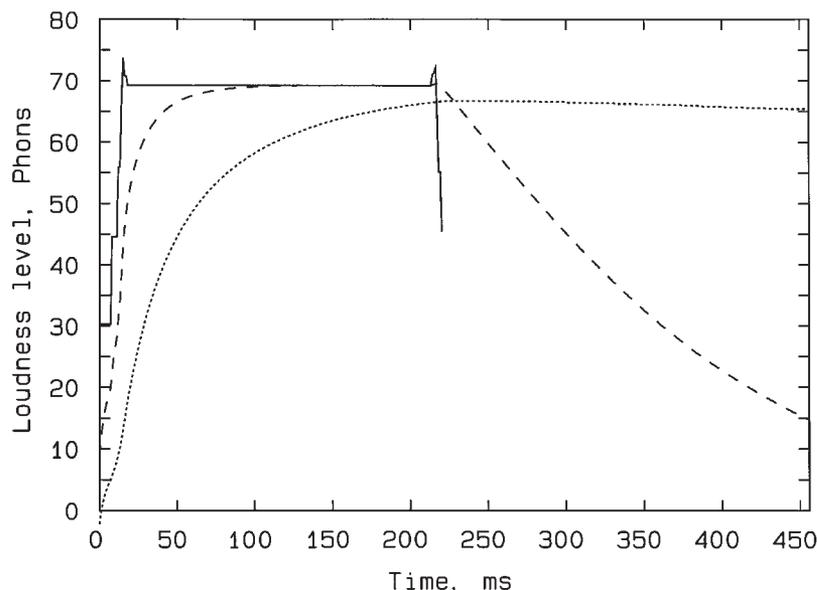


Fig. 5. Output of model in response to 4000-Hz 200-ms tone, showing instantaneous loudness (—), short-term loudness (---), and long-term loudness (···), plotted as a function of time.

a role over a wide range of durations and the slope is generally less than 3 dB per doubling of duration. Also, the loudness is always greater for the signal without ramps than for the signal with ramps. This happens because the ramps decrease the spectral spread of the signal.

In conclusion, the model accounts well for the empirically observed features of the temporal integration of loudness.

3.3 Long-Term Loudness of Amplitude-Modulated Sounds

As described in the Introduction, there is some controversy in the literature over the loudness of amplitude-modulated sounds. However, the combined results across studies indicate that for carriers that are amplitude modulated at very low rates, the perceived loudness corresponds to a level between the rms level and the peak level. For sounds that are modulated at intermediate rates, the loudness corresponds to a level close to but slightly below the rms level. To generate model predictions of the loudness of amplitude-modulated sounds, we determined the level of a steady tone that gave the same predicted long-term loudness as a modulated tone of the same frequency.

For modulation rates up to 10 Hz, the long-term loudness estimate fluctuates slightly, although at 10 Hz the fluctuation amounts to only about 0.5 phon; the time constants used in the second averager are not sufficient completely to remove these fluctuations. This corresponds to the subjective impression of listeners when attempting to judge the overall loudness of sounds that are amplitude modulated at low rates. The listeners complain that it is difficult to make a judgment of overall loudness because the loudness is continually changing. To make predictions for these cases, we have used the mean value of the long-

term loudness produced by the amplitude-modulated tone.

The solid line in Fig. 7(a) shows predictions of the model for a 4000-Hz carrier that is 100% sinusoidally amplitude modulated at rates from 2 to 1000 Hz. The figure shows the difference in rms level between the modulated and the unmodulated tones required to give equal loudness. If loudness were determined by the peak level for low modulation rates, the level difference would be -4.2 dB. In fact, the difference is about -2.5 dB for the 2-Hz modulation rate and increases to slightly positive values for rates from 30 to 100 Hz. This is in good correspondence with the empirical data reviewed in the Introduction. For even higher rates the level difference decreases and becomes negative, because the spectral sidebands are resolved and a loudness summation effect across frequency occurs. Again, this accords well with the experimental data [12], [18]. The results for a 1000-Hz carrier are shown by the dashed line. They are similar to those for the 4000-Hz carrier, except that the level difference starts to become negative at a smaller bandwidth. This happens because the ERB of the auditory filter is smaller at 1000 than at 4000 Hz, so spectral sidebands are resolved at lower modulation rates at 1000 Hz.

The solid line in Fig. 7(b) shows predictions for a 4000-Hz carrier that was sinusoidally amplitude modulated on a decibel scale; we refer to this as dB modulation. The peak-to-valley ratio of the modulation was 60 dB. With this modulation depth, the ratio of the peak value of the envelope to the rms value is 8.1 dB. Loudness matches to sounds of this type were obtained by Moore et al. [16], [17]. For the 2-Hz modulation rate, the level of the modulated sound is about 3 dB below the level of the unmodulated sound at the point of equal loudness. For the 40-Hz

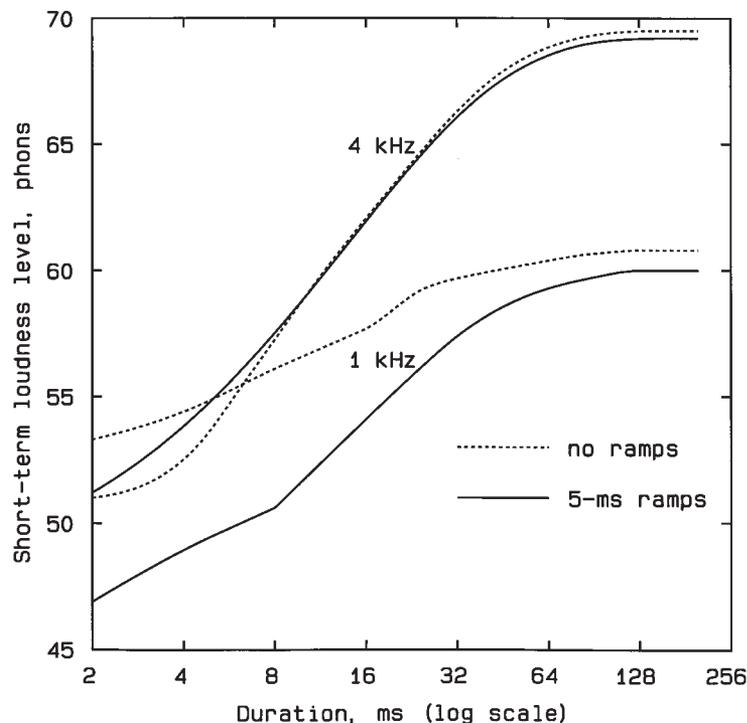


Fig. 6. Short-term loudness level as a function of duration of steady-state portion of tone pulses with frequencies of 1000 and 4000 Hz. Tone pulses were either gated on and off abruptly (\cdots) or had raised-cosine 5-ms ramps (—).

rate, the level of the modulated sound is about 3 dB above the level of the unmodulated sound at the point of equal loudness. These results are broadly consistent with the empirical data. For high modulation rates, the difference in level decreases and becomes negative. This can be attributed to the spectral spread of the dB-modulated signal, which starts to become significant for the 100-Hz modulation rate. Again, results are similar for the 1000-Hz carrier (dashed line) except that the level difference becomes negative at a smaller bandwidth than for the 4000-Hz carrier.

4 LIMITATIONS OF THE MODEL

Some of the limitations of our earlier model [1] also apply to the present model, but will not be repeated here. Although the present model can be applied to time-varying sounds, the method of calculating a short-term spectrum to derive an excitation pattern does not represent accurately the way that excitation patterns are evoked in the human auditory system, via a bank of level-dependent overlapping filters. In particular, our model does not take into account the fact that the auditory filters have a phase characteristic with significant curvature [48], [49]. Because of this curvature, harmonic complex sounds with identical power spectra can give rise to waveforms on the basilar membrane with very different peak factors (ratio of peak amplitude to rms amplitude), depending on their phase spectra [48], [49]. This in turn may lead to differences in loudness [50], [51]. The present model does not correctly predict the effects of the relative phases of the

components on loudness. It should be noted, however, that such effects are mainly apparent for synthesized sounds presented via earphones. When listening in a room, the phase spectrum is to a large extent randomized by room reflections, except when the listener is close to the sound source [52].

5 IMPLEMENTATION OF THE MODEL

The model requires specification of the conditions of presentation. These affect the FIR filter used to simulate the effect of transmission through the outer and middle ear. The options at present are free-field (frontal incidence), diffuse-field, or flat response at the eardrum. The last option is appropriate for earphones designed to have a flat response at the eardrum (such as Etymotic Research ER2 insert earphones). For earphones designed to have a diffuse-field response (such as Etymotic Research ER4 or ER6, Sennheiser HD 414 and HD 580, and many others), the diffuse-field option can be used.

The model also requires specification of a reference level. For example, it can be “told” that a full-scale 16-bit sinusoid corresponds to a free-field level (that is, the level measured at the point corresponding to the center of the listener’s head, the listener having been removed from the sound field) of, say, 100 dB.

6 CONCLUSIONS

We have described a model that can be used to predict the loudness of a sound from its waveform. The model has

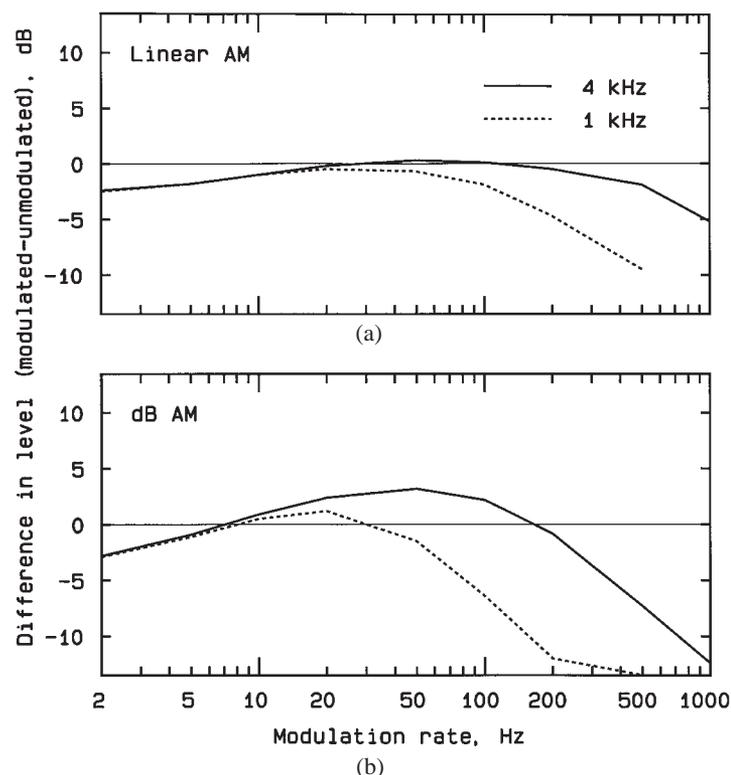


Fig. 7. Difference in rms level required for predicted equal loudness of modulated and unmodulated tones, plotted as a function of modulation rate. Carrier frequency: — 4000 Hz; · · · · 1000 Hz. (a) Sinusoidal modulation on linear scale with 100% modulation. (b) Modulation on dB scale with peak-to-valley ratio of 60 dB.

the advantage that it can be applied directly to the sound picked up by a microphone, and does not require an intermediate stage of spectral analysis. The model can be applied both to steady sounds, such as ventilation and air-conditioning noise, and to time-varying sounds, such as speech and music. For such sounds the model gives estimates both of the momentary short-term loudness and of the long-term overall loudness impression. This may be useful for monitoring the loudness of broadcast sounds, including advertisements. The model gives accurate results for sounds with both broad-band and narrow-band spectra, and mixtures of the two.

7 ACKNOWLEDGMENT

This work was supported by the Medical Research Council (UK). The authors wish to thank Michael Stone and Thomas Baer for their substantial contributions to this work. They also thank two anonymous reviewers for helpful comments.

8 REFERENCES

- [1] B. C. J. Moore, B. R. Glasberg, and T. Baer, "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness," *J. Audio Eng. Soc.*, vol. 45, pp. 224–240 (1997 Apr.).
- [2] H. Fletcher and W. A. Munson, "Relation between Loudness and Masking," *J. Acoust. Soc. Am.*, vol. 9, pp. 1–10 (1937).
- [3] E. Zwicker, "Über psychologische und methodische Grundlagen der Lautheit," *Acustica*, vol. 8, pp. 237–258 (1958).
- [4] E. Zwicker and B. Scharf, "A Model of Loudness Summation," *Psychol. Rev.*, vol. 72, pp. 3–26 (1965).
- [5] E. Zwicker, H. Fastl, and C. Dallmayr, "BASIC-Program for Calculating the Loudness of Sounds from Their 1/3-Octave Band Spectra According to ISO 532B," *Acustica*, vol. 55, pp. 63–67 (1984).
- [6] E. Zwicker and H. Fastl, *Psychoacoustics—Facts and Models* (Springer, Berlin, 1990).
- [7] E. Zwicker, "Temporal Effects in Simultaneous Masking by White-Noise Bursts," *J. Acoust. Soc. Am.*, vol. 37, pp. 653–663 (1965).
- [8] E. Zwicker, "Procedure for Calculating Loudness of Temporally Variable Sounds," *J. Acoust. Soc. Am.*, vol. 62, pp. 675–682 (1977).
- [9] H. Fastl, "Loudness Evaluation by Subjects and by a Loudness Meter," in *Sensory Research—Multimodal Perspectives*, R. T. Verrillo, Ed. (Erlbaum, Hillsdale, NJ, 1993), pp. 199–210.
- [10] M. A. Stone, B. C. J. Moore, and B. R. Glasberg, "A Real-Time DSP-Based Loudness Meter," in *Contributions to Psychological Acoustics*, A. Schick and M. Klatte, Eds. (Bibliotheks- und Informationssystem der Universität Oldenburg, Oldenburg, Germany, 1997), pp. 587–601.
- [11] ISO 532, "Acoustics—Method for Calculating Loudness Level," International Organization for Standardization, Geneva, Switzerland (1975).
- [12] H. Bauch, "Die Bedeutung der Frequenzgruppe für die Lautheit von Klängen," *Acustica*, vol. 6, pp. 40–45 (1956).
- [13] B. R. Glasberg and B. C. J. Moore, "Derivation of Auditory Filter Shapes from Notched-Noise Data," *Hear. Res.*, vol. 47, pp. 103–138 (1990).
- [14] R. P. Hellman, "Perceived Magnitude of Two-Tone-Noise Complexes: Loudness, Annoyance, and Noisiness," *J. Acoust. Soc. Am.*, vol. 77, pp. 1497–1504 (1985).
- [15] H. Fastl, "Loudness and Masking Patterns of Narrow Noise Bands," *Acustica*, vol. 33, pp. 266–271 (1975).
- [16] B. C. J. Moore, S. Launer, D. Vickers, and T. Baer, "Loudness of Modulated Sounds as a Function of Modulation Rate, Modulation Depth, Modulation Waveform and Overall Level," in *Psychophysical and Physiological Advances in Hearing*, A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis, Eds. (Whurr, London, 1998), pp. 465–471.
- [17] B. C. J. Moore, D. A. Vickers, T. Baer, and S. Launer, "Factors Affecting the Loudness of Modulated Sounds," *J. Acoust. Soc. Am.*, vol. 105, pp. 2757–2772 (1999).
- [18] C. Zhang and F. G. Zeng, "Loudness of Dynamic Stimuli in Acoustic and Electric Hearing," *J. Acoust. Soc. Am.*, vol. 102, pp. 2925–2934 (1997).
- [19] W. Jesteadt, "An Adaptive Procedure for Subjective Judgments," *Percept. Psychophys.*, vol. 28, pp. 85–88 (1980).
- [20] G. Grimm, V. Hohmann, and J. L. Verhey, "Loudness of Fluctuating Sounds," *Acustica—Acta Acustica* (2002) (in press).
- [21] B. Scharf, "Loudness," in *Handbook of Perception*, vol. IV: *Hearing*, E. C. Carterette and M. P. Friedman, Eds. (Academic Press, New York, 1978), pp. 187–242.
- [22] M. Florentine, S. Buus, and T. Poulsen, "Temporal Integration of Loudness as a Function of Level," *J. Acoust. Soc. Am.*, vol. 99, pp. 1633–1644 (1996).
- [23] S. Buus, M. Florentine, and T. Poulsen, "Temporal Integration of Loudness, Loudness Discrimination, and the Form of the Loudness Function," *J. Acoust. Soc. Am.*, vol. 101, pp. 669–680 (1997).
- [24] E. A. G. Shaw, "Transformation of Sound Pressure Level from the Free Field to the Eardrum in the Horizontal Plane," *J. Acoust. Soc. Am.*, vol. 56, pp. 1848–1861 (1974).
- [25] H. Fletcher, "Auditory Patterns," *Rev. Mod. Phys.*, vol. 12, pp. 47–65 (1940).
- [26] B. C. J. Moore and B. R. Glasberg, "Formulae Describing Frequency Selectivity as a Function of Frequency and Level and Their Use in Calculating Excitation Patterns," *Hear. Res.*, vol. 28, pp. 209–225 (1987).
- [27] R. J. Baker, S. Rosen, and A. M. Darling, "An Efficient Characterisation of Human Auditory Filtering across Level and Frequency that Is also Physiologically Reasonable," in *Psychophysical and Physiological Advances in Hearing*, A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis, Eds. (Whurr, London, 1998), pp. 81–87.
- [28] S. Rosen, R. J. Baker, and A. Darling, "Auditory Filter Nonlinearity at 2 kHz in Normal Hearing Listeners," *J. Acoust. Soc. Am.*, vol. 103, pp. 2539–2550 (1998).
- [29] B. R. Glasberg and B. C. J. Moore, "Frequency Selectivity as a Function of Level and Frequency Mea-

sured with Uniformly Exciting Notched Noise,” *J. Acoust. Soc. Am.*, vol. 108, pp. 2318–2328 (2000).

[30] B. C. J. Moore and B. R. Glasberg, “Suggested Formulae for Calculating Auditory-Filter Bandwidths and Excitation Patterns,” *J. Acoust. Soc. Am.*, vol. 74, pp. 750–753 (1983).

[31] C. Giguère and P. C. Woodland, “A Computational Model of the Auditory Periphery for Speech and Hearing Research. I. Ascending Path,” *J. Acoust. Soc. Am.*, vol. 95, pp. 331–342 (1994).

[32] T. Irino and R. D. Patterson, “A Compressive Gammachirp Auditory Filter for Both Physiological and Psychophysical Data,” *J. Acoust. Soc. Am.*, vol. 109, pp. 2008–2022 (2001).

[33] X. Zhang, M. G. Heinz, I. C. Bruce, and L. H. Carney, “A Phenomenological Model for the Responses of Auditory-Nerve Fibers: I. Nonlinear Tuning with Compression and Suppression,” *J. Acoust. Soc. Am.*, vol. 109, pp. 648–670 (2001).

[34] R. Meddis and L. O’Mard, “A Computational Algorithm for Computing Nonlinear Auditory Frequency Selectivity,” *J. Acoust. Soc. Am.*, vol. 109, pp. 2852–2861 (2001).

[35] A. Kohlrausch, R. Fassel, and T. Dau, “The Influence of Carrier Level and Frequency on Modulation and Beat-Detection Thresholds for Sinusoidal Carriers,” *J. Acoust. Soc. Am.*, vol. 108, pp. 723–734 (2000).

[36] B. C. J. Moore and B. R. Glasberg, “Temporal Modulation Transfer Functions Obtained Using Sinusoidal Carriers with Normally Hearing and Hearing-Impaired Listeners,” *J. Acoust. Soc. Am.*, vol. 110, pp. 1067–1073 (2001).

[37] G. K. Yates, “Cochlear Structure and Function,” in *Hearing*, B. C. J. Moore, Ed. (Academic Press, San Diego, CA, 1995), pp. 41–73.

[38] M. A. Ruggero, N. C. Rich, A. Recio, S. S. Narayan, and L. Robles, “Basilar-Membrane Responses to Tones at the Base of the Chinchilla Cochlea,” *J. Acoust. Soc. Am.*, vol. 101, pp. 2151–2163 (1997).

[39] A. J. Oxenham, “Forward Masking: Adaptation or Integration?,” *J. Acoust. Soc. Am.*, vol. 109, pp. 732–741 (2001).

[40] ISO 389-7, “Acoustics—Reference Zero for the Calibration of Audiometric Equipment. Part 7: Reference

Threshold of Hearing under Free-Field and Diffuse-Field Listening Conditions,” International Organization for Standardization, Geneva, Switzerland (1996).

[41] ISO 226, “Acoustics—Normal Equal-Loudness Contours,” International Organization for Standardization, Geneva, Switzerland (1987).

[42] Y. Suzuki and T. Sone, “Frequency Characteristics of Loudness Perception: Principles and Applications,” in *Contributions to Psychological Acoustics*, A. Schick, Ed. (Bibliotheks- und Informationssystem der Universität Oldenburg, Oldenburg, Germany, 1994), pp. 193–221.

[43] B. Gabriel, B. Kollmeier, and V. Mellert, “Influence of Individual Listener, Measurement Room and Choice of Test-Tone Levels on the Shape of Equal-Loudness Level Contours,” *Acustica—Acta Acustica*, vol. 83, pp. 670–683 (1997).

[44] E. Zwicker, G. Flottorp, and S. S. Stevens, “Critical Bandwidth in Loudness Summation,” *J. Acoust. Soc. Am.*, vol. 29, pp. 548–557 (1957).

[45] B. Scharf, “Complex Sounds and Critical Bands,” *Psychol. Bull.*, vol. 58, pp. 205–217 (1961).

[46] B. Scharf, “Critical Bands,” in *Foundations of Modern Auditory Theory*, J. V. Tobias, Ed. (Academic Press, New York, 1970), pp. 157–202.

[47] P. Bonding and C. Elberling, “Loudness Summation across Frequency under Masking and in Sensorineural Hearing Loss,” *Audiology*, vol. 19, pp. 57–74 (1980).

[48] A. Kohlrausch and A. Sander, “Phase Effects in Masking Related to Dispersion in the Inner Ear. II. Masking Period Patterns of Short Targets,” *J. Acoust. Soc. Am.*, vol. 97, pp. 1817–1829 (1995).

[49] A. Recio and W. S. Rhode, “Basilar Membrane Responses to Broadband Stimuli,” *J. Acoust. Soc. Am.*, vol. 108, pp. 2281–2298 (2000).

[50] R. P. Carlyon and A. J. Datta, “Excitation Produced by Schroeder-Phase Complexes: Evidence for Fast-Acting Compression in the Auditory System,” *J. Acoust. Soc. Am.*, vol. 101, pp. 3636–3647 (1997).

[51] H. Gockel, B. C. J. Moore, and R. D. Patterson, “Influence of Component Phase on the Loudness of Complex Tones,” *Acustica—Acta Acustica* (2002) (in press).

[52] R. Plomp and H. J. M. Steeneken, “Place Dependence of Timbre in Reverberant Sound Fields,” *Acustica*, vol. 28, pp. 50–59 (1973).

THE AUTHORS



B. R. Glasberg



B. C. J. Moore

Brian R. Glasberg received B.Sc. and Ph.D. degrees in applied chemistry from Salford University in 1968 and 1972, respectively.

He then worked as a chemist refining precious metals. He spent some time as a researcher into process control before joining the laboratory of Brian C. J. Moore at the University of Cambridge, initially as a research associate and then as a senior research associate.

Dr. Glasberg's research focuses on the perception of sound in both normally hearing and hearing-impaired people. He also works on the development and evaluation of hearing aids, especially digital hearing aids. He is a member of the Acoustical Society of America and has published 95 research papers and book chapters.



Brian C. J. Moore received a B.A. degree in natural sciences in 1968 and a Ph.D. degree in psychoacoustics in 1971, both from the University of Cambridge, UK.

He is currently professor of auditory perception at the University of Cambridge. He has also been a visiting professor at Brooklyn College, the City University of New York, and the University of California at Berkeley and

was a van Houten Fellow at the Institute for Perception Research, Eindhoven, The Netherlands.

Dr. Moore's research interests include perception of sound, mechanisms of normal hearing and hearing impairments, relationship of auditory abilities to speech perception, design of signal processing hearing aids for sensorineural hearing loss, methods for fitting hearing aids to the individual, and design and specification of high-fidelity sound-reproducing equipment. He is a fellow of the Acoustical Society of America and the Academy of Medical Sciences, and is an honorary fellow of the Belgian Society of Audiology and the British Society of Hearing Aid Audiologists. He is a member of the Experimental Psychology Society (U.K.), the British Society of Audiology, the American Speech-Language Hearing Association, the American Auditory Society, the Acoustical Society of Japan, the Audio Engineering Society, and the Association for Research in Otolaryngology. He is president of the Association of Independent Hearing Healthcare Professionals (UK). He has published 9 books and over 350 scientific papers and book chapters. He is wine steward of Wolfson College, Cambridge.