Audio Engineering Society

Convention Paper 8873

Presented at the 134th Convention
2013 May 4–7 Rome, Italy

A Practical Step-by-Step Guide to the Time-Varying Loudness Model of Moore, Glasberg and Baer (1997; 2002)

Andrew J. R. Simpson[1], Michael J. Terrell[1], and Joshua D. Reiss[1]

[1] Centre for Digital Music, Queen Mary University of London, London E1 4NS, UK

Andy.Simpson@eecs.qmul.ac.uk; Micheal.Terrell@eecs.qmul.ac.uk; Josh.Reiss@eecs.qmul.ac.uk

ABSTRACT

In this tutorial article we provide a condensed, practical step-by-step guide to the excitation pattern loudness model of Moore, Glasberg and Baer [J. Audio Eng. Soc. **45**, 224–240, 1997; J. Audio Eng. Soc. **50**, 331–342, 2002]. The various components of this model have been separately described in the well known publications of Patterson *et al.* [J. Acoust. Soc. Am. **72**, 1788–1803, 1982], Moore [Hearing (Academic Press), 161-205, 1995], Moore *et al.* (1997) and Glasberg and Moore (2002). This paper provides a consolidated and concise introduction to the complete model for those who find the disparate and complex references intimidating and who wish to understand the function of each of the component parts. Furthermore, we provide a consolidated notation and integral forms. This introduction may be useful to the loudness theory beginner and to those who wish to adapt and apply the model for novel, practical purposes.

1. INTRODUCTION

The loudness model of Moore, Glasberg and Baer [1], later extended to include time-varying sounds by Glasberg and Moore [2], is ubiquitous in several sound related research fields. One possible reason for the popularity of this model is the fact that it is often necessary to account for psychoacoustic phenomena such as the frequency-dependent hearing thresholds, level-dependent compression and masking. This empirical model accounts for all the major behavioral phenomena of psychoacoustics and so provides an ideal entry point to the application of auditory models to a variety of modeling problems relating to auditory perception. Furthermore, the model employs common signal processing and computational techniques, and multiple implementations are freely available to download. However, the model has seen a long and fragmented development over a period of more than twenty years [1-4], from the rounded exponential 'roex' filter defined by Patterson *et al.* in 1982 to the time-varying model of Glasberg and Moore in 2002. As such, the complexity and heritage of this model provide a somewhat intimidating first step into the mature field of psychoacoustics. In this article, we attempt to lay it out as simply and completely as possible for those who wish to understand its workings well enough to apply it.

2. THE EXCITATION PATTERN MODEL

Sound pressure waves pass through the outer and middle ear and enter the inner ear (cochlea), causing the basilar membrane to resonate at a given location along its length that depends on the frequency of the exciting sound. Resonance of the basilar membrane causes the displacement (shearing) of inner hair cells arranged along the basilar membrane. The extent of the shearing of each hair cell is then converted into a pulsed electrical signal by neurons attached to the hair cell. This neural representation of the pattern of resonance on the basilar membrane, caused by a given sound, is known as its excitation pattern. The electrical signal, produced by the population of neurons, is sent up the auditory nerve to the brain. This gives rise to the concept of the auditory filter, which specifies the shape of the excitation pattern for a sound of given frequency and level. To make things more complicated, there are also outer hair cells which contribute little in the way of signals sent to the brain, but which are motile and act in synchrony with the corresponding inner hair cell to amplify the basilar membrane excitation at low levels. This produces the effect of changing the shape of the auditory filter with level.

The excitation pattern model of loudness is based on the assumption that the total area of excitation along the length of the basilar membrane is integrated (on the auditory nerve) in the calculation of loudness. However, consistent with what is known of the cochlear amplifier and of neural transduction, the excitation is locally compressed before being integrated. The role of the auditory filter is to provide a summation of energy at local frequencies (i.e., within the auditory filter), and subsequent compression of the sum energy at the output of the auditory filter. The output of the auditory filter is known as specific loudness (loudness per filter). This mechanism results in energetic (simultaneous) masking because the specific loudness resulting from the compressed sum of excitation at two nearby locations (i.e., within a single auditory filter) contributes less to overall loudness than a linear sum of the specific loudness resulting from the same excitation at two distant locations (i.e., within two separate auditory filters).

2.1 Definitions

Loudness is the perceived intensity (I) of a sound. Intensity is defined in terms of sound pressure (x) squared;

$$I = kx^2 \tag{1}$$

where k is a constant that represents the specific acoustic impedance of air. To calculate I for a sound described by $x(t)$, from time t=0 to T, Eq. 1 is then integrated over time (T);

$$I = \frac{1}{T} k \int_0^T x^2(t) dt \tag{2}$$

Intensity may then be defined in terms of a ratio, with respect to a reference (e.g., I_{ref} = 20 μW/cm^2), in decibels. This is known as the intensity *level* (L_I);

$$L_I = 10 \log_{10} \left(\frac{I}{I_{ref}} \right) \tag{3}$$

The use of intensity levels allows us to drop the absolute reference, and with it the k parameter, which simplifies the following notation.

3. EQUIVALENT RECTANGULAR BANDWIDTH

The equivalent rectangular bandwidth (ERB) gives a measure of auditory filter width, such that the sum excitation that falls within any given ERB will be equivalently compressed and result in an equivalent contribution to total loudness. Thus, the ERB provides a mechanism by which compression, masking and loudness are related. The mapping between frequency, f (Hz), and ERB (Hz) shown in Fig. 1a is achieved using the following formula (see Moore, [4]):

$$ERB = 24.7(0.00437 f + 1) \tag{4}$$

In order to relate ERB to frequency, the ERB number (n) for a given centre frequency (f_c) - as shown in Fig. 1b - can be calculated as (see Moore, [4]):

$$n = 21.4 \log_{10}(0.00437 f_c + 1) \tag{5}$$

Given frequency bounds defined in terms of centre frequencies between 50 – 15,000 Hz (see [1]), the ERB numbers of the respective upper and lower bounding auditory filters may be calculated and intervening filters specified at arbitrary ERB-scaled intervals. To this end, Eq. 5 may be rewritten, as follows;

$$f_C = \frac{10^{(n/21.4)} - 1}{0.00437} \quad (6)$$

Using Eq. 5, auditory filters at ERB intervals between the known range of the basilar membrane (50 - 15,000 Hz) may be specified for the later excitation pattern calculation and using Eq. 6 the centre frequencies may be calculated at ERB-spaced intervals between.

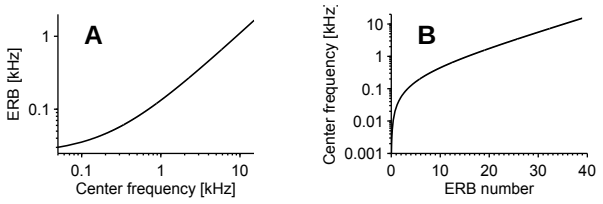

Figure 1. (a) Illustration of Eq. 4 which relates ERB to centre frequency. (b) Illustration of Eq.s 5 & 6 which relates centre frequency to ERB number.

4. MODEL FOR STEADY SOUNDS

The first stage of the model represents the transformation of sound pressure through the outer and middle ear to the inner ear (cochlea). This transformation is represented by a fixed, linear filter, with a frequency dependent gain, *y*, as follows;

$$L_{I_1} = L_{I_0} \cdot y \quad (7)$$

where L_{I_0} is the input sound intensity, L_{I_1} is the intensity reaching the inner ear and *y* is the gain of the filter at that frequency. Fig. 2 provides an illustration of the combined transfer function. Because the loudness model of Moore *et al.* is generally intended for diffuse-field sound, phase information is discarded (see [2] for discussion).

From this point onwards it is important to note that the input sound signal is defined as an intensity level (Eq. 3) at a specific frequency, wherever a dB measure is used. Furthermore, excitation is defined in terms of excitation level (L_E) as an intensity ratio with respect to the excitation reference of a 1 kHz sinusoidal signal at 0 dB SPL (presented in the free field and at frontal incidence);

$$L_E = L_{I_1} - E(0_{1kHz}) \quad (8)$$

where $E(0_{1kHz})$ is the reference excitation level.

4.1 The rounded exponential (roex) filter

The excitation pattern, which represents the basilar membrane response, is calculated using a set of ERB-spaced auditory filters. The auditory filter is based on the rounded-exponential ('roex') form proposed by Patterson *et al.* [3]. The roex filter is defined as;

$$w(g) = (1 + pg)e^{-pg} \quad (9)$$

where for a given centre frequency, f_c, the normalized frequency relationship between the f_c and a given frequency, *f*, (i.e., of the input signal) is given by;

$$g = |f - f_c| / f_c \quad (10)$$

where f_c is evaluated for a given ERB number (*n*) using Eq. 6, and where *p* determines bandwidth and slope of the filter and is defined in relation to the ERB as follows [4]:

$$p = \frac{4 f_C}{ERB} \quad (11)$$

Larger values of *p* lead to more narrowly tuned filters. Thus, given an input at frequency *f*, *w*(*g*) can be used to calculate the attenuation of the input at frequency *f* within the roex filter at centre frequency f_c.

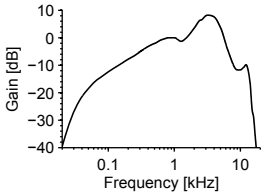

Figure 2. Illustration of combined outer and middle ear transfer function, with zero dB gain at 1 kHz.

Eq. 9 gives a symmetrical auditory filter [*w*(*g*)]. However, the auditory filter is known to be asymmetrical and so Eq. 9 is broken down into two such expressions, the choice of which depends on whether the input frequency (*f*) is above or below the centre frequency (f_c) for the auditory filter of interest;

$$w(g) = \begin{cases} (1 + p_l g)e^{-p_l g} & for \quad f <= f_c \\ (1 + p_u g)e^{-p_u g} & for \quad f > f_c \end{cases} \qquad (12)$$

p_l and p_u replace p to represent the parameters for input frequencies (f) below or above the centre frequency (f_c) respectively. This conditional aspect is necessary because although the auditory filter is roughly symmetrical when the excitation level per ERB is around 51 dB [5], the low-frequency 'skirt' of the auditory filter becomes less sharp with increase in level. This excitation level dependent relationship is accommodated in terms of the p_l value as follows;

$$p_l(L_E) = p_l(51) - 0.35\left(p_l(51) / p_l(51_{1kHz})\right)(L_E - 51) \qquad (13)$$

where $p_l(L_E)$ is the value of p_l for the input excitation level of L_E, in dB, at f, and $p_l(51)$ is the value of p (Eq. 11) at the centre frequency (f_c) for an input level of 51 dB (i.e., where the filter is symmetrical), and where $p_l(51_{1kHz})$ is the value of p_l for a 51 dB input excitation level at 1 kHz. Figure 3 provides an illustration of the level dependent roex filter shape for excitation at 1 kHz at levels between 10 and 100 dB in 10 dB intervals.

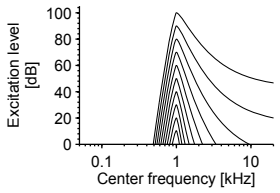

Figure 3. Illustration of roex filter shapes (Eq. 12) for excitation levels between 10 and 100 dB in 10 dB intervals.

4.2 The excitation pattern

For each ERB number (n), the excitation pattern, E, is defined as the pattern of outputs from the ERB-spaced auditory filters. For a given frequency, f, and for an input excitation level (Eq. 8), the excitation pattern, E, is then defined as:

$$E(n) = w\left(g(n)\right) \cdot L_E \qquad (14)$$

where ERB number (n) is related to f_c by Eq. 6.

4.3 Specific loudness

To reflect the production of neural signals in response to inner hair cell displacement caused by excitation of the basilar membrane, the excitation pattern is then transformed from excitation level into specific loudness (loudness per ERB) for the nth auditory filter by calculating the specific loudness in each filter according to three possible conditional expressions, which relate to the excitation level as follows in Eq. 15 (above). Since loudness is later notated as N, specific loudness is notated as N', to reflect the later integration (over frequency) of specific loudness to form loudness.

Frequency dependence (denoted with parameter n) refers to the f_c for the nth auditory filter. T_Q represents the threshold excitation in quiet and is frequency dependent as shown in Fig. 4c. The parameter G represents low-level gain in the cochlear amplifier, relative to the gain at 500 Hz and above, and is also frequency dependent. For a given centre frequency, f_c, G (in dB) is related to T_Q (in dB) with a simple subtraction;

$$G = T_Q(500) - T_Q(f_c) \qquad (16)$$

The parameter A is used to bring the input-output function close to linear around the absolute threshold, and is dependent on the value of G as shown in Fig. 4a. The compressive exponent α is dependent on the value of G as shown in Fig. 4b. At frequencies below 500 Hz T_Q rises as frequency decreases and the value ranges between 28 dB at 50 Hz and 3.7 dB at 500 Hz. Above 500 Hz T_Q is constant and equal to T_Q at 500 Hz. α is also frequency-dependent and a similar lookup table is employed such that α varies between 0.27 and 0.2,

$$N'(n) = \begin{cases} C \cdot \left(\dfrac{2E(n)}{E(n) + T_Q(n)}\right)^{1.5} \cdot \left((G \cdot E(n) + A)^{\alpha} - A^{\alpha}\right) & for \ E(n) \le T_Q(n) \\ C \cdot \left((G \cdot E(n) + A)^{\alpha} - A^{\alpha}\right) & for \ 10^{10} \ge E(n) > T_Q(n) \\ C \cdot \left(\dfrac{E(n)}{1.04x10^6}\right)^{0.5} & for \ E(n) > 10^{10} \end{cases} \qquad (15)$$

depending on the value of *G*. *C* is a constant which scales the loudness to conform to the sone scale, where the loudness of a 1 kHz tone at 40 dB SPL is 1 sone. *C* is equal to 0.047. Figure 4d shows the result of Eq. 16 used to transform excitation at levels between 0 and 120 dB to specific loudness for a 1 kHz signal. Finally, specific loudness is integrated, over the arbitrarily (*dn*) spaced auditory filters, between ERB numbers n_{min} and n_{max}, to produce loudness, *N*;

$$N = \int_{n_{\min}}^{n_{\max}} N'(n)dn \tag{17}$$

where n_{min} and n_{max} may be calculated, from centre frequencies of 50 and 15,000 Hz respectively, using Eq. 5.

For a complex sound, loudness is calculated from a linear sum of excitation patterns calculated from each input sound component.

4.4 Energetic masking

A formal definition of loudness allows us to derive a formal definition of energetic (simultaneous) masking with respect to two arbitrary excitation patterns for the target, $E_{t,}$ and the masker, E_m. The two excitation patterns may then be used to evaluate the degree of masking by comparing the sum of loudness for each excitation pattern alone [$N(E_m)$ + $N(E_t)$] and the loudness of the linear sum of the two excitation patterns [$N(E_m+E_t)$]. This provides a loudness ratio ($N_{masking}$, in sones);

$$N_{masking} = \frac{N(E_t + E_m)}{N(E_t) + N(E_m)} \tag{18}$$

5. MODEL FOR TIME-VARYING SOUNDS

The time-varying model [2] is an extension of the 1997 model for steady (state) sounds. In the earlier model, the sounds are defined in terms of steady sound components, which are then combined within the excitation pattern to produce an overall loudness. In the time-varying model, the excitation pattern is typically calculated, from a time-domain input signal, on an instantaneous sliding-window basis, giving a time-varying excitation pattern.

The time-varying excitation pattern is then resolved into a corresponding time-varying specific loudness function and hence is integrated to form a time-varying intermediate stage known as 'instantaneous loudness'. Instantaneous loudness is essentially an intensity-like temporal integration of specific loudness over an arbitrarily small time interval. The 'small' time interval is typically in the order of 1 ms, which may be considered small with respect to the integration time constants of the auditory system (usually much longer). Thus, instantaneous loudness is calculated as 'steady loudness' over a very small time scale.

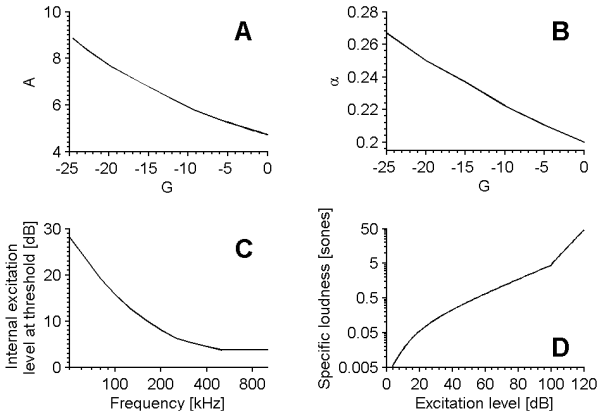

Figure 4. (a) Illustrating the relation between parameters A and G. (b) Illustrating the relationship between the parameters α and G. (c) illustrating the internal excitation level at threshold as a function of frequency (showing increased thresholds at low frequencies). (d) Specific loudness as a function of excitation level between 0 and 120 dB at 1 kHz, illustrating the conditional effects of Eq. 15.

$$N'(n,t)=\begin{cases} C\cdot\left(\dfrac{2E(n,t)}{E(n,t)+T_Q(n,t)}\right)^{1.5}\cdot\left((G\cdot E(n,t)+A)^{\alpha}-A^{\alpha}\right) & for\ E(n,t)\le T_Q(n) \\ C\cdot\left((G\cdot E(n,t)+A)^{\alpha}-A^{\alpha}\right) & for\ 10^{10}\ge E(n,t)>T_Q(n) \\ C\cdot\left(\dfrac{E(n,t)}{1.04x10^{6}}\right)^{0.5} & for\ E(n,t)>10^{10} \end{cases} \quad (20)$$

Intensity, for sound of a given integration time (Δt), is then defined in terms of an integral with respect to time (t);

$$I(t)=\frac{1}{\Delta t}k\int_{t}^{t+\Delta t}x^2(t)dt \quad (19)$$

which may again be resolved into intensity level as in Eq. 4, and hence excitation level as in Eq. 8, for substitution into Eq. 15 to give Eq. 20.

Eq. 17 is then extended to integrate the result of Eq. 20 with respect to ERB number (n), to produce a time-varying instantaneous loudness [$N(t)$];

$$N(t)=\int_{n_{\min}}^{n_{\max}}N'(n,t)dn \quad (21)$$

5.1 Temporal integration

Loudness of brief sounds increases with duration up to a limit of around 200 ms [6]. This is known as the temporal integration of loudness. A further phenomenon captured in the time-varying loudness model is forward masking, which has a similar time scale. In order to account for these phenomena, the instantaneous loudness function is smoothed with an exponential sliding window.

To predict the decay of loudness after a sound has ceased, given an initial loudness value (N_0), the decaying value of loudness at time t may be calculated as;

$$N(t)=N_0e^{-t/\tau} \quad (22)$$

where τ is the time constant. This represents the decay of loudness (i.e., forward masking). To predict the accumulation of loudness with duration of a steady (fixed intensity) sound, loudness at time t is calculated as;

$$N(t)=N_{\infty}(1-e^{-t/\tau}) \quad (23)$$

where N_{∞} represents the asymptotic loudness. The values of N_0 and N_{∞} may be calculated in terms of instantaneous loudness for a given signal and used to predict the effects of temporal integration.

In order to provide a time-varying output function, Eq. 22 is re-arranged in order to relate it to the time step of the model (Δt) and used to calculate a smoothing coefficient (β);

$$\beta=e^{-\Delta t/\tau} \quad (24)$$

To smooth the time-varying instantaneous loudness function, N_0 and N_{∞} are replaced and β is applied to calculate STL (N_{ST}) with respect to instantaneous loudness [$N(t)$];

$$N_{ST}(t)=(1-\beta)\cdot N(t)+\beta\cdot N_{ST}(t-\Delta t) \quad (25)$$

And to calculate LTL (N_{LT}) with respect to STL;

$$N_{LT}(t)=(1-\beta)\cdot N_{ST}(t)+\beta\cdot N_{LT}(t-\Delta t) \quad (26)$$

where

$$\tau=\begin{cases} 22 & for\ \ N(t)>N_{ST}(t-\Delta t) \\ 50 & for\ \ N(t)<N_{ST}(t-\Delta t) \\ 100 & for\ \ N_{ST}(t)>N_{LT}(t-\Delta t) \\ 2000 & for\ \ N_{ST}(t)<N_{LT}(t-\Delta t) \end{cases} \quad (27)$$

The value of τ (and hence β) is conditional such that separate values of τ are assigned depending on whether the function is in the attack or release phase (see [2]). As can be seen from the values of τ shown above, convergence is much faster for attack than for release in

both cases of STL and LTL. This is intended to reflect disparity in forward and backwards masking.

Finally, Glasberg and Moore [2] specify that the loudness of brief duration sounds (i.e., gated tones) should be calculated as the peak (maximum) value in the STL time series and that the loudness of continuous sounds (e.g., amplitude modulated tones) should be calculated as the mean (average) of the LTL time series.

5.2 Temporal masking

Eq. 18 may be extended to provide a time-varying definition of energetic masking, in terms of instantaneous loudness, as follows;

$$L_{masking}(t) = \frac{N\left(E_t(t) + E_m(t)\right)}{N\left(E_t(t)\right) + N\left(E_m(t)\right)} \tag{28}$$

However, the stated purpose of Eq.s 25 & 26 is to provide temporal integration (or summation) of loudness, at the two respective time scales. This means that forward and backwards masking may not be quantified in terms of Eq. 28, and are therefore outside the scope of this tutorial paper.

6. CONCLUSIONS

In this paper we have provided a condensed, practical step-by-step description of the excitation pattern loudness model which consolidates descriptions found in the multiple original articles. We have included a brief description of the function of, and rationalisation for, each modelling component. This introduction may be useful to those who wish to implement, or adapt the model to novel usage or useful to students new to loudness theory and modelling.

REFERENCES

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

[2] Glasberg, B. R., and Moore, B. C. J. "A model of loudness applicable to time-varying sounds," J. Audio Eng. Soc. 50, pp. 331–342 (2002).

[3] Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. 72, pp. 1788–1803 (1982).

[4] Moore, B. C. J. "Frequency Analysis and Masking," in Hearing, B. C. J. Moore, Ed., (Academic Press, San Diego, California), pp. 161-205 (1995).

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

[6] Munson, W. A. "The growth of auditory sensation," J. Acoust. Soc. Am. 19, pp. 584-591 (1947).