Loudness Scaling in a Digital Synthesis Library

Jesse Guessford*, Hans Kaper†, and Sever Tipei*

*Computer Music Project, School of Music, University of Illinois at Urbana-Champaign

guessfor@uiuc.edu; s-tipei@uiuc.edu

†Mathematics and Computer Science Division, Argonne National Laboratory

kaper@mcs.anl.gov

Abstract

The paper describes the algorithms used in MOSS, an additive synthesis library, to realize sounds with a prescribed loudness. Loudness is a perceived attribute of sound that depends both on the amount of energy carried by the sound wave and on the way this energy is processed by the ear-brain system. The loudness routines in MOSS translate the target loudness of a sound into amplitudes of the constituent partials. The algorithms are based on equal-loudness data published by the International Standards Organization and critical-band data.

1 Introduction

Loudness is one of the most challenging concepts to implement in a digital synthesis library. Loudness is a *perceived* attribute of sound that has as much to do with the amount of energy that the sound wave carries as with the listening environment and the processing of this energy in the ear-brain system of the listener. Moreover, in a complex sound, hundreds of waves interact, and specifying the amplitude of the aggregate is simply not enough to achieve a desired loudness.

In this paper, we describe the algorithms used in MOSS (Music Object-oriented Sound Synthesis) to realize sounds with a prescribed loudness. *Perceived loudness* is one of the degrees of freedom that the user can control. The loudness routines in MOSS translate the target loudness of a sound into amplitudes of the constituent partials. The algorithms are based on equal-loudness data published by the International Sandards Organization and critical-band data.

2 Loudness of a Pure Tone

In MOSS, a sound is constructed as a superposition of *partials*—sinusoidal waves, each with a well defined frequency and amplitude (which may vary with time).

The definition of loudness of a pure tone (a single partial) is based on the *intensity* of the sound wave. The intensity is the energy flux across a unit area. If the intensity is I (watts/m^2), then the *sound intensity level* is

$$\mathrm{SIL} = 10\,\log_{10}(I/I_0). \tag{1}$$

Here, I_0 is a reference value, taken to be $1.0 \cdot 10^{-12}$ watt/m^2.

The sound intensity level is distinct from the *sound pressure level*, which is the quantity that is measured in a sound pressure meter. If Δp is the average pressure variation (newton /m^2), then the sound pressure level is

$$\mathrm{SPL} = 20\,\log_{10}(\Delta p/\Delta p_0), \tag{2}$$

where $\Delta p_0 = 2.0 \cdot 10^{-5}$ newton/m^2. For a pure tone, the average pressure variation is equal to the amplitude of the wave divided by $\sqrt{2}$.

For a free progressive wave in air (that is, a plane wave traveling down a tube or a spherical wave traveling outward from a point source), the intensity is proportional to the square of the amplitude, $I \sim (\Delta p)^2$, so the SIL and SPL coincide, but this is not true in general. Nevertheless, the two are very close, and it is a good approximation to identify the SIL and the SPL. We denote their common value by the symbol L in decibel (dB) units,

$$\mathrm{SIL} \approx \mathrm{SPL} = L. \tag{3}$$

In MOSS, Δp is represented by the amplitude A of the sound envelope. The latter is scaled to the interval $(0, 1)$. Thus, if L_m is the maximum value of L, then

$$L = L_m - 20\,\log_{10}(1/A). \tag{4}$$

The sensation of loudness is strongly frequency dependent. For example, although an SPL of 50 dB at 1,000 Hz is considered *piano*, the same SPL is barely audible at 60 Hz. In 1933, Fletcher and Munson (Fletcher and Munson 1933) published a diagram representing the results of a number of loudness-matching experiments under free-field conditions of listening; see Fig. 1. (A free field is an environment without reflections.) The *equal-loudness level contours*—curves showing the SPL required to make single, sustained, pure tones equally loud—show clearly that, to be perceived as equally loud, very low and very high frequencies require higher intensity levels than frequencies in the middle range of the spectrum of audible sounds.

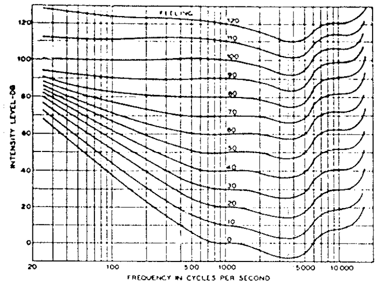

Figure 1: Equal-LL contours (Fletcher and Munson 1933, Fig. 4).

Loudness levels are measured in *phons* (singular *phon*). The Loudness Level (LL) in phons, which we denote by L_p, is numerically equal to the SPL in decibels at 1,000 Hz. The Fletcher–Munson curves range from $L_p = 0$ (threshold of hearing) to $L_p = 120$ (limit of pain) over a frequency range from 25 to 16,000 Hz. The Fletcher–Munson curves are reproduced in the monographs of Jeans (Jeans 1968, Fig. 59) and Roederer (Roederer 1995, Fig. 3.13). Another set of equal-LL contours was given in 1937 by Churcher and King (Churcher and King 1937). The importance of the age of the listener was shown by Robinson and Dadson (Robinson and Dadson 1956). The contours for the younger age group, which show local minima around 4,000 and 12,500 Hz, are reproduced in the monograph of Rossing (Rossing 1982, Figure 6.4). The same contours were recommended for ages 18 to 30 over the frequency range from 20 to 12,500 Hz (up to the second minimum) by the International Organization for Standardization (ISO 1987; ISO 2003). New Standards were published in 2003 (ISO 2003).

It is important to note that these equal-LL curves and their counterparts, the equal-SPL curves (showing the loudness level at constant SPL as a function of the frequency), represent statistical averages. Moreover, they are the results of experiments in a reflection-free environment, so the loudness level perceived in any other listening environment may very well vary.

The quantity L_p still does not measure loudness in an absolute manner: doubling L_p does not lead to the perception that the tone is twice as loud. An absolute measure is given by the LL in *sones* (singular *sone*), here denoted by L_s. It is defined in terms of L_p by the formula (Rossing 1982, Section 6.5)

$$L_s = 2^{(L_p - 40)/10}. \tag{5}$$

This definition assumes that the perceived loudness doubles whenever the loudness level increases by 10 dB, as found experimentally by Stevens (Stevens 1955).

In MOSS, loudness scaling is done on the basis of sones. To achieve a given loudness, the software solves the *inverse problem* of finding the envelope amplitude that yields the target LL at the given frequency. The solution requires three steps, going from a given value of L_s first to L_p, then to L, and finally to Δp or A. The formulae for the first and last step are obtained by inverting Eqs. (4) and (5),

$$L_p = 40 + 10\, \log_2 L_s, \quad A = 10^{-(L_m - L)/20}. \tag{6}$$

The second step, obtaining L from L_p, requires a functional relationship between L and L_p, which needs to be obtained from the equal-LL contours. Various approximations schemes have been proposed in the literature.

Stevens (Stevens 1955) proposed a formula implying that equal increments of L give equal increments of L_p and vice versa,

$$L_p = a + L, \quad L = -a + L_p, \tag{7}$$

with a a constant. This approximation corresponds to the power law approximation of L_s,

$$L_s = k(I/I_0)^{0.3}, \quad L_s = k(\Delta p/\Delta p_0)^{0.6}, \tag{8}$$

with a constant $k = 2^{(a-40)/10}$. The exponent $0.3 = \log_{10} 2$ was established experimentally for a pure tone of 1,000 Hz (cf. Eq. (5)), and the suggestion was made that "for all levels greater than 50 dB the loudness of continuous noises may be calculated from the [same] equation." The first of the two expressions [the one above] is quoted in Rossing (Rossing 1982, Section 6.5). In a later publication, Stevens (Stevens 1970) suggested that the exponent 0.6 in the second formula should be $\frac{2}{3}$, so the law would fit into a general theory that sensory perceptions obey universal power laws with exponents that are simple fractions (like $\frac{2}{3}$).

Steven's law yields equidistant isocurves, clearly not a realistic result in view of Fig. 1. A more realistic approximation is obtained if expressions (7) are generalized to

$$L_p = a + bL, \quad L = -a/b + (1/b)L_p, \tag{9}$$

and both a and b are allowed to vary with f. Thus L_p still varies linearly with L, but not necessarily at the same rate for all frequencies. In this case, Steven's power law (8) becomes

$$L_s = k(I/I_0)^{0.3b} = k(\Delta p/\Delta p_0)^{0.6b}. \tag{10}$$

Robinson and Dadson (Robinson and Dadson 1956) proposed a quadratic approximation,

$$L_p = a + bL + cL^2, \tag{11}$$

with frequency-dependent coefficents a, b, c. This approximation is not significantly better than the linear approximation (9); moreover, the computation of L from L_p requires the evaluation of a square root.

The ISO publication (ISO 1987) gives a rational approximation,

$$L_p = 4.2 + \frac{a(L-T)}{1 + b(L-T)}, L = T + \frac{L_p - 4.2}{a - b(L_p - 4.2)}. \tag{12}$$

Here, T is the threshold value of the SPL at the given frequency f, and a and b are frequency-dependent parameters; 4.2 phon is identified as the threshold of hearing at 1,000 Hz.

The approximation (12) can be modified in several ways. For example, a simple rational approximation, which requires a minimum amount of arithmetic, is

$$L_p = \frac{a + bL}{1 + cL}, \quad L = \frac{-a/b + (1/b)L_p}{1 - (c/b)L_p}. \qquad (13)$$

For MOSS, we have chosen the expressions (13) over the entire range of frequencies, 20 to 12,500 Hz, with table lookups for the coefficients. We constructed the tables by fitting straight lines through the data in (ISO 1987, Table 3) (L_p as a function of L at the tabulated values of f and using a cubic spline interpolation procedure). In 2003, the International Standards Organization revised these data (ISO 2003). The new data do not affect the algorithm but may affect the values of the coefficients in the expressions (13). Figure 2 shows the equal-SPL curves at 2 dB intervals and the equal-LL contours at 2 phon intervals, both extended to the full frequency range.

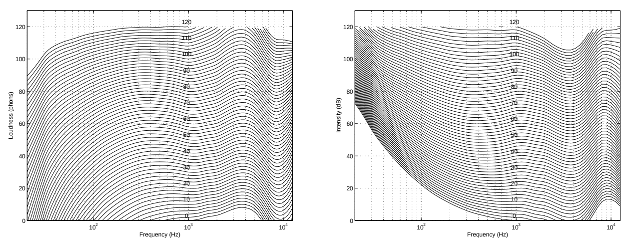

Figure 2: Top: Equal-SPL curves. Bottom: Equal-LL curves. From Eq. (13).

3 Critical Bands

The concept of loudness becomes more complicated when two or more tones are superimposed in a sound, because the way in which their individual loudnesses combine depends on the closeness of their frequencies.

The loudness level of a sound composed of partials within a *critical band* is computed from the total intensity I (Rossing 1982, Section 6.7),

$$I = \sum_i I_i. \qquad (14)$$

Here, I_i is the intensity of the ith partial within the critical band. The total intensity I is associated with the center frequency of the critical band, and the LL for the critical band in phones, L_p, is found from one of the approximations given earlier.

The width of a critical band varies with frequency but is independent of intensity (Zwicker, Flottorp, and Stevens 1957). A table of critical bandwidths was given by Zwicker et al. in (Zwicker, Flottorp, and Stevens 1957, Table I), and later used by Zwicker (Zwicker 1961) to propose a partition of the frequency range from 20 to 15,500 Hz into 24 critical bands. An analytical expression to fit the bandwidth to the data of (Zwicker 1961) was given by Zwicker and Terhardt (Zwicker and Terhardt 1980).

In MOSS, the inverse problem is solved: Given the composition of a sound that consists of partials belonging to a single critical band, the loudness level of the composite sound, and the *relative* loudness levels of the partials, find the amplitudes of the constituent partials.

Consider a sound consisting of n partials, where the frequency of the ith partial is f_i. The target loudness of the sound is specified in sones, L_s, and the relative contribution to the loudness of the ith partial is specified as a fraction of some as yet unknown parameter $L_{s,1}$ (for example, the loudness of the first partial),

$$L_{s,i} = \gamma_i L_{s,1}, \quad i = 1, \ldots, n. \qquad (15)$$

We assume that loudness follows the general power law (10),

$$L_{s,i} = k_i (I_i/I_0)^{0.3b_i}, \quad i = 1, \ldots, n, \qquad (16)$$

where, k_i and b_i are the values of k and b at the frequency f_i. We invert this relation and use Eq. (15) to find an expression for the quantity I_i,

$$I_i = I_0 \left(\gamma_i L_{s,1}/k_i\right)^{1/0.3b_i}, \quad i = 1, \ldots, n. \qquad (17)$$

The sum of the intensities I_i gives the total intensity I of the critical band,

$$I = I_0 \sum_{i=1}^{n} \left(\gamma_i L_{s,1}/k_i\right)^{1/0.3b_i}. \qquad (18)$$

This total intensity must yield the desired loudness L_s at the center frequency f_{CB} of the critical band. Therefore, if k_{CB} and b_{CB} are the values of k and b, at f_{CB}, we also have the relation

$$L_s = k_{\mathrm{CB}} (I/I_0)^{0.3b_{\mathrm{CB}}}. \qquad (19)$$

Substituting the expression (18), we obtain the identity

$$L_s = k_{\mathrm{CB}} \left(\sum_{i=1}^{n} \left(\frac{\gamma_i L_{s,1}}{k_i} \right)^{1/0.3b_i} \right)^{0.3b_{\mathrm{CB}}}. \qquad (20)$$

This equation defines $L_{s,1}$. Once $L_{s,1}$ is known, we find the loudness in sones of each partial from Eq. (15). From the loudness $L_{s,i}$ and the frequency f_i we compute the size A_i in the usual manner.

A considerable simplification occurs if we assume that k and b are constant over the critical band, $b_i = b_{\mathrm{CB}} = b$ and $k_i = k_{\mathrm{CB}} = k$ for $i = 1, \ldots, n$. Then

$$L_s = \gamma L_{s,1}, \qquad (21)$$

where γ is a weighted mean of the fractions γ_i,

$$\gamma = \left(\sum_{i=1}^{n} \gamma_i^{1/0.3b} \right)^{0.3b}. \qquad (22)$$

Combining with Eq. (15), we thus find the loudness in sones of each partial,

$$L_{s,i} = (\gamma_i/\gamma)L_s, \quad i = 1, \ldots, n. \tag{23}$$

4 Beyond the Critical Band

If the range of frequencies in a sound exceeds a critical band, the resulting loudness is more than that of the critical band, but less than the sum of the loudness contributions from adjacent critical bands. If the frequencies are densely distributed, the sound is more like white noise, and Stevens's formula (Stevens 1956) applies,

$$L_s = L_m + F \sum_{i \neq m} L_{s,i}, \quad L_m = \max_i L_{s,i}. \tag{24}$$

Here, it is assumed that the frequency spectrum is partitioned into bands of equal length (on the logarithmic scale), $L_{s,i}$ is the LL in sones of the ith band, and the sum extends over all bands except the loudest. The factor F increases with the bandwidth; $F = 0.13$ for third-octave bands, $F = 0.18$ for half-octave bands, and $F = 0.3$ for octave bands. Rossing (Rossing 1982, Section 6.7), assuming a partition in octave bands, gives Eq. (24) with $F = 0.3$. If the frequency spectrum is partitioned in a nonuniform way, Eq. (24) must be modified,

$$L_s = L_{s,m} + \sum_{i \neq m} F(\Delta_i) L_{s,i}. \tag{25}$$

The loudness routine in MOSS implements an algorithm of the inverse problem: Given the composition of a sound with partials whose frequencies are spread over one or more critical bands, find the amplitudes of the individual partials that collectively give the sound its specified (perceived) loudness.

The algorithm requires a partition of the frequency range. The choice of this partition is to some degree arbitrary. In the current version of MOSS, the entire frequency range is partitioned once and for all into nonoverlapping bands. The bands can be of fixed length (octave, half-octave, or third-octave bands) or of variable length (the partition suggested by Zwicker et al. (Zwicker, Flottorp, and Stevens 1957) or the more generously rounded partition suggested by Zwicker (Zwicker 1961).) Other methods, which take the actual distribution of frequencies into account, are under consideration.

Once the frequency range has been partitioned, the next problem is how to find the size of each partial so the complex sound has the desired loudness. This problem can be solved if we assume that the partition is such that k and b in Eq. (10) are constant within each band, so the loudness of each band can be computed by means of the simplified formulas (22) and (23).

Consider a sound composed of partials distributed over n bands. Assume that the ith band has n_i partials. Let the target loudness of the complex sound be given in sones, L_s, and let the loudness of the jth partial in the ith band be specified as a fraction γ_{ij} of, for example, the loudness of the very first partial,

$$L_{s,ij} = \gamma_{ij} L_{s,11}, \quad j = 1, \ldots, n_i, \; i = 1, \ldots, n. \tag{26}$$

Let b_i be the value of the exponent in Eq. (10) within the ith band. Then the loudness in sones of the jth partial in the ith band is given by

$$L_{s,ij} = \frac{\gamma_{ij}}{\gamma_m + \sum_{i \neq m} F(\Delta_i)\gamma_i} L_s. \tag{27}$$

Here, γ_i is a weighted average of the γ_{ij} over the ith band,

$$\gamma_i = \left(\sum_{j=1}^{n_i} \gamma_{ij}^{1/0.3b_i} \right)^{0.3b_i}, \quad i = 1, \ldots, n. \tag{28}$$

From the loudness $L_{s,ij}$ and the frequency f_{ij} we compute the amplitude A_{ij} in the usual manner.

References

Churcher, B. G. and A. J. King (1937). *Journal of the Institute of Electrical Engineers 1937*, 57–90.

Fletcher, H. and W. A. Munson (1933). Loudness, its defi nition, measurement and calculation. *Journal of the Acoustical Society of America 5*, 82–108.

ISO (1987). Acoustics – normal equal-loudness level contours. In *Standard ISO E*. International Standards Organization.

ISO (2003). *Acoustics – Normal equal-loudness level contours*. International Standards Organization.

Jeans, S. J. (1937 and 1968). *Science & Math*. New York: Cambridge University Press and Dover Publishing.

Robinson, D. W. and R. S. Dadson (1956). A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics 7*, 166–181.

Roederer, J. G. (1995). *The Physics and Psychophysics of Music* (3 ed.). New York: Springer-Verlag.

Rossing, T. D. (1982). *The Science of Sound*. Addison-Wesley Publishing Company.

Stevens, S. S. (1955). Measurement of loudness. *Journal of the Acoustical Society of America 27*, 815–829.

Stevens, S. S. (1956). Calculation of the loudness of complex tones. *Journal of the Acoustical Society of America 28*, 807–832.

Stevens, S. S. (1970). Neural events and the psychophysical law. *Science 170*, 1043–1050.

Zwicker, E. (1961). Subdivisions of the audible frequency range into critical bands. *Journal of the Acoustical Society of America 33*, 248.

Zwicker, E., G. Flottorp, and S. S. Stevens (1957). Critical band width in loudness summation. *Journal of the Acoustical Society of America 29*, 548–557.

Zwicker, E. and E. Terhardt (1980). Analytical expression for critical-band and critical band-width as a function of frequency. *Journal of the Acoustical Society of America 68*, 1523–1525.