A REVISED METHOD OF CALCULATING AUDITORY EXCIATION PATTERNS AND LOUDNESS FOR TIME-VARYING SOUNDS

Zhangli Chen and Guangshu Hu

Department of Biomedical Engineering, Medical School, Tsinghua University, Beijing, China, 100084
E-mail: chenzl03@gmail.com

ABSTRACT

Previously we described a method of calculating auditory excitation patterns and loudness for steady sounds, based on a nonlinear filterbank. Here the method is extended to deal with time-varying sounds. Firstly, the input waveform is transformed to short-term spectrum by a structure with six FFTs, which using longer signal segments for low frequencies and shorter segments for higher frequencies. Secondly, the excitation patterns are calculated from the short-term spectrum, and the summation of the excitation gives a value for the instantaneous loudness. Thirdly, the short-term 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 long-term loudness is calculated from the short-term loudness using a similar averaging mechanism, but with longer attack and release time. The method gives good predictions of loudness for both steady sounds and time-varying sounds.

Index Terms— Auditory excitation patterns; loudness; steady sounds; time-varying sounds.

1. INTRODUCTION

In an earlier paper [1] we developed a method of calculating auditory excitation patterns and loudness for steady sounds. The method is based on a nonlinear filterbank in which each filter is the sum of a broad passive filter and a sharp active filter. All filters have a rounded-exponential (ROEX) shape. For each center frequency (CF), the gain of the active filter is controlled by the output of the passive filter. The parameters of the model were derived from large sets of notched-noise masking data obtained from human subjects. Excitation patterns derived using the new filterbank include compression effects comparable to those observed on the basilar membrane. Loudness can be calculated directly from the area under the excitation pattern when plotted in intensity-like units on an ERB_N-number scale; no transformation from excitation to *specific loudness* [2] is required. The new method predicts loudness as a function of level (loudness growth function), the standard equal-loudness contours and loudness as a function of bandwidth (spectrum loudness summation) with good accuracy.

However, the above method had two limitations. First it was only applicable to steady sounds. Many everyday sounds, like speech and music, are time-varying. For these sounds, there are two aspects of loudness impression: one is short-term loudness, for example, the loudness of a specific syllable; the other is long-term loudness, such as the overall loudness of a sentence. Second it used spectrum of the sound as input. This meant that spectra had to be calculated from time waveforms of the sounds. If a simple FFT transform is used, the temporal resolution and frequency resolution will not be both optimal.

In order to overcome these two limitations, the method described here is extended from the previous method [1] with following procedure. Firstly the input waveform is transformed to short-term spectrum by a structure with six FFTs, which using longer signal segments for low frequencies and shorter segments for higher frequencies, to give adequate spectral resolution at low frequencies combined with adequate temporal resolution at high frequencies. Then the excitation patterns are calculated from the short-term spectrum and summed to give a value for the instantaneous loudness. Finally short-term loudness can be calculated from the instantaneous loudness using an averaging mechanism similar to an automatic gain control system, with attack and release times; and long-term loudness can be calculated from the short-term loudness using a similar averaging mechanism, but with longer attack and release time.

A similar procedure was reported by Glasberg and Moore [3]. However, the way of calculating excitation patterns and loudness in this paper is largely different from that described by Glasberg and Moore; and the parameters used in the FFT structure and averaging mechanism are also specified here. We hope the previous method in [1] can be extended to account for both steady sounds and time-varying sounds. The method can also be easily extended to account for cochlear hearing loss, as described in [4].

2. DETAILS OF THE METHOD

Fig. 1 shows a block diagram of the method. A sample rate of 32 kHz is assumed.

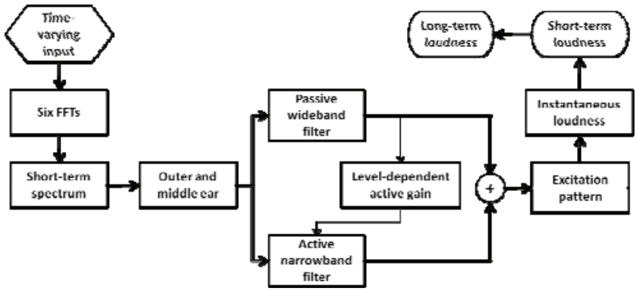

Fig. 1. Block diagram of the model.

2.1. Six-FFTs parallel structure

The cochlea can be characterized as containing a bank of bandpass filters whose center frequencies (*CFs*) span the range from about 40 to 16,000 Hz; the bandwidths of the filters increase with increasing *CF* [1]. To obtain spectral resolution at low *CFs* comparable to that of the auditory system, the analysis of relatively long segments of the input signal is required. However, for high *CFs*, the use of such long segments would limit temporal resolution in a way that does not occur in the auditory system. To give adequate temporal resolution at high *CFs*, analysis using much shorter signal segments is required.

To deal with these problems, a six-FFTs parallel structure is used. The structure is based on Hanning-windowed segments with duration of 4, 8, 16, 32, 64 and 128 ms. Note that these values are different from those in [3] (as 2, 4, 8, 16, 32 and 64 ms). Because we found that if using values in [3], for example, when input is a 1-kHz tone, the calculated spectrum bandwidth is larger than critical-band, which will cause the calculated loudness larger than the actual. Each FFT is used to calculate spectral magnitudes over a specific frequency range, as shown in Table 1.The windowed segments are zero padded and all FFTs are based on 2048 sample points. All FFs are updated at 1-ms intervals.

Table 1. Frequency regions of the FFTs and window.

F_{low} (Hz)	F_{high} (Hz)	Window size (ms)
20	80	128
80	500	64
500	1250	32
1250	2540	16
2540	4050	8
4050	16000	4

2.2. Excitation patterns and instantaneous loudness

The magnitude of the output of each filter in response to a given sound, plotted as a function of *CF*, is called excitation pattern. For each *CF*, the passive wideband filter $W_{pf}(CF,f)$ can be described as:

$$\begin{cases} W_{pf}(CF, f \le CF) = \left(1 + \frac{CF - f}{CF} t_l(CF)\right)\exp\left(-\frac{CF - f}{CF} t_l(CF)\right) \\ W_{pf}(CF, f > CF) = \left(1 + \frac{f - CF}{CF} t_u(CF)\right)\exp\left(-\frac{f - CF}{CF} t_u(CF)\right) \end{cases} \quad (1)$$

where $t_l(CF)$ and $t_u(CF)$ are parameters describing the shape of the filter for frequencies below and above *CF*, respectively. For a short-term spectrum input $X(T,f)$ at time interval of *T*, the output of the passive filter is given by:

$$E_{pf}(T,CF) = \int X(T,f) \cdot W_{pf}(CF,f)\,df \quad (2)$$

The excitation at the output of the passive filter $E_{pf}(T,CF)$ is then transformed to decibels, giving $E_{pf}dB(T,CF)$. This output is used to control the gain of the active filter, according to the following equations:

$$\begin{cases} GdB(T,CF) = GdBm(CF)\begin{Bmatrix} 1 - \frac{1}{1+\exp\left[-0.05\left[E_{pf}dB(T,CF) - \left(100\text{-}GdBm(CF)\right)\right]\right]} \\ + \frac{1}{1+\exp\left[0.05\left(100\text{-}GdBm(CF)\right)\right]} \end{Bmatrix} \\ When\ E_{pf}dB(T,CF) > 30, \quad GdB(T,CF) = GdB(T,CF) - 0.003\left[E_{pf}dB(T,CF) - 30\right]^2 \end{cases} \quad (3)$$

where $GdBm(CF)$ is the maximum gain in dB at CF.

The active narrowband filter $W_{af}(CF,f)$ can be described as:

$$\begin{cases} W_{af}(T,CF, f \le CF) = G(T,CF)\left(1 + \frac{CF - f}{CF} p_l(CF)\right)\exp\left(-\frac{CF - f}{CF} p_l(CF)\right) \\ W_{af}(T,CF, f > CF) = G(T,CF)\left(1 + \frac{f - CF}{CF} p_u(CF)\right)\exp\left(-\frac{f - CF}{CF} p_u(CF)\right) \end{cases} \quad (4)$$

where $G(T,CF)$ is the linear gain corresponding to $GdB(T,CF)$, and $p_l(CF)$ and $p_u(CF)$ are parameters describing the shape of the filter for frequencies below and above *CF*, respectively. For an input with power spectrum $X(T,f)$, the output of the active filter is given by:

$$E_{af}(T,CF) = \int X(T,f) \cdot W_{af}(T,CF,f)\,df \quad (5)$$

Summation of the outputs of the passive and active filters gives the overall excitation:

$$E(T,CF) = E_{pf}(T,CF) + E_{af}(T,CF) \quad (6)$$

The value of $E(T,CF)$ as a function of *CF* is the excitation pattern at time interval of *T*.

It is assumed that instantaneous loudness (*IL*) is directly proportional to the area under the excitation pattern when plotted on a Cam (the units of the ERB_N-number) scale:

$$IL(T) = C \cdot \sum_{Cam(CF)=1.5}^{40.2} E(T,CF) \quad (7)$$

where *C* is a constant. Details of the derivation of the parameters that characterize the model are given in [1].

2.3. Short-term and long-term loudness

The short-term loudness (*STL*) is calculated from *IL* using a form of temporal integration like the automatic gain control:

$$\begin{cases} if\ IL(T) > IL(T-T_0) \\ \quad STL(T) = a_1 IL(T) + (1-a_1) STL(T-T_0) \\ else \\ \quad STL(T) = b_1 IL(T) + (1-b_1) STL(T-T_0) \end{cases} \quad (8)$$

Where a_1 and b_1 are constants related to attack time (*AT*) and release time (*RT*) respectively:

$$\begin{cases} a_1 = 1 - \exp\left(-\frac{T_0}{AT_1}\right) \\ b_1 = 1 - \exp\left(-\frac{T_0}{RT_1}\right) \end{cases} \quad (9)$$

Where T_0 is the time interval of *IL* (1 ms in this paper).The value a_1=0.06 and b_1=0.03, corresponding to *AT*=16 ms and *RT*=32 ms, respectively. These values are chosen to give reasonable predictions of temporal integration of loudness for single tone bursts, as shown in section 3.1.

The long-term loudness (*LTL*) is calculated from the *STL* using a similar strategy as equations (8) and (9), but with longer *AT* (100 ms) and *RT* (2000 ms). The *LTL* is assumed to correspond to the loudness of amplitude-modulated sounds, as shown in section 3.2.

3. PREDICTIONS FOR TIME-VARYING SOUNDS

For time-varying sounds, we assume the loudness of a brief sound (like a syllable) is determined by the maximum value of short-term loudness; the loudness of a modulated sound is determined by the mean value of long-term loudness.

3.1. Loudness as a function of duration

Data in this topic show considerable variability across studies, which reflect the great difficulty for comparing loudness of sounds with different durations. However, it is generally agreed that, loudness of a fixed-intensity sound increases with increasing duration up to 100-200 ms; and increases by 3 phons per doubling of duration when duration is less than 100 ms.

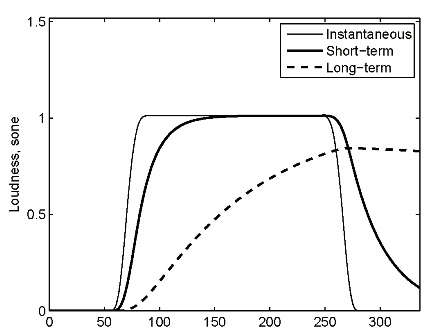

Fig. 2. Response to a 1-kHz tone burst.

Fig. 2 shows the instantaneous loudness, the short-term loudness and long-term loudness of a 200-ms 1-kHz tone burst with a level of 40 dB SPL, ramped with 10 ms. The asymptote of short-term loudness is 1 sone, consistent with the definition of loudness unit.

Solid line in Fig. 3 shows the predictions for a 1-kHz tone with durations of 5, 10, 20, 40, 80, 160, 320 and 640 ms. The predictions are consistent with the conclusions of previous results.

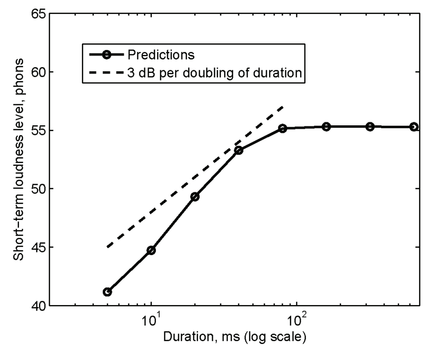

Fig. 3. Loudness as a function of duration for 1-kHz tone.

3.2. Loudness as a function of amplitude-modulation

There is some controversy in previous studies on the loudness of amplitude-modulated sounds. However, three points are basically agreed across study: (1) for modulated-rate smaller than 10 Hz, loudness corresponds to a level between the RMS level and peak level; (2) for mid-rate modulation, loudness corresponds to a level close to the RMS level; (3) for modulated-rate higher than half of the critical-band, loudness increased as the components were resolved by the auditory system.

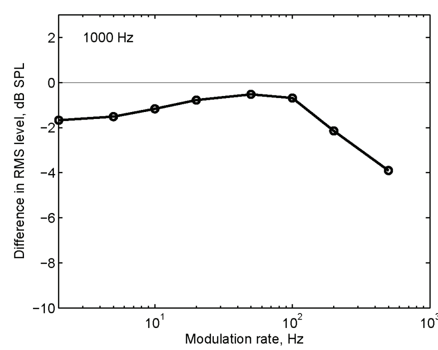

Fig. 4. Loudness as a function of amplitude-modulation

Fig. 4.shows the difference in RMS level required for predicted equal loudness of modulated and unmodulated 1-kHz tones, plotted as a function of modulation rate. The predicted results are broadly consistent with the empirical data.

4. PREDICTIONS FOR STEADY SOUNDS

The revised method should be also able to predict loudness for steady sounds. Here we used signals of relatively long duration (typically greater than 2000 ms) to determine long-term loudness as the loudness of steady sounds.

4.1. Loudness as a function of level

Solid line in Fig. 5 shows predicted loudness in sones (log scale) for binaural listening as a function of the input level of a 1-kHz tone presented in free field with frontal incidence. Dashed line shows comparable predictions of the model in ANSI S3.4-2007 [5]. The correspondence is good.

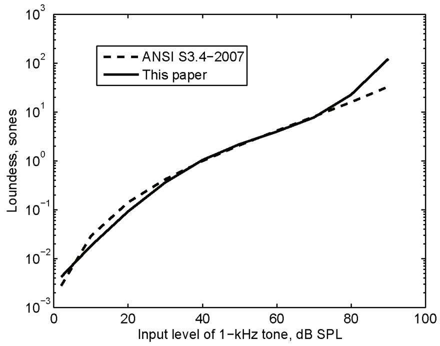

Fig. 5. Loudness as a function of level of a 1-kHz tone.

4.2. Absolute threshold and equal-loudness contours

Fig. 6 shows Comparison of equal-loudness contours from ISO 226-2003 [6] (dashed curves) and predictions of the model (solid curves). The lowest curve is the absolute threshold curve for binaural listening. The predictions are very good.

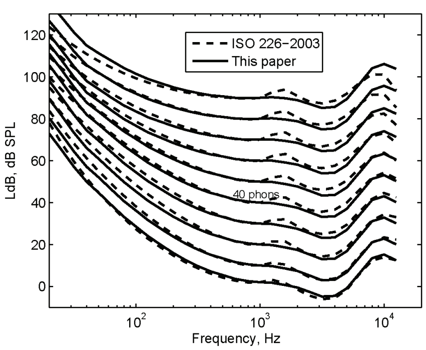

Fig. 6. Absolute threshold and equal-loudness contours.

4.3. Loudness as a function of bandwidth

Fig. 7 shows the level of a 210-Hz wide noise required to match the loudness of a noise of variable bandwidth, plotted as a function of the variable bandwidth. The bands of noise were geometrically centered at 1420 Hz and had an overall level of 30, 50 or 80 dB SPL. The circles show the experimental data [7] and the solid lines show predictions of the model.

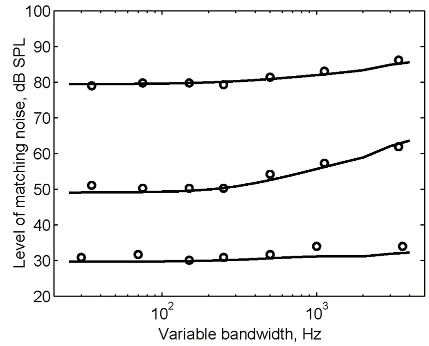

Fig. 7. Loudness matching between a fixed-bandwidth noise and a variable-bandwidth noise

5. CONCLUSIONS

A method of calculating excitation patterns and loudness for steady sounds is extended to account for time-varying sounds. The revised method uses a six-FFTs structure to transform waveform to spectrum in an interval of 1 ms; then it calculates excitation patterns and instantaneous loudness from the short-time spectrum; with a averaging mechanism like automatic gain control, short-term and long-term loudness can be calculated. In general, predicted results by the method are comparable to published data in literature.

6. ACKNOWLEDGMENTS

The work was supported by the National Natural Science Foundation of China and Tsinghua-Yu-Yuan Medical Sciences Fund.

7. REFERENCES

[1] Z. Chen, G. Hu, B. R. Glasberg, and B. C. J. Moore, “A new method of calculating auditory excitation patterns and loudness for steady sounds,” Hearing Research, in press, 2011.

[2] B. C. J. Moore, B. R. Glasberg, and T. Baer, “A model for the prediction of thresholds, loudness, and partial loudness," Journal of Audio Engineering Society, vol. 45(4), pp. 224-240, 1997.

[3] B. R. Glasberg and B. C. J. Moore, “A model of loudness applicable to time-varying sounds," Journal of Audio Engineering Society, vol. 50, no. 5, pp. 331-342, 2002.

[4] Z. Chen, G. Hu, B. R. Glasberg, and B. C. J. Moore, “A new model for calculating auditory excitation patterns and loudness for cases of cochlear hearing loss,” Hearing Research, accepted, 2011.

[5] ANSI S3.4, “Procedure for the Computation of Loudness of Steady Sounds,” American National Standards Institute, New York, 2007.

[6] ISO 226, “Acoustics - Normal Equal-Loudness Contours,” International Organization for Standardization, Geneva, 2003.

[7] E. Zwicker, G. Flottorp, and S. S. Stevens, “Critical bandwidth in loudness summation,” Journal of Acoustical Society of America, vol. 29, pp. 548-557, 1957.