인간 청각 모델의 설계 및 성능 평가

류승완, 김수궐, 박정렬, 신재호

동국대학교 전자공학과

Human Auditory Model Design and Quality Assessment

Seung-Wan Ryu, Su-Kweor Kim, Jeong-Yeol Park, Jaeho Shin

Department of Electronic Engineering, Dongguk University
E-mail : standing@dongguk.edu

Abstract

Objective quality measurement schemes that incorporate properties of the human auditory system. The basilar membrane (BM) acts as a spectrum analyzer, spatially decomposing the signal into frequency components. Filterbanks were used to complementing the linearity of BM. Each filterbank is an implementation of the Equivalent Rectangular Bandwidth (ERB), gammachirp function. This filterbank is level-dependent asymmetric compensation filters. And for the validation of the auditory model, we calculate the calculated perceived difference(CPD).

1. Introduction

Recently coding technique of audio signal has significantly developed. Especially the perceptual coding technique which is used human auditory properties, so the most of audio coder has achieved high compression ratio and high quality. Perceptual coding technique has psychoacoustic model applying to the human auditory masking effect. There are many audio coding technique but we have to evaluate the technique in terms of complexity, sound quality etc. In aspect of sound quality of the coding result is usually measured by means of subjective tests(i.e. SG, SDG). But such tests are expensive, time-consumed and inconvenient, so objective measurement method is needed that will model the sensory and cognitive process underlying subjective ratings[1]. The masking effect which is one of human auditory properties in psychoacoustic model play a important role to reduce bit-rate but it generate the noise which human is not disturbed. Therefore, a simple SNR(Signal to Noise Ratio) measurement is not at all sufficient to assess the audio quality. So we need advanced and precise technique to measure the sound quality[2].

2. The Human Auditory System

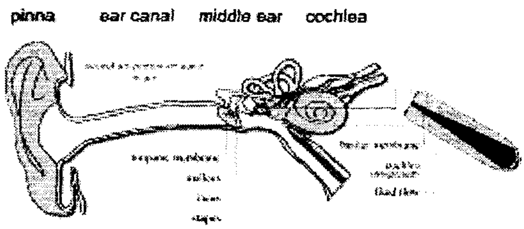

Figure 1. The main components of the human auditory system

Incoming sounds are funneled into the ear canal by the pinna. This is the flap of cartilage and skin found on each side of the head. The frequency responses of the sound reaching the ear canal (via the pinna), from sources at various angles around a listener are called Head Related Transfer Functions (HRTFs).

Ear Canal is the resonant cavity between the outer and middle ear. It has a resonance at around 3-5 kHz, hence it attenuates higher and lower frequencies.

The middle ear consists of the tympanic membrane (ear drum), malleus (hammer) and stapes(stirrup). This transmits the sound pressure wave from the ear canal into the cochlea.

The fluid-filled cochlea is a coil within the ear, partially protected by bone. The cochlea is semi-partitioned along its length by a thin flap called the basilar membrane.

The basilar membrane (BM) vibrates with the incoming sound, and acts as a spectrum analyzer, spatially decomposing the signal into frequency components. The movement of the BM is in the form of a traveling wave. The amplitude envelope of this traveling wave varies along the BM. The resonant frequency of a point on the BM varies with stiffness, thus sounds of different frequencies cause different parts of the BM to vibrate.

The inner hair cells on the BM fire when the BM movies upwards, so transducing the sound wave at each point into a signal on the auditory nerve.

There are approximately 12000 outer hair cells distributed along the length of the BM. They react to feedback from the brainstem, altering their length to change the resonant properties of the BM. This causes the frequency response of the BM to be amplitude dependent.

3. Gammachirp Filter

The complex impulse response of the gammachirp is

$$g_c(t) = at^{n-1}\exp(-2\pi bERB(f_r)t)\exp(j2\pi f_r t + jc\ln t + j\phi) \tag{1}$$

where time t > 0

- a: amplitude, f_r : the asymptotic frequency
- n, b : parameters defining the envelope of the gamma distribution
- c: a parameter for the frequency modulation or the chirp rate
- ϕ: the initial phase, $\ln t$: a natural logarithm of time
- $ERB(f_r)$: equivalent rectangular bandwidth of an auditory filter at f_r

The Fourier transform of the gammachirp is

$$G_C(f) = \frac{a\Gamma(n+jc)e^{j\phi}}{2\pi bERB(f_r) - j2\pi(f-f_r)^{n+jc}}$$

$$= \bar{a}\left[\frac{1}{\{2\pi\sqrt{\bar{b}^2+(f-f_r)^2}\}^n}e^{-jn\theta}\right]\cdot\left[e^{c\theta}\cdot e^{-jc\ln\{2\pi\sqrt{\bar{b}^2+(f-f_r)^2}\}^2}\right] \tag{2}$$

$$\theta = \arctan\frac{f-f_r}{\bar{b}}$$

$\bar{a} = a\Gamma(n+jc)e^{j\phi}$ and $\bar{b} = bERB(f_r)$

The first term $\bar{a}$ is a constant. The second term is known as the Fourier spectrum of the gammatone,. The third term represents an asymmetric function.

When the amplitude is normalized($\bar{a}=1$), the frequency response of the gammachirp is

$$G_c(f;n,b,c,f_c) = G_T(f;n,b,f_r)\cdot H_A(f;b,c,f_r) \tag{3}$$

A gammachirp filter can be implemented by cascading a gammatone filter and an asymmetric filter[3], [4].

4. Auditory Model

4.1 Outer-Middle Ear Filtering

Above 1 ㎑, the magnitude response of the outer-middle ear filter is similar to the inverted absolute threshold curve. This is base on the assumption that the inner ear is equally sensitive to all frequencies above 1 ㎑. The absolute threshold curve is probably influenced by the low frequency internal noise of the inner ear and therefore does not reflect the transmission through the outer and middle ear below 1 ㎑ is reflected shape of an equal-loudness contour at a high loudness level.

The transfer function of the outer-middle ear filter consists of a cascade of a recursive lowpass filter of order 8 and a parameterized recursive highpass filter of order2.

$$H_{OM}(z) = H_{LP}(z) * H_{HP}(z) \tag{4}$$

$$H_{LP}(z) = \frac{0.109z^7(z+1)}{z^8+a_1z^7+a_2z^6+a_3z^5+a_4z^4+a_5z^3+a_6z^2+a_7z^1+a_8} \tag{5}$$

$$\begin{pmatrix} a_1 = 2.5359, a_2 = 3.9295, a_3 = 4.7532, a_4 = 4.7251 \\ a_5 = 3.5548, a_6 = 2.1396, a_7 = 0.9879, a_8 = 0.2836 \end{pmatrix}$$

$$H_{HP}(z) = \frac{z^2-2z+1}{z^2-2Rz+R^2} \tag{6}$$

We are using an outer-middle ear filter with R=0.989 and we assume the inverted absolute threshold contour with

R=0.957 at the moment. The random firing of the inner hair cells, combined with the blood flow within the ear, gives rise to an internal noise that limits our absolute hearing threshold[5].

4.2 Basilar Membrane Filtering

The topic of basilar membrane(BM) filterbanks are nonuniform bandwidth bandpass filterbanks designed to imitate the frequency resolution of human hearing. BM filterbanks have also been based more directly on psychoacoustical measurements and more recently the gammatone filterbank. The gammachirp filterbank further adds a level-dependent asymmetric correction to the basic gammatone channel frequency response, thereby providing a more accurate approximation to the auditory frequency response.

Moore and Glasberg suggested that the excitation pattern of a given sound could be thought of as the output of the BM filter as a function of their center frequency. The BM filter shapes for five center frequencies. Each filter is symmetrical on the linear frequency scale used, but the bandwidths of the filters increase with increasing center frequency.

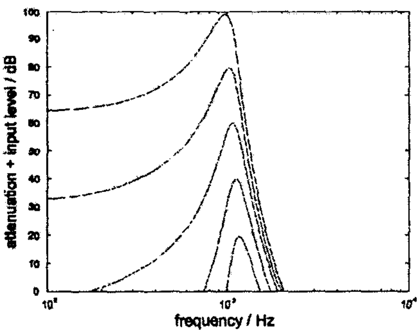

Figure 2. BM filtering via amplitude dependent filterbank

Moore and Glasberg examined the available data and concluded that the input level to the filter is the primary factor controlling its shape. The BM filter shapes for each center frequency and masker level were calculated according to the assumptions outlined above. The high frequency side of each excitation pattern is determined by the low frequency branches of the BM filters and vice versa.

A bank of amplitude dependent filters simulates the response of the BM. Each filter is an IIR implementation of the gammachirp, and simulates the response of the BM at a given point. The filters are linearly spaced on the Bark frequency scale, which itself accurately describes the spacing of resonant frequencies along the BM.

As the BM changes its resonant properties in response to the amplitude of the incoming signal, so the shape of each gammachirp filter is dependent on the signal amplitude at the output of the filter. The envelope of the signal is derived by rectification, and low pass filtering. This envelope is then used to adjust the filter coefficients.

When the sound level is sufficiently high, the cochlear filter has a broad bandwidth and behaves like a passive and linear filter. The filter gain increases and the bandwidth becomes narrower because of the active processes.

4.3 Outer Hair Cell

The output of each gammachirp filter is processed to provide the amplitude information necessary to tune the response of that individual filter. This feedback causes the Q of each filter to be proportional to the amplitude of the signal passing through it. This corresponds to the tuning mechanism mediated by the outer hair cells.

The process is as follows:

1. Rectification
2. Peak detection and holding
3. Low pass filtering

The value of c is

$$c(t) = 3.29 - 0.059 \times 20\log_{10}(env(t) + ina) \tag{7}$$

$env(t)$ is the low-pass filtered envelope, and ina is present to prevent the calculation of $\log_{10}(0)$ during a silent input signal. The value of ina is below the minimum audible threshold, and has a negligible effect on supra-absolute-threshold calculations.

4.4 Inner Hair Cell

The hair cell contains a quantity of the "free transmitter", $q(t)$, which leaks through a permeable membrane into the synaptic cleft. The permeability, $k(t)$, fluctuates as a function of the instantaneous amplitude of the acoustic stimulation, $s(t)$, which we linearize for our implementation :

$$k(t) = \begin{cases} 0 & for \quad s(t)+a \le 0 \\ s(t)+32 & for \quad 0 < s(t) \le 1348 \\ \frac{s(t)}{8} + \frac{s(t)}{128} + 1200 & for \quad s(t)+a > 1348 \end{cases} \tag{8}$$

5. Simulation Results

One possible method of calculating the perceived difference is to simply take the difference between the two sets of outputs from the auditory model. This is a time-varying calculated perceived difference(CPD). The threshold of perceiving a difference between two signals is known for a wide variety of possible signals, from various psychoacoustic experiments using human subjects. If the difference between two signals is greater than this threshold value, then a human listener will perceive a difference. If it is less, then the difference, though present, will be imperceptible.

Rather than looking at the difference on a sample-by- sample basis, can be summed over time. Then, using average interval.

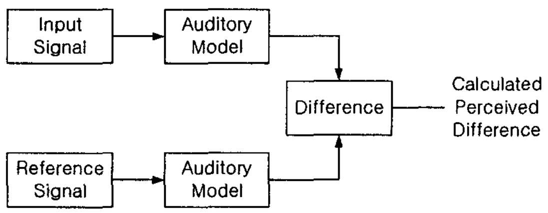

Figure 3. Method for calculating the perceived difference

To simple test how accurately the model, we will use the model to assess the quality of some audio streams generated by an MPEG-1 layer-2 codec. The audio sample is taken from EBU-SQAM CD. Each coded stream was analyzed by the auditory model, and compared with the original.

Table 1. Simulation result.

Bit-Rate	Number of Bands in which CPD is greater than threshold	Human perceive a difference
256	0	None
160	5	Slight
128	29	Much
64	53	More

The results indicate that, in this particular test, the model predicted human perception well. At the high bit rate the auditory model perceives small difference between the original and coded signals. At the lower bit rates, the model predicts that the difference between the original and coded signals will be well.

6. Conclusions

In this paper, we design the auditory model for objective audio quality measurement. The outer middle ear filter consists of a cascade of recursive lowpass filter and a parameterized recursive highpass filter. The BM filter is implemented using a gammachirp function.

The gammachirp function is approximated by the combination of a gammatone function and an IIR asymmetric compensation filter. This implementation reduces the computational cost for time varying filtering because both of the filters can be implemented with a few filter coefficients.

The output of this model is analyzed to detect perceptible differences between reference signal and test signal.

References

[1] C. Colomes, C. Schmidmer, "Perceptual Quality Assessment For digital Audio: PEAQ - The New ITU Standard for Objective Measurement of the Perceived Audio Quality," AES 17th International Conference, pp. 17-33, Sep. 1999.

[2] David J M Robinson & Malcolm J Hawksford, "Time-Domain Auditory Model for The Assessment of High-Quality Coded Audio," 107th convention of AES, Preprint 5017, Sep. 1999.

[3] IRINO Toshio "An Optimal Auditory Filter," NTT Basic Research Laboratories 3-1, Morinosato Wakamiya, Atsugi-shi Kanagawa. Vol 243-01, JAPAN.

[4] Toshio Irino, Masashi Unoki, "An Analysis/Synthesis Auditory Filterbank Based on an IIR Implementation of the Gammachirp" J. Acoust. Soc. Vol 20, pp.397-406, Nov. 1999.

[5] M. Pflueger, R. Hoeldrich, W. Riedler, "A Nonlinear Model of The Peripheral Auditory System,". 103th Convention of the AES, New York, Sep. 1997.