Jennifer Hsu
Music 421
Winter 2012

An overview of auditory filter banks

Introduction

Auditory filter banks attempt to model the frequency resolution of the human auditory system as a filter bank of non-uniform bandpass filters. We can categorize these filter banks into those that are modeled after psychoacoustic and/or physiological data and those that are not modeled off of either of those measures. The classic auditory filter bank that was not directly based on psychoacoustic experiments or physiological measures is the constant-Q filter bank, which we can break down into the third-octave filter bank and the auditory wavelet transform filter bank. Those modeled after psychoacoustic and physiological measures can be categorized into three families: rounded exponential (roex), gammatone, and filter cascades. Through an exploration of the different models, we will see how these auditory filter banks are related to each other and also how each one differs in its strengths and weaknesses.

We will first begin with a brief explanation of how the ear is believed to work and consider the desired characteristics of an auditory filter bank. Then we will examine the different families of proposed auditory filter bank models as well as their strengths and weaknesses.

The Ear

An understanding of the way humans process sound is necessary when designing auditory filter banks. When an acoustic signal enters the ear, it travels through the outer, middle, and inner ear. It triggers a mechanical bone response which affects the fluid of the cochlea and creates traveling waves in the basilar membrane. Each frequency that we can hear triggers a response on the basilar membrane at a specific location. The ear decomposes the time domain signal into a frequency representation on a nonlinear scale. The hair cells along the basilar membrane convert this frequency response into electrical signals that are propagated to the brain [10]. The nonlinear scale has been modeled previously on the ERB and Bark scales [17].
Many of the following filter banks begin as linear models, but have been incorporated with nonlinear measures and parameters including level-dependence and instantaneous nonlinearities [4, 11, 13]. An important feature of the cochlea is its ability to compress a wide range of sound levels to a narrow range of cochlear motion. This property is modeled by level-dependent filters that exhibit the quality where the signal level increases are paired with filter gain decreases [14]. Instantaneous nonlinearities include combination tones such as the sound that emanates from the ear itself [15, 16]. It has also been shown that the filter of the human ear is approximately symmetric for low frequency input but asymmetric for higher frequency input where there is a shallower low frequency skirt and a steeper high frequency skirt [4].

Desired characteristics for auditory filter banks

A review of the literature covering auditory filter banks reveals that each author has his or her own criteria for a good auditory filter bank. When evaluating these models, it is important to keep in mind the motivation for the existence of the model. For some researchers, it is the simplicity of the description in either domain that is the most important characteristic while for others it is perceptual or physiological accuracy that is paramount. In even other models, it is the invertibility of the transform that holds higher importance over perceptual or physiological preciseness. Although the criteria for the filter bank models differ, they seem to all agree on the idea that the nonlinearities of the cochlea play a large role in the way we hear. Additionally, because each model borrows ideas from other models, we will still find similar results in auditory filter banks that were designed with different goals in mind.

Constant Q

The constant Q transform filter bank has been widely used to approximate the human auditory system. Because of the constant difference in both frequency and resolution of the Fourier transform, the Fourier transform cannot be directly used to approximate the human auditory filter bank. The constant Q transform, on the other hand, has a similar calculation to the discrete Fourier transform, but the ratio of center frequency to resolution is kept constant, making it easier to map signals to what we would think of as musical frequencies or notes [1]. In this model, the time and frequency resolutions are nonuniform and bandwidth increases as the frequency increases. A smaller window size is used in the high frequency area which translates to better time resolution in the high frequencies. For that reason, there are usually more samples in the higher frequencies.

Third Octave

The third octave filter bank is also a model that has been widely accepted as an imitation of the human auditory system. In this filter bank, the input signal is decomposed into octaves and these octave bands are further decomposed into third octave sections. The third octave filter bank is an approximation to the critical bands of hearing as it models the idea that signals are perceived to mask one another when they occur within a critical band of one another [8].

Auditory Wavelet Transform

The wavelet transform has also been used to model an auditory filter bank. By choosing the correct analyzing wavelet from the impulse response of the auditory model, this filter bank approximates the human auditory system. It has been claimed that the auditory wavelet transform could be used to both analyze and resynthesize a signal with the help of a Hilbert transform. Because the wavelet transform analyzes a signal on a log-linear frequency scale with constant Q, it has been thought to be a good model for simulating the nonlinear frequency response of the human auditory system [3].

Models motivated from psychoacoustic and physiological data

Auditory filter banks based off of psychoacoustic data include the results seen from tone-masking experiments and notched noise masker experiments. These type of experiments have revealed the desired skirt shape for auditory filter simulations [2,18]. Those that are modeled after physiological data involve the reproduction of measured mechanical responses of the basilar membrane as well as neural responses of the auditory nerve. Models based on psychoacoustic data are not expected to agree with those based on physiological measures as there is much neural processing in between the basilar membrane response to signals and the perception of the signal itself. However, we can still expect the nonlinear cochlear processing to be reflected in both types of models.

Rounded Exponential (roex) Family

The roex family of auditory filter banks is described by the power frequency response. The simplest model of the roex family of auditory filter banks just takes into account one parameter, the shape parameter or the bandwidth. Other related roex models take bandwidth and skirt shape into account. Another model has been proposed to control the low side as well as the high side of the skirt to take asymmetry into account. Several other roex filter have been proposed with as many as six adjustable parameters. These filter banks have been criticized for having too much flexibility. The roex family of auditory filter banks are not easily implemented nor are they used very often. They are best used for describing an auditory filter's magnitude transfer function shape [16].

Gammatone Family

The gammatone family of auditory filter banks includes the gammatone filter, which is modeled after matching the impulse response of the filter to the impulse response of the auditory system in animals, and filters based on the gammatone filter with nonlinearities added. The gammatone filter can be thought of as an asymmetric envelope and is simply a gamma distribution envelope multiplied with a tone in the time-domain. The peak and skirt shape of these filters are better than those of the roex family filters, but these filters are linear, which are undesirable for auditory filter banks [19]. Unfortunately, by nature, the passband of these filters are almost symmetric, which is not desired for an auditory filter bank [6]. Additionally, they are hard to implement based on the Laplace-domain poles and zeros [13, 16].

The compressive gammachirp filter is a generalization of the gammatone filter. It was designed based on the observation that the auditory filter changes shape depending on input signal level in psychoacoustic masking experiments. This auditory filter includes a chirp that is often observed in basilar membrane response. It also meets the goal of being a nonlinear filter bank capable of both analysis and resynthesis [5]. In the frequency domain, the compressive gammachirp filter is more realistic and controllable than the gammatone filter as it exhibits asymmetry. Again, this filter is difficult to implement and approximate due to the lack of a pole-zero decomposition [16]. The all-pole gammatone filter approximates the gammatone filter by removing all the zeros [20]. The one-zero gammatone filter extends the all-pole gammatone filter and adds an extra real zero to give more precise control of the low-frequency tail. It also uses automatic gain control which is a feedback loop that keeps the output level from varying too much. The output level feeds back through parameters so that higher output will lead to lower filter gains. Both of these two filters fix the problems of the gammatone filter and the gammachirp filter by simplifying the Laplace-domain description and implementation while still preserving asymmetry in the frequency domain [6].

Filter Cascade

Filter cascade models are described by poles and zeros so they are inherently more easily implemented than the previous filter banks mentioned. They are modeled as linear wave propagations and nonlinearities are incorporated afterwards. They approximate the magnitude and phase along the basilar membrane at various frequencies [12, 15]. The all-pole filter cascade is similar to the all-pole gammatone filter. It is a tapped filter cascade structure that is efficient with easily computed properties and allows for level-dependent nonlinearity with the variation of a single parameter. Two-pole filters in cascade have been shown to be good enough to approximate the human auditory system [14]. Its drawbacks include a high-side that is not sharp enough and an unrealistically long delay when tuned for frequencies that are required by the auditory system [16].

The pole-zero filter cascade is a cascade of two-pole two-zero second-order filter sections. Each output channel has one of these filter sections. Nonlinearity has been added through an automatic gain control feedback system. The pole-zero filter cascade has been shown to be quite good at matching both psychoacoustic experimentation results and physiological results. It is argued that the pole-zero filter cascade is realistic in both the time domain and frequency domain. It is strong where the all-pole filter cascade is not as it allows for control of the delay, high side steepness, and low frequency tail of the filter [14, 15].

Newer models

There is much literature surrounding the above auditory filter banks, but there also exist two other models that have recently emerged. The first is Li's auditory-based transform in which the goal was to create an invertible transform that is closer to the way the auditory system works than the Fourier transform and the wavelet transform. The auditory-based transform is actually quite similar to the gammatone filter banks, but slightly different because its main aim was the invertibility of the transform and not emulation of the human hearing system [9, 10]. Another recently proposed model is called the synchrony-capture filter bank which is based on an observed property of auditory nerve fibers. In this feature, when multiple harmonics are present, the auditory nerve fibers work in a "winner-takes-all" fashion and lock onto the dominant harmonic only. The model uses a gammatone filter bank and each filter is cascaded with three bandpass filters. The center frequencies of the bandpass filters are adjusted and tuned by feedback [7].

Conclusion

We have covered a wide range of auditory filter banks with the goal of giving a glimpse of the thought process behind the design of a model that emulates human hearing. From the patterns seen through all the previously mentioned filter banks, it seems that future models will continue to extend former models and give a better understanding of the way human hearing works.

Bibliography

[1] J.C. Brown, "Calculation of a constant Q spectral transform," *J. Acoust. Soc. Am.,* vol. 89, no. 1, pp. 425-434, Jan 1991.

[2] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research,* vol. 47, pp. 103-138, Aug. 1990.

[3] T. Irino and H. Kawahara, "Signal Reconstruction from Modified Auditory Wavelet Transform," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3549-3554, Dec. 1993.

[4] T. Irino and R. D. Patterson, "A compressive gammachirp auditory filter for both physiological and psychophysical data," *J. Acoust. Soc. Am.*, vol. 109, no. 5, pp. 2008-2022, May 2001.

[5] T. Irino and R. D. Patterson, "A Dynamic Compressive Gammachirp Auditory Filterbank," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 2222-2232, Nov. 2006.

[6] A. G. Katsiamis, E. M. Drakakis, R. F. Lyon, “Introducing the Differentiated All-Pole and One-Zero Gammatone Filter Responses and their Analog VLSI Log-domain Implementation,” 49th *IEEE International Midwest Symposium on Circuits and Systems*, vol. 1, pp. 561-565, Aug. 2006.

[7] R. Kumaresan, V.K. Peddinti, and P. Cariani, “Synchrony Capture Filterbank (SCFB): An Auditory Periphery Inspired Method for Tracking Sinusoids,” *IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2012.

[8] S. Levine, “Critically Sampled Third Octave Filter Banks,” *ICMC Proceedings*, pp. 301-304, 1996.

[9] Q. Li, “An Auditory-Based Transform for Audio Signal Processing,” *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 2009.

[10] Q. Li, "Auditory-Based Time Frequency Transform," in *Speaker Authentication*, Berlin: Springer, ch. 7, sec. 1-7, pp. 111-134, 2012.

[11] E. A. Lopez-Poveda and R. Meddis, “A human nonlinear cochlear filterbank,” *J. Acoust. Soc. Am.*, vol. 110, no. 6, pp. 3107-3118, Dec. 2001.

[12] R. F. Lyon, “All-pole Models of Auditory Filtering,” *Diversity in Auditory Mechanics*, Lewis et al. (eds.), World Scientific Publishing, Singapore, pp. 205-211, 1997.

[13] R. F. Lyon, “Cascades of two-pole-two-zero asymmetric resonators are good models of peripheral auditory function. *J. Acoust. Soc. Am.*, vol. 130, no. 6, pp. 3893-3904, Dec. 2011.

[14] R. F. Lyon, “Filter Cascades as Analogs of the Cochlea,” *Neuromorphic Systems Engineering: Neural Networks in Silicon,* Springer US, pp. 3-18, 1998.

[15] R. F. Lyon, “Machine Hearing: An Emerging Field [Exploratory DSP],” *Signal Processing Magazine, IEEE*, vol. 27, no. 5, pp. 131-139, Sept. 2010.

[16] R. F. Lyon, A. G. Katsiamis, and E. M. Drakakis, “History and future of auditory filter models,” *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp 3809-3812, May 2010 - June 2010.

[17] B. C. J. Moore and B. R. Glasberg, “Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,” *J. Acoust. Soc. Am.*, vol. 74, no. 3, pp. 750-753, Sept. 1983.

[18] R. D. Patterson, "Auditory filter shapes derived with noise stimuli," *J. Acoust. Soc. Am.*, vol. 59, no. 3, pp. 640-654, Mar. 1976.

[19] R. D. Patterson, M. H. Allerhand and C. Giguere, “Time-domain modeling of peripheral

auditory processing: A modular architecture and a software platform," *J. Acoust. Soc. Am.,* vol. 98, no. 4, pp. 1890-1894, Oct. 1995.

[20] M. Slaney, "An Efficient Implementation of the Patterson-Holdsworth Auditory Filterbank," Apple Computer, Inc., Technical Report #35, 1993.