AUDITORY WAVELET TRANSFORM

Y.Salimpour[1,3], M.D.Abolhassani[1,2] ,H.Soltanian-Zadeh[3,4]

[1] Research Centre for Science and Technology in Medicine, Tehran, Iran
[2] Tehran University of Medical Sciences, Tehran, Iran
[3] Institute for Studies in Theoretical Physics and Mathematics, Tehran, Iran
[4]Tehran University, Tehran, Iran
salimpour@ipm.ir

Abstract: The auditory periphery system recieves a one dimensional acoustical signal that describe how the local pressure varies with time. However, this one dimentional signal information is then somehow unfolded into a two dimentional time-frequency plane that tells us when which frequency occurs. Therefore, hearing process is thus based on some compromise between time localization and frequency localization and a kind of time-frequency or wavelet type transformation is done in auditory signal processing. In this study the similarities between auditory transform based on auditory physiological process and wavelet transform is introduced. Specially band pass filter bank property and variable time and frequency resolution with the signal frequency are considered.

Introduction

In the inner ear or cochlea, sound is detected by an array of several thousand hair cells that convert mechanical vibrations into electrical activities. The cochlea is often thought of as a bank of filters because it performs frequency analysis using a frequency to place mapping along the basilar membrane. That is, each place along the membrane has a characteristic frequency, f_c, for which it is maximally displaced when a pure tone of that frequency is presented as an input. As a filter bank, the cochlea exhibits the following characteristics: (a) Non-uniform filter bandwidths; Frequency resolution is higher at the lower frequencies (near the apical end of the cochlea) than at high frequencies (near the basal end of the cochlea). For an equivalent filter bank representation, this implies narrower filters that are more closely spaced together for low frequencies, and broader filters that are spaced further apart for high frequencies. (b) Asymmetric frequency response of individual filters; for a particular place along the basilar membrane with characteristic frequency f_c, the response to $f_c+\Delta f$ is lower than the response to $f_c-\Delta f$. For a bandpass filter centered at f_c, this can be interpreted as an asymmetric magnitude response, with sharper cutout on the high frequency side. (c) Level-dependent frequency response of individual Filters; as mentioned in the previous section, basilar membrane motion is compressive and non-linear, meaning that doubling the input stimulus intensity does not result in doubling of membrane displacement. From a Filtering perspective, this implies that the peak gain of the filter centered at f_c decreases as the level of the input stimulus increases. Another observation is that the magnitude response becomes broader and more symmetric with increasing sound levels.

The individual hair cells, and the auditory nerve fibers to which they are connected, are tuned to specific frequencies. [1] The population of auditory nerve fibers, thus, provides us with a frequency analysis of sound waveforms in the environment. Each auditory nerve fiber may be considered as a filter that signals information about the temporal structure of stimuli are within its preferred frequency range. As engineers have understood for years, the design of a filter involves an inevitable trade-off between the precision of frequency tuning and temporal tuning. A tone consists of cyclical fluctuations of air pressure, and to obtain an accurate frequency estimate, many cycles must be integrated. But a longer integration period means a decrease in the temporal accuracy of the filter in other words, a filter cannot signal both the frequency and the timing of a sound with arbitrary precision. Yet discriminations of the real world sounds often require accurate measurements of both frequency and timing. Precise temporal information is also important for the sound localization, which in many cases depends on time-of-arrival differences between the two ears. The challenge for the auditory system, then, is to find the right trade-off between timing and frequency analysis [2].

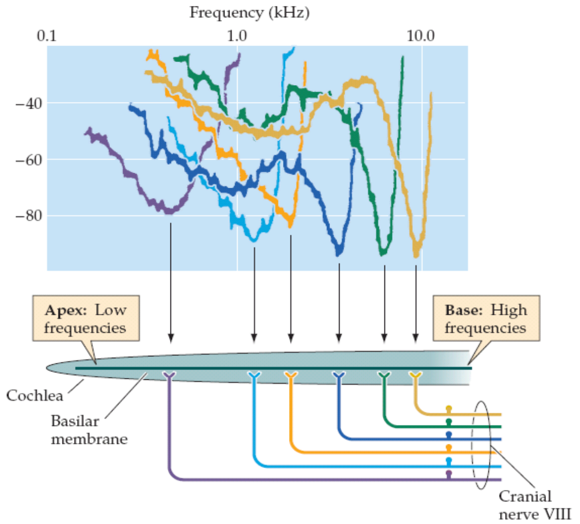


Figure 1: The frequency tuning curves of auditory nerve fibers superimposed and aligned with their approximate relative points of innervations along the basilar membrane

Materials and Methods

The membrane displacement and fluid pressure in the lower chamber are shown schematically in Figure 2. The wave is said to be in the long-wave region when its wavelength is long with respect to the height of the duct. In this region, the fluid particle motion is constrained to be essentially horizontal, like a wall of fluid moving back and forth in a pipe. When the wavelength becomes short with respect to the height of the duct, the wave is said to have entered the short-wave region. At this point, the wave propagates more like ripples on the surface of a deep pond, where the fluid particles trace out elliptical trajectories, with greater amplitude near the surface. Finally, the wave dies out in the highly damped cut-of region. [3]

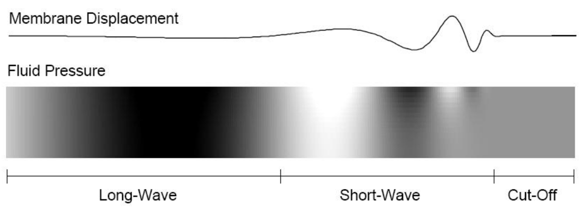
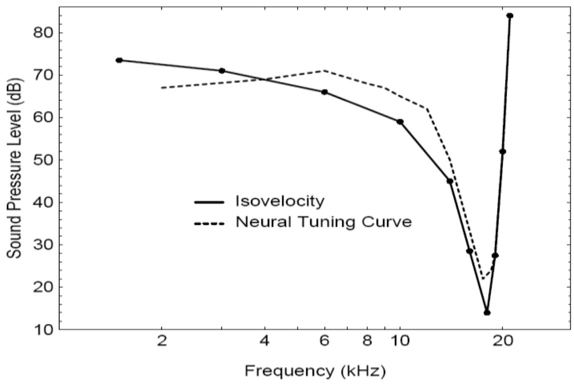


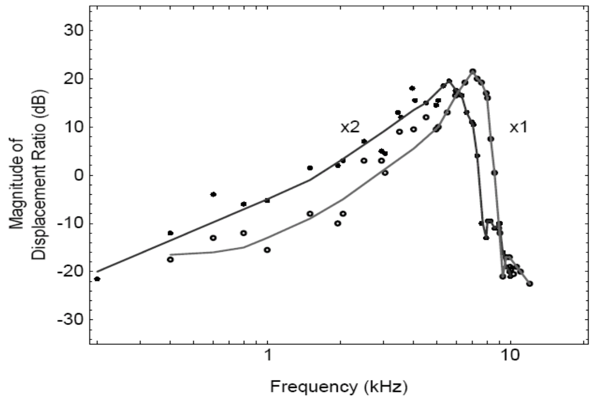
Figure 2. Detail of wave propagation, showing the membrane displacement and fluid pressure along a vertical slice through the lower chamber, for a sinusoidal stapes vibration. The amplitude of the membrane displacement wave is small near the base, reaches a peak at the best place, and dies out quickly in the cut-of region. Deviations in fluid pressure from the resting pressure are shown as dark or light deviations from gray. The amplitude of the fluid pressure wave is large near the base, and gradually decays through the long-wave and short-wave regions, and dies out quickly in the cut-of region. In the short-wave region, the amplitude of the pressure wave decreases approximately exponentially away from the partition.

In Figure 3 (b) the isovelocity curve from a point on the guinea-pig cochlea is compared to neural isoresponse curve from a spiral ganglion cell in the guinea pig. This famous measurement, [5], shows that the sharp tuning of an auditory nerve fiber is determined at the mechanical level of the basilar-membrane vibration. Since the system is nonlinear, these isoresponse tuning curves are not directly comparable to transfer function data, as pointed out by Lyon [7].

Figure 4 shows tuning curves of auditory periphery and the combinations of band pass filter bank .In comparison with wavelet filter bank it could be concluded that both systems are decomposing input signal into different frequency bandwidth and the coefficient of bandpass filters considered as a representation of the signal. The frequency response of the tuning curves indicate that like wavelet mother function and daughter functions, each frequency response of the tuning curve could be obtain by shifting and translation of certain tuning curve frequency response.[7], [8]



(a)



(b)

Figure 3.(a) A comparison of isovelocity response from a guinea-pig basilar membrane and neural isoresponse from a guinea pig spiral ganglion cell. Both curves show the level of input stimulation required to maintain a constant output response (b) Rhode's data, taken from a live squirrel monkey using the Mossbauer technique. The two curves indicate responses of the basilar membrane at two different positions, x_1 and x_2, on the basilar membrane, where x_1 is 1.5 mm closer to the apex than x_2. The best fit lines in the amplitude figure were drawn by Rhode.

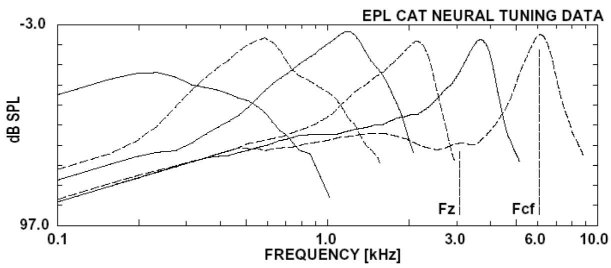


Figure 4: Cat neural tuning curves from Eaton Peabody Lab. The pressure scale, in dB, has been reversed to make the curves look like filter transfer functions. The response "tail" for the 6 kHz neuron is the "flat" region between 0.1 kHz and frequency in the tail the sound must be above 65 dB SPL (which on this scale is down) before the neuron will respond.

It is well known from Fourier theory that a signal can be expressed as the sum of a possibly infinite, series of sines and cosines. This sum is also referred to as a Fourier expansion. However, the big disadvantage of a Fourier expansion is that it has only frequency resolution and no time resolution. This means that although we might be able to determine all the frequencies present in a signal, we do not know when they are present. To overcome this problem in the past decades several solutions have been developed which are more or less able to represent a signal in the time and frequency domain at the same time. The *wavelet* transform or wavelet analysis is probably the most recent solution to overcome the shortcomings of the Fourier transform. In wavelet analysis the use of a fully scalable modulated window solves the signal-cutting problem. The window is shifted along the signal and for every position the spectrum is calculated. Then this process is repeated many times with a slightly shorter (or longer) window for every new cycle. At the end, the result will be a collection of time-frequency representations of the signal, all with different resolutions. Because of this collection of representations we can speak of a multiresolution analysis. In the case of wavelets we normally do not speak about time-frequency representations but about time-scale representations. Scale being in a way the opposite of frequency, because the term frequency is reserved for the Fourier transform. The wavelet analysis described in the introduction is known as the *continuous wavelet transform* or *CWT*. More formally it is written as:

$$\gamma(s,\tau) = \int f(t)\psi^{*}_{s,\tau}(t)dt \qquad (1)$$

where * denotes complex conjugation.

This equation shows how a function ***f(t)*** is decomposed into a set of basis functions $\psi_{s,\tau}(t)$, called the wavelets. The variables ***s*** and τ, scale and translation, are the new dimensions after the wavelet transform. The wavelets are generated from a single basic wavelet $\psi(t)$, the so-called *mother wavelet*, by scaling and translation:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right) \qquad (2)$$

In (2) ***s*** is the scale factor, τ is the translation factor and the factor $s^{-1/2}$ is for energy normalization across the different scales. It can be shown that the transform which is done in the cochlea could be estimated by a kind of wavelet transform. The two main roles of the cochlea are to separate the input acoustic signal into overlapping frequency bands, and to compress the large acoustic intensity range into the much smaller mechanical and electrical dynamic range of the inner hair cell.

Variations of air pressure at the ear are mechanically transferred into movement of the basilar membrane which is located in the cochlea. The basilar membrane is equipped with hair cells that react on deviation of the membrane from its rest position if the cochlea is imagined unrolled the basilar membrane extend along real axis. Sound information at any point can be represented as real function **B(x,t)**, the deviation of the membrane inside the cochlea at position x and time **t**. Experimental measurements show that for a sinusoidal stimulus the response might be:

$$B(x,t) = \psi_{\omega}(x - \log(\omega))e^{j\omega t} \qquad (3)$$

Which it means that the dependency of $\psi\omega$ ( ) is approximately a logarithmic shift so for input acoustic signal **f(t)** which has got its Fourier Transform **F(ω)**

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(\omega)e^{j\omega t}d\omega \qquad (4)$$

$$F(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-j\omega t}dt \qquad (5)$$

By substituting **f(t)** in (3) and putting (4) as **f(t)** the **B(x,t)** function could be written as follow:

$$B(x,t) = \frac{1}{2\pi}\int F(\omega)\,\psi_{\omega}(x - \log(\omega))e^{j\omega t}d\omega \quad (6)$$

$$B(x,t) = \frac{1}{2\pi}\iint f(\tau)e^{-j\omega\tau}d\tau \; \psi_{\omega}(x - \log(\omega))e^{j\omega t}d\omega \quad (7)$$

By some mathematical manipulation of (7), following term will appear:

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}\psi_{\omega}(x - \log\omega)e^{j\omega(t-\tau)}d\omega \qquad (8)$$

By considering **x=-log s,** the final equation for basilar membrane movement could be written as:

$$B(t,-\log s) = \int f(t)\frac{1}{s}\psi'(\frac{t-\tau}{s})dt \qquad (9)$$

Therefore, the transform of acoustic signal from the eardrum to the cochlea with a logarithmic scale along the basilar membrane could be approximated by continuous wavelet transform. Physiological observation justify that, the auditory system has got sort of wavelet like transform behavior. Neural tuning is measured by measuring the spiking activity in an auditory nerve fiber as a function of the frequency and intensity of a probe search tone. The locus of threshold intensities that cause the neuron to fire slightly above its spontaneous rate is called the neural tuning curve. The superscript indicates that the probe intensity is at threshold. Each neuron has such a tuning curve, which is tuned to its “best” characteristic frequency.[8]

Results

As it was shown the wavelet transform perform the log-linear frequency analysis and constant quality factor and can be used as an approximation of auditory acoustic signal transform. The cochlear impulse response was used for choosing the analyzing wavelet transform. The impulse response at 20mm from the oval window was selected as a wavelet function because its peak frequency about 1000Hz and in log-linear scale this frequency is almost at the center of the audible range. By this consideration an auditory wavelet transform could be realized.

Discussion

A comparison between the auditory periphery acoustic signal transform and wavelet transform shows that, also there are similarities especially in band-pass filter bank property and variable time and frequency resolution with the signal frequency, but the experimental measurement shows that some differences exist, and the main difference is in the quality factor. Wavelet transform is a filter bank with constant quality factor, but physiological research in hearing system found the quality factor which is changing and highly influenced by the activities of the hair cells.

Conclusion

The wavelet transform performs the log-linear frequency analysis with constant quality factor filtering and can therefore simulate the auditory model. The cochlear tuning curve is used for choosing the analyzing wavelets or mother functions which determine the overall filter shape. The impulse response at medium distance from oval window could be chosen as an analyzing wavelet because its peak frequency is almost at the center of audible range (on a log-linear scale) Figure 1 shows the impulse responses (in inverse scale) of auditory tuning curves. These responses satisfy the admissibility condition so they can be used as a mother function. So analysis and synthesis stage could be implemented.

References

[1] Barbour, D. L., Wang D. X., “Contrast tuning in auditory cortex”, *Science*, 299: 1073–1075, 2003.

[2] Crawford, A. C., R. Fettiplace, “An electrical tuning mechanism in turtle cochlear hair cells”, *J. Physiol.*, 312: 377–412.1981.

[3] Pickles, J. O. An Introduction to the Physiology of Hearing, London: Academic Press, 1988.

[4]. Sellick P. M,. Patuzzi, Johnstone B. M.,” Measurement of basilar membrane motion in the guinea pig using the Mossbauer technique”, *Journal of the Acoustical Society of America*, 72:131{141, 1982.

[5] Lyon R. F., “Automatic gain control in cochlear mechanics”, In P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, C. R. Steele, editors, The Mechanics and Biophysics of Hearing, Springer-Verlag, Berlin, 1990.

[6] Irino T., Patterson R. D., “A time-domain, level-dependent auditory filters: the gammachirp", *J. Acoust. Soc. Am.*, Vol. 101, No. 1, pp. 412{419, January 1997.

[7] Irino T., Patterson R. D., “A compressive gammachirp auditory filter for both physiological and psychophysical data", *J. Acoust. Soc. Am.*, Vol. 109, No. 5, pp., May 2001.

[8] Irino T., Unoki M., “An analysis/synthesis auditory filter bank based on an IIR implementation of the gammachirp," *J. Acoust. Soc. Jap.*, Vol. 20, No. 5, pp. 397{406, November 1999.