NOVEL BRICK-WALL FILTERS BASED ON THE AUDITORY SYSTEM

Amitava Biswas. Department of Speech and Hearing Sciences
Indiana University, Bloomington, IN 47408, USA
E-mail: abiswas@indiana.edu

ABSTRACT

A novel class of narrow band filters is presented that offers great rejection of out of band noise and a flat top at the peak. The filter is unconventional, as the output of such a filter is a series of spikes, like the action potential in the auditory nerve. The product of its time and frequency window is less than unity even using 100% output cutoff points. A bank of such filters has been used to compute a spectrogram like display that has no streaks as seen in conventional spectrograms. This model differs from Lyon's cochlea in several areas particularly energy management and damping factor requirements. Its Q factor depends on the signal intensity, and it produces acoustic cubic distortion products similar to the auditory system, however, it is basically linear with a very wide dynamic range and its dynamics can be analyzed using linear filter theory.

1. INTRODUCTION

The human auditory system can discriminate very small frequency differences in very short intervals of time [9], but such capability is not easily achieved in practical filter designs. This paper addressed a novel class of filters that gives superior time-bandwidth product. This could be a solution of the auditory system to maintain faster time response than Helmholtz's lightly damped model, and narrower frequency response than Bekesy's heavily damped model of the inner ear at the same time [3]. Other than that, a smaller product of time and frequency resolutions is important in many practical applications including but not limited to spectrogram [6], speech recognition, and speech assessment.

The basic building block of our filter unit is a mass spring unit. Unlike conventional analog filters, and similar to the neuronal spikes in the auditory nerve [10], output of the proposed filter unit is a series of spikes depending on the intensity of stimulus. As soon as the internal energy level in the filter unit reaches a threshold, a spike is released and its internally stored energy is cleared. A similar filter has been described by the author [2] where the spike production was dependent on threshold amplitude of the spring compression instead of the total internal energy as in the current design. If no subsequent stimulus is applied, internal energy can not reach the threshold any more and no further spike is produced by the unit, giving a brick-wall shaped cutoff in time response. In a sense, this is like 'impulse response is an impulse'! On the other hand, an ordinary analog filter would drop in output after the excitation is removed, but continues to produce a little output indefinitely before discharging its internal energy completely. The frequency response of the proposed filter unit is brick-wall shaped too. If a sinusoidal stimulus is applied at the natural frequency of the unit, it resonates with the input excitation, and its amplitude of oscillation grows linearly and its internal energy reaches the threshold limit eventually, even if the input excitation is small but maintained for sufficiently long time. When a sinusoidal stimulus differs from the natural frequency of the unit, the amplitude of oscillation can not grow indefinitely. The envelope of the amplitude of oscillation keeps on rising and falling at the beat rate which is the difference between the input frequency and the natural frequency. Therefore the internal energy level periodically rises to a maximum even if the input excitation intensity is maintained at a fixed level. This maximum possible internal energy level drops as the difference between the applied frequency and the natural frequency increases. If the frequency difference is more than a certain limit, no spikes could be produced, resulting into a brick-wall shaped frequency response. The frequency range over which spikes will be produced depends on the intensity of applied excitation. If the intensity is higher then the allowable frequency range is wider. On the other hand, as the intensity is increased the time period between consecutive spikes, which is also the response time of the unit, is reduced proportionately. We will show in the next section that the product of the time period and the frequency range is less than unity. For an ordinary analog filter neither the time response nor the frequency response has well defined cutoff points. If we take their cutoff points where most of the energy is confined, the product of frequency and time resolution exceeds unity [7].

2. THE MODEL

Our undamped mass-spring model has a mass m, with a spring of stiffness k, therefore, its natural frequency is $\omega_0 = \sqrt{k/m}$. There is no internal energy at time $t = 0$, and as soon as the internal energy reaches the threshold E_{th}, a spike is discharged. It appears unwise to keep that energy 'hanging around' any longer, as it would masquerade subsequent energy influx. Therefore the internal energy is immediately voided. The action will be referred to as a DV to emphasize the need to 'discharge *and* void'.

2.1 Behavior at Resonance

When the input force is given by

$$f = A\sin(\omega_0 t + \theta) \tag{1}$$

From earlier work by the author [2], the displacement and velocity are given by

$$x = \frac{-tA\cos(\omega_0 t + \theta)}{2\omega_0 m} + \frac{A\cos(\theta)\sin(\omega_0 t)}{2\omega_0^2 m} \tag{2}$$

And

$$v = \frac{tA\sin(\omega_0 t + \theta)}{2m} + \frac{A\sin(\theta)\sin(\omega_0 t)}{2\omega_0 m} \tag{3}$$

The total internal energy at any instant can be obtained from the potential energy in the spring component and the kinetic energy in the mass component. Therefore the total energy is given by

$$E = \tfrac{1}{2}mv^2 + \tfrac{1}{2}kx^2 = \frac{t^2A^2}{8m} + \tag{4}$$

$$+\frac{tA^2\sin(\omega_0 t)\cos(\omega_0 t + 2\theta)}{4\omega_0 m} + \frac{A^2\sin^2(\omega_0 t)}{8\omega_0^2 m}$$

The first term in above expression is larger than other two terms, when $t\omega_0 > 2$. Therefore, if the DV rate is below the natural frequency of the unit, the expression may be approximated to the first term and it equals to the threshold energy level when the next DV takes place.

$$E_{th} \cong \frac{\Delta t^2 A^2}{8m} \tag{5}$$

Therefore the DV period is given by

$$\Delta t \cong \frac{2\sqrt{2mE_{th}}}{A} \tag{6}$$

and intensity

$$A \cong \frac{2\sqrt{2mE_{th}}}{\Delta t} \tag{7}$$

The earlier work by this author [2] derived a similar expression but an important difference is that now the threshold is energy based instead of amplitude of displacement or velocity. A DV can now take place at any time the energy threshold is exceeded. Therefore the strong tendency to phase lock seen earlier [2] is not apparent in this case. On the other hand, large phase jitter could happen from one DV to the next DV in this model depending on Δt.

2.2 When not at Resonance

When the frequency ω_1 of input stimulus differs from the natural frequency ω_0 of the model, amplitude of oscillation can not grow indefinitely but shows a beat phenomenon and reaches a maximum value periodically at the beat frequency [2].

$$x_{\max} = \frac{A}{\omega_0(\omega_1 - \omega_0)m} \tag{8}$$

This implies a maximum level of internal energy that can accumulate at any time. Therefore, to obtain a DV, we need to have

$$E_{th} \le \tfrac{1}{2}kx_{\max}^2 = \frac{A^2}{2(\omega_1 - \omega_0)^2 m} \tag{9}$$

This decides the maximum and minimum values of input frequency to get a DV.

$$\omega_{1(\max/\min)} = \omega_0 \pm \frac{A}{\sqrt{2mE_{th}}} \tag{10}$$

Total bandwidth in Hz is given by

$$\Delta f = \frac{A}{\pi\sqrt{2mE_{th}}} \tag{11}$$

2.3 Time-Frequency Product

Therefore from equations 6 and 11

$$\Delta t.\Delta f \cong \frac{2}{\pi} \tag{12}$$

There are some interesting consequences -- the product of instantaneous DV period and the instantaneous bandwidth is independent of the filter parameters although each of the two factors depends on the input level and the filter parameters. Therefore, if a unit gives a certain DV rate, no matter what its natural frequency is, its bandwidth is predictable from the DV rate alone. If the DV rate is lower then the bandwidth is narrower. If the DV rate is higher then the bandwidth is wider. The tuning curve of neurons in the auditory nerve is consistent with this phenomenon [10]. On the other hand setting another threshold can change the DV rate and the bandwidth. The efferent path to the outer hair cells in the cochlea probably has a similar function [10].

3. COMPUTER RESULTS

To verify the mathematical formulation, a computer model was simulated using a mass spring damper system under a sinusoidal excitation. The model parameters were m = 1, natural frequency = 1000, threshold energy = potential energy at displacement of 0.001. The internal energy level was monitored to execute a DV as soon as the threshold is reached. Figure 1 shows that in the absence of any damping factor there is a linear relationship between input intensity level and average DV rate over a very broad range. If the damping factor is negative then the DV rate drops too slowly for lower intensity levels. If the damping factor is positive then the DV rate drops to zero too quickly. All three traces get closer at top right corner as the DV rate nears the natural frequency. Difference from our previous design [2] is seen in the smoothness of the trace of average DV rate for zero damping factor, whereas it was like a staircase due to strong phase locking tendency [2]. Small wavy envelopes are drawn around the trace for zero damping near the top right corner indicating one standard deviation of the DV rate from the mean level.

Figure 2 shows the brick-wall shaped frequency response in DV/second of a unit of natural frequency 1000 Hz, using several input stimuli frequencies and at three amplitude levels. As expected from analytic formulation, the pass band is broader for higher intensity levels. This may have important implications in auditory modeling because all known models use external means to reduce the Q-factor at higher intensity levels, which may not be necessary with this approach. Another interesting point is that for auditory nerves the slope of the tuning curve is extremely steep, of the order of 200-600 dB/octave and attempts to model such a steep slope using conventional means do not give sufficient bandwidth at the peak [10]. This reason is -- the shape of the tuning curve is still like a bell. It is only compressed horizontally, giving a sharp pointed peak and never as rectangular as in this model.

Figure 3 shows cubic distortion product level of the amplitude of oscillation for several units. Cubic distortion products (DP) have been observed in healthy ear [5][1][10], therefore we decided to measure DP values in our model. DP is commonly measured by applying two frequencies F2 and F1 simultaneously and monitoring amplitude of distortion product at 2F2-F1. To be consistent with the format reported by Harris et al [4], for every F2/F1 ratio, the values of F2 and F1 were chosen to produce 2F2-F1 at 1000 Hz. Natural frequency of the filter unit was set at F1. 2F2-F1 and noise floor was calculated using spectrum of the displacement waveform. Input intensity was set at 8 different levels for each F2/F1 ratio to get 8 equally spaced DV rates from 175 to 350 per second. Similar to the report by Harris et al [4], we notice nonmonotonic change in DP level with F2/F1 ratio. Interestingly, DP levels in our model peaks when F2-F1 difference is close to the DV rate. For higher F2/F1 ratio F2-F1 difference was larger and best DV rate to create maximum DP was higher in our model. This appears to be consistent with the trend seen in humans [4]. Interestingly, our model demonstrated that the acoustic cubic distortion products can be generated by a system that is basically linear with a very wide dynamic range and its dynamics can be analyzed using linear filter theory. Whereas current models of the auditory system that produce acoustic cubic distortion products are based on nonlinear elements and require nonlinear analysis [11].

Figures 5 and 6 compare a conventional spectrogram and a DV history plot. 250 filter units tuned at equally spaced intervals from 20 Hz to 5000 Hz produced the DV history plot. The threshold of each unit was set to produce about 280 DV/second when the input signal is at resonance with that unit. The input waveform for figures 5 and 6, is shown in figure 4. The conventional spectrogram in figure 5 shows vertical streaks but the DV history plot in figure 6 is remarkably clean without any postprocessing.

4. DISCUSSION

The presented approach is remarkably different from the commonly used cochlear model, originally developed by Lyon [8], particularly in internal energy management and damping factor requirements. It is interesting to compare the implications of choosing the basis of DV threshold between energy and displacement. As shown by our current and previous [2] models, both approaches ensure brick-wall shaped frequency response and same value of time-frequency response product. The energy threshold approach shown in this paper gives very smooth relationship between input intensity and output DV rate but implementation mechanism of the displacement threshold approach could be simpler. The payoff for the displacement threshold is a staircase like relationship between input intensity and output DV rate due to phase locking tendency to integral number of cycles of the input waveform [2]. In mammalian cochlea, phase locking tendency in the firing rate of auditory neurons is very evident [10]. In that case neuronal firing rate alone could be inadequate to convey the intensity coding. Some recovery of intensity information may be possible by evaluating the phase of the phase locked condition. At present we are unable to offer any practical solution for our model to interpret the phase information, that has a parallel in the mammalian cochlea. We hope to find a solution in near future. Probably it would need some kind of a phase sampling device or a reference clock mechanism. In the mammalian cochlea the efferent projection to the cochlea via the olivocochlear bundle probably perform a similar function. No conclusively supportive or contrary physiological evidence is currently available because of several technical limitations and also because the current experiments are conducted on anaesthetized or decerebrated animals [10]. Even though we do not know yet how far a real cochlea follows our model, we hope the principles presented here could be used in man-made filters for various purposes.

5. ACKNOWLEDGEMENT

Steven M. Barlow, Ann Clock Eddins, David A. Eddins, Charles S. Watson, Larry E. Humes, Ward R. Drennan, Richard D. Andreatta and Donald S. Finan gave encouragement and support. Work partially supported by Neuro Logic, Inc.

6. REFERENCES

[1] Allen J. "Why is the 2f1-f2 distortion product maximum at f2/f1=1.2?". *Journal of the Acoustical Society of America,* 91:2410, 1992.

[2] Biswas A. "A Spectrography Algorithm based on Cochlear Mechanics". *Proc. Of ICSPAT,* vol. 1, San Diego, September 1997, pages 246-250.

[3] Dallos P. "The Active Cochlea". *The Journal of Neuroscience,* 12 (12):4575-4585, 1992.

[4] Harris B. L., Konsbury-Martin B. L., Stagner B. B., Coats A. C., and Martin G. K. "Acoustic distortion products in humans: Systematic changes in amplitude as a function of f2/f1 ratio". *Journal of the Acoustical Society of America,* 85:220-229, 1989.

[5] Kemp D. T. "What are otoacoustic emissions?". *International Symposium on Otoacoustic Emissions*, Kansas City, 1991, page 7.

[6] Kent R. D., and Read C. *The Analysis of Speech*. Singular Publishing Group, San Diego, CA. 1992.

[7] Kharkevich A. A. *Spectra and Analysis*. Consultants Bureau, New York, 1960.

[8] Lyon R. F., Mead C. "An Analog Electronic Cochlea". *IEEE Transactions on Acoustics, Speech, and Signal Processing,* 36(7):1119:1133, 1988.

[9] Moore B. C. J. *An Introduction to the Psychology of Hearing*. Academic Press, San Diego, CA, 1989.

[10] Pickles J. O. *An Introduction to the Physiology of Hearing*. Academic Press. San Diego, CA, 1996.

[11] Zwicker E. "Suppression and (2f1-f2)-difference tones in a nonlinear cochlear preprocessing model with active feedback". *Journal of the Acoustical Society of America,* 80:163-176, 1986.

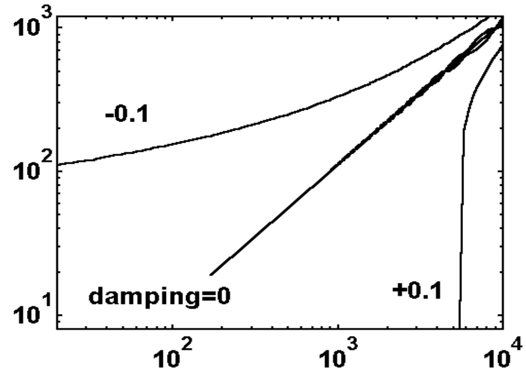

Figure 1. DV rate (Y) at resonance as a function of stimulus intensity (X). The damping factors are –0.1, 0, and +0.1 as marked in figure.

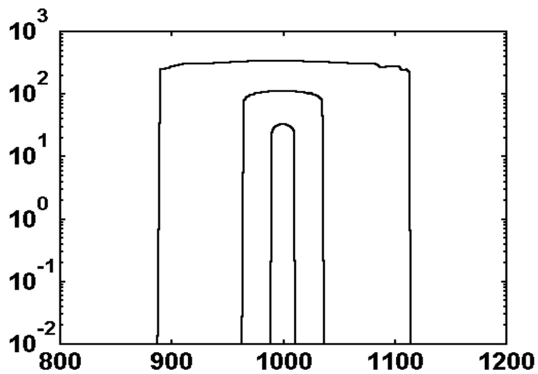

Figure 2. DV rate (Y) with stimulus frequency (X) at 300 (inner), 1000, and 3000 (outer) stimulus intensity. Damping factor is zero.

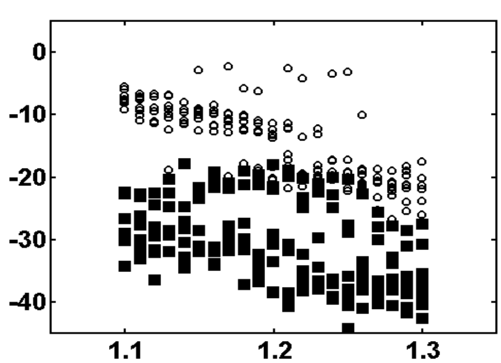

Figure 3. Distortion product amplitude in dB (Y) as a function of F2/F1 ratio (X). Circles indicate DP level, squares indicate noise floor.

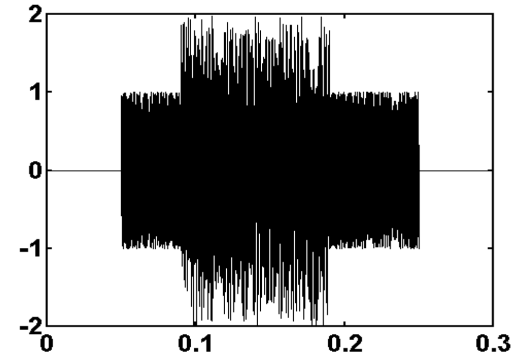

Figure 4. Composite signal, amplitude (Y) with time (X) in seconds. A 0.1 second long signal is added to a 0.2 second long signal.

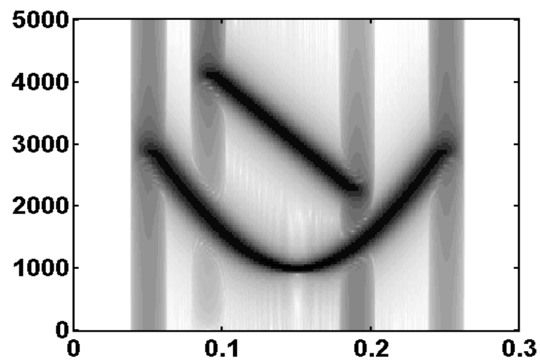

Figure 5. Spectrogram of the signal shown in figure 4, using conventional short term Fourier transforms. Note four vertical streaks.

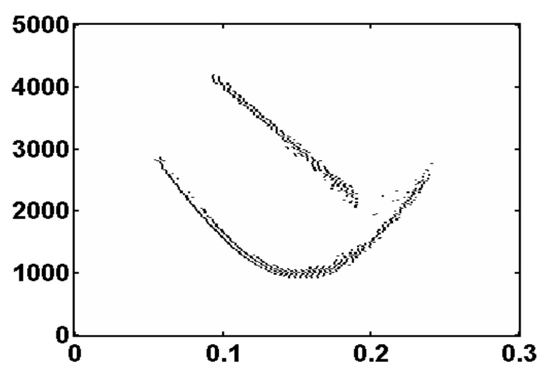

Figure 6. DV history of 250 units equally spaced from 20 to 5000 Hz. Each unit gives about 280 DV/second when it resonates with the input.