NATURAL HEARING MODEL BASED ON DYADIC WAVELET

Amr M. Gody*
Cairo University

The logarithmic nature of the dyadic wavelet has been used to generate a human like hearing model. The model is used to represent the Arabic vowels as an example. It is a highly discriminative and simple model. The speech information has been split into components in a logarithmic frequency bands as the manner in the human auditory system. This human like analogy gives the extracted information a great immunity against unwanted noise.

1. Introduction

It is believed that the best auditory system have been ever discovered is the human auditory system. It is highly efficient in speech understanding even in a very bad environment of low signal to noise ratios.
Many researchers spent a lot of time to study and to understand how this system works and how it models the speech information. It is discovered that the human speech understanding system starts at the ear. The received sound is split into frequency components. The components are logarithmically related as in figure 1 [1].

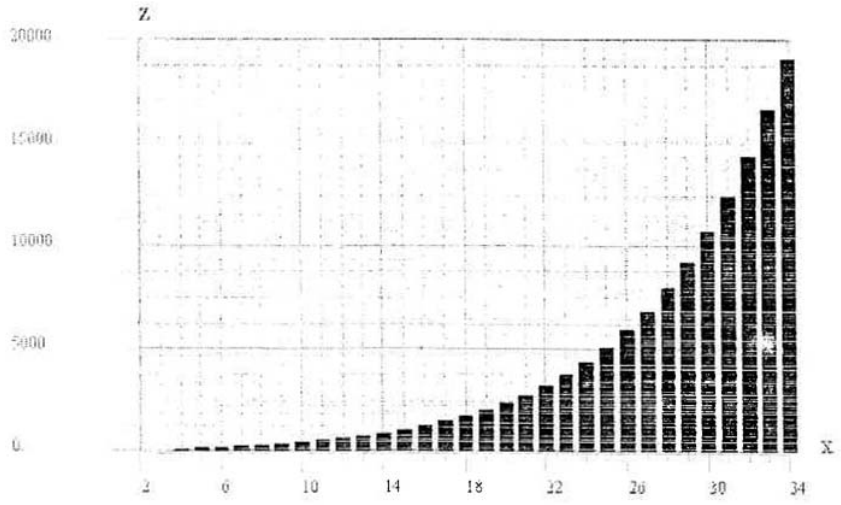

Figure 1. Triggering frequencies of the human ear's sensors. X-axis is the sensor number and Y-axis is the triggered frequency.[1]

* Department of Electronics and communication Engineering, Faculty of Engineering , Cairo University , Fayoum Branch, El-fayoum, EGYPT, E-mail: agody@ieee.org.

In simple words, human ear acts as a spectrum analyzer. Brain receives the spectral information from the auditory sensors and makes the decision based on the pre-trained models stored inside.

In the last 3 decades, many mathematicians have put rules for the multiresolution analysis [2…7]. Wavelet transform is the baby of this effort. Instead of decomposing the signal into frequency components it is rather projected into different frequency bands using the scaling property of the wavelet transform. Dyadic wavelet gives a logarithmic bands nature, see figure 2.

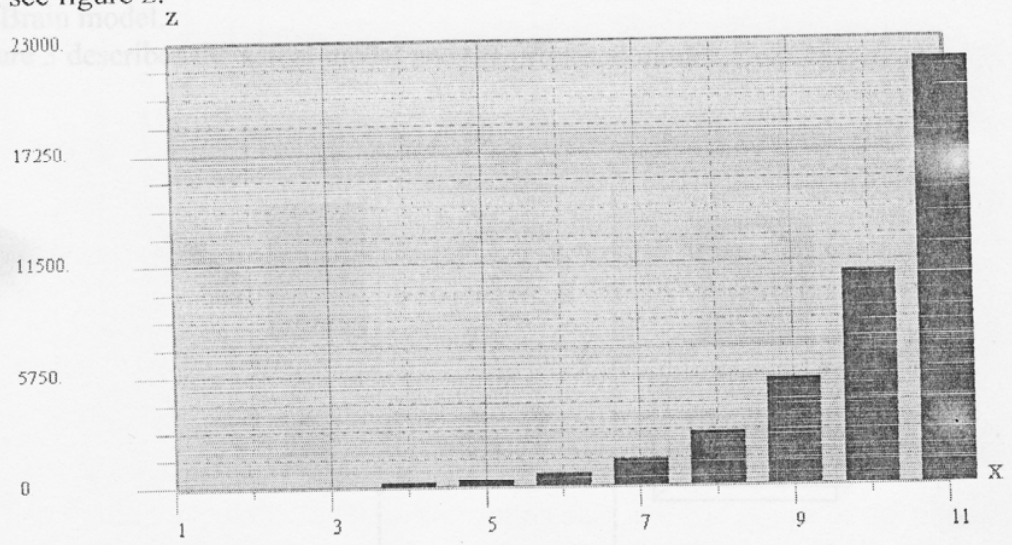

Figure 2. Dyadic wavelet bandwidths. The x- axis is the band title and the y-axis is the bandwidth in Hz.

In this work, a trial to use the above analogy to construct new speech parameters that are highly immune to noise is done. Arabic vowels representations using the proposed parameters are shown as an example.

2. Noise immunity concept

In the human, it is found that triggering frequency steps are small in the lower frequency ranges, as shown in figure 1. This gives the human auditory system a high immunity against noise hit. Brain can compensate noise hit by correlating the information from adjacent sensors.

The logarithmic nature of wavelet representation makes it highly analogous to the human ear system. As it was shown in figure 2, smaller bands are equivalent to human sensors in the low frequency area. In human speech, a great portion of speech information is concentrated in the lower frequency bands [9]. For simplicity consider the white noise. It is considered as a constant level in all bands in the frequency domain. Hence, smaller

bands have smaller bandwidth and smaller noise power contribution than higher bands. The result is that the wavelet parameters in the lower bands are approximately noise filtered compared to those in the higher bands.
In addition, the signal representation in the lower bands is highly correlated due to the small frequency stepping between bands in the low frequency ranges. In the same time the noise power level is attenuated by half in any adjacent bands.

3. SYSTEM MODEL

Hearing / understanding system is divided into 3 basic parts:

1. Signal detection
2. Spectrum analyzer
3. Brain model.

Figure 3 describes the actual model and the proposed model.

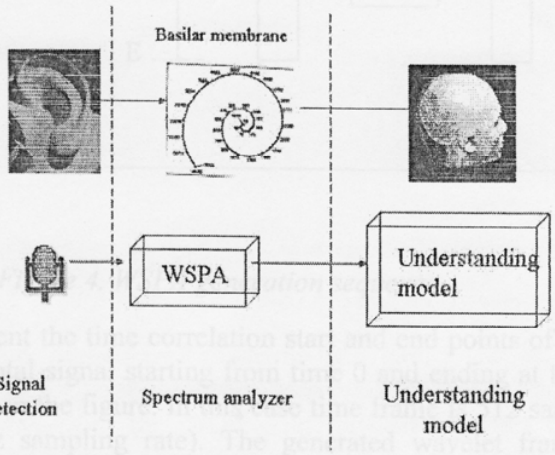

Figure 3. System model. WSPA is for Wavelet-based Spectral Analyzer.

WSPA is the wavelet based spectral Analyzer. Speech signal is sampled and framed. Frame duration is 45 ms. Daubechies wavelets of 4 points is applied to the time frames to generate the associated parametric wavelet frames. The details will be discussed in the next section.
Figure 3. illustrates the analogy between the proposed model and the human understanding model. This work is focused on WSPA parameters. The advantages of WSPA parameters will be discussed in a later section.

4. Wavelet-Based Basilar membrane model

In this section the process of generating WSPA parameters will be discussed. Figure 4 illustrates the sequence.

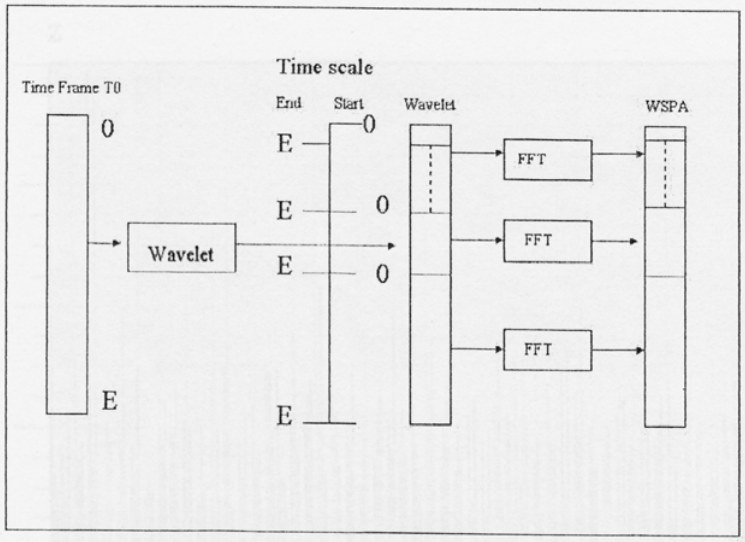

Figure 4. WSPA generation sequence

0 and E in figure 4 represent the time correlation start and end points of the time frame. Each band represent the total signal starting from time 0 and ending at time E. Wavelet frame is divided as shown in the figure. In this case time frame is 512 samples (about 45 ms in case of 11025 Hz sampling rate). The generated wavelet frame is also 512 parametric vector. The parameters in the wavelet vector are distributed as in table 1.

Table 1 Correlation between parametric wavelet frame and frequency bands. Sampling rate = 11025Hz, frame length = 512.

Band#	Parameters range	Frequency band
0	256 – 511	2756 – 5512 Hz
1	128 – 256	1378 – 2756 Hz
2	64 – 128	689 – 1378 Hz
3	32 – 64	344 – 689 Hz
4	16 – 32	172 – 344 Hz
5	8 – 16	86 – 172 Hz
6	0 – 8	0 – 86 Hz

5. WSPA advantages over short time FFT.

As illustrated in section 2 in this paper, the logarithmic nature of wavelet parameters gives the lower frequency bands a good immunity to noise. Signal in the lower bands have a lot of speech information and a low noise contribution. Wavelet phase makes an

implied noise rejection stage before spectral analysis. Then FFT is applied to the reduced parameters set that represents the whole signal as illustrated in figure 4. That makes it faster and worthy (it spectrally analyzes a highly informative signal rather than the whole signal in case of direct FFT analysis). Figure 5. is a typical WSPA frame for a certain vowel.

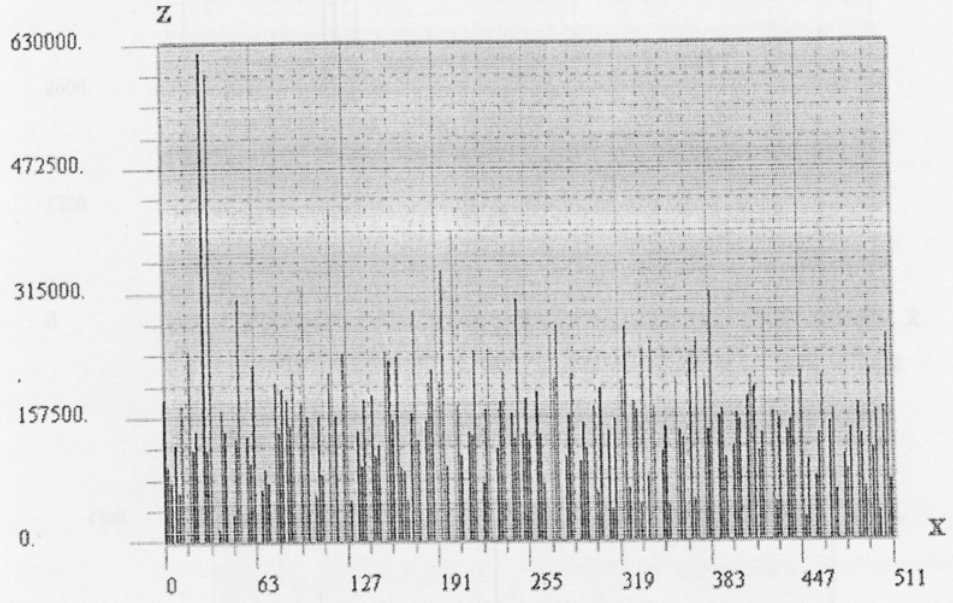

Figure 5. WSPA frame for a part of vowel /a/(فتحة) .

Starting from WSPA frame, human understanding like model can be derived. In this work a closer touch is made for Arabic vowels. Vowels have a low frequency nature so bands 0 and 1 are excluded from the study. The following figures represent typical WSPA frames in the required frequency range for Vowels discrimination.

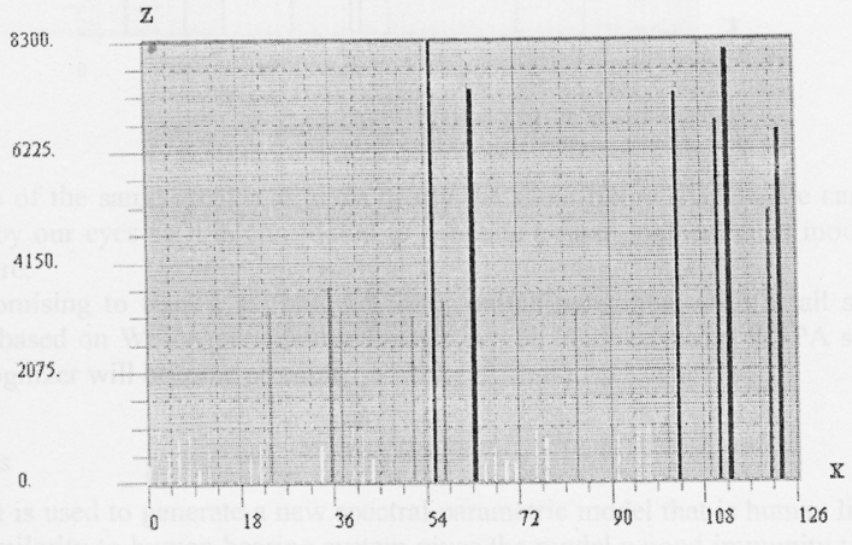

Figure 6. WSPA for vowel /a/(فتحة)

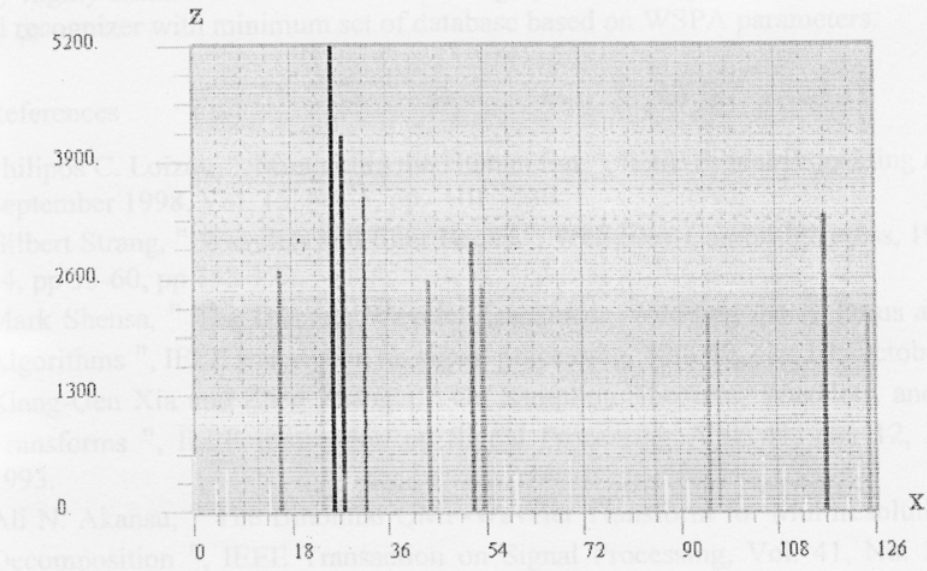

Figure 7. WSPA for vowel /i/(كسرة)

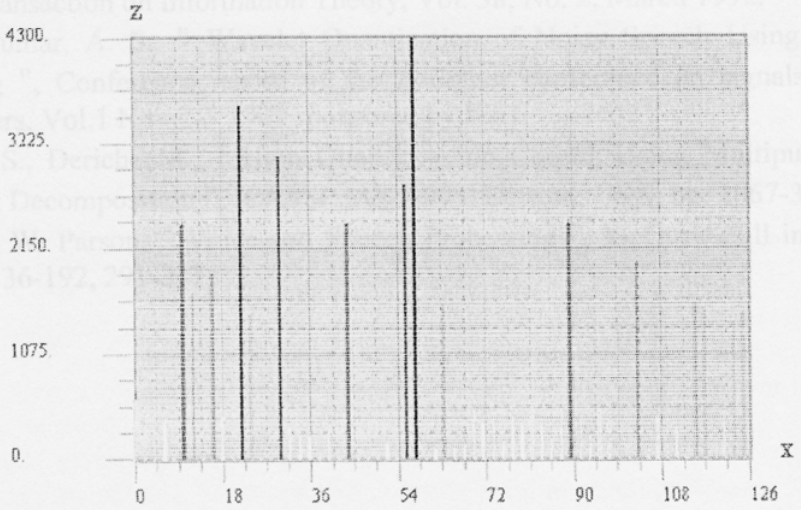

Figure 8. WSPA for vowel /o/(ضمة)

For many trials of the same speaker it gives nearly the same figure shapes. We can see the difference by our eyes so it is considered to generate a good mathematical model in the nearest future.

It is highly promising to train a simple and faster online recognizer with small set of database units based on WSPA parameters. Phones can be modeled using WSPA so the real online recognizer will become possible.

6. Conclusions

Dyadic wavelet is used to generate a new spectral parametric model that is human like in nature. This similarity to human hearing system gives the model a good immunity to any noise added. WSPA parameters give a new domain for studying speech understanding by

a way highly similar to human understanding system. It is promising to realize a high-speed recognizer with minimum set of database based on WSPA parameters.

7. References

[1] Philipos C. Loizou, " Mimicking the Human Ear ", IEEE Signal Processing magazine, September 1998, Vol. 15, No. 5, pp. 101 –130.

[2] Gilbert Strang, " Wavelets and filter Banks ", Wellesley-Cambridge press, 1996, pp 1-34, pp 53-60, pp 155-172.

[3] Mark Shensa, " The Discrete Wavelet Transform: Wedding the A Trous and Mallat Algorithms ", IEEE transaction on signal processing, Vol. 40, No. 10, October 1992.

[4] Xiang-Gen Xia and Zhen Zhang, " On Sampling Theorem, Wavelets, and Wavelet Transforms ", IEEE transaction on Signal Processing, Vol. 41, No. 12, December 1993.

[5] Ali N. Akansu, " The Binomial QMF-Wavelet Transform for Multiresolution Signal Decomposition ", IEEE Transaction on Signal Processing, Vol. 41, No. 1, January 1993.

[6] Ahmed H. Tewfik, " On the Optimal Choice of a Wavelet for Signal Representation ", IEEE Transaction on Information Theory, Vol. 38, No. 2, March 1992.

[7] Madhukumar, A. S., " Wavelet Quantization of Noisy Speech Using Constrained Filtering ", Conference record of the Asilomar conference on signals, systems & computers, Vol.1 Nov 2-5 1997 sponsored by IEEE.

[8] Boland S., Deriche M., " High Quality Audio Coding Using Multipulse LPC and Wavelet Decomposition ", ICASSP May 1995, Detroid, USA, pp. 3067-3069

[9] Thomas W. Parson, " Voice and Speech Processing ", McGraw-Hill inc., 1987, pp. 57-98, 136-192, 291-317