Optimal Wavelets for Speech Signal Representations

Shonda L. Walker and Simon Y. Foo
Electrical Engineering Department, Florida A&M University
2525 Pottsdamer Street, Tallahassee, FL 32310
{swalker,foo}@eng.fsu.edu

ABSTRACT

It is well known that in many speech processing applications, speech signals are characterized by their voiced and unvoiced components. Voiced speech components contain dense frequency spectrum with many harmonics. The periodic or semi-periodic nature of voiced signals lends itself to Fourier Processing. Unvoiced speech contains many high frequency components and thus resembles random noise. Several methods for voiced and unvoiced speech representations that utilize wavelet processing have been developed. These methods seek to improve the accuracy of wavelet-based speech signal representations using adaptive wavelet techniques, superwavelets, which uses a linear combination of adaptive wavelets, gaussian methods and a multi-resolution sinusoidal transform approach to mention a few. This paper addresses the relative performance of these wavelet methods and evaluates the usefulness of wavelet processing in speech signal representations. In addition, this paper will also address some of the hardware considerations for the wavelet methods presented.

Keywords: Speech recognition, adaptive wavelets, super wavelets, quadrature spline wavelets, speech coding.

1. INTRODUCTION

Traditional techniques for speech signal analysis use Fourier methods for signal processing. Fourier analysis, however, only details the spectral content of a signal in the frequency domain. The time domain information for a particular event is lost during Fourier transformations because preservation of time instances is not considered. This condition can be overlooked if the signal is stationary. However, for nonstationary signals, like speech, time *and* frequency domain information is necessary to avoid any loss of significant information in the signal. Wavelet analysis provides an alternative method to Fourier analysis for signal processing. Wavelets apply the concept of multi-resolution analysis (i.e., time and frequency scale representations) to produce precise decompositions of signals for accurate signal representation. They can reveal detailed characteristics, like small discontinuities, self-similarities, and even higher order derivatives that may be hidden by the conventional Fourier analysis.

Speech can be classified as information carrying nonstationary acoustical signals. Based on this classification, speech is a good candidate for wavelet analysis because its variability in speech styles changes rapidly over time depending on the environment and speaker characteristics. This nonstationary behavior requires the use of signal processing techniques that adequately preserve information to avoid speech degradation over communication channels.

Consider the basic definition of a wavelet: Given a signal $f(t)$, it can be represented as

$$\tilde{f}(t) = \sum_{n=1}^{N} c_n \cdot w\left(\frac{t - b_n}{a_n}\right)$$

where $\tilde{f}(t)$ is the wavelet representation of $f(t)$, $w(t)$ is the mother wavelet and $w\left(\frac{t - b_n}{a_n}\right)$ represents a set of basis functions called the daughter wavelets with $a_n, b_n, and\ c_n$ as the dilation, translation and coefficient parameters of the mother wavelet, respectively. By correlating these wavelet parameters to speech components that represent voiced and unvoiced speech, an optimal wavelet representation of the signal can be determined. Furthermore, an appropriate choice for the mother wavelet, for example Daubechies, Morlet, or Haar, can be chosen such that the accuracy of the characterized speech is optimized and the speech intelligibility of the signal is preserved.

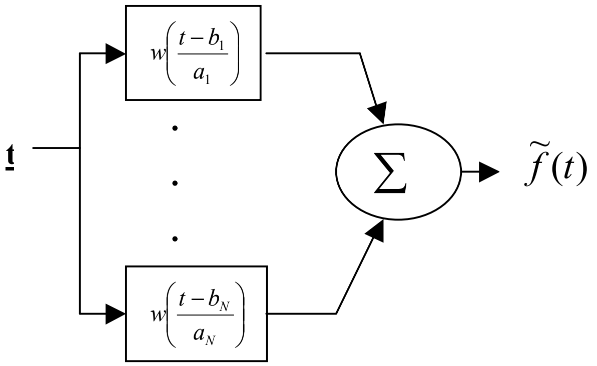
Section 2 will describe several methods that are being used for speech coding including adaptive wavelets, superwavelets, multiresolution wavelets and quadrature spline wavelets. Section 3 gives some hardware considerations for implementing speech recognition applications and section 4 presents some conclusions.

2. SPEECH CODING USING WAVELETS

Speech can be characterized as a combination of dense periodic frequency spectrum (voiced components) and high frequency content resembling random noise (unvoiced components). Using these classifications, wavelet processing can be used to code speech signals. Four wavelet based speech coding techniques are summarized in this paper: *adaptive wavelets, superwavelets, mulitresolution wavelets* and *qudrature spline wavelets.*

Adaptive Wavelets

Adaptive wavelets use the error function, $E = \left(f(t) - \tilde{f}(t)\right)^2$ to adaptively update the wavelet parameters (a_n, b_n, c_n) until error is minimized and an optimal speech signal is identified [1]. By constantly varying the parameters based on minimal error, the speech signal that best represents the original speech is found. Speaker variability and speaker rate can be addressed using this method since wavelet parameters are updated to better approximate the signal. The authors in [1] have implemented a speech coder which models voiced speech using the Morlet wavelet, $w(t) = \cos(1.75t) \cdot \exp(-t^2/2)$ as the mother wavelet. A single pitch period of voiced speech is captured and a nonlinear approach is applied. The coder implements a neural network architecture to produce the wavelet approximations and minimizes the error using the gradient descent algorithm. Figure 1 shows the architecture model used. Figure 2 shows an example of the original and reconstructed speech sound /A/. Using a vector quantization scheme, a low-bit rate speech coder was developed.



$w\left(\frac{t-b_1}{a_1}\right)$
t
Σ
$\tilde{f}(t)$
$w\left(\frac{t-b_N}{a_N}\right)$

Figure 1-Neural Network Architecture for Adaptive Wavelet Implementation.

Superwavelets

Szu, Telfer, et al. [2,3] used the concept of a super-wavelet, which is a linear combination of adaptive wavelets, to allow the fundamental shape, as opposed to the parameters, of the wavelet to vary adaptively. Adaptive wavelet processing was extended to include unvoiced speech characterization using super-wavelets. The authors used a third order Daubechies wavelet to model high frequency speech. They illustrate that the Daubechies wavelet technique provides a good approximation to the original signal. However, the neural network implementation is computationally intensive.

Multiresolution Wavelets

Anderson in [4] uses a multiple resolution sinusoidal transform (MRST) based on the wavelet concept of multiresolution, multi-scale analysis. This novel approach to speech coding derives a MRST that uses quadrature mirror filters (QMF) banks to obtain finer time resolution at high frequencies and better spectral resolution at lower frequencies. This method was shown to provide natural speech signal decomposition with good speech variability and sound quality. The MRST combines many sinusoids to produce signal decompositions that are comparable to speech. The signal is filtered at different levels using low pass and high pass filters as illustrated in Figure 3. By applying this approach iteratively, finer details in the signal can be filtered out and a decomposition of low and high frequencies of the original signal can be achieved. This approach allows good spectral resolution however sinusoids are used as a basis function as opposed to a wavelet function, thus variability and accuracy is compromised using this technique.

Quadrature Spline Wavelets

Griebel et al. [5] examine quadrature spline wavelets for use in wavelet analysis of speech in the presence of echoes or in a reverberant environment using a multiple microphone array to capture speech signals. This technique uses an event-based model of the glottal closure instance (GCI), a period in voiced sounds where the vocal cords are forced open due to air pressure, to discriminate impulses in speech signals. The authors illustrate that by identifying coherent signal structures over an array of microphones, the original signal can be reconstructed using the quadrature spline wavelet. The quadrature spline wavelet can efficiently identify discontinuities in the signal as a result of its definition as a derivative of smoothing functions. In particular, the quadrature spline wavelet resembles the following function

$$\Psi(\omega) = -\frac{j\omega}{4} e^{-j\frac{\omega}{2}} \left(\frac{\sin\frac{\omega}{4}}{\frac{\omega}{4}} \right)^4$$

where $\Psi(\omega)$ is the Fourier Transform of $w(t)$ and $j = \sqrt{-1}$. Figure 4 illustrates the results of this technique.

3. HARDWARE CONSIDERATIONS

Hardware considerations help to determine the practical efficiency of any algorithmic design. Identifying efficient hardware approaches to speech analysis systems can improve the quality of speech signals and provide state-of-the-art ASR systems. Since most reconfigurable architectures are algorithm-specific, a unique and optimal speech system architecture can be implemented in the FPGA environment. The computational requirements of speech recognition algorithms require hardware capable of supplying this processing power. Conventional architectures included digital signal processors (DSPs), however, field programmable logic devices, like FPGAs, are becoming increasingly more powerful and desirable for signal processing applications. FPGAs can exceed the performance of DSPs by utilizing parallel architectures and minimizing on-chip resources [6] via design synthesis. FPGAs are programmable like DSPs, they support reconfigurability, and they can be optimized to support a wide range of applications, not just digital signal processing [7].

Conventional speech recognition systems perform well under ideal environmental conditions however, high performance involves the recognition of speech under more practical conditions like high levels of background noise, reverberant acoustic environments, and even spontaneous conversations [8]. These conditions require more complex and robust speech recognition, which in turn requires an optimal architecture. FPGAs address many of the hardware considerations that can become problematic in speech designs, thus its use as a target platform in speech recognition systems seems advantageous.

4. CONCLUSIONS

The adaptive wavelet technique by Kadambe et al. used the Morlet function as the mother wavelet to represent voiced speech. Though adaptive processing increases the computational complexity of the algorithm, its implementation resulted in a low-bit rate speech coder, which improved speech variability and speaker rate. The super-wavelet was an extension of the adaptive wavelet technique. It addresses speech coding of unvoiced speech. The MRST technique used the wavelet concept be applying QMF filter banks to the original signal. Good spectral resolution was achieved using this method, but speech variability was compromised. The quadrature spline technique requires a priori knowledge of a mathematical model of speech production to measure voiced events. It is assumed that a linear model is used but this limits that speech accuracy of the original signal. However, this method is efficient in discriminating impulses in the original signal.

Future work will investigate other wavelet techniques that can be used to overcome some of the deficiencies in the methods presented. For example, the derivation of a unique mother wavelet may provide good accuracy and speaker variability and capture the essence of speech. Also, hardware implementations will also be studied to implement practical wavelet based speech recognition systems.

5. REFERENCES

1. S. Kadambe and P. Srinivasan. “Applications of Adaptive Wavelets for Speech Coding”. Proceedings of IEEE Signal Processing, International Symposium on Time-Frequency and Time-Scale Analysis. pp. 623-635. October 1994.

2. H. Szu, B. Tefler and S. Kadambe, “Neural Network Adaptive Wavelets for Signal Representation and Classification”. Journal of Optical Engineering, Volume 31, pg. 1907-1916. September 1992.

3. S. Kadambe, P. Srinivasan, B. Telfer and H. Szu. “Representation and classification of Unvoiced sounds using Adaptive Wavelets”. Proceedings of IEEE, ICASSP Volume 5, pp. 3417-3420. May 1991.

4. D. Anderson. ”Speech Analysis and Coding Using a Multi-Resolution Sinusoidal Transform”. Proceedings of IEEE, ICASSP, pp. 1037-1040, 1996.

5. S. Griebel and M. Brandstein. ”WaveletTransform Extrema Clustering for Multi-channel Speech Dereverberation”. IEEE Workshop on Acoustic Echo and Noise Control. , pp. 52-55, September 1999.

6. Application Notes, “Xilinx Virtex Series: Redefining FPGAs”. Xilinx Corporation. www.xilinx.com.

7. Product Overview, “Virtex Series FPGAs: System Timing”. Xilinx Corporation. www.xilinx.com.

8. S. Greenberg. “On the Origins of Speech Intelligibility in the Real World”. Proceedings of the ESCA Workshop on Robust Speech Recognition for Unknown Communication Channels. April 1997. pg. 23-52.

9. S. Griebel. ”Multi-Channel Wavelet Techniques for Reverberant Speech Analysis and Enhancement”. Technical Report. Harvard University – Division of Engineering and Applied Sciences. 1999.

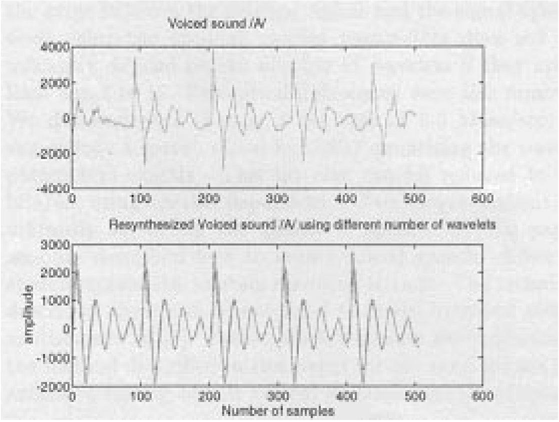


Figure 2-Original speech signal and resynthesized signal [1].

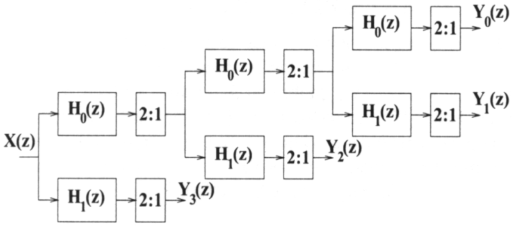


Figure 3-QMF analysis filter bank used in MRST [4].

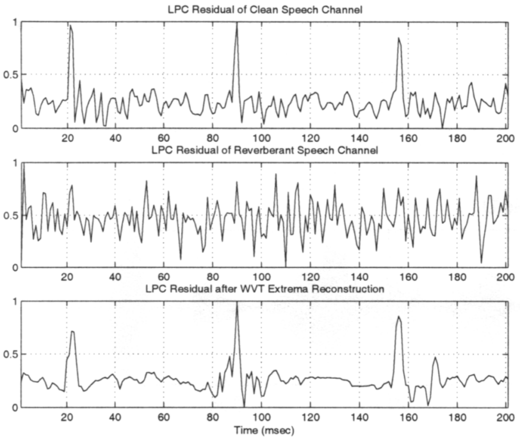


Figure 4-Reconstructed speech using quadrature spline wavelet [5].