Chapter 19
Amplification of Signal Features Using Variance Fractal Dimension Trajectory

Witold Kinsner
University of Manitoba, Canada

Warren Grieder
University of Manitoba, Canada

ABSTRACT

This paper describes how the selection of parameters for the variance fractal dimension (VFD) multiscale time-domain algorithm can create an amplification of the fractal dimension trajectory that is obtained for a natural-speech waveform in the presence of ambient noise. The technique is based on the variance fractal dimension trajectory (VFDT) algorithm that is used not only to detect the external boundaries of an utterance, but also its internal pauses representing the unvoiced speech. The VFDT algorithm can also amplify internal features of phonemes. This fractal feature amplification is accomplished when the time increments are selected in a dyadic manner rather than selecting the increments in a unit distance sequence. These amplified trajectories for different phonemes are more distinct, thus providing a better characterization of the individual segments in the speech signal. This approach is superior to other energy-based boundary-detection techniques. Observations are based on extensive experimental results on speech utterances digitized at 44.1 kilosamples per second, with 16 bits in each sample.

INTRODUCTION

This work is motivated by the need for better human-machine interaction. We have reached a point at which our ability to create large computing systems and information has exceeded our ability to maintain and manage them manually. The developments towards cognitive machines and systems are intended to increase utilization of resources, as well as to ease their maintenance and their human-centric interactions (Kinsner, 2007, January). The semantic Web is intended to provide applications to alleviate the problem of manual keyword searches and blog surfing

DOI: 10.4018/978-1-4666-1743-8.ch019

through natural-language and semantic tools to determine the required content from any textual material through cognitive informatics (Wang, 2002) and other new developments in intelligent signal processing (Haykin, Principe, Sejnowski, & McWhirter, 2007), (Kinsner, 2007, April). However, non-textual materials such as non-annotated music, sounds, images, video, mathematical expressions, and chemical formulae are still largely inaccessible. More importantly, the vast amount of recorded acoustic speech and voices in their analog form (e.g., magnetic recordings) and digital form (uncompressed or compressed recordings) is largely inaccessible for browsing or real-time automatic translation at this stage.

This paper provides a small, but significant step towards not only an automatic speech recognition and translation into text, but also speech browsability (searching for specific phrases) and speech understanding. The technique described here is based on a multiscale information-theoretic approach to dealing with band-limited and broadband signals contaminated by noise. This approach differs fundamentally from the traditional energy-based approaches.

One major difficulty in the automatic recognition of speech by a machine is the unlimited number of possible utterances in a given language. This is further compounded by the non-stationary and variable nature of speech, where even two utterances of the same phrase, while perceptually identical to a human listener, contain several differences that make template matching of complete phrases impractical due to the extensive memory and search time requirements for a system with a large vocabulary. Thus, it is necessary to segment utterances into smaller units such as words, syllables, or phonemes. When the size of the segmentation units is reduced, there are fewer possible matching patterns, thus reducing both the size of the search for matching elements and the amount of storage required for pattern templates. Although still large on the word and syllable level, pattern matching becomes more manageable on the phonemic level.

Many speech pre-processing techniques, which utilize various parameterization of speech, have been developed that can perform speech recognition (Bristow, 1986) with varying degrees of success. Among the parameterizations used are linear predictive coding (LPC) with dynamic time warping (Parsons, 1987; Rabiner & Schafer, 1978), used in formant tracking techniques, and fast Fourier transforms, used in the neural phonetic typewriter (Kohonen, 1988). Many segmentation and recognition techniques use wavelets (Mallat, 1998; Wornell, 1996), fractals (Al-Akaidi, 2004; Kinsner, 2007, October), higher-order statistics (Haykin, Principe, Sejnowski, & McWhirter, 2007), neuro-fuzzy classifications techniques (Haykin & Kosko, 2001, Ch. 5; Bishop, 2006). With a more accurate identification scheme of the segmentation boundaries, further analysis using these and other techniques on the speech segments may result in increased performance.

In a previous paper (Kinsner & Grieder, 2008), we have shown how the variance fractal dimension (VFD) algorithm can be used to calculate a dimension value of the speech samples contained in a window within the speech utterance. By shifting this window to different positions spaced at regular intervals along the utterance and calculating the dimension for each window, a trajectory for the speech in the variance fractal dimension domain can be obtained, which can then be analyzed to provide a segmentation of the utterance. This variance fractal dimension trajectory (VFDT) follows a path that is determined by the characteristics of the utterance, with each phoneme exhibiting a characteristic trajectory pattern and transitions between adjacent phonemes generally indicated by a change in the trajectory. The beginning and end of an utterance can be determined by transitions between silence, which contains low level noise with a high dimension, and the more correlated speech waveform, which has a lower dimension. Most importantly, transitions