IMPORTANCE OF PROSODIC FEATURES IN LANGUAGE IDENTIFICATION

Niraj Kr. Singh
Research Scholar (M.Tech’)
Vivekananda Global University, Jaipur
Jaipur, India
Connect2vit@gmail.com

Anoop Singh Poonia
Professor, E & C Engineering
Vivekananda Global University, Jaipur
Jaipur, India
anoopsingh_25@rediffmail.com

Abstract: Earlier Researches in LID systems found that inclusion of prosodic features (alike, speech rate, Fundamental Frequency and Syllable timing) offered a little to develop the performance of their systems. Focused study on the utility of prosodic feature is attempted to evaluate parameters to behold the fundamental frequency and amplitude contours on syllable – by-syllable bases. The timing relationship is computed from the F0 whereas histogram of features (feature pairs) are collected through amplitude information followed by computations of loglikelihood ratio function for evaluating the unknown utterance (word/speech) in a pair wise discrimination task. The consequence reflects prosodic parameters to be useful in discriminating languages. Recent works in LID using prosody have proven appreciating results with better accuracy. This paper describes the importance of prosodic features in comparison with other possible features and methods for language identification.

Keywords: Prosody, Utterance, Modeling, Expectation Maximization, GMM.

I. INTRODUCTION

In the discussion of related work, we focus on previous work in sign language recognition. For coverage of gesture recognition, the survey in [24] is an excellent starting point. Other, more recent work is reviewed in [35]. Much previous work has focused on isolated sign language recognition with clear pauses after each sign, although the research focus is slowly shifting to continuous recognition. These pauses make it a much easier problem than continuous recognition without pauses between the individual signs, because explicit segmentation of a continuous input stream into the individual signs is very difficult. For this reason, and because of coarticulation effects, work on isolated recognition often does not generalize easily to continuous recognition.

Erensthteyn and colleagues used neural networks to recognize finger spelling [6]. Waldron and Kim also used neural networks, but they attempted to recognize a small set of isolated signs [34] instead of finger spelling. They used Stokoe’s transcription system [29] to separate the
handshape, orientation, and movement aspects of the signs. Kadous used Power Gloves to recognize a set of 95 isolated Auslan signs with 80% accuracy, with an emphasis on computationally inexpensive methods [13]. Grobel and Assam used HMMs to recognize isolated signs with 91.3% accuracy out of a 262-sign vocabulary. They extracted 2D features from video recordings of signers wearing colored gloves [9]. Braffort described ARGo, an architecture for recognizing French Sign Language. It attempted to integrate the normally disparate fields of sign language recognition and understanding [2]. Toward this goal, Gibet and colleagues also described a corpus of 3D gestural and sign language movement primitives [8]. This work focused on the syntactic and semantic aspects of sign languages, rather than phonology. Most work on continuous sign language recognition is based on HMMs, which offer the advantage of being able to segment a data stream into its constituent signs implicitly. It thus bypasses the difficult problem of segmentation entirely.

II. BACKGROUND

Speech recognition technology made major strides in the 1970s, thanks to interest and funding from the U.S. Department of Defense. The DoD’s DARPA Speech Understanding Research (SUR) program, from 1971 to 1976, was one of the largest of its kind in the history of speech recognition, and among other things it was responsible for Carnegie Mellon’s "Harpy" speech-understanding system. Harpy could understand 1011 words, approximately the vocabulary of an average three-year-old. Over the next decade, i.e. 1980s, a new approaches to understanding what people say, speech recognition vocabulary jumped from about a few hundred words to several thousand words, and had the potential to recognize an unlimited number of words. One major reason was a new statistical method known as the HMM (Hidden Markov Model). As the time scope was limited and to be able to focus on more specific issues than HMM in general, the Hidden Markov Model toolkit (HTK) was used. HTK is a toolkit for building Hidden Markov Models (HMMs). HMMs can be used to model any time series and the core of HTK is similarly general-purpose.

III. PROSODY FOR LID

Speech is one of the oldest and most natural means of information exchange between human beings. We as humans speak and listen to each other in human-human interface. For centuries people have tried to develop machines that can understand and produce speech as humans do so naturally (Pinker, 1994 [20]; Deshmukh et al., 1999 [5]). Obviously such an interface would yield great benefits (Kandasamy, 1995) [12]. Attempts have been made to develop vocally interactive computers to realise voice/speech recognition. In this case a computer can recognize text and give out a speech output (Kandasamy, 1995) [12]. Speech recognition can be defined as the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words (Zue et al., 1996 [36]; Mengjie, 2001 [17]).Automatic speech recognition (ASR) is one of the fastest developing fields in the
framework of speech science and engineering. As the new generation of computing technology, it comes as the next major innovation in man-machine interaction, after functionality of text-to-speech (TTS), supporting interactive voice response (IVR) systems.

IV. DISCUSSIONS

The implementation purposes the following methods were taken into practice: (a) Building the task grammar, (b) Constructing a dictionary for the models (c) Recording the data (d) Creating transcription files for training data (e) Encoding the data (feature processing) (f) (Re-) training the acoustic models (g) Evaluating the recognizers against the test data (h) Reporting recognition results. These are exhaustive representation of a LID system and do require more robustness for better performance.

REFERENCES


