Chapter 7
Voice Biometric for Learner Authentication: Biometric Authentication

Kamaraj A.
Mepco Schlenk Engineering College, India

Selva Nidhyananthan S
Mepco Schlenk Engineering College, India

Kalyana Sundaram C.
Mepco Schlenk Engineering College, India

ABSTRACT

The objective of this chapter is to verify the identity of the claimed learner by extracting the prosodic features of the speech signal. TIMIT Acoustic-Phonetic Continuous Speech Corpus is used for learner verification using prosodic and articulation features such as energy, pitch, and formants. The prosodic feature includes pitch (F0), and articulation feature includes formants (F1-F7). From this database, for this project in the training phase, 200 learners were used and in the testing phase 160 learners were used. The pitch and formants were extracted using linear predictive analysis. The first seven formants were used for verification purpose. The feature set consists of eight features. The features are fed into the Gaussian mixture model. In the Gaussian mixture model, parameters are estimated from the training and testing data using the iterative expectation-maximization. Log likelihood score is computed using these parameters, and then these scores are normalized to make decisions. The decision is made based on the threshold.

DOI: 10.4018/978-1-5225-7724-9.ch007
INTRODUCTION

Learner verification system must decide if a Learner is the person he/she claims to be. The person wishing to be verified first enters his claimed identity, and then on request from the verification system utters his verification phrase, and requests some decisions to be made in the event he is verified. The spoken utterance is processed to obtain a pattern which is compared to the stored reference pattern and then based on decision, accept or reject the claimed Learner. The major advantage of Learner verification system is voice mail, telephone banking, security purpose, authentication etc. The most important application is the forensics.

Speech Prosody

In linguistics, prosody is the rhythm, stress, and intonation of speech. Prosody may reflect various features of the learner or the utterance: the emotional state of the learner; the form of the utterance (statement, question, or command); the presence of irony or sarcasm; emphasis, contrast, and focus; or other elements of language that may not be encoded by grammar or choice of vocabulary.

In terms of acoustics, the prosodics of oral languages involve variation in syllable length, loudness and pitch. In sign languages, prosody involves the rhythm, length, and tension of gestures, along with mouthing and facial expressions. Prosody is typically absent in writing, which can occasionally result in reader misunderstanding. Orthographic conventions to mark or substitute for prosody include punctuation (commas, exclamation marks, question marks, scare quotes, and ellipses), and typographic styling for emphasis.

The details of a language’s prosody depend upon its phonology. For instance, in a language with phonemic vowel length, this must be marked separately from prosodic syllable length. In similar manner, prosodic pitch must not obscure tone in a tone language if the result is to be intelligible. Although tone languages such as Mandarin have prosodic pitch variations in the course of a sentence, such variations are long and smooth contours, on which the short and sharp lexical tones are superimposed. If pitch can be compared to ocean waves, the swells are the prosody, and the wind-blown ripples in their surface are the lexical tones, as with stress in English. The word dessert has greater stress on the second syllable, compared to the noun desert, which has greater stress on the first (in its “arid land” meaning, but not in its “thing which is deserved” meaning); but this distinction is not obscured when the entire word is stressed by a child demanding “Give me dessert!” Vowels in many languages are likewise pronounced differently (typically less centrally) in a careful rhythm or when a word is emphasized, but not so much as to overlap with the formant structure of a different vowel. Both lexical and prosodic information are encoded in rhythm, loudness, pitch, and vowel formants.

Prosodic features are supra segmental. They are not confined to any one segment, but occur in some higher level of an utterance. These prosodic units are the actual phonetic “spurts”, or chunks of speech. They need not correspond to grammatical units such as phrases and clauses, though they may; and these facts suggest insights into how the brain processes speech.

Prosodic units are marked by phonetic cues. Phonetic cues can include aspects of prosody such as pitch, pauses, and accents, all of which are cues that must be analyzed in context, or in comparison to other aspects of a sentence. Pitch, for example, can change over the course of a sentence. In English, falling intonation indicates a declarative statement while rising intonation indicates an interrogative statement. Pauses are important prosodic units because they can often indicate breaks in a thought and can also
Related Content

An Evaluation of Neurogames®: A Collection of Computer Games Designed to Improve Literacy and Numeracy
[www.igi-global.com/article/evaluation-neurogames-collection-computer-games/53859?camid=4v1a](www.igi-global.com/article/evaluation-neurogames-collection-computer-games/53859?camid=4v1a)

Using a Classification of Psychological Experience in Social-Networking Sites as a Virtual Learning Environment
[www.igi-global.com/article/using-classification-psychological-experience-social/74839?camid=4v1a](www.igi-global.com/article/using-classification-psychological-experience-social/74839?camid=4v1a)

Virtual Learning Environment
[www.igi-global.com/chapter/virtual-learning-environment/63115?camid=4v1a](www.igi-global.com/chapter/virtual-learning-environment/63115?camid=4v1a)

Online Learning: Guidelines for Team Effectiveness
[www.igi-global.com/chapter/online-learning/100595?camid=4v1a](www.igi-global.com/chapter/online-learning/100595?camid=4v1a)