Prosody Based Speech Analysis

(No. T4-1856)

Principal investigator

Tirza Biron

Physics
Department of Physics of Complex Systems

Principal investigator

Elisha Moses

Faculty of Physics
Department of Physics of Complex Systems

Principal investigator

David Harel

Faculty of Mathematics and Computer Science
Department of Computer Science and Applied Mathematics

Overview

A major part of human communication is conveyed through intonation, or prosody â€“ the music of speech. It is the least conscious, most instinctive activity of speech conveying all non-verbal linguistic information including sentiment, emphasis, conversation action and a large variety of signals: whether we are happy, irritated or even truthful can be detected through the way in which we utter the words. Remarkably, this aspect of communication has not yet been analyzed in a comprehensive, quantitative manner. A group of researchers from the Weizmann Institute of Science has created a model for analyzing prosodic patterns that enables access to non-verbal communication in speech, based on computational analysis of intonation-unit modulation. The model relies on a segmentation method that adds a compact layer of basic algorithmic analysis within a speech recognition engine, adding very little computational load. This provides the basis for universally describing prosody and grants access to non-verbal information in speech.

The Need

The study of prosody is largely absent from the rapidly-evolving field of automatic speech processing applications. Current approaches to computerized analysis of spoken text are generally heuristic and struggle for robust means of contextualization. They rely heavily on narrow semantic searches, resulting in a very partial interpretation of language and text. Specifically, technologies for automatic recognition of spoken intent still depend on vocabulary and key expressions (e.g. "ridiculous" as a dissatisfaction marker). Furthermore, they are unable to parse conversation correctly - to the point of failing to recognize where commas, question marks or periods occur. These extensively used technologies (including speech-related man-machine interfaces, speechrecognition engines and auto-translation) would be revolutionized using a computerized linguistic analysis technology that provides information regarding the context of the communication, the attitude of the speaker, emphasis and other contextual information.

The Solution

A new linguistic model for prosody analysis, based on the automatic classification of intonation units.

Technology Essence

The technology developed by the group of researchers requires a multidisciplinary approach, combining linguistic theory, signal processing, and artificial neural network and deep-learning methodologies. Deep machine learning algorithms for pattern recognition are combined with Saussurian pattern analysis and contextualization methodologies, for the detection of prosodic patterns and, in time, their syntactic constellations. Identification of prosodic audio features is essentially done through voice pitch, intensity, rhythm, timbre and voice quality. The features are then assigned to their contextualized "meaning" by comparison to a database, or prosodic â€œdictionaryâ€. The end result can be then compressed and presented in multiple formats including xml or json.

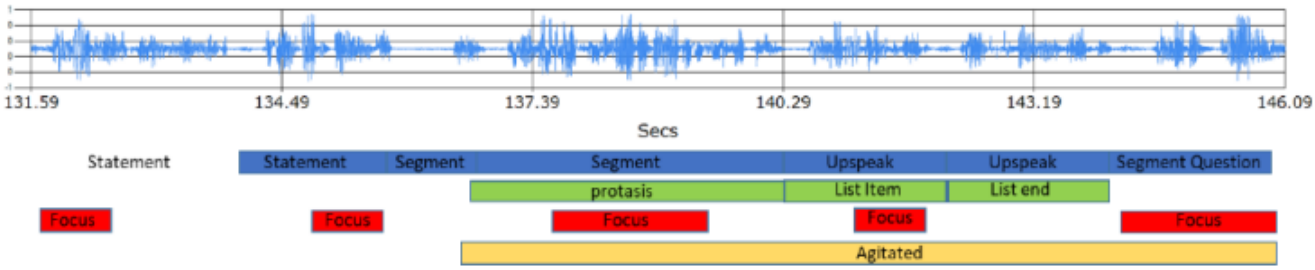

Figure 1: An illustration of implementation of automatic classification of speech segments in a recorded sentence. A multi-layered model: each segment is described by multiple attributes.

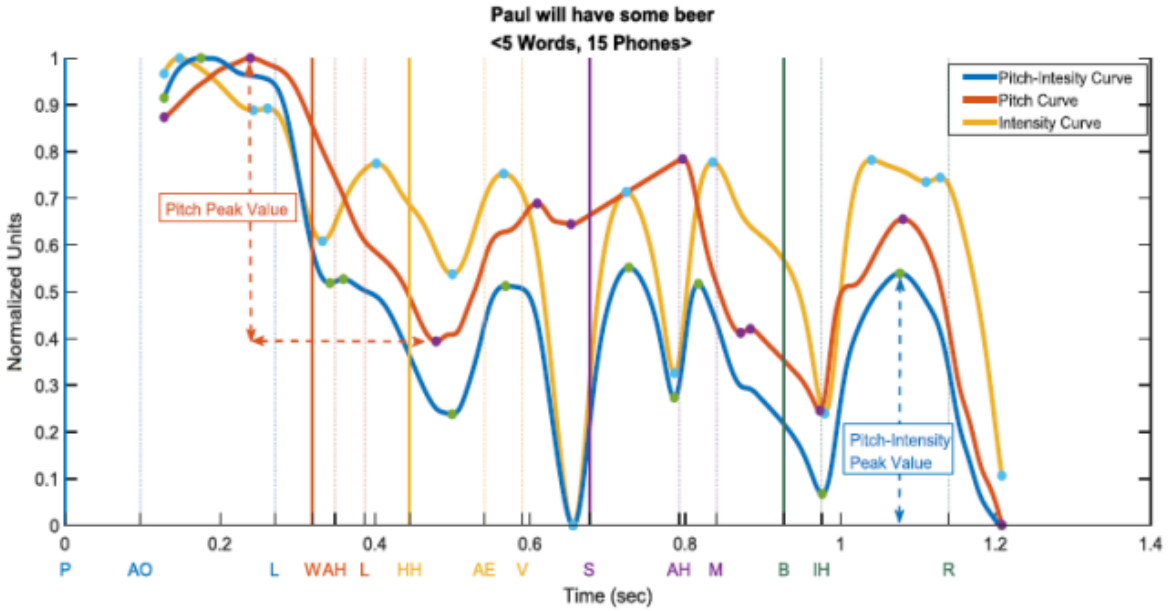

Figure 2: Examples of peak values in pitch and "intensity-pitch" curves. Interpolated pitch curve (red), intensity curve (yellow). Curve extrema are marked. The "intensity-pitch" curve (blue) is estimated by multiplying the scaled pitch and intensity. Peak values measure the distance between the maxima and the nearby minima.

Applications and Advantages

Advantages

- Compact segmentation algorithm
- Easy integration and implementation into existing speech recognition and speech analysis technologies
- Universal model for language-specific prosodic dictionary
- Up to 90% accuracy in identifying emphasized words in a sentenc

Â

Applications

- Improved human-machine interactions (e.g., personal assistants, in-car speech recognition devices, etc.)Â
- Upgraded speech synthesis systemsÂ
- Improved automatic translationÂ
- Upgraded NLP syntactic processingÂ
- Speech intonation recognition devices for people with autismÂ
- Improved mass data mining (interception purposes)

Development Status

The researchers have developed an algorithm for the detection of prosodic unit boundary, a strong emphasis detection algorithm and an analytical scheme. In addition, they are in the process of creating a "prosodic dictionary", containing a database and pattern search and currently applicable to audio files/texts. Examples of current progress are an automated classification of yes/no vs. wh- questions, advanced quantification of parentheticals, and quantification of the difference between "up-speak" and questions. Further developments are planned to make this invention applicable to written texts and to combine it with speech recognition and speech analytics, enabling a more comprehensive assessment of language and text. The invention is patent-protected.

Market Opportunity

In 2018, the global speech and voice recognition market size was valued at approximately $7 billion and is projected to reach approximately $28 billion by 2026, exhibiting a CAGR of nearly 20%. A technology that will supplement existing speech recognition technologies and enable extraction of speaker intentions is expected to increase technology value and be integratable in multiple fields, including:

- Speech recognition
- Next-generation call centers
- Intelligence and cyber systems
- Transcription technologies
- Speech-generating devices

Patent Status

USA Granted: 11,600,264