
IMPLEMENTATION OF DEAL FOR ANGER DETECTION THROUGH STRESS ON SPEECH USING BAYES NAÏVE CLASSIFIER

Vikas Jain

Assistant Professor

Dept. of Computer Application

C.C.S. University, Meerut

ABSTRACT

Recent approaches to text classification have used two different first-order probabilistic models for classification, both of which make the naive Bayes assumption. Some use a multi-variate Bernoulli model, that is, a Bayesian Network with no dependencies between words and binary word features. Others use a multinomial model, that is, a uni-gram language model with integer word counts. More than a decade has passed since research on automatic recognition of emotion from speech has become a new field of research in line with its 'big brothers' speech and speaker recognition. Speech is the primary mode of conveying the message in human-human communication. It also contains paralinguistic information like speaker, stress, acoustic environment, person's intention, language, accent and dialect. Psychological research on human emotions claimed that there are six basic expressions of emotions that are universally displayed and expressed: happiness, anger, sadness, surprise, disgust and fear. This implies that these basic emotions are probably displayed and recognized cross culturally and speaker-independently.

KEYWORDS: *Naive Bayes, stress*

INTRODUCTION

Automatic recognition of human emotion (anger) in speech aims at recognizing the underlying emotional state of a speaker from the speech signal. The area has received rapidly increasing research interest over the past few years. However, designing powerful spectral features for high-performance speech emotion recognition (SER) remains an open challenge. Our purpose was to investigate whether emotion (anger), as perceived by a panel of listeners, were observable in various acoustic cues of the speech signal. The cues were chosen, by examining earlier studies on the same subject and they were: the syllable rate, the minimum, maximum, median and mean of the pitch, the amplitude and the first six formants, % jitter and %shimmer. Application is first trained on person voice will provide signal if detected as angry, and a person, who is signaled that his/her manner of speaking is classified as angry, becomes aware of his mental state and could regulate his way of expressing his thoughts. It is of great advantage in this situation if the corrective comes from a machine that does not play a part in the situation and shows no emotions in itself.

Speech emotion analysis refers to the use of various methods to analyze vocal behavior as a marker of affect, focusing on the nonverbal aspects of speech. The basic assumption is that there is a set of objectively measurable voice parameters that reflect the affective state a person is currently experiencing. This assumption appears reasonable given that most affective states involve physiological reactions, which in turn modify different aspects of the voice production process. For example, the sympathetic arousal associated with an anger state often produce changes in respiration and an increase in muscle tension, which influence the vibration of the vocal folds and vocal tract shape, affecting the acoustic characteristics of the speech, which in turn can be used by the listener to infer the respective state.

BASIC CONCEPTS OF SOUND PRODUCTION**Speech Organs**

Produce the many sounds needed for language. Organs used include the lips, teeth, tongue, alveolar ridge, hard palate, velum (soft palate), uvula and glottis.

Speech organs—or *articulators*—are of two types:

Passive articulators:

Passive articulators remain static during the articulation of sound. Upper lips, teeth, alveolar ridge, hard palate, soft palate, uvula, and pharynx wall are passive articulators.

Active articulators:

Move relative to these passive articulators to produce various speech sounds, in different manners. The most important active articulator is the tongue. The lower lip and glottis are other active articulators

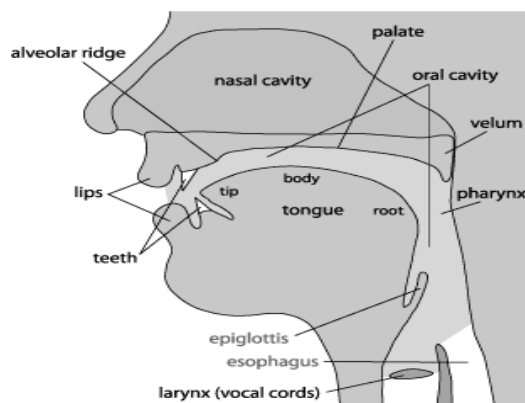


Fig.-1: Speech Articulators

VOCAL FOLDS

The vocal folds, also known popularly as vocal cords, are composed of twin infolding of mucous membrane stretched horizontally across the larynx. They vibrate, modulating the flow of air being expelled from the lungs during phonation. Another name for the airway at the level of the vocal cords is the glottis, and the opening between the cords is called the glottic chink. The size of the glottic chink is important in respiration and phonation.

CHALLENGES IN EMOTION DETECTION THROUGH SPEECH IN COMPUTING

In the growing range of interactive interfaces, the research of emotional voice is still at an early stage, not to mention a paucity of literatures on real applications. To make the mainstream use of speech emotion recognition a reality, the industry must deliver robust and high-recognition accuracy close to human-like. Although progress has been impressive, there are still several hurdles that automatic emotion recognition from speech. Here we are going to put eye on technical challenges in automatic emotion recognition through speech.



Fig.-2: Image of Normal Vocal Cord, courtesy of the Milton J. Dance

ARTICULATION

Articulation refers to the production of the speech sounds. Accurate articulation involves precise movement of the articulators including the tongue, lips, alveolar ridge, velum, and jaw coordinated with correct air flow and voicing

PLACE OF ARTICULATION

- Refers to the relative positions of the lips, teeth and tongue.
- There are six distinct types of classification bilabial, labiodental, interdental, alveolar, alveo-palatal, and velar.
- The six places of articulation describe the parts of the vocal tract which are responsible for the obstruction of the air flow from the lungs the degree of obstruction the airstream incurs must also be considered.

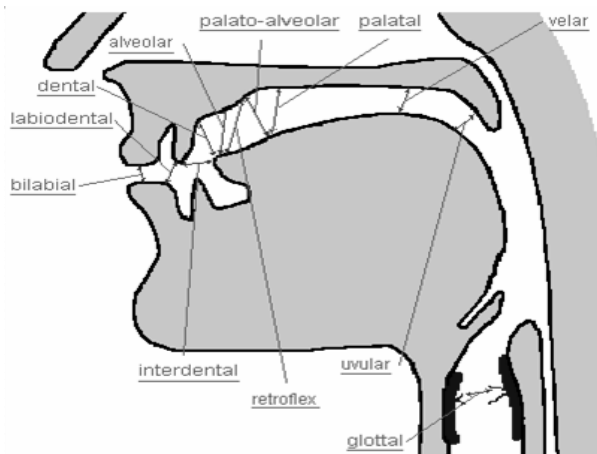


Fig.-3: The Places of Articulation

EFFECTS OF EMOTION ON HUMAN VOCAL SYSTEM

There are many situations where people perceive stress and emotion. These include heavy workload, adverse environment and social problems. Stress has impact on the body as well as on the mind of the person and this in turn affects the vocal system. There are physiological reasons for the acoustic cues changing whenever a speaker changes his or her emotional state. A Study done by Kienast at the University of Berlin examined the spectral and segmental changes caused by the articulatory behaviour of a person feeling an emotion (Kienast et al., 2000).

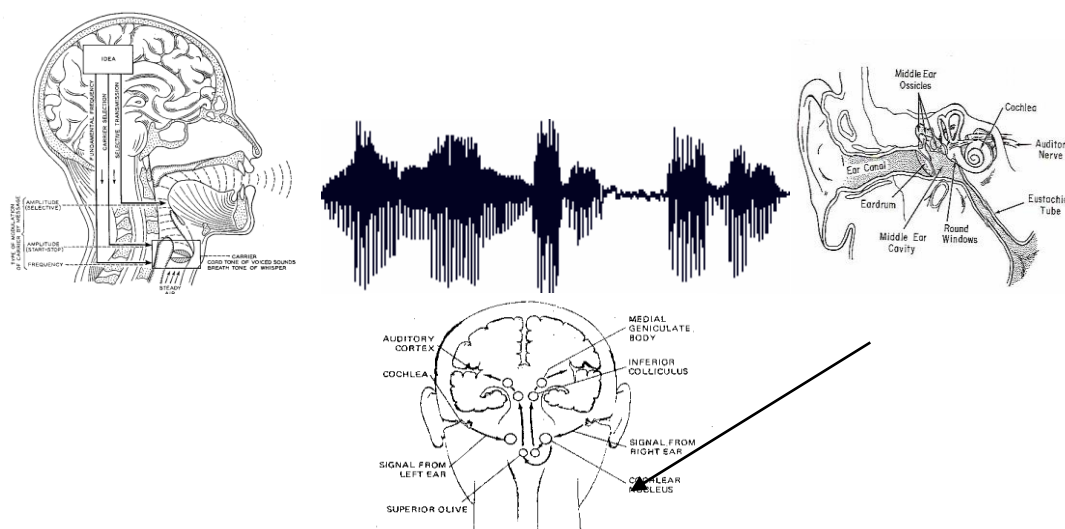


Fig.-4: Speech Production and Perception

When people are undergoing under any kind of negative emotion such as anger or stress, their bodily resources are automatically changed to prepare an attack or to run away from danger. If the situation persists, considerable strain may be placed on the body and affects a person’s ability to perform including producing speech. Stress is observed even in positively toned emotions. For example, Anger, Anxiety, Guilt and Sadness are regarded as stressed emotions.

Positive emotions of Joy, Pride and Love are also frequently associated with stress. For example, when people are in happy mood, they may fear that the favorable conditions provoking their happiness will end. The research studies that have emphasized especially only on psychological, biological, and linguistic aspects of several emotional states can be found in. From the psychological perspective, of particular interest is the *cause-and-effect* of emotion. The activation-evaluation space provides a simple approach in understanding emotions. In a nutshell, it considers the stimulus that excites the emotion, the cognition ability of the agent to appraise the nature of the stimulus and subsequently his/her mental and physical responses to the stimuli. The mental response is in the form of emotional state.

Thus, emotion has a broad sense and a narrow sense effect. The broad sense reflects the underlying long-term emotion and the narrow sense refers to the short-term excitation of the mind that prompts people to action. In automatic recognition of emotion, a machine will not distinguish if the emotional states are due to 15 long-term or short-term effects so long as it is reflected in the speech or facial expression. From the perspective of physiology in the production of speech. As a result, heart rate and blood pressure increase, the mouth becomes dry and there are occasional muscle tremors. On the other hand, with the arousal of the parasympathetic nervous system, as with Sadness, heart rate and blood pressure decrease and salivation increases, producing slow speech.

The corresponding effects on speech of such physiological changes thus show up vocal system modifications and affect the quality and characteristics of the utterances. The acoustic characteristics that are altered during stressed and emotional speech production are studied in the following section.

CONCLUSIONS

System for anger recognition and classification is proposed. Evaluations that concentrate in identifying the effect of anger on vocal system are carried out. It is found that the characteristics of speech utterances are altered when producing stress or emotion. From this knowledge, the best acoustic features that are important for stress and emotion (Anger) detection are selected from several traditional features. The features such as pitch, amplitude, spectral distribution and speaking rate parameters function as basic acoustic parameters to characterize emotion.

REFERENCES

1. Allen, N, 2013. "Automatic Emotion Recognition in Speech Signal Using Teager Energy Operator and MFCC Features", In Jianhong Zhou (ed.) Proceedings of the 3rd International Conference on Computer Technology and Development (ICCTD 2011) Volume 1, United States, 25-27 November, 2011, pp. 2315-2320.
2. Hansen, J.H.L, 2014. "Morphological Constrained Feature Enhancement with Adaptive Cepstral Compensation for Speech Recognition in Noise and Lombard Effect", IEEE Transactions on Speech Audio Processing, Vol. 2, no. 4, pp. 598-614, Oct. 1.
3. Heinz J., 2017, "English Phonology: An introduction", in Cambridge University Press, ISBN-10: 0521336031 | ISBN-13: 978-0521336031.
4. J.H.L. Hansen, 2012. "Speech Under Stress Conditions: Overview of the Effect of Speech Production and on System Performance", IEEE ICASSP-99, Inter. Conf. on Acoustics, Speech, and Signal Processing, Vol. 4, pp. 2079-2082, Phoenix, Arizona, March.
5. J.N. Gowdy, 2016. "Subband Based Classification of Speech Under Stress", Proceedings of the IEEE Inter. Conf. on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 569 -572.
6. Lazarus, R.S., 2010. "Emotion Adaptation". In Annual review of psychology, New York: Oxford Univ. Press. (pp. 1-21).
7. Nicholson, J., K. Takahashi and R. Nakatsu, 2013. "Emotion Recognition in Speech Using Neural Networks", 6th Inter. Conf. on Neural Information Processing, ICONIP '99, Vol 2, pp. 495 .501.
8. Tin Lay Nwea, Say Wei Foob, Liyanage C. De Silvaa, 2003. "Speech emotion recognition using hidden Markov models", in Elsevier Speech Communication Journal ,Vol. 41, Issue 4, pp. 603-623, November.
9. Yongming HUANG, Guobao ZHANG, Feipeng DA, 2017. "Speech Emotion Recognition with New Spectral Features and Its Application to Pet Robot", In: Journal of Computational Information Systems 7: 13 (2011) 4915-4922.