IOSR Journal of Computer Engineering (IOSRJCE)
ISSN: 2278-0661, ISBN: 2278-8727 Volume 5, Issue 4(Sep-Oct. 2012), PP 10-15
www.iosrjournals.org

Impact of Emotion on Prosody Analysis

Padmalaya Pattnaik
(Dept. of Information Technology, C.V. Raman College of Engineering, Bhubaneswar, Odisha, India)

Abstract: *Speech can be described as an act of producing voice through the use of the vocal folds and vocal apparatus to create a linguistic act designed to convey information. Linguists classify the speech sounds used in a language into a number of abstract categories called phonemes. Phonemes are abstract categories, which allow us to group together subsets of speech sounds. Speech signals carry different features, which need detailed study across gender for making a standard database of different linguistic, & paralinguistic factors. When people interact with others they convey emotions. Emotions play a vital role in any kind of decision in affective, social or business area. The emotions are manifested in verbal, facial expressions but also in written texts. The objective of this study is to verify the impact of various emotional states on speech prosody analysis.*

Keywords: *Duration, Emotion, Jitter, Prosody, Shimmer*

I. Introduction

To make Information Technology (IT) relevant to rural India, voice access to a variety of computer based services is imperative. Although many speech interfaces are already available, the need is for speech interfaces in local Indian languages. Application specific Indian language speech recognition systems are required to make computer aided teaching, a reality in rural schools. This paper presents the preliminary work done to demonstrate the relevance of an Oriya Continuous Speech Recognition System in primary education. Automatic speech recognition has progressed tremendously in the last two decades. There are several commercial Automatic Speech Recognition (ASR) systems developed, the most popular among them are Dragon Naturally Speaking, IBM Via voice and Microsoft SAPI. Speech is a complex waveform containing verbal (e.g. phoneme, syllable, and word) and nonverbal (e.g. speaker identity, emotional state, and tone) information. Both the verbal and nonverbal aspects of speech are extremely important in interpersonal communication and human-machine interaction Each spoken word is created using the phonetic combination of a set of vowel semivowel and consonant speech sound units. Different stress is applied by vocal cord of a person for particular emotion.

Stress is a psycho-physiological state which is characterized by subjective strain, dysfunctional physiological activity of the speech signal. Stress may be induced by external factors such as workload, noise, vibration, sleeps loss, etc. and by internal factors like emotion, fatigue, etc. Physiological consequences of stress are respiratory changes. This respiratory change can be increased respiration rate, irregular breathing, increased muscle tension of the vocal cords, etc. The increased muscle tension of the vocal cords and vocal tract can directly or indirectly and adversely affect the quality of speech.

We use emotions to express and communicate our feelings in everyday life. Our experience as speakers as well as listeners tells us that the interpretation of meaning or intention of a spoken utterance can be affected by the emotions that area expressed and felt. With recent advances in man-machine communication technologies, through automatic spoken dialog management systems, the question of how human emotion is encoded by a speaker and decoded by a listener has now attained practical importance. It is expected that by being able to detect emotion in user's speech and inject emotion into automatically generated system response depending on the dialogue situation.

II. Literature

An emotion is a mental and physiological state associated with a wide variety of feelings, thoughts, and internal (physical) or external (social) behaviors. Love, hate, courage, fear, joy, sadness, pleasure and disgust can all be described in both psychological and physiological terms. An emotion is a psychological arousal with cognitive aspects that depends on the specific context. According to some researcher, the emotions are cognitive processes. Emotion is a process in which the perception of a certain set of stimuli, follows cognitive assessment which enables people to label and identify a particular emotional state. At this point there will be an emotional physiological, behavioral and expressive response. For example, the primordial fear, that alerts us as soon when we hear a sudden noise, allows to react to, dangerous situations and provides instantly resources to face them as escape or close the door. The emotional stimuli may be an event, a

scene, a face, a poster, an advertising campaign. These events, as a first reaction, put on alert the organism with somatic changes as heart rate, increase of sweat, acceleration of respiratory rhythm, rise of muscle tensions.

Emotions give an immediate response that often don't use cognitive processes and conscious elaboration and sometimes they have an effect on cognitive aspects as concentration ability, confusion, loss, alert and so on. This is what is asserted in evaluation theory, in which cognitive appraisal is the true cause of emotions [2]. Two factors that emerge permanently are those related to signals of pleasure and pain and characterizing respectively the positive and negative emotions. It's clear that these two parameters alone are not sufficient to characterize the different emotions. Many authors debate on primary and secondary emotions other on pure and mixed emotions, leaving the implication that emotions can somehow be composed or added.

The systems based on the analysis of physiological response as blood pressure, heart rate, respiration change present an initial phase where the signals are collected in configurations to be correlated with different emotional states and a subsequently recognition basing on the measure of indicators. One of the interesting early works on the emotions was that one of Ortony [3]. From this work, through componential analysis, other authors constructed an exhaustive taxonomy on affective lexicon. According to Ortony, stimuli that cause emotional processes are of three basic types: events, agents and objects corresponding to three classes of emotions: satisfied/ unsatisfied (reactions to events), approve/disapprove(reaction to agents), appreciate/unappreciate (reaction to objects). According to Osgood [4] an emotion consists of a set of stages: stimulus (neural and chemical changes), appraisal and action readiness. Continuing the studies of Charles Darwin, the Canadian psychologist Paul Ekman [5] has confirmed that an important feature of basic emotions is that they are universally expressed, by everybody in any place, time and culture, through similar methods. Some facial expressions and the corresponding emotions are not culturally specific but universal and they have a biological origin. Ekman, analyzed how facial expressions respond to each emotion involving the same type of facial muscles and regardless of latitude, culture and ethnicity. This study was supported by experiments conducted with individuals of Papua New Guinea that still live in a primitive way.

III. Emotions

Human emotions are deeply joined with the cognition. Emotions are important in social behavior and to stimulate cognitive processes for strategies making. Emotions represent another form of language universally spoken and understood. Identification and classification of emotions has been a research area since Charles Darwin's age. In this section we consider facial, vocal and textual emotional expressions.

3.1 Facial Expressions

Facial expression recognition [8] [9], coupled with human psychology and neuroscience, is an area which can bridge psychology and computations. Expressions of a human face can be captured through facial features. There are two types of facial expression features, transient (wrinkles and bulges) and transient (mouth, eyes and eyebrows). The feature points of a face, for recognizing facial expression, are located at eyebrows, eyelids, cheeks, lips, chin and forehead.

The first and the most important step in feature detection is to track the position of the eyes. Thereafter, the symmetry property of the face with respect to eyes is used for tracking the rest of the features like eyebrows, lips, chin, cheeks and forehead. The systems for treatment of facial expressions [10] are based on computational images. The model can contain information on the geometry of the face and facial muscles or on movements of various portions of the face during a change of expression.

3.2 Vocal Expressions

In the case of voice analysis, the parameters considered are typically volume, speed, regularity of speech. The vocal expression is also strongly influenced from the mood of the speaker, context and culture. For example, a hold orator, engaged in a major speech, hardly shows any tension level. He takes the same behavior in any context.

3.3 Textual Emotions

The core of our project is to recognize the emotion sensing from textual information. This field or research is known as emotion recognition or emotion computing. Human-machine Interface technology has been investigated for several decades. Recent research has placed more emphasis on the recognition of non verbal information. Textual information can be collected from many sources, such as books, newspapers, web pages, e-mail messages, etc. Nowadays Internet is the most popular communication medium also rich in emotion. With the help of natural language processing techniques, emotions can be extracted from textual input by analyzing punctuation, emotional keywords, syntactic structure and semantic information. We believe that text is a particularly important modality for emotion sensing because the most

important of user interfaces today are textually based. With textual emotions recognition [11] the text-based user interfaces become socially intelligent.. For textual emotional recognition it's better to focus on concepts, rather than words, so that words are related to emotional states through a structure of conceptual representation.

In the natural language, there are many words of a language that contain, in their semantic representation, information about an emotional state. If we split the phrase in part of speech (names, verbs, adjectives), we can consider different emotional terms: names (fear, awe, gratitude, disorientation), verbs (admire, hate, get angry, rejoice), adjectives (angry, furious, sad, happy), adverbs (sadly, joyful, interjections (ooh, perbacco). In textual case it is necessary to extract affective terms, relative to emotions or context, from statements in natural language. With the lexical approach it's possible to infer the properties of emotions from the analysis of linguistic labels.

To organize the natural language relating to emotions, the first step is to gather statements and terms from the vocabularies or corpus of statements extracted from emotional literary or journalistic texts. Inside affective terms it is possible extract emotional terms. The emotional terms set is splitted in groups of synonyms labeled by the term most representative. These terms are marked with properties (attributes or parameters). These parameters derive from lists containing up to 200 affective adjectives; with statistical techniques it is possible reduces the number of latent factors.

It has long been known that speech prosody[7], that is patterns in pitch and amplitude modulation and segmental durations carry emotional information in the acoustic speech signal [4,5]. The investigation of the phonetic characteristics of the emotional speech can be carried out only after making the proper classification of the emotional states. Emotion classifications of the researchers differ according to the goal of the research and the field .Emotion technology [1] is an important component of artificial intelligence, especially for human-computer communication. For emotion recognition by an artificial intelligent system we must take into account different contexts. Many kinds of physiological characteristics aroused to extract emotions, such as voice, facial expressions, hand gestures, body movements, heartbeat, blood pressure and textual information. The face and the verbal language can reflect the outside deepest emotions: a trembling voice, a tone altered, a sunny smile, the face corrugated.

Emotion identification is used in different applications such as Lie detector and can be used as voice tag in different database access systems. This voice tag is used in telephony shopping, ATM machine as a password for accessing that particular account. Formants are resonances of vocal tract and estimation of their location and frequencies at that location which is important in emotion recognition. Duration parameter is calculated using extraction of vowel, semivowel and consonant duration in that speech signals for emotion recognition. Vocal tract spectrum estimation or analysis is considered on the basis of bandwidth calculations of speech signal in frequency domain.

Emotion classifications of the researchers differ according to the goal of the research and the field .Also the scientist's opinion about the relevance of dividing different emotions is important. There is no standard list of basic emotions. However, it is possible to define list of emotions which have usually been chosen as basic, such as: erotic(love)(shringar), pathetic (sad) (karuNa), wrath (anger)(roudra), quietus (shAnta), normal(neutral).

The objective of this study is to analyze impact for different emotions on vowels in terms of certain parameters for stage prosody analysis.

IV. Prosody

In linguistics, prosody is the rhythm, stress, and intonation of speech. Prosody may reflect various features of the speaker or the utterance: the emotional state of the speaker; 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. Prosody has long been studied as an important knowledge source for speech understanding and also considered as the most significant factor of emotional expressions in speech [16]. Emotional prosody is the expression of feelings using prosodic elements of speech

V. Parametric Measurements of Acoustics Signals

Four acoustic signal features such as Vowel Duration, Pitch, Jitter and Simmer were used to parameterize the speech.

5.1 Duration

Utterance durations, vowel durations were measured from the corresponding label files produced by a manual segmentation procedure. On an average, utterance durations become longer when speech is emotionally elaborated.

5.2 Fundamental Frequency (pitch)

We calculated the pitch contours of each utterance using speech processing software. Global level statistics related to F0 such as minimum, maximum, mean were calculated from smoothed F0 contours.

5.3 Jitter & Shimmer

Jitter and Shimmer are related to the micro-variations of the pitch and power curves. In other words, Shimmer and Jitter are the cycle-to-cycle variations of waveform amplitudes and fundamental periods respectively. The Jitter & Shimmer occur due to some undesirable effect in audio signal. Jitter is the period frequency displacement of the signal from the ideal location. Shimmer is the deviation of amplitude of the signal from the ideal location. Mathematically Jitter is expressed as:

$$\text{Jitter} = \frac{\text{Average absolute difference between Consecutive period}}{\text{Average period}} \quad (1)$$

$$\text{Shimmer} = \frac{\text{Average abs diff. between amplitude of Consecutive period}}{\text{Average amplitude}} \quad (2)$$

VI. Methodology and Experimentation

There are many features available that may be useful for classifying speaker affect: pitch statistics, short-time energy, long-term power spectrum of an utterance, speaking rate, phoneme and silence durations, formant ratios, and even the shape of the glottal waveform [8, 11, 12, 9, 10]. Studies show, that prosody is the primary indicator of a speaker's emotional state [1, 13, 12]. We have chosen to analyze prosody as an indicator of affect since it has a well-defined and easily measureable acoustical correlate -- the pitch contour. In order to validate the use prosody as an indicator for affect and to experiment with real speech, we need to address two problems: First, and perhaps most difficult, is the task of obtaining a speech corpus containing utterances that are truly representative of an affect. Second, what exactly are the useful features of the pitch contour in classifying affect? Especially as many factors influence the prosodic structure of an utterance and only one of these is speaker's emotional state [6, 7, 9].

The data analyzed in this study were collected from semi-professional actors and actress and consists of 30 unique Odia language sentences that are suitable to be uttered with any of the five emotions i.e., roudra, shringar, shanta, karuna, and neutral. Some example sentences are "Jayantta jagi rakhhi kaatha barta kaara", "*Babu chuata mora tinni dina hela khaaini*". The recordings were made in a noise free room using microphone. For the study, samples have been taken from three male & three female speakers. The utterances are recorded at the bit rate of 22,050Hz. The Vowels are extracted from the words consisting of 3 parts i.e.CV, V, VC. CV stands for Consonant to Vowel transition, V for steady state vowel, VC for Vowel to Consonant transition.

VII. Experimental Results

For experimental analysis data samples were created from recordings of male and female speakers in various emotions (mood). Vowels are then extracted and stored in a database for analysis. From this database after analysis the result of the utterance. Duration, fundamental frequency & variations of pitch of every vowel are measured and compared to give following results.

7.1 Duration

It is observed from the duration Table 1 that speech associated with the emotion "love(Shringar)"has higher duration with the vowel /i/ gets elongated both for male and female speakers where as the emotion "sad (karuna)" for male speakers vowel(/a/,/i/) gets elongated whereas for female (/i/,/u/) gets elongated.

Table (1): Average duration of vowels for speakers in different emotions

Emotions	Duration in Millie Seconds (male-average)				
	/a/	/i/	/u/	/e/	/o/
Neutral	67	70	60	39	53
Santa	58	75	60	58	43
Karuna	**106**	**106**	61	56	54
Raoudra	64	50	43	49	56
Shringar	**100**	**113**	**90**	54	56
Emotions	Duration in Millie Seconds (female-average)				
	/a/	/i/	/u/	/e/	/o/
Neutral	50	64	66	40	35
Santa	53	**99**	79	43	50
Karuna	**83**	**105**	**101**	**86**	63
Raoudra	50	43	58	38	41
Shringar	**80**	**161**	**118**	55	54

7.2 Fundamental Frequency

Figure1shows the analysis result that the mean pitch for male speaker associated with emotion karuna for vowel /i/ has dominance where as in female the speech associated with emotion shringar for vowel /a/ plays dominant role.

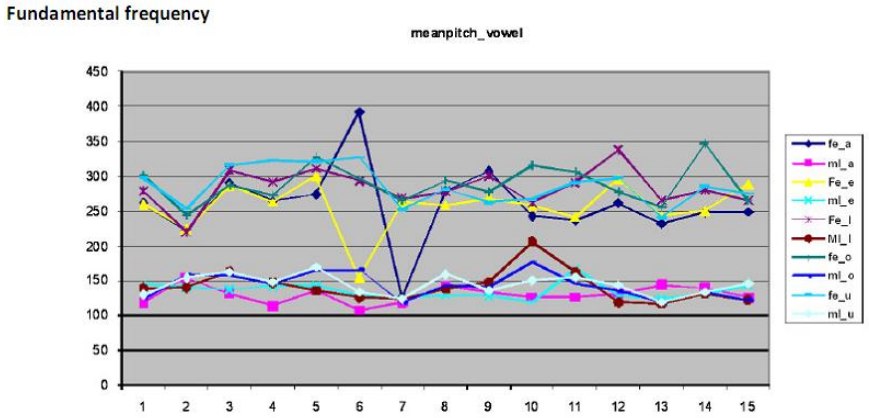

Figure1: Mean pitch of vowels of speakers in different emotions

7.3 Jitter

Figure2 shows that the emotion "anger" of vowel /i/ is having a dominant role in jitter for male speaker. The jitter value for female speaker has dominant vowel /u/ for the emotion "love".

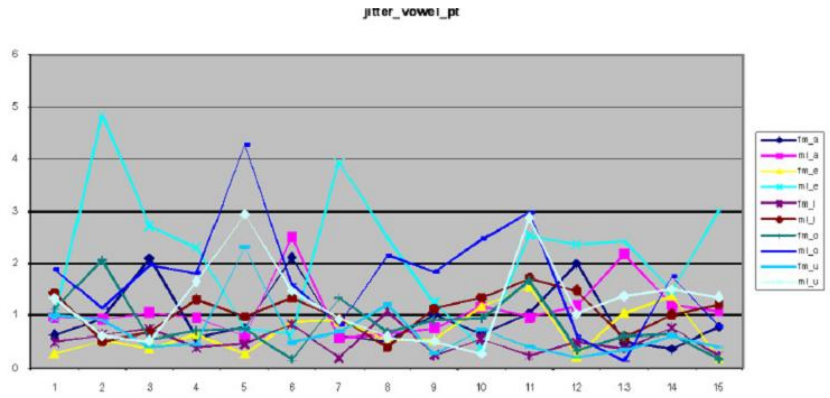

Figure2: Jitter (local) for vowels in different emotions

7.4 Shimmer

It is observed from figure3 that the shimmer in the emotion "anger" of vowel /o/ has dominant role for males & for female in emotion "anger" of vowel /i/ has dominant role.

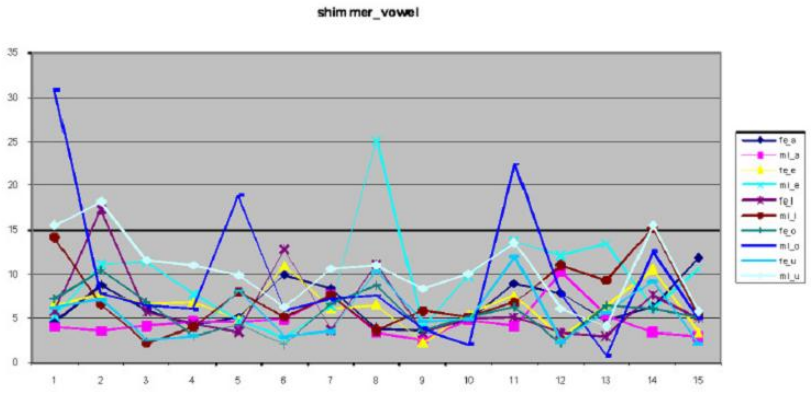

Figure3: Shimmer (local) for vowels in different emotions

VIII. Conclusion

In this study, we investigate acoustic properties of speech prosody associated with five different emotions (love (shringar)), pathetic (sad) (karuNa), wrath (anger) (roudra), quietus (shAnta),normal (neutral) intentionally expressed in speech by male and female speakers. Results show speech associated with love (shringar) and sad (karuna) emotions are characterized by longer utterance duration, and higher pitch. However we observed that for jitter anger or love has dominance over others, whereas for shimmer the emotion anger plays a vital role. Future works of my research are the following. We have to collect synthetic speech and put emotion labels on them. We have to reconsider how to estimate emotion in speech using parallel programming.

References

[1] P. Olivier and J. Wallace, Digital technologies and the emotional family, International Journal of Human Computer Studies, 67 (2), 2009, 204-214.

[2] W. L. Jarrold, Towards a theory of affective mind: computationally modeling the generativity of goal appraisal, Ph.D. diss., University of Texas, Austin, 2004.

[3] C. G. Ortony and A. Collins, The cognitive structure of emotions, (Cambridge University Press: New York, 1990).

[4] M. M. C.E. Osgood and W.H. May, Cross-cultural Universals of Affective Meaning, (Urbana Chaimpaigne: University of Illinois Press, 1975).

[5] E. Paul, Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life, (NY: OWL Books, 2007).

[6] A. Ichikawa and S. Sato, Some prosodical characteristics in spontaneous spoken dialogue, International Conference on Spoken Language Processing, v. 1, 1994, 147-150.

[7] R. Collier, A comment of the prediction of prosody, in G. Bailly, C. Benoit, and T.R. Sawallis (Ed.), Talking Machines: Theories, Models, and Designs, (Amsterdam: Elsevier Science Publishers, 1992).

[8] H. Kuwabara and Y. Sagisaka, Acoustic characteristics of speaker individuality: Control and conversion, Speech Communication, 16(2), 1995, 165-173.

[9] K. Cummings and M. Clements, Analysis of the glottal excitation of emotionally styled and stressed speech, Journal of the Acoustical Society of America, 98(1), 1995, 88-98.

[10] D. Roy and A. Pentland, Automatic spoken affect classification and analysis, Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition, 1996, 363-367.

[11] A. Protopapas and P. Lieberman, Fundamental frequency of phonation and perceived emotional stress, Journal of Acoustical Society of America, 101(4), 1997, 2267-2277.

[12] X.Arputha Rathina and K.M.Mehata, Basic analysis on prosodic features in emotional speech, nternational Journal of Computer Science, Engineering and Applications (IJCSEA) , Vol.2, No.4, August 2012,99-107

[13] D. Hirst, Prediction of prosody: An overview, in G. Bailly, C. Benoit, and T.R. Sawallis (Ed.), Talking Machines: Theories, Models, and Designs, (Amsterdam: Elsevier Science Publishers, 1992).

[14] L. Rabiner and R. Shafer, Digital Processing of Speech Signals, (New York: Wiley and Sons, 1978)

[15] Wavesurfer, http://www.speech.kth.se/wavesurfer/

[16] Masaki Kuremastsu et eal, An extraction of emotion in human speech using speech synthesize and classifiers foe each emotion , International journal of circuits systems and signal processing, 2008.