Emotiphons: Emotion markers in Conversational Speech - Comparison across Indian Languages

Nandini Bondale[1]   Thippur Sreenivas[2]

(1) School of Tech. and Comp. Sci., Tata Institute of Fundamental Research, Mumbai, 400005, India.
(2) Dept. of ECE, Indian Institute of Science, Bangalore, 560012, India.

drnandini.bondale@gmail.com, tvsree@ece.iisc.ernet.in

ABSTRACT

In spontaneous speech, emotion information is embedded at several levels: acoustic, linguistic, gestural (non-verbal), etc. For emotion recognition in speech, there is much attention to acoustic level and some attention at the linguistic level. In this study, we identify paralinguistic markers for emotion in the language. We study two Indian languages belonging to two distinct language families. We consider Marathi from Indo-Aryan and Kannada from Dravidian family. We show that there exist large numbers of specific paralinguistic emotion markers in these languages, referred to as *emotiphons*. They are inter-twined with prosody and semantics. Preprocessing of speech signal with respect to *emotiphons* would facilitate emotion recognition in speech for Indian languages. Some of them are common between the two languages, indicating cultural influence in language usage.

KEYWORDS : Emotion recognition, *emotiphons,* emotion markers, Indian languages

1   Introduction

One of the important reasons for communication is the desire on the part of the members to express their emotions (Millar 1951). Language is the effective tool to carry out this task and speech is the most efficient mode of language communication between humans. In the recent years, emotion recognition in human speech is more important because of human computer interaction as in automatic dialogue systems or robotic interactions (Cowie et al., 2001). In interactive applications, detection of emotions such as frustration, boredom or annoyance in the speaker's voice helps to adapt the system response, making the system more effective. Speech carries a lot of information over and above the text content in the language. Speaker's voice expresses the physical and emotional state, sex, age, intelligence and personality (Kramel, 1963). Emotion is intimately connected with cognition and many physiological indices change during emotion arousal (Lindsay and Norman, 1972).   The task of speech emotion recognition is challenging as it is not clear which speech features are effective in distinguishing a large range and shades of emotions over a range of human voices and context. How a certain emotion is expressed generally depends on the speaker, his or her culture and environment (Ayadi et al., 2011). Therefore, integration of acoustic and linguistic information has been tried out. (Lee and Pieraccini, 2002, Schuller et al. 2004). Spoken dialogue and written language are very different due to many paralinguistic aspects such as the *emotiphons*, defined and discussed in this paper.

In this study, we examine specific lexical expressions in Indian languages conveying emotion, referred to as *emotiphons*. This is the first attempt of its kind to list and study these lexical expressions. We consider two Indian languages, namely Marathi from Indo-Aryan family and Kannada from Dravidian family, whose people are culturally very connected. This data across languages and their acoustic correlates would throw light on the flow of information from the prosodic level to the highest cognitive level of speech processing, in general, and emotional speech processing in particular.

The following section describes the role of *emotiphons* in emotion recognition. Section 3 lists *emotiphons* in Marathi and Kannada. Section 4 mentions the observations along with discussion. Section 5 states conclusions.

2   Speech and emotion

Cowie and Cornelius (2003) have described issues related to speech and emotion in great details, covering the basic concepts and relevant techniques to study conceptual approaches. It is well recognized that emotion analysis in human communication is multi-faceted and varied. It is also intertwined with the culture of the language users.

2.1   *Emotiphons*

In this study, we identify specific lexical expressions referred to as *Emotiphons*, used to communicate emotions, in Indian languages. *Emotiphons* are essentially short lexical expressions in conversational speech conveying emotions by modifying the prosody of the utterance. Use of *emotiphon* keeps the body of the lexical content unaltered in a sentence, but explicitly brings out the intended emotion. Yet they are not considered as part of lexicon always; hence can be referred to as paralinguistic markers. *Emotiphons* are analogous to emoticons of printed text that have become so essential in email communication. In contrast to affect-bursts which are non-

speech in nature (Schroder 2003), *emotiphons* are phonetic in nature and blend well with the lexical phonetics and sentence structure of the language. Being short and specific, *emotiphons* disambiguate subtle emotions and help to convey emotions better and stronger.

2.2   Emotion recognition

In databases for emotion recognition, it is common to record the same sentence with different emotions, thus reducing the effect of lexical content on perceived emotions. This suggests that, a particular lexical content can be expressed in more than one type of emotion. In such cases, it can be seen from our data mentioned in section 3, that suitable *emotiphon* can be used by speakers, to effectively express the respective emotion.

Improvement in speech emotion recognition performance has been attempted by combining other information such as facial expressions or specific words along with acoustic correlates. It has been shown that searching for emotional keywords or phrases in the utterances and integrating linguistic classifier with acoustic classifier have improved emotion classification accuracy (Ayadi et al., 2011). Computational techniques used in these approaches could be varied depending on the sophistication of the system application. The *emotiphons* discussed in this paper would be an additional source for emotion recognition. The presence of *emotiphons* heavily affects the prosody and convey emotions effectively. Some of the *emotiphons* are stand-alone and hence may be identified through a pre-processing stage, such as keyword spotting, whereas other *emotiphons* would have to be viewed along with prosody. Stochastic model based recognition would be required in most cases, because of the subjective variability of pronunciation.

3   *Emotiphons* in two Indian Languages

The Indian subcontinent is a good example of a sprachbund (Emeneau, 1956), because there are two distinct language families, Indo-Aryan and Dravidian. However, there is a lot of interaction and similarity across the languages belonging to these two families due to centuries of language and culture contact. While grouping Indian languages using machine learning techniques, based on their text, it is observed that, Marathi is the closest language to the Southern zone consisting of the languages from Dravidian family and can be grouped with Hindi, Punjabi and Gujarati. The grouping corresponds well with the geographic proximity also (Ghosh et al., 2011).

In the following tables we give a sample list of *emotiphons* used in Marathi and Kannada, the languages of Maharashtra and Karnataka states respectively. We have categorized *emotiphons* in different groups. Phonetic representation of e*motiphons* is given using IPA symbols. Additional characteristics are mentioned wherever they are significant. (All the emotions mentioned in the following tables are indicative and may change from region to region where the language is spoken).

Table 1 lists the *emotiphons* that are smallest expressions consisting a single vowel or a diphthong. They are 'stand-alone' expressions, i.e., can be used in isolation to express the respective emotions and may not need conversation mode. In all the tables, 'K' stands for Kannada & 'M' and Marathi.

Phonetic representation	Language	Pitch &/or loudness	Emotion description
1. [ɑː] or [ɑ̃ː]	K& M	a) Falling b) Rising	a) Pain b) Request to repeat
2. [i]	K&M	Flat	Disgust, dislike
3. [ũː]	K&M	a) Falling b) Rising c) Low & louder	a) Exhausted b) Pain c) Disapproval
4. [e]	K&M	a) High b) Low	a) Rude alert b) Affectionate alert
5. [o]	K&M	a) Rising b) Flat	a) Exclamation b) Mild amazement
6. [ei]	K&M		Derogatory challenge

Table 1 - 'Stand-alone' *emotiphons*; (Vowels, diphthong)

Table 2 lists *emotiphons* which are fricative like and can be used as isolated expressions to express the respective emotion.

Phonetic representation	Language	Additional characteristic	Emotion description
1. [t͡ʃʰeː]	a) K b) M		a) Sadness b)Displeasure
2. [t͡ʃet͡ʃeː]	a) K b) M		a) Repenting b)Disapproval
3. a) [ʃiː] b) [cʰiː]	a) M b) K		Disgust
4. [ǀ] (dental click)	M	Single or multiple utterances with breaks	Frustration or repenting or disappointment or sadness

Table 2 - 'Stand-alone' *emotiphons;* (Fricatives, click)

Tables 3a and 3b list some of the *emotiphons* which involve multiple-phons.

Emotiphons in Table 3a can be used in isolation as in Table 1 and 2.

Phonetic representation	Language	Additional characteristic	Emotion description
1. [oho:]	K & M	a) No stress b) Stress 1st/last vowel	a) Surprise b)Surprise with sarcasm
2. [ohoho….]	K & M		Enjoying the surprise
3. [ʈʰu]	K & M		Dirty, disgust
4. a) [əjjo:] b) [əjjəjjo:]	K		a) Pain b)Severe pain
5. [ɑigə:]	M		Pain
6. [ɑiggə:]	M		Boredom

Table 3a- 'Stand-alone*' emotiphons;* (Multi-phons)

Table 3b lists the multi-phon *emotiphons* that are used only in conversational mode in contrast to 'stand-alone' expressions.

Phonetic representation	Language	Pitch or other characteristic	Emotion description
1. [ə̃hə̃]	K&M		Emphatic disapproval, disagreement
2. a) [pɑpə] b) [ərere]	a) K b) M		Sympathy
3. a) [kəɳo] b) [re]	a) K b) M	Singular, masculine	Affectionate address
4. a) [kəɳe] b) [gə]	a) K b) M	Singular, feminine	Affectionate address

Table 3b – Conversation mode *emotiphons ;* (Multi-phons)

4   Observations and discussion

It can be seen from the tables above that the number of emotions covered by *emotiphons* are far more than those expressed in the databases mentioned in the literature for emotion recognition. Common emotions covered by the databases are anger, fear, joy, sadness, disgust, surprise and neutral, mainly in the prosody at the acoustic level. However, *emotiphons* express many more shades and nuances of emotions like affection, pain, disbelief, sympathy, boredom and so on, which are all important for the semantic context.

We have classified e*motiphons* grossly into three categories; Vowel like, Fricative like and Multi-phons as mentioned in Table 1, 2 and 3 respectively. *Emotiphons* in Table 1, 2 and 3a are 'Stand-alone', self-expressive, conveying a specific emotion. They are exclamatory in nature. 'Stand-alone' *emotiphons* are unaffected by the linguistic parameters such as gender and number. They cover large number of emotions. (All are not listed due to space limit). *Emotiphons* in Table 3b are used in conversation mode only.

Although we believe that many emotions would be common across all humans which is a Darwinian perspective (Hozjan and Kacic, 2003), we feel that expression of emotion is dependent on culture and society. As seen in Table 1 and 2, many *emotiphons* are common across Marathi and Kannada, suggesting that people using these languages share similar cultural values, although the languages belong to two different families. We feel that *emotiphons* would be common across other Indian languages too. (Study of *emotiphons* for other Indian languages is in progress). The common *emotiphons* across Marathi and Kannada are of 'stand-alone' type and are independent of linguistic parameters such as gender and number. *Emotiphons* which depend on linguistic parameters are expected to vary across the languages and is evident from Table 3b.

Conclusions

We identified that there exist many emotion markers, referred to as *emotiphons* in two Indian languages, Marathi and Kannada belonging to Indo-Aryan and Dravidian language family, respectively. We find that *emotiphons* are short lexical expression used in conversational speech to convey many different specific emotions explicitly and effectively. Although Marathi and Kannada are from two different language families, we notice that there are many common *emotiphons* across the two languages. Commonality of *emotiphons* across the languages would lead us to understand cognitive aspects of the emotion communication as well as the linguistic evolution. *Emotiphons* would play a major role in identification of emotion in speech processing, adding naturalness to synthesized speech, and in design of dialogue systems.

References

Ayadi, M. E., Kamel, M.S. and Karray, F. (2011). Survey on speech emotion recognition: features, classification schemes, and databases. *Pattern Recognition*, 44:572-587.

Cowie, R. and Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech, *Speech Communication,* Vol. 40, pp 5-32.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Kollias, S., Fellenz, W., and Taylor, J. (2001). Emotion recognition in human–computer interaction. *IEEE Signal Processing Magazine*, 18, 32–80.

Emeneau, M. B. (1956). India as a linguistic area. *Language*, Vol. 32, No. 1.

Ghosh, S. K. Girish, K. V. and Sreenivas, T. V. (2011). Relationship between Indian languages using long distance bigram language models, In *Proceedings of ICON-2011: 9th International Conference on Natural Language Processing,* pp 104-113, Macmillan Publishers, India.

Hozjan, V and Kacic, Z. (2003). Context-independent multi-lingual emotion recognition from speech signals. *Int. J. Speech Tech.,* 6:311-320.

Lee, C. and Pieraccini, R. Combining acoustic and language informationfor emotion recognition, In *Proceedings of the ICSLP 2002,* 873-876.

Kramel E. (1963). Judgment of personal characteristics and emotions from nonverbal properties of speech, *Psych. Bulletin,* 60:408-420.

Lindsay, P. H. and Norman, D. A. (1972). Human Information Processing, Academic Press, New York and London.

Millar, G. A. (1951). Language and Communication. McGraw-Hill Book Company, New York, U.S.

Schroder, M. (2003). Experimental study of affect bursts. *Speech Communication*, Vol. 40, pp 99-116.

Schuller, B., Rigoll G., Lang, M. (2004). Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture, In *Proceedings of the ICASSP,* vol 1, pp.577-580.