Establishing Relationship between Emotional Prosody and Affective Pragmatics

Rana Muhammad Basharat Saeed

M.Phil. Scholar (Applied Linguistics)

The University of Lahore

Sargodha, Punjab, Pakistan

engr.basharatrana@gmail.com

Muhammad Umair Ashraf

B.S. (Hons) English

University of Sargodha

Sargodha, Punjab, Pakistan

ranau0592@gmail.com

Abstract

This work aims to provide theoretical grounds to establish a relationship between emotional prosody and affective pragmatics. Affective pragmatics is a significant yet least studied area of research. This theoretical framework focuses how emotional expressions are constructive in channelizing the pragmatic meaning under the umbrella of affective pragmatics and also encompasses the speech engineering that conveys unabridged abstract emotions in the phenomenal process of emotion recognition. Since speech act theory focuses meaning at

DOI: 10.24113/ijohmn.v9i4.271 Received 10 May 2023, Revised 25 July Accepted 10 August 2023

utterance level and not at emotional level, thus, there is a need to reflect on emotional expressions that function as paralinguistic features. There are several studies carried out on identification of emotions using prosodic modeling; however, there is no meticulous study that shows relationship between emotions and their pragmatic meanings. The current study will be an application of Theory of Affective Pragmatics (TAP) proposed by Andrea Scarantino in 2017, a theory analogous to speech act theory. The objectives of the study can be achieved through analysis of quantitative data procured taking advantage of the emotional recordings from Emotional Prosody Speech and Transcripts (EPST), a worldwide database used in emotion recognition processes, along with employment of appropriate prosodic features using a Hidden Markov Model (HMM), a popular emotion recognition statistical model.The outcomes of this study will contribute to benefit new researchers in the field of linguistics to understand affective pragmatics at a profound level as a novel area of research.

Keywords: Theory of Affective pragmatics, Illocutionary Acts, Communicative moves, Hidden Markov Model, Pragmatic Meaning.

1. Introduction

Human communication is amazingly multimodal. The linguistic utterances are generally accompanied by non-verbal signals, therefore, one can draw inferences based on non-verbal expressions such as emotional expressions(Wu et al., 2021). Prosody, in linguistics, is concerned with larger units of speech such as stress, intonation, and rhythm, and these elements are known as suprasegmentals(Prieto & Roseano, 2018). Prosody not only reflects the form of utterance but also the emotional state of speaker. To understand the prosodic aspects, it is inevitable to distinguish between subjective impressions (auditory measures) and objective measures (acoustic properties). The major prosodic variables in auditory terms are pitch of voice, length of sounds,

loudness or prominence, and timbre. The acoustic properties closely correspond to the fundamental frequency, duration, intensity, and spectral characteristics. The behavioral analysis of prosodic variables is generally done either as contours or boundaries (Lai & Gooden, 2016). Prosodic features are also instrumental in signaling emotions and attitudes, yet context clues are equally important (Mozziconacci, 2002).

Pragmatics, in linguistics, is the study of meaning of an utterance in context and pragmatic competence is the ability to understand speaker's intended meaning(Domaneschi & Bambini, 2020). Prosodic features play fundamental role in conveying the pragmatic meaning of an utterance. (Pronina, Hübscher, Vilà-Giménez, & Prieto, 2021). Prosodic pragmatics is central in identifying the intentions behind the utterances produced to distinguish between the individual features of speaker and the features that knit the web of meaning (Wichmann, Dehé, & Barth-Weingarten, 2009). In prosody, emotional expressions have their own particular natural meaning and ability to make communicative moves. The meanings of prosodic features become ambiguous once isolated from the context. TAP emphasizes that emotional expressions have natural meanings that rely on statistical correlations instead of non-natural meanings that depend on reflexive intentions of signalers (Scarantino, 2017b).

To study emotional expressions ata profound level, Andrea Scarantino proposed a novel theory in 2017 named theory of affective pragmatics (TAP) introducing a new framework of affective pragmatics. According to him, "as linguistic pragmatics focuses on what utterances mean in a context, affective pragmatics focuses on what emotional expressions mean in a context". He emphasized two principal insights in TAP. The first has long been neglected by the Basic Emotion theorists, is that "emotional expressions do much more than simply expressing emotions". The second is that "it is possible to engage in analogs of speech acts without using

language at all". Correspondingly, hemanifested two principal objectives of TAP. The first principleaims to advance experimental research to analyze emotional expressions by introducing a new taxonomy of communicative moves. The second one aims to set the stage for better understanding of the evolutionary aspect of language by unveiling commonalities and dissimilarities between verbal and non-verbal communication. To explicate TAP, he asserted that there could be genuine analogs for the speech acts performed without using language. In this regard, he proposed a general taxonomy of communicative moves for emotional expressions. For this purpose, he borrowed Searl's taxonomy of illocutionary acts which is an expansion of Austen's concept of speech act theory in "How to Do Things with Words". There are two main parts of TAP, the first is the distinction between the three speech acts named locutionary, illocutionary, and perlocutionary acts in terms of emotional expression (the nonverbal behavior *of* expressing emotion E), communicative moves (what one does *in* expressing emotion E), and communicative effects (what one does *by* expressing emotion E) respectively. The second part accompanies the analysis of these proposed dimensions of emotional communication; the analysis of nature and function of emotional expressions, communicative moves, and communicative effects. (Scarantino, 2017a)

Now, to understand the pragmatic meanings of emotions in affective pragmatics, we need first to identify emotions through emotional prosody. The prosodic analysis of emotions can be accomplished using some prosody modeling technique that determines how the superasegmental elements of speech (generated automatically) communicate the structure and meaning of utterances (Pan, 2002). In prosody, different prosodic features such as pitch, energy, duration, etc., are extracted to get additional information beyond utterances to better understand the pragmatic meaning of speech. It has been found that prosody helps in automatic identification of

significant events such as sentence boundaries, disfluencies, discourse markers, and emotions in dialogues thatultimately enrich speech recognition process (Liu, 2006). Prosody modeling is a process of producing prosodic variations in a synthesized speech automatically by building different computational models. For speech synthesis, Hidden Markov Model (HMM) is considered one of the best models among other speech synthesis systems (Rajeswari & Uma, 2012). HMM based on statistical parametric synthesis functions corresponding to maximum likelihood criterion in which frequency spectrum ofvocal tract, fundamental frequency of voice source, and duration asprosodic feature are simultaneously modeled in wave forms (Zen et al., 2007).

2. Literature Review

The role of prosody is priceless in disambiguating the syntactic and pragmatic information of utterances and is also instrumental in providing linguistic information of suprasegmentals (Applebaum, Coppola, & Goldin-Meadow, 2014). It is complex sphere of linguistics that establishes connection among grammatical, pragmatic, and affective levels in a language (Martin, 2008). The relationship between prosody and pragmatics is somehow hard to elucidate to non-native speakers, however, it is intelligible for the native speakers (Levinson, 1983).

Charles Darwin was the first person who took emotional prosody into account in “The Descent of Man” to predate the evolutionary journey of human language. He opined that monkeys express their feelings of anger, impatience, fear, pain, and happiness in different low and high notes (Darwin, 2008). In the late twentieth century, Ekmanand Fridlund took inspiration from Darwin’s book “The Expression of Emotions in Men and Animals” (1872), and further focused on bodily vehicles of emotional expressions (Scarantino, 2017b). Ekman, a proponent of

Basic Emotion Theory, agreed with Darwin's view point that "all facial expressions of emotion are involuntary". He suggested that the emotions specific to facial expressions are involuntary and are caused by basic affective programs such as anger, fear, sadness, happiness, disgust, and surprise. His experimental evidences divulged that facial expressions are universal and culturally structured. His experimental work suggests that the involuntary facial expression of an emotion is a prime communicative clue that facial expressions are independent of context. Fridlund, a proponent of the Behavioral Ecology View refuted both the Ekman's assumptions asserting that facial expressions are voluntary and context dependent. He shed light on complexities of Darwin's analysis of emotional expressions. Voluntariness of facial expression and context dependency are the central disagreements between the proponents of two different schools of thought. Darwin understood basic emotions as feelings by following the trend of time. While, to cope with persistent evolutionary issues, Ekman proposed a more sophisticated framework of understanding the arche typal emotions as basic affect program. In this basic affect program, the archetypal emotions may engross feelings, however, it is probable that feelings may not be accompanied by the archetypal emotions under certain circumstances (Scarantino, 2017a).

A. Fischer and D. Sauter appreciated TAP proposed by Andrea Scarantino as an integrative theory of emotional expressions that aims to bridge the gap between the two opposing theories by Ekman and Fridulund. Nonetheless, they also rasised some questions about voluntariness of emotional expressions; either voluntary, or involuntary? How statistical correlations are said to be true for voluntarily produced expressions, however these may be true on perception level but not on production level? A query raised for emotional clarity of imperatives that emotions can be dramatically different depending on the context. i.e., from a sincere confession to ahostile mind-set. They agreed with Scarantino for relying on contextual

information to fully understand the pragmatic meaning of an emotionin an utterance, but they also look for more insightful articulation of emtions in TAP in terms of production and perception (Fischer & Sauter, 2017).

Scarantino argued that emotional expressions can be both voluntary and involuntary holding the notion that the topic is too complex to formulize for quick treatment. Voluntary actions depend more on mental representations of intended goal and anticipated effect and cannot be fully concluded by an immediate stimulus. On the contrary, involuntary actions can fully be determined by an immediate stimulus. It has been suggested through experimental paradigm that expression of involuntary emotions are the key communicative point and stay unaffected by the context. (Scarantino, 2017b)

According to Scarantino, the framework offered in TAP to understand affective pragmatics is not only valid for facial expression but also applicable to emotional expressions of all kinds. The fundamental proposal of TAP is that we can engage in a variety of communicative moves just as we engage in multiple illocutionary acts presented in speech act theory. Scarantino is to discover the potential of natural meanings from emotional expressions. That is why; he highlights the difference between the natural meanings of emotional expressions and the non-natural meanings of linguistic utterances. There is a vague concept of “social motives” propositioned by behavioral ecologists, TAP improves upon differentiating between varieties of things we do with emotional expressions. To get full account of the communicative moves, TAP is needed to be extended to other forms of non-linguistic communication such as gestures, spatial positioning, orientation, etc. (Scarantino, 2017a)

3. Methodology

3.1 Theoretical Underpinnings

Speech act theory offers distinction among three speech acts; locutionary acts (the acts *of* saying something), illocutionary acts (the acts one does *in* saying something), and perlocutionary acts (the acts one does *by* saying something). Analogously, TAP distinguishes among three emotional acts; the emotional expressions (the nonverbal behavior *of* expressing the emotion), the communicative moves(what one does *in* expressing the emotion), and the communicative effects (what one does *by* expressing emotion E) respectively (Scarantino, 2017a).

The core thesis of TAP is that emotional expressions can act as four communicative moves analogous to first four out of five types of illocutionary acts. Scarantino labeled Declaratives$_{EE}$ as analogues of Assertives, Imperatives$_{EE}$ as analogues of Directives, Commissives$_{EE}$ as analogues of Commissives, and Expressives$_{EE}$ as analogues of Expressives. Remarkably, the emotional expressions are deemed to be the analogues of speech acts by dint of their natural information they carry (Scarantino, 2017b).

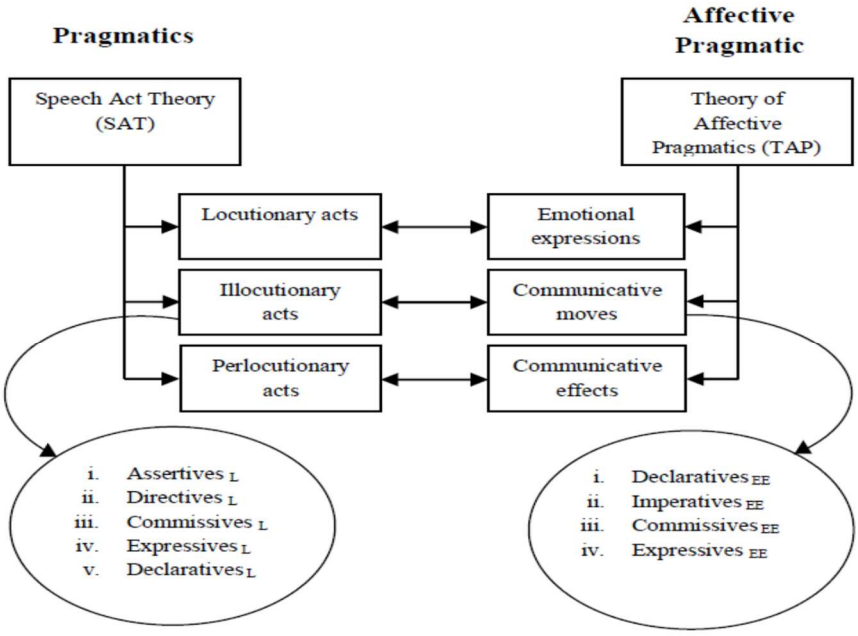

Figure 1 *Theoretical Framework of TAP*

The influential taxonomy of communicative moves is built around three basic features of illocutionary acts; illocutionary point, direction of fit, and sincerity condition. Likewise, in TAP, the respective three features of communicative moves are emotional expression $_{EE}$, direction offit, and communicative effect (Scarantino, 2017a).

Table 1 *Features of Communicative Moves*

Emotional Expression (EE)	Direction of Fit	Communicative Effect
Expressive $_{EE}$	Null	The recipient formation of the belief that the signaler is in a certain emotional state
Imperative $_{EE}$	world-to-mind	The recipient does what the signaler demands
Declarative $_{EE}$	mind-to-world	The recipient formation of the belief that the world is as signaler represents it to be
Commissive $_{EE}$	world-to-mind	The recipient comes to expect the behavior of the signaler commits to

3.2 Research Design

The objectives of the study can be achieved through analysis of quantitative data obtained by HMM emotion recognition statistical model. The data can be generated by use of emotional recordings from Emotional Prosody Speech and Transcripts (EPST) as EPST is a worldwide database used in emotion recognition processes. EPST database contains speech data comprising nine hours and includes fifteen categories of emotions named hot anger, cold anger, panic-anxiety, despair, sadness, elation, happiness, interest, boredom, shame, pride, disgust, and contempt (Ververidis & Kotropoulos, 2012).

In emotional prosody, the selection of the features to be employed is one the prime factors to be considered while building a HMM based emotion recognition framework. To get well estimated results of prosody as per global statistics (means, median, standard deviations,

etc.), it is emphasized that the prosodic information obtained must fit the HMM structure. The features that can easily be estimated in real time framework are preferred to be employed because these are useful in conceiving better estimated output of the emotional state in an utterance. Spectral measures usually provide complex information to be characterized as the spectrum greatly depends on phonetic content of the utterance. Contrary to spectral measures, pitch and energy are the features that belong to broader classes of sound i.e., superasegmentals, with no dependency on phonemes. The dependency of spectral measures on phonetic content is believed to be a flaw while building language-independent emotion recognizer. Hence, the spectral features may be neglected in the process of emotion recognition. To get valuable information about an emotion, the consideration of syllabic contours of pitch with instantaneous levels is a fine technique. For this purpose, a simple auto-correlation is executed at every frame to characterize instantaneous pitch. It has been found that a HMM based approach for emotion recognition is more useful when employed with short time low level features. For both energy and pitch, instantaneous features provide better results. Pitch features function better when compared to energy features. Therefore, the best feature combination to be employed is instantaneous pitch (Nogueiras, Moreno, Bonafonte, & Mariño, 2001).

In emotional prosody, the selection of the features to be employed isone the prime factors to be considered while building a HMM based emotion recognition framework. To get well estimated results of prosody as per global statistics (means, median, standard deviations, etc.), it is emphasized that the prosodic information obtained must fit the HMM structure.The features that can easily be estimated in real time framework are preferred to be employed because these are usefulin conceiving better estimated output ofthe emotional state in an utterance. Spectral measures usually provide complex information to be characterized as the spectrum greatly

depends on phonetic content of the utterance. Contrary to spectral measures, pitch and energyare the features thatbelong to broader classes of soundi.e., superasegmentals,with no dependency on phonemes. The dependency of spectral measures on phonetic content is believed to be a flaw while building language-independent emotion recognizer. Hence, the spectral features may be neglected in the process of emotion recognition. To get valuable information about an emotion, the consideration of syllabic contours of pitch with instantaneous levels is a finetechnique. For this purpose, a simple auto-correlation is executed at every frame to characterize instantaneous pitch. It has been found that a HMM based approach for emotion recognition is more useful when employedwith short time low level features. For both energy and pitch, instantaneous features provide better results. Pitch featuresfunction better when compared to energy features. Therefore, the best featurecombination to be employed is instantaneous pitch(Nogueiras, Moreno, Bonafonte, & Mariño, 2001).

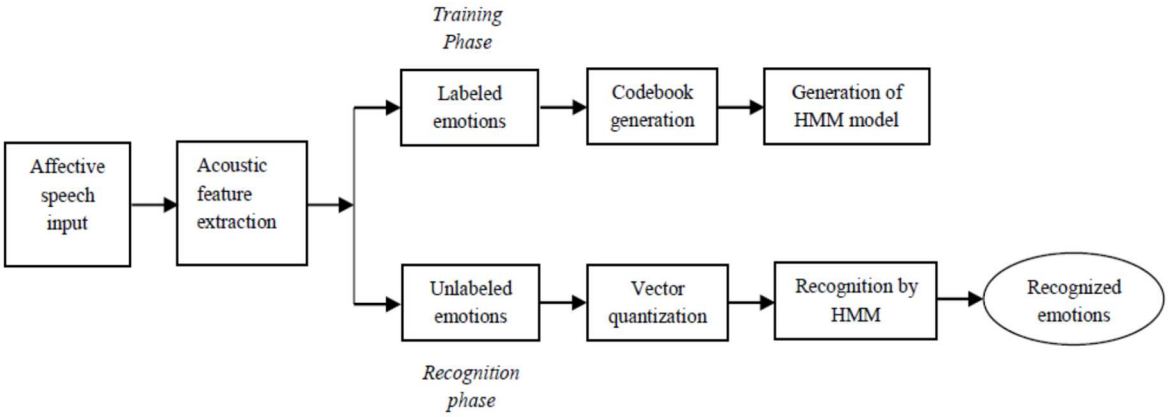

Figure 2 *HMM based Emotion Recognition Process*

3.3 Conceptual Framework

Considering the above theoretical grounds, the conceptual framework for the current research is presented in the following block diagram. This framework will be instrumental in understanding the practical role of affective pragmatics.

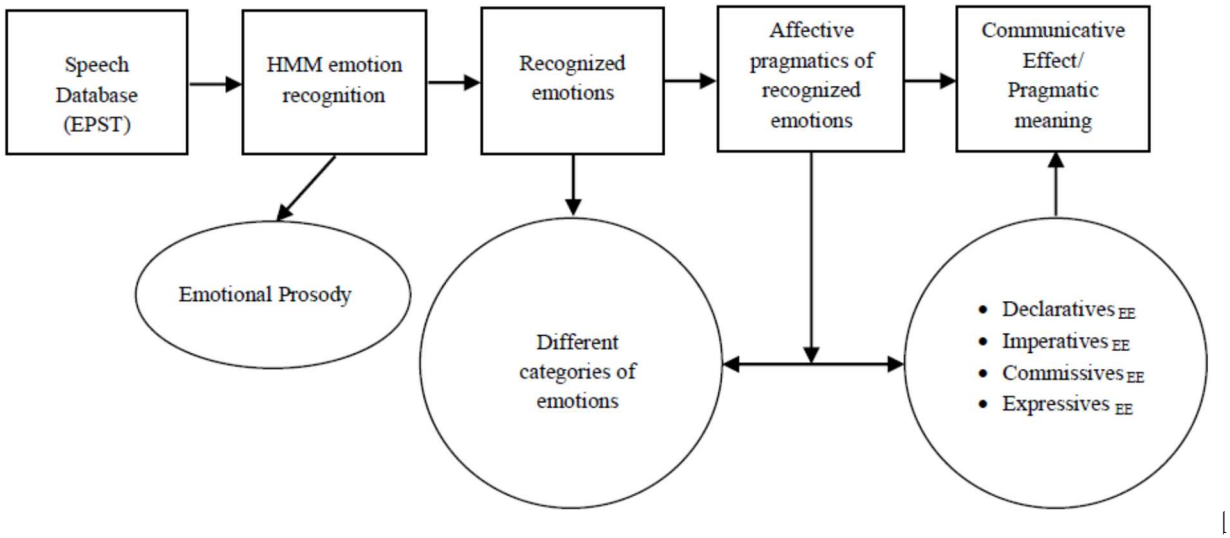

Figure 3 *Conceptual Framework of the Study*

expressions. Andrea Scarantino introduced TAPas a new framework of affective pragmatics that establishes a relationship between illocutionary forces and emotional expressions, however, in the current framework, TAP is applied to explore its practical essence in establishing relationship between emotional prosody and affective pragmatics.

4. Discussion

There are countless studies based on speech act theory to understand the pragmatic meanings of utterances. Speech act theory focuses the discernment of meaning at utterance level but not at emotional level. There are numerous studies available on identification of emotions through prosodic modeling; however, there exist almost no study that centers emotions accompanied by utterances to derive deeper meaning at affective level. The current study is an

application of TAP, a theory analogous to speech act theory that functions at affective level. This study will shed light on different aspects of affective pragmatics as a novel area of research in pragmatics.

For, emotional prosody of speech, it is important to understand the significance of speech synthesis process. At University of Portsmouth, UK, it was reported that the use of vocal features along with emotional content is advantageous in synthesizing a natural speech and the employment of recorded voice for listeners could be a better option to determine the state of the speaker(Drahota et al., 2008).A strong correlation between affective state of the speaker and statistical measures of speech has been revealed in numerous studies; the most common measures are energy, pitch, articulation, and spectral configuration. For example, Sadness is associated with slow speaking rates and low standard deviation of pitch, while anger is usually connected with fast speaking rates and high standard deviation of pitch. Many valuable efforts have been done for emotion recognition in speech. Usually, mean, median, standard deviation, and percentile are the commonly employed statistical measures. The accuracy in order to achieve the emotional content is about 50% which is close to that of human judgment (Nogueiras et al., 2001).

In speech recognition systems, various techniques are employed for acoustic modeling, however, the successful ones are based on Hidden Markov Models (Rajeswari & Uma, 2012). In this study, HMM based synthesis using low level features will be employed for emotion recognition and further statistical parametric analysis. It is the most widely used speech synthesis system but it somehow lacks variations in prosodic parameters. Improved approaches to HMM based synthesis have now been introduced with high level linguistic feature extraction, and in this way, the quality of prosodic modeling has been improved. It has been found that the use of

explicit and implicit duration models in combination is instrumental in improving the quality of HMM based speech synthesis (Rajeswari & Uma, 2012). It has also been revealed that the HMM based on syllabic synthesis of utterance is good at estimating prosodic features because it improves the synthesis of missing units that make the algorithm unsupervised (Ronanki et al., 2014).

5. Conclusion

Theory of affective pragmatics (TAP) is a novel theory in pragmatics laying foundations of an independent sphere of knowledge known as affective pragmatics. Affective pragmatics is what emotional expressions mean in context. TAP is all about conceiving emotional expressions and deriving meaning in context. It foregrounds that emotional expressions have their own natural meanings and a natural meaning generally relies on statistical correlations (Scarantino, 2017a). Affective pragmatics is the least focused and ill-conceived area as its importance is yet not well understood and emphasized by the researchers. The studies available in this researchable domain are scarce and still there is no well-known study available on affective pragmatics that shows relationship between emotions and their pragmatic meanings. Hence, the abstruse power of emotional expressions is yet to be explored. From communicative point of view, the take home-message of TAP is that emotional expressions can do what exactly words can. The research on the present conceptual delineation will be influential in establishing a relationship between emotions and their pragmatic meanings to understand affective pragmatics at a deeper level.

6. Recommendations

The research on the similar outlines should determine how modern technology is instrumental in recognizing emotions and comprehending the knowledge of emotional prosody.

Various researches should also be carried out to know the pros and cons of usage of software technology in estimating emotional prosody. Further similar researches should broaden the horizon of affective pragmatics as an important sphere of linguistics and open up new doors of research in the field of pragmatics.

Conflict of Interest: The corresponding author, on behalf of all authors, confirms that there are no conflicts of interest to disclose.

References

Applebaum, L., Coppola, M., & Goldin-Meadow, S. (2014). Prosody in a communication system developed without a language model. *Sign Language & Linguistics, 17*. doi: 10.1075/sll.17.2.02app

Darwin, C. (2008). *The descent of man, and selection in relation to sex*: Princeton University Press.

Domaneschi, F., & Bambini, V. (2020). *Pragmatic Competence*.

Drahota, A., Costall, A., & Reddy, V. (2008). The Vocal Communication of Different Kinds of Smile. *Speech Communication, 50*, 278-287. doi: 10.1016/j.specom.2007.10.001

Fischer, A., & Sauter, D. (2017). What the Theory of Affective Pragmatics Does and Doesn't Do. *Psychological Inquiry, 28*, 190-193. doi: 10.1080/1047840X.2017.1338100

Lai, L.-F., & Gooden, S. (2016). Acoustic cues to prosodic boundaries in Yami: A first look. *Speech Prosody 2016*, 624-628.

Levinson, S. C. (1983). Pragmatics Cambridge University Press. *Cambridge UK*.

Liu, Y. (2006). Modeling prosody in speech processing. *The Journal of the Acoustical Society of America, 120*, 3006. doi: 10.1121/1.4787018

Martin, J. (2008). Prosodic ‘structure’: grammar for negotiation Prosodic ‘structure’: grammar for negotiation. *Ilha do Desterro*.

Mozziconacci, S. (2002). Prosody and emotions.

Nogueiras, A., Moreno, A., Bonafonte, A., & Mariño, J. B. (2001). *Speech emotion recognition using hidden Markov models.* Paper presented at the Seventh European conference on speech communication and technology.

Pan, S. (2002). *Prosody modeling in concept-to-speech generation*: Columbia University.

Prieto, P., & Roseano, P. (2018). Prosody: Stress, rhythm, and intonation (pp. 211-236).

Pronina, M., Hübscher, I., Vilà-Giménez, I., & Prieto, P. (2021). Bridging the Gap Between Prosody and Pragmatics: The Acquisition of Pragmatic Prosody in the Preschool Years and Its Relation With Theory of Mind. *Frontiers in Psychology*, *12*. doi: 10.3389/fpsyg.2021.662124

Rajeswari, K., & Uma, M. (2012). Prosody modeling techniques for text-to-speech synthesis systems–a survey. *International Journal of Computer Applications*, *39*(16), 8-11.

Ronanki, S., Watts, O., King, S., & Clark, R. (2014). Syllable based models for prosody modeling in HMM based speech synthesis. *criterion*, *9*(10), 11.

Scarantino, A. (2017a). How to do things with emotional expressions: The theory of affective pragmatics. *Psychological Inquiry*, *28*(2-3), 165-185.

Scarantino, A. (2017b). Twelve questions for the theory of affective pragmatics. *Psychological Inquiry*, *28*(2-3), 217-232.

Ververidis, D., & Kotropoulos, C. (2012). A State of the Art Review on Emotional Speech Databases.

Wichmann, A., Dehé, N., & Barth-Weingarten, D. (2009). Where Prosody Meets Pragmatics: Research at the Interface (pp. 1-20).

Wu, Y., Tessler, M. H., Asaba, M., Zhu, P., Gweon, H., & Frank, M. C. (2021). *Integrating emotional expressions with utterances in pragmatic inference.* Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society.

Zen, H., Nose, T., Yamagishi, J., Sako, S., Masuko, T., Aw, B., & Tokuda, K. (2007). The HMM-based Speech Synthesis System Version 2.0. 131-136.