DOI: 10.54392/ijll2513

Asian Research Association

Indian Journal of
LANGUAGE AND LINGUISTICS

Development of Test Stimuli and Assessing Acoustic Correlates of Affective Prosody in Kannada-Speaking Young Female Adults

S. Shalini [a,*], S. Srushti [a], Jayashree S Bhat [a]

[a] Nitte Institute of Speech and Hearing, Mangalore, Karnataka, India

[*] Corresponding author Email: shalini.sdss22@gmail.com

DOI: https://doi.org/10.54392/ijll2513

Received: 24-11-2024; Revised: 24-02-2025; Accepted: 07-03-2025; Published: 19-03-2025

Abstract: Purpose: Affective prosody, which conveys emotions through variations in stress and pitch, is crucial for social interactions and emotional communication. This study investigates the acoustic correlates of expressive affective prosody in Kannada-speaking young female adults. Methods: The study involved 31 Kannada-speaking young female adults performing an expressive affective prosody task (reading sentences) for four emotions: happy, sad, anger, and fear. The data collected was acoustically analysed based on a few acoustic parameters such as fundamental frequency (F0) and its range and contour; intensity (I0) and its range; and rate of speech (ROS) across the four emotions. Results: Statistical analysis revealed significant differences across emotions for all parameters, with rate of speech and I0 range showing the most variance. The frequency measures showed significant differences among almost all the comparisons. For the F0 contour, happy and anger exhibited rising and falling patterns, while sad and fear showed falling and rising patterns. The mean rate of speech was highest for anger and lowest for happy. Intensity measures showed significant differences, particularly between happy vs. sad, and sad vs. anger. These observations are consistent with previous research but also had a few contradictory findings. Conclusion: The study highlights distinct acoustic profiles for different emotions in Kannada among young female adults, aiding in the understanding of normal affective expression. This normative data on prosodic features can serve as a benchmark for identifying deviations in individuals with neuro communication disorders affecting their emotional prosody. These acoustic parameters can be used for objective assessments in clinical settings to detect early changes in emotional expression. Future research should include diverse age groups, both genders and various dialects to enhance the generalizability of findings.

Keywords: Affective Prosody, Emotional Prosody, Emotional Speech, Acoustic Correlates, Speech Prosody, Emotions

1. Introduction

Speech prosody is often considered to be 'the music of speech' (Hellerman, 2003). Prosodic cues play a major role in providing various information about the speaker such as age, gender, emotion, and situation. Affective prosody helps to convey speaker's emotional state. It is crucial for understanding and expressing emotions during conversation and to develop an emotional connection (Leentjens *et al.,* 1998). One of the important parts of social skill development is the ability to understand emotions during social interactions which require appropriate reception and expression of affective prosody (Brunswik, 2020; Scherer, 1994) and also to form emotional attachments with close relationships (Eggum *et al.,* 2011).

The expression of affective prosody for each emotion varies as each emotion has its own acoustic profile mainly characterized by variations in pitch, intensity and other related parameters. These acoustic correlates of affective prosody for different emotions have been started to be explored since the 1960s. Pitch and loudness acts as an important cue to convey the emotional state of the speaker. There are other acoustic cues which also help to express the emotion (Banse & Scherer, 1996) such as speaking rate, duration and intensity (Dellaert *et al.,* 1996; Grimm *et al.,* 2007; Morrison *et al.,* 2007; Paidi *et al.,* 2016). A review study (Paidi et al., 2016) on the acoustic profile of basic emotions concluded that anger and happy emotions generally exhibit higher average F0, wider F0 range, and higher intensity, while sad speech shows lower F0 and narrower intensity range. Fearful speech is characterised by much higher F0, much irregular F0 contour, normal intensity, and a faster speaking rate. However, one of the main disadvantages of previous studies is that the speech samples analyzed were mostly enacted by professional and non-professional actors rather than derived from natural settings. Therefore, generalizing these results is questionable.

Given these limitations, researchers have explored ways to improve the ecological validity of emotion recognition studies. A study compared a natural data set collected from a call centre (mainly anger and neutral emotion) and a pre-recorded available database (consisting of recordings from 12 non-professional actors and actresses conveying 6 types of emotions) (Banse & Scherer, 1996). The data were fed to an automatic emotion recognition system which classifies emotions automatically based on the frequency and intensity-related parameters. The natural set's classification rates were lower than anticipated due to various factors, including unbalanced data, the presence of background noise, and the difficulty in precisely measuring specific emotions. These findings highlight the challenges of accurately capturing emotional expressions in real-world contexts. Using data collected from real-world scenarios can offer a better representation of the expression of emotions. The induction of emotions in non-actors can assist in achieving better generalization, and a control over the variables. This underscores the need for understanding the acoustic correlates of affective prosody in the general population (non–actor). Such knowledge is also essential to establish a baseline against which deviations in disordered populations (aprosodia) can be identified and analyzed. A comprehensive understanding of the acoustic features that underlie affective prosody in the general population can serve as a valuable starting point in the exploration of disordered populations.

Building on this need, various standardized tests have been developed to assess affective prosody. The identification of the acoustic cues that correspond to various emotions can be achieved through the administration of a range of tests, often requiring participants to perform perceptual, reading or repetition tasks. However, studies focused on developing or validating test stimuli for assessing affective prosody are limited (Filippa *et al.,* 2022). The PEPS-C evaluates prosodic skills, including affective prosody, in individuals above 4 years (Filipe *et al.,* 2017; Martínez-Castilla & Peppé, 2008; Zanchi *et al.,* 2022). The Florida Affect Battery (FAB), designed for individuals above 17 years, assesses affective prosody through multiple subtests. The Aprosodia Battery distinguishes prosodic deficits due to hemispheric damage, focusing on production, comprehension, and discrimination tasks. Additionally, databases with prerecorded speech stimuli and validated sentence lists have been developed for assessing affective prosody in English (Castro & Lima, 2010; Darcy & Fontaine, 2020; Gong *et al.,* 2022). However, there are no such standardized test materials available in Indian languages for the assessment of affective prosody.

Recent studies have also examined how different populations perceive and process emotional prosody. One such (Darshan & Shilpashri N, 2023) investigated how Kannada-speaking adults perceive emotions such as happiness, sadness, anger, questions, and statements. The researchers developed five-to-six-word sentences for each emotion, spoken by a female native Kannada speaker. These sentences were validated by experienced speech-language pathologists and presented to participants for emotion identification. The study found that both male and female participants accurately identified emotions, with minor variations across age groups. Additionally, another study (Kao & Zhang, 2024) explored how listeners automatically process emotional prosody in speech, focusing on different neural responses for various prosodic categories and potential sex differences. The study found that both men and women exhibit distinct neural responses to different emotional tones, highlighting the importance of considering sex differences in prosody research. These findings further emphasize the need for research that systematically investigates expressive affective prosody in Kannada, particularly among young female adults, to bridge the gap between perceptual understanding and production-based acoustic analyses.

Since there is a research gap in this area, particularly in Indian languages, the current study aimed to explore the acoustic correlates of expressive affective prosody in Kannada-speaking young female adults by developing test stimuli and collecting data on a reading task for the expression of four basic emotions: 'happy,' 'sad,' 'anger,' and 'fear.' Given the lack of standardized assessment tools in Kannada, this study will contribute to the development of culturally and linguistically appropriate test materials, aiding both clinical and research applications in affective prosody assessment.

2. Methodology

The study was carried out in a sound-treated room in a multidisciplinary hospital set-up. The study design was descriptive research with a convenience sampling method. The number of participants included in the study was 31 females, in the age group of 18 to 25 years (Mean age= 20 years ± 1.4 years) (young adults) (Arnett, 2000), with representation from different departments (Dental Sciences, Physiotherapy and Anaesthesiology departments). The participants were selected through convenience sampling and were recruited only if they matched the inclusion criteria. All the participants were native Kannada speakers. The sample size was calculated using the G power analysis based on the pilot study.

The study's inclusion criteria required individuals without speech and language disorders to have normal hearing, cognitive, and mental abilities, ascertained by the WHO DAS (Disability Assessment Schedule) 2.0 questionnaire (*WHO Disability Assessment Schedule (WHODAS 2.0)*, n.d.). Additionally, participants had to be native Kannada speakers, with education above PUC and could read Kannada. Professional or non-professional actors, trained vocal artists, singers, and Speech-language pathologists were excluded from the study as they might not represent a natural population since they are trained in these aspects.

2.1 Procedure

The study was conducted only after obtaining ethical clearance from the University [REG. NO.:-EC/NEW/INST/2022/KA/0174].

2.1.1 Development of test stimuli

Three sentences for each emotion (happy, sad, anger, fear) in Kannada were created. All the sentences had emotion-laden words which depicted the target emotion. Content validation of the developed stimuli was carried out using the Delphi technique to choose the single best sentence for each emotion as it validates content systematically through expert review, ensuring it aligns with theoretical frameworks. In total, 12 sentences (3 for each emotion) were validated by five experts in the field (Speech-language pathologists). The experts were asked to rate the relevance of the stimuli to the target emotion using a 4-point rating scale (1- not relevant, 2-somewhat relevant, 3-relevant, 4- highly relevant). The content validation index (CVI) and Content validation ratio (CVR) were calculated. Sentences with a CVI of 0.78 or higher and a CVR greater than 0.49 were considered. Among them, the sentence with the highest value of CVI and CVR was taken as stimuli (one sentence for each emotion).

Following the selection of the most relevant sentence as a stimulus for each emotion based on CVR and CVI scores, the sentences with the next highest value of CVR and CVI scores were employed as a model for the participants during the administration of the original stimuli. The model sentences were produced by vocal artists. Five vocal artists (3 females and 2 males) with at least 2 years of performing experience were requested to produce the chosen sentences in the appropriate emotion and it was recorded in a sound-treated room. Post-recording, the audio recordings were validated. Three skilled vocal artists (minimum of two years experience) and three experts in the field (speech-language pathologists with a minimum of two years experience) assessed the recordings using a 4-point rating scale (1- not relevant, 2- somewhat relevant, 3- relevant, 4- highly relevant) and ranking methodology. The content validation index and ratio were calculated to select the best recording for the model.

2.1.2 Administration of test stimuli on young adults

The recording was carried out in a sound-treated room using the BESTOR professional condenser microphone using the PRAAT software with an ambient noise level of 30dB SPL. The participants were asked to sit comfortably in the chair and the microphone was placed 10cm away from the participant's mouth. The clinician first played the audio model (the pre-recorded speech) through a loudspeaker for the first emotion followed by which the sentence for the target emotion was shown on the laptop screen in a white background. The participants were asked to read the sentence with the target emotion 3 times. The model was played again if required but was not repeated more than once. The model for the next emotion was played only after recording the response for the previous emotion.

The instructions during the recording were given as follows, 'I will be playing an audio for you, You have to listen to it carefully and repeat the sentence presented on the screen three times in the same way as you heard before'.

2.1.3 Acoustic analysis

Acoustic analysis was carried out by PRAAT software version 6.1. 29. The best of three recordings for each emotion was taken for acoustic analysis, based on six acoustic parameters. The parameters include average F0 (Average Fundamental frequency), F0 range (Fundamental Frequency range), F0 contour (Fundamental frequency contour), Average I0 (Average Intensity), I0 range (Intensity range) and Rate of speech (ROS). These parameters were previously examined in a review study, and significant acoustical differences in speech across emotions were found (Paidi *et al.,* 2016). Average F0 is defined as the lowest frequency at which the speech signal repeats itself (O'Shaughnessy, 2000). The F0 range is the difference between the maximum and minimum F0. F0 contour is the shape or pattern of F0. Average intensity is the Volume or intensity of the speech. The intensity contour provides information that can be used to differentiate sets of emotions. The intensity range is the difference between the maximum and minimum intensity. ROS is the number of syllables produced per second. The difference in average F0, F0 range, F0 contour, average intensity, intensity range and ROS among the four emotions were analysed. Other parameters, such as jitter, shimmer, and formant frequencies, were not included as they were not found to be significant in representing differences between emotions (Paidi *et al.,* 2016).

2.1.4 Statistical analysis

The data was entered into SPSS version 20 and was subjected to statistical analysis. Descriptive statistics were done to compute the mean, median and standard deviation of the data. Normality tests including the Shapiro-Wilk test were carried out to know if the data was falling within normal distribution. The level of significance considered was 0.05. Since the data failed to meet the normality, non-parametric tests were conducted using the Friedman test to find the significant difference in average F0, F0 range, F0 contour, average intensity, intensity range and ROS among the four emotions.

3. Results

Descriptive analyses were carried out to find the mean, standard deviation, median and interquartile range for all the parameters (F0, F0 range, I0, I0 range, ROS) for each of the emotions. The results for the emotions 'happy', 'sad', 'anger', and 'fear' have been compiled in Table 1.

Table 1. Results of descriptive analysis for the four emotions

Emotions	**Happy**		**Sad**		**Anger**		**Fear**	
Parameters	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
F0	264.13	31.78	244.05	27.09	315.43	295.45	290.81	306.81
F0 range	171.32	77.78	154.53	79.87	218.08	82.80	161.59	81.33
I0	71.31	7.23	71.52	6.62	88.75	84.51	86.45	85.29
I0 range	42.38	7.09	31.18	8.75	44.34	6.37	41.18	7.57
ROS	6.25	1.03	7.25	1.2	7.25	1.2	7.3	1.17

Abbreviations: SD = Standard Deviation, IQR = Interquartile Range, F0 = Fundamental frequency, F0 range = Fundamental frequency range, I0 = Intensity, I0 range = Intensity range, ROS = Rate of speech

The Friedmann test was used to compare the following acoustic parameters - F0, F0 range, I0, I0 range, and ROS across four different emotions (Happy, Sad, Anger, and Fear). Emotions were compared in six primary pairs: 'happy' vs. 'sad,' 'happy' vs. 'anger', 'happy' vs. 'fear', 'sad' vs. 'anger', 'sad' vs. 'fear', and 'anger' vs. 'fear'. (Table 2) as detailed below.

Table 2. Comparison of different acoustic parameters

EMOTIONS	PARAMETERS – Chi-square value									
	F0	Cohen's d	F0 range	Cohen's d	I0	Cohen's d	I0 range	Cohen's d	ROS	Cohen's d
Happy vs. Sad	1.032*	0.21	0.645*	0.17	0.226	0.10	1.355**	0.24	1.097*	0.22
Happy vs. Anger	0.065	0.05	0.774*	0.18	0.806*	0.19	0.581	0.16	2.081**	0.30
Happy vs. Fear	1.419**	0.25	0.452	0.14	0.774*	0.18	0.194	0.09	1.016*	0.21
Sad vs. Anger	1.097*	0.22	1.419**	0.25	1.032*	0.21	1.935**	0.29	0.984*	0.21
Sad vs. Fear	0.387	0.13	0.194	0.09	0.548	0.15	1.161**	0.22	0.081	0.06

Abbreviations: F0 = Fundamental frequency, F0 range = Fundamental frequency range, I0 = Intensity, I0 range = Intensity range, ROS = Rate of speech.
Asterisks indicate the significance level: *P< 0.05, **P< 0.001.
Bold indicates higher effect size.

3.1 Fundamental Frequency [F0]

The mean F0 for females was higher for the emotion 'anger' followed by 'fear', 'happy', and 'sad' as shown in Table 1. The Friedmann test results showed a significant difference in F0 only when compared between 'happy' vs. 'sad', 'happy' vs. 'fear', 'sad' vs. 'anger', and 'anger' vs. 'fear' (Table 2).

3.2 Fundamental Frequency [F0] range

The emotion 'sad' had a narrower mean F0 range and 'anger' had a wider mean F0 range (Refer to Table 1 for the mean values). The Friedmann test revealed that the F0 range showed a significant difference when comparing emotions 'happy' vs. 'sad', 'happy' vs. 'anger,' 'sad vs. 'anger', and 'anger' vs. 'fear' (Table 2).

3.3 Intensity [I0]

The mean I0 for the emotions 'anger' and 'fear' was comparatively higher than other emotions (Refer to Table 1 for the mean values). The Friedmann test results showed a significant difference in I0 when compared between 'happy' vs. 'anger', 'happy' vs. 'fear', 'sad' vs. 'anger', and 'anger' vs. 'fear'. However, no significant difference was noticed when compared between 'happy' vs. 'sad', and 'sad' vs. 'fear' (Table 2).

3.4 Intensity [I0] range

The average I0 range for the emotion 'sad' was found to be narrower than other emotions, like males (Refer to Table 1 for the mean values). The only significant difference in the I0 range was observed when comparing 'happy' vs. 'sad', 'sad' vs. 'anger', 'sad' vs. 'fear', and 'anger' vs. 'fear', as shown in Table 2.

3.5 Rate of speech [ROS]

The ROS was higher for the emotion 'anger' and lowest for the emotion 'happy' in females (Refer to Table 1 for the mean values). The Friedmann test revealed significant differences in speech rate across all emotions except when compared between 'sad' and 'fear' (Table 2).

3.6 F0 contour

Three different patterns of F0 contour were recorded for the four emotions, which included rising and falling, falling and rising, and flat. The frequency of rising and falling patterns was high for the emotions 'happy' (71.0%) and 'anger' (90.3%), while the frequency of falling and rising patterns was high for the emotions 'sad' (77.4%) and 'fear' (51.6%). The frequency of flat patterns was found to be high for the emotion 'fear' (48.4%). To compare the F0 contour across the four emotions, Kendall's tau-b test was conducted.

A significant difference was found in the F0 contour for almost all comparisons of the emotions, but no significant difference was found when compared between 'happy' vs. 'anger' and 'sad' vs. 'fear' (Table 3).

Table 3. Results of the Kendall's tau-b test

EMOTIONS	F0 contour T^b
Happy vs. Sad	**4.909****
Happy vs. Anger	1.873
Happy vs. Fear	**6.689****
Sad vs. Anger	**6.992****
Sad vs. Fear	1.209
Anger vs. Fear	**13.537****

Abbreviations: Asterisks indicate the significance level: *$P< 0.05$, **$P< 0.001$.

Effect size for each parameter across emotion pairs were calculated by measuring Cohen's d value. The differences between emotions in acoustic parameters showed statistically small effect. Happy vs. Anger and Anger vs. Fear show the strongest effects, suggesting these pairs have more noticeable speech differences. The Rate of Speech (ROS) and Intensity (I0) parameters show relatively higher effects, indicating rate of speech and loudness may be more affected by emotion than pitch.

In summary, the study results found significant F0 differences between all pairs except 'happy' vs. 'anger,' and 'sad' vs. 'fear' emotions. F0 range differences were more pronounced between 'sad' vs. 'anger' and 'anger' vs. 'fear'. Intensity differences in males were significant for all emotion pairs 'happy' vs. 'sad' and 'sad' vs. 'fear.' The I0 range was significantly different for all pairs except 'happy' vs. 'anger'. ROS showed significant differences across most pairs. F0 contour patterns revealed high occurrences of rising and falling in 'happy' and 'anger' emotions and falling and rising in 'sad' and 'fear', with Kendall's tau-b test showing significant differences across most emotion pairs.

4. Discussion

The current study aims to explore the acoustic correlates of expressive affective prosody in Kannada-speaking female young adults, by employing speech samples during the reading task. Acoustic parameters such as F0, F0 range, I0, I0 range, and ROS were compared across different emotions. Emotions were compared with six primary pairs: 'happy' vs. 'sad,' 'happy' vs. 'anger', 'happy' vs. 'fear', 'sad' vs. 'anger', 'sad' vs. 'fear', and 'anger' vs. 'fear'. The sections below discuss how each acoustic parameter varied across different emotions.

4.1 Fundamental frequency [F0]

The emotion 'anger' had the highest F0, followed by 'fear', 'happy', and 'sad'. These findings align with the previous studies (Banse & Scherer, 1996; Hammerschmidt & Jürgens, 2007; Juslin & Laukka, 2001; Morrison *et al.,* 2007; Murray & Arnott, 1993; Paidi *et al.,* 2016). Few studies (Juslin & Laukka, 2001; Morrison *et al.*, 2007; Paidi *et al.,* 2016) have also noted that the emotion 'fear' has a higher F0 compared to the emotion 'happy' due to the greater intensity intended when expressing 'fear', a pattern also observed in the current study. When conducting pairwise comparisons, a significant difference in F0 was found when comparing 'happy' vs. 'sad', 'happy' vs. 'fear', 'sad' vs.

'angry', and 'angry' vs. 'fear'. However, no significant difference was observed when comparing 'happy' vs. 'angry', and 'sad' vs. 'fear'. Since the emotions 'happy' and 'anger' have almost similar acoustic characteristics with respect to F0, no significant difference is noted, as observed before (Banse & Scherer, 1996; Hammerschmidt & Jürgens, 2007; Juslin & Laukka, 2001; Morrison *et al.,* 2007; Murray & Arnott, 1993; Paidi *et al.,* 2016). However, both 'fear' and 'sadness' resulting in a lower F0 could not be explained. One probable reason could be that the model [male speaker] portrayed both 'sad' and 'fear' at a lower frequency, causing participants to often confuse these emotions. This might explain why there was no noticeable difference in F0 between 'sad' and 'fear'. The mean F0 for 'fear' was higher compared to 'sad' as stated before, but the standard deviation was large due to mixed performances by the participants. These differences can be attributed to the lack of control over the intensity of emotion expressed and how participants imitated the pre-recorded model leading to mixed results.

4.2 Fundamental frequency range [F0 range]

The next parameter that was compared across emotions was the F0 range. The results showed that there was a narrower mean F0 range for 'sad' and a wider mean F0 range for 'anger'. The result of pairwise comparison indicated a significant difference in the F0 range when comparing 'happy' vs. 'sad', 'sad' vs. 'anger', 'happy' vs. 'anger,' and 'anger' vs. 'fear.' These results are consistent with prior research (Hammerschmidt & Jürgens, 2007; Morrison et al., 2007; Murray & Arnott, 1993; Paidi et al., 2016) supporting the present findings There is no significant difference in the F0 range when comparing the emotions 'sad' vs. 'fear', as well as 'happy' vs. 'fear'. This lack of distinction can be attributed to inconclusive findings about the F0 range from the emotion 'fear', as the supporting studies (Morrison et al., 2007; Paidi et al., 2016) have also indicated that the F0 range of fear can be both wider and narrower, contributing to the absence of a conclusive finding.

4.3 Intensity [I0]

When comparing the mean I0 across different emotions a higher I0 for 'anger' and 'fear' was noticed differing from past findings (Banse & Scherer, 1996; Hammerschmidt & Jürgens, 2007; Juslin & Laukka, 2001; Morrison *et al.,* 2007; Murray & Arnott, 1993; Paidi *et al.,* 2016) that reported a higher I0 for 'happy' and 'anger', a lower I0 for 'sad', and a normal or lower I0 for fear. This discrepancy may be due to differences in participant characteristics, as previous studies often included professional actors, whereas this study involved non-actors. The intensity variations observed in actors may be attributed to their professional training, which could be artificial, whereas these variations may not be noticeable in a natural context. Participants may have produced fear-related expressions with greater intensity to emphasize the emotional salience of fear, aligning with findings from acoustic studies, including (Scherer, 2005), which suggest that fear is often conveyed through a combination of rising and falling intonations. This prosodic pattern not only enhances the perceptibility of fear but also reflects its urgent yet uncertain communicative intent, reinforcing the natural tendency to use heightened intensity as a signal of emotional significance. However, further research on the general population is required to fully understand this aspect. Pairwise comparisons revealed significant differences in I0 for the most emotion pairs. However, no significant difference emerged when comparing 'happy' vs. 'sad,' and 'sad' vs. 'fear,' further reflecting participant variability.

4.4 Intensity range [I0 range]

The difference in the intensity range [I0] across different emotions has not been extensively studied by researchers. The current study found a narrower I0 range for 'sad,' in line with a study done by Paidi et al., which reported a narrower range for 'sad' and a wider range for 'happy' and 'anger' (Paidi *et al.,* 2016). Although his study did not explicitly provide a specific reason for the observed differences among emotions, one plausible explanation inferred from it and applicable to the current research, is that high-arousal speech, produced by speakers in emotionally intense situations typically shows higher levels of vocal intensity and range. The results of the pairwise comparison demonstrated significant differences in nearly all comparisons, with the exception of 'happy' vs. 'anger' and 'happy' vs. 'fear'. The absence of differentiation between 'happy' vs. 'anger' may be rationalized by the broader I0 range shared by both emotions (Paidi *et al.,* 2016). Nevertheless, these findings lack empirical support and warrant further research.

4.5 Rate of speech [ROS]

The next parameter considered was the ROS, which was measured as the number of syllables produced per second. When the ROS was compared across emotions, the results revealed that the mean ROS was higher for 'anger', consistent with prior research (Banse & Scherer, 1996; Juslin & Laukka, 2001; Morrison *et al.,* 2007; Murray & Arnott, 1993; Paidi *et al.*, 2016). However, the lower ROS for 'happy' contradicts past studies associating happiness with faster speech. Similarly, pairwise comparisons revealed significant differences for most emotion pairs, except for 'sad' vs. 'fear,' These contradictions could be attributed to differences in the length of stimuli for different emotions. The stimuli used in the present study to assess the 'happy' emotion have fewer syllables compared to other stimuli, which may have contributed to the overall reduction in ROS. Another reason could be the linguistic differences between Kannada (Syllable timed) and English (Stress timed). Although syllable-timed languages sound faster, the actual rate of information transfer (measured in words per second) may be lower due to longer words and less compression compared to stress-timed languages. Research by Arvaniti (Arvaniti, 2009) supports this perspective, indicating that languages such as Greek, Spanish, and Italian, which are considered syllable-timed, exhibit higher syllable rates but may not necessarily convey information more rapidly than stress-timed languages like English. This is due to the inclusion of more syllables per interstress interval. However, further research is needed to confirm these findings.

4.6 Fundamental frequency [F0] contour

The parameter F0 contour exhibits three distinct patterns across the four emotions: rising and falling, falling and rising, and flat. Analysis revealed that the frequency of rising and falling patterns was predominant in emotions such as 'happy' and 'anger', whereas the frequency of falling and rising patterns was prevalent in emotions like 'sad' and 'fear'. These results were consistent with prior research findings found in studies conducted in English, Swedish, and Punjabi languages (Dawood *et al.,* 2004; Juslin & Laukka, 2001). Conversely, studies (Mini & Nataraja, 2000; Sandhya & Nataraja, 2000; Varshini & Nataraja, 2001) examining the F0 contour in South Indian Dravidian languages (Tamil, Telugu) revealed divergent outcomes. These languages predominantly exhibited a final falling pattern across all emotions, except for 'fear' and 'surprise'. Additionally, in the current study, a flat pattern was consistently observed across all emotions, especially for the emotion 'fear.' This phenomenon may be attributed to limited frequency variations, as the participants in the current study constituted of the general population [non–actor] thus unable to replicate precise modulations as elicited by the model. When the F0 contour was compared between different emotions, significant differences were found in almost all comparisons, except 'sad' vs. 'fear' in females, and 'happy' vs. 'anger'. While the similarity between 'happy' and 'anger' is expected, the lack of distinction between 'sad' and 'fear' contradicts previous research (Juslin & Laukka, 2001; Paidi *et al.,* 2016). This discrepancy could be because some participants had difficulty expressing 'fear' and 'sad' distinctly, even with the assistance of the model.

Overall, all the parameters exhibited significant differences across emotions which is consistent with previous research. However, some discrepancies may arise from differences in participant characteristics, as previous studies often recruited professional actors, while this study focused on non-actors. The findings contribute to the growing understanding of affective prosody in Kannada and highlight the need for further research with varied participant groups to enhance generalizability.

5. Conclusion

The present study represents the initial attempt to explore the acoustic correlates for affective prosody related to four fundamental emotions in Kannada, aiding in forming a database to distinguish individuals with normal affective expression and those with affective aprosodia. This normative data on prosodic features can serve as a benchmark for identifying deviations in individuals with neuro communication disorders (Parkinson's disease etc.,) affecting their emotional prosody. These acoustic parameters can be used for objective assessments in clinical settings to detect early changes in emotional expression by designing emotion recognition softwares.

However, the current study has a few limitations. The first major limitation of the study is the lack of male participants, which should be addressed in future studies. The length of the stimuli used could potentially act as a limitation, as it can impact the results, especially for ROS. Since the participants were recruited from the Mangalore region only, the obtained results cannot be generalized to all Kannada speakers due to dialectal variations. Future research can recruit participants, both males and females, from different parts of Karnataka, and analysis can be done separately for different dialects (Dakshina Kannada, Uthra Kannada and Madhya Kannada) for generalizing the results. Further, the study can be replicated in the clinical population, such as patients with Parkinson's disease, to identify the clinical validity of the test tool.

References

Arnett, J.J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American psychologist*, 55(5), 469. https://psycnet.apa.org/doi/10.1037/0003-066X.55.5.469

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, 66(1-2), 46-63. https://doi.org/10.1159/000208930

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology*, 70(3), 614-636. https://psycnet.apa.org/doi/10.1037/0022-3514.70.3.614

Brunswik, E. (2020). Perception and the Representative Design of Psychological Experiments. In Perception and the Representative Design of Psychological Experiments. *University of California Press*. https://doi.org/10.1525/9780520350519

Castro, S.L., Lima, C.F. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody. *Behavior Research Methods*, *42*(1), 74-81. https://doi.org/10.3758/BRM.42.1.74

Darcy, I., Fontaine, N.M.G. (2020). The Hoosier Vocal Emotions Corpus: A validated set of North American English pseudo-words for evaluating emotion processing. *Behavior Research Methods,* 52(2), 901–917. https://doi.org/10.3758/s13428-019-01288-0

Darshan, D., Shilpashri, H.N. (2023). Perception of Vocal Expression of Emotions in Kannada Speaking Healthy Adults. *International Journal of Health Sciences and Research,* 13(3), 75–79. https://doi.org/10.52403/ijhsr.20230308

Dawood, H., Shahid, R., Ahmed, T. (2004). Intonation Patterns in Punjabi.

Dellaert, F., Polzin, T., Waibel, A. (1996). Recognizing emotion in speech. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96, IEEE,* USA. https://doi.org/10.1109/ICSLP.1996.608022

Douglas O'Shaughnessy, (2000) "Speech Analysis," in Speech Communications: Human and Machine , IEEE, 173-227, https://doi.org/10.1109/9780470546475.ch6

Eggum, N.D., Eisenberg, N., Kao, K., Spinrad, T.L., Bolnick, R., Hofer, C., Kupfer, A.S., Fabricius, W.V. (2011). Emotion understanding, theory of mind, and prosocial orientation: Relations over time in early childhood. *The Journal of Positive Psychology,* 6(1), 4–16. https://doi.org/10.1080/17439760.2010.536776

Filipe, M.G., Peppé, S., Frota, S., Vicente, S.G. (2017). Prosodic development in European Portuguese from childhood to adulthood. *Applied Psycholinguistics,* 38(5), 1045–1070. https://doi.org/10.1017/S0142716417000030

Filippa, M., Lima, D., Grandjean, A., Labbé, C., Coll, S.Y., Gentaz, E., Grandjean, D.M. (2022). Emotional prosody recognition enhances and progressively complexifies from childhood to adolescence. *Scientific Reports,* 12(1), 17144. https://doi.org/10.1038/s41598-022-21554-0

Gong, B., Li, N., Li, Q., Yan, X., Chen, J., Li, L., Wu, X., & Wu, C. (2022). The Mandarin Chinese auditory emotions stimulus database: A validated set of Chinese pseudo-sentences. *Behavior Research Methods,* 55(3), 1441–1459. https://doi.org/10.3758/s13428-022-01868-7

Grimm, M., Kroschel, K., Mower, E., Narayanan, S. (2007). Primitives-based evaluation and estimation of emotions in speech. *Speech Communication,* 49(10), 787–800. https://doi.org/10.1016/j.specom.2007.01.010

Hammerschmidt, K., Jürgens, U. (2007). Acoustical Correlates of Affective Prosody. *Journal of Voice*, 21(5), 531–540. https://doi.org/10.1016/j.jvoice.2006.03.002

Hellerman, J. (2003). Ann K. Wennerstrom, The music of everyday speech: Prosody and discourse analysis. *Oxford University Press*, New York. https://doi.org/10.1017/S0047404503244055

Juslin, P.N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. Emotion, 1(4), 381–412. https://doi.org/10.1037/1528-3542.1.4.381

Kao, C., Zhang, Y. (2024). Sex Differences in Processing Emotional Speech Prosody: Preliminary Findings from a Multi-Feature Oddball Study. *Brain Sciences*, 14(12), 1216. https://doi.org/10.3390/brainsci14121216

Leentjens, A.F.G., Wielaert, S.M., Van Harskamp, F., Wilmink, F.W. (1998). Disturbances of affective prosody in patients with schizophrenia; a cross sectional study. *Journal of Neurology, Neurosurgery & Psychiatry,* 64(3), 375–378. https://doi.org/10.1136/jnnp.64.3.375

Martínez-Castilla, P., Peppe, S. (2008). Developing a test of prosodic ability for speakers of Iberian Spanish. *Speech Communication*, 50(11), 900–915. https://doi.org/10.1016/j.specom.2008.03.002

Mini, N., Nataraja, N.P. (2000) Intonation in Malayalam - Some aspects. Research at A.I.I.S.H Dissertation Abstracts, 90-91.

Morrison, D., Wang, R., De Silva, L.C. (2007). Ensemble methods for spoken emotion recognition in call-centres. *Speech Communication*, 49(2), 98–112. https://doi.org/10.1016/j.specom.2006.11.004

Murray, I.R., & Arnott, J.L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, 93(2), 1097–1108. https://doi.org/10.1121/1.405558

Paidi, G., Kadiri, S., Yegnanarayana, B. (2016). Analysis of Emotional Speech—A Review. Toward Robotic Socially Believable Behaving Systems, 205–238. https://doi.org/10.1007/978-3-319-31056-5_11

Sandhya, C. S., & Nataraja, N. P. (2000). Intonation in Telugu—Some aspects, Research at AIISH Dissertation Abstracts, 90-91, Vol IV. (D 403).

Scherer, K.R. (1994). Affect bursts. In Emotions: Essays on emotion theory. *Lawrence Erlbaum Associates*, 161-193.

Scherer, K.R. (2005). What are emotions? and how can they be measured?. Social Science Information, 44(4), 695–729. https://doi.org/10.1177/0539018405058216

Varshini, M. J., & Nataraja, N. P. (2001). Some aspects of intonation in Tamil. Dissertation submitted to. University of Mysore, Mysore.

WHO. (2023). Measuring health and disability : manual for WHO Disability Assessment Schedule (WHODAS 2.0). *World Health Organization*.

Zanchi, P., Peppé, S., Zampini, L. (2022). Prosodic development from 4 to 10 years: Data from the Italian adaptation of the PEPS-C. Speech Communication, 144, 10–19. https://doi.org/10.1016/j.specom.2022.08.007

Author contributions

S. Shalini: Conceptualisation, Data curation, formal analysis, investigation, methodology, Project administration, Writing – original draft. S. Srushti: Conceptualisation, Writing – Review & Editing. Jayashree S Bhat: Conceptualisation, Writing – Review & Editing.

Informed consent and consent for publication

The Ethical Committee of Nitte University, Mangalore, India [REG] approved this study [NO.: EC/NEW/INST/2022/KA/0174]. All subjects voluntarily participated in the study and signed an informed consent form.

Funding

This study did not receive any funding.

Declaration of generative AI in scientific writing

While preparing this work, the author[s] used ChatGPT [https://chat.openai.com/] for writing assistance at the drafting stage of the writing process.

Has this article been screened for Similarity?

Yes

Conflict of interest

The Authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

About The License