

# THE INTONATION OF THAI EMOTIONAL SPEECH

Apiluck Tumtavitikul and Kanlayarat Thitikannara

Linguistics Dept., Kasetsart University, Bangkok, Thailand 10900  
email: fhumalt@ku.ac.th

## Abstract

This paper studies the acoustics of Thai intonation in four types of emotional speech; anger, surprise, happiness and sadness. The subjects are radio performers, one male and one female aged 30 - 40 years. The results confirm that the speaking rate, average pitch, and pitch range of the intonation as well as amplitude of the utterance varies according to the type of emotion involved. Anger and surprise are in agreement with Luksaneeyanawin (1983). Anger and sadness are in agreement with Cahn (1988). The pitch for the male speaker shows two separated registers; anger, surprise and happiness are separate in a higher pitch level whereas neutral speech and sadness employ a lower pitch level. Thus, the overall pitch range for all emotional types combined for the male speaker is broader than that of the female speaker. However, when considered separately, the pitch range for each emotional type is quite comparable between male and female speakers. The Highs and Lows of the intonation vary greatly among utterances of the same emotional type unlike the intonation of Thai declarative sentences without emotion superimposed where the Highs and Lows are clearly rule-governed. Nevertheless, the results of this study have implications for speech technology. Sentences can be manipulated in the dimensions of time, fundamental frequency and fundamental frequency range as well as amplitude contour according to the emotional type targeted for vividness and naturalness.

## 1. Introduction

Emotions are universally expressed both verbally and non-verbally. Emotional speech is usually coupled with facial expressions and may also include body gestures. Emotion involves our entire beings, the brain, heart beat rate, blood pressure, muscles movements, etc. Emotions can be pleasant or unpleasant. Most common emotions found in the 32 multilingual databases compiled by Ververidis and Kotropoulos (2003) are: anger, sadness, happiness, fear, disgust, surprise, boredom and joy. This paper studies Thai Intonation in four of these most commonly expressed emotions in speech, namely, anger, sadness, happiness and surprise.

## 2. Data

The data were collected from two native speakers, one male and one female aged 30-40 years. Both are professional radio performers. Four one mono-syllabic word utterances (two proper nouns and two verbs), and three declarative utterances with three, five and six mono-syllabic words, all together seven utterances were said normally without emotion and in all four emotions; anger, surprise, sadness, and happiness, five times each. A total of 175 (7 x 5 x 5) test-tokens were collected from each speaker. All the 350 (175x2) test-tokens from both speakers were recorded using Sonic Foundry Vegas Video 3.0 program at a professional studio. The data were then analyzed acoustically with Kay Elemetrics

CSL 3700 program. For each test-token, fundamental frequency (F0), duration, and amplitude were investigated and measured at end points and at every 10% of the entire utterance. The data was then analyzed statistically. Some of the results are shown in tables 1-5 and figures 1-4 in the following section.

Table 1: Average utterance duration (in second) for all types of utterances combined in neutral speech and in each emotional type; anger, surprise, happiness, and sadness. (male and female combined)

Duration (second)	Neutral	Anger	Surprise	Happiness	Sadness
Average	0.59	0.52	0.49	0.73	0.61
S.D.	0.36 (n =70)	0.35 (n = 70)	0.32 (n =70)	0.35 (n=70)	0.42 (n=70)

Table 2: Average Pitch Range (in semi-tone) for all types of utterances combined in neutral speech and in each emotional type; anger, surprise, happiness, and sadness. (male and female combined)

Pitch Range (semitone)	Neutral	Anger	Surprise	Happiness	Sadness
Average	5.40	7.11	5.48	7.49	5.49
S.D.	2.51 (n=70)	2.20 (n=70)	2.12 (n=70)	3.40 (n=70)	1.75 (n=70)

Table 3: Average F0 (in Hertz) for one-word utterances (2 proper names and 2 verbs combined) and multiple-word utterances (3, 5 and 6 word utterances combined). (male and female)

F0 (Hertz)	Utterances	Neutral	Anger	Surprise	Happiness	Sadness
m	one-word	118.75	210.35	200.12	208.70	125.94
	multiple-word	133.62	228.27	227.28	226.72	128.84
f	one-word	205.67	260.04	204.67	189.74	193.53
	multiple-word	207.47	235.72	209.11	217.22	202.21

Table 4: Maximum and Minimum Amplitude (dB) for all utterances combined in all types of emotion. (male and female)

Speaker	Amplitude (dB)	Neutral	Anger	Surprise	Happiness	Sadness
male	maximum	72.84	83.60	75.46	79.07	71.54
	minimum	46.51	50.78	50.45	44.47	43.10
female	maximum	67.30	75.85	72.97	70.73	67.80
	minimum	35.93	46.55	42.42	38.79	33.14

Table 5: Average Amplitude (dB) for all utterances combined in each type of emotions. (male and female)

Speaker	Amplitude (dB)	Neutral	Anger	Surprise	Happiness	Sadness
male	average	62.32	72.40	65.87	67.24	58.71
	S.D.	2.04 (n=35)	3.10 (n=35)	2.70 (n=35)	3.87 (n=35)	3.16 (n=35)
female	average	57.80	66.58	62.89	61.88	54.19
	S.D.	1.82 (n=35)	3.70 (n=35)	3.09 (n=35)	2.72 (n=35)	4.08 (n=35)

### 3. Discussion

Table 1 and figure 1 show the duration for each emotional type compared with normal speech for one-word and multiple word utterances combined for both speakers combined. Happiness is expressed with the longest utterance duration and surprise is expressed with the shortest utterance duration. Sadness seems to lean towards longer duration than neutral speech and anger is shorter than neutral speech.

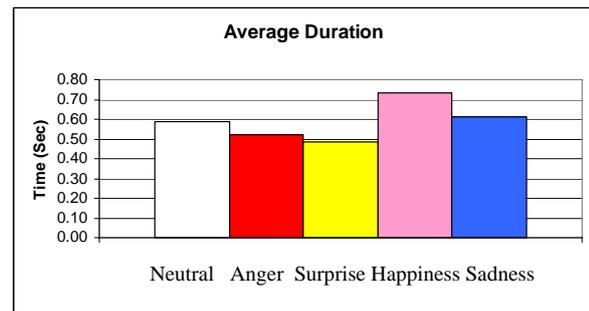


Figure 1: Average utterance duration (in seconds) for all type of utterances, in neutral speech and in each emotional type. (male and female combined)

When the average number of utterances per second is calculated, the precedence order for speaking rate, for both speakers combined, and for all types of utterances is:

**happiness > sadness > neutral > anger > surprise**  
**slowest -----> fastest speaking rate**  
**(utterance/second)**

Table 2 and figure 2 show the average pitch range in semitones for each emotional type for both speakers, in both one-word and multiple-word utterances. Anger is expressed with larger pitch range when compared with other types of emotion. On the other hand, sadness is expressed with smaller pitch range when compared with other emotional types. Happiness is expressed with the largest pitch range (7.49S). However, happiness seems to fluctuate with either very large or small pitch range (S.D. =3.40S) whereas surprise leans toward median and smaller pitch range. The precedence order for average pitch range for both speakers, all types of utterances combined, may be summarized as follows.

**happiness > anger > sadness / surprise > neutral**  
**largest -----> smallest pitch range**  
**(semitones)**

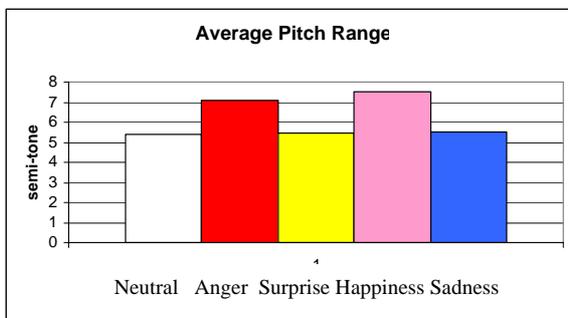


Figure 2: Average Pitch Range (in semi-tone) for all type of utterances in neutral speech and in each emotional type; anger, surprise, happiness, and sadness.(male and female combined)

From table 3, the average fundamental frequency (F0) for each emotional type for the male speaker in precedence order when compared with neutral speech is anger > happiness > surprise > sadness > neutral for one-word utterances. For multiple-word utterances, anger > surprise > happiness > neutral > sadness. For the female speaker, the precedence order is anger > neutral > surprise > sadness > happiness for one-word utterances. And, anger > happiness > surprise > neutral > sadness for multiple-word utterances. For both speakers, in both one-word and multiple-word utterances, anger is expressed with highest average pitch and sadness is expressed with lower average pitch when compared with other emotional types. Happiness and surprise lean towards a higher pitch level than neutral speech. The precedence order for average pitch, for both speakers, in all types of utterances, may be summarized as follows.

**anger > happiness/surprise > neutral/sadness**  
**highest -----> lowest pitch**  
**(Hertz)**

It is noticeable that the average pitch for the male speaker separates into two Fo levels ; anger, happiness and surprise average around ±217 Hz. whereas sadness and neutral speech average around ±127 Hz. The overall pitch range in semitones for all emotional types combined for the male speaker is broader than that of the female speaker. However, when considered separately, the pitch range in semitones for each emotional type is quite comparable between male and female speakers.

From figure 3 and table 4, the highest maximum amplitude in all types of utterances combined is found in the expression of anger in both male and female speakers. And the lowest minimum amplitude is found in the expression of sadness in both male and female speakers.

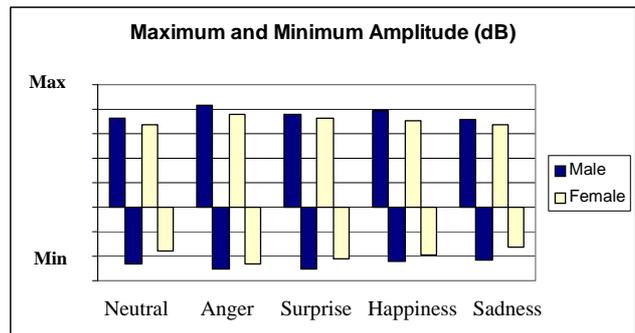


Figure 3: Maximum and minimum amplitude (in deci-Bel) for all types of utterances (one-word and multiple-word) combined in neutral speech and in each emotional type. (male and female).

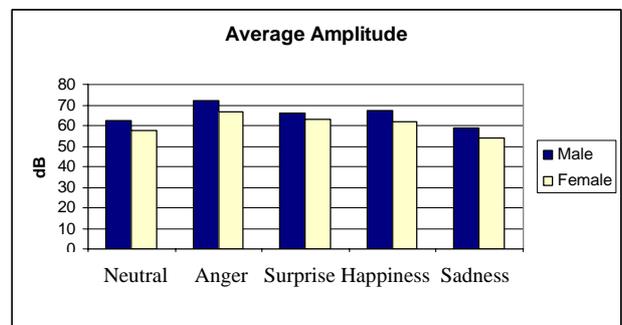


Figure 4: Average amplitude (in deci-Bel) for all types of utterances (one-word and multiple-word) combined in neutral speech and in each emotional type. (male and female).

From figure 4 and table 5, the average amplitude is highest in the expression of anger in both male and female speakers in all utterances combined. And sadness is found to be with lowest average amplitude. The precedence order of average amplitude in all types of emotions is as follows:

**anger > happiness/surprise > neutral > sadness**  
**highest -----> lowest amplitude**  
**(dB)**

## 4. Conclusion

Luksaneeyanawin (1983) in her studies of Thai intonation found that in expressing anger, the average pitch is either higher (or lower) than neutral speech, with a wider pitch range, either longer (or shorter) duration and a very high degree of loudness. For surprise, the average pitch is higher than neutral speech, with a narrower pitch range, either shorter (or longer) duration and a higher (or lower) degree of loudness.

Cahn (1988) in her studies of emotions expressed in English intonation found that anger is expressed in a faster and louder speech with a larger pitch range and higher average pitch than in neutral speech. The pitch contour fluctuates more with greatest energy found in higher frequencies. And for grief, the speech is found to be slow with low pitch and weak high frequencies.

The data found in our two Thai speakers as shown in tables 1-5 and figures 1-4 above are, more or less, in agreement with Luksaneeyanawin (1983) for anger and surprise. Our data are also in agreement with Cahn (1988) for anger and sadness. Happiness awaits comparison in continuing work. For our two speakers, anger is expressed with highest amplitude, highest average pitch, larger pitch range, and faster speaking rate when compared with other emotional types and neutral speech. Surprise is expressed with a higher average pitch, higher amplitude, and shorter duration than neutral speech. Our data show a smaller pitch range when compared with other emotional types but larger than neutral speech for surprise. Sadness is expressed with lowest average pitch, lowest amplitude, medium average pitch range, and slower speaking rate when compared with other types of emotion. The average pitch for sadness is comparable to that of neutral speech in Thai. Happiness is found to be expressed with slowest speaking rate, largest pitch range, higher average pitch and higher amplitude than neutral speech in our Thai speakers. This is yet to be compared with other data

As for the  $F_0$  contour, there is no uniformity in the present data for each type of the emotional speech observed in either male or female speaker. Unlike the  $F_0$  contour of neutral declarative statements studied by Tumtavitikul and Thitikannara (2006) where the highs and lows conform to the universal tendency of  $F_0$  declination in declarative sentences (Hirst and Di Crito, 1998) and the highs and lows are clearly rule-based even with focus shifted. The highs and lows in the present data do not form a unified pattern in any type of the four emotions observed. The shapes of the contour vary greatly among the same type of emotion in both male and female speakers. This is no surprise since emotions are subject to mood and attitude as well as environments. Emotional pitch contours may vary greatly within and among individuals (Ladefoged, 2006).

Nevertheless, it may be induced from the comparable features found between Thai and English in the two types of emotional speech, anger and sadness, that emotional speech may have a universal tendency of expression. The uniformity of expression may not be found in the intonation contour per se but in other acoustic cues e.g., speaking rate,  $F_0$  range, average  $F_0$ , and amplitude contour. This is readily explainable since emotions, which affect humans mentally and psychologically are impulses or natural responses to stimuli.

Ververidis, Kotropoulos and Pitas (2004) in their studies of automatic classification of emotional speech in the Danish language on five emotional states: anger, happiness, neutral, sadness, and surprise, reported accuracy rates of classification between 20-52%. The set of features used for speech generations in their studies were mainly abstracted from the pitch and energy contours of natural speech.

Vroomen, Collier and Mozziconacci (1993) in their acoustics studies of intonation in emotional speech in Dutch found that intonation and duration are sufficient to express emotions by re-synthesizing the intonation contour copied from natural speech, with proper time alignment, onto monotonous utterances. The rule-based manipulations of pitch and duration were found to suffice in representing emotions. The parameters for their synthesis were the type of emotion targeted, excursion size, the key in the register, and the optimal duration of the utterances.

The findings of this study on Thai intonation in four emotional ways of speaking have implications for Thai speech technology. Utterances can be manipulated in the dimensions of time for utterance duration, average fundamental frequency and fundamental frequency range as well as amplitude to represent emotional speech according to the type of emotion targeted. Conversely, these features can be used as parameters for automatic recognition/classification of emotional speech. Further studies are necessary to derive the rules governing the alignments of these parameters with the utterances.

Tumtavitikul and Thitikannara (2006) show that the highs and lows of the pitch contour of Thai intonation in neutral declarative sentences are rule-governed and ToBI can be used as a model for labeling and re-synthesizing the intonation contour. In this same manner, it is hoped that future research in emotional speech in Thai intonation will arrive at such rules.

## 5. Acknowledgements

This paper is a part of the research project, Thai Intonation in Different Emotions, funded by Kasetsart University Research and Development Institute grant no.  $\pi$ - $\nu$  (uu.) 1.47 to Apiluck Tumtavitikul. Prior presentations of the data and findings were made at the ICSTLL-39, Seattle, Washington, USA, 14-17 September, 2006, the Thailand Research Expo 2006, 9-13 September 2006, Bangkok, Thailand and the 6<sup>th</sup> National Symposium on Graduate Research, October 13-14, 2006, Bangkok, Thailand. Samples of stylized fundamental frequency contour and sample sound files are available at <http://pirun.ku.ac.th/~fhumalt/Exhibition/ICSTLL-39%20present-Sept%202006/Thai%20Intonation%20in%20Four%20Emotions.ppt>

## 6. References

- Cahn, Janet. (1988). From Sad to Glad: Emotional Computer Voices. *Proceedings of Speech Tech '88, Voice Input/Output Applications Conference and Exhibition.* ( pp. 35-37). New York City.

- Hirst, D. and Di Cristo, A. (1998). A Survey of Intonation Systems. In Hirst and A. Di Cristo (Eds.), *Intonation Systems*. Cambridge: Cambridge University.
- Kanlayarat, Thitikannara. (2006). *Thai Intonation in Different Emotions*. Unpublished MA Thesis, Kasetsart University.
- Ladfoegd, Peter. (2006). *A Course in Phonetics* (5<sup>th</sup> ed.). Boston: Thomson Wadsworth.
- Luksaneeyanawin, Sudaporn. (1983). *Intonation in Thai*. Unpublished Doctoral Dissertation, University of Edinburgh.
- Tumtavitikul, Apiluck and Thitikannara, Kanlayarat. (2006). Stress and Intonation in Thai. *Journal of Language and Linguistics*, 24.2.59-76. (in Thai)
- Ververidis, Dimitrios, Kotropoulos, Constantine, and Pitas, Ioannis. (2004). Automatic Emotional Speech Classification. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2004*. Montreal, Canada.
- Ververidis, Dimitrios and Kotropoulos, Constantine. (2003). *A State of the Art Review on Emotional Speech Databases*, from <http://poseidon.dsd.auth.gr/EN/index.html>
- Vroomen, Jean, Collier, Rene, and Mozziconacci, Sylie (1993). Duration and Intonation in Emotional Speech. *Proceedings of Eurospeech 1993*, (pp.577–580). Berlin, Germany.