# Thai Intonation in Four Emotions

Apiluck Tumtavitikul and Kanlayarat Thitikannara
Linguistics Dept., Kasetsart University
Bangkok, Thailand 10900
*email: fhumalt@ku.ac.th*

# Our Quests

- How do we express emotions in speech in a tonal language such as Thai?

- Other than words, what are the features in speech that are cues for emotions expressed by the speaker and taken up by the listener?

- Are these acoustic cues universal or language specific?

# Assumptions

- There is uniqueness in an overall pattern of each type of emotional intonation.

- Intonation is the cue to emotional speech.

# Investigations

- Four types of emotions most commonly found in 32 languages databases were selected;

  Anger, Surprise, Happiness, Sadness.

- Subjects: two radio performers, male and female aged 30-40 years.

Tumtavitikul & Thitikarnnara 2006

Test tokens:  4 one-word  and 3 multiple- word utterances all said in neutral and in four emotions, five times each, a total of 350 (7x5x5x2) utterances were investigated.

Acoustics measurments and calculations:
   - average Fo (Hertz),
   - Fo range (semitone),
   - maximum, minimum amplitude (dB),
   - average amplitude (dB),
   - utterance duration (seccond),
   - speaking rate (utterance/second)

# Results:
# Average Utterance Duration

| Duration | Neutral | Anger | Surprise | Happiness | Sadness |
|---|---|---|---|---|---|
| **Average** | **0.59** | **0.52** | **0.49** | **0.73** | **0.61** |
| S.D. | **0.36**<br>**(n =70)** | **0.35**<br>**(n = 70)** | **0.32**<br>**(n =70)** | **0.35**<br>**(n=70)** | **0.42**<br>**(n=70)** |

**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)
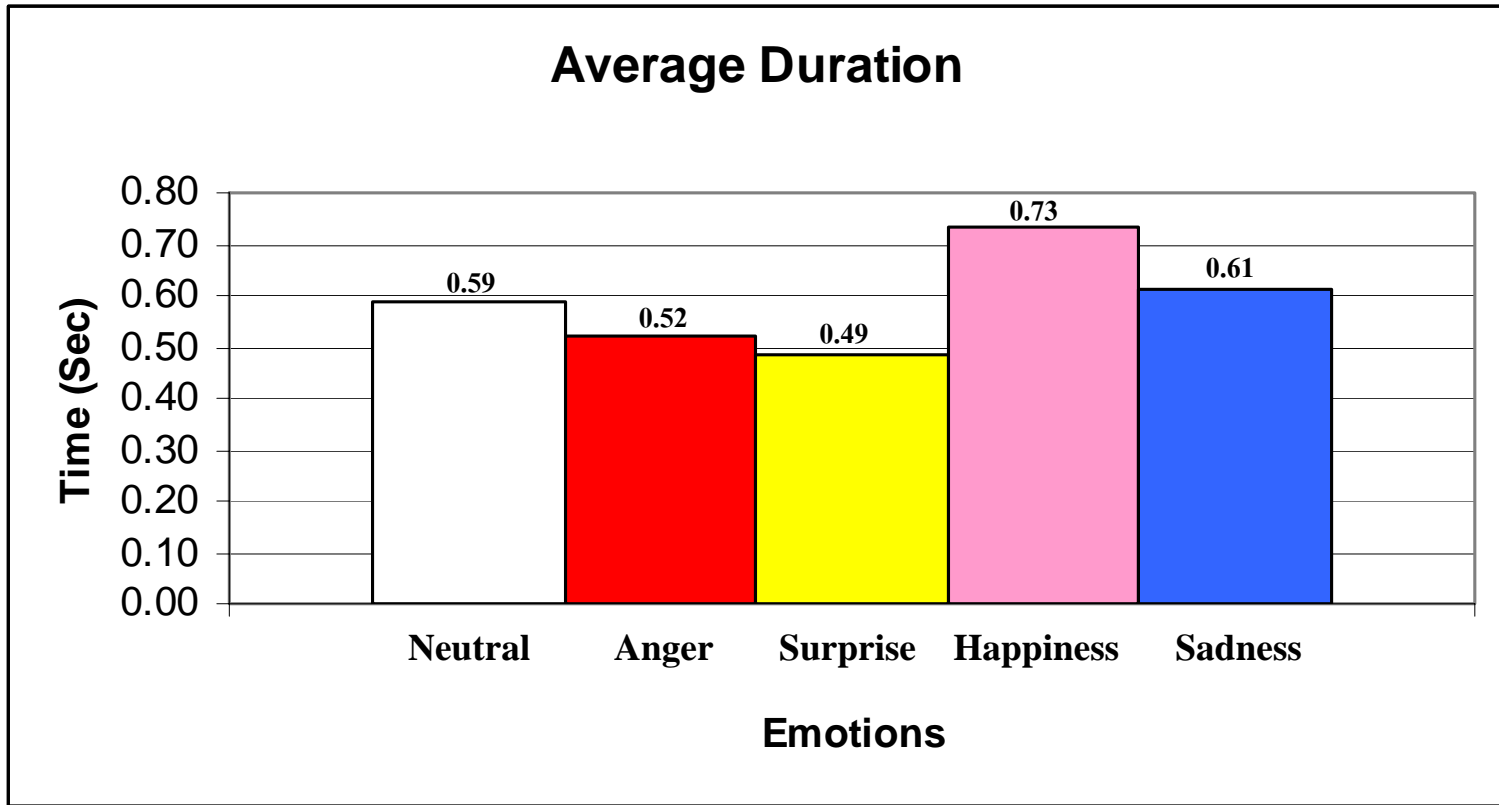
Tumtavitikul & Thitikarnnara 2006

**Figure 1**:: Average utterance duration (in second) for all types of utterances combined (male and female combined)

Tumtavitikul & Thitikarnnara 2006

# Speaking Rate

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

Tumtavitikul & Thitikarnnara 2006

# Average Pitch Range

| Pitch Range | 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 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)
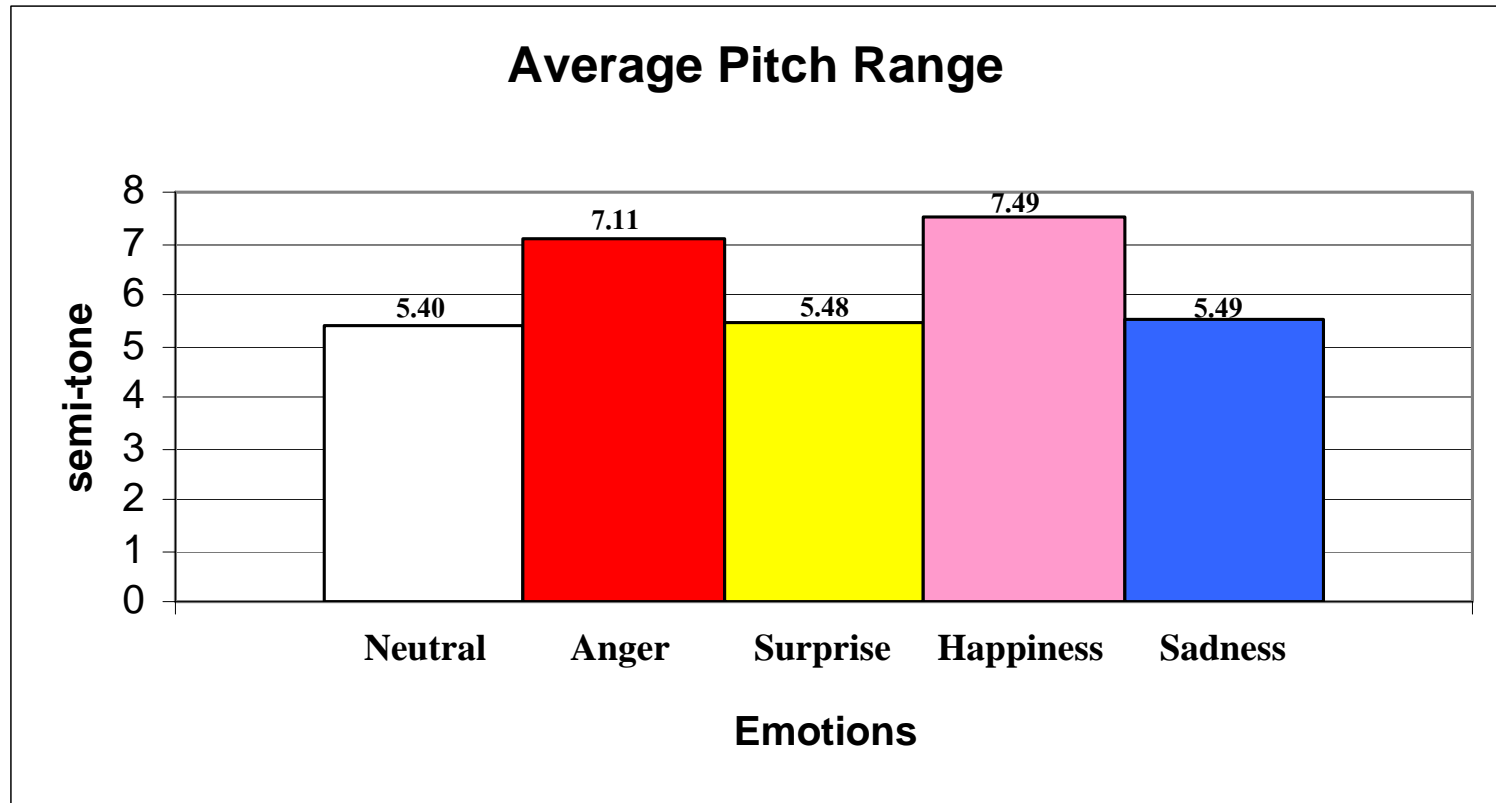
Tumtavitikul & Thitikarnnara 2006

**Figure 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)

Tumtavitikul & Thitikarnnara 2006

# Average Pitch Range

happiness > anger > sadness/surprise > neutral

largest           -------------->                smallest   pitch range

(semitones)

# Average Pitch

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

Tumtavitikul & Thitikarnnara 2006

# Average Pitch

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

# Maximum & Minimum Amplitude

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

Tumtavitikul & Thitikarnnara 2006
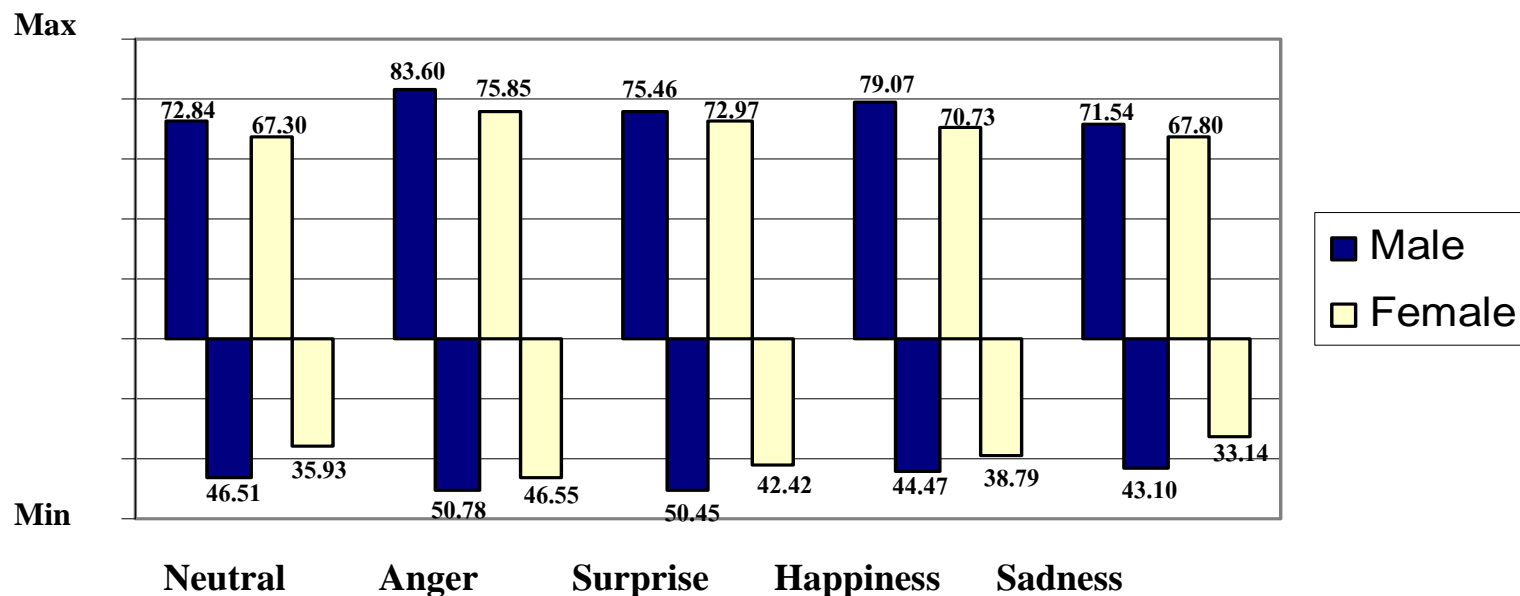
**Figure 3**: Maximum and Minimum Amplitude (dB) for all utterances combined in all types of emotion. (male and female)

# Average Amplitude

| Speaker | Amplitude | Neutral | Anger | Surprise | Happiness | Sadness |
|---|---|---|---|---|---|---|
| Male | $\overline{X}$ | 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 | $\overline{X}$ | 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) |

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

Tumtavitikul & Thitikarnnara 2006
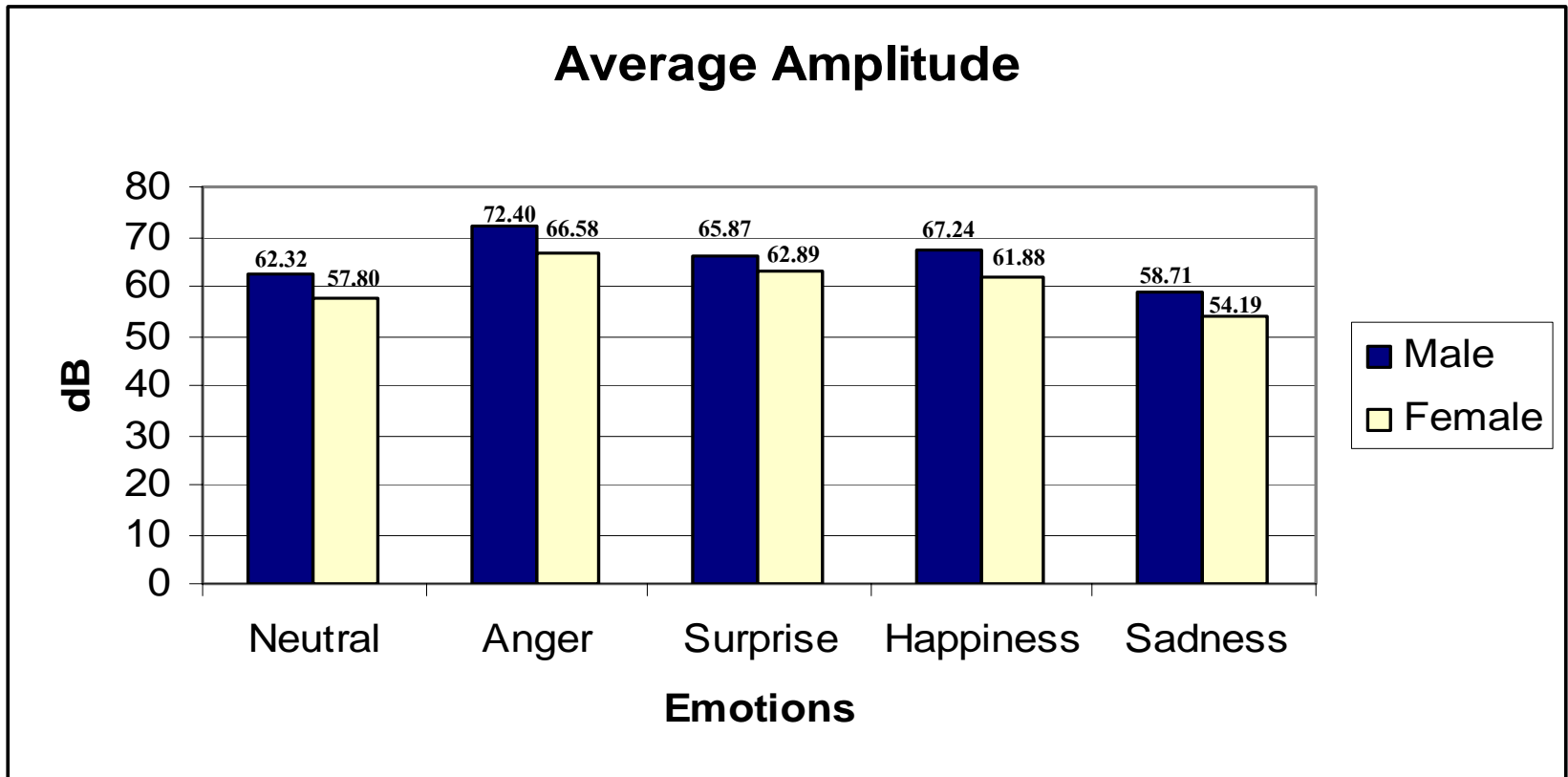
**Figure 4:** Average Amplitude (dB) for all utterances combined in each type of emotions. (male and female)

# Average Amplitude

anger > happiness/surprise > neutral > sadness

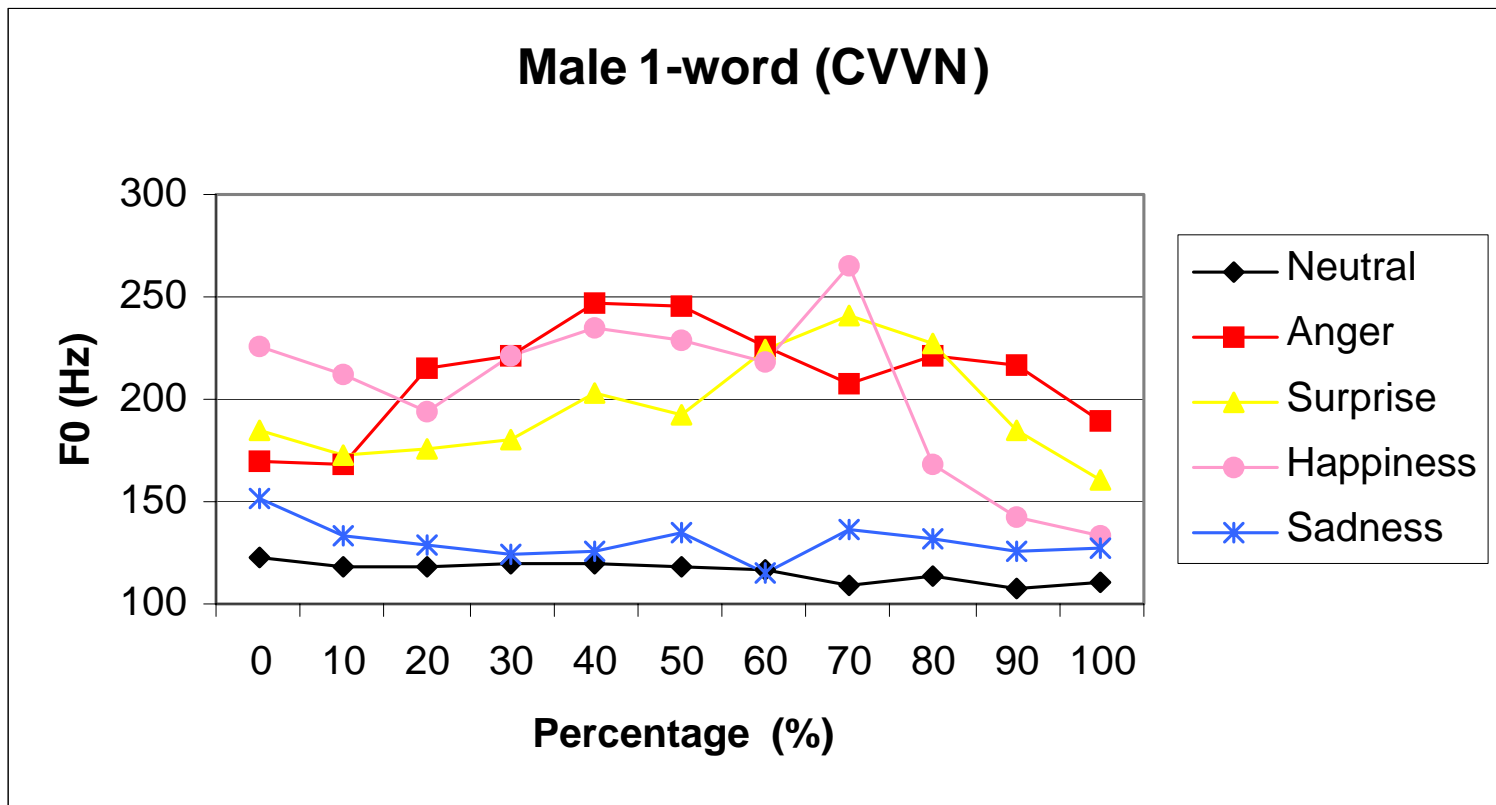highest　　　--------->　　　　　　lowest amplitude

(dB)

**Figure 5:** Stylized Fundamental Frequency (in Hertz) and neutralized duration (in Percentage) for one-word utterances. (male)
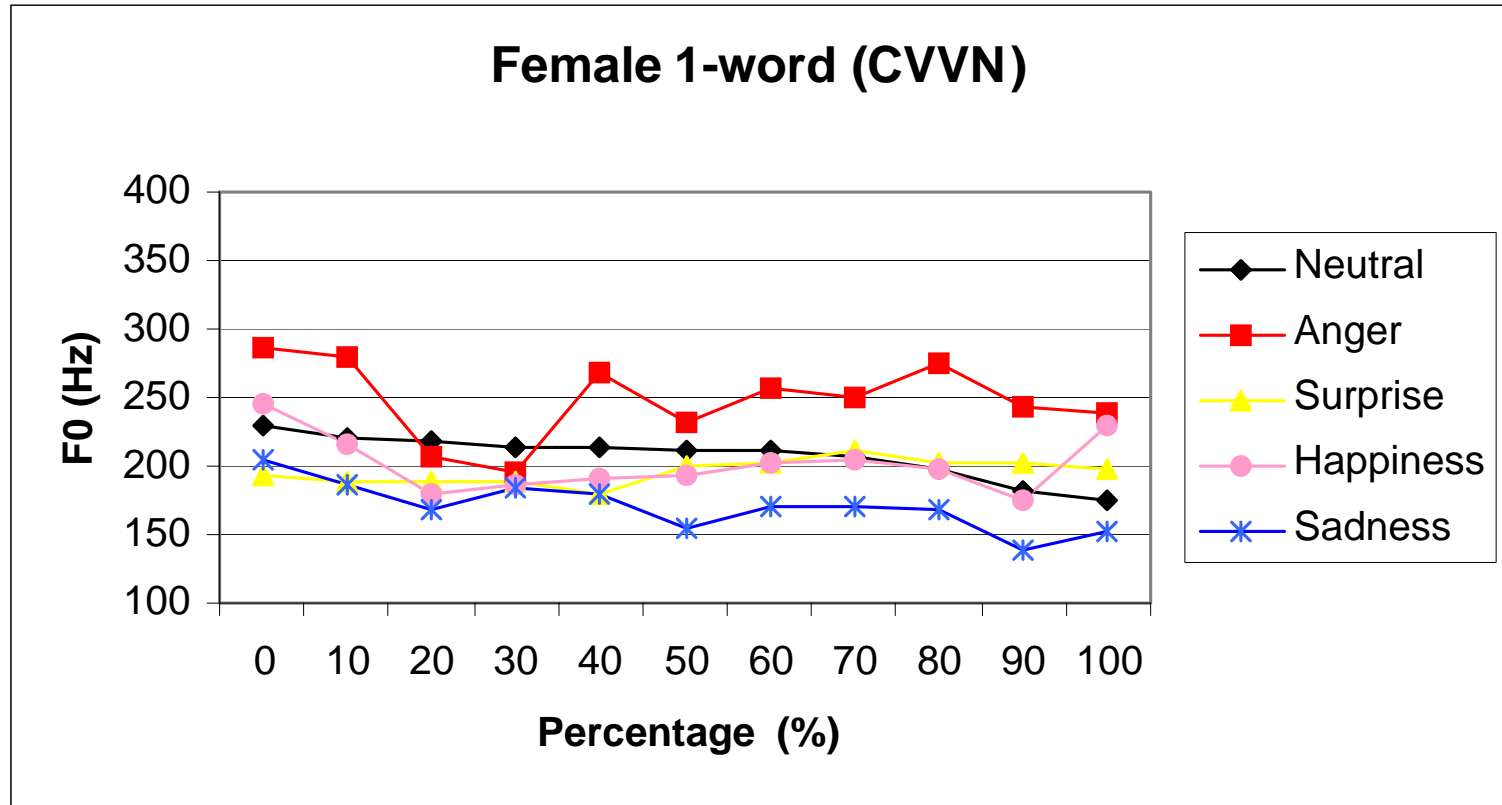
**Figure 6:** Stylized Fundamental Frequency (in Hertz) and neutralized duration (in Percentage) for one-word utterances. (female)
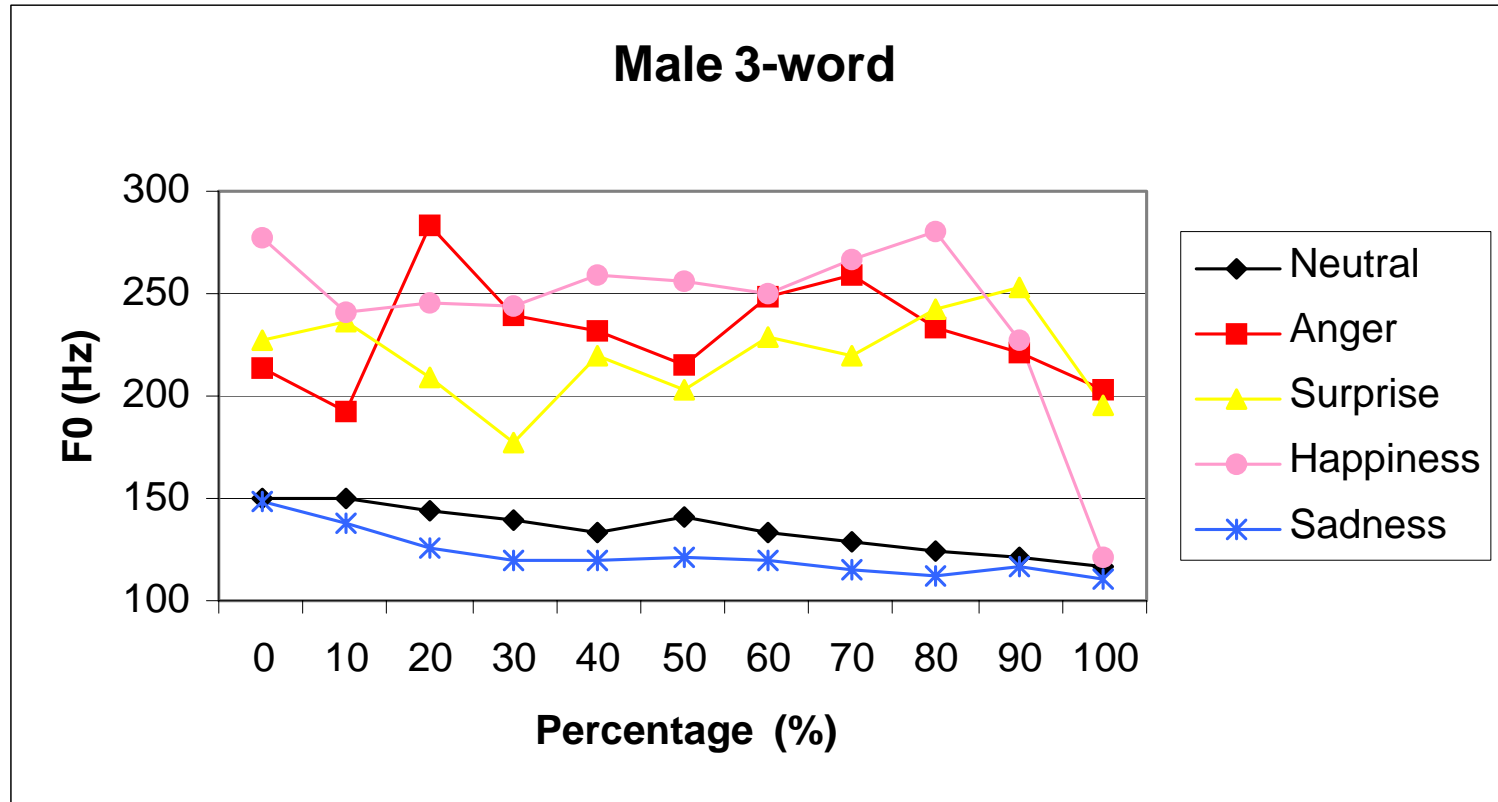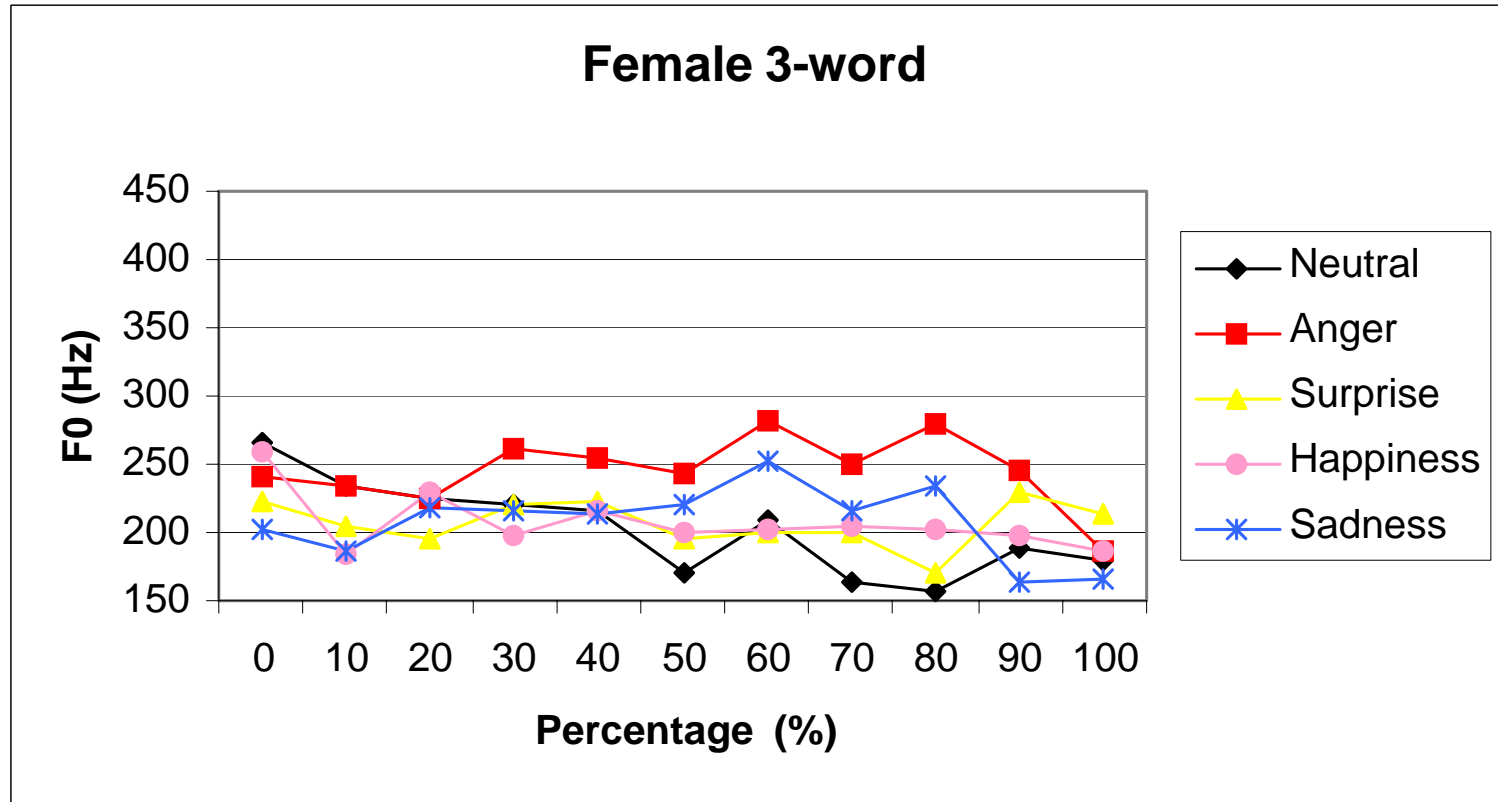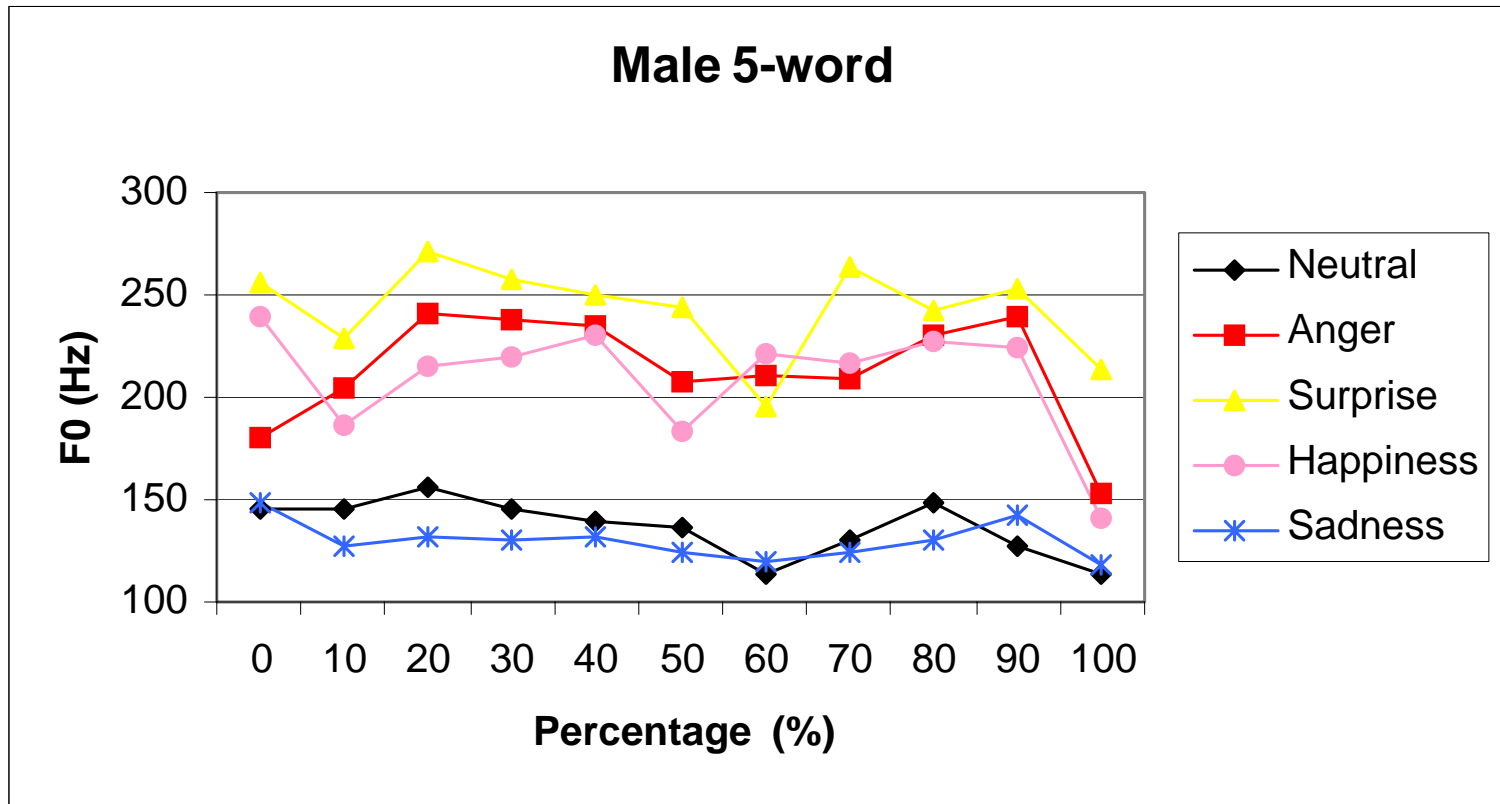
Tumtavitikul & Thitikarnnara 2006

**Figure 7:** Stylized Fundamental Frequency (in Hertz) and neutralized duration (in Percentage) for three-word utterances. (male)

Tumtavitikul & Thitikarnnara 2006

**Figure 8:** Stylized Fundamental Frequency (in Hertz) and neutralized duration (in Percentage) for three-word utterances. (female)

Tumtavitikul & Thitikarnnara 2006

**Figure 9:** Stylized Fundamental Frequency (in Hertz) and neutralized duration (in Percentage) for five-word utterances. (male)

Tumtavitikul & Thitikarnnara 2006

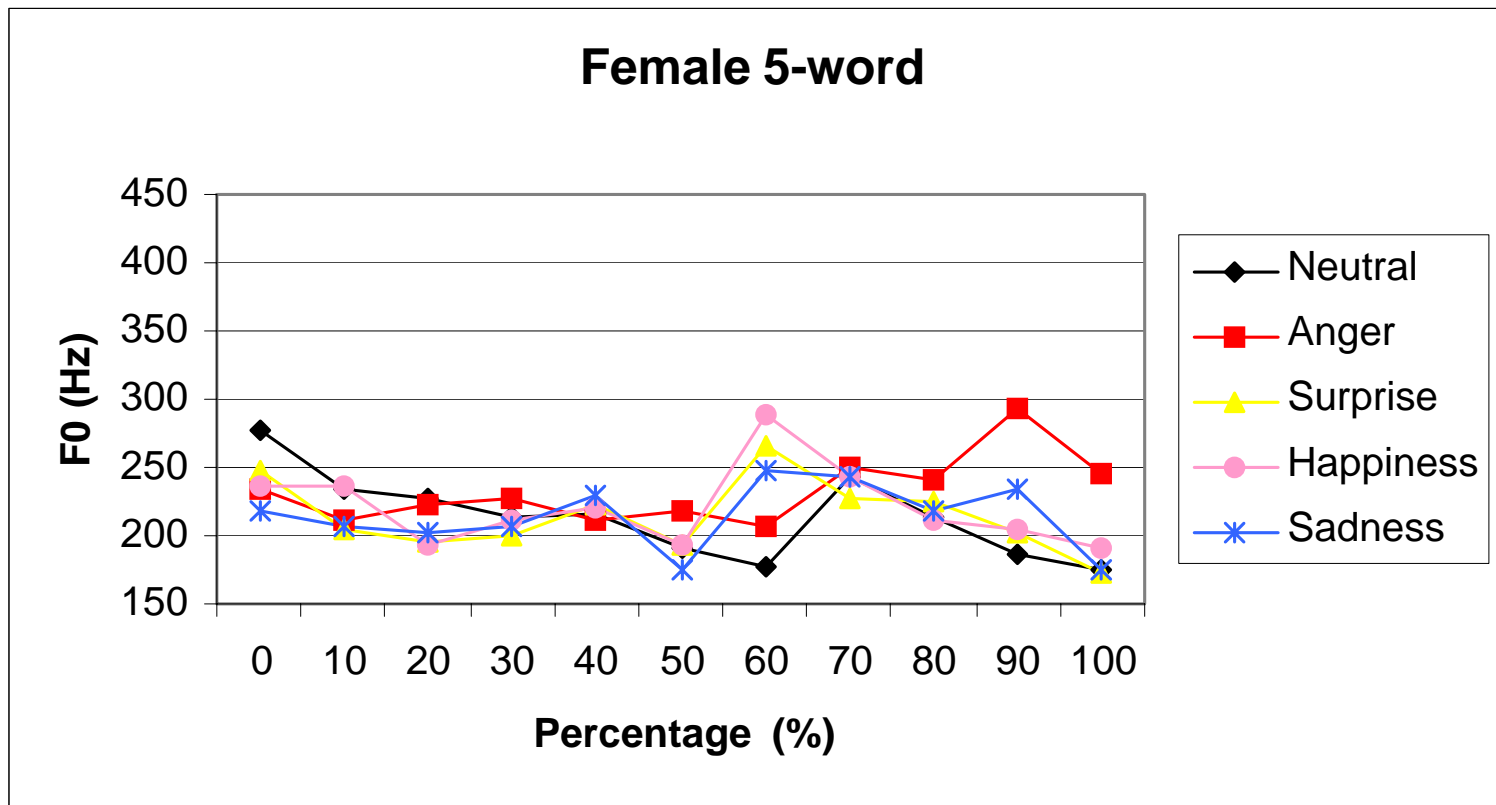**Figure 10:** Stylized Fundamental Frequency (in Hertz) and neutralized duration (in Percentage) for five-word utterances. (female)
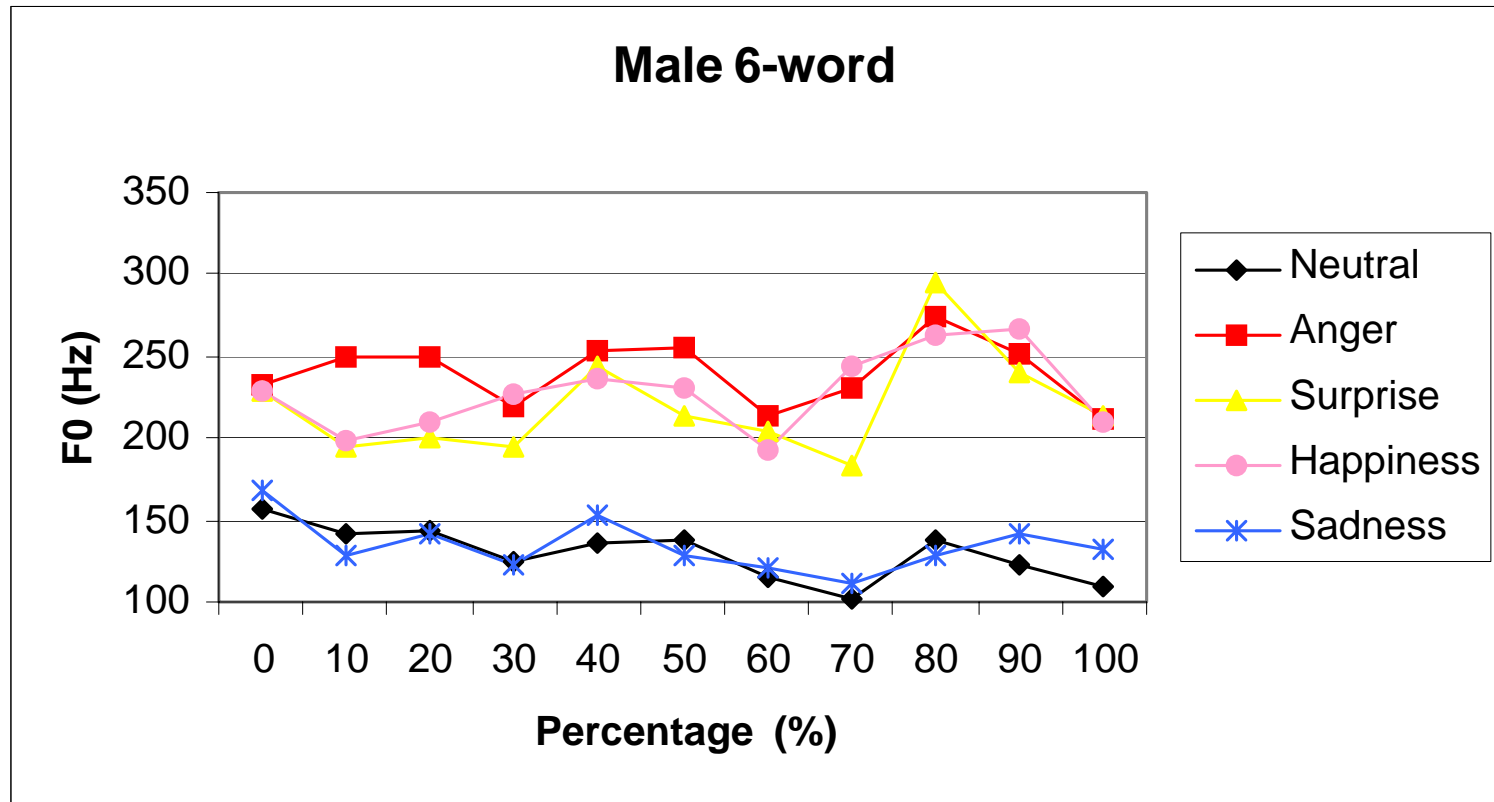
Tumtavitikul & Thitikarnnara 2006

**Figure 11:** Stylized Fundamental Frequency (in Hertz) and neutralized duration (in Percentage) for six-word utterances. (male)
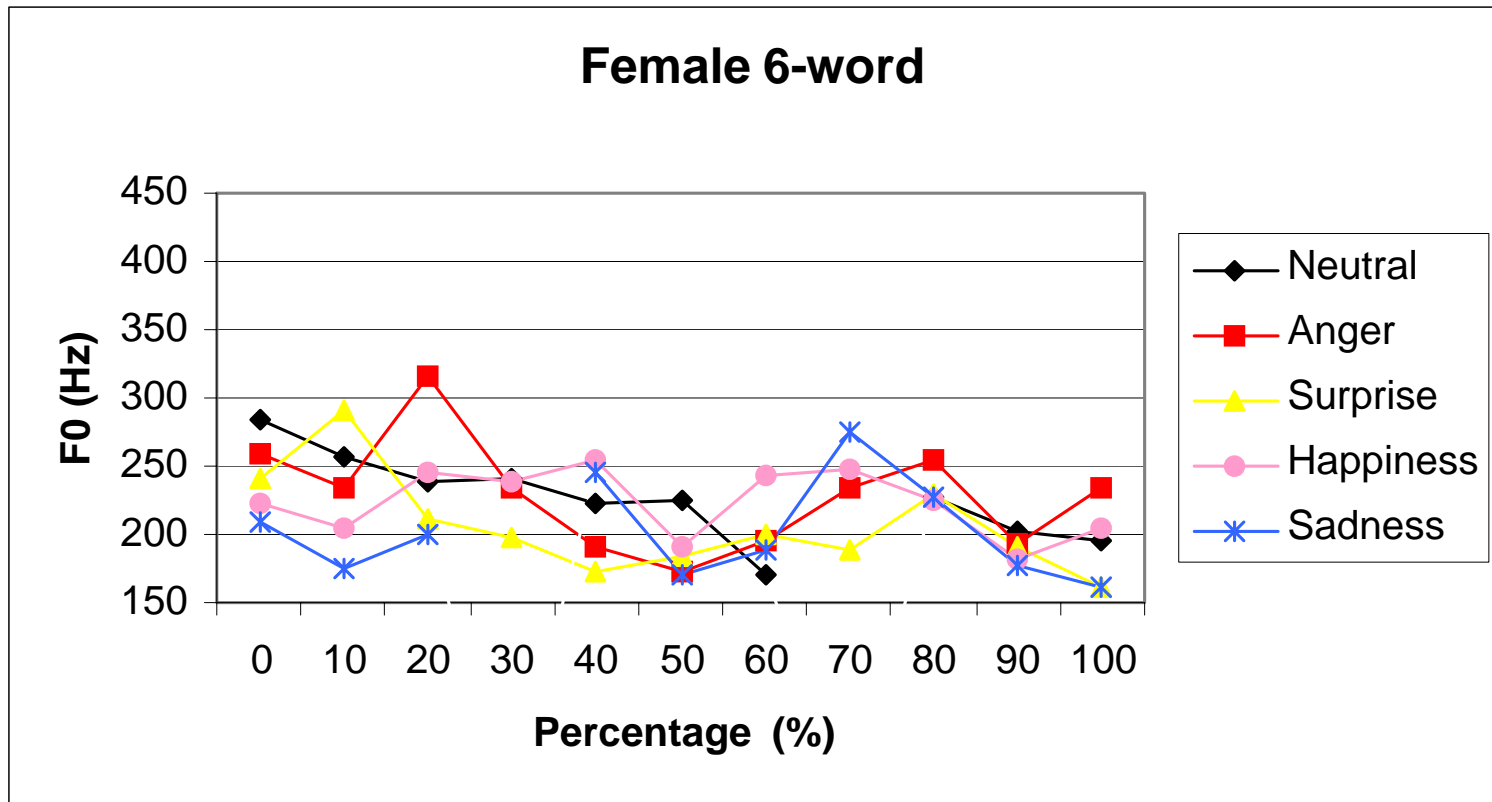
Tumtavitikul & Thitikarnnara 2006

**Figure 12:** Stylized Fundamental Frequency (in Hertz) and neutralized duration (in Percentage) for six-word utterances. (female)

Tumtavitikul & Thitikarnnara 2006

# Discussions

Luksaneeyanawin (1983) in her studies of Thai intonation found that in expressing <span style="color:red">anger,</span> the average pitch is either higer (or lower) than neutral speech, with a wider pitch range, either longer (or shorter) duration and a very high degree of loudness.  For <span style="color:yellow">surprise,</span> 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 1-12 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 is yet to be compared).

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.

For Fo 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 Fo contour of neutral declarative statements (Tumtavitikul and Thitinarnnara 2006) where the Highs and Lows conform to the universal tendency of Fo declination in declarative sentences (Hirst and Di Crito 1988) 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 emontion in both male and female speakers.

This is not surprising since emotions are subject to mood and attitude as well as environments.  Emotional pitch contours may vary greatly within and among individuals (Ladefoged 2006).

It may be induced from the comparable features found in Luksaneeyanawin, our Thai and Cahn's English data for the three types of emotional speech, anger, surprise and sadness, that emotional speech may have a universal tendency of expression.  The unifomity of expression may not be found in the intonation contour per se but in other acoustic cues e.g.,  speaking rate, F0 range, average F0, amplitude contour.

This is readily explainable since emotions which affect humans mentally and psychologically  are impulses or natural responses to stimuli.

# Conclusion

Ververidis, Kotropoulos and Pitas (2004) in their studies of automatic classification of emotional speech in Danish 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 Ververidis, Kotropoulos and Pitas (2004) 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-synthsizing the intonation contour copied from natural speech, with proper time alignments, onto monotonous utterances. The rule-based manipulations of pitch and duration were found to suffice in representing emotions.

The parameters in Vroomen, Collier and Mozziconacci (1993) for their synthesis were the type of emotion targeted, excursion size, the key in the register, and the optimal duration of the utterances.

Our findings 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.

Tumtavitikul & Thitikarnnara 2006

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.

# Acknowledgement

Tumtavitikul & Thitikarnnara 2006

# References

- Cahn, Janet. 1983. From Sad to Glad: Emotional Computer Voices. *Proceedings of Speech Tech '88, Voice Input/ Ouput Applications Conference and Exhibition.* New York City. April, 1988. Pages 35- 37.

- Dimitrios Ververidis, Constantine Kotropoulos, Ioannis Pitas. 2004. Automatic Emotional Speech Classification. *Procddings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2004.* Montreal, Canada.

- Hirst, D. and Di Cristo, A. 1998. A Survey of Intonation Systems. *Intonation Systems*, ed. by D. Hirst and A. Di Cristo. Cambridge. Camgridge University.

- Kanlayarat Thitikannara. 2006. *Thai Intonation in Different Emotions*. Unpublished MA Thesis, Kasetsart University.

- Ladfoged, Peter. 2006. *A Course in Phonetics,* 5th ed. Boston: Thomson Wadsworth.

- Luksaneeyanawin, Sudaporn. 1983. *Intonation in Thai.* Unpublished Doctoral Dissertation, University of Edinburgh.

- Tumtavitikul, Apiluck and Thitikannara, Kanlayarat. 2006. Stress and Intontaion in Thai. *Journal of Language and Linguistics* 24.2.59-76.

- Ververidis, Dimitrios and Kotropoulos, Constantine. 2003. *A State of the Art  Review on Emotional Speech Databases.* (http://poseidon.dsd.auth.gr/ EN/index.html)

- Vroomen, Jean, Collier, Rene, and Sylie Mozziconacci. 1993. Duration and Intonation  in Emotional Speech. *Proceedings of Eurospeech 1993* (pp.577–580).  Berlin, Germany.

# Thai Intonation in Four Emotions

Apiluck Tumtavitikul and Kanlayarat Thitikannara
Linguistics Dept., Kasetsart University
Bangkok, Thailand 10900
email: fhumalt@ku.ac.th
http://pirun.ku.ac.th