Role of Intonation Patterns in Conveying Emotion In Speech

H. Mesut Meral*, Hazim K Ekenel, A. Sumru Ozsoy***
***Department of Western Languages and Literatures**
****Department of Electronics Engineering**
Bogazici University, Bebek, 80815 Istanbul
{mesut.meral, ekenelha, ozsoys}@boun.edu.tr

Abstract

The aim of this paper is to present an analysis on four basic emotions – anger, happiness, sadness, unemotional state- that are expressed during the speech, by using fundamental frequency curves (henceforth F0). In this experimental study, the sentences that can be expressed with four emotions were read by non-professional actors with four different emotions. The F0 curves for each utterance are analyzed by looking at the effects of each emotion on curves.

1. Introduction:

As is stated by Mozziconacci, S.J & D.J. Hermes (1999:2001), speech not only conveys the strictly linguistic content of the sentences but also the expression of the attitudes and emotions of the speaker. In recent years, the analysis of emotions expressed in speech has become a complex subject which attracts the attentions of different disciplines and areas of study like phonetics, psychology and human-computer interaction. Not only the sequence of words and phrases, but also the features like intonation, emotion, pitch, and stress have become very important for the studies in communication, and the studies on these subjects have increased in number.

Studies on these topics mainly focus on the effects of emotions that are expressed in speech on the F0 curves. They are based on the analysis of the databases which include the recordings of ordinary people or (non)professional actors who read some sentences with certain emotions as they are instructed or the analysis of some parts that include emotional expressions extracted from the movies. Bosch (2003) argues to the possibility of emotional description by using speech signal. He focuses on the analysis of acoustic features in speech signal and the possibility of recognizing the emotional state of the speaker from these features. Bosch concludes that the acoustic features in speech signal can be used in defining the emotions that are expressed in speech.

Abelin & Allwood (2000) studies the degree of stability in the way different emotions and attitudes are interpreted by the use of prosodic patterns, whether this interpretation is dependent on the listener's cultural and linguistic background and whether there is any reoccurring relation between acoustic and semantic properties of the stimuli. A Swedish speaker's utterances with different emotions were interpreted by people with different L1's who were asked to label the emotional content of the expressions. The results of Abelin & Allwood' study showed that some emotions like anger, fear, and sadness are interpreted in accordance with the intended emotions in a greater degree than the others. They also

concluded that emotions with similar semantic content e.g. “anger” and “dominance” or “fear” and “shyness” have similar acoustic properties e.g. short duration and strong intensity.

Mozziconacci, & Hermes (1999) focused on the role of intonation patterns in conveying emotions and attitudes in the production and in the perception of speech. Their aim is to investigate whether specific intonation patterns contribute to the perception of particular emotions. Results of their experiments are analyzed in terms of pitch level-pitch range combinations, tone patterns and listeners’ judgments. They concluded that there are interactions between pitch range-pitch level combinations and listeners’ judgments.

In this study, by using speech signal, the effects of emotions that are expressed in speech on the F0 curves are analyzed in Turkish. The general aim of the analysis is to investigate whether there are some correlations between the emotional speech and the F0 curves or not, and if there is any, to analyze these correlations.

2. The method:

The present study analyses four basic emotions- anger, happiness, sadness, unemotional state- by taking into account the changes in F0 curves. The changes in F0 curves are mainly the changes in pitch level and in pitch range. The reason why the changes in F0 are analyzed is that the emotions in speech affect mostly the F0 curves as the previous studies indicate.

The data used in the study consists of the utterances produced by nonprofessional actors. The same sentences were produced with four different emotions. The reason why actors were used is that, the overall success of ordinary people in conveying emotions in speech is very low as stated in the previous studies. One alternative method to collect data for these studies is to get emotional expressions from movies. But this method has disadvantages in having different sentence structures, restricting the number of different structures that can be used in studies. The sentences used in this study are:

1. Dışarıda kar yağıyor.

   outside-LOC snow-PROG-AGR

   “It is snowing outside”

2. Sınavdan yetmiş aldım.

   exam-ABL seventy take-PAST-AGR

   “I got 70 on the exam”

3. Hoca bana yetmiş verdi.

   instructor me 70 give-PAST-AGR

   “The instructor gave me 70” (on the exam)

4. Galatasaray maçı iki sıfır kazandı.

galatasaray match-ACC two-zero win-PAST-AGR

"Galatasaray won the game 2-0"

5. Beni çok şaşırttın.

me much surprise-CAUS-AGR

"You surprised me"

6. Telefonum çalıyor.

phone-POSS ring-PROG-AGR

"My phone is ringing"

7. Kurs yarın bitiyor.

course tomorrow end-PROG-AGR

"The course ends tomorrow"

8. Yarın kar yağacakmış.

Tomorrow snow-FUT-MOD-AGR

"They say that it will be snowing tomorrow"

9. Dersler iki hafta ertelendi.

course-PL two week-PL postpone-PASS-PAST-AGR

"The courses were postponed two weeks"

10. Kapı açık kalmış.

door open left-MOD-AGR

"The door has been left open"

11. Dersi sadece iki kişi geçemedi.

course-ACC only two person pass-NEG ABIL-PAST-AGR

"Only two people could not pass the course"

The sentences are affirmative-declarative sentences which have syntactically different structures. Negative sentences and questions are excluded from the data. Sentences contain different parts of speech words like noun, verb, adjective, adverb and pronoun. Ten of the verbs are in active voice and one is in passive voice. Four sentences have transitive verbs, six intransitive and one ditransitive verb. Sentences are simple and have different word orders that Turkish allows like adverb-subject-verb, adverb-direct object-verb, subject-indirect object-direct object-verb, subject-direct object-adverb-verb, subject-verb, subject-adverb-verb, indirect object-adverb-verb, direct object-subject-verb.

3. Experiments:

In the study, eleven sentences were read by seven women and four men with four different emotions. All the participants were university students and were recorded in different settings in which one of the researchers present. The utterances of one emotion were recorded at different times in order not to be affected by others. The sentences were shown to the subjects before the recording, the subjects were asked to read them independently from each other. At the end of the recordings, we had a database which consists of 484 voice files. Background noises like breathing were deleted prior to analysis. The F0 curves of the recordings were identified and the words labeled with Hidden Markov Model. The control group listened to the recordings and labeled them as -anger, happiness, sadness, unemotional-. The aim of having the control group was to check the overall success of the interpretation.

3. 1 F0 analysis:

The effects of the emotions on F0 curves are analyzed on the recordings that are the most successful in perception by the second group (90% and above). The main focus of the analyses is the F0 patterns in sentence initial and sentence final positions. The effects of the emotions on the F0 curves can be summarized as follows:

The F0 curves of the sentences read with anger, happiness and unemotional state show, in most cases, rise and fall in sentence initial position. The sentences read with sadness, on the other hand, show either fall and rise or hold pattern in this position. The basic tendency of F0 curves of the utterances in sentence final position is fall in anger and unemotional state cases and rise in the happiness and sadness cases. According to the statistical values of the F0 curves in the utterances of these four different emotions, happiness has the highest average F0 value, while unemotional state has the lowest. The emotion that has the highest range and standard deviation is anger; unemotional state and sadness have low range and standard deviation compared to other emotions.

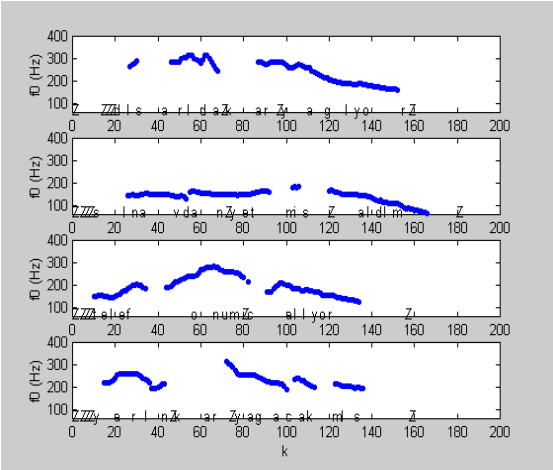


Figure 1: F0 curves of the anger

The general tendency of the F0 curves in the utterances with anger is to fall towards the end of the utterance. We recognized 81% fall at the end of the sentences. The percentage of rise at the end is 10% and of hold 9%. The tendency at the beginning of the sentences is not clear. However, we recognized more rise and fall than fall and rise and hold. Anger utterances have 48% rise and fall, 29% fall and rise and 23% hold at the beginning of the sentence. On these observations, we can say that the F0 curves of anger utterances tend to fall at the end of the sentence and rise and fall at the beginning of the sentence.

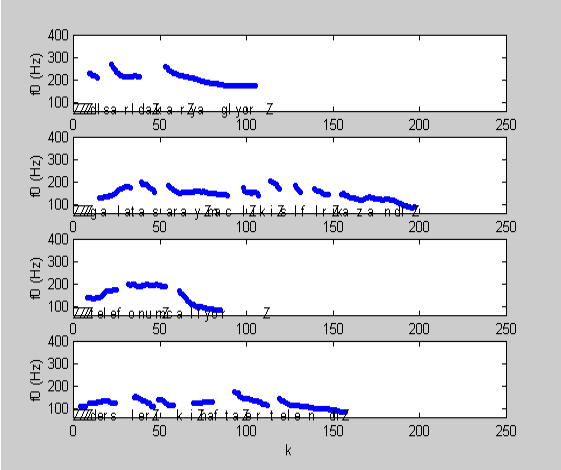


Figure 2: F0 curves of unemotional utterances

The main tendency in F0 curves of the unemotional sentences is to fall at the end of the sentence as in anger utterances. However, rise and hold tendencies are also possible at the end of the sentences. We recognized more rise and fall case than fall and rise at the beginning of the sentences. The unemotional utterances have 49% rise and fall, 16% fall and rise and 35% hold cases at the beginning of the sentences. According to these observations, the F0 curves of unemotional utterances tend to fall at the end of the sentence and have rise and fall at the beginning as in anger utterances. The main difference between them is that in anger utterances, the pitch range value is greater than unemotional state.

F0 curves of the sentences that are expressed with happiness differ from those of anger and unemotional sentences. They have rise at the end of the sentences. The percentage of rise case is 59%. F0 curves that tend to fall at the same position have the percentage 28% and 13% of total is hold. The tendency of the F0 curves at the beginning of the sentences is more accurate than the one in anger and unemotional cases. Utterances with happiness rise and fall in the beginning of the sentences with the percentage 71%. Those with rise and fall is 22%, and those with hold is 7%.

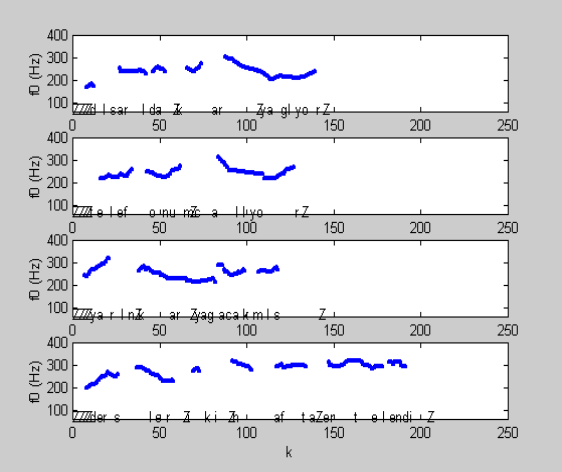


Figure 3: F0 curves of the happiness utterances

As can be seen in figure 3 above, F0 curves of the utterances with happiness tend to rise at the end of the sentences, unlike those of anger and unemotional utterances. However, just like the curves on anger and unemotional utterances, they tend to rise and fall at the beginning of the sentences.

Sentences that are produced with sadness rise at the end of the sentences as in happiness. However, this tendency is not as clear as it is in the case of happiness. 61% percentage of the sentences have rise, 22% percentage fall and 17% have hold at the end of the sentence. The tendency at the beginning of the sentences is not clear either. The F0 curves have more falls and rises than hold or rise and fall. The percentage of fall and rise is 48% and the percentage of hold is 39%. The curves rise and fall with the percentage 13%.

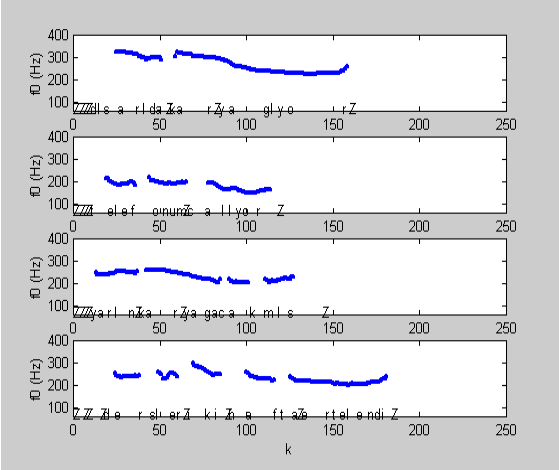


Figure 4: F0 curves of sadness expressions

By looking at the figure 4 above and the previous observations, it is possible to claim that F0 curves of the utterances of sadness tend to have rise at the end of the sentences. At the beginning, the utterances can either have fall and rise or hold.

	Unemotional	**Anger**	**Sadness**	**Happiness**
Maximum	281	332	321	343
Minimum	156	180	196	208
Range	125	152	125	135
Average	214	251	259	274
Median	214	250	259	273
Standard deviation	33,7	39,3	34,5	35,6

Table 1: Statistical values for F0 curves

Table 1 shows various statistical values calculated from the F0 curves of four different emotions. According to these values, happiness has the highest average F0 value while the unemotional state has the lowest one. The pitch range

and standard deviation are highest in anger and lowest in unemotional state and sadness.

3.2 Perceptual experiments

In the second stage of the study, a control group of 10 between the ages 18-30 listened to the recordings and classified them as expressing anger, happiness, sadness and unemotional. The voice files were fixed randomly and 3 different sets of files were formed from these sentences. The listeners listened to the sets of files at different settings between half day and one week intervals.

emotion	Unemotional	Anger	Sadness	Happiness
Unemotional	85,71	6,40	6,77	1,12
Anger	3,70	92,37	1,12	2,81
Sadness	2,92	3,03	90,80	3,25
Happiness	5,69	5,80	6,43	82,08

Table 2: confusion matrix (%)

Table 2 indicates the confusion matrix resulting from the perceptual experiments. In this matrix, expressed emotions are shown horizontally, those perceived are shown vertically. As can be seen in confusion matrix, the most correctly perceived emotion is anger. This can be explained by the fact that the sentences that are expressed with anger have more precise tendencies. The least correctly perceived emotion is happiness. Most of the errors are involved with unemotional state and sadness. The main reason in this confusion is that both of these emotions have lowest pitch range and standard deviation values.

Experiment Sets	Error Percentages
Set 1	14,90
Set 2	12,21
Set 3	9,68
Average	12,26

Table 3: Error percentages (%)

To determine whether there is a learning process in perception of emotions, the perception errors in each set were analyzed. Table 3 indicates that as the set

number increases, the percentage of error decreases. This fact indicates that there is a learning process in perception of emotions.

4. Results:

In this study, the effects of four basic emotions on the F0 curves are analyzed. As a result of the analysis the following are observed:

- The F0 curves of the anger and unemotional utterances tend to have rise and fall at the beginning of the sentence and fall at the end of the sentence.
- The difference between unemotional state and anger is that the pitch range is higher in the former and lower in the latter.
- The F0 curves of the happiness utterances have generally rise and fall at the beginning of the sentences and rise at the end of the sentences.
- The F0 curves of the utterances that are expressed with sadness have either fall and rise or hold at the beginning of the sentences and rise at the end of the sentences.
- Happiness has the highest average F0 value and unemotional state has the lowest F0 value.
- The pitch range and standard deviation are highest in anger.
- Unemotional state and sadness have the lowest pitch range and standard deviation compared to the others.
- The most correctly recognized emotion is anger and the least correctly recognized emotion is happiness. There is a learning process in interpretation of emotions.

5. Thanks:

We would like to express our thanks to "Bogazici University Actors" -Canan Tanır, Deniz Aydın, Didem Karanfil, Erdoğan Çete, Eser Dilsöz, Ozan Uysal, Sezin Gündoğan- and to Emel Kahraman, Işıl Eğrikavuk, İlker Öztop, Umay Karabaş, A. Hakan Büyükbayrak, Beste Can Yolcu-Kamali, Duygu Şimşek, E. Edip Çiçekli, E. Pınar Cen, Erdem Ulukan, Gülbin Akgün, Görkem Şarkış, Murat Erman, Onur Sarkan, Özge Çağlar, and to the participants of Linguistics Baglunch of Bogazici University for their contributions and comments.

6. References:

Abelin, A & J. Allwood (2000) "Cross Linguistic Interpretation of Emotional Prosody", *Proceedings of the ISCA Workshop on Speech and Emotion.* http://www.qub.ac.uk/en/isca/proceedings

Bosch, L.T (2003) Emotions, Speech and the ASR Framework, *Speech Communication,* c. 40, s. 213-215.

Burkhardt, F, & W.F Sendlmeier (2000) Verification of Acoustical Correlates of

Emotional Speech Using Formant-Synthesis, *Proceedings of the ISCA Workshop on Speech and Emotin.* http://www.qub.ac.uk/en/isca/proceedings

Demircan, Ö (1980) Türkçe Ezgilemeye Giriş, *TDAY-Belleten* p.65-79.

Kornfilt, J (1997) *Turkish*, London: Routledge.

Lee, C.M, S. Narayanan, R. Pierraccini (2001) Recognition of Negative Emotions from the Speech Signal, *Automatic Speech Recognition and Understanding Workshop*. http://asru01.itc.it/paper/abs/a01cl099.htm

McGilloway, S, R. Cowie, E. Douglas-Cowie, S. Gielen, M. Westerdijk, S. Stroeve (2000) Approaching Automatic Recognition of Emotion from Voice: A Rough Benchmark, *Proceedings of the ISCA Workshop on Speech and Emotion.* http://www.qub.ac.uk/en/isca/proceedings

Montero, J. M, J. Gutierrez-Arriola, J. Colas, E.Enriquez & J.M Pardo (1999) Analysis and Modelling of Emotional Speech in Spanish, *ICPhS'99.* http://lorien.die.upm.es/~macias/doc/pubs/eurosp99/submitted/m058.pdf

Mozziconacci, S.J & D.J. Hermes (1999) Role of Intonation Patterns in Conveying Emotion in Speech, *ICPhS'99.* http://www.lpl.univ-aix.fr/projects/sprosig/guests/mozzic/icphs99.pdf

Oudeyer, P.Y (2002) The Production and Recognition of Emotions in Speech: Features and Algorithms, *International Journal of Human Computer Interaction.* http://www.csl.sony.fr/downloads/papers/2002/oudeyer02b.pdf

Petrushin, V.A (1999) Emotion in Speech: Recognition and Application to Call Centers, *Artificial Neural Networks in Engineering (ANNIE '99),* p. 7-10.