FUNDAMENTAL FREQUENCY
IN ESTONIAN EMOTIONAL READ-OUT SPEECH

Kairi Tamuri
Institute of the Estonian Language

Abstract. Fundamental frequency (F0, perceived as pitch) is an important prosodic cue of emotion. The aim of the present study was to find out if sentence emotion has any influence detectable in the F0 height and range of Estonian read-out speech. Thus the F0 of each vowel found in Estonian read-out sentences was measured, and its median for three emotions (anger, joy, sadness) and for neutral speech was calculated. In addition, the F0 range was measured for emotional and neutral speech, as well as the height of F0 for sentence-initial and sentence-final positions. The results revealed that in the investigated material, F0 was highest for joy and lowest for anger. The F0 range, however, was widest for anger and narrowest for sadness. The differences in F0 height at the beginning versus the end of sentences were not statistically significant, either for pairs of emotions or for emotions compared with neutral speech.

1. Introduction

Speech is never monotonous (Ariste 1953: 99), its parameters are constantly changing by forming acoustic patterns. Both neutral and emotional speech have its characteristic pattern.

Estonian emotion acoustics have mainly been addressed in the context of speech synthesis. Previous analysis has been focused on speech rate and pauses (Tamuri and Mihkla 2012, Tamuri 2010), formants and precision of articulation (Tamuri 2012a) and intensity of speech (Tamuri 2012b), the emotions involved being joy, sadness and anger. The overall objective of the studies is the creation of an acoustic model of Estonian emotions, based on real acoustic cues, in order to improve the naturalness, expressiveness and subsequently the intelligibility of Estonian synthetic speech.

Recent developments in speech technology have brought a lot of attention to the acoustics of emotional speech, particularly in the context of speech synthesis and emotion identification. Research has established the acoustic cues of basic emotions (incl. joy, sadness and anger), which enables recognisable synthesis of those emotions as well as their identification in oral speech (Koolagudi and Rao 2012, Schuller et al. 2011, El Ayadi et al. 2011, Nose and Kobayashi 2011, Schröder 2009).

A speech sound has numerous acoustic variables enabling the expression of emotions. One of these variables is fundamental frequency, whose height and range are significant parameters for distinguishing between vocal emotions (Arias et al. 2014, Busso et al. 2009, Pell et al. 2009). However, the results of other languages cannot be applied to Estonian automatically, because acoustic expression of emotions differs across cultures, and correct identification of emotions from vocal cues is only possible within a culture (Altrov 2013, Soto and Levenson 2009).

Fundamental frequency (F0) derives from the rate of vibration of the speaker's vocal cords. The listener perceives F0 as pitch. The F0 of an average male voice is 80–150 Hz and that of an average female voice is 150–300 Hz.

In the 1980s, Klaus R. Scherer developed a theoretical model to predict the influence of emotions on vocal expression. His *Component Process Model* (CPM) takes account of various psychological and physiological aspects of emotional expression, stating the existence of certain emotion-specific acoustic patterns (see Scherer 1986). Scherer has also tested the validity of the CPM (see Scherer and Meuleman 2013, Scherer 2009). According to the results, vocal expression of emotions is indeed subject to certain regularities that are predictable by a theoretical model.

This study will introduce the predictions of the CPM for the effects of certain emotions on fundamental frequency, and give a survey of the available results on the F0 of emotional speech in English, German, Swedish, Hindi and Arabic.

1.1. Effects of anger on fundamental frequency

As concerns the fundamental frequency of anger speech, emotional speech studies do not always agree. As the reason may lie in the speech material, a difference is now made between *hot anger* or *rage* and *cold anger* or *irritation*.

For cold anger, the CPM predicts that, depending on the type of voice (lax or tense), F0 may be either higher or lower than for neutral speech, but the F0 range will decrease. For hot anger, too, F0 may be either lower or higher than neutral, but the F0 range will increase (Scherer 1986).

According to earlier studies, the anger sentences have a higher F0 than either joy, sadness or neutral sentences (in English, German and Hindi), while in English the difference is significant; in Arabic, however, the F0 of anger sentences is significantly lower than that of joy sentences, but significantly higher than for sadness or neutral speech (Pell et al. 2009, Murray and Arnott 2008). A study of authentic Swedish emotional speech has shown that in Swedish anger sentences, the F0 was higher than for sadness or neutral speech (Laukka et al. 2011).

Regarding the F0 range of anger sentences, previous studies have shown that in English and Hindi, the F0 range is wider for anger sentences than for joy, sadness or neutral speech, whereas in German and Arabic it is narrower for anger than for joy. In Hindi and German, the F0 range of anger sentences differs significantly from that of sadness sentences or neutral speech, while in English it also differs significantly from that of joy sentences (Pell et al. 2009, Murray and Arnott 2008).

1.2. Effects of joy on fundamental frequency

Like anger, joy is also divided into two states: *enjoyment* and *elation*. The first is defined as quiet happiness, while the other is defined as elated joy. The differences of the two states of emotion are reflected in opposite tendencies in F0, intensity and speech rate (Scherer 1986). The CPM predicts that enjoyment generates a low F0 and a narrow F0 range, whereas elation brings about a high F0 and a wide F0 range.

As not all studies distinguish between the two states of joy, the corresponding results may sometimes be inconsistent. However, joy is generally associated with a high value and wide range of fundamental frequency, as seen in, for example, the results on English emotional speech as obtained by Murray and Arnott (2008). Pell and his colleagues (2009) found that although in English, German and Hindi, joy sentences have a lower F0 than anger sentences, their F0 is significantly higher than in sadness sentences or in neutral speech. In Arabic, however, F0 is highest in joy sentences (Pell et al. 2009).

Regarding the F0 range, it is widest for joy sentences in German and Arabic, while in English and Hindi it is narrower for joy than for anger, but wider for sadness and neutral speech (Pell et al. 2009).

1.3. Effects of sadness on fundamental frequency

The CPM predicts that in comparison with neutral speech, the sadness sentences will have a lower F0, but depending on the voice type (e.g. tense or lax voice) it may also be higher; regardless, the F0 range will certainly narrow down for sadness (Scherer 1986).

Previous studies have also shown that sadness goes with a low fundamental frequency. In Swedish, the F0 for sadness is lower than for anger, but higher than for neutral speech (Laukka et al. 2011). In German and Arabic, sadness has the lowest F0 compared with joy, anger and neutral speech, whereas in English and Hindi the F0 is higher for sadness than for neutral speech (Pell et al. 2009).

Regarding the F0 range of sadness sentences, previous results have been generally consistent, scoring the narrowest in comparison with joy, anger and neutral speech, at least in German, Arabic and Hindi. In English, the F0 range is narrower for sadness than for joy or anger, but wider than for neutral speech (Pell et al. 2009). The small F0 range of English sadness sentences has also been pointed out by Murray and Arnott (2008).

A comparison between our results, the CPM predictions and earlier studies will be presented in the Discussion section.

2. Research question and material

The aim of the present study was to find out if sentence emotion has any effect on the height and range of the fundamental frequency of Estonian read-out speech.

The research material comes from the *Estonian Emotional Speech Corpus*[1] of the Institute of the Estonian Language. The corpus contains journalistic passages as read by a female voice, which have been segmented into sentences. The corpus has been designed on the principle that the reader is not dictated or prompted with the correct emotion, as it should be inferable from the contents of the passage. All corpus

---

1   See http://peeter.eki.ee:5000/

sentences are different, while their emotion (anger, joy, sadness) or neutrality has been determined by perception tests. An emotion was considered to be identified if agreed upon by no fewer than 51% of the listeners. The corpus does not contain full-blown, prototypical emotions, but rather emotion-related states. Thus *anger* also covers displeasure, irony, distaste, disdain, malignant delight, rage; *joy* = gratitude, happiness, pleasure, enthusiasm; *sadness* = loneliness, disconsolation, concern, hopelessness; *neutral* = ordinary speech without special emotions. (Altrov and Pajupuu 2013)

To answer the research question, I analysed 329 sentences whose emotion or neutrality had been perceived from voice by no less than 51% of the listeners (more than two times better than chance probability) (see Table 1).

Table 1. Material of F0 analysis

Emotion	No. of sentences
Anger	79
Joy	60
Sadness	87
Neutral	103
Total	329

3. Method

In order to see if sentence emotion can affect the F0 height and/or range, the Praat program (Boersma and Weenink 2014) was used to measure the fundamental frequency in sentences representing the emotion of anger, joy or sadness, as well as in neutral sentences. This done, the F0 median and range (difference between the maximum and minimum F0 values), as well as the F0 height in sentence-initial and sentence-final positions, were computed for each emotion and for neutral speech. A Wilcoxon rank-sum test was run to see if the differences between the pairs of emotions as well as between each emotion and neutral speech were statistically significant.

4. Results

4.1. Fundamental frequency in emotional and in neutral speech

The F0 measurements for three emotions and neutral speech are presented in Table 2.

Table 2. Fundamental frequency in emotional and in neutral speech (Hz)

	Anger	Joy	Sadness	Neutral
Min	134.5	133.3	139.4	120.7
Q1	165.7	171.3	172.6	171.9
Median	180.8	187.7	185.4	185.4
Q3	208.0	214.8	206.6	208.5
Max	270.4	279.5	257.1	263.3

The results reveal that in the material investigated, the F0 median is highest for joy sentences (187.7 Hz) and lowest for anger sentences (180.8 Hz), while neutral speech and sadness have a similar F0 of 185.4 Hz.

To find out if the F0 differences between the emotions and neutral speech were statistically significant, a Wilcoxon rank-sum test was run, the results of which are presented in Table 3.

Table 3. Wilcoxon rank-sum test results for F0 by emotion pairs and in comparison with neutral speech

	Anger	Joy	Neutral
Joy	.001***		
Neutral	.001***	.350	
Sadness	.001***	.240	.610

Note. P value adjustment method: Holm, $*p < .05$, $**p < .01$, $***p < .001$

The test revealed that in Estonian, the differences in F0 height are statistically significant for anger, whose F0 is significantly lower than in the case of the other emotions or neutral speech. No other difference between the emotions or in comparison with neutral speech was statistically significant.

4.2. F0 range in emotional and in neutral speech

In order to find the F0 range for emotional and neutral speech, the difference between the maximum and minimum F0 values was calculated for each sentence. For the results, see Table 4.

Table 4. F0 range (Hz) in emotional and in neutral speech

	Anger	Joy	Sadness	Neutral
Min	51.4	42.9	39.3	45.2
Q1	91.1	84.0	75.4	81.7
Median	105.1	105.0	89.1	94.8
Q3	124.6	124.2	102.7	107.0
Max	166.4	181.0	142.3	135.3

According to the results, the median of the F0 range is the widest in anger and joy sentences (105.1 Hz and 105.0 Hz, respectively) and the narrowest in sentences of sadness (89.1 Hz). The Wilcoxon rank-sum test demonstrated that the F0 range shows a significant difference between anger and sadness as well as neutral speech, and also between joy and sadness (see Table 5).

Table 5. Wilcoxon rank-sum test results on the F0 range in pairs of emotions and in comparison with neutral speech

	Anger	Joy	Neutral
Joy	.519		
Neutral	.013*	.179	
Sadness	.001***	.009**	.109

Note. P value adjustment method: Holm, *p < .05, **p < .01, ***p < .001

4.3. Sentence-initial and sentence-final F0 in emotional and in neutral speech

The fundamental frequency of emotional versus neutral speech was also investigated at the beginning and end of sentences. For that purpose, the F0 was measured (in Hz) for the first vowel of the first and last words of each sentence. The results are presented in Table 6.

Table 6. F0 in sentence-initial and sentence-final position in emotional and in neutral speech

	Anger *Beginning / end of sentence*	Joy *Beginning /end of sentence*	Sadness *Beginning / end of sentence*	Neutral *Beginning / end of sentence*
Min	142.3 / 145.3	163.2 / 141.1	142.0 / 143.9	163.1 / 139.4
Q1	198.6 / 156.6	201.2 / 158.4	197.3 / 157.2	203.7 / 157.1
Median	214.3 / 161.4	225.8 / 164.4	213.1 / 164.4	231.6 / 164.9
Q2	245.0 / 165.9	252.5 / 171.7	234.8 / 173.1	249.3 / 170.7
Max	313.1 / 177.1	314.3 / 187.7	283.9 / 192.2	284.7 / 186.8

According to the results, the sentence-initial F0 is highest in neutral Estonian speech and lowest in sentences of sadness. The sentence-final F0 is also highest in neutral speech, but lowest in anger sentences.

The Wilcoxon rank-sum test showed, however, that the F0 differences in sentence-initial and sentence-final positions were not statistically significant, either for pairs of emotions or between emotional and neutral sentences ($p > .05$).

5. Discussion

The reported analysis demonstrated that within the material investigated, sentence emotion did affect the height as well as the range of the fundamental frequency.

Subsequently, results for the Estonian language are compared with the predictions of Scherer's CPM and with previous results for English, Swedish, German, Hindi and Arabic.

For anger, the CPM predicted that depending on the type of voice (e.g. lax or tense), the F0 of anger sentences can be either higher or

lower than in neutral speech, while the same is true for both hot and cold anger. Previous studies have shown that in comparison with sentences of joy and sadness as well as with neutral speech, anger sentences have the highest F0 in English, German and Hindi, while in English the difference is statistically significant. However, in some languages, such as Arabic for example, the F0 of anger sentences is significantly lower than that of joy sentences, but significantly higher than for sadness or neutral speech. In Swedish, the F0 is higher for anger than for sadness or neutral speech. (Laukka et al. 2011, Pell et al. 2009, Murray and Arnott 2008, Scherer 1986)

The results of our F0 study differ from those described above; notably, in Estonian emotional read-out speech, the F0 for anger was the lowest compared to the F0 for joy, sadness or neutral speech, while the difference was statistically significant.

The CPM prediction for the F0 range was that it would get narrower for cold anger and wider for hot anger. According to previous results, in English and Hindi the F0 range for anger is the widest in comparison with joy, sadness and neutral sentences, whereas in German and Arabic it is narrower for anger than for joy, remaining still wider than in sadness and neutral sentences. In German and Hindi, the difference between the F0 range for anger and that for sadness and neutral sentences is statistically significant, while in English it also differs significantly from the F0 range for joy. (Pell et al. 2009, Murray and Arnott 2008, Scherer 1986)

The results of the present study partially coincide with previous results in that the Estonian sentences of anger have the widest F0 range, differing significantly from that for sadness or for neutral speech.

It is hard to say whether the anger of the research material was hot or cold. Although the wide F0 range should be typical of hot anger, the low height of F0 could, according to CPM, associate with hot as well as with cold anger.

As for joy, it is also divided into two emotions by the CPM, which are enjoyment (quiet happiness) and elation. The CPM predictions for the F0 height are low for quiet happiness and high for elation. According to previous studies of English, German and Hindi, joy sentences have a lower F0 than anger sentences; this is, however, significantly higher than the F0 observed for sadness or for neutral speech. In Arabic, the joy sentences have the highest F0 compared to either anger, sadness or neutral sentences. (Pell et al. 2009, Murray and Arnott 2008, Scherer 1986)

Our results also showed that for joy sentences, the F0 is high. In read-out Estonian speech, the F0 of joy sentences was indeed the highest as compared with other emotions and neutral speech, but the difference in F0 height was significant only in comparison with anger, which has the lowest F0 in Estonian.

The CPM predictions for the F0 range are that it will get narrower for quiet happiness and wider for elated joy. Earlier studies have observed that in German and Arabic, the F0 range is the widest in joy sentences as compared with anger, sadness and neutral sentences, whereas in English and Hindi it is narrower than in anger sentences, but wider than for either sadness or neutral speech. (Pell et al. 2009, Murray and Arnott 2008, Scherer 1986)

The results of the present study revealed that the F0 range is relatively wide for Estonian joy, meaning that it is wider than for sadness and neutral speech, but narrower than for anger. The difference in F0 range is statistically significant only between joy and sadness.

According to the CPM description, the joy investigated may have been elated, as it had both a high F0 and a wide F0 range.

According to the CPM prediction for sadness, the F0 in sadness sentences depends on the type of voice (e.g. lax or tense), thus being respectively either low or high. Previous results indicate that in German and Arabic, the F0 is lowest for sadness, being lower than for joy, anger, or neutral speech, whereas in English and Hindi, the F0 for sadness is higher than for neutral speech (Pell et al. 2009, Scherer 1986). A low F0 in sadness sentences has also been observed for English by Murray and Arnott (2008) and for Swedish by Laukka et al. (2011).

The F0 results for the present study revealed that in Estonian read-out speech, the emotion of sadness shares the same F0 height with neutral sentences, which is lower than for joy and higher than for anger. Statistical significance was demonstrated only for the F0 difference between sadness and anger.

As for the F0 range, the CPM predicts that sadness would make it decrease. This has been demonstrated by several previous studies; notably, in German, Arabic and Hindi, sadness has the narrowest F0 range in comparison with joy, anger and neutral speech. At the same time, the English F0 for sadness is higher than for neutral speech. (Pell et al. 2009, Scherer 1986) The small F0 range of English sadness sentences has also been pointed out by Murray and Arnott (2008).

The results of the present study also show that sadness sentences have the narrowest F0 range compared with other emotions and neutral

speech, while its difference from the F0 range of anger and joy sentences is statistically significant.

In general, the present results obtained on emotional Estonian read-out speech indicate that Estonian joy and sadness affect the F0 parameters more or less similarly to other languages, but for anger, the similarity is only partial. Note that the difference in the observed behaviour of F0 height may be due either to the specificity of the linguistic and cultural space or to the research material.

One of the possible F0 patterns of Estonian emotions identifiable from voice is joy > neutral = sadness > anger, where the F0 is highest for joy and lowest for anger, while the F0 range is widest for joy and anger and narrowest for sadness (joy = anger > neutral > sadness).

As the present study used the recordings of just one reader, a follow-up study using the speech of several readers would be welcome in order to show whether the described model is dominant in Estonian or if there are other equally relevant patterns in vocal expression of Estonian emotions.

6. Conclusion

The aim of the study was to find out if sentence emotion has any effect on the height and/or range of the fundamental frequency (F0) of read-out Estonian speech. The analysis revealed that in the material investigated, the anger sentences have the lowest F0 compared with neutral speech as well as with the emotions of anger and joy, while anger also has the widest F0 range. The highest F0 is associated with joy. The F0 range for sentences of joy is narrower than for anger, yet wider than for neutral speech or sadness. Sadness sentences have a similar F0 height to neutral speech, which is lower than that of either joy or anger. Sentences of sadness have the narrowest F0 range of all.

The most distinctive Estonian emotion, as revealed by the F0 parameters used, was anger, as the respective difference of both its F0 height and range from those of joy, sadness and neutral speech was statistically significant in most cases.

Address:
   Kairi Tamuri
   Institute of the Estonian Language
   Roosikrantsi 6
   10119 Tallinn, Estonia
E-mail: kairi.tamuri@eki.ee

References

Altrov, Rene (2013) "Aspects of cultural communication in recognizing emotions". *Trames* 17, 159–174.

Altrov, Rene and Hille Pajupuu (2013) "Estonian emotional speech corpus: content and options". In J. Bamford, S. Cavalieri, and G. Diani, eds. *Variation and Change in Spoken and Written Discourse*, Amsterdam, 109–122.

Arias, Juan P., Carlos Busso, and Nestor B. Yoma (2014) "Shape-based modeling of the fundamental frequency contour for emotion detection in speech". *Computer Speech and Language* 28, 278–294.

Ariste, Paul (1953) *Eesti keele foneetika.* Tallinn: Eesti Riiklik Kirjastus.

Boersma, Paul and David Weenink (2014) *Praat: doing phonetics by computer. [Computer program] version 5.3.85.* Available online at <http://www.praat.org/>. Accessed on 05.12.2014.

Busso, Carlos, Sungbok Lee and Shrikanth Narayanan (2009) "Analysis of emotionally salient aspects of fundamental frequency for emotion detection". *IEEE Trans Audio Speech Lang Process* 17, 4, 582–596.

El Ayadi, Moataz, Mohamed S. Kamel, and Fakhri Karray (2011) "Survey on speech emotion recognition: features, classification schemes and databases". *Pattern Recognition* 44, 572–587.

Koolagudi, Shashidhar G. and K. Sreenivasa Rao (2012) "Emotion recognition from speech: a review". *International Journal of Speech Technology* 15, 2, 99–117.

Laukka, Petri, Daniel Neiberg, Mimmi Forsell, Inger Karlsson, and Kjell Elenius (2011) "Expression of affect in spontaneous speech: acoustic correlates and automatic detection of irritation and resignation". *Computer Speech & Language* 25, 1, 84–104.

Murray, Iain R. and John L. Arnott (2008) "Applying an analysis of acted vocal emotions to improve the simulation of synthetic speech". *Computer Speech and Language* 22, 2, 107–129.

Nose, Takashi and Takao Kobayashi (2011) "Recent development of HMM-based expressive speech synthesis and its applications". *Proceedings of APSIPA ASC 2011*. Available online at <http://www.apsipa.org/proceedings_2011/pdf/APSIPA189.pdf>. Accessed on 05.12.2014.

Pell, Marc D., Silke Paulmann, Chinar Dara, Areej Alasseri, and Sonja A. Kotz (2009) "Factors in the recognition of vocally expressed emotions: a comparison of four languages". *Journal of Phonetics* 37, 417–435.

Scherer, Klaus R. (1986) "Vocal affect expression: a review and a model for future research". *Psychological Bulletin* 99, 2, 143–165.

Scherer, Klaus R. (2009) "The dynamic architecture of emotion: evidence for the component process model". *Cognition and Emotion* 23, 7, 1307–1351.

Scherer, Klaus R. and Ben Meuleman (2013) "Human emotion experiences can be predicted on theoretical grounds: evidence from verbal labeling". *PLoS ONE* 8, 3, e58166.

Schröder, Marc (2009) "Expressive speech synthesis: past, present, and possible futures". *Affective Information Processing*, 111–126.

Schuller, Björn, Anton Batliner, Stefan Steidl, and Dino Seppi (2011) "Recognising realistic emotions and affect in speech: state of the art and lessons learnt from the first challenge". *Speech Communication* 53, 9–10, 1062–1087.

Soto, Jose Angel and Robert W. Levenson (2009) "Emotion recognition across cultures: the influence of ethnicity on empathic accuracy and physiological linkage". *Emotion* 9, 6, 874–884.

Tamuri, Kairi (2010) "Kas pausid kannavad emotsiooni?". *Eesti Rakenduslingvistika Ühingu Aastaraamat* 6, 297–306.

Tamuri, Kairi (2012a) "Kas formandid peegeldavad emotsioone?". *Eesti Rakenduslingvistika Ühingu Aastaraamat* 8, 231–243.

Tamuri, Kairi (2012b) "Intensity of estonian emotional speech". *Human Language Technologies – The Baltic Perspective*, Proceedings of the Fifth International Conference Baltic HLT 2012 (IOS Press), 238–246.

Tamuri, Kairi and Meelis Mihkla (2012) "Emotions and speech temporal structure". *Linguistica Uralica* 3, 209–217.

Kokkuvõte. Kairi Tamuri: Põhitoon eestikeelses emotsionaalses etteloetud kõnes. Põhitoon on oluline emotsiooni kandev akustiline tunnus. Põhitooni uurimuse eesmärk oli teada saada, kas lause emotsioon mõjutab eestikeelse etteloetud kõne põhitooni kõrgust ja selle ulatust. Selleks mõõdeti etteloetud emotsionaalse ning neutraalse kõne lausetes kõigi vokaalide põhitooni ning arvutati välja kolme emotsiooni (viha, rõõmu, kurbuse) ja neutraalse kõne põhitooni mediaan. Lisaks mõõdeti emotsionaalse ja neutraalse kõne põhitooni ulatust, samuti lausealguse ja -lõpu põhitooni kõrgust. Tulemustest selgus, et uuritud eestikeelses etteloetud emotsionaalses kõnes on kõige kõrgem põhitoon rõõmul, kõige madalam vihal. Põhitooni ulatus on kõige laiem vihal, kõige kitsam aga kurbusel. Põhitooni kõrguse erinevused lause alguses ja lõpus polnud emotsioonipaariti ja võrdluses neutraalse kõnega statistiliselt olulised.

Märksõnad: eesti keel, emotsionaalne kõne, kõne akustika, põhitoon, sünteeskõne, emotsioonide akustiline mudel