



Kairi Tamuri
Institute of the Estonian Language

Purpose

Currently the Estonian Emotional Speech Corpus is investigated for the distinctive acoustic parameters of three emotions – anger, joy and sadness – and neutral speech, with a view to recognizable synthesis of emotions in Estonian speech. This research is focused on intensity as one of the parameters vital for emotion synthesis.

Intensity is a physical characteristic of voice, which reflects the volume of energy in the speech wave and the effort of speaking, perceived as loudness and measured in decibels.

Research question

- Whether the intensity of read Estonian emotional speech is in any way affected by emotions?

Material

The material comes from the Estonian Emotional Speech Corpus (<http://peeter.eki.ee:5000>) of the Institute of the Estonian Language. The corpus contains journalistic passages read by a female voice. The emotions are not acted but *elicited* (i.e. it is semantic content of the passage that elicits the reader's emotion). The read passages have been segmented into sentences, words and speech sounds. The corpus sentences have passed both the listening and the reading test. The sentences, all with a different semantic content, have been classified into anger, joy, sadness and neutral. An emotion is considered as recognized if at least 51% of the listeners have been unanimous over it. Those sentences where the semantic content has not affected emotion identification (i.e. where the listening results differ from the results of the reading test) serve as material for the present study (see table below).

Emotion	No of sentences	No of vowels in the sentences
<i>anger</i>	79	1435
<i>joy</i>	60	973
<i>sadness</i>	87	1807
<i>neutral</i>	103	2194
<i>TOTAL</i>	329	6409

Method

The mean intensities of sentences and the intensity range between max and min were computed for each emotion class. In addition intensity was measured at the beginning and end of all sentences.

The mean intensities were computed from the intensities of the middle of all vowels. In order to find the intensity range the difference between highest and lowest intensities measured for each sentence and a median was taken for each emotion and the neutral group. In order to find the intensity at the beginning and end of sentences the intensity was measured at the first (stressed) vowel of the first word of the sentence and the first (stressed) vowel of the sentence-final word.

The measurements were done using the Praat program and the EMU speech database system, using R for statistical analysis.

Results

• *The mean intensity in emotional and in neutral speech:*

neutral (median 71.6 dB) > anger (71.1 dB) > joy (70.6 dB) > sadness (70.3 dB).

Emotion	Anger	Joy	Sadness	Neutral
<i>min</i>	58.9	60	57.9	61.9
<i>Q1</i>	67.9	67.7	66.9	69.2
<i>median</i>	71.1	70.6	70.3	71.6
<i>Q3</i>	74	73.1	73.1	74.1
<i>max</i>	82.9	81.1	82.2	81.3

In order to see whether there is significant difference between the mean intensities of different emotion groups a Wilcoxon test was used. According to the results intensity difference is statistically significant both for emotion pairs and in comparison with neutral speech.

Pairs	p-value
<i>anger vs. joy</i>	0.008
<i>anger vs. joy</i>	0.001
<i>anger vs. joy</i>	0.001
<i>anger vs. joy</i>	0.020
<i>anger vs. joy</i>	0.001
<i>anger vs. joy</i>	0.001

• *Intensity range in emotional and in neutral speech:*

sadness (median 14.7 dB) > anger (14.3 dB) > neutral (13.7 dB) > joy (13.2 dB).

Emotion	Anger	Joy	Sadness	Neutral
<i>min</i>	6.6	5.6	4.2	6.5
<i>Q1</i>	12.5	11	11.7	12.2
<i>median</i>	14.3	13.2	14.7	13.7
<i>Q3</i>	18.5	17.2	19.7	17.6
<i>max</i>	25.8	25.9	30.8	24.9

A Wilcoxon test was run to see whether the pairwise ranges obtained differed significantly. The results reveal no significant difference either between the intensity ranges of emotions or between those of emotions and neutral speech.

• *Intensity at the beginning vs. end of a sentence in emotional and in neutral speech:*

Intensity at the beginning of sentences:

neutral (median 74.7 dB) > joy (73.9 dB) > anger (73.8 dB) > sadness (72.3 dB).

Intensity at the end of sentences:

neutral (65.3 dB) > anger (64.9 dB) > joy (64.5 dB) > sadness (63.6 dB).

To find out whether emotion groups differ from each other significantly by beginning and end characteristics, a Wilcoxon test was used. According to the results intensity at the beginning and end of sentences differ significantly only in pair *neutral vs. sadness* (both $p = 0.002$). Other pairs do not have statistically significant difference.

Conclusions

- **The highest intensity** was typical of neutral speech, while the lowest intensity signalled of sadness. The differences between the mean intensities were also significant statistically, both between emotions examined pairwise and in comparison with neutral speech.
- An analysis of the **range of intensity** showed that the widest amplitude of intensity was characteristic of sadness-sentences, while the narrowest range was typical of joy-sentences. The differences of intensity for emotion pairs as well as in comparison with neutral speech were not statistically significant.
- **At the beginning and end of a sentence intensity** was the highest for neutral speech and the lowest for sadness. Those two, neutral and sadness, were the only emotions whose intensity differences was statistically significant, both at the beginning and end of a sentence.