Human Language Technologies – The Baltic Perspective
A. Tavast et al. (Eds.)

doi:10.3233/978-1-61499-133-5-238

Intensity of Estonian Emotional Speech

Kairi TAMURI[1]
Institute of the Estonian Language, Tallinn

Abstract. 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 article is focused on intensity as one of the parameters vital for emotion synthesis. The research question is whether the intensity of Estonian read speech is in any way affected by emotions. The Estonian Emotional Speech Corpus was used as the acoustic basis of the study. The intensity analysis comprised calculations of the means and ranges of the intensities of emotional and neutral speech. In addition, pairwise studies were applied to find out whether intensity differs across emotions and in comparison with neutral speech in utterance-initial and utterance-final positions. The results revealed that mean intensities make a significant difference between concrete emotions as well as in comparison with neutral speech. The highest intensity was measured in neutral speech and the lowest in the utterances of sadness. Intensity ranges, however, were not significantly different between the utterance groups analysed. Intensity at the beginning and end of utterance was also the highest in neutral speech and the lowest with sadness. Those two groups displayed the only statistically significant differences between the intensities of utterance beginnings as well as ends.

Keywords. emotional speech, speech acoustics, intensity, Estonian

Introduction

One of the recent challenges of Estonian text-to-speech synthesis consists in adding emotionality to synthetic text [1]. For this purpose we are developing a model of emotional speech, which should provide an acoustic description of the three major emotions – anger, joy and sadness – and define a combination of acoustic characteristics enabling perceivable distinction between those three emotions in speech.

According to many studies (e.g. [2], [3], [4], [5], [6]), pitch and intensity are relevant acoustic parameters for the detection of emotion, as well as segmental duration and speaking rate and accuracy of articulation (the list is not closed).

For Estonian speech we have so far been able to examine pause duration [7] and speaking rate [8] and (in connection with accuracy of articulation) the quality (F1, F2) of the vowels *a*, *i* and *u* [9]. The results, compared with other relevant studies (e.g. [10],[11],[12],[13],[14]), have hitherto revealed relatively universal tendencies, notably,

[1] Corresponding Author: Analyst, Institute of the Estonian Language, Roosikrantsi 6, Tallinn, 10119 Estonia; E-mail: kairi.tamuri@eki.ee.

active / high-arousal emotions[2] such as anger and joy are expressed at a higher speech rate than neutral, whereas passive / low-arousal emotions such as sadness is expressed at a considerably slower speech rate than neutral. As for Estonian vowel quality it is relatively little affected by emotion, only in the case of sadness a loss of quality in *a* was observed, while the vowel approached to the centroid of the vowel space.

The present study is focused on intensity. 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. There is a relation between the speaker's state of emotion and the loudness of his/her speaking ([2], [15], [6]). The present article investigates the intensity means and ranges (between max and min) in the utterances expressing anger, joy or sadness and in neutral utterances.

Although it has been established that higher arousal is accompanied by higher intensity and vice versa, ([2], [15], [6]), this does not provide knowledge about how this or that particular emotion is expressed in this or that particular culture: emotions traditionally expressed loudly in one culture may be expressed with much more reserve in another. Thus, in addition to universal tendencies there are emotion patterns, which are language and culture dependent ([16], [17], [18[), and that is why the ways of expressing emotions should be studied specifically for each language and each culture.

In order to improve studying and mutual comparison of emotions Scherer [19] has elaborated a theoretical model to predict the influence of an emotion on acoustic expression. Scherer's *component process model* (CPM) considers both psychological and physiological factors accompanying emotional expression and causing changes in the acoustics of a speech signal. E.g. a sensation of something extremely unpleasant often causes a tightness felt in the pharynx and larynx, resulting in tension in the vocal tract and a higher pitch of the outcoming [6]. The CPM can also predict how different emotions may affect the intensity of the speech signal (loudness of speech). Scherer has tested and improved his model on the basis of a literature survey of acoustic-phonetic evidence as well as carried out emotion studies of his own ([6], [19]).

The present study explores the role of intensity as a characteristic of the vocal expression of anger, joy and sadness in Estonian speech. On the basis of the CPM and the available results we may, of course, speculate on what intensity strength is likely to characterize this or that emotion, but as the available studies have used very different material (acted or elicited emotions, professional or non-professional reading etc.) and also different languages the results can be interpreted as merely orienting.

For **anger** the CPM predicts that both hot anger / rage and cold anger / irritation cause a rise in all intensity parameters (mean, range and variability[3]). In this study both variants of anger are regarded as belonging to one and the same class labelled “anger”. Most studies of emotional speech have observed a rise of intensity in anger-utterances: Banse and Scherer [6] made such an observation for acted German speech; Murray and Arnott's [2] analogous observation concerns the content words in acted English speech;

[2] In research on emotions dominate two approaches: categorical and dimensional. In categorical approach emotions are treated as categorically distinct variables (sadness, anger, joy, disgust etc.). Dimensional approach describes emotions in multidimensional space. The two most commonly used dimensions are arousal (high vs. low) and valence (positive vs. negative). In arousal/valence space joy = high arousal and positive valence, anger = high arousal and negative valence, sadness = low arousal and negative valence.

[3] Mean – energy values for a speech sound wave averaged over an utterance; range – difference between the highest and lowest intensity values in an utterance; variability – measure of the dispersion of intensity values in an utterance (e.g., standard deviation) [19].

heightened intensity in the case of anger has also been pointed out by Johnstone and Scherer [15] (meta-analytic review).

Of **pleasant** emotions the CPM contains two categories: joy/elation and happiness/enjoyment, which in our treatment both belong to the class labelled as "joy". CPM predicts that joy brings a rise in all intensity parameters (mean, range and variability), whereas in the case of happiness a fall of all three is in order. Murray and Arnott [2], however, point out heightened intensity in the case of happiness. Heightened intensity for happiness has also been confirmed by Banse and Scherer [6]. Again, Johnstone and Scherer [15] have found heightened intensity in joy vs falling intensity in happiness.

For **sadness** the CPM predicts a fall in all intensity parameters (mean, range and variability). Murray and Arnott [2] confirmed that sadness is characterized by low intensity that falls during the utterance. Falling intensity has also been pointed out by Banse and Scherer [6] and Johnstone and Scherer [15].

The aim of the present study of the intensity of read Estonian emotional speech is to find out if there is an intensity difference between the anger, joy, sadness and neutral sentences in Estonian speech.

1. Material and Method

The material studied comes from the Estonian Emotional Speech Corpus[4] (EESC) of the Institute of the Estonian Language. The corpus contains journalistic passages read by a female voice. Passage selection is based on the principle that it is the semantic content of the passage that elicits the reader's emotion and thus the reader is never dictated what emotion she should use when reading. [20] The reader is a non-professional with a pleasant female voice. The sentences of the read passages, all with a different semantic content, have been classified into anger, joy, sadness and neutral sentences by subjects of listening tests. An emotion is considered as recognized if at least 51% of the listeners have been unanimous over it. [21]

As the emotion is not acted but *elicited* the expression in speech is moderate, often hardly perceptible; neither have we got full-blown emotions, but rather *emotion related states*. Thus our *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. In order to make a difference between the corpus sentences where the semantic content may have affected emotion identification by ear and those where it has not, all corpus sentences have passed a reading test requiring identification of the emotion from the written text, without hearing it. As a result the corpus sentences are additionally classified (see Table 1) into:

• sentences where the content did not affect emotion identification (the results of reading tests differ from the results of listening tests);

• sentences where the content might have affected emotion identification (the results of reading tests coincide with the results of listening tests).

The corpus contains 1,234 sentences that have passed both the listening and the reading test. 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)

[4] http://peeter.eki.ee:5000/

serve as material for the present study (Table 1, sentence types 1 and 2); for the number of the sentences investigated see Table 2.

Table 1. Classification of emotions in the corpus by emotion identification in reading and listening tests (test results in %) [22]

Tests	Joy	Anger	Sadness	Neutral	Not sure	Sentence type in corpus
1. Ehkki Ott minu olemasolust midagi ei teadnud. [Although Ott knew nothing of my existence.]						
By listening	**87.5**	0.0	0.0	12.5	-	**Joy**, no content influence
By reading	4.0	0.0	32.0	32.0	32.0	
2. Ükskõik, mida ma teen, ikka pole ta rahul! [Whatever I do, he is never satisfied!]						
By listening	0.0	14.3	**80.0**	5.7	-	**Sadness**, no content influence
By reading	0.0	**64.3**	35.7	0.0	0.0	
3. Täiesti mõistetamatu! [Completely incomprehensible!]						
By listening	0.0	**100.0**	0.0	0.0	-	**Anger**, content influence
By reading	0.0	**83.0**	0.0	11.0	5.6	

Johnstone and Scherer [15] have noted that although intensity is not difficult to measure, it is sensitive to recording conditions, therefore it is important to take account of the distance of the microphone from the reader, whether the recording room is quiet enough, absence of background noise, whether the recording device has been calibrated etc. All these circumstances may affect the measurements. The corpus sentences used in the present study have been recorded in a quiet room, using a digital recorder and a high-quality microphone located at about 50 cm from the speaker (wav-format, 44.1 KHz, 16Bit, Mono), and segmented into words and speech sounds using Praat (textgrid-format) [23]. In this study the mean intensities of utterance 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 utterances. The mean intensities were computed from the intensities of the middle of all vowels used in the utterances making up a particular group (anger, joy, sadness, neutral) (for material see Table 2). In order to find the intensity range of the groups of utterances the difference between highest and lowest intensities measured for each utterance were calculated and a median was taken for each emotion and the neutral group. In order to find the intensity at the beginning and end of utterances the intensity was measured at the first (stressed) vowel of the first word of the utterance and the first (stressed) vowel of the utterance-final word.

Table 2. Material used for intensity analysis

Emotion	No of utterances	No of vowels in the utterances
anger	79	1435
joy	60	973
sadness	87	1807
neutral	103	2194
TOTAL	329	6409

The measurements were done using the Praat[5] program and the EMU[6] speech database system, using R[7] for statistical analysis. The results were compared with Scherer's CPM and with other studies.

2. Results

2.1. The mean intensity in emotional and in neutral speech

In order to find out whether the emotion of an utterance may affect the volume of the speech signal intensity was first measured on all vowels, whereupon a median was taken for each emotion group (see Table 3).

Table 3. Intensity in emotional and in neutral speech (in dB)

	anger	joy	sadness	neutral
min	58.9	60	57.9	61.9
Q1	67.9	67.7	66.9	69.2
median	71.1	70.6	70.3	71.6
Q3	74	73.1	73.1	74.1
max	82.9	81.1	82.2	81.3

According to the results the mean intensity is the highest in neutral speech (median 71.6 dB), followed by anger (71.1 dB) and joy (70.6 dB). The lowest intensity reading is characteristic of sadness (70.3 dB). In order to see whether there is significant difference between the mean intensities of different utterance groups a Wilcoxon test was used, the results of which are presented in Table 4 below.

[5] http://www.praat.org/
[6] http://emu.sourceforge.net/
[7] http://www.r-project.org/

Table 4. Wilcoxon test results for the mean intensities in emotion pairs and in comparison with neutral speech ($p < 0.05$ refers to statistically significant difference)

Pairs	p-value
anger vs. joy	0.008
anger vs. sadness	0.001
anger vs. neutral	0.001
joy vs. sadness	0.020
joy vs. neutral	0.001
sadness vs. neutral	0.001

As can be seen from Table 4 intensity difference is statistically significant both for emotion pairs and in comparison with neutral speech.

2.2. Intensity range in emotional and in neutral speech

Besides the mean intensities the range of intensity within an utterance was examined. The medians of the intensity range for emotional and neutral speech are presented in Table 5.

Sadness has the widest range of intensity (median 14.7 dB), followed by anger (14.3 dB) and neutral speech (13.7 dB). The narrowest range of intensity was observed in joy (13.2 dB). A Wilcoxon test was run to see whether the pairwise ranges obtained differed significantly (see Table 6).

Table 5. Range of intensity (dB) in emotional and in neutral speech

	anger	joy	sadness	neutral
min	6.6	5.6	4.2	6.5
Q1	12.5	11	11.7	12.2
mediaan	14.3	13.2	14.7	13.7
Q3	18.5	17.2	19.7	17.6
max	25.8	25.9	30.8	24.9

Table 6. Wilcoxon test results for the intensity ranges in emotion pairs and in comparison with neutral speech ($p < 0.05$ refers to statistically significant difference)

Pairs	p-value
anger vs. joy	0.280
anger vs. sadness	0.930
anger vs. neutral	0.800
joy vs. sadness	0.250
joy vs. neutral	0.800
sadness vs. neutral	0.800

The results reveal no significant difference either between the intensity ranges of emotions or between those of emotions and neutral speech.

2.3. Intensity at the beginning vs. end of an utterance in emotional and in neutral speech

Apart from the mean intensities and intensity ranges it was studied whether intensity differs at the beginning and end of an utterance. The measurements for emotion pairs and in comparison with neutral speech are presented in Table 7.

Table 7. Intensity at the beginning and end of utterances in emotional and in neutral speech (dB)

	Beginning of utterance / end of utterance			
	anger	**joy**	**sadness**	**neutral**
min	65.6/58.2	68.2/59.1	64.3/54.7	65.3/56.6
Q1	70.8/62.6	71.6/63	70/61.1	72/62.9
median	73.8/64.9	73.9/64.5	72.3/63.6	74.7/65.3
Q3	76.7/67.1	75.8/66.3	74.9/65.5	76.4/67.5
max	82.7/73.7	81.5/69.1	81.8/71	81.1/73.2

Table 7 shows that intensity at the beginning of utterances is the highest in neutral speech (median 74.7 dB), followed by joy (73.9 dB), anger (73.8 dB) and sadness (72.3 dB). Intensity at the end of utterances is the lowest in neutral speech (65.3 dB), followed by anger (64.9 dB), joy (64.5 dB) and 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 utterances differ significantly only in pair sadness vs. neutral (both $p = 0.002$). Other pairs do not have statistically significant difference.

3. Discussion

According to the results of this study the most intensive utterances were those of neutral speech, whereas the lowest intensity was characteristic of sadness. The rank order of the four utterance groups when placed along the intensity dimension was as follows: neutral > anger > joy > sadness. Like in the case of mean intensities, the intensity at the beginning of an utterance was the highest in neutral speech and the lowest in sad utterances, the rank order being: neutral > joy > anger > sadness. Almost the same applies to the end of utterances, where neutral showed the highest intensity and sadness the lowest, with a slight change in the central part of the rank order: neutral > anger > joy > sadness. The intensity range was the widest for sadness and the narrowest for joy, the rank order being: sadness > anger > neutral > joy.

Thus in Estonian neutral speech is read louder than emotional speech. A closer look at the measurements reveals that in neutral speech the reader's voice strength varied less than in emotional speech, i.e. the variation of intensity during an utterance was considerably smaller for neutral speech than, for example, for sadness or anger. It may well be a peculiarity of Estonian speech that neutral speech is louder and keeping a more constant loudness than emotional speech.

For anger (both cold and hot anger) the CPM predicts a rise in the mean intensity as well as in the intensity range. That hypothesis has been tested out in several studies of emotional speech ([2], [15], [6]). In Estonian emotional speech, too, anger goes with

high intensity (highest of the three emotions studied) and a wide range of intensity (only sadness having a still wider one).

As for utterances carrying the positive emotions of joy and happiness the CPM predicts a rise of the mean and range of intensity only for joy, whereas for happiness a fall in both parameters is predicted. Similar results have been reported in several other studies ([2], [15], [6]). In Estonian emotional speech the joy-utterances have a low intensity (lower than anger and higher than sadness) and a narrow range (the narrowest of all emotions). This suggests that the joy of the present study is more like happiness.

For sadness the CPM predicts a fall in the intensity mean as well as range. The CPM prediction has been confirmed by several studies of emotional speech (e.g. [2], [15], [6]). According to the present results, however, only the intensity falls with sadness, whereas the range of intensity gets wider, making sadness the emotion with the greatest amplitude of intensity (loudness) in Estonian speech.

The results on Estonian read emotional speech enable the conclusion that mean intensity is an important parameter in emotion distinction (and, consequently, in emotion modelling and synthesis). Although the measured differences in intensity range were not statistically significant either in emotion pairs or in comparison with neutral speech, the range was wide enough, both in emotional and neutral speech, to be considered in synthesis for the sake of naturalness.

4. Conclusion

The aim of the study was to find out whether intensity might be a parameter enabling a distinction between anger, joy, sadness and neutrality in Estonian utterances.

The measurements demonstrated that 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-utterances, while the narrowest range (the least variation) was typical of joy-utterances. 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 an utterance intensity was the highest for neutral speech and the lowest for sadness. Those two, neutral and sadness, were the only utterance groups whose intensity was significant statistically, both at the beginning and end of an utterance.

Acknowledgements

The study was supported by the Estonian Ministry of Education and Research (grants SF0050023s09 and EKT11001).

References

[1] M. Mihkla, I. Hein, M-L. Kalvik, I. Kiissel, R. Sirts, K. Tamuri, Estonian speech synthesis: Applications and challenges, *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue"* **11**, 18 (2012), 443–453.

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

[3] L. ten Bosch, Emotions, speech and the ASR framework, *Speech Communication* **40**, 1–2 (2003), 213–226.

[4] A. Iida, N. Campbell, F. Higuchi, M. Yasumura, A corpus-based speech synthesis system with emotion, *Speech Communication* **40**, 1–2 (2003), 161–187.

[5] R. Picard, *Affective Computing*. Cambridge, Massachusetts: The MIT Press, 1997.

[6] R. Banse, K.R. Scherer, Acoustic profiles in vocal emotion expression, *Journal of Personality and Social Psychology* **70**, 3 (1996), 614–636.

[7] K. Tamuri, Kas pausid kannavad emotsiooni?, *Eesti Rakenduslingvistika Ühingu Aastaraamat* **6** (2010), 297–306.

[8] K. Tamuri, M. Mihkla, Emotions and speech temporal structure, *Linguistica Uralica* (2012), forthcoming.

[9] K. Tamuri, Kas emotsioonid peegelduvad formantides?, *Eesti Rakenduslingvistika Ühngu Aastaraamat* **8** (2012), 231–243.

[10] F. Burkhardt, N. Audibert, L. Malatesta, O. Türk, L.M. Arslan, V. Auberge, Emotional prosody – does culture make a difference?, *Proceedings of Speech Prosody. Dresden, Germany* (May 2-5 2006).

[11] G. McIntyre, R. Göcke, Researching emotions in speech, P. Warren, C. I. Watson (Eds.), *Proceedings of the 11th Australian International Conference on Speech Sciences & Technology. New Zealand: University of Auckland* (2006), 264–269.

[12] J. Wilting, E. Krahmer, M. Swerts, Real vs. acted emotional speech, *Proceedings of Interspeech ICSLP, Pittsburgh, PA, USA* (2006), 805–808.

[13] E. Douglas-Cowie, N. Campbell, R. Cowie, P. Roach, Emotional speech: Towards a new generation of databases, *Speech Communication* **40**, 1–2 (2003), 33–60.

[14] K.R. Scherer, Vocal communication of emotion: A review of research paradigms, *Speech Communication* **40**, 1–2 (2003), 227–256.

[15] T. Johnstone, K.R. Scherer, Vocal communication of emotikon, *M. Lewis, J. Haviland (Eds.). Handbook of Emotion, 2nd ed. New York: Guilford*, 220–235.

[16] H. Elfenbein, N. Ambady, Cultural Similarity's Consequences: A Distance Perspective on Cross-Cultural Differences in Emotion Recognition, *Journal of Cross-Cultural Psychology* **34**, 1 (2003), 92–110.

[17] R. Altrov, H. Pajupuu, Estonian Emotional Speech Corpus: Culture and age in selecting corpus testers. I. Skadiņa, A. Vasiļjevs (Eds.), *Human Language Technologies – The Baltic Perspective – Proceedings of the Fourth International Conference Baltic HLT*, *Amsterdam: IOS Press* (2010), 25–32.

[18] N. Kamaruddin, A. Wahab, C. Quek, Cultural dependency analysis for understanding speech emotion, *Expert Systems with Applications* **39**, 5 (2012), 5115–5133.

[19] K.R. Scherer, Vocal affect expression: a review and a model for future research, *Psychological bulletin* **99**, 2 (1986), 143–65.

[20] R. Altrov, Eesti emotsionaalse kõne korpus: teoreetilised toetuspunktid, *Keel ja Kirjandus* **4** (2008), 261–271.

[21] R. Altrov R, H. Pajupuu, The Estonian Emotional Speech Corpus: Release 1, Čermak F, Marcinkevičienė R, Rimkutė E, Zabarskaitė J (Eds.), *Proceedings of the Third Baltic Conference on Human Language Technologies.* Vilnius: Vytauto Didžiojo Universitetas, Lietuviu Kalbos Institutas (2008), 9–15.

[22] R. Altrov, H. Pajupuu, Estonian Emotional Speech Corpus: Theoretical base and implementation, L. Devillers, B. Schuller, A. Batliner, P. Rosso, E. Douglas-Cowie, R. Cowie, C. Pelachaud (Eds), *4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals (ES3)*, Istanbul (2012), 50–53.

[23] P. Boersma, D. Weenink, *Praat: doing phonetics by computer [Computer program]. Version 5.3.18*, retrieved 15 June 2012 from http://www.praat.org/.