

KAIRI TAMURI

Basic emotions in read Estonian speech:  
acoustic analysis and modelling





**KAIRI TAMURI**

Basic emotions in read Estonian speech:  
acoustic analysis and modelling



UNIVERSITY OF TARTU  
Press

University of Tartu, Institute of Estonian and General Linguistics

Dissertation accepted for the commencement of the degree of Doctor of Philosophy on June 26<sup>th</sup>, 2017 by the Committee of the Institute of Estonian and General Linguistics, Faculty of Philosophy, University of Tartu

Supervisors: Academician Professor Karl Pajusalu  
(University of Tartu)

Leading Researcher Hille Pajupuu  
(Institute of the Estonian Language)

Opponent: Professor Jean Léo Léonard (Université Paris-Sorbonne)

Academician Professor Jaan Ross (Estonian Academy of Music  
and Theatre)

Commencement: October 6<sup>th</sup>, 2017 at 14.15, Senat Hall in the University main building, Ülikooli 18–204, Tartu

This study has been supported by the Graduate School of Linguistics, Philosophy and Semiotics; funded by the (European Union) European Social Fund.



European Union  
European Social Fund



Investing  
in your future

ISSN 1406-1325  
ISBN 978-9949-77-539-2 (print)  
ISBN 978-9949-77-540-8 (pdf)

Copyright: Kairi Tamuri, 2017

University of Tartu Press  
[www.tyk.ee](http://www.tyk.ee)

## ACKNOWLEDGEMENTS

When I luckily arrived at the Institute of the Estonian Language, they had just started creating an emotion corpus. My first study addressed pauses in read-aloud texts of fiction, which provided apt material for my Master's thesis. Further on, my attention was attracted by the acoustics of emotional speech, with the first focus on pauses, of course. This was the beginning of my doctoral studies the fruits of which have now been gathered between the covers you are presently holding.

While writing the thesis, I constantly sensed the presence of certain people whose wise advice, firm support, endless patience, safeguarding benevolence and wonderful example never failed to guide, encourage and inspire me. These are my honorable supervisors Leading Researcher Hille Pajupuu ja Academician Professor Karl Pajusalu. I am extremely grateful to both of them. I also owe a lot of thanks to PhD Meelis Mihkla and PhD Heete Sakkai for thinking along and helping me in my endeavour.

My deep thanks also go to Academician Professor Jaan Ross and Professor Jean Léo Léonard for their valuable observations and helpful comments. I also thank all my considerate colleagues from the Institute of the Estonian Language and the University of Tartu. I am also grateful to Sirje Ainsaar and Jaan Pajupuu for their specific contributions, as well as to all those obliging persons who granted their valuable time to take my Listening tests.

A warm feeling of gratitude envelops me every time I think of my family – my parents, my brother, and my daughter Ingrid – and of my dear friends. Thank you so much for your faith, support and empathy.

My research for the present dissertation has received financial support from several project sources, notably including the project Estonian Emotional Speech Corpus under the National Programme for the Estonian Language, the specifically funded project Modelling Intermodular Phenomena in Estonian, the Estonian Science Foundation research grant project Bidirectional Relationships between Estonian Grammatical and Lexical Structures and the Speech Rhythm, the project Statistical Models of the Emotionality of Speech and Written Text under the National Programme for the Estonian Language, and the institutional research funding project Speech Styles, Sentence Prosody, Phonological Variation: Description. Theory and Modelling. My genuine thanks!

## EESSÕNA

Kui mina Eesti Keele Instituuti jõudsin, oli emotsioonikorpusse loomine juba alanud. Instituudis alustasin tööd ettelõetud ilukirjanduslike tekstide pauside uurimisega, mis oli ühtlasi minu magistr töö teema. Sealt edasi hakkasin tegelema juba emotsionaalse kõne akustikaga ja muidugi alustasin ma taas pausidest. Nii said alguse minu doktoriõpingud, mille viljad on nüüd koos nende kaante vahel, mida käes hoiate.

Väitekirja kirjutades on minu kõrval olnud inimesed, kes on mind oma targa nõu, kindla toe, katkematu kannatuse, turvalisust loova heatahtlikkuse ning imelise eeskujuga suunanud, julgustanud ja innustanud. Need on minu juhendajad juhtivteadur dr Hille Pajupuu ja akadeemik professor Karl Pajusalu. Olen neile mõlemale südamest tänulik. Samuti võlgnen palju tänu dr Meelis Mihklale ja dr Heete Sahkaile kaasamõtlemise ja abi eest.

Minu suur tänu kuulub ka professor akadeemik Jaan Rossile ja professor Jean Léo Léonardile väärtuslike tähelepanekute ja nõuannete eest. Samuti tänan kõiki abivalmis kolleege Eesti Keele Instituudist ja Tartu Ülikoolist. Olen tänulik ka Sirje Ainsaarele ja Jaan Pajupuule nende panuse eest väitekirja valmimisse ning samuti kõigile neile vastutulelikele inimestele, kes ohverdasid oma hinnalist aega, et vastata kuulamistestide küsimustele.

Soe tänutunne valdab mind ka siis, kui mõtlen oma perekonnale – vanematele, vennale ja tütar Ingridile – ning oma kallitele sõpradele. Aitäh usu, toe ja kaasaalamise eest!

Oma väitekirja valmimisele olen saanud toetust mitmest projektist ja teadusteemast: riikliku programmi „Eesti keele keeletehnoloogiline tugi“ projekt „Eesti emotsionaalse kõne korpus“, sihtfinantseeritav teadusteema „Eesti keele alusuuringud keeletehnoloogiliste rakenduste teenistuses“, Eesti Teadusfondi uurimistoetuse projekt „Eestikeelse kõne rütmilisuse peegeldused grammatilistes ja leksikaalsetes struktuurides (ja *vice versa*)“, riikliku programmi „Eesti keeletehnoloogia“ projekt „Kõne ja teksti emotsionaalsuse statistilised mudelid“ ning institutsionaalse uurimistoetuse projekt „Kõnestiilid, lauseprosoodia ja fonoloogiline varieerumine: kirjeldus, teooria ja modelleerimine“. Suur tänu!

## CONTENTS / SISUKORD

TABLES / TABELID .....	10
FIGURES .....	13
LIST OF AUTHOR'S PUBLICATIONS / AUTORI PUBLIKATSIOONIDE LOETELU .....	14
1. INTRODUCTION .....	15
1.1. Objectives .....	15
1.2. Structure of the thesis .....	17
1.3. Overview of publications and the author's contribution .....	17
2. THE INFLUENCE OF JOY, SADNESS AND ANGER ON SPEECH ACOUSTIC PARAMETERS .....	19
2.1. Pauses and how they may be influenced by emotions .....	21
2.2. Formant frequencies and articulation precision and how they may be influenced by emotions .....	21
2.3. Speech rate and how it may be influenced by emotions .....	22
2.4. Intensity and how it may be influenced by emotions .....	23
2.5. Fundamental frequency and how it may be influenced by emotions .....	23
3. MATERIAL AND METHODS .....	25
3.1. Material and methods used in the analysis of speech acoustic parameters .....	25
3.1.1. Material and method used in studying pauses .....	27
3.1.2. Material and method used in studying formant frequencies ...	29
3.1.3. Material and method used in studying speech rate .....	30
3.1.4. Material and method used in studying intensity and fundamental frequency .....	31
3.2. Material and method used to create acoustic models of synthetic emotional speech .....	32
4. RESULTS .....	35
4.1. Results on the acoustic parameters of emotional speech .....	35
4.1.1. Results on pauses .....	35
4.1.2. Results on formant frequencies .....	40
4.1.3. Results on speech rate .....	42
4.1.4. Results on intensity .....	43
4.1.5. Results on fundamental frequency .....	46
4.2. Results of the experiment to create acoustic models for emotional synthetic speech .....	49
4.2.1. Results of the evaluation of test models of emotional speech synthesis for a male voice .....	49
4.2.2. Results of the evaluation of the test models of emotional speech synthesis for a female voice .....	50

5. DISCUSSION .....	53
5.1. Speech acoustic parameters .....	53
5.1.1. Pauses .....	53
5.1.2. Formant frequencies and articulatory precision .....	55
5.1.3. Speech rate .....	57
5.1.4. Intensity .....	58
5.1.5. Fundamental frequency .....	59
5.1.6. Acoustic patterns of joy, sadness and anger .....	60
5.2. Acoustic models of Estonian emotional synthetic speech .....	61
6. CONCLUSION .....	64
7. SUMMARY IN ESTONIAN. Põhiemotsioonid eestikeelses ettelõetud kõnes: akustiline analüüs ja modelleerimine .....	66
7.1. Väitekirja eesmärk .....	66
7.2. Väitekirja struktuur .....	68
7.3. Väitekirja publikatsioonide ülevaade ja autori panusest kaasautoriga artiklites .....	68
7.4. Rõõmu, kurbuse ja viha mõju kõne akustilistele parameetritele .....	69
7.4.1. Pausid ja emotsioonide mõju nendele .....	71
7.4.2. Formantsagedused ja artikulatsiooni täpsus ning emotsioonide mõju nendele .....	72
7.4.3. Kõnetempo ja emotsioonide mõju sellele .....	73
7.4.4. Intensiivsus ja emotsioonide mõju sellele .....	73
7.4.5. Põhitoon ja emotsioonide mõju sellele .....	74
7.5. Materjal ja meetodid .....	75
7.5.1. Kõne akustiliste parameetrite analüüsi materjal ja meetodid .....	76
7.5.1.1. Pauside uurimismaterjal ja -meetod .....	78
7.5.1.2. Formantsageduste ja artikulatsiooni täpsuse uurimismaterjal ja -meetod .....	79
7.5.1.3. Kõnetempo uurimismaterjal ja -meetod .....	80
7.5.1.4. Intensiivsuse ja põhitooni uurimismaterjal ja -meetod .....	81
7.5.2. Emotsionaalse sünteeskõne akustiliste mudelite loomise materjal ja meetod .....	82
7.6. Tulemused .....	85
7.6.1. Emotsionaalse kõne akustiliste parameetrite uurimistulemused .....	85
7.6.1.1. Pauside uurimuse tulemused .....	85
7.6.1.2. Formantsageduste ja artikulatsiooni täpsuse uurimuse tulemused .....	89
7.6.1.3. Kõnetempo uurimuse tulemused .....	92
7.6.1.4. Intensiivsuse uurimuse tulemused .....	93
7.6.1.5. Põhitooni uurimuse tulemused .....	95
7.6.2. Emotsionaalse sünteeskõne akustiliste mudelite loomise eksperimendi tulemused .....	97



7.6.2.1. Emotsionaalse sünteeskõne katsemudelite hindamise tulemused meessünteesääle põhjal .....	97
7.6.2.2. Emotsionaalse sünteeskõne katsemudelite hindamise tulemused naissünteesääle põhjal .....	99
7.7. Diskussioon .....	100
7.7.1. Kõne akustilised parameetrid .....	100
7.7.1.1. Pausid .....	100
7.7.1.2. Formantsagedused ja artikulatsiooni täpsus .....	102
7.7.1.3. Kõnetempo .....	104
7.7.1.4. Intensiivsus .....	105
7.7.1.5. Põhitoon .....	106
7.7.1.6. Rõõmu, kurbuse ja viha akustilised mustrid .....	108
7.7.2. Emotsionaalse sünteeskõne akustilised mudelid .....	108
7.8. Kokkuvõte .....	110
REFERENCES / VIIDATUD KIRJANDUS .....	112
APPENDIX 1 / LISA 1 .....	117
APPENDIX 2 / LISA 2 .....	125
PUBLICATIONS / PUBLIKATSIOONID .....	137
CURRICULUM VITAE .....	233
ELULOOKIRJELDUS .....	235

## TABLES / TABELID

<b>Table 1.</b>	Material used for the study of pauses .....	27
<b>Table 2.</b>	Material used for formant analysis .....	29
<b>Table 3.</b>	Material for studying the speech rate .....	30
<b>Table 4.</b>	Material for studying the speech intensity and fundamental frequency .....	31
<b>Table 5.</b>	Parametric models of Estonian emotional synthetic speech .....	33
<b>Table 6.</b>	Number and location of pauses .....	35
<b>Table 7.</b>	Commas .....	36
<b>Table 8.</b>	Breathing pauses .....	36
<b>Table 9.</b>	Proportion of the breathing pauses coinciding with punctuation marks .....	37
<b>Table 10.</b>	The results of the perception test (rate of identification) .....	38
<b>Table 11.</b>	The average durations of pause groups in milliseconds .....	38
<b>Table 12.</b>	The average values of the first and second formant frequencies of the vowels <i>a</i> , <i>i</i> and <i>u</i> , measured in Hertz/Bark for four emotion groups .....	40
<b>Table 13.</b>	The results of ANOVA on the $F_1$ of the short vowels <i>a</i> , <i>i</i> and <i>u</i> in emotional and neutral sentences .....	41
<b>Table 14.</b>	The results of ANOVA on the $F_2$ of the short vowels <i>a</i> , <i>i</i> and <i>u</i> in emotional and neutral sentences .....	41
<b>Table 15.</b>	Euclidean distance between vowels and the central sound <i>x</i> (Hertz/Bark) .....	42
<b>Table 16.</b>	The average speech rates for emotional and neutral speech (sounds per second) .....	42
<b>Table 17.</b>	The results of ANOVA on average speech rate differences in emotion pairs .....	43
<b>Table 18.</b>	The level of intensity (dB) in emotional and neutral speech .....	44
<b>Table 19.</b>	The results of the Wilcoxon rank-sum test on the level of intensity across emotion groups .....	44
<b>Table 20.</b>	The range of intensity (dB) for emotional and neutral speech ....	45
<b>Table 21.</b>	The results of the Wilcoxon rank-sum test for the intensity range in emotion pairs and in comparison with neutral speech ...	45
<b>Table 22.</b>	The level of intensity (dB) at sentence beginning and sentence end in emotional and neutral speech .....	45
<b>Table 23.</b>	The results of the Wilcoxon rank-sum test on the level of intensity at sentence beginning and sentence end in emotional and neutral speech .....	46
<b>Table 24.</b>	The fundamental frequency level (Hz) in emotional and neutral speech .....	46
<b>Table 25.</b>	The results of the Wilcoxon rank-sum test on the $F_0$ level in emotion pairs and in comparison with neutral speech .....	47

<b>Table 26.</b> The range of fundamental frequency (Hz) in emotional and neutral speech .....	47
<b>Table 27.</b> The results of the Wilcoxon rank-sum test on the $F_0$ range between emotions and in comparison with neutral speech .....	48
<b>Table 28.</b> The $F_0$ level value (Hz) at sentence beginning vs sentence end in emotional and neutral speech .....	48
<b>Table 29.</b> Confusion matrix (synthetic male voice, Tests A/B). Emotion perception from a synthetic male voice in three different test models, in percent of responses .....	50
<b>Table 30.</b> Confusion matrix (female synthetic voice, Test A/B). Emotion perception from a synthetic female voice in three different test models, in percent of responses .....	51
<b>Table 31.</b> The preferred emotion models for a male and a female voice synthesised parametrically for synthetic Estonian speech .....	63
<b>Tabel 32.</b> Pauside uurimuse materjal .....	78
<b>Tabel 33.</b> Formantanalüüsi materjal .....	80
<b>Tabel 34.</b> Kõnetempo uurimismaterjal .....	81
<b>Tabel 35.</b> Intensiivsuse ja põhitooni uurimismaterjal .....	82
<b>Tabel 36.</b> Eestikeelse emotsionaalse sünteeskõne parameetrilised mudelid .....	83
<b>Tabel 37.</b> Pauside arv ja asukoht .....	86
<b>Tabel 38.</b> Komad .....	86
<b>Tabel 39.</b> Hingamispausid .....	86
<b>Tabel 40.</b> Hingamispauside osakaal kirjavahemärkide kohal .....	87
<b>Tabel 41.</b> Tajukatse tulemused (tuvastusprotsendid) .....	88
<b>Tabel 42.</b> Pausirühmade keskmised kestused millisekundites .....	88
<b>Tabel 43.</b> Vokaalide $a$ , $i$ ja $u$ esimese ja teise formantsageduse keskmised (Hertzi skaala / Barki skaala) emotsiooniti .....	90
<b>Tabel 44.</b> Dispersioonanalüüsi (ANOVA) tulemused emotsionaalse ja neutraalse kõne lausete lühikeste vokaalide $a$ , $i$ ja $u$ esimese formantsageduse ( $F_1$ ) kohta .....	90
<b>Tabel 45.</b> Dispersioonanalüüsi (ANOVA) tulemused emotsionaalse ja neutraalse kõne lausete lühikeste vokaalide $a$ , $i$ ja $u$ teise formantsageduse ( $F_2$ ) kohta .....	91
<b>Tabel 46.</b> Eukleidiline kaugus vokaalide ja neutraalse hääliku $x$ vahel (Hertsi skaala / Barki skaala) .....	91
<b>Tabel 47.</b> Emotsionaalse ja neutraalse kõne tempo keskmised (häälikut sekundis) .....	92
<b>Tabel 48.</b> Dispersioonanalüüsi (ANOVA) tulemused kõnetempo erinevuste kohta emotsioonipaariti .....	92
<b>Tabel 49.</b> Intensiivsuse tase (dB) emotsionaalses ja neutraalses kõnes .....	93
<b>Tabel 50.</b> Wilcoxon'i astaksummatesti tulemused intensiivsuse taseme kohta emotsioonipaariti ja võrdluses neutraalse kõnega .....	93
<b>Tabel 51.</b> Intensiivsuse ulatus (dB) emotsionaalses ja neutraalses kõnes ...	93
<b>Tabel 52.</b> Wilcoxon'i astaksummatesti tulemused intensiivsuse ulatuse kohta emotsioonipaariti ja võrdluses neutraalse kõnega .....	94

<b>Tabel 53.</b> Lause alguse ja lõpu intensiivsuse tase (dB) emotsionaalses ja neutraalses kõnes .....	94
<b>Tabel 54.</b> Wilcoxon'i astaksummatesti tulemused emotsionaalse ja neutraalse kõne intensiivsuse taseme kohta lause alguses ja lõpus .....	95
<b>Tabel 55.</b> Põhitooni kõrgus (Hz) emotsionaalses ja neutraalses kõnes .....	95
<b>Tabel 56.</b> Wilcoxon'i astaksummatesti tulemused põhitooni kõrguse kohta emotsioonipaariti ja võrdluses neutraalse kõnega .....	95
<b>Tabel 57.</b> Põhitooni ulatus (Hz) emotsionaalses ja neutraalses kõnes .....	96
<b>Tabel 58.</b> Wilcoxon'i astaksummatesti tulemused põhitooni ulatuse kohta emotsioonipaariti ja võrdluses neutraalse kõnega .....	96
<b>Tabel 59.</b> Lause alguse ja lõpu põhitooni kõrgus (Hz) emotsionaalses ja neutraalses kõnes .....	97
<b>Tabel 60.</b> Eksimismatriks (meessüntheeshäl, A-/B-test). Emotsioonide tajumine meessüntheeshälest kolme erineva katsemudeli põhjal, vastused protsentides .....	98
<b>Tabel 61.</b> Eksimismatriks (naissüntheeshäl, A-/B-test). Emotsioonide tajumine naissüntheeshälest kolme erineva katsemudeli põhjal, vastused protsentides .....	99
<b>Tabel 62.</b> Eelistatud emotsioonimudelid mees- ja naissüntheeshäele eestikeelse parameetrilise kõnesüntheesi jaoks .....	110

## FIGURES

<b>Figure 1.</b> Commas in emotional speech, where the reader made a pause (% of all commas) .....	36
<b>Figure 2.</b> The overall speech rate in emotional and neutral speech .....	43
<b>Figure 3.</b> The level of intensity in emotional and neutral speech (medians) .....	44
<b>Figure 4.</b> The level of fundamental frequency in emotional and neutral speech (medians) .....	46
<b>Figure 5.</b> The range of fundamental frequency (medians) in emotional and neutral speech .....	47

## LIST OF AUTHOR'S PUBLICATIONS / AUTORI PUBLIKATSIOONIDE LOETELU

- [P1] Tamuri, Kairi. 2010. Kas pausid kannavad emotsiooni? *Eesti Rakenduslingvistika Ühingu aastaraamat*, 6, 297–306.
- [P2] Tamuri, Kairi. 2012. Kas formandid peegeldavad emotsioone? *Eesti Rakenduslingvistika Ühingu aastaraamat*, 8, 231–243.
- [P3] Tamuri, Kairi & Mihkla, Meelis. 2012. Emotions and speech temporal structure. *Linguistica Uralica*, 3, 209–217.
- [P4] Tamuri, Kairi. 2012. Intensity of Estonian emotional speech. *Human Language Technologies – The Baltic Perspective – Proceedings of the Fifth International Conference Baltic HLT 2012*, IOS Press, 238–246.
- [P5] Tamuri, Kairi. 2015. Fundamental frequency in Estonian emotional read-out speech. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics*, 6 (1), erinumber / special issue: *Papers from the conference “Finnic Languages, Cultures, and Genius Loci”*, 9–21.
- [P6] Tamuri, Kairi & Mihkla, Meelis. 2015. Expression of basic emotions in Estonian parametric text-to-speech synthesis. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics*, 6(3), erinumber / special issue: *Kõneuurimise suundi / Aspects of speech studies*, 145–168.

# 1. INTRODUCTION

## 1.1. Objectives

The contemporary history of studying the vocal expression of emotions goes back to the end of the 19th century, when it was mainly an interest in psychology and psychiatry. Owing to technological development, research opportunities have since widened, and emotional speech has attracted the interest of some additional areas, including linguistics, primarily in connection with speech technology (see, e.g., Burkhardt & Campbell, 2015; Juslin & Scherer, 2005; Scherer, 2003; Schröder, 2001).

Speech technology includes speech synthesis as one of its branches, the task of which is to convert a written text to a spoken one, imitating human speech. Speech synthesis is applied in many areas, such as human-machine interaction, multimedia, and aids for the people with visual, reading or speech impairments – hence, the necessity for synthetic speech to sound as natural or human-like as possible, in all aspects.

Expression of emotions is one of the essential aspects of human speech. Being ever present in human speech, emotions should also be perceptible in the synthetic speech. There are several ways to add emotions to synthetic speech. One is to provide the synthesiser with acoustic models of emotional speech (see, e.g., Audibert, Aubergé & Rilliard, 2005; Iriondo, Alías, Melenchón & Llorca, 2004). To manage that, one needs to know, on the one hand, which values of acoustic parameters are influenced by emotions to what extent and in what direction; on the other hand, the limitations of the parametric tuning of a concrete synthesiser have to be considered.

According to the results of emotion acoustics, every emotion has a characteristic combination of acoustic parameter values, or an acoustic pattern, which distinguishes the concrete emotion from other emotions (see, e.g., Banse & Scherer, 1996; Juslin & Scherer, 2005; Paeschke, Kienast & Sendlmeier, 1999). We also know that listeners can guess speech emotions from mere sound, without seeing the speaker's face (like in telephone conversations) (Bachorowski, 1999). Moreover, it is possible to guess the emotion of an utterance consisting solely of meaningless words (Goudbeek, Goldman & Scherer, 2009). The fact that listeners can reliably recognise emotions from voice only leads to the assumption that voice must indeed carry some information of the speaker's emotions, and the vocal expressions of those emotions have specific acoustic patterns.

A speech wave consists of numerous acoustic parameters that can be influenced by emotions. Part of those parameters have been studied more thoroughly than others for the main reason that they are more easily accessible to measurement and analysis, being also successfully applicable in modern speech synthesis<sup>1</sup> and speech recognition (El Ayadi, Kamel & Karray, 2011;

---

<sup>1</sup> Examples of emotions synthesised for various languages can be heard at <http://emosamples.syntheticspeech.de/> (last checked 2017, August 21).

Koolagudi & Rao, 2012; Nose & Kobayashi, 2011; Schröder, 2009; Schuller, Batliner, Steidl & Seppi, 2011). The parameters most frequently mentioned in research as conveying emotions are the fundamental frequency, speech rate and intensity. Besides those three, many others have been pointed out, including pauses, formants and articulatory precision (Juslin & Scherer, 2005; Koike, Suzuki & Saito, 1998; Montero, Gutierrez-Arriola, Palazuelos, Enriquez, Aguilera & Pardo, 1998; Murray & Arnott, 2008; Toivanen, Waaramaa, Alku, Laukkanen, Seppänen, Väyrynen & Airas, 2006).

The acoustic expression of emotions can differ across languages and cultures, while vocal emotions are easier to recognise from voice within a language or culture (see, e.g., Altrov & Pajupuu, 2015; Elfenbein, 2013; Kamaruddin, Wahab & Quek, 2012; Paulmann & Uskul, 2014; Soto & Levenson, 2009). Consequently, there cannot be any automatic takeover of the results based on other languages or cultures, and a special study is needed to create acoustic models for an Estonian synthesiser of emotional speech.

In Estonia, emotional speech was first addressed in a speech technological context in 2006, when the project of creating an Estonian Emotional Speech Corpus was launched at the Institute of the Estonian Language.<sup>2</sup> There were two basic principles: first, the Corpus was intended as a reliable database for emotion research and, second, it had to be able to serve speech technological applications (Altrov, 2014; Altrov & Pajupuu, 2012). As the Institute was, at the same time, engaged in the development of Estonian speech synthesis (see Mihkla, 2009; Mihkla, Piits, Nurk & Kiissel, 2008), synthesis of emotional speech was seen as the primary high-tech application of the Corpus (Pajupuu, 2012).

By 2008, the Institute of the Estonian Language had provided a background enabling me to start research on the expression of emotions in Estonian speech. My aim was to acquire some basic knowledge of emotional speech acoustics and, based on that knowledge, to create some acoustic models helping an Estonian speech synthesiser to produce emotionally coloured speech.

**My doctoral dissertation has two objectives: (a) to find out and describe the acoustic expression of three basic emotions – joy, sadness and anger – in read Estonian speech and (b) to create some acoustic models for parametric synthesis of emotional speech in an Estonian speech synthesiser, which would help the synthesiser recognisably express joy, sadness and anger.**

Based on the above aims, I set the following **research questions**: (1) To what extent, if any, and in what direction do three emotions (joy, sadness and anger), as manifested in read Estonian speech and compared to each other and to neutral speech, influence the values of certain acoustic parameters (pauses, formants, intensity, speech rate and fundamental frequency)? (2) Which of

---

<sup>2</sup> The project Estonian Emotional Speech Corpus under the National Programme for Estonian Language Technology (2006–2010). For the Corpus see <http://peeter.eki.ee:5000/> (last checked 2017, August 21).



those acoustic parameters are distinctive of the three emotions and a neutral attitude in read Estonian emotional speech? (3) How should the results in the creation of acoustic models of emotional speech be applied for Estonian text-to-speech synthesis?

## **1.2. Structure of the thesis**

The dissertation consists of an introductory part and six publications (see [P1], [P2], [P3], [P4], [P5], [P6]). The introductory part is divided into six chapters. Chapter 1 presents the objectives of the study, describes the structure of the dissertation and gives a short survey of the publications included and the author's contribution to joint articles. Chapter 2 presents the theoretical starting points about the influence of three basic emotions – joy, sadness and anger – on speech acoustic parameters. Chapter 3 describes the material and method used in the acoustic analysis of the three basic emotions and the material and method used to create the acoustic models for emotional synthetic speech. Chapter 4 presents the results of the measurements of the acoustic parameters of Estonian emotional speech and the experimental results of creating acoustic models for emotional synthetic speech. Chapter 5 contains a comparison of the results of the present and previous studies, suggestions for emotion modelling for parametric speech synthesis, an evaluation of our experiment to create acoustic models for emotional synthetic speech and the most preferred emotion models for three basic emotions and two synthetic voices used in Estonian parametric speech synthesis. Chapter 6 summarises the dissertation. The introductory part is followed by Chapter 7, which presents an Estonian introduction to the dissertation. Chapter 7 is followed by the list of references.

## **1.3. Overview of publications and the author's contribution**

The six publications presented fall into two thematic groups. The first five address the acoustics of read Estonian emotional speech. Those articles provide a survey of the extent, if any, and direction of the influence of the emotion of a sentence on the values of the acoustic parameters (pauses, formants, articulatory precision, speech rate, intensity and fundamental frequency) of speech. The sixth publication describes the process and results of an experiment performed to create acoustic models for Estonian emotional synthetic speech. Two publications ([P3], [P6]) have been written jointly with Meelis Mihkila, who has read and accepted the present surveys of those publications.

[P1] discusses pauses in read Estonian emotional speech. The article provides an answer to two research questions: (1) Do the number, position, nature and duration of pauses depend on text emotion? (2) Are pauses important in distinguishing emotions, and can the emotion of a speech passage be recognised from pause differences only?

[P2] concentrates on the formants and articulatory precision of read Estonian emotional speech. The article investigates whether emotions have any influence on the first and second formant frequencies of vowels and on the precision of the speaker's articulation, and whether differences in the values of those parameters can make a difference between the emotions as well as between emotional and neutral speech.

[P3] focuses on speech temporal structure and how it could be influenced by emotions. Two research questions are answered: (1) Do emotions influence speech rate (and are differences in speech rate distinctive of different emotions as well as of neutral speech)? (2) Do emotions make differences in word prosody? The contribution of the first author of the article concerns speech rate – in particular, its theoretical aspects, the description of the research material and method and the presentation and analysis of measurements.

[P4] reports a study of the intensity in read Estonian emotional speech. The article answers two research questions: (1) Do emotions influence the intensity level and intensity range of the speaker's voice? (2) Are the differences between the values of those parameters significantly distinctive of the three emotions and neutral speech?

[P5] analyses the fundamental frequency in read Estonian emotional speech. The article answers two questions: (1) Does sentence emotion influence the level and range of fundamental frequency? (2) Are the differences between the values of those parameters significantly distinctive of the three emotions and neutral speech?

[P6] describes an attempt to create, based on the results of ([P3], [P4], [P5]), some acoustic models of emotional speech that would help Estonian parametric speech synthesis to express three basic emotions (joy, sadness and anger), both in male and female synthetic voices. The first author of the article has written the theoretical part concerning the acoustics of Estonian emotional speech and, together with the second author, has created the experimental models, compiled the perception tests, carried out the tests and analysed the results.

## 2. THE INFLUENCE OF JOY, SADNESS AND ANGER ON SPEECH ACOUSTIC PARAMETERS

Human speech is never stable but prone to variation in the height, range and strength of the speaking voice, as well as in the speech rate. One of the reasons behind such variation is the speaker's emotions (the rest may be due, for example, to the speaker's sex, age, social status or health condition) (Juslin & Scherer, 2005).

Studies of emotional speech conducted for speech technology are mainly focused on basic emotions (Cowie, Douglas-Cowie, Tsapatsoulis, Votsis, Kollias & Fellenz, 2001; Iida, Campbell, Higuchi & Yasumura, 2003; Juslin & Laukka, 2003), which mostly include joy, sadness, anger, fear, disgust and surprise (Ekman, 1992).

Before starting on acoustic models of emotional speech, one has to find out how exactly emotions are vocally expressed. In 1986, Klaus R. Scherer devised a model to predict emotional influence on vocal expression. Scherer's *component process model* (CPM) considers the psychological and physiological factors involved in emotional expression, demonstrating the existence of emotion-specific acoustic patterns. Scherer has tested and improved his model, not only on the basis of acoustic and phonetic facts found in the relevant literature but also based on his own studies of emotion (Banse & Scherer, 1996; Scherer, 1986; Scherer, 2009; Scherer & Meuleman, 2013).

Scherer (1986) describes an emotion as a series of adaptive changes; namely, stimulated by an emotional impulse, the nervous system will influence breathing as well as the muscular tension in speech organs, which leads to changes in the acoustics of the speech signal. For example, the sensation of something very unpleasant will usually cause a tightness or pressure in the throat and neck area, resulting in tension in the vocal tract and hence a higher frequency of the emitted voice (Thompson & Balkwill, 2006). Acoustic changes in a speech signal can also be elicited by facial expressions, such as smiling (Tartter, 1980).

Although different studies of emotion acoustics have shown that every emotion has an acoustic pattern of its own, the results on emotion-carrying parameters are far from homogeneous and sometimes even controversial (Cowie, Douglas-Cowie, Tsapatsoulis, Votsis, Kollias & Fellenz, 2001; Murray & Arnott, 1993). This can be due to various reasons. One of the reasons may lie in different research materials (acted vs spontaneous vs read speech with elicited emotion). For emotion acoustics, each of them has its pros and cons as research material. Acted speech, for example, is usually recorded in controlled conditions in a sound studio, where there is no background noise and the speaker's distance from the microphone can be controlled to avoid the adverse effects of its fluctuation on the values of speech intensity. Also, it is likely that in acted speech we find more emotionally enhanced values of acoustic parameters than in read or spontaneous speech. Moreover, an actor need not always act emotions as natural and spontaneous as they come in real life. As a result, emotions of

acted speech may display a different acoustic pattern than the spontaneous emotions of real life (Juslin & Scherer, 2005; McIntyre & Göcke, 2006; Scherer, 2013; Wilting, Krahmer & Swerts, 2006).

In natural speech, emotions are expressed spontaneously. Such material is authentic and thus best for the synthesis of natural-sounding speech. However, natural speech can also have its flaws regarding emotion studies: namely, the speakers can feel disturbed by the recording situation, which makes them suppress their emotions; also, recordings of spontaneous emotions need not always be of high quality (due to, for example, background noise or simultaneous talk of the interlocutors) (Douglas-Cowie, Campbell, Cowie & Roach, 2003; Juslin & Scherer, 2005; McIntyre & Göcke, 2006).

Elicited emotions fall midway between acted and spontaneous speech in that the speaker is given a text containing a certain emotion, which is elicited while reading and reflected in the reading voice. Elicited emotions can also be recorded in the controlled conditions of a sound studio, which makes the recording good material for acoustic analysis. However, it can happen that elicited emotions are but moderately expressed, which weakens the distinctive powers of their acoustic parameters (Juslin & Scherer, 2005; McIntyre & Göcke, 2006; Scherer, 2003).

Another reason behind the differences in emotion acoustics results may lie in the language and culture investigated; namely, speaking is a linguistic/cultural act, while every community has its own rules for emotional expression (Altrov & Pajupuu, 2015; Douglas-Cowie, Campbell, Cowie & Roach, 2003; Johnstone & Scherer, 2000). In addition, emotional expression can be influenced by a speaker's individual specifics (Juslin & Scherer, 2005).

Still another reason for different results may arise from the definition of emotions, the reference of emotion terms: notably, anger can be understood as rage as well as resentment; joy can range from contentment to elation, while sadness may cover disappointment as well as sorrow. Even if the general term (e.g., joy or anger) may be the same, contentment and elation or resentment and rage, respectively, can display quite different acoustic patterns (Burkhardt, Audibert, Malatesta, Türk, Arslan & Auberge, 2006; Douglas-Cowie, Campbell, Cowie & Roach, 2003; McIntyre & Göcke, 2006; Scherer, 2003; Wilting, Krahmer & Swerts, 2006).

In order to be able to create acoustic models of speech synthesis, one needs to know to what extent, if any, and in what direction emotions may shift the values of speech acoustic parameters compared to those of neutral speech. Despite the occasional contradictions in the results on emotion acoustics, research has shown the existence of certain general tendencies between emotions and their acoustic correlates (Goudbeek, Goldman & Scherer, 2009; Iida, Campbell, Higuchi & Yasumura, 2003; Kienast & Sendlmeier, 2000; Murray & Arnott, 2008; Paeschke, Kienast & Sendlmeier, 1999).

In what follows, I describe the mutual relations between three basic emotions (joy, sadness and anger) and five acoustic parameters (pauses, formants, speech rate, intensity and fundamental frequency).

## 2.1. Pauses and how they may be influenced by emotions

*Pause* is a voiceless interval (silence) in a sentence or between two sentences. Pauses are usually measured in milliseconds (ms). Research has shown that across emotions pauses may differ by duration as well as by frequency of occurrence, while in some cases, listeners can tell emotions apart from mere pause differences. The most distinctive pauses are those of sadness, being longer and more frequent than those of joy or anger utterances (Murray & Arnott, 2008; Schröder, Cowie, Douglas-Cowie, Westerdijk & Gielen, 2001; Tisljár-Szabó & Pléh, 2014; Yildirim, Bulut, Lee, Kazemzadeh, Deng, Lee, Narayanan & Busso, 2004). Scherer's CPM does not provide pause prediction.

## 2.2. Formant frequencies and articulation precision and how they may be influenced by emotions

*Formant frequencies* are frequency areas where most acoustic energy is concentrated in the speech spectrum. The lowest of such frequency areas is called  $F_1$ , the next  $F_2$ , and so on. The formant frequencies are in Hertz (Hz) and Barks (a psychoacoustic scale). All voiced sounds have a specific formant structure. The positions of the formant frequencies are never fixed, depending on the volume and shape of the vocal tract. That volume and shape can, inter alia, be varied by emotions: Whenever a change occurs in a speaker's emotional state, it brings about changes in their facial expression as well as the position and muscular tension of their vocal organs (e.g., tongue, jaw and lips), which, in turn, leads to changes in the formant structure of the speech sounds emitted. Anger and joy, for example, bring more exertion to the speaking effort as the speaker tries to articulate more precisely. Muscular tension, however, causes a tightness in the throat. The tension and tightness lead to a shortening of the vocal tract, which may result in a higher  $F_1$  and a rise in articulatory precision. In the case of joy, the vocal tract may also be shortened by smiling. For sadness, however, there is no such muscular tension in the throat, and the speaker acts in a more relaxed manner. This may lead to a lower  $F_1$  and less precision in articulation (Johnstone & Scherer, 2000; Juslin & Laukka, 2003; Juslin & Scherer, 2005; Tartter, 1980; Tolkmitt & Scherer, 1986; Ververidis & Kotropoulos, 2006).

According to some results, the vowels of joy utterances show a rise in  $F_1$  and  $F_2$  values, while in most of the vowels,  $F_1$  rises more than  $F_2$  (Goudbeek, Goldman & Scherer, 2009; Juslin & Laukka, 2003; Kienast & Sendlmeier, 2000; Tartter, 1980). Some other scholars, however, have argued that pleasant emotions (e.g., joy) make  $F_1$  fall (Pribil, Pribilova & Matousek, 2013).

For the emotion of sadness, a fall of vowel  $F_1$  has been observed (Juslin & Laukka, 2003). It has also been noticed that the vowels of sadness utterances have a lower  $F_2$  than those of joy (Goudbeek, Goldman & Scherer, 2009).

For the emotion of anger, a relatively higher  $F_1$  and a lower  $F_2$  have been observed in vowels (Goudbeek, Goldman & Scherer, 2009; Juslin & Laukka, 2003; Murray & Arnott, 2008; Pribil, Pribilova & Matousek, 2013).

Scherer's CPM predicts that, in comparison with neutral speech, a rise in  $F_1$  and a fall in  $F_2$  is typical of the voiced sounds in utterances expressing hot anger (rage) as well as cold anger (irritation) and also in sadness utterances. For joy, the CPM predicts a fall in  $F_1$ , both for enjoyment and elation.

*Articulatory precision.* Each vowel has its own position in a two-dimensional vowel space, which is defined by vowel height or the vertical dimension ( $F_1$ ) and the front-back or horizontal dimension ( $F_2$ ). Those two dimensions depend on the position of the tongue when the vowel is articulated. For clear speech (precise articulation), the vowel space is large, and the vowels lie far enough from each other so that the listener can easily distinguish between them. If the vowel space is reduced, the speaker's articulatory precision is reduced as well, and the vowels lose in quality. It has been found that a stressed or depressed person will articulate voiced sounds with less effort than is normal for neutral speech. In this case, the speech sounds are reduced, and their articulation is less precise (Harrington, 2010: 46, 190–193; Tolkmitt & Scherer, 1986).

According to previous studies on emotion acoustics, the vowels of sadness utterances have low articulatory precision, in contrast to anger utterances, where the vowels are pronounced with higher precision (Juslin & Laukka, 2003; Kienast & Sendlmeier, 2000; Murray & Arnott, 2008; Paeschke, Kienast & Sendlmeier, 1999). However, results on anger differ. According to Iida, Campbell, Higuchi and Yasumura (2003) emotions accompanied by a heightened voice and a higher speech rate (like anger) go with a restricted vowel space and reduced vowels. Results on the articulatory precision of the vowels in joy utterances are also inconsistent, being reported either as rising (Johnstone & Scherer, 2000; Juslin & Laukka, 2003; Murray & Arnott, 2008) or similar to that of neutral speech (Kienast & Sendlmeier, 2000).

### **2.3. Speech rate and how it may be influenced by emotions**

*Speech rate*, or speaking rate, is measured by the number of speech segments produced per unit of time – sounds per second, for example. Results on emotion acoustics have proven that variation in speech rate may, inter alia, be indicative of the speaker's emotional state. A very slow speaking rate, for example, can be a signal of the speaker being sad or depressed. However, one should note that speech rate is also a rather subjective parameter, which, apart from emotional state, depends on the speaker's sex, age, style of speaking, language, culture, communicative situation, etc. (Laver, 1994: 534).

Although the available studies of emotional speech have used material from different languages and cultures, the results have been consistent in the following: Compared to neutral speech, the speech rate is lower in sadness utterances but higher than neutral for anger and joy. This suggests that, for those

three basic emotions, speech rate variation tends to be similar across languages and cultures (Juslin & Laukka, 2003; Murray & Arnott, 2008; Scherer, 2003; Scherer, Sundberg, Tamarit & Salomão, 2015; Yildirim, Bulut, Lee, Kazemzadeh, Deng, Lee, Narayanan & Busso, 2004).

Scherer's CPM predicts that, compared to neutral, the speech rate will rise for hot anger (rage) but fall for sadness. As for joy utterances, the CPM predicts a fall in the speech rate for enjoyment but a rise for elation.

## **2.4. Intensity and how it may be influenced by emotions**

*Intensity* shows the amount of energy contained in the speech wave, the effort made when speaking. Intensity level is perceived as volume or loudness and is measured in decibels (dB). Although intensity level is easy to measure, one has to consider its sensitivity to recording conditions, in particular, the distance between the speaker and the microphone, background noises, calibration of the recording device, etc., which can all affect the measurements.

*Intensity range* (the difference between the maximum and minimum intensity values within a speech unit) shows the variation of intensity level: the wider the range, the higher the variation in the intensity of speech.

According to earlier results on speech acoustics, there is a correlation between a speaker's emotional state and the intensity of his/her speech; notably, anger and elation are usually associated with a higher-than-neutral intensity level, whereas a lower-than-neutral intensity level, rather, goes with sadness and enjoyment (Banse & Scherer, 1996; Johnstone & Scherer, 2000; Juslin & Laukka, 2003; Murray & Arnott, 2008).

Scherer's CPM predicts a rise in the intensity level's parameters (mean and range) for elation, but a fall for enjoyment. The CPM prediction for sadness is a fall in the intensity parameters. For anger (both hot and cold), the CPM predicts a rise in the intensity parameters.

## **2.5. Fundamental frequency and how it may be influenced by emotions**

*Fundamental frequency* ( $F_0$ ) shows the frequency with which the speaker's vocal cords vibrate during speech. The listener perceives the fundamental frequency as pitch. The speech units produced at a higher vibration rate of the vocal cords have a higher fundamental frequency level, and for a listener, they sound as if they have a higher pitch. In terms of physics, pitch is frequency, which is usually measured in Hertz (Hz). The average level of fundamental frequency of a male voice remains between 80–150 Hz, while that of a female voice ranges from 150–300 Hz.

*The range of fundamental frequency* (the difference between the maximum and minimum values within a speech unit) characterises  $F_0$  variation, namely, the larger the range, the more dynamic the  $F_0$  of speech.

Fundamental frequency is one of the classical parameters addressed in studies of emotional speech, because both its level and range play an important part in telling apart emotions from each other and from neutral speech (Arias, Busso & Yoma, 2014; Banse & Scherer, 1996; Busso, Lee & Narayanan, 2009; Pell, Paulmann, Dara, Allasseri & Kotz, 2009; ten Bosch, 2003).

Results on the  $F_0$  of our three basic emotions are not quite consistent across scholars. *The emotion of joy* is usually associated with a high  $F_0$  level and a wide  $F_0$  range, but even here we can see contradictory results. The  $F_0$  of the joy emotion can be either lower or higher than that of the anger emotion, but it remains higher than that of the sadness emotion or of neutral speech. The range of the  $F_0$  for the joy emotion can be either wider or narrower than that of the anger emotion, remaining, however, wider than in the sadness emotion or in neutral speech (Murray & Arnott, 2008; Pell, Paulmann, Dara, Allasseri & Kotz, 2009). Scherer's CPM predicts that enjoyment will have a lower  $F_0$  and narrower  $F_0$  range than neutral speech, while elation utterances are likely to have a higher  $F_0$  and a wider  $F_0$  range (Scherer, 1986).

Earlier results on *the emotion of sadness* have shown that in a general case sad speech has a low  $F_0$  level. More precisely, it is lower than for the emotions of joy or anger, yet it can be either lower or higher than the  $F_0$  level of neutral speech. It has been observed that, compared to the  $F_0$  range for joy, anger and neutral speech, the sadness emotion has the narrowest range of  $F_0$ . However, some studies have found that the  $F_0$  range of speech expressing sadness may also be wider than that of neutral speech (Laukka, Neiberg, Forsell, Karlsson & Elenius, 2011; Murray & Arnott, 2008; Pell, Paulmann, Dara, Allasseri & Kotz, 2009). According to the prediction of Scherer's CPM, the  $F_0$  level is lower for sadness than for neutral speech, but depending on the type of voice (e.g., tense voice), it may even be higher. The  $F_0$  range will decrease for the sadness emotion (Scherer, 1986).

*The anger emotion* has been found to have the highest  $F_0$  level compared to the emotions of joy or sadness or to neutral speech. But it can also be lower than the  $F_0$  level of the joy emotion. The  $F_0$  range for anger has been described as being wider than in the case of sadness, joy or neutral speech. But it has also been described as narrower than the  $F_0$  range of the joy emotion (Laukka, Neiberg, Forsell, Karlsson & Elenius, 2011; Murray & Arnott, 2008; Pell, Paulmann, Dara, Allasseri & Kotz, 2009). Scherer's CPM predicts that, for cold anger (irritation), the  $F_0$  can be either higher or lower than in neutral speech, while the  $F_0$  range will decrease. For hot anger (rage), the  $F_0$  can also be either higher or lower than in neutral speech, but the  $F_0$  range will increase (Scherer, 1986).



### 3. MATERIAL AND METHODS

The dissertation consists of two major parts – first, a study of the acoustic parameters of emotional speech and, second, an experiment to create some acoustic models of emotional speech for an Estonian speech synthesiser. Each of the following two subsections, 3.1 and 3.2, addresses the material and method of each part, respectively.

#### 3.1. Material and methods used in the analysis of speech acoustic parameters

*The research material* of the acoustic expression of emotions derives from the Estonian Emotional Speech Corpus<sup>3</sup> (Altrov, 2014; Altrov & Pajupuu, 2012). The Corpus contains read-aloud Estonian<sup>4</sup> journalistic passages segmented into sentences, words and speech sounds. In total, the Corpus includes 1,234 sentences, which are all different. The Corpus uses a nonprofessional female reader of early middle age who has correct pronunciation and a pleasant voice (Altrov & Pajupuu, 2013).

The speech material of the Estonian Emotional Speech Corpus has been recorded in a quiet room, using a digital recorder and a high-quality microphone placed at a 50-cm distance from the speaker (WAV format, 44.1 kHz, 16 bit, mono).

The Estonian Emotional Speech Corpus rests on the principle that the content of the passage elicits an emotion in the reader and that emotion is reflected in the reader's voice. Otherwise stated, the Corpus emotions are not acted but elicited by text. In their definition of emotion, the authors of the Corpus have relied on the idea widely spread among speech technologists that emotion is something always present in speech except when the speaker is emotionless (Altrov, 2014; Cowie, Sussman & Ben-Ze'ev, 2011; Schuller, Batliner, Steidl & Seppi, 2011). The emotions caught in the Corpus are not "full-blown", being rather expressed modestly, but they also cover close emotions and emotion-like states (Altrov, 2014). Three basic emotions – joy, sadness and anger – have been chosen for the Estonian Emotional Speech Corpus, because it is those three emotions that are most needed in speech technological applications (see Altrov, 2014: 35). The Corpus material consists

---

<sup>3</sup> See <http://peeter.eki.ee:5000/> (last checked 2017, August 21).

<sup>4</sup> Estonian belongs to the Finnic subgroup of the Finno-Ugric family of the Uralic languages. Estonian is spoken as a mother tongue by nearly one million people, most of which live in Estonia, where Estonian is the national language. At the beginning of 2016 there were 883 707 native Estonian speakers in Estonia, which makes up 68% of the total population of Estonia (see <https://statistikaamet.wordpress.com/tag/emakeel/>, last checked 2017, August 21).

of journalistic passages, because the Estonian text-to-speech synthesiser is mainly meant for reading out journalistic texts (see Altrov, 2014: 36).

All Corpus sentences have been separated from the context (i.e., the passages included in the Corpus have been segmented into sentences). The concrete emotion of each sentence (joy, sadness, anger or neutral) has been detected by evaluators (30 evaluators for each sentence) who took perception tests (Listening and Reading). The evaluators were instructed to consider that the three above emotions were more like “umbrella emotions”, each covering several closely related emotions. Joy, for example, includes gratitude, happiness, pleasantness and exhilaration; sadness covers loneliness, disconsolation, concern and hopelessness, while anger includes resentment, rage, irony, reluctance, contempt and malice. Neutral speech was described to the evaluators as normal speech without any significant emotions. Evaluators were asked to guess the emotion of the sentences, not only in a Listening test but also in a separate Reading test, where the evaluators had to decide about the emotion just on the basis of the verbal content of the sentence, without hearing its vocal presentation (Altrov, 2014; Altrov & Pajupuu, 2012; Pajupuu, Pajupuu, Tamuri & Altrov, 2015).

Based on the results of the Reading and Listening tests, the Corpus sentences have been divided into two groups: (1) sentences where the verbal content and the voice convey a similar emotion (the same emotion has been identified both by readers and listeners); (2) sentences where the tone of the voice has changed the emotion (readers and listeners have identified different emotions) (see Pajupuu, Pajupuu, Tamuri & Altrov, 2015).

My research material consisted of sentences whose emotion had been identified from listening experience only (second group of the Corpus sentences, see before), that is, (1) sentences in the case of which the emotion guessed from the verbal content differed from the emotion identified from the audio presentation (see Ex. 1) and (2) sentences whose emotion was impossible to identify from the written version but was recognised while listening (see Ex. 2) (see also Pajupuu, Pajupuu, Tamuri & Altrov, 2015).

**Example 1 (sadness):** *Ükskõik, mida ma teen, ikka pole ta rahul!* [Whatever I do, he is never satisfied.] The emotion of the written sentence was classified under anger by 64.3% of the evaluators, whereas listening to the read sentence, 80.0% of the evaluators sensed sadness.

**Example 2 (joy):** *Ehkki Ott minu olemasolust midagi ei teadnud.* [Although Ott was not even aware of my existence.] The evaluators could not agree upon the emotion of the written sentence, but upon listening, 87.5% of the evaluators sensed joy.

The rate of Listening identification for the emotions or neutrality in the sentences used as research material was not less than 51% (i.e., more than two times better than chance).

The number of sentences<sup>5</sup> analysed differed for different tasks. Besides sentences, the Corpus could be searched for passages, which was necessary for the study of pauses (see 3.1.1).

*Methods.* The present study being a corpus-based one, the results have been subjected to various methods of statistical analysis (ANOVA, Wilcoxon's rank-sum test, confusion matrix and binary logistic regression analysis). The acoustic parameters were measured using the Praat<sup>6</sup> program (Boersma & Weenink, 2016) and the EMU Speech Database System<sup>7</sup>. The measurements were analysed using the statistical environment R<sup>8</sup>.

The following description addresses the material and methods for each acoustic parameter separately.

### 3.1.1. Material and method used in studying pauses

Pauses in emotional speech are the focus of article [P1], which consists of two parts: (1) a study trying to find out whether the number, location, nature and duration depend on text emotion and whether pauses are distinctive of emotions and (2) a perception test conducted to determine whether the emotion of a speech passage can be identified from mere pause differences.

In order to analyse the number, location, (including coordination with punctuation marks<sup>9</sup>), nature (breathing vs non-breathing pauses<sup>10</sup>) and duration of pauses, I selected 16 Corpus passages, on average, for each emotion<sup>11</sup>, while the passages contained 120–123 sentences<sup>12</sup> (see Table 1).

---

<sup>5</sup> The sentences of different emotions, which were used as research material are presented in Appendixes 1 and 2. The sentences can also be read and heard on the Corpus website at [peeter.eki.ee:5000/reports/segments?lg=en](http://peeter.eki.ee:5000/reports/segments?lg=en) (last checked 2017, August 21).

<sup>6</sup> Praat: <http://www.praat.org/> (last checked 2017, August 21).

<sup>7</sup> EMU <http://ips-lmu.github.io/EMU.html/> (last checked 2017, August 21).

<sup>8</sup> R <http://www.r-project.org/> (last checked 2017, August 21).

<sup>9</sup> I grouped the pauses by their coincidence with punctuation marks (e.g., full stop pause or comma pause).

<sup>10</sup> A breathing pause consists of an intake preceded and/or followed by silence. A non-breathing pause is a period of silence between words or sentences, which lasts at least 30 ms. I chose 30 ms to make my results comparable to previous studies of Estonian pauses, where 30 ms is the minimum length of a pause (see Pajupuu & Kerge, 2006; Tamuri, 2007: 14).

<sup>11</sup> In this study, emotion data were not compared with those of neutral speech because of the lack of reference material in the 2010 Emotional Speech Corpus.

<sup>12</sup> The material used in my pause analysis is presented in Appendix 1.

**Table 1.** Material used for the study of pauses (see [P1])

Emotion	No. of passages	No. of sentences	No. of word tokens	No. of punctuation marks
Joy	16	120	966	181
Sadness	15	120	1053	187
Anger	18	123	1124	201
TOTAL	49	363	3143	569

The pauses were measured and analysed using the Praat program (Boersma & Weenink, 2016). For each speech passage, I fixed the number, location and nature of pauses, whereupon I grouped the pauses by nature and location. I also calculated the duration statistics of the pause groups in milliseconds (the minimum value, the first quartile, mean, median, the third quartile and the maximum value) and found out which pause groups are statistically significant in telling apart emotions by pause duration.

In order to find out whether the emotion of a sentence can be identified based on mere pause differences, I picked a four-sentence neutral Corpus passage<sup>13</sup> and, using the program Sound Forge Pro 9,<sup>14</sup> modified the duration, location and nature of the passage pauses, approximating the parameters of emotional speech obtained during my pause analysis. This resulted in eight passages with different pause patterns (2 x neutral, anger, sadness and joy). Four of those eight had a simple pause pattern, where only sentence-end pauses had been modified. The other four had a more complex structure, as modification had been applied to pause duration, nature and location. I left neutral passages untouched.

Then I conducted a perception test, asking ten subjects to listen<sup>15</sup> to the eight passages differing in pause patterns and to determine from sound alone what emotion they perceived, either joy, sadness, anger or no particular emotion (neutral). The group of evaluators consisted of two men and eight women, aged 30 to 70 years. All evaluators were ethnic Estonians with Estonian as their mother tongue. The perception tests were conducted in August 2009.

---

<sup>13</sup> “About three centimetres is poured out. Just a little bit more than is offered to taste by a respectable sommelier. Seventy kroons, please. Thank you.”

<sup>14</sup> See <http://www.sonycreativesoftware.com/> (last checked 2017, August 21).

<sup>15</sup> Instruction to the Listening test: Please listen to the following eight passages and mark, for each passage, what emotion it conveys in your opinion. There are no right or wrong answers for this test. On the contrary, your answer will help us find how well an emotion can be recognised in this or that passage. You need not complete the whole test at once, instead, you can take a pause, saving the interim results for later resume (continue and/or change).

### 3.1.2. Material and method used in studying formant frequencies

Formant frequencies of emotional speech are the focus of article [P2]. The study consists of two parts: (1) the influence of sentence emotion on the first and second formant frequencies of short vowels and (2) the influence of sentence emotion on the speaker's articulatory precision. The aim of the study was to find out if emotions could influence the first and second formant frequencies in vowels and the speaker's articulatory precision, and if the differences in the values of those parameters are distinctive of three emotions and neutral speech.

For formant analysis, I decided to use the Estonian vowels *a*, *i* and *u*, which were picked from sentences of joy, sadness and anger, as well as from neutral speech, 2,395 vowel tokens<sup>16</sup> in total (see Table 2). Those three vowels can be seen as the corner vowels of a triangular vowel space (*a* being a low-back vowel, *i* a high-front vowel and *u* a high-back vowel), and they enable a valid description of the precision of a speaker's articulation. Since it is known that emotion is carried by stressed syllables (see Seppi, Batliner, Steidl, Schuller & Nöth, 2010) and the sounds of non-stressed syllables tend to lose in quality (see, e.g., Moon & Lindblom, 1994), I picked the vowels to be analysed from word-initial stressed positions, such as VC, CVC or CCVC<sup>17</sup>.

First, I measured the first and second formant frequencies of the short stressed vowels *a*, *i*, and *u*. The measuring point was the vowel centre.

**Table 2.** Material used for formant analysis (see [P2])

Vowel	Emotion	No. of vowels
<i>a</i>	Joy	279
	Sadness	244
	Anger	395
	Neutral	319
	TOTAL	1237
<i>i</i>	Joy	146
	Sadness	154
	Anger	220
	Neutral	145
	TOTAL	665
<i>u</i>	Joy	133
	Sadness	115
	Anger	138
	Neutral	107
	TOTAL	493
<i>a, i, u</i>	TOTAL	2395

<sup>16</sup> The material used to study the formant frequencies is presented in Appendix 2.

<sup>17</sup> V – vowel; C – consonant.

The formant frequencies were measured using the Praat program (Boersma & Weenink, 2016) and the EMU Speech Database System. I manually corrected the errors in the automatic calculations of the formant frequency values. Then I calculated the mean values of the first and second formant frequencies of each of the three vowels for each emotion and for neutral speech. After that, I used the ANOVA to determine which differences between the formant frequencies values are statistically significant to distinguish between emotion groups (joy, sadness, anger, neutral).

The next step was to measure the speaker's articulatory precision by Euclidean distance. For that purpose, I first found the centre of the speaker's vowel space, that is, her neutral sound  $x$ . After that, I computed the Euclidean distance between the means of the vowels  $a$ ,  $i$  and  $u$  as used in her emotional and neutral speech and her neutral sound  $x$ , using the formula  $d(x, V_{em}) = \sqrt{(F_1x - F_1V_{em})^2 + (F_2x - F_2V_{em})^2}$ , where  $x$  is the neutral vowel and  $V_{em}$  is the vowel mean of the emotion under scrutiny.

### 3.1.3. Material and method used in studying speech rate

My study of the influence of emotions on speech rate is reported in article [P3]. The study consists of two parts, the first of which is focused on speech rate. In particular, I investigate whether emotions may influence a reader's speech rate and whether differences in speech rate values are significant in distinguishing emotions from each other and from neutral speech.

The material of my choice for studying speech rate consisted of 314 sentences<sup>18</sup> at least three words long, some of which represented joy, some sadness, some anger, and some were neutral (see Table 3).

**Table 3.** Material for studying the speech rate (see [P3])

Emotion	No. of sentences
Joy	55
Sadness	84
Anger	77
Neutral	98
TOTAL	314

First, I used the Praat program (Boersma & Weenink, 2016) to measure the speech rate in emotional and in neutral speech.<sup>19</sup> As the Corpus sentences were

<sup>18</sup> The material used to study speech rate is presented in Appendix 2.

<sup>19</sup> There are two approaches to speech rate measuring: (a) with pauses and/or (b) without pauses (Braun & Oba, 2007). For the present study, pauses have been omitted from the material while measuring the speaker's speed of articulation.

all different, I decided that sounds per second would make a good measure of speech rate. Long phonemes were counted double, assuming that phonologically long speech sounds are actually sequences of two short phonemes (Eek, 2008).

Then I computed the mean speech rate for each of the three emotions and for neutral speech and carried out a pairwise comparison. In addition, I tested whether speech rate varied within a sentence. For that purpose, I took separate measurements of the speech rate of words within the phrase<sup>20</sup> and of the phrase-final word.

Using the ANOVA, I determined whether the pairwise differences in speech rates were significant enough statistically to distinguish the three emotions from each other and from neutral speech.

### 3.1.4. Material and method used in studying intensity and fundamental frequency

The intensity of emotional speech is the focus of article [P4]; fundamental frequency is addressed in article [P5]. The former tries to find out if emotions influence the level and range of the intensity of a speaker's voice and if the differences in those parameter values are significant enough to distinguish emotions from each other and from neutral speech. The latter study tries to ascertain if emotions have any influence on the level and range of the fundamental frequency of a speaker's voice and if the differences between the values of those parameters are significant enough to distinguish emotions from each other and from neutral speech.

In order to gauge the values of the level and range of intensity and to measure the fundamental frequency level and range, I picked 329 Corpus sentences<sup>21</sup>, which contained 6,409 vowels in total (see Table 4).

**Table 4.** Material for studying the speech intensity and fundamental frequency (see [P4, P5])

Emotion	No. of sentences	No. of vowels in sentences
Joy	60	973
Sadness	87	1807
Anger	79	1435
Neutral	103	2194
TOTAL	329	6409

<sup>20</sup> In terms of the present study, a phrase means a prosodic unit lasting from pause to pause, while a pause means a period of silence between words or sentences, lasting at least 30 ms.

<sup>21</sup> The material used to study the levels of intensity and fundamental frequency is presented in Appendix 2.

Then I measured the intensity and  $F_0$  of the emotion groups (joy, sadness, anger and neutral speech) at the midpoint of all vowels of the 329 sentences. After that, I computed, for each emotion group, the minimum value, first quartile, median, third quartile and maximum value of the fundamental frequency level and its range as well as of the intensity level and its range.

To gauge the level of intensity and  $F_0$  for sentence beginning and sentence end, I measured those parameters at the first stressed vowel of the first and last words of every emotional and neutral sentence. Based on those measurements, I computed, for each emotion group, the minimum value, the first quartile, median, third quartile and maximum value of the level of intensity and  $F_0$ .

The measurements of intensity and fundamental frequency were made using the Praat program (Boersma & Weenink, 2016) and the EMU Speech Database System. The Wilcoxon rank-sum test helped me to decide whether the differences in intensity and  $F_0$  level values were significant enough statistically to distinguish emotions from each other and from neutral speech.

### **3.2. Material and method used to create acoustic models of synthetic emotional speech**

The experiment to create synthetic emotional speech is described in article [P6]. The article consists of two parts: (1) a theoretical part introducing the principles of emotion modelling and (2) an experimental part aimed at creating some emotion models.

Together with my co-author (see [P6]), we started on emotion models for synthetic speech relying on the recent results on the acoustic expression in human emotional Estonian (see 4.1.). In order to develop the best acoustic models for expressing joy, sadness and anger in Estonian parametric speech synthesis, for male as well as female voices,<sup>22</sup> we first compiled three test models (considering the options of parametric tuning of the concrete synthesiser) for each basic emotion. The test models used synthetic voices created with the help of the HTS method and trained on a corpus of neutral speech. The parameter values of the synthetic neutral speech were controlled by the model used by a statistical parametric synthesiser (see Mihkla, Hein, Kalvik, Kiissel, Sirts & Tamuri, 2012) and did not depend on my own results of neutral speech.

The test models<sup>23</sup> operated with four parameters, namely, the  $F_0$  level,  $F_0$  range and intensity level of the voice, and the speech rate. Every test model contained a combination of parametric values for joy, sadness and anger. In

---

<sup>22</sup> Both synthetic voices were HTS voices, based on hidden Markov models and trained on voluminous speech material (read text). The male training corpus contained about 2,300 sentences read by a professional radio and TV announcer; the female training corpus consisted of about 2,000 sentences read by an actress.

<sup>23</sup> Recorded examples of emotional speech synthesised using the test models can be accessed at [https://www.eki.ee/heli/index.php?option=com\\_content&view=article&id=7&Itemid=494](https://www.eki.ee/heli/index.php?option=com_content&view=article&id=7&Itemid=494) (last checked 2017, August 21).



Model 2 (M2) the values were on the optimal level (based on my study results); in Model 1 (M1), the M2 values were lowered by approximately 15% from the direction characterising the emotion (towards neutrality), whereas in Model 3 (M3) the M2 values were raised by 15% in the emotion’s characteristic direction (away from neutrality) (see Table 5).

**Table 5.** Parametric models of Estonian emotional synthetic speech (see [P6])

Parameter	Neutral	Joy	Sadness	Anger
<i>MODEL 1 (M1)</i>				
Speech rate		1.10	0.90	1.24
Intensity level		0.90	0.85	0.94
F <sub>0</sub> level		1.50	-3.00	-4.00
F <sub>0</sub> range		1.50	-0.90	2.10
<i>MODEL 2 (M2)</i>				
Speech rate	1.00	1.15	0.80	1.40
Intensity level	1.00	0.85	0.70	0.90
F <sub>0</sub> level	0.00	2.50	-4.00	-5.00
F <sub>0</sub> range	0.00	2.50	-1.15	2.60
<i>MODEL 3 (M3)</i>				
Speech rate		1.20	0.70	1.56
Intensity level		0.80	0.55	0.86
F <sub>0</sub> level		3.50	-5.00	-6.00
F <sub>0</sub> range		3.60	-1.40	3.10

Table 5 presents the values of emotion parameters relative to parameters of neutral speech. As the acoustic analysis of Estonian emotional human speech was only based on a female voice, the human results had to be transformed for the test models.

The following two hypotheses were set for the test models – H1: The test persons will do best in the identification of emotions synthesised by Model 3, in both male and female voices; and H2: The test persons will recognise the neutral speech produced by means of the acoustic model used in the synthesiser.

In order to test the hypotheses and evaluate the test models, we compiled four perception tests, two for each synthetic voice. The first, Test A, consisted of ten three-sentence synthetic passages<sup>24</sup> of neutral content, whose acoustic parameters had been modified according to an emotion (joy, sadness or anger) and a model (M1, M2 or M3). One passage out of the ten was left neutral, acoustically (the parameter values were provided default by the synthesiser<sup>25</sup>).

<sup>24</sup> A passage of the Test A: “In the evening we went to a restaurant. We made an order and waited. When the food was brought and we saw it, we went all speechless.”

<sup>25</sup> The mean values of neutral voices: *male voice* F<sub>0</sub> level 119 Hz, F<sub>0</sub> range 78–158 Hz (11.9 semitones), intensity level 74.5 dB, and speech rate 162 words per minute; *female voice* F<sub>0</sub> level 177 Hz, F<sub>0</sub> range 122–237 Hz (11.4 semitones), intensity level 73.6 dB, and speech rate 148 words per minute.

The second, Test B, also consisted of ten three-sentence synthetic passages<sup>26</sup> of neutral content, but here the passages began with an acoustically neutral sentence and changed, if at all, to joyous, sad or angry from the second sentence on. In nine passages, the values of the acoustic parameters were modified according to an emotion (joy, sadness or anger) and to a model (M1, M2 or M3). One passage out of the ten was left neutral, acoustically (the parameter values were provided default by the synthesiser).

The aim of the compilation of Test A and Test B was to establish whether a change of emotion within a passage could be helpful in identifying the emotion.

I carried out the web-based perception tests in the environment of the Estonian Emotional Speech Corpus (cf. Altrov & Pajupuu, 2012). The evaluators of the test models were asked to listen to 4 x 10 passages of synthetic speech – 2 x male (Tests A and B) and 2 x female voices (Tests A and B) – and to determine the emotion or neutrality for each passage. Every test was introduced by instructions on how to perform the test. In Test A, the testee was to listen to the passage and answer the question “What is the emotion of the passage?”, while in Test B, the question was “What is the emotion of the final part of the passage?” The multiple-answers list provided “joy”, “sadness”, “anger” and “neutral”. The listeners could run each passage as many times as they wished. If necessary, the listeners could change their earlier answers or discontinue taking the test to re-enter it later.

The test models were evaluated by ten male and ten female persons, aged 30 to 73 (average 43.3) years. All evaluators were ethnic Estonians whose mother tongue was Estonian. I personally carried out the perception tests in September 2015.

In order to establish what acoustic features were dominant in the emotion models and what their role was in the identification of the target emotion, my co-author, M. Mihkla, subjected the results of the perception tests to binary logistic regression analysis. The dependent variable of the binary logistic regression was the (non-)identification of the emotion or of the neutral speech (true vs false), while the argument features were the values of the speech rate, voice intensity level,  $F_0$  level and  $F_0$  range parameters.

---

<sup>26</sup> A passage of the Test B: “Wednesday night I got a phone call from someone. I could never have guessed what he had to say. Things can take a different turn altogether.”

## 4. RESULTS

The chapter consists of two major parts: (1) results on the acoustic parameters of emotional speech (see 4.1) and (2) results of an experiment to create acoustic models for emotional synthetic speech (see 4.2).

### 4.1. Results on the acoustic parameters of emotional speech

#### 4.1.1. Results on pauses

The results on pauses are presented in article [P1].

*Number and location of pauses.* In the research material, for all three emotions, the number of pauses mainly depended on the number of punctuation marks where most frequently the reader would pause (see Table 6). Extra pauses<sup>27</sup> made up 8–10% of all pauses, with slight variation across the emotions. For all three emotions, in about half of the cases, the extra pauses immediately preceded the conjunctions *ja* or *ning* (Est 'and'), or proper (person or place) names. The rest of the extra pauses were occasional, possibly depending either on text contents or on the reader's personal characteristics (e.g., her lung volume).

**Table 6.** Number and location of pauses (see [P1])

Emotion	No. of pauses	Pauses coinciding with punctuation marks (% of all pauses)	Extra pauses (% of all pauses)
Joy	181	92	8
Sadness	165	92	8
Anger	195	90	10

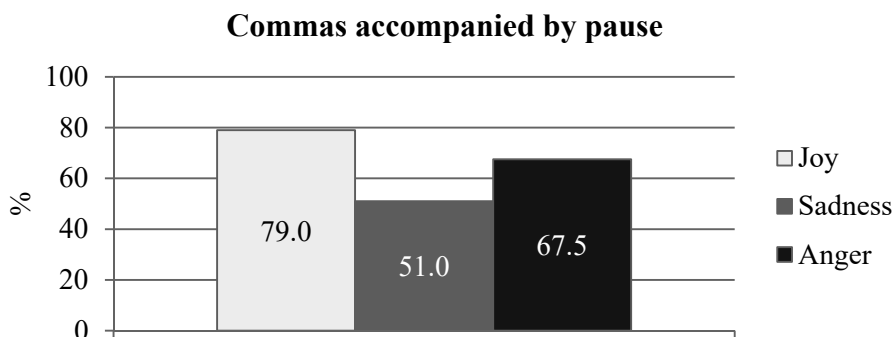
For all three emotions, the *comma* was the only punctuation mark at which the reader sometimes missed the pause. In sadness sentences, there was a comma pause in 51% of the commas, while for anger, the proportion was 67.5%, and for joy it was 79% (see Table 7 and Figure 1).

---

<sup>27</sup> An extra pause is a pause not coinciding with any punctuation mark.

**Table 7.** Commas (see [P1])

Emotion	No. of commas	Commas accompanied by pause (% of all commas)
Joy	67	79.0
Sadness	72	51.0
Anger	80	67.5

**Figure 1.** Commas in emotional speech, where the reader made a pause (% of all commas)

*Nature of pauses (breathing vs. non-breathing pauses).* All emotional, non-neutral passages had more breathing pauses than non-breathing ones (see Table 8). The greatest number of pauses occurred in anger passages, slightly surpassing that of sadness or joy. The number of breathing pauses of punctuation were more or less similar for joy, sadness and anger passages. A significantly distinctive difference occurred in extra pauses with breathing, which were much more frequent in the case of anger than for sadness or joy. Moreover, here the difference between joy and sadness was also notable.

**Table 8.** Breathing pauses (see [P1])

Emotion	No. of breathing pauses	Breathing pauses (% of all pauses)	Breathing pauses of punctuation (% of all punctuation pauses)	Extra pauses with breathing (% of all extra pauses)
Joy	96	53.0	56.0	21.0
Sadness	90	54.5	57.0	31.0
Anger	115	59.0	61.0	55.0

*Correlation between punctuation and breathing pauses.* Punctuation-coordinated end-of-sentence breathing pauses occurred more frequently in sadness and joy passages (68% and 67%, respectively) and less in anger passages (58%) (see Table 9). Commas were found to coincide with a breathing pause most frequently in anger passages (67%) and least frequently in sadness passages (27%). The rest of the punctuation marks (i.e., apart from sentence end and comma) coincided with breathing pauses mostly in sadness passages and the least in joy passages.

**Table 9.** Proportion of the breathing pauses coinciding with punctuation marks (see [P1])

Emotion	End of sentence (% of end-of-sentence pauses)	Comma (% of comma pauses)	Other punctuation marks (% of other punctuation pauses)
Joy	67	36	40
Sadness	68	27	60
Anger	58	67	56

*Duration of pauses.* Measurements of pause duration revealed that, across pause groups, the longest pauses were typical of sadness and the shortest of anger (see Table 11 below). According to average durations<sup>28</sup>, all three emotions had their longest pauses at the full stop (for sadness and joy those were pauses with breathing, for anger without breathing), whereas the non-breathing extra and comma pauses were the shortest of all.

As it turned out, non-breathing as well as breathing full-stop pauses and breathing exclamation mark pauses are statistically significant<sup>29</sup> in distinguishing emotions:

- Non-breathing full-stop pause
  1. sadness (944 ms) vs joy (638 ms) ( $p = 0.010$ )
  2. sadness (944 ms) vs anger (691 ms) ( $p = 0.006$ )
- Breathing full-stop pause
  1. sadness (969 ms) vs joy (774 ms) ( $p = 0.001$ )
  2. sadness (969 ms) vs anger (664 ms) ( $p = 0.001$ )
  3. joy (774 ms) vs anger (664 ms) ( $p = 0.003$ )
- Exclamation mark pauses with breathing
  1. sadness (911 ms) vs anger (625 ms) ( $p = 0.017$ )

The duration of the rest of the pause groups did not differ significantly across emotions.

<sup>28</sup> Without those pause groups that were represented fewer than three times.

<sup>29</sup> The Student's t-Test was used. The null hypothesis was that the difference between the means of two emotions is zero ( $p > 0.05$ ).

*Results of the perception test.* As revealed by the perception test, the emotion of a passage was not identified from mere pause differences (see Table 10). The passages with a simpler pause pattern (modification concerned end-of-sentence pauses only) were evaluated as neutral, and so were the passages with a more complex pause pattern, where modifications had been applied to pause duration, location and nature.

**Table 10.** The results of the perception test (rate of identification) (see [P1])

Passages with a simple pause pattern				
<i>Passage emotion</i>	<i>Multiple choice</i>			
	Joy	Sadness	Anger	Neutral
Joy	0	0	10	90
Sadness	0	10	20	70
Anger	10	10	10	70
Neutral	0	10	20	70
Passages with a complex pause pattern				
<i>Passage emotion</i>	<i>Multiple choice</i>			
	Joy	Sadness	Anger	Neutral
Joy	10	10	10	70
Sadness	10	0	20	70
Anger	10	0	10	80
Neutral	10	0	10	80

**Table 11.** The average durations of pause groups in milliseconds (see [P1])

Type of pause	Emotion	No.	Min	Q1	Median	Mean	Q3	Max
<i>Extra pause</i>	Joy	11	34	103	117	128	127	359
	Sadness	9	54	91	111	122	133	251
	Anger	9	91	119	133	141	145	267
<i>Extra pause with breathing</i>	Joy	3	365	416	467	448	490	513
	Sadness	4	300	366	436	417	488	497
	Anger	11	289	326	399	386	412	555
<i>Comma pause</i>	Joy	34	52	137	222	252	339	636
	Sadness	28	43	113	203	280	379	795
	Anger	21	75	145	207	225	299	428
<i>Comma pause with breathing</i>	Joy	19	280	407	440	480	538	724
	Sadness	9	296	394	531	511	615	721
	Anger	33	205	337	408	412	449	732
<i>Colon pause</i>	Joy	2	139	181	222	222	264	306
	Sadness	0	–	–	–	–	–	–
	Anger	1	364	364	364	364	364	364
<i>Colon pause with breathing</i>	Joy	1	438	438	438	438	438	438
	Sadness	1	777	777	777	777	777	777
	Anger	2	296	366	437	437	507	578

Type of pause	Emotion	No.	Min	Q1	Median	Mean	Q3	Max
<i>Dash pause</i>	Joy	4	393	526	606	628	708	906
	Sadness	2	837	838	839	839	841	842
	Anger	5	236	292	525	441	545	606
<i>Dash pause with breathing</i>	Joy	3	373	537	702	604	719	736
	Sadness	5	533	987	989	936	1033	1138
	Anger	8	316	375	448	526	693	865
<i>Ellipsis pause</i>	Joy	1	724	724	724	724	724	724
	Sadness	5	407	412	786	742	972	1133
	Anger	2	878	878	879	879	879	880
<i>Ellipsis pause with breathing</i>	Joy	1	474	474	474	474	474	474
	Sadness	8	578	883	934	912	987	1085
	Anger	0	–	–	–	–	–	–
<i>Full-stop pause</i>	Joy	25	257	524	594	638	710	1137
	Sadness	30	211	638	913	944	1156	2136
	Anger	34	426	534	689	691	805	1006
<i>Full-stop pause with breathing</i>	Joy	53	413	637	772	774	882	1364
	Sadness	54	528	757	884	969	1076	2366
	Anger	43	353	542	667	664	746	1016
<i>Exclamation mark pause</i>	Joy	7	256	445	540	571	708	893
	Sadness	0	–	–	–	–	–	–
	Anger	6	384	626	635	624	655	802
<i>Exclamation mark pause with breathing</i>	Joy	16	383	526	705	696	843	996
	Sadness	6	709	813	835	911	955	1285
	Anger	13	341	600	644	625	706	719
<i>Question mark Pause</i>	Joy	1	687	687	687	687	687	687
	Sadness	1	941	941	941	941	941	941
	Anger	1	532	532	532	532	532	532
<i>Question mark pause with breathing</i>	Joy	0	–	–	–	–	–	–
	Sadness	2	479	602	724	724	847	970
	Anger	1	623	623	623	623	623	623
<i>Question exclamation mark pause</i>	Joy	0	–	–	–	–	–	–
	Sadness	0	–	–	–	–	–	–
	Anger	1	652	652	652	652	652	652
<i>Question exclamation mark pause with breathing</i>	Joy	0	–	–	–	–	–	–
	Sadness	1	1260	1260	1260	1260	1260	1260
	Anger	4	608	619	638	653	671	727

### 4.1.2. Results on formant frequencies

The results of the study of vowel formant frequencies are presented in article [P2].

*The influence of emotions on the values of the first formant frequency ( $F_1$ ) of the vowels a, i and u.* The vowel *a* showed a significant difference between neutral speech and the emotions of sadness and anger: for both sadness and anger, the average values of the  $F_1$  of *a* were lower than for neutral speech. As revealed by pairwise analysis, all emotions differed significantly from each other on average  $F_1$  values (see Tables 12 and 13). In the case of the vowel *i*, all three emotions also differed significantly from neutral speech, while for joy, sadness as well as anger sentences, the average values of the  $F_1$  of *i* were lower than for neutral speech. Pairwise, the vowel *i* made a significant difference between joy and sadness (the  $F_1$  of joy was higher) and between anger and sadness (the  $F_1$  of anger was higher). As for the vowel *u*, its average values of  $F_1$  did not differ significantly across the four emotion groups analysed (joy, sadness, anger, neutral).

*The influence of emotions on the values of the second formant frequency ( $F_2$ ) of the vowels a, i and u.* For all three emotions as well as for neutral speech, the differences between the  $F_2$  average values of the three vowels analysed were small (dwindling to none in the case of *i*) (see Table 12), let alone statistical significance (see Table 14). In pairwise analysis, too, the differences in average  $F_2$  values were small and statistically insignificant. The only exception was the vowel *a*, in which the  $F_2$  for anger differed significantly from that for sadness on the Hertz scale (sadness had a higher  $F_2$ ).

**Table 12.** The average values of the first and second formant frequencies of the vowels *a*, *i* and *u*, measured in Hertz/Bark for four emotion groups (see [P2])

Vowel	Emotion	$F_1$	$F_2$
<i>a</i>	Joy	644 / 6.1	1440 / 10.8
	Sadness	558 / 5.4	1460 / 10.9
	Anger	615 / 5.9	1412 / 10.7
	Neutral	635 / 6.0	1448 / 10.9
<i>i</i>	Joy	382 / 3.8	2658 / 14.9
	Sadness	359 / 3.6	2642 / 14.9
	Anger	386 / 3.9	2642 / 14.9
	Neutral	403 / 4.0	2660 / 14.9
<i>u</i>	Joy	418 / 4.2	1272 / 10.0
	Sadness	410 / 4.1	1334 / 10.3
	Anger	427 / 4.3	1298 / 10.1
	Neutral	430 / 4.3	1274 / 10.0



**Table 13.** The results of ANOVA on the  $F_1$  of the short vowels *a*, *i* and *u* in emotional and neutral sentences (see [P2])

Vowel	Emotion pair	Pr > $F$ Hertz scale	Pr > $F$ Bark scale
<i>a</i>	anger vs joy	0.010**	0.030*
	anger vs neutral	0.050*	0.040*
	anger vs sadness	0.001***	0.010**
	joy vs neutral	0.430	0.690
	joy vs sadness	0.001***	0.010**
	neutral vs sadness	0.001***	0.010**
<i>i</i>	anger vs joy	0.500	0.440
	anger vs neutral	0.010**	0.020*
	anger vs sadness	0.001***	0.010**
	joy vs neutral	0.010**	0.010**
	joy vs sadness	0.001***	0.010**
	neutral vs sadness	0.001***	0.010**
<i>u</i>	anger vs joy	0.500	0.390
	anger vs neutral	0.740	0.780
	anger vs sadness	0.150	0.140
	joy vs neutral	0.380	0.310
	joy vs sadness	0.540	0.600
	neutral vs sadness	0.110	0.110

*Note.* Statistically significant differences are marked by asterisks: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

**Table 14.** The results of ANOVA on the  $F_2$  of the short vowels *a*, *i* and *u* in emotional and neutral sentences (see [P2])

Vowel	Emotion pair	Pr > $F$ Hertz scale	Pr > $F$ Bark scale
<i>a</i>	anger vs joy	0.189	0.280
	anger vs neutral	0.055	0.050
	anger vs sadness	0.036*	0.070
	joy vs neutral	0.720	0.490
	joy vs sadness	0.460	0.500
	neutral vs sadness	0.618	0.930
<i>i</i>	anger vs joy	0.793	0.840
	anger vs neutral	0.667	0.630
	anger vs sadness	0.661	0.400
	joy vs neutral	0.895	0.810
	joy vs sadness	0.575	0.400
	neutral vs sadness	0.490	0.300
<i>u</i>	anger vs joy	0.402	0.380
	anger vs neutral	0.476	0.400
	anger vs sadness	0.286	0.320
	joy vs neutral	0.962	0.960
	joy vs sadness	0.072	0.070
	neutral vs sadness	0.112	0.090

*Note.* Statistically significant differences are marked by asterisks: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

*The influence of emotions on the speaker's articulatory precision.* The results revealed that, with a single exception (see below), the articulatory precision of the vowels *a*, *i* and *u* did not significantly differ for emotional vs neutral speech (see Table 15). Nor were the differences remarkable in pairwise analysis. The only exception was the *a* of sadness, which was considerably closer to the central vowel *x* than in the case of the other emotions or neutral speech; in other words, for sadness, the articulatory precision of *a* changed, and the vowel lost in quality.

**Table 15.** Euclidean distance between vowels and the central sound *x* (Hertz/Bark) (see [P2])

Vowel	<i>x</i> – joy	<i>x</i> – sadness	<i>x</i> – anger	<i>x</i> – neutral
<i>a</i>	332.7 / 1.5	288.8 / 0.9	348.9 / 1.4	321.8 / 1.4
<i>i</i>	922.1 / 3.4	909.8 / 3.4	915.6 / 3.3	921.3 / 3.3
<i>u</i>	482.1 / 1.8	423.2 / 1.6	454.8 / 1.7	477.9 / 1.8

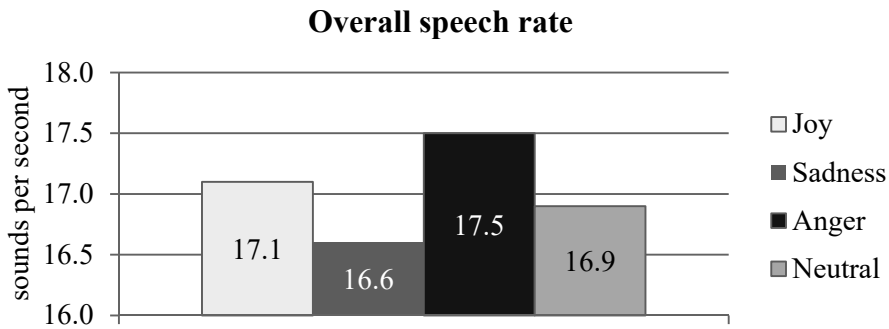
### 4.1.3. Results on speech rate

The results on speech rate are presented in article [P3].

As for the *overall average speech rate*, it was the fastest in anger sentences (17.5 s/s – sounds per second) and the slowest in sadness sentences (16.6 s/s) (see Table 16 and Figure 2). The same holds for the *average intra-phrase speech rate* (anger 18.4 s/s; sadness 17 s/s). *The last word of a phrase* was articulated the fastest in sentences of anger and of sadness (14.4 s/s) and the slowest in sentences of joy (14.0 s/s).

**Table 16.** The average speech rates for emotional and neutral speech (sounds per second) (see [P3])

Emotion	Overall speech rate	Intra-phrase speech rate	Phrase-final speech rate
Joy	17.1	17.6	14.0
Sadness	16.6	17.0	14.4
Anger	17.5	18.4	14.4
Neutral	16.9	17.5	14.1



**Figure 2.** The overall speech rate in emotional and neutral speech

The overall average rate of emotional speech was significantly different from that of neutral speech only in anger sentences (see Table 17). Pairwise, however, all the differences were statistically significant. In the case of the average intra-phrase rate, the emotions of anger and sadness differed significantly from neutral speech. Pairwise, intra-phrase differences were also significant. The average tempo differences between emotional and neutral speech in the phrase-final word were not statistically significant; the same holds for emotion pairs.

**Table 17.** The results of ANOVA on average speech rate differences in emotion pairs (see [P3])

Emotion pair	<i>P</i> -value		
	<i>overall speech rate</i>	<i>intra-phrase speech rate</i>	<i>phrase-final word speech rate</i>
Joy vs anger	0.010**	0.043*	0.452
Joy vs sadness	0.039*	0.036*	0.317
Joy vs neutral	0.557	0.810	0.985
Sadness vs anger	0.001***	0.001***	0.837
Sadness vs neutral	0.107	0.033*	0.237
Anger vs neutral	0.031*	0.008**	0.370

*Note.* Statistically significant differences are marked by asterisks: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

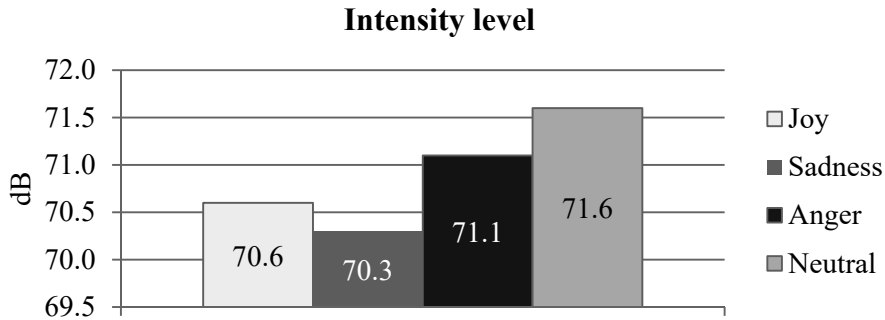
#### 4.1.4. Results on intensity

The results on speech intensity are presented in article [P4].

*Level of intensity in emotional and neutral speech.* According to the results, the intensity level was the highest in neutral speech (median 71.6 dB) and the lowest in sadness sentences (median 70.3 dB) (see Table 18 and Figure 3).

**Table 18.** The level of intensity (dB) in emotional and neutral speech (see [P4])

Emotion	Minimum	Q1	Median	Q3	Maximum
Joy	60.0	67.7	70.6	73.1	81.1
Sadness	57.9	66.9	70.3	73.1	82.2
Anger	58.9	67.9	71.1	74.0	82.9
Neutral	61.9	69.2	71.6	74.1	81.3



**Figure 3.** The level of intensity in emotional and neutral speech (medians)

According to a statistical analysis of the results, all differences in the level of intensity were significant, between emotions as well as compared to neutral speech (see Table 19).

**Table 19.** The results of the Wilcoxon rank-sum test on the level of intensity across emotion groups (see [P4])

Emotion	<i>P</i> -value
Joy vs anger	0.008**
Joy vs sadness	0.020*
Joy vs neutral	0.001***
Sadness vs anger	0.001***
Sadness vs neutral	0.001***
Anger vs neutral	0.001***

*Note.* Statistically significant differences are marked by asterisks: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

*Range of intensity in emotional and neutral speech.* The results revealed that the widest intensity range belonged to sadness (median 14.7 dB), while the narrowest was observed in the case of the joy emotion (median 13.2 dB) (see Table 20).

**Table 20.** The range of intensity (dB) for emotional and neutral speech (see [P4])

Emotion	Minimum	Q1	Median	Q3	Maximum
Joy	5.6	11.0	13.2	17.2	25.9
Sadness	4.2	11.7	14.7	19.7	30.8
Anger	6.6	12.5	14.3	18.5	25.8
Neutral	6.5	12.2	13.7	17.6	24.9

Statistical analysis showed that there was no significant difference in the intensity range, neither between the emotions nor in comparison with neutral speech (see Table 21).

**Table 21.** The results of the Wilcoxon rank-sum test for the intensity range in emotion pairs and in comparison with neutral speech (see [P4])

Emotion pair	<i>P</i> -value
Joy vs anger	0.280
Joy vs sadness	0.250
Joy vs neutral	0.800
Sadness vs anger	0.930
Sadness vs neutral	0.800
Anger vs neutral.	0.800

*Note.* Statistically significant differences are marked by asterisks: \**p* < .05, \*\**p* < .01, \*\*\**p* < .001.

*Level of intensity at sentence beginning and sentence end in emotional and neutral speech.* The results showed that at both positions, that is, at the beginning as well as at the end of the sentence, the intensity level was the highest for neutral speech (median 74.7/65.3 dB) and the lowest for sadness (median 72.3/63.6 dB) (see Table 22).

**Table 22.** The level of intensity (dB) at sentence beginning and sentence end in emotional and neutral speech (see [P4])

Emotion	Sentence beginning / Sentence end				
	Minimum	Q1	Median	Q3	Maximum
Joy	68.2 / 59.1	71.6 / 63.0	73.9 / 64.5	75.8 / 66.3	81.5 / 69.1
Sadness	64.3 / 54.7	70.0 / 61.1	72.3 / 63.6	74.9 / 65.5	81.8 / 71.0
Anger	65.6 / 58.2	70.8 / 62.6	73.8 / 64.9	76.7 / 67.1	82.7 / 73.7
Neutral	65.3 / 56.6	72.0 / 62.9	74.7 / 65.3	76.4 / 67.5	81.1 / 73.2

However, the difference in the intensity level was statistically significant, at both sentence beginning and sentence end, only for the pair of sadness vs neutral speech (see Table 23), not for any other pair.

**Table 23.** The results of the Wilcoxon rank-sum test on the level of intensity at sentence beginning and sentence end in emotional and neutral speech (see [P4])

Emotion pair	Sentence beginning ( <i>P</i> -value)	Sentence end ( <i>P</i> -value)
Joy vs anger	0.917	0.932
Joy vs sadness	0.091	0.112
Joy vs neutral	0.532	0.512
Sadness vs anger	0.053	0.053
Sadness vs neutral	0.002**	0.002**
Anger vs neutral	0.917	0.932

*Note.* Statistically significant differences are marked with asterisks: \**p* < .05, \*\**p* < .01, \*\*\**p* < .001.

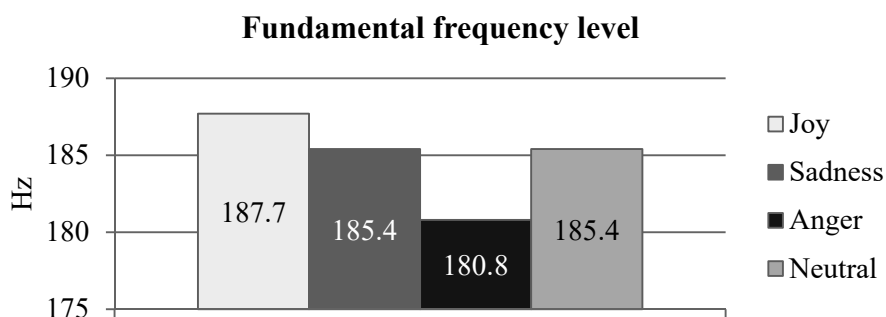
#### 4.1.5. Results on fundamental frequency

The results on fundamental frequency ( $F_0$ ) are presented in the article [P5].

*Fundamental frequency level in emotional and neutral speech.* According to the results, the  $F_0$  level was the highest in joy sentences (median 187.7 Hz) and the lowest in anger sentences (median 180.8 Hz), while sadness sentences and neutral speech shared a mid-level  $F_0$  (median 185.4 Hz) (see Table 24 and Figure 4).

**Table 24.** The fundamental frequency level (Hz) in emotional and neutral speech (see [P5])

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



**Figure 4.** The level of fundamental frequency in emotional and neutral speech (medians)

A statistical analysis revealed that the  $F_0$  level's differences were only significant for anger, where the  $F_0$  level was significantly lower than for the other emotions or for neutral speech (see Table 25). The differences between the rest of the emotions as well as between the emotions and neutral speech were not significant.

**Table 25.** The results of the Wilcoxon rank-sum test on the  $F_0$  level in emotion pairs and in comparison with neutral speech (see [P5])

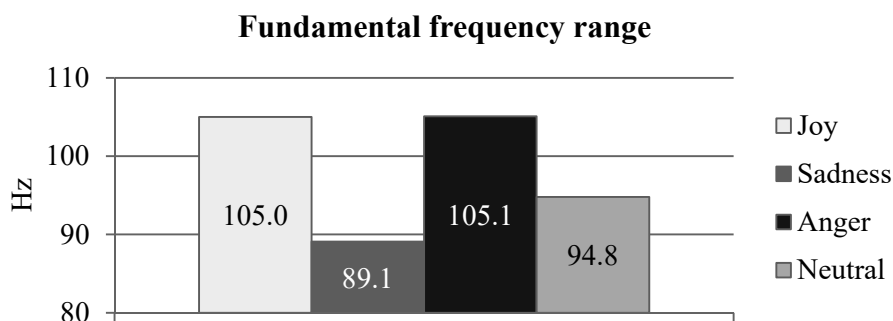
Emotion pair	<i>P</i> -value
Joy vs anger	0.001***
Joy vs sadness	0.240
Joy vs neutral	0.350
Sadness vs anger	0.001***
Sadness vs neutral	0.610
Anger vs neutral.	0.001***

*Note.* Statistically significant differences are marked by asterisks: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

*F<sub>0</sub> range in emotional and neutral speech.* The  $F_0$  range turned out to be the widest in anger sentences (median 105.1 Hz) and the narrowest in sadness sentences (median 89.1 Hz) (see Table 26 and Figure 5).

**Table 26.** The range of fundamental frequency (Hz) in emotional and neutral speech (see [P5])

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



**Figure 5.** The range of fundamental frequency (medians) in emotional and neutral speech

According to a statistical analysis, the  $F_0$  range of the sadness emotion significantly differed from that of the emotions of joy and anger, and the  $F_0$  range for anger significantly differed from that of neutral speech (see Table 27).

**Table 27.** The results of the Wilcoxon rank-sum test on the  $F_0$  range between emotions and in comparison with neutral speech (see [P5])

Emotion	<i>P</i> -value
Joy vs anger	0.519
Joy vs sadness	0.009**
Joy vs neutral	0.179
Sadness vs anger	0.001***
Sadness vs neutral	0.109
Anger vs neutral	0.013*

*Note.* Statistically significant differences are marked with asterisks: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

*The  $F_0$  level of sentence beginning and sentence end in emotional and neutral speech.* The results showed that, at the beginning of a sentence, the  $F_0$  level was the highest in neutral speech (median 231.6 Hz) and the lowest in sadness sentences (median 213.1 Hz). At the end of a sentence, the highest  $F_0$  level value was also found in neutral speech (median 164.9 Hz), but the lowest  $F_0$  belonged to anger sentences (median 161.4 Hz). There was no difference between the  $F_0$  values at sentence end when joy or sadness were expressed (median 164.4 Hz) (see Table 28).

**Table 28.** The  $F_0$  level value (Hz) at sentence beginning vs sentence end in emotional and neutral speech (see [P5])

Emotion	Sentence beginning / Sentence end				
	Minimum	Q1	Median	Q3	Maximum
Joy	163.2 / 141.1	201.2 / 158.4	225.8 / 164.4	252.5 / 171.7	314.3 / 187.7
Sadness	142.0 / 143.9	197.3 / 157.2	213.1 / 164.4	234.8 / 173.1	283.9 / 192.2
Anger	142.3 / 145.3	198.6 / 156.6	214.3 / 161.4	245.0 / 165.9	313.1 / 177.1
Neutral	163.1 / 139.4	203.7 / 157.1	231.6 / 164.9	249.3 / 170.7	284.7 / 186.8

According to the Wilcoxon rank-sum test, the  $F_0$  differences were not significant, either between emotions or in comparison with neutral speech, neither at sentence beginning nor end.



## 4.2. Results of the experiment to create acoustic models for emotional synthetic speech

Results of the evaluation of the test models of emotional speech are presented in article [P6].

### 4.2.1. Results of the evaluation of test models of emotional speech synthesis for a male voice

Table 29 presents the results of evaluators trying to capture emotions from a synthetic male voice, as presented to them in three test models. The perception is regarded as successful (emotion correct) only if its average rate of identification exceeds the chance probability (25%) and is higher than that of the other emotions among the responses.

*The emotion of joy.* According to the perception test results, the joy emotion in the synthetic male voice of Model 1 was not identified correctly by the evaluators. In both Test A and Test B, it was mostly mistaken for sadness. The joy emotion of Model 2 was mistaken for anger or sadness in Test A and for sadness in Test B. Model 3 enabled recognition of the joy emotion in Test B by 55% of the evaluators, whereas in Test A, it was still mistaken for sadness (see Table 29).

*The emotion of sadness.* According to the results of perception tests, the evaluators failed to recognise the sadness of Model 1 in a male synthetic voice. Both in Tests A and B, it was mostly mistaken for neutral. The sadness of Model 2 was recognised in Test B by 50% of the evaluators, while in Test A it was perceived as neutral. Model 3 enabled correct recognition of sadness (80% of the evaluators) only in Test B, whereas in Test A it was mistaken for neutral (see Table 29).

*The emotion of anger.* According to the results of perception tests, the anger of Model 1 was not recognised correctly from a male synthetic voice. It was considered neutral in both Tests A and B. Nor was anger recognised when Model 2 was applied. In both tests, it was mostly mistaken for neutral. The anger emotion modelled according to Model 3, however, was recognised from the male synthetic voice by 60% of the evaluators in Test A and by 50% in Test B (see Table 29).

*Neutral speech.* The neutral speech of the male voice had been synthesised using the existent model of the synthesiser. According to the perception test results, the evaluators could not recognise neutral speech in the male synthetic voice, either in Test A or in Test B. In both tests, it was mostly mistaken for sadness (see Table 29).

**Table 29.** Confusion matrix (synthetic male voice, Tests A/B<sup>30</sup>). Emotion perception from a synthetic male voice in three different test models, in percent of responses (see [P6])

Target emotion	Multiple choice			
<b>Model 1</b>	<i>Joy</i>	<i>Sadness</i>	<i>Anger</i>	<i>Neutral</i>
Joy	10/30	45/40	15/10	30/20
Sadness	10/5	30/15	5/0	55/80
Anger	15/20	10/15	25/10	50/55
<b>Model 2</b>	<i>Joy</i>	<i>Sadness</i>	<i>Anger</i>	<i>Neutral</i>
Joy	5/10	35/70	35/10	25/10
Sadness	0/5	30/50	0/0	70/45
Anger	0/10	10/0	35/40	55/50
<b>Model 3</b>	<i>Joy</i>	<i>Sadness</i>	<i>Anger</i>	<i>Neutral</i>
Joy	25/55	35/25	25/15	15/5
Sadness	0/0	35/80	0/0	65/20
Anger	5/5	10/0	60/50	25/45
<b>Neutral</b>	10/5	65/55	10/5	15/35

*Note.* Correctly perceived emotions are in bold type.

*The summary of perception test results for the male synthetic voice.* The results showed that for the male synthetic voice the emotions synthesised according to Model 3 were best recognised by the evaluators. Neutrality was not recognised; instead, it was confused with sadness. A comparison of the results of Tests A and B revealed that, for the male synthetic voice, joy and sadness were better recognised in Test B, while anger scored better in Test A.

#### 4.2.2. Results of the evaluation of the test models of emotional speech synthesis for a female voice

Table 30 presents the responses of evaluators trying to capture emotions from a synthetic female voice, as presented to them in three test models.

*The emotion of joy.* The results of the perception tests showed that the joy emotion modelled according to Model 1 was correctly recognised in Test A by 40% of the evaluators (see Table 30). In Test B, joy was mistaken for sadness. In the case of Model 2, the evaluators failed to capture joy in the female synthetic voice, confusing it mostly with sadness, both in Tests A and B. The joy emotion modelled according to Model 3 was correctly recognised in Test A by 30% of the evaluators. In Test B, however, it was mostly confused with sadness.

<sup>30</sup> In Test A, the whole sound clip conveyed a single emotion. In Test B, the clip began with an acoustically neutral sentence but its colouring could change, from the second sentence on, either to joyous, sad or angry or remained neutral.

*The emotion of sadness.* As revealed by perception tests, the sadness emotion modelled according to Model 1 was not recognised by the evaluators from the female synthetic voice, being mostly mistaken for neutral, both in Tests A and B (see Table 30). When Model 2 was used, the sadness was correctly recognised in Test A by 60% of the evaluators, whereas in Test B, it was rather confused with neutral. The sadness emotion modelled according to Model 3 was correctly recognised, both in Tests A and B, by 50% and 75% of the evaluators, respectively.

*The emotion of anger.* According to the perception test results, the anger emotion modelled according to Model 1 was not recognised from the female synthetic voice: in Test A, the evaluators mistook it for joy and in Test B for neutral (see Table 30). When Model 2 was used, the anger emotion was correctly recognised in both Tests A and B by 45% and 50% of the evaluators, respectively. In the case of Model 3, the anger emotion was recognised even better, that is, by 65% of the evaluators in Test A and by 60% in Test B.

*Neutral speech.* The neutral speech of the female voice had been synthesised according to a current model of the synthesiser. The perception tests showed that neutral speech was correctly recognised, both in Tests A and B, by 50% and 40% of the evaluators, respectively (see Table 30).

**Table 30.** Confusion matrix (female synthetic voice, Test A/B). Emotion perception from a synthetic female voice in three different test models, in percent of responses (see [P6])

Target emotion	Multiple choice			
<b>Model 1</b>	<i>Joy</i>	<i>Sadness</i>	<i>Anger</i>	<i>Neutral</i>
Joy	<b>40/5</b>	20/55	25/20	15/20
Sadness	10/10	<b>30/30</b>	10/0	50/60
Anger	45/25	10/5	<b>25/25</b>	20/45
<b>Model 2</b>	<i>Joy</i>	<i>Sadness</i>	<i>Anger</i>	<i>Neutral</i>
Joy	<b>25/10</b>	40/50	5/35	30/5
Sadness	0/10	<b>60/35</b>	0/0	40/55
Anger	30/5	0/30	<b>45/50</b>	25/15
<b>Model 3</b>	<i>Joy</i>	<i>Sadness</i>	<i>Anger</i>	<i>Neutral</i>
Joy	<b>30/10</b>	25/50	20/40	25/0
Sadness	5/10	<b>50/75</b>	5/0	40/15
Anger	15/20	5/10	<b>65/60</b>	15/10
<b>Neutral</b>	15/25	25/30	10/5	<b>50/40</b>

*Note.* Correctly perceived emotions are in bold type.

*Summary results of the perception tests using a female synthetic voice.* The evaluators scored best in recognising sadness and anger emotions in the passages synthesised according to Model 3, but the joy emotion was best recognised when Model 1 was involved. Neutral speech was also recognised

correctly. Comparing the results of Tests A and B, we can see that the anger and sadness of the female synthetic voice were recognised equally well in both tests, while joy and neutral speech received higher recognition scores in Test A.

The *binary logistic regression analysis* applied to the results yielded by the acoustic test models of emotional synthetic speech demonstrated that, both for the male and female synthetic voices, correct perception of the emotion significantly depended on the intensity level and  $F_0$  range of the synthetic speech. The speech rate was also important, but only in the recognition of sadness. The pitch register, or  $F_0$  level, was not found to be an essential parameter in emotion perception.

## 5. DISCUSSION

In this chapter I will compare the results on the acoustics of Estonian emotional speech with those of earlier studies and with CPM predictions and offer recommendations for emotion modelling in parametric speech synthesis. In addition, I will give an assessment to the experiment of creating acoustic models of emotional synthetic speech and present the preferred models for three basic emotions and two synthetic voices.

### 5.1. Speech acoustic parameters

#### 5.1.1. Pauses

The aim of the pause study was to find out (1) whether the emotion of a sentence has any influence on the number, location, nature and duration of the pauses made by the speaker and (2) whether pauses are significant in distinguishing between emotions to the extent enabling listeners to recognise the emotion of a speech passage from pause differences alone.

As for the *number of pauses*, the results showed that in read Estonian emotional speech, the emotion of a sentence was not a contributing factor to the number of pauses, which – despite being different across emotions – mainly depended on the number of punctuation marks in the sentence. For all three emotions investigated, the reader mostly paused at punctuation marks. The comma was the only punctuation mark not always inducing a pause in reading whatever the emotion involved. The percentage of comma pauses was the highest in joy sentences and the lowest in sadness sentences. The differences across emotions were remarkable enough to enable the conclusion that all three emotions affected the number of comma pauses and that the number of comma pauses was significant in distinguishing between the emotions.

It is known from earlier studies (see, e.g., Murray & Arnott, 2008; Schröder, Cowie, Douglas-Cowie, Westerdijk & Gielen, 2001; Tisljár-Szabó & Pléh, 2014; Yildirim, Bulut, Lee, Kazemzadeh, Deng, Lee, Narayanan & Busso, 2004) that emotions can differ in pause frequency. Of the three emotions analysed, sadness has the highest number of pauses compared to joy and anger. The reason why my results on pause frequency differ from those of some earlier studies may well lie in different research material (read vs spontaneous speech, elicited vs acted emotions vs natural speech – see Chapter 2). It can be assumed that, in producing read speech, the reader keeps track of the punctuation marks available in the text and, with great probability, coordinates his/her pauses with those marks.

The only punctuation mark that, for any of the three emotions, did not always make the reader react with a pause was comma. Most frequently commas would elicit pauses in the case of joy sentences and least frequently in sadness sentences. The differences across emotions were significant enough,

thus enabling the conclusion that in Estonian read-aloud emotional speech comma pauses do have an emotion-distinctive power.

My study of *the nature of pauses* (breathing vs non-breathing) showed that although there are more breathing than non-breathing pauses in Estonian emotional speech, the frequency of breathing pauses need not be distinctive of different emotions, as for each of the three emotions analysed, the number of breathing pauses slightly surpassed that of the non-breathing ones (53–59%). However, a division of the pauses into punctuation-coordinated pauses and extra pauses revealed that, although the percentage of punctuation pauses was more or less similar for different emotions (56–61%), the extra pauses made a significant difference, occurring much more frequently in anger sentences (55%) than in those of sadness (31%) or joy (21%). Consequently, the rate of extra breathing pauses can be regarded as a significant parameter to distinguish an anger emotion from those of joy and sadness in read Estonian emotional speech.

An analysis of the coordination of punctuation and breathing pauses revealed that for the emotions of sadness and joy, the reader produced significantly more of end-of-sentence reathing pauses (68 % and 67 %) than for anger emotions (58%). At the samet ime, in anger sentences, the percentage of breathing pauses (67%) amongst comma pauses was significantly higher than in the sentences of joy (36%) or sadness (27%). The reason may be that in the case of anger the reader inhaled remarkably less frequently at sentence end than for a joy or sadness emotion, and so, falling short of oxygen, she had to inhale more frequently in the middle of the sentence. After all, the expression of anger is more consumptive of speaker energy than that of sadness or joy (see Ch. 5.1.4. below), which increases the speaker's demand of oxygen. Moreover, the pauses in anger sentences are shorter (see the results on *pause duration* below) than those accompanying sadness or joy, which reduces oxygen intake time for expressing anger.

According to the results on *pause duration*, the longest pauses were characteristic of sadness sentences and the shortest of anger sentences (see Ch. 4.1.1). This is in harmony with earlier results, stating that sadness brings longer pauses than joy or anger (Murray & Arnott, 2008; Schröder, Cowie, Douglas-Cowie, Westerdijk & Gielen, 2001; Tisljár-Szabó & Pléh, 2014; Yildirim, Bulut, Lee, Kazemzadeh, Deng, Lee, Narayanan & Busso, 2004).

Based on my results, it can be argued that emotions do influence the duration of certain types of pauses, such as full-stop pauses (breathing as well as non-breathing) and breathing pauses at exclamation marks (see Ch. 4.1.1). The duration of non-breathing full-stop pauses distinguishes sadness from joy and anger, while the duration of breathing full-stop pauses is distinctive of all three emotions analysed, and the duration of breathing exclamation-mark pauses tells apart sadness from anger.

To summarise the the results on pauses, here are my suggestions on how to model pauses for Estonian emotional synthetic speech:

(1) Location of pauses: For all three emotions, better insert the pauses at punctuation marks except the comma, which can be used as a signal to pause in sentences of joy and anger, but rather not in sadness sentences. As the rate of extra pauses was stably low (8–10%) depending, in half of the cases, either on the content of the text or on the person of the reader, it is unnecessary to model extra pauses.

(2) Nature of pauses: For all three emotions, at least half of the pauses should be modelled as breathing ones. The pauses should be placed at reasonable intervals, computable, for example, from the supposed duration of the declination. For anger, the coordination of punctuation and breathing is hardly relevant, whereas in joy sentences, breathing should rather occur at sentence end, and in sadness sentences, breathing and pauses at commas should be avoided.

(3) Duration of pauses: The duration of pauses should better be correlated with the speech rate appropriate for the particular emotion (see Ch. 4.1.3). For all three emotions, the longest pauses should occur at sentence end and the shortest at commas.

Although in places my own study of pauses showed significant differences between some emotions, the *perception test* revealed that listeners could not differentiate between emotions from mere pause differences. This result is contrary to those of Tisljár-Szabó and Pléh (2014), which state that they can. However, we cannot claim that pauses are totally irrelevant in distinguishing between emotions in Estonian emotional speech, because they can become important in combination with some other acoustic parameters, such as speech rate, fundamental frequency and intensity.

### 5.1.2. Formant frequencies and articulatory precision

The aim of the formant study was to find out (1) whether the emotion of a sentence may influence (a) the two first formant frequencies of Estonian short vowels and (b) the precision of the speaker's articulation and (2) whether that influence is enough to distinguish emotions from each other and from neutral speech. According to the results, emotions do affect both the  $F_1$  and  $F_2$  values of the vowels scrutinised, while the  $F_1$  values were influenced significantly more than those of  $F_2$ . As for the speaker's articulatory precision, it was significantly lower in sadness sentences than in neutral speech or in the case of joy and anger.

*Influence of emotions on the first and second formant frequencies in vowels.* According to several earlier studies, in sentences of joy, the  $F_1$  and  $F_2$  values in vowels may rise, while  $F_1$  is affected more than  $F_2$  (see Goudbeek, Goldman & Scherer, 2009; Juslin & Laukka, 2003; Kienast & Sendlmeier, 2000; Tartter, 1980). Some other studies, however, report a fall in  $F_1$  values compared to anger and sadness (see Pribil, Pribilova & Matousek, 2013). Scherer's CPM also predicts a fall in the  $F_1$  of vowels compared to neutral speech, both for enjoyment and elation.

According to my results on Estonian emotional speech, the  $F_1$  values significantly differed from neutral only in the vowel *i*, where joy had a lower  $F_1$  than neutral speech. Thus, the result for *i* confirms the CPM prediction. The same cannot be said about my results on *a* and *u*: For the emotion of joy, the  $F_1$  of *a* and *u* was similar to that found in neutral speech. A comparison of the  $F_1$  values of the vowels in joy sentences with other emotions revealed that significant differences concerned the vowels *a* and *i*. The  $F_1$  values of *a* showed a significant difference between joy and anger as well as between joy and sadness, while the average  $F_1$  value was the highest in the case of joy emotions. The  $F_1$  values of *i* showed a significant difference between joy and sadness, while the average  $F_1$  value was higher for joy emotions. These results seem to support the previous results, which argue for a rise in the  $F_1$  for joy emotions.

As for the  $F_2$  values, my results did not reveal any significant difference between the vowels in joy sentences and in neutral speech, or, as a matter of fact, in any other emotion. Hence, my results on the  $F_2$  values of the vowels in joy sentences differ from the previous results as well as from the CPM prediction.

As for *the emotion of sadness*, earlier research has stated a fall in the values of the  $F_1$  of vowels (Juslin & Laukka, 2003). Also, it has been noticed that the  $F_2$  of the vowels in sadness sentences is lower than in joy sentences (Goudbeek, Goldman & Scherer, 2009). Scherer's CPM predicts that, compared to neutral speech, sadness will raise the  $F_1$  and lower the  $F_2$  of voiced sounds.

According to my results on the  $F_1$  values of vowels, sadness, compared to the other emotions and to neutral speech, has the lowest  $F_1$  values in all three vowels analysed, while in the cases of *a* and *i*, the differences are statistically significant. As for the  $F_2$  values of the vowels, a significant difference was noticed only in the case of the vowel *a*, whose  $F_2$  for sadness was significantly higher than for anger. Consequently, my results on the formant values for sadness largely agree with previous results, but they do not confirm the CPM prediction.

In the case of *the emotion of anger*, earlier studies have observed a higher  $F_1$  and a lower  $F_2$  in vowels (Goudbeek, Goldman & Scherer, 2009; Juslin & Laukka, 2003; Murray & Arnott, 2008; Pribil, Pribilova & Matousek, 2013). Scherer's CPM also predicts that, compared to neutral speech, the  $F_1$  of voiced sounds will rise, and the  $F_2$  will fall, both for hot and cold anger. My results showed that for anger the  $F_1$  values of the Estonian vowels *a*, *i* and *u* were lower than in neutral speech, while in the cases of *a* and *i*, the differences were statistically significant. The  $F_1$  values of *a* and *i* were also significant in distinguishing anger from other emotions. In anger sentences, both *a* and *i* had a higher  $F_1$  than in sadness sentences. The  $F_1$  of the vowel *a* was lower for anger than for joy, but there was no significant difference between the  $F_1$  values of *i* in the sentences of anger and joy. Most of the  $F_2$  values of the anger vowels did not differ significantly from neutral speech or other emotions; the only exception was *a*, whose  $F_2$  was significantly lower for anger than for sadness. Thus, my results on the vowels of anger emotions partly agreed with earlier



results for  $F_1$  but disagreed for  $F_2$ . Nor did my results on vowel formants support the CPM prediction.

In conclusion, the results on the first and second formant frequencies have shown that, in Estonian speech, emotions mainly affect the first formant frequency of the vowels *a* and *i*. Specifically, in the case of *a*, differences in the values of  $F_1$  turned out to be significant in distinguishing the emotions of sadness and anger from neutral speech and in telling apart all three emotions. In the case of *i*, differences in the  $F_1$  values turned out to be significant in distinguishing all three emotions from neutral speech as well as in discriminating sadness from joy and anger.

*Articulatory precision.* According to earlier results, sadness goes with a low precision in articulation. Results differ for anger as well as joy: the precision may either rise for both emotions (Johnstone & Scherer, 2000; Juslin & Laukka, 2003; Kienast & Sendlmeier, 2000; Murray & Arnott, 2008; Paeschke, Kienast & Sendlmeier, 1999), fall for anger (Iida, Campbell, Higuchi, & Yasumura, 2003) or, for joy, remain level with the articulatory precision typical of neutral speech (Kienast & Sendlmeier, 2000).

According to my results, the only remarkable difference in articulatory precision was observed in sadness sentences, which were articulated with a far lower precision than neutral speech or any other emotion. As for the emotions of joy and anger, their articulatory precision remained similar to that of neutral speech. The results suggest that a decrease in the precision of articulation tends to characterise the expression of sadness in many languages. Results differ, however, for anger and joy. The reason may lie, for example, in different research material or in the linguistic or cultural specifics manifested in the expression of those emotions.

Based on the results of the formant study, I suggest that, for the synthesis of sadness sentences, the  $F_1$  values of vowels should be lowered in comparison with neutral speech as well as with other emotions, and the articulation should be made less precise.

### 5.1.3. Speech rate

The objective of the speech rate study was to learn whether emotions may influence the tempo of reading a text aloud and whether speech rate is a significant factor to differentiate between emotions as well as between emotions and neutral speech.

My results on the tempo of read Estonian emotional speech showed that, in the material analysed, emotions did influence the speech rate, both overall and intra-phrase. Specifically, the fastest speech was typical of anger, followed by joy then neutral speech, while sadness sentences were the slowest of all. The overall average speech rate was distinctive of each of the three emotions, and it also distinguished anger from neutral speech. The same can be stated about the intra-phrase speech rate, plus it also distinguishes sadness from neutral speech. The duration of the phrase-final word was not significantly affected by emotions.

The results on the rate of read Estonian emotional speech largely coincided with Scherer's (1986) CPM prediction (see Ch. 2.1) in that, compared to neutral speech, the average speech rate would rise in anger sentences and fall in sadness sentences. As for the average speech rate for joy, the results differed from the CPM prediction, as in joy sentences, the average speech rate remained similar to that of neutral speech.

A comparison of the above results on Estonian emotional speech with earlier ones (see Ch. 2.1.) reveals a large coincidence in that, compared to neutral speech, anger would make speech faster, whereas sadness would slow it down. Also, the tempo of anger and joy sentences would be quicker than for sadness.

The results on speech rate suggest that in Estonian the reflection of emotions on the average speech rate is similar to that in other languages; namely, the tempo is higher for anger and slower for sadness. (A similar result was received for pauses, pause duration being the other temporal parameter analysed: In Estonian emotional speech, pauses are shorter in anger sentences than in sadness sentences – see Ch. 4.1.1).

Based on the results on speech rate, I suggest that, for Estonian emotional synthetic speech, the speech rate should be modelled so that anger sentences are uttered the fastest and sadness sentences the slowest (descending speech rate order: anger > joy > neutral > sadness).

#### **5.1.4. Intensity**

The purpose of the study of intensity was to learn (1) whether emotions may influence the intensity level of the speaker's voice and its variation (range of intensity) and (2) whether the level and range of intensity are significant to distinguish between different emotions and between emotions and neutral speech.

My results on the intensity of read Estonian emotional speech revealed that, in the material analysed, the level of intensity did distinguish between the emotions as well as between the emotions and neutral speech, whereas the range of intensity had no such distinctive function. The level of intensity at sentence beginning and sentence end distinguished the emotion of sadness from neutral speech. The overall intensity level was the highest in neutral speech and the lowest in the sentences of sadness. Those results do not entirely coincide either with Scherer's (1986) CPM predictions or the intensity results obtained for some other languages (see Banse & Scherer, 1996; Johnstone & Scherer, 2000; Juslin & Laukka, 2003; Murray & Arnott, 2008), which argued that the intensity parameters would rise (compared to neutral) for elation and anger and fall for sadness and enjoyment.

The intensity results on read Estonian emotional speech differed the most from the respective data on other languages in the case of neutral speech. According to intensity measurements (see Ch. 4.1.4), for neutral (compared to emotional) speech, the reader's strength of voice stayed on a relatively even

level; in other words, during a neutral sentence, the intensity varied considerably less than during a sentence of, say, sadness or anger. It can be a specific feature of Estonian that neutral speech is louder and has a more even intensity level (i.e., with less variation) than emotional speech. In addition, the result could have been influenced by the reading style.

Based on the above results, I suggest that, when modelling Estonian synthetic speech, intensity level should be held the highest for neutral speech and the lowest for sad speech (descending intensity order: neutral > anger > joy > sadness). Although the results showed that the intensity range did not have a distinctive role, the range may still be important in combination with some other emotion-distinguishing acoustic features. Therefore, I recommend to model the intensity range the widest for sadness sentences and the narrowest for joy (descending order of intensity range: sadness > anger > neutral > joy).

### 5.1.5. Fundamental frequency

The objective of the study of fundamental frequency was to find out (1) whether emotions may influence the  $F_0$  level and  $F_0$  range of the speaker's voice and (2) whether the  $F_0$  level and  $F_0$  range are relevant parameters to distinguish between different emotions and between emotions and neutral speech.

According to my results on  $F_0$  in read Estonian emotional speech, emotions do affect the reader's  $F_0$  level and  $F_0$  range; notably, the  $F_0$  level distinguished the anger emotion from joy, sadness and neutral speech, while the  $F_0$  range distinguished sadness from the emotions of joy and anger, and the anger emotion from neutral speech. The  $F_0$  level at sentence beginning and sentence end did not have a distinctive function.

A comparison of my results with Scherer's (1986) CPM prediction (see Ch. 2.5.) revealed that the only coincidence can be observed in anger, which was the only Estonian emotion distinguished from neutral speech both by  $F_0$  level and  $F_0$  range: Compared to neutral speech, the anger emotion had a lower  $F_0$  and a narrower  $F_0$  range (as is predicted for cold anger).

If we compare the  $F_0$  results of Estonian emotional speech with those of other languages, we have to discuss the  $F_0$  level and  $F_0$  range separately. As stated above, in Estonian emotional speech, anger was the only emotion whose  $F_0$  level differed from that of the other emotions as well as of neutral speech. Earlier results on emotion acoustics have shown that the  $F_0$  level may distinguish not only the anger emotion but also, for example, joy and sadness from neutral speech, as well as those emotions from each other. As for the  $F_0$  level for anger, it has been observed as higher or lower than for joy, but higher than for sadness or for neutral speech (see Laukka, Neiberg, Forsell, Karlsson & Elenius, 2011; Murray & Arnott, 2008; Pell, Paulmann, Dara, Allasseri & Kotz, 2009). My results on the  $F_0$  level of anger do not agree with the earlier results (except where it was revealed that the  $F_0$  can be lower for anger than for joy). Thus, it can be assumed that, in Estonian emotional speech, the only

salient  $F_0$  value can be observed in the case of anger, which is lower than that of neutral speech or of any other emotion.

The measurements of the  $F_0$  range revealed that, in Estonian emotional speech, the  $F_0$  range distinguishes (1) sadness from joy and anger (sadness has the narrowest  $F_0$  range) and (2) anger from neutral speech (anger has a wider  $F_0$  range). Earlier results have shown that the  $F_0$  range can also distinguish joy and sadness from neutral speech as well as emotions from each other. As for the emotion of sadness, the previous results suggest that its  $F_0$  range can either be wider or narrower than in neutral speech, but it is narrower than for joy or anger (see Laukka, Neiberg, Forsell, Karlsson & Elenius, 2011; Murray & Arnott, 2008; Pell, Paulmann, Dara, Alasseri & Kotz, 2009). This means a partial coincidence with my results on sadness, according to which, the reader's fundamental frequency varied significantly less for sadness than for joy or anger. As for the emotion of anger, the earlier results have suggested that its  $F_0$  range can be either narrower or wider than for joy, but certainly wider than for neutral speech or sadness (see Laukka, Neiberg, Forsell, Karlsson & Elenius, 2011; Murray & Arnott, 2008; Pell, Paulmann, Dara, Alasseri & Kotz, 2009). My results on anger coincide with the previous ones to a rather small extent; notably, the  $F_0$  range for anger differed significantly (was wider) only from that of neutral speech. Thus, it can be assumed that, in read Estonian emotional speech, the reader's fundamental frequency varies (a) significantly less for sadness than for anger or joy and (b) significantly more for anger than for sadness or neutral speech.

By way of conclusion, my study of fundamental frequency parameters demonstrated that in read Estonian emotional speech the only emotion to influence the  $F_0$  level is anger, whereas the  $F_0$  range is affected by all three emotions, distinguishing anger from neutral and sadness from anger and joy. Although the differences between the  $F_0$  level or  $F_0$  range of the emotions investigated were not always statistically significant (both when analysed pairwise or compared with neutral speech), in my material, there still was a tendency for the  $F_0$  to differentiate between the emotions, and therefore, I suggest that the  $F_0$  level in synthetic Estonian speech should be modelled so that it is kept the highest for joy emotion and the lowest for anger emotion, while sadness and neutral speech can do with an equal  $F_0$  level (scheme: joy > neutral = sadness > anger). As for the  $F_0$  range, it should be modelled as widest for anger and narrowest for sadness (anger > joy > neutral > sadness).

### **5.1.6. Acoustic patterns of joy, sadness and anger**

In what follows, I will provide a short description of the acoustic patterns of joy, sadness and anger, compared to neutral speech, in read Estonian emotional speech, as discovered in my research material.

*Joy.* Compared to neutral speech, the acoustic pattern of Estonian joy is characterised by a similar articulatory precision, a faster speech rate, a lower intensity level and a narrower intensity range, a higher  $F_0$  level and a larger  $F_0$  range.

*Sadness.* Compared to neutral speech, the acoustic pattern of Estonian sadness is characterised by a lower articulatory precision, slower speech rate, lower intensity level and wider intensity range, similar  $F_0$  level and smaller  $F_0$  range.

*Anger.* Compared to neutral speech, the acoustic pattern of Estonian anger is characterised by a similar articulatory precision, higher speech rate, lower intensity level and wider intensity range, lower  $F_0$  level and larger  $F_0$  range.

## 5.2. Acoustic models of Estonian emotional synthetic speech

The purpose of the experiment to create acoustic models for emotional synthetic speech was to find the best models enabling an Estonian speech synthesiser to express joy, sadness and anger in the parametric synthesis of a male and a female voice.

As was revealed by perception tests, either of the two test hypotheses set for the test models of three basic emotions (H1: In the case of the male as well as female synthetic voices, the evaluators will best recognise the emotions if these are synthesised according to Model 3, with amplified parameters; H2: In the case of the male as well as female synthetic voices, the evaluators will best recognise neutral speech if it is synthesised according to the synthesiser's own model) were partly confirmed. Notably, both from the male and female synthetic voices, the listeners recognised emotions the best if the emotions had been modelled according to Model 3, except for joy in the female synthetic voice, which scored best if Model 1<sup>31</sup> was used. Neutral speech was not recognised from the male synthetic voice, being mistaken for sadness, but it was recognised correctly from the female synthetic voice. As for confusion, in the

---

<sup>31</sup> This result of the female synthetic voice was highly unexpected. After all, in the male synthetic voice joy was best recognised with Model 3, quite as expected. For the female synthetic voice, the scores of recognised joy in Models 1 and 3 were not too different: 40% and 30%, respectively. For the male synthetic voice, the best score of joy recognition was 55% (Model 3). For comparison: the best scores for other emotions were 65–80% (Model 3). Maybe one of the reasons for the unexpected score lies in our failure to create an adequate model of joy: In the female synthetic voice joy was, after all, most often confused with sadness. Another reason for the surprising result could lie in our modification of the intensity level, which in Model 1 moved the intensity level of joy closer to that of sadness, whereas in Model 3, on the contrary, the two intensity levels were moved further apart. The rest of the parameter values also changed so that Model 1 took the joy emotion closer to the sadness emotion, whereas Model 3 enhanced their mutual distance. Hence the change introduced in the intensity level may well have caused the superiority of Model 1 in the recognition of joy. Another reason for the surprising result could lie in the female donor voice. Maybe there was something specific in the voice, which made Model 1 more appropriate for the expression of joy than Model 3? As research of emotion acoustics in Estonian speech is still in its infancy we cannot give an exhaustive answer to that question at present.

case of both synthetic voices, joy was mostly confused with sadness, while sadness and anger were confused with neutral speech.

In the case of Model 3, which was the best test model for the Estonian speech synthesiser, the mean recognition rate (with male and female voice results aggregated) was 60 for sadness, 59 for anger and 30 for joy. For neutral speech, the recognition rate was 35.

According to the perception tests, the synthesiser could satisfactorily express sadness and anger in both male and female voices (as long as Model 3 was used) but failed for joy. The synthetic expression of neutral speech was satisfactory in the female voice only.

The average recognition rate of emotions, across all three models, was 32% for the male and 38% for the female synthetic voice. Thus, the synthesiser was better in adding emotions to the female voice than to the male one. One of the reasons may be that the original acoustic analysis lying at the basis of the test models had been based on a female voice.

A comparison of the results of Tests A and B leads to the conclusion that, in the case of the male synthetic voice, the emotions were better recognised in Test B, where the speech passage started with acoustically neutral speech but changed from the second sentence on, becoming either joyous, sad or angry or remained neutral. So it seems that, in the male synthetic voice, a change of emotion within the speech passage would contribute to correct recognition of the emotion. In the case of the female synthetic voice, the anger and sadness emotions were recognised more or less equally well, both in Tests A and B, while joy and neutrality were recognised better in Test A, where the whole passage carried one and the same emotion or stayed neutral. Thus, in contrast to the male synthetic voice, a change of emotion within the passage was not relevant for emotion recognition from the female synthetic voice.

An overall analysis of the results of the perception tests, using logistic regression, showed that emotion perception strongly depended on the parametric values of the intensity level and  $F_0$  range used in the test models. Speech rate was also relevant, but its contribution to emotion recognition was rather modest (except for sadness, where the contribution was remarkable). Pitch initialisation or register was not a relevant parameter for emotion recognition. However, other research has shown that  $F_0$  level is one of the main parameters distinctive of emotions (see, e.g., Arias, Busso & Yoma, 2014; Banse & Scherer, 1996; Busso, Lee & Narayanan, 2009; Pell, Paulmann, Dara, Allasseri & Kotz, 2009; ten Bosch, 2003). In Estonian emotional speech, the  $F_0$  level distinguishes anger from joy and sadness emotions as well as from neutral speech (see Ch. 4.1.5). Therefore, the  $F_0$  level for anger in the acoustic model for Estonian emotional speech synthesis needs some additional specification. The same holds for the values of the speech rate for the emotions of joy and anger.

A comparison of the results of logistic regression analysis and those of human acoustics research revealed that for human speech the speech rate has a significant role in distinguishing emotions from each other as well as from

neutral speech, whereas for synthetic speech, the speech rate was distinctive for sadness only. As for the  $F_0$  level, in human speech, it distinguished anger from neutral speech and from the other emotions, but in synthetic speech, it did not have a distinctive role for emotion groups. As for the  $F_0$  range, it distinguished sadness from the other emotions and anger from neutral speech in human speech, and it also had a significant distinctive role for the emotion groups. The level of intensity served to distinguish between different emotions as well as between emotions and neutral speech, both in human and synthetic speech.

Based on the results of the test models, I will now present the preferred emotion models for three basic emotions and two voices synthesised parametrically for synthetic Estonian speech (see Table 31).

**Table 31.** The preferred emotion models for a male and a female voice synthesised parametrically for synthetic Estonian speech (see [P6])

Emotion	Male synthetic voice	Emotion	Female synthetic voice
Joy	M3	Joy	M1
Sadness	M3	Sadness	M3
Anger	M3	Anger	M3

*Note.* M3 – a model enhancing the parameter values of human speech; M1 – a model decreasing the parameter values of human speech.

## 6. CONCLUSION

This study of the acoustics of Estonian emotional speech had two objectives: (a) to find out how the three basic emotions – joy, sadness and anger – are expressed acoustically in read Estonian speech and (b) to develop, based on the acoustic results, some acoustic models of Estonian emotional speech for a parametric speech synthesiser so that it could recognisably express joy, sadness and anger.

According to the objectives of the study, I set three research questions: (1) To what extent, if any, and in what direction do the emotions of joy, sadness and anger affect the values of the acoustic parameters (pauses, formants, speech rate, intensity and fundamental frequency) in a read Estonian text? (2) Which of those parameters enable the distinction between different emotions as well as between emotions and neutral speech in read Estonian emotional speech? (3) How could the results of the analysis be applied to create acoustic models of emotional speech for Estonian parametric text-to-speech synthesis?

The first two questions are answered in articles [P1], [P2], [P3], [P4] and [P5], while the third is answered in article [P6].

The study of pauses ([P1]) showed that emotions may influence the duration of certain pause groups and that those differences play a distinctive role for emotions. According to the perception test, however, listeners cannot differentiate between emotions from mere pause differences.

The study of formants ([P2]) established that, in Estonian emotional speech, emotions did affect the first formant frequency of the vowels *a* and *i*, while the emotion-induced changes in the formant values were significant (except in a couple of cases) in distinguishing emotions from each other and from neutral speech. An analysis of the articulatory precision demonstrated that sadness sentences were articulated with noticeably less precision than neutral speech or the other emotions investigated.

The study of speech rate ([P3]) revealed that emotions may indeed influence the speech rate, which was, notably, the highest for anger and the lowest for sadness. Differences in speech rate distinguished emotions from each other and from neutral speech.

The study of intensity ([P4]) proved that, in read Estonian emotional speech, the intensity level of the voice was indeed influenced by emotions: The intensity level was the highest for neutral speech and the lowest for sadness. Intensity level turned out to be another significant parameter to distinguish emotions from each other and from neutral speech. Emotions did not exert a noticeable influence on the variation of the intensity level of the reader's voice.

The study of the fundamental frequency ( $F_0$ ) ([P5]) showed that the  $F_0$  level only depended on anger: In anger sentences, the  $F_0$  level was significantly lower than for the other emotions or for neutral speech. The  $F_0$  range was influenced by all of the three emotions: In sadness sentences, the reader's fundamental frequency displayed significantly less variation than for joy or anger, while in



anger sentences, the fundamental frequency varied significantly more than for sadness or for neutral speech.

The model-creating experiment ([P6]) proved that emotion models based on acoustic analysis enable the Estonian speech synthesiser to satisfactorily express sadness as well as anger. For joy, the results were, alas, less commendable. As my study of emotion acoustics was confined to the texts read by a single speaker, my results refer to just one possible way to express those emotions in Estonian, which means that the emotion models based on those results are hardly the only way to represent those emotions in Estonian synthetic speech. Nevertheless, my studies were the very first attempt to make the Estonian speech synthesiser express joy, sadness and anger, thus creating a possible starting point for further development of acoustic models for Estonian emotional speech synthesis.

## 7. SUMMARY IN ESTONIAN

### Põhiemotsioonid eestikeelses ettelõetud kõnes: akustiline analüüs ja modelleerimine

#### 7.1. Väitekirja eesmärk

Emotsioonide häälelist väljendumist on uusimal ajal uuritud alates 19. sajandi lõpust, peamiselt psühholoogia ja psühhiaatria valdkonnas. Tänu tehnoloogia arengule on uurimisvõimalused üha enam suurenenud ning emotsionaalse kõne uurimine on laienenud ka teistesse valdkondadesse, sealhulgas keeleteadusesse, kus viimastel aastakümnetel on hakatud emotsioonidega tegelema ennekõike seoses kõnetehnoloogiaga (vt nt Burkhardt & Campbell, 2015; Juslin & Scherer, 2005; Scherer, 2003; Schröder, 2001).

Kõnetehnoloogia üks haru on kõnesüntees, mille abil muudetakse kirjalik tekst suuliseks, matkides inimkõnet. Kõnesünteesi rakendatakse paljudes valdkondades, näiteks inimese ja masina suhtluses, multimeedias või nägemis-, lugemis- ja kõnepuudega inimeste abivahendites. Seepärast on väga oluline, et sünteeskõne kõlaks loomulikuna, võimalikult inimese rääkimise moodi kõigis selle aspektides.

Inimkõne juurde kuulub muu hulgas emotsioonide väljendamine. Emotsioonid on inimkõnes alati olemas ja seepärast peaksid need olema tajutavad ka sünteeskõnes. Võimalusi, kuidas emotsioone sünteeskõnele lisada, on mitu. Üks neist on luua kõnesüntesaatori jaoks emotsionaalse kõne akustilised mudelid (vt nt Audibert, Aubergé & Rilliard, 2005; Iriondo, Alias, Melenchón & Llorca, 2004). Selleks on vaja ühelt poolt teada, milliste akustiliste parameetrite väärtusi millisel määral ja mis suunas emotsioonid mõjutavad, teisalt tuleb arvesse võtta seda, millised on konkreetse kõnesüntesaatori parameetrilise häälestamise võimalused.

Emotsiooniakustika uurimustest on teada, et igal emotsioonil on teda teistest emotsioonidest eristav akustiliste parameetrite väärtuste kombinatsioon ehk akustiline muster (vt nt Banse & Scherer, 1996; Juslin & Scherer, 2005; Paeschke, Kienast & Sendlmeier, 1999). Kuulajad suudavad emotsioonid kõnes ära tunda ainuüksi heli põhjal, ilma kõneleja nägu nägemata (näiteks telefonivestlustes) (Bachorowski 1999). Samuti ollakse võimelised määrama kõnelõigu emotsiooni siis, kui kõik selle lõigu sõnad on tähenduseta (Goudbeek, Goldman & Scherer, 2009). See, et kuulajad suudavad ka ainult hääle järgi emotsioone usaldusväärsetl ära tunda, toob kaasa oletuse, et häälel kannab informatsiooni kõneleja emotsioonide kohta ja et emotsioonide häälelistel väljendustel on erinevad akustilised muustrid.

Kõnelaine koosneb paljudest akustilistest parameetritest, mida emotsioonid võivad mõjutada. Osa neist parameetritest on põhjalikumalt uuritud kui teisi. Põhjus, miks osale parameetritest rohkem tähelepanu pööratakse kui teistele, peitub selles, et neid saab kergemini mõõta ja analüüsida ning nüüdisaegses

kõnesünteesis<sup>32</sup> ja -tuvastuses edukalt rakendada (El Ayadi, Kamel & Karray, 2011; Koolagudi & Rao, 2012; Nose & Kobayashi, 2011; Schröder, 2009; Schuller, Batliner, Steidl & Seppi, 2011). Uurimustes enim mainitud emotsiooni kandvad akustilised parameetrid on põhitoon, kõnetempo ja intensiivsus. Nende kõrval on nimetatud ka palju teisi parameetreid, muu hulgas pause ning formante ja artikulatsiooni täpsust (Juslin & Scherer, 2005; Koike, Suzuki & Saito, 1998; Montero, Gutierrez-Arriola, Palazuelos, Enriquez, Aguilera & Pardo, 1998; Murray & Arnott, 2008; Toivanen, Waaramaa, Alku, Laukkanen, Seppänen, Väyrynen & Airas, 2006).

Kuna emotsioonide akustiline väljendumine võib erineda keeleti/kultuuriti ning emotsioone tuntakse heli järgi paremini ära sama keele/kultuuri sees (vt nt Altorv & Pajupuu, 2015; Elfenbein, 2013; Kamaruddin, Wahab & Quek, 2012; Paulmann & Uskul, 2014; Soto & Levenson, 2009), siis ei saa teiste keelte/kultuuride uurimistulemusi automaatselt eesti keelde üle võtta. Eestikeelsele kõnesüntesaatorile emotsionaalse kõne akustiliste mudelite loomiseks on vaja eraldi uurimust.

Eestis hakati emotsionaalse kõnega kõnetehnoloogia kontekstis tegelema 2006. aastal, kui Eesti Keele Instituudis alustati Eesti emotsionaalse kõne korpuse loomist<sup>33</sup>. Korpus loodi kahest eesmärgist lähtuvalt: see pidi olema usaldusväärne andmekogu emotsioonide uurimiseks ning rakendatav kõnetehnoloogilistes ülesannetes (Altrov, 2014; Altrov & Pajupuu, 2012). Kuna Eesti Keele Instituudis tegeleti samal ajal ka eestikeelse kõnesünteesi arendamisega (vt Mihkla, Piits, Nurk & Kiissel, 2008; Mihkla, 2009), siis nähtigi korpuse esmast kõnetehnoloogilist rakendust just emotsioone väljendavas kõnesünteesis (Pajupuu, 2012).

Eesti Keele Instituudis oli seega 2008. aastaks olemas sobiv kontekst, mis lubas mul hakata uurima emotsioonide väljendumist eestikeelses kõnes. Sihiks oli saada eestikeelse emotsionaalse kõne akustika kohta baastadmisi ning püüda neile uurimistulemustele tuginedes luua eestikeelse sünteeskõne tarvis akustilised mudelid, mis aitaksid kõnesüntesaatoril emotsioone sünteeskõnes väljendada.

**Minu doktoritööl on kaks eesmärki: a) saada teada, milline on kolme põhiemotsiooni – rõõmu, kurbuse ja viha – akustiline väljendumine eestikeelses ettelõetud kõnes ning b) luua eestikeelsele kõnesüntesaatorile parameetrilise sünteesi jaoks emotsionaalse kõne akustilised mudelid, mis aitaksid süntesaatoril äratuntavalt väljendada rõõmu, kurbust ja viha.**

Nendest eesmärkidest lähtudes püstitasin järgmised **uurimisküsimused**: 1) kas, millisel määral ja mis suunas mõjutavad eestikeelses ettelõetud emotsionaalses kõnes emotsioonid (rõõm, kurbus ja viha) akustiliste parameetrite

---

<sup>32</sup> Eri keeltes sünteesitud emotsioonide näiteid saab kuulata aadressil <http://emosamples.syntheticsspeech.de/> (viimati kontrollitud 21.08.2017).

<sup>33</sup> Riikliku programmi „Eesti keele keeletehnoloogiline tugi 2006–2010“ projekt „Eesti emotsionaalse kõne korpus“. Korpus asub aadressil <http://peeter.eki.ee:5000/> (viimati kontrollitud 21.08.2017).

(pausid, formandid, intensiivsus, kõnetempo ja põhitoon) väärtusi võrreldes üksteise ja neutraalse kõnega; 2) millised neist akustilistest parameetritest on eestikeelses ettelõetud emotsionaalses kõnes need, mis võimaldavad emotsioone üksteisest ja neutraalsest kõnest eristada ning 3) kuidas rakendada saadud uurimistulemusi eesti tekst-kõne sünteesile emotsionaalse kõne akustiliste mudelite loomisel.

## **7.2. Väitekirja struktuur**

Väitekirja koosneb sissejuhatavast osast ja kuuest publikatsioonist (vt [P1], [P2], [P3], [P4], [P5], [P6]). Sissejuhatav osa on jagatud kuueks peatükiks. Esimene peatükk tutvustab uurimuse eesmärke, kirjeldab väitekirja struktuuri ning annab lühikese ülevaate publikatsioonidest ja autori panusest kaasautoriga artiklites. Teine peatükk esitab väitekirja teoreetilised lähtekohad kolme uuritud põhi-emotsiooni – rõõmu, kurbuse ja viha – mõjust kõne akustilistele parameetritele. Kolmas peatükk kirjeldab põhiemotsioonide akustilise analüüsi uurimismaterjali ja -meetodit ning emotsionaalse sünteeskõne akustiliste mudelite loomise materjali ja meetodit. Neljas peatükk tutvustab eestikeelse emotsionaalse kõne akustiliste parameetrite mõõtmistulemusi ja emotsionaalse sünteeskõne akustiliste mudelite loomise eksperimendi tulemusi. Viies peatükk sisaldab võrdlust käesoleva töö uurimistulemuste ning varasemate uurimuste vahel, soovitusi emotsioonide modelleerimiseks parameetrilise kõnesünteesi jaoks, hinnangut emotsionaalse sünteeskõne akustiliste mudelite loomise eksperimendile ning eestikeelse parameetrilise kõnesünteesi eelistatuid emotsioonimudeleid kolme põhiemotsiooni ja kahe sünteeshäälte kohta. Kuues peatükk sisaldab väitekirja kokkuvõtet. Sissejuhatavale osale järgneb seitsmes peatükk, mis sisaldab väitekirja sissejuhatust eesti keeles. Sellele järgneb loetelu teadustekstidest, millele on väitekirjas viidatud.

## **7.3. Väitekirja publikatsioonide ülevaade ja autori panusest kaasautoriga artiklites**

Kuus publikatsiooni jagunevad teemade järgi kaheks. Esimesed viis artiklit käsitlevad eestikeelse ettelõetud emotsionaalse kõne akustikat. Need annavad ülevaate selle kohta, kas, mil määral ning mis suunas mõjutab lause emotsioon kõne akustiliste parameetrite (pausid, formandid, artikulatsiooni täpsus, kõnetempo, intensiivsus ja põhitoon) väärtusi. Kuues publikatsioon tutvustab eestikeelse emotsionaalse sünteeskõne akustiliste mudelite loomise eksperimenti ja selle tulemusi. Kahel publikatsioonil ([P3], [P6]) on kaasautor (Meelis Mihkla), kes on nimetatud publikatsioonide ülevaadetega tutvunud ning on nendega nõus.

[P1] käsitleb pause eestikeelses ettelõetud emotsionaalses kõnes. Artiklis antakse vastus kahele uurimisküsimusele: kas pauside arv, asukoht, iseloom ja

kestus sõltuvad teksti emotsioonist ja kas pausid on emotsioonide üksteisest eristamisel olulised ning kas kõnelõigu emotsioon on ära tuntav ainult pausi-erinevuste põhjal.

[P2] keskendub eestikeelse etteloetud emotsionaalse kõne formantidele ja artikulatsiooni täpsusele. Artiklis uuritakse, kas emotsioonid mõjutavad vokaalide esimest ja teist formantsagedust ja kõneleja artikulatsiooni täpsust ning kas nende parameetrite väärtuste erinevused eristavad emotsioone üksteisest ja neutraalsest kõnest.

[P3] fookuses on kõne ajaline struktuur ja emotsioonide mõju sellele. Artiklis saavad vastuse kaks uurimisküsimust: kas emotsioonid mõjutavad kõnetempot (ja kas kõnetempo erinevused eristavad emotsioone üksteisest ja neutraalsest kõnest) ning kas emotsioonid tekitavad erinevusi sõnaprosoodias. Artikli esimese autori panus puudutab kõnetempot: selle teooriaosa, uurimismaterjali ja -meetodi kirjeldust ning mõõtmistulemuste esitust ja analüüsi.

[P4] tutvustab eestikeelse etteloetud emotsionaalse kõne intensiivsuse uurimust. Artiklis saavad vastuse kaks uurimisküsimust: kas emotsioonid mõjutavad kõneleja hääle intensiivsuse taset ja ulatust ning kas nende parameetrite väärtuste erinevused on emotsioonide eristamisel üksteisest ja neutraalsest kõnest olulised.

[P5] analüüsib eestikeelse etteloetud emotsionaalse kõne põhitooni. Artiklis vastatakse küsimustele, kas lause emotsioon mõjutab kõneleja hääle põhitooni kõrgust ja ulatust ning kas nende parameetrite väärtuste erinevused on emotsioonide eristamisel üksteisest ja neutraalsest kõnest olulised.

[P6] kirjeldab katset luua eesti emotsionaalse kõne akustika uurimistulemustele ([P3], [P4], [P5]) tuginedes emotsionaalse kõne akustilised mudelid, mis aitaksid eestikeelsel parameetrilisel kõnesünteesil väljendada põhiemotsioone (rõõmu, kurbust ja viha) nii mees- kui ka naissünteesihääle baasil. Artikli esimene autor on kirja pannud eesti emotsionaalse kõne akustika teooriaosa ning koostöös artikli teise autoriga loonud katsemudelid, koostanud tajutestid ja need läbi viidud ning analüüsinud tulemusi.

## **7.4. Rõõmu, kurbuse ja viha mõju kõne akustilistele parameetritele**

Inimkõne ei ole stabiilne, vaid varieeruv: kõnelemisel muutuvad muu hulgas nii hääle kõrgus, ulatus, tugevus kui ka kõnekiirus. Üks põhjus, miks need muutused tekivad, on kõneleja emotsioonid (teised põhjused võivad olla näiteks kõneleja sugu, vanus, sotsiaalne staatus ja tervislik seisund) (Juslin & Scherer, 2005).

Emotsionaalse kõne uurimused tegelevad kõnetehnoloogia valdkonnas peamiselt põhiemotsioonidega (Cowie, Douglas-Cowie, Tsapatsoulis, Votsis, Kollias & Fellenz, 2001; Iida, Campbell, Higuchi, & Yasumura, 2003; Juslin & Laukka, 2003), milleks loetakse enamasti rõõmu, kurbust, viha, hirmu, vastikust ja üllatust (Ekman, 1992).

Selleks et luua kõnesüntesaatorile emotsionaalse kõne akustilisi mudeleid, peab esmalt kindlaks tegema, kuidas emotsioonid hääles väljenduvad. Klaus R. Scherer töötas 1986. aastal emotsiooniakustika valdkonnas välja mudeli, mis ennustab emotsiooni mõju häälelisele väljendusele. Schereri *component process model* e CPM võtab arvesse psühholoogilisi ja füsioloogilisi mõjusid, mis emotsiooni väljendamisega kaasnevad, ning näitab, et on olemas emotsioonispetsiifilised akustilised mustrid. Scherer on kontrollinud ja täpsustanud oma mudelit kirjanduses leiduvate akustilis-foneetiliste tõendite baasil ja ka ise läbi viinud emotsiooniuurimusi (Banse & Scherer, 1996; Scherer, 1986; Scherer, 2009; Scherer & Meuleman, 2013).

Scherer (1986) kirjeldab emotsiooni kui kohanemismuutuste sarja: emotsiooni tekkimisest impulsi saanud närvisüsteem mõjutab nii kõneleja hingamist kui ka tema kõneorganite lihaspinget ning see toob kaasa erinevusi helisignaali akustikas. Näiteks midagi väga ebameeldivat avaldub tihti neelu ja kõri pitsituses, mille tagajärjel läheb kõnetrakt pingesse ning väljatulev hääl on kõrgema sagedusega (Thompson & Balkwill, 2006). Ka näoilmed võivad tekitada kõnesignaalis akustilisi muutusi, näiteks naeratamine (Tartter, 1980).

Kuigi erinevatest emotsiooniakustika uurimustest on teada, et igal emotsioonil on just talle omane akustiline muster, pole emotsioone kandvate akustiliste parameetrite uurimistulemused homogeensed: mõnikord on need üksteisele lausa vastandlikud (Cowie, Douglas-Cowie, Tsapatsoulis, Votsis, Kollias & Fellenz, 2001; Murray & Arnott, 1993). Seda võivad põhjustada mitu asjaolu. Näiteks võib põhjus peituda erinevas uurimismaterjalis (näideldud *vs.* spontaanne *vs.* etteleotud esilekutsutud emotsiooniga kõne). Igalühel neist on emotsiooniakustika uurimismaterjalina oma plussid ja miinused. Näiteks salvestatakse näideldud kõnet peamiselt helistuudio kontrollitud tingimustes, kus puudub taustamüra ning saab jälgida kõneleja kaugust mikrofonist (kauguse muutumine võib mõjutada kõne intensiivsuse väärtusi). Samuti on näideldud kõnes emotsioonid tõenäoliselt akustiliste parameetrite väärtusi enam võimendanud kui etteleotud või loomulikus kõnes. Samal ajal ei pruugi näitleja emotsioone alati esitada nii, nagu need esinevad päriselus: spontaansete ja loomulike. Nii võib näideldud kõne emotsioonidel olla teistsugune akustiline muster kui päriselu spontaansetel emotsioonidel (Juslin & Scherer, 2005; McIntyre & Roland, 2006; Scherer, 2013; Wilting, Kraemer & Swerts, 2006).

Loomulikus kõnes väljendatakse emotsioone tavaliselt spontaanselt. Selline materjal on autentne ning oleks loomulikkust taotlevale kõnesünteesile parim. Kuid loomulikul kõnel võib emotsioonide uurimismaterjalina olla ka puudusi: kõnelejaid võib häirida teadmine, et neid salvestatakse, ja nad suruvad seetõttu oma emotsioone alla, samuti võivad spontaanse kõne emotsioonide salvestused olla halva kvaliteediga (seda põhjustavad näiteks taustamüra või vestluspartnerite ühel ajal rääkimine) (Douglas-Cowie, Campbell, Cowie & Roach, 2003; Juslin & Scherer, 2005; McIntyre & Roland, 2006).

Esilekutsutud emotsioonid jäävad näideldud ja spontaanse kõne vahepeale: kõnelejale antakse teatud emotsiooniga tekst, mille ettelugemine kutsub temas esile soovitud emotsiooni ja see peegeldub tema hääles. Esilekutsutud emot-

sioone saab samuti salvestada helistuudio kontrollitud tingimustes, mis teeb sellest hea akustilise analüüsi materjali. Samal ajal võivad esilekutsutud emotsioonid olla tagasihoidlikult väljendatud ja seepärast võivad nende akustiliste parameetrite väärtused emotsioone üksteisest ja neutraalsest kõnest nõrgemalt eristada (Juslin & Scherer, 2005; McIntyre & Roland, 2006; Scherer 2003).

Emotsiooniakustika uurimistulemuste erinevuste põhjus võib peituda ka uuritavas keeles ja kultuuris: kõnelemine on keeleline/kultuuriline tegevus ja igas ühiskonnas on emotsioonide väljendamisel oma reeglid (Altrov & Pajupuu, 2015; Douglas-Cowie, Campbell, Cowie & Roach, 2003; Johnstone & Scherer, 2000). Lisaks võivad emotsioonide väljendamist mõjutada kõneleja individuaalsed eripärad (Juslin & Scherer, 2005).

Veel võib uurimistulemuste erinevuste põhjus olla emotsioonide määratlus, neile antav sisu. Viha all võib mõista nii raevu kui ka lihtsalt pahameelt, rõõmuks võib pidada nii rahulolu kui ka vaimustust ning kurbuseks lugeda nii pettumust kui ka muret. Kuigi emotsiooni nimetus on sama (näiteks rõõm või viha), võivad rahulolul ja vaimustusel või pahameelel ja raevul olla erinevad akustilised muustrid (Burkhardt, Audibert, Malatesta, Türk, Arslan & Auberge, 2006; Douglas-Cowie, Campbell, Cowie & Roach, 2003; McIntyre & Roland, 2006; Scherer, 2003; Wilting, Kraemer & Swerts, 2006).

Kõnesüntesaatorile akustiliste mudelite loomiseks peab teadma, kas, millisel määral ja mis suunas emotsioonid akustiliste parameetrite väärtusi võrreldes neutraalse kõnega liigutavad. Vaatamata emotsionaalse kõne akustika uurimistulemustes aeg-ajalt esinevale vastandlikkusele on uurimused näidanud, et esinevad siiski mingid üldised tendentsid teatud emotsioonide ja nende akustiliste korrelaatide vahel (Goudbeek, Goldman & Scherer, 2009; Iida, Campbell, Higuchi & Yasumura, 2003; Kienast & Sendlmeier, 2000; Murray & Arnott, 2008; Paesche, Kienast & Sendlmeier, 1999).

Järgnevalt kirjeldan kolme põhiemotsiooni (rõõmu, kurbuse ja viha) ning viie akustilise parameetri (pauside, formantide, kõnetempo, intensiivsuse ja põhitooni) omavahelisi seoseid.

#### **7.4.1. Pausid ja emotsioonide mõju nendele**

*Paus* on helitu intervall (vaikus) ühe lause sees või kahe lause vahel. Paused mõõdetakse enamasti millisekundites. Uurimistulemused on näidanud, et pausid võivad emotsiooniti erineda nii kestuse kui ka esinemise sageduse poolest ning kuulajad suudavad emotsioone eristada ka ainult pausierinevuste põhjal. Kõige selgemini eristuvad kurbuselausete pausid: võrreldes näiteks rõõmu- või vihalause pausidega on nende esinemise sagedus suurem ja kestus pikem (Murray & Arnott, 2008; Schröder, Cowie, Douglas-Cowie, Westerdijk & Gielen, 2001; Tisljár-Szabó & Pléh, 2014; Yildirim, Bulut, Lee, Kazemzadeh, Deng, Lee, Narayanan & Busso, 2004).

Pauside kohta Schereri CPM ennustusi ei sisalda.

## 7.4.2. Formantsagedused ja artikulatsiooni täpsus ning emotsioonide mõju nendele

*Formantsagedused* on akustilise energia kontsentratsiooni piirkonnad kõnespektris. Kõige alumist piirkonda nimetatakse  $F_1$ , järgmist  $F_2$  jne. Formantsagedusi mõõdetakse hertsides (Hz) ja barkides (Bark, psühhoakustiline skaala). Kõikidel helilistel häälikutel on oma formantstruktuur. Formantsageduste asukohad ei ole fikseeritud, need sõltuvad kõnetrakti suurusest ja kujust. Kõnetrakti suurust ja kuju muudavad muu hulgas ka emotsioonid: kui kõneleja emotsionaalne seisund muutub, siis muutuvad tema näoilme ning häälduselundite (näiteks keel, lõug ja huuled) asend ja lihaspinge, mis omakorda tekitab muutusi häälikute formantstruktuuris. Näiteks panevad viha ja rõõm kõneleja rääkimisel rohkem pingutama (ta püüab selgemalt artikuleerida). Lihaspinge aga tekitab kõris pitsituse. Pingutamine ja pitsitus viivad kõnetrakti lühenemiseni ning see võib tingida kõrgemad  $F_1$  väärtused ja täpsema artikulatsiooni. Rõõmu puhul muudab kõnetrakti lühemaks ka naeratamine. Kurbuse puhul pole aga kõris sellist lihaspinget ja kõneleja räägib lödvestunumalt. See võib kaasa tuua madalamad  $F_1$  väärtused ja ebatäpsema artikulatsiooni (Johnstone & Scherer, 2000; Juslin & Laukka, 2003; Juslin & Scherer, 2005; Tartter, 1980; Tolkmitt & Scherer, 1986; Ververidis & Kotropoulos, 2006).

Osa emotsionaalse kõne akustika uurimistulemusi on näidanud, et rõõmulause vokaalidele on omane  $F_1$  ja  $F_2$  väärtuste tõus, kusjuures  $F_1$  on enamikul vokaalidel rohkem mõjutatud kui  $F_2$  (Goudbeek, Goldman & Scherer, 2009; Juslin & Laukka, 2003; Kienast & Sendlmeier, 2000; Tartter, 1980). Samal ajal on teised uurijad väitnud, et meeldivatel emotsioonidel (näiteks rõõm)  $F_1$  väärtused langevad (Pribil, Pribilova & Matousek, 2013).

Kurbuse-emotsiooni puhul on märgatud vokaalide  $F_1$  väärtuste langust (Juslin & Laukka, 2003). Märgitud on ka seda, et kurbuselauses on vokaalide  $F_2$  väärtused madalamad kui rõõmu-emotsiooni vokaalidel (Goudbeek, Goldman & Scherer, 2009).

Viha puhul on täheldatud vokaalide  $F_1$  väärtuste tõusu ning  $F_2$  väärtuste langust (Goudbeek, Goldman & Scherer, 2009; Juslin & Laukka, 2003; Murray & Arnott, 2008; Pribil, Pribilova & Matousek, 2013).

Schereri CPM ennustab, et võrreldes neutraalse kõnega tõuseb nii kuuma viha (raevu) kui ka külma viha (pahameele) ja kurbuselausete heliliste häälikute  $F_1$  väärtused ja langevad  $F_2$  väärtused. Rõõmu puhul ennustab CPM, et nii vaikselt kui ka lustlikul rõõmul  $F_1$  väärtused kahanevad.

*Artikulatsiooni täpsus.* Igal vokaalil on oma asupaik kahemõõtmelises vokaaliruumis, mida iseloomustavad vokaali kõrgus ehk vertikaalne dimensioon ( $F_1$ ) ja vokaali eespoolsus-tagapoolsus ehk horisontaalne dimensioon ( $F_2$ ). Need kaks mõõdet on seotud keele asukohaga suus vokaali moodustamisel. Selge kõne (täpse artikulatsiooni) puhul on vokaaliruum suur ja vokaalid asetsevad üksteisest piisavalt kaugel ning kuulaja suudab erinevate häälikute vahel kergesti vahet teha. Kui vokaaliruum muutub kitsamaks, siis kõneleja artikulatsiooni täpsus väheneb ning vokaalid, mida ei hääldata enam selgelt, redutseeruvad, st



nad kaotavad oma kvaliteeti. Emotsionaalse kõne artikulatsiooni täpsuse uurimustest on selgunud, et inimene, kes on näiteks stressis või depressioonis, ei artikuleeri helilisi häälikuid sama jõupingutusega nagu neutraalse kõne puhul. Sellisel juhul häälikud redutseeruvad ja nende artikulatsiooni täpsus langeb (Harrington, 2010: 46, 190–193; Tolkmitt & Scherer, 1986).

Varasematest emotsiooniakustika uurimustest on teada, et kurbuselauseste vokaalidele on omane madal artikulatsiooni täpsus, viha-emotsiooni vokaalidele aga kõrgem artikulatsiooni täpsus (Juslin & Laukka, 2003; Kienast & Sendlmeier, 2000; Murray & Arnott, 2008; Paeschke, Kienast & Sendlmeier, 1999). Viha kohta on saadud ka teistsuguseid tulemusi. Iida, Campbell, Higuchi ja Yasumura (2003) uurimusest selgus, et kõrgema põhitooni ja kiirema tempoga emotsioonidel (nagu viha) on vokaaliruum kitsenenud ning vokaalid rohkem redutseerunud. Rõõmulauseste vokaalide artikulatsiooni täpsuse kohta on saadud samuti kahesuguseid andmeid: artikulatsiooni täpsus kas tõuseb (Johnstone & Scherer, 2000; Juslin & Laukka, 2003; Murray & Arnott, 2008) või on sarnane neutraalse kõne artikulatsiooni täpsusega (Kienast & Sendlmeier, 2000).

### **7.4.3. Kõnetempo ja emotsioonide mõju sellele**

*Kõnetempo* väljendab kõnelemise kiirust, näitab kõnesegmentide hulka ühe ajaühiku kohta, näiteks häälikute arvu sekundis. Emotsionaalse kõne akustika uurimustest on teada, et kõnetempo varieerumine osutab muu hulgas ka kõneleja emotsionaalsele seisundile: väga aeglane tempo võib viidata kurvale või depressioonis kõnelejale. Samas tuleb meeles pidada, et kõnetempo on üsna subjektiivne tunnus, mis sõltub peale emotsionaalse seisundi veel kõneleja soost, vanusest, kõnestiilist, keelest, kultuuriruumist, suhtlussituatsioonist jms (Laver, 1994: 534).

Kuigi emotsionaalse kõne uurimustes on kasutatud erinevat kõnematerjali eri keeltest ja kultuuridest, on uurijad saanud kõnetempo kohta järjekindlalt sarnaseid tulemusi: võrreldes neutraalse kõnega langeb kõnetempo kurbuselausestes, tõuseb aga viha- ja rõõmulausestes. Võib oletada, et nende kolme põhiemotsiooni puhul kaldub kõnetempo olema keele- ja kultuuriruumi sarnane (Juslin & Laukka, 2003; Murray & Arnott, 2008; Scherer, 2003; Scherer, Sundberg, Tamarit & Salomão, 2015; Yildirim, Bulut, Lee, Kazemzadeh, Deng, Lee, Narayanan & Busso, 2004).

Schereri CPM ennustab, et võrreldes neutraalse kõnega tõuseb kõnekiirus kuuma viha lausungites, langeb aga kurbuselausestes. Rõõmu kohta ennustab CPM, et vaikse rõõmu puhul kõnetempo langeb, lustliku rõõmu puhul aga tõuseb.

### **7.4.4. Intensiivsus ja emotsioonide mõju sellele**

*Intensiivsus* näitab kõnelaines sisalduva energia hulka, näitab pingutust, mida kõnelemisel tehakse. Intensiivsuse taset tajutakse helitugevusena ning seda

mõõdetakse detsibellides (dB). Kuigi intensiivsuse taset on kerge mõõta, peab arvestama, et intensiivsus on tundlik salvestustingimustele: kui kaugel on mikrofon kõnelejust, kas salvestuskoht on vaikne (st puudub taustamüra), kas salvestustehnika on kalibreeritud jms. Kõik need asjaolud võivad mõõtmistulemusi mõjutada.

*Intensiivsuse ulatus* (maksimum- ja miinimumväärtuse vahe ühe kõneüksuse piires) näitab intensiivsuse taseme varieerumist: mida laiem on ulatus, seda enam intensiivsuse tase kõnelemise jooksul muutub.

Varasematest kõneakustika uurimistulemustest on teada, et kõneleja emotsionaalse seisundi ja tema kõne intensiivsuse taseme vahel on seos: lustlikku rõõmu ja viha seostatakse neutraalsest kõnest kõrgema intensiivsuse tasemega ning kurbust ja vaikset rõõmu neutraalsest kõnest madalama intensiivsuse tasemega (Banse & Scherer, 1996; Johnstone & Scherer, 2000; Juslin & Laukka, 2003; Murray & Arnott, 2008).

Schereri CPM ennustab lustlikule rõõmule intensiivsuse taseme näitajate (keskmine ja ulatus) tõusu, vaigse rõõmu puhul aga langust. Kurbuse puhul ennustab CPM intensiivsuse taseme näitajate langust. Viha puhul ennustab CPM nii kuumale vihale kui ka külmale vihale intensiivsuse taseme näitajate tõusu.

#### 7.4.5. Põhitoon ja emotsioonide mõju sellele

*Põhitoon* (põhisagedus,  $F_0$ ) näitab sagedust, millega kõneleja häälepaelad vibreerivad. Kuulaja tajub põhitooni helikõrgusena: kõneüksustel, mille moodustamisel vibreerivad häälepaelad kiiremini, on põhitoon kõrgem ning need üksused on kuulaja jaoks kõrgema helikõrgusega. Helikõrgus vastab füüsikas sagedusele ja seda mõõdetakse hertsides (Hz). Keskmine meeshääle põhitooni kõrgus on 80–150 Hz ja naishäälel 150–300 Hz.

*Põhitooni ulatus* (maksimum- ja miinimumväärtuse vahe ühe kõneüksuse piires) näitab kõneleja hääle põhitooni kõrguse varieerumist: mida laiem on ulatus, seda enam põhitooni kõrgus kõnelemise ajal üles-alla liigub.

Põhitoon on üks klassikalistest parameetritest, mida emotsionaalse kõne uurimustes mõõdetakse, sest põhitooni kõrgusel ja ulatusel on oluline roll emotsioonide eristamisel üksteisest ja neutraalsest kõnest (Arias, Busso & Yoma, 2014; Banse & Scherer, 1996; Busso, Lee & Narayanan, 2009; Pell, Paulmann, Dara, Allasseri & Kotz, 2009; ten Bosch, 2003).

Emotsiooniakustika uurijad on kõigi kolme emotsiooni põhitooni kohta saanud paiguti erinevaid tulemusi. Rõõmu-emotsiooni seostatakse üldiselt kõrge põhitooni ja laia ulatusega, kuid ka siin on uurijad saanud vastuolulisi tulemusi. Rõõmu-emotsiooni põhitoon võib olla nii madalam kui ka kõrgem kui viha-emotsiooni põhitoon, kuid kõrgem kui kurbuse-emotsiooni ja neutraalse kõne põhitoon. Rõõmu-emotsiooni põhitooni ulatus võib olla samuti kas laiem või kitsam kui viha-emotsiooni põhitooni ulatus, kuid laiem kui kurbuse-emotsiooni ja neutraalse kõne põhitooni ulatus (Murray & Arnott, 2008; Pell, Paulmann, Dara, Allasseri & Kotz, 2009). Schereri CPM ennustab rõõmu-emotsiooni põhi-

tooni kohta, et võrreldes neutraalse kõnega on vaikset rõõmu väljendaval kõnel madalam põhitoon ja kitsam ulatus ning lustlikku rõõmu väljendaval kõnel kõrgem põhitoon ja laiem ulatus (Scherer, 1986).

Kurbuse-emotsiooni kohta on varasemad uurimistulemused näidanud, et üldiselt on põhitoon kurbust väljendavas kõnes madal. See on madalam kui rõõmu- ja viha-emotsiooni puhul, kuid võrreldes neutraalse kõnega võib see olla nii kõrgem kui ka madalam. Kurbuse-emotsiooni põhitooni ulatuse kohta on varasemad uurimused saanud tulemuseks, et võrreldes rõõmu- ja viha-emotsiooni ning neutraalse kõnega on see kõige kitsam. Samal ajal on täheldatud, et kurbust väljendava kõne põhitooni ulatus võib olla ka laiem kui neutraalsel kõnel (Laukka, Neiberg, Forsell, Karlsson & Elenius, 2011; Murray & Arnott, 2008; Pell, Paulmann, Dara, Alasseri & Kotz, 2009). Schereri CPM ennustab, et võrreldes neutraalse kõnega on kurbuse-emotsiooni puhul põhitoon madalam, kuid olenevalt hääletüübist (näiteks pinges häääl) võib see olla ka kõrgem. Põhitooni ulatus kurbuse-emotsiooni puhul väheneb (Scherer, 1986).

Viha-emotsiooni põhitooni kohta on märgitud, et võrreldes rõõmu- ja kurbuse-emotsiooni ning neutraalse kõne põhitooniga on see kõige kõrgem. Samal ajal võib see olla ka madalam kui rõõmu-emotsiooni põhitoon. Viha-emotsiooni põhitooni ulatuse kohta on öeldud, et see on laiem kui rõõmu- ja kurbuse-emotsioonil ning neutraalsel kõnel. Samas on ka väidetud, et viha-emotsiooni põhitooni ulatus on rõõmu-emotsiooni põhitooni ulatusest kitsam (Laukka, Neiberg, Forsell, Karlsson & Elenius, 2011; Murray & Arnott, 2008; Pell, Paulmann, Dara, Alasseri & Kotz, 2009). Schereri CPM ennustab viha-emotsiooni põhitooni kohta, et võrreldes neutraalse kõnega võib põhitoon olla külma viha puhul kas kõrgem või madalam, kuid põhitooni ulatus väheneb. Kuuma viha puhul võib põhitoon olla samuti neutraalse kõne põhitoonist kas kõrgem või madalam, kuid põhitooni ulatus suureneb (Scherer, 1986).

## 7.5. Materjal ja meetodid

Doktoriväitekiri koosneb kahest suuremast osast: emotsionaalse kõne akustiliste parameetrite uurimisest ning eestikeelsele kõnesüntesaatorile emotsionaalse kõne akustiliste mudelite loomise eksperimendist. Järgnevalt tutvustan eraldi kummagi osa meetodeid ja materjali. Alapeatükis 3.1. käsitlen kõne akustiliste parameetrite uurimuse materjali ja meetodeid ning alapeatükis 3.2. emotsioonimudelite loomise eksperimendi materjali ja meetodit.

### 7.5.1. Kõne akustiliste parameetrite analüüsi materjal ja meetodid

Emotsioonide akustilise väljendumise *uurimismaterjal* on pärit Eesti emotsionaalse kõne korpusest<sup>34</sup> (Altrov, 2014; Altrov & Pajupuu, 2012). Korpus sisaldab eestikeelseid<sup>35</sup> ettelõetud ajakirjanduslikke tekstilõike, mis on segmenteeritud lauseteks, sõnadeks ja häälikuteks. Korpuses on kokku 1234 lauset, mis kõik on erinevad. Diktorina on kasutatud hea artikulatsiooni ja meeldiva häälega nooremas keskeas naisterahvast (ei ole näitleja) (Altrov & Pajupuu, 2013).

Eesti emotsionaalse kõne korpuses sisalduv kõnematerjal on salvestatud vaikselt ruumis digitaalse salvestaja ning kvaliteetse mikrofoniga, mis asus kõnelejust 50 cm kaugusel (WAV-vorming, 44,1 kHz, 16 bitti, mono).

Eesti emotsionaalse kõne korpuse loomisel on lähtunud printsibist, et tekstilõigu sisu kutsub lugejas esile emotsiooni ning see emotsioon kajastub lugeja hääles. Seega ei ole korpuses sisalduvad emotsioonid näideldud, vaid teksti poolt esile kutsutud. Emotsiooni defineerimisel on korpuse loojad lähtunud kõnetehnoloogias laialt kasutatavast tähendusest, et emotsioon on see, mis esineb kõnes kogu aeg, välja arvatud juhul, kui kõneleja on emotsioonitu (Altrov, 2014; Cowie, Sussman & Ben-Ze'ev, 2011; Schuller, Batliner, Steidl & Seppi, 2011). Korpuses sisalduvad emotsioonid ei ole täismahulised, vaid on väljendatud pigem tagasihoidlikult, ning need hõlmavad ka lähedasi emotsioone ja emotsioonilaadseid seisundeid (Altrov, 2014). Eesti emotsionaalse kõne korpusesse on valitud kolm põhiemotsiooni – rõõm, kurbus ja viha – seepärast, et neid kolme emotsiooni vajatakse kõnetehnoloogilistes rakendustes kõige enam (vt Altrov, 2014: 35). Kuna eesti tekst-kõne süntesaator on mõeldud peamiselt ajakirjandustekstide ettelugemiseks, siis on korpuse materjaliks võetud ajakirjanduslikud tekstilõigud (vt Altrov, 2014: 36).

Korpuses olevad laused on kõik kontekstist eraldatud (st tekstilõigud on korpuses segmenteeritud üksikuteks lauseteks). Selle üle, mis emotsiooni (rõõm, kurbus, viha või neutraalne) mingi lause kannab, on otsustanud hindajad (igat lauset on kuulunud ja selle emotsiooni määranud 30 hindajat), kes osalesid tajukatsetes (kuulamis- ja lugemistestid). Hindajaid juhendati mõtlema nii, et kõik kolm emotsiooni hõlmavad ka mitut teist sellega lähedalt seotud emotsiooni. Nii näiteks hõlmab rõõm ka tänuikkust, õnnelikkust, meeldivust ja vaimustust; kurbus üksindust, trööstitust, murelikkust ja lootusetust ning viha pahameelt, raevu, irooniat, vastumeelsust, põlgust ja kahjurõõmu. Neutraalset kõnet kirjeldati hindajatele kui normaalset, ilma eriliste emotsioonideta kõnet. Emotsiooni üle otsustasid hindajad ka neid lauseid lugedes, st lause verbaalse sisu järgi,

<sup>34</sup> Vt <http://peeter.eki.ee:5000/> (viimati kontrollitud 21.08.2017).

<sup>35</sup> Eesti keel kuulub Uurali keelte soome-ugri keelkonna läänemeresoome keelte allrühma. Eesti keelt räägib emakeelena ligi üks miljon inimest, kellest enamik elab Eestis, kus see on riigikeel. 2016. aasta alguses kõneles Eestis eesti keelt emakeelena 883 707 inimest, mis teeb 68% kõigist Eesti elanikest (vt <https://statistikaamet.wordpress.com/tag/emakeel/>, viimati kontrollitud 21.08.2017).

ilma heli kuulmata (Altrov, 2014; Altrov & Pajupuu, 2012; Pajupuu, Pajupuu, Tamuri & Altrov, 2015).

Korpuses on laused kuulamis- ja lugemistestide tulemuste põhjal jagatud kaheks rühmaks: 1) laused, mille verbaalne sisu kannab häälega samasugust emotsiooni (st hindajad on lause emotsiooniks määranud sama emotsiooni nii lugemis- kui ka kuulamistestis), ning 2) laused, kus hääletoon muudab emotsiooni, mida arvatakse lausel verbaalse sisu põhjal olevat (st hindajad on lugemistestis ja kuulamistestis määranud lausele erineva emotsiooni) (vt Pajupuu, Pajupuu, Tamuri & Altrov, 2015).

Minu uurimismaterjal koosnes lausetest, mille puhul on hääletoon muutnud lause verbaalses sisus peituvat emotsiooni (korpuse teine rühm, vt eespool) (näide 1) ning mille puhul lugedes pole olnud võimalik lause emotsiooni tuvastada, kuid hääle järgi on emotsioon ära tuntud (näide 2) (vt ka Pajupuu, Pajupuu, Tamuri & Altrov, 2015).

**Näide 1 (kurbus):** *Ükskõik, mida ma teen, ikka pole ta rahul!*

64,3% hindajaist oli lugedes määranud lause vihaks, kuid kuulates olid selle 80,0% hindajatest määranud kurbuseks.

**Näide 2 (rõõm):** *Ehkki Ott minu olemasolust midagi ei teadnud.*

Lause emotsiooni ei osanud hindajad lugemise põhjal määrata, kuid kuulates määras 87,5% hindajaist selle rõõmuks.

Uurimismaterjalina kasutatud lausete emotsiooni või neutraalsuse tuvastusprotsent oli kuulamise järgi vähemalt 51% (st kaks korda parem juhuslikust valikust).

Erinevate ülesannete puhul varieerus analüüsitavate lausete<sup>36</sup> hulk. Lisaks lausetele oli korpusest võimalik kasutamiseks valida ka löike. Selline vajadus tekkis kõnepauside uurimisel, vt 3.1.1.

*Uurimismeetod.* Kuna tegu on korpuspõhise uurimusega, siis on tulemusi analüüsitud statistilise analüüsi meetoditega (dispersioonanalüüs (ANOVA), Wilcoxon'i astaksummatest, eksimismatriksi ja binaarne logistiline regressioonanalüüs). Akustiliste parameetrite mõõtmisel kasutasin programmi Praat<sup>37</sup> (Boersma & Weenink, 2016) ning kõne andmebaaside süsteemi EMU<sup>38</sup>. Mõõtmistulemusi on analüüsitud statistikakeskkonnas R<sup>39</sup>.

Alljärgnevalt kirjeldan iga uuritud akustilise parameetri uurimismaterjali ja -meetodit eraldi.

---

<sup>36</sup> Uurimismaterjalina kasutatud eri emotsiooniga laused on esitatud lisades 1 ja 2 ning neid saab lugeda ja kuulata ka korpuse lehelt aadressil <http://peeter.eki.ee:5000/reports/segments?lg=et> (viimati kontrollitud 21.08.2017).

<sup>37</sup> Praat: <http://www.praat.org/> (viimati kontrollitud 21.08.2017).

<sup>38</sup> EMU <http://ips-lmu.github.io/EMU.html/> (viimati kontrollitud 21.08.2017).

<sup>39</sup> R <http://www.r-project.org/> (viimati kontrollitud 21.08.2017).

### 7.5.1.1. Pauside uurimismaterjal ja -meetod

Emotsionaalse kõne pauside uurimisele on keskendunud artikkel [P1], mis koosneb kahest osast: (1) uurimusest, kas pauside arv, asukoht, iseloom ja kestus sõltuvad teksti emotsioonist ning kas pausid on olulised emotsioonide eristamisel üksteisest, ning (2) tajukatsest, kas kõnelõigu emotsioon on ära tuntav ainult pausierinevuste põhjal.

Pauside arvu, asukoha (sh vastavust kirjavahemärkidele<sup>40</sup>), iseloomu (hingamisega vs. hingamiseta pausid<sup>41</sup>) ning kestuse uurimiseks valisin iga emotsiooni<sup>42</sup> kohta korpusest välja keskmiselt 16 lõiku, milles oli 120–123 lauset<sup>43</sup> (vt tabel 32).

**Tabel 32.** Pauside uurimuse materjal (vt [P1])

Emotsioon	Lõikude arv	Lausete arv	Sõnede arv	Kirjavahemärkide arv
Rõõm	16	120	966	181
Kurbus	15	120	1053	187
Viha	18	123	1124	201
KOKKU	49	363	3143	569

Pauside mõõtmiseks ja analüüsimiseks kasutasin programmi Praat (Boersma & Weenink, 2016). Tegin kindlaks tekstilõikudes esinenud pauside arvu, asukoha ja iseloomu ning jagasin pausid nende iseloomu ja asukoha järgi rühmadesse. Samuti mõõtsin pausirühmade kestusi millisekundites (miinimumväärtuse, esimese kvartiili, keskmise, mediaani, kolmanda kvartiili ja maksimumväärtuse) ning selgitasin välja, millised pausirühmad on kestuse järgi emotsioonide üksteisest eristamisel statistiliselt olulised.

Selleks et teada saada, kas lause emotsioon on äratuntav ainult pausierinevuste põhjal, valisin korpusest välja neljalausealise neutraalse lõigu<sup>44</sup> ning programmi Sound Forge Pro 9<sup>45</sup> abil muutsin selles lõigus pauside kestuse,

<sup>40</sup> Jagasin pausid teksti kirjavahemärkidele vastavuse järgi rühmadeks (punktipaus, komapaus jne).

<sup>41</sup> Hingamisega paus koosneb sissehingamisest, millele eelneb ja/või järgneb vaikus. Hingamiseta paus on selline sõnade või lausete vahel olev vaikus kõnes, mille kestus on vähemalt 30 ms. 30 ms valisin seepärast, et materjal oleks võrreldav varasemate eesti keelt puudutavate pauside uurimustega, kus pausi miinimumpikkuseks on võetud 30 ms (vt Pajupuu & Kerge, 2006; Tamuri, 2007: 14).

<sup>42</sup> Selles uurimuses ei võrrelnud ma emotsioone neutraalse kõnega, sest vastav võrdlusmaterjal emotsionaalse kõne korpuses 2010. aastal puudus.

<sup>43</sup> Pausiuurimuses kasutatud materjal on esitatud lisa 1.

<sup>44</sup> „Klaasipõhja valatakse seda vahest kolm sentimeetrit. Ainult tirtsukese rohkem, kui enesest lugupidav sommeljee mekkimiseks pakub. Seitsekümmend krooni, palun. Aitäh.“

<sup>45</sup> Vt [www.sonycreativesoftware.com](http://www.sonycreativesoftware.com) (viimati kontrollitud 21.08.2017).

asukoha ning iseloomu emotsionaalsele kõnele vastavaks pausianalüüsis saadud tulemuste põhjal. Moodustasin kaheksa erineva pausimustriga lõiku (2 x neutraalne, viha, kurbus ja rõõm). Neli neist olid lihtsama pausimustriga, kus olin muutnud ainult lauselõpupaaside kestusi. Teised neli olid keerulisema struktuuriga, seal olin muutnud nii paaside kestust, iseloomu kui ka asukohta. Neutraalsetes lõikudes ei muutnud ma midagi. Seejärel viisin läbi tajutesti, kus palusin kümnel katseisikul kuulata<sup>46</sup> kaheksat erineva pausimustriga lõiku ning heli põhjal määrata, mis emotsioon lõigus kõlab: kas rõõm, kurbus, viha või on lõik neutraalne, st ilma erilise emotsioonita. Hindajateks olid kaks meest ja kaheksa naist vanuses 30 kuni 70 eluaastat. Kõik hindajad olid rahvuselt eestlased ning nende emakeel oli eesti keel. Tajutestid viisin läbi augustis 2009.

#### 7.5.1.2. Formantsageduste ja artikulatsiooni täpsuse uurimismaterjal ja -meetod

Emotsionaalse kõne formantsageduste uurimisele on keskendunud artikkel [P2]. Uurimus koosneb kahest osast: (1) lause emotsiooni mõju lühikeste vokaalide esimesele ja teisele formantsagedusele ning (2) kõneleja artikulatsiooni täpsusele. Uurimuse eesmärk on teada saada, kas emotsioonid mõjutavad vokaalide esimest ja teist formantsagedust ja kõneleja artikulatsiooni täpsust ning kas nende parameetrite väärtuste erinevused eristavad emotsioone üksteisest ja neutraalsest kõnest.

Formantanalüüsiks valisin rõõmu-, kurbuse- ja vihalausetest ning neutraalsest kõnest kokku 2395<sup>47</sup> *a*-, *i*- ja *u*-vokaali (vt tabel 33). Need kolm vokaali moodustavad kolmnurkse vokaaliruumi (*a* on madal tagavokaal, *i* kõrge eesvokaal ja *u* kõrge tagavokaal) ning nende vokaalide abil saab hästi kirjeldada kõneleja artikulatsiooni täpsust. Kuna on teada, et emotsiooni kannavad rõhulised silbid (vt Seppi, Batliner, Steidl, Schuller & Nöth, 2010) ning mitterõhuliste silpide häälikud kipuvad oma kvaliteeti kaotama (vt nt Moon & Lindblom, 1994), siis valisin uuritavad vokaalid rõhulistest sõnaalgulistest positsioonidest VC, CVC või CCVC<sup>48</sup>. Esmalt mõõtsin lühikeste rõhuliste vokaalide *a*, *i*, ja *u* esimest ja teist formantsagedust. Mõõtmiskohaks valisin vokaalide keskpunkti.

---

<sup>46</sup> Kuulamistesti juhend: „Palun kuula kaheksat kõnelõiku ja märgi iga lõigu kohta, millist emotsiooni see Sinu arvates kannab. Selles testis ei ole õigeid ega valesid vastuseid. Vastupidi, Sinu arvamus põhjal saame teada, kui hästi ühes või teises lõigus emotsioon äratuntav on. Sa ei pea testi korraga lõpuni tegema, võid vahepeal salvestada ja hiljem edasi teha või varemtehtut muuta.“

<sup>47</sup> Formantsageduste uurimismaterjal on esitatud lisa 2.

<sup>48</sup> V – vokaal; C – konsonant.

**Tabel 33.** Formantanalüüsi materjal (vt [P2])

Vokaal	Emotsioon	Vokaalide arv
<i>a</i>	Rõõm	279
	Kurbus	244
	Viha	395
	Neutraalne	319
	KOKKU	1237
<i>i</i>	Rõõm	146
	Kurbus	154
	Viha	220
	Neutraalne	145
	KOKKU	665
<i>u</i>	Rõõm	133
	Kurbus	115
	Viha	138
	Neutraalne	107
	KOKKU	493
<i>a, i, u</i>	KOKKU	2395

Formantsageduste mõõtmisel kasutasin programmi Praat (Boersma & Weenink, 2016) ning kõne andmebaaside süsteemi EMU. Formantsageduste automaatse arvutamise vead korrigeerisin käsitsi. Iga emotsiooni ja neutraalse kõne kohta arvutasin välja esimese ja teise formantsageduse keskmised väärtused. Dispersioonanalüüsi (ANOVA) abil tegin kindlaks, millised formantsageduste erinevused on emotsioonirühmade (rõõm, kurbus, viha ja neutraalne kõne) üksteisest eristamisel statistiliselt olulised. Seejärel mõõtsin kõneleja artikulasiooni täpsust eukleidilise kaugusega. Selleks määrasin esmalt kõneleja vokaaliruumi keskpunkti ehk arvutasin välja kõneleja neutraalse hääliku  $x$ . Pärast seda arvutasin välja eukleidilise kauguse emotsionaalse ja neutraalse kõne vokaalide  $a$ ,  $i$  ja  $u$  keskmiste ning neutraalse hääliku  $x$  vahel, kasutades valemit  $d(x, V_{em}) = \sqrt{(F_1x - F_1V_{em})^2 + (F_2x - F_2V_{em})^2}$ , kus  $x$  on neutraalne vokaal ja  $V_{em}$  vaadeldava emotsiooni vokaali keskmine.

### 7.5.1.3. Kõnetempo uurimismaterjal ja -meetod

Emotsioonide mõju uurimist kõnetempole kajastab artikkel [P3]. Uurimus koosneb kahest osast, millest esimeses, kõnetempot puudutavas osas uurin, kas emotsioonid mõjutavad ettelugeja kõne kiirust ning kas kõnetempo väärtuste erinevused on emotsioonide eristamisel üksteisest ja neutraalsest kõnest olulised.



Kõnetempo uurimismaterjaliks valisin rõõmu, kurbust ja viha ning neutraalset kõnet väljendavad vähemalt kolmesõnalised laused, mida oli kokku 314 lauset<sup>49</sup> (vt tabel 34).

**Tabel 34.** Kõnetempo uurimismaterjal (vt [P3])

Emotsioon	Lausete arv
Rõõm	55
Kurbus	84
Viha	77
Neutraalne	98
KOKKU	314

Esmalt mõõtsin programmiga Praat (Boersma & Weenink, 2016) emotsionaalse ja neutraalse kõne tempot<sup>50</sup>. Kuna korpuses olevad laused on kõik erinevad, siis valisin kõnetempo mõõtmiseks häälikute arvu sekundis. Pikki foneeme arvestasin topelt, sest eeldasin, et fonoloogiliselt pikad häälikud on kahe lühikese foneemi jasad (Eek, 2008).

Seejärel arvutasin iga emotsiooni ja neutraalse kõne kohta välja keskmise kõnetempo ning võrdlesin tulemusi emotsioonipaariti ja neutraalse kõnega. Lisaks uurisin seda, kas kõnetempo varieerub ka ühe lause piires. Selleks mõõtsin nii emotsionaalse kui ka neutraalse kõne tempot eraldi fraasi<sup>51</sup> sees olevatel sõnadel ning fraasi viimasel sõnal.

Dispersioonanalüüsi (ANOVA) abil tegin kindlaks, kas kõnetempo erinevused emotsioonipaariti ja võrreldes neutraalse kõnega on üksteisest eristamisel statistiliselt olulised.

#### 7.5.1.4. Intensiivsuse ja põhitooni uurimismaterjal ja -meetod

Emotsionaalse kõne intensiivsuse uurimisele on keskendunud artikkel [P4] ning põhitooni uurimisele artikkel [P5]. Esimeses uuritakse, kas emotsioonid mõjutavad kõneleja hääle intensiivsuse taset ja ulatust ning kas nende parameetrite väärtuste erinevused on emotsioonide eristamisel üksteisest ja neutraalsest kõnest olulised, ning teises, kas emotsioonid mõjutavad kõneleja hääle põhitooni

<sup>49</sup> Kõnetempo uurimismaterjal on esitatud lisa 2.

<sup>50</sup> Kõnetempo mõõtmiseks on uurimustes kasutatud tavaliselt kaht viisi: a) kõnetempot mõõdetakse koos pausidega ja/või b) kõnetempot mõõdetakse ilma pausideta (Braun & Oba, 2007). Käesolevas uurimuses on kõnematerjalist pausid välja jäetud ning mõõdetud on kõneleja artikuleerimiskiirust.

<sup>51</sup> Fraasiks pean siin uurimuses prosoodilist fraasi, mis kestab pausist pausini. Pausiks loetakse sõnade- või lausetevahelist vaikust kõnes, mis kestab vähemalt 30 ms.

kõrgust ja ulatust ning kas nende parameetrite väärtuste erinevused on emotsioonide eristamisel üksteisest ja neutraalsest kõnest olulised.

Intensiivsuse taseme ja ulatuse ning põhitooni kõrguse ja ulatuse mõõtmiseks valisin korpusest välja 329 lauset<sup>52</sup>, milles oli kokku 6409 vokaali (vt tabel 35).

**Tabel 35.** Intensiivsuse ja põhitooni uurimismaterjal (vt [P4, P5])

Emotsioon	Lausete arv	Vokaalide arv lausetes
Rõõm	60	973
Kurbus	87	1807
Viha	79	1435
Neutraalne	103	2194
KOKKU	329	6409

Intensiivsuse taset ja põhitooni kõrgust mõõtsin emotsioonirühmade (rõõmu, kurbuse, viha ja neutraalse kõne) kõigi lausete kõigi vokaalide keskkohast. Seejärel arvutasin iga emotsioonirühma kohta välja intensiivsuse taseme ja ulatuse ning põhitooni kõrguse ja ulatuse miinimumväärtuse, esimese kvartiili, mediaani, kolmanda kvartiili ja maksimumväärtuse.

Lause alguse ja lõpu intensiivsuse taseme ning põhitooni kõrguse arvutamiseks mõõtsin intensiivsuse taset ja põhitooni kõrgust kõigi emotsionaalse ja neutraalse kõne lausete esimese sõna esimesel rõhulisel vokaalil ning lause viimase sõna esimesel rõhulisel vokaalil. Seejärel arvutasin iga emotsioonirühma kohta välja intensiivsuse taseme ja põhitooni kõrguse miinimumväärtuse, esimese kvartiili, mediaani, kolmanda kvartiili ja maksimumväärtuse.

Intensiivsuse ja põhitooni mõõtmisel kasutasin programmi Praat (Boersma & Weenink, 2016) ning kõnetöötlaste süsteemi EMU. Wilcoxon'i astaksummatesti abil tegin kindlaks, kas intensiivsuse taseme ja ulatuse ning põhitooni kõrguse ja ulatuse väärtuste erinevused on emotsioonipaariti ja võrreldes neutraalse kõnega statistiliselt olulised.

### **7.5.2. Emotsionaalse sünteeskõne akustiliste mudelite loomise materjal ja meetod**

Emotsionaalse sünteeskõne akustiliste mudelite loomise eksperimenti kajastab artikkel [P6], mis koosneb kahest osast: (1) emotsioonimudelite loomise aluseid tutvustav teoreetiline osa ning (2) emotsioonimudelite loomise katse.

Kõnesüntesaatori emotsioonimudelite loomisel lähtusime uurimuse kaasautoriga (vt [P6]) eestikeelse emotsionaalse inimkõne akustilise väljendumise

<sup>52</sup> Intensiivsuse taseme ja põhitooni kõrguse uurimismaterjal on esitatud lisas 2.

uurimistulemustest (vt ptk 4.1.). Selleks et leida sobivaim akustiline mudel rõõmu, kurbuse ja viha väljendamiseks eestikeelsel parameetrilisel kõnesünteesil nii mees- kui ka naissünteesihäälele<sup>53</sup>, koostasime esmalt iga põhiemotsiooni kohta kolm katsemudelit (võttes arvesse konkreetse kõnesüntesaatori parameetrilise häälestamise võimalusi). Katsemudelites kasutasime HTS-meetodil loodud sünteesihääli, mida on treenitud neutraalse kõne korpuste baasil. Neutraalse sünteesikõne parameetrite väärtused olid määratud statistilis-parameetrilise süntesaatori (vt Mihkla, Hein, Kalvik, Kiissel, Sirts & Tamuri, 2012) enda kõnemudeligana ning need ei sõltunud emotsionaalse kõne uurimistöö tulemustest neutraalse kõne kohta.

Katsemudelid<sup>54</sup> komponeerisid nelja parameetriga: kõnekiiruse, intensiivsuse taseme, põhitooni kõrguse ja põhitooni ulatusega. Iga katsemudel sisaldas rõõmule, kurbusele ja vihale omast parameetrite väärtuste kombinatsiooni kas optimaalsel tasemel (M2, tugineb inimkõne analüüsi tulemustele), vähendatud kujul (M1, milles on M2s olevaid väärtusi vähendatud 15% emotsioonile omases suunas võrreldes neutraalse kõnega) või võimendatud kujul (M3, milles on M2s olevaid väärtusi suurendatud 15% emotsioonile omases suunas võrreldes neutraalse kõnega) (vt tabel 36).

Tabelis 36 on esitatud emotsioonide parameetrite väärtused neutraalse kõne parameetrite suhtes. Kuna eestikeelse emotsionaalse inimkõne akustiline analüüs oli tehtud ainult naishääle põhjal, siis tuli inimkõne uurimistulemusi katsemudelite jaoks transformeerida.

**Tabel 36.** Eestikeelse emotsionaalse sünteesikõne parameetrilised mudelid (vt [P6])

Parameeter	Neutraalne	Rõõm	Kurbus	Viha
<i>MUDEL 1 (M1)</i>				
Kõnetempo		1,10	0,90	1,24
Intensiivsuse tase		0,90	0,85	0,94
Põhitooni kõrgus		1,50	-3,00	-4,00
Põhitooni ulatus		1,50	-0,90	2,10
<i>MUDEL 2 (M2)</i>				
Kõnetempo	1,00	1,15	0,80	1,40
Intensiivsuse tase	1,00	0,85	0,70	0,90

<sup>53</sup> Mõlemad sünteesihääled on HTS-sünteesihääled, põhinevad peidetud Markovi mudelitel ja on treenitud mahukal kõnematerjalil (keelejuhi ettelõetud tekst). Meeshääle (professionaalne raadio- ja telediktor) treeningkorpus sisaldas ca 2300 lauset, naishääle (näitleja) treeningkorpus sisaldas ca 2000 lauset.

<sup>54</sup> Katsemudelite järgi sünteesitud emotsionaalset kõnet saab kuulata aadressil [https://www.eki.ee/heli/index.php?option=com\\_content&view=article&id=7&Itemid=494](https://www.eki.ee/heli/index.php?option=com_content&view=article&id=7&Itemid=494) (viimati kontrollitud 21.08.2017).

Parameeter	Neutraalne	Rõõm	Kurbus	Viha
Põhitooni kõrgus	0,00	2,50	-4,00	-5,00
Põhitooni ulatus	0,00	2,50	-1,15	2,60
<i>MUDEL 3 (M3)</i>				
Kõnetempo		1,20	0,70	1,56
Intensiivsuse tase		0,80	0,55	0,86
Põhitooni kõrgus		3,50	-5,00	-6,00
Põhitooni ulatus		3,60	-1,40	3,10

Seadsime katsemudelite kohta kaks hüpoteesi. H1: katseisikud tunnevad nii mees- kui ka naissünteeshäälest kõige paremini ära mudeli 3 järgi sünteesitud emotsioonid ning H2: katseisikud tunnevad ära kõnesüntesaatori kõnemudeli järgi sünteesitud neutraalse kõne.

Hüpoteeside kontrollimiseks ja katsemudelite hindamiseks koostasime neli tajutesti: mõlema sünteeshääle kohta kaks testi. Esimene, A-test, koosnes kümnest kolmelauselise neutraalse sisuga sünteeskõnelõigust<sup>55</sup>, mille akustilisi parameetreid oli muudetud vastavalt emotsioonile (kas rõõm, kurbus või viha) ja mudelile (kas M1, M2 või M3). Üks lõik kümnest oli jäetud akustiliselt neutraalseks (parameetrite väärtused andis vaikimisi ette kõnesüntesaator<sup>56</sup>).

Teine, B-test, koosnes samuti kümnest kolmelauselise neutraalse sisuga sünteeskõnelõigust<sup>57</sup>. B-testi puhul algas helilõik akustiliselt neutraalse kõnega, kuid muutus alates lõigu teisest lausest kas rõõmsaks, kurvaks, vihaseks või jäi neutraalseks. Iga helilõigu akustilisi parameetreid oli muudetud vastavalt emotsioonile (kas rõõm, kurbus või viha) ja mudelile (kas M1, M2 või M3). Üks lõik kümnest oli jäetud akustiliselt neutraalseks (parameetrite väärtused andis vaikimisi ette kõnesüntesaator).

Kahe erineva testi, A- ja B-testi koostamise eesmärk oli teada saada, kas emotsiooni muutumine kõnelõigu jooksul aitab kaasa emotsiooni tuvastamisele.

Veebipõhised tajutestid viisin läbi eesti emotsionaalse kõne korpuse keskkonnas (vrd Altrov & Pajupuu, 2012). Katsemudelite hindajatel tuli kuulata 4 x 10 sünteeskõnelõiku (2 x meeshäält (A- ja B-test) ja 2 x naishäält (A- ja B-test)

<sup>55</sup> Lõigu tekst A-testis: „Käisime õhtul restoranis. Tellisime toidu ja jäime ootama. Kui toit meile lauda toodi ja me seda nägime, siis jäime kõik sõnatuks.“

<sup>56</sup> Neutraalsete hääle akustiliste parameetrite keskmised väärtused: meeshäälel põhitoon kõrgus 119 Hz, põhitooni ulatus 78–158 Hz (11,9 pt), intensiivsuse tase 74,5 dB ja kõnetempo 162 sõna minutis; naishäälel põhitoon kõrgus 177 Hz, põhitooni ulatus 122–237 Hz (11,4 pt), intensiivsuse tase 73,6 dB ja kõnetempo 148 sõna minutis.

<sup>57</sup> Lõigu tekst B-testis: „Sain kolmapäeva õhtul ühelt inimeselt telefonikõne. Seda, mis ta rääkis, ei osanud ma oodata. Asjad võivad minna täiesti teistmoodi.“

ning määrata iga kõnelõigu emotsioon või neutraalsus. Iga testi alguses oli esitatud juhend, kuidas testi täita. Katseisikul tuli kuulata helilõiku ning vastata A-testi puhul küsimusele „Milline on lõigu emotsioon?“ ja B-testi puhul „Milline on lõigu lõpuosa emotsioon?“. Vastusevariandid olid valikutena ette antud: rõõm, kurbus, viha, neutraalsus. Kuulaja sai igat lõiku kuulata nii mitu korda, kui ise soovis, ning vajadusel varemvastatut muuta või jätta test pooleli ning jätkata hiljem.

Katsemudelite hindajateks olid kümme meest ja kümme naist vanuses 30 kuni 73 eluaastat (hindajate keskmine vanus oli 43,3 eluaastat). Kõik hindajad olid rahvuselt eestlased ning nende emakeel oli eesti keel. Tajutestid viisin läbi septembris 2015.

Selleks et otsustada, millised akustilised tunnused emotsioonimudelites kõige enam tooni annavad ja milline on nende osakaal õige emotsiooni tuvastamisel, rakendas uurimuse kaasautor M. Mihkla tajutestide tulemustele binaarset logistilist regressioonanalüüsi. Binaarse logistilise regressiooni funktsioonitunnus oli emotsiooni ja neutraalse kõne tuvastamise õigsus (õige vs. vale) ja argument-tunnusteks parameetrilised väärtused.

## 7.6. Tulemused

Peatükk koosneb kahest suuremast osast: 1) emotsionaalse kõne akustiliste parameetrite uurimistulemused (vt ptk 4.1.) ja 2) emotsionaalse sünteeskõne akustiliste mudelite loomise eksperimendi tulemused (vt ptk 4.2.).

### 7.6.1. Emotsionaalse kõne akustiliste parameetrite uurimistulemused

#### 7.6.1.1. Pauside uurimuse tulemused

Pauside uurimuse tulemusi kajastab artikkel [P1].

*Pauside arv ja asukoht.* Uuritavas materjalis sõltus kõigi kolme emotsiooni puhul pauside arv peamiselt kirjavahemärkide arvust: kõige sagedamini tegi lugeja pausi just kirjavahemärgi kohal (vt tabel 37). Lisapausid<sup>58</sup> moodustasid pauside arvust olenevalt emotsioonist 8–10%. Kõigi kolme emotsiooni puhul langesid lisapausid ligi pooltel kordadel sidesõnade *ja/ning* ette ning päris- ja kohanimede ette. Ülejäänud lisapausid olid juhuslikku laadi ja võisid sõltuda kas teksti sisust või lugejast endast (näiteks tema kopsumahust).

---

<sup>58</sup> Lisapaus on paus, mis esineb mujal kui kirjavahemärgi kohal.

**Tabel 37.** Pauside arv ja asukoht (vt [P1])

Emotsioon	Pauside arv	Pausid kirjavahemärkide kohal (% kõigist pausidest)	Lisapausid (% kõigist pausidest)
Rõõm	181	92	8
Kurbus	165	92	8
Viha	195	90	10

Kõigi kolme emotsiooni puhul oli *koma* ainus kirjavahemärk, mille kohal lugeja pausi alati ei teinud. Kurbuselauses tegi lugeja koma kohal pausi 51%-l, vihalauses 67,5%-l ning rõõmulauses 79%-l juhtumitest (vt tabel 38 ja joonis 1 ptk 4.1.1.).

**Tabel 38.** Komad (vt [P1])

Emotsioon	Komade arv	Komad, mille kohal tegi lugeja pausi (% kõigist komadest)
Rõõm	67	79,0
Kurbus	72	51,0
Viha	80	67,5

*Pauside iseloom* (hingamisega vs. hingamiseta pausid). Nii rõõmu-, kurbuse- kui ka vihalõikudes oli hingamispause rohkem kui hingamiseta pause (vt tabel 39). Kõige enam oli hingamispause vihalõikudes: neid oli seal mõnevõrra rohkem kui kurbuse- ja rõõmulõikudes. Hingamisega kirjavahemärgipause oli nii rõõmu-, kurbuse- kui ka vihalõikudes samuti enam-vähem võrdselt. Emotsiooniti oluline erinevus ilmnes hingamisega lisapauside juures, kus lugeja tegi viha puhul märgatavalt rohkem lisapause kui kurbuse või rõõmu puhul. Ka rõõmu- ja kurbuse-emotsiooni omavaheline erinevus oli seal märkimisväärne.

**Tabel 39.** Hingamispausid (vt [P1])

Emotsioon	Hingamis- pauside arv	Hingamis- pausid (% kõigist pausidest)	Hingamisega kirjavahemärgipausid (% kõigist kirjavahemärgipausidest)	Hingamisega lisapausid (% kõigist lisapausidest)
Rõõm	96	53,0	56,0	21,0
Kurbus	90	54,5	57,0	31,0
Viha	115	59,0	61,0	55,0

*Kirjavahemärkide ja hingamispauside vaheline seos.* Lause lõpus tegi lugeja kirjavahemärgi kohal hingamispause enam kurbuse- ja rõõmulõikudes (vastavalt 68% ja 67%), vähem vihalõikudes (58%) (vt tabel 40). Koma kohal tegi

lugeja hingamispausi kõige sagedamini vihalõikudes (67%), kõige harvem kurbuselõikudes (27%). Muude kirjavahemärkide (st v.a lause lõpp ja koma) kohal tegi lugeja kõige enam hingamispause kurbuselõikudes ja kõige vähem rõõmulõikudes.

**Tabel 40.** Hingamispauside osakaal kirjavahemärkide kohal (vt [P1])

Emotsioon	Lause lõpp (% lauselõpupausidest)	Koma (% komapausidest)	Muud kirjavahemärgid (% muudest kirjavahemärgipausidest)
Rõõm	67	36	40
Kurbus	68	27	60
Viha	58	67	56

*Pauside kestus.* Pauside kestuse mõõtmine näitas, et pausirühmade üldkokkuvõttes olid kõige pikemad pausid kurbusel ja lühimad vihal (vt allpool tabel 42). Kõigil kolmel emotsioonil olid kestuste keskmiste järgi<sup>59</sup> pikimad pausid punktitaunid (kurbusel ja rõõmul hingamisega punktitaunid, vihal hingamiseta punktitaunid) ning lühimad hingamiseta lisa- ja komapausid.

Emotsiooniti statistiliselt oluliselt erinevaks<sup>60</sup> osutusid hingamiseta ja hingamisega punktitaunid ning hingamisega hüüümärgipausid.

- Hingamiseta punktitaun
  1. kurbus (944 ms) vs. rõõm (638 ms) ( $p = 0,010$ )
  2. kurbus (944 ms) vs. viha (691 ms) ( $p = 0,006$ )
- Hingamisega punktitaun
  1. kurbus (969 ms) vs. rõõm (774 ms) ( $p = 0,001$ )
  2. kurbus (969 ms) vs. viha (664 ms) ( $p = 0,001$ )
  3. rõõm (774 ms) vs. viha (664 ms) ( $p = 0,003$ )
- Hingamisega hüüümärgipausid
  1. kurbus (911 ms) vs. viha (625 ms) ( $p = 0,017$ )

Ülejäänud pausigrupid kestuse järgi emotsiooniti oluliselt ei erinenud.

*Tajukatse tulemused.* Tajukatse näitas, et kõnelõigu emotsiooni ainult pausi-erinevuste põhjal ära ei tuntud (vt tabel 41). Lihtsama pausimustriga lõike (kus oli muudetud ainult lauselõpupauside kestusi) pidasid hindajad neutraalseks. Samuti hindasid nad neutraalseks keerulisema struktuuriga lõigud (kus oli muudetud nii pauside kestust, asukohta kui ka iseloomu).

<sup>59</sup> Arvestatud ei ole nende pausirühmade andmeid, mille esinemissagedus jäi alla kolme juhtumi.

<sup>60</sup> Kasutati Student's t-Testi. Püstitati 0-hüpotees: kahe emotsiooni keskmiste vahe on null ( $p > 0,05$ ).

**Tabel 41.** Tajukatse tulemused (tuvastusprotsendid) (vt [P1])

Lihtsama pausimuustriga lõigud				
<i>Lõigu emotsioon</i>	<i>Vastusevariandid</i>			
	Rõõm	Kurbus	Viha	Neutraalne
Rõõm	0	0	10	90
Kurbus	0	10	20	70
Viha	10	10	10	70
Neutraalne	0	10	20	70
Keerulisema pausimuustriga lõigud				
<i>Lõigu emotsioon</i>	<i>Vastusevariandid</i>			
	Rõõm	Kurbus	Viha	Neutraalne
Rõõm	10	10	10	70
Kurbus	10	0	20	70
Viha	10	0	10	80
Neutraalne	10	0	10	80

**Tabel 42.** Pausirühmade keskmised kestused millisekundites (vt [P1])

Pausi tüüp	Emot-sioon	Arv	Min	Q1	Mediaan	Keskmine	Q3	Max
<i>Lisapaus</i>	Rõõm	11	34	103	117	128	127	359
	Kurbus	9	54	91	111	122	133	251
	Viha	9	91	119	133	141	145	267
<i>Hingamisega lisapaus</i>	Rõõm	3	365	416	467	448	490	513
	Kurbus	4	300	366	436	417	488	497
	Viha	11	289	326	399	386	412	555
<i>Komapaus</i>	Rõõm	34	52	137	222	252	339	636
	Kurbus	28	43	113	203	280	379	795
	Viha	21	75	145	207	225	299	428
<i>Hingamisega komapaus</i>	Rõõm	19	280	407	440	480	538	724
	Kurbus	9	296	394	531	511	615	721
	Viha	33	205	337	408	412	449	732
<i>Koolonipaus</i>	Rõõm	2	139	181	222	222	264	306
	Kurbus	0	–	–	–	–	–	–
	Viha	1	364	364	364	364	364	364
<i>Hingamisega koolonipaus</i>	Rõõm	1	438	438	438	438	438	438
	Kurbus	1	777	777	777	777	777	777
	Viha	2	296	366	437	437	507	578
<i>Mõttekriipsu-paus</i>	Rõõm	4	393	526	606	628	708	906
	Kurbus	2	837	838	839	839	841	842
	Viha	5	236	292	525	441	545	606
<i>Hingamisega mõttekriipsu-paus</i>	Rõõm	3	373	537	702	604	719	736
	Kurbus	5	533	987	989	936	1033	1138
	Viha	8	316	375	448	526	693	865
<i>Mõttepunktide paus</i>	Rõõm	1	724	724	724	724	724	724
	Kurbus	5	407	412	786	742	972	1133
	Viha	2	878	878	879	879	879	880



Pausi tüüp	Emot-sioon	Arv	Min	Q1	Mediaan	Keskmine	Q3	Max
<i>Hingamisega mõttepunktide paus</i>	Rõõm	1	474	474	474	474	474	474
	Kurbus	8	578	883	934	912	987	1085
	Viha	0	–	–	–	–	–	–
<i>Punktipaus</i>	Rõõm	25	257	524	594	638	710	1137
	Kurbus	30	211	638	913	944	1156	2136
	Viha	34	426	534	689	691	805	1006
<i>Hingamisega punktipaus</i>	Rõõm	53	413	637	772	774	882	1364
	Kurbus	54	528	757	884	969	1076	2366
	Viha	43	353	542	667	664	746	1016
<i>Hüüumärgi-paus</i>	Rõõm	7	256	445	540	571	708	893
	Kurbus	0	–	–	–	–	–	–
	Viha	6	384	626	635	624	655	802
<i>Hingamisega hüüumärgi-paus</i>	Rõõm	16	383	526	705	696	843	996
	Kurbus	6	709	813	835	911	955	1285
	Viha	13	341	600	644	625	706	719
<i>Küsimärgi-paus</i>	Rõõm	1	687	687	687	687	687	687
	Kurbus	1	941	941	941	941	941	941
	Viha	1	532	532	532	532	532	532
<i>Hingamisega küsimärgi-paus</i>	Rõõm	0	–	–	–	–	–	–
	Kurbus	2	479	602	724	724	847	970
	Viha	1	623	623	623	623	623	623
<i>Küsi-hüüu-märgipaus</i>	Rõõm	0	–	–	–	–	–	–
	Kurbus	0	–	–	–	–	–	–
	Viha	1	652	652	652	652	652	652
<i>Hingamisega küsi-hüüu-märgipaus</i>	Rõõm	0	–	–	–	–	–	–
	Kurbus	1	1260	1260	1260	1260	1260	1260
	Viha	4	608	619	638	653	671	727

### 7.6.1.2. Formantsageduste ja artikulatsiooni täpsuse uurimuse tulemused

Formantsageduste uurimuse tulemusi käsitleb artikkel [P2].

*Emotsioonide mõju vokaalide a, i ja u esimese formantsageduse väärtustele.* Vokaali *a* puhul erinesid neutraalsest kõnest oluliselt kurbuse- ja viha-emotsioon, mõlemal oli  $F_1$  väärtuste keskmine madalam kui neutraalsel kõnel. Emotsioonipaariti erinesid vokaali *a*  $F_1$  väärtuste keskmiste järgi üksteisest oluliselt kõik emotsioonid (vt tabel 43 ja tabel 44). Vokaali *i* puhul erinesid neutraalsest kõnest oluliselt kõik kolm emotsiooni, kusjuures nii rõõmu-, kurbuse- kui ka vihalause vokaalide  $F_1$  väärtuste keskmine oli madalam kui neutraalsel kõnel. Emotsioonipaariti erines vokaali *i* puhul oluliselt rõõm kurbusest (rõõmul oli kõrgem  $F_1$ ) ja viha kurbusest (vihal oli kõrgem  $F_1$ ). Vokaali *u* puhul ei erinenud ühegi emotsiooni vokaalide  $F_1$  väärtuste keskmine oluliselt ei neutraalse kõne ega mõne teise emotsiooni vokaalide  $F_1$  väärtuste keskmisest.

*Emotsioonide mõju vokaalide a, i ja u teise formantsageduse väärtustele.* Kõigi kolme emotsiooni ja neutraalse kõne vokaalide  $F_2$  väärtuste keskmiste erinevused olid väikesed (vokaali *i* puhul erinevused puudusid) (vt tabel 43)

ning need erinevused ei olnud ka statistiliselt olulised (vt tabel 45). Ka emotsioonipaariti olid  $F_2$  väärtuste keskmiste erinevused väikesed ning statistiliselt ebaolulised. Ainuke erand oli vokaal  $a$ , mille puhul erines viha-emotsiooni vokaalide  $F_2$  oluliselt kurbuselausete vokaalide  $F_2$ -st Hertzi skaalal (kurbusel oli kõrgem  $F_2$ ).

**Tabel 43.** Vokaalide  $a$ ,  $i$  ja  $u$  esimese ja teise formantsageduse keskmised (Hertzi skaala / Barki skaala) emotsiooniti (vt [P2])

Vokaal	Emotsioon	$F_1$	$F_2$
$a$	Rõõm	644 / 6,1	1440 / 10,8
	Kurbus	558 / 5,4	1460 / 10,9
	Viha	615 / 5,9	1412 / 10,7
	Neutraalne	635 / 6,0	1448 / 10,9
$i$	Rõõm	382 / 3,8	2658 / 14,9
	Kurbus	359 / 3,6	2642 / 14,9
	Viha	386 / 3,9	2642 / 14,9
	Neutraalne	403 / 4,0	2660 / 14,9
$u$	Rõõm	418 / 4,2	1272 / 10,0
	Kurbus	410 / 4,1	1334 / 10,3
	Viha	427 / 4,3	1298 / 10,1
	Neutraalne	430 / 4,3	1274 / 10,0

**Tabel 44.** Dispersioonanalüüsi (ANOVA) tulemused emotsionaalse ja neutraalse kõne lausete lühikeste vokaalide  $a$ ,  $i$  ja  $u$  esimese formantsageduse ( $F_1$ ) kohta (vt [P2])

Vokaal	Emotsioonipaar	$Pr > F$ Hertzi skaala	$Pr > F$ Barki skaala
$a$	viha vs. rõõm	0,010**	0,030*
	viha vs. neutraalne	0,050*	0,040*
	viha vs. kurbus	0,001***	0,010**
	rõõm vs. neutraalne	0,430	0,690
	rõõm vs. kurbus	0,001***	0,010**
	neutraalne vs. kurbus	0,001***	0,010**
$i$	viha vs. rõõm	0,500	0,440
	viha vs. neutraalne	0,010**	0,020*
	viha vs. kurbus	0,001***	0,010**
	rõõm vs. neutraalne	0,010**	0,010**
	rõõm vs. kurbus	0,001***	0,010**
	neutraalne vs. kurbus	0,001***	0,010**
$u$	viha vs. rõõm	0,500	0,390
	viha vs. neutraalne	0,740	0,780
	viha vs. kurbus	0,150	0,140
	rõõm vs. neutraalne	0,380	0,310
	rõõm vs. kurbus	0,540	0,600
	neutraalne vs. kurbus	0,110	0,110

*Märkus.* Statistiliselt olulised erinevused on tähistatud tärniga: \* $p < 0,05$ , \*\* $p < 0,01$ , \*\*\* $p < 0,001$ .

**Tabel 45.** Dispersioonanalüüsi (ANOVA) tulemused emotsionaalse ja neutraalse kõne lausete lühikeste vokaalide *a*, *i* ja *u* teise formantsageduse ( $F_2$ ) kohta (vt [P2])

Vokaal	Emotsioonipaar	Pr > $F$ Hertzi skaala	Pr > $F$ Barki skaala
<i>a</i>	viha vs. rõõm	0,189	0,280
	viha vs. neutraalne	0,055	0,050
	viha vs. kurbus	0,036*	0,070
	rõõm vs. neutraalne	0,720	0,490
	rõõm vs. kurbus	0,460	0,500
	neutraalne vs. kurbus	0,618	0,930
<i>i</i>	viha vs. rõõm	0,793	0,840
	viha vs. neutraalne	0,667	0,630
	viha vs. kurbus	0,661	0,400
	rõõm vs. neutraalne	0,895	0,810
	rõõm vs. kurbus	0,575	0,400
	neutraalne vs. kurbus	0,490	0,300
<i>u</i>	viha vs. rõõm	0,402	0,380
	viha vs. neutraalne	0,476	0,400
	viha vs. kurbus	0,286	0,320
	rõõm vs. neutraalne	0,962	0,960
	rõõm vs. kurbus	0,072	0,070
	neutraalne vs. kurbus	0,112	0,090

*Märkus.* Statistiliselt olulised erinevused on tähistatud tärniga: \* $p < 0,05$ , \*\* $p < 0,01$ , \*\*\* $p < 0,001$ .

*Emotsioonide mõju kõneleja artikulatsiooni täpsusele.* Uurimistulemustest selgus, et kui üks erand välja arvata (vt allpool), siis emotsionaalse ja neutraalse kõne vokaalide *a*, *i* ja *u* artikulatsiooni täpsus märkimisväärselt ei erinenud (vt tabel 46). Ka emotsioonipaariti olid artikulatsiooni täpsuse erinevused väikesed. Erand oli kurbuse-emotsiooni vokaal *a*, mis oli võrreldes neutraalse kõne ja teiste emotsioonidega tsentraalsele häälikule *x* märkimisväärselt lähemal: kurbuse-emotsiooni *a* artikulatsiooni täpsus oli muutunud ning vokaal oma kvaliteeti kaotanud.

**Tabel 46.** Eukleidiline kaugus vokaalide ja neutraalse hääliku *x* vahel (Hertsi skaala / Barki skaala) (vt [P2])

Vokaal	<i>x</i> – rõõm	<i>x</i> – kurbus	<i>x</i> – viha	<i>x</i> – neutraalne
<i>a</i>	332,7 / 1,5	288,8 / 0,9	348,9 / 1,4	321,8 / 1,4
<i>i</i>	922,1 / 3,4	909,8 / 3,4	915,6 / 3,3	921,3 / 3,3
<i>u</i>	482,1 / 1,8	423,2 / 1,6	454,8 / 1,7	477,9 / 1,8

### 7.6.1.3. Kõnetempo uurimuse tulemused

Kõnetempo uurimuse tulemusi kajastab artikkel [P3].

Kõige kiirem *üldine kõnetempo* oli vihalausetes (17,5 h/s) ja kõige aeglasem kurbuselusetes (16,6 h/s) (vt tabel 47 ja joonis 2 ptk 4.1.3.). Sama kehtib *fraasisisese kõnetempo* kohta (viha 18,4 h/s; kurbus 17 h/s). *Fraasi viimast sõna* artikleeris lugeja kõige kiiremini viha- ja kurbuselusetes (14,4 h/s) ning kõige aeglasemini rõõmulausetes (14 h/s).

**Tabel 47.** Emotsionaalse ja neutraalse kõne tempo keskmised (häälikut sekundis) (vt [P3])

Emotsioon	Üldine tempo	Fraasisisene tempo	Fraasi viimase sõna tempo
Rõõm	17,1	17,6	14,0
Kurbus	16,6	17,0	14,4
Viha	17,5	18,4	14,4
Neutraalne	16,9	17,5	14,1

Emotsionaalse kõne üldine tempo erines neutraalse kõne tempost oluliselt vaid vihalausetes (vt tabel 48). Emotsioonipaariti olid kõik erinevused statistiliselt olulised. Fraasisisese kõnetempo puhul erinesid neutraalsest kõnest oluliselt viha- ja kurbuse-emotsioon. Samuti olid erinevused olulised ka emotsioonipaariti. Fraasi viimase sõna kestuse erinevused emotsionaalses ja neutraalses kõnes statistiliselt olulised ei olnud, samuti polnud erinevused olulised emotsioonipaariti.

**Tabel 48.** Dispersioonanalüüsi (ANOVA) tulemused kõnetempo erinevuste kohta emotsioonipaariti (vt [P3])

Emotsioonipaar	Tõenäosusväärtus		
	<i>üldine tempo</i>	<i>fraasisisene tempo</i>	<i>fraasi viimase sõna tempo</i>
Rõõm vs. viha	0,010**	0,043*	0,452
Rõõm vs. kurbus	0,039*	0,036*	0,317
Rõõm vs. neutraalne	0,557	0,810	0,985
Kurbus vs. viha	0,001***	0,001***	0,837
Kurbus vs. neutraalne	0,107	0,033*	0,237
Viha vs. neutraalne	0,031*	0,008**	0,370

*Märkus.* Statistiliselt olulised erinevused on tähistatud tärniga: \* $p < 0,05$ , \*\* $p < 0,01$ , \*\*\* $p < 0,001$ .

#### 7.6.1.4. Intensiivsuse uurimuse tulemused

Intensiivsuse uurimuse tulemusi käsitleb artikkel [P4].

*Intensiivsuse tase emotsionaalses ja neutraalses kõnes.* Uurimistulemustest selgus, et kõige kõrgem oli intensiivsuse tase neutraalsel kõnel (mediaan 71,6 dB) ja kõige madalam kurbuse-emotsioonil (mediaan 70,3 dB) (vt tabel 49 ja joonis 3 ptk 4.1.4.).

**Tabel 49.** Intensiivsuse tase (dB) emotsionaalses ja neutraalses kõnes (vt [P4])

Emotsioon	Miinumum	Q1	Mediaan	Q3	Maksimum
Rõõm	60,0	67,7	70,6	73,1	81,1
Kurbus	57,9	66,9	70,3	73,1	82,2
Viha	58,9	67,9	71,1	74,0	82,9
Neutraalne	61,9	69,2	71,6	74,1	81,3

Mõõtmistulemuste statistiline analüüs näitas, et emotsioonipaariti ja võrreldes neutraalse kõnega olid kõik intensiivsuse taseme erinevused olulised (vt tabel 50).

**Tabel 50.** Wilcoxon'i astaksummatesti tulemused intensiivsuse taseme kohta emotsioonipaariti ja võrdluses neutraalse kõnega (vt [P4])

Emotsioon	Tõenäosusväärtus
Rõõm vs. viha	0,008**
Rõõm vs. kurbus	0,020*
Rõõm vs. neutraalne	0,001***
Kurbus vs. viha	0,001***
Kurbus vs. neutraalne	0,001***
Viha vs. neutraalne	0,001***

*Märkus.* Statistiliselt olulised erinevused on tähistatud tärniga: \* $p < 0,05$ , \*\* $p < 0,01$ , \*\*\* $p < 0,001$ .

*Intensiivsuse ulatus emotsionaalses ja neutraalses kõnes.* Uurimistulemused näitasid, et kõige laiem oli intensiivsuse ulatus kurbuse-emotsioonil (mediaan 14,7 dB) ja kõige kitsam rõõmu-emotsioonil (mediaan 13,2 dB) (vt tabel 51).

**Tabel 51.** Intensiivsuse ulatus (dB) emotsionaalses ja neutraalses kõnes (vt [P4])

Emotsioon	Miinumum	Q1	Mediaan	Q3	Maksimum
Rõõm	5,6	11,0	13,2	17,2	25,9
Kurbus	4,2	11,7	14,7	19,7	30,8
Viha	6,6	12,5	14,3	18,5	25,8
Neutraalne	6,5	12,2	13,7	17,6	24,9

Mõõtmistulemuste statistiline analüüs näitas, et intensiivsuse ulatus ei erinenud oluliselt ei emotsioonipaariti ega ka emotsioonide võrdluses neutraalse kõnega (vt tabel 52).

**Tabel 52.** Wilcoxon'i astaksummatesti tulemused intensiivsuse ulatuse kohta emotsioonipaariti ja võrdluses neutraalse kõnega (vt [P4])

Emotsioonipaar	Tõenäosusväärtus
Rõõm vs. viha	0,280
Rõõm vs. kurbus	0,250
Rõõm vs. neutraalne	0,800
Kurbus vs. viha	0,930
Kurbus vs. neutraalne	0,800
Viha vs. neutraalne	0,800

*Märkus.* Statistiliselt olulised erinevused on tähistatud tärniga: \* $p < 0,05$ , \*\* $p < 0,01$ , \*\*\* $p < 0,001$ .

*Lause alguse ja lõpu intensiivsuse tase emotsionaalses ja neutraalses kõnes.* Uurimistulemustest selgus, et nii lause alguse kui ka lõpu intensiivsuse tase oli kõige kõrgem neutraalsel kõnel (mediaan 74,7/65,3 dB) ja kõige madalam kurbuse-emotsioonil (mediaan 72,3/63,6 dB) (vt tabel 53).

**Tabel 53.** Lause alguse ja lõpu intensiivsuse tase (dB) emotsionaalses ja neutraalses kõnes (vt [P4])

Emotsioon	Lause algus / lause lõpp				
	Miinumum	Q1	Mediaan	Q3	Maksimum
Rõõm	68,2 / 59,1	71,6 / 63,0	73,9 / 64,5	75,8 / 66,3	81,5 / 69,1
Kurbus	64,3 / 54,7	70,0 / 61,1	72,3 / 63,6	74,9 / 65,5	81,8 / 71,0
Viha	65,6 / 58,2	70,8 / 62,6	73,8 / 64,9	76,7 / 67,1	82,7 / 73,7
Neutraalne	65,3 / 56,6	72,0 / 62,9	74,7 / 65,3	76,4 / 67,5	81,1 / 73,2

Mõõtmistulemuste statistiline analüüs näitas, et nii lause alguse kui ka lõpu intensiivsuse tase erines oluliselt vaid paaris kurbus vs. neutraalne kõne (vt tabel 54). Teiste paaride puhul ei olnud lause alguse ega lõpu intensiivsuse taseme erinevused statistiliselt olulised.

**Tabel 54.** Wilcoxon'i astaksummatesti tulemused emotsionaalse ja neutraalse kõne intensiivsuse taseme kohta lause alguses ja lõpus (vt [P4])

Emotsioonipaar	Lause algus (tõenäosusväärtus)	Lause lõpp (tõenäosusväärtus)
Rõõm vs. viha	0,917	0,932
Rõõm vs. kurbus	0,091	0,112
Rõõm vs. neutraalne	0,532	0,512
Kurbus vs. viha	0,053	0,053
Kurbus vs. neutraalne	0,002**	0,002**
Viha vs. neutraalne	0,917	0,932

*Märkus.* Statistiliselt olulised erinevused on tähistatud tärniga: \* $p < 0,05$ , \*\* $p < 0,01$ , \*\*\* $p < 0,001$ .

### 7.6.1.5. Põhitooni uurimuse tulemused

Põhitooni uurimuse tulemusi kajastab artikkel [P5].

*Põhitooni kõrgus emotsionaalses ja neutraalses kõnes.* Uurimistulemused näitasid, et põhitooni oli kõrgeim rõõmlausetes (mediaan 187,7 Hz) ja madalaim vihalausetes (mediaan 180,8 Hz), neutraalses kõnes ja kurbuselausetes oli põhitooni kõrgus ühesugune (mediaan 185,4 Hz) (vt tabel 55 ja joonis 4 ptk 4.1.5.).

**Tabel 55.** Põhitooni kõrgus (Hz) emotsionaalses ja neutraalses kõnes (vt [P5])

Emotsioon	Miinumum	Q1	Mediaan	Q3	Maksimum
Rõõm	133,3	171,3	187,7	214,8	279,5
Kurbus	139,4	172,6	185,4	206,6	257,1
Viha	134,5	165,7	180,8	208,0	270,4
Neutraalne	120,7	171,9	185,4	208,5	263,3

Mõõtmistulemustele tehtud statistiline analüüs näitas, et põhitooni kõrguse erinevused olid olulised vaid viha-emotsiooni puhul, mille põhitooni oli võrreldes teiste emotsioonide ja neutraalse kõnega oluliselt madalam (vt tabel 56). Teiste emotsioonide erinevused omavahel ja võrdluses neutraalse kõnega olulised ei olnud.

**Tabel 56.** Wilcoxon'i astaksummatesti tulemused põhitooni kõrguse kohta emotsioonipaariti ja võrdluses neutraalse kõnega (vt [P5])

Emotsioonipaar	Tõenäosusväärtus
Rõõm vs. viha	0,001***
Rõõm vs. kurbus	0,240
Rõõm vs. neutraalne	0,350
Kurbus vs. viha	0,001***
Kurbus vs. neutraalne	0,610
Viha vs. neutraalne	0,001***

*Märkus.* Statistiliselt olulised erinevused on tähistatud tärniga: \* $p < 0,05$ , \*\* $p < 0,01$ , \*\*\* $p < 0,001$ .

*Põhitooni ulatus emotsionaalses ja neutraalses kõnes.* Uurimistulemustest selgus, et põhitooni ulatus oli kõige laiem vihalausetes (mediaan 105,1 Hz) ning kõige kitsam kurbuselausetes (mediaan 89,1 Hz) (vt tabel 57 ja joonis 5 ptk 4.1.5.).

**Tabel 57.** Põhitooni ulatus (Hz) emotsionaalses ja neutraalses kõnes (vt [P5])

Emotsioon	Miinumum	Q1	Mediaan	Q3	Maksimum
Rõõm	42,9	84,0	105,0	124,2	181,0
Kurbus	39,3	75,4	89,1	102,7	142,3
Viha	51,4	91,1	105,1	124,6	166,4
Neutraalne	45,2	81,7	94,8	107,0	135,3

Põhitooni ulatuse mõõtmistulemustele tehtud statistiline analüüs näitas, et kurbuse-emotsiooni põhitooni ulatus erines oluliselt rõõmu- ja viha-emotsiooni põhitooni ulatusest ning viha-emotsiooni põhitooni ulatus erines oluliselt neutraalse kõne põhitooni ulatusest (vt tabel 58).

**Tabel 58.** Wilcoxon'i astaksummatesti tulemused põhitooni ulatuse kohta emotsiooni-paariti ja võrdluses neutraalse kõnega (vt [P5])

Emotsioon	Tõenäosusväärtus
Rõõm vs. viha	0,519
Rõõm vs. kurbus	0,009**
Rõõm vs. neutraalne	0,179
Kurbus vs. viha	0,001***
Kurbus vs. neutraalne	0,109
Viha vs. neutraalne	0,013*

*Märkus.* Statistiliselt olulised erinevused on tähistatud tärniga: \* $p < 0,05$ , \*\* $p < 0,01$ , \*\*\* $p < 0,001$ .

*Lause alguse ja lõpu põhitooni kõrgus emotsionaalses ja neutraalses kõnes.* Uurimistulemused näitasid, et lause alguses oli põhitoon kõrgeim neutraalses kõnes (mediaan 231,6 Hz) ning madalaim kurbuselausetes (mediaan 213,1 Hz). Lause lõpus oli põhitoon kõrgeim samuti neutraalses kõnes (mediaan 164,9 Hz), kuid madalaim vihalausetes (mediaan 161,4 Hz). Rõõmu- ja kurbuselausetes oli põhitooni kõrgus lause lõpus ühesugune (mediaan 164,4 Hz) (vt tabel 59).



**Tabel 59.** Lause alguse ja lõpu põhitooni kõrgus (Hz) emotsionaalses ja neutraalses kõnes (vt [P5])

Emotsioon	Lause algus / lause lõpp				
	Miinumum	Q1	Mediaan	Q3	Maksimum
Rõõm	163,2 / 141,1	201,2 / 158,4	225,8 / 164,4	252,5 / 171,7	314,3 / 187,7
Kurbus	142,0 / 143,9	197,3 / 157,2	213,1 / 164,4	234,8 / 173,1	283,9 / 192,2
Viha	142,3 / 145,3	198,6 / 156,6	214,3 / 161,4	245,0 / 165,9	313,1 / 177,1
Neutraalne	163,1 / 139,4	203,7 / 157,1	231,6 / 164,9	249,3 / 170,7	284,7 / 186,8

Mõõtmistulemustele tehtud statistiline analüüs (Wilcoxon'i astaksummatest) näitas, et lause alguse ja lõpu põhitooni kõrguse erinevused emotsiooniti ja võrdluses neutraalse kõnega olulised ei olnud.

## 7.6.2. Emotsionaalse sünteeskõne akustiliste mudelite loomise eksperimendi tulemused

Emotsionaalse kõne akustiliste mudelite katsemudelite hindamise tulemusi käsitleb artikkel [P6].

### 7.6.2.1. Emotsionaalse sünteeskõne katsemudelite hindamise

Tabelis 60 on esitatud hindajate vastuste jagunemine kolme katsemudeli järgi modelleeritud emotsioonide tajumise kohta meessünteeshäälest. Emotsioon on loetud õigesti tajutuks ainult siis, kui selle tuvastusprotsent on ületanud juhusliku valiku tõenäosuse ehk 25% ning on võrreldes teiste emotsioonide tuvastusprotsentidest suurem.

*Rõõmu-emotsioon.* Tajutestide tulemustest selgus, et mudeli 1 järgi modelleeritud rõõmu-emotsiooni hindajad meessünteeshäälest õigesti ära ei tundnud. Nii A- kui ka B-testis pidasid nad seda kõige enam kurbuseks. Mudeli 2 järgi modelleeritud rõõmu-emotsiooni pidasid hindajad A-testis kas vihaks või kurbuseks ja B-testis kurbuseks. Mudeli 3 järgi modelleeritud rõõmu-emotsioon tunti õigesti ära B-testis (55% vastanutest), A-testis peeti seda kurbuseks (vt tabel 60).

*Kurbuse-emotsioon.* Tajutestide tulemused näitasid, et mudeli 1 järgi modelleeritud kurbuse-emotsiooni hindajad meessünteeshäälest ära ei tundnud. Nii A- kui ka B-testis peeti seda kõige sagedamini neutraalseks kõneks. Mudeli 2 järgi modelleeritud kurbuse-emotsiooni tundsid hindajad ära B-testis (50% vastanutest), A-testis peeti seda neutraalseks kõneks. Mudeli 3 põhjal tunti kurbuse-emotsioon õigesti ära samuti vaid B-testis (80% vastanutest), A-testis peeti seda neutraalseks kõneks (vt tabel 60).

*Viha-emotsioon.* Tajutestide tulemustest nähtus, et mudeli 1 järgi modelleeritud viha-emotsiooni hindajad meessünteeshäälest õigesti ära ei tundnud. Nii

A- kui ka B-testis peeti seda neutraalseks kõneks. Mudeli 2 puhul ei tundnud hindajad samuti viha-emotsiooni ära. Mõlemas testis peeti seda kõige enam neutraalseks kõneks. Mudeli 3 järgi modelleeritud viha-emotsiooni tundsid hindajad meessüntheeshäälest ära: A-testis 60% ja B-testis 50% vastanutest (vt tabel 60).

*Neutraalne kõne.* Meessüntheeshääle neutraalne kõne oli sünteesitud kõne-süntesaatori olemasoleva mudeli järgi. Tajutestide tulemustest selgus, et nii A- kui ka B-testi puhul ei tundnud hindajad meessüntheeshääles neutraalset kõnet ära. Mõlemas testis pidasid hindajad seda kõige sagedamini kurbuseks (vt tabel 60).

**Tabel 60.** Eksimismaatriks (meessüntheeshääle, A-/B-test<sup>61</sup>). Emotsioonide tajumine meessüntheeshäälest kolme erineva katsemudeli põhjal, vastused protsentides (vt [P6])

Sihtemotsioon	Vastusevariandid			
	<i>Rõõm</i>	<i>Kurbus</i>	<i>Viha</i>	<i>Neutraalne</i>
<b>Mudel 1</b>				
Rõõm	10/30	45/40	15/10	30/20
Kurbus	10/5	30/15	5/0	55/80
Viha	15/20	10/15	25/10	50/55
<b>Mudel 2</b>				
Rõõm	5/10	35/70	35/10	25/10
Kurbus	0/5	30/50	0/0	70/45
Viha	0/10	10/0	35/40	55/50
<b>Mudel 3</b>				
Rõõm	25/55	35/25	25/15	15/5
Kurbus	0/0	35/80	0/0	65/20
Viha	5/5	10/0	60/50	25/45
<b>Neutraalne</b>	10/5	65/55	10/5	15/35

*Märkus.* Poolpaksus kirjas on märgitud õigesti tajutud emotsioonid.

*Tajutestide tulemuste kokkuvõtte meessüntheeshääle kohta.* Tajutestide tulemustest selgus, et hindajad tundsid meessüntheeshäälest kõige paremini ära M3 järgi sünteesitud emotsioonid. Neutraalsust õigesti ära ei tuntud, seda peeti kurbuseks. Kui võrrelda A- ja B-testi tulemusi, siis meessüntheeshääle puhul tunti rõõm ja kurbus paremini ära B-testis, viha A-testis.

<sup>61</sup> A-testis kandis terve helilõik ühte emotsiooni, B-testis algas helilõik akustiliselt neutraalse kõnega ja muutus alates lõigu teisest lausest kas rõõmsaks, kurvaks, vihaseks või jäi neutraalseks.

### 7.6.2.2. Emotsionaalse sünteeskõne katsemudelite hindamise tulemused naissünteeshääle põhjal

Tabelis 61 on ära toodud hindajate vastuste jagunemine kolme katsemudeli järgi modelleeritud emotsioonide tajumise kohta naissünteeshäälest.

*Rõõmu-emotsioon.* Tajutestide tulemused näitasid, et mudeli 1 järgi modelleeritud rõõmu-emotsiooni tundsid hindajad õigesti ära A-testis (40% vastanutest) (vt tabel 61). B-testis rõõmu ära ei tundud, seda peeti kurbuseks. Mudeli 2 puhul hindajad rõõmu-emotsiooni naissünteeshäälest ära ei tundnud: nii A- kui ka B-testis pidasid nad seda kõige enam kurbuseks. Mudeli 3 järgi modelleeritud rõõmu-emotsiooni tundsid hindajad õigesti ära A-testis (30% vastanutest). B-testis pidasid nad seda kõige sagedamini kurbuseks.

*Kurbuse-emotsioon.* Tajutestide tulemustest selgus, et mudeli 1 järgi modelleeritud kurbuse-emotsiooni hindajad naissünteeshäälest ära ei tundnud, nad pidasid seda nii A- kui ka B-testis eeskätt neutraalseks kõneks (vt tabel 61). Mudeli 2 puhul tunti kurbuse-emotsioon õigesti ära A-testis (60% vastanutest), B-testis peeti seda pigem neutraalseks kõneks. Mudeli 3 järgi modelleeritud kurbuse-emotsiooni tundsid hindajad õigesti ära nii A- kui ka B-testis (vastavalt 50% ja 75% vastanutest).

*Viha-emotsioon.* Tajutestide tulemustest selgus, et mudeli 1 järgi modelleeritud viha-emotsiooni hindajad naissünteeshäälest ära ei tundnud: A-testis pidasid nad seda rõõmuks ja B-testis neutraalseks kõneks (vt tabel 61). Mudeli 2 puhul tunti viha-emotsioon õigesti ära nii A- kui ka B-testis (vastavalt 45% ja 50% vastanutest). Ka mudeli 3 järgi modelleeritud viha-emotsiooni tundsid hindajad ära mõlemas testis: A-testis 65% ja B-testis 60% vastanutest.

*Neutraalse kõne.* Naissünteeshääle neutraalne kõne oli sünteesitud kõne-süntesaatori olemasoleva mudeli järgi. Tajutestide tulemused näitasid, et nii A- kui ka B-testis tundsid hindajad neutraalse kõne õigesti ära (vastavalt 50% ja 40% vastanutest) (vt tabel 61).

**Tabel 61.** Eksimismatriks (naissünteeshääle, A-/B-test). Emotsioonide tajumine naissünteeshäälest kolme erineva katsemudeli põhjal, vastused protsentides (vt [P6])

Sihtemotsioon	Vastusevariandid			
	Rõõm	Kurbus	Viha	Neutraalne
<b>Mudel 1</b>				
Rõõm	40/5	20/55	25/20	15/20
Kurbus	10/10	30/30	10/0	50/60
Viha	45/25	10/5	25/25	20/45
<b>Mudel 2</b>				
Rõõm	25/10	40/50	5/35	30/5
Kurbus	0/10	60/35	0/0	40/55
Viha	30/5	0/30	45/50	25/15
<b>Mudel 3</b>				
Rõõm	30/10	25/50	20/40	25/0
Kurbus	5/10	50/75	5/0	40/15
Viha	15/20	5/10	65/60	15/10
<b>Neutraalne</b>	15/25	25/30	10/5	50/40

*Märkus.* Poolpaksus kirjas on märgitud õigesti tajutud emotsioonid.

*Tajutestide tulemuste kokkuvõte naissüntheeshääle kohta.* Hindajad tundsid kurbuse- ja viha-emotsiooni kõige paremini ära M3 järgi sünteesitud lõikudes, rõõmu-emotsioon tunti kõige paremini ära mudeli 1 järgi. Neutraalse kõne tundsid hindajad samuti õigesti ära. Kui võrrelda A- ja B-testi tulemusi, siis naissüntheeshääle viha- ja kurbuse-emotsioon tunti ühtviisi hästi ära nii A- kui ka B-testis, rõõmu ja neutraalsust tunti aga paremini ära A-testis.

Emotsionaalse sünteeskõne akustiliste katsemudelite tulemustele tehtud *binaarne logistiline regressioonanalüüs* näitas, et nii mees- kui ka naissüntheeshääle puhul mängisid intensiivsuse tase ja põhitooni ulatus tajutestides õige emotsiooni tuvastamisel olulist rolli. Mõju oli ka kõnetempol, peamiselt küll ainult kurbuse tuvastamisel. Põhitooni kõrguse register ehk algväärtustamine emotsioonide tajumisel oluliseks parameetriks ei osutunud.

## 7.7. Diskussioon

Selles peatükis esitan võrdluse eestikeelse emotsionaalse kõne akustika uurimistulemuste ning varasemate uurimuste ja CPMi ennustuste vahel ning jagan soovitusi emotsioonide modelleerimiseks parameetrilisel kõnesüntheesil. Lisaks annan hinnangu emotsionaalse sünteeskõne akustiliste mudelite loomise eksperimendile ja esitan kolme põhiemotsiooni ja kahe sünteeshääle kohta eelista-tuimad emotsioonimudelid.

### 7.7.1. Kõne akustilised parameetrid

#### 7.7.1.1. Pausid

Pauside uurimuse eesmärk oli teada saada, kas lause emotsioon mõjutab kõneleja tehtud pauside arvu, asukohta, iseloomu ja kestust ning kas pausid on emotsioonide üksteisest eristamisel olulised ja kas kuulajad suudavad kõnelõigu emotsiooni ära tunda ainult pausierinevuste põhjal.

*Pauside arvu* kohta selgus uurimistulemustest, et eestikeelses ettelõetud emotsionaalses kõnes lause emotsioon pauside arvu ei mõjutanud. See oli küll emotsiooniti erinev, kuid sõltus eeskätt lausetes olevate kirjavahemärkide arvust: kõigi kolme emotsiooni puhul tegi lugeja pausi enamasti kirjavahemärgi kohal.

Varasematest uurimustest (vt nt Murray & Arnott, 2008; Schröder, Cowie, Douglas-Cowie, Westerdijk & Gielen, 2001; Tisljár-Szabó & Pléh, 2014; Yildirim, Bulut, Lee, Kazemzadeh, Deng, Lee, Narayanan & Busso, 2004) on teada, et pauside esinemissagedus võib emotsiooniti olla erinev. Enim eristuvad kurbuse-emotsiooni pausid: võrreldes rõõmu- ja viha-emotsiooniga esineb kurbuselausetes pause rohkem. Põhjus, miks minu uurimistöö tulemused pauside esinemissageduse kohta erinevad varasemate uurimuste tulemustest, võib peituda erinevas uurimismaterjalis (etteeloetud vs. spontaanne kõne ning esilekutsutud vs. näidel-

dud vs. loomulik kõne, vt ptk 2). Võib oletada, et ettelõetud kõne puhul jälgib lugeja teksti kirjavahemärke ning teeb pausi suure tõenäosusega nende kohal.

Ainus kirjavahemärk, mille kohal lugeja kõigi kolme emotsiooni puhul alati pausi ei teinud, oli koma. Kõige enam kutsus koma pausi esile rõõmlausetes ja kõige vähem kurbuselausetes. Erinevused emotsiooniti olid märkimisväärsed ning võib öelda, et koma kohal tehtud pauside arv eristab eestikeelses ettelõetud emotsionaalses kõnes emotsioone üksteisest.

*Pauside iseloomu* (hingamisega vs. hingamiseta) kohta näitasid minu uurimistöö tulemused, et eestikeelses emotsionaalses kõnes on hingamispaused rohkem kui hingamiseta pause, kuid hingamispauside esinemissagedus ei pruugi olla emotsioone eristav omadus: kõigil kolmel emotsioonil olid natuke üle poolte pausidest hingamispausid (53–59%).

Hingamispauside jagamine kirjavahemärgipausideks ja lisapausideks näitas, et kirjavahemärgi kohal tehtud hingamispauside osakaal oli emotsiooniti enam-vähem võrdne (56–61%), kuid oluline erinevus ilmnes lisapauside juures, kus vihalausetes tegi lugeja neid märgatavalt enam (55%) kui kurbuse- (31%) ja rõõmlausetes (21%). Seega võib pidada hingamisega lisapause oluliseks tunnuseks, mis eristab eestikeelses ettelõetud emotsionaalses kõnes viha-emotsiooni rõõmu- ja kurbuse-emotsioonist.

Kirjavahemärkide ja hingamispauside vahelist seost analüüsid selgus, et kurbuse- ja rõõmu-emotsiooni puhul tegi lugeja lause lõpus hingamisega pause märgatavalt enam (68% ja 67%) kui viha-emotsiooni puhul (58%). Samal ajal hingas kõneleja vihalausetes koma kohal oluliselt rohkem (67%) kui rõõmu- (36%) ja kurbuselausetes (27%). Põhjus võib seisneda selles, et viha-emotsiooni puhul hingas lugeja lause lõpus sisse harvem kui rõõmu- ja kurbuse-emotsiooni puhul ning seega oli kõnelejal hapnikku vähem ning tal tekkis lause keskel tihedamini vajadus hingata. Viha väljendamine nõuab kõnelejalt ka rohkem energiat kui rõõmu või kurbuse väljendamine (vt allpool ptk 5.1.4.) ning see suurendab omakorda kõneleja hapnikutarvidust. Samuti on vihalause pausid kestuselt lühemad kui kurbuse- või rõõmlause pausid (vt allpool *pauside kestus*), mis tähendab, et aega hapniku sissehingamiseks on vähem.

Pauside uurimuse tulemused *pauside kestuse* kohta näitasid, et pikimad pausid olid kurbuse- ja lühimad vihalausetes (vt ptk 4.1.1.). See, et kurbuse-emotsioonil on pikemad pausid kui rõõmu- ja viha-emotsioonil, läheb kokku varasemates uurimustes saadud tulemustega, mis väitsid, et kurbuselausete pausid on pikema kestusega kui rõõmu- ja vihalause pausid (Murray & Arnott, 2008; Schröder, Cowie, Douglas-Cowie, Westerdijk & Gielen, 2001; Tisljár-Szabó & Pléh, 2014; Yildirim, Bulut, Lee, Kazemzadeh, Deng, Lee, Narayanan & Busso, 2004).

Minu uurimistöö tulemuste põhjal saab väita, et emotsioonid mõjutavad teatud pausirühmade kestusi: hingamiseta ja hingamisega punktipaused ning hingamisega hüüumärgipausid (vt ptk 4.1.1.). Hingamiseta punktipaused eristab kurbuse-emotsiooni rõõmu- ja viha-emotsioonist, hingamisega punktipaused eristab kõiki kolme emotsiooni üksteisest ning hingamisega hüüumärgipausi kestus eristab kurbuse-emotsiooni viha-emotsioonist.

Pauside uurimuse tulemusi kokku võttes esitan mõned soovitud pauside modelleerimiseks emotsionaalse sünteeskõne jaoks:

1) pauside asukoht – kõigil kolmel emotsioonil tasub pausid panna kirja-vahemärkide kohale, v.a koma puhul, kus rõõmu- ja vihalausetes võib koma kohale pausi panna, kuid kurbuselausetes pigem mitte. Kuna lisapauside osakaal oli emotsiooniti väike (8–10%) ja sõltus pooltel juhtudel kas teksti sisust või lugejast endast, siis lisapause modelleerida ei ole oluline;

2) pauside iseloom – kõigil kolmel emotsioonil võib vähemalt pooled pausid modelleerida hingamispausideks. Pausid tuleb teha mõistliku vahemaa tagant, intervalli võib arvutada näiteks deklinatsiooni oletatava kestuse järgi. Kui viha puhul pole oluline, millise kirjavahemärgi kohale hingamine panna, siis rõõm-lausetes sobib hingamine pigem lause lõppu ning kurbuselausetes tuleb jälgida, et hingamine (ja paus) ei satuks koma kohale;

3) pauside kestus – pauside kestus tasub siduda emotsioonile omase kõne-tempoga (vt ptk 4.1.3.). Pikimateks pausideks maksab modelleerida lauselõpu-pausid ja lühimateks komapausid.

Kuigi minu pauside uurimus näitas mõnes kohas pauside vahel emotsiooniti olulisi erinevusi, selgus *tajutestist*, et ainult pausierinevuste põhjal ei suuda kuulajad siiski emotsioone üksteisest eristada. See tulemus lahkneb Tisljár-Szabó ja Pléh' (2014) uurimuse tulemusest, kus selgus, et kuulajad suudavad emotsioone üksteisest eristada ka ainult pauside kestuse erinevuse põhjal. Samal ajal ei saa ka väita, et pausidel pole eestikeelses emotsionaalses kõnes emotsioonide eristamisel üldse mingit tähtsust: pausid võivad muutuda oluliseks kombinatsioonis koos teiste emotsioone eristavate akustiliste parameetritega (näiteks kõnetempo, põhitoon ja intensiivsus).

### 7.7.1.2. Formantsagedused ja artikulatsiooni täpsus

Formandiuurimuse eesmärk oli teada saada, kas lause emotsioon mõjutab lühikeste vokaalide esimest kaht formantsagedust ja kõneleja artikulatsiooni täpsust ning kas need on emotsioonide eristamisel üksteisest ja neutraalsest kõnest olulised. Uurimistulemustest selgus, et emotsioonidel on olemas mõju nii vokaalide  $F_1$  kui ka  $F_2$  väärtustele, kusjuures  $F_1$  väärtused on oluliselt rohkem mõjutatud kui  $F_2$  väärtused. Kõneleja artikulatsiooni täpsus oli võrreldes neutraalse kõnega märgatavalt muutunud kurbuse-emotsiooni puhul: kurbuselausetes artikuleeris kõneleja võrreldes neutraalse kõne ning rõõmu- ja viha-emotsiooniga ebatäpsemalt.

*Emotsioonide mõju vokaalide esimesele ja teisele formantsagedusele.* Varasemad uurimused on näidanud, et *rõõmu-emotsiooniga* lausungites võivad vokaalide  $F_1$  ja  $F_2$  väärtused tõusta, kusjuures  $F_1$  väärtused on rohkem mõjutatud kui  $F_2$  väärtused (vt Goudbeek, Goldman & Scherer, 2009; Juslin & Laukka, 2003; Kienast & Sendlmeier, 2000; Tartter, 1980). Samal ajal on rõõmu puhul täheldatud ka võrreldes viha ja kurbusega  $F_1$  väärtuste langust (vt Pribil, Pribilova & Matousek, 2013). Schereri CPM ennustab, et nii lustliku kui ka

vaikse rõõmu puhul vokaalide  $F_1$  väärtused võrreldes neutraalse kõnega kaha-nevad.

Minu uurimuse tulemused näitasid, et rõõmu-emotsiooni vokaalide  $F_1$  väärtused erinesid neutraalsest kõnest oluliselt vaid vokaali *i* puhul, kus rõõmul oli madalam  $F_1$  kui neutraalsel kõnel. Vokaali *i* kohta saadud tulemus kinnitab CPMi ennustust. Vokaalide *a* ja *u* kohta saadud tulemus sellega kokku ei lange: rõõmu-emotsiooni vokaalide  $F_1$  väärtused olid sarnased neutraalse kõne vokaalide  $F_1$  väärtustega. Kui võrrelda rõõmulausete vokaalide  $F_1$  väärtusi teiste emotsioonidega, siis selgus, et olulised erinevused olid vokaalide *a* ja *i* puhul. Vokaali *a*  $F_1$  väärtuste järgi erines rõõm oluliselt nii vihast kui ka kurbusest, kusjuures rõõmu-emotsioonil oli neist kolmest emotsioonist kõige kõrgem  $F_1$  väärtuste keskmine. Vokaali *i*  $F_1$  väärtuste järgi erines rõõm oluliselt kurbusest, kusjuures rõõmu-emotsiooni vokaalidel oli kurbuse-emotsiooni vokaalidest kõrgem  $F_1$  väärtuste keskmine. Need tulemused lähevad suuresti kokku nende varasemate uurimustega, mis ütlesid, et rõõmu-emotsioonil  $F_1$  väärtused tõusevad.

Rõõmulausete vokaalide  $F_2$  väärtused ei erinenud oluliselt ei neutraalse kõne ega ka teiste emotsioonide vokaalide  $F_2$  väärtustest. Seega minu uurimuse tulemused rõõmulausete vokaalide  $F_2$  väärtuste kohta ei lange kokku varasemate uurimistulemuste ega CPMi ennustusega.

*Kurbuse-emotsiooni* kohta on varasemates uurimustes märgitud vokaalide  $F_1$  väärtuste langust (Juslin & Laukka, 2003). Täheldatud on ka madalamaid  $F_2$  väärtusi kui rõõmu-emotsiooni vokaalidel (Goudbeek, Goldman & Scherer, 2009). Schereri CPM ennustab, et võrreldes neutraalse kõnega kurbuselausetes heliliste häälikute  $F_1$  väärtused tõusevad ja  $F_2$  väärtused langevad.

Minu uurimistulemused kurbuse-emotsiooni vokaalide  $F_1$  väärtuste kohta näitavad, et need on võrreldes neutraalse kõne ja teiste emotsioonidega kõigi kolme vokaali puhul kõige madalamad, kusjuures vokaalide *a* ja *i* puhul on need erinevused ka statistiliselt olulised. Kurbuselausete vokaalide  $F_2$  väärtused erinesid oluliselt vaid vokaali *a* puhul viha-emotsiooni vokaalide  $F_2$  väärtustest, kusjuures kurbuselausete vokaalidel oli kõrgem  $F_2$  kui vihalausete vokaalidel. Seega minu uurimuse tulemused langevad suure osas kokku varasemate uurimistulemustega, kuid ei kinnita CPMi ennustust.

*Viha-emotsiooni* puhul on varasemates uurimustes täheldatud vokaalide kõrgemat  $F_1$  ning madalamat  $F_2$  (Goudbeek, Goldman & Scherer, 2009; Juslin & Laukka, 2003; Murray & Arnott, 2008; Pribil, Pribilova & Matousek, 2013). Schereri CPM ennustab samuti, et võrreldes neutraalse kõnega tõusevad nii kuuma viha kui ka külma viha lausungites heliliste häälikute  $F_1$  väärtused ja langevad  $F_2$  väärtused. Minu uurimuse tulemused näitasid, et viha-emotsiooni vokaalide  $F_1$  väärtused olid neutraalse kõne vokaalide  $F_1$  väärtustest madalamad ning need erinevused olid statistiliselt olulised vokaalide *a* ja *i* puhul. Teiste emotsioonide vokaalidest erinesid viha-emotsiooni vokaalide  $F_1$  väärtused oluliselt samuti vokaalide *a* ja *i* puhul. Mõlemal juhul oli vihalausete vokaalide  $F_1$  kõrgem kui kurbuselausete vokaalide  $F_1$ . Vokaali *a* puhul oli vihalausete vokaalidel madalam  $F_1$  kui rõõmu-emotsioonil, vokaali *i* puhul viha- ja rõõmu-emotsioon teineteisest oluliselt ei erinenud. Vihalausete vokaalide  $F_2$  väärtused

neutraalsest kõnest ja teistest emotsioonidest oluliselt ei erinenud, ainus erand oli vokaal *a*, mille puhul erinesid oluliselt viha-emotsiooni vokaalide  $F_2$  väärtused kurbuselause vokaalide  $F_2$  väärtustest, kusjuures vihalause vokaalidel oli madalam  $F_2$  kui kurbuselause vokaalidel. Minu uurimistulemused viha-emotsiooni vokaalide kohta langesid osaliselt kokku varasemate uurimustega  $F_1$  puhul, kuid ei kattunud  $F_2$  kohta saadud tulemustega. Samuti ei kinnitanud minu formandiuurimuse tulemused CPMi ennustust.

Esimese ja teise formantsageduse uurimistulemuste põhjal saab öelda, et eestikeelses emotsionaalses kõnes mõjutasid emotsioonid eeskätt vokaalide *a* ja *i* esimest formantsagedust: vokaali *a* puhul osutusid muutused  $F_1$  väärtustes olulisteks kurbuse- ja viha-emotsiooni eristamisel neutraalsest kõnest ja kõigi kolme emotsiooni eristamisel üksteisest. Vokaali *i* puhul osutusid muutused  $F_1$  väärtustes olulisteks kõigi kolme emotsiooni eristamisel neutraalsest kõnest ja kurbuse-emotsiooni eristamisel rõõmu- ja viha-emotsioonist.

*Artikulatsiooni täpsus.* Varasemad uurimistulemused on näidanud, et kurbuse-emotsioonile on omane madal artikulatsiooni täpsus. Viha- ja rõõmu-emotsiooni kohta on saadud kahesuguseid tulemusi: artikulatsiooni täpsus võib viha ja rõõmu puhul tõusta (Johnstone & Scherer, 2000; Juslin & Laukka, 2003; Kienast & Sendlmeier, 2000; Murray & Arnott, 2008; Paeschke, Kienast & Sendlmeier, 1999), aga ka viha puhul langeda (Iida, Campbell, Higuchi, & Yasumura, 2003) ning rõõmu puhul jääda sarnaseks neutraalse kõne artikulatsiooni täpsusega (Kienast & Sendlmeier, 2000).

Minu uurimistöö tulemused artikulatsiooni täpsuse kohta näitasid, et neutraalsest kõnest ja teistest emotsioonidest erines vaid kurbuselause artikulatsioon, mis oli võrreldes teiste emotsioonirühmadega märgatavalt ebatäpsem. Rõõmu- ja viha-emotsioon artikulatsiooni täpsuse poolest neutraalsest kõnest ja teistest emotsioonidest oluliselt ei erinenud ning nende artikulatsiooni täpsus jäi neutraalse kõne omaga sarnaseks. Saadud tulemuste põhjal võib väita, et keeleti on tendents väljendada kurbust ühtemoodi: ebatäpsema artikulatsiooniga. Rõõmu ja viha artikulatsiooni kohta on keeleti saadud erinevaid tulemusi. Selle põhjuseks võib olla näiteks erinev uurimismaterjal või nende emotsioonide keele- ja kultuurispetsiifiline väljenduslaad.

Formandiuurimuse tulemuste põhjal esitan soovitusena langetada emotsionaalses sünteeskõnes kurbuse-emotsiooniga lausetes vokaalide formantsageduse  $F_1$  väärtusi võrreldes neutraalse kõne ja teiste uuritud emotsioonidega ning muuta madalamaks ka artikulatsiooni täpsus.

### 7.7.1.3. Kõnetempo

Kõnetempo uurimuse eesmärk oli teada saada, kas emotsioonid mõjutavad ettelugeja kõne kiirust ning kas kõnetempo on emotsioonide üksteisest ja neutraalsest kõnest eristamisel oluline.

Minu uurimistulemused eestikeelse etteloetud emotsionaalse kõne tempo kohta näitasid, et uuritud materjal mõjutasid emotsioonid nii üldist kõnetempot kui ka fraasisisest kõnetempot: kõige kiirem kõnetempo oli viha-



emotsioonil, seejärel rõõmu-emotsioonil, neutraalsel kõnel ja kõige aeglasem kurbuse-emotsioonil. Üldine kõnetempo eristas kõiki kolme emotsiooni üksteisest ning viha-emotsiooni neutraalsest kõnест. Fraasisisese kõnetempo kohta saab öelda sama, kuid lisaks eristas see veel kurbuse-emotsiooni neutraalsest kõnест. Fraasi viimase sõna kestusele emotsioonid olulist mõju ei avaldanud.

Eestikeelse etteloetud emotsionaalse kõne tempo uurimistulemused kattusid suuresti ka Schereri (1986) CPMi ennustusega (vt ptk 2.1.): võrreldes neutraalse kõnega tõusis kõnetempo vihalausetes ja langes kurbuselausetes. Rõõmu-emotsiooni tempo kohta saadud tulemused CPMi ennustusega kokku ei langenud: rõõmulausetes oli kõnetempo neutraalse kõne tempoga sarnane.

Kui kõrvutada eesti emotsionaalse kõne tempo uurimistulemusi varasemate uurimuste tulemustega (vt ptk 2.1.), siis selgub, et suurelt osalt langesid eesti emotsionaalse kõne tempo uurimistulemused nendega kokku: võrreldes neutraalse kõnega oli kõnetempo vihalausetes kiirem ja kurbuselausetes aeglasem. Samuti oli viha- ja rõõmulausete tempo kiirem kui kurbuselausete tempo.

Kõnetempo uurimuse tulemuste põhjal võib oletada, et eesti keeles väljenduvad emotsioonid kõnetempos teiste keeltega sarnaselt: kõneleja kõnetempo on viha-emotsiooni puhul kiirem ja kurbuse-emotsiooni puhul aeglasem. (Ka teise temporaaalse parameetri – pauside – kohta sain sama tulemuse: eestikeelses emotsionaalses kõnes on viha-emotsiooniga lõikudes pauside kestused lühemad kui kurbuse-emotsiooniga lausetes (vt ptk 4.1.1.))

Kõnetempo uurimuse tulemuste põhjal esitan soovitusel modelleerida eestikeelses emotsionaalses sünteeskõnes kõnetempo nii, et see oleks kõige kiirem vihalausetes ja kõige aeglasem kurbuselausetes (viha > rõõm > neutraalne > kurbus).

#### 7.7.1.4. Intensiivsus

Intensiivsuse uurimuse eesmärk oli teada saada, kas emotsioonid mõjutavad kõneleja hääle intensiivsuse taset ja selle varieerumist (intensiivsuse ulatust) ning kas intensiivsuse tase ja ulatus on emotsioonide eristamisel üksteisest ja neutraalsest kõnест olulised.

Minu uurimistulemused eestikeelse etteloetud emotsionaalse kõne intensiivsuse kohta näitasid, et uuritud materjal eristas intensiivsuse tase kõiki kolme emotsiooni nii üksteisest kui ka neutraalsest kõnест, kuid intensiivsuse ulatusel eristavat rolli ei olnud. Lause alguse ja lõpu intensiivsuse tase eristas kurbuse-emotsiooni neutraalsest kõnест. Kõige kõrgem oli intensiivsuse tase neutraalses kõnes ja kõige madalam kurbuselausetes. Need tulemused kattuvad vaid osaliselt Schereri (1986) CPMi ja teiste keelte kohta saadud intensiivsuse taseme uurimistulemustega (Banse & Scherer, 1996; Johnstone & Scherer, 2000; Juslin & Laukka, 2003; Murray & Arnott, 2008), mis ütlesid, et võrreldes neutraalse kõnega tõusevad intensiivsuse taseme näitajad lustliku rõõmu ja viha puhul ning langevad kurbuse ja vaikse rõõmu puhul.

Eestikeelse etteloetud emotsionaalse kõne intensiivsuse taseme uurimistulemused erinesid teiste keelte kohta saadud uurimuste tulemustest kõige enam

neutraalse kõne puhul. Intensiivsuse taseme mõõtmistulemustest (vt ptk 4.1.4.) on näha, et võrreldes emotsionaalse kõnega püsis lugeja hääle intensiivsus neutraalse kõne puhul ühtlasemal nivool: intensiivsuse taseme muutumine lause jooksul oli neutraalsel kõnel märkimisväärselt väiksem kui näiteks kurbusel ja vihal. See võib olla eestikeelse kõne eripära, et neutraalne kõne on kõrgema intensiivsuse taseme ja ühtlasema helinivooga kui emotsionaalne kõne. Samuti võis tulemust mõjutada ka ette loetud kõne lugemisstiil.

Nende uurimistulemuste põhjal esitan soovitusel modelleerida eestikeelses sünteeskõnes intensiivsuse tase kõige kõrgemaks neutraalses kõnes ning kõige madalamaks kurvas kõnes (neutraalne > viha > rõõm > kurbus). Kuigi uurimistulemused näitasid, et intensiivsuse ulatusel pole eristavat rolli, võib see osutada siiski oluliseks parameetrikaks koos teiste emotsioonide eristavate akustiliste tunnustega. Seepärast soovitatakse modelleerida intensiivsuse ulatus kõige laiemaks kurbuselaulsetes ja kõige kitsamaks rõõmu puhul (kurbus > viha > neutraalne > rõõm).

#### 7.7.1.5. Põhitoon

Põhitooni uurimise eesmärk oli teada saada, kas emotsioonid mõjutavad kõneleja põhitooni kõrgust ja selle varieerumist (põhitooni ulatust) ning kas põhitooni kõrgus ja ulatus on emotsioonide eristamisel üksteisest ja neutraalsest kõnest olulised.

Minu uurimistulemused põhitooni kohta eestikeelses ette loetud emotsionaalses kõnes näitasid, et uuritud materjalil oli emotsioonidel kõneleja põhitooni kõrgusele ja ulatusele mõju olemas: põhitooni kõrgus eristas viha-emotsiooni rõõmu- ja kurbuse-emotsioonist ja neutraalsest kõnest ning põhitooni ulatus eristas kurbuse-emotsiooni rõõmu- ja viha-emotsioonist ja viha-emotsiooni neutraalsest kõnest. Lause alguse ja lõpu põhitooni kõrgusel eristavat rolli ei olnud.

Kui kõrvutada minu uurimistulemused Schereri (1986) CPMi ennustusega (vt ptk 2.5.), siis selgub, et need kattuvad ainult viha-emotsiooni puhul, mis ainukesel emotsioonina eristus põhitooni kõrguse ja ulatuse poolest neutraalsest kõnest: võrreldes neutraalse kõnega oli viha-emotsioonil põhitoon madalam ja selle ulatus kitsam (nagu on ennustatud külma viha kohta).

Kui võrrelda eesti emotsionaalse kõne põhitooni uurimuse tulemusi teiste keelte kohta saadud uurimistulemustega, siis tuleb vaadelda põhitooni kõrgust ja ulatust eraldi. Nagu eespool märkisin, oli eestikeelses emotsionaalses kõnes viha-emotsioon ainuke, mille põhitooni kõrgus erines teiste emotsioonide ja neutraalse kõne põhitooni kõrgusest. Varasemad emotsiooniakustika uurimistulemused on põhitooni kõrguse kohta näidanud, et neutraalsest kõnest eristab põhitooni kõrgus peale viha-emotsiooni ka näiteks rõõmu- ja kurbuse-emotsiooni ning neid emotsioone ka omavahel. Viha-emotsiooni põhitooni kohta on öeldud, et see võib võrreldes rõõmu-emotsiooniga olla kas kõrgem või madalam, kuid võrreldes kurbuse-emotsiooni ja neutraalse kõnega kõrgem (vt Laukka, Neiberg, Forsell, Karlsson & Elenius, 2011; Murray & Arnott, 2008;

Pell, Paulmann, Dara, Allasseri & Kotz, 2009). Minu uurimistulemused viha-emotsiooni põhitooni kõrguse kohta varasemate uurimistulemustega kokku ei lange (v.a uurimused, mille tulemustest selgus, et viha-emotsiooni põhitoon võib olla rõõmu-emotsiooni põhitoonist madalam). Uurimistulemustest võib järeldada, et eestikeelses emotsionaalses kõnes erineb kõneleja põhitoon vaid viha-emotsiooni puhul, mil see on võrreldes neutraalse kõne ja teiste emotsioonidega madalam.

Põhitooni ulatuse mõõtmistulemused näitasid, et eestikeelses emotsionaalses kõnes eristab põhitooni ulatus esiteks kurbuse-emotsiooni rõõmust ja vihast (kurbuse põhitooni ulatus oli kõige kitsam) ning teiseks viha-emotsiooni neutraalsest kõnest (viha-emotsiooni põhitooni ulatus oli laiem). Varasemates uurimustes on saadud tulemuseks, et neutraalsest kõnest eristuvad põhitooni ulatuse poolest ka rõõmu- ja kurbuse-emotsioon ning samuti emotsioonid omavahel. Kurbuse-emotsiooni kohta on varasemad uurimused saanud tulemuseks, et selle põhitooni ulatus võib võrreldes neutraalse kõnega olla kas laiem või kitsam, kuid rõõmu- ja viha-emotsiooniga võrreldes on kurbuse-emotsiooni põhitooni ulatus kitsam (vt Laukka, Neiberg, Forsell, Karlsson & Elenius, 2011; Murray & Arnott, 2008; Pell, Paulmann, Dara, Allasseri & Kotz, 2009). Kurbuse-emotsiooni puhul langevad minu uurimuse tulemused ja varasemad uurimistulemused osaliselt kokku: kurbuse-emotsiooni puhul muutub kõneleja põhitooni kõrgus võrreldes rõõmu- ja viha-emotsiooniga oluliselt vähem. Viha-emotsiooni põhitooni ulatuse kohta on varasemates uurimustes saadud tulemuseks, et võrreldes rõõmu-emotsiooni põhitooni ulatusega võib see olla nii kitsam kui ka laiem, kuid võrreldes neutraalse kõne ja kurbuse-emotsiooniga laiem (vt Laukka, Neiberg, Forsell, Karlsson & Elenius, 2011; Murray & Arnott, 2008; Pell, Paulmann, Dara, Allasseri & Kotz, 2009). Viha-emotsiooni puhul langevad minu uurimuse tulemused ja varasemad uurimistulemused kokku vaid väikeses osas: viha-emotsiooni põhitooni ulatus erines oluliselt vaid neutraalse kõne põhitooni ulatusest, olles sellest laiem. Seega võib oletada, et eestikeelses ettelõetud emotsionaalses kõnes varieerub kõneleja põhitooni kõrgus kurbuse-emotsiooni puhul oluliselt vähem kui viha- ja rõõmu-emotsiooni puhul ning viha-emotsiooni puhul oluliselt rohkem kui kurbuse-emotsiooni ja neutraalse kõne puhul.

Põhitooni uurimus näitas, et eestikeelses ettelõetud emotsionaalses kõnes avaldab põhitooni kõrgusele ainukesena mõju viha-emotsioon, põhitooni ulatust mõjutavad võrreldes neutraalse kõnega viha-emotsioon ja võrreldes teiste emotsioonidega kurbuse-emotsioon. Kuigi põhitooni kõrguse ja ulatuse erinevused ei olnud emotsioonipaariti ja võrreldes neutraalse kõnega alati statistiliselt olulised, näitas analüüsitud materjal siiski, et põhitoonil on tendents emotsioone eristada ning seepärast esitan soovitusel modelleerida põhitoon sünteeskõnes nii, et see oleks kõrgeim rõõmu-emotsioonil ja madalaim viha-emotsioonil, neutraalse kõne ja kurbuse-emotsiooni põhitooni kõrgused võib modelleerida ühesuguseks (rõõm > neutraalne = kurbus > viha). Põhitooni ulatuse tasub modelleerida kõige laiemaks vihal ning kõige kitsamaks kurbusel (viha > rõõm > neutraalne > kurbus).

#### 7.7.1.6. Rõõmu, kurbuse ja viha akustilised mustrid

Järgnevalt kirjeldan lühidalt, millised on minu uurimismaterjali põhjal rõõmu, kurbuse ja viha akustilised mustrid eestikeelses etteloetud emotsionaalses kõnes võrreldes neutraalse kõnega.

*Rõõm.* Rõõmu-emotsiooni akustilist mustrit iseloomustavad võrreldes neutraalse kõnega sarnane artikulatsiooni täpsus, kiirem kõnetempo, madalam intensiivsuse tase ja kitsam intensiivsuse taseme varieerumine ning kõrgem põhitoon ja suurem põhitooni kõrguse varieerumine.

*Kurbus.* Kurbuse-emotsiooni akustilist mustrit iseloomustavad võrreldes neutraalse kõnega ebatäpsem artikulatsioon, aeglasem kõnetempo, madalam intensiivsuse tase ja laiem intensiivsuse taseme varieerumine ning sarnane põhitooni kõrgus ja väiksem põhitooni kõrguse varieerumine.

*Viha.* Viha-emotsiooni akustilist mustrit iseloomustavad võrreldes neutraalse kõnega sarnane artikulatsiooni täpsus, kiirem kõnetempo, madalam intensiivsuse tase ja laiem intensiivsuse taseme varieerumine ning madalam põhitoon ja suurem põhitooni kõrguse varieerumine.

### 7.7.2. Emotsionaalse sünteeskõne akustilised mudelid

Emotsionaalse sünteeskõne akustiliste mudelite loomise eksperimendi eesmärk oli leida eestikeelsele kõnesüntesaatorile sobivaimad akustilised mudelid rõõmu, kurbuse ja viha väljendamiseks eestikeelsel parameetrilisel kõnesünteesil nii mees- kui ka naissünteeshäälele.

Tajutestide tulemustest selgus, et põhiemotsioonide katsemudelite kohta seatud kaks hüpoteesi (H1: katseisikud tunnevad nii mees- kui ka naissünteeshäälest kõige paremini ära võimendatud parameetrite väärtustega mudeli 3 järgi sünteesitud emotsioonid ning H2: kõnesüntesaatori enda kõnemudeli järgi sünteesitud neutraalse kõne) leidsid mõlemad kinnitust osaliselt: kuulajad tundsid nii mees- kui ka naissünteeshäälest kõige paremini ära mudeli 3 ehk võimendatud parameetrite väärtustega mudeli järgi sünteesitud emotsioonid, v.a naissünteeshääle rõõmu-emotsioon, mis tunti kõige paremini ära mudeli 1<sup>62</sup>

---

<sup>62</sup> Selline naissünteeshääle tulemus oli ootamatu. Seda enam, et meessünteeshääles tunti rõõm kõige paremini ära just mudeli 3 puhul, nagu oli eeldatud. Naissünteeshääle rõõmu-emotsiooni tuvastusprotsentide erinevused mudeli 1 ja 3 vahel polnud väga suured: vastavalt 40% ja 30%. Meessünteeshääle rõõmu-emotsiooni parim tuvastusprotsent oli 55 (mudel 3). Võrdluseks: teistel emotsioonidel olid parimad tuvastusprotsendid 65–80 (mudel 3). Võib-olla peitub ootamatu tulemuse üks põhjus selles, et me ei suutnud rõõmule head emotsioonimudelit luua: naissünteeshääle puhul aeti rõõm kõige enam segamini kurbusega. Veel võis sellise üllatava tulemuse anda intensiivsuse taseme muutmine, mille tulemusel nihkus mudeli 1 puhul rõõmu-emotsiooni intensiivsuse tase kurbuse-emotsiooni intensiivsuse tasemele lähemale, mudeli 3 puhul vastupidi – sellest kaugemale. Teiste parameetrite väärtused muutusid nii, et mudel 1 viis rõõmu-emotsiooni kurbuse-emotsioonile lähemale ja mudel 3 kaugemale. Seega võis intensiivsuse taseme muutus põhjustada selle, et rõõm tunti paremini ära mudeli 1 järgi. Veel võis

järgi. Neutraalset kõne meessüntheeshäälest ära ei tuntud, seda peeti kurbuseks; naissüntheeshääle puhul tunti neutraalsus õigesti ära. Kõige enam aeti mõlema sünteeshääle puhul rõõm segamini kurbusega ning kurbus ja viha segamini neutraalse kõnega.

Eestikeelse kõnesüntesaatori peal kõige paremini töötava katsemudeli ehk emotsioonimudeli 3 puhul oli keskmine tuvastusprotsent (mees- ja naissüntheeshääle tulemused koos) kurbusel 60, vihal 59 ja rõõmul 30. Neutraalse kõne tuvastusprotsent oli 35.

Emotsionaalse sünteeskõne tajutestide tulemuste põhjal saab järeldada, et kõnesüntesaator suutis nii mees- kui ka naissüntheeshääle puhul rahuldavalt väljendada kurbust ja viha (mudeli 3 järgi), aga mitte rõõmu. Neutraalse kõne väljendamisega sai kõnesüntesaator rahuldavalt hakkama ainult naissüntheeshääle puhul.

Kõigi mudelite põhjal tundsid hindajad emotsioonid meessüntheeshääles õigesti ära keskmiselt 32%-l juhtumitest ja naissüntheeshääles 38%-l juhtumitest. Seega suutis kõnesüntesaator väljendada emotsioone paremini naissüntheeshääles. Üks põhjus võib siin olla see, et inimkõne akustiline analüüs, mille tulemused olid katsemudelite loomise aluseks, oli tehtud just naishääle põhjal.

Kui võrrelda A- ja B-testi tulemusi, siis meessüntheeshääle puhul tunti sünteestitud emotsioonid paremini ära B-testis, kus helilõik algas akustiliselt neutraalse kõnega ja muutus alates lõigu teisest lausest kas rõõmsaks, kurvaks, vihaseks või jäi neutraalseks. Seega aitas meessüntheeshääle puhul emotsiooni vahetumine helilõigu keskel kaasa õige emotsiooni tuvastamisele. Naissüntheeshääle puhul tunti viha- ja kurbuse-emotsioon enam-vähem võrdselt ära nii A- kui ka B-testis, rõõmu ja neutraalsust tunti paremini ära A-testis, kus terve lõik kandis sama emotsiooni või neutraalsust. Seega vastupidiselt meessüntheeshäälele ei olnud naissüntheeshääle puhul emotsiooni vahetumine helilõigu keskel tähtis õige emotsiooni tuvastamiseks.

Tajutestide tulemuste üldine analüüs logistilisel regressioonil näitas, et emotsioonide tuvastamisel mängisid katsemudelites olulist rolli kõne intensiivsuse taseme ja põhitooni ulatuse parameetrite väärtused. Kõnetempo oli samuti oluline, kuid selle panus emotsiooni tuvastamisse oli pigem tagasihoidlik (v.a kurbuse-emotsioon, mille tuvastamisele aitas kõnetempo märkimisväärselt kaasa). Põhitooni algväärtustamine ehk register emotsiooni tuvastamise seisukohalt oluline parameeter ei olnud. Samal ajal on uurimustest teada, et põhitooni kõrgus on üks peamine emotsioone eristav tunnus (vt nt Arias, Busso & Yoma, 2014; Banse & Scherer, 1996; Busso, Lee & Narayanan, 2009; Pell, Paulmann, Dara, Alasseri & Kotz, 2009; ten Bosch, 2003). Eestikeelses emotsionaalses kõnes eristab põhitooni kõrgus viha-emotsiooni rõõmu- ja kurbuse-

---

üllatava tulemuse põhjus peituda naissüntheeshääle doonorhääles. Ehk oli seal midagi sellist, mille tõttu sobis tema puhul rõõmu väljendamiseks paremini mudel 1, mitte mudel 3. Eestikeelse kõne emotsioonide akustika uurimine ja emotsioonide modelleerimine on alles algusjärgus ning ammendavat seletust sellisele tulemusele ei ole praegu võimalik anda.

emotsioonist ning neutraalsest kõnest (vt ptk 4.1.5.). Seega vajavad põhitooni kõrguse väärtused eestikeelse emotsionaalse sünteeskõne akustilises mudelis viha-emotsiooni puhul täpsustamist. Samuti vajaksid täpsustamist kõnetempo väärtused rõõmu- ja viha-emotsiooni puhul.

Kui võrrelda logistilise regressiooni analüüsi tulemusi ja inimkõne akustika uurimistulemusi, siis selgub, et kõnetempol oli inimkõne puhul oluline roll nii emotsioonide üksteisest kui ka neutraalsest kõnest eristamisel, sünteeskõnes aga eristas kõnetempo vaid kurbuse-emotsiooni. Põhitooni kõrgus eristas inimkõne puhul viha-emotsiooni neutraalsest kõnest ja teistest emotsioonidest, sünteeskõnes põhitooni kõrgusel emotsioone eristavat rolli ei olnud. Põhitooni ulatus eristas inimkõnes kurbuse-emotsiooni teistest emotsioonidest ja viha-emotsiooni neutraalsest kõnest ning ka sünteeskõne puhul oli põhitooni ulatusel oluline roll emotsioonirühmade üksteisest eristamisel. Intensiivsuse tase eristas emotsioone üksteisest ja neutraalsest kõnest hästi nii inimkõnes kui ka sünteeskõnes.

Katsemudelite testitulemustele tuginedes esitan kolme põhiemotsiooni ja kahe sünteeshääle kohta eelistatuimad emotsioonimudelid eestikeelse parameetrilise kõnesünteesi jaoks (vt tabel 62).

**Tabel 62.** Eelistatud emotsioonimudelid mees- ja naissünteeshäälele eestikeelse parameetrilise kõnesünteesi jaoks (vt [P6])

Emotsioon	Meessünteeshäääl	Emotsioon	Naissünteeshäääl
Rõõm	M3	Rõõm	M1
Kurbus	M3	Kurbus	M3
Viha	M3	Viha	M3

*Märkus.* M3 – inimkõne parameetrite väärtusi võimendav mudel; M1 – inimkõne parameetrite väärtusi vähendav mudel.

## 7.8. Kokkuvõte

Eesti emotsionaalse kõne akustika uurimisel oli kaks eesmärki: a) saada teada, kuidas kolm põhiemotsiooni – rõõm, kurbus ja viha – eestikeelses ettelõetud kõnes akustiliselt väljenduvad, ning b) luua analüüsi tulemustele tuginedes eestikeelsele kõnesüntesaatorile parameetrilise sünteesi tarvis emotsionaalse kõne akustilised mudelid, mille abil võiks süntesaator äratuntavalt väljendada rõõmu, kurbust ja viha.

Uurimuse eesmärkidest lähtuvalt püstitasin kolm uurimisküsimust: 1) kas, millisel määral ja mis suunas mõjutavad rõõmu-, kurbuse- ja viha-emotsioon eestikeelses ettelõetud kõnes akustiliste parameetrite (pausid, formandid, kõnetempo, intensiivsus ja põhitoon) väärtusi; 2) millised neist parameetritest on eestikeelses ettelõetud emotsionaalses kõnes need, mis võimaldavad emotsioone üksteisest ja neutraalsest kõnest eristada, ning 3) kuidas rakendada

akustikaanalüüsi tulemusi eesti tekst-kõne parameetrilisele sünteesile emotsionaalse kõne akustiliste mudelite loomisel.

Kahele esimesele uurimisküsimusele vastavad artiklid [P1], [P2], [P3], [P4] ja [P5], kolmandale uurimisküsimusele vastab artikkel [P6].

Pauside uurimus ([P1]) näitas, et emotsioonid mõjutavad teatud pausirühmade kestusi ning nendel erinevustel on ka emotsioone üksteisest eristav roll. Samas selgus tajutestist, et ainult pausierinevuste põhjal ei suutnud kuulajad emotsioone üksteisest eristada.

Formandiuurimusest ([P2]) selgus, et eestikeelses emotsionaalses kõnes oli emotsioonidel vokaalide *a* ja *i* esimesele formantsagedusele mõju olemas ning emotsioonide tekitatud muutused osutusid emotsioonide üksteisest ja neutraalsest kõnest eristamisel olulisteks (v.a paar erandit). Artikulatsiooni täpsuse analüüs näitas, et kurbuselausete artikulatsioon oli märgatavalt ebatäpsem kui neutraalsel kõnel või teistel uuritud emotsioonidel.

Kõnetempo uurimusest ([P3]) nähtus, et emotsioonid mõjutavad ka kõnetempot: kõige kiirem oli kõnetempo viha-emotsiooni puhul ja kõige aeglasem kurbuselausetes. Kõnetempo erinevused eristasid emotsioone üksteisest ning viha-emotsiooni neutraalsest kõnest.

Intensiivsuse uurimusest ([P4]) selgus, et eestikeelses etteloetud emotsionaalses kõnes oli emotsioonidel kõneleja hääle intensiivsuse tasemele mõju olemas: kõige kõrgem oli intensiivsuse tase neutraalse kõne puhul ning kõige madalam kurbuselausetes. Intensiivsuse tase osutus oluliseks parameetriks emotsioonide eristamisel nii üksteisest kui ka neutraalsest kõnest. Intensiivsuse taseme varieerumisele emotsioonid märgatavat mõju ei avaldanud.

Põhitooni uurimus ([P5]) näitas, et põhitooni kõrgust mõjutas emotsioonidest oluliselt ainult viha: vihalausetes oli põhitoon võrreldes teiste emotsioonide ja neutraalse kõnega oluliselt madalam. Põhitooni ulatust mõjutasid kõik kolm emotsiooni: põhitooni kõrgus varieerus kurbuselausetes oluliselt vähem kui rõõmu- ja vihalausetes ning vihalausetes oluliselt rohkem kui kurbuselausetes ja neutraalse kõne puhul.

Emotsioonimudelite loomise eksperiment ([P6]) kinnitas, et akustikaanalüüsi tulemustele tuginevate emotsioonimudelitega suudab eestikeelne kõnesüntesaator rahuldavalt väljendada nii kurbust kui ka viha. Rõõmu puhul paraku tulemused nii head ei olnud. Kuna minu emotsiooniakustika uurimus tugines vaid ühe ettelugeja kõnematerjalil, siis kajastavad uurimistulemused ainult ühte võimalikku viisi nimetatud emotsioone eesti keeles väljendada ning nende põhjal loodud emotsioonimudelid ei ole ainus tee, kuidas mainitud emotsioone eestikeelses sünteeskõnes esitada. Minu uurimistööga on tehtud esimene katsetus panna eestikeelne kõnesüntesaator väljendama rõõmu, kurbust ja viha ning see võiks olla lähtepunkt edasisele eestikeelse emotsionaalse sünteeskõne akustiliste mudelite arendamisele.

## REFERENCES / VIIDATUD KIRJANDUS

- Altrov, R. (2014). *The Creation of the Estonian Emotional Speech Corpus and the Perception of Emotions*. Tartu: University of Tartu Press.
- Altrov, R., & Pajupuu, H. (2012). Estonian Emotional Speech Corpus: theoretical base and implementation. In 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)*, (pp. 50–53). Istanbul.
- Altrov, R., & Pajupuu, H. (2013). Estonian Emotional Speech Corpus: Content and options. In J. Bamford, S. Cavalieri, & G. Diani (Eds.), *Variation and Change in Spoken and Written Discourse: Perspectives from Corpus Linguistics. Dialogue Studies 21* (pp. 109–122). Amsterdam: John Benjamins.
- Altrov, R., & Pajupuu, H. (2015). The influence of language and culture on the understanding of vocal emotions. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics*, 6, 3, 11–48, 10.12697/jeful.2015.6.3.01.
- Arias, J. P., Busso, C., & Yoma, N. B. (2014). Shape-based modeling of the fundamental frequency contour for emotion detection in speech. *Computer Speech and Language* 28, 278–294.
- Audibert, N., Aubergé, V., & Rilliard, A. (2005). The prosodic dimensions of emotion in speech: the relative weights of parameters. *Proceedings of the 9th International Conference on Speech Communication and Technology (INTERSPEECH 2005)*, 525–528.
- Bachorowski, J.-A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science*, 8, 2, 53–57. doi:10.1111/1467-8721.00013.
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70, 3, 614–636.
- Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.19, retrieved 13 June 2016 from <http://www.praat.org/>.
- Braun, A., & Oba, R. (2007). Speaking tempo in emotional speech – a cross-cultural study using dubbed speech. *ParaLing'07*, 77–82.
- Burkhardt, F., Audibert, N., Malatesta, L., Türk, O., Arslan, L. M., & Auberge, V. (2006). Emotional prosody – does culture make a difference? *Proceedings of Speech Prosody*. Dresden, Germany. May 2–5, 2006. doi: 10.1.1.148.6152.
- Burkhardt, F., & Campbell, N. (2015). Emotional Speech Synthesis. In R. Calvo, S. D'Mello, J. Gratch, & A. Kappas (Eds.), *The Oxford Handbook of Affective Computing* (pp. 286–295). Oxford University Press, New York.
- Busso, C., Lee, S. & Narayanan, S. (2009). Analysis of emotionally salient aspects of fundamental frequency for emotion detection. *IEEE Trans Audio Speech Lang Process*, 17, 4, 582–596.
- Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., & Fellenz, W. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18, 32–80.
- Cowie, R., Sussman, N., & Ben-Ze'ev, A. (2011). Emotions: Concepts and definitions. In R. Cowie, C. Pelachaud, & P. Petta (Eds.), *Emotion-oriented systems: The HUMAINE handbook* (pp. 9–31). Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-15184-2\_2.
- Douglas-Cowie, E., Campbell, N., Cowie, R., & Roach, P. (2003). Emotional speech. Towards a new generation of databases. *Speech Communication*, 40, 1/2, 33–60. doi:10.1016/S0167-6393(02)00070-5.



- Eek, A. (2008). *Eesti keele foneetika I*. Tallinn : Tallinna Tehnikaülikooli Kirjastus.
- Ekman, P. (1992). Are there basic emotions? *Psychological Review*, 99, 3, 550–553.
- El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: features, classification schemes and databases. *Pattern Recognition*, 44, 572–587.
- Elfenbein, H. A. (2013). Nonverbal dialects and accents in facial expressions of emotion. *Emotion Review*, 5, 90–96.
- Goudbeek, M., Goldman, J. P., & Scherer, K. R. (2009). Emotion, dimensions and formant position. *INTERSPEECH-2009*, 1575–1578.
- Harrington, J. (2010). *Phonetic Analysis of Speech Corpora*. Chichester: Wiley-Blackwell.
- Iida, A., Campbell, N., Higuchi, F., & Yasumura, M. (2003). A corpus-based speech synthesis system with emotion. *Speech Communication*, 40, 1–2, 161–187. [http://dx.doi.org/10.1016/S0167-6393\(02\)00081-X](http://dx.doi.org/10.1016/S0167-6393(02)00081-X).
- Iriondo, I., Alias, F., Melenchón, J., & Llorca, M. A. (2004). Modeling and synthesizing emotional speech for Catalan text-to-speech synthesis. In E. André, L. Dybkjær, W. Minker, & P. Heisterkamp (Eds.). *Affective dialogue systems: tutorial and research workshop; ADS 2004*, pp. 197–208. Berlin et al.: Springer.
- Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis, J. Haviland (Eds.). *Handbook of Emotion*, 2nd ed, pp. 220–235. New York: Guilford.
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129, 5, 770–814. <http://dx.doi.org/10.1037/0033-2909.129.5.770>.
- Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. In J. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research*. (pp. 65–135). New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780198529620.003.0003.
- Kamaruddin, N., Wahab, A., & Quek, C. (2012). Cultural dependency analysis for understanding speech emotikon. *Expert Systems with Applications*, 39, 5, 5115–5133. doi:10.1016/j.eswa.2011.11.028.
- Kienast, M., & Sendlmeier, W. F. (2000). Acoustical analysis of spectral and temporal changes in emotional speech. In R. Cowie, E. Douglas-Cowie, & M. Schroeder (Eds.), *Speech and Emotion: Proceedings of the ISCA workshop* (pp. 92–97). Newcastle, Co. Down.
- Koike, K., Suzuki, H., Saito, H. (1998). Prosodic parameters in emotional speech. *ICSLP-1998*, 679–682.
- Koolagudi, S. G., & Rao, K. S. (2012). Emotion recognition from speech: a review. *International Journal of Speech Technology* 15, 2, 99–117.
- Laukka, P., Neiberg, D., Forsell, M., Karlsson, I., & Elenius, K. (2011). Expression of affect in spontaneous speech: acoustic correlates and automatic detection of irritation and resignation. *Computer Speech & Language* 25, 1, 84–104.
- Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.
- McIntyre, G., & Göcke, R. (2006). Researching emotions in speech. *Proceedings of the 11th Australian International Conference on Speech Sciences & Technology*, 264–269.
- Mihkla, M. (2009). Kõnesüntees kui keele iseäralik võlu. J. Ross, A. Krikman (Toim.). *Teadusmõte Eestis (V)*, Humanitaarteadused (pp. 55–62). Tallinn: Eesti Teaduste Akadeemia.

- Mihkla, M., Hein, I., Kalvik, M.-L., Kiissel, I., Sirts, R., & Tamuri, K. (2012). Estonian speech synthesis: applications and challenges / Синтез речи эстонского языка: применение и вызовы. In A. E. Kibrik (Ed.), *Computational Linguistics and Intellectual Technologies, Papers from the Annual International Conference "Dialogue"* (pp. 443–453). Moskva: ПГТУ.
- Mihkla, M., Piits, L., Nurk, T., & Kiissel, I. (2008). Development of a unit selection TTS system for Estonian. In F. Čermak, R. Marcinkevičienė, E. Rimkutė, & J. Zabarskaitė (Eds.), *Proceedings of the Third Baltic Conference on Human Language Technologies: The Third Baltic Conference on Human Language Technologies* (181–187). Kaunas Lithuania, October 4–5, 2007. Vilnius.
- Montero, J. M., Gutierrez-Arriola, J. M., Palazuelos, S., Enriquez, E., Aguilera, S., & Pardo, J. M. (1998). Emotional speech synthesis: from speech database to TTS. *ICSLP-1998*, 3, 923–926.
- Moon, S.-J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America*, 96 (1), 40–55. <http://dx.doi.org/10.1121/1.410492>.
- Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, 93, 1097–108.
- Murray, I. R., & Arnott, J. L. (2008). Applying an analysis of acted vocal emotions to improve the simulation of synthetic speech. *Computer Speech and Language*, 22, 107–129. doi:10.1016/j.csl.2007.06.001.
- Nose, T., & Kobayashi, T. (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](http://www.apsipa.org/proceedings_2011/pdf/APSIPA189)>.
- Paeschke, A., Kienast, M., Sendlmeier, W. F. (1999). F0-contours in emotional speech. *Proceedings. ICPHS 99, San Francisco*, 2, 929–933.
- Pajupuu, H. (2012). Emotsioonid – kõnetehnoloogia olevik ja tulevik [Emotions – the present and the future of speech technology]. *Keel ja Kirjandus*, 8–9, 629–643.
- Pajupuu, H., & Kerge, K. (2006). Hingav süntesaator ja pausid tekstiliigiti [A breathing synthesizer: Breaks across different functional styles]. *Keel ja Kirjandus*, 3, 202–210.
- Pajupuu, H., Pajupuu, J., Tamuri, K., & Altrov, R. (2015). Influence of verbal content on acoustics of speech emotions. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, August 6–10, 2015, 1–5.
- Paulmann, S., & Uskul, A. K. (2014). Cross-cultural emotional prosody recognition: evidence from Chinese and British listeners. *Cognition and Emotion*, 28, 2, 230–244.
- Pell, M. D., Paulmann, S., Dara, C., Alasser, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: a comparison of four languages. *Journal of Phonetics*, 37, 417–435.
- Pribil, J., Pribilova, A., & Matousek, J. (2013). Comparison of formant features of male and female emotional speech in Czech and Slovak. *Elektronika ir Elektrotechnika, ISSN 1392-1215*, 19, 8. <http://dx.doi.org/10.5755/j01.eee.19.8.1739>.
- Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 2, 143–165. <http://dx.doi.org/10.1037/0033-2909.99.2.143>.

- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*, 1–2, 227–256. [http://dx.doi.org/10.1016/S0167-6393\(02\)00084-5](http://dx.doi.org/10.1016/S0167-6393(02)00084-5).
- Scherer, K. R. (2009). The dynamic architecture of emotion: evidence for the component process model. *Cognition and Emotion, 23*, 7, 1307–1351.
- Scherer, K. R. (2013). Vocal markers of emotion: comparing induction and acting elicitation. *Computer Speech & Language, 27*, 1, 40–58.
- Scherer, K. R., & Meuleman, B. (2013). Human emotion experiences can be predicted on theoretical grounds: evidence from verbal labeling. *PLoS ONE, 8*, 3, e58166.
- Scherer, K. R., Sundberg, J., Tamarit, L., & Salomão, G-L (2015). Comparing the acoustic expression of emotion in the speaking and the singing voice. *Computer Speech & Language, 29*, 1, 218–235. <http://dx.doi.org/10.1016/j.csl.2013.10.002>.
- Schröder, M. (2001). Emotional speech synthesis: A review. *Proceedings of the 7th European Conference on Speech Communication and Technology, 1. Eurospeech 2001, September 3–7, 2001* (pp. 561–564). Aalborg, Denmark: International Speech Communication Association.
- Schröder, M. (2009). Expressive speech synthesis: past, present, and possible futures. *Affective Information Processing*, 111–126.
- Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., & Gielen, S. (2001). Acoustic correlates of emotion dimensions in view of speech synthesis. *EURO-SPEECH-2001*, 87–90.
- Schuller, B., Batliner, A., Steidl, S., & Seppi, D. (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.
- Seppi, D., Batliner, A., Steidl, S., Schuller, B., & Nöth, E. (2010). Word accent and emotion. *Proceedings of Speech Prosody*, Chicago.
- Soto, J. A., & Levenson, R. W. (2009). Emotion recognition across cultures: The influence of ethnicity on empathic accuracy and physiological linkage. *Emotion, 9*, 6, 874–884. doi:10.1037/a0017399.
- Tamuri, K. (2007). Pausid ettelotud ilukirjandustekstis. *Magistritöö*. Käsikiri Tallinna Ülikooli humanitaarteaduste instituudis.
- Tamuri, K. (2010). Kas pausid kannavad emotsiooni? *Eesti Rakenduslingvistika Ühingu aastaraamat, 6*, 297–306.
- Tamuri, K. (2012). Intensity of Estonian emotional speech. *Human Language Technologies – The Baltic Perspective – Proceedings of the Fifth International Conference Baltic HLT 2012* (pp. 238–246). IOS Press: Amsterdam, The Netherlands.
- Tamuri, K. (2012). Kas formandid peegeldavad emotsioone? *Eesti Rakenduslingvistika Ühingu aastaraamat, 8*, 231–243.
- Tamuri, K. (2015). Fundamental frequency in Estonian emotional read-out speech. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics, 6* (1), erinumber / special issue: Papers from the conference “Finnic Languages, Cultures, and Genius Loci”, 9–21.
- Tamuri, K., & Mihkla, M. (2012). Emotions and speech temporal structure. *Linguistica Uralica, 3*, 209–217.
- Tamuri, K., & Mihkla, M. (2015). Expression of basic emotions in Estonian parametric text-to-speech synthesis. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics, 6*, 3, erinumber / special issue: Kõneuurimise suundi / Aspects of speech studies, 145–168.

- Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception and Psychophysics*, 27, 1, 24–27. <http://dx.doi.org/10.3758/BF03199901>.
- ten Bosch, L. (2003). Emotion, speech and the ASR framework. *Speech Communication*, 40, 1/2, 213–225. doi:10.1016/S0167-6393(02)00083-3.
- Thompson, W. F., & Balkwill, L.-L. (2006). Decoding speech prosody in five languages. *Semiotica*, 158, 1/4, 407–424. <http://dx.doi.org/10.1515/SEM.2006.017>.
- Tisljár-Szabó, E., & Pléh, C. (2014). Ascribing emotions depending on pause length in native and foreign language speech. *Speech Communication*, 56, 35–48. doi:10.1016/j.specom.2013.07.009.
- Toivanen, J., Waaramaa, T., Alku, P., Laukkanen, A.-M., Seppänen, T., Väyrynen, E., & Airas, M. (2006). Emotions in [a]: a perceptual and acoustic study. *Logopedics Phoniatrics Vocology*, 31, 43–48.
- Tolkmitt, F. J., Scherer, K. R. (1986). Effect of experimentally induced stress on vocal parameters. *Journal of Experimental Psychology: Human Perception and Performance*, 12, 3, 302–313. <http://dx.doi.org/10.1037/0096-1523.12.3.302>.
- Ververidis, D., & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication*, 48, 9, 1162–1181. <http://dx.doi.org/10.1016/j.specom.2006.04.003>.
- Wilting, J., Kraemer, E., & Swerts, M. (2006). Real vs. acted emotional speech. *Proceedings of Interspeech 2006 ICSLP*, 805–808.
- Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Deng, Z., Lee, S., Narayanan, S., & Busso, C. (2004). *An acoustic study of emotions expressed in speech*. INTERSPEECH-2004, 2193–2196.

## APPENDIX 1 / LISA 1

### Material used in studying pauses / Pausiurimuse materjal

#### *Joy / Rõõm*

*Passage 1 / Lõik 1:* Kui ma Otti esimest korda nägin, siis mulle tundus kohe, et selle inimesega ma abiellun. Ehkki Ott minu olemasolust midagi ei teadnud... Oli juhuseid, kui mõtlesin, et tahaksin teda näha, ja ootamatult nägingi! See ei olnudki nagu otsus, vaid äratundmine. Ott tundus kuidagi tuttav ja ligitõmbav. Tema esimesed sõnad mulle olid: „Häid pühi!“ Pärast seda oli mu suu mitu tundi kõrvuni – ma lihtsalt ei saanud seda sirgeks[!]<sup>63</sup>

*Passage 2 / Lõik 2:* Minu elu kõige olulisem inimene on minu tütar. Ma tegelikult ei tea, kuidas lapsi kasvatada... Olen kasvatanud teda tunde järgi. Igal hommikul ütlen talle, kuidas ma teda armastan ja et ma ei jäta teda mitte iialgi maha! Ma olen oma lapsega kogu aeg üdini aus. Räägin talle absoluutselt kõike, sest meie peres on ikkagi ainult kaks naist ja me elame naiste elu. Mingit ümbernurgajuttu ei ole. Kõrvalt vaadates on see ehk imelik, aga ma arvan, et meie peres see toimib[.]

*Passage 3 / Lõik 3:* Kui keegi paluks mul nimetada suursuguseid inimesi, siis ühena esimestest mainiksin ma Kadrit. Ta on suursugune oma hoiakutelt ja käitumiselt. Ja kõik see on tema juures täiesti loomulik! Kadri on minu jaoks midagi enam kui sõbranna või koostööpartner. Ta on isiksus, kellega ühes mõtteruumis olla on privileeg. Kadriga vesteldes tunned, et ka enda mõttelend läheb kõrgemaks[.]

*Passage 4 / Lõik 4:* Olen õpetaja Kristjanile palju võlgu. Tänu temale ma selle kooli üldse ära lõpetasin. Tõesõna, ma käisin ikka jube vähe koolis. Kogu aeg tuli midagi vahele. Ka pärast mitme kuu pikkuseks veninud kooli ei andnud õpetaja järele. Ta helistas koguni mu vanaemale ja ärgitas kooli tagasi tulema. Tagantjärele võin öelda, et püüdsin just seepärast kooliga hakkama saada, et õpetaja saaks minu üle uhke olla[.]

*Passage 5 / Lõik 5:* Allan on väga tasakaalukas ja rahulik, ta on perekeskne inimene. Sedasorti inimene, kes suudab ise kõõgimööbli ehitada. Võrreldes minuga elab ta märksa selgemat ja korrektsemat elu. Ta ei satu kunagi mingitesse hullumeelsetesse situatsioonidesse. Ta ei pruugi peaaegu kunagi alkoholi. Allan teeb geniaalset muusikat, minu arvates on ta üks paremaid Eesti lauljaid[.]

*Passage 6 / Lõik 6:* Maailmas ei ole ilusamat ja paremat asja kui armastus. Armastus teeb inimese tõeliselt kauniks. Armumine on edasiviiv jõud, ta paneb sind mõtlema. Ja kui kannatama paneb, siis las paneb. Siis on see ilus kannatus[.]

---

<sup>63</sup> The punctuation marks at the end of the passages did not belong to the test material. / Lõikude lõpus olevaid kirjavahemärke pole materjali hulka loetud.

*Passage 7 / Lõik 7:* Olen selle aja jooksul jõudnud mängida juba neljas filmis. Olen väga õnnelik uute pakkumiste üle. Aga praegune film on mulle kõige armsam. No saage aru! Ma tõesti armastan seda filmi, mulle väga meeldib see lugu. Ja ma armastan neid inimesi, kes selle filmiga seotud on. Kord soovisin öösel kell kolm tulist mannaputru. Ja viieteist minuti pärast oli taldrik tulise pudruga olemas! Öösel kell kolm! Mul tulid õnnest pisarad silma – millised inimesed siin on[!]

*Passage 8 / Lõik 8:* Tead noorena mõtlesin, et kuuekümmeselt on kõik möödas, kuid õnneks nii see ei ole. Arvan, et oleme nii vanad, kui end ise tunneme. Kui noorus on läbi, saab määravaks vaimus. Vananemine on loomulik protsess. Kui sellest aru saad, siis on see üks väga huvitav teekond. Vananemises pole midagi traagilist! Kui nii mõtled, siis on lihtsalt tore elada[.]

*Passage 9 / Lõik 9:* Meie kõigi argipäeva eesmärk võiks ju olla elada ja särada! Selle asemel et mõelda sellele, mis olemas, ja õnnelik olla, mõeldakse hoopis sellele, mida pole. Mul on olnud huvitav lapsepõlv, on kallid perekonnaliikmed, head sõbrad, toredad töökaaslased ja lisaks veel väga huvitav töö. Elu on teinud mulle kingitusi nii inimeste, reise kui ka palju muu näol. Ma olen selle kõige eest väga tänulik. Ja kui inimene on õnnelik, siis ta ka särab[!]

*Passage 10 / Lõik 10:* Kõige rohkem rõõmustab, et jalgpallurid tegid oma töö ära. Professionaalne esitlus – võitsime mängu esimese 20 minutiga. Riietusruumis valitseb pärast viimaseid võite enesekindlus, usk ja hea suhtumine. Aga Eesti koondist tuleb austada. Hästi organiseeritud, distsiplineeritud. Seda liini hoides võivad nad meid veel kunagi võita[.]

*Passage 11 / Lõik 11:* Olen eluaeg püüdnud vältida millestki sõltuvusse sattumist, aga siin ma nüüd olen – sõltuvuses oma abikaasast. Eks ma otsisin teda ka kaua. Hea et ta polnud teine-kolmas inimene minu elus. Sain enne Ennuga kohtumist selgeks, mida ma tahan ja mida mitte. Temas on väga palju sellist, mida ma tahan. Ja mõned omadused, mida ma vajan, kuigi võib-olla ei tahagi. Me tasakaalustame teineteist täiesti. Tunnen Ennust isegi siis puudust, kui päeval tööl olen. Pooljoostes lähen vahel koju[!]

*Passage 12 / Lõik 12:* Laura on elujõuline. Kui ta tegeleb tööga, mis teda innustab ja talle korda läheb, siis ta süüvib sellesse väga. Kohe väga põhjani. Laura puhul meeldib mulle väga tema eelarvamuste vaba suhtumine ümbritsevasse. Talle võivad ühtäkki väga ootamatud asjad huvi pakkuda. Ta tahab kõiksugu erinevaid asju proovida. Pole mingi ime, kui ta jälle mõnest uuest asjast vaimustusse satub[.]

*Passage 13 / Lõik 13:* Võin öelda küll, et minu lapsepõlv oli õnnelik. Minu arvates on kogu mu elu õnnelik olnud. Ma olen üks neist, kes tunnevad, et saatus hoiab kuidagi eriliselt. Olen üle elanud autoavarii ja tulekahju: need on olnud nagu meeldetuletused, et ei tohi hooletuks minna. Kõike võib juhtuda. Mõelge sellele! Minul on põhjust saatusele tänulik olla[.]

*Passage 14 / Lõik 14:* Uhiuus staadion on täiuslik koht. Ilus! See on maailma üks paremaid võistluskohti. Rajad, staadionile tulek, avarus – kõik on täpselt nii nagu peab. Ei suuda midagi paremat ette kujutada. See on õige koht tähtsaimate võistluste pidamiseks. Minu jaoks on sel staadionil tugev võlujõud – olen reisinud terves maailmas, kuid pole kunagi ühelgi teisel staadionil tundnud end nii rahuloleva ja õnnelikuna. Ja milline publik! Fantastiline staadion ja fantastiline publik! Suur tänu kõikidele häälekatele fännidele[!]

*Passage 15 / Lõik 15:* Möödunud nädala parim üllatus teleekraanil oli minu jaoks Kadri Kukk. Esmaspäeva hilisõhtul etendus Kadri reportaaž Afganistanist. Ja saade oli ka vaatamist väärt! Suurepärased esinejad, mõtestatud tekstid, voolav esitus, hea tempo. Tõsi küll, pildikeel polnud midagi „moodsat“. Aga kas peakski saates, kus on tõesti midagi ka vaadata, pilt tõmblema nagu mõnes sisutihjas noorteprogrammis? Ei peagi! Ma ei tunne Kadri Kukke isiklikult. Aga saates nägin tema silmis arukust ja soojust. Hea Kadri, ma soovin, et Sul elus hästi läheks[!]

*Passage 16 / Lõik 16:* Aga nüüd, daamid ja härrad, palume tähelepanu! Me tahame teile tutvustada õhtu staari! Võtke lahkesti vastu tema majesteet – kadakaga suitsetatud pardifilee! Aplaus! Tormilised ovatsioonid! Ovatsioonid jätkuvad! Publik möllab! Söögisaalist kostavad hõiked „Braavo!“ Õhus lendlevad roosiõied[!]

### ***Sadness / Kurbus***

*Passage 1 / Lõik 1:* Mul oli oma vanema vennaga harukordselt head suhted. Ta teenis Prantsuse Võõrleegionis. Ühel päeval tuli aga teade, et õppustel oli juhtunud... lihtsalt õnnetus. Venda ei olnud enam... Ma läksin liimist lahti... ei osanud midagi teha... See oli murdepunkt. Päevapealt tulin töölt ära. Ma ei saanud üldse püsti, pool aastat ei mäleta õieti midagi... Ma olin 22 ja äkki lõppes minu jaoks elu ära. See oli nii suur kaotus! Ma elasin vennaga kahekesi, sest mu vanemad olid juba ammu lahus. Ta oli sel hetkel ainuke inimene, kes oli alati seitse päeva nädalas 24 tundi minu jaoks olemas. Mu hing läks katki[...]

*Passage 2 / Lõik 2:* Viimasel ajal näib, et mu isa ei armasta mind. Miks? Sest ta karjub mu peale pidevalt. Ükskõik, mida ma teen, ikka pole ta rahul. Näiteks käsib nõud kappi panna, kui need pole nii, nagu tema tahab, saab vihaseks. Tean, et tal on tööl jamad, kuid ikka jääb tunne, et äkki olen mina süüdi või et ta tahab meie juurest ära minna[...]

*Passage 3 / Lõik 3:* Nägin, kuidas Nõmme linnaosa vanem eilsel Rabajooksul kokku kukkus. Mees on kõigest 26-aastane! Ja seal ta siis lamas... Oli täitsa külili. Nagu surnukeha. Kurb oli see, et abi polnud kusagilt võtta. Mõni meditsiinitöötaja võiks ju metsas raja ääres ikka olla. Jumal tänatud, et pealtvaatajad oskasid meest turgutada ja ise kiirabi välja kutsusid[.]

*Passage 4 / Lõik 4:* Aastate jooksul olen liiga palju näinud, kuidas säravad anded selles hallide seintega majas on lämbunud... Keegi pole läinud sellest majast hea tundege. Ma ei suuda kunagi unustada, kuidas ETV esimene diktor Kalmer Tennosaar saadeti piimakombinaati väravavalvuriks... Näib, justkui ei hooliks see maja inimestest, kes oma südame sinna matnud. Ega olukord ajaga paremaks pole läinud. Ja ei usu mina enam, et uus juhtkondki suudab seda needust murda[...]

*Passage 5 / Lõik 5:* Minu sportlastee kõige suurem pettumus on Ateena olümpiamängudel. See pani mõtlema, kas üldse suudan suurvõistlustel läbi lüüa. Olin ju nii kindel, et suudan! Ei teagi, mis sellel võistlusel juhtus... Olin 33. kohaga lõpetamisest nii löödud, et selle nahka läks ka järgmine hooaeg. Vahetasin siis treenerit, kuid tegelikult oli viga ilmselt minus, mitte juhendajas[.]

*Passage 6 / Lõik 6:* Kõige rohkem kardan seda, et ma ei saa ennast tõestada ega näidata, milleks olen võimeline. Et mulle ei anta seda võimalust. See hirmutab mind kõige rohkem. Ma ei karda seda, et äkki jään vabakutseliseks näitlejaks. Sel ei ole nii suurt tähtsust. Võimaluse andmine on kõige olulisem. Kuid praegu ei ole mulle veel võimalusi antud. Minuga ei olda siirad[.]

*Passage 7 / Lõik 7:* Aga tänaseks ma näen, et mind ei ole selles saalis enam vaja. Ma ei tunne, et minust midagi sõltuks. Ma ei näe, et ma millelegi kaasa aitaks. Minu nägemuse kohaselt on tänane saal kummitempel. Mul on kahju teile öelda, aga ma olen näinud läbi 15 aasta selles saalis suurepäraseid kõnesid. Ma ei taha öelda, et siin nüüd ei ole suurepäraseid kõnelejaid. On küll. Aga need kõnelejad enam ei kõnele – kõik otsused on kusagil varem valmis tehtud. Ja siinne parlament täidab vaid käske. Mul on teile suur palve. Palun, ärge jätkake samaviisi! Muidu veenabki keegi ühiskonna ära – milleks nii palju inimesi nii väikeste otsuste tegemiseks[?!]

*Passage 8 / Lõik 8:* Meie peres ei ole kunagi probleemidest räägitud. Meil polegi perekonda kui sellist – elu on nagu hotellis. Elame küll ühe katuse all, kuid omavahel ei suhtle. Kõige tähtsam asi korteris on televiisor. Vanemad tulevad koju ööseks vaid magama. Aga vahel juhtub, et nad ei jõuagi koju[...]

*Passage 9 / Lõik 9:* Minu jaoks ei ole Tiina nagu surnud, lihtsalt kadunud. Üle ei saa oma lapse kaotamise valust mitte kunagi, ehkki selle teadmisega võib aja jooksul õppida elama. Aga see on ebanormaalne, kui lapsed oma vanematest kauem ei ela[!]

*Passage 10 / Lõik 10:* Eks proovige siin 600 loomaga toime tulla, kui töölisi ei ole! Miks ei koolitata meile noori abiks? Tahaks ju tõesti, et Eesti põllumees söödaks ise oma rahva ära. Me võiksime kartulit, piima, liha, kõike toota, aga meil pole ju töökäsi. Keegi siia tööle ei tule, ütlevad, et see on kolgas! Ma ei saa ju kolida oma laudaga Tallinna külje alla, pean leppima sellega, et elan ääremaal[.]



*Passage 11 / Lõik 11:* Elan koos oma invaliidist tütreaga nüüd Tallinnas ning tunneme end siin palju mugavamalt kui Narvas. Siin pole enam paljusid probleeme, mis tekitasid varem peavalu. Ka rongidega on kõik lihtsam. Varem olin ma sunnitud last ravile viies ta ratastoolist otse räpasele perroonile istuma panema: kõigepealt tõstsin välja ratastooli ning seejärel lapse. Isegi ratastooli või kepiga liikuvad täiskasvanud ei saa rongi kasutada. Vagunisse ronimine võtab nii palju aega, et rong võib vahepeal ära sõita. Sageli ei tule teised reisijadki appi – paremal juhul hoiavad kotti. Ka bussidega on raske. Tavaliselt kasutatakse isiklikku transporti või renditakse auto – teisel korral ei taha enam keegi rongiga sõita[.]

*Passage 12 / Lõik 12:* Juhani kaotamine on nii suur tühjus, et seda on raske kirjeldada. Nagu oleks Eestisse jäänud üks suur auk, kuhu on koos Juhaniga langenud suur osa meie inimlikkusest. See on suur, suur auk. Ta andis edasi nii palju inimlikkust, armastust inimese vastu. Juhani jaoks ei olnud elu kunagi kerge. Ta lasi läbi oma südame kogu Eesti saatuse ja oma lähedaste saatuse, kõik need vaevad ja häbid – ning jäi suureks oma inimlikkuses ja mõistmises. Tema vaim mõjutab meid veel väga pikalt. Tema vaim mõjutab meid veel väga pikalt. Tema vaim mõjutab meid veel väga pikalt[.]

*Passage 13 / Lõik 13:* Mul ei ole enam kodu. Korter öeldi üles. Ega mul raha ka õieti ei ole. Ma ei tea, kuidas edasi... Olen ikka väga täbaras olukorras, aga parem on nüüd midugi kui vanglas. Tervis on halb. Liigesevalu piinab, käedjalad kisuvad krampi. Üsna õudne on ikka olla. Tulevik on esialgu täiesti lootusetu[.]

*Passage 14 / Lõik 14:* Mis see neli aastat teinud on... Olin enne kooli lõbus poiss, tunduvalt mängulisem kui nüüd. See kool on pannud peale mingid piirid. Piirid, mis on mul midagi kinni pannud. Ma ei julge teha asju, mida varem julgesin. Mida aeg edasi, seda rohkem kontrollin ennast. Mida ma tohin teha ja mida ei tohi. Enam ei tea, mis on õige ja mis on vale. Ma ei julge enam midagi kursusekaaslaste ja õppejõudude ees teha. Ennast näidata. Mul pole enam seda esinemisjulgust, mis oli enne kooli[.]

*Passage 15 / Lõik 15:* Ei teagi, kuidas saaks inimesi lapse sünniks ette valmistada... Kas me tõesti enam ilma koolituseta mitte millegagi hakkama ei saa? Meeskonnatööd ja müümise kunsti õpime kursustel, lastekasvatust õpime kursustel... Kuhu on jäänud terve talupojamõistus?! Tundub, et vanemarolli on astunud nii öelda saajate põlvkond – need on siiani lihtsalt läbi elu kulgenud. Nad on harjunud tegema vaid seda, mis nendele endile meeldib ja kasulik on. Raske on mõista selliseid lapsevanemaid. Või äkki on nad jäänudki ise lasteks, kelle ümber kogu maailm tiirlema peab[?]

## *Anger / Viha*

*Passage 1 / Lõik 1:* Ma ei saa aru, kuidas saavad ministrid kogu aeg rääkida, et hinnatõus tuleb väike, aga palgatõus kompenseerib kõik mured. Nii ju ei lähe! Ma ei tea kedagi, kelle palk tõuseks. Vaid riigiametnikele lubatakse kahekümneprotsendilist palgatõusu. Kuidas nii hakkama saada[?!]

*Passage 2 / Lõik 2:* Paneb imestama, miks meie poliitikud ärimeeste annetusi nii kergekäeliselt vastu võtavad. Jutt sellest, et tegemist on isikliku lapsepõlvesõbraga ja hea tuttavaga, on täiesti süüdimatu. Kuidas poliitikud küll nii müüdavavad on?! Siiber on sellest, et kõik läheb vanamoodi edasi – kõik ongi nii ja midagi ei muutu. Nagu öeldakse, ei au ei häbi[!]

*Passage 3 / Lõik 3:* Esmaspäeval oli halvim päev Tartu liikluses. No kellel tuli mõttesse kõik Tartu suuremad sillad kinni panna?! Täiesti mõistetamatu! Kui tavaliselt saab Annelinnast kesklinna kümne minutiga, siis nüüd pidin tund aega ummikus passima. Olen 30 aastat autoga sõitnud – ei ole veel sellist asja juhtunud, et korraga mitut silda asfalteerima hakatakse. Tartus ei peaks olema nagu Leningradis. Seal on ka tavaks korraga kõik sillad kinni panna, ja inimestel kästakse siis lihtsalt kodus püsida[.]

*Passage 4 / Lõik 4:* Meie ühiskond ei hinda ei au ega austust. Näiteks, täna hommikul kuulsin, et üks noorpoliitik, B-klassi särav täht, on kõrvaldanud suure summa raha. Ja selle raha on ta kulutanud lihtsalt hilpudele... Meie ühiskonnas peab oma eesmärkide saavutamiseks oleme ilus ja hästi riides. Peab olema klantsajakirjade pildil. See on see, mida ühiskond tahab näha. Selleks võib kasvõi suure hulga võõrast raha pihta panna. Ja seda ei panda pahaks! Vanasti oli lihtne, siis nimetati seda selge sõnaga – VARGUS[!]

*Passage 5 / Lõik 5:* Kui sõjas on paha, miks sinna minna? Peale raha ei leia ma mingit motiivi. Ja ega see raha ka teab kui kõva ei ole. Kaif ja seiklus, mis muud. Võitlesid nad seal Afganistanis ja Iraagis Eesti Vabariigi eest või ei, härra reservkolonelleitnant? Sotti ei saa! Palun mõelge see Eesti sõjajõu ja sõjaväe asi veel kord läbi[!]

*Passage 6 / Lõik 6:* Mind häirib Eestis teenindus. Halb pole üksnes teenindamise tase, vaid teenindajate hoiak. Üks teenindaja ütles ajalehes, et tal on kogu aeg halb tuju, sest kõrtsis käib palju soomlasi. Või et Pärnus on nüüd sügisel hea ja rahulik olla – ei kuule enam soome keelt. Eestis ei ole aru saadud, et turist otsib siit teeninduskultuuri – elamust, mis muudaks õhtu eriliseks. Eestis ei hinnata teenindaja ametit. Kui mees, töötab restoranis, siis mõeldakse, et ju tal on mingi viga küljes. Eestis pole olnud kunagi teenindav ühiskond, ei olnud Nõukogude ajal, ei ole selleks ka nüüd veel saanud. Loodan, et lubatud majanduskrahhist tuleb Eestile hea ja vajalik õppetund[.]

*Passage 7 / Lõik 7:* No mis sa tahad nendest naistest saada?! Suurem jagu, kes mehele trügivad, on ju liiga paksud. Ja uimased nagu haid. No on, noh! Ma ei hakka siin midagi keerutama. Ja pealegi nad ju valetavad oma vanuse kohta –

mitte paar-kolm aastat, vaid kohe väga palju. Kannavad ees musti päikeseprille ja algul ei saagi aru, kui vanad nad siis ikkagi on. Aga seda karmim on tõde hiljem! Ja nad on liiga laisad! Magavad kella 9–10ni ja ei taha tööd teha. Käivad siin ringi nagu mingid staarid. No mida ma teen nendega?! Parem elan üksi edasi, kui selliseid söödan ja poputan[.]

*Passage 8 / Lõik 8:* Paluks mitte ühtegi piiksu kriisist ja katastroofist! Vähemalt mitte enne, kui Eesti peaks Andorralt kolaka saama. Põhjendamatult kõrged ootused võivad tõepoolest kriisini viia. Juba ei olda rahul uue peatreeneri mängujoonise ja kes teab millega veel. Kujutan ette, kui keeruline oleks palluritel lühikese aja jooksul hakata ümber õppima. Loodetavasti antakse taanlasele võimalus ennast tõestada. On andestamatu, et tema hollandlasest eelkäia sattus just eestlaste kärsituse ohvriks[.]

*Passage 9 / Lõik 9:* Jutt tuleb tänapäeva Eesti meesinimestest. Väga sageli võib kohata pikajuukselisi, isegi patsiga mehi. Habemeajamisega ei näi tegelevat vähemalt kolmandik kesk- ja vanemaealistest. Karvasena juhitakse riiki ja riigiasutusi, samuti ülikoole ja gümnaasiume. Karvaseid isendeid võib kohata ettevõtetejuhtide hulgas ning nüüdisaegne põllumajanduski pole neist vaba. Muusikuil ja kunstnikel lausa meeldib karvus olles kaaskodanikega suhelda. Lausa arusaamatu, miks inimesed tahavad olla palju hirmsamad, kui nad tegelikult on[?!]

*Passage 10 / Lõik 10:* Ma ei saa aru, kuidas saavad ministrid kogu aeg rääkida, et hinnatõus tuleb väike, aga palgatõus kompenseerib kõik mured. Nii ju ei lähe! Ma ei tea kedagi, kelle palk tõuseks. Vaid riigiametnikele lubatakse kümneprotsendilist palgatõusu. Kuidas nii hakkama saada[?!]

*Passage 11 / Lõik 11:* Mind väsitab moodne kohustus olla ikka ja alati roosa ning rõõmus. Nii enda kui ka teiste puhul. Ühel suvisel lõunasöögil ajas see mind lausa närvi. Üks väliseestlasest amerikaniseerunud daam hüüatas iga minuti tagant: „Otepää on super! Pärnu on võrratu! Haapsalu – sõnu ei jätku! Tartu on maailma parim!“ Kõik oli nii isikupäratu, et ma ei saanudki aru, kas ta oli üldse kuskil käinud[.]

*Passage 12 / Lõik 12:* Katsume kindlasti kohe alguses värava lüüa ega kavatse anda eestlastele mingit võimalust! Ei mingit lootusekiirt! Peame olema halastamatud ja nad lihtsalt kohe alguses ära tapma. Pealegi jäävad nad meile kõiges alla: jõulisuses, taktikas, kiiruses ja kehavõitluses. Ütlen teile, nad on kõnnid. Kobad, noh[!]

*Passage 13 / Lõik 13:* Praegune tants alkoholipiirangute ümber muutub juba labaseks ja naeruväärseks tsirkuseks! Soovitan Kristiine Prismale – astuge Keskerakonda, siis on kõik lubatud. Kes tahab juua, see joob ikka. Arvan, et keegi ei saa ega suuda joomarlust käskude ja keeldudega pidurada. Organiseerige noortele parem vaba aja veetmise võimalusi[.]

*Passage 14 / Lõik 14:* Söögikohta tullakse ju ometigi sooviga saada värsket toitu, ilma et seda ise valmistama peaks. Mitte keegi ei soovi eilseid jahtunud, ja siis taassoajendatud pannkooke. On arusaamatu, kas toidukohad püüavad nõnda kokku hoida või on tegu lihtsalt mõttelaiskusega. Kui kokad pole südamega asja juures, tunneb klient seda kohe ja enam tagasi tulla ei taha[.]

*Passage 15 / Lõik 15:* Eilsel kontserdil suitses Padar nagu korsten. Turvamehed ei teinud rokkarile isegi märkust! Kui mina oleksin rahva hulgas suitsu teinud, oleks mind klubist välja visatud. Mis ajast tohivad avaliku elu tegelased niimoodi reegleid rikkuda?! Pealegi on see lihtsalt tobe – ma ei leia, et selline lapsikus Padarist suurema staari teeks[.]

*Passage 16 / Lõik 16:* Olen täiesti šokeeritud meie riigikogu liikmete häbematust suhtumisest maksumaksja rahasse. Seda kasutatakse pidevalt oma eralõbudeks. Niisugune asi peab lõppema! Tuleb taotleda, et need riigikogulased tagasi astuksid, sest meie – maksumaksjad – oleme neis pettunud – kuigi meie ju neid valisime. Niisugused inimesed, kellele töö ei olegi nii tähtis kui tantsimine, võiksidki jääda tantsima. Need riigikogulased, kes tõesti on huvitatud rahva ja riigi heaolust, tegeleks siis selle raske ja tõsise tööga[.]

*Passage 17 / Lõik 17:* Natukene punast veini tahaks ka ju toidu kõrvale. Saab. Aga ainult natukene. Oleme punast veini näinud küll, aga nii vähe mitte kunagi. Klaasipõhja valatakse seda vahest kolm sentimeetrit. Ainult tirtsukese rohkem, kui enesest lugupidav sommeljee mekkimiseks pakub. Seitsekümmend krooni, palun. Aitäh. Eh sõbrad, nii see ikka ei käi! Nüüd lähme vinoteeki, et juua klaasike HEAD veini ja järele mõelda: mis asi see nüüd kõik kokku ikkagi oli[...]

*Passage 18 / Lõik 18:* Ma olin väga noor sportlane, kui jõudsin meedia huvi-orbiiti. Ajakirjanikud hakkasid päris kiirelt mind ära kasutama ja tasapisi aina jultunumalt. Asi läks tõesti mitu korda liiale – varrukast sikutamised, nõudmised ja suvalisel ajal tehtud telefonikõned kasvasid üle pea. Olen pannud tähele, et kõikvõimalikke pilte minust maalivad reeglina sellised „tarkpead“, kes on tippspordist väga-väga kaugel ja kes pole minu ega mu treeneriga kunagi ühtegi sõna vahetanud. Olen endiselt arvamusel, et sportlase isiklik mobiil-telefoninumber ei ole mõeldud kõikidele ajakirjanikele enesestmõistetavaks kasutamiseks[.]

## APPENDIX 2 / LISA 2

### Material used in studying formants, speech tempo, intensity and fundamental frequency<sup>64</sup> / Formantsageduste, kõnetempo, intensiivsuse ja põhitooni uurimismaterjal<sup>65</sup>

#### *Joy / Rõõm*

- 51%<sup>66</sup> Allan on väga tasakaalukas ja rahulik, ta on perekeskne inimene.  
51% Korraldus on ikka väga „kõva“.  
52% Krissu on olnud kogu aeg kange iseloomuga tüdruk.  
53% Katsume kindlasti kohe alguses värava lüüa ega kavatse anda eestlastele mingit võimalust!  
53% Mõelge sellele!  
53% Räägin talle absoluutselt kõike, sest meie peres on ikkagi ainult kaks naist ja me elame naiste elu.  
55% Appi, kuradi magus on!  
55% Ometi, kui suurde poodi minna, on igal tootjal ka vähem magus või – elagu ime! – lausa hapukas variant toote portfellis.  
55% Samas, kõrgharidusega mehed elavad sama kaua kui naised.  
55% Suurhalli väljakul tegutses vaid üks meeskond, Tartu Rock.  
55% Ta ei satu kunagi mingitesse hullumeelsetesse situatsioonidesse.  
57% Tunnistan, et olin pessimist ja jõudsin juba Mardilegi kuulutada, et me ei võida olümpial ühtegi medalit.  
57% Tõesõna, ma käisin ikka jube vähe koolis.  
57% Täna öölaulupeo nimeks on „Märkamisaeg“.  
58% Ta tahab kõiksugu erinevaid asju proovida.  
58% Talle võivad ühtäkki väga ootamatud asjad huvi pakkuda.  
60% Aga ega seal suurel laval käimine kedagi ka ära tapa.  
60% Oma rahva usaldust ei tohi petta!  
60% Varem oleme tagantjärele ikka leidnud esinejaid, kes siia väga hästi ei sobinud.  
63% Kodune võistlus on sootuks teistsugune.  
63% Mees, sa ei tea, kui kiiresti ma saen!  
63% Õige – ajaloo prügikastis!  
64% Kui noorus on läbi, saab määravaks vaimsus.  
64% Mees, sa ei tea, kui kiiresti ma saen!  
64% Mõelge sellele!

<sup>64</sup> The number of Corpus sentences used for research was different for each acoustic parameter, as the Corpus was constantly growing.

<sup>65</sup> Uurimismaterjalina kasutatud korpuse lausete hulk oli iga akustilise parameetri puhul erinev (korpus täienes pidevalt).

<sup>66</sup> The rate of Listening identification for the emotions or neutrality. / Emotsioonide või neutraalse kõne tuvastusprotsent kuulamise järgi.

- 64% See ei olnudki nagu otsus, vaid äratundmine.
- 64% Selleks on üks hea põhjus.
- 65% Ja kõik see on tema juures täiesti loomulik!
- 66% Kui ta tegeleb tööga, mis teda innustab ja talle korda läheb, siis ta süüvib sellesse väga.
- 66% On kohe näha, et ta tunneb võistlemisest rõõmu.
- 66% Saab.
- 66% See, mis juhtus, oli uskumatu!
- 66% Ta vaatab spordile nüüd hoopis teise pilguga kui varem.
- 66% Öösel kell kolm!
- 67% Kord soovisin öösel kell kolm tulist mannaputru.
- 68% Olen õpetaja Kristjanile palju võlgu.
- 69% Arvan, et oleme nii vanad, kui me end ise tunneme.
- 69% Kui noorus on läbi, saab määravaks vaimsus.
- 69% Ma ei karda seda, et äkki jään vabakutseliseks näitlejaks.
- 69% Tunnen Ennust isegi siis puudust, kui päeval tööl olen.
- 71% Ma ei jõua seda masu ära kiita.
- 71% Oleks nii siiras.
- 71% Paar vene lehte on juba helistanud ja palunud toetust Sotšile.
- 72% Seekord sellist küsimust ei teki!
- 73% Pooljoostes lähen vahel koju!
- 75% Ta ei pruugi peaaegu kunagi alkoholi.
- 75% Tagantjärele võin öelda, et püüdsin just seepärast kooliga hakkama saada, et õpetaja saaks minu üle uhke olla.
- 76% Võrreldes minuga elab ta märksa selgemat ja korrektsemat elu.
- 77% Alati tunneb ta sügavat huvi igasuguste pisiasjade vastu – Eesti hari-  
duse, Eesti inimese vastu, mida praegu loetakse, mida praegu kirjuta-  
takse, kuidas Eesti elu praegu läheb, kuidas näeb välja Eesti raha.
- 77% Täpselt sama moodi oli mu enda lapsepõlves, kõik kordub.
- 78% Saatsin nad seenele!
- 78% Uus trall hakkab pihta Sotši talimängudega.
- 80% Aga ainult natukene.
- 80% Meie kõigi argipäeva eesmärk võiks ju olla elada ja särada!
- 80% Selline värin on sees.
- 80% Siis on see ilus kannatus.
- 82% Ta helistas koguni mu vanaemale ja ärgitas kooli tagasi tulema.
- 84% Ehkki Ott minu olemasolust midagi ei teadnud...
- 84% Vananemises pole midagi traagilist!
- 88% Aga Eesti koondist tuleb austada.
- 88% Tegelikult tunnevad kõik peale minu end väga mõnusalt.
- 90% Aga – ega võitja nimi seekord nii tähtis ei olegi.
- 90% Võtame seda kui noorte muusikute ajateenistust.
- 92% Lähme otsime!

### ***Sadness / Kurbus***

- 51% Esmalt vaatasin ta palatikaaslast ja sel ajal, täiesti ootamatult, virutas rohtu nõudnu mulle jalaga ribidesse.
- 51% Kas läheb veel kuuskümmend aastat?!
- 51% Ma küsin: mille eest meid sinna viidi?
- 51% Sellel on ju pidevalt probleeme – kord on neil bussid katki ja puuduvad juhid, siis kurdavad nad vähese dotatsiooni üle.
- 51% Tänapäeva õpetaja on nagu ori vanas Egiptuses, ilma igasuguste õigusteta.
- 52% Ma ei oskagi öelda, miks.
- 52% Sageli ei tule teised reisijadki appi – paremal juhul hoiavad kotti.
- 53% Juba ei olda rahul uue peatreeneri mängujoonise ja kes teab millega veel.
- 53% Meie ühiskond ei hinda ei au ega austust.
- 53% Mida ma tohin teha ja mida ei tohi.
- 53% Võimaluse andmine on kõige olulisem.
- 54% Ikka ja jälle jõutakse üheskoos tõdemuseni, et meie riigis kehtivad seadused vaid paberil.
- 54% Olen kindel, et ülemus tegi seda kõike ainult kiusamiseks.
- 54% Ütlen veel kord: kui saja lapse koolipõlv nahka pannakse, et nende arvelt miljoneid teenida, siis ei vääri see lugupidamist.
- 55% Kui arvestada linna senist laenukoormust ja lisada juurde laenatav 500 miljonit krooni ja erafirmade kaudu miljardite eest ehitatavad koolimajad, siis ongi Savisaar suutnud linna pankrotti ajada.
- 55% Küünlaid on?
- 55% Ma ei taha pealkirju „Nad langesid kodumaa eest“.
- 55% Ta mureseb väga nii Venemaa kui Eesti ühiskonna pärast.
- 55% Teised sõjamehed temaga aga ühe laua taga enam istuda ei oska või ei taha.
- 56% Tuleb välja, et ma olen tööandja oma!
- 57% Aga mis me räägime uutest teedest: olemasolevate tänavate olukord on jube.
- 57% Enam ma niimoodi laste peale häält ei tõsta, kui vanasti.
- 57% Enne ma nii oma elu ei väärtustanud.
- 57% Eriti hullud on toiduainete reklaamid, kus vastav produkt värvitakse isuäratavamaks.
- 57% Haridust peaks saama mitte vähema vaevaga, vaid teisiti.
- 57% Ja sööstis mööda tänavat lumme.
- 57% Sest vaadake, mis ümberringi toimub.
- 57% Siin pole enam paljusid probleeme, mis tekitasid varem peavalu.
- 57% Vaata et hoopis uut sõda ei puhke!
- 57% Vana Pärnu maantee ja Soo tänav on kõige hullemad.
- 58% Me ei räägi enam olukorrast, kui liinilt kaob paar bussi, vaid et päevas jääb ära 20–30 väljumist.

- 58% Miks ülemus mulle ei helistanud, vaid niimoodi nuhkima pidi?
- 60% Ma küsin, mille eest meid sinna viidi.
- 60% Ta andis edasi nii palju inimlikkust, armastust inimese vastu.
- 60% Tahan minna tööle tagasi, aga ülemus ütles, vabandust, Sulle pole enam kohta!
- 61% Enne ma nii oma elu ei väärtustanud.
- 61% Katkev internet, tarduv ja ruuduline telepilt, kättesaamatu tehnilise toe telefon, mitte töötav koduleht...
- 61% Kord on Harjumaa bussiliikluses täiesti käest!
- 61% Mul olid oma vanema vennaga harukordselt head suhted.
- 61% Tundsin, et suremine oleks liigvarane ja vale.
- 62% Mida aeg edasi, seda rohkem kontrollin ennast.
- 63% On pühapäev.
- 64% Arukamalt, inimlikumalt.
- 64% Kas midagi jäi kooli maha?
- 64% Laps hakkabki uskuma, et see on tõsi – seda ütleb ju üks tähtsamaid inimesi tema elus.
- 64% Mis see abielu muud kui pliidi ees seismine on?
- 64% Need hoolivad, igapäevased küsimused kõlaga sagedamini kui seni.
- 66% Kristina ei tulnud ju suurde suusatamisse tagasi selleks, et võistelda kümnendate kohtade pärast.
- 66% Kui satun juuksuris Kroonikat lugema, siis pärast seda on küll süda paha.
- 66% Ma küsin: mille eest meid sinna viidi?
- 66% Minu noorimal lapsel on olnud teistega võrreldes tunduvalt kergem.
- 66% Minu poeg on 190 cm kõrge, 120 kg raske, on praegu 16 aastat vana.
- 66% Poodides on enamik lihast valesti lõigatud ja müüjad on tihti asjatundmatud.
- 66% Raudselt ei tule siin 30 aastat mingit korda.
- 66% See oli alanduste ja solvangute jada, mida kogesid ühel või teisel kombel kõik kongressi delegaadid ja vaotlejad.
- 66% Või tema silmad muretsevad, kui räägime kodututest lastest, kui räägime aidsist.
- 66% Ülikooli õppejõu töötasu ulatub murdosani sellest.
- 70% Laps hakkabki uskuma, et see on tõsi – seda ütleb ju üks tähtsamaid inimesi tema elus.
- 70% Sa pole midagi väärt, keegi pole sind tahtnud, sa oled rumal ja küündimatu.
- 70% See oli ikka õudne, mida inimestega tehti.
- 71% Kohustused on ikka samad.
- 71% Küsisin, kas sa kaotasid midagi.
- 71% Pilt on, aga häält ei ole.
- 71% Pole mõtet minna sinna riiki ülikooli, kus professor on nii väsinud.
- 72% Mitte aga saada õigust olla hall anonüümne mass.
- 73% Ja siis ma porisengi.



- 74% Need ei pääse ju üldse liikuma!
- 76% Küsisin, kus on käpikud.
- 76% Oma häda pole kellelegi kaevata ja Elionist öeldakse, et jah, teame ja teeme korda nii kiiresti kui suudame.
- 76% See on suur, suur auk.
- 76% Ükskõik, mida ma teen, ikka pole ta rahul.
- 77% Ma ei taha oma poega sinna lasta.
- 78% Ma ei saa ju kolida oma laudaga Tallinna külje alla, pean leppima sellega, et elan ääremaal.
- 78% Mida ma tohin teha ja mida ei tohi.
- 78% Mis see neli aastat teinud on...
- 80% Miks surutakse meile peale just sellist ühiskonda, kus puuduvad nimed, kus puudub vastutus, kus puudub kodanikutunne.
- 83% Ja kui tema ära pöörab või pööratakse, mis siis saab?
- 83% Ja mis seal salata, isadel kah...
- 83% Kui mulle oleks varem öeldud, et inimene suudab ka niimoodi elada, siis ma poleks seda uskunudki.
- 83% Pidin sellest loobuma, sest lihtsalt ei saa enam head liha.
- 83% Ta oli käinud just poisil sõjaväes külas.
- 84% Elame küll ühe katuse all, kuid omavahel ei suhtle.
- 84% Meie peres ei ole kunagi probleemidest räägitud.
- 84% Raske lõik.
- 85% Ma ei uskunud oma silmi.
- 87% Mul tuli Siberis veeta kokku 17 aastat!
- 88% Kaks aastat veel ja tal on lipu alla minek.
- 90% Imetlusest, pettumusest, hoolivusest, kurbusest.
- 90% Päevapealt tulin töölt ära.
- 92% Tema vaim mõjutab meid veel väga pikalt.
- 93% Neil pole ju raha!
- 96% Ta lasi läbi oma südame kogu Eesti saatuse ja oma lähedaste saatuse, kõik need vaevad ja häbid – ning jäi suureks oma inimlikkuses ja mõistmises.
- 96% Ta oli sel hetkel ainuke inimene, kes oli alati seitse päeva nädalas 24 tundi minu jaoks olemas.
- 100% Ma väga loodan, et tal jätkub aru sinna mitte minna.

### ***Anger / Viha***

- 51% Ei peagi!
- 51% Järjekordsed valimised on mõne poliitiku pea täiesti pööraseid mõtteid täis tuupinud.
- 51% No kui ei sobi, siis ei sobi!
- 52% Eriti paigast ära oli esimesel päeval teenindus.
- 53% Kelle jõud, selle õigus.
- 53% Kõik me tahame kuskile minna ja sinna tervelt kohale jõuda.

- 53% Minu lõogist ei jäänud kohe kindlasti mitte mingisugust sinikat või vigastust.
- 53% Oleme punast veini näinud küll, aga nii vähe mitte kunagi.
- 53% On küll.
- 53% Raske on mõista selliseid lapsevanemaid.
- 53% Suudate seda ette kujutada?!
- 53% Tahaks ju tõesti, et Eesti põllumees söödaks ise oma rahva ära.
- 53% Tartus ei peaks olema nagu Leningradis.
- 53% Vaid riigiametnikele lubatakse kahekümneprotsendilist palgatõusu.
- 54% Keel kaob.
- 55% Aga ikkagi.
- 55% Ja siis ma lõugangi.
- 55% Me võiksimme kartulit, piima, liha, kõike toota, aga meil pole ju töökäsi.
- 55% Mida näevad siin minu lapsed?!
- 55% Nokaut!
- 55% Reklaamidesse ei saa tõsiselt suhtuda.
- 55% See võib anda sügisel üpris valusad vitsad.
- 55% Uskuge, kes uskuda suudab.
- 56% Ma ei taha öelda, et siin nüüd ei ole suurepäraseid kõnelejaid.
- 56% Meie suusaliidus ei muretse, kas kergejõustiku meistrivõistlused toimuvad.
- 57% Jah, lahutada ja uut kaasat otsida on probleemide ilmnemisel alati lihtsam kui suhet hoida.
- 57% Joodik tuuakse tänavalt otseteed pidi, kiirabiga kohale.
- 57% Kolm-neli aastat tagasi ei huvitanud kedagi, kust raha tuleb.
- 57% Kui astun kellelegi varbale, vabandab tema, et jäi mulle ette.
- 57% Kustkohast korruga säärane tusk ja talveuni.
- 57% Millal Sina viimati märkasid, et Eesti on vaba?
- 60% Eh, sõbrad, nii see ikka ei käi!
- 60% Kaste on lihtne joonistada, aga inimesed tuleb tööle panna.
- 60% Reaalset tulemust see nagunii ei anna.
- 60% See on nagu alasti inimene.
- 60% Õppisin ära, et sellise kambaga ei saa oodata – põruta, kuni jõuad!
- 60% Ükskõik mis teenindaja-ametisse ei saa palgata oskamatut inimest.
- 61% Ja ma ei näe, et olukord hakkaks paranema!
- 61% Kõik on ju juriidiliselt korrektne!
- 61% Mees on kõigest 26-aastane!
- 61% Minu nägemuse kohaselt on tänane saal kummitempel.
- 61% Need riigikogulased, kes tõesti on huvitatud rahva ja riigi heaolust, tegeleks siis selle raske ja tõsise tööga.
- 61% Nii enda kui ka teiste puhul.
- 61% Taganemist enam ei tule.
- 61% Üks väliseestlasest amerikaniseerunud daam hüüatas iga minuti tagant: „Otepää on super! Pärnu on võrratu! Haapsalu – sõnu ei jätku! Tartu on maailma parim!“

- 62% Kui mees töötab restoranis, siis mõeldakse, et ju tal on mingi viga küljes.
- 62% Kõik oli nii isikupäratu, et ma ei saanudki aru, kas ta oli üldse kuskil käinud.
- 63% Absurd!
- 63% Ja ma ei tee niisama suuri sõnu!
- 63% Mis ma teile küll teinud olen?!
- 64% Arvan, et keegi ei saa ega suuda joomarlust käskude ja keeldudega pidurada.
- 64% Ei ole kohti!
- 64% Kui oled suitsetaja – lõpetame ära.
- 64% Kuidas nii hakkama saada?!
- 64% Suudate seda ette kujutada?!
- 64% Taganemist enam ei tule.
- 65% Kas me tõesti enam ilma koolituseeta mitte millegagi hakkama ei saa?
- 66% Aga kõik on kuradi magus.
- 66% Elva linn küll oma kodanikele seda kingitust ei tee.
- 66% Ja uimased nagu haid.
- 66% Kas me oleme nendest targemad?
- 66% Mõelgu-mõelgu!
- 66% Suudate seda ette kujutada?!
- 69% Raske on mõista selliseid lapsevanemaid.
- 70% Kas tõesti on inimesi, kes üldse ei hooli?
- 70% Kuid kui mina juba kuhugi sisse saan, siis mind sealt enam niisama välja ei aja.
- 70% Mitu lumesahka ma selle aja jooksul nägin?
- 71% Kas ma pean kedagi aitama, kui tean, et selle inimese abistamine toob kaasa kellegi teise haiguse?
- 71% Vabadus ei ole iseenesestmõistetav.
- 73% Mõelgu, mõelgu!
- 74% Minu nägemuse kohaselt on tänane saal kummitempel.
- 74% Ootaks valitsuse otsust, aga valitsust ju ei ole!
- 75% Miks ei koolitata meile noori abiks?
- 75% Nagu surnukeha.
- 75% Seal on ka tavaks korruga kõik sillad kinni panna, ja inimestel kästakse siis lihtsalt kodus püsida.
- 76% Ma ei taha öelda, et siin nüüd ei ole suurepäraseid kõnelejaid.
- 76% Ma ei tea kedagi, kelle palk tõuseks.
- 77% Kõigest kolm tuhat krooni tund!
- 77% Ma ei huvita tootjaid.
- 77% Peale meid tulgu või veeuputus!
- 78% Kui sõjas on paha, miks sinna minna?
- 80% Ja miks pidi president taas sellist kahtlast värvi kikäpsu kandma?
- 80% Muusikuil ja kunstnikel lausa meeldib karvus olles kaaskodanikega suhelda.

- 80% Vaid riigiametnikele lubatakse kümneprotsendilist palgatõusu.
- 80% Õpetaja on nagu ori vanas Egiptuses, õpetajal on ainult kohustused.
- 81% Aga mis nüüd meie jutt...!
- 83% Poeg on nüüd pähe võtnud, et tema läheb Afganistani.
- 84% Kuidas nii hakkama saada?!
- 86% Mitte ükski härrasmees mitte üheski riigis ei käi daami kõrval, käed mantli taskus!
- 87% Peaminister peab tegema oma suu lahti ja midagi ütleva.
- 87% Peaminister peab tegema oma suu lahti ja midagi ütleva.
- 90% Miks neid ei takistata?
- 90% Saan aru, et kikilips on meie presidendi stiil.
- 92% Mõni meditsiinitöötaja võiks ju metsas raja ääres ikka olla.
- 93% See ei ole normaalne!
- 96% Ometi leidis linnavalitsus, et just Tokman on sobiv inimene eestikeelset kooli juhtima.
- 100% Aga täiesti maitsetud ja lõhnatud.
- 100% Kas talle saab olla uudiseks, et teda taga otsitakse?!
- 100% Võin vanduda, ilma midagi tegemata ollakse sügisel sama küna ees.

### *Neutral / Neutraalne*

- 51% Aga neile on ju elu ja surma küsimus taas parlamenti pääseda!
- 51% Esmaspäeval oli halvim päev Tartu liikluses.
- 52% Ka bussidega on raske.
- 52% Tema esimesed sõnad mulle olid: „Häid pühi!“
- 52% Üldiselt andsin kohe vastu.
- 53% Ei teagi, kuidas saaks inimesi lapse sünniks ette valmistada...
- 53% Habemeajamisega ei näi tegelevat vähemalt kolmandik kes- ja vanema- ealistest.
- 53% Kadri on minu jaoks midagi enam kui sõbranna või koostööpartner.
- 53% Talle võivad ühtäkki väga ootamatud asjad huvi pakkuda.
- 54% Julgust olla ilus ja sentimentaalne.
- 54% Maa ei ole enam ammu külmunud, aga nüüd on meie peaministril kärss kärnas.
- 54% Meie suusaliidus ei muretse, kas kergejõustiku meistrivõistlused toimuvad.
- 54% Mõelgu, et ta on olümpiavõitja.
- 54% Seal on minu koht, kus on minu juured ja traditsioonid.
- 54% Sest Eurovisiooni lauluvõistlus on ikka naljasaade.
- 54% Tundub, nagu juhiksid õpilased kooli!
- 54% Tõesõna, me ei otsi konflikte, vaid tahame lahendust.
- 54% Ühiskond, kus inimene ei häbene oma nime ega sõna ega tegu.
- 55% Aga mitte piisavalt!
- 55% Eesti liha kvaliteet on halb.
- 55% Ja palju.
- 55% Kuidas talitaksite?!

- 55% Kuri ema lööb välja perioodidel, mil mind pole tihti kodus ja kogu see kamp on seal tegevuseta.
- 55% Need minu 30 pudelit ehk 400 krooni kuus on nende jaoks marginaalsed.
- 55% Sama nime ja hinnaga toode on ühel päeval ühe, teisel päeval teise kvaliteediga.
- 55% Seega – Eesti mehe kõige suurem probleem on harimatus.
- 55% Selle maailma vägevad pole Tartu rahule pööranud kõige vähematki tähelepanu.
- 55% Sest minusuguseid on veel.
- 55% Teate küll seda koerte, kollaste kilpide ja kirsasaabaste värki – Tartu rahu 1988.
- 55% Üritatakse lüüa massiga.
- 56% Minu esimesed muljed on, kuidas seda öeldagi – segased.
- 56% Organiseerige noortele parem vaba aja veetmise võimalusi.
- 56% Sõna vastutus on selle juures täiesti naeruväärse kõlaga.
- 57% Jopeta ja külmast kange.
- 57% Lambaliha kvaliteet on halb.
- 57% Olen mõelnud, et mina ei viitsiks kõike uuesti alata.
- 57% Oma mehega mängime teisi rolle...
- 57% See on see, mida ühiskond tahab näha.
- 57% Võtke näiteks enesetapja!
- 58% Aga ehk on tegu hoopis isikliku solvumisega?
- 58% Esmalt vaatasin ta palatikaaslast ja sel ajal, täiesti ootamatult, virutas rohtu nõudnu mulle jalaga ribidesse.
- 58% Tundub, et aeg kiirustas tagant ja kohvik tuli avada enne, kui asjad lõplikult paigas.
- 60% Ja ma usun, et linnateatrilgi on teda väga vaja.
- 60% Kas me teame, mis on meile parem?
- 60% Kool peaks olema kultuurikeskus.
- 60% Kui Irina Tokman kahe aasta eest kooli direktoriks määrati, polnud tal ei eesti keele kõrgtaseme tunnistust ega juhtimiskogemust.
- 60% Kui me neile „puudustele” otsa vaatame ja neid tundma õpime, võime neist kasugi saada.
- 60% Kui pensionile jään, siis kolime mehega metsa elama.
- 60% Lapsed on vaesusrisk number üks – kui koledalt ja küüniliselt see ka ei kõla.
- 60% Ma tahan olla kaitstud riigi, ametnike ja ajakirjanike omavoli kui ka kuritarvituste eest.
- 60% Mees, sa ei tea, kui kiiresti ma saen!
- 60% Meie olime sullejopedes ja puhvaikades, aga tema oli elegantses tviidpintsakus, prantslaslikult sõlmitud sall ümber kaela...
- 60% Mulle on Tartu rahu õige mitmel põhjusel vastik.
- 60% Nüüd on tuhanded inimesed ja apteekrid hädas.
- 60% See muusika on selge ja aus ning kõlbab ka hiljem tarvitada.

- 60% Sinikad olid tol ajal elu lahutamatu osa.
- 60% Sinul on.
- 60% Sõda ei ole hea.
- 60% Tardunud õhku.
- 60% Tänu Eesti Laulule on nüüd kuidagi ausam ja kuivem tunne.
- 60% Vigu on süsteemis nii palju, et sellisel kujul seda käiku anda ei tohi.
- 63% Keskerakonnal polegi vaja mingeid uusi nippe välja mõelda, piisab nendest, mis neile ette mängitakse.
- 63% Kohta, kus puudub vastutus ja sisuliselt ka puudub vabadus.
- 63% Ma arvan, et avalikkus peaks politseilt iga viie tunni tagant küsima, millised on edusammud.
- 63% Ma usun, et selline graatsilisus on alati nõudnud omajagu häbematumust.
- 63% Mina tahan, et andmekaitse kaitseks mind, minu puutumatumust spämmi ja muu inforünde eest.
- 63% Paraku valitseb liigagi paljudes kodudes pidevalt kalk õhkkond.
- 63% Samas tundub, et need õied räägivad alati väga inimlikest asjadest.
- 64% Bussi sattus tihtilugu sihukesi vendi, kes hõlmast kinni võtsid ja tänavale rebisid.
- 64% Elan koos oma invaliidist tütreaga nüüd Tallinnas ning tunneme end siin palju mugavamalt kui Narvas.
- 64% Mart vedas näo viltu.
- 64% Puudutasin kätt, käed jääkülmad.
- 65% Nägin, kuidas Nõmme linnaosa vanem eisel Rabajooksul kokku kukkus.
- 66% „Ära pööranud,“ ütles sõber.
- 66% Eestis ei ole head lambaliha saada.
- 66% Järjekordsed valimised on mõne poliitiku pea täiesti pööraseid mõtteid täis tuupinud.
- 66% Kes hõõrub teisele rahu nina alla, see mahitab salakaubana sõda.
- 66% Mul oli kolm aastat menüüs Eesti vasikas.
- 66% Mul on tunne, et läheb Eesti armee.
- 66% Neid on päris palju.
- 66% Paljud kliendid on leppinud, et teenustasud kipuvad ikka ja jälle pärmina kerkima.
- 66% Sedasorti inimene, kes suudab ise köögimööbli ehitada.
- 66% Tüüpiline Eesti mees on keskmiselt vähem haritud kui naine ning sageli ebatervislike eluviisidega – riskib palju ja hukkub õnnetustes, joob palju ja jääb haigeks, ei liiguta ennast ja jääb haigeks.
- 66% Varavalges ilmuvad Vova, Boba ja Slava ning viskavad katuse lumest puhtaks ja raiuvad lahti 10 cm paksuse jääkamaka seal all.
- 66% Veel mõni aeg tagasi tõstis Hansapank teatud kliendid „võtmekliendi“ staatusesse ja pakkus soodsamaid tingimusi.
- 67% Või ei ole Nool elus oma kohta leidnud?
- 68% Näiteks, täna hommikul kuulsin, et üks noorpoliitik, B-klassi särav täht, on kõrvaldanud suure summa raha.
- 69% Põhjendamatult kõrged ootused võivad tõepoolest kriisini viia.

- 69% Tundub, et aeg kiirustas tagant ja kohvik tuli avada enne, kui asjad lõplikult paigas.
- 70% Ainus küsitav element Eesti Laulu puhul on peaauphind – sõit Eurovisiooni lõppkontserdile.
- 70% Eilsel kontserdil suitsetas Padar nagu korsten.
- 70% Mis siis soome-ugri maailmakongressil TEGELIKULT juhtus?
- 71% Eriti kesklinnas.
- 71% See on sisserändajate maa: mustad ja kollased, venelased ja hiinlased.
- 74% Üks teenindaja ütles ajalehes, et tal on kogu aeg halb tuju, sest kõrtsis käib palju soomlasi.
- 76% Aga seda karmim on tõde hiljem!
- 76% Ei teagi, kuidas saaks inimesi lapse sünniks ette valmistada...
- 76% Varem olin ma sunnitud last ravile viies ta ratastoolist otse rāpasele perroonile istuma panema: kõigepealt tõstsin välja ratastooli ning seejärel lapse.
- 77% Eks see lõugamine olegi rohkem iseenda abitus.
- 77% Korraga tekivad lühised, plahvatavad lambipirnid, seinad kobrutavad valguvast veest.
- 77% Ma ei tea kedagi, kes teaks kedagi, kes tahaks Venemaaga sõdida.
- 77% Ostetakse välimuse, mitte maitse järgi.
- 77% Sõnaga – mulle meeldib juua mittealkohoolseid karastusjooke.
- 78% Krissu on olnud kogu aeg kange iseloomuga tüdruk.
- 80% Jõutakse selleni, et kommenteerijate tuvastamisnõue riivab sõnavabadust.
- 80% Kodanikuühiskond on tugevate isiksuste, ausate ja avameelsete inimeste ühiskond.
- 80% Leo Kunnas võtab väsimatult sõna teemal, kuidas Eesti kaitsevāes on asjad hapud.
- 80% Lõin talle lahtise käega vastu õlga.
- 80% Meile kummalegi ei meeldi nutt ega virin.
- 80% Mul on kodus klaveri kohal üks akvarell – kärbse ning suure kollase õisikuga.
- 80% Võib olla üsna kindel, et lõviosa finaali jõudnud lugudest oleksid sündinud niikuinii.
- 83% Arvan, et õpetajaid mõnitavate õpilaste hulk kasvab iga aastaga.
- 85% Kõikidesse suhtutakse ühtemoodi.
- 88% Sõber tundus murelik.
- 93% Mul tuli Siberis veeta kokku 17 aastat!
- 100% Näiteks kadedus võib sundida ennast ületama või arendama.
- 100% Ühe miljardi krooni juurde laenamine tähendab iga tallinlase kohta 2500 krooni ehk ühte tavalist piiks-piiks laenu.





## **PUBLICATIONS / PUBLIKATSIOONID**

## CURRICULUM VITAE

**Name:** Kairi Tamuri  
**Date of birth:** January 7, 1978  
**Citizenship:** Estonian  
**Address:** Institute of the Estonian Language, Roosikrantsi 6, Tallinn  
10119, Estonia  
**Telephone:** (+372) 506 5572  
**E-mail:** kairi.tamuri@eki.ee  
**Current position:** Institute of the Estonian Language, Analyst –  
Junior Researcher

### Education

- University of Tartu, since 09.2008 PhD Student (Estonian and Finno-Ugric Linguistics)
- Tallinn University, 2007, Master's Degree (Linguistics)
- Tallinn University, 2005, Bachelor's Degree (Estonian Philology)

### Professional career

Since 2016 Analyst – Junior Researcher, Institute of the Estonian Language  
2006–2016 Analyst, Institute of the Estonian Language

### Field of research

Speech acoustics and perception

### Projects

#### *Projects in progress*

- Speech styles, sentence prosody, phonological variation: description, theory and modelling (project no. IUT35-1), 01.01.2015 – 31.12.2020
- Center of Estonian Language Resource (project no. IUT49-1), 01.01.2016 – 31.12.2021

#### *Completed projects*

- CLARIN PLUS (project no. MLTAT16284R), 01.05.2016 – 01.11.2016
- Statistical models for emotional speech and text (project no. EKT1), 01.01.2011 – 31.12.2014
- Modelling intermodular phenomena in Estonian (project no. SF0050023s09), 01.01.2009 – 31.12.2014
- The bidirectional relationships between the speech rhythm and Estonian grammatical and lexical structures (project no. ETF7998), 01.01.2009 – 31.12.2012
- Estonian Emotional Speech Corpus (project no. EKKTT06-1), 01.01.2006 – 31.12.2010

### **A list of publications**

- Tamuri, Kairi & Mihkla, Meelis. 2015. Expression of basic emotions in Estonian parametric text-to-speech synthesis. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics*, 6(3), erinumber / special issue: *Kõneuurimise suundi / Aspects of speech studies*, 145–168.
- Tamuri, Kairi. 2015. Fundamental frequency in Estonian emotional read-out speech. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics*, 6 (1), erinumber / special issue: *Papers from the conference “Finnic Languages, Cultures, and Genius Loci”*, 9–21.
- Pajupuu, Hille, Pajupuu, Jaan, Tamuri, Kairi, & Altrov, Rene. 2015. Influence of verbal content on acoustics of speech emotions. *Proceedings of the 18th International Congress of Phonetic Sciences*. Editors (abbr.) The Scottish Consortium for ICPHS 2015. Glasgow, UK: The University of Glasgow, 1–5.
- Tamuri, Kairi. 2012. Kas formandid peegeldavad emotsioone? *Eesti Rakenduslingvistika Ühingu aastaraamat*, 8, 231–243.
- Tamuri, Kairi & Mihkla, Meelis. 2012. Emotions and speech temporal structure. *Linguistica Uralica*, 3, 209–217.
- Tamuri, Kairi. 2012. Intensity of Estonian emotional speech. *Human Language Technologies – The Baltic Perspective – Proceedings of the Fifth International Conference Baltic HLT 2012*, IOS Press, 238–246.
- Mihkla, Meelis, Hein, Indrek, Kalvik, Mari-Liis, Kiissel, Indrek, Sirts, Risto, & Tamuri, Kairi. 2012. Estonian speech synthesis: applications and challenges / Синтез речи эстонского языка: применение и вызовы. *Computational Linguistics and Intellectual Technologies, Papers from the Annual International Conference “Dialogue”*. Editor (abbr.) A. E. Kibrik, Moskva: РГГУ, 443–453.
- Tamuri, Kairi. 2010. Kas pausid kannavad emotsiooni? *Eesti Rakenduslingvistika Ühingu aastaraamat*, 6, 297–306.
- Kerge, Krista, Pajupuu, Hille, Tamuri, Kairi, & Meier, Heidi. 2008. Kõnetehnoloogia vajab žanrilist lähenemist [Speech technology needs a genre-based approach]. *Eesti Rakenduslingvistika Ühingu aastaraamat*, 4, 53–65.
- Kerge, Krista, Pajupuu, Hille, & Tamuri, Kairi. 2008. Where should TTS-synthesizer pause and breath? In: *The Third Baltic Conference on Human Language Technologies*, Vilnius: Vytauto Didžiojo Universitetas, Lietuvių kalbos institutas, 143–149.

### **Other professional activities**

Since 2009 Member of the Estonian Mother Tongue Society

Since 2008 Member of the Estonian Association for Applied Linguistics

## ELULOOKIRJELDUS

**Nimi:** Kairi Tamuri  
**Sünniaeg:** 7. jaanuar 1978  
**Kodakondsus:** Eesti  
**Aadress:** Eesti Keele Instituut, Roosikrantsi 6, Tallinn 10119, Eesti  
**Telefon:** (+372) 506 5572  
**E-post:** kairi.tamuri@eki.ee  
**Töökoht:** Eesti Keele Instituut, analüütik-nooremteadur

### Haridus

- Tartu Ülikool, alates september 2008 doktorant (eesti ja soome-ugri keeleteadus)
- Tallinna Ülikool, 2007, magistrikraad (lingvistika)
- Tallinna Ülikool, 2005, bakalaureusekraad (eesti filoloogia)

### Teenistuskäik

2016 analüütik-nooremteadur, Eesti Keele Instituut  
2006–2016 analüütik, Eesti Keele Instituut

### Peamised uurimisvaldkonnad

Kõneakustika ja -taju

### Projektid

#### *Jooksvad projektid*

- Kõnestiilid, lauseprosoodia ja fonoloogiline varieerumine: kirjeldus, teooria ja modelleerimine (projekt nr IUT35-1), 01.01.2015 – 31.12.2020
- Eesti Keeleressursside Keskus (projekt nr IUT49-1), 01.01.2016 – 31.12.2021

#### *Lõppenud projektid*

- CLARIN PLUS (projekt nr MLTAT16284R), 01.05.2016 – 01.11.2016
- Kõne ja teksti emotsionaalsuse statistilised mudelid (projekt nr EKT1), 01.01.2011 – 31.12.2014
- Eesti keele alusuuringud keeletehnoloogiliste rakenduste teenistuses (projekt nr SF0050023s09), 01.01.2009 – 31.12.2014
- Eestikeelse kõne rütmilisuse peegeldused grammatilistes ja leksikaalsetes struktuurides (ja *vice versa*) (projekt nr ETF7998), 01.01.2009 – 31.12.2012
- Eesti emotsionaalse kõne korpus (projekt nr EKKTT06-1), 01.01.2006 – 31.12.2010

### **Publikatsioonide loetelu**

- Tamuri, Kairi & Mihkla, Meelis. 2015. Expression of basic emotions in Estonian parametric text-to-speech synthesis. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics*, 6(3), erinumber / special issue: *Kõneuurimise suundi / Aspects of speech studies*, 145–168.
- Tamuri, Kairi. 2015. Fundamental frequency in Estonian emotional read-out speech. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics*, 6 (1), erinumber / special issue: *Papers from the conference "Finnic Languages, Cultures, and Genius Loci"*, 9–21.
- Pajupuu, Hille, Pajupuu, Jaan, Tamuri, Kairi, & Altrov, Rene. 2015. Influence of verbal content on acoustics of speech emotions. *Proceedings of the 18th International Congress of Phonetic Sciences*. Editors (abbr.) The Scottish Consortium for ICPHS 2015. Glasgow, UK: The University of Glasgow, 1–5.
- Tamuri, Kairi. 2012. Kas formandid peegeldavad emotsioone? *Eesti Rakenduslingvistika Ühingu aastaraamat*, 8, 231–243.
- Tamuri, Kairi & Mihkla, Meelis. 2012. Emotions and speech temporal structure. *Linguistica Uralica*, 3, 209–217.
- Tamuri, Kairi. 2012. Intensity of Estonian emotional speech. *Human Language Technologies – The Baltic Perspective – Proceedings of the Fifth International Conference Baltic HLT 2012*, IOS Press, 238–246.
- Mihkla, Meelis, Hein, Indrek, Kalvik, Mari-Liis, Kiissel, Indrek, Sirts, Risto, & Tamuri, Kairi. 2012. Estonian speech synthesis: applications and challenges / Синтез речи эстонского языка: применение и вызовы. *Computational Linguistics and Intellectual Technologies, Papers from the Annual International Conference "Dialogue"*. Editor (abbr.) A. E. Kibrik, Moskva: РГГУ, 443–453.
- Tamuri, Kairi. 2010. Kas pausid kannavad emotsiooni? *Eesti Rakenduslingvistika Ühingu aastaraamat*, 6, 297–306.
- Kerge, Krista, Pajupuu, Hille, Tamuri, Kairi, & Meier, Heidi. 2008. Kõnetehnoloogia vajab žanrilist lähenemist [Speech technology needs a genre-based approach]. *Eesti Rakenduslingvistika Ühingu aastaraamat*, 4, 53–65.
- Kerge, Krista, Pajupuu, Hille, & Tamuri, Kairi. 2008. Where should TTS-synthesizer pause and breath? In: *The Third Baltic Conference on Human Language Technologies*, Vilnius: Vytauto Didžiojo Universitetas, Lietuvių kalbos institutas, 143–149.

### **Teadusorganisatsiooniline tegevus**

2009 alates Emakeele Seltsi liige

2008 alates Eesti Rakenduslingvistika Ühingu liige

## DISSERTATIONES PHILOLOGIAE ESTONICAE UNIVERSITATIS TARTUENSIS

1. **Ülle Viks.** Eesti keele klassifikatoorne morfoloogia. Tartu, 1994.
2. **Helmi Neetar.** Deverbaalne nominaaltuletus eesti murretes. Tartu, 1994.
3. **Ülo Valk.** Eesti rahvausu kuradi-kujutelm. Tartu, 1994.
4. **Arvo Eek.** Studies on quantity and stress in Estonian. Tartu, 1994.
5. **Reet Kasik.** Verbid ja verbaalsubstantiivid tänapäeva eesti keeles. Tartu, 1994.
6. **Silvi Vare.** Nimi- ja omadussõnatuletus tänapäeva eesti kirjakeeles. Tartu, 1994.
7. **Heiki-Jaan Kaalep.** Eesti keele ressursside loomine ja kasutamine keele- tehnoloogilises arendustöös. Tartu, 1998.
8. **Renate Pajusalu.** Deiktikud eesti keeles. Tartu, 1999.
9. **Vilja Oja.** Linguistic studies of Estonian colour terminology. Tartu, 2001.
10. **Külli Habicht.** Eesti vanema kirjakeele leksikaalsest ja morfosüntaktilisest arengust ning Heinrich Stahli keele eripärast selle taustal. Tartu, 2001.
11. **Pire Teras.** Lõunaeesti vokaalisüsteem: Võru pikkade vokaalide kvaliteedi muutumine. Tartu, 2003.
12. **Merike Parve.** Väited lõunaeesti murretes. Tartu, 2003.
13. **Toomas Help.** Sõnakeskne keelemudel: Eesti regulaarne ja irregulaarne verb. Tartu, 2004.
14. **Heli Laanekask.** Eesti kirjakeele kujunemine ja kujundamine 16.–19. sajandil. Tartu, 2004.
15. **Peeter Päll.** Võõrnimed eestikeelses tekstis. Tartu, 2005.
16. **Liina Lindström.** Finiitverbi asend lauses. Sõnajärg ja seda mõjutavad tegurid suulises eesti keeles. Tartu, 2005.
17. **Kadri Muischnek.** Verbi ja noomeni püsiühendid eesti keeles. Tartu, 2006.
18. **Kanni Labi.** Eesti regilaulude verbisemantika. Tartu, 2006.
19. **Raili Pool.** Eesti keele teise keelena omandamise seaduspärasusi täis- ja osasihitise näitel. Tartu, 2007.
20. **Sulev Iva.** Võru kirjakeele sõnamuutmissüsteem. Tartu, 2007.
21. **Arvi Tavast.** The translator is human too: a case for instrumentalism in multilingual specialised communication. Tartu, 2008.
22. **Evar Saar.** Võrumaa kohanimed analüüs enamlevinud nimeosade põhjal ja traditsioonilise kogukonna nimesüsteem. Tartu, 2008.
23. **Pille Penjam.** Eesti kirjakeele *da-* ja *ma-*infinitiiviga konstruktsioonid. Tartu, 2008.
24. **Kristiina Praakli.** Esimese põlvkonna Soome eestlaste kakskeelne keelekasutus ja koodikopeerimine. Tartu, 2009.
25. **Mari Mets.** Suhtlusvõrgustikud reaajas: võru kõnekeele varieerumine kahes Võrumaa külas. Tartu, 2010.

26. **Karen Kuldnook.** Militaarne retoorika. Argumentatsioon ja keeleline mõjutamine Eesti kaitsepoliitilises diskursuses. Tartu, 2011.
27. **Kai Tafenau.** Uue Testamendi tõlkimisest Rootsi ajal: käsikirjad, tõlkijad ja eesti kirjakeel. Tartu, 2011.
28. **Külli Prillop.** Optimaalsusteoreetiline käsitlus eesti keele fonoloogilisest kujunemisest. Tartu, 2011, 261 lk.
29. **Pärtel Lippus.** The acoustic features and perception of the Estonian quantity system, Tartu, 2011, 146 p.
30. **Lya Meister.** Eesti vokaali- ja kestuskategooriad vene emakeelega keelejuhtide tajus ja häälduses. Eksperimentaalfoneetiline uurimus. Tartu, 2011, 145 lk.
31. **Kersti Lepajõe.** Kirjand kui tekstiliik. Riigieksamikirjandite tekstuaalsed, retoorilised ja diskursiivsed omadused. Tartu, 2011, 141 lk.
32. **Tiit Hennoste.** Grammatiliste vormide seoseid suhtlustegevustega eesti-keelses suulises vestluses. Tartu, 2013, 236 lk.
33. **Helena Metslang.** Grammatical relations in Estonian: subject, object and beyond. Tartu, 2013, 443 p.
34. **Kristel Uihoaed.** Verbiühendid eesti murretes. Tartu, 2013, 227 lk.
35. **Helen Plado.** Kausaalsuhete adverbiallaused eesti keeles. Tartu, 2013, 244 lk.
36. **Annika Küngas.** Pragmatiliste markerite kujunemine ja funktsioonid eesti keeles *It*-sõnade näitel. Tartu, 2014, 200 lk.
37. **Maarika Teral.** Arvutipõhine eesti keele õpe: vahendid ja hinnangud nende efektiivsusele Tartu ülikooli keelekursuste näitel. Tartu, 2015, 175 lk.
38. **Anni Jürine.** The development of complex postpositions in Estonian: a case of grammaticalization via lexicalization. Tartu, 2016, 315 p.