

RENE ALTROV

The Creation of the Estonian Emotional
Speech Corpus and the Perception
of Emotions



RENE ALTROV

The Creation of the Estonian Emotional
Speech Corpus and the Perception
of Emotions

University of Tartu, Institute of Estonian and General Linguistics

Dissertation accepted for the commencement of the degree of Doctor of Philosophy on March 4, 2014 by the Committee of the Institute of Estonian and General Linguistics, Faculty of Philosophy, University of Tartu

Supervisors: Professor Urmas Sutrop, Doctor of Philosophy,
University of Tartu, and the Institute of the Estonian Language,
Tallinn

Hille Pajupuu, Doctor of Philosophy,
Institute of the Estonian Language, Tallinn

Opponent: Professor Jens Allwood, Doctor of Philosophy,
University of Gothenburg

Commencement: June 2, 2014 at 14.15



European Union
European Social Fund



Investing in your future

ISSN 1406–5657
ISBN 978–9949–32–554–2 (print)
ISBN 978–9949–32–555–9 (pdf)

Copyright: Rene Altrov, 2014

University of Tartu Press
www.tyk.ee

ACKNOWLEDGEMENTS

I am indebted to a number of people whose support, dedication and encouragement have made it possible for me to undertake this thesis.

Sincerest thanks to my supervisors Dr Hille Pajupuu and Professor Urmas Sutrop for teaching, encouraging and supervising me.

I would also like to thank Jaan Pajupuu for the technical support and for all those heart-warming cups of tea.

I am very grateful to editors Maria-Maren Sepper and Penny Mountain, who helped to bring clarity and flow into my writing.

I wish to thank Professor Jens Allwood and Professor Krista Kerge for their pre-reviews, advice and useful suggestions.

Thanks, too, to Tiia Margus, coordinator of the Institute of Estonian and General Linguistics of the University of Tartu, for her help in several matters.

I am extremely grateful to my colleagues in the Institute of the Estonian Language for their enthusiasm and support.

Thanks to all 190 testers of the Estonian Emotional Speech Corpus.

The thesis was completed with support from the National Programme for Estonian Language Technology projects – “Estonian Emotional Speech Corpus”, “Statistical Models of the Emotionality of Speech and Written Text” and the target funded research project “Modelling intermodular phenomena in Estonian”.

CONTENTS

LIST OF PUBLICATIONS.....	7
1. INTRODUCTION.....	8
1.1. Objectives.....	8
1.2. Structure OF THE THESIS.....	9
1.3. Overview of Publications and the Author's Contribution.....	10
1.4. Terminological Issues.....	11
2. ESTONIAN EMOTIONAL SPEECH CORPUS: THEORETICAL BASE AND IMPLEMENTATION	13
3. PERCEPTION OF EMOTION	19
3.1. Age Influence on the Identification of Emotion.....	20
3.2. Cultural Influence on the Identification of Emotion	21
3.3. The role of Empathy in the Identification of Emotion	22
4. RESULTS AND DISCUSSION	25
4.1. The Estonian Emotional Speech Corpus among other Corpora of Emotional Speech.....	25
4.2. The Influence of Age, Culture, and Empathy on the Identification of Emotion	27
5. CONCLUSION	30
6. SUMMARY IN ESTONIAN. Eesti emotsionaalse kõne korpuse loomine ja emotsioonide taju	32
6.1. Väitekirja eesmärk ja lähted.....	32
6.2. Eesti emotsionaalse kõne korpus: teoreetiline alus ja teostus	34
6.3. Emotsioonide taju.....	39
6.3.1. Vanuse mõju emotsioonide identifitseerimisele.....	40
6.3.2. Kultuuri mõju emotsioonide identifitseerimisele	41
6.3.3. Empaatia mõju emotsioonide identifitseerimisele.....	42
6.4. Tulemused ja diskussioon	44
6.4.1. Eesti emotsionaalse kõne korpus teiste kõneemotsioonikorpuste seas	44
6.4.2. Vanuse, kultuuri ja empaatia mõju emotsioonide identifitseerimisele	47
REFERENCES	50
PUBLICATIONS	59
CURRICULUM VITAE	141
ELULOOKIRJELDUS.....	143

LIST OF PUBLICATIONS

- [P1] Altrov, Rene. 2008. Eesti emotsionaalse kõne korpus: teoreetilised toetuspunktid. *Keel ja Kirjandus*, 4, 261–272.
- [P2] Altrov, Rene, & Pajupuu, Hille. 2008. The Estonian Emotional Speech Corpus: Release 1. In F. Čermak, R. Marcinkevičienė, E. Rimkutė, & J. Zabarskaitė (Eds.), *Proceedings of the Third Baltic Conference on Human Language Technologies* (pp. 9–15). Vilnius: Vytauto Didžiojo Universitetas, Lietuvių Kalbos Institutas.
- [P3] Altrov, Rene, & Pajupuu, Hille. 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.
- [P4] Altrov, Rene, & Pajupuu, Hille. 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.
- [P5] Altrov, Rene, & Pajupuu, Hille. 2010. Estonian Emotional Speech Corpus: Culture and Age in Selecting Corpus Testers. In I. Skadina, & A. Vasiljevs (Eds.), *Human Language Technologies – The Baltic Perspective – Proceedings of the Fourth International Conference Baltic HLT 2010* (pp. 25–32). Amsterdam: IOS Press.
- [P6] Altrov, Rene. 2013. Aspects of cultural communication in recognizing emotions. *Trames: Journal of the Humanities and Social Sciences*, 17(67/62), 159–174.
- [P7] Altrov, Rene, Pajupuu, Hille, & Pajupuu, Jaan. 2013. The role of empathy in the recognition of vocal emotions, In *INTERSPEECH-2013*, 1341–1344.

I. INTRODUCTION

I.1. Objectives

In the last decade affective computing and research of speech emotions have shifted from the exotic to the mainstream. This has created a need for emotional speech corpora. The development of an Estonian corpus of emotional speech started in 2006 under the National Programme for Estonian Language Technology (2006–2010). The theoretical base of this corpus was created as a result of the present study.

The Estonian Emotional Speech Corpus (henceforth the Corpus) was designed to fulfil the following two objectives:

- 1) To serve as a reliable database for studying emotions manifested in spoken and written texts;
- 2) To be applicable in speech technological tasks. Initially, the Corpus material was to be applied in unit selection based text-to-speech systems to synthesize emotional speech¹(See publications [P1], [P2], [P3], [P4]).

By the time the Estonian Emotional Speech Corpus was initiated, some corpora of emotional speech already existed for other languages. Most of these were datasets rather than databases. They were comparatively small-scale collections of material, typically created to examine a single issue and thus not widely available while in some cases development had ceased altogether (Douglas-Cowie, Campbell, Cowie & Roach, 2003). Moreover, most of the available corpora of emotional speech were, and still are, based on acted material; that is, on sentences or passages read by professional actors who have been instructed to simulate different emotions. The use of actors to collect emotional expressions has been continually challenged, because acted emotional expressions are different from authentic and genuine emotions (see Douglas-Cowie, Campbell, Cowie & Roach, 2003; Laukka, Neiberg, Forsell, Karlsson, Elenius 2010; Scherer, 2013), and because of overacting (Wilting, Kraemer & Swerts, 2006).

The aim of the thesis is to create a theoretical base for the Estonian Emotional Speech Corpus and to verify the validity of the theoretical starting points. With that aim in mind, I scrutinized such limited research and published material on the available corpora as existed at the time, and by examining the corpus descriptions and the positive and negative lessons accumulated by their authors, I developed a theoretical starting-point according to which the Corpus was subsequently implemented. The basic principles thus developed were tested on corpus material (see [P1], [P2], [P3], [P4], [P5], [P6], [P7]).

¹ In the early stages of corpus creation, unit selection based text-to-speech synthesis was used in Estonia. However, after a short while, statistical parametric speech synthesis came into use. The creators of the Corpus had to adapt to these changes and ensure that the Corpus contained enough material for training statistical models. Emotional speech synthesis is still new in Estonia (see Mihkla et. al., 2012).

Based on the analysis of previous studies I established the following basic principles for the Estonian Emotional Speech Corpus:

- 1) The Corpus should contain authentic emotions, not acted ones.
- 2) The Corpus should contain speech units for the emotion which has been identified (attributed) by listeners (testers).

In reference to the above two decisions, I followed the example of the corpus created for the Japanese emotional speech synthesis system CHATR. Their practice meant abandoning the use of actors and having emotions identified by listeners, which was novel at the time but nevertheless a method that produced acceptable results and appeared valid (see Iida, Campbell, Higuchi & Yasumura, 2003).

- 3) The Corpus should be able to be extended as required for research and language technological purposes.
- 4) The Corpus should be publicly available in all development stages.

When the Corpus was started, there was no certainty of how well moderately expressed non-acted emotions could be identified from the voice (cf. Cowie & Cornelius, 2003). As the emotions of the speech units (sentences) selected for the Estonian corpus were to be established by listeners (testers), a lot of attention had to be paid to the criteria of listener selection. I used perception (listening) tests to check the relevance of the criteria. As one of the aims of the thesis bore upon the perception of emotion, I formulated the following research questions:

- 1) Can the moderately expressed emotion of a sentence be identified from the voice without the speaker being seen?
- 2) Does the identification of emotions depend on the listener's age?
- 3) Is the identification of emotion culturally bound?
- 4) What is the possible role of empathy in the identification of emotion?

The results of the study reported in the present thesis serves as the theoretical base of the now existing Estonian Emotional Speech Corpus. Several theoretical starting-points of the Corpus have been tested in the course of this study. The Corpus is publicly available at <http://peeter.eki.ee:5000/>.

I.2. Structure of the Thesis

The present thesis consists of an introductory part and seven publications (see [P1], [P2], [P3], [P4], [P5], [P6], [P7]). The introductory part is divided into seven chapters. Chapter one provides an introduction to the thesis, describes the structure of the thesis, and offers a short overview of related publications and the author's contribution to co-authored papers. Chapter two provides a survey of the theoretical starting-points of the Corpus, describes the design and structure of the Corpus, and explains the major decisions made prior to starting the Corpus. Chapter three focuses on the perception of emotions, surveying the extent to which emotion identification can be affected by a person's age, culture, and empathic ability. Chapter four summarizes the results of the seven

publications emerging from the Corpus. Chapter five concludes the thesis and suggests some further research objectives. Chapter six is an overview of the thesis in Estonian and chapter seven lists the references cited throughout the thesis.

I.3. Overview of Publications and the Author's Contribution

The seven publications can be divided into two major groups: the first four articles describe the design and development of the Estonian Emotional Speech Corpus while the other three deal with the perception of emotion. The co-authors of the papers are aware of the following descriptions and agree with them.

[P1] summarizes the theoretical starting-points considered at the development of the Corpus. This is followed by a description of the stages of Corpus development, including the selection of emotions, reading material, and speakers/readers as well as the perception tests carried out for emotion identification.

[P2] is an extended version of [P1]. It presents a survey of the stages of Corpus development and of the results of tests measuring perception. Detailed attention is given to the method used to differentiate between the sentences in which emotion is carried by the voice only and those in which the identification of the emotion may have been influenced by the verbal-semantic content of the sentence. The first author deals with the theoretical base of the Corpus, its development, and the procedure of the perception tests.

[P3] is a follow-up of [P1] and [P2]. Here, the focus is on the validity of the choices made in the course of the elaboration of the Corpus's theoretical base, considering how appropriate it was to compile the Corpus material from non-acted speech and whether it was appropriate to classify the sentences into two categories: those with emotion carried just by the voice, and those where verbal-semantic content may, in some way, have affected the identification of emotion. The first author deals with the Corpus structure, material recording, the conduct of perception tests, and the analysis of the results of identifying emotion.

[P4] introduces the theoretical base and structure of the Corpus and reports some research results. The first author focuses on the implementation of the theory and describes the material used.

[P5] gives an overview of the Corpus and the principles used in selecting its testers. The paper investigates how appropriate it was to choose the testers from among the ethnic Estonians aged over thirty. The first author is responsible for the description of the theoretical base of the Corpus, a survey of previous research on age-related effects on emotion recognition, the conduct of the tests, an introduction of the material and method, and writing the discussion component.

[P6] discusses how people who have lived in a specific cultural environment for a long period have learned to understand the emotions of representatives of this culture, when these emotions are not expressed in the formers' first language (L1) but in their second language, and determines if their perception of emotions resembles more L1 or L2.

[P7] addresses the role of empathy in the recognition of vocal emotions. The first author of the article has been involved in surveying previous research on the relationship between empathy and the identification of emotions, in conducting the tests, and in writing the discussion component of the paper.

I.4. Terminological Issues

The articles included in the dissertation have been produced throughout the relatively long period from 2008–2013. As the study of speech emotions is a new and continuously evolving area of research, the relevant terminology has been subject to gradual specification, crystallization and unification in my own articles as well as in the publications of other scholars.

My approach to the concept of emotion relies on the wider sense of the word, which is common in speech technology: 'emotion' is "what is present in most of life but absent when people are emotionless" (Cowie, Sussman & Ben Ze'ev, 2011; Schuller, Batliner, Steidl & Seppi, 2011). In the present study, when joy, anger, or sadness are mentioned, I mean not only full-blown basic emotions but rather 'families' of emotions which cover several similar emotions and emotion-related states (vt Cowie & Cornelius 2003; Iida, Campbell, Higuchi & Yasumura, 2003; Scherer, 1999). Also, the subjects of perception tests, when asked to identify joy, anger, or sadness, were instructed to think of those words as referring to such families and not to anticipate a stereotypical basic emotion. Specifically, joy also covered gratitude, happiness, pleasantness, and exhilaration; anger included resentment, irony, reluctance, contempt, malice, and rage; sadness also referred to loneliness, disconsolation, concern, and hopelessness (cf. Iida, Campbell, Higuchi & Yasumura, 2003). For this study, emotionless speech – speech conveying no particular emotion – is defined as neutral speech. Listeners also decided which sentences in particular qualified as emotionally 'neutral'.

As regards emotion tests, I distinguish between the terms *identify* and *recognition*. I use *identify* when there is no right answer. I use *recognition* when comparing the test results of two groups and the right answer is known. I consider the answers given by Estonian listeners to be right if they score two times better than chance probability.

A wide variety of terms are used in publications to refer to text content, such as *semantic content* (Kienast, Paeschke & Sendlmeier, 1999; Lee & Narayanan, 2005; Scherer, 2010; Dupuis & Pichora-Fuller, 2010; Rodero, 2011), *verbal content* (Kramer, 1964; Scherer, 1986; Hawk, van Kleef, Fischer & van der Schalk, 2009; Tawari & Trivedi, 2010; Scherer, 2013), *linguistic content*

(Devillers, Vidrascu & Lamel 2005; Campbell, 2005; Batliner, Steidl, Hacker & Nöth, 2008; Pell, Paulmann, Dara, Alasseri & Kotz, 2009; Campbell, 2010), *lexical content* (Fragopanagos & Taylor, 2005; Busso & Narayanan, 2009; Barkhusen, Krahmer & Swerts, 2010; Polzehl, Schmitt, Metze & Wagner, 2011; Grichkovtsova, Morel & Lacheret, 2012), and *verbal-semantic content* (Paulmann, Pell & Kotz, 2008). In this thesis I use the term verbal-semantic content to refer to the content of written text.

2. ESTONIAN EMOTIONAL SPEECH CORPUS: THEORETICAL BASE AND IMPLEMENTATION

The Estonian Emotional Speech Corpus is a publicly available collection of the emotions expressed in the Estonian language. The creation of the Corpus began by formulating theoretical starting-points (see [P1]; [P2]; [P3]; [P4]), based on overviews of existing emotion corpora and previous emotion research (e.g. Campbell, 2000; Cowie & Cornelius, 2003; Douglas-Cowie, Campbell, Cowie & Roach, 2003; Scherer, Banse & Wallbott, 2001; Ververidis & Kotropoulos, 2006). The Corpus was initiated in 2006, at which time the number of publicly available corpora based on reliable data was still quite small. The Reading-Leeds, which was the first of the relatively large datasets of emotional speech, was started in 1994 (Greasley et al., 1995). Yet most of the rest were ad hoc and small-scale datasets, which were not meant for public use. The main problems of the available datasets had to do with scale, validity, and generalisability of the data (Douglas-Cowie, Campbell, Cowie & Roach, 2003).

Thus, after considering the available surveys of, and research on, the previous emotion corpora, decisions regarding the scope of the study were based upon the following questions related to data selection:

- 1) Which emotions should the corpus cover?
- 2) Should the corpus contain spontaneous, elicited, or acted emotions?
- 3) Should the texts in the corpus be spoken or read?
- 4) Which texts should be selected and of what length, content, and context?
- 5) Should the texts be presented by professional or trained speakers (actors, announcers), or by non-professionals (ordinary people)?
- 6) What size should the corpus be?
- 7) How many readers/speakers should be used?
- 8) Who and how many people should be used as emotion evaluators (listeners) in the perception tests?

The theoretical starting-points developed as an answer to the above questions formed a foundation for the practical creation of the Estonian Emotional Speech Corpus, on which these starting-points were subsequently tested.

<p style="text-align: center;">I. CHOICE OF EMOTIONS Joy Anger Sadness Neutral</p>
<p style="text-align: center;">II. CHOICE OF READING MATERIAL Journalistic text passages; Identification of emotion solely from writing (without hearing the text) minimum of 10 testers</p>
<p style="text-align: center;">III. CHOICE OF READERS Good articulation, pleasant voice, empathy; Reading and recording passages; Segmenting passages into sentences</p>
<p style="text-align: center;">IV. CHOICE OF LISTENERS Adult Estonians with good empathic abilities Minimum of 30 testers</p>
<p style="text-align: center;">V. LISTENING TEST Identification of sentence emotion by audition; Joy? Anger? Sadness? Neutral?</p>
<p style="text-align: center;">VI. READING TEST Identification of sentence emotion by silent reading; Joy? Anger? Sadness? Neutral?</p>
<p style="text-align: center;">VII. CLASSIFICATION OF SENTENCES Comparison of the results of listening and reading tests for the classification of sentences (see Table 1): 1) Verbal-semantic content may affect the identification of sentence emotion; 2) Verbal-semantic content does not affect the identification of sentence emotion</p>

Figure 1. Creation of the Estonian Emotional Speech Corpus (see [P1], [P2], [P3], [P4])

The theoretical starting-points for the Corpus design and its implementation was as follows ([P1], [P2], [P3], [P4], see also Figure 1):

TO START THE CORPUS WITH THREE EMOTIONS – joy, anger, and sadness – represent the three most essential emotions in speech technological applications and neutral speech (see Campbell, 2000; Iida, Campbell, Higuchi & Yasumura, 2003). (See Figure 1, Stage I). For Corpus purposes, the three emotions are not only interpreted as full-blown emotions but also as covering moderate emotion-related states. This is due to the fact that, in ordinary life, emotions are mostly expressed moderately while full-blown expressions are rare (Cowie & Cornelius, 2003).

NOT TO USE SIMULATED EMOTIONS AND ACTORS. The use of actors is the most straightforward way of generating emotional speech, as good actors under adequate guidance are quite capable of delivering expressions and conveying emotions which are identified by the listeners as authentic (Scherer, 2013). However, a study of existing research during the early stages of Corpus creation gave reason to believe, that actors may easily overact and the resulting expression of emotion may turn out to be too intense and prototypical in comparison with speech produced by a speaker experiencing a genuine emotion (Campbell,

2000; Iida, Campbell, Higuchi & Yasumura, 2003; Scherer, 2003). Thus it was decided that the authentic and moderately expressed emotions necessary for the Corpus should be obtained from written text passages read by ordinary people. As we assumed that the reader is stimulated by the verbal-semantic content to express the emotion latent in the text, we found it necessary to avoid dictating to the readers what emotion to express. (cf. Iida, Campbell, Higuchi & Yasumura, 2003; Navas et al., 2004). Thus the Corpus emotions can be called ‘elicited emotions’ as they have been elicited by the text.

TO RECORD TEXT PASSAGES, NOT SEPARATE SENTENCES, because the message of a passage helps the reader to better achieve the appropriate state of emotion (see Iida, Campbell, Higuchi & Yasumura, 2003). Before reading the texts out loud and recording them, a test has to be applied in which a group of testers read written passages silently to themselves to determine their emotion, so that only those passages conveying joy, anger, or sadness were picked for the Corpus (see Figure 1, Stage II). We decided to draw the material from the press, because 1) journalistic texts represent the average parameters of the Estonian textual space (Kerge, 2002; Kerge, Pajupuu, Tamuri & Meier, 2008) and also because, 2.) the primary purpose of the Corpus was an emotional text-to-speech synthesis for the Estonian Text-to-Speech-synthesizer, which was mainly designed for the audio representation of journalistic texts (Mihkla, Piits, Nurk & Kiissel, 2008).

TO MAKE A CAREFUL CHOICE OF THE READERS, chosen from a group of non-professional readers, they should have clear articulation and a pleasant voice (height and timbre). As empathic people are known to better understand emotions expressed by others, I presumed that they would also be better at expressing emotions themselves (cf. Baron-Cohen & Wheelwright, 2004; Chakrabarti, Bullmore & Baron-Cohen, 2006). The quality of reader candidates’ articulation was judged by experts. Their empathic ability was tested by means of Baron-Cohen & Wheelwright’s Empathy Quotient (or EQ, 2004). In addition, a voice test was carried out enabling the listeners to choose the reader with the most pleasant voice (see [P2], Figure 1, Stage III). As a result, a female reader was chosen whose empathy was higher than average and whose voice sounded pleasant to most of the listeners. She read 130 passages for the Corpus without being told which emotion to use. The instruction to the reader was: “Read the text so that you render its mood.” The word “emotion” was not used or referred to in any way. Thus the reader decided how exactly to read these passages. The recorded passages were segmented into sentences which I compiled, using the test interface of the Corpus, as tests to be presented to listeners for the identification of emotion from vocal cues in the sentence.

TO DETERMINE THE EMOTION OF THE CORPUS SENTENCES THROUGH LISTENING TESTS, MAKING A CAREFUL CHOICE OF THE LISTENERS (TESTERS). When the Corpus was being created, there was no full consensus on how well a listener could identify emotions from non-acted speech without the speaker being seen. Therefore careful consideration was given to the choice of

the subjects listening to the tests, in the expectation of raising the probability of the emotions being identified.

Previous research contained discussions linking more mature listeners to a better recognition of emotions from vocal cues than younger ones (e.g. students), which is because emotion recognition is a culture-specific skill that can be acquired only with time (see Toivanen, Väyrynen & Seppänen, 2004). It was decided to choose the listeners (testers) from ethnic Estonians aged over thirty who lived in Estonia and whose first language was Estonian. (See [P5], [P6] and Figure 1, Stage IV).

According to some studies, empathy, besides age, can also have an important role in emotion recognition (Baron-Cohen & Wheelwright, 2004; Chakrabarti, Bullmore & Baron-Cohen, 2006). Thus assuming that empathic people are better identifiers of emotions from vocal cues than non-empathic ones (see Keen, 2006), the empathic ability of the testers was measured by Baron-Cohen & Wheelwright's EQ (2004; for more detail of the EQ see [P7]). The testers were also optionally subjected to the EPIP-NEO questionnaire, which is a personality test (for its Estonian version NEO-PI-R, see Möttus, Pullmann & Allik, 2006). This enabled a closer look at the correlations, if any, between personality traits and the identification of emotion.

The Corpus database contains the following data for 190 registered testers: sex, age, education, nationality, mother tongue, language of education, work experience, empathy quotient, and personality profile.

Of the 1234 sentences contained in the passages recorded for the Corpus, 14 online listening tests were compiled. The underlying principle of the tests was that the content of any two successive sentences would not form a logical sequence.

The participants of the listening test were asked to listen to separate sentences, without context, and to decide without seeing the written text, which emotion they heard. The available choices were the three emotion families – joy, anger, sadness, and a neutral attitude. I used the same test method for the research on emotion perception (see Chapter 3). Each sentence was listened to by at least 30 testers (see Figure 1, Stage V). An emotion was considered as successfully identified if more than 50% of listeners agreed (i.e. emotion is identified over two times better than chance probability) (see [P3], [P4]).

TO VERIFY THE ROLE OF THE VERBAL-SEMANTIC CONTENT IN IDENTIFYING THE EMOTION OF THE SENTENCE HEARD. Regarding the listening tests, it was necessary to find out how much impact verbal-semantic content on the identification of emotions. For this purpose, 14 reading tests were compiled out of the listening sentences. Here, the testers had to decide emotion or neutrality by reading the written sentences; that is, by using the verbal-semantic content without hearing the sound. The reading tests did not include the same people who had identified emotions audibly by listening (see Figure 1, Stage VI). Table 1, which is a modification of the tables of [P2], [P3], [P4], [P5], shows how the Corpus sentences were divided into two groups based on the comparison of the listening and reading test:

- 1) Sentences in which the identification of emotion may have been influenced by verbal-semantic content. These are the sentences where one and the same emotion was found in both listening and reading tasks. Here, one cannot be confident that the emotion is carried solely by the sound of the voice as it may have been identified as well from the verbal-semantic content of the sentence (see example 3 in Table 1).
- 2) Sentences in which the emotion has been identified from vocal cues only. These are the sentences in which different emotions were detected in listening and reading tasks. In the case of this group it is sure that the emotion is carried by the voice, i.e. vocal expression has altered the meaning of the verbal-semantic content of written text (examples 1 and 2 in Table 1).

As far as we know, the Estonian Emotional Speech Corpus is unique among the publicly available corpora of emotional speech in that its material has been divided into two categories depending on whether the vocal expression corresponds to the verbal-semantic content of the written text or not (see Figure 1, stage VII).

Based on the theoretical starting-points listed above, 1234 sentences (all different) have been collected in the Corpus. All those sentences have been run through both the listening and reading tests. In 908 sentences out of the 1234, more than 50% of listeners identified one and the same emotion or neutral status (see [P4]). Table 2 contains a classification of the current Corpus sentences (see Table 2).

Table 1. Sentence classification in the Corpus according to how its emotion was identified in listening vs. reading tests (the test results are presented as a percentage)

Tests	Emotions					Sentence type in corpus
	Joy	Anger	Sadness	Neutral	Not sure	
<i>1. Ehkki Ott minu olemasolust midagi ei teadnud. [Although Ott knew nothing of my existence.]</i>						
By listening	87.5	0.0	0.0	12.5		Joy, no verbal-semantic content influence
By reading	4.0	0.0	32.0	32.0	32.0	
<i>2. Ükskõik, mida ma teen, ikka pole ta rahul! [Whatever I do, he is never satisfied!]</i>						
By listening	0.0	14.3	80.0	5.7		Sadness, no verbal-semantic content influence
By reading	0.0	64.3	35.7	0.0	0.0	
<i>3. Täiesti mõistetamatu! [Completely incomprehensible!]</i>						
By listening	0.0	100.0	0.0	0.0		Anger, verbal-semantic content influence
By reading	0.0	83.3	0.0	11.1	5.6	

Note. The category ‘not sure’ was added for cases when the subject finds it hard to decide on one particular emotion in a sentence and feels that the emotion instead depends on how the sentence is read out loud.

Table 2. Classification of the Corpus sentences (see [P4])

Emotion	Sentences	Verbal-semantic content influence on identification	No verbal-semantic content influence on identification
Joy	232	163	69
Anger	277	177	100
Sadness	191	88	103
Neutral	208	87	121
Unable to identify	326		
Total	1234	515	393

Although such double testing of each Corpus sentence is rather time consuming, it works as a measure of validation for the Corpus. As a result, Corpus users can be sure that Corpus sentences contain identifiable emotions (see [P4]). Users can select either sentences where the emotion is rendered by voice only or sentences where the emotion is also rendered by verbal-semantic content.

As the emotions of the sentences contained in the Estonian Emotional Speech Corpus were determined by listeners, some issues of emotion perception came to the fore. Chapter 3 will demonstrate if the choices of testers was justified (see [P5], [P6], [P7]).

3. PERCEPTION OF EMOTION

The theoretical base of the Corpus required verification regarding the choice of testers, who were to identify sentence emotion. When designing the Corpus, I assumed that different people do not perceive emotions in the same fashion. So I tried to find testers who would identify emotions as effectively as possible. While creating the Corpus, I decided to pick testers from among empathic adults aged over thirty who lived in Estonia and spoke Estonian as their mother tongue. In my study, I wished to confirm whether the principles of tester selection were appropriate, which led me to ask the following research questions:

- 1) Is sentence emotion identifiable purely from vocal cues, without the speaker being seen (see [P3])?
- 2) Can age affect the identification of emotion (see [P5])?
- 3) Is the identification of emotion culturally bound (see [P5]), [P6])?
- 4) Does identification depend on the listeners' empathy (see [P7])?

For each of the above problems, the material and methods have been described in detail in subchapters 3.1, 3.2, 3.3 and the corresponding articles ([P3], [P5], [P6]) and [P7]).

For the first question asking if the emotion of a sentence can be identified from non-acted vocal expression without the speaker being seen, results confirmed the supposition that listeners can recognize the moderate expression of non-acted emotions from the voice of a non-professional reader: over 50% of the listeners agreed on the emotion or neutrality in 73.5% of the Corpus sentences (see [P3], [P4]), see Table 3.

Table 3. Statistics of the emotional and neutral sentences identified by the listening test (see [P4])

Listening response	Joy	Anger	Sadness	Neutral
Number of emotional sentences identified by more than 50% of listeners	232	277	191	208
Mean percentage of identification and SD	75.4 14.5	73.3 14.6	72.1 14.7	68.3 11.9

In order to find out how a person's age, culture, and level of empathy could possibly affect their ability to identify emotion, I compiled a web-based **listening test**. For that purpose, I searched the Estonian Emotional Speech Corpus for 10 sentences of joy, 10 sentences of anger, 10 sentences of sadness, and 5 neutral sentences, all with an emotion identification score of over 65% (i.e. over 2.5 times higher than chance probability) in which the emotion was identifiable from vocal cues (i.e. from how the sentence is read out) rather than from verbal-semantic content. When compiling the test, I made sure that no two consecutive

sentences would make a logical whole. The list of the test sentences and the identification rate of emotion in each sentence in the Corpus are presented in [P6].

The testers were asked to listen to isolated sentences without the written text being seen and decide which emotion they heard in each sentence. The choice was between the families of joy, anger, sadness, and neutrality (for elaboration, see Chapter 1, section 1.4). The test was done with the web-based testing system of the Estonian Emotional Speech Corpus. The testers were free to listen to sentences as many times as they wished. (See [P5], [P6], [P7])

3.1. Age Influence on the Identification of Emotion

Age has been considered one of the factors affecting perceptions of emotion. On the one hand, it has been shown that understanding emotion decreases with age (Mill, Allik, Realo & Valk, 2009; Paulmann, Pell & Kotz, 2008). Paulmann, Pell & Kotz (2008), for example, have found considerable differences between the classification of emotions through voice detection by middle-aged subjects (mean age, 42.6 years) and younger subjects (mean age, 23.4 years). Although mistakes were made by both groups, the younger subjects were less prone to error. Moreover, Mill, Allik, Realo & Valk (2009) found that even young adults between 21–30 are slightly less capable of recognizing sadness and anger than their juniors, while a considerable decline in the perception of negative emotions occurred in the group aged 31–40. No significant differences between the age groups were revealed for neutral expressions (Mill, Allik, Realo & Valk, 2009). On the other hand, the expression and interpretation of emotions are culturally bound and can be learned with time (see Dewaele, 2010; Nazir et al., 2009, [P5], [P6]). Based on studies which have demonstrated cultural specificity in recognizing emotions, I decided not to have the Corpus emotions identified by young testers under the age of 30. This presumes that older people will recognize emotions better because they have a longer experience of living in the culture, which is necessary for obtaining the culture-specific expression and decoding of emotions. In [P5], I focused on the problem of how appropriate it was to let the sentence emotions be defined by testers aged over 30, leaving younger listeners aside. For that purpose, I investigated the influence of both age and culture on the perception of emotion. I arranged an experiment with two groups of subjects participating: 1) young adult Estonians aged 20–28 who spoke Estonian as their mother tongue; and 2) Latvians aged over 30 who could not speak Estonian. The results were compared with those of the Corpus testers who were aged over 30 (in particular, 30–62) and spoke Estonian as their mother tongue. Latvians were used as testers to find out how important it is to live in a specific culture in order to be able to recognize emotions from vocal expression. Latvians were chosen because Estonia and Latvia are neighbouring countries and their cultures are based on more or less similar values (Huettigen, 2008), which allowed me to assume that the emotional expression of the two

peoples is not too different to enable recognition of each other's emotions from vocal cues.

The testers were asked to listen to 35 Estonian sentences and determine the emotion of each sentence from what was heard (see the test description in Chapter 3). The results revealed significant differences in the identification of emotion. The difference between Estonian adults and young adults was not significant in the cases of sadness and neutrality. As for anger, the difference was not significant between Estonian young adults and Latvians. Both the Estonian adults and young adults identified the target emotion in more than 50% of the sentences, but the group of young adults scored lower than the adults. The young adults determined many more of the sentences to be neutral.

The Latvians surpassed the 50% threshold only in identifying sadness, yet they differed significantly from Estonians in their precise identification. Like Estonian young adults, the Latvians often perceived the sound of the sentences as neutral. It shows that in order to successfully identify an emotion in voice, one has to have a longer experience in how emotions are expressed in a certain culture (cf. Pell, Paulmann et al., 2009).

Thus the results support the decision that the emotions of the sentences in the Estonian Emotional Speech Corpus should be determined by Estonian adults aged over 30 who speak Estonian as their native language because they are more likely to have acquired the skills for decoding the culture-specific expression of emotions.

The Latvians' ability to identify Estonian emotions by vocal expression was relatively low, which seems to confirm that emotion recognition from the voice is culturally bound. (See [P5])

3.2. Cultural Influence on the Identification of Emotion

Intercultural studies on emotion recognition support the general view that culture has a say in how we express and understand emotions and what emotional signals are considered socially acceptable (Soto & Levenson, 2009). Recent intercultural results on emotion recognition from vocal cues have revealed the existence of cultural influence on emotion identification as well as universal tendencies (Bryant & Barrett, 2008; Pell, Paulmann et al., 2009; Scherer, Clark-Polner & Mortillaro 2011; Thompson & Balkwill, 2006). As was proved in [P5], emotions expressed in a non-native language might not be recognized – even if the speaker comes from a geographically close area and has a cultural background formed by many historical similarities and characterized by similar values.

However, while the majority of studies on cultural differences in the recognition of vocal expressions of emotion have compared members of different ethnic groups living in different countries, my intention in [P6] was to find out how emotions are recognized by people who speak different languages but live in the same country. More precisely, I explored how emotional expression

is recognized in a situation where the emotions are conveyed not in one's first language (L1) but in their second (L2). The focus group consisted of self-defined ethnic Russians living in Estonia, for whom Estonian is the second language. I wished to know: 1) if the ethnic Russians living in Estonia can recognize emotions expressed in Estonian; 2) if they recognize them in the same way as the Estonians; or 3) if not, whether their perception of emotion is more like that of the ethnic Russians living in Russia. The test was thus taken by three groups of adults: 36 native speakers of Estonian living in Estonia (mean age $M=41.3$, $SD=9.2$); 16 native speakers of Russian living in Estonia (mean age $M=35.8$ years, $SD=8.1$) who could speak Estonian (L2) on a high (C1) level; and 16 native speakers of Russian living in Russia (mean age $M = 33.1$, $SD = 5.0$), who could not speak Estonian. All of the testers were asked to listen to 35 Estonian sentences and determine the emotion of each sentence without seeing the text (see the test description in Chapter 3). The comparison of the listening results demonstrated that language skills and an opportunity for intercultural communication have an influence on emotion recognition. The Russians living in Estonia were able to recognize all Estonian emotions with an accuracy of more than 50%, while there was no significant difference between the two groups with regard to anger or sadness. But for joy and neutrality, the difference was significant: the Russians living in Estonia displayed a lower accuracy of recognition than ethnic Estonians and confused joy with neutrality. The native speakers of Estonian differed significantly from the Russians living in Russia in the recognition of all three emotions as well as neutrality like the Russians living in Estonia and Russians living in Russia. Thus the recognition of Estonian emotions by the Russians living in Estonia resembled that of the Estonians rather than the Russians living in Russia. The Russians living in Russia recognized only sadness in more than 50% of the cases. They interpreted Estonian joy as a neutral expression and could not tell the difference between Estonian anger and neutrality. These results imply that the understanding of emotions depends on cultural factors and social interactions, including the social norms specific to one culture. The interpretation of emotional messages is therefore learned in the course of social interactions. (See [P6])

3.3. The role of Empathy in the Identification of Emotion

Another question was if empathic ability may play a part in the recognition of vocal emotions when they are moderately expressed in speech which is listened to without the speaker being seen. I had, after all, chosen empathic people to determine the emotions of the Corpus sentences presuming that they can recognize emotions better than the non-empathic ones. The study is reported in [P7]. The research question is whether a person's empathy level (low, medium, high) can affect their ability to recognize vocal emotions. That is to say, how appropriate it is to have only empathic listeners to determine the Corpus

emotions and to leave aside the results of the non-empathic listeners, as they possibly do not recognize emotion from voice.

The available studies on the relationship of empathy to the ability of recognizing purely vocal emotions serves mainly the clinical purpose of investigating how emotions are recognized by people whose diagnosis involves a low level of empathy (e.g. autism, Asperger's Syndrome, psychopathy). Most clinical studies have indicated that a low empathy level may well be accompanied by difficulties in the recognition of emotions and mental states from the voice (Golan, Baron-Cohen & Hill, 2006; Golan, Baron-Cohen, Hill & Rutherford, 2007; Rutherford, Baron-Cohen & Wheelwright, 2002).

The study reported in [P7] differs from the earlier ones in two aspects: (1) the testers have no diagnosed disorders that might reduce one's empathy; and (2) the emotions presented for recognition are not acted but elicited. In addition, these are not stereotypically expressed full-blown emotions, which are rarely met in everyday life, but moderately expressed emotions belonging to the families of joy, anger, and sadness. The listening test was taken by 67 adults (29 males and 38 females, aged 30–60, $M=42.0$, $SD=10.7$). Their empathic ability was measured through an Estonian translation of the EQ self-report questionnaire designed by Baron-Cohen & Wheelwright (2004).

According to empathy scores, the participants of the listening test were first divided into three groups: low empathy (under 32 points), medium empathy (32–51 points) and high empathy (over 51 points) (cf. Bate, Parris, Haslam & Kay, 2010).

In our sample, only men belonged to the low-empathy group and only women had high empathy. This is not surprising in light of other studies that have also revealed that women's empathic ability tends to be higher than men's (Baron-Cohen & Wheelwright, 2004; Lawrence et al., 2004). Next, the group with medium scores was divided into male and female subgroups. Test groups emerged as follows:

- I. Low empathy: EQ 18–31, $M=26.2$, $SD=3.2$
14 men aged 32–44, $M=36.7$, $SD=4.5$
- II. Medium empathy: EQ 32–51, $M=42.5$, $SD=5.7$
 - a) 15 men aged 31–56, $M=39.0$, $SD=7.6$; EQ 32–50, $M=41.5$, $SD=5.8$
 - b) 22 women aged 31–60, $M=38.5$, $SD=6.4$; EQ 34–51, $M=43.1$, $SD=5.1$
- III. High empathy: EQ 52–60, $M=54.9$, $SD=2.6$
16 women aged 30–58, $M=44.1$, $SD=10.5$

In the listening test, the testers were asked to listen to 35 isolated sentences and decide what emotion was heard (see the test description in Chapter 3). As the test groups differed in gender – the low empathy group consisted of men and the high empathy group consisted of women – the first task was to find out the possible influence of gender on the recognition of emotions. For that purpose, we analysed the men and women with medium empathy since their group was the only group containing both sexes. The results did reveal some gender influence on the recognition of two emotions, joy and anger. Therefore the emotion scores were only compared within genders, comparing the male groups

(low versus medium empathy) with each other and the female groups (medium versus high empathy). According to the results, there was no significant difference either between the emotion recognition scores of the male groups or between those of the female groups. That is to say, in test sentences with moderately expressed emotions, joy, anger, sadness, and neutrality were identified in more than 50% of the cases, regardless of the groups' gender or empathic ability.

By generalizing our results, it can be stated that, in the recognition of emotion from vocal cues, empathy is less important than clinical results would suggest. In conducting emotion studies for speech technological purposes, it is obviously unnecessary to exclude non-empathic people from the testers for the reason that they may not recognize the emotions expressed if their low empathy level is not due to mental or developmental disorders. At the same time, it would be recommended to involve both male and female subjects for the differences in their emotion recognition. (See [P7].)

4. RESULTS AND DISCUSSION

4.1. The Estonian Emotional Speech Corpus among other Corpora of Emotional Speech

As outlined previously, the aim of this study was to develop a theoretical base for the Estonian Emotional Speech Corpus and to test the validity of the theoretical starting-points on Corpus material.

The Corpus is now ready as designed (see [P1], [P2], [P3], [P4], <http://peeter.eki.ee:5000>). The results of my research reveal the importance of detailed planning and of the design elements of the Corpus. The theoretical starting-points of the study are relevant and applicable in real situations. Therefore these results could be taken into consideration in the creation of other emotional speech corpora.

What makes this Corpus unique among the other corpora of its kind is the fact that its sentences have different labels according to whether their emotion is carried just by the sound of the sentence or whether the recognition of their emotion from vocal expression may be influenced by the verbal-semantic content. This classification enables the research of emotions both in speech as well as in writing.

The following overview of emotional speech corpora is intended to provide a comparative international background.

According to how the recorded speech has been obtained, emotional speech corpora can be divided into three major groups (see the surveys by Pajupuu, 2012; Schuller, Batliner, Steidl & Seppi, 2011). The largest group consists of corpora containing acted emotions. While collecting the material the speakers/readers are instructed which (basic) emotion they should express. Often one and the same sentence is asked to be read with different emotions (see, e.g., the Berlin Emotional Speech Database (Burkhardt et al., 2005) and the Danish Emotional Speech Database (Engberg, Hansen, Andersen & Dalsgaard, 1997)).

The second group consists of corpora containing elicited emotions. In order to obtain such material, an emotive situation is arranged. The material of the German FAU Aibo Emotion Corpus, for example, has been collected from human-computer interaction (Batliner et al., 2004). Emotions can also be elicited by a written text. This method was used in building up the Japanese emotional speech synthesis system CHATR (Iida, Campbell, Higuchi & Yasumura, 2003). Elicitation usually produces emotion-related states (e.g. resentment, indifference, irritation) rather than full-blown emotions.

The third group consists of corpora containing emotions collected from real-life situations, such as, for example, the German *Vera am Mittag* (VaM) (Grimm, Kroschel & Narayana, 2008). This is a multi-modal corpus in which the voice and facial expressions are tagged.

The list of emotion corpora and databases by J. Pitterman, Pitterman & Minker (2010: 237–250) contains 103 emotional speech corpora and databases for sixteen languages: English (33), German (16), Chinese (10), Japanese (6),

French (5), Dutch (4), Spanish (3), Swedish (3), Basque (2), Korean (2), Burmese (1), Danish (1), Finnish (1), Greek (1), Hebrew (1), Russian (1), multi-lingual (13). Only a minor part of them (15) are multi-modal. The speech unit recorded is a sentence or utterance in 64% of all the corpora or databases. In 57% of cases, the emotions are acted or simulated. Most of the corpora or databases of emotional speech are small ad-hoc datasets, described only in the article reporting that particular study. Several project-based corpora and databases have ceased to be developed after the project was over. There are few corpora or databases open to public access. See Table 4 (Raport 2011).

Table 4. Available corpora and databases of emotional speech as compared with the Estonian Emotional Speech Corpus (Raport 2011)

The Estonian Corpus vs Others	No. of Emotions	No. of Speakers	No. of Sentences/ Utterances	Acted or Simulated Emotions	Contains Neutral Utterances	Publicly available
Min	1	1	15	57%	70%	10%
Q1	3	2	263			(may
median	4	8	732			require
Q2	6	36	2333			registration
Max	15	784	20013			or
				None	Yes	payment)
The Estonian Corpus	3	1	1234	(emotions elicited from reading)		Yes (publicly available, no payment)

In comparison with other corpora of emotional speech, the Estonian Emotional Speech Corpus is one of the few freely available documented ones that reviews moderately expressed emotions. The Corpus abandoned acted emotions because of their possible stereotypicality and overactedness. The sentences recorded for the Corpus were read out by a so-called ordinary person, who was not dictated what emotion to use while reading. The following general guideline was given: “Read out the text so that you render its mood.” The reader could decide how to interpret the text. Hence, emotions were not mentioned. In this aspect of data collection, the Corpus coincides with the Japanese emotional speech corpus in the assumption that the emotion latent in the reading material will be elicited by the reader and detected in the reader’s voice (see Iida, Campbell, Higuchi & Yasumura, 2003). This method of collecting emotional expressions was effective, and as a result emotions were sufficiently well identified by listeners.

As for corpus size, the Estonian Emotional Speech Corpus is bigger than average but hitherto there is only one speaker. The use of just one speaker was due to the fact that our first task was to create a corpus sufficient for emotional

unit selection based speech synthesis. However, the theoretical base of the Corpus was developed with a multi-purpose view, considering research as well as various speech technological applications. The size of the Corpus enables its use in speech technology for statistical modelling of the prosody of synthesized speech. In addition, the Corpus design enables extension through new speakers, video material, pictures, and spontaneous speech. Also, the emotions can be divided into additional categories and determined on several dimensions. For example, by now all of the Corpus sentences have passed not only the emotion tests (categorizing them into sadness, joy, anger, and neutrality) but also the dimension test of valence. This divides them into negative, positive, and neutral (both in listening and reading) as well as the dimension test of activity or passivity (in listening only). By testing the emotions of the 1234 Corpus sentences, our testers were able to categorize 908 sentences as belonging either to joyous, angry, sad, or neutral ones, while 326 sentences were left unlabelled. As for valence, more were classified – resulting in the classification of 1165 sentences out of 1234: 269 as positive, 675 as negative, and 211 as neutral. A combination of the categorization and dimensional classification of the emotions enables the researcher to study, for example, the negative sad sentences against the positive sad sentences, and so on.

4.2. The Influence of Age, Culture, and Empathy on the Identification of Emotion

In the course of this study, I tested the validity of three starting-points underlying the theoretical base of the Estonian Emotional Speech Corpus. The findings demonstrated the following:

- 1) Adults (aged over 30) are better at emotion recognition from vocal expression than young people (aged 20–30);
- 2) Emotion recognition from vocal expression is culturally bound;
- 3) Empathic level does not affect the recognition of vocal emotions; see [P3], [P5], [P6], [P7].

The emotion labels attached to the Corpus sentences are based on perception tests conducted with adult testers aged over 30. In order to be identified as such, the emotion has to be agreed upon by more than 50% of the testers aged over 30.

According to the results of [P5], there is no single answer to the question of whether it was justified to use emotion testers aged over 30, hoping they would recognize emotions better than young adults. On the one hand, there was a significant difference between those aged over 30 and the younger group. On the other hand, neither group had problems in the identification of emotions from vocal expression: the consensus on what emotion can be heard in this or that sentence is high for both groups (over 50%, see Chapter 3, section 3.1). Yet there is a tendency among younger subjects to define more sentences as neutral. As for each Corpus sentence, the emotion label comes from listeners and there

is no way to judge which group responded correctly. There are, however, two circumstances indirectly supporting the choice using people over 30 testers: 1) the ability of the adult testers to identify emotions from voice has not decreased with age (sadness is identified similarly with the young adults' group (cf. Mill, Allik, Realo & Valk, 2009); and 2) the adults' group interpreted more sentences as emotional than the group of young adults who preferred 'neutral' more often, which indicates that the older testers had better skills of decoding culture-specific expressions of emotion. The importance of culture in the identification of emotion has also been confirmed by experiments with Latvians, whose ability to identify Estonian emotions from voice was relatively low.

[P6] was focused on how culture may influence the recognition of moderately expressed emotions which are not conveyed in the first language (L1) of the listeners but in their second language (L2). In particular, I sought an answer to the questions of whether native speakers of Russian who live in Estonia and speak Estonian as their second language are able to recognize emotions expressed in Estonian and whether they recognize emotions in similar patterns to ethnic Estonians or to Russians who do not know Estonian. A comparison of the test results of the native speakers of Estonian, the ethnic Russians living in Estonia, and the Russians living in Russia demonstrated that the Russians living in Estonia could recognize all Estonian emotions with an accuracy of more than 50%, whereas the Russians living in Russia did not score over 50% in any category except for sadness.

Research has shown that negative emotions like anger and sadness are better recognised across cultures (see, e.g., Pell, Monetta et al., 2009; Scherer, Clark-Polner & Mortillaro, 2011; Thompson & Balkwill, 2006). The fact that the subjects from Russia managed to recognise Estonian sadness is in accord with those results. At the same time, the subjects from Russia were unable to recognize Estonian anger – confusing it (and joy) with neutrality. The reason could lie in the difference of the Russian and Estonian acoustic models for expressing anger, joy, and neutrality (see [P6]). On the other hand, the facts that the Russians living in Estonia did not differ from the Estonians in their recognition of Estonian anger and that they moreover mistook joy for confusion less frequently than the subjects from Russia is evidence that they have culturally acquired the Estonian way of expressing emotion. The results suggest that understanding emotions is largely dependent on whether a person lives in the culture where the language is spoken and whether the person is able to interact in second language (L2).

In [P7] I investigated the influence of empathy on the recognition of moderately expressed vocal emotions where the decision has to be made from the voice, without the speaker being seen. Although most of the previous studies have demonstrated that people with a low empathic ability have difficulties in the recognition of emotions from speech (Golan, Baron-Cohen & Hill, 2006; Golan, Baron-Cohen, Hill & Rutherford, 2007; Rutherford, Baron-Cohen & Wheelwright, 2002), the present study provides evidence that there is no significant difference in emotion recognition by people of low, medium, and

high empathy if there is no time limit on their decision making. The previous studies had been conducted with people suffering from conditions involving a lowered empathy level. According to my knowledge, the present study is the first to address the ability of healthy people who differ in their empathy level to recognize moderately expressed emotions from the voice.

A side result of [P7] was that men and women differ in their recognition of anger and joy, which indicates that both male and female testers should be used in emotion studies. In the case of the Estonian Emotional Speech Corpus, 58% of the 190 registered testers were female, so the requirement is fulfilled.

In summarizing the results of the perception tests, it can be stated that the principle of picking the testers from native speakers of Estonians who live in Estonia is slightly too restrictive, particularly when the group is further limited to those who are empathic and over 30. Notably, living long-term in a culture is useful for the identification of emotion identification, but empathic ability is not important enough to be measured before the selection of testers. However, the theoretical base was improved by the knowledge that Corpus testers should include people of both genders.

The work done on the Corpus has opened up some new lines of research. In future studies, more of the principles underlying the theoretical base could be examined. For example, a study has already been started on whether it is absolutely necessary for the identification of vocal emotions to eliminate verbal influence or, in other words, whether the acoustic parameters of emotions differ in sentence groups (1) in which the verbal-semantic content affects the listener's identification of emotion and (2) in which it does not (see [P4]).

As the Corpus can be easily extended, the addition of spontaneous speech could enable a comparison of elicited (by reading) and spontaneous emotions. The study would bring some new knowledge to stand alongside the comparative study of acted versus authentic expression done by Scherer (2011).

In addition, the Corpus also contains data on the testers' personality traits as assessed by the EPIP-NEO test (see Mõttus, Pullmann & Allik, 2006). In principle, this enables future studies of the possible personality effects on emotion recognition like, for example, how emotions are perceived by introverted, extroverted, or neurotic people.

Finally, the Corpus continues to be developed according to the requirements of new research directions. As the Corpus is publicly available and accessible for free, its data can be used for tackling different research challenges.

5. CONCLUSION

In the framework of the thesis, a theoretical base for the Estonian Emotional Speech Corpus has been created and the underlying theoretical starting points have been tested. The results of the studies reported in the thesis have been surveyed in seven publications, four of which ([P1], [P2], [P3] and [P4]) describe the theoretical base, implementation, and options of the Corpus, while the remainder ([P5], [P6], [P7]) are focused on the perception of emotions in vocal expression.

On the foundation of this theoretical base, the Estonian Emotional Speech Corpus was constructed which is now used by researchers as well as by speech technologists. The Corpus material has been obtained from passages read by a non-professional reader. It was assumed that every text elicits a mood in the reader and this mood is somehow expressed in the reading voice. Thus this Corpus does not contain acted emotions. The 130 passages read aloud were segmented into 1234 sentences, isolated from context, and arranged into fourteen listening tests – each of which was presented to no fewer than thirty listeners (Corpus testers) who were asked to identify the emotions perceived in each sentence and, accordingly, to classify them into four groups: joy, anger, sadness, and neutrality. Prior to classification, the testers had been instructed that they were not expected to hear full-blown basic emotions and that each classification was to cover related emotions. All the sentences have also been passed through a reading test in which the emotion had to be determined without the sound of the sentence being heard. The aim was to find the sentences in which emotion can be identified from vocal expression without any influence from verbal-semantic content. Thus each of the Corpus sentences passed both listening and reading tests. To my knowledge, the Estonian Emotional Speech Corpus is the only freely available corpus of vocal emotions where the material has been classified according to whether the results of the listeners who have been asked to identify emotions from voice can or cannot be affected by the verbal-semantic content of the sentence.

As the Corpus emotions have been determined by using perception (listening) tests, I wondered if it was appropriate to pick the testers (the persons determining the Corpus emotions) from among empathic adult native speakers of Estonian who live in Estonia and are aged over thirty. While working on the present thesis, I took a closer look at whether the recognition of moderately expressed vocal emotions is affected by age, culture, and empathic ability. The results indicated that both a person's age and culture can affect their understanding of emotions. As for empathy, however, the results did not confirm its importance in the recognition of vocal emotions by a non-clinical population. Consequently, it was acceptable to pick the testers from among Estonian adults aged over thirty while the exclusion of non-empathic people from the tester candidates is not necessary. The study also revealed that men and women identify emotions differently. In future, this aspect might be taken into account in emotion recognition research.

Overall, the creation of a theoretical base for the Corpus, in advance of its practical development, was fully justified. The methodology and results of this study may be usefully applied to other studies into the perception of emotions and to the compilation of speech corpora.

6. SUMMARY IN ESTONIAN

Eesti emotsionaalse kõne korpuse loomine ja emotsioonide tajumine

6.1. Väitekirja eesmärk ja lähted

Viimasel kümnendil on kõneemotsioonide uurimine ja töötlus muutunud eksootilisest uurimisvaldkonnast peavooluteemaks. See on toonud kaasa vajaduse emotsionaalse kõne korpuste järele. Eestis algas emotsionaalse kõne korpuse loomine 2006. aastal riikliku programmi „Eesti keele keeletehnoloogiline tugi (2006–2010)” raames. Selle korpuse teoreetiline alus on käesoleva uurimuse tulemus.

Eesti emotsionaalse kõne korpusele sai pandud kaks peamist ülesannet:

- 1) olla usaldusväärne andmekogu kõnes ja kirjas avalduvate emotsioonide uurimiseks;
- 2) olla rakendatav kõnetehnoloogilistes ülesannetes. (Vt artiklid [P1]; [P2]; [P3]; [P4].)

Selleks ajaks, kui Eestis emotsionaalse kõne korpuse loomine algas, olid maailmas juba mõned kõneemotsioonikorpused olemas. Enamasti olid need väikesemahulised andmestikud, mis olid mõeldud üksikjuhtude uurimiseks, ja polnud laialdaselt kättesaadavad või oli nende arendamine sootuks lõpetatud (Douglas-Cowie, Campbell, Cowie & Roach, 2003). Samuti põhinesid sel ajal, ja põhinevad ka praegu, suur osa kättesaadavaid emotsionaalse kõne korpuseid näideldud materjalil, st üksiklausetel või tekstidel, mida on lugenud professionaalsed näitlejad, kel on palutud simuleerida erinevaid emotsioone. Näitlejate kasutamist emotsionaalse kõne saamiseks on aga jätkuvalt kritiseeritud, põhjuseks tuuakse näideldud emotsiooniväljenduste vähest autentsust ning loomulikkust (vt Douglas-Cowie, Campbell, Cowie & Roach, 2003; Laukka, Neiberg, Forsell, Karlsson, Elenius, 2010; Scherer, 2013) või ülemängimist (Wilting, Kraemer & Swerts, 2006).

Minu doktoritöö eesmärk oli luua teoreetiline alus Eesti emotsionaalse kõne korpusele ning kontrollida teooria kehtivust. Selleks töötasin läbi olemasolevate emotsioonikorpuste kohta avaldatud materjalid ja uurimused ning, arvestades nende valmiskorpuste kirjeldusi ja saadud kogemusi, kujundas teoreetilised seisukohad, mille järgi Eesti emotsionaalse kõne korpus praktiliselt teostati. Neid seisukohti kontrollisin loodud korpuse materjali põhjal. (Vt [P1]; [P2]; [P3]; [P4]; [P5]; [P6]; [P7].)

Varasemaid uurimusi analüüsid võtsin Eesti emotsionaalse kõne korpuse jaoks vastu järgmised põhiotsused.

- 1) Mitte võtta korpusesse näideldud kõnet, et saada korpusesse autentsemad ja ülemängimata emotsioonid.
- 2) Korpus peab koosnema kõneüksustest, mille emotsiooni on määranud kuulajad (korpuse testijad).

Eeskuju nende otsuste langetamiseks andis tol ajal uudsetel alustel loodud jaapani emotsionaalse kõne sünteesisüsteemi CHART korpus, kus näitlejatest

loobumine ning salvestatud kõne emotsiooni identifitseerimine kuulajate poolt oli end õigustanud (Iida, Campbell, Higuchi & Yasumura, 2003).

3) Korpus peab olema laiendatav sõltuvalt uurimisülesandest ja keele- tehnoloogilisest eesmärgist.

4) Korpus peab olema avalikult kättesaadav kõikides arenguetappides. Väitekirja artiklid on kirjutatud küllaltki pika perioodi 2008–2013 jooksul. Kuna kõneemotsioonide uurimine on noor, kuid pidevalt arenev valdkond, siis on sarnaselt teiste uurijate publikatsioonides olevale mõistekasutusele ka minu artiklites kõnesoleva aja jooksul terminoloogia täpsustunud ja ühtlustunud.

Emotsiooni defineerimisel olen lähtunud kõnetehnoloogias levinud laia- tähendusest: emotsioon on see, mis on enamasti kogu aeg olemas, ent puudub, kui inimesed on emotsioonitud (Cowie, Sussman & Ben-Ze'ev, 2011; Schuller, Batliner, Steidl & Seppi, 2011). Rääkides selles töös emotsioonikategoriatest rõõm, viha ja kurbus, ei mõelda nende all ainult täismahulisi põhiemotsioone, vaid pigem emotsiooniperesid, mis hõlmavad nii lähedasi emotsioone kui ka emotsioonilaadseid seisundeid. Ka tajukatsetes, kui kuulajatel tuli identifitseerida rõõm, viha või kurbus, juhendati neid mõtlema nendest emotsioonidest kui emotsiooniperedest, mitte ootama kõlamas stereotüüpset põhiemotsiooni. Nii tuli rõõmu all mõelda ka hääles väljenduvat tänulikkust, õnnelikkust, meeldivust, vaimustust; viha all pahameelt, irooniat, vastumeelsust, põlgust, kahju- rõõmu, raevu; kurbuse all üksindust, trööstitust, murelikkust, lootusetust (vrd Iida, Campbell, Higuchi & Yasumura, 2003). Neutraalne kõne on siin töös defineeritud kui ilma eriliste emotsioonideta kõne. Millised laused selliseks klassifitseeritud, otsustasid samuti kuulajad.

Korpuse loomist alustades ei olnud täit kindlust, kui hästi tuntakse häälest ära mõõdukalt väljendunud mittenäideldud emotsioone. Kuna Eesti korpuse kõneüksuste (lausete) emotsiooni otsustasid kuulajad, siis kuulajate valimise kriteeriumid olid teoreetilises aluses olulisel kohal. Kas need kriteeriumid end õigustasid, seda kontrollisin tajukatsetega (kuulamistestidega). Uurimuse emotsioonitajuga seotud eesmärgiks oli teada saada:

- 1) kas lauses mõõdukalt väljendunud emotsioon on ainult hääle järgi, ilma kõnelejat nägemata identifitseeritav;
- 2) kas inimese vanus mõjutab emotsiooni identifitseerimist;
- 3) kas emotsiooni identifitseerimine on kultuurisõltlik;
- 4) milline on empaatia roll emotsioonide identifitseerimisel.

Käesoleva töö tulemusele – korpuse teoreetilisele alusele – tuginedes on loodud Eesti emotsionaalse kõne korpus. Mitmed korpuse teoreetilised seisukohad on selle töö raames kontrollitud (vt [P1]; [P2]; [P3]; [P4]; [P5]; [P6]; [P7]). Korpus on avalikult kättesaadav <http://peeter.eki.ee:5000>.

6.2. Eesti emotsionaalse kõne korpus: teoreetiline alus ja teostus

Eesti emotsionaalse kõne korpus on avalik eesti keele emotsioone peegeldav korpus, mille teoreetiline alus (vt [P1]; [P2]; [P3]; [P4]) on kujundatud emotsioonikorpuste ja emotsioonide kohta tehtud uurimuste (nt Campbell, 2000; Cowie & Cornelius, 2003; Douglas-Cowie, Campbell, Cowie & Roach, 2003; Scherer, Banse & Wallbott, 2001; Ververidis & Kotropoulos, 2006) analüüsi põhjal. Eesti korpuse loomine algas 2006. aastal, mil usaldusväärsetel andmetel põhinevaid ja avalikult kasutatavaid korpusi oli küllaltki vähe. Esimese suurema emotsionaalse kõne andmekogu The Reading-Leeds loomist oli alustatud 1994. aastal (Greasley jt, 1995), enamasti aga oli kokku koondatud vaid väikesemahuline andmestik mingi üksikjuhu uurimiseks, mis polnud mõeldud avalikuks kasutamiseks. Olemasolevate kogude puudused oli põhiliselt seotud andmete vähesuse, valiiduse ja üldistatavusega (Douglas-Cowie, Campbell, Cowie & Roach, 2003).

Võttes arvesse varasemate emotsioonikorpuste ülevaateid ja emotsioonide kõnes avaldumise kohta tehtud uurimusi, tuli leida vastused küsimustele, mis puudutasid korpuse mahtu ning andmete valikut:

- 1) mis emotsioonid peaksid korpuses olema;
- 2) millist emotsionaalset kõnet peaks korpus sisaldama, kas spontaanseid, esilekutsutud või näideldud emotsioone;
- 3) kas korpuse materjal/tekst peaks olema kõneldud või loetud;
- 4) millised tekstid tuleks korpusesse valida; mis oleks nende sisu, pikkus ja kontekst;
- 5) kas tekste peaksid esitama professionaalsed lugejad (nt teadustajad, näitlejad) või võiksid need olla n-õ tavaliste inimeste ettekantud;
- 6) kui mahukas peaks korpus olema;
- 7) kui palju kõnelejaid/lugejaid peaks kasutama;
- 8) kui palju inimesi peaks kasutama emotsioonide hindamisel/tajukatsetes.

Nendele küsimustele vastuseks kujundatud teoreetiliste seisukohtade põhjal algas Eesti emotsionaalse kõne korpuse praktiline teostamine, ja hiljem, kui korpus loodud, teoreetiliste seisukohtade õigsuse kontrollimine. Teostamisetappide ülevaade on esitatud joonisel 2.

I. EMOTSIOONIDE VALIK rõõm, viha, kurbus, neutraalne
II. LUGEMISMATERJALI VALIK ajakirjanduslikud tekstilõigud lõikude emotsiooni määramine kirjaliku teksti põhjal min 10 testijat
III. LUGEJATE VALIK hea artikulatsioon, meeldiv hääl, empaatiavõime tekstilõikude lugemine, salvestamine, segmenteerimine lauseteks
IV. KUULAJATE VALIK täiskasvanud empaatilised eestlased min 30 testijat
V. KUULAMISTEST heli põhjal lause emotsiooni määramine rõõm? viha? kurbus? neutraalne?
VI. LUGEMISTEST kirjaliku teksti põhjal lause emotsiooni määramine rõõm? viha? kurbus? neutraalne?
VII. LAUSETE KLASSIFITSEERIMINE kuulamis- ja lugemistesti tulemuste põhjal lausete klassifitseerimine (vt tabel 5): <ul style="list-style-type: none"> • lause verbaalne sisu mõjutab emotsiooni identifitseerimist • lause verbaalne sisu ei mõjuta emotsiooni identifitseerimist

Joonis 2. Eesti emotsionaalse kõne korpuse loomise etapid (vt [P1]; [P2]; [P3]; [P4])

Korpuse loomiseks kujundatud teoreetilised seisukohad ja nende teostus oli järgmine ([P1]; [P2]; [P3]; [P4]; vt ka joonis 2).

- **VALIDA KORPUSESSE KOLM EMOTSIOONI:** rõõm, viha ja kurbus, kuna neid emotsioone on kõige enam vaja kõnetehnoloogilistes rakendustes, ning lisaks neutraalne kõne (vt Campbell, 2000; Iida, Campbell, Higuchi & Yasumura, 2003). (Vt joonis 2, etapp I.) Neid kolme emotsiooni ei käsitata korpuses täismahuliste emotsioonidena, vaid nendega seostatakse ka leebemaid emotsioonilaadseid seisundeid. Selline käsitlus lähtub asjaolust, et tavaelus väljendatakse emotsioone enamasti mõõdukalt ning täismahulisi emotsioone kohtab harva (vt Cowie & Cornelius, 2003).

- **LOOBUDA SIMULEERITUD EMOTSIOONIDEST JA NÄITLEJATE KASUTAMISEST.** Emotsionaalse kõne saamiseks on lihtsaim viis kasutada näitlejaid, kuna head näitlejad suudavad erineva tehnika ja juhendamise abil esitada kõnet, mille emotsiooni kuulajad küllaltki hästi ära tunnevad ja mis nende arvates ei kõla võltsilt (Scherer, 2013). On aga põhjust arvata, et näitlejad võivad emotsioone väljendades üle mängida ning nende emotsioonid olla liiga intensiivsed ja prototüüpsed ning seega liialt erineva ehtsatest kogetud emotsioonidest

(Campbell, 2000; Iida, Campbell, Higuchi & Yasumura, 2003; Scherer, 2003). Eesti korpuse jaoks langetati otsus autentsed mõõdukalt väljendatud emotsioonid kätte saada tavalise inimese ette loetud tekstilõikudest. Võis eeldada, et teksti sisu stimuleerib lugejat väljendama selles peituvat emotsiooni, mistõttu talle ei ole vaja öelda, millise emotsiooniga tuleb teksti lugeda (vrd Iida, Campbell, Higuchi & Yasumura, 2003; Navas jt, 2004). Sellisel korpusesse kogutud emotsiooni nimetan esilekutsutud emotsiooniks: emotsiooni esilekutsujaks on loetav tekst. Loomulikku kõnet ei valitud uurimismaterjaliks seetõttu, et korpuse loomise alguse kontekstis polnud väga palju teada spontaanse kõne korpuste uurimusi, millele oleks saanud Eesti emotsionaalse korpuse loomisel tugineda.

• **VALIDA KORPUSESSE SALVESTAMISEKS EMOTSIOONE VÄLJENDAVATE ÜKSIKLAUSETE ASEMELE TEKSTILÕIGUD**, sest lõigu sõnum aitab teksti ettelugejal paremini saavutada konkreetsele lõigule kohast emotsionaalset seisundit (vt Iida, Campbell, Higuchi & Yasumura, 2003). Enne lõikude ettelugemist ja salvestamist on vajalik korraldada test, milles rühm inimesi loeks lõike vaikselt omaette ja määraks nende emotsiooni. Korpusesse salvestamiseks valitakse neist välja ainult need, mille emotsioonis – rõõm, viha või kurbus – on testijad ühel meelel. (Vt joonis 2, etapp II.) Ajakirjanduslikud tekstilõigud neis peituvate emotsioonidega valiti korpuse materjaliks seetõttu, et 1) ajakirjandus esindab eesti tekstiruumi keskmisi näitajaid (Kerge, 2002; Kerge, Pajupuu, Tamuri & Meier, 2008) ning 2) korpuse esmast rakendust nähti emotsionaalses tekst-kõne sünteesis: eesti tekst-kõne süntesaator on mõeldud peamiselt ajakirjandustekstide ettelugemiseks (Mihkla, Piits, Nurk & Kiissel, 2008).

• **VALIDA HOOLIKALT TEKSTILÕIKUDE ETTELUGEJAT**, sest kuigi tegu on mitteprofessionaalse ettelugejaga, peaks tal siiski olema hea artikulatsioon ja meeldiv hää (kõrgus ja tämber). Ettelugejalt eeldati emotsionaalse teksti lugemiseks ka empaatiavõimet, kuna empaatilised inimesed saavad teiste emotsioonidest paremini aru, siis eeldasin, et nad on ka ise paremad emotsioonide väljendajad (cf. Baron-Cohen & Wheelwright, 2004; Chakrabarti, Bullmore & Baron-Cohen, 2006). Eesti emotsionaalse kõne korpuse tekstilugeja artikulatsiooni üle otsustasid eksperdid. Ettelugeja empaatiavõime määrati eesti keelde tõlgitud Baron-Cohen & Wheelwright'i (2004) empaatiatestiga. Lisaks sellele viidi kandidaatide hääle meeldivuse hindamiseks läbi hääletest, milles kuulajatel tuli välja valida kõige meeldivama häälega lugeja (vt [P2]; joonis 2, etapp III). Korpuse tekstilõike võeti lugema naisisik, kelle empaatia oli üle keskmise ning kelle hääle kuulajatele kõige enam meeldis. Ta luges korpusesse 130 tekstilõiku, kusjuures talle ei öeldud ette, millise emotsiooniga lõike lugeda. Talle anti vaid üldine juhised: lugeda teksti nii, et selle meeleolu esile tuleks. Seega sai ettelugeja ise otsustada, kuidas ta tekste loeb. Salvestatud lõigud segmenteeriti 1234 lauseks, millest koostas 14 kuulamistesti lausete emotsiooni määramiseks.

• **MÄÄRATA KORPUSE LAUSETE EMOTSIOON KUULAMISTESTIDEGA JA VALIDA HOOLIKALT KUULAMISTESTIS OSALEJAJD (KORPUSE TESTIJAJD)**. Korpuse loomise ajal ei olnud täit kindlust, kui hästi mittenäideldud kõnest ilma

kõnelejat nägemata emotsioone suudetakse identifitseerida, mistõttu tuli hoolikalt kaaluda, keda kuulamistestides osalejateks valida, et emotsioonide identifitseerimise tõenäosust suurendada.

Varasemates uurimustes oli diskuteeritud teemal, kas täiskasvanud võivad suuta emotsioone paremini ära tunda kui noored (nt tudengid) põhjusel, et emotsioonide äratundmine on kultuurispetsiifiline, mis saadakse selgeks kultuuris pikemalt elades (vt Toivanen, Väyrynen & Seppänen, 2004), Sellepärast otsustati valida korpuse testijateks üle 30-aastased Eestis elanud eestlased, kelle esimene keel on eesti keel. (Vt [P5]; [P6] ja joonis 2, etapp IV.)

Uurimused on näidanud, et kõrvuti vanusega võib emotsioonide äratundmisel olla oluline roll ka empaatial (Baron-Cohen & Wheelwright, 2004; Chakrabarti, Bullmore & Baron-Cohen, 2006). Eeldades, et empaatilised inimesed suudavad emotsioone hääle järgi paremini ära tunda kui mitteempaatilised (vt Keen, 2006), määrati kuulamistestis osalejate empaatiavõime Baron-Cohen & Wheelwright'i (2004) empaatiatestiga (testist pikemalt vt [P7]). Kuulajatele tehti soovi korral ka isiksusetest EPIP-NEO (NEO-PI-R eesti versioon vt Mõttus, Pullmann & Allik, 2006), et tekiks võimalus uurida isiksuseomaduste ja emotsioonide äratundmise seoseid.

Korpuses on 190 registreeritud testijat, kelle kohta on järgmised andmed: sugu, vanus, haridus, rahvus, emakeel, hariduskeel, töökogemus, empaatiatase ja isiksuse profiil.

Korpusesse salvestatud tekstilõikud segmenteeriti laustekes ning saadud 1234-st lausest koostas 14 veebipõhist kuulamistesti. Kuulamistesti põhimõte oli, et järjekorras laused ei moodustaks semantiliselt ühtset loogilist tervikut. Testimine viidi läbi korpusega integreeritud testiliidese abil. Testis osalejal tuli kuulata kontekstita üksiklauseid, ilma teksti nägemata, ja otsustada kuuldu põhjal, mis emotsioon lauses kõlas. Valida sai kolme emotsioonipere (rõõm, viha, kurbus) ja neutraalsuse vahel. (Sama testimismeetodit kasutasin ka emotsioonide tajuvõrdlevas uuringus, vt ptk 6.3.) Igat lauset kuulas vähemalt 30 testijat (vt joonis 2, etapp V). Emotsioon loeti identifitseerituks, kui üle 50% kuulajatest otsustas ühe ja sama emotsiooni kasuks (emotsiooni identifitseerimine ületab kaks korda juhusliku valiku tõenäosuse) (vt [P3]; [P4]).

• **TEHA KINDLAKS LAUSE SISU MÕJU LAUSE EMOTSIONI IDENTIFITSEERIMISELE.** Kuulamistestidega seoses vajab lahendamist küsimus, kui suur on lause verbaalse sisu roll emotsioonide identifitseerimisel. Selle kindlakstegemiseks on kuulamistesti lausetest koostatud 14 lugemistesti, milles testijatel tuli otsustada emotsiooni või neutraalsuse üle lauseid lugedes, st lause emotsioon tuli määrata verbaalse sisu põhjal ilma heli kuulmata. Lugemistestis osalesid teised inimesed kui need, kes identifitseerisid emotsioone heli põhjal (vt joonis 2, etapp VI). Tabel 5, mis on modifikatsioon [P2], [P3], [P4], [P5] tabelitest, näitab, kuidas on korpuse laused jagatud kahte rühma kuulamis- ja lugemistestide tulemuste võrdluse põhjal.

- 1) Laused, mille emotsiooni identifitseerimist võib olla mõjutanud lause verbaalne sisu. Sellesse rühma kuuluvad laused, milles testijad on identifitseerinud nii lugedes kui ka kuulates ühe ja sama emotsiooni

(kuulamistesti ja lugemistesti tulemused langevad kokku). Selle rühma puhul ei saa kindlalt väita, et emotsiooni kannab heli, emotsioon võib olla ära tuntud pelgalt verbaalsele sisule tuginedes (tabelis 5 näitelause 3).

- 2) Laused, mille emotsiooni identifitseerimist kuulamise järgi ei ole mõjutanud sisu. Sellesse rühma kuuluvad laused, milles testijad on identifitseerinud lugedes ja kuulates erineva emotsiooni (kuulamistesti tulemused erinevad lugemistesti omadest). Selle rühma puhul saab olla kindel, et emotsioon on helis (tabelis 5 näitelause 1 ja 2).

Tabel 5. Emotsioonide klassifitseerimine korpuses kuulamis- ja lugemistestide tulemuste põhjal (testitulemused on protsentides) (vt [P2], [P3]; [P4]; [P5])

Test	Emotsioon					Lausetüüp korpuses
	Rõõm	Viha	Kurbus	Neutraalne	Ei oska öelda	
<i>1. Ehkki Ott minu olemasolust midagi ei teadnud.</i>						
Kuulamistest	87,5	0,0	0,0	12,5		Rõõm, verbaalne sisu ei mõjuta
Lugemistest	4,0	0,0	32,0	32,0	32,0	
<i>2. Ükskõik, mida ma teen, ikka pole ta rahul!</i>						
Kuulamistest	0,0	14,3	80,0	5,7		Kurbus, verbaalne sisu ei mõjuta
Lugemistest	0,0	64,3	35,7	0,0	0,0	
<i>3. Täiesti mõistetamatu!</i>						
Kuulamistest	0,0	100,0	0,0	0,0		Viha, verbaalne sisu mõjutab
Lugemistest	0,0	83,3	0,0	11,1	5,6	

Märkus. Lugemistesti vastusevariantidesse lisati valik „ei oska öelda” juhuks, kui testitaval on raske ühte emotsiooni välja valida ja talle tundub, et emotsioon sõltub pigem sellest, kuidas lauset öeldakse.

Teadaolevalt on Eesti emotsionaalse kõne korpus ainus avalikult kättesaadav kõneemotsioonikorpus, kus materjal on jagatud kahte rühma selle järgi, kas verbaalne sisu võib olla kuulajaid mõjutanud emotsiooni identifitseerimisel või mitte (vt joonis 2, etapp VII).

Eelloetletud teoreetilistele seisukohtadele tuginedes on korpusesse praeguseks kogutud 1234 erinevat lauset, mis kõik on läbinud nii kuulamis- kui ka lugemistesti. Kuulamistestides identifitseeris rohkem kui 50% testijatest sama emotsiooni või neutraalsuse 1234-st korpuse lausest 908-s lauses, 326 lause tuvastusprotsent jäi alla 50% (vt [P4]). Tabelis 6 on esitatud korpuse praegune lausete arv rühmiti.

Tabel 6. Eesti emotsionaalse kõne korpuse lausete arv rühmiti (vt [P4])

Emotsioon	Lausete arv	Verbaalne sisu mõjutab identifitseerimist	Verbaalne sisu ei mõjuta identifitseerimist
rõõm	232	163	69
viha	277	177	100
kurbus	191	88	103
neutraalne	208	87	121
ei suudetud identifitseerida	326		
Kokku	1234	515	393

Kuigi selline kahekordne testimine on töömahukas ja aeganõudev, töötab see kui korpuse valideerija: korpuse kasutaja saab olla kindel, et selles olevate lausete emotsioon on äratuntav (vt [P4]), ning valida korpusest lauseid, kus emotsioon on ainult hääles või emotsiooni äratundmist võib olla mõjutanud lause sisu.

Kuna Eesti emotsionaalse kõne korpuses olevate lausete emotsioonid on määratud etteantud emotsiooniperede piires kuulajad testidega, said töös oluliseks emotsioonide tajuga seotud küsimused. Ptk-s 6.3 näidatakse, kas korpuse testijate valik on ennast õigustanud (vt [P5]; [P6]; [P7]).

6.3. Emotsioonide tajut

Eesti emotsionaalse kõne korpuse teoreetilises aluses vajasid kontrollimist seisukohad, mis puudutasid korpuse testijate valikut ehk lausete emotsiooni määrajaid. Korpust kavandades eeldasin, et inimesed ei tajut emotsioone samal moel, ning püüdsin testijate valikuga tagada emotsioonide võimalikult kõrge identifitseerimise. Seega, korpust luues sai tehtud otsus kasutada korpuse testijatena üle 30-aastaseid empaatilisi täiskasvanuid, kes elasid Eestis ja rääkisid eesti keelt emakeelena. Väitekirja raames soovisin saada kinnitust, kas testijate valikuprintsiibid on end õigustanud, ehk otsisin vastust küsimustele:

- 1) kas lause emotsioon on ainult hääle järgi, ilma kõneleajat nägemata identifitseeritav (vt [P3]);
- 2) kas inimese vanus mõjutab emotsiooni identifitseerimist (vt [P5]);
- 3) kas emotsiooni identifitseerimine on kultuurisõltlik (vt [P5], [P6]);
- 4) kas empaatia mängib rolli emotsioonide identifitseerimisel (vt [P7]).

Iga eeltoodud uurimisprobleemi materjali ja analüüsimeetodit on põhjalikult kirjeldatud peatükkides 6.3.1, 6.3.2, 6.3.3 ja vastavates artiklites: [P3]; [P5]; [P6]; [P7].

Otsides vastust esimesele küsimusele, kas lause emotsioon on äratuntav ka tavalise inimese esitatud mittenäideldud kõnest ilma kõneleajat nägemata, leidis kinnitust oletus, et kuulajad suudavad hästi ära tunda mõõdukalt väljendatud emotsioone mitteprofessionaalse lugeja häälest: üle 50% kuulajatest määras 73,5% korpuse lausetele sama emotsiooni või neutraalsuse (vt [P3]; [P4]; tabel 7).

Tabel 7. Lausete emotsiooni või neutraalsuse määramise statistika kuulamistestide põhjal (vt [P4])

Kuulamistest	Rõõm	Viha	Kurbus	Neutraalne
Lausete arv, mille emotsioon identifitseeriti üle 50%-lise kindlusega	232	277	191	208
Keskmine tuvastusprotsent ja standardhälve	75,4 14,5	73,3 14,6	72,1 14,7	68,3 11,9

Saamaks teada, millist rolli mängib emotsioonide äratundmises inimese vanus, kultuur ja empaatia, koostas inimese veebipõhise kuulamistesti, milles osalesid uurimisküsimusest sõltuvalt eri sihtrühmad. Kuulamistesti valisin Eesti emotsionaalse kõne korpusest 10 rõõmu, 10 viha, 10 kurbust ja 5 neutraalsust väljendavat lauset, mille 1) tuvastusprotsent oli vähemalt 65 (üle 2,5 korra parem juhuslikust valikust) ning 2) emotsiooni määramisel oli olnud otsustavaks hääl (see, kuidas lauset öeldakse), mitte lause verbaalne sisu. Lausete järjestamisel testiks arvestasin põhimõttega, et kaks järjestikust lauset ei moodustaks sisult ühtset loogilist tervikut. Testis kasutatud lausete loetelu ja nende emotsiooni tuvastusprotsent korpuses on esitatud [P6]-s. Testitavatel tuli kuulata kontekstita lauseid ilma teksti nägemata ja otsustada, millise emotsiooniga konkreetse lause puhul tegemist on. Valida sai rõõmu, viha, kurbuse emotsioonipere ja neutraalsuse vahel (vt ptk 6.1). Kuulamistesti läbiviimiseks kasutati emotsioonikorpuse testiliidest. Testitavad said kuulata lauseid nii mitu korda, kui soovisid. (Vt [P5]; [P6]; [P7].)

6.3.1. Vanuse mõju emotsioonide identifitseerimisele

Emotsioonidest arusaamise üheks mõjutajaks on peetud vanust. Ühelt poolt on uurimused näidanud, et vanemaks saades emotsioonidest arusaamine väheneb (Mill, Allik, Realo & Valk, 2009; Paulmann, Pell & Kotz, 2008). Nt Paulmann kolleegidega (2008) täheldas keskealistel (keskmine vanus 42,6) ja noorematel (keskmine vanus 23,4) inimestel märkimisväärsed erinevusi emotsioonide klassifitseerimises häälte põhjal: kuigi klassifitseerimisvigu tegid mõlemad rühmad, oli see noortel väiksem (Paulmann, Pell & Kotz, 2008). Veelgi enam, Mill kolleegidega (2009) leidis, et juba 21–30-aastased on vähem võimekad kurbuse ja viha määramisel, kui nendest ealiselt nooremad, ning vanuses 31–40 on kurbuse ja viha emotsioonide taju langenud oluliselt. Neutraalsete väljendite tajumisel silmapaistvaid erinevusi vanusegruppide vahel ei ole (Mill, Allik, Realo & Valk, 2009). Teiselt poolt on aga leitud, et emotsioonide väljendamine ja tõlgendamine on kultuurikohane ning aja jooksul õpitav (vt Dewaele, 2010; Nazir jt, 2009; [P5]; [P6]).

Eesti emotsionaalse kõne korpuse teoreetilises aluses otsustasin lähtuda nendest uurimustest, mis on näidanud emotsioonide kultuurispetsiifilisust, ning noori, alla 30-aastaseid testijaid emotsioonide määrajana mitte kasutada, eeldades, et vanemad inimesed tunnevad emotsioone noortest paremini ära,

kuna nad on vastavas kultuuris pikemalt elanud ning kultuurikohase emotsiooniväljenduse ja -dekodeerimise omandanud. [P5]-s keskendusin küsimusele, kui õigustatud on otsus valida korpuse lausete emotsiooni määrajateks üle 30-aastased inimesed ja jätta kõrvale nooremad. Selleks uurisin nii vanuse kui ka kultuuri mõju emotsioonide tajumisele. Katses osales kaks rühma – eesti keelt emakeelena rääkivad noored täiskasvanud eestlased (20–28-aastased) ja eesti keelt mitteoskavad üle 30-aastased lätlased, kelle tulemusi võrdlesin korpuse üle 30-aastaste (30–62-aastaste) eesti keelt emakeelena rääkivate testijate tulemustega. Lätlased kaasasin selleks, et välja selgitada, kui oluline on emotsioonide hääle järgi identifitseerimisel kultuuris elamise kogemus. Valiku lätlaste kasuks langetasin põhjusel, et Eesti ja Läti on naaberriigid ning nende kultuurid on väärtustelt sarnased (vrd Huettigen, 2008), mistõttu võiks eeldada, et nad väljendavad emotsioone sarnaselt ja on suutelised üksteise emotsioone ka hääle järgi ära tundma.

Testitavatel tuli kuulata 35 eestikeelset lauset ja määrata kuuldu põhjal lause emotsioon (vt testikirjeldust ptk 6.3). Tulemused näitasid emotsioonide määramisel statistiliselt olulisi rühmadevahelisi erinevusi. Täiskasvanud ja noored täiskasvanud ei erinenud oluliselt vaid kurbuse ja neutraalsuse määramises. Viha määramises ei erinenud oluliselt noored täiskasvanud ja lätlased. Täiskasvanutel ja noortel täiskasvanutel ületas sihtemotsiooni tuvastusprotsent 50%, kuid noortel jäi see vanemate omast madalamaks. Noored täiskasvanud määrasid märksa enam lauseid neutraalseteks.

Lätlaste rühmas ületas 50% künnise vaid kurbuse identifitseerimine, ent selleski erinesid lätlased oluliselt eestlastest, kuna ka lätlaste jaoks kõlasid eesti emotsioonilauseid tihti neutraalsetena. Neutraalseks identifitseeritud lausete suur osakaal lätlaste ja noorte täiskasvanute rühmal viitab sellele, et emotsioonide hääle järgi identifitseerimiseks on oluline pikaajaline kultuuris elamise kogemus, mida on vaja kultuurispetsiifilise emotsiooniväljenduse omandamiseks (vrd Pell, Paulmann jt, 2009). Lätlaste võime ära tunda eestikeelseid emotsioone oli suhteliselt madal, mis näib kinnitavat, et emotsioonide äratundmine häälest on kultuurisõltlik. (Vt [P5].)

Seega saadud uurimistulemused toetavad otsust valida Eesti emotsionaalse kõne korpuse lausete emotsiooni määrajateks üle 30-aastased eesti keelt emakeelena rääkivad täiskasvanud eestlased, kuna nad suudavad paremini kultuurikohaselt dekodeerida sõnumi emotsiooni.

6.3.2. Kultuuri mõju emotsioonide identifitseerimisele

Kultuuridevahelised emotsioonide äratundmise uurimused kinnitavad üldist seisukohta, et kultuuril on osa selles, kuidas me emotsioone väljendame, neist aru saame ja millised emotsionaalsed signaalid on sotsiaalselt aktsepteeritavad (Soto & Levenson, 2009). Viimase aja kultuuridevaheliste uuringute tulemused, mis keskenduvad emotsioonide äratundmisele ainult häälest, näitavad nii kultuuri mõju emotsioonide tuvastamisele kui ka universaalseid tendentse. Nt on uurimused näidanud, et negatiivsed emotsioonid nagu viha ja kurbus

tuntakse kultuuridevaheliselt paremini ära (Bryant & Barrett, 2008; Pell, Paulmann jt, 2009; Scherer, Clark-Polner & Mortillaro, 2011; Thompson & Balkwill, 2006). [P5]-s leidis kinnitust, et ka sarnase ajaloo ja väärtustega kultuuridest pärit ning geograafiliselt lähedal olevad inimesed ei pruugi teises keeles väljendatud emotsioone ära tunda.

Samas, kui suur osa kultuuridevahelisi emotsiooniuurimusi on käsitlenud erinevaid etnilisi gruppe erinevatelt maadelt, soovisin mina [P6]-s teada saada, kuidas inimesed, kes küll räägivad eri keeli, kuid elavad ühel maal, emotsioonidest aru saavad. Täpsemalt, kuidas tuntakse emotsiooniväljendusi ära olukorras, kui emotsioone ei edastata mitte esimeses (L1), vaid teises keeles (L2). Teise keele ja emotsioonide äratundmise vaheliste seoste analüüsimiseks valisin sihtgrupiks Eestis elavad venelased, kelle teine keel (L2) on eesti keel. Soovisin teada, kas Eesti venelased tunnevad ära eesti keeles väljendatud emotsioone ja kas nad tunnevad need ära nagu eestlased või sarnaneb nende emotsioonitaju Venemaal elavate venelaste omale. Selleks testisin kolme rühma täiskasvanuid: 36 eestlast (keskmine vanus $M = 41,3$, $SD = 9,2$), 16 Eestis elavat venelast (keskmine vanus $M = 35,8$, $SD = 8,1$), kelle esimene keel (L1) on vene keel ja kes oskavad eesti keelt (L2) kõrgtasemel (C1). Kolmanda rühma moodustasid 16 Venemaal elavat venelast (keskmine vanus $M = 33,1$, $SD = 5,0$). Testitavatel tuli kuulata 35 eestikeelset lauset ja määrata kuuldu põhjal, ilma teksti nägemata, lause emotsioon (vt testikirjeldust ptk 6.3).

Eestis elavate venelaste ja Venemaal elavate venelaste kuulamistulemusi võrdlesin omavahel ja eesti keelt emakeelena rääkivate eestlaste tulemustega. Kolme uurimisrühma tulemuste võrdlus näitas kultuuridevahelise suhtluse mõju emotsioonide äratundmisele. Eesti venelased olid võimelised ära tundma kõik eesti emotsioonid üle 50% kindlusega, seejuures ei erinenud nad eestlastest statistiliselt oluliselt viha ja kurbuse tuvastuses, kuid erinesid rõõmu ja neutraalsuse tuvastuses, jäädes nende tuvastuses eestlastest täpsuselt alla ja ajades rõõmu segi neutraalsusega. Eestlased ja Venemaa venelased erinesid oluliselt kõigi kolme emotsiooni ja neutraalsuse tuvastamises, samuti erinesid Eesti venelased ja Venemaa venelased omavahel. Seega, Eesti venelased sarnanesid eesti emotsioonide tuvastuselt pigem eestlastele kui Venemaa venelastele. Venemaa venelased tundsid üle 50% juhtudest ära üksnes kurbuse. Nad pidasid eesti rõõmu neutraalsuseks ega teinud vahet eesti vihal ja neutraalsusel. Saadud tulemused vihjavad sellele, et emotsioonidest arusaamine sõltub kultuurist ning suhtlemisest, st vastavale kultuurile omased sotsiaalsed normid, sh emotsionaalsest sõnumist arusaamine saadakse selgeks suhtlemise käigus. (Vt [P6].)

6.3.3. Empaatia mõju emotsioonide identifitseerimisele

Otsisin ka vastust küsimusele, kas empaatia mängib rolli emotsioonide identifitseerimisel, kui emotsiooni üle tuleb otsustada ainult hääle järgi, ilma kõnelejat nägemata. Korpuse lausete emotsiooni määrajaks olin valinud empaatilised inimesed eeldades, et nemad tunnevad emotsioone paremini ära kui mitte-empaatilised. Teemat käsitleb artikkel [P7], milles püstitatakse uurimisküsimus,

kas empaatiatase (madal, keskmine, kõrge) mõjutab emotsioonide tuvastamist häälest. Ehk kui õigustatud on otsus arvestada korpuse lausete emotsiooni määrajatena ainult empaatilisi kuulajaid ja jätta arvestamata mitteempaatiliste kuulajate määramistulemused, kuna mitteempaatilised inimesed ei pruugi emotsioone häälest ära tunda.

Senised uurimused empaatiavõime ja emotsioonide häälest äratundmise seoste kohta teenivad peamiselt kliinilist eesmärki: on uuritud, kuidas tunnevad emotsioone ära inimesed, kellel diagnoosiga kaasneb madal empaatia, nt autism, Aspergeri sündroom, psühhopaatia. Kliinilistel eesmärkidel tehtud tööd on valdavalt näidanud, et madala empaatiaga võivad kaasneda emotsioonide ja mentaalsete seisundite häälest äratundmise raskused (Golan, Baron-Cohen & Hill, 2006; Colan, Baron-Cohen, Hill & Rutherford, 2007; Rutherford, Baron-Cohen & Wheelwright, 2002).

[P7]-s kajastatud uurimus erineb nendest uurimustest kahes aspektis: 1) uuritavatel ei ole diagnoositud empaatiat vähendavaid haigusi; 2) tuvastamiseks antud emotsioonid pole näideldud, vaid esilekutsutud, st tegemist pole täismahuliste stereotüüpselt väljendatud põhiemotsioonidega, milliseid igapäevaelus kohtab harva, vaid mõõdukalt väljendatud emotsioonidega, mis kuuluvad rõõmu, viha ja kurbuse emotsiooniperesse. Kuulamistestis osales 67 täiskasvanut (29 meest ja 38 naist vanuses 30–60, $M = 42,0$, $SD = 10,7$). Kõikide osalejate empaatiavõimet mõõdeti eesti keelde tõlgitud Baron-Cohen ja Wheelwright'i (2004) enesehinnangu küsimustikuga Empathy Quotient (EQ).

Kuulamistestis osalejad jagasime esmalt kolme rühma empaatiatesis saadud punktide põhjal: madal empaatia alla 32 punkti, keskmine empaatia 32–51 punkti ja kõrge empaatia üle 51 punkti (vrd Bate, Parris, Haslam & Kay, 2010). Kogu valimist olid madala empaatiaga ainult mehed ja kõrge empaatiaga ainult naised, mis pole üllatav, kuna ka teised uurimused on näidanud, et naiste empaatia on kõrgem (Baron-Cohen & Wheelwright, 2004; Lawrence jt, 2004). Keskmise empaatiaskoori saanud jagasime meeste ja naiste alarühmaks:

- I. madal empaatia: $EQ\ 18-31$, $M = 26,2$, $SD = 3,2$
14 meest vanuses 32–44, $M = 36,7$, $SD = 4,5$
- II. keskmine empaatia: $EQ\ 32-51$, $M = 42,5$, $SD = 5,7$
 - a) 15 meest vanuses 31–56, $M = 39,0$, $SD = 7,6$; $EQ\ 32-50$, $M = 41,5$, $SD = 5,8$
 - b) 22 naist vanuses 31–60, $M = 38,5$, $SD = 6,4$; $EQ\ 34-51$, $M = 43,1$, $SD = 5,1$
- III. kõrge empaatia: $EQ\ 52-60$, $M = 54,9$, $SD = 2,6$
16 naist vanuses 30–58, $M = 44,1$, $SD = 10,5$

Kuulamistestis tuli testitavatel kuulata 35 üksiklauset ja otsustada kuuldu põhjal, mis emotsioon lauses kõlab (vt testikirjeldust ptk 6.3). Et uuritavad rühmad erinesid sooliselt kuuluvuselt – madala empaatiaga rühm koosnes ainult meestest ja kõrge empaatiaga rühm ainult naistest –, siis esmalt tuli kindlaks teha soo mõju emotsioonide tuvastamisele. Selleks vaatlesime keskmiselt empaatilise rühma mehi ja naisi, so ainsat rühma, kus oli nii mehi kui ka naisi. Tulemused näitasid soo mõju kahe emotsiooni – rõõmu ja viha – tuvastamisele.

Seetõttu võrdlesime eri empaatiarühmi sootunnuste järgi eraldi: madala ja keskmise empaatiaga meeste rühmi omavahel ning keskmise ja kõrge empaatiaga naiste rühmi omavahel. Tulemused näitasid, et madala ja keskmise empaatiaga mehed ei erinenud statistiliselt oluliselt emotsioonide tuvastamises, samuti ei erinenud emotsioonide tuvastamises statistiliselt oluliselt keskmise ja kõrge empaatiaga naised. Seega mõõdukalt väljendatud emotsioonidega testilausetes tunti rõõm, viha, kurbus ja neutraalsus ära üle 50% juhtudest olenemata soost ja empaatiatasemest.

Üldistades saadud tulemusi, võib öelda, et empaatial ei ole emotsioonide tuvastamisele häälest nii suur roll, kui kliinilistel eesmärkidel tehtud uurimuste põhjal võis arvata. Kõnetehnoloogia valdkonna jaoks tehtavatest emotsiooni-uurimustest ei ole ilmselt vaja testijate seast välja arvata mitteempaatilisi inimesi põhjusel, et nad ei pruugi emotsioone ära tunda, kui nende madal empaatia pole tingitud psüühilistest või arenguhäiretest. Küll aga oleks otstarbekas kaasata emotsiooniuurimustesse nii mehi kui naisi nende emotsioonituvastuse erinevuse tõttu. (Vt [P7].

6.4. Tulemused ja diskussioon

6.4.1. Eesti emotsionaalse kõne korpus teiste kõneemotsioonikorpuste seas

Nagu eelnevalt kirjeldatud, oli väitekirja raames tehtud uurimuse eesmärk luua Eesti emotsionaalse kõne korpus teoreetiline alus ja kontrollida loodud korpus materjali põhjal teoreetiliste seisukohtade õigsust. Minu uurimus näitas, kui oluline on korpus enne selle loomist hoolikalt planeerida ja tulemust analüüsida. Saadud teadmisi saavad rakendada nii emotsiooniuurijad kui ka kõne-korpuste arendajad.

Korpus on niisugusena, nagu ta teoreetiliselt kavandati, olemas (vt [P1]; [P2]; [P3]; [P4]; <http://peeter.eki.ee:5000>). Eesti korpus teeb teiste kõne-emotsioonikorpuste seas ainulaadseks asjaolu, et lausete emotsioon on märgendatud selle järgi, kas emotsiooni kannab lause heli või mõjutab emotsiooni äratundmist häälest lause verbaalne sisu. Selline jaotus teeb võimalikuks emotsioonide uurimise nii kõnes kui kirjas.

Milline on Eesti emotsionaalse kõne korpus teiste samalaadsete seas, saab ettekujutuse, võrreldes seda teiste kõneemotsioonikorpustega.

Kõne laadi järgi jagatakse korpusi kolmeks suuremaks rühmaks (vt ülevaateid Pajupuu, 2012; Schuller, Batliner, Steidl & Seppi, 2011). Suurima rühma moodustavad näideldud emotsioonidega korpused. Nendes öeldakse korpus materjali kogudes kõnelejatele, millist (põhi)emotsiooni väljendada. Sageli lastakse korrata üht ja sama lauset erineva emotsiooniga (vt nt Berliini emotsionaalse kõne andmebaas (Burkhardt jt, 2005) ja Taani emotsionaalse kõne andmebaas (Engberg, Hansen, Andersen & Dalsgaard, 1997)).

Teise rühma moodustavad esilekutsutud emotsioonidega korpused. Selleks kavandatakse emotsioone tekitav olukord, nt inimese ja masina suhtlus, nagu on

kogutud saksa AIBO emotsioonikorpus (Batliner jt, 2004). Emotsioone võib esile kutsuda ka kirjaliku tekstiga, nagu on loodud jaapani emotsionaalse kõne sünteesisüsteemi CHART korpus (Iida, Campbell, Higuchi & Yasumura, 2003). Esilekutsutult saadakse korpusesse pigem emotsioonilaadseid seisundeid (nt pahameel, ükskõiksus, ärritus) kui täismahulisi emotsioone.

Kolmanda rühma moodustavad reaalses situatsioonides kogutud emotsioonidega korpused, nagu saksakeelne Vera am Mittag (VaM) (Grimm, Kroschel & Narayana, 2008). Selles korpuses on lisaks häälele märgendatud ka näoilmed (nn multimodaalne korpus).

J. Pitterman, Pitterman & Minker (2010: 237–250) koostatud loendis on 103 kõneemotsioonikorpust ja -andmebaasi 16 keele jaoks: inglise (33), saksa (16), hiina (10), jaapani (6), prantsuse (5), hollandi (4), hispaania (3), rootsi (3), baski (2), korea (2), burmese (1), taani (1), soome (1), kreeka (1), heebrea (1), vene (1), mitmekeelne (13). Neist väike osa on audiovisuaalsed (15). Kõigist korpustest ja andmebaasidest 64%-l on kõneüksuseks lause või lausung, 57%-l on tegemist näideldud või simuleeritud emotsioonidega. Valdavalt on kõneemotsioonikorpused ja -andmebaasid väikesed andmekogud, mis on loodud mingi üksikuurimuse jaoks ja mille kirjelduse leiab vaid seda konkreetset uurimistööd puudutavas artiklis. Mitmed neist on projektipõhised, st pärast projekti lõppu arenduse lõpetanud, vähestele korpustele ja andmebaasidele on vaba juurdepääs. Vt tabel 8 (Raport, 2011).

Eesti emotsionaalse kõne korpus on üks väheseid esilekutsutud mõeldukalt väljendunud emotsioone sisaldavaid kõnekorpuseid, mis on dokumenteeritud, avalikult ja tasuta kättesaadav. Korpuses loobuti näideldud emotsioonidest nende võimaliku stereotüüpsuse ja ülepaisutatuse tõttu. Korpuse jaoks salvestati n-õ tavalise inimese ettelõetud tekstid, kellele ei öeldud, millise emotsiooniga ta tekste lugema peaks. Eeltoodutest sarnaneb Eesti emotsionaalse kõne korpus materjali kogumise poolest jaapani kõnesünteesi jaoks loodud emotsioonikorpusega – eeldatakse, et lugemismaterjali emotsioon tekitab ka ettelugejas vastava emotsiooni, mis kajastub tema hääles (vt Iida, Campbell, Higuchi & Yasumura, 2003). Meetod nii emotsioone korpusesse koguda osutus efektiivseks, kuna kuulajad suutsid lauses emotsioone identifitseerida.

Mahult on Eesti emotsionaalse kõne korpus üle keskmise, kuid kõnelejaid on vaid üks. Keskendumise ühele kõneleajale tingis asjaolu, et esimese ülesandena peeti vajalikuks luua piisava suurusega korpus üksuste valikul põhineva emotsionaalse kõnesünteesi jaoks. Samas, korpuse teoreetilist alust luues sai silmas peetud korpuse võimalikku kasutamist paljudel eesmärkidel, nii uurimistööks kui ka mitmesugusteks kõnetehnoloogilisteks rakendusteks. Korpuse maht on selline, et seda on võimalik kasutada kõnetehnoloogias emotsionaalse sünteesikõne statistiliste prosoodiamudelite väljatöötamiseks. Ka on korpus kavandatud selliselt, et seda on võimalik laiendada uute kõneleajatega, lisada videomaterjali, pilte, spontaanset kõnet; jagada emotsioone enamatesse kateooriatesse ja määratleda neid dimensioonidel. Riiklikku programmi „Eesti keeletehnoloogia (2011–2017)” raames oleme alustanud projekti „Kõne ja teksti emotsionaalsuse statistilised mudelid”, mis võimaldab korpust edasi arendada.

Tabel 8. Eesti emotsionaalse kõne korpus võrdluses teiste teadaolevate kõneemotsioonikorpusete ja -andmebaasidega (Raport, 2011)

Muud vs. Eesti emotsionaalse kõne korpus	Emotsioonide arv	Kõnelejate arv	Lausete/ lausungite arv	Näide/diud või simuleeritud emotsioonid	Sisaldab ka neutraalseid lausungeid	Avalikult kättesaadav
min	1	1	15			10%
Q1	3	2	263			(võib vajada registreerimist või ollatasuline)
mediaan	4	8	732	57%	70%	
Q2	6	36	2333			
max	15	784	20013			
Eesti emotsionaalse kõne korpus	3	1	1234	Ei (sisaldab loetud esilekutsutud emotsioone)	Jah	Jah (avalik, tasuta)

Nt praeguseks on kõik 1234 korpuse lauset läbinud lisaks määrangule emotsioonikategooriateks rõõm, viha, kurbus ja neutraalsus ka kaks dimensioonitesti: valentsitestis on laused määratud negatiivseteks, positiivseteks ja neutraalseteks (nii kuulates kui ka lugedes) ning aktiivsustestis aktiivseteks ja passiivseteks (ainult kuulates). Korpuse testijad suutsid kategooriatesse rõõm, viha, kurbus ja neutraalne määrata 1234-st lausest 908 lauset, kusjuures 326 jäi määranguta. Edaspidi on võimalik anda määranguta jäänud lauseid kuulajatele testimiseks ilma etteantud emotsioonikategooria valikvariantideta, s.t anda kuulajale endale võimalus vabalt valida, millise emotsioonikategooria või emotsionaalse seisundiga on tegemist.

Valentsidimensioonil klassifitseerusid 1234-st lausest 1165 lauset: 269 positiivseks, 675 negatiivseks ja 211 neutraalseks. Emotsioonide klassifitseerimine kategooriateks ja dimensioonideks avab võimaluse uurida nende kombinatsioone, nt kurbi negatiivseid, kurbi positiivseid emotsioone jne.

Korpus sisaldab praegu ühe naishääle lauseid, mis on klassifitseeritud vihaks, rõõmuks, kurbuseks ja neutraalsuseks. Laused on kuvatud tekstina ja neil klikates on võimalik lauseid kuulata. Iga lause juures on näidatud emotsionaalsuse (või neutraalsuse) tajuprotsent. Lausete heli (wav-formaat: 44.1 KHz, 16Bit, mono) saab alla laadida ja salvestada. Segmneteeritud ja märgendatud on kolm tasandit: laused, sõnad, häälikud. Tavakasutajal on võimalik teha päringuid nii lausete kui ka testide kohta. Alla saab laadida ka kogu korpuse materjali.

6.4.2. Vanuse, kultuuri ja empaatia mõju emotsioonide identifitseerimisele

Väitekirja raames kontrollisin Eesti emotsionaalse kõne korpuse teoreetilise aluse kolme seisukoha kehtivust. Saadud uurimistulemused näitasid, et:

- 1) üle 30-aastased täiskasvanud tunnevad emotsioone hääle järgi paremini ära kui noored täiskasvanud (20–30-aastased);
- 2) emotsioonide äratundmine hääle järgi on kultuurisõltlik;
- 3) empaatiatase ei mõjuta emotsioonide tuvastamist häälest; vt [P3]; [P5]; [P6]; [P7].

Eesti emotsionaalse kõne korpuse laused on emotsioonimärgendi saanud üle 30-aastaste empaatiliste testijate määrangute põhjal. Emotsioon on loetud identifitseerituks, kui enam kui 50% üle 30-aastastest testijatest on konkreetse emotsiooni kasuks otsustanud.

[P5] tulemused näitasid, et uurimuses püstitatud küsimusele, kas on õige kasutada korpuse lausete emotsiooni määrajatena üle 30-aastaseid testijaid lootuses, et nad tunnevad emotsioone noortest paremini ära, ei saa anda ühest vastust. Ühest küljest näitavad tulemused, et üle 30-aastased ja nooremad täiskasvanud erinevad emotsioonide identifitseerimisel üksteisest statistiliselt oluliselt. Teisalt ei tekita kummalegi rühmale probleeme emotsioonide identifitseerimine hääle järgi: konsensus selles, mis emotsioon ühes või teises lauses

kõlab, on mõlemal rühmal suur (üle 50%, vt ptk 6.3.1). Küll aga identifitseerivad nooremad täiskasvanud rohkem lauseid neutraalseks. Kuna korpuse iga lause emotsiooni otsustavad kuulajad, siis ei ole juhtudel, kui eestlaste erinevad rühmad (antud juhul noorematest või vanematest koosnevad) määravad emotsiooni kuulamistestis erinevalt, võimalik öelda, kummal rühmal on õigus. Üle 30-aastaste valikut toetavad kaudselt kaks seika: 1) vanemate testijate võime hääle järgi emotsioone määrata pole vanuse kasvades kahanenud: nad määravad kurbust noortega sarnaselt (vrd Mill, Allik, Realo & Valk, 2009); 2) vanemad testijad määravad rohkem lauseid emotsionaalseks kui nooremad, kelle valikusse kuulub sagedamini määratlus neutraalne, mis näitab vanemate inimeste suuremat oskust dekodeerida kultuuriomast emotsiooniväljendust. Kultuuri olulisust emotsioonide identifitseerimisel kinnitavad ka katsed lätlastega, kelle võime eesti emotsioone hääle järgi ära tunda on suhteliselt madal.

[P6] käsitles kultuuri mõju kõnes mõõdukalt väljendunud emotsioonide äratundmisele tingimustes, kus emotsiooni edastatakse mitte esimeses (L1), vaid teises keeles (L2). Otsisin vastust küsimusele, kas Eestis elavad venelased, kelle teine keel on eesti keel, tuvastavad eesti keeles väljendatud emotsioone ja tunnevad emotsioone ära nagu eestlased või nagu Venemaa venelased. Eestlaste, Eesti venelaste ja Venemaa venelaste võrdlus näitas, et Eesti venelased olid võimelised ära tundma kõik eesti emotsioonid üle 50% kindlusega, Venemaa venelased tundsid üle 50%-lise kindlusega ära üksnes kurbuse.

Uurimused on näidanud, et negatiivsed emotsioonid nagu viha ja kurbus tuntakse kultuuridevahelisel paremini ära (vt nt Pell, Paulmann jt, 2009; Scherer, Clark-Polner & Mortillaro, 2011; Thompson & Balkwill, 2006). See, et Venemaa venelased tundsid ära eesti kurbuse, on nende uurimustega kooskõlas. Samas ei olnud Venemaa venelased võimelised ära tundma eesti viha, ajades selle segi neutraalsusega. Samuti ajasid nad neutraalsusega segi rõõmu. Põhjuseks võib olla eestlaste ja venelaste erinev viha, rõõmu ja neutraalsuse väljendamise akustiline mudel (vt [P6]). Et Eesti venelased ei erinenud eestlastest eesti viha äratundmisel ja ajasid rõõmu vähem segi neutraalsusega kui Venemaa venelased, näitab, et nad on samas kultuurikeskkonnas suheldes õppinud ära eesti tava emotsioone väljendada. Seega tulemus kinnitab, et vastavas kultuuris elamine ja võime suhelda teises keeles (L2) on emotsioonidest arusaamiseks olulised.

[P7]-s uurisin empaatia mõju kõnes mõõdukalt väljendunud emotsioonide äratundmisele tingimustes, kus otsustada tuleb ainult hääle põhjal, ilma kõnelejat nägemata. Kui varasemad uurimused on enamasti näidanud madala empaatiaga inimeste raskusi emotsioone kõnest ära tunda (Golan, Baron-Cohen & Hill, 2006; Golan, Baron-Cohen, Hill & Rutherford, 2007; Rutherford, Baron-Cohen & Wheelwright, 2002), siis käesolev uurimus näitas, et madala, keskmise ning kõrge empaatiaga inimesed ei erine oluliselt emotsioonide tuvastamises, kui neil on piisavalt aega emotsiooni üle otsustada. Nimetatud eelnevad uurimused olid tehtud inimestega, kel oli diagnoositud mõni empaatiat vähendav haigus. Käesolev uurimus on teadaolevalt esimene, mis käsitleb terve,

kuid erineva empaatiatasemega inimeste võimet häälest ära tunda mõõdukalt väljendatud emotsioone.

Uurimus [P7] andis kõrvaltulemuse, et mehed ja naised tuvastavad viha ja rõõmu erinevalt, millest järeldub, et emotsiooniuurimustesse peaks testijatena kaasama nii mehi kui ka naisi. Eesti emotsionaalse kõne korpuse registreeritud 190-st testijast on 58% naised, nii on see nõue täidetud.

Tajukatsete tulemuste põhjal võib kokkuvõtvalt öelda, et korpuse teoreetilise aluse seisukoht võtta testijateks üle 30-aastased Eestis elanud eesti emakeelega empaatilised eestlased on veidi piirav: kui pikaajaline kultuuris elamine tuleb igati kasuks emotsioonide identifitseerimisele, siis empaatiavõime ei oma nii suurt tähtsust, et seda peaks mõõtma ning tulemuste järgi testijaid valima. Teoreetiline alus täienes aga seisukohaga, et korpuse testijateks tuleks võtta mõlemast soost inimesi.

Emotsioonikorpuse materjal on avanud täiendavaid uurimissuundi. Edaspidi võiks kontrollida korpuse teoreetilise aluse muidki seisukohti. Nt oleme alustanud uurimust, kas kõneemotsioonide identifitseerimiseks on vaja lausest verbaalse sisu mõju kõrvaldada, ehk kas emotsioonide akustilised parameetrid erinevad lauserühmades, mille verbaalne sisu 1) mõjutab ja 2) ei mõjuta kuulajat emotsiooni identifitseerimisel (vt ka [P4]).

Kuna Eesti emotsionaalse kõne korpus on kergesti laiendatav, võiks korpusesse lisada ka spontaanset kõnet, mis võimaldaks võrrelda, kas ja kuidas erinevad loetava tekstiga esilekutsutud ja spontaansed emotsioonid. See uurimus tooks lisateadmist Scherer'i (2011) näideldud ja spontaanse kõne võrdleva uurimuse kõrvale.

Veel, korpusesse EPIP-NEO testiga (vt Mõttus, Pullmann & Allik, 2006) kogutud testijate isiksuseomaduste andmed võimaldavad edaspidi uurida isiksuseomaduste mõju emotsioonide äratundmisele, nt kuidas tunnevad emotsioone ära kõrge avatusega, ekstravertsed või neurootilised inimesed.

Ja viimaks, korpust arendatakse edasi vastavalt uutele uurimissuundadele. Kuna korpus on avalikult kasutatav ja tasuta ligipääsetav, on korpuse materjali võimalik kasutada väga mitmesuguste ja eri valdkondade uurimistülesannete lahendamiseks.

REFERENCES

- Abe, M., Sagisaka, Y., Umeda, T., & Kuwabara, H. (1990). *ATR technical report TR-I-0116 speech database user's manual*. Kyoto: ATR Interpreting Telephony Research Laboratories.
- Abelin, Å., & Allwood, J. (2000). Cross linguistic interpretation of emotional prosody. In *Proceedings of the ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion: Speech Emotion-2000* (pp. 110–113). Newcastle: International Speech Communication Association.
- Allik, J., Realo, A., & Konstabel, K. (Eds.). (2003). *Isiksusepsühholoogia*. [Psychology of Personality]. Tartu: Tartu University Press.
- Allison, C., Baron-Cohen, S., Wheelwright, S., Stone, M. H., & Muncer, S. J. (2011). Psychometric analysis of the empathy quotient (EQ). *Personality and Individual Differences*, *51*(7), 829–835. doi:10.1016/j.paid.2011.07.005
- Al-Moubayed, S., Zliekha, M. A., Ghneim, N., Al-Hakawati Al-Dakkak, O. (2005). Emotion inclusion in an Arabic text-to-speech. In *Proceedings of the 13th European Signal Processing Conference (EUSIPCO 2005): September 4–8, Antalya, Turkey* (pp. 2070–2073). Red Hook, NY: Curran Associates, Inc.
- Altrov, R. (2007). *Emotsionaalse kõne korpuse loomine eesti keele tekst-kõne sünteesi jaoks. Tekstimaterjali evalvatsioon viha näitel*. [Development of an emotional speech corpus for Estonian text-to speech synthesis: Evaluation of text material on the example of anger]. (Unpublished master's thesis). University of Tartu, Institute of Estonian and General Linguistics, Tartu.
- Altrov, R. (2008). Eesti emotsionaalse kõne korpus: teoreetilised toetuspunktid. [Theoretical underpinning of the corpus of Estonian emotional speech]. *Keel ja Kirjandus*, *4*, 261–272.
- Altrov, R. (2013). Aspects of cultural communication in recognizing emotions. *Trames: Journal of the Humanities and Social Sciences*, *17*(2), 159–174. doi:10.3176/tr.2013.2.04
- Altrov, R., & Pajupuu, H. (2008). The Estonian Emotional Speech Corpus: Release 1. In F. Čermak, R. Marcinkevičienė, E. Rimkutė, & J. Zabarskaitė (Eds.), *Proceedings of the 3rd Baltic Conference on Human Language Technologies* (pp. 9–15). Vilnius: Vytauto Didžiojo Universitetas, Lietuvių Kalbos Institutas.
- Altrov, R., & Pajupuu, H. (2010). Estonian Emotional Speech Corpus: Culture and age in selecting corpus testers. In I. Skadina, & A. Vasiljevs (Eds.), *Frontiers in Artificial Intelligence and Applications: Vol. 219. Human Language Technologies – The Baltic Perspective. Proceedings of the Fourth International Conference Baltic HLT 2010* (pp. 25–32). Amsterdam: IOS 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: European Language Resources Association.
- 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., & Pajupuu, J. (2013). The role of empathy in the recognition of vocal emotions. In *INTERSPEECH-2013* (pp. 1341–1344). Lyon: International Speech Communication Association.
- Anolli, L., Wang, L., Mantovani, F., & De Toni, A. (2008). The voice of emotion in Chinese and Italian young adults. *Journal of Cross-Cultural Psychology*, *39*(5), 565–598. doi:10.1177/0022022108321178
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Barkhuysen, P., Krahmer, E., & Swerts, M. (2010). Crossmodal and incremental perception of audiovisual cues to emotional speech. *Language and Speech*, *53*(1), 3–30. doi:10.1177/0023830909348993
- Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient. An investigation of adults with Asperger syndrome or high functioning autism and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. doi:10.1023/B:JADD.0000022607.19833.00
- Barra-Chicote, R., Yamagishi, J., King, S., Montero, J. M., & Macias-Guarasa, J. (2010). Analysis of statistical parametric and unit selection speech synthesis systems applied to emotional speech. *Speech Communication*, *52*(5), 394–404. doi:10.1016/j.specom.2009.12.007
- Bate, S., Parris, B., Haslam, C., & Kay, J. (2010). Socio-emotional functioning and face recognition ability in the normal population. *Personality and Individual Differences*, *48*(2), 239–242. doi:10.1016/j.paid.2009.10.005
- Batliner, A., Hacker, C., Steidl, S., Nöth, E., D’Arcy, S., Russell, M., & Wong, M. (2004). “You stupid tin box” – children interacting with the AIBO robot: A cross-linguistic emotional speech corpus. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)* (pp. 171–174). Lisbon: European Language Resources Association.
- Batliner, A., Steidl, S., Hacker, C., & Nöth, E. (2008). Private emotions versus social interaction: a data-driven approach towards analysing emotion in speech. *User Modeling and User-Adapted Interaction*, *18*(1/2), 175–206. doi:10.1007/s11257-007-9039-4
- Boersma, P., & Weenink, D. (2007). Praat: Doing phonetics by computer (Version 4.6.31) [Computer program]. Retrieved from <http://www.praat.org/>
- Busso, C., & Narayanan, S. (2009). Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection. *IEEE Transactions on Audio, Speech, and Language Processing*, *17*(4), 582–596. doi:10.1109/TASL.2008.2009578
- Bryant, G. A., & Barrett, C. H. (2008). Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture*, *8*(1/2), 135–148. doi:10.1163/156770908X289242
- Buck, R. (1999). Biological affects. A typology. *Psychological Review*, *106*(2), 301–336. doi:10.1037/0033-295X.106.2.301
- Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., & Weiss, B. (2005). A database of German emotional speech. In *INTERSPEECH-2005* (pp. 1517–1520). Lisbon: International Speech Communication Association.
- Calder, A. J., Keane, J., Manly, T., Sprengelmeyer, R., Scott, S., Nimmo-Smith I., & Young, A. W. (2003). Facial expression recognition across the adult life span. *Neuropsychologia*, *41*(2), 195–202. doi:10.1016/S0028-3932(02)00149-5
- Campbell, N. (2000). Databases of emotional speech. In *Proceedings of the ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion: SpeechEmotion-2000* (pp. 34–38). Newcastle: International Speech Communication Association.

- Campbell, N. (2005). Getting to the Heart of the Matter: Speech as the Expression of Affect; Rather than Just Text or Language. *Language Resources and Evaluation*, 39(1), 109–118. doi:10.1007/s10579-005-2699-y
- Chakrabarti, B., Bullmore, E., & Baron-Cohen, S. (2006). Empathizing with basic emotions: Common and discrete neural substrates. *Social Neuroscience*, 1(3/4), 364–384. doi:10.1080/17470910601041317
- Cheang, H. S., & Pell, M. D. (2008). The sound of sarcasm. *Speech Communication*, 50(5), 366–381. doi:10.1016/j.specom.2007.11.003
- Costa, P. T. Jr., & McCrae, R. R. (1995). Domains and facets. Hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment*, 64(1), 21–50. doi:10.1207/s15327752jpa6401_2
- Cowie, R., & Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech Communication*, 40(1/2), 5–32. doi:10.1016/S0167-6393(02)00071-7
- 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
- Dawel, A., O'Kearney, R., McKone, E., & Palermo, R. (2012). Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience and Biobehavioral Reviews*, 36(10), 2288–2304. doi:10.1016/j.neubiorev.2012.08.006
- Dellaert, F., Polzin, T., & Waibel, A. (1996). Recognizing emotion in speech. In *Proceedings of the International Congress on Spoken Language Processing* (pp. 786–789). Philadelphia.
- Devillers, L., Vidrascu, L., & Lamel, L. (2005). Challenges in real-life emotion annotation and machine learning based detection. *Neural Networks*, 18(4), 407–422. doi:10.1016/j.neunet.2005.03.007
- Dewaele, J.-M., & Pavlenko, A. (2002). Emotion vocabulary in interlanguage. *Language Learning*, 52(2), 263–322. doi:10.1111/0023-8333.00185
- Dewaele, J.-M. (2010). *Emotions in multiple languages*. New York: Palgrave Macmillan. doi:10.1057/9780230289505
- 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
- Douglas-Cowie, E., Cowie, R., & Schröder, M. (2000). A new emotion database: Consideration, sources and scope. In *Proceedings of the ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion: SpeechEmotion-2000* (pp. 39–44). Newcastle: International Speech Communication Association.
- Dupuis, K., & Pichora-Fuller, M. K. (2010). Use of affective prosody by young and older adults. *Psychology and Aging*, 25(1), 16–29. doi:10.1037/a0018777
- Dziobek, I. (2012). Comment: Towards a more ecologically valid assessment of empathy. *Emotion Review*, 4(1), 18–19. doi:10.1177/1754073911421390
- Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation 1971* (pp. 207–283). Lincoln, NE: University of Nebraska Press.
- Ekman, P. (1999). Basic emotions. In T. Dalgleish, & M. Power (Eds.), *Handbook of cognition and emotion* (pp. 301–320). New York: John Wiley.

- Engberg, I. S., & Hansen, A. V. (1996). *Documentation of the Danish emotional speech database (DES)* (Technical AAU report). Retrieved from Aalborg University, Center for PersonKommunikation.
- Fragopanagos, N., & Taylor, J. G. (2005). Emotion recognition in human–computer interaction. *Neural Networks, 18*(4), 389–405.
- Frijda, N. H., Suprapti, M., Kaori, S., & Reinout, W. (1995). Emotions and emotion words. In J. A. Russell, J.-M. Fenandez-Dols, A. S. R. Manstead, & J. C. Wellenkamp (Eds.), *Everyday concepts of emotion. An introduction to the psychology, anthropology and linguistics of emotion* (pp. 121–143). Dordrecht: Kluwer Academia.
- Giovannella, G., Conflitti, D., Santoboni, R., & Paoloni, A. (2009). Transmission of vocal emotion: Do we have to care about the listener? The case of the Italian speech corpus EMOVO. In *Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction: ACII 2009* (pp. 494–499). Amsterdam: Institute of Electrical and Electronics Engineers.
- Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge mindreading (CAM) face-voice battery: Testing complex emotion recognition in adults with and without Asperger syndrome. *Journal of Autism and Developmental Disorders, 36*(2), 169–83. doi:10.1007/s10803-005-0057-y
- Golan, O., Baron-Cohen, S., Hill, J., & Rutherford, M. D. (2007). The “Reading the Mind in the Voice” test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders, 37*(6), 1096–1106. doi:10.1007/s10803-006-0252-5
- Greasley, P., Setter, J., Waterman, M., Sherrard, C., Roach, P., Arnfield, S., & Horton, D. (1995). Representation of prosodic and emotional features in a spoken language database. In *Proceedings of the 13th International Congress of Phonetic Sciences, Vol. 1* (pp. 242–245). Stockholm: International Phonetic Association.
- Grichkovtsova, I., Morel, M., & Lacheret, A. (2012). The role of voice quality and prosodic contour in affective speech perception. *Speech Communication, 54*(3), 414–429. doi:10.1016/j.specom.2011.10.005
- Grimm, M., Kroschel, K., & Narayanan, S. (2008). The Vera am Mittag German audio-visual emotional speech database. In *Proceedings of the 2008 IEEE International Conference on Multimedia and Expo (ICME 2008)* (pp. 865–868). Hannover: Institute of Electrical and Electronics Engineers. doi:10.1109/ICME.2008.4607572
- Hawk, S. T., van Kleef, G. a, Fischer, A. H., & van der Schalk, J. (2009). “Worth a thousand words”: absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion, 9*(3), 293–305. doi:10.1037/a0015178
- Heikkinen, J., Jansson-Verkasalo, E., Toivanen, J., Suominen, K., Väyrynen, E., Moilanen, I., & Seppänen, T. (2009). Perception of basic emotions from speech prosody in adolescents with Asperger’s syndrome. *Logopedics, Phoniatrics, Vocology, 35*(3), 113–120. doi:10.3109/14015430903311184
- Huber, R., Batliner, A., Buckow, J., Noth, E., Warnke, V., & Niemann, H. (2000). Recognition of emotion in a realistic dialogue scenario. In *Proceedings of the Sixth International Conference on Spoken Language Processing: ICSLP 2000 / INTERSPEECH-2000* (pp. 665–668). Beijing: International Speech Communication Association.
- Huettigen, M. (2008). Cultural dimensions in business life: Hofstede’s indices for Latvia and Lithuania. *Baltic Journal of Management, 3*(3), 359–376. doi:10.1108/17465260810902414

- Iida, A., Campbell, N., Higuchi, F., & Yasumura, M. (2003). A corpus-based speech synthesis system with emotion. *Speech Communication, 40*(1/2), 161–187. doi:10.1016/S0167-6393(02)00081-X
- Isaacowitz, D. M., Löckenhoff, C. E., Lane, R. D., Wright, R., Sechrest, L., Riedel R., & Costa, P. T. (2007). Age differences in recognition of emotion in lexical stimuli and facial expressions. *Psychology and Aging, 22*(1), 147–159. doi:10.1037/0882-7974.22.1.147
- 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. doi:10.1037/0033-2909.129.5.770
- 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
- Keen, S. (2006). A theory of narrative empathy. *NARRATIVE, 14*(3), 207–236. doi:10.1353/nar.2006.0015
- Kerge, K. (2002). Kirjakeele kasutusvaldkondade süntaktiline keerukus. [Complicacy of Syntax in the domains of Standard Estonian]. In R. Kasik (Ed.), *Tekstid ja taustad. Artikleid tekstianalüüsisist*. [Texts and backgrounds. Papers on text analysis]. *Publications of the Department of Estonian of the University of Tartu, Vol. 23* (pp. 29–46). Tartu: Tartu University Press.
- Kerge, K., Pajupuu, H., Tamuri, K., & Meier, H. (2008). Kõnetehnoloogia vajab žanrlist lähenemist. [Speech technology needs a genre-based approach]. *Estonian Papers in Applied Linguistics, 4*, 53–65. doi:10.5128/ERYa4.04
- Kienast, M., Paeschke, A., & Sendlmeier, W. (1999). Articulatory reduction in emotional speech. In *Proceedings of the Sixth European Conference on Speech Communication and Technology. EUROSPEECH-1999* (pp. 117–120). Budapest: International Speech Communication Association.
- Kovescses, Z. (2000). *Metaphor and emotion. Language, culture, and body in human feeling*. Cambridge: Cambridge University Press.
- Kramer, E. (1964). Elimination of verbal cues in judgements of emotion from voice. *Journal of Abnormal Psychology, 68*(4), 390–396.
- Lazarus, R. S. (1999). *Stress and emotion: A new synthesis*. New York: Springer.
- Laukka, P., & Juslin, P. N. (2007). Similar patterns of age-related differences in emotion recognition from speech and music. *Motivation and Emotion, 31*(3), 182–191. doi:10.1007/s11031-007-9063-z
- Laukka, P., Neiberg, D., Forsell, M., Karlsson, I., & Elenius, K. (2010). Expression of affect in spontaneous speech: Acoustic correlates and automatic detection of irritation and resignation. *Computer Speech & Language, 25*(1), 84–104. doi:10.1016/j.csl.2010.03.004
- Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: Reliability and validity of the empathy quotient. *Psychological Medicine, 34*(5), 911–924. doi:10.1017/S0033291703001624
- Lee, C. M., & Narayanan, S. S. (2005). Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing, 13*(2), 293–303. doi:10.1109/TSA.2004.838534
- Lindner, J. L., & Rosén, L. A. (2006). Decoding of emotion through facial expression, prosody and verbal content in children and adolescents with Asperger's syndrome. *Journal of Autism and Developmental Disorders, 36*(6), 769–777. doi:10.1007/s10803-006-0105-2

- Makarova, V., & Petrushin, V. A. (2003). Phonetics of emotion in Russian speech. In D. Sole, & J. Romero (Eds.), *Proceedings of the XVth International Conference of Phonetic Sciences* (pp. 2857–2860). Barcelona: International Phonetic Association.
- Makarova, V., & Petrushin, V. A. (2012). Phonetics. Tracing emotions in Russian vowels. In V. Makarova (Ed.), *Russian language studies in North America: New perspectives from theoretical and applied linguistics* (pp. 3–42). Anthem Series on Russian, East European and Eurasian Studies. London, NY: Anthem Press.
- Markus, H. R., & Kitayama, S. (2001). The cultural construction of self and emotion. Implications for social behavior. In W. G. Parrot (Ed.), *Emotions in Social Psychology. Essential Readings* (pp. 119–137). New York/Succex: Taylor and Francis, Psychology Press.
- Matte, A. C. F. (2006). Neutral Speech Corpora – a test for neutrality. In *Proceedings of the 3rd International Conference on Speech Prosody, Vol. 1.* (pp. 1–4). Dresden: TUD Press. Retrieved from http://www.isca-speech.org/archive/sp2006/papers/sp06_099.pdf
- Mihkla, M., Meister, E., & Eek, A. (2000). Eesti keele tekst-kõne süntees: grafeem-foneem teisendus ja prosoodia modelleerimine. [Estonian text-to-speech synthesis: grapheme-to phoneme conversion and prosody modelling]. In T. Hennoste (Ed.), *Arvutuslingvistikalt inimesele. [From computational linguistics to people]. Publications of the Department of Estonian of the University of Tartu, Vol. 1* (pp. 306–320). Tartu: Tartu University Press.
- 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 3rd Baltic Conference on Human Language Technologies* (pp. 181–187). Vilnius: Vytauto Didžiojo Universitetas, Lietuvių Kalbos Institutas.
- Mihkla, M.; Hein, I., Kalvik, M., 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.
- Mill, A., Allik, J., Realo, A., & Valk, R. (2009). Age related differences in emotion recognition ability: A cross-sectional study. *Emotion, 9*(5), 619–630. doi:10.1037/a0016562
- Mitchell, R. L. C. (2007). Age-related decline in the ability to decode emotional prosody: Primary or secondary phenomenon? *Cognition & Emotion, 21*(7), 1435–1454. doi:10.1080/02699930601133994
- Montero, J. M., Gutierrez-Arriola, J., Palazuelos, S., Enriquez, E., Aguilera, S., & Pardo, J. M. (1998). Emotional speech synthesis. From speech database to TTS. In *Proceedings of the 5th International Conference on Spoken Language Processing: ICSLP 1998* (pp. 923–926). Sydney: International Speech Communication Association.
- Muncer, S. J., & Ling, J. (2006). Psychometric analysis of the empathy quotient (EQ) scale. *Personality and Individual Differences, 40*(6), 1111–1119. doi:10.1016/j.paid.2005.09.020
- 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 Acoustic Society of America, 93*(2), 1097–1108. doi:10.1121/1.405558
- Mõttus, R., Pullmann, H., & Allik, J. (2006). Toward more readable Big Five Personality Inventories. *European Journal of Psychological Assessment, 22*(3), 149–157. doi:10.1027/1015-5759.22.3.149

- Navas, E., Castelruiz, A., Luengo, I., Sánchez, J., & Hernáez, I. (2004). Designing and recording an audiovisual database of emotional speech in Basque. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)* (pp. 1387–1390). Lisbon: European Language Resources Association.
- Nõlvak, A., & Valk, R. (2003). Isiksus ja emotsioonid. [Personality and emotions]. In J. Allik, A. Realo, & K. Konstabel (Eds.), *Isiksusepsühholoogia* (pp. 169–191). [Psychology of personality]. Tartu: Tartu Ülikooli Kirjastus.
- Pajupuu, H. (2012). Emotsioonid – kõnetehnoloogia olevik ja tulevik. [Emotions – the present and future of speech technology]. *Keel ja Kirjandus*, 8/9, 629–643.
- Paulmann, S., Pell, M. D., Kotz, S. A. (2008). How aging affects the recognition of emotional speech. *Brain and Language*, 104(3), 262–269. doi:10.1016/j.bandl.2007.03.002
- Pavlenko, A. (2002). Bilingualism and emotions. *Multilingua*, 21(1), 45–78. doi:10.1515/mult.2002.004
- Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, 33(2), 107–120. doi:10.1007/s10919-008-0065-7
- Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, 37(4), 417–435. doi:10.1016/j.wocn.2009.07.005
- Phillips, L. H., MacLean, R. D. J., & Allen, R. (2002). Age and the understanding of emotions: Neuropsychological and sociocognitive perspectives. *Journal of Gerontology: Psychological Sciences*, 57B(6), 526–530. doi:10.1093/geronb/57.6.P526
- Pittermann, J., Pittermann, A., & Minker, W. (2010). *Handling emotions in human-computer dialogues*. Berlin: Springer. doi:10.1007/978-90-481-3129-7
- Polzehl, T., Schmitt, A., Metze, F., & Wagner, M. (2011). Anger recognition in speech using acoustic and linguistic cues. *Speech Communication*, 53(9/10), 1198–1209. doi:10.1016/j.specom.2011.05.002
- Raport 2011=Eesti emotsionaalse kõne korpus. *Projekti lõpparuanne*. [Estonian Emotional Speech Corpus. Final report of the project]. (2011). Tallinn: Eesti Keele Instituut.
- Rodero, E. (2011). Intonation and emotion: influence of pitch levels and contour type on creating emotions. *Journal of Voice: Official Journal of the Voice Foundation*, 25(1), 25–34. doi:10.1016/j.jvoice.2010.02.002
- Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders*, 32(3), 189–194. doi:10.1023/A:1015497629971
- Scherer, K. R. (1986). Vocal affect expression. A review and a model for future research. *Psychological Bulletin*, 99(2), 143–165. doi:10.1037/0033-2909.99.2.143
- Scherer, K. R. (1989). Vocal correlates of emotion. In H. Wagner, & A. Manstead (Eds.), *Handbook of Psychophysiology. Emotion and Social Behavior* (pp. 165–197). London: Wiley.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1/2), 227–256. doi:10.1016/S0167-6393(02)00084-5
- Scherer, K. R., 2010. Emotion and emotional competence: conceptual and theoretical issues for modeling. In: K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.),

- Blueprint for Affective Computing: A Sourcebook* (pp. 3–20). Oxford: Oxford University Press.
- Scherer, K. R. (2013). Vocal markers of emotion: Comparing induction and acting elicitation. *Computer Speech & Language*, 27(1), 40–58. doi:10.1016/j.csl.2011.11.003
- Scherer, K. R., Banse, R., & Walcott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32(1), 76–92. doi:10.1177/0022022101032001009
- Scherer, K. R., Clark-Polner, E., & Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology*, 46(6), 401–435. doi:10.1080/00207594.2011.626049
- Schröder, M. (2001). Emotional speech synthesis: A review. In P. Dalsgaard, B. Lindberg, H. Benner, & Z.-H. Tan (Eds.), *Proceedings of the EUROSPEECH 2001 Scandinavia, 7th European Conference on Speech Communication and Technology, 2nd INTERSPEECH Event* (pp. 561–564). Aalborg, Denmark: International Speech Communication Association.
- Schröder, M. (2009). Expressive speech synthesis: Past, present and possible futures. In J. Tao, & T. Tan (Eds.), *Affective information processing* (pp. 111–126). London: Springer. doi:10.1007/978-1-84800-306-4_7
- 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. doi:10.1016/j.specom.2011.01.011
- Shigeno, S. (1998). Cultural similarities and differences in the recognition of audio-visual speech stimuli. In R. Mannell, & J. Robert-Ribes (Eds.), *Proceedings of the 5th International Conference on Spoken Language Processing: ICSLP 1998, Vol. 1* (pp. 281–284). Sydney: International Speech Communication Association.
- Skinner, E. (1935). A calibrated recording and analysis of the pitch, force and quality of vocal tones expressing happiness and sadness. *Speech Monographs*, 2(1), 81–137. doi:10.1080/03637753509374833
- Slobodan, T. J., Zorka, K., Miodrag, D., & Mirjana, R. (2004). Serbian emotional speech database. Design, processing and evaluation. *Proceedings of the 9th International Conference Speech and Computer (SPECOM 2004)*. St.Petersburg, Russia.
- 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
- Zhao, L., Lu, W., Jiang, Y., & Wu, Z. (2000). A study on emotional feature recognition in speech. In *Proceedings of the 6th International Conference on Spoken Language Processing: ICSLP-2000 / INTERSPEECH-2000* (pp. 961–964). Beijing: International Speech Communication Association.
- Zovato, E., Pacchiotti, A., Quazza, S., & Sandri, S. (2004). Towards emotional speech synthesis: A ruled based approach. In A. W. Black, & K. Lenzo (Eds.), *Proceedings of the 5th ISCA ITRW Speech Synthesis* (pp. 219–220). Pittsburgh: International Speech Communication Association.
- Tamuri, K. (2010). Kas pausid kannavad emotsiooni? [Do the pauses in read text carry emotion?]. *Estonian Papers in Applied Linguistics*, 6, 297–306. doi:10.5128/ERYa6.18
- Tamuri, K. (2012). Intensity of Estonian emotional speech. In A. Tavast, K. Muischnek, & M. Koit (Eds.), *Frontiers in Artificial Intelligence and Applications: Vol. 247*.

- Human Language Technologies – The Baltic Perspective. Proceedings of the Fifth International Conference Baltic HLT 2012* (pp. 238–246). Amsterdam: IOS Press.
- Tamuri, K., & Mihkla, M. (2012). Emotions and speech temporal structure. *Linguistica Uralica*, 48(3), 209–217. doi:10.3176/lu.2012.3.07
- Tao, J., & Tan, T. (Eds.). (2009). *Affective information processing*. London: Springer. doi:10.1007/978-1-84800-306-4
- Tawari, A., & Trivedi, M. M. (2010). Speech emotion analysis: exploring the role of context. *IEEE Transactions on Multimedia*, 12(6), 502–509.
- 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. doi:10.1515/SEM.2006.017
- Toivanen, J., Väyrynen, E., & Seppänen, T. (2004). Automatic discrimination of emotion from spoken Finnish. *Language & Speech*, 47(4), 383–412. doi:10.1177/00238309040470040301
- Tuovila, S. (2005). *Kun on tunteet. Suomen kielen tunnesanojen semantiikkaa*. Acta Universitatis Ouluensis B 65. Oulu: Oulun Yliopisto.
- Vainik, E. (2002). Millest on tehtud eestlaste emotsioonisõnavara? [What is the Estonians' emotion vocabulary made of?]. *Keel ja Kirjandus*, 8, 537–553.
- Vainik, E. (2010). Kuidas õpetada kõnesüntesaatorile empaatiat? Emotsiooni automaatse tuvastamise võimalustest eestikeelses kirjalikus lauses sisalduva info põhjal. [How to teach empathy to a speech synthesizer? On the possibility of identifying emotion solely from written Estonian sentences]. *Estonian Papers in Applied Linguistics*, 6, 327–347. doi:10.5128/ERYa6.20
- Vainik, E., & Orav, H. (2005). Tee tööd ja näe vaeva, ... aga ikka oled vihane. [Work hard and moil... but still in anger]. *Keel ja Kirjandus*, 4, 257–277.
- Valk, R. (1999). *Biases in recognition of emotions*. (Unpublished master's thesis). University of Tartu, Institute of Psychology, Tartu.
- Walter, H. (2012). Social cognitive neuroscience of empathy: Concepts, circuits, and genes. *Emotion Review*, 4(1), 9–17. doi:10.1177/1754073911421379
- van Goozen, S., & Frijda, N. H. (1993). Emotion words used in six European countries. *European Journal of Social Psychology*, 23(1), 89–95. doi:10.1002/ejsp.2420230108
- van Rijn, S., Aleman, A., van Diessen, E., Berckmoes, C., Vingerhoets, G., & Kahn, R. S. (2005). What is said or how it is said makes a difference: Role of the right fronto-parietal operculum in emotional prosody as revealed by repetitive TMS. *The European Journal of Neuroscience*, 21(11), 3195–3200. doi:10.1111/j.1460-9568.2005.04130.x
- Ververidis, D., & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication*, 48(9), 1162–1181. doi:10.1016/j.specom.2006.04.003
- Wierzbicka, A. (1999). *Emotions across cultures*. Oxford: Oxford University Press. doi:10.1017/CBO9780511521256
- Wilting, J., Kraemer, E., & Swerts, M. (2006). Real vs. acted emotional speech. In *INTERSPEECH-2006 / ICSLP-2006* (pp. 805–808). Pittsburgh: International Speech Communication Association.
- Vihalemm, T. (2011). Keelepraktikad, kollektiivne identiteet ja mälu. [Linguistic practices, collective identity, and memory]. In *Eesti ühiskonna integratsiooni monitooring 2011* (pp. 113–156). Tallinn: Kultuuriministeerium.

PUBLICATIONS

CURRICULUM VITAE

Name: Rene Altrov
Date of birth: 24 December, 1973
Citizenship: Estonian
Address, phone, e-mail: Koidu 69–4, Tallinn 10139, 522 7203,
rene.altrov@eki.ee
Current position: The Institute of the Estonian Language, Researcher

Education:

University of Tartu, since 09.2007 PhD Student
University of Tartu, 2007, Master's Degree
Estonian Institute of Humanities, 1999, Bachelor's degree

Language skills:

Estonian – mother tongue
English – C1
German – B2

Professional career:

Since 2011 Researcher, Institute of the Estonian Language
2007–2010 Extraordinary Researcher, Institute of the Estonian
Language
2006–2007 Analyst, Institute of the Estonian Language
01.2006–09.2006 Office Manager Mebius IT OÜ
1999–12.2005 Human Resource Specialist, Abobase Systems AS

Main fields of research:

- 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
- Estonian as a modelling communicative system: code, translation, and
history (project no. SF0050042s08), 01.01.2008–31.12.2008
- Estonian Emotional Speech Corpus (project no. EKKTT06–1), 01.01.2006–
31.12.2010

A list of publications:

Altrov, R., & Pajupuu, H. (2013). Estonian Emotional Speech Corpus: Content and options. In J. Bamford, S. Cavalieri, & G. Diani (Eds.), *Dialogue Studies: Vol. 21. Variation and change in spoken and written discourse: Perspectives from corpus linguistics* (pp. 109–122). Amsterdam: John Benjamins.

- Altrov, R., Pajupuu, H., & Pajupuu, J. (2013). The role of empathy in the recognition of vocal emotions. In INTERSPEECH-2013 (pp. 1341–1344). Lyon: International Speech Communication Association.
- Altrov, R. (2013). Aspects of cultural communication in recognizing emotions. *Trames: Journal of the Humanities and Social Sciences*, 17(2).
- 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: European Language Resources Association.
- Pajupuu, H., Kerge, K., & Altrov, R. (2012). Lexicon-based detection of emotion in different types of texts: preliminary remarks. *Eesti Rakenduslingvistika Ühingu Aastaraamat*, 8, 171–184.
- Pajupuu, H., Kerge, K., & Altrov, R. (2012). Detecting emotional valence of text by using a small dictionary. In I. Elorza, O. Carbonell i Cortés, R. Albarrán, B. García Riaza, & M. Pérez-Veneros (Eds.), *Empiricism and Analytical Tools for 21st Century Applied Linguistics. Selected Papers from the XXIX International Conference of the Spanish Association of Applied Linguistics (AESLA)* (229–242). Salamanca: Universidad de Salamanca.
- Altrov, R., & Pajupuu, H. (2010). Estonian Emotional Speech Corpus: Culture and age in selecting corpus testers. In I. Skadina, & A. Vasiljevs (Eds.), *Frontiers in Artificial Intelligence and Applications: Vol. 219. Human Language Technologies – The Baltic Perspective. Proceedings of the Fourth International Conference Baltic HLT 2010* (pp. 25–32). Amsterdam: IOS Press.
- Altrov, R., & Pajupuu, H. (2008). The Estonian Emotional Speech Corpus: Release 1. In F. Čermak, R. Marcinkevičienė, E. Rimkutė, & J. Zabarskaitė (Eds.), *Proceedings of the 3rd Baltic Conference on Human Language Technologies* (pp. 9–15). Vilnius: Vytauto Didžiojo Universitetas, Lietuvių Kalbos Institutas.
- Altrov, R. (2008). Eesti emotsionaalse kõne korpus: teoreetilised toetuspunktid. *Keel ja Kirjandus*, 4, 261–272.
- Kerge, K., Pajupuu, H. & Altrov, R. (2007). Tekst, kontekstuaalsus ja kultuur. *Keel ja Kirjandus*, 8, 624–637.
- Altrov, Rene (2007). Kultuuri mõju eestlaste ja leedulaste ruumikasutusele. *Eesti Rakenduslingvistika Ühingu Aastaraamat*, 3, 7–34.

Other professional activities:

Since 2013 Member of ISCA (International Speech Communication Association)

Scholarships:

2011 Ilse Lehiste's scholarship of the Estonian American Fund

ELULOOKIRJELDUS

Nimi: Rene Altrov
Sünniaeg: 24. detsember, 1973
Kodakondsus: Eesti
Aadress, tel, e-mail: Koidu 69–4, Tallinn 10139, 522 7203,
rene.altrov@eki.ee
Praegune töökoht: Eesti Keele Instituut, teadur

Haridus:

Tartu Ülikool, 09.2007 alates doktorant
Tartu Ülikool, 2007, magistrikraad (teaduskraad)
Eesti Humanitaarinstituut, 1999, bakalaureusekraad

Keelteoskus:

Eesti keel – emakeel
Inglise keel – C1
Saksa keel – B2

Teenistuskäik:

2011 alates teadur, Eesti Keele Instituut
2007–2010 erakorraline teadur, Eesti Keele Instituut
2006–2007 analüütik, Eesti Keele Instituut
01.2006–09.2006 bürojuhataja, Mebius IT OÜ
1999–12.2005 personalispetsialist, Abobase Systems AS

Peamised uurimisvaldkonnad :

- Kõne ja teksti emotsionaalsuse statistilised mudelid, 01.01.2011 – 31.12.2014
- Eesti keele alusuuringud keeletehnoloogiliste rakenduste teenistuses, (projekt nr SF0050023s09), 01.01.2009–31.12.2014
- Eesti keel kui modelleeriv kommunikatiivne süsteem: kood, tõlge ja ajalugu (projekt nr SF0050042s08), 01.01.2008–31.12.2008
- Eesti emotsionaalse kõne korpus (projekt nr EKKTT06–1), 01.01.2006–31.12.2010

Publikatsioonide loetelu:

Altrov, R., & Pajupuu, H. (2013). Estonian Emotional Speech Corpus: Content and options. Rmt-s J. Bamford, S. Cavalieri, & G. Diani (Toim), *Dialogue Studies: Vol. 21. Variation and change in spoken and written discourse: Perspectives from corpus linguistics* (lk 109–122). Amsterdam: John Benjamins.

Altrov, R., Pajupuu, H., & Pajupuu, J. (2013). The role of empathy in the recognition of vocal emotions. Rmt-s INTERSPEECH-2013 (lk 1341–1344). Lyon: International Speech Communication Association.

- Altrov, R. (2013). Aspects of cultural communication in recognizing emotions. *Trames: Journal of the Humanities and Social Sciences*, 17(2).
- Altrov, R., & Pajupuu, H. (2012). Estonian Emotional Speech Corpus: Theoretical base and implementation. Rmt-s L. Devillers, B. Schuller, A. Batliner, P. Rosso, E. Douglas-Cowie, R. Cowie, & C. Pelachaud (Toim) 4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals (ES3) (lk 50–53). Istanbul: European Language Resources Association.
- Pajupuu, H., Kerge, K., & Altrov, R. (2012). Lexicon-based detection of emotion in different types of texts: preliminary remarks. *Eesti Rakenduslingvistika Ühingu Aastaraamat*, 8, 171–184.
- Pajupuu, H. Kerge, K., & Altrov, R. (2012). Detecting emotional valence of text by using a small dictionary. Rmt-s I. Elorza, O. Carbonell i Cortés, R. Albarrán, B. García Riaza, & M. Pérez-Veneros (Toim), *Empiricism and Analytical Tools for 21st Century Applied Linguistics. Selected Papers from the XXIX International Conference of the Spanish Association of Applied Linguistics (AESLA)* (lk 229–242). Salamanca: Universidad de Salamanca.
- Altrov, R., & Pajupuu, H. (2010). Estonian Emotional Speech Corpus: Culture and age in selecting corpus testers. Rmt-s I. Skadina, & A. Vasiljevs (Toim), *Frontiers in Artificial Intelligence and Applications: Vol. 219. Human Language Technologies – The Baltic Perspective. Proceedings of the Fourth International Conference Baltic HLT 2010* (lk 25–32). Amsterdam: IOS Press.
- Altrov, R., & Pajupuu, H. (2008). The Estonian Emotional Speech Corpus: Release 1. Rmt-s F. Čermak, R. Marcinkevičienė, E. Rimkutė, & J. Zabarskaitė (Toim), *Proceedings of the 3rd Baltic Conference on Human Language Technologies* (lk 9–15). Vilnius: Vytauto Didžiojo Universitetas, Lietuvių Kalbos Institutas.
- Altrov, R. (2008). Eesti emotsionaalse kõne korpus: teoreetilised toetuspunktid. *Keel ja Kirjandus*, 4, 261–272.
- Kerge, K., Pajupuu, H. & Altrov, R. (2007). Tekst, kontekstuaalsus ja kultuur. *Keel ja Kirjandus*, 8, 624–637.
- Altrov, Rene (2007). Kultuuri mõju eestlaste ja leedulaste ruumikasutusele. *Eesti Rakenduslingvistika Ühingu Aastaraamat*, 3, 7–34.

Teadusorganisatsiooniline tegevus:

2013 alates Rahvusvahelise kõnekommunikatsiooni assotsiatsiooni ISCA liige

Saadud stipendiumid:

2011 Eesti Ameerika Fondi Ilse Lehiste stipendium

DISSERTATIONES LINGUISTICAE UNIVERSITATIS TARTUENSIS

1. **Anna Verschik.** Estonian yiddish and its contacts with coterritorial languages. Tartu, 2000, 196 p.
2. **Silvi Tenjes.** Nonverbal means as regulators in communication: socio-cultural perspectives. Tartu, 2001, 214 p.
3. **Iлона Trangel.** Eesti keele tuumverbid. Tartu, 2003, 196 p.
4. **Einar Meister.** Promoting Estonian speech technology: from resources to prototypes. Tartu, 2003, 217 p.
5. **Ene Vainik.** Lexical knowledge of emotions: the structure, variability and semantics of the Estonian emotion vocabulary. Tartu, 2004, 166 p.
6. **Heili Orav.** Isiksuseomaduste sõnavara semantika eesti keeles. Tartu, 2006, 175 p.
7. **Larissa Degel.** Intellektuaalsfäär intellektuaalseid võimeid tähistavate sõnade kasutuse põhjal eesti ja vene keeles. Tartu, 2007, 225 p.
8. **Meelis Mihkla.** Kõne ajalise struktuuri modelleerimine eestikeelsele tekst-kõne sünteesile. Modelling the temporal structure of speech for the Estonian text-to-speech synthesis. Tartu, 2007, 176 p.
9. **Mari Uusküla.** Basic colour terms in Finno-Ugric and Slavonic languages: myths and facts. Tartu, 2008, 207 p.
10. **Petar Kehayov.** An Areal-Typological Perspective to Evidentiality: the Cases of the Balkan and Baltic Linguistic Areas. Tartu, 2008, 201 p.
11. **Ann Veismann.** Eesti keele kaas- ja mäarsõnade semantika võimalusi. Tartu, 2009, 145 p.
12. **Erki Luuk.** The noun/verb and predicate/argument structures. Tartu, 2009, 99 p.
13. **Andriela Rääbis.** Eesti telefonivestluste sissejuhatus: struktuur ja suhtlusfunktsioonid. Tartu, 2009, 196 p.
14. **Liivi Hollman.** Basic color terms in Estonian Sign Language. Tartu, 2010, 144 p.
15. **Jane Klavan.** Evidence in linguistics: corpus-linguistic and experimental methods for studying grammatical synonymy. Tartu, 2012, 285 p.
16. **Krista Mihkels.** Keel, keha ja kaardikepp: õpetaja alagatatud parandussekventsides multimodaalne analüüs. Tartu, 2013, 242 lk.
17. **Sirli Parm.** Eesti keele ajasõnade omandamine. Tartu, 2013, 190 lk.