

XXIX Fonetiiikan Päivät Espoo 20.-21.3.2015

Julkaisut - Papers

Mona Lehtinen & Unto K. Laine (toim.)

XXIX Fonetikan päivät, Espoo 20.- 21.3.2015

Julkaisut - Papers

Mona Lehtinen & Unto K. Laine (toim.)

Aalto-yliopiston julkaisusarja
TIEDE + TEKNOLOGIA 7/2015

© Einar Meister & Lya Meister
Janne Savela & Osmo Eerola
Riikka Ylitalo
Tommi Jantunen
Teija Waaramaa, Jarkko Niemi & Jari Eerola
Tommi Nieminen & Tommi Kurki

ISBN 978-952-60-6279-2 (pdf)
ISSN-L 1799-487X
ISSN 1799-487X (painettu)
ISSN 1799-4888 (pdf)
<http://urn.fi/URN:ISBN:978-952-60-6279-2>

Unigrafia Oy
Helsinki 2015

Sisälllys / Table of Contents

Esipuhe / Preface	1
Einar Meister & Lya Meister: F0 Characteristics of Estonian Adolescents	3
Janne Savela & Osmo Eerola: The effect of Vowel Harmony on Identifications and Goodness Ratings of Synthetic Vowels	8
Riikka Ylitalo: Aksentoitujen CV.CV-rakenteisten sanojen F0 ja segmenttikestot Etelä-Pohjanmaalla ja Pohjois-Savossa	12
Tommi Jantunen: Rinnastuksen prosodiaa suomalaisessa viittomakielessä	15
Teija Waaramaa, Jarkko Niemi & Jari Eerola: Emotionaalisen puheen intonaatiopiirteiden graafisen representaation mahdollisuuksia	22
Tommi Nieminen & Tommi Kurki: Prosovar-hankkeen väliraportti: Puheaineiston keruusta verkossa sekä havaintoja aineistosta	29

ESIPUHE

Vuonna 1971 Turun yliopistossa alkanut Fonetikan päivien järjestämisen perinne jatkuu vireänä. Päiviä on pidetty noin puolentoista vuoden välein. Merkillepantavaa on, että Fonetikan päivät ovat syntyneet, jatkuneet ja kehittyneet puhtaasti spontaanin vapaaehtoistyön voimin. Päivien organisoinnista ei siis huolehdi mikään yksittäinen instituutio tai yhdistys. Ehkä juuri tästä johtuen päivät eivät voi koskaan kangistua kaavoihin, koska niitä ei ole, vaan ne jatkavat tätä elävää perinnettä alati uusiutuvassa muodossaan. Nyt Otaniemeen järjestettyjen XXIX päivien yhteydessä tehtiin päätös jopa tiivistää tahtia. Seuraavien päivien vetovastuun sai Oulun yliopisto.

Me XXIX päivien järjestäjät olimme ilahtuneita, että päiville saapui yli 70 osallistujaa. Puheenvuoroja kuultiin 17 kpl ja postereita oli esillä 14 kpl. Tämä kertoo päivien tarpeellisuudesta. Kommunikaation tutkimus fonetiikan, viittomakielen ja puheteknologian eri osa-alueilla voi maassamme kohtalaisen hyvin. Sen saatoimme kuulla ja nyttemmin myös lukea päivien esityksistä ja niiden pohjalta laadituista julkaisuista. Pieni kansakunta, joka kuljettaa mukanaan pientä ja erikoista kieltä tarvitsee yhteyksiä muihin kieliin ja kulttuureihin. Ehkä juuri tästä nousee motivaatiota tutkia syvemmin ja laajemmin puhetta ja kommunikaatiota.

Kiitän lämpimästi kaikkia päiville osallistuneita ja tähän koosteeseen julkaisun laatineita. Erityisesti kiitän tohtoriopiskelijä Mona Lehtistä päivien järjestelyiden ja myös tämän julkaisun laadinnan avustamisesta. Kiitokset myös André Mansikkaniemelle kokouksen [www-sivujen](http://www.sivujen) laatimisesta ja ylläpidosta. Päivien käytännön järjestelyissä olivat apunani myös Seppo Fagerlund ja Okko Räsänen. Henkistä ja myös taloudellista tukea sain professoreilta Paavo Alku ja Mikko Kurimo. Kaikille teille kaunis kiitos!

Espoon Otaniemessä 1.6.2015

Unto K. Laine

PREFACE

In 1971, the University of Turku hosted the 1st Phonetics Symposium - the start of a biennial tradition that is still going strong. Remarkably, the Phonetics Symposium has always functioned solely on the basis of spontaneous volunteerism. No individual organization or association is responsible for the arrangements. This has perhaps led to a lack of rigidity, however, and may be one of the reasons why the conference can easily be renewed and will never get stuck in its ways. During the XXIX Symposium in Otaniemi a decision to arrange these symposiums more often was made, and the next one will be held in Oulu.

We as the organizers were pleasantly surprised to see over 70 people attend. There were altogether 17 presentations and 14 posters, which speaks of the necessity of this type of a meeting.

Communication research in different fields of Phonetics and Language Technology, as well as in sign-language -related Linguistics, is doing quite well in Finland. This can be seen in the presentations and proceedings of the Symposium. A small nation, with a 'small' national language, has to develop connections to other language groups and cultures. Maybe this will motivate further research into speech and communication.

I would like to give warm thanks to the participants of the XXIX Phonetics Symposium and the authors of the articles in this publication. Furthermore, I would like to thank Mona Lehtinen for her help in the organization of the event and in the editorial process of the proceedings. Many thanks also to André Mansikkaniemi for creating and administrating the symposium's Web-pages. Seppo Fagerlund and Okko Räsänen have kindly helped me with the practicalities. I also received moral and financial support from professors Paavo Alku and Mikko Kurimo. Once again, thank you all!

Otaniemi, Espoo
1st June 2015

Unto K. Laine

F0 Characteristics of Estonian Adolescents

Einar Meister, Lya Meister

Laboratory of Phonetics and Speech Technology
Institute of Cybernetics at Tallinn University of Technology
Estonia

einar@ioc.ee, lya@phon.ioc.ee

Abstract

We will report some preliminary data on speaking fundamental frequency characteristics of Estonian adolescents. 175 girls and 134 boys in the age range from 9 to 18 years were recorded while reading a text corpus containing linguistically diverse material.

F0 mean, median, minimum, maximum, and standard deviation were calculated from read utterances. For male speakers, F0 mean decreases gradually from 230 Hz to 190 Hz at the age from 9 to 13 years, due to puberty voice mutation it drops down ca 60 Hz at the age 13–14, and then it lowers further from 130 Hz to 110 Hz at the age from 14 to 18 years. For female speakers, F0 mean shows a gradual change from 240 Hz (9 years) to 210 Hz (18 years).

Index Terms: adolescent speech, fundamental frequency, speech corpus

1. Introduction

Since the 1980s, the creation of large speech corpora has become increasingly important for phonetic analysis and even more important for training and testing speech technology systems. Especially, in speech recognition the approach – *there is no data like more data* – is still valid in the current state-of-the-art technology. The existing Estonian speech resources [1], [2], [3] available for the training of speech recognition systems include the recordings of adult subjects only. Therefore, child and adolescent speech remains challenging for the existing speech recognition systems, as reported for different languages [4], [5], [6].

In addition, there exist no suitable corpora for sociophonetic studies addressing the development of voice and speech of this age group. These were the two main motives for the development of the Corpus of Estonian Adolescent Speech that was started in 2011 under the national program Estonian Language Technology (2011-2017) [7].

The few known corpora of adolescent speech include the Ph@tSessionz speech corpus [8], the VOYS speech corpus [9], the JASMIN corpus [10], the CMU Kids' Speech Corpus [11], the OGI Kids' Speech Corpus [12], and the PF.STAR Children's Speech Corpus [13].

The Ph@tSessionz speech corpus contains recordings of 1019 adolescent speakers of German in the age range 12 – 20. The recordings were performed via the WWW in public schools in 45 locations in Germany.

The VOYS (Voices of Young Scots) corpus comprises recordings of 300 adolescent Scottish speakers in the age range from 13 to 18 years.

The JASMIN corpus includes a subset of contemporary

Dutch spoken by native and non-native children between 7 and 16 years, in total 222 native and 105 non-native subjects.

The CMU Kids' Speech Corpus comprises English sentences read aloud by children in the age from 6 to 11, the number of subjects being 76.

The OGI Kids' Speech Corpus is composed of both prompted and spontaneous speech from 1100 children from kindergarten (5-6 years of age) through grade 10 (15-16 years of age).

The PF.STAR Children's Speech Corpus is a multilingual corpus containing native-language speech in British English, German and Swedish (children aged from 4 to 12), and non-native English speech from German, Italian and Swedish children (children aged from 10 to 11), in total 611 children.

The German and Scottish corpora have served as prototypes for the Estonian corpus.

It is well known that acoustic and linguistic characteristics of children's speech differ widely from those of adult speech [14], [15], e.g., children's speech features higher pitch and formant frequencies. In addition, the acoustic characteristics of children's speech vary as a function of age due to the anatomical and physiological changes. Research on the development of human vocal tract has found anatomic gender differences in the oral and pharyngeal areas of the vocal tract in prepubertal, pubertal and postpubertal age groups (see [16] and references therein). As the acoustic implication of these anatomical changes, age and gender related differences in fundamental frequency and in vowel formants have been documented in different languages (see [17] and references therein). No research on the developmental characteristics of F0 in Estonian adolescents has been previously carried out.

In the paper we will first introduce the design and development of the corpus, and second, the fundamental frequency characteristics of Estonian adolescents measured from read speech.

2. Corpus specification

2.1. Corpus design

The goal was set to collect the corpus of native Estonian subjects in the age range from 9 to 18 years at least from 300 subjects with balanced gender and age distribution, about 20 minutes of speech from each subject. The corpus aims at: (1) covering the age group that is not at all represented in the existing Estonian speech corpora, (2) serving as a resource for speech technology (mainly for training of speech recognition systems) as well as for sociolinguistic and phonetic studies.

2.2. Corpus content

The corpus contains linguistically diverse material: digits, numbers, phone numbers, time and date expressions, IT terms, sentences with place, person and institution names, phonetically rich sentences, two longer passages (see Table 1). Samples of spontaneous speech are elicited with prompts including self-introduction and topic suggestions for storytelling (about school, hobbies, etc.), and pictures to be described. The diversity of the corpus contents ensures compatibility with the existing Estonian adult speech corpora and makes it possible to compare adolescent and adult speech using coherent and partly verbatim speech material. Phonetically rich sentences of the current corpus are derived from the Estonian Babel Corpus [1], and from the Estonian Foreign Accent Corpus [18]. They involve all Estonian vowels and consonants as well as frequent diphthongs and consonant clusters in two-syllable word structures representing Estonian quantity contrasts. Application-oriented items (time and date expressions, digits, PIN-codes, telephone numbers, IT terms) are compatible with the Estonian SpeechDat Corpus [2]. In total, 70 items are recorded from each subject.

Table 1: *Corpus content.*

Item description	Count	
	per session	total
Spontaneous items	10	10
Time and date expressions	5	90
Digits, numbers, PIN-codes	9	270
Phone numbers	4	150
IT terms	5	150
Phonetically rich sentences	21	90
Sentences with place names	3	200
Sentences with person names	3	200
Sentences with object names	5	150
Passages	2	60
Pictures	3	15
Total	70	1385

3. Data collection

3.1. Speaker recruitment

For the recruitment of subjects several schools in the capital area and in different dialectal areas of Estonia were approached. The schools were selected on the basis of their willingness to cooperate with the recording team and of the availability of a suitable room for recordings (usually a quiet classroom). Finally, ten schools – three in Tallinn, one in Harju county (Kehra), two in North-East Estonia (Jõhvi and Narva), two in South-East Estonia (Põlva and Vastseliina), and two in Saaremaa (Kuressaare and Salme) – were selected for recordings. The selected schools are located in the Mid, the Coastal/North-Eastern, the Võru, and the Insular dialectal areas, respectively (Figure 1).

The selection of subjects was made by the schools, typically by a teacher of Estonian according to given criteria (native Estonian, balance by age and gender, no hearing and speaking disorders, fluency in reading of unfamiliar texts). Approvals from head teacher and from parents were obtained beforehand. From each subject the following data was collected: date of birth, gender, school and grade, mother tongue, other languages studied, place of living in early childhood, and current place of living.

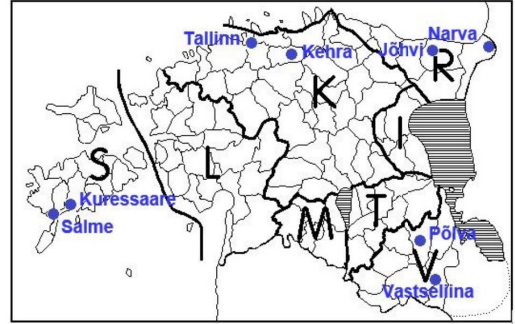


Figure 1: *Dialectal areas and recording places. K – Mid, L – Western, S – Insular, R – Coastal/North-Eastern, I – Eastern, M – Mulgi, T – Tartu, and V – Võru dialectal areas [19].*

3.2. Recordings

The recordings were carried out using a mobile recording set including a laptop with BAS SpeechRecorder software [20], a microphone preamplifier with USB interface (MAudio Mobile Pre), two microphones (desktop microphone Audio-Technica ATM33a and close-talking microphone Sennheiser ME3), and an external monitor to show the prompts. The signals were stored directly to the hard disc in wav format (sampling at 44.1 kHz, resolution 16 bit). At the beginning of each recording session the recording process was explained to the subject and some test prompts were recorded in order to adjust the signal levels from both microphones. The recording session was supervised by a technical operator, and when necessary, additional explanations were given to the subject during the session.

Since almost all subjects had previously never participated in such recordings, they were mostly cooperative and rather excited about the opportunity to participate in the project. However, especially younger pupils were often slightly stressed and worried about their performance in reading unfamiliar texts and too shy in tasks involving story telling or description of pictures. These situations called for a personal approach and encouragement by the recording team in order to help the subject relax and quickly adapt to the recording situation. All subjects were awarded a box of chocolate at the end of the session.

In total, 309 subjects (57% female, 43% male, Table 2) were recorded.

Table 2: *Distribution of recorded subjects in different age groups.*

Age	Male	Female	Total
9	2	2	4
10	12	12	24
11	18	22	40
12	12	23	35
13	18	32	50
14	17	28	45
15	21	15	36
16	16	20	36
17	8	11	19
18	10	10	20
Total	134	175	309

3.3. Quality checking

In general, the quality of most recordings is good. However, as a post-recording quality check of randomly selected signals revealed, in some recorded signals several dropped frames at irregular intervals were found. Possible explanations for these dropped frames are discussed at BAS SpeechRecorder website [20]. In addition, in some signals, occasional clipping and rather high level of background noise were found. The clipping has occurred due to variable reading styles of subjects, usually greater variability of loudness occurring in spontaneous speech. In school environment the background noise is picked up by the microphones especially during lesson breaks when pupils run and communicate loudly in corridors, also the room acoustics (e.g. signal reflection from the walls) causes additional degradation of signal quality.

4. F0 characteristics

For F0 analysis of each subject a subset of read items (20 sentences) were used, the acoustic analysis was carried out in Praat [21] using autocorrelation method for F0 extraction.

A two-pass procedure of F0-extraction suggested in [22] was implemented in a Praat script used to calculate F0 mean, median, minimum, maximum, and standard deviation. Different settings of lower and upper frequency for male and female subjects were used. In the case of male subjects specific adjustment of F0-settings depending on the age was necessary. F0 statistics were calculated first for each utterance, and then pooled over utterances to obtain the results for each subject. The results by age and gender are reported in Table 3 and in Figures 2 and 3.

Table 3: F0 statistics (in Hz) of adolescent read speech for female and male subjects in different age groups.

Gender	Age	Min	Max	Mean	Median	Sd	Range
F	9	178	319	239	238	19.0	141
F	10	190	323	244	242	18.8	134
F	11	177	318	235	232	19.6	141
F	12	173	310	228	226	27.6	137
F	13	175	300	223	221	19.0	126
F	14	170	294	217	215	18.2	124
F	15	175	302	223	220	17.4	128
F	16	170	289	214	212	13.5	119
F	17	170	296	218	214	18.0	125
F	18	168	297	212	208	15.4	129
M	9	183	299	230	230	14.9	116
M	10	163	295	219	217	18.8	131
M	11	172	297	221	219	24.3	124
M	12	164	281	212	211	35.6	118
M	13	143	249	186	184	36.5	106
M	14	97	173	128	126	36.1	76
M	15	81	159	114	111	29.2	78
M	16	89	157	117	114	26.3	68
M	17	83	147	110	105	27.0	64
M	18	83	150	110	106	15.8	67

For female speakers, F0 mean/median shows a gradual change from ca 240 Hz (9 years) to ca 210 Hz (18 years). In contrast, male speakers' F0 mean/median decreases gradually from 230 Hz to 185 Hz in the age from 9 to 12 years, due to puberty voice mutation it drops down ca 60 Hz at the age 13–14,

and then it lowers further from ca 125 Hz to 110 Hz at the age from 14 to 18 years (Figure 2). Also the other F0 characteristics (min, max, and median) show similar patterns (Figure 3).

The standard deviation (SD) of F0 means shows the largest values in males of age 12 to 14 years; in general, male groups show larger values of SD than female groups. In females, F0 range slightly narrows (ca 15 – 20 Hz) with age; in males it narrows significantly after the age of 13.

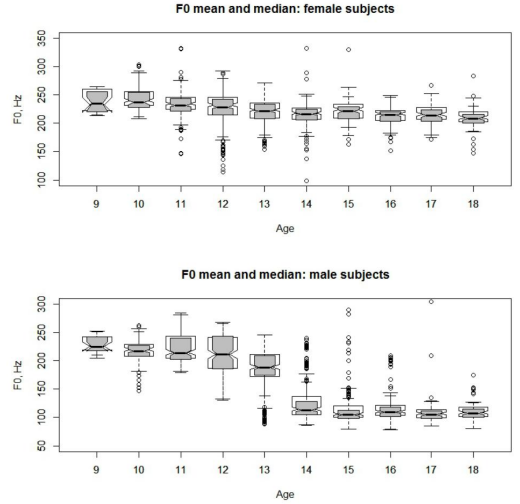


Figure 2: Boxplots of F0 mean and median (gray) for female (upper) and male (bottom) subjects in different age groups.

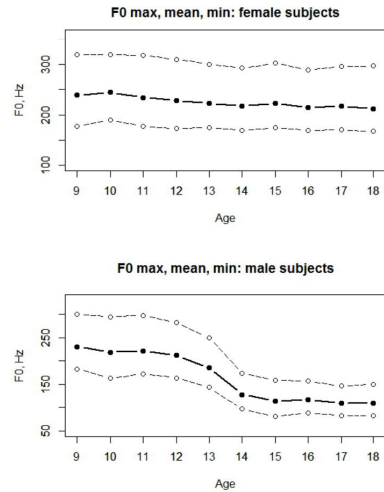


Figure 3: F0 min (upper dashed line), mean (mid solid line) and max (bottom dashed line) for female (upper) and male (bottom) subjects in different age groups.

4.1. Individual differences

There are individual differences in F0 characteristics within each age group. Large values of the SD of F0 mean in males of age 12 to 14 (see Table 3) suggest that the subjects of these age groups are in the process of voice mutation.

The F0 mean for the age group of 12 years is 212 Hz (SD=35.6); however, for 5 subjects (out of 12) the F0 mean has dropped down under 200 Hz (subjects K_203, K_204, K_205, K_207, and K_212) (Figure 4, top).

In the age group of 13 years the F0 mean is 186 Hz (SD=36.5), but two subjects (out of 18) show F0 means <160 Hz (subjects K_193 and K_211), and two subjects as low as 111 Hz (subject K_271) and 101 Hz (subject K_283) (Figure 4, mid).

The age group of 14 years manifests ca 60 Hz decrease of F0 mean in comparison to the 13 years' group: F0 mean = 128 Hz (SD=36.1). For most of the subjects in this group the voice mutation process has taken place; however, in the case of three subjects (K_248, K_269, and K_300) the process is delayed and they have still a rather high F0 mean: 225 Hz, 164 Hz, and 194 Hz, respectively (Figure 4, bottom).

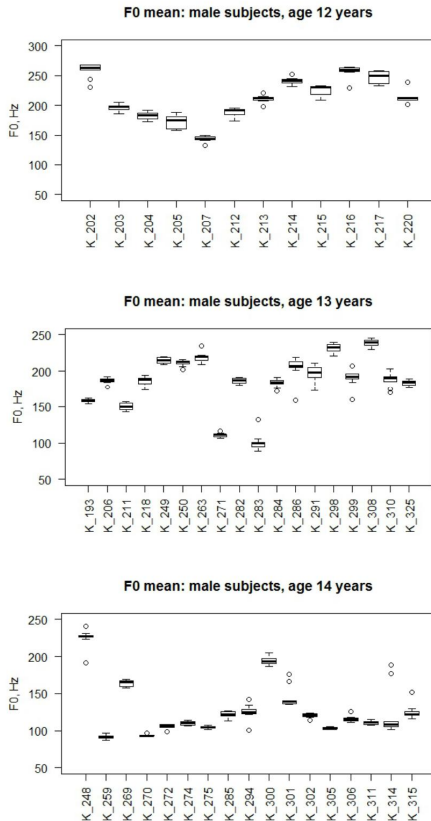


Figure 4: Boxplots of F0 mean for male subjects in the age of 12 (top), 13 (mid) and 14 (bottom) years.

5. Discussion

The study has shown that female F0 decreases linearly throughout development, while male F0 gradually lowers until around the age 12–13, and then drops ca 60 Hz during a year, resulting in F0 mean around 110–115 Hz at the age of 15 years.

The results are in line with earlier studies reporting similar F0 characteristics in different languages [23], [24], [15], [25]. However, the voice mutation period is observed earlier in Estonian males, occurring as it does in the age interval from 12 to 14 years, whereas in the case of German males it takes place between 13 and 15 years of age. Also the post-pubertal F0 mean value is higher in German males (ca 126 Hz) than in Estonian males (110 Hz) [25].

6. Conclusions

We have introduced the work in progress on the development of the Corpus of Estonian Adolescent Speech and presented preliminary results on F0 analysis of male and female subjects depending on age.

The corpus will be available for registered users via the Center of Estonian Language Resources (<http://keelelressursid.ee/eng/>) and internationally via the EU CLARIN infrastructure (<http://www.clarin.eu/>).

7. Acknowledgements

We thank the schools that made the data collection possible and all volunteer speakers. The work has been supported by the National Program for Estonian Language Technology.

8. References

- [1] A. Eek and E. Meister, "Estonian speech in the BABEL multi-language database: Phonetic-phonological problems revealed in the text corpus." in *Proceedings of LP'98*, O. Fujimura, Ed., Prague: The Karolinum Press, 1999, pp. 529–546.
- [2] E. Meister, J. Lasn, and L. Meister, "SpeechDat-like Estonian database." in *Text, Speech and Dialogue: 6th International Conference, TSD 2003*, ser. Lecture Notes in Artificial Intelligence 2807, V. Matoušek and P. Mautner, Eds., Berlin: Springer, 2003, pp. 412–417.
- [3] E. Meister, L. Meister, and R. Metsvahi, "New speech corpora at IoC." in *XXVII Fonetikain pivi 2012 - Phonetics Symposium 2012*, E. Meister, Ed., Tallinn: TUT Press, 2012, pp. 30–33.
- [4] M. Gerosa, D. Giuliani, and F. Brugnara, "Acoustic variability and automatic recognition of childrens speech." *Speech Communication*, vol. 49, no. 10–11, pp. 847–860, 2007.
- [5] S. Narayanan and A. Potamianos, "Creating conversational interfaces for children." *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 10, no. 2, pp. 65–78, 2002.
- [6] S. D'Arcy and M. Russell, "A comparison of human and computer recognition accuracy for childrens speech." in *proceedings of Interspeech*, 2005.
- [7] "National Program for Estonian Language Technology." <https://www.keeletehnoloogia.ee/>.
- [8] C. Draxler and K. Jänsch, "Speech recordings in public schools in Germany the perfect show case for web-based

- recordings and annotation.” in *Proceedings of LREC 2006*, Genova, 2006.
- [9] C. Dickie, F. Schaeffler, C. Draxler, and K. Jänsch, “Speech recordings via the internet: An overview of the VOYS project in Scotland.” in *Proceedings of Interspeech 2009*, Brighton, 2009.
- [10] C. Cucchiari and H. Van hamme, “The JASMIN speech corpus: Recordings of children, non-natives and elderly people.” in *Essential Speech and Language Technology for Dutch, Theory and Applications of Natural Language Processing*, P. Spyns and J. Odijk, Eds., 2013, pp. 43–59.
- [11] M. Eskenazi, J. Mostow, and D. Graff, “The CMU Kids Corpus LDC97s63,” <https://catalog.ldc.upenn.edu/LDC97s63>, 1997, Philadelphia: Linguistic Data Consortium.
- [12] K. Shobaki, J.-P. Hosom, and R. A. Cole, “The OGI kids’ speech corpus and recognizers,” in *In ICSLP*, 2000.
- [13] A. Batliner, M. Blomberg, D. Elenius, D. Giuliani, M. Gerosa, C. Hacker, M. Russell, S. Steidl, and M. Wong, “The PF STAR Children’s Speech Corpus,” 2005.
- [14] S. Lee, A. Potamianos, and S. Narayanan, “Acoustic of children’s speech: Developmental changes of temporal and spectral parameters.” *Journal of Acoustical Society of America*, vol. 105, no. 3, pp. 1455–1468, 1999.
- [15] J. E. Huber, E. T. Stathopoulos, G. M. Curione, T. A. Ash, and K. Johnson, “Acoustic of children’s speech: Developmental changes of temporal and spectral parameters.” *Journal of Acoustical Society of America*, vol. 106, no. 3, pp. 1532–1542, 1999.
- [16] H. Vorperian, S. Wang, E. Schimek, R. Durtschi, R. Kent, L. Gentry, and M. Chung, “Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study.” *Journal of Speech, Language, and Hearing Research*, vol. 54, pp. 995–1010, 2011.
- [17] S. Xue, R. Cheng, and L. Ng, “Vocal tract dimensional development of adolescents: An acoustic reflection study.” *International Journal Of Pediatric Otorhinolaryngology*, vol. 74, no. 8, pp. 907–912, 2010.
- [18] L. Meister and E. Meister, “Aktsendikorpus ja vöörkeele aktsendi uurimine.” *Keel ja Kirjandus*, vol. 55, no. 8–9, pp. 696–714, 2012.
- [19] “Eesti murdered.” Retrieved 4 May 2015 from <http://www.eki.ee/murded/>.
- [20] “Speechrecorder [computer program].” Version 2.8.4, retrieved 2 May 2014 from <http://www.bas.unimuenchen>.
- [21] P. Boersma and D. Weenink, “Praat: doing phonetics by computer [computer program].” Version 5.4.08, retrieved 24 March 2015 from <http://www.praat.org/>.
- [22] D. Hirst, “A Praat plugin for Momel and Intsint.” in *Proceedings of ICPHS 2007*, Saarbrücken, 2007.
- [23] H. Hollien, R. Green, and K. Massey, “Longitudinal research on adolescent voice change in males.” *Journal of Acoustical Society of America*, vol. 96, no. 5, pp. 2646–2654, 1998.
- [24] M. Harries, S. Hawkins, J. Hacking, and I. Hughes, “Changes in the male voice at puberty: vocal fold length and its relationship to the fundamental frequency of the voice.” *Journal of Laryngology & Otology*, vol. 112, no. 5, pp. 451–454, 1998.
- [25] C. Draxler, F. Schiel, and T. Ellbogen, “F0 of adolescent speakers — first results for the German Ph@ttSessionz database.” in *Proceedings of LREC 2008*, Marrakech, 2008.

The effect of Vowel Harmony on Identifications and Goodness Ratings of Synthetic Vowels

Janne Savela¹, Osmo Eerola²

¹ Department of Information Technology, University of Turku

² Business, ICT and Chemical Engineering, Turku University of Applied Sciences

janne.savela@utu.fi, osmo.eerola@turkuamk.fi

Abstract

Three vowel listening experiments were conducted, one with isolated vowels, and two in word-context, using a large vowel grid covering all Finnish vowels. Subjects identified and rated the goodness of the vowel stimuli. Vowel prototypes in the word context condition seemed to shift in the F1-F2 space. The vowel harmony affected the identification of the latter vowels, especially /e/.

Index Terms: vowel perception, prototypes, contrast-effect, vowel harmony

1. Introduction

This study has two aims: First, to study the effect of word context vs. isolated vowels in vowel perception. In the isolated condition the vowels are identified alone. In the word context they are identified in the second syllable of two syllable words: hahV and hihV. Since vowel prototypes are considered as stable marks of vowel identity [1], it is interesting to explore their use in word context.

It is well known phenomenon (for review [2]), that different pre-cursor effects do affect the identification of vowels (e.g. [3][4][5][6][7][8]). The basic contrast effect implies, that the area of precursor sound decreases compared to identification of same stimuli in isolated situation.

In these studies either one or two continua are used. Fry et al. [7] used vowels from /i-ε-æ/ continuum in ABX triads. Repp et al. [5] used /i-i-ε/ in AX identification and discrimination task.

In study by Ainsworth [8] adding a precursor vowel to hVd context changed the vowel to the direction of the first vowel. In Kanamori et al. [4] study the vowels were presented in isolation and in VCV word context. The prototypical first vowel produced larger contrast effect than non-prototypical first vowel.

In the present study the entire vowel space is used, and it is assumed that the area (in F1-F2 space) of /a/ decreases in hahV context and the area of /i/ decreases in hihV context. It will be examined whether this change affects the neighboring vowel categories or all categories similarly.

Secondary question concerns vowel harmony. The Finnish vowel harmony restricts the possible vowels of secondary syllables. The /y, ø, æ/ and /u, o, a/ cannot be grammatically in the same word. Furthermore /i, e/ are neutral. It can be assumed that in hihV context there could be larger contrast effects than

in hahV context, if there is bias to prefer answers in compliance with the restrictions of vowel harmony.

2. Methods

2.1. Stimuli and procedure

The grid of 196 single vowel stimuli was synthesized with KLATT synthesizer (Figure 1).

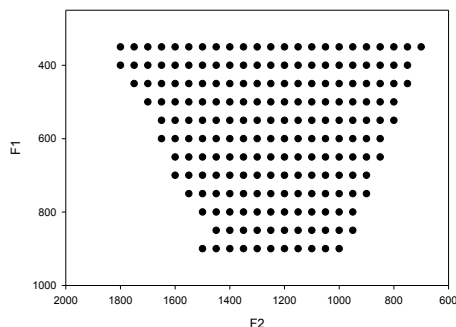


Figure 1: The grid of stimuli in F1-F2 space.

The F1 varied between 350 – 900 mels, and the F2 varied between 700 – 1800 mels. The F3 was constantly 1875 mels. The duration of the stimuli was 300 ms. The F0 increased in beginning of the stimulus (until 100ms) and decreased during 200-300 ms of the stimulus.

The same set of formant values were used both in the isolated and word-context conditions.

There were two sets of word context: hahV context and hihV context (Figure 2).

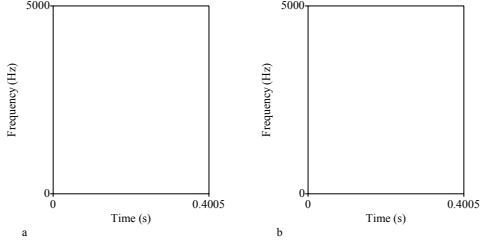


Figure 2: The spectrogram of *hahV(a)* and *hihV(b)* stimuli.

In word contexts the stimuli were synthesized with KLATT synthesizer. The entire stimulus had a duration of 400 ms (both vowel parts were 100ms).

The first vowel in the syllable was a weighted prototype of an earlier study [9]. Hence they are called Prototype vowels.

The three sets of stimuli were presented in randomized order.

The task of each participant was to identify the vowel to one of the eight Finnish vowel categories and rate it then on the scale of 1-7. In the word-context conditions they were asked to identify and rate the vowel of the second syllable.

2.2. Subjects

There were 10 subjects (3 female, average age 38,4). They come from various parts of Finland.

3. Results

The weighted prototypes for different categories were computed (for method [7]). The shift in vowel space was plotted for Single vowel vs. Hah_ context (Figure 3), and Single vowel vs. Hih_ context. (Figure 4).

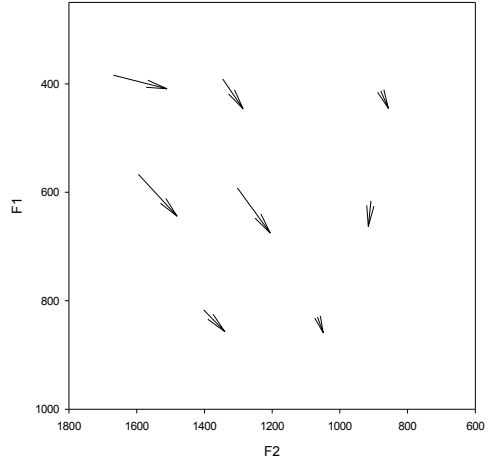


Figure 3: The shift of vowel prototypes between isolated vowels (arrow beginning) and vowels in *Hah* context (arrow top).

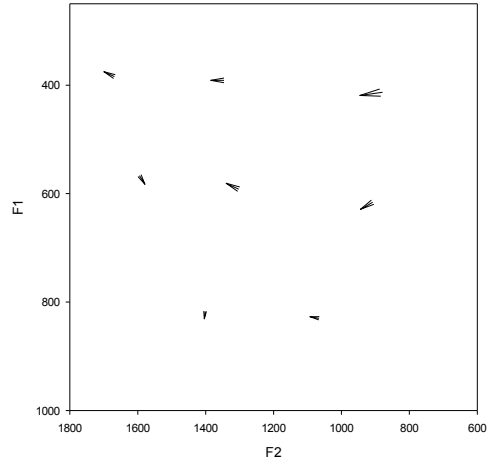


Figure 4: The shift of vowel prototypes between isolated vowels (arrow beginning) and vowels in *Hih* context (Arrow top).

The Euclidean distances between the Prototype vowels and the target vowels were then computed. Prototype vowel is the first vowel in the word context (for /a/, F1=792, and F2=1045 mels, and for /i/, F1=376 and F2=1689 mels (Table 2).

Table 1. *The Euclidean distances in mels (SD in parentheses) between Prototype vowels and the target vowels in the three listening conditions (Single vowel = isolated condition , HahV and HihV = word context conditions).*

	Proto /a/ > Single vowel	Proto /a/ > HahV context	Proto /i/ > Single vowel	Proto /i/ > HihV context
α	47 (25)	72 (17)	772 (17)	748 (27)
e	595 (65)	462 (90)	215 (66)	237 (54)
i	746 (76)	607 (94)	69 (53)	42 (18)
o	224 (37)	183 (22)	818 (13)	788 (24)
u	415 (34)	397 (29)	811 (49)	745 (33)
y	508 (68)	435 (63)	344 (107)	305 (49)
æ	360 (40)	303 (36)	527 (52)	538 (46)
ø	329 (63)	206 (71)	444 (66)	406 (53)

In HahV case the grand average shift over all categories was 70 mels and for HihV 24 mels.

In HahV context Anova showed a significant effect of context and vowel on Euclidean distance between prototype of first and second syllable (for context, (F (1,7)= 31.151, p= 0.014), for vowel (F (7,49)= 221.274, p< 0.001). The interaction between vowel type and context is significant (F (7,63)= 6.680, p= 0.004). Differences in individual vowels are significant for mid vowels (e,ø,o) p< 0.002 (Boniferroni corrected).

In HihV context Anova showed a significant effect of context and vowel on Euclidean distance between prototype of first and second syllable (for context, (F (1,9)= 9.161, p= 0.036), for vowel (F (7,63)= 532.691, p< 0.001). The interaction between vowel type and context is significant (F (7,63)= 3,450, p= 0.025). Using Boniferroni correction none of the t-tests are significant.

The results for 8 vowel categories in three experimental conditions were calculated (Table 1).

Table 2. *The percentage of different vowel categories in the identification experiment with isolated and word context experiment. Two latter columns show the difference in share of vowel identifications (in percentage points) between isolated (S) and word context experiments.*

	S	HahV	HihV	Δ S>HahV	Δ S>HihV
α	11,7	6,5	12,5	-5,2	0,8
e	8,9	18,7	13,8	9,8	4,9
i	8,7	25,3	5	16,6	-3,7
o	10,2	7,6	11,9	-2,6	1,7
u	16,7	15,2	21,1	-1,5	4,4
y	10,6	7,1	10,3	-3,5	-0,3
æ	20,8	11	16,6	-9,8	-4,2
ø	14	10,4	10,6	-3,6	-3,4

An Anova was made using Vowel type and Context type as main factors. For Single vs. Hah Context type had no main effect (F (1,9)= 0.310, p= 0.591). The Vowel type had significant effect (F (7,63)= 5.274, p= 0.011) on difference between identifications in isolated and word context vowels. Finally, interaction between Context and Vowel type was significant (F (7,63)= 11.714, p= 0.000). The pair-wise comparisons of vowel types showed that the shift in percentage with /a/ and /e/ are significant (using Boniferroni correction).

An Anova was made using Vowel type and Context type as main factors. For Single vs. Hih Context, none of the factors are significant.

4. Discussion

The contrast effect was found in both vowel contexts (/a/ and /i/. The prototypical vowel in first syllable decreased the Euclidean distance between it and the weighted prototypes of target vowels in the word context experiments. The largest shifts were in mid vowels. In present study there was smaller shift in hihV context than in hahV context. No separate role of prototypicality of first vowel was explored. Interestingly, some earlier studies [10] showed that /i/ should have larger contrast effect than /e/. In our study /i/ had smaller contrast effect than /a/.

Secondly, Vowel harmony seemed to affect the identification of target vowels. The /e/ answers had larger share of results in the /Hah/ condition than in the rest of conditions, and results more vowel harmony neutral answers. It indicates that the phonological awareness affect the contrast and prototypes besides the basic prototype level perceptual mapping.

According to criterion setting theory by Treisman, [10] three features affect the category boundaries. Firstly, the criteria are based on experience. Secondly, they are based on the probabilities of sounds within an experiment, stabilization (an

idea that the vowels are likely to belong to different categories) and tracking (idea that two stimuli should be similar). According to Rosner [2], the speech experience suggest that vowels in typical study should be different, hence producing a contrast effect, but in language with vowel harmony and lot of multisyllable words the situation may be different

In our experiment the first vowel was always prototypical and same, and suggested therefore a large contrast effect. Interestingly, there was large shift between rounded front vowel /ø/ and open back vowel /ɑ/ than between vowels belonging to same vowel harmony class (/ɑ/ and /o/). The difference between the first two vowels was 124 mels whereas the second two vowels had only shift of 41 mels.

In general the vowel harmony does affect the language experience of Finnish listeners, since the shifts in vowel prototypes and identification were based on the appropriateness of the stimuli in the context of vowel harmony. However, it was larger in vowel harmony sensitive context than in the neutral context.

References

- [1] Savela, J., Eerola, O., Aaltonen, O. (2014) "Weighted vowel prototypes in Finnish and German." *Journal of the Acoustical Society of America*, Vol. 135 (3). Pages 1530-1540.
- [2] Rosner, B.S. & Pickering, J.B. (1994). "Vowel Perception and Production." New York, Oxford University Press. 432 p.
- [3] Lindner, G. (1966). "Veränderung der Beurteilung synthetischer Vokale unter dem Einfluss des Sukzessivkontrastes." *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 19, 287–307.
- [4] Kanamori, Y., Kasuya, H., Arai, S., and Kido, K. (1971). „Effect of context on vowel perception". *Proceedings of the Seventh International Congress on Acoustics, Budapest*, Vol. 3, pp. 37–40. Akademiai Kiado, Budapest.
- [5] Repp, B. H., Healey, A. F., and Crowder, R. G. (1979). "Categories and context in the perception of isolated steady-state vowels." *Journal of Experimental Psychology: Human Perception and Performance*, 5, 129–45.
- [6] Traunmüller, H. and Lacerda, F. (1987). "Perceptual relativity in identification of two-formant vowels." *Speech Communication*, 6, 143–57.
- [7] Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. (1962). "The identification and discrimination of synthetic vowels." *Language and Speech*, 5, 171–89.
- [8] Ainsworth, W. A. (1974). "The influence of precursive sequences on the perception of synthesized vowels." *Language and Speech*, 17, 103–9.
- [9] Eerola, O., Savela, J. (2011) "Differences in Finnish front vowel production and weighted perceptual prototypes in the F1–F2 space." In: *Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, China, 2011*. Pages 631-634.
- [10] Treisman, M. (1985). "The magical number seven and some other features of category scaling: Properties of a model for absolute judgment." *Journal of Mathematical Psychology*, 29, 175–230.
- [11] Morse, P. A., Kass, J. E., and Turkienicz, R. (1976). "Selective adaptation of vowels." *Perception and Psychophysics*, 19, 137–43.

Aksentoitujen CV.CV-rakenteisten sanojen F_0 ja segmenttikestot Etelä-Pohjanmaalla ja Pohjois-Savossa

Riikka Ylitalo¹

¹ Oulun yliopisto

riikka.ylitalo@oulu.fi

Tiivistelmä

Tutkimuksessa tarkastellaan aksentoitujen CV.CV-rakenteisten sanojen perustaajuutta ja segmenttikestoja Etelä-Pohjanmaalta ja Pohjois-Savosta kotoisin olevien henkilöiden puheessa. Tutkimuksessa todetaan paikkansapitäviksi aiemmat havainnot, joiden mukaan Pohjois-Savo kuuluu puolipidennysmurteiden alueeseen, mutta Etelä-Pohjanmaa ei: tutkimusaineiston V_2 :n kesto on Pohjois-Savossa keskimäärin 158,4 % ja Etelä-Pohjanmaalla 82,1 % V_1 :n kestoista. Lausepainoa toteuttavan perustaajuuden nousu-lasku-kuvion huippu on Pohjois-Savossa merkitsevästi korkeampi kuin Etelä-Pohjanmaalla. Huippu sijaitsee Etelä-Pohjanmaalla keskimäärin V_1 :ssä, Pohjois-Savossa C_2 :ssa lähellä segmentin puoltavaliä. Vasta toisessa tavussa sijaitseva perustaajuushuippu ei siis ole pelkästään lounaismurteisiin rajoittuva ilmiö.

Asiasanat: areaalilingvistiikka, perustaajuus, prosodiikka, puolipidennys, segmenttikesto

1. Taustaa

CV.CV(X) yksi harvoista suomen sanarakenteista, joiden ensi tavu on lyhyt eli yksimorainen. CV.CV(X)-rakenteeseen on vanhastaankin havaittu liittyvän murteiden välistä äännesegmenttien kesto vaihtelua: kyseessä on ns. puolipidennysaseman sisältävä sanarakenne, jonka V_2 on puolipidennysmurteissa V_1 :tä pidempikestoinen; puolipidennysksetömissä murteissa V_1 päinvastoin on V_2 :tä pidempikestoinen. Suomen sana- ja lausepainon toteutumisasiala on kutakuinkin sanan kahden ensimmäisen moran rajaama jakso [1, s. 123–136], [2, s. 39–61], [3, s. 326], [4, s. 86–90, 105–121], ja CV.CV(X)-rakenteeseen liittyvät prosodiset vaihtelut saattavatkin olla seurausta siitä, että suomen eri varieteetit käsittelevät eri tavoin erikoista tilannetta, jossa sanan toinen mora sijaitsee vasta sen toisessa tavussa. Väitöskirjassani [4] vertailtiin useita eri sanarakenteita Oulun, Tampereen ja Turun seuduilta kotoisin olevien henkilöiden puheessa, mutta ehdottomasti eniten varieteettien välisiä segmenttikesto- ja perustaajuuseroja ilmeni juuri CV.CV(X)-rakenteessa.

Paitsi perinteisesti tunnetumpia vokaalien kestoeroja, CV.CV-sanoihin liittyi väitöskirjani aineistossa myös konsonanttien kestoeroja ja aksentoiduissa sanoissa perustaajuuseroja. C_1 oli lyhempi Oulussa kuin muissa varieteeteissa ja C_2 pitempi Turussa kuin muissa varieteeteissa. V_1 oli Tampereella pitempi kuin V_2 ja Oulussa ja Turussa V_2 pitempi kuin V_1 , eli Oulun ja Turun varieteetit odotetusti osoittautuivat olevan puolipidennysvarieteetteja, Tampereen varieteetti ei. Aksentoitujen sanojen puolipidennys kuitenkin saatiin Turussa aikaan hieman eri tavoin kuin

Oulussa: aksentoitujen sanojen V_1 oli lyhempi Turussa kuin muissa varieteeteissa, kun taas Oulussa puolipidennys syntyi nimenomaan pitkän V_2 :n avulla. Lausepainoa toteuttavassa F_0 :n nousu-lasku-kuviossa perustaajuus nousi korkeimmalle Oulussa, sitä matalammalle Turussa ja jäi kaikkein matalimmaksi Tampereella. Lisäksi CV.CV(X)-sanojen F_0 :n huippukohta sijaitsi kauempana sanassa Turussa kuin muissa varieteeteissa: Turussa se sijaitsi C_2 :n puolivälin tienoilla sanan 2. tavussa C_2 :n ollessa soinnillinen, soinnittoman C_2 :n sisältävissä sanoissa vasta V_2 :n alussa. Muissa varieteeteissa perustaajuushuippu sijaitsi V_1 :ssä. [4, s. 51–57, 105–121, 136–138.]

2. Tavoite, aineisto ja menetelmät

Koska suurin osa väitöskirjassani todetuista varieteettien välisistä prosodisista eroista liittyi aksentoitujen CV.CV(X)-rakenteisten sanojen kahteen ensimmäiseen tavuun, katson, että kaiken kaikkiaan tarkoituksenmukaisin materiaali suomen alueellisten sanatason prosodisten erojen kartoittamiseen ovat aksentoidut CV.CV-rakenteiset sanat. Tässä tutkimuksessa laajennan aksentoitujen CV.CV-rakenteisten sanojen segmenttikestojen ja perustaajuuden järjestelmällistä tarkastelua Etelä-Pohjanmaalle ja Pohjois-Savoon. Aiempien puolipidennys tutkimusten perusteella ennako-oletuksena on, että Etelä-Pohjanmaalla tutkittavien sanojen V_1 on V_2 :tä pidempikestoinen, Pohjois-Savossa tilanne saattaa olla päinvastainen [5, s. 298–306], [6, s. 570–573]. Perustaajuutta aksentoiduissa CV.CV(X)-rakenteisissa sanoissa näillä murrealueilla ei ole aiemmin tutkittu.

Tutkimuksen koehenkilöt ovat Kuopiosta tai Seinäjoelta samaan murrealueeseen kuuluvine ympäristökuntiineen kotoisin olevia, äänityshetkellä 20–26-vuotiaita Oulun yliopiston opiskelijoita. Molemmilta murrealueelta koehenkilöitä on viisi. He ovat syntyneet ja asuneet koko ikänsä edustamallaan murrealueella lukuun ottamatta korkeintaan muutaman vuoden asumista aikuisiässä opiskelupaikkakunnalla, joka kaikilla koehenkilöillä on Oulu. Koska tutkimusaineiston on tarkoituksenmukaista olla mahdollisimman vertailukelpoista väitöskirjani aineiston kanssa, kaikki tämänkin tutkimuksen koehenkilöt ovat naispuolisia. Koehenkilöiden tietokoneen ruudulta lukemat kontrastiivisesti aksentoituun asemaan sijoitetut kohdesanat yleiskielisine kehysvirkkeineen on tallennettu talvella 2014–15 Oulun yliopiston logopedian ja fonetiikan äänitysstudioissa suoraan tietokoneen kovalevylle (44,1 kHz, 16 bittiä). Kultakin koehenkilöltä on äänitetty 30 CV.CV-rakenteista nimityä, joiden konsonantit on valittu joukosta /l m n/, esim. *lana*, *moni*, *nimi*. Sanat kehysvirkkeineen ovat samat kuin väitöskirjatyössäni Turun varieteetista hankitussa lisäaineistossa. Kyseinen lisäaineisto hankittiin, koska

väitöskirjani alkuperäisessä aineistossa C₂ oli soinniton, ja aksentoitujen sanojen F₀:n keskimääräisen huipun sijoituttua Turun varieteetissa C₂:een oli selvítettävä, missä kyseisen varieteetin perustaajuushuippu sijaitsee, jos C₂ on soinnillinen. Sama materiaali valittiin käytettäväksi myös tässä tutkimuksessa paitsi vertailukelpoisuuden, myös sen takia, että tutkimuksessa käytettävien kohdesanojen C₂:ien on oltava soinnillisia siltä varalta, että jommassakummassa nyt tarkasteltavassa varieteetissa F₀:n huippu sijoittuu C₂:een.

Esimerkki virkkeestä, jonka koehenkilöt lukivat, on *Sanoin että hyvä NIMI puuttuu, en sanonut että hyvä AJATUS puuttuu*, jossa kohdesana on *nimi*. Kaikkien kohdesanojen kehysvirkkeet ovat rakenteeltaan esimerkkivirkettä vastaavia. Kohdesanojen aksentoimisen varmistamiseksi koehenkilöitä pyydettiin ennen äänitystä ”korostamaan” isoin kirjaimin kirjoitettuja sanoja. Kohdesanojen perustaajuus mitattiin Praat-ohjelman [7] avulla hertsin kymmenesosan tarkkuudella kohdesanan ensimmäisen tavun alusta, puolivälistä ja lopusta sekä alun ja keskikohdan puolivälistä ja keskikohdan ja lopun puolivälistä, vastaavista viidestä kohdasta sen toista tavua, kohdesanan perustaajuushuipusta sekä kohdesanaa edeltävän ja sitä seuraavan tavun alusta ja lopusta. Kohdesanojen äännesegmenttien kestot mitattiin millisekunnin tarkkuudella. IBM SPSS -ohjelman avulla tehdyin kahden riippumattoman otoksen t-testein ja yksisuuntaisin varianssianalyysin selvitetiin, miten eri murrealueita edustavien koehenkilöiden tuottamien kohdesanojen äänteiden kestot ja perustaajuuden kulku eroavat toisistaan. Tapauksissa, joissa post hoc -testien tekeminen oli mahdollista, käytettiin Tukeyn post hoc -testiä 5 prosentin riskitasolla.

3. Tulokset ja pohdintaa

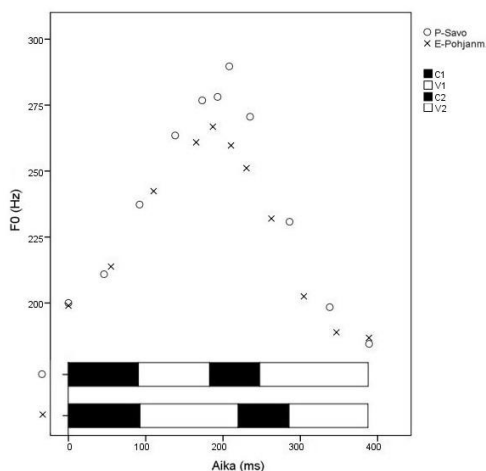
Sanan kokonaiskesto on varieteeteissa tilastollisesti sama: sekä Pohjois-Savossa että Etelä-Pohjanmaalla keskimäärin 389 ms. Varieteetit eivät eroa toisistaan myöskään C₁:n kestossa, joka on Pohjois-Savossa 91 ms ja Etelä-Pohjanmaalla 93 ms. V₁:n kesto on merkitsevästi pitempi Etelä-Pohjanmaalla kuin Pohjois-Savossa [t(298) = 16.48, p < 0.001]. V₁:n kesto on Pohjois-Savossa 92 ms ja Etelä-Pohjanmaalla 127 ms. C₂:n kesto on varieteeteissa tilastollisesti sama: Pohjois-Savossa 65 ms ja Etelä-Pohjanmaalla 66 ms. V₂:n kesto on pitempi Pohjois-Savossa kuin Etelä-Pohjanmaalla [t(298) = 17.19, p < 0.001]. V₂:n kesto on Pohjois-Savossa 141 ms ja Etelä-Pohjanmaalla 103 ms. V₂:n keston prosentuaalinen osuus V₁:n kestoista on varieteeteissa merkitsevästi erilainen [t(197.66) = 20.69, p < 0.001]. V₂:n kesto on Pohjois-Savossa 158,4 % ja Etelä-Pohjanmaalla 82,1 % V₁:n kestoista; Pohjois-Savossa näin ollen tapahtuu vahva puolipidennys, mutta Etelä-Pohjanmaalta puolipidennysilmiö puuttuu. Tämä tulos oli aiempien havaintojen perusteella odotuksenmukainen. Segmenttikestojen analyysit osoittivat myös, että toisin kuin Oulun, Tampereen ja Turun varieteettien välillä, Pohjois-Savon ja Etelä-Pohjanmaan varieteettien välillä ei ole CV.CV-sanojen konsonanttien kestoeroja. Kuvasta 1 näkyy, että varieteettien segmenttikestot ovatkin keskenään lähes samanlaiset siltä erotuksella, että pitempi vokaalisegmentti sijaitsee Pohjois-Savossa toisessa tavussa ja Etelä-Pohjanmaalla ensimmäisessä tavussa, lyhempi Pohjois-Savossa ensimmäisessä ja Etelä-Pohjanmaalla toisessa tavussa.

Ennen F₀-analyysien tekemistä aineiston F₀-arvot normalisoitiin. Tämä tapahtui vähentämällä kustakin mitatusta F₀-arvosta kyseistä kohdesanaa edeltävän tavun alusta mitattu

F₀-arvo ja lisäämällä erotukseen koko aineiston kohdesanoja edeltävien tavujen alun F₀:n keskiarvo. Kohdesanaa edeltävän tavun lopussa varieteettien välillä on melkein merkitsevä perustaajuusero [t(292) = 1.90, p = 0.059]: F₀ on Pohjois-Savossa 195,9 Hz ja Etelä-Pohjanmaalla 193,8 Hz. Kohdesanan ensimmäisen tavun ensimmäisessä, toisessa, kolmannessa ja neljännessä mittauskohdassa varieteettien välillä ei ole tilastollista perustaajuuseroa. Pohjois-Savossa näiden mittauskohtien keskimääräiset F₀-arvot ovat ensimmäisestä neljanteen kohtaan lueteltuina 200,0, 210,9, 237,3 ja 263,5 Hz, Etelä-Pohjanmaalla vastaavat arvot ovat 199,0, 213,8, 242,4 ja 260,9 Hz.

Kohdesanan ensimmäisen tavun lopussa eli sen viidennessä mittauskohdassa varieteettien välillä on merkitsevä perustaajuusero [t(247.40) = 3.53, p < 0.001]. F₀ on Pohjois-Savossa 276,8 Hz ja Etelä-Pohjanmaalla 259,7 Hz. Toisen tavun ensimmäisessä mittauskohdassa F₀ on Pohjois-Savossa korkeampi kuin Etelä-Pohjanmaalla [t(226.26) = 5.92, p < 0.001]. F₀ on Pohjois-Savossa 278,1 Hz, Etelä-Pohjanmaalla 251,1 Hz. Myös toisen tavun toisessa mittauskohdassa F₀ on Pohjois-Savossa korkeampi kuin Etelä-Pohjanmaalla [t(235.35) = 8.83, p < 0.001]: F₀ on Pohjois-Savossa 270,6 Hz ja Etelä-Pohjanmaalla 232,0 Hz. Samoin toisen tavun kolmannessa mittauskohdassa F₀ on Pohjois-Savossa korkeampi kuin Etelä-Pohjanmaalla [t(277.64) = 7.60, p < 0.001], F₀ on Pohjois-Savossa 230,8 Hz ja Etelä-Pohjanmaalla 202,5 Hz. Edelleen toisen tavun neljännessä mittauskohdassa F₀ on Pohjois-Savossa korkeampi kuin Etelä-Pohjanmaalla [t(281) = 2.99, p < 0.01]. Tässä mittauskohdassa F₀ on Pohjois-Savossa 198,4 Hz, Etelä-Pohjanmaalla 188,9 Hz. Toisen tavun viidennessä mittauskohdassa varieteettien perustaajuudet eivät eroa toisistaan tilastollisesti, F₀ on Pohjois-Savossa 184,5 Hz ja Etelä-Pohjanmaalla 186,8 Hz. Kohdesanaa seuraavan tavun alussa F₀ on Etelä-Pohjanmaalla merkitsevästi korkeampi kuin Pohjois-Savossa [t(265) = 2.44, p < 0.05]: F₀ on Pohjois-Savossa 181,0 Hz ja Etelä-Pohjanmaalla 187,4 Hz. Kohdesanaa seuraavan tavun lopussa varieteettien perustaajuusarvot eivät eroa tilastollisesti toisistaan; F₀ on Pohjois-Savossa 177,4 Hz ja Etelä-Pohjanmaalla 181,7 Hz.

Perustaajuuden huippu on merkitsevästi korkeampi Pohjois-Savossa kuin Etelä-Pohjanmaalla [t(255.46) = 4.39, p < 0.001]. F₀:n huippuarvo on Pohjois-Savossa 289,7 Hz ja Etelä-Pohjanmaalla 266,8 Hz. Perustaajuushuipun sijainti kohdesanan alusta lukien on varieteeteissa merkitsevästi erilainen [t(268.99) = 4.30, p < 0.001]. Huippu sijaitsee Etelä-Pohjanmaalla 187 ms sanan alusta, Pohjois-Savossa 208 ms sanan alusta. Koska C₁:n ja V₁:n keskimääräinen yhteenlaskettu kesto on Etelä-Pohjanmaalla 220 ms ja Pohjois-Savossa 183 ms, F₀:n keskimääräinen huippukohta sijaitsee Pohjois-Savossa vasta sanan toisessa tavussa. F₀:n huipun sijainti suhteessa tavunrajaan on myös tilastollisesti merkitsevästi varieteeteissa erilainen [t(273.66) = 14.17, p < 0.001]. Huippu sijaitsee Pohjois-Savossa keskimäärin 24,6 ms tavunrajan jälkeen ja Etelä-Pohjanmaalla 33,3 ms ennen tavunrajaa.



Kuva 1: Aksentoitujen CV.CV-rakenteisten sanojen keskimääräiset F_0 -arvot ja segmenttikestot Pohjois-Savossa ja Etelä-Pohjanmaalla.

Kuvassa 1 näkyy myös F_0 :n kulku Pohjois-Savossa ja Etelä-Pohjanmaalla aksentoitujen CV.CV-sanojen aikana. Etelä-Pohjanmaalla F_0 :n huippu sijaitsee ensimmäisessä tavussa ja on selvästi matalampi kuin Pohjois-Savossa. Perustaajuushuippu sijaitsee Pohjois-Savossa melko lähellä C_2 :n puoltaväliä, kuten myös Turussa väitöskirjani lisäaineistossa [4, s. 121]. Väitöskirjani lisämateriaalin ja tämän tutkimuksen aineiston analysointi osoittaa, että Turun ja Pohjois-Savon varieteeteissa huipun sijainti on tilastollisesti sama, samoin huipun sijainti suhteessa tavunrajaan on näissä varieteeteissa tilastollisesti sama. Huippu sijaitsee Pohjois-Savossa keskimäärin 24,6 ms tavunrajan jälkeen ja Turussa keskimäärin 29,3 ms tavunrajan jälkeen. Tämä tulos on hieman yllättävä, koska CV.CV(X)-sanoissa vasta toisessa tavussa sijaitseva perustaajuushuippu on aiemmin havaittu [8, s. 215–224] ja väitöskirjassani tieteellisin menetelmin todettu Turun seudulla puhuttavassa suomessa, mutta muilla murrealueilla vastaavaa ilmiötä ei ole osattu edes epäillä olevan, ilmeisesti ainakin osittain siksi, että ilmiö on tullut mm. ”kiekaisun” nimellä leimallisesti tunnetuksi juuri lounaismurteisiin kuuluvana. Tämä tutkimus, jossa ensimmäisen kerran tarkastellaan aksentoitujen CV.CV-sanojen perustaajuusarvoja itämurteissa, osoittaa ilmiön esiintyvän myös ainakin Pohjois-Savossa. F_0 kuitenkin laskee huippukohdan jälkeen jyrkemmin Pohjois-Savossa kuin Turussa [vrt. 4, s. 121], ja jää tulevaisuudessa tutkittavaksi, missä F_0 :n huippu Pohjois-Savossa sijaitsee CV.CV-sanoissa, joiden C_2 on soinniton. Turussa tällaisissa sanoissa F_0 :n keskimääräinen huippu sijoittuu heti V_2 :n alkuun [4, s. 106–109, 119–120].

Väitöskirjani aineiston aksentoiduissa CV.CV-rakenteisissa sanoissa, joiden C_2 on soinniton, F_0 :n huippu sijaitsee Oulun varieteetissa 158 ms sanan alusta ja Tampereen varieteetissa 186 ms sanan alusta. Etelä-Pohjanmaalla tämän tutkimuksen aineistossa huippu sijaitsee 187 ms sanan alusta, ja huipun sijainti sanan alusta lukien on tilastollisesti sama Etelä-Pohjanmaalla ja väitöskirjani aineiston Tampereen varieteetin aksentoiduissa CV.CV-sanoissa. Oulussa huippu sijaitsee aiemmin sanassa [$F(2,266) = 7.41, p < 0.01$], mutta

Oulussa puhuttava suomi on puolipidennysvarieteetti, jossa ensi tavun kesto on lyhempi kuin Etelä-Pohjanmaan ja Tampereen puolipidennysksettömissä varieteeteissa [$F(2,267) = 18.17, p < 0.001$]. Ensimmäisen ja toisen tavun rajaan suhteutettuna F_0 :n huipun sijainti on sekä Oulussa, Etelä-Pohjanmaalla että Tampereella tilastollisesti sama. Huippu sijaitsee Oulussa keskimäärin 30,7 ms, Seinäjoella 33,3 ms ja Tampereella 25,2 ms ennen tavunrajaa.

4. Lopuksi

Toistaiseksi tutkituista varieteeteista saatujen tulosten perusteella näyttää siltä, että suomen murrealueilla on olemassa ainakin kaksi erilaista lausepainoa toteuttavan F_0 :n nousu-lasku-kuvion ajoitustapaa CV.CV-rakenteisten sanojen aikana. Toisessa näistä tavoista, joka on käytössä Etelä-Pohjanmaan, Oulun ja Tampereen varieteeteissa, perustaajuushuippu sijaitsee V_1 :n keskikohdan ja lopun välillä. Toisessa tavassa, jota käytetään Pohjois-Savossa ja Turussa, F_0 :n huippu sijaitsee soinnillisessa C_2 :ssa lähellä sen puoltaväliä. Jos C_2 on soinniton, huippu sijaitsee Turun varieteetissa vasta V_2 :n alussa, mutta Pohjois-Savosta CV.CV-sanoja, joiden C_2 on soinniton, ei ole vielä tutkittu. Ottaen huomioon F_0 :n huippukohdan jälkeisen jyrkän laskun Pohjois-Savossa voidaan pitää mahdollisena, että huippu sijoittuisi tällaisissa sanoissa V_2 :n, mutta asia täytyy tutkia empiirisesti. Aksentoitujen CV.CV-sanojen prosodiikan tutkimusta on edelleen mielekästä laajentaa myös kokonaan uusille murrealueille.

5. Lähteet

- [1] K. Suomi, J. Toivanen ja R. Ylitalo, “Durational and tonal correlates of accent in Finnish,” *Journal of Phonetics*, vol. 31, s. 113–138, 2003.
- [2] K. Suomi ja R. Ylitalo, “On durational correlates of word stress in Finnish,” *Journal of Phonetics*, vol. 32, s. 35–63, 2004.
- [3] M. Vainio ja J. Järvikivi, “Tonal features, intensity, and word order in the perception of prominence,” *Journal of Phonetics*, vol. 34, s. 319–342, 2006.
- [4] R. Ylitalo, *The realisation of prominence in three varieties of standard spoken Finnish*. Väitöskirja. Acta Universitatis Ouluensis B 88. <http://herkules.oulu.fi/isbn9789514291142>. Oulu: Oulun yliopisto, 2009.
- [5] K. Wiik, “Suomen murteiden vokaalien kestoista,” O. Aaltonen ja T. Hulkko (toim.), *Fonetikan päivät – Turku 1985. XIII Fonetikan päivillä Turun yliopistossa pidetyt esitelmät*, s. 253–317. Turku: Turun yliopisto, 1985.
- [6] K. Wiik ja I. Lehist, “Vowel Quantity in Finnish Disyllabic Words,” P. Ravila (toim.), *Congressus Secundus Internationalis Fenno-Ugristarum. Helsingiae habitus 23.–28. VIII. 1965. Pars I. Acta Linguistica*, s. 569–574. Helsinki: Suomalais-Ugrilainen Seura, 1968.
- [7] P. Boersma ja D. Weenink, *Praat: doing phonetics by computer* [tietokoneohjelma], versio 5.4.08, haettu 24.3.2015 osoitteesta <http://www.praat.org/>. 2015.
- [8] K. Wiik, “ F_0 :n huipun sijainti suomessa,” M. Karjalainen ja U. Laine (toim.), *Fonetikan päivät – Espoo 1988*, s. 215–229. Teknillinen Korkeakoulu, Akustikan julkaisusarja 31. Otaniemi: Teknillinen Korkeakoulu, 1988.

Rinnastuksen prosodiaa suomalaisessa viittomakielessä

Tommi Jantunen

Jyväskylän yliopisto, kielten laitos, PL 35, 40014 Jyväskylän yliopisto

tommi.j.jantunen@jyu.fi

Tiivistelmä

Artikkelissa käsitellään ei-manuaalisen prosodian ja syntaksin yhteispeliä suomalaisen viittomakielen lauserinnastusrakenteissa. Tietokonenäköpohjaisesti käsitellyn videoaineiston ja äidinkielisten viittojen tekemien kielipöllisuusarvioiden perusteella argumentoidaan, että tunnusmerkittömässä disjunktivisessa ('tai') ja kontrastiivisessa ('mutta') lauserinnastuksessa on varsin selkeät prosodiset erityispiirteet, kun taas additiivisessa ('ja') lauserinnastuksessa prosodia on vaikeammin yksilöitävissä. Konjunktioittomasti ja täten tunnusmerkisesti ilmenevän disjunktivisen ja kontrastiivisen rinnastuksen perusteella artikkeli tarjoaa lisäevidenssiä väitteelle, jonka mukaan manuaalisten merkistinten poissaolo kasvattaa ei-manuaalisen prosodian roolia lauserinnastuksessa.

Asiasanat: prosodia, ei-manuaalisuus, rinnastus, lause, tietokonenäkö, suomalainen viittomakieli

1. Johdanto

Tämän artikkelin aiheena on lauserinnastuksen prosodia suomalaisessa viittomakielessä. *Lauserinnastus* tarkoittaa kahden tai useamman samankaisen lauseen peräkkäistä yhteenliittämistä [1] (viittomakielessä lauseita voidaan rinnastaa myös samanaikaisesti, mutta tähän ei tässä artikkelissa puututa). *Prosodian* käsitteellä puolestaan viitataan viittomakielentutkimuksessa yleisesti siihen kielen ja kielenkäytön osaluokkaan, joka määrittää, miten viittoja sanoo sanottavansa [2]. Riippuen tarkasteluun otettavista artikulaattoreista, viittomakielen prosodia voidaan jakaa *manuaaliseen* eli käsien toiminnan kautta ilmenevään prosodiaan ja *ei-manuaaliseen* eli esimerkiksi kehon ja pään kerrosteisen toiminnan kautta ilmenevään prosodiaan [2,3]. Tässä artikkelissa keskitytään näistä jälkimmäiseen eli ei-manuaaliseen prosodiaan. Käytännössä huomio kohdistetaan siihen, miten kehon ja pään toimintaa käytetään merkitsemään lauserinnastuksen kannalta keskeisten syntaktisten yksiköiden aloja ja rajoja.

Artikkeli on teemaltaan osa laajempaa lauserinnastusta suomalaisessa viittomakielessä kartoittanutta tutkimusta [4]. Toistaiseksi lauserinnastuksen prosodiaa ei ole yhtä laajassa mielessä tarkasteltu missään muussa viittomakielessä, mutta yleisesti ottaen ei-manuaalisella prosodialla on argumentoitu olevan tärkeä rooli viittomakielten kompleksisten lauseiden muodostamisessa ja merkinnässä. Rinnastuksen kannalta keskeisiä havaintoja ovat tehneet muun muassa Davidson [5] sekä Tang ja Lau [6]. Näistä ensimmäinen on kuvannut, kuinka yhdysvaltalaisen viittomakielen disjunktivisessa eli *tai*-tyyppisessä rinnastuksessa rinnastettavat lauseet sijoitetaan kehon sivusuuntaisella liikkeellä hivenen eri kohtiin viittomailaa. Jälkimmäisessä tutkimuksessa on puolestaan esitetty, että rinnasteisten lauseiden raja on yhdysvaltalaisessa ja hongkongilaisessa viittomakielessä usein merkitty joko pään kestollisesti

pidennetyllä nyökkäysliikkeellä tai ylävartalon asennon muutoksella. Yleisemmin Tang ja Lau [6] ovat argumentoineet, että manuaalisen materiaalin (esim. konjunktioviittomien) poissaolo lisäisi ei-manuaalisen prosodian roolia viittomisessa.

Artikkeli lähestyy lauserinnastusta ja prosodiaa kahdesta näkökulmasta. Näistä toinen on funktionaalisen-typologisesti väärittyneen deskriptiivisen syntaksin näkökulma ja toinen viittomakielen fonetiikan näkökulma. Funktionaalisen-typologisen deskriptiivisen syntaksin on kielellisten ilmiöiden merkityslähtöistä, omaehtoista ja tarkkaan kuvaukseen pohjaavaa analyysia, jossa hyödynnetään laajan kielen vertailevan tutkimuksen tuloksia [7,8]. Viittomakielen fonetiikka tarkoittaa puolestaan viittomakielen tuottamisen, välittymisen ja havaitsemisen tarkastelua, kuvausta ja mittausta mahdollisimman matalalla abstraktiotasolla [9,10]. Alana viittomakielen fonetiikka on saanut nostetta etenkin viimeaikoina tapahtuneesta teknologisesti kehityksestä, jonka myötä esimerkiksi videolle taltioituja artikulaattoreiden liikkeitä on voitu alkaa mitata hyvin tarkkaan. Tietokonenäköpohjaisesti toteutettu videoaineiston analyysi on keskeisessä roolissa myös tässä tutkimuksessa.

2. Aineisto ja menetelmä

Tutkimuksen perusaineiston muodostaa laajemmasta korpuksesta [11–13] erotettu monikuvakulmainen, pääosin HD-laatuinen (1920x1080, 25–50 fps), jatkuva viittomista sisältävä videomateriaali, jonka kokonaiskesto on noin 45 minuuttia. Aineisto sisältää tuotoksia yhteensä viideltä äidinkieliseltä, iältään noin 20–60 vuotiaalta suomalaiselta viittojalta, jotka suorittavat pareittain erilaisia kerronta- ja keskustelutehtäviä (mm. sarjakuvista kertominen, keskustelut harrastuksista, kielitaidusta ja opinnoista). Aineisto on annotoitu kielipöllisen ja prosodisen toiminnan osalta *Max Planck* -instituutissa Hollannissa kehitetyllä ELAN-ohjelmalla [14], joka mahdollistaa sekä annotaatiomerkintöjen hyvin tarkan ajallisen kytkemisen videolla näkyvään toimintaan että erilaisten sekundaariaineistojen liittämisen primaariaineistoon. Kielipöllisen perusannotaatio kattaa viittomien, lauseiden ja virkkeiden identifioinnin ja rakenneanalyysin. Prosodinen annotaatio tarkoittaa pään ja kehon liikutyyppien tokenisointia (esim. pään nyökkäys ja kääntö, kehon erilaiset kallistuliikkeet).

Aineiston erityispiirre on sen tietokonenäköpohjainen, SLMotion-ohjelmalla [15] toteutettu pään liikkumisen jatkoanalyysi, joka on tehty noin 40 minuutilla videomateriaalia. Käytännössä tämä jatkoanalyysi tarkoittaa, että jokaisesta videoruudusta, joka on tallennettu kutakin viittojaa kohtisuorasti kuvanneella kameralla, on automaattisesti laskettu pään asento kolmessa ulottuvuudessa. Nämä ovat pään kääntöliikkeitä kuvaava *yaw*-ulottuvuus, pään nyökkäysliikkeitä kuvaava *pitch*-ulottuvuus ja pään kallistelu- eli *roll*-ulottuvuus [16]. Ruutukohtainen pääasentodata on tuotu sekundaariaineistona ELANiin, jossa se on visualisoitu ajan funktiona.

Työnkulultaan varsinainen tutkimusosuus oli seuraava: Annotoitu ja tietokonenäköpohjaisesti käsitelty videoaineisto läpikäytiin ELANissa ensin useamman kerran ja useamman tutkijan voimin. Aineistosta identifioitiin erilaisia rinnastuspauksia, joista koottiin lopulta viidenkymmenen rinnastusesimerkin edustava otos. Näistä esimerkeistä keskusteltiin äidinkielisten viittojen kanssa, ja heiltä saadut palautteet dokumentoitiin. Esimerkit analysoitiin laadullisesti.

3. Rinnastuksen kielioppi suomalaisessa viittomakielessä ja tutkimuskysymykset

Kolme kielitypologisesti keskeisintä lauserinnastustyyppiä ovat merkityksen näkökulmasta *additiivi* eli *ja*-tyyppinen rinnastus, *kontrastiivi* eli *mutta*-tyyppinen rinnastus ja *disjunktii* eli *tai*-tyyppinen rinnastus [1,17,18]. Kielet koodaavat näitä merkitystyyppisiä erilaisilla rakenteilla, joissa yksi vaihtelua aiheuttava parametri on *konjunktio* elementin eli linkkerin mukana- tai poissaolo. Suomalaisessa viittomakielessä additiivinen lauserinnastus on perustapauksessa konjunktio- eli rinnastettavat lauseet tuotetaan *jukstaposition* rakenteella, käytännössä peräkkäin. Kontrastiivisessa ja disjunktii- vassa lauserinnastuksessa suomalainen viittomakieli sen sijaan perustapauksessa hyödyntää konjunktioita. Kontrastiivisessa lauserinnastuksessa käytetty konjunktio on viittoma MUTTA, disjunktii- vassa rinnastuksessa viittoma TAI (ks. kuvio 1). Kuten suomessa, myös suomalaisessa viittomakielessä konjunktio analysoituvat osaksi jälkimmäistä lausetta [4].

Suomalaisen viittomakielen rinnastusrakenteissa on konjunktioiden käytön suhteen kuitenkin variaatiota ja perustapauksen tunnusmerkittömille rakenteille on olemassa myös tunnusmerkkiset vaihtoehdot. Kontrastiivisen ja disjunktii- vinnastuksen osalta tunnusmerkkinen vaihtoehto on olla käyttämättä konjunktioviittomia eli hyödyntää jukstapositiona- kennetta. Additiivisen rinnastuksen osalta tunnusmerkkinen vaihtoehto on puolestaan käyttää konjunktioita. Jos additiivinen rinnastus tuotetaan tunnusmerkkisesti, niin konjunktio toimii yleensä viittoma MYÖS tai PLUS (ks. kuvio 1). Suomalaisessa viittomakielessä on olemassa myös itsenäinen viittoma, jonka merkitys on yksiselitteisesti 'ja'. Tämä viittoma kuitenkin kuuluu hyvin lähellä suomea olevaan rekisteriin, eikä sitä käytännössä äidinkielisten viittojen arkikielessä käytetä.

Taulukko 1. Lauserinnastuksen päätyypit suomalaisessa viittomakielessä. Jukstaposition käyttö kaikkien semanttisten rinnastustyyppien koodaukseen antaa aiheen olettaa, että ei-manuaalisella prosodialla on tärkeä rooli lauserinnastuksessa.

Semanttinen tyyppi	Merkitön rakenne	Merkkinen rakenne
Additiivi ('ja')	Jukstapositio [A] [B]	Konjunktio [A] [MYÖS/PLUS B]
Kontrastiivi ('mutta')	Konjunktio [A] [MUTTA B]	Jukstapositio [A] [B]
Disjunktii ('tai')	Konjunktio [A] [TAI B]	Jukstapositio [A] [B]

Tunnuksmerkittömien ja -merkisten rakenteiden olemasolosta aiheutuu suomalaisen viittomakielen lauserinnastukseen kieliopillinen paradoksi: jukstaposition rakenteella voidaan



Kuvio 1: Suomalaisen viittomakielen konjunktioina käytettyjä viittomia. Ylärivillä (vasemmalla oikealle) viittomat MYÖS ja PLUS. Alarivillä (vasemmalla oikealle) viittomat MUTTA ja TAI. Kuvat [19, 20].

ilmaista kaikki kolme semanttista lauserinnastustyyppiä (ks. taulukko 1). Puhtaasti kielen ymmärtämisen kannalta tilanne on epäedullinen, sillä pelkkä merkityspohjainen rinnasteisuuden tulkinta kuormittaa turhaan vastaanottajan kognitiivista kapasiteettia (miten lopulta olisi tulkittava vaikkapa virkkeen [[TYTTÖ NAURAA] [POIKA ITKEE]] täsmällinen merkitys?). Onkin oletettavaa, että kielen tuottoprosessissa viittojat pyrkivät kompensoimaan tilannetta kieliopin ulkopuolisilla keinoilla, ja varteenotettava hypoteesi on, että nämä keinot liittyvät ei-manuaaliseen prosodiaan [2,3,5,6].

Seuraavassa tarkastellaan lauserinnastuksen prosodiaa edellä kuvatun kieliopillisen paradoksin motivoimana. Täsmällisesti ottaen tarkoituksena on vastata kahteen kysymykseen: (i) Millainen on ei-manuaalinen prosodia lauserinnastuksen tunnusmerkittömissä rakenteissa? ja (ii) Miten prosodia muuttuu, kun tunnusmerkittömät konjunktiorakenteet (so. kontrastiivi ja disjunktii) tuotetaan tunnusmerkkisinä jukstaposition rakenteina? Kysymyksiä käsitellään alla omista luvuista (4 ja 5).

4. Tunnuksmerkittömien rinnastusrakenteiden prosodia

Ei-manuaalinen prosodia – tässä yhteydessä kehon ja pään toiminta, joka merkitsee kielen yksiköiden aloja ja rajoja – tunnusmerkittömissä lauserinnastusrakenteissa varioi suuresti niin viittojen kesken kuin yhden viittoajan eri tuotoksissa. Rinnastusesimerkkien tarkan laadullisen analyysin ja äidinkielisten viittojen kieli-intuition perusteella kullekin semanttiselle lauserinnastustyyppille on kuitenkin osoitettavissa tietyt prosodiset erityispiirteet. Selkeimmät nämä ovat disjunktii- vassa ja kontrastiivisessa lauserinnastuksessa (tässä järjestyksessä), hämärimmät additiivisessa lauserinnastuksessa. Semanttisten lauserinnastustyyppien prosodiset erityispiirteet on niputettu yhteen tunnusmerkittömien rakenteiden osalta taulukossa 2. Kutakin erityispiirrettä esitellään ja pohditaan tarkemmin luvuissa 4.1–4.3.

Taulukko 2. Yhteenvedo ei-manuaalisen prosodian erityispiirteistä suomalaisen viittomakielen tunnusmerkitömissä lauserinnastusrakenteissa.

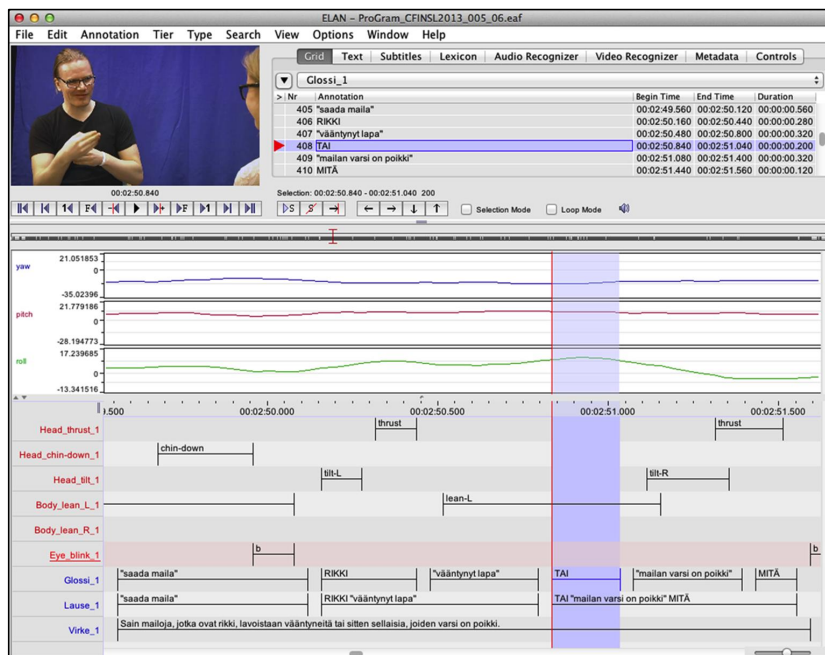
Semanttinen tyyppi	Merkitön rakenne	Prosodin erityispiirre
Additiivi ('ja')	Jukstapositio [A] [B]	Pään sivusuuntainen kallistusliike, joka sitoo lauseet kokonaisuuksiksi.
Kontrastiivi ('mutta')	Konjunktio [A] [MUTTA B]	Pään syvyyssuuntainen liike konjunktioviittoman MUTTA aikana.
Disjunktii ('tai')	Konjunktio [A] [TAI B]	Pää ja/tai keho liikuvat TAI-viittoman aikana sivusuunnassa viittomatiilan yhdestä paikasta toiseen paikkaan.

4.1. Disjunktii

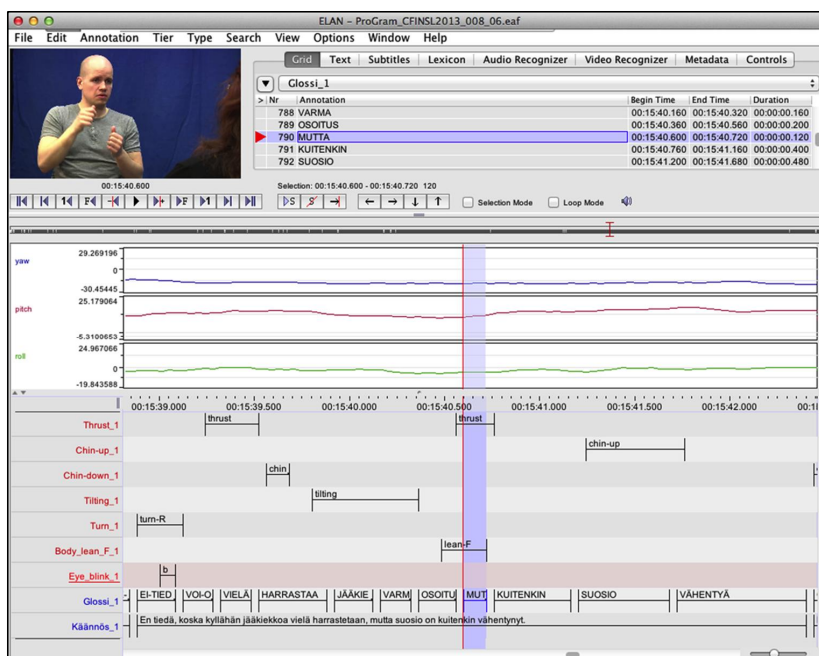
Ei-manuaalinen prosodia oli kaikkein yksiselitteisintä disjunktiiisessa lauserinnastuksessa. Sekä videoaineisto että äidinkielen viittojen mielipiteet osoittivat, että tunnusmerkittömässä disjunktiiisessa rinnastuksessa viittotjat pyrkivät tuottamaan rinnastettavat lauseet viittomatiilan eri paikoissa siten, että pää ja/tai keho liikkuvat lauserajalla – TAI-viittoman aika-

na – sivusuunnassa viittomatiilan paikasta toiseen. Käytännössä suomalaisen viittomakielen disjunktiiisen lauserinnastuksen prosodia on samankaltainen kuin yhdysvaltalaisen viittomakielen vastaavista rakenteista dokumentoitu prosodia [5]. Samanlaista kehon ja pään toimintaa käytetään myös puheen ohessa korostamaan sanallisesti ilmaistujen propositioiden vaihtoehtoisuutta.

Kuviossa 2 on esimerkki yhdestä videoaineiston disjunktiiisesta TAI-viittoman avulla toteutetusta lauserinnastuspauksesta. Esimerkissä rinnastettuna on kaksi kompleksista lausetta, '(Sain mailoja, jotka olivat) rikki siten, että lupa oli vääntynyt' ja 'mailojen varsi oli poikki' (kaksinkertaisten lainausmerkkien sisään kirjoitetut glossit viittaavat kuvioissa niin kutsuttuihin kuvaileviin viittomiin, jotka kääntyvät suomeen usein kokonaisilla lauseilla). Prosodin annotaatio osoittaa, kuinka viittojan keho alkaa kallistua ensimmäisen lauseen lopussa vasemmalle (rivillä *Body lean_L_1* oleva annotaatiolu *lean-L*) ja kuinka keho on palautunut perusasentoon seuraavan lauseen alkuun mennessä (yksi kehon liikkeiden annotaatiolu kattaa tässä kehon liikkeen sen perusasennosta takaisin perusasentoon). Kehon liikettä seuraa välittömästi oikealle suuntautuva pään kallistusliike (rivin *Head tilt_1* annotaatiolu *tilt-R*). Se, että kehon ja pään asennon sivuttainen muutos rytmittyy juuri lauserajan mukaan on havaittavissa vihreästä *roll*-kuvaajasta, johon muodostuu symmetrinen huippu viittoman TAI kohdalle (kaikki kuvaajat esittävät pään liikkumista, mutta koska pää on anatomisesti osa kehoa, kuvaaja heijastaa tässä myös kehon liikkumista). Viittojan todellinen asento punaisen (vertikaalisen) kohdistusviivan merkittävällä hetkellä näkyy vasemman yläkulman videorudussa.



Kuvio 2: Ruutukaappaus ELANista, jossa tarkastelussa kaksi disjunktiiisesti TAI-viittomalla rinnastettua lausetta. Tummenne osoittaa lauseiden juntuurin. Annotointiriveillä näkyy kielipiillistä (siniset rivit) ja prosodista annotaatiota (punaiset rivit). Kuvaajat esittävät pään liikkumista yaw- (sininen), pitch- (punainen) ja roll-ulottuvuudessa (vihreä kuvaaja).



Kuvio 3: ELAN-ruutukaappaus kontrastiivisesta MUTTA-viittoman avulla toteutetusta lauserinnastuksesta. Tummennus osoittaa jälleen lauseiden rajan, jonka kohdalla toteutuva pään (ja kehon) eteenpäin suuntautuva työntöliike (thrust) näkyy punaisen pitch-käyrän muodossa.

4.2. Kontrastiivi

Ei-manuaalinen prosodia oli varsin selkeää myös kontrastiivisen lauserinnastuksen tunnusmerkittömissä rakenteissa. Tässä rinnastustyyppissä viittojan pää pyrki liikkumaan konjunktioviittoman MUTTA aikana syvyyssulottuvuudessa suorittaen joko eteenpäin suuntautuvan työntöliikkeen tai taaksepäin suuntautuvan vetoliikkeen; liike saattoi toteutua syvyyssakselilla myös nyökkäysmäisesti. Kehon ja pään anatomis-fysiologisen yhteyden perusteella keho usein myötäili pään liikettä, mutta kieltenoppaiden arvioiden perusteella kehon liike oli näissä tapauksissa pelkästään pään liikettä vahvistava elementti.

Kuviossa 3 on annettu yksi esimerkki MUTTA-viittoman avulla toteutetusta kontrastiivisesta lauserinnastuksesta. Esimerkin prosodin annotaatio osoittaa, miten pää ja keho työntyvät lauserajalla eteenpäin. Tämä syvyyssuuntainen liike on taltioitunut myös pään pitch-ulottuvuudessa toteutuvaa liikumista kuvaavaan (punaiseen) kuvaajaan, joka suuntautuu viittoman MUTTA aikana ylöspäin. Kuvaajan perusteella pään työntöliike on esimerkissä varsin lievä, mutta silti havaittava. Videoruutu kertoo viittojan asennon työntöliikkeen alussa.

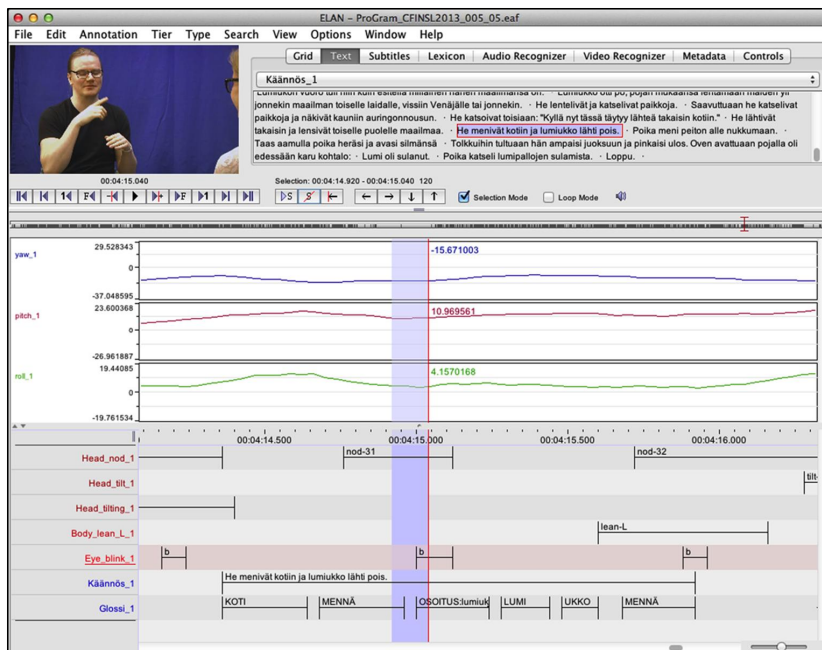
Huomionarvoista on, että pään liike kontrastiivisessa lauserinnastuksessa ei ole tulkittavissa viittoman MUTTA rakenteelliseksi ominaisuudeksi: viittoma MUTTA tuotettiin aineistossa myös ilman pään (ja kehon) liikettä, ja toisaalta tapauksissa, joissa viittomaa MUTTA ei käytetty (ks. luku 5), pää ja keho liikkuvat yhä samantyyppisesti. Toisaalta voi olla, että viittoman MUTTA kanssa esiintyvä pään liike ei ole myöskään puhtaasti prosodin liike: etenkin pään vetoliike signaloi osit-

tain myös semanttista eksluusiota [21], ja kontrastiivisessa lauserinnastuksessa sen voidaan tulkita ilmaisevan myös kontrastin luomiselle tärkeää näkökulman vaihtoa.

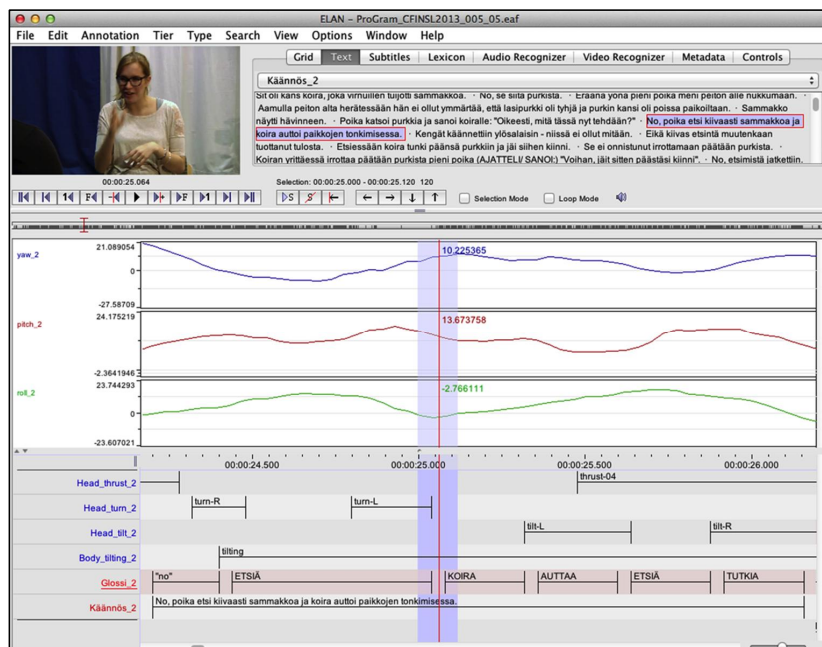
4.3. Additiivi

Rinnastustyypeistä additiivi oli ei-manuaaliselta prosodialtaan kaikkein vaihtelevin ja moninaisin luokka. Äidinkielisten viittojen konsultaatiossa tunnusmerkittömiä additiivirakenteita ei mielletty millään erityisellä ei-manuaalisella prosodialla merkityksi, vaan prosodian ajateltiin pikemminkin olevan riippuvaista kontekstista. Tietokonenäköpohjaisesti analysoidun videoaineiston tarkastelu kuitenkin paljasti, että tunnusmerkittömiä additiivirakenteita on löydettävissä myös systemaattisempia prosodisia tendenssejä: useiden additiivirakenteiden osalta oli havaittavissa, että rinnastettavat lauseet pyrittiin sitomaan prosodisiksi kokonaisuuksiksi ja kytkemään toisiinsa pään kallistusliikkeellä. Tätä kallistusliikettä havainnollistavat vihreät roll-kuvaajat kuvioissa 4 ja 5.

Käytännössä kuvioiden 4 ja 5 havainnollistama pään liike on neutraalisennosta sivulle ja takaisin toteutuva symmetrinen kallistus. Kuviossa 4 tämä liike tuotetaan suhteessa selkeästi isompana ensimmäisen lauseen aikana, kun taas kuviossa 5 pään kallistuminen ja paluu neutraalisentoon ovat varsin vahvoja ja samalajaisia molemmissa lauseissa (liikkeen vahvuudesta on osoituksena myös se, että pään yleinen liikkuminen on kuviossa 5 jaettu prosodisella annotaatiolla itsenäisiin liiketyppeihin). Huomionarvoista on, että kallistusliike merkitsee hyvin täsmällisesti rinnastettujen lauseiden aloja, ja li-



Kuvio 4: Tunnusmerkitöntä additiivista lauserinnastusrakennetta havainnollistava ELAN-ruutukaappaus. Vihreä roll-kuvaaja osoittaa, kuinka pään kallistusliike sitoo peräkkäiset lauseet prosodisiksi kokonaisuuksiksi.



Kuvio 5: Toinen ELAN-ruutukaappaus, joka havainnollistaa pään kallistusliikkeen roolia additiivisessa lauserinnastuksessa (vihreä roll-kuvaaja).

säksi se, että pää ei missään vaiheessa pysähdy lauseiden välillä. Esimerkit havainnollistavat myös, kuinka lauseisiin sisältyy paljon muutakin ei-manuaalista toimintaa: esimerkiksi kuviossa 4 lauserajalle osuu muun muassa prosodinen nyökkäys, vaikka yleisesti nyökkäystä ei additiivisessa lauserinnastuksessa junktuurin merkintään käytettykään (vrt. yhdysvaltalainen ja hongkongilainen viittomakieli [5]).

Kuvioissa 4 ja 5 havainnollistettua pään kallistusliikettä ei löydetty tunnusmerkittömistä kontrastiivi- ja disjunktiiivirakenteista (todennäköisesti näihin rakenteisiin liittyvän muun prosodisen toiminnan vuoksi). Tämä ei kuitenkaan tarkoita, että kallisteluliike olisi lähtökohtaisesti additiivisen lauserinnastuksen ominaispiirre. Todennäköistä on, että kyseistä ei-manuaalista prosodista toimintaa käytetään viittomakielissä yleisemminkin lauseita ja niiden osasia yhteensitovana koesiivisena keinona [3].

Additiivisten jukstapositionirakenteiden osalta on aiheellista kysyä, miten ne on mahdollista erottaa tapauksista, joissa tuotetaan peräkkäin kaksi erillistä, syntaktisesti yhteenkuulumatonta lausetta? Vastaus on, että käytännössä tätä eroa ei voida suomalaisessa viittomakielessä koskaan täydellisesti diagnosoida. Manuaalinen prosodia antaa tulkintaan kuitenkin vihjeitä: jos kaksi lausetta tuotetaan suhteellisen nopeasti peräkkäin ilman merkittävää prosodista katkosta (kuten esimerkeissä 4 ja 5), on todennäköistä, että lauseet on rinnastettu. Jos lauseiden välissä taas on tauko, niin todennäköisyys lauseiden syntaktiselle irrallisuudelle kasvaa. Manuaalisen prosodian lisäksi lauseiden rinnasteiselle statukselle on olemassa myös erilaisia kielipillisiä indikaattoreita [6].

Ei-manuaalisen prosodian tarkastelu tunnusmerkittömissä additiivirakenteissa oli haasteellista sikäli, että ei-manuaalisen prosodian ohella viittojat tuottivat additiivirakenteissa keholla ja päällä myös suhteellisen paljon kielipilliseksi mielletävää informaatiota. Tyypiesimerkki aineiston sisältämästä ei-manuaalisesta kielipillisestä toiminnasta oli niin kutsuttu *konstruoitu toiminta* eli konventionaalistunutta pantomiimia hyödyntävä esiintyminen, jota käytetään viittomakielissä muun muassa suoran ja epäsuoran kerronnan erottamiseen. Metodinen lähtökohta kehon ja pään toiminnan analyysissä olikin, että ei-manuaalista prosodista toimintaa ei voida koskaan täysin erottaa ei-manuaalisesta kielipillisestä toiminnasta [21].

5. Tunnusmerkkisten kontrastiivi- ja disjunktiiivirakenteiden prosodia

Toinen tässä artikkelissa tarkasteltava tutkimuskysymys liittyi prosodiseen eroon tunnusmerkittömien ja -merkkisten rinnastusrakenteiden välillä, ja oli seuraava: Miten prosodia muuttuu, kun tunnusmerkittömät konjunktiorakenteet (so. kontrastiivi- ja disjunktiiivirakenne) tuotetaan tunnusmerkkisinä jukstapositionirakenteina? Tangin ja Laun [6] yhdysvaltalaisesta ja hongkongilaisesta viittomakielestä esittämän argumentin nojalla oletus oli, että *konjunktioellipsi* lisää ei-manuaalisen prosodian roolia viitonnassa (ks. luku 1). Tässä tutkimuksessa käytetty videoaineisto ja natiiviviittojien tuotokset osoittivat, että juuri näin tapahtui tunnusmerkkisten jukstapositionirakenteiden kohdalla. Taulukko 3 vetää yhteen ei-manuaalisen prosodian erityispiirteet merkkisissä kontrastiivi- ja disjunktiiivirakenteissa.

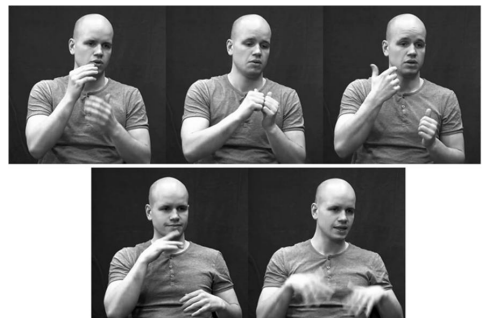
Ilman MUTTA-viittomaa tuotetuissa kontrastiivisissa lauserinnastusrakenteissa havaittiin normaalin syvyysuuntaisen

pään liikkeen korostumista ja tapauskohtaisesti myös kehon liikkeen entistä prominentimpaa mukaantuloa. Tätä on havain-

Taulukko 3. *Yhteenveto ei-manuaalisen prosodian erityispiirteistä suomalaisen viittomakielen tunnusmerkkisissä kontrastiivisissa ja disjunktiiivisissa lauserinnastusrakenteissa.*

Semanttinen tyyppi	Merkkinen rakenne	Prosodinen erityispiirre
Kontrastiivi ('mutta')	Jukstapositio [A] [B]	Normaalin pään liikkeen korostaminen, kehon liikkeen (prominentinpi) mukaantulo.
Disjunktiiivi ('tai')	Jukstapositio [A] [B]	Normaalin kehon ja/tai pään liikkeen jopa liioittelu korostaminen.

nollistettu kahdella videoaineistosta poimitulla esimerkillä kuviossa 6. Kuvion 6 ylärivin ruudut esittävät viittomia KYLLÄ, MUTTA ja NYT siten, kun ne ilmenevät lauseiden rajalla virkkeessä 'Se auttoi kyllä, mutta nyt siihen ei ole enää aikaa.' Alarivin ruudut esittävät puolestaan viittomia MUKAVA ja VÄSYNYT siten, kun ne esiintyvät virkkeessä 'Pojalla oli ollut mukavaa, mutta nyt hän oli totaalisen väsynyt.' Kuten viittomata-son esityksestä käy ilmi, yläriivi edustaa tunnusmerkitöntä konjunktiorakennetta, kun taas alarivi edustaa tunnusmerkkistä jukstapositionirakennetta. Ylärivin viitonnassa on heikosti nähtävissä pieni pään eteenpäin suuntautuva työntöliike viittoman MUTTA aikana, mutta ei juuri muuta prosodisesti merkittävää toimintaa. Alarivin viitonnassa, jossa konjunktioita MUTTA ei ole, työntöliike on huomattavasti selkeämpi. Tämän lisäksi koko keho liikkuu toisen lauseen alussa etuviistoon; osaltaan kehon liike vahvistaa myös viittoman VÄSYNYT liikettä.



Kuvio 6: Tunnusmerkittömän (yläriivi) ja tunnusmerkkisen (alarivi) kontrastiivirakenteen prosodia. Tunnusmerkkisessä rakenteessa pään ja kehon toiminta on korostuneempaa.

Ilman konjunktioviittomaa TAI tuotettujen disjunktiiivisten lauserinnastusrakenteiden kohdalla havainto oli, että normaalia kehon ja/tai pään liikkeen sivusuuntaista liikettä pyrittiin korostamaan jopa liioitellusti. Tätä prosodista korostamista on havainnollistettu kuvion 7 esimerkissä, jossa viittoja tuottaa tunnusmerkkiselle lauserajalle sijoittuvat viittomat TUTTU ja

KUURO osana virkettä 'Asettiin vanhojen tuttujen luona tai menttiin kuurojen luokse.' Ensimmäisen lauseen aikana viittojan keho on taipunut suhteellisen voimakkaasti vasemmalle. Lauserajalla keho liikkuu nopeahkosti viittojan oikealle sivulle ja samalla viittoja työntyy myös hieman eteenpäin. Tämä eteenpäin suuntautuva pää ja kehon liike aiheutuu tässä tapauksessa ensisijaisesti siitä, että viittojan ei ole istuma-asennossa mahdollista korostaa toista lausetta pelkällä sivusuuntaisella kehon liikkeellä.



Kuvio 7: Tunnusmerkkisen disjunktiivirakenteen prosodiaa. Viittoja korostaa semanttisten propositionien vaihtoehtoisuutta suhteellisen laajalla kehon ja pään liikkeellä.

Ei-manuaalisen prosodian korostaminen tunnusmerkkisissä kontrastiivi- ja disjunktiivirakenteissa häivyttää lauserinnastuksen kielipiillisen paradoksin, joka syntyi siitä, että kaikki kolme semanttista rinnastustyyppiä on mahdollista ilmaista jukstapositionarakenteella (ks. taulukko 1). Kun ei-manuaalinen prosodia tehdään hyvin selkeäksi lauseiden tuottamisen tasolla, tulkintaongelmaa jukstapositionarakenteiden täsmällisistä merkityksistä ei synny.

6. Lopuksi

Tässä artikkelissa on käsitelty ei-manuaalisen prosodian ja syntaksin yhteistoimintaa suomalaisen viittomakielen rinnastetuissa lauseissa. Artikkelissa on osoitettu, että tunnusmerkittömällä disjunktiiivisella ('tai') ja kontrastiivisella ('mutta') lauserinnastuksella on suomalaisessa viittomakielessä varsin yksiselitteiset prosodiset hahmot: disjunktiiivisessä rinnastuksessa lauseet tuotetaan hivenen eri kohdissa viittomatilaa siten, että keho liikkuu lauserajalla sivusuunnassa viittomatilan yhdestä paikasta toiseen paikkaan, ja kontrastiivisessä rinnastuksessa pää tekee lauserajalla syvyysuuntaisen liikkeen. Additiivinen rinnastus ('ja') on prosodialtaan hämäämpi: artikkelissa on kuitenkin osoitettu, että pään sivusuuntaisella kallistusliikkeellä on tunnusmerkittömässä additiivisessä rinnastuksessa tärkeä rooli lauseiden yhdistäjänä. Konjunktioittomasti ja täten tunnusmerkkisesti ilmenevän disjunktiiivisen ja kontrastiivisen rinnastuksen perusteella artikkelissa on lisäksi tarjottu evidenssiä väitteelle, jonka mukaan manuaalisten merkitsinten poissaolo kasvattaa ei-manuaalisen prosodian roolia lauserinnastuksessa. Kokonaisuutena artikkelissa kuvattu tutkimus on osoittanut, että syntaksin ja ei-manuaalisen prosodian suhde ei ole suomalaisen viittomakielen lauserinnastuksessa sattumanvarainen, vaan lähtökohtaisesti systemaattinen.

7. Kiitokset

Kirjoittaja kiittää kollektiivisesti kaikkia tutkimuksessa mukana olleita henkilöitä ja tahoja. Artikkelissa kuvattu tutkimus on rahoitettu Suomen Akatemian projekteista 269089 ja 273408.

8. Lähteet

- [1] Haspelmath, M. 2007. Coordination. — T. Shopen (toim.), *Language typology and syntactic description: Complex constructions. Volume 2*, 1-51. Toinen painos. Cambridge: Cambridge University Press.
- [2] Sandler, W. 2012. Visual prosody. — R. Pfau, M. Steinbach & B. Woll (toim.), *Sign language: An international handbook*, 55-76. Berlin: Mouton De Gruyter.
- [3] Pfau, R. & Quer, J. 2010. Nonmanuals: Their grammatical and prosodic roles. — D. Brentari (toim.), *Sign languages: A Cambridge language survey*, 381-402. Cambridge: Cambridge University Press.
- [4] Jantunen, T. 2015. Clausal coordination in Finnish Sign Language. Julkaistavaksi tarjottu käsikirjoitus, tammikuu 2015.
- [5] Davidson, K. 2013. 'And' or 'or': General use coordination in ASL. *Semantics & Pragmatics* 6(4), 1-44.
- [6] Tang, G. & Lau, P. 2012. Coordination and subordination. — R. Pfau, M. Steinbach & B. Woll (toim.), *Sign language: An international handbook*, 340-365. Berlin: Mouton De Gruyter.
- [7] Dixon, R. M. W. 2010. *Basic Linguistic Theory. Vol. 1*, Methodology. Oxford: Oxford University Press.
- [8] Jantunen, T. 2009. *Tavu ja lause: tutkimuksia kahden sekventiaalisen perusyksikön olemuksesta suomalaisessa viittomakielessä*. Jyväskylä Studies in Humanities 117.
- [9] Crasborn, O. 2012. Phonetics. — R. Pfau, M. Steinbach & B. Woll (toim.), *Sign language: An international handbook*, 4-20. Berlin: Mouton De Gruyter.
- [10] Jantunen, T. 2011. Foneettisen viittoman ala. *Puhe ja kieli* 31:2, 51-65.
- [11] Jantunen, T., Burger, B., De Weerd, D., Seilola, I. & Wainio, T. 2012. Experiences collecting motion capture data on continuous signing. — *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions Between Corpus and Lexicon*, 75-82. Paris: ELRA.
- [12] Puupponen, A., Jantunen, T., Takkinen, R., Pippuri, O. & Wainio, T. 2014. Taking non-manuality into account in collecting and analyzing Finnish Sign Language video data. — *Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel*, 143-148. Paris: ELRA.
- [13] Pippuri, O., Salonen, J., Takkinen, R., Puupponen, A. & Jantunen, T. 2015. Annotation conventions in ProGram and CFINSL projects. Posterisitelmä DiG-työpajassa University College Londonissa Iso-Britanniassa 31.3.2015.
- [14] Crasborn, O. & Sletjes, H. 2008. Enhanced ELAN functionality for sign language corpora. — *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, 39-43. Paris: ELRA.
- [15] Karppa, M., Viitanen, V., Luzardo, M., Laaksonen, J. & Jantunen, T. 2014. SLMotion - An extensible sign language oriented video analysis tool. — *Proceedings of the 9th international conference on Language Resources and Evaluation (LREC 2014)*, 1886-1891. Paris: ELRA.
- [16] Luzardo, M., Karppa, M., Laaksonen, J. & Jantunen, T. 2013. Head pose estimation for Sign Language video. — J.-K. Kamarainen & M. Koskela (toim.), *Image Analysis*, 349-360. Springer.
- [17] Velupillai, V. 2012. *An introduction to linguistic typology*. Amsterdam: John Benjamins.
- [18] Gast, V. & H. Diessel 2012. The typology of clause linkage: Status quo, challenges, prospects. — V. Gast & H. Diessel (toim.), *Clause Linkage in cross-linguistic perspective: Data-driven approaches to cross-clausal syntax*, 1-36. Berlin: Mouton De Gruyter.
- [19] *Viittomakielen kuvasanakirja*. 1977. Helsinki: Kuurojen Liitto.
- [20] *Suomalaisen viittomakielen perussanakirja*. 1998. Helsinki: KL Support Oy.
- [21] Puupponen, A., Wainio, T., Burger, B. & Jantunen, T. 2014. Head movements in the dimension of depth in FinSL: nods, nodding, head thrusts and head pulls. Julkaistavaksi tarjottu käsikirjoitus, lokakuu 2014.

Emotionaalisen puheen intonaatiopiirteiden graafisen representaation mahdollisuuksia

Teija Waaramaa, Jarkko Niemi, Jari Eerola

Viestinnän, median ja teatterin yksikkö

Tampereen yliopisto, Tampere

teija.waaramaa@uta.fi

Abstrakti

Aiemmissa empiirisissä kansainvälisissä kuuntelukokeissa havaittiin, että tietyt näyttelijöiden tuottamat emootio-älyt tunnistettiin hyvin kuuntelijan kulttuuri- tai kielitaidasta riippumatta. Kuuntelukokeet oli tehty Aasiassa, Afrikassa, Euroopassa ja Pohjois-Amerikassa. Käsillä olevalla tutkimuksella haluttiin selvittää, miltä näyttävät graafisessa muodossa kuuntelukokeissa tunnistetut älyt ja miten ne eroavat niistä älyistä, joita oli vaikea tunnistaa. Näytteet analysoitiin Praat-ohjelmalla, jonka avulla älyistä piirrettiin intonaatio- ja intensiteettikuvaajat. Lisäksi mitattiin kesto. Tämän lisäksi näytteet nuotinnettiin kuulonvaraisesti ja kirjoitettiin nuottitranskriptio. Analysoitavia emootioita olivat ilo, suru, viha ja pelko sekä referenssinä käytetty neutraali tunnetila. Tulokseksi saatiin, että intonaatio (perustaajuuden, F0:n), intensiteetin (voimakkuuden, dB:n) ja keston vaihtelut tulee olla riittävän suuria sekä emootio-älyn sisällä että eri emootioiden välillä, jotta ne erottuisivat toisistaan ja olisivat kuulonvaraisesti havaittavissa. Tämä oli toteutunut aiemmissa kuuntelukokeissa hyvin tunnistettujen emootio-älyiden kohdalla, mutta ei niiden näyttelijöiden kohdalla, joita oli ollut vaikea tunnistaa. Tuloksena havaittiin myös, että musiikillisen notaation avulla voidaan nostaa esiin joitakin puheen piirteitä, jotka havainnollistavat puhe-signaalissa tuotettuja emootioita ja niiden välisiä eroja. Musiikkia hyödynnettäessä sen ulkopuolisilla tieteenalogoilla, on kuitenkin syytä tarkistaa, kuten tässä tapauksessa, puheentutkimukseen ja musiikkiin liittyvä terminologia ja sen mahdolliset eroavuudet näillä tieteenalogoilla. Samoin on laita analyysissä käytettävien parametrien suhteen. Edelleen tulee varmistaa, että käsitteistö on kulttuurienvälisesti validia.

1. Johdanto

Tässä tutkimuksessa etsitään mahdollisuuksia kuvata emotionaalista puhetta graafisilla ilmaisukeinoilla sekä arvioidaan näitä kuvausmuotoja. Tutkimuksella halutaan selvittää, miltä näyttävät graafisessa muodossa sellaiset puhenäytteet, joiden sisältämät emootiot (tunteet) tunnistettiin hyvin aiemmissa empiirisissä kansainvälisissä kuuntelukokeissa ja vastaavasti, miltä näyttävät sellaiset puhenäytteet, joista emootioita oli vaikea tunnistaa.

Tutkimuskysymyksen esittämiseen ovat johdatelleet ajatukset puheen ja musiikin yhteisestä alkuperästä ja puhuttujen emootio-älyiden melodisuudesta. Fónagy (1981) on määritellyt vokaalisen ilmaisun melodisuuden seuraavalla tavalla: *“the perceptual response to the higher or lower degree of regularity/continuity/predictability of the fundamental frequency curve within each syllable”* [1].

Levitinin ja Tirovolasin (2010) mukaan ihminen muistaa yleisen melodiarakenteen paremmin kuin tarkat yksittäiset intervallit eri sävelten välillä [2]. Puheen prosodia onkin saattanut toimia mallina emotionaalille ja niin myös musiikilliseen ilmaisuun [3], [4], eikä ilman tätä adaptaatiota musiikkia olisi ehkä koskaan syntynyt ihmisen evoluutiossa [5]. Richmanin (2001) mukaan puhe ja musiikki olivat alun perin sama asia, ne olivat kollektiivisia, reaaliajassa tapahtuvia, kaavamaisia peräkkäisiä toistoja [6]. Thompson ym. (2004) ovat huomauttaneet, että ei ole todennäköistä, että ihmisen evoluutiossa olisi ollut syytä kehittyä kaksinkertaista mekanismia samalle asialle eli mekanismia äänenkorkeuden ja tempon vaihtelulle erikseen musiikissa ja erikseen emootioilmaisuissa [7]. Edelleen Juslin & Laukka (2003) ovat esittäneet, että musiikin emotionaalinen ilmaisevuus perustuu ihmisen ääni-ilmaisuun ja siihen liittyvien akustisten ominaisuuksien samankaltaisuuksiin [4]. Emotionaalinen musiikki ja puhe saattavat näin ollen kytkeytyä aivojen samoihin neuralaasiin prosessointeihin [8].

Perusemootioiden (viha, suru, ilo, pelko, inho ja hämmästyminen) [9] ajatellaan olevan universaaleja tunteita, mikä tarkoittaa sitä, että ne ilmaistaan ja vastaanotetaan samalla tavalla kaikkialla maailmassa. Perustunteiden ajatellaan syntyneen ja saaneen ilmaisumuotonsa osana ihmisen evoluutiota ja henkiinjäämistä [10]. Näiden tunteiden vokaalinen ilmaisu näyttää perustuvan ensisijain geneettiseen perimään ja toiseksi kulttuurisesti opittuihin ilmaisutapoihin [11]. Samalla tavalla musiikissa ilmaistut emootiot perustuvat sekä perittyihin ominaisuuksiin että kulttuuriseen oppimiseen [12].

Emootioiden tunnistamista mittaavissa kuuntelukokeissa on tyypillisesti sekoitettu keskenään ne tunteet, joiden valenssi on sama. (Valenssi tarkoittaa tunteen laatua akselilla positiivinen – neutraali – negatiivinen.) Vaikka esimerkiksi pelko on onnistettu tunnistamaan hyvin [13], [14], [15], se on silti sekoitettu usein toisen negatiivisen emootion, surun, kanssa [16], sillä näiden tunteiden akustisissa ominaisuuksissa on useita samankaltaisuuksia ja niiden molempien ilmaisutavat sisältävät paljon katkoksia äännessä (voice breaks). Tässä tulee esille negatiivisille emootioilmaisuille tyypillinen säännöttömyys (ks. myös [4]. Myös Kotlyar ja Morozov (1976) [17], ks. myös [18] ovat raportoineet pitkiä taukoja tavujen välillä ja vastaavasti lyhyitä tavujen kestoja pelko-ilmaisuissa.

Aiemmissa tutkimuksissa on havaittu, että positiivisten emootioiden ilmaisu sisältävät vähemmän katkoksia äännessä ja ne oli ilmaistu ennustettavammalla rytmisyydellä kuin negatiiviset emootiot [13], [19]. Ystävällinen puhetapa sekä esimerkiksi lapsille suunnattu puhe ovat niin ikään melodisempia ja rytmisempiä kuin negatiivinen puhe [1], [20]. Melodisuutta voitaisiin näin ollen pitää yhtenä äänenlaadun parametrina, joka välittää emootioita tai valenssia. Muita emootioilmaisuun liittyviä äänen parametreja on todettu olevan intensiteetti (SPL, sound pressure level), perusaanentaajuus (F0), formanttitaajuudet F3 ja F4, EGG

(electroglottography) ja äännön katkokset [13], [19], [21], [22], [23].

Yllä mainittu Fónagyn määritelmä melodisuudesta on puheen taajuusvaihteluiden kontekstissa varsin selkeäräjäinen. Sinänsä melodisuuden käsite viittaa kuitenkin myös musiikkiin, ja tämä määrittely voi olla globaalilla, kulttuurirelativistisella tasolla jo huomattavasti hankalampaa. Siihen on kuitenkin käsillä olevassa kokeilussa otettava kantaa, koska tässä esitellään ääni-ilmaisun graafisen representaation mahdollisuuksia myös musiikin merkitsemiseen käytetyn länsimaisen konventionaalisen notaatiosymboliikan avulla.

Länsieurooppalaisen taidemusiikki-instituution taustasta syntyneiden käsitteiden ja ymmärryksen on pitkälti katsottu olevan soveliaita koko maapallon musiikkikulttuurien ymmärtämiseen. Nykyäänkin saatetaan esimerkiksi puhua musiikin säveljärjestelmien yhteydessä ”duurista” ja ”mollista”, vaikka ne kuuluvat nimenomaan länsieurooppalaisen taidemusiikin kulttuurihistoriaan ja siinäkin vain muutaman viime vuosisadan ajalle. Tämä anakronismi vastaa sitä, että ajateltaisiin keinotekoisen esperanton kielen olevan indoeurooppalaisten kielten piirteitä selittävä tekijä. Esperantoa kuitenkin ei koskaan liene tarjottu jonkin indoeurooppalaisen kielimuodon selitykseksi, kun taas duuri-molli-ajattelulla on useinkin yritetty perustella esimerkiksi eurooppalaisia perinнемusiikin ilmiöitä.

Etnomusikologi John Blacking (1973) korostaa – eteläisen Afrikan heimopohjaisiin kyläkulttuureihin liittyvän kenttätutkimuksensa perusteella – musiikillisen toiminnan olevan ensisijaisesti sosiaalista vuorovaikutusta, ei niinkään pelkkää ääntä, kuten länsimaissa on tapana ajatella. Näin Blackingille musiikki on inhimillisesti ja siis sosiaalisesti organisoitua ääntä. [24, x–xi.] Tästä huolimatta paikalliskulttuureissa syntyneet musiikillisen ajattelun kulttuurihistoriat ovat aina peräisin jostain, niillä tuskin on omaa, kulttuurin tason ylittävää universaaliluonnetta. Kuten Blackingin [24, s. 25] toteaa: ”musiikin ominaispiirteet ovat musiikin tuottavan yhteisön ja kulttuurin ominaispiirteitä”. Näiden ominaispiirteiden keskinäisen suhteen osoittaminen ja tulkinta on etnografinen projekti, niille ei ole olemassa automaattista globaalia selitystä. Tämän tilanteen voidaan tosin ajatella muuttuvan sitä mukaa, mikäli paikalliskulttuurien omat äänet katoavat globaalina, standardisoidun eurooppalaislähtöisen mediakulttuurin tieltä. Musiikin määrittely jää kuitenkin aina varsin suhteelliseksi ja kulttuurisidonnaiseksi kysymykseksi. Tämä on hyvä pitää mielessä, kun pohditaan esimerkiksi puheen ja laulun (= musiikin) välisiä eroavaisuuksia.

2. Materiaali ja metodit

Aiemmin tallennetusta materiaalista [13], [19] valittiin käsillä olevaan tutkimukseen yhden nais- ja yhden miesnäyttelijän tuottamat emotionaaliset. Näyttelijät lukivat lingvistisesti merkitykselliset lauseet:

”Elki neiku kootsa, fonta tegoa viififfi askepanna äspa. Fissafii teeki staaku porkas talu.”

Lauseet ilmaistiin kahdeksassa eri tunnetilassa, joista tähän tutkimukseen valittiin neutraalin ilmaisuuden lisäksi neljä tunnetta, ilo, pelko, suru ja viha. Tutkimuksen materiaali oli tallennettu ammattistudiossa Sony Sound Forge 9.0 –äänitys-

ja editointiohjelmalla. Puhujien etäisyys mikrofonista (Røde NTK) oli 40 cm.

Naispuhujan ilmaisemat tunteet oli tunnistettu hyvin aiempaan tutkimukseen liittyvissä kansainvälisissä kuuntelukokeissa riippumatta kuuntelijan kulttuuri- tai kielitautasta (tunnistusprosentti oli n. 90 %). Sen sijaan miespuhujan positiivisiksi tunteiksi tarkoitetut näytteet olivat valtaosin jääneet tunnistamatta (tunnistusprosentti oli 20–30 %). Miespuhujan negatiivisiksi tarkoitetut tunnenäytteet tunnistettiin n. 70 %:n tarkkuudella. Kansainväliset kuuntelukokeet oli järjestetty neljässä eri maanosassa: Aasia/Kiina, Afrikka/Egypti, Eurooppa/Ruotsi, Suomi ja Viro ja Pohjois-Amerikka/Yhdysvallat. Kokeisiin osallistui yhteensä 350 kuuntelijaa (50 % naisia, 50 % miehiä) [19]. Käsillä olevassa tutkimuksessa haluttiin vertailla, millainen intonaatio, intensiteetti ja kesto on niissä näytteissä, jotka oli tunnistettu kansainvälisesti hyvin ja vastaavasti niissä, jotka oli ollut vaikea tunnistaa.

Näytteiden intonaatio, intensiteetti ja kesto analysoitiin Praat-ohjelmalla [25], ks. myös [26]. Alkuperäiset ääninäytteet oli tallennettu kaksikanavaisina (stereo). Ne muutettiin yksikanavaisiksi (mono) (Convert/Convert to mono). Tiedostoihin asetettiin puhutut tavut paikoilleen (Annotate/To Text Grid) Edit-ikkunassa. Tämän jälkeen muodostettiin analysoitava pitch-objekti (Analyse/Periodicity/To Pitch (ac)). Miespuhujan F0:ksi oli mitattu neutraalissa ilmaisussa 137 Hz ja naispuhujan 178 Hz (Taulukko 1, Kuvat 1 ja 2). Nämä taajuudet pujoitettiin pystyakselin puolisävelaskelikon 0-arvoiksi. Näytteiden intonaatioakäyrien transponoinnit tehtiin formula-toiminnolla (Modify/Formula) niin, että kuvaajien intervallisuhteet säilyivät ennallaan. Mukaan liitettiin myös intensiteetti-kuvaaja.

Puhenäytteet nuotinettiin Transcribe!- (<http://www.seventhstring.com>) ja NoteWriter-ohjelmilla (<https://www.opusonemusic.net/hamel.html>). Transcribe!-ohjelmaa käytettiin äänitiedoston kuunteluun. Puhenäytteiden intonaatioista tehtiin kuulonvarainen tulkinta, ja tuloksena syntyneet taajuusarvot merkittiin likimääräisinä intervallihavaintoina. Viimeistelyvaiheessa kuitenkin joitakin taajuuksia tarkistettiin koneellisesti. Varsinaiseen nuottitranskriptioon käytettiin NoteWriter-ohjelmaa, jonka keskeisin ominaisuus on ohjelmassa käytettävien symbolien asemoinnin vapaus.

Tässä kokeilussa puheen graafisessa transkriptiossa käytettiin siis konventionaalista länsimaista notaatiota [27, s. 110–152]. Yksi syy länsimaisen notaation käyttöön oli luottavuus: äänen spektri tuotiin koneellisen luennan jälkeen taas ihmiskorvan tulkittavaksi ja tämän perusteella tehtiin subjektiivinen ja kuulonvaraisesti määräytynyt nuottitranskriptio. Nuotinnossymboliikka yksinkertaistaa soivaa todellisuutta, mutta näin syntyneen uuden symbolisen abstraktiotason avulla on helpompi tutkia analysoitavan materiaalin sisäisiä rakennepiirteitä.

Nuottien asemoinnissa käytettiin näytteiden yhtäläisyyksien hahmottamista helpottavaa paradigmaattista vertikaalilinjausta [28], menetelmästä ja sen strukturalistisesta taustasta ks. [29]. Tämän asemoinnin tavoitteena oli jäsentää näytteiden taajuus- ja kesto-suhteet motiiviparadigmoiksi. Tässä viitataan paradigmaalla vain puhe- tai musiikkiesityksen (saussurelaisen parole-tason) äänellisiin (intonatiivisiin tai metrisiin) ja esityksellisiin perusrakenteen elementteihin, eikä esimerkiksi

proosamuotoisen, ei-metrisen kielenaineksen (langue-tason) kieliopillisiin paradigmoihin. Rakenteellisten motiiviparadigmojen hahmottaminen on mahdollista lähinnä vain metrisiin prominenssielementteihin kiinnittyneestä puhe- tai musiikki-ilmaisusta; ei-metrisestä puheesta motiiviparadigmoja on yleensä hedelmättömästi etsiä.

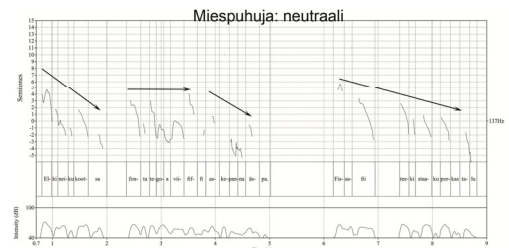
3. Tulokset

Naispuhujan näytteissä oli selvästi laajempi intonaation, intensiteetin ja keston vaihtelu kuin miespuhujan näytteissä. Näiden parametrien laaja variaatio emootioiden välillä vaikutti positiivisesti emootioiden tunnistamiseen aiemmissa kuuntelukokeissa.

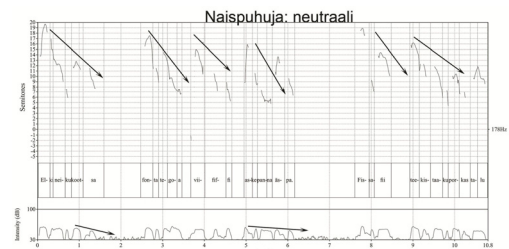
Taulukko 1: Nais- ja miespuhujan mitatut keskiarvot perusäänentaajuudesta (F0), perusäänentaajuuden minimistä ja maksimista, intensiteetistä (SPL, sound pressure level), intensiteetin minimistä ja maksimista sekä kehosta koko emootioilmaisusta mitattuina.

Sukupuoli	Emootio	F0 (Hz)	F0 min (Hz)	F0 max (Hz)	SPL (dB)	SPL min (dB)	SPL max (dB)	Kesto (s)
Nainen	Ilo	252	78	495	58	25	71	8,47
Nainen	Neutraali	178	94	258	51	23	61	10,76
Nainen	Pelko	203	95	266	46	19	61	12,53
Nainen	Suru	232	77	500	51	21	62	13,79
Nainen	Viha	218	81	496	69	17	82	8,2
Mies	Ilo	205	169	425	69	26	80	6,95
Mies	Neutraali	137	77	460	61	25	73	8,03
Mies	Pelko	174	98	499	62	26	72	8,89
Mies	Suru	158	75	492	60	24	70	9,19
Mies	Viha	154	125	403	66	26	76	8,03

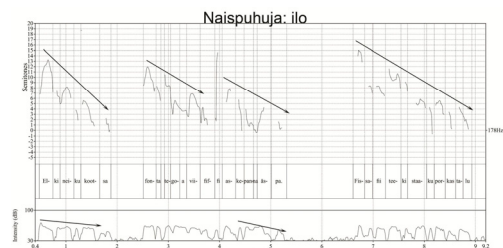
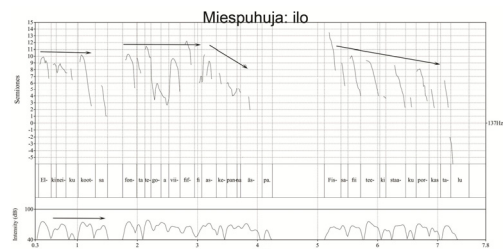
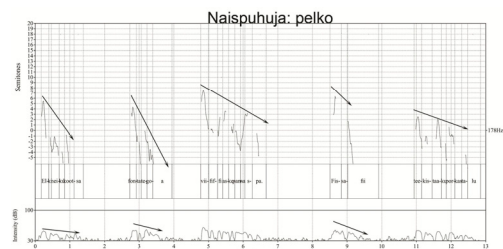
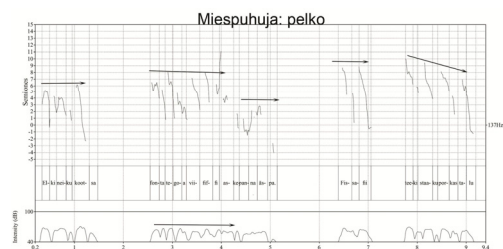
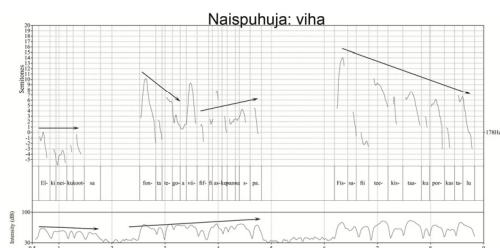
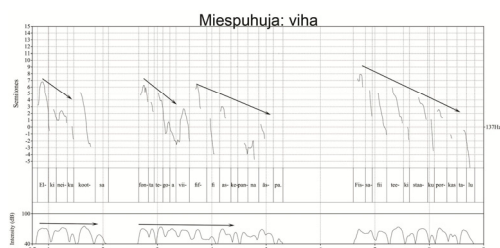
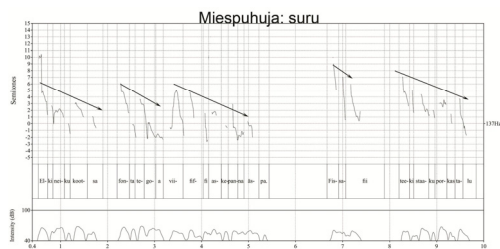
Suru-näytteiden kesto oli emootioista pisin molemmilla puhujilla. Miespuhujan surun intonaatio varioidi 2–5 puolisävelaskelen välillä, eikä intensiteetissä ollut suuria muutoksia (Kuva 3). Naispuhujan intonaatio varioidi vähiten surussa, kolmea puolisävelaskelta vastaavalla taajuusalueella ja se sijoittui tasaisesti noin 5 puolisävelaskelta F0:n yläpuolelle. Intensiteetissä on paljon vaihtelua. (Kuva 4.) Miespuhujan vihassa intensiteetti pysyi lähestulkoon samana alusta loppuun. Miespuhujan intonaatio oli samansuuntainen kaikissa emootionäytteissä ja muistutti siksi myös esimerkiksi suru-näytettä. (Kuva 5.) Naispuhujan vihaa ilmaisevassa emootiossa intensiteetin ja intonaation vaihtelu on suuri. Intensiteetti näytti kasvavan lauseen loppua kohden. Näyte jakautui kolmeen erilaiseen fraasiin, joista ensimmäisessä intensiteetti oli matalin ja viimeisessä fraasissa voimakkain. (Kuva 6.) Miespuhujalla pelkoa ilmaisevassa näytteessä intonaatio varioidi 10 puolisävelaskelteen saakka, mutta fraasien sisällä oli vain vähän vaihtelua (Kuva 7). Naispuhujan pelko-näytteen intonaatiossa oli havaittavissa jyrkkiä vaihteluita, yli oktaavia vastaavalla taajuusalueella. Näyte jakautui useisiin pieniin fraaseihin, joissa intensiteetti alkoi voimakkaasti ja laski jyrkästi (Kuva 8). Miespuhujan ilossa intonaatio ja intensiteetti vaihtelivat vain vähän. Huomattavaa tässä on ilon F0:n suhde puhujan neutraalin näytteen F0:aan: ilon F0 oli lähes kauttaaltaan oktaavin verran neutraalin F0:n yläpuolella. (Kuva 9.) Naispuhujan ilossa intonaatio hypähteli ylös ja laski sitten melko jyrkästi. Vaihtelua oli paljon, jopa 13 puolisävelaskelta. Intensiteetti seuraili intonaatiota ja oli suurimmillaan lauseen alussa (Kuva 10).



Kuva 1: Miespuhujan neutraali-näyte



Kuva 2: Naispuhujan neutraali-näyte



Nuottitranskriptio osoitti puhenäytteistä omalla abstraktiotasollaan ehkä joitakin itsestäänselvyyksiä, mutta myös joitakin tuoreita näkökulmia (Kuva 11). Kuten voitiin odottaa, äänen säveltaajuudet vaihtelivat puheintonaatiossa varsin ”epämusiikillisesti”, mutta molempien puhujien kaikissa emotioninäytteissä tavupituuksien peruskategoriat (lyhyt, pitkä) heijastuivat transkriptiossa ja nuottinotossymboliikassa yllättävän tarkasti.



Kuva 11: Nais- ja miespuhujan neljän emootio- sekä yhden neutraalinäytteen transkriptiot. Miesäänen karkeasti arvioitu perussävel on nuotin "e¹" (tässä 164,81 Hz) ja naisäänen "a¹" (tässä 220 Hz). (Huom., että kuvan 9. nuotinnokset on oktaavitransponoitu, jolloin ne tulee lukea oktaavin verran tavanomaista luenta matalammalta.) Näitä molemmat taajuusarviot ovat samansuuntaisia, kuin puhujilta koneellisissa analyysissä mitatut F0-tasot.

Perusemootioista esimerkiksi surun nuotinnosta luonnehti intonaation kapea-alaisuus, mikä heijastui nuotinnoksessa kapeana sävelalana, etenkin naispuhujalla. Esimerkiksi karjalaisista tai inkeriläisistä itkusävelmistä piirretyt nuottikuvat muistuttavat tätä sävelalansa puolesta (esim. [30]). Ilon intonaatio näytti puolestaan olevan laaja-alaisempi. Sitä ei niinkään voi kuvata ainakaan näiden näytteiden perusteella esimerkiksi "duurin" kielikuvaan, vaan pikemminkin suurten intervallihyppyjen (kvartti, kvintti) avulla.

4. Pohdinta

Suomalaisten ilmaisemat emootiot tunnistettiin puhenäytteistä kansainvälisesti sängen hyvin, mikä ei tue sitä stereotyyppistä käsitystä, että suomen kieli olisi monotonista, huonosti ilmaisevaa ja prosodisesti vähäeleistä. Kun intonaatio, intensiteetti ja kesto varioivat riittävän laajasti emootioiden välillä niin, että ne erottelevat emootioita toisistaan, myös kuuntelukokeissa emootioilmaisujen tunnistaminen ja erottelu on silloin mahdollista.

Tutkimuksessa käytettiin aiempien kuuntelukokeiden näytteitä kokonaisuudessaan, koska juuri nämä olivat ne signaalit, jotka kuuntelijat olivat kuulleet. Tästä syystä myös parametrien mittaukset tehtiin koko näytteistä. Mittausten ongelmaksi tuli näin ollen se, että Praat-ohjelma laski kaikki mitattavissa olevat arvot näytteen signaalista, muutkin kuin puhesignaalin arvot. Tällöin näytteen sekä intensiteetin pienin arvo että taajuuden korkein arvo saattoivat tulla mitatuiksi esimerkiksi sisäänhengityksestä. Näitä arvoja ei kuitenkaan otettu huomioon näytteiden nuotinnoksessa. Mitatut arvot ja nuotinnos eivät näin ollen vastaa täysin toisiaan, vaan nuotinnos kuvaa pelkästään puhesignaalin taajuuksia ja kestoja.

Ilon tunnistaminen on kuuntelukokeissa osoittautunut hankalammaksi kuin muiden emootioiden tunnistaminen [13], [16], [19], [31], [32]. Tämä voi johtua paitsi ilon-ilmaisuksen akustisesta rakenteesta niin myös siitä, että ilon tunnistamistarve ei ehkä ole liittynyt evoluutiossa ihmisen selviytymiseen ja henkinjäämiseen, toisin kuin esimerkiksi aggressioilmaisujen tunnistaminen. Ilon ja vihan [33] sekä myös inhon [34] havaitseminen saattaa olla muita perusemootioita vahvemmin sidoksissa visuaaliseen informaatioon. Tosin vihan ja inhon on todettu sekoittuvan keskenään myös visuaalisissa kokeissa [11]. Matsumoto onkin esittänyt, että näiden emootioiden semantiikka on varsin samanlainen, joten sama ärsyke voi herättää kumman tahansa näistä tunteista.

Vaikka tässä tutkimuksessa käytetyn kuuntelukokeen puhujat olivat ammattinäyttelijöitä, tilastollisesti merkitseviä eroja havaittiin puhujakohtaisissa emootiotunnistuksessa. Kun kokeessa on mukana yksi näytteenantaja, joka tuottaa selvästi helpommin tunnistettavia näytteitä kuin muut, koetulos saattaa vinoutua tämän johdosta. Samanlaisia tilanteita on raportoitu aiemminkin [35]. Tämä antoi kuitenkin mahdollisuuden tarkastella tässä tutkimuksessa, millaisia ovat ne emootioilmaisut, jotka pystytään tunnistamaan kuuntelijoiden kulttuuri- ja kielitautasta riippumatta.

Graafiset esitystavat havainnollistivat emootioiden variaatioita, mutta esitystapojen luettavuus ja analyttisen hyödyntämisen mahdollisuudet ovat erilaisia. Äänestä koneellisesti tuotetun spektrogrammin voidaan ajatella edustavan *deskriptiivistä*, todellisuuden mahdollisimman "tarkkaa" kuvaa tai raakadataa edustavaa transkriptiota, kun taas konventionaalinen länsimainen notaatio symbolisesti äärimmilleen suhteistettua *preskriptiivistä*, oman luentatapansa hallintaa vaativaa, ikään kuin jo esitulkittua transkriptiota [36]. Konventionaalista nuotinnosta käytettiin nostamaan esille emotionaalisen puheen tärkeimmät ominaisuudet vokaalisissa ilmaisuissa. Tämä edellytti kuitenkin nuotinnoksen suhteellisuuden ja preskriptiivisyyden huomioimista. Analysoitujen ilmaisujen motiivi- tai lauseriymien paradigmoja voitiin näin vertailla helpommin keskenään.

Nuottisymboliikka on joka tapauksessa (deskriptiivistä) todellisuutta abstraktimpaa ja yleistävämpää, mikä mahdollistaa kompleksisten ilmiöiden vertailun, kunhan voidaan kohtuullisesti perustella, miten abstrahointiprosessin aikana kadonneeseen informaatioon suhtaudutaan. Nuotinnoksen käyttö rakenneprosessin välineenä sen sijaan on tehokasta, koska nuottisymboleja voidaan helposti siirrellä ja ryhmitellä uudelleen, mikä puolestaan mahdollistaa motiivisten paradigmojen tunnistamisen ja vertailun.

Kuulonvaraisen säveltaajuusaistimuksen siirtäminen nuotinnossymboliikaksi on puheen intonaation tapauksessa kuitenkin yleisesti ottaen ongelmallista, koska puheintonaatio ei ole niin mallintunutta kuin se eurooppalainen musiikki-ilmaisu, jota kuvaamaan eurooppalainen notaatiokin on vuosisatojen kuluessa kehittynyt [vrt. 37, s. 153–164]. Ilmiö on kehämäinen: lisäksi eurooppalaisen musiikin, kuten muunkin kulttuurin kirjallistuminen kanonisoi tässä prosessissa vähitellen itsensäkin. Niinpä notaatiokielestään on vähitellen tullut myös osa eurooppalaisen "puhtaan" tai "oikeaoppisen" musiikillisen ilmaisun normistoa sekä lopulta esimerkiksi säveltäjän teoksen yksi olomuoto. Tästä johtuen konventionaalisen notaatiosymboliikan ongelmana on paradoksaalisesti yhtäältä sen karkeus: esimerkiksi portaattomasti muuntelevia taajuuksia on hankala kuvata. Tällainen taajuuden muuntelu näkyy spektrissä diagonaaleina

tai kaarevina mutkina. Notaatio sen sijaan kuvaa äntä pistemäisten, ääni-ilmiön taajuus- ja kesto-suhteita ilmentävien symbolien avulla. Tästä syntyy representaation ongelma: mikä käyrän mutkasta irrotettu piste kuvaa parhaiten tuota mutkaa? Toisaalta taas notaatiosymboliikka luo todellisuudesta taajuus- ja kesto-suhteitaan ”hyperkorrektina” näyttäytyvän ideaalikuva, mikä puolestaan heijastuu nuottinnoksen luetaan ja esittämiseen [29], [38, s. 16]. Länsimaisen notaation ongelmallisuus tulee hyvin esille monissa alkuperäiskansojen laulutyylilien tutkimuksissa, joissa tarkastellaan nuottisymboliikan edustaman mallintuneisuuden ja kulttuurisen esityksen ”realistisuuden” eroja [29], [39], [40], [41].

Etnomusikologit ovat jo 1900-luvun puolivälistä lähtien korostaneet musiikkikulttuurin kiinteää yhteyttä muuhun kulttuurin kokonaisuuteen. Esimerkiksi Alan Merriam toteaa musiikkilisten äänten olevan kulttuurinsa muokkaamia ja kulttuuri puolestaan on yksilöiden ja yksilöryhmien ylläpitämää. Näin inhimillinen musiikillinen toiminta on siis ensisijaisesti sosiaalista vuorovaikutusta, musiikkia ei voida tyhjentävästi määritellä pelkästään ääneksi tai esitykseksi. Toisaalta musiikkiliseksi kutsuttavien kulttuuristen ilmaisumuotojen universaalius lienee miltei ainoa maapallomme eri kulttuureja yhdistävä tekijä. [42, s. 26–28], ks. myös [43, s. 17–23].

5. Johtopäätökset

On ilmeistä, että puheen ja musiikillisen notaation välille ei voida rakentaa yksi-yhteen –vastaavuutta. Notaatiolla voidaan kuitenkin nostaa esiin joitakin puheen piirteitä, joilla voidaan havainnollistaa puheen sisältämien emootioilmaisujen välisiä eroja.

Mikäli musiikkia hyödynnetään monitieteisissä projekteissa tai sovellusympäristöissä, on tarpeen tarkistaa itse musiikkiin liittyvä ja kulttuurienvälisesti pätevä käsitteistö, jotta varmistetaan esimerkiksi analyysin parametrien validius.

6. Kiitokset

Tätä tutkimusta ovat tukenet Suomen Akatemia ja Suomen Kulttuurirahasto.

7. Lähteet

- [1] I. Fónagy. Emotions, voice and music. Teoksessa: Sundberg J. (toim.): Research Aspects on Singing. Royal Swedish Academy of Music 33, s. 51–79, 1981. Alkuperäisjulkaisu: I. Fónagy, K. Magdics. A beszéd dallama. Akadémiai kiado. Budapest, 1967.
- [2] D.J. Levitin, A.K. Tirovolas. Music cognition and perception. Sage Encyclopedia of Perception. Thousand Oaks, CA: Sage Publications, s. 599-606, 2010.
- [3] I. Cross. Music, cognition, culture and evolution. Annals of the New York Academy of Sciences 930, s. 28-42, 2010.
- [4] P.N. Juslin, P. Laukka. Communication of emotions in vocal expression and music performance: Different channels, same code? Psychological Bulletin, 129, s. 770-814, 2003.
- [5] J. Panksepp. The emotional antecedents to the evolution of music and language. *Musicae Scientiae*, Special Issue, s. 229-259, 2009-2010.
- [6] B. Richman. How music fixed “nonsense” into significant formulas: on rhythm, repetition and meaning. In: The origins of music. Eds. Wallin NL, Merker B and Brown S. Massachusetts Institute of Technology, USA, s. 301-314, 2001.
- [7] W.E. Thompson, E.G. Schellenberg, G. Husain. Decoding speech prosody: Do music lessons help? *Emotion*, 4, 1, s. 46-64, 2004.
- [8] P.N. Juslin, D. Västfäll. Emotional responses to music: The need to consider underlying mechanisms *Behavioral and Brain Sciences*, 31, s. 559-575, 2008.
- [9] I.R. Murray, J.L. Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustic Society of America*, 93, 2, s. 1097-1108, 1993.
- [10] C.E. Izard. Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science* 2, 3, s. 260-280, 2007.
- [11] D. Matsumoto, B. Franklin, J.-W. Choi, D. Rogers, H. Tatani. Cultural influences on the expression and perception of emotion. Teoksessa: Gudykunst, W.B., Mody, B. (toim.) *Handbook of international and intercultural communication*. 2nd ed. USA: Sage Publications, Inc., 2002.
- [12] S.J. Morrison, S.M. Demorest. Cultural constraints on music perception and cognition. Teoksessa: Chiao, J.Y. (toim.) *Progress in Brain Research. Cultural Neuroscience: Cultural Influences on Brain Function*. The Netherlands: Elsevier, 178, s. 67 – 77, 2009.
- [13] T. Waaramaa, T. Leisiö. Perception of emotionally loaded vocal expressions and its connection to responses to music. A cross-cultural investigation: Estonia, Finland, Sweden, Russia and the USA. *Frontiers in Emotion Science*, 2013. http://www.frontiersin.org/Emotion_Science/10.3389/fpsyg.2013.00344/abstr act.
- [14] Å. Abelin. Anger or fear? Cross-cultural multimodal interpretations of emotional expressions. Teoksessa: K. Izdebski (toim.) *Emotions in the Human Voice*. Vol. I. Plural Publishing. San Diego, s. 65-73, 2008a.
- [15] Å. Abelin. Seeing glee but hearing fear? Emotional McGurk effect in Swedish. *Proceedings of Speech Prosody*. May 6-9, Campinas, Brazil, 2008b.
- [16] K.R. Scherer, R. Banse, H.G. Wallbott. Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology* 32, 1, s. 76-92, 2001.
- [17] G.M. Kotlyar, V.P. Morozov. Acoustical correlates of the emotional content of vocalized speech. *Soviet Psychology and Acoustics* 22, s. 208-211, 1976.
- [18] K.R. Scherer. Expression of emotion in voice and music. *Journal of Voice* 9, 3, s. 235-248, 1995.
- [19] T. Waaramaa (2014). Perception of emotional nonsense sentences in China, Egypt, Estonia, Finland, Russia, Sweden and the USA. *Logopedics Phoniatrics Vocology*, May 27, s. 1-7, 2014. <http://www.ncbi.nlm.nih.gov/pubmed/24861103>.
- [20] S.E. Trehub. The developmental origins of musicality. *Nature Neuroscience* 6, 7, s. 669-673, 2003.
- [21] A.-M. Laukkanen, E. Vilkman, P. Alku, H. Oksanen. On the perception of emotions in speech: the role of voice quality. *Scandinavian Journal of Logopedics, Phoniatrics, Vocology* 22: 4, s. 157-168, 1997.
- [22] T. Waaramaa, P. Alku, A.-M. Laukkanen. The role of F3 in the vocal expression of emotions. *Logopedics Phoniatrics Vocology*, 31, 4, s. 153-156, 2006.

- [23] T. Waaramaa, E. Kankare. Acoustic and EGG analyses of emotional utterances. *Logopedics Phoniatrics Vocology* 38:1, s. 11-18, 2013. doi: 10.3109/14015439.2012.679966.
- [24] J. Blacking. *How Musical is Man?* Seattle-London: University of Washington Press, 1973.
- [25] Praat. [http://www.fon.hum.uva.nl/praat/manual/Sound To Pitch ac.html](http://www.fon.hum.uva.nl/praat/manual/Sound%20To%20Pitch.html), 2015. (Luettu 4.4.2015.)
- [26] W. van der Meer. PRAAT Manual for musicologists. Academia.edu. [http://www.academia.edu/10404719/PRAAT Manual for musicologists](http://www.academia.edu/10404719/PRAAT_Manual_for_musicologists), 2015. (Luettu 31.3.2015.)
- [27] T. Ellingson. Transcription. Teoksessa: Helen Myers (toim.) *Ethnomusicology: an Introduction*. London: MacMillan, s. 110–152, 1992b.
- [28] N. Ruwet. Methods of analysis in musicology. *Music Analysis* 1–2/1987, s. 11–36, 1987. (Alkuperäisjulkaisu: *Méthodes d'analyse en musicologie. Revue belge de musicologie*. Vol. 20, s. 65–90, 1966.)
- [29] J. Niemi, M. Joust. Musiikin paradigmaattinen analyysi. Teoksessa Pirkko Moisala & Elina Seye (toim.) *Musiikki kulttuurina. Suomen Etnomusikologisen Seuran julkaisuja* 21. Helsinki: Suomen Etnomusikologinen Seura, s. 173–200, 2013.
- [30] J. Niemi. Inkeriläisten itkuvirsiens musiikilliset rakenteet. Teoksessa: A. Nenola: *Inkerin itkuvirret – Ingrian Laments*. Helsinki: Suomalaisen Kirjallisuuden Seura, s. 694–707, 2002. (Käännös samassa julkaisussa: *Musical Structures of Ingrian Laments*, s. 708–728.)
- [31] J. Toivanen, T. Waaramaa, P. Alku P, A.-M. Laukkanen, T. Seppänen, E. Väyrynen, M. Airas. Emotions in [a:]: A perceptual and acoustic study. *Logopedics Phoniatrics Vocology* 31; 1, s. 43–48, 2006.
- [32] T. Waaramaa, A.-M Laukkanen, P. Alku, E. Väyrynen. Mono-pitched expression of emotions in different vowels. *Folia Phoniatrica et Logopaedica*, 60, 5, s. 249–255, 2008.
- [33] P. Ekman. *Emotions Revealed Recognizing faces and feelings to improve communication and emotional life*. New York: Owl Books, 2004.
- [34] R. Banse, K.R. Scherer. Acoustic Profiles in Vocal Emotion Expression. *Journal of Personality and Social Psychology* 70, 3, s. 614–636, 1996.
- [35] K.R. Scherer, R. Banse, H.G. Wallbott, T. Goldbeck. Vocal cues in emotion encoding and decoding. *Motivation and Emotion* 15, 2, s. 123–148, 1991.
- [36] C. Seeger. *Prescriptive and Descriptive Music Writing. Studies in Musicology 1935–1975*. Berkeley-Los Angeles-London: University of California Press, 1977. Alkuperäisjulkaisu: *Prescriptive and Descriptive Music Writing. Musical Quarterly* XLIV, 2, s. 184–195, 1958.
- [37] T. Ellingson. Notation. *Ethnomusicology: an Introduction*, toim. Helen Myers. London: MacMillan, s. 153–164, 1992a.
- [38] N. Cook. *A Guide to Musical Analysis*. Oxford: Oxford University Press 1995, [1987].
- [39] J. Niemi, M. Joust. Teesejä pohjoisen laulun analyysiin: tarkastelussa nenetit ja saamelaiset. Teoksessa Antti-Ville Kärjä (toim.) *Etnomusikologian vuosikirja* 14, Helsinki: Suomen Etnomusikologinen Seura, s. 187–190, 2003.
- [40] J. Niemi, A. Lapsui. Network of songs. Individual songs of the Ob' Gulf Nenets. Music and local history as sung by Maria Maksimovna Lapsui. *Mémoires de la Société Finno-Ougrienne* 248. Helsinki: Société Finno-Ougrienne, 2004.
- [41] M. Joust. Tullälcalmaaš kirdáccej 'tulisilmillä lenteli' – Inarinsaamelainen 1900-luvun alun musiikkikulttuuri paikallisen perinteen ja ympäröivien kulttuurien vuorovaikutuksessa. *Acta Universitatis Tampereensis* 1650. Tampere: Tampere University Press, 2011.
- [42] A.P. Merriam. *The Anthropology of Music*. Evanston: Northwestern University Press, 1964.
- [43] B. Nettl. *The Study of Ethnomusicology: Thirty-One Issues and Concepts*. Urbana-Chicago: University of Illinois Press, 2005.

Prosovar-hankkeen väliraportti: Puheaineiston keruusta verkossa sekä havaintoja aineistosta

Tommi Nieminen¹, Tommi Kurki²

¹Itä-Suomen yliopisto

²Turun yliopisto

tommi.nieminen@uef.fi, tommi.kurki@utu.fi

Abstract

The Prosovar project or *Dialectal and social variation of Finnish prosody* was started by in 2013 and financed by the Kone Foundation in the period 2013-15. The project has applied in April 2015 for further funding from the Academy of Finland. The project has three objectives: (a) To analyse the prosodic phenomena of spoken Finnish, taking into account regional and social factors. (b) To form a speech corpus especially targeted for the study of prosody and variation (The Corpus of Prosodic Variation in Finnish). (c) To develop and test new, at least partially crowdsourced methods of collecting and analyzing spoken language data via the Internet. Obtaining data truly started in the autumn 2014. At the moment we have over 700 registered users 300 of which have actually contributed to the project by recording their speech samples.

Tommi Kurjen (TY) ja Tommi Niemisen (ISY) luotsaama Suomen kielen prosodian alueellisen ja sosiaalisen variaation hanke Prosovar alkoi 2013 ja sitä rahoitti Koneen säätiö 2013-15. Entisestään laajentuneelle hankkeelle on jo anottu uutta rahoitusta Suomen Akatemiasta. Hankkeella on ollut kolme tavoitetta: (a) Tutkia puhe-suomen kielen prosodisia piirteitä ja siinä ilmenevää alueellista ja sosiaalista variaatiota. (b) Muodostaa erityisesti prosodian ja sen variaation tutkimukseen soveltuva puhe-suomen korpus. (c) Kehittää ja testata uusia, vähintäänkin osin joukkoistettuja menetelmiä puheaineistojen keräämiseksi ja tutkimiseksi internetin välityksellä. Aineistonkeruu alkoi toden teolla syksyllä 2014, ja kirjoitusajankohtaan mennessä yli 700 henkilöä on jo luonut keruusivustolle käyttäjätunnuksen ja heistä 300 tehnyt äänityksiä tutkimuksen tarpeisiin.

Avainsanat: prosodia, puheaineiston keruu, joukkoistaminen, puhekorpuset

1. Johdanto

Prosovar tai virallisemmin *Suomen prosodian alueellinen ja sosiaalinen variaatio* on ollut FT, dos. Tommi Kurjen (Turun yliopisto) ja FT Tommi Niemisen (Itä-Suomen yliopisto) ideoitu sosiofoneettinen puheaineistonkeruu- ja tutkimushanke, jota Koneen säätiö on rahoittanut 1.8.2013-31.5.2015. Hankkeen johtajana on toiminut Kurki, ja sitä on sen mukaisesti hallinnoitu Turun yliopistosta. Mainittujen tutkijoiden lisäksi hankkeessa ovat toimineet M.Sc. Hamid Behravan (Turun yliopisto) hankkeen keruusivuston suunnittelijana marraskuun alusta 2013 lokaan loppuun 2014 sekä FM Heini Kallio (Helsingin yliopisto) väitöskirjantekijänä vuoden 2014. Hankkeella on ollut alusta asti myös yhteistyökumppaneita, joiden asiantuntemusta on varsinaisesti tarkoitus hyödyntää ennen kaikkea meneillään olevan keruuvaiheen jälkeen. Hanketta on eri puolilta ja eri yleisöille esitelty kolmessa aiemmassa artikkelissa [1, 2, 3].

Prosovar-hankkeen tavoitteita ovat olleet:

1. kehittää uusi, joukkoistukseen perustuva puheaineiston keruumenetelmä
2. kerätä suomesta prosodian tutkimukseen soveltuva puhetietokanta sekä
3. tutkia suomen prosodian alueellista ja sosiaalista variaatiota.

Tavoitteista ensimmäinen on toteutunut (ks. lukua 2), mikä on antanut hyvän pohjan toisenkin tavoitteen enenevälle toteutumiselle jo lähiaikoina (luku 3). Kolmas tavoite on kiinni kahdesta ensimmäisestä, mutta siinäkin ollaan jo hyvässä alussa (luku 4). Tavoitteiltaan, menetelmiltään ja tutkijapohjaltaan entisestään laajentuneelle hankkeelle on jo haettu lisärahoitusta Suomen Akatemian huhtikuussa 2015 päättyneessä Digital humanities -haussa. Uuden, Akatemian asiantuntijoiden arvioitava olevassa hankkeessa on edellä mainittujen lisäksi myös tietojenkäsittelytieteeseen ja kieliteknologiaan asetettuja tavoitteita. Yllä mainitut kielitieteelliset tavoitteet ovat silti käytännössä edelleen samat, koska kuten jäljempänä nähdään, sinänsä toimivaksi osoittautuneessa ja tuottoisassa keruumenetelmissäkin on yhä kehitettävää.

2. Keruukone verkossa

Prosovar-hanke käynnistyi virallisesti 1.8.2013, jolloin Tommi Nieminen aloitti ensimmäinen tutkimuskautensa. Hankkeen tässä vaiheessa keskityttiin lähinnä suomen prosodiaa ja etenkin sen variaatiota koskevan tutkimuskirjallisuuden kartoittamiseen, hyvin oleellisen verkkosivuston toteuttavan projektitutkijan valintaan sekä sivuston aineistonkeruutehtävien ideointiin ja kehittelyyn; keruutehtävät tosin ovat jatkuva projekti, koska vanhoja käyttäjiä ei saada palaamaan sivulle, ellei siellä ole aina heille jotakin uutta tehtävää. Hankkeen varsinaisen työvaiheen voi sanoa alkaneen pikemmin 1.11.2013, jolloin projektitutkijaksi valittu Hamid Behravan alkoi valmistella hankkeen pääkeruumenetelmäksi suunniteltua keruukonetta verkkoon. Alkukeväästä 2014 sivusto oli jo hankkeen jäsenten sisäisessä testikäytössä (VPN:n kautta), ja huhtikuukokouksessa samana vuonna sivusto julkistettiin osoitteessa <https://puhu.utu.fi/>, jossa se sijaitsee ja on aktiivinen edelleen. Hieman harmillisesti tosin sivuston julkistus osui juuri aikaan, jolloin yliopistojen kevätlukukaudet olivat vaihtumassa kesälomiin, joten sivuston mainostus venyi kesän molemmin puolin, mikä hidasti keruun käyntiin saattamista. Aktiivinen keruu pääsikin vauhtiin ennen kaikkea syksyllä 2014.

Sivustoa on sittemmin mainostettu paitsi kielentutkijoiden ja foneetikkojen kotimaisissa konferensseissa (kuten Kielitieteen ja Fonetikan päivillä) ja yliopistojen omissa tutkijata-

paamisissa myös sosiaalisessa mediassa (ennen kaikkea Suomen eri maakuntien ja seutujen ja paikkakuntien Facebook-ryhmissä, joista yli 90:ssä on tiedotettu hankkeesta ja sivustosta). Markkinointia on porrastettu alueittain ja mainoskampanjat rytmitetty sykleittäin. Sivustolla kävijöiden määrässä on aina ollut havaittavissa selvä piikki kunkin mainostusjakson jälkeen, mikä jälleen kerran osoittaa, ettei verkkoon saattaminen yksin riitä saavutettavuuden takaamiseksi: aiotulle kävijäkunnalle on myös tavalla tai toisella saatava viedyksi tieto verkkosivuston olemassaolosta. Käytännössä tämä voi olla hankalaa pelkästään hankkeen budjetin kannalta, koska varsinaiseen mainostukseen ei ole varattu erillistä rahoitusta.

Yksi puoli asiasta on, että mitä suuremmalle yleisölle tiedon saa kulkeutumaan, sitä todennäköisemmiksi käyvät myös erilaiset häirintä- tai jopa väärinkäyttöyritykset, mutta toisaalta jos viestintä suunnataan pelkästään tutkijayhteisölle, sivustolle onnistuu myös houkuttamaan lähinnä tutkijayhteisöön kuuluvia tai siihen läheisessä kosketuksessa olevia käyttäjiä (kuten opiskelijoita), mikä olisi omiaan vääristämään kertyvän aineiston niin alueellista kuin sosiaalista jakaamaa.

Mainostamisen ja markkinoinnin näkökulmasta pelkästään tiedon välittäminen sivuston olemassaolosta potentiaalisille osallistujille ei riitä. Kaikki, jotka kiinnostuvat aiheesta ja sivustosta, eivät käy sivuilla heti tai ainakaan tee äänityksiä välittömästi. Osa käyttäjätunnuksen luoneistakin - siitä riippumatta, ovatko jo äänittäneet kaikki tarjolla olevat tehtäväsarjat vai ei - unohtaa sivuston, ellei heitä muistuteta sivuston olemassaolosta. Tässä mielessä ei ole käyttäjien kannalta houkuttelevaa kertoa vain, että sivustolle odotetaan edelleen osallistujia, vaan sivuilla pitää olla jotakin uutta kiinnostavaa. Kiinnostavaksi tiedoksi taas ei helposti katsota pelkästään rekisteröityneiden käyttäjien tai kertyneiden äänitteiden kasvua, paitsi sikäli että se tarkoittaa sivuilla mahdollisesti olevan uutta aineistoa myös tavallisten käyttäjien kuunneltavaksi (esim. murteen tunnistuspeleissä, ks. myöhemmin).

Toteutetulla keruusivustolla on julkinen ja vain rekisteröityneille käyttäjille - eli informanteille - tarkoitettu alueensa. Julkinen, kaikille avoin osuus on hyvin tiivis ja sisältää lähinnä lyhyet, jo vanhaan tutkimuskirjallisuuteen perustuvat esittelyt suomen vanhoista murrealueista. Käytännössä tiedetään, että tämä tieto on aikansa elänyt, ja sellaiseksi se sivulla myös tunnustetaan: hankkeen julkinen, tiedeyhteisön ulkopuolelle annettava kuva on nimenomaan se, että hankkeella pyritään kartoittamaan suomen nykyisten aluepuhekielen tila. Millään lailla ei ole pyritty erikseen piilottamaan näkyvistä, että keskeinen tutkimuskohde on nimenomaan prosodia, päinvastoin tämä näkyy jo sivuston tervetulo-tervehdyksessä (Tervetuloa Suomen kielen prosodian tutkimushankkeen sivuille) samoin kuin sivuston otsakkeessa (Suomen prosodian tutkimushanke — Suomen prosodia). Prosodian vanhasta variaatiosta ei kuitenkaan tarjota olemassa olevaa tutkimustietoa, koska sellaista ei yksinkertaisesti juurikaan ole olemassa. Prosodiaa koskeva suomalainen tutkimus, joka on kaiken kaikkiaan vähäistä, on pitkälti kosketellut niin sanottua yleispuhekieltä.

Puhu.utu.fi on ollut aito yritys joukkoistaa keruuta, ei vain hyödyntää verkkoa keruun menetelmänä. Tavoitteena oli luoda sivusto, jonne ihmiset voisivat palata yhä uudelleen äänittämään. Keinoina olivat erilaiset tehtäväsarjat, pelimaisuus ja tiedon jatkuva karttuminen. Vaikka sivuston tärkein tehtävä on mahdollistaa aineistonkeruu eri puolilta Suomea ja erilaisia suomen muotoja puhuvilta ihmisiltä ja joukkoistaa massatallentamaan puhettua tutkimuksen tarpeisiin, tarkoitus on saada heidät mahdollisuuksien mukaan osallistumaan hankkeeseen muutenkin: osallistujat muun muassa kuuntelevat toistensa

näytteitä ja raportoivat niiden mahdollisista teknisistä ongelmista (esim. heikosta äänenvoimakkuudesta, taustahälystä, rahinasta, puheen epäselvyydestä) ja näin keventävät tutkimusryhmän taakkaa aineiston läpikäynnissä. Samoin pelatessaan murteen tunnistuspeliä (ks. tarkemmin seuraavaa lukua) he kartuttavat samalla tutkijoiden käyttöön tavallisten suomalaisten kansanlingvistisiä käsitteitä puhutusta suomesta ja sen eri varieteeteista. Tulevaisuudessa haaveena on myös tarjota verkossa tiedon paluukanava tutkijoilta kerääjille siten, ettei tieto jäisi yksin odottamaan tulevia julkaisuja vaan tavoitaisi informantit asianosaisina nopeammin ja vuorovaikutteisemmin.

Erilaisten tehtäväsarjojen ajatuksena on ollut, että sivusto tarjoaisi pelaajalle joka kerralla erilaiset tehtävät ja pitäisi näin informanttien mielenkiintoa yllä. Samalla helpotettaisiin sivuston käyttöä niinkin, että käyttäjä voi aina keskeyttää tehtävien tekemisen ja palatessaan saa eteensä tekemättömiä niin kauan kuin niitä on tarjota. Kaikissa tehtävissä on pyritty löytämään tehtävän taustaksi tavallisia kielenkäyttötilanteita ja arkista toimintaa, joka houkuttaisi käyttämään nimenomaan aitoa arkikieltä, ei muodollisempaa yleispuhekieltä, jotta aineisto kokonaisuudessaan voisi tavoittaa saman asiaa eri murteen. Vastapainona tehtävien taustatekijät on kuitenkin hankkeen päätavoitteiden täyttämiseksi pyritty vakioimaan niin, että puhunnokset olivat mahdollisimman pitkälle vertailukelpoisia. Tehtävinä on esimerkiksi käydyt tietokoneen esittämän toisen osapuolen kanssa torilla keskustelua marjojen ostamisesta, kuvailla tapahtumia sivulla esitetyllä videolla sekä opastaa henkilöä paikasta toiseen (kuvitteellisen kaupungin kartan avulla).

Täysin tavoitteissa ei ilmeisesti ole onnistuttu, koska havaintojen perusteella liian moni äänittäneistä ei enää sivustolle palaa tekemään loppuja tehtäviä. Tulevaisuutta varten on siis kysyttävä, miten vuorovaikutteisuutta voitaisiin lisätä ja joukoistamisesta tehdä aidompaa. Tavoitteena ei tulisi olla vain käyttää internetiä helppona ja nopeana keruukanavana, vaan tutkimuksen kulku tulisi saada mahdollisimman paljon näkyville, kuitenkin kaiken aikaa huolehtien myös tutkijan eettisestä vastuusta kuten informanttien yksityisyyden suojelemisesta. Jos saamme uudelle hankkeelle rahoituksen, yksi sivuston kehitysteemoista tulee olemaan se, kuinka sivuja kyettäisiin uudistamaan määräajoin ja kaikille halukaille käyttäjille tarjoamaan uusia tehtäviä - ja pelejä - niin, että sivuston vetovoima säilyy ja informantit sitoutuvat kiinteämmin sen (ja samalla hankkeemme) tavoitteisiin. Tarkoituksena on joka tapauksessa, että käyttäjät eivät vierailisi sivulla vain kertaluontoisesti vaan palaisivat sinne uudestaan, jotta kultakin informantilta saataisiin riittävän paljon ääninäytteitä ja mielellään useampana eri äänityskertana.

Todettakoon kuitenkin erikseen, että vaikka selvä enemmistö tähänastisista informanteista on vieraillut sivustolla tekemässä äänitystehtäviä vain kerran, ovat tähän enemmistöönkin kuuluvat tehneet yleensä useita eri äänitystehtäviä ja heiltä on saatu useita äänitteitä. Tällä hetkellä yhtä informanttia kohden on saatu keskimäärin 13 ääninäytettä (ääninäytteiden kokonaismäärä 6.5.2015 on 3886), kun enimmillään niitä on ollut tehtävissä 25, tavallisesti 20. Sellaisia informanteja, jotka ovat tallentaneet vähemmän kuin viisi omaa ääninäytettään on vain 8 % kaikista informanteista. Sen sijaan 60 % informanteista on jättänyt tietokantaan tutkijoiden käyttöön vähintään 15 omaa ääninäytettään.

Eräänä ajatuksena joukoistamisnäkökulman edistämiseksi on ollut lisätä sivustolle pelimaisuutta, joka saisi paitsi käyttäjät palaamaan myös tarjoaisi heille jotain nopeasti takaisin. Toiseksi tästä on toteutettu murteen tunnistuspeli (kuva 1), jossa informantit saavat kuultavakseen anonymisoituja näytteitä

ja heidän tehtävänsä on pyrkiä sijoittamaan kuulemiensa näytteiden puhujat Suomen kartalle. Mitä paremmin ja tarkemmin he pystyvät paikantamaan näytteet, sitä enemmän he saavat pisteitä. Vastavuoroisesti kukin osallistuja pystyy näkemään omista käyttäjätiedoistaan, mistä päin Suomea muut tunnistuspeliä pelanneet ovat arvelleet hänen olevan. Tiedot ovat nähtävissä kustakin osallistujan ääninäytteestä erikseen, ja sivustolla esitetään graafisesti jakauma siitä, millaisia paikannusyritykset ovat olleet.

Kuten edellä mainittiin, on huomattava osa tähänastisista äänityksiä tehneistä informanteista on käynyt sivustolla vain kerran. Vaikemme ole ennen kaikkea käytännön syistä pystyneet vielä uudistamaan sivuston sisältöä ja tehtäviä niin tiheästi kuin on ollut tarkoitus, informanteja on palannut sivustolle vielä ensi kerran jälkeenkin. Saamamme palautteen perusteella osa on kaivannut lisää äänitystehtäviä ja pitänyt tehtäviä koukuttavina, osa on puolestaan palannut sivustolle kuullakseen näytteitä toisten suomalaisten puheesta, ja osaa ovat houkuttaneet nimenomaan pelimaisuus ja pisteiden karttaminen. Muutamalle osallistujalle on selvästi ollut tärkeää, että he ovat tunnistuspelin pistetaulukossa parhaimpien joukossa. Näytteiden kuuleminen on kiinnostanut muutamaa sellais-takin vierailijaa, jotka ovat yrittäneet äänittää lyhyitä sanatomia omia äänityksiään vain päästäkseen kuulemaan toisten äänikatkelmia. Tällaiset osallistujat olemme kuitenkin pystyneet torjumaan.

Sivuston tällä hetkellä käytössä olevista tehtävistä valtaosa näkyy kuvassa 2. Lomakeskustelutehtävässä (ylävasemmalla) tarkoituksena on poimia kehoitteesta annetuista lauseista omaan puheenparteen aina parhaiten sopiva vaihtoehto tai ehdottaa omaa avoimeksi jätettyyn tekstikenttään. Kaksi vaihtoehtoa esittämällä pyritään rajaamaan informantin mielikuvitusta. Vaihtoehdot ovat aina sellaiset, että toinen niistä edustaa varsin prototyyppistä länsi- ja toinen itämurretta. Kuvan erojen selitystehtävässä (yläoikealla) informantin puheen sisältöön on pyritty ennakoimaan kuvan avulla. On tietenkin mahdollista ennustaa, missä järjestyksessä informantti kuvan eroja havaitsee, mutta kutakin eroa koskevan kielennöksen voisi olettaa olevan prosodisilta (joskaan ei esimerkiksi sanastollisilta) ominaisuuksiltaan joksenaan samantapainen. Karttaopastustehtävässä (keskellä) on tarkoituksena opastaa henkilö kuvitteellisen kaupungin kartalla paikasta toiseen. Informantin valintoja rajaavat tässä kulkuväylien suhteet toisiinsa. Torikeskustelutehtävässä (alina) on kaksi osaa - ostajan ja myyjän -, joista kuvassa näkyy ensimmäinen. Yleiskielisessä kuvauksessa kerrotaan, mitä informantin tulisi omalla murteellaan sanoa. - Kuvasta puuttuvat vielä videoesitystehtävä, jossa tarkoituksena on kuvailla samaan aikaan näkyvillä juoksevan videon tapahtumia, sekä nyttemmin vain rekisteröitymisvaiheessa käytettävä ja kertyvään aineistoon tallentumaton tehtävä, jossa mm. luetellaan pieniä numeroita sekä kuukausien nimiä ja kerrotaan aamun toimista. Viimeksi mainittu tehtävä toimii nykyään vain käyttäjän mahdollisuutena tarkistaa laitteistonsa toiminta ja mielellään myös äänitystaso (joskaan jälkimmäistä mahdollisuutta kaikki eivät tunnu käyttäneen hyväkseen, ks. myöh.).

3. Kertyneestä aineistosta

Aineistoa on keruukoneen ensimmäisen toimintavuoden aikana kertynyt varsin hyvin: sivustolla on jo yli 700 rekisteröitynyttä käyttäjää. Heistä kuitenkin osallistuneita eli äänittäneitä informanteja on selvästi vähemmän, 300 (tilanne 6.5.2015). Myönteistä on ollut se, että useimmat äänittäneistä ovat kuitenkin tehneet useamman kuin yhden äänityksen (ks. edell. luku)

ja pahantahtoiset väärinkäyttäjät (sekä pelleillen äänittäneet että sivustolle murtautumista yrittäneet) ovat olleet hyvin harvassa.

Ensimmäinen huolenaihe on tietenkin se, miksi oikeasti tehtäviä tehneitä eli puhunnoksiaan äänittäneitä käyttäjiä on niin selvästi vähemmän kuin rekisteröityneitä käyttäjiä kaikkiaan. Tehdäänkö rekisteröityminen vain, jotta päästään katsomaan, millaisia tehtävät ovat? Onko joukossa sittenkin murtautumis- tai muita häirintäyrityksiä? Vai eikö sivuston tekniikka kaikille käyttäjille edelleenkin toimi?

Asiallista arviota eri syiden suhteesta on vaikea esittää. Sivuston sen paremmin kuin sitä majoittavan palvelimien ylläpitäjiltä ei ole ainakaan tullut viestiä murto- tai häirintäyrityksistä. On myös vaikea uskoa, että sivuston häirintä laajemmin kiinnostaisi, koska mihinkään arkaluontoiseen tietoon ei näin edes pääsisi käsiksi; sivustollehan ei tallenneta käyttäjistä edes oikeita nimiä paitsi sikäli kuin ne sähköpostiosoitteista mahdollisesti näkyvät. Kaksi muuta syytä jäävät siis jäljelle. Jossakin määrin on - ennen kaikkea sivustoa hyödyntäneiden kouluvierailujen yhteydessä - voitu havaita, että joskus äänitys ei vain toimi tai ainakaan käyttäjä ei saa sitä toimimaan. Esimerkiksi silloin tällöin on voinut huomata, että kun Flash-tekniikalla toteutettu äänitysovellus pyytää käyttäjää vahvistamaan äänityslaitteiston käytön, kyselyikkuna jää selainikkunan alle niin, ettei käyttäjä sitä edes huomaa. Toisaalta aina ei ole ollut kyse tästäkään vaan sovellusta ei yksinkertaisesti ole saanut koneella käyntiin. Erityisongelmana ovat olleet mobiililaitteet, joita hanke alun perin ei edes pyrkinyt tavoittamaan mutta jotka kuluttajalaitteiston nykyjakautuman muistaen olisi ehkä pitänyt ottaa paremmin huomioon; nykytilanteessa sovellusta ei voi lainkaan käyttää iOS-laitteissa, ja Android-käyttö muistuttaa enemmän hankalaa offline-toiminnallisuutta kuin sujuvaa verkkosovellusta. Mobiililaitteille keruusivusto on tarkoitus toteuttaa kokonaan omana sovelluksenaan, jos hankkeelle saadaan jatkorahoitusta.

Alueellinen jakauma näyttää kartalle sijoitettuna (kuva 3) jo hyvältä, mutta kuvaa haalistaa toistaiseksi vielä se, että muutamilta alueilta äänittäneiden informanttien määrät ovat äärimmäisen pieniä (eritoten peräpohjalais- ja eteläpohjalaismurteiden alueelta; ks. taulukkoa 1). Useilta vanhoilta murrealueilta aineistoa on jo riittävästi, jos tarkastelussa on tarkoitus keskittyä nimenomaan alueellisiin eroihin kiinnittämättä sen tarkemmin huomiota vielä kunkin alueen sisäisiin sosiaalisiin eroihin. Jos alueellisen variaation tarkastelu halutaan ulottaa koskemaan myös vanhojen murrealueiden eri alaryhmiä, tarvitaan useiden alaryhmien alueelta vielä edustavampi otos.

Ikäkausittaisia variaatiota lukuun ottamatta on muuhun sosiaaliseen variaatioon toistaiseksi epämielikäsiä tarkastella, koska aineistoa on siihen liian vähän. Tähänastisista informanteista suuri enemmistö on naisia. Koulutustaustan mukainen jakauma painottuu aikuisissa puolestaan korkeakoulututkinnon suorittaneisiin (yliopistot, teknilliset korkeakoulut ja ammattikorkeakoulut), ja lisäksi - kuten tunnettua - on koululaisten ja opiskelijoiden luokitteluun koulutuksen perusteella suhteessa muihin informantteihin kaikkea muuta kuin ongelmantonta. Jälkimmäisenä mainittu koskee luonnollisesti yhtä lailla ammattitaitaustaa. Ammattillisen taustan kirjavuutta kuvastaa hyvin se, että tällä hetkellä, kun informantteja on 300, on heidän ilmoittamiensa ammattinimikkeen yhteismäärä 100.

Kansainvälisessä sosiolingvistiksessä tutkimuksessa kielennukkoisista (sosiaalisista) tekijöistä on tutkimuksessa yleisimmin otettu huomioon ikä, sukupuoli ja sosioekonominen status, joka on määritelty usein juuri joko ammatin, koulutuksen tai kummankin perusteella. Toisinaan statuksen määrittelemisessä



Kuva 1: Murteentunnistuspeili.

	alle 16	16-18	18-20	21-30	31-40	41-50	51-60	61-70	71-80	tunt.	YHT.
eteläpohj.m.		1	1		4	2		1			9
hämäläism.		5	4	8	5	7	5	1			35
kaakkoism.			5	1	1	1	2				10
keski- ja pohjoispohj. m.		1	1	10	7	6					25
lounaism.	9	1		5	4	3	1			1	24
peräpohj.m.				3	2	1					6
pääkaupunkiseutu		2	7	28	14	9	6	1		1	68
ruotsink. alueet				1		1					6
savolaism.	1	2	1	26	9	11	3	1	1		55
siirtymäm.	11	21	10	12	4	1	1				62
ulkomaat				3	1						4

Taulukko 1: Aineiston ikäryhmittäinen ja murrealueittainen jakauma.

on käytetty hyväksi myös asuinalueita tai tuloja. Suomalaisissa tutkimuksissa sosiaaliluokkaa tai sosioekonomista status-ta ei ole 1970-luvun Nykysuomalaisen puhekielen murros -hanketta lukuun ottamatta yleensä syystäkään huomioitu, koska Suomessa ei ole katsottu muodostuneen samanlaista luokkayhteiskuntaa kuin esimerkiksi Brittein saarilla. Yleisimminkin tämän kielenulkoisen tekijän määrittelyä on pidetty hankalana, ja sen sijaan on tarkasteltu esimerkiksi informanttien sosiaalisia verkostoja (ks. esim. [4, 39-41], [5, 81-82]). Sosiaaliryhmää (1. sosiaaliluokkaa 1. sosioekonomista statusta) on Suomessa perinteisesti ollut vaikea määrittää ammatin ja koulutuksen perusteella; esimerkiksi 1970-luvun Tampereen puhekielen tutkimuksessa saatettiin sosiaaliluokka samastaa informantin - tai alaikäisen tapauksessa hänen vanhempiensa - koulutustasoon [6, 15], mikä ymmärrettävästi on karkea yksinkertaistus. Hankkeemme keruusiivustolla osallistujilta kysytään rekisteröitymisvaiheessa perustiedot heidän koulutus- ja ammatitilustaan siltä varalta, että halutessaan tutkija voi ottaa nämä seikat huomioon, kun hän käyttää korpuksen aineistoja.

Kun siis tähän asti karttunutta aineistoa tarkastellaan sen alueellisen ja sosiaalisen edustavuuden näkökulmasta, keruun voi katsoa edenneen vähintäänkin tyydyttävästi. Aineistoa on kokonaisuudessaan kertynyt kiitettävästi, ja äänitteitä on paljon useimmilta vanhoilta murrealueilta. Jotta aineistoa edustaisi riittävän hyvin kaikkia murrealueita, jatkossa keruusi-

vustoa on keskityttävä markkinoimaan erityisesti Lapissa ja Etelä-Pohjanmaalla. Toisaalta koska aineistoa on kuitenkin kertynyt tähänkin saakka runsaasti, olisi houkuttelevaa pyrkiä saamaan edustavat aineistot vielä vanhojen eri murreryhmien alamurteistoista. Sosiaalisen edustavuuden saavuttaminen vaatii sekin enemmän aineistoa. Toistaiseksi kertyneen aineiston pohjalta pystytään tarkastelemaan määrättyjen rajoitusten ikävariaatiota. Korpuksen edustavuuden parantamiseksi on tehty jatkosuunnitelma, jonka mukaan aineistoa kerrytetään sivuston lisäksi myös perinteisin menetelmin keräämällä täydennysaineistot kasvokkaisuäänityksin fonetiikan laboratoriossa Turussa, Tampereella, Jyväskylässä, Joensuussa ja Oulussa. Näillä äänityksillä parannetaan ja täydennetään myös korpuksen aineiston äänenlaadullista edustavuutta.

Jo hanketta suunniteltaessa ja ensi kertaa esiteltäessä ulkopuoliset nostivat usein esiin kysymyksen siitä, riittääkö vaihtelevissa olosuhteissa (kotona, työpaikalla, mahdollisesti ulkosalakin) verkon ylitse tehtävien äänitteiden laatu mitenkään soiofoneettiseen tutkimukseen. Vastauksenamme oli aina, että lähtökohtanamme on tutkittavien prosodisten piirteiden (F0:n, A0:n ja keston) robustius akustisina piirteinä: analyysiin riittää usein tieto piirteiden suhteellisesta arvosta kontekstissaan ja absoluuttinen arvo on varsin merkityksetön.

Äänitteiden äänitystasoa on toistaiseksi arvioitu vain kuvailevalla asteikolla, mutta keskiarvon voi sanoa olevan tyy-

dyttävä. Kuvissa 4-5 näkyy esimerkkejä asteikon äärilaidoilta (kummassakin kuvassa Praatin amplitudi on säädetty vertailun helpottamiseksi vakioiksi välille -151). Kuvassa 4 äänitystaso on erinomainen: äänitteen laatu on parempi kuin useimmissa kotiolosuhteissa tehdyissä murrenauhoitteissa ja riittäisi hyvin niin perinteisen murteen tutkimuksen kuin segmentaalisen foneettisenkin tutkimuksen käyttöön. Kuvassa 5 taas äänitystaso on niin heikko, että jo egmentointi on hankalaa, mikä tekee prosodiankin havainnoimisesta vaikeaa ellei täysin mahdotonta. Kummankin ääripään on tietenkin täysin mahdollista esiintyä aineistossa, mutta onneksi valtaosa todellisista äänitteistä on kuitenkin huomattavasti lähempänä kuvaa 4 kuin kuvaa 5. Äänitystasoon olisi myös informantin usein kohtuullisen helposti mahdollista vaikuttaa esim. säätämällä mikrofonin ottoäänenvoimakkuutta tai puhumalla lähempänä mikrofonia, ja tähän sivustolla on myös annettu yleisluontoiset ohjeet. Näyttää kuitenkin siltä, että sivuston ensimmäistä, aineistoon kertymätöntä äänitystason testaukseen toteutettua tehtäväsarjaa ei juurikaan aiotuun tarkoitukseen ole käytetty. Käyttäjät tuntevat tyytyvän siihen, että edes jotakin ääntä tallentuu, ja jatkavat asetuksia muuttamatta. Osa ongelmia voi olla se, että ohjeissa on mahdotonta tarkalleen neuvoa, miten mikrofonin äänenvoimakkuutta voi kaikissa eri käyttöympäristöissä ja -liittymässä muuttaa. Tilanne helpottuisi huomattavasti, jos keruukoneeseen itsessään voisi toteuttaa äänenvoimakkuuden säätimet, jolloin niitä ei tarvitsisi käyttäjän etsiä muualta.

Äänitystason lisäksi ja ehkä ensisijaisestikin on otettava huomioon muut äänitykseen laatuun liittyvät ongelmat: huono S/N-suhte (taustahäly, usein jo itse tallentavasta laitteesta), huono taajuusvaste (etenkin korkeat taajuudet huonoja: esim. frikatiivien analyysi mahdotonta) sekä muut, oudommat mahdollisesti laitteistoon tai äänitysympäristöön liittyvät ongelmat.

4. Missä mennään?

Ensimmäinen osatavoite on saada kultakin äänittäneeltä luotettavasti segmentoitua sanat, tavut ja foonit vähintään yhdestä äänitteestä. Pientä hävikkiä on luvassa siitä, etteivät aivan kaikki äänitteet ole ehkä segmentoitavissa laatusyistä.

Pelkkä aineiston segmentointi riittää jo kesto-suhteiden - ja niiden mahdollisen variaation - analyysiin. Taulukossa 2 näkyy esimerkki lyhyen ensitavun jälkeisen toisen vokaalin keston variaation alustavista havainnoista (aineistona videonkuvailutehtävä). Kaikki esimerkkitapaukset valitun sanan (*dino*) mitatut tapaukset eivät ole täsmälleen samoista lauseasemista ja ovat muutenkin erilaisia; tarkoituksena ei tässä ole vielä esittää mittaustuloksia vaan lähinnä havainnollistaa, mitä aineistosta löytyy ja mihin suuntiin tulevaa tutkimusta on tarkoitus viedä. Vaihtelu on häkellyttävän suurta, olkoonkin että joitakin tapauksia voisi selittää esimerkiksi puhetahdin loppupidennyksellä tai muilla oletettavasti kaikkiin puhemuotoihin vaikuttavilla syillä.

Seuraava tavoite on merkitä äänitteisiin prominenssit, vähintäänkin aksentit eli lausepainot. Intonaatiota voi tarkastella jo Praatin muokkausikkunoissa, mutta intonaation lingvistisen merkityksen arviointi edellyttää ihmiskuulijan havaintoa siitä, miten F0:n kulkua käytetään hyväksi. Jo vanhoista tutkimuslähteistä löytyy yksittäisiä mainintoja siitä, että murteittain voi olla tarpeen erottaa ekspiratorinen eli aksenttiin liittyvä ja melodinen eli intonaation kulkuun liittyvä korko eli sävelkorkeuden nousu (esim. [7, 8, 9]), vaikka nämä funktiot yleispuhekielessä kulkisivat yhdessä.

Aineistosta ei siis toistaiseksi ole tehty minkään piirteen laajempaa kartoitusta, koska sen karttumista odotetaan yhä ja monet rutiinitehtävät laadun tarkistuksesta segmentointiin

(käytännössä Praatin textgrid-objektien luomiseen) sijaitsevat työjärjestyksessä ensin, mutta yksittäisiä havaintoja on totta kai olemassa myös muista kuin kestoilmiöistä, ja ne vievät tutkimusta kaiken aikaa eteenpäin. Esimerkiksi useamman informantin laaja ja selvä nousevan intonaation käyttö kuvailevassa puheessa (kuva 6) osoittaa varmasti tämän tähän asti lähinnä tilanteisesti tarkastellun piirteen ansaitsevan tulla tarkastelluksi myös alueellisen jakaumansa kannalta.

5. Lopuksi

Prosovar-hanke on osoittanut kaksi asiaa: ensinnäkin, että tutkimustarkoituksiin riittävän hyvälaatuista puheaineistoa on mahdollista kerätä verkossa, ja toiseksi, että näin toimittaessa aineistoa kertyy huomattavasti nopeammin ja vähemmällä tutkijan rutiiniväsymällä kuin perinteisissä menetelmissä. Käytännössä on luultavasti vain niin, että ellei aineistonkeruuseen ole varattavissa runsaasti ihmistyövoimia, ainoa mahdollinen tapa laajojen puhekorpuksen keräämiseen on pyrkiä levittämään työtä joukkoistuksen avulla informanteille itselleen. Tavallaan tässä ei edes ole muuta uutta kuin internetin läsnäolo, koska sanoja ja muoto-opin aineistoa Suomessa on aiemminkin kerätty myös tavallisten ihmisten vapaa-aikanaan tekemin keruin.

Tämä ei tarkoita sitä, ettei verkkokeruussa olisi jäljellä olevia ja mahdollisesti ratkaisemattomiakin ongelmia. Osa liittyy informanttien saavutettavuuteen - pelkkä läsnäolo internetissä ei riitä, vaan sivustosta pitää saada tieto kulkeutumaan mahdollisille käyttäjille -, osa äänitysteknisiin ongelmiin - ei niinkään epäilyihin kuluttajätietokoneiden äänityslaitteiston laadusta vaan itse äänityspaikan sopivuuteen tarkoitukseensa. Aito joukkoistaminen edellyttäisi ehkä myös merkittävämpää roolia itse käyttäjille myös sivuston kehittämisessä, mutta tässä tulevat helposti eettiset kysymykset vastaan eli mahdottomuus pitää aineistoa käyttäjiltä salassa, jos sivuston tekninen toteutus heille paljastetaan.


Edelleen voisi pohtia sitä, olisiko joukkoistetusta keruusta mahdollista päästä myös joukkoistettuun tutkimukseen. Toistaiseksi analyysivaihe jää täysin siihen koulutetuille tutkijoille, koska automaattiset menetelmät eivät yleensä joko toimi riittävän hyvin (etenkään jos aineiston laatu vaihtelee näin suuresti) tai niitä ei ole vielä kehitettykään kuin kokeilukäyttöön laboratoriotasoisille normalisoiduille äänitteille. Kaukana tulevaisuudessa on vielä mahdollisuuskin ajatella puheentutkimusta tehtävän yksityisille tietokoneille hajautettuna jonkin SE-TI@Homen tai vastaavan hankkeen tapaan.

id	V1	V2	V2/V1 (%)
298	0,117317	0,113847	97,04 %
205	0,073803	0,094247	127,70 %
218	0,074110	0,095098	128,32 %
333	0,073463	0,099031	134,80 %
240	0,087630	0,118438	135,16 %
103	0,065067	0,112725	173,24 %
300	0,062389	0,115403	184,97 %
181	0,060069	0,123763	206,03 %
282	0,055163	0,116267	210,77 %
238	0,074896	0,171522	229,01 %

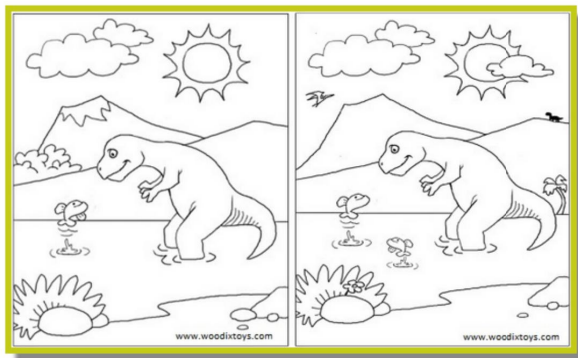
Taulukko 2: Esimerkki CVCV-tyyppisen sanan ensi ja toisen tavun lyhyiden vokaalien keston suhteista aineistossa.

6. References

- [1] T. Nieminen, T. Kurki, H. Kallio, and H. Behravan, “Uusi puhe-suomen variaatiota tarkasteleva hanke: Katse kohti prosodisia ilmiöitä,” *Sananjalka*, vol. 56, p. 186–195, 2014.
- [2] T. Nieminen, T. Kurki, H. Behravan, and H. Kallio, “Tutkimushanke suomen prosodian variaatiosta,” in *XXVIII Fonetikan päivät: Turku 25.–26. lokakuuta 2013: Konferenssijulkaisu*, K. Jähi and L. Taimi, Eds. Turku: Fonetikka, LAB-laboratorio, Turun yliopisto, 2014, pp. 56–64. [Online]. Available: <http://urn.fi/URN:ISBN:978-951-29-5980-8>
- [3] T. Nieminen, “Puhesuomen prosodiaa tutkimassa,” *Kielikuvia*, no. 2, pp. 2–9, 2014.
- [4] H. Lappalainen, *Variaatio ja sen funktiot: Erään sosiaalisen verkoston jäsenten kielellisen variaation ja vuorovaikutuksen tarkastelua*, ser. Suomalaisen Kirjallisuuden Seuran Toimituksia. Helsinki: Suomalaisen Kirjallisuuden Seura, 2004, no. 964.
- [5] T. Kurki, *Yksilön ja ryhmän kielen reaaliaikainen muuttuminen: Kielenmuutosten seuraamisesta ja niiden tarkastelussa käytettävistä menetelmistä*, ser. Suomalaisen Kirjallisuuden Seuran Toimituksia. Helsinki: Suomalaisen Kirjallisuuden Seura, 2005, no. 1036.
- [6] L. Mustanoja, *Idiolekti ja sen muuttuminen: Reaaliaikatuutus Tampereen puhkielestä*, ser. Acta Universitatis Tamperensis. Tampere: Tampere University Press, 2011, no. 1605.
- [7] H. Ojansuu, *Suomen lounaismurteiden äännehistoria I: Vokaalioppi: Descendentti esitys*, Helsinki, 1901.
- [8] M. Airila, *Äännehistoriallinen tutkimus Tornion murteesta: Murteen suhdetta suomen muihin murteisiin silmälläpitäen*. Helsinki: Sana, 1912.
- [9] J. Laurosela, *Foneettinen tutkimus Etelä-Pohjanmaan murteesta*, ser. Suomi. Helsinki: Suomalaisen Kirjallisuuden Seura, 1922, no. V: 1.



- Mää varasim meille hotellihuoneel las palmasista kesäks.
- Mie varasim meille hotellihuoneen las palmasista kesäks.
- Kirjoita ja sano omin sanoin:



1st 2nd 3rd 4th 5th



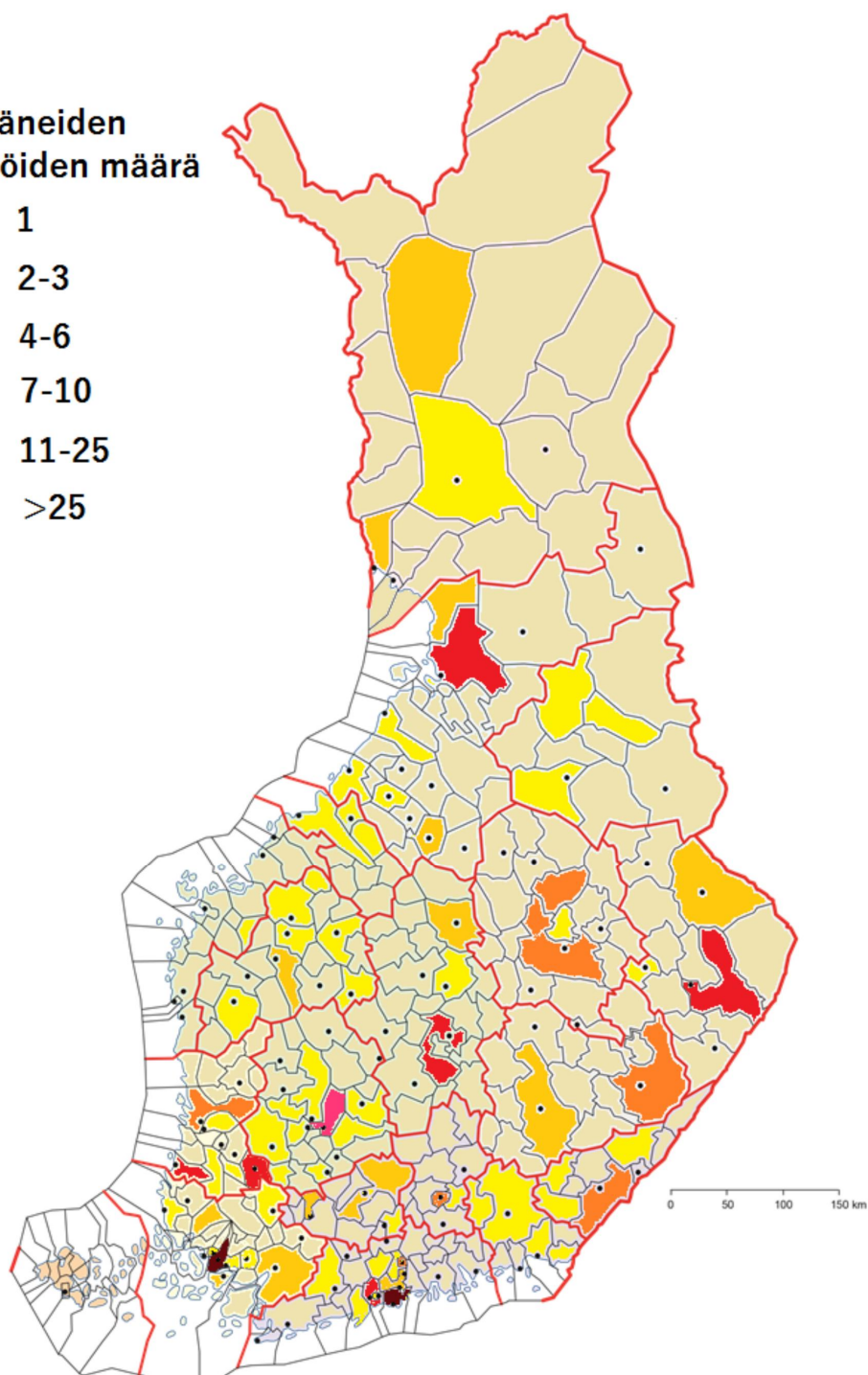
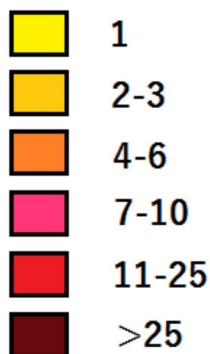
Haluat ostaa marjoja kauppiaalta. Kysy, mitä marjoja hänellä on myytävänä. Esitä kysymys omin sanoin niin kuin esittäisit sen omalla arkikielelläsi tutulle kauppiaille/myyjälle.



Esitä kysymys omin sanoin niin kuin esittäisit omalla arkikielelläsi tutulle kauppiaille/myyjälle

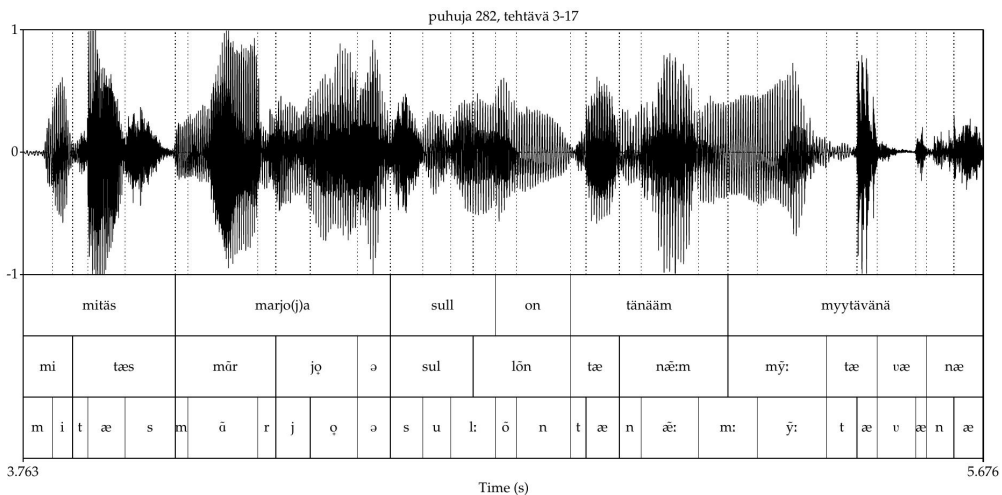
Kuva 2: Tehtävät asetelmana (videonselitystehtävää lukuun ottamatta). Ylävasemmalla lomakeskustelu, yläoikealla kuvien eron selitys, keskellä karttatehtävä ja alla torikeskustelu.

Äänittäneiden henkilöiden määrä

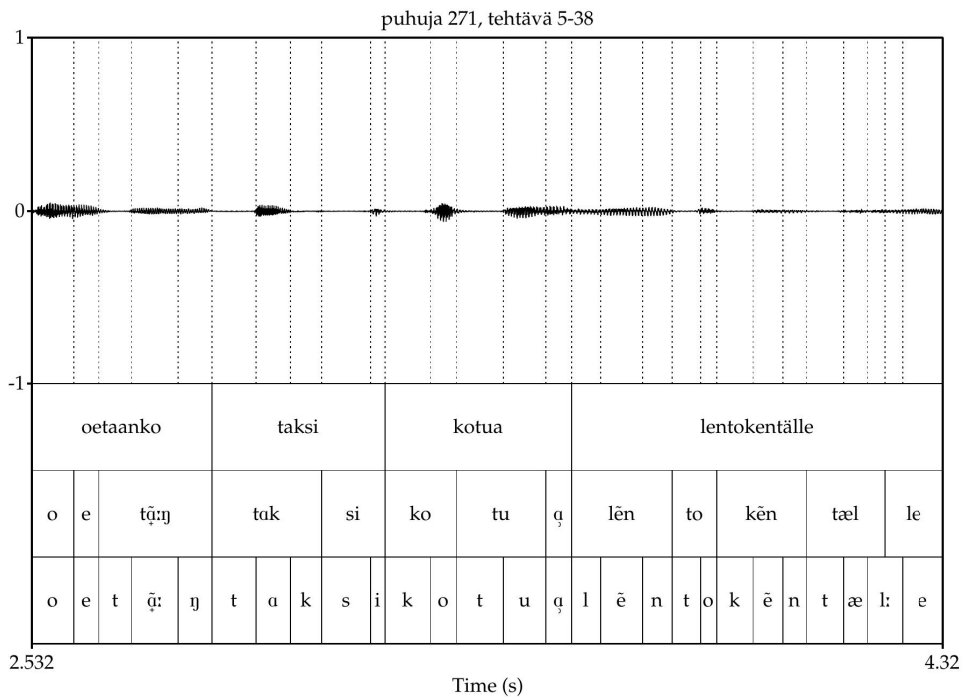


© MAANMITTAUSLAITOS. LANTMÄTERVERKET 2015

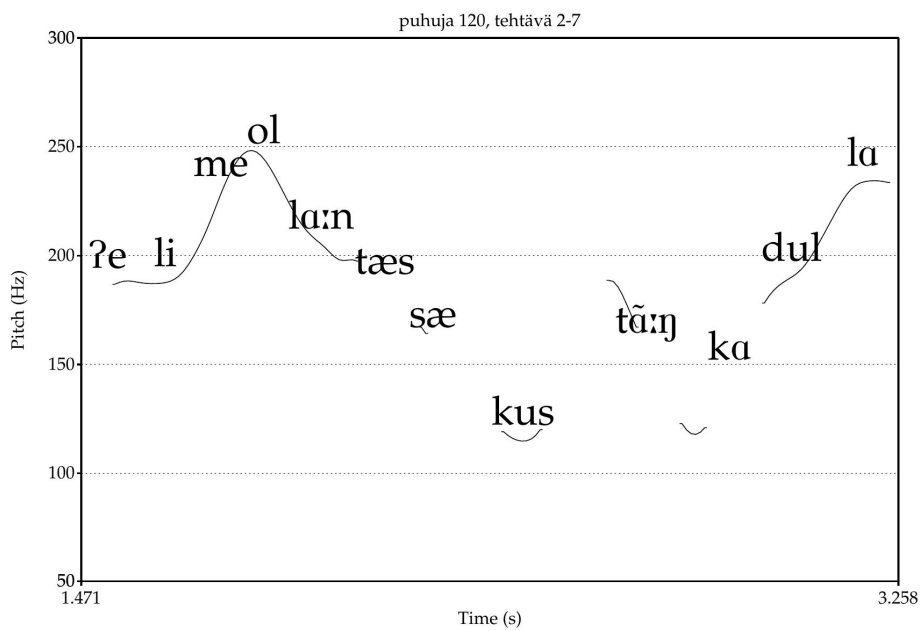
Kuva 3: Kartuntakartta.



Kuva 4: Esimerkki erinomaisesta - vaikka osin ehkä jo aavistuksen leikkaantuvastakin - äänitystasosta. (Tehtävä: torikeskustelu.)



Kuva 5: Esimerkki heikosta äänitystasosta. (Tehtävä: lomamatkakeskustelu.)



Kuva 6: Esimerkki puhunnoksenloppuisen nousevan intonaation käytöstä karttaselitystehtävässä.

ISBN 978-952-60-6279-2 (pdf)
ISSN-L 1799-487X
ISSN 1799-487X (painettu)
ISSN 1799-4888 (pdf)

Aalto-yliopisto

Signaalinkäsittely ja Akustiikka
www.aalto.fi

**KAUPPA +
TALOUS**

**TAIDE +
MUOTOILU +
ARKKITEHTUURI**

**TIEDE +
TEKNOLOGIA**

CROSSOVER

**DOCTORAL
DISSERTATIONS**