TAMPERE
UNIVERSITY
PRESS

Fonetiikan päivät 2008

XXV Fonetiikan päivillä Tampereen yliopistossa
11.–12.1.2008 pidetyt esitelmät

Michael L. O'Dell & Tommi Nieminen, toim.

Tampere Studies in Language, Translation and Culture, Series B 3

Tampere University Press
Tampere 2009

Taitto:
Michael L. O'Dell ja Tommi Nieminen

ISBN 978-951-44-7580-1
ISSN 1795-1208

Tampere University Press
Tampere 2009

Sisältö

Toimittajien saatesanat

Järjestyksessä 25. Fonetiikan päivät järjestettiin Tampereella 11.–12. tammikuuta 2008. Kyseessä oli neljäs kerta, kun Fonetiikan päivät pidettiin Tampereen yliopistossa: edelliset kerrat olivat vuosina 1974, 1981 ja 1994.

Osallistujia päivillä oli seitsemisenkymmentä ja esitelmiä ja postereita kaikkiaan 24, joista 18 julkaistaan tässä kokoelmassa. Lisäksi seuraavat kuusi esitelmää ja posteria esitettiin päivillä:

- Ville Hautamäki: Maximum a posteriori adaptation of the centroid model for speaker verification
- Tomi Kinnunen: Puhujantunnistus modulaatiospektripiirteillä
- Anna Lantee & Tommi Nieminen: Monitulkintaisten yhdyssanojen sivupaino
- Tuomo Raitio, Antti Suni, Martti Vainio & Paavo Alku: Äänilähteen käänteissuodatukseen perustuva HMM-pohjainen suomen kielen puhesynteesi
- Riikka Ruonamo, Anna-Maija Korpijaakko-Huuhka, Risto Kontio & Marjo Rönkkö: Suusyöpäleikkauksen vaikutus puhemotoriikkaan ja konsonanttien ääntämiseen
- Jyrki Tuomainen, Vicky Knowland, Mike Coleman & Stuart Rosen: Attention modulates audio-visual speech integration

Päivillä järjestettiin myös paneelikeskustelu fonetiikan oppiaineen tilanteesta Suomessa. Paneelin puheenjohtajana toimi emeritusprofessori Antti Iivonen ja sihteerinä Riikka Ullakonoja. Panelisteina toimivat Jussi Niemi (professori, Joensuun yliopisto), Patrik Scheinin (professori, käyttäytymistieteellisen tiedekunnan dekaani, Helsingin yliopisto), Sirkka Saarinen (professori, Turun yliopisto), Jouni Isoaho (professori, Turun yliopisto; varapuheenjohtaja, KITES ry.) ja Marja-Liisa Niemi (opetusneuvos, opetusministeriön korkeakoulu- ja tiedeyksikkö).

Paneeli alkoi professori Iivosen alkusanoilla, mitä seurasi kunkin panelistin puheenvuoro. Lopuksi sana oli vapaa. Tavoitteena oli saada aikaan rakentavaa keskustelua sekä tavoitella yhteisymmärrystä oppiaineen toiminnan selkeyttämiseksi ja parantamiseksi Suomessa. Kooste paneelista löytyy Tampereen yliopiston kieli- ja kään-

nöstieteen laitoksen sivuilta osoitteesta `http://www.uta.fi/laitokset/kielet/yht/tutkimus/fp2008/paneeli.html`.

Kiitämme lämpimästi järjestelytoimikunnan muita jäseniä, Anne-Maria Laukkasta, Kari Leinosta ja Teija Waaramaata, sekä opiskelija-avustajia Jarmo Haapaharjua, Hennariikka Kairannevaa, Maria Kunnasta, Anna Lanteeta ja Minna Saarta. Kiitokset myös kaikille päivien osallistujille ja artikkelien kirjoittajille antoisasta tieteellisestä dialogista.

Tampereella marraskuussa 2008

Michael O'Dell Tommi Nieminen

Michael O'Dell & Tommi Nieminen (eds.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 7–16.

Placing Zyryan vowels in the vowel chart

Dennis Estill
University of Helsinki

Abstract

In Zyryan, a Komi language spoken in the northern part of European Russia, word stress generally falls on the first syllable, there are seven vowels, and when stressed these are, with one possible exception /o/ and /ö/ which are relatively close to each other, fairly evenly spread out on the vowel chart. Therefore, Zyryan has quite a typical vowel arrangement. When these vowels are unstressed the pattern changes, and four of these /i/, /y/, /o/ and /ö/, fall into a position otherwise occupied by another vowel. All vowels, with the exception of /a/, centralise on the F1 axis, with /a/ centralising on the F2. From the results of the experiment described below, the acoustic parameters related to centralisation in Zyryan are, to a limited extent duration, to a more general extent intensity and, although possibly speaker-related, fundamental frequency (F0). This conclusion was reached on the basis of an experiment, the results of which were obtained from measurements taken from the recorded speech of four Zyryan speakers. These participants read the same text, and this contained at least 100 words. In their speech, these informants represented the standard language dialect, or a dialect close to the standard language. This paper also provides data concerning the combinations of acoustic parameters characteristic of individual speakers.

Keywords: Zyryan, vowels, centralisation

1 Introduction

The purpose of this article is to describe an analysis designed to measure Zyryan vowels acoustically and locate them on the vowel chart. A further aim of the exercise is to determine which factors are related to the centralisation of unstressed Zyryan vowels.

Zyryan is a Uralic language and one of the two (or three, depending on one's convictions concerning what is to be classified as a language) Komi languages. Most Zyryan speakers are located in the Republic of Komi, in the north-eastern corner of European Russia, and the number of Komi speakers in Russia is 293,406, of which

about 130,000 are Zyryans, according to the latest census.[1] The remainder are mostly Permyak Komis, although there is also a small minority of Komi Yaźva speakers. The closest linguistic relative of Komi is Udmurt, and these together form the Permic language group.

The literary language is based on the dialect spoken in the capital area, Syktyvkar, and was introduced as such by G. S. Lytkin (1838–1875). There are approximately ten major dialects. A description of Zyryan word stress from the chronological point of view can be found in Lytkin (1970) and Estill (2006). Word stress nowadays is predominantly on the initial syllable, although if the speaker prefers to stress another syllable, this would not generally be regarded as incorrect by native speakers (Lytkin 1955, Estill 2006).

Table 1: Zyryan system of vowels according to V. I. Lytkin (1955).

i	y	u
e	ö	o
	a	

The traditionally accepted Zyryan (also Komi) system of vowels is shown in Table 1. Like most other Finnic languages Zyryan avoids strings of consonants particularly at the beginnings of words and the possible syllabic phoneme combinations are V/VC/CV/CVC. Zyryan has a wide range of consonants, 26 in all. Like the other Permic languages, Zyryan is non-quantitative (this concerns both vowels and consonants). Unlike Udmurt, Komi, although very inflectional, has generally developed shorter words during the course of its separate historical development.

While little experimental research has been carried out respecting the Komi vowel system as a whole, a comparison has been made between the vowels of Zyryan and those of Finnish[2] (Savela 1999). For the purposes of Savela's study stressed and unstressed vowels were not differentiated, whereas this article treats stressed and unstressed as two separate categories in order to describe the centralisation process. Finnish and Zyryan both have word stress on the initial syllable.[3] The contact language to have had the greatest influence is Russian, as might well be expected. It is

[1] These figures have been taken from the last census, 2000 (http//:www.perepis2000.ru). The round figure of 130,000 was reached by deduction. For the first time it would now seem that the Permyak population outnumbers the Zyryan. In view of the complications with censuses, care must be taken in interpreting the figures.

[2] Finnish has eight vowels. In addition to the vowels described for Zyryan, Finnish has an open front vowel /ä/.

[3] While word stress may fall on any syllable in Zyryan, that is, it may be termed preferential, in practice word stress can today be described as occurring regularly on the first syllable (Estill 2006).

hoped that this paper will help clarify the nature of the vowels, whether and to what extent centralisation takes place and, if it does, what the related acoustic parameters are.

2 Informants

The informants for this experiment were JG, LC, OM and OO. All were females between the ages of 20–30 permanently resident in the Republic of Komi, and all were either university students or teachers. Even though the informants were all well educated and thus not entirely representative of the population as a whole, it must be remembered that an experiment of the kind that was carried out for this study demands excellent reading capabilities. It was also felt that by using persons of the same sex (and education for that matter) problems of formant interpretation might be made easier. Three speakers were from the same dialect area, that is, верневычегодский. The other informant spoke another dialect, присыктывкарский. All of these dialect areas are, nevertheless, located in the vicinity of Syktyvkar and the speech of all the informants was close to the standard dialect.

3 Description of experimental procedure

All seven vowels were measured acoustically in stressed position as either the nuclear vowel in stressed syllables in a polysyllabic word or as the nuclear vowel in a monosyllabic word. Sentential stress was not taken into consideration. These vowels, with the exception of /e/,[4] were also measured in unstressed position. Measurements were made of F1 and F2 for determining the placement of the vowels, and of duration, intensity and fundamental frequency (F0) for studying those features related to centralisation. The same text read by all informants consisted of at least 100 words (the number varied slightly according to speaker), and the informants were, of course, quite unaware of the purpose of the experiment. Further measurements were taken if it was considered that a particular stressed or unstressed vowel was not sufficiently well represented in the material, as was the case with unstressed /u/. The text extract used was chosen at random from a modern Zyryan novel Эжва педымса зонка by V. Timin (2000). The method used in the experiment was not the same as that of Savela (1999) who presented his informants with pre-prepared short sentences for later analysis. For this experiment it was felt that the reading should be as natural as the conditions allow, although at the same time controlled in such a way that comparison between speakers would be possible. The recording room was a small auditorium at the Department of Finno-Ugrian Studies, University of Helsinki. All doors and windows were closed and electrical interference was kept to a minimum.

[4] In Zyryan /e/ is never unstressed.

A Plextalk recorder was used for recording and the sound (in WAV form) later calibrated for fidelity. The microphone used was an AKG D 660S and the acoustic measurements made using the Praat 4.0 program.

4 Results of the experiment

A vowel chart for Zyryan based on the results of the experiment is shown in Figure 1. This agrees well with Table 1, although it will be observed that /ö/ is a little further back and thus closer to /o/, than rough descriptions imply. Otherwise, the vowels are quite linearly aligned on the F1 axis. The figure also demonstrates that Zyryan vowels are, with the exception of /a/, all close or fairly close and that the mid-open position is not characteristic of any vowel. The positions of the vowels shown on the chart are based on averages and do not, of course, reflect the wide variation that occurs in practice. However, only by dealing with averages can comparisons between the placement of stressed and unstressed vowels be made.

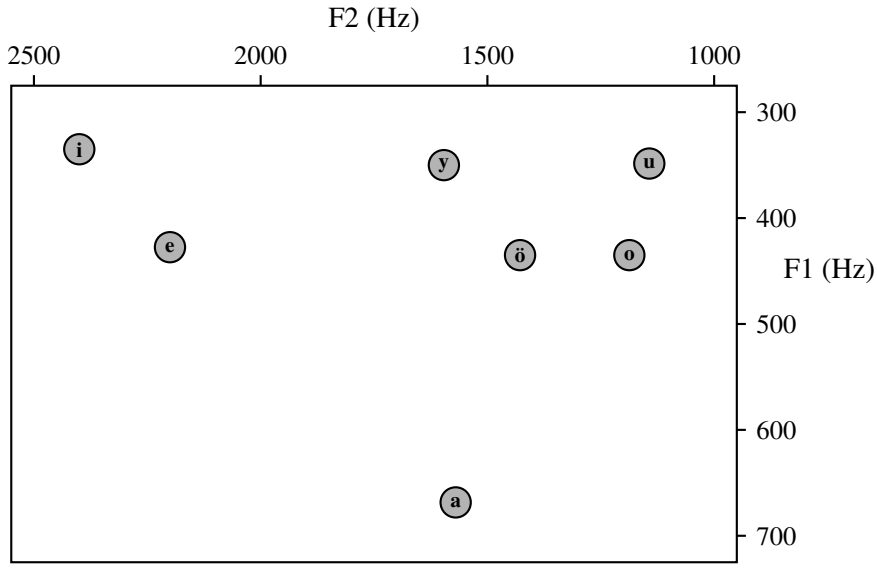


Figure 1: Placement of Zyryan vowels in the vowel chart based on measurements obtained from four informants

Figure 2 shows both the position of stressed vowels on the vowel chart in Zyryan and is thus the same as Figure 1 in this respect, and the position of the unstressed vowels. Several interesting features can be observed. Firstly, in the case of all vowels except /a/ movement from stressed to unstressed vowel takes place on the F2 scale,

with little or no change of position on the F1 axis. The greatest change in position is that of /i/ which moves from the normal position of a stressed /i/ in Zyryan into the centre of the chart. On the other hand, the back close vowel /u/ only changes its position slightly. Otherwise, centralisation is to be observed in the case of all Zyryan vowels with the exception of /e/, which is always, as stated earlier and at least as far as this investigation goes, unstressed.

Secondly, the re-placement of unstressed vowels causes "collisions," even when averages are considered. Unstressed /i/ and /y/ meet, as do stressed /o/ and unstressed /ö/. However, it will be observed that some vowels tend to retain their own distinctiveness when unstressed, that is, /a/, /u/ and /ö/.

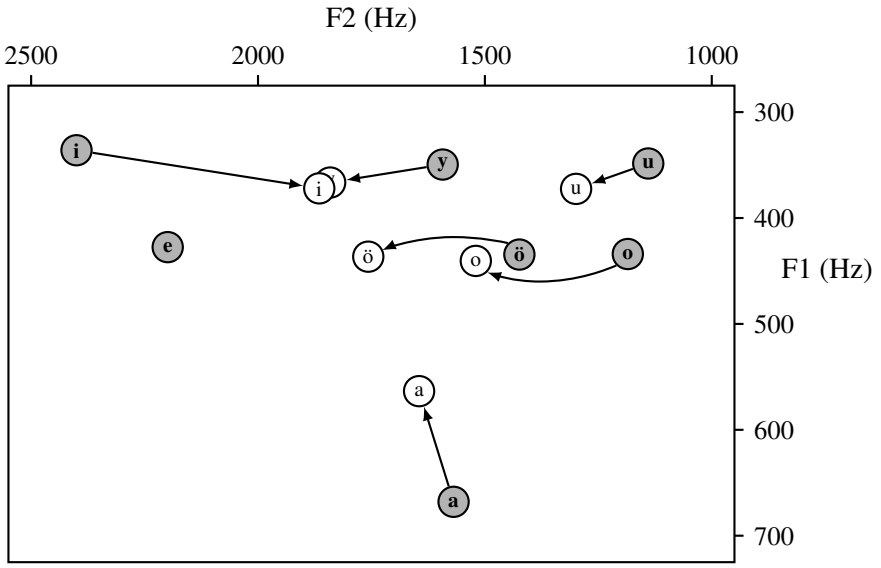


Figure 2: Position of stressed and unstressed vowels in Zyryan vowel chart with arrows indicating movement of vowel position in chart as the result of the loss of stress. Based on measurements obtained from four Zyryan informants.

The question next arises as to what acoustic parameters can be found that differentiate stressed and unstressed vowels. The characteristic most generally associated with the feature of vowel centralisation, or reduction, is vowel length, yet considering this to be the reason per se may be misleading. Other factors must be considered, too. While centralisation is shown in vowel charts as the change in relationship between F1 and F2, its production is determined by vowel length, pitch and loudness of voice (Laver 1994, p. 157). I shall therefore now turn to consider the acoustic parameters of these three features as they occur in Zyryan vowels. The first of these is duration.

4.1 Duration

Table 2 presents a summary of the duration of both stressed and unstressed vowels for the four speakers with summary totals in the last two columns. The average lengths of unstressed vowels as a percentage of stressed vowels are also shown.

There were individual peculiarities in the average duration of vowels as uttered by the speakers. For example, OM had noticeably shorter vowels and OO's pronunciation was longer than the others. With two exceptions the average duration of stressed vowels was longer than that of the unstressed.

Table 2: Average duration of vowels in ms for four Zyryan speakers and percentage ratio between unstressed vowel and stressed together with totals.

speaker		/a/	/e/	/i/	/o/	/u/	/y/	/ö/
JG	stressed (ms)	98.7	77.4	84.0	89.8	100.9	69.5	77.7
	unstressed (ms)	82.1		72.9	75.3	54.4	64.3	70.9
	ratio (%)	83.3		86.8	83.9	53.9	92.5	91.2
LC	stressed (ms)	99.2	100.0	93.5	104.0	97.4	79.8	97.3
	unstressed (ms)	83.7		61.1	60.7	76.9	65.9	70.6
	ratio (%)	84.4		65.3	58.0	79.0	82.6	72.6
OM	stressed (ms)	80.4	68.5	65.8	72.3	72.6	63.2	67.5
	unstressed (ms)	71.6		60.5	73.8	50.2	59.1	63.7
	ratio (%)	89.1		91.9	102.1	69.1	93.5	94.4
OO	stressed (ms)	122.9	99.5	112.4	106.2	102.0	78.2	103.2
	unstressed (ms)	92.2		71.1	87.6	65.6	82.0	82.3
	ratio (%)	75.0		63.3	82.5	64.3	104.0	79.7
total	stressed (ms)	100.7	86.0	89.3	92.8	94.7	72.4	87.2
	unstressed (ms)	82.3		66.9	73.2	61.0	67.5	72.0
	ratio (%)	81.7		74.9	78.9	64.4	93.2	82.6

Table 3 shows the total average length of stressed and unstressed vowels for all speakers and the extent, as a percentage, to which the unstressed vowel is shorter than the stressed. This table reveals that the length of unstressed vowels was 79 per cent that of stressed vowels, on average. With the average length of an unstressed vowel being 72 ms, a JND (Lehiste 1970, p. 13) would be audible if a compared stressed vowel were 87 ms. Since stressed vowels were 91.1 ms in length on average, duration

clearly has some impact on the production of stress and, similarly, on the production of centralisation.

Table 3: Summary of length of stressed (+) and unstressed (−) syllables for four Zyryan speakers with −/+ shown as percentage. /e/ is not included.

stressed (+)	unstressed (−)	−/+
91.1 ms	72.0 ms	79.0 %

For the purposes of this study the monosyllabic vowels in the extract analysed were added to the stressed vowels in polysyllabic words and the averages are shown in Table 4. These figures demonstrate that mono- and polysyllabic word-stressed vowels are approximately the same in length and it can be inferred that together they provide a contrasting class to unstressed vowels.

Table 4: Division of vowels lengths shown as stressed in Table 3, according to whether word was monosyllabic or polysyllabic, for four Zyryan speakers.

	polysyllabic words (*a*)	monosyllabic words (*b*)	*b*/*a*
vowel length	88.1 ms	95.5 ms	108.4 %
n. of samples	387	158	

Word length then is to some extent a factor that would seem to have at least some bearing on centralisation.

4.2 Intensity

The extent to which a decrease in loudness affects vowel quality is summed up in Table 5, which shows the intensity of the vowels, measured in dB. These results are fairly uniform, although in the case of /u/ it is questionable whether a change in intensity between stressed and unstressed vowel occurs. Table 6 presents a summary of the data shown in Table 5.

According to Backus (1977), a general statement can be made that it takes a change of about one decibel to be heard as a difference. Tables 5 and 6 confirm that in Zyryan the JND for intensity was considerably exceeded for all speakers, and it can safely be said that intensity is noticeably greater in the case of stressed vowels than of unstressed, and that a drop in intensity is at least concomitant with centralisation.

Table 5: Details of intensity (dB) readings on syllables stressed and unstressed for four Zyryan speakers, according to phoneme.

speaker		/a/	/e/	/i/	/o/	/u/	/y/	/ö/
JG	stressed	71.3	73.4	72.4	72.7	73.0	73.2	73.9
	unstressed	69.6		70.3	71.8	71.7	70.0	70.0
LC	stressed	67.8	70.5	71.3	69.8	72.3	71.7	72.3
	unstressed	65.6		69.3	68.5	71.2	66.8	67.3
OM	stressed	69.2	69.8	71.9	69.0	69.4	71.2	74.5
	unstressed	67.7		68.6	68.1	70.7	68.9	67.4
OO	stressed	71.8	73.9	75.0	72.3	72.5	73.9	75.3
	unstressed	69.9		71.9	69.3	72.8	71.9	70.5
total	stressed	69.9	71.9	72.6	70.9	71.8	72.5	74.0
	unstressed	68.3		70.0	69.5	71.6	69.4	68.8

Table 6: Summary of intensity (dB) of stressed and unstressed vowels for four Zyryan speakers. Totals are also shown. /e/ is not included. Based on measurements of 885 vowels (of which 334 stressed and 551 unstressed).

	JG	LC	OM	OO	total
stressed	72.6	70.1	70.3	73.0	71.5
unstressed	70.0	67.1	68.1	70.9	69.0

4.3 Fundamental frequency

The third acoustic parameter for consideration is fundamental frequency (F0), which is heard as pitch. Measurements of F0 were made for all of the informants and the results are shown in Table 7. Table 7 is a summary and the details are found in Table 8.

The problem with assessing the role of F0 is that there is no consensus concerning what the optimal JND might be. It would seem that at the very least a JND would be one semitone, although the threshold could be much smaller, as Martti Vainio's studies of Finnish prosody suggest (Vainio 2001). Bearing this in mind, the level of F0 for the informants taking part in this experiment would appear to indicate the involvement of F0 as a feature related to centralisation, although the great variation

Table 7: Summary of fundamental frequency (Hz) on stressed and unstressed syllables for four Zyryan speakers with corresponding semitone differences. Totals are shown. /e/ is not included. Based on 927 vowels (of which 338 stressed and 589 unstressed).

	JG	LC	OM	OO	total
stressed (Hz)	290.5	268.1	255.8	272.0	271.5
unstressed (Hz)	266.1	235.3	242.2	253.9	249.1
difference (semitones)	1.52	2.26	0.95	1.19	1.49

Table 8: Details of fundamental frequency (Hz) readings on vowels stressed and unstressed for four Zyryan speakers, according to phoneme.

speaker		/a/	/e/	/i/	/o/	/u/	/y/	/ö/
JG	stressed	283.8	281.1	282.7	279.4	290.6	305.1	312.4
	unstressed	264.6		279.7	266.9	296.9	261.0	260.7
LC	stressed	247.5	266.8	258.3	267.4	301.5	284.2	272.8
	unstressed	267.4		259.7	233.1	262.0	231.8	231.4
OM	stressed	256.8	244.2	260.7	244.7	266.5	257.7	270.3
	unstressed	239.2		246.6	235.7	267.7	243.0	241.1
OO	stressed	268.0	273.6	272.8	260.5	294.1	271.5	297.7
	unstressed	247.1		270.9	240.1	284.3	258.0	247.8

in the data suggests this could be somewhat speaker-related.

5 Conclusions

The vowel charts compiled on the basis of the acoustic measurements of the Zyryan vowels showed that the traditional placements of the vowels in the chart still hold. These results were based on the stressed vowels. When the unstressed are added to the chart, centralisation becomes apparent, such that some vowels move into positions occupied by other vowels. Thus, unstressed /i/ and /y/ merge, and unstressed /o/ moves into the position already occupied by stressed /ö/.

In this study the acoustic parameters involved as being related to centralisation were also considered. The factors taken into account were duration, intensity and F0.

From the results it was determined that the duration of vowels is longer when under stress although, it would seem, not to the extent that might be expected. The presence of greater intensity in the case of stressed vowels, on the other hand, was significant and this was found to be the case for all informants. Further investigation revealed that the effect of F0 as a feature of centralisation while being difficult to define could be speaker-related.

References

BACKUS, John 1977: *The Acoustical Foundations of Music*. 2nd Edition. New York: W. W. Norton.

ESTILL, Dennis 2006: Some observations on Zyryan word stress, past and present. – *Suomalais-ugrilaisen seuran aikakauskirja*, **91**: 71–80.

LAVER, John 1994: *Principles of Phonetics*. Cambridge: Cambridge University Press.

LEHISTE, Ilse 1970: *Suprasegmentals*. Cambridge MA: Massachusetts Institute of Technology Press.

LYTKIN, V. I. 1955: *Sovremennyj komi yazyk, fonetika, leksika, morfologiya*. Syktyvkar: Komi knizhnoye izdatel'stvo.

LYTKIN, V. I. 1970: Problema leksicheskogo udareniya v finno-ugorskikh yazykakh. – *ACTT Linguistic Academiae Scientiarum Hungaricae*, **20**(3–4): 245–263.

SAVELA, Janne 1999: Tutkimus komisyrjäänin ja suomen vokaalifoneemien rakenteesta. – *Sananjalka*, **41**: 167–176.

TIMIN, V. 2000: *Èzhva pedymsa zonka*. Syktyvkar: Èsköm Izdatel'stvo.

VAINIO, Martti 2001: *Artificial Neural Network Based Prosody Models for Finnish Text-to-Speech Synthesis*. Helsingin yliopiston fonetiikan laitoksen julkaisuja 43. University of Helsinki.

Michael O'Dell & Tommi Nieminen (eds.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 17–32.

Inter- and intra-speaker variability of LTAS and F0 in GSM and microphone speech

Tuija Niemi-Laitinen[1] & Kirsi Harinen[2]
[1]Crime Laboratory/NBI, [2]University of Helsinki

Abstract

The aim of this study was to find out differences in speech analysis results when GSM speech and microphone speech are compared. Limits for inter- and intra-speaker variability were also studied. A reading passage was recorded simultaneously with a microphone and GSM phone. Five male and six female speakers read one story twice. The speaking fundamental frequency (F0) statistics as well as Long-Term-Average Spectrum (LTAS) characteristics were analysed with the Praat program.

The results show that F0 average and median values are higher with GSM speech than microphone speech. The difference for F0 average value is about 1–4 Hz for male speakers and 3–14 Hz for female speakers. On the other hand, intra-individual difference between two consecutive reading sessions was 2–5 Hz for male and 0.25–7 Hz for female speakers.

LTAS-analysis shows that the radio channel transmission can create random artefacts to sound files. GSM distortion can be seen especially in the frequency band 2000–3400 Hz. Both visual inspection and correlations of the spectra reveal that there is more variation in consecutive recordings of the same reading passage in GSM recordings than in microphone recordings. The amount of this variation varies depending on the random nature of the radio channel transmission. The same channel intra- and inter-individual correlation values (standardised and centered) of LTAS do not overlap (with the whole material), excluding the worst case of GSM distortion.

Keywords: fundamental frequency, Long-Term-Average Spectrum, GSM

1 Background

1.1 GSM speech

GSM codecs used in phones vary in different countries. One of the following codecs may be used: RPE-LTP, 13 kbit/s, VSELP, 5.6 kbit/s (not used in Finland very

much), EFR, 12.2 kbit/s, AMR-NB, 4.75–12.2 kbit/s (8 different speeds), AMR-WB (9 different speeds, used in 3G, but may also be used in GSM phones, but not with all the 9 speeds).

A bit rate does not always explain the quality of the codec: The EFR-codec is better in quality than the RPE-LTP codec although the signal is more compressed. The radio channel transmission changes in a function of time. If a GSM phone is in a moving car, the situation could be much worse: even the bit rate may change during one GSM conversation. The smaller the bit rate, the bigger the changes between the original (such as a DAT recording) and the GSM recording are. The radio channel affects most the spectral representation i.e. the energy distribution along the frequency axis. Differences occur mainly in the upper part of the spectrum (over 2 kHz). This may be due to several reasons. The spectrum starts to decline from 2 kHz (with voiced segments) and it is possible that there is more noise when the bit rate is low. The spectrum is flattened over 2 kHz, and measuring e.g. higher formants might not be possible. It is also possible that the compression (codec) has changed during the test and consecutive recordings introduce a different amount of digital quantisation noise.

The radio channel creates errors into the bit flow, and the GSM channel coder tries to correct these errors. Although the bit rate has been the same during the recordings, it is possible that recordings differ from each other due to the bit errors. When the signal is compressed (which is always the case with GSM speech), it is possible that a certain segment may differ more in (consecutive) GSM recordings than in (consecutive) DAT recordings. Two very close compressed signals differ from each other more that two uncompressed signals do in a real situation.[1]

According to Künzel (2001) the most serious problem caused by GSM transmission is distortion induced by the data-reduction algorithm used for breaking down and reassembling the speech signal at either end of the transmission path. The telephone filtering technique (band-pass 350–3400 Hz) affects most the lower and upper parts of speech that are outside the range of band-pass area. It means that for example, the first formant and higher formants as well as voice quality characteristics are affected most. Also phonemes such as [s] and [f] that have energy higher than the upper limit of telephone filtering are affected quite dramatically. Künzel (2001) made an experiment studying the influence of telephone transmission on the measurement of formant frequencies. He found out that the first formant is affected most: it is shifted upwards. Künzel used land-line telephone (ISDN transmission line) in his study.

1.2 Fundamental frequency in GSM speech

Fundamental frequency is a speech parameter that has been studied in various contexts (e.g. Atkinson 1976, Braun 1995, Künzel 2000, Rose 2002). It is a well known

[1] Paavo Alku, personal communication, 15.8.2005.

fact that F0 can vary a lot intra-individually and that it is also easy to disguise. Still, it is a robust speech feature in forensic phonetics. It has been tested that in automatic speaker recognition, combining F0 analysis results with spectral features (MFCC) improves the accuracy of the recognizer in noisy conditions (Kinnunen & González-Hautamäki 2005).

1.3 Long-Term Average Spectrum in GSM speech

LTAS is a widely used, almost always available and easily measurable parameter in speaker recognition. However, its interpretation and numeric expression in comparisons is more difficult, because it cannot be (easily and reliably) expressed with only one numeric value. Finding an efficient measure for comparing the spectra is a challenge, too.

LTAS shows the energy distribution of any particular sound along the frequency axis. The advantages of LTAS are that is saturates sufficiently quite quickly (at its best) and it is usually always available and easy to measure. It is also text-independent (if the sample is long enough) and quite efficient and consistent within one speaker (low intra-individual variation). On the other hand, the disadvantages of LTAS are that its profile is very similar with most of the population (quite low inter-individual variation) and there are some facts that have to be acknowledged (because they can cause variation in LTAS) before starting to do the LTAS-analysis and especially before interpreting the results of it. It is also quite difficult to find the parameters (other than similarity and distance measures) to express the results of LTAS analysis and compare them numerically.

LTAS carries information about the speaker's voice quality. Unfortunately a) at least until these days the researchers have not been able to extract the voice quality components from other information (Nolan 1999) and b) the voice quality parameters are considered to situate at higher frequencies that are filtered out in GSM speech. For example, voice quality features such as breathiness or creak can be seen in LTAS (especially in spectral tilt, Biemans 2000), but if these features/attributes are weak, it is possible that LTAS cannot differentiate them from the so-called normal voice or from each other (Wolfe & Martin 1997). Recording distance, recording level and/or speaking volume may also have the same kind of effects to LTAS (Nordenberg & Sundberg 2003, Keller 2005). In addition, both telephone band pass filtering and random distortion effects of the radio channel transmission make interpreting of LTAS results even more challenging. Background noise and other disturbances in the speech signal can also affect LTAS.

Even though expressing the spectrum itself numerically is easy, it is much more difficult to find methods to quantitatively compare several spectra with each other. The most widely used measures are based on either similarities/dissimilarities or distances of the spectra. The most common measure of similarity is the Pearson corre-

lation coefficient and the most common distance measure is the Euclidian distance (ED). There are also lots of other distance measures, such as the Mahalanobis distance, but the algorithms can be quite complex ones. Only the correlation measure is used in this study, because it yields better results than ED with this material.

1.4 Centering and standardising

The problem with the LTAS correlation is that all the values can be quite high, because the common shape of the spectra is very much alike within the whole population due to its similar origin mechanisms (source-filter-theory). To eliminate (or at least to reduce) this effect the problem is solved here by centering and/or standardising the raw points used in correlation calculation. Centering is done by subtracting the sample mean from every observed raw point by frequency bands. Standardising is done by dividing the centered scores by the standard deviation of the sample. Standardising can also be called as "normalising" or "z-score transformation". Mathematically the scale is converted so that the transformed scores will necessarily have a mean of zero (mean = 0) and a standard deviation of one (stdev = 1). The z-scores can be computed a) from a population or b) from a sample. Here z-scores are understandably computed from the sample, because we do not know the mean or standard deviation of the population (in the same conditions, etc.), so we use the following formula: $\tilde{x}_i = (x_i - \bar{x}_i)/\sigma_x$.

It should be noted that the correlations that have been computed with the transformed scores give much lower values than if the correlations were computed with the raw points. For the above mentioned reasons the sample (its mean and standard deviation) affects the results so that in some comparison in box plot representation the deviating value can be marked as an extreme (more than 3 box lengths from upper or lower edge). In another comparison the deviating value can be marked as an outlier (1.5–3 box lengths from the upper or lower edge). It depends on the sample which one of the used methods (standardising or centering) functions better.[2]

2 Study of LTAS and F0 in GSM and microphone speech

2.1 Speech material, speakers and recording arrangements

In this research, GSM speech and microphone speech were recorded simultaneously. The aim was to study how reliably LTAS and F0 analysis can be performed on GSM speech data. For this study, a total of 11 speakers were recorded, namely 6 female and 5 male speakers. The age of the speakers varied from 26 to 63 (mean 39). All speakers have Finnish as their mother tongue, and no one was bilingual. They come from different parts of Finland and all but one work at the Dept. of Speech Sciences

[2] See e.g. http://espse.ed.psu.edu/statistics/Chapters/Chapter6/Chap6.html.

at the University of Helsinki. Two consecutive recordings of the same paragraph (in Finnish, about 2–3 minutes) were made to find out the intra-individual differences.

One part of the speech data was recorded with a GSM phone and stored on a computer. Microphone speech was recorded simultaneously on a DAT recorder with a condenser microphone (AKG C451E). All the recordings were made at the Department of Speech Sciences, University of Helsinki. Recording arrangements were as follows (see Figure 1). The microphone speech was recorded in a quiet, partly soundproof office room. The speaker was sitting on a chair and leaned on a table holding a GSM phone in one hand. The microphone was placed on the table. The distance between the speaker and the microphone was approximately 10 centimetres. The microphone was attached to a DAT recorder in the next room. The distance between the DAT recorder and the sending GSM was approximately 2 metres, and there was a thick wall between the GSM apparatus and the recorder. The DAT recorder was placed in another room in order to avoid the GSM interference.

A Tascam DA-DI digital audio tape recorder was used to record the microphone speech. The sample rate and the bit rate of the recordings were 44.1 kHz and 24 bits. The speaker spoke to the sending GSM phone (Nokia 6100) at the same time as to the microphone. The receiving GSM phone (Nokia 5350) was situated in a third room. This GSM phone was attached to a computer with a specially made cable through Roland UA-30 USB audio interface. SoundForge 6.0 was used to record the incoming call. GSM speech files were stored in a wave format in 44.1 kHz, 16 bit, linear. The DAT recordings (microphone speech) were transferred to the same PC in a wave format in 44.1 kHz, 16 bit, linear.

The duration of the samples varied from 124 to 212 seconds. The second reading took usually shorter time than the first one, because the speakers were already familiar with the material and could read the text faster.

2.2 Methods

The Praat program (version 4.2.25) with a special script developed for it was used for the F0 measurements. The Praat program performs pitch analyses based on the autocorrelation method (see Praat Manual, www.praat.org). The whole reading passage was taken for the F0 analysis (124–212 seconds long, depending on the speaker's speech rate). The measured F0 parameters include average, median, minimum and maximum as well as standard deviation values. The Praat script calculates the duration of measured samples and also gives the number of voiced samples out of all the samples used for calculations. This script is developed so that it uses an adaptive frequency scale with pre-analysis.

The spoken material was handled as a whole chunk in Praat's LTAS-analysis, and in this study the whole passage was taken into the analysis. The sub-band bin width was 43 Hz. The similarities and dissimilarities of the spectra were compared by visual

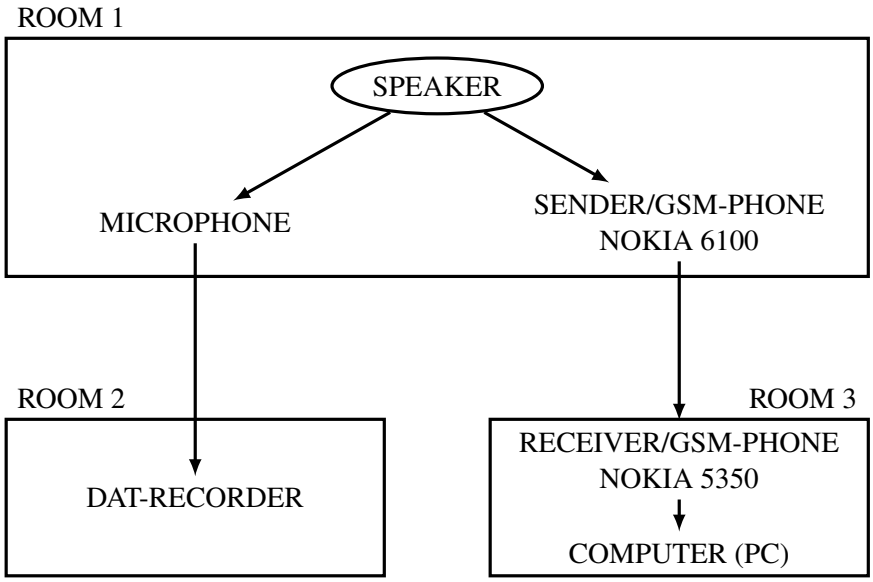


Figure 1: Recording situation for the study.

inspection and with correlations. The '1' in DAT1/GSM1 refers to the first recording and the '2' in DAT2/GSM2 refers to the second recording. Comparisons between the channels, GSM and the DAT, were either intra- or inter-individual. Centering and standardising of the raw points of the spectra were used to eliminate the "big trend" of the spectra.

3 Results: LTAS

The main, but quite expected, result is that there is usually more variation in the GSM recordings than in the microphone recordings. This variation is due to the random properties of the radio channel transmission and it occurs in upper frequencies of the telephone band. In one of these 11 samples, variation between consecutive repetitions in the GSM recordings was very large even though there was not any variation in the simultaneous recordings of the microphone speech (see Figure 2). This was quite striking and important result and it could not have been found out if we had only the GSM recordings without comparable microphone recordings, which is usually the case in forensic fieldwork. This kind of variation could also be interpreted wrongly as intra-individual. On the other hand, there can also be quite large inter-individual differences in the amount of variation. Among some speakers the intra-individual

variation is markedly larger than among others. Unfortunately the majority of the speaker dependent dispersion is also located in the upper frequencies.

In our material female speakers as a group showed much more homogenous results than male speakers. In the female group the radio channel distortion effects were also minimal.

3.1 LTAS—visual inspection

Visual inspection is not an efficient enough method to compare the spectra. The differences on the intensity levels can be quite large, which on the other hand does not affect the correlation calculation. Visual inspection is necessary, because it reveals the differences in the shape and slope of the spectra.

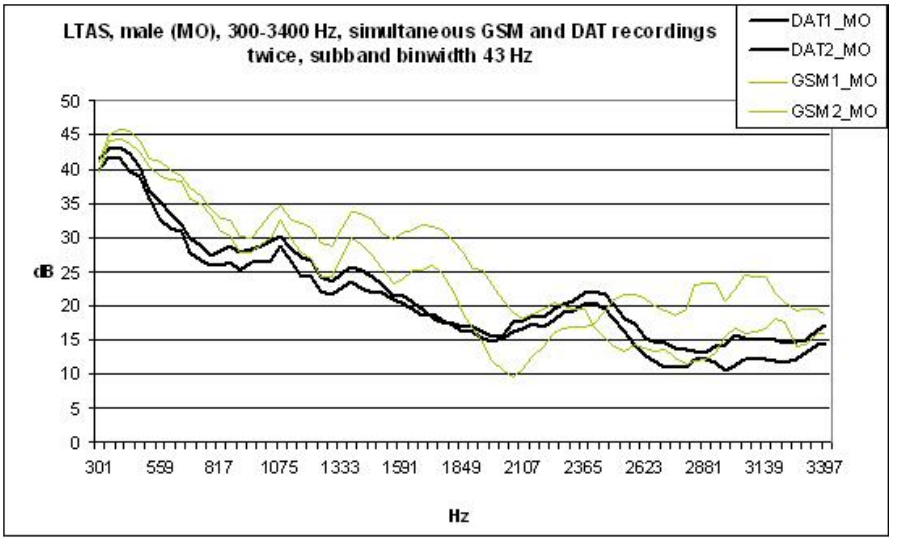


Figure 2: Variation in LTAS in two consecutive recordings of GSM and DAT. This figure shows the "worst case" of the random effect of radio channel transfer.

3.2 LTAS correlation measurements

Although there are many things that could deteriorate the results of LTAS comparison, we achieved very good results in respect to the whole material with correlation (centered and standardised points) in comparing intra- and inter-individual variation. This included both the microphone and GSM-speech.

Figure 3 shows the results as box plot presentation in DAT-recordings. The situation is much the same in GSM-recordings. There is only one extreme value in the

presentation of the GSM-speech: this is the result of a male speaker (MO, also seen in Figure 2), which is the worst case of the radio channel transmission effect.

Table 1: The numerical information of the intra- and inter-individual variation of LTAS of 11 (5 male and 6 female) speakers in microphone speech. Correlations of LTA-spectra are calculated of centered points.

	N	Min	Max	Mean	SD	Range
Intra-individual	11	0.83	0.98	0.91	0.05	0.15
Inter-individual	220	−0.70	0.73	−0.09	0.36	1.43

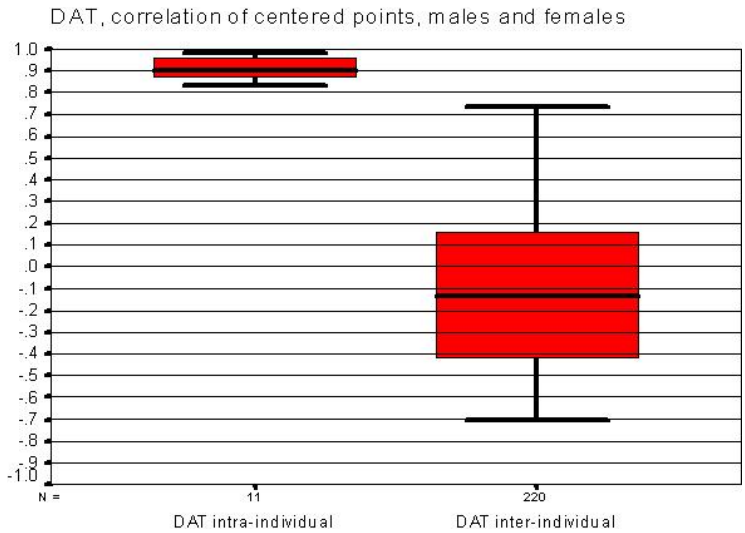


Figure 3: This box plot representation and the corresponding Table 1 show the intra- and inter-individual variation of LTAS of 11 (5 male and 6 female) speakers in microphone speech. Correlations of LTA-spectra are calculated of centered points. From total number of 11 speakers, 11 intra-individual and 220 inter-individual comparisons were made for this analysis.

As can be seen in Figure 3 and Table 1, there is no overlap with the intra- and inter-individual groups in the microphone recordings (DAT). On the other hand, in the GSM-recordings there is one extreme value −0.29, which in real forensic fieldwork would be problematic. The next worst intra-individual GSM correlation is 0.85 and the inter-quartile range is very small. The range and standard deviation of the inter-individual correlations are quite the same with the GSM and microphone speech.

Intra-individual groups would behave similarly, if one deviating value (−0.29) would be excluded.

Table 2: The numerical information of the intra- and inter-individual LTAS variation of 11 (5 male and 6 female) speakers measured from the GSM-speech. Correlations of LTA-spectra are calculated of standardised points.

	N	Min	Max	Mean	SD	Range
Intra-individual	11	−0.29	0.94	0.79	0.36	1.23
Inter-individual	220	−0.70	0.77	−0.09	0.33	1.47

3.3 Grouping of the LTAS results

When LTAS of the speech of different channels is compared (male and female speakers separately) with centered or standardised values, correlations have to be calculated separately for both channels. The results can then be combined in a graphical representation (a box plot). The groups to be compared are:

DAT intra:	intra-individual comparison in microphone (DAT) speech
GSM intra:	intra-individual comparison in GSM-speech
DAT vs. GSM intra:	intra-individual comparison of different channel speech
DAT inter:	inter-individual comparison of microphone (DAT) speech
GSM inter:	inter-individual comparison of GSM-speech
DAT vs. GSM inter:	inter-individual comparison of different channel speech

In an ideal situation the results of the groups 1 and 2, as well as in the groups 4 and 5 should be exactly the same, as they are from the simultaneous (but different channel) recordings. In an ideal situation all the correlations in group 3 should be 1.

Figure 4 shows that one of the male speakers has a deviating intra-individual value and that he has been marked as an outlier in this representation. In Figure 4 also two other values have been marked as outliers in the GSM inter -group. The most interesting group in this comparison is DAT vs. GSM intra. In this group the speakers are compared with themselves, only the recordings are via different channel. The minimum value of this group is the same as in the group GSM intra, but the inter-quartile range is situated much lower in the correlation scale, and the median of the group is approximately 0.24. This value in the group GSM intra is nearly 0.9. On the other hand, the same inter-quartile range and median values are clearly higher in the scale than in all of the inter groups.

In the group of female speakers the situation is slightly different (Table 4): There are neither outliers nor extremes. Both of the intra groups are very homogenous, the

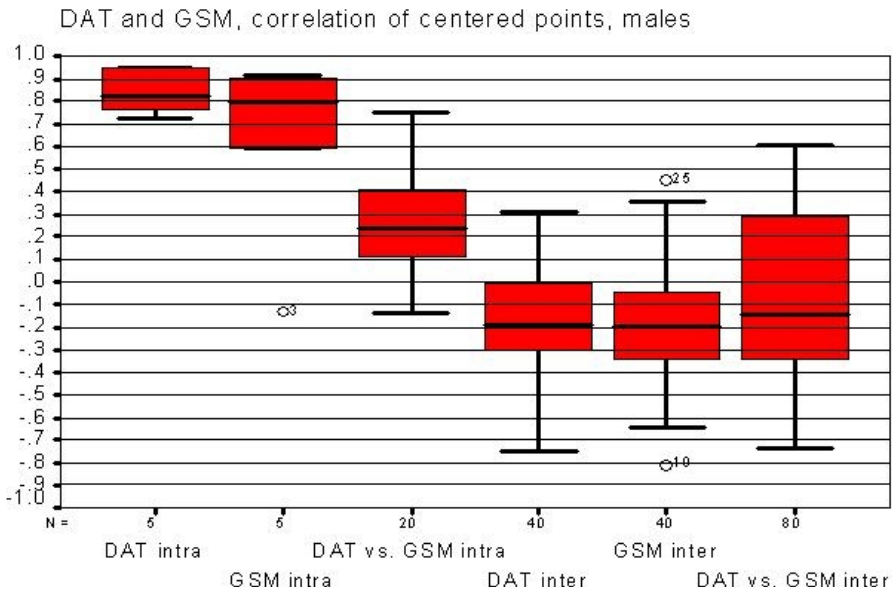


Figure 4: This figure and the corresponding Table 4 show the different comparisons of LTAS correlations with centered points for 5 male speakers.

dispersion is very small and the lowest value of the group intra GSM (0.79) overlaps only minimally with the highest value of all the inter groups (0.80). The median and the mean of the different-channel comparison of the same speaker (DAT vs. GSM intra) are almost the same as in the male group, but the inter-quartile range is larger. Even though there are more higher correlations in the between-channel comparison in the group of female speakers, there are also more lower correlations than in the group of male speakers. This means that samples via different channels are sometimes quite comparable, sometimes not, but which is the case in each situation, is random. It can be clearly seen that the division into two groups, intra- and inter-individual based on the mean of each group, is very distinctive. The mean values do not overlap with each other (with a very good marginal). In speaker recognition research the mean value is not very useful because every single value is important.

4 Results: F0

Statistical analysis of fundamental frequency of speech was expressed in different ways in this study. Mean, median minimum, maximum and standard deviation of the average F0 were measured.

Table 3: Different comparisons of LTAS correlations with centered points for 5 male speakers.

	N	Min	Max	Mean	SD	Range
DAT intra	5	0.72	0.95	0.84	0.10	0.23
GSM intra	5	−0.14	0.92	0.61	0.44	1.05
DAT-GSM intra	20	−0.14	0.75	0.27	0.27	0.89
DAT inter	40	−0.75	0.31	−0.19	0.29	1.06
GSM inter	40	−0.81	0.45	−0.19	0.27	1.26
DAT-GSM inter	80	−0.74	0.61	−0.06	0.35	1.34

Table 4: Different comparisons of LTAS correlations with centered points for 6 female speakers.

	N	Min	Max	Mean	SD	Range
DAT intra	6	0.84	0.94	0.89	0.04	0.10
GSM intra	6	0.79	0.94	0.88	0.05	0.14
DAT-GSM intra	24	−0.40	0.77	0.25	0.41	1.17
DAT inter	60	−0.76	0.55	−0.19	0.35	1.31
GSM inter	60	−0.92	0.80	−0.18	0.51	1.71
DAT-GSM inter	120	−0.68	0.65	−0.05	0.34	1.33

Fundamental frequency characteristics were first analysed to find out the intra-individual variation between two consecutive reading sessions. The microphone speech results were then compared to the results of the GSM recordings. For male speakers, the results are shown in Figure 5.

It can be seen in Figure 5 that the intra-individual variation between session 1 and 2 (aver-DAT1 and aver-DAT2) is quite small for every speaker, about 2–5 Hz. On the other hand, the difference between the DAT and GSM recordings show that the difference is even smaller than the intra-individual variation, about 1–4 Hz for male speakers.

The F0 average difference between two consecutive recordings is bigger with the GSM than DAT recordings. The reason for this is that GSM coding affects the signals in different ways according to the recording time.

For female speakers, the results are shown in Figure 6. It can be seen that for every speaker the intra-individual variation between session 1 and 2 (aver-DAT1 and

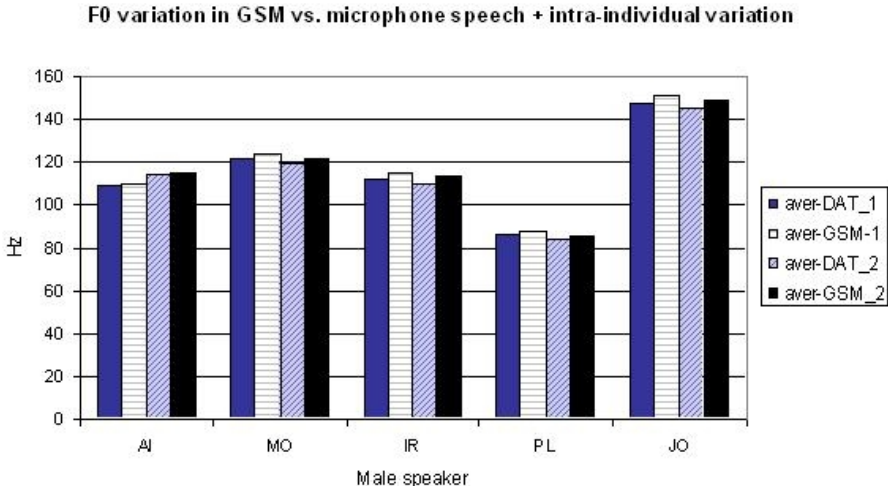


Figure 5: F0 average variation for five male speakers. Two consecutive reading sessions with a microphone (DAT1 and DAT2) as well as two simultaneous recordings with a GSM phone (GSM1 and GSM2).

aver-DAT2) is also small for female speakers as it was for male speakers, about 0.25 to 3 Hz (1st) and 1–7 Hz (2nd reading session). On the other hand, the comparison between the DAT and GSM recordings shows that the difference is about 4–14 Hz for the first reading session and 3-9 Hz for the second reading session.

It seems that the GSM has more influence to the analysis results of female speakers than male speakers. With female speakers, the difference between the DAT and GSM recordings were bigger than intra-individual variation between two consecutive reading sessions.

Statistical comparison of the differences and their significances in F0 analysis results between microphone and GSM recordings are shown in Table 6.

The number of voiced samples measured was higher for the DAT recordings than the GSM recordings; about 1000 points. F0 average and median values act similarly. The GSM values were higher than the microphone-DAT values. The difference is not significant. F0 standard deviation is bigger with the GSM speech, but the difference is not significant either for male or female speakers. The fewer points are taken into the analysis, the higher the F0 average. It means that with the GSM speech the script finds more high points and leaves out the lower ones. The only significant differences were found in F0 range both for male and female speakers as well as in F0 maximum for female speakers (one-way ANOVA, $p < 0.005$).

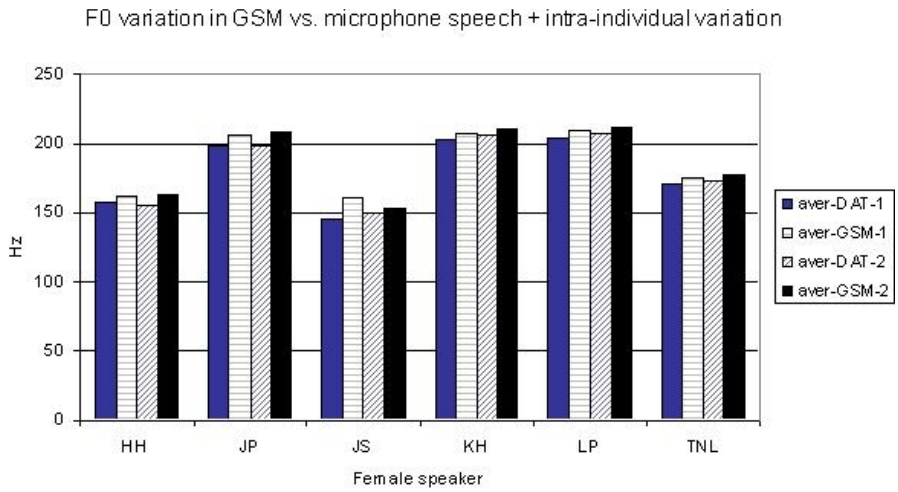


Figure 6: F0 average in Hertz for six female speakers. Two consecutive reading sessions with a microphone (DAT1 and DAT2) as well as two simultaneous recordings with a GSM phone (GSM1 and GSM2).

5 Conclusions

In this study, the GSM channel effects were studied using LTAS and F0 parameters. The intra-individual variation of LTAS correlation is very small in both microphone (DAT) and GSM-recordings and the correlations are high. There may, though, be cases where the GSM-distortion totally corrupts the results so that in the consecutive GSM recordings there is much variation while there is not any in microphone recordings. This effect is random and can be quite striking, but it cannot be anticipated. In this sense the channel distortion affects more LTAS-analysis than F0 analysis. When centered and/or standardised scores are used for LTAS correlation calculation and the material is grouped so that the same channel recordings of male and female speakers are grouped together, there is no overlap with the intra- and inter-individual variation. In this study LTAS of microphone-recordings seems to be very efficient parameter and classifies well the two consecutive samples as coming from the same speaker.

The average fundamental frequency value for every tested speaker differed very little intra-individually. The second reading passage had slightly higher F0 average and median values than the first one. Inter-individual variation, instead, was high. The voices of the highest male and the lowest female speakers could be very well recognized in forensic case work.

The GSM channel specific differences in F0 analysis results were not significant despite the F0 range (females and males) and maximum values (females). For ev-

Table 5: Average differences in several F0 parameters and their statistical significances in the results of DAT (microphone) and GSM recordings measured with SPSS program.

	DAT-GSM	Females		Males	
Parameter	Direction of diff.	Aver. diff.	Signif.	Aver. diff.	Signif.
N of calc. points	DAT > GSM	1200	0.014	971	0.113
F0 average	GSM > DAT	6.0 Hz	0.536	2.2 Hz	0.816
F0 median	GSM > DAT	4.2 Hz	0.690	1.5 Hz	0.864
F0 minimum	DAT > GSM	13.3 Hz	0.187	10.8 Hz	0.071
F0 maximum	GSM > DAT	46.3 Hz	0.003*	20.6 Hz	0.163
F0 range	GSM > DAT	59.6 Hz	0.002*	31.4 Hz	0.005*
F0 SD	GSM > DAT	4.4 Hz	0.038	2.0 Hz	0.154

ery speaker and in both reading sessions the average F0 was somewhat higher in the GSM speech than in the DAT recordings. This difference is about the same as the intra-individual difference between two consecutive reading sessions. The difference found in F0 average value between two consecutive recordings is bigger with the GSM than the DAT recordings. The reason for this is that GSM coding affects the signals in different ways according to the time of the recording. This is quite interesting result. If it is not known, what a certain GSM codec does to the speech signal, the differences in the analysis results may be explained in different ways. The differences might be either intra- or inter-individual, due to different channels (microphone-GSM), or due to the GSM phone type, or due to GSM coding at certain time. The last reason is very difficult to explain and its effect is uncontrollable.

A multiparametric approach for speaker identification could be the solution for this kind of problems found in this study. The artefacts that the GSM channel may create can be minimized when the results of several speech parameters are measured and pooled together (see e.g. Iivonen *et al.* 2003, Kinnunen *et al.* 2003; 2006, Hautamäki *et al.* 2007). It has also been tested that in automatic speaker recognition, combining F0 analysis results with spectral features (MFCC) improves the accuracy of the recognizer in noisy conditions (Kinnunen & González-Hautamäki 2005). New ideas are also presented about GMM and kernel functions (Lee *et al.* 2007).

The results of this study can be used in the forensic speech analysis to achieve more reliable results and it also helps in formulating conclusions. It seems that GSM-speech is less stable than microphone speech—at least in higher frequencies of the telephone band. The random effect of GSM channel variation to the LTAS analysis is worth consideration.

References

ATKINSON, J. E. 1976: Inter- and intraspeaker variability in fundamental voice frequency. – *Journal of the Acoustical Society of America*, **60**:440–445.

BIEMANS, Monique 2000: *Gender Variation in Voice Quality*. Utrecht: LOT Publications. http://www.lotpublications.nl/publish/issues/Biemans/index.html.

BRAUN, A. 1995: Fundamental frequency—how speaker-specific is it? – A. Braun & J.-P. Köster (eds.), *Studies in Forensic Phonetics*. Trier: Wissenschaftlicher Verlag Trier. 9–23.

HAUTAMÄKI, V., TUONONEN, M., NIEMI-LAITINEN, T. & FRÄNTI, P. 2007: Improving speaker verification by periodicity based voice activity detection. – *Proceedings of the International Conference on Speech and Computer (SPECOM'2007), Moscow, Russia*, vol. 2. 645–650.

IIVONEN, A., HARINEN, K., KEINÄNEN, L., KIRJAVAINEN, J., MEISTER, E. & TUURI, L. 2003: Development of a multiparametric speaker profile for speaker recognition. – M. J. Solé, D. Recasens & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, vol. 1. Universitat Autònoma de Barcelona, Spain. 695–698.

KELLER, Eric 2005: The analysis of voice quality in speech processing. – Gerard Chollet, Anna Esposito, Marcos Faundez-Zanuy & Maria Marinaro (eds.), *Nonlinear Speech Modeling and Applications: Advanced Lectures and Revised Selected Papers*. Berlin: Springer-Verlag. 54–73.

KINNUNEN, T. & GONZÁLEZ-HAUTAMÄKI, R. 2005: Long-term F0 modeling for text-independent speaker recognition. – *Proceedings of the International Conference on Speech and Computer (SPECOM'2005), Patras, Greece*. 567–570.

KINNUNEN, T., HAUTAMÄKI, V. & FRÄNTI, P. 2003: On the fusion of dissimilarity-based classifiers for speaker identification. – *Proceedings: Eurospeech 2003—Geneva*. 2641–2644.

KINNUNEN, T., HAUTAMÄKI, V. & FRÄNTI, P. 2006: On the use of long-term average spectrum in automatic speaker recognition. – *Proceedings: 5th International Symposium on Chinese Spoken Language Processing (ISCSLP 2006)*, vol. II. 559–567.

KÜNZEL, Hermann J. 2000: Effects of voice disguise on speaking fundamental frequency. – *The International Journal of Speech, Language and the Law*, **7**(2):149–179.

KÜNZEL, Hermann J. 2001: Beware of the 'telephone effect': The influence of telephone transmission on the measurement of formant frequencies. – *The International Journal of Speech, Language and the Law*, **8**(1):80–99.

LEE, K.-A., YOU, C., LI, H. & KINNUNEN, T. 2007: A GMM-based probabilistic sequence kernel for speaker verification. – *Interspeech, Antwerpen, Belgium, August 2007*. 294–297.

NOLAN, F. 1999: Speaker recognition and forensic phonetics. – W. J. Hardcastle & J. Laver (eds.), *The Handbook of Phonetic Sciences*. Blackwell Publishers Ltd. 744–767.

NORDENBERG, Maria & SUNDBERG, Johan 2003: Effect on LTAS of vocal loudness variation. – *Speech, Music and Hearing Quarterly Progress and Status Report*, **45**(1):93–100. http://www.speech.kth.se/qpsr/tmh/2003/03-45-093-100.pdf (11.10.2005).

ROSE, P. 2002: *Forensic Speaker Identification*. London: Taylor & Francis.

WOLFE, V. & MARTIN, D. 1997: Acoustic correlates of dysphonia: Type and severity. – *Journal of Communication Disorders*, **30**:403–416.

Michael O'Dell & Tommi Nieminen (eds.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 33–39.

Following speech changes in ALS: A call for co-operation

Tanja Makkonen[1,2] & Anna-Maija Korpijaakko-Huuhka[2]
[1]Tampere University Hospital, [2]University of Tampere

Abstract

The purpose of this study is to get more information about how communication and swallowing problems proceed in adults with Amyotrophic Lateral Sclerosis (ALS) during the follow-up and also to find out specific factors that may predict the progression pattern—or patterns—of communication difficulty and swallowing. In this presentation we focus on changes of speech and voice.

Keywords: communication disorders, Amyotrophic Lateral Sclerosis, speech

1 Background

ALS, also known by the name motor neuron disease, causes progressive speech and swallowing disorders. ALS is generally considered a relentlessly and rapidly progressive disease, which destroys motor neurons in both brain and spinal cord (Somer 2006). The deterioration proceeds individually, and little is known about factors predicting its course. Its etiology is unknown, and there is no cure for it.

The symptoms of ALS are generally divided into two groups: bulbar features and spinal features. In two thirds of the patients with ALS, the spinal deterioration affecting upper and lower extremity functions predominate. In one third of the patients, however, the disease starts with bulbar symptoms as the deterioration of the corticobulbar tract affects the innervations of muscles responsible for speech and swallowing functions. Dysarthria is one of the first symptoms of ALS in about 25–30 % of the patients (Yorkston *et al.* 1993, Duffy 1995, Freed 2000).

ALS affects both lower and upper motor neurons (Somer 2006). Upper motor neuron involvement affects neurons originating from cortex and ending in the anterior horn cells of the spinal cord. The typical upper motor neuron lesion results in slow and spastic movements and increased muscles tone which causes, for example, a strangled voice quality. The lower motor neuron involvement affects neurons originating from the anterior horn cells and ending in the muscles. The typical symptoms

of lower motor neuron involvement are atrophy, fasciculation and weakness, which can cause, for example, a breathy voice quality.

As the disease progresses, the patient's speech becomes less understandable because of dysarthria (e.g. DePaul & Brooks 1993, Dworkin *et al.* 1980, Kent *et al.* 1992, Langmore & Lehman 1994, Mulligan *et al.* 1994), the most common communication problem of patients with ALS. The patients with lower motor neuron involvement will demonstrate flaccid dysartria, and those with predominantly upper motor neuron involvement will show spastic dysarthria. These clear dysarthria types may be present only at the earliest stages of the disease. Because ALS represents a mixed lower and upper motor neuron disease, the most common motor speech problem of patients with ALS is mixed flaccid-spastic dysarthria (Duffy 1995, Yorkston *et al.* 2004, Freed 2000). Typically dysarthria influences all levels of speech production: phonation, respiration, articulation, resonance and prosody.

Along the progression of ALS, speech becomes less understandable, and the adults with ALS need means of augmentative and alternative communication (AAC) to interact efficiently and understandably. Although the linguistic skills and cognition of ALS patients are relatively intact, in selecting the most functional means of AAC, it is important to evaluate the neuropsychological status of the patient in order to gain the best possible match between the patient's needs and communication aids available.

Decline in speech rate is often the first symptoms of dysarthria in people with ALS (Watts & Vanryckeghem 2001, Kent *et al.* 1991, Mulligan *et al.* 1994). Speech intelligibility may remain good even if speech rate and maximum repetition rate decline (Nishio & Niimi 2000). Patients with ALS seem to use slowered speech rate as a compensatory strategy to maintain speech intelligibility when the ability to articulate gets worse. However, even when speech rate (pauses included in total time of speaking) declines, speech may remain understandable as long as articulation rate (speech time without pauses) is not affected; the slowing down of articulatory movements and the increased number and duration of pauses result in unavoidable decrease of speech intelligibility (Nishio & Niimi 2000). Speech intelligibility may decline rather fast. In a case reported by Watts & Vanryckeghem (2001), dysarthria progressed rapidly and resulted in the decrease of sentence intelligibility from 98 % to 10 % during the six months follow-up. Mulligan *et al.* (1994) have reported word intelligibility decrease from 95 % to 88 % during a six months period. From the acoustic attributes the F2 slope seems to be a sensitive index of lingual function (Kent *et al.* 1991; 1992) and useful index for making predictions about intelligibility impairments (Weismer *et al.* 2001) for people with ALS.

Articulatory changes can also be verified acoustically, for example, as reduced vowel space (Weismer *et al.* 2001, Turner & Tjaden 1995). Disruptions of some specific phonetic features, such as velopharyngeal valving, lingual functions for producing consonant contrasts of place and manner, and syllable shape, have been shown

to interact with the decrease of word intelligibility in dysarthric ALS speakers (Kent *et al.* 1992) Thus, complex relationships of speech intelligibility and acoustic measures have been reported but more research is needed.

The bulbar symptoms and declination of vital capacity (see Mulligan *et al.* 1994) may cause phonatory weakness and changes in the quality of voice, for example breathy or stained-strangled voice. Prolonged intervals, prolonged phonemes and inappropriate silences may result from respiratory changes (Yorkston *et al.* 2004). Dysarthric speakers with ALS might have increased phonatory instability and reduce phonatory limits (Raming *et al.* 1990). Phonatory characteristics might vary greatly among the cases (Strand *et al.* 1994).

2 Speech therapy for patients with ALS

In setting the goals for speech therapy, it is elementary to understand that there is no cure for ALS, and that often the speech function deteriorates very quickly, especially in the bulbar type of ALS (Watts & Vanryckeghem 2001). The main goal of speech therapy is to maintain the patients' ability to communicate using natural speech or alternative and augmentative communication (AAC) means. Speech therapy aims at helping adults with ALS to maintain their active communicator role. Speech therapy should focus on the maintenance of functional communication rather than on trying to reduce the speech impairment (Yorkston *et al.* 2004).

As the decline of speech function may be fast, it is important to start the intervention at an early state of the disease, as early as possible, to provide the patient with communication means that enhance the maintenance of quality of live. The unavoidable changes in speech may be compensated for by using different strategies (Yorkston *et al.* 1993). When the intelligibility of natural speech is starting to decline, speech must be accompanied by augmentative approaches to communication. Frequent monitoring of speech and early introduction of AAC means are essential in ensuring that decisions about the AAC-related technology for people with ALS are based on communication preferences (Ball *et al.* 2004).

3 Aims

The clinical work as speech and language therapists has motivated us to study communication and swallowing problems of adults with ALS. As the deteriorations proceed individually, it's difficult to know how rapidly communication and swallowing become impaired. In order to be able to plan the speech and language therapy process adequately, it is most important to know if there are some specific factors predicting the course of symptoms in these patients.

The purpose of this doctoral study is to get more information about how communication and swallowing problems proceed in adults with Amyotrophic Lateral Sclerosis (ALS) during the follow-up and also to find out what kind of means of AAC, including technological solutions, adults with ALS can use, and if these means are efficient. Based on these findings, we aim to find out specific factors that may predict the progression pattern—or patterns—of deteriorating communication difficulty. In this presentation we focus on changes of speech and voice.

4 Materials and methods

Participants in this study will be 20–30 adult patients with the diagnosis of probable ALS or ALS definitive. The follow-up starts at the first appointment with the speech therapists in the neurological department, and will last about two years for each patient. The research data will accumulate during clinical management of patients with ALS at the Department of Neurology and rehabilitation of the Tampere University Hospital. Follow-up frequency is every 3–4 months for patients with bulbar ALS and every 3–6 months for patients with spinal ALS. Data collection has started during the fall 2007.

To describe speech changes during the follow-up we are collecting versatile material. As there is no standard method for the assessment of speech and communication skills in patients with ALS, we chose to look at the symptom progression from four different angles: physiological, speech production, speech intelligibility and communication. Physiological changes are followed by using oral motor tasks and measuring maximum phonation time. Speech production tasks will include story telling based on cartoon frames, sentence and word production and diadochokinetic rate. Speech samples will be audio-recorded at each visit. Speech and articulation rate are measured from story telling and sentence production tasks. The severity of the speech disorder is assessed by using the ALS Severity Scale of Speech. Speech intelligibility of spontaneous speech will be perceptually evaluated on 4-point scale by a group of listeners. The listeners will also hear single sentences and words read by the patient with ALS. The proportion of correctly perceived words will serve as the second estimate of the speech intelligibility. The communicative ability of the patients will be evaluated by classifying the need and usage of means of augmentative communication. Modified CETI will be used to determinate the communicative effectiveness of each of the communication tools used in different social situations.

Parameters and their operationalisations:

Physiological oral motor functions on a 4-point scale, phonation time

Speech production speech rate, articulation rate, diadochokinetic rate, ALS Severity Scale of Speech (Yorkston *et al.* 1993)

Speech intelligibility perceptually evaluated on a 4-point scale by a group of listeners on the basis of spontaneous speech; the proportion of correctly perceived words based on the sentence and word production tasks

Communication skills a classification of the need and usage of AAC means (Yorkston *et al.* 1993); communication tools in use, Modified CETI (The Communication Effectiveness Index) (Lomas *et al.* 1989, Ball *et al.* 2004)

The data will mainly be analysed according to clinical criteria but acoustic analyses of speech and voice are also possible in this audiotaped material.

Audiotaped material:

- spontaneous speech in a story generation task (see Korpijaakko-Huuhka & Aulanko 1994)
- sentence production task from Speech Examination (Keller 1990; Finnish version Werner *et al.* 1994)
- word production task from Speech Examination (Keller 1990; Finnish version Werner *et al.* 1994)
- repeated monosyllables and syllable sequences (e.g. [pɑtɑkɑ], [tɑpɑkɑ])
- maximum phonation time (prolonged [ɑː])

We call on students and researchers within phonetics and vocology for co-operation!

References

BALL, L. J., BEUKELMAN, D. R. & PATTEE, G. L. 2004: Communication effectiveness of individuals with amyotrophic lateral sclerosis. – *Journal of Communication Disorders*, **37**: 197–215.

DEPAUL, R. & BROOKS, B. R. 1993: Multiple orofacial indices in amyotrophic lateral sclerosis. – *Journal of Speech and Hearing Research*, **36**: 1158–1167.

DUFFY, J. R. 1995: *Motor Speech Disorders: Substrates, Differential Diagnosis and Management*. St. Louis: Mosby.

DWORKIN, J. P., ARONSON, A. E. & MULDER, D. W. 1980: Tongue force in normal and dysarthric patients with amyotrophic lateral sclerosis. – *Journal of Speech and Hearing Research*, **23**: 828–837.

FREED, D. B. 2000: *Motor Speech Disorders: Diagnosis and Treatment*. San Diego: Singular.

KELLER, E. 1990: Instructions for scoring the speech examination (SE). Version 2.0, August 1990 (unpublished manuscript).

KENT, J. F., KENT, R. D., ROSENBEK, J. C., WEISMER, G., MARTIN, R., SUFIT, R. & BROOKS, B. R. 1992: Quantitative description of the dysarthria in women with amyotrophic lateral sclerosis. – *Journal of Speech and Hearing Research*, **35**: 723–734.

KENT, R. D., SUFIT, R. L., ROSENBEK, J. C., KENT, J. F., WEISMER, G., MARTIN, R. E. & BROOKS, B. R. 1991: Speech deterioration in amyotrophic lateral sclerosis: A case study. – *Journal of Speech and Hearing Research*, **34**: 1269–1275.

KORPIJAAKKO-HUUHKA, A.-M. & AULANKO, R. 1994: Auditory and acoustic analyses of prosody in clinical evaluation of narrative speech. – R. Aulanko & A.-M. Korpijaakko-Huuhka (eds.), *Proceedings of the Third Congress of the International Clinical Phonetics and Linguistics Association, 9–11 August 1993, Helsinki*, Publications of the Department of Phonetics, University of Helsinki 39. 91–98.

LANGMORE, S. E. & LEHMAN, M. E. 1994: Physiologic deficits in the orofacial system underlying dysarthria in amyotrophic lateral sclerosis. – *Journal of Speech and Hearing Research*, **37**: 28–37.

LOMAS, J., PICKARD, L., BESTER, S., ELBARD, H., FINLAYSON, A. & ZOGHAIB, C. 1989: The communicative effectiveness index: Development and psychometric evaluation of a functional communication measure for adult aphasia. – *Journal of Speech and Hearing Disorders*, **54**: 113–124.

MULLIGAN, M., CARPENTER, J., RIDDEL, J., DELANEY, M. K., BADKER, G., KRUSINSKI, P. & TANDAN, R. 1994: Intelligibility and the acoustic characteristics of speech in amyotrophic lateral sclerosis (ALS). – *Journal of Speech and Hearing Research*, **37**: 496–504.

NISHIO, M. & NIIMI, S. 2000: Changes over time in dysathric patients with amyotrophic lateral sclerosis (ALS): A study of changes in speaking rate and maximum repetition rate (MRR). – *Clinical Linguistics & Phonetics*, **14**: 485–497.

RAMING, L. O., SCHERER, R. C., KLASNER, E. R., TITZE, I. R. & HORII, Y. 1990: Acoustic analysis of voice in amyotrophic lateral sclerosis: A longitudinal case study. – *Journal of Speech and Hearing Disorders*, **55**: 2–14.

SOMER, H. 2006: Selkäytimen taudit ja oireyhtymät. – S. Soinila, M. Kaste & H. Somer (eds.), *Neurologia*. Helsinki: Duodecim.

STRAND, E. A., MILLER, R. M., YORKSTON, K. M. & RAMING, L. O. 1994: Differential phonatory characteristics of four women with amyotrophical lateral sclerosis. – *Journal of Voice*, **8**: 327–339.

TURNER, G. S. & TJADEN, K. 1995: The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. – *Journal of Speech and Hearing Research*, **38**: 1001–1013.

WATTS, C. R. & VANRYCKEGHEM, M. 2001: Laryngeal dysfunction in amyotrophic lateral sclerosis: A review and case report. – *BMJ Ear, Nose and Throat Disorders*, **1**(1).

WEISMER, G., JENG, J.-Y., LAURES, J. S., KENT, R. D. & KENT, J. F. 2001: Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. – *Folia Phoniatrica et Logopedica*, **53**: 1–18.

WERNER, S., TUOMAINEN, J. & LEHTIHALMES, M. 1994: The speech examination (SE). (unpublished Finnish translation).

YORKSTON, K. M., MILLER, R. M. & STRAND, E. A. 2004: *Management of Speech and Swallowing in Degenerative Diseases*. 2nd Edition. Texas: Pro-ed.

YORKSTON, K. M., STRAND, E., MILLER, R., HILLEL, A. & SMITH, K. 1993: Speech deterioration in amyotrophic lateral sclerosis: Implications for the timing of intervention. – *Journal of Medical Speech-Language Pathology*, **1**: 35–46.

Michael O'Dell & Tommi Nieminen (toim.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 41–48.

Puoliautomaattisten puhujantunnistusohjelmien suorituskyvyn vertailu

Tuija Niemi-Laitinen[1] & Mona Lehtinen[2]
[1]Rikostekninen laboratorio/KRP, [2]Helsingin yliopisto

Tiivistelmä

Puhujantunnistuksen automatisointi kiinnostaa monia tahoja. Turvallisuusala, biometria ja rikostutkinta hyötyvät tämän alan kehityksestä. Puhujantunnistuksen automatisointi ei välttämättä ole ratkaisu tuntemattomien puhujien ongelmaan, mutta se on yksi menetelmä muiden joukossa. Menetelmästä on hyötyä erityisesti silloin, kun joudutaan hakemaan tuntematonta puhujaa useiden kymmenien tai satojen ehdokkaiden joukosta. Jos taas halutaan tarkkaa analyysia muutaman puhenäytteen välillä, ovat tutkijan tekemät mittaukset, kuulonvarainen tutkimus sekä johtopäätökset eri virhelähteiden vaikutuksista näytteisiin edelleen tärkeä tekijä. Puhujantunnistuksen täysautomaattisuus on mahdollista biometriassa, puhujan todentamisen yhteydessä. Rikosten tutkinnassakin ohjelman avulla voidaan etsiä oikeaa ääntä tietokannasta, mutta näytteiden vertailuun se ei yksin riitä. Mukana täytyy olla myös tutkija, joka tekee muitakin mittauksia ja tulkitsee tulokset.

Keskusrikospoliisi yhdessä monen muun yritystahon kanssa on ollut mukana rahoittamassa TEKESin PUMS-projektia, jossa on kehitetty automaattinen puhujantunnistusohjelma, WinSProfiler (http://cs.joensuu.fi/pages/pums/).

Tässä testissä vertailtiin WinSProfilerin ja neljän kaupallisen ohjelman suorituskykyä ja tuloksia. Testimateriaalina käytettiin mm. autenttista rikoksiin liittynyttä äänimateriaalia. Kaikkiaan 61 miespuhujalta valittiin kaksi näytettä, joista toinen mallinnettiin tietokantaan ja toista käytettiin ns. tuntemattomana näytteenä. Jokainen näyte oli eri puhelusta, joten taustahälyt saattoivat vaihdella tilanteesta toiseen. Tarkoituksena oli selvittää, kuinka hyvin kukin ohjelma löytää tuntematonta puhenäytettä vastaavan saman puhujan puhenäytteen tietokannasta. Parhaimmillaan ohjelmat tunnistivat oikein 74 % näytteistä, huonoin ohjelma vain 29 %. Viiden ensimmäisen joukkoon oikea puhenäyte sijoittui parhaimmillaan 100 % tarkkuudella ja huonoimmillaan vain 52 % tarkkuudella. Algoritmeista MFCC vaikuttaa tämän testin perusteella parhaiten toimivalta.

Avainsanat: puhujantunnistus, MFCC

1 Taustaa

Tässä kuvailtava projekti on TEKES-rahoitteisen *Puhe- ja kieliteknologian uudet menetelmät ja sovellukset* -projektin (PUMS) osaprojekti. Keskusrikospoliisi oli kiinnostunut lähinnä puhujan- ja puheentunnistussovellusten kehittämisestä. Tässä osaprojektissa testattiin erilaisia puhujantunnistussovelluksia tarkoituksena vertailla niiden suorituskykyä. Myös ohjelmien käytettävyyttä ja soveltuvuutta poliisiorganisaation käyttöön arvioitiin. Seuraavissa kappaleissa esitellään käytetyt ohjelmat, näytteet ja tulokset.

2 Ohjelmat

2.1 Verrattujen ohjelmien ominaisuuksia

Testissä käytettyjä ohjelmia oli 5: WinSProfiler, VoiceNet, FreeSpeech, Batvox ja ASIS. Jälkimmäiset neljä ovat kaupallisia. VoiceNetin kehittäjä on pietarilainen Speech Technology Center, FreeSpeechin takana on israelilainen Verint, Batvox ja ASIS ovat espanjalaisen Agnition tuotteita. Jokainen ohjelma tarjoaa mahdollisuuden luoda tietokanta joko palvelimelle tai yksittäiselle työasemalle. Osa ohjelmista mahdollistaa eräajon, jolloin montaa eri näytettä voidaan vertailla koko tietokantaan kerralla. Jokainen ohjelma myös tuottaa vertailuista ns. score-listan. Ohjelmista Batvox ja ASIS ilmoittavat tulokset myös LR-muodossa (*likelihood ratio*, uskottavuusosamäärä), eli ne siis suorittavat bayesilaista päättelyä sen suhteen, onko kahdella näytteellä sama puhuja vai ei. Batvox myös piirtää LR-käyrät JPEG-kuvaksi. Ohjelmat käyttävät osittain eri algoritmeja. Näytteiden minipituuden, tiedostomuodon, sekä signaali/kohina-suhteen vaatimukset eroavat ohjelmien välillä. Jotkin ohjelmat kieltäytyivät käsittelemästä näytteitä, jotka eivät vastanneet tiettyjä vaatimuksia. Vertailuissa pyrittiin valitsemaan näytteitä, joita kaikki ohjelmat pystyisivät käsittelemään. Joistain ohjelmista oli käytössä vain demoversio, mikä vaikutti lopulliseen arvioon. Jokaisen ohjelman kehittelijätahoon oltiin yhteydessä testausten aikana. WinSProfilerista koekäytettiin myös sen UNIX-versiota SProfileria.

2.2 Taustamallinnus

Puhujantunnistusta tehtäessä luodaan yleensä taustamalli (UBM *Universal Background Model*), eli muodostetaan eräänlainen keskiarvopuhuja monista eri puhujista. Taustamallia tarvitaan puhujantunnistuksessa vertailukohdaksi, jotta puheen keskivertoiset, yleisesti esiintyvät piirteet voitaisiin normalisoida. Kaikissa testatuissa ohjelmissa vertailun tekeminen edellytti taustamallin luomista. Yleensä malleja luotiin kolme: mies- ja naismalli sekä sukupuoleton malli. Sukupuoleton malli on erittäin yleinen malli, joka luodaan samasta määrästä mies- ja naispuhujia. Tämä malli luo-

daan siitä syystä, että tuntemattoman puhujan sukupuoli ei välttämättä aina ole täysin selvä, joten ei myöskään ole varmaa, kumpaan taustamalliin tätä pitäisi verrata.

ASIS ja Batvox käyttävät tilastollisiin *t*- ja *z*-testeihin pohjautuvia menetelmiä tarkastaakseen, onko taustamalli sopiva ko. testiin vai ei.

Näissä kokeissa taustamallina käytettiin näytteitä, jotka oli kerätty edellisen TEKES-projektin aikana vuonna 2001.

2.3 Algoritmit

Puheen perustaajuuteen perustuvaa analyysia suoritetaan sekä WinSProfilerissa että VoiceNet-ohjelmassa. Pitkäaikaiskeskiarvospektrin LTAS:n laskenta on yhtenä ominaisuutena WinSProfilerissa.

WinSprofiler sekä ASIS ja Batvox käyttävät analyysissaan myös mel-kepstrikertoimia (MFCC). Nämä kertoimet muodostetaan ikkunoimalla signaali ja laskemalla ikkunoista Fourier-spektri. Spektri suodatetaan kolmiomaisilla ikkunoilla (ihmisen kuulokykyä mallintavaan mel-skaalaan perustuva suodatus). Jokaisesta suodinikkunasta saadut signaaliarvot keskiarvoistetaan ja keskiarvoista otetaan logaritmit. Tälle lukujoukolle tehdään vielä diskreetti kosinimuunnos (Harinen *et al.* 2006, Kinnunen 2005). MFCC on havaittu yhdeksi parhaiten puhujia erottelevaksi piirteeksi (Young 1996 Kinnusen 1999 mukaan). Samansuuntaisiin tuloksiin ovat päätyneet myös Harinen *et al.* (2006).

Puhujantunnistuksessa jokaista puhujaa kuvaa suuri joukko signaalista saatuja piirrevektoreita. WinSProfilerissa näiden piirrevektoreiden käsittelyyn voidaan soveltaa GMM:ää ja VQ:ta. GMM viittaa gaussilaiseen mikstuuramalliin (Gaussian Mixture Model), VQ:lla puolestaan tarkoitetaan vektorikvantisointia, tilastollista datapisteiden luokittelua (Fränti *et al.* 2007). Edellä mainitut ovat tiedonpakkausalgoritmeja. Niiden tarkoitus on käsitellä piirrevektoreita ilmaisemalla ne pienempänä joukkona, koodivektoreina (Kinnunen 1999; 2005).

3 Näytteet

Tutkimuksessa käytettiin sekä autenttisia rikosääninäytteitä että laboratorio-olosuhteissa nauhoitettua puhetta. Testeissä käytettiin pääosin GSM-suodatuksen läpikäynyttä puhetta. Näytteenottotaajuus oli 8 kHz, bittisyys 16, näytteiden kesto vaihteli kymmenestä sekunnista muutamiin minuutteihin.

3.1 Laboratorio-olosuhteissa tallennetut näytteet

Ensimmäinen osa-aineisto oli aiemmassa TEKESin USIX-projektissa tallennettuja puhenäytteitä (vuodelta 2001). Puhujat olivat eri-ikäisiä miehiä ja naisia, jotka edustivat eri murrealueita. Jokaiselta puhujalta otettiin mukaan kaksi näytettä: spontaania

puhetta (kuvakerrontaa) sekä lukupuhetta. Spontaanin puheen näytteet olivat 1–5 minuuttia pitkiä. Lukupuhunnan näytteiden kesto oli 2,3–4 minuuttia. Nämä näytteet oli äänitetty joko KRP:n rikosteknisessä laboratoriossa tai Helsingin yliopiston puhetieteiden laitoksella. Kaikkiaan näytteitä otettiin mukaan 34 mies- ja 52 naispuhujalta, yhteensä siis 68 ja 104 kpl. Näytteitä käytettiin lähinnä taustamallien tekemiseen ja joidenkin alustavien testien tekemiseen.

3.2 Autenttiset näytteet

Mukana testeissä oli myös KRP:n rikosteknisen laboratorion omia näytteitä. Projektiin mukaan otettiin miespuhujien yli kymmenen sekunnin mittaiset näytteet vuosilta 2000–2006. Näillä näytteillä tehtiin lähinnä kokeita, joilla selvitettiin WinSProfilerin eri versioiden välisiä eroja. Näytteitä käytettiin myös tehtäessä joitain tietokantaajoja.

Projektin edetessä mahdollistui myös uudempien autenttisten rikosääninäytteiden hankkiminen. Näytteet oli tallennettu uudemmalla tekniikalla ja ne olivat parempilaatuisia kuin edeltävät. Näitä näytteitä oli 122 kappaletta, kaksi jokaiselta 61:ltä eri miespuhujalta. Tässä artikkelissa mainitut tulokset on saatu testaamalla juuri tällä 61 puhujan aineistolla.

4 Testauksen kulku

Testauksen aluksi poistettiin testaukseen kelpaamattomat näytteet ja siistattiin mukaan otettavat näytteet sopivaan muotoon. Näytteet mallinnettiin jokaista ohjelmaa varten erikseen. Itse testauksessa ohjelmaa sitten tarkasteltiin monilla eri asetuksilla ja aineistoilla. Tuloslistojen käsittely on olennainen osa tulosten raportointia ja vie oman aikansa. Ohjelmat yleensä tulostavat listan ”raakoja” score-arvoja, joita sinällään on hankala tulkita. Tuloslistoja käsiteltiin eri skripteillä, jotta niistä saataisiin haluttu tulos eli oikein tai lähes oikein tunnistuneiden näytteiden määrä prosentteina. Saadut tulokset ja käytettävyysarviot taulukoitiin ja raportoitiin.

5 Tuloksista

5.1 Tekninen suorituskyky

Kaikki testatut ohjelmat antavat suorittamistaan vertailuista ns. score-arvon. Mitä suurempi tuo arvo on, sitä suuremmalla todennäköisyydellä vertailluilla näytteillä puhuu sama puhuja. Parhaimmillaan ohjelmat tunnistivat oikein 74 % näytteistä, huonoimmillaan tunnistustulos oli vain 29 %. Viiden ensimmäisen joukkoon oikea puhenäyte sijoittui parhaimmillaan 100 % tarkkuudella, huonoin ohjelma ylsi 52 %:iin.

Taulukossa 1 on ohjelmien antamia tunnistustuloksia.

Taulukko 1: Eri ohjelmien antamat tulokset autenttisella 61 puhujan aineistolla.

Ohjelman nimi	N	hylättyjä	sija 1	sijat 1–3	sijat 1–5
VOICENET	38	24	29 %	47 %	52 %
ASIS	51	11	67 %	86 %	92 %
FREESPEECH	61	0	74 %	98 %	98 %
WINSPROFILER	51	0	53 %	86 %	100 %

Batvox ja ASIS käyttävät samaa perusteknologiaa, joten tähän vertailuun valittiin ASIS sen nopeamman käytön vuoksi.

Aineisto koostui 61 puhujasta ja näytteet olivat autenttisia rikosnäytteitä (ks. 3.2). Kaksi ohjelmaa hyväksyi näytteistä vain osan. Esimerkiksi VoiceNet kelpuutti keston tai signaali/kohina-suhteen puolesta vain 38 näytettä, minimikeston ollessa kaikilla näytteillä 16 sekuntia. ASIS hyväksyi 51 näytettä 61:stä, koska rajana tietokantaan meneville tiedostoille on 35–40 sekuntia, testattaville 7 sekuntia. WinSProfiler ei periaatteessa hylkää yhtään näytettä, mutta haluttaessa verrata sitä lähinnä vastaavaan kaupalliseen tuotteeseen ASISiin valittiin samat näytteet kuin em. ohjelma oli analysoinut. WinSProfilerin kohdalla vertailuun valittiin parhaimman tunnistustuloksen antanut MFCC/GMM.

Eri algoritmeilla saavutettujen tulosten välillä on eroja. Esimerkki tästä näkyy taulukossa 2, jossa on eritelty WinSProfilerin eri algoritmeilla saavutettuja tulokset.

Taulukko 2: Ohjelman WinSProfiler version 2.11 eri algoritmien antamat tulokset. Näytteet laboratorio-olosuhteissa tallennettuja (N = 86).

Parametri	1. sija	sijat 1–5
LTAS	49,4 %	73,0 %
F0/GMM	28,8 %	69,2 %
F0/VQ	33,6 %	70,7 %
MFCC/GMM	74,3 %	86,8 %
MFCC/VQ	74,6 %	87,6 %
Fuusio	80,6 %	92,7 %

Taulukosta 2 huomataan, että paras yksittäinen algoritmi on MFCC. Huonoimpana puhujia erottelevana puheen piirteenä on tässä tapauksessa perustaajuus F0 yksinään mitattuna. Myöskään pitkäaikaiskeskiarvospektri ei yllä yhtä luotettavaan tun-

nistustasoon kuin mel- kepstrikertoimet. Kohta ”Fuusio” on WinSProfilerin tarjoama painotettu keskiarvo eri algoritmeilla saavutetuista tuloksista.

5.2 Käytettävyys

WinSProfilerin käyttöliittymä on yksinkertainen ja selkeä. Näytteiden ajo tietokantaan on verrattain vaivatonta ja nopeaa. Sadan näytteen analysoiminen loppuun asti sujuu alle tunnissa. Käytännöllinen ominaisuus on se, että ohjelma sallii monen vertailunäytteen lisäämisen tietokantaan samalla kertaa. Näytteiden vertailu on nopeaa. Tulosten analysointi on hitain ja työläin työvaihe. Jos on käytetty eräajo-ominaisuutta, ohjelma tulostaa score-listan, jossa on eritelty jokaisen vertailun tulokset kaikkien optioiden (MFCC+VQ, MFCC+GMM, F0+VQ, F0+GMM jne.) suhteen. Tämä score-lista on tekstitiedosto, jossa voi olla tuhansia rivejä. Listan jatkokäsittelyssä voidaan käyttää taulukkolaskentaohjelmaa. Dataa tulee kuitenkin paljon, joten sen käsittely on haasteellista ja hidasta. Lisäksi score-listat ovat jokseenkin vaikeatulkintaisia. Listat kannattaa jatkokäsitellä esimerkiksi skriptaamalla (Unix/Linux) tai jollain tehokkaalla tilasto-ohjelmalla (R, SPSS jne.).

FreeSpeech-ohjelma ei suorittanut vertailuja yhtä nopeasti kuin WinSprofiler, SProfiler tai VoiceNet. Myös näytteiden vieminen tietokantaan oli hidasta, vaikka ohjelman demoversiossa voidaan käyttää eräajomahdollisuutta. Tutkimuksen aikana ei saatu tietoa siitä, ajetaanko GUI-ohjelmassa näytteet tietokantaan yksitellen vai onko mahdollista suorittaa eräajo. Jos näytteet on ajettava tietokantaan yksitellen, käytettävyys heikkenee huomattavasti. Taustamallin luominen on FreeSpeechissä hitaampaa ja monimutkaisempaa kuin esim. WinSProfilerissa. Ohjelman tuottamien score-listojen käsittelyyn on oltava tietty valmius tämänkin ohjelman kohdalla.

VoiceNetin käytettävyys on tutkimuksen aikana syntyneen käsityksen pohjalta muita huonompi. Jokainen puhuja on lisättävä tietokantaan erikseen, mikä hidastaa prosessia. Käyttäjä ei voi vaikuttaa ohjelman asetuksiin juuri millään tavalla. Jokaiselle tietokantaan ajettavalle näytteelle on tehtävä ns. puhujakortti, joka liittää äänitiedoston johonkin henkilöllisyyteen (tässä tapauksessa näytteen arbitraariseen tunnistenumeroon). VoiceNet käyttäytyy siten, että jos hakuvaiheessa haettava näyte osoittautuu liian lyhyeksi tai huonolaatuiseksi, haku keskeytyy ja näyte hylätään. Haettavillekin näytteille voidaan jo ennen vertailua tehdä puhujakortit, jolloin ohjelma hylkää huonot ja lyhyet näytteet jo tässä vaiheessa. Tämä on kuitenkin tietyssä mielessä hankalaa, koska jokainen kortti on nimettävä ja luotava erikseen.

Batvox osoittautui kaikista vertailluista ohjelmista hitaimmaksi. Hidaskäyttöisyys ilmenee myös siinä mielessä, että näytteet on lisättävä ja tunnettujen ääninäytteiden mallit luotava yksi kerrallaan. Batvoxin käyttöliittymä ei ole aivan yksinkertainen. Sitä ei myöskään ole tarkoitettu suuren näytemäärän vertailuun, sillä se ei pysty laskemaan ja vertaamaan kovin montaa näytettä kerrallaan. Toisaalta, Batvoxin

tulokset ja niiden esittämistapa ovat huomattavasti paremmat kuin esimerkiksi VoiceNetin.

ASIS-ohjelmassakin toimitaan myös näyte kerrallaan sekä tietokantaan lisäämisen että sieltä hakemisen suhteen, mutta itse hakuprosessi on nopeampi kuin samaa perustekniikkaa käyttävässä Batvoxissa. ASISin käyttöliittymä on yksinkertainen ja selkeä. Ohjelma onkin tarkoitettu ennen kaikkea suuntaa-antavaksi ja rajaavaksi työkaluksi.

6 Johtopäätökset

Testauksen perusteella voidaan päätellä, että laboratoriotyöhön sopii parhaiten tarkka Batvox-ohjelma, jolla voidaan tutkia aina yhden jutun näytteet kerrallaan ja vertailla tuloksia taustamalliin. WinSprofiler (kuten samankaltaisuutensa vuoksi myös ASIS) sopii käsittelyn nopeutensa vuoksi laboratorioon suuntaa-antavaksi hakuohjelmaksi, jolloin tietokannasta etsitään sopivaa ehdokasta tuntemattomalle puhujalle. WinSprofilerin tulosten esittäminen ei toistaiseksi ole vielä niin yksityiskohtaista kuin Batvoxin, joka ilmoittaa tulokset sekä score-listoina että graafisesti. Tulostensa puolesta WinSprofiler ei jää yhtään jälkeen kaupallisesta ASIS-ohjelmasta. Päinvastoin, viiden parhaan joukkoon WinSprofiler löysi enemmän näytteitä kuin ASIS. FreeSpeech-ohjelmalla saatiin parhaat tulokset (ks. taulukko 1), mutta sen käytettävyydestä ei pelkästään demoversion perusteella voi sanoa mitään. Windows-komentorivillä toimivana se ei ainakaan sovellu rikosteknisen laboratorion käyttöön.

Puoliautomaattiset puhujantunnistusohjelmat ovat kehittyneet vakuuttavalle tasolle, niitä käytetäänkin rikoslaboratorioissa ympäri maailmaa ja kehitystyö jatkuu. Nämä ohjelmat eivät kuitenkaan ole erehtymättömiä eivätkä täydellisiä ja niiden suorituskykyyn vaikuttavat monet seikat. Kaikkein hyödyllisin tapa lähestyä tuntemattomien äänien ongelmaa forensisessa tutkimuksessa on käyttää hyväksi uusinta teknologiaa yhdessä perinteisen auditiivisen puhujantunnistuksen kanssa.

Viitteet

FRÄNTI, P., SAASTAMOINEN, J., KÄRKKÄINEN, I., KINNUNEN, T., HAUTAMÄKI, V. & SIDOROFF, I. 2007: Implementing speaker recognition system: From Matlab to practice. Report A-2007-4, University of Joensuu, Department of Computer Science and Statistics.

HARINEN, K., KIRJAVAINEN, J. & IIVONEN, A. 2006: Segmenttikohtaisen tiedon merkitys puhujien erottelussa. – Reijo Aulanko, Leena Wahlberg & Martti Vainio (toim.), *Fonetiikan päivät 2006 / The Phonetics Symposuim 2006*, Helsingin yliopiston puhetieteiden laitoksen julkaisuja 53. 34–43.

KINNUNEN, Tomi 1999: Automaattinen puhujantunnistusutkielma. Pro gradu –tutkielma. Tietojenkäsittelytieteen laitos, Joensuun yliopisto.

KINNUNEN, Tomi 2005: *Optimizing Spectral Feature Based Text-Independent Speaker Recognition*. Department of Computer Science, University of Joensuu.

YOUNG, S. 1996: A review of large-vocabulary continuous-speech recognition. – *IEEE Signal Processing Magazine*, (Sept. 1996):45–57.

Michael O'Dell & Tommi Nieminen (toim.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 49–56.

Äännemallien diskriminatiivinen opettaminen puheentunnistuksessa

Matti Varjokallio, Janne Pylkkönen, Mikko Kurimo
Teknillinen korkeakoulu

Tiivistelmä

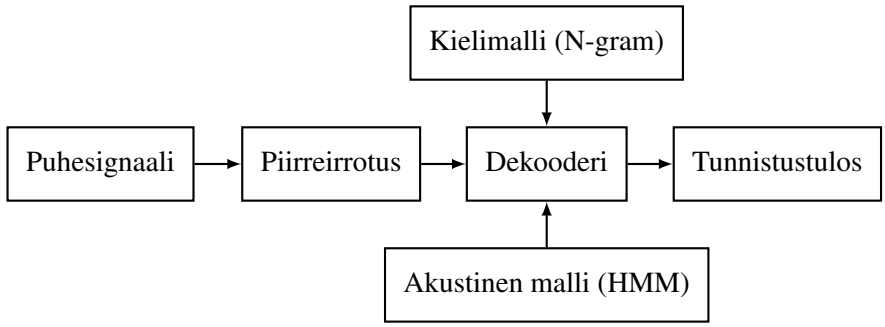
Jatkuvassa laajan sanaston puheentunnistuksessa käytetään yleensä kontekstiriippuvia trifoneita äänteiden mallinnukseen. Mallin parametrien hyvä opettaminen on tärkeää, jotta saavutetaan korkea tunnistustarkkuus. Perinteinen tapa mallin parametrien opettamiseen on niin sanottu suurimman uskottavuuden menetelmä, jossa mallin parametrit asetetaan maksimoimaan opetusnäytteiden uskottavuus. Tässä tapauksessa kaikkia malleja käsitellään erillisinä yksiköinä, joten äänteiden erottuvuutta ei voida ottaa huomioon. Diskriminatiivisessa opetuksessa puolestaan äännemallien parametrit opetetaan kasvattamaan oikean luokituksen todennäköisyyttä. Tässä työssä verrataan suurimman uskottavuuden menetelmällä opetettua äännemallia kahteen diskriminatiivisilla menetelmillä, suurimman keskinäisinformaation menetelmällä sekä pienimmän äännevirheen menetelmällä, opetettuihin malleihin. Eri opetusmenetelmillä saavutettuja tunnistustarkkuuksia verrataan suomenkielisessä sanelutehtävässä.

Avainsanat: puheentunnistus, suurin uskottavuus, diskriminatiivinen opetus

1 Johdanto

Laajan sanaston automaattinen puheentunnistus on haastava ongelma. Jotta saavutetaan hyvä tunnistustulos, täytyy tunnistimella olla akustista informaatiota, kielellistä informaatiota ja tehokas hakualgoritmi erilaisten tunnistusvaihtoehtojen pisteyttämiseen. Tyypillisen laajan sanaston puheentunnistimen rakenne on esitetty kuvassa 1. Tyypillisesti puhesignaalista irrotetaan Mel-kepstriin perustuvia piirteitä, akustisessa mallinnuksessa käytetään kätkettyä Markov-mallia ja kielimallinnuksessa N-gram -mallia. Dekooderi yrittää sovittaa akustisiin piirteisiin hypoteeseja käyttäen pisteytykseen sekä akustista mallia että kielimallia.

Suomen kielen erityispiirteenä on sen agglutinatiivisuus, mikä aiheuttaa ongelmia tunnistussanaston valintaan: sanayksiköiden suuren määrän takia sanoja ei sellaisenaan voida käyttää tunnistuksessa. Viime vuosina onkin saatu hyviä tuloksia käytettäessä tunnistusyksiköinä ns. tilastollisia morfeja (Hirsimäki *et al.* 2006, Creutz



Kuva 1: Laajan sanaston puheentunnistimen rakenne

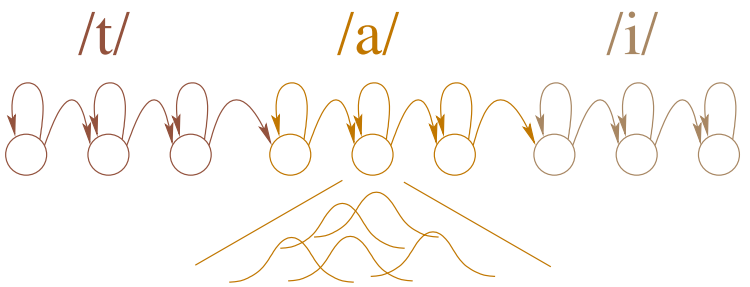
2006). Tässä työssä vertaillaan eri opetuskriteereitä akustisen mallin parametrien opettamiseen ja testit suoritetaan suomenkielisessä sanelutehtävässä. Tunnistustehtävä on siis laajan sanaston jatkuvan puheen tunnistusta (Large Vocabulary Continuous Speech Recognition, LVCSR), jossa käytetään morfeihin perustuvaa laajan sanaston kielimallia.

Artikkeli rakentuu seuraavasti. Luvussa 2 käsitellään äänteiden mallinnusta yleisesti. Luvussa 3 keskitytään äännemallien diskriminatiiviseen opettamiseen käyttäen suurimman keskinäisinformaation menetelmää (Maximum Mutual Information, MMI; Normandin *et al.* 1994) sekä pienimmän äännevirheen menetelmää (Minimum Phone Frame Error, MPFE; Zheng & Stolcke 2005). Luvussa 4 raportoimme tulokset suomenkielisessä sanelutehtävässä ja lopuksi luvussa 5 on pohdiskeleva yhteenveto.

2 Äännemallinnus

2.1 Trifonit

Tunnistimelle täytyy valita yksiköt, joita halutaan mallintaa. Kontekstiton eli ns. monofonimalli on puheentunnistuskäyttöön riittämätön, koska se ei huomioi koartikulaatiota lainkaan. Tavuihin perustuvan akustisen mallinnuksen ongelmaksi muodostuu puolestaan riittämätön opetusaineisto (Siivola *et al.* 2002). Automaattista puheentunnistusta varten hyvä tasapaino näiden kahden tapauksen väliltä saavutetaan käyttämällä akustisina yksiköinä kontekstiriippuvia trifoneita. Trifoni tuntee välittömän kontekstinsa eli edeltävän ja seuraavan foneemin. Mallinnuksessa pystytään näin ottamaan huomioon myös jonkin verran koartikulaatiota ja yksiköiden määrä pysyy vielä mallinnettavissa.



Kuva 2: Esimerkki kätketystä Markov-mallista

2.2 Kätketty Markov-malli

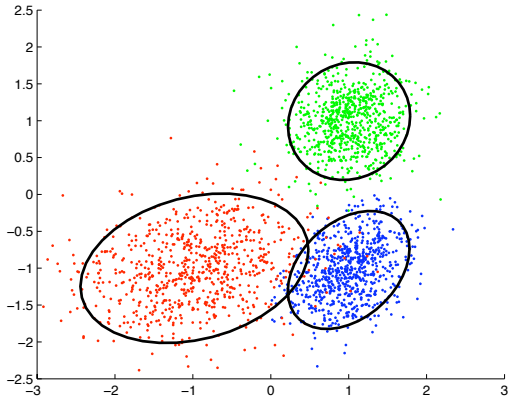
Akustinen mallinnus perustuu tyypillisesti kätkettyihin Markov-malleihin (Hidden Markov Model, HMM; Rabiner 1989). Suurin syy tähän lienee se, että HMM-viitekehys on teoreettisesti hyvin tunnettu ja HMM-malleille on olemassa tehokkaat opetus- ja tunnistusalgoritmit. Markov-mallissa on tiloja, joiden välillä siirrytään tietyillä siirtymätodennäköisyyksillä ja lisäksi järjestelmän tila on aina tiedossa. Kätketyssä Markov-mallissa senhetkinen tila ei puolestaan ole tiedossa vaan siitä saadaan vain viitteitä havaintojen perusteella. Äännemallinnuksessa jokaiselle trifonille opetetaan tyypillisesti kolmitilainen kätketty Markov-ketju siirtymätodennäköisyyksineen ja emissiotodennäköisyysjakaumineen. Emissiotodennäköisyysjakauman avulla voidaan Mel-kepstri-piirteistä laskea, kuinka todennäköistä on, että ollaan tietyssä tilassa kyseisellä hetkellä.

2.3 Gaussin mikstuurimalli

Kätketyn Markov-mallin yhteydessä käytetään emissiotodennäköisyysjakaumana tyypillisesti Gaussin mikstuurimallia (Gaussian Mixture Model, GMM; Rabiner 1989). Gaussin mikstuuri on yhdistelmä monesta moniulotteisesta Gaussin jakaumasta. Moniulotteisten Gaussisten jakaumien tasatodennäköisyyspinnat ovat hyperellipsejä. Tästä siis saadaan geometrinen intuitio ellipsien sovittamisesta sopivasti opetusnäytteisiin. Kuvassa 2.3 on esimerkki kaksiulotteisesta kolmikomponenttisesta Gaussin mikstuurista. Oikeassa tapauksessa komponenttien määrä olisi luokkaa 10–150 ja ulotteisuus tyypillisesti n. 40.

3 Äännemallien opettaminen

HMM-mallin rakenne selostettiin edellisessä luvussa. Tässä luvussa keskitymme mallin parametrien opettamiseen, kun käytössä on akustinen tietokanta transkriptioineen.



Kuva 3: Esimerkki Gaussin mikstuurimallista

3.1 Suurimman uskottavuuden menetelmä

Perinteinen tapa HMM:n parametrien opettamiseen on ns. suurimman uskottavuuden menetelmä (Maximum Likelihood, ML). Opetuksessa pyritään maksimoimaan seuraavaa lauseketta:

$$\mathscr{F}_{ML}(\lambda) = \sum_{r=1}^{R} \log p_\lambda \left(\mathscr{O}_r | \mathscr{M}_{w_r} \right), \tag{1}$$

jossa $\mathscr{O}$ ovat opetushavainnot, $\mathscr{M}$ malli, λ mallin parametrit ja r opetuslauseen indeksi. Alaindeksi w_r on lausekkeessa mukana korostamassa sitä, että käsittelyn kohteena on vain oikea transkriptio. Tavoitteena on siis asettaa mallin parametrit maksimoimaan opetusnäytteiden uskottavuutta keskiarvoistettuna koko tietokannan yli. Tähän tarkoitukseen on olemassa tehokas Baum-Welch -algoritmi Rabiner (1989). Kriteeri on toimivaksi todettu ja sillä saadaan opetettua hyvä perusmalli. Sen heikkoutena puolestaan on, että eri äännemallien parametrit opetetaan toisistaan riippumatta. Puheentunnistus on kuitenkin luokittelutehtävä, joten parempiakin menetelmiä on olemassa.

3.2 Suurimman keskinäisinformaation menetelmä

Suurimman keskinäisinformaation menetelmässä (Maximum Mutual Information, MMI, Woodland & Povey 2002) pyritään maksimoimaan seuraavaa lauseketta:

$$\mathscr{F}_{MMIE}(\lambda) = \sum_{r=1}^{R} \log \frac{p_\lambda \left(\mathscr{O}_r | \mathscr{M}_{w_r} \right) P(w_r)}{\sum_{\hat{w}} p_\lambda (\mathscr{O}_r | \mathscr{M}_{\hat{w}}) P(\hat{w})}, \tag{2}$$

jossa $\mathcal{O}$ ovat opetushavainnot, $\mathcal{M}$ malli, λ mallin parametrit, w hypoteesi, w_r oikea hypoteesi ja r opetuslauseen indeksi kuten suurimman uskottavuuden menetelmässäkin. Lisäksi huomataan, että mukaan ovat tulleet kielimallitodennäköisyydet $P(w)$ sekä nimittäjään summa kaikkien mahdollisten (myös virheellisten) hypoteesien $\hat{w}$ yli. Tavoitteena on siis opettaa parametrit niin, että oikeiden hypoteesien uskottavuuden suhde virheellisten hypoteesien uskottavuuteen maksimoidaan keskiarvoistaen jälleen koko opetusaineiston yli. Kaikkia virheellisiä vaihtoehtoja $\hat{w}$ ei luonnollisesti voida ottaa huomioon, vaan nimittäjää täytyy jotenkin approksimoida. Käytännössä tämä tapahtuu tunnistamalla opetusaineisto olemassa olevalla mallilla ja käyttäen sopivaa määrää tunnistimen parhaiksi arvioimia hypoteeseja. Opettaminen tapahtuu käyttäen ns. Extended Baum-Welch -algoritmia (EBW; Normandin *et al.* 1994) tunnistimelta saatavan tunnistusverkon yli.

3.3 Pienimmän äännevirheen menetelmä

Suurimman keskinäisinformaation menetelmä otti jo huomioon äänteiden välisiä riippuvuuksia, mutta tämä tapahtui eri hypoteesien todennäköisyyksien avulla. Olisi kuitenkin hyvä ottaa jollain tavalla huomioon suoraan eri hypoteeseissa tapahtuvat virheet. Tällaisten kriteerien voidaan ajatella minimoivan Bayes-luokitteluriskiä eli olevan MBR (Minimum Bayes Risk) -estimaatteja. Kaksi tärkeintä puheentunnistukseen ehdotettua MBR-kriteeriä ovat MCE (Minimum Classification Error; McDermott *et al.* 2007) sekä MPE (Minimum Phone Error; Povey & Woodland 2002). Oma järjestelmämme käyttää erästä jälkimmäisen kriteerin varianttia nimeltään MPFE (Minimum Phone Frame Error; Zheng & Stolcke 2005). Se pyrkii minimoimaan seuraavaa lauseketta:

$$\mathcal{F}_{MPFE}(\lambda) = \sum_{r=1}^{R} \sum_{s} \frac{p_\lambda(\mathcal{O}|s)^\kappa P(s)^\nu}{\sum_u p_\lambda(\mathcal{O}_r|u)^\kappa P(u)^\nu} Accuracy(s, s_r), \tag{3}$$

jossa $\mathcal{O}$ ovat opetushavainnot, r opetuslauseen indeksi, s hypoteesin indeksi, κ ja ν vakioita ja $P(\cdot)$ kielimallitodennäköisyys. Accuracy-termi on hypoteesin s virhefunktio verrattuna oikeaan reittiin s_r. Koska Accuracy-termi oletetaan vakioksi aina yhden iteraation ajaksi, voidaan lausekkeen minimoimisen ajatella muuttavan mallia niin, että keskimääräinen kielimallilla painotettu virhe pienenee koko opetusaineiston yli. Opetukseen voidaan käyttää Extended Baum-Welch -algoritmia kuten MMI-kriteerin tapauksessakin. Hypoteesien määrää rajoitetaan myös samalla tavalla kuin MMI-kriteerin yhteydessä.

4 Tulokset

Testit suoritettiin laitoksen suomenkielisellä laajan sanaston puheentunnistimella. Kielimalli opetettiin CSC:n 150 miljoonan sanan tekstiaineistosta, joka sisältää mm.

STT:n uutisia, sanomalehtiaineistoa ja romaaneja. Akustisen mallin opetukseen käytettiin SPEECON-aineistoa (Iskra *et al.* 2002), joka sisältää yksittäisiä sanoja ja lauseita erilaisissa kohinaolosuhteissa. Näistä valitsimme foneettisesti rikkaat lauseet vähäkohinaisissa olosuhteissa. Yhteensä opetusaineistoa oli 20 tuntia ilman hiljaisuuksia reilulta kahdeltasadalta puhujalta. Testauksessa käytettiin samaa aineistoa ja sitä oli 1,2 tuntia ilman hiljaisuuksia 40 puhujalta, jotka eivät olleet opetusaineistossa.

Opetus tapahtui seuraavasti. Ensiksi opetettiin perusmalli suurimman uskottavuuden menetelmällä. Tällä mallilla koko opetusaineisto tunnistettiin ja saatujen tunnistusverkkojen perusteella opetettiin MMI- sekä MPFE -kriteereillä neljä iteraatiota kutakin käyttäen ML-perusmallia lähtökohtana. Neljä iteraatiota on havaittu riittäväksi määräksi molemmille opetusmenetelmille. Näin saatujen kolmen mallin tunnistustulokset ovat taulukossa 4. Virheprosentilla tarkoitetaan tässä yhteydessä lisäys-, poisto- ja korvausvirheiden suhdetta todelliseen kirjain- tai sanamäärään.

Perusmalli oli siis jo itsessään varsin hyvälaatuinen ja sillä saavutettiin kirjainvirhe 2,83 prosenttia. Molemmilla diskriminatiivisilla menetelmillä tulos parani edelleen jonkin verran. MMI-kriteerillä saatiin kirjainvirheeksi 2,75 prosenttia, joka tarkoittaa noin kolmen prosentin suhteellista parannusta lähtötuloksesta. MPFE-kriteerillä puolestaan kirjainvirheeksi tuli 2,59 prosenttia, joka on reilun kahdeksan prosentin suhteellinen parannus lähtötuloksesta.

Taulukko 1: Tulokset suomenkielisessä sanelutehtävässä

Opetusmenetelmä	Kirjainvirhe (%)	Sanavirhe (%)
ML	2,83	12,4
MMI	2,75	12,1
MPFE	2,59	11,7

Absoluuttiset parannukset eivät siis ole kovin suuria, johtuen jo valmiiksi hyvästä lähtötasosta. Suhteellisesti parannukset ovat kuitenkin kohtuullisen suuria yhdellä menetelmällä. Tilastollinen merkitsevyys testattiin käyttäen Wilcoxon signed rank -testiä. Parannus ML-tuloksesta MPFE-tulokseen on tilastollisesti merkitsevä, kun taas ML–MMI- ja MMI–MPFE-parit eivät ole tilastollisesti merkitseviä.

5 Yhteenveto

Tässä työssä tutkittiin äännemallien diskriminatiivista opettamista laajan sanaston puheentunnistukseen. Kahta kriteeriä sovellettiin suomenkielisessä sanelutehtävässä. Menetelmillä saavutettiin kohtuullinen parannus vertailutulokseen nähden. Pie-

nimmän äännevirheen menetelmä havaittiin paremmaksi kuin suurimman keskinäisinformaation menetelmä.

Menetelmän hyödyllisyyttä luultavasti rajoitti vertailumallin jo valmiiksi hyvä taso sekä opetusdatan määrä. Diskriminatiivisesta opetuksesta saatava hyöty riippuu varsin voimakkaasti opetusdatan määrästä per opetettava parametri ja suomenkielinen opetusaineistomme ei ollut kovinkaan iso. Alustavat tulokset hieman isommalla suomenkielisellä Speechdat-puhelinaineistolla näyttävät vahvistavan tätä olettamusta ja alustavat suhteelliset parannukset ovat selkeästi suurempia kuin Speecon-aineistolla.

Akustisten mallien diskriminatiivinen opettamisen toteuttaminen on varsin haastavaa, koska opetuksessa statistiikkoja täytyy kerätä monesta hypoteesista ja tyypilliset opetustietokannat ovat varsin isoja. Nopeusvaatimukset täytyy siis ottaa tarkkaan huomioon opetusvaiheessa. Tunnistusvaiheessa puolestaan diskriminatiivisesti opetetut mallit toimivat aivan samalla lailla kuin ML-kriteerilläkin opetetut mallit eivätkä aseta mitään lisävaatimuksia järjestelmälle, joka on varsin hyvä ominaisuus.

Viitteet

CREUTZ, M. 2006: *Induction of the Morphology of Natural Language: Unsupervised Morpheme Segmentation with Application to Automatic Speech Recognition*. väitöskirja, Helsinki University of Technology.

HIRSIMÄKI, T., CREUTZ, M., SIIVOLA, V., KURIMO, M., VIRPIOJA, S. & PYLKKÖNEN, J. 2006: Unlimited vocabulary speech recognition with morph language models applied to Finnish. – *Computer Speech and Language*, **20**: 515–541.

ISKRA, D., GROSSKOPF, B., MARASEK, K., VAN DEN HEUVEL, H., DIEHL, F. & KIESSLING, A. 2002: SPEECON—speech databases for consumer devices: Database specification and validation. Tekninen raportti.

MCDERMOTT, E., HAZEN, T. J., LE ROUX, J., NAKAMURA, A. & KATAGIRI, S. 2007: Discriminative training for large vocabulary speech recognition using minimum classification error. – *IEEE Transactions on Audio, Speech and Language Processing*, **15**: 203–223.

NORMANDIN, Yves, LACOUTURE, Roxane & CARDIN, Regis 1994: MMIE training for large vocabulary continuous speech recognition. – *ICSLP-1994*. 1367–1370.

POVEY, D. & WOODLAND, P. C. 2002: Minimum phone error and I-smoothing for improved discriminative training. – *Proceedings of the ICASSP*, osa 1. 105–108.

RABINER, L. R. 1989: A tutorial on hidden Markov models and selected applications in speech recognition. – *Proceedings of the IEEE*, **77**(2): 257–286.

SIIVOLA, V., HIRSIMÄKI, T. & KURIMO, M. 2002: Äännemallien vertailua jatkuvassa suuren sanaston puheentunnistuksessa. – P. Korhonen (toim.), *Fonetiikan Päivät 2002 / The Phonetics Symposium 2002*. Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing. 75–82.

WOODLAND, P. C. & POVEY, D. 2002: Large scale discriminative training of hidden Markov models for speech recognition. – *Computer Speech and Language*, **16**: 25–47.

ZHENG, Jing & STOLCKE, Andreas 2005: Improved discriminative training using phone lattices. – *INTERSPEECH-2005*. 2125–2128.

Michael O'Dell & Tommi Nieminen (toim.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 57–64.

Puheen äänittäminen magneettiresonanssikuvauksen aikana

Teemu Lukkari, Jarmo Malinen & Pertti Palo
Teknillinen Korkeakoulu

Tiivistelmä

Kuvaamme mittausjärjestelyn puheen äänittämiseksi koehenkilön ääntöväylän anatomisen MRI:n aikana. Järjestelyllä hankittua kuva- ja äänidataa käytämme jatkosssa ääntöväylän numeerisen mallin rakentamiseen ja validoimiseen.

Avainsanat: Puheen äänittäminen, MRI

1 Johdanto

Tässä artikkelissa raportoimme suomen kielisten vokaalien FEM-pohjaisen numeerisen simulaattorin kehitystyön edistymistä. Anatomian geometrisen mallin pohjaksi ja akustisen mallin validoimiseen tarvitsemme formantteja puhesignaalista, joka on äänitetty yhtäaikaa MRI kuvauksen aikana.

Magneettiresonanssikuvausta (MRI) on käytetty ääntöväylän kuvantamiseen jo pitkään (Baer *et al.* 1987). Nykyisin koko ääntöväylä voidaan kuvata reilusti alle 30 sekunnissa Engwall (2004). MRI:llä tuotettu anatominen data sopii laskennallisen verkon rakentamiseen elementtimenetelmää varten (FEM). FEM-ratkaisijoita aaltoyhtälölle on käytetty simuloitaessa normaalin puheentuoton akustiikkaa (Hannukainen *et al.* 2007, Lu *et al.* 1993, Švancara *et al.* 2004) sekä anatomisten poikkeamien ja suu- ja leukakirurgian vaikutuksia (Dedouch *et al.* 2002, Nishimoto *et al.* 2004, Švancara & Horáček 2006).

Suoritamme mittaukset Siemens Magnetom Avanto 1.5 T -laitteella. MRI-huoneen ympäristö on äänittämisen kannalta haastava. MRI-laitteen käämin sisällä on 1.5 T staattinen magneettikenttä ja huoneessa vallitseva hajakenttäkin voi olla merkittävä. Kuvaussekvenssi tuottaa ≈ 64 MHz sähkömagneettisen kentän, koska protonien Larmor-taajuus on 42.58 MHz/T. Huipputeho voi nousta useisiin kilowatteihin. Ylimääräisenä lisävaikeutena laite tuottaa sekvenssin aikana akustista melua, jonka voimakkuus on noin 90 dB (SPL) laajalla taajuuskaistalla, joka osuu epäedullisesti juuri formanttien taajuuksille.

Melu estää koehenkilö kuulemasta omaa ääntään kuvauksen aikana normaalilla tavalla. Tästä syystä olisi hyödyllistä syöttää melusta puhdistettu viivästymätön äänisignaali takaisin koehenkilön kuulokkeisiin puheen luonnollisuuden parantamiseksi. Koska kokeissa on mukana ihminen koehenkilönä, turvallisuus- ja mukavuustekijat pitää ottaa huolellisesti huomioon.

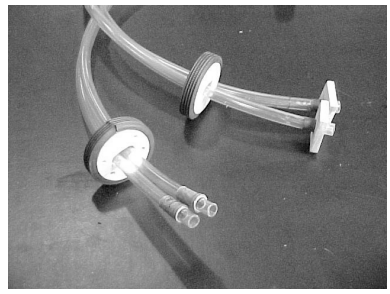
Karkeasti ottaen, tehtävämme on erottaa tasoaalto (eli puhesignaali) sylinterisymmetrisestä melulähteestä (eli ympäristöstä), kun samaan aikaan otetaan huomioon yllä kuvatut haasteet.

2 Vaatimusmäärittely

Magneettikentän taki ferromagneettisia materiaaleja voidaan käyttää MRI-huoneeseen menevissä laitteiston osissa vain mitättömän pieniä määriä eikä niitä sallita ollenkaan MRI-käämin sisään sijoitettavassa äänenkerääjässä. Kaikki elektroniikka kuvaushuoneessa pitää suojata ylijännitettä ja radiotaajuuksia vastaan. Luonnollisesti johtavien materiaalien suljettuja silmukoita on huolellisesti vältettävä.

2.1 Äänenkerääjä ja akustiset aaltojohdot

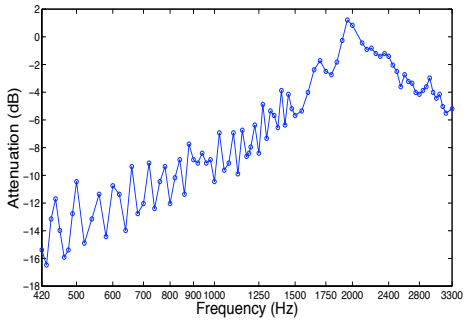
Käytämme kaksikanavaista äänenkerääjä, jossa toinen kanava on puhe- ja toinen melunäytettä varten. Äänenkerääjän mittojen tulee olla pienet verrattuna formanttien aallonpituuksiin ja kerääjän pitää mahtua MRI-laitteiston sisään.



Kuva 1: Akustiset aaltojohdot ja niiden ripustusjärjestely

Äänisignaalit kuljetetaan mikrofonilevylle akustisilla aaltojohdoilla (Katso kuvaa 1. Ne on tehty pehmeästä PVC-putkesta, jonka sisä halkaisija on 9 mm. Kunkin aaltojohdon pituus on 3.0 m ja ne on ripustettu parittain siten, että ulkopuoliset häiriöt saadaan kumottua.

Väliaine kerääjässä ja aaltojohdoissa on ilma. Äänen johtuminen aaltojohtojen seinissä vaikuttaa olevan hyvin pientä verrattuna ilmassa johtuvaan ääneen. Kuvassa 2 on aaltojohdon taajuusvaste taajuuskaistalle 0.42–3.3 kHz. Aaltojohdon pitkittäiset resonannssit näykyvät kuvassa 2 matalemmilla taajuuksilla. Alle 1.5 kHz nä-



Kuva 2: Aaltojohtojen taajuusvaste

kyy myös oktaavia kohden $\approx$ 4dB vaimennus, joka voidaan helposti kompensoida esimerkiksi RC-suodattimella.

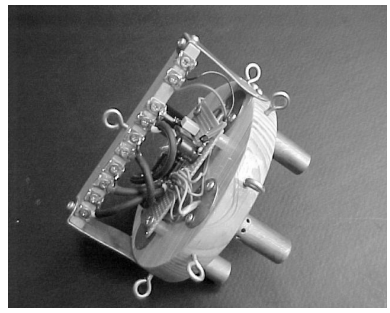
2.2 Suojattu mikrofonilevy ja kaapelointi

Mikrofonilevy on sijoitettu Faradayn häkkiin. Häkki on 6 mm alumiinilevyä, joka on riittävän paksua, jottei se taivu tai resonoi. Vuoraamme häkin sisäpuolelta äänieristysmateriaalilla. Akustiset aaltojohdot viemme sisään häkkiin sähkömagneettisten aaltojohtojen, jotka on suunniteltu olemaan läpinäkymättömiä taajuuskaistalla 10–100 MHz, läpi.

Mikrofonilevy (katso kuvaa 3) koostuu neljästä Panasonicin WM-62 kondensaattorimikrofonista (herkkyys -45 ± 4 dB re 1V/Pa @ 1 kHz, $\varnothing$ 9 mm) sekä niiden jännitelähteestä. Valmistajan datalehden antama mikrofonien nimellinen taajuusvaste on käytännössä suora sovellukseemme tarvitulla taajuuskaistalla. Pintapuolisen mittauksen mukaan tällaisten mikrofonien herkkyydet ja taajuusvasteet eivät näytä eroavan merkittävästi toisistaan. Näinollen jätimme tarkemmat kalibrointimittaukset tekemättä.

Mikrofonit on upotettu levyyn joka on eristetty akustisesti ja sähköisesti Faradayn häkin seinistä. Äänisignaalit viedään mikrofoneille yksinkertaisten, säädettävien akustisten impedanssisovittimien kautta (katso oikeaa alalaitaa kuvassa 3). Nämä sovittimet viritetään kokeellisesti sulkemalla joitakin putkien seinässä olevista rei'istä ($\varnothing$ 2 mm). Virittäminen on tarpeen, jotta mikrofonilevyltä aaltoputkiin heijastuva aalto saadaan minimoitua. Tilanne vastaa sähköisten siirtolinjojen terminointia. Tuloksena aaltojohtojen pitkittäiset resonanssit vaimenevat osin (katso kuvaa 2).

Riippuen impedanssisovittiemien suljettujen reikien määrästä useiden desibelien energian vaimentuminen näkyy järjestelmän taajuusvasteessa. Koska sovittimet koostuvat sekä suljetusta että avoimesta aaltojohdon päätöksestä, vastaavasti jäännösheі-



Kuva 3: Mikrofonilevy

jastus koostuu sekä vaiheensäilyttävästä että -kääntävästä osasta. Huomaamme, että tämä vastaa täsmälleen mitattujen piikkien määrää kuvassa 2.

Signaalit siirretään MRI-huoneesta kahdta microfonikaapelia pitkin (Tasker C116 4x0.14-26AWG); kaksi kanavaa kummassakin. Kaikki kaapelien päätteet on suojattu ylijännitteeltä diodeilla. Koska vain kaksi kanavista on äänenkerääjän käytössä, loput ovat varalla.

2.3 Melunperuutusvahvistin ja CMRR-käyrät

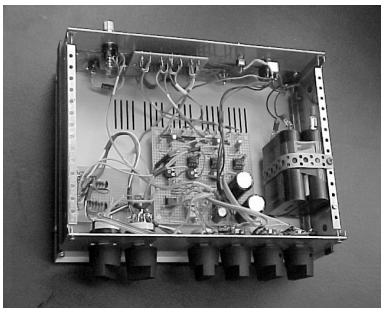
Koehenkilön pitää kuulla puhdistettu signaali reaaliaikaisena. Tämän mahdollistamiseksi olemme toteuttaneet melunperuutusjärjestelmän analogisella elektroniikalla. Laite on summausvahvistin (katso kuvaa 4), jossa on yksi suora kanava (signaalille) ja kolme säädettävää käänteisessä vaiheessa summattavaa kanavaa (mahdollistaen korkeintaan kolmen melusignaalin vähentämisen). Ennen äänitystä summauspainot säädetään manuaalisesti kuuntelemalla vahvistimen ulostuluo. Laitteen tärkeimmät komponenentit ovat kuusi LM741:tä, ja sen tuloimpedanssi on 3 kΩ.

Vahvistimen taajuusvaste on suora kaistalla 0.2–5 kHz. Sen optimaalinen yhteismuodon vaimennussuhde (CMRR) välillä 0.42–3.3 kHz on kuvattu alimmalla suhteellisen sileällä käyrällä kuvassa 5. CMRR:ää voidaan parantaa pienentämällä vahvistimen elektrolyyttikondensaattorien toleransseja.

Ylin, melko karkean näköinen käyrä kuvassa 5 on koko järjestelmän mitattu CMRR. Tässä vaiheessa järjestelmä koostuu aaltojohdoista sekä akustisista impedanssisovittimista aaltojohtojen molemmissa päissä. Kuvan 5 kahden käyrän välinen ero syntyy enimmäkseen aaltojohtojen fyysisistä ominaisuuksista ja — valitettavasti — mittauksissa käytetyn äänilähteen heikosta laaudusta.

2.4 Tietokonelaitteisto ja signaalikäsittely

Vahvistimelta tuleva puhdistettu signaali digitoidaan MacBook Pro (2,2 GHz) -tietokoneella, jossa on käyttöjärjestelmänä Mac OS X 10.4.9. Tarvittu signaalinkäsittely



Kuva 4: Melunperuutusvahvistin

ja formanttien poimiminen tehdään Matlabin versiolla 7.4, Signal Processing Toolboxilla sekä erikseen tähän tarkoitukseen kirjoittamillamme ohjelmilla. Erityisesti kuvassa 2 näkyvät pitkittäiset resonanssit voidaan kompensoida Matlabilla. On syytä huomata, että taajuusvaste on mitattava uudelleen lopullisessa koejärjestelyssä, koska aaltoputkien taivuttaminen muuttaa niiden resonansseja (Sondhi 1986).

3 Mittaukset

Seuraavassa kerromme yksityiskohtaisemmin, miten suoritimme mittaukset joihin kuvat 2 ja 5 perustuvat.

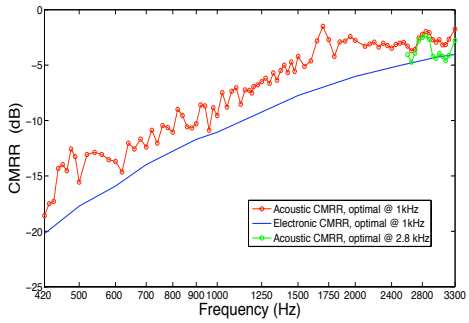
3.1 Järjestely ja laitteisto

Sinigeneraattori (Taylor 192A) liitettiin kaksikanavaiseen äänilähteeseen (katso kuvaa 6), ja äänenpaine lähteen päässä pidettiin manuaalisesti 94 dB:n (SPL) vakiotasolla taajuuskaistalla 0.42–3.3 kHz. Tämä oli mahdollista mittaamalla äänilähteen sisäisten referenssimikrofonien antama signaali analogisella jännitemittarilla (Heathkit V-7 A) mikrofoniesivahvistimen (Resound CVS908) kautta. Oskilloskooppia käytimme mahdollisten vääristymien tarkkailuun visuaalisesti.

Tuotetut äänisignaalit syötettiin mikrofonilevylle (katso kuvaa 3) aaltojohtojen läpi (katso kuvaa 1). Aaltojohdot suoristettiin täysin mittausten ajaksi, ja ympäristön akustisten häiriöiden kontrollointiin käytettiin erinäisiä keinoja.

Mikrofonilevyltä, molemmat signaalit tuotiin melunperuutusvahvistimen suoraan ja käänteisen vaiheen kanaviin. Suoran kanavan vahvistus asetettiin 45 dB:n tasolle. Käänteisen kanavan vahvistus asetettiin siten, että vahvistimen lähtötaso oli minimissään kun 1 kHz:n sinisignaali syötettiin molempiin kanaviin.

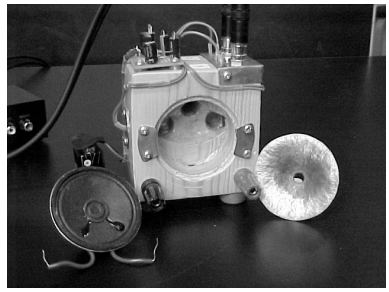
Kaikki data kuvia 2 ja 5 varten mitattiin toisella analogisella jännitemittarilla (Goerz Unigor 226221) vahvistimen lähdöstä. Kaikilla mitatuilla taajuuksilla tulokset mitattiin sekä käänteisen kanavan ollessa kytkettynä että sen ollessa irti.



Kuva 5: Optimaaliset CMRR käyrät vahvistimelle (alempi) ja akustisille aaltojohdoille (ylin)

3.2 Äänilähde

Jos ajatellaan edellä kuvattuja mittauksia, ideaalisen äänilähteen tulisi pystyä tuottamaan kaksi siniäänisignaalia, joilla on sama amplitudi ja vaihe. Molempien kanavien tulisi olla akustisesti riippumattomia ja niiden akustisten impedanssien tulisi olla identtiset. Kaikki tämä pitäisi saada aikaan ilman vääristymiä laajalla taajus- ja äänenpaineen amplitudialueella.



Kuva 6: Purettu äänilähde

Meidän äänilähteemme (katso kuvaa 6) koostuu kaiuttimesta (∅ 50 mm, impedanssi 8 Ω) ja symmetrisestä kammiosta, joka jakaa painekentän kahteen kanavaan. Kummankin kanavan seinään on upotettu referenssimikrofoni tyypiltään Panasonic WM-62.

Äänilähteessä on myös saman tyyppinen yksinkertainen akustinen impedanssisovitin kuin mikrofonilevyssä. Sen tarkoituksena on matkia kvalitatiivisesti todellisen, MRI-laitteiston kanssa käytettävän, äänenkerääjän impdanssia. On syytä huomata, että äänenkerääjän ja äänilähteen akustiset impedanssit ovat erilaiset, ja että tällä on kvantitatiivinen vaikutus kuvan 2 kaltaiseen taajusvastekäyrään.

Äänilähde kärsii sekä kammion että itse kaiuttimen resonansseista, joiden lähellä tuotetut äänisignaalit ovat eri vaiheessa. Tuloksena saatujen CMRR -mittausten tulokset ovat huonompia kuin todellinen CMRR olisi. Vähentääksemme erityisen ikävää ≈ 1.7 kHz resonanssia, valmistimme torven kuparipellistä ja asetimme sen kammion ja kaiuttimen väliin. Torvi näkyy kuvassa 6 oikealla. Emme saaneet CMRR -dataa korkeille taajuuksille, koska kammio alkaa resonoida ≈ 3.5 kHz:n taajuudella. Toisaalta, alle 0.4 kHz taajuudet pitää mitata ilman torvea, koska se vääristäisi signaalia matalilla taajuuksilla.

1.7 kHz:n, 2.85 kHz:n, ja 3.3 kHz:n huiput ylimmässä CMRR -käyrässä kuvassa 5 selittyvät ainakin osittain äänilähteen kanavien välisellä vaihe-erolla, jotka varmistettiin oskilloskoopilla tehdyllä Lissajous -mittauksella. Sen sijaan 1.95 kHz huippu ei johdu vaihe-erosta.

2 kHz:n yläpuolella äänilähteen kanavat alkavat ajautua epätasapainoon, koska kaiutin ei ole symmetrinen. Kun tämä epätasapaino kompensoitiin säätämällä melunperuutusvahvistinta, saimme paljon paremman CMRR -käyrän taajuskaistalle 2.6–3.3 kHz. Tämäkin käyrä on piirretty kuvaan 5.

Yhteenvetona toteamme, että aaltojohtojen todellinen CMRR on merkittävästi parempi korkeilla taajuuksilla kuin mitä kuvasta 5 voisi päätellä. Hyvä laatuisen monikanavaisen äänilähteen suunnittelu ja rakentaminen on edelleen haastava akustisen insinöörityön tehtävä.

4 Johtopäätökset

Olemme kuvanneet melunperuutus-, äänensiirto- ja äänitystekniikat akustisten aaltojohtojen välityksellä vaikeissa oloissa kuten MRI-huoneessa.

Akustiset aaltojohdot muuttavat ääneen laatua. Ne lisäävät puheeseen naksahduksia ja muuttavat sen hieman käheäksi. Tästä huolimatta puhe on edelleen hyvin ymmärrettävää ilman minkäänlaista, esimerkiksi numeerista, aaltojohtojen resonanssien korjausta. Johtopäätöksenä odotamme saavamme äänitettyä erityyppisistä puhesignaaleista hyvälaatuisia näytteitä, joista esimerkiksi formanttien poimimisen pitäisi olla mahdollista.

Kiitokset

Haluamme kiittää TkT A. Laaksoa ja TkT K. Rytsölää (Fysiikan laboratorio, TKK) arvokkaista keskusteluista ja laboratorio laitteiden sekä -tilan lainaamisesta mittauksia varten.

Suomen Akatemia on rahoittanut Teemu Lukkarin työtä. Tämä artikkeli on aiemmin julkaistu englanninkielisenä MAVEBA 2007 -konferenssissa (Lukkari *et al.* 2007).

Viitteet

BAER, T., GORE, J. C., BOYCE, S. & NYE, P. W. 1987: Application of MRI to the analysis of speech production. – *Magnetic Resonance Imaging*, **5**: 1–7.

DEDOUCH, K., HORÁČEK, J., VAMPOLA, T. & ČERNÝ, L. 2002: Finite element modelling of a male vocal tract with consideration of cleft palate. – *Forum Acusticum*, Sevilla, Spain.

ENGWALL, Olov 2004: Speaker adaptation of a three-dimensional tongue model. – Soon Hyob Kim & Dae Hee Youn (toim.), *ICSLP 2004*, osa I, Jeju Island, Korea. 465–468.

HANNUKAINEN, A., LUKKARI, T., MALINEN, J. & PALO, P. 2007: Vowel formants from the wave equation. – *Journal of the Acoustical Society of America Express Letters*, **122**(1): EL1–EL7.

LU, C., NAKAI, T. & SUZUKI, H. 1993: Finite element simulation of sound transmission in vocal tract. – *J. Acoust. Soc. Jpn. (E)*, **92**: 2577–2585.

LUKKARI, T., MALINEN, J. & PALO, P. 2007: Recording speech during magnetic resonance imaging. – *MAVEBA 2007*, Florence, Italy. 163–166.

NISHIMOTO, H., AKAGI, M., KITAMURA, T. & SUZUKI, N. 2004: Estimation of transfer function of vocal tract extracted from MRI data by FEM. – *The 18th International Congress on Acoustics*, osa II, Kyoto, Japan. 1473–1476.

SONDHI, M. M. 1986: Resonances of a bent vocal tract. – *J. Acoust. Soc. Am.*, **79**(4): 1113–1116.

ŠVANCARA, P. & HORÁČEK, J. 2006: Numerical modelling of effect of tonsillectomy on production of Czech vowels. – *Acta Acustica united with Acustica*, **92**: 681–688.

ŠVANCARA, P., HORÁČEK, J. & PEŠEK, L. 2004: Numerical modelling of production of Czech vowel /a/ based on FE model of the vocal tract. – *Proceedings of International Conference on Voice Physiology and Biomechanics*.

Michael O'Dell & Tommi Nieminen (eds.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 65–72.

Gender and expression of emotions

Teija Waaramaa[1], Anne-Maria Laukkanen[1] & Paavo Alku[2]
[1]University of Tampere, [2]Helsinki University of Technology

Abstract

Some earlier results suggest gender differences in the use of various acoustic parameters related to expression of emotions. The present study tested this hypothesis for simulated emotional expressions by 9 Finnish professional actors (5 males, 4 females). They read a prose extract expressing joy, tenderness, sadness, anger and a neutral emotional state. Stress carrying vowel [ɑː] was extracted for analyses (average sample duration 143 ms, N = 171. Fundamental frequency (F0), equivalent sound level (L_{eq}), duration, voice quality reflecting alpha ratio (SPL(1–5 kHz) − SPL(50 Hz–1 kHz)) and formant frequencies F1–F4 were measured. The samples were inverse filtered using the IAIF method, and the normalized amplitude quotient (NAQ) was calculated ($NAQ = f_{ac}/(d_{peak}T)$) to quantify the closing phase of the glottal pulse form. Differences between the genders were studied by calculating Student's Unpaired *t*-tests (normally distributed parameters) and Mann-Whitney *U*-tests (parameters with skewed distribution) for all subjects' individual ranges and standard deviations of the acoustic parameters. Variation of F0 and F1–F4 was studied as percentages of each subject's lowest values. According to the results, females had a significantly larger variation in NAQ, Alpha ratio and F1–F3. Females appeared to use a wider scale of both voice source and resonance characteristics in the expression of emotions. This may reflect different phonatory and articulatory patterns adopted to optimize linguistic message transfer and may also be related to cultural aspects, such as females' tendency to express emotions more freely compared to males.

Keywords: voice quality, inverse filtering, gender differences, emotions, formants

1 Introduction

Some earlier results suggest gender differences in the use of various acoustic parameters (Whiteside 2001, Diehl *et al.* 1996). Also in the expression of emotions gender differences have been reported for the acoustic parameters (Airas & Alku 2006). Most of the earlier studies, however, have concentrated on the prosodic characteristics in the expressions of emotional content in speech, such as fundamental

frequency (F0) and range of its variation, sound pressure level (SPL) speech rate and phoneme duration (Murray & Arnott 1993, Laukkanen *et al.* 1996, Lieberman & Michaels 1962). F0, SPL and duration have been found to be important discriminating characteristics in the production and perception of emotional information (e.g. Laukkanen *et al.* 1997, Mozziconazzi 1998, Wayland & Jongman 2003). Voice quality has been reported to interact with tempo: Speech produced with hypofunctional voice quality tends to have a longer duration than speech produced with hyperfunctional voice quality (Wayland & Jongman 2003). However, few studies have focused on the individual role of the acoustic voice quality (voice source and filter characteristics) (Fant 1970) in the expressions of emotions and their valence (neutrality/positivity/negativity) (Airas & Alku 2006, Laukkanen *et al.* 1997, Scherer & Wallbott 1994, Gobl & Ní Chasaide 2003, Waaramaa *et al.* 2006; forthcoming).

Voice source characteristics possibly related to phonation type are also known to co-vary with F0 and SPL (Sonesson 1960). Raising F0 increases the open quotient (OQ) of the glottis, while higher intensity decreases it. In general, the speed quotient (SQ) and relative closed time of the glottis (CQ) increase, and the closing time (ClQ) decreases with higher intensity (Sonesson 1960, Sundberg *et al.* 2005; 1993). Faster closing speed of the glottis increases the glottal flow declination rate. Thus, the spectral slope becomes less steep (e.g. Gauffin & Sundberg 1989). Hypofunctional ('breathy') voice quality is characterized by a steep spectral slope and an almost sinusoidal glottal flow pulse shape while a flatter slope of the spectrum and a steeper glottal pulse shape are characteristic to hyperfunctional ('pressed') voice quality (Laukkanen *et al.* 1996, Gauffin & Sundberg 1989). The glottal pulse shape can also be affected by vocal tract inertia (Fant & Lin 1987).

The length of the vocal tract may vary with F0 (e.g. Shipp & Izdebski 1975); raising the F0 typically leads to elevated laryngeal position and thus a shortened vocal tract with higher formant frequencies. Formant frequencies, especially F1, are used in intensity control: Increased mouth opening raises F1 closer to F2, which increases SPL (Fant 1970).

Speech rate may affect voice quality parameters indirectly, e.g. by increasing or decreasing the articulatory movements and thus affecting formant frequencies. The tendency for simultaneous variation of duration and voice quality parameters (Mozziconazzi 1998) seems to suggest a centrally controlled coding of the psychophysiological activity level related to an emotional state.

In the present study, the gender differences in voice source and filter parameters of emotional expressions were focused on by testing this hypothesis in simulated emotional expressions of Finnish professional actors.

2 Subjects and samples

Nine professional actors (5 males and 4 females, aged 26–45) read a paragraph of a Finnish prose text in an anechoic room expressing joy, tenderness, sadness, anger and a neutral emotional state. The main stress carrying vowel [ɑː] was extracted from the word *taakkahan* ([tɑːkːɑɦɑn] 'a burden indeed') for further analyses. The average duration of the samples was 143 ms. The total number of the samples was 171, of which 99 were produced by males and 72 by females. Males' samples were: Neutral (N = 22), sadness (N = 17), joy (N = 22), anger (N = 21), and tenderness (N = 17). Females' samples were: Neutral (N = 18), sadness (N = 17), joy (N = 14), anger (N = 8) and tenderness (N = 16).

3 Analyses

3.1 Listening test

The randomized [ɑː] vowels (N = 171) were replayed to 50 listeners. The listening test was arranged four times altogether in a sound-treated studio, in free field (the listeners did not wear headphones). A digital recorder and Genelec Biamp1019 A loudspeaker were used. The listeners were seated approximately 2.5 meters from the loudspeaker. They heard each sample only once at a time, and their task was to identify the emotions expressed. Twenty one of the samples were repeated randomly in the course of the test in order to enable calculation of intrarater reliability.

3.2 Acoustic analyses

The samples were analyzed for fundamental frequency (F0), duration, equivalent sound level (L_{eq}), and alpha ratio (SPL(1–5 kHz)−SPL(50 Hz–1 kHz), Frøkjær-Jensen & Prytz 1973). Formant frequencies F1–F4 were measured using LTAS and spectrograms. Analyses were made with a signal analysis system named Intelligent Speech Analyser (ISA), developed by Raimo Toivonen, M.Sc.Eng. The vowels were inverse filtered using the IAIF method (Alku 1992). Normalized amplitude quotient, ($NAQ = f_{ac}/(d_{peak}T)$) was calculated to characterize the closing phase features of glottal volume velocity waveform (Alku *et al.* 2002).

3.3 Statistical treatment

The gender differences were studied by calculating Student's unpaired *t*-tests (normally distributed parameters: NAQ, F1–F4, L_{eq}, duration and alpha ratio) and Mann-Whitney *U*-tests (parameter with skewed distribution: F0) for all subjects' individual ranges and standard deviations of F0, L_{eq}, duration, alpha ratio, NAQ and F1–F4.

Variation of F0 and F1–F4 was studied as percentages of each subject's lowest values. SPSS 15 (SPSS Inc., Chicago, IL) was used in the analyses.

4 Results and conclusions

The interrater reliability of the listening test was 0.95 (Cronbach's alpha). The intrarater reliability was satisfactory ($r = 0.417$, $p < 0.001$, Spearman's correlation).

Table 1: Means and standard deviations of variations in F0, L_{eq}, duration, alpha ratio, NAQ and formant frequencies F1–F4, measured from individual subjects' standard deviations in the parameters in emotional expressions. Significance of differences (Student's unpaired *t*-test; Mann-Whitney *U*-test), ns = non-significant = $p > 0.05$.

	Males		Females		
	Mean	(SD)	Mean	(SD)	Significance (p)
F0 (%)	46.7	(16.4)	51.5	(11.4)	ns
L_{eq} (dB)	5.7	(1.2)	6.2	(0.5)	ns
Dur (ms)	28.8	(4.2)	34.7	(5.2)	ns
Alpha (dB)	3.1	(0.9)	5.7	(1.4)	0.015
NAQ	0.023	(0.005)	0.036	(0.010)	0.029
F1 (%)	8.4	(3.4)	14.3	(2.5)	0.025
F2 (%)	5.0	(1.2)	14.5	(8.4)	0.014
F3 (%)	5.5	(2.2)	23.3	(25.5)	0.050
F4 (%)	9.6	(4.7)	10.9	(5.7)	ns

Number of samples from each subject: 8–22.

Table 1 shows the means of the variation (measured as standard deviation) in various acoustic parameters and statistical significance of the differences between the genders in the variation. Results were similar when variation was studied as ranges instead of standard deviations. For instance, the average of 46.7 % for F0 in males means that the standard deviation of F0 among each male subject's emotional samples was on average 46.7 % of each subject's lowest F0 in these samples (typically found in the expression of sadness).

Females had significantly wider variation in NAQ (Figure 1), which is in line with earlier findings (Airas & Alku 2006), and in alpha ratio and F1–F3 (Figure 2). Females appeared to use a wider scale of both voice source and resonance characteristics in the expression of emotions. This may reflect different phonatory and articulatory patterns adopted to optimize linguistic message transfer (e.g. Whiteside 2001, Diehl *et al.* 1996) and may also be related to cultural aspects, such as females'

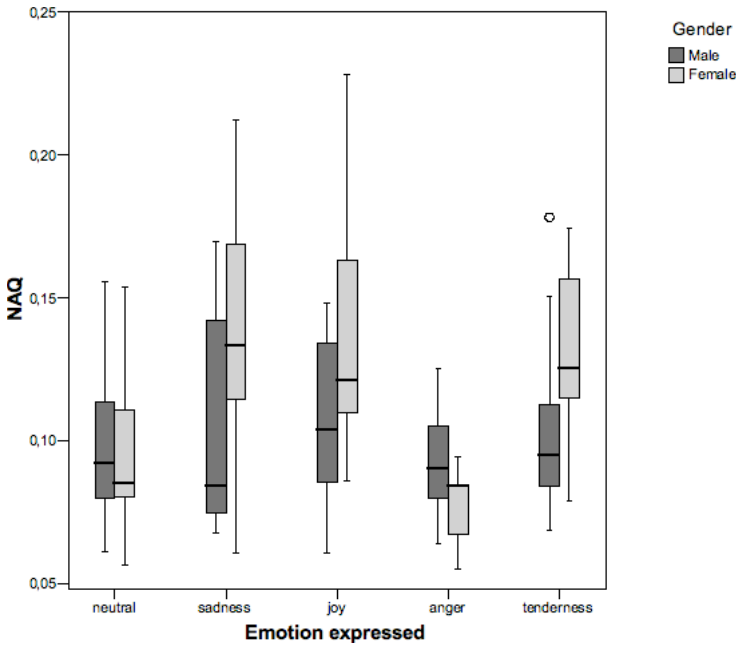


Figure 1: The results of NAQ values of the emotions expressed, both genders.

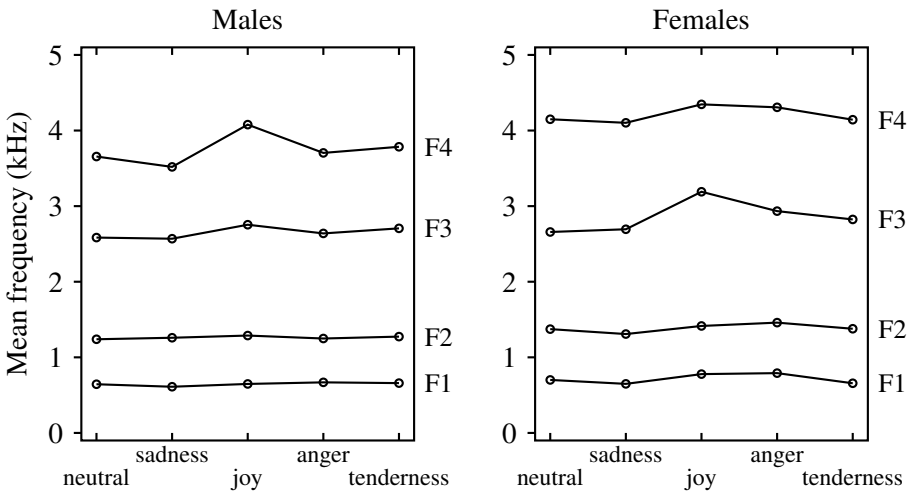


Figure 2: Graph of average formant frequencies in males' (left panel) and in females' (right panel) emotional expression.

tendency to express emotions more freely compared to males (see e.g. Tolkmitt & Scherer 1986). The results need to be interpreted with caution, since the accuracy of formant detection and thus also of inverse filtering is lower for females. However, the results obtained for the more robust alpha ratio measurement support a true gender effect. The differences in NAQ values are shown in Figure 1. It can be observed that females' scale was wider and they tended to use more extremes than males (see Airas & Alku 2006). Mean formant frequencies are presented in Figure 2.

It can be seen in Figure 2 that in both genders, the differences in formant frequencies between emotional expressions are greater for F3 and F4 than for F1 and F2. This seems rational since the lower two formants are more determined by the vowel expressed while the two higher ones may carry more information about e.g. emotional states.

Acknowledgements

This study was supported by the Academy of Finland (grant nos. 200807 and 200859).

References

AIRAS, M. & ALKU, P. 2006: Emotions in vowels segments of continuous speech: Analysis of the glottal flow using the normalised amplitude quotient. – *Phonetica*, **63**: 26–46.

ALKU, Paavo 1992: Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. – *Speech Communication*, **11**(2–3): 109–117.

ALKU, Paavo, BÄCKSTRÖM, Tom & VILKMAN, Erkki 2002: Normalized amplitude quotient for parametrization of the glottal flow. – *Journal of the Acoustical Society of America*, **112**(2): 701–710.

DIEHL, R. L., LINDBLOM, B., HOEMEKE, K. A. & FAHEY, R. P. 1996: On explaining certain male-female differences in the phonetic realization of vowel categories. – *Journal of Phonetics*, **24**: 187–208.

FANT, G. & LIN, Q. 1987: Glottal source-vocal tract acoustic interaction. – *STL-QPSR*, **1**: 13–27.

FANT, Gunnar 1970: *Acoustic Theory of Speech Production, with calculations based on X-ray studies of Russian articulations*. 2nd Edition. The Hague: Mouton.

FRØKJÆR-JENSEN, B. & PRYTZ, S. 1973: Registration of voice quality. – *Brüel & Kjær Technical Review*, **3**: 3–17.

GAUFFIN, J. & SUNDBERG, J. 1989: Spectral correlates of glottal voice source waveform characteristics. – *Journal of Speech and Hearing Research*, **32**: 556–565.

GOBL, C. & NÍ CHASAIDE, A. 2003: The role of voice quality in communicating emotion, mood and attitude. – *Speech Communication*, **40**(1-2): 189–212.

LAUKKANEN, A.-M., VILKMAN, E., ALKU, P. & OKSANEN, H. 1996: Physical variations related to stress and emotional state: A preliminary study. – *Journal of Phonetics*, **24**: 313–335.

LAUKKANEN, A.-M., VILKMAN, E., ALKU, P. & OKSANEN, H. 1997: On the perception of emotions in speech: The role of voice quality. – *Scandinavian Journal of Logopedics, Phoniatrics, Vocology*, **22**(4): 157–168.

LIEBERMAN, P. & MICHAELS, S. B. 1962: Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. – *Journal of The Acoustical Society of America*, **34**(7): 922–927.

MOZZICONAZZI, S. J. L. 1998: *Speech Variability and Emotion: Production and Perception*. Technische Universiteit Eindhoven, Netherlands.

MURRAY, I. R. & ARNOTT, J. L 1993: Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. – *Journal of The Acoustical Society of America*, **93**(2): 1097–1108.

SCHERER, K. R. & WALLBOTT, H. G. 1994: Evidence for universality and cultural variation of differential emotion response pattering. – *Journal of Personality and Social Psychology*, **66**(2): 310–328.

SHIPP, T. & IZDEBSKI, K. 1975: Vocal frequency and vertical larynx positioning by singers and nonsingers. – *Journal of The Acoustical Society of America*, **58**: 5.

SONESSON, B. 1960: *On the Anatomy and Vibratory Pattern of the Human Vocal Folds, with special reference to a photo-electrical method for studying the vibratory movements*. Acta Oto-laryngologica, supplementum 156. Lund, Sweden: The Department of Anatomy and the Department of Otolaryngology, University of Lund.

SUNDBERG, J., FAHLSTEDT, E. & MORELL, A. 2005: Effects on the glottal voice source of vocal loudness variation in untrained female and male voices. – *Journal of The Acoustical Society of America*, **117**: 879–885.

SUNDBERG, J., TITZE, I. & SCHERER, R. 1993: Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. – *Journal of Voice*, **7**: 15–29.

TOLKMITT, F. J. & SCHERER, K. R. 1986: Effect of experimentally induced stress on vocal parameters. – *Journal of Experimental Psychology: Human Perception and Performance*, **12**(3): 302–313.

WAARAMAA, Teija, ALKU, Paavo & LAUKKANEN, Anne-Maria 2006: The role of F3 in the vocal expression of emotions. – *Logopedics, Phoniatrics, Vocology*, **31**(4): 153–156.

WAARAMAA, Teija, LAUKKANEN, Anne-Maria, AIRAS, Matti & ALKU, Paavo forthcoming: Perception of emotional valences and activity levels from vowel segments of continuous speech. – *Journal of Voice*.

WAYLAND, R. & JONGMAN, A. 2003: Acoustic correlates of breathy and clear vowels: The case of Khmer. – *Journal of Phonetics*, **31**(2): 181–201.

WHITESIDE, S. P. 2001: Sex-specific fundamental and formant frequency patterns in a cross-sectional study. – *Journal of the Acoustical Society of America*, **110**(1): 464–478.

Michael O'Dell & Tommi Nieminen (toim.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 73–78.

Onko maahanmuuttajien vieraalla aksentilla väliä?

Minnaleena Toivola, Mietta Lennes & Eija Aho
Helsingin yliopisto

Tiivistelmä

Vuoden 2008 alussa käynnistyneen ProoF-tutkimushankkeen tavoitteena on lisätä foneettista tietämystä suomen ääntämisestä toisena kielenä. Tutkimuksen kohteena ovat pääkaupunkiseudulla asuvat aikuiset maahanmuuttajat. Puhetta äänitetään eri äidinkieliä puhuvilta. Lisäksi äänitetään suomenkielistä vertailuaineistoa. Tutkimuksessa kerätään monipuolinen puheaineisto, josta voidaan tutkia myös mahdollisia ääntämisessä tapahtuvia segmentaalisia ja prosodisia muutoksia pitemmällä aikavälillä.

Avainsanat: suomi toisena kielenä, ääntäminen, aikuiset maahanmuuttajat

1 Johdanto

Suomi toisena kielenä -opetuksen tavoitteena on toiminnallinen kielitaito, joka auttaa oppijaa selviytymään suomen kielellä jokapäiväisessä elämässään. Ääntämisen oppimiselle ei ole asetettu erillisiä tavoitteita. Tämä johtunee osittain siitä, että lasten kouluopetus on etusijalla S2-opetuksen suunnittelussa. Lapsillahan ääntämisen oppiminen ei ole useinkaan ongelma. Ääntämisen oppimisesta ja opetuksesta on kuitenkin keskusteltu vilkkaasti (ks. esim. Iivonen 2005a;b, Vihanta 1990), mutta systemaattista foneettista tutkimusta asiasta ei Suomessa vielä ole tehty.

Suomalaiset hyväksyvät yleensä erilaiset suomen kielen ääntämistavat. Silti maahanmuuttajien kielitaidolla on vaikutusta heidän hyvinvoinnilleen ja selviytymiselleen arkielämän tilanteissa. Monella alalla yhteiskunnassamme näyttää menestymisen edellytyksenä olevan lähes virheetön suomen kielen suullinen taito. Vaikka kielitaito muutoin olisi hyvä, poikkeava ääntäminen voi estää työnhakijaa saamasta koulutustaan ja ammatillista osaamistaan vastaavaa työtä. Maahanmuuttajan puheessa esiintyvä vieras aksentti saattaa häiritä tai ärsyttää kuulijaa vaikka hän puhuisikin ymmärrettävää suomea. Tutkimuksissa on todettu, että esimerkiksi englannin ääntämisessä ja intonaatiossa esiintyvät epäröinnit ja ei-äidinkieliset piirteet häiritsevät natiivia kuulijaa enemmän kuin kielioppivirheet (Fayer & Krasinski 1987).

Ääntäminen ja vieras aksentti liittyvät oleellisesti kielitaitoon. Silti asiaa ei ole huomioitu kielenopetuksessa riittävästi Suomessa, eikä muuallakaan. Tämä johtuu

ainakin Suomessa pitkälti puutteellisesta tutkimustiedosta. Foneettista tietoa esimerkiksi englannista toisena kielenä on jonkin verran, mutta sitäkään ei ole osattu hyödyntää tarpeeksi toisen kielen opettamisen ja oppimisen tutkimuksessa (Derwing & Munro 2005). Suurin osa vieraan aksentin vahvuuden ja ymmärrettävyyden tutkimuksesta on tähän mennessä keskittynyt äännepiirteiden tutkimukseen. Mutta on edelleenkin epäselvää, miten äidinkielen ja opittavan kielen väliset fonologiset erot ilmenevät aikuisilla puhujilla toisen kielen oppimisen eri vaiheissa. Asian selvittäminen edellyttää pitkittäistutkimusta (Han & Selinker 2005). Aikuisten puhujien toisen kielen ääntämisen systemaattista pitkittäistutkimusta on toistaiseksi tehty hyvin vähän (ks. esim. Hansen 2006). Toisen kielen prosodiaa on tähän mennessä tutkittu vielä vähemmän kuin äännepiirteitä (ks. esim. Bannert 1979, Munro 1995, Magen 1998, Jilka 2000, Trofimovich & Baker 2006, Aho & Toivola 2008).

2 Taustaa

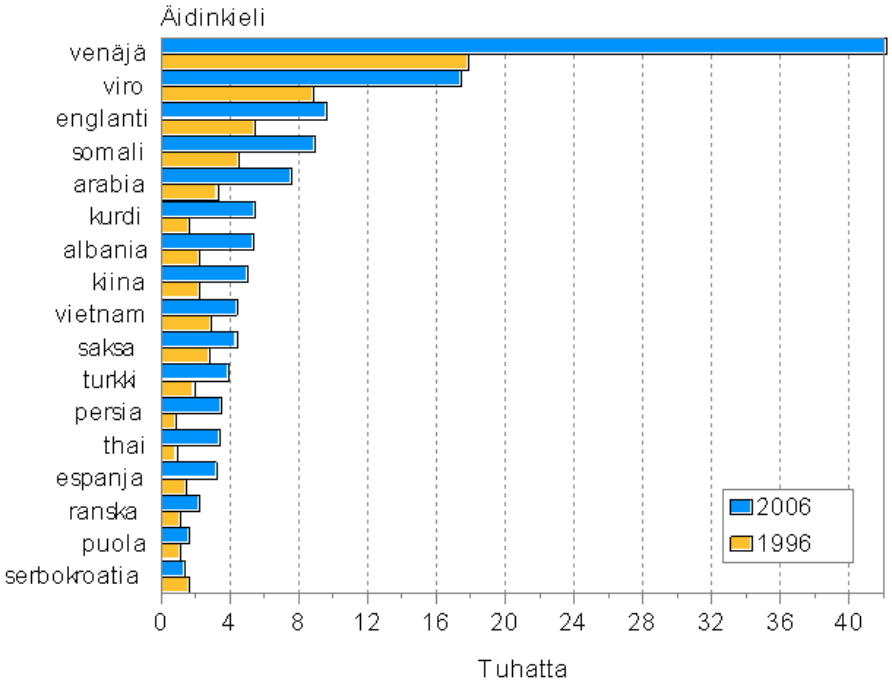
Vuonna 2006 Suomessa asui 156 827 äidinkielenään muuta kieltä kuin suomea, ruotsia tai saamea puhuvia. Näistä 42 182 puhui venäjää, 17 489 puhui viroa, 8 990 somalia ja 7 564 arabiaa äidinkielenään. Lisäksi 80 602 puhui yhtä noin sadasta muusta kielestä. (Tilastokeskus 2006.) Maahanmuuttajien määrä on kasvanut nopeasti viime vuosien aikana (kuva 1). Maahanmuuttajat ovat pääasiassa työikäisiä.

Suomalaiset suhtautuvat eri kulttuureista tuleviin maahanmuuttajiin eri tavoin. Suoranaista syrjintää voivat puutteellisen kielitaidon lisäksi aiheuttaa puhujan etninen tausta, uskonto tai muut kulttuuriset tekijät. Tutkimusten mukaan äidinkieleltään venäläiset ja somalit kokivat Suomessa useammin syrjintää työnhakutilanteissa kuin äidinkieleltään vietnamilaiset tai turkkilaiset (Suomen venäjänkielisen väestönosan kysymyksiä 2002). Sen sijaan äidinkieleltään virolaisia kohdellaan täällä yleensä paremmin (Jaakkola 2005).

3 Tutkimushankkeen esittely

Vuoden 2008 alussa käynnistyi Suomen Akatemian rahoittama tutkimushanke ProoF – Pronunciation of Finnish by immigrants in Finland, jossa tutkimme pääkaupunkiseudulla asuvien aikuisten maahanmuuttajien suomen ääntämistä. Hankkeen päätavoitteena on selvittää, mitä ääntämisongelmia esiintyy puhuttaessa suomea toisena kielenä. Tutkimushankkeen tavoitteet:

- Eri äidinkieliä puhuvien suomen ääntämisessä ilmenevien tyypillisten segmentaalisten ja prosodisten piirteiden kartoittaminen
- Vieraan aksentin arviointi
- Suomenkielisen keskustelupuheen fonetiikan kuvaus
- Suomen kielen ääntämisen oppimateriaalin laadinta



Kuva 1: Suurimmat vieraskielisten ryhmät 1996–2006 (Tilastokeskus 2006).

4 Tutkittavat kielet

Tutkimushankkeen aikana keräämme korpuksen erikielisten maahanmuuttajien puhumasta suomesta. Kielet valittiin maahanmuuttajien määrän perusteella mahdollisimman monesta kieliryhmästä. Kielet ovat maahanmuuttajien lukumäärän mukaisessa järjestyksessä seuraavat:

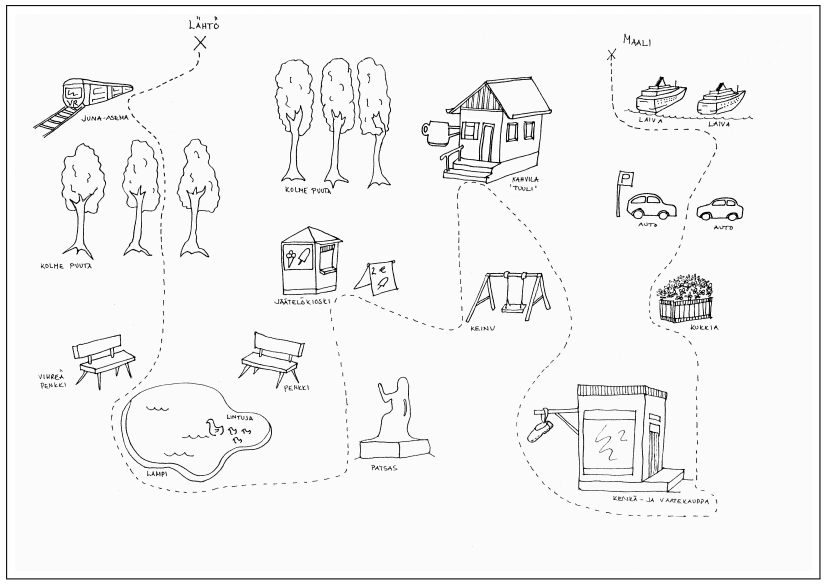
venäjä, somali, arabia, kurdi, kiina, vietnam, turkki, thai, hindi, tagalog

5 Tutkimusmenetelmät ja aineisto

Äänitämme puhetta noin kolmeltakymmeneltä maahanmuuttajalta sekä kymmeneltä syntyperäiseltä suomalaiselta. Seurantatutkimuksena äänitämme puolen vuoden välein kymmentä äidinkieleltään venäläistä ja kiinalaista. Pyrimme selvittämään pysyvätkö puhujan äidinkieleen liittyvät ääntämispiirteet samoina vai vähenevätkö ne vaiheittain suomen kielen oppimisen myötä. Keräämme kaikilta puhujilta mahdollisimman tarkat lingvistiset ja sosiaaliset taustatiedot.

Teemme äänitykset pääasiassa puhetieteiden laitoksen studiossa Helsingissä. Äänitettävä aineisto sisältää tekstien ääneen lukua, kuvakerrontaa, karttatehtäviä (kuva 2) ja spontaania keskustelua. Annotoimme puhekorpuksen Praat-ohjelmalla (Boersma & Weenink 2008). Kerätty puhekorpus on tarkoitus saattaa muiden tutkijoiden käyttöön tutkimuksemme jälkeen.

Karttatehtävillä pyritään innostamaan keskusteluun, jossa puhujien huomio kiinnittyy tehtävään, eikä omaan puheeseen. Kaksi puhujaa katsoo samanaikaisesti hieman erilaisia karttoja. Toiseen karttaan on merkitty reitti, jota pitkin molempien on tarkoitus päästä samaan päätepisteeseen. Tehtävän aikana puhujat keskustelevat siitä, miten perille päästään. Puhujat eivät tiedä, että kartat ovat erilaisia. Tämä aiheuttaa yleensä vilkkaan ja vapautuneen keskustelun.



Kuva 2: Esimerkki karttatehtävästä (mukaillen Brown *et al.* 1983), joissa kuvien erilaisuus houkuttelee spontaaniin keskusteluun.

Viitteet

AHO, Eija & TOIVOLA, Minnaleena 2008: Venäläisten maahanmuuttajien suomen prosodiasta. – *Virittäjä*, **112**:3–23.

BANNERT, Robert 1979: *Ordprosodi i invandrarundervisningen*. Praktisk lingvistik 3. Institutionen för lingvistik. Lund: Lunds universitet.

BOERSMA, Paul & WEENINK, David 2008: Praat—doing phonetics by computer. <http://www.fon.hum.uva.nl/praat/>.

BROWN, G., ANDERSON, A., SHILLCOCK, R. & YULE, G. 1983: *Teaching Talk: Strategies for Production and Assessment*. Cambridge: Cambridge University Press.

DERWING, Tracey M. & MUNRO, Murray J. 2005: Second language accent and pronunciation teaching: A research-based approach. – *TESOL Quarterly*, **39**:379–397.

FAYER, J.M. & KRASINSKI, E. 1987: Native and nonnative judgements of intelligibility and irritation. – *Language Learning*, **37**:313–326.

HAN, Z. & SELINKER, L. 2005: Fossilization in L2 learners. – E. Hinkel (toim.), *Handbook of Research in Second Language Teaching and Learning*. Mahwah: Erlbaum. 455–468.

HANSEN, Jette. G. 2006: *Acquiring a Non-native Phonology: Linguistic Constraints and Social Barriers*. London: Continuum.

IIVONEN, Antti 2005a: Fonetiikan merkitys kielenomaksumisessa ja -opetuksessa. – Antti Iivonen, Reijo Aulanko & Martti Vainio (toim.), *Monikäyttöinen fonetiikka*. Fonetiikan laitoksen monisteita 21. Helsingin yliopisto. 45–64.

IIVONEN, Antti (toim.) 2005b: *Puheen salaisuudet: Fonetiikan uusia suuntia*. Gaudeamus. ISBN 951-662-918-0.

JAAKKOLA, Magdalena 2005: Suomalaisten suhtautuminen maahanmuuttajiin vuosina 1987–2003. Työpoliittinen tutkimus 286, Helsinki: Työministeriö. <http://www.mol.fi/mol/fi/99_pdf/fi/06_tyoministerio/06_julkaisut/06_tutkimus/tpt286.pdf>.

JILKA, Matthias 2000: *The Contribution of Intonation to the Perception of Foreign Accent*. Arbeiten des Instituts für Maschinelle Sprachverarbeitung (AIMS) 6(3). University of Stuttgart.

MAGEN, Harriet 1998: The perception of foreign-accented speech. – *Journal of Phonetics*, **26**:381–400.

MUNRO, Murray. J. 1995: Nonsegmental factors in foreign accent: Ratings of filtered speech. – *Studies in Second Language Acquisition*, **17**:17–34.

SUOMEN VENÄJÄNKIELISEN VÄESTÖNOSAN KYSYMYKSIÄ 2002: Venäjän ja Itä-Euroopan instituutti, Etnisten suhteiden neuvottelukunnan asettaman työryhmän raportti.

TILASTOKESKUS 2006: Väestörakenne 2006. <http://www.stat.fi/til/vaerak/2006/vaerak_2006_2007-03-23_tie_001.html>.

TROFIMOVICH, Pavel & BAKER, Wendy 2006: Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. – *Studies in Second Language Acquisition*, **28**:1–30.

VIHANTA, V. V. 1990: Suomi vieraana kielenä foneettiselta kannalta. – J. Tommola (toim.), *Vieraan kielen ymmärtäminen ja tuottaminen. AFinLA:n vuosikirja 1990*. AFinLA. 199–225.

Michael O'Dell & Tommi Nieminen (toim.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 79–82.

Puheen ominaispiirteet neurofibromatoosi 1 -potilailla: Alustava tutkimus

Lotta Alivuotila[1], Jussi Hakokari[1], Vivian Visnapuu[1,2],
Sirkku Peltonen[2], Juha Peltonen[1],
Risto-Pekka Happonen[1,2] & Olli Aaltonen[3]
[1]Turun yliopisto, [2]Turun yliopistollinen keskussairaala, [3]Helsingin yliopisto

Tiivistelmä

Neurofibromatoosi 1 (NF1) on pääasiassa hermostoon ja ihoon vaikuttava perinnöllinen sairaus, jonka tiedetään aiheuttavan myös erilaisia kielellisiä ongelmia. Aikaisemmissa tutkimuksissa sairauteen liittyviä puheen ominaispiirteitä on käsitelty melko pintapuolisesti, eikä perusteellista foneettista katsausta aiheeseen ole vielä tehty. Tämän alustavan tutkimuksen tavoitteena oli luoda yleiskuvaus NF1-sairauteen liittyvistä puheen ominaisuuksista fonetiikan näkökulmasta. Tutkimusaineistona oli 19 suomalaista miespotilasta. Alustavat tulokset osoittavat, että tyypillisimpiä sairauteen liittyviä puheen ominaispiirteitä ovat poikkeavuudet fonaatiossa, intonaation säätelemisen ongelmat sekä puheen sujuvuuden ja artikulaation häiriöt.

Avainsanat: puhehäiriöt, NF1, intonaatio, artikulaatiovaikeudet

1 Johdanto

Neurofibromatoosi 1 (NF1) on vallitsevasti periytyvä useisiin elinjärjestelmiin vaikuttava sairaus. Sen esiintyvyys on noin 1/3 500, ja Suomessa NF-potilaita on arviolta 1 500. NF1:n taudinkuva on hyvin vaihteleva. Tyypillisiä oireita ovat iholla esiintyvät maitokahviläiskät, taivealueiden kesakkoisuus, Lischin nodulukset (hamartoomat silmän värikalvossa), ja neurofibroomat, jotka ovat ääreishermoissa sijaitsevia hyvänlaatuisia kasvaimia. Sairauteen liittyy myös kognitiivisia ongelmia, kuten oppimisvaikeuksia sekä sosiaalisen hahmottamisen ongelmia. Häiriöitä esiintyy myös kielen eri tasoilla, esim. syntaksiin, semantiikkaan ja fonologiaan liittyen. (Peltonen & Jaakkola 1991, Young *et al.* 2002, Lorch *et al.* 1999).

Aikaisemmissa tutkimuksissa NF1-potilaiden kielellisiä ongelmia on käsitelty melko vähän ja puheen ominaisuuksia on tarkasteltu melko pintapuolisesti (ks. esim.

Lorch *et al.* 1999). Tämän alustavan tutkimuksen tarkoituksena oli tarkastella suomalaisten NF1-miesten puheen ominaisuuksia foneettisesta näkökulmasta. Tutkimus on osa Turun yliopiston NF-tutkimuskonsortion laajempaa tutkimushanketta.

2 Aineisto ja menetelmät

Tutkittavana oli 19 äänenmurroksen ohittanutta suomalaista NF1-miespotilasta. Potilaat rekrytoitiin Turun yliopistollisen keskussairaalan NF-klinikalta sekä Suomen neurofibromatoosiyhdistyksen tilaisuuksista eri puolilta Suomea. Potilaat olivat iältään 16–65-vuotiaita.

Tutkimustilanteessa potilaiden kanssa keskusteltiin ensin vapaasti, minkä jälkeen he suorittivat tutkijan opastuksella puheen ominaisuuksia kartoittavan testin. Tutkimuksessa käytettiin soveltuvin osin suomenkielisiä versioita logopedisista testeistä Scoring Speech Examination (Keller 1990) sekä Frenchay Dysarthria Assessment (Enderby 1981). Tutkimuspaikka vaihteli, mutta tutkimus tehtiin aina suljetussa, rauhallisessa huoneessa ilman ulkoisia häiriötekijöitä. Tutkimustilanteessa olivat läsnä vain potilas ja tutkija. Tutkimus kesti keskusteluineen noin 20–25 minuuttia. Tutkimustilanne nauhoitettiin ja kaksi foneetikkoa analysoi aineiston jälkikäteen sekä auditiivisesti että akustisesti. Analyysit tehtiin toisistaan riippumatta, ja tuloksia vertailemalla muodostettiin lopullinen yhteenveto tutkimuksen tuloksista.

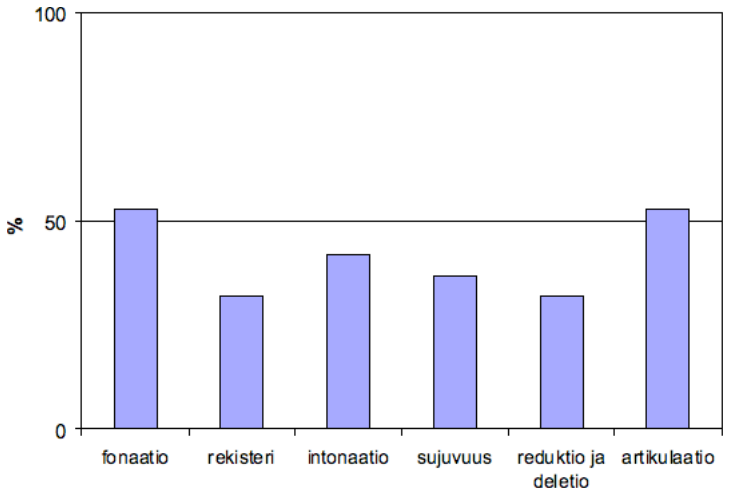
3 Tulokset ja yhteenveto

Tutkimuksessa puhetta tarkasteltiin eri näkökulmista ja poikkeavuuksia havaittiin puheen eri osa-alueilla. Yksilökohtainen vaihtelu oli huomattavaa; osalla potilaista puheen poikkeavuuksia ei esiintynyt juuri lainkaan, kun taas toisilla puheen ongelmat olivat huomattavia.

Kuvassa 1 on esitetty alustavat arviot tyypillisistä neurofibromatoosiin liittyvistä puheen poikkeavuuksista.

Noin puolella tutkittavista oli havaittavissa erilaisia poikkeavuuksia fonaatiossa. Potilaiden äänessä esiintyi mm. käheyttä ja vuotoisuutta. Osasävelrakenteessa havaittiin vääristymiä, minkä vuoksi perustaajuuden (F0) määrittäminen oli toisinaan hankalaa. Osalle potilaista pitkäkestoisen fonaation ylläpitäminen oli haasteellista, ja pitkäkestoiset yksittäiset vokaalit tuotettiin usein epätasaisesti. Myös nasaalisuutta esiintyi. Noin 30 %:lla potilaista äänirekisteri oli poikkeuksellisen kapea ja noin 40 %:lla oli ongelmia intonaation säätelemisessä. Vaikeus säädellä intonaatiota erilaisissa kielellisissä yhteyksissä näyttää olevan tyypillinen NF1-puheeseen liittyvä piirre.

Vajaalla 40 %:lla potilaista havaittiin poikkeavuutta puheen sujuvuudessa ja suunnilleen kolmanneksella oli taipumus redusoida tai deletoida äänteitä, tavuja tai jopa



Kuva 1: Alustava arvio tyypillisistä NF1-potilailla esiintyvistä poikkeavuuksista puheen eri osa-alueilla.

suurempia sanan osia. Noin puolella potilaista esiintyi jonkinlaisia poikkeavuuksia artikulaatiossa. Poikkeavuudet tyypillisesti koskivat äänteitä [r] tai [s], sekä jossakin määrin myös äännettä [k]. Poikkeava [r] tuotettiin tyypillisesti täryttömänä, soinnillisena tai soinnittomana approksimanttina. Poikkeava [s] oli yleensä ”suhumainen”, ja energia sijoittui poikkeuksellisen matalille taajuuksille. Osalla potilaista [k]:n sulkeuma oli toisinaan epätäydellinen, minkä seurauksena energiaa saattoi esiintyä koko sulkeuman ajan. Tavallisesti soinnittomassa klusiilissa sulkeuman aikana ei esiinny lainkaan energiaa.

Yleisesti voidaan todeta, että NF1-sairauteen näyttää liittyvän erilaisia puheen häiriöitä, joista tyypillisimpiä ovat poikkeavuudet fonaatiossa, perustaajuuden säätelyssä, puheen sujuvuudessa ja artikulaatiossa. Edellä esitetyt havainnot ovat vielä alustavia ja suuntaa antavia. Ne kuitenkin osoittavat, että perusteellinen, laajemmalla potilasaineistolla tehty foneettinen tutkimus on tarpeen.

Kiitokset

Haluamme kiittää professori Anna-Maija Korpijaakko-Huuhkaa (Logopedia, Tampereen yliopisto) logopedisen asiantuntemuksen tarjoamisesta.

Viitteet

ENDERBY, P. 1981: Frenchay dysarthria assessment. Frenchay Hospital; Bristol BS16 1LE, England, UK. Lehtihalmes, Matti 1981 (suom.), Korpijaakko-Huuhka, Anna-Maija 2006 (suomennoksen päivitys).

KELLER, E. 1990: Instructions for scoring the speech examination (SE). Versio 2.0. Julkaisematon käsikirjoitus. Werner, Stefan, Tuomainen, Jyrki & Lehtihalmes, Matti 1994 (julkaisematon suomenkielinen versio), Korpijaakko-Huuhka, Anna-Maija 2007 (uusi puhtaaksikirjoitus).

LORCH, M., FERNER, R., GOLDING, J. & WHURR, R. 1999: The nature of speech and language impairment in adults with neurofibromatosis 1. – *Journal of Neurolinguistics*, **12**(3–4):157–165.

PELTONEN, Juha & JAAKKOLA, Sirkku 1991: Neurofibromatoosi. – *Duodecim*, **107**:1126–1134.

YOUNG, Helen, HYMAN, Shelley & NORTH, Kathryn 2002: Neurofibromatosis 1: Clinical review and exceptions to the rules. – *Journal of Child Neurology*, **17**(8):613–621.

Michael O'Dell & Tommi Nieminen (toim.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 83–96.

Makro- ja mikroprosodian yhteensovittaminen perusävelkontuurissa

Antti Iivonen[1], Tapio Seppänen[2],
Kai Noponen[2] ja Juhani Toivanen[2]
[1]Helsingin yliopisto, [2]Oulun yliopisto

Tiivistelmä

Mikroprosodinen variaatio aiheuttaa perussävelkontuurissa (F0) katkoksia ja vaihteluita. Niitä voidaan selittää kontekstuaalisilla tekijöillä tai äänteiden luontaisilla ominaisuuksilla. Makroprosodisena muotona tarkastellaan irrallaan äännettyjen suomenkielisten sanojen yleistä F0-kontuurimuotoa ja verrataan sitä irrallisen lauseen muotoon. Näissä F0-alkuhuippu edustaa tauonjälkeistä uutta aloitusta sekä aloitushuippuun sattuvaa painollista tavua. Mikroprosodian tarkastelu koskee konsonanttien (*h*, *j*, *k*, *l*, *m*, *n*, *p*, *r*, *s*, *t*, *v*) ja vokaalien (*a* ja *i*) vaikutuksia. Mies- ja naispuhujan tuottamien sanojen äänitietokannasta valittiin mikroprosodisen tutkimuksen kannalta kiinnostavia ryhmiä. Kunkin ryhmän perustaajuuskontuurien laadittiin tilastollinen jakauma 10 ms:n välein 200 ms:n ajalta käyttäen TVRPVA-ohjelmaa. Ryhmien jakaumia vertailemalla todettiin, että ensimmäisellä äänteellä on selvä ja säännönmukainen vaikutus vaihteluun, mikä selittää alun suurta vaihtelua. Mies- ja naispuhujalla konsonanttien mikroprosodinen vaikutus oli joistakin eroista huolimatta hämmästyttävän samansuuntainen. Se ilmeni tauonjälkeisessä F0-kontuurissa etenkin siten, että korkeampitaajuinen F0-aloitus esiintyy obstruenttien jälkeen verrattuna resonantteihin (kuvat 5–6). Mies- ja naispuhujalle yhteistä oli myös, että resonanttikonsonanttien jälkeinen F0-kontuuri oli matala ja suunnaltaan laskeva. Vokaalit *a* ja *i* olivat molemmilla nousevia. Eri ryhmissä sanojen alun F0-huipun mediaanikäyrän vaihtelu osoittautui vähäiseksi, mikä voidaan tulkita makroprosodian ominaisuudeksi.

Avainsanat: makroprosodia, mikroprosodia, perustaajuus, suomi

1 Johdanto

Tavallinen havainto on, että *mikro*prosodinen variaatio aiheuttaa perussävelkontuurissa (F0) katkoksia ja vaihteluita. Niitä voidaan selittää kontekstuaalisilla tekijöillä tai äänteiden *inherenteillä*, luontaisilla eli ominaispiirteillä. Kontekstuaalisiin tekijöihin kuuluu soinnittomien klusiilien jälkeisen laskevan F0-piikin esiintyminen. Vokaalien väljyysasteeseen liittyvä F0:n korkeuden vaihtelu on esimerkki luontaisista

piirteistä (Meister & Werner 2006). Termi *intrinsinen allofoni* (*intrinsic allophone*; Wang & Fillmore 1961, Trask 1996, 184) viittaa tapauksiin, joiden foneettinen luonne on kokonaan selitettävissä kontekstuaalisen ympäristön perusteella (etisempi [k] etuvokaalien edellä), kun taas *ekstrinsinen allofoni* (*extrinsic allophone*; Trask 1996, 138) on säännönalainen (englannin tumma *l* sanoissa *full* ja *film* erotuksena kirkkaalle *l*:lle: *left*, *loop*). F0:n inherentit ilmiöt on tunnettu jo kauan (Lehiste 1976), mutta mikroprosodian käsitteen lienee ottanut käyttöön vasta Di Cristo (1978; 1982). Yhteistä inherenteillä ja intrinsisillä ominaisuuksilla on, että ne ovat puhumisprosessissa fonologisten sääntöjen jälkeisiä ja johtuvat puhe-elinten toimintatavasta.

Perussävelen (F0:n) *makro*prosodinen kontuurimuoto määräytyy puhujan asettamista funktionaalisista tavoitteista. Mikäli kaikki funktionaaliset ja pragmaattiset tekijät ovat samat kahdessa ilmauksessa, mutta toisessa on vain soinnillisia äänteitä ja toisessa myös soinnittomia, ilmenee niiden F0-kontuureissa suuria eroja mikroprosodian vuoksi. Niiden makroprosodisen (prosodisen fonologian mukaisen) muodon voidaan väittää olevan silti identtinen. Tarkastelun kohteena on tässä, miten mikroprosodia vaikuttaa makroprosodiaan ja miten niiden yhteensovittaminen ilmenee.

Makroprosodisena muotona tarkastelemme irrallaan äännettyjen suomenkielisten sanojen yleistä F0-kontuurimuotoa ja vertaamme sitä irrallisen lauseen muotoon. Näissä F0-alkuhuippu edustaa tauonjälkeistä uutta aloitusta sekä aloitushuippuun sattuvaa painollista tavua. Mikroprosodian tarkastelu koskee konsonanttien (*h*, *j*, *k*, *l*, *m*, *n*, *p*, *r*, *s*, *t*, *v*) ja vokaalien (*a* ja *i*) vaikutuksia. Sana-aineiston puhujina olivat miespuolinen TP ja naispuolinen MW. Lauseaineiston puhuja oli mies ToPa.

2 Menetelmät ja puheaineistot

Analyyseissä on käytetty seuraavia ohjelmia.

- PRAAT-ohjelmalla (Boersma & Weenink 2008) voidaan monipuolisesti käsitellä äänitiedostoja ja F0-käyriä (kuvantaa, segmentoida, annotoida, laskea tilastollisia ominaisuuksia ym.). Analyyseissä oli käytössä versio 4.4. F0-analyyseissä sovellettiin esisuodatuksen min/max-rajoina 50–170 Hz (TP) ja 100–330 Hz (MW). Koko $2 \times 10\,000$ sanan äänitiedostot muutettiin Praatilla PitchTier TextFile short-muotoisiksi tiedostoiksi, joita TVRPVA-ohjelma lukee.

- Java-kielisellä TVRPVA-ohjelmalla voidaan käsitellä suuria F0-kontuurimassoja ja visualisoida valitun F0-kontuurijoukon jakauma Hz- tai puolisävelasteikolla (psa) ajan suhteen. Kukin käyrä alkaa samasta ajallisesta nollapisteestä. Jakaumasta voidaan esittää muun muassa valittu joukko F0-käyriä, valitun joukon persentiilijakauma, minimi- ja maksimikäyrä, kahden joukon jakaumat vertaillen yhteisessä kuvassa ja yksittäisen kontuurin sijainti jakaumassa (Iivonen *et al.* 2004). Analyyseissä käytettiin optioita linear/data points, puolisävelasteikko kantalukuna 16,35, persentiilijakauma 5, 25, 50, 75, 95 % (valkoinen

viiva kuvissa = mediaani), tarkasteluaika enimmäkseen ensimmäiset 200 ms, puolisävelasteikon min/max-rajat 18–40 psa (= noin 48–170 Hz; TP) ja 31–51 psa (= noin 100–320 Hz; MW). Ohjelma lukee Praatin PitchTier-tiedostoja tyyppiä ”ooTextFile short”. Niissä F0 on mitattu 10 ms:n askelin. Prosessoinnissa TVRPVA-ohjelma laskee F0:n tilastollisen jakauman jokaisen 10 ms:n kohdalla. Aineiston käsittely on erittäin nopeaa: yli tuhannenkin tiedoston luku ja prosessointi visualisoinniksi vie muutaman sekunnin.

Puheaineistona oli Suomen Akatemian suomen kielen fonetiikkaa koskevassa hankkeessa (1999–2000) äänitetty sanatietokanta (puhujat TP ja MW) sekä Tekesin Fenix/Usix-hankkeessa äänitetty lausetietokanta (puhuja ToPa). Sana-aineiston koostamisessa oli mukana Pirkko Kukkonen. Äänityksen tekivät Martti Vainio ja joukko fonetiikan opiskelijoita. Puhujan ToPa lauseet äänitti Liisa Vilhunen. Puhujat ja äänitysominaisuudet olivat:

- Nais- (MW) ja miespuhujan (TP) $2 \times 10\,000$ suomenkielistä sanaa. Puhujat ovat ammattipuhujia: MW oli äänitysten aikana TV-kuuluttajana ja TP on muun muassa TV:n suomeksi käännettyjen asiaohjelmien suomenkielinen puhuja. He saivat ohjeen hengittää sanojen välillä ja välttää luettelointonaatiota. Tämä onnistui hyvin. Yhdellä äänityskerralla puhuttujen sanojen maksimimäärä oli noin 500.
- Äänitys: Äänitysstudio; sankamikrofoni AKG CC420; DAT-nauhoitin
- Miespuhujan (ToPa) lukemat suomenkielistä 100 lausetta. Ne oli keinotekoisesti muokattu eräästä arkkitehtuuria käsittelevästä julkaistusta niin, että kussakin lauseessa oli 7 yksinkertaista sanaa. Lauseilla oli siis semanttisia yhtymäkohtia, mutta lauseet oli sekoitettu, ja puhuja sai ohjeen lukea ne toisistaan riippumatta. Tämä onnistui toivotulla tavalla.
- Äänitys: Äänitysstudio; sankamikrofoni AKG C444L; tietokonetallennus (PowerMacintosh G3).

3 Makro- ja mikroprosodia

Makroprosodinen yksikkö on tässä irrallaan, kontekstia vailla olevan sanan F0-kontuuri. Sen samankaltaisuus kontekstittoman lausuman kontuurin kanssa ilmenee seuraavasta vertailusta (kuva 1). Vertailussa kohteina ovat:

1. Puhujan TP tuottamien 100 irrallisen sanan F0-kontuurien 90 %:n persentiilijakauma. Sanat olivat 1–4-tavuisia, enemmistö 2- ja 3-tavuisia. Puolet sanoista alkoi *a*:lla ja toinen puoli *i*:llä. Maksimikesto oli noin 0,8 s.

2. 50 irrallaan äännettyä yhdyssanaa, joissa oli vähintään 3 yhdysosaa (esim. *elin+keino+rakenne*). Maksimiaika 1,5 s. Puhuja TP.

3. Miespuhujan ToPa lukemat 100 irrallista lausetta.

Makrokontuuri toteutuu seuraavasti. Kussakin visualisointikuvassa ilmenee F0-käyrien jakaumassa korkea aloitushuippu, jota seuraa ensin nopea, vähintään 3,5 psa:n lasku ja sen jälkeen lasku loppurajalle. Huippuarvo on mitattu kohdasta, jossa F0-mediaani saavuttaa ensi kerran huippukorkeuden. Tarkempi vertailu esitetään taulukossa 1.

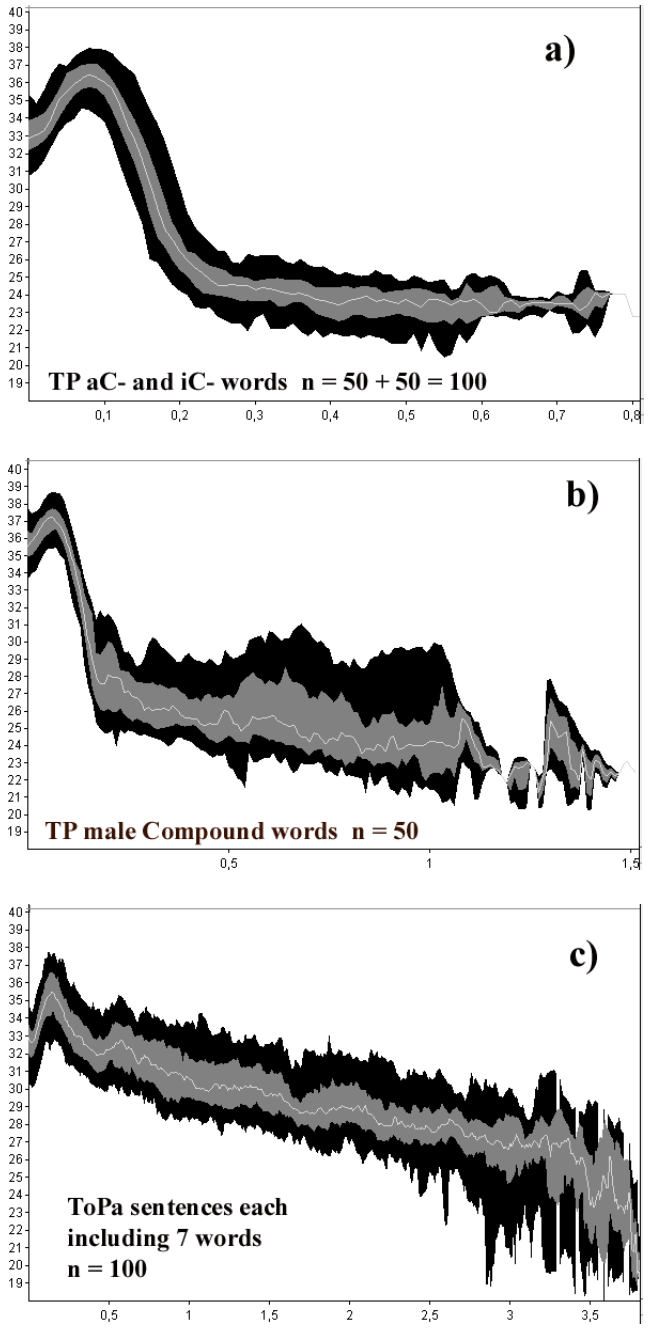
Puhujan ToPa ääniala ei poikkea suuresti TP:n äänialasta, joten hänen jakaumansa on varsin hyvin vertailukelpoinen. Käyrämuotojen voidaan katsoa edustavan tyypillistä neutraalia (*default*) muotoa, joka toteutuu sekä kontekstittomissa irrallaan äännetyissä sanoissa että yksinkertaisissa lauseissa. Niissä alkuhuippu edustaa tauonjälkeistä uutta aloitusta — lauseissa uutta topiikkia — sekä aloitushuippuun sattuvaa painollista tavua. Yhdyssana-aineiston (b) jakaumaan vaikuttavat yhdysosien sivupainot, jotka aiheuttavat huippuja F0-käyriin ja persentiilijakauman leventymiä sen keskivaiheilla. Lausejakaumassa (c) lopun sekavuus johtuu siitä, että lauseet päättyvät eri aikoina, jolloin muun muassa mediaanikäyrän laskentaan on eri aikapisteissä niukemmin tapauksia kuin jakauman alkupuolella.

Taulukko 1: Kuvan 1 mediaanikäyrien mukaiset kääntöpisteet (psa ja ms).

	alku	huippu		nopea lasku huipusta		loppu	
	(psa)	(ms)	(psa)	(ms)	(psa)	(s)	(psa)
a	33	70	36,5	210	26	0,8	24
b	36	50	37,5	200	28	1,5	23,5
c	33	140	35,5	425	32	4,0	24

	huipun ja alun erotus (psa)	huipun ja alkulaskun erotus (psa)	huipun ja lopun erotus (psa)
a	3,5	10,5	12,5
b	1,5	9,5	14,5
c	2,5	3,5	11,5

Mittausarvoissa huomattavaa ovat muun muassa seuraavat yksityiskohdat. Yhdyssana-aineistossa (b) mediaanin alkukorkeus on 3 psa korkeammalla kuin 1–4-tavuisten yksinkertaisten sanojen aineistossa (a). Edellisissä F0-huippu on 20 ms aikaisempi ja lauseissa paljon myöhäisempi: 140 ms. F0-huipun jakaumalaajuus (90 % tapauksista huipun kohdalla) on kapeampi yhdyssana-aineistossa. Aineistoissa b ja c

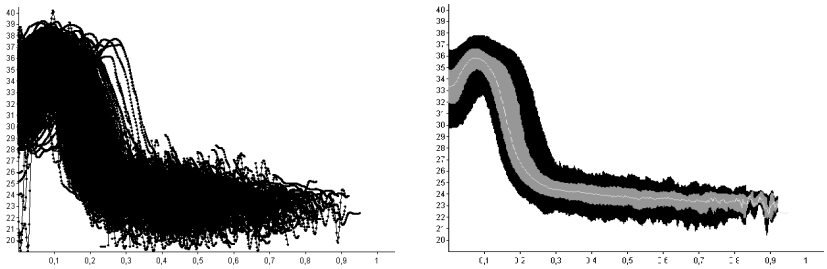


Kuva 1: a) Sadan *a*- ja *i*-alkuisen sanan F0-käyrien persentiilijakauma (90 %), b) Viidenkymmenen yhdyssanan jakauma (a ja b: puhuja TP) ja c) sadan lauseen jakauma (puhuja ToPa).

ei mediaanin loppukorkeutta voi täsmällisesti mitata, mutta ilmeisen osuva on arvio, että kaikissa kolmessa aineistossa mediaanin huipun ja lopun erotus on noin oktaavi tai hieman yli (a = 12,5, b = 14,5 ja c = 11,5 psa).

Mikäli F0-käyrän alku edustaa muita funktioita, korkeus vaihtelee suuresti. Siihen vaikuttavat esimerkiksi se, onko ensimmäinen sana substantiivi vai verbi, sisältösana vai funktiosana, subjekti vai objekti, kontrastiivinen aksentti sanalla vai ei, esiintyykö kontrastiivinen aksentti myöhemmin ilmauksessa, kuinka monta sanaa lauseessa on, onko ensimmäinen vokaali suljetussa vai avotavussa (Iivonen 1984).

Kuvassa 2 esitetään puhujan TP 750 balansoidun sana-aineiston perussävelkontuurien jakauma. Balansointi tarkoittaa tässä sitä, että aineistoon valittiin tasaisesti eri konsonanteilla ja vokaaleilla alkavia sanoja. Vasemmalla kaikkien sanojen F0-käyrät on asetettu päällekkäin. Analyysi osoitti kontuurien aloituskorkeuksissa huomattavaa vaihtelua: noin 10 puolisävelaskelta (psa) alussa ja noin 9 psa huipussa. Alun vaihteluväli noin 28–38 psa vastaa Hz-asteikolla väliä noin 85–150 Hz eli 65 Hz. Lisäksi huippujen ajoituksessa on suuria poikkeamia. Käyrien loppuosan leveyskin osoittaa myös huomattavia korkeuseroja. Voidaan osoittaa, että vaihtelun laajuus ei ole satunnaista ja pelkästään puhujan vapausasteista riippuvaa, vaan mikro- ja makroprosodiset tekijät selittävät sitä. Esimerkiksi sanoissa (N = 144) jotka alkavat /ta/:lla oli alku keskimäärin korkeammalla kuin sanoissa (N = 200), jotka alkavat /a/:lla. Sanoissa, jotka alkavat /ta/:lla oli nousu pienempi ja F0:n huippu aikaisempi kuin sanoissa, jotka alkavat /a/:lla. Vokaalien ominaiskorkeuden vaikutus ilmeni puolestaan seuraavasti. Sanoissa (N = 250), jotka alkavat /i/:lla oli alku keskimäärin hieman korkeammalla kuin sanoissa, jotka alkavat /a/:lla. Kuvassa 2 oikealla esitetään samasta aineistosta 90 %:n persentiilijakauma, mikä poistaa kaikkein poikkeavimmat tapaukset. Siinä alun jakauma on kaventunut 6,5 psa:een ja huippu 5,2 psa:een.



Kuva 2: Balansoidun, 750 sanaa käsittävän aineiston perussävelkontuurien jakauma. Vasemmalla kaikki F0-kontuurit päällekkäin (aika/psa). Oikealla 90 %:n persentiilijakauma 5, 25, 50 (= mediaani), 75, 95 %. Puhuja TP.

Keskeisenä tutkimuskohteenamme on, miten makrokontuurin muoto varioi mikroprosodian tasolla, kun kontuuri alkaa tauon jälkeen samanlaisten tekijöiden valli-

tessa sanan alkaessa konsonanteilla *h*, *j*, *k*, *l*, *m*, *n*, *p*, *r*, *s*, *t* tai *v* tai vokaaleilla *a* tai *i*. Miten mainitut äänteet vaikuttavat F0-käyrän alkukorkeuden vaihteluun sekä käyrän F0-huipun korkeuteen ja ajoitukseen?

4 Virhelähteiden eliminointi

F0-käyrien mittauksen luotettavuus. F0-kontuurien alkukorkeuden laskenta Praatilla edellyttää, että ensimmäistä äänihuuliperiodia edeltää vähintään 30 ms:n segmentti. Äänitiedostojen tallennuksessa tämä otettiin huomioon. F0-kontuurien luotettavuus voidaan tarkistaa asettamalla spektrogrammi Praatin Edit-ikkunassa kontuurin taustaksi siten, että pitch-kontuurille ja spektrogrammille (aikaikkuna 30 ms) asetetaan näytössä 500 Hz:n yläraja. Tällöin pitch-kontuurin ja alimman osasävelen tulee asettua päällekkäin.

Sanan pituuden vaikutus. Kuvasta 1 ilmenee, että yhdyssanoissa F0-alku asettuu korkeammalle kuin yksinkertaisissa sanoissa. Sanan pituudella tai jollakin muulla äännekestoja säätelevällä yksiköllä on siis vaikutusta siihen, mille korkeudelle F0:n alku asettuu. Näin taustalla olisikin siis jokin puhujan tahallinen, ekstrinsinen, ei mikroprosodinen tekijä. Aineiston alaryhmittelyyn on siis kiinnitettävä erityistä huomiota.

Vertailuryhmien rajaaminen. Koska yksittäisäännöksen F0-tasoon voi siis samanaikaisesti vaikuttaa useita eri tekijöitä, on näiden tekijöiden määrää vähennettävä. Vaikuttaako se, että ensimmäistä vokaalia *seuraava* konsonantti vaihtelee? Tutkittiin 3-tavuisten sanojen aineistoa, jossa vokaalia seuraava konsonantti on *l* tai *t* (/kal.CV$\cdots$/, esim. *kalkita* tai /kat.CV$\cdots$/, esim. *katsomo*). Ensi tavu /kaC/ on siis kaksimorainen. Aineistoa löytyi 11 sanaa ryhmää kohti. Mediaanikäyrä saavutti huippunsa (60 versus 65 ms) kummassakin ryhmässä ensimmäisen moran aikana toteuttaen näin Suomen, Toivasen ja Ylitalon (2003, 2006) havaitseman säännönmukaisuuden. Mediaanin F0-huippu oli 1,2 psa korkeampi /kat.CV$\cdots$/-ryhmässä (37,2 psa) kuin /kal.CV$\cdots$/-ryhmässä (36 psa). Aloituskorkeudessa ilmeni samansuuntainen yhden psa:n ero (/kat.CV$\cdots$/ = 35,8 ja /kal.CV$\cdots$/ = 34,8 psa). Aineiston määrä oli kuitenkin pieni. Evidenssi on siis, että rakenteessa /kat.CV$\cdots$/ toimii mikroprosodinen tekijä, jonka mukaan vokaalinjälkeisen soinnittoman *t*-klusiilin edellä vokaalin F0-alku ja huippu ovat noin 1 psa:n verran korkeampia verrattuna rakenteeseen /kal$\cdots$/, jossa vokaalia seuraa soinnillinen *l*-konsonantti. Huipun ajoitusero 5 ms sen sijaan lienee merkityksetön. Jäljempänä esitellään systemaattisemmin konsonanttien aiheuttamaa mikroprosodiaa.

Vertailuryhmien koon vaikutus. Yksittäiskontuurien erot ovat melko suuria, kuten edelliset esimerkit osoittavat (ks. etenkin kuva 2). Mikroprosodisten tekijöiden perusteella valitut ryhmät osoittavat, että laajaa variaatiota voidaan selittää ja rajoittaa, jolloin kontuurien mikroprosodinen säännöllisyys paljastuu etenkin mediaanikäyrissä. Puhujan vapausasteita jää silti jäljelle. Taulukossa 2 (alempana) esitetään 927 *h*-alkuisen sanan mittaustietoja, joiden mukaan ryhmän arvot eivät enää muutu paljon, vaikka ryhmää rajoitetaan rakenne-eroja täsmentämällä. Tilastollisissa analyyseissä 30 yksittäistapauksen määrää pidetään alarajana, jotta ryhmien välisiä eroja alkaa paljastua luotettavasti. Jäljempänä tarkastellaan ryhmiä, joissa valitun ryhmäkriteerin täyttäviä tapauksia on 100.

5 Sananalkuisen konsonantin ja vokaalin vaikutus F0-kontuuriin

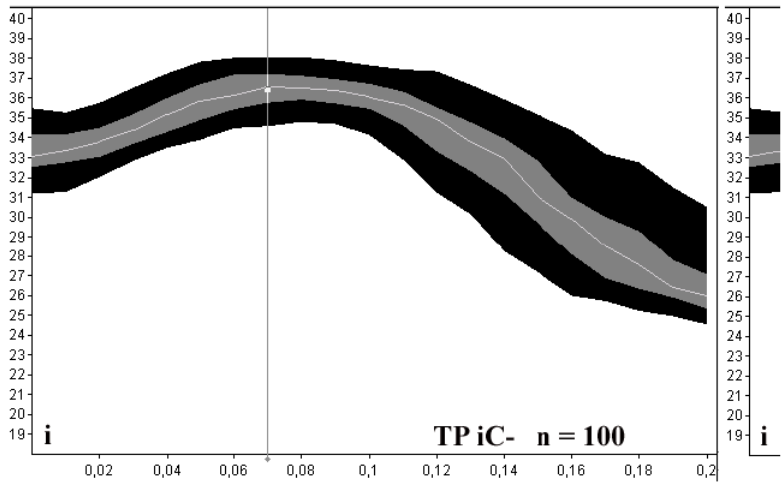
Seuraavat analyysit kohdistuvat sananalkuisten konsonanttien (*h*, *j*, *k*, *l*, *m*, *n*, *p*, *r*, *s*, *t* ja *v*) ja vokaalien (*a* ja *i*) vaikutuksien osoittamiseen. Kussakin ryhmässä kontuureja oli 100 ja ryhmiin otettiin etupäässä 2- ja 3-tavuisia sanoja, mutta ryhmän kokoon saamiseksi tarvittiin mukaan joitakin 4-tavuisiakin. Yhdyssanoja ei valittu.

Kuva 3 esittää tällaisen ryhmän kontuurijakauman *i*-vokaalin osalta. Lisäksi mediaanikäyrän F0-huipun ensimmäinen ajallinen aikapiste on merkitty. Näyte kustakin yksittäiskontuurista on rajattu 200 ms:iin. Taajuus esitetään puolisävelasteikon mukaan (18–40 psa) kantaluvun ollessa 16,35 Hz. Lisäksi saman jakaumakuvan alusta on otettu 10 ms:n näytepala (kuvassa 3 oikealla). Alun persentiilijakauman ominaisuudet esitetään taulukossa 2.

Kuva 4 esittää vertailukohtina sadan /iC/-, /kVC/- ja /vVC/-alkuisen sanan jakaumat ja F0-huippujen sijainnit. Taulukossa 2 esitetään vastaavasti jakauma-alkujen tiedot ja F0-huippujen ajoitus. Seuraavat päätelmät voidaan tehdä. Mediaanikäyrien alku on korkein rakenteessa /kVC$\cdots$/ (35,2 psa), seuraavaksi korkein rakenteessa /iC$\cdots$/ (33 psa) ja matalin rakenteessa /vVC$\cdots$/ (31,4 psa). Korkeimman ja matalimman mediaanialun erotus on 3,8 psa. Arvot vastaavat Hz-asteikolla lukemia 126, 110 ja 100 Hz, ja korkeimman ja matalimman mediaanialun erotus on siis 26 Hz. Merkille pantavaa on myös, että jakaumien laajuudessa on eroja: /vVC/-alkuisissa on laajin vaihtelu.

Vertailun vuoksi mukana on myös 927 /hV(V)C/-alkuisen sanan jakauman mittausarvot. Vaikka aineiston määrä on yli 9-kertainen ja rakenteessa on sallittu ensi tavussa yksinkertainen vokaali, pitkä vokaali tai diftongi, ei jakauman laajuus ylitä juurikaan muiden rakenteiden arvoja ja on lähinnä rakenteen /kVC$\cdots$/ kaltainen.

F0-huipun ajoituksessa on *v*-alkuisten (ja yleensä soinnillisella konsonantilla alkavien) sanojen myöhempi huippu selitettävissä sillä, että alun konsonantilla on oma kestonsa ja F0-huippu sijoittuu sen jälkeiseen vokaaliseen segmenttiin.



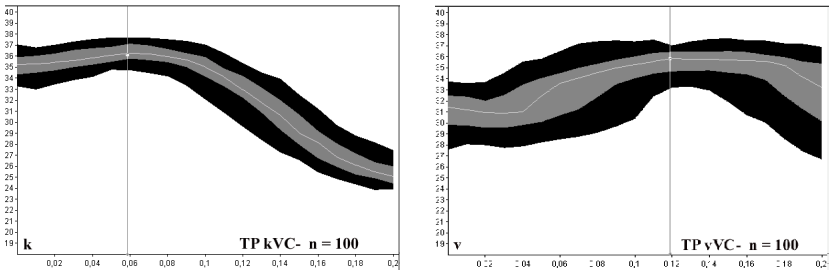
Kuva 3: Sadan /iC/-alkuisen sanan F0-kontuurien persentiilijakauma (90 %) puolisävelasteikolla 200 ms:iin asti. Oikealla lisäksi samasta jakauman alusta 10 ms:n näytepala varustettuna omalla asteikolla. Kuvan vasempaan nurkkaan on merkitty sanaryhmän sanojen alkuäänne. F0-huippukohta on myös merkitty mediaanikäyrään.

Kaikkien tutkittujen rakenteiden jakaumien (N = 100) alkujen perusteella on laadittu kuvat 5 (miespuhuja TP) ja 6 (naispuhuja MW). Näissä ryhmissä alkukonsonanttia seurasi balansoidusti kaikki suomen fonologisesti lyhyet vokaalit. Mukana on myös *a*- ja *i*-alkuisten sanojen sata sanaa käsittävät ryhmät.

Taulukko 2: 100 tapauksen persentiilijakaumien ominaisuuksia rakenteissa /iC⋯/, /kVC⋯/ ja /vVC⋯/: jakauman maksimi (= 95 %:n raja), mediaani, minimi (5 %), maksimin ja minimin erotus sekä mediaanin ajallisen huippukorkeuden paikka (ms). Vertailun vuoksi mukana on /hV(V)C/-alkuisten sanojen (N = 927) jakauman mittausarvot.

	/iC⋯/	/kVC⋯/	/vVC⋯/	/hV(V)C⋯/
maksimi (psa)	35,5	37	33,7	37,3
mediaani (psa)	33	35,2	31,4	35
minimi (psa)	31,3	33,3	27,6	32,8
erotus max-min (psa)	3,7	3,7	6,1	4,5
F0-huippu (ms)	65	57	138	55

Kuvan 5 mukaan miespuhujalla TP alkukorkeudet madaltuvat järjestyksessä *s*, *t*,



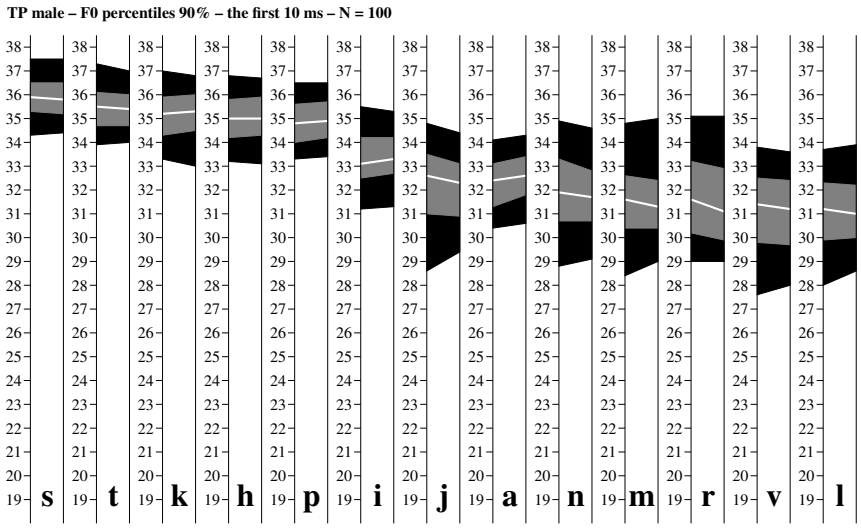
Kuva 4: Vasemmalla: Sadan /kVC/-alkuisen sanan F0-kontuurien persentiilijakauma (90 %). Puhuja TP. Oikealla: Sadan /vVC/-alkuisen sanan jakauma. F0-huippukohta on myös merkitty mediaanikäyriin. Evidenssi: /kVC/-alkuisissa sanoissa alku on korkea ja nouseva, kun taas /vVC/-alkuisissa alku on *v*:n aikana matala ja madaltuva. Jälkimmäisissä jakauman 90 %:n jakauma on laajempi ja kaventuu vain huipun kohdalla.

k, *h*, *p*, *i*, *j*, *a*, *n*, *m*, *r*, *v* ja *l*. Vaihteluväli on 4,7 psa. Alut jakaantuvat kolmeen ryhmään 1) obstruentit (mukana soinniton *h*; kaikilla tasainen tai nouseva F0), 2) vokaalit *i* ja *a* (joilla on nouseva F0) sekä 3) resonantit (approksimantit, nasaalit ja likvidat, joilla on laskeva F0). Obstruentit ja *h* muodostavat korkeamman ryhmän ja vokaalit ja resonanttikonsonantit matalan ryhmän. Vokaalien *a* ja *i* alkukorkeuksien ero on 0,6 psa.

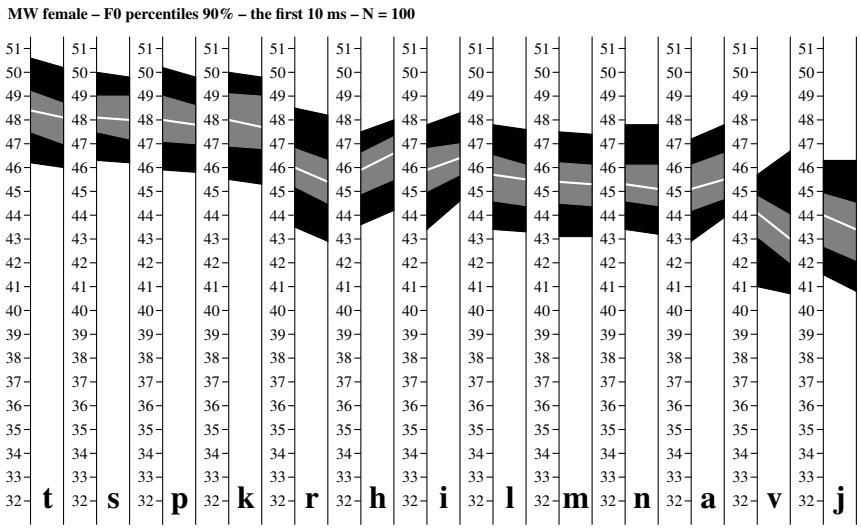
Kuvan 6 mukaan naispuhujalla MW alkukorkeudet madaltuvat hieman eri järjestyksessä *t*, *s*, *p*, *k*, *r*, *h*, *i*, *l*, *m*, *n*, *a*, *v* ja *j*. Vaihteluväli on 4,6 psa. Muita eroja ovat seuraavat. Obstruenttien F0-alut ovat laskevia. Konsonantti *h* kuuluu vokaalien *i* ja *a* ohella nousevien joukkoon. Psa-asteikolla mitattuna vaihtelun laajuus on ryhmissä suurempi kuin TP:llä. Pääryhmittelyn mukaan obstruenttien lähtökorkeus on korkeampi kuin muilla, joihin kuuluvat *h*, vokaalit ja resonanttikonsonantit. Vokaalien *a* ja *i* alkukorkeuksien ero on 0,7 psa. Inherenttiä vaikutusta on siis kummallakin puhujalla.

6 Mikroprosodian vaikutus sananalkuisen huipun korkeuteen ja ajoitukseen

Kustakin 100 sanan ryhmästä mitattiin F0-mediaanin huipun etäisyys mediaanikäyrän alusta. Kaikissa tapauksissa huippu sattui hyvin 200 ms:n sisäpuolelle ja oli maksimissaan 138 ms (TP *r*-äänne). Taulukosta 3 ilmenee, että huipun vaihteluväli on TP:llä 48–138 ms ja MW:llä 14–105 ms. Pidemmät alut (keskiarvot 121,0 ja 87,7 ms) selittyvät sillä, että F0-käyrä kattaa myös soinnillisen resonanttikonsonantin. Alku on lyhyt, jos vokaalia edeltää soinniton konsonantti *h*, *k*, *p*, *s*, *t* (ka = 61,8 ja 34,6 ms) tai jos sana alkaa vokaalilla (ka = 80 ja 48,5 ms). MW on tavoittanut F0-huipun suh-



Kuva 5: Persentiilijakaumien alun ensimmäiset 10 ms konsonantti- ja vokaalialkuisissa sanoissa. Miespuhuja TP.



Kuva 6: Persentiilijakaumien alun ensimmäiset 10 ms konsonantti- ja vokaalialkuisissa sanoissa. Naispuhuja MW.

teellisesti aikaisemmin, mutta ajallinen korrelaatio on huomattava: Pearsonin korrelaatiokerroin TP ja MW $r = 0{,}78$. Ilman *s*-äännettä korrelaatio on 0,91.

Taulukko 3: Persentiilijakaumien mediaanien alkuhuippujen maksimikorkeuden saavuttamisen ajakohdat (ms) puhujilla TP ja MW sadan sanan konsonantti- ja vokaaliryhmissä.

	h	*j*	*k*	*l*	*m*	*n*	*p*	*r*	*s*	*t*	*v*	*a*	*i*
TP	77	117	59	123	115	113	48	138	57	68	120	88	72
MW	36	78	18	93	87	85	27	78	78	14	105	50	47

Kustakin 100 sanan ryhmästä mitattiin myös F0-huipun korkeus mediaanikäyrästä (taulukko 4).

Taulukko 4: Persentiilijakaumien mediaanien alkuhuippujen F0:n maksimikorkeudet (psa) puhujilla TP ja MW sadan sanan konsonantti- ja vokaaliryhmissä.

	h	*j*	*k*	*l*	*m*	*n*	*p*
TP	36,2	36,1	36,1	36,2	36,6	36,1	36,1
MW	47,2	46,8	47,9	46,9	47,2	46,8	47,8

	r	*s*	*t*	*v*	*a*	*i*
TP	35,9	36,2	36,2	36,0	36,0	36,7
MW	47,2	47,1	48,0	46,6	46,8	47,4

Tässä havaitaan se erikoisuus, että mediaanikäyrän huippukorkeus vaihtelee hyvin vähän: TP:llä vaihteluväli on $35{,}9–36{,}7 = 0{,}8$ psa ja MW:llä $46{,}6–48{,}0 = 1{,}4$ psa. Tulkinta on, että makroprosodian mukaista on, että huipun korkeus on suhteellisen vakio ja sanan alkuäänteen vaikutuksesta riippumaton. Huipun ajoitus vaihtelee siis paljonkin, mutta sen korkeus näyttää olevan mediaanien mukaan hyvin stabiili. Huomattavaa on, että puhujalla TP F0:n huippukorkeus *a*:lla 36 psa on 0,7 psa alempi kuin *i*:llä (36,7 psa). Vastaava ero MW:llä on 0,6 psa (46,8 ja 47,4 psa). Vokaaleilla inherentti korkeus on siten kummallakin puhujalla samansuuruinen sekä F0:n alku- että huippukorkeudessa.

7 Tulosten pohdinta

Kontekstista irrallaan äännetyn, tauonjälkeisen sanan F0-kontuurilla on *makro*prosodinen muoto, jonka hallitsevia piirteitä ovat alun korkea huippu ja sitä seuraava lasku puhujan äänialan alimpiin taajuuksiin (kuva 1). Kontuurin foneettinen realisaatio vaihtelee sanan fonologisen rakenteen ja sanan painotuskaavan (mahdollisten sivupainojen) mukaan. Myös F0-kontuurin mikroprosodia aiheuttaa kontuuriin vaihteluita. Suuren balansoidun aineiston yksittäisten F0-kontuurien jakauma osoitti miespuhujan aloituskorkeuksissa ja huippukorkeuksissa suurta vaihtelua: noin 10 puolisävelaskelta (psa) alussa (vastaa noin 65 Hz:n vaihteluväliä) ja noin 9 psa huipussa (kuva 2). Puhujan vapausasteet selittävät osaltaan tätä vaihteluita. Aineiston ryhmittely sanan ensimmäisen äänteen mukaisesti osoitti kuitenkin, että ensimmäisellä äänteellä on osaltaan selvä ja säännöllinen vaikutus vaihteluun.

Mies- ja naispuhujalla konsonanttien mikroprosodinen vaikutus oli joistakin eroista huolimatta hämmästyttävän samansuuntainen. Se ilmeni F0-kontuurissa etenkin siten, että korkeampitaajuinen *F0-aloituskorkeus* esiintyy obstruenttien jälkeen verrattuna resonantteihin (kuvat 5, 6). Mies- ja naispuhujalle yhteistä oli myös, että resonanttikonsonanttien jälkeinen F0-kontuuri oli paitsi matalampi myös suunnaltaan laskeva. Vokaalit *a* ja *i* olivat molemmilla nousevia. Seuraavia eroja ilmeni. Miespuhujalla TP alkukorkeudet madaltuvat järjestyksessä *s*, *t*, *k*, *h*, *p*, *i*, *j*, *a*, *n*, *m*, *r*, *v* ja *l* ja naispuhuja MW:llä järjestyksessä *t*, *s*, *p*, *k*, *r*, *h*, *i*, *l*, *m*, *n*, *a*, *v* ja *j*. Miespuhujalla obstruenttien F0 oli nouseva, naispuhujalla laskeva. Konsonantti *h*:n vaikutus ilmeni eri tavoin: TP:llä se kuului obstruenttien ryhmään korkeuden ja F0:n suunnan mukaan, MW:llä vokaalien ryhmään. Vokaaleilla *a* ja *i* inherentti korkeus osoittautui kummallakin puhujalla samansuuruiseksi (0,6–0,7 psa) sekä F0:n alku- että huippukorkeudessa.

F0-huipun ajoitus ilmeni olevan riippuvainen siitä, alkaako sana klusiililla tai vokaalilla (suhteellisen varhainen huippu) vai soinnillisella resonanttikonsonantilla tai *s*:llä ja *h*:lla (myöhemmin ilmenevä huippu). Tämä johtuu jälkimmäisten osalta siitä, että niillä on oma kokonainen segmentti, jota tauonjälkeisellä klusiililla ei ole. Sen sijaan *F0-huipun korkeuteen* edellä mainitut tekijät eivät mediaanikäyrästä mitattuna näytä juurikaan vaikuttavan, mikä voidaan tulkita makroprosodian ominaisuudeksi. Yksittäiskontuureissa ilmenee kuitenkin melkoista vaihtelua. Kaikkia mahdollisia vaikutustekijöitä ei tutkittu muun muassa siksi, että suurikaan sanamateriaali ei tarjonnut riittävästi sanoja riippuvuuksien todentamiseksi. Alustavat kokeet näyttivät, että fonotaksin (mm. tavurakenteen) mukaisella ryhmittelyllä on ehkä mahdollista selittää lisää vaihtelua.

Huomattakoon, että kuvan 2 suurta vaihtelua voidaan tutkia edelleen muiden tekijöiden avulla. Tuloksilla on merkitystä yleensä F0:aa koskevassa tutkimuksessa. Mittauksissa on aiheellista kiinnittää huomiota aineiston äännerakenteen vaikutuksiin. Kuvatuilla ilmiöillä saattaa olla sekundaarinen funktionaalinen merkitys prosodian

ulkopuolella: äänteiden ja foneemien tunnistamisessa.

Viitteet

BOERSMA, Paul & WEENINK, David 2008: Praat—doing phonetics by computer. <http://www.fon.hum.uva.nl/praat/>.

DI CRISTO, Albert 1978 [1985]: *De la Microprosodie à l'Intonosyntaxe*. Publications de l'Université de Provence.

DI CRISTO, Albert 1982: *Prolégomènes à l'étude de l'intonation: micromélodie*, osa 2 sarjasta *Travaux de l'Institut de phonétique d'Aix-en-Provence, Collection "Sons et parole"*. Paris: Centre National de la Recherche Scientifique.

IIVONEN, Antti 1984: On explaining the initial fundamental frequency in Finnish utterances. – C.-C. Elert, I. Johansson & E. Strangert (toim.), *Nordic Prosody III: Papers from a Symposium*, Acta Universitatis Umensis, Umeå Studies in the Humanities 59. Stockholm: Almqvist and Wiksell. 107–119.

IIVONEN, Antti, SEPPÄNEN, Tapio, NOPONEN, Kai & TOIVANEN, Juhani 2004: Puhujan temporaalisen äänialan visualisointisovellus. – *Puhe ja kieli*, **24**(1):31–39.

LEHISTE, Ilse 1976: Suprasegmental features of speech. – N. Lass (toim.), *Contemporary Issues in Experimental Phonetics*. New York/San Francisco/London: Academic Press. 225–239.

MEISTER, Einar & WERNER, Stefan 2006: Intrinsic microprosodic variations in Estonian and Finnish: Acoustic analysis. – Reijo Aulanko, Leena Wahlberg & Martti Vainio (toim.), *Fonetiikan päivät 2006 / The Phonetics Symposuim 2006*, Helsingin yliopiston puhetieteiden laitoksen julkaisuja 53. 103–112.

SUOMI, Kari, TOIVANEN, Juhani & YLITALO, Riikka 2003: Durational and tonal correlates of accent in Finnish. – *Journal of Phonetics*, **31**(1):113–138.

SUOMI, Kari, TOIVANEN, Juhani & YLITALO, Riikka 2006: *Fonetiikan ja suomen äänneopin perusteet*. Helsinki: Gaudeamus.

TRASK, R.L. 1996: *A Dictionary of Phonetics and Phonology*. London/New York: Routledge.

WANG, W. S.-Y. & FILLMORE, C. J. 1961: Intrinsic cues and consonant perception. – *Journal of Speech and Hearing Research*, **4**:130–136.

Michael O'Dell & Tommi Nieminen (eds.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 97–109.

Speech rate as an indicator of fluency in the Russian of Finnish learners

Riikka Ullakonoja
University of Jyväskylä

Abstract

This study focuses on the speech rate development of 12 Finnish university students of Russian during their 3.5-month-study abroad experience. Speech and articulation rates are measured in phonetic words per second and syllables per second in the Russian read-aloud speech of the subjects. This is done at three recordings: prior to, during and following their stay in Russia. The results are compared to their read-aloud Finnish speech. The students are also compared depending on the residence (host-family vs. dormitories) in Russia. The study shows that speech and articulation rates correlate with the evaluated fluency of the speech samples. It was found that speech rate is a better indicator of fluency than articulation rate in non-native read-aloud speech. The results also show that articulation rate in mother tongue (Finnish) and foreign language (Russian) correlate with each other more than speech rate.

Keywords: speech rate, fluency, Finnish (L1), Russian (L2)

1 Introduction

When asking foreign language learners what aspects they consider important in learning the new language, their answers might include a desire to become fluent in that language. Also in the words of their teacher, in the syllabus and in also the Common European framework of reference for languages (Council for Cultural Co-operation. Education Committee, Modern Languages Division, Strasbourg and Council of Europe 2001) the term fluency and its derivations occur frequently. However, when teaching oral skills, it is perhaps not the fluent features of speech that are in the focus of attention, but instead the grammatical and lexical features or the pronunciation of segments. The purpose of the study is to follow the fluency development of 12 Finnish students of Russian during their 3.5-month-stay in Russia by studying their speech and articulation rates and comparing them to fluency evaluations of teachers.

Fluency can be defined in a number of ways, e.g. by studying pausing (pause frequency, duration and placement), hesitations or tempo (se e.g. Cucchiarini *et al.* 2002,

Lauranto 2005, for a review). In this study speech rate is regarded as an important factor of fluency. Cucchiarini *et al.* (2002) have shown that speech rate and pause frequency are the most important factors in read-aloud speech fluency perception. Also Riggenbach (1991) concluded that the central elements of foreign language (L2) fluency are pausing, speech rate and repairs. Moreover, several researchers (Riggenbach 1991, Freed 1995, Towell *et al.* 1996) have found that as L2 fluency increases, the speech rate increases also. My previous study (Ullakonoja 2008) focused on pausing and its relationship to foreign language fluency. In this paper, the same data is studied, but speech and articulation rates are regarded as acoustic correlates of fluency.

The speech rate (tempo) indicates the total time of a speaker uttering his speech, including pauses whereas the term articulation rate is commonly used to refer to the speech rate without pauses. In this study speech rate refers to reading rate. There are multiple factors affecting the habitual speech rate of individual speakers, and speakers can also vary their speaking rate in different situations (see Trouvain 2003 for a review). In this study the speaking context and content are the same for all speakers at all recording sessions. The speech and articulation rates of a L2 learner are often shown to be slower than these of a native speaker (e.g. Riggenbach 1991, Cenoz 2000, Paananen-Porkka 2007). In addition, learners possibly transfer the prosodic characteristics (e.g. stress) of their mother tongue to the language they are learning:

> When the Finn transfers the habit of pronouncing all of the syllables of each word unreduced and manifesting word boundaries with phonetical juncture segments (instead of linking) the rate of his speech is inevitably slower (Lehtonen 1981, p. 331).

A foreign language learner often has the impression that native speakers of the language speak very fast (Abercrombie 1967, p. 96). Also, when native speakers are listening to L2 speech, they would often prefer about 10 % faster speech rate than what the learner is producing (Munro & Derwing 2001, p. 464).

It has been found in several studies (Simoes 1996, Freed *et al.* 2004, Lafford 2004, Trofimovich & Baker 2006) that a good way to improve fluency in L2 is to spend some time in the country where L2 is spoken. For example Segalowitz & Freed (2004) established that the students who studied abroad improved their fluency more (on several measures including speech rate) than the students who stayed at home. Trofimovich & Baker (2006) found that L2 learners could not achieve a native speech rate no matter how long they stayed in the country of the L2 language. On the contrary, a study by Freed *et al.* (2004) suggests that the study abroad did not result in better fluency than an "intensive domestic immersion" context. In their study it was in fact the immersion context that turned out to be the most effective in fluency learning. To summarize, all the studies show the positive influence of L2 context to the fluency development.

There have been a few studies (e.g. Lehtonen 1979, Iivonen *et al.* 1995, Moore & Korpijaakko-Huuhka 1996, Suomi 2007) about speech rate in native Finnish speech. In Russian, pausing and its influence on prosodic phrasing and speech rate have been researched also in spontaneous speech (e.g. Shtern 1988, Volskaya forthcoming). To my knowledge the current paper is the first study investigating non-native speech rate in Russian and comparing it to the speakers' native language, Finnish, and contrasting different stages of learning. The aim of this study was to find out, firstly, whether speakers who are considered fluent speak/read aloud faster than disfluent speakers (both in terms of speech and articulation rates). In other words, speakers with faster speech or/and articulation rates are evaluated more fluent than slower speakers. Secondly, the speech and articulation rates in Finnish (mother tongue, L1) were compared to speech and articulation rate in Russian (L2) to find any similarities between the two.

2 Material

12 native Finnish students of Russian read two Russian and one Finnish dialogue in pairs. The reading was recorded in different stages of their university studies: prior to, in the middle of and following their stay in Russia. Only the longest turn of the Russian dialogues and two turns of the Finnish dialogue were analyzed of each student. The Russian material, hence, includes the reading of the same text three times (c. 11 minutes in total), whereas the Finnish material is from the first recording session (c. 3 minutes in total). The students are undergraduate major students of Russian who have studied Russian for 1–10 years prior to university studies. At the beginning of their 2nd year of university studies they participated in a 3.5-month-study-abroad-program. Half of the students (subjects Fi3, Fi4, Fi5, Fi7, Fi9 and Fi10) resided in the dormitories for foreign students during their stay in Russia with the remaining (subjects Fi1, Fi2, Fi6, Fi8, Fi11 and Fi12) living with a host family. The two groups were compared for speech and articulation rates development where applicable.

3 Methods

For evaluating the perceptual fluency of the speech samples, 30 Russian as a foreign language teachers in Finland were asked to determine the fluency of each sample on 1–5 scale (1 = not fluent, 5 = very fluent). Teachers listened to the samples in a random order without knowing that multiple samples of the same speaker were included. The reliability of the fluency ratings was good (Cronbach's alpha = 0.92). The procedure of the fluency evaluation task is more thoroughly reported in a parallel study (Ullakonoja 2008).

Segmentation and acoustic analysis of the samples were completed in Praat (Boersma & Weenink 2008). The segmentation consisted of annotation of phonetic words and syllables. The term 'phonetic word' comes from the Russian research tradition (e.g. Avanesov 1956, p. 61), and usually corresponds to a lexical word, but also to some two word combinations, where e.g. a preposition is pronounced together with the main word and where there is only one lexical stress. For example, in this data the preposition and pronoun *k nam* [knɑm] ('to us') are treated as a phonetic word. The term prosodic word has sometimes been used to describe the same phenomena in Finnish (see e.g. Aho & Yli-Luukko 2005). In Finnish, I decided that lexical words always correspond phonetic words in the annotation. The syllables were determined according to auditory analysis, hence the syllable means a realized syllable. Syllable nuclei were determined and proportioned with time (counting syllable nuclei instead of syllables has been used e.g. by Simoes 1996). In Russian the number of syllables corresponded the number of vowels in the utterance. In Finnish, single vowels were treated similarly as in Russian, as a syllable nucleus. Vowels in the vowel combinations in Finnish were mostly pronounced very closely together and consequently, they were also regarded as one syllable. Sometimes the syllabification in Finnish did not respect the traditional (or textual) syllabification, if e.g. the word *teorioita* ('theories (partitive case)') was pronounced [teoriotɑ], it was considered trisyllabic: teo-rio-ta (speaker Fi7). Similarly also the phrase *mä en oo* ('I'm not') was pronounced mostly as [mæeno], [mæeo] or [menːoː] and in all cases it only had two syllables. Syllable omission was quite frequent in Finnish, e.g. *no en* [non] ('well no', Fi7), *huomenna* [huomen] ('tomorrow', Fi7).

The duration of phonetic words was measured with a script in Praat. Phonetic words per second and syllables per second were used for measuring speech and articulation rates (i.e. speech rate without pause time). Both measures were used in order to find out the differences, if any, between them and to make the language comparison as thorough as possible. Based on earlier results of a comparative study of English and Finnish speech rate (Lehtonen 1981), it was expected that the comparison of syllable-timed Finnish and stress-timed Russian would yield different results depending on the measure chosen. Syllables per second would show the influence of hesitation better, since hesitation is often not only one or two syllables but one phonetic word. Also syllables per second as a measure would show mispronunciations (e.g. omission of a syllable, see examples above) better than phonetic words per second. For example, following her stay in Russia speaker Fi12 has much hesitation in her speech and the segmentation gives quite different results depending on the measure chosen (Figure 1). The sentence has 6 phonetic words and 18 syllable nuclei, when the original text only had 5 phonetic words and 13 syllable nuclei.

Microsoft Excel was used for calculating speech rate and articulation rate as well as for the graphical representation of the results. SPSS was used to determine the correlations in the data and their statistical significances. The existence of linear

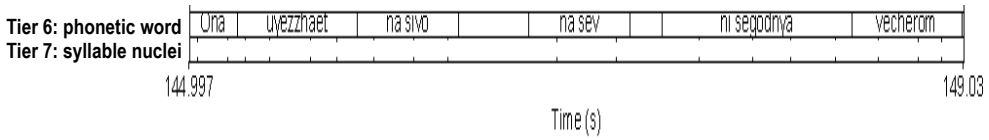


Figure 1: An example of the segmentation of the corpus *Ona uyezzhaet ne segodnya vecherom* 'She will leave not today at night' into phonetic words.

correlation was verified in scatterplot graphs. Paired samples t-test was used to find out the differences between different stages of learning. Speech and articulation rates of each sample were compared to its average fluency rating in order to determine the connection between speech and/or articulation rates and fluency. When comparing Finnish (L1) with Russian (L2) the individual variations in speech and articulation rates were minimized by comparing the within group ranking of each student in both languages (i.e. seeing whether the 2nd fastest student in Russian was also the 2nd fastest in Finnish etc.).

4 Results

In a previous study (Ullakonoja 2008), it was found that the majority of the speakers ($9/12$) developed in terms of their read-aloud fluency during the first half of their stay in Russia, and slightly over a half of them ($7/12$) further increased their perceived fluency during the rest of their stay. Furthermore, the study showed that pausing was closely related to read-aloud fluency in a foreign language.

4.1 Speech and articulation rates development during study abroad

In all subjects' speech the *speech rate* increased during the first half of their 3.5-month stay in Russia (0.2 phonetic words per second or 0.5 syllables per second on average) (Figures 2, 3; SR). Also, the majority of the subjects had a faster speech rate following their stay than before it (0.2 phonetic words per second or 0.5 syllables per second on average). Hence, the speech rate increases as the amount of experience increases. The development in speech rate is statistically significant ($p < 0.05$) when comparing before the stay results with middle of stay and before the stay results with after the stay in both phonetic words and syllables per second. However, the speech rate of some students ($4/12$ students when measuring phonetic words per second, $6/12$ students when measuring syllables per second) decreased slightly between the recordings done in the middle and after their stay. This decline is possibly due to the

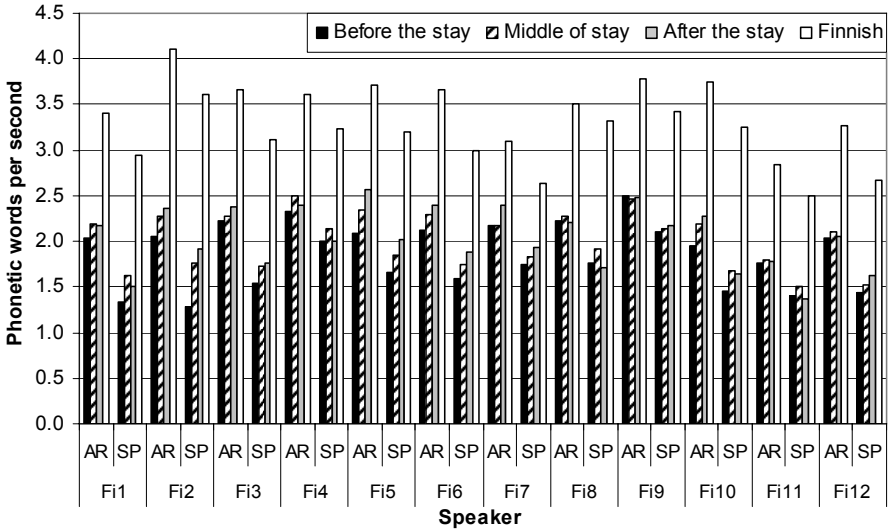


Figure 2: Articulation rate (AR) and speech rate (SR) in phonetic words per second in Finnish (L1) and in Russian (L2) at different stages of learning.

fact that their Russian reading was more "activated" while in the Russian speaking context than in the recording done following their stay.[1]

The measurement of *articulation rate* indicated a tendency similar to speech rate (Figures 2, 3; AR). Articulation rate also increased (0.1 phonetic words per second or 0.3 syllables per second on average) during the first half of the stay in the speech of most students ($9/12$). Between the 2nd and 3rd recordings, the articulation rate further increased for the majority ($7/12$) of the students (0.1 phonetic words per second on average), but also decreased or remained the same for some subjects. When comparing only the recordings done prior to and following the stay in Russia, it can be seen that the majority ($9/12$) of the students had a faster articulation rate after their stay than before it (0.2 phonetic words per second on average). The increase in articulation rate was statistically significant ($p < 0.05$) between before the stay and middle of stay results and between before the stay and after the stay results in both phonetic words and syllables per second.

The students were also divided into two groups according to their *residence* in Russia (host family vs. dormitories). The groups were neither balanced nor equal in their speech rate before their stay in Russia. When measuring phonetic words, students residing with a host family did not increase their speech rate on average

[1]The last recording was completed approximately one month after the students returned to Finland from Russia. It is possible that they had somewhat "forgotten" their Russian during that month, because some students had not used Russian at all after returning to Finland.

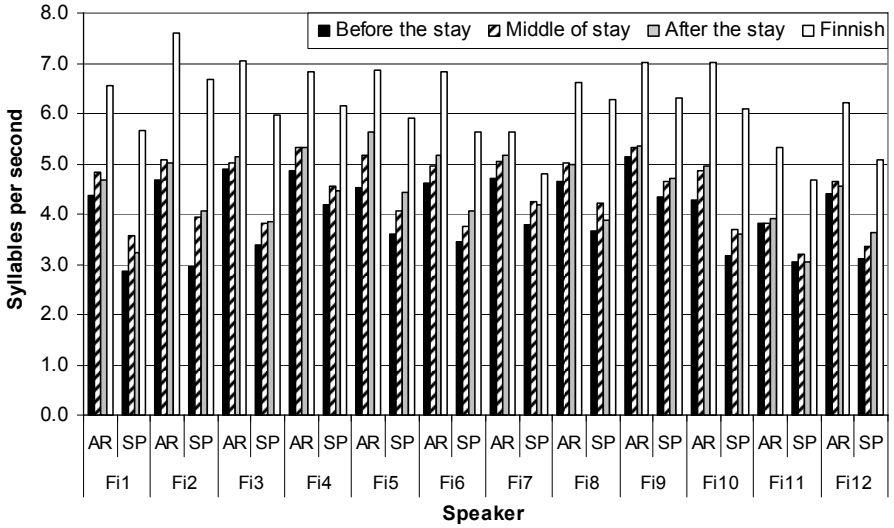


Figure 3: Articulation rate (AR) and speech rate (SR) in syllables per second in Finnish (L1) and in Russian (L2) at different stages of learning.

more than students living in the dormitories (Table 1). Contrary to what might have been expected, in syllables per second the dormitories group increased their speech rate more than the host-family group both during the first half and the whole length of their stay. In fact, the students residing with a host family had on average a slower speech rate at all recording sessions but as they also had a slower rate in Finnish, it seems that this is a random result. Similarly as in speech rate, the results of the articulation rate do not indicate that residence in the host family would make students speak faster during their stay in Russia. As a matter of fact, students residing in the dormitories increased their articulation rate more during the second half of their stay and during their entire stay in Russia (Table 1). The dormitories group might have had a better Russian competence and motivation already before the stay, which might have also been reflected in their speech rate.

4.2 Speech and articulation rates and fluency

What then is the relationship between speech or articulation rates and L2 fluency? The comparison of speech and articulation rates with perceived mean fluency rating flagged significant correlations (Table 2). The correlation was stronger between the speech rate and fluency rather than articulation rate and fluency. This indicates that pausing (hesitations and total pause time) also affects the fluency perception. The samples were also studied at the individual level where it was also noted that speech

Table 1: Mean speech and articulation rate of the students living with a host family and in the dormitories.

Residence	Before the stay	Middle of stay	After the stay	Finnish
	Speech rate: Phonetic words per second			
Host-family	1.47	1.68	1.67	3.01
Dormitories	1.75	1.89	1.92	3.14
	Speech rate: Syllables per second			
Host-family	3.18	3.68	3.65	5.66
Dormitories	3.74	4.17	4.20	5.87
	Articulation rate: Phonetic words per second			
Host-family	2.04	2.16	2.16	3.47
Dormitories	2.21	2.33	2.41	3.60
	Articulation rate: Syllables per second			
Host-family	4.42	4.72	4.72	6.52
Dormitories	4.73	5.12	5.27	6.73

rate correlates more reliably with the perceived fluency rating. For example, it was found that the least fluent (evaluated fluency = 1.3) sample was the speaker Fi2 prior to the stay. She was also the slowest of all speakers when measuring speech rate in phonetic words (Figure 2) and the second slowest when measuring speech rate in syllables (Figure 3). However, her articulation rate was not the slowest; in fact it was just below the average (Figures 2, 3). Correspondingly, the speaker who was evaluated the most fluent was Fi9 following their stay in Russia, who was also found to be the fastest of all speakers in speech rate and among the two fastest in articulation rate (Figures 2, 3).

4.3 Speech and articulation rates in Russian (L2) and Finnish (L1)

Next, speech and articulation rates in Finnish (L1) and Russian (L2) were compared. It was found that speech rate in Finnish correlates with the speech rate in Russian (Table 3). The correlation is however stronger between the articulation rate than speech rate in L1 and L2. This suggests that it is the amount of pause time that differs in L1 and L2, because the articulation rate indicates the speed of "uttering sounds," whereas speech rate includes pauses. As mentioned above, when comparing the in-

Table 2: Pearson correlations (R) between mean perceived fluency rating and speech and articulation rate.

	N cases	Correlation (R)	Significance (p)
Mean perceived fluency rating and articulation rate:			
Phonetic words/s	36	0.484	0.003
Syllables/s	36	0.416	0.012
Mean perceived fluency rating and speech rate:			
Phonetic words/s	36	0.722	< 0.001
Syllables/s	36	0.697	< 0.001

terspeaker performance, the speakers were ranked by speech rate and articulation rate from slowest to fastest in Finnish and at each recording session in Russian in order to be able to normalize the effect of differences in the structure of the two languages.

In Finnish (L1) the differences were small between syllables per second and phonetic words per second in articulation rate and speech rate. An individual speaker almost always received the same ranking position among the speakers in L1. In speech rate, $6/12$ speakers received a similar (maximum difference between ratings being 2) rating on average in Russian and in Finnish. In articulation rate $8/12$ speakers (when measuring phonetic words) and $7/12$ speakers (when measuring syllables) were ranked similarly in Finnish and Russian. This also indicates, that articulation and speech rates in L1 and L2 are related. Hence, speech rate seems to be a speaker-specific rather than a language-specific phenomenon.

5 Discussion and Conclusions

Overall, the majority of the students increased their L2 speech and articulation rates during their 3.5-month-stay in Russia statistically significantly as their perceived fluency increased also. This clearly shows that students seem to benefit from their stay in Russia so that they become faster and more fluent in Russian. Consistently with Towell *et al.* (1996, p. 103) the increased speech rate was found to be more significant than articulation rate in determining the L2 fluency of the speakers. When comparing the results with Lehtonen's (1978) study, it was found that the L1 Finnish reading rate was faster in this study when measuring phonetic words, but speech rates in syllables were similar in both studies.

The comparison of the students who stayed with a host family and students who resided in the dormitories was not very yielding as it turned out that the dormitories

Table 3: Pearson correlations for articulation rate (AR) and speech rate (SR) in phonetic words/s (pw) and syllables/s (syll) in Russian (L2) and Finnish (L1).

	Russian				Finnish		
	AR pw	AR syll	SR pw	SR syll	AR pw	AR syll	SR pw
Russian							
AR pw	1	0.966**	0.868**	0.861**	0.579**	0.556**	0.577**
AR syll	0.966**	1	0.811**	0.848**	0.586**	0.557**	0.574**
SR pw	0.868**	0.811**	1	0.985**	0.333*	0.282	0.424**
SR syll	0.861**	0.848**	0.985**	1	0.335*	0.279	0.423*
Finnish							
AR pw	0.579**	0.586**	0.333*	0.335*	1	0.985**	0.931**
AR syll	0.556**	0.557**	0.282	0.279	0.985**	1	0.913**
SR pw	0.577**	0.574**	0.424**	0.423*	0.931**	0.913**	1
SR syll	0.559**	0.552**	0.381*	0.376*	0.922**	0.929**	0.989**
N	36	36	36	36	36	36	36

** $p < 0.001$, * $p < 0.05$

group was already faster prior to the stay. Still, the results showed that in fact the students in the dormitories increased their speech and articulation rates more than the students living with host families. It can also be concluded that the speech and articulation rates in L1 are related to the speech and articulation rates in L2, consistently with Towell *et al.*'s study (1996, p. 96), where a strong correlation in L1 and L2 speech rate was established. Not surprisingly, the results also show that L1 is spoken faster than L2 (see e.g. Paananen-Porkka 2007).

The rhythmical features of speech were not taken into the account in this study. However, it is possible that the speech rate varies across the speech sample in the way as e.g. Deese (1980, pp. 74–76) has found that the majority of the faster sequences of speech occur either at sentence initial or terminal position. This study included recordings in Finnish only at the beginning and it was assumed that speech and articulation rates do not change significantly over time in one's L1 in the same reading task.

It has to be acknowledged that, naturally, there are other factors influencing speech and articulation rates and perceived fluency than the study abroad. Firstly, there is much individual variation in reading rate (even in L1). Also, in a reading task the subject might read very fast without comprehending everything being read (Lehtonen 1981, pp. 328–329; Perfetti 1985, p. 10) The student's motivation and interest are essential in L2 learning, therefore in this study also e.g. the motivation of the student towards Russian oral skills in general might have increased during the

stay in Russia. Furthermore, the findings concern only read-aloud speech in a laboratory setting and the analysis of spontaneous speech in a real communicative situation might have yielded different results.

It can be concluded that faster L2 speech (either in measures of speech or articulation rate) is perceived more fluent than slower L2 speech and that speech and articulation rates come closer to L1 speech and articulation rates as experience with L2 increases. Because native speakers of a language have been found to evaluate fast speech rate in non-native speech more positively than a slower speech rate (Munro & Derwing 1998; 2001, Paananen-Porkka 2007, p. 340), L2 teaching should pay more attention to practising appropriate speech rate in order to improve the communicative competence of the learners.

References

ABERCROMBIE, David 1967: *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

AHO, Eija & YLI-LUUKKO, Eeva 2005: Intonaatiojaksoista. – *Virittäjä*, **109**(2): 201–220. (abstract: Intonation units).

AVANESOV, Ruben I. 1956: *Fonetika sovremennogo russkogo literaturnogo jazyka*. Moscow: Izdatel'stvo Moskovskogo Universiteta.

BOERSMA, Paul & WEENINK, David 2008: Praat—doing phonetics by computer. <http://www.fon.hum.uva.nl/praat/>.

CENOZ, Jasone 2000: Pauses and hesitation phenomena in second language production. – *ITL: Review of Applied Linguistics*, **127–128**: 53–69.

CUCCHIARINI, Catia, STRIK, Helmer & BOVES, Lou 2002: Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. – *The Journal of the Acoustical Society of America*, **111**: 2862–2873.

COUNCIL FOR CULTURAL CO-OPERATION. EDUCATION COMMITTEE, MODERN LANGUAGES DIVISION, STRASBOURG AND COUNCIL OF EUROPE 2001: *Common European framework of reference for languages: learning, teaching, assessment.* Cambridge: Cambridge University Press.

DEESE, James 1980: Pauses, prosody, and the demands of production in language. – Hans W. Dechert & Manfred Raupach (eds.), *Temporal Variables in Speech*. The Hague: Mouton. 69–84.

FREED, Barbara F. 1995: What makes us think that students who study abroad become fluent? – Barbara F. Freed & Charles A. Ferguson (ed.), *Second Language Acquisition in a Study Abroad Context*. Amsterdam: Benjamins. 123–148.

FREED, Barbara F., SEGALOWITZ, Norman & DEWEY, Dan P. 2004: Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. – *Studies in Second Language Acquisition*, **26**: 275–301.

IIVONEN, Antti, NIEMI, Tuija & PAANANEN, Minna 1995: Comparison of prosodic characteristics in English, Finnish and German radio and TV newscasts. – K. Elenius & P. Branderud (eds.), *Proceedings of the XIII International Congress of Phonetic Sciences, Stockholm, 13–19 Aug. 1995*, vol. 2. 382–385.

LAFFORD, Barbara A. 2004: The effect of the context of learning on the use of communication strategies by learners of Spanish as a second language. – *Studies in Second Language Acquisition*, **26**: 201–225.

LAURANTO, Yrjö 2005: Sujuvuuden mittoja. – Leena Kuure, Elise Kärkkäinen & Maarit Saarenkunnas (eds.), *AFinLA Yearbook 2005*. Jyväskylä: Jyväskylän yliopisto. 127–147.

LEHTONEN, Jaakko 1978: On the problems of measuring fluency. – Matti Leiwo & Anne Räsänen (eds.), *AFinLA Yearbook 1978*. Jyväskylä: Jyväskylän yliopisto. 53–68.

LEHTONEN, Jaakko 1979: Speech rate and pauses in the English of Finns, Swedish-speaking Finns and Swedes. – Rolf Palmberg (ed.), *Perception and Production of English: Papers on Interlanguage*. Åbo Akademi. 3–19.

LEHTONEN, Jaakko 1981: Problems of measuring fluency and normal rate of speech. – Jean-Guy Savard & Lorne Laforge (eds.), *Actes du 5e Congrès de l'Association internationale de linguistique appliquée*. Montreal: Presses de l'Université Laval. 322–332.

MOORE, Kate & KORPIJAAKKO-HUUHKA, Anna-Maija 1996: The clinical assessment of Finnish fluency. – Martin J. Ball & Martin Duckworth (eds.), *Advances in Clinical Phonetics*. Amsterdam: John Benjamins. 171–196.

MUNRO, Murray J. & DERWING, Tracey M. 1998: The effects of speaking rate on listener evaluations of native and foreign-accented speech. – *Language Learning*, **48**: 159–182.

MUNRO, Murray J. & DERWING, Tracey M. 2001: Modelling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. – *Studies in Second Language Acquisition*, **23**: 451–468.

PAANANEN-PORKKA, Minna M. 2007: *Speech Rhythm in an Interlanguage Perspective: Finnish Adolescents Speaking English*. University of Helsinki.

PERFETTI, Charles A. 1985: *Reading Ability*. New York: Oxford University Press.

RIGGENBACH, Heidi 1991: Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. – *Discourse Processes*, **14**: 423–441.

SEGALOWITZ, Norman & FREED, Barbara F. 2004: Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. – *Studies in Second Language Acquisition*, **26**: 173–199.

SHTERN, A. S. 1988: Nekotorye statisticheskie harakteristiki russkoj spontannoj rechi. – Natalia D. Svetozarova (ed.), *Fonetika spontannoj rechi*. Leningrad: Izdatel'stvo Leningradskogo universiteta. 196–224.

SIMOES, Antonio R. M. 1996: Phonetics in second language acquisition: An acoustic study of fluency in adult learners of Spanish. – *Hispania*, **79**: 87–95.

SUOMI, Kari 2007: Accentual tonal targets and speaking rate in Nortern Finnish. – *Proceedings of Fonetik, TMH-QPSR, 50(1)*. 109–112.

TOWELL, Richard, HAWKINS, Roger & BAZERGUI, Nives 1996: The development of fluency in advanced learners of French. – *Applied Linguistics*, **17**: 84–119.

TROFIMOVICH, Pavel & BAKER, Wendy 2006: Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. – *Studies in Second Language Acquisition*, **28**: 1–30.

TROUVAIN, Jürgen 2003: *Tempo Variation in Speech Production: Implications for Speech Synthesis*. Institute of Phonetics, Saarland University.

ULLAKONOJA, Riikka 2008: Pausing as an indicator of fluency in the Russian of Finnish learners. – Plínio A. Barbosa, Sandra Madureira & César Reis (eds.), *Speech Prosody 2008: Fourth Conference on Speech Prosody, Campinas, Brazil*. 339–342.

VOLSKAYA, Nina forthcoming: Prosodic features of Russian spontaneous and read-aloud speech. – Viola de Silva & Riikka Ullakonoja (eds.), *Phonetics of Russian and Finnish, General Introduction: Spontaneous and Read-Aloud Speech*. Bern: Peter Lang.

Michael O'Dell & Tommi Nieminen (eds.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 111–121.

The categorisation of synthetic vowels by Swedish speaking listeners in Finland

Janne Savela[1], Ilkka Raimo[1], Esa Uusipaikka[1],
Olli Aaltonen[2] & Tapio Salakoski[1]
[1]Universiy of Turku, [2]University of Helsinki

Tiivistelmä

Turvotes-aineiston tavoitteena on ollut kuvata vokaalien tunnistamista eri kielissä. Yhtenä tutkituista kielistä on ollut suomenruotsi. Aineistoa on analysoitu monella tavalla. Ensin on piirretty kategorisointia ja hyvyysarvioita kuvaavat "vokaalikartat". Vokaalikarttojen katsotaan heijastavan niitä psykoakustisen avaruuden alueita, jotka heijastavat foneettisten liikkeiden suunnittelun perusteita. Aineisto koostui koehenkilöistä eri murrealueilta. Pääosa koehenkilöistä oli Turun seudulta. Lisäksi koehenkilöitä oli Pohjanmaan ja Uudenmaan murrealueilta. Koehenkilöittäinen tarkastelu keskittyi ensi sijassa /ø/:n alueeseen ja hyvien /ʉ/-vastausten alueeseen. Tämän ohessa aineistoon on kohdennettu log-lineaarinen regressioanalyysi, jonka tavoite on ollut selvittää mitkä akustiset parametrit selittävät suomenruotsalaisten vokaalien tunnistusta. Tutkimusta varten aineisto jaettiin hyviin (koehenkilöiden välinen yksmielisyys yli 85), melko hyviin (yksimielisyys 65–85 %) ja huonohin (yksimielisyys alle 65 %). Eri spektraalisten piirteiden, formantit tai spektrimomentit, selittävyyttä verrattiin keskenään. Todettiin spektrimomenttien lisäämisen malliin parantavan sen selittävyyttä suhteessa pelkästään formantteihin pohjaaviin malleihin. Tulokset ovat saman suuntaisia kuin aiemmat havainnot Turvotes-aineistosta.

Keywords: vowel perception, Swedish, Finnish

1 Introduction

The categorisation of synthetic vowels has remained a substantial method for studying vowel identification in different languages. It allows describing the criteria that listeners use in identification of the stimuli in a controlled paradigm. The present study is based on this type of experiment, namely on the data of the Turku Vowel Test (Turvotes). In this data base the identification data of different languages are compared between languages using same set of language-neutral synthetic stimuli.

Vowel studies on the Swedish spoken in Finland have been presented by Määttä (1983). He compared the identification of synthetic vowels in word context, manipulated with the OVEIII synthesizer. In Määttä's study there were 29 subjects representing the region of Jakobstad. The main purpose of the study was to compare the vowel identification of Finnish and Swedish listeners. Intra-group variation was not considered more specifically.

Since the study of Määttä (1983), the understanding of vowel perception has developed. Although Määttä already made a distinction between perceptually central and noncentral areas, the inner-structure of the vowel categories has provided interesting aspects of vowel identification. Since the 1990's vowel studies have been focused on individual vowel categorisation strategies. This means that the identification of subjects is the result of a process, in which the perceptual space of the subject emerges as the process of developing categorisation structure (e.g. Kuhl 2004). Research on pattern recognition models has provided techniques for understanding the vowel maps in terms of general log-linear models (Nearey & Kiefte 2003). Furthermore, the alternative spectral measures have been revisited in the vowel studies (Ito *et al.* 2001, Kiefte & Kluender 2005) showing alternative solutions to traditional peak picking paradigms. In Kiefte & Kluender's study the manipulation of spectral tilt affected the boundaries between [u] and [i]. Therefore the present study is interested in the three questions: Firstly, the differences between individual vowel categorisation charts between languages, secondly, the possible effects of the socially relevant languages in different areas of the language community, thirdly, modelling the identification data in terms of regression analyses.

2 Methods

2.1 Stimuli

The test consisted of synthetic vowels which covered the entire vowel space except for diphthongs and nasal vowels (Figure 1). The stimuli were synthesized with the Klatt serial synthesizer. The vowel space was created by varying F1 from 250 to 800 Hz in steps of 30 mel and F2 from 600 to 2 800 Hz in steps of 50 mel. F3 is 2 500 Hz as long as F2 is 2 000 Hz or below and 200 mel higher when F2 is above 2 000 Hz. The duration of the vowel stimuli was 350 ms and their pitch rose first from 100 Hz to 120 Hz (for the first 120 ms) and then fell to 80 Hz during the rest of the stimulus.

2.2 Subjects

The number of subjects was 18 (11 male and 8 female). The mean age was 34,8. Five subjects originated from Ostro-Bothnia (see Figure 2). Seven subjects were from the

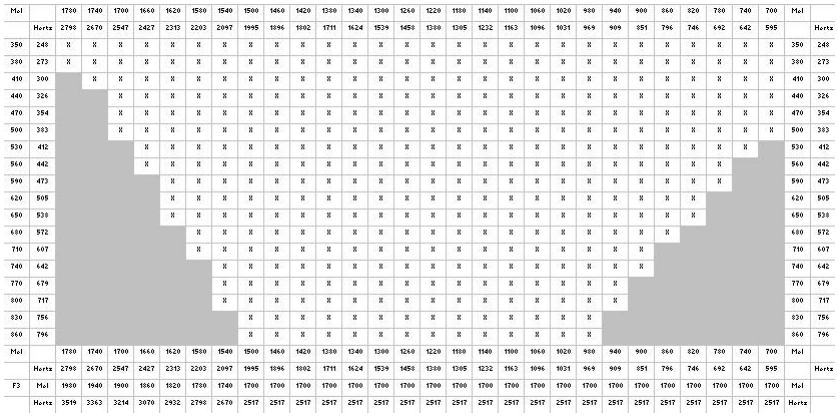


Figure 1: Turvotes Stimuli.

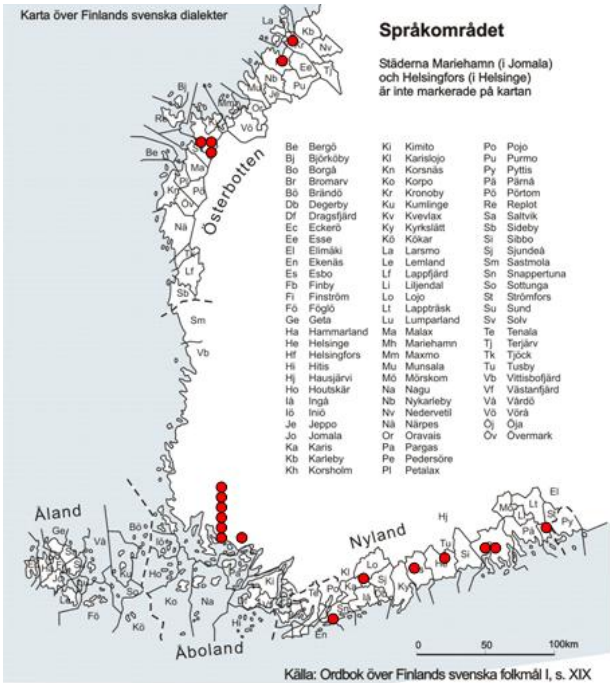


Figure 2: Subjects of the Study, places of the origin.

coast of Uusimaa. Finally, seven subjects were from the region surrounding Turku in South-Western Finland. All subjects had spent some time in their life in Turku.

2.3 Procedure

The tests were made in the language laboratory of the University of Turku. The test took about 45 minutes. Subjects were instructed, firstly, to identify the stimuli according to the letters on the screen, and secondly, to rate the stimuli they heard on a scale of 1 to 7. In the case of Swedish (spoken in Finland) there were nine vowel categories (a distinction between /e/ and /æ/ occurring before the sound /r/).

3 Results

3.1 Descriptive analysis

The results were analysed on the basis of two types of vowel maps. The first revealed the categorisation strategies (Figure 3) of the subjects. The second type was based on the goodness ratings of the subject (Figure 4).

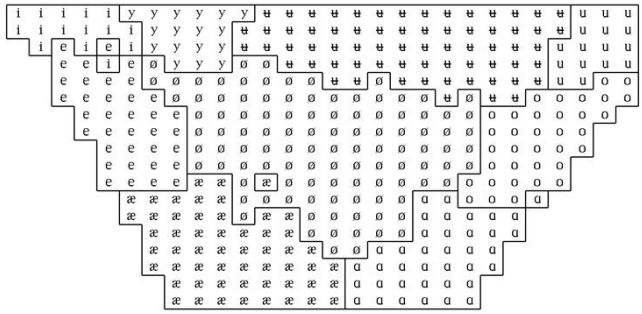


Figure 3: The relative majority of answers on different points of the vowel space.

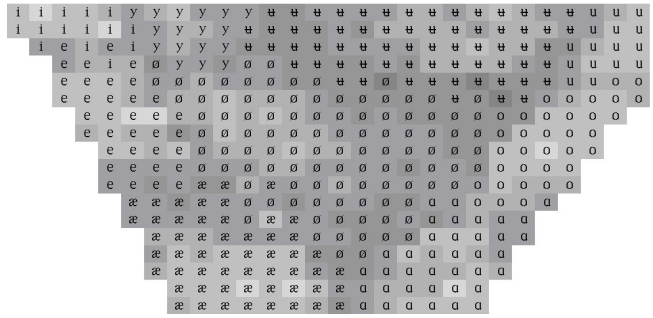


Figure 4: The average goodness rating of the subjects. The lighter the area the higher the average rate of the particular stimulus.

In order to show the possible differences between individual listeners the identification maps for different dialect groups were studied.

The first group consisted of speakers in Ostrobothnia (Figure 5). Some remarks can be made on the answers of the Ostrobothnian listeners. In comparison with the average vowel chart, some listeners have accepted low values of F1 as belonging to category /ø/. There is also relatively large variation in the categorisation of /ʉ/ sounds. This may be based on subjective differences.

The next figure shows the answers of people from South-Western Finland (mainly from Turku; Figure 6). The results are distinct from the earlier group. Some subjects had their /ø/ area with a higher F2. This may have been an idiosyncratic feature for them. Subject 2 also had an interesting pattern for /e/ which included areas with higher F2 values.

The third group comprised the people from the Uusimaa area. Their results were also some what mixed. In the case of /ø/, it did not was as open as in the Ostrobothnia area. In the case of /ʉ/, Subject 6 had a smaller area of the responses in terms of F1 than the other subjects. Other subjects preferred a wider range of F2 values.

Altogether the vowel charts differed between subjects. Interestingly, some subjects had vowel charts similar to Sweden Swedish (Figure 8). Identification of both critical vowels /ø/ and /ʉ/ was close to Stockholm Swedish. On the other hand some subjects had a chart similar to Finnish, e.g. Subject 9. Her map greatly resembles the Finnish chart, having a /y/ area similar to that of the Finnish average map (Figure 9).

This result may be due to the effect of Finnish in the different idiolects of Finland Swedish. It is possible that in Finnish speaking areas the Swedish speaking subjects have adapted their vowel identification to Finnish, whereas the areas with links to Sweden Swedish have remained closer to Sweden Swedish.

3.2 Statistical analysis

The second aim of the study was to compare the role of spectral moments in different languages. The aim of these studies is to compare the relative strength of different spectral attributes in explaining the identification data of the Swedish speakers. The method was a log-linear regression model based on frequencies. In this methodology a dependent variable (vowel category) is predicted on the basis of continuous and/or categorical independent variables (formant and spectral moments). The relative importance of independent variables can be investigated and represented in terms of Wald statistics to show the significance of individual independent variables.

In addition to the two lowest formants, four spectral moments were calculated for each stimuli (Figure 10). They were measured from the Turvotes vowel stimuli using the PRAAT analysis system. They describe the shape of the energy distribution throughout the whole spectrum and are determined by all the frequencies of the spectral envelope.

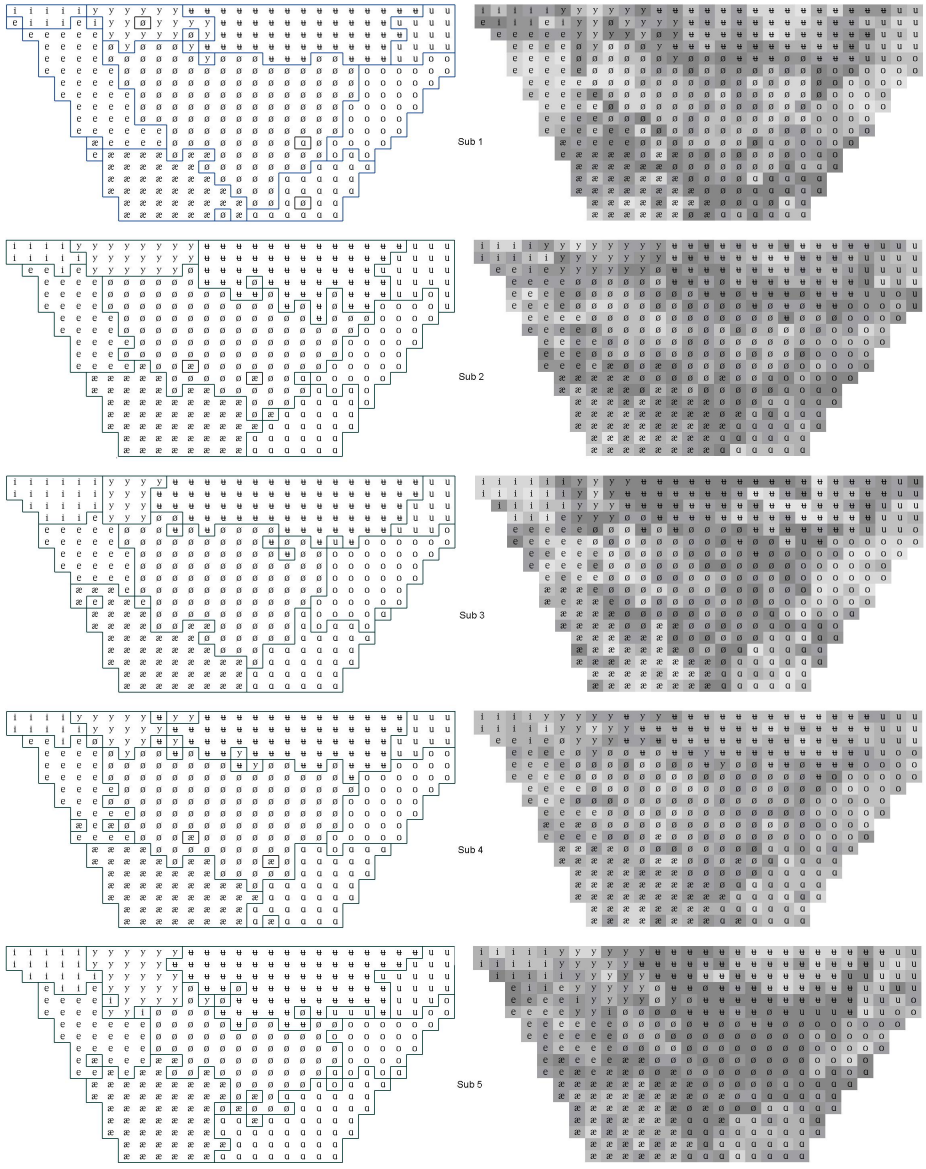


Figure 5: Identification charts (left column) and categorisation charts (right column). The home areas of the subjects were: Subject 1 Korsholm, Subject 2 Vasa, Subject 3 Korsholm, Subject 4 Kronoby, Subject 5 Pedersöre.

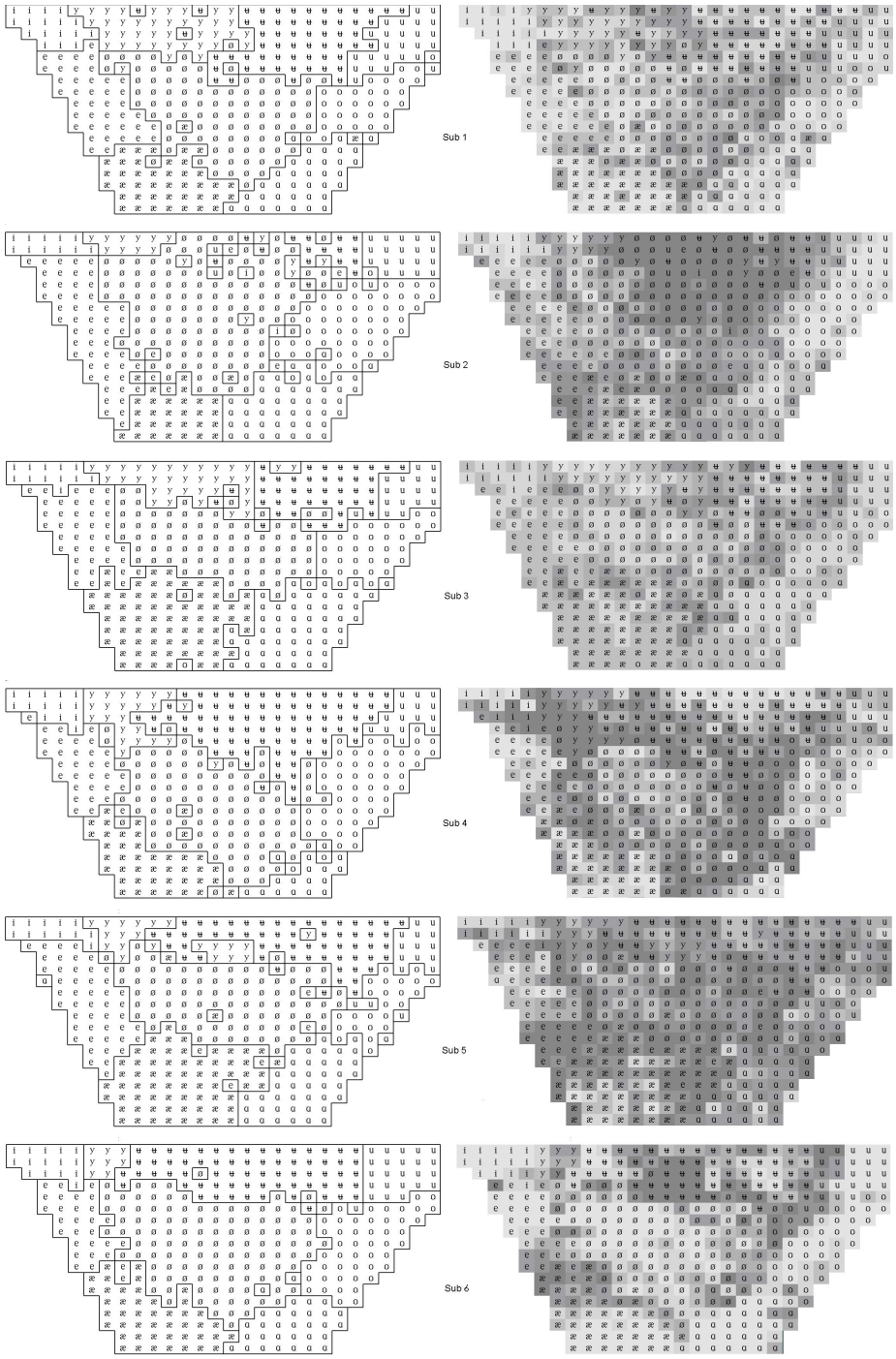


Figure 6: The answers of the Swedish speaking people in Ostrobothnia. All subjects were from Turku, except Subject 1, who was from Kaarina neighbouring Turku.

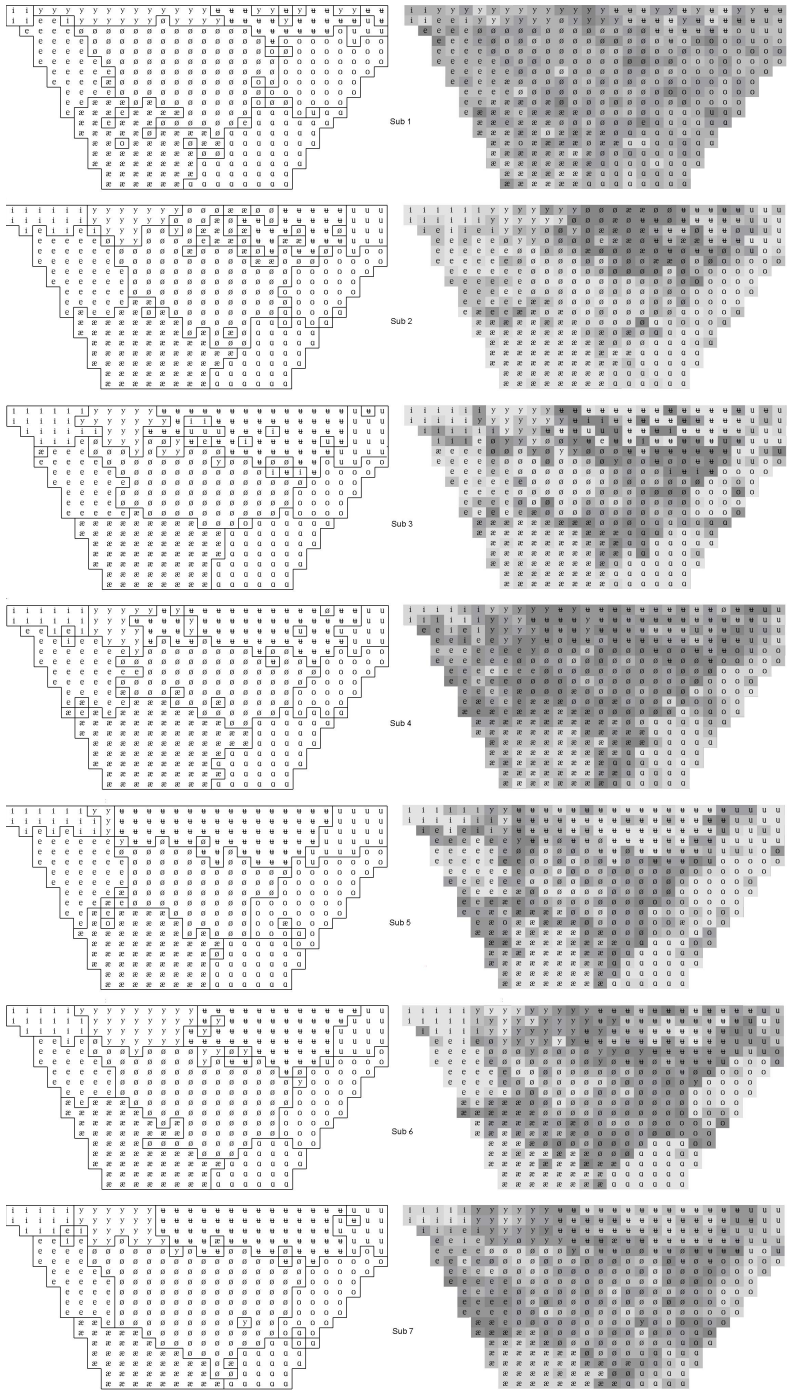


Figure 7: The answers of the subjects from Uusimaa. The three first subjects were from Central Uusimaa (Kauniainen and Vantaa). Two next subjects came from relatively unilingual Snappertuna and less unilingual Virkkala, which are both located in Western Uusimaa. Finally, there were three subjects from Eastern Uusimaa (Porvoo and Loviisa).

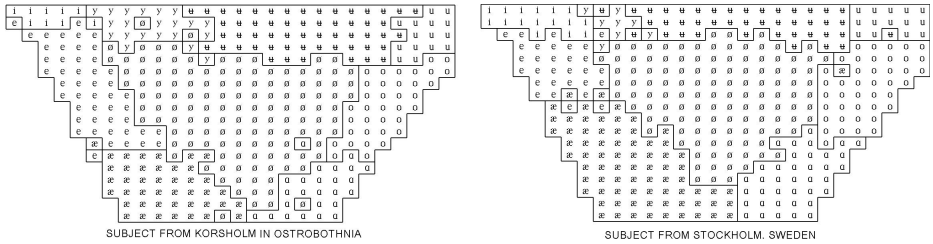


Figure 8: The identification charts of one Finland Swedish and one Swedish Swedish subject.

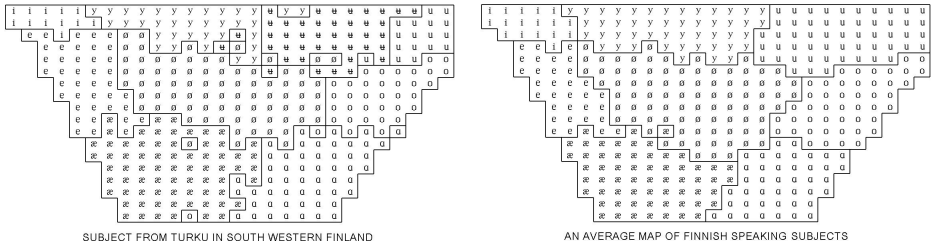


Figure 9: The identification charts of one Finland Swedish and average of 67 Finnish subjects.

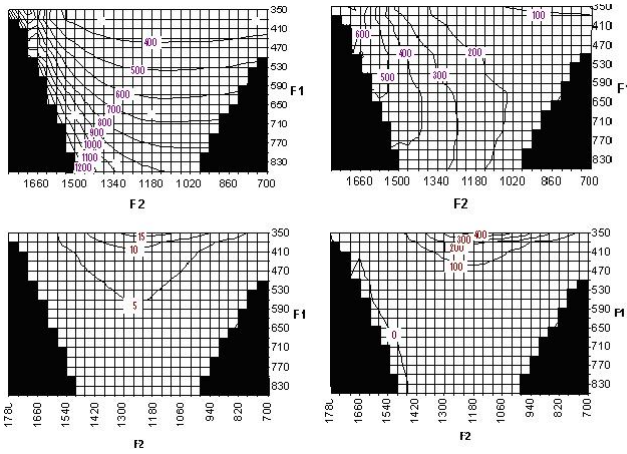


Figure 10: The spectral moments plotted against the loci in the F1-F2 vowel grid. The values of formants and Centre of Gravity (upper left) are in mels and values of standard deviation is in Hertz (upper right). Finally, the skewness (lower left) and kurthosis (lower right) are described in terms of coefficient.

Two models (formants and formants with spectral moments) were used in the analysis. Furthermore the data was divided into good, medium and poor data sets. To emphasize the goodness ratings, the data was weighted with the goodness ratings, meaning higher unanimity for areas with a large number of high rates. In the first data set, the mode of each stimulus received over 85 % unanimity and in the poor data set the unanimity was under 65 %.

The first analysis used a $-2\log$ criterion for each type of vowel data. This value makes it possible to compare models using data with different numbers of observations (Table 1). A lower value means a better model.

Table 1: The $-2\log$ values of each vowel set.

	Formants	Formants and spectral moments
Good	1558.558	1523.054
Medium	3814.824	3777.102
Poor	15295.252	15066.445

A closer look at the relative strength of different factors was also made (Table 2). Although the models with more spectral components were better, the results showed no significance for spectral features other than formants.

Table 2: The relative strength of different explaining factors. The Wald Chi-Square (χ^2) is a statistical measure explaining the relative strength of a particular variable.

		Good		Medium		Poor	
Effect	DF	Wald χ^2	Sig.	Wald χ^2	Sig.	Wald χ^2	Sig.
F1	8	30.4745	0.0002	45.8194	<0.0001	563.7905	<0.0001
F2	8	63.2809	<0.0001	78.1227	<0.0001	649.6217	<0.0001
$F1^2$	8	80.3700	<0.0001	61.2073	<0.0001	122.8588	<0.0001
$F2^2$	8	83.6867	<0.0001	92.2668	<0.0001	594.9176	<0.0001
$F1 \times F2$	8	19.1292	0.0078	24.5999	0.0018	367.9450	<0.0001
cog_mel	8	21.2289	0.0034	17.0422	0.0297	40.0389	<0.0001
std_Hz	8	18.2751	0.0193	16.7067	0.0333	37.5707	<0.0001
kurtosis	8	16.6419	0.0341	23.6027	0.0027	65.3252	<0.0001
skewness	8	21.3084	0.0064	28.7138	0.0004	45.0175	<0.0001

4 Discussion

The present study surveyed the identification data of speakers of Swedish in Finland. In contrast to earlier studies (Määttä 1983), the present data included data from different regional varieties of Swedish in Finland. The general picture showed that all listeners were able to identify the vowels proposed for them in the vowel test. There were differences between dialectal groups of the different listeners. Some of the listeners seemed to have a more Swedish Swedish vowel system and some listeners had a vowel chart closer to Finland Swedish listeners. This aspect should be studied more thoroughly.

The statistical models showed that the additional spectral features, especially spectral moments, did not make the log-linear models more effective. However, this could be due the sample that was used in the statistical analysis. The division of the data into different groups based on the regional/dialect variation would have made the analysis more efficient. The study showed that in the case of Swedish in Finland, the traditional understanding of the formants as the primary carrier of vowel identity holds.

References

ITO, Mashadi, TSUCHIDA, Juri & YANO, Masafumi 2001: On effectiveness of whole spectral shape of vowel perception. – *Journal of the Acoustical Society of America*, **117**(3):1395–1404.

KIEFTE, Michael & KLUENDER, Keith R. 2005: The relative importance of spectral tilt in monophtongs and diphthongs. – *Journal of the Acoustical Society of America*, **110**(2):1141–1149.

KUHL, Patricia 2004: Early language acquisition: Cracking the speech code. – *Nature Reviews Neuroscience*, **5**(11):831–843.

MÄÄTTÄ, Taisto 1983: *Hur finskspråkiga uppfattar svenskans vokaler: En studie i kontrastiv fonetik med naturligt och syntetiskt tal*. Umeå Universitet.

NEAREY, Terrance & KIEFTE, Michael 2003: Comparison of several proposed perceptual representations of vowel spectra. – *Proceedings of the XVth International Congress of Phonetic Sciences*. 1005–1008.

Michael O'Dell & Tommi Nieminen (toim.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 123–137.

Artikulaation demonstraatio-ohjelma

Antti Iivonen
Helsingin yliopisto

Tiivistelmä

Esittelyn kohteena oleva tietokoneohjelma ArticulationDemo on tarkoitettu edistämään opittavan kielen artikulaation tiedostamista ja oppimista. Toistaiseksi ohjelma näyttää suomen kielen vokaalien ja konsonanttien äänteiden dominantit tavoiteasennot sivusta katsottuna. Ne esitetään puhujan artikulaatioelimistön liikkuvana tasosivukuvana, "tomografisena filminä" neutraalivokaalin asennosta kuvattavan äänteen tavoiteasemaan. Erikoistapauksia ovat ovat *h*, *l* ja *r*. Konsonantti *h*:lla ei ole dominanttia tyyppiä suuren kontekstuaalisen variaation takia. Konsonantti *l*:n tavoiteasennon lateraaliaukkoja ei ole mahdollista esittää kaksiulotteisessa lateraalikuvassa. Konsonantti *r* on täryinen, ja siksi ohjelma näyttää sen tavoiteasemassa kielen kärjen liikettä, täryjä, joita ei stillkuvassa voi näyttää. Käyttäjän valikossa on seuraavat optiot: vokaalin tai konsonantin näyttö, kahden äänteen artikulaatiokonfiguraation vertailu, artikulaatioterminologiaa koskeva sivu ja infosivu. Ohjelman kehittelyn ratkaisuja esitellään. Taustaksi esitetään muutamia historiallisia tietoja ja tietoja muista opetusteknologisista artikulaatio-ohjelmista. Myös artikulatorisen tiedon alkuperää käsitellään. Vokaalien ja konsonanttien määrittelyt sekä ohjelman toiminta ja vuokaavio esitellään.

Avainsanat: opetusteknologia, fonetiikan opetus, artikulaatio, suomi

1 Johdanto

Renessanssiajalta alkanut ihmisen anatomian tutkimus teki vähitellen mahdolliseksi puhe-elimistönkin tarkan tuntemuksen. Kielten, etenkin ranskan ja englannin, äänteellinen muuttuminen kiinnitti huomiota 1500-luvulla kirjoituksen ja puheen ristiriitaisuuksiin, mikä edisti puhetietoisuuden kehittymistä. Tähän vaikutti myös kuuromykkien puheen opetus. Hellwagin (1781) saksaa koskevan vokaalikolmion kuvaus osoitti äänteiden auditiivisiin piirteisiin nojaavan kuvauksen mahdollisuuden. Äänteiden ja yleensä puheen akustiikka on nykyäänkin asiantuntijoita lukuun ottamatta kokonaan tietoisuuden ulkopuolella. Useimmille kielenkäyttäjille riittää se, että osaa puhua. Tietoinen kiinnostus nousee vasta tilanteessa, jossa puhe ei onnistukaan normaalisti. Tämä tapahtuu häiriötapauksissa ja uuden kielen opiskelussa. Foneettisen

opetuksen suunnittelijoiden on siis syytä ottaa tämä huomioon, koska jokainen oppija joutuu mielessään käymään samantapaista tietoisuuden kehittämistä kuin mitä on tapahtunut ihmiskunnan kehityksen aikana. Fonetiikassakin perinteisin tapa kuvata äänteitä nojaa artikulaatiopiirteisiin. Johtopäätös on siten, että opettajan ja opetettavan on aiheellista aloittaa kysymyksestä, mitä artikulaatio oikeastaan on.

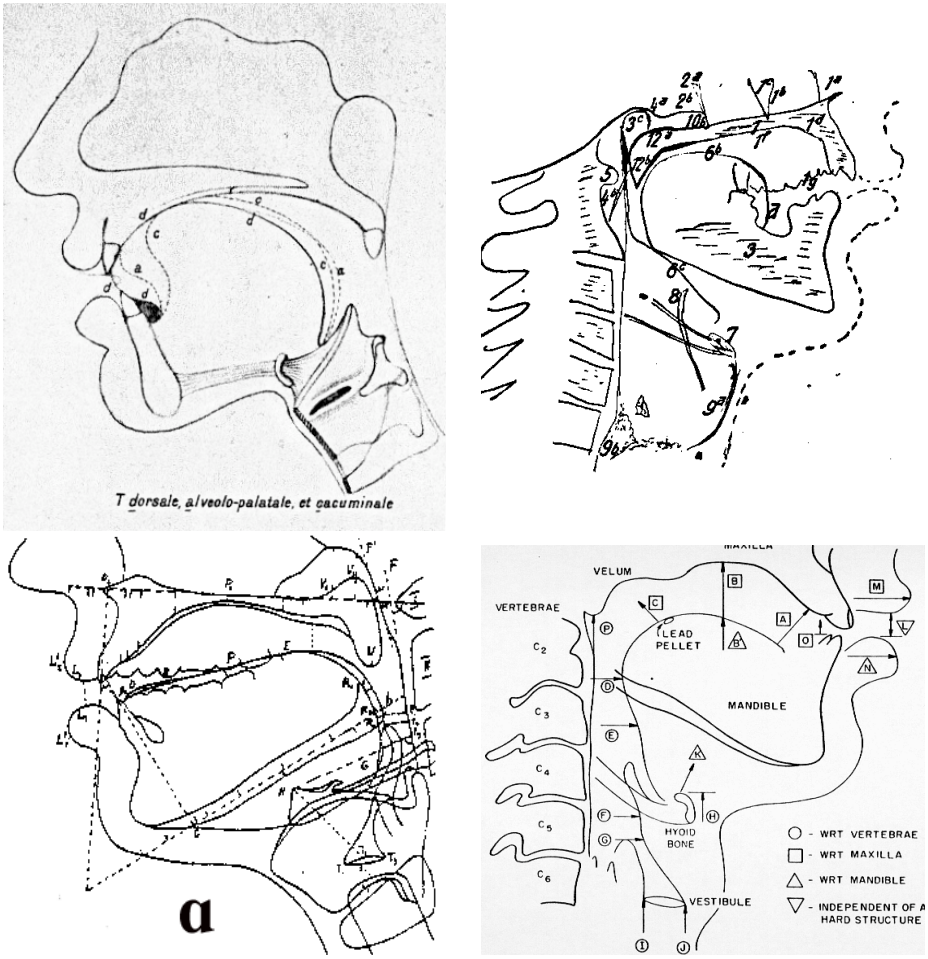
2 Mistä tieto artikulaatiosta on peräisin?

Artikulatorisen tiedon lähteet ovat introspektio, tunto- ja kinestesia-aisti, motorinen kuuleminen ja kokeellinen tutkimus. Introspektiolla kohdistamme tietoisen tarkkailun kohteeseen ja kohteen nimittämiseen (esim. suomen *a*:n artikulaatioon). Hyviä tuntopalautteita saamme huulista ja kielen etuosasta mutta emme samassa määrin kielen selästä, kielen juuresta tai velumista. Kinestesia eli asentoaisti saattaa auttaa jossakin määrin vokaalien artikulaation havaitsemisessa. Motorinen kuuleminen viittaa siihen, että kuullessamme puhuttuja ilmauksia muodostuu samalla aivojen motorisella alueella vasteita, joiden perusteella voimme ”päätellä”, miten ko. ilmaus on tuotettu ja voimme yrittää matkia ilmausta. Peilin avulla voi tarkkailla etenkin huulia ja kielen etuosaa. Kokeellisessa tutkimuksessa evidenssiä tarjoavat huulten valokuvaus ja videointi edestä ja sivulta, röntgen sivulta ja edestä, artikulografia ja (funktionaalinen; dynaaminen) magneettiresonanssikuvantaminen (MRI) edestä ja sivulta (ks. myös Ridouane 2006). Tasoröntgenillä voidaan saada kohdennettuja kuvia kiinnostavilta alueilta. Voidaan esimerkiksi tutkia, esiintyykö *s*:n aikana todellakin kielen selässä pitkittäiskouru.

3 Historian välähdyksiä

Artikulatorisen tiedon tie on ollut pitkä ja vaivalloinen. Historiasta tässä vain muutama välähdys ääntöväylän (*vocal tract*) ja äänteiden välisen suhteen tutkimuksesta. Ääntöväylän anatomisen sivukuvan lienee ensimmäisenä esittänyt Leonardo da Vinci 1500-luvun alussa (Panconcelli-Calcia 1961, s. 91). Nenäväylää kuvassa ei ole. Kolmessa lisäpiirroksessa esitetään vokaalien *a*, *o* ja *u* huuliartikulaatiot. Kuvaus ei ole tarkka, mutta ajankohtaan nähden huomattava saavutus. Muistettakoon, että tuolloin ei ollut esimerkiksi selvillä, että soinnillinen ääni saadaan aikaan äänihuulilla.

Merkel (1866) esitti suuren määrän artikulaation sivukuvia. Ne ovat huomattavan tarkkoja ja anatomisesti korrekteja. Merkel oli lääketieteilijä ja tunsi hyvin ihmisen anatomisen tutkimuksen. Sen sijaan äänteiden kuvaus oli vasta oraalla. Kuvan 1 esimerkissä nenäportti, kova ja pehmeä suulaki, kurkunpään rustot, kurkunkansi ja kieliluu ovat korrektisti kohdallaan. Kielen osalta Merkel esittää dorsaalisen, alveolo-palataalisen ja kakuminaalisen artikulaation. Niillä hän tarkoittaa kuvassa kielen etuosan ja selän kontaktia hampaisiin, hammasvalliin ja palatumiin (”dorsaa-



Kuva 1: Ääntöväylän foneettisen kuvauksen historiallisia esimerkkejä. Vasemmalla yllä: kolme artikulaatioasentoa (Merkel 1866). Oikealla yllä: vokaali *u* Scheierin röntgentutkimuksen perusteella (Poirot 1911, s. 34) Vasemmalla alla: suomen *a*-vokaali (Sovijärvi 1938). Oikealla alla: kielen ja huulten etäisyyksien mittaus toisista artikulaatioelimistä (Perkell 1969).

linen" = d), siitä takaisempaan kohtaan hammasvallissa ja siitä taaksepäin ("alveolopalataalinen" = a) sekä artikulointia kitakupuun (*cacumen*, kovan suulaen kaareva kohta; "kakuminaalinen" = c). Termit eivät vastaa täysin nykyisiä termejä. Ottaen huomioon, ettei röntgenkuvausta ollut olemassa, Merkelin kuvat ovat merkkipaalu äänteiden artikulaation kuvauksessa.

Röntgenkuvaus keksittiin vuonna 1885, ja pian sitä sovellettiin myös foneettisiin tutkimuksiin (Scheier 1897). Kuvan 1 röntgenkuvaan perustuva piirros *u*-vokaalista

(Poirot 1911, s. 34; Poirot viittaa virheellisesti tekijänimeen ”Schleier”) on jo artikulaationkin osalta varsin todenmukainen. Jones (1917) käytti röntgenkuvia kardinaalivokaalien neljän nurkkavokaalin määrittelyyn (jotka on julkaistu myös esim. Jones 1958: nimiösivua edeltävä sivu; ks. myös www-osoitteet tämän kirjoituksen lopussa), mutta muut kardinaalivokaalit hän määritteli auditiivisina etäisyyksinä noista neljästä. Alkuperäisten röntgenkuvien perusteella nurkkapisteet eivät muodosta nelikulmiota vaan ellipsin, joka on julkaistu muun muassa edellä mainitussa Wikipedian osoitteessa. Elliptiseen muotoon on saattanut vaikuttaa se, että englannin tiukka ja usein pitkä [u] on enemmänkin keski- kuin takavokaali (vrt. englannin *room* versus suomen *ruuma*). Kirjallisuusviitteiden jälkeen mainitussa www-osoitteessa (>Kardinaalivokaalit) voi kuulla, kuinka yllättäviltä suomenkielisen korvissa osa Jonesin itsensä vuonna 1956 ääntämistä kardinaalivokaaleista kuulostavat. Ladefoged (1967) on kommentoinut kriittisesti kardinaalivokaalijärjestelmää.

Sovijärven (1938) väitöskirja sisältää röntgentekniikkaan perustuvat suomen vokaalien ja nasaalien artikulaatiokuvat (*a*-vokaali kuvassa 1). Artikulaatioelinten ääriviivat ovat hyvin näkyvissä. Teknisesti perinteinen röntgentekniikka eteni vielä röntgenfilmin keksimisellä, mutta foneettisesti sen huippu lienee saavutettu jo mainitussa julkaisussa. Huomattavaa on, että alkuperäisestä röntgenkuvasta saadaan käyttökelpoinen kuva vain jäljentämällä se eli kalkioimalla. Wängler (1974) on julkaissut kirjan, jossa alkuperäisten röntgenkuvien päälle on asetettu läpinäkyvä jäljennöspiirros. Näitä tarkastelemalla saa vaikutelman, että tulkinnalle on jäänyt pelivaraa. Tämä koskee Wänglerin nimikkeellä ”alveolar-coronal” varustettuja äänteitä *d*, *t*, *n*, *l* ja *r*, koska kielen etuosa ei röntgenkuvissa erotu. Jäljennekuvat tosin näyttävät osoittavan, että kielen kärjen kosketus mainituissa äänteissä on etupäässä apiko-alveolaarinen, r:ssä se olisi mahdollista tulkita jopa retrofleksiseksi, siis subapikaaliseksi. Termiin ”apikaalinen” Wängler viittaa vain alaviitteessä s. 30 sekä s. 41 (”apikaalinen *s*”).

Perkell (1969) jalosti vielä röntgentekniikkaan perustuvaa artikulaatiotutkimusta mittaustekniikan osalta (kuva 1). Kuvien laatu on korkea ja mittauskohteet on merkitty selvästi. Säännöllisin aikavälein otetuista kuvista on mahdollista rekonstruoida artikulaatioliike. Kuvattuun korkeaan laatuun nähden eräät Perkellin myöhemmin (1997, s. 351) julkaisemat kuvat englannin konsonanteista ovat yllättävän epätarkkoja. Röntgenin yksi heikkous onkin ollut nimenomaan kielen kärkeä ja lapaa koskevan kuvauksen epätarkkuus. Tähän ovat syynä muun muassa kuvauslinjalla oleva hammasrivi ja hampaiden paikat.

Röntgentekniikan jälkeen on kehitetty uusia menetelmiä: sähkömagneettinen artikulografi (EMA; esim. Hoole & Nguyen 1997) ja magneettiresonanssikuvantaminen (MRI). Moore (1992) esitti jo kiinnostavia MRI-kuvia amerikanenglannin vokaaleista ja /r/-foneemin artikulaatiosta. Tuolloin äänteet oli äännettävä pitkitettyinä (*sustained*) heikon aikaresoluution vuoksi. Jatkuvan puheen kuvantamiseen 10 ms:n resoluutio olisi tarpeen. Koejärjestelyn viiden puhujan tulokset poikkeavat toisistaan,

mikä osoittaa sen, että saman foneemin artikuloinnin yksilökohtainen vaihtelu on tosiasia. Se on yksi todiste invarianssin puuttumisesta.

Kiintoisia ovat myös pyrkimykset ennustaa artikulaatio akustiikan perusteella (Richmond *et al.* 2003).

4 Artikulaation kuvantaminen oppimateriaalina

Vanhastaan tunnettuja ovat eri kielten fonetiikan oppikirjojen äännekuvat (Jones 1958, Ward 1929, Sovijärvi 1963, Wängler 1974). Sovijärvi (1968) kehitti jo 1960-luvulla artikulaation demonstraatio-ohjelman *Apparatus for Demonstration Articulatory Movements* (A.D.A.M.). Se perustui magneettiseen ketjuun, joka kiinnitettiin esitystaululle ja jonka konfiguraatiota voitiin käsin muuttamalla osoittaa äänteiden artikulaation sivukuva. Internetissä on nykyisin oppimateriaalia, joiden joukossa on hyvin toimiva ja monipuolinen amerikanenglannin äänteiden opetusohjelma (ks. www-osoitteet kirjoituksen lopussa). Saksan ääntämistä simuloiva SpeechTrainer-ohjelma on myös verkossa (ks. www-osoitteet). Sen kehittäjä Berndt Kröger oli mukana myös Kölnin yliopiston fonetiikan laitoksessa Georg Heiken käynnistämän DYNAMO-ohjelman (Heike *et al.* 1986) kehitystyössä. Tekesin PUMS-ohjelmassa on Helsingin yliopiston Puhetieteiden laitoksessa kehitetty myös brittienglantia ja suomea koskeva opetusohjelma (ks. www-osoitteet). Verkossa on myös IPA:n äänteitä demonstroiva ohjelma. Sen äännemallit ovat varsin viitteellisiä (ks. www-osoitteet). Huomattavan kunniahimoinen on japanilainen PHAROS- ohjelma (www-listassa: Dang Pharos) sekä huulten artikulaatiota koskeva ohjelma (www-listassa: Huulten artikulaatio). Ohjelmien kehitystyötä tarvitaan vielä paljon todenmukaisuuden ja käytettävyyden parantamiseksi.

5 Artikulaation demo-ohjelma

Esiteltävässä ohjelmassa seuraavat valinnat ovat mahdollisia: (1) Äännevalikosta voi valita minkä tahansa äänteen. Aloitusajankohdassa puhe-elimet ovat neutraalivokaalin asennossa, nenäväylä suljettuna, huuliaukko avoimena. Liike konsonanttiin tai vokaaliin esitetään, jonka jälkeen seuraa paluu neutraalivokaalin asemaan. (2) Käyttäjä voi vertailla kuvassa kahden eri äänteen päällekkäin asetettua tavoiteasemaa. (3) Käyttäjää auttanee, että hän voi valita kohdan, jossa ohjelma näyttää artikulaatioelimet ja niiden nimet. (4) INFO-ikkuna antaa tietoja ohjelmasta suomeksi ja englanniksi. Ohjelma on toteutettu Macintosh-ympäristössä (G3, G4) FutureBasic II -kielellä. Windows-ympäristöä varten on olemassa erillinen PowerPoint-demo. Ohjelman kehittelyä on edistänyt Tekesin Fenix/Pums-4-tutkimusohjelma vuosina 2006–07.

Ohjelman toiminnan vuokaavio on esitetty kuvassa 2. Olennaiset toimintaratkaisut ovat seuraavat:

1. Käytettävyyden ja informatiivisuuden varmistamiseksi ohjelma hakeutuu nopean aloituksen jälkeen suoraan päävalikkoon. Käyttäjä voi valita siitä tarvittaessa INFO-ikkunan tai artikulaatioelimiä käsittävän ikkunan. Elimen nimi ilmestyy ruutuun ja samalla elimen kohdalle ilmestyy vastaava numero.

2. Kutakin liikkuvaa ja liikkumatonta artikulaatioelimistön osaa vastaa 25 koordinaattapisteparia (tämä ei koske kuitenkaan kurkunkantta, kurkunpään elimiä, ruokatorvea ja nielun takaseinämää, joissa vähäisempi määrä riittää). Äänteiden todenkaltaisuutta tavoittelevat artikulaatioprofiilit muodostettiin suomen äänteitä koskevan röntgenkuva-aineiston (etenkin Sovijärvi 1938; 1963) perusteella käyttäen erillistä apuohjelmaa. Puhujakohtaisen anatomisen vaihtelun vuoksi oli sovellettava puhujanormalisointia. Toistaiseksi ohjelma käyttää miehen anatomista sivuprofiilia, jonka mallina on todellinen röntgenkuva (Bonnot *et al.* 1989, s. 297). Etenkin kielen etuosan artikulaatiota koskevan tiedon hankinnassa apuna olivat myös kirjallisuudessa esitetyt verbaaliset luonnehdinnat ja omat taktiiliset havaintoni.

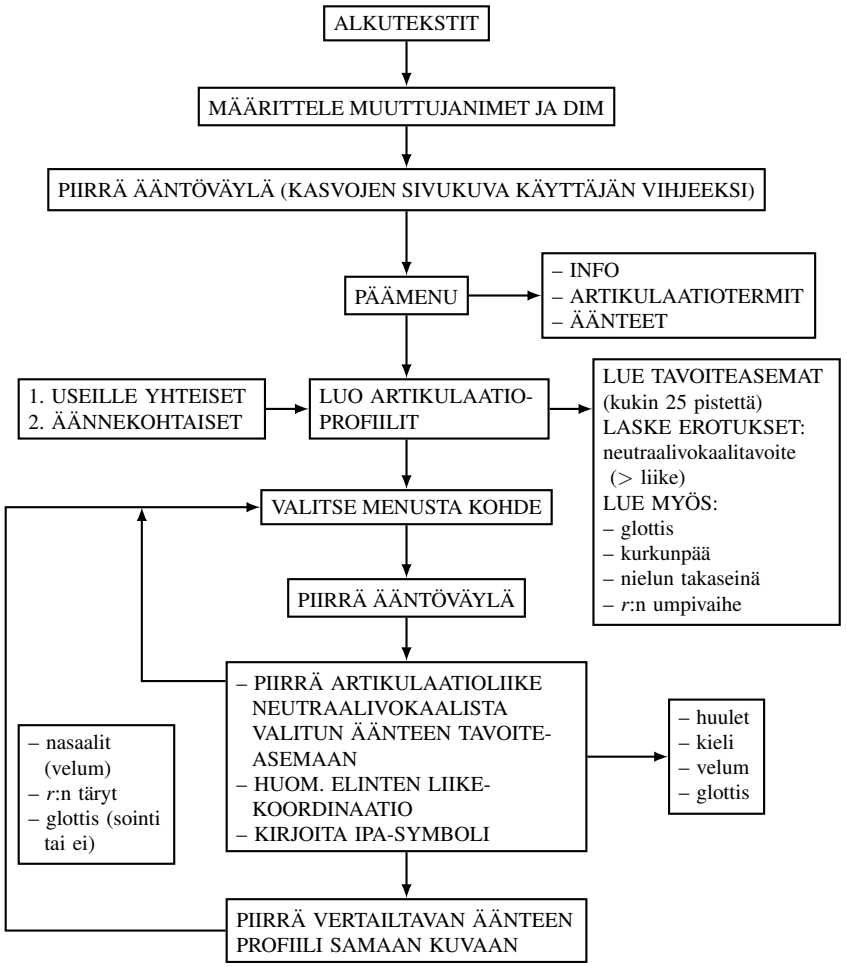
3. Koska useilla äänteillä on osin yhteisiä ominaisuuksia, näitä ja äännekohtaisia ominaisuuksia yhdistelemällä ohjelma piirtää tavoitekonfiguraation.

4. Lähtö- ja tavoiteaseman välisen, ajallisesti peräkkäisen liikkeen muodostamiseksi kunkin vastepisteen välinen matka jaettiin kuuteen välivaiheeseen. Silmän heikkoa reagointikykyä harhautettiin piirtämällä ja poistamalla nopeasti kukin välivaihe, jolloin silmä tajuaa liikkeen.

5. Artikulaatioelinten simultaanisen liikekoordinaation (kieli, huulet ja velum) demonstroimiseksi kunkin äännetyypin tuottamiseen tarvittavien elinten, esimerkiksi ylä- ja alahuulen, liikkeiden kukin välivaihe ohjelmoitiin `FOR...TO`-lausekkeisiin, jolloin tietokoneen prosessointinopeuden ansiosta katsoja näkee elinten liikkeet samanaikaisina. Nasaalien nenäportin avautuminen versus sulkeutuminen sijoitettiin nasaalin alkuun versus loppuun. /r/:n reaalistumisen täryt ilmaistaan vain kolmella täryllä umpi- ja avovaiheen välillä. Glottiksen värähtely ilmaistaan pikemminkin symbolisesti soinnillisten äänteiden aikana värähtelynä glottiksen avo- ja sulkuvaiheen välillä. /h/:n demonstraation vaikeudesta enemmän jäljempänä.

6. Äänteiden profiilien vertailussa ensimmäinen äänne kuvataan liikkeenä neutraalivokaalista tavoiteasemaan. Vertailtavasta äänteestä piirretään eri värillä samaan kuvaan päällekkäin vain tavoiteasema.

7. Äänneprofiiliin piirretään aina myös kansainvälisen foneettisen kirjoituksen mukainen äännesymboli. Jos kysymyksessä on vertailu, merkit piirretään rinnakkain samalla värillä kuin niitä esittävä profiiliviiva.



Kuva 2: Artikulaation demo-ohjelman vuokaavio.

6 Suomen kielen äännemallit

Suomessa on vokaalifoneemit /i e æ y ø a o u/ ja konsonanttifoneemit: /p t d k m n ŋ l r s v j h/ ja lisäksi lainakonsonantit /b g f/. Foneemi /š/ on erittäin marginaalinen suomessa. Niiden tarkemmat foneettiset symbolit ilmenevät äännekuvista (kuva 3 ja liite 1).

Vokaalien artikulaatiopiirteet esitetään taulukossa 1 ja artikulaatiomallit kuvassa 3. Piirteiden ”korkea / keskikorkea / matala” synonyymejä ovat ”suppea / välinen / väljä” (ks. terminologiaa ja suomen äänteiden määrittelyjä myös Suomi *et al.* 2006). Distinktiivien piirteiden teoriassa erotetaan vain dikotominen vastakohtaisuus

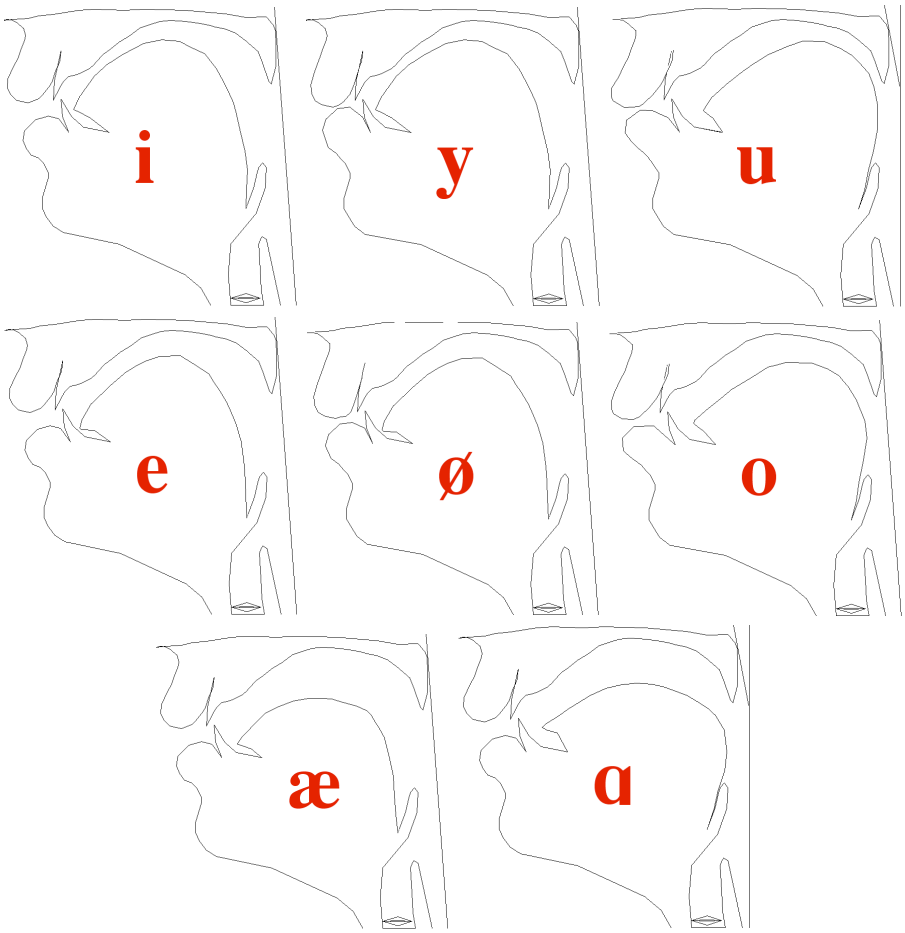
Taulukko 1: Suomen vokaalien artikulaatiopiirteet. Plus-merkki ilmaisee tässä piirteen voimakkaampaa astetta.

/V/	kielen asema vaakatasossa	kielen korkeusasema	huuliaukon tila
i	+++ etinen	korkea	lavea
e	++ etinen	keskikorkea	lavea
æ	+ etinen	matala	lavea
y	+++ etinen	korkea	++ pyöreä; huulet pyöristyvät toisiaan kohti
ø	++ etinen	keskikorkea	+ pyöreä; huulet pyöristyvät toisiaan kohti
a	+ takainen	matala	lavea
o	++ takainen	keskikorkea	+ pyöreä; huulet pyöristyvät eteenpäin
u	+++ takainen	korkea	++ pyöreä; huulet pyöristyvät eteenpäin

lavea/pyöreä. Foneettisesti kuitenkin pyöreydessä on graduaalisia eroja. Takavokaalit *u* ja *o* eivät aina ole pyöreitä spontaanissa puheessa, koska pyöreille takavokaaleille ei suomessa ole vastakohtaa laveat takavokaalit. Vokaaleissa *y* ja *ö* pyöreys toteutuu omassa selvässä artikulaatiossani huulten toisiaan lähentymisellä (vrt. piirre *compression*; Ladefoged & Maddieson 1996, s. 295), kun taas vokaaleissa *u* ja *o* pyöreys saadaan aikaan huulten eteenpäin työntämisellä ja aukon supistamisella (vrt. *protrusion*). Tämä on ymmärrettävää selvässä artikulaatiossa, koska protruusiolla saadaan aikaan ääntöväylän pidennys ja formanttien aleneminen ja siten selvempi ero etu- ja takavokaalien kesken. Vokaalit *y* ja *ö* voidaan silti tuottaa myös protruusiolla. Huuliaukon laveudessakin on aste-eroja, mutta niitä ei ohjelmassa ole huomioitu.

Konsonanttien dominanttien tyyppien artikulaatiopiirteet esitetään taulukossa 2 ja artikulaatiokuvat liitteessä 1. Niiden monentyyppinen allofoninen vaihtelu ei saa ohjelmassa sijaa. Kuvassa 4 esitetään kaksi esimerkkiä äännemallien vertailusta.

Kolmiulotteisuus olisi hyödyllinen etenkin sellaisissa äänteissä kuin *l*, *s* ja *š*. /h/:ta ei voi esittää yhdellä artikulaatiokuvalla, koska sillä ei ole yhtä ainoata tyypillistä ja hallitsevaa artikulaatiota. /h/:ta voidaan kuitenkin luonnehtia yleisemmin näin: sille on tyypillistä glottiksen avautumisliike ja ääntöväylässä ilmenevä, rajavo-



Kuva 3: Suomen vokaalien artikulaatiomallit.

kaalin ahtaimman paikan artikulaatioon mukautuva väylähäly. Sen variantit assimiloituvat voimakkaasti ympäristön vokaalin mukaan. /h/:lla on kaksi pääkontekstia: (1) tavunalkuinen ja (2) tavunloppuinen:

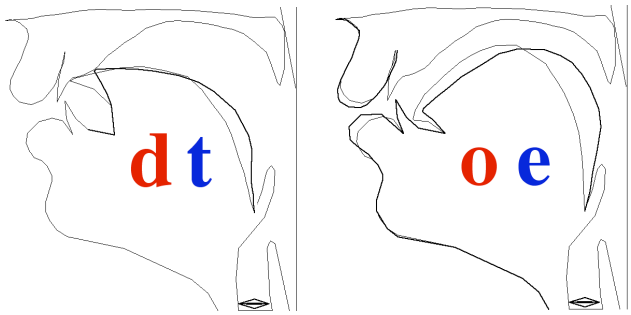
(1) *hame*, *heti*, *home*, *hyvä*; *piha*, *pahe*, *vähän*, *lyhyt*

(2) *pahvi*, *lyhty*, *suihku*, *haahka*.

Relevantilla artikulaatiopaikalla ilmavirran aukko on kuitenkin suppeampi kuin vastaavalla vokaalilla. Aukossa syntyy hankaushälyä. Mikäli vokaali on *y* tai *u*, hankaushäly syntyy huuliaukossa. On tyypillisempää, että /h/:n variantit ovat soinnittomia, ja äänirako pyrkii avautumaan. Mitä enemmän ympäristö on soinnillinen, sitä enemmän [h]-variantti on henkäyssoinnillinen. Henkäyssoinnissa äänihuulet värähtelevät, mutta samalla rustorako on auki. Soinnillinen ympäristö voi olla vokaali (*piha*,

Taulukko 2: Suomen konsonanttien artikulaatiopiirteet.

/K/	sointi	huuliaukko	kieli	artikulaatiopaikka	artikulaatiotapa
p	−	kiinni	neutraali	labiaalinen	klusiili
b	+	kiinni	neutraali	labiaalinen	klusiili
t	−	neutraali	lapa	hammasvallin etuosa	klusiili
d	+	neutraali	kärki	hammasvallin keskiosa	klusiili
k	−	neutraali	selkä	kova ja pehmeä suulaki	klusiili
g	+	neutraali	selkä	kova ja pehmeä suulaki	klusiili
m	+	kiinni	neutraali	labiaalinen	nasaali
n	+	neutraali	kärki	hammasvallin keskiosa	nasaali
ŋ	+	neutraali	takaselkä	pehmeä suulaki	nasaali
l	+	neutraali	kärki	hammasvallin keskiosa	lateraali
r	+	neutraali	lapa	hammasvallin etuosa	tremulantti
f	−	labiodentaa-linen rako	neutraali	labiodentaalinen	frikatiivi
s	−	neutraali	lapa	hammasvallin etuosa	frikatiivi
ʃ	−	neutraali	lapa	hammasvallin etuosa	frikatiivi
ʋ	+	kuten *f*	neutraali	labiodentaalinen	approksimantti
j	+	neutraali	etuselkä	kova suulaki	approksimantti
h	allofonisten vaihteluiden vuoksi dominanttia artikulaatiota ei voi määritellä				



Kuva 4: Kaksi äännemallivertailua. Vasemmalla *d* (vahvennettu viiva) ja *t*. Oikealla *o* ja *e*.

paha) tai soinnillinen konsonantti (*lahja*, *lähde*).

7 Päätelmät ja jatkonäkymät

Kuvatun artikulaatio-ohjelman päätavoite on edistää foneettista tietoisuutta ja (suomen) äänteiden artikulaation oppimista. Fonetiikkakin tyypittelee, abstrahoi, vaikka foneettinen tutkimus osoittaa suurta variaation määrää puhujien, sosiolektien ja puhetyylien mukaan. IPA:n äännesymbolitkin merkitsevät jonkinasteista abstraktiota. Opetusohjelmissa vaihtelua ei voi ainakaan ensi vaiheessa ottaa huomioon. Ääntöväylän puhujakohtainen anatominen vaihtelu on suurta. Kirjallisuudessa julkistuissa röntgenkuvissa kuitenkin usein yksi henkilö on kerran ääntänyt tutkittavan äänteen. Opetusohjelman visuaalinen puhujakin on väistämättä yksilö, ja ohjelman käyttäjän odotetaan mukautuvan tähän. On erittäin stimuloivaa etsiä evidenssiä artikulaatiokonfiguraatioiden todenmukaisuudelle. On myös aiheellista muistaa, että artikulaatioiden tarkoitus on tuottaa kuulon avulla toisistaan erotettavia äänteitä, siis ääntä.

Esiteltävä ohjelma on prototyyppi. Jatkokehittelyn mahdollisuuksia on paljon. Tavoitteena on siirtää ohjelma Java-kielelle. Seuraavat lisäykset ovat mahdollisia: (1) ääniesimerkit, (2) videoklipit edestä ja sivulta, (3) äänteiden luokituspiirteet, (4) keskeiset koartikulaatiovariantit, (5) tyyliltelty huuliaukko edestäpäin, (6) kolmiulotteinen esitystapa ja (7) jatkuvan puheen simulointi (artikulaatiosynteesi). Kohdassa 7 esitetyn tavoitteen toteutus vaatii vähintään jonkinasteisen äänteiden ajoituksen ja siirtymävaiheiden sekä koartikulaation suunnittelua.

Viitteet

BONNOT, Jean-François, BOTHOREL, André & CHEVRIE, Claude 1989: De l'invariance relationelle a la modélisation de la variabilité: Application au cas de groupes de consonnes hétérorganiques. – *Travaux de l'Institut de phonétique de Strasbourg*, **21**: 265–302.

HEIKE, G., PHILLIPP, J. E. & HILGER, S. 1986: Computergraphikdarstellung von Artikulationsbewegungen zur Unterstützung des Artikulationstrainings. – *Sprache—Stimme Gehör*, **10**: 4–8.

HELLWAG, C. F. 1781: *Dissertatio de formatione loquelae*. Tübingen.

HOOLE, Philip & NGUYEN, Noel 1997: Electromagnetic articulography in coarticulation research. – *Forschungsberichte des Instituts für Phonetik und Spachliche Kommunikation der Universität München*, **35**: 177–184.

JONES, Daniel 1917: *An English Pronouncing Dictionary*. London: J. M. Dent & Sons.

JONES, Daniel 1958: *The Pronunciation of English*. Neljäs laitos. Cambridge: University Press. (tarkistettu ja laajennettu 1956; ensimmäinen painos 1909).

LADEFOGED, Peter 1967: *Three Areas of Experimental Phonetics*. London: Oxford University Press.

LADEFOGED, Peter & MADDIESON, Ian 1996: *The Sounds of the World's Languages*. Oxford, UK/Malden, U.S.A.: Blackwell.

MERKEL, C. L. 1866: *Physiologie der menschlichen Sprache (physiologische Laletik)*. Leipzig: Otto Wigand.

MOORE, Christpher, A. 1992: The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images. – *Journal of Speech and Hearing Research*, **35**: 1009–1023.

PANCONCELLI-CALCIA, G. 1961: *3000 Jahre Stimmforschung*. Marburg: N. G. Elwert.

PERKELL, J. S. 1969: *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*. Research Monograph 53. Cambridge, Mass./London, England: The M.I.T. Press.

PERKELL, J. S. 1997: Articulatory processes. – J. Laver & W. J. Hardcastle (toim.), *The Handbook of Phonetic Sciences*. Oxford: Blackwell. 333–370.

POIROT, Jean 1911: *Die Phonetik*. Leipzig: Hirzel.

RICHMOND, Korin, KING, Simon & TAYLOR, Paul 2003: Modelling the uncertainty in recovering articulation from acoustics. – *Computer Speech and Language*, **17**(2–3): 153–172.

RIDOUANE, Rachid 2006: Investigating speech production: A review of some techniques. `http://lpp.univ-paris3.fr/equipe/rachid_ridouane/Ridouane_Investigating.pdf` (24.3.2008).

SCHEIER, M. 1897: *Die Anwendung der Röntgenstrahlen für die Physiologie der Stimme und Sprache*. Deutsche Medizinische Wochenschrift 25.

SOVIJÄRVI, Antti 1938: *Die gehaltenen, geflüsterten und gesungenen Vokale und Nasale der finnischen Sprache*, osa B 44 sarjasta *Annales Academiae Scientiarum Fennicae*.

SOVIJÄRVI, Antti 1963: *Suomen kielen äännekuvasto*. Jyväskylä: Gummerus.

SOVIJÄRVI, Antti 1968: Äänteiden siirtymistä röntgenelokuvien ja spektrogrammien valossa. – *Esitelmät ja pöytäkirjat*. Helsinki: Suomalainen Tiedeakatemia.

SUOMI, Kari, TOIVANEN, Juhani & YLITALO, Riikka 2006: *Fonetiikan ja suomen äänneopin perusteet*. Helsinki: Gaudeamus.

WARD, Ida 1929 [1962]: *The Phonetics of English*. Cambridge: W. Heffer & Sons. First published 1929, entirely revised edition 1939, minor corrections 1948, reprinted 1962.

WÄNGLER, Hans-Heinrich 1974: *Atlas deutscher Sprachlaute*. Berlin: Akademie-Verlag.

WWW-osoitteet:

Magneettinen resonanssikuvantaminen (artikulaatio, ääntöväylä)

Dang Pharos `<http://iipl.jaist.ac.jp/dang-lab/ja/>` (16.3.2008)

Kardinaalivokaalit

`<http://en.wikipedia.org/wiki/Cardinal_vowel>` (15.3.2008)
`<http://www.phonetics.ucla.edu/course/chapter9/cardinal/cardinal.html>` (15.3.2008)

Artikulaatio ja auditiiviset äännemallit (16.3.2008)

Amer. englanti
<http://www.uiowa.edu/acadtech/phonetics/english/frameset.html>

Suomi ja brittienglanti
<http://www.edu.helsinki.fi/fon_demo/fonzie_demo.swf>

Saksan äänteet SpeechTrainer
<http://www.speechtrainer.de/software.htm> (17.3.2008)

IPA:n äänteet
<http://www.chass.utoronto.ca/ danhall/phonetics/sammy.html>

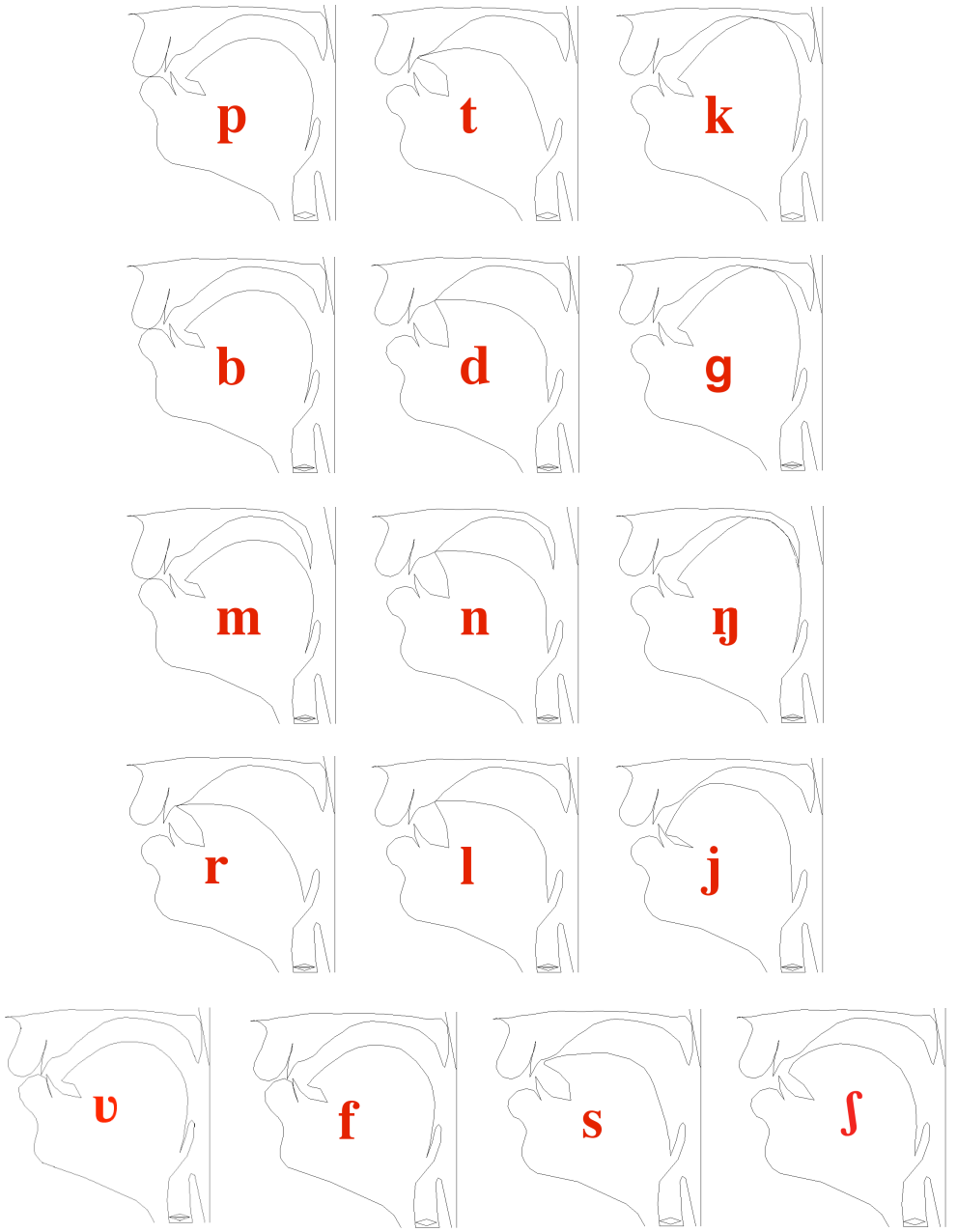
Röntgentietokanta
<http://psyc.queensu.ca/%7Emunhallk/05_database.htm>

Akustiikasta artikulaatioon
<http://www.cstr.ed.ac.uk/downloads/publications/2003/
richmond2003.pdf>

Huulten artikulaatio

<http://citeseer.ist.psu.edu/lee99formant.html> (16.3.2008)

Liite 1: Suomen konsonanttien artikulaatiomallit.



Michael O'Dell & Tommi Nieminen (toim.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 139–146.

Viittomien koartikulaatiosta

Stina Ojala[1], Tapio Salakoski[1] & Olli Aaltonen[2]
[1]Turun yliopisto, [2]Helsingin yliopisto

Tiivistelmä

Puheessa koartikulaation ilmenemismuodot ovat olleet tutkimuskohteena ja -suuntauksena jo pidemmän aikaa. Viittomisen koartikulaatiota on pästy tutkimaan vasta viimeaikaisen teknisen kehityksen kautta. Koartikulaation tutkiminen viittomisessa on sikäli helpompaa, että tutkimuksen kohde on koko ajan näkyvissä — toisin kuin puheessa, jossa tutkimus on kohdistunut suurilta osin akustisessa signaalissa tapahtuvien koartikulaatiomuutosten tutkimiseen. Tässä tutkimuksessa keskitytään sormien liikelaajuuksien ja niiden keskinäisen koordinaation tutkimukseen. Tietoa saadaan myös viittomien rytmityksestä.

Avainsanat: koartikulaatio, viittomakieli

1 Johdanto

Puheessa äänteet eivät ole diskreettejä selvärajaisia segmenttejä vaan puhe on jatkuvaa artikulaatioelinten liikettä, jossa edelliset ja seuraavat äänteet vaikuttavat toisiinsa. Tätä jatkuvaa liikesarjaa äänteestä toiseen nimitetään koartikulaatioksi. Koartikulaatio on tapa tehdä artikulaatioelinten liikkeiden tuottaminen mahdollisimman helpoksi. Tällöin noudatetaan *artikulaation ekonomisuus* -periaatetta (mm. Lindblom 1981, Shariatmadari 2006), mikä helpottaa ja nopeuttaa puhumista.

Ääninäytenauhoituksia tehtäessä koartikulaation vaikutusta puheeseen pyritään toisinaan minimoimaan käyttämällä ns. sana- ja lausekonteksteja niin, että eri äänteet ovat aina samassa äänneympäristössä. Tämä käytäntö helpottaa koehenkilöiden ja yksittäisten toistojen välistä vertailua, sillä tällöin kaikki koehenkilöt ääntävät jokaisessa toistossa kyseiset äänteet samassa ympäristössä. Koartikulaation kautta saadaan lisätietoa siitä, miten eri äänteet edustuvat eri äänneympäristöissä. Koartikulaatiota puheessa tutkitaan pääosin akustiikan kautta, koska artikulaatioelimistön liikkeet eivät ole näkyvissä. Viittomisen tutkimuksessa tilanne on toinen: artikulaatioelinten liikkeet ovat suoraan näkyvissä ja näin ollen koartikulaation tutkiminen on siltä osin suoraviivaisempaa.

Koartikulaatiolla on samanlainen tehtävä viitottaessa kuin puhuttaessakin: viestin välittämisen helpottaminen ja viestin tiivistäminen. Viitottaessa viestin tiivistämisen tärkeys korostuu, sillä kädet ovat hitaampia artikulaattoreita kuin puhe-elimet:

puhenopeus on suomen kielessä 90–160 sanaa minuutissa, kun taas tämän tutkimuksen puitteissa kerätyn aineiston perusteella viitottaessa nopeus on noin 20–30 viittomaa minuutissa. Viittomisessa koartikulaatioon osallistuvat molemmat kädet erikseen minkä lisäksi kädet yhdessä muodostavat koartikulaatioyksikön. Tällöin koartikulaatiotutkimuksessa tutkitaan kolmea yhtäaikaista koartikulaatioyksikköä: oikea käsi, vasen käsi ja molempien käsien vuorovaikutus. Molempien käsien sisäinen koartikulaatio sormien välillä muodostaa alemman tason koartikulaatioyksikön kun taas molempien käsien vuorovaikutus on koartikulaation tutkimuksen kannalta ylemmän hierarkiatason ilmiö. Koartikulaatiota on myös kasvoissa, mutta tässä tutkimuksessa ilmeet ja kasvojen eleet on rajattu tutkimusalueen ulkopuolelle. Yleisesti puhutaan manuaalisesta ja ei-manuaalisesta koartikulaatiosta. Tämän tutkimuksen puitteissa kuvataan manuaalista koartikulaatiota.

Viittomien tutkimuksessa koartikulaation kontrollointi tapahtuu tehtävänannon kautta. Pohjapiirros ja karttatehtävät toimivat analogisesti lausekontekstien kanssa. Tehtävänanto on viittomien tutkimuksessa ainoa tapa kontrolloida tuotettua viestiä, koska viittomakielistä ei ole olemassa kirjoitettua muotoa. Tässä tutkimuksessa on kerätty sekä viittomakielistä lukupuhuntaa (pohjapiirros ja karttatehtävä) että spontaania kerrontaa.

Viittomien koartikulaation tutkimus alkoi sormittamisen (*fingerspelling*) tutkimuksesta. Sormittaminen tarkoittaa tekstin muuttamista näkyväksi sormiaakkosten avulla. Se on koartikulaatioltaan hyvin rajattua: siinä on mukana vain toinen käsi ja käsi ei yleensä liiku paljonkaan. Silti koartikulaation vaikutus on huomattava myös sormittamisessa (Wilcox 1992). Aivan viime aikoina on saatu ensimmäiset tutkimukset viittomisen koartikulaatiosta: koartikulaatio vaikuttaa viittomisessa aivan samoin kuin puheessa — helpottaen artikulaatiota sekä artikulaatiopaikkojen että käsimuotojen osalta (Lindblom *et al.* 2006, Mauk 2003, Ann 1996). Myös viittomisessa näyttäisi toteutuvan täten *artikulaation ekonomisuus* -periaate.

Viittomisen artikulaatioon vaikuttaa suuresti käsien ja käsimuotojen osalta erityisesti kämmenen ja sormien anatomia ja fysiologia. Viittomien artikulaatiota tutkittaessa artikulaation ekonomisuus on pitkälti riippuvainen siitä, millaiset ulottuvuudet ja liikelaajuudet ovat mahdollisia käden ja sormien osalta. Fysiologisten tutkimus-

ten perusteella voidaan sanoa, että peukalo on käden sormista kaikkein liikkuvin ja nimettömän sormen liikeradat ovat kaikkein rajoittuneimpia (Ann 1996).

2 Materiaali ja metodit

Tutkimusaineisto kerättiin osana laajempaa tutkimusta, jossa kerätään tietoa suomalaisen viittomakielen tuottamisesta ja havaitsemisesta sekä niihin liittyvistä prosesseista. Aineisto koostui koehenkilöiden viittomista vastauksista testikysymyksiin, joissa pyydettiin viittomaan vastaukset kysymysten, tehtävänantojen ja vapaan kerronnan perusteella. Koetilanne oli analoginen puheentutkimuksen ääninäytteille, jossa tehtävänantona ovat erilaiset kontekstilauseet (lukupuhunta) ja vapaa kerronta (spontaani tilanne; informantteja pyydettiin kertomaan vapaasti mitä he tekivät viime lomallaan tai vaihtoehtoisesti informantilla oli mahdollisuus kertoa pieni, muutaman minuutin kestävä tarina haluamastaan aiheesta). Tuotokset kerättiin digitaaliselle videonauhalle, jotka syötettiin tietokoneelle tarkempaa analyysiä varten.

Tähän osatutkimukseen otettiin materiaaliksi yhden koehenkilön tuotos TÄSSÄ ASUNTO *Tässä on asunto, joka on suorakaiteen muotoinen*[1]. Viitottu lause sisältää yhteensä 3 viittomaa ja 6 rytmijaksoa. Rytmijaksot määriteltiin visuaalisesti viittomien sisältämien hidastumisten ja nopeutumisten vuorotteluista. Koartikulaation analyysi tehtiin kuva kuvalta, siis 42 ms:n välein. Jokaisesta kuvasta määriteltiin koartikulaatiopisteet, joiden perusteella molempien käsien liikkeitä ajassa tutkittiin.

Koartikulaatiopisteitä määriteltiin 10 molemmista käsistä, jotta saadaan mahdollisimman tarkka tieto siitä, miten eri käsimuodot edustuvat jatkuvassa viittomisessa. Näissä alustavissa tuloksissa on mukana 6 koartikulaatiopistettä molemmista käsistä. Muita pisteitä käytetään pääosin käden orientaation määrittämiseen sikäli kuin se on mahdollista. Käden orientaatiolla tässä tutkimuksessa tarkoitetaan kämmenpohjan suuntaa suhteessa viittojan kehoon. Koartikulaatiopisteiden pikselikoordinaatit kirjattiin matriisiin, jonka perusteella piirrettiin koartikulaation kulku aikapisteittäin kolmiulotteiseen kuvaajaan sekä yksittäisten koartikulaatiopisteiden liikenopeudet kaksiulotteiseen kuvaajaan. Liikenopeuden mittarina käytettiin suuretta pikseliä kuvaa kohden, eli kaksiulotteisista kuvista laskettiin kunkin koartikulaatiopisteen muutos pikseleinä kyseisestä kuvasta seuraavaan. Tämän perusteella laskettiin myös muutosnopeuden mediaani, joka näkyy kuvaajassa paksuna yhtenäisenä vaakasuorana viivana. Molemmat kuvaajat tehtiin Jyri Paakkulaisen (Informaatioteknologia/TY) kirjoittamien Matlab-skriptien avulla.

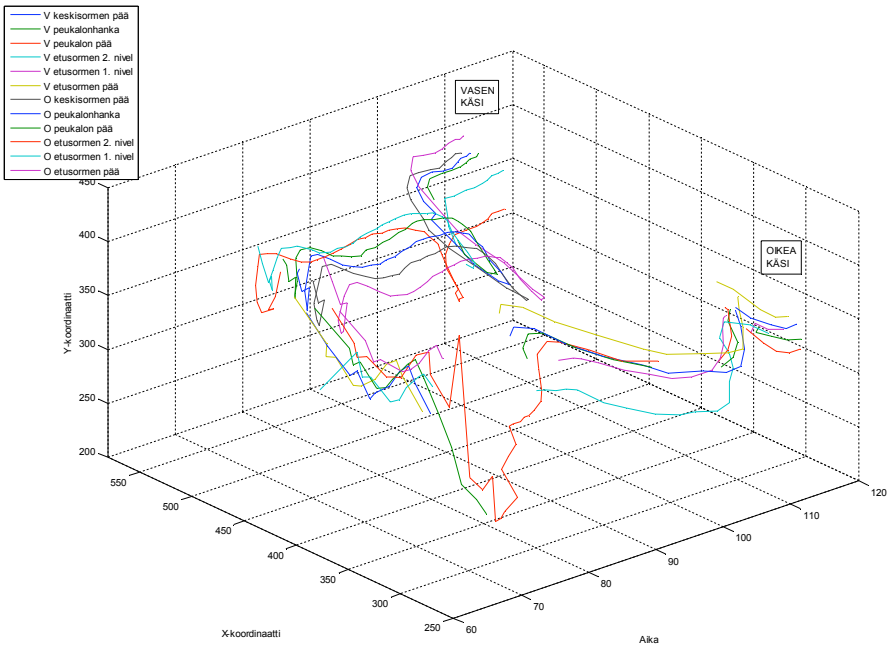
[1] Yleistä käytäntöä viittomien transkriptiosta ei ole, joten tässä artikkelissa käytetään viittomien merkityksen glossausta, joka on yleinen käytäntö Suomessa. Glossauksessa viittomat kirjoitetaan kapiteeleilla ja viittomien merkitykset suomennettuina kursiivilla.

Kuva 1: Koartikulaatiopisteet koko tutkimuksessa (tummat ja vaaleat timantit) ja tässä osatutkimuksessa (vaaleat timantit). Kyseiset pisteet on haettu molempien käsien osalta.

3 Tulokset

Viittomisen koartikulaatiota voi seurata sekä käsien keskinäisen koordinaation kannalta (interkoartikulaatio) että yksittäisen käden kannalta. Interkoartikulaatio näyttäisi vaikuttavan niin, että kumpikin käsi liikkuu nopeammin käsien ollessa erillään toisistaan kun taas käsien ollessa lähekkäin käden liikkeet hidastuvat. Suurilta osin kaikki sormet liikkuvat tässä materiaalissa samalla nopeudella ja samanlaajuisesti, mutta molempien etusormien liikkeet ovat laajemmat ja nopeammat kuin muiden sormien liikkeet. Etusormien liikkeitä verrattaessa oikean käden etusormen liikkeet ovat hieman laajempia ja nopeampia kuin vasemman käden etusormen liikkeet. Sama kätisyysefekti on huomattavissa peukalon osalta, mutta muiden sormien osalta kyseistä efektiä ei ole löydettävissä. Muiden kuin etusormien liikelaajuudet ja nopeudet pienenevät pikkusormeen päin mentäessä kuitenkin niin, että pikkusormen liikelaajuus ja nopeus ovat suuremmat kuin nimettömän. Toisin sanoen nimetön sormi on kunkin käden vähiten liikkuva sormi tässä aineistossa. Peukalon liikelaajuus on isompi kuin etusormen, mutta toisaalta peukalon liikkeet ovat hitaampia kuin etusormen. Molempien käsien sormet näyttävät liikkuvan hyvin samanaikaisesti ja liikelaajuudet näyttävät samankaltaisilta kummankin käden osalta.

Liikkeen muutosnopeutta seurattiin myös aikapisteittäin. Kuvaajasta näkyy sekä yksittäisten sormien liikkeiden nopeudet että yksittäisten sormien liikkeiden nopeuden muutokset suhteessa toisiinsa. Suurimman osan aikaa tässä materiaalissa sormien välinen hetkittäinen nopeusero on niin pientä, että kuvaajaan riittäisi pelkästään yksittäisen sormen liikenopeuden kuvaus, mutta poikkeuksiakin on. Tämän hetkisten

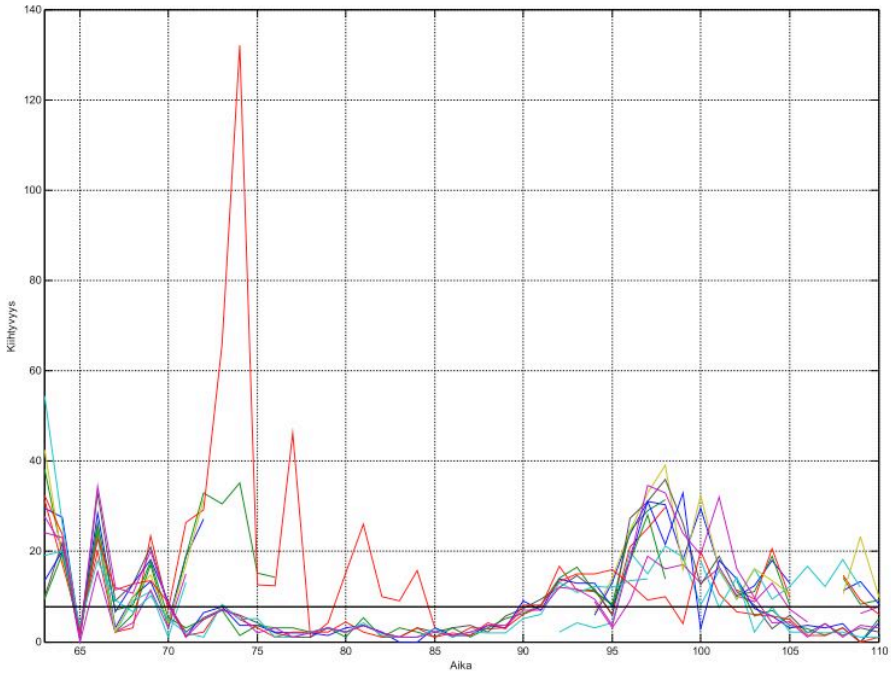


Kuva 2: Kaikkien sormien yhteinen kolmiulotteinen liikekuvaaja aikapisteittäin. Koartikulaatiopisteiden liikekäyrän epäjatkuvuudet johtuvat siitä, että kyseinen koartikulaatiopiste ei tuolloin ole ollut näkyvissä kameralle.

alustavien tulosten perusteella näyttäisi siltä, että yksittäiset vertikaaliset liikkeet ovat nopeampia kuin horisontaalitasossa liikkuvat liikkeet, mutta havainto kaipaa vielä lisäselvityksiä. Kuvaajassa näkyy myös pyrkimys liikkeiden pitämiseen mahdollisimman hitaina, tosin materiaalin vähyys saattaa vääristää tuloksia.

4 Yhteenveto

Liikkeiden hidastumisten ja nopeutumisen vuorottaisuus näyttäisi samanlaiselta prosessilta kuin puheessakin, siinäkin on rytmejä, jotka limittyvät toisiinsa eri tasoilla. Näiden rytmien perusteet ovat ainakin osittain kielikohtaisia (mm. O'Dell & Nieminen 2001). Puheen rytmi saavutetaan artikulaatioliikkeiden koordinaatiolla ja konsonanttien ja vokaalien vuorottelun mahdollistuminen puheessa saattaa olla jopa yksi puheen evoluution kulmakivistä (MacNeilage 1998). Tässä alustavassa tutkimuksessa on keskitytty viittomisen alemman hierarkiatason liikkeiden muutosnopeuksien tutkimukseen. Aiemmin viittomisen liikkeiden nopeutumisten ja hidastumisten vuorottelua on tutkinut mm. Loomis työryhmineen (Loomis *et al.* 1983). Viittomisen liikkeet näyttäisivät rytmillisesti toteuttavan samanlaista syklistä muotoa kuin puheen ar-



Kuva 3: Kaikkien sormien yhteinen liikenopeuskuvaaja aikapisteittäin.

tikulaatioliikkeet. Viittomisen ylemmän hierarkiatason perusrytmi toteutuu viittoman sisäisten ja viittomien välisten liikkeiden ja pysähtymisten vuorottelussa (Liddell & Johnson 1989, Kita *et al.* 1998).

Tämän aineiston perusteella etusormi näyttää olevan suurelta osin määräävänä tekijänä viittomien liikelaajuuden ja nopeuden suhteen. Muut sormet näyttävät seuraavan etusormen liikkeitä mutta suppeammin. Etusormella on erityinen asema muihin sormiin nähden: osoitusten tekeminen. Osoituksella on tärkeä osa sekä puhetta että viittomia — se on kätevä tapa viitata näkyvissä olevaan asiaan tai esineeseen. (Osoituksista ks. esim. Corballis 2002.) Peukalo tosin näyttää liikkuvan itsenäisemmin muihin sormiin nähden. Peukalon itsenäinen liikkuminen osoittaa, että viittomisessa pyritään hyödyntämään yksittäisten sormien liikkeitä niin laajasti kuin mahdollista ja toisaalta taas nimettömän sormen rajoittuneimmat liikelaajuudet osoittavat, että viittomien tuottamisessa pyritään välttämään sitä, mikä artikulatorisesti on vaikeaa. Viittominen kommunikaatiotapana ottaa huomioon käden fysiologiset rajoitukset ja toimii niiden määrittämissä puitteissa.

Vielä tutkimattomia kysymyksiä viittomakielen koartikulaatiosta on paljon, esim. miten viittomisnopeus vaikuttaa interartikulaatioon; mikä on koartikulaation ja interartikulaation suhde. Tämä tutkimus on osaltaan mukana jatkamassa koartikulaation

tutkimuksen laajentamista viittomakielten tutkimuksen puolelle.

Kiitokset

Kiitokset Jyri Paakkulaiselle (Informaatioteknologia/TY) Matlab-skriptin kirjoittamisesta, joka mahdollisti koartikulaatiopisteiden kuvaamisen ajassa sekä liikenopeuksien laskemisen.

Viitteet

ANN, Jean 1996: On the relation between ease of articulation and frequency of occurrence of handshapes in two sign languages. – *Lingua*, **98**:19–41.

CORBALLIS, Michael C. 2002: *From Hand to Mouth*. Princeton: Princeton University Press.

KITA, Sotaro, VAN GIJN, Ingeborg & VAN DER HULST, Harry 1998: Movement phases in signs and co-speech gestures, and their transcription by human coders: Lecture notes in computer science. – *Proceedings of the Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop, Bielefeld, Germany, September 1997*. Heidelberg: Springer Verlag.

LIDDELL, Scott K. & JOHNSON, Robert E. 1989: American Sign Language: The phonological base. – *Sign Language Studies*, **64**:195–277.

LINDBLOM, Björn 1981: Economy of speech gestures. – Peter MacNeilage (toim.), *The Production of Speech*. New York: Springer Verlag. 217–246.

LINDBLOM, Björn, MAUK, Claude & MOON, Seung-Jae 2006: Dynamic specification in the production of speech and sign. – Pierre L. Divenyi, Steven Greenberg & Georg Meyer (toim.), *Dynamics of Speech Production and Perception*, osa 374 sarjasta *NATO Science Series: Life and Behavioural Sciences*.

LOOMIS, Jeffrey, POIZNER, Howard, BELLUGI, Ursula, BLAKEMORE, Alynn & HOLLERBACH, John 1983: Computer graphic modelling of American Sign Language. – *Computer Graphics*, **17**(3):105–114.

MACNEILAGE, Peter 1998: The frame/content theory of evolution of speech production. – *Behavioral and Brain Sciences*, **21**:499–511.

MAUK, Claude 2003: *Undershoot in Two Modalities: Evidence from Fast Speech and Fast Signing*. University of Texas, Austin.

O'DELL, Michael & NIEMINEN, Tommi 2001: Speech rhythms as cyclical activity. – S. Ojala & J. Tuomainen (toim.), *21. Fonetiikan päivät Turku 4.–5.1.2001*, Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja / Publications of the Department of Finnish and General Linguistics of the University of Turku 67. 159–168.

SHARIATMADARI, David 2006: Sounds difficult? Why phonological theory needs 'ease of articulation'. – *SOAS Working Papers in Linguistics*, **14**:207–226.

WILCOX, Sherman 1992: *Phonetics of Fingerspelling*. Studies in speech pathology and clinical linguistics 4. John Benjamins.

Michael O'Dell & Tommi Nieminen (toim.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 147–155.

Puheen perustaajuusjakaumat: Alustavia tuloksia

Mietta Lennes[1], Daniel Aalto[2] & Pertti Palo[2]
[1]Helsingin yliopisto, [2]Teknillinen korkeakoulu

Abstract

The typical pitch in the speech of a particular speaker is often described with basic statistical parameters such as the mean, the median or the standard deviation of the fundamental frequency (F0). The distribution of different F0 values can also be illustrated with a histogram. However, the exact shapes of F0 distributions for different speakers have not been thoroughly studied or compared, although they might provide useful information for, e.g., forensic phonetics. This study is a first attempt to compare the shapes of pitch distributions between different speakers, speaking styles and languages. Differences between spontaneous and read-aloud Finnish are investigated and statistically analysed. In addition, the pitch distributions of speakers of Finnish, Russian and Dutch are visually inspected and compared.

Avainsanat: perustaajuusjakauma, F0, sävelkorkeus, puhujantunnistus, kielentunnistus

1 Johdanto

Puheen perustaajuuden eli F0:n jakauma kuvastaa sitä, kuinka paljon puhuja käyttää puheessaan eri sävelkorkeuksia. Jakaumaan vaikuttavat osaltaan puhujan fysiologiset ominaisuudet ja hänen yksilölliset tapansa käyttää äänihuuliaan. F0-jakaumien analyysi onkin ollut erityisesti forensisen fonetiikan mielenkiinnon kohteena (mm. Braun 1995, ks. myös yhteenveto perustaajuuden jakaumaa ja forensisia sovelluksia koskevasta kirjallisuudesta, Lindh 2006). Yksilöllisten piirteiden ja erilaisten muuttuvien ei-kielellisten ja kielellisten tekijöiden rooleja F0-jakauman muodostumisessa ei kuitenkaan tarkasti tunneta.

Voidaan olettaa, että puhujat käyttävät puheessaan eniten sellaisia sävelkorkeuksia, joiden tuottaminen vaatii heiltä vähiten energiaa. Toisaalta tiedetään, että sana- tai lausepainollisten tavujen tai muiden kuulijan havainnon kannalta prominenttien puheen osien kohdalla esiintyy usein perustaajuushuippu (mm. Gussenhoven *et al.*

1997). Puhujat siis todennäköisesti hyödyntävät äänialaansa tehokkaasti tuottaessaan erilaisia prosodisia kuvioita.

Puheen F0-jakaumia koskevissa tutkimuksissa on raportoitu useimmiten vain tärkeimmät tilastolliset tunnusluvut kuten keskiarvo, hajonta tai mediaani. Jakaumien muoto on sen sijaan jätetty vähemmälle huomiolle. Perustaajuusjakaumien on tosin osoitettu olevan oikealle eli positiivisesti vinoja. Pääosa puheesta siis painottuu puhujan modaalirekisterin matalille taajuuksille.

Tässä tutkimuksessa tarkastellaan F0-jakaumien muotoa eri puhujilla ja eri puhetyyleissä. F0-jakaumia vertaillaan suomenkielisessä vapaassa keskustelupuheessa ja ääneen luetussa puheessa ja selvitetään, miten χ^2-testi soveltuu jakaumien muodon vertailuun. Lisäksi tutkitaan alustavasti F0-jakaumien mahdollisia eroja suomen-, venäjän- ja hollanninkielisen spontaanin puheen välillä.

2 Menetelmät

2.1 Aineisto

Tässä työssä on käytetty suomen-, venäjän- ja hollanninkielisiä puhekorpuksia. Suomen puhujat (7 naista ja 5 miestä) olivat äänityshetkellä iältään 21–42-vuotiaita ja he olivat asuneet koko ikänsä Helsingissä tai Espoossa. Suomenkielisiin naispuhujiin viitataan jatkossa kirjaimella F ja suomenkielisiin miespuhujiin kirjaimella M. Venäjänkieliset puhujat (2 naista ja 5 miestä) olivat 20–50-vuotiaita ja kotoisin Pietarin seudulta. Hollannin puhujat (4 naista ja 4 miestä) olivat 15–66-vuotiaita amsterdamilaisia (*IFA-korpus*, ks. van Son *et al.* 2001). Kaikki tätä tutkimusta varten analysoidut puhujat olivat yksikielisiä natiiveja ja joko korkeakoulututkinnon suorittaneita tai korkeakouluopiskelijoita.

Suomenkielisen korpuksen äänitykset suoritettiin joko Teknillisen korkeakoulun akustiikan ja äänenkäsittelytekniikan laboratorion kaiuttomassa huoneessa tai ammattikäyttöön tarkoitetussa äänitysstudiossa. Jokaista puhujaa äänitettiin kaksi kertaa. Ensimmäiseen äänitykseen puhujat osallistuivat pareittain, kukin hyvän ystävänsä kanssa, ja kaverusten annettiin keskustella vapaasti keskenään 40–60 minuutin ajan. Toisella kerralla jokaista puhujaa äänitettiin yksin ja hänen tehtävänään oli lukea tekstiä ääneen. Kunkin puhujan puhe äänitettiin AKG HSC-200-kuuloke-mikrofoniyhdistelmällä omalle raidalleen DAT-nauhalle (kaiuttomassa huoneessa) tai suoraan studiolaitteiston tietokoneelle (äänitysstudiossa). Puheääni näytteistettiin 16-bittisenä ja 22,05 kHz:n taajuudella.

Venäjän ja hollannin puhekorpukset äänitettiin myös studio-olosuhteissa. Tässä tutkimuksessa käytetään venäjän ja hollannin osalta vain spontaanin puheen aineistoja. Näissä äänitystilanne oli kuitenkin haastattelumaisempi kuin suomen vapaan keskustelupuheen aineistossa.

2.2 Perustaajuusanalyysi

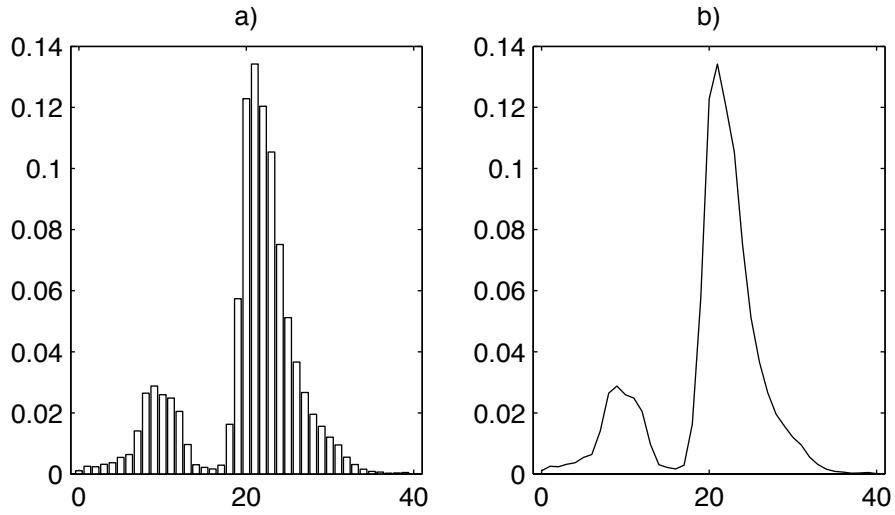
Perustaajuusanalyysi suoritettiin Praat-ohjelman (Boersma & Weenink 2007) standardialgoritmilla. Perustaajuus laskettiin 20 ms aika-askelilla kunkin puhujan kaikista jatkuvista puhunnoksista. Tauot jätettiin analyysin ulkopuolelle. Yksittäinen puhunnos saattoi siten sisältää mielivaltaisen määrän sanoja, lauseita, epäröintiääniä ja niin edelleen, kunhan puhuja jatkuvasti tuotti puhetta. Suomenkielisestä spontaanin keskustelupuheen aineistosta saatiin tällä tavalla kokoon 42 165–73 332 F0-mittauspistettä puhujaa kohti, ääneen luetusta suomenkielisestä puheesta 37 927–73 043 mittauspistettä, venäjänkielisestä spontaanin puheen aineistosta 4 214–18 419 mittauspistettä sekä hollanninkielisestä spontaanipuheesta 5 163–63 812 mittausta puhujaa kohti.

Jotta F0-mittauksen tuottamat arvot olisivat tulkittavissa, on analyysin parametrit valittava järkevästi. Normaalitapauksessa on säädettävä erityisesti F0-analyysin taajuusrajoja tarpeen mukaan. Jos rajat on asetettu liian väljiksi, tuottaa analyysialgoritmi helposti ns. ”oktaavivirheitä” l. taajuusarvoja, jotka sijaitsevat oktaavia alempana tai ylempänä kuin kuulohavainnon ja perustaajuuskäyrän suunnan perusteella olisi ennakoitavissa. Alustavien kokeilujen perusteella jokaiselle puhujalle yksilöllisesti säädetyt taajuusrajat kuitenkin vaikuttivat heidän puheestaan syntyviin perustaajuusjakaumiin yllättävän vähän. Lisäksi alaspäin suuntautuvat oktaavivirheet ovat sinänsä kiinnostavia, koska ne kokemuksemme mukaan liittyvät yleensä puhujan narinaiseen äänenlaatuun, joka taas on mahdollisesti yksilöllinen tai kieliriippuva puheen piirre. Oktaavivirheitä voidaan siis hyödyntää arvioitaessa modaalisen vs. narinaisen äänenlaadun esiintymistä tietyllä puhujalla. Tässä tutkimuksessa päätettiinkin käyttää kaikkien puhujien kohdalla täsmälleen samoja analyysiparametrejä, jotta analyysi olisi mahdollisimman yksinkertainen ja objektiivinen ja jotta eri puhujien perustaajuusjakaumat olisivat paremmin vertailtavissa. Perustaajuusminimiksi valittiin 50 Hz ja -maksimiksi 500 Hz.

2.3 Tilastollinen analyysi

Usein F0-jakaumat esitetään histogrammeina. Tätä tarkoitusta varten kunkin puhujan perustaajuusaineistosta muodostettiin aluksi luokitellut frekvenssijakaumat. Luokan leveytenä käytettiin yhtä puolisävelaskelta. Esimerkki tällaisesta F0-histogrammista näkyy kuvassa 1 a. Eri puhujien F0-jakaumien muodon visuaalinen vertailu on kuitenkin helpompaa, jos yksittäiset jakaumat esitetään jatkuvina käyrinä kuten kuvassa 1 b.

Kuvassa 2 on esitetty kaikkien suomenkielisten puhujien F0-jakaumat jatkuvina käyrinä vapaan keskustelupuheen ja ääneen luetun puheen osalta. Silmämääräisesti tarkasteltuna lähes kaikilla puhujilla on ero näiden kahden puhetyylin F0-jakaumien välillä.



Kuva 1: Sama F0-jakauma esitettynä histogrammina a) ja jatkuvana käyränä b).

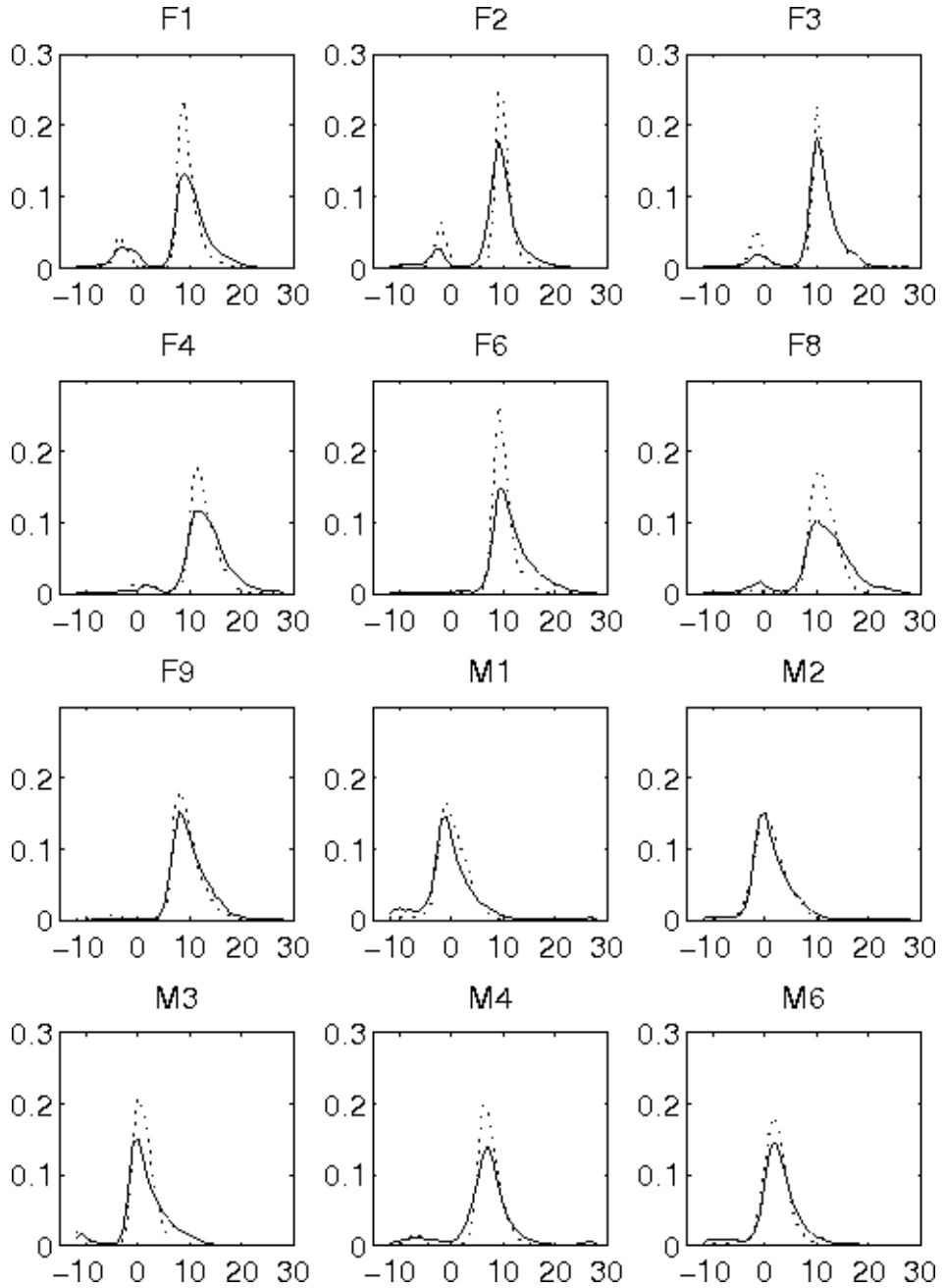
Kaikki tässä tutkimuksessa tarkastellut F0-jakaumat sisältävät selvästi yhden korkean huipun, jota voidaan käyttää kiintopisteenä jakaumien muodon vertaamisessa. Siksi jakaumat normalisoitiin moodin suhteen siirtämällä kunkin jakauman huippu puolisävelasteikon nollakohtaan. Kuvassa 3 on esitetty suomen, venäjän ja hollannin puhujien F0-jakaumat tällä tavoin siirrettynä.

Suomenkielisten puhujien osalta F0-jakaumia tutkittiin myös tilastollisesti. Aluksi verrattiin spontaani- ja lukupuheen F0-jakaumien huippujen korkeutta. Vertailu suoritettiin parittaisella Studentin t-testillä, jossa parit muodostuivat kunkin puhujan spontaani- ja lukupuheen moodiluokkien suhteellisista osuuksista.

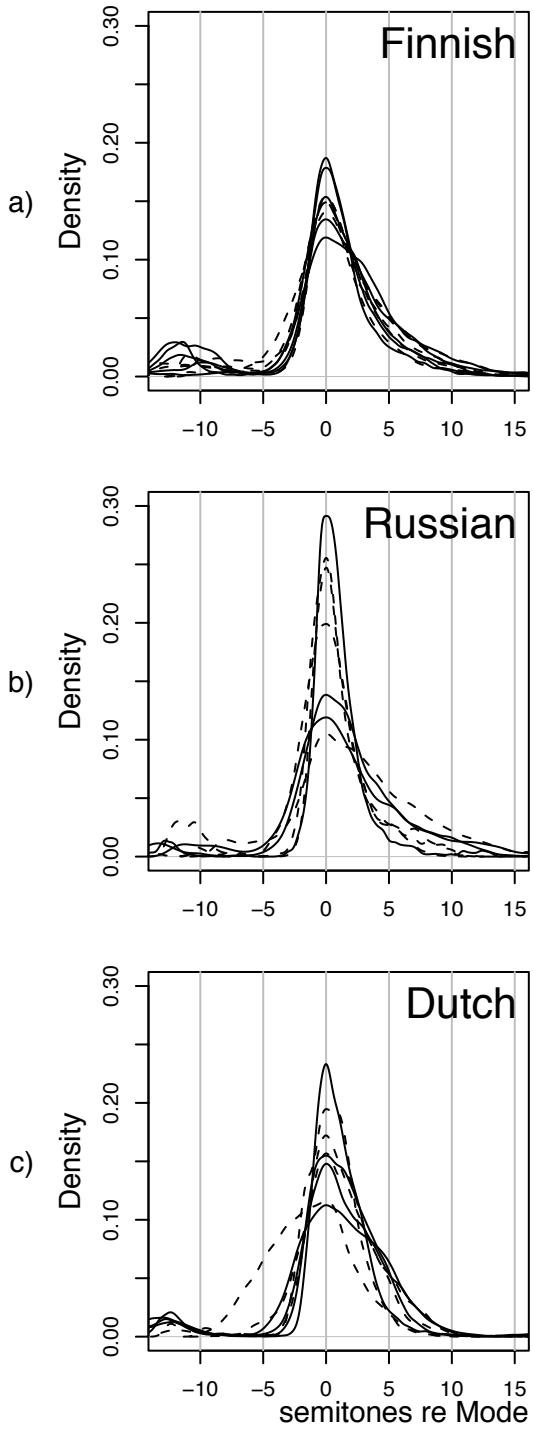
Lisäksi suomalaisten puhujien F0-jakaumien muotoa verrattiin tarkemmin χ^2-homogeenisuustestillä, jolla testattiin kolmea eri nollahypoteesia:

- Kaikkien puhujien moodinormalisoidut F0-jakaumat ovat samanmuotoisia eli peräisin samasta jakaumasta.
- Puhujan spontaanin ja luetun puheen moodinormalisoidut F0-jakaumat ovat peräisin samasta jakaumasta.
- Puhujan F2 parittomien ja parillisten puhunnosten moodinormalisoidut F0-jakaumat ovat peräisin samasta jakaumasta.

Viimeisen nollahypoteesin parittomat ja parilliset puhunnokset valikoitiin puhujan F2 äänitysjärjestyksessä numeroiduista puhunnoksista.



Kuva 2: Seitsemän suomenkielisen nais- ja viiden miespuhujan perustaajuusjakaumat tiheyskäyrinä vapaassa keskustelupuheessa (yhtenäinen viiva) ja ääneen luetussa puheessa (katkoviiva). Vaaka-akseli osoittaa perustaajuuden puolisävelaskeleina 100 Hz:iin nähden. Pystyakseli osoittaa eri sävelkorkeuksien puhujakohtaista todennäköisyyttä.



Kuva 3: Moodinormalisoidut F0-jakaumat suomen-, venäjän- ja hollanninkielisillä puhujilla. Naispuhujien jakaumat on merkitty yhtenäisillä viivoilla, miespuhujien jakaumat katkoviivoilla.

3 Tulokset ja pohdintaa

Kuvassa 2 on puhujien spontaanin ja ääneen luetun puheen perustaajuusjakaumat. Kuvasta voidaan selvästi nähdä, että lukupuheen huippu on keskimäärin korkeampi kuin spontaanin puheen. Tämä havainto vahvistettiin *t*-testillä, jonka *p*-arvo oli 0.00026543. Kuvan 2 perusteella ääneen luetussa puheessa käytetään korkeita taajuuksia vähemmän kuin spontaanissa puheessa.

Kuvan 2 kaikilla muilla puhujilla paitsi naispuhujilla F6 ja F9 ja miespuhujalla M2 on moodin ympärillä sijaitsevan suurimman kasauman alapuolella erotettavissa toinen, pienempi kasauma. Koska tämä matalampi huippu on noin oktaavin (= 12 puolisävelaskeleen) päässä korkeammasta huipusta, se lienee syntynyt aiemmin mainituista ns. oktaavivirheistä, joissa F0-algoritmi on poiminut perustaajuusarvon oktaavia alempaa kuin kuulohavainto tai perustaajuuskäyrän jatkuvuus antaisivat olettaa. Koska alaspäin suuntautuvia oktaavivirheitä tapahtuu usein narinaisen äänenlaadun kohdalla, voidaan ajatella, että perustaajuusjakaumassa erottuva matalampi kumpare kuvastaa suureksi osaksi juuri puhujan taipumusta käyttää narinaa puheessaan. Useimmilla suomalaispuhujilla narinaa näyttää siis jonkin verran esiintyvän sekä vapaassa keskustelussa että ääneen lukiessa. Narinan ajallisesta jakaumasta tai kielellisestä käytöstä nämä jakaumakuvat eivät tietenkään kerro vielä mitään.

Kuvat 2 ja 3 a antaisivat silmämääräisesti olettaa, että ainakin suomenkielisten puhujien F0-jakaumat ovat lähes samanmuotoisia. Tästä huolimatta kaikkia kolmea nollahypoteesia koskevien χ^2-homogeenisuustestien *p*-arvot olivat häviävän pieniä ja siten nollahypoteesit jouduttiin jokaisessa testissä hylkäämään. Näin ollen edes saman puhujan kahtia jaetun aineiston puoliskot eivät testin mukaan olisi peräisin samasta jakaumasta. Voidaan kuitenkin epäillä, että testin tulos johtuu tutkimuksen suuresta otoskoosta, eikä se siis välttämättä tarkoita, ettei jakaumista löytyisi lainkaan samanmuotoisuutta. Homogeenisuustestien tarkoitus on nimittäin yleensä toimia päätöskriteerinä tapauksissa, joissa dataa on käytettävissä vähän.

Kuvassa 3 näkyvät suomen, venäjän ja hollannin puhujien F0-jakaumat puolisävelasteikolla, jonka nollakohta vastaa kunkin puhujan aineistosta laskettua yksilöllistä moodia. Pystyakselin osoittama tiheys ilmaisee eri sävelkorkeuksien todennäköisyyttä kunkin puhujayksilön aineiston sisällä. Kaikkien kolmen kielen jakaumissa on yhteneviä piirteitä, mutta lähempi tarkastelu osoittaa myös eroja. Matalat, luultavasti narinan muodostamat kasaumat, jotka sijaitsevat oktaavia alempana kuin moodi eli kuvan 3 puolisävelasteikon nollapiste, ovat suomen kohdalla jonkin verran selvempiä kuin hollannissa tai venäjässä. Erityisesti venäläisillä naispuhujilla (yhtenäiset viivat kuvassa 3 b) narinaa näyttäisi olevan tuskin lainkaan. Venäjän ja hollannin puhujien F0-jakaumissa näyttäisi olevan enemmän yksilöiden välistä vaihtelua kuin suomenkielisten puhujien jakaumissa etenkin korkeampien taajuuksien suhteellisen käytön osalta. Venäjän ja hollannin kaikilta puhujilta ei kuitenkaan ollut käytettävissä yhtä paljon F0-mittauspisteitä, mikä voi vaikuttaa hieman tiheyskäyriin.

4 Johtopäätökset

Tässä tutkimuksessa tarkasteltiin alustavasti F0-jakaumien muodon mahdollisia puhetyylistä riippuvia eroja suomenkielisessä puheessa sekä kielikohtaisia eroja suomen-, venäjän- ja hollanninkielisessä spontaanissa puheessa. Suurimmalla osalla suomenkielisistä puhujista todettiin selvä ero spontaanin keskustelupuheen ja ääneen luetun puheen perustaajuusjakaumien huippujen korkeuden välillä. Ilmiö voidaan todennäköisimmin selittää ääneen luetun puheen pienemmällä vaihtelulla. Tämä ei ole kuitenkaan ainoa mahdollinen selitys, sillä myös pienellä taajuusalueella tapahtuva nopea F0-vaihtelu voi tuottaa samantyyppisen tuloksen. Jälkimmäinen vaihtoehto ei kuitenkaan vaikuta todennäköiseltä.

Kielten välisistä F0-jakaumien eroista ei vielä tämän tutkimuksen perusteella voida päätellä paljoakaan, sillä puhujia on melko vähän ja aineiston analyysi on kesken. Silmämääräisesti arvioiden suomen ja hollannin perustaajuusjakaumat olisivat muodoltaan keskimäärin lähempänä toisiaan kuin suomen ja venäjän jakaumat. Yksittäisten puhujien välillä esiintyy kuitenkin vaihtelua, eikä mikään yksinkertainen tunnusluku näytä suoraan erottavan kieliä toisistaan.

Tässä tutkimuksessa perustaajuusjakaumien muodon selvittely on päässyt vasta alkuun. Voidaan todeta, että χ^2-testi ei selvästikään sovellu F0-jakaumien vertailuun. Näin ollen tähän tarkoitukseen on löydettävä muita menetelmiä. Tilastollisen testaamisen sijaan luonnollinen seuraava askel on datan käsittely tiedonlouhintamenetelmillä kuten klusteroinnilla ja erityyppisillä komponenttianalyysimenetelmillä. Kiinnostavaa olisi myös liittää F0-jakaumaan F0:n muutosten jakauma, jolloin mukaan tarkasteluun saataisiin myös puheen sävelkulkuun liittyviä dynaamisia ominaisuuksia.

Viitteet

BOERSMA, Paul & WEENINK, David 2007: Praat: doing phonetics by computer (versio 4.6.36) [Tietokoneohjelma]. URL: http://www.praat.org/, haettu 3.11.2007.

BRAUN, A. 1995: Fundamental frequency — how speaker-specific is it? – A. Braun & J.-P. Köster (toim.), *Studies in forensic phonetics*. Trier: Wissenschaftlicher Verlag Trier. 9–23.

GUSSENHOVEN, Carlos, REPP, B. H., RIETVELD, A., RUMP, H. H. & TERKEN, J. 1997: The perceptual prominence of fundamental frequency peaks. – *Journal of the Acoustical Society of America*, **102**(5):3009–3022.

LINDH, Jonas 2006: Preliminary descriptive F0-statistics for young male speakers. – *Papers from FONETIK 2006*, Working Papers 52. Lund University, Centre for Languages and Literature, Department of Linguistics and Phonetics. 89–92.

VAN SON, R. J. J. H., BINNENPOORTE, Diana, VAN DEN HEUVEL, Henk & POLS, Louis C. W. 2001: The IFA corpus: a phonemically segmented Dutch "open source" speech database. – Paul Dalsgaard, Børge Lindberg, Henrik Benner & Zheng hua Tan (toim.), *Proceedings of Eurospeech 2001, Aalborg, Denmark*.

Michael O'Dell & Tommi Nieminen (eds.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 157–165.

Microdurational variations and perception of vowel quality

Einar Meister[1], Lya Meister[1] & Stefan Werner[2]
[1]Tallinn University of Technology, [2]University of Joensuu

Abstract

A study by Kouznetsov (2001) for Russian has reported that native speakers of languages without phonemic quantity opposition also perceive changes in vowel quality when vowel duration is modified. In a recent paper (Werner & Meister 2008) it has been shown that vowel duration can affect vowel quality perception also in quantity languages like Estonian and Finnish—in the case of a vowel quality close to the perceptual boundary shorter stimuli were perceived as closed vowels and longer stimuli as mid vowels. The stimuli used in these two studies involved only stimuli around the /i/–/e/ category boundary while microdurational effects on vowel quality perception at other vowel category boundaries were not tested.

In the present paper we report the results from a further study where synthetic stimuli close to different vowel category boundaries (/y/–/ø/, /e/–/æ/, /u/–/o/, /o/–/ɑ/) with modified durations from 60 to 140 ms were judged by Estonian listeners in ABX/BAX listening tests administered with Praat's multiple forced-choice test facility. As expected, microduration tends to affect listeners' judgements in a systematic way but we also found considerable inter-subject differences and an effect of stimulus presentation order.

Keywords: microprosody, microduration, quantity, intrinsic duration, categorical perception

1 Introduction

Generally accepted microprosodic rules (see e.g. Black 1949, Peterson & Lehiste 1960, Di Cristo 1978), assume that open vowels tend to have lower F0, higher intensity and longer duration than close vowels. Studies like Kouznetsov (2001) for Russian have shown that native speakers of languages without a phonemic quantity opposition also perceive changes in vowel quality when vowel duration is modified. Whether and to what extent vowel duration can affect vowel quality perception also

in quantity languages like Estonian and Finnish where one would expect the long-short opposition to override less crucial "intrinsic" duration effects has been investigated for both Estonian and Finnish exemplars of the /i/–/e/ vowel pair in Werner & Meister (2008); the results show a clear effect of segment duration influencing listeners' vowel perception: the longer the duration of the stimulus, the more often it is identified as the more open of the two vowels. The present study is intended to reconfirm the outcome of this earlier experiment and to extend its empirical base to include additional vowels along the close-open axis. All five vowel pairs examined are: 1. /e/–/æ/, 2. /u/–/o/, 3. /i/–/e/, 4. /y/–/ø/, 5. /o/–/ɑ/.

2 Method

As in the earlier experiment, the main test was preceded by a preparatory test to identify the boundary formant values for distinctive perception of the two vowels in each pair. Interpolating stepwise between prototypical values of the first three formants for each vowel pair we calculated evenly spaced steps through the F1/F2/F3 continua and synthesized the respective eighteen to twenty gradually changing stimuli with a Klatt (1980)-type formant synthesizer, KlattWorks[1]; the formant prototype values were taken from an evaluation of the Estonian BABEL database (Eek & Meister 1998). Listeners had to decide on vowel quality in a binary identification task listening test with non-primed single stimuli. The test results for four native Estonian subjects are shown in Figure 1, with markings for the boundaries and ambiguous transition regions of the various vowel pairs. Accordingly, we selected stimuli 9–12 for our vowel pairs 1–3 and stimuli 8–11 for vowel pairs 4 and 5.

The main perception test then used an ABX setup with isolated vowel stimuli of constant F0 and intensity and with four different formant structures, all from the ambiguous regions identified in the preparatory test (see Figure 2). The stimuli's durations were varied from 60 to 140 ms, in 20 ms steps, and presented each six times (three times in AB and three times in BA contexts) which made for a total number of 120 (five durations × four formant structures × six presentations) decisions for every vowel pair. Like the pre-test, it was administered with Praat's multiple forced-choice test facility using balanced permutation for stimulus randomization. The same four Estonian subjects from the pre-test also took the main test; the question they had to answer—either by clicking in one of two response boxes on the screen or by pressing one of two alternative character keys on the computer keyboard—after listening to every set of three signals was "Does the sound you heard last resemble more the first or the second vowel?".

[1] `http://www.psychology.uiowa.edu/faculty/mcmurray/KlattWorks/`

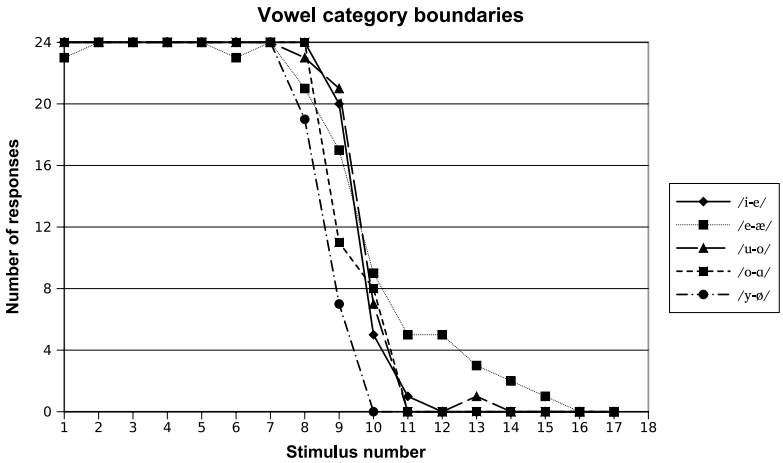


Figure 1: Vowel category boundaries.

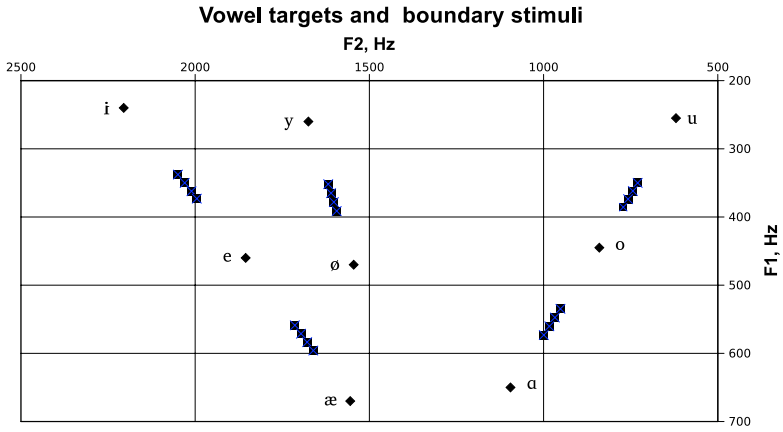


Figure 2: Formants 1 and 2 of target and stimulus vowels.

3 Results

As in the earlier test on /i/ and /e/ alone, there was considerable variation between subjects. Nevertheless the overall picture is clear: perceived vowel openness correlates positively with stimulus duration (adjusted $R^2 = 0.977$, $p < 0.001$), as shown in Figure 3.

While subjects 1 and 2 produced consistent results for all five vowel pairs (see Figures 5 and 6, subjects 3 and 4 produced similar results only for three out of the five pairs, but not for /e/–/æ/ and /o/–/ɑ/ (Figures 7 and 8). And even for vowel pairs with the expected correlation of vowel duration and vowel quality perception, duration-based category boundaries vary among subjects, as can be seen from the comparison of different subjects' response distributions in Figure 9. There are also effects of both stimulus presentation order and formant structure (see Figure 4).

4 Conclusion

Our experiments showed that changes in segment duration tend to predictably affect Estonian listeners' vowel perception: the longer the formant-wise ambiguous stimulus, the more likely it is to be categorized as closer to the more open vowel of a pair. This trend was manifest more or less clearly depending on the test subject and lends support to the hypothesis that the influence of microdurational variation on vowel quality perception represents a universal tendency since we could establish it also in a sample of native speakers of a quantity language which is a far from self-evident outcome.

However, much more work is needed in order to confirm our preliminary results. Apart from the obvious need of larger samples and subjects from more languages, at least the following issues should and will be addressed in further studies:

- subject-specific differences in vowel formant category boundaries,
- adequateness of synthetic stimuli used (sound artefacts, noise ...),
- phonological relevance of the stimuli (isolated sounds versus e.g. words),
- optimization of the test setup (compensation for presentation order effect etc.).

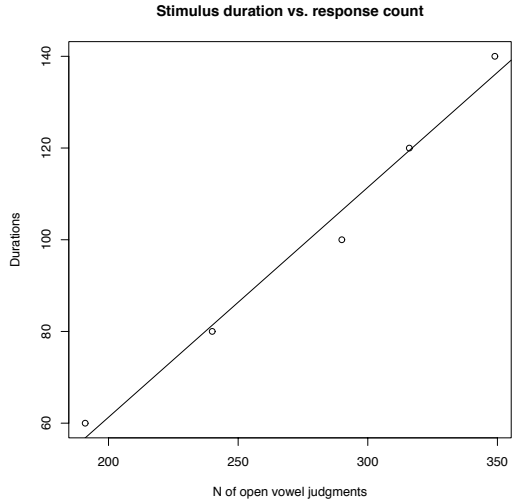


Figure 3: Stimulus duration vs. response count for more open vowel alternatives, all subjects.

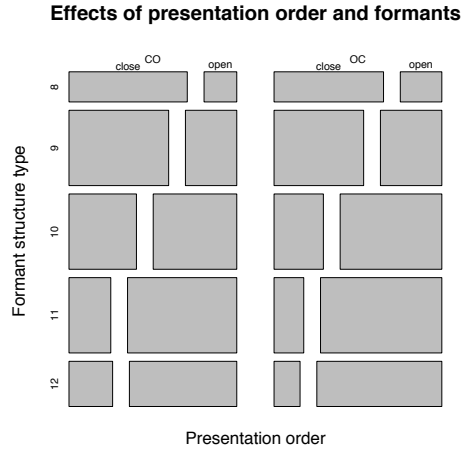


Figure 4: Also stimulus presentation order (CO, close–open vs. OC, open–close) and formant structure (numbers from the stimulus formant transition continuum, see section 2) influence vowel category judgments.

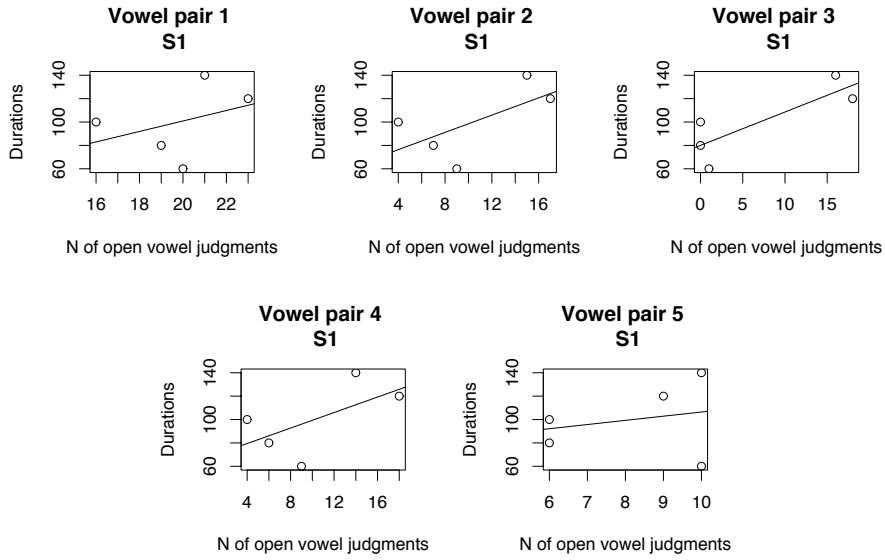


Figure 5: Test results for subject 1.

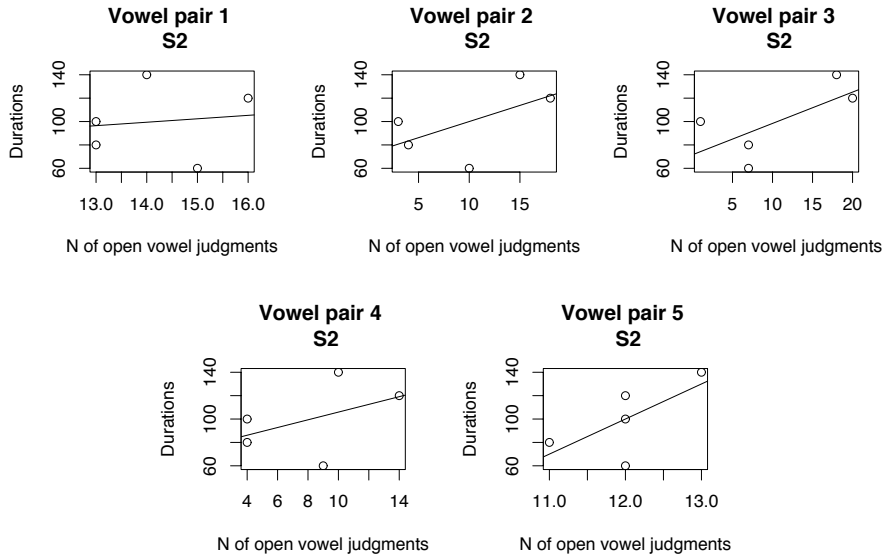


Figure 6: Test results for subject 2.

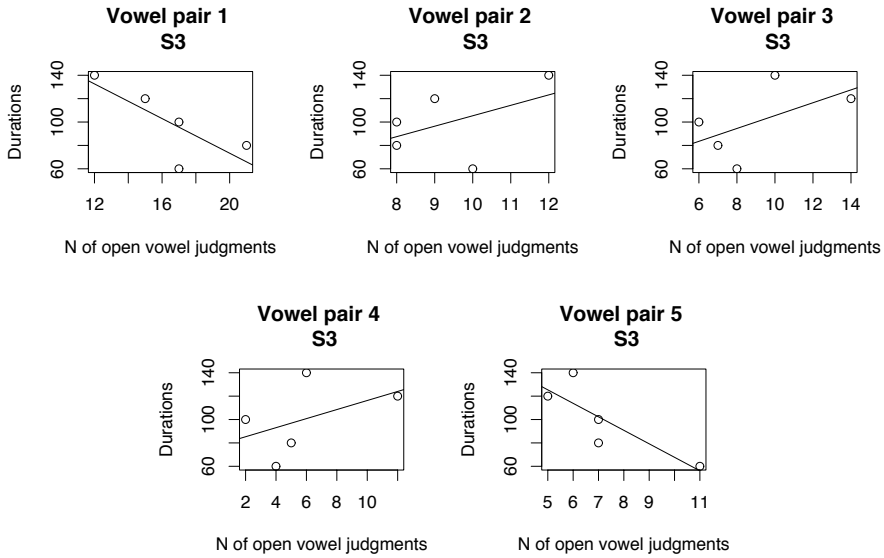


Figure 7: Test results for subject 3.

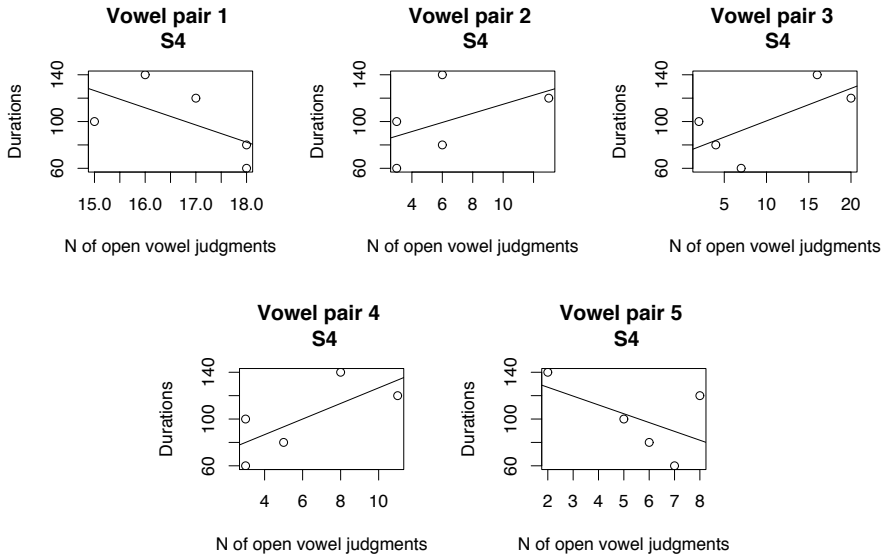


Figure 8: Test results for subject 4.

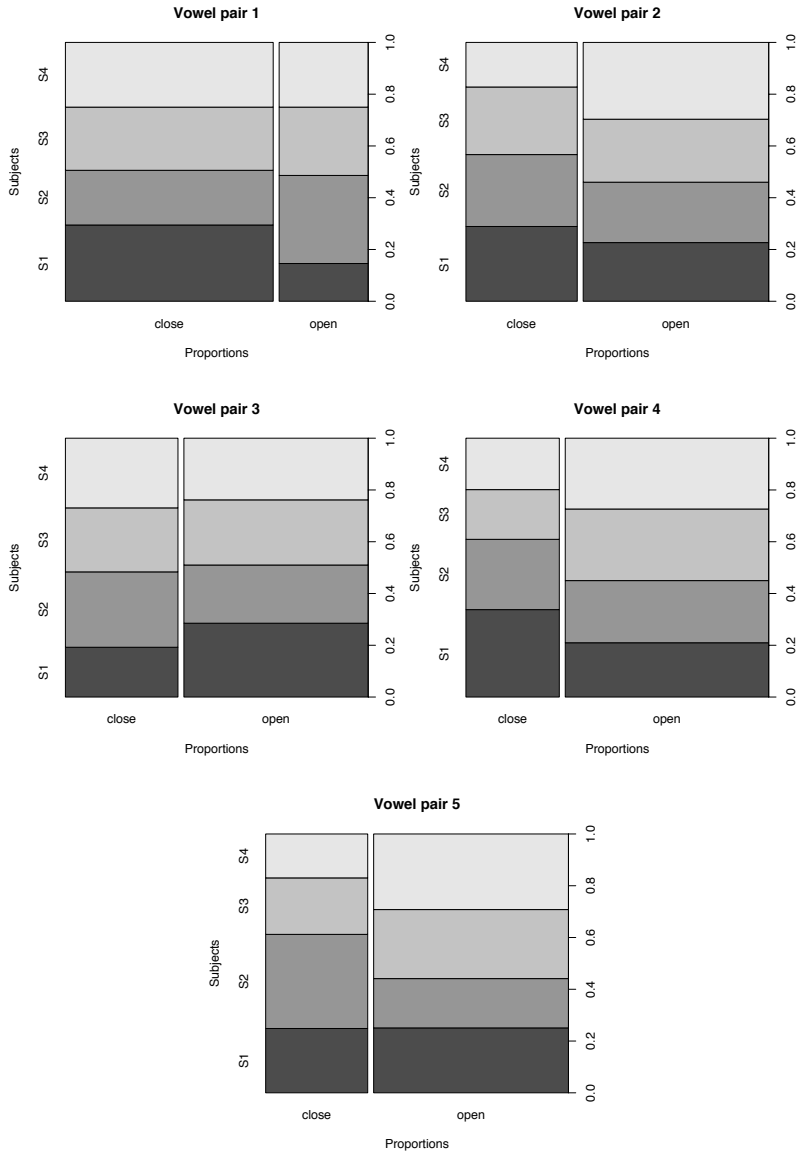


Figure 9: Overall response distributions for subject and vowel judgment, five vowel pairs.

References

BLACK, J. W. 1949: Natural frequency, duration, and intensity of vowels in reading. – *Journal of Speech and Hearing Disorders*, (14):216–221.

DI CRISTO, Albert 1978: *De la microprosodie à l'intonosyntaxe*. Thèse d'Etat. Université de Provence, Aix-en-Provence.

EEK, Arvo & MEISTER, Einar 1998: Quality of standard Estonian vowels in stressed and unstressed syllables of the feet in three distinctive quantity degrees. – *Proceedings of the Finnic Phonetics Symposium Pärnu, 11.-14.08.1998*, Linguistica Uralica. 226–233.

KLATT, Dennis 1980: Software for a cascade/parallel formant synthesizer. – *Journal of the Acoustical Society of America*, **67**(3):971–995.

KOUZNETSOV, Vladimir 2001: Perceptual role of inherent vowel duration as distinctive feature in Russian. – *Proceedings of the XI Session of the Russian Acoustical Society, Moscow, November 19–23, 2001*. Retrieved June 8, 2007, from http://www.akin.ru/Rao/sess11/sect5sp.htm, 443–447.

PETERSON, G. E. & LEHISTE, Ilse 1960: Duration of syllable nuclei in English. – *Journal of the Acoustical Society of America*, **32**(6):693–703.

WERNER, Stefan & MEISTER, Einar 2008: Microdurational influences on perceived vowel quality. – *Proceedings of the 3rd Baltic Conference on Human Language Technologies, October 4–5, 2007, Kaunas*, Kaunas (in print).

Michael O'Dell & Tommi Nieminen (toim.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 167–180.

Ikääntyneille suunnatun puheen prosodisia piirteitä kognitiivisesti vaativassa tilanteessa

Terhi Hautala & Taisto Määttä
Oulun yliopisto

Tiivistelmä

Tämä tutkimus liittyy Oulun yliopiston monitieteisen geronteknologisen ryhmän ja Soveltavan ergonomian ja geronteknologian laboratorion tutkimus- ja tuotekehitysprojektiin, jossa kehitettiin VoiceBit Oy:n NextInfo®-tuotteeseen perustuvaa automaattista puhelinpalvelujärjestelmää ikääntyneille käyttäjille. Kehitystyötä tehtiin yhdessä Oulun kaupungin sosiaali- ja terveystoimen sekä palvelujen käyttäjien kanssa. Tutkimusaineiston muodostavat automaattisen puhelinpalvelujärjestelmän käytettävyystutkimuksiin (1996) osallistuneet henkilöt. Tutkimuksen päätavoitteena oli tarkastella puheen ominaisuuksien merkitystä ikääntyneiden kuuntelijoiden puheen vastaanotossa kognitiivisesti vaativassa tilanteessa. Siinä analysoitiin järjestelmän neljän puhujan puheen ja äänen ominaisuuksia. Akustisilla mittauksilla tutkittiin ikääntyneille suunnatun puheen prosodisia piirteitä ja tarkasteltiin niiden merkitystä käytettävyystutkimuksen tehtävissä suoriutumisen kannalta. Ikääntyneet kuuntelijat antoivat myös subjektiiviset arviot puhujista. Kaikilla puhujilla oli havaittavissa ikääntyneille suunnatun puheen piirteitä, joilla he pyrkivät helpottamaan kuuntelijoiden toimintaa. Erot tehtävissä suoriutumisessa eri puhujien äänellä olivat kuitenkin pieniä. Vain yhden puhujan äänellä tehtävät kuulleiden suoriutuminen erosi tilastollisesti merkitsevästi muita puhujia kuunnelleiden suoriutumisesta.

Avainsanat: prosodia, puheen ymmärrettävyys, ikääntyneille suunnattu puhe, geronteknologia, käytettävyystutkimus

1 Johdanto

Teknologiasta on lähdetty etsimään sovelluksia, joita voitaisiin hyödyntää kehitettäessä sosiaali- ja terveydenhuollon palveluja kasvavan ikääntyneen väestönosan tarpeisiin. Geronteknologia on monitieteistä ikääntymisen tuntemiseen pohjautuvaa teknologista tutkimusta, jossa ikääntymisen tutkimusta yhdistetään teknologiseen tutkimukseen ja tuotekehitykseen (Bouma *et al.* 2000). Käyttäjänäkökulman huomiointi on suunnittelussa keskeistä. Äänen käytettävyysominaisuuksia on tutkittu erittäin vähän (Sinkkonen *et al.* 2002) ja vähäinenkin tieto siitä on usein vierasta käyttöliittymien suunnittelijoille (Gardner-Bonneau 1992, Sinkkonen *et al.* 2002). Jotta ääneen

perustuvan järjestelmän käyttäjä pystyisi optimaaliseen suoritukseen, suunnittelijoiden olisi huomioitava ympäristöhälyn vaikutus, puhujan puhetyyli, artikulaatio ja sopiva puhenopeus (Gardner-Bonneau 1992). Käyttöliittymää suunniteltaessa on huomioitava käyttäjien työmuistin ja informaation käsittelyn rajallisuus, puheäänen hetkellisyys sekä äänen vaikuttavuus vastaanottajan tunteisiin (Gardner-Bonneau 1992, Sinkkonen *et al.* 2002).

Ikääntymiseen liittyvät fyysiset ja kognitiiviset muutokset vaikuttavat auditiiviseen puheen prosessointiin, kuullun muistamiseen ja ohjeiden mukaiseen toimintaan. Ikäkuulo on iän ja kuulovian kombinaatio, ja siksi sen seuraukset näkyvät auditiivisen järjestelmän eri tasoilla (Frisina & Frisina 1997). Perifeeriset ja kognitiiviset tekijät selittävät yhdessä ikääntymiseen liittyvää sanojen tunnistamisen vaikeutta (Pilotti *et al.* 2001). Ikääntyneiden ihmisten puheen ymmärtämisen kannalta keskeisiä kognitiivisia tekijöitä ovat prosessointinopeus, työmuisti (Cohen 1987, Wingfield & Stine-Morrow 2000) ja eksekutiivinen kontrolli huomion suuntaamiseksi ja jakamiseksi sekä häiriötekijöiden ehkäisemiseksi (Tun *et al.* 2002, Zacks *et al.* 2000). Ikääntyneiden ihmisten puheen ymmärtäminen vaikeutuu viestin kompleksisuuden ja esittämisnopeuden kasvaessa (Kemper & Kemtes 1999, Wingfield & Stine-Morrow 2000). Prosodisten kontekstivihjeiden on todettu olevan ikääntyneille kuuntelijoille tärkeitä lauseiden syntaktisen rakenteen jäsentämisen ja muistamisen tukena (Cohen & Faulkner 1986, Gordon-Salant & Fitzgibbons 1997, Wingfield & Stine-Morrow 2000). Ikääntyneille suunnattu puhe on luonteeltaan yksinkertaistettua puhetta, jolle on tyypillistä liioiteltu prosodiikka (suurempi sävelkorkeuden vaihtelu ja korostettu painotus) sekä suurempi äänen voimakkuus ja huolellinen artikulaatio (Cohen & Faulkner 1986, Kemper & Harden 1999). Ikääntyneet kuuntelijat eivät kuitenkaan pitäneet nuorten heille suuntaamasta puhetavasta, jossa puhetempo oli hitaampi ja kielelliset rakenteet olivat yksinkertaistettuja (Kemper *et al.* 1996). Ohjeiden antamisessa vahva prosodisten piirteiden korostus ei myöskään miellyttänyt heitä, vaikka se auttoi kuultujen ohjeiden muistamista (Gould & Dixon 1997).

Tutkimuksen päätavoitteena oli selvittää puheen ominaisuuksien merkitystä ikääntyneiden kuuntelijoiden puheen vastaanotossa kognitiivisesti vaativassa tilanteessa. Tavoitteena oli etsiä ja analysoida ikääntyneiden puheen vastaanottoa helpottavia tekijöitä. Tutkimuksen pääkysymykseen etsittiin vastausta seuraavien alakysymysten kautta: 1) Miten ikääntyneet tutkimushenkilöt suoriutuvat järjestelmän käytettävyystutkimuksen tehtävistä eri puhujien äänillä? 2) Onko kuuntelijoiden subjektiivisesti arvioimilla puheen ominaisuuksilla (äänen voimakkuus, miellyttävyys, puheen selvyys ja puhenopeus) yhteyttä heidän suoriutumiseensa tehtävissä? 3) Onko puhelinpalvelujärjestelmän valikkoäänitteistä mitatuilla puheen prosodisilla ominaisuuksilla (tempo, tauotus ja painotus) yhteyttä ikääntyneiden kuuntelijoiden suoriutumiseen käytettävyystutkimuksen tehtävissä? 4) Onko puhujilla havaittavissa ikääntyneille suunnatun puheen piirteitä; minkälaisia mahdollisesti puheen vastaanottoa helpottavia strategioita he käyttävät? 5) Millaisia ikääntyneet kuuntelijat ovat äänen ja

puheen arvioijina? Tässä artikkelissa tutkimusaihetta on tarkasteltu keskittyen ikääntyneille suunnatun puheen prosodisiin piirteisiin sekä tehtävissä suoriutumiseen.

2 Tutkimusaineisto ja tutkimusmenetelmät

Projektin tavoitteena oli kehittää palvelutaloihin sopiva puhelinpalvelusovellus. Siksi tutkimushenkilöiksi valittiin palvelutalojen asukkaita (92 %) tai ikääntyneitä henkilöitä (8 %), jotka käyttivät läheisen palvelutalon palveluita. Tutkimusaineiston muodostavat automaattisen puhelinpalvelujärjestelmän käytettävyystutkimuksiin (1996) osallistuneet ikääntyneet tutkimushenkilöt (N = 36, keski-ikä 78 v, ikäjakauma 70–91 v). Naisia heistä oli 78 % ja miehiä 22 %. Tutkimuksiin osallistujille suoritettiin ennen varsinaista käytettävyystutkimusta terveyttä ja puhelimenkäyttöä selvittävä esihaastattelu (Pirinen *et al.* 1998). Tutkimushenkilöistä 53 %:lla oli käden toimintaan vaikuttavia sairauksia, 28 %:lla oli aivoverenkiertoperäisiä sairauksia ja 36 %:lla oli kuuloon vaikuttavia sairauksia. Haastatelluista 22 %:lla oli näköön vaikuttavia sairauksia. Tutkimushenkilöistä vain 19 %:lla (N = 7) ei oman ilmoituksen mukaan ollut mitään edellisistä sairauksista.

Puheen ja äänen laadullisten ominaisuuksien haluttiin lisäävän järjestelmäsovelluksen käytettävyyttä ja helppokäyttöisyyttä ikääntyneille käyttäjille. Käytettävyystutkimuksessa käytettiin neljän eri puhujan ääntä. Molemmat naispuhujat ovat puheterapeutteja, heistä FemA oli tutkimusajankohtana 31-vuotias ja FemB oli 33-vuotias. Miespuhujista MalA (23 v) oli aiemmin ansioitunut Yleisradion oppikirjanauhoituksissa ja MalB (52 v) on pitkään opetustyötä tehnyt fonetiikan lehtori. NextInfo®-järjestelmän lokitiedostoon kirjautuivat käytettävyystutkimukseen osallistuneiden tekemät onnistuneet toiminnot sekä tehdyt virheelliset valinnat. Siten pystyttiin analysoimaan tutkimushenkilöiden menestymistä järjestelmän tehtävissä eri puhujien äänillä ja havainnoimaan kasaantuivatko käyttäjien tekemät virheet johonkin tiettyyn valikkokäsikirjoituksen kohtaan. Tutkittavat kuulivat järjestelmän tehtävät satunnaistetusti yhden puhujan äänellä, ja lisäksi he kuulivat lyhyen ääninäytteen kaikilta neljältä puhujalta.

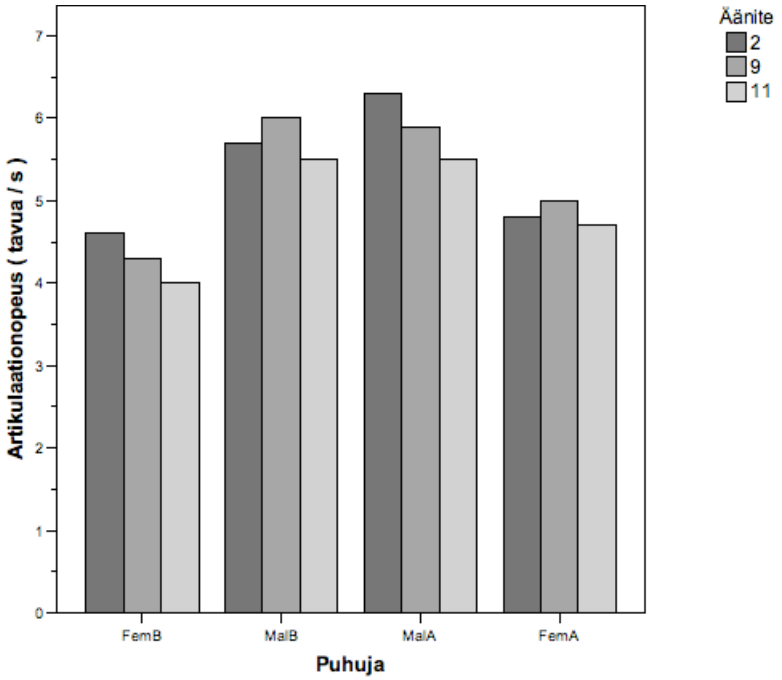
Puhelinpalvelujärjestelmän käytettävyystutkimuksen jälkeen tutkimushenkilöt arvioivat subjektiivisesti sen puhujan puheen ominaisuuksia (äänen voimakkuus, äänen miellyttävyys, puheen selvyys ja puhenopeus), jolla kuulivat järjestelmän tehtävät. Lisäksi he arvioivat lyhyen ääninäytteen kaikilta neljältä puhujalta. Kyselylomakkeen lopussa he antoivat puheäänelle kouluarvosanan (asteikko 4–10). Sen tarkoituksena oli saada kuullusta puhenäytteestä numeerinen kokonaisarvio, joka edustaisi puhujasta saatua kokonaisvaikutelmaa. Kouluarvosanan ajateltiin olevan ikääntyneille kuuntelijoille tuttu arviointiasteikko. Puhelinpalvelujärjestelmän valikkoäänitteistä valittiin analysoitaviksi kolme äänitettä, jotka olivat keskeisiä kuulijoiden toiminnan ohjaamisen kannalta käytettävyystutkimuksen tehtävien suorittamisessa.

Nämä valikkoäänitteet ovat myös tekstirakenteeltaan pidempiä ja yhtenäisempiä sisältäen sekä pää- että sivulauseita. Pidemmissä puhejaksoissa esiintyy tauotusta ja puheen prosodiset piirteet tulevat paremmin esille. Äänen ja puheen analysointiohjelmalla (CSL) mitattiin perusäänen taajuutta valikoissa annettujen ohjeiden avainfraasien prominenssihuippujen sijoittumisen ja laajuuden selvittämiseksi. Mittauskohteiksi valittiin sanat, jotka välittivät keskeisen informaation valittavista vaihtoehdoista sekä toimintaohjeen valinnan toteuttamiseksi. Mittaukset tehtiin lausepainollisessa asemassa olevista reemasanoista mittaamalla edellisen sanan aikainen F0:n minimiarvo, ja maksimiarvo mitattiin sanan F0:n huippukohdasta. F0 arvot muunnettiin puolisävelaskeliksi käyttäen Bakenin (1987) esittämää kaavaa. SoundEdit®-ohjelmalla mitattiin puheen temporaalisia ominaisuuksia. Hieke *et al.* (1983) pitävät yleensä käytettyjä tauon kynnysarvoja liian korkeina, koska lyhyemmät 0,13–0,25 sekunnin tauot ovat yhteydessä psykologisiin ja tekstuaalisiin tekijöihin puheessa. He suosittelevat 0,10 sekunnin vähimmäisrajaa tauon kriteeriksi. Tässä tutkimuksessa lyhyidenkin taukojen arvioitiin antavan tietoa puhujien välisistä strategiaeroista tauotuksen suhteen. Tauon jälkeisen klusiilin kestoksi arvioitiin mittauksissa 100 ms, joka rajattiin pois tauon kestosta. Suomen kielessä on mitattu sananalkuisen konsonantin kestoksi yksinäisvokaalien edellä 92 ms (Suomi 2006). Tutkimuksen aikana Oulun aluetyöterveyslaitoksen työhygieenikko mittasi päivittäin äänitteiden äänen voimakkuudet puhelimen kuulokkeesta, jotta äänen voimakkuus pysyi vakiona eri tutkimuspäivinä.

3 Tulokset

Käytettävyystutkimuksen puhetilanne oli valikkokäsikirjoituksen mukaista lukemista, mutta puhujat tiesivät kuuntelijoiden olevan ikääntyneitä ihmisiä, joiden pitäisi pystyä toimimaan kuulemiensa ohjeiden mukaisesti. Puhujien prosodisessa tyylissä oli eroja, mutta kaikilla puhujilla oli havaittavissa ikääntyneille suunnatun puheen piirteitä. Varsinaisia puhujaprofiileja tai puhestrategioita ja niiden tilastollista merkitsevyyttä tehtävissä suoriutumisen kannalta ei näin pienellä puhuja-aineistolla voinut tehdä. Ikääntyneille suunnatun puheen prosodisten piirteiden tarkastelu perustuu pääasiassa akustisiin mittauksiin ja laadulliseen analyysiin. Ikääntyneille suunnatun puheen määrässä, eri keinojen käytössä sekä puhetavan johdonmukaisuudessa oli havaittavissa eroja puhujien välillä. Heillä oli havaittavissa ikääntyneille suunnatun puheen piirteitä prosodisten ominaisuuksien osalta puhe- ja artikulaationopeudessa, tauotuksessa (kestot, lukumäärä, fokusointi) ja painotuksessa.

Puhujien artikulaationopeudet vaihtelivat välillä 4,32–5,93 tavua/s, keskiarvon ollessa 5,22 tavua/s. Hitain keskimääräinen artikulaationopeus (4,32 tavua/s) oli puhujalla FemB. Tempoltaan seuraavaksi hitain puhuja oli FemA, jonka nopeus oli 0,53 tavua/s suurempi kuin FemB:n. Jo selvästi havaittava ero artikulaationopeuksissa oli

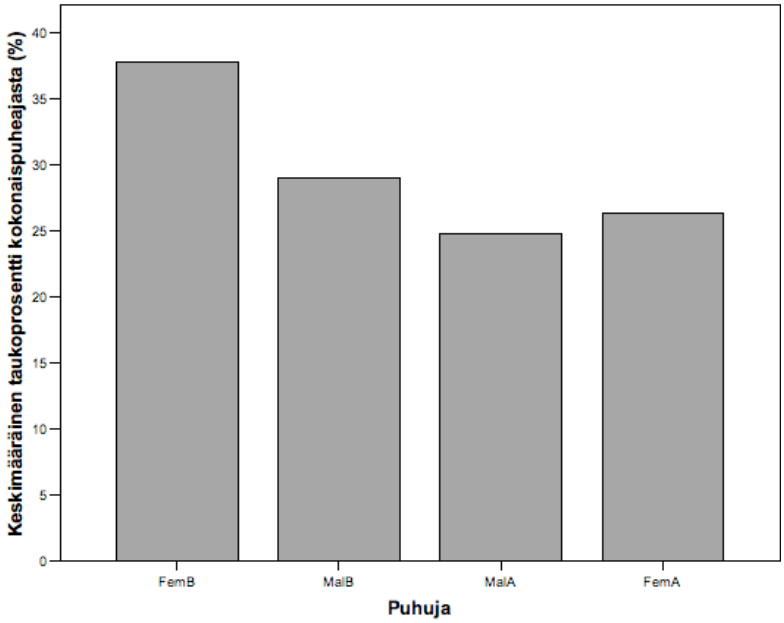


Kuva 1: Puhujien keskimääräiset artikulaationnopeudet äänitteittäin.

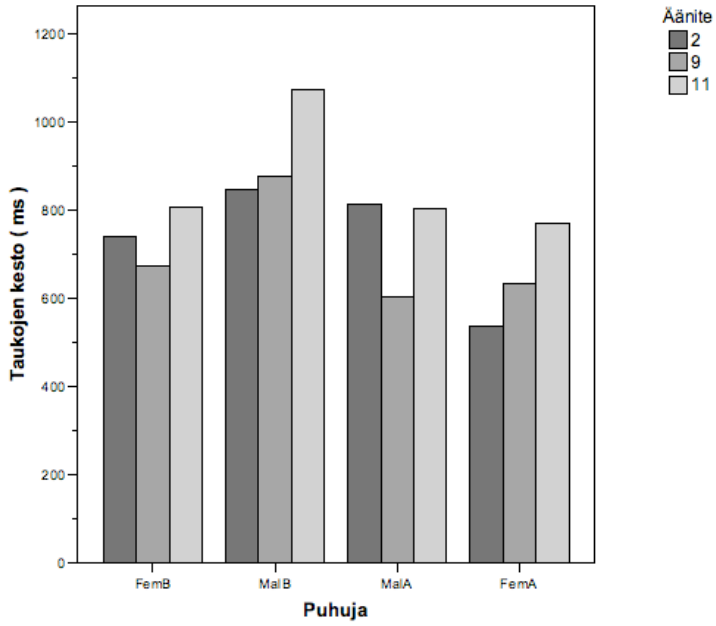
puhujien FemB ja MalB välillä (ero 1,43 tavua/s). Vielä suurempi ero artikulaationopeuksissa oli puhujien FemB ja MalA välillä (ero 1,61 tavua/s). Puheen temporaalisissa tekijöissä oli havaittavissa yksilöllistä vaihtelua eri puhujilla. Keskimääräinen puhe- tai artikulaationopeus ei ollut kuitenkaan tilastollisesti merkitsevä tekijä tehtävissä suoriutumisen kannalta. Naispuhujilla puhetempo oli hitaampi kuin miehillä, ja kaikkien puhujien keskimääräinen puhenopeus ja artikulaationopeus hidastuivat kuuntelijoille vaikeimmaksi osoittautuneessa äänitteessä 11.

Kokonaispuheajasta prosentuaalisesti eniten taukoja oli puhuja FemB:llä eri äänitteissä. Kuulijoille vaikeimmassa äänitteessä 11 hänen taukoprosenttinsa oli 42,3 % kokonaispuheajasta. Hänellä oli myös erittäin paljon fokusointitaukoja, ja hän sijoitti tauot systemaattisesti aina ennen puhelinpalvelujärjestelmässä annettavia valintavaihtoehtoja tai toimintaohjeita. Seuraavaksi suurin taukoprosentti oli MalB:llä ja hänelläkin se oli suurin äänitteessä 11. Pienin taukoprosentti oli puhujalla MalA, ja sekä hänellä että FemA:lla taukoprosentti kuulijoille vaikeimmassa äänitteessä 11 oli pienempi kuin äänitteessä 2, jossa heillä oli suurin taukoprosentti. Miespuhujat tauottivat kieliopillisia rajoilla ja sijoittivat tauot tekstissä samoihin kohtiin.

Puhuja MalB:n taukojen keskimääräiset kestot olivat suurimpia verrattuna muihin puhujiin kaikissa äänitteissä. Hän käytti lukumääräisesti vähemmän taukoja kuin



Kuva 2: Puhujien keskimääräinen taukoprosentti kokonaispuheajasta.



Kuva 3: Taukojen keskimääräiset kestot valikkojärjestelmän äänitteissä.

Taulukko 1: Puhujien puolisävelaskelmuutosten (psa) keskiarvot ja luottamusvälit äänitteittäin.

Puheääni	Äänite	Keskiarvo psa	Keskihajonta	95 % luottamusväli
FemB	2	4,9	3,5	(2,3–7,4)
	9	5,9	5,1	(2,2–9,6)
	11	3,9	6,3	(−0,9–8,7)
MalB	2	6,1	3,4	(3,5–8,7)
	9	6,2	3,8	(3,3–9,1)
	11	7,9	3,5	(4,7–11,1)
MalA	2	6,2	6,5	(1,8–10,6)
	9	4,3	3,7	(1,6–6,9)
	11	4,3	5,4	(0,7–7,9)
FemA	2	4,0	2,1	(2,5–5,6)
	9	4,4	1,9	(3,0–5,9)
	11	4,8	1,7	(3,2–6,4)

FemA ja FemB, mutta ne olivat kestoltaan vastaavasti pidempiä. Määrällisesti samanverran taukoja käyttäneen MalA:n taukojen kestot olivat MalB:n taukoja lyhyempiä. FemB, jolla oli lukumäärällisesti eniten taukoja, käytti niihin keskimäärin enemmän aikaa kuin FemA kaikissa äänitteissä. FemA:lla taukoja oli enemmän kuin MalA:lla ja MalB:llä, mutta hänen taukojensa kesimääräiset kestot olivat heihin verrattuna lyhyempiä.

Puhujien välillä oli havaittavissa selviä yksilöllisiä eroja painotuksen käytössä. Perusäänen taajuuden muutokset mitattiin valikkokäsikirjoituksen keskeisimmistä ohjeista kaikilta puhujilta. Yleensä 1–1,5 puolisävelaskeleen nousua/laskua pidetään kuulijalle jo havaittavana muutoksesta (Rietveldt & Gussenhoven 1985, Iivonen *et al.* 1998).

Suurinta perusäänen taajuuden muutosta ilmausten fokuskohdissa käytti puhuja MalB ja hänellä puolisävelaskelmuutosten keskiarvo oli suurin kuuntelijoille vaikeimmassa äänitteessä 11. Pienempiä perusäänen taajuuden muutoksia käyttivät puhujat MalA ja FemA. Puhuja FemB:llä oli vaihtelua äänitteiden kesken puolisävelaskelmuutosten käytössä, esimerkiksi vaikeimmassa äänitteessä 11 puolisävelmuutosten keskiarvo oli pienin.

Ikääntyneiden tutkimushenkilöiden suoriutumista puhelinpalvelujärjestelmän käytettävyystutkimuksen tehtävistä tarkasteltiin kahdella eri tavalla. Suoraan aineistosta lasketut onnistumisprosentit olettavat havaintojen olevan toisistaan riippumattomia.

Taulukko 2: Kuuntelijoiden tehtävistä suoriutuminen eri puhujien äänillä ja odotettu onnistuminen eri äänillä (suoraan aineistosta laskettu suoriutuminen).

	Oikein		Väärin		Kaikki		Odotettu onnistuminen	95 % luottamusväli
Puhuja	N	%	N	%	N	%	%	
FemB	17	68,0	8	32,0	25	100	87,4	(68,2–106,5)
MalB	15	60,0	10	40,0	25	100	69,2	(50,1–88,4)
MalA	14	56,0	11	44,0	25	100	64,0	(44,8–83,1)
FemA	9	37,5	15	62,5	24	100	24,1	(4,4–43,7)
Kaikki	55	55,6	44	44,4	99	100		

Riippumattomuusehto ei kuitenkaan täyty, koska äänitteissä onnistumisen mittauksissa arvioidaan toistuvasti saman tutkimushenkilön suoriutumista. Siksi tutkimuksessa käytettiin eri puhujien äänellä menestymisen tilastollisessa tarkastelussa myös toistettujen mittausten sekamallia. Suoriutumista tarkasteltiin sekä suoraan aineistosta lasketuilla prosenttiosuuksilla että sekamallilla lasketuilla odotetuilla onnistumisprosenteilla.

Tarkasteltaessa erikseen puhujan vaikutusta tehtävästä suoriutumiseen vain FemA:n äänellä tehtävät kuulleiden odotettu onnistumisprosentti 24,1 % 95 % lv (4.4–43.7) erosi tilastollisesti merkitsevästi muiden suoriutumisesta. Parhaiten tehtävissä menestyttiin puhuja FemB:n äänellä, vaikka ero ei ollut tilastollisesti merkitsevä. Miespuhujia kuunnelleiden tutkimushenkilöiden suoriutumisessa ei ollut suurta eroa.

Tuloksissa on nähtävissä puheen arviointien osalta jonkinlaista ärtymystä puheen mukauttamisen suhteen, joka tuli esille puheen miellyttävyyden arvioinneissa ja annetuissa kouluarvosanoissa. Puhenäytteiden kuuntelutilanteessa 28 kuuntelijaa (N = 144) eli 19,4 % piti ääntä hieman epämiellyttävänä tai erittäin epämiellyttävänä. Kuuntelijoista 6 (N = 36) eli 16,7 % arvioi FemB:n äänen hieman epämiellyttäväksi. MalB:n äänen arvioi hieman epämiellyttäväksi 5 kuuntelijaa eli 13,9 % ja MalA:n arvioi vastaavasti vain 2 kuuntelijaa. Puhenäytteiden arvioinnissa käytettiin myös kategoriaa erittäin epämiellyttävä. MalA:n ääntä kukaan ei arvioinut erittäin epämiellyttäväksi. Molemmat naispuhujat saivat yhdeltä kuuntelijalta arvion erittäin epämiellyttävä. MalB:n äänen arvioi erittäin epämiellyttäväksi kaksi kuuntelijaa. Puheen selvyyttä arvioitiin puhenäytteissä kriittisemmin, 26 kuuntelijaa (N = 144) eli 18,1 % arvioi puheen hieman epäselväksi. Miespuhujien saamat arviot eivät poikenneet merkittävästi käytettävyystutkimustilanteen arvioista. Puhenäytteiden arviointitilanteessa FemB:n puheen hieman epäselväksi arvioi 9 puhujaa (N = 36) eli 25 %

kuuntelijoista. FemA:n äänen arvioi liian hiljaiseksi 13 kuuntelijaa eli 36,1 %. Myös työterveyslaitoksen mittauksissa hänen äänenvoimakkuutensa oli hiljaisin. Mahdollisesti pienempi äänen voimakkuus vaikutti myös puheen miellyttävyyden ja selvyyden arvioihin, koska hänestä annetut arviot niiden osalta poikkesivat muista puhujista. Kuuntelijoista 11 eli 30,6 % arvioi hänen äänensä hieman epämiellyttäväksi. Hieman epäselväksi hänen puheensa arvioi 14 kuuntelijaa eli 38,9 %.

4 Pohdinta

Kahden puhujan FemB:n ja MalB:n äänillä suoriuduttiin tehtävistä parhaiten. Heidän puheessaan oli havaittavissa eniten ikääntyneille suunnatun puheen piirteitä. Kuitenkin erot eri puhujia kuunnelleiden suoriutumisessa olivat pieniä. Vain voimakkuudeltaan heikoimmalla FemA:n äänellä tehtävät kuulleet suoriutuivat tilastollisesti merkitsevästi heikommin kuin muita puhujia kuunnelleet. Kaikkia puhujia voidaan pitää hyvinä puhujina, joten erojen vähäisyys käytettävyystutkimuksessa suoriutumisessa voi johtua myös siitä. Tutkimuksessa löytyi viitteitä siihen suuntaan, että voimakas avainsanojen painotus ja taukojen pitkät kestot (puhuja MalB) helpottivat puheen vastaanottoa. Myös hidas puhetempo, fokusointitaukojen suuri määrä sekä niiden sijoittaminen ennen valintavaihtoehtoja ja toimintaohjeita (puhuja FemB) näyttivät tukevan ikääntyneiden kuuntelijoiden toimintaa käytettävyystutkimuksessa. Puhuja FemB käytti samanaikaisesti monia ikääntyneille suunnatun puheen piirteitä. MalB puolestaan käytti korostettua painotusta ja taukojen kestoja hyvin johdonmukaisesti.

FemB:n ja MalB:n kuuntelijat suoriutuivat hieman paremmin tehtävissä. Kuitenkin menestymisessä eri äänitteissä eri puhujaa kuunnelleiden välillä oli eroja. Äänitteessä yksitoista MalA:n äänellä odotettu onnistumistodennäköisyys oli suurin, tosin ero ei ollut suuri. Suoraan aineistosta lasketuissa arvoissa hänen ja FemB:n kuuntelijoiden virheprosentti oli sama (57,1 %) äänitteessä 11. Jokin tekijä FemB:n puheessa on heikentänyt kuuntelijoiden suoriutumista tässä vaativimmassa äänitteessä. On mahdollista, että ikääntyneille suunnatun puheen piirteiden suuri määrä on jo kääntynyt vastaanottoa haittaavaksi tekijäksi. Äänitteen 11 fokusointitaukojen erittäin suuri määrä (10 kpl) yhdistyneenä taukojen keston kasvamiseen sekä hidas artikulaationopeus ovat voineet vaikeuttaa puheen vastaanottoa rikkomalla lauseiden prosodista kontuuria. Ikääntyneiden puheen vastaanottoa on todettu helpottavan prosessointiajan lisääminen lingvististen rakenteiden rajakohtiin, kun taas taukojen lisääminen sanojen väliin lauseissa ei helpota vastaanottoa (Gordon-Salant & Fitzgibbons 1997, Wingfield *et al.* 1999). Ylimukautetulla puhetavalla esitetyn lääkkeen käyttöohjeiden muistaminen oli parempaa sellaisten ikääntyneiden kuulijoiden ryhmässä, joilla oli parempi työmuisti (Gould & Dixon 1997). Ilmauksen keskipituuden vähentäminen voi vaikeuttaa ymmärtämistä hajaannuttamalla informaatiota liian moniin lauseisiin rasittaen työmuistia (Kemper & Harden 1999). Työmuisti on voinut kuormittua

FemB:n puheen hitaan tempon ja runsaan tauotuksen vuoksi. Puhuja MalB:n osalta aineistosta laskettu virheprosentti oli äänitteessä yksitoista varsin suuri (71, 4 %). Tämä näkyy myös siinä, että hänen äänellään odotettu onnistumisprosentti kyseisessä äänitteessä oli kolmanneksi alhaisin. Myös hänellä ikääntyneille suunnatun puheen piirteiden osuus kasvoi äänitteiden vaikeutuessa, mikä näkyi taukojen keston pidentymisenä ja fokuskohtien painotuksen lisääntymisenä. Odotetun onnistumisprosentin heikkenemisen syitä äänitteessä 11 on vaikea tietää. Mukauttiko hänkin jo liikaa puhettaan, vai vaikuttiko heikompaan menestykseen hänen kuuntelijoidensa korkeaan ikään liittyvät tekijät tai kuuntelijoiden heikoin kuulo- ja puheenerotuskyky? Oliko FemB:n ja MalB:n puhetyyli äänitteessä yksitoista tulkittavissa jo ylimukautetuksi?

Heikoimmista puhuja-arvioinneista (kouluarvosanan keskiarvo) huolimatta FemB:n kuuntelijat menestyivät käytettävyystutkimuksen tehtävissä parhaiten. Kuuntelijoita on voinut ärsyttää hänen runsas puheen mukauttamisensa, ja FemA:n arviointeihin vaikutti heikompi äänen voimakkuus. Mahdollisesti FemB:n erittäin hidas puhetempo aiheutti puhujille kokemuksen siitä, että heidän puheen vastaanottokykyään aliarvioidaan kuten Kemperin *et al.* (1996) tutkimuksessa. Kemper & Harden (1999) totesivat suuremman sävelkorkeuden vaihtelun jopa heikentävän kuuntelijoiden suoriutumista. Tämä ei kuitenkaan toteutunut MalB:n kohdalla, vaikka hän käytti koetilanteessa suurinta sävelkorkeuden vaihtelua ilmausten fokuskohdissa. Hänen äänellään onnistuttiin tehtävissä toiseksi parhaiten. Miellyttävimpänä ja parhaimpana puhujana kouluarvosanalla arvioiden pidettiin puhuja MalA:ta, jonka puheessa oli havaittavissa vähiten ikääntyneille suunnatun puheen piirteitä. Miespuhujista pidettiin enemmän kuin naispuhujista yleisarvioina annettujen kouluarvosanojen perusteella molemmissa arviointitilanteissa. Tämä on mielenkiintoista, koska puhestrategian suhteen miespuhujat poikkesivat toisistaan selvästi. MalA helpotti kuuntelijoiden puheen vastaanottoa prosodisin keinoin vähemmän kuin MalB. Käytettävyystutkimuksen kuuntelijoista suurin osa oli naisia (78 %), joten miesäänen kuunteleminen voi olla heistä miellyttävämpää. Valon (1994) tutkimuksen äänenarvioinnissa (64 % naisia) miesääniin liittyi joiltakin osin positiivisempia mielleyhtymiä kuin naisääniin, ja samantyyppisten ääninäytteiden parista mies arvioitiin positiivisemmin kuin nainen. Markhamin ja Hazanin (2004) tutkimuksessa puheen selvyyden ja ymmärrettävyyden osalta naispuhujien äänellä suoriuduttiin kuuntelutehtävistä hieman paremmin kuin miesäänillä. Myös Dearbornin *et al.* (2006) tutkimuksessa ikääntyneistä kuuntelijoista (69–91 v) sekä naiset että miehet vastaanottivat ja muistivat suuremman määrän kuulemastaan informaatiosta naispuhujan kuin miespuhujan äänellä. Tässä työssä naisäänellä onnistuttiin tehtävissä parhaiten, vaikka miesääniä pidettiin miellyttävämpinä. Puheen miellyttävyystekijät ja puheen vastaanottoon liittyvät tekijät eivät aina liity toisiinsa. Tosin on muistettava, että järjestelmän tehtävissä suoriutumisessa FemB:n, MalB:n ja MalA:n kuuntelijoiden välillä ei ollut suuria eroja.

Puheen ominaisuuksien erottelu toisistaan on vaikeaa, jotta voisi osoittaa mikä tietyistä prosodisista piirteistä olisi erityisen tärkeä ikääntyneiden puheen vastaanot-

tamisen kannalta. Puheen prosodisilla keinoilla tuotettavien erojen havaitseminen perustuu useamman akustisen parametrin yhdistelmiin (Suomi *et al.* 2006, Warren 1999). Puhujan prosodisten keinojen valintaan ovat vaikuttamassa samanaikaisesti monet tekijät (Aho & Yli-Luukko 2005, Shattuck-Hufnagel & Turk 1996, Suomi *et al.* 2006, Tuomainen 2001). Näistä tekijöistä sekä niiden suhteellisesta tärkeydestä ei ole tarkkaa tietoa (Shattuck-Hufnagel & Turk 1996). On vaikea esittää luotettavaa vastausta siihen, miten puheen ominaisuudet ja mitkä puheen ominaisuudet todella vaikuttivat tehtävissä suoriutumiseen. Puhujia olisi pitänyt olla moninkertaisesti enemmän, jotta prosodisia piirteitä tai niiden yhdistelmiä olisi voinut tilastollisesti mallintaa esimerkiksi erilaisiksi puhestrategioiksi. Tällöin olisi voitu saada luotettavampia tuloksia siitä, onko olemassa jokin puheen ominaisuuksien yhdistelmä, joka tukisi ikääntyneiden kuuntelijoiden puheen vastaanottoa kognitiivisesti vaativassa tilanteessa? Nyt ikääntyneille suunnatun puheen ominaisuuksien tarkastelu jäi yksittäisistä puhujista tehtyjen akustisten mittausten ja laadullisen analysoinnin sekä kuuntelijoiden arvioiden varaan. Tutkimuksessa ei tarkasteltu ikääntyneisiin kuuntelijoihin liittyviä piirteitä eikä myöskään kuunneltavan tekstin lingvistisen rakenteen mahdollisia vaikutuksia ikääntyneiden suoriutumiseen puhelinpalvelujärjestelmän tehtävissä. Siten kovin pitkälle meneviä johtopäätöksiä ei voi tehdä pelkästään puhujiin liittyvien tekijöiden pohjalta, vaan on tarkasteltava valikkotekstin rakenteen merkitystä sekä ennen kaikkea ikääntyneisiin kuuntelijoihin liittyviä tekijöitä. Viitteitä valikkotekstin merkityksestä saatiin tässäkin tutkimuksessa, koska ikääntyneiden kuuntelijoiden suoriutuminen heikkeni työmuistin kannalta vaativimmassa (kolme eri valintavaihtoehtoa, pidempi teksti) äänitteessä. Tulevaa tutkimusta varten samaa ikääntyneiden ryhmää tutkittiin uudelleen ja siinä tarkastellaan puhujien lisäksi valikkoteksteihin sekä ikääntyneisiin kuuntelijoihin liittyviä tekijöitä puhelinpalvelujärjestelmän tehtävissä suoriutumisen kannalta. Siinä tarkastellaan kuulijaan liittyvien tekijöiden merkitystä (kuulokyky, koulutustaso, subjektiivinen terveydentila, kognitiivinen suoriutuminen, puheen ymmärtäminen) puheen vastaanoton kannalta. Tavoitteena on tutkia myös kuultavan viestin lingvistisen kompleksisuuden ja muistin kuormituksen merkitystä tehtävissä suoriutumisen kannalta.

Olisi tärkeää pystyä erittelemään ikääntyneille suunnatusta puheesta ne tekijät, jotka todella helpottavat puheen vastaanottoa ja toisaalta tekijät, jotka laukaisevat kuuntelijoiden negatiiviset itsearviot omista kommunikaatiokyvyistään (Kemper & Harden 1999, Kemper & Kemtes 1999). Ääneen perustuvien käyttöliittymien suunnittelussa ikääntyneitä varten joudutaan tekemään jonkinlaisia kompromisseja näiden keskenään ristiriidassa olevien tekijöiden välillä, jotta järjestelmän käytettävyys sekä käytön miellyttävyys olisivat mahdollisimman hyviä.

Viitteet

AHO, Eija & YLI-LUUKKO, Eeva 2005: Intonaatiojaksoista. – *Virittäjä*, **109**(2):201–220. (abstract: Intonation units).

BAKEN, R. J. 1987: *Clinical Measurement of Speech and Voice*. Boston: Little, Brown and Company Inc.

BOUMA, H., FOZARD, J. L., HARRINGTON, T. L. & KOSTER, W. G. 2000: Overview of the field. – T. L. Harrington & M. K. Harrington (toim.), *Gerontechnology Why and How*. Maastricht: Shaker Publishing. 7–37.

COHEN, G. 1987: Speech comprehension in the elderly: The effects of cognitive changes. review article. – *British Journal of Audiology*, **21**:221–226.

COHEN, G. & FAULKNER, D. 1986: Does "Elderspeak" work? The effect of intonation and stress on comprehension and recall of spoken discourse in old age. – *Language and Communication*, **6**:91–98.

CRAIK, F. I. M. & SALTHOUSE, T. A. (toim.) 2000: *The Handbook of Aging and Cognition*. Lawrence Erlbaum Associate Publishers.

DEARBORN, J. L., PANZER, V. P., BURLESON, J. A., HORNUNG, F. E., WAITE, H. & INTO, F. H. 2006: Effect of gender on communication of health information to older adults. – *Journal of American Geriatrics Society*, **54**:637–641.

FRISINA, D. & FRISINA, R. 1997: Speech recognition in noise and presbyacusis: Relations to possible neural mechanisms. – *Hearing Research*, **106**(1-2):95–104.

GARDNER-BONNEAU, D. 1992: Human factors problems in interactive voice response (IVR) applications: Do we need a guideline/standard? – *Proceedings of the Human Factors Society 36th Annual Meeting*, osa 1. Santa Monica USA: Human Factors and Ergonomics Society. 222–226.

GORDON-SALANT, S. & FITZGIBBONS, P. J. 1997: Selected cognitive factors and speech recognition performance among young and elderly listeners. – *Journal of Speech, Language and Hearing Research*, **40**(2):423–431.

GOULD, O. N. & DIXON, R. A. 1997: Recall of medication instructions by young and elderly adult woman: Is overaccommodative speech helpful? – *Journal of Language and Social Psychology*, **16**(1):50–69.

HIEKE, A., KOWAL, S. & O'CONNEL, D. 1983: The trouble with "articulatory" pauses. – *Language and Speech*, **26**:203–214.

IIVONEN, A., NIEMI, T. & PAANANEN, M. 1998: Do F0 peaks coincide with lexical stresses? – S. Werner (toim.), *Nordic Prosody: Proceedings of the VIIth Conference, Joensuu 1996*. Frankfurt am Main: Peter Lang GmbH. 141–158.

KEMPER, S. & HARDEN, T. 1999: Experimentally disentangling what's beneficial about elderspeak from what's not. – *Psychology and Aging*, **14**(4):656–670.

KEMPER, S. & KEMTES, K. 1999: Aging and message production and comprehension. – N. Schwarz, D. Park, B. Knäuper & S. Sudman (toim.), *Cognition, Aging and Self-reports*. Taylor & Francis. 229–244.

KEMPER, S., OTHICK, M., WARREN, J., GUBARCHUK, J. & GERHING, H. 1996: Facilitating older adults' performance on a referential communication task through speech accommodations. – *Aging, Neuropsychology, and Cognition*, **3**(1):37–55.

MARKHAM, D. & HAZAN, V. 2004: The effect of talker- and listener-related factors on intelligibility for a real-word, open-set perception test. – *Journal of Speech, Language, and Hearing Research*, **47**(4):725–737.

PILOTTI, M., BEYER, T. & YASUNAMI, M. 2001: Encoding tasks and the processing of perceptual information in young and older adults. – *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, **56**(2):119–128.

PIRINEN, M., HAUTALA, T., PRYKÄRI, T., MÄÄTTÄ, T., KIVELÄ, E–M., MIELONEN, P., VIROKANNAS, H., VÄYRYNEN, S. & SAAJANTO, E. 1998: Automaattisen puhelinpalvelujärjestelmän kehittäminen ikääntyneille. – A. Oikarinen, J. Sinisammal, V. Tornberg & S. Väyrynen (toim.), *Geronteknologian perusteita ja sovelluksia*, Oulun yliopiston työtieteen jaoksen hankeraportteja 4. Oulun yliopistopaino. 66–130.

RIETVELDT, A. & GUSSENHOVEN, C. 1985: On the relation between pitch excursion size and prominence. – *Journal of Phonetics*, **13**(3):299–308.

SHATTUCK-HUFNAGEL, S. & TURK, A. E. 1996: A prosody tutorial for investigators of auditory sentence processing. – *Journal of Psycholinguistic Research*, **25**(2):193–247.

SINKKONEN, I., KUOPPALA, H., PARKKINEN, J. & VASTAMÄKI, R. 2002: *Käytettävyyden psykologia*. Helsinki: IT Press.

SUOMI, Kari 2006: Suomen segmenttikestojen määräytymisestä. – *Virittäjä*, **110**(4):483–503.

SUOMI, Kari, TOIVANEN, Juhani & YLITALO, Riikka 2006: *Fonetiikan ja suomen äänneopin perusteet*. Helsinki: Gaudeamus.

TUN, P., O'KANE, G. & WINGFIELD, A. 2002: Distraction by competing speech in young and older adult listeners. – *Psychology and Aging*, **17**(3):453–467.

TUOMAINEN, Jyrki 2001: *Language Specific Cues to Lexical Segmentation of Spoken Words in Finnish: Behavioral and Event-Related Brain Potential Studies*. Tilburg University.

VALO, M. 1994: *Käsitykset ja vaikutelmat äänestä: Kuuntelijoiden arviointia radiopuhujien äänestä*. Studia Philologica Jyväskyläensia 33. Jyväskylän yliopisto.

WARREN, P. 1999: Prosody and language processing. – S. Garrod & M. Pickering (toim.), *Language Processing*. UK: Psychology Press Ltd. 155–188.

WINGFIELD, A. & STINE-MORROW, E. A. L. 2000: Language and speech. – Craik & Salthouse (2000). 359–416.

WINGFIELD, A., TUN, P., KOHN, K. & ROSEN, M. J. 1999: Regaining lost time: Adult aging and the effect of time restoration on recall of time-compressed speech. – *Psychology and Aging*, **14**(3):380–389.

ZACKS, R. T., HASHER, L. & LI, K. Z. H. 2000: Human memory. – Craik & Salthouse (2000). 293–357.

Michael O'Dell & Tommi Nieminen (toim.), *Fonetiikan päivät 2008*.
Tampere Studies in Language, Translation and Culture, Series B 3.
Tampere: Tampere University Press, 181–188.

Klusiilin kvantiteetin vaikutus edeltävän vokaalin fonaatioon

Maria Kunnas & Michael O'Dell
Tampereen yliopisto

Tiivistelmä

Tutkimuksen tavoitteena oli selvittää, vaikuttavatko klusiilin ääntöpaikka ja kvantiteetti sitä edeltävän vokaalin H1 − H2-arvoon. Suuri H1 − H2-arvo viittaa siihen, että äänihuulet ovat auki suhteellisen suuren osan glottispulssista. Tämä taas voi viitata siihen, että valmistellaan äänihuulten avaamista klusiilin okkluusion ajaksi. H1 − H2-arvon avulla tutkitaan tässä siis sitä, onko klusiilia edeltävässä vokaalissa akustisia merkkejä siitä, että äänihuulet avattaisiin aktiivisesti okkluusion ajaksi.

Tutkittavina olivat suomen yksinäisklusiilit *p*, *t* ja *k* sekä geminaattaklusiilit *pp*, *tt* ja *kk*. Klusiileja tutkittiin nauhoituksista, joissa koehenkilöt lukivat lauseita, joihin kyseiset klusiilit oli sijoitettu. Tutkimuksessa havaittiin, että ääntöpaikka vaikutti yksinäisklusiilia edeltävän vokaalin H1 − H2-arvoon. Geminaattaklusiilia edeltävän vokaalin H1 − H2-arvo sen sijaan oli verraten suuri riippumatta ääntöpaikasta. Tulokset viittaavat siihen, että äänihuulia avataan aktiivisesti geminaattaklusiilien yhteydessä ja että tämä avaaminen vaikuttaa soinnin laatuun jo edeltävän vokaalin lopussa, ja toisaalta siihen, että äänihuulia ei avata aktiivisesti yksinäisklusiilien yhteydessä.

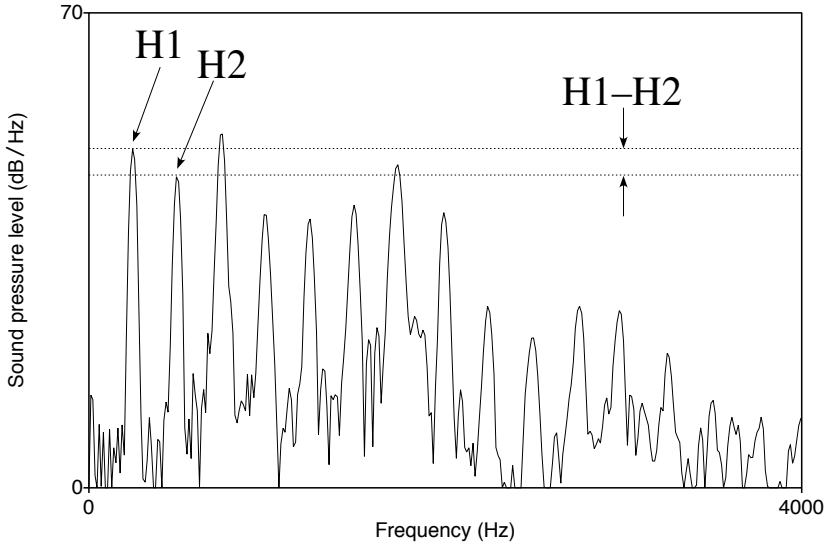
Avainsanat: kvantiteetti, fonaatio, geminaatta, klusiili, suomi

1 Johdanto

Tutkimuksen taustalla ovat Suomen (1980) ja Iivosen (1975) tutkimukset. Suomi huomasi, että suomessa lyhyen, vokaalien välisen klusiilin okkluusion alussa sointi jatkuu sitä pidempään, mitä etisemmästä klusiilista on kysymys. Tämä on luonnollista, jos äänihuulia ei avata aktiivisesti. Intraoraalinen paine kasvaa hitaammin etisemmissä tapauksissa. Englannin klusiileissa ei tutkimuksessa havaittu vastaavaa vaikutusta, mikä johtunee siitä, että englannin tapauksessa sointi lopetetaan aktiivisesti avaamalla äänihuulet. Iivonen puolestaan oli glottiksen liikkeitä tutkiessaan havainnut, että hänen koehenkilöillään glottis avautui geminaattaklusiilin tai muun tavunloppuisen klusiilin aikana, mutta ei yksinäisklusiilin aikana.

Iivosen havainto herätti useita kysymyksiä: Jos äänihuulet avautuvat ja sulkeutuvat geminaattaklusiilin aikana, liikettä ei mitä ilmeisimmin ole mahdollista kuulla. Mahdollinen aktiivinen avaaminen on kuitenkin opittua toimintaa, joten sen pitäisi olla jotenkin kuultavissa. Jos aktiivinen avaus siis tehdään, mistä se on havaittavissa? Jatkokysymys edelliseen on, valmistellaanko äänihuulien avaamista jo edeltävän äänteen aikana, jos äänihuulet avataan geminaattaklusiilin aikana? Ja jos avaamista valmistellaan jo edeltävän äänteen aikana, niin onko alkavasta avaamisesta akustisia merkkejä kyseisessä äänteessä?

Tarkoituksena oli tutkia, onko klusiilia edeltävän vokaalin fonaatiossa merkkejä siitä, että äänihuulet alkaisivat avautua. Tutkittavina olivat yksinäisklusiilit *p*, *t* ja *k* sekä geminaattaklusiilit *pp*, *tt* ja *kk*. Varsinaisen tarkastelun kohteena olivat klusiilia edeltävän vokaalin spektristä lasketut H1 − H2-arvot. H1 − H2-arvolla tarkoitetaan desibeleinä ilmoitettavaa arvoa, joka saadaan, kun lasketaan kahden ensimmäisen osasävelen erotus (ks. kuva 1).



Kuva 1: Esimerkki H1 − H2:n laskemisesta

Suuri H1 − H2-arvo viittaa siihen, että äänihuulet ovat auki suhteellisen suuren osan glottispulssista eli *Open Quotient* -arvo (OQ-arvo) on suuri. Tämä voi puolestaan viitata siihen, että äänihuulten avaamista valmistellaan (ks. esim. Ní Chasaide & Gobl 1993, Hanson 1997, Hanson & Chuang 1999).

2 Aineisto ja tutkimusmenetelmät

Tutkimusta varten tehtiin studionauhoituksia Tampereen yliopiston puheopin laitoksen puheentutkimuslaboratorion äänitysstudiossa. Nauhoitukset tehtiin 11.10.2007. Koehenkilöitä oli yhteensä 5 (4 naista ja 1 mies), ja he olivat 20–35-vuotiaita Tampereen yliopiston opiskelijoita. Koehenkilöistä 2 koki puhuvansa hämäläismurretta ja 1 itäistä murretta. Loput kaksi eivät kokeneet puhuvansa mitään tiettyä murretta.

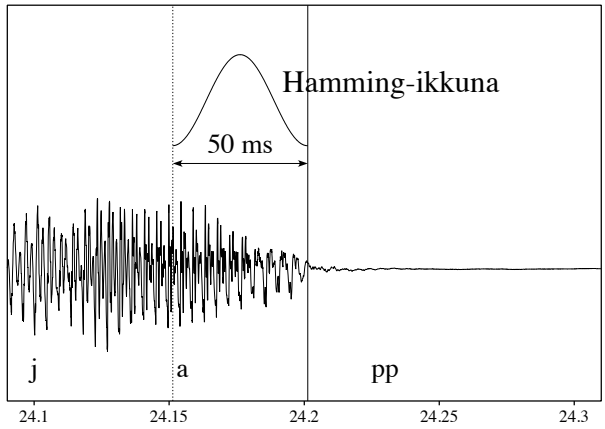
Taulukko 1: Koesanat kehyslauseessa

	japa	jappa	
	jata	jatta	
	jaka	jakka	
Onkos	_____		tuttu sana?

Nauhoitustilanteessa koehenkilöt äänsivät lyhyitä ja pitkiä klusiileja sisältäviä epäsanoja: *japa*, *jappa*, *jata*, *jatta*, *jaka*, *jakka*. Sanat oli sijoitettu kehyslauseeseen *Onkos _____ tuttu sana?* (vrt. taulukko 1). Kukin koehenkilö luki yhteensä 60 lausetta (6 koesanaa × 10 toistoa) paperista, jolle lauseet oli sijoitettu satunnaisjärjestyksessä. Koehenkilöitä ohjeistettiin lukemaan lauseet ilman sanojenvälisiä miettimistaukoja. Heille kerrottiin myös, että on mahdollista sanoa lause uudestaan, jos lauseen aikana tulee takeltelua, yskimistä tai muuta vastaavaa. Kukaan koehenkilöistä ei kuitenkaan tarvinnut tätä mahdollisuutta.

Kun nauhoituksia tarkasteltiin jälkeenpäin, huomattiin, että yksi koehenkilöistä oli ääntänyt yhdessä lauseessa *jaka*, vaikka tekstissä luki *jakka*. Varmuuden vuoksi tämä lause jätettiin pois akustisista mittauksista. Tämän vuoksi mitattavia tapauksia oli yhteensä 299.

Aineiston käsittelyyn käytettiin Praat-ohjelmaa (Boersma & Weenink 2007). Lauseet siirrettiin digitaalisessa muodossa (44,1 kHz) DAT-nauhalta tietokoneeseen. Aluksi merkittiin tutkittavien klusiilien alku- ja loppukohdat. Alkukohdaksi tulkittiin kohta, jossa osasävelrakenne vaikutti hajoavan. Loppukohdaksi merkittiin kohta, jossa sointi alkoi uudelleen. Seuraavaksi erotettiin signaalista okkluusion alkua edeltävä 50 millisekunnin jakso. Erottamiseen käytettiin Hamming-ikkunaa (ks. kuva 2). Tästä 50 millisekunnin jaksosta laskettiin spektri, josta laskettiin edelleen H1 − H2-arvo desibeleinä. Spektrin ensimmäinen ja toinen osasävel haettiin automaattisesti Praat-skriptillä, joka ensin arvioi laskettavan jakson keskimääräistä F0:aa (alueelta 75–600 Hz). Osasävelien tarkat sijainnit ja dB-lukemat skripti löysi etsimällä spektrin paikallista huippua arvioitua taajuutta ympäröivän 100 Hz:n alueelta. Näin saatuja H1 − H2-arvoja analysoitiin käyttäen perinteistä varianssianalyysia (ANOVA) sekä bayesiläistä analyysia.



Kuva 2: Esimerkki mitattavan jakson erottamisesta

3 Tulokset

3.1 Perinteinen ANOVA

Perinteisen ANOVA:n tulokset esitetään taulukossa 2. Koehenkilö (lyhenne KH) on satunnaisvaikutus. Ääntöpaikka (ÄP) ja kvantiteetti (KV) ovat kiinteisiä vaikutuksia.

Merkitsevin vaikutus varianssianalyysissa oli koehenkilöiden väliset erot (KH, $F(4,269) = 58.4035$, $p < 0.0001$). Merkitseviä olivat myös koehenkilön ja kvantiteetin interaktio (KH×KV, $F(4,269) = 3.7283$, $p < 0.01$) sekä ääntöpaikan ja kvantiteetin interaktio (ÄP×KV, $F(2,8) = 5.5061$, $p < 0.05$). Muut vaikutukset eivät olleet merkitseviä ($p > 0.05$).

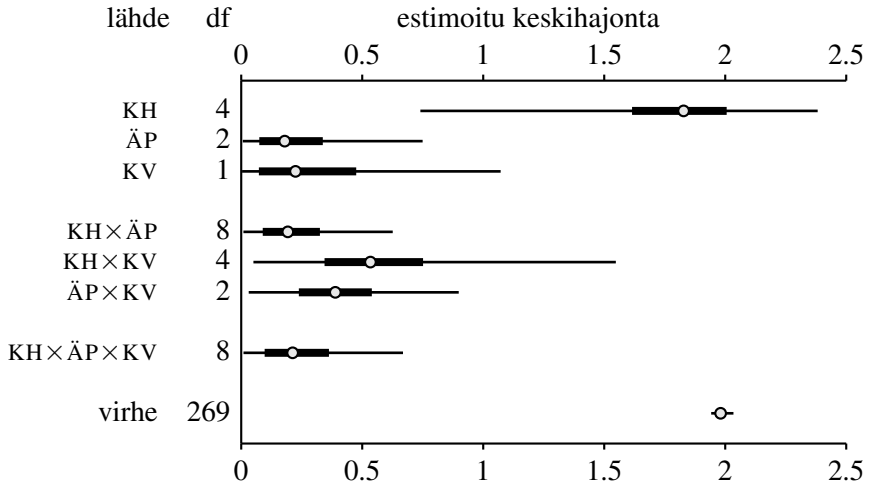
3.2 Bayesiläinen analyysi

Bayesiläinen ANOVA-malli (Gelman 2005a) sovitettiin dataan WinBUGS-ohjelmalla (Spiegelhalter *et al.* 2005). Ryhmien pienen koon takia parametriryhmien (vaikutusten) keskihajontojen priorina käytettiin puoli-Cauchy-jakaumaa (Gelman 2005b). Kuvassa 3 esitetään vaikutusten keskihajontien estimaattien mediaanit (pienet ympyrät), 50 % -välit (paksummat viivat) ja 95 % -välit (ohuemmat viivat).

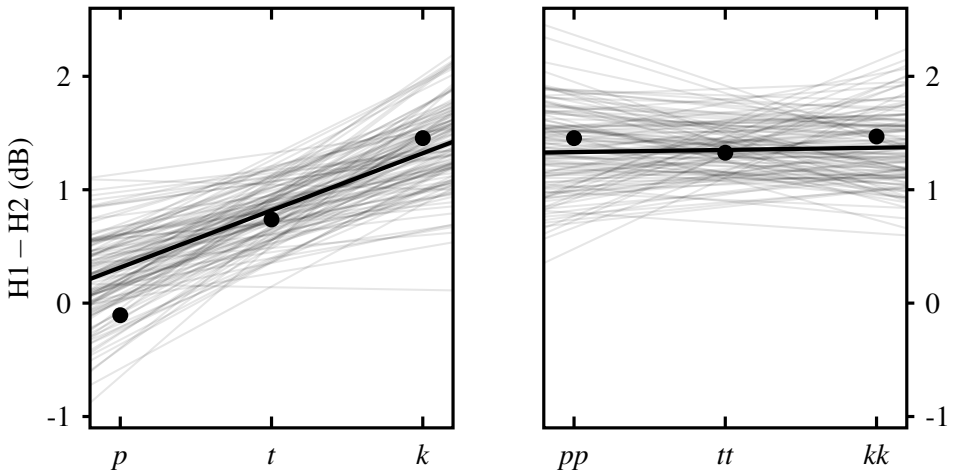
Tulokset sopivat yhteen perinteisen ANOVA:n kanssa. Suurin vaikutus oli koehenkilöiden väliset erot (KH). Koehenkilön ja kvantiteetin interaktio (KH×KV) oli myös merkitsevä. Selittämätön variaatio (virhe) oli verraten suuri — samaa luokkaa kuin koehenkilöiden välinen variaatio.

Taulukko 2: H1 – H2:n ANOVA-tulokset

	df	neliösumma	keskineliö	F-ratio	p-arvo
KH	4	919.335	229.834	58.4035	$<$ **0.0001**
ÄP	2	27.374	13.687	2.5873	0.1360
KV	1	45.754	45.754	3.1185	0.1522
KH×ÄP	8	42.322	5.290	1.3443	0.2216
KH×KV	4	58.687	14.672	3.7283	**0.0057**
ÄP×KV	2	35.302	17.651	5.5061	**0.0313**
KH×ÄP×KV	8	25.646	3.206	0.8146	0.5902
virhe	269	1058.590	3.935		
yhteensä	298	2213.010			



Kuva 3: Bayesiläisen ANOVA-mallin vaikutusten keskihajontaestimaattien mediaanit, 50 % -välit ja 95 % -välit



Kuva 4: Ääntöpaikan lineaariset kontrastit lyhyissä klusiileissa (vasemmalla) ja pitkissä (oikealla). Tummat viivat osoittavat mediaaneja, vaaleat taustaviivat kontrastien epävarmuutta ja mustat pisteet koko aineiston keskiarvoja.

Ääntöpaikan ja kvantiteetin päävaikutukset eivät olleet kovin suuria, mutta niiden interaktio oli merkitsevä. Toisin sanoen molemmat vaikuttivat H1 – H2:een, mutta toisistaan riippuvalla tavalla.

Interaktio selittyy siten, että ääntöpaikka vaikutti H1 – H2-arvoon merkitsevästi lyhyen klusiilin tapauksessa, mutta ei pitkän klusiilin tapauksessa. Ääntöpaikan vaikutukset (lineaariset kontrastit) esitetään kuvassa 4. Lyhyissä klusiileissa ääntöpaikan vaikutuksen (*p–k*-erotus tai kulmakerroin) mediaani oli 1.01 dB, 95 % -väli (0.03117, 2.041), kun taas pitkissä klusiileissa ääntöpaikan vaikutuksen (*pp–kk*-erotuksen) mediaani oli 0.0367 dB, 95 % -väli (-0.7962, 0.9049).

4 Johtopäätökset

Tutkimuksessa havaittiin, että ääntöpaikka vaikutti lyhyttä klusiilia edeltävän vokaalin H1 – H2-arvoon. Tämä tuo mieleen Suomen mainitseman ääntöpaikan vaikutuksen soinnin kestoon okkluusion alussa (Suomi 1980). Ilmiöiden taustalla saattavat mahdollisesti olla samat syyt, mikäli nopeammin tasaantuva transglottaalinen paine takaisemman klusiilin sulkeuman muodustuessa pyrkisi kasvattamaan OQ:ta ja H1 – H2-arvoa.

Geminaattaklusiilien kohdalla H1 – H2 oli verraten suuri kaikissa ääntöpaikoissa, mikä viittaa siihen, että niissä sointi katkaistaan aktiivisesti. Tulokset siis tuke-

vat hypoteesia, jonka mukaan äänihuulia avataan aktiivisesti tavunloppuisen klusiilin yhteydessä ja tämä avautuminen vaikuttaa soinnin laatuun jo edeltävän vokaalin lopussa.

Mikäli tämä ero vokaalin laadussa on kuultavissa, se on teoriassa myös opittavissa, esim. kielen omaksumisessa. Tämän olettamuksen mukaan suomenkielinen lapsi on siis oppinut avaamaan äänihuulensa aktiivisesti tavunloppuisessa klusiilissa pääasiassa sen takia, että lopputulos silloin kuulostaa oikealta. Jatkossa olisi siis ensisijaisen tärkeää testata ilmiön kuultavuutta myös empiirisesti havaintokokeilla.

Jatkossa olisi myös syytä tutkia ilmiötä suuremmalla aineistolla erityisesti siksi, että koehenkilöiden välinen variaatio oli suurta. Olisi mielenkiintoista kokeilla myös muita mittareita kuin H1 − H2-arvo sekä muita konteksteja kuin *ja___a*. Lisäksi voisi tutkia muitakin kieliä ja myös murre-eroja suomessa. On jo jonkin verran näyttöä siitä, että muistakin kvantiteettikielistä löytyisi vastaavia ilmiöitä (ks. esim. Ridouane *et al.* 2006). Mutta, jos kyseessä on opittu käyttäytyminen (eikä esim. fysiologinen välttämättömyys), ilmiön pitäisi kuitenkin olla kielikohtainen ja mahdollisesti myös murrekohtainen.

Viitteet

BOERSMA, Paul & WEENINK, David 2007: Praat: doing phonetics by computer (version 4.6.02) [tietokoneohjelma]. Haettu 18.5.2007 osoitteesta http://www.praat.org/.

GELMAN, Andrew 2005a: Analysis of variance—why it is more important than ever (with discussion). – *The Annals of Statistics*, **33**(1):1–53.

GELMAN, Andrew 2005b: Prior distributions for variance parameters in hierarchical models. – *Bayesian Analysis*, **1**(2):1–19.

HANSON, Helen M. 1997: Glottal characteristics of female speakers: Acoustic correlates. – *Journal of the Acoustical Society of America*, **101**(1):466–481.

HANSON, Helen M. & CHUANG, Erika S. 1999: Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. – *Journal of the Acoustical Society of America*, **106**(2):1064–1077.

IIVONEN, Antti 1975: Ääniraon avauma-asteen suuruudesta suomen konsonanteilla. – *Fonetiikan paperit — Helsinki 1975*, Publications of the Institute of Phonetics 27. University of Helsinki. 43–61.

NÍ CHASAIDE, Ailbhe & GOBL, Christer 1993: Contextual variation of the vowel voice source as a function of adjacent consonants. – *Language and Speech*, **36**(2, 3):303–330.

RIDOUANE, R., FUCHS, S. & HOOLE, P. 2006: Laryngeal adjustments in the production of voiceless obstruent clusters in Berber. – J. Harrington & M. Tabain (toim.), *Speech Production: Models, Phonetic Processes and Techniques*. Sydney, Australia: Psychology Press. 249–267.

SPIEGELHALTER, David, THOMAS, Andrew, BEST, Nicky & LUNN, Dave 2005: *WinBUGS User Manual, Version 2.10*. Cambridge: Medical Research Council Biostatistics Unit.

SUOMI, Kari 1980: *Voicing in English and Finnish Stops: A Typological Comparison with an Interlanguage Study of the Two Languages in Contact*. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 10. University of Turku.

Osallistujat

Olli Aaltonen
Eija Aho
Lotta Alivuotila
Paavo Alku
Heljä Asikainen
Reijo Aulanko
Päivikki Eskelinen-Rönkä
Dennis Estill
Jarmo Haapaharju
Jussi Hakokari
Kirsi Harinen
Terhi Hautala
Ville Hautamäki
Elina Helander
Antti Iivonen
Jouni Isoaho
Hennariikka Kairanneva
Heini Kallio
Tomi Kinnunen
Emmi Kivistö
Anna-Maija Korpijaakko-Huuhka
Maria Kunnas
Kaisa Kurki
Annamari Lahtinen
Anna Lantee
Anne-Maria Laukkanen
Mona Lehtinen
Heidi Lehtola
Kari Leinonen
Mietta Lennes
Tanja Makkonen
Einar Meister
Lya Meister
Elinita Mäki
Tuija Niemi-Laitinen
Tommi Nieminen
Mirja Nissinen
Michael O'Dell
Stina Ojala
Jan Paakkanen
Annu Paganus
Pertti Palo
Heidi-Maria Puolakka
Tuomo Raitio
Riikka Ruonamo
Veera Ryynänen
Minna Saari
Tuomo Saarni
Tapio Salakoski
Janne Savela
Hanna Silén
Tuire Silvennoinen
Antti Suni
Kari Suomi
Suvi Syrjänen
Henna Tamminen
Elina Tergujeff
Minnaleena Toivola
Riikka Ullakonoja
Teija Waaramaa
Leena Wahlberg
Martti Vainio
Annukka Vanhanen
Matti Varjokallio
Veijo Vihanta
Eeva Yli-Luukko
Riikka Ylitalo