



Turun yliopisto
University of Turku

LAB
lab

XXVIII FONETIIKAN PÄIVÄT

Turku 25.-26. lokakuuta 2013

Konferenssijulkaisu



FONETIIKKA, LAB-LABORATORIO, TURUN YLIOPISTO

XXVIII FONETIIKAN PÄIVÄT

Turku 25.–26. lokakuuta 2013

Konferenssijulkaisu

Katri Jähi ja Laura Taimi (toim.)

Turun yliopisto

Fonetiikka, Kieli- ja käännöstieteiden laitos 2014

ISBN 978-951-29-5980-8

Taitto: Katri Jähi

Kannen kuva: Katri Jähi

Sisällys

Alkusanat

Pekka Lintunen

Kuinka kielenoppija eroaa foneetikosta? Ääntämistaidon opettamisesta ja itsearviointista 1

Riikka Ullakonoja, Hannele Dufva, Mikko Kuronen, Pertti Hurme

How to imitate an unknown language? Russians imitating Finnish..... 10

Katrin Leppik, Pärtel Lippus

A comparative study of Estonian and Spanish vowels in L1 and L2 production..... 19

Kätlin Aare, Pärtel Lippus, Juraj Šimko

Creaky voice in spontaneous spoken Estonian 27

Juraj Šimko, Michael O'Dell, Martti Vainio, Mona Lehtinen & Tommi Nieminen

Effects of anatomical differences on relative timing of tongue and lips..... 36

Dennis Estill

Causal factors and indicators of word stress in Meadow Mari..... 46

Tommi Nieminen, Tommi Kurki, Hamid Behravan, Heini Kallio

Tutkimushanke suomen prosodian variaatiosta..... 56

Terhi Peltola, Pertti Palo, Olli Aaltonen

A short term study of Hungarians learning Finnish vowels 65

Heete Sahkai, Mari-Liis Kalvik, Meelis Mihkla

Perception of focus size and focus type in Estonian 72

Antti Saloranta, Henna Tamminen, Paavo Alku & Maija S. Peltola

Foneettiset ääntämisohjeet osana kuuntele ja toista –harjoitusta..... 80

Joonas Vakkilainen, Michael O'Dell

Tremulantin kvantiteetti suomessa..... 89

XXVIII FONETIIKAN PÄIVÄT

Turku 25.–26. lokakuuta 2013

Fonetiikan päivät järjestysnumerolla 28 pidettiin syksyisessä Turussa lokakuun lopulla 2013. Mukana oli 20 esitelmää sekä 12 posteria; osallistujia ilmoittautui yli 80. Esitelmissä ja postereissa näkyi jälleen alamme laaja kirjo, sillä aiheet vaihtelivat puhesynteesin erityispiirteistä intonaation analyysiin, mutta tarjolla oli erityisen runsaasti oppimistutkimuksia ja kontrastiivisia aiheita. Ehkäpä painottumiseen oli ainakin osasyynä Turun fonetiikka, joka on viime vuosina voimakkaasti profiloitunut juuri oppimistutkimukseen. Tässä julkaisussa esitellään otos tämänkertaisista esitelmistä ja postereista – teemapainotus näkyy myös teoksen sisällössä, sillä mukana on useita oppimistutkimuksia.

Meillä turkulaisilla oli myös ilo esitellä uutta Sanako Oy:n tuella perustettua Learning, Age and Bilingualism –laboratoriotamme (LAB-lab), joka sijaitsee kauniissa tiloissamme Turun yliopiston kampuksen keskellä vanhassa kuurojen koulussa.

Virallisen ohjelman lisäksi ehdimme hieman myös rentoutua posterisession jälkeen järjestetyssä cocktail-tilaisuudessa sekä Panimoravintola Koulun illallisella. Tunnelmaa voisi kuvailla rennoksi ja keskustelut jatkuivat vilkkaina lähes tauotta seuraavaan seminaaripäivään. Kyseessä oli siis jälleen kerran mukava tilaisuus yhdistää tiedettä ja iloisia jälleennäkemisiä.

Haluan kiittää kaikkia osallistujia sekä kirjoittajia siitä panoksesta, jonka annoitte Turun Fonetiikan päiville ja koko suomalaiselle puheentutkimukselle. On ilo työskennellä tällaisessa joukossa ja odotan innolla seuraavia, Otaniemessä järjestettäviä päiviä. Otaniemessä tavataan!

Turussa, Signumin kellarissa, 18.11.2014

Maija S. Peltola

Kuinka kielenoppija eroaa foneetikosta? Ääntämistaidon opettamisesta ja itsearviinnista

Pekka Lintunen

Englannin kieli, Turun yliopisto

pekka.lintunen@utu.fi

Yksi vieraan kielen oppimisen tavoitteista on kasvattaa oppijoiden tietoisuutta kohdekielen fonologiasta ja fonetiikasta. Tämä on erityisen tärkeää suomalaisessa kontekstissa, jossa äidinkielen oikeinkirjoituksella ja ääntämisellä on läheinen suhde, mutta opittavassa vieraassa kielessä näin ei usein ole. Tietoisuus ja uusien piirteiden havaitseminen on nostettu vieraan kielen oppimisen teorioissa tärkeäksi oppimista auttavaksi tekijäksi, jopa oppimisen edellytykseksi. Oppijoiden foneettista tietoisuutta voi kehittää monin tavoin. Yliopistotasolla tämä tehdään usein eksplisiittisesti opettamalla teoreettisia kursseja kohdekielen äännejärjestelmästä tai järjestämällä ääntämiskursseja harjoitusryhmissä. Tutkimuksessa koehenkilöitä pyydettiin arvioimaan ääntämisongelmiaan englannin kielessä. Sen jälkeen koehenkilöt osallistuivat sekä teoreettisiin luentokursseihin fonetiikasta että käytännölliseen ääntämiskurssiin. Lisäksi koehenkilöt osallistuivat ääntämistestiin. Tutkimuksessa verrataan oppijoiden itsearviointeja ääntämisongelmista heidän ääntämistituloksiinsa. Lisäksi verrataan itsearviointeja ennen ja jälkeen fonetiikan ja ääntämisen opetusta. Tulosten mukaan koehenkilöt arvioivat ääntämisongelmiaan lähinnä sanatasolla. Arviot paljastavat myös ongelma-alueita, joihin ei yleensä kiinnitetä paljon huomiota, kun suomenkielisiä oppijoita opetetaan ääntämään englantia. Tulosten perusteella voidaan argumentoida oppijoiden analysoivan omaa ääntämistään erilailla kuin mihin ääntämisen tai fonetiikan opetus yleensä keskittyy.

Asiasanat: vieraan kielen ääntäminen, itsearviointi, fonetiikka

1. Johdanto

Tämä artikkeli raportoi tutkimuksesta, jossa verrattiin englannin yliopisto-opiskelijoiden mielipidettä heille vaikeista englannin äänneistä heidän ääntämistaitoonsa ja itsearviinnissa tapahtuneita muutoksia eksplisiittisen fonetiikan ja ääntämisen opettamisen jälkeen. Artikkelissa pohditaan myös tutkimuksen tuloksia ja niiden implikaatioita vieraan kielen ääntämisen ja fonetiikan opettamisen kannalta. Tutkimuksen alkuperäinen idea syntyi Szpyra-Kozłowskan (2010, 2012) tutkimussarjan pohjalta. Hän oli kiinnostunut puolalaisten englannin oppijoiden ääntämisestä ja erityisesti sanoista, jotka oppijat olivat oppineet virheellisessä foneemisessa muodossa. Szpyra-Kozłowska (2010) analysoi 250 oppijoiden mainitsemaa vaikeaa sanaa. Oletuksen mukaisesti eniten vaikeuksia tuottivat sanapaino ja konsonanttiyhdistelmät sekä sanat, joissa oikeinkirjoituksen ja

ääntämisasun suhde ei ole selvä (esim. *surface*), ja sanat, jotka olivat muodoltaan mutta eivät ääntämisasultaan samanlaisia äidin- ja kohdekielellä (esim. useat maantieteelliset nimet kuten *Japan*). Yllättävimpinä havaintoina oli, että ääntämistä vaikeuksia aiheuttivat myös pitkät sanat, sanat, joissa oli suppeita etuvokaaleja tai monta likvidääntä, sekä sanat, joiden ääntämisessä oli morfofonologista vaihtelua (esim. *south/southern*).

Ääntäminen on yksi englannin kielen haastavimpia piirteitä suomenkieliselle englannin oppijalle. Englannin kohdalla tilannetta sekoittaa edelleen englannin asema maailmankielenä. Englannin kieltä puhutaan ensimmäisenä ja vieraana kielenä ympäri maailmaa (näkökulmasta riippuen voidaan puhua myös ensimmäisen, toisen ja vieraan kielen puhujista). Siksi on vaikea määrittää, millaista englantia oikeastaan on, koska puhujia ja potentiaalisia malleja oppijoille on monia. Onkin siis vaikea sanoa nykymaailmassa, milloin englantia äännetään oikein, koska myös äidinkielisten puhujien englantia vaihtelee niin suuresti. Tärkeä ja tarpeellinen keskustelu englannin ääntämismalleista ja tarkkuudesta vieraan kielen ääntämisen vaatimuksena jatkuu Suomessa ja maailmalla (esim. Jenkins 2000, Dzuibalska-Kołodziej & Przedlacka 2005).

Laajasta variaatiosta huolimatta yksi englannin ääntämisen haasteista kuitenkin koskettaa erityisesti suomenkielisiä englannin oppijoita: englannin ja suomen oikeinkirjoituksen suhde kielen foneemiseen tasoon on hyvin erilainen. Suomessa ääntämisen ja ortografian suhde on läpinäkyvä, vaikkakin tiettyjen foneettisten poikkeusten vuoksi ei voida sanoa, että foneemit järjestelmällisesti vastaavat grafeemeja (esim. Karlsson 1982: 77–79; Suomi et al. 2008: 141). Englannissa ortografian ja äänteiden suhde on kuitenkin monimutkaisempi. Tästä syystä äidinkieleltään suomalaisille oppijoille on erityisen tärkeää, että vieraan kielen oppimisessa ja opettamisessa huomioidaan myös kielen foneeminen ja foneettinen taso, koska ortografian on havaittu vaikuttavan käsityksiin ääntämisestä ja sanojen foneemisesta rakenteesta (esim. Celce-Murcia, Brinto & Goodwin 2010: 419–447; Iivonen 1998). Erityisesti suomalaiset oppijat saattavat analysoida kohdekieltä kirjoituksen kautta ja esimerkiksi tallentaa vieraan kielen sanoja muistiinsa enemmän niiden ortografisessa kuin foneettisessa muodossa (esim. Ringbom 1977: 106). Järjestelmällisesti kohdekielen fonetiikkaa ja ääntämistä opetetaan Suomessa vasta yliopistotasolla. Tätä alemmilla tasoilla opettajan ja oppikirjan vaikutus valintoihin on huomattava (Tergujeff 2013).

Ääntämisen oppimisen kannalta fonetiikan opettaminen auttaa oppijan tietoisuuden lisääntymisessä. Tämä on erityisen tärkeää, kun oppijoiden äidinkielen järjestelmä on läpinäkyvämpi foneemien ja grafeemien välisissä suhteissa kuin kohdekielen järjestelmä, esimerkiksi kun suomenkielinen oppija oppii englantia. Vieraan kielen oppimisessa tietoisuus ja opittavien asioiden ja rakenteiden

tiedostaminen ja tunnistaminen on merkittävässä asemassa (esim. Schmidt 1990). On todettu (esim. Ortega 2009; Van Lier 2001), että sellaiset asiat opitaan nopeammin, joihin myös kiinnitetään huomiota (*noticing*). Toisaalta ajatellaan, että jos jokin piirre kohdekielessä jää oppijalta huomaamatta, sitä ei myöskään ole helppo oppia (vaikkakin kielenoppimisessa omaksumista voi tapahtua myös tiedostamatta). Ääntämisen oppimisen kannalta sama idea toistuu perinteissä ajattelussa, että havainto edeltää tuotosta. Hyödyllinen opetus tähtääkin siihen, että oppija osaa opettamisen tukemana huomioida kohdekielen piirteitä, jotka hän on kenties ennen sivuuttanut (esim. Dalton & Seidlhofer 1994; Seidlhofer 2001). Vieraan kielen oppimisen teorioiden mukaisesti tällöin eksplisiittisellä opetuksella on vaikutus oppimisnopeuteen ja tarkkuuteen, vaikkakin oppimisjärjestys tai kohdekielen vaikeimmiksi koetut piirteet yleensä pysyvätkin samana (esim. Ortega 2009: 139–140). Aiemmissä tutkimuksissa on todettu, että fonetiikan opettaminen parantaa vieraan kielen ääntämisen oppimista (esim. Aliaga-Garcia 2007; Lord 2005; Pourhosein 2012). Foneettisten käsitteiden ja ilmiöiden opettaminen saa oppijan tietoisemmaksi kohdekielen foneettisesta tasosta, jota oppija pystyy hyödyntämään oppimisprosessissaan (esim. Lintunen 2005). Siksi foneetikot ovat usein myös huolissaan foneettisen tietouden vähyydestä kielenopetuksessa ja opettajainkoulutuksessa (esim. Derwing 2009; Henderson et al. 2012; Iivonen & Tella 2009: 269).

2. Tutkimuksen toteutus

Tutkimus toteutettiin kolmessa osassa. Koehenkilöt osallistuivat ääntämistestiin ja vastasivat kahteen kyselyyn foneettisesti vaikeista sanoista. Ääntämistesti pidettiin ennen opetuksen alkua. Ensimmäinen kysely pidettiin ensimmäisellä opetusviikolla syyslukukaudella ja jälkimmäinen kevätlukukauden alussa. Testi tehtiin kahdelle vuosikurssille. Ääntämistestin tarkoituksena oli selvittää, kuinka oppijoiden mielipiteet vastasivat heidän ääntämisvaikeuksiaan. Koehenkilöt olivat ensimmäisen vuoden englannin kielen sivu- tai pääaineopiskelijoita yliopistossa. Jokaisen äidinkieli oli suomi. Yhteensä koehenkilöitä oli 156.

Ääntämistestissä oppijoiden piti lukea ääneen lyhyt teksti ja sanalista. Testissä oli mukana kaikki englannin foneemit, ja sanalistassa painotettiin erityisesti sellaisia äänneitä, jotka on aiemmin todettu haastaviksi suomenkielisille englannin oppijoille (esim. Lintunen 2004; Morris-Wilson 2004). Testi pyrki olemaan äänneellisesti haastava, mutta sanastoltaan melko tavallinen. Koehenkilöille annettiin 10 minuuttia aikaa harjoitella ja tutustua sanoihin ja tekstiin. Tämän jälkeen heidän tuotoksensa äänitettiin. Koehenkilöjoukosta valittiin satunnaisotannalla 69 opiskelijaa tutkimuksen tähän osaan. Tulokset arvioi auditiivisesti kaksi kokenutta ääntämisen opettajaa.

Koehenkilöiden mielipiteet foneettisesti vaikeista sanoista kerättiin kahden kyselylomakkeen avulla (kuten Szpyra-Kozłowska 2010). Koehenkilöitä pyydettiin luettelemaan sanoja, joiden ääntäminen oli heistä vaikeaa, ja alleviivaamaan sanasta se kohta, joka on vaikea ääntää. Tällä tavalla sanoista pystyttiin tunnistamaan se äänne tai äännejono, joka oli koehenkilölle ongelmallinen. Ensimmäinen kyselylomake täytettiin syyslukukauden ensimmäisellä opetusviikolla ja toinen kevätlukukauden ensimmäisellä opetusviikolla. Opetussuunnitelman mukaisesti opiskelijoiden piti syyslukukauden aikana suorittaa Foneetiikan perusteiden luentokurssi (2 op) ja englanninkielinen English Phonetics -luentokurssi (2 op). Tämän lisäksi suurimman osan opiskelijoista piti suorittaa englannin ääntämisen harjoituskurssi (4 op), jossa oli yhteensä 45 tuntia kontaktiopetusta. Jälkimmäisessä kyselyssä opiskelijoiden siis oletettiin olevan tietoisia englannin kielen äännejärjestelmästä ja omasta ääntämisestään. Ensimmäisessä kyselyssä taas oletettiin, että tietoisuus ei ole yhtä vahva. Jälkimmäisestä kyselystä huomioitiin vain niiden opiskelijoiden vastaukset, jotka olivat opiskelleet opetussuunnitelman mukaisesti. Jälkimmäisestä kyselystä otettiin mukaan 78 koehenkilöä, mikä tarkoittaa, että tasan 50 % ensimmäisen kyselyn osallistujista ei osallistunut toiseen kyselyyn tai he eivät olleet opiskelleet opetussuunnitelman mukaisesti. Kyselyt olivat nimettömiä, joten keskityn jatkossa koko ryhmän tuloksiin. Yksittäisten oppijoiden kehityksen ja itsearvioinnin vertailu olisi vaatinut erilaisen metodologisen ratkaisun.

3. Tulokset

3.1 Ääntämisvirheet

Ääntämisvirheet perustuvat 69 opiskelijan tuotosten auditiiviseen analyysiin. Ääntämisen arvioi kaksi kokenutta englannin ääntämisen opettajaa, joita pyydettiin kiinnittämään huomiota foneemeihin, foneemisiin oppositioihin ja muihin mahdollisiin ääntämisongelmiin. Sujuvuuteen ja prosodiaan liittyvät ongelmat rajattiin analyysin ulkopuolelle, koska niitä olisi vaikea osoittaa tämän tutkimuksen sanalistakyselyssä. Analyysin perusteella laadittiin äänneiden vaikeusjärjestys tälle koehenkilöjoukolla. Ääntämisvaikeudet on luokiteltu vaikeusjärjestyksessä taulukossa 1.

Taulukko 1. Todetut ääntämisvaikeudet.

		% koehenkilöistä (n=69)
1	v/w	69,6%
2	ʒ	63,8%
3	ŋg	58,0%
4	dʒ	52,2%
5	z	39,1%
6	θ, ð	34,8%
7	ʃ	24,6%
8	tʃ	23,2%
9	ɪ	14,5%
10	aspiraatio	13,0%

Kuten taulukko 1 osoittaa, ääntämisvaikeudet vastaavat melko tarkkaan aiempia tutkimuksia suomalaisten oppijoiden englannin oppimisesta. Suurin osa vaikeuksista liittyi konsonantteihin. Suurin vaikeus oli oppositio v/w. Myös sibilantit (paitsi /s/), affrikaatat ja dentaalifrikatiivit havaittiin tavallisesti vaikeiksi äänteiksi.

3.2 Vaikeat sanat ennen opetusta

Kun opiskelijoilta (n=156) kysyttiin vaikeista sanoista, 133 (85,3 %) pystyi antamaan vastaukseksi ainakin yhden sanan. Yhteensä mainittiin 423 sanaa (keskimäärin 2,7 sanaa per opiskelija). Sanoista analysoitiin ne äänteelliset kohdat, jotka oli alleviivaamalla osoitettu vaikeiksi. Kuten taulukko 2 osoittaa, tavallisimmin vaikeiksi äänteiksi mainittiin likvidat /l/ ja /r/. Kiinnostavasti Szpyra-Kozłowska (2010) sai saman tuloksen. Toinen vaikea piirre, jota ei todettu useasti ääntämistestissä, oli sanapaino. Tämä johtunee siitä, että arvioitsijoita pyydettiin jättämään prosodiset seikat huomioimatta ääntämisten arvioinnissa. Muuten vaikeina piirteinä mainittiin samoja äänteitä, oppositioita ja muita piirteitä kuin ääntämistestissäkin havaittiin. Lintunen (2013) raportoi tarkemmin yksittäisistä sanoista, jotka mainittiin vaikeiksi. Osan tuloksista selittää se, että tietyt sanat esiintyivät useammin kuin toiset.

Taulukko 2. Itsearvioidut ääntämisvaikeudet ennen opetusta.

		% koehenkilöistä (n=156)
1	l/r	28,2%
2	j	24,4%
3	dʒ	19,9%
4	sanapaino	18,6%
5	θ, ð	16,0%
6	z	12,2%
7	tʃ	11,5%
8	v/w	7,7%
9	aspiraatio	2,6%
10	ʒ	1,9%

3.3. Vaikeat sanat opetuksen jälkeen

Kun opiskelijoilta (n=78) kysyttiin vaikeista sanoista opetuksen jälkeen, 72 (92,3 %) vastasi ainakin yhden sanan. Yhteensä mainittiin 229 sanaa (keskimäärin 2,9 sanaa per opiskelija). Jälleen sanoista analysoitiin alleviivatut kohdat. Tavallisimmat vaikeudet on listattu taulukossa 3. Kuten taulukko osoittaa, likvidat /l/ ja /r/ mainittiin taas kaikkein tavallisimmin. Muuten piirteet vastasivat melko hyvin ääntämisongelmia. Kuten prosenttiluvut osoittavat, variaatio vastausten välillä oli hyvin laaja, koska tavallisimmankin piirteen mainitsi vain alle 25 % opiskelijoista.

Taulukko 3. Itsearvioidut ääntämisvaikeudet opetuksen jälkeen.

		% koehenkilöistä (n=78)
1	l/r	24,4%
2	v/w	22,0%
3	θ, ð	21,8%
4	j	16,7%
5	dʒ	15,4%
6	tʃ	14,1%
7	z	12,8%
7	sanapaino	12,8%
9	ʒ	11,5%
10	ŋg	3,8%

Jos verrataan vastauksia ennen ja jälkeen opetuksen, voidaan huomata, että likvidat on molemmissa testeissä arvioitu vaikeimmiksi äänteiksi. Opetuksen jälkeen opiskelijat mainitsivat v/w opposition, dentaalifrikatiivit ja soinnillisen palatoalveolaarisen sibilantin suhteellisesti useammin vaikeaksi kuin ennen opetusta. Tämä mahdollisesti heijastaa opiskelijoiden parempaa tietoisuutta heidän omasta ääntämisestään. On kuitenkin myös huomioitava, että kyselyiden välillä ei tapahtunut suuria eroja. Tämä saattaa johtua siitä, että opiskelijat olivat hyvin tietoisia ääntämisongelmistaan jo ennen opetusta tai että opetus pystyi vain rajallisesti muuttamaan opiskelijoiden mielipidettä lähemmäksi ääntämisestissä havaittuja ongelmia. Tarkemmassa analyysissa toki havaittiin myös, että jälkimmäisessä kyselyssä koehenkilöt keskittyivät enemmän äänne- kuin sanatasoon, mikä myös heijastaa opetuksen vaikutusta.

4. Pohdintaa tuloksista

Kun tutkimuksen tuloksia katsotaan kokonaisuutena, kysymys likvidoista herättää kysymyksiä. Suomessa ja englannissa on molemmissa /l/ ja /r/ foneemit, joissa on foneettisella tasolla paljon eroja kielten välillä, ja erityisesti allofonisella tasolla ne ovat haastavia suomenkielisille englannin oppijoille (esim. Morris-Wilson 2004). Näitä äänteitä ei kuitenkaan usein harjoitella ääntämisen opetuksessa, koska niiden mallista poikkeava laatu ei yleensä muuta sanan merkitystä. Toki nämä äänteet ovat foneettisesti lähellä toisiaan, ja opiskelijoiden vaikeudet ilmenivät erityisesti sanoissa, joissa oli monta likvidaa (esim. *rural*, *literally*), jotka ovat haasteellisia tuottaa tarkasti ja nopeasti. Ääntämisharjoituksissa keskitytään usein soinnillisiin tai aspiroituihin konsonantteihin ja vokaalin laatuun tai kohdekielen uusiin äänteisiin, usein vielä sanan eri asemissa. Harjoituksissa on kuitenkin usein keskiössä äänne, kun taas sanataso tai erilaiset äännekombinaatiomahdollisuudet sanoissa jäävät vähemmälle huomiolle. Onnistuneen oppimiskokemuksen kannalta on kuitenkin tärkeää, että opetuksen sisällöt vastaavat oppijoiden omaa näkemystä oppimisen haasteista. Siksi vieraan kielen ääntämisen opetuksen kannalta on erittäin tärkeää saada tietoa siitä, kuinka oppijat itse kokevat oman ääntämisensä. Oppijan näkemys prioriteeteista tai haastavista asioista saattaa erota fonetiikan ja ääntämisen opettamisen ammattilaisen näkemyksestä. Onnistunut oppimiskokemus vaatiikin dialogia oppijan ja oppimisen ohjaajan välillä. Oppija saattaa kaivata harjoitusta motorisesti haastavista tavallisista sanoista, kun taas foneetikko saattaa olla kiinnostuneempi siitä, että oppija hallitsee ja pystyy tuottamaan kohdekielen foneemi-inventaarion.

Tämän tutkimuksen tarkoituksena oli verrata oppijoiden itsearviointia ääntämisvaikeuksista ääntämistituloksiin ja verrata itsearvioinnin tuloksia ennen ja jälkeen ääntämisen ja fonetiikan opetuksen. Ääntämisen itsearviointi suoritettiin pyytämällä oppijoita listaamaan sanoja, jotka

tuntuvat heistä vaikeilta. Näin oppijoita ohjasi jo tehtävänasettelussa sanataso. Tästä rajoituksesta huolimatta on tärkeä huomata, että usein oppijat ajattelevat vaikeita sanoja kokonaisuuksina eivätkä he välttämättä osaa segmentoida niitä äänneiksi. Fonetikan ja ääntämisen opetuksen tavoitteena on lisätä tietoisuutta sanojen äännekuoista, jolloin oppijoiden olisi helpompi myös analysoida kielen foneemista tasoa. Tulosten perusteella opetuksella oli vaikutusta oppijoiden itsearviointiin. Vaikka suurta muutosta ei tapahtunut opetusjakson aikana, pitää myös muistaa, että opetuksella on vaikutuksia myös testiajankohdan jälkeen. On mahdollista, että oppijat oppivat vasta myöhemmin analysoimaan omaa ääntämistään tarkemmin. Tutkimuksen tärkein anti foneetikoille ja ääntämisen opettajille on se havainto, että oppijan oma näkemys ääntämisen vaikeuksista ei välttämättä aina vastaa opettajien näkemystä ääntämisen ongelmista.

Lähteet

Aliaga-García, C. (2007). The role of phonetic training in L2 speech learning. *Proceedings of the Phonetics Teaching and Learning Conference*. London: University College London. [Available: http://www.phon.ucl.ac.uk/ptlc/proceedings/ptlcpaper_32e.pdf]

Celce-Murcia, M., Brinto, D.M. & Goodwin, J.M. (2010). *Teaching pronunciation*. Cambridge: Cambridge University Press.

Dalton, C. & Seidlhofer, B. (1994). *Pronunciation*. Oxford: Oxford UP.

Derwing, T. (2009). Utopian goals for pronunciation teaching. In: Levis, J. & LeVelle, K. (eds.) *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference*. Ames: Iowa State University, pp. 24-37.

Dzuibalska-Kolaczyk, K. & Przedlacka, J. (2005). *English pronunciation models: a changing scene*. Bern: Peter Lang.

Henderson, A., Frost, D., Tergujeff, E., Kautzsch, A., Murphy, D., Kirkova-Naskova, A., Waniek-Klimczak, E., Levey, D., Cunningham, U. & Curnick, L. (2012). The English pronunciation teaching in Europe survey: selected results. *Research in Language*, 10, pp. 5-27.

Iivonen, A. (1998). Nykynäkymiä fonetiikan asemasta kielenomaksumisessa ja -oppimisessa. In: Iivonen, A. & Nevalainen, T. (eds.) *Vieraan kielen fonetiikan opetuksen näkökohtia*. Helsingin yliopiston fonetiikan laitoksen julkaisuja 41. Helsinki: Helsingin yliopisto, pp. 15-30.

Iivonen, A. & Tella, S. (2009). Vieraan kielen ääntämisen ja kuulemisen opetus ja harjoittelu. In: Aaltonen, O., Aulanko, R., Iivonen, A., Klippi, A. & Vainio, M. (eds.). *Puheviestinnän perusteet*. Helsinki: Otava, pp. 269-281.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.

Karlsson, F. (1982). *Suomen kielen äänne- ja muotorakenne*. Porvoo: WSOY.

- Lintunen, P. (2004). *Pronunciation and phonemic transcription: a study of advanced Finnish learners of English*. Turku: University of Turku.
- Lintunen, P. (2005). Ääntäminen, tietoisuus ja oppiminen. In: Virtanen, A., Merenluoto, K. & Pöyhönen, P. (eds.). *Ainedidaktiikan ja oppimistutkimuksen haasteet opettajankoulutukselle*. Turun yliopiston kasvatustieteiden tiedekunnan julkaisu B:75. Turku: Turun yliopisto, pp. 342-349.
- Lintunen, P. (2013). The effect of phonetic knowledge on evaluated pronunciation problems. In: Przedlacka, J., Maidment, J. & Ashby, M. (eds.) *Proceedings of the Phonetics Teaching and Learning Conference*. London: Phonetics Teaching and Learning Conference, pp. 55-58. [Available: http://www.ucl.ac.uk/psychlangsci/ptlc/proceedings_2013/proceedings_2013.pdf]
- Lord, G. (2005). (How) can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania*, 88, pp. 557-567.
- Morris-Wilson, I. (2004). *English segmental phonetics for Finns*. Oulu: Oulu University Press.
- Ortega, L. (2009). *Understanding second language acquisition*. London: Hodder Education.
- Pourhosein G.A. (2012). The significance of pronunciation teaching in English language teaching. *English Language Teaching*, 5, pp. 96-107.
- Ringbom, H. (1977). Spelling errors and foreign language learning strategies. In: Palmberg, R. & Ringbom, H. (eds.) *Papers from the Conference on Contrastive Linguistics and Error Analysis*. Turku: The Research Institute of the Åbo Akademi Foundation, pp. 101-111.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, pp. 129-158.
- Seidlhofer, B. (2001). Pronunciation. In: Carter, R. & Nunan, D. (eds.) *The Cambridge guide to teaching English to speakers of other languages*. Cambridge: Cambridge University Press, pp. 56-65.
- Suomi, K., Toivanen, J. & Ylitalo, R. (2008). *Finnish sound structure. Phonetics, phonology, phonotactics and prosody*. Oulu: University of Oulu.
- Szpyra-Kozłowska, J. (2010). Phonetically difficult words in intermediate learners' English. In: Dziubalska-Kołaczyk, K., Wrembel, M. & Kul, M. (eds.) *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010*, pp. 481-486. [Available: http://ifa.amu.edu.pl/newsounds/files/proceedings/proceedings_quotable_version.pdf]
- Szpyra-Kozłowska, J. (2012). Mispronounced lexical items in Polish English of advanced learners. *Research in Language*, 10, pp. 243-256.
- Tergujeff, E. (2013). *English pronunciation teaching in Finland*. Jyväskylä: University of Jyväskylä.
- Van Lier, L. (2001). Language awareness. In: Carter, R. & Nunan, D. (eds.) *The Cambridge guide to teaching English to speakers of other languages*. Cambridge: Cambridge University Press, pp. 160-165.

How to imitate an unknown language? Russians imitating Finnish

Riikka Ullakonoja¹, Hannele Dufva², Mikko Kuronen², Pertti Hurme³

¹ Centre for Applied Language Studies, University of Jyväskylä, ² Department of Languages, University of Jyväskylä, ³ Department of Communication, University of Jyväskylä

riikka.ullakonoja@jyu.fi, hannele.dufva@jyu.fi, mikko.j.kuronen@jyu.fi, pertti.hurme@jyu.fi

The study focuses on how Russians without any knowledge of Finnish imitate Finnish three-word utterances. Perceptual judgments of utterance and word comprehensibility of the imitations were compared to the subjects' results in a working memory test. The results showed differences between the utterances but also between the subjects. A strong correlation between the score of the working memory test and comprehensibility of the imitated utterance was found. Further, the initial word of the utterance was the easiest to imitate. The results suggest a need for further studies.

Keywords: imitation, Finnish, Russian

1. Background

This paper reports findings of an experiment in which Russian students imitated read-aloud utterances of Finnish, a language they did not have any former knowledge of. The experiment was carried out as a pilot study for investigating the ability to imitate an unknown language, that is, an ability to perceive and replicate *what* and *how* is said. Evolutionarily old (Donald, 1991), and connected with the workings of mirror neurons (Arbib, 2012), the ability to imitate is manifest in a wide range of mimicking behaviours, ranging from unconscious processes of accommodation in social interaction (Giles et al, 1991) to intentional uses for parody and entertainment (Zetterholm, 2006). There is a wide range of studies on imitation ranging from neurosciences, cognitive science and phonetics (e.g. Alivuotila et al, 2008) to social sciences (Krauss and Pardo, 2006).

The views on the role of imitation in first and second/foreign language learning have varied over the years. While behaviorism underlined its centrality and the Chomskyan, cognitivist views denied its importance, the sociocultural thinking has continued to see imitation as an essential element of learning (e.g. Lantolf and Thorne, 2006: 166-176). Further, if imitation turns out to be an important tool for learning, this is a highly relevant finding for second/foreign language teaching as well (Wik and Preben, 2010; Chappell, 2012).

There are few studies on the correlation between L2 pronunciation and the aptitude for oral mimicry. Purcell and Suter (1980) carried out a statistical reanalysis of Suter's earlier study (1976) in which 14 native judges rated the pronunciation accuracy of 61 non-native speakers of English. Only four of the 20 variables had significance for pronunciation accuracy. The subject's first language and his/her aptitude for oral mimicry were the most important predictors.

While a good number of studies related to imitation and repetition of speech have been carried out (e.g. Gathercole, 2006; Peschke et al, 2009), experiments where an unknown language is imitated are scarce - and the existing studies often deal with vowels only (e.g. Alivuotila et al, 2008). As early as the 1960s, the research group at the Pavlov Institute of Physiology, Academy of Sciences of the USSR, Leningrad, conducted experiments in speech imitation and shadowing (Kozhevnikov and Chistovich, 1967). Inspired by their research, Hurme (1975) designed a study to find out how well people can imitate a language unknown to them. Speakers of Russian were asked to imitate recorded Finnish sentences consisting of 5 to 9 syllables, immediately after hearing them through headphones. The results show that the length of the utterance (as measured by the number of syllables) was clearly related to the quality of the imitation (short utterances imitated better). Also, in longer utterances the initial and final parts of utterances were imitated better than those in-between. Further, while the imitations showed some Russian characteristics, the subjects were fully able to produce non-Russian features, often rather Finnish-sounding.

The current experiment repeats the design of Hurme (1975) with an additional working memory task. As the subjects did not have any formal knowledge of the language they were asked to imitate – neither its phonological system nor its phonetic aspects – they supposedly need to rely their capacity of imitation without resorting to any L2 knowledge. Below, we will report results on the success of the subjects' performance as evaluated by Finnish judges, discussing also the relationship between the subjects' performance and the results of their working memory test. Finally, we will also shortly comment on the possible role of L1.

2. Material and methods

2.1. Tasks

The data was elicited in a task, where the participants were asked to imitate a set of 30 auditory stimuli. The responses were recorded with an Edirol by Roland R-09 digital recorder and a high quality Koss headset (sample rate 44.1 kHz, 16 bit resolution). The stimuli were three-word Finnish utterances (with 4–11 syllables) read-aloud by a female Finnish speaker with a standard pronunciation. The stimuli were created for the purposes of this experiment aiming at semi-

spontaneous speech (see Table 1 for examples). Two different random orders were used in presenting the stimuli in order to avoid fatigue and learning effects. Before the task, the participants were given a practise stimulus. In the test itself, each stimulus was heard only once.

Prior to the task, the subjects filled in a background questionnaire. As the language learning is likely to be linked to working memory (e.g. Gathercole et al, 1999), the participants also completed a working memory task, a backwards digit span test in Russian, the mother tongue of the subjects (see National Longitudinal Surveys 2014 for more information). In this task, the subjects heard auditory stimuli that consisted of series of random numbers which they were supposed to repeat in a reverse order. In the test, the amount of numbers is gradually increased from two until the participant makes two mistakes at the same level. The test is considered to be a good indicator of the working memory span (e.g. Baddeley, 2003).

2.2. Participants

The participants were six native speakers of Russian (20-26-year-old university students) with knowledge of English and some French. None had any previous knowledge of Finnish, and most reported hearing Finnish for the first time in the experiment. None reported having spent longer periods abroad but one participant (R2) had been to Finland for a 6-week holiday. All participants but one (R6) had had some phonetic training in their university studies. None reported hearing problems.

2.3. Auditive judgments

The imitations were submitted to auditive evaluations by the four authors, all experts in phonetics. All judges evaluated the imitations in different randomized orders by listening always first to the Finnish stimulus and then each imitation three times for giving the rating. The judges rated both the imitated utterance as a whole and each word separately on a 5-point Likert scale. The rating criteria were mutually agreed upon: 0=missing imitation, 1=completely against the model, 2=not completely incomprehensible, 3=comprehensible, 4=rather good, 5=near-native. The judges were also asked to comment verbally on the particularities noticed.

2.4. Statistic methods

Cronbach's alpha was used to investigate the consistency between the raters. In addition, paired samples t-test was used to compare comprehensibility measures in different parts of the utterance. Pearson's correlation was computed for investigating the relationship between working memory

and comprehensibility of the utterances. Effect size was calculated using Cohen’s *d* for determining the magnitude of the significant difference between the groups.

3. Results

3.1. Utterance comprehensibility ratings

Analysis of the comprehensibility ratings showed that judges were highly consistent in their evaluations: Cronbach’s alpha was .91 in the utterance ratings, and .86 for the first word and .92 for the second and third words.

Our analysis showed there were differences in comprehensibility both between the speakers and between the utterances. First, there were utterances that were “easy” – i.e. with an average comprehensibility score between 3–5 received by all or most speakers – and those that were “difficult” – i.e. with an average score between 1–1.5 received by half of the speakers (Table 1).

Table 1. Examples of “easy” and “difficult” utterances.

	Utterance	Comprehensible production
“Easy” utterances	Voi tuota kissaa!	6/6 speakers rated over 3
	No päätä nyt!	5/6 speakers rated over 3
	Tili tuli tänään.	4/6 speakers rated over 3
	Tuleeko tulli kohta?	4/6 speakers rated over 3
“Difficult” utterances	Jouko asuu Jyväskylässä.	3/6 speakers rated under 1.5
	Onpas perunalaatikko löysää.	
	Pörheä kuusitiainen palelee.	
	Äiti laittaa kaalisoppaa.	
	Näin mökillä mäyrän.	

A closer examination of the “easy” utterances tells that they all contain voiceless plosives, and sounds such [s], [l], and [n]. These sounds can be considered unmarked: they exist in the majority of world’s languages and fill the criteria proposed by Lindblom and Maddieson (1988) of maximal perceptual distinctiveness at minimum of articulatory cost. Also, the “easy” utterances are rather short and they also contain rhythmically similar words (as e.g. *tili tuli* showing a simple two-syllable structure).

Among the segmental features judges considered as deviant were 1) an exaggerated [r] (too long [r]; too many trills), 2) an occasional voicing of unvoiced consonants, e.g., [bø̌l:ø̌] instead of [pø̌l:ø̌] (*owl*), 3) a lateral l-sound untypical of Finnish (probably postalveolar or palatal instead of alveolar), 4) vowel reversal in a diphthong, e.g., [liepæ:] instead of [leipæ:] (*bread*) and 5) omitted segment in a diphthong, e.g., [joko] instead of [jouko] (*a proper name*).

Prosodic features such as intonation and rhythm were usually imitated fairly successfully. However, the segmental length seemed to disturb comprehension to a higher extent. This is expected (Ullakonoja et al, 2014), considering that segmental length is distinctive in Finnish. As to the potential influence of L1, the judges commented rather upon the lack of a particularly Russian accent (e.g., palatalization, different features of signaling word stress) than its presence.

3.2. Word comprehensibility ratings

We also investigated the comprehensibility ratings of each word in the three-word utterances (that is, in initial, medial and final positions). On average, the initial word was the easiest to imitate, followed by the final one, leaving the word in medial position as the most difficult (Figure 1). The differences were statistically significant (paired samples t-test beginning vs. end: $t(179)=2.268$, $p<.5$; beginning vs. middle: $t(179)=6.341$ $p<.0001$; middle vs. end $t(179)=-3.616$ $p<.001$). Also the effect size was large (beginning vs. end 0.18, beginning vs. middle 0.43 and middle vs. end 0.26). Some individual differences were found: one speaker (R3) had the mean comprehensibility rating of near or over three (i.e. comprehensible) in the beginning, middle and end of the utterance, whereas for others the means were below three.

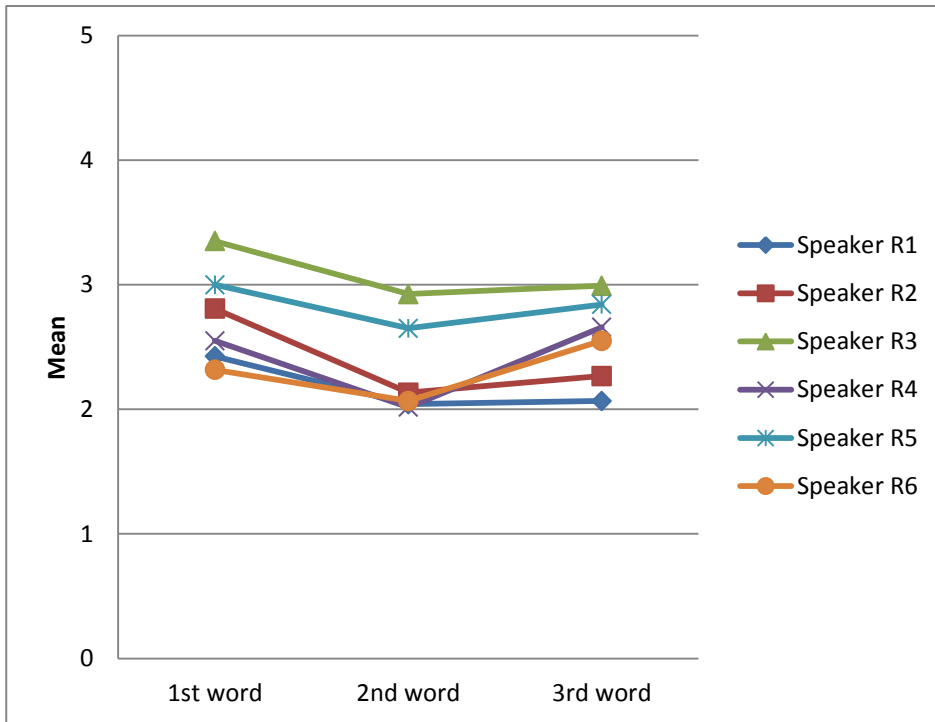


Figure 1. Word comprehensibility ratings

3.3. Comprehensibility ratings and working memory

Further, the comprehensibility ratings were compared with the results of the working memory test (see Figure 2). The working memory test gives both the *level* that indicates how many digits the participants were able to repeat in a reversed order and the *score* that indicates the number of correctly repeated stimuli in the whole test. All participants had a normal working memory span, ranging between levels 5–6. The comprehensibility ratings correlated with the working memory test level (Pearson’s correlation=.19) and score (Pearson’s correlation=.226). For one participant (R1), these results are not fully comparable due to a problem in the procedure of managing the task. There were again individual differences in the performances, but only few were significant.

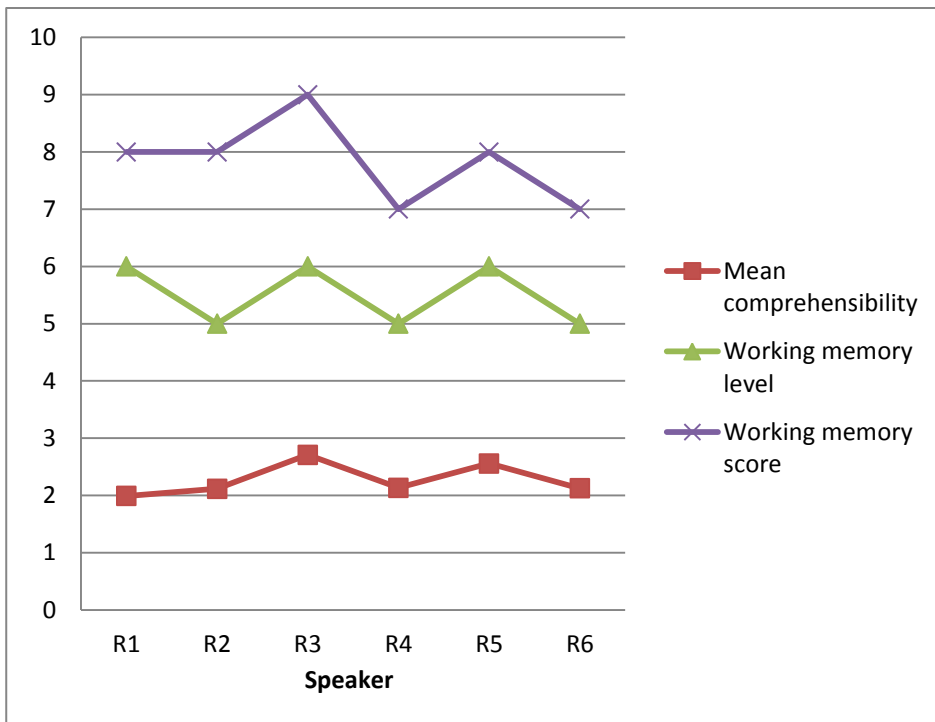


Figure 2. Comprehensibility ratings compared with working memory test results

4. Discussion and conclusions

First, as far as the features of utterances themselves are concerned, our findings are in line with Hurme (1975) in that the initial and final parts of the utterance were imitated better than the middle part and that the beginning of the utterance was easiest to imitate. It also turned out here that some of the short utterances were actually difficult (Table 1) while some longer ones were easy. Further, also the segmental content and syllable structure seem to influence imitation. These findings will be examined in more detail in future.

Second, as to the effect of working memory, an expected link was found between digit span test results and the ability to imitate, a finding that speaks for individual differences in coping with the imitation task (see e.g. Baddeley, 2003). Third, although language specific features were not directly investigated here, the lack of notable influence of L1 seems to suggest a need for further studies, and, e.g., inclusion of different L1s. To conclude with, our preliminary findings seem to open a variety of research avenues and help to pose new research questions to be subjected to rigorous investigation.

Acknowledgements

The authors wish to thank the students and staff of the St. Petersburg State University for the help in conducting the experiment. We also acknowledge the travel grant (n:o 260539) from the Academy of Finland to the first author, which enabled this collaboration.

References

- Alivuotila, L., Savela, J. and Aaltonen, O. (2008). Kielitaustan vaikutus vokaaleja matkittaessa, *Puhe ja kieli*, 28 (3), pp. 129–140.
- Arbib, M. (2012). *How the brain got language: The mirror neuron hypothesis*. Oxford: Oxford University Press.
- Baddeley, A. (2003). Working memory: looking back and looking forward, *Nature Reviews Neuroscience*, 4, 829–839.
- Chappell, P. (2012). A sociocultural account of the role of imitation in instructed second language learning, *JLLT*, 3 (1), pp. 61–91.
- Donald, M. (1991). *Origins of the Modern Mind. Three stages in the evolution of culture and cognition*. Harvard: Harvard University Press.
- Gathercole, S. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics* 27, pp. 513–543.
- Gathercole, S., Service, E., Hitch, G., Adams, A-M, and Martin, A. (1999). Phonological Short-term Memory and Vocabulary Development: Further Evidence on the Nature of the Relationship. *Applied Cognitive Psychology*, 13, pp. 65–77.
- Giles, H., Coupland, J. and Coupland, N. (1991). Accommodation Theory: Communication, Context, and Consequence. In: Giles, H., Coupland, J. and Coupland, N. (eds.) *Contexts of Accommodation*. NY: Cambridge University Press. pp. 1–68.
- Hurme, P. (1975). Oudon kielen matkimisesta: Ihmisen kyvyistä ja rajoituksista imitoida oudon kielen lauseita. In: *Fonetiikan paperit - Helsinki 1975. Publications of the University of Helsinki Department of Phonetics* 27, pp. 19–35.
- Kozhevnikov, V. A. and Chistovich L. A. (1967). *Speech: Articulation and Perception*. U.S. Department of Commerce, Clearinghouse for Federal Scientific and Technical Information, Joint Publications Research Service.
- Krauss, R. and Pardo, J. S. (2006). Speaker Perception and Social Behavior: Bridging Social Psychology and Speech Science. In: van Lange, P. A. M. (ed.), *Bridging Social Psychology: The Benefits of Transdisciplinary Approaches*. Hillsdale, NJ: Erlbaum. pp. 273–278.
- Lantolf, J. and Thorne, S. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.

- Lindblom, B. and Maddieson, I. (1988). Phonetic Universals in Consonant Systems. In: Hyman, L. and Li, Ch. (eds.) *Language, Speech and Mind*. London: Routledge. pp. 62–78.
- National Longitudinal Surveys. (2014). *Wechsler Intelligence Scale for Children - Memory for Digit Span*. [Electronic] <https://www.nlsinfo.org/content/cohorts/nlsy79-children/topical-guide/assessments/wechsler-intelligence-scale-children> Accessed 8 April 2014
- Purcell, E.T. and Suter, R.W. (1980). Predictors of Pronunciation Accuracy: A Reexamination. *Language Learning*, 30 (2), pp. 271–287.
- Peschke C., Ziegler W., Kappes J. and Baumgaertner A. (2009). Auditory-motor integration during fast repetition: the neuronal correlates of shadowing. *Neuroimage* 47, pp. 392–402.
- Suter, R.W. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning*, 26, pp. 233–253.
- Ullakonoja, R., Kuronen, M., Hurme, P. and Dufva, H. (2014). Segment Duration in Finnish as Imitated by Russians. *Proceedings of the Speech Prosody 2014 conference*, Dublin.
- Wik, P. and Granström B. (2010). Simicry - A mimicry-feedback loop for second language learning. *Proceedings of Second Language Studies: Acquisition, Learning, Education and Technology*. [Electronic] http://www.gavo.t.u-tokyo.ac.jp/L2WS2010/papers/L2WS2010_O4-04.pdf Accessed 30 April 2014
- Zetterholm, M. (2006). Same speaker - different voices. A study of one impersonator and some of his different imitations. *Proceedings of the 11th Australian International Conference on Speech Science & Technology*. [Electronic] <http://assta.org/sst/2006/sst2006-41.pdf> Accessed 30 April 2014

A comparative study of Estonian and Spanish vowels in L1 and L2 production

Katrin Leppik¹, Pärtel Lippus^{1,2}

¹ Institute of Estonian and General Linguistics, University of Tartu, ² Institute of Behavioural Sciences, University of Helsinki

leppik.katrin@gmail.com, partel.lippus@ut.ee

In this paper we compare the Estonian and Spanish vowel systems in the production of L1 and L2 speakers of the both languages. The Speech Learning Model (SLM) suggests that obtaining L2 categories that are “new” compared to L1 is easier than obtaining those that are “similar” to L1 categories. Spanish has 5 vowels /i, e, a, o, u/. Estonian has 9 vowels /i, y, e, ø, æ, a, o, x, u/. The vowels /i, u, e, o/ being identical in both languages, while Estonian /æ/ and /a/ are similar to Spanish /a/ and Estonian /y, ø, x/ do not have corresponding vowels in Spanish. In this paper we examine how the two L2 groups obtain the “new” and the “similar” categories compared to the categories that are identical in both languages. Additionally we show that the learning time has a great effect on obtaining the L2 vowel categories.

Keywords: Estonian vowels, Spanish vowels, L2 vowel production

1. Introduction

Languages differ in their vowel inventory size while there are certain patterns how the vowel space is covered in most languages (Liljencrants & Lindblom 1972). All languages use at least three different vowels /a, i, o/ or /a, i, u/. Languages tend to use those three vowels because they are far apart in the vowel space. It also seems that languages prefer to use odd number of vowels, probably because the vowel space has a triangular shape (Ladefoged 2005).

Cross-linguistic studies have shown the perceptual space of each vowel category is expanded to cover the space evenly in the case of smaller number of vowels in a language (e.g. Näätänen et al. 1997). On the other hand a smaller number of categories does not necessarily consequence greater within-category variation in production, and the distribution of the vowels depends more on language-specific pronunciation base than a universal maximal contrast in the available space (Bradlow 1995).

The Speech Learning Model (SLM) suggests that obtaining L2 categories that are “new” compared to L1 is easier than obtaining those that are “similar” to L1 categories (Flege et al. 1997). In this

paper we compare the Estonian and Spanish vowel systems in the production of L1 and L2 speakers of the both languages. Spanish has 5 vowels /i, e, a, o, u/ that are symmetrically spread over the vowel space (Table 1). The Estonian vowel system (Table 2) on the contrary is more crowded, having 9 vowels /i, y, e, ø, æ, ɑ, o, ɤ, u/, that additionally interact with the quantity. Estonian has as rather complex three-way quantity system, which functions over a trochaic disyllabic foot. Together with the vowel duration, the vowel quality changes from more central in short (Q1) to more peripheral place of articulation in long (Q2) and overlong (Q3) quantity degrees. In unstressed syllables, the quantity pattern is reversed, the vowels being longer in Q1 and shorter in Q3, and while the unstressed vowels are reduced in general, similar quantity-related variation of vowel quality can be seen as in the stressed syllables (Eek & Meister 1998; Lippus et al. 2013). In Spanish the quantity is not relevant, the longer duration is a correlate of stress (Quilis & Fernández 1996).

Table 1. Spanish vowels system.

	Front	Central	Back
High	i		u
Mid	e		o
Low		a	

Table 2. Estonian vowel system.

	Front	Central	Back
High	i, y		u
Mid	e, ø	ɤ	o
Low	æ		ɑ

In this paper we are analysing Spanish and Estonian vowels in both L1 and L2 production. We expect that:

- The Estonian L1 speakers are more sensitive to vowel duration and the Spanish L1 speakers produce the vowels with a greater variation in duration. At the same time we expect the Spanish stressed vowel duration correspond to Estonian long rather than short vowel duration.

- The vowels /i, e, u, o/ are the same in both languages and there is no difference in their place of articulation.
- Estonian L1 speakers produce Spanish /a/ more back than Spanish L1 speakers, while Spanish L1 speakers merge Estonian /ɑ/ and /æ/ into a central low vowel.
- Spanish L1 speakers merge Estonian /y, ø, ʏ/ into an ambiguous central vowel.
- As there is a smaller number of vowels in Spanish, a greater dispersion of vowel quality could be expected in Spanish L1 production than in Estonian L1 production.

2. Materials and Methods

The data was recorded from 12 speakers: six Estonian L1 Spanish L2 speakers and six Spanish L1 Estonian L2 speakers (three females and three males in both groups). The background information of the subjects in the Spanish L1 group is given in Table 3. They were aged from 16 to 42 and have learned Estonian from 5 months to 12 years while living in Estonia. One of the speakers attends to Estonian high school and uses Estonian in her everyday life. Other speakers use mainly Spanish because they are involved in the local Spanish community.

Table 3. Spanish L1 group.

Gender	M	M	M	F	F	F
Year of birth	1974	1988	1988	1997	1971	1978
Place of birth	Cataluña, Spain	Valencia , Spain	Sevilla, Spain	Cádiz, Spain	Cádiz, Spain	Cádiz, Spain
Other languages	English, French, Slovenian, Catalan	French	English, French	English, French, Italian	Italian	English, Italian
Estonian studies	2.5 years	1.5 years	0.5 years	12 years	2 years	2 years
Lived in Estonia	3.5 years	2 years	8 months	12 years	12 years	6 years

The subjects in the Estonian L1 group are described in Table 4. They were aged from 21 to 28 and all of them were students of University of Tartu. Most of them have studied Spanish at language courses, some have studied Spanish on their own. Their Spanish studying time ranges from 1.5 to 3.5 years. Four subjects have Spanish as their main subject at University of Tartu.

Table 4. Estonian L1 group.

Gender	M	M	M	F	F	F
Year of birth	1990	1990	1985	1991	1992	1992
Place of birth	Tallinn, Estonia	Tartu, Estonia	Tallinn, Estonia	Viljandi, Estonia	Tartu, Estonia	Paide, Estonia
Other languages	English, Russian, Italian	English, Russian, Polish, German	English, Russian, Finnish	English, Russian, German, Italian, Latin	English, German, French	English, Russian, French, Italian
Spanish studies	3.5 years	1.5 years	2 years	2 years	1.5 years	1.5 years

The data consists of the 12 speakers' production of the five Spanish and nine Estonian vowels. Each target vowel was placed in the stressed first syllable of words with CVCV structure, which was embedded in a carrier sentence. The test words were placed in the final position of the sentence to receive the focal stress. To eliminate the possible quantity-related variation in Estonian vowel quality, only long vocalic quantity words were selected. As discussed above, we decided for using Estonian long rather than short quantity because in Spanish the duration is a strong correlate of stress and we expected the duration of Spanish stressed vowels to be the most comparable to Estonian long vowels.

The experiment was carried out in a recording booth at the University of Tartu using the SpeechRecorder software. There were 10 sentences for each target vowel, thus each subject read $5 \times 10 + 9 \times 10 = 140$ sentences. The sentences were presented to the subjects in randomized order in two blocks for Spanish and Estonian. The target vowels were manually tagged in Praat (Boersma & Weenink 2014). The measurements of the duration, and the first and the second formants from the midpoint of the target vowel were extracted with a Praat script.

3. Results and Discussion

Figure 1 shows the vowel duration in Spanish and Estonian produced by the two groups. The hypothesis that the Spanish stressed vowel duration corresponds to the duration of Estonian long vowels can be rejected: both groups produced the Spanish vowels as short and Estonian vowels as long and there is no greater temporal variation in the Spanish L1 group. However, there is a significant interaction of the language and the mother tongue: in both languages the vowels produced by the L1 group are shorter than by the ones produced by the L2 group.

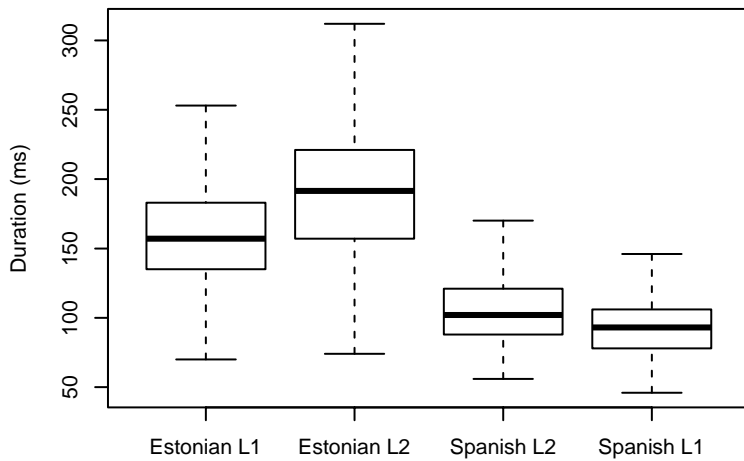


Figure 1. The vowel duration in Estonian and Spanish produced by the two groups.

In order to compare the vowels of the two languages produced by the two groups, the formant values were log-scaled normalized to z-scores for each speaker. Figure 2 plots the normalized values in F1-F2 space.

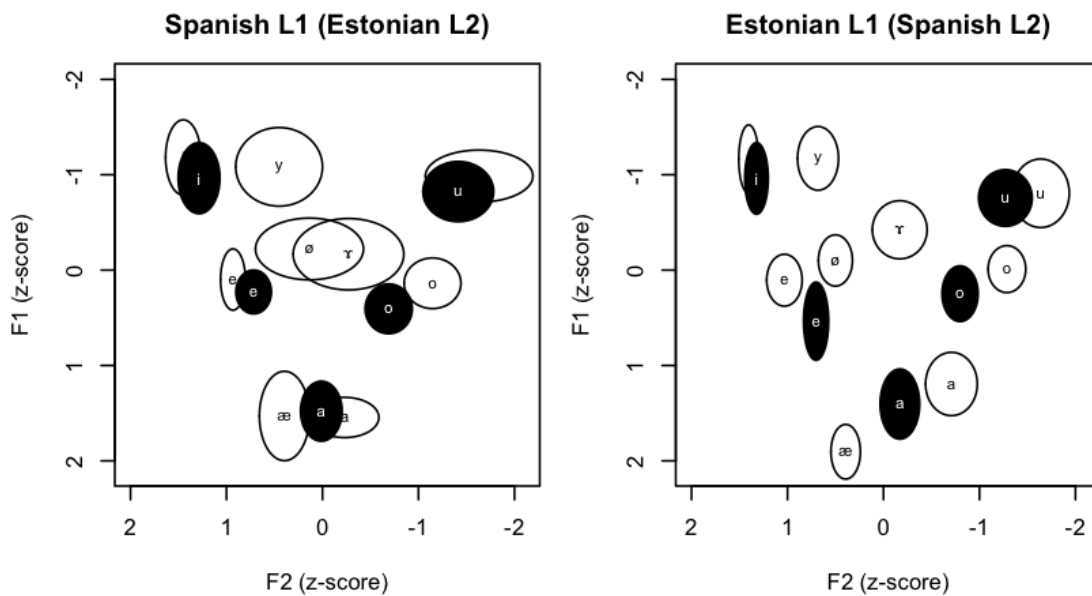


Figure 2. Spanish (black) and Estonian (white) vowels produced by the Spanish L1 group in the left and the Estonian L1 group in the right panel, the ellipses showing the standard deviation of the normalized formant values.

Figure 2 shows that the L1 vowels in both languages have approximately the same amount of within-category variation. The non-low vowels /i, e, u, o/ that fall in the same vowel category in

both languages are produced more peripheral in Estonian and more central in Spanish by both speaker groups. This difference between the two languages follows the same pattern in the production of both speaker groups and there seems to be no L1 bias or language learning effects. This difference can be an effect of the vowel duration, as both groups produced the Estonian vowels with a considerably longer duration than the Spanish vowels.

The Spanish vowels are produced in an equally spaced triangular pattern by both groups except that the low vowel /a/ is perfectly in the centre for the L1 group while the L2 group has produced it more back. However, it is difficult to say whether the more central position of Spanish /a/ compared to Estonian /a/ produced by Spanish L2 speakers is the result of language learning or is it just the effect of shorter duration that we could see in the case of the non-low vowels /i, e, u, o/.

In the L2 production of the Estonian vowel system, we can firstly see that the dispersion of the vowel quality is considerably larger for vowels /æ, ø, ɤ, y/. Spanish L1 speaker had to create new categories for Estonian vowels /ø, ɤ, y/. Those three vowels are relatively close to each other, making it difficult for L2 speakers to distinguish them. There is a vast overlap of /ø/ and /ɤ/, which are produced as an ambiguous mid-vowel by the Estonian L2 group. The vowel /a/ in Estonian L2 production is produced more front compared to the L1 group, and /æ/ is higher and more back. Thus Estonian L2 speakers merge Estonian vowels /a/ and /æ/ into Spanish vowel /a/.

It seems that the more complicated Estonian vowel system is more difficult for the Spanish L1 Estonian learners than the more sparsely occupied Spanish vowel system is for the Estonian L1 Spanish learners. However, this difference in the two L2 groups could be a result of the differences in learning time. In order to assess the successfulness of obtaining the L2 categories, the Euclidian distance of each L2 speakers vowels was calculated from the mean value of the L1 group. The distance of L2 from L1 vowels as a function of learning time is plotted in Figure 3.

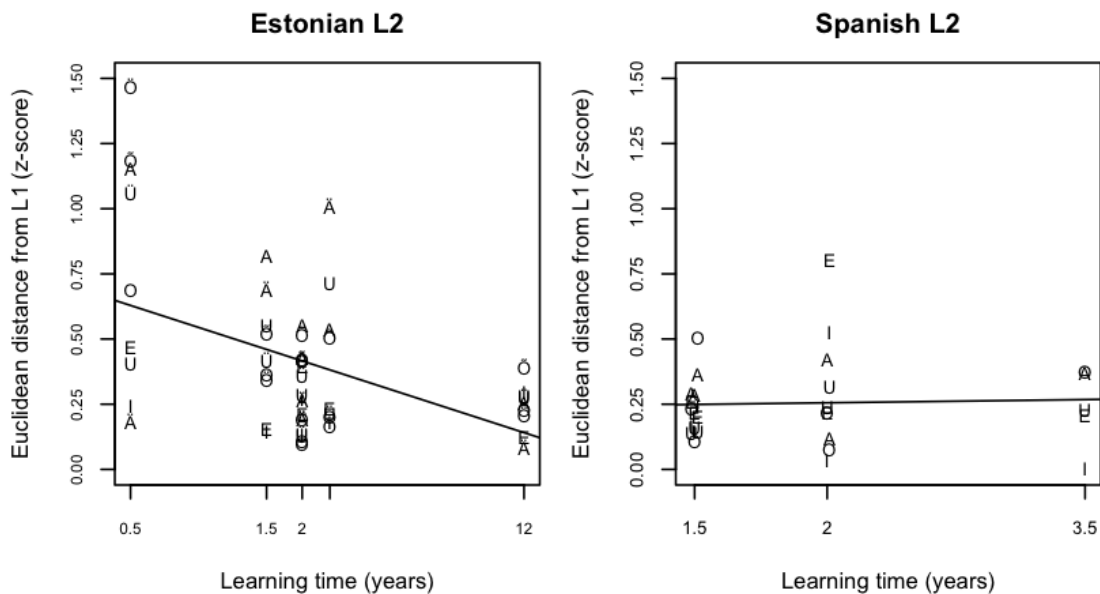


Figure 3. The Euclidean distance of the L2 speakers' vowels from the L1 mean value as a function of learning time and the fitted linear regression line.

Figure 3 shows that in the case of Estonian L2 group, there is a significant correlation between the Euclidean distance of the vowels from the L1 mean and the learning time ($r = -0.813$). This means that with a longer language learning time the vowels are produced closer to the L1 target values.

In case of Spanish L2 there is no significant correlation between the distance from L1 target vowels and learning time. Regrettably the timespan in Spanish L2 group was much more homogeneous than in Estonian L2 group and one might speculate that a wider range could possibly reveal an effect of learning time to the vowel category learning also in Spanish L2 group. On the other hand the Spanish L2 distances from the L1 target values are already rather low with considerably short learning time.

4. Conclusions

The results show that the four vowels /i, u, e, o/ are identical in both languages. As expected from the SLM, if there are two categories in one language that are close to but not the same with, as it is the case with Spanish /a/ and Estonian /æ/ and /ɑ/, the L2 speakers have difficulty obtaining the L2 category without confusing it with their L1. Estonian L1 speakers produce Spanish low vowel /a/ more back, close to their L1 corresponding vowel /ɑ/. Spanish L1 speakers merge Estonian low vowels /æ/ and /ɑ/ into Spanish /a/.

The three non-high vowels /ø, ɤ and y/ that do not occur in Spanish were produced with a greater dispersion by the L2 speakers. They are difficult for Estonian L2 speakers to discriminate and especially the mid-high vowels /ø/ and /ɤ/ were often produced as an ambiguous mid-vowel.

In conclusion we can say that learning has an important role in the acquisition of L2. The results suggest that the Spanish system with 5 vowel categories is more easily obtained by the Estonian L1 speakers whose native vowel system is more complicated, but the Spanish L1 group showed a strong effect of learning time: the longer they have learned Estonian the closer they reach to the Estonian target vowels.

Acknowledgements

We are grateful to the test subjects who volunteered to participate in this study. The study was supported by the Estonian Research Council grant IUT2-37.

References

- Boersma, P. & Weenink, D., 2014. *Praat: doing phonetics by computer* [Computer program]. Available at: <http://www.praat.org>.
- Bradlow, A.R., 1995. A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America*, 97(3), pp. 1916–1924.
- Eek, A. & Meister, E., 1998. Quality of standard Estonian vowels in stressed and unstressed syllables of the feet in three distinctive quantity degrees. *Linguistica Uralica*, 34(3), pp. 226–233.
- Flege, J.E., Bohn, O.-S. & Jang, S., 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), pp. 437–470.
- Ladefoged, P., 2005. *Vowels and consonants: An introduction to the sounds of languages* 2nd ed., Malden, MA: Blackwell Publishing.
- Liljencrants, J. & Lindblom, B., 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 48(4), pp. 839–862.
- Lippus, P., Asu, E.L., Teras, P. & Tuisk, T., 2013. Quantity-related variation of duration, pitch and vowel quality in spontaneous Estonian. *Journal of Phonetics*, 41(1), pp. 17–28.
- Näätänen, R. et al., 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(6615), pp. 432–434.
- Quilis, A. & Fernández, J.A., 1996. *Curso de fonética y fonología españolas: para estudiantes angloamericanos* 15. ed. revisada y aumentada., Madrid: Consejo superior de investigaciones científicas.

Creaky voice in spontaneous spoken Estonian

Kätlin Aare¹, Pärtel Lippus^{2,3}, Juraj Šimko³

¹ Department of Linguistics, Stockholm University, ² Institute of Estonian and General Linguistics, University of Tartu, ³ Institute of Behavioural Sciences, University of Helsinki

k2tlin.a@hotmail.com, partel.lippus@ut.ee, juraj.simko@helsinki.fi

Creaky voice has many functions in spoken communication. In a number of languages (e.g. in Finnish, Swedish, English) it has been observed to function as a turn-yielding discourse marker. The use of voice quality has been associated with sociolinguistic variation, e.g. it is more frequent in informal settings. The use of different voice qualities has been ascribed to specific social groups, e.g., in American English creak is a characteristic of young educated women. Here we observe the use of creaky voice in Estonian. The data consist of spontaneous monologues and dialogues from the Phonetic Corpus of Estonian Spontaneous Speech. The results show that creak appears mainly in informal situations. They reveal an interesting sociolinguistic pattern where creakiness depends on the age and gender of dialogue partners. While creak tends to occur more likely towards the end of a turn, it does not seem to serve directly as a turn taking cue; our data suggest that it in fact suppresses overlaps by interlocutors.

Keywords: creaky voice, sociolinguistic variation, discourse patterns

1. Introduction

Creaky voice is a phonation type typically associated with tightly adducted vocal folds open enough along a portion of their length to allow for voicing. Acoustically, it is a series of irregularly spaced vocal pulses with a decreased acoustic intensity and a lower fundamental frequency compared to modal phonation, although the lowering effect for f_0 is not a universal feature (Gordon & Ladefoged 2001).

Cross-linguistically, creaky voice has many functions. A common example is turn-yielding. In Finnish, creaky voice is used systematically in places where it is relevant for turn transition to occur (Ogden 2001). A comparable feature has been reported from Swedish – utterance fragments with final creak signal boundaries more strongly (Carlson et al. 2005). When *yeah* is said in a creaky voice in American English, it signals reciprocity and requests for a change in topic, thus occurring speech-finally (Grivičić & Nilep 2004). Creaky voice is also a turn ending marker in English (RP) used regularly with a lowering fundamental frequency to cue the turn (Laver 1994).

Different situations demand a different language use and phonation types are used accordingly. Creaky voice is most prominent in informal situations (Podesva 2007). In Hungarian, creakiness can refer to emotions with a lower emotional activity, such as boredom or relaxation (Kane et al. 2011). In Estonian, creak is also used most often in informal spontaneous dialogues (Aare 2013). If creaky voice has been used throughout the whole utterance in English (RP) it is perceived as boredom, but could be habitual and a part of the speaker's identity (Laver 1994).

Voice quality can distinguish between social groups. For example, young female speakers of American English in California use creakiness significantly more than men and Japanese women with comparable backgrounds (Yuasa 2010). Another study shows that college-aged young women in Virginia have very creaky voices (Lefkowitz & Sicoli 2007). Generally, creakiness has been interpreted as a sign of masculinity or authority and is associated with high social status. Young educated females might thus use creaky voice with a very low fundamental frequency to project an authoritative image (Yuasa 2010).

Speakers exhibit significant variation in voice quality across situations. Podesva (2007) examined intraspeaker variation in the speech of a gay student who uses phonation types depending on the situation: modal voice quality for professional purposes and when communicating with his family, but falsetto to construct a diva persona. He suggests that this speaker uses phonation types to expand his f_0 range to the physical extreme, with falsetto at the high end and creaky voice at the low. Another example are the girls belonging to the Chicana gang in Northern California who use creaky voice when they need to construct a hardcore persona (Mendoza-Denton 2011).

Previous studies have shown that voice quality, especially creaky voice, is exploited differently across languages and among various social groups and communicative situations. While there is evidence of certain groups using it more, it is unclear whether it is the same for Estonian or whether creakiness is used only when speaking with somebody from a specific social group. In addition, so far there is no plausible proof of creaky voice being used as a turn-transition marker in Estonian.

2. Material and Methods

The data comes from the University of Tartu Phonetic Corpus of Estonian Spontaneous Speech (<http://www.keel.ut.ee/foneetikakorpus>). This corpus consists of recordings of monologues and dialogues, hand labelled on phoneme level. The dialogues are between two speakers and carried out in a sound-detected room, the monologues are academic lectures recorded in a lecture hall with a

head-mounted microphone. All speakers were aware of being recorded for the corpus. The recordings have been made 2006–2013.

Table 1. Number of speakers in dialogues and monologues grouped by gender and age.

	Dialogue		Monologue	
	F	M	F	M
older	11	14	5	5
younger	19	13	1	1

For this study 47 hours of speech was used: 39 hours of dialogues and 8 hours of monologues from 64 speakers (Table 1). Five speakers participated in both a dialogue and a monologue, 7 speakers participated in more than one dialogue. The duration of each recording is approximately 30 minutes, varying from 15–50 minutes. The age ranges from 20 to 85 years, the mean being 36.5 years. The speakers were divided into two age groups based on the year of birth 1973.

The creaky voice in the recordings has been assessed in three steps. Firstly, the labellers have noted the voice quality during the phoneme level segmentation in Praat. Then, creaky voice has been detected with the Voice Analysis Toolkit (https://github.com/jckane/Voice_Analysis_Toolkit). Finally, the manual and automatic labelling was manually checked and corrected. There were misses and false alarms in both manual and automatic creak detection, agreement between the two methods was approximately 70%.

3. Results

For each speaker, the amount of creakiness was calculated as a percentage of being creaky over the time being in voice. Figure 1 compares the percentage of creakiness between speaking style (dialogue vs. monologue) and speaker gender and age group.

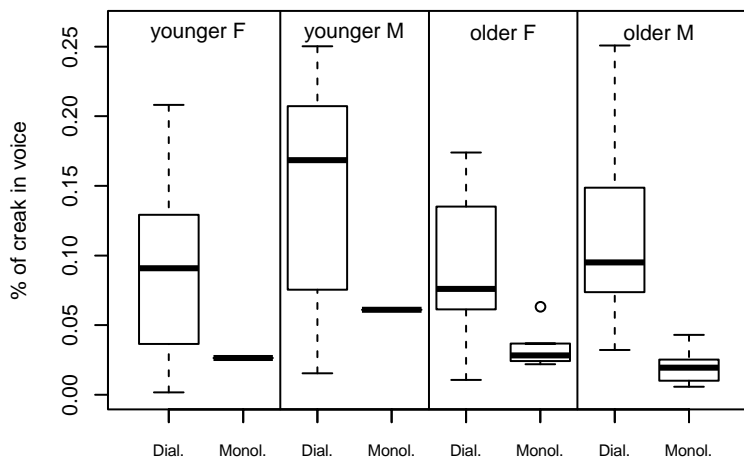


Figure 1. The amount of being creaky when being in voice as a function of speaking style, gender and age group.

Table 2. The minimal linear mixed effect model of creakiness in monologues and dialogues with significant interactions.

	Estimate	<i>t</i> value
Intercept	0.095	9.686
gender <i>M</i>	0.033	2.295
Style <i>monologue</i>	-0.056	-2.689
gender <i>M</i> :style <i>monologue</i>	-0.058	-2.003

The amount of creakiness varied greatly from speaker to speaker from 0.1% to 25%, but there was no speaker who did not creak at all. The amount of creakiness was tested with a mixed model with *age*, *gender* and *style* as fixed effects and the speaker as a random effect. Table 2 contains the significant main effects and interactions revealed in a minimal model. The speaker *age* did not have a significant effect. The intercept yields the estimate of average percentage of creak for female speakers in dialogues (about 10%). There was a significant effect of *gender*: in dialogues, the male speakers creak about 3% more than females. The significant main effect of *style* shows less amount of creakiness in monologues than in dialogues (decrease by 5.6% for females). As shown by the interaction the male speakers decrease the percentage of creakiness in monologues significantly more than the female speakers (by approximately further 6%).

Looking at dialogues we asked to what extent was the amount of creakiness influenced by characteristics of the interlocutors. Figure 2 shows the percentage of creaky voice as a function of the speaker's (SP1) age and gender depending on the speaking partner's (SP2) age and gender.

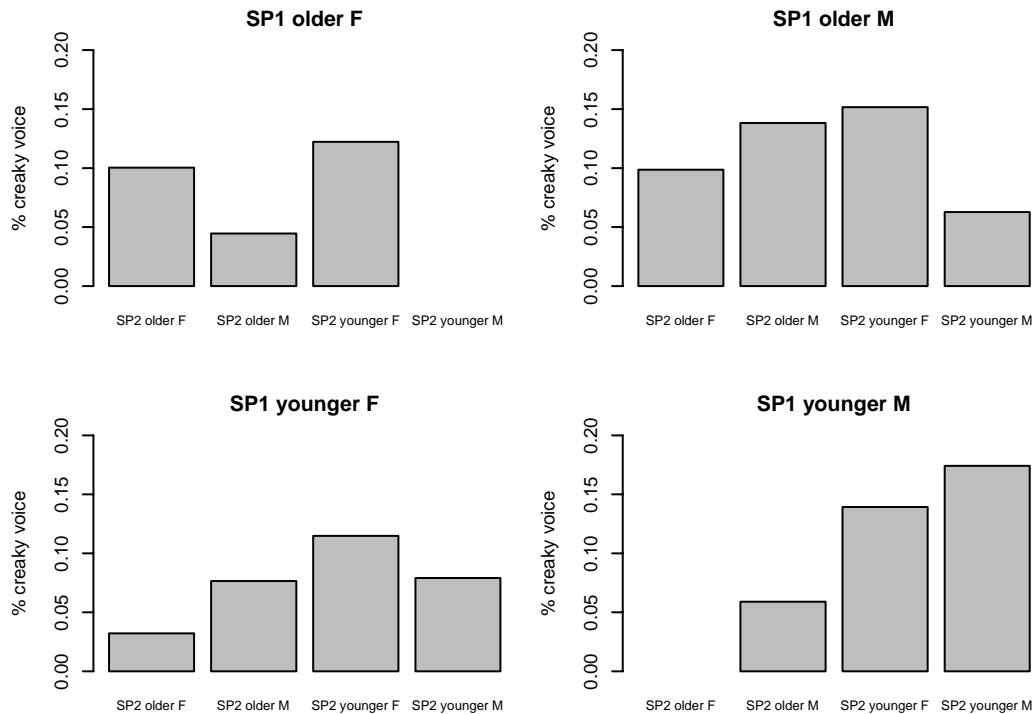


Figure 2. The percentage of creakiness depending of the speaking partner's age and gender.

The rather complicated patterns shown in Figure 2 are illuminated by a linear mixed model with *SP1 gender* and *age* and *SP2 gender* and *age* as fixed effects and the speaker as a random effect, see Table 3. Consistently with the above analysis, there was a strong main effect of *SP1 gender*, male speakers creaking considerably more than females. The main effects of speakers' *age* were not significant, although the results suggest that creakiness tends to decrease when speaking to older partners. There were two significant interactions. The *SP1 age:SP2 gender M* interaction shows that when speaking to males the percentage of creaky voice significantly decreased with speakers age, older speakers creaked significantly less. On the other hand, *SP1 gender M:SP2 age* indicates that male subjects were creaking significantly less when speaking to older dialogue partners.

Table 3. The minimal linear mixed effect model including both dialogue partners with significant interactions.

	Estimate	<i>t</i> value
(Intercept)	0.1196	5.104
SP1 gender <i>M</i>	0.1099	3.442
SP1 age	0.0002	0.307
SP2 age	-0.0006	-1.597
SP1 age:SP2 gender <i>M</i>	-0.0006	-2.708
SP1 gender <i>M</i> :SP2 age	-0.0021	-2.786

Next we ask how is creakiness distributed within a dialogue turn, and how is this distribution linked to overlaps between the partners. We look at the relative occurrence of overlaps and its dependence on creakiness. The speech was divided into Inter-Pausal Units (IPU) containing speech between two pauses longer than 15 ms. A typical IPU in our data ranged between 0.5 and 2 seconds and contained 2–8 words.

For each IPU, we extracted 100 points equally distributed in time which were checked for creakiness. Also for each of these points we registered whether the second speaker was in voice. Similarly, we logged whether SP2 was speaking during the 200 ms following the SP1's IPU. For both the SP1's IPU and the SP2 action we also checked, whether it contained lexical words or backchannels (fillers, words such as *ahah*, *mhmh*, *ja*). These measurements were used to estimate the likelihood of creakiness and overlaps within the normalized IPUs by dividing the number of cases with creaky voice and overlap, respectively, by the total number of all relevant IPUs.

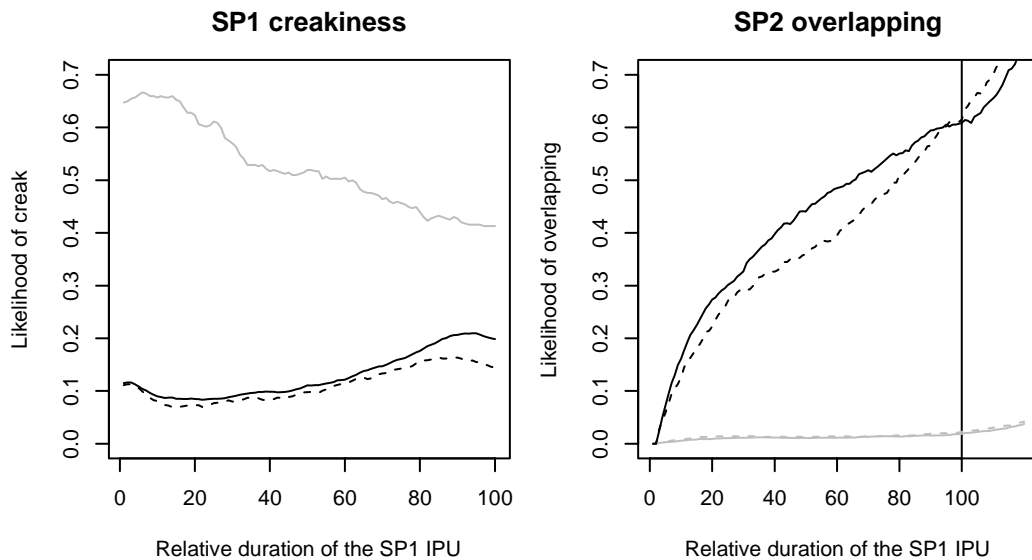


Figure 3. Left: Probability of SP1 being creaky during lexical speech when SP2 is not overlapping (black solid line) and when SP2 is overlapping (black dashed line), and when SP1 is backchanneling (grey). Right: Probability of SP2 overlapping during SP1’s IPU with lexical speech when SP1 is modal (black solid line) and creaky (black dashed line), and with backchanneling (grey).

Figure 3 depicts the results, the left panel shows the likelihood of creakiness within speaker’s IPU and the right panel the probability of overlapping speech by SP2. For lexical speech (non-backchannels), the creakiness increases towards the end of IPU. The likelihood of speech being creaky also seems to be lowered when there is an overlap by the second speaker within the IPU. In contrast with lexical speech, there is more creakiness at the beginning than at the end of SP1’s own backchannels.

The likelihood that the second speaker overlaps SP1’s lexical IPU with lexical speech is greater when the SP1 speaks with modal voice than when there is a creak within his or her IPU. The presence and distribution of backchannels does not seem to be influenced by creak.

This analysis was also performed separately for SP1 gender and age groups, but the results were qualitatively the same as those pictured in Figure 3, with the exception of older male speakers for which the percentage of creakiness was greater when there was an overlap within their IPU.

4. Discussion and Conclusions

Our data show that creakiness occurred far more often in dialogues than in monologues. It should be noted that the monologues were recorded in a lecture hall before audience, while the dialogues were recorded in a studio with speakers sitting close to each other. Consequently, the settings are

not equal; the monologues can be less creaky just because the speaker must speak up to be heard. When speaking louder and with a higher pitch the voice is less likely to creak.

Our results do not reveal a straightforward link between creakiness and turn-taking. The tendency to creak towards the end of turns might be associated with the declining pitch and intensity within a phrase. Our findings suggest that creakiness in fact somewhat inhibited the overlapped speech and the likelihood of overlap was greater during turns containing only modal voice. A possible explanation is that creaky speech is more difficult to parse and therefore requires greater attention from the interlocutor and impedes her propensity for taking the floor.

The age and gender influence the creakiness in a manner that is open to sociolinguistic interpretations. Creak might be associated with informal attitude and reflect the traditional social hierarchy. Indeed, at least male subjects creaked significantly less when speaking to older dialogue partners. Also, the tendency of older speakers creaking more to females can be seen as supporting this interpretation. However, it must be noted that everybody does creak to some extent and before reaching general conclusions, more thorough analysis needs to be performed. For example, the current version of corpus is not perfectly balanced (missing some age/gender combinations of partners).

Acknowledgements

The study was supported by the Estonian Research Council grant IUT2-37, and the National Program for the Estonian Language Technology Project EKT4.

References

- Aare, K., 2013. *Kärin eesti keele spontaanses kõnes*. BA Thesis. Tartu: University of Tartu. Available at: <http://hdl.handle.net/10062/30862>.
- Carlson, R., Hirschberg, J. & Swerts, M., 2005. Cues to upcoming Swedish prosodic boundaries: Subjective judgment studies and acoustic correlates. *Speech Communication*, 46(3-4), pp. 326–333.
- Gordon, M. & Ladefoged, P., 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics*, 29(4), pp.383–406.
- Grivičić, T. & Nilep, C., 2004. When phonation matters: The use and function of yeah and creaky voice. *Colorado Research in Linguistics*, 17(1). Available at: http://www.colorado.edu/ling/CRIL/Volume17_Issue1/paper_GRIVICIC_NILEP.htm.
- Kane, J. et al., 2011. On the use of creak in Hungarian spontaneous speech. In W.-S. Lee & E. Zee, eds. *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong: Department of Chinese, Translation and Linguistics, City University of Hong Kong, pp. 1014–1017.

- Laver, J., 1994. *Principles of phonetics*, Cambridge ; New York, NY: Cambridge University Press.
- Lefkowitz, D. & Sicoli, M., 2007. Creaky voice: Constructions of gender and authority in American English conversation. In 106th annual meeting of the American Anthropological Association. Washington, DC.
- Mendoza-Denton, N., 2011. The semiotic hitchhiker's guide to creaky voice: Circulation and gendered hardcore in a Chicana/o gang persona. *Journal of Linguistic Anthropology*, 21(2), pp. 261–280.
- Ogden, R., 2001. Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association*, 31(1), pp. 139–152.
- Podesva, R.J., 2007. Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics*, 11(4), pp. 478–504.
- Yuasa, I.P., 2010. Creaky voice: A new feminine voice quality for young urban-oriented upwardly mobile American women? *American Speech*, 85(3), pp. 315–337.

Effects of anatomical differences on relative timing of tongue and lips

Juraj Šimko¹, Michael O'Dell², Martti Vainio¹, Mona Lehtinen¹ & Tommi Nieminen³

¹ University of Helsinki, ² University of Tampere, ³ University of Eastern Finland

juraj.simko@helsinki.fi, michael.odell@uta.fi, martti.vainio@helsinki.fi,

mona.lehtinen@helsinki.fi, tommi.nieminen@uef.fi

In recent research regarding the relative timing of tongue and lip articulations in Finnish we found evidence that our test subjects use two distinct patterns associated with consonant quantity. While quantity effects (eg. tongue movement onset from /a/ to /i/ relative to lip closure onset in /tapi/ vs. /tappi/) were evident for all speakers, they were relatively large for our male speakers and relatively small for our female speakers (compared to other effects). The Embodied Task Dynamics (ETD) model we employed correctly predicted the quantity effect on lip-to-tongue coordination as a result of exploiting a natural bifurcation in optimal timing plausibly associated with long vs. short /p/. Here we discuss a possibility to replicate the two quantity related timing patterns exhibited by our subjects by systematically varying anatomical parameters (such as jaw size and mass) in the ETD model.

Keywords: articulatory timing, coordination, speaker differences, quantity

1. Introduction

Anatomical differences among individual speakers often have sizable effects on many aspects of speech articulation. Some of the largest and most obvious differences are related to gender. For instance Simpson (2009) discusses a range of various male vs. female differences in speech, mostly related to anatomical characteristics, such as larynx size (affecting phonation and F0), vocal tract size (vowel articulation, formants), mass and speed of articulators, but also learned (sociophonetic) behavior and interactions of the above. In this article we concentrate on one articulatory feature, the *relative timing of tongue and lip articulations*, which, as far as we know, has not been previously investigated in this connection. While it is certainly possible that anatomical speaker differences influence coordination of gestures, it is not immediately obvious just what the results of such influence would be.

1.1. Lip-tongue coordination: V2LAG

Löfqvist & Gracco (1999) reported that, in relation to tongue body movement associated with vowel transition from one syllable to the next, a lip closing gesture associated with an intervocalic

bilabial consonant starts earlier when the vowel transition is from /a/ to /i/ (/a–i/ condition) rather than /i/ to /a/ (/i–a/ condition). This difference can be quantified using a measure called V2LAG: How long after the onset of bilabial consonant gesture does the vowel transition begin?

Since the work of Löfqvist & Gracco (1999) similar results have been reported in several other studies for a variety of languages (e.g. Šimko, O’Dell & Vainio 2014 for Finnish; Beňuš & Šimko, 2014 for Slovak; see also Löfqvist, 2006, for Japanese). The main effect has been that V2LAG is larger in the /a–i/ condition than in the /i–a/ condition. In what follows we call this the VOWEL effect.

In earlier work (O’Dell, Šimko, Nieminen, Vainio & Lehtinen 2011a, O’Dell, Šimko, Nieminen, Vainio & Lehtinen 2011b, Šimko *et al.* 2014) we have studied this phenomenon for Finnish using articulatory and test words such as *tapi*, *pippa*, etc. (for details we refer the reader to those articles). We assess various effects in our data with Bayesian inference using a hierarchical (or ANOVA-type) model (Gelman & Hill 2007). All empirical measures are included in the model simultaneously, with any logical (i.e. non-empirical) restrictions holding between them imposed on the prior distributions (e.g. the fact that velocity of lip movement at movement onset and movement offset is zero and less than or equal to maximum velocity everywhere in between). Using this model we can estimate an average trajectory (and accompanying posterior distribution) for any combination of effects.

Figure 1 shows examples of such “average trajectories” corresponding to lip closing gestures (lip aperture), aligned in time at the onset of tongue movement for the change of vowel gesture. Cross-hairs located at three points (movement onset, maximum velocity, movement offset) indicate 95% credible intervals (CI) and thus give some indication of the uncertainty of the average trajectory.

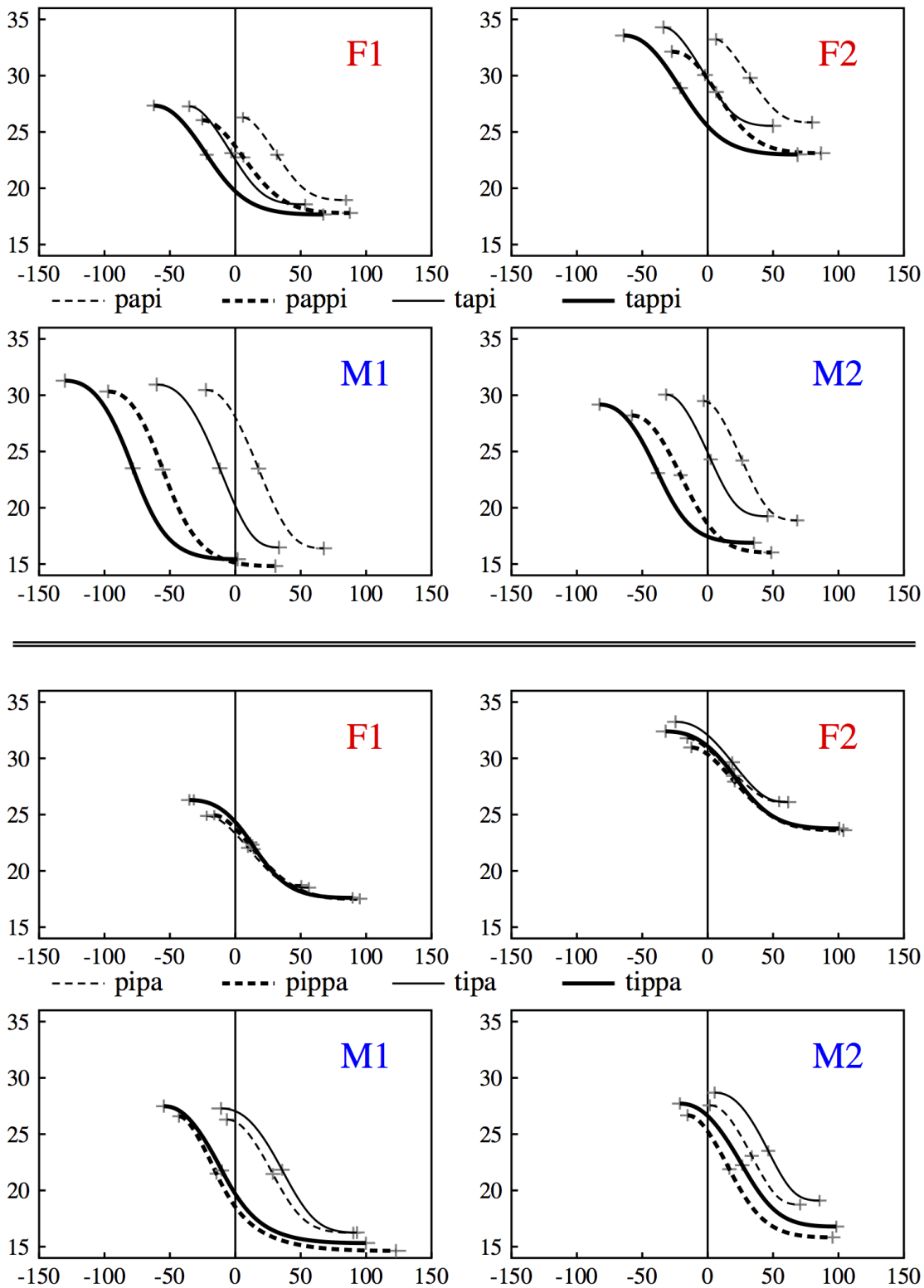


Figure 1. Lip closing gestures for female speakers F1, F2 & male speakers M1, M2: Contexts /a-i/ (top), and /i-a/ (bottom).

As mentioned above, our research has shown, first of all, that V2LAG is indeed larger in the /a-i/ condition than in the /i-a/ condition in Finnish. In addition to this VOWEL effect we have found that V2LAG is sensitive to several other factors as well. For instance, V2LAG is on average larger when

the *previous consonant* is /t/ rather than /p/ (CONSONANT effect). This effect can be seen in Fig. 1 by noticing that in the /a-i/ condition (upper half of Fig. 1) the dashed line trajectories occur later than the solid line trajectories. The *quantity* of the intervocalic consonant (singleton /p/ vs. geminate /pp/) also has a significant effect on V2LAG (QUANTITY effect). This effect can be seen in Fig. 1 by noticing that the thick line trajectories occur earlier than the thin line trajectories. It is also clear that there are significant interactions between the three effects, for instance the CONSONANT effect is close to zero in the /i-a/ condition (lower half of Fig. 1).

2. V2LAG and speaker differences

In addition to these three main effects, we also noticed clear differences among our four speakers, most easily seen by comparing the relative sizes of effects for separate speakers. The CONSONANT effect (e.g. *papi* vs. *tapi*) is roughly similar for all speakers, about 40 ms in the /a-i/ condition and close to 0 ms in the /i-a/ condition. The QUANTITY effect (e.g. *tapi* vs. *tappi*), however, is quite different for our female and male speakers. It is relatively small for our two female speakers (slightly smaller than the CONSONANT effect), but relatively large for our two male speakers (much greater than the CONSONANT effect).

Although this difference is fairly clear in Fig. 1, we can also look directly at the posterior distribution for the difference between V2LAG in the QUANTITY effect and the CONSONANT effect, i.e., the difference in time of onset of lip closing for the test words *pappi* and *tapi* (/a-i/ condition) and *pippa* and *tipa* (/i-a/ condition). The relevant posterior distributions are shown in Fig. 2. It is clear that the speaker differences, especially the females (F1 and F2) vs. the males (M1 and M2), are very robust. The posterior probability that the difference is positive (i.e. QUANTITY effect larger than CONSONANT effect) in the /a-i/ condition and in the /i-a/ condition, respectively, is $p = 0.003$ and $p < 0.001$ for speaker F1, $p = 0.058$ and $p = 0.004$ for speaker F2, $p > 0.999$ and $p > 0.999$ for speakers M1 and M2.

Of course, with only four speakers, the apparent division of females vs. males could be just a coincidence, but it does raise the question as to what extent anatomical differences might affect the optimal timing of tongue and lips.

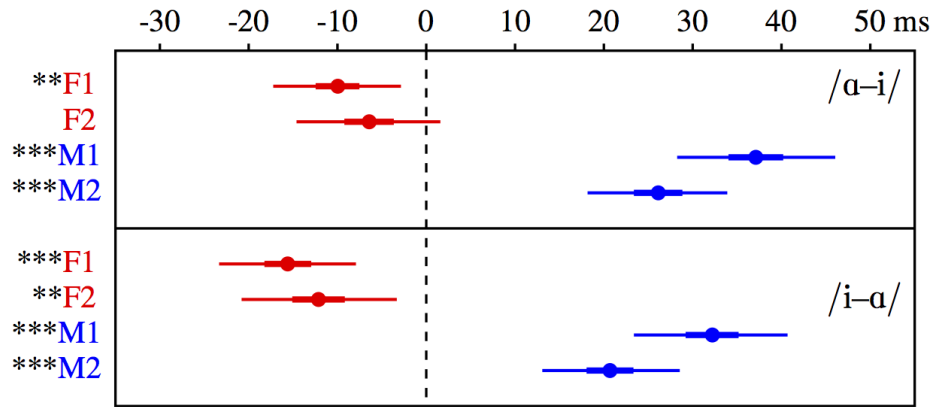


Figure 2. Difference in QUANTITY effect and CONSONANT effect for four speakers in the /a-i/ condition (above) and /i-a/ condition (below).

3. Embodied Task Dynamics

3.1. Overview

Embodied Task Dynamics (ETD, Šimko & Cummins, 2011) is a modeling paradigm designed for computing inter-gestural sequencing patterns satisfying demands of production efficiency and perceptual efficacy. The platform is grounded in the theory of Articulatory Phonology: each utterance is represented by a gestural score, a temporal prescription of activation intervals of target-oriented articulatory actions called gestures (Browman & Goldstein, 1992). A gesture is a primitive dynamical unit driving a set of relevant articulators towards a target. For instance, a lip closing gesture drives the lips towards each other; given an appropriate dynamical target, the lips eventually collide and a bilabial stop is produced.

The requirements of efficient articulation and sufficient contrast impose, as a rule, conflicting demands on the dynamics of movement of individual articulators and on their coordination. For instance, the requirement of a greater contrast between vowels /a/ and /i/ entails more extreme positions for the tongue body within the oral cavity. What is gained in terms of perceptual clarity is paid for by greater exertion of muscles and possibly also in terms of longer durations of the movements.

This intuition is captured in ETD by quantifiable cost components evaluating effort (E), duration (D) and a parsing cost conceptualized as the listener's effort needed to parse the intended utterance (P). A linear combination of these three measures yields the overall cost

$$C = \alpha_E E + \alpha_P P + \alpha_D D . \quad (1)$$

Deploying an optimization procedure, the optimal gestural score for a given utterance, i.e., the sequence of gestures minimizing the overall cost, can be computed.

The dynamics driving the model is based on the Task Dynamics implementation of Articulatory Phonology (Saltzman & Munhall, 1989) adjusted to incorporate physiological and physical properties of the vocal tract. The characteristics of articulation depend on adjustable masses and dimensions of individual articulators that can be *individualized for different model speakers*. These characteristics allow articulatory effort E to be estimated as an overall magnitude of all forces acting on the articulators.

Parsing cost P is defined as a combination of durational and articulatory characteristics of individual gestures (see Šimko *et al.*, 2014, for details). The third cost component D is simply the duration of realization of the given gestural score.

The weights α_E , α_P and α_D in Equation 1 implement intentional control over the relative influences of the components and can be adjusted globally (spanning the entire utterance) or locally (pertaining to an individual gesture in the sequence).

3.2. Embodied Task Dynamics, Quantity and V2LAG

As discussed in detail in Šimko *et al.* (2014), localized increase of the weight α_P , limited to the bilabial /p/, elicits lengthening of the closure duration corresponding to gemination. The increase in closure duration accompanying the increase of the local weight α_P is not linear: initially, the duration increases very slowly, at some point it jumps to a considerably greater value and again continues to increase slowly. In fact, the model exhibits a hysteretic jump dividing continuous duration into discrete categories of short vs. long /p/ as shown schematically in Figure 3. This discontinuity allows the model to distinguish between singletons and geminates: it is assumed that optimal solutions lying on the short branch correspond to singletons, solutions on the long branch to geminates.

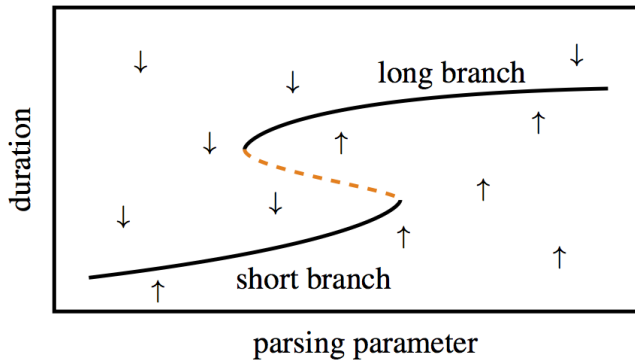


Figure 3. Hysteretic jump dividing continuous duration into discrete categories (small arrows point toward more optimal durations for a given parsing parameter, the solid line indicates durations which are locally most optimal).

The continuous adjustment of the parameter also leads to an abrupt change in the coordination between the bilabial gesture and the following (co-produced) vocalic gesture captured quantitatively by the variable V2LAG introduced above. As the control parameter changes, V2LAG and duration both follow the same pattern of slow change, qualitative jump and subsequent further increase. Thus the model provides an explanation in terms of optimization for the QUANTITY effect observed in our subjects.

In quantitative modeling of the utterances listed in Section 1, the ETD model correctly predicts a considerable number of coordination details and kinematic characteristics of articulation including the VOWEL, CONSONANT and QUANTITY effects. As these results are reported in detail elsewhere (Šimko *et al.*, 2014) we focus here on a particular aspect of the comparison between the empirical data and modeling, namely the relative size of the CONSONANT and QUANTITY effects introduced in Section 2.

4. Embodied Task Dynamics and speaker differences

Because ETD has been so successful in explaining nontrivial effects in lip-tongue coordination, it is natural to ask whether it might also be able to account for the speaker differences we have uncovered. Since the optimal solutions found by the model depend on anatomical parameters, it is certainly plausible that different values for these parameters might lead to different optimal solutions.

With the physiological parameters used by Šimko *et al.* (2014) the model rather closely mimics the behavior of our two female speakers. Could changes in physiological parameters of the model elicit behavior more closely resembling that of our male speakers? In spite of extensive searching of the

space of anatomical parameters, the best answer we can offer so far is that changes in a single parameter do not yield the required results—we have yet to produce V2LAG patterns similar to those of our male speakers.

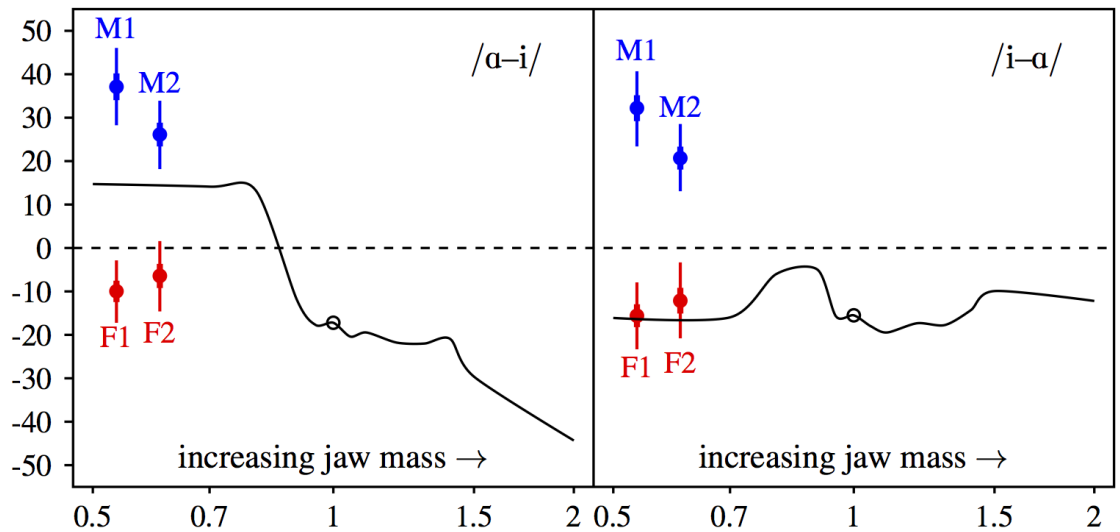


Figure 4. Difference in optimal QUANTITY effect and CONSONANT effect (ms) as computed by ETD for varying jaw mass. For comparison, the estimates for our four speakers are repeated from Fig. 2.

As an example of this, Fig. 4 illustrates how the difference measure depends on jaw mass, a parameter which has been one of the most effective in changing model behavior. The small circle in the figure indicates the value for the original mass parameter (set to 550 g in the Šimko *et al.* (2014) simulations), the solid curve shows roughly how this measure changes as jaw mass is varied. It can be seen that in this case a radical adjustment in jaw mass does manage to make the difference positive (indicating that the QUANTITY effect is larger than the CONSONANT effect), if only in the /a-i/ condition. However, the direction of the change is a bit surprising: it appears that a lighter jaw can lead to a more male-like coordination! This is perhaps not as counterintuitive as it initially sounds, because we haven't adjusted other parameters such as sizes and muscle forces, which would normally covary in real individuals. In all likelihood individual differences such as those observed in our speakers involve complex relative differences in mass, size and muscle force. It is possible that, taking these into consideration, male jaws might in some sense be “lighter” than female jaws *relative to various other dimensions* such as muscle strength.

On the other hand, even adjustment of jaw mass fails to produce a male-like pattern in the /i-a/ condition. Similar failures are obtained by adjusting other single anatomical features, masses, muscle force, or sizes of articulators and the oral cavity.

5. Discussion

On the positive side, our extensive search of anatomical parameter space has given us even more confidence in the robustness of the main effects of the model. Even radical changes in model parameters do not alter the three main effects (VOWEL effect, CONSONANT effect, QUANTITY effect) predicted by the model, and these main effects are exhibited by all our speakers.

There are many possible reasons for our failure to replicate the observed speaker differences involving the relative size of the main effects. It may be simply that multiple parameters need to be manipulated simultaneously. Due to the large number of parameters involved even in the simplified ETD model a thorough search of parameter space is a daunting (but ongoing) task.

It may be, however, that the ETD model is not capable of reproducing the observed effects exactly because it is vastly simplified. For example, the model is still missing an active lip-opening gesture (see Beňuš & Šimko, 2014 for discussion).

Also, in the present form of the model the approximation to perceptual parsing cost (P) utilized is based on articulation (e.g. degree of target approximation), which means that perception must change in the model along with any anatomical changes implemented. Thus, in a sense, the model can only represent an average speaker/hearer of a speech community (which may partly account for its success). It may be, therefore, that a deeper understanding of the role of individual anatomical differences will require incorporation of a more sophisticated model of speech perception.

Of course it may simply be that some individuals (for instance our two male speakers) are less optimal (in the sense defined by the model) than others. This in turn could be the result of the particular history of speech experienced by each individual, again in relation to the background of an entire speech community. In the future more individuals need to be observed, males and females, and preferably at various stages of linguistic and anatomical development.

All of these speculations present huge challenges to the ETD model, which will not easily be overcome. In the light of these challenges the success this relatively simple optimization model has already achieved is all the more intriguing.

References

Beňuš, Š. & Šimko, J. (2014) Emergence of prosodic boundary: Continuous effects of temporal affordance on inter-gestural timing, *Journal of Phonetics* 44:110–129

- Browman, C. P. & Goldstein, L. (1992). Articulatory Phonology: An Overview. *Phonetica*, 49:155–180.
- Gelman, A. & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.
- Löfqvist, A. (2006). Interarticulator programming: Effects of closure duration on lip and tongue coordination in Japanese. *The Journal of the Acoustical Society of America*, 120(5 Pt 1):2872.
- Löfqvist, A. & Gracco, V. L. (1999). Interarticulator programming in VCV sequences: Lip and tongue movements, *Journal of the Acoustical Society of America* 105(3): 1864–1876.
- O'Dell, M., Šimko, J., Nieminen, T., Vainio, M. & Lehtinen, M. (2011a). Timing of intervocalic consonant gestures in Finnish, in S. Werner & T. Kinnunen (eds), *XXVI Fonetiikan päivät 2010*, Itä-Suomen yliopisto, pp. 16–21.
- O'Dell, M., Šimko, J., Nieminen, T., Vainio, M. & Lehtinen, M. (2011b). Relative timing of bilabial gesture in Finnish, in W.-S. Lee & E. Zee (eds), *Proceedings of the 17th International Congress of Phonetic Sciences*, City University of Hong Kong, pp. 1518–1521.
- Saltzman, E. L. & Munhall, K. G. (1989). A Dynamical Approach to Gestural Patterning in Speech Production. *Ecological Psychology*, 1(4):333–382.
- Šimko, J., & Cummins, F. (2011). Sequencing and optimization within an embodied task dynamic model. *Cognitive Science*, 35(3): 527–562.
- Šimko, J., O'Dell, M. & Vainio, M. (2014). Emergent consonantal quantity contrast and context-dependence of gestural phasing, *Journal of Phonetics* 44: 130–151.
- Simpson, A. P. (2009). Phonetic differences between male and female speech, *Language and Linguistic Compass* 3(2): 621–640.

Causal factors and indicators of word stress in Meadow Mari

Dennis Estill

University of Helsinki, Department of Finnish, Finno-Ugrian and Scandinavian Studies

dennis.estill@helsinki.fi

While it may be accepted that fundamental frequency (F_0) and/or intensity are instrumental in producing perceived stress, the interpretation of duration as a contributing factor would appear to be complicated to define in Meadow Mari. While, on the one hand, increased F_0 and intensity may, under specific circumstances, increase the duration of, especially, vowels, on the other, extended duration could be intentional, bearing in mind that in this language vowel quantity is not differentiated. Complications resulting from the interaction between these three parameters seem evident in Meadow Mari, although this is probably true also of many other languages. The objective in the case examined is to take an overall look at the situation in Meadow Mari, in order to determine whether F_0 or intensity or perhaps even duration (or the interaction between these) is the cause of stress, with particular attention given to the part duration plays in the process, that is, is this an accompanying feature or a cause? At the same time it would be useful to know which of the three factors under discussion best identifies stress, so as to determine which factors should we examine in ascertaining how word stress is perceived in Meadow Mari.

It was found that fundamental frequency and intensity are certainly involved in the production of stress in Meadow Mari and probably duration, while the best indicator of stress is definitely duration.

Keywords: Meadow Mari, fundamental frequency, intensity, duration, causal and indicative factors

1. Introduction

As is known, the three acoustic factors most commonly related to the production of stress in any language are fundamental frequency (F_0), intensity and duration, and combinations of these. The main cause of stress is often regarded to be fundamental frequency, although in some languages intensity may be more salient, for example, Greek and Komi (Botinis, 1998; Estill, 2010).

While it may be accepted that F_0 and/or intensity are instrumental in producing stress, the interpretation of duration as a contributing factor is complicated to define. On the one hand, increased F_0 and intensity may, under specific circumstances, increase the duration of, especially, vowels as an accompanying feature, while on the other, extended duration could be intentional, particularly in languages that do not differentiate vowel quantity. Meadow Mari is one such language. The objective in the case of Meadow Mari on hand is to determine whether F_0 or intensity (or the interaction of both) is the cause of word stress, and whether duration could be added, or even

singled out, as a relevant acoustic factor. At the same time it would be useful to know which of the three factors under discussion identifies stress that is to say, which factors should we examine in order to determine where stress in Mari lies, since this aspect should not be confused with the production of stress. These are the targets for the ensuing examination.

2. The Meadow Mari vocalic system

For reference purposes the diagram below displays the placement of vowels in the Meadow Mari vowel chart from Estill (2012), in which several earlier variations are also described. Strings of consonants do not occur in Meadow Mari and the syllable pattern is (C)V(C). Unlike certain other Uralic languages, Meadow Mari does not display vowel harmony as a systematic feature, although it is still present to some extent in the modern language (Alhoniemi, 1985: 20).

	front→		←back	
close	<i>i</i>	<i>y</i>		<i>u</i>
mid-close	<i>e</i>	<i>ö</i>	<i>ə</i>	<i>o</i>
mid-open				
open		<i>a</i>		

3. Word stress

Although there are certain exceptions word stress in Meadow Mari falls on the last full vowel in a word or, in the absence of a full vowel, on the first reduced vowel /ə/. In the Meadow Mari teaching manual for Russian speakers (Mariyskiy, 2000) from which the material for this study was taken, word stress was marked, thus providing an initial basis for comparison.¹ However, in the experiments described below, the findings accord with the actual realisation of the placement of word stress based on the evidence of two referees, even though very little discrepancy was found to exist.

¹ Although it may seem obvious, the text from which the informants read was not marked for word stress.

4. Description of experiment

Measurements were taken of intensity, F_0 and duration based on the readings from the Meadow Mari language manual for Russian students (referred to above) of two female informants, EA and LS, whose ages and places of birth were respectively 39, Zvenigovskij *raion*, and 38, Volzhskij *raion* in the Republic of Mari El. Both were teachers of the Mari languages. For the experiment the standard dialect was considered. The chosen text was composed of several short readings from the manual consisting of a total of about 500 words. The recordings took place in an auditorium in the then Department of Finno-Ugrian Studies of the University of Helsinki where every attempt was made to reduce outside noise to a minimum. All doors and windows were closed and electrical interference eliminated as far as possible. A Plectalk recorder was used for the recordings and the sound quality later calibrated for fidelity. Although the quality of the sound was considered satisfactory, it did fall a little short of what might be expected in perfect laboratory conditions. The material was read once by each speaker, after they had become sufficiently acquainted with the material. The microphone used was an AKG D 660S and the acoustic measurements made using the Praat 5.143 program.

Figure 1 shows the distribution of vowels for the experiment described below. For convenience the figures are based on the number of tokens analysed for fundamental frequency, there being a slight difference for intensity and duration.

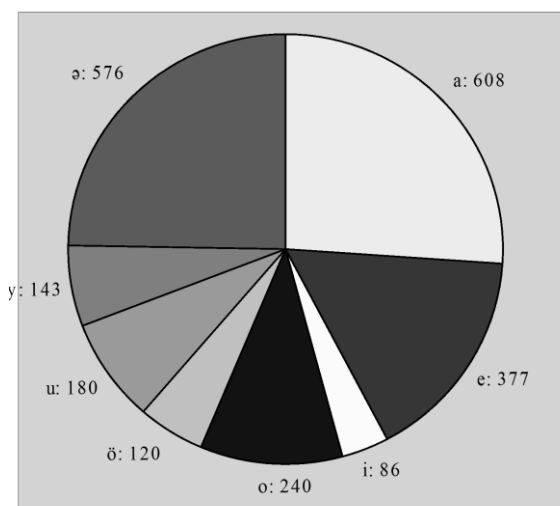


Figure 1. Distribution of Meadow Mari vowels in experiment described below. The slices of the chart show the vowels and corresponding numbers of tokens.

Although originally intensity and F_0 were measured from one to three points, ultimately only the peak readings were used for determining the results. Four classes of vowel were examined, namely, stressed, “indefinite” vowels about which there was disagreement between the two listener referees concerning the location of stress, unstressed vowels and vowels occurring in monosyllabic words. The results were compiled and averages determined for the three acoustic variables in question.

5. Findings

The complete results are shown in Table 1.²

Table 1. Table of complete results for experiment showing averages for three variables, intensity, fundamental frequency and duration for two informants, EA and LS. These averages cover all seven Meadow Mari vowels in the material, stressed, possibly (indefinite) stressed, unstressed and those occurring in monosyllabic words. The number of tokens for each sample is shown in rounded brackets. The total number of tokens is 2330.

	int (dB)		F0 (Hz)		dur (ms)	
	ea	ls	ea	ls	ea	ls
<i>a</i> stress	69.8 (147)	64.7 (151)	246.4 (153)	204.9 (142)	96 (146)	97 (147)
<i>a</i> indefinite	70.1 (60)	63.5 (45)	243 (60)	181.8 (44)	79 (60)	76 (46)
<i>a</i> no stress	70.1 (83)	63.9 (94)	242.7 (83)	180.5 (95)	79 (82)	72 (94)
<i>a</i> monosyllabic	71.3 (14)	64.5 (14)	241.2 (15)	178.6 (14)	115 (14)	107 (13)
<i>e</i> stress	70.5 (64)	66.8 (75)	256.2 (73)	209.2 (73)	86 (64)	94 (73)
<i>e</i> indefinite	67.1 (30)	63.1 (26)	201.7 (29)	171.2 (24)	99 (29)	77 (26)
<i>e</i> no stress	69.3 (78)	64.9 (78)	234.6 (78)	186.4 (78)	69 (78)	65 (80)
<i>e</i> monosyllabic	72.4 (13)	65.6 (13)	258.8 (13)	199.3 (13)	81 (11)	87 (13)
<i>i</i> stress	72.3 (11)	67.2 (11)	287.5 (12)	210.1 (12)	68 (10)	84 (12)
<i>i</i> indefinite	72.1 (7)	61.9 (2)	266.2 (7)	179.9 (2)	62 (7)	64 (2)

² This information would be much better presented visually with line charts, but because of the limitations on space, only the values are shown. This also applies to the presentation of the data contained in Table 2.

<i>i</i> no stress	70.8 (21)	65.6 (24)	255.3 (20)	203.1 (24)	53 (20)	58 (24)
<i>i</i> monosyllabic	71.9 (5)	67.8 (5)	275.2 (5)	212.2 (5)	76 (5)	68 (5)
<i>o</i> stress	70.1 (42)	66.5 (35)	256 (47)	204.3 (37)	91 (41)	89 (36)
<i>o</i> indefinite	69.1 (11)	61.8 (14)	224.3 (13)	164.7 (14)	76 (12)	61 (15)
<i>o</i> no stress	69.4 (55)	66.3 (58)	249.5 (55)	199.3 (57)	65 (56)	63 (60)
<i>o</i> monosyllabic	72 (12)	68.9 (13)	271.5 (12)	235.3 (13)	83 (12)	81 (13)
<i>ö</i> stress	73.6 (7)	67.5 (7)	275.7 (8)	188.5 (7)	83 (7)	74 (7)
<i>ö</i> indefinite		66.2 (1)		206.3 (1)		49 (1)
<i>ö</i> no stress	70.4 (17)	64.6 (16)	249.1 (17)	178.8 (15)	48 (17)	53 (14)
<i>u</i> stress	69.7 (39)	66.1 (33)	261.9 (41)	205.9 (33)	62 (36)	59 (33)
<i>u</i> indefinite	68.3 (6)	65.4 (12)	232.7 (6)	203.9 (12)	49 (6)	58 (12)
<i>u</i> no stress	69.7 (42)	64.7 (40)	245.4 (44)	190.6 (40)	56 (44)	52 (40)
<i>u</i> monosyllabic	70.9 (4)	65.3 (4)	258.3 (3)	183.1 (4)	93 (4)	73 (4)
<i>y</i> stress	73.1 (20)	67.8 (24)	279.9 (22)	207.7 (24)	70 (23)	59 (24)
<i>y</i> indefinite	72.4 (7)	63.1 (5)	260.6 (7)	184.2 (5)	42 (7)	49 (5)
<i>y</i> no stress	70.2 (30)	65 (31)	240.6 (29)	190.3 (30)	54 (29)	45 (31)
<i>y</i> monosyllabic	73.9 (13)	70.3 (13)	304.7 (13)	242.4 (13)	68 (12)	69 (13)
<i>ə</i> stress	70.6 (8)	65.3 (16)	272.1 (18)	206.7 (18)	65 (17)	44 (18)
<i>ə</i> indefinite	65.5 (11)	64.1 (14)	186.6 (12)	177.5 (14)	53 (12)	62 (12)
<i>ə</i> no stress	68.5 (253)	63.3 (240)	230.9 (247)	189 (223)	47 (245)	43 (241)
<i>ə</i> monosyllabic	70.4 (17)	66.8 (17)	248.9 (18)	218.5 (17)	61 (19)	53 (17)

The above table presents averages for each of the acoustic variables under consideration for both informants. Looking at the intensity figures first, agreement between the informants is evident. Generally speaking, in the tables as a whole monosyllabic-word vowel patterns are close to those of stressed vowels and, despite the differences in the numbers of tokens, these measurements strongly support such similarity. My so-called indefinite vowels come close to the unstressed class. The best indicators of stress would therefore appear to be found by comparison of (confirmed) stressed

vowels and the unstressed. Therefore, as far as intensity is concerned, a clear correlation between stress and a rise in this parameter is can observed.

Although the means for *a* and *o* are almost the same, and this feature is mentioned again the next paragraph since it also concerns F_0 , the mean intensity readings for stressed as against unstressed vowels are nevertheless approximately 2 dB higher. Given this, it would seem that increased intensity is a feature of stressed vowels. On the other hand, a close examination of the intensity values of the first three sentences read by the informants, the values for which are shown in below in Table 2, might give rise to some doubt as to whether reading from averages always reveals the truth or whether some variables are being missed, and may give support to an opinion that intensity is simply an accompanying feature. It could be that acoustic parameters are distributed, or accumulate, onto certain vowels in certain environments, a tendency which can not easily be interpreted from the values. It would seem necessary to match the results between the means and the actual production of stress in individual cases. Unlike the results for duration, counting the readings of the peaks and troughs provides no clue as to the incidence of intensity, let alone its cause. It should of course be remembered when interpreting the relevant charts, that the numbers of tokens for /i/, /y/ and, especially /ö/, were smaller.

The middle columns in Table 1 concern F_0 , and displays mean vowel readings for EA and LS. Although with F_0 there appears to be a difference between the speakers in the amount of extra energy involved in the production of, particularly, /i/ and /y/ and this might be put down to the smaller samples, the pattern for F_0 is otherwise almost identical to that of intensity.

Returning to the coincidence of /a/ and /o/ it can be observed that the tendency in the case of F_0 seems to follow much the same pattern as that of intensity, with the exception of EA's *a*. Of course, this could be a speaker-specific feature, since there is considerable divergence between he two speakers in the case of /a/. Of further interest when analysing the roles of F_0 and intensity is that those vowels concerning which there is disagreement between listeners regarding the placement, or not, of stress on a particular vowel, in most instances display lower readings than those vowels than are unstressed. Yet this is not noticeable in the case of duration (discussed below). Looking at the complete figures in Table 1, it can be ascertained that, overall, stressed and monosyllabic vowels displayed higher readings than unstressed and indefinite vowels. In the case of F_0 the difference was generally in the range of 10–20 Hz.

The results compiled from the readings for intensity and F_0 show that for both speakers both properties rise when a vowel is stressed and, largely, follow the same pattern. It would seem that these acoustic features are mixed in the production of stress in Meadow Mari. This combined effect of rises in F_0 and intensity would appear to be a cause of stress. What part then does duration play in the process?

Interpretation of the columns on the right in Table 1 indicate a clear difference between the degrees of stress on stressed compared to unstressed vowels as do the values for intensity and F_0 . So similar are they that simply on the basis of these figures alone it is impossible to single out any determining acoustic factor(s), and in addition duration could be just an accompanying feature. However, the part played by intensity and F_0 in regulating stress can still be assumed, but to what extent do these affect duration? Since the speaker exerts more energy in the production of a stressed than an unstressed sound, the resulting lengthening of the vowel would appear to be an accompanying feature. Yet, under certain circumstances vowel lengthening can be deliberately used to produce stress, for example, in the case of emphatic stress, when several acoustic alternatives are available to achieve the desired effect, even a lessening of such features, such as a marked drop in F_0 . Examination of Table 2 as a line chart would demonstrate clearly that stress can be rather easily located when referring to the duration curves, because stressed vowels nearly always appear at the top and unstressed in or near the troughs at the bottom of the curves, something which is not evident from corresponding intensity and F_0 charts. Even a close examination of the intensity and F_0 values in Table 1 gives little indication of the placement of stress.

Table 2. Table of duration values for title and first three sentences of recorded Meadow Mari speech used in the study for two informants EA and LS. The original text and transliteration are shown below the table.

EA						LS					
ma*	82	ne	39	mas	70	ma*	37	ne	53	mas*	68
rij	92	chap	51	tar*	77	rij	82	chap	43	tar	37
tyr#	97	la	71	eng*	64	tyr#	52	la*	48	eng*	142
ma*	64	na*	59	vla	93	ma*	43	na	71	vla	67
rij	83	ku*	65	kən	38	rij	65	ku*	48	kən	44
ka*	88	rəm	37	saj*	67	ka*	92	rəm	42	saj*	89
lək	47	gəch#	47	ən	62	lək	44	gəch#	47	ən	74

ak	75	ku*	66	əsh*	67	ak	52	ku*	47	əsh*	29
ret*	53	rə	54	tə	21	ret*	36	rə	77	tə	27
god	52	məsh	47	me	58	god	54	məsh	47	me	57
sek*	67	tu*	46	yz	63	sek*	66	tu*	59	ga*	85
tyr*	47	kəm	22	ga*	105	tyr*	32	kəm	60	rəsht	26
lö	45	gəch#	58	rəsht	57	lö	52	gəch#	67	mem	79
pu	47	tu*	59	mem	74	pu	48	tu*	64	nan*	91
sheng*	103	ku	30	nan*	60	sheng*	115	ku	52	i	60
ge	64	məsh	51	i*	52	ge	173	məsh	50	ləsh*	30
kyrt*	58	ka*	111	ləsh	45	kyrt*	27	ka*	105	nam*	53
njö	25	lək	56	nam*	63	njö	26	lək	46	sö	41
pu*	83	is	46	sö	23	pu*	88	is	32	ras	60
dər	38	kusst*	44	ras	59	dər	40	kuss(t)*	58	ta	31
go	69	(v)ən	43	ta*	41	go	65	(v)ən	44	rat*	108
gəch#	52	tra	68	ra(t)	86	gəch#	49	tra	45	da#	71
yz	65	di*	73	da#	85	yz	63	di*	81	jeng*	53
gar	60	tsij	36	jeng*	54	gar	48	(t)sij	26	vlak	59
lam*	90	zhe	49	vlak	60	lam*	88	zhe	32	lan	62
əsh*	40	kus	48	lan	63	əsh*	43	kus	46	ku	55
ten	67	nen*	78	ku	40	ten	49	nen*	88	a*	93
mosh*	67	vi	75	a*	109	mosh*	72	vi	73	nəm	43
tə	20	jang*	70	nəm	32	tə	37	jang*	127	kon*	88
məzh	31	to	81	kon	60	məzh	28	to	56	dat	84
de*	86	lesh*	109	dat*	119	de*	50	lesh*	94		

* = stressed syllable, # = divided opinion on stress. Values are in ms.

Original text

Марий Түр. Марий калык акрет годсек түрлө пушенге, күртнө пудырго гыч үзгарлам ыштен моштымыж дене чаплана. Курым гыч курымыш, тукум гыч тукумыш калык искусствын традицийже куснен, вияг толеш. Мастар ег-влакын сайын ыштыме үзгарышт мемнан илышнам сөрастарат да ег-влаклан куаным кондат.

Transiteration

ma rij tyr ma rij ka lək ak ret god sek tyr lö pu sheng ge kyrt njö pu dər go gəch yz gar lam əsh ten mosh tə məzh de ne chap la na. ku rəm gəch ku rə məsh tu kəm gəch tu ku məsh ka lək is kusst vən tra di tsij zhe kus nen vi jan to lesh. mas tar eng vla kən saj ən əsh tə me yz ga rəsht mem nan i ləsh nam sō ras ta rat da jeng vlak lan ku a nəm kon dat.

6. Conclusion

The interaction between and relative significance of specific acoustic parameters would appear to be quite complicated in Meadow Mari. Perhaps from the forgoing it may be tentatively concluded that intensity and F_0 together as a combination rise when stress occurs in Meadow Mari and that, as a result, vowel length is increased. The rise in intensity of, on average, two decibels is clearly audibly distinct. A significant difference is also evident in F_0 , usually in the range of 10–30 Hz. However, an examination of these acoustic parameters in Table 2 reveals the difficulty of deciding which words are actually stressed, even if the averages in Table 1 reveal these features as the cause. The Table 2 values show that neither intensity nor F_0 show particular prominence when individual words are examined, although it is clear that duration is a reliable indicator of stress.

The experiment described above suggests that word stress in Meadow Mari is indicated by vowel length, and the cause is a combination of intensity and F_0 , and that neither parameter seems to have an over-riding effect. This conclusion does not leave out the possibility that an increase in duration may also be causal and not only indicative, or at least in some instances.

References

Alhoniemi, Alho, (1985) *Marin kielioppi*, Helsinki: Suomalais-Ugrilainen Seura.

Botinis, Antonius. 1998. *Intonation in Greek*. In: Hirst, Daniel and Albert Di Cristo (eds), *Intonation Systems. A Survey of Twenty Languages*. Cambridge: Cambridge University Press, 288–310.

Estill, Dennis (2010) The acoustic correlates of Zyryan stress, *Congressus XI. Internationalis Fenno-Ugristarum*, Piliscsaba, Pars IV 9–14. VIII. 2010.

Estill, Dennis (2012) Revisiting the Meadow Mari vocalic system, *Linguistica Uralica*, Tallinn, 3/228–237.

Mariyskiy (2000) = Зорина З, Г. и др., Марийский язык для всех, Част 2/3 139, Йошкар-Ола: Марийское книжное издательство.

Tutkimushanke suomen prosodian variaatiosta

Tommi Nieminen², Tommi Kurki¹, Hamid Behravan¹, Heini Kallio³

¹ Turun yliopisto, ² Itä-Suomen yliopisto, ³ Helsingin yliopisto

tommi.nieminen@uef.fi, tommi.kurki@utu.fi, hamid.behravan@utu.fi, heini.kallio@helsinki.fi

”Suomen kielen alueellinen ja sosiaalinen variaatio” on Koneen säätiön rahoittama, elokuussa 2013 käynnistynyt tutkimushanke, joka pyrkii keräämään uusin menetelmin verkon yli puheaineistoa prosodian tutkimuksen tarpeisiin. Toukokuussa 2014 verkkosovellus saatiin julkiseen internetiin, ja tarkoitus on jatkaa aineiston keruuta näillä näkymin ainakin vuoden 2014 loppuun. Tässä artikkelissa esittelemme hankkeen taustaa, tavoitteita, menetelmiä sekä yhteistyökuvioita.

Asiasanat: prosodia, variaatio, puhekorpuksset

1. Uuden tutkimushankkeen taustaa

Elokuussa 2013 käynnistyi Koneen säätiön rahoittama tutkimushanke *Suomen kielen alueellinen ja sosiaalinen variaatio*, jonka tavoitteena on koota prosodian tutkimiseen sopiva ja yleiseen käyttöön tuleva puhekorpus sekä tutkia suomen prosodisten kielenpiirteiden variaatiota. Hankkeen perustivat vuoden 2013 keväällä FT, dosentti Tommi Kurki (joka toimii hankkeen vastuullisena johtajana) Turun yliopistosta ja FT Tommi Nieminen Itä-Suomen yliopistosta. Marraskuussa 2013 tutkimusryhmään liittyi keruusovelluksesta vastaava projektitutkija M. Sc. Hamid Behravan, ja vuoden 2014 alusta apurahatutkijana aloitti FM Heini Kallio Helsingin yliopistosta.

Johtavana ajatuksena on ollut se tosiasia, että vaikka puhuttua suomea ja siinä esiintyvää variaatiota on tutkittu paljon, prosodian eri osa-alueita (Iivonen 1998: 317) ja etenkin sen kielellistä variaatiota on tarkasteltu toistaiseksi niukasti. Fennistisessä sosiolingvistiikassa sekä murteen- ja variaationtutkimuksessa on käsitelty tyypillisesti fonologiaa, morfofonologiaa tai morfologiaa (Lappalainen 2004: 16; Nuolijärvi & Sorjonen 2005: 124–125,131; Kurki 2007: 152), toisinaan myös syntaktisia (ks. esim. Ikola, Palomäki & Koitto 1989; Forsberg 1998) ilmiöitä. Sen sijaan foneettinen variaationtutkimus, josta viime vuosikymmeninä on maailmalla kehkeytynyt uusi (osa-) tieteenalakin, sosiofonetiikka, on jäänyt suomalaisessa tutkimuksessa marginaaliin. Sosiolingvistien ja foneetikkojen yhteistyötä on ollut valitettavan vähän (vrt. kuit. esim. Mustanoja & O’Dell 2007). Tilanne on tässä suhteessa kovin erilainen kuin vielä ennen sotia.

Viime aikoina prosodisia kielenpiirteitä lienee tarkasteltu tai vähintään sivuttu useimmin keskusteluntutkimuksessa (esim. Ogden & Routarinne 2005), mutta prosodian ei silti voi sanoa olleen keskustelunanalyttisen tutkimuksen tai sen vastaanoton keskipisteessä (vrt. Laakso & Laitinen 2005); sitä on myös harvoin todennettu kokeellisin menetelmin. Tosin Sara Routarinne, jonka väitöskirjaa (2004) voikin variaationtutkimuksen näkökulmasta pitää yhtenä 2000-luvun kiinnostavimmista keskustelunanalyttisistä tutkimuksista, käsittelee tutkimuksessaan myös sitä, miten prosodisilla keinoilla voidaan säädellä ja jäsentää vuorovaikutusta. Keinojen alueelliseen tai sosiaaliseen variaatioon hänenkään tutkimuksensa ei ulotu. Hänen aineistonsa pääkaupunkiseudun teinityttöjen puheelle on ominaista lausumanloppuinen sävelkorkeuden loppunousu, jonka hän tulkitsee keskustelijoiden yhteisen riittävän maaperän ja seurattavuuden varmistamisen keinoksi; hän pitää kuitenkin selvänä, että loppunousun käyttökontekstit ovat tätä laajemmat ja että esimerkiksi koulumaailma tarjoaisi oivan mahdollisuuden tutkia ilmiön levinneisyyttä ja leviämissuuntaa (ks. Routarinne 2008: erit. 139–141). Hänen havaintojensa perusteella voi tosin sanoen olettaa kielenpiirteen ilmentävän pääkaupunkiseudun sosiaalista (sukupuoleen ja ikään perustuvaa) variaatiota, mutta sen alueellinen – tai laajempi sosiaalinen – variaatio jäävät selvittämättä.

Prosodian tutkimus on Suomessa enimmäkseen kohdistunut yleiskieleen tai lähinnä funktionaaliseen (diatyyppiseen) variaatioon (poikkeuksena esim. Ylitalo 2004, 2009). Jonkin verran on tehty tutkimuksia, joissa on tarkasteltu yksittäistä prosodista ilmiötä jostakin aluepuhesuomen muodosta, mutta yleensä ei ole päästy tarkastelemaan ilmiötä alueellisesti laajemmin. Koko kieliyhteisön kattavan prosodian korpuksen puuttuessa huomio on kiinnittynyt yksittäisen aluepuhekielen ominaisedustukseen.

Suomen tilanne eroaa olennaisesti naapurimaista Virossa ja Ruotsissa, joissa on onnistuttu luomaan nimenomaan prosodian tutkimukseen soveltuvia puhetietokantoja. Ruotsissa näitä on ollut toistakymmentä vuotta (Aho 2010: 14) ja Virossakin tutkimus on ollut vireää (esim. The Finno-Ugric Prosody -hanke³, Lehist 2007). Suomesta vastaava korpus puuttuu yhä.

2. Hankkeen tavoitteet

Hankkeellamme on kolme tavoitetta:

1. Puhesuomen prosodian variaation tarkastelun tarpeisiin koostetun korpuksen muodostaminen. Suomesta on suuria puhekorpuksia, mutta ne on koostettu ensisijaisesti muiden kielenilmiöiden

3 Verkko-osoite: <http://www.murre.ut.ee/phonetic-corpus/>.

kuin prosodian tarkasteluun. Murrehaastattelujen vapaahkosti tuotetusta tekstimassasta on vaikea poimia eri puhujilta funktionaalisesti toisiaan vastaavia puhunnoksia. Tutkijan ainoaksi mahdollisuudeksi on näin jäänyt tutustua aineistoon niin perinpohjaisesti, että hän on muistinvaraisesti pystynyt poimimaan siitä tyypillisiksi katsomiaan piirteitä (esim. Yli-Luukko 2001, 2010). Tämä vie tavattoman kauan aikaa.

Tämän hankkeen tavoitteena on panna alulle nimenomaisesti prosodian tutkimukseen sopiva ja alueellisesti edustava puhekorpus. Alkuvaiheessa korpuksen pyritään saamaan keskenään vertailukelpoiset osa-aineistot ainakin eri päämurrealueilta sekä pääkaupunkiseudulta. Aineistoa ei kerätä vain murrehaastattelujen tapaan vaan enimmäkseen äänittämällä informanttien kielelliset reaktiot erilaisiin kielellisiin tai kuvallisiin ärsykkeisiin. Tätä keruutapaa varten verkkoon on luotu hankkeen oma keruusivusto (ks. tavoitetta 3). Kun aineisto kerätään näin, varmistetaan osa-aineistojen keskinäinen vertailukelpoisuus. Tutkimustarpeiden mukaan on mahdollista hyödyntää myös aiempia korpuksia ja niiden äänitteitä. Keruusivuston välityksellä kerättävän aineiston lisäksi osa hankkeen aineistosta kerätään edelleen aidoissa kasvokkaisviestintätilanteissa. Syntyvä korpus sijoitetaan hankekauden ajaksi Turun yliopiston kieli- ja käännöstieteiden laitokseen, mutta se on tarkoitus koostaa siten, että se olisi yhteensopiva CSC:n kielentutkimuksen aineistojen ja LAT-ohjelmiston kanssa. Valmistuttuaan se saatetaan FIN-Clarinin kautta Kielipankkiin kielentutkijoiden yleiseen käyttöön.

2. Puhesuomen prosodisten ilmiöiden tarkastelu niin, että huomioon tulevat myös ekstralingvistiset eli alueelliset ja sosiaaliset tekijät. Tätä varten tarvitaan korpus, jossa eri puhemuodoista esiintyy mahdollisimman vertailukelpoisia puhunnoksia (”sama asia eri murtein”). Jo olemassa oleviin, vapaahkoa puhetta käsittäviin korpuksiinkin päästään käsiksi paremmin, kun ensin on saatu vertailukelpoiset parametrit kontrolloidusta aineistosta. Vertailukohtana voi käyttää tuoretta sveitsinsaksan puherytmin murrevariaatiosta tehtyä tutkimusta (Leemann ym. 2012), jossa neljältä päämurrealueelta kultakin valitut viisi informanttia (yhteensä siis 20) tuottivat samat puhunnokset tutkijoiden valmiiksi murteella kirjoittamista elisitoinneista.

Pyrimme siis vastaamaan kysymykseen: Millaisia mahdollisesti varioivia prosodisia piirteitä tai piireryppäitä nykysuomessa esiintyy? Tätä voi tarkentaa esimerkiksi seuraavin kysymyksiin: Millaisia jaksottavia sävelkuvioita esiintyy? Eroavatko eri puhesuomen muotojen äänialat toisistaan ja jos, niin miten? Toteutetaanko paino kaikissa suomen murteissa sävelkorkeuden eli F0:n (lausepaino tai ”aksentti”) ja keston (sanapaino eli tahdin ensitavu) yhdistelmällä? Seuraavatko sävelkorkeus- ja intensiteettikäyrät F0 ja A0 kaikissa murteissa uskollisesti toisiaan vai tuotetaanko

jossakin päin Suomea sävelkorkeus ”melodisemmin” tai ”dynaamisemmin”? Onko puherytmissä paino-, tavu-, tahti- tai mora-ajotteisempia murteellisia tai sosiaalisia variantteja?

3. Uusien puhekielen tutkimuksen keruumenetelmien kehittäminen ja testaaminen. Jo vanhastaan käytössä olleiden, tutkijan ja informantti(e)n samanaikaiseen läsnäoloon perustuvien keruumenetelmien rinnalle pyritään kehittämään joukkoistettua keruutapaa internetiä ja etenkin ns. Web 2.0:aa hyödyntäen; tästä lisää seuraavassa luvussa.

3. Puheaineiston joukkoistettu keruu verkossa?

Hankkeen yhtenä päätavoitteena on siis kehittää uusia puheaineistonkeruumenetelmiä. Keskeisenä toimintaperiaatteena on Web 2.0 -hengessä joukkoistaa (engl. *crowdsource*) keruuta niin, että aineistoa tuottaisivat maallikkokäyttäjät suhteellisen itsenäisesti. On tärkeää kuitenkin korostaa eroa ulkoistamiseen: tarkoitus ei ole yksinkertaisesti teettää työtä muilla vaan saada verkkoyhteisö myös jossain mielessä osallistumaan tutkimukseen ja hyötymään siitä.

Hankkeelle on verkkoon luotu sivusto⁴, jonka julkisella puolella kuka tahansa pääsee tutustumaan yleisluontoisiin, kirjallisuudesta poimittuihin murretietoihin sekä (tuonnempana; tätä osaa ei ole vielä toteutettu) tutkijaryhmän valikoimiin ääninäytteisiin ja karttuvan aineiston levikkitietoihin. Keruusivusto on avattu yleisölle 1.5.2014, ja keruu verkossa jatkuu ensi vaiheessa ainakin vuoden 2014 loppuun.

Julkisella puolella esitettävät tiedot koskevat pääosin – joskaan eivät pelkästään – vanhoja murteita ja vastaavat siis sisällöltään jo kirjallisuudessakin julki tuotua tietoa. Päästäkseen tutustumaan uusiin ääninäytteisiin pitää käyttäjän rekisteröityä sivustolle omalla tunnuksellaan, hyväksyä sivuston käyttöehdot sekä tuottaa ja tallentaa itse ääninäytteitä, jotka sivustolta kerätään erilliselle tallennuspalvelimelle.

4 Sivuston verkko-osoite on <https://puhu.utu.fi/>.



Kuva 1. Keruusovellussivuston julkinen etusivu.

Kuvassa 1 näkyy esimerkkinä sivuston etusivu. Vasemmanpuoleinen kartta esittää suomen vanhoja murrealueita. Karttaa napsauttamalla pääsee toiselle julkiselle sivulle, jossa käyttäjää on ensimmäisenä vastassa lyhyt kartan kuvaus. Kartan alueita voi edelleen napsauttaa ja saada näin lyhyen kuvauksen kustakin murrealueesta.

Keruussa käytettäviä elisointitehtäviä on tarkoitus olla useita erilaisia; toistaiseksi näitä on toteutettu neljä erilaista. Joka kerta sivustolle kirjautuessaan käyttäjä saa tehdä uuden tehtäväsarjan. Joissakin tehtävissä informantin tarvitsee vain lukea ruutuun tuleva teksti; tällainen on esimerkiksi rekisteröitymisen yhteydessä läpi käytävä äänityssarja, jonka pääasiallinen tarkoitus on testata laitteiston toimivuus ja mahdollisesti (tuonnempana) antaa informantille myös yleisluontoista palautetta äänitystasosta ja tekijöistä, joilla hän voi helposti vaikuttaa äänityksen laatuun. Sarjassa pyydetään esimerkiksi luettelemaan numerot yhdestä kymmeneen, jolloin saadaan käyttäjiltä myös näyte luettelointonaatiosta. Etukäteen voi jo olettaa vaihtelua tulevan käyttäjien välillä siinä, luetellaanko numerot ”täysin”, yleiskielen mukaisina (*yksi, kaksi...*) vai laskettaessa usein käytetyissä ”typistetyissä” puhekielen muodoissa (esim. *yy, kaa...*). Oleellista on kuitenkin, että vaikka variaatiota tulee, se pysyy tutkimuksen kannalta hallituissa rajoissa, ja jos aineistoa kertyy tarpeeksi – eli tarpeeksi moni ottaa osaa tutkimukseen –, kutakin vaihtoehtoa kertyy riittävän paljon.

Jotkin tehtävät kuten jo mainittu rekisteröitymisen yhteydessä käytettävä sarja, ovat yksinkertaisesti sellaisia, että kehote sanoo suoraan, mitä informantin tulisi sanoa. Tällöinkin kehotteessa

tyypillisesti pyydetään informanttia käyttämään omaa puheenparttaan eikä niinkään suoraan lukemaan näkyvää tekstiä. Toisissa tehtäväsarjoissa ruutuun tuleva kehote sen sijaan vain opastaa, mitä informantin tulee omin sanoin tehdä; yhdestä tällaisesta tehtävästä on näyte kuvassa 2. Näissä tehtävissä kehotteen ja informantin tuottamaksi toivotun puhunnoksen etäisyyttä pidentämällä pyritään paremmin simuloimaan tilannetta, jossa informantti pystyisi käyttämään omaa kieltään.



Kuva 1: Esimerkki kuvallisesta tehtävästä, jossa informantilta ei suoraan vaadita tiettyä puhunnosta vaan kuvan perusteella tuottamaan oma versionsa.

Samalla kun keruusovellus toimii keruun välineenä, se osaltaan auttaa myös testaamaan erilaisten tehtävämuotojen toimivuutta joukkoistetussa keruussa. Juuri tämän vuoksi tavoitteena on tuottaa ei vain monia tehtäväsarjoja vaan nimenomaan monia *tyypiltään erilaisia* tehtäviä. Myöhemmässä vaiheessa on sitten mahdollisuus lisätä tai karsia tehtävätyyppejä sen mukaan, mitkä ovat käytännössä osoittautuneet toimiviksi. Tässä yhteydessä on myös mahdollista käyttää hyväkseen maallikkokäyttäjiltä tulevaa suoraa palautetta, mikä tukee joukkoistuksen ideaa.

Vain tutkijoiden käytössä olevalta varsinaiselta tallennuspalvelimelta näytteitä siirretään valikoiden keruusivuston julkiseen ja puolijulkiseen eli tavallisille käyttäjille tarkoitettuun osioon, jossa ääninäytteen antaneet pääsevät niihin tutustumaan.

Keruuseen osallistuvat maallikot äänittävät kielenäytteitä siis kotikoneeltaan tai muulta vastaavalta käytössään olevalta laitteelta, joka on verkossa. Tärkein laitteistoon (verkkoyhteyden lisäksi) kohdistuva vaatimus on, että tietokoneessa on mikrofoni, mikä yleensä pitää paikkansa kaikkien nykyaikaisten kannettavien tietokoneitten osalta, ja pöytäkoneisiinkin voi monen käyttäjän olettaa hankkineen jonkinlaisen mikrofonin Skypen tai vastaavan internetpuhelin- tai chat-ohjelman käyttöä varten. Kehitysalustana on HTML 5 ja Flash, joista jälkimmäinen ei ehkä ole aivan

optimaalinen mutta joka tällä hetkellä on ainoa riittävän hyvin tuettu alusta, jolla äänitys verkon yli on mahdollista. Mobiilikäyttöympäristöissä Flash tosin ei välttämättä ole käytettävissä lainkaan, ja tämän takia mobiiliympäristöihin (niin Android- kuin iPhone-alustoillekin) on kehitetty varsinaisen sivuston ohessa omia toimintatapojaan; nämä ovat toistaiseksi enemmän kesken.

Näytteet anonymisoidaan tallennettaessa siten, että sivustolla näytteen antajasta kerrotaan enimmilläänkin vain informantin asuinpaikkakunta tai -seutu, sukupuoli ja ikäryhmä. Tutkimustavan mielekäs toteuttaminen vaatii jokaista informanttia kirjautumaan sivustolle omilla henkilötiedoillaan ensikirjautumisen yhteydessä, mutta näitä taustatietoja ei luovuteta tutkimusryhmän ulkopuolelle. Korpuksen valmistuttua aineistoa käyttävä tutkija saa käyttää niitä vain tutkimuskäyttöön ja siinäkin tapauksessa vain, jos tutkija pystyy perustelemaan, mihin hän tarvitsee taustatietoja. Vastanneiden taustatietoja ei säilytetä keruusivustolla eikä missään verkossa tutkimushankkeen aikana. Korpuksen valmistuttua, jolloin korpus siirtyy tutkijoiden käyttöön yleisemmin, pääsyä vastaajien taustatietoihin valvotaan asianmukaisesti ja vastaajien yksityisyydensuojaa kunnioitetaan. Sivulle rekisteröityessään vastaaja voi myös rajoittaa ääninäytteidensä ja taustatietojensa käyttöä.

Sen lisäksi, että keruusivustolle kirjautuneet pääsevät tutustumaan omia näytteitään koskeviin tuloksiin ja suhteuttamaan tuotoksiaan muiden vastaajien tuotoksiin, annetaan heille mahdollisuus osallistua näytteiden arviointiin. Vaikka siis tutkimusryhmä vastaa viime kädessä korpuksen muodostamisesta ja sen keruun koordinoinnista, vapaaehtoisesti mukaan tulevilla aktiivisilla maallikoilla ja asiaan vihkiytyneillä harrastajilla ja alan ammattilaisilla on merkittävä osa hankkeessa.

Verkossa kerättävän aineiston lisäksi hankkeen tutkimuksessa on tarkoitus hyödyntää myös muuta aineistoa. Näin päästään todennäköisesti paremmin käsiksi sellaisiin prosodian ilmiöihin kuin spontaanille puheelle ominaisiin laajoihin intonaatiojaksoihin. Testiympäristö on aina sikäli laboratoriomainen, että tutkimuksen informantin vastapuolena on aina vain ennalta käsikirjoitettu koneen vastaus tai ennalta rakennettu tehtävä, mistä seuraa, ettei tilanne ole erityisen spontaani eikä vuorovaikutuksellinen. Tätä kuitenkin tasapainottaa se, että prosodisten piirteiden tietoinen kontrolli on vähäistä. Kielenpuhujalla on yleensä niukasti tietoa siitä, miten esimerkiksi painollinen tavu tuotetaan – mikä ei ole yllättävää varsinkin, kun havaitsee tutkimuskirjallisuudenkin monilta osin ristiriitaiseksi. Esimerkiksi suomen painon on toisaalta kuvattu toimivan ensisijaisesti F0:n (tai F0:n ja A0:n liki yhtäaikaisen) nousun avulla mutta toisaalta myös keston avulla. Samoin intonaation kielellisen käytön (kysymyksen tai jatkuvuuden merkitseminen) ja dialektaalisen ja diatyyppisen

käytön (murretaustan tai sosiaalisen tilanteen merkitsemisen) täsmällinen suhde suomen puhemuodoissa on toistaiseksi tutkijoillekin tuntematonta aluetta.

Verkon yli äänittämiseen liittyy myös ilmeisiä teknisiä vaikeuksia, mikä heikentää kerättävän aineiston teknistä laatua. Tätä puolestaan tasapainottaa se, että tutkimuskohteina olevat prosodiset piirteet ovat varsin robusteja: perussävel ja kestot pystytään selvittämään tarvittavalla tarkkuudella varsin huonolaatuisestakin materiaalista. Siitä huolimatta osa aineistosta tulee pakostakin olemaan teknisesti liian heikkoa, että sitä voitaisiin analysoida kovin hienojakoisesti. Kaikkia mahdollisia kielenpiirteitä ei ehkä voida tutkia lainkaan. Toisaalta, kun rajoitus tiedostetaan jo etukäteen, tarkasteltavat piirteet ja ilmiöt voidaan valikoida samoin kuin verkkosovelluksessa käytettävät ärsykkeet suunnitella niin, että mahdolliset vaikeudet ja ongelmat minimoidaan.

4. Tutkimusyhteistyö

Hanke pyrkii aktiivisesti eri tutkimusalojen väliseen yhteistyöhön. Hankkeen ydinryhmä (Kurki ja Nieminen) edustavat sosiolingvistiikan, (fennistisen) fonologian ja fonetiikan tuntemusta, ja ensimmäisenä välitavoitteena kurotaankin nimenomaan umpeen variaation tutkimuksen ja fonetiikan välille Suomessa auennutta kuilua. Projektitutkijana toimiva Hamid Behravan puolestaan tuo tutkimusryhmään tietotekniikan ja laskennallisten analyysimenetelmien tuntemusta ja väitöskirjantekijä Heini Kallio vahvistaa fonetiikan tuntemusta. Hankkeeseen nimettyinä tutkimusyhteistyökumppaneina ovat Michael L. O'Dell Tampereen yliopistosta, Karl Pajusalu Tarton yliopistosta, Maija S. Peltola Turun yliopistosta sekä Stefan Werner Itä-Suomen yliopistosta. Hankkeeseen haetaan myös muita tutkimusyhteistyökumppaneita. Tavoitteena on päästä uusilla menetelmillä kerätyn aineiston avulla tarkastelemaan mahdollisimman laajasti suomen prosodiaa ja sen variaatiota.

Lähteet

Aho, Eija 2010: *Spontaanin puheen prosodinen jaksottelu*. Helsingin yliopiston nykykielten laitos, yleinen kielitiede.

Forsberg, Hannele 1998: *Suomen murteiden potentiaali: Muoto ja merkitys*. Suomalaisen Kirjallisuuden Seuran Toimituksia 720. Helsinki: Suomalaisen Kirjallisuuden Seura.

Iivonen, Antti 1998: Intonation in Finnish. – Daniel Hirst & Albert Di Cristo (toim.): *Intonation systems: A survey of twenty languages*. Cambridge: Cambridge University Press.: 311–327.

Ikola, Osmo, Ulla Palomäki & Anna-Kaisa Koitto 1989: *Suomen murteiden lauseoppia ja tekstikielioppia*. Suomalaisen Kirjallisuuden Seuran Toimituksia 511. Helsinki: Suomalaisen Kirjallisuuden Seura.

- Kurki, Tommi 2007: Variaation tutkimuksen nykynäkymiä. *Sananjalka* 49: 143–162.
- Laakso, Minna & Lea Laitinen 2005: Keskustelujen kielioppia tyttöjen kielessä. *Virittäjä* 109. (Arvostelu Sara Routarinteen väitöskirjasta *Parenteesit ja nouseva sävelkulku kertojien vuorovaikutuskeinona*, 2004.) 117–123.
- Lappalainen, Hanna 2004: *Variaatio ja sen funktiot: Erään sosiaalisen verkoston jäsenten kielellisen variaation ja vuorovaikutuksen tarkastelua*. Suomalaisen Kirjallisuuden Seuran Toimituksia 964. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Leemann, Adrian ym. 2012: Dialectal typology based on speech rhythm. Julkaisematon poster: Perspectives on Rhythm and Timing, Glasgow (Iso-Britannia) 19.–21.7.2012.
- Lehiste, Ilse 2007: The Finno-Ugric Prosody Project. *Linguistica Uralica* XLIII.1: 1–10.
- Mustanoja, Liisa & Michael L. O’Dell 2007: Suomen d ja r sosiofoneettisessa kentässä. *Virittäjä* 111.1: 56–67.
- Nuolijärvi, Pirkko & Marja-Leena Sorjonen 2005: *Miten kuvata muutosta? Puhutun kielen tutkimuksen lähtökohtia murteenseuruuhankkeen pohjalta*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 133. Helsinki: Kotimaisten kielten tutkimuskeskus.
- Ogden, Richard & Sara Routarinne 2005: The communicative functions of final rises in Finnish intonation. *Phonetica* 62: 160–175.
- Routarinne, Sara 2004: *Tytöt äänessä: Parenteesit ja nouseva sävelkulku kertojan vuorovaikutuskeinona*. Suomalaisen Kirjallisuuden Seuran Toimituksia 903. Helsinki: Suomalaisen Kirjallisuuden Seura.
- 2008: Miksi intonaatio nousee? – Sara Routarinne & Tuula Uusi-Hallila (toim.): *Nuoret kielikuvassa: Kouluikäisten kieli 2000-luvulla*. Tietolipas 220. Helsinki: Suomalaisen Kirjallisuuden Seura: 125–145.
- Yli-Luukko, Eeva 2001: Yläsatakuntalainen intonaatiokuvio. *Virittäjä* 105.1: 2–21.

A short term study of Hungarians learning Finnish vowels

Terhi Peltola¹, Pertti Palo², Olli Aaltonen¹

¹ Helsingin yliopisto, ² Queen Margaret University, Edinburgh

terhi.peltola@helsinki.fi, ppalo@qmu.ac.uk, olli.aaltonen@helsinki.fi

A group of Hungarian students (n=10) participated in a Finnish phonetics and conversation course during the first 3 months of their language studies. During the course, the students trained in the allophonic variation of Finnish speech sounds, comparing them to Hungarian sounds and participating in group conversation exercises. We call the method used on the course conscious phonetic training of foreign language speech sounds. Additionally, the students participated in one-on-one imitating exercises, which were recorded for the current study. We followed the participants' foreign sounds pronunciation development during the first semester of their studies and compared it to their peers (n=4). The results suggest that participating in the course affected the students' pronunciation skills towards the end of the three-month course, whereas, at the beginning, both of the groups' pronunciation was more similar.

Keywords: foreign language learning, native language interference, conscious phonetic training, acoustic phonetics

1. Introduction

Learning foreign sound categories similar to native speech sounds is a difficult task. Perception and production take place without conscious control, and native categories control the learners' perception and production of foreign sounds on a preattentive level (e.g. Peltola et al. 2003). Thus, mastering new foreign sounds is easier than learning foreign sounds similar to native sounds (Flege 1987).

Native sounds' governance of the preattentive perception of foreign sounds can also be considered native language interference on foreign language. However, native language sounds' interference can be interfered, and perhaps even diminished or changed, with phonetic training of foreign languages sounds.

In the current paper, we discuss the possibility of effecting the native sounds' patterns of governance over the similar foreign language sounds by using conscious phonetic training as a classroom teaching method. We also discuss what type of training seems to be effective and how much training is needed, as well as over how long period of time that training needs to take place.

The teaching method used in the current study involves the students in simple pronunciation exercises, which could perhaps be described as drills, as well as longer conversation exercises. We wanted to pursue the issue of whether there is any difference in using passive listen-repeat exercises compared to using exercises that encourage the students to learn by doing in a normal classroom environment. This approach relies on the motor theory of language, according to which speech sounds are encoded in the human memory as a series of the movements they employ (Lieberman & Mattingly 1985). However, the sizes of the participant groups (n=10 and 4) still leave the results rather uncertain.

2. Hungarian and Finnish vowels in comparison

The Finnish phonemic differences causing the most problems (among the vowels) for the Hungarian learners are the ones between the front vowels /i-e-æ/. Comparing the Hungarian /e:-ɛ/ categories to the Finnish /e-æ/ equivalents reveals that there is a large gap in the Hungarian vowel space, whereas in Finnish, the front vowels are more evenly located over that corner of the vowel space.

There are also some phonetic differences in the back vowels. However, these differences do not cause as significant difficulty to the learner as front vowels, since there are more open back vowels in Hungarian than in Finnish (over differentiating). Sometimes differentiating between the Finnish front /æ/ and back /ɑ/ is also difficult for the Hungarians, because the category boundaries are not on the same area.

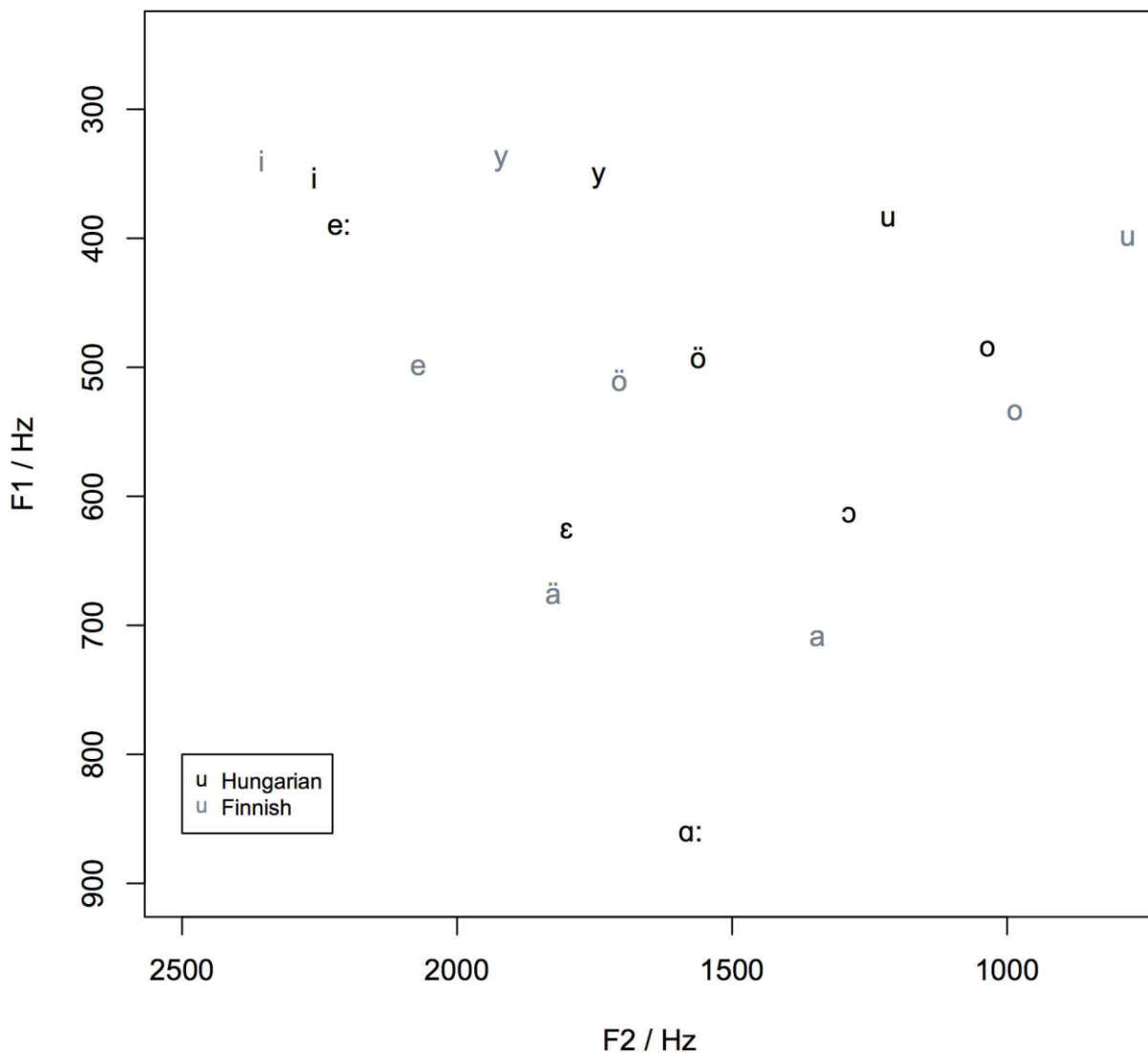


Figure 1. Finnish and Hungarian vowels in comparison: The average formant values of Hungarian vowels (in black) from the subjects of this study and those of Finnish vowels (in grey, according to Wiik 1965).

3. Materials and methods

In this chapter, we describe the data collection methods, classroom teaching, methods of acoustic phonetics used in the study, as well as statistical methods.

3.1. Data collection and processing

Hungarian university students participated in an optional Finnish phonetics and conversation exercise course during the first semester of studying Finnish language and culture. The students trained in correct pronunciation twice per week for 45 minutes in the lessons, which consisted of studying phonetics, conscious phonetic training of Finnish speech sounds and their allophones, and imitation and conversation exercises with native informants – both with the whole group and in smaller groups of three to five students.

In the beginning of the course, half way through, and at the end, the students participated in individual exercises, which were recorded, processed, and analyzed for this study. The exercises consisted of seven blocks of pseudo and real Finnish two-syllable words. The first and the last blocks were read (baselines) and the five in between were imitated by following the native instructor closely.

We followed the development of the students pronunciation during the course. Additionally, we compared the students participating in the phonetics course with the ones who did not. The peer group only took part in the data collection, i.e. reading and imitation exercises, without any phonetic or conversation exercises.

The first and the last (read) blocks of words of the one-on-one exercises were segmented and vowel formants were extracted automatically with a Praat script (Lennes 2003).

3.2. Statistical models

The formant data were analysed with mixed effects models (e.g. Baayen et al. 2008) in R (R Core Team 2012). Four separate models were fitted to the formant data (F1 and F2) of short vowels of the students who participated in the exercises and those who did not. Each combination of formant and group was first fitted with a model with the factors' gender, vowel segment, recording session and first/last (indicating whether the sample was from the first or the last baseline of the session) as fixed effects and the subject as a random effect.

The final models were chosen by first stepping up model complexity (number of parameters) and then stepping down as insignificant effects were identified based on the corresponding t-values. Inclusion/exclusion of independent variables was decided by testing for significance with a likelihood ratio test.

4. Results

All final mixed effect models had the same effects: gender and an interaction of session with the segment and an interaction of the sole random effect – subject – with the segment. The mainly interesting effect of session/vowel interaction is illustrated in Figure 2. The plot was produced by plotting for both groups the interaction effect of session and segment for both analysed formants. The figure shows arrows corresponding to the predicted change in formant values of each vowel from the first recording session to the last.

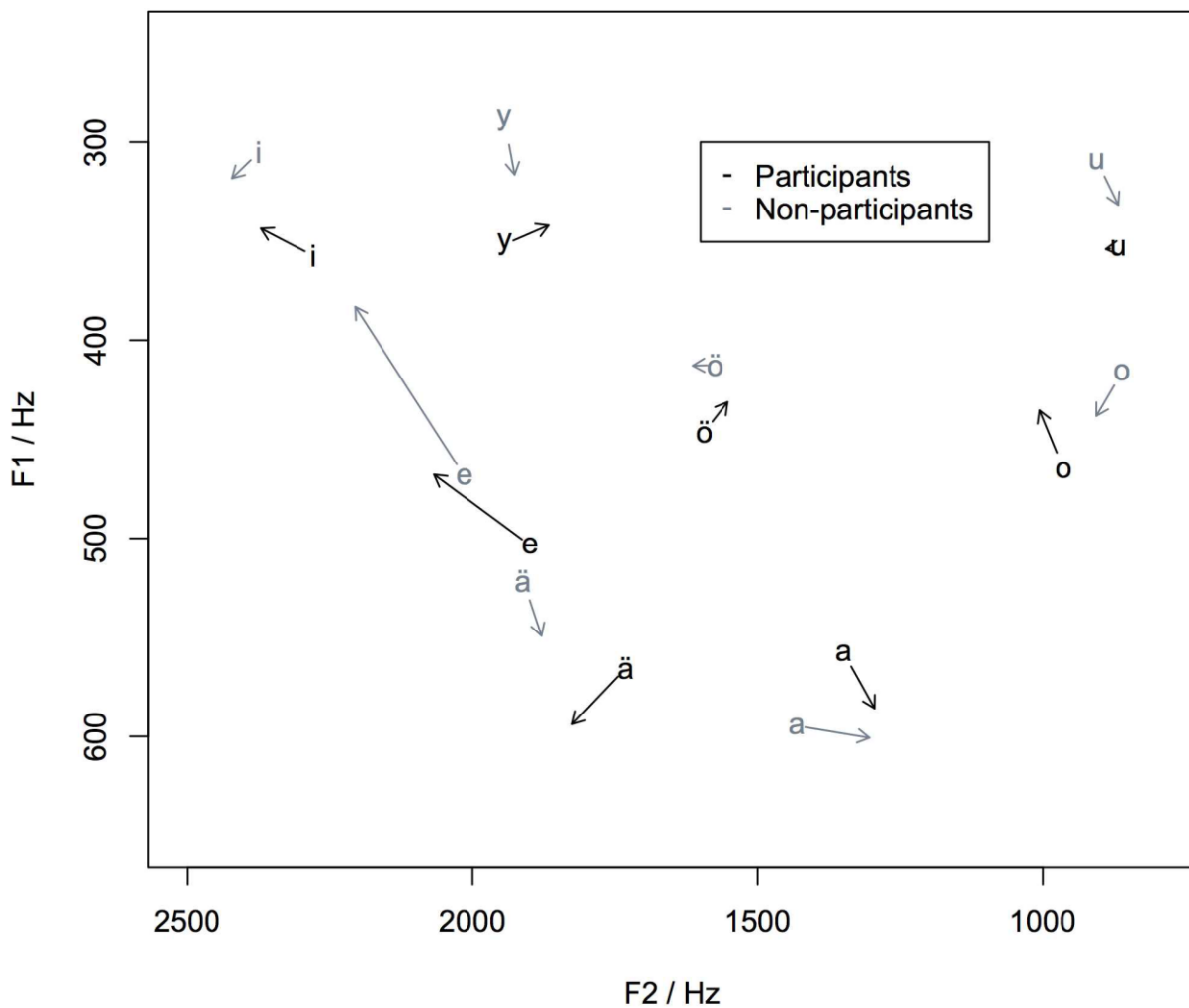


Figure 2. Interaction effects of segment (vowel) and session for both participants and non-participants in F2/F1 space.

4.1. Changes from the beginning to the end of the semester

The imitation exercises did not effect the formant qualities immediately. That is, the baselines from the beginning and the end of a given exercise did not differ statistically. The change was rather gradual and happened during a longer period of time.

Overall, the students who participated on the course have slightly more centralized vowels compared to their peers.

The data seem to indicate that the students who did not participate on the course performed better at the beginning of the course, and rather surprisingly, their /e/ pronunciations drew nearer to the closer Hungarian equivalent towards the end of the first semester language studies.

5. Discussion

It seems pronunciation training used in this study has two types of effect on pronunciation. Firstly, the students' pronunciation becomes more *fluent* as some of the vowels' F1 and F2 values are slightly more centralized.

Secondly, pronunciation is more *stable* over a period of time. It appears that all the students were motivated in the beginning of the training, and successfully imitated the model. However, only the ones participating in the pronunciation training course twice per week managed to maintain the differences between the native categories and similar foreign language categories at the end of the course.

Thus, it seems the conscious phonetic training during the first semester of foreign language studies helps the students to create new categories for the foreign vowels /e-æ/ instead of using their own L1 categories over a longer period of time. However, it is impossible to say with these data whether the effect is due to the motivation of the students or conscious phonetic training method and training of the motorics involved in the pronunciation of the given sounds.

References

Baayen, R.H.; Davidson, D.J.; Bates, D.M (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59: 390–412.

Felge, J. (1987). The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15. Pp. 47-65.

Lennes, M. (2003). The Speech Corpus Toolkit for Praat. <http://www.helsinki.fi/~lennes/praat-scripts/>

Lieberman, A. M., Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21 (1): 1-36.

Peltola, M. S., Kuntola, M., Tamminen, H., Hämäläinen, H. & Aaltonen, O. (2005). Early exposure to non-native language alters preattentive vowel perception. *Neuroscience Letters*, 388, 121-125.

R Core Team (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Wiik, K. (1965). Finnish and English Vowels. A comparison with special reference to the learning problems met by native speakers of Finnish learning English. *Annales Universitatis Turkuensis, Series B*, 94. University of Turku. 94.

Perception of focus size and focus type in Estonian

Heete Sahkai, Mari-Liis Kalvik, Meelis Mihkla

Institute of the Estonian Language, Tallinn, Estonia

Heete.Sahkai@eki.ee, Mari-Liis.Kalvik@eki.ee, Meelis.Mihkla@eki.ee

The paper presents the results of a perception study aimed at establishing whether broad and narrow focus, and narrow contrastive and information focus are perceptually distinguishable in Estonian. The study verifies the results of two previous production studies. The data consist of identical sentences produced with different information structure in a semi-spontaneous production task. The sentences were elicited by asking questions about pictures, and the perception task consisted in matching the answers with questions. The results suggest that Estonian conventionally uses emphasis to distinguish narrow focus from broad focus, but does not distinguish contrastive focus from information focus. The study also suggests that no distinction is made between different sizes of broad focus.

Keywords: Estonian, information structure, perception

1. Introduction

The paper reports the results of a perception study that aimed to verify some of the findings of our two previous production studies on the prosodic effects of information structure in Estonian (Sahkai, Kalvik & Mihkla 2013a,b). The studies examined the prosodic correlates of focus size (broad and narrow focus) and focus type (information focus and contrastive focus), assuming the autosegmental-metrical framework (e.g. Ladd 2008).

The previous studies found that the marking of narrow focus depends on the clause-final *versus* non-final position of the focused constituent. While sentences with non-final narrow focus were distinguished from broad-focus sentences by phonological features like pitch accent distribution, pitch accent type on the focal constituent, and prosodic phrasing, sentences with final narrow focus differed from broad-focus sentences only in the level of prominence strength: they displayed a significant lengthening and, for some informants, a larger pitch range in the focal constituent, and a shortening in the pre-focal constituent. It is cross-linguistically common that emphasis is used to disambiguate information structural meanings in phonologically identical sequences (Ladd 2008:254-257). However, the linguistic status of emphasis is controversial: it may be interpreted categorically, but it is perceived gradually (cf. Ladd & Morton 1997). A perception test is therefore

necessary in order to verify whether the level of prominence strength is indeed interpreted as a conventional marker of a categorical difference in focus size.

Our previous studies also checked for a possible distinction between different sizes of broad focus, specifically, sentence focus (all-new) and VP focus. It has been found that the phonetic correlates of focus size may vary in function of the size of broad focus, but it seems that such differences tend not to be significant (cf. Baumann et al. 2006). Our results confirmed this as we found no significant difference in prominence strength between the two broad-focus conditions. However, all-new sentences differed from narrow-focus and VP-focus sentences in that they exhibited a smaller proportion of upstepped nuclear accents. A perception test is therefore again needed in order to verify whether this could affect the interpretation of focus size.

Regarding focus type, the aim of our previous study was to establish whether the positing of a separate category of contrastive focus is warranted in Estonian (for an overview of the discussion on the status of contrastive focus see e.g. Hartmann 2008). The study concentrated on narrow focus. It found that in case of non-final narrow focus, sentences with contrastive focus exhibit a stronger prominence on the focal noun. They also exhibited a potential phonological difference as post-focal constituents were much more often labelled as deaccented than in information-focus sentences (in 45% vs. 10% of cases). However, the labelling of post-focal accentuation was based on preliminary criteria that need further verification. Again, a perception study would permit to verify whether the difference in prominence strength distinguishes between focus types and whether information-focus and contrastive-focus sentences are likely to differ in post-focal accentuation.

As for final narrow focus, our previous study hypothesised that prominence strength would not distinguish between contrastive and information focus as it had already been found to distinguish between broad and narrow focus. Namely, earlier production and perception studies suggest that prominence strength makes binary distinctions (Hanssen et al. 2008, Ladd & Morton 1997), and it is therefore not expected to make a three-way distinction between broad and narrow information focus on the one hand, and narrow information and contrastive focus on the other hand. This prediction was borne out to the extent that the sentences with contrastive *versus* non-contrastive narrow focus did not differ in the duration or pitch range of the focal and pre-focal constituents. Interestingly however, a three-way distinction was nevertheless obtained by way of different distributions of binary durational distinctions in the different constituents, suggesting that final narrow and contrastive focus might still be distinguishable (for details see Sahkai et al. 2013b). A perception study is therefore needed in order to establish whether this kind of global variation in

patterns of prominence strength could be a conventional means for expressing information structural meanings.

To summarise, the goal of the present study is to verify the following four hypotheses:

1. Sentence-final narrow focus is distinguishable from broad focus.
2. Broad foci of different sizes are not perceptually distinguishable from each other.
3. Non-final narrow contrastive focus is distinguishable from non-final narrow information focus.
4. Final narrow contrastive focus is not distinguishable from final narrow information focus.

Additionally, the study is expected to clarify the status of emphasis, upstep and patterns of prominence strength as potential correlates of information structural meanings, and to contribute to the study of post-focal (de)accentuation in Estonian. It will also contribute to the theoretical discussion as to whether contrastive focus should be considered a separate linguistic category.

2. Data and procedure

The study is based on semi-spontaneous production data elicited for the purposes of the two previous production studies. They consist of 8 sentences with the structure NOUN - VERB - NOUN, each produced by 9 informants with 6 different focus structures (see Table 1).

Table 1. Focus conditions, illustrated with the translations of the questions that were used to elicit the sentence *Mees joob kokat* ‘the man is drinking coke’ with the 6 different focus structures.

Focus size Focus type	Broad focus	Narrow focus	
		Final	Non-final
Information focus	1. Sentence focus (What is happening?) 2. VP focus (What is the man doing?)	3. Object focus (What is the man drinking?)	4. Subject focus (Who is drinking coke?)
Contrastive (corrective) focus	-	5. Contrastive object focus (Is the man drinking wine?)	6. Contrastive subject focus (Is the woman drinking coke?)

The sentences were elicited using pictures and elicitation questions based on the “Questionnaire on Information Structure” of Skopeteas et al. (2006); for more details see Sahkai et al. (2013a,b).

The data were subjected to evaluation by informants via an on-line survey application in the form of three separate tasks. The task of the informants was to listen to the sentences and to match them with the elicitation questions. There was no time limit, the informants could listen to the sentences several times and correct their answers. The tasks were performed respectively by 15, 8 and 14 informants.

The first task tested for the perception of focus size (hypotheses 1 and 2) and was conducted on a subset of the four types of information focus sentences (two randomly selected sentences of each type from each participant of the production task, altogether 72 sentences): sentence focus, VP focus, object focus, subject focus. The subject focus sentences were included as a benchmark: since they are categorically different from all the others, their recognition rate permits to interpret those of the other three conditions. The informants listened to the sentences one by one in a random order and were asked to attribute each sentence to one of the four elicitation questions.

The second test was ran in order to verify the results of the first task. It involved just two conditions, non-final narrow focus and sentence focus (7 sentences of each type from each participant of the production task, altogether 126 sentences). The subjects listened to sentence pairs involving the same sentence spoken once with narrow focus and once with broad focus by the same speaker. Their task was to judge which sentence was the answer to the narrow-focus elicitation question. In this task, all the sentences produced by the same speaker were presented in sequence.

The third task tested for the perception of focus type (hypotheses 3 and 4). It was conducted on a subset of the contrastive and non-contrastive subject- and object-focus sentences (5 randomly selected sentences of each type from 8 participants of the production task, altogether 160 sentences). The design was the same as in the second task: the informants listened to contrastive and non-contrastive sentence pairs spoken by the same speaker and were asked to pick out the answer to the contrastive focus elicitation question. Again, all the sentences produced by the same speaker were presented in sequence.

3. Results

The results of the first task confirm that the two broad focus conditions are not distinguishable from each other, but narrow focus is distinguishable from broad focus. Table 2 shows for each focus condition (in rows) the overall percentage of each judgment assigned to it by the informants (in

columns). Since the data contained obvious production errors the most outlying sentence of each condition was discarded. The bold figures on the diagonal show the overall recognition rate of each condition. Both the non-final and final narrow focus sentences received a proportion of 67% of correct judgments (i.e. assignments to the correct elicitation question). The identical recognition rate of the phonologically marked non-final narrow focus and the phonetically marked final narrow focus confirms that emphasis does indeed serve to signal focus size. However, the categorical difference of the non-final narrow focus from all the other conditions is reflected in the fact that the sentences in the other conditions were very rarely judged as answers to the subject-focus elicitation question; this is not the case in the final narrow focus condition, which is only gradually distinct from broad focus.

For VP-focus and sentence-focus sentences, the proportion of correct judgments was respectively 34% and 26%, suggesting that they are not distinguished from each other.

Table 2. Recognition rate of non-final narrow focus, final narrow focus, VP focus and sentence focus.

Focus conditions	Judgments of informants			
	Non-final narrow focus	Final narrow focus	VP focus	Sentence focus
Non-final narrow focus	67.2	5.3	14.6	13.0
Final narrow focus	1.2	67.3	18.7	13.2
VP focus	3.0	43.5	33.6	20.2
Sentence focus	3.7	35.0	35.4	26.3

The second test confirmed that final narrow focus and sentence focus are indeed distinguished from each other: the overall percentage of correct judgements was 66% (cf. Table 3). Also, 70% of the tested sentences were correctly recognised by 60% or more of the informants.

Table 3. Recognition rate of final narrow focus.

Focus conditions	Judgments of informants	Narrow focus
	Narrow focus	
Broad focus		34%

The results of the third task suggest that neither non-final nor final narrow contrastive focus is reliably distinguished from narrow information focus. However, the hypotheses were correct to the extent that non-final contrastive focus was slightly more often correctly recognised than final contrastive focus. Table 4 presents the overall distribution of informants' judgments (in columns) between the focus conditions (in rows). While in non-final position the percentage of correct judgments is 56.9%, in final position it is 49.1%.

Table 4. Recognition rate of narrow contrastive focus in sentence-final and non-final position.

Focus conditions	Judgments of informants	
	Non-final contrastive focus	Final contrastive focus
Non-final information focus	43.1%	
Non-final contrastive focus	56.9%	
Final information focus		50.9%
Final contrastive focus		49.1%

The slightly better recognition rate in non-final position is also illustrated by the fact that the percentage of sentences that were correctly recognised by 60% or more of the informants is 50% in the non-final focus condition and only 27.50% in the final focus condition. However, even in the non-final condition the recognition of contrastive focus is weaker than that of narrow focus, suggesting that it is not conventionally distinguished from information focus.

4. Discussion and conclusions

The results confirm the hypothesis that Estonian uses the level of prominence strength to distinguish between broad and narrow focus. They also confirm that no distinction is made between

different sizes of broad focus and hence that their difference in the proportion of upstepped accents is not meaningful.

Contrastive and information focus, on the other hand, were not distinguished from each other. This result does not mean necessarily that there are no perceptible differences between the contrastive and non-contrastive sentences. Rather, it means that the possible differences are not conventionally associated with a particular meaning. Estonian data thus supports the position that contrastive focus is not a separate category.

The study suggests that emphasis is meaningful only when it is conventionally associated with a linguistic category: it distinguishes (final) narrow focus from broad focus, but not (non-final) contrastive focus from information focus, although significant differences in prominence strength were found in both cases.

The results also show that the three-way difference in prominence strength distribution between broad focus, final narrow information focus and final narrow contrastive focus that was discovered in the production study is not a distinctive feature.

Finally, the fact that sentences with contrastive and non-contrastive non-final focus were not distinguished from each other suggests that they do not differ in post-focal accentuation and that the existence of post-focal accentuation in Estonian requires further study.

References

Baumann, S., Grice, M., Steindamm, S. (2006). Prosodic marking of focus domains – categorical or gradient? In: Hoffmann, R. & Mixdorff, H (eds.) *Proceedings of Speech Prosody: 3rd international conference, Dresden, May 2-5, 2006*, Dresden: TUDpress, pp. 301-304.

Hanssen, J., Peters, J., Gussenhoven, C. (2008). Prosodic effects of focus in Dutch declaratives. *Proceedings of Speech Prosody*, pp. 609-612. (Online: <http://sprogis.isle.illinois.edu/sp2008/>, accessed on 31 March 2014).

Hartmann, K. (2008). Focus and Emphasis in Tone and Intonational Languages. In: Steube, A. (ed.) *The Discourse Potential of Underspecified Structures*. Berlin; New York: Walter de Gruyter, pp. 389-411.

Ladd, D.R. (2008). *Intonational phonology*. Cambridge: Cambridge University Press.

Ladd, D.R. & Morton, R. (1997). The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics* 25, pp. 313-342.

Sahkai, H., Kalvik, M.-L. & Mihkla, M. (2013a). Prosodic effects of Information Structure in Estonian. In: Asu, E. L. & Lippus, P. (eds.) *Proceedings of International Conference: Nordic Prosody XI, Tartu, 15.-17.08.2012*. Frankfurt am Main: Peter Lang Verlag, pp. 323-332.

Sahkai, H., Kalvik, M.-L. & Mihkla, M. (2013b). Prosody of contrastive focus in Estonian. In: *Proceedings of INTERSPEECH-2013: 14th Annual Conference of the International Speech Communication Association, Lyon, France, August 25-29, 2013*, pp. 315-319.

Skopeteas, S., Fiedler, I., Hellmuth, S., Schwarz, A., Stoel, R., Fanselow, G., Féry, C., Krifka, M. (2006). Questionnaire on Information Structure: Reference Manual. *Interdisciplinary studies on information structure*, Vol. 4. Universitätsverlag Potsdam. (Online: <http://opus.kobv.de/ubp/volltexte/2007/1241/>, accessed on 15 Feb 2012.)

Foneettiset ääntämisohjeet osana kuuntele ja toista –harjoitusta

Antti Saloranta^{1,2}, Henna Tamminen^{1,2}, Paavo Alku³ & Maija S. Peltola^{1,2}

¹ Fonetikka, Turun yliopisto, ² LAB-lab, fonetiikka, Turun yliopisto, ³ Signaalinkäsittelyn ja akustiikan laitos, Aalto-yliopisto

antti.saloranta@utu.fi, henna.tamminen@utu.fi, paavo.alku@aalto.fi, maija.peltola@utu.fi

Tutkimuksessa oli tavoitteena selvittää, voidaanko artikulatorisilla ääntämisohjeilla tehostaa kuuntele ja toista -harjoitusta. Tällainen tehostettu harjoittelu voisi hyödyntää puheen kontrollimekanismeja tavalla, joka on verrattavissa lasten kielenomaksumiseen. Harjoittelu toteutettiin kahden päivän aikana. Ohjeiden tarkoituksena oli auttaa koehenkilöitä tuottamaan vieras äänne tutun suomalaisen äänteen ääntymäpaikkaa lähtökohtana käyttäen. Oppimiskontrastiksi valittiin ruotsin /y/-/w/ joka ei esiinny suomessa. Koehenkilöinä oli 9 äidinkieleltään suomalaista 18–30-vuotiasta aikuista, jotka eivät opiskele kieliä eivätkä ole asuneet muissa Pohjoismaissa. Ääntämisohjeet annettiin samanlaisina jokaiselle koehenkilölle ennen kutakin harjoitussessiota. Tulokset eivät olleet yksiselitteisiä. Ryhmätasolla formanttiarvoissa ei ollut merkittäviä muutoksia kummassakaan vokaalissa, mutta yksittäisillä koehenkilöillä /w/:n F2:ssa tapahtui kymmenien prosenttien muutoksia. Keskihajontojen tilastollinen ero muuttui kokeen aikana merkitseväksi, mikä selittyi erityisesti /w/ F2:n hajonnan koehenkilöiden välisellä vaihtelulla. Muutokset voivat selittyä muutoksella puheen tuoton sisäisessä mallissa. Tutkimus tarjosi selviä viitteitä käytetyn menetelmän mahdollisuuksista. Tulevissa tutkimuksissa tulisi selvittää erityisesti muutosten näkyvyyttä neuraalisella tasolla.

Asiasanat: kontrollimekanismit, vieraan kielen oppiminen, puheen tuottaminen

1. Johdanto

Puheen ymmärtäminen ja tuottaminen ovat tiiviisti toisiinsa linkittyneitä ilmiöitä. On esitetty (esim. Guenther 2003), että mekanismi jolla lapset oppivat puhumaan on kaksivaiheinen: ensin lapset saavat äidinkielistä puhealtistusta ympäristöstään ja oppivat siten äidinkieltensä äännekategoriat, jonka jälkeen niitä käytetään havaittavina kohteina joihin puheentuotossa pyritään. Lapsi siis vertaa omia ääntämyksiään oppimiinsa äännekategorioihin, ja muodostaa vähitellen käsityksen siitä, millaisia motorisia komentoja kunkin äänteen tuottamiseen tarvitaan. Tämä prosessi hyödyntää erilaisia tunnettuja neuraalisia palautemekanismeja, ja toimii täysin automaattisesti. Aikuisilla tämän mekanismi jää taka-alalle ja sen tehtävänä on lähinnä äännejärjestelmän ylläpito.

Tässä artikkelissa esitellään tutkimus, jossa selvitettiin foneettisten ääntämisohjeiden vaikutusta vieraan kielten äänneiden oppimiseen. Koeasetelma perustuu sekä aiempiin tutkimustuloksiin että

tuoreisiin kielten omaksumiseen ja oppimiseen liittyviin malleihin. Koeasetelmalla on kaksi tavoitetta: ensinnäkin kiinnittää oppijan huomio uuden äännekontrastin kannalta relevantteihin akustisiin piirteisiin, ja toiseksi tarjota valmis lähtökohta uuden puheentuotossa havaittavan kohteen muodostamiseksi.

2. Vieraan kielen äänneiden oppimisen mallit

Vieraan kielen äänneiden oppiminen on tärkeimpiä kielitieteen tutkimusaiheita. Ihmisten erittäin vaihtelevat oppimistulokset ovat yleisesti tiedossa, mutta niiden syyt ovat huonommin tunnettuja, ja aihetta on lähestytty kirjallisuudessa usealta kannalta. Tässä tutkielmassa keskitytään kuitenkin pääasiallisesti kielten äännejärjestelmien vertailuun.

2.2. Kategorinen kuuleminen

1950-luvulla kehitetty teoria, kategorinen kuuleminen, on noussut erääksi oppimistutkimuksen kulmakivistä. Kategorisen kuulemisen mukaan ihmiset eivät käsittele kuulemaansa jatkuvaa äänivirtaa sellaisenaan, vaan jakavat sen selkeisiin kategorioihin äidinkieltensä foneemien perusteella. Käytännössä kaikki nykyiset oppimismallit perustuvat kategoriseen havaitsemiseen.

2.3. Speech Learning Model ja Perceptual Assimilation Model

Speech Learning Model (SLM) on James Flegen kehittämä yleisesti käytössä oleva vieraan kielen oppimisen malli. Se kehitettiin 1980-luvulla vastineeksi kontrastiiviselle analyysille, kategoriselle kuulemiselle sekä kriittisen iän hypoteesille, jonka mukaan oppiminen muuttuu lähes mahdottomaksi tietyn iän jälkeen (Flege 2005). Kuten vanhat teoriat, SLM perustuu kielten äännejärjestelmien vertailulle. Se jakaa vieraan kielen äänneet kolmeen kategoriaan, tässä helpoimmasta opittavasta vaikeimpaan: identtiset (Identical), uudet (New) sekä samankaltaiset (Similar) (Flege 1988). Vieräs äänne voi samankaltaisuudestaan riippuen joko assimiloitua johonkin äidinkielen foneemikategoriaan tai muodostaa kokonaan uuden.

Perceptual Assimilation Model (esim. Best 1994a) on perusominaisuuksiltaan SLM:n kaltainen, mutta monipuolisempi, sillä se käsittelee assimilaatiota kahdella tasolla. Yksittäiset vieraan kielen foneemit joko: assimiloituvat natiivikategoriaan hyvin tai huonosti; eivät assimiloitu, mutta mielletään silti puheäänteiksi; tai niitä ei pidetä lainkaan puheena. PAM:n pääteema on kontrastien assimiloituminen, eli miten vieraan kielen äänneparit (tai ryhmät) sijoittuvat äidinkieleen nähden. Tähän on neljä vaihtoehtoa: pari assimiloituu vastaavaan kahteen kategoriaan äidinkielessä; kumpikin vieräs äänne assimiloituu yhteen äidinkielen kategoriaan yhtä hyvin tai huonosti;

kumpikin vieras äänne assimiloituu yhteen äidinkielen kategoriaan epätasaisen hyvin tai huonosti; kumpikaan äänneistä ei assimiloitu mihinkään kategoriaan (Best 1994). Vaikeinta oppimistilannetta näistä edustavat yhteen kategoriaan assimiloitumiset.

Kummallekin mallille on yhteistä se, että assimilaatio korreloi erottelukyvyn kautta, ja että vaikeimpia opittavia ovat ne äänneet, jotka muistuttavat äidinkielen äänneitä mutta kuitenkin eroavat niistä systemaattisesti. Tilannetta kuvaavat SLM:ssä samankaltaiset äänneet, ja PAM:ssä yhteen kategoriaan assimiloitumiset. Kumpikaan malleista ei tarjoa keinoja arvioida yksittäisten vieraiden äänneiden samankaltaisuutta äidinkieleen. Niiden perusteella tehdyt ennustukset ovat kuitenkin usein yhteneväisiä, ja yhdessä käytettynä niillä voidaan melko hyvin arvioida odotettavissa olevia oppimisongelmia koetilanteissa.

3. Puheen tuoton kontrollimekanismit

Puheen tuotto on monimutkainen prosessi, joka vaatii saumatonta yhteistyötä jopa kymmeniltä pään ja kaulan alueen osilta. Tässä luvussa käydään läpi joitain puheen tuottoon liittyviä malleja, sekä tapoja, joilla puhetta kontrolloidaan neuraalisesti.

3.1. Palautemekanismit

Palautemekanismeilla tarkoitetaan tässä niitä tapoja, joilla keskushermosto saa tietoa kulloinkin tuotettavista puhesegmenteistä. Fysiologisia palautemekanismeja on kolme: auditorinen, taktiilinen sekä proprioseptinen. Auditorinen palaute sisältää ääniaaltojen kautta saadun palautteen, eli normaalin kuulon sekä äänen johtumisen kallon luita pitkin. Taktiilinen palaute aktivoituu mekaanisesta kosketuksesta, esimerkiksi kielen koskiessa kitalakeen. Proprioseptinen palaute, joka on nopeinta näistä kolmesta, tarkoittaa aistimuksia artikulaatioelimistön asennoista. Palautemekanismin keskinäinen tärkeys riippuu kulloinkin tuotettavasta äänneestä (Hardcastle 1976).

3.2. Kvantaaliteoria ja H&H-hypoteesi

Kvantaaliteorian (Stevens 1989) mukaan monet puheäänneet ovat luonteeltaan kvantaalisia, eli suhteellisen pienet muutokset artikulaatiossa muuttavat niiden akustiikkaa merkittävästi. Foneemit eivät siis muutu lineaarisesti, vaan hyppäyksittäin, ja hyppäysten välissä on vakaita alueita. Tämä tarkoittaa sitä, että artikulaation ei tarvitse olla täysin sama tietyn foneemin jokaisella tuottokerralla. Ominaisuus on korostunut tietyillä äänneillä, mikä on saattanut johtaa helpommin tuotettavien äänneiden korostumiseen kielten äänneinventaariorissa.

H&H-hypoteesi (Lindblom 1990) pyrkii selittämään vaihteluita yksittäisten puhujien välillä. Hypoteesin mukaan puhe perustuu selkeyden ja tuoton taloudellisuuden tasapainoon, ja näitä kahta varioidaan tilanteen mukaan. Puheen tuottoon tarvittavat liikkeet suunnitellaan etukäteen, ja ne pyritään toteuttamaan mahdollisimman tehokkaasti, eli tuottamalla haluttu äänne mahdollisimman vähäisillä artikulaatioelimistön liikkeillä.

3.3. DIVA-malli

DIVA (Direction Into Velocities of Articulators) (Guenther 2003; Perkell 2012) on malli, joka kuvaa puheen kontrolloinnista vastaavaa neuraalista verkostoa ja sen osien vuorovaikutuksia. Mallin mukaan aivoihin muodostuu puheen tuotolle eräänlainen sisäinen malli, joka käyttää feedforward-komentoja. Malliin on tallentunut motoriset komennot sekä perseptuaaliset kohteet kaikille äidinkielen äänneille, ja puhe muodostuu kun niitä yhdistetään sarjoiksi. Malli muodostuu varhaislapsuudessa, alkaen foneemikategorioista jotka muodostuvat kun vauva kuulee puhetta. Jokerteluiässä näitä kategorioita käytetään malleina omien ääntämysten luokitteluun auditorisen palautteen avulla. Tuotettu äänne kategorisoidaan joksikin äidinkielen foneemiksi, ja sen tuottamisessa käytetyt artikulaatiot tallentuvat kyseisen foneemin kohdalle. Toiston myötä aivoihin lopulta muodostuu foneemispesifiset motoriset komentosarjat kullekin äänneelle ja malli on valmis (Guenther 2003). Tämän jälkeen puhetta ohjataan feedforward-komennoilla, ja palautemekanismit omaksuvat ylläpitävän ja korjaavan roolin (Perkell 2012).

DIVA-malli tarjoaa selkeän ja tehokkaan mallin puheen oppimiselle ja kontrolloinnille. Se yhdistää kvantaaliteorian ja H&H-hypoteesin perusteemoja tarjoamalla keinon jolla puhe voidaan tuottaa taloudellisesti ja vakaita alueita hyödyntäen. Mallia ei ole toistaiseksi sovellettu vieraan kielen oppimiseen, mutta ottaen huomioon palautemekanismin sisäistä mallia ylläpitävän vaikutuksen niiden käyttöä mallin muokkaamiseen oppimistilanteessa voidaan pitää uskottavana vaihtoehtona. Tässä tutkimuksessa käytetty koeasetelma pohjaakin teoreettiselta taustaltaan juuri näiden mekanismien hyväksikäyttöön.

4. Aiempia tutkimuksia harjoittelun vaikutuksesta

Vieraan kielen oppiminen on eräs tutkituimpia kielitieteen osa-alueita. Aihetta on lähestytty tutkimalla sekä jo saavutettuja oppimistuloksia, että erilaisten harjoitteiden vaikutusta. Tässä luvussa esitellään joitakin tämän tutkielman kannalta oleellisia harjoitetutkimusten tuloksia.

4.1. Havaitsemisharjoitteet

Havaitseminen ja tuottaminen ovat kiinteästi yhteydessä toisiinsa (esim. Kuhl 2000; Guenther 2003), joten havaitsemisen harjoittamista voidaan pitää hyvänä menetelmänä vieraan kielen äänteiden opettamisessa. Hyviä tuloksia on saavutettu pakotettua valintaa sekä samanlainen/erilainen luokittelua käyttävillä havaitsemisharjoituksilla (Flege 1995) sekä liioittelemalla erottelun kannalta relevantteja akustisia piirteitä keinotekoisesti (McCandliss et al. 2002). Eräs erityisen onnistunut tutkimussarja käytti suuren vaihtuvuuden foneettiseksi harjoitteluksi kutsuttua menetelmää, jossa japaninkieliset englanninoppijat tekivät minipari-identifikaatioharjoitteita natiivipuhujien ääntämistä ärsykkeistä (esim. Logan, Lively & Pisoni 1991; Lively et al. 1994). Menetelmän etuna pidettiin erityisesti suurta määrää erilaisia saman kategorian edustajia joille oppijat altistuivat harjoittelun aikana. Mikä mielenkiintoisinta, tämä harjoittelu paransi myös oppijoiden tuottokykyä (Bradlow et al. 1997; Bradlow et al. 1999). Tutkimukset siis osoittivat, että havaitseminen edeltää tuottamista ja vaikuttaa siihen myös aikuisten vieraan kielen opiskelussa, ei ainoastaan lapsuuden puheenomaksumisvaiheessa.

4.2. Tuottoharjoitteet ja huomion ohjaaminen

Vaikka suurin osa vieraan kielen opetustutkimuksista koskeekin havaitsemisharjoitteita, myös tuottotutkimuksia on tehty. Catford & Pisoni (1970) osoittivat, että artikulaatioharjoituksella, jossa oppijoille ohjeistettiin oikeat artikulaatioelimistön asennot kunkin vieraan äänteen tuottamiseksi, saatiin huomattavasti parempia oppimistuloksia tuotossa kuin pelkällä kuuntele ja toista – harjoituksella. Harjoituksessa kutakin vierasta äännettä lähestyttiin käyttäen lähtökohtana jotakin samankaltaista äidinkielen äännettä. Vaikka tämä harjoite tuotti hyviä tuloksia, uudemmat tuottoon keskittyvät tutkimukset ovat lähinnä esitelleet erilaisia koneellisia ääntämisharjoituksia sen sijaan että tuottoa käytettäisiin tieteellisen tutkimuksen välineenä.

Mielenkiintoisena harjoitusmenetelmänä on mainittava myös koehenkilöiden huomion ohjaaminen harjoitusvaiheessa. Menetelmää on kokeiltu kahdessa tutkimuksessa (Guion & Pederson 2007; Pederson & Guion-Anderson 2010). Harjoitteessa koehenkilöiden huomio kiinnitetään ohjeilla oppimisen kannalta olennaisiin piirteisiin ärsykeissä. Kontrollina oli sanojen merkitykseen keskittyvä ryhmä. Tulosten mukaan tällä menetelmällä saavutetaan erityisen hyviä tuloksia äänteissä, jotka ovat koehenkilöiden äidinkielen takia vaikeimpia.

Edellä esitetyt tulokset osoittavat selkeästi, että harjoittelu jossa yhdistetään havaitsemis- ja tuottoharjoitus voi tuottaa erityisen hyviä tuloksia, varsinkin jos siihen lisätään huomionkiinnityskeino jolla koehenkilöiden erottelukykyä voidaan keinotekoisesti lisätä. Näitä harvemmin käytettyjä menetelmiä käytetään tässä tutkielmassa esitellyssä tutkimuksessa.

5. Tutkimus

Tutkimuksen tarkoituksena on selvittää, voidaanko foneettisilla ääntämisohjeilla tehostaa tavallista kuuntele ja toista -harjoitusta. Teoreettisena perustana on DIVA-malli, jonka mukaan foneemikategoriat opitaan ensin puhetta kuuntelemalla, ja näitä kategorioita käytetään tuoton opiskelussa malleina, joiden perusteella muodostetaan foneemikohtainen motorinen kohde artikulaatioelimistölle. Tämä mekanismi pyritään aktivoimaan uudelleen ohjeilla, joiden tarkoitus on sekä tarjota valmis lähtökohta uuden motorisen kohteen muodostumiselle, että kohdistaa koehenkilön huomio oppimisen kannalta relevantteihin piirteisiin opittavassa äänneessä. Koeasetelma pyrkii siis nopeuttamaan DIVA-mallissa esitettyä kategorianmuodostusta tarjoamalla samanaikaisesti mallin sekä äännekategorian että motorisen kohteen muodostumiselle.

5.1. Metodit

Koehenkilöinä oli yhdeksän (viisi naista) suomalaista nuorta aikuista, iältään 21–30 (keskiarvo 24,3). Kaikki puhuivat äidinkielenään suomea, olivat normaalikuuloisia, eivät opiskelleet kieliä yliopistossa eivätkä olleet asuneet muissa Pohjoismaissa kuin Suomessa. Kielitaidoltaan koehenkilöt vastasivat keskivertosuomalaisia: kaikki ilmoittivat osaavansa englantia ja ruotsia, ja suurin osa lisäksi jotain muuta kieltä.

Ärsykkeiksi valittiin pseudosanat /tʌ:ti/ and /ty:ti/, ja ne valmistettiin semisynteettisesti Aalto-yliopiston Signaalinkäsittelyn ja akustiikan laitoksella (lisätietoa valmistustavasta Alku, Tiitinen & Näätänen 1999). Vain ruotsissa esiintyvä vokaali /ʉ/ edustaa suomalaisen kannalta erittäin vaikeaa oppimistilannetta, sillä se sijoittuu akustisesti noin suomen /y/ ja /u/ -kategorioiden puoliväliin, ja voidaan puhujasta riippuen kategorisoida kummaksi tahansa niistä. Tämä on vaikea oppimistilanne sekä SLM:n että PAM:in kautta tarkasteltuna.

Koe suoritettiin Turun yliopiston fonetiikan oppiaineen LAB-laboratoriossa käyttäen Sanako Oy:n Lab 100-ohjelmistoa. Kaksipäiväinen koe sisälsi kaksi nauhoitusta ja kaksi harjoitusta kumpanakin päivänä. Ensimmäisenä päivänä järjestyksenä oli nauhoitus-harjoitus-nauhoitus-harjoitus, ja toisena harjoitus-nauhoitus-harjoitus-nauhoitus. Sekä harjoituksissa että nauhoituksissa koehenkilöt toistivat nauhalta vuoron perään kuulemiaan pseudosanoja (ISI 3000 ms). Harjoituksessa toistoja tuli 30/sana, nauhoituksessa 10/sana. Kunkin harjoituskerran aluksi koehenkilöille luettiin ääntämisohjeita, jotka on lueteltu seuraavassa:

1. Kuulet vuorotellen kahta sanaa, toisessa on suomalainen ja toisessa ruotsalainen vokaali.
2. Ruotsin vokaali on ikään kuin suomen /y/:n ja /u/:n välimuoto.

3. Koeta pitää suu muuten samassa asennossa kuin /y/:tä sanoessa, mutta siirrä kieltä suussa hieman taaksepäin.
4. Huulten pyöreyydessä on pieniä eroja. /y/:ssä huulet ovat tiukemmin supussa kuin /u/:ssa.

Nauhoituksista eristettiin formanttiarvot Praat-ohjelmalla (Boersma & Weenink 2013), jonka jälkeen tuloksista tehtiin tilastolliset analyysit toistettujen mittausten varianssianalyysillä (ANOVA).

5.2. Tulokset

Keskimääräisten formanttiarvojen tilastollinen analyysi ei paljastanut oppimistuloksia. Sanat olivat formanttiarvoiltaan johdonmukaisesti erilaisia koko tutkimuksen ajan, eikä mittauskertojen välillä havaittu minkäänlaisia tilastollisesti merkitseviä vaikutuksia. Selkein formanttiarvoissa tapahtunut ryhmätason muutos oli /t̥a:ti/:n F2-formantin keskihajonnan voimakas laskeminen kokeen aikana. Tämä viittaa ryhmän tuotosten yhdenmukaistumiseen. Yksilökohtaisten keskihajontojen tilastollinen analyysi sen sijaan paljasti osoituksen muutoksesta: kaikkien sessioiden analyysissä (sessio(4) × sana(2) × formantti(2)) sanan päävaikutus sekä sana × formantti interaktio olivat tilastollisesti merkitseviä, eli sanojen hajonnat olivat johdonmukaisesti erilaisia, mutta yksittäisten sessioiden analyysi paljasti että kumpikaan näistä vaikutuksista ei ollut merkitsevää vielä ensimmäisessä sessiossa. Kahden ensimmäisen session yhteisanalyysissä kumpikin oli jo tilastollisesti merkitsevää. Tämä viittaa siihen, että harjoitus ja artikulaatio-ohjeet vaikuttivat koehenkilöiden tuottoon heti alkamisestaan lähtien ja muuttivat keskihajontaa, eli tuoton yhdenmukaisuutta, merkittäväällä tavalla.

Yksittäisten koehenkilöiden datan tarkastelu osoitti että koehenkilöiden lähtötasossa oli suurta hajontaa. Osa osasi selvästi tuottaa harjoiteltavan vokaalin jo kokeen alussa ennen harjoittelua, mikä väärästi ryhmän tulosta koska oppimista ei näiden henkilöiden kohdalla voinutkaan tapahtua. Osalla koehenkilöistä uuden vokaalin tuotto sen sijaan muuttui kokeen aikana selvästi, mikä oli havaittavissa erityisesti keskihajontojen muutoksissa. Hajonnoissa esiintyi sekä selvää laskua, joka viittaa tuoton yhdenmukaistumiseen ja motorisen kohteen vahvistumiseen, että voimakasta nousua, joka taas viittaa alkavaan mutta kesken jääneeseen tuoton muuttumiseen. Kumpikin näistä muutoksista sopii hypoteesiin, ja viittaa siihen että eri koehenkilöt tarvitsevat lähtötilanteestaan riippuen eri määrän aikaa oppiakseen uuden kategorian ja motorisen kohteen. Kaikilla kokeessa käytettyä kontrastia osaamatomilla koehenkilöillä selvää oppimismuutosta alkoi kuitenkin tapahtua.

6. Yhteenveto

Tässä tutkielmassa esitelty tutkimus tarjosi viitteitä siitä, että kuuntele ja toista – harjoitetta voidaan tehostaa foneettisilla ääntämisohjeilla. Oppimistuloksia havaittiin niin tilastollisessa analyysissä kuin yksittäisten koehenkilöiden datan tarkastelussa. Vaikuttaa siis siltä, että DIVA-mallin mukainen kielenomaksumismenetelmä on ainakin osittain mahdollista aktivoida aikuisiässä uudelleen oikeanlaisilla harjoitteilla, jotka sekä ohjaavat huomiota ja parantavat mahdollisesti erottelukykyä, että tarjoavat valmiin lähtökohdan uuden motorisen kohteen muodostumiselle. Aikuisilla tämä vaikutus voidaan myös saada esiin melko nopeasti verrattuna pieniin lapsiin joita DIVA-malli alun perin käsittelee. Yhteenvetona voidaan sanoa, että tutkimus tarjoaa alustavan osoituksen siitä, että siinä kuvatulla harjoittelulla on mahdollista opettaa aikuisille vieraan kielen äänteitä. Jatkossa erityistä huolta tulisi kiinnittää opeteltavan kontrastin huolelliseen valintaan ja koehenkilöiden kielitaustan tarkkaan kontrollointiin tilastollista analyysia vaikeuttavien tekijöiden vähentämiseksi.

Lähteet

Alku, P., Tiitinen, H. & Näätänen, R. 1999. A method for generating natural sounding speech stimuli for cognitive brain research. *Clinical Neurophysiology* 110 (8): 1329-1333.

Best, C.T. 1994a Learning to perceive the sound pattern of English. In Rovee-Collier, C. Lip-sitt, L.P. (Eds.). 1994. *Advances in Infancy Research* 217-304. Norwood, NJ: ABLIX Publishing Corporation.

Boersma, P. & Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Version 5.2.1, retrieved 12 November 2011 from <http://www.praat.org/>

Bradlow, A., Akahane-Yamada, R., Pisoni, D.B. & Tohkura, Y. 1999. Training Japanese listeners to identify English /r/ and /l/: Long-term retention in perception and production. *Perception & Psychophysics* 61(5): 977-985.

Bradlow, A.R., Pisoni, D.B., Akahane-Yamada, R. & Tohkura, Y. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101(4): 2299-2310.

Catford, J.C. & Pisoni, D.B. 1970. Auditory vs. articulatory training in exotic sounds. *The Modern Language Journal* 54(7): 477-481.

Flege, J.E. 1995. Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics* 16: 425-442.

Guenther, F. 2003. Neural control of speech movements. Teoksessa Schiller, N.O. & Meyer, A.S. (Eds.). 2003. *Phonetics and phonology in language comprehension and production* 209-239. New York: Mouton de Gruyter.

- Guion, S. & Pederson, E. 2007. Investigating the Role of Attention in Phonetic Learning. In Bohn, O-S. & Munro, M. (Eds.) 2007. *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* 55-77. Amsterdam: John Benjamin.
- Hardcastle, W.J. 1976. *Physiology of speech production*. London: Academic Press Inc.
- Kuhl, P.K. 2000. A new view of language acquisition. *Proceedings of the National Academy of Sciences* 97(22): 11850-11857.
- Lively, S.E., Pisoni, D.B., Yamada, R.A., Tohkura, Y., & Yamada, T. 1994. Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America* 96(4): 2076-2087.
- Logan, J.S., Lively, S.E. & Pisoni, D.B. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America* 89(2): 874-886.
- McCandliss, B.D., Fiez, J.A., Protopapas, A., Conway, M. & McLelland, J.L. 2002. Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience* 2(2): 89-108.
- Pederson, E. & Guion-Anderson, S. 2010. Orienting attention during phonetic training facilitates learning. *Journal of the Acoustical Society of America* 127(2): 54-59.
- Perkell, J. 2012. Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics* 25: 382-407.
- Stevens, K.N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17: 3-46.

Tremulantin kvantiteetti suomessa

Joonas Vakkilainen, Michael O'Dell

Tampereen yliopisto

joonas.vakkilainen@uta.fi, michael.odell@uta.fi

Tremulantit koostuvat täryistä, ja näin ollen niiden kvantiteetin määräytymisen voisi olettaa olevan erilaista muihin konsonantteihin verrattuna. Suomen /r/ voi esiintyä sekä yksi- että monitäryisenä. Tutkimuksemme selvittää, millaisia tärymääriä /r/ sisältää eri konteksteissa, onko yksitäryinen variantti oma erillinen äänteensä ja poikkeako /r/ kestoltaan muista konsonanteista. Tutkittavat /r:/t saatiin kehyslauseessa olleista koeksanoista, joissa varioitiin vokaaliympäristöä (/a/, /i/) ja ensimmäisen vokaalin kvantiteettia. Tulosten perusteella lyhyt /r/ on yleisimmin yksitäryinen ja pitkä /rr/ monitäryinen. Tärymäärät ovat kuitenkin osittain päällekkäisiä eri kestoissa. Näin ollen yksitäryinen variantti näyttäisi olevan osa tremulantin kvantiteetin jatkumoa eikä erillinen äänne. Muista konsonanteista poiketen /r/ on lyhyempi /i:/n kuin /a:/n jäljessä. Muiden konsonanttien tavoin /r/ on pidempi lyhyen kuin pitkän vokaalin jäljessä.

Asiasanat: suomen kieli, tremulantti, kvantiteetti, täryjen määrä

1. Taustaa

Tremulantit ovat aerodynaamisista syistä muita konsonantteja vaikeampia tuottaa (Iivonen 1993, Ladefoged & Maddieson 1996). Tremulantti koostuu täryistä, jotka synnytetään pulmonisella ilmavirralla. Jos tremulantti sisältää vain yhden täryn, puhutaan puolitremlantista.

Suomen /r:/n katsotaan olevan apikaalinen tremulantti, jolla on yksi- ja monitäryiset variantit. Perinteisen käsityksen mukaan /r/ toteutuu yleensä monitäryisenä; puolitremlanttia [r] (SUT δ) pidetään ainoastaan /t:/n heikon asteen varianttina. Näin ollen moni- ja yksitäryistä varianttia on pidetty eri äänteinä. Täryjen määrä näyttäisi kuitenkin liittyvän /r:/n keston eikä laatueroon. Yksitäryinen variantti olisi tällöin osa tremulantin kvantiteetin jatkumoa (Mustanoja & O'Dell 2007). Puolitremlantti onkin kuin tremulantin yksi täry ja monitäryinen tremulantti kuin sarja puolitremlantteja (Lindau 1985).

Suomen konsonantit ovat yleensä pitkän vokaalin jäljessä lyhyempiä kuin lyhyen vokaalin jäljessä (Lehtonen 1970, esim. s. 122). Tätä ei kuitenkaan ole tutkittu /r:/n tapauksessa. Toinen konsonanttien keston yleisesti vaikuttava tekijä on ympäröivien vokaalien laatu. Suomessa, kuten maailman kielissä yleisestikin, suppea vokaali kuten /i/ on samassa kontekstissa keskimäärin

lyhyempi kuin väljä vokaali kuten /a/. Viereiset konsonantit kompensoivat tätä eroa: konsonantti on keskimäärin pitempi *i*:n vieressä kuin /a/:n vieressä (ks. esim. Lehtonen 1970, s. 86). Tremulantti ei kuitenkaan välttämättä käyttäydy tässä samalla tavalla kuin muut konsonantit, sillä /i/:n yhteydessä täryjä on vaikeampi tuottaa kuin /a/:n yhteydessä (suomessa vrt. esim. Sovijärvi 1963, s. 54; myös Mustanoja & O'Dell 2004 huomasivat mahdollisen yhteyden yksitäryisyyden ja palataalistumisen välillä Tampereen puhekielessä). Monet tutkijat ovat huomanneet, että [i]-konteksti tai palataalistuminen usein lyhentää tremulanteja tai vähentää täryjä (Ladefoged & Maddieson 1996, Barry 1997, Kavitskaya 1997, Recasens & Pallarès 1999, Solé 2002, Iskarous & Kavitskaya 2010). Esim. venäjässä palataalistumaton /r/ on yleensä monitäryinen, mutta palataalistunut /r^j/ jää usein yksitäryiseksi (Ladefoged & Maddieson 1996, s. 221). Solén tutkimuksessa (Solé 2002) tremulantit kontekstissa [i__i] olivat lyhyempiä ja sisälsivät vähemmän täryjä kuin kontekstissa [a__a]. Tämän arvellaan johtuvan siitä, että apikaalinen tremulantti edellyttää aerodynaamisista syistä laajaa tilaa kielen kärjen takana (kielen selkä alhaalla ja takana), mikä on ristiriidassa [i]:n artikulaation kanssa (Solé 2002, Iskarous & Kavitskaya 2010).

2. Tutkimuskysymykset

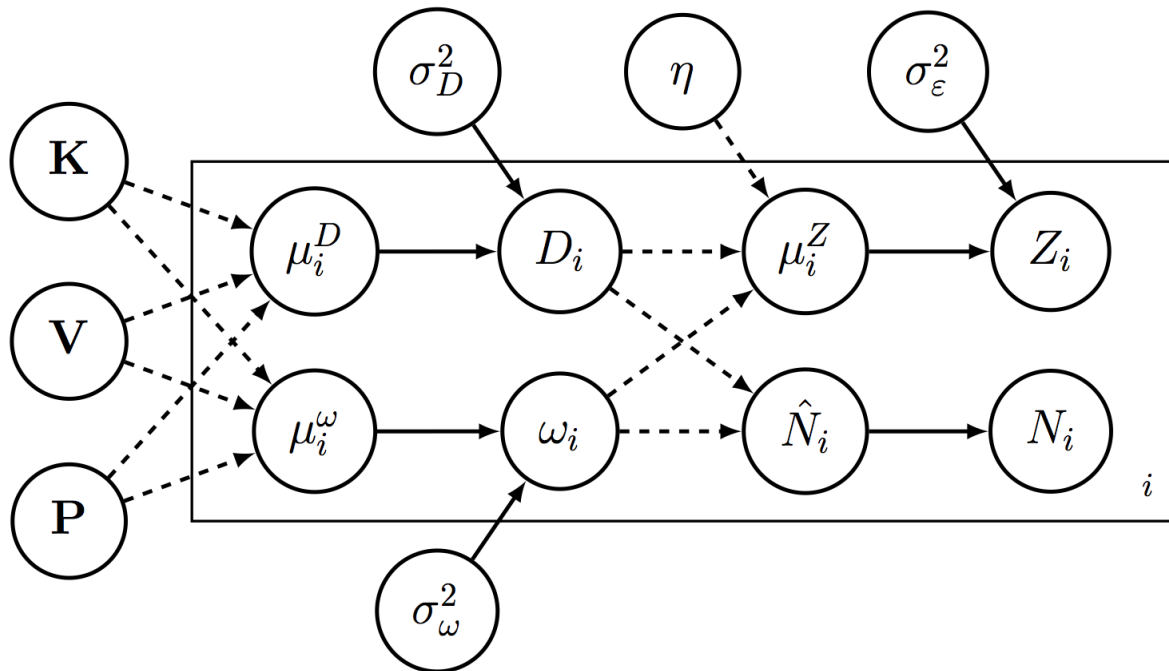
Tutkimuksen tarkoitus on selvittää suomen tremulantin kvantiteetin määräytymistä. Tarkoituksena on tarkastella, ovatko yksi- ja monitäryiset /r/:n variantit omia erillisiä äänteitään, vai ovatko ne saman foneemin kestojatkomon osia. Näin ollen tutkimuksen kohteena on myös tremulantin täryjen määrä eri konteksteissa. Lisäksi on tarkoitus selvittää, poikkeako /r/ muista konsonanteista kvantiteetin tai vokaaliympäristön suhteen.

3. Menetelmät

Tutkimuksen aineiston tuottivat neljä suomenkielistä koehenkilöä (kolme naista ja yksi mies). Koehenkilöt lukivat kehyslauseita, jotka sisälsivät varioitavan koesanan. Koesanassa varioitiin ensimmäisen tavun vokaalin pituutta, /r/:n pituutta ja kummankin tavun vokaaleja, jotka olivat /a/ ja /i/. Koesanoina toimivat *kiri, kira, kari, kara, kiiri, kiira, kaari, kaara, kirri, kirra, karri, karra, kiirri, kiirra, kaarri* ja *kaarra*, jotka oli sijoitettu kehyslauseeseen *Sano sana _____ taas kerran*. Jokaisella koesanalla oli kymmenen toistoa, joten tutkittavien lauseiden määrä yhdellä puhujalla oli 160. Lisäksi alkuun ja loppuun oli sijoitettu intonaation vaikutuksen minimoimiseksi ylimääräinen lause, jota ei analysoitu.

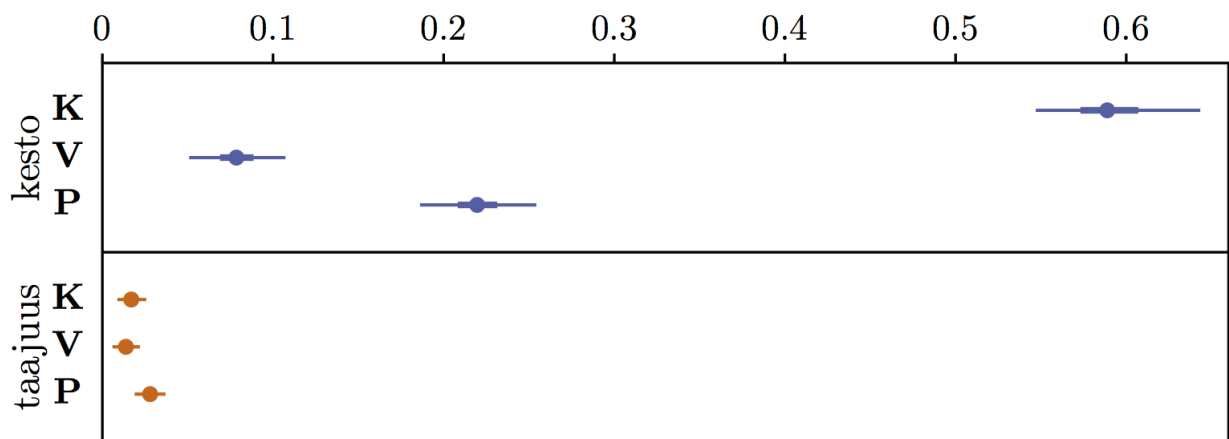
Koesanoista mitattiin /r/:n kesto ja täryjen määrä. Mittauksiin käytettiin *Praat*-skriptiä, ja mittauksia tarkistettiin käsin. 640 koesanasta 611:tä voitiin käyttää tutkimuksessa. Pois jätettiin tapaukset,

jotka eivät sisältäneet mitattavissa olevia täryjä. Tämä johtui mm. r :n frikatiivimaisesta toteutumisesta tai muusta epäselvyydestä. Tutkimus käsittelee vain täryjä sisältäviä tapauksia.



Kuva 1. Tilastollinen malli. Yhtenäiset nuolet edustavat stokastisia suhteita, katkoviivalliset laskennallisia suhteita. Symbolit selitetään tekstissä.

Tilastollisessa analyysissä käytettiin bayesilaista inferenssiä hierarkkisessa (ANOVA-tyyppisessä) mallissa (ks. esim. Gelman & Hill 2007). Mallia havainnollistetaan kuvassa 1. Faktoreina toimivat kvantiteettihahmo (kuvassa **K**, $n = 4$ kategorioita: CVCV, CVVCV, CVCCV, CVVCCV), vokaaliympäristö (**V**, $n = 4$ kategorioita: $i-i$, $i-a$, $a-i$, $a-a$) sekä puhuja (**P**, $n = 4$). Tremulantin jaksottaisuus vaikuttaa voimakkaasti empiirisesti mitattuun keston (kuvassa Z_i), koska aktivoinnin keston tasaisesti kasvaessa viimeinen täry jää usein vajaaksi ja mitattu kesto ryhmittyy kokonaistärymäärien mukaan. Tämä otettiin huomioon mallissa tärytaajuuden tahdissa vaihtelevalla porrastusfunktiolla, joka antaa odotettua mitattua kestoa (kuvassa μ_i^Z) tremulantin aktivoinnin keston (D_i) funktiona. Faktorit siis vaikuttavat mallissa suoraan odotettuun aktivoinnin keston (kuvassa μ_i^D), kun taas mitattu kesto riippuu sekä aktivoinnin kestosta että tärytaajuudesta (kuvassa ω_i ; myös porrastusfunktiolla on mallissa portaiden jyrkkyyttä säättävä parametri η , jonka arvo määräytyy bayesilaisessa inferenssissä datan mukaan). Samat faktorit (**K**, **V**, **P**) vaikuttavat mallissa myös odotettuun tärytaajuuteen (kuvassa μ_i^ω), ja empiirisesti havaittu tavumäärä (N_i) riippuu niin ikään aktivoinnin kestosta (D_i) ja tärytaajuudesta (ω_i). Kiinnostavimpia parametreja mallissa tutkimuksen kannalta ovat faktorien vaikutukset keston ja tärytaajuuteen.



Kuva 2. Vaikutusten (geometriset) keskihajonnat

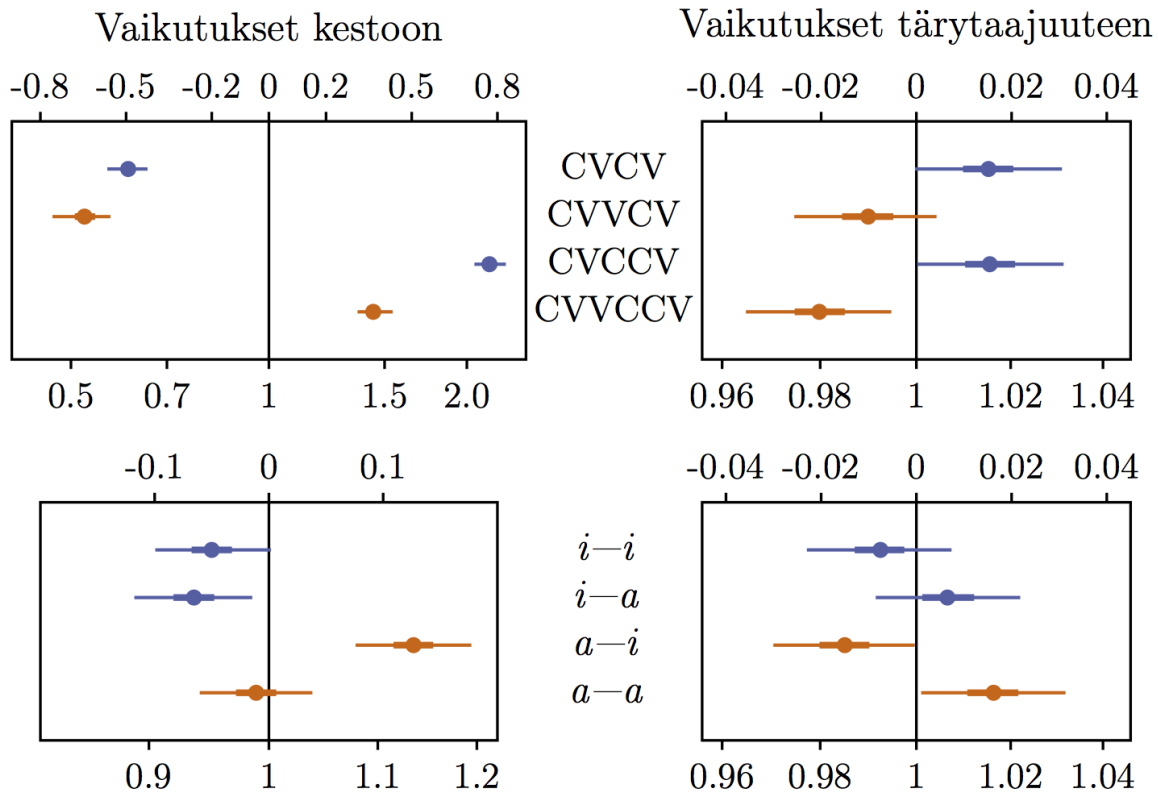
4. Tulokset

4.1. Kesto ja tärytaajuus

Bayesilaisessa mallissa arvioituja vaikutusten arvoja (posteriorijakaumat) esitetään kuvissa 2 ja 3. Näissä kuvissa jokainen jakauma on pelkistetty yhteen janaan: janan keskellä oleva piste edustaa jakauman mediaania, paksumpi viiva 50% ja ohuempi viiva 95% luottamusväliä.

Kuvassa 2 näkyvät keskihajonnat auttavat arvioimaan vaikutuksien suuruutta ja tärkeyttä — suuri keskihajonta kertoo, että faktorin kategoriat ovat hyvin erilaisia. Kaikki janat kuvassa 2 ovat selvästi erillään nolasta, joten kaikki faktorit (**K**, **V**, **P**) ovat merkitsevästi vaikuttaneet sekä keston että tärytaajuuteen. Vaikutukset ovat kuitenkin selvästi hyvin eri suuruisia.

Selvästi eniten tremulantin keston vaikuttaa sanan kvantiteettihahmo. Vähiten vaikuttaa vokaaliympäristö. Vokaaliympäristön ja puhujan vaikutukset ovat lähellä toisiaan verrattuna kvantiteettihahmon vaikutukseen. Täryjen taajuuteen muuttujien vaikutukset ovat verrattain pieniä.



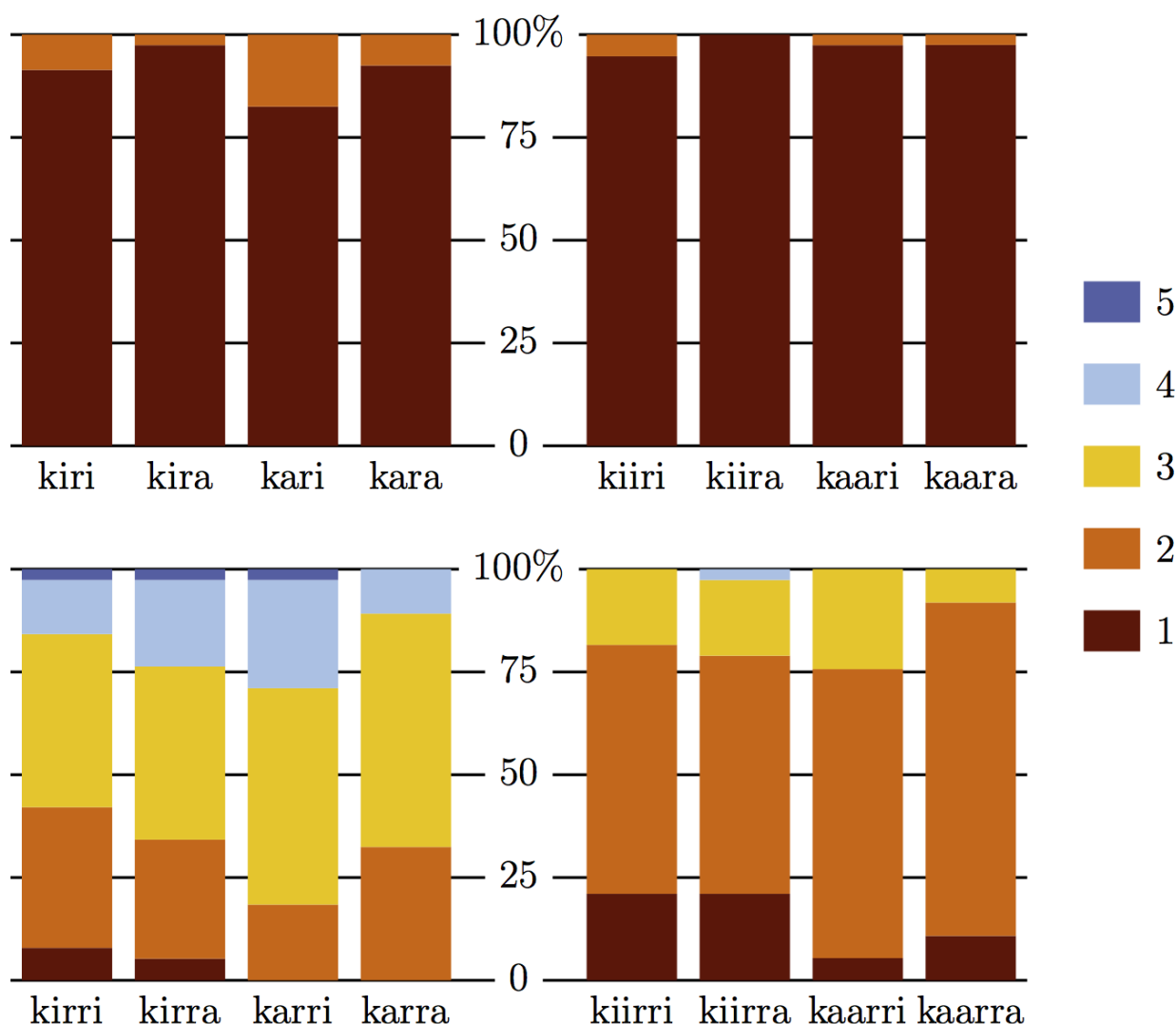
Kuva 3. Kvantiteettihahmon (**K**, yläriivi) ja vokaaliympäristön (**V**, alarivi) vaikutukset kestoon (vasemmalla) ja tärytaajuuteen (oikealla).

Kuvassa 3 esitetään faktorien eri kategorioiden vaikutukset poikkeamina (geometrisesta) keskiarvosta. Jokaisen osakuvion yläpuolella on logaritminen asteikko ja alapuolella vastaava lineaarinen asteikko. Nolla logaritmisella asteikolla (yksi lineaarisella asteikolla) tarkoittaisi, että kyseinen kategoria ei poikkea keskiarvosta, ja esim. arvo 2.0 lineaarisella asteikolla tarkoittaisi, että kategoria on keskiarvoon verrattuna kaksinkertainen.

Odotetusti /r/:n oma kvantiteetti vaikuttaa sen kestoon eniten. Muiden konsonanttien tavoin tremulantti lyhenee pitkän vokaalin jäljessä. Kestoero on suurempi pitkien kuin lyhyiden tremulanttien välillä. Vokaalin kvantiteetti vaikuttaakin erityisesti pitkän /r/:n kestoon. Tremulantti on pisimmillään pitkässä kvantiteetissa lyhyen vokaalin jäljessä ja lyhimmillään lyhyessä kvantiteetissa pitkän vokaalin jäljessä. Vokaalin kvantiteetti vaikuttaa myös tärytaajuuteen. Pitkä vokaali aiheuttaa hitaampia täryjä, mutta itse /r/:n kvantiteetti ei juuri vaikuta tärytaajuuteen.

Vokaaliympäristöistä *a—i*-yhdistelmä kasvattaa tremulantin kestoa selvästi. Muiden vokaaliyhdistelmien vaikutuksilla ei ole suurta keskinäistä eroa. Tremulantti on lyhyempi /i/-alkuisissa vokaaliympäristöissä, mutta ero verrattuna *a—a*-ympäristöön on pieni. Yhdistelmässä

a—i on tärytaajuus kuitenkin pienin, suurimmillaan se on *a—a*-yhdistelmässä. Vokaaliyhdistelmät, jotka loppuvat /i:/hin, aiheuttavat hitaampia täryjä kuin /a:/han loppuvat yhdistelmät.



Kuva 4. Tärymäärien yleisyys eri sanatyypeissä.

4.2. Täryjen määrä

Tärymäärien suhteellisia osuuksia eri sanatyypeissä esitetään kuvassa 4. Lyhyt /r/ on lähes aina yksitäryinen. Lyhyen vokaalin jäljessä esiintyy lyhyessä /r:ssä enemmän monitäryisyyttä kuin pitkän vokaalin jäljessä, mutta ero on pieni. Pitkän /rr/:n tärymäärät ovat vaihtelevampia, mutta myös pitkässä /rr:ssä on yleisimmin enemmän täryjä lyhyen vokaalin jäljessä. Lyhyen vokaalin jäljessä yleisin määrä on kolme ja pitkän vokaalin jäljessä kaksi. Lyhyen vokaalin jäljessä pitkä /rr/ on joskus myös kaksi- tai nelitäryinen, ja viisitäryisiäkin tapauksia on hieman. Pitkän vokaalin jäljessä pitkä /rr/ on puolestaan joskus myös kolmitäryinen ja erittäin harvoin nelitäryinen. Pitkässä

/rr/:ssä esiintyy jopa yksitäryisyyttä, joka on yleisempää pitkän vokaalin kuin lyhyen vokaalin jäljessä. Suurimmat tärymäärät esiintyvät *a—i*-ympäristössä.

5. Johtopäätökset

Pitkä /rr/ sisältää odotuksenmukaisesti yleensä enemmän täryjä kuin lyhyt /r/. Lyhyt /r/ on enimmäkseen yksitäryinen ja pitkä /rr/ monitäryinen. Kuitenkin myös pitkä /rr/ voi olla yksitäryinen. Suurimmat tärymäärät esiintyvät sekä pitkässä että lyhyessä kvantiteetissa lyhyen vokaalin jäljessä. Vokaalin kvantiteetti vaikuttaa pitkään /rr/:ään enemmän kuin lyhyeen. Näin ollen yksitäryinen [r] näyttäisi olevan osa tremulantin kestojatkuoa eikä oma erillinen äänteensä, jota vanhassa kirjallisuudessa on esitetty. Kvantiteetin suhteen /r/:n kesto käyttäytyy muiden konsonanttien tavoin. Muista konsonanteista poiketen /r/ on /i/:n jäljessä lyhyempi kuin /a/:n jäljessä. Odotusten vastaisesti /r/ on pisimmillään *a—i*-kontekstissa eikä *a—a*-kontekstissa. Tämä saattaa viitata siihen, että kielen selän suppeneva liike on täryjen muodostamisen kannalta oleellisempaa kuin suppeus tai väljyys sinänsä.

Edelleen voisi selvittää tärymäärän ja keston suhdetta. Pelkkä täryjen määrä ei yksiselitteisesti kerro tremulantin kvantiteettia, sillä sekä lyhyt /r/ että pitkä /rr/ voivat sisältää kaksi täryä ja pitkä /rr/ voi puolestaan olla jopa yksitäryinen. Tärymäärien päällekkäisyyksien vuoksi havainto /r/:n pituudesta saattaakin perustua itse /r/:n keston riippumatta täryjen määrästä. Näin ollen jatkotutkimuksena voisi selvittää, mitkä seikat vaikuttavat kuulijan havaintoon /r/:n pituudesta.

Lähteet

Barry, W. J. (1997). Another r-ticle, *Journal of the International Phonetic Association* 27(1, 2): 35–45.

Gelman, A. & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.

Iivonen, A. (1993). Tremulanttien fonetiikkaa ja typologiaa, teoksessa A. Iivonen & R. Aulanko (toim.), *Fonetiikan päivät — Helsinki 1992 / Papers from the 17th Meeting of Finnish Phoneticians*, Helsingin yliopiston fonetiikan laitoksen julkaisuja 36, s. 41–78.

Iskarous, K. & Kavitskaya, D. (2010). The interaction between contrast, prosody, and coarticulation in structuring phonetic variability, *Journal of Phonetics* 38(4): 625–639.

Kavitskaya, D. (1997). Aerodynamic constraints on the production of palatalized trills: The case of Slavic trilled [r], teoksessa G. Kokkinakis, N. Fakotakis & E. Dermatas (toim.), *Proceedings of Eurospeech 1997, Rhodes, Greece*: University of Pratas, s. 751–754.

- Ladefoged, P. & Maddieson, I. (1996). *The Sounds of the World's Languages*, Oxford, UK/Malden, U.S.A.: Blackwell.
- Lehtonen, J. (1970). *Aspects of Quantity in Standard Finnish*, Studia Philologica Jyväskyläensia VI, University of Jyväskylä.
- Lindau, M. (1985). The story of /r/, teoksessa V. A. Fromkin (toim.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, Academic Press, s. 157–168.
- Mustanoja, L. & O'Dell, M. (2004). Tremulantinäätämys tamperelaisittain, teoksessa T. Seppänen, K. Suomi & J. Toivanen (toim.), *Fonetiikan päivät 2004 / The Phonetics Symposium 2004*, MediaTeam Oulu, Suomen kielen, informaatiotutkimuksen ja logopedian laitos, Oulun Yliopisto, s. 21–25.
- Mustanoja, L. & O'Dell, M. (2007). Suomen *d* ja *r* sosiofoneettisessa kentässä, *Virittäjä* 1/2007: 56–67.
- Recasens, D. & Pallarès, M. D. (1999). A study of /r/ and /r/ in the light of the 'DAC' coarticulation model, *Journal of Phonetics* 27(2): 143–169.
- Solé, M.-J. (2002). Aerodynamic characteristics of trills and phonological patterning, *Journal of Phonetics* 30(4): 655–688.
- Sovijärvi, A. (1963). *Suomen kielen äännekuvasto*, Jyväskylä: Gummerus.