

# PUHujan TEMPORAALISEN ÄÄNIALAN VISUALISOINTISOVELLUS

Antti Iivonen, Helsingin yliopisto, Puhetieteiden laitos  
antti.iivonen@helsinki.fi

Tapio Seppänen, Oulun yliopisto, Sähkö- ja tietotekniikan osasto,  
MediaTeam, tapio.seppanen@ee.oulu.fi

Kai Nojonen, Oulun yliopisto, Sähkö- ja tietotekniikan osasto,  
MediaTeam, kai.nojonen@ee.oulu.fi

Juhani Toivanen, Oulun yliopisto, Sähkö- ja tietotekniikan osasto,  
MediaTeam, juhani.toivanen@ee.oulu.fi

*Kirjoituksessa esitellään uutta puhujan pragmaattisen temporaalisen äänialaprofiilin kuvantamissovellusta. Kehitystyön taustalla on Helsingin yliopiston fonetiikan laitoksessa laadittu TVRP-ohjelma, jolla voidaan kuvata puhujan tuottamien ilmausten temporaalinen ääniala. TVRPVA-ohjelma (1.0) on kehitetty Oulun yliopiston sähkö- ja tietotekniikan osaston MediaTeamissa. Java-kielillä laadittu sovellus mahdollistaa FO-kontuuriaineiston vuorovaikutteisen visualisoinnin ja analyysin. Temporaalista äänialaa voidaan kuvata kolmella eri visualisaatiolla, joita ovat 'contour plot', 'min-max plot' ja 'percentile plot'. Taajuus voidaan esittää Hz- tai puolisävelasteikolla. Lisäksi käytetyn kontuuriaineiston tilastollisia ominaisuuksia voidaan tarkastella myös histogrammina. Jokaisen tuotetun äänialan päälle voidaan piirtää yksi tai useampi ilmaus ja näin suhteuttaa niiden ominaisuudet äänialaan ja ilmauksen funktioihin. Halutut ilmaukset voidaan joko poimia listalta tai itse äänialaprofiilista. Sovellusaloja ovat ainakin prosodian, puhetyyliin, tunteiden ja asenteiden tutkimus, voice-tutkimus, puhujan- ja kielen tunnistus, kielten vertailu ja murretutkimus.*

**Avainsanat:** temporaalinen ääniala, prosodia, äänitutkimus, temporaalisen äänialan visualisointiohjelma

## JOHDANTO

Puheen perustaaajuutta koskevissa tutkimuksissa on jo pitkään käytetty puhujan äänialaa ilmaisevia ohjelmia (esim. Orlikoff & Kahane 1993; Orlikoff & Baken 1996). Niis-

sä ääniala esitetään ilman aikaulottuvuutta ja ensisijaisesti puhujan ääniresurssien kuvauksena, ja tavallinen sovellusala on äänihäiriöiden tutkimus. Äänialan temporaalisen ilmaisun tarve on kuitenkin olemassa etenkin puheen prosodian tutkimuksissa. Ääniala voi vaihdella puhuja-, tyyli ja kielikohtaisesti niin, että ajallisella dimensiolla on merkitystä. Jokainen äännetty ilmaus sijoittuu puhujan äänialan sisään ja sijoituksella on funktioita, joista toiset ovat selvemmin kie-

Kirjeenvaihto-osoite:  
Antti Iivonen  
Puhetieteiden laitos  
PL 35 (Vironkatu 1)  
00014 Helsingin yliopisto

len systeemin ydinalueeseen kuuluvia, toiset pikemminkin pragmaattisia (vuorovaikutukseen kykeytyviä), tyyllisiä, emotionaalisia, sosiaalisia, jopa esteettisiä. Pragmaattisuus viittaa siihen, että tutkimuskohteiksi voidaan valita aidoissa puhetilanteissa äänitettyjä puhenäytteitä; sillä varauksella, että äänitysolosuhteet ovat riittävän hyvät. Puhujan ääniala on tärkeä tekijä puheen prosodiassa olipa tarkastelukohteena intonaatio, rytmi tai painotus, luettu puhe, haastattelu, spontaani monologi tai keskustelu, kielten vertailu tai murretutkimus. Niiden tutkimuksen ohella menetelmä tarjoaa metodin ainakin seuraaville aloille: tunteiden ja asenteiden tutkimus, *voice*-tutkimus, puhujan- ja kielen tunnistus ja puhetyylien tutkimus. Perustaajuusvaihteluun latautuu erityisen paljon funktionaalisuutta (esim. noin 600 kirjallisuuslähteen katsaus Iivonen, Nevalainen, Aulanko & Kaskinen 1987; Iivonen 1998; 20 kielen intonaatiokuvaus: Hirst & Di Cristo 1998; Toivanen 2001).

Esittelemme uutta puhujan pragmaattisen temporaalisen äänialaprofilin kuvantamisovellusta (*Temporal Voice Range Profile Visualisation Application; TVRPVA*). Sitä edelsi FutureBasic II -kielellä laadittu *Temporal Voice Range Profile (TVRP)*; Iivonen 1999, 2001a; sovellusesimerkkejä myös Iivonen 2001b, c; uuden ohjelman alustava esittely: Iivonen, Seppänen, Noponen & Toivanen 2002). Uudella ohjelmalla voidaan koota joukko ilmauksia yhteiseen visualisointiin, laskea kuvan perustaajuuskontuurien aikapisteiden, käyttäjän määriteltävissä olevat persenttiilit ja asettaa yksittäisen valitun ilmauksen perustaajuuskontuuri äänialakuvan päälle, jolloin voidaan osoittaa sen täsmällinen paikka puhujan profiilissa. Taajuus voidaan ilmaista hertsi- (Hz) tai puolisävelasteikolla (semitone; ST). Lisäksi puhujan ääniala voidaan ilmaista histogrammina ilman aikaulottuvuutta, ja käyttäjä voi määritellä taajuusasteikon luokkavälin leveyden. Pragmaattisuus

viittaa siihen, että aineistona voidaan käyttää kaikilla tyyllilajeilla tuotettua puhetta puhujan todellisissa puhetilanteissa käyttämäästä puheesta. Puhe voi näin ollen olla yhtä hyvin spontaania puhetta, lukupuhuntaa kuin laboratoriopuhetta. On tietenkin mahdollista käyttää aineistona tutkimusta varten suunniteltua puhetta, vaikkapa tarkoituksellisesti huudettua tai hiljaista puhetta. Äänitystekniikan tasovaatimukset voi käyttäjä itse valita haluamikseen.

## TVRPVA-SOVELLUSOHJELMAN OMINAISUUDET

Uusi temporaalisen äänialan visualisointisovellus TVRPVA 1.0 on laadittu Java-ohjelmointikielellä alustariippumattomuuden saavuttamiseksi. Sovellus mahdollistaa F0-kontuuriaineiston vuorovaikutteisen visualisoinnin ja analyysin. Tarvittavien F0-kontuurien laskentaan voidaan käyttää esimerkiksi Praat-ohjelmaa, jonka tuottamia *Pitch Tier-ShortText* -tiedostoja se osaa lukea. Sovellus keskittää kaikki käytetyt ilmaukset automaattisesti alkamaan ajanhetkestä nolla. Tämä mahdollistaa käytettyjen äänitiedostojen karkeamman segmentoinnin ilmauksiksi.

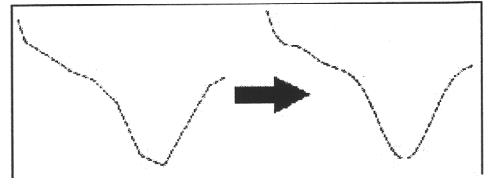
Temporaalista äänialaa voidaan kuvata kolmella eri visualisaatiolla, joita ovat perustaajuuskäyrästä ('contour plot'), ääriarvokäyrät ('min-max plot') ja persenttiilikuvain ('percentile plot'). Näiden kolmen äänialaprofilin lisäksi käytetyn kontuuriaineiston tilastollisia ominaisuuksia voidaan tarkastella histogrammivisualisaatiolla. Ensimmäisessä visualisaatiossa äänialaprofiili muodostetaan superpositioperiaatteella piirtämällä kaikki kontuurit tai niitä vastaavat aineistopisteet päällekkäin samaan kuvaan ajan funktiona. Tämä visualisaatio vastaa aiemman sovelluksen toimintaa (esim. Iivonen 1999). Toinen äänialaprofiilivisualisaatio, 'min-max plot', koostuu F0-kontuurien minimi- ja maksimiarvojen rajaamasta aluees-

ta aikatasossa. Kolmas visualisaatio taas kuvaa kontuurien jakauman ajan funktiona halutuim persenttiivälein, kuten esimerkiksi kvartileit-tain eli 25 % luokkavälein. Tässä visualisaati-ossa aineistolle vieraat havainnot (*outliers*) voi-daan poistaa suodattamalla (*filter*) asettamalla tarkasteluvälin jakauman minimi ja maksimi-rajat halutuiksi. Samaan visualisaatioon voi-daan piirtää useita eri äänialaprofileja ja ver-tailla niitä toisiinsa. Näin voidaan tutkia esi-merkiksi puhujien tai vaikkapa tunnetilojen välisiä eroavuuksia.

Jokaisen tuotetun äänialan päälle voidaan helposti piirtää yksi tai useampi ilmaus ja näin suhteuttaa niiden ominaisuudet äänia-laan ja ilmauksen funktioihin. Halutut il-maukset voidaan joko poimia listalta tai itse äänialaprofiilista, ja ne voidaan halutta-essa kuunnella helposti. Havaittuja ominai-suuksia voidaan sitten korostaa käyttämäl-lä värejä. Ilmausaineistoa voidaan suodat-taa poistaen tiettyjä taajuuksia pienemmät ja suuremmat F0-arvot, jos esimerkiksi ää-nitysteknisistä syistä tai F0-kontuurien las-kennassa käytetyn algoritmin vuoksi aineis-tossa on F0-pisteitä, joka halutaan siitä pois-taa. Näin voi käydä esimerkiksi, jos puheessa on narinaa. Kaikki visualisaatiot voidaan piir-tää joko hertsi- tai puolisävelasteikolla. No-lanin (2003) mukaan kuulohavaintoa vas-tasi parhaiten puolisävelasteikko, kun loga-ritmisen, erb- ja puolisävelasteikon soveltu-vuutta evaluoitiin.

Sen, millaisia ilmauksia sijoitetaan ohjel-maan niin, että niillä on yhteinen ajallinen nollapiste, riippuu käyttäjän tavoitteista. Il-maukset voivat olla esimerkiksi erillisiä täy-dellisiä lauseita, tauon jälkeisiä lausumia, keskustelun puheenvuoroja tai funktioltaan identtisiä lausumia (esimerkiksi kysymyssa-nalla alkavia kysymyksiä). Funktiot voidaan ymmärtää esimerkiksi prosodisten teorioiden, puheaktiteorian tai keskusteluanalyy-sin kategorioiksi.

Tuotetut visualisaatiot voidaan tallentaa halutulla tarkkuudella *Tag Image File Format* (TIFF) muotoisina kuvina käyttäen *Pack-Bits*-pakkausta. Valitusta tarkkuudesta riip-puen ohjelma automaattisesti joko interpo-loi tai desimoi perustaajuuskontuurien ai-neistoa. Aineiston interpolointiin on käy-tettävissä kaksi erilaista menetelmää: lineaarinen ja *spline*-interpolointi. Linearisessa interpoloinnissa aineistoon sovitetaan suoria ja *spline*-interpoloinnissa paloittain matala-asteisia sileitä interpolointipolynomeja. Ku-vassa 1 on esimerkki F0 kontuurista, jonka vasemman puoleinen osa esittää lineaarisel-la interpoloinnilla tuotettua ja oikealla puo-lella oleva osa *spline*-interpoloatiolla tuotet-tua kuvaajan osaa. *Spline*-interpoloatio tuot-taa jatkuvia sileitä käyriä, joissa ei ole terä-viä kulmia. Suurilla aineistomäärillä työsken-neltäessä nopeampaa lineaarista interpoloin-tia voidaan käyttää esikatseluun, ja lopullis-ta kuvaa tuotettaessa voidaan siirtyä käyttä-mään *spline*-interpoloatiota.



Kuva 1. Linearisesti interpoloimalla tuotettu F0-kontuuri vasemmalla ja *spline*-interpoloimalla tuotettu kontuuri oikealla.

Perustaajuusarvojen muunto puolisävelas-keliksi voi tapahtua eri algoritmien avulla. Erilaisuutta aiheuttaa muun muassa se, mikä Hz-asteikon arvo otetaan viitearvoksi (= *base*). Voidaan lähteä nollatasolta (0 Hz), tai alimmasta kuultavissa olevasta taajuudesta (16,35 Hz; ”standardiääni”, Orlikoff & Ba-ken 1993: 72) tai tavallisesta miesäänen pe-rustaajuuden keskiarvosta (100 Hz). Muun-noskaavoja ovat seuraavat:

(1)

$$st(n) = 39,86 * \text{LOG}_{10}(f(n)/\text{base})$$

(Orlikoff ja Baken 1993:72)

(2)

$$st(n) = 12 * \text{LN}(f(n)/\text{base}) / \text{LN}(2)$$

(esim. Praat-puheenkäsittelyohjelmassa)

Niissä

f(n) = mitattu Hz arvo

st(n) = f(n):n arvo puolisävelinä

base = referenssitaajuudeksi asetettu Hz-arvo

Ero on kuitenkin näennäinen, sillä laskukaavat ovat samat logaritmien laskusääntöjen perusteella. Kun muodostetaan lukusarjan 1–440 (Hz) arvot puolisävelaskeliksi kummallakin kaavalla, saadaan lähes identtiset puolisävelaskelleet. TVRPVA-ohjelmassa muunto tapahtuu kaavalla (2) ja käyttäjä itse voi asettaa referenssiarvon.

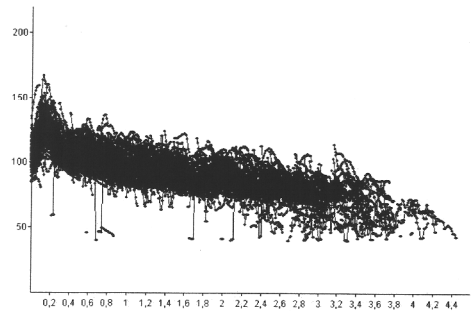
## SOVELLUSESIMERKKEJÄ

Esimerkkisovelluksessamme esitellään yhden mies- ja yhden naispuhujan luetun puheen ääniala. 30-vuotias miespuhuj M30 on syntynyt Espoossa ja asunut siellä aina. 27-vuotias naispuhuj N27 on syntynyt Forssassa ja opiskellut Helsingin seudulla sekä oleskellut paljon Turussa. Kumpikin luki tekstin, joka käsitti 100 lausetta. Lauseet oli poimittu yhtenäisestä vanhaa kaupunkiarkkitehtuuria kuvaavasta kirjoituksesta, mutta ne kaikki muokattiin 7-sanaisiksi, eikä yhdyssanoja sallittu. Lauseilla oli tekstuaalisia yhteyksiä, mutta puhujat saivat ohjeen välttää niitä, koska esimerkkiaineiston tarkoituksena on pikemminkin vain kuvata tässä vaiheessa TVRPVA:n perusominaisuuksia kuin tutkia erilaisia käyttömahdollisuuksia. Lauseet luettiin siten irrallisina kontekstistaan. Naispuhujan lauseissa on kuultavissa jonkin verran merkitysisältöön liittyvää äänen ”elävöittämistä”. Luettaessa puhe tallennettiin pääpantaan kiinnitetyn mikrofonin (kondensaattorimikrofoni AKG C444L) ja (PowerMacintosh G3) 16-bittisen äänikortin kautta suoraan tietokoneeseen käyttäen SoundEdit 16.2 -ohjelmaa. Näytteenottotaa-

juus oli 44100 Hz. Lauseet editoitiin kukin omaksi äänitiedostoksekseen. Lauseiden laadinnan, äänitykset ja äänitiedostojen editoinnin suoritti Liisa Vilhunen.

Praat-puheenkäsittelyohjelmalla (4.0.30) laskettiin lauseiden perustaajuuskontuurit ja muodostettiin niistä *PitchTier*-tiedostot käyttäen *short-text*-tallennusta. Laskennan esisuodatusrajat voivat vaikuttaa tilastollisiin arvoihin. Perustaajuuden (F0) laskennassa sovellettiin miespuhujalle esisuodatusrajoja 40–160 Hz ja naispuhujalle 90–310 Hz.

Praat-puheenkäsittelyohjelmalla lasketut perustaajuuden tilastolliset tunnusluvut (minimi, mediaani, keskiarvo, maksimi ja hajonta) kuvatuista kahdesta aineistosta ilmenevät taulukosta 1. Siitä näkyy, että miespuhujan lauseiden kestojen keskiarvo on 3,815 sekuntia (vaihteluväli oli 2,554–4,876 s) ja naispuhujan vastaava arvo 3,914 sekuntia (vaihteluväli 2,606–5,433 s).



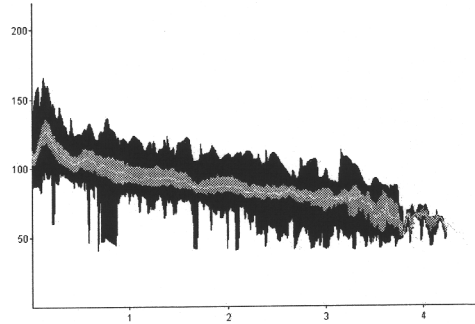
Kuva 2. Miespuhuj M30 lukeman 100 lauseen perustaajuuskontuurit (*ContourPlots*) asetettuina samaan kuvaan päällekkäin aika (s)/taajuus (Hz) -asteikolle.

Kuva 2 näyttää miespuhujan 100 lauseen perustaajuuskontuurit (*ContourPlots*) asetettuina samaan kuvaan päällekkäin. Vaaka-asteikko kuvaa aikaa sekunneissa (s), pystyasteikko taajuutta (Hz). Kuva osoittaa aineiston jakautuman pääalueen sijoittuvan noin 70–150 Hz:n väliin. Samalla käy kuitenkin selväksi, että äänialan temporaalinen esitys näyttää realistisemmin äänialan ominaisuuksia verrattuna



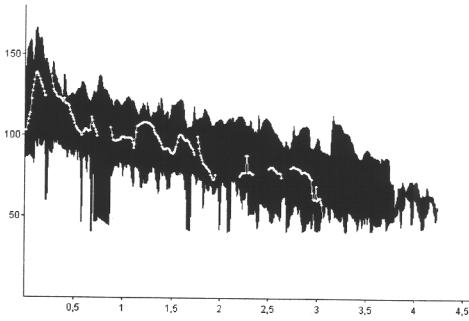
taulukon 1 tilastoon. Jakauma-alue osoittaa selvää deklinaatiota eli laskua loppua kohti, ja lasku tapahtuu suhteellisen suhteellisen leveänä, noin 40 Hz leveänä nauhana. Ääniala itse asiassa siis liikkuu ajan mukana. Jakau- man alussa on selvä huippu noin 150 ms:n kohdalla. Jakauma kapenee loppua kohden ja lauseiden loput laskevat 50 Hz:n tasolle, osin mahdollisesti esisuodatuksen alarajan asetuk- sen alapuolellekin (40 Hz). Perustaajuuspistei- den haja-arvoja sattuu pääjakauman ylä- ja ala- puolelle, mutta selvää kuvaa siitä, kuinka pal- jon tällaisia tapauksia on, ei kontuurikuvauk- sesta saa. Koska lauseiden kesto vaihteli, ovat niiden lopputaajuudet eri aikapisteissä, mistä johtuu jakauman lopun hajonta.

Äänialan persentiilijakaumalla voidaan pa- remmin kuvata yleisempiä ja poikkeavampia kontuurien pisteitä ja jaksoja. Kuva 3 näyttää miespuhujan 100 lauseen jakauman s/Hz-as- teikolla. Harmaa alue käsittää 50 % jakaumas- ta (persentiilit 0, 25, 50, 75, 100 %). Valkoi- nen viiva kuvaa mediaania (sen kummallakin puolella on 50 % koko jakaumasta). Puhujan lausumien sävy oli neutraali, ja voidaan arvi- oida, että jakaumakuva edustaa hyvin hänen tavanomaista neutraalin luetun puheensa ää- nialaa. Puolet hänen äänialastaan osuu var- sin kapealle, noin 20 Hz:n alalle, ja vastaa- va deklinaatio on havaittavissa kuin kuvassa 2 sekä tämän 50% :n jakauman että koko ja- kauman suhteen. Alakvartiili osoittaa jonkin verran narinarekisterin esiintymiä, jotka erot- tuvat ehkä yksittäisinä tapauksina paremmin kuvassa 2. Koko äänialan loppupuolella nari- naesiintymiä ilmenee runsaammin, mikä on ymmärrettävää lausuman loppurajan lähes- tyessä. Koska pisimmät lausumat ovat harvi- naisia, on äänialajakauman lopun persentiili- jakauma tapausten lukumäärän vähäisyyden vuoksi epämääräinen.



Kuva 3. Miespuhujan 100 lauseen jakauma Hz-astei- kolla (percentile plot). Harmaa alue käsittää 50 % ja- kaumasta (persentiilit 0, 25, 50, 75, 100 %). Valkoi- nen viiva kuvaa mediaania (sen kummallakin puolel- la on 50 % koko jakaumasta).

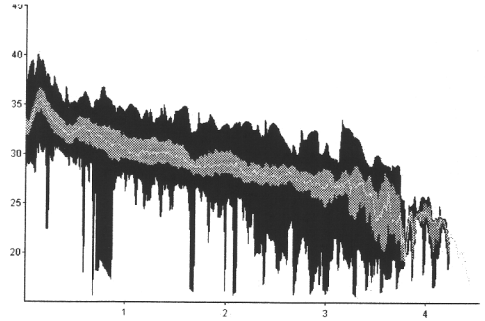
Pääosa äänialasta muodostuu lausumien aksenttien nousujen ja laaksojen muodosta- masta F0:n vaihtelusta. Suomenkielisen pu- heen monisanaisen lausuman tavanomai- nen kontuuri näkyy kuvassa 4, jossa yksit- täisen lauseen ”*Tämä sarja esittelee vanhoja Helsingistä purettuja taloja.*” F0-kontuuri nä- kyy valkoisena juovana äänialan päällä. On tunnettua, että painon tärkeä korrelaatti on äkillinen lyhytkestoinen perussävelen nou- su- ja laskukuvio. Painotetut tavut muodos- tavatkin kuvan 4 käyrän huiput ja painot- tomat käyrän laaksot. Kuvan 4 kontuuri ei kuitenkaan ulotu koko äänialan reunoille, jo- ten reunat muodostuvat esimerkiksi tapauk- sista, joissa painotus on ollut voimakkaam- paa. Soinnittomien ääniteiden kohdalla käy- rässä on tietenkin katkos. Tarkempi analyysi paljastaisi lausuman rytmijaksottelun ja sen, mitkä sanat ovat painotettuja. On ymmärret- tävää, että painohuippujen ja -laaksojen ra- jaama alue vaihtelee painotusasteiden mu- kaan, ja tehotetut painot ja emotionaaliset tavut nostavat F0-kontuuria erityisen kor- kealle. Siksi temporaalinen ääniala ei täysin vastaa Gårdingin ja Bručen kehittämän niin sanotun Lundin mallin *grid*-käsitettä, mut- ta on sitä lähellä (ks. Gårding 1982; Iivonen ym. 1987: 255).



Kuva 4. Miespuhujan 100 lauseen jakauma Hz-asteikolla (*percentile plot*; musta alue = 100 %; persenttiilit 0, 100 %). Äänialan päällä sijaitseva valkoinen juova esittää yhden erillisen lauseen kontuuria.

Kuvassa 5 miespuhujan aineisto on sijoitettu puolisävelasteikolle. Muunnoksella pyritään simuloimaan kuulijan havaintoa. Puolisävelasteikon käyttö mahdollistaa myös puhujien vertailun riippumatta heidän perustaajuuden keskiarvojen erilaisuudesta. Keskiarvoltaan korkeampi ääniala tosin on puolisävelasteikollakin korkeammalla, mutta suhteelliset erot ovat vertailukelpoisia. Asteikko paljastaa kuvassa 5 selvemmin kuin kuvassa 3 sen, että ilmausten lopussa noin 3,7 sekunnin kohdalla 50 %:n keskialue on laajempi kuin alkupuolella, mikä aiheutuu siitä,

että lausumien loppumarkkerin pitää sijoitua suhteellisen kapealle alalle eli mahdollisuuksia ylä- ja alakvartiilin muodostumiselle ei enää ole yhtä suuressa määrin.



Kuva 5. Miespuhujan 100 lauseen jakauma puolisävelasteikolla (*percentile plot*; persenttiilit 0, 25, 50, 75, 100 %). Viitearvo (*base*) oli 16,35 Hz.

Naispuhujan 100 lauseen ääniala esitetään Hz-asteikolla kuvassa 6. Hänellä ei juuriakaan ilmene F0:n deklinaatiota, jollainen miehellä selvästi on. Äänialan loppuosa näyttää osoittavan nousevia kontuureja. Yksittäisten lauseiden F0-nousuja voi tutkia 33 ms:n aikaikkunalla spektrogrammeista Praatilla, jolloin yläsävel näkyvät. Niistä ilmeni, että hänellä on varsin usein lievä narina lopussa, johon liittyy todellakin yläsävelten nou-

Taulukko 1. Käytetyn mies- ja naispuhujan 100 lauseen perustaajuuden tilastolliset tunnusluvut. Taulukko esittää lausekohtaisista arvoista lasketut keston tai taajuuden keskiarvot ja hajonnat.

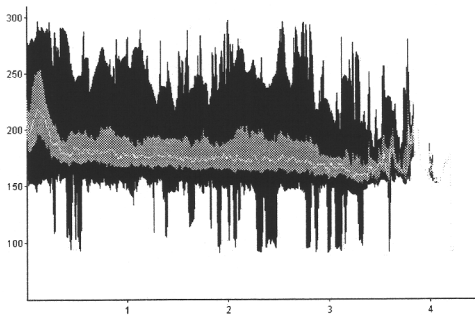
Puhuja		lausekesto (s)	minimi	mediaani	keskiarvo	maksimi	hajonta
M30	keskiarvo	3,815	52,6	90,9	92,5	132,9	15,9
	hajonta	0,504	10,1	3,4	3,2	10,7	2,3
N27	keskiarvo	3,914	130,5	178,1	183,7	262,1	23,8
	hajonta	0,597	27,0	11,3	10,1	27,0	6,4

Taulukko 2. Esisuodatusrajojen vaikutus F0-tilastoon: rajojen 90–130 ja 130–310 vertailu.

esisuodatusrajat		minimi	mediaani	keskiarvo	maksimi	hajonta
90-310 Hz	keskiarvo Hz	130,5	178,1	183,7	262,1	23,8
	hajonta Hz	27,0	11,3	10,1	27,0	6,4
130-310 Hz	keskiarvo Hz	145,5	178,0	183,6	254,2	22,6
	hajonta Hz	9,4	11,4	10,3	28,6	6,5

sevia kulkuja. Mahdollisesti loppunousujen taustalla on lievässä muodossa loppunousumalli, joka nykyisin tavataan tietyissä nuorten puhujien puheessa (Routarinne 2003). Luettelointonaatiosta ei ole kysymys, koska koehjeena oli nimenomaan sen välttäminen, eikä sitä kuuntelun perusteella esiintynyt. Kummallekin puhujalle on yhteistä alussa noin 150 ms:n kohdalla oleva lauseiden aloitushuippu.

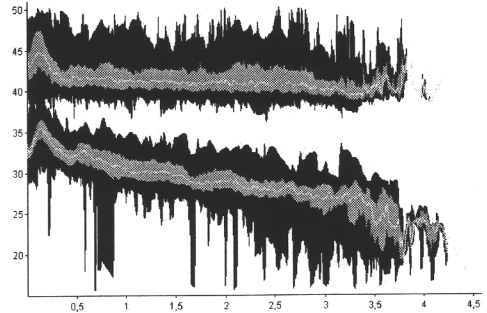
Prosodia ja äänitutkimus (*vokologia; voice research*) kohtaavat sikäli toisensa, että soinnillinen (modaalinen) ääni korvautuu usein puheen tavuissa soinnittomuudella, narinalla tai puristeisella äänellä, joten niiden hallinta prosodian tutkimuksessa tulee välttämättömäksi. Laryngaalinen äänenlaatu (*voice quality*) muodostuukin näin ollen prosodian ”neljänneksi ulottuvuudeksi” keston, äänen korkeuden ja voimakkuuden ohella (Campbell 2003).



Kuva 6. Naispuhujan 100 lauseen jakauma Hz-asteikolla (*percentile plot*). Puhuttu aineisto oli sama tekstiaineisto kuin miespuhujalla (persentiilit 0, 25, 50, 75, 100 %).

Kuvassa 6 näkyvät N27:n noin 140 Hz:n alapuolella ja noin 90 Hz:iin ulottuvat piikit edustavat muusta aineistosta poikkeavia arvoja, ja ne aiheutunevat pääosin narinasta. Kun koko N27:n aineisto laskettiin uudelleen esisuodatusarvoilla 130–310 Hz, saatiin taulukon 2 F0-tilasto. Voidaan havaita, että mediaani, keskiarvo ja hajonta ovat muuttuneet vain hieman. Poikkeavien arvojen lukumäärä on siis niin pieni, että ne eivät näy te-

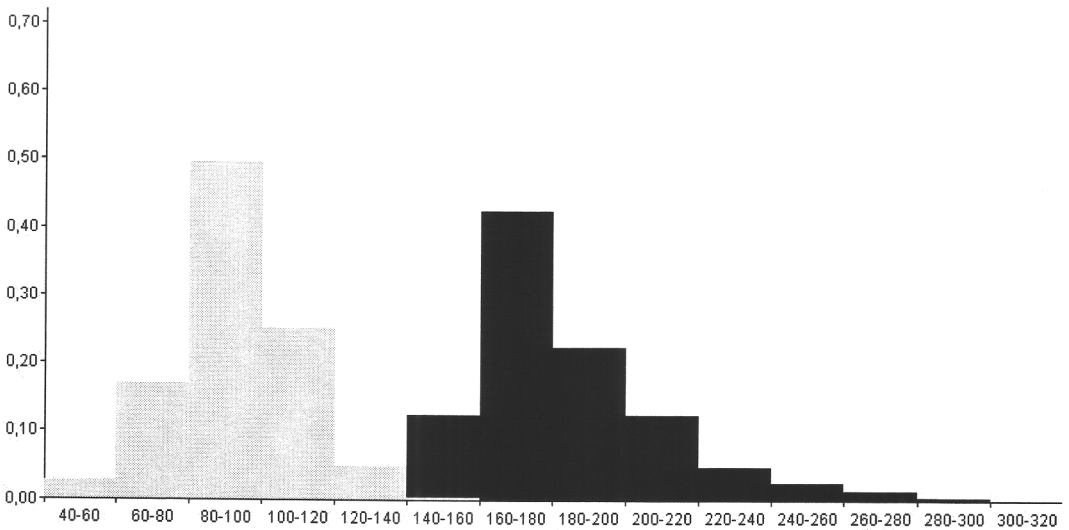
merkittävää osaa puhujan F0-tilastossa. Vain minimiarvo ja sen hajonta 100 lauseessa ovat ymmärrettävästi muuttuneet paljon. Yläkvartiiliin sattuu jonkin verran tapauksia (yksittäisiä teräviä piikkejä), jotka ovat selitettävissä esisuodatuksen liian korkeasta ylärajasta. Tämä seikka riippuu kuitenkin Praat-ohjelman toiminnasta, ei TVRPVA-ohjelmasta, ja kyseinen virhelähde paljastuu selvästi juuri TVRPVA-ohjelmalla. Virhettä voidaan korjata muun muassa käyttämällä F0:n laskennassa Praatin optiota, jolla voidaan vähentää sitä Hz-arvoa, joka sallitaan vierekkäisten F0-arvojen erotukseksi.



Kuva 7. Mies- ja naispuhujan äänialan vertailu samassa kuvassa (*percentile plot*) puolisyvelasteikolla. N27:n äänialasta on jälkisuodatettu alle 130 Hz:n F0-pisteet, jolloin poikkeukselliset tapaukset eivät näy kuvassa.

Kuvassa 7 mies- ja naispuhujan 100 lauseen äänialat on esitetty samassa kuvassa. Naispuhujan aineistosta jälkisuodatettiin TVRPV:lla alle 130 Hz:n F0-pisteet, jolloin narinarekisteri on jäänyt pois. Äänialan siirto puolisyvelasteikolle osoittaa, että naispuhujan äänialassa yläkvartiili on laajempi kuin miespuhujalla, mikä johtunee puhujan pyrkimyksestä elävöittää lauseiden ilmaisuja. Puhujien äänialat leikkaavat toisiaan vain hyvin vähän. Miesäänen äänialan deklinaatio ja lopun lasku erottuvat hyvin naisäänen äänialan tasaisuudesta ja lopun noususta. Koska pisimmät lauseet ovat aineistossa poikkeuksellisia, niiden loput eivät enää näy visualisaatiossa.

Kuvan 8 histogrammissa näkyvät sekä nais-



Kuva 8. Mies- ja naispuhujan (tummempi alue) aineistojen yhdistetty histogrammi. Hz-asteikko 40–320 Hz, bin-luku 14, esisuodatus 40–160 (mies) ja 90–310 (nainen) Hz.

että miespuhujan aineistot. Taajuusasteikko on jaettu 20 Hz:n luokkaväleihin (*bin*-luku kaistalla 40–320 Hz = 14). Taulukon 1 yhteydessä mainittuja esisuodatusrajoja sovellettiin. Miespuhujan aineisto jakaantuu kuuteen ensimmäiseen luokkaan ja limittyy hieman naispuhujan histogrammin kanssa luokassa 140–160 Hz. Naispuhujan hieman vasemmalle vino jakauma osuu alaosastaan miespuhujan kanssa, mutta limittymisalueelle sattuu vain muutamia esiintymiä, eikä tämä alue näy kuvassa.

## PÄÄTELMÄ

Kirjoitus on uuden TVRPVA-ohjelman perusominaisuuksien esittely. Kuvaus koskee menetelmän ensimmäistä versiota, jota on tarkoitus kehittää edelleen. Ohjelma mahdollistaa sen, että yksilöllisten puhujien tai puhujaryhmienkin äänialasta saadaan aikaisidonnainen profiili ja profiileja voidaan vertailla keskenään. Kirjoituksessa osoitetaan, että näin saavutetaan etuja, joilla puheaineistojen perustaajuusominaisuudet saadaan selkeästi ilmi suhteessa puhujan äänialaan. Menetelmä on pragmaattinen, koska puhetylien rajoituksia ei ole, eikä

tarvitse tyytyä laboratoriossa äänitettyyn puheeseen. Puhetylien eroja voidaan havainnollistaa. Yksittäinen perustaajuuskontuuri voidaan sijoittaa puhujan äänialan muodostamaa taustaa vasten, mikä mahdollistaa päätelmien teon tutkittavan kontuurin akustisten ominaisuuksien ja kielellisen funktion yhteydestä suhteessa äänialaan. Myös mahdolliset virhelähteet perustaajuuslaskennassa havaitaan paremmin. Kovin pienistä aineistomääristä profiilia ei kuitenkaan kannata laskea, eikä teknisesti heikoktasoisia äänitteitä ole syytä käyttää. Puolisävelasteikon käyttö mahdollistaa sen, että kuulijan auditiivisia vasteita perustaajuuteen voidaan simuloida.

## VIITTEET

- Campbell, N. (2003). Voice quality: the 4th prosodic dimension. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona 3-9 Aug. 2003*, (s. 2417–2420).
- Gärding, E. (1982). Swedish prosody. Summary of a project. *Phonetica*, 39, (s. 288–301).
- Iivonen, A. (1998). Functional interpretation of prosody within the linguistic system. *The 1998 Yearbook of the Linguistic Association of Finland*, (s. 69–91).

- Iivonen, A. (1999). F0 contours of utterances superimposed on the temporal voice range profile of the speaker. Proceedings of the XIV International Congress of Phonetic Sciences 1999, San Francisco 1–7 Aug., (s. 953–956).
- Iivonen, A. (2001a). Pragmatic temporal voice range profile as a tool in the research of speech styles. *VII Eurospeech 2001 Scandinavia, Aalborg 3–7.9.2001. Vol. I*, (s. 103–106).
- Iivonen, A. (2001b). Kommunikatiivisten aktien ja puhetyylien prosodiasta. S. Ojala & J. Tuomainen (toim.), *21. Fonetikan päivät, Turku 4.–5.1.2001*, (s. 1–18). Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 67.
- Iivonen, A. (2001c). Intonation of Finnish questions. W.A. van Dommelen & T. Fretheim (toim.), *Nordic Prosody. Proceedings of the VIIIth Conference, Trondheim 2000*, (s. 137–151).
- Iivonen, A., Nevalainen, T., Aulanko, R. & Kasanen, H. (1987). *Puheen intonaatio*. Helsinki: Gaudeamus.
- Iivonen, A. & Seppänen, T., Noponen, K. & Toivanen, J. (2002). Puhujan pragmaattisen temporaalisen äänialan visualisointimenetelmä ja -ohjelmisto. Petri Korhonen (toim.), *Fonetikan päivät 2002*. Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Report 67, (s. 123–132).
- Nolan, F. (2003). Intonational equivalence: An experimental evaluation of pitch scales. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona 3–9 Aug. 2003*, (s. 771–774).
- Orlikoff, R.W. & Kahane, J.C. (1996). Structure and function of the larynx. N. Lass (toim.), *Principles of Experimental Phonetics*. Mosby/St. Louis/Baltimore/ym., (s. 112–181).
- Orlikoff, R.F. & Baken, R.J. (1993). *Clinical Speech and Voice Measurement Laboratory Exercises*. San Diego: Singular Publishing Group.
- Routarinne, S. (2003). *Tytöt äänessä. Parenteesit ja nouseva sävelkulku kertojan vuorovaikutuskeinoina*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Toivanen, Juhani (2001). *Perspectives on Intonation: English, Finnish and English Spoken by Finns. Forum Linguisticum 37*. Perustuu Oulun yliopistossa 1999 hyväksytyyn väitöskirjaan. Frankfurt am Main ym.: Peter Lang.

## VISUALISATION APPLICATION OF SPEAKER'S TEMPORAL VOICE RANGE

Antti Iivonen, Department of Speech Sciences, University of Helsinki, Finland

Tapio Seppänen, Kai Noponen, Juhani Toivanen, Department of Electrical Engineering, Computer Engineering Laboratory, MediaTeam, University of Oulu, Finland

This paper describes a new visualization application for the pragmatic temporal voice range profile of the speaker: TVRPVA. The application is based on the TVRP program developed at the Department of Phonetics, University of Helsinki. The TVRPVA 1.0 program was developed in MediaTeam Language and Audio Technology Group, University of Oulu. TVRPVA was implemented with Java language, and it enables an interactive visualization and analysis of F0-contour data. The temporal voice range profile can be displayed with three different visualizations: the contour plot, the min-max plot and the percentile plot. Frequency can be displayed on the Hz scale or on the semitone scale. In addition, the statistics of the contour data can be displayed as histograms. It is possible to superimpose one or several utterances on each voice range, thus showing their characteristics in relation to the voice range and the functions of the utterance. The utterances can be selected from the list or from the voice range profile itself. TVRPVA has applications at least in the following fields: research on prosody, speech styles, emotions and attitudes, voice research, recognition of speaker and language, comparison of languages and study of dialects.

**Keywords:** temporal voice range, prosody and voice research, program for the visualisation of temporal voice range