https://doi.org/10.7592/MT2017.68.pajupuu_altrov

Arvutiparalingvistika väljakutsed ja eesti hääle meeldivus

Hille Pajupuu
Eesti Keele Instituudi juhtivteadur
eki@eki.ee

Jaan Pajupuu
tarkvaraarendaja
eki@eki.ee

Rene Altrov
Eesti Keele Instituudi teadur
eki@eki.ee

Teesid: Artiklis heidetakse pilk viimase kümnendi suundumustele arvutiparalingvistikas: kõneleja omaduste ja seisundite tuvastamisele häälest ning nõuetest sellega seonduvatele kõnekorpustele. Tutvustatakse Eesti häälekorpuse olemust ning võimalust hääle meeldivust akustiliselt iseloomustada ja automaatselt tuvastada, kasutades Genfi minimaalset akustiliste parameetrite laiendatud kogumit eGeMAPS.

Märksõnad: arvutiparalingvistika, eGeMAPS, häälekorpus, hääle meeldivus, kõneakustika

Viimasel kümnendil on kõneuurimises toimunud suured muutused. Selle taga on eeskätt rahastuse ja arvutivõimsuse kasv (suurt hulka kõnet on võimalik töödelda tuhandetes kordades kiiremini kui varem). Ka on kõneuurimisel selged rakenduslikud eesmärgid. Suured firmad, nagu Google, Microsoft ja IBM on panustanud palju ressurssi kõnetöötluse tarkvarasse, ja mis oluline, palju sellest on kättesaadav vabavarana. Kõnetuvastuse ja kõnesünteesi ühisrakenduse heaks näiteks on Skype'i tõlkemoodul. Skype'i teel vesteldes on võimalik peaaegu reaalajas kõnet tõlkida kaheksasse keelde (inglise, prantsuse, saksa, hiina, itaalia, hispaania, portugali, araabia, vene).

Esilekerkinud valdkond – arvutiparalingvistika – keskendub kõneleja hääle analüüsile arvuti abil, et saada sealt infot kõneleja individuaalsete omaduste

ja seisundite kohta ning nende automaatsele tuvastusele ja genereerimisele (vt Schuller & Steidl *et al.* 2015).

Hääles on oluline info kõnelejast, tema püsivamatest omadustest ja hetkeseisunditest (vt Schuller & Batliner 2014: 23–24):

- püsivamad omadused: sugu, vanus, emakeel (L1), sotsiaalne staatus, isikuomadused (suur viisik: neurootilisus, ekstravertsus, avatus, sotsiaalsus, meelekindlus), meeldivus jne;
- keskmise püsivusega omadused-seisundid: sõbralikkus, positiivne/negatiivne suhtumine, unisus, tervislik seisund, joove, meeleolu (depressioon), huvi, viisakus jne;
- lühiajalised seisundid: kõnestiil ja hääle kvaliteet, emotsioonid (täismahulised, prototüüpsed) ning emotsioonilaadsed seisundid (stress, usaldus, ebakindlus, frustratsioon, valu jne).

Meile on antud võime suhtluses pidevalt analüüsida ja ümber hinnata vestluspartneri omadusi ja seisundeid ning kasutada seda infot partneri kavatsuste tõlgendamisel ja oma vestlusstrateegia kohandamisel. Sellist sotsiaalset kompetentsi enamikust tänapäeva häält kasutavatest tehnilistest lahendustest veel ei leia, kuid selleni jõudmine on seatud eesmärgiks (vt Schuller & Weninger 2012).

Alates 2010. aastast on rahvusvahelise kõnekommunikatsiooni assotsiatsiooni iga-aastasel konverentsil Interspeech üks sessioon pühendatud arvutiparalingvistikale – nn arvutiparalingvistika väljakutse (ingl *Computational Paralinguistics Challenge*). Selle eesmärk on tutvustada ja võrrelda erinevaid meetodeid mõne paralingvistilise nähtuse automaatses tuvastuses. Sessioonideks valmistumiseks antakse kõigile osaleda soovijaile kasutada üks ja sama kõnekorpus ning neil tuleb leida meetod ja valida akustiliste tunnuste komplekt, millega automaatselt klassifitseerida kõnelejad hääle järgi võimalikult õigesti etteantud rühmadesse. 2010. aastal oli üks ülesanne klassifitseerida kõnelejad hääle järgi nelja rühma: lapsed (7–14aastased), noored (15–24aastased), täiskasvanud (25–54aastased) ja seeniorid (55–80aastased). Võitjaks osutus uurijate rühm Brno Tehnoloogiaülikoolist, kelle kasutatud meetodil tuvastati vanused õigesti 52,4% juhtudest (Kockmann & Burget *et al.* 2010). Aastate jooksul on häälest tuvastatud väga mitmeid paralingvistilisi omadusi ja seisundeid (vt tabel 1).

Lisaks Interspeechi arvutiparalingvistika sessioonide ülesannetele on uuritud, kas ja kuidas kajastuvad hääles inimese pikkus ja kaal, stress, usaldatavus, depressioon, haridus jm (vt Schuller & Weninger 2012).

Tabel 1. Arvutiparalingvistika sessioonide temaatika Interspeechil.

Aasta	**Tuvastatavad omadused ja seisundid**
2010	vanus; sugu; huvitatuse tase (Schuller & Steidl *et al.* 2010)
2011	joove (üle või alla 0,5 promilli); unisus (Schuller & Steidl *et al.* 2011)
2012	isikuomadused (suur viisik); kõneleja meeldivus; patoloogilise kõne arusaadavus (Schuller & Steidl *et al.* 2012)
2013	sotsiaalsed signaalid (naer, ohkamine); konflikt; autism; emotsioonid (dimensioonidel negatiivne-positiivne ja aktiivne-passiivne) (Schuller & Steidl & Batliner & Vinciarelli *et al.* 2013)
2014	vaimne koormus; füüsiline koormus (treening/puhkus) (Schuller & Steidl *et al.* 2014)
2015	teise keele (L2) loomulikkuse tase; neuroloogiline seisund Parkinsoni haiguse korral; söömistingimused (kas ja mis tüüpi toitu süüakse) (Schuller & Steidl *et al.* 2015)
2016	emakeel L2 põhjal; valetamine; siirus (Schuller & Steidl *et al.* 2016)
2017	kõne adresseeritus (kas räägitakse lapse või täiskasvanuga); külmades tingimustes rääkimine; norskamise liigid (4) (Schuller & Steidl *et al.* 2017)

Et õpetada arvuteid häälest kõneleja omadusi ja seisundeid ära tundma, on vaja vastavaid kõnekorpusi. Paraku on avalikke realistlikke andmeid sisaldavaid hästi märgendatud ja kirjeldatud korpusi väga vähe (vt Schuller & Steidl & Batliner & Burkhardt *et al.* 2013). Enamik avalikke korpusi sisaldab ideaalsetes tingimustes salvestatud näideldud kõnet. Neist üks tuntumaid ja rahvusvaheliselt palju kasutust leidnud on Berliini emotsionaalse kõne andmebaas (Burkhardt & Paeschke *et al.* 2005). Kui aga õpetada arvutit emotsioone ära tundma näideldud kõne korpuse materjalil, siis tunnebki arvuti ära näideldud, mitte aga loomulikus spontaanses kõnes esiletulevaid emotsioone (vt Schuller & Steidl *et al.* 2009). Seetõttu on oluline, et treeningkorpuse kõnematerjal oleks võimalikult sarnane sellele, mida hakatakse kasutama loodavas rakenduses (Schuller & Batliner 2014: 25–29).

Kõnekorpuste loomine on aeganõudev. Tuleb otsustada, millist kõnematerjali koguda, kuidas, kellelt ja kui palju. Kogutud materjali paralingvistiliste tunnuste märgendamiseks on vaja kõnet segmenteerida (näiteks sõnadeks) ja läbi viia mitmesuguseid teste (näiteks lasta rühmal inimestel hinnata hääle meeldivust või teha testidega kindlaks kõneleja isikuomadused). Vajalik on metainformatsioon kõneleja kohta (vanus, sugu, haridus, pikkus, kaal, emakeel vms) ja materjali kohta (nt salvestuskoht, kõnestiil). Kõik korpusega seotu

peaks olema korralikult dokumenteeritud, et korpust saaksid kasutada mis tahes maade uurijad. Selliseid avalikke arvutiparalingvistikas kasutatavaid korpusi on äärmiselt vähe.

Paralingvistilised ilmingud võivad olla universaalsed või kultuurisõltlikud. Näiteks suulaelõhega laste hääl on samade akustiliste tunnustega kultuurideüleselt, kuid emotsioonide hääleline väljendumine erineb kultuuriti: mõnes kultuuris püütakse emotsioone vaos hoida, teises aga mitte (vt ka Schuller & Batliner 2014: 41–42). Uurimused on näidanud, et eesti keelt kuuldes ei pruugi muude kultuuride inimesed aru saada emotsioonidest, eriti kui need pole näideldud, vaid on igapäevased mõõdukalt väljendunud (Altrov & Pajupuu 2015). Seega läheb vaja oma kultuuri kohaseid korpusi.

Eestis on olemas emotsionaalse kõne korpus, mille avalik materjal sisaldab mõõdukalt väljendunud emotsioonidega lauseid. Lausete emotsioon on kuulamistestidega määratud nii kategooriatasandil (rõõm, kurbus viha, neutraalne) kui ka dimensioonidel negatiivne-positiivne, aktiivne-passiivne (Altrov & Pajupuu 2012). Korpuse materjali põhjal on tehtud uurimusi emotsioonide akustikast (Pajupuu, H. & Pajupuu *et al.* 2015), kuulaja vanuse ja empaatia ning kultuuri ja keele osatähtsusest emotsioonide äratundmisel (Altrov & Pajupuu 2010; Altrov & Pajupuu 2015; Altrov & Pajupuu *et al.* 2013). Korpust on kasutatud ka emotsionaalse kõne sünteesiks (Tamuri & Mihkla 2015) ning emotsioonide automaatse tuvastuse ühe treeningbaasina[1]. Kui emotsionaalse kõne korpuse materjal on kasutatav pigem kuulaja emotsioonitaju mõjutavate tegurite uurimiseks, siis kõneleja omaduste ja seisundite uurimiseks on vaja teistsuguse materjali ja märgendusega korpust.

Selles artiklis kirjeldame loodava häälekorpuse hetkeseisu ning võtame käsitleda kõneleja ühe paralingvistilise omaduse – hääle meeldivuse.

Häälekorpus

Häälekorpus on mõeldud peaasjalikult häälest kõneleja omadusi ja seisundeid äratundvate automaatsete klassifitseerijate loomiseks.

Korpusesse kogume erinevas eas mees- ja naishääli eri fonožanritest.[2] Praeguseks on korpuses 60 meeshäält vanuses 27–81 ja 50 naishäält vanuses 25–71. Salvestisi (22050 Hz, 16 bit, mono, eesti keel) on igalt kõnelejalt kaks: 3–5minutilised ja 5sekundilised katkendid Tallinna Tehnikaülikooli loengukorpusest (vt Meister, E. & Meister, L. & Metsvahi 2012) ja raadiosaadetest. Hääle metaandmeteks on kõneleja sugu, vanus, helilõigu fonožanr ning selle tunnused (vt tabel 2).

Tabel 2. Korpuse fonožanrite iseloomustus.

Fonožanr	**Ettevalmistusaste**	**Auditoorium**	**Esitusviis**
Raadiokommentaarid	ettevalmistatud	kaudne	monoloog
Loengud ja konverentsiettekanded	poolspontaanne	otsene	monoloog
Vestlussaated	spontaanne	kaudne	dialoog

Raadiokommentaarid on lühikesed arvamuslood, mida esitavad nii professionaalsed raadiotöötajad kui ka erinevate valdkondade spetsialistid, kellele raadios esinemine ei ole igapäevatöö. Kõneleja loeb teksti, mis on eelnevalt kirja pandud. Mõeldud on need suurele auditooriumile (raadiokuulajatele), kuid vahetu kontakt publikuga puudub. Tegemist on monoloogidega.

Loengud ja konverentsiettekanded on poolspontaansed, s.t kõneleja on teema ette valmistanud, kuid kannab seda ette vabalt. Mõeldud on nad suurele auditooriumile, kontakt publikuga on vahetu. Tegemist on monoloogidega.

Vestlussaadetest on korpusesse võetud saatekülalise kõne. Kõne on spontaanne – puudub ettevalmistatud tekst, mida järgida. Vestlussaated on suunatud suurele auditooriumile (raadiokuulajatele), kuid vahetu kontakt publikuga puudub. Tegemist on dialoogiga (vestluspartneriks on saatejuht).

Kuulamistestid hääle meeldivuse märgendamiseks

Hääle meeldivuse hindamiseks oleme läbi viinud kaks veebipõhist kuulamistesti. Ühes tuli kuulata ja hinnata 50 naishäält, teises 60 meeshäält, iga hääl 5 sekundit. Kõik kõnelõigud olid erinevad. Igat häält võis kuulata nii palju kordi, kui soovi oli. Hinnata tuli 7pallisel skaalal, kus 1 = *ei meeldi üldse* .... 7 = *meeldib väga*.

Hindajaid oli kokku 82:

- naised alla 35 a (*N*=17, vanus 24–34)
- naised üle 35 a (*N*=25, vanus 36–60)
- mehed alla 35 a (*N*=20, vanus 20–35)
- mehed üle 35 a (*N*=20, vanus 37–63)

Hindajate usaldusväärsuse (ingl *inter-rater reliability*) kindlakstegemiseks kasutasime intraklass korrelatsiooni (vrd Goy & Pichora-Fuller *et al.* 2016). Intraklass korrelatsiooni koefitsiendi (ICC2k) arvutasime iga hindajarühma jaoks nii nais- kui ka meeshäälte hindamisel. Mõlemas vanuserühmas olid nais- ja meeshindajate ICC-väärtused suuremad 0,8st, mis näitab, et igas rühmas käitusid tema liikmed hindamisel sarnaselt (vt tabel 3 ja 4).

Tabel 3. Nelja kuulajarühma intraklass korrelatsiooni koefitsiendid naishääle hindamisel.

Hindajate rühmad	Naishääle meeldivuse hindamine			
	ICC	Alumine piir	Ülemine piir	*p*
Naised alla 35 a	0,89	0,84	0,93	0,0001
Naised üle 35 a	0,93	0,89	0,95	0,0001
Mehed alla 35 a	0,90	0,85	0,94	0,0001
Mehed üle 35 a	0,91	0,87	0,94	0,0001

Tabel 4. Nelja kuulajarühma intraklass korrelatsiooni koefitsiendid meeshääle hindamisel.

Hindajate rühmad	Meeshääle meeldivuse hindamine			
	ICC	Alumine piir	Ülemine piir	*p*
Naised alla 35 a	0,88	0,83	0,92	0,0001
Naised üle 35 a	0,93	0,91	0,96	0,0001
Mehed alla 35 a	0,95	0,94	0,97	0,0001
Mehed üle 35 a	0,90	0,86	0,93	0,0001

Rühmadevahelise erinevuse kindlakstegemiseks hääle hindamisel kasutasime Pearsoni korrelatsiooni, mis näitas et rühmade hinnang hääle meeldivusele ei erinenud üksteisest oluliselt. Rühmadevaheline korrelatsioon oli > 0,80 (p<0,0001). Seega pidasid kõik rühmad meeldivamateks ja vähem meeldivamateks samu hääli.

Korpuses on iga hääle juures hääle meeldivuse testi tulemus ja andmed hindajate kohta (vanus, sugu). Anonüümsuse tagamiseks ei tehta meeldivuse hinnangut avalikkusele kättesaadavaks koos helifailiga, vaid heli asemel esitatakse hääle akustiliste tunnuste komplekt, mille järgi kõneleja ei ole tuvastatav.

Hääle meeldivuse akustilised tunnused

Meeldivat häält seostatakse usaldatavuse ja kompetentsiga (McAleer & Todorov *et al.* 2014; Nesler & Storr *et al.* 1993). Meeldiv hääl on vajalik mitmete elukutsete puhul (lektorid, poliitikud, müüjad, uudistelugejad, tugiteenuste töötajad), häält kasutavad ka erinevad tehnilised lahendused, nagu nutitelefonid, ekraanilugejad, e-raamatud, autod jm (vt ka Eyben & Weningen *et al.* 2013; Pinto-Coelho & Braga *et al.* 2013). Hääle meeldivust ja selle akustilisi tunnuseid on veel vähe uuritud (Schuller & Steidl *et al.* 2015).

Meie eesmärk oli kindlaks teha, millised hääled eestlastele meeldivad, millised on meeldivaid ja mittemeeldivaid hääli eristavad olulised akustilised tunnused ning katsetada hääle meeldivuse automaatset tuvastust.

Häälte akustiliseks analüüsiks kasutasime openSMILE'i tarkvara (Eyben & Weninger *et al.* 2013). Ekstraheerisime kõnest 88 parameetrit, mis moodustavad nii-nimetatud laiendatud Genfi minimaalse akustiliste parameetrite kogumi (eGeMAPS) (Eyben & Scherer *et al.* 2016). Need 88 parameetrit on eGeMAPSi võetud kolmel põhjusel: (1) potentsiaal eristada emotsioonidest tingitud füsioloogilisi muutusi hääles; (2) tulemuslikkus senistes uuringutes ja automaatne ekstraheeritavus; (3) teoreetiline tähtsus. Parameetrid grupeeruvad nelja rühma: sagedusega seotud parameetrid, energia ja amplituudiga seotud parameetrid, spektriparameetrid ja tempo parameetrid. eGeMAPSi on soovitatud kasutada automaatses hääle analüüsis, nagu kõne paralingvistika analüüs või kliiniline analüüs (vt Eyben & Scherer *et al.* 2016).

Meeldivaid ja mittemeeldivaid hääli eristavate parameetrite leidmiseks kasutasime ANOVAt.

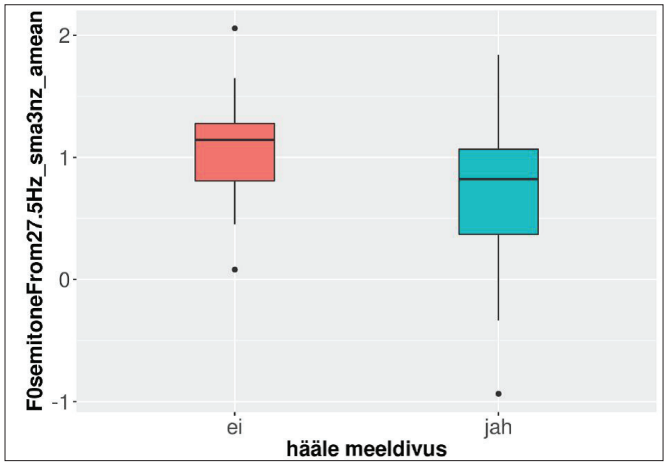
ANOVA põhjal osutusid 88st eGeMAPSi parameetrist meeldivaid ja mittemeeldivaid naishääli oluliselt eristavateks seitse, mis kuulusid kahte rühma: põhitooniga seotud parameetrid ning energia ja amplituudiga seotud parameetrid (vt tabel 5).

Tabel 5. Hääle meeldivuse olulised akustilised parameetrid naishäälte puhul ANOVA tulemuste põhjal.

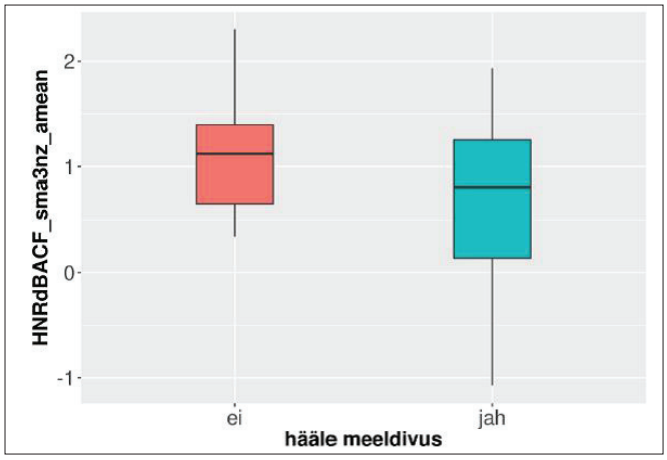
eGeMAPSi parameetrid	**Kirjeldus**	***F*-statistik**
Sagedusega seotud parameetrid		
F0semitoneFrom27.5Hz_sma3nz_amean	keskmine põhitooni sagedus pooltoonides	4,9*
F0semitoneFrom27.5Hz_sma3nz_percentile20.0	põhitooni 20. pertsentiil	6,1*
F0semitoneFrom27.5Hz_sma3nz_percentile50.0	põhitooni 50. pertsentiil	5,4*
F0semitoneFrom27.5Hz_sma3nz_stddevFallingSlope	põhitooni langeva osa kalde standardhälve	5,7*
Energiaga/Amplituudiga seotud parameetrid		
HNRdBACF_sma3nz_amean	harmooniliste ja müra energia suhte keskmine	5,5*
HNRdBACF_sma3nz_stddevNorm	harmooniliste ja müra energia suhte normaliseeritud standardhälve	4,3*
shimmerLocaldB_sma3nz_amean	hääletugevuse võbelemise keskmine	5,1*

Märkus. * $p < 0,05$.

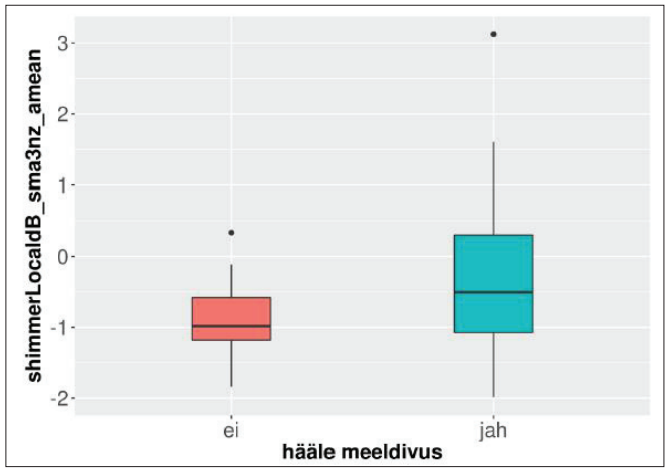
Paraku on eGeMAPSi paljusid parameetreid raske seostada tajutavate kõneomadustega. Naishäälte puhul on lihtsamini tõlgendatavaid kolm: keskmine põhitooni sagedus, kähedus ja hääletugevuse võbelemine[3] (ingl *shimmer*). Kuulajatele meeldisid rohkem madalamad ja vähem kähedad naishääled (vt joonised 1–2). Hääletugevuse võbelemine lisas meeldivust (vt joonis 3).



Joonis 1. Meeldivate ja mittemeeldivate naishäälte põhitooni sageduse keskmine normaliseeritud skaalal.



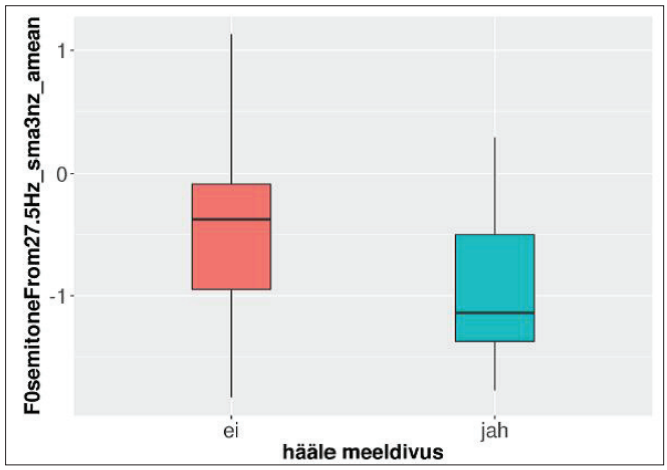
Joonis 2. Meeldivate ja mittemeeldivate naishäälte kähina keskmine normaliseeritud skaalal.



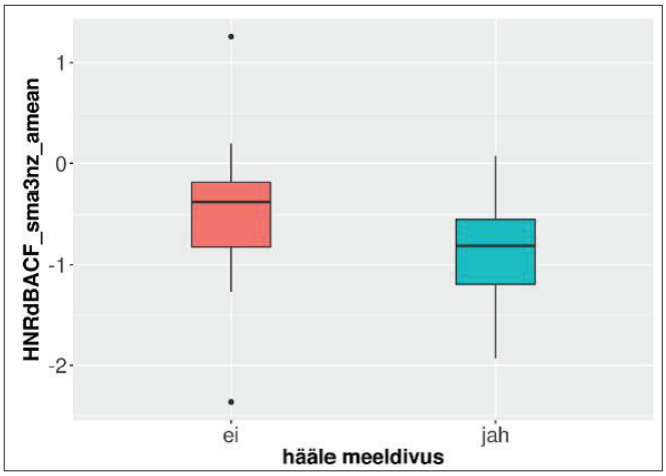
Joonis 3. Hääletugevuse võbelemise keskmine normaliseeritud skaalal.

Meeshäälte puhul oli meeldivaid ja mittemeeldivaid hääli eristavaid parameetreid rohkem – 18. Need olid sageduse, energia ja amplituudi ning spektriga seotud parameetrid (vt tabel 6). Eristavate hulgas ei olnud tempo parameetreid.

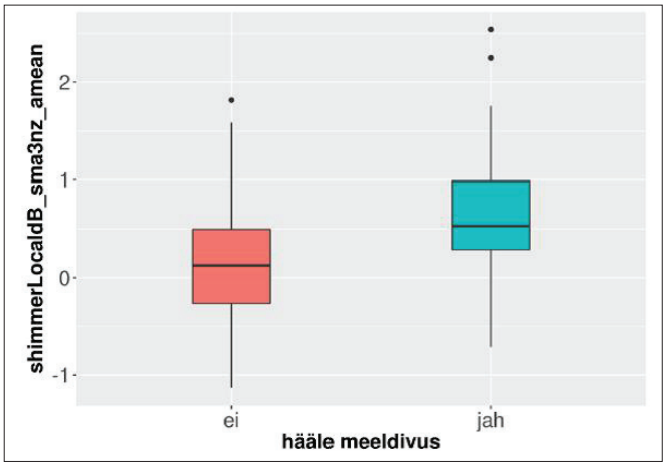
Nii nagu naishäälte puhul, meeldivad ka meeshäälte puhul enam madalamad ja vähem kähedad hääled. Sarnane on ka hääletugevuse võbelemise mõju: võbelevad hääled meeldisid rohkem. Erinevalt naishäältest, osutus meeshäälte puhul oluliseks hääletugevus. Valjemad hääled meeldisid vähem (vt joonised 4–7).



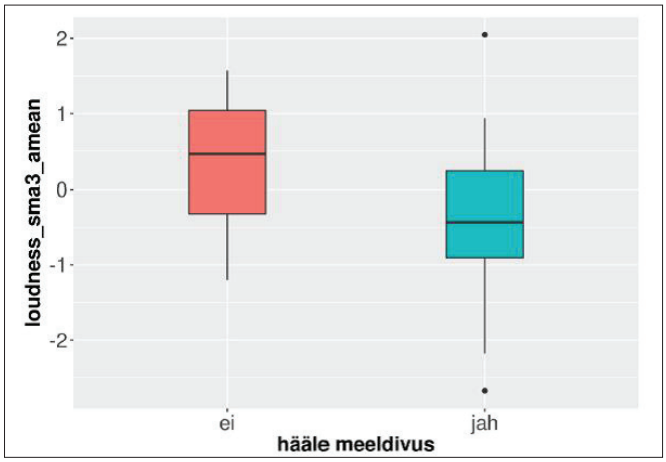
Joonis 4. Meeldivate ja mittemeeldivate meeshäälte põhitooni sageduse keskmine normaliseeritud skaalal.



Joonis 5. Meeldivate ja mittemeeldivate meeshäälte kähina keskmine normaliseeritud skaalal.



Joonis 6. Hääletugevuse võbelemise keskmine normaliseeritud skaalal.



Joonis 7. Hääletugevuse keskmine normaliseeritud skaalal.

Tabel 6. Hääle meeldivuse olulised akustilised parameetrid meeshäälte puhul ANOVA tulemuste põhjal.

eGeMAPSi parameetrid	**Kirjeldus**	***F*-statistik**
Sagedusega seotud parameetrid		
F0semitoneFrom27.5Hz_sma3nz_amean	keskmine põhitooni sagedus pooltoonides	10,4**
F0semitoneFrom27.5Hz_sma3nz_percentile20.0	põhitooni 20. pertsentiil	16,5****
F0semitoneFrom27.5Hz_sma3nz_percentile50.0	põhitooni 50. pertsentiil	10,7**
F0semitoneFrom27.5Hz_sma3nz_percentile80.0	põhitooni 80. pertsentiil	6,6*
F3frequency_sma3nz_stddevNorm	F3 normaliseeritud standardhälve	8,3**
Energiaga/amplituudiga seotud parameetrid		
shimmerLocaldB_sma3nz_amean	hääletugevuse võbelemise keskmine	4,5*
HNRdBACF_sma3nz_amean	harmooniliste ja müra energia suhte keskmine	6,4*
loudness_sma3_amean	keskmine hääletugevus	10,5**
loudness_sma3_pctlrange0.2	hääletugevuse võõrväärtused	6,7*
loudness_sma3_percentile50.0	hääletugevuse 50. pertsentiil	6,5*
loudness_sma3_percentile80.0	hääletugevuse 80. pertsentiil	10,7**
Spektri parameetrid		
mfcc2V_sma3nz_amean	MFCC heliliste segmentide 2. koefitsiendi keskmine	4,6*
mfcc4_sma3_amean	MFCC helitute segmentide 4. koefitsiendi keskmine	14,5***
mfcc4V_sma3nz_amean	MFCC heliliste segmentide 4. koefitsiendi keskmine	13,9***
mfcc2_sma3_amean	MFCC helitute segmentide 2. koefitsiendi keskmine	4,7*
slopeV0.500_sma3nz_amean	spektri 0–500 Hz piirkonna võimsuse regressioonikoefitsiendi keskmine	8,1**
slopeV500.1500_sma3nz_amean	spektri 500–1500 Hz piirkonna võimsuse regressioonikoefitsiendi keskmine	4,5*
slopeV500.1500_sma3nz_stddevNorm	spektri 500–1500 Hz piirkonna võimsuse regressioonikoefitsiendi standardhälve	7,6**

Märkus. * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$, **** $p < 0,0001$.

Hääle meeldivuse automaatne tuvastamine

Akustiline analüüs näitas, et eGeMAPSi parameetritega on võimalik eristada meeldivaid ja mittemeeldivaid hääli, kuid nais- ja meeshäälte puhul kattuvad need parameetrid vaid osaliselt. Seega tuleks hääle meeldivuse automaatseks tuvastuseks esmalt kindlaks teha, kas tegu on nais- või meeshäälega.

Hääle meeldivuse automaatsele tuvastusele oli pühendatud 2012. aasta Interspeechi arvutiparalingvistika sessioon (Schuller & Steidl *et al.* 2012). Selles osalejatele oli antud kasutada korpus, mis sisaldas 800 eri vanuses mehe ja naise telefonihäält (digitaliseeritud sagedusel 8 kHz), iga kõneleja kohta üks lause. Häälte meeldivus oli hinnatud seitsmepallisel skaalal. Samuti oli kasutada openSMILE'i tunnuste ekstraheerija, mis võimaldas kõnest kätte saada 6125 tunnust. Eesmärk oli teada saada, millise meetodi ja tunnustega on võimalik saada parimaid tulemusi häälte klassifitseerimisel meeldivaiks ja mittemeeldivaiks. Selle ülesande lahendamises osales kümme uurijarühma. Parim tulemus saadi SVM-klassifitseerijaga (ingl *Support Vector Machine*) – 65,8% (vt Montacié & Caraty 2012; Schuller & Steidl *et al.* 2015).

Meie kasutasime tuvastuses minimaalset tunnuste kogumit eGeMAPS ning katsetasime SVM-klassifitseerijat (vt Chang & Lin 2011). Materjal koosnes 60 mees- ja 50 naishäälest, mis kuulamistesti tulemusel olid märgendatud meeldivaks ja mittemeeldivaks vastavalt sellele, kas hääle keskmine hinne oli üle või alla kõigi häälte keskmise hinde.

Esmalt klassifitseerisime hääled mees- ja naishäälteks. Klassifitseerimistäpsuseks saime 93%.

Hääle meeldivuse automaatseks klassifitseerimiseks võtsime kummastki rühmast juhuslikult 75% treeninghäälteks ja jätsime 25% kontrolliks ning treenisime SVM-mudeli. Kuna andmeid oli vähe, siis selleks, et saada realistlikum hinnang, kordasime kogu protseduuri sada korda. Mudeli keskmiseks täpsuseks saime meestel 64% ja naistel 58%. Need esmased tulemused näitavad, et siit on võimalik edasi minna: suurendada korpust, tegelda tunnuste valikuga ja proovida ka teisi meetodeid.

Kokkuvõte

Kõneleja omaduste ja seisundite automaatne tuvastus häälest on tõusnud arvutiparalingvistika keskseks teemaks. On hulk kõnetehnoloogilisi rakendusi, kus kõneleja klassifikatsiooni võiks kasutada. Näiteks, klienditoe kõnekeskustes suunata klient häälest automaatselt tuvastatud omaduste või seisundite põhjal sobiva profiiliga teenindajale või kohandada nende tunnuste järgi kliendiga

käitumist: valida sobiv kõnetempo teise emakeelega või vanade inimestega rääkides, olla valmis suhtlema vihase kliendiga jne. Eestis on arvutiparalingvistikaga tegeldud veel vähe, kuna puuduvad mudelite treeninguks vajalikud märgendatud kõnekorpused. Artiklis andsime lühiülevaate sellest, milliseid kõneleja omadusi ja seisundeid on proovitud häälest tuvastada ja milliseid kõnekorpusi selleks vajatakse. Kirjeldasime Eestis loodavat häälekorpust ja demonstreerisime selle materjalil hääle meeldivuse tuvastust. Edaspidised tööd keskenduvad korpuse laiendamisele, sest ilma mitmekesiselt märgendatud kõnekorpusteta on arvutiparalingvistika võimatu.

Tänusõnad

Uurimust on toetanud Euroopa Liit Euroopa Regionaalarengu Fondi kaudu (Eesti-uuringute Tippkeskus), see on seotud Eesti Haridus- ja Teadusministeeriumi uurimisprojektiga IUT 35-1.

Kommentaarid

[1] https://github.com/EKT1/emotional

[2] Fonožanr – situatsioonist sõltuv kõnestiil.

[3] Hääletugevuse võbelemine – hääletugevuse kiire perioodiline muutumine.

Kirjandus

Altrov, Rene & Pajupuu, Hille 2010. Estonian Emotional Speech Corpus: Culture and age in selecting corpus testers. *Frontiers in Artificial Intelligence and Applications*, 219: *Human Language Technologies – The Baltic Perspective*, lk 25–32 (doi: 10.3233/978-1-60750-641-6-25).

Altrov, Rene & Pajupuu, Hille 2012. Estonian Emotional Speech Corpus: theoretical base and implementation. Devillers, Laurence & Schuller, Björn & Batliner, Anton & Rosso, Paolo & Douglas-Cowie, Ellen & Cowie, Roddy & Pelachaud, Catherine (toim). *The 4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals (ES3)*, lk 50–53.

Altrov, Rene & Pajupuu, Hille & Pajupuu, Jaan 2013. The role of empathy in the recognition of vocal emotions. *Interspeech 2013*. 14th Annual Conference of the International Speech Communication Association, Lyon, France, lk 1341–1344.

Altrov, Rene & Pajupuu, Hille 2015. The influence of language and culture on the understanding of vocal emotions. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics* 6 (3), lk 11–48 (doi: 10.12697/jeful.2015.6.3.01).

Burkhardt, Felix & Paeschke, Astrid & Rolfes, Miriam & Sendlmeier, Walter & Weiss, Benjamin 2005. A Database of German Emotional Speech. *Interspeech 2005*, lk 1517–1520.

Chang, Chih-Chung & Lin, Chih-Jen 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2 (3), nr 27 (doi: 10.1145/1961189.1961199).

Eyben, Florian & Scherer, Klaus & Schuller, Björn & Sundberg, Johan & Andre, Elisabeth & Busso, Carlos & Devillers, Laurence & Epps, Julien & Laukka, Petre & Narayanan, Shikanth & Truong, Khiet 2016. The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing* 7 (2), lk 190–202 (doi: 10.1109/TAFFC.2015.2457417).

Eyben, Florian & Weninger, Felix & Marchi, Erik & Schuller, Björn 2013. Likability of human voices: A feature analysis and a neural network regression approach to automatic likability estimation. *Proceedings of the 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, lk 1–4 (doi: 10.1109/WIAMIS.2013.6616159).

Goy, Huiwen & Pichora-Fuller, Kathleen M. & Lieshout, Pascal van 2016. Effects of age on speech and voice quality ratings. *Journal of the Acoustical Society of America* 139 (4), lk 1648–1659 (doi: 10.1121/1.4945094).

Kockmann, Marcel & Burget, Lukáš & Černocký, Jan 2010. Brno university of technology system for Interspeech 2010 paralinguistic challenge. *Interspeech 2010.* 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26–30, lk 2822–2825.

McAleer, Phil & Todorov, Alexander & Belin, Pascal 2014. How do you say "hello"? Personality impressions from brief novel voices. *PLoS ONE* 9 (3), lk 1–10. (doi: 10.1371/journal.pone.0090779).

Meister, Einar & Meister, Lya & Metsvahi, Rainer 2012. New speech corpora at IoC. Meister, Einar (koost). *XXVII Fonetiikan päivät 2012 = Phonetics Symposium 2012: 17–18 February 2012.* Tallinn: TUT Press, lk 30–33.

Montacié, Claude & Caraty, Marie-José 2012. Pitch and intonation contribution to speakers' traits classification. *Interspeech 2012*. 13th Annual Conference of the International Speech Communication Association in Portland, Oregon, lk 526–529.

Nesler, Mitchell S. & Storr, Dawn M. & Tedeschi, James T. 1993. The Interpersonal Judgment Scale: A measure of liking or respect? *The Journal of Social Psychology* 133 (2), lk 2237–2242 (doi: 10.1080/00224545.1993.9712141).

Pajupuu, Hille & Pajupuu, Jaan & Tamuri, Kairi & Altrov, Rene 2015. Influence of verbal content on acoustics of speech emotions. *Proceedings of the 18th International Congress of Phonetic Sciences.* The Scottish Consortium for ICPhS 2015. Glasgow, UK: The University of Glasgow, lk 1–5 (https://www.researchgate.net/publication/281004592_Influence_of_verbal_content_on_acoustics_of_speech_emotions – 4. oktoober 2017).

Pinto-Coelho, Luis & Braga, Daniela & Sales-Dias, Miguel & Garcia-Mateo, Carmen 2013. On the development of an automatic voice pleasantness classification and intensity estimation system. *Computer Speech and Language* 27 (1), lk 75–88 (doi: 10.1016/j.csl.2012.01.006).

Schuller, Björn & Batliner, Anton 2014. *Computational Paralinguistics. Emotion, Affect and Personality in Speech and Language Processing.* John Wiley & Sons, Ltd.

Schuller, Björn & Weninger, Felix 2012. Ten recent trends in computational paralinguistics. Esposito, Anna & Esposito, Antonietta M. & Vinciarelli, Alessandro & Hoffmann, Rüdiger & Müller, Vincent C. (toim). *4th COST 2102 International Training School on Cognitive Behavioural Systems* 7403/2012, lk 35–49 (doi: 10.1007/978-3-642-34584-5_3).

Schuller, Björn & Steidl, Stefan & Batliner, Anton 2009. The Interspeech 2009 Emotion Challenge. *Interspeech 2009*, lk 312–315 (http://emotion-research.net/sigs/speech-sig/emotion-challenge/INTERSPEECH-Emotion-Challenge-2009_draft.pdf – 4. oktoober 2017).

Schuller, Björn & Steidl, Stefan & Batliner, Anton & Nöth, Elmar & Vinciarelli, Alessandro & Burkhardt, Felix & Son, Rob van & Weninger, Felix & Eyben, Florian & Bocklet, Tobias & Mohammadi, Gelareh & Weiss, Benjamin 2015. A survey on perceived speaker traits: Personality, likability, pathology, and the first challenge. *Computer Speech and Language* 29 (1), lk 100–113 (doi: 10.1016/j.csl.2014.08.003).

Schuller, Björn & Steidl, Stefan & Batliner, Anton & Burkhardt, Felix & Devillers, Laurence & Müller, Christian & Narayanan, Shrikanth S. 2010. The Interspeech 2010 paralinguistic challenge. *Interspeech 2010*, lk 2794–2797 (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.331.6236&rep=rep1&type=pdf – 4. oktoober 2017).

Schuller, Björn & Steidl, Stefan & Batliner, Anton & Schiel, Florian & Krajewski, Jarek 2011. The Interspeech 2011 speaker state challenge. *Interspeech 2011*, lk 3201–3204 (http://emotion-research.net/sigs/speech-sig/The%20INTERSPEECH%202011%20Speaker%20State%20Challenge.pdf/ – 4. oktoober 2017).

Schuller, Björn & Steidl, Stefan & Batliner, Anton & Nöth, Elmar & Vinciarelli, Alessandro & Burkhardt, Felix & Son, Rob van & Weninger, Felix & Eyben, Florian & Bocklet, Tobias & Mohammadi, Gelareh & Weiss, Benjamin 2012. The Interspeech 2012 speaker trait challenge. *Interspeech 2012*, lk 254–257 (http://emotion-research.net/sigs/speech-sig/IS2012-Speaker-Trait-Challenge.pdf/ – 4. oktoober 2017).

Schuller, Björn & Steidl, Stegan & Batliner, Anton & Burkhardt, Felix & Devillers, Laurencce & Müller, Christian & Narayanan, Shrikanth 2013. Paralinguistics in speech and language – State-of-the-art and the challenge. *Computer Speech and Language* 27 (1), lk 4–39 (doi: 10.1016/j.csl.2012.02.005).

Schuller, Björn & Steidl, Stefan & Batliner, Anton & Vinciarelli, Alessandro & Scherer, Klaus & Ringeval, Fabien & Chetouani, Mohamed & Weninger, Felix & Eyben, Florian & Marchi, Erik & Mortillaro, Marcello & Salamin, Hugues & Polychroniou, Anna & Valente, Fabio & Kim, Samuel 2013. The Interspeech 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. *Interspeech 2013*, lk 148–152 (http://emotion-research.net/sigs/speech-sig/is2013_compare.pdf – 4. oktoober 2017).

Schuller, Björn & Steidl, Stefan & Batliner, Anton & Epps, Julien & Eyben, Florian & Ringeval, Fabien & Marchi, Erik & Zhang, Yue 2014. The Interspeech 2014 computational paralinguistics challenge: cognitive & physical load. *Interspeech 2014*, lk 427–431 (http://emotion-research.net/sigs/speech-sig/is2014_compare.pdf – 4. oktoober 2017).

Schuller, Björn & Steidl, Stefan & Batliner, Anton & Hantke, Simone & Hönig, Florian & Orozco-Arroyave, J. R. & Nöth, Elmar & Zhang, Yue & Weninger, Felix 2015. The Interspeech 2015 computational paralinguistics challenge: nativeness, Parkinson's & eating condition. *Interspeech 2015*, lk 478–482 (http://emotion-research.net/sigs/speech-sig/is2015_compare.pdf – 4. oktoober 2017).

Schuller, Björn & Steidl, Stefan & Batliner, Anton & Hirschberg, Julia & Burgoon, Judee K. & Baird, Alice & Elkins, Aaron & Zhang, Yue & Coutinho, Eduardo & Evanini, Keelan 2016. The Interspeech 2016 computational paralinguistics challenge: Deception, sincerity & native language. *Interspeech 2016*, lk 2001–2005 (doi: 10.21437/Interspeech.2016-129).

Schuller, Björn & Steidel, Stefan & Batliner, Anton & Bergelson, Elika & Krajewski, Jarek & Janott, Christoph & Amatuni, Andrei & Casillas, Marisa & Seidl, Amanda & Soderstrom, Melanie & Warlaumont, Anne S. & Hidalgo, Guillerma & Schnieder, Sebastian & Heiser, Clemens & Hohenhorst, Winfried & Herzog, Michael & Schmitt, Maximilian & Qian, Kun & Zhang, Yue & Trigeorgis, George & Tzirakis, Panagiotis & Zafeiriou, Stefanos 2017. The Interspeech 2017 Computational paralinguistics challenge: Addressee, cold & snoring. *Interspeech 2017*, lk 3442–3446 (doi: 10.21437/Interspeech.2017-43).

Tamuri, Kairi & Mihkla, Meelis 2015. Expression of basic emotions in Estonian parametric text-to-speech synthesis. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics* 6 (3), lk 145–168 (doi: 10.12697/jeful.2015.6.3.06).

Summary

Computational paralinguistics challenges and Estonian voice likability

Hille Pajupuu
leading researcher, Institute of the Estonian Language
eki@eki.ee

Jaan Pajupuu
Software developer
eki@eki.ee

Rene Altrov
researcher, Institute of the Estonian Language
eki@eki.ee

Keywords: computational paralinguistics, eGeMAPS, speech acoustics, speech corpora, voice likability

This article looks into tendencies of the last decade in computational paralinguistics: ascertaining of speaker traits and states in voice, and the requirements set for the related speech corpora. It introduces the Estonian voice corpus and the ability to acoustically characterize voice likability and identify it automatically, using the expanded Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) for voice research and affective computing.