

Varieeruva vältega sõnad: häälduseelistused ja määramisraskused

Mari-Liis Kalvik

Eesti Keele Instituudi vanemleksikograaf ja
eksperimentaalfoneetika teadur
Mari-Liis.Kalvik@eki.ee

Liisi Piits

Eesti Keele Instituudi teadur kõnetehnoloogia alal
Liisi.Piits@eki.ee

Teesid: Artikli üks eesmärke on välja selgitada vältesõnade hääldustava, aga selle kõrval on vaatluse all varieeruva vältega sõnade välde tajumise ja määramise problemaatika.

Analüüsitakse 50 keelejuhilt kogutud 49 varieeruva vältega sõna 2438 hääldusjuhtu, mille välde hindas vähemalt kaks inimest, lahkneva hinnangu puhul kaasati kolmas vältemääraja. Esimese ja teise vältemääraja hinnanguid võrreldes selgub, et kõigi 2438 hääldusjuhu puhul lahkesid kahe määraja hinnangud 375 korral ehk 15% juhtudest.

Tegime kindlaks, et sõnade *meetod* ja *keiserlik* häälduse hindamisel oli kõige rohkem erimeelsusi.

Varieeruva vältega sõnade häälduseelistusi uurides võib 42 sõna puhul välja tuua häälduseelistuse, ainult seitsme sõna puhul ei õnnestunud domineerivat välde välja selgitada.

Märksõnad: fonoloogiline varieerumine, välte varieerumine, välte taju, lugemiseksperiment, eesti keel

Sissejuhatus

Selle artikli lähteks on fonoloogilise varieerumise uurimisel, täpsemalt varieeruva vältega sõnade käsitlemisel esile kerkinud küsimused. Artiklis otsime neile küsimustele vastust, kirjeldades ühtlasi oma uurimust laiema. Esiteks huvitab meid, millise varieerumisega tuleb kokku puutuda välte kuuldelisel määramisel – st kuulamise järgi sõnade kohta välteotsustuste tegemisel. Põhiküsimus on siin, kuidas saada võimalikult objektiivseid tulemusi varieeruva vältega

sõnade hääldusjuhtude hindamisel. Tahame teada, kui palju sõltub kuuldelse hinnangu andmisel välde määrajast ja kui palju esialgseid vältelhinnanguid hindamise käigus muudetakse. Nendele küsimustele vastuste leidmine aitab jõuda meie uurimuse ühe peaesmärgini, milleks on vahelduva vältega sõnade häälduseelistuse kindlaks tegemine ning saadud tulemuste rakendamine sünteeskõnes.

Lugemiseksperimenti kirjeldus

Käesolev artikkel on osa laiemast uurimusest, mille eesmärk on analüüsida häälduse fonoloogilist varieerumist. Meid huvitavad kolm nähtust: välte vaheldumine sõnades, mille puhul on lubatud nii teise- kui ka kolmandavälte-line hääldus, *h* hääldamine sõnaalgulises positsioonis ning palatalisatsioon *i*-tüvelistes ühesilbilistes sõnades pika vokaali järel asuvates konsonantides. Varieerumise võimalike põhjuste ja mõjutajatena uurime nii lingvistilisi, sõnastruktuurist ja esinemissagedusest tulenevaid kui ka kõnelejust tulenevaid (lugeja vanus, päritolupiirkond, haridus) põhjuseid.

Alates 2015. aastast oleme kogunud materjali lugemiseksperimentidega, mille käigus tuleb lugejal ette lugeda 53 lauset. Need laused sisaldavad 144 sihtsõna kõigi kolme meid huvitava nähtuse uurimiseks. Uuritavad sihtsõnad on paigutatud võimalikult loomulikku lausekonteksti. Allolevates näitelausestes on varieeruva vältega sihtsõnad paksus kirjas.

- (1) *Ka **soodus** asi võib olla **maitsekas** või heldima panev.*
- (2) ***Toimekas** sepp tagus **kangelase** mõõka, kuni ääs hõõgus.*
- (3) ***Tallinlane** on **kaitsetu**, kui ta **hilja** õhtul pimedas ringi kõnnib.*

Põhjalikumalt on kogu eksperimenti ning sellesse valitud materjali kirjeldatud artiklis “Lugemiseksperiment fonoloogilise varieerumise uurimiseks” (Kalvik & Piits 2015).

2017. aasta sügiseks oleme salvestanud 151 inimese poolt ettelõetud laused, millest kolmandik on ka segmenteeritud ja vaatlusalused sõnad märgendatud. Salvestusi oleme seni teinud Alatskivil, Karksi-Nuias, Meremäel, Märjamaal, Palamusel, Põlvas, Pürksis, Tallinnas, Tartus, Tõrvas, Vainupeal, Võrus jt Eesti paikades, otsides eeskätt keelejuhte, kes oleksid üles kasvanud kohas, kust on pärit ka nende vanemad. Sellise valiku alusel saame ühtlasi edaspidi uurida inimese päritolupiirkonna ja häälduse seoseid.

Salvestamine ja segmenteerimine

Salvestused jagunevad kaheks: stuudiosalvestused Tallinnas ja Tartus ning salvestused välitingimustes erinevais Eesti paigus. Sellest tulenevalt on kasutatud ka erinevat aparatuuri. Eesti Keele Instituudi stuudios kasutame helipulti Mackie Onyx 1640i ja mikrofoni Neumann TLM 102, Tartu ülikooli foneetikalabori stuudios Sound Devices USBPre helikaarti ja mikrofoni Beyerdynamic MC 930. Stuudiosalvestused on tehtud suletud salvestusruumis, kus keelejuht luges arvutiekraanile ükshaaval ilmuvaid lauseid. Lugemiseksperimentide läbiviimiseks kasutame Eesti Keele Instituudis välja töötatud programmi “Korpussalvestuste ja indekseerimise kiosk” (KJIK), mis võimaldab salvestajal salvestusprotsessi juhtida: kuvada lugejale sobivas tempos ette lauseid ja lugemistõrgete korral lauseid uuesti salvestada. Programm salvestab kõik laused erinevate helifailidena, kusjuures lause uuesti lugemisel jääb alles uusim variant. Väljaspool stuudiot salvestades oleme valinud ettelugemiseks võimalikult vaikse ruumi. Välitöödel oleme kasutanud kas ainult diktofoni Zoom H4 või Zoom H1 (keelejuht loeb lauseid paberilt) või sülearvutit ja KJIK programmi (keelejuht loeb lauseid arvutiekraanilt), kusjuures diktofon on sel juhul kasutuses mikrofoni ja helikaardina.

Kõik eksperimendis osalejad täidavad ankeedi, milles on küsimused soo, vanuse, hariduse ja elukoha kohta (märkida tuleb nii praegune elukoht kui ka peamine elukoht esimese kümne eluaasta jooksul). Lisaks tuleb kirja panna vanemate päritolupiirkond, mille juhul, kui vanemad on pärit samast murdepiirkonnast, seome murdealaga. Isikuandmete kaitsest tulenevalt on iga katses osaleja hilisemas uurimistöös tähistatud numbrikombinatsiooniga. Ankeeti allkirjastades annab eksperimendis osaleja nõusoleku salvestist kasutada analüüsiks, kuulamistestides ja kõnekorpuste koosseisus.

Salvestused on segmenteeritud sõna- ja häälikutasandil eestikeelse kõne autosegmenteerija (<http://www.phon.ioc.ee/dokuwiki/doku.php?id=projects:tuvastus:est-align.et>) abil. Programmis Praat (Boersma & Weenink 2015) on automaatse märgendusega sõna- ja häälikutasandi kõrvale loodud iga uuritava nähtuse (palatalisatsiooni, välte ja sõnaalgulise *h*) jaoks lisatasand. Kõigi analüüsi haaratud salvestuste märgendused on üle kontrollitud ning autosegmenteerija loodud piire on vajadusel käsitsi kohendatud. Varieeruva vältega sihtsõnade puhul märgime vastavale segmentatsioonitasandile kestussuhete automaatseks arvutamiseks pearõhulise silbi tuuma (V1) ja kooda olemasolul ka pearõhulise silbi kooda (C1) ning järgneva silbi tuuma (V2).

Varieeruva vältega sõnad

Varieeruva vältega sõnade määratlemisel oleme lähtunud õigekeelsussõnaraamatu (ÕS 2013) normingutest. ÕSis esineb 295 sõna, mille alg- ehk sõnastikuvormi ja muutevorme võib hääldada nii teises kui ka kolmandas vältes. Kui muutmistüüp jääb mõlema häälduse puhul samaks, siis on sõnastikuvormis vältemärk esitatud sulgudes (nt (^)andekas <2: -ka, -kat> anderikas), aga kui muutmistüüp sõltub vältest, siis on paralleelvormid esitatud eraldi märksõnadena (nt ihne <6: `ihne, ihnet>, `ihne <1: `ihne, `ihnet> kitsi). Selliseid sõnu peamegi varieeruva vältega sõnadeks ja sagedamad neist on valitud sihtõnadena meie uurimusse. Uurimuses on vaatluse all järgmised 49 sõna: *ahne, alles, ammu, andekas, augustis, aula(sse), eile, enne, haiguslik, hilja, hirmus, ihne, ilma, ilmetu, jaanuar, kaine, kaitsetu, kangelase, keiserlik, kiire, kirju, kuulus, karme, looduslik, maitsekas, meetod, mullu, nooruslik, nõiduslik, pealik, peenelt, piknikuga, politseinik, rahvuslik, rõõmus, saatanlik, saatuslik, soodus, tallinnlane, teaduslikes, teisal, terve, toimekas, täiuslik, ungarlane, võimekas, võimetu, äärmuslikult ja ümbrikus*. Sõnad on esitatud samas käändevormis, milles nad lausetes esinevad. Suurim osa sõnu esineb nimetavas käändes, aga on ka omastavas, sisseütlevas, seesütlevas ja kaasaütlevas käändes sõnu. Sõna *aula* puhul on sisseütleva käände lõpp esitatud sulgudes, kuna see sõna esineb olenevalt lugemistekstist, mida salvestuste käigus kohendasime, nii omastavas kui ka sisseütlevas käändes. Nagu loendist näha, on enamjaolt tegemist käändsõnadega, lisaks on määrsõnad *alles, ammu, eile, enne, hilja, mullu, peenelt, teisal, äärmuslikult* ning kaassõna *ilma*.

Lisaks sisaldab ÕS 2013 veel ligi 20 sõna (nt tegusõnad *kustuma* ja *säilima* või käändsõnad *peen* ja *luule*), mille sõnastikuvormi hääldus ei varieeru, aga mille mõnd muutevormi saab hääldada nii teises kui ka kolmandas vältes. Need üksikute muutevormide häälduskõikumised on praegusest uurimusest välja jäänud. Kindlasti on varieerumine ka suurem kui ÕSis sätestatud juhtudel. Näiteks on küllalt levinud õigekeelsussõnaraamatu järgi vaid teisevälteliste sõnade *pilved, taimed, jõulud, räimed* kolmandavältiline hääldus. Kuid kuna meie eesmärk on välja selgitada eelkõige nende sõnade hääldustava, mis sõnaraamatus on saanud kahesuguse vältemäärangu, siis on kindlalt teise- või kolmandavältiliseks normitud sõnad jäänud uurimusest välja.

Varieeruva vältega sõnu on jagatud raskuskategooriasse selle järgi, kui võrd palju mõjutab valde seda, kuidas sõna morfoloogilises süsteemis käitub (Hint 1968). Normimise seisukohast tekitavad kõige vähem probleeme sellised paralleelvormid, mille vältest ei sõltu muutmistüüp (nt muutumatud sõnad *enne, ammu* ja *hilja* või astmevahelduseta käändsõnad *kaitsetu, andekas* ja *looduslik*). Kõige rohkem tekitavad probleeme paralleelvormid, mille vältest sõltuvad nii

muuttuüp kui ka muutelõpud (nt *pealik*, *soodus* ja *aula*). Automaatses tekstianalüüsis aga tekitavad suuremaid raskusi just sõnavormid, kus muutelõpu järgi sõna vädet kindlaks määrata ei saa. Kuna meie uurimisprojekti peamine eesmärk pole vältenorme sätestada ega normimise ebakõlasid vähendada, vaid selgitada välja varieeruva vältega sõnade hääldustava, millest tekst-kõne sünteesis lähtuda, siis pakuvad ühtviisi huvi nii need sõnad, mille vältest muutevormid sõltuvad, kui ka need, kus vältest ei olene midagi. Peamine eesmärk on välja selgitada, kumba vädet iga uuritava sõna puhul eelistatakse. Selleks tuleb aga määrata ettelõetud sihtsõnade välde.

Esialgne idee oli määrata iga ettelõetud sihtsõna ehk hääldusjuhu välde automaatselt, arvutades välja sõna pearõhulise ja rõhutu silbi kestussuhte. Eesti keele kolme vädet iseloomustavad suhteliselt kindlaks kujunenud kestussuhted ning nende arvestamine on välte määramisel laialt levinud (Lehiste 1960, 1997; Liiv 1961; Eek 1983; Krull 1991; Eek & Meister 1997; Asu & Lippus *et al.* 2009; Lippus 2011). Silpide kestussuhete põhjal oleme hääldustrende analüüsinud oma varasemas artiklis “Varieeruva vältega sõnade hääldusüüringud kõnesünteesi teenistuses”, kus silpide kestussuhe saadakse rõhulise silbi riimi kestuse jagamisel teise silbi tuuma kestusega (vt täpsemalt, kuidas kestussuhet arvutatakse Piits & Kalvik 2017: 125–126). Selle analüüsi tulemusel jäi aga suure hulga uuritavate sõnade välde välja selgitamata, s.t hääldusjuhte kestussuhete järgi liigitades ei saavutanud paljude sõnade puhul kumbki välde vajalikku ülekaalu. Seetõttu tegime lisaks katse, kus samu hääldusjuhte hindasime kuuldeliselt, määrares sõna välte kuulamise järgi. Selgus, et kui enamike sõnade puhul langes kuuldeline hinnang arvutatud kestussuhtega enam-vähem kokku, siis sõnade *maitsekas*, *andekas* ja *kaine* puhul osutus ainult silbikestuste järgi saadud tulemus vastupidiseks kuuldelisele hinnangule. Kui sõnu *andekas* ja *maitsekas* võis kestussuhte järgi pidada kolmandavältelisteks, siis kuuldeliselt määrasime need sõnad üle 80 protsendil juhtudest teisevälteliseks. Iga hääldusjuhtu eraldi vaadates selgus, et sõna *maitsekas* 23 hääldusjuhust läks ainult neljal juhul kestussuhte järgi saadud välde kokku kuuldeliselt määratud vältega. Nende hääldusjuhtude puhul ei olnud ka tegemist kirderanniku murdealalt pärit keelejuhtidega, kelle kõnes võib ka tänapäeval kirjakeelepärane teise ning kolmanda välte eristus puududa (nt Kalvik 2005) ning seetõttu võib vädete kategoriseerimine olla keerukas. Nt sõna *maitsekas* kõik kuus suurima kestussuhtega (5,4; 4,4; 4,2; 3,9; 3,8; 3,7) hääldusjuhtu määrasime kuuldeliselt ikka teisevälteliseks. Sõna *kaine* võis pidada kestussuhte järgi teisevälteliseks, samas kui kuuldelise hinnangu järgi arvati ta kolmandavälteliseks. Ligi pooltel sõna *kaine* hääldusjuhtudel ei läinud kestussuhtel põhinev välde kokku kuuldelise vältemääranguga: nt klassikaliselt teisevältelise kestussuhtega¹ 1,5 hääldusjuhud määrati koguni

neljal korral kuulamise põhjal kolmandaväلتelisteks (Piits & Kalvik 2017: 131). Sellised lahknevused kinnitavad ilmekalt, et silbituumade kestussuhe on küll peamine, aga sugugi mitte ainus väلتet määrav tunnus (Lehiste 1997; Lippus 2011 jt). Tajukatsete põhjal on ka täpsustatud (nt Lippus & Pajusalu *et al.* 2009), et väلتe tajumiseks on olulisim sõna temporaalne struktuur, ent kolmandat väلتet on raske tajuda, kui põhitoonitunlus on eksitav. Eesti väلتetele iseloomulike kestussuhete varieeruvusele ning tinglikkusele on osutatud ka uuemas eesti keele häälduse käsitluses (Asu & Lippus *et al.* 2016: 134–135). Kuigi me sihtsõnade põhitoniandmeid ei ole analüüsinud, arvame, et väلتe tajumisel on oluline roll nii ajalistel kui ka tonaalsetel tunnustel, samuti konteksti olemasolul või puudumisel.

Materjal ja meetod

Siinne uurimus põhineb 50 keelejuhiga läbi viidud kõnematerjali salvestusel, mille käigus luges iga keelejuht ette 53 lauset. Uurimuses osalenud 36 naist ja 14 meest kõnelevad kõik eesti keelt emakeelena. Keelejuhtide keskmine vanus on 45 aastat, noorim neist 18- ja vanim 72aastane. Katseisikud on pärit erinevatest Eesti piirkondadest. Etteloetud laused sisaldavad 49 varieeruva väلتega sõna ehk sihtsõna. Kõigi lugejate peale kokku kogunes 2438 hääldusjuhtu, mida siinses artiklis analüüsimise.

Meie uurimuse lõppeesmärk on välja selgitada nende 49 sõna hääldustava ehk uurida, kummas väلتes sõna rohkem hääldatakse. Peame võimalikuks, et vähemalt poolte sõnade puhul on välja kujunenud häälduseelistus. Otsustasime, et sõna väلتe-eelistus on määratud, kui vähemalt 2/3 hääldusjuhtudest hääldatakse ühes väلتes. Oma töös nimetame seda domineerivaks väلتeks. Väلتe-eelistuse väljaselgitamiseks tuleb aga määrata iga hääldusjuhu väلت. Eespool kirjeldasime, kuidas varasemalt (Piits & Kalvik 2017) avastasime, et silpide kestussuhet ei saa väلتemääramisel lõpuni usaldada ning selle järgi antud väلتemäärang ja kuuldeline hinnang võisid minna vastuollu. Hääldusjuhtude kuulamisel ning väلتe määramisel aga leidsime, et kuuldeline hinnang sisaldab teatavat subjektiivsust ja sama hääldusjuhu hindamisel võivad erineda nii erinevate määrajate otsustused kui ka sama määraja hinnang erineval ajahetkel. Seega otsustasime rohkem uurida ka kuuldelse väلتemääramise enda varieerumist ja võrdlesime erinevaid väلتemääranguid.

Välte määrajad ja määramisprotsess

Välteid on kokku määranud viis inimest, kellest üks (P) on otsustanud vaid ühe keelejuhi väldete üle ning tema tulemusi edaspidi pikemalt ei kirjeldata. Piirkondliku päritolu varieeruvust vältemäärajate valimisel pole arvestatud, kõik nad on pärit Tallinnast. Kolm määrajat (M, L1 ja K) on filoloogilise haridusega, neljandal (L2) varasemad kokkupuuted väldete määramisega puudusid.

Välte määramise protsess koosnes neljast osast. Esimesena määras välte märgendaja segmenteerimise ja märgendamise käigus, teiseks kuulas kõik sama keelejuhi vältesõnad üle teine vältemääraja. Määranguid võrreldes selgusid esialgsed lahknevused. Kolmandaks tuli esimesel määrajal lahknevad hääldusjuhud uuesti üle kuulata ning jällegi välteotsustus teha, seejuures oma esialgset hinnangut arvesse võtmata. Neljandaks kaasasime kolmanda määraja, kes jäi lõplikuks otsustajaks kõigil neil juhtudel, kus esimese ja teise määraja hinnangud olid jäänud lahknevaks.

Seega on iga hääldusjuhtu kuulanud vähemalt kaks vältemäärajat ja lahknevuste korral on kaasatud otsustamisprotsessi ka kolmas. Pidasime sellist hindajate arvu optimaalseks, kuna hääldusjuhtude arv oli ligi 2500. Jättes välja mõned praakhääldused (kus inimene hääldas sihtsõna asemel midagi muud või oli hääldus muul moel ebaloomulik) selgus, et analüüsi objektiks oli 2438 erinevat hääldusjuhtu, mida kuulasid kaks hindajat. Neist 197 juhtu hindas veel kolmas vältemääraja. Seega kokku hinnati erinevaid hääldusjuhte 5073 korral. Seetõttu leidsime, et 50 keelejuhi materjali hindamine kolme inimese poolt on piisav ja enamate vältemäärajate kaasamine oleks olnud liialt töömahukas.

Kuuldeline hinnang kannab endas riski eelkõige kõnetaju kategoriaalsuse tõttu (Schouten & Gerrits *et al.* 2003; Salveste 2013) ehk väga palju oleneb kuulajast (mõjutajad: vanus, haridus, päritolupiirkond jms), kuidas ta vastavat akustilist signaali kategoriseerib. Lisaks võib kuulaja enda hinnang ka muutuda ja oleneda erinevatest teguritest. Välte määramise käigus selguski, et mõnel juhul oleneb välteotsustus sellest, kas hindaja kuulas ainult määratavat sõna või sõna koos lausekontekstiga. Artikli autorite tähelepanekute põhjal tundus, et sõna koos lausekontekstiga kuulates kiputi välteks määrama oma keeletundele sobivamat varianti, aga ainult määratavat sõna kuulates suudeti rohkem ära tunda ka oma keeletundele võõrast väldet. Seda tähelepanekut selles uurimuses siiski ei ole kontrollitud. Varasemalt on konteksti mõju kohta välte tajumisel avaldatud vastakaid tulemusi. Ühest küljest on osutunud, et kontekst on kolmanda välte tajumisel väga oluline, nt Diana Krull (1998: 170) jõudis tulemusele, et ilma kontekstita ei suudetud spontaanselt kõnest lõigatud sõnade puhul seda väldet tuvastada. Teisalt on Arvo Eek (1980) leidnud oma uurimuses, et lausekontekst ei avaldanud mõju välte tajumisele ning vältetaju

sõltus eelkõige silpide kestussuhtest. Nele Salveste (2010: 58) aga on vastupidi oletanud, et probleeme kolmanda välte äratundmisel tekitas asjaolu, et tema tajukatses olid sõnad esitatud raamlausetes ja sellest tulenevalt võis terve lausungi põhitooni kontuur maskeerida väikesed erinevused sõna põhitooni kontuuris.

Meie võimaldasime vähemäärajatel ise otsustada, kas kuulata hääldusjuhtu koos kontekstiga või ilma. Samuti ei piiranud me kuulamiskordade arvu: sama hääldusjuhtu võis kuulata piiramatult arv kordi nii lausekontekstis kui ilma. Peamine oli lõpuks otsustada, milline on konkreetse sihtsõna välde.

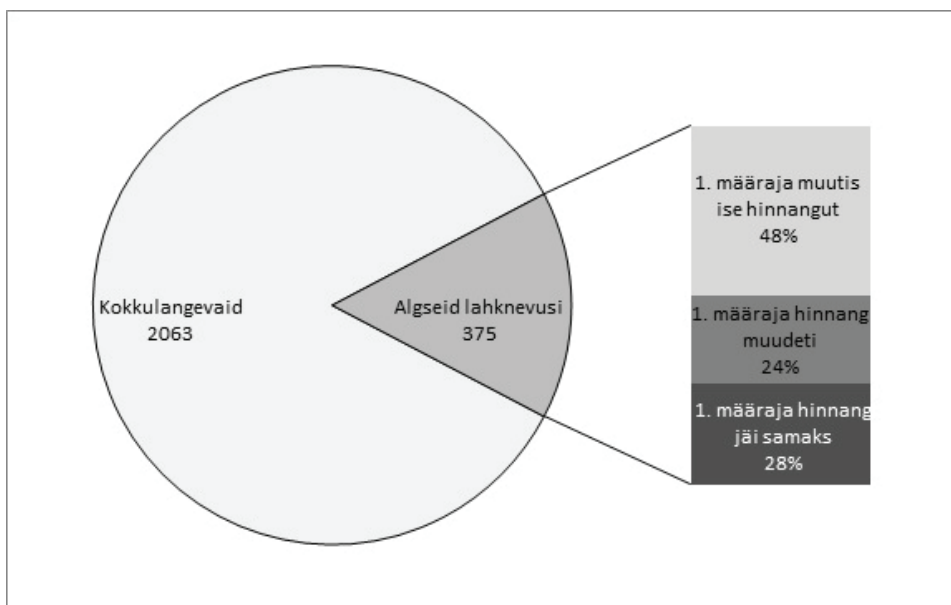
Tulemused

Kõigepealt anname ülevaate, kui palju kuuldeline hinnang erinevatel vähemäärajatel erines ja milliste sihtsõnade puhul oli lahknevusi kõige rohkem. Seejärel kirjeldame vähemäärangute põhjal selgunud välte-eelistusi.

Vähemäärangute lahknemine

Nagu eespool kirjeldatud, tegi esimese välteotsustuse märgendaja, seejärel kuulus sõnad üle teine määraja. Esimese ja teise vähemääraja hinnanguid võrreldes selgus, et kõigi 2438 hääldusjuhu puhul lahknesid kahe hindaja välteotsustused 375 korral ehk 15% juhtudest. Lahknevuse puhul tuli esimesel määrajal vastavaid sõnu uuesti kuulata ning uus hinnang anda, seejuures teadmata, milline oli ta esialgne arvamus. Jooniselt 1 on näha, et lahknevate juhtude ülekuulamisest muutis esimene määraja keskmiselt 48% ehk pea pooltel juhtudel ise hinnangut.

Ilmselt on juba nende 375 erineva vähemäärangu saanud hääldusjuhu puhul tegu keerukamate piiripealsete juhtudega, kus on rohkem tõlgendamisvõimalusi ja seetõttu lahknesid ka vähemäärajate arvamusd enam. Kindlasti oleks ülekuulamisest muudetud hinnangute protsent olnud tunduvalt väiksem, kui uus hinnang oleks tulnud anda ka neile hääldusjuhtudele, kus erimeelsused puudusid. Praegusel juhul ei pidanud kõigi hääldusjuhtude välteid uuesti otsustama, vaid ainult nende omad, kus teine määraja kuulis teistsugust välde. Siiski näitab ka see nn lahknevate juhtude ülekuulamisest saadud 48%, et vähemääraja ei pruugi oma esialgsele hinnangule kindlaks jääda. Siin tuleks rõhutada, et välte uuesti hindamisel ei olnud määraja teadlik oma eelmisest otsusest, ehk kedagi ei survestatud oma hinnanguid muutma. Hindamiste vahel jäi ajaline distants ning hindajate sõnul nad oma esialgset hinnangut ei mäletanud ning seega nende esialgne vähemäärang uut otsust ei mõjutanud.



Joonis 1. Kokkulangevused ja lahknevused määrajate hinnangutes.

Tabelis 1 on esitatud vältehinnangute muutused ka erinevate määrajate kaupa. Sealt nähtub, et vältetähtsuse K (määratud kokku kuue keelejuhi sihtsõnade välteid) on lahkneva hinnangu saanud hääldejuhte üle kuulades oma otsust muutnud koguni 73% juhtudest. Kõige vähem on oma vältetähtsust muutnud L1 (määratud kokku 13 keelejuhi sihtsõnade välteid): ainult 31% juhtudest. Seega esineb erinevusi ka määrajati.

Tabel 1. Erinevate vältetähtsuste hinnangute muutumine ja muutmine.

vältetähtsuse	mitu keelejuhti määratud	esialgsete lahknevuste % kõigist hinnangutest	muudatuste % lahknenud juhtude uuesti hindamisel	enda muudetud vältetähtsuste % lahknenud juhtude uuesti hindamisel
M	18	14%	90%	58%
L1	13	15%	53%	31%
L2	12	15%	64%	36%
K	6	22%	81%	73%

Edasi on joonisel 1 näha need juhud, kus esimene määraja jäi endale kindlaks, aga tema hinnang läks lahku nii teise kui ka kolmanda määraja hinnangust ning seega muudeti. Sellised juhud moodustasid keskmiselt 24% algselt lahknenuid juhtudest. Seega võib öelda, et keskmiselt 72% algselt lahknenuid hinnangute puhul muutis hindaja vältemäärangut ise või muudeti tema määrangut ja ainult 28% vältemäärangute puhul jäi esimese määraja hinnang muutmata, sest tema hinnangud läksid kokku kolmanda vältemääraja hinnangutega. Vaadeldes seda varieeruvust erinevate vältemäärajate lõikes on näha, et vältemääraja M (määratud kokku 18 keelejuhi sihtsõnade välteid) muutis ise või muudeti tema määratud väldeid koguni 90% vältemäärangute puhul. Kõige vähem (53% esialgselt lahknemist vältemäärangutest) muutis ise või muudeti vältemääraja L1 määranguid.

Seega võib öelda, et kuuldalise hinnangu arvestamisel tuleb arvesse võtta teatud subjektiivsus. Siiski muudab kolme vältemääraja kaasamine tulemust ehk veidi objektiivsemaks. Algselt lahknenuid 375 vältemäärangust muutus 72%, mis tähendab, et kõigi hääldusjuhtude kohta muudetakse keskmiselt 12% juhtudest vältemäärangut. Kui oleksime piirdunud ühe vältemäärajaga, siis oleksid jäänud need vältemäärangud muutmata.

Lahknemise vältemäärangutega sihtsõnad

Järgmisena vaatleme, milliste sõnade puhul kõige enam välteotsustused lahknud. Tabelist 2 on näha 15 sõna, mille välde üle otsustamisel tekkis kõige rohkem erinevusi.

Kõige enam erinevusi tekitas sõna *meetod*, mille hääldusjuhtudest 38% juhtudest olid kaks vältemäärajat eriarvamusel. Tabelis on nii selliseid sõnu, mille puhul domineeris teine välde (*ungarlane, politseinik, andekas, täiuslik*), kui selliseid, mille puhul domineeris kolmas välde (*piknikuga, nõiduslik, nooruslik, hirmus, jaanuar, pealik, looduslik*). Silma hakkab nende sõnade suur osakaal, mille puhul domineerivad väldeid ei õnnestunudki välja selgitada (*meetod, keiserlik, kaitsetu, ilmetu, enne*), ehk siis sõnad, mille hääldus kõige rohkem varieerub, tekitavad vältemäärajates kõige rohkem erinevusi.

Tabelis 2 esitatud sõnade struktuuri analüüsides selgub, et peamiselt on esindatud erinevad tuletised: nt lik-liitelised keiserlik, nõiduslik, nooruslik, täiuslik, pealik ja looduslik ning muud tuletised nagu ungarlane, kaitsetu, ilmetu, politseinik, andekas ja hirmus. Peeter Päll (1986: 13) on tuletiste puhul rõhutanud, et need hälbivad tihti sõna struktuuritüübi põhjal moodustunud produktiivsetest vältemallidest ja loovad üldtüübist erinevalt käituvaid alltüüpe.

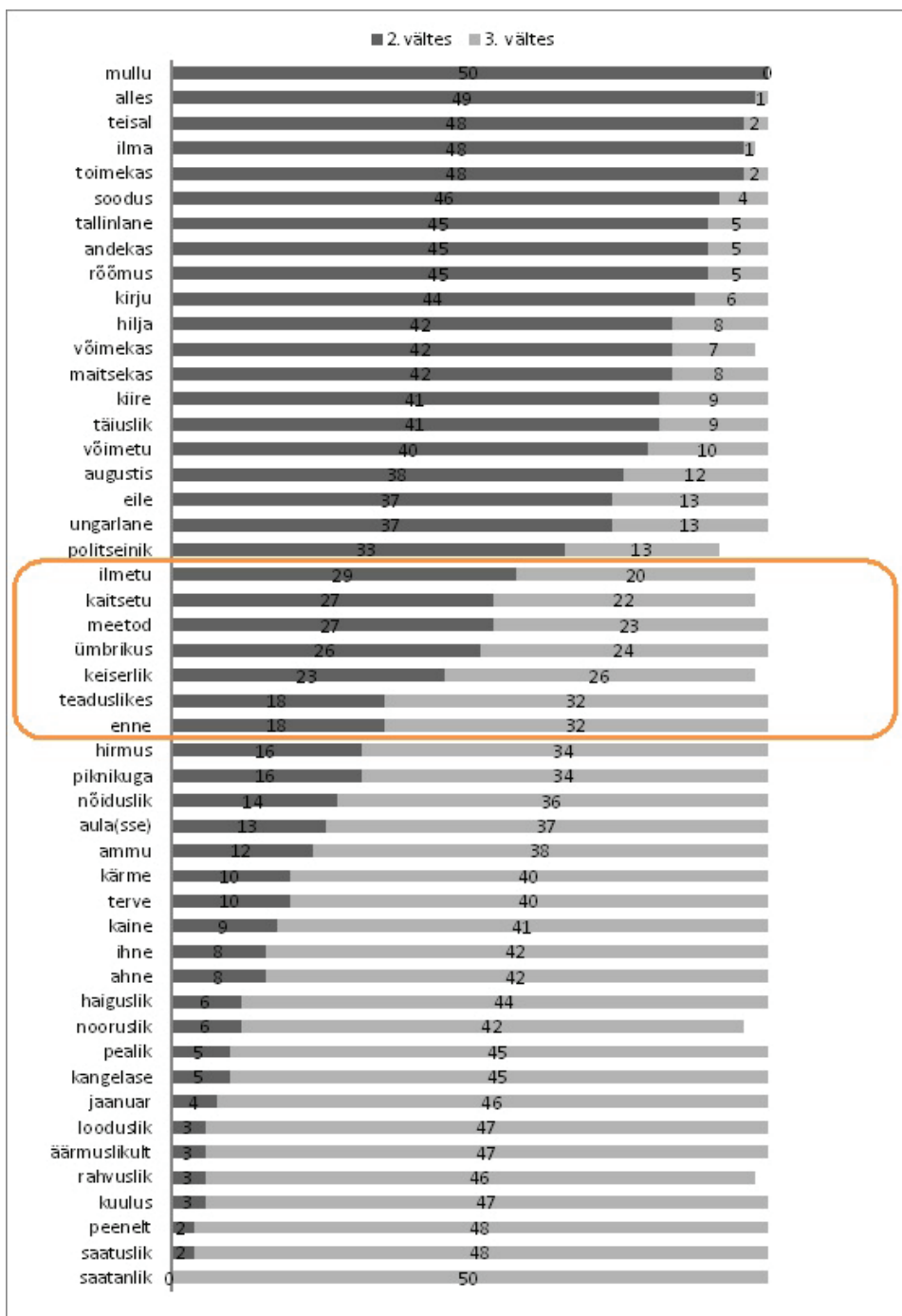
Tabel 2. Sõnad, mille hääldejuhtude hindamisel tekkis kahe vältetäpi hinnangutes kõige enam lahknevusi. Domineeriv välte tähendab, et sõna hääldejuhtudest vähemalt 2/3 on ühes või teises vältes hääldeandud: 2.v = teine välte, 3.v = kolmas välte. Küsimärk viitab, et kummaski vältes hääldejuhtud ei moodustanud 2/3 suurust ülekaalu.

sõna	algseid lahknevusi	domineeriv välte
meetod	19 (38%)	?
keiserlik	18 (37%)	?
ungarlane	17 (34%)	2.v
kaitsetu	16 (33%)	?
politseinik	15 (33%)	2.v
piknikuga	15 (30%)	3.v
nõiduslik	14 (28%)	3.v
ilmetu	12 (24%)	?
andekas	12 (24%)	2.v
nooruslik	11 (23%)	3.v
hirmus	11 (22%)	3.v
täiuslik	11 (22%)	2.v
enne	11 (22%)	?
jaanuar	10 (20%)	3.v
pealik	10 (20%)	3.v
looduslik	10 (20%)	3.v

Hääldejuhtude väljaselgitamine

Järgnevalt vaatleme, millised olid varieeruva vältega sõnade hääldejuhtude moodused 50 keelejuhi kõne põhjal. Joonisel 2 on iga sõna kohta näha, mitu inimest hääldeandis seda kuulamishinnangute põhjal teises ja mitu kolmandas vältes.

Diagrammi ülemisel poolel on esitatud sõnad, kus domineeris teine välte: loend algab sõnaga *mullu*, mille kõik 50 hääldejuhtu hinnati vältetäpi vaheliste poolt teisest vältest, ja lõpeb sõnaga *politseinik*, mille hääldejuhtudest 33 juhtu ehk 72% hääldeandis teisest vältest. Järgneb grupp sõnu, mille puhul ei õnnestunud domineerivat välde välja selgitada, st kummaski vältes hääldejuhtud ei saavutanud 2/3 suurust ülekaalu. Need seitse sõna on *ilmetu*, *kaitsetu*, *meetod*, *ümbrikus*, *keiserlik*, *teaduslikes* ja *enne*. Diagrammi alumises osas on esitatud sõnad, kus domineeris kolmas välte. Loetelu algab sõnaga *hirmus*, mille 34



Joonis 2. Varieeruva vältega sõnade hääldejuhtude vältemäärangud. Joonega on ümbritsetud sõnad, mille hääldejuhtudest ei saavutanud 2/3 suurust ülekaalu kummaski vältes juhud.

hääldusjuhtu ehk 68% on hinnatud kolmandavältiliseks ja lõpeb sõnaga *saa-tanlik*, mille kõik hääldusjuhud ehk 100% on hinnatud kolmandavältiliseks.

Seega võib öelda, et suure hulga sõnade puhul on häälduseelistus välja selgitatud ja ainult seitsme sõna puhul ei domineeri kumbki välde. Samas võib alati arutleda selle üle, kas 2/3 suurune ehk 66,6 % ülekaal on piisavalt suur või tuleks eesmärgiks võtta, et häälduseelistus on välja selgitatud, kui vähemalt 75% hääldajatest hääldab seda ühes vältes. Kuna praktiliseks eesmärgiks pole inimestele vältenorme sätestada, vaid valida tekst-kõne sünteesi jaoks üks eelistatum variant, siis otsustasime, et 2/3 on piisav ülekaal.

Varasemalt (vt Piits & Kalvik 2017) oleme 23 keelejuhi hääldatud välteid arvutanud kestussuhete põhjal ja sel moel välte-eelistusi püüdnud välja selgitada. Siis selgus, et silpide kestussuhete järgi väldeid automaatselt arvutades jäi välja selgitamata 15 sõna välde. Siinses artiklis jäi häälduseelistus leidmata seitsme sõna puhul. Seega tundub, et kuuldeliselt hääldusjuhte hinnates õnnestub sõnu teise- ja kolmandavältiliseks liigitada paremini kui kestussuhete põhjal. Samas on kestussuhete järgi välte arvutamine objektiivsem ega sõltu vältetähtsuse subjektiivsetest otsustustest. Ilmselt tasuks objektiivse kriteeriumina edaspidi uurida ka põhitoni.

Kokkuvõte

Artiklis analüüsiti 50 keelejuhi etteloetud lausetes sisalduvaid varieeruva vältega sõnu. Kõigi lugejate peale kokku kogunes 2438 hääldusjuhtu, millele antud vältetähtsuseid selles artiklis võrdlesime.

Peamiseks eesmärgiks oli välja selgitada 49 sõna hääldustava ehk uurida, kummas vältes iga sõna rohkem hääldatakse. Lisaks pöörati palju tähelepanu välte kuuldelisele hindamisele: kui palju sõltub kuuldelise hinnangu andmisel välde määrast ja kui palju esialgseid vältetähtsuseid hindamise käigus muudetakse.

Iga hääldusjuhu väldeid hindas vähemalt kaks inimest, lahkneva hinnangu puhul kaasati kolmas vältetähtsuse määraja. Esimese ja teise vältetähtsuse määrajate hinnanguid võrreldes selgus, et kõigi 2438 hääldusjuhu puhul lahknesid kahe määraja hinnangud 375 korral ehk 15% juhtudest. Lahknevate juhtude puhul andsime esimesele määrajale võimaluse uuesti väldeid määrata. Selgus, et oma esialgset määrangut muudeti keskmiselt poolte ülekuulatud hääldusjuhtude puhul. Määrajati esines küll erinevusi, nt kõige vähem ehk 31% ülekuulatud hääldusjuhtudest muutis oma esialgset otsust vältetähtsuse L1 ja kõige rohkem ehk 73% ülekuulatud juhtudest muutis otsust vältetähtsuse K. Nii suur

eneseperanduste protsent näitab, et iga kord ei olegi lihtne üheselt vältet määrata ja samale hääldusjuhule võib kuulaja ka ise anda erinevaid hinnanguid. Just vältehindajad ise muutsid hääldusjuhte üle kuulates kõige rohkem oma esialgseid hinnanguid, alles seejärel muudeti veel neid juhte, kus kaks teist vältemäärajat olid esimese vältemäärajaga eriarvamusel.

Algselt lahknenuid 375 vältemäärangust muutus lõpuks 72%. See tähendab, et kõigi 2438 hääldusjuhtu kohta muutis ise või muudeti esialgset vältemäärangut keskmiselt 12% juhtudest. Tegime kindlaks need sõnad, mille hääldusjuhtude hindamisel oldi kõige rohkem eriarvamusel. Sõnade *meetod* ja *keiserlik* häälduse hindamisel oli kõige rohkem erimeelsusi.

Varieeruva vältega sõnade häälduseelistusi uurides moodustus kolm gruppi: esimesse gruppi kuulus 20 sõna, mille puhul domineeris teine välde, teise gruppi kuulus 22 sõna, mille puhul domineeris kolmas välde ja kolmandasse gruppi kuulus seitse sõna, mille puhul ei õnnestunud domineerivat vältet välja selgitada. Usume, et suurem ja eripärasem kõnematerjali hulk aitab välte-eelistusi veelgi selgemaks teha, nii et neid võiks hakata kõnesünteesis ka rakendama. Samamoodi võimaldaks suurem materjalihulk otsida seoseid välte-eelistuse ning keelejuhi päritolupiirkonna, vanuse ja muude tegurite vahel, millest välte varieerumine võiks sõltuda.

Tänu sõnad

Artikli valmimist on toetanud Euroopa Liit Euroopa Regionaalarengu Fondi kaudu (Eesti-uuringute Tippkeskus), see on seotud Eesti Haridus- ja Teadusministeeriumi uurimisprojektiga IUT 35-1 “Kõnestiilid, lauseprosoodia ja fonoloogiline varieerumine: kirjeldus, teooria ja modelleerimine”.

Kommentaar

¹ Teise välte keskmine kestussuhe on 1,5 ja kolmandat vältet iseloomustab kahest suurem keskmine kestussuhe (Lehiste 1997).

Kirjandus

- Asu, Eva Liina & Lippus, Pärtel & Teras, Pire & Tuisk, Tuuli 2009. Realization of Estonian Quantity Characteristics in Spontaneous Speech. Vainio, Martti & Aulanko, Reijo & Aaltonen, Olli (toim). *Nordic Prosody: Proceedings of the Xth Conference, Helsinki 2008*. Frankfurt: Peter Lang Verlag, lk 49–56.
- Asu, Eva Liina & Lippus, Pärtel & Pajusalu, Karl & Teras, Pire 2016. *Eesti keele hääldus*. Eesti keele varamu II. Tartu: Tartu Ülikooli Kirjastus.
- Boersma, Paul & Weenink, David 2015. *Praat: doing phonetics by computer*. [Arvutiprogramm.] Versioon 5.4.09 (<http://www.praat.org/> – 12. oktoober 2017).
- Eek, Arvo 1980. Further information on the perception of Estonian quantity. Eek, Arvo (toim). *Estonian Papers in Phonetics: EPP*. Tallinn: Academy of Sciences of the Estonian SSR, lk 31–57.
- Eek, Arvo 1983. Kvantiteet ja rõhk eesti keeles (I). Fonoloogiliste tõlgenduste kriitikat. *Keel ja Kirjandus* 9, lk 481–489.
- Eek, Arvo & Meister, Einar 1997. Simple Perception Experiments on Estonian Word Prosody: Foot Structure vs. Segmental Quantity. Lehiste, Ilse & Ross, Jaan (toim). *Estonian Prosody: Papers from a Symposium*. Tallinn: Institute of Estonian Language, lk 77–99.
- Eestikeelse kõne autosegmenteerija. [Arvutiprogramm.] (<http://www.phon.ioc.ee/dokuwiki/doku.php?id=projects:tuvastus:est-align.et> – kasutatud 1. veebruar 2017).
- Hint, Mati 1968. Ortoeepia normeerimise probleeme. *Keel ja struktuur* 2. Töid struktuuraalse ja matemaatilise lingvistika alalt. Tartu: Tartu Riiklik Ülikool, lk 2–123.
- Kalvik, Mari-Liis 2005. Väited muutuvus rannikumurdes. *Keel ja Kirjandus* 3, lk 209–222.
- Kalvik, Mari-Liis & Piits, Liisi 2015. Lugemiseksperiment fonoloogilise varieerumise uurimiseks. [Reading experiment for discovering phonological variation.] *Aspects of Speech Studies. Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics* 6 (3), lk 49–77 (doi: 10.12697/jeful.2015.6.3.02).
- Krull, Diana 1991. Stability in some Estonian duration relations. *Experiments in speech processes*. PERILUS (Phonetic Experimental Research, Institute of Linguistics, University of Stockholm) No XIII. Papers from the Fifth National Phonetics Conference, May 29–31 1991, Stockholm, lk 57–60.
- Krull, Diana 1998. Perception of Estonian Word Prosody. A Study of Words Extracted from Conversational Speech. *Linguistica Uralica* XXXIV 3, lk 167–171.
- Lehiste, Ilse 1960. Segmental and syllabic quantity in Estonian. Sebeok, Thomas A. (toim). *American Studies in Uralic Linguistics* 1. Bloomington: Indiana University.
- Lehiste, Ilse 1997. Search for phonetic correlates in Estonian prosody. Lehiste, Ilse & Ross, Jaan (toim). *Estonian Prosody: Papers from a Symposium. Proceedings of the International Symposium on Estonian Prosody*. Tallinn: Institute of the Estonian Language, lk 11–35.

Liiv, Georg 1961. Eesti keele kolme vältusastme kestus ja meloodiatüübid. *Keel ja Kirjandus* 7–8, lk 412–424 ja 480–490.

Lippus, Pärtel & Pajusalu, Karl & Allik, Jüri 2009. The tonal component of Estonian quantity in native and non-native perception. *Journal of Phonetics* 37 (4), lk 1049–1052 (doi: 10.1016/j.wocn.2009.07.002).

Lippus, Pärtel 2011. *The acoustic features and perception of the Estonian quantity system*. Dissertationes philologiae estonicae Universitatis Tartuensis 29. Tartu: Tartu University Press.

Piits, Liisi & Kalvik, Mari-Liis 2017. Varieeruva vältega sõnade hääldeuuringud kõnesünteesi teenistuses. *Eesti Rakenduslingvistika Ühingu Aastaraamat* 13. Tallinn: Eesti Keele Sihtasutus, lk 123–140.

Päll, Peeter 1986. *Eesti noomeni silbistruktuur ja aktsent*. Tallinn: Eesti NSV Teaduste Akadeemia ühiskonnateaduste osakond.

Salveste, Nele 2010. *Väldete akustiliste tunnuste varieerumine tajus*. Magistritöö. Käsikiri Tartu Ülikooli eesti keele ja soome-ugri keeleteaduse instituudis (http://www.murre.ut.ee/arhiiv/naita_pilt.php?materjal=kasikiri&materjal_id=D1624&sari=D – 12. oktoober 2017).

Salveste, Nele 2013. Kõnetaju kategoriaalsus ehk hüpotees sellest, kuidas me keelelisi üksusi tajume. Theoretical linguistics in Estonia III. *Eesti ja soome-ugri keeleteaduse ajakiri ESUKA* 4 (1). *Journal of Estonian and Finno-Ugric Linguistics JEFUL* 4 (1). Tartu: Tartu Ülikooli Kirjastus, lk 127–143.

Schouten, Bert & Gerrits, Ellen & Hessen, van Arjan 2003. The end of categorical perception as we know it. *Speech Communication* 41, lk 71–80 (doi: 10.1016/S0167-6393(02)00094-8).

ÕS 2013 = *Eesti õigekeelsussõnaraamat ÕS 2013*. Tallinn: Eesti Keele Sihtasutus.

Summary

Words with variable quantity degrees: Auditory assessment and pronunciation preferences

Mari-Liis Kalvik

senior lexicographer and researcher of experimental phonetics
Institute of the Estonian Language
Mari-Liis.Kalvik@eki.ee

Liisi Piits

speech technology researcher
Institute of the Estonian Language
Liisi.Piits@eki.ee

Keywords: auditory assessment, Estonian, phonological variation, reading experiment, variation of quantity degrees

For the studies of quantity degrees, words were chosen which the Dictionary of Standard Estonian (ÕS 2013) lists as being pronounced with both the second and third quantity degree. As Estonian text-to-speech synthesis relies in its determination of pronunciation on this dictionary and automatic text analysis cannot handle multiple outputs, the aim is to find out which variant is more common among language users, to give the preference to one of the pronunciation variants.

This study is based on a reading experiment conducted with 50 informants (36 women and 14 men), in which each informant read 52 sentences aloud. These sentences contained 49 target words, i.e., words of variable quantity degrees; in total, the study yielded 2438 pronunciation instances to examine. Each pronunciation instance got an audio assessment made by 2 listeners. If their quantity degree assessment were conflicted (one listener gave second and another gave third quantity assessment) the third listener was then ultimate decider.

There are two main questions we would like to answer. First of all, we are interested in finding out how much an auditory assessment depends on listeners and how much the estimations are changed during the evaluation. Finding answers to these questions will help us to reach the main goal of our study: to find out the main tendencies in how the words with variable quantity degrees are pronounced.

Comparing the auditory assessment of the first and second listener, it turned out that of the 2438 pronunciations, the assessments of two listeners differed in 375 cases (15% of all the pronunciation instances). In case of inconsistencies, we gave the first listener the opportunity to make a new assessment. It turned out that the first estimation was changed on average in 50% of the instances.

These 375 cases in which the first and second listener's quantity degree assessments were conflicted were in the end changed by 72% of cases. This means that for all the 2438 pronunciations, the listeners changed their first assessment by an average of 12% of the cases.

On the basis of the informants' pronunciation, the words were grouped into three categories: the second quantity degree (words in which pronunciation with the second quantity degree dominated), variable quantity degree (where neither the second nor the third quantity degree accounted for more than 2/3 of all pronunciations), and the third quantity degree (words in which pronunciation with the third quantity degree dominated).

Based on the auditory assessment, 20 words fell into the second quantity degree group, in which words were pronounced predominantly with the second quantity degree. These words were *mullu* 'yester-year', *alles* 'only', *teisal* 'elsewhere', *ilma* 'without', *toimekas* 'expeditious', *soodus* 'favourable', *tallinlane* 'citizen of Tallinn', *andekas* 'talented', *rõõmus* 'jolly', *kirju* 'varicolored', *hilja* 'late', *võimekas* 'capable', *maitsekas* 'tasteful', *kiire* 'fast', *täiuslik* 'perfect', *võimetu* 'incapable', *augustis* 'in August', *eile* 'yesterday', *ungarlane* 'Hungarian', *politseinik* 'policeman'. Listing starts with a word which has the largest amount of second quantity degree pronunciation instances (100% of readers pronounced *mullu* as a word with the second quantity degree) and ends with the smallest amount of second quantity degree pronunciation instances (72% of readers pronounced *politseinik* as a word with the second quantity degree).

The third quantity degree group contained 22 words. These words are *saatanlik* 'satanic', *saatuslik* 'fateful', *peenelt* 'finely', *kuulus* 'famous', *rahvuslik* 'national', *äärmuslik* 'extreme', *looduslik* 'natural', *jaanuar* 'January', *kangelane* 'hero', *pealik* 'chief', *nooruslik* 'youthful', *haiguslik* 'diseaseful', *ahne* 'greedy', *ihne* 'penny-pinching', *kaine* 'sober', *terve* 'healthy', *kärme* 'fast', *ammu* 'erst', *aula* 'hall', *nõiduslik* 'bewitching', *piknikuga* 'with picnic', *hirmus* 'horrible'.

The variable quantity degree group contained 7 words: *ilmetu* 'inexpressive', *kaitsetu* 'defenceless', *meetod* 'method', *ümbrikus* 'in envelope', *keiserlik* 'imperial', *teaduslik* 'scientific', *enne* 'before'.