



## **AKUSTISEN ÄÄNENLAATUINDEKSIN (AVQI) VERSION 03.01 VALIDOINTI SUOMENKIELISILLE PUHUJILLE**

Eliina Kankare, Foniatrian poliklinikka,  
Tampereen yliopistollinen sairaala

Leena Rantala, Puheen- ja äänentutkimuksen laboratorio,  
Logopedia, Yhteiskuntatieteiden tiedekunta,  
Tampereen yliopisto

Tero Ikävalko, Puheen- ja äänentutkimuksen laboratorio,  
Yhteiskuntatieteiden tiedekunta,  
Tampereen yliopisto

Ben Barsties v. Latoszek, Speech-Language Pathology,  
SRH University of Applied Health Sciences, Düsseldorf,  
Germany ja Department of Phoniatics and Pediatric  
Audiology, University Hospital Münster, Westphalian  
Wilhelm University, Münster, Germany

Anne-Maria Laukkanen, Puheen- ja äänentutkimuksen  
laboratorio, Yhteiskuntatieteiden tiedekunta,  
Tampereen yliopisto

Akustinen äänenlaatuindeksi (AVQI) on objektiivinen, kuuteen eri akustiseen muuttujaan perustuva äänen arviointimenetelmä. Se käyttää analyysissään vokaali- ja jatkuvan puheen ääninäytettä, ja siksi AVQI validoidaan ennen käyttöönottoa eri kielille. Tässä tutkimuksessa validoitiin AVQI:n uusi, sisäisesti yhdenmukaisempi, pidempää jatkuvan puheen näytettä käyttävä versio (03.01) suomenkielisille puhujille. Tutkimuksessa oli osallistujina 197 vapaaehtoista, äidinkielenään suomea puhuvaa henkilöä, joista 111

oli sairaalan foniatriksen yksikön potilaita ja 86 terveäänisiä. Osallistujilta tallennettiin luenta- ja vokaaliääninäytteet. Luentanäytteistä mitattiin keskimääräinen kolmen sekunnin vokaaliääntöä vastaava tavujen määrä. Kuusitoista asiantuntijakuuntelijaa arvioi neliportaisella asteikolla äänen yleislaatua. Tilastollisilla menetelmillä arvioitiin AVQI-analyysin diagnostista kykyä erotella terve- ja äänihäiriöäänäni toisistaan suomenkielisellä aineistolla. Kolmen sekunnin vokaaliääntöä vastaavaksi tavumääräksi määritettiin 31 tavua. AVQI-tulosten ja kuunteluarvioiden välinen yhteys oli vahva (Spearmanin rho 0.77,  $p=0.01$ ). Paras erottelevuus terveen- ja äänihäiriöäänänen välille saatiin raja-arvolla 1,83. Tutkimus osoitti AVQI 03.01FIN -version toimivan hyvin äänihäiriöiden diagnostisena työkaluna.

**Avainsanat:** akustinen analyysi, kuunteluanalyysi, äänihäiriö

## 1 JOHDANTO

Äänihäiriöt jaetaan karkeasti kolmeen luokkaan, elimellisiin, neurologisiin ja toiminnallisiin häiriöihin (Stemple, Roy & Klaben, 2014). Elimellisiä äänihäiriöissä äänihuulikudoksessa on äänentuottoa haittaava muutos, kuten äänihuulikyhmy tai polyyppi. Neurologisissa äänihäiriöissä kurkunpään toiminta häiriintyy neurologisista syistä kuten esimerkiksi äänihuulihalvauksessa tapahtuu. Toiminnallisissa äänihäiriöissä kurkunpään rakenteet ovat normaalit, mutta kurkunpään toiminta on syystä tai toista muuttunut aiheuttaen häiriön äänentuottoon. Äänihäiriössä äänen laatu heikkenee eikä ääni riitä käyttäjänsä tarpeisiin. Äänen laadun arviointi on tärkeä osa äänihäiriön diagnosointia. Äänen laatua arvioidaan tavallisimmin kahdella tavalla: kuuntelemalla ja analysoimalla äänen piirteitä akustisesti (Stemple ym., 2014). Näistä äänen akustinen analyysi on objektiivinen, havainnoitsijasta riippumaton menetelmä, kun taas kuulohavaintoarvioon vaikuttavat väistämättä kuuntelijan ominaisuudet, kuten esimerkiksi kokemus äänihäiriöiden hoidossa (Eadie & Baylor, 2006) ja ammatillinen tausta (De Bodt, Wuyts, Van de Heyning & Croux, 1997). Niin ikään arvioon vaikuttavat puhenäyte (Zraick, Wendel & Smith-Olinde, 2005), äänihäiriön vaikeusaste (Rabinov, Kreiman, Gerratt & Bielamowicz, 1995) ja arvioinnissa käytetty mittari (Karnell ym., 2007; Yu, Revis, Wuyts, Zanaret & Giovanni, 2002). Parhaimmillaan kuuntelijat arvioivat äänen häiriöisyyttä hyvinkin yhdenmukaisesti (Kreiman, Gerratt, Kempster, Erman & Berke, 1993), mutta etenkin lievien äänihäiriöiden arvioinnissa havainnot saattavat poiketa toisistaan (Rabinov ym., 1995).

Kuulonvarainen arvio äänenlaadusta on parhaimmillaankin kuuntelijan subjektiiv-

vinen arvio äänestä. Vaikka äänenlaadun arvioinnissa kokeneen kuuntelijan *korvaa ei mikään korvaa*, tarvitaan sen rinnalle sekä tutkimuskäyttöön että kliiniseen työhön objektiivista äänenlaadun arviointia (Roy ym., 2013). Tarve on tunnistettu jo sata vuotta sitten, mistä lähtien akustista signaalia mittaavia algoritmejakin on kehitelty runsaasti – Buderin (2000) katsauksen mukaan yli sata. Ainoakaan algoritmi ei ole kuitenkaan yksinään kyennyt mittaamaan äänen häiriöisyysastetta ja erottelemaan riittävän luotettavasti erilaisia äänenlaatuja. Sen sijaan lupaavimpia ovat olleet ne algoritmit, jotka ovat sisältäneet useita äänenlaatuja kuvaavia muuttujia (Awan & Roy, 2009; Wuyts ym., 2000).

Uusin tulokas tällaisista monimuuttujaisista algoritmeista on Acoustic Voice Quality Index (AVQI), akustinen äänenlaatuindeksi (Maryn, Bodt & Roy, 2010; Maryn, Corthals, Van Cauwenberge, Roy & De Bodt, 2010.) AVQI on jo käytössä useassa maassa ympäri maailmaa ja esimerkiksi Belgiassa lääkintäviranomaiset vaativat sen käyttöä kaikilta äänipotilaiden kanssa työskenteleviltä. Suomessa AVQI:a käytetään tällä hetkellä kliinisessä työssä ainakin kahdessa yliopistosairaalassa ja lisäksi muutamat yksityisen sektorin puhe-terapeutit käyttävät sitä äänipotilaiden äänen ja terapiatulosten arvioinnissa. Tutkimuskäytössä AVQI on Suomessa Tampereen ja Oulun yliopistoissa. AVQI:n vahvuutena on, että se sisältää informaatiota kahdesta erilaisesta ääninäytteestä: pidennetystä [a:]-fonaatiosta ja jatkuvasta puheesta (luenta). Näiden kahden näytteen hyödyntäminen tekee AVQI:sta ekologisesti validin, sillä kumpikin ääninäyte antaa erilaista tietoa äänenlaadusta. Lisäksi menetelmän kehitystyö on osoittanut, että muutaman lauseen luentanäyte riittää jatkuvan puheen näytteeksi (Maryn, Corthals ym., 2010; Zraick ym., 2005; Barsties & Maryn, 2015). Esimerkiksi adduktorisessa spasmodisessa dysfoniassa spasmit tulevat herkemmin

esiin jatkuvassa puheessa kuin pidennetyssä fonaatiossa, joka saattaa olla suhteellisen normaali (Roy, Gouse, Mauszycki, Merrill & Smith, 2005). Myös äänen häiriöisyyden aste voi olla paremmin kuultavissa jatkuvasta puheesta kuin pidennetystä fonaatiosta (Revis, Giovanni, Wuyts, & Triglia, 1999). Pidennetty fonaatio on puolestaan ainoa ääninäyte, josta voi mitata luotettavasti äänihuulijaksojen epätasaisuuden/huojunnan eli akustisesti kutsuttuna jitterin ja shimmerin (perustaajuuden ja amplitudin eli intensiteetin huojunta).

AVQI:n kehittäminen käynnistyi meta-analyysistä, jossa tutkijaryhmä Maryn, Roy, De Bodt, Van Cauwenberge ja Corthals (2009) kävi läpi tutkimuksia, joissa verrattiin subjektiivisesti arvioitua ja objektiivisesti mitattua äänenlaatua toisiinsa. Ainoastaan 25 tutkimusta 85:stä täytti tutkijaryhmän asettamat kriteerit. Ääninäytteistä vokaaliääntö (*sustained vowels*) ylsi puhenäytettä (*continuous speech*) useammin tutkijoiden edellyttämään tarkkuuteen (vokaaliääntö 21 tutkimuksessa; jatkuva puhe 7:ssä). Meta-analyysi osoitti, että subjektiivisten ja objektiivisten arviointien yhdenmukaisuus vaihteli paljon. Analyysinsä perusteella ryhmä poimi akustiseen äänenlaatuindeksiin viisi eri parametria, joista kahta lasketaan kahdella eri tavalla (Maryn, Corthals ym., 2010). Valitut parametrit olivat tasoitetun kepstrin huipun erottuvuus (*smoothed cepstral peak prominence*, CPPS), signaali–kohinasuhde eli äänen akustisen jaksollisuuden aste (*harmonics-to-noise ratio*, HNR), kaksi amplitudiperturbaatiota (shimmer) ja kaksi pitkäaikaisspektrin kaltevuutta mittaavaa parametria *slope* (yleinen spektrin jyrkkyys) ja *tilt* (spektrin kallistuminen) (Maryn, Corthals ym., 2010).

Kepstri on logaritmisen (tehollis)spektrin spektri (Hillenbrand, Cleveland & Erickson, 1994). Se kertoo, miten nopeita muutoksia spektrikaistojen energiassa tapahtuu. Kepstri

erottelee hitaat muutokset (formantit) nopeista muutoksista (perustaajuus) ja hälyn periodisesta signaalista. Kepstrihuipun prominenssi eli erottuvuus kertoo, miten hyvin kepstrin amplitudimaksimi erottuu taustakohinasta. Käytännössä tähän taustakohinaan liittyy niin signaalin perturbaatio kuin turbulenssihälykin, jota aiheuttaa esimerkiksi epätäydellinen äänihuulisulku. Tasoitetun kepstrihuipun prominenssi (CPPS) saadaan siten, että lasketaan ero huipun amplitudin ja koko tasoitetun kepstrin keskimääräisen amplitudin välillä (tasoitetun kepstrin amplitudi lasketaan regressiosuorasta). Mitä suuremman arvon CPPS saa, sitä periodisempi eli kuulohavainnoltaan selkeämpi ja hälyttömämpi ääni on.

HNR mittaa äänen harmonisen ja epäharmonisen energian suhdetta. Pieni HNR viittaa hälyisempään ääneen, jota voi aiheuttaa esimerkiksi äänihuulien sulkuvajaus (Awan & Roy, 2006). Myös äänihuulivärähtelyn epätasaisuus eli perturbaatio lisää hälyisyyttä. Lisäksi ääninäytteen voimakkuus ja sävelkorkeus saattavat vaikuttaa niin hälypitoisuuden kuin perturbaation määrään siten, että molemmat pienenevät voimakkuuden kasvaessa ja korkeuden noustessa (Glaze, Bless & Susser, 1990; Koike, 1973; Lieberman, 1963; Orlikoff & Kahane, 1991).

Shimmer kertoo äänisignaalin perättäisten jaksoiden amplitudien (värähdyslaajuuksien) poikkeaman suuruuden keskimääräisestä amplitudista. AVQI:ssa shimmer lasketaan sekä prosentteina että desibeleinä. Mitä suurempi shimmerin arvo on, sitä enemmän äänihuulivärähdysten amplitudit poikkeavat toisistaan. Riittävän epätasainen värähtely saa aikaan kuulohavainnon karheasta (*rough*) äänenlaadusta. Epätasaisen äänihuulivärähtelyn syynä on usein äänihuulten limakalvon massassa tapahtunut muutos, kuten turvotus tai äänihuulikyhyt (Hammarberg, Fritzell, Gaufin, Sundberg & Wedin, 1980). Varsin-

kin jos limakalvomuutos on vain toisessa äänihuulella, shimmerin määrä kasvaa (Rosen, Lombard & Murry, 2000) tai toinen äänihuuli on liikkumaton äänihuulihalvauksen takia (mm. Oguz, 2007).

Pitkäaikaisspektrin kaltevuus ilmaisee kuulohavaintoa äänenlaadusta jatkumolla hypo-hyperfunktionaalinen. AVQI mittaa spektrin kaltevuutta kahdella luvulla: yleinen spektrin jyrkkyys ja spektrin kallistuminen. Yleinen spektrin jyrkkyys lasketaan AVQI:ssa kahden taajuuskaistan (1–10 kHz ja 0–1 kHz) äänenpainetasojen välisenä suhteena. Koska äänenpainetaso on logaritminen suure, kaistojen energiamäärän suhde lasketaan äänenpainetasojen erotuksena. Spektrin kallistuminen kuvaa samaa asiaa, mutta siinä kallistuman arvo on sen suoran kaltevuus, joka kulkee pitkäaikaisspektrin läpi. Suoran kaltevuus lasketaan spektrin kallistumassa samojen kaistojen suhteena kuin yleisessä spektrin jyrkkyudessa. Koska energiaa on enemmän spektrin matalissa taajuuksissa kuin korkeissa, molempien laskentatapojen tuottamat lukuarvot ovat negatiivisia. Mitä suurempi arvo (itseisarvoltaan pieni luku) on, sitä voimakkaampia ovat spektrin yläsävelet. Voimakkaat yläsävelet voivat liittyä äänen voimistamiseen, tai ne voivat kertoa äänen hyperfunktionaalisuudesta (puristeisuudesta).

AVQI-laskenta antaa tuloksen janalle 0 ja 10 välillä (kuvio 1). Mitä suuremman arvon indeksi saa ylitettyään terveen ja sairaan äänen rajan, sitä poikkeavampi äänenlaatu on. Koska AVQI-analyysissä on mukana jatkuvaa puhetta, ainakin jossakin määrin puhujan kieli vaikuttaa indeksin arvoon ja siihen, mihin terveen ja häiriintyneen äänen raja asettuu.

AVQI:sta on kaksi versiota (versiot 02.02 ja 03.01). Vokaaliäännöt ovat versioissa samanpituisia, mutta puhenäytteiden (luenta) pituudet poikkeavat toisistaan (Barsties & Maryn, 2015). Puhenäytteiden pituus AVQI 02.02 -versiossa on vaihdellut eri kielten vali-

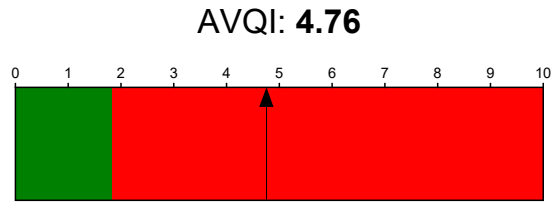
doinneissa 17 tavusta 23:een. Version 02.02 AVQI-laskennassa [a:]-äännön laatu on korostunut, sillä AVQI-analyysin skripti poistaa puhenäytteestä taout ja sen myötä esimerkiksi myös soinnittomat konsonantit, jolloin puhenäyte lyhenee. Parantaakseen AVQI:n sisäistä yhdenmukaisuutta (*internal consistency*) Barsties ja Maryn (2015, 2016) suosittelivat, että puhenäytteen pituutta kasvatetaan vastaamaan paremmin vokaaliäännön kolmen sekunnin kestoa. Uuden AVQI -03.01 version puhenäytteisiin lisättiin siis tavuja, ja esimerkiksi hollannin kielessä tavuja on nyt 34 (Barsties & Maryn, 2015), ranskassa 27 (Pommée, Maryn, Finck & Morsomme, painossa) ja japanissa 30 (Hosokawa ym., 2019).

Tähän mennessä AVQI 03.01 -versio on validoitu viidelle kielelle, ja Brasilian portugalilainen versio on tekeillä (Barsties & Maryn, 2015; Pommée ym., painossa; Hosokawa ym., 2019; Hernández ym., 2018; Barsties v. Latoszek ym., painossa). AVQI 02.02 on validoitu suomalaisille puhujille, ja se sisältää 23 tavua ”Pohjantuuli ja aurinko” -tekstin alusta (Kankare ym., 2019). Sen sijaan version 03.01 validointi puuttuu suomenkielisiltä puhujilta. Tämän tutkimuksen tavoitteena onkin tämän version validointi.

Tässä artikkelissa kuvattu validointitutkimus jakautui kahteen osaan. Ensimmäiseksi selvitettiin, kuinka monta tavua sisältää luentanäyte, joka keskimäärin vastaa ajallisesti kolmen sekunnin vokaaliääntöä. Kun optimaalinen luentatekstin pituus oli selvitetty, testattiin AVQI 03.01FIN -version kykyä erotella terve ja häiriöinen ääni toisistaan kolmen sekunnin vokaaliääntöä vastaavalla luentanäytteellä. Validoinnissa selvitettiin myös AVQI 03.01FIN -versiolla terveen ja dysfonisen äänen kynnyсарvo suomenkielisille puhujille.

**ACOUSTIC VOICE QUALITY INDEX (AVQI) v.03.01**

Smoothed cepstral peak prominence (CPPS): <b>8.46</b>
Harmonics-to-noise ratio: <b>16.76 dB</b>
Shimmer local: <b>7.44 %</b>
Shimmer local dB: <b>0.72 dB</b>
Slope of LTAS: <b>-28.26 dB</b>
Tilt of trendline through LTAS: <b>-11.08 dB</b>



KUVIO 1. Esimerkki AVQI 03.01FIN Praat-skriptin tulostuokuvasta. Kuvaajassa vasemmalla on taulukko AVQI:n kuuden parametrin tuloksista (osallistujan numero 32, dg. äänihuulipareesi). Taulukon vieressä oikealla näkyy AVQI-tulos 0-10 janalla 4.76. AVQI -raja- arvo terve ja äänihäiriöäänänen välillä on 1.83 (näkyvä kuviossa vaalean ja tumman harmaan rajana). Kuunteluanalyysin keskiarvo (G = grade, skaalalla 0-3) osallistujan numero 32 ääninäytteenä oli 1,9 (16 asiantuntijakuuntelijaa).

## 2 MENETELMÄT

Tämä tutkimuksen suorittamiseen on saatu lupa Tampereen yliopistollisen sairaalan eettiseltä toimikunnalta (koodi R15014).

### 2.1 Ääninäytteiden tallennus

Ääninäytteinä oli terveäänisten ja äänihäiriöisten puhujien tallenteet. Terveäänisten näytteet kerättiin Tampereen yliopistossa ja äänihäiriöisten Tampereen yliopistollisessa sairaalassa. Tallennusolosuhteet vaihtelivat: 159 osallistujan näytteet tallennettiin tavallisissa toimisto- tai vastaanottohuoneissa ja 38:n näytteet studio-olosuhteissa. Osallistujilta tallennettiin kaksi ääninäytettä. Ensimmäinen näyte oli 77 sanan mittainen tarina ”Pohjantuuli ja aurinko”, ja toisena näytteenä kolmesti toistettu, noin 5 sekunnin mittainen [a:]-vokaali. Osallistujaa ohjeistettiin käyttämään ääninäytteissä itselleen sopivaa puheäänänen korkeutta ja voimakkuutta. Mikäli osallistuja tuotti vokaalinäytteen laulunomaisesti, pyydettiin häntä toistamaan näyte

useaan kertaan, jotta se vastaisi hänen tavantomaista äänenkorkeuttaan ja puheenomaista äänenlaatuaan. Tarvittaessa osallistujaa kehoitettiin etsimään itselleen sopiva korkeus hymähtämällä tai toistamalla lyhyt sana ennen vokaaliääntöä.

Äänentallennuslaitteistona oli AKG C544L- pääpantamikrofoni ja Focusrite iTrack Solo -äänikortti ja PC-tietokoneelle asennettu Praat-ohjelma (versio 6.0.39) (Boersma & Weenink, 2013). Mikrofonin oli 45 asteen kulmassa 4 cm:n etäisyydellä henkilön huulikulmasta, ja sen etäisyys tarkastettiin jokaisessa äänityksessä viivaimella. Tallennuksessa käytettiin 44,1 kHz:n näytteenottotaajuutta, ja amplitudin erottelutarkkuus oli 16 bittiä. Näytteet tallennettiin wav-tiedostoina. Huoneakustiikan sopivuus arvioitiin mittaamalla äänitallenteista SNR-suhdeluku (*signal-to-noise ratio*). Kaikissa näytteissä SNR oli suositusten mukainen eli suurempi kuin 30 dB (KA 39,8 dB, KH 5,6 dB), eli ne soveltuivat akustiseen analyysiin (Deliyski, Shaw & Evans, 2005).

## 2.2 Luentanäytteen tavumäärän määrittäminen

### *Osallistujat*

Luentanäytteen tavumäärän määrittämiseen käytettiin tutkimuksen 137:n ensimmäisen osallistujan luentaääninäytettä. Kolmen osallistujan näyte jouduttiin poistamaan lukuvirheen takia. Jäljelle jääneiden 134 osallistujan (ikä KA 48,1 vuotta, hajonta 15,5, VV 19–84 vuotta) näytteiden avulla selvitettiin suomenkielisen luentatekstin ”Pohjantuuli ja aurinko” kolmen sekunnin ääntöä vastaava keskimääräinen tavumäärä. Miesten ääninäytteitä oli 32 ja naisten 102. Näytteistä 83 oli sairaalassa tallennettuja äänihäiriöpotilaiden näytteitä ja loput 51 terveäänisten näytteitä.

### *Tavumäärän standardointi ja sen tilastollinen analyysi*

Suomen kielelle validoitavan AVQI 03.01-version tavumäärän haku tehtiin samalla tavalla kuin Barsties ja Maryn (2015) ovat kuvanneet selvittäessään hollanninkielisen näytteen tavumäärää. Koska analyysissä poistetaan soinnittomat konsonantit ja tauot, mitattiin ensin tauottomien luentanäytteiden kokonaiskesto (KA 18,1 sekuntia, KH 3,3 sekuntia). Tämän jälkeen mitattiin jokaisen luentanäytteen alusta yksilöllisesti kolmen sekunnin mittainen tauoton osa, ja niistä laskettiin kolmessa sekunnissa olevien tavujen määrä. Puhenopeudesta johtuen osallistujien luennan kolmen sekunnin tavumäärä vaihteli 12–50 tavun välillä. Kestoltaan parhaiten kolmen sekunnin vokaaliääntöä vastaavaksi ääninäytteeksi saatiin 95 %:n todennäköisyydellä 30 tavua. Kolmaskymmenes tavu tarinan alusta päättyy sanan ”näkevät” tavuun ”ki”. Koska keskelle sanaa päättyvä tavumäärä on visuaalisesti vaikeampi erottaa akustisesta signaalista kuin sanan lopusta, mikä hidastaa näin käytännön työtä, päätettiin tilastollisesti testata 31 tavun käyttöä puhenäytteenä

AVQI-analyysissa. Kaikki ääninäytteet editoitiin yksilöllistä kolmen sekunnin ääntöä vastaavaan tavupituuteen (KA = 2,99 s., KH = 0,06) ja 31 tavun pituuteen (KA = 3,09 s., KH = 0,60). Näytteet editoitiin Praat-äänianalyysiohjelmalla (versio 6.0.39).

AVQI-analyysi suoritettiin 134 osallistujan molemmilla tavumäärillä ja tulosten välistä yhdenmukaisuutta arvioitiin Wilcoxonin merkittyjen sijaintilukujen testillä (Windows-pohjainen SPSS 22.0 -tilastoanalyysiohjelma, IBM Corp., Armonk, NY, USA). Merkitsevää eroa ei 31 tavulla laskettujen AVQI-tulosten ja yksilöllisesti editoitujen kolmen sekunnin tavumäärällä laskettujen tulosten välillä ollut  $Z = 1,277, p = 0,204$ ). Sisäkorrelaation mukaan kahden tavumäärältään eripituisen puhenäytteen AVQI-tulokset olivat lähes identtiset ( $r = 0,99$ ; kuvio 2), joten standardoiduksi tavumääräksi valittiin käytännön työtä helpottava, kokonaiseen sanaan päättyvä, 31 tavun ilmaus. Näin suomenkieliseen AVQI 03.01 -analyysiin käytettäväksi luentanäytteeksi tuli ilmaus ”Pohjantuuli ja aurinko väittelivät kummalla olisi enemmän voimaa, kun he samalla näkivät”.

## 2.3 AVQI 03.01FIN -version validointi

### *Osallistujat*

Tutkimuksen validointiosioon osallistui 197 (ikä KA = 46,7 vuotta, hajonta 15,4, vaihteluväli 19–84 vuotta) vapaaehtoista, suomea äidinkielenään puhuvaa henkilöä. Osallistujista 111 (ikä KA = 50,8 vuotta, hajonta 15,2, vaihteluväli 19–84 vuotta) oli potilaina Tampereen yliopistollisessa sairaalassa foniatrian poliklinikalla. Tutkimuksen potilasryhmä muodostui suurelta osaltaan foniatrian poliklinikan yhden puheterapeutin vastaanotolla vuoden aikana käyneistä äänihäiriöpotilaista. Terveäänisiä osallistujia oli 86 (ikä KA 41,5 vuotta, hajonta 14, vaihteluväli 19–67 vuotta). He olivat yliopiston henkilökuntaa, opis-

kelijoita ja yliopistolla vierailevia henkilöitä. Kenelläkään yliopistolta rekrytoituista osallistujista ei ollut äänihäiriödiagnoosia, mutta kahdeksan heistä sai VAPP-kyselystä (*Voice Activity and Participation Profile*) (Ma & Yiu 2001) pistemääräksi yli 38, mitä pidetään kyselyssä äänihäiriön pisterajana (Kleemola, Helminen, Rorarius, Isotalo & Sihvo, 2011). Naisia tähän tutkimukseen osallistujista oli 151 (77 %) ja miehiä 46 (23 %). Sukupuolijakauma noudattaa hyvin tämän artikkelin ensimmäisen kirjoittajan kliinistä kokemusta puheterapeutin potilaiden sukupuolijakaumasta foniatriisella poliklinikalla.

#### TAULUKKO 1.

Osallistujat listattuna diagnoosiryhmittäin ja terve äänisten ryhmään, diagnoosit ICD-10 koodein.

Diagnoosi	ICD10 koodi	Muuta selittävää	Määrä
Osallistujat, joilla ei diagnosoitua äänihäiriötä			86
Toiminnallinen äänihäiriö	R49.01		29
Spasmodinen dysfonia	R49.02	1 abduktorinen 23 adduktorista	24
Äänihuulipareesi/ paralyysi	J38.0	Eri asteisia äänihuulihalvauksia	23
Krooninen laryngiitti	J37.0		8
Äänihuulikyhmyt	J38.2		5
Muu äänihuulisairaus	J38.3	Esim. Kysta, verisuinipullistuma, turvotusta äänihuulissa	5
Muu määrittelemätön äänihäiriö	R49.08	Kranuloma, monitekijäinen äänihäiriö	5
Muu kurkunpään sairaus	J38.7	Myokloonus, kurkunpään spasmi, toiminnallinen äänihuulisalpaus, yliärtyvä kurkunpää	5
Transseksualismi	F64.0	Miehestä naiseksi muutos ja siihen liittyvä äänen häiriö	2
Äänihuulipolyyppi	J38.1		1
Yskä	R05	+ äänioireet	1
Toiminnallinen nielemis- häiriö	R13	+ äänioireet	1
Ehlers-Danlos syndrooma	Q79.6	+ äänioireet	1
Kurkunpäätrauma	Y96.0, W17		1
		Foniatrian poliklinikan potilaat	111
		Osallistujat yhteensä	197



### Ääninäytteet ja AVQI-analyysi

Kaikki ääninäytteet analysoitiin siten, että analyysissa yhdistyi uusi jatkuvan puheen pidennetty tavumäärä (ks. luku 2.2) ja 3 sekunnin vokaalinäyte. Ennen analyysia näytteet editoitiin sovitun pituisiksi ja nimettiin ohjelman vaatimalla tavalla: luentanäyte cs:kksi (lyhenne termistä *connected speech*) ja vokaalinäyte sv:kksi (lyhenne termistä *sustained vowel*). Analyysissa käytettiin Praat-ohjelmaan ladattua Barstiesin ja Marynin (2015) laatimaa AVQI 03.01 laskentakaavaa  $(4.152 - (0.177 \times \text{CPPS}) - (0.006 \times \text{HNR}) - (0.037 \times \text{Shim}) + (0.941 \times \text{ShdB}) + (0.01 \times \text{Slope}) + (0.093 \times \text{Tilt})) \times 2.8902$ . Analyysi yhdistää luetun tekstin ääntöosuuden ja vokaalinäytteen automaattisesti, ja tulokseksi saadaan yksi AVQI-lukuarvo asteikolla 0–10.

### Kuunteluanalyysi

AVQI 03.01 -version validoimiseksi suomenkielisille puhujille ääninäytteet arvioitiin kuulonvaraisesti. Ääninäytteen pituus oli noin 6 sekuntia, joka oli sama, mistä AVQI -analyysi tehtiin (noin 3 sekuntia luentaa ja 3 sekuntia [a:]-vokaalia). Kuusi puheterapeuttia, viisi foniatria ja viisi vokologia kuuntelivat yhteensä 221 näytettä. Sisäisen reliabiliteetin arvioimiseksi 24 näytettä kuunneltiin sekoitettussa järjestyksessä kahteen kertaan. Kaikki kuunteluanalyysiin osallistuneet puheterapeutit ja foniatrit työskentelivät foniatristen potilaiden parissa, ja kaikilla vokologeilla oli paljon kokemusta kuunteluanalyysin teosta. Kuunteluanalyysissa käytettiin neliportaisen GRBAS-asteikolla G-arviota (*Grade*), joka kuvaa äänen kokonaispoikkeavuuden astetta (Hirano, 1981). Arviointiasteikon ääripäinä olivat 0 (terve ääni) ja 3 (erittäin poikkeava äänenlaatu).

Kuuntelijat saivat ääninäytteet muistitukilla, jossa näytteet olivat satunnaistetussa järjestyksessä. He kuuntelivat ääninäytteet

omalta tietokoneeltaan kuppikuulokkeilla ja merkitsivät arviointituloksensa laskentataulukkaan. Kuunteluanalyysin aluksi kuuntelijat kuuntelivat niin sanotut ankkuriäänet, joita oli yhteensä kahdeksan: kaksi näytettä esimerkkinä kustakin G:n häiriötasosta (G 0–3). Ankkuriäänet oli valittu viiden puheterapeutin kuunteluanalyysin perusteella. Kuuntelijoita pyydettiin antamaan yksi yhteinen arvio molemmista ääninäytteistä eli luennasta ja vokaalinäytteestä. Jotta kuuntelijoiden tarkkaavaisuus ei herpaantuisi eikä kuuloaisti väsyisi, laskentataulukossa oli muistutus kuunnella ankkuriäänet uudelleen aina 25 kuuntelunäytteen jälkeen, ja näytteen 110 kohdalla oli kehoitus pitää pieni lepotauko.

### 2.4 Tilastollinen analyysi

Tilastolliset analyysit tehtiin Windows-pohjaisella SPSS 22.0 -tilastoanalyysiohjelmalla (IBM Corp., Armonk, NY, USA). AVQI03.01FIN -version validointiprosessissa arvioitiin ensin kuuntelijoiden arvioiden yhtäpitävyyttä samojen näytteiden toistoarvioinnissa (sisäinen reliabiliteetti) Cohenin kappa ( $C_k$ ) -analyysillä, ja tämän jälkeen kuuntelijoiden välisiä kuunteluanalyysin tulosten reliabiliteettia (yhdenmukaisuutta) Fleissin kappa ( $F_k$ ) -kertoimella. Tarkemmin  $\kappa$ -kertoimien käyttöön voi perehtyä Landisin ja Kochin (1977) artikkelissa.

Tutkimukseen valittiin kuuntelijoiksi ne, joiden arviointi oli ollut luotettava. Heidän arviointinsa täytyi täyttää seuraavat kuuntelun yhdenmukaisuuden edellytykset: 1) kuuntelijat eivät saaneet erota toisistaan merkittävästi sisäisen reliabiliteetin suhteen 2) sisäinen  $C_k$  reliabiliteettiarvo  $\geq 0,41$  (Landis & Koch 1977), 3) kuuntelijoiden välinen reliabiliteetti arvioitiin  $F_k$ -kertoimella ( $F_k \geq 0,41$ ) (Landis & Koch 1977) kaikkien niiden kuuntelijoiden osalta, joiden sisäinen reliabi-

liteetti todettiin riittäväksi. Fleissin kappa-analyysissa poistettiin yksi kuuntelija kerrallaan ja arvioitiin, oliko  $F\kappa$ -tulos merkitsevästi parempi poiston seurauksena. Testaus aloitettiin kuuntelijoista, joiden kuuntelutulos erosi eniten kaikkien kuuntelijoiden  $F\kappa$ -tuloksesta. Testausta jatkettiin, kunnes jäljellä olevien  $F\kappa$ -tulos oli minimissään  $\geq 0,41$ .

Validointia jatkettiin tutkimalla kuuntelu-arvion keskiarvojen ja AVQI 03.01FIN

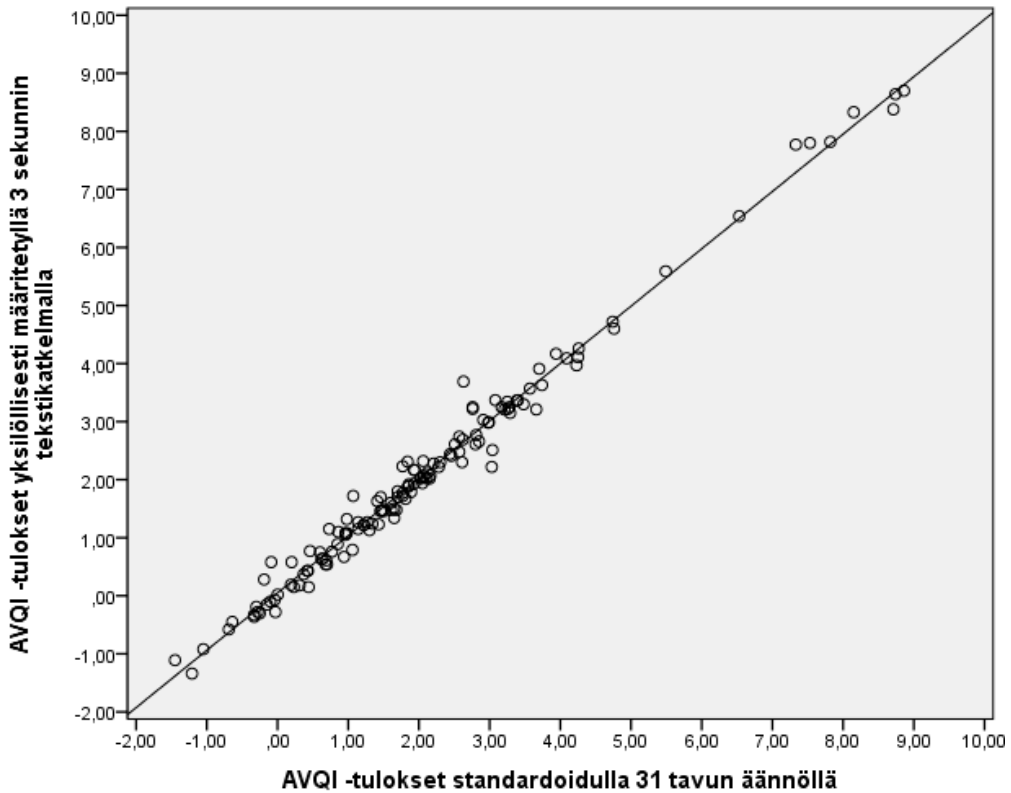
-arvojen yhteyttä Spearmanin (Spearman's rho,  $r_s$  ja  $r^2$ ) järjestyskorrelaatiokertoimella  $\rho$  (Frey, Botan, Friedman, & Kreps, 1991). Lopuksi indeksin diagnostista tarkkuutta arvioitiin testin erottelukykyä mittaavalla Roc-käyrällä (receiver operating characteristic), jota käytetään etenkin lääketieteellisessä tutkimuksessa (Forsström, 1995; Portney & Watkins, 2000).

Diagnostista tarkkuutta tutkittiin selvittämällä, miten herkästi AVQI 03.01FIN erottelee äänihäiriöäänät ja terveistä äänistä (sensitiivisyys) ja miten tarkasti se tunnistaa äänet, joissa ei ole häiriötä (spesifisyys). Ei-käheäksi häiriöttömäksi ääneksi määriteltiin ääni, jonka kuuntelu-arvion G:n keskiarvo sijoittui välille 0–0,49. AVQI 03.01FIN-luvun paras kynnsarvo terveen ja dysfonisen äänen välille määriteltiin ROC-käyrän ja Youden-indeksin avulla. Youden-indexillä laskettiin paras uskottavuusosamäärä sensitiivisyyden ja spesifisyyden suhteen (paras sensitiivisyys + spesifisyys – 1). Käyttökelpoisin kynnsarvo haettiin tasapainotetusti uskottavuusosamäärän positiivisten ( $LR+$ ) ja negatiivisten ( $LR-$ ) tulosten kautta (sensitiivisyys/(1–spesifisyys) ja (1–sensitiivisyys)/spesifisyys). Uskottavuusosamäärän katsotaan olevan korkea, kun  $LR+ \geq 10$  ja  $LR- \leq 0,1$  (Dollaghan CA 2007). Lisäksi AVQI 03.01FIN -luvun kykyä erotella normaali ja dysfoninen ääni arvioitiin ROC-käyrän alle jäävän pinta-alan avulla ( $A_{ROC}$ ). Täydellinen  $A_{ROC} = 1$ , ja kelvottoman testin raja on  $A_{ROC} = 0,5$  (Forsström, 1995; Portney & Watkins, 2000).

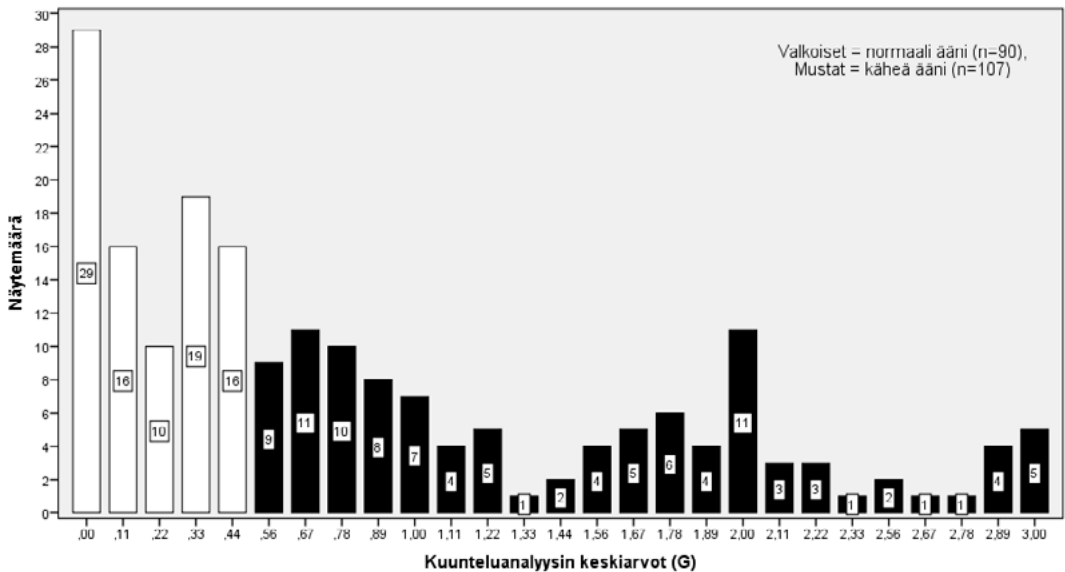
### 3 VALIDOINNIN TULOKSET

Kuuntelu-analyysin sisäinen yhtäpitävyys eli kuulijoiden samanmielisyyssä itsensä kanssa toistoarvioinnissa vaihteli paljon ( $C\kappa$ -kertoimen vaihtelu 0,26–0,92). Kahdella kuuntelijalla yhdenmukaisuutta kuvaava  $\kappa$ -kertoimen arvo jäi alle kohtalaisen ( $< 0,41$ ), minkä vuoksi nämä kuuntelijat poistettiin analyysistä. Seuraavaksi testattiin kuuntelijoiden välistä yhdenmukaisuutta, kohtalainen taso saavutettiin, kun kuuntelijoista poistettiin neljä ( $F\kappa = 0,50$ ,  $t = 17,559$ ;  $p = 0$ ). Lopulliseen analyysiin jäi 10 kuuntelijan arviot ääninäytteiden yleislaadusta. Kuvio 3 esittää annettujen kuuntelu-arvioiden keskiarvot ja keskiarvon mukaan ei terveiksi arvioidut näytteet ( $G KA > 0,49$ ).

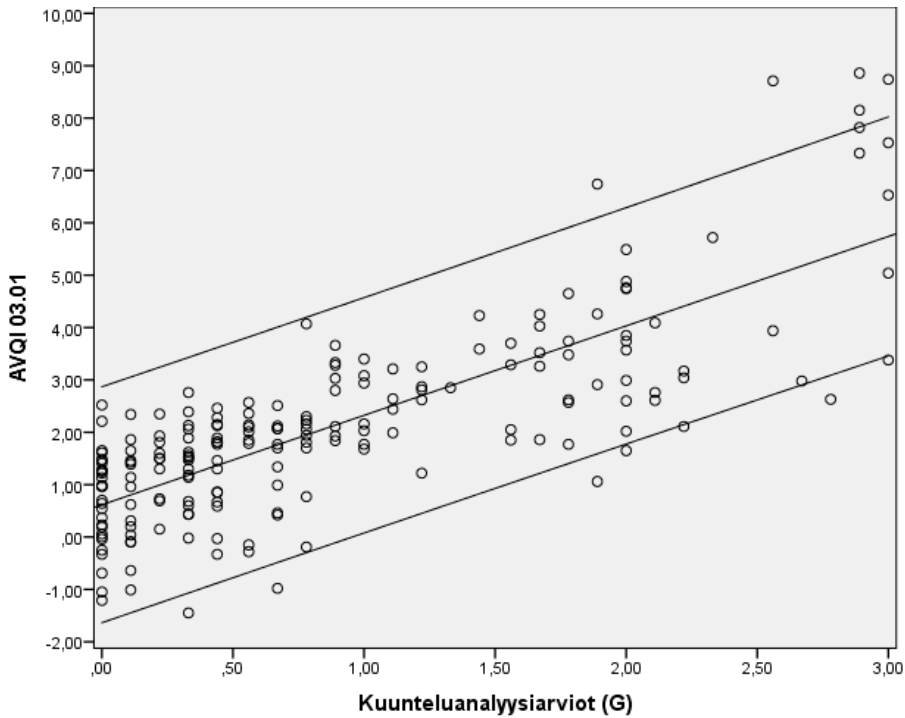
AVQI 03.01FIN -analyysin tulokset näkyvät taulukossa 2. Suomalaisen AVQI 03.01-version laskeman AVQI-tuloksen ja kuulohavaintoarvioiden välillä oli vahva yhteys (Spearmanin rho 0,77,  $p = 0,01$ ) (kuvio 4). Terveen ja häiriöisen äänen rajaa etsittiin ROC-analyysillä, joka osoitti, että AVQI 03.01:n paras erottelevuus ( $A_{ROC}$ ) 0,813 eli 81,3 % saatiin raja-arvolle 1,83, jolla oli myös korkein Youden-indeksi 0,635 (kuvio 5). Tällä raja-arvolla saavutettiin 81,3 % erottelukyky ja 82,2 % tarkkuus. Raja-arvon ylittävät arvot viittaavat häiriöiseen ääneen. Uskottavuusosamäärällä ei saavutettu yleisiä tilastollisia suositusarvoja ( $LR+ = 4,57$ ,  $LR- = 0,23$ ).



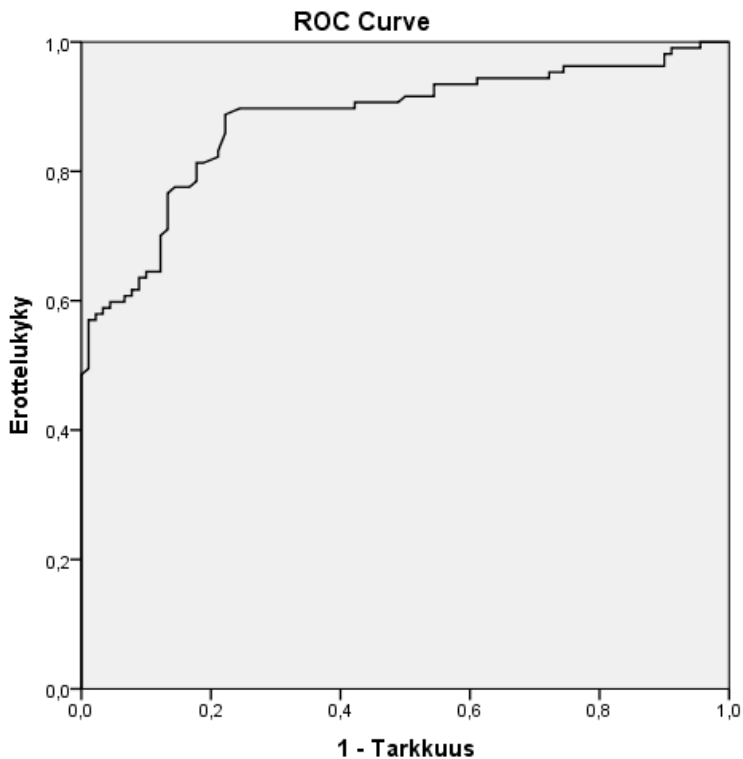
KUVIO 2. Pisteparvikuvaaja 31 tavulla laskettujen ja yksilöllisesti editoitujen kolmen sekunnin tavumäärällä laskettujen AVQI-tulosten välisestä korrelaatiosta ( $r = 0,99$ ).



KUVIO 3. Kuunteluanalyysin 10 kuuntelijan keskiarvotulokset. Pystyakselilla näytteiden määrä, vaaka-akselilla kuuntelutulosten keskiarvo. Kuunteluanalyysi G 0–0,49 = terve ääni (valkoiset pylväsdiagrammit), G 0,50–3 = äänihäiriöääni (tummat pylväsdiagrammit).



KUVIO 4. Pisteparvi ja lineaariset regressiolinjat kuunteluanalyysin ja AVQI 03.01FIN:n tulosten välillä. Keskimääräinen regressiosuorista selittää parhaiten kuunteluanalyysin ja AVQI:n suhdetta. Ylempi ja alempi suora kuvaavat kumpikin 95 %:n vaihteluvälin.



KUVIO 5. Roc -käyrä kuvaa AVQI-analyysin diagnostista erottelukykyä ja tarkkuutta.

TAULUKKO 2. Osallistujien AVQI-tulosten keskiarvo, hajonta, minimi ja maksimi.

	N	AVQI 03.01FIN KA	HA	Minimi	Maksimi
Kaikki osallistujat	197	2,12	1,83	-1,45	8,86
Terveet yliopistolta	86	1,13	1,04	-1,45	3,66
Foniatrian poliklinikan potilaat	111	2,89	1,95	-0,98	8,86

#### 4 POHDINTA

Tämän tutkimuksen tarkoituksena oli akustisen äänenlaatuindeksi AVQI:n sisäisesti yhdenmukaisen version 03.01 validointi suomenkielisille puhujille. Tutkimuksessa määritettiin suomen kielellä AVQI-analyysissä käytettäväksi luennan tavumääräksi 31 tavua. Suomen 31 tavua on lähellä aikaisemmissa tutkimuksissa eri kielillä saatujen tulosten keskiarvoa, kun hollannin tavumääräksi on määritelty 34 (Barsties & Maryn 2015), espanjan 33 (Hernández, Gómez, Jiménez, Izquierdo & Barsties v. Latoszek, 2018), japanin 30 (Hosakawa ym., 2019), ranskan (Hernández ym., 2018; Pommée ym., painossa) ja saksan 27 (Barsties v. Latoszek, Lehnert & Janotte, painossa).

Kuunteluanalyysin yhdenmukaisuus jäi tässä tutkimuksessa kohtalaiselle tasolle ja matalammaksi kuin muissa vastaavissa tutkimuksissa. Tutkimusessamme kuunteluanalyysiin osallistui ääneen erikoistuneita ammattilaisia kolmesta eri ammattiryhmästä. Aikaisemmissa muilla kielillä tehdyissä AVQI 03.01 -validoinneissa kuunteluryhmät ovat olleet pienempiä (3–5 henkilöä) ja homogeenisempia siten, että kuuntelijat ovat olleet vain puheterapeutteja (esim. Barsties & Maryn, 2015) tai kahdessa tutkimuksessa otolaryngologeja (Barsties v. Latoszek ym., painossa; Hosokawa ym., 2019). Toisaalta AVQI 02.02FIN -version validointitutkimuksen kuunteluraadissa oli vain yhden ammattiryhmän edustajia

(viisi puheterapeuttia), kuuntelun yhdenmukaisuus oli vain hiukan parempi ( $F_{\kappa} = 0,55$ ,  $t = 6,624$ ,  $p = 0,159$ ) (Kankare ym. 2019) kuin isommalla kuuntelijaryhmässämme tässä tutkimuksessa ( $F_{\kappa} = 0,50$ ,  $t = 17,559$ ;  $p=0$ ). Kohtalaiselle tasolle jäänyt kuuntelijoiden yhdenmukaisuus saattoi johtua myös siitä, että osa kuuntelijoista saattoi painottaa arviossaan jompaakumpaa ääninäytteen osaa, vaikka ohjeena oli molempien osien yhteisarvio. AVQI:n 03.01-version analysoi tasapuolisesti luenta- ja vokaalinäytteet ja antaa niistä yhden tuloksen.

Kuten aikaisemmissakin tutkimuksissa, myös tässä tutkimuksessa kuunteluanalyysin ja AVQI-tulosten välinen korrelaatio oli vahva. Korrelaatio (Spearmanin  $\rho = 0,77$ ,  $p = 0,01$ ) jäi arvoltaan matalammaksi kuin muilla kielillä tehdyissä AVQI 03.01 validointitutkimuksissa (Barsties & Maryn 2015; Barsties v. Latoszek ym., painossa; Hernández ym., 2018; Hosakawa ym., 2019; Pommée ym., painossa), mutta se oli parempi kuin AVQI 02.02FIN -validointitulosten korrelaatio (Spearmanin  $\rho = 0,74$ ,  $p = 0,01$ ) (Kankare ym., 2019).

Tutkimuksessa testattiin AVQI 03.01FIN -version diagnostista tarkkuutta ja validiteettia 31 tavun näytteillä käyttämällä Barstiesin ja Marynin (2015) laatimaa AVQI03.01 -laskentakaavaa. Tulokset varmistivat AVQI 03.01:n erotteluvan hyvin terveen ja äänihäiriöäänän raja-arvolla 1,83 (erottelukyky 81,3 %, tarkkuus 82,2 %). Raja-arvo on suomenkielisellä

AVQI 03.01 -versiolla (1,83) korkeampi kuin japanin kielen (1,41) 03.01-versiossa ja lähes yhtäläinen saksan raja-arvon (1,85) kanssa. Hollannin (2,43), ranskan (2,33) ja espanjan (2,28) kielillä validoiduissa versioissa terveen ja äänihäiriöäänänen AVQI tuloksen raja-arvot ovat korkeammat. Erojen AVQI-raja-arvoissa eri kielillä voidaan olettaa johtuvan eroista kielten foneettisessa rakenteesta.

Suomalaisen AVQI 03.01 -version diagnostinen terveen ja äänihäiriöäänänen erottelukyky oli hyvä (81,3 %). Se jäi alle alkuperäisen hollantilaisen (93,6 %) version erottelukyvyn, mutta oli selkeästi parempi kuin espanjalaisessa (72,1 %), japanilaisessa (64,6 %), saksalaisessa (72 %) ja ranskalaisessa (59,8 %) versiossa. Diagnostinen tarkkuus (82,2 %) jäi suomalaisversiossa kuitenkin alle kaikkien aikaisemmin validoitujen AVQI 03.01 -versioiden (Barsties & Maryn 2015; Hernández ym. 2018; Hosakawa ym., 2019; Pommée ym., painossa). Huonompi diagnostinen tarkkuus saattoi johtua tutkimuksessa olevasta suuresta adduktorisen spasmoidisen dysfonia potilaiden ryhmästä (23 osallistujaa). Muiden maiden AVQI 03.01 -validoinneissa on ollut spasmoidisia dysfoniapotilaita vähän (ranskalaisessa 3, hollantilaisessa 1) (Barsties & Maryn, 2015; Pommée ym., painossa;) tai ei lainkaan (Espanja, Japani, Saksa,) (Barsties v. Latoszek ym., painossa; Hernández ym. 2018; Hosakawa ym., 2019).

Adduktorisessa spasmoidisessa dysfoniassa potilaan äänen laatu voi olla pätkivä ja huojuva ja äänihuulivärähtelyyn voi tulla hetkellisiä katkoksia/pausseja, eikä ääni ole välttämättä käheä, karhea tai hälyinen, vaan monesti ääni on ponnisteisesti ja puristeisesti tuotettu. AVQI-analyysi ei analysoi katkoksia äänestä, mutta kuulonvaraisessa analyysissä ne vaikuttavat kokonaisarvioon. Aikaisemman

AVQI-version 02.02 validointitutkimuksessa (Kankare ym. 2019) todettiin, että suuri vaihtelu kuulonvaraisen arvion  $G = 3$  ja AVQI-tuloksen välillä oli juuri spasmoidinen dysfonia potilaiden ääninäytteiden kohdalla. Tutkimuksessa testattiin näiden näytteiden poistamisen vaikutusta kuulonvaraisen arvion ja AVQI-tulosten väliseen korrelaatioon. Näytteiden poistaminen nosti korrelaatiokerrointa vain yhden sadasosan ( $r$  nousi 0,74:stä 0,75:een). Vaikka kokeilu ei osoittanutkaan, että diagnostinen epätarkkuus olisi johtunut juuri spasmodisesta dysfoniasta kärsivien potilaiden ääninäytteistä, todennäköistä kuitenkin on, että AVQI-analyysi ei välttämättä tavoita kaikkia äänen laatuun liittyviä tekijöitä juuri tässä äänihäiriössä. Kliinisen kokemuksemme mukaan tämän potilasryhmän AVQI-tulosten tulkinta vaatii kuitenkin erityistä tarkkuutta. Tämän lehden Rantala, Ikävalko, Laukkanen & Kankare artikkelissa ”Akustinen äänenlaatuindeksi äänenlaadun arvioinnissa” pohditaan muun muassa äänentuottotavan vaikutusta AVQI:n analyysitulokseen. Tutkimuksessa todetaan, että huokoisesti ja hiljaa tuotettu ääni voi aiheuttaa AVQI-analyysiin niin kutsutun vääran positiivisen tuloksen.

AVQI 03.01FIN -validoinnissa kuunte-luanalyysin ja AVQI-tulosten sekä terveen ja äänihäiriöäänänen erottelukyky ja tarkkuus paranivat verrattuna AVQI 02.02FIN -version vastaaviin tuloksiin. Tällä perusteella voidaan suositella jatkossa käytettäväksi suomenkielisillä puhujilla nyt tasapainotetulla äänimateriaalilla (3 s jatkuva puhe ja 3 s vokaaliääntö) validoitua AVQI 03.01FIN -versiota. Tulevissa tutkimuksissa olisi hyvä selvittää, eroavatko erityyppiset äänihäiriöt AVQI-analyysillä toisistaan.

## 5 LOPUKSI

Tämä tutkimus validoi äänihäiriöpotilaiden arviointiin tarkoitetun akustisen äänenlaatuindeksin (AVQI) version 03.01 suomenkielisille puhujille. Menetelmää käytetään kliinisessä työssä maailmalla laajalti äänihäiriöpotilaiden äänen arviointiin ja terapian tuloksellisuuden seurantaan ja sitä voidaan suositella käytettäväksi myös Suomessa.

Suomen AVQI 03.01 -analyysiin valittiin käytettäväksi luentatekstistä Pohjantuuli ja aurinko tekstikatkelma ”Pohjantuuli ja aurinko väittelivät, kummalla olisi enemmän voimaa, kun he samalla näkivät”. Raja-arvoksi terveen ja äänihäiriöäänänen välille saatiin 1,83 (janalla 0–10). Tutkimus osoitti, että AVQI 03.01FIN on menetelmänä validi ja että se erottelee terveen äänen häiriöisestä hyvin.

## LÄHTEET

- Awan, S. N. & Roy, N. (2009). Outcomes measurement in voice disorders: Application of an acoustic index of dysphonia severity. *Journal of Speech, Language, and Hearing Research*, 52, 482–499.
- Barsties, B. & Maryn Y. (2015). The improvement of internal consistency of the Acoustic Voice Quality Index. *American Journal of Otolaryngology*. 36, 647–656.
- Barsties, B. & Maryn, Y. (2016). External validation of the Acoustic Voice Quality Index version 03.01 with extended representativity. *Annals of Otolaryngology, Rhinology & Laryngology*, 125, 571–583.
- Barsties v. Latoszek, B., Lehnert, B. & Janotte, B. Validation of the Acoustic Voice Quality Index Version 03.01 and Acoustic Breathiness Index in German. *Journal of Voice*, painossa.
- Boersma, P. & Weenink, D. (2013). Praat [computer program]: *Doing Phonetics by Computer (Version 5.3.57)*. Amsterdam, The Netherlands: Institute of Phonetic Sciences.
- Buder, E. H. (2000). Acoustic analysis of voice quality: A tabulation of algorithms 1902–1990. Teoksessa R. D. Kent ja M. J. Ball (toim.), *Voice quality measurement*, (s. 119–244). San Diego, CA: Singular.
- De Bodt, M. S., Wuyts, F. L., Van de Heyning, P. H. & Croux, C. (1997). Test-retest study of the GRBAS scale: Influence of experience and professional background on perceptual rating of voice quality. *Journal of Voice*, 11, 74–80.
- Deliyski, D. D., Shaw, H. S. & Evans, M. K. (2005). Adverse effects of environmental noise on acoustic voice quality measurements. *Journal of Voice*, 19, 15–28.
- Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore, USA: MD Books.
- Eadie, T. L. & Baylor, C. R. (2006). The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice. *Journal of Voice*, 20, 527–544.
- Forsström, J. (1995). Testien diagnostisen arvon mittaaminen ROC-käyrän avulla. *Lääketieteellinen aikakauskirja Duodecim*, 111, 237.
- Frey, L. R., Botan, C. H., Friedman, P. G. & Kreps, G. L. (1991). *Investigating communication: An introduction to research methods*. Englewood Cliffs, NJ, USA: Prentice Hall.
- Glaze, L. E., Bless, D. M. & Susser, R. D. (1990). Acoustic analysis of vowel and loudness differences in children's voice. *Journal of Voice*, 4, 37–44.

- Hammarberg, B., Fritzell, B., Gaufin, J., Sundberg, J. & Wedin, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities. *Acta Oto-Laryngologica*, 90, 441–451.
- Hernández, J.D., Gómez, NML., Jiménez, A., Izquierdo, LM. & Barsties v. Latoszek, B. (2018). Validation of the Acoustic Voice Quality Index Version 03.01 and the Acoustic Breathiness Index in the Spanish language. *Annals of Otolaryngology, Rhinology & Laryngology*, 127, 317–326.
- Hillenbrand, J., Cleveland, R. A. & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, 37, 769–778.
- Hirano, M. (1981). Psycho-acoustic evaluation of voice. Teoksessa G. E. Arnold, F. Winckel & B. D. Wyke B. D. (toim.), *Disorders of human communication 5. Clinical examination of voice*. (s. 81–84). Vienna: Springer-Verlag.
- Hosokawa, K., Barsties v Latoszek, B., Iwahashi, T., Iwahashi, M., Iwaki, S., Kato, C., Yoshida, M., Sasai, H., Miyachi, A., Matsushiro, N., Inohara, H., Ogawa, M. & Maryn, J. (2019). The Acoustic Voice Quality Index Version 03.01 for the Japanese-speaking Population. *Journal of Voice*, 33, 125e1–e12.
- Kankare, E., Barsties V. Latoszek, B., Maryn, Y., Asikainen, M., Rorarius, E., Vilpas, S., Ilomäki, I., Tyrmi, J., Rantala, L. & Laukkanen, A. M. (2019). The Acoustic Voice Quality Index Version 02.02 in the Finnish-speaking Population. *Logopedics Phoniatrics Vocology*, 30, 1–8.
- Kleemola, L., Helminen, M., Rorarius, E., Isotalo, E. & Sihvo M. (2011). Voice Activity and Participation Profile in assessing the effects of voice disorders on quality of life: Estimation of the validity, reliability and responsiveness of the Finnish version. *Folia Phoniatrica et Logopaedica*, 63, 113–121.
- Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A. & Berke, G. S. (1993). Perceptual evaluation of voice quality review, tutorial, and a framework for future research. *Journal of Speech, Language, and Hearing Research*, 36, 21–40.
- Karnell, M. P., Melton, S. D., Childes, J. M., Coleman, T. C., Dailey, S. A. & Hoffman, H. T. (2007). Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *Journal of Voice*, 21, 576–590.
- Koike, Y. (1973). Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Studia Phonologica*, 7, 17–23.
- Landis, J. R & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
- Lieberman, P. (1963). Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. *The Journal of the Acoustical Society of America*, 35, 344–353.
- Maryn, Y., Corthals, P., Van Cauwenberge, P., Roy, N. & De Bodt, M. (2010). Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *Journal of Voice*, 24, 540–555.
- Maryn, Y., De Bodt, M. & Roy, N. (2010). The Acoustic Voice Quality Index: Toward improved treatment outcomes assessment in voice disorders. *Journal of Communication Disorders*, 43, 161–174.
- Maryn, Y., Roy, N., De Bodt, M., Van Cauwenberge, P. & Corthals, P. (2009). Acoustic measurement of overall voice quality: A meta-analysis a. *The Journal of the Acoustical Society of America*, 126, 2619–2634.
- Oguz, H., Demirci, M., Safak, MA., Arslan, N., Islam, A. & Kargin, S. (2007). Effects of unilateral vocal cord paralysis on objective voice measures obtained by Praat. *European Archives of Oto-Rhino-Laryngology* 264, 257–261.
- Pommée, T., Maryn, Y., Finck, C. & Morsomme, D. Validation of the Acoustic Voice Quality Index, Version 03.01, in French. *Journal of Voice*, painossa.
- Portney, LG. & Watkins, MP. (2000). *Foundations of clinical research, applications to practice. 2.th ed.* Prentice Hall Health, Upper Saddle River, NJ, USA.
- Rabinov, C. R., Kreiman, J., Gerratt, B. R. & Bielamowicz, S. (1995). Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *Journal of Speech, Language, and Hearing Research*, 38, 26–32.



- Revis, J., Giovanni, A., Wuyts, F., & Triglia, J. M. (1999). Comparison of different voice samples for perceptual analysis. *Folia Phoniatica et Logopaedica*, 51, 108–116.
- Rosen, C. A., Lombard, L. E. & Murry, T. (2000). Acoustic, aerodynamic, and videostroboscopic features of bilateral vocal fold lesions. *Annals of Otolaryngology, Rhinology & Laryngology*, 109, 823–828.
- Roy, N., Barkmeier-Kraemer, J., Eadie, T., Sivasankar, M. P., Mehta, D., Paul, D. & Hillman, R. (2013). Evidence-based clinical voice assessment: A systematic review. *American Journal of Speech-Language Pathology*, 22, 212–226.
- Roy, N., Gouse, M., Mauszycki, S. C., Merrill, R. M. & Smith, M. E. (2005). Task specificity in adductor spasmodic dysphonia versus muscle tension dysphonia. *Laryngoscope*, 115, 311–316.
- Stemple, J. C., Roy, N. & Klaben, B. K. (2014). *Clinical voice pathology: Theory and management* (5. painos). San Diego: Plural Publishing.
- Wuyts, F. L., De Bodt, M. S., Molenberghs, G., Remacle, M., Heylen, L., Millet, B., Van Lierde, K., Raes, J. & Van de Heyning, P. H. (2000). The dysphonia severity index: An objective measure of vocal quality based on a multiparameter approach. *Journal of Speech, Language, and Hearing Research*, 43, 796–809.
- Yu, P., Revis, J., Wuyts, F. L., Zanaret, M. & Giovanni, A. (2002). Correlation of instrumental voice evaluation with perceptual voice analysis using a modified visual analog scale. *Folia Phoniatica et Logopaedica*, 54, 271–281.
- Zraick, R. I., Wendel, K. & Smith-Olinde, L. (2005). The effect of speaking task on perceptual judgment of the severity of dysphonic voice. *Journal of Voice*, 19, 574–581.

## VALIDATION OF THE ACOUSTIC VOICE QUALITY INDEX 03.01FIN (AVQI) IN FINNISH SPEAKING POPULATION

*Elina Kankare, Department of Phoniatics, Tampere University Hospital*

*Leena Rantala, Speech and Voice Research Laboratory, Unit of Logopedics, Faculty of Social Sciences, Tampere University*

*Tero Ikävalko, Speech and Voice Research Laboratory, Faculty of Social Sciences, Tampere University*

*Ben Barsties v. Latoszek, Speech-Language Pathology, SRH University of Applied Health Sciences, Düsseldorf, Germany and Department of Phoniatics and Pediatric Audiology, University Hospital Münster, Westphalian Wilhelm University, Münster, Germany*

*Anne-Maria Laukkanen, Speech and Voice Research Laboratory, Faculty of Social Sciences, Tampere University*

The Acoustic Voice Quality Index (AVQI) is an objective tool based on six acoustic parameters. It uses sustained vowel and continuous speech in the analyses and therefore it must be validated in different languages. In this study, the newest version of AVQI (03.01) with an extended continuous speech sample and improved internal consistency was validated to Finnish-speaking population. The study included 197 native Finnish-speaking voluntary participants, out of which 111 were patients from a phoniatic clinic and 86 were healthy voice users. A sustained vowel and a reading sample were recorded. Mean number of the syllables comparable to the 3 second sustained vowel was calculated from the reading samples. Sixteen voice specialists evaluated the overall voice quality of the voice samples with a four-point scale. Statistic analyses were performed to test the diagnostic accuracy between healthy and disordered voices in Finnish-speaking population. The number of syllables, comparable to 3 seconds of sustained vowel, was 31. The correlation between the AVQI scores and the overall voice quality was strong (Spearman's rho 0.77,  $p=0.01$ ). The AVQI score 1.83 was the best to distinguish healthy and dysphonic voices. The study confirmed the AVQI03.01FIN version to be a good tool in voice disorder diagnostics in Finnish speaking population.

**Keywords:** acoustic analysis, dysphonia, Finnish language, perceptual evaluation, voice disorder