

Kaisa Tivonen

**LOGOPEDIAN OPISKELIJOIDEN VÄLINEN
YHTENEVYYS LASTEN R-ÄÄNTEEN
LUOKITTELUSSA**
Reliabiliteettitutkimus

Yhteiskuntatieteiden tiedekunta
Logopedian kandidaattitutkielma
Kesäkuu 2023

TIIVISTELMÄ

Kaisa Tivonen: Logopedian opiskelijoiden välinen yhtenevyys lasten r-äänteen luokittelussa: Reliabiliteettitutkimus
Kandidaatintutkielma
Tampereen yliopisto
Logopedian tutkinto-ohjelma
Kesäkuu 2023

Puheterapian heikko saatavuus Suomessa on johtanut huoltajien keskeiseen rooliin lasten äänneharjoittelussa. Samanaikaisesti kuntoutusteknologioiden käyttö on yleistynyt myös puheterapiakuntoutuksessa. Esimerkiksi mobiilipelit ja niiden osana automaattinen puheentunnistus (englanniksi *Automatic Speech Recognition, ASR*) voivat tarjota käyttäjilleen palautetta ja näin tukea harjoittelun toteuttamista. Jotta tekoälyä ja automaattista puheentunnistusta voidaan puheterapeuttisessa kontekstissa hyödyntää, on niille aluksi tarjottava opetusdataa siitä, miten automaattisen puheentunnistuksen toivotaan toimivan, minkä lisäksi kulttuuriset ja kielierot on huomioitava. Jotta automaattinen puheentunnistus voi toimia tarkoituksenmukaisesti, on sille tarjottavan opetusdatan oltava luotettavaa ja tasalaatuista. Tätä varten opetusaineiston luokittelijoiden keskinäistä arvioitsijareliabiliteettia on syytä tarkastella.

Tämän tutkimuksen tarkoituksena on selvittää, kuinka yhdenmukaisesti logopedian 3.vuoden opiskelijat luokittelevat suomalaislasten /r/-äänteitä. Aineisto koostui kolmen logopedian opiskelijan luokittelemista puhenäytteistä, ja se toimi opetusdatana kehitteillä olevalle /r/-äänteen harjoitteluun suunnatulle puheentunnistusta hyödyntävälle mobiilipelille. Yhden sanan mittaiset puhenäytteet olivat peräisin TEFLON-projektiin osallistuneilta yhdeltätoista 5–10-vuotiaalta suomalaislapselta ja niitä oli yhteensä 286 kappaletta. Opiskelijoiden käyttämä luokitteluasteikko oli kuusiportainen, ja sen luokat oli luotu mobiilipeliä varten. Opiskelijoiden luokittelema aineisto analysoitiin hyödyntäen Fleissin Kappa-kerrointa.

Logopedian opiskelijoiden luokittelun yhtenevyys kaikki luokat huomioiden oli $\kappa = .68$ (95 % LV .64–.71), $p < .001$. Käytettäessä arvon tulkintaan perinteistä Cohenin Kappa-kertoimen tulkintatapaa voidaan yhtenevyyden tulkita olevan hyvällä tasolla. Yksittäisiä luokkia tarkasteltaessa Kappa-kertoimen arvo vaihteli välillä $\kappa = .37$ –.88. Tulokset logopedian opiskelijoiden arvioitsijareliabiliteetin tasosta on hyvin samansuuntaisia aikaisemman tutkimuskirjallisuuden kanssa, vaikka niissä luokiteltava kohde on ollut jokin muu. Kansallista tutkimustietoa puheterapeuttien ja puheterapeuttiopiskelijoiden arviointien yhtenevyydestä on niukasti. Arvioitsijareliabiliteettia voitaisiin jatkossa tarkastella erityisesti sellaisissa kliinisesti yleisissä puheterapeuttisissa tilanteissa, joihin ei ole olemassa standardoituja testejä. Kuntoutusteknologioiden yleistyminen ja kuntoutujien oman aktiivisen roolin korostuminen kuntoutusprosessissa koskee myös puheterapia-alaa. Automaattinen puheentunnistus ja erityisesti sen tarjoama palaute voivat kannustaa asiakasta itsenäiseen harjoitteluun ja tukea kuntoutuksen tavoitteiden toteutumista. Toimiakseen puheterapeuttisissa sovelluksissa tarkoituksenmukaisesti automaattinen puheentunnistin tarvitsee kieli- ja tavoitekohtaisesti räätälöityä opetusdataa. Sen muodostamiseen tarvitaan yhdenmukaista puheterapeuttista osaamista, minkä vuoksi arvioitsijareliabiliteetin tarkastelu on tärkeää.

Avainsanat: artikulaatiokuntoutus, äännevirhe, automaattinen puheentunnistus, reliabiliteetti, puheterapeuttiopiskelija, mobiilipeli

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

SISÄLLYSLUETTELO

1	JOHDANTO	1
2	TUTKIMUKSEN TEOREETTISIA LÄHTÖKOHTIA	2
2.1	Artikulaatio	2
2.1.1	Suomalaislasten äänteellinen kehitys	2
2.1.2	/r/-äänteen variaatiot.....	3
2.2	Artikulaatiokuntoutus.....	4
2.2.1	Artikulaatiokuntoutuksen perusteet	4
2.2.2	Teknologiat artikulaatiokuntoutuksessa.....	5
2.3	Automaattinen puheentunnistus	6
2.3.1	Automaattisen puheentunnistuksen perusteet	6
2.3.2	Automaattinen puheentunnistus puheterapiassa	7
2.4	Reliabiliteetti puheterapiassa	8
3	TUTKIMUKSEN TAVOITE JA TUTKIMUSKYSYMYS.....	10
4	TUTKIMUSMENETELMÄT.....	11
4.1	Tutkimushenkilöt	11
4.2	Aineiston muodostuminen	11
4.3	Aineiston analyysi.....	12
4.4	Tutkimuksen eettisyys.....	12
5	TULOKSET	14
6	POHDINTA	16
6.1	Tulosten tarkastelu	16
6.2	Menetelmän pohdinta.....	18
6.3	Työn kliininen merkitys ja jatkotutkimusaiheita.....	19
	LÄHDELUETTELO	21

1 JOHDANTO

Erilaiset teknologiset sovellukset ovat tulleet osaksi terveydenhuoltoa ja kuntoutusta viime vuosikymmeninä (Saeedi ym., 2022; Salminen ym., 2016a, s. 9). Kuntoutuksessa sovelluksia käytetään muun muassa toimintakyvyn tutkimiseen, kuntoutuksen etenemisen seurantaan sekä itse kuntoutuksen toteutukseen esimerkiksi erilaisia interaktiivisia pelejä hyödyntäen (Salminen ym., 2016b, s. 11–13). Myös puheterapiakuntoutuksessa erilaiset pelit ovat yleistyneet, ja monet niistä tarjoavat käyttäjälleen palautetta suoriutumisesta hyödyntäen automaattista puheentunnistusta (englanniksi *Automatic Speech Recognition, ASR*) (Saeedi ym., 2022).

Valtakunnallinen pula puheterapeuteista ja kasvanut puheterapian kysyntä on johtanut siihen, että puheterapiaa on suunnattu haastavampien puhehäiriöiden kuntoutukseen, ja yksittäisen artikulaatiovirheen kuntoutukseen voi olla vaikeampi saada puheterapiaa (Pitkänen, 2020; Yle, 2023). Puheterapian saatavuushaasteet johtavat huoltajien roolin korostumiseen äänneharjoittelussa. Artikulaatiokuntoutuksessa aisteilla ja niiden kautta saatavalla palautteella on keskeinen merkitys (Savinainen-Makkonen & Kunnari, 2012, s. 179). Perinteisesti suullisesta palautteesta on vastannut lapsen kanssa harjoitteita tekevä aikuinen. Automaattista puheentunnistusta hyödyntävät sovellukset voivat tukea harjoittelua esimerkiksi tarjoten mallin oikeanlaisesta tuotoksesta, kannustaen kuuloerotteluun toistamalla lapsen puhunnoksen sekä antamalla palautetta. Kuntoutuksessa on parhaat edellytykset edetä silloin, kun kuntoutuja on itse motivoitunut työskentelemään tavoitetta kohti (Ahmed ym., 2018; Savinainen-Makkonen & Kunnari, 2012, s. 178). Digitaalinen pelialusta voi perinteisen artikulaatiokuntoutuksen rinnalla tarjota mahdollisuuden mielekkääseen harjoitteluun. On havaittu, että pelit voivat lisätä lapsen kiinnostusta ja sitoutumista harjoitteluun, vahvistaa pystyvyyden tunnetta sekä olla nautinnollisia (Malaquias & Malaquias, 2021). Tämä edellyttää kuitenkin peliltä helppokäyttöisyyttä, interaktiivisuutta, haasteellisuutta, palautetta ja palkkioita.

Vaikka pelisovelluksissa tekoäly huolehtiikin puheentunnistuksesta, on sen äänneharjoittelussa toimiakseen aluksi opittava, miltä tavoiteltu äänne kuulostaa ja millaisia variaatioita äänten harjoittelu voi sisältää (Ylinen & Kurimo, 2017, s. 64). Kielten äännejärjestelmät ovat erilaisia, ja sanasto sekä kunkin kielen haastavimmat äänteet kieli- ja kulttuurisidonnaisia. Tekoälylle tarjottavan opetusaineiston on oltava luotettavaa ja riittävän yhdenmukaista, mihin tarvitaan puheterapeuttista asiantuntemusta. Tässä tutkielmassa tarkastellaan logopedian 3.vuoden opiskelijoiden tekemän /r/-äänneluokittelun yhteneväisyyttä. Luokiteltava aineisto toimii automaattisen puheentunnistuksen opetusmateriaalina pilotoitaessa pelisovellusta /r/-äänneharjoitteluun.

2 TUTKIMUKSEN TEOREETTISIA LÄHTÖKOHTIA

2.1 Artikulaatio

2.1.1 Suomalaislasten äänteellinen kehitys

Artikulaatiolla tai ääntämisellä tarkoitetaan puheäänteiden tuottamista ja käyttämistä (Korpinen & Nasreddin, 2011, s. 60). Foneemi puolestaan on pienin äännejärjestelmän yksikkö, joka erottaa merkityksiä (Tieteen termipankki, 2021). Esimerkiksi suomen kielessä /r/- ja /l/-äänteet erottavat sanojen merkityksiä, kuten "rokki" ja "lokki". Puheen selkeys voidaan määritellä esimerkiksi niin, että puhujalla on käytössään sellainen foneeminen variantti, ettei sekoittumista toiseen foneemiin tapahdu (Savinainen-Makkonen & Kunnari, 2012, s. 173).

Kielen ja puheen kehitys on yksilöllistä, ja varsinkin varhaisessa vaiheessa kielelliset taidot voivat poiketa toisistaan paljonkin samanikäisillä lapsilla (Savinainen-Makkonen & Kunnari, 2009, s. 114). Lapsi alkaa äännellä heti syntymänsä jälkeen, ja esimerkiksi jokeltelua pidetään tärkeänä myöhemmän kielenkehityksen kannalta (Törölä, 2009, s. 107). Puheen ja puhunnosten vaihtelevuus on osa pienten lasten normaalia puheen kehitystä (esimerkiksi Macrae, 2013; Sosa, 2015). Lapsi oppii äidinkielen äänteitä kuuntelemalla, katselemalla sekä matkimalla ympäristön mallia (Korpinen & Nasreddin, 2011, s. 61).

Moninaiset, hiljalleen kehittyvät verbaalimotoriset taidot kehittyvät lapsen kasvaessa, eikä siksi pienten lasten kohdalla ole perusteltua puhua äännevirheistä (Savinainen-Makkonen & Kunnari, 2012, s. 173). Lapsi oppii tyypillisesti ensin vokaaliäänteitä, ja hallitsee ne pääosin noin kolmivuotiaana (Kunnari ym., 2012). Nelivuotiaana lapsen puhetta ymmärtävät yleensä vieraatkin, vaikka äänneinventaarista vielä joitakin äänteitä puuttuisikin (Savinainen-Makkonen & Kunnari, 2009, s. 119). Larmolan mukaan (1982) äänteiden puuttumista tai poikkeavuutta on reilulla kolmanneksella 4–5-vuotiaista. Aalto (2023) havaitsi väitöskirjatutkimuksessaan 3;6-vuotiaiden suomalaislasten puhuvan selkeästi – tyypillisimmin puuttuvia äänteitä olivat harvoin suomen kielessä esiintyvä /d/ sekä motorisesti haastava /r/, minkä lisäksi sananalkuiset konsonanttiyhtymät olivat lapsille haastavia. Puheen selkeys oli merkittävästi yhteydessä kuulonvaraiseen erotteluun, tuottavan sanaston kokoon sekä lukivalmiustaitoihin.

2.1.2 /r/-äänteen variaatiot

Yleisin puheterapeutin arvioon tulossyy lapsilla on normeista poikkeava artikulaatio eli niin sanottu artikulaatiovirhe (myös dyslalia, äännevirhe, foneettinen virhe) (Korpinen & Nasredtin, 2011 s. 60; Lass & Pannbacker, 2008). Se tarkoittaa äänten tuottamista jollakin äidinkielestä poikkeavalla variantilla, äänten korvaamista toisella äänneellä tai äänten puuttumista (Savinainen-Makkonen & Kunnari, 2012, s. 174). Äännevirhe on kuitenkin syytä erottaa tavanomaiseen kehitykseen kuuluvasta vaihtelusta (Savinainen-Makkonen & Kunnari, 2009, s. 118–119). Äänneistä, niiden varianteista ja äännevirheistä puhuttaessa on hyvä tiedostaa, että ne ovat kieli- ja kulttuurisidonnaisia. Van Riper (1979) määrittelee ääntämisen virheelliseksi silloin, jos puheen poikkeavuus hankaloittaa kommunikointia, siitä seuraa puhujalle mielipahaa tai se herättää huomiota. Yksiselitteistä syytä äännevirheen kehittymiselle ei ole kyetty nimeämään, mutta perinteisesti sitä on pidetty puhemotoriikan pulmana (Fey, 1992; Korpinen & Nasredtin 2009, s. 74). Taustalla voivat olla esimerkiksi kuuloerottelun haasteet tai ääntöelinten eli kielen, suulaen, huulten ja kurkunpään liikkeiden puutteellinen hallinta sijoittumisen, liikkeen suunnan, sen nopeuden tai ajoittamisen suhteen (Korpilahti, 2011, s. 42; Korpinen ja Nasredtin, 2011, 64–65).

Suomen kielessä erityisesti /r/- ja /s/-äänteet ovat motorisesti haastavia (Korpinen & Nasredtin, 2011, s. 60; Martikainen ym. 2019; Savinainen-Makkonen & Kunnari, 2012). /r/-äänne tuotetaan suomen kielessä yläetuhampaiden takana hammasvallilla, eli kyseessä on alveolaarinen äänne (Korpinen & Nasredtin, 2009, s. 66). Ääntötapa on tremulantti eli täryäänne, joka tuotetaan kielenkärjellä (Savinainen-Makkonen, 2012). Kyetäkseen tuottamaan /r/-äänteen lapsen on samanaikaisesti kyettävä hallitsemaan uloshengitystä ja lihasjännitystä (s. 173–175; 180).

5-vuotiaista suomalaislapsista noin kolmannes ja 6-vuotiaista reilu kymmenes tuottaa /r/-äänteen suomen kielestä poikkeavasti (Kunnari ym., 2012). Poikkeavat tuottotavat voidaan jakaa foneemiin ja foneettisiin virheisiin (Korpinen & Nasredtin, 2009, s. 66). Foneemisessa äännevirheessä äänne korvautuu toisella, esimerkiksi /r/-äänne /l/-äänneellä, joskus myös vokaalilla (esimerkiksi [puakka] *purkka*) (Savinainen-Makkonen & Kunnari, 2012, s. 174). Foneettinen virhe on kyseessä silloin, kun tuotettu äänne ei ole suomen kielen mukainen, mutta se ei kuitenkaan sekoitu toiseen suomen kielen foneemiin (Korpinen & Nasredtin, 2009, s. 66). Äänne voi myös omissoitua, eli lapsi voi jättää sen kokonaan sanomatta (esimerkiksi [otta] *rotta*) (Savinainen-Makkonen & Kunnari, 2012). Mitä vanhempi lapsi on, sitä epätodennäköisempää täysi omissio on, sillä lapset ovat hyvin kekseliäitä pyrkiessään korvaamaan haastavasti äännettävää /r/-äännettä (s. 174; 177).

Tyypillisimpiä foneettisia variaatioita /r/-äänteen tuottamiseksi ovat frikatiivinen, uvulaarinen, lateraalinen ja bukkaalinen (Korpinen-Nasredtin, 2009, s. 66; Savinainen-Makkonen & Kunnari, 2012, s. 177). Frikatiivisessa /r/:ssä (ð) syntyy täryn sijaan hankaushälyä (Savinainen-Makkonen & Kunnari, 2012). Tällöin kieli voi olla oikealla ääntöpaikalla ylätuhammasvallilla tai se voi työntyä hampaiden väliin (interdentaalinen). Uvulaarisessa (ʀ), niin sanotussa kurkkuärrässä on täry, mutta se tehdään kielen kärjen sijaan kitakielekkeellä eli uvulalla kurkussa. Lateraalinen /r/ viittaa tilanteeseen, jossa täry muodostetaan kielen laidoilla. Bukkaalisessa /r/-äänteessä täry muodostetaan poskessa (s. 177).

2.2 Artikulaatiokuntoutus

2.2.1 Artikulaatiokuntoutuksen perusteet

Yksittäinen poikkeavasti tuotettu äänne ei yleensä heikennä puheen ymmärrettävyyttä, mutta yhteiskunnalliset normit ja odotukset luovat usein paineita ja halua oppia tuottamaan äänne ”oikein” eli suomen kielen mukaisesti (Savinainen-Makkonen & Kunnari, 2012, s. 186). Äännevirheet ovatkin yleisin syy puheterapiaan hakeutumiseen lasten kohdalla (Lass & Pannbacker, 2008; Mullen & Schooling, 2010; Savinainen-Makkonen & Kunnari, 2012). Esikouluikä ja ensimmäiset kouluvuodet ovat otollista aikaa yksittäisten äänteiden harjoitteluun, sillä tuolloin lapsi on usein tietoinen puheensa puutteista ja ympäristön normeista, kiinnostunut kielen rakenteista ja motivoitunut harjoittelemaan (Korpinen & Nasredtin, 2009, s. 60, 66; Savinainen-Makkonen & Kunnari, 2012, s. 178). Savinainen-Makkosen ja Kunnarin (2012, s. 178) mukaan artikulaatioterapiaa tarvitsevat ensisijaisesti lapset, joiden mahdollisuuksia ilmaista itseään poikkeuksellinen artikulaatio haittaa. Päätöstä puheterapian saatavuudesta ohjaavat käytännössä resurssit, jolloin asiakasta ja hänen tarpeitaan suhteutetaan muihin asiakkaisiin ja heidän tarpeisiinsa. Kruegerin (2019) mukaan lapsen kokemusta sosiaalisesta haitasta kuitenkin usein aliarvioidaan tehtäessä päätöstä kuntoutustarpeesta.

Puheterapiapalveluiden vähäisten resurssien vuoksi kaikki halukkaat eivät saa puheterapiaa, joten esimerkiksi vanhemman rooli artikulaatio-ohjauksessa on keskeinen (Savinainen-Makkonen, 2012). Tämä ei sinänsä ole ongelmallista, sillä artikulaatio-ohjausta voi antaa kuka tahansa aikuinen, joka hallitsee periaatteet äänteen oppimisesta ja jolla on aikaa lapselle. Ensimmäisinä kouluvuosina myös erityisopettaja voi tarjota artikulaatio-ohjausta (puheopetusta) esimerkiksi ryhmämuotoisena. Kun harjoittelu aloitetaan, keskeisiä tekijöitä ovat halu oppia, kiinnostus sekä riittävä määrä toistoja.

Äännettä opeteltaessa on perusteltua hyödyntää kaikkia aisteja – kuuloa, liike- ja kosketustuntoaistia sekä näköä (s. 178–179; 187). Onnistumisen kokemukset ovat tärkeitä, joten tavoitetta pilkotaan pieniksi välitavoitteiksi, ja toteutuksessa hyödynnetään lapsen kiinnostuksen kohteita (Korpinen & Nasredtin, 2009, s. 71–72; Savinainen-Makkonen, 2012, s. 178).

Kotimaista tutkimustietoa artikulaatioterapiamenetelmien vaikuttavuudesta ei ole tarjolla, ja yleisestikin tietoa niiden kuntoutuksesta on niukasti (Savinainen-Makkonen & Kunnari, 2012). Menetelmät pohjautuvat kliinisessä työssä tehtyihin huomioihin. Varsinaiset ääntämisharjoitteet aloitetaan yleensä kielen ja suun alueen motorisilla harjoitteilla, sillä yleisesti ajatellaan /r/:n onnistumisen vaativan onnistuakseen kielen ja leuan liikkeiden eriytymistä (s. 180–181). Vaikka oraalmotoriikan harjoitteet ovat laajasti käytettyjä, ei näyttö niiden käytön hyödyllisyydestä artikulaatiokuntoutuksessa ole kiistatonta (esimerkiksi Lass & Pannbacker, 2008; Lof & Watson, 2008). On syytä tarkistaa myös lapsen kuuloerottelu, ja vahvistaa sitä tarpeen mukaan (Korpinen & Nasredtin, 2009, s. 72). Apuäänteenä käytetään usein /d/-äännettä ja sen avulla tehdään erilaisia harjoituksia, sillä ne aktivoivat kieltä napakkaan ja täsmälliseen liikkeeseen ja vahvistavat /r/:n ääntöpaikkaa (Savinainen-Makkonen & Kunnari, 2012, s. 181). Kuntoutuksessa edetään vaiheittain äänteen oppimisen edellytyksien hallitsemisesta kohti äänteen hallintaa apuäänteen avulla, sitten yksittäisenä äänteenä, tuottamiseen sanoissa eri positioissa sekä viimein kohti yleistymistä kaikkiin sanoihin sekä spontaanipuheeseen (Brenthal, ym. 2009; Savinainen-Makkonen, 2012). Perinteisesti artikulaatioterapiaa ohjataan kotona tehtävästi erilaisten monisteiden ja niissä olevien pelien avulla.

2.2.2 Teknologiat artikulaatiokuntoutuksessa

Maailmanlaajuinen pula puheterapeuteista aiheuttaa osaltaan tarvetta kehittää tietokonepohjaisia puheterapeuttisia menetelmiä. Furlong ja kumppanit (2017) tarkastelivat katsauksessaan tietokonepohjaisten puheterapeuttisten ohjelmien tehokkuutta lasten puheen äänteellisen epäselvyyden kuntoutuksessa. Katsauksen tulosten mukaan tietokonepohjainen puheterapia on yhteydessä positiivisiin klinisiin muutoksiin osalla lapsista, joiden puhe on epäselvää. Furlong ja muut (2017) korostavat, että moniammatillista yhteistyötä tarvitaan erityisesti sovellusten kehittämisvaiheessa. Erityisesti pelisovelluksissa harjoittelun mielekkyys ja aktiivinen toimijarooli voivat lisätä lapsen kiinnostusta ja sitoutumista harjoitteluun (Malaquias & Malaquias, 2021). Peleihin kohdistuu kuitenkin myös vaatimuksia, sillä niiden pitäisi olla helppokäyttöisiä ja sopivan haasteellisia, niiden pitäisi tarjota palkkioita ja palautetta sekä olla interaktiivisia.

Wrenin ja Roulstonen (2008) artikulaatiokuntoutusta tarkastelevassa tutkimuksessa vertailtiin perinteistä monisteharjoittelua ja tietokonepohjaista harjoittelua, minkä lisäksi mukana oli verrokkiryhmä, joka ei saanut kuntoutusta. Interventioryhmissä kuntoutuksen toteuttajana oli puheterapeutti, ja niissä tehtiin samankaltaisia harjoitteita joko monisteen tai tietokonepelin avulla. Tutkimuksessa ei havaittu merkitseviä eroja ryhmien välillä, mutta jokainen ryhmistä edistyi. Jesuksen ja muiden (2019) tulokset olivat samansuuntaisia monisteharjoittelun ja tablettisovelluksen suhteen – lapset, joilla oli puheen äänteellistä epäselvyyttä, harjoittelivat puheterapeutin kanssa joko monisteen tai tablet-tietokoneen avulla, ja molemmat ryhmistä edistyivät.

Findley ja Gasparyan (2022) hyödynsivät tutkimuksessaan automaattisen puheentunnistuksen tarjoamaa mahdollisuutta litteroida puhuttu tekstiksi ja tarjota näin palautetta lapselle hänen puhunnoksestaan. Tutkimukseen osallistui kolme 7–9-vuotiasta lasta, joilla oli jokin äännevirhe. Jokaisen osallistujan kohdefoneemin tuotto parantui harjoittelujakson aikana sekä seurannassa. McAllister Byun (2017) tutki visuaalisen ja akustisen palautteen vaikuttavuutta osana äännekuntoutusta ja havaitsi, että visuaalinen ja akustinen palaute voivat parantaa intervention tehoa osalla lapsista.

Ahmed ja muut (2018) kartoittivat tutkimuksessaan lasten ja puheterapeuttien kokemuksia puheentunnistusta hyödyntävästä mobiilipelistä. Tutkimukseen osallistui lapsia, joilla oli puheen dyspraksiaa, tavanomaisesti kehittyneitä verrokkeja sekä puheterapeutteja. Lapset kokivat pelien olevan kiinnostavia ja hauskoja, ja erityisesti he pitivät pelissä palkinnoista ja eritasoisista haasteista. Puheterapeutit kokivat, että peleistä voisi olla erityisesti hyötyä kotiharjoittelun muotona, minkä lisäksi he pitivät tärkeinä seikkoina puheentunnistuksen tarjoamaa palautetta ja mahdollisuutta kuunnella lapsen tuotosta.

2.3 Automaattinen puheentunnistus

2.3.1 Automaattisen puheentunnistuksen perusteet

Automaattisella puheentunnistuksella (englanniksi *Automatic Speech Recognition*, ASR) tarkoitetaan hahmontunnistuksen sovellusta, jossa tietokoneen tavoitteena on tunnistaa ihmisen puhetta hyödyntäen tekoälyä ja koneoppimista (Kurimo, 2008, s. 73; Li ym., 2016). Siinä missä puhesignaalin tunnistaminen on ihmiselle luontaista, on se tietokoneelle haastavaa esimerkiksi kielen

monimuotoisuuden ja äänisignaalin laadun takia (Koskenniemi, 2013, s. 76). Signaalin laadukaskaan käsittely tietokoneella ei vastaa ihmiskorvaa, mutta automaattisen puheentunnistuksen laatu on kehittynyt niin, että sen soveltaminen erilaisten tehtävien ratkaisuun on mielekästä (Koskenniemi, 2013, 76; Li ym., 2016). Automaattisen puheentunnistuksen sovelluksia ovat esimerkiksi laitteiden puheohjaus sekä kieltenopiskeluun tarkoitetut sovellukset (Kurimo 2009, s. 336; Li ym. 2016).

Puheentunnistus alkaa äänen tallentamisesta ja digitaalisen äänidatan muokkaamisesta suodattamalla ja pienempiin osiin jakamalla (Koskenniemi, 2013, s. 97; Kurimo, 2009, s. 336–337). Seuraavaksi muodostetaan taajuusspektri ja sen lukuarvojen perusteella edelleen piirvektori (Kurimo, 2009, s. 337). Varsinainen automaattinen puheentunnistus tapahtuu tilastollisen mallin avulla piirvektoria hyödyntäen (Kurimo, 2008, 74; Li ym., 2016). Tyypillisesti puheentunnistusjärjestelmä sisältää useita samanaikaisesti hyödynnettäviä malleja (Kurimo, 2008). Kielimalli kuvaa sanojen ja sanayhdistelmien todennäköisyyksiä, ja sanastomalli tarjoaa tietoa sanoista sekä niissä olevista foneemeista. Foneemimallia käyttäen lasketaan todennäköisyys sille, että jokin tietty puheen segmentti edustaa tiettyä foneemia (s. 76–77).

Jotta puheentunnistuksessa voidaan käyttää tilastollista mallia, tarvitsee se opettaa kieli- ja äännejärjestelmäkohtaisesti (Mansikkaniemi ym., 2017). Tähän tarvitaan runsaasti opetusdataa eli puhenäytteitä. Puhesignaalin jatkuvuus, ylimääräisen aineksen erottelu puheesta sekä kielessä harvakseltaan esiintyvien taivutusmuotojen tunnistaminen haastavat automaattista puheentunnistusta (Kurimo, 2008, s. 74). Suomen kieli on agglutinoiva eli runsaasti taipuva, mikä osin lisää puheentunnistimeen kohdistuvia vaatimuksia (Hirsimäki ym., 2006, 516). Toistaiseksi puheentunnistimen tarkkuus on yhteydessä tunnistimelle asetettuihin vaatimuksiin, ja rajallisissa tehtävissä, esimerkiksi vaihtoehtojen akustisten erojen ollessa selkeitä, puheentunnistin toimii usein luotettavasti (Kurimo, 2008, s. 79).

2.3.2 Automaattinen puheentunnistus puheterapiassa

Yksi automaattisen puheentunnistuksen sovelluksista on ääntämisen laadun automaattinen arviointi, ja siihen perustuva palautteenanto käyttäjälle (Ylinen & Kurimo, 2017). Automaattiseen puheentunnistukseen kohdistuu runsaasti vaatimuksia – sen on oltava nopea, yhdenmukainen ja tarkka, sillä välitön palaute parantaa oppimistulosta. Laadun arvioinnissa tunnistetaan aluksi tallennetusta puhesignaalista sanat ja jaetaan ne äännteiksi. Tämän jälkeen äännteistä johdetaan numeerisia arvoja, joita verrataan malliääntämisen arvoihin. Tunnuslukuja verrataan asiantuntijoiden

tekemien luokitteluiden, eli opetusdatan ja koeälyn oppimisprosessin perusteella saatuihin tunnuslukuihin. (s. 64, 66). Automaattinen puheentunnistin on väsymätön, ja tarkoituksenmukaisesti toimiessaan sen palaute on tasalaatuista. Nämä ovat tärkeitä piirteitä myös silloin, kun puheentunnistusta käytetään /r/-äänteen harjoittelussa.

Kun automaattista puheentunnistusta ja sen tarjoamaa palautetta halutaan hyödyntää äänteen harjoitteluun, on puheentunnistimen tiedettävä esimerkiksi, millaisia variaatioita äänneestä on olemassa, ovatko toiset variaatiot toisia suotavampia, millainen äänneasu on tavoiteltava, ja miltä äänne kuulostaa muiden äänneiden ympäröimänä. Puheentunnistin yhdistää asiantuntijan puhenäytteistä antaman arvioinnin ja koneoppimista, jotta se voi antaa palautetta käyttäjän suorituksesta (Mulfari ym., 2022). Puheentunnistimen luotettava toiminta edellyttää asiantuntijoiden puhenäytteistä antaman arvioinnin riittävää yhdenmukaisuutta.

2.4 Reliabiliteetti puheterapiassa

Luotettavuutta ja toistettavuutta voidaan tarkastella käsitteen *reliabiliteetti* avulla (Tilastokeskus, 2023). Se kuvaa, kuinka luotettavasti ja toistettavasti haluttua ilmiötä kyetään mittaamaan valitulla mittarilla. Terveydenhuollossa samaa mittaria käyttävät useat eri henkilöt, joten eri mittaaajien välinen toistettavuus on tärkeää (Terveyden ja hyvinvoinnin laitos, 2023). Reliabiliteetin tarkastelu on puheterapiassa perusteltua yhdenmukaisten käytänteiden edistämiseksi, minkä lisäksi esimerkiksi automaattista puheentunnistusta hyödyntävien sovellusten kehittämisessä opetusdatan tulee olla laadultaan yhtenäistä.

Arvioitsijareliabiliteetilla tarkoitetaan eri arvioitsijoiden arvioiden välistä yhdenmukaisuutta (englanniksi *Inter-Rater Reliability*) (Anttila, 2014). Arvioitsijareliabiliteetin tilastolliseen tarkasteluun on kehitetty muutamia menetelmiä – esimerkiksi Cohenin Kappa on luokitteluasteikollisille muuttujille soveltuva yhteneväisyyden mitta, kun arvioitsijoita on kaksi (Cohen, 1960; McHugh, 2012). Fleissin Kappa on Cohenin Kapan muunnos, jossa arvioitsijoita on useampi kuin kaksi, eikä se sisällä oletusta aineiston normaalijakautuneisuudesta (Fleiss, 1971). Kappa-kertoimen (κ) arvo voi vaihdella välillä $-1-1$, jossa arvo 1 tarkoittaa arvioitsijoiden täydellistä yhtenevyyttä ja arvo 0 täydellistä sattumaa (McHugh, 2012). Tyypillisesti korkean yhdenmukaisuusarvon on mielletty viittaavan luotettavuuteen (Johnson & Christensen, 2008, s. 504). Landis ja Koch (1977) ovat ehdottaneet Kappa-kertoimen tulkintaa taulukon 1 mukaisesti.

Esimerkiksi McHugh (2012) on esittänyt kritiikkiä koskien perinteistä Kappa-kertoimen tulkintaa, sillä se sallii huomattavasti erimielisyyttä luokassa hyvä. Hän onkin ehdottanut, että esimerkiksi Kappa-arvo .60–.79 tulkittaisiin kohtalaiseksi, ja tuolloin prosentuaalisesti datasta luotettavaa olisi 35–63 %. Fleissin Kappa-kertoimen tulkintaan sovelletaan yleisesti Cohenin Kappa-kertoimen mukaista tulkintaa (Laerd Statistics, 2019).

Taulukko 1. Perinteinen Cohenin Kappa-kertoimen tulkinta (Landis & Koch, 1977).

Kappa-arvo κ	Yhtenevyyden taso
.01–.2	ei yhtenevyyttä tai vähäinen
.21–.4	heikko
.41–.6	kohtalainen
.61–.8	hyvä
.81–.1	erinomainen

Puheterapiassa tehdään runsaasti puheterapeutin henkilökohtaisiin havaintoihin perustuvia arviointeja (Jung ym., 2022). Esimerkiksi arvioidessaan lapsen ääntämisen oikeellisuutta puheterapeutit muodostavat arvioitaan perustuen kuulohavaintoon (Jing & Grigos, 2022). Erityisesti kliinisesti rajallisen kokemuksen omaavalle, hiljattain puheterapeutiksi valmistuneelle henkilölle tehtävä voi olla haastava (Jung, ym. 2022; Munson ym., 2012). On viitteitä siitä, että pidemmän kliinisen kokemuksen omaavat puheterapeutit ovat luokittelussaan tarkempia. Jing ja Grigos (2022) ovat soveltaneet painotettua Kappa-kerrointa tutkiessaan puheterapeuttien havaintojen yhtenevyyttä luokiteltaessa lasten puhunnosten oikeellisuutta. Tutkimukseen osallistui 30 puheterapeuttia, ja keskimääräisesti heidän luokittelunsa yhtenevyys kolmiportaisella asteikolla oli hyvä ($\kappa = .69$). Jung ja muut (2022) tarkastelivat juuri valmistuneiden puheterapeuttien luokittelun yhtenevyyttä kokeneempien puheterapeuttien luokittelun kanssa samaisella kolmiportaisella luokitteluasteikolla. Keskimääräisesti luokittelun yhtenevyys oli hyvällä tasolla ($\kappa = .62$).

3 TUTKIMUKSEN TAVOITE JA TUTKIMUSKYSYMYS

Puheterapiassa käytetään standardoituja testejä, mutta merkittävä osa arvioinneista perustuu puheterapeuttien tekemiin havaintoihin ja tulkintoihin niistä (esimerkiksi Jing ym., 2022; Munson ym. 2012). On ihmiselle ominaista, että eri henkilöt kiinnittävät huomiota erilaisiin ja osin toisistaan poikkeaviin puheen piirteisiin (Ylinen & Kurimo, 2017, s. 65). Havaintoihin perustuvia päätelmiä, esimerkiksi luokittelua, tarkastelemalla voidaan arvioida puheterapeuttien tekemien arviointien yhdenmukaisuutta ja edelleen edistää tasalaatuista toimintaa arviointikäytänteitä kehittämällä. Tämän tutkimuksen tarkoituksena on tarkastella kolmen 3.vuoden logopedian opiskelijan luokittelun yhtenevyyttä 5–10-vuotiaiden suomalaislasten /r/-äänneaineistossa. Tutkimus pyrkii edistämään kehitteillä olevan äänneharjoittelupelin osana toimivan automaattisen puheentunnistuksen ja näin ollen pelin antaman palautteen laatua ja yhdenmukaisuutta.

Tutkimuskysymys:

Miten yhteneväisesti logopedian 3. vuoden opiskelijat luokittelevat suomalaislasten /r/-äänteitä?

4 TUTKIMUSMENETELMÄT

Tutkimus on kvantitatiivinen tutkimus, jonka aineiston analyysissä on hyödynnetty aineiston kvantifiointia sekä tilastollisia menetelmiä. Tutkimusaineisto on kerätty osana TEFLON (Technology-enhanced foreign and second-language learning of Nordic languages) -projektia vuonna 2023.

4.1 Tutkimushenkilöt

Tämän tutkimuksen tutkimushenkilöinä toimi kolme Tampereen yliopiston logopedian oppiaineen 3. vuoden opiskelijaa. Tutkimushetkellä opiskelijat olivat kandidaatin tutkinnon suorittamisen loppuvaiheessa ja osallistuivat osana opintojaan TEFLON-projektiin.

4.2 Aineiston muodostuminen

Tutkimuksen aineistona toimi kolmen logopedian opiskelijan tekemä luokitteluaineisto suomalaislasten /r/-äänteiden laadusta. /r/-äänneaineisto on koottu osana TEFLON-projektia. Kokonaisuudessaan /r/-äänneaineisto sisälsi 1 329 puhenäytettä yhdeltätoista 5–10-vuotiaalta suomalaislapselta, jotka eivät tutkimushetkellä saaneet puheterapiaa tai artikulaatio-ohjausta. Aineisto tallennettiin Audacity-ohjelman versiolla 3.2.1. käyttäen pääpantamikrofonia ja tietokoneen äänikorttia. Puhenäytteiden keruussa lapsia ohjeistettiin puhumaan heille tavanomaiseen tapaan. Yhden puhenäytteen pituus oli yksi sana, ja yhden lapsen puhenäytteitä aineistossa oli keskimäärin 121 kappaletta. Opiskelijat luokittelivat kukin samat 286 puhenäytettä, mikä on 21,5 % luokiteltavasta aineistosta.

Opiskelijat luokittelivat lasten tuottamat /r/-äänteet kuuteen kategoriaan äänten laadun mukaisesti (taulukko 2). Luokiteltua aineistoa oli tarkoitus hyödyntää /r/-äänten harjoitteluun sovelletun mobiilipelin kehittämisessä, joten luokitteluasteikko kehitettiin pelin tarpeisiin. Aluksi opiskelijat luokittelivat kukin itsenäisesti 55 puhenäytettä, minkä jälkeen he keskustelivat luokittelun toimivuudesta yhdessä pelikehittäjien kanssa, ja luokitteluun tehtiin pieniä muutoksia. Lopullinen

luokitteluasteikko on kuvattu taulukossa 2. Tämän jälkeen opiskelijat luokittelivat loput tutkimukseen valitut puhenäytteet itsenäisesti. Kukin opiskelija tallensi luokittelunsa Microsoft Excel -työkirjaan.

Taulukko 2. /r/-äänteen luokitteluasteikko.

Arvo	Kuvaus
0	hylätty
1	/r/-äänne omissoituu tai korvautuu jollakin muulla äänteellä kuin /d/-äänteellä, esimerkiksi /l/
2	virheellinen ääntöpaikka, esimerkiksi uvulaarinen täryllinen /r/
3	/r/-äänne korvataan /d/-äänteellä
4	/r/-äänteen paikka oikea, täryn kestossa tai voimakkuudessa poikkeavuutta
5	suomen kielen mukainen /r/-äänne

4.3 Aineiston analyysi

Aineisto analysoitiin hyödyntäen Microsoft Excel -taulukkolaskentaohjelmaa sekä IBM SPSS Statistics -tilastointiohjelman versiota 28. Aluksi kolmen opiskelijan luokittelut yhdistettiin yhteen työkirjaan Microsoft Excel -taulukkolaskentaohjelmassa, laskettiin luokittelun frekvenssit ja piirrettiin kuvaaja.

Arvioitsijareliabiliteetin arvioimiseksi tarkasteltiin aineiston jakautuneisuutta IBM SPSS Statistics-tilastointiohjelmalla. Aineisto ei ollut normaalijakautunut, joten normaalijakautuneisuusoletusta edellyttäviä tilastollisia testejä ei voitu soveltaa. Tämän jälkeen aineiston reliabiliteettia tarkasteltiin Kappa-kertoimen avulla. Tilastollisessa testauksessa hyödynnettiin Fleissin Kappa-kerrointa, joka soveltuu usean arvioitsijan luokittelun yhtenevyyden tarkasteluun.

4.4 Tutkimuksen eettisyys

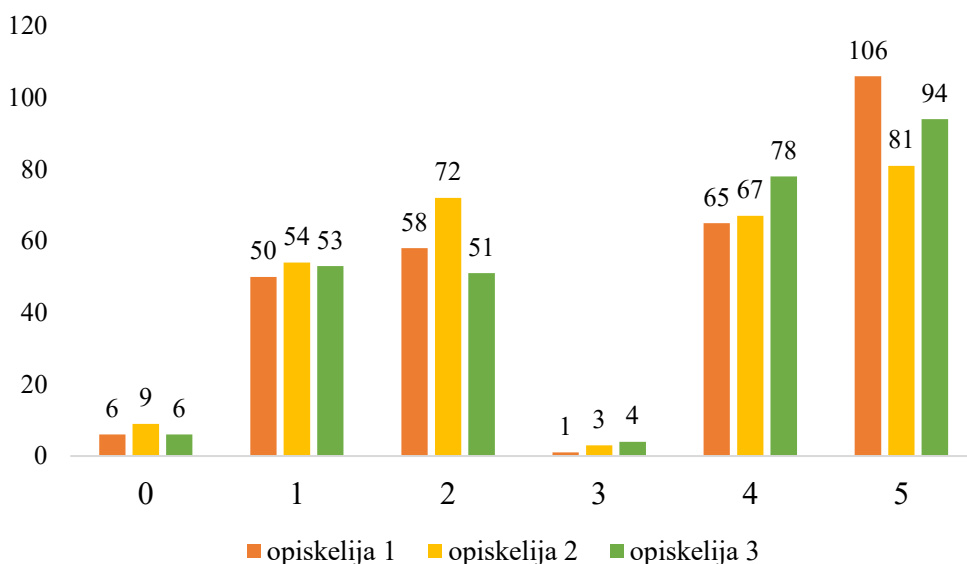
TEFLON-projektissa suoritettuun aineistonkeruuseen ei tarvittu erillistä eettistä ennakoarviointia. Aineistonkeruuta varten TEFLON-projektilla oli Tampereen kaupungilta tutkimuslupa, jossa huomioitiin aineistonkeruun eettiset näkökohdat sekä tietosuojat. Osallistujat olivat aineistonkeruuseen osallistuneiden henkilöiden henkilökohtaisten kontaktien kautta hankittuja. Lapsen huoltajat ilmoittivat lapsen tutkimukseen ja allekirjoittivat tutkimusluvan. Luokittelijoina

toimineet opiskelijat eivät osallistuneet tässä tutkimuksessa käytetyn aineiston keruuseen, vaan sen suorittivat TEFLON-projektin tutkimusavustaja ja TEFLON-projektiin osanaan opintoja osallistunut logopedian opiskelija. Puhenäytteet oli nimetty anonymisti, mutta ääni itsessään on tunnistetieto, jota ei voi anonymisoida. Puhenäytteet lähetettiin opiskelijoille hyödyntäen tietoturvallista Funet FileSender -tiedostonjakopalvelua. Logopedian opiskelijoiden osallistuminen tutkimukseen oli vapaaehtoista.

5 TULOKSET

Kaaviossa 1 on esitetty frekvenssit logopedian 3. vuoden opiskelijoiden /r/-äänteen luokittelusta. Yksittäiseen luokkaan tehtyjen havaintojen yhteenlaskettu määrä vaihtelee huomattavasti, sillä luokkaan 3 on luokiteltu 8 näytettä ja luokkaan 5 yhteensä 281 näytettä. Voidaan havaita, että valtaosa opiskelijoiden luokitteluista on sijoittunut kategorioihin 4 ja 5, vastaavasti kategoriat 0 ja 3 ovat hyvin pieniä. Luokittelun yhtenevyys kaikki luokat huomioiden Fleissin Kappaa käytettäessä oli $\kappa = .68$ (95 % LV .64–.71), $p < .001$. Taulukossa 3 on esitetty eriteltynä jokaisen luokan yhtenevyysarvot. Taulukkoon 4 on koottu jokaisen luokan Kappa-kerroin, kyseiseen luokkaan arvioitsijoiden tekemien havaintojen määrän vaihteluväli sekä perinteinen Cohenin mukainen tulkinta yhtenevyydestä.

Kaavio 1. /r/-äänteen luokittelun frekvenssit.



Taulukko 3. Luokittelun yhtenevyys Fleissin Kappa-kertoimella.

	κ	Keskivirhe	z	p-arvo	95 % luottamusväli	
					alaraja	yläraja
kaikki luokat	.68	.019	35.69	< .001	.64	.71
0	.85	.034	25.00	< .001	.79	.92
1	.88	.034	25.64	< .001	.81	.94
2	.65	.034	19.04	< .001	.58	.72
3	.37	.034	10.81	< .001	.30	.44
4	.47	.034	13.78	< .001	.40	.54
5	.73	.034	21.39	< .001	.66	.80

Taulukko 4. Yhtenevyyden taso.

Luokka	Kappa-kerroin κ	n*	Yhtenevyyden taso**
kaikki	.68	[286]	hyvä
0	.85	[6...9]	erinomainen
1	.88	[50...54]	erinomainen
2	.65	[51...72]	hyvä
3	.37	[1...4]	heikko
4	.47	[65...78]	kohtalainen
5	.73	[81...106]	hyvä

*n = kyseisen luokan arvojen vaihteluväli eri arvioitsijoilla

**perinteinen Cohenin Kappa -tulkinta

6 POHDINTA

Tämän tutkimuksen tarkoituksena oli tarkastella sitä, kuinka yhtenevästi logopedian 3.vuoden opiskelijat luokittelevat tavanomaisesti kehittyneiden, 5–10-vuotiaiden suomalaislasten tuottamia /r/-äänteitä. Tutkimukseen osallistui kolme logopedian opiskelijaa Tampereen yliopistosta. Aineistona toimi yhteensä 286 puhenäytettä, joiden kesto oli yksi sana ja jotka jokainen opiskelija luokitteli. Luokitteluasteikko oli luotu /r/-äänteen harjoitteluun tarkoitetun mobiilipelin tarpeisiin. Opiskelijoiden luokittama aineisto toimi mobiilipelin automaattisen puheentunnistimen opetusmateriaalina. Luokittelun yhtenevyyttä arvioitiin käyttämällä usean arvioitsijan tapaukseen soveltuvaa Fleissin Kappa-menetelmää.

6.1 Tulosten tarkastelu

Tässä tutkimuksessa Kappa-kertoimen arvot luokittelun yhtenevyydestä olivat tilastollisesti merkitsevät tarkasteltaessa sekä koko aineistoa että yksittäisiä luokkia. Koko aineistossa yhtenevyyden voidaan tulkita olevan hyvällä tasolla. Tässä tutkimuksessa havaittu yhtenevyyden taso on hyvin samansuuntainen kuin aiemmissa puheterapeuttien tai puheterapeuttiopiskelijoiden arviointien yhtenevyyttä tarkastelevissa tutkimuksissa (Jing & Grigos, 2022; Jung ym., 2022; Miles, 2017). Näiden tulosten perusteella voitaisiin todeta, että tyypillisesti ja erilaisissa konteksteissa tehtyjen puheterapeuttien arvioiden yhtenevyys on Kappa-kertoimen avulla tarkasteltaessa hyvällä tasolla. Toisaalta, jos sovelletaan McHughin (2012) esittämää tulkintaa Kappa-kertoimen arvoista, on yhtenevyys tällöin enää kohtalainen, ja prosentuaalisesti tarkasteltuna luotettavuus vaihtelisi 35–63 prosentin välillä.

Tarkasteltaessa yksittäisiä luokkia havaitaan, että yhtenevyydeltään heikoimman Kappa-kertoimen arvon saanut luokka 3 (/r/-äänteen korvautuminen /d/-äänteellä). Se sisältää vain kahdeksan luokiteltua havaintoa, ja niiden määrät arvioitsijoittain vaihtelevat välillä 1–4. Näytteiden pieni määrä ja siihen suhteutettu suuri vaihtelevuus arvioitsijoittain heikentää tämän luokan yhtenevyyden tasoa merkittävästi. On lisäksi huomioitava, että suurta aineistoa luokiteltaessa inhimilliset virheet ovat mahdollisia, ja nämä virheet korostuvat erityisesti vähän havaintoyksikköjä sisältävissä luokissa. Luokittelusta tehtävien havaintojen perusteella voitaneen kuitenkin odotetusti todeta, että suomalaislapset korvaavat spontaanipuheessaan /r/-äännettä hyvin harvoin /d/-äänteellä. /d/-äänteen

lisääminen omaksi luokakseen perustui sen käyttöön yhtenä välivaiheena /r/-äännettä harjoiteltaessa sen alveolaarisen ääntöpaikan vuoksi.

Korkeimman yhtenevyyden tason logopedian opiskelijat saavuttivat luokissa 0 (hylätty näyte) sekä 1 (/r/-äänne omissoituu tai korvautuu jollakin muulla äänneellä kuin /d/-äänneellä, esimerkiksi /l/-äänne). Tulokset ovat yhtenevyyden osalta samansuuntaisia kuin Jingillä ja Grigosilla (2022), jotka havaitsivat puheterapeuttien olevan luokittelussaan vahvasti samanmielisiä silloin, kun luokiteltavat tuotokset olivat huomattavasti virheellisiä.

Luokan 2 (virheellinen ääntöpaikka) yhtenevyyden voidaan tulkita olevan hyvällä tasolla. Luokittelujen frekvenssejä tarkasteltaessa voidaan havaita, että opiskelija 2 on valinnut muita luokittelijoita useammin luokan 2. Luokkiin 4 ja 5 hän on vastaavasti sijoittanut näytteitä toisia vähemmän. On mahdollista, että kyseinen opiskelija on tulkinnut osan puhenäytteistä virheellisessä paikassa tuotetuksi, kun taas toiset opiskelijat ovat tulkinneet samat näytteet joko oikein tai lähes oikein tuotetuksi ja sijoittaneet ne luokkiin 4 tai 5. Jos näin on, on havainto samansuuntainen kuin Jingin ja Grigosin (2022) havainto siitä, että hienovaraisempien virheiden yhdenmukainen luokittelu on puheterapeuteille haastavampaa kuin selvästi virheellisten tai oikeiden tuotosten luokittelu.

Luokat 4 (/r/-äänneen paikka oikea, täryn kestossa tai voimakkuudessa poikkeavuutta) ja 5 (suomen kielen mukainen /r/-äänne) ovat luokista lähimpinä toisiaan määrittelyn suhteen, ja niihin on myös luokiteltu suurin osa havainnoista, peräti 57 %. Luokassa 4 yhtenevyys oli kohtalaisella ja luokassa 5 hyvällä tasolla. Sekä Jung & Grigos (2022) että Jing ja muut (2022) havaitsivat omissa tutkimuksissaan, että täysin oikeiden suoritusten luokittelu oli yhteneväisintä. Tämän tutkimuksen tulokset poikkesivat aiemmista tutkimustuloksista, sillä suurin yhtenevyys oli merkittävästi virheellisissä suorituksissa. Tulosta voi selittää osaltaan se, että luokkien 4 ja 5 määrittelyt olivat lähellä toisiaan.

6.2 Menetelmän pohdinta

Tässä tutkimuksessa tarkasteltiin logopedian opiskelijoiden luokittelun yhtenevyyttä luokitteluasteikolla, jossa luokkia oli yhteensä kuusi. Luokkien syntyyn ja määrään vaikuttivat ensisijaisesti tarpeet, joita luokittelua hyödyntävä mobiilipeli asetti. On tyypillistä, että erilaisissa arviointikonteksteissa käytetään viisiportaista asteikkoa. Sen eduksi voidaan katsoa esimerkiksi kolmiportaaiseen asteikkoon verrattuna parempi erottelukyky, minkä ajatellaan edistävän oikeudenmukaisuutta (Opetushallitus, 2018). Kehitteillä olleessa mobiilipelissä haluttiin antaa lapsen tuotoksesta tähtiä välillä 0–5 sen mukaan, kuinka oikeanlainen lapsen tuottama /r/-äänne sanassa on. Näin ollen myös automaattisen puheentunnistuksen opetusaineistossa hyödynnettiin viisiportaista, tarkemmin määriteltynä kuusiportaista asteikkoa, kun osa luokiteltavasta aineistosta hylättiin arvolla 0 esimerkiksi tallenteen ollessa tyhjä.

Aiemmissa tutkimuksissa puheterapeuttien puhunnoksen oikeellisuuden arviointiin käyttämät luokitteluasteikot ovat pääasiassa olleet kolmiportaisia niin, että luokkina ovat olleet huomattavan virheellinen, jollakin tapaa virheellinen sekä oikea suoritus (esimerkiksi Jing ym., 2022; Jung & Grigos, 2022). Miles (2017) käytti kaksiportaista asteikkoa nielemisen arviointiin keskittyvässä tutkimuksessaan. Kuusi luokkaa sisältävä asteikko vaikuttaa siis melko hienovaraiselta työkalulta puheterapeuttiseen arvioon.

Luokkien määrän vaikutusta Kappa-kertoimen laskemiseen ja edelleen arvoon on tarkasteltu kriittisesti. Esimerkiksi Hallgren (2012) esittää, että sekä luokkien määrällä että luokiteltavien kohteiden määrä voi vaikuttaa Kappa-kertoimen suuruuteen. Vaikuttaisi siltä, että Kappa-kerroin on tyypillisesti korkeampi silloin, kun luokkia on vähemmän. Fleissin Kappa-kertoimen arvolle ei ole esitetty tulkintaan liittyviä arvoja, joten työssä on sovellettu Landisin ja Kochin (1977) esittelemää tulkintaa Cohenin Kappa-kertoimelle. Kappa-kertoimen yhteydessä on syytä huomioida, että merkitsevä p-arvo itsessään ei kerro luokittelun yhtenevyydestä, vaan osoittaa vain Kappa-kertoimen poikkeavan tilastollisesti merkitsevästi arvosta 0, joka merkitsee täydellistä sattumaa (Laerd Statistics, 2019).

On huomioitava, että pelin tarpeisiin rakennettu luokittelu johti siihen, että luokkien havaintoyksikköjen määrä vaihteli huomattavasti, ja havainnot painoutuivat luokkiin 4 ja 5. Tämä ei sinänsä vaikuta Kappa-kertoimen käyttöön, sillä normaalijakaumaoletusta ei ole. Luokkien määrään kohdistuvan kritiikin sekä aineiston epätasaisen jakautumisen vuoksi analyysivaiheessa ajettiin läpi

myös vaihtoehtoisia luokitteluja joko poistamalla pienimmät luokat 0 ja 3 aineistosta tai yhdistelemällä luokkia. Tulokset eivät kuitenkaan muuttuneet alkuperäisen luokittelun antamista tuloksista merkittävästi, ja toisaalta luokkien yhdisteleminen olisi ristiriidassa mobiilipelin ja siinä käytettävän automaattisen puheentunnistuksen opetusdatan vaatimusten kanssa. Näin ollen tulokset päädyttiin raportoimaan alkuperäistä luokitteluasteikkoa vastaavalla tavalla.

Tässä tutkimuksessa luokittelijoina toimi kolme 3. vuoden logopedian opiskelijaa. Luokittelijoiden määrä on ollut pieni. Tutkimukseen osallistuneet opiskelijat olivat kaikki samalta vuosikursilta, joten tästäkään syystä tulokset eivät ole yleistettävissä koskemaan kaikkia logopedian opiskelijoita tai jo työssä toimivia puheterapeutteja. Opiskelijoiden luokittelemien näytteiden osuus, 21,5 % koko aineistosta on kohtuullinen, mutta isompi otoskoko olisi voinut lisätä tulosten luotettavuutta. Lisäksi on huomioitava, että aineisto kokonaisuudessaan on pieni automaattisen puheentunnistuksen tarpeisiin. Luokkien määrään ja Kappa-kertoimen arvon tulkintaan liittyvästä epävarmuudesta huolimatta voidaan todeta, että tässä tutkimuksessa saadut tulokset puheterapeuttiopiskelijoiden suorittaman luokittelun yhtenevyydestä ovat samansuuntaiset kuin aiemmissa tutkimuksissa.

6.3 Työn kliininen merkitys ja jatkotutkimusaiheita

Puheterapian kontekstissa reliabiliteettia on tarkasteltu tyypillisesti testaus- ja arviointitilanteisiin liittyen. Tutkimusten tavoitteena on ollut selvittää, kuinka yhdenmukaisesti puheterapeutit toimivat määrittäessään jonkin toiminnon, esimerkiksi puheen tai nielemisen, tavanmukaisuutta tai poikkeavuutta ja näiden perusteella arvioida kuntoutustarvetta tai kuntoutuksen vaikuttavuutta. Tässä työssä selvitettiin puheterapeuttiopiskelijoiden tekemän luokittelun reliabiliteettia koskien suomalaislasten /r/-äännettä. Usein puheterapeuttinen arvio lapsen artikulaatiovarianttia ja kuntoutustarvetta määritettäessä sisältää kuvauksen lapsen äänteentuottavasta ja -paikasta. Tämän tutkimuksen luokittelun tavoitteena ei kuitenkaan ollut vain ääntämisen oikeellisuuden tarkasteleminen, vaan luokittelu automaattisen puheentunnistuksen tarpeisiin. Tällainen työ puheterapiakontekstissa kansallisesti on uutta. Tämän tutkimuksen perusteella voidaan todeta, että logopedian opiskelijoiden suomalaislasten /r/-äänteiden luokittelun yhtenevyyden voidaan tulkita olevan hyvällä ($\kappa=,68$) tasolla. Tämän tutkimuksen perusteella ei kuitenkaan voida ottaa kantaa siihen, kuinka toimiva luokittelu on mobiilipelin käyttötarkoituksessa, ja riittääkö esimerkiksi kyseinen aineisto automaattiselle puheentunnistukselle opetusmateriaaliksi.

Kansallista tutkimustietoa puheterapeuttien ja puheterapeuttiopiskelijoiden arviointien yhtenevyydestä on niukasti. Arvioitsijareliabiliteettia voitaisiin jatkossa tarkastella erityisesti sellaisissa kliinisesti yleisissä puheterapeuttisissa tilanteissa, joihin ei ole olemassa standardoituja testejä. Se voi myös auttaa havaitsemaan ja kehittämään ammatillisen uransa eri vaiheissa olevien puheterapeuttien työn yhdenmukaistamista. Arvioitsijareliabiliteetin tarkastelu edesauttaa osaltaan toteuttamaan lakivelvoitetta sosiaali- ja terveystalvelujen yhdenvertaisesta toteuttamisesta (Laki sosiaali- ja terveydenhuollon järjestämisestä 1:4 §).

Kuntoutusteknologioiden yleistyminen ja kuntoutujien oman aktiivisen roolin korostuminen kuntoutusprosessissa koskee myös puheterapia-alaa (esimerkiksi Saeedi ym., 2022; Salminen ym., 2016a). Automaattinen puheentunnistus ja erityisesti sen tarjoama palaute voivat kannustaa asiakasta itsenäiseen harjoitteluun, ja edesauttaa kuntoutuksen tavoitteiden toteutumista. Toimiakseen tarkoituksenmukaisesti automaattinen puheentunnistin tarvitsee kieli- ja tavoitekohtaisesti räätälöityä opetusdataa, ja tämän tiedon muodostamiseen tarvitaan puheterapeuttista osaamista. Jatkossa puhunnosten luokittelu erilaisin kriteerein puheentunnistuksen käyttöön yhteistyössä koneoppimisen kanssa voi kuulua myös puheterapeutin tehtäviin.

LÄHDELUETTELO

- Aalto, E. (2023). *Phonological skills in children acquiring Finnish at 3;6. The associations with lexical, auditory word recognition, and pre-reading skills* (Acta Universitatis Ouluensis, B, Humaniora, 207) [väitöskirja, Oulun yliopisto]. JULTIKA Oulun yliopiston julkaisuarkisto. <http://jultika.oulu.fi/files/isbn9789526235868.pdf>
- Ahmed, B., Monroe, P., Hair, A., Tien Tan, C., Gutierrez-Osuna, R. & Ballard, K. J. (2018). Speech-driven mobile games for speech therapy: User experiences and feasibility. *International Journal of Speech-Language Pathology*, 20 (6), 644–658. <https://doi.org/10.1080/17549507.2018.1513562>
- Anttila, P. (2014). *Tutkimisen taito ja tiedon hankinta*. Haettu 24.4.2023. <https://metodix.fi/2014/05/17/anttila-pirkko-tutkimisen-taito-ja-tiedon-hankinta/>
- Brenthal, J.E., Bankson, N.W. & Flipsen Jr., P. (2009). Treatment approaches. Teoksessa J.E. Bernthal, N.W. Bankson & P. Flipsen Jr. (toim.) *Articulation and phonological disorders – speech sound disorders in children* (6. painos, s. 278–330). Pearson.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20 (1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Fey, M. E. (1992). Articulation and Phonology. Inextricable Constructs in Speech Pathology. *Language, Speech, and Hearing Services in Schools*, 23 (3), 225–232. <https://doi.org/10.1044/0161-1461.2303.225>
- Findley, B. R. & Gasparyan, D. (2022). *Perspectives of the ASHA special interest groups* 7(3), 926–937. https://doi.org/10.1044/2022_PERSP-21-00276.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76 (5), 378–382.
- Furlong, L., Erickson, S. & Morris, M. E. (2017). Computer-based speech therapy for childhood speech sound disorders. *Journal of Communication Disorders* 68, 50–69. <https://doi.org/10.1016/j.jcomdis.2017.06.007>
- Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology*, 8 (1), 23–34. <https://doi.org/10.20982/tqmp.08.1.p023>
- Hirsimäki, T., Creutz, M., Siivola, V., Kurimo, M., Virpioja, S. & Pylkkönen, J. (2006). Unlimited vocabulary speech recognition with morph language models applied to Finnish. *Computer Speech and Language*, 20 (4), 515–541. <https://doi.org/10.1016/j.csl.2005.07.002>.
- Jesus, L. M. T, Martinez, J., Santos, J., Hall, A. & Joffe, V. Comparing Traditional and Tablet-Based Intervention for Children With Speech Sound Disorders: A Randomized Controlled Trial. *Journal of Speech, Language, and Hearing Research*, 62 (11), 4045–4061. https://doi.org/10.1044/2019_JSLHR-S-18-0301

- Jing, L. & Grigos, M. I. (2022). Speech-Language Pathologists' Ratings of Speech Accuracy in Children With Speech Sound Disorders. *American Journal of Speech-Language Pathology* 31 (1), 419–430. https://doi.org/10.1044/2021_AJSLP-20-00381
- Johnson, B. & Christensen, L. B. (2008). *Educational research: quantitative, qualitative, and mixed approaches*. SAGE.
- Jung, S., Jing, L. & Grigos, M. (2022). Graduate Student Clinicians' Perceptions of Child Speech Sound Errors. *Perspectives of the ASHA Special Interest Groups*, 7 (4), 1275–1283. https://doi.org/10.1044/2022_PERSP-21-00332
- Korpilahti, P. (2011). Kielihäiriöiden luokittelu. Teoksessa K. Launonen & A-M. Korpijaakko-Huuhka (toim.) *Kommunikoinnin häiriöt* (s. 39–58). Gaudeamus.
- Korpinen, L. & Nasredtin, F. (2011). Artikulaatio ja äännevirheet. Teoksessa K. Launonen & A-M. Korpijaakko-Huuhka (toim.) *Kommunikoinnin häiriöt* (s. 59–76). Gaudeamus.
- Koskenniemi, K. (2013). *Johdatus kieliteknologiaan, sen merkitykseen ja sovelluksiin*. Helsingin yliopiston nykykielten laitos.
- Krueger, B. I. (2019). Eligibility and Speech Sound Disorders: Assessment of Social Impact. *Perspectives of the ASHA Special Interest Groups* 4 (1), 85–90. https://doi.org/10.1044/2018_PERS-SIG1-2018-0016
- Kunnari, S., Savinainen-Makkonen, T. & Saaristo-Helin, K. (2012). *Fonologiatesti. Lasten äänteellisen kehityksen arviointimenetelmä*. Niilo Mäki Instituutti.
- Kurimo, M. (2008). Puheentunnistus. *Puhe ja kieli*, 28 (2), 73–83.
- Kurimo, M. (2009). Puheentunnistus. Teoksessa O. Aaltonen, R. Aulanko, A. Iivonen, A. Klippi & M. Vainio (toim.). *Puhuva ihminen: puhetieteiden perusteet* (s. 336–343). Otava.
- Laerd Statistics. (2019). Fleiss' kappa using SPSS Statistics. Statistical tutorials and software guides. Haettu 6.6.2023. <https://statistics.laerd.com/spss-tutorials/fleiss-kappa-in-spss-statistics.php>
- Laki sosiaali- ja terveydenhuollon järjestämisestä 29.6.2021/612.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33 (1), 159–174.
- Larmola, M. (1982). *Kouluikäisten kieli*. Suomen kirjallisuuden seura.
- Lass, N. J. & Pannbacker, M. (2008). The Application of Evidence-Based Practice to Nonspeech Oral Motor Treatments. *Language, Speech and Hearing Services in Schools*, 39 (3), 408–421. [https://doi.org/10.1044/0161-1461\(2008/038\)](https://doi.org/10.1044/0161-1461(2008/038))
- Li, J., Deng, L., Häb-Umbach & Gong. (2016). *Robust Automatic Speech Recognition*. Academic Press.

- Lof, G. L. & Watson, M. M. (2008). A Nationwide Survey of Nonspeech Oral Motor Exercise Use: Implications of Evidence-Based Practice. *Language, Speech & Hearing Services in Schools*, 39 (3), 392–407. [https://doi.org/ 10.1044/0161-1461\(2008/037\)](https://doi.org/10.1044/0161-1461(2008/037))
- Macrae, T. (2013). Lexical and child-related factors in word variability and accuracy in infants. *Clinical Linguistics & Phonetics*, 27 (6–7), 497–507. <https://doi.org/10.3109/02699206.2012.752867>
- Malaquias, R. F. & Malaquias, F. F. (2021). A literature review on the benefits of serious games to the literacy process of children with disabilities and learning difficulties. *Technology and Disability*, 33 (4), 273–282. [https://doi.org/ 10.3233/TAD-210339](https://doi.org/10.3233/TAD-210339)
- Mansikkaniemi, A., Smit, P., Kurimo, M. (2017). Automatic Construction of the Finnish Parliament Speech Corpus. *Interspeech*, 3762–3766. <https://doi.org.10.21437/Interspeech.2017-1115>.
- Martikainen, A.-L., Savinainen-Makkonen, T. & Kunnari, S. (2019). Intra-word consistency and accuracy in Finnish children aged 3–6 years. *Clinical Linguistics & Phonetics*, 33 (9), 815–830. <https://doi.org/10.1080/02699206.2019.1576770>
- McAllister Byun, T. (2017). Efficacy of Visual–Acoustic Biofeedback Intervention for Residual Rhotic Errors: A Single-Subject Randomization Study. *Journal of Speech, Language, and Hearing Research* 69 (5), 1175–1193. [https://doi.org/ 10.1044/2016_JSLHR-S-16-0038](https://doi.org/10.1044/2016_JSLHR-S-16-0038)
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22 (3), 276–282. <https://doi.org/10.11613/BM.2012.031>
- Miles, A. Inter-rater reliability for speech–language therapists’ judgement of oesophageal abnormality during oesophageal visualization. *International Journal of Language & Communication Disorders*, 52 (4), 450–455. <https://doi.org/10.1111/1460-6984.12283>
- Mullen, R. & Schooling, T. (2010). The national outcomes measurement system for pediatric speech-language pathology. *Language, Speech, and Hearing Services in Schools*, 41 (1), 44–60. [https://doi.org/10.1044/0161-1461\(2009/08-0051\)](https://doi.org/10.1044/0161-1461(2009/08-0051))
- Munson, B., Johnson, J. M., & Edwards, J. (2012). The role of experience in the perception of phonetic detail in children’s speech: A comparison between speech-language pathologists and clinically untrained listeners. *American Journal of Speech-Language Pathology* 21(2), 124–139. [https://doi.org/10.1044/ 1058-0360\(2011/11-0009\)](https://doi.org/10.1044/1058-0360(2011/11-0009))
- Opetushallitus. (2018). *Osaamisen arviointi*. Haettu 27.4.2023. <https://eperusteet.opintopolku.fi/#/fi/opas/4343283/tekstikappale/4414100>.
- Pitkänen, V. (2020). Puheterapian koulutuspaikkojen lisäys kannattaa. *Kela. Tutkimusblogi*. Haettu 3.2.2023. <https://tutkimusblogi.kela.fi/arkisto/5534>

- Saeedi, S., Bouraghi, H., Seifpanahi, M-S. & Ghazisaeedi, M. (2022). Review Article. Application of Digital Games for Speech Therapy in Children: A Systematic Review of Features and Challenges. *Journal of Healthcare Engineering*, 2022. <https://doi.org/10.1155/2022/4814945>
- Salminen, A.-L., Hiekkala, S. & Stenberg, J.-H. (2016a). Johdanto. Teoksessa A.-L. Salminen, S. Hiekkala & J.-H. Stenberg (toim.) *Etäkuntoutus*, (s. 9–10). Kela.
- Salminen, A.-L., Heiskanen, T., Hiekkala, S., Naamanka, J., Stenberg, J.-H. & Vuononvirta, T. (2016a). Etäkuntoutuksen ja siihen läheisesti liittyvien termien määrittelyä. Teoksessa A.-L. Salminen, S. Hiekkala & J.-H. Stenberg (toim.) *Etäkuntoutus*, (s. 11–18). Kela.
- Savinainen-Makkonen, T. & Kunnari, S. (2009). Puheen, kielen ja kommunikoinnin kehitys lapsuudessa. Teoksessa O. Aaltonen, R. Aulanko, A. Iivonen, A. Klippi & M. Vainio (toim.) *Puhuva ihminen: puhetieteiden perusteet* (s. 114–121). Otava.
- Savinainen-Makkonen, T. & Kunnari, S. (2012). Äännevirheet ja niiden kuntoutus. Teoksessa S. Kunnari & T. Savinainen-Makkonen (toim.) *Pienten sanat: Lasten äänteellinen kehitys* (s. 173–187). PS-kustannus.
- Sosa, A. V. (2015). Intraword variability in typical speech development. *American Journal of Speech-Language Pathology*, 24 (1), 24–35. https://doi.org/10.1044/2014_AJSLP-13-0148
- Terveiden ja hyvinvoinnin laitos. (2023). *Mittariarvioinnin käsitteet*. Haettu 13.6. osoitteesta <https://thl.fi/fi/web/toimintakyky/etusivu/toimia-tietokanta/toimian-kasikirja/mittariarvioinnin-toteutus/mittariarvioinnin-kasitteet>
- Tieteen termipankki. (2021). *Foneemi*. Haettu 18.4.2023 osoitteesta <https://tieteentermipankki.fi/wiki/Kielitiede:foneemi>
- Tilastokeskus. (2023). *Reliabiliteetti*. Haettu 22.4.2023 osoitteesta <https://www.stat.fi/meta/kas/reliabiliteetti.html>
- Törölä, H. (2009). Ääntelyn kehitys ensimmäisen ikävuoden aikana. Teoksessa O. Aaltonen, R. Aulanko, A. Iivonen, A. Klippi & M. Vainio (toim.) *Puhuva ihminen: puhetieteiden perusteet* (s. 107–113). Otava.
- Van Riper, C. (1978). *Speech correction principles and methods*. 6. painos. Prentice-Hall.
- Wren, Y. & Roulstone, S. (2008). A comparison between computer and tabletop delivery of phonology therapy. *International Journal of Speech-Language Pathology*, 10 (5), 346–363. <https://doi.org/10.1080/17549500701873920>
- Yle. (2023). Matilda Knuutti, 7, harjoittelee r-äännettä kotona, koska puheterapeuteista on pula kaikkialla – ”On tullut olo, että on jätetty yksin”, sanoo äiti. *Yle Uutiset*. Haettu 3.2.2023. <https://yle.fi/a/74-20013462>

Ylinen, S. & Kurimo, M. (2017). Kielenoppiminen vauhtiin puheteknologian avulla. Teoksessa H. Savolainen, R. Vilkkonen & L. Vähäkylä (toim). *Oppimisen tulevaisuus* (s. 67–69). Gaudeamus.