

Metakognitiivinen kyky havaita virhe kuulopalautteessa

Tarkastelussa signaalidetektio teoriaan perustuva koeasetelma

Rita Heikkilä ja Liinu Paajanen

Pro gradu -tutkielma

Ohjaaja: Henry Railo

Turun yliopisto

Yhteiskuntatieteellinen tiedekunta

Psykologian ja logopedian laitos

Logopedia

20.12.2023

Turun yliopiston laatu järjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu

Turnitin OriginalityCheck -järjestelmällä.

TURUN YLIOPISTO

Psykologian ja logopedian laitos / Yhteiskuntatieteellinen tiedekunta

HEIKKILÄ, RITA & PAAJANEN, LIINU: Metakognitiivinen kyky havaita virhe kuulopalautteessa, Tarkastelussa signaalidetektiteoriaan perustuva koeasetelma

Pro gradu -tutkielma, 36 s.

Logopedia

Joulukuu 2023

Metakognitiolla, eli tiedolla ja tietoisuudella omista kognitiivisista ilmiöistä, on keskeinen rooli puheterapiassa, vaikka siihen keskittyvää tutkimustietoa on logopedian alalla rajallisesti. Tämän pro gradu -tutkielman tavoitteena oli selvittää, kuinka hyvin ihmiset pystyvät metakognitiivisesti arvioimaan virheitä kuulopalautteessa. Lisäksi tarkastellaan, kuinka luotettavasti käyttämämme menetelmä toimii metakognitiivisen tehokkuuden mittaamisessa. Aihetta tutkittiin koeasetelmalla, jossa koehenkilö lausui /u/-äännettä ja kuulopalautetta muokattiin reaaliaikaisesti nostamalla hetkellisesti äänenkorkeutta. Koehenkilön tuli vastata, tapahtuiko muokkaus eli perturbaatio ensimmäisessä vai toisessa aikaikkunassa. Saman aikaisesti koehenkilön tuli antaa vastausvarmuus omasta vastauksestaan. Koeasetelman avulla koehenkilöille laskettiin signaalidetektiteoriaan perustuen arvot sensitiivisyydelle d' , metakognitiiviselle sensitiivisyydelle $meta-d'$ sekä metakognitiiviselle tehokkuudelle $Mratio$.

Ihmisten kykyä metakognitiivisesti arvioida virhe kuulopalautteessa tutkittiin vertaamalla sensitiivisyyttä metakognitiiviseen sensitiivisyyteen. Samalla tarkasteltiin, onko yksilöiden välillä vaihtelua metakognitiivisessa suoriutumisessa. Koeasetelman luotettavuutta tarkasteltiin test-retest-reliabiliteetin kautta. Tulokset laskettiin sekä Guggenmosin (2021) että Leen ja työryhmän (2018) suosittamia aineiston rajauksia käyttäen kaikille kolmelle tunnusluvulle. Tutkimukseen osallistui 59 koehenkilöä, jotka olivat iältään 19–32-vuotiaita. Koehenkilöt suorittivat kokeen kahteen kertaan.

Tulosten perusteella koehenkilöiden metakognitiivinen tehokkuus oli keskimäärin alle optimaalisen tason ja havaittavissa oli vaihtelua metakognitiivisessa suoriutumisessa. Test-retest-reliabiliteetin osalta positiivinen korrelaatio oli löydettävissä osalle tunnusluvuista testikertojen välillä. Kuitenkin sekä Guggenmosin (2021) että Leen ja kumppaneiden (2018) rajauksilla reliabiliteetit jäivät matalaksi ja kauas yleisistä reliabiliteetin tavoitearvoista. Koeasetelmassa tulisi tulevaisuudessa kiinnittää huomiota erityisesti yksilöllisen vaikeustason riittävän tarkkaan määrittämiseen sekä toistojen määrän vaikutukseen.

Asiasanat: metakognitio, metakognitiivinen sensitiivisyys, metakognitiivinen tehokkuus, perturbaatio, äänenkorkeus, d' , $meta-d'$, $Mratio$, signaalidetektiteoria, test-retest-reliabiliteetti

Sisällysluettelo

1	Johdanto	1
1.1	Signaalidetektio teoria	2
1.2	Metakognitio	3
1.2.1	Vastausvarmuus	5
1.2.2	Metakognition mittaaminen ja tunnusluvut	5
1.3	Koeasetelman rakentaminen	6
1.3.1	Perturbaatioparadigma	6
1.3.2	Pakkovalinta	7
1.4	Test-retest-reliabiliteetti metakognition mittaamisessa	8
2	Tutkimuskysymykset	10
2.1	Hypoteesit	10
3	Menetelmät	12
3.1	Koehenkilöt	12
3.2	Laitteisto ja lomakkeet	13
3.3	Koeasetelma	13
3.4	Kokeen kulku	14
3.4.1	Kokeen osiot	16
3.5	Tilastolliset analyysit	17
3.5.1	Metakognitiivinen suoriutuminen	17
3.5.2	Test-retest-reliabiliteetin analysointi	18
3.6	Tutkimuksen eettisyys	19
4	Tulokset	20
4.1	Erot perturbaatioiden koossa	20
4.2	Metakognitiivinen arvio omasta suoriutumisesta	21
4.3	Test-retest-reliabiliteetti	22
4.3.1	Test-retest-reliabiliteetti Guggenmosin (2021) rajausten mukaan	22
4.3.2	Test-retest-reliabiliteetti Leen ja työryhmän (2018) rajausten mukaan	23
5	Pohdinta	25
5.1	Tulosten kuvaus ja arviointi	25
5.2	Menetelmän kehittämiskohteet	28
5.3	Tutkimuksen merkitys ja jatkotutkimusehdotukset	30
	Lähteet	32

1 Johdanto

Metakognitiolla tarkoitetaan niitä prosesseja, joita ihminen käyttää reflektoidessaan omia kognitiivisia toimintojaan (Koriat, 2007). Metakognition tutkimuksessa keskeisessä asemassa on, kuinka tätä tietoa hyödynnetään informaation prosessoinnin säätelyssä. Tämän pro gradu -tutkielman tavoitteena on selvittää, kuinka hyvin ihmiset pystyvät metakognitiivisesti arvioimaan omaa ääntämistään. Aihetta tarkastellaan muokkaamalla ääntämisestä saatua kuulopalautetta. Lisäksi tarkastelemme, kuinka käyttämämme signaalidetektio teoriaan perustuva koeasetelma toimii metakognition mittaamisessa. Kokeessa muokataan lyhyesti ja reaaliaikaisesti koehenkilön äänenkorkeutta /u/-äännön aikana ja pyydetään valitsemaan, tapahtuiko muokkaus ensimmäisessä vai toisessa aikaikkunassa. Tämän lisäksi koehenkilöä pyydetään arvioimaan, kuinka varma hän oli antamastaan vastauksesta.

Metakognition tutkimus logopedian alalla on toistaiseksi vähäistä. Kuitenkin tarvetta laajemmalle tutkimukselle on, sillä metakognitio on keskeisessä osassa useissa puheterapiassa käytetyissä interventioissa, kuten kehityksellisten kielellisten haasteiden ja fonologian kuntoutuksessa (Gaile & Adams, 2018). Law ja tutkimusryhmä (2008) havaitsivat tutkimuksessaan, että lapsen kehittyessä terapian fokus usein siirtyy varsinaisesta taidon omaksumisesta taidon käytön metakognitiiviseen puoleen eli strategioihin ajatella ja käyttää kieltä. Tätä tukevat myös Gaile ja Adamsin (2018) tekemät havainnot siitä, että terapiassa käytetyt tehtävät asettavat vaatimuksia sekä taidon oppimiselle että taidon käytön monitoroinnille.

Metakognitiolla on siis keskeinen rooli puheterapiassa ja näin ollen sen laajempi ymmärtäminen olisi myös kliiniseltä kannalta hyödyllistä. Esimerkiksi sitä, miten hyvin ihmiset pystyvät arvioimaan metakognitiivisesti tuottamansa äänteen oikeellisuutta, ei ole vielä laajasti tutkittu. Kliinisestä näkökulmasta käyttämämme koeasetelma voisi tarjota tietoa siitä, onko mahdollisilla oman puheen havainnoinnin vinoumilla vaikutusta esimerkiksi puhehäiriöiden hoidossa.

Tässä tutkimuksessa koehenkilöinä toimivat neurologisesti terveet nuoret aikuiset. Tulevaisuudessa samankaltaista koeasetelmaa voidaan mahdollisesti hyödyntää esimerkiksi Parkinsonin tautia sairastavien ihmisten oman äänen havaitsemisen tutkimuksessa. Kliiniset havainnot viittaavat Parkinsonin tautia sairastavien henkilöiden kyvyn tunnistaa oman

äänenvoimakkuuden heikentymisen olevan alentunut (Clark ym., 2014; Fox ym., 2002; Ho ym., 2000). Luotettavaa menetelmää tämän tutkimiseen ei kuitenkaan ole vielä pystytty kehittämään.

1.1 Signaalidetektio teoria

Havaitsemiseen liittyvissä tehtävissä tulee usein tunnistaa ennalta määritelty signaali eri ärsykevaihtoehtojen väliltä. Tällaisesta ärsykkeiden tunnistamiseen liittyvästä tehtävästä käytetään myös nimeä tyypin 1 tehtävä (Maniscalco & Lau, 2012). Suoriutumista havaitsemisen tehtävissä kuvataan usein sensitiivisyyden mittareilla (Macmillan & Creelman, 2005). Korkea sensitiivisyys tarkoittaa, että yksilö pystyy tunnistamaan hyvin eron eri ärsykkeiden välillä. Näin ollen esimerkiksi äänen havaitsemisen tehtävässä korkean sensitiivisyyden omaava kuulija kuulee ja tunnistaa eron äänessä, epäsensitiivinen ei.

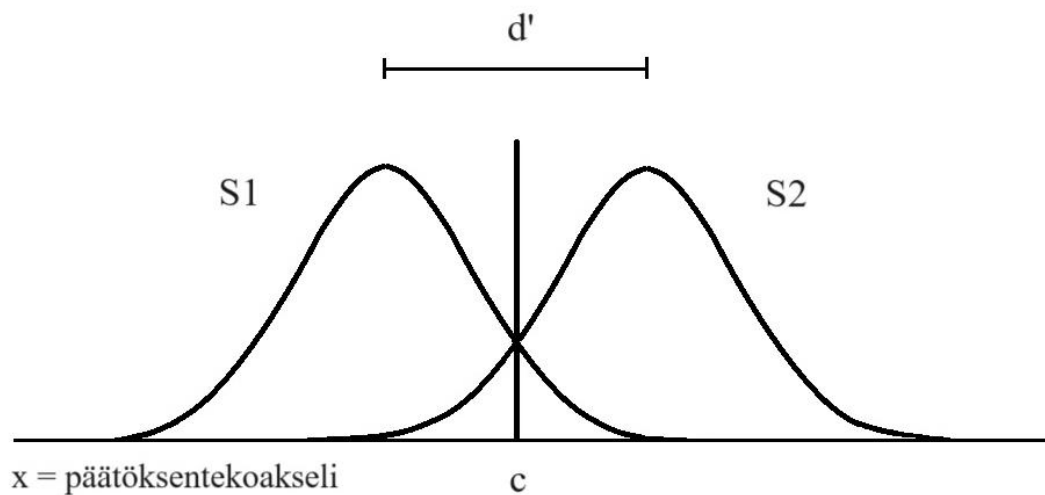
Signaalidetektio teoriassa on tietyt perusolettamukset (Salmela ym., 2006). Ensinnäkin havaintoprosessin kohteena olevan signaalin kulku on kohinaista, jolloin signaali ei välity virheettömästi. Signaalin laatua voi heikentää esimerkiksi ulkoinen kohina tai muu satunnainen taustahäly. Signaali on siis jo aistikanavaan välittyessään usein heikentynyt. Myöskään hermostossa signaali ei välity täydellisesti, sillä esimerkiksi hermoston tausta-aktivaation seurauksena neuraaliseen signaaliin välittyy lisää kohinaa.

Toinen keskeinen oletus käsittelee havaintoprosessin jakoa kahteen osaprosessiin, jotka ovat havainnointi ja päätöksenteko (Salmela ym., 2006). Havaintoprosessi on kohinan takia epävarmaa, mikä heijastuu päätöksentekoon ja eri vaihtoehtoihin liittyvien riskien arviointiin. Kolmantena olettamuksena on edellytys sille, että havainnoinnin ja edelleen päätöksenteon tulos voidaan ilmaista yhtenä lukuna.

Tyypillisessä signaalidetektio teoriaan perustuvassa tehtävässä havaittajan tulee erottaa toisistaan ärsykkeet, jotka kuuluvat joko luokkaan S1 tai S2 (Rounis ym., 2010). Nämä luokat kuvaavat tehtävästä riippuen esimerkiksi vastausvaihtoehtoja A ja B tai kyllä–ei. Päätöksenteko akselille muodostuu päätöksenteon kriteeri c , jonka ylittävät ärsykkeet kuuluvat luokkaan S2 ja alittavat luokkaan S1 (kuva 1). Kriteeri c kuvaa yksilön vastauskriteeriä eli toisin sanoen sitä, kuinka herkästi yksilö vastaa esimerkiksi ”kyllä” (Macmillan & Creelman, 2005). Näin ollen c :n sijainti päätöksenteko akselilla voi vaihdella henkilöiden välillä.

Jokainen ärsyke tuottaa arvon, joka päätöstä tehtäessä kallistuu enemmän joko luokkaan S1 tai S2 (Rounis ym., 2010). Nämä arvot ja niihin liittyvä todistusaineisto muodostavat päätöksenteko akselille normaalisti jakautuneet käyrät sekä luokalle S1 että S2. Näiden

normaalisti jakautuneiden käyrien normalisoitua etäisyyttä toisistaan kuvataan tunnusluvulla d' ja se kertoo, kuinka hyvin havaitsija pystyy erottamaan luokat S1 ja S2 toisistaan. Tunnusluku d' toimii siis sensitiivisyyden mittarina. Lisäksi d' on riippumaton vastausvinoumasta, jolloin tunnusluku pysyy samana riippumatta siitä, onko tutkittavalla taipumusta antaa korostuneesti esimerkiksi kyllä-vastauksia (Fleming & Lau, 2014).



Kuva 1. Signaalidetektio teoriaan perustuva malli päätöksenteosta kahden vaihtoehdon välillä. Tunnusluku d' kuvaa signaalia, joka on saatavilla kulloistakin tehtävää varten (Rounis ym., 2010).

Signaalidetektio teoriassa hyödynnetään siis tietoa oikeista ja vääristä vastauksista, jotta saadaan laskettua tunnusluku yksilön sensitiivisyydelle d' ja päätöksenteon kriteerille c (Rounis ym., 2010). Teorian perusteella voidaan siis tehdä päätelmiä yksilön herkkyydestä havaita tietty signaali ja siitä, mitä yksilö havaintonsa perusteella mahdollisesti vastaa. Signaalidetektio teorian avulla voidaan analysoida ja kuvailla päätöksiä tai valintoja, jotka on tehty epäselvissä tai epävarmoissa tilanteissa (Wickens, 2002). Vaikka tyypillisesti päätöksentekoon liittyvät valinnat jakautuvat kyllä ja ei -vaihtoehtoihin, ovat vaillinaisen informaation takia virheelliset vastaukset hyvin mahdollisia.

1.2 Metakognitio

Metakognitiolla on käsitteenä pitkä historia ja sitä voidaan kuvailla esimerkiksi ajattelemisen ajattelemiseksi. Flavell (1979) kuvaa metakognition olevan tietoa ja tietoisuutta kognitiivisista ilmiöistä. Metakognitioon liittyvä tutkimus käsittelee Koriatin (2007) mukaan sitä prosessia, jota ihminen käyttää reflektoidessaan omia kognitiivisia toimintojaan. Erityisen tärkeää on se,

miten yksilö käyttää tätä tietoa säädellössään omaa informaation prosessointiaan ja käyttäytymistään. Flemingin (2024) mukaan oman toiminnan säätelmissen osalta on keskeistä, miten arviot omasta kognitiivisesta tai fyysisestä suoriutumisesta vastaavat todellisuutta. Jos yksilöllä on esimerkiksi haasteita tunnistaa oman kuulon heikkeneminen, voi olla haastavaa omaksua tapoja tilanteen helpottamiseksi.

Ihminen hyödyntää metakognitiota myös puhuessaan (Sharwood, 2017). Kun esimerkiksi ääntämisestä tai kielellisistä ilmauksista tulee aktiivisen reflektion ja analyysin kohteita, voidaan puhua metakognitiosta. Metakognitiivisen tarkastelun alla voi olla niin fonologinen ja syntaktinen tieto kuin semantiikkaan ja pragmatiikkaan liittyvä osaaminen. Tyypillisessä puhetilanteessa esimerkiksi sanojen, sanajärjestyksen ja aikamuodon valinta tapahtuu alitajuisesti. Usein puheeseen liittyvää tietoista reflektointia tapahtuu silloin, kun kohtaamme jonkin yllättävän ongelman, kuten vaikeuden kuvailla jotain asiaa tilanteeseen sopivilla sanoilla, tai kun havaitsemme virheen omassa ääntämisessämme.

Puheen tuottoa monitoroidaan sekä sisäisesti ennen tuottoa että ulkoisesti tuoton aikana ja sen jälkeen (Gauvin ym., 2016). Sisäistä monitorointia on pohdittu tapahtuvan esimerkiksi sanojen valintaan ja syntaktiseen rakentamiseen liittyen (Postma, 2000). Ulkoiseen monitorointiin lukeutuu esimerkiksi kuulopalautteena saadun tiedon käsittely (Gauvin ym., 2016). Ihminen korjaa omaa puhettaan keskustelujen ja monologiensa aikana. Esimerkiksi Manfra ja työryhmä (2016) mainitsevat, että näitä korjauksia tapahtuu sekä tiedostetusti että tiedostamattomasti. Puheessa esiintyviä virheitä verrataan sisäisiin kriteereihin, jolloin niiden onnistunut korjaaminen on mahdollista (Postma, 2000). Virheitä voi esiintyä esimerkiksi käsitteen, syntaksin, äänten tai prosodian tasolla.

Vaikka metakognitiota on tutkittu paljon esimerkiksi muistin ja tarkkaavaisuuden osalta, on sen systemaattinen tutkimus kielen osalta toistaiseksi vähäistä (Kartushina, 2023). Kokonaisuudessaan logopedian näkökulmasta tutkimus on hyvin rajallista. Lähelle kyseistä tieteenalaa osuu esimerkiksi lähiaikoina tehty metakognitiivinen tutkimus vieraan kielen äänteiden tuottamisesta ja havaitsemisesta (Kartushina ym., 2023). Logopedian kannalta olennaista olisi tutkia esimerkiksi sitä, miten ihmiset arvioivat tuottamansa äänteet. Onko yksilön mielestä tuottamansa ääni tai äänne esimerkiksi sellainen, kuin sen on tarkoituskin? Tähän aukkoon tutkimustiedossa pyrimme osin vastaamaan selvittämällä, kuinka ihmiset metakognitiivisesti arvioivat äänenkorkeuden muutosta tuottamassaan äännössä.

1.2.1 Vastausvarmuus

Oman suoriutumisen arviointia käytetään tuomaan lisätietoa havaitsemisesta ja päätöksenteosta. Koehenkilöä voidaan pyytää arvioimaan omaa vastaamistaan esimerkiksi kokeen jälkeen tai aina jokaisen päätöksenteon jälkeen. Tällaista metakognitiivista tehtävää, jossa koehenkilön tulee antaa vastausvarmuus tekemälleen valinnalle tai vastaukselle, kutsutaan myös tyypin 2 tehtäväksi (Maniscalco & Lau, 2012).

Kun tarkastellaan sitä, kuinka hyvin vastausvarmuus ennustaa aitoa suoriutumista, saadaan tietoa henkilön metakognitiivisesta sensitiivisyydestä (Maniscalco & Lau, 2012). Metakognitiivinen sensitiivisyys eli tyypin 2 sensitiivisyys ilmaisee, kuinka hyvin ja millä varmuudella yksilö pystyy erottelemaan oikeat vastaukset vääristä. Tyypin 2 sensitiivisyys poikkeaa siis tyypin 1 sensitiivisyydestä, jolla viitataan yksilön kykyyn tunnistaa ja erottaa ärsykkeet. Korkeat vastausvarmuudet oikeiden vastausten jälkeen sekä matalat vastausvarmuudet väärin vastausten jälkeen kertovat korkeasta metakognitiivisesta sensitiivisyydestä (Fleming, 2017). Lisäksi vastausvarmuus antaa osviittaa siitä, mihin havaittu signaali asettuu päätöksentekoakselilla (Wickens, 2001). Jos havainto asettuu kauas päätöksenteon kriteeristä c , voidaan päätöksen päätellä olleen helppo.

1.2.2 Metakognition mittaaminen ja tunnusluvut

Maniscalco & Lau (2012) esittelevät artikkelissaan tyypin 2 sensitiivisyydelle mittarin $meta-d'$, joka linkittyy suoraan signaalidetektiteorian tyypin 1 arvoon d' . $Meta-d'$ on tunnusluku, joka havainnollistaa sitä määrää signaalista, mikä yksilöllä on metakognitiivisesti hyödynnettävissä kulloisessakin metakognitiota mittaavassa testissä (Rounis ym., 2010).

$Meta-d'$ ilmaistaan samalla skaalalla kuin d' , mikä mahdollistaa näiden tunnuslukujen vertailun (Maniscalco & Lau, 2012). Yksilön voidaan tulkita suoriutuvan metakognitiivisesti optimaalisella tavalla, kun $meta-d'=d'$. Tällöin tehtävän suorittamisessa on hyödynnetty kaikki saatavilla oleva informaatio. Tilanteissa, joissa $meta-d' < d'$, metakognitiivisen prosessin vuoksi ikään kuin menetetään osa siitä signaalista, joka on hyödynnettävissä tyypin 1 tehtävässä (Rounis ym., 2010).

Vertailu mahdollistaa myös metakognitiivisen tehokkuuden mittaamisen (Fleming & Lau, 2014). Metakognitiivinen tehokkuus kuvaa yksilön metakognitiivista tasoa, kun otetaan huomioon yksilön suoriutuminen tietyssä tehtävässä. Tätä voidaan tarkastella vertaamalla

Meta-d':tä suhteessa d':hen. Kun $\text{Meta-d}'/d' = 1$, on kyseessä ideaali metakognitiivinen tehokkuus. Tätä tehokkuutta kuvaavaa meta-d':n ja d':n suhdelukua kutsutaan Mratioksi. Mratio on metakognition tarkastelussa informatiivisempi kuin esimerkiksi meta-d' yksinään, sillä Mratio suhdelukuna neutralisoi tyypin 1 suoriutumisen vaikutusta. Mratiota on aikaisemmin hyödynnetty esimerkiksi tutkittaessa metakognitiivisten prosessien käsittelyä aivoissa (Baird ym., 2013; McCurdy ym., 2013) ja kaksikielisyyden vaikutusta metakognitiiviseen prosessointiin (Folke ym., 2016; Ordin ym., 2020).

Meta-d':llä on etunsa verrattuna muihin tyypin 2 sensitiivisyyden mittareihin. Useiden mittareiden kohdalla on ongelmallista, että ne eivät erottele toisistaan tyypin 2 sensitiivisyyttä ja tyypin 2 vinoumaa (Maniscalco & Lau, 2012). Tyypin 2 vinoumalla eli metakognitiivisella vinoumalla tarkoitetaan esimerkiksi tilannetta, jossa itsevarmuus vastauksen oikeellisuudesta on korostuneen korkea tai matala (Fleming & Lau, 2014). Metakognitiivista suoriutumista mittaavat menetelmät ovat myös vaihtelevasti riippuvaisia tyypin 1 suoriutumisesta (Guggenmos, 2021). Metakognitiiviseen sensitiivisyyteen voi siis vaikuttaa jo pelkästään mittarina käytetyn kokeen tai tehtävän luonne, jolloin yksilön metakognitiivinen sensitiivisyys voi vaihdella riippuen siitä, onko koe hänelle helppo vai vaikea (Fleming & Lau, 2014). Barret ja tutkimusryhmä (2013) määrittelevät artikkelissaan meta-d':n olevan signaalidetektio teoriaan pohjautuvista metakognition mittareista teoreettisesti luotettavin ja empiirisesti hyödyllisin verrattuna muihin esitettyihin mittareihin. Nykyään metakognitiivisen sensitiivisyyden tunnusluku meta-d' sekä meta-d':stä ja d':stä johdettu metakognitiivisen tehokkuuden tunnusluku Mratio ovat käytössä lukuisissa metakognitiota mittaavissa tutkimuksissa. Nämä tunnusluvut ovat käytössä myös tässä tutkimuksessa.

1.3 Koeasetelman rakentaminen

1.3.1 Perturbaatioparadigma

Ihminen monitoroi ja muokkaa omaa ääntään kuulopalautteen perusteella (Guenther, 2016). Kun omassa äänessä havaitaan virhe, pyritään ääni palauttamaan ennalleen pienillä korjauksilla. Näitä korjauksia tapahtuu myös tiedostamattomasti (Franken ym., 2018; Hafke, 2008). Esimerkiksi tilanteessa, jossa oma ääni kuulostaa korkeammalta kuin oli tavoiteltu, on ihmisten havaittu muokkaavan ääntään automaattisesti matalammaksi (Burnett ym., 1998).

Perturbaatiolla viitataan tässä tutkimuksessa kuulopalautteen muokkaamiseen, jota yleisesti toteutetaan esimerkiksi nostamalla tai laskemalla äänenvoimakkuutta tai äänenkorkeutta.

Tutkimustilanteissa käytetään usein pidennettyä vokaalin ääntöä, jonka aikana perturbaatio esiintyy (Chen ym., 2006). Perturbaatiot voivat olla esimerkiksi lyhyitä ja odottamattomia (mm. Bauer ym., 2006; Schenck ym., 2023) tai kestoiltaan hieman pidempiä ja helpommin havaittavissa olevia (mm. Behroozmand ym., 2011; Hain ym., 2000). Menetelmää käytetään paljon tutkimuksissa, joissa pyritään selvittämään, miten ihmiset korjaavat omaa puhettaan tai ääntämistään kuulopalautteen virheen seurauksena (esim. Bauer ym., 2006; Burnett ym., 1998; Chen ym., 2006; Larson ym., 2007; Schenck ym., 2023).

Äänen tuoton korjaukseen keskittyvän tutkimuksen lisäksi menetelmää on käytetty laajasti esimerkiksi kuulopalautteen virheen neuraalisen perustan tutkimuksissa (esim. Behroozmand ym., 2009; Flagmeier ym., 2014; Korzyukov ym., 2012; Parkinson ym., 2013). Perturbaatioparadigma mahdollistaa myös oman äänen metakognitiivisen arviointikyvyn tutkimisen, kun tutkittavaa pyydetään antamaan vastauksen lisäksi myös vastausvarmuus. Tässä tutkimuksessa tarkastellaan metakognitiota kyseisestä näkökulmasta. Käyttämämme perturbaatioparadigma pyrkii mittaamaan yksilön metakognitiivista tehokkuutta äänenkorkeuden perturbaatioiden havaitsemisen kautta. Yhtenä tutkimuksen tavoitteena on tutkia, kuinka hyvin ihmiset pystyvät metakognitiivisesti arvioimaan virheitä kuulopalautteessa, ja esiintyykö tässä kyvyssä vaihtelua.

1.3.2 Pakkovalinta

Havaitsemista voidaan mitata esittämällä joko yksi ärsyke, joka sisältää tai ei sisällä haettua ominaisuutta eli signaalia, tai useampi ärsyke, joista valitaan oikea vaihtoehto (Wickens, 2001). Kun käytetään useampaa ärsykettä, joista on valittava oikea vaihtoehto, puhutaan pakkovalinnasta. Pakkovalinnassa ärsykeitä verrataan toisiinsa, kun taas yhden ärsykkeen tehtävissä ärsykettä arvioidaan riippumattomana kaikesta muusta. Paljon käytetty pakkovalinta-asetelma on 2AFC (two-alternative-forced choice), jossa kahdesta ärsykkeestä valitaan toinen ennalta määriteltujen kriteerien pohjalta (Aleci, 2021). Yhden ärsykkeen tehtävällä puolestaan viitataan esimerkiksi kyllä–ei-asetelmaan.

Lee ja kollegat (2018) tutkivat julkaisussaan 2AFC ja kyllä–ei-vastausvaihtoehtojen käytön vaikutusta metakognitiivisen tehokkuuden mittaamisessa. Metakognitiivisen tehokkuuden korrelaatioissa näkyi eroja riippuen siitä, mitä vastaustapaa oli käytetty. Vastaustapoja tutkittiin käyttämällä niitä kahdessa erilaisessa metakognitiivista tehokkuutta mittaavassa tehtävässä. Kun vastaukset annettiin tehtäviin 2AFC-asetelmalla, korreloivat tulokset keskenään tarkoittaen koehenkilöiden metakognitiivisen tehokkuuden olleen samankaltaista tehtävästä

riippumatta. Kyllä–ei-vaihtoehtoja tai molempia vastaustapoja käytettäessä korrelaatiota ei kuitenkaan löytynyt. Tuloksia tukevat myös aikaisemmat tutkimukset, joissa molempia vastaustapoja käytettäessä korrelaatiota ei ollut havaittavissa (Baird ym., 2013; Baird ym., 2015; Sadeghi ym., 2017). Näin ollen esimerkiksi Lee ja kumppanit (2018) suosittlevat 2AFC-asetelman käyttöä metakognition tutkimuksessa.

Lee ja kollegat (2018) totesivat kyllä–ei-vastauksien vaikuttavan metakognitiiviseen arvioon etenkin ei-vastauksen jälkeen. Syyksi esitettiin esimerkiksi päätöksen monimutkaisuutta sen vaatiessa enemmän työmuistilta kuin 2AFC-asetelman, jossa ärsykeitä verrataan suoraan keskenään. Myös Maniscalco ja Lau (2011) ovat havainneet metakognitiivisen sensitiivisyyden olevan heikompaa ei-vastausten jälkeen. Wickens (2001) toteaa kirjassaan asetelmien eroavaisuudesta sen, ettei koehenkilön tarvitse pakkovalinnassa ottaa huomioon vastausvinoumaa samalla tavalla kuin kyllä–ei-asetelmassa, sillä pakkovalinnassa signaali on aina läsnä. Sen sijaan saattaa esiintyä ajallista vinoumaa eli esimerkiksi taipumusta valita useammin ensimmäisenä esitetty vaihtoehto.

1.4 Test-retest-reliabiliteetti metakognition mittaamisessa

Yhtenä tutkimuksemme tavoitteena on tutkia, toimiiko käyttämämme koeasetelma yksilön metakognitiivisen tehokkuuden mittaamisessa. Toimiakseen menetelmän tulee olla luotettava. Tätä ominaisuutta tutkitaan tarkastelemalla test-retest-reliabiliteettia. Test-retest-menetelmässä sama testi tehdään samalle koehenkilölle kahtena eri ajankohtana, jolloin voidaan laskea korrelaatio kahden testauskerran tuloksille (Carmines & Zeller, 1979). Oletuksena on, että tulokset korreloivat keskenään testauskertojen välissä kuluneesta ajasta riippumatta, sillä testit mittaavat samaa muuttujaa. Todellisessa tilanteessa testausten välinen korrelaatio ei kuitenkaan ole täydellinen.

Yleisesti ottaen minimivaatimus kliinisen testin reliabiliteetille on $r = .90$ ja tutkimuskäyttöön tarkoitetun testin reliabiliteetin odotetaan olevan ainakin tasoa $r = .70$ (Sellbom & Suhr, 2019). Toisaalta sopivan test-retest-reliabiliteetin tason määrittely ei aina ole tutkittavasta aiheesta tai käytetystä menetelmästä riippuen kovin yksinkertaista. Todellisuudessa kaikki mitattavat ominaisuudet ja niissä käytetyt mittarit eivät välttämättä täytä ajallisen pysyvyyden vaatimuksia. Hunsley ja Mash (2008) määrittelevät riittäväksi mittarin test-retest-reliabiliteetiksi vähintään $r = .70$, kun testauskertojen välillä on useita päiviä tai useita viikkoja.

Guggenmosin (2021) mukaan myös metakognition tutkimuksessa kahden eri testauksen tulosten tulisi olla keskenään vastaavanlaisia, jos testauskerrat ovat muuten samanlaiset. Metakognitiivista suoriutumista mitattaessa mittausvirheiden voi kuitenkin olettaa olevan suhteellisen korkeita, sillä tulokseen vaikuttavat mittausvirheet niin tyypin 1 kuin tyypin 2 suoriutumisessa. Tämän lisäksi tyypin 1 suoriutuminen jo itsessään voi olla suhteellisen ailahteleva mittari, sillä se lasketaan muuttujasta, jolla on kaksi arvoa. Toisin sanoen tyypin 1 suoriutuminen lasketaan sen perusteella, onko vastaus oikein vai väärin.

Guggenmos (2021) perustelee artikkelissaan M_{ration} test-retest-reliabiliteetin olevan riippuvainen sekä tyypin 1 suoriutumisesta että toistojen määrästä. Tyypin 1 suoriutumisen ollessa lähellä arvaustasoa, esimerkiksi sen ollessa 60 % tai alle, on Pearsonin korrelaatiokerroin matala ($r \approx .40$), vaikka toistojen määrä olisi 400–600. Samalla toistomäärällä ja tyypin 1 suoriutumisen ollessa 80 %, on kokeen test-retest-reliabiliteetti huomattavasti korkeampi ($r \approx .80$). Koetoistojen määrällä on myös oma vaikutuksensa M_{ration} test-retest-reliabiliteettiin. Tyypin 1 suoriutumisen ollessa matala tulisi toistojen määrän olla suhteessa korkeampi, jotta matala suoriutumistaso saataisiin kompensoitua.

Viimeaikaiset julkaisut, jotka eivät vielä ole käyneet läpi vertaisarviointia, viittaavat siihen, ettei korkeaa test-retest-reliabiliteettia M_{ration} lle ole saavutettu (Kopčanová ym., 2023; Rahnev, 2023). Rahnevin (2023) tutkimuksessa M_{ration} korrelaatiot vaihtelivat heikosta keskinkertaiseen riippuen toistojen määrästä (50 toistoa $r = .17$, 100 toistoa $r = .24$, 200 toistoa $r = .29$ ja 400 toistoa $r = .42$). Myös Kopčanován ja kumppaneiden (2023) tutkimuksessa test-retest-reliabiliteetit jäivät heikoiksi. Havaitsemisen tehtävässä M_{ration} sisäkorrelaatiokerroin (ICC) sai arvon $r = .27$ ja tietämykseen liittyvässä tehtävässä vastaava korrelaatio oli $r = .18$. Vaikka M_{ration} on yksi tämän hetken käytetyimmistä metakognition tunnusluvuista, ei kattavia tutkimustuloksia M_{ration} test-retest-reliabiliteetista ole vielä juurikaan saatavilla. Osittain tähän perustuen tulemme tekemään test-retest-reliabiliteetin analyysit käyttäen kahta erilaista lähestymistapaa, sillä aikaisempaa tietoa toimivista aineiston rajauksista tai perusteluista rajauksille ei ollut juurikaan saatavilla. Analyyseissa sovelletaan Leen ja kumppaneiden (2018) ja Guggenmosin (2021) tutkimusartikkeleissa esiteltyjä aineiston rajauksia.

2 Tutkimuskysymykset

Tämän tutkimuksen tavoitteena on tarkastella sitä, kuinka hyvin ihmiset pystyvät metakognitiivisesti arvioimaan virheitä kuulopalautteessa. Metakognitiivisia arvioita omasta suoriutumisesta analysoidaan vertailemalla d' :tä ja meta- d' :tä keskenään, jolloin saadaan tietoa siitä, kuinka optimaalista suoriutuminen on. Lisäksi tavoitteena on arvioida käyttämämme perturbaatiomenetelmän luotettavuutta kuulopalautteen virheen havaitsemiseen liittyvässä tutkimuksessa. Käyttämämme menetelmän toimivuutta metakognitiivisen tehokkuuden mittaamiseen arvioidaan test-retest-reliabiliteetin kautta. Tarkastelemme test-retest-reliabiliteettia kahteen eri näkökulmaan perustuvien rajausten kautta (Guggenmos 2021; Lee ym., 2018). Guggenmosin (2021) tapauksessa rajaukset tehdään M ration arvoihin perustuen, kun taas Leen ja kumppaneiden (2018) artikkelissa rajaukset painottuvat oikeinvastausprosenttiin. Käytetyistä poissulkukriteereistä kerrotaan tarkemmin Tilastolliset analyysit -luvussa.

Aihetta ei ole aikaisemmin tutkittu, ja sen voidaan nähdä tarjoavan arvokasta tietoa puheen ja äänen metakognitiivisesta arvioimisesta. Tämän lisäksi tutkimaamme menetelmää voidaan mahdollisesti tulevaisuudessa käyttää erilaisten asiakasryhmien kanssa esimerkiksi puheen havaitsemisen vinouman tutkimisessa. Koeasetelma on kuvattu tarkemmin Menetelmä-osiossa.

Tutkimuskysymykset ovat:

1. *Kuinka hyvin ihmiset pystyvät metakognitiivisesti arvioimaan virheitä kuulopalautteessa?*
2. *Toimiiko käyttämämme menetelmä yksilön metakognitiivisen tehokkuuden mittaamisessa? Mikä on käyttämämme menetelmän test-retest-reliabiliteetti?*

2.1 Hypoteesit

Erityisesti visuaaliseen havaitsemiseen liittyvissä tutkimuksissa, joissa on käytetty meta- d' :tä metakognitiivisen sensitiivisyyden mittarina, on meta- d' :n usein havaittu olevan keskimäärin pienempi kuin d' (Fleming ym., 2014; Maniscalco & Lau, 2012; McCurdy ym., 2013, Vlassova ym., 2014). Tämän perusteella hypoteesimme on, että metakognitiivisen tehokkuuden tunnusluvun eli M ration arvo tulee olemaan keskimäärin alle 1 eli toisin sanoen metakognitiivinen suoriutuminen on alle optimaalisen tason. Aikaisempiin artikkeleihin

perustuen on kuitenkin odotettavissa yksilöllistä vaihtelua sekä metakognitiivisessa sensitiivisyydessä kuin tehokkuudessa.

Yleisesti ottaen metakognition tutkimiseen liittyy test-retest-reliabiliteetin osalta haasteita. Tutkimusmenetelmässämme ja tutkittavassa aiheessa on useita eri tasoja ja vaiheita, jotka voivat vaikuttaa osaltaan test-retest-reliabiliteettiin. Vaikka tulemme käyttämään ennalta suositeltuja rajoituksia (Guggenmos, 2021; Lee ym., 2018) aineiston tarkastelussa, ei test-retest-reliabiliteetti todennäköisesti tule olemaan kovin korkea. Oletuksena on, että yleisesti kliinisesti käytetyissä testeissä edellytetty reliabiliteetti $r > .90$ ei käyttämämme koeasetelman osalta toteudu. Näin ollen esimerkiksi Hunsleyn ja Mashin (2008) asettama riittävän korrelaation vähimmäistavoite $r = .70$ olisi koeasetelmamme osalta realistisempi vertailukohta. Toisaalta aikaisempiin tutkimustuloksiin (Guggenmos, 2021) ja aihetta tarkasteleviin toistaiseksi vertaisarvioimattomiin julkaisuihin (Kopčanová ym., 2023; Rahnev, 2023) viitaten on mahdollista, että erityisesti metakognitiivisen tehokkuuden korrelaatio jäisi kauas myös tästä tavoitteesta. Odotettavissa kuitenkin on, että kahden eri testauskerran välillä on löydettävissä positiivinen korrelaatio niin sensitiivisyyden eli d' :n, metakognitiivisen sensitiivisyyden eli $meta-d'$:n kuin metakognitiivisen tehokkuuden eli $Mratio$ arvojen osalta.

Hypoteesimme tutkimuskysymyksittäin ovat seuraavat:

- 1. Metakognitiivisen suoriutumisen taso on keskimäärin alle optimaalisen tason.*
- 2. Kahden eri testauskerran välillä on havaittavissa positiivinen korrelaatio d' , $meta-d'$ ja $Mratio$ -arvoille. Odotettavissa on, että $r < .70$ metakognitiivisen tehokkuuden osalta.*

3 Menetelmät

3.1 Koehenkilöt

Tutkimukseen rekrytoitiin yhteensä 60 koehenkilöä. Rekrytoinnissa hyödynnettiin Turun yliopiston psykologian ja logopedian laitoksen koehenkilöjärjestelmää (psykoehlot.utu.fi) sekä mukavuusotantaa. Sisäänottokriteerinä tutkimukseen osallistumiselle oli 18–35 vuoden ikä. Edellytyksenä osallistumiselle oli, että tutkittavalla ei ollut kuuloon, puheeseen tai ääneen vaikuttavaa sairautta tai häiriötä, kurkunpään tai nielun sairautta eikä havaitsemiseen merkittävästi vaikuttavaa neurologista sairautta. Poissulkukriteerinä oli aikaisempi osallistuminen saman tutkimusprojektin tutkimukseen.

Yhteensä 59 koehenkilöä suoritti kokeen kahteen kertaan. Iät vaihtelivat 19–32 vuoden välillä ja keskiarvo oli noin 25 vuotta. Koehenkilöistä 48 oli naisia ja 11 miehiä. Toinen testauskerta suoritettiin noin 7 vuorokauden jälkeen ensimmäisestä testauskerrasta ja tavoitteena oli, että toinen testauskerta suoritettaisiin 5–9 päivän sisällä ensimmäisestä testauskerrasta. Aikataulullisista ja terveydellisistä tekijöistä johtuen testausväli vaihteli 5–19 vuorokauden välillä. Keskiarvo testausvälille oli 8,20 ja keskihajonta 3,08 päivää. Metakognitiivisen tehokkuuden voidaan ajatella olevan suhteellisen pysyvä ominaisuus, joten pienellä vaihtelulla ei pitäisi olla vaikutusta tuloksiin. Tutkimukseen osallistuminen oli täysin vapaaehtoista ja kaikilta tutkittavilta pyydettiin kirjallinen suostumus osallistua tutkimukseen. Tutkittavat saivat osallistumisesta rahallisen korvauksen tai vaihtoehtoisesti merkinnän suorituksesta koehenkilöjärjestelmään. Rahallinen korvaus oli suuruudeltaan 10 euroa tutkimuskertaa kohden.

Lopulliseen aineistoon otettiin mukaan 56 koehenkilöä. Kolmen koehenkilön tulokset jätettiin pois analyyseistä perustuen joko puutteellisiin tuloksiin, alle 50 % oikeinvastausprosenttiin, jolloin tulos on ollut alle arvaustason, tai 100 % oikeinvastausprosenttiin, jolloin suoriutuminen tyyppiin 1 tehtävässä on ollut virheetöntä. Kahdessa viimeisessä tapauksessa perturbaation koko oli suuruudeltaan moninkertainen verrattuna muuhun aineistoon. Lopullista 56 koehenkilön aineistoa käytettiin tarkasteltaessa perturbaation kokoa sekä koehenkilöiden metakognitiivista arviota omasta suoriutumisesta. Test-retest-reliabiliteettiä liittyvissä analyyseissä aineistoa tarkasteltiin perustuen Guggenmosin (2021) sekä Leen ja kumppaneiden (2018) artikkeleissaan tekemiin rajauksiin. Guggenmosin (2021) kriteereitä vastasi 47 koehenkilöä ja Leen ja

kumppaneiden (2018) kriteereitä 39 koehenkilöä. Päällekkäisiä koehenkilöitä, jotka täyttivät molemmat kriteerit ja siten sisällytettiin kumpaankin analyysiin, oli yhteensä 35.

3.2 Laitteisto ja lomakkeet

Kokeet suoritettiin Turun yliopiston tiloissa erillisessä koehuoneessa, jossa sijaitsivat käytettävät laitteistot ja materiaalit. Tutkija oli läsnä koetilanteen ajan ja vastasi laitteiden käytöstä ja toimivuudesta.

Koehenkilön tuottama ääntö nauhoitettiin käyttäen mikrofonia (Audio Technica AT2035), joka oli kytketty MOTU-äänikorttiin. Mikrofonin säädettiin noin 2–3 cm etäisyydelle koehenkilöstä. Koehenkilö kuuli tuottamansa äännön kuulokkeiden (Beyerdynamic DT 770M 80ohm) kautta äänenvoimakkuuttaan vastaavalla voimakkuudella. Kuulokkeista saatuun auditiiviseen palautteeseen oli lisätty myös taustakohinaa, joka oli voimakkuudeltaan noin 68 dB. Taustakohinan avulla pyrittiin minimoimaan ilma- ja luujohteinen auditiivinen palaute. Koehenkilöt seurasivat kokeen kulkua 27 tuumaisesta näytöstä (Dell LED display), joka oli sijoitettu alle metrin etäisyydelle koehenkilöistä. Vastaukset annettiin erillisen näppäimistön kautta (Cedrus response pad). Koe oli ohjelmoitu Psychtoolbox-3 sovellukseen (Brainard, 1997; Kleiner ym., 2007) osana MATLAB R2022a -ohjelmistoa (The MathWorks Inc.). Perturbaatioiden toteutuksessa hyödynnettiin Audapter-ohjelmistoa (Cai ym., 2008).

Koehenkilöt täyttivät tutkimuskäynneillä kaksi lomaketta. Ensimmäisellä käynnillä täytettiin musikaalisuuskyselely ja toisen käynnin päätteeksi suomennettu MSAS-kyselylomake (Metacognition Self-Assessment Scale) (Pedone ym., 2017), jossa tutkittava arvioi omia kykyjään tunnistaa ja kuvailla ajatuksiaan, tunteitaan ja sosiaalisia suhteitaan. Näitä kyselyitä ei käsitellä tässä tutkielmassa.

3.3 Koeasetelma

Käyttämämme koeasetelma on kahden intervallin pakkovalintatehtävä: koehenkilö ääntää pitkään vokaalia /u/, ja perturbaatio eli äänenkorkeuden äkillinen nosto tulee joka kerta joko ensimmäisessä (A) tai toisessa (B) aikaikkunassa. Koehenkilöä pyydetään vastaamaan kummassa aikaikkunassa perturbaatio hänen mielestään esiintyi ja kuinka varma hän oli vastauksestaan. Vastaukset annetaan kuusiosaisella asteikolla, joka sisältää vaihtoehdot varma, melko varma ja epävarma sekä A että B vaihtoehdolle. Vastausvaihtoehdon A tai B

valitseminen sekä vastausvarmuus annetaan samalla painalluksella, sillä sen on todettu parantavan reliabiliteettia (Guggenmos, 2021).

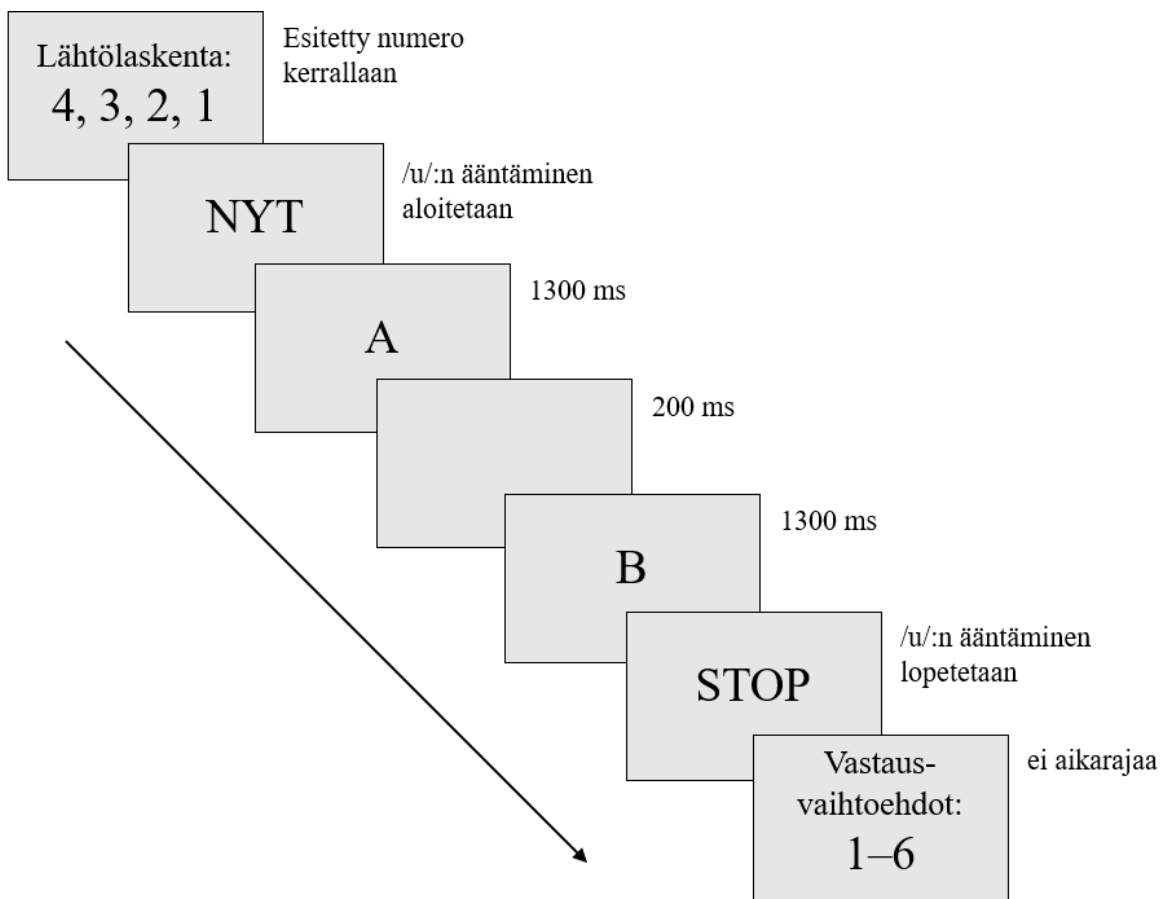
Koeasetelmassamme käytetään äänenkorkeuden perturbaatiota, sillä sitä on tutkittu melko paljon. Muokkaukset äänenkorkeudessa on suhteellisen helppo havaita ja varsinainen havaitsemistehtävä on haluttu pitää mahdollisimman yksinkertaisena, sillä kokeessa metakognitiivisen suoriutumisen mittaaminen on ollut keskiössä. Äännettäväksi vokaaliksi valikoitui /u/, mutta on mahdollista, että jokin toinen vokaali toimisi ominaisuuksiltaan yhtä hyvin tai jopa paremmin. Toisaalta ei ole tietoa, onko vokaalin valinnalla merkitystä metakognition mittaamisen kannalta.

Perturbaation koko määritellään erikseen jokaiselle koehenkilölle käyttäen Bayesilaista adaptiivista QUEST-portaikkoa (Watson & Pelli, 1983). Koska ihmisten sensitiivisyydessä erottaa ärsykkeet esiintyy vaihtelua, odotetaan yksilöllisen perturbaation koon tekevän tehtävästä kaikille koehenkilöille vaativuudelta samankaltaisen. Kun suoriutuminen vakioidaan, saadaan paremmin esiin yksilöiden välisiä eroja metakognitiivisessa sensitiivisyydessä. Perturbaatio pidetään samankokoisena läpi varsinaisen kokeen, jolloin tarkkuuden ja varmuuden muutosten voidaan olettaa johtuvan koehenkilöstä riippuvista tekijöistä (Fleming & Lau, 2014). Perturbaatio mitataan erikseen molemmille tutkimuskerroille, sillä tarkastelussa on koko kokeen toimivuus ja test-retest-reliabiliteetti. Näin ollen perturbaation mittaamisen tulisi myös toimia molemmilla kerroilla samankaltaisesti ja perturbaation koon johtaa samankaltaiseen suoriutumiseen. Lisäksi koeasetelman pilotoinnissa oli havaittavissa tietynasteista oppimista perturbaation tunnistamisessa, joten samaa perturbaatiota käytettäessä voisi toisella testauskerralla tehtävän taso olla hieman liian helppo.

3.4 Kokeen kulku

Koehenkilöt osallistuivat tutkimuskäynnille kahtena erillisenä kertana ja molemmilla kerroilla suoritettiin sama perturbaatiokoe. Tutkimuskäynti koostui kolmesta osiosta: harjoitusosio, portaikko ja varsinainen koe, jotka pysyivät samana sekä ensimmäisellä että toisella tutkimuskäynnillä. Ensimmäisen käynnin alussa tutkija kertoi suullisesti tutkimuksen kulusta ja antoi koehenkilölle tutkimustiedotteen luettavaksi. Tutkittava täytti kaksi lomaketta, joista toinen oli suostumus tutkimukseen osallistumisesta ja toinen kartoitus tutkittavan musikaalisuudesta. Ennen kokeen alkua koehenkilö luki tietokoneen näytöltä tarkemman ohjeistuksen tutkimuksen kulusta ja sai halutessaan kysyä lisäkysymyksiä.

Koetilanteessa koehenkilö äänsi mikrofoniin jatkuvaa /u/-äännettä noin neljän sekunnin ajan kuullen samalla reaaliaikaisesti oman äänensä sekä taustakohinan kuulokkeista. Jokaista ääntä edelsi tietokoneen näytöllä esitetty lähtölaskenta 4, 3, 2, 1. Tämän jälkeen näytölle ilmestyi peräkkäin ohjeistukset NYT, A, B ja STOP. Koehenkilön tuli aloittaa ääntäminen nähdessään NYT ja jatkaa, kunnes näytölle ilmestyi STOP. Ääntäminen kesti kokonaisuudessaan noin neljä sekuntia, jonka aikana sekä A että B vaihtoehdot olivat näkyvillä kumpikin 1300 ms ajan (kuva 2). A vaihtoehto edusti ajallisesti yhden toiston ensimmäistä puolikasta ja B jälkimmäistä puolikasta. Jokaisen äännön aikana koehenkilön äänenkorkeutta muokattiin korkeammaksi joko vaihtoehdon A tai B aikana. Perturbaatio oli ohjelmoitu kestämään 200 ms ja sen sijainti oli satunnaistettu vastausvaihtoehtojen välillä sekä alkamaan joko 1200–1600 ms tai 2700–3100 ms jälkeen äännön aloittamisesta.



Kuva 2. Kaaviossa esitetty yhden toiston kulku. Koehenkilön vastattua alkoi uusi toisto taas lähtölaskennasta. A ja B osioiden välissä näkyi 200 ms ajan tyhjä näyttö erottamaan osiot toisistaan.

Ääntämisen jälkeen tutkittavaa pyydettiin vastaamaan, tapahtuiko muutos vaihtoehdon A vai B aikana sekä arvioimaan, kuinka varma oli antamastaan vastauksesta. Molempiin kysymyksiin

vastattiin samalla painalluksella. Vastaus annettiin näppäimistöllä, jossa oli vastausvaihtoehdot 1, 2, 3, X, 4, 5 ja 6. Näistä kolme ensimmäistä edusti vaihtoehtoa A ja kolme viimeistä vaihtoehtoa B. X näppäin ei ollut käytössä, joten tutkittavan oli valittava aina joko A:ta tai B:tä edustava numero. Vastausvaihtoehdot olivat seuraavat: 1 = varma A, 2 = melko varma A, 3 = epävarma A, 4 = epävarma B, 5 = melko varma B ja 6 = varma B. Vastauksen antamiselle ei ollut aikarajaa ja vaihtoehdot olivat näkyvillä auki selitettynä aina vastattaessa.

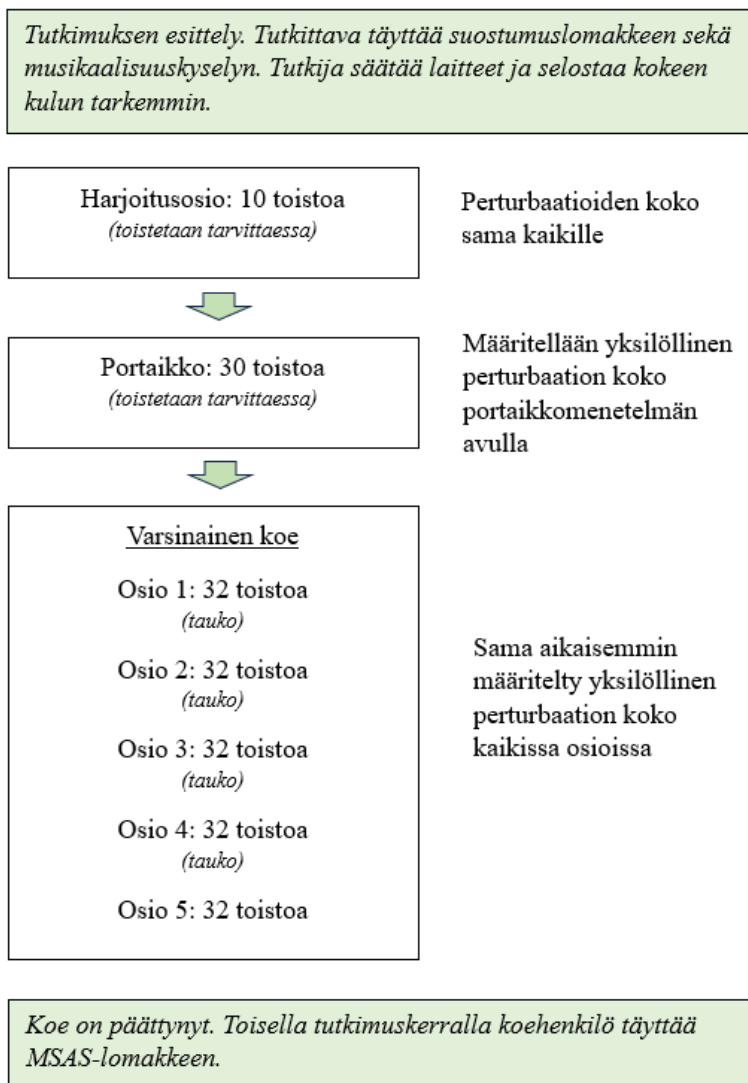
3.4.1 Kokeen osiot

Harjoitusosio sisälsi kymmenen ääntöä, ja perturbaatiot laskivat 240:stä 60 senttiin (C) muuttuen haastavammin havaittaviksi. Osio oli kaikille koehenkilöille samanlainen ja oletuksena oli, että perturbaatiot olivat suhteellisen helposti havaittavissa. Harjoituksessa tutkittava sai palautetta vastausten oikeellisuudesta ja tehtävän sai tarvittaessa suorittaa uudestaan. Jokaisen vastauksen jälkeen tietokoneen näytöltä näkyi, oliko annettu vastaus oikein vai väärin.

Harjoitusosion jälkeen koehenkilölle määriteltiin optimaalinen perturbaation koko eli kynnysarvo portaikkomenetelmällä. Tässä tutkimuksessa kynnysarvo oli määritetty tasolle, jolla koehenkilön odotettiin havaitsevan äänenkorkeuden perturbaation noin 75 % kerroista. Kullekin tutkittavalle laskettiin sopiva kynnysarvo QUEST-portaikkoa (Watson & Pelli, 1983) hyödyntäen. QUEST-portaikkoon pohjautuen järjestelmä voimisti perturbaatiota eli nosti äänenkorkeutta jokaisesta väärästä vastauksesta ja laski jokaisesta oikeasta vastauksesta. Toisin sanoen jokaisen oikean vastauksen jälkeen seuraava perturbaatio oli vaikeammin havaittavissa ja jokaisen väärän vastauksen jälkeen helpommin havaittavissa kuin edellinen perturbaatio. Perturbaatio oli aina korkeampi kuin koehenkilön oma äänenkorkeus. Portaikkomenetelmä sisälsi 30 toistoa ja tarvittaessa se suoritettiin uudestaan, kunnes portaikko tasaantui ja sopiva perturbaatio löytyi. Toistomäärä osoittautui pilotoinnissa pääsääntöisesti riittäväksi tasaamaan vastaukset. Menetelmän avulla laskettu perturbaation koko oli käytössä koko varsinaisen kokeen ajan.

Varsinainen koevaihe sisälsi viisi samanlaista osiota, joista jokaisessa oli 32 toistoa. Perturbaation koko pysyi samansuuruisena kaikissa osioissa ja toistoissa. Osioista toiseen siirryttäessä koehenkilöä kehoitettiin pitämään halutessaan taukoa. Kokeen ohjeistus näkyi näytöllä ennen jokaisen osion alkamista ja koehenkilö sai aloittaa uuden osion, kun oli siihen valmis. Koeasetelmat olivat lomakkeita lukuun ottamatta molemmilla kerroilla identtiset. Toisen tutkimuskäynnin lopuksi tutkittava täytti suomennetun MSAS-lomakkeen.

Tutkimuskäynnin kulku on esitetty kuvassa 3. Yhden käynnin kesto oli kokonaisuudessaan noin yksi tunti.



Kuva 3. Kaavio tutkimuskäynnin kulusta. Koeosuus, sisältäen harjoitusosion, portaikon ja varsinaisen kokeen, oli molemmilla tutkimuskerroilla samanlainen.

3.5 Tilastolliset analyysit

3.5.1 Metakognitiivinen suoriutuminen

Metakognitiivista suoriutumista analysoitiin toistettujen mittausten t-testillä. Molempien testauskertojen d' :tä sekä meta- d' :tä verrattiin suhteessa toisiinsa, jotta saatiin selville, kuinka hyvin koehenkilöt pystyivät metakognitiivisesti arvioimaan virheen kuulopalautteessa. Vertailussa oli mukana molemmilta tutkimuskäynneiltä saatu aineisto, jolloin jokaista tutkittavaa kohden saatiin mahdollisimman monta toistoa mukaan analyysiin.

3.5.2 Test-retest-reliabiliteetin analysointi

Test-retest-reliabiliteetin analysoinnissa käytettiin Pearsonin korrelaatiokerrointa (r) sekä toistettujen mittausten t -testiä. Korrelaatiokertoimet laskettiin ensimmäisen ja toisen testauskerran d' :lle, meta- d' :lle sekä M ratioille. Aineistoa analysoitiin hyödyntäen Guggenmosin (2021) sekä Leen ja kumppaneiden (2018) tutkimuksissaan käyttämiä rajoituksia.

Guggenmosin (2021) artikkelin aineistosta on poistettu M ration arvot, jotka ovat joko negatiivisia tai ylittävät 1,6:n rajan. Negatiivisten arvojen poistamista perustellaan sillä, että M ratio käyttäytyy ongelmallisella tavalla meta- d' :n ollessa negatiivinen. Tällaisessa tilanteessa d' :n kasvaessa M ratio saa korkeamman arvon eli metakognitiivinen tehokkuus kasvaa, vaikka tilanteen tulisi olla päinvastainen. Esimerkiksi tilanteessa, jossa meta- $d' = -1$ ja $d' = 2$ saa M ratio arvon $-0,5$, kun taas tilanteessa, jossa meta- $d' = -1$ ja $d' = 4$ saa M ratio arvon $-0,25$. Jälkimmäisessä esimerkissä metakognitiivinen suoriutuminen on vähemmän optimaalista eli d' :n ja meta- d' :n arvot ovat kauempana toisistaan, vaikka M ratio on arvoltaan korkeampi. Rajoituksen yläraja 1,6 puolestaan perustuu M ration keskiarvoon, joka oli Guggenmosin (2021) analyysissä noin 0,8. Näin ollen rajoituksen ääriarvot 0 ja 1,6 ovat yhtä kaukana keskiarvosta. Guggenmos (2021) suositukseen perustuen myös tässä tutkimuksessa päädyttiin samoihin raja-arvoihin, sillä M ration keskiarvo oli 0,82. Koehenkilöt, joiden M ratio jäi skaalan ulkopuolelle, jätettiin tässä vaiheessa pois analyysistä. Jäljelle jäi 47 koehenkilöä.

Leen ja kumppaneiden (2018) artikkelissa aineistoa rajataan useamman muuttujan osalta. Tutkijat päätyivät jättämään ulkopuolelle koehenkilöt, joiden d' oli alle 0,5, meta- d' alle 0 tai jos heidän oikeinvastausprosenttinsa poikkesi yli 15 % määritellystä tasosta. Lee ja kollegat (2018) selittivät d' :n pienen arvon vaikuttavan M ratioon ja tuottavan tälle mahdollisesti kohtuuttoman suuria arvoja. Oikeinvastausprosentin rajaamisella pyrittiin välttämään liian suurta variaatiota. Tutkijat olivat asettaneet portaikolle 75 % tavoitteen, mikä oli käytössä myös meidän koeasetelmassamme. Näin ollen Leen ja kumppaneiden (2018) rajausten pohjalta tehdyistä analyysistä poistettiin koehenkilöt, joiden oikeiden vastausten osuus oli alle 60 % tai yli 90 %. Kyseiset d' :tä, meta- d' :tä ja oikeinvastausprosenttia koskevat kriteerit täyttivät yhteensä 39 koehenkilöä.

Molemmat näkökulmat päädyttiin sisällyttämään analyysiin niiden erilaisten lähtökohtien perusteella. Guggenmos (2021) pyrkii rajaamaan aineistosta M ration ääriarvoja sillä perusteella, että niiden taustalla ovat usein esimerkiksi matala suoriutumisen taso tai vähäinen toistojen määrä. M ration rajausten seurauksena voidaan d' :n ja meta- d' :n test-retest-

reliabiliteetin odottaa olevan vahvempi, sillä jäljelle jäävä vaihtelu on näiden arvojen vaihtelua. Leen ja kumppaneiden (2018) tekemät rajaukset taas pyrkivät vakioimaan tyypin 1 suoriutumisen, jolloin oletettavasti yksilölliset erot metakognitiivisessa suoriutumisessa olisivat selkeämmin havaittavissa. Toisin sanoen tässä rajauksessa korostuu vaihtelu metakognitiivisessa tehokkuudessa eli Mratiossa. Koeasetelmissamme tavoitteena oli vakioida suoriutuminen lähelle asetettua 75 % tasoa. Reliabiliteettia ajatellen Leen ja kumppaneiden (2018) asettama skaala sisällyttää analyysiin kuitenkin runsaasti tuloksia, jotka ovat alle tavoitellun tason. Tämä on puolestaan vastoin Guggenmosin (2021) näkemystä tyypin 1 suoriutumisen vaikutuksesta reliabiliteettiin: kun tyypin 1 suoriutuminen paranee, arvausten suhteellinen osuus vähenee ja Mration reliabiliteetti kasvaa. Vaikka Leen ja kumppaneiden (2018) rajaukset tukevat enemmän asettamiamme tavoitteita tyypin 1 suoriutumisen vakioimisesta, on Guggenmosin (2021) tutkimuksessa fokus nimenomaan metakognition mittaamisen test-retest-reliabiliteetissa.

3.6 Tutkimuksen eettisyys

Tämä tutkimus on toteutettu osana Suomen Akatemian rahoittamaa tutkimusprojektia ”Oman puheen metakognitiivinen arviointi Parkinsonin tautia sairastavilla ja neurologisesti terveillä ihmisillä: Interventiotutkimus”. Tutkimusprojekti on saanut puoltavan lausunnon Varsinais-Suomen hyvinvointialueen eettiseltä toimikunnalta.

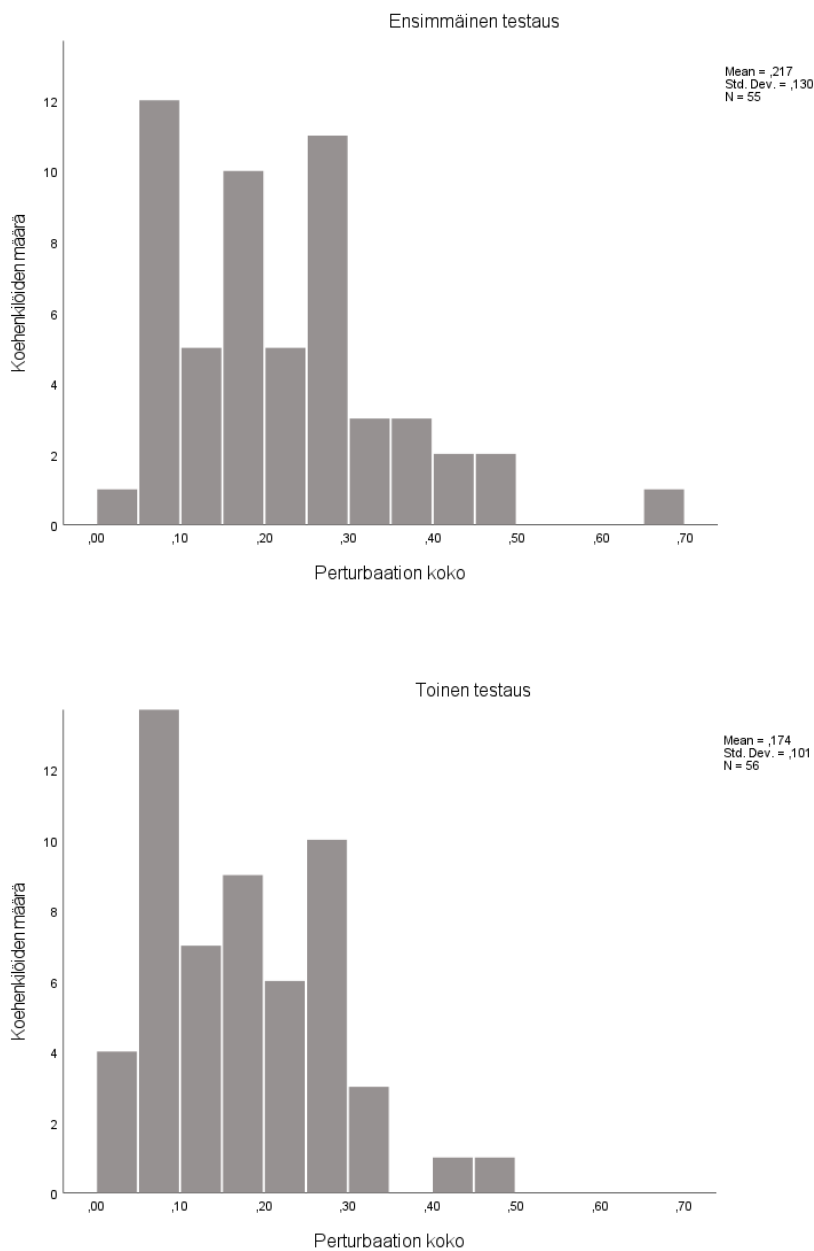
Tutkimukseen osallistuminen oli täysin vapaaehtoista ja koe oli mahdollista keskeyttää milloin tahansa joko tutkittavan tai tutkijan toimesta. Ennen kokeen alkamista tutkittaville selostettiin lyhyesti kokeen kulku ja tarkoitus, minkä jälkeen tutkittavalle annettiin tutkimustiedote luettavaksi. Tämän jälkeen tutkittavaa pyydettiin täyttämään suostumuslomakkeet tutkimukseen osallistumiselle. Jokainen tutkittava täytti kaksi lomaketta, yhden tutkijoille ja yhden itselleen.

Koehenkilöitä koskevat tunnistettavat tiedot olivat ainoastaan tutkimuksesta vastaavien tutkijoiden tiedossa, ja koehenkilöitä oli informoitu tutkijoiden salassapitovelvollisuudesta. Kerättyjä tietoja käsiteltiin koodattuna eikä koehenkilöitä ole mahdollista tunnistaa tuloksista, selvityksistä tai julkaisuista. Turun yliopisto toimii tutkimuksen toimeksiantajana ja tutkimusrekisterin rekisterinpitäjänä. Tutkimustiedotteessa koehenkilöille oli kerrottu rekisterin sisällöstä ja mahdollisuudesta halutessaan tilata rekisteriseloste.

4 Tulokset

4.1 Erot perturbaatioiden koossa

Koeasetelmassamme on pyritty vakioimaan suoriutumisen taso, joten on tärkeää tarkastella, onko perturbaation koko määräytynyt samalle tasolle molemmilla tutkimuskerroilla. Sillä, että koe on helpompi jommalla kummalla testauskerralla, voi olla huomattava vaikutus myös kokeen test-retest-reliabiliteettiin.

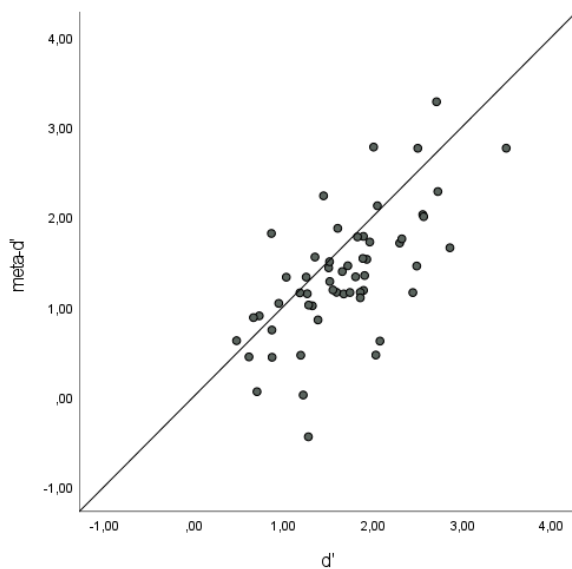


Kuva 4. Histogrammit perturbaatioiden jakautumisesta eri testauskerroilla. Kuvassa 0.10 kokoinen perturbaatio vastaa 10 senttiä (100 senttiä= yksi puolisävelaskel).

Pearsonin korrelaatiokertoimella ensimmäisen ja toisen testauskerran perturbaatioiden koon välillä esiintyi positiivinen lineaarinen yhteys ja tulos oli tilastollisesti merkitsevä ($r(56) = .60$, $p < .001$). Tarkastellessa perturbaation koon eroja toistettujen mittausten t-testillä havaittiin, että perturbaatio oli keskimäärin suurempi ensimmäisellä ($m = 0.22$, $SD = 0.13$) kuin toisella testauskerralla ($m = 0.17$, $SD = 0.10$) ja tulos oli tilastollisesti merkitsevä keskikokoisella efektillä ($t(55) = 3.02$, $p = .004$, $d = 0.40$). Perturbaatioiden kokoja on esitetty tarkemmin kuvassa 4. Perturbaatioiden lisäksi myös oikeinvastausprosentti oli suurempi ensimmäisellä ($m = .80$, $SD = .11$) kuin toisella testauskerralla ($m = .76$, $SD = .09$).

4.2 Metakognitiivinen arvio omasta suoriutumisesta

Metakognitiivisen suoriutumisen arvioinnissa verrattiin d' :n ja meta- d' :n suhdetta toisiinsa (kuva 5). Vertailussa yhdistettiin kahdelta tutkimuskerralta saadut tulokset jokaiselle koehenkilölle ja laskettiin näiden keskiarvo. Toistettujen mittausten t-testin perusteella d' ($m = 1.68$, $SD = 0.64$) oli keskimäärin korkeampi kuin meta- d' ($m = 1.35$, $SD = 0.70$). Tulos oli tilastollisesti merkitsevä keskisuurella efektillä ($t(55) = 4.50$, $p < .001$, $d = 0.60$). Yhteensä 14 koehenkilön ($N=56$) kohdalla meta- d' sai suuremman arvon kuin d' , ja keskiarvo molempien testauskertojen M ratioille alitti optimaalisen tehokkuuden ($m = 0.82$, $SD = .38$).



Kuva 5. Sirontakuvi koehenkilöiden meta- d' ja d' tunnusluvuista. Vinoviiva kuvaa optimaalista metakognitiivista suoriutumista testissä eli meta- $d' = d'$. Havaittavissa on vaihtelua koehenkilöiden välillä metakognitiivisessa suoriutumisessa.

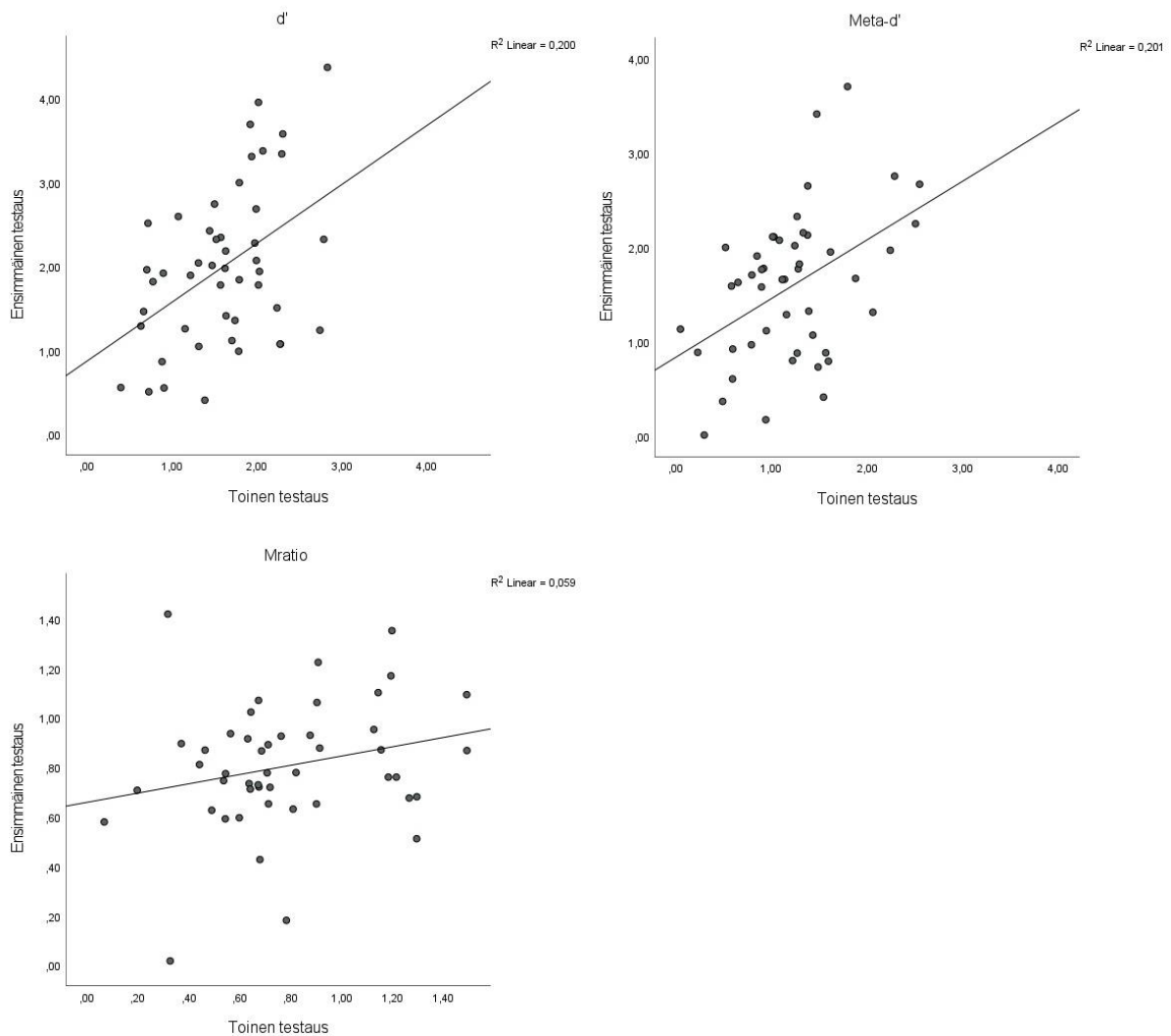
4.3 Test-retest-reliabiliteetti

4.3.1 Test-retest-reliabiliteetti Guggenmosin (2021) rajausten mukaan

Ensimmäisen ja toisen testauskerran välillä d' :tä tarkastellessa löytyi positiivinen lineaarinen yhteys ja tulos oli tilastollisesti merkitsevä ($r(47) = .45, p = .002$). Toistettujen mittausten t-testin perusteella d' :n arvot olivat suurempia ensimmäisellä ($m = 1.99, SD = 0.95$) kuin toisella testauskerralla ($m = 1.60, SD = 0.60$). Saatu tulos oli tilastollisesti merkitsevä keskikokoisella efektillä ($t(46) = 3.10, p = .003, d = 0.45$).

Meta- d' :n osalta havaittiin tilastollisesti merkitsevä positiivinen lineaarinen yhteys testauskertojen välillä ($r(47) = .45, p = .002$). Toistettujen mittausten t-testi viittaa meta- d' :n arvojen olleen suurempia ensimmäisellä ($m = 1.58, SD = 0.79$) kuin toisella testauskerralla ($m = 1.20, SD = 0.57$), ja tulos oli tilastollisesti merkitsevä keskikokoisella efektillä ($t(46) = 3.54, p < .001, d = 0.52$).

Myös Mratiota tarkastellessa havaittiin positiivinen lineaarinen yhteys testauskertojen välillä. Tulos ei kuitenkaan ollut tilastollisesti merkitsevä ($r(47) = 0.24, p = .101$). Toistettujen mittausten t-testi osoittaa Mration olleen keskimäärin suurempi ensimmäisellä ($m = 0.80, SD = 0.26$) kuin toisella testauskerralla ($m = 0.79, SD = 0.33$). Tulos ei ollut tilastollisesti merkitsevä ($t(46) = 0.35, p = .73, d = 0.05$). Kaikkien tunnuslukujen sirontakuviot on esitetty kuvassa 6.



Kuva 6. Sirontakuvi ensimmäisen ja toisen testauskerran d' :n, meta- d' :n ja $Mratio$ n arvoista, kun aineistosta poistettiin koehenkilöt, joiden $Mratio$ arvot olivat joko alle 0 tai yli 1,6 Guggenmosin (2021) suosituksiin perustuen.

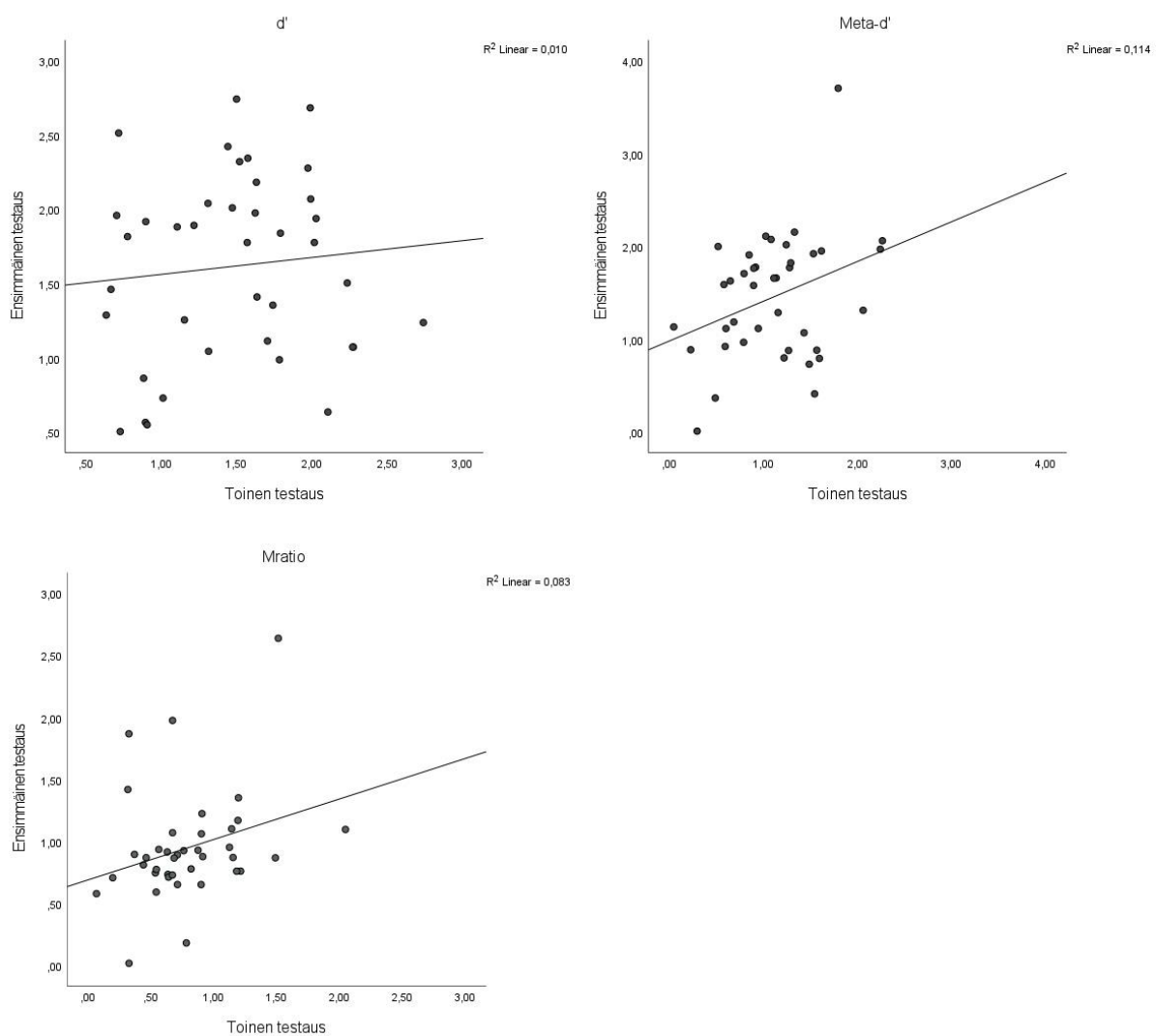
4.3.2 Test-retest-reliabiliteetti Leen ja työryhmän (2018) rajausten mukaan

Verratessa eroja ensimmäisen ja toisen testauskerran välillä Lee ym. (2018) rajausten perusteella, ei d' :n arvoissa ollut havaittavissa korrelaatiota ($r(39) = .10, p = .55$). Toistettujen mittausten t-testin perusteella ensimmäisen testauskerran arvot ($m = 1.61, SD = 0.62$) olivat suuremmat kuin toisella testauskerralla ($m = 1.48, SD = 0.54$). Tulos ei ollut tilastollisesti merkitsevä ja efekti oli heikko ($t(38) = 1.11, p = .28, d = 0.18$).

Meta- d' :tä tarkastellessa testauskertojen väliltä löytyi positiivinen lineaarinen yhteys ja tulos oli tilastollisesti merkitsevä ($r(39) = .34, p = .04$). Toistettujen mittausten t-testin mukaan meta-

d' oli suurempi ensimmäisellä ($m = 1.45$, $SD = 0.66$) kuin toisella testauskerralla ($m = 1.10$, $SD = 0.52$). Tulos oli tilastollisesti merkitsevä keskisuurella efektikoolla ($t(38) = 3.14$, $p = .003$, $d = 0.50$).

M ration osalta testauskertojen välillä oli positiivinen lineaarinen yhteys, mutta tulos ei ollut tilastollisesti merkitsevä ($r(39) = .29$, $p = .08$). Toistettujen mittausten t -testi osoitti arvojen olleen suurempia ensimmäisellä ($m = 0.95$, $SD = 0.45$) kuin toisella testauskerralla ($m = 0.79$, $SD = 0.40$). Vaikka pieni efekti oli havaittavissa, ei tulos ollut kuitenkaan tilastollisesti merkitsevä ($t(38) = 1.92$, $p = .06$, $d = 0.31$). Tunnuslukujen sirontakuviot on esitetty kuvassa 7.



Kuva 7. Sirontakuviot ensimmäisen ja toisen testauskerran d' :n, meta- d' :n ja M ration arvoista, kun aineistosta on poistettu ulkopuolelle jääneet arvot Leen ja kumppaneiden (2018) käyttämien rajausten mukaan.

5 Pohdinta

Tämän tutkimuksen tarkoituksena oli arvioida käyttämämme perturbaatioparadigmaan perustuvan menetelmän soveltuvuutta yksilön metakognitiivisen tehokkuuden tutkimiseen. Tarkastelun alla oli erityisesti menetelmän test-retest-reliabiliteetti, jota käsiteltiin sekä Guggenmosin (2021) että Leen ym. (2018) ehdottamien rajausten perusteella. Tähän tutkimusmenetelmään perustuen tarkoituksena oli myös selvittää, kuinka hyvin ihmiset pystyvät metakognitiivisesti arvioimaan virheitä kuulopalautteessa. Tarkastelussa olivat sensitiivisyyden tunnusluku d' , metakognitiivisen sensitiivisyyden tunnusluku $\text{meta-}d'$ sekä metakognitiivista tehokkuutta kuvaava M ratio. Hypoteeseina oli, että tunnuslukujen d' , $\text{meta-}d'$ sekä M ratio osalta kahden testauksen väliltä löytyy positiivinen yhteys, ja että metakognitiivinen suoriutuminen on keskimäärin alle optimaalisen tason, mutta yksilöllistä vaihtelua esiintyy. Positiivinen korrelaatio löytyi lähes kaikkien tunnuslukujen kohdalta, joskaan se ei ollut kaikissa tilanteissa tilastollisesti merkitsevä. Tämä vastasi osittain asetettua hypoteesia. Metakognitiivisessa suoriutumisessa oli näkyvissä vaihtelua, ja M ratio oli keskiarvoltaan alle yksi eli metakognitiivisen tehokkuuden optimaalinen taso alittui, mikä tuki asetettua hypoteesia.

5.1 Tulosten kuvaus ja arviointi

Kahden eri testauskerran tulosten vertailussa käy ilmi, että niin d' :n, $\text{meta-}d'$:n kuin M ration arvot olivat keskimäärin korkeampia ensimmäisellä kuin toisella testauskerralla. Lisäksi perturbaation kokoa sekä oikeiden vastausten määrää tarkasteltaessa selvisi, että perturbaation koko oli keskimäärin suurempi ja oikeinvastausprosentti korkeampi ensimmäisellä testauskerralla. Sillä, että koe on ollut eri tekijöiden takia helpompi ensimmäisellä testauskerralla, on ollut todennäköisesti vaikutusta myös metakognitiiviseen suoriutumiseen tehtävässä.

Koehenkilöiden metakognitiivisen arvion antaminen omasta suoriutumisesta oli keskimäärin heikompaa kuin heidän varsinainen kokeesta suoriutumisensa. Tämä tukee aikaisempia tutkimuksia, joissa $\text{meta-}d'$:n on havaittu olevan keskimäärin pienempi kuin d' (Fleming ym., 2014; Maniscalco & Lau, 2012; McCurdy ym., 2013). Ero suoriutumisissa johti siihen, että M ratio oli keskimäärin alle optimaalisen tason. Metakognitiivinen tehokkuus oli 82 % siitä, mikä olisi ollut odotettavissa perustuen tyypin 1 suoriutumiseen. Tehokkuus oli samaa luokkaa esimerkiksi Flemingin ym. (2014) ja Maniscalcon ja Laun (2012) visuaalisen havaitsemisen

tehtävistä saatujen tutkimustulosten kanssa, joissa vastaavat arvot olivat 77 % ja 88 %. On kuitenkin mielenkiintoista havaita, että yhteensä 15 koehenkilön kohdalla metakognitiivinen suoriutuminen ylitti optimaalisen tason. Käyttämämme tutkimusmenetelmä toi siis esiin yksilöiden välisiä eroja metakognitiivisen tehokkuuden osalta. Tulokset tukivat hypoteesiamme, jonka mukaan metakognitiivisen suoriutumisen tason odotettiin olevan keskimäärin alle optimaalisen.

Tarkasteltaessa tuloksia Guggenmosin (2021) ja Leen ja kumppaneiden (2018) rajausten valossa nähdään, että sekä kahden testauksen d' että meta- d' osoittautuivat test-retest-reliabiliteeteiltaan korkeammaksi Guggenmosin (2021) rajauksia noudatettaessa. Leen ja kumppaneiden (2018) rajausten mukaisissa tuloksissa d' :n osalta ei voida todeta, että koehenkilöt olisivat suoriutuneet tyypin 1 tehtävässä samankaltaisesti ensimmäisellä ja toisella testauskerralla. Tähän todennäköisesti vaikuttaa se, että aineiston rajausten takia d' :n arvoissa ei esiinny tarpeeksi vaihtelua riittävän reliabiliteetin saavuttamiseksi. Guggenmosin (2021) rajauksiin perustuen taas näyttäisi siltä, että testauskertojen tyypin 1 suoriutumiset vastasivat toisiaan keskinäisesti. Metakognitiivinen sensitiivisyys eli meta- d' puolestaan näyttäytyi olevan samansuuntaista testauskertojen välillä rajauksesta riippumatta.

Mration kohdalla tilanne on kuitenkin päinvastainen. Metakognitiivisen tehokkuuden test-retest-reliabiliteetti oli testauskertojen välillä suurempi Leen ja kumppaneiden (2018) rajauksilla. Ero rajausten välillä ei kuitenkaan ollut erityisen suuri. Vaikka positiiviset yhteydet olivat löydettävissä, jäivät ne odotettua matalammiksi eivätkä tulokset olleet tilastollisesti merkitseviä kummassakaan tapauksessa. Test-retest-reliabiliteettien tulokset tukevat pääosin asetettua hypoteesia siitä, että positiivinen korrelaatio on löydettävissä kaikkien kolmen tunnusluvun osalta.

Analysoinnissa käytimme tukena Guggenmosin (2021) sekä Leen ja kollegoiden (2018) käyttämiä rajauksia, mutta tulosten perusteella kumpikaan rajauksista ei toiminut optimaalisella tavalla. Tarkasteltaessa d' :n ja meta- d' :n arvoja ainoastaan test-retest-reliabiliteetin näkökulmasta voidaan todeta Guggenmosin (2021) suositusten toimineen paremmin. Toisaalta esimerkiksi Mration reliabiliteetti on voinut heiketä, jos osa analyysistä jätetyistä yli 1,6 ylittävistä arvoista ovat olleet todellisia Mration arvoja eikä esimerkiksi mittausvirheiden aiheuttamia.

Leen ja työryhmän (2018) rajauksia käyttämällä metakognitiivisen tehokkuuden test-retest-reliabiliteetti oli hieman korkeampi kuin Guggenmosilla (2021), mutta ei merkittävästi.

Kuitenkin Leen ja kumppaneiden (2018) suosituksia käyttämällä voidaan todeta Mration reliabiliteetin olleen parempi d' ja meta- d' arvojen kustannuksella, sillä rajausten perusteella pyrittiin vakioimaan tyypin 1 suoriutumista. Molemmissa suosituksissa on siis omat vahvuutensa ja heikkoutensa. Tyypin 1 suorituksen vakioiminen olisi voinut toimia test-retest-reliabiliteetin osalta paremmin, jos huomioon otettaisiin myös Guggenmosin (2021) tutkimuksen tulokset siltä osin, että tyypin 1 suoriutumisen taso määriteltäisiin alun perin tarpeeksi korkealle. Tällöin esimerkiksi puhtaiden arvausten osuus vähenisi. Leen ja kollegoiden (2018) oikeinvastausprosentin rajausten alaraja oli 60, joka on Guggenmosin (2021) tekemän analyysin perusteella liian matala antamaan luotettavia tuloksia Mration test-retest-reliabiliteetille.

Tutkimustulosten perusteella voidaan todeta, että käyttämämme koeasetelma ei tällaisenaan toimi metakognition mittaamisessa riittävän hyvin. Jos tuloksia verrataan puhtaasti psykologiassa käytettyihin $r \geq .90$ tai $r \geq .70$ suuruisiin tavoitearvoihin, voidaan todeta, että tulokset jäävät kauas tavoitteista. Toisaalta käyttämämme menetelmä on huomattavasti monimutkaisempi kuin esimerkiksi perinteiset kyselylomakkeet. Tarkastelun alla olleet aiheet eli virheen havaitseminen kuulopalautteessa sekä metakognitiivinen suoriutuminen kyseisessä tehtävässä aiheuttavat tutkimukselle omat haasteensa. Esimerkiksi Hunsley ja Mash (2008) kuvailevat test-retest-reliabiliteetin olevan monimutkainen ilmiö, johon vaikuttavat esimerkiksi niin tutkittavan aiheen ominaisuudet kuin testauskertojen välinen aika. Hypoteesin mukaisesti tämän tutkimuksen osalta $r \geq .70$ suuruinen tavoitearvo ei ollut tavoitettavissa.

On siis haasteellista määritellä, mikä olisi sopivan suuruinen tavoitearvo test-retest-reliabiliteetille metakognitiota mittaavassa tutkimuksessa. Vaikka esimerkiksi psykologisille testeille on kehitetty tiettyjä standardeja, ei ole kuitenkaan tarkkaan määritelty, mitä psykometrisiä arvoja mittarin tulisi saavuttaa aina tietyssä tarkoituksessa ja tietyllä otoksella (Hunsley & Mash, 2008). Esimerkiksi Mration osalta aikaisemman tutkimukset (Guggenmos, 2021; Rahnev ym., 2023) antavat viitteitä siitä, että Mration test-retest-reliabiliteetin voidaan odottaa jäävän heikoksi erityisesti silloin, kun toistojen määrä on matala. Mielenkiintoista on kuitenkin havaita, että 160 toistolle saamamme Mration korrelaatiot ($r = .24$, $r = .29$) olivat hyvin vastaavanlaisia Rahnevin (2023) julkaisussa 100 ($r = .24$) ja 200 ($r = .29$) toistolle saatujen Mration korrelaatioiden kanssa.

Metakognition mittaamisessa ja sen test-retest-reliabiliteetin määrittelyssä haasteellista on eri tasojen vaikutus lopullisiin tuloksiin esimerkiksi Mration osalta. Testauskertojen välisten

perturbaatioiden korrelaatiokertoimia tarkasteltaessa voidaan test-retest-reliabiliteetin todeta jääneen sen osalta vajaaksi tavoitearvosta. Toisin sanoen, jos test-retest-reliabiliteetti jäi matalaksi jo testin alkuvaiheessa eli jo vaikeustasoa määriteltäessä, heijastuu vaikutus mahdollisesti myös myöhempien vaiheiden ja tunnuslukujen test-retest-reliabiliteettiin. Näin ollen on monitulkintaista, toimiiko käyttämämme tutkimusmenetelmä metakognitiivisen tehokkuuden mittaamisessa – tekemällä esimerkiksi pieniä muutoksia alun kynnsarvon määrittelyyn voisi tähän tutkimuskysymykseen vastaaminen olla jo selkeämpää.

5.2 Menetelmän kehittämiskohteet

Aikaisemman tutkimustiedon pohjalta voidaan todeta, että toistojen määrä oli tutkimuksessamme pieni. Guggenmos (2021) toteaa näin käyvän usein metakognition tutkimuksessa, jolloin test-retest-reliabiliteetti jää väistämättä pieneksi, eikä vaadittua reliabiliteetin tasoa silloin saavuteta. Hän suosittelee siten metakognitiivisen tehokkuuden tutkimuksissa käytettävän vähintään 400 toistoa. Koeasetelmissamme yksi testauskerta sisälsi 160 toistoa, mikä on huomattavasti vähemmän kuin Guggenmosin (2021) suosituksessa. Voidaan siis todeta, että toistomäärää lisäämällä tutkimuksen reliabiliteetti olisi parempi. Toisaalta suuri toistomäärä asettaa haasteita käyttämämme koeasetelman osalta esimerkiksi siinä, että kokeen pitkä kesto lisää kokeen rasittavuutta. Testaukseen ei saisi kulua liikaa aikaa, sillä pitkät testaukset altistavat tarkkaavaisuuden herpaantumiselle, joka osaltaan vaikuttaa testin reliabiliteettiin (Aleci, 2021). Tutkimuksemme aikana useampi koehenkilö mainitsi jatkuvan /u/-äänteen toistamisen käyvän raskaaksi ja osalla oli vaikeuksia ylläpitää riittävää äänenvoimakkuutta koko tutkimuksen ajan.

Äänenvoimakkuutta kontrolloitiin ohjeistamalla koehenkilöitä käyttämään normaalia äänenvoimakkuutta. Koehenkilöiden arviot normaalista äänenvoimakkuudesta poikkesivat kuitenkin toisistaan kuulonvaraisesti arvioituna. Pyytämällä kaikkia koehenkilöitä ääntämään tietyllä ennalta määritetyllä voimakkuudella ja mittaamalla tämä desibelimitarilla olisi pystytty varmistamaan, että kaikki saavat myös kuulopalautteen samalla voimakkuudella. Samasta syystä myös etäisyys mikistä olisi pitänyt vakioida.

Guggenmos (2021) suosittelee artikkelissaan metakognitiivista suoriutumista tutkittaessa käytettävän noin 80 % oikeinvastausprosenttia tyypin 1 suoriutumisessa reliabiliteetin lisäämiseksi. Tutkimuksessamme päädyttiin tavoittelemaan 75 % oikeinvastausprosenttia, millä on mahdollisesti voinut olla vaikutus reliabiliteettiin. Kyseistä tavoitetasoa oli kuitenkin

käytetty Leen ja kumppaneiden (2018) kokeessa, jonka pohjalta myös rajasimme pois koehenkilöt, joiden oikeiden vastausten osuus oli joko alle 60 % tai yli 90 %.

QUEST-portaikkoa (Watson & Pelli, 1983) pidetään luotettavana menetelmänä laskea sopiva kynnsarvo. Koetilanteiden pohjalta kuitenkin havaittiin, että käyttämämme 30 toistoa ei ollut kaikissa tilanteissa riittävä tasaamaan perturbaatiota, jolloin perturbaation koko jäi osalle koehenkilöistä joko liian haastavaksi tai liian helpoksi. Tähän on voinut entisestään vaikuttaa se, kuinka suuri ero harjoituksen ja portaikon perturbaation välillä oli. Harjoituksessa perturbaatiot olivat pienimmillään 60 senttiä, kun taas portaikkoharjoitus alkoi 28,5 sentistä. Kontrasti on ollut suuri, minkä johdosta useat koehenkilöt kommentoivat hämmentyneensä vaikeustason muutoksesta. Pakkovalinnan tehtävissä kynnsarvo määräytyy helposti kauas todellisesta kynnsarvosta, jos tapahtuu sarja onnekkaita tai epäonnekkaita arvauksia (Alec, 2021). Jos heti portaikon alussa on tullut väärää vastauksia, perturbaatio on mahdollisesti hypännyt jo alussa hyvin korkealle, minkä jälkeen portaikon tasaantuminen vie entistä enemmän aikaa.

Erot koekertojen välisessä vaikeustasossa ovat voineet vaikuttaa oikeiden vastausten määrään ja vastausvarmuuteen sekä yhä edelleen laskettujen tunnuslukujen arvoihin. Näin ollen myös käyttämämme Guggenmosin (2021) ja Leen ja kumppaneiden (2018) rajausten ulkopuolelle jäi huomattava määrä koehenkilöitä, mikä pienensi yhä ennestään otoskokoa. Lisäksi perturbaation koko oli keskimäärin suurempi ensimmäisellä kuin toisella tutkimuskerralla, mikä voi osittain selittyä oppimisvaikutuksella. Koeasetelman ollessa jo tuttu on vaikeustaso voinut määrittyä suhteessa hankalammaksi toisella testauskerralla. Kuitenkin löytyi myös tuloksia, joissa toisen kerran perturbaatio oli suurempi kuin ensimmäisen.

Alecin (2021) mukaan aistijärjestelmän sensitiivisyyden heilahtelu, tarkkaavuuden herpaantuminen, muutokset subjektiivisissa kriteereissä sekä mahdollinen oppimis- tai väsymysefekti aiheuttavat tietyissä määrin vaihtelua yksilön tehtävässä suoriutumiseen. Koeasetelmamme vaati aktiivista keskittymistä ja tarkkaavuuden ylläpitoa, joten esimerkiksi testauksen ajankohdan kontrollointi olisi voinut osaltaan vaikuttaa reliabiliteettiin. Koska koehenkilöille annettiin mahdollisuus varata testikerta itsenäisesti, on koehenkilö saattanut suorittaa toisen käynnin esimerkiksi aikaisin aamulla ja toisen myöhemmin iltapäivällä. Koehenkilöiden kertoman palautteen perusteella vireystiloissa on voinut esiintyä huomattavia eroavaisuuksia kahden eri testauskerran välillä. Osa koehenkilöistä esimerkiksi kertoi väsymyksestä tai keskittymisen herpaantumisesta pitkän päivän päätteeksi. Lisäksi

motivaatiolla on voinut olla vaikutusta kokeessa suoriutumiseen. Toisella testauskerralla koehenkilöt ovat tienneet, mitä on tulossa ja esimerkiksi kokemus yksitoikkoisesta kokeesta on voinut osaltaan vaikuttaa suoriutumiseen.

Kaikki koehenkilöt eivät hyödyntäneet vastausvarmuutta annettaessa koko vastausvaihtoehtojen skaalaa. Alkuperäisessä aineistossa kuusi koehenkilöä jätti kokonaan käyttämättä varma A ja varma B vaihtoehtoja. Näiden lisäksi yksi koehenkilö vastasi ainoastaan käyttäen epävarma A ja epävarma B vaihtoehtoja johtaen meta-d':n ja Mration negatiivisiin arvoihin molempien testausten osalta. Kyseessä on kuitenkin voinut olla koehenkilön aito arvio omasta suoriutumisesta ja epävarmuuteen on voinut esimerkiksi vaikuttaa se, että koe on voinut jäädä vaikeustasoltaan liian haastavaksi. Annetut vastausvaihtoehdot vaikuttivat kuitenkin yleisesti ottaen selkeiltä ja tutkittavat oivalsivat hyvin nopeasti käytännön antaa vastausvaihtoehto ja vastausvarmuus samalla painikkeella.

5.3 Tutkimuksen merkitys ja jatkotutkimusehdotukset

Tämä pro gradu -tutkielma arvioi perturbaatioparadigman luotettavuutta tutkittaessa metakognitiivista kykyä havaita muutos kuulopalautteessa. Lisäksi tarkasteltiin sitä, kuinka hyvin ihmiset pystyvät metakognitiivisesti arvioimaan virheitä kuulopalautteessa. Tutkimuksen tulokset viittaavat siihen, että perturbaatioparadigma toimi kahden testauskerran välillä osittain samankaltaisesti meta-d':n ja Mration osalta. Test-retest-reliabiliteetti ei hypoteesin mukaisesti tuottanut lähelle $r \geq .70$ korrelaatioita, mutta positiivinen yhteys oli testikertojen välillä havaittavissa. Vaikka reliabiliteetit jäivät heikoiksi, olivat tulokset samankaltaisia aikaisempien test-retest-reliabiliteettia tarkastelevien tutkimustulosten kanssa.

Metakognitiivisen arvion osalta koehenkilöiden havaittiin arvioivan virheitä kuulopalautteessa suhteellisen hyvin, mutta ei kuitenkaan optimaalisesti. Otokseen mahtui koehenkilöitä, joiden metakognitiivinen tehokkuus oli huomattavasti alle optimaalisen tason, sekä koehenkilöitä, jotka ylittivät tason. Tämä vahvistaa ajatusta siitä, että metakognitiivinen tehokkuus vaihtelee tehtävässä yksilöittäin, jolloin aihetta ja sen vaikutuksia on mielekästä sekä tarpeellista tutkia enemmän. Maniscalcon ja Laun (2014) mukaan meta-d' on tunnuslukuna kuvaileva, mutta ei kuitenkaan selittävä. Toisin sanoen meta-d' ei itsessään selitä, mitä metakognitiivisen sensitiivisyyden vaihtelun taustalla on. Mielenkiintoista olisi siis tutkia, mikä yksilöllisiä eroja metakognitiivisessä suoriutumisessa aiheuttaa. Lisäksi voisi olla kiinnostavaa tutkia, miten pitkä tai lyhyt aikaväli tutkimuskertojen välissä vaikuttaa metakognitiiviseen suoriutumiseen.

Ylipäänsä metakognition tutkimus rajoittuu kielen ja puheen osalta pääosin kaksikielisyyteen, toisen kielen oppimiseen sekä metakognition osuuteen lasten puheterapiassa (Folke ym., 2016; Gaile & Adams, 2018; Kartushina ym., 2023; Ordin ym., 2020). Metakognition osuutta kuulopalautteen arvioimisessa ei ole aikaisemmin tutkittu eikä esimerkiksi tiedetä, kuinka hyvin ihmiset pystyvät metakognitiivisesti arvioimaan omaa ääntämistään tai puhettaan. Kuitenkin metakognitiolla on suuri merkitys oman puheen säätelyssä: jos puhe on epäselvää tai virheellistä, pyrkii ihminen tavallisesti korjaamaan omaa puheen tuottamistaan. Tutkimuksessamme havaittu metakognitiivinen vaihtelu kuulopalautteen arvioinnissa herättää kysymyksiä siitä, voisiko havainnolla olla myös kliinistä merkitystä puheterapian kannalta.

Kliinisten havaintojen mukaan Parkinsonin tautia sairastavat ihmiset eivät aina tunnista oirekuvaan kuuluvaa oman äänen voimakkuuden heikentymistä, vaan mieltävät puhuvansa normaalilla äänen voimakkuudella (Clark ym., 2014; Fox ym., 2002; Ho ym., 2000). Esimerkiksi tällainen mahdollinen vinouma oman äänen havaitsemisessa voisi vaikuttaa myös häiriön hoitoon. Jos koeasetelmalla pystytään luotettavasti tutkimaan terveiden ihmisten metakognitiota ääntämisvirheiden kohdalla, voisi tutkimusta laajentaa koskemaan myös tiettyjä puheterapian asiakasryhmiä.

Jos koeasetelma toimisi toivotulla tavalla, se voisi auttaa ymmärtämään metakognition roolia oman puheen tuottamisen ja monitoroinnin osalta kattavammin. Tutkimukseemme perustuen voimme myös todeta yksilöllä olevan vaihtelua metakognitiivisessa tehokkuudessa. Näiden havaintojen laajemman ymmärtämisen kautta voisi mahdollisesti avautua uusia näkökulmia ja keinoja myös puheterapian kliinisen työn saralle. Voidaan pohtia, voisiko yksilöllisten metakognitiivisten taitojen tunnistaminen ja tarvittaessa niiden vahvistaminen edesauttaa yksilön kuntoutumista. Jos yksilöllä on esimerkiksi heikot metakognitiiviset taidot, kuntoutusta voisi mahdollisesti lähestyä tästä näkökulmasta. Flemingin (2024) ajatusta aistijärjestelmän kyvystä havaita virhe voidaan soveltaa myös meidän aiheeseemme: jos ihminen ei kykene tunnistamaan virhettä omassa puheessaan, miten hän pystyy korjaamaan sen?

Lähteet

- Aleci, C. (2021). *Measuring the Soul: Psychophysics for Non-Psychophysicists* (1. painos). Les Ulis: EDP Sciences. doi:10.1051/978-2-7598-2518-9
- Baird, B., Cieslak, M., Smallwood, J., Grafton, S. T., and Schooler, J. W. (2015). Regional white matter variation associated with domain-specific metacognitive accuracy. *J. Cogn. Neurosci.* 27, 440–452. doi:10.1162/jocn_a_00741
- Baird, B., Smallwood, J., Gorgolewski, K. J., & Margulies, D. S. (2013). Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. *The Journal of Neuroscience*, 33(42), 16657–16665. doi:10.1523/JNEUROSCI.0786-13.2013
- Barrett, A. B., Dienes, Z., & Seth, A. K. (2013). Measures of metacognition on Signal-Detection Theoretic models. *Psychological Methods*, 18(4), 535–552. doi:10.1037/a0033268v
- Bauer, J. J., Mittal J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: an automatic mechanism for stabilizing voice amplitude. *The Journal of the Acoustical Society of America*, 119(4), 2363-71. doi:10.1121/1.2173513
- Behroozmand, R., Karvelis, L., Liu, H., & Larson, C. R. (2009). Vocalization-induced enhancement of the auditory cortex responsiveness during voice F 0 feedback perturbation. *Clinical Neurophysiology*, 120(7), 1303–1312. doi:10.1016/j.clinph.2009.04.022
- Behroozmand, R., Liu, H., & Larson, C. R. (2011). Time-dependent neural processing of auditory feedback during voice pitch error detection. *Journal of Cognitive Neuroscience*, 23(5), 1205–1217. doi:10.1162/jocn.2010.21447
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436. doi:10.1163/156856897X00357
- Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, 103(6), 3153–3161. doi:10.1121/1.423073
- Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H. & Perkell, J. S. (2008). A system for online dynamic perturbation of formant frequencies and results from perturbation of the mandarin triphthong /iau/. *Proceedings of the 8th International Seminar on Speech Production*, 65–68.

- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Newbury Park, California: SAGE.
- Chen SH., Liu H., Xu Y. & Larson CR. (2006). Voice F0 responses to pitch-shifted voice feedback during English speech. *The Journal of the Acoustical Society of America*, 2007(121), 1157–63.
- Clark, J. P., Adams, S. G., Dykstra, A. D., Moodie, S., & Jog, M. (2014). Loudness perception and speech intensity control in Parkinson’s disease. *Journal of Communication Disorders*, 51, 1–12. doi:10.1016/j.jcomdis.2014.08.001
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *The American Psychologist*, 34(10), 906–911. doi:10.1037/0003-066X.34.10.906
- Fleming, S. M. (2017). HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness*, 2017(1). doi:10.1093/nc/nix007
- Fleming, S. M. (2024). Metacognition and confidence: A review and synthesis. *Annual Review of Psychology*, 75(1). doi:10.1146/annurev-psych-022423-032425
- Fleming, S. M. & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8, 443–443. doi:10.3389/fnhum.2014.00443
- Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain (London, England : 1878)*, 137(10), 2811–2822. doi:10.1093/brain/awu221
- Folke, T., Ouzia, J., Bright, P., De Martino, B., & Filippi, R. (2016). A bilingual disadvantage in metacognitive processing. *Cognition*, 150, 119–132. doi:10.1016/j.cognition.2016.02.008
- Fox, C., Morrison, C. E., Ramig, L. O., & Sapir, S. (2002). Current perspectives on the Lee Silverman voice treatment (LSVT) for individuals with idiopathic Parkinson disease. *American Journal of Speech-Language Pathology*, 11(2), 111–123. doi:10.1044/1058-0360(2002/012)
- Franken, M. K., Eisner, F., Acheson, D. J., McQueen, J. M., Hagoort, P., & Schoffelen, J.-M. (2018). Self-monitoring in the cerebral cortex: Neural responses to small pitch shifts in auditory feedback during speech production. *NeuroImage*, 179, 326–336. doi:10.1016/j.neuroimage.2018.06.061
- Gaile, J., & Adams, C. (2018). Metacognition in speech and language therapy for children with social (pragmatic) communication disorders: implications for a theory of therapy.

- International Journal of Language & Communication Disorders*, 53(1), 55–69.
doi:10.1111/1460-6984.12326
- Gauvin, H. S., De Baene, W., Brass, M., & Hartsuiker, R. J. (2016). Conflict monitoring in speech processing: An fMRI study of error detection in speech production and perception. *NeuroImage*, 126, 96–105. doi:10.1016/j.neuroimage.2015.11.037
- Guenther, F. H. (2016). Neural control of speech. *The MIT Press*.
- Guggenmos, M. (2021). Measuring metacognitive performance: type 1 performance dependence and test-retest reliability. *Neuroscience of Consciousness*, 2021(1). doi:10.1093/nc/niab040
- Hafke, H. Z. (2008). Nonconscious control of fundamental voice frequency. *The Journal of the Acoustical Society of America*, 123(1), 273–278. doi:10.1121/1.2817357
- Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., & Kenney, M. K. (2000). Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Exp. Brain Res.* 130, 133–141. doi: 10.1007/s002219900237
- Ho, A. K., Bradshaw, J. L., & Iannakou, R. (2000). Volume perception in Parkinsonian speech. *Movement Disorders*, 15(6), 1125–1131. doi:10.1002/1531-8257(200011)15:6<1125::AID-MDS1010>3.0.CO;2-R
- Hunsley, J., & Mash, E. J. (2008). *A Guide to Assessments that Work*. Cary: Oxford University Press. doi:10.1093/med:psych/9780195310641.001.0001
- Kartushina, N., Soto, D., & Martin, C. (2023). Metacognition in Second Language Speech Perception and Production. *Language Learning*, 73(2), 508–542. doi:10.1111/lang.12549
- Kleiner, M., Brainard, D. & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, 36(14), 1–16.
- Kopčanová, M., Ince, R. A. A., & Benwell, C. S. Y. (2023) Two distinct stimulus-locked EEG signatures reliably encode domain-general confidence during decision formation (Preprint). bioRxiv. doi:10.1101/2023.04.21.537831
- Koriat, A. (2007). Metacognition and consciousness. In *The Cambridge Handbook of Consciousness*. Cambridge University Press. doi:10.1017/CBO9780511816789.012
- Korzyukov, O., Karvelis, L., Behroozmand, R., & Larson, C. R. (2012). ERP correlates of auditory processing during automatic correction of unexpected perturbations in voice auditory feedback. *International Journal of Psychophysiology*, 83(1), 71–78. doi:10.1016/j.ijpsycho.2011.10.006

- Larson, C. R., Sun, J., & Hain, T. C. (2007). Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control. *The Journal of the Acoustical Society of America*, *121*(5), 2862-2872. doi:10.1121/1.2715657
- Law, J., Campbell, C., Roulstone, S., Adams, C., & Boyle, J. (2008). Mapping practice onto theory: the speech and language practitioner's construction of receptive language impairment. *International Journal of Language & Communication Disorders*, *43*(3), 245–263. doi:10.1080/13682820701489717
- Lee, A. L. F., Ruby, E., Giles, N., & Lau, H. (2018). Cross-domain association in metacognitive efficiency depends on first-order task types. *Frontiers in Psychology*, *9*, 2464–2464. doi:10.3389/fpsyg.2018.02464
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Mahwah: Psychology Press. doi:10.4324/9781410611147
- Manfra, L., Tyler, S. L., & Winsler, A. (2016). Speech monitoring and repairs in preschool children's social and private speech. *Early Childhood Research Quarterly*, *37*, 94–105. doi.org/10.1016/j.ecresq.2016.04.004
- Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430. doi:10.1016/j.concog.2011.09.021
- Maniscalco, B., & Lau, H. (2011). On a distinction between detection and discrimination: metacognitive advantage for signal over noise. *Journal of Vision*, *11*(11), 163–163. doi:10.1167/11.11.163
- Maniscalco, B., & Lau, H. (2014). Signal detection theory analysis of type 1 and type 2 data: meta-d, response-specific meta-d, and the unequal variance SDT Model. In S. M. Fleming & C. D. Frith (Eds.), *The cognitive neuroscience of metacognition* (25-66). New York, NY: Springer. doi:10.1007/978-3-642-45190-4_3
- McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. *The Journal of Neuroscience*, *33*(5), 1897–1906. doi:10.1523/JNEUROSCI.1890-12.2013
- Ordin, M., Polyanskaya, L., & Soto, D. (2020). Metacognitive processing in language learning tasks is affected by bilingualism. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *46*(3), 529–538. doi:10.1037/xlm0000739

- Parkinson, A. L., Korzyukov, O., Larson, C. R., Litvak, V., & Robin, D. A. (2013). Modulation of effective connectivity during vocalization with perturbed auditory feedback. *Neuropsychologia*, *51*(8), 1471–1480. doi:10.1016/j.neuropsychologia.2013.05.002
- Pedone, R., Semerari, A., Riccardi, I., Procacci, M., Nicolò, G., and Carcione, A. (2017). Development of a self-report measure of metacognition: The Metacognition Self-Assessment Scale (MSAS). Instrument description and factor structure. *Clinical Neuropsychiatry*, *14*, 185–194. doi:10.1037/t82174-000
- Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition*, *77*(2), 97–132. doi.org/10.1016/S0010-0277(00)00090-1
- Rahnev, D. (2023). Measuring metacognition: A comprehensive assessment of current methods (Preprint). PsyArXiv. doi:10.31234/osf.io/waz9h
- Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience*, *1*(3), 165–175. doi:10.1080/17588921003632529
- Sadeghi, S., Ekhtiari, H., Bahrami, B., & Ahmadabadi, M. N. (2017). Metacognitive deficiency in a perceptual but not a memory task in methadone maintenance patients. *Scientific Reports*, *7*(1), 7052–7052. doi:10.1038/s41598-017-06707-w
- Salmela, V., Kurki, I., & Laurinen, P. (2006). Psykofysiikka. Teoksessa H. Hämäläinen, M. Laine, O. Aaltonen, & A. Revonsuo (toim.), *Mieli ja aivot: kognitiivisen neurotieteen oppikirja*. (s. 33–44). Turun yliopisto, kognitiivisen neurotieteen tutkimuskeskus
- Schenck, A., Hilger, A. I., Levant, S., Kim, J. H., Lester-Smith, R. A., & Larson, C. (2023). The effect of pitch and loudness auditory feedback perturbations on vocal quality during sustained phonation. *Journal of Voice*, *37*(1), 37–47. doi:10.1016/j.jvoice.2020.11.001
- Sellbom, M., & Suhr, J. A. (2019). *The Cambridge handbook of clinical assessment and diagnosis*. Cambridge: Cambridge University Press. doi:10.1017/9781108235433
- Sharwood Smith, M. (2017). *Introducing language and cognition: a map of the mind*. Cambridge: Cambridge University Press.
- Vlassova, A., Donkin, C., & Pearson, J. (2014). Unconscious information changes decision accuracy but not confidence. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *111*(45), 16214–16218. doi:10.1073/pnas.1403619111
- Watson, A. B. & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120. doi:10.3758/BF03202828
- Wickens, T. D. (2001). *Elementary Signal Detection Theory*. New York: Oxford University Press. doi:10.1093/acprof:oso/9780195092509.001.0001