

KOHINAT, HÄLINÄT JA KAIUNNAT - HÄLYPUHETESTIEN VANHAT JA UUDET PEITTOÄÄNET

Antti Kuusinen¹ ja Tapio Lokki¹

Aalto-yliopiston elektroniikan ja sähkötekniikan korkeakoulu
Signaalinkäsittelyn ja akustiikan laitos
Otakaari 5, 02150 Espoo
¹antti.kuusinen@aalto.fi

Tiivistelmä

Hälypuhetestit, kuten kansainvälinen lausematriisitestit, mittaavat ihmisen kykyä ymmärtää puhetta hälyssä. Tavanomaisesti testissä käytetään peittoäänenä tasaista kohinaa, jonka keskimääräinen taajuusrakenne vastaa kuunneltavan puheen spektriä. Spektrisovitetun kohinan lisäksi peiteääninä voidaan käyttää myös erilaisia hälinöitä, ja virtuaaliakustiikan avulla hälypuheteihin voidaan tuoda monimuotoisuutta esimerkiksi lisäämällä testiäänin kaihuntaa. Tässä artikkelissa käymme läpi yleisimpiä peiteäänityyppejä ja esitämme myös uuden tavan muodostaa hälypuhetta käyttäen hyväksi jo olemassa olevia lausematriisitestin puheääniä ja Fourier-muunnoksen perusominaisuuksia. Esitämme myös kuuntelukoetuloksia kaiunna vaikutuksesta hälypuhetein tuloksiin eri peiteäänityypeillä normaalikuuloisilla kuuntelijoilla.

1 JOHDANTO

Hälypuheteit ovat tärkeä osa nykyistä kuulontutkimusta ja kuulo-ongelmien diagnostiikkaa [1, 2]. Hälypuheteissä kuuntelijalle esitetään kohdepuhetta jonkin peiteäänin kanssa. Muuttamalla adaptiivisesti puheen ja peiteäänin signaali-kohinasuhdetta voidaan mitata esimerkiksi 50 prosentin puheentunnistuksen kynnystaso. Peiteäänin ominaisuuksilla on oleellinen merkitys kokeen tuloksiin, kynnystasoihin ja luotettavuuteen. Vaikka puheen ja kuulon tutkimuksissa käytetään monenlaisia peiteääniä ja tutkimuksissa niitä räätälöidään tutkimuskysymyksen perusteella, ei kliinisessä diagnostiikassa peiteääninä pääsääntöisesti käytetä muuta kuin tasaista puheen spektriin sovitettua kohinaa. Mahdollisia peiteääniä on kuitenkin monia ja tässä artikkelissa tarkastellaan lyhyesti erilaisia kohinoita ja hälinöitä sekä kaiunna vaikutusta testituloksiin. Lisäksi esitämme myös yhden uuden tavan muodostaa hälypuhetta jo olemassa olevista suomalaisen lausematriisitestin testimateriaaleista [3].

2 KOHINAT

Hälypuhetein tyypillisin peiteääni on tasainen kohina, jonka spektri vastaa kuunneltavan puheen pitkäaikaisspektriä. Tällainen kohina voidaan tuottaa esim. summaamalla

Copyright ©2021 Antti Kuusinen, Tapio Lokki. Tämä on avoimesti julkaistu teos, joka noudattaa Creative Commons BY 4.0 Kansainvälinen –lisenssiä (CC BY 4.0). Teosta saa kopioida, levittää, näyttää ja esittää julkisesti ja siitä saa luoda johdannaisteoksia, kunhan tekijän nimi ja lähde mainitaan asianmukaisesti.

ja satunnaisesti limittämällä suuri määrä kuunneltavia puhesignaaleja. Vaihtoehtoisesti spektrisovitettu kohina voidaan tuottaa käyttäen hyväksi Fourier-muunnosta ja satunnais-tamalla puhesignaalin vaiheet taajuutasossa ja kääntämällä vaihesatunnaistettu signaali takaisin aika-tasoon [4]. Eri kielten pitkäaikaisspektrien ollessa hyvin samankaltaisia, on ehdotettu käytettäväksi myös ”universaalia” spektrisovitettua kohinaa [5], joka voisi helpottaa esim. kuulokojeiden sovitusta sekä tulosten vertailua eri kielillä. Jossain tilanteissa voidaan peiteääninä käyttää myös valkoista tai pinkkiä kohinaa, mutta hälypuhetesteissä näitä käytetään nykyään hyvin harvoin.

Spektrisovitetun kohinan etuja ovat sen ennustettavuus, johdonmukaisuus ja toistettavuus. Psykometriset kynnyskäyrät signaali-kohinasuhteen muuttuessa ovat jyrkkiä, jolloin hälypuhetestin tuloksena saadaan tarkat estimaatit puheenymmärrettävyyden kynnystasoista [6]. Spektrisovitetun kohinan energettisen peiteprofiilin voidaan ajatella olevan maksimaalinen suhteessa kuunneltavaan puheeseen, eikä kohinassa ole myöskään ajallisia muutoksia. Yleisesti ajatellaan, että ajallisesti tasainen kohina aiheuttaa vain ns. energettistä maskausta, mutta tasaisen kohinan peittoefekti ei itseasiassa olekaan niin yksinkertainen [7, 8]. Jopa tasaisessa kohinassa on mukana (satunnaisia) modulaatioita, jotka aiheuttavat jonkin verran myös ns. modulaatiomaskausta [7, 8]. Tyypillisemmin modulaatiomaskausta tapahtuu ajallisesti vaihtelevien peiteäänien vaikutuksesta, jolloin puheen ymmärrettävyyden kannalta tärkeät modulaatiokomponentit peittyvät tai sotkeutuvat peiteäänien modulaatioiden kanssa. Seuraavassa kappaleessa käsitellään ajallisesti vaihtelevia peiteääniä, joita tässä artikkelissa kutsutaan termillä ”hälinät”.

3 HÄLINÄT

Hälinänä voidaan pitää peiteääntä, joka enemmän tai vähemmän vaihtelee ajallisesti. Tutkimuskirjallisuudesta löytyy paljon erilaisia peiteäänihälinöitä, jotka usein räätälöidään tutkimuskysymyksen perusteella aiheuttamaan tietynlaisia peittovaikutuksia. Esimerkiksi amplitudimoduloidulla kohinalla voidaan kartoittaa kuulijan kykyä käyttää hyväksi hiljaisempia kohtia peittoäänessä. Verrattuna tasaisiin peittoääniin, amplitudimodulaatio helpottaa puheen ymmärrettävyyttä normaalikuuloisilla henkilöillä, mutta kuulon alenema tai ikäkuulo voi merkittävästi vähentää tätä hyötyä [9]. Varsinkin ikäkuulo voi vähentää kuulon herkkyyttä äänien ajallisille vaihteluille.

Amplitudimodulaatio voidaan toteuttaa monella tapaa, mm. sinifunktioiden avulla. Tyypillinen tapa on kuitenkin käyttää hyväksi jonkin puhesignaalin verhoikäyrää, jolloin modulaatio mukailee oikean puheen ajallista vaihtelua. Tällaisia hälinäpeiteääniä ovat muun muassa ”ICRA”-hälinät [10], jotka kuulostavat hieman puheelta, mutta ovat käytännössä kuitenkin täysin käsittämättömiä. Vaikka ICRA-äänien ominaisuuksiltaan ovat samankaltaisia oikean puheen kanssa, niistä puuttuu puheen tonaaliset ja harmoniset rakenteet, jotka määrittävät erityisesti vokaaleja. ICRA-hälinöitä on muutamaa eri tyyppiä, joissa puhujien määrä ja sukupuoli vaihtelee signaalien välillä.

Kansainvälinen hälypuhesignaali (eng. *The international speech test signal*, ISTS) [11] on myös kiinnostava hälypuhepeiteääni. Se on muodostettu segmentoimalla ja yhdistämällä puhetta kuudesta eri kielestä (englanti, arabia, kiina, ranska, saksa, espanja). ISTS on täysin puheen kaltaista, mutta käytännössä täyttä siansaksaa.

Ehkä yksinkertaisin ja yleisin tapa muodostaa hälypuhetta (eng. *babble talk*) on summata limitäin N -kappaletta puhesignaaleja toistensa kanssa. Lisäämällä puhujien lukumäärää, hälypuheesta tulee tiheämpää ja lopulta tasaista kohinaa. Hälypuhenäytteiden voidaan ajatella muodostavan hälypuhejatkumon yhdestä puhujasta äärettömään moneen puhujaan. Hälypuheen ollessa ymmärrettävää, esim. yhden tai vain muutaman puhujan tapauksessa, puhutaan usein ns. informatiivisesta maskauksesta, jossa peiteäänen informatiivinen luonne vaikeuttaa kuulijan tarkkaavaisuutta ja keskittymistä kohdepuheeseen. Tutkimukset ovat todenneet, että hälypuhe on häiritsevintä, kun siinä on noin neljästä kuuteen puhujaa [12, 13]. Usein jokapäiväiset taustahälinät tuottavat enemmän tai vähemmän sekä energettistä että informatiivista maskausta ja usein niitä on vaikea erottaa toisistaan. Informatiivista maskausta voi tapahtua myös muilla kuin puheäänillä [14].

Seuraavassa kappaleessa esitetään yksi vaihtoehtoinen tapa muodostaa hälypuhetta käyttäen hyväksi Fourier-muunnoksen perusominaisuuksia.

3.1 Hälypuheen tuottaminen vaiheviiveen vaihtelulla

Vaiheviiveen vaihtelulla (eng. phase jitter) voidaan hajottaa signaalin aikatason rakenne siten, että signaalin magnitudivaste pysyy koskemattomana. Fourier-muunnoksen aikatason siirto-ominaisuuden mukaan signaalin siirtäminen aikatasossa vastaa vaiheiden yhteenlaskua taajuustasossa [15]:

$$\mathcal{F}\{x(n-k)\} = X(\omega)e^{-j\omega k} \quad (1)$$

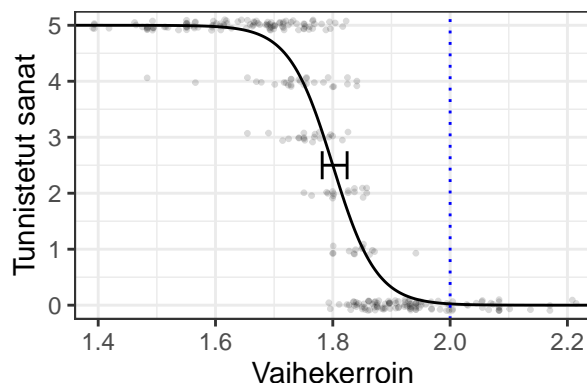
$$= |X(\omega)|e^{j(\angle X(\omega) - \omega k)} \quad (2)$$

Jos parametri k muuttuu eri taajuuskomponenttien välillä, siirtyvät komponentit myös ajassa eri tavoin, jolloin signaali vääristyy riippuen k arvoista. Käyttämällä täysin satunnaisia k :n arvoja eli toisin sanoen satunnaistamalla signaalin vaiheet, voidaankin tuottaa vaikka loppumattomia kohinasignaaleja [4]. Tämänlaista vaiheviiveen vaihtelua on jo aikaisemmin käytetty puheen ymmärtämisen tutkimuksessa monella eri tavoin [16, 17], mutta pääasiassa prosessointi on tehty lyhyissä aikaikkunoissa (eng. short-time FT, STFT). Tekemällä vaiheviiveen vaihtelua lyhyissä aikaikkunoissa voidaan esimerkiksi tuottaa signaaleja, joilla on sama verhoikäyrä kuin alkuperäisellä puhesignaalilla, mutta joiden ajallinen hienorakenne (eng. temporal fine structure) on muuttunut. On selvää, että puheen ymmärrettävyyden kannalta vaiheinformaatiolla on olennainen merkitys [18].

Tässä työssä ehdotamme, että vaiheviiveen vaihtelua / aikatason siirto-ominaisuutta voidaan käyttää hyväksi myös pitkäaikaisen Fourier-muunnoksen kautta, jolloin taajuustason vaiheinformaatio vastaa koko käsiteltävää signaalia. Vaiheiden yhteenlaskun tai satunnaistamisen sijasta hälypuhetta voidaan tuottaa kertomalla puhesignaalin vaiheet kertoimella c :

$$y(n) = \mathcal{F}^{-1}\{|X(\omega)|e^{j(c\angle X(\omega))}\} \quad (3)$$

Kertolasku tekee vaiheviiveen vaihtelusta riippuvaisen alkuperäisen signaalin vaiheinformaatiosta, jolloin riippuen kertoimen c arvosta ja alkuperäisten vaiheiden suhteista, enemmän samassa vaiheessa olevat taajuuskomponentit siirtyvät ajassa enemmän samankaltaisesti kuin eri vaiheissa olevat komponentit. Diskreetin Fourier-muunnoksen epävarmuusperiaatteen mukaisesti muunnoksen taajuusresoluutio riippuu signaalin pituudesta,



Kuva 1: Vaihekerrotun puheen ymmärrettävyys kertoimen c eri arvoilla. Psykometrisen funktion 50 prosentin kynnsarvo: 1.8; kulmakerroin ($/0.1$): -0.77. Pystysuora viiva merkitsee kertoimen arvoa kaksi, koska sitä käytettiin tuottamaan hälypuhetta hälypuheteestiin, jonka tuloksia on esitetty kuvassa 3.

jolloin oleellinen vaatimus vaihekerrotun hälypuheen tuottamiselle on, että prosessoitava puhesignaali on riittävän pitkä. Tällöin vaiheviiveen vaihtelu toimii tarpeeksi tiheällä taajuusresoluutiolla ja lopputulos kuulostaa hälypuheelta.

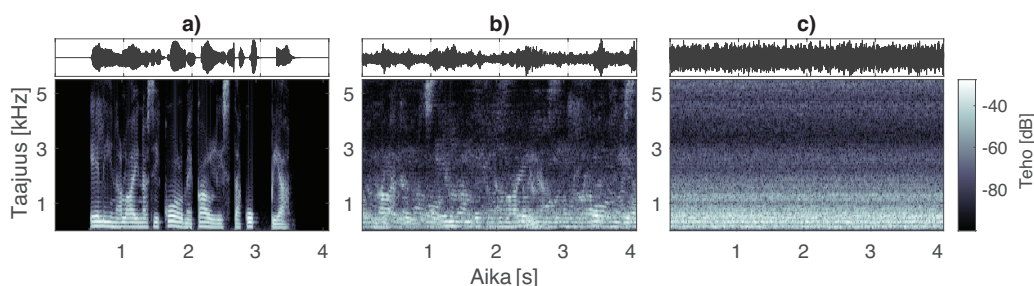
Kun prosessia sovelletaan suomalaisessa lausematriisitestissä [2], riittävän pitkä puhesignaali saadaan aikaiseksi esimerkiksi siten, että poimitaan lausemateriaalista satunnaisesti kolme lausetta ja liitetään ne peräkkäin ennen vaihekerrotun prosessia.

Koska signaali vääristyy sitä enemmän mitä suurempi c :n arvo on, myös vaihekerrotun puhesignaalin ymmärrettävyys muuttuu. Lausematriisitestissä peittoäänenä käytettävän hälypuheen tulisi olla puheenkaltaista, mutta mahdollisimman vähän ymmärrettävää, jotta lausematriisitestin tulokset pysyvät yksiselitteisinä.

Muutosta puheen ymmärrettävyydessä eri c arvoilla on tutkittu kuuntelukokeella, jossa lausematriisitestin testiprotokolla adaptoitiin mittaamaan hälypuheen ymmärrettävyyttä. Kuva 1 esittää vaihekerrotun hälypuheen ymmärrettävyyttä kertoimen c eri arvoilla. (Tästä kuuntelukokeesta ja tutkimuksesta voi lukea lisää tulevasta vertaisarvioidusta artikkelista [3].) Kuvassa 2 on esitetty aika-amplitudi ja spektrogrammi-kuvaajat suomenkielisestä puhesignaalista (a), jonka vaiheet on joko tuplattu (b) tai kokonaan satunnaistettu (c). Vaihetuplatun signaalin kuvaajissa on havaittavissa puheenkaltaisia aika-taajuusvaihteluita, vaikka tästä puhehälystä ei enää tunnista yksittäisiä sanoja. Vaihesatunnaistettu signaali on tasaista kohinaa.

4 KAIUNNAT

Kaiunnan tiedetään vaikeuttavan puheen ymmärrettävyyttä kun äänien jälkikaiuntahännät peittävät korviin seuraavaksi saapuvia ääniä riippuen suoran äänen ja kaiunnan välisestä suhteesta. Jos puhuja on hyvin lähellä kuuntelijaa, suora ääni on voimakas suhteessa kaiuntaan, eikä kaiunnan kestolla tai määrällä ole juurikaan vaikutusta puheen ymmärrettävyyteen [19]. Kun äänilähteen ja kuuntelijan etäisyys kasvaa, kaiunnalla (ml. varhaiset



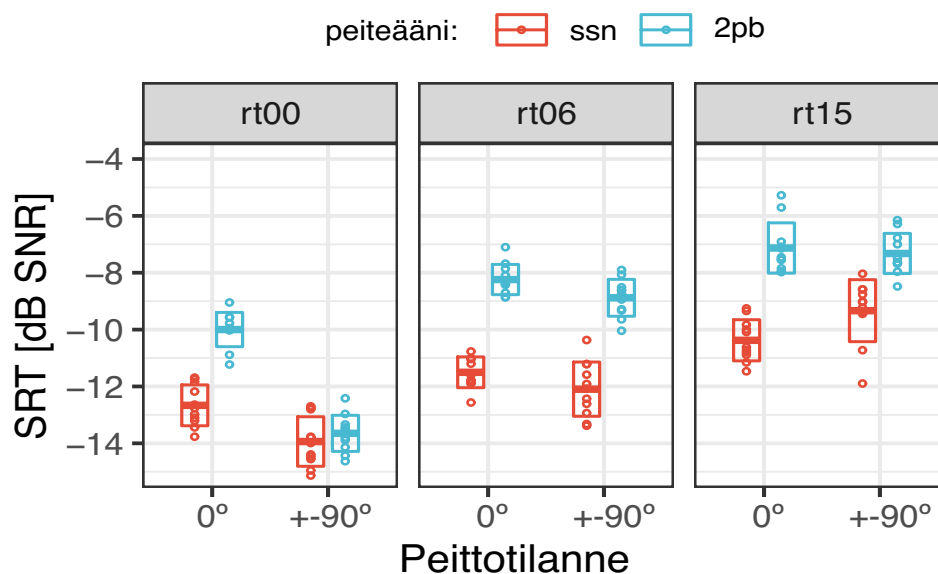
Kuva 2: Aika-amplitudi ja spektrogrammikuvaajat suomenkielisestä lauseesta, joka on vaihetuplattu ja vaihesatunnaistettu. a) Alkuperäinen puhe signaali; b) Vaihetuplattu; c) Vaihesatunnaistettu.

heijastukset) on myös lisääntyvä merkitys puheen ymmärrettävyyden kannalta. Vaikka jälkikaiuntahännät eivät suoraan peittäisikään seuraavia ääniä, kaiunta tasoittaa puheen amplitudimodulaatioita, jolloin puheen ymmärrettävyys heikkenee. Kaiunta myös dekorreloi ääniä korvien välillä, jolloin kuulojärjestelmän on vaikeampi saada kiinni ei toivotusta melun piirteistä ja kyky erottaa kohdeäänit ja melu toisistaan heikkenee [20].

Kaiunnan tiedetään häiritsevän enemmän kuulon alenemasta tai vaurioista kärsiviä kuuntelijoita kuin normaalikuuloisia verrokkeja [21]. Kuulontutkijat ja erityisesti kuulolaitteiden valmistajat ovatkin yhä enemmän kiinnostuneita realististen ääniympäristöjen tuomista mahdollisuuksista kuulolaitteiden kehitystyössä ja sovitusten parantamisessa [22].

Omissa tutkimuksissamme [19] olemme auralisoineet oikeita tiloja kuvalähdehajotelmalta (eng. *Spatial Decomposition Method, SDM* [23]). Kuvassa 3 on esitetty 50 prosentin puheentunnistuksen kynnystasoja mitattuna normaalikuuloisilta henkilöiltä käyttämällä joko spektrisovitettua (ssn)- tai vaihetuplattuakohinaa (2pb) peiteääninä [24]. Hälypuhetestinä näissä mittauksissa käytettiin suomalaista lausematriisitestiä ja kynnyksarvot mitattiin kaiuttomassa tilanteessa ilman auralisaatiota sekä kahdessa kaiuntaisessa tilanteessa, jotka auralisoitiin käyttämällä mitattuja huoneimpulssivasteita akustiikan laboratorion muunneltavan akustiikan huoneesta. Kaiuntaiset olosuhteet vastaavat keskitaajuuksilla noin 0,6 sekunnin (rt06) ja noin 1,5 sekunnin (rt15) jälkikaiunta-aikoja. Kohdepuhuja auralisoitiin kahden metrin etäisyydelle kuulijan eteen ja peiteäänilähteet auralisoitiin joko puhujan kanssa samaan paikkaan (0°) tai kahden metrin etäisyydelle kuulijan vasemmalle ja oikealle puolelle ($\pm 90^\circ$). Kuuntelukoe suoritettiin kuulokkeilla, ja koe suoritettiin siten, että koehenkilöt antoivat vastauksensa merkitsemällä tunnistetut sanat graafisen käyttöliittymän avulla (ns. ”suljetun setin”, ”closed-set” asetelma). Kokeeseen osallistui 11 suomalaista henkilöä, joilla kaikilla oli normaalikuulo äänesaudiometrialla mitattuna.

Referenssituloksena voidaan pitää kaiutonta ssn-rt00-(0°) tulosta (ka: -12.7 dB SNR). Verrattuna tähän referenssitulokseen näyttää siltä, että kaiuttomassa tilanteessa 2pb-kohina aiheuttaa n. 3 dB:ä huonomman tuloksen, kun peiteääni tulee samasta paikasta kohdepuheen kanssa. Tämä ero kuitenkin häviää, kun 2pb esitetään dikoottisesti ($\pm 90^\circ$) ja 2pb-peiteäänellä kaiuttomassa tilanteessa kuulijalle on etua, jos peiteääni on dekorreloitu korvien välillä. Kaiuntaisissa tilanteissa tämä etu 2pb-hälinän kohdalla häviää ja kaiuntaisissa tilanteissa n. 3 dB:n ero peiteäänien välillä on havaittavissa myös $\pm 90^\circ$ -tilanteissa. Huomion arvoista on myös, että lisäämällä referenssitestiin puolitoista sekuntia jälki-



Kuva 3: Puheentunnistuskynnystasoja mitattuna kahdella eri peiteäänellä (spektrisovitettu kohina (ssn), vaihetuplattuhälypuhe (2pb)), kolmessa eri kaiunnassa (ei kaiuntaa (rt00), n. 0,6 sekunnin jälkikaunta (rt06) ja n. 1,5 sekunnin jälkikaiunta (rt15)), ja kahdessa eri peittotilanteessa (puhe ja peiteääni samassa paikassa kuuntelijan edessä (0°), puhe edessä, mutta peiteääni vasemmalla ja oikealla ($\pm 90^\circ$)).

kaiuntaa saadaan aikaan myös noin 3 dB heikennys (kts. ssn: rt00-(0°) vs. rt15-(0°)). Vaikuttaa siltä, että 2pb-peiteäänien avulla voi olla mahdollista approksimoida kaiunnan vaikutusta hälypuhetestiin ilman, että kaiuntaa oikeasti lisätään testiääniin. Isoin kontrasti referenssitestitulokseen verrattuna saadaan kuitenkin käyttämällä 2pb-hälinää ja lisäämällä testiin myös kaiuntaa. Tämä testiasetus on lupaava kuulo-ongelmien diagnostiikan kannalta, joten on mielenkiintoista jatkaa tutkimusta tulevaisuudessa myös huonokuuloisten potilaiden kanssa.

5 YHTEENVETO

Tässä artikkelissa käsiteltiin lyhyesti hälypuhetesteissä ja kuulon tutkimuksessa yleisimmin käytettyjä peiteääniä ja niiden ominaisuuksia. Yleisimmin hälypuhetestien peiteääninä käytetään kohdepuhujan pitkäaikaisspektriin sovitettua tasaista kohinaa, mutta myös erilaiset amplitudimoduloidut kohinat sekä hälypuhehälinät ovat tyypillisiä peiteäänityyppejä. Myös auralisoidut, ns. realistiset ääniympäristöt ovat jo suhteellisen yleisessä tutkimuskäytössä erityisesti kuulolaitteiden valmistajilla. Artikkelissa tuotiin myös esille uusi vaihtoehtoinen tapa tuottaa hälypuhetta jo olemassa olevista puheäänistä käyttäen hyväksi Fourier-muunnoksen aikasiirto ominaisuutta. Uudella peittoäänityypillä ja virtuaaliakustiikalla voidaan luoda kontrolloitavissa olevia ääniympäristöjä kliiniseen kuulontutkimukseen siten, että peittoääni simuloi realistista hälyä huonetilassa. Tutkimustyö jatkuu ja lisää voi lukea tulevista julkaisuista [3, 24].

VIITTEET

- [1] V. Sivonen, T. Willberg, S. Sinkkonen, A. Aarnisalo, and A. Dietz. Suomenkielinen puheaudiometria ja uudet hälypuhetestit. *Suomen lääkärilehti*, 72(41):2032–2037, 2017.
- [2] A. Dietz, M. Buschermöhle, A. Aarnisalo, A. Vanhanen, T. Hyyrynen, O. Aaltonen, H. Löppönen, M. Zokoll, and B. Kollmeier. The development and evaluation of the finnish matrix sentence test for speech intelligibility assessment. *Acta otolaryngologica*, 134(7):728–737, 2014.
- [3] Antti Kuusinen. Phase multiplied babble noise for speech-in-noise tests (submitted manuscript under review). *The Journal of the Acoustical Society of America Express Letters*, 2021.
- [4] V. Välimäki, J. Rämö, and F. Esqueda. Creating endless sounds. In *Proc. 21st Int. Conf. Digital Audio Effects (DAFx-18)*, Aveiro, Portugal, pages 32–39, 2018.
- [5] D. Byrne, K. Dillon, H. and Tran, S. Arlinger, K. Wilbraham, R. Cox, B. Hagerman, R. Hetu, J. Kei, C Lui, et al. An international comparison of long-term average speech spectra. *The journal of the acoustical society of America*, 96(4):2108–2120, 1994.
- [6] B. Kollmeier, A. Warzybok, S. Hochmuth, M. Zokoll, V. Uslar, T. Brand, and K. Wagener. The multilingual matrix test: Principles, applications, and comparison across languages: A review. *International Journal of Audiology*, 54(sup2):3–16, 2015.
- [7] M. Stone, C. Füllgrabe, and B. Moore. Notionally steady background noise acts primarily as a modulation masker of speech. *The Journal of the Acoustical Society of America*, 132(1):317–326, 2012.
- [8] Finn Dubbelboer and Tammo Houtgast. The concept of signal-to-noise ratio in the modulation domain and speech intelligibility. *The Journal of the Acoustical Society of America*, 124(6):3937–3946, 2008.
- [9] H. Gustafsson and S. Arlinger. Masking of speech by amplitude-modulated noise. *The Journal of the Acoustical Society of America*, 95(1):518–529, 1994.
- [10] W. Dreschler, H. Verschuure, C. Ludvigsen, and S. Westermann. Iera noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. *Audiology*, 40(3):148–157, 2001.
- [11] I. Holube, S. Fredelake, M. Vlaming, and B. Kollmeier. Development and analysis of an international speech test signal (ists). *International journal of audiology*, 49(12):891–903, 2010.
- [12] S. Simpson and M. Cooke. Consonant identification in n-talker babble is a non-monotonic function of n. *The Journal of the Acoustical Society of America*, 118(5): 2775–2778, 2005.

- [13] I. Pollack and JM. Pickett. Stereophonic listening and speech intelligibility against voice babble. *The Journal of the Acoustical Society of America*, 30(2):131–133, 1958.
- [14] G. Kidd Jr, C. Mason, T. Rohtla, and P. Deliwala. Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *The Journal of the Acoustical Society of America*, 104(1):422–431, 1998.
- [15] J. Proakis and D. Manolakis, editors. *Digital signal processing : principles, algorithms, and applications*. Prentice Hall, Upper Saddle River, N.J., 3. ed edition, 1996.
- [16] M. Elhilali, T. Chi, and S. Shamma. A spectro-temporal modulation index (stmi) for assessment of speech intelligibility. *Speech communication*, 41(2-3):331–348, 2003.
- [17] A. Chabot-Leclerc, S. Jørgensen, and T. Dau. The role of auditory spectro-temporal modulation filtering and the decision metric for speech intelligibility prediction. *The Journal of the Acoustical Society of America*, 135(6):3502–3512, 2014.
- [18] A. Oppenheim and J. Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541, 1981.
- [19] A. Kuusinen, E. Saariniemi, V. Sivonen, A. Dietz, A. Aarnisalo, and T. Lokki. An exploratory investigation of speech recognition thresholds in noise with auralisations of two reverberant rooms. *International Journal of Audiology*, 60(3):210–219, 2021.
- [20] N. Durlach. Equalization and cancellation theory of binaural masking-level differences. *The Journal of the Acoustical Society of America*, 35(8):1206–1218, 1963.
- [21] A. Neuman, M. Wroblewski, J. Hajicek, and A. Rubinstein. Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. *Ear and hearing*, 31(3):336–344, 2010.
- [22] J. Cubick and T. Dau. Validation of a virtual sound environment system for testing hearing aids. *Acta Acustica united with Acustica*, 102(3):547–557, 2016.
- [23] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki. Spatial decomposition method for room impulse responses. *Journal of the Audio Engineering Society*, 61(1/2):17–28, 2013.
- [24] A. Kuusinen and T. Lokki. Effects of reverberation and babble noise on speech-in-noise recognition thresholds with normal hearing listeners in three different languages. *Unpublished manuscript*, 2022.