

LASKOSTUMISEN HAVAITSEMINEN SAHA-AALLOSSA

Heidi-Maria Lehtonen¹, Jussi Pekonen² ja Vesa Välimäki¹

¹ Aalto-yliopisto
Sähkötekniikan korkeakoulu
Signaalinkäsittelyn ja akustiikan laitos
PL 13000, 00076 AALTO
heidi-maria.lehtonen@aalto.fi

² Itsenäinen tutkija
Espoo
jussi@pekonen.cc

Tiivistelmä

Tässä artikkelissa esittelemme tutkimustuloksia kuuntelukokeista, joissa tutkittiin kuinka paljon laskostuneita taajuuskomponentteja perustaajuuden ylä- ja alapuolella pitää vaimentaa, jotta emme kuulisi niitä. Testiääninä käytimme saha-aaltoa kuudella eri perustaajuudella (415, 932, 1480, 2093, 3136 ja 3951 Hz), jotka soitettiin koehenkilöille 60 dB äänipainetasolla käyttäen kuulokkeita. Tulokset osoittavat, että perustaajuuden yläpuolella laskostumista on vaimennettava 0-41 dB ja vastaavasti perustaajuuden alapuolella 0-12 dB, perustaajuudesta riippuen. Esitämme lisäksi, että A-painotettu kohina-peittosuhde (noise-to-mask ratio) sopii laskostumista sisältävien saha-aaltosignaalien subjektiivisen laadun mittariksi. Lopuksi esitämme yleiset suunnittelusäännöt laskostumisvapaille digitaalisille saha-aalto-oskillaattoreille.

1 JOHDANTO

Saha-aalto on klassinen geometrinen aaltomuoto, jota käytettiin yleisesti elektronisissa syntetisaattoreissa 1960- ja 1970-luvuilla [1]. Nämä syntetisaattorit käyttivät vähentävää synteesiä, jossa ideana on suodattaa spektriltään rikasta lähdesignaalia, kuten saha- ja kanttiaaltoa, aikamuuttuvalla suodattimella. Nykyään syntetisaattorit mallintavat näitä aaltomuotoja digitaalisesti. Ongelmaksi muodostuu usein kuitenkin analogisen aaltomuodon epäjatkuvuudet, jotka tuottavat äärettömän monta harmonista komponenttia. Kun tällainen signaali näytteistetään, Nyquist-taajuuden yläpuolella olevat komponentit laskostuvat takaisin audiokaistalle. Nämä laskostuneet komponentit ovat selkeästi kuultavissa särö-tyyppisenä tai kohinamaisena häiriönä [2]. Monet digitaalisissa vähentävässä äänisynteesissä käytetyt menetelmät pyrkivät vähentämään laskostumista eri tavoin. Lisäksi tiedetään, että osa laskostuneista komponenteista jää kuulumattomiin taajuuspeittoilmiön takia ja matalilla taajuuksilla kuulokynnyksen ansiosta [3, 4]. Harmonisten sarja peittää spektrihiikkien läheisyydessä olevia laskostuneita komponentteja, kun taas kuulokynnys rajoittaa häiriön kuulumista perustaajuuden alapuolella.

Aiemmin vähentävässä äänisynteesissä on arvioitu laskostumista laskennallisilla menetelmin [5]. Tässä artikkelissa lähestymistapa on toisenlainen, sillä esittelemme tuloksia kuuntelukokeesta, joissa määritettiin laskostumisen peittokuulokynnys triviaalin saha-aallon tapauksessa [6]. Kuuntelukokeessa tutkittiin erikseen laskostumisen havaitsemista perustaajuuden ylä- ja alapuolella, joissa vastaavasti kuulokynnys ja taajuuspeittoilmiö määrittävät kuinka herkästi laskostumista havaitaan. Tutkimuksen tuloksia voidaan hyödyntää oskillaattorialgoritmien suunnittelussa ja arvioinnissa.

2 TESTISIGNAALIEN SYNTEESI

Kuuntelukokeissa esitettiin koehenkilöille kahdenlaisia signaaleja: referenssiääniä, jotka olivat additiivisella synteesillä generoituja saha-aaltoja, ja testiääniä, jotka sisälsivät referenssiäänien, johon oli summattu laskostunut komponentti. Referenssiäänien $s_{\text{ref}}(n)$ voidaan ilmaista seuraavasti:

$$s_{\text{ref}}(n) = -\frac{2}{\pi} \sum_{k=1}^K \frac{1}{k} \sin\left(\frac{2\pi nk f_0}{f_s}\right), \quad (1)$$

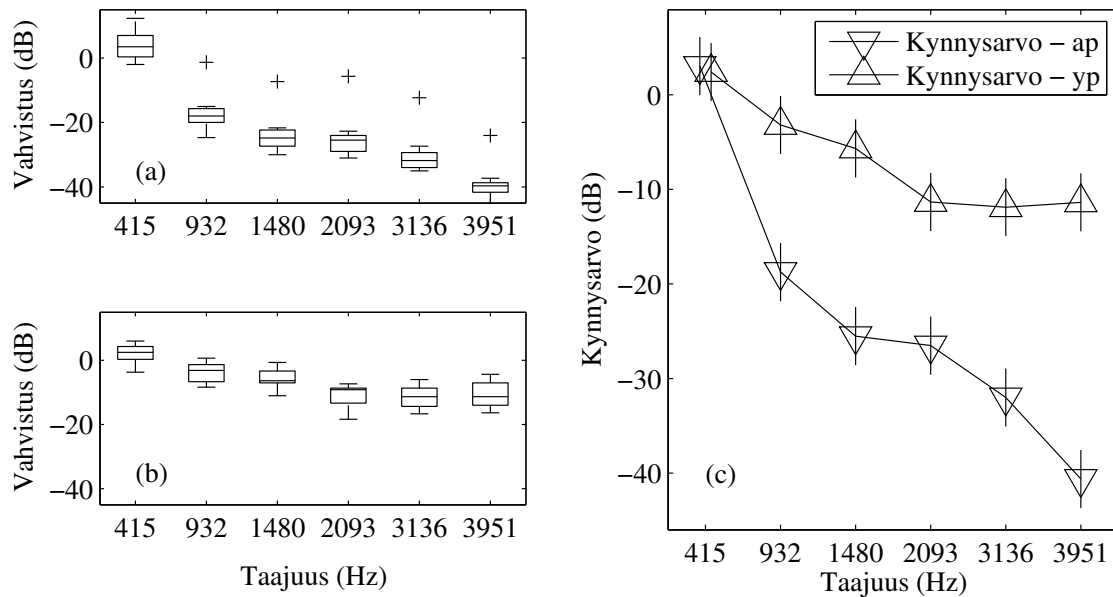
missä n on näyteindeksi, k on harmonisen indeksi, f_0 on perustaajuus, f_s on näytetaajuus (44100 Hz) ja parametri $K = \text{floor}(f_s/f_0)$. Laskostunut komponentti $s_{\text{lask}}(n)$ voidaan laskea triviaalista saha-aallosta $s_{\text{lask}}(n) = s_{\text{triv}}(n) - s_{\text{ref}}(n)$. Testiääniä oli kahdenlaisia, ja ne sisälsivät laskostuneen komponentin joko perustaajuuden ylä- tai alapuolella. Tämän lähestymistavan avulla voitiin tutkia erikseen taajuuspeittoilmiön ja kuulokynnyksen vaikutusta laskostumisen havaitsemiseen. Perustaajuuden ylä- ja alapuoliset komponentit saatiin suodattamalla signaali $s_{\text{lask}}(n)$ yli- ja alipäästösuodattimilla, joiden rajataajuus oli sama kuin äänen perustaajuus. Suodattamisessa käytettiin 10000-asteista lineaarivaiheista FIR-suodatinta, joka suunniteltiin käyttäen Chebyshev-ikkunointia 100 dB:n sivukeilavaimennuksella. Testiäänien voidaan kirjoittaa:

$$s_{\text{testi}}(n) = s_{\text{ref}}(n) + 10^{G/20} \hat{s}_{\text{lask}}(n), \quad (2)$$

missä $\hat{s}_{\text{lask}}(n)$ on joko yli- tai alipäästösuodatettu laskostunut komponentti ja G on sen vahvistus desibeleissä. Referenssiäänien toistovoimakkuus kalibroitiin C-painotuksella 60 dB äänipainetasolle käyttämällä telinettä, johon kiinnitettiin testissä käytetyt kuulokkeet (Sennheiser HD-650) ja äänipainetasomittari siten, että äänipainetaso saatiin mitattua kuulokkeen ovaalinmuotoisen tyynyn sisältä. Itse testin aikana säädettiin vain testiäänien laskostuneen komponentin tasoa (käytännössä kaavan (2) muuttujaa G).

3 MENETELMÄT

Kuuntelukoe koostui neljästä osasta, joista jokaiseen kuului kolme eri perustaajuista ääntä, joissa laskostunut komponentti oli joko perustaajuuden ylä- tai alapuolella. Testi suoritettiin Aalto-yliopiston signaalinkäsittelyn ja akustiikan laitoksen kuunteluhuoneessa, ja siihen osallistui kymmenen 22-29-vuotiaasta normaalikuuloista opiskelijaa. Testimenetelmänä käytettiin 3AFC-menetelmää (three-alternative forced-choice) [7].



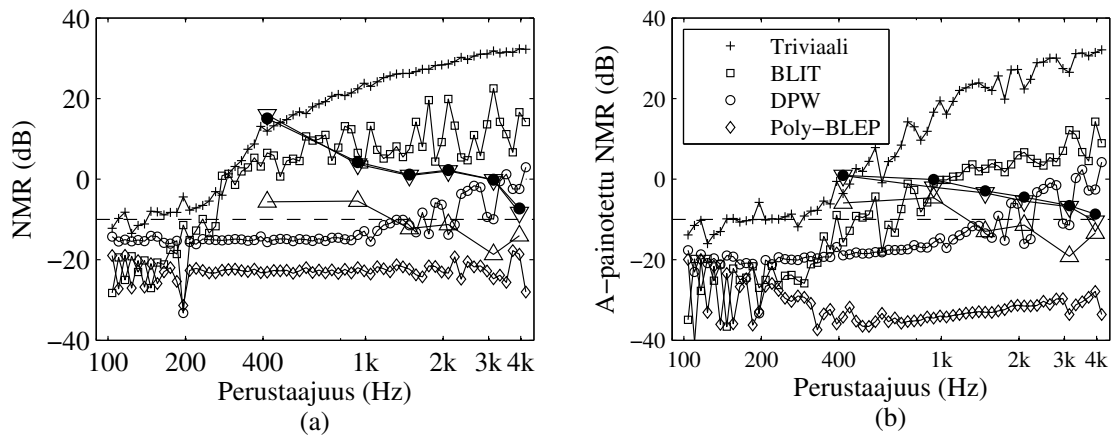
Kuva 1: Kuuntelukokeen tulokset perustaajuuden (a) alapuolisen ja (b) yläpuolisen komponentin tapauksessa. Laatikon keskiviiva osoittaa mediaanin sekä reunat 25 % ja 75 % pisteet. Datajoukosta selvästi erottuvat havainnot on merkitty +-merkillä. (c) Kynnysarvojen ryhmäkeskiarvot perustaajuuden alapuolella (ap) ja yläpuolella (yp). Pystyviivat näyttävät 95 % luottamusvälit.

Kaksi äänistä oli referenssiäniä ja kolmas testiäni, ja koehenkilön tehtävä oli löytää testiäni yhden kuuntelukerran jälkeen. Kunkin äänen pituus oli 700 ms ja ne oli erotettu toisistaan 500 ms hiljaisuudella. Testimenetelmä seurasi adaptiivista “kaksi-alas, yksi-ylös” sääntöä, eli kaavan (2) G -arvoa pienennettiin kun koehenkilö oli löytänyt testiäniäni kaksi kertaa peräkkäin oikein, mutta sitä kasvatettiin jo yhden väärän vastauksen jälkeen. G -parametrin alkuarvo saatiin pilottikokeesta, jonka perusteella saatiin kynnysarvoihin lisättiin 15 dB. Testi päättyi kahdeksan käännöspisteen jälkeen, joista kahden ensimmäisen ajan askelkokona käytettiin 6 dB ja tämän jälkeen 2 dB. Lopullinen kynnysarvo laskettiin kuuden viimeisen käännöspisteen keskiarvona. Kussakin osassa olevat testitapaukset limitettiin, ja niiden toistojärjestys satunnaistettiin.

4 TULOKSET

Kuvat 1 (a) ja (b) esittävät kuuntelukokeen tulokset perustaajuuden ala- ja yläpuolisen laskostuneen komponentin tapauksessa. Tulokset analysoitiin kaksisuuntaisen ANOVA:n avulla. Tekijöinä kaksisuuntaisessa ANOVA-analyysissä olivat äänen perustaajuus sekä se, oliko laskostunut komponentti perustaajuuden ylä- vai alapuolella. Analyysi osoitti, että sekä perustaajuus [$F(5, 96) = 184, 6; p \ll 0, 05$] että laskostuneen komponentin taajuusalue [$F(1, 96) = 736, 2; p \ll 0, 05$] olivat merkittäviä tekijöitä. Myös niiden yhteisvaikutus todettiin merkittäväksi [$F(5, 96) = 43, 3; p \ll 0, 05$].

Kuva 1 (c) esittää ryhmäkeskiarvot ja niiden 95% luottamusvälit. Ylös- ja alaspäin osoittavat kolmiot esittävät tapaukset, jossa laskostuneet komponentit ovat perustaajuuden



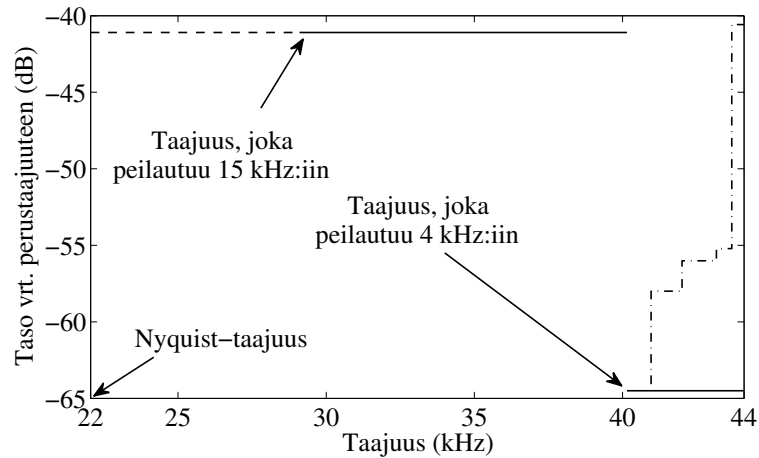
Kuva 2: (a) NMR- ja (b) ANMR-arvot neljälle saha-aalto-oskillaattorille ja testiäänille, joiden laskostunut komponentti on juuri kynnsarvon tasolla. Alaspäin ja ylöspäin osoittavat kolmiot viittaavaat laskostuneeseen komponenttiin perustaajuuden ala- ja yläpuolella, vastaavasti, ja pisteet osoittavat tapaukset, joissa molemmat laskostuneet komponentit on huomioitu. Katkoviiva osoittaa NMR-mittarin mukaisen kuulokynnyksen.

ylä- ja alapuolella, vastaavasti. Kuvasta 1 (c) voidaan nähdä, että perustaajuuden 415 Hz tapauksessa kynnsarvot ovat tilastollisesti yhtä suuret. Lisäksi perustaajuuden alapuolisen komponentin tapauksessa luottamusvälit menevät päällekkäin perustaajuuksilla 1480 ja 2093 Hz. Perustaajuuden yläpuolisen komponentin tapauksessa perustaajuuksien 932 ja 1480 Hz sekä 2093, 3136 ja 3951 Hz luottamusvälit menevät päällekkäin. Nämä tulokset varmistettiin Tukeyn Honestly Significant Difference -testillä.

5 A-PAINOTETTU NMR JA SUUNNITTELUSÄÄNNÖT

NMR:ää on ehdotettu laskostumista sisältävien signaalien laadun mittariksi [2], ja se määritellään häiriön ja hyötysignaalin yksinkertaistetun peittokuulokynnyksen tehojen suhteena [8]. Kuvassa 2 (a) on esitetty NMR-arvot kuudelle testiäänelle. Alas- ja ylöspäin olevat nuolet viittaavat tapauksiin, joissa laskostuminen on perustaajuuden ala- ja yläpuolella, vastaavasti, ja pisteet viittaavat tapaukseen, jossa molemmat laskostuneet komponentit on lisätty referenssiääneseen. Laskostuneiden komponenttien taso on asetettu kuuntelukokeista saadun kynnsarvojen tasolle. Lisäksi kuvassa on esitetty NMR-luvut triviaaleille saha-aallolle sekä kolmella eri synteesimenetelmällä (BLIT (bandlimited impulse train) [9], DPW (differentiated parabolic waveform) [10] ja Poly-BLEP (polynomial bandlimited step function) [11]) approksimoituille signaaleille.

Audiokoodauksen sovelluksissa yli -10 dB:n NMR-arvot viittaavat, että häiriö todennäköisesti on kuultavissa (katkoviiva kuvassa 2 (a)) [8]. Kuvasta 2 (a) nähdään, että tässä työssä esitettyjen kuuntelukokeiden perusteella kynnsarvojen kohdalla NMR-arvot vaihtelevat -20 ja -5 dB:n välillä, kun laskostunut komponentti on perustaajuuden yläpuolella. Tämä on yllättävää, sillä määritelmän mukaan juuri havaittavien tapauksien NMR-arvojen pitäisi olla -10 ja 0 dB:n välillä. Lisäksi perustaajuuden alapuolisen komponentin tapauksessa NMR-arvot ovat 15 dB:n luokkaa, mikä on yllättävän paljon kuuntelukokeista saadun tuloksen valossa. Nämä erot saattavat johtua NMR:n käyttämisestä



Kuva 3: Suurimmat sallitut tasot ensimmäisen kertaluvun laskostuneille komponenteille kun $f_s = 44,1$ kHz. Yhtenäinen viiva esittää konservatiivisen tason ja pistekatkoviiva realistisemman kynnysarvon perustaajuuden alapuolelle laskostuville komponenteille.

liian yksinkertaisesta kuulokynnysmallista. Erot kuuntelukokeista saatujen tulosten ja NMR:n antamien laskennallisten arvojen välillä pienenevät kun signaalit prosessoidaan A-painotussuotimella [6]. Tämä valinta tehtiin, sillä A-painotuskäyrä on yleisesti käytetty standardi ja lisäksi sen spektrin muoto on käännteinen kuulokynnykseen nähden. Kuva 2 (b) esittää A-painotetun NMR:n testiäänille ja tutkituille algoritmeille, ja siitä voidaan nähdä, että A-painotus tuo kuuntelukokeiden tulokset ja laskennalliset tulokset lähemmäs toisiaan erityisesti perustaajuuden alapuolisen komponentin tapauksessa.

Laskostuneiden komponenttien vaimennukselle sekä perustaajuuden ylä- että alapuolella voidaan johtaa minimivaatimukset skaalaamalla triviaalin saha-aallon amplitudiltaan suurimmat laskostuneet komponentit kuuntelukokeista saaduilla kynnysarvoilla. Nämä komponentit ovat peräisin ensimmäisen kertaluvun laskostuneesta sarjasta ja ne sijoittuvat perustaajuuden alapuolelle sekä perustaajuuden ja toisen harmonisen väliin. Lisäksi niiden amplitudit voidaan laskea [6]. Kun minimivaimennus tunnetaan, tiedetään myös jatkuva-aikaisen saha-aallon taajuuskomponenttien suurin sallittu taso Nyquist-taajuuden yläpuolella. Tätä tietoa voidaan taas käyttää hyväksi digitaalisten oskillaattorialgoritmien suunnittelussa valitsemalla algoritmin parametreiksi arvot, jotka täyttävät nämä ehdot. Kuva 3 esittää maksimiampplitudit ensimmäisen kertaluvun laskostuneelle sarjalle, kun näytetaajuus on 44,1 kHz. Yhtenäinen viiva on konservatiivinen raja, joka on saatu kuuntelukokeiden tulosten perusteella perustaajuuden ylä- ja alapuolella, ja se ottaa huomioon myös kynnysarvojen luottamusvälit. Perustaajuuden alapuolella tämä raja on liian konservatiivinen, ja pistekatkoviiva esittää realistisemman kynnysarvon. Katkoviiva esittää taajuusalueen, joka peilautuu 15 kHz:n yläpuolelle. Tämä voidaan jättää huomiotta, sillä kuulo ei ole herkkä tällä alueella.

6 JOHTOPÄÄTÖKSET

Tässä artikkelissa esittelimme tuloksia kuuntelukokeista, joissa etsittiin laskostuneen komponentin tasolle kynnysarvoa saha-aallon tapauksessa. Tulokset osoittavat, että pe-

rustaajuudella 415 Hz laskostunutta komponenttia ei tarvitse vaimentaa perustaajuuden ylä- tai alapuolella. Korkeampien äänten tapauksessa perustaajuuden alapuolella olevaa komponenttia on vaimennettava 19-41 dB ja perustaajuuden yläpuolella olevaa komponenttia 3-12 dB. Osoitimme myös, että A-painotettu NMR on hyvä laskennallinen mittari laskostumista sisältävien saha-aaltojen subjektiivisen laadun arviointiin. Lopuksi esitimme suunnittelusäännöt laskostumisvapaille digitaalisille sala-aalto-oskillaattoreille.

KIITOKSET

Heidi-Maria Lehtosen työtä rahoittaa Suomen Kulttuurirahasto.

VIITTEET

- [1] MOOG R A, Voltage-controlled electronic music modules, *J. Audio Eng. Soc.*, **13**(1965) 3, 200–206.
- [2] VÄLIMÄKI V & HUOVILAINEN A, Antialiasing oscillators in subtractive synthesis, *IEEE Signal Process. Mag.*, **24**(2007) 2, 116–125.
- [3] ZWICKER E & FASTL H, *Psychoacoustics – Facts and Models*, pages 15–19 and 56–102, Springer-Verlag, Berlin, Germany, 1990.
- [4] MOORE B C J, *An Introduction to the Psychology of Hearing*, pages 49–53, 93–94, and 103–114, Academic Press, London, UK, 4th edition, 1997.
- [5] NAM J, VÄLIMÄKI V, ABEL J S, & SMITH J O, Efficient antialiasing oscillator algorithms using low-order fractional delay filters, *IEEE Trans. Audio, Speech, Language Process.*, **18**(2010) 4, 773–785.
- [6] LEHTONEN H M, PEKONEN J, & VÄLIMÄKI V, Audibility of aliasing distortion in sawtooth signals and its implications for oscillator algorithm design, *J. Acoust. Soc. Am.*, **132**(2012) 4, 2721–2733.
- [7] LEVITT H, Transformed up-down methods in psychoacoustics, *J. Acoust. Soc. Am.*, **49**(1971) 2, 467–477.
- [8] BRANDENBURG K, Evaluation of quality for audio encoding at low bit rates, in *Proc. 82nd AES Convention*, London, UK, 1987.
- [9] STILSON T & SMITH J, Alias-free digital synthesis of classic analog waveforms, in *Proc. Int. Computer Music Conf.*, pages 332–335, Hong Kong, 1996.
- [10] VÄLIMÄKI V, Discrete-time synthesis of the sawtooth waveform with reduced aliasing, *IEEE Signal Process. Lett.*, **12**(2005) 3, 214–217.
- [11] VÄLIMÄKI V, PEKONEN J, & NAM J, Perceptually informed synthesis of band-limited classical waveforms using integrated polynomial interpolation, *J. Acoust. Soc. Am.*, **131**(2012) 1, 974–986.