
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Fagerström, Jon; Välimäki, Vesa
Realistisen vaihtelun tuottaminen iskuäänille

Published in:
Akustiikkapäivät 2021

Julkaistu: 22/11/2021

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Fagerström, J., & Välimäki, V. (2021). Realistisen vaihtelun tuottaminen iskuäänille. teoksessa *Akustiikkapäivät 2021* (Sivut 259-264). (Akustiikkapäivät). Akustinen Seura ry. https://www.akustinenseura.fi/wp-content/uploads/2021/11/akustiikkapaivat_2021_s259.pdf

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

REALISTISEN VAIHTELUN TUOTTAMINEN ISKUÄÄNILLE

Jon Fagerström ja Vesa Välimäki

Aalto-yliopiston sähkötekniikan korkeakoulu
Signaalinkäsittelyn ja akustiikan laitos, akustiikan laboratorio
Otakaari 5, 02150 Espoo
jon.fagerstrom@aalto.fi

Tiivistelmä

Tässä artikkelissa esitellään algoritmi, joka tuottaa hienovaraisia muutoksia iskuääniin. Tavoitteena on parantaa nämplyksen laatua, missä yksittäisen äänitteen peräkkäinen toisto kuulostaa epäaidolta ja toistuvalla. Äänen spektrien eroja tutkittiin kahdessa äänitteessä, joista ensimmäisessä lyötiin toistuvasti virvelirumpua ja toisessa koputettiin ovea. Spektrianalyysia hyödynnettiin algoritmin kalibroimisessa eri äänilähteille. Kehittämämme algoritmi hyödyntää lyhyttä ja harvaa samettikohinasuodinta ja bassohyllykorjainta satunnaisilla parametreilla, ja tuottaa loputtoman määrän uusia realistisia muunnelmia yhdestä äänitteestä. Muokattujen äänen realistisuutta tutkittiin kuuntelukokeen avulla, jossa verrattiin ehdottamamme menetelmää aiempaan. Tulokset osoittavat esitetyn algoritmin tuottavan vähintään yhtä realistisen lopputuloksen kuin aiempi menetelmä. Ehdottamamme menetelmä on kuitenkin tehokkaampi, koska se käyttää 77% vähemmän laskuoperaatioita. Uusi menetelmä soveltuu videopelien äänitehosteiden ja tietokone musiikin elävöittämiseen.

1 JOHDANTO

Sämply (engl. sampling) on yleinen äänisynteesitekniikka, jota käytetään syntetisaattoreissa, rumpukoneissa ja videopelien äänitehosteissa [1]. Sämply perustuu äänitteiden soittamiseen ja mahdollistaa luonnollisen äänenlaadun. Yksittäisen äänitteen peräkkäinen toisto kuulostaa kuitenkin epäaidolta ja toistuvalla. Tätä ilmiötä kutsutaan konekiväärifektiksi (engl. “machine-gun effect”) [2]. Monilla signaalinkäsittelymenetelmillä voidaan muokata ääniä dramaattisesti [1, 3]. Hienovaraisten äänenvärien muutosten tuottamiseen sämplytyn ääneen ei ole kuitenkaan esitetty helppoa ratkaisua.

Videopelin ääniefekteihin luodaan vaihtelua usein kiertovuorottelulla (engl. round-robin playback). Siinä jokaisesta äänitehosteesta talletetaan useita esimerkkejä, joita toistetaan vuorotellen satunnaisessa järjestyksessä. Tämä tehdään kaikkia tarvittavia äänilähteitä varten [4]. Erilaisia morfaus- eli muodonmuutostekniikoita (engl. morphing) voidaan käyttää hienovaraisten muutosten tuottamiseen, kun käytettävissä on vähintään kaksi esimerkkiäänitettä. [5, 6]. Myös kahden tai useamman nämplyn ristisynteesillä, joka

Copyright ©2021 Jon Fagerström et al. Tämä on avoimesti julkaistu teos, joka noudattaa Creative Commons NIMEÄ 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.

hyödyntää lineaariennustusta, voidaan tuottaa variaatioita [7]. Edellä mainitut menetelmät eivät kuitenkaan takaa realististen muutosten koko kirjoa.

Korkea-asteista lineaariennustusta ($N = 1000$) voidaan hyödyntää variaatioiden tuottamiseen laskemalla yksittäisen äänitteen käänteissuodin ja korvaamalla jäännössignaali jokaista variaatioita kohden satunnaiskohinalla. [8]. Lloyd ym. hyödynsivät kuoppasuotimia satunnaisilla parametreilla tuottaakseen uskottavia muutoksia iskuääniin [9]. Heidän algoritminsa käyttää kymmentä toisen asteen suodinta satunnaisella keskitaajuudella, vahvistuksella ja Q -arvolla (vrt. kaistanleveys). Vaihteluvälejä satunnaishparametrien arvoille ei raportoitu.

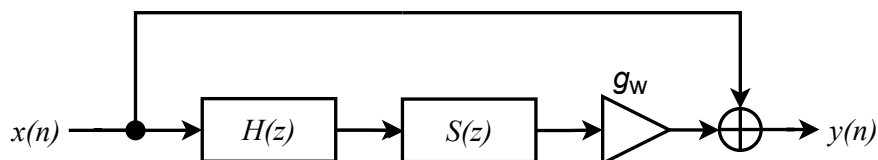
Tässä artikkelissa esitellään uusi ratkaisu ongelmaan, jolloin saatavilla on vain yksi esimerkkiäänite, josta halutaan tuottaa realistisia variaatioita tehokkaasti ja tosiaikaisesti. Uusi menetelmä hyödyntää lyhyttä samettikohinasuodinta [10, 11] ja bassohyllysuodinta spektrimuutosten tuottamiseksi. Samettikohina- ja bassohyllysuodattimen parametrit määritetään esimerkkiäänitteiden spektrijakauman perusteella. Äänenlaatu varmistetaan kuuntelukokeen avulla. Tämän tutkimuksen tuloksia on esitelty aiemmin DAFX21-konferenssissa [12].

2 UUSI MENETELMÄ

Ehdotetun variaatiosuotimen lohkokaavio esitetään kuvassa 1. Se koostuu suorasta ja suodatetusta signaalireitistä, missä signaali käsitellään sarjaankytketyillä ensimmäisen asteen bassohyllysuotimella $H(z)$ ja samettikohinasuotimella $S(z)$. Lisäksi vahvistuskerroin g_w mahdollistaa variaation voimakkuuden säädön. Tutkimuksessa käytetyt suodatinparametrit on ilmoitettu taulukossa 1.

Samettikohinasuodattimen impulssivaste on lyhyt ($2 \leq L_s \leq 4$ ms) ja sisältää vain kahdeksan impulssia ($M = 8$). Loput impulssivasteen näytteet ovat nollija. Suotimen impulssivaste vaimenee loppua kohti ja parametri L_{dB} ilmoittaa kokonaisvaimennuksen desibeleinä. Bassohyllysuodin mahdollistaa matalien taajuuksien vaimentamisen valitun rajataajuuden alapuolella, jolloin samettikohinasuotimen tuottamat variaatiot saadaan kohdistettua suurille taajuuksille. Rajataajuuden F_c kohdalla bassohyllysuodin saavuttaa puolet valitusta maksimivaimennuksesta G_{10} desibeleinä.

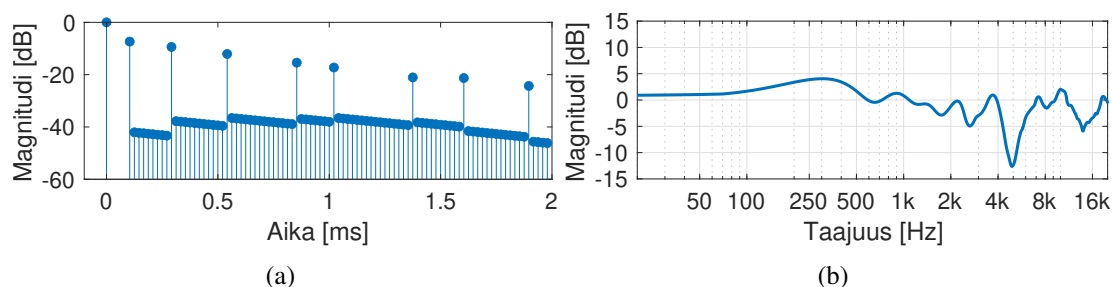
Kuvissa 2(a) ja 2(b) näytetään esimerkki variaatiosuodattimen impulssivasteesta ja Bark-tasoitetusta magnitudivasteesta. Suoran reitin osuus nähdään impulssivasteessa hetkellä 0 ja sitä seuraa bassohyllysuodatettu samettikohinasuotimen vaste.



Kuva 1: Ehdotetun variaatiosuotimen lohkokaavio, jossa bassohyllysuodin on $H(z)$ ja samettikohinasuodin on $S(z)$ [12].

Taulukko 1: Vertailussa käytetyt uuden menetelmän suodatinparametrit.

Äänilähde	F_c	G_{lo}	L_s	M	L_{dB}	g_w
Virvelirumpu	100 Hz	-5 dB	4 ms	8	20 dB	0.2
Koputus	50 Hz	-25 dB	2 ms	8	20 dB	0.4



Kuva 2: (a) Esimerkki variaatiosuodattimen impulssivasteesta ja (b) sitä vastaavasta Bark-tasoitetusta magnitudivasteesta.

3 TULOKSET

Tässä kappaleessa esitetään objektiivisen spektrianalyysin ja subjektiivisen kuuntelukokeen tulokset. Näissä analyyseissä vertailtiin uuden menetelmän äänenlaatua aiempaan satunnaissuodatusmenetelmään [9]. Alkuperäistoteutuksesta poiketen pelkästään vahvistuskertoimet satunnaistettiin aiemmassa menetelmässä. Näin vältettiin useampien suodattimien keskittyminen päällekkäisille taajuusalueille. Lisäksi verrataan menetelmien laskentakuormia.

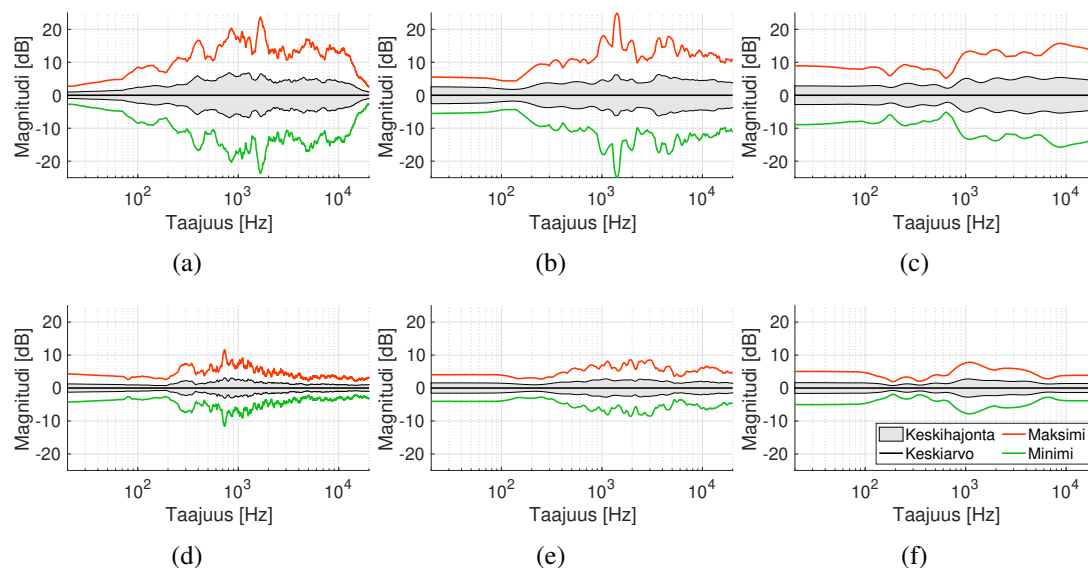
3.1 Spektrianalyysi

Uuden ja aiemman menetelmän [9] tuottamia magnitudivasteen muutoksia verrattiin muutoksiin, joita havaittiin äänitteissä (virvelirumpu ja koputus). Tässä analyysissä 30 sämplin magnitudivasteen erotukset laskettiin pareittain kaikilla mahdollisilla permutaatioilla ja niistä laskettiin tilastolliset arvot, ks. kuva 3. Jokainen magnitudivaste Bark-tasoitettiin, jotta magnitudivasteiden resoluutio vastaisi ihmisen kuuloa.

Kuvissa 3 (a), (b) ja (c) näytetään spektrianalyysin tulokset koputusäänelle. Uusi menetelmä (kuva 3(b)) tuottaa aiempaa menetelmää (kuva 3(c)) yksityiskohtaisempia muutoksia magnitudivasteeseen aiemman menetelmän seurattessa äänitteissä havaittavien variaatioiden tasoitettua muotoa. Sama ilmiö toistuu tarkasteltaessa virvelirummun spektrianalyysin tuloksia kuvassa 3 (d), (e) ja (f). Virvelirummun tapauksessa aiempi menetelmä (kuva 3(f)) vastaa paremmin äänitteissä havaittavien muutosten (kuva 3(d)) suuruutta korkeilla taajuuksilla verrattaessa uuden menetelmän tuottamiin muutoksiin (kuva 3(e)). Koputusääniä ja virvelirummun ääniä vertailtaessa on myös selvästi havaittavissa, että koputusäänet sisältävät huomattavasti suurempia spektrivaihteluja.

3.2 Laskentamäärä

Laskutoimitusten lukumäärät näytettä kohti on esitetty taulukossa 2. GEQ viittaa aiempaan satunnaissuodatusmenetelmään, joka käyttää kymmentä toisen asteen suodinta [9].



Kuva 3: Spektrivaihtelun jakaumat koputukselle ja virvelirummulle (a), (d) äänitteissä, (b), (e) uudella menetelmällä suodatettuna ja (c), (f) aiemmalla menetelmällä [9]. [12].

Kukin kymmenestä toisen asteen suotimesta tarvitsee neljä yhteen- ja viisi kertolaskua eli yhteensä 90 laskuoperaatiota. Uudessa menetelmässä samettikohinasuodin vaatii kahdeksan yhteen- ja kahdeksan kertolaskua ja hyllykorjain lisäksi kaksi yhteen- ja kolme kertolaskua eli yhteensä 21 operaatiota. Uusi menetelmä käyttää siis 77% vähemmän laskutoimituksia kuin aiempi menetelmä.

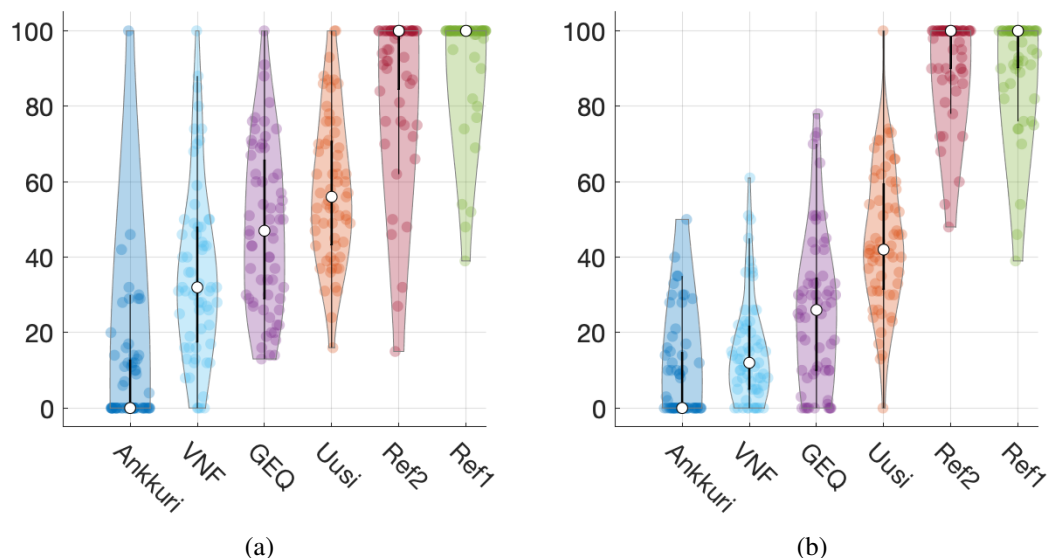
3.3 Kuuntelukoe

Suodatettujen sämplöjen realistisuutta tutkittiin MUSHRA-kuuntelukokeessa (engl. Multiple Stimuli with Hidden Reference and Anchor). Kokeeseen osallistui 21 henkilöä, joilla oli normaali kuulo. Heistä 20 oli osallistunut aiemmin kuuntelukokeeseen. Kuuntelukoe ohjelmoitiin webMUSHRA-sovelluksella [13] ja toteutettiin Aalto-yliopiston akustiikan laboratorion äänieristetyissä kuuntelukopeissa Sennheiser HD-650 -kuulokkeilla.

Kokeessa osallistujien tuli arvioida testisignaalien äänenvärien variaatioiden realistisuutta verrattuna referenssisignaaliin. Testisignaalit kuvataan taulukossa 3, jossa “VNF” viittaa samettikohinasuodatukseen ilman bassohyllysuodinta ja suoraa signaalireittiä. Jokainen testisignaali oli 20 sämplin aikakvantisoitu sekvenssi, joissa yksittäisten äänten kokonaisenergia oli normalisoitu. Ajallisen kvantisoinnin ja energianormalisoinnin tarkoituksena oli poistaa ajoituksen (rytmin) ja äänekkyuden vaihtelut, jotta koe keskittyi pelkästään äänenväriin. Kokeen arviointiasteikko oli 0–100. Arviointiasteikon ylintä väliä 80–100 vastasi “realistinen”, jota seurasivat “lähes realistinen” (60–80), “hieman realistinen” (40–60), “epärealistinen” (20–40) ja viimeisenä “erittäin epärealistinen” (0–20).

Taulukko 2: Uuden ja aiemman GEQ menetelmän laskentakuormien vertailu.

Menetelmä	Yhteenlaskut	Kertolaskut	Yhteensä	Säästö
GEQ [9]	40	50	90	Vertailuarvo
Uusi	10	11	21	77%



Kuva 4: Kuuntelukokeen tulokset (a) koputus- ja (b) virvelirumppuäänille [12].

Kuuntelukokeen tulokset esitetään koputusäänelle kuvassa 4(a) ja virvelirummulle kuvassa 4(b). Osallistujat tunnistivat molemmista äänistä keskimäärin täydellisesti niin ankkurin (mediaani 0) kuin myös piilotetun referenssin Ref1 (mediaani 100). Myös referenssin sekoitettu versio (Ref2) arvioitiin keskimäärin realistiseksi. Kaikki vertailut suodatusmenetelmät pärjäsivät keskimääräisesti paremmin koputusäänen kohdalla, mutta koputusäänten pisteiden jakauma oli laajempi kuin rumppuäänten. Uusi menetelmä oli tilastollisesti parempi kuin ankkurisignaali (luottamusvälit eivät kohtaa kuvassa 4) ja vähintään yhtä hyvä kuin aiempi menetelmä (jakaumat menevät päällekkäin).

4 YHTEENVETO

Tässä artikkelissa esiteltiin uusi menetelmä hienovaraisten muutosten tuottamiseksi sämplätyihin iskuääniin [12]. Uusi menetelmä hyödyntää lyhyttä ja harvaa samettikohinasuodinta satunnaisten spektrimuutosten tuottamiseen. Lisäksi menetelmä vaatii ensimmäisen asteen bassohyllysuotimen, jolla tuotetut muutokset kohdistetaan korkeille taajuuksille, missä vaihteluita esiintyy eniten myös äänitteissä.

Uuden menetelmän osoitettiin tuottavan yksityiskohtaisia ja realistisia spektrimuutoksia, vaikkakin aiempi menetelmä seurasi äänitteistä analysoitujen muutosten laajuutta tarkem-

Taulukko 3: Kuuntelukokeen testisignaalit.

Testisignaali	Kuvaus
Ref1	Referenssisekvenssi, jossa soitetaan peräkkäin 20 äänitettyä sampleä
Ref2	Ref1-sekvenssin 20 sampleä satunnaisessa järjestyksessä
Ankkuri	Ref1:n ensimmäinen sample toistetaan 20 kertaa
VNF	20 yksinkertaistetulla menetelmällä suodatettua sampleä
GEQ	20 GEQ-menetelmällä [9] suodatettua sampleä
Uusi	20 uudella menetelmällä suodatettuna sampleä

min. Kuuntelukokeen perusteella uusi menetelmä paransi selvästi realismia verrattuna yksittäisen äänitteen peräkkäiseen toistoon. Lisäksi uusi menetelmä oli kuuntelukokeessa yhtä hyvä tai niukasti parempi aiempaan menetelmään verrattuna, vaikka uusi menetelmä käyttää 77% vähemmän laskuoperaatioita.

VIITTEET

- [1] J. O. Smith. Viewpoints on the history of digital synthesis. In *Proc. Int. Computer Music Conf.*, Montreal, Canada, Oct. 1991.
- [2] K. J. Werner, J. S. Abel, and J. O. Smith III. A physically-informed, circuit-bendable, digital model of the Roland TR-808 bass drum circuit. In *Proc. Int. Conf. Digital Audio Effects (DAFx)*, pages 159–166, Erlangen, Germany, Sept. 2014.
- [3] M. Ilmoniemi, V. Välimäki, and M. Huotilainen. Subjective evaluation of musical instrument timbre modifications. In *Proc. Baltic-Nordic Acoustics Meeting (BNAM)*, Mariehamn, Åland, June 2004.
- [4] B. Schmidt. Interactive mixing of game audio. In *Proc. Audio Eng. Soc. 115th Conv.*, New York, NY, USA, Oct. 2003. paper 5857.
- [5] W. Ahmad, H. Hacıhabiboglu, and A. M. Kondo. Morphing of transient sounds based on shift-invariant discrete wavelet transform and singular value decomposition. In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pages 297–300, Taipei, Taiwan, Apr. 2009.
- [6] S. Siddiq. Real-time morphing of impact sounds. In *Proc. Audio Eng. Soc. 139th Conv.*, New York, NY, USA, Oct. 2015. paper 9407.
- [7] J. Wikström. Improved sound sampling. Master’s thesis, Aalto University, 2017.
- [8] V. Mäntyniemi, R. Mignot, and V. Välimäki. REMES final report: Realistic machine and environmental sounds for a training simulator to improve safety at work. Aalto Univ. publication series SCIENCE + TECHNOLOGY 16/2014.
- [9] D. B. Lloyd, N. Raghuvanshi, and N. K. Govindaraju. Sound synthesis for impact sounds in video games. In *Proc. Symp. Interactive 3D Graphics and Games (I3D)*, pages 55–62, San Francisco, CA, USA, Feb. 2011.
- [10] M. Karjalainen and H. Järveläinen. Reverberation modeling using velvet noise. In *Proc. Audio Eng. Soc. 30th Int. Conf. Intell. Audio Environ.*, Saariselkä, Mar. 2007.
- [11] B. Alary, A. Politis, and V. Välimäki. Velvet-noise decorrelator. In *Proc. Int. Conf. Digital Audio Effects (DAFx-17)*, pages 405–411, Edinburgh, UK, Sept. 2017.
- [12] J. Fagerström, S. J. Schlecht, and V. Välimäki. One-to-many conversion for percussive samples. In *Proc. Int. Conf. Digital Audio Effects (DAFx-21)*, pages 129–135, Vienna, Austria, Sep. 2021.
- [13] M. Schoeffler, S. Bartoschek, F.-R. Stöter, M. Roess, S. Westphal, B. Edler, and J. Herre. WebMUSHRA—A comprehensive framework for web-based listening tests. *J. Open Res. Softw.*, 6(1):1–8, Feb. 2018.