

NOLLAVAIHEINEN ÄÄNI

Vesa Välimäki, Roope Salmi

Aalto-yliopiston sähkötekniikan korkeakoulu
Akustiikan laboratorio, Informaatio- ja tietoliikennetekniikan laitos
Otakaari 5A, 02150 Espoo
vesa.valimaki@aalto.fi, rpsalmi@gmail.com

Tiivistelmä

Nopea Fourier-muunnos (FFT) mahdollistaa nykyisin erittäin pitkien signaalien käsittelyn, mikä on avannut uusia mahdollisuuksia äänenkäsittelyssä. Tässä artikkelissa esitellään nollavaiheinen muunnos, jossa äänen spektrin vaihekomponentti nollataan ja suoritetaan käänteinen FFT. Tuloksena on muokattu ääni, jolla on alkuperäinen taajuussisältö eli sävy, mutta eri aikarakente. Nollavaiheinen ääni on palindrominen eli se kuulostaa samalta etu- ja takaperin toistettuna. Transientit muuttuvat, mutta toistuvat rytmikuviot säilyvät. Nollatäyttö parantaa nollavaiheisen muunnoksen äänenlaatua. Nollavaiheistusmenetelmän muunnelmalla voidaan tuottaa vastaava stereosignaali. Menetelmä voidaan toteuttaa reaaliajassa jakamalla signaali lyhyisiin kehyksiin, joiden vaihe nollataan, mikä mahdollistaa live-äänenkäsittelyn. Menetelmällä ajatellaan olevan sovelluksia äänisuunnittelussa musiikki-, peli- ja elokuvatuotannossa.

1 JOHDANTO

Nopea Fourier-muunnos (*Fast Fourier Transform*, FFT) voidaan nykyään laskea hetkessä erittäin pitkästä äänisignaalista. Testissä kesällä 2025 ajoimme 1.6 miljardin pisteen FFT:n yli 10 h pituisesta äänisignaalista alle minuutissa. Tämä kehitys on johtanut uuden menetelmäperheen syntyyn, josta käytetään nimitystä jättiläis-FFT (engl. *giant FFT*). Sen aiempia sovelluksia ovat lyhyen äänen pidentäminen [1], hälypuheen tuottaminen [2] ja taajuusalueen näytetaajuusmuunnos [3][4].

Hammer ja Sundt esittelivät vuonna 1999 tietokoneohjelman nimeltä Mammut, joka muokkaa äänitiedostoa pitkällä FFT:llä, joka ei siihen aikaan voinut olla läheskään yhtä pitkä kuin nykyisin [5]. Mammut tarjoaa useita mahdollisuuksia äänenkäsittelyyn FFT-spektriä muokkaamalla. Tämän artikkelin menetelmää eniten muistuttava Mammutin tekniikka kertoo signaalin vaiheen vakiolla, joka voi olla positiivinen tai negatiivinen [5].

Tietokonemusiikin säveltäjä Theodore Burt sovelsi samantapaista menetelmää vuosikymmentä myöhemmin väitöstyössään, jossa hän tuotti musiikkia korvaamalla FFT-spektrin vaiheen vakiolla [6]. Lähdemateriaalina Burt käytti muiden tuottamaa musiikkia

Copyright ©2025 Vesa Välimäki ja Roope Salmi. 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.

ja huomasi, että musiikkimateriaalin äänensävy ja rytmi säilyivät, vaikka itse musiikki ja sen rakenne muuttuivat tunnistamattomiksi [6].

Tämän tutkimuksen lähtökohta on Stefan Bilbaon kokeiluissa, joita hän teki pitkällä FFT:llä IRCAM-keskuksessa Pariisissa vuonna 1992 [7]. Hän havaitsi mielenkiintoisen muutoksen äänessä nollattuaan FFT-spektrin vaiheen ja palautettuaan signaalin aika-alueeseen käänteis-FFT:n avulla. Tässä artikkelissa esittelemme nollavaiheistuksen periaatteen ja näytämme esimerkin pitkän musiikkiaänitteen muokkaamisesta. Menetelmä on alunperin esitelty DAFx25-konferenssissa Italiassa syyskuussa 2025 [7].

2 ÄÄNEN NOLLAVAIHEISTUS

Mikä tahansa digitaalinen äänite $s(n)$ voidaan nollavaiheistaa, kun ensin määritetään sen diskreetti Fourier-muunnos (engl. *discrete Fourier transform*, DFT):

$$S(k) = \sum_{n=0}^{L-1} s(n)e^{-j2\pi kn/L}, \quad k = 0, 1, 2, \dots, L-1, \quad (1)$$

missä n on diskreetti aika, k on diskreetti taajuus, L on signaalin $s(n)$ pituus näytteinä, e on Neperin luku ja j on imaginaariyksikkö. DFT lasketaan yleensä FFT-algoritmin avulla. Keskeistä jättiläis-FFT:n kannalta on, että signaalin pituus L voi nykyisin olla suuri luku, kuten $L = 2\,646\,000$, kun äänitteen kesto on 1 min näytetaajuudella 44,1 kHz.

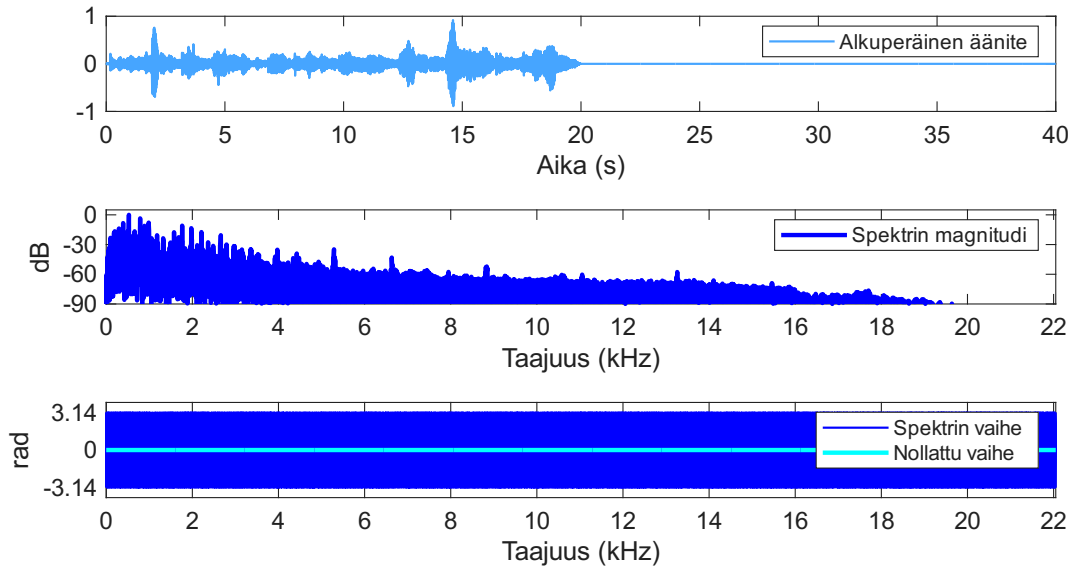
Nollavaiheistus toteutetaan käytännössä [7] laskemalla signaalin kompleksiarvoisesta spektristä $S(k)$ itseisarvo ja palauttamalla näin saatu reaaliarvoinen spektri aika-alueeseen käänteis-DFT:n avulla:

$$s_0(n) = \frac{1}{L} \sum_{k=0}^{L-1} |S(k)|e^{j2\pi kn/L}, \quad n = 0, 1, 2, \dots, L-1, \quad (2)$$

missä alaindeksi "0" viittaa nollavaiheistukseen. Tavallisesta käänteis-DFT:stä yhtälö (2) poikkeaa vain itseisarvomerkkien osalta.

Digitaaliäänitettä kannattaa pidentää nollanäytteillä (nollatäyttö) ainakin 2 kertaa alkuperäisen pituiseksi [7]. Tämä johtuu siitä, että nollavaiheistuksen jälkeen signaali leviää aika-alueessa nollahetkestä molempiin suuntiin ja pidentyy noin kaksinkertaiseksi [7]. Ilman nollatäyttöä saadaan aikalaskostunut nollavaiheinen signaali, joka myös voi sopia joihinkin tarkoituksiin.

Kuvan 1 esimerkissä (ylin) alkuperäistä äänitettä pidennetään kaksinkertaiseksi nollatäytöllä ($P = 2$). Äänite on 20 s katkelma Martin Arnoldin jousikvartetosta. Koko 40 s aika-alueen signaalista, joka sisältää 1,76 miljoonaa näytettä, lasketaan FFT-spektri. Kuvan 1 keskiosassa nähdään spektrin itseisarvo eli magnitudi ja alaosassa vaihe, joka vaihtelee arvojen $-\pi$ ja π välillä. Spektrin magnitudi sisältää tiedon äänitteen energiajakaumasta eri taajuuksilla, kun taas spektrin vaihe määrittää taajuuksien esiintymisen eri ajan hetkillä. Vaiheen yksityiskohdilla ei nollavaiheistuksessa ole merkitystä, sillä vaiheen arvot korvataan nolilla, ks. kuvan 1 alaosan sinivihreä suora viiva. Tämän jälkeen spektriin jää vain reaaliosa, joka on sama kuin alkuperäisen spektrin itseisarvo.



Kuva 1: (Ylin) Alkuperäisen äänitteen aaltomuoto nollatäytöllä (kohdasta 20 s loppuun), (keskellä) sen magnitudispektri ja (alin) vaihespektri, jonka arvot vaihtelevat välillä $[-\pi, \pi]$, sekä nollattu vaihe.

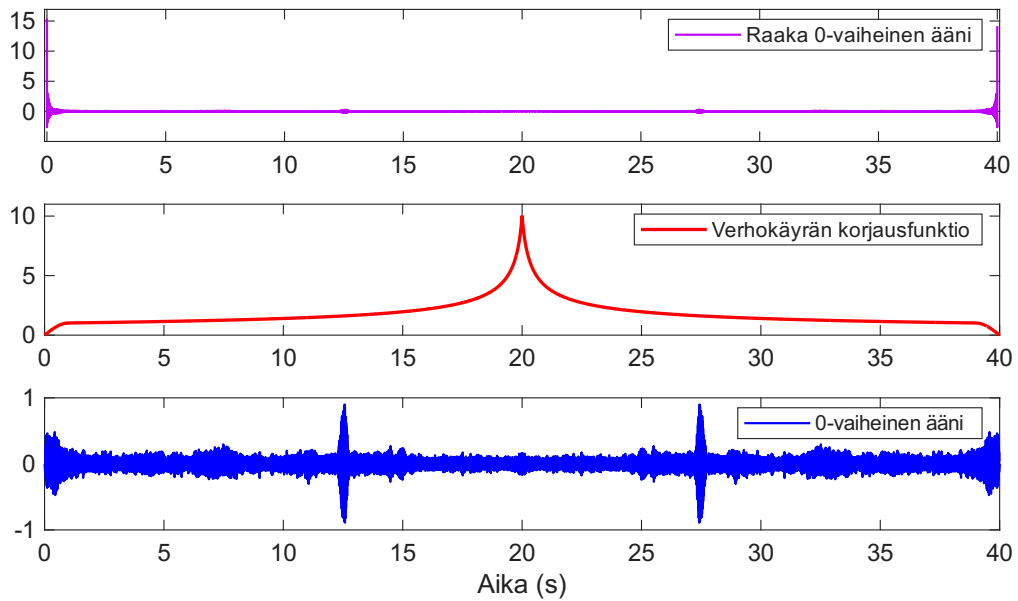
2.1 Nollavaiheisen äänen verhokäyrän korjaus

Kuvan 2 yläosa näyttää käänteismuunnoksen tuottaman signaalin $s_0(n)$, joka on raakaversio nollavaiheisesta äänestä. Tämä signaali on harvoin sellaisenaan käyttökelpoinen, koska energia keskittyy sen alkuun ja loppuun. Alku- ja loppupiikkien lisäksi signaali hiljenee keskikohdassa lähes nollaan, mikäli alkuperäiseen signaaliin tehdään nollatäyttö [7]. Hiljaisissa kohdissa on kuitenkin mielenkiintoista sisältöä, joka voidaan tuoda kuuluviin verhokäyrää säätämällä.

Alun ja lopun piikit voidaan vaimentaa neljäsosilla sini- ja kosinifunktioista. Piikkien kesto vaihtelee, mutta yleensä sopiva sisään- ja uloshäivytyksen pituus on 1 s [7]. Alku- ja loppuhäivytyksifunktiot näkyvät kuvan 2 keskimmaisessä käyrässä väleillä 0–1 s ja 39–40 s.

Keskikohdan hiljenemistä voidaan mallintaa kohinan avulla. Jos alkuperäinen signaali $s(n)$ on valkoista kohinaa, piikkien ulkopuolella raakavaiheisen signaalin $s_0(n)$ hetkellinen teho heikkenee likimain lineaarisesti [7]. Keskellä hetkellinen teho on lähes nolla. Kaksinkertaisen nollatäytön yhteydessä sopiva verhokäyrän korjaus on $1/\sqrt{p(n)} + \varepsilon$, missä $p(n)$ on paloittain lineaarinen rampifunktio, jolle $p(0) = p(2L) = 1$ ja $p(L) = 0$ [7]. Vakion ε arvoksi valitaan pieni positiivinen luku, esimerkiksi $\varepsilon = 0,01$. Kuvan 2 keskimmaisesta käyrästä nähdään, että keskikohta hetkellä 20 s vahvistetaan eniten. Jos nollatäyttöä ei tehdä, keskikohdan verhokäyrän korjausta ei tarvita.

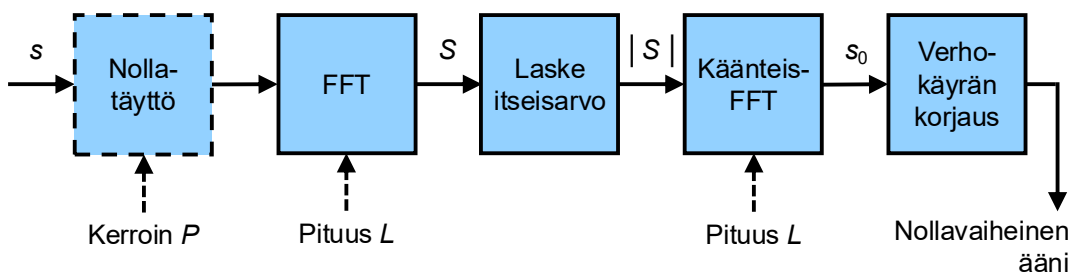
Saatu aikaverhokäyrän korjausfunktio kerrotaan pisteittäin kuvan 2 ylimmän raakasignaalin kanssa. Näin saadaan kuvan 2 alin signaali, joka on korjattu, valmis nollavaiheinen ääni. Sen sävy on hyvin samankaltainen kuin alkuperäinen, mutta tapahtumat ovat erilaisia. Nollavaiheinen signaali on palindrominen eli se on symmetrinen ajassa keski-



Kuva 2: (Ylin) Käänteis-FFT:n antama raaka nollavaiheinen ääni $s_0(n)$, (keskellä) verhoikäyrän korjausfunktio, kun $\varepsilon = 0,01$, ja (alin) valmis nollavaiheinen ääni, josta voimakkuusvaihtelut on korjattu.

pisteensä suhteen ja kuulostaa samalta etu- ja takaperin soitettuna. Tämä näkyy selvästi siitä, että kuvan 2 alin käyrä on täsmälleen symmetrinen keskikohdan 20 s suhteen. Tämä ääniesimerkki on kuunneltavissa verkkosivullamme [8].

Kuva 3 kokoaa nollavaiheistuksen operaatiot. Aluksi digitaalista tulosignaalia $s(n)$ voidaan pidentää nollatäytöllä. Parametrinarvolla $P = 2$ signaali pitenee kaksinkertaiseksi. FFT lasketaan koko signaalista, jonka pituus L eli näytteiden määrä annetaan parametrina. Tulos on kompleksiarvoinen spektri $S(k)$. Siitä määritetään jokaisen kompleksiluvun itseisarvo, joka on magnitudispektri $|S(k)|$. Kun lasketaan käänteinen FFT, aika-alueeseen palautuu signaali $s_0(n)$, joka on raakaversio nollavaiheisesta äänestä. Kun verhoikäyrän suuret vaihtelut korjataan, valmistuu kuvan 3 prosessin lähtösignaali, joka on lopullinen nollavaiheinen ääni.



Kuva 3: Nollavaiheisen äänenkäsittelyn vaiheet. Nollatäyttö on valinnainen.

3 MENETELMÄN MUUNNELMIA

Seuraavassa esittelemme kaksi nollavaiheistuksen muunnelmia. Ensimmäinen niistä mahdollistaa stereoäänien luomisen ja toinen reaaliaikaisen äänenkäsittelyn.

3.1 Nollavaiheinen stereoääni

Saamme kaksi eri nollavaiheista signaalia määrittämällä spektrin $S(k)$ reaali- ja imaginaariosista itseisarvot erikseen ja laskemalla niistä käänteis-DFT [7]:

$$s_{0,1}(n) = \frac{1}{L} \sum_{k=0}^{L-1} |\operatorname{Re}[S(k)]| e^{j2\pi kn/L} \quad (3)$$

ja

$$s_{0,2}(n) = \frac{1}{L} \sum_{k=0}^{L-1} |\operatorname{Im}[S(k)]| e^{j2\pi kn/L}, \quad (4)$$

missä $\operatorname{Re}[z]$ merkitsee kompleksiluvun z reaali-osaa ja $\operatorname{Im}[z]$ sen imaginaariosaa. Käytämme signaaleja $s_{0,1}(n)$ ja $s_{0,2}(n)$ stereosignaalin vasempana ja oikeana kanavana verhoikäyrän korjauksen jälkeen.

Vaikka alkuperäinen signaali $s(n)$ on yksikanavainen monosignaali, lopputuloksessa on miellyttävän avara stereokuva. Signaalit $s_{0,1}(n)$ ja $s_{0,2}(n)$ ovat molemmat samantapaisia mutta kohinaisempia kuin yksikanavainen versio $s_0(n)$ [7]. Erilliset tapahtumat, jotka juontuvat alkuperäisen signaalin rytmistä, esiintyvät samaan aikaan vasemmassa ja oikeassa kanavassa. Muualla kuuluva tasaisempi ääni ja kohina korreloivat kanavien välillä vaihtelevasti, mikä luo laajan stereovaikutelman [7].

3.2 Reaaliaikainen toteutus lyhyissä ikkunoissa

Menetelmää voidaan muokata reaaliaikaista äänenkäsittelyä varten käyttämällä lyhytaikaista Fourier-muunnosta. Kehysten (ikkunoiden) pituus ei nyt voi olla kovin suuri, vaan se on esimerkiksi 1 s [7]. Tämäkin on suurempi kuin lyhytaikaisessa Fourier-muunnoksessa yleensä. Kun kehysten verran äänisignaalia on vastaanotettu, se voidaan muuntaa FFT:llä taajuusalueeseen, vaihe nolataan ja kehys palautetaan aika-alueeseen käänteis-FFT:n avulla [7]. Tämän jälkeen verhoikäyrä on syytä korjata ennen kuin signaali soitetään kuulijoille. FFT-laskennan kehysten täyttämisen vuoksi nollavaiheistuksen reaaliaikainen muunnelmä aiheuttaa vähintään yhden kehysten verran latenssia eli toimii hiukan jäljessä reaaliajasta.

4 DISKUSSIO

Tässä artikkelissa näytimme yhden esimerkin äänen nollavaiheistuksesta. Verkkosivullemme on tarjolla useita ääniesimerkkejä [8], joista ilmenee kuinka alkuperäisen äänen sävy ja rytmi säilyvät nollavaiheisessa muunnoksessa. Puheäänitteestä syntyy puheen kaltaista ”pulinaa”, jossa sanoista ei saa selvää, kun taas useiden puhujien yhdistäminen tuottaa ”cocktail-kutsukeskustelua”. Nollavaiheistettu laulu muistuttaa usein ääretöntä kaikua. Verkkosivulla on myös kappaleen 3 stereo- ja ikkunointimuunnelmien avulla tuotettuja ääniesimerkkejä [8].

Nollavaiheistuksen avulla äänitteistä voidaan tehdä korkealaatuisia muunnelmia, millä on sovelluksia luovilla aloilla. Musiikkikappaleen materiaalista voidaan tuottaa nollavaiheisia versioita ja käyttää osana miksausta. Tietokonepeleissä, tv-sarjoissa ja elokuvissa kaivataan jännittäviä tehosteääniiä, joskus pitkiäkin. Nollavaiheisen äänen etuja näitä sovelluksia ajatellen ovat yksinkertaisen äänenkäsittelyn lisäksi korkea äänenlaatu, jossa ei ole napsahduksia, säröä tai muita ylimääräisiä häiriöitä, ja mahdollisuus tuottaa hyvinkin pitkiä äänitiedostoja.

5 YHTEENVETO

Artikkelimme kuvaa menetelmän, jossa alkuperäisestä äänestä lasketaan pitkä FFT ja sen magnitudispektristä tehdään käänteis-FFT. Tuloksena saadaan nollavaiheinen signaali, joka on reaaliarvoinen ja symmetrinen – eli se muodostaa akustisen palindromin. Äänen alkuun ja loppuun suosittelemme häivytystä, koska ne ovat voimakkaimpia kohtia. Nollatäyttö ennen FFT-operaatiota pidentää ääntä, esimerkiksi tekijällä 2, mutta lopputuloksen keskiosa hiljenee. Tämä voidaan kompensoida korjausfunktiolla. Nollavaiheistuksessa tarvittava laskenta on hyvin yksinkertaista ja nopeaa nykyisillä tietokoneilla.

VIITTEET

- [1] V. Välimäki, J. Rämö ja F. Esqueda. Creating endless sounds. In *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, s. 32–39, Aveiro, Portugal, syyskuu 2018.
- [2] A. Kuusinen ja T. Lokki. Kohina, hälinät ja kaiunnat – Hälypuhetestien vanhat ja uudet peittoäänät. In *Akustiikkapäivät*, s. 277–284, Turku, marraskuu 2021.
- [3] V. Välimäki ja S. Bilbao. Giant FFTs for sample rate conversion. *Journal of the Audio Engineering Society*, 71(3):88–99, maaliskuu 2023.
- [4] V. Välimäki ja S. Bilbao. Näytetaajuuden vaihto nopealla Fourier-muunnoksella. In *Akustiikkapäivät*, s. 103–108, Tampere, marraskuu 2023.
- [5] Ø. Hammer ja H. Sundt. Musical applications of decomposition with global support. In *Proceedings of the Second COST-G6 Workshop on Digital Audio Effects (DAFx)*, Trondheim, Norja, joulukuu 1999.
- [6] T. Burt. Composition folio. Väitöskirja, Music Research Centre, University of York, York, Yhdistynyt kuningaskunta, syyskuu 2012.
https://etheses.whiterose.ac.uk/id/eprint/3817/1/TheoBurt_Ph D_Commentary.pdf
- [7] V. Välimäki, R. Salmi, S. Bilbao, S. J. Schlecht ja D. Zicarelli. Zero-phase sound via giant FFT. In *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, s. 290–297, Ancona, Italia, syyskuu 2025.
- [8] V. Välimäki, R. Salmi, S. Bilbao, S. J. Schlecht ja D. Zicarelli. Verkkoliite julkaisuun ”Zero-phase sound via giant FFT”. Tarkistettu 14.10.2025.
<http://research.spa.aalto.fi/publications/papers/dafx25-zero-phase/>