

Teknillinen Korkeakoulu
Sähkö- ja tietoliikennetekniikan osasto

Jutta Katariina Mahkonen

**Äänen laadun parantaminen
puheensirrossa keinotekoisella
taajuuskaistan laajennuksella**

Diplomityö, joka on jätetty opinnäytteenä tarkastettavaksi diplomi-insinöörin
tutkintoa varten Espoossa 13.12.1999.

Työn valvojana ja ohjaajana toimi prof. Paavo Alku.

Teknillinen korkeakoulu

DIPLOMITYÖN TIIVISTEMÄ

| | |
|-------------|---|
| Tekijä: | Jutta Katariina Mahkonen |
| Työn nimi: | Äänen laadun parantaminen puheensierrossa keinotekoisella taajuuskaistan laajenuksella |
| Päivämäärä: | 13.12.1999 |
| Sivumäärä: | 88 |

| | |
|--------------|---|
| Osasto: | Sähkö- ja tietoliikennetekniikan osasto |
| Professuuri: | Akustiikka ja äänenkäsittelytekniikka, S-89 |

| | |
|---------------|------------------|
| Työn valvoja: | prof. Paavo Alku |
|---------------|------------------|

Tässä diplomityössä kehitettiin uusi menetelmä 300 – 3400 Hz:n taajuudet sisältävän puhelinäänen taajuuskaistan laajentamiseksi. Menetelmä on suunniteltu toimimaan jatkuva-aikaisesti ja periaatteena oli algoritmin toimintavarmuus kaikissa häiriöolosuhteissa. Taajuuskaistan laajennus toteutetaan menetelmässä hallitun kaistalaajennuksen ja adaptiivisen vaimennusalgoritmin avulla.

Työssä on myös esitelty puhelinkaistan keinotekoiseen laajentamiseen aiemmin suunniteltuja menetelmiä. Menetelmien toimintaperiaatteen ymmärtämisen helpottamiseksi on aluksi käsitelty lyhyesti niitä digitaalisen signaalinkäsittelyn algoritmeja, joita voidaan hyödyntää keinotekoisien taajuuskaistan laajennuksen toteuttamisessa.

Erilaisia toimintatapoja äänen laadun arvioimiseksi on tutkittu, ja selvitetty kehitetyn taajuuskaistan laajennusmenetelmän äänenlaadun parannuskykyä subjektiivisin kuuntelukokein.

Avainsanat:

puheenkäsittely, taajuuskaistan laajennus, subjektiivinen kuuntelukoe, PCM, LPC

Helsinki University of Technology

ABSTARACT OF THE MASTER'S THESIS

| | |
|--|--|
| Author: | Jutta Katariina Mahkonen |
| Name of the Thesis: | Improving speech quality by artificial bandwidth expansion |
| Date: | 13.12.1999 |
| Number of pages: | 88 |
| Faculty: | Department of Electrical and Communications Engineering |
| Professorship: | Acoustics and Audio Signal Processing |
| Supervisor: | prof. Paavo Alku |
| <p>In this study a method for the artificial expansion of the frequency band of a voice signal with 300 – 3400 Hz spectrum was developed. The method is designed to operate in real time, and reliability of the method in all circumstances was used as a guideline in design process. The expansion of the frequency band is carried out with controlled aliasing and an adaptive attenuating algorithm.</p> <p>Also some previously designed methods for expanding the frequency band have been studied. For better understanding of operations in these methods some basic digital signal processing procedures are discussed briefly.</p> <p>Different ways to evaluate the speech quality have been studied, and the ability of the designed method to enhance the quality of a speech signal has been examined with subjective listening tests.</p> | |
| Keywords: | speech processing, frequency band expansion, subjective listening test, PCM, LPC |

ALKULAUSE

Olen tehnyt tämän diplomityön Teknillisen korkeakoulun sähkö- ja tietoliikennetekniikan osastolla Akustiikan ja äänenkäsittelytekniikan laboratoriossa. Työn teettäjänä ja ohjaajanani oli prof. Paavo Alku. Hänelle kuuluvat kiitokset selkeästä tehtäväkuvauksesta ja ystävällisistä ohjeista työvaiheiden suoritukseen.

Haluan myös kiittää miestäni Marko Mahkosta työn aikana saamastani henkisestä tuesta ja käytännön ohjeista.

13.12.1999

Jutta Mahkonen

SISÄLLYSLUETTELO

| | |
|---|------------|
| ALKULAUSE | IV |
| SISÄLLYSLUETTELO | V |
| SYMBOLI- JA LYHENNELUETTELO | VII |
| 1 JOHDANTO | 1 |
| 2 TAAJUUSKAISTAN LAAJENNUKSESSA HYVÄKSI KÄYTETTÄVIÄ SIGNAALINKÄSITTELYMENETELMIÄ | 4 |
| 2.1 Näytetaajuusmuunnokset | 4 |
| 2.1.1 Alinäytteistys | 5 |
| 2.1.2 Ylinäytteistys | 5 |
| 2.1.3 Näytteenottotaajuuden muuttaminen muulla kuin kokonaislukukertoimella | 7 |
| 2.2 Hallittu laskostaminen | 7 |
| 2.2.1 Toteutus aika-alueessa | 8 |
| 2.2.2 Spektrin käyttäytyminen | 8 |
| 2.3 Ikkunointi | 9 |
| 2.3.1 Ikkunoinnin vaikutukset | 10 |
| 2.3.2 Ikkunafunktion valinta | 11 |
| 2.4 Lineaarinen ennustus | 15 |
| 2.4.1 Signaalin LP-suodatus | 15 |
| 2.4.2 Virhesignaalin ominaisuuksia | 16 |
| 2.4.3 LP-kerrointen laskenta | 16 |
| 2.4.4 Puheen tuottaminen LP-parametrien avulla | 17 |
| 2.5 Puhesignaalin koodaus | 19 |
| 2.5.1 Pulssikoodimodulaatio | 19 |
| 2.5.2 CELP | 21 |
| 3 ÄÄNEN LAADUN SUBJEKTIIVINEN TESTAUS | 23 |
| 3.1 Mitä on äänen laatu? | 23 |
| 3.1.1 Objektiivinen äänenlaadun mittaust | 24 |
| 3.1.2 Subjektiivinen äänenlaatu | 25 |
| 3.2 Kuuntelukokeiden muuttujia | 26 |
| 3.2.1 Kuuntelijoiden valinta | 26 |
| 3.2.2 Koeympäristö | 27 |
| 3.2.3 Testattavien laitteiden tai algoritmien valinta | 28 |
| 3.2.4 Testissä käytettävä äänimateriaali | 28 |
| 3.2.5 Arvioitavien ominaisuuksien valinta | 29 |

| | | |
|------------|--|------------------|
| 3.2.6 | Kuuntelukoetilanne | 30 |
| 3.3 | Testausmenetelmiä | 30 |
| 3.3.1 | Sokkotestit | 31 |
| 3.3.2 | Yksittäiskuuntelukoe | 31 |
| 3.3.3 | Parivertailutestit | 32 |
| 3.3.4 | Arviointiasteikkoja | 32 |
| 3.4 | Kuuntelukoetulosten analysointi | 35 |
| 4 | <u>TUNNETTUJA TAAJUUSKAISTAN LAAJENNUSMENETELMIÄ</u> | <u>37</u> |
| 4.1 | Perinteiseen signaalinkäsittelyyn perustuvia taajuuskaistan laajennusmenetelmiä | 37 |
| 4.1.1 | Laskostamiseen ja suodatukseen perustuva menetelmä | 38 |
| 4.1.2 | Tasasuuntausta hyväksi käyttävä menetelmä | 40 |
| 4.1.3 | LPC-kerrointen ekstrapolointiin perustuva menetelmä | 41 |
| 4.1.4 | Menetelmien ongelmia | 42 |
| 4.2 | Puheen tilastollisiin ominaisuuksiin ja koodikirjoihin perustuvat menetelmät | 43 |
| 4.2.1 | Tilastollinen menetelmä, joka olettaa yläkaistan joukoksi satunnaislähteitä | 43 |
| 4.2.2 | LP-analyysiin ja koodikirjan käyttöön perustuva menetelmä | 45 |
| 4.2.3 | CELP-koodatun äänen laadunparannus monikaistaisen herätemallin avulla | 47 |
| 4.2.4 | Pohdintaa | 49 |
| 5 | <u>TOTEUTETTU TAAJUUSKAISTAN LAAJENNUSMENETELMÄ</u> | <u>50</u> |
| 5.1 | Algoritmin kuvaus | 50 |
| 5.2 | Menetelmän toteutus Matlab-ohjelmistolla | 54 |
| 5.2.1 | Funktiot | 54 |
| 5.2.2 | Muuttujien oletusarvot | 56 |
| 6 | <u>KEHITETYN MENETELMÄN ÄÄNEN LAADUN PARANNUSKYVYN MÄÄRITYS KUUNTELUKOKEIN</u> | <u>58</u> |
| 6.1 | Ensimmäinen suuntaa antava kuuntelukoe | 58 |
| 6.1.1 | Kokeessa käytetyt äänet | 58 |
| 6.1.2 | Tulosten analysointi | 59 |
| 6.2 | Toinen, perusteellisempi kuuntelukoe | 62 |
| 6.2.1 | Kokeessa käytetyt lauseet | 62 |
| 6.2.2 | Tulosten analysointi | 63 |
| 6.2.3 | Yhteenveto ja pohdintaa | 65 |
| 7 | <u>YHTEENVETO</u> | <u>67</u> |
| | <u>LÄHTEET</u> | <u>69</u> |
| | <u>LIITE 1</u> | <u>72</u> |

SYMBOLI- JA LYHENNELUETTELO

| | | |
|-------------------|---|--|
| Matlab | | matemaattinen laskentaohjelma |
| Nyqvistin taajuus | | puolet digitaalisen signaalin näytteenottotaajuudesta |
| perkeptuaalinen | | ihmiskuulon ominaisuuksiin perustuva |
| wav | | tiedostotyyppi äänitiedostoille |
| ACELP | algebraic code exited linear prediction | CELP-koodaukseen perustuva puheenkoodaus-algoritmi |
| CD | compact disc | CD-levy |
| CELP | code exited linear prediction | koodikirjan avulla toteutettava lineaariseen ennustukseen perustuva puheenkoodausalgoritmi |
| CMOS | comparison mean opinion score | vertaileva MOS-pisteytys |
| CS-ACELP | conjugate structure algebraic code exited linear prediction | ITU-T standardin G.729 mukainen CELP-koodaukseen perustuva puheenkoodausalgoritmi |
| dB | desibeli | logaritminen mittayksikkö |
| DCR | degradation category rating | huonontuman luokitus |
| DMOS | degradation mean opinion score | huonontuman MOS-pisteytys |
| DSP | digital signal processor | digitaalinen signaaliprosessori |
| EFR | enhanced full rate | GSM-järjestelmässä käytety koodekkityyppi |
| FFT | fast Fourier transformation | nopea Fourier-muunnos |
| FIR | finite impulse response | äärellisen impulssivasteen suodatin |
| F _s | sampling frequency | näytteenottotaajuus |
| GSM | global system for mobile communications | maailmanlaajuinen digitaalinen matkapuhelinjärjestelmä |
| IFFT | inverse FFT | käänteinen FFT |

| | | |
|---------|--|--|
| INRS | l'institut national de la recherche scientifique | Quebecin yliopiston kansallinen tieteellinen tutkimusinstituutti |
| IEC | international electrical committee | sähkötekniikan alan kansainvälinen standardointijärjestö |
| ITU-T | telecommunications standardization sector of the international telecommunication union | tele-alan kansainvälinen standardisointijärjestö |
| LPC, LP | linear predictive coding, linear prediction | lineaarinen ennustus |
| LSP | line spectral pair | viivaspektripari |
| MBE | multiband excitation | monikaistainen heräte |
| MOS | mean opinion score | mielipiteen mittausasteikko |
| MOS-LE | MOS for loudness preference | mielipiteen mittausasteikko äänenvoimakkuudelle |
| MOS-LP | MOS for listening effort | mielipiteen mittausasteikko kuuntelemisen työläydelle |
| MSE | mean squared error | neliösumman keskiarvo |
| NTT | Nippon telegraph and telephone corporation | japanilainen teleyhtiö |
| PC | personal computer | henkilökohtainen tietokone |
| PCM | pulse code modulation | pulssikoodimodulaatio |
| TAFC | two alternative forced choice | pakotettu valinta kahdesta vaihtoehdosta |

1 JOHDANTO

Äänen taajuuskaistan keinotekoinen laajennus tarkoittaa sitä, että ääni, jonka taajuuskaista on jossakin vaiheessa jouduttu kaventamaan, halutaan takaisin laajakaistaiseksi. Nykyisin olemme tottuneet erittäin korkealaatuisiin ääniin modernien ja digitaalisten äänentoistolaitteiden myötä, joten kovin kapeakaistainen ääni kuulostaa meistä helposti tukkoiselta.

Ääntä tallennettaessa ja siirrettäessä digitaalisesti asettaa näytteenottotaajuus rajan äänen taajuuskaistalle, sillä Nyqvistin näytteenottoeteoreeman [Carlson 1986] mukaan äänen taajuuskaista voi sisältää taajuuksia nolasta korkeintaan puoleen näytteenottotaajuuteen asti. Digitaalista ääntä tallennettaessa on ajateltava äänen viemää muistitilaa esimerkiksi CD-levyllä tai tietokoneen muistissa. Tällöin näytteenottotaajuutta ei voida rajattomasti nostaa, koska tallennukseen tarvittava muistitila kasvaa samassa tahdissa. Digitaalista ääntä siirrettäessä linjojen siirtokapasiteetti asettaa erittäin tiukkoja rajoja siirrettävän tiedon määrälle. Tätä tiedonsiirron ongelmaa on näytteenottotaajuuden pienentämisen lisäksi kierretty kehittämällä äänenkoodausalgoritmeja, jotka minimoivat siirrettävän tiedon määrää.

Osa äänenkoodausalgoritmeista säilyttää kaikki äänen ominaisuudet, jolloin algoritmin käyttö signaalin siirrossa ei millään tavalla huononna äänen laatua. Tällainen on esimerkiksi PCM [Owen 1982], pulssikoodimodulaatiokoodaus, joka on perustana perinteisessä puhelimen välittämässä 300 – 3400 Hz:n taajuudet sisältävän äänen siirrossa. Toiset äänenkoodausalgoritmit sen sijaan hävittävät osan äänen vähempiarvoisista ominaisuuksista ja niin pystyvät pienentämään siirrettävän tiedon määrää huomattavasti tehokkaammin kuin tietoa hävittämättömät algoritmit. Yleisimmin käytetty tietoa hävittävä äänenkoodausalgoritmi on CELP (code excited linear prediction). Siihen perustuvaa ACELP (algebraic code excited linear prediction) – algoritmia [ETSI 1997] käytetään esimerkiksi GSM-järjestelmän EFR (enhanced full rate)-koodausmenetelmässä.

Digitaalisen puhesignaalin näytteenottotaajuus vakiintui 8 kHz:iin ensimmäisen digitaalisen puheenkoodausmenetelmän, pulssikoodimodulaation eli PCM:n myötä. 8 kHz:n näytetaajuus merkitsee sitä, että korkein mahdollinen puhelinlinjan läpäisevä ääni olisi 4 kHz. Käytännössä analogisella suotimella saatiin korkeille taajuuksille riittävä vaimennus aikaan, kun suodin alkoi vaimentaa signaalia jo 3.4 kHz:n kohdalla. Näistä historiallisista syistä puhelimen välittämä taajuuskaista on nykyisellään 300 – 3400 Hz.

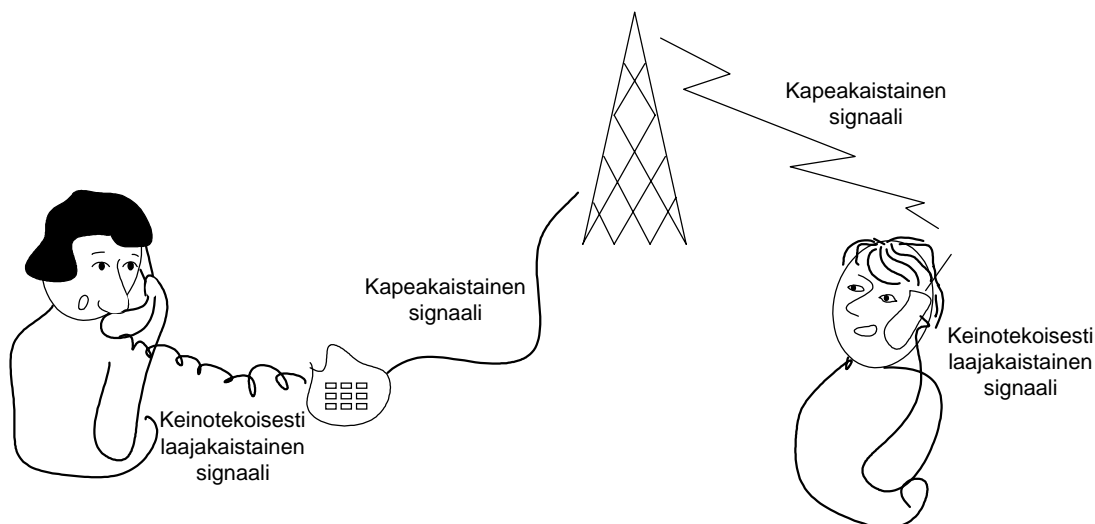
Luonnollisen puheen taajuuskaista on kuitenkin moninkertainen edellä mainittuun puhelinkaistaan verrattuna. Kokemuksemme mukaan puhelimen välittämä ääni ei ole yhtä kirkas kuin luonnollinen ääni, mikä johtuu juuri korkeiden taajuuksien puuttumisesta. Puhelinlinjojen läpi tuleva puhe on ymmärrettävää ja useimmiten puhuja

on tunnistettavissa äänen perusteella. Puutteita äänen laadussa on helppo huomata kuitenkin mm. frikatiivien, kuten /s/:n tai /f/:n välittymisessä tai soitettaessa musiikkia puhelinlinjan läpi.

Tässä diplomityössä on tutkittu mahdollisuuksia puhelimen välittämän äänen taajuuskaistan laajentamiseksi keinotekoisesti, jotta siirretyn puheäänien laatua voitaisiin parantaa muuttamatta tietoliikenneyhteyden, lähettimen tai siirtotapahtuman toimintoja. Taajuuskaistan laajennus on toteutettava päätelaitteessa keinotekoisesti, koska siirtolinjaa ei haluta kuormittaa ylimääräisellä laajempaa taajuuskaistaa koskevalla informaatiolla. Kuten kuva 1 havainnollistaa, kulkee puhe siirtolinjalla kapeakaistaisena. Vasta päätelaitteessa suoritetaan taajuuskaistan laajennus, joka tuottaa ääneen korkeataajuisia komponentteja ja näin kirkastaa ääntä ja parantaa sen laatua.

Mahdollisuuksia äänen taajuuskaistan laajennukseen keinotekoisesti on tutkittu jonkin verran aiemminkin. Tunnettuja taajuuskaistan laajennusmenetelmiä käsittelevässä luvussa 4 esitellään muutamia tällaiseen äänen laadun parantamiseen keskittyviä ehdotelmia. Pääosin esiteltävät menetelmät suuntautuvat PCM-koodatun puheen taajuuskaistan laajentamiseen. Mukana on kuitenkin yksi menetelmä, joka keskittyy parantamaan erityisesti CELP-koodatun äänen laatua. Toiset menetelmistä käyttävät hyväkseen perinteisen signaalinkäsittelyn keinoja ja toisissa menetelmissä oikeanlainen ylempi taajuuskaista etsitään tilastollisesti opettujen koodikirjojen avulla.

Tässä diplomityössä kehitetty menetelmä keskittyy juuri PCM-koodatun puheen äänen laadun parantamiseen ja luottaa perinteisen signaalinkäsittelyn toimintavarmuuteen. Periaate menetelmän kehitystyössä on ollut luoda algoritmi, joka parantaisi äänen laatua havaittavasti, eikä missään häiriöolosuhteissa pilaisi ääntä. Menetelmän on oltava luotettava kaikille äänneille ja puhujille.



Kuva 1. Puhelinsovelluksessa puheen taajuuskaistan laajennus suoritetaan päätelaitteen vastaanottavassa osassa esimerkiksi signaaliprosessorilla. Siirtotiellä puhesignaali kulkee kapeakaistaisena.

Jotta algoritmi toimisi reaaliaikaisesti, on signaali käsiteltävä riittävän lyhyinä paloina, jotta näytteiden puskuroinnista ei aiheudu häiritsevää viivettä. Lisäksi algoritmin vaatimat laskentaoperaatiot on kyettävä suorittamaan kerralla käsiteltävän näytekehyksen mittaisessa ajassa, jotta uusia näytteitä pystytään tuottamaan vähintään tulevien näytteiden taajuudella.

Tässä diplomityössä kehitetty menetelmä on suunniteltu siten, että se on sellaisenaan toteutettavissa DSP-prosessorille tai ASIC-piirille ja kykenee reaaliaikaisen signaalin käsittelyyn. Tämän työn puitteissa menetelmä on kuitenkin toteutettu PC-tietokoneessa Matlab-ohjelmistolla, jolloin se ei luonnollisestikaan toimi reaaliajassa. Kehitetty taajuuskaistan laajennusalgoritmi käyttää hyväkseen spektrin hallittua laskostamista, mikä luo ylemmät taajuuskomponentit ja samalla nostaa näytteenottotaajuuden 16 kHz:iin. Ylempien taajuuksien jatkokäsittely on toteutettu adaptiivisesti siten, että kunkin signaalinpätkän yläkaista muotoillaan yksilöllisesti.

Jotta keinotekoisien kaistalaajennuksen onnistumista voitaisiin arvioida, on menetelmän äänen laadun parannuskykyä testattava jollakin tavalla. Tavallisin keino äänen laadun arviointiin on subjektiivinen kuuntelukoe. Työn puitteissa on tutkittu kuuntelukokeiden järjestämistä yleisesti, ja kehitetyn menetelmän äänen laadun parannuskykyä on selvitetty subjektiivisin kuuntelukokein.

2 TAAJUUSKAISTAN LAAJENNUKSESSA HYVÄKSI KÄYTETTÄVIÄ SIGNAALINKÄSITTELYMENETELMIÄ

Tässä kappaleessa esittelen niitä digitaalisen signaalinkäsittelyn osa-alueita, joita kehitetyssä taajuuskaistan laajennusmenetelmässä käytetään hyväksi.

Hallittu laskostaminen, joka perustuu uudelleennäytteistykseen, muodostaa taajuuskaistan laajennusalgoritmile perustan. Siksi aluksi käsitellään uudelleennäytteistystä ja sen vaikutuksia signaalin spektriin. Tämän jälkeen tutustutaan hallitun laskostamisen periaatteeseen.

Reaaliaikaisessa signaalinkäsittelyssä joudutaan aina ikkunoimaan signaali ennen varsinaista käsittelyä. Jatkuvasta tietovirrasta on jonkin ikkunafunktion avulla otettava sopivan mittaisia paloja käsiteltäväksi, ja käsittelyn jälkeen nuo palat on sovitettava yhteen siten, että ikkunointiprosessi vaikuttaisi käsiteltävään aika-alueen signaaliin mahdollisimman vähän. Siksi tässä vertaillaan erilaisten ja eri mittaisten ikkunafunktioiden vaikutuksia alkuperäiseen signaaliin.

Lineaarinen ennustus (LP, linear prediction) on digitaalisen signaalin käsittelymenetelmä, jota käytetään laajasti hyväksi äänenkäsittelyssä. Lineaarinen ennustus muodostaa perustan monille puheenkoodausalgoritmeille ja sitä on käytetty hyväksi myös kahdessa luvussa 3 esiteltävistä taajuuskaistan laajennusalgoritmeista.

Puheenkoodaus koskee taajuuskaistan laajennusta siltä osin, että suurin osa taajuuskaistan laajennusalgoritmeista on suunniteltu puhelinsovelluksiin. Puheenkoodausalgoritmeista tässä kappaleessa esitellään lyhyesti PCM ja CELP.

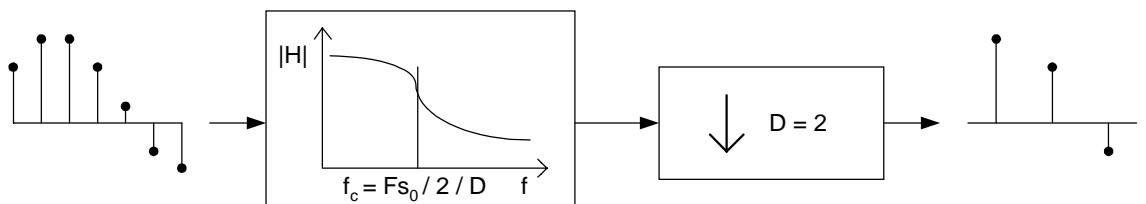
2.1 Näytetaajuusmuunnokset

Uudelleennäytteistys tarkoittaa sitä, että digitaalisen signaalin näytteenottotaajuutta muutetaan. Signaali voidaan alinäytteistää, mikä merkitsee näytteenottotaajuuden madaltamista, tai ylinäytteistää, missä näytteenottotaajuutta nostetaan. Uudelleennäytteistystä tarvitaan mm. alkuperäisen näytteenottotaajuuden ollessa sopimaton laitteelle, mihin signaalia pitäisi syöttää.

2.1.1 Alinäytteistys

Aika-alueessa signaalin alinäytteistäminen eli näytteenottotaajuuden madaltaminen tarkoittaa sitä, että osa näytteistä poistetaan. Esimerkiksi kun alinäytteistyskerroin $D = 2$, kuten kuvassa 2 eli puolitettaessa näytteenottotaajuutta, poistetaan joka toinen näyte. Vastaavasti pudotettaessa näytteenottotaajuus kolmasosaan, jolloin $D = 3$, jätetään joka kolmas näyte jäljelle.

Taajuusalueessa näytteenottotaajuuden laskeminen aiheuttaa sen, että kaikki alkuperäisen näytteenottotaajuuden välein esiintyvät spektrin monikerrat laskeutuvat matalammille taajuuksille uuden näytteenottotaajuuden mukaisesti. Näytteenottotaajuutta madallettaessa myös Nyqvistin taajuus, eli näytteenottotaajuuden puolikas laskee. Jos alkuperäisessä signaalissa on uutta Nyqvistin taajuutta korkeampia taajuuksia, ne laskostuvat alkuperäisen spektrin päälle aivan kuten yritettäessä näytteistää laajakaistaista signaalia liian matalalla näytteenottotaajuudella. Laskostumisen estämiseksi on signaali alipäästösodatettava ennen näytteiden poistoa siten, että uuden Nyqvistin taajuuden yläpuolella olevat spektrikomponentit vaimenevat riittävästi.



Kuva 2. Alinäytteistyksessä on ensin suodatettava tulevaa Nyqvistin taajuutta korkeammat spektrikomponentit pois, jonka jälkeen ylimääräiset näytteet voidaan poistaa.

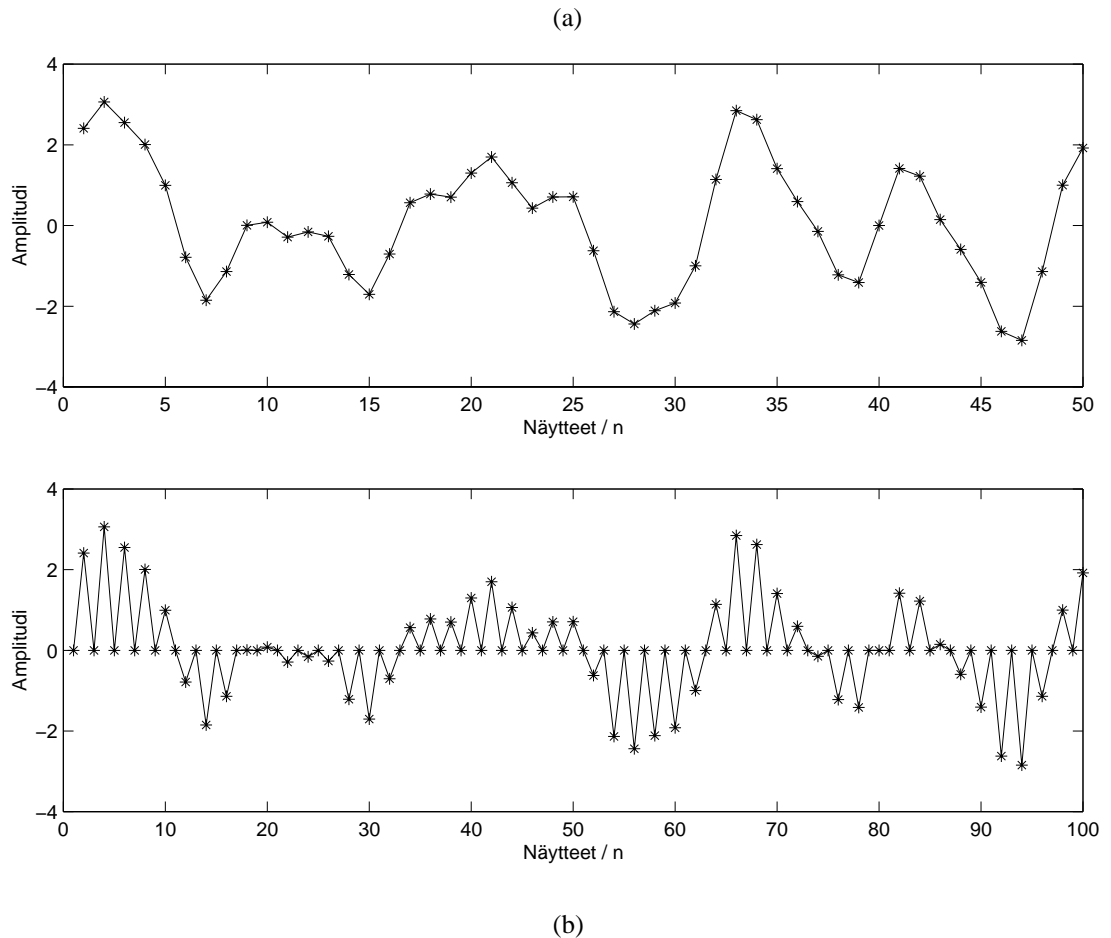
2.1.2 Ylinäytteistys

Ylinäytteistys eli näytteenottotaajuuden korottaminen toteutetaan siten, että olemassa olevien signaalinäytteiden väliin lisätään uusia näytteitä tarvittava määrä. Lisättyjen näytteiden arvot lasketaan interpoloimalla alkuperäisten näytteiden avulla. Interpolointi voidaan määrittellä aika-alueessa monella tavalla.

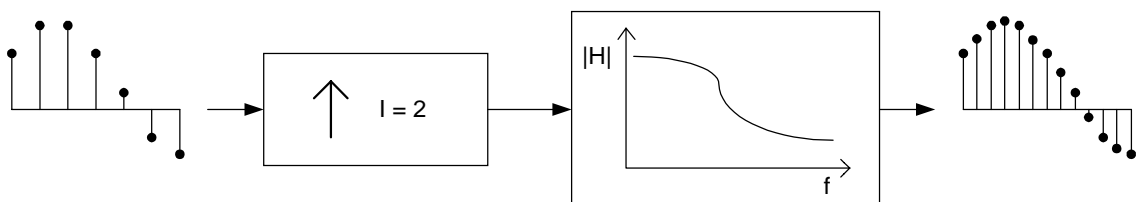
Signaalin näytteenottotaajuus nostetaan signaalinkäsittelyn keinoin siten, että aluksi lisätään näytteiden väliin riittävä määrä uusia nolla-arvoisia näytteitä, kuten kuvassa 3 on tehty ylinäytteistyskertomella $I = 2$. Kaikki aika-alueen interpoloinnin määrittelyt johtavat alipäästösodatukseen, joten nolla-arvoisten näytteiden lisäyksen jälkeen uusi signaali alipäästösodatetaan. Tämä alipäästösodatusta toteuttaa välinäytteiden interpoloinnin. Kuva 4 esittää lohkokaaaviona ylinäytteistys-

työvaiheet.

Taajuusalueessa nollien lisääminen näytteiden väliin merkitsee sitä, että signaalin näytteenottotaajuus nousee, mutta alkuperäisen näytteenottotaajuuden mukaiset spektrin monikerrat jäävät paikoilleen. Näin useita spektrikuviota jää – riippuen ylinäytteistyskertoimesta I – uuden Nyqvistin taajuuden alapuolelle. Hallitun laskostamisen tapauksessa näin laskostuneita taajuuksia käytetään hyödyksi, mutta tavanomaisessa näytteenottotaajuuden nostamisessa ne suodatetaan pois.



Kuva 3. (a) Alkuperäinen signaali aika-alueessa. (b) Signaali, jossa joka näytteen jälkeen on lisätty nolla-arvoinen näyte.

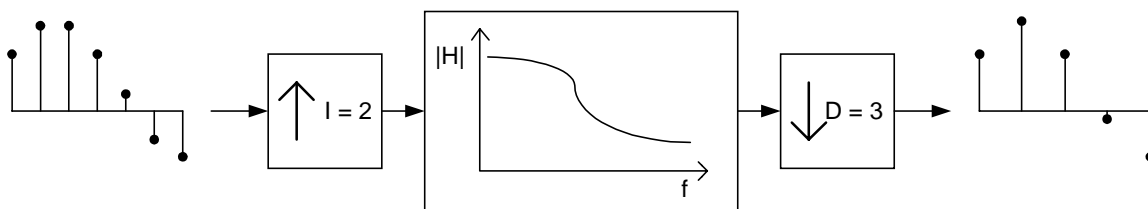


Kuva 4. Ylinäytteistyksessä lisätään nolla-arvoisia näytteitä jokaisen signaalinäytteen perään, jonka jälkeen näytearvojen interpolointi toteutetaan alipäästösuotimella.

2.1.3 Näytteenottotaajuuden muuttaminen muulla kuin kokonaislukukertoimella

Edellä esitellyt ylinäytteistys ja alinäytteistys toteuttavat näytteenottotaajuuden muunnoksen ainoastaan kokonaislukukertoimella. Kun näytteenottotaajuutta halutaan muuttaa jonkin murtolukukertoimen mukaisesti, on suoritettava peräkkäin ylinäytteistys ja alinäytteistys – tässä järjestyksessä, jotta tietoa häviäisi mahdollisimman vähän.

Ylinäytteistyksessä lisättyjen nollien interpoloinnin toteuttava alipäästösuodatus ja alinäytteistystä varten suoritettava laskostumista estävä alipäästösuodatus voidaan yhdistää yhdellä suotimella toteutettaviksi, kuten kuvan 5 lohkokaaviossa on esitetty. Tällöin suotimen päästökaistan rajataajuuden määrää joko alkuperäisen näytteenottotaajuuden mukainen Nyqvistin taajuus tai lopullisen uuden signaalin Nyqvistin taajuus, riippuen siitä, kumpi näistä on matalampi.



Kuva 5. Näytetaajuuden muuttaminen kertoimella $2/3$. Ensin suoritetaan näytteiden lisäys, minkä jälkeen signaalista suodatetaan uutta Nyqvistin taajuutta korkeammat spektrikomponentit pois. Lopuksi poistetaan välinäytteet.

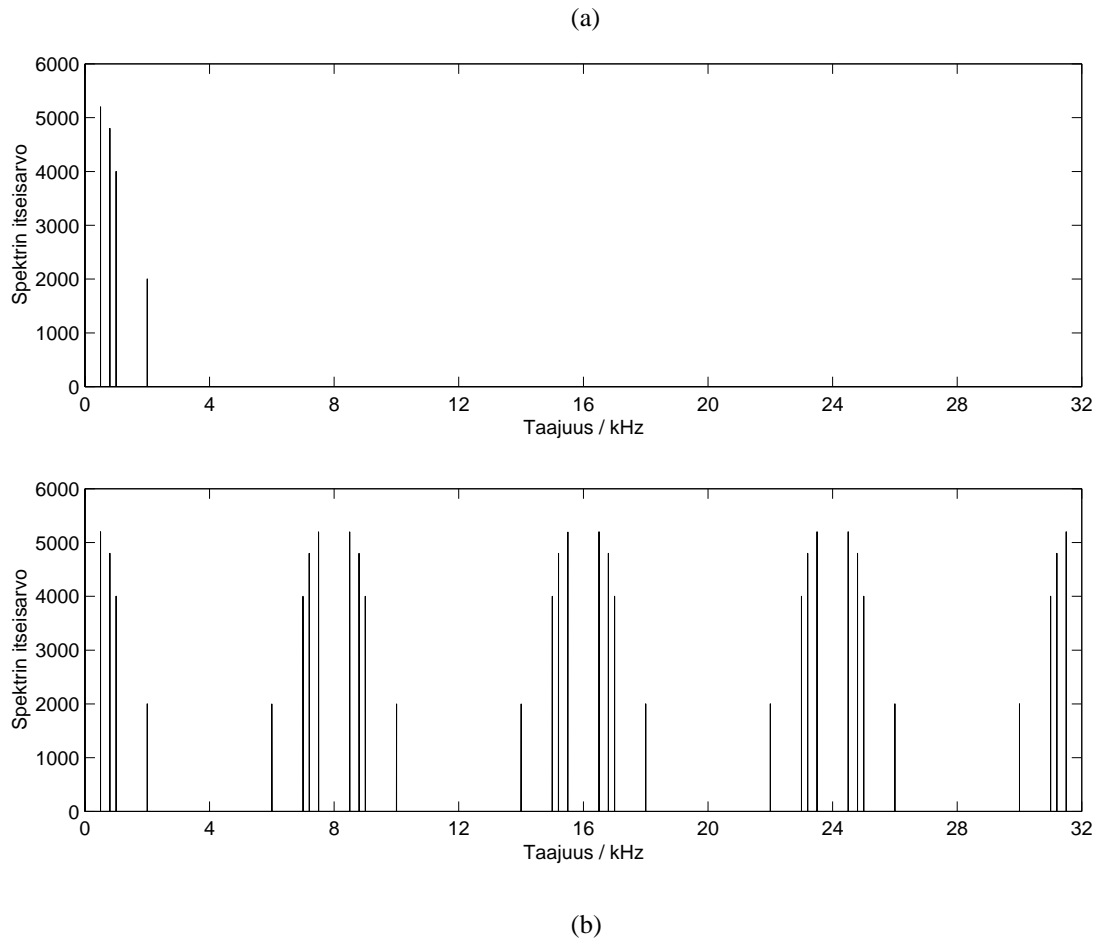
2.2 Hallittu laskostaminen

Hallittu laskostaminen perustuu siihen näytteistetyn signaalin ominaisuuteen, että alkuperäisen signaalin spektrikuviot on kopioitunut näytteistykseen yhteydessä näytteenottotaajuuden välein toistuvaksi jatkumoksi näytteistetyn signaalin spektriin kuvan 6 mukaisesti. Hallittu laskostaminen käyttää hyväkseen näitä toistuvia spektrikuviota, jotka tavallisesti koetaan haitallisina.

Tavallisesti uudelleennäytteistykseen yhteydessä, sekä ylinäytteistyksessä että alinäytteistyksessä on totuttu siihen, että laskostuminen on haitallista. Siksi on tarkasti huolehdittava signaalin esi- ja jälkisuodatuksesta, jotta spektri ei pääsisi laskostumaan.

Hallittu laskostaminen sen sijaan toteutetaan ylinäytteistykseenä, jossa otetaan hyötykäyttöön syntyvä laskostumisspektri. Tällöin jälkisuodatusta ei toteuteta, eli uuden Nyqvistin taajuuden alapuolelle jääneet spektrin kopiot jätetään signaaliin.

Hallittua laskostamista käytetään tässä työssä hyväksi siten, että näytteenottotaajuus nostetaan kaksinkertaiseksi, jolloin alkuperäisen spektrin mukaisesti yksi ylimääräinen spektrikuviot jää uuden Nyqvistin taajuuden alapuolelle.



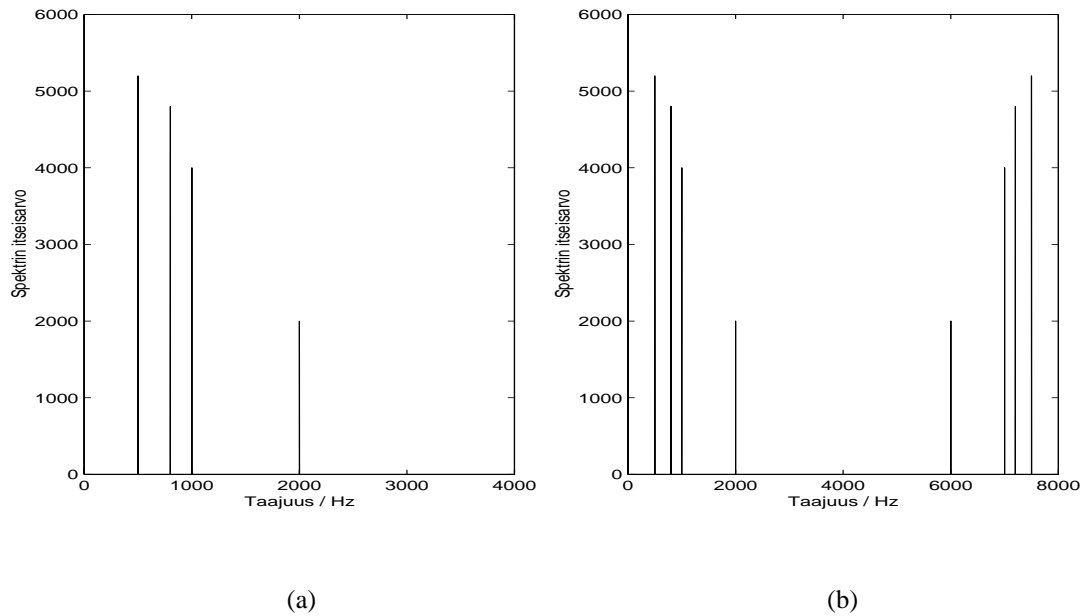
Kuva 6. (a) Jatkuvan ts. näytteistämättömän signaalin spektri sisältää ainoastaan ne taajuudet, joista signaali todellisuudessa muodostuu. (b) Signaali 8kHz:n näytetaajuudella näytteistettynä, jolloin alkuperäinen spektrikuvio toistuu näytteenottotaajuuden välein.

2.2.1 Toteutus aika-alueessa

Hallittu laskostaminen toteutetaan aika-alueessa siten, että jokaisen signaalinäytteen jälkeen lisätään yksi nolla-arvoinen näyte. Hallitun laskostamisen tapauksessa toimitaan siis samoin kuin ylinäytteistyksessä kertoimella $I = 2$, vain jälkisuodatus jätetään toteuttamatta. Signaalin näytetaajuus kasvaa kaksinkertaiseksi kuvan 3 mukaisesti.

2.2.2 Spektrin käyttäytyminen

Kun näytetaajuus nousee kaksinkertaiseksi nolla-arvoisten näytteiden lisäyksen seurauksena, jää yksi ylimääräinen alkuperäisen spektrin monikerta uuden Nyqvistin taajuuden alapuolelle, kuten kuvassa 7 on esitetty.



Kuva 7. (a) Alkuperäisen signaalin spektrin itseisarvo ($F_s = 8000\text{Hz}$). (b) Laskostetun signaalin spektrin itseisarvo ($F_s = 16000\text{Hz}$).

2.3 Ikkunointi

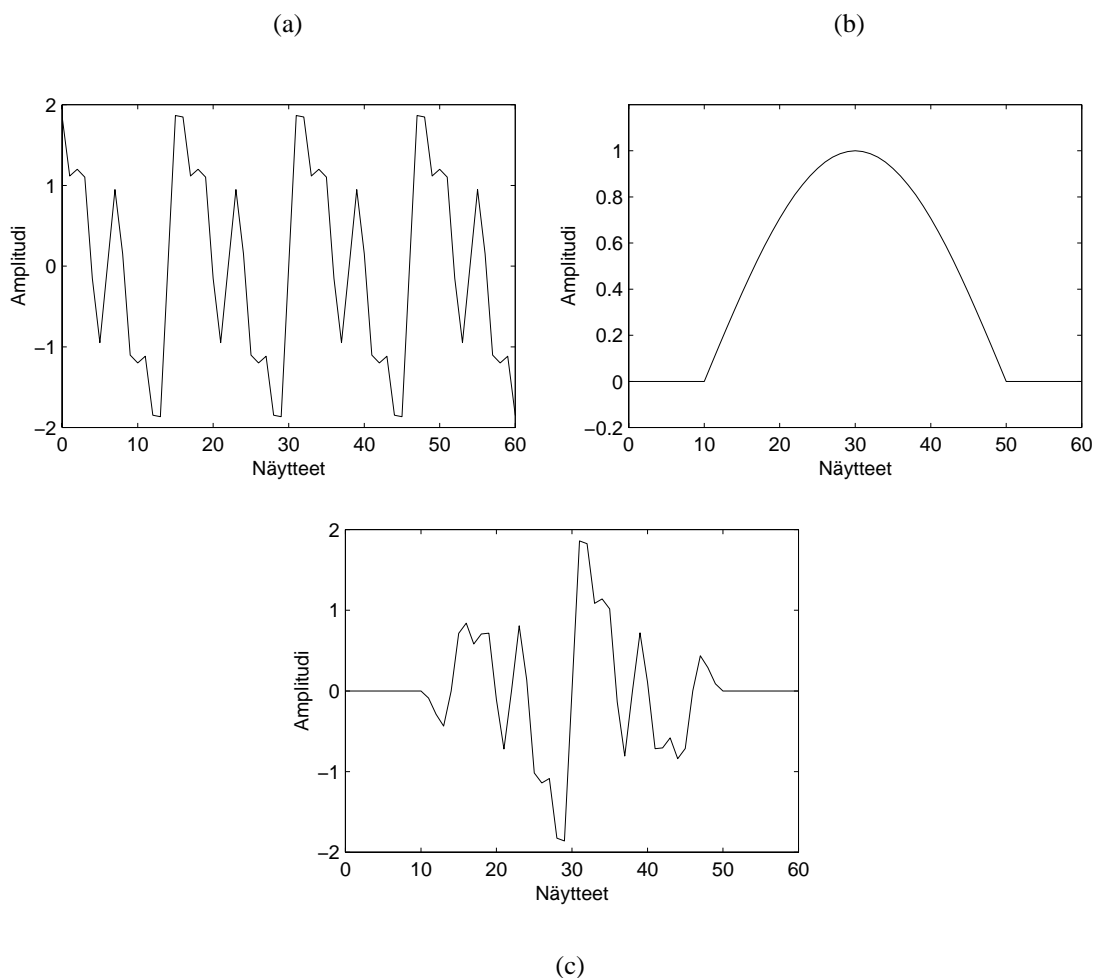
Jatkuvaa signaalia käsiteltäessä on tulevasta näytevirrasta katkaistava sopivia näytejonoja käsiteltäväksi. Ikkunoinnin tarkoitus on siis leikata signaali sopivan mittaisiin paloihin muuta käsittelyä varten. Käsittelyn jälkeen signaalinpalat jälleen yhdistetään jatkuvaksi näytejonoksi.

Mitä lyhyempinä osina signaali käsitellään, sen paremmin algoritmi toteuttaa jatkuva-aikaisuuden käsitteen, sillä sitä vähemmän syntyy viivettä halutun mittaiseen otokseen tarvittavia näytteitä odotellessa. Kuitenkaan hyvin lyhyestä näytepalasta on vaikea mitata signaalin ominaisuuksia, kuten ikkunoinnin vaikutuksia käsittelevässä kappaleessa selviää, joten äärettömästi ei voida kerralla käsiteltäväksi otettavan näytejonon pituutta lyhentää. Pidempi otos antaa paremmat mahdollisuudet tulkita signaalia oikein ja siten signaalia käsittelevä algoritmi muokkaa näytejonoa enemmän siten, miten sen on suunniteltu signaalia muuttavan. Pitkän näytejonon varastointi kuitenkin aiheuttaa viiveen signaalin etenemiseen.

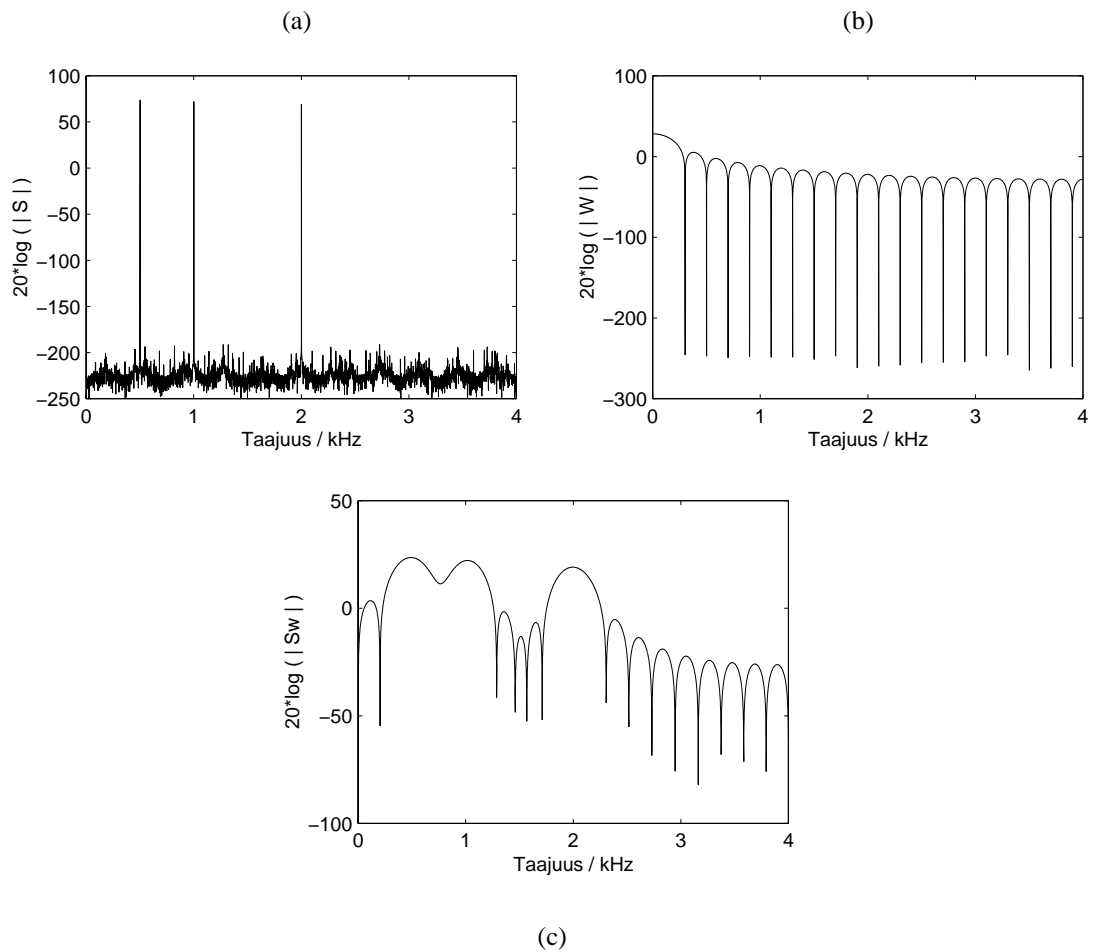
2.3.1 Ikkunoinnin vaikutukset

Ikkunafunktio ei ainoastaan jaa signaalia sopivanmittaisiin osiin, vaan vaikuttaa kuhunkin katkaisemaansa osaan muuttaen sen ominaisuuksia sekä aika- että taajuusalueessa [Oppenheim 1989]. Kuva 8 kertoo aika-alueen muutoksista ja kuva 9 taajuusalueen muutoksista.

Matemaattisesti ikkunointi aika-alueessa merkitsee alkuperäisen signaalin ja ikkunafunktion kertomista keskenään, jota kuva 8 esittää. Kahden näytejonon kertominen aika-alueessa merkitsee taajuusalueessa näytejonojen spektrien konvoloimista. Näin ikkunoitavan signaalin spektri konvoloituu ikkunafunktion spektrin kanssa, kuten kuvassa 9 on esitetty. Konvoloituminen merkitsee sitä, että ikkunafunktion spektri kopioituu kunkin alkuperäisessä signaalissa esiintyvän taajuuskomponentin kohdalle sen voimakkuuden mukaisesti ja summautuu kokonaisspektriin.



Kuva 8. Ikkunointi aika-alueessa. Näytetaajuus on 8000Hz ja ikkunan pituus 40 näytettä. (a) Alkuperäinen signaali, (b) ikkunafunktio, ja (c) ikkunoitu signaali.



Kuva 9. Ikkunoinnin vaikutus signaalin spektriin. Näytetaajuus on 8000Hz ja ikkunan pituus 40 näytettä. (a) Ikkunoitavan signaalin spektri. (b) Ikkunafunktion spektri. (c) Ikkunoidun signaalin spektri, eli (a) ja (b) –kuvien spektrien konvoluutio.

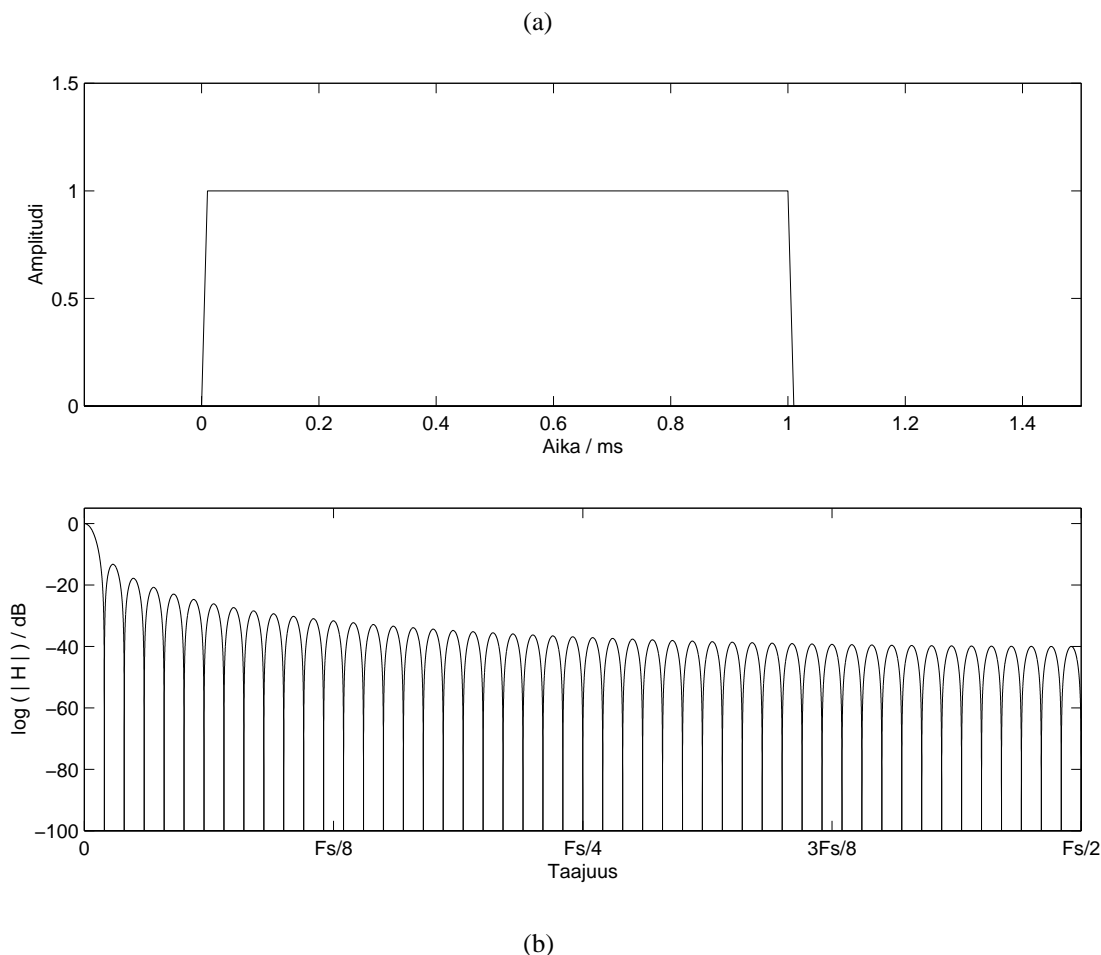
Käänteisikkunointi ja palojen yhdistäminen, mitkä suoritetaan signaalin käsittelyn jälkeen, palauttavat periaatteessa signaalin ominaisuudet ennalleen. Kuitenkin kun ikkunoinnissa vääristynyttä signaalia muokataan, ei käänteisikkunoitava signaali enää vastaa alunperin ikkunoitua signaalia. Näin käänteisikkunoinnin vääristymiä kompensoivat ominaisuudet eivät vaikuta ikkunoidun signaalin, vaan erilaisen – muokatun - signaalin ominaisuuksiin, jolloin osa vääristymistä saattaa vaimentua ja osa voimistua.

2.3.2 Ikkunafunktion valinta

Kuten edellisessä kappaleessa todettiin, ei spektrin muuttumista voida välttää signaalia paloiteltaessa, joten ikkunoinnin aiheuttamien vääristymien välttäminen on tärkeää. Kaikilta osin täydellistä ikkunafunktiota ei ole, vaan kuhunkin sovellukseen on

löydettävä sopiva ikkunafunktio arvioimalla sen aika- ja taajuusalueen ominaisuuksia.

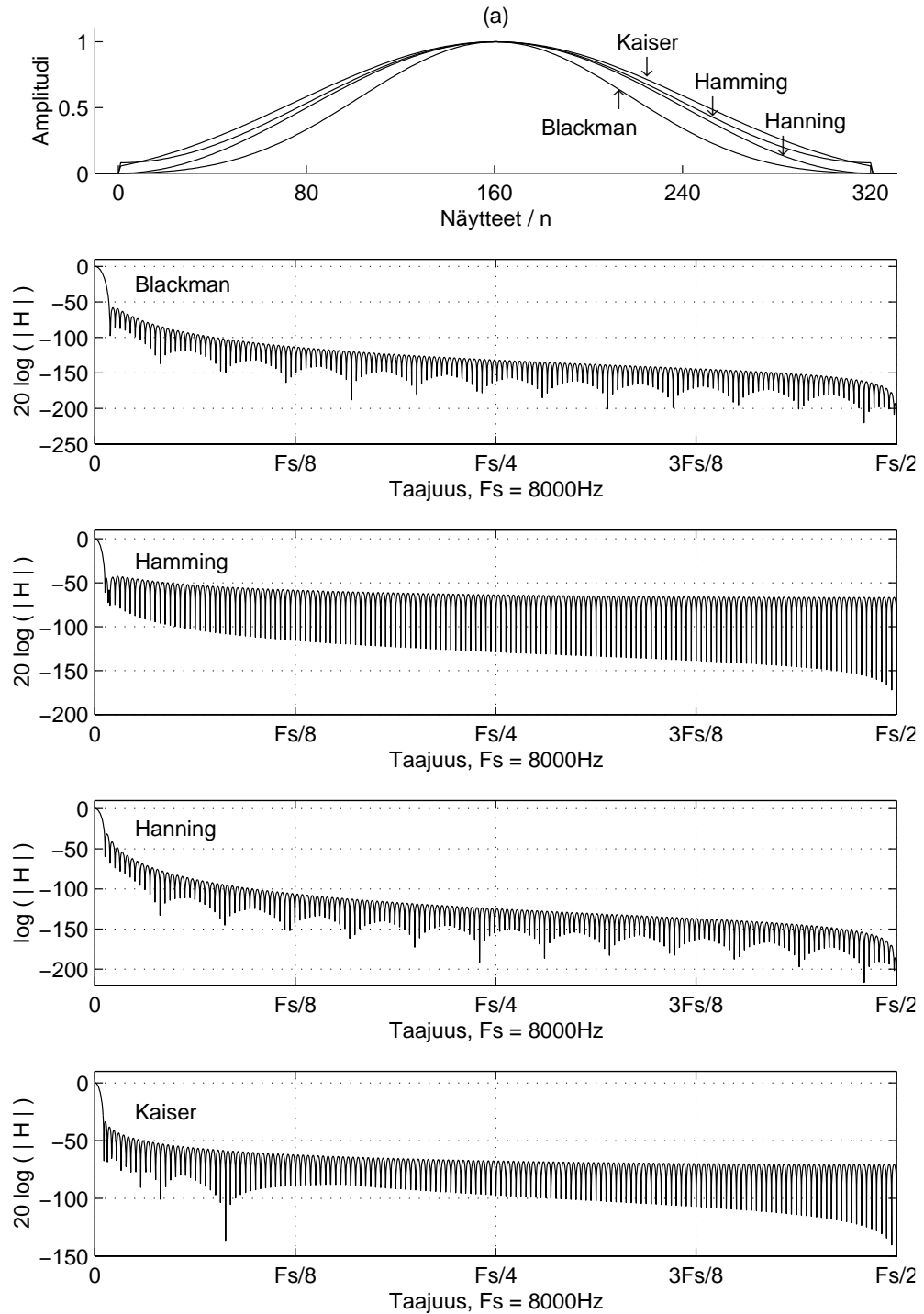
Aika-alueessa ideaalinen ikkunafunktio olisi kuvan 10a mukainen laatikko, joka katkaisee signaalista pätkän muuttamatta mitenkään sen näytteiden arvoja. Kuvassa 10b on kuvattu tämän aika-alueessa ideaalisen ikkunafunktion spektri. Lienee ilmeistä, että konvoloihtaessa alkuperäisen signaalin spektri tällä laatikkofunktion spektrillä muuttuu alkuperäisen signaalin spektri melkoisesti.



Kuva 10. (a) Aika-alueessa ideaalinen ikkunafunktio katkaisee signaalista halutun mittaisen pätkän muuttamatta näytteiden arvoja. (b) Laatikkofunktion taajuusvaste sen sijaan hävittää lähes kaiken tiedon alkuperäisen signaalin taajuusominaisuuksista konvoluution seurauksena.

Taajuusalueessa ideaalisen ikkunafunktion pitäisi toistaa jokainen taajuuskomponentti samana kuin se alkuperäisessä signaalissa esiintyy kääntämättä sen vaihetta ja muuttamatta sen voimakkuutta. Jotta ikkunafunktio ei konvoloinnissa muuttaisi alkuperäisen signaalin spektriä, olisi ikkunafunktion spektrin sisällettävä vain yksi itseisarvoltaan yhden yksikön korkuinen piikki nolllataajuudella. Tuollainen ikkuna on kuitenkin aika-alueessa äärettömän mittainen, joten käyttökelpoinen ikkuna on löydettävä sellaisten funktioiden joukosta, joiden aika- ja taajuusalueen ominaisuudet muodostavat sopivan kompromissin.

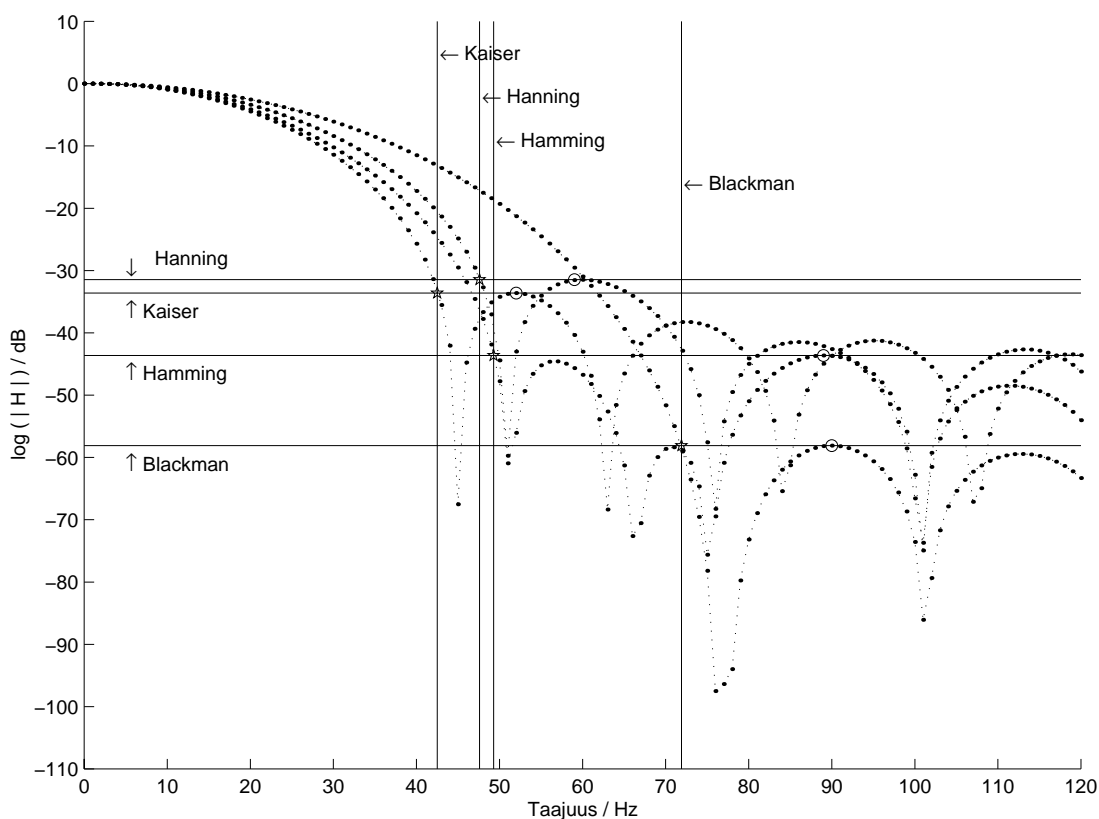
Ikkunafunktiota etsittäessä tärkeitä huomioon otettavia parametrejä ovat spektrin pääkomponentin leveys ja sen sivuilla olevien sivukomponenttien vaimennus. Mitä kapeampi pääkomponentti on, ja mitä vaimeampia ovat sivukomponentit, sitä enemmän ikkunafunktio muistuttaa taajuusalueessa ideaalista ikkunaa. Nämä ominaisuudet ovat sitä helpompia saavuttaa, mitä pidempi ikkuna on.



Kuva 11. (a) Blackman, Hamming, Hanning ja Kaiser -ikkunat aika-alueessa. (b) Blackman-ikkunan, (c) Hamming-ikkunan, (d) Hanning-ikkunan ja (e) Kaiser-ikkunan spektrit.

Yleisesti käytössä olevia ikkunafunktioita ovat esimerkiksi Blackman, Hamming, Hanning ja Kaiser –ikkunat, joiden ominaisuuksia on esitelty kuvassa 11. Kaikki kuvan 11 ikkunat ovat 40 ms:n, eli 320 näytteen mittaisia, kun näytteenottotaajuus on 8000Hz. Kuvaan 12 on piirretty näiden ikkunafunktioiden spektreistä merkittävin osa, eli pääkomponentit ja kustakin muutamia voimakkaimpia sivukomponentteja. Kuvan 12 perusteella voi havaita, että on vaikea toteuttaa ikkunaa, jolla olisi samanaikaisesti sekä kapea pääkomponentti, että voimakas vaimennus sivukomponenttien osalta.

Ikkunafunktio vaikuttaa konvoluutiossa myös signaalin vaiheeseen, mikä olisi toisinaan otettava huomioon. Audiosovelluksissa kuitenkin signaalin eri taajuuskomponenttien vaiheiden keskinäisellä vaihtelulla ei ole paljon merkitystä, sillä ihmiskorva ei juurikaan kuule vaihevaihteluita.



Kuva 12. Blackman, Hamming, Hanning ja Kaiser-ikkunoiden spektrien pääkomponentit ja voimakkaimmat sivukomponentit. Pääkomponenttien leveydet sivukomponentin korkeuden mukaisesti on merkitty tähdellä ja voimakkaimmat sivukomponentit on merkitty ympyröillä.

2.4 Lineaarinen ennustus

LP (linear prediction) eli lineaarinen ennustus on eräs digitaalisen signaalin käsittelymenetelmä, jolla puheen heräte ja ääniväylän taajuusominaisuudet voidaan karkeasti erottaa toisistaan. Puheen spektri voidaan parametrisoida LP-suotimen ennustuskertoimien avulla pieneen tietomäärään, mitä monet puheenkoodausalgoritmit käyttävät hyväkseen. Myös jotkin taajuuskaistan laajennusmenetelmät toteutetaan lineaarisen ennustuksen avulla.

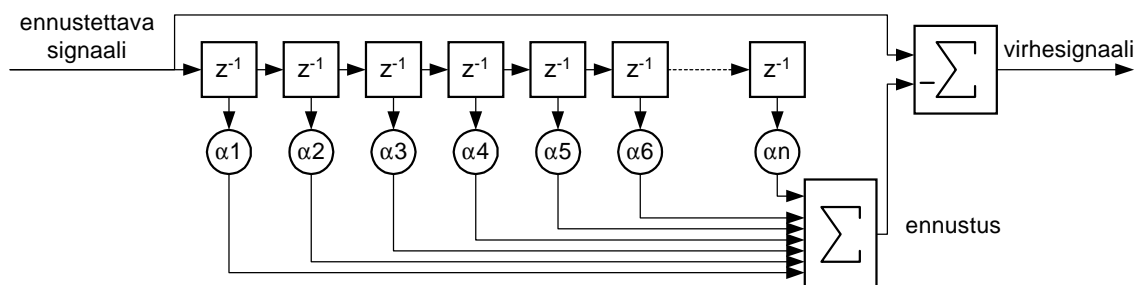
Lineaarinen ennustus nimensä mukaisesti ennustaa tulevia signaaliarvoja lineaarisesti edellisten signaaliarvojen perusteella. LPC-koodauksen kyky pakata ääntä perustuu ennustuksen onnistumiseen, jolloin äänen sijasta voidaan lähettää vastaanottopäähän mahdollisimman pieni ennustusvirhesignaali LPC-parametreineen.

2.4.1 Signaalin LP-suodatus

Lineaarinen ennustus voidaan toteuttaa kuvan 13 LP-suodinrakenteessa esiintyvällä FIR-suotimella. Suotimen kertoimet toimivat painotuskertoimina, joilla edellisiä näytteitä painotetaan seuraavaa näytettä ennustettaessa. Ennustus syntyy siis suodinkertoimilla painotettuna lineaarikombinaationa edellisistä näytteistä. Ennustettu näytetarvo vähennetään uuden näytteen todellisesta arvosta, jolloin LP-suodin tuottaa virhesignaalin, joka kertoo ennustuksen onnistumisesta. Suotimen siirtofunktio on lausekkeen (1) mukainen.

$$A(Z) = 1 - \sum_{k=1}^p \alpha_k \cdot z^{-k} \quad (1)$$

Suotimen kertoimet optimoidaan käsiteltävän signaalinpalan osalle siten, että virhesignaali muodostuu mahdollisimman pieneksi. Mitä pidempi FIR-suodin on, eli mitä suurempi on suotimen asteluku p , sitä yksityiskohtaisemmin se kykenee parametrisoimaan signaalin spektrin.



Kuva 13. Lineaarisen ennustuksen ja LP-suodatuksen toteuttava rakenne.

Jotta kaikki formanttien ja äänihuulten ominaisuudet tulisivat riittävän tarkasti kuvatuiksi, riittää 8 kHz:n tajuudella näytteistetyyn puheäänen kuvaamiseksi 8 – 12 suodinkerrointa [Rabiner & Schafer 1978].

2.4.2 Virhesignaalin ominaisuuksia

Ennustimen tarkoitus olisi luonnollisesti ennustaa seuraava signaaliarvo niin tarkasti, ettei mitään virhesignaalia syntyisi. Yksinkertainen lineaarinen ennustin ei kuitenkaan kykene täydelliseen suoritukseen, jolloin todellisen näytteen ja ennustetun arvon vähennyslaskusta jää jäljelle pieni virhesignaalinäyte, joka muodostuu lausekkeen (2) mukaisesti.

$$e(n) = s(n) - \sum_{k=1}^p \alpha_k \cdot s(n-k) \quad (2)$$

On kehitetty myös monimutkaisempia ennustimia, jotka yksinkertaisen lyhyen ajan ennustuksen lisäksi ennustavat ääntä pidemmällä aikavälillä. Tällöin virhesignaali jää pienemmäksi ja se kertoo signaalista hieman eri asioita kuin pelkän lyhyen aikavälin ennustimen virhesignaali.

Puheen lyhyen ajan ennustuksessa virhesignaali sisältää pienen kohinan lisäksi useimmiten pulsseja äänenkorkeuden jaksojen välein. Pulssit aiheutuvat äänihuulten sulkeutumishetkellä syntyvästä suuresta ennustusvirheestä.

Kohinamaisen virhesignaalin spektri on tasainen, siitä on hävinnyt äänen formanttirakenne. Virhesignaaleissa esiintyvät äänenkorkeuden jaksoista johtuvat pulssit synnyttävät spektriin kampamaisen rakenteen.

2.4.3 LP-kerrointen laskenta

Koodattaessa signaalia LPC-algoritmillä on näytejono analysoitava aluksi kehyksittäin, jotta löydetään LP-suotimelle oikeat painotuskertoimet.

LP-suotimen ennustuskertoimet voidaan määrittää monella tavalla. Kaikki laskentatavat kuitenkin tähtäävät samaan lopputulokseen, eli mahdollisimman tarkkaan signaaliarvojen ennustukseen. Kaikkia menettelytapoja käytettäessä on kuitenkin huomioitava, että riittävän pitkä – muutamia äänenkorkeuden jaksoja – otos signaalista valitaan kerralla analysoitavaksi, jotta kerrointen laskenta onnistuisi luotettavasti.

Tärkeimpiä ja perusideoiltaan poikkeavia LP-suotimen kerrointen määritysmenetelmiä ovat autokorrelaatiomenetelmä, kovarianssimenetelmä ja hilamenetelmät. Muiden menetelmien erot koskevat lähinnä laskentaoperaatioiden yksityiskohtia [Rabiner & Schafer 1978].

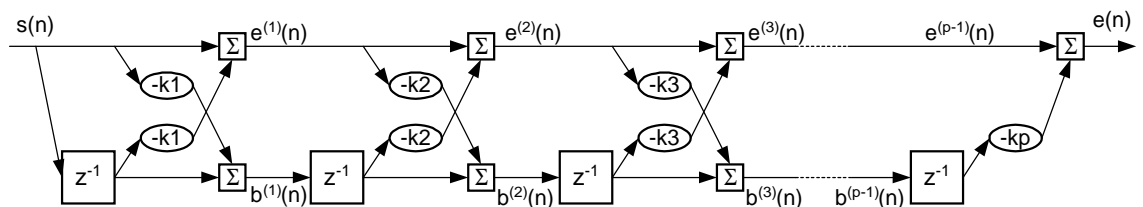
Autokorrelaatiomenetelmässä oletetaan analysoitavan näytejonon ulkopuolelle

jäävien signaalinäytteiden olevan nolla-arvoisia. Laskentavirheen minimoimiseksi analysoitava pala leikataan signaalista ikkunafunktiolla, joka vaimentaa otoksen reunat alueet taipumaan kohti nollaa.

Kovarianssimenetelmässä kertoimien optimointiin käytetään vain aitoja signaalinäytteitä, joten muuten autokorrelaatiomenetelmän kanssa samankaltaisen laskennan kautta päädytään hieman erilaisiin suodinkertoimiin kuin autokorrelaatiomenetelmällä.

Sekä autokorrelaatio- että kovarianssimenetelmässä päädytään yhtälöryhmiin, jotka voidaan esittää matriiseina ja ratkaista lineaarialgebran tarjoamin keinoin.

Hilamenetelmät ovat rekursiivisia algoritmeja suodinkertoimien laskemiseksi. Nimensä hilamenetelmät saavat hilarakenteen muodostavasta kuvan 14 lohkokaaavion mukaisesta toteutuksesta, jota ne käyttävät FIR-suotimella toteutetun rakenteen sijaan. Hilarakenteen heijastuskertoimet eivät ole samat kuin FIR:llä toteutetun LP-suotimen kertoimet, mutta ne ovat suoraan laskettavissa LP-suodinkertoimista.



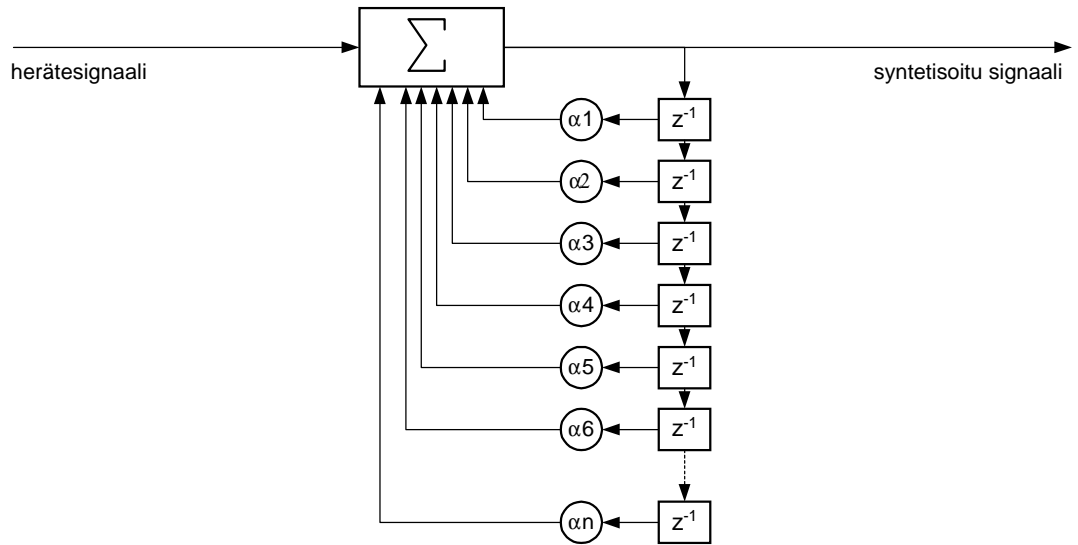
Kuva 14. Hilamenetelmän lohkokaavio [Rabiner & Schafer 1978].

2.4.4 Puheen tuottaminen LP-parametrien avulla

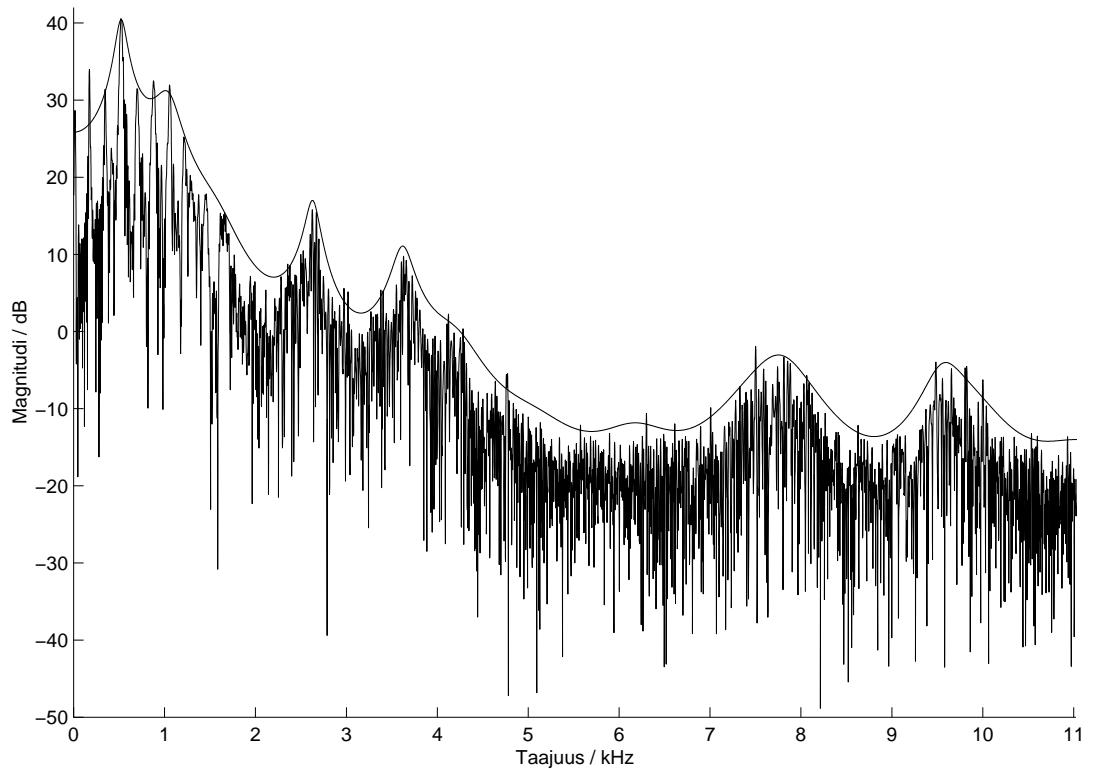
Signaalin LP-parametrit sisältävät jäännössiinaalin ja LP-suodinkertoimet. Näin herätteen ja ääniväylän ominaisuuksien avulla ilmaistu signaali voidaan tuottaa takaisin puheeksi syöttämällä herätesignaali LP-analyysisuotimen käänteissuotimeen, eli LP-synteesisuotimeen.

Kuvassa 15 on esitetty yksinkertainen LP-synteesisuodin. Sama laskentaoperaatio on tosin mahdollista toteuttaa myös muunlaisen rakenteen avulla. Kuvan 15 synteesisuodinkertoimet vastaavat kuvan 13 LP-analyysisuotimen kertoimia. LP-synteesisuodin toteuttaa LP-analyysisuotimelle vastakkaisen operaation, eli tuottaa ääntä virhe- eli herätesignaalin ja kehyksittäin muuttuvien suodinkertoimien avulla.

Synteettistä puhetta voidaan tuottaa LP-synteesisuotimen avulla käyttämällä soinnillisissa äänissä herätteenä pulssijonoa ja soinnittomissa ääniteissä valkoista kohinaa ja muuttelemalla suodinkertoimia [Rabiner & Schafer 1978]. Tosin näin yksinkertaisesti tuotettu puhe kuulostaa hyvin synteettiseltä. Taajuuskaistan laajennusmenetelmät, jotka on toteutettu LP-synteesin avulla, käyttävät herätteenä kohinan ja pulssijonon sijaan LP-analyysisuodatuksessa syntynyttä jäännössiinaalia.



Kuva 15. IIR-tyyppinen LP-synteesisuodin, jonka kertoimet $\alpha_1 - \alpha_n$ vastaavat FIR-suotimen avulla toteutetun LP-analyysisuotimen kertoimia.



Kuva 16. Signaalin FFT-spektrin ja LPC-spektrin eli käänteisen LP-suotimen siirtofunktion vastaavuus.

Taajuusalueessa LP-synteesisuotimen siirtofunktio noudattaa mahdollisuuksiensa mukaan kuvattavan systeemin siirtofunktiota. Kerrointen optimointiperiaatteesta johtuen IIR-tyyppisen LP-synteesisuotimen taajuusvaste vastaa systeemin siirtofunktiota paremmin voimakkaiden kuin vaimeiden spektrikomponenttien kohdalla, kuten kuvassa 16 on havaittavissa.

2.5 Puhesignaalin koodaus

Tiedonsiirtojärjestelmissä pyritään käyttämään siirtolinjoja mahdollisimman tehokkaasti, jotta jatkuvasti kasvavia tietovirtoja pystyttäisiin käsittelemään. Digitaalista tietoa voidaan pakata monella tavalla, jotta se veisi vähemmän muistitilaa ja siirtokapasiteettia. Jokaisella tiedon lajilla on sille ominaiset piirteet, joita hyväksi käyttämällä tieto saadaan pakattua ja siirrettyä pienemmällä bittimäärällä kuin mitä alkuperäinen data sisälsi.

Puhelimet ovat yksi suurimmista siirtolinjoja – sekä kiinteitä, että ilmaita – kuormittavista laitteista, minkä vuoksi puheäänien tehokas koodaus on erityisen tärkeää. Lähettävässä päässä ääni pakataan eli koodataan mahdollisimman pieneen bittimäärään jonkin algoritmin avulla, ja vastaanottopäässä pakkaus puretaan algoritmin käänteisoperaatiolla.

Toiset äänenkoodausalgoritmit toimivat häviöttömästi, jolloin ääni saadaan vastaanottopäässä konstruoitua täydellisesti, eikä äänen laatu millään tavalla kärsi koodausalgoritmin käytöstä.

Toiset algoritmit taas hävittävät osan tiedosta päästäkseen pienempään bittimäärään. Häviöllisissä algoritmeissa käytetään hyväksi mm. ihmiskorvan peitto-ominaisuuksia ja puheen laadun kannalta olennaisia parametreja valittaessa säilytettäviä ja hävitettäviä äänen piirteitä, jotta häviöllisyydestä huolimatta äänen laatu pystyttäisiin säilyttämään mahdollisimman hyvänä.

Tässä kappaleessa esitellään puheenkoodausalgoritmeista ensin PCM (pulse code modulation), joka ei varsinaisesti pakkaa tietoa, vaan on luotu äänen muuttamiseksi analogisesta muodosta digitaaliseen muotoon. Toinen esiteltävä algoritmi on CELP (codebook excited linear prediction), joka edustaa modernia puheenkoodausta. CELP:iä käytetään mm. GSM-puhelimien tiedonsiirrossa.

2.5.1 Pulssikoodimodulaatio

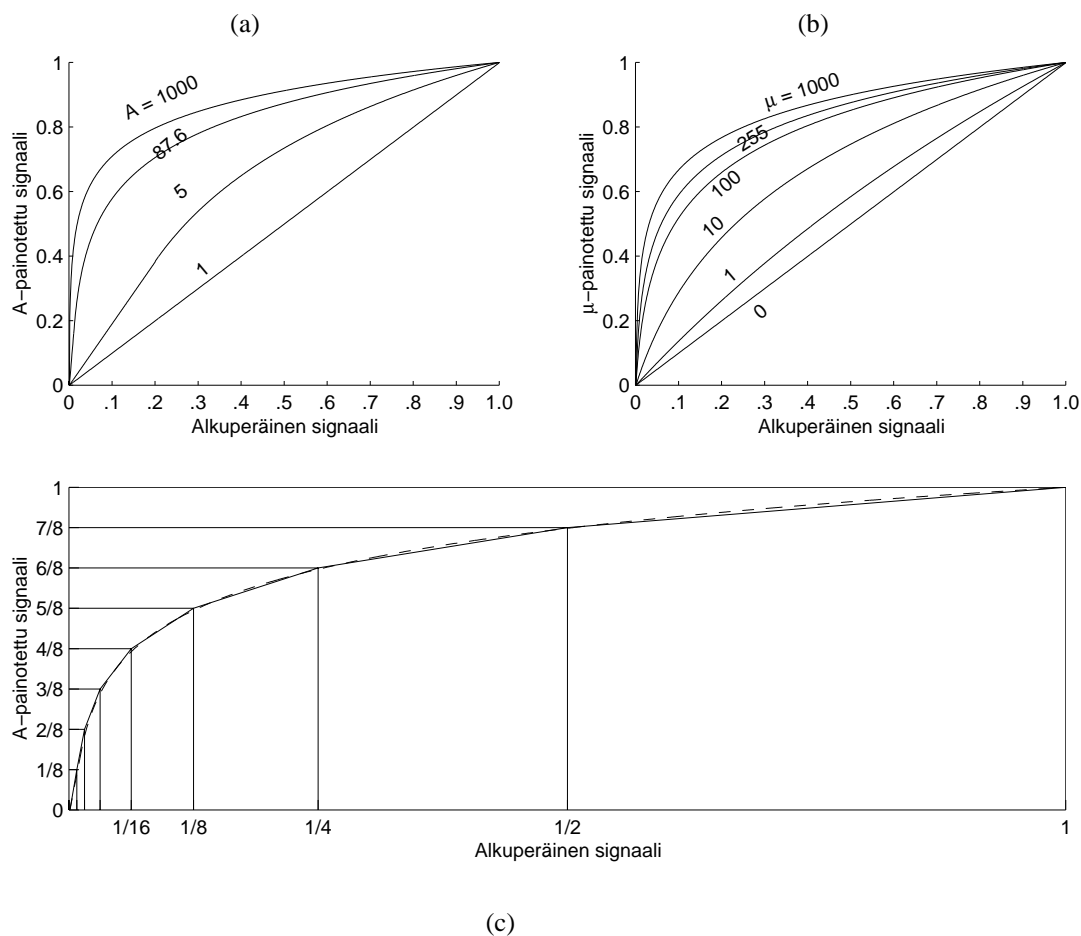
Siirryttäessä analogisesta tiedonsiirrosta digitaaliseen tiedonsiirtoon kehittyi PCM (pulse code modulation) eli pulssikoodimodulaatio [Owen 1982]. PCM on yksinkertainen ja nykymittapuun mukaan varsin tehoton tapa koodata ääntä, mutta se kuitenkin kykenee yksinkertaista digitaalisten signaaliarvojen välittämistä parempaan

äänenlaatuun. PCM on säilynyt pitkään käytössä, koska se vakiinnutti paikkansa heti digitaalisen aikakauden alkutaipaleella puhelimen äänensiirrossa.

Siirtyminen analogisen signaalin lähetyksestä signaalin siirtoon digitaalisessa esitysmuodossa oli jo sinänsä suuri harppaus kohti parempaa äänenlaatua, sillä siirtotien häiriöt oli huomattavasti helpompi poistaa digitaalisessa kuin analogisessa muodossa olevasta signaalista.

Analogisen signaalin koodaus digitaaliseen PCM-muotoon

PCM:ssä käytetään historiallisista syistä näytteenottotaajuutena 8 kHz:iä, mikä riittää kohtuullisen puheäänien laadun säilyttämiseen. Ennen näytteistystä analogisesta signaalista on laskostumisen estämiseksi suodatettava yli 4 kHz:n taajuudet pois. PCM:n välittämä taajuuskaista on vakiintunut 300 – 3400 Hz:iin, koska 4 kHz:n kohdalla riittävä vaimennus saatiin analogisella suotimella, joka alkoi vaimentaa signaalia voimakkaasti jo 3,4 kHz:n tienoilla. PCM:ssä koodataan kukin näyte 8 bitillä, joten tiedonsiirtonopeudeksi syntyy 64 kbit/s



Kuva 17. Signaalin painotus PCM-koodauksessa. (a) A-lain mukaisia painotuskäyriä; (b) μ -lain mukaisia painotuskäyriä; ja (c) A-lain mukainen paloittain lineaarinen painotuskäyrä ($A = 87,6$).

Signaalin pieniamplitudiset osuudet sisältävät ymmärrettävyyden kannalta kaikkein tärkeimmän informaation, joten PCM:ssä käytetään epälineaarista painotusta tämän alueen korostamiseksi. Euroopassa, Afrikassa, Etelä-Amerikassa ja kansainvälisillä reiteillä käytetään A-lain mukaista painotusta A:n arvolla 87,6 ja Pohjois-Amerikassa ja Japanissa on käytössä μ -lain mukainen painotus μ :n arvolla 255 [Alahuhta 1979]. Kuvassa 17(a) ja (b) on esitelty näiden lakien mukaisia painotuskäyriä muutamilla A:n ja μ :n arvoilla [Owen 1982].

Painotuskäyriä käytetään paloittain lineaarisessa muodossa, jolloin signaaliarvojen rekonstruointi on helpompaa kuin olisi käytettäessä kokonaan epälineaarista painotuskäyriä. Kuvassa 17 (c) on esitetty käytössä oleva A-lain mukainen paloittain lineaarinen painotuskäyrä, joka on jaettu 8:aan lineaariseen segmenttiin.

Esisuodatuksen ja painotuksen jälkeen signaali kvantisoidaan digitaaliseen muotoon. Kukin signaalinäyte koodataan kuvan 18 mukaisesti 8 bitin tietovuoksi. 8:sta bitistä ensimmäinen kertoo signaalinäytteen etumerkin. Kolmella seuraavalla bitillä merkitään segmentti, jonka alueelle signaalinäytteen itseisarvo osuu. Kukin segmentti jaetaan 16:een samankokoiseen osaan ja neljä viimeistä bittiä käytetään tämän segmentin sisäisen arvon koodaukseen [Alahuhta 1979]. Näin 8 kHz:n näytetaajuudella PCM tuottaa välitettävää dataa 64 kbit/s.

| Etumerkki | Segmentti | Segmentti | Segmentti | Arvo segmentin sisällä | Arvo segmentin sisällä | Arvo segmentin sisällä | Arvo segmentin sisällä |
|-----------|-----------|-----------|-----------|------------------------|------------------------|------------------------|------------------------|
| | | | | | | | |

Kuva 18. PCM-näytteen bittien merkitykset.

2.5.2 CELP

CELP (code excited linear prediction) edustaa modernia puheenkoodausmenetelmää, jossa ääni analysoidaan synteessin avulla ja äänilähde parametrisoidaan pieneen määrään tunnuslukuja, jotka voidaan siirää huomattavasti PCM:ää pienemmällä tiedonsiirtonopeudella.

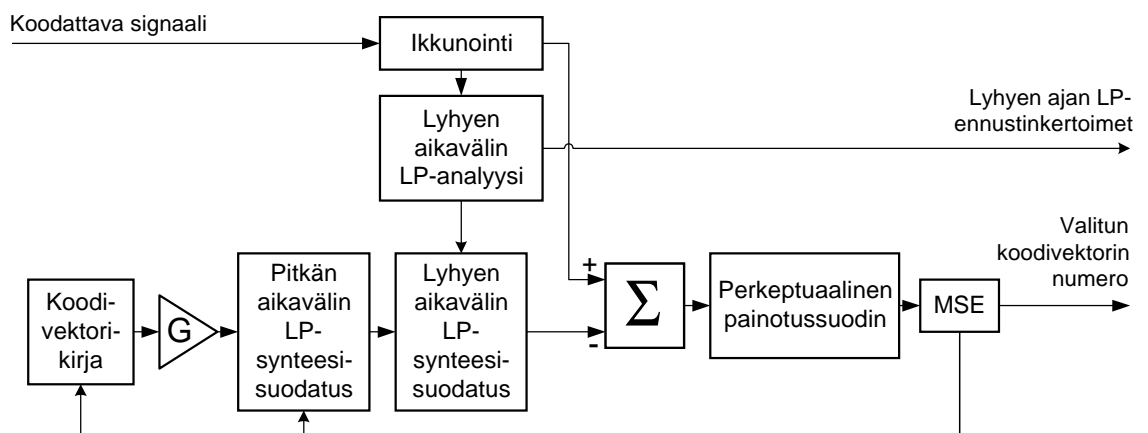
Alkuperäisen CELP-menetelmän [Schröder, Atal 1985] vuokaavio on esitetty kuvassa 19. CELP käsittelee signaalin muutaman kymmenen millisekunnin mittaisina paloina, jotka leikataan tulevasta näytejonosta sopivan ikkunafunktion avulla. Kukin pala analysoidaan lyhyen ajan LP-analyysin avulla ja LP-kertoimet lähetetään parametreina vastaanottopäähän.

Koodikirja sisältää käsiteltävien palojen mittaisia näytevektoreita, joista valitaan kullekin äänenpätkälle vektori, joka tuottaa pienimmän virheen alkuperäisen ja näin syntetisoidun äänen välille. Koodikirjasta valitun vektorin teho ja pitkän aikavälin LP-

kertoimet sovitetaan minimoimalla neliöityä ennustusvirhettä siten, että valittu vektori tuottaa lyhyen ajan LP-synteesisuotimessa mahdollisimman tarkasti alkuperäisen kaltaisen signaalin. Näin syntetisoitu ääni vähennetään alkuperäisestä äänestä, jolloin jäljelle jää pieni virhesignaali. Tämä suodatetaan vielä perkeptuaalisen eli ihmiskuulon ominaisuudet huomioon ottavan painotussuotimen läpi, jonka tarkoitus on siirtää virhe formanttien välisiltä taajuusalueilta formanttien alle.

Koodikirja voidaan valita hyvin monella tavalla. Se voidaan luoda satunnaisesti noudattaen esimerkiksi Gaussin jakaumaa, joka on osoittautunut hyvin samankaltaiseksi kuin ennustusvirheen jakauma [Schröder, Atal 1985]. Voidaan käyttää myös esimerkiksi koodikirjoja, joiden vektorit ovat toistensa viivästettyjä kopioita [Trancoso, Ribeiro, Almeida, Oliveira, Marques 1990].

Alkuperäisestä CELP-menetelmästä on kehitetty monia muunnelmia, esimerkiksi GSM-järjestelmässä käytössä oleva ACELP (algebraic code excited linear prediction) –koodausjärjestelmään perustuva EFR (enhanced full rate) –kooderi toimii tiedonsiirtonopeudella 12,2 kbit/s ja kykenee välittämään puheen laadultaan MOS-asteikolla arvioituna noin arvosanalla 4, kun bitinsiirrossa ei tapahdu virheitä [Honkanen, Vainio, Järvinen, Haavisto 1997]. Vuonna 1995 ITU-T standardoi CS-ACELP (conjugate-structure algebraic code excited linear prediction) –menetelmän, joka toimii bittinopeudella 8 kbit/s [ITU-T G.729].



Kuva 19. CELP (code excited linear prediction) –menetelmän mukaisen puheenkoodauksen toteuttava rakenne. Koodattava signaali parametrisoidaan koodikirjan ja LP-suotimien avulla pienellä tiedonsiirtonopeudella välitettäväksi.

3 ÄÄNEN LAADUN SUBJEKTIIVINEN TESTAUS

Kun jonkin äänenkäsittelyjärjestelmän toimintaa halutaan arvioida syntyvän äänen laadun kannalta, ovat vaihtoehtoina äänen laadun mittaaminen subjektiivisin tai objektiivisin menetelmin. Ennen digitaalisen signaalinkäsittelyn aikakautta äänen laatua mitattiin yksinomaan subjektiivisin kuuntelukokein, joissa ihmiset kertovat mielipiteensä äänestä. Digitaalisten signaalinkäsittelymenetelmien kehittyttyä on ollut mahdollista määritellä objektiivisia äänen laadun mittareita, jotka perustuvat äänen aika- ja taajuustason ominaisuuksista laskettaviin matemaattisiin suureisiin.

Täyttää yksimielisyyttä siitä, onko objektiivinen vai subjektiivinen äänen laadun mittaaminen suositeltavampaa, ei olla saavutettu. Molempien puolesta ja kumpaakin vastaan voidaan esittää näkökulmia. Joka tapauksessa voidaan väittää subjektiivisten ja objektiivisten testitulosten kertovan samoista asioista [Toole 1985].

Äänen laadun subjektiivinen testaus on sikäli ongelmallista, että kokeiden toteutuksessa erilaiset ympäristömuuttujat ja inhimilliset tekijät saattavat vaikuttaa koetuloksiin. Lisäksi mielipiteitä äänestä voidaan esittää hyvin monesta eri näkökulmasta. Monet arvostelevat subjektiivisten kuuntelukokeiden tuloksia siksi, että he uskovat jokaisen ihmisen kuuloaistin olevan niin erilainen, etteivät tulokset päde muulle kuin koekuuntelijajoukolle. On totta, että ihmiset kokevat kuulemansa hieman eri tavalla riippuen heidän mieltymyksistään, kokemuksestaan kuuntelijana, kulttuuritaustastaan yms. Jokaisella on kuulemistaan äänistä henkilökohtainen mielipide, mutta tuolla mielipiteellä ei ole muille mitään merkitystä, ellei voida osoittaa, että mielipide perustuu objektiiviseen havaintoon ja että mielipide on vakaa [Lipshitz 1981].

Jotta eri aikoina eri paikoissa suoritettujen kuuntelukokeiden tuloksia olisi mahdollista verrata keskenään, on kuuntelukokeiden yhtenäistämiseksi ollut välttämätöntä luoda standardeja, joissa määritellään kaikki kuuntelijoita, koetilannetta ja tulosten analysointia koskevat näkökohdat. Puhelinyhteyden äänenlaadun subjektiivista testausta käsittelee mm. ITU-T:n Recommendation P.800.

3.1 Mitä on äänen laatu?

Äänen laadun mittausta ajatellen äänen laatu on sitä, kuinka paljon arvioitava ääni muistuttaa vastaavaa luonnollista ääntä. Toisaalta joskus saattaa synteettinen ääni kuulostaa paremmalta kuin vastaava luonnollinen. Tällöin herää kysymys siitä, mikä

määrittää äänen laadun, ellei sen luonnollinen vastine? Matemaattisesti puhdas ääni taas voi ihmiskorvaan kuulostaa hyvinkin epämiellyttävältä, joten matemaattisinkaan keinoin on vaikea määrätä mittaa äänen laadulle.

Ihmiskorvan, hermoston ja aivojen toimintaa tutkitaan jatkuvasti, jotta ymmärrettäisiin, miten ihminen kokee kuulemansa äänet. Digitaalinen signaalinkäsittely on mahdollistanut kuulohavainnon määrittelyn jossain määrin matemaattisesti. Fysiologisten tekijöiden lisäksi kokemukseen äänestä vaikuttavat kunkin yksilön psykologiset ominaisuudet kuten myös ympäristötekijät ja kuulijan kokemusmaailma. On siis mahdotonta määrätä mitään yhtä suuretta, joka kertoisi äänen absoluuttisen laadun.

Äänellä on monia ominaisuuksia ja ääntä tuotetaan moneen eri tarkoitukseen, näin eri ääniltä odotetaan eri asioita. Äänentoistolaitteilla tuotettavan musiikin odotetaan täydellisimminkin jäljittelevän alkuperäistä musiikkiesitystä, eikä äänessä saa esiintyä laitteistosta eikä äänitystilanteesta syntyneitä häiriöitä. Vähemmän sen sijaan odotetaan puhelimen välittämältä ääneltä, jonka tarkoitus on siirtää puhe ymmärrettävässä muodossa kuulijalle. Äänen laatu riippuu siis äänen objektiivisten ominaisuuksien lisäksi siitä, kuinka hyvin se täyttää sille asetetut vaatimukset.

Äänen vaikutukset ihmiseen voidaan jakaa ainakin seuraaviin kolmeen pääkategoriaan: 1) fysiologiset vaikutukset, 2) informaatiovaikutukset ja 3) elämykselliset vaikutukset [Karjalainen 1998]. PCM-koodatun puheen kohdalla näistä merkittävimmäksi nousevat informaatiovaikutukset – onhan puheäänien välittämisen perimmäisenä tarkoituksena siirtää informaatiota. Myös elämyksellinen vaikutus on siinä mielessä merkittävä, kuinka hyvin puhuja on tunnistettavissa ja kuinka kuulija kokee mahdollisesti PCM-koodauksen vääristämät äänenpainot ja äänensävyyn vivahteet, joilla puhuja ilmentää tunnetilojaan ja todellisia ajatuksiaan. PCM-koodattu signaali kykenee välittämään puheen sanojen informaatioisisällön kiitettävästi, mutta taajuuskaistan laajennuksella voidaan äänen laatua parantaa sen elämykselliseltä kannalta.

3.1.1 Objektiivinen äänenlaadun mittaus

Koska objektiivinen äänen laadun määrittäminen on monin verroin helpompaa, nopeampaa ja edullisempaa kuin subjektiivisten kuuntelukokeiden järjestäminen, ja lisäksi objektiivisin menetelmin saadut tulokset ovat helpommin vertailtavissa keskenään kuin subjektiivisista kokeista saadut tulokset, määritetään tulevaisuudessa äänen laatua yhä enemmän objektiivisin menetelmin.

Tällä hetkellä käytössä olevia puheäänien objektiivisia äänenlaadunmittareita ovat esimerkiksi meluannos tiettyinä aikavälinä eli ekvivalenttitaso, artikulaatioindeksi ja laskennallinen konsonanttihäviöprosentti [Karjalainen 1998]. Nämä suureet voidaan laskea digitaalisen signaalin ominaisuuksista, joten tulokset ovat täysin yksiselitteisiä ja eri äänten tunnusluvut voidaan helposti laittaa järjestykseen. Tosin objektiivisin

menetelmin havaittuja äänen ominaisuuksia ei aina osata yhdistää oikeisiin kuulohavaintoihin [Lipshitz 1981], eikä näin osata sanoa mikä tunnusluvun arvo merkitsee parasta äänenlaatua.

Tällä hetkellä objektiiviset menetelmät eivät vielä anna luotettavaa kuvaa todellisesta äänen laadusta, eivätkä kerro paljoakaan siitä, miten ihminen äänen kokee. Jatkuvasti kuitenkin kehitetään uusia objektiivisia menetelmiä äänen laadun määrittämiseen sitä mukaa, kun ihmisen kuulemisen fysiologian ja hermoston tuntemus kasvaa. Askel kohti kuulon mukaista äänen laadun mittausta ovat ajassa, taajuudessa ja dynamiikassa koettavat peittoilmiöt huomioon ottavien mallien kehittäminen.

3.1.2 Subjektiiivinen äänenlaatu

Subjektiiivinen äänenlaatu on pohjimmiltaan sitä, millaisena kuulija äänen kokee. Äänestä voidaan erotella kuuleman perusteella monia ominaisuuksia, kuten esimerkiksi äänekkyyys, äänenkorkeus, terävyys, karheus, vaihteluvoimakkuus ja tonaalisuus. Kunkin ominaisuuden olemuksesta äänessä jokaisella kuulijalla on oma mielipiteensä, mutta ihmiset pystyvät kuitenkin vertailemaan eri ääniä näiden ominaisuuksien suhteen kohtalaisen yksimielisesti.

Äänen laadulle on olemassa myös paljon muita määritelmiä. Floyd Toole on subjektiivista äänen laatua käsittelevässä artikkelissaan [Toole 1985] ehdottanut seuraavia ominaisuuksia subjektiivisen äänenlaadun mittareiksi:

Selvyys kertoo kuinka hyvin äänestä ovat erotettavissa eri äänilähteet ja kunkin äänilähteen tuottamat erilliset äänet.

Pehmeys viittaa korkeataajuisten äänten laatuun, joiden tulisi olla luonnollisen pehmeitä; ei liian vaisuja eikä puuroutuneita, eikä myöskään liian teräviä, kirskuvia tai kovia.

Täyteläisyys kuvaa matalataajuisia ääniä - niiden tasapainoa keski- ja ylätaajuuksien kanssa. Hyvä ääni ei saa olla liian ohutta, eikä liian täyteläistä.

Kirkkaus mittaa korkeataajuisten äänten suhteellista osuutta keski- ja matalataajuisiin ääniin. Korkeat äänet eivät saa soida liian kirkkaasti, eivätkä myöskään saa jäädä matalampien taajuuksien peittoon.

Miellyttäväisyys on äänen yleismittari, joka kertoo onko äänessä häiritseviä piirteitä vai onko sen kuunteleminen miellyttävä kokemus.

Tarkkuus on myöskin äänen laatua yleisesti arvioiva mittari, joka kertoo kuinka lähellä täydellistä arvioitava ääni on, eli kuinka hyvin tuotettu ääni vastaa kyseisen äänen luonnollista asua.

Kaikille normaalikuuloisille ihmisille yhteiset kokemukset äänestä syntyvät korvan ja hermoston fysiologisista ominaisuuksista. Ihmiskorva kuulee äänen suhteellisen pieniä lineaarisia eroja, kuten eroja äänekkyydessä, signaalin polaarisuudessa, taajuus- ja vaihevasteessa [Lipshitz 1981]. Kuultavien äänien ominaisuuksiin vaikuttavat myös korvan peitto-ominaisuudet aika- ja taajuusalueessa [O'Shaughnessy 1987]. Lisäksi kuulokokemukseen vaikuttavat monet psykologiset ja kokemusperäiset syyt.

3.2 Kuuntelukokeiden muuttujia

Edellisessä kappaleessa lueteltujen näkökulmien paljoudesta huolimatta kuuntelukokeiden avulla on mahdollista tutkia äänen laatua tieteellisesti hyväksyttävästi, kun kaikki kokeisiin vaikuttavat tekijät otetaan huomioon ja raportoidaan asianmukaisella tavalla. On myös olemassa subjektiivisten kuuntelukokeiden järjestelyjä koskevia standardeja, joiden käyttö kuuntelukokeita suunniteltaessa on erittäin suositeltavaa. Standardien mukaisten kuuntelukokeiden tulokset ovat vertailtavissa keskenään toisin kuin erilaisissa ympäristöissä vaihtelevin menetelmin suoritettujen kuuntelukokeiden tulokset.

Kaikki koetilanteeseen ja tulosten analysointiin vaikuttavat pienetkin tekijät on otettava huomioon ja toteutettava harkiten, sillä koetulos riippuu pääasiallisesti koko prosessin heikoimmasta osasta. Kaikkien kokeessa käytettävien laitteiden ja ympäristömuuttujien vaikutukset on tarkastettava ennen kokeen suorittamista, jotta kuunneltavan äänen laatu ei huononisi koeympäristön vajavaisuuden takia.

Jotta kuuntelukoe olisi ideaalinen, se on oltava toistettavissa eri paikassa, eri aikaan ja eri kuuntelijoilla, se mittaa ainoastaan tutkittavan järjestelmän kuultavia ominaisuuksia ja koe määrittää selkeästi kuultavien erojen suuruudet tai absoluuttisen hyvyyden sopivaa subjektiivista asteikkoa hyväksi käyttäen [Toole 1982].

Ongelmallisia asioita kokeen suunnittelussa ovat arvioitavien parametrien ja riittävän yksinkertaisten arviointimenetelmien valinta ja lopulta yksilöllisten arvioijien tulosten analysointi [Toole 1985].

3.2.1 Kuuntelijoiden valinta

Kuuntelijan arviointikykyyn vaikuttavat monet tekijät. Tällaisia vaikuttajia ovat mm. kuunneltavien äänten tuntemus, kuuntelukoemenetelmän ja kuunteluhuoneen tuntemus, kokemukset kuuntelutehtävästä, arvostelukyky ja -taipumus, kuuntelijan fyysinen kuulokynnyks, kokeneisuus koekuuntelijana ja kuuntelijaryhmän keskinäinen kommunikointi [Toole 1985].

Tärkein vaatimus kuuntelijalle lienee normaali kuulo, joka yleensä määritetään kuulokynnyksen perusteella. Kuuntelijoiden kuulon herkkyys vaihtelee huomattavasti erityisesti korkeataajuisten äänten kohdalla. Kuitenkin kuulokynnyks matalilla, alle 1000Hz:n taajuuksilla vaikuttaa erityisesti koevastausten vaihteluun. Alle 1000:Hz:n alueella koekuuntelijan kuulokynnyksen tulisi olla lähellä 0dB:ä [Toole 1985].

Jos henkilö on jossain elämänsä vaiheessa ollut alttiina voimakkailla äänillä pitkiä aikoja, on hänen kuulokynnyksensä todennäköisesti kohonnut, ja samalla hänen kuulonsa tarkkuus on heikentynyt. Monessa yhteydessä on todistettu, että henkilöt, joiden kuulo on heikentynyt tai vaurioitunut, arvioivat ääntä vajavaisesti tai poikkeavasti normaalikuuloisiin kuuntelijoihin verrattuna [Toole 1985]. Epänormaali

kuulokyky aiheuttaa poikkeavia arvioita erityisesti taajuuksista, äänenpaineesta ja äänenvoimakkuudesta.

Myös kuuntelijan ikä vaikuttaa kuuntelun tarkkuuteen. Kuuntelukokeissa [Toole 1985] on yleisesti ottaen havaittavissa korrelaatiota kuuntelijan iän ja arvioinnin hajonnan kesken.

Kuulokynnys on silti vain yksi koehenkilön arviointikykyyn vaikuttava tekijä, joten pelkästään sen perusteella ei voida tehdä johtopäätöksiä.

Toinen erittäin tärkeä tulosten luotettavuutta parantava koekuuntelijan ominaisuus on kokemus. Kokeneet kuuntelijat pystyvät erittelemään äänten eroja huomattavasti tarkemmin ja luotettavammin kuin kokemattomat, sillä kokemattomilla kuuntelijoilla on taipumus keskittyä vain kaikkein ilmeisimpiin äänen piirteisiin, jolloin äänen muut ominaisuudet jäävät arvioimatta.

Koekuuntelu on laji, jota voidaan harjoitella ja oppia aivan kuten mitä tahansa muutakin taitoa. Samoin tarkasti kuuntelemisen taito on kyky, joka on toisilla luonnostaan parempi kuin toisilla [Toole 1982].

3.2.2 Koeympäristö

Huone, jossa kuuntelukokeet järjestetään, muodostaa merkittävän osan kuuntelijan koetilanteesta havaitsemasta ympäristövaikutelmasta. Huoneen akustiset ominaisuudet kuten kaikuisuus ja resonanssit täytyy pitää määrättyjen raja-arvojen sisällä, jotta kuunteluhuoneesta ei aiheudu vääristymiä arvioitaviin ääniin. Asuinhuoneiden kokoisten kuunteluhuoneiden suurimpia ongelmia ovat yleensä resonanssit ja reunaefektit matalilla taajuuksilla [Toole 1982], vaikka kuunteluhuoneet yleisimmin ovatkin käytännön syistä juuri tätä kokoluokkaa. Resonansseja poistetaan tällöin akustiikkalevyin ja matalien taajuuksien reunaefektejä voidaan lieventää seiniin asennettavin resonaattorein.

Kuulokkeilla kuunneltaessa huoneen akustiikka ei merkitse yhtä paljon kuin kaiuttimilla kuunneltaessa. Molemmissa tapauksissa tärkeää on kuitenkin se, että huone on rauhallinen ja hiljainen, eivätkä huoneen ulkopuolelta tulevat äänet häiritse kuuntelukoetta. Huoneen olisi hyvä olla lisäksi viihtyisä, jotta kuuntelija voi rentoutua koetta tehdessään eivätkä huoneen visuaaliset ominaisuudet häiritse hänen keskittymistään.

Merkittävä vaikutus kokeen onnistumiselle on niin ikään äänentoistolaitteilla. Toteutettiinpa koe kaiuttimilla tai kuulokkeilla, on varmistettava, että niiden äänentoisto-ominaisuudet ovat kaikilla taajuuksilla riittävät ja tasalaatuiset kokeen tarkoitusta varten. Kaiuttimien tai kuulokkeiden lisäksi ääneen vaikuttavat epätäydellisyydet ja häiriöt vahvistimessa ja soittimessa sekä näiden välisissä liitännöissä, joiden kaikkien kunto ja laatu on syytä selvittää koetta suunniteltaessa.

3.2.3 Testattavien laitteiden tai algoritmien valinta

Kuuntelukokeiden avulla voidaan testata laitteiden tai algoritmien, esimerkiksi kaiuttimien tai koodekkien, paremmuutta toisiinsa nähden vertailemalla niiden tuottamia ääniä. Voidaan myös etsiä jonkin äänenkäsittelyalgoritmin parametreille oikeita arvoja vertailemalla hieman eri parametrein tuotettuja ääniä.

Kuuntelukokeeseen mukaan otettavien laitteiden, algoritmien tai parametrien valinnassa täytyy ottaa huomioon ihmisen suhteellinen tapa arvioida kuulemaansa. Kuuntelija sopeuttaa arvostelunsa kuulemiensa äänten mukaisesti siten, että arviot keskittyvät kokeessa mukana olevien äänten keskimääräisen laadun mukaisen painopisteen molemmin puolin niin laajalle kuin testimateriaalin laatuero vaativat. Jotta eri testien tulokset ja testeissä havaitut erot olisivat vertailtavissa keskenään on kokeeseen valittava jokseenkin koko laatuasteikon kattava joukko laitteita, algoritmeja tai parametreja [Toole 1985].

Lisäksi on hyvä ottaa testijoukkoon mukaan joitakin ääniä, joiden laatu on ennalta tunnettu. Nämä lisä-äänit ankkuroivat kuuntelijoiden käyttämät laatuasteikot, jotka voidaan tulosten analysointivaiheessa normalisoida laadultaan tunnetun ääninäytteen avulla. Näin voidaan vähentää eri kuuntelijoiden arvostelutapojen hajontaa.

3.2.4 Testissä käytettävä äänimateriaali

Kuunneltava aineisto on syytä valita tarkoituksenmukaisesti siten, että testitulokset kertovat täyttääkö laite tai algoritmi sille asetetut vaatimukset. On eri asia testata elävän äänen laatua tavoittelevia huippukaiuttimia tai mahdollisimman pieneen bittimäärään pyrkiviä äänenkoodausalgoritmeja. Ihmiskuulon peitto-ominaisuudet saattavat estää joitakin ominaisuuksia kuulumasta tietyillä äänimateriaaleilla, mutta toisilla äänillä ominaisuudet tulevat esiin. Näin ollen äänimateriaalin tulee olla kaikissa tapauksissa riittävän vaativaa ja monipuolista, jotta testattavan laitteen tai algoritmin kaikki ominaisuudet tulevat testissä esille.

Ensimmäinen vaatimus kuuntelukokeissa käytettävälle äänimateriaalille on mahdollisimman hyvä tallenteen laatu. Jos tallenne on epätäydellisempi, kuin mitä arvioitava laite tai menetelmä kykenee toistamaan, niin silloin kritiikki voi koskea yhtä hyvin tallennetta kuin arvioitavana olevaa laitettakin.

Jos kuuntelukokeen tarkoituksena on verrata tuotettua ääntä ns. täydelliseen eli alkuperäiseen ääneen, on hyvä, jos koekuuntelijoilla on mahdollisuus olla paikalla tallennustilanteessa. Toisaalta ihmisen pitkäaikaismuisti ei välttämättä riitä muistamaan alkuperäisestä äänestä paljoakaan kuuntelukoetilanteessa. Ennestään tuttu äänimateriaali kuitenkin poistaa yhden muuttujan koetilanteesta [Toole 1982].

Kysymykseen: ”Miten pitkiä kuunneltavien ääninäytteiden tulisi olla?” ei ole täysin tyydyttävää vastausta olemassa. Pitkä ääni antaa kuuntelijalle mahdollisuuden sopeutua

ja kuulla tarkasti kuunneltavan äänen ominaisuudet, mutta hyvin pitkä ääni peittää muut muistissa olleet äänet ja näin vaikuttaa seuraavien äänten vertailuun. Liian lyhyt ääninäyte taas estää kuuntelijaa tekemästä tarkkoja havaintoja äänestä [Toole 1985].

On havaittu, että äänenvoimakkuus ja kuuntelijoiden antamat arvosanat korreloivat merkittävästi. Tämän takia on erityisen tärkeää, että kaikki kokeessa käytettävät äänet ovat yhtä voimakkaita [Toole 1985]. Äänenvoimakkuuden on oltava riittävä, jotta ääni kaikkine vivahteineen tulee hyvin kuulluksi. On kuitenkin varottava liian voimakasta ääntä, jotta vältetään kuulijoiden hetkellinen kuulokynnyksen nousu. Äänenvoimakkuuden tulisi olla realistinen – luonnollista ääntä jäljittelevä [Toole 1982]. Käytännössä kuitenkin särö ja tallenteiden kohina pakottavat käyttämään pienempää äänenvoimakkuutta, joka voidaan valita kuuntelijoiden mieltymyksen mukaan. Äänenvoimakkuutta voidaan mitata esimerkiksi käyttämällä A-painotettua vaaleanpunaista kohinaa. Keskimääräinen äänenvoimakkuus on toisaalta subjektiivinen kokemus, jonka määrittäminen matemaattisin keinoin ei välttämättä aina tuota toivottua tulosta.

3.2.5 Arvioitavien ominaisuuksien valinta

On tärkeää valita selvästi ne parametrit, joista halutaan kuuntelijan antavan arvionsa. Samoin on tärkeää määrittää täsmällisesti millaisin asteikoin ja kuinka tarkasti kutakin ominaisuutta arvioidaan. Lisäksi on hyvä kuvata sanallisesti arviointiasteikkoa, jotta jokainen kuuntelija käyttäisi arviointiasteikkoa edes jokseenkin samalla tavalla.

Vaikka arvioitaviksi halutut tärkeimmät ominaisuudet listataan, on hyvä jättää kuuntelijoille mahdollisuus muihinkin kommentteihin, ettei vastauskaavake estä kuuntelijoita välittämstä todellista kokemustaan äänestä kokeen järjestäjälle [Toole 1982].

Kuitenkaan vastauskaavaketta ei saa jättää liian avoimeksi. Jos ei selvästi määrätä mitä äänen ominaisuuksia halutaan arvioitavan, saattavat kuuntelijat keskittyvät vain kaikkein selkeimpiin äänten ominaisuuksiin, jolloin äänen hienovaraisemmat piirteet jäävät arvioimatta. Lisäksi kuuntelijoiden vapaasti muotoilemia sanallisia vastauksia, joista osa on kuvainnollisia ja osa teknisiä ilmauksia, on vaikea kääntää selkeästi analysoitavaan muotoon. Testi on sitä parempi, mitä yksinkertaisempia arvioita kuuntelijalta vaaditaan [Lipshitz 1981].

Äänen ominaisuuksia kuvaavia adjektiiveja, joita käsiteltiin äänen laatua koskevassa kappaleessa, on vertailtu tilastollisesti ja todettu, että monet niistä korreloivat voimakasti keskenään [Toole 1982] [Gabrielsson & Sjögren 1979]. Tällaisista äänen ominaisuuksia kuvaavista adjektiiveista vain kahdeksan on todettu riittävän riippumattomiksi toisistaan. Nämä äänen ominaisuudet ovat: selvyys, jota voidaan kuvata myös sanalla tarkkuus; terävyys, jota voidaan kutsua myös kovuudeksi, ja jonka vastakohtana on äänen pehmeys; kirkkaus, jonka vastakohtana muodostaa äänen tummuus; täyteläisyys, jonka puute aiheuttaa äänen ohuuden; tilantuntu; läheisyys;

häiritsevät äänet ja äänenvoimakkuus. Monimutkaisin, mutta samalla yleisin tapa arvioida äänenlaatua on määrittää yksinomaan sen tarkkuutta [Toole 1982].

3.2.6 Kuuntelukoetilanne

Kuuntelukokeiden toteutus on suunniteltava kiireettömäksi tilanteeksi, jossa kuuntelijat saavat häiriytymättä arvioida äänten ominaisuuksia. Kuunteluaikatauluun on varattava aikaa kuuntelijoiden lepäämiselle, sillä tarkassa kuuntelutyössä keskittymiskyky herpaantuu nopeasti. Jos kuunteluistunnot venyvät liian pitkiksi, tulee arvioista epätarkkoja ja jopa virheellisiä.

Kuuntelijoille pitää antaa selkeät ohjeet kokeen kulusta, ja on varmistettava, ettei heillä ole epäselvyyttä kokeen suorittamisesta eikä äänien arviointitavasta. Jos mahdollista, heille olisi annettava mahdollisuus harjoitella kuuntelua ja parametrien arvioimista ennen varsinaisen kokeen aloittamista.

Olisi myös suotavaa, että kuuntelija voisi itse määrätä milloin ottaa uuden äänen kuunneltavakseen. Riippuu tosin kokeen luonteesta onko se mahdollista. Kuuntelija voi paremmin rentoutuneena keskittyä jokaiseen ääneen, jos liukuhihnamaisesti etenevä äänten jono ei pakota häntä harkitsemattomiin ja äkkinäisiin arvioihin. Lisäksi jos erityisestä syystä ei ole tarvetta rajata mahdollisuuksia kuunnella ääniä uudestaan, on hyvä jos kuuntelija voi vapaasti kuunnella kunkin äänen niin monta kertaa kuin kokee tarvitsevansa luotettavaa arviointia varten.

Kuuntelijoihin ei pidä vaikuttaa millään tavalla kuuntelukokeen aikana, vaan heille on taattava täydellinen keskittymisrauha kokeen suorituksen ajaksi. Jos kuuntelukokeet suoritetaan ryhmissä, on ryhmän keskinäistä kommunikaatiota rajoitettava, jotta mittauksessa ei syntyisi ryhmäpaineen aiheuttamaa virhettä. Paras vaihtoehto on kuunteluttaa koehenkilöt yksitellen [Toole 1985].

Ihminen arvioi äänen ominaisuuksia yleensä viimeisimpiin muistissa oleviin ääniin vertaamalla. Kunkin äänen saamaan arvosteluun vaikuttaa se, minkä laatuista ääniä kuuntelijalla on lyhytmuistissaan vertailukohtina. Vertailtavia ääniä tulee tästä syystä soittaa kuuntelijoille vaihtelevassa järjestyksessä, jotta mikään laite tai algoritmi ei saa etulyöntiasemaa paremmasta sijainnistaan kuuntelujärjestyksessä.

3.3 Testausmenetelmiä

Haluttaessa arvioida äänen laatua subjektiivisesti voidaan äänten kuuntelu ja arviointi järjestää monella tavalla. Ääniä voidaan kuunnella joko yksitellen tai monta kerrallaan ja arvioinnin toteuttamiseen on olemassa monia eri arvointitapoja ja arvosana-asteikoita. Kuuntelijalle voidaan myös antaa ennakkotietoja kuulemistaan äänistä tai hänen voidaan antaa arvioida äänet tietämättä äänten alkuperää.

Kuten muutkin kuuntelukokeisiin liittyvät muuttajat, on testausmenetelmä valittava kunkin sovelluksen testaukseen erikseen siten, että sen avulla parhaiten selviävät ne äänen ominaisuudet, jotka kuuntelukokein halutaan määritellä.

3.3.1 Sokkotestit

Kuuntelijoiden tai kokeenjärjestäjien ennakkosenteet arvioitavia laitteita tai menetelmiä kohtaan voivat vääristää voimakkaasti kokeen tuloksia, ellei asenteiden vaikutusmahdollisuuksia oteta huomioon ja poisteta. Jotta ennakkomieliä ei pääsisi vaikuttamaan arviointeihin, voidaan kokeet järjestää sokkotesteinä. Sokkotestissä ei tiedetä missä järjestyksessä arvioitavia ääniä kuunnellaan. Kun mielipiteitä nostattavaa laitetta tai algoritmia ei tunnusteta arvioitavien äänten joukosta, eivät ennakkomieliä pääse vaikuttamaan arviointiin. On todettu [Lipshitz 1981], että kokeet on toteutettava vähintään yksöissokkotesteinä, mutta mieluiten kaksoissokkotesteinä, jotta tulokset olisivat luotettavia.

Kaksoissokkotesti tarkoittaa sitä, etteivät niin kuuntelijat kuin kokeen järjestäjäkään tiedä, missä järjestyksessä arvioitavat menetelmät tai laitteet esiintyvät kuunneltavien äänten joukossa. Näin kuuntelijan on suoritettava arviointinsa tietämättä mistä arvosteltava ääni on peräisin ja voimatta käyttää arvostelussa ennakkotietojansa – asenteitansa. Samoin kokeen järjestäjänkään ennakkotiedot ja –asenteet eivät voi kaksoissokkotestissä vaikuttaa esimerkiksi siihen, missä järjestyksessä hän kuunteluttaa ääniä tai siihen, mitä hän sanattomasti viestittää kokeen suorittajalle. Kaksoissokkotesti poistaa asenteiden vaikutusmahdollisuuden näin molempien osapuolien suunnalta.

Yksöissokkotestissä kuuntelija ei tiedä missä järjestyksessä ääniä kuunnellaan, mutta kokeen järjestäjällä on tieto äänten kuuntelujärjestyksestä. Koe voidaan järjestää yksöissokkotestinä, jos kaksoissokkotestin toteuttaminen olisi kohtuuttoman hankalaa. Asenteiden vaikutusmahdollisuutta voidaan yksöissokkotestissä vähentää mm. minimoimalla kokeen aikaista kommunikointia kokeen järjestäjän ja suorittajan välillä.

3.3.2 Yksittäiskuuntelukoe

Yksinkertaisin tapa arvioida äänenlaatua subjektiivisesti on toteuttaa kuuntelukoe, jossa kuuntelijalle sovitetaan yksi ääni kerrallaan ja hän arvioi kuulemansa äänen laadun valitun arviointiasteikon mukaisesti. Helppoutensa takia yksittäisarviointi on hyvin tavallinen tapa toteuttaa subjektiivinen kuuntelukoe.

Yksittäiskuuntelukokeessa näytteiden keskinäisen järjestyksen on erityisesti oltava sellainen, että kaikenlaatuiset näytteet ovat jakautuneet satunnaisesti koko äänijoukkoon, sillä kuulija arvostelee kunkin äänen verraten niitä juuri aiemmin kuulemiinsa ääniin. Mieluiten äänten järjestys saisi olla jokaisen koehenkilön kohdalla

erilainen, jotta kuuntelujärjestyksen aiheuttamat vääristymät poistuisivat tilastollisesti kaikkien koehenkilöiden tuloksia yhdistettäessä.

Yksittäiskuuntelukokeiden arviointikaavakkeissa on erityisen tärkeää kuvailla kutakin arvosanaa sanallisesti, jotta kuuntelijan olisi helpompi asettaa jokainen näyte oikeaan kohtaan arviointiasteikossa. Ilman sanallisia arvosanakuvauksia kuuntelijat mitoittavat arvostelunsa arviointiasteikolle hyvin erilaisilla tavoilla.

3.3.3 Parivertailutestit

Yksittäisarviointikokeen lisäksi yleisimpiä kuuntelukokeita ovat parivertailutestit, jossa arvioidaan kahden äänen keskinäisiä laatueroja [Toole 1982]. Parivertailutesteissä ääniä kuunnellaan yleisimmin pareittain tai kolmen ryhmässä. Parivertailutesteiksi voidaan kutsua myös kuuntelukokeita, joissa vertaillaan useampia näytteitä kerrallaan ja laitetaan ne jonkin ominaisuuden mukaiseen järjestykseen tai valitaan joukosta esimerkiksi miellyttävin ääni.

Tyypillisimpiä parivertailutestejä ovat AB ja ABX-menetelmät [Toole 1985]. Yksinkertaisin parivertailukoe on AB-menetelmä, jossa kuuntelija valitsee kahdesta näyteäänestä toisen jollakin perusteella tai antaa arvionsa molemmista näytteistä. AB-koe voidaan toteuttaa myös niin, että toinen näytteistä on referenssinäyte, johon toista verrataan. Kuulijalle voidaan kertoa, kumpi näytteistä on referenssi, jolloin hän arvioi, kuinka paljon toinen näyte muistuttaa sitä, tai voidaan myös jättää kertomatta, kumpi näyte on referenssi, jolloin kuulijan on arvosteltava molemmat näytteet.

ABX-menetelmässä koekuuntelija kuulee kolme näyteääntä. X-näyte on aina sama ääni kuin joko A- tai B-näyte ja kuuntelijan on pääteltävä kumman, A:n vai B:n, kanssa X-ääni on identtinen. Koe voidaan toteuttaa myös niin, että kuuntelijan ei ole yksiselitteisesti kerrottava, kumman hän epäilee olevan identtinen X-näytteen kanssa, vaan hänen on annettava arvionsa sekä A-, että B-näytteistä.

TAFC (two alternatives forced choice) –menetelmä on kuuntelukoe, jossa kuuntelijalla kuuntelutetaan kahta ääntä, joiden jotakin parametria muutetaan, kunnes kuuntelija ei enää huomaa eroa äänten välillä.

Epätieteellisissä testeissä vertaillaan usein monia ääniä samanaikaisesti ja ympäristömuuttujia ei raportoida lainkaan. Tällaisilla kuuntelukokeilla ei luonnollisesti ole tieteellistä merkitystä.

3.3.4 Arviointiasteikkoja

Parametrien arvioinnissa voidaan käyttää monenlaisia asteikoita. Eräs suositeltava (IEC 1995) asteikko on arvosana 0 – 10 siten, että arvosana annetaan korkeintaan yhden desimaalin tarkkuudella [Toole 1982]. Arvosana 0 tarkoittaa, että arvioidulla äänellä ja

sen ideaalisella vastineella ei ole mitään yhteistä. Arvosana 10 taas kertoo, että ääni on täsmälleen ideaalisen kaltainen, eikä sitä olisi enää mitenkään mahdollista parantaa. Näitä ääriarvosanoja ei tulisi lainkaan käyttää todellisissa kuuntelukokeissa.

Havaittavuustesteissä, joissa tutkitaan jonkin ominaisuuden kuultavuutta arvioitavassa äänessä, sopivat vastausvaihtoehdot ovat: häiritsevä, havaittava ja ei-havaittava [Koistinen 1998].

MOS (Mean Opinion Score) –luokituksia on olemassa monenlaisia. ITU-T:n suosituksessa P.800 on esitetty muutamia puhelinkaistasovelluksien testaukseen soveltuvia MOS-arviointiasteikoita, jotka on esitelty taulukoissa 1-3. MOS-arviointiasteikko sisältää arvosanat 1 – 5.

Äänen laadun arviointiin yleisesti sopii tavanomainen MOS-arviointiasteikko, joka on esitetty taulukossa 1. Arvosana 5 tarkoittaa täydellisesti luonnollisen tai alkuperäisen mukaista ääntä, jonka laatua ei ole enää mahdollista parantaa ja arvosana 1 vastaa puheenlaatua, jossa sanoma jää ymmärtämättä. Puhelimen välittämän äänen laatu on MOS-asteikolla arvioituna noin 3,5 riippuen käytettävästä puheenkoodausmenetelmästä.

Taulukko 1. Subjektiiivisissa kuuntelukokeissa käytettävä MOS (mean opinion score) –asteikko puheen laadun arviointia varten. Arvosana 5 tarkoittaa täydellisesti luonnollisen tai alkuperäisen mukaista ääntä, jonka laatua ei ole enää mahdollista parantaa ja arvosana 1 vastaa puheenlaatua, jossa sanoma jää ymmärtämättä.

| Puheen laatu | MOS |
|---------------------|------------|
| Erinomainen | 5 |
| Hyvä | 4 |
| Tyydyttävä | 3 |
| Välttävä | 2 |
| Huono | 1 |

Taulukko 2. Laadultaan heikon puheäänien laadun arvioimiseen sopiva kuuntelemisen työläyttä määrittävä MOS-LE (mean opinion score for listening effort) –asteikko.

| Kuuntelemisen työläys (listening effort) | MOS-LE |
|---|---------------|
| Täydellinen rentoutuminen mahdollista; ei vaadi minkäänlaista ponnistelua | 5 |
| Ei vaadi erityistä ponnistelua | 4 |
| Vaatii kohtuullista ponnistelua | 3 |
| Vaatii huomattavaa ponnistelua | 2 |
| Sanoma jää ymmärtämättä ponnistelusta huolimatta | 1 |

Taulukko 3. Äänekkyyden arvioimisessa käytettävä MOS-LP (mean opinion score for loudness preference) -arvosana-asteikko. Äänenvoimakkuus on määriteltävissä matemaattisesti, mutta ihmisen kokemaa äänekkyyttä on useimmiten erilainen kuin laskettu äänenvoimakkuus.

| Äänekkyyys (loudness preference) | MOS-LP |
|---|---------------|
| Hyvin paljon äänekkäämpi kuin haluttu | 5 |
| Äänekkäämpi kuin haluttu | 4 |
| Sopivan äänekäs | 3 |
| Hiljaisempi kuin haluttu | 2 |
| Hyvin paljon hiljaisempi kuin haluttu | 1 |

Taulukossa 2 on esitelty kuuntelemisen työläyden arviointia varten räätälöity MOS-LE (listening effort) –asteikko, jota voidaan käyttää laadultaan heikkojen äänten arviointiin.

Äänekkyyden arviointia varten on olemassa MOS-LP (loudness preference) –asteikko, jonka avulla voidaan määrätä kuuntelijoiden suosima äänekkyytystaso. Vaikka äänenvoimakkuuden määrittämiseksi on olemassa matemaattisia keinoja, eivät ne kuitenkaan kerro ihmisen kokemaa äänekkyyttä. MOS-LP –arvosana-asteikko on esitelty taulukossa 3.

DCR (degradation category rating) –menetelmää, eli huonontuman luokitus –menetelmää käytetään A-B –testeissä, joissa toinen kuunneltavista näytteistä on hyvälaatuinen referenssinäyte ja toisen äänenlaatua arvioidaan taulukossa 4 esitetyn DMOS (degradation mean opinion score) –asteikon mukaisesti.

CMOS (comparison mean opinion score) –asteikkoa, joka on esitelty taulukossa 5, käytetään vertailevassa luokituksessa, jossa arvioitavan näytteen laatu voi olla joko parempi tai huonompi kuin referenssinäytteen äänenlaatu. CMOS-asteikkoa käytettäessä saattaa syntyä ongelmia siitä, että arvosanojen merkityksiä ei ole yksiselitteisesti määritelty, joten kuuntelijat saattavat mitoitaa arvosanaskaalan hyvin erilaisilla tavoilla.

Taulukko 4. DMOS (degradation mean opinion score) -arviointiasteikko, jota käytetään äänen laadun huonontumaa määrittelevässä DCR (degradation category rating) –menetelmässä, jossa arvioitavan äänen laatua verrataan alkuperäiseen referenssiääneseen. Arvosana 5 vastaa täysin alkuperäisen kaltaista ääntä.

| Huonontuma | DMOS |
|-----------------------------------|-------------|
| Ei kuultavissa | 5 |
| Kuultavissa, mutta ei häiritsevää | 4 |
| Hieman häiritsevää | 3 |
| Häiritsevää | 2 |
| Erittäin häiritsevää | 1 |

Taulukko 5. Puheen laadun subjektiiviseen arviointiin käytettävä CMOS (comparison mean opinion score) –asteikko, jota voidaan käyttää vertailtaessa arvioitavan äänen laatua alkuperäiseen referenssiääneen.

| Puheen laatu | CMOS |
|---------------------|-------------|
| Paljon parempi | +3 |
| Parempi | +2 |
| Hieman parempi | +1 |
| Yhtä hyvä | 0 |
| Hieman huonompi | -1 |
| Huonompi | -2 |
| Paljon huonompi | -3 |

3.4 Kuuntelukoetulosten analysointi

Koetulosten analysoinnissa huomioon otettavien muuttujien lukumäärä riippuu mm. käytetystä koemenetelmästä, arvioiduista parametreista ja arvioinnissa käytetyistä arvosana-asteikoista. Myös kuuntelijoiden lukumäärä vaikuttaa tilastollisen analysoinnin tulosten luotettavuuteen. Kaiken kaikkiaan tulosten luotettavuuteen vaikuttavat monet suorat ja epäsuorat kuuntelukokeen muuttujat, joiden vaikutuksia olisi hyvin vaikea huomioida tulosten analysoinnissa. Niinpä subjektiivisten kuuntelukokeiden tuloksiin on aina suhtauduttava tietyllä varauksella ja ymmärrettävä koetulosten suurpiirteisyys.

Koetulosten analysointia vaikeuttaa se tosiasia, että sama fysikaalinen mekanismi voi eri aikoina aiheuttaa hyvinkin erilaisia vastauksia koekuuntelijoilta. Sama ääni voidaan kokea sekä positiivisena että negatiivisena, jolloin eri kuuntelijoiden arviot voivat olla keskenään selvästi ristiriitaisia [Toole 1982]. On huomattava, että jotkut hyvin erilaiselta näyttävät arviot saattavat todellisuudessa olla ilmauksia samasta kuulohavainnosta.

Koehenkilöiden kuulohavainnoissa esiintyvät virheet voidaan jakaa kahteen ryhmään. Ensimmäisen ryhmän virheet ovat sellaisia, joissa kuuntelija kuulee äänessä jonkin ominaisuuden, jota siinä ei ole olemassa. Toisen ryhmän kuulovirheissä taas jää kuulijalta jokin äänessä oleva ominaisuus huomaamatta.

Arvosanojen normalisointi

Eri kuuntelijoilla on erilaisia mielipiteitä siitä, millaisella lukualueella erilaatuisia ääniä tulisi arvioida. Toiset varaavat 0 – 10 lukualueesta luvut 4 – 9 huippukaiuttimien arviointiin ja mahdolltavat puhelinlaadun, koodekkien välittämän äänen ja kannettavien

radioiden äänenlaadun lukuihin 1 – 3. Toiset taas käyttävät lukualuetta hieman lineaarisemmin [Toole 1982]. Pienen otoksen kokeissa tämä voi muodostua ongelmaksi, mutta suuremmissa otoksissa käytettyjen lukualueiden vaihtelulla ei yleensä ole suurta merkitystä.

Koetulokset voidaan normalisoida kuuntelijoiden käyttämien erilaisten arviontiskaalojen tasavertaistamiseksi [Toole 1985]. Arviointiasteikoiden uudelleenmitoituksessa on tapauskohtaisesti tutkittava, onko asteikoita syytä muuttaa lisäämällä tai kertomalla, vai ovatko eri henkilöiden käyttämät asteikot logaritmisessa suhteessa toisiinsa.

Binomijakauman avulla arvioitu ABX-testin tulosten luotettavuus

ABX-kokeiden tulosten arvioinnissa on tärkeää erottaa todelliset oikeat vastaukset oikein arvatuista vastauksista [Burstein 1989]. Osa arvioijan vastauksista on oikein siksi, että hän on kuullut selvästi, kumpi arvioitavista näytteistä on sama kuin vertailunäyte. Osa vastauksista taas osuu oikeaan, vaikka arvioija ei olekaan kuullut eroa arvioitavien näytteiden välillä, vaan on arvannut vastauksen oikein. Arvaamisen tapauksessa voidaan olettaa, että todennäköisyys oikeaan osumiselle on $1/2$. Koehenkilön arvioinnin luotettavuudesta voidaan laskea piste-estimaatti kaavan (3) mukaisesti olettamalla että tasan puolet koehenkilön arvauksista on osunut oikeaan ja tasan puolet väärään, ja loput oikeaan osuneet tapaukset hän on todella kuullut. P_c on oikeaan osuneen vastauksen kokonaistodennäköisyys ja P_k on todella oikein kuullun vastauksen todennäköisyys.

$$P_c = P_k + \frac{1}{2}(1 - P_k) = \frac{1}{2}(P_k + 1) \quad (3)$$

Kun lausekkeeseen (3) sisällytetään P_c :n estimaatti c/n , jossa c on oikeiden vastausten lukumäärä ja n on vertailujen lukumäärä, saadaan lasketuksi estimaatti P_k :lle kaavan (4) mukaisesti.

$$P_k = 2\frac{c}{n} - 1 \quad , \text{ kun } c/n \geq \frac{1}{2} \quad (4)$$

Binomijakauman avulla määritellään P_k :n estimaatin ympärille halutulla luotettavuudella varmuusalue, jolle todellisten oikein kuultujen äänten lukumäärä osuu. Taulukoiden [Burstein 1971] avulla, tehtyjen kokeiden lukumäärän ja oikein osuneiden vastausten perusteella voidaan arvioida oikein kuultujen äänien lukumäärää halutulla varmuudella.

4 TUNNETTUJA TAAJUUSKAISTAN LAAJENNUSMENETELMIÄ

Tässä luvussa käsitellään joitakin sellaisia taajuuskaistan laajennusmenetelmiä, jotka oli esitelty ennen tässä diplomityössä kehitettyä menetelmää.

Esitellyt menetelmät voidaan karkeasti jakaa kahteen pääryhmään. Toinen taajuuskaistan laajennusmenetelmien joukko käyttää hyväkseen perinteisiä signaalinkäsittelymenetelmiä kuten uudelleennäytteistystä, suodatusta ja lineaarista ennustusta (linear predictive coding, LPC) [Rabiner&Schafer 1978], ja toinen ryhmä pohjautuu tilastollisten koodikirjojen käyttöön.

Perinteiseen signaalinkäsittelyyn perustuvat taajuuskaistan laajennusmenetelmät eivät ota kantaa siihen, mitä ylemmillä taajuuksilla todellisuudessa on ollut. Sen sijaan menetelmien toimivuus perustuu siihen, ettei tällä yläkaistalla ole juurikaan merkitystä puheen ymmärrettävyyden kannalta, vaan sattumanvaraisetkin korkeat taajuuskomponentit tekevät puheesta luonnollisempaa, ja siten kuuntelukokeiden perusteella parempilaatuista kuin kapeakaistainen puhe. Pääosaa perinteiseen signaalinkäsittelyyn perustuvissa taajuuskaistan laajennusmenetelmissä esittää hallittu spektrin laskostaminen.

Kaikki kolme esiteltävää tilastollista taajuuskaistan laajennusmenetelmää perustuvat viime kädessä koodikirjoihin. Menetelmän suunnitteluvaiheessa tuotetaan tilastollisin menetelmin koodikirjat, jotka kertovat päätelaitteeseen tulevan kapeakaistaisen puhesignaalin perusteella millainen signaalin yläkaista todennäköisimmin olisi. Koodikirjamenetelmin on mahdollista saavuttaa parannuksia kapeakaistaisen puheen laatuun laboratorio-olosuhteissa, mutta meluisassa ympäristössä niiden toimintavarmuutta on erittäin vaikea taata.

4.1 Perinteiseen signaalinkäsittelyyn perustuvia taajuuskaistan laajennusmenetelmiä

Perinteisellä signaalinkäsittelyllä tässä tarkoitetaan suoraviivaista taajuuskaistan muotoilua mm. hallitun laskostamisen, suodatuksen, uudelleennäytteistysten, tasasuuntauksen ja lineaarisen ennustuksen keinoin.

Tässä luvussa esitellyistä menetelmistä ensimmäinen ja toinen tuottavat ylätaajuudet hallitun laskostamisen avulla sekä muotoilevat yläkaistan luonnollisen kaltaiseksi. Kolmas menetelmä tuottaa tasasuuntauksen avulla yläkaistalle alkuperäisen signaalin harmonisia taajuuksia. Neljäntenä esiteltävä menetelmä arvioi LPC-kerrointen avulla

signaalin spektrin kaltevuutta ja luo tämän tiedon pohjalta uudet LPC-kertoimet, joiden avulla laajakaistainen spektri voidaan tuottaa.

Nämä kaikki perinteiseen signaalinkäsittelyyn perustuvat taajuuskaistan laajennusmenetelmät ovat japanilaisen Hiroshi Yasukawan esittelemiä NTT Transmission Systems Laboratories:ilta. Hän ei ole suunnitellut menetelmiään jatkuva-aikaisen signaalin käsittelyyn, vaan hän on tutkinut tajuuskaistan laajentamista koko signaalin osalta ikkunoimatta signaalia mitenkään.

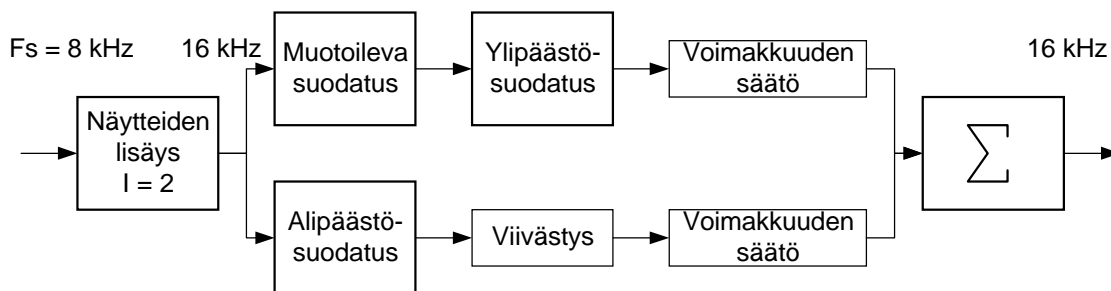
4.1.1 Laskostamiseen ja suodatukseen perustuva menetelmä

Ensimmäisen artikkelinsa puhelimen keinotekoiseen taajuuskaistan laajennukseen liittyen [Yasukawa 1994] Yasukawa on julkaissut vuonna 1994. Hänen ehdottamansa menetelmä on hyvin yksinkertainen ja parantaa puhesignaalin laatua esitettyjen kuuntelukoetulosten perusteella jonkin verran. Esitellyissä kuuntelukokeissa on tosin mitattu lähinnä ylemmän ja alemman taajuuskaistan voimakkuustasojen erotuksen vaikutusta subjektiiviseen äänen laatuun.

Menetelmän toteutus

Kaavio ensimmäisestä Yasukawan esittelemästä menetelmästä näkyy kuvassa 20. Kaikki operaatiot suoritetaan aika-alueessa. Aluksi otoksen näytteenottotaajuus nostetaan kaksinkertaiseksi ja tuotetaan 4 – 8 kHz:n taajuuskaista hallitun laskostamisen keinoin. Tämän jälkeen käsitellään alempi (0 – 4kHz) ja ylempi (4 – 8kHz) taajuuskaista erikseen. Lopuksi taajuuskaistat yhdistetään yhdeksi laajakaistaiseksi puhesignaaliiksi. Käytettävät suotimet ovat lineaarisia FIR-suotimia.

Ylempi taajuuskaista, 4 – 8 kHz muodostetaan siten, että aluksi hallitun laskostamisen avulla tuotettu laajakaistainen spektri muokataan muotoilusuotimella, jotta se muistuttaisi puheen spektriä vastaavilla taajuuksilla. Sen jälkeen spektristä erotetaan ylipäästösuotimella ylempi, eli 4 – 8 kHz:n taajuuskaista ja säädetään sen äänenvoimakkuus sopivaksi.



Kuva 20. Hiroshi Yasukawan vuonna 1994 esittelemä taajuuskaistan laajennusmenetelmä.

Alempi taajuuskaista otetaan käyttöön sellaisenaan alipäästösuodattamalla se 16 kHz:n signaalista. Näytteiden lisäys ja alipäästösuodatus yhdessä toteuttavat alkuperäiselle signaalille ylinäytteistyksen kertoimella $I = 2$. Ylemmän ja alemman taajuuskaistan prosessointien välinen aikaero joudutaan tasoittamaan viivästyttämällä alempaa taajuuskaistaa ennen signaalien summausta, sillä ylemmän taajuuskaistan tuottamiseen tarvittavat kaksi suodatusta vaativat pidemmän ajan kuin alemman kaistan yksi suodatus. Tämän jälkeen säädetään alemman taajuuskaistan äänenvoimakkuus halutulle tasolle. Lopuksi ylempi ja alempi taajuuskaista yhdistetään summaamalla signaalit.

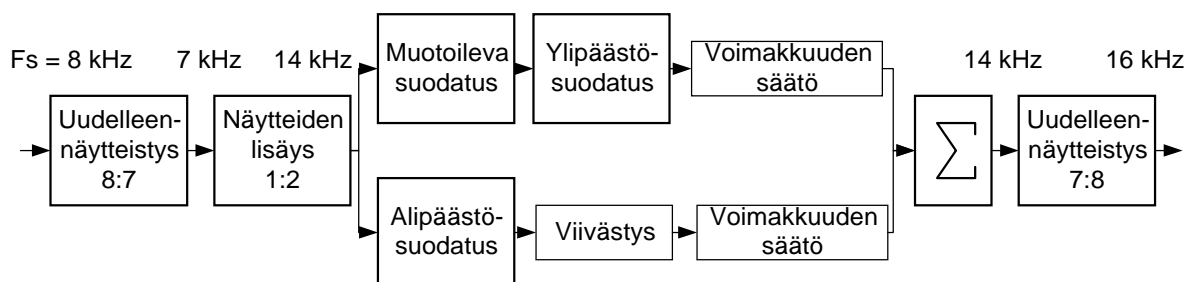
Yasukawa on mitannut menetelmän toimivuutta kuuntelukokein, joissa vertailujen avulla etsittiin sopivaa äänenvoimakkuuseroa alemman ja ylemmän taajuuskaistan välille. Yasukawa on tulosten perusteella päätellyt, että äänen laatu paranee eniten silloin, kun ylemmän kaistan voimakkuustaso asetetaan 0 – 12 dB vaimeammaksi kuin alemman.

Menetelmän parannettu versio

Koska puhelimen siirtotieltä saatava puhesignaalin spektri loppuu jo 3400 Hz:n kohdalla, jää spektriin välttämättä 1200 Hz:n levyinen aukko 3400 – 4600 Hz:n välille laajennettaessa signaalin taajuuskaistaa tämän Yasukawan kehittämän menetelmän avulla. Vuotta myöhemmin julkaisemassaan artikkelissa [Yasukawa 1995] Hiroshi Yasukawa on kehittänyt edelliseen taajuuskaistan laajennusmenetelmäänsä parannuksen, jonka avulla kavennetaan spektrin alemman ja ylemmän taajuuskaistan väliin jäävää aukkoa.

Kuvassa 21 esitetyn kaavion mukaisesti menetelmään on lisätty taajuusmuunnokset ennen ja jälkeen varsinaisen taajuuskaistan laajennuksen. Tässä tapauksessa on kyse todellisesta taajuusmuunnoksesta, jolloin laskostunut spektri suodatetaan pois.

Ennen varsinaista taajuuskaistan laajennusta näytetaajuus lasketaan 7000 Hz:iin, jolloin PCM-signaalin 300-3400 Hz:n taajuuskaistan peilikuva asettuu 3600 - 5700 Hz:n alueelle, ja alemman ja ylemmän taajuuskaistan väliin jäävä aukko kapenee näin 200 Hz:n levyiseksi. Taajuuskaistan laajennuksen yhteydessä näytetaajuus nousee 14000 Hz:iin ja lopuksi signaalin näytetaajuus nostetaan vielä 16000 Hz:iin.



Kuva 21. Yasukawan taajuuskaistan laajennusmenetelmäänsä esittelemä parannus.

Tämänkin menetelmän toimivuutta on tutkittu kuuntelukokein, jossa vertailuin etsittiin optimaalista eroa taajuuskaistojen äänenvoimakkuudelle. Parhaaseen tulokseen päästiin, kun ylempi kaista säädettiin 0 – 18 dB vaimeammaksi kuin alempi kaista.

4.1.2 Tasasuuntausta hyväksi käyttävä menetelmä

Vuonna 1996 Hiroshi Yasukawa jälleen julkaisi artikkelin [Yasukawa 1996a], jossa hän esitteli uuden perinteiseen signaalinkäsittelyyn pohjautuvan taajuuskaistan laajennusmenetelmän.

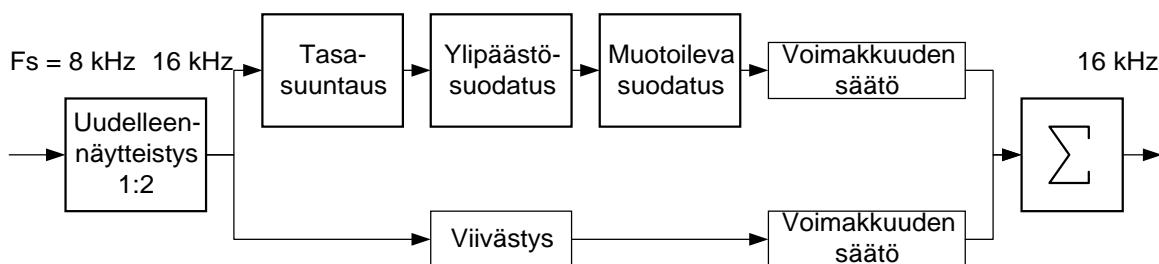
Yasukawan aikaisemmin kehittämissä menetelmissä korkeat taajuudet tuotettiin hallitun laskostamisen keinoin olemassa olevasta 300 – 3400Hz:n taajuuskaistasta. Tässä uudessa menetelmässä ylempi taajuuskaista tuotetaan tasasuuntaamalla alkuperäinen signaali. Muilta osin tämä menetelmä toimii samoin periaattein kuin Yasukawan aikaisempikin menetelmä.

Tasasuuntaus tuottaa alkuperäisessä signaalissa olevien taajuuksien harmonisia taajuuksia, joiden intensiteetti pienenee korkeampiin taajuuksiin päin mentäessä. Tällainen spektrirakenne muistuttaa ainakin vokaalien spektriä, joten teoriassa menetelmä saattaisi tuottaa suhteellisen luonnollista spektriä korkeille taajuuksille.

Toteutus

Kaavio tasasuuntaukseen perustuvasta taajuuskaistan laajennusmenetelmästä on esitetty kuvassa 22. Aluksi kapeakaistainen signaali uudelleennäytteistetään 16 kHz:n taajuudelle. Alempaa taajuuskaistaa ei tarvitse muokata muutoin kuin viivästämyllä näytteitä ylempään kaistan viiveen verran ja säätämällä äänenvoimakkuus sopivaksi.

Ylemmät taajuudet tuotetaan tasasuuntaamalla alkuperäinen signaali, ja syntyneestä spektristä suodatetaan alempi taajuuskaista, eli alle 3400 Hz:n taajuudet, ylipäästösuotimen avulla pois. Muotoileva suodin huolehtii siitä, että spektrin verhokäyrä taipuu kohti nollaa ennen 8 kHz:n taajuutta, etteivät korkeat taajuudet jää häiritsevän voimakkaiksi.



Kuva 22. Kaavio tasasuuntausta hyväksi käyttävästä menetelmästä.

Käytettävät suotimet ovat lineaarisen vaiheen FIR-suotimia, joiden rajataajuus on 3,4 kHz:n tienoilla. Käytettävän ylipäästösuotimen estokaista vastaa uudelleennäytteistyksessä tarvittavan alipäästösuotimen päästökaistaa.

Optimaalista voimakkuustasojen erotusta Yasukawa ei ole tämän menetelmän osalta tutkinut, mutta mainitsee, että se voidaan selvittää joko kuuntelukokein, tai minimoimalla ulostulevan signaalin säröä.

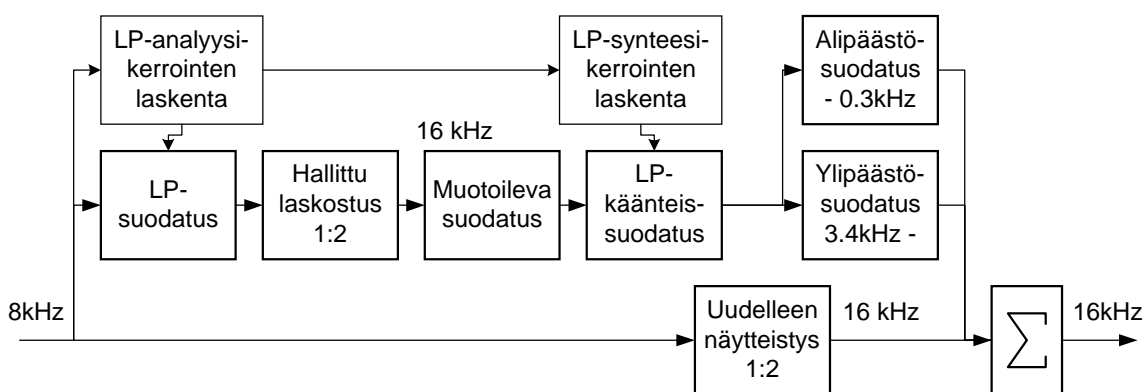
Tasasuuntausta käyttävän menetelmän etuina edellisiin menetelmiin verrattuna voidaan mainita ainakin kaksi asiaa. Tasasuuntauksen avulla tuotetut taajuudet ovat alempien taajuuksien harmonisia, mikä parantaa niiden sopivuutta lisättäviksi ylätaajuuksiksi. Toinen etu on se, että spektrissä ala- ja ylätaajuuskaistojen väliin ei jää aukkoa, mikä oli ongelmana Yasukawan aikaisemmissa menetelmissä.

4.1.3 LPC-kerrointen ekstrapolointiin perustuva menetelmä

Artikkelissaan [Yasukawa 1996b] Hiroshi Yasukawa käyttää taajuuskaistan laajennuksessa hyväkseen lineaarisen ennustuksen (LP, linear prediction) antamia mahdollisuuksia. Menetelmä perustuu lineaariseen spektrin ekstrapolointiin, joka aikaansaadaan LP-kertoimia käsittelemällä ja jäännössignaalia laskostamalla. Tämän menetelmän erikoispiirteenä Yasukawan aikaisempiin menetelmiin verrattuna on taajuuksien lisääminen myös matalalle, alle 300 Hz:n alueelle.

Menetelmän toimintaperiaate

Kuvan 23 kaavio kertoo menetelmän toimintaperiaatteen. Laajakaistainen signaali muodostetaan kolmesta taajuuskaistasta. Alkuperäinen PCM-signaali sisältää taajuudet 300 – 3400 Hz, jonka lisäksi luodaan taajuudet 3400 – 8000 Hz sisältävä ylempi taajuuskaista ja taajuudet 0 – 300Hz sisältävä alempi taajuuskaista.



Kuva 23. Kaaviokuva spektrin ekstrapolointiin perustuvasta menetelmästä.

Alkuperäisen signaalin sisältämät tajuudet 300 – 3400 Hz yhdistetään sellaisinaan laajakaistaiseen signaaliin. Ainoa toimenpide on näytteenottotaajuuden nostaminen 8 kHz:stä 16 kHz:iin.

Lisättävät ylä- ja alataajuudet tuotetaan LP-analyysin avulla. Aluksi alkuperäinen signaali suodatetaan LP-suotimella, jonka kertoimet lasketaan LP-analyysin avulla signaalin ominaisuuksista.

LP-suodatuksen jälkeen jäännösignaaliin tuotetaan taajuudet 4600 – 7700 Hz:n alueelle nostamalla näytteenottotaajuus 16 kHz:iin hallitun laskostamisen avulla. Muotoileva suodatus taivuttaa jäännösignaalin spektrin verhokäyrän kohti nollaa 8 kHz:ä lähestyttäessä, jotta ylätaajuudet eivät jää häiritsevän voimakkaiksi.

Käänteisen LP-suodatuksen avulla tuotetaan muokatusta jäännösignaalista laajakaistainen signaali, josta ali- ja ylipäästösuotimien avulla leikataan sopivat spektrilaajennukset alkuperäiseen signaaliin lisättäväksi.

LP-kerrointen laskenta

LP-analyysisuotimen kertoimet lasketaan alkuperäisen signaalin avulla optimoimalla virhesignaali mahdollisimman pieneksi. Käänteisen LP-suotimen kertoimet lasketaan interpoloimalla LP-analyysisuotimen kertoimien mukaista signaalin spektrin verhokäyrää.

Laajakaistaisen signaalin LP-kerrointen selvittämiseksi siirrytään aluksi taajuusalueeseen, jossa LP-analyysisuotimen kertoimista lasketaan spektrin verhokäyrän keskimääräinen kaltevuus. Arvioidun kaltevuuden mukaisesti spektriä ekstrapoloidaan lineaarisesti halutuille ylä- ja alataajuuksille. Lopuksi palataan takaisin aika-alueeseen.

Yasukawa mainitsee spektrin ekstrapolointiin perustuvan menetelmän eduksi pienemmän spektrivääristymän kuin edellisissä menetelmissä. Koekuuntelua hän ei kerro järjestäneensä, joten menetelmän kyvystä parantaa äänen laatua ei voida sanoa mitään.

4.1.4 Menetelmien ongelmia

Yasukawa käyttää menetelmissään [Yasukawa 1994] [Yasukawa 1995] [Yasukawa 1996a] [Yasukawa 1996b] hyvin pikiä digitaalisia suotimia, jotka aiheuttavat äänen etenemiseen pitkän viiveen. Esimerkiksi puhelinverkossa viiveet ovat erittäin häiritseviä, joten näitä menetelmiä ei ole mahdollista käyttää tällaisissa reaaliaikaisissa laitteistoissa.

Pitkistä suotimista johtuu myös toinen jatkuvan signaalin prosessointia haittaava tekijä. Suotimien pituuden takia jouduttaisiin reaaliaikaisessa toteutuksessa valitsemaan prosessointikehykselle kohtuuton pituus, mistä johtuva puskuroinnista aiheutuva viive on liian pitkä reaaliaikaiseen toteutukseen.

Artikkeleissaan Yasukawa ei kerro signaalin ikkunonnista lainkaan. Kuuntelukokeita

varten tuottamansa äänimateriaalin hän on todennäköisesti tuottanut ikkunoimatta signaalia lainkaan, jolloin mitään ikkunoinnista johtuvia häiriöitä ei ole jouduttu käsittelemään.

4.2 Puheen tilastollisiin ominaisuuksiin ja koodikirjoihin perustuvat menetelmät

Kuten aikaisemmin on jo todettu, perinteisin signaalinkäsittelymenetelmin tuotettu taajuuskaistan laajennus ei ota kantaa puheen yläkaistan todelliseen sisältöön, joten lisätyt ylätaajuudet voivat poiketa hyvin paljon puheen todellisesta spektristä. Tähän ongelmaan käyvät käsiksi tilastolliset taajuuskaistan laajennusmenetelmät.

Puheen spektriä tilastollisesti tutkimalla voidaan selvittää millainen yläkaistan spektri vastaa tiettyä kapeakaistaista ääntä. Näin ollen ideaalisissa olosuhteissa voidaan päästä erittäin hyviin tuloksiin puheen laadun parannuksen suhteen. Todellisuudessa ei kuitenkaan voida olettaa puheäänien olevan puhdasta, ja näin kaikkien mahdollisten taustäänien häiritsemän puheen mallintaminen tilastollisesti on lähes mahdotonta.

Esittelen tässä kappaleessa kuitenkin kolme menetelmää, jotka laajentavat äänen taajuuskaistaa tilastollisesti luodun koodikirjan avulla. Näistä menetelmistä on yksi paneutunut erityisesti CELP-koodatun äänen laadunparannukseen.

4.2.1 Tilastollinen menetelmä, joka olettaa yläkaistan joukoksi satunnaislähteitä

Yan Ming Cheng, Douglas O'Shaugnessy ja Paul Mermelstein Kanadan INRS-Telecommunications:lta ja Bell-Northern Research:stä ovat kehittäneet tilastolliseen spektrin luokitteluun perustuvan menetelmän [Cheng, O'Shaugnessy, Mermelstein 1992] [Cheng, O'Shaugnessy, Mermelstein 1994]. Menetelmä olettaa puheen syntyvän äärellisestä määrästä – käytännössä joistakin kymmenistä – satunnaislähteitä.

Kapeakaistaisesta signaalista saadun informaation perusteella tilastollinen korjausfunktio määrää kullekin yläkaistan lähteelle todennäköisyyden, jolla se osallistuu äänen tuottamiseen. Todennäköisyysmäärittelyn jälkeen kapeakaistaiseen puheeseen lisätään yläkaista, jossa jokainen satunnaislähde on mukana todennäköisyytensä mukaisella painokertoimella kerrottuna.

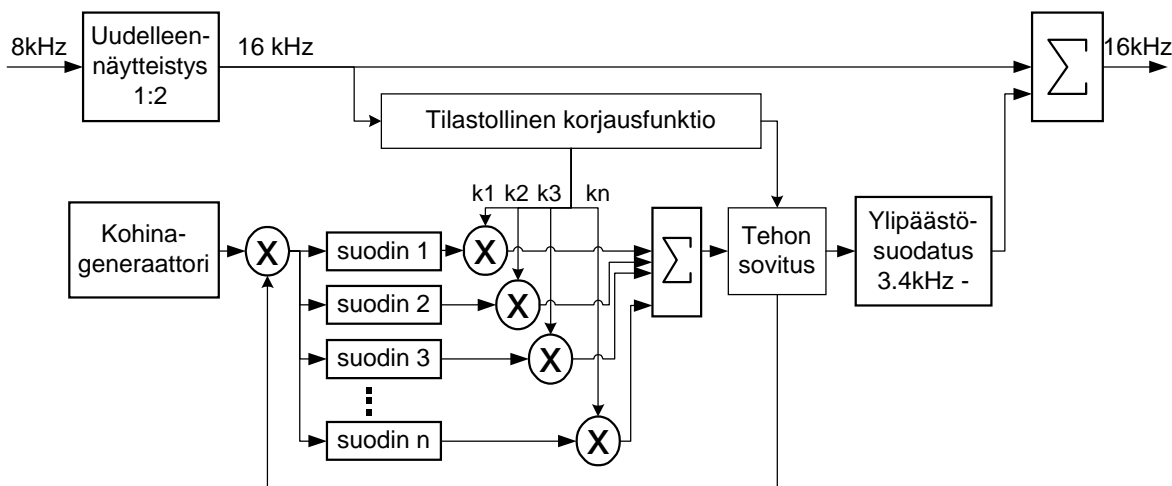
Menetelmän toteutus

Aluksi signaali on uudelleennäytteistettävä 16 kHz:n näytetaajuudella, mikäli on kyse puhelimesta tulevasta signaalista, jonka näytteenottotaajuus on 8 kHz. Signaali

käsitellään 20 ms:n mittaisina pätkinä, ja jokaisen näytepalan osalta etsitään kapeakaistaisen signaalin ja tilastollisen korjausfunktion avulla todennäköisyydet yläkaistan satunnaislähteille.

Menetelmän alkuoletuksen mukaiset satunnaislähteet, joille kullekin on määrätty sopiva taajuuskaista, tuotetaan yläkaistalle kohinageneraattorin ja suodinpankin avulla kuvan 24 mukaisesti. Kohinageneraattorin tuottamasta valkoisesta spektristä suodatetaan satunnaislähteiden taajuuskaistat. Yläkaistaa muodostettaessa kustakin suotimesta tulevan lähdekohinan voimakkuus kerrotaan todennäköisyyskertoimella, jonka tilastollinen korjausfunktio määrää alakaistan oletettujen satunnaislähteiden perusteella. Näin aikaansaatuja satunnaislähteiden äänet yhdistetään, niiden yhteisenergia sovitetaan ja saadusta spektristä suodatetaan alle 3,4 kHz:n alueelle levinnyt spektrin osa pois.

Lopuksi alkuperäinen alakaista ja tuotettu yläkaista yhdistetään laajakaistaiseksi signaaliksi.



Kuva 24. Kaavio tilastollisesta menetelmästä.

Tilastollinen korjausfunktio

Tilastollinen korjausfunktio, joka käytännössä toteutetaan koodikirjana, luodaan opetusaineiston avulla. Opetusalgoritmi optimoi ylä- ja alakaistat vastaamaan toisiaan halutulla määrällä satunnaislähteitä. Koetulosten perusteella riittäväksi lähteiden määräksi alakaistalle on todettu 64 ja yläkaistalle 16. Opetusalgoritmi tuottaa joukon vektoreita, joiden kertoimet määrittelevät kunkin satunnaislähteen olemassaolotodennäköisyyden. Satunnaislähteiden todennäköisyyksien lisäksi korjausfunktio määrittelee ala- ja yläkaistojen energioiden suhteen.

Korjausfunktion tuottava harjoitusalgoritmi käy opetusaineiston läpi 20 ms:n mittaisina otoksina. Algoritmi etsii todennäköisimmät satunnaislähteiden yhdistelmät päivittämällä tarvittavia suureita kunkin signaalinpätkän kohdalla.

Algoritmi laskee ala- ja yläkaistan lähteiden yhteistodennäköisyydet sekä alakaistan lähteiden todennäköisyydet käsiteltävänä olevan signaalikeyhyksen puitteissa. Näiden avulla päivitetään kaikille lähdepareille todennäköisyys, joka kertoo osallistuuko jokin yläkaistan lähde signaalin tuottamiseen, kun tietty alakaistan lähde osallistuu. Jokaisen näytekeyhyksen kohdalla päivitetään myös kunkin alakaistan lähteen a priori – todennäköisyys.

Lisäksi joka kierroksella lasketaan jokaisen sekä ala-, että yläkaistan lähteen sekvenssien autokorrelaatiot. Sovellusvaiheessa irrallisina yhtä todennäköisistä laajakaistavektoreista voidaan autokorrelaation avulla valita kokonaisuuden kannalta todennäköisin laajakaistainen vektori. Lopuksi kullekin kapeakaista – laajakaista vektoriparille päivitetään ala- ja yläkaistojen energioiden suhde, jonka avulla toteutusvaiheessa määritellään yläkaistan lähteiden tuottama kokonaisenergia.

Käytyään koko opetusaineiston läpi, harjoitusalgoritmin tuottama korjausfunktio voidaan tallettaa koodikirjaksi, jolla tilastollinen taajuuskaistan laajennusmenetelmä toteutetaan.

4.2.2 LP-analyysiin ja koodikirjan käyttöön perustuva menetelmä

Vuonna 1994 Yuki Yoshida ja Masanobu Abe julkaisivat artikkelin [Yoshida, Abe 1994] koodikirjan ja lineaarisen ennustuksen avulla toteutettavasta taajuuskaistan laajennusmenetelmästä. Menetelmän toimintaperiaate on hyvin samanlainen kuin Hiroshi Yasukawan lineaariseen ennustukseen perustuvassa ehdotelmassa [Yasukawa 1996b]. Yoshidan ja Aben menetelmä valitsee LP-synteesikertoimet todellisen puhemateriaalin avulla luodun koodikirjan avulla, kun Yasukawan menetelmä laskee mekaanisesti alakaistan LP-kertoimista välittämättä yläkaistan todellisesta luonteesta.

Koodikirjaa käyttämällä saadaan tuotettua mekaanista menetelmää todenmukaisempi spektri yläkaistalle, koska opetusaineiston avulla voidaan etsiä ala- ja yläkaistan spektrien välinen todellinen korrelaatio.

Yoshidan ja Aben menetelmä tuottaa yläkaistan, eli 3,4 – 7,3 kHz:n lisäksi matalan alakaistan 50 - 300 Hz:n alueelle.

Menetelmän toteutus

Kuvassa 25 on esitetty kaavio koodikirjan avulla toteutetusta LP-kertoimiin perustuvasta menetelmästä. Kuvien 23 ja 25 menetelmät muistuttavat hyvin paljon toisiaan, sillä molemmat perustuvat lineaarisen ennustuksen suodatuskerrointen laajentamiseen kapeakaistaisen signaalin LP-kerrointen perusteella.

Normaali puhelinkaista 0,3 – 3,4 kHz siirretään sellaisenaan, vain sen näytteenottotaajuus nostetaan 16 kHz:iin.

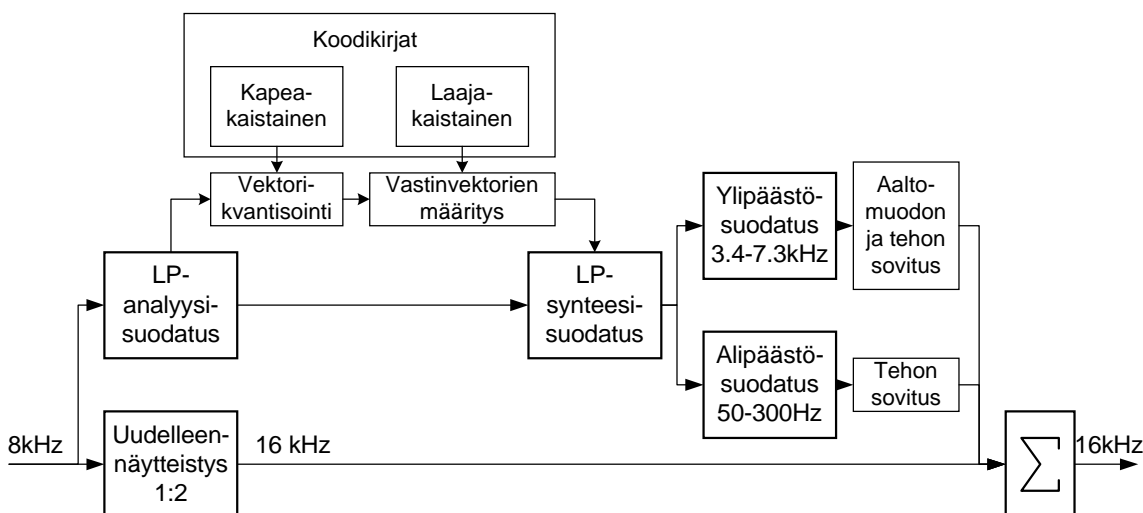
Ylä- ja alakaistojen, 3,4 – 7,3 kHz ja 50 – 300 Hz, tuottamiseksi suoritetaan aluksi

LP-analyysi. LP-analyysikertoimet vektorikvantisoidaan kapeakaistaisen signaalin koodikirjan mukaisesti. Laajakaistaisen signaalin koodikirjasta etsitään kapeakaistaisen kerroinvektorin vastine, josta saatavia kertoimia käytetään LP-synteesisuotimessa laajakaistaisen signaalin tuottamiseksi.

Alakaista 50 – 300 Hz suodatetaan erilleen LP-syntetisoidusta signaalista. Matalat taajuudet täytyy vahvistaa sovituskertoimella, jotta alakaistan teho olisi riittävä laajakaistaisen äänen muodostamiseksi.

Yläkaista 3.4 – 7.3 kHz suodatetaan samoin erilleen LP-syntetisoidusta signaalista. Suodatettu signaali kerrotaan cosini-funktiolla LP-synteesin aiheuttamien pulssien vaimentamiseksi. Signaalin teho normalisoidaan vielä ennen kuin taajuuskaistat voidaan yhdistää laajakaistaisen signaalin aikaansaamiseksi.

Yläkaistan tuottamiseksi Yoshida ja Abe kokeilivat myös toista, hieman monimutkaisempaa menetelmää kuin tässä esitetty. Kuuntelukokeiden perusteella he totesivat menetelmät kuitenkin yhtä hyväksi äänen laadun parannuskyvyn suhteen.



Kuva 25. Kaavio lineaarista ennustusta ja koodikirjaa käyttävästä menetelmästä.

Koodikirjan luominen

Yoshidan ja Aben taajuuskaistan laajennusmenetelmä käyttää kahta koodikirjaa. Koodikirjoihin talletetaan tiedot kapeakaistaisten ja laajakaistaisten näytteiden spektrien ominaisuuksista. Koodikirjat luodaan yhteistä harjoitusmateriaalia käyttäen, jolloin spektrimuotojen jakaumat vastaavat suoraan toisiaan.

Kapeakaistainen harjoitusmateriaali luodaan suodattamalla puhelinkaistainen signaali laajakaistaisesta äänimateriaalista. Koodikirjojen luomiseksi analysoidaan sekä kapea-, että laajakaistaisten näytteiden spektrien verhoikäyrät.

Laajakaistaisten spektrien verhoikäyrät talletetaan koodikirjaan LBG-algoritmilla [Linde 1980] koodattuina.

Kapeakaistaisen puheen spektrien verhokäyrät luokitellaan klustereiksi laajakaistaisen koodikirjan vektorikvantisoitujen spektrien ja näytteiden alkuperäisen ajallisen sijainnin avulla. Klustereihin kasautuneiden kapeakaistaisten spektrien verhokäyrät keskiarvoistetaan ja talletetaan koodivektoreina kapeakaistaisen signaalin koodikirjaan.

4.2.3 CELP-koodatun äänen laadunparannus monikaistaisen herätemallin avulla

Cheung-Fat Chan ja Wai-Kwong Hui ovat kehittäneet Hong Kongin yliopistossa monikaistaisen herätemallin (MBE, MultiBand Excitation) ja siihen perustuvan taajuuskaistan laajennusmenetelmän [Chan, Hui 1996]. Menetelmä keskittyy erityisesti CELP-koodatun äänen restaurointiin. Taajuuskaistan laajennuksen lisäksi menetelmä parantaa äänen laatua vähentämällä siitä häiriöitä ja korjaamalla signaalin jaksollisuutta.

Kaistalaajennus toteutetaan kahden koodikirjan avulla, joihin opetusvaiheessa talletetaan laajakaistaisen signaalin MBE-mallin parametrit. MBE-analyysin avulla puheesta saatuja parametreja verrataan koodikirjojen parametreihin, joista valitaan sopivimmat laajakaistaista MBE-synteesiä varten. Äänen jaksollisuus palautetaan syntetisoimalla puheen jaksolliset osat harmonisella MBE- syntetisaattorilla, jolloin niihin liittynyt häiriö poistuu.

MBE-malli

MBE-mallissa puheen spektri jaetaan kapeisiin kaistoihin, jotka keskittyvät äänen formanttien ja niiden harmonisten taajuuksien ympärille. Jokainen kaista määritellään joko soinnilliseksi tai soinnittomaksi.

MBE-malli mahdollistaa soinnillisuuden ja soinnittomuuden yhdistämisen koottaessa yläkaistan spektrikomponentteja. MBE-parametrit sisältävät soinnillisuus-soinnittomuus-jakauman lisäksi tiedot kaistojen voimakkuuksista ja vaiheista, sekä äänenkorkeuden.

Äänen vaiheominaisuuksien takia MBE-synteesi luo soinnilliset kaistat aika-alueessa, kun soinnittomat voidaan tuottaa taajuusalueessa.

Menetelmän toteutus

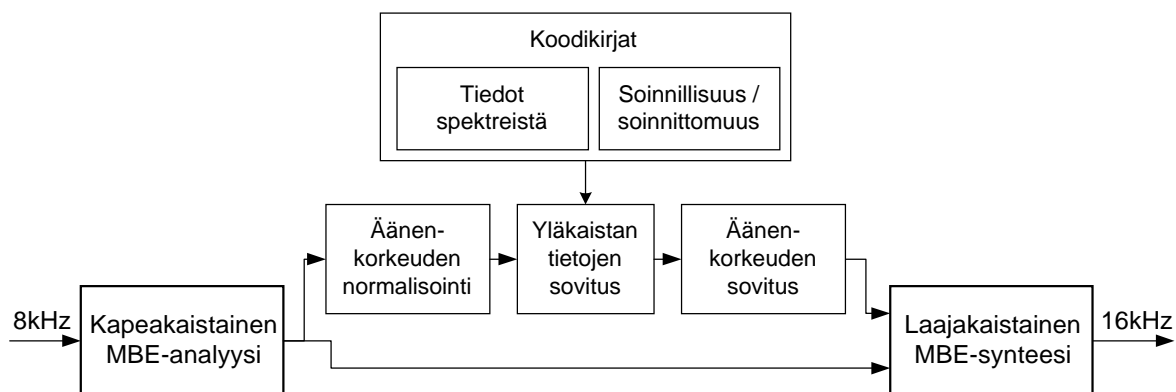
Kuvassa 26 on esitetty kaavio MBE-malliin perustuvasta taajuuskaistan laajennusmenetelmästä.

Kapeakaistaisen signaalin ominaisuudet tutkitaan MBE-analysaattorissa ja spektrikaistojen voimakkuus ja soinnillisuus/soinnittomuus-informaatio sekä puhujan äänenkorkeus siirretään MBE-syntetisaattoriin. Äänen jaksollisuuden parantaminen toteutetaan MBE-syntetisaattorissa analyysissä havaittujen tietojen perusteella. Soinnillisiksi todetut äänet syntetisoidaan puhtaalla signaalilla, joten häiriökohina

saadaan näiltä osin poistetuksi. Soinnittomiksi todetut alakaistan äänet toistetaan sellaisinaan. Tästä johtuen virheellisesti soinnittomaksi todetun äänen soinnillisuus ei katoa, ainoastaan menetelmän häiriönpoistokyky kärsii.

Ylempää taajuuskaistaa luotaessa normalisoidaan aluksi kapeakaistaisen MBE-analysoidun signaalin äänenkorkeus, jotta sen spektri olisi vertailukelpoinen koodikirjoihin talletettujen spektrikuvioiden kanssa. Tämän jälkeen yläkaistan MBE-parametrit etsitään spektri- ja soinnillisuus/soinnittomuus-koodikirjoista alakaistan vastaavien tietojen perusteella. Ennen kuin spektriparametrejä toimitetaan MBE-syntetisaattorille, painotetaan ne vastaamaan puhujan äänenkorkeutta.

MBE-syntetisaattori yhdistää analysaattorilta saamansa alakaistan tiedot ja puhujan äänenkorkeudelle siirretyn koodikirjoista etsityn yläkaistan tiedot ja tuottaa niiden perusteella laajakaistaista puhetta 0 – 8 kHz:n alueella 16 kHz:n näytetaajuudella. Alkuperäinen signaali on jäljellä enää vain alakaistan soinnittomissa äänissä.



Kuva 26. Kaavio MBE-tekniikkaan perustuvasta taajuuskaistan laajennusmenetelmästä.

Koodikirjat

Koodikirjat luodaan laajakaistaista puhemateriaalia analysoimalla. Spektrit jaetaan MBE-mallin mukaisesti taajuuskaistoihin, joiden ominaisuudet tallennetaan koodikirjaan viivaspektripareja (LSP, line spectral pair) sisältävinä koodivektoreina. Chan ja Hui tallettivat sovelluksessaan 1024 vektoria, joista kukin sisälsi 18 viivaspektriparia.

Toiseen koodikirjaan talletetaan tiedot äänen spektrikaistojen soinnillisuudesta tai soinnittomuudesta. Soinnillisuutta arvioidaan laajakaistaisen spektrin energian jakaumasta. Oletetaan, että spektrin verhokäyrä ja soinnillisuusenergia korreloivat voimakkaasti – soinnillisten äänteiden formanteilla on suuri energia, kun taas soinnittomien äänten energia on pieni ja niiden spektri painottuu korkeille taajuuksille.

4.2.4 Pohdintaa

Tilastollisin menetelmin tuotettujen koodikirjojen avulla on mahdollista tuottaa hyvinkin tarkasti alkuperäisen kaltaista puhetta häiriöttömissä olosuhteissa. Rajoittavana tekijänä koodikirjojen koon ja tuottamisen suhteen ovat ihmisten erilaiset ääntämistavat, joita on lukematon määrä ja joiden täydellinen arkistointi koodikirjoihin on lähes mahdoton tehtävä. Lisäksi eri kielissä ilmenevät sadat äänneet ja äänneaset aiheuttavat luonnollisesti ongelman koodikirjan koon ja hakunopeuden suhteen.

Jos ajatellaan puhelintoteutusta, kuten menetelmien suunnittelijat ovat selvästi tehneet, ei äänneiden määrä rajoitu vain noihin muutamiin satoihin äänneasetiin, vaan puhelimen käyttöympäristöstä aiheutuvat häiriöäännet voivat olla millaisia tahansa. Jos puhelinyhteyden äännet tulkitaan koodikirjan perusteella häiriöiden takia joksikin toiseksi kuin mitä puhuja on sanonut, voi puheen laatu huonontua dramaattisesti taajuuskaistan laajennusalgoritmin takia.

5 TOTEUTETTU TAAJUUSKAISTAN LAAJENNUSMENETELMÄ

Tavoitteenamme taajuuskaistan laajennusalgoritmin kehitystyössä on ollut sovellettavuus käytännön puhelinjärjestelmään. Käytännön tilanteissa puhe on häiriöistä, jolloin on varmistettava, että taajuuskaistan laajennus ei missään tilanteessa pilaa eikä huononna puheen laatua. Tavoitteena ei ole ollut täydellisen alkuperäisen puheen rekonstruointi, vaan puheen laadun parantaminen.

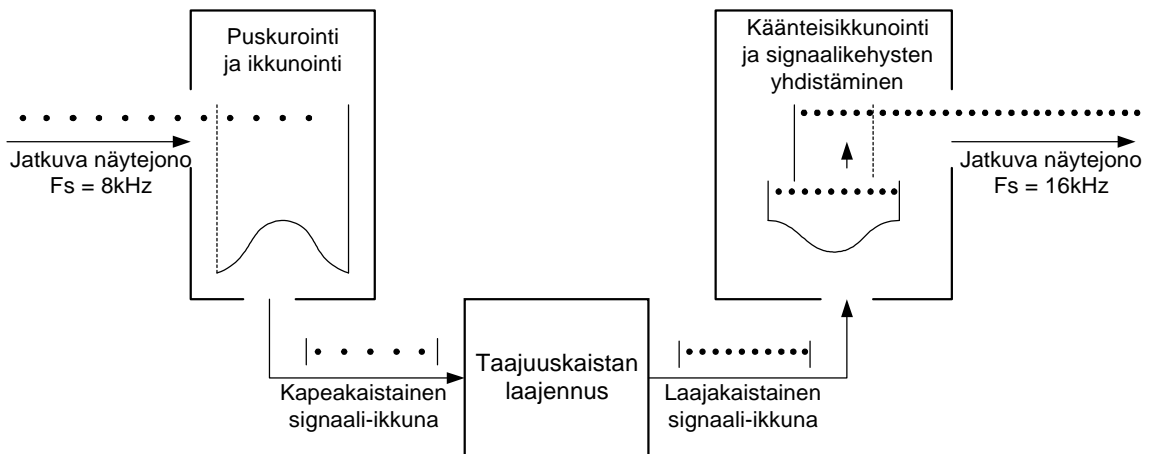
Olemme toteuttaneet menetelmän PCM-koodatun puheen taajuuskaistan laajentamista varten siten, että jatkuva-aikaisen signaalin käsittely on mahdollista. Ensimmäinen vaatimus jatkuva-aikaisuudelle on signaalin ikkunointi. Toinen edellytys toimivalle jatkuva-aikaiselle toteutukselle on se, että algoritmin vaatima laskenta on mahdollista suorittaa DSP-prosessorilla signaali-ikkunan puitteissa. Näiden vaatimusten mukaisesti algoritmi on suunniteltu ja toteutettu Matlab-ohjelmistoa hyväksi käyttäen. Konekielinen DSP-toteutus on jätetty jatkotutkimusten aikana tehtäväksi.

Toteuttamamme taajuuskaistan laajennusmenetelmä ryhmittyy toisessa luvussa esiteltujen menetelmien joukossa perinteisen signaalinkäsittelyn keinoin toteutettujen menetelmien ryhmään. Perinteisen signaalinkäsittelyn keinoin on tilastollisia menetelmiä helpompi saavuttaa algoritmin toimintavarmuus kaikissa tilanteissa. Menetelmämme käyttää hyväkseen hallittua laskostamista kuten osa neljännen luvun menetelmistäkin. Yläkaistan vaimennus on toteutettu adaptiivisesti siten, että kullekin äänelle sovitetaan yksilöllinen vaimennus.

Olemme testanneet menetelmän toimivuutta subjektiivisin kuuntelukokein, joiden perusteella voidaan sanoa puheen laadun paranevan tämän taajuuskaistan laajennusmenetelmän ansiosta.

5.1 Algoritmin kuvaus

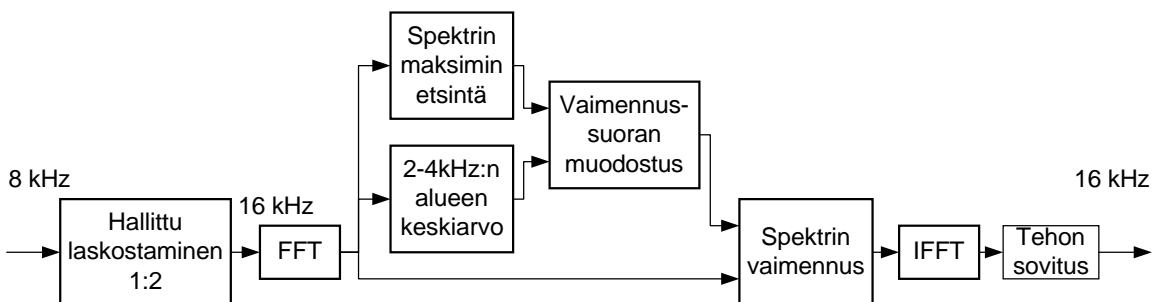
Jatkuva-aikaisessa puhelintoteutuksessa signaalinäytteitä joudutaan odottamaan ja puskuroimaan aina yhden ikkunanpituuden verran. Kun riittävä määrä näytteitä on saatu kokoon, voidaan puskuriin kertyneistä näytteistä ottaa ikkunan mittainen pala käsiteltäväksi. Ikkunat leikataan jatkuvasta signaalivuosta osittain päällekkäin, sillä käyttämällä vain ikkunan keskialueille osuvia signaalinäytteitä pysyy käsitellyn signaalin laatu parempana. Tässä menetelmässä käytämme 30 ms:n mittaista Hamming-ikkunaa signaalivuon leikkaamiseen. Kuvassa 27 on esitetty puskuroinnin ja ikkunoinnin periaate.



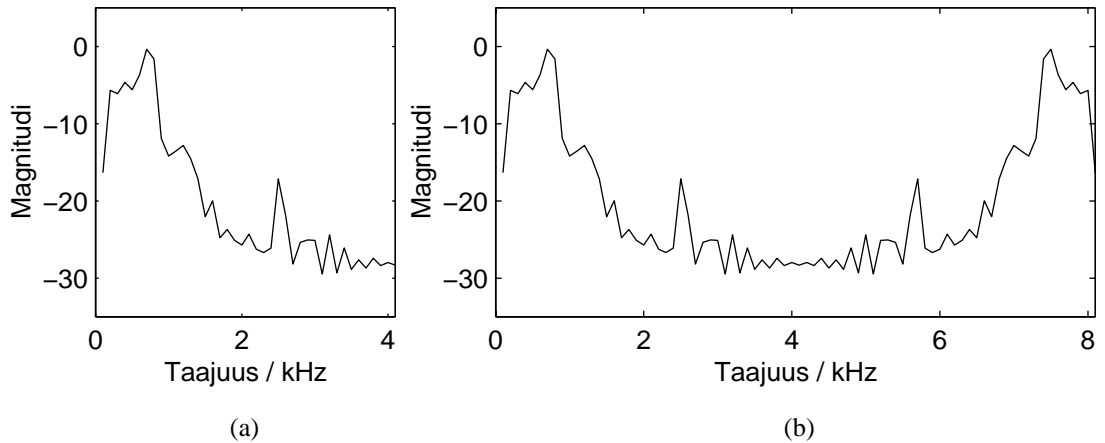
Kuva 27. Puskurointi ja ikkunointi ennen taajuuskaistan laajennusta ja käänteisikkunointi ja kehysten irrotus ja yhdistäminen jatkuvaksi näytejonoksi taajuuskaistan laajennusoperaation jälkeen.

Samoin kuvassa 27 näkyvät operaatiot, jotka signaali-ikkunalle on suoritettava taajuuskaistan laajennuksen jälkeen, jotta saadaan jälleen aikaan jatkuva signaali. Ikkunoinnin käänteisoperaationa suoritetaan käänteisikkunointi, joka poistaa ikkunoinnin aiheuttaman amplitudin vaimenemisen ikkunan reuna-alueilta. Käänteisikkunoidusta signaalinpätkästä irrotetaan hyväksi käytettävä osuus ja liitetään se laajakaistaiseen jatkuvaan näytejonoon. Signaali-kehäksiä yhdistettäessä keskiarvoistetaan näytteitä jonkin matkaa perättäisten kehysten raja-alueen molemmilta puolilta, jotta välttyttäisiin signaali-kehysten raja-alueiden häiriöiltä.

Kuvassa 28 näkyy kaavio taajuuskaistan laajennusalgoritmista, jolla signaali-ikkunoiden spektrit laajennetaan. Menetelmässä käytetään hyväksi hallittua laskostamista, mikä toteutuu nolla-arvoisen näytteen lisäyksellä jokaiseen näyteväliin. Hallitun laskostamisen avulla nostetaan signaalin näytetaajuus 8 kHz:stä 16 kHz:iin ja samalla tuotetaan 4 – 8 kHz:n taajuusalueelle kapeakaistaisen spektrin peilikuvan mukaiset taajuuskomponentit, kuten kuvassa 29 näkyy.



Kuva 28. Ikkunoittain suoritettava taajuuskaistan laajennus kehittämässämme menetelmässä.



Kuva 29. (a) Alkuperäisen 8000 Hz:n taajuudella näytteistetyin signaali-ikkunan spektri. (b) Spektri signaalista, jonka näytetaajuus on nostettu 16000 Hz ja laskostuneet ylätaajuudet on jätetty suodattamatta.

Kun näytteenottotaajuus on hallitulla laskostamisella nostettu 16000 Hz:iin, siirretään laskenta FFT:n avulla taajuusalueeseen. Tässä vaiheessa signaali on jo periaatteessa laajakaistainen, mutta hallitun laskostamisen avulla kopioidut spektrikomponentit ovat aivan liian voimakkaista kuulostaakseen millään tavalla luonnollisilta, joten näitä yläkaistan taajuuksia on vaimennettava voimakkaasti.

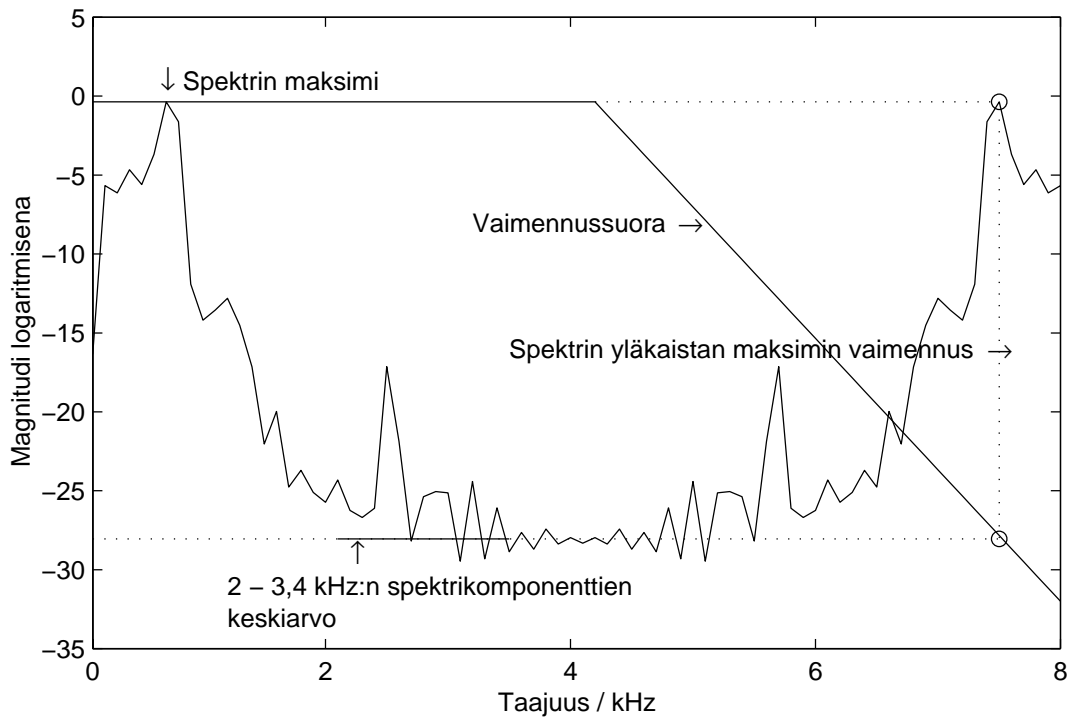
Vaimennus toteutetaan taajuusalueessa logaritmisella asteikolla määriteltävän vaimennussuoran avulla. Kuten kuvan 28 kaaviossa näkyy, etsitään FFT:llä lasketusta spektristä magnitudin maksimikohta ja lasketaan 2 – 4 kHz:n taajuusalueelta magnitudin keskiarvo. Suuret lasketaan lineaarisen asteikon mukaisesta spektrin magnitudista ja ne siirretään logaritmiselle asteikolle vaimennussuoraa määrättäessä.

Näiden tietojen perusteella määrätään kullekin signaali-ikkunalle yksilöllinen vaimennussuora kuvan 30 mukaisesti siten, että 4 – 8 kHz:n ylätaajuuksien spektrin maksimikohta vaimenee 2 – 4 kHz:n spektrin magnitudin keskiarvon tasalle.

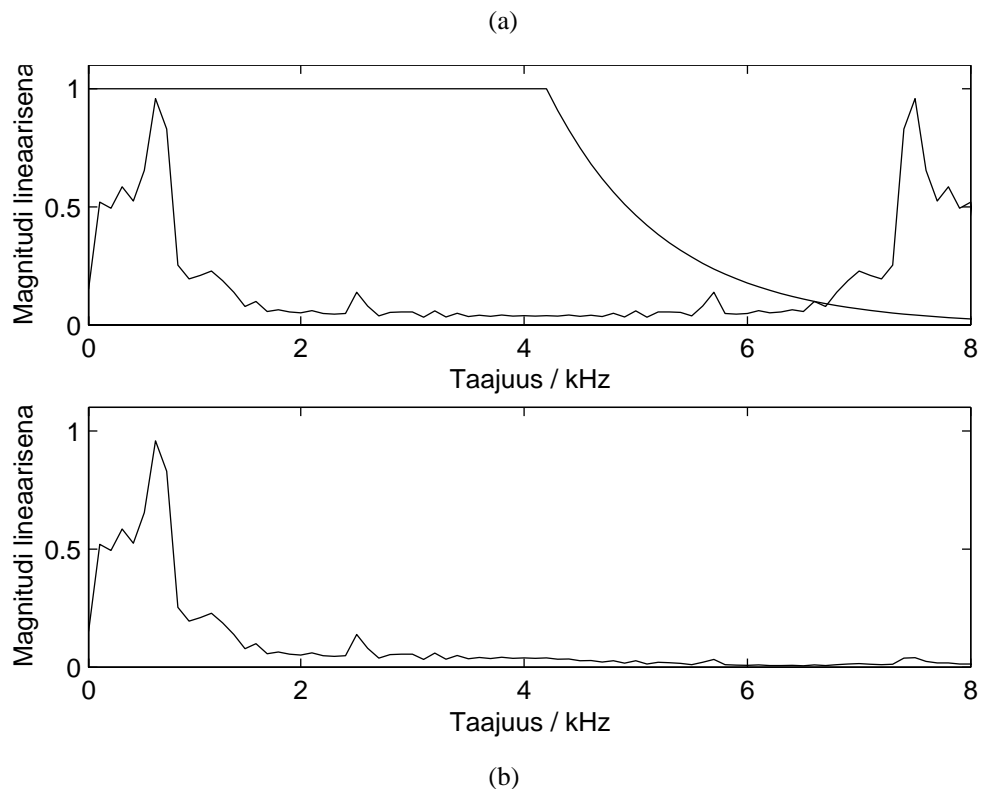
Vaimennussuoran kulmaa voidaan vielä säätää algoritmin toimintaa määrävän vaimennuskertoimen avulla. Kunkin taajuuden desibeleinä ilmaistua vaimennusta kerrotaan vaimennuskertoimella, joten yhtä suuremmilla vaimennuskertoimen arvoilla vaimennussuoran kulma jyrkkenee ja vaimennus on voimakkaampi.

Koska FFT:llä laskettua spektriä ei ole syytä siirtää logaritmiselle asteikolle, laskettu vaimennussuora siirretään lineaariselle asteikolle, jolloin suora taipuu kuvan 31 (a) mukaisesti.

Näin lasketulla vaimennussuoralla kerrotaan laskostettu spektri vaimennuksen toteuttamiseksi, kuten kuvan 28 kaaviossa on esitetty. Kuvassa 31 (b) näkyy vaimennettu laajakaistainen spektri.



Kuva 30. Vaimennussuoran määrittäminen logaritmisella asteikolla esitetyllä signaali-ikkunan spektrille. Vaimennussuora mitoitetaan vaimentamaan yläkaistan voimakkain spektrikomponentti 2 - 4 kHz:n taajuusalueen keskimagnitudin tasolle.



Kuva 31. (a) Vaimennussuoraa käytetään lineaarisen magnitudiasteikon mukaisessa muodossa, jolloin sen vaimennus kasvaa eksponentiaalisesti kohti 8 kHz:n Nyquistin taajuutta. (b) Vaimennussuoran avulla vaimennettu laajakaistaisen signaali-ikkunan spektri.

Kun yläkaista on vaimennettu, siirrytään takaisin aika-alueeseen IFFT:n avulla. Ennen signaali-ikkunoiden käänteisikkunointia ja kehysten yhdistämistä laajakaistaisten signaali-ikkunoiden äänenvoimakkuutta on vielä vaimennettava. Kun spektri laskostetaan hallitusti, syntyy kuultavaan ääneen uusia taajuuksia, jotka lisäävät signaalin tehoa. Signaalin tehon lisäys on kompensoitava, jotta äänenvoimakkuus olisi taajuuskaistan laajennuksen jälkeen sama kuin alkuperäisellä kapeakaistaisella äänellä.

5.2 Menetelmän toteutus Matlab-ohjelmistolla

Menetelmän kehitystyö on tehty PC-tietokoneessa toimivan Matlab-ohjelmiston avulla. Matlab on matemaattiseen työskentelyyn tarkoitettu laskentaohjelma, jolle on mahdollista ohjelmoida uusia funktioita tavallisen ohjelmointikielen tapaan.

Matlabin funktioiden muodossa käskyjä suorittavasta luonteesta johtuen ei ole ollut mahdollista luoda todellisuudessa reaaliaikaisesti toimivaa ohjelmaa. Matlab:lla toteutettu ohjelma toimii siten, että se ottaa vastaan valmiin mielivaltaisen mittaisen kapeakaistaisen signaalin ja tuottaa siitä laajakaistaisen version. Toiminnallisesti Matlab-ohjelma kuitenkin käsittelee signaalia aivan kuten DSP:llä toteutettu jatkuva-aikaisesti toimiva ohjelmakin käsittelee.

5.2.1 Funktiot

Taajuuskaistan laajennuksen toteuttava Matlab-ohjelma käyttää tässä kappaleessa esiteltyjä funktioita tehtävänsä suorittamiseen. Pääohjelma on nimeltään `fbe`, joka muodostuu sanoista 'frequency band expand'. Muut funktiot ovat sen avuksi kirjoitettuja rutiineja, joiden tehtävänä on jakaa taajuuskaistan laajennusalgoritmin työvaiheet tarkoituksen mukaisiin osiin ja selkeyttää ohjelman hahmottamista.

Tässä aliluvussa esiteltyjen Matlab-funktioiden ohjelmakoodit löytyvät liitteestä 1.

`wavfbe`

`wavfbe` on pieni kehysfunktio, joka on kirjoitettu helpottamaan `wav`-tyyppisten äänitiedostojen käsittelyä. `wavfbe` käyttäytyy täysin `fbe`:n tavoin lukuunottamatta sitä, että se käsittelee `.mat`-tyyppisen signaalimatriisin sijasta `.wav`-tyyppistä ääntä.

Funktio tarvitsee argumentteinaan `wav`-tyyppisen lukutiedoston ja kirjoitustiedoston nimet. Näiden lisäksi sille on mahdollista antaa argumentteina kaikki algoritmia koskevat parametrit, jotka `fbe`:kin voi ottaa vastaan. `wavfbe` ainoastaan lukee `wav`-tyyppisen lukutiedoston näytteet vektoriin, jonka se välittää `fbe`:lle. Saatuaan laajakaistaisen signaalin vektorimuodossa `fbe`:ltä `wavfbe` kirjoittaa sen `wav`-muodossa kirjoitustiedostoon.

fbe

Argumenttina `fbe` tarvitsee vaakavektorimuodossa signaalin, jonka taajuuskaista on tarkoitus laajentaa. Funktio olettaa näytteenottotaajuuden olevan 8000 Hz, koska puhelimen välittämä ääni, jota varten taajuuskaistan laajennusmenetelmä on suunniteltu, näytteistetään nykyään juuri tuolla taajuudella. Lisäksi `fbe` tarvitsee argumenttina vaimennuskertoimen, jonka mukaisesti ylätaajuuksien vaimennus sovitetaan. Mahdollisia lisäargumentteja `fbe` -funktiolle ovat `fft:n` koko, ikkunan pituus, kustakin ikkunasta tehollisesti käytettävän näytejonon pituus, ikkunoiden yhdistämisessä käytettävän keskiarvoistusliuskan pituus ja spektrin huippukohta etsittäessä keskiarvoistettavien spektrikomponenttien lukumäärä.

`fbe` luo argumenttina saamansa pituuden tai oletusarvon mittaisen Hamming-ikkunan signaalin pilkkomista varten. Se pilkkoo kapeakaistaisen signaalin sopivanmittaisiin osiin `cutframes`-funktion avulla. `expand`-funktion avulla `fbe` laajentaa `cutframes`-funktiolta saamiensa kehysten taajuuskaistat, jolloin laajakaistaisen signaalikehysten näytetaajuus nousee 16000Hz:iin. Lopuksi `fbe` poistaa signaalikehyksistä ikkunoinnin käänteisellä Hamming-ikkumalla ja yhdistää kehukset yhdeksi signaalivektoriksi `cascade`-funktion avulla.

`fbe` palauttaa vaakavektorina signaalin, jonka taajuuskaista on laajennettu ja näytetaajuus noussut 16000Hz:iin.

cutframes

Argumentteina `cutframes` tarvitsee pilkottavan signaalin vaakavektorimuodossa, samoin vaakavektorina ikkunafunktion, jolla signaali tulee paloitella ja ikkunasta hyväksi käytettävän kehyksen pituuden.

`cutframes` leikkaa ikkunavektorin avulla signaalivektorista paloja. Palat leikataan signaalista siten, että hyväksi käytettävä kehys osuu ikkunan keskelle ja ikkunat jäävät osittain päällekkäin siten, että hyväksi käytettävät kehukset seuraavat toinen toistaan saumattomasti.

`cutframes` palauttaa matriisiin, jossa leikatut ikkunat ovat vaakariveillä. Matriisiin vaakarivien määrä riippuu kehyksen ja signaalin pituuksien suhteesta.

expand

`expand`-funktio laajentaa argumenttina vastaanottamansa signaali-ikkunan spektrin laajakaistaiseksi. Lisäksi `expand` tarvitsee argumenttikseen vaimennuskertoimen. Haluttaessa voidaan vielä antaa argumenteiksi FFT:n koko ja spektrin maksimikohta etsittäessä keskiarvoistettavien näytteiden lukumäärä. Muutoin näille parametreille käytetään oletusarvoja.

Aluksi `expand` laskostaa hallitusti signaali-ikkunan spektrin lisäämällä nolla-arvoisen näytteen kuhunkin näyteväliin ja tämän jälkeen siirtää laskennan taajuusalueeseen FFT:n avulla. Hallitun laskostamisen avulla tuotetun yläkaistan spektri

vaimennetaan attenuate-funktiolla, jonka jälkeen expand palauttaa signaalin aika-alueeseen IFFT:llä. Ennen laajakaistaisen signaalin palauttamista expand:ia kutsuneelle fbe-funktiolle expand vielä normalisoi laajakaistaisen signaalin tehon vastaamaan kapeakaistaisen signaalin tehoa.

attenuate

Kuvassa 12 näkyvät attenuate-funktion työvaiheet. attenuate vastaanottaa sitä kutsuvalta expand-funktiolta signaali-ikkunan spektrin, joka on laajennettu ulottumaan 8000 Hz:iin asti hallitun laskostamisen avulla. Lisäksi attenuate tarvitsee argumentteina vaimennuskertoimen ja spektrin maksimikohtaa etsittäessä keskiarvoistettavien näytteiden lukumäärän. Näiden tietojen perusteella attenuate etsii spektrin maksimikohdan ja laskee spektrin keskiarvon 2 – 4 kHz:n alueelta.

Logaritmisen vaimennussuoran attenuate määrittää logaritmisten spektrin maksimikohtatiedon ja keskitaajuuksien keskiarvon avulla. Vaimennussuoran kallistus sovitetaan vaimennuskertoimen avulla halutuksi ja lopuksi laajakaistaisen spektrin ylätaajuudet vaimennetaan lineaariselle asteikolle siirretyn vaimennusvektorin avulla.

cascade

cascade-funktion tarkoitus on yhdistää taajuuskaistan laajennusoperaatioiden jälkeen signaalikehykset takaisin jatkuvaksi näytejonoksi. Tätä varten cascade tarvitsee argumenteikseen laajakaistaiset signaali-ikkunat sisältävän matriisin, ikkunasta hyödyksi käytettävän näytejonon pituuden ja kehysten reuna-alueilla keskiarvoistettavien näytteiden määrän.

cascade selvittää mitkä näytteet signaali-ikkunoista jäävät täysin käyttämättä, mitä käytetään keskiarvoistuksessa ja mitkä liitetään lopulliseen signaaliin sellaisenaan. Hyväksi käytettävien kehysten reuna-alueet sovitetaan toisiinsa keskiarvoistusliuskosten avulla ennen kuin kehykset irrotetaan signaali-ikkunoista ja asetetaan peräkkäin.

cascade palauttaa sitä kutsuneelle fbe-funktiolle valmiin laajakaistaisen signaalin.

5.2.2 Muuttujien oletusarvot

Argumenteille on määrätty oletusarvot, joita taajuuskaistan laajennusmenetelmän toimivuutta arvioivien kuuntelukokeidenkin äänten tuottamisessa on käytetty.

Ikkunanpituutena on käytetty 30 ms:ia, mikä tarkoittaa 8000 Hz:n näytteenottotaajuudella 240 näytteen mittaista ikkunaa. Tämä ikkunanpituus ei vielä aiheuta liian suurta viivettä reaaliaikaisessa toteutuksessa käytettäessä ja lyhyempi ikkuna olisi toisaalta hämärtänyt taajuusalueessa suoritettavaa laskentaa tarpeettomasti. Ikkunafunktiona on aina käytetty Hamming-ikkunaa. Ikkunavalintaa ei ole edes sisällytetty fbe-funktion argumenttien luetteloon, koska Hamming-ikkuna todettiin

sopivaksi tähän sovellukseen. Ikkunasta hyväksi käytettävän osuuden, eli kehyksen oletuspituudeksi on asetettu 20 ms eli 160 näytettä.

FFT:n pituutena on käytetty 512:ta taajuuskomponenttia, mikä tarkoittaa 240 näytteen mittaisella signaali-ikkunalla sitä, että signaali on täydennettävä 512 näytteen mittaiseksi nolla-arvoisilla näytteillä ennen FFT:n laskemista. IFFT:n yhteydessä nämä nollanäytteet poistetaan signaalin lopusta.

Spektrin maksimikohtaa etsittäessä keskiarvoistettavien spektrikomponenttien lukumäärän oletusarvoksi on asetettu 5. Tämä arvo estää yhden komponentin satunnaisten piikkien dominoinnin ja siksi etsii varmemmin suuritehoista taajuusaluetta. Tätä suurempaa lukua ei toisaalta kannata valita, sillä suuritehoiset spektrialueet eivät välttämättä ole kovin laajoja.

Signaalikehysten yhdistämisessä käytettävän keskiarvoistusliuskan pituutena on käytetty 100:aa näytettä. Tämä siis tarkoittaa sitä, että kahta peräkkäin olevaa kehystä yhdistettäessä etummaisesta kehystä viimeiset 50 näytettä ja jälkimmäisen kehystä 50 ensimmäistä näytettä kuuluvat keskiarvoistusalueelle. Tämä arvo todettiin riittäväksi estämään kehysten alku- ja loppupäissä syntyvien transienttien aiheuttamien häiriöiden syntyä.

6 KEHITETYN MENETELMÄN ÄÄNEN LAADUN PARANNUSKYVYN MÄÄRITYS KUUNTELUKOKEIN

Taajuuskaistan laajennusmenetelmän toiminnan arvioimiseksi järjestetään kaksi kuuntelukoea, joista ensimmäisen on tarkoitus kertoa suuntaa antavasti siitä, miten kuulijat kokevat lisätyt ylätaajuudet. Toisen kokeen tarkoituksena oli tarkentaa ensimmäisen kokeen tuloksia, jotta saataisiin selville, millaisilla vaimennuskertoimen arvoilla äänen laatu paranee.

Kuuntelukokeet toteutettiin toimistohuoneessa PC-tietokoneella. Äänet kuunneltiin kuulokkein (Sennheiser HD580), koska kuulokkeet vastaavat kaiuttimia paremmin puhelimen käyttötilanteen akustiikkaa. Äänet tuotettiin wav-tyyppisiksi, joita kuuntelijan on helppo kuunnella haluamassaan järjestyksessä niin monta kertaa kuin kokee tarpeelliseksi arvioidakseen äänen.

6.1 Ensimmäinen suuntaa antava kuuntelukoe

Ensimmäinen kuuntelukoe toimi pioneerikokeena toista, laajempaa testausvaihetta varten. Koe järjestettiin yksinkertaiseksi, jotta osittain kokemattomat kuuntelijat onnistuisivat arviointitehtävässään. Kuuntelijoita oli yhteensä kymmenen, neljä naista ja kuusi miestä.

Kokeessa kustakin sanasta oli tuotettu neljä erilaatuista versiota, joista kuulijan tuli valita miellyttävin. Kokeessa esitetty kysymys ennen ääninäytteiden kuuntelemista oli: ”Mikä vaihtoehto kustakin sanasta on mielestäsi selkein/luontevin/miellyttävin/paras?”

6.1.1 Kokeessa käytetyt äänet

Ensimmäisessä kuuntelukokeessa ääninäytteinä oli 16 ääntä, jotka sisälsivät lukusanoja yhdestä viiteen viiden eri henkilön lausumina. Puhujista oli naisia kaksi ja miehiä kolme.

Kustakin kuunneltavasta sanasta oli tuotettu neljä erilaatuista versiota, kuten taulukossa 6 on esitetty. Ääniversio A oli suodatettu sisältämään taajuudet noin 300 Hz:stä noin 2,8 kHz:iin. Se oli laadultaan selvästi tukkoisempi ja mumisevampi kuin versio B, joka vastasi PCM-koodattua ääntä, eli sisälsi taajuudet 0,3 – 3,4 kHz.

Ääniversion C taajuuskaista oli laajennettu kehitetyllä menetelmällä, mutta ylempiä taajuuksia oli vaimennettu verraten voimakkaasti, vaimennuskertoimella 1,8. Ääniversion D taajuuskaista oli myöskin laajennettu ja ylempien taajuuksien vaimennus oli jätetty vähäisemmäksi käyttämällä vaimennuskerrointa 1,2.

Taulukko 6. Ensimmäisessä kuuntelukokeessa käytettyjen äänten laatuversioiden taajuusominaisuudet. Vaimennuskerroin kertoo kehitetyn menetelmän mukaisesti keinotekoisesti tuotetun 4000 – 7700 Hz:n yläkaistan vaimennuksen voimakkuuden. Suurempi vaimennuskertoimen arvo vastaa voimakkaampaa vaimennusta.

| Äänen laatuversio | Taajuuskaista | Vaimennuskerroin |
|-------------------|---------------|------------------|
| A | 300-2800Hz | - |
| B | 300-3400Hz | - |
| C | 300-7700Hz | 1,8 |
| D | 300-7700Hz | 1,2 |

6.1.2 Tulosten analysointi

Koetulosten perusteella voidaan sanoa, että taajuuskaistan keinotekoinen laajennus kehitetyllä menetelmällä parantaa jonkin verran kapeakaistaisen puheäänien laatua. Naisäänten kohdalla parannus on selvempi kuin miesäänten kohdalla, koska ylätaajuudet kuulostavat luonnollisemmilta naisääniin kuin miesääniin lisätynä.

Kahden kuuntelijan kohdalla tulokset olivat ristiriitaiset muiden kuulijoiden mielipiteisiin verrattuina ja monessa kohtaa he olivat valinneet sanan laatuversioista ensimmäisen, minkä taajuuskaista oli suodatettu vielä puhelinkaistaakin kapeammaksi. Tämä osoittaa, että kuulijan keskittymiskyky on herpaantunut tai että jokin muu tekijä kuin äänen taajuuskaistan laajuus vaikuttaa kokemukseen kyseisen äänen laadusta, joten tulokset on laskettu sekä kaikkien tulosten kesken, että siten, että näiden kahden kuuntelijan tulokset on jätetty huomiotta.

Tulokset on laskettu keskiarvoina kuuntelijoiden miellyttävimmiksi arvioimien äänten mukaan siten, että ääniversio A vastaa 1:tä pistettä, B 2:ta, C 3:a ja D 4:ää pistettä. Keskiarvon käyttö tämän kaltaisen kuuntelukokeen tulosten arvioinnissa on perusteltua, koska kuuntelijoiden vastausten hajonta oli hyvin pientä eli tilannetta, jossa toinen kuuntelija olisi arvioinut ääniversion A parhaaksi ja toinen kuuntelija ääniversion D, ei analysoitavassa aineistossa ollut.

Keskiarvo 2,0 siis merkitsee sitä, että puhelinkaistainen ääni on koettu miellyttävämmäksi kuin äänet, joiden taajuuskaista on laajennettu. Keskiarvo 3,0 taas

merkitsee äänen olevan miellyttävimmillään, kun äänen taajuuskaista on laajennettu ja laajennus on vaimennettu kertoimella 1,8.

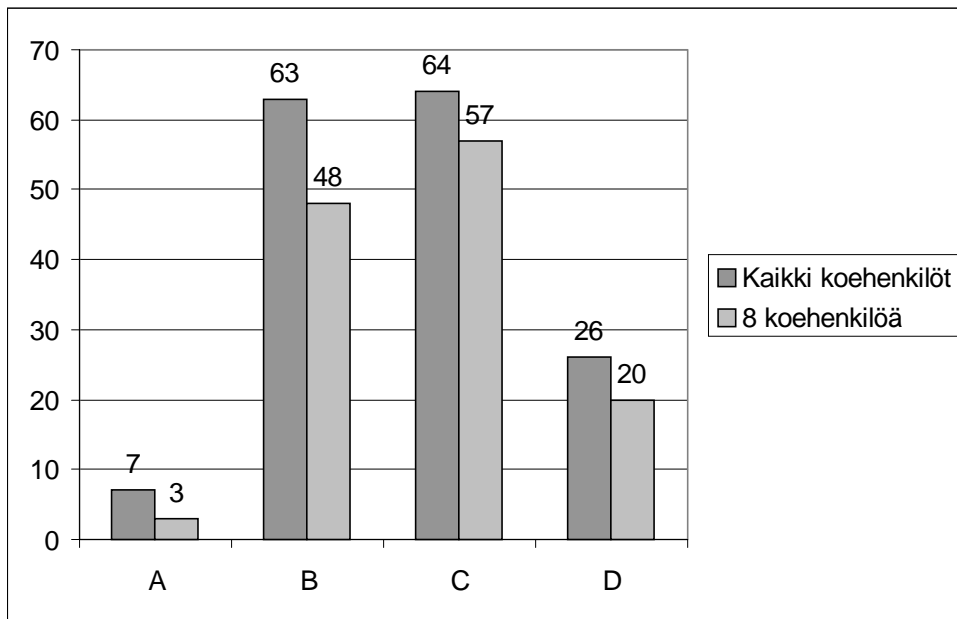
Seuraavassa taulukossa 7 näkyvät eri sanoille annetut keskiarvot sekä kaikkien kymmenen että kahdeksan onnistuneeksi katsotun kuuntelusuorituksen osalta. Kokonaiskeskiarvoon kahden kuuntelijan pudottaminen laskuista ei vaikuttanut merkittävästi, sen sijaan yksittäisten sanojen saamiin arvioihin kuuntelusuoritusten hylkääminen on paikoitellen vaikuttanut huomattavasti.

Taulukko 7. Ensimmäisen kuuntelukokeen tuloksia.

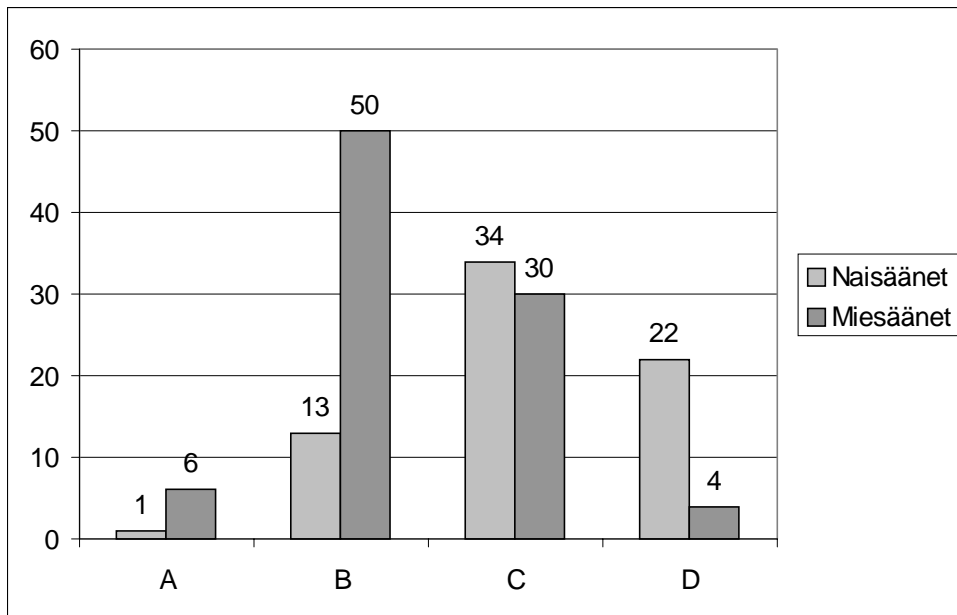
| | Nais- /Mies- ääni | Sanalle annettu keskiarvo: | Keskiarvo ilman poikkeavia tuloksia: |
|-------------------|-------------------------|-------------------------------|--|
| Sana nro 1 | N | 3.10 | 3.25 |
| Sana nro 2 | M | 2.20 | 2.25 |
| Sana nro 3 | N | 3.20 | 3.38 |
| Sana nro 4 | N | 3.10 | 3.00 |
| Sana nro 5 | M | 2.90 | 3.00 |
| Sana nro 6 | M | 2.30 | 2.38 |
| Sana nro 7 | M | 2.20 | 2.13 |
| Sana nro 8 | N | 3.10 | 3.00 |
| Sana nro 1 | M | 2.90 | 2.88 |
| Sana nro 2 | M | 2.30 | 2.38 |
| Sana nro 3 | N | 2.40 | 2.63 |
| Sana nro 4 | M | 1.90 | 2.00 |
| Sana nro 5 | N | 3.10 | 3.13 |
| Sana nro 6 | M | 2.20 | 2.38 |
| Sana nro 7 | M | 2.30 | 2.38 |
| Sana nro 8 | N | 3.70 | 3.63 |
| Keskiarvo: | | 2.68 | 2.73 |

Kuvassa 32 on esitetty, miten äänten eri laatuversioiden valinnat ovat jakautuneet sekä kaikkien koekuuntelijoiden että kahdeksan kuuntelijan kesken.

Huomattava ero syntyy laskettaessa naisäänten ja miesäänten saamiin arvioiden keskiarvoja. Naisäänten saamiin arvioiden keskiarvo nousee yli 3,1:een lasketavasta riippuen ja miesäänten arviot jäävät 2,4:n tuntumaan. Kuvasta 33 nähdään, kuinka miesäänten kohdalla valinta on hyvin usein osunut äänen laatuversioon B eli puhelimen kapeakaistaiseen ääneen. Naisäänten osalta sen sijaan useimmin on valittu ääniversio C tai D, joiden taajuuskaistat olivat laajennetut. Näiden lukujen valossa näyttää siltä, että naisääniä taajuuskaista laajennus parantaa huomattavasti, mutta miesääniä hieman epävarmemmin.



Kuva 32. Kuuntelijoiden valintojen lukumäärät koeäänten neljästä eri laatuversioista (A, B, C ja D) ensimmäisessä kuuntelukokeessa. Ensimmäinen palkki kertoo kyseisen version valintojen määrän kaikkien kuuntelijoiden osalta ja toinen palkki osoittaa valintojen määrää kahdeksan kuuntelijan osalta.



Kuva 33. Kuuntelijoiden valintojen lukumäärät ensimmäisessä kuuntelukokeessa kuunneltujen nais- ja miesääninäytteiden neljästä eri laatuversioista. Ensimmäinen palkki kertoo kyseisen version valintojen määrän naisääninäytteiden osalta ja toinen palkki miesääninäytteiden osalta, kun kaikkien koekuuntelijoiden tulokset on otettu huomioon.

6.2 Toinen, perusteellisempi kuuntelukoe

Ensimmäisen kuuntelukokeen perusteella kehitetty taajuuskaistan laajennusmenetelmä parantaa puheäänien laatua. Toisessa kuuntelukokeessa oli tarkoitus selvittää tarkemmin millaisella ylätaajuuksien vaimennuksella paras äänen laatu saavutetaan.

Toinen koe oli selvästi ensimmäistä vaativampi, arviointiperiaate oli kuitenkin sama. Ääniä oli huomattavasti enemmän ja ne ovat pidempiä kuin ensimmäisessä kokeessa, joten ne on jaettu kolmeen ryhmään, jotta kuuntelija ei väsyisi. Kuuntelija valitsi tässä, samoin kuin ensimmäisessäkin kokeessa, annetuista äänen laatuversioista miellyttävimmän. Toisen kokeen äänet arvioi 12 koehenkilöä.

6.2.1 Kokeessa käytetyt lauseet

Toisessa kuuntelukokeessa arvioitavana oli yhteensä 39 lausetta. Puhujia oli lauseilla yhteensä 10, joista kuusi miestä ja neljä naista. Ääninäytteet sisälsivät taulukossa 8 lueteltuja lauseita.

Taulukko 8. Toisessa kuuntelukokeessa käytettyjä lauseita.

| |
|---|
| Golf on hauska peli. |
| Kip Keino juoksi varpaansa verille. |
| Budjettivaje on täytetty. |
| Mooses uhkasi lyödä lihoiksi koko pässin. |
| Mäyräkoiramme söi tabaskoa. |
| Patikointi vuoristossa virkistää väsyneen mielen. |
| Perheloma autoillen on yhtä tuskaa. |
| Talvi saapuu vaihtelevin tavoin. |

Kustakin lauseesta oli tuotettu viisi eri versiota taulukon 9 mukaisesti. Kunkin kuunneltavan lauseen laatuversio A sisälsi lauseen suodatettuna siten, että se muodostui taajuuksista 300 – 3200 Hz. Laatuversio B sisälsi lauseen puhelinkaistaisena. Laatuversion C taajuuskaista oli laajennettu ja laajennusta on vaimennettu kertoimella 2,0. Jokaisen kuunneltavan lauseen laatuversion D taajuuskaistan laajennusta oli vaimennettu kertoimella 1,5, ja laatuversion E taajuuskaistan laajennusta kertoimella 1,2.

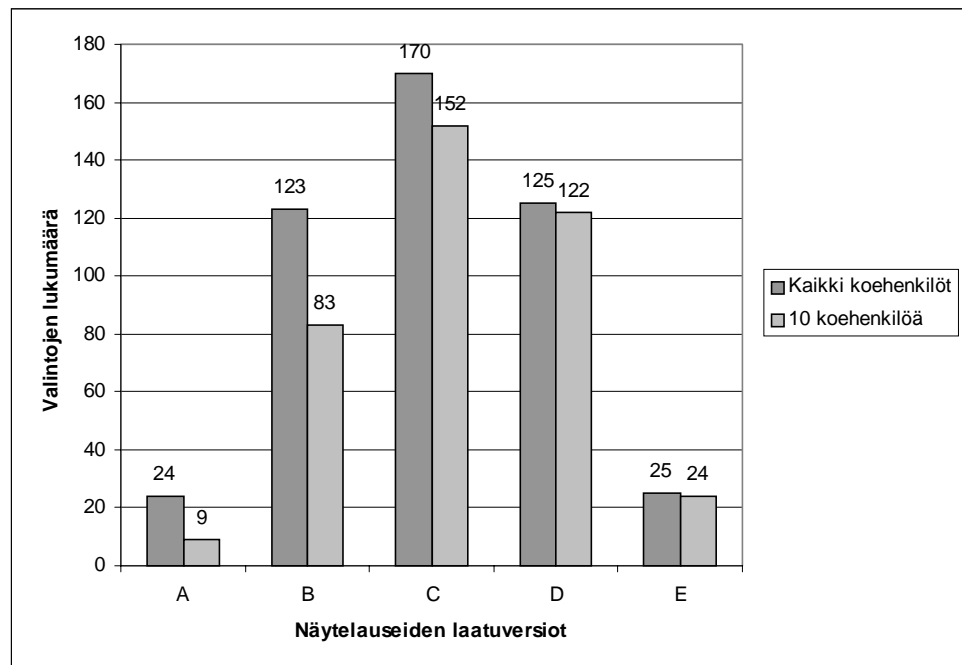
Lauseet ja puhujat oli sijoitettu kokeeseen sattumanvaraisessa järjestyksessä, jotta säännönmukaisuudesta ei aiheutuisi virheitä koetuloksiin.

Taulukko 9. Toisessa kuuntelukokeessa käytettyjen lauseiden eri laatuversioiden taajuusominaisuudet. Vaimennuskertoimen kertoo kehitetyn menetelmän mukaisesti keinotekoisesti tuotetun 4000 – 7700 Hz:n yläkaistan vaimennuksen voimakkuuden. Suurempi vaimennuskertoimen arvo kertoo voimakkaammasta vaimennuksesta.

| Äänen laatuversio | Taajuuskaista | Vaimennuskertoimen |
|-------------------|---------------|--------------------|
| A | 300-3200Hz | - |
| B | 300-3400Hz | - |
| C | 300-7700Hz | 2,0 |
| D | 300-7700Hz | 1,5 |
| E | 300-7700Hz | 1,2 |

6.2.2 Tulosten analysointi

Toisen kuuntelukokeen suoritti 12 henkilöä. Näiden tulosten perusteella vahvistuu päätelmä siitä, että taajuuskaistan laajennusmenetelmä parantaa äänen laatua. Samoin vahvistuu havainto siitä, että naisäänten kohdalla taajuuskaistan laajennus koetaan luonnollisempuna kuin miesäänten kohdalla.



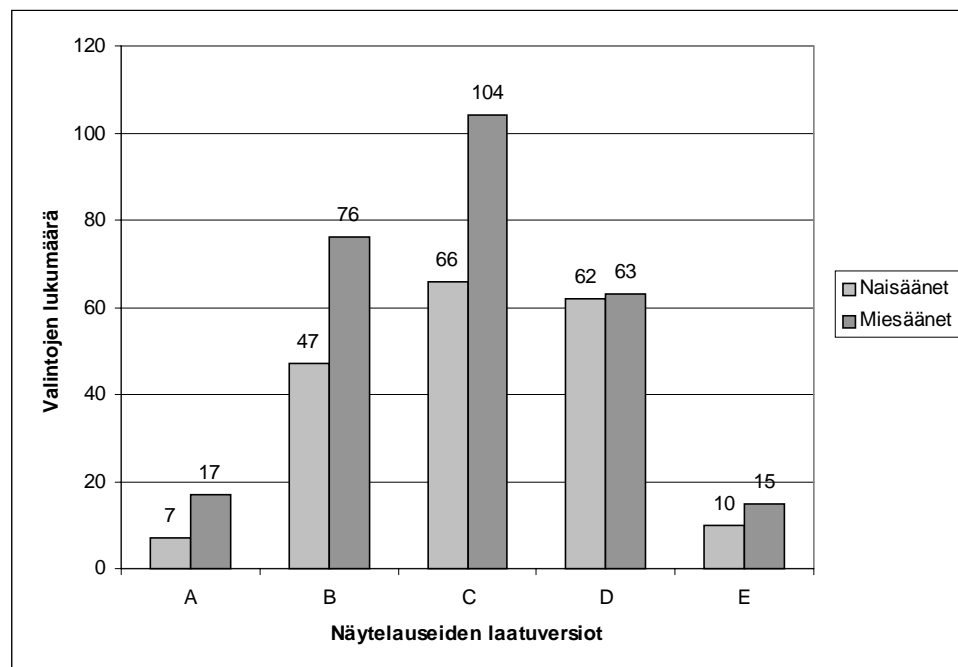
Kuva 34. Kuuntelijoiden valintojen jakauma näytelauseiden eri laatuversioiden (A, B, C, D ja E) kesken toisessa kuuntelukokeessa. Ensimmäinen palkki kertoo valintojen lukumäärän kaikkien koehenkilöiden valintojen perusteella ja toinen palkki 10 koehenkilön kesken.

Myös tässä kokeessa jouduttiin hylkäämään kahden kuuntelijan tulokset, sillä he olivat valinneet valitettavan useasti koeäänen laatuversion A, jonka taajuuskaista oli kavennettu puhelinkaistaakin kapeammaksi. Version A valinta saattaa johtua siitä, että puhelinääneltä ei yleensä odoteta kirkkautta, vaan on totuttu tukkoisempaan ääneen, joka siksi puhelimessa tuntuu kuulostavan luonnollisemmalta.

Kuvassa 34 on esitetty näytelauseiden eri laatuversioiden valintojen jakaumat sekä kaikkien koehenkilöiden tulosten että kymmenen kuuntelijan kesken.

Toisen kuuntelukokeen tulokset on analysoitu siten, että äänen laatuversio A vastaa pistemäärää 1, laatuversio B pistemäärää 2, laatuversio C 3:a, laatuversio D 4:ää ja laatuversio E 5:ttä pistettä. Myös tässä kuuntelukokeissa vastausten hajonta koeäänten eri laatuversioiden kesken oli hyvin pientä, joten tulosten analysointi keskiarvojen perusteella on mahdollista. Toisen kuuntelukokeen perusteella lasketut keskiarvot eivät kuitenkaan ole suoraan vertailtavissa ensimmäisen kuuntelukokeen vastaaviin lukuihin, mikä johtuu näytelauseissa käytettyjen vaimennuskertoimien eroista.

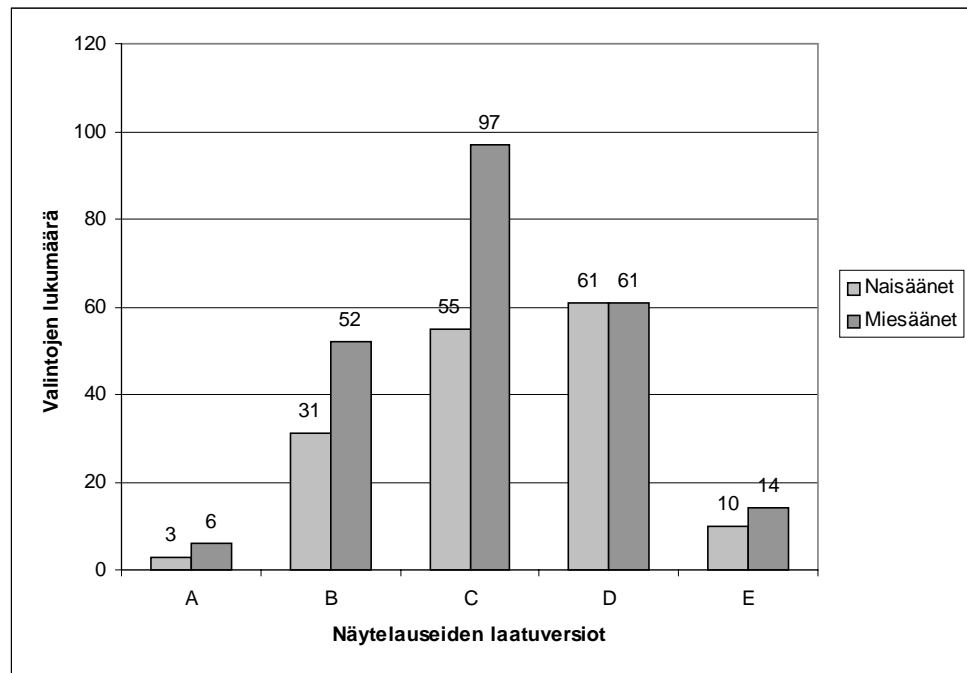
Kaikkien kuuntelijoiden arvioiden keskiarvoksi muodostuu tässä kokeessa 3,0, mikä kertoo, että miellyttävimmäksi on keskimäärin koettu äänen laatuversio C, jonka taajuuskaista oli laajennettu ja laajennusosaa oli vaimennettu kertoimella 2,0. Kun kaksi usein äänen laatuversiota A valinnutta kuuntelijaa jätetään pois tuloksista, saadaan keskiarvoksi 3,2. Optimaalinen vaimennuskerroin taajuuskaistan laajennetulle osalle voisi siis olla hieman pienempi kuin 2,0. Ei kuitenkaan niin pieni kuin 1,5, mitä on käytetty äänen laatuversiossa D.



Kuva 35. Näytelauseiden eri laatuversioiden (A, B, C, D ja E) valintojen lukumäärät nais- ja miesääninäytteiden suhteen toisessa kuuntelukokeessa, kun kaikkien koehenkilöiden arviot on huomioitu.

Naisäänille laskettu keskiarvo kymmenen kuuntelijan tuloksista on 3,3, ja vastaava luku miesäänille 3,1. Kaikkien kuuntelijoiden arviot huomioon otettaessa keskiarvot ovat 0,2 yksikköä pienempiä, niiden keskinäinen suhde on kuitenkin sama.

Kuvassa 35 näkyvät näytelauseiden eri laatuversioiden valintojen jakaumat naisäänten ja miesäänten osalta, kun kaikkien koehenkilöiden tulokset on otettu huomioon. Kuva 36 taas kertoo vastaavat naisäänten ja miesäänten jakaumat kymmenen koekuuntelijan tulosten perusteella.



Kuva 36 Näytelauseiden eri laatuversioiden (A, B, C, D ja E) valintojen lukumäärät kuunneltavien lauseiden naisäänten ja miesäänten osalta toisessa kuuntelukokeessa, kun vain kymmenen koehenkilön arviot on huomioitu.

6.2.3 Yhteenveto ja pohdintaa

Näiden kuuntelukokeiden perusteella voidaan siis sanoa, että kehitetty taajuuskaistan laajennusalgoritmi parantaa kapeakaistaisen puheäänien laatua, ja että sopiva vaimennuskerroin olisi hieman pienempi kuin 2,0. Algoritmi käyttää vaimennuskerrointa siten, että menetelmän mukaisesti määrätty logaritminen asteikon vaimennus kerrotaan vaimennuskertoimella. Logaritminen asteikko vastaa lineaarista paremmin sitä, miten ihminen äänet kuulee. Karkeasti voidaan siis sanoa, että vaimennuskertoimella 2,0 käsitellyt ylätaajuudet kuulostavat puolta vaimeammilta kuin vaimennuskertoimella 1,0 tuotetut ylä-äänit.

Se, että algoritmin avulla näytetään saavutettavan naisäänten kohdalla suurempi

äänen laadunparannus kuin miesäänten kohdalla tuntuu luonnolliselta, sillä onhan naisten äänissä luonnollisestikin enemmän korkeita taajuuksia kuin miesten äänissä. Joskus tosin ongelmaa saattaa aiheuttaa se, jos jokin formantti tai osa frikatiivin taajuuksista ovat jääneet puhelinkaistan 3,4 kHz:n yläpuolelle. Kun taajuuskaistan laajennusmenetelmällä käsitellyltä puheelta odotetaan tasaista laadun paranemista, kuulostavat äänteet, joista nuo kriittiset taajuudet puuttuvat ja jotka muuten ovat laajakaistaisia, helposti luonnottomilta ja häiritseviltä. Tämä ongelma nousee ehkä useammin esille naispuhujien kuin miespuhujien kohdalla.

Kuuntelukokeiden tulokset ovat kuitenkin rohkaisevia kertoessaan äänen laadun parantuvan kehitetyn menetelmän avulla.

7 YHTEENVETO

Tässä diplomityössä on tarkasteltu keinotekoisista taajuuskaistan laajennusta menetelmänä parantaa puhelinkaistaisen puhesignaalin laatua. Aluksi on käsitelty muutamia yleisiä digitaalisessa signaalinkäsittelyssä käytettyjä menettelytapoja, jotta lukijan olisi helpompi ymmärtää esiteltyjen taajuuskaistan laajennusmenetelmien toimintaperiaatteet.

Taajuuskaistan laajennusmenetelmän kehitystyön lisäksi diplomityössä on tutustuttu lähemmin subjektiivisiin kuuntelukokeisiin, koska äänenkäsittelymenetelmien kehitystyössä aikaansaadun äänen laadun määrittely on välttämätöntä, eikä suurpiirteisesti toteutettujen kuuntelukokeiden tuloksilla ole arvoa. Ennen uuden taajuuskaistan laajennusmenetelmän kuvausta työssä on tutustuttu vielä tätä aiemmin kehitettyjen taajuuskaistan laajennusmenetelmien toimintaperiaatteisiin, jotta katsaus äänen kaistalaajennuksen maailmaan olisi nyt kehitettyä menetelmää monipuolisempi.

Signaalinkäsittelyn digitalisoitumisen myötä aukenivat laajat mahdollisuudet uusien menetelmien kehittämiseen monen muun sovellusalueen lisäksi myös äänenkäsittelyssä. Äänen laadun parantuessa muillakin äänenkäsittelyn aloilla ovat paineet puhelimen välittämän äänen laadun parantamiseen alkaneet kasvaa. Erityisesti langaton tiedonsiirto kaipaa tehokkaiden äänenkoodausalgoritmien lisäksi keinoja parantaa välitetyn äänen laatua. Monet äänenkoodausalgoritmit kykenevät nykyisin välittämään puheääntä hyvin pienellä tiedonsiirtonopeudella. Nämä useimmiten häviölliset koodausalgoritmit tarvitsevat kuitenkin keinotekoisien taajuuskaistan laajennuksen lisäksi muitakin äänen laadunparannuskeinoja täyttääkseen jatkuvasti kasvavat äänen laatuvaatimukset.

Taajuuskaistan laajennus on tällä hetkellä vielä pioneerityötä, johon kenelläkään ei ole varmaa ratkaisua. Monenlaisia menetelmiä on kehitetty käyttäen hyväksi sekä digitaalisen signaalinkäsittelyn että informaatiotekniikan keinoja. Tässä diplomityössä on lähdetty etsimään taajuuskaistan laajentamiseksi menetelmää, joka olisi riittävän yksinkertainen ollakseen toteutettavissa mm. matkapuhelimen DSP:llä ja riittävän luotettava toimiakseen kaikissa häiriöolosuhteissa.

Tässä diplomityössä kehitetty taajuuskaistan laajennusmenetelmä on suunniteltu siten, että jatkuva-aikaisuuden asettamat vaatimukset tulevat täytetyiksi, eivätkä meluisat puheolosuhteet häiritse algoritmin toimintaa. Menetelmä perustuu taajuuskaistan hallittuun laskostamiseen, jonka avulla korkeat taajuudet luodaan ääneen. Yläkaistan tehon ja taajuuskomponenttien suhteellisten voimakkuuksien säätöön liittyvä laskenta on toteutettu taajuusalueessa, mikä vaatii vähemmän laskentaoperaatioita ja aiheuttaa signaaliin vähemmän viivettä kuin laskennan suoritus aika-alueessa aiheuttaisi.

Äänenkäsittelymenetelmän toiminnan arvioimisessa aikaansaadun äänen laadun arviointi on välttämätöntä. Koska objektiiviset äänen laadun arviointimenetelmät eivät

vielä nykyisin kykene kokonaisvaltaiseen laadun määrittämiseen, on subjektiivisilla kuuntelukokeilla tärkeä tehtävä arvioitaessa uuden äänenkäsittelymenetelmän toimivuutta. Subjektiivisten kuuntelukokeiden järjestäminen huolellisesti on ehdottoman tärkeää, jotta tutkimustulokset olisivat uskottavia. Tässä työssä toteutettiin subjektiiviset kuuntelukokeet kahdessa vaiheessa. Jo ensimmäinen koekierros osoitti, että menetelmän avulla saavutetaan parannus puhelinkaistaisen äänen laatuun. Toisella koekierroksella etsittiin algoritmille äänen luonnollisuuden kannalta optimaalisia parametreja.

Tässä diplomityössä kehitetyn taajuuskaistan laajennusmenetelmän ansiosta olemme päässeet taas yhden askeleen eteenpäin digitaalisen puheenkäsittelyn saralla. Äänen taajuuskaistan laajennusmenetelmien etsinnässä on vielä suuri osa digitaalisen signaalinkäsittelyn, informaatiotekniikan ja mahdollisesti uudentyyppisten digitaalisen tiedon käsittely- ja analysointimenetelmien tarjoamista mahdollisuuksista tutkimatta. Tässä työssä kehitettyä menetelmää on tarkoitus jatkossa jalostaa entistä tehokkaammaksi ja sen toimintakykyä tullaan testaamaan häiriöisemmillä puhesignaaleilla. Algoritmi voidaan kirjoittaa DSP-prosessorille, jolloin menetelmää päästään käyttämään enemmän todellisuutta vastaavissa tilanteissa.

LÄHTEET

- [Alahuhta 1979] Alahuhta, Matti. 1979 *Pulssikoodimodulaation (PCM) perusteita*. PCM-järjestelmät. INSKO, 1979. Julkaisu 113-79.
- [Bech 1991] Bech, Soren. 1991. *Selection and Training of Subjects for listening Tests on Sound-Reproducing Equipment*. Journal of Audio Engineering Society, Vol. 40, No. 7/8, 1992 July/August, ss. 590-609.
- [Burstein 1971] Burstein, Herman. 1971. *Attribute Sampling: Tables and Explanations*. McGraw-Hill, New York, 1971.
- [Burstein 1989] Burstein, Herman. 1989. *Transformed Confidence Limits for Listening Tests*. Journal of Audio Engineering Society, Vol. 37, No. 5, 1989 May, ss. 363-367.
- [Carlson 1989] Carlson, Bruce A. *Communication Systems*. McGraw-Hill 1986.
- [Chan, Hui 1996] Chan, Cheung-Fat; Hui, Wai-Kwong. *Wideband re-synthesis of Narrowband CELP-coded speech using Multiband Excitation Model*. Proceedings of ICSLP 96. ss. 322-325.
- [Cheng, O'Shaugnessy, Mermelstein 1992] Cheng, Yan Ming; O'Shaugnessy, Douglas. Mermelstein, Paul. 1992. *Statistical Recovery of Wideband Speech from Narrowband Speech*. Proceedings of ICSLP 92. ss. 1577-1580
- [Cheng, O'Shaugnessy, Mermelstein 1994] Cheng, Yan Ming; O'Shaugnessy, Douglas. Mermelstein, Paul. 1994. *Statistical Recovery of Wideband Speech from Narrowband Speech*. IEEE Transactions on Speech and Audio Processing. Vol. 2. No. 4. October 1994. ss. 544-548.
- [ETSI 1997] ETSI Recommendation Draft EN 301 245 v4.0.0 (1997-08). *Digital cellular telecommunications system (Phase 2); Enhanced Full Rate (EFR) speech transcoding (GSM 06.06 version 4.0.0)*
- [Gabrielsson, Sjögren 1979] Gabrielsson, A.; Sjögren, H.. *Perceived Sound Quality of Sound Reproducing Systems*. Journal of Acoustical Society of America, Vol. 65, 1979 April, ss. 1919-1933.
- [Honkanen, Vainio, Järvinen, Haavisto 1997] Honkanen, T.; Vainio, J.; Järvinen, K.; Haavisto, P. *Enhance full rate speech codec for IS-139 digital cellular system*. Proceedings of ICASSP 97. Ss.731-734.

- [ITU-T G.729] ITU-T Recommendation G.729. *Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)*. March 1996.
- [ITU-T P.800] ITU-T Recommendation P.800. *Series P: Telephone Transmission Quality, Methods for Objective and Subjective Assessment of Quality, Methods for Subjective Determination of Transmission Quality*. August 1996.
- [Karjalainen 1998] Karjalainen, Matti. *Äänenlaatu- monta näkökulmaa*. Karjalainen, M; Välimäki, V. (toim.) *Äänenlaatu*. Akustiikan seminaari 1998. Otaniemi 1998/Raportti 49.
- [Koistinen 1998] Koistinen, Tommi. *Kuuntelumenetelmät äänenlaadun arvioinnissa*. Karjalainen, M; Välimäki, V. (toim.) *Äänenlaatu*. Akustiikan seminaari 1998. Otaniemi 1998/Raportti 49.
- [Linde 1980] Linde, Y.; Buzo, A.; Gray, R.M. *An algorithm for vector quantizer design*. IEEE Transactions on Communication, COM-28, 1. ss. 84-95. January 1980.
- [Oppenheim 1989] Oppenheim, A. V.; Schaffer, R. W. 1989. *Discrete-Time Signal Processing*. Prentice Hall, Englewood Cliffs, N.J.
- [O'Shaughnessy 1987] O'Shaughnessy, Douglas. 1987. *Speech Communication, Human and Machine*. Reading, MA Addison-Wesley.
- [Owen 1982] Owen, Frank F. E. 1982. *PCM and Digital Transmission Systems*. McGraw-Hill.
- [Rabiner, Schaffer 1978] Rabiner, L. R.; Schaffer, R. W. 1978. *Digital Processing of Speech Signals*. Prentice Hall, Englewood Cliffs, N.J.
- [Schroeder, Atal 1985] Schroeder, Manfred R.; Atal, Bishnu S. 1985. *Code-Excited Linear Predictive (CELP) High-Quality Speech at Very Low Bit Rates*. Proceedings of ICASSP, pp. 937-940.
- [Soulodre 1998] Soulodre, Gilbert A.; Grusec, Theodore; Lavoie, Michel; Thibault, Louis. *Subjective Evaluation of State-of-the-Art 2-Channel Audio Codecs*. 104th Convention, Amsterdam, the Netherlands, Pre No. 4740, 1998 May 16-19.
- [Srednicki 1988] Srednicki, Mark. *A Bayesian Analysis of A-B Listening Tests*. Journal of Audio Engineering Society, Vol. 36, No. 3, 1988 March, ss. 143-146.
- [Toole 1982] Toole, F. E. *Listening Tests – Turning Opinion into Fact*. Journal of the Audio Engineering Society. Vol. 30. No. 6. S. 431-445.
- [Toole 1985] Toole, Floyd E. *Subjective Measurements of Loudspeaker Sound Quality and Listener Performance*. Journal of Audio Engineering Society, Vol. 33, No. 1/2, 1985 January/February, ss. 2-32.

- [Trancoso,Ribeiro,Almeida,Oliveira,Marques 1990]. Trancoso, I; Ribeiro, C; Almeida, L; Oliveira, L; Marques, J. *CELP: A Candidate for GSM Half-Rate Coding*. Int.Conf. on Acoust.,Speech and Sign.Proc. April 3-6, 1990. pp. 469-472.
- [Yasukawa 1994] Yasukawa, Hiroshi. *Quality Enhancement of Band Limited Speech by Filtering and Multirate Techniques*. Proceedings of ICSLP 94, Yokohama. pp. 1607-1610.
- [Yasukawa 1995] Yasukawa, Hiroshi. *Enhancement of Telephone Speech Quality by Simple Spectrum Extrapolation Method*. Proceedings of EUROSPEECH'95. pp.1545-1548.
- [Yasukawa 1996a] Yasukawa, Hiroshi. *Signal Restoration of Broad Band Speech Using Nonlinear Processing*. Proceedings of EUSIPCO'96. pp. 987-990.
- [Yasukawa 1996b] Yasukawa, Hiroshi. *Restoration of Wide Band Signal from Telephone Speech Using Linear Prediction Error Processing*. Proceedings of ICSLP 96. pp. 901-904.
- [Yoshida&Abe 1994] Yoshida, Yuki; Abe, Masanobu. *An Algorithm to Reconstruct Wideband Speech from Narrowband Speech Based on Codebook Mapping*. ICSLP 94, Yokohama. pp. 1591-1594.

LIITE 1

```
function wavfbe(snbfile,swbfile,att,Nfft,winln,frameln,slidln,aveln)

%WAVFBE Frequency band expansion for type .wav voice signal files.
% wavfbe(snbfile,swbfile,att,Nfft,winlen,framelen,slidlen,avelen)
% expands the frequency band of the narrowband signal in file
% snbfile and saves the corresponding wideband signal in file
% swbfile. The sampling frequency of the input signal is expected
% to be 8000Hz and the sampling frequency of the output signal
% will be 16000Hz.
% In addition to input and output filenames, a required input
% argument is att, which adjusts the highband attenuation to
% desired level.
% Optional arguments are:
% Nfft, the number of points in fast Fourier transformation,
% winln, the length of window used for cutting the input signal,
% frameln, the length of the usable frame in window,
% slidln, the number of samples averaged when cascading the
% frames and,
% aveln, the number of frequency components averaged when
% searching the maximum level of the spectrum.

% Author(s): J.K. Mahkonen, 1999

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                                                                                   VARIABLES AND ARGUMENTS
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
[snb, Fs, bits] = wavread(snbfile);
if(nargin < 8),
    aveln = 5; %Number of frequency components averaged
    if(nargin < 7),
        slidln = 100; %Length of the slide for averaging samples
        if(nargin < 6),
            winln = 240; %Length of the window 30ms = 240 samples
            if(nargin < 5),
                frameln = 160; %Length of the frame 20ms = 160 samples
                if(nargin < 4),
                    Nfft = 512; %Number of points in FFT
            end
        end
    end
end,end,end,end,end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                                                                                   EXPANDING THE FREQUENCY BAND AND WRITING THE RESULT IN A FILE
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
swb = fbe(snb', att, Nfft, winln, frameln, slidln, aveln);
wavwrite(swb', 2*Fs, bits, swbfile);
```



```

function swb = fbe(snb, att, Nfft, winlen, framelen, slidlen, avelen)

%FBE Frequency band expansion for 8kHz voice signal files.
%   swb = fbe(snb, att, Nfft, winlen, framelen, slidlen, avestart)
%   expands the frequency band of the narrowband signal snb
%   and returns a corresponding wideband signal in swb.
%   The sampling frequency of the input signal is expected to
%   be 8000Hz and the sampling frequency of the output signal
%   will be 16000Hz.
%   In addition to input and output signals, a required input
%   argument is att, which adjusts the highband attenuation to
%   desired level.
%   Optional arguments are:
%   Nfft, the number of points in fast Fourier transformation,
%   winlen, the length of window used for cutting the input signal,
%   framelen, the length of the usable frame in window,
%   slidlen, the number of samples averaged when cascading the
%   frames and,
%   avelen, the number of frequency components averaged when
%   searching the maximum level of the spectrum.

%   Author(s): J.K. Mahkonen, 1999

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%                               VARIABLES AND ARGUMENTS
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
if(nargin < 8),
    avelen = 5;                %Number of frequency components averaged
    if(nargin < 7),
        slidlen = 100;       %Length of the slide for averaging samples
        if(nargin < 6),
            winlen = 240;     %Length of the window 30ms = 240 samples
            if(nargin < 5),
                framelen = 160; %Length of the frame 20ms = 160 samples
                if(nargin < 4),
                    Nfft = 512;                %Number of points in FFT
                end
            end
        end
    end
end,end,end,end,end

if slidlen > 2*(winlen-framelen),    %Multiply by 2 for 16kHz signal
    error('Variable "slidlen" is too large.');
```

```
swb_frames = [];  
for i = 1:Nframes,  
    swb_frames(i,:) = unwin .* swb_frames_win(i,:);  
end  
  
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%  
%                                                                    CASCADING THE FRAMES  
swb = Cascade(swb_frames, 2*framelen, slidlen);
```

```

function [sig_frames_win, Nframes] = cutframes(sig, win, framelen)

%CUTFRAMES
% [sig_frames_win, Nframes] = cutframes(sig, win, framelen)
% cuts the input signal sig into peaces with the window win
% given as an argument so that the input argument framelen
% defines the number of samples that the window is forwarded for
% each cut peace.

% Author(s): J.K. Mahkonen, 1999

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% SOLVING VARIABLES
winlen      = length(win);           %Length of the window
slidelen    = (winlen-framelen)/2;   %Unused part of the window
siglen      = length(sig);          %Length of the signal
Nframes     = ceil(siglen/framelen); %Number of frames
Lastframelen = rem(siglen,framelen); %Number of samples
                                                %in the last frame

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% CUTTING FRAMES
frontz = zeros(1, slidelen); %Zeros to be added in front and back..
backz  = zeros(1, (framelen-lastframelen) + slidelen);%.of the signal
sig_z  = [frontz sig backz];    %Prepared signal to be
windowed
for i = 1:Nframes,
    for n = 1:winlen,
        sig_frames_win(i,n) = win(n) .* sig_z((i-1)*framelen + n);
    end
end

```

```

function swb = expand(snb, att, Nfft, avelen)

%EXPAND
%   swb = expand(snb, att, Nfft, avelen)
%       expands the frequency band of the given signalframe snb with
%       arguments given.
%   att, which regulates the attenuation level of high frequencies
%       has to be defined, but the other two arguments are optional.
%   Nfft tells the number of points in fft and avelen defines
%       how many frequency components are averaged when searching
%       for the maximum level of the spectrum.

%   Author(s): J.K. Mahkonen, 1999

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%                               VARIABLES AND ARGUMENTS
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
if(nargin < 4),
    avelen = 5;                               %Number of frequency components averaged
    if(nargin < 3),
        Nfft = 512;                           %Number of points in FFT
    end,end

lensnb = length(snb);                         %Length of the narrowband signal
lenswb = 2*lensnb;                            %Length of the wideband signal

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%                               CONTROLLED ALIASING OF THE FREQUENCY BAND
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
swb_ali = zeros(1, 2*lensnb);
for n = 1:lensnb,
    swb_ali(2*n-1) = snb(n);
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%                               FFT
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Swb_ali = fft(swb_ali, Nfft);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%                               ATTENUATING THE HIGHBAND
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Swb_att = Attenuate(Swb_ali, att, avelen);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%                               IFFT AND CUTTING OFF THE EXTRA ZEROS ADDED BY FFT IF NEEDED
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
swb_z = real(ifft(Swb_att));

if lenswb >= Nfft,
    swb_pow = swb_z;
else
    for n = 1:lenswb,
        swb_pow(n) = swb_z(n);
    end,end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%                               COMPENSATING THE EXTRA POWER BROUGHT BY ALIASING
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
power_nb = mean(snb.^2);
power_wb = mean(swb_pow.^2);

swb = swb_pow .* sqrt((power_nb/power_wb));

```

```

function S_att = attenuate(S_ali, att, avelen)

%ATTENUATE
%   S_att = attenuate(S_ali, att, avelen)
%   attenuates the frequencies from 4 to 8 kHz in spectrum S_ali
%   using att, which regulates the attenuation level and optional
%   argument avelen, which defines how many frequency components
%   are averaged when searching for the maximum level of the
%   spectrum.

%   Author(s): J.K. Mahkonen, 1999

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               VARIABLES AND ARGUMENTS
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
if(nargin < 3),
    avelen = 5;                %Number of frequency components averaged
end
Nfft = length(S_ali);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               SEARCHING FOR THE MAXIMUM LEVEL OF THE SPECTRUM
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
len = length(S_ali( (Nfft/4+2):(Nfft/2+1) ));
maxave = -Inf;
for i = (Nfft/4+2) : (Nfft/4+2)+len-avelen-1,
    ave = mean(abs(S_ali( i:(i+avelen-1) )));
    if ave > maxave,
        maxave = ave;
        nstart = i;
    end
end
[maxf, nmax] = max(abs(S_ali( nstart:nstart+avelen-1 )));
nmaxf      = nstart + nmax -1; %The number of the maximal component
maxflg     = 20*log10(abs(maxf)); %The maximum level in logarithm

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               CALCULATING THE AVERAGE LEVEL OF THE SPECTRUM IN 2-4kHz
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
avehf      = mean(S_ali(Nfft/8:floor(Nfft/4*3.4/4)));
avehflg    = 20*log10(max(avehf)); %The average level in logarithm

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               DEFINING THE ATTENUATION LINE
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
attnlg     = maxflg - avehflg; %Logarithmic attenuation for the highest...
%                               %...frequency component
attminlg   = - att * ((Nfft/4)./(nmaxf-Nfft/4) .* atnlg);
%                               %Logarithmic attenuation at 8kHz
attslidelg = attminlg .* [0:Nfft/4-2] ./(Nfft/4-2) ;
%                               %Logarithmic attenuation slide for 4-8kHz
attlinelg  = [zeros(1,Nfft/4+1) attslidelg attminlg rot90(rot90(
%                               attslidelg)) zeros(1,Nfft/4)];
%                               %The logarithmic attenuation line
attline    = 10.^(attlinelg./20); %The attenuation line in linear scale

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               ATTENUATING THE SPECTRUM
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
S_att      = S_ali .* attline;

```

```

function swb = cascade(swb_frames, framelen, slidlen)

%CASCADE
%   swb = cascade(swb_frames, framelen, slidlen)
%   cascades the signal windows given in argument swb_frames such
%   that the argument framelen tells how many samples the window is
%   for forwarded each window. slidlen is an optional argument,
%   which defines the number of samples averaged around the edges
%   of frames when cascading them.

%   Author(s): J.K. Mahkonen, 1999

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               VARIABLES AND ARGUMENTS
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
if(nargin < 3),
    slidlen = 100;           %Length of the slide for averaging samples
end

winlen = size(swb_frames, 2);           %Length of the window
frstart = (winlen-framelen)/2+1;       %The first sample of a frame
frend = winlen-frstart;                %The last sample of a frame
slidhalf= slidlen/2;                   %Number of samples to be averaged...
                                        %...in one end of a frame
edge(1) = frstart - slidhalf; %The first sample of the left side slide
edge(2) = frstart - 1;         %The last sample before the frame starts
edge(3) = frstart;            %The first sample of the frame
edge(4) = frstart + (slidhalf-1); %The last sample of left side slide
edge(5) = frstart + slidhalf;  %The first sample to be left untouched
edge(6) = frend - slidhalf;    %The last sample to be left untouched
edge(7) = frend - (slidhalf-1); %The first sample of right side slide
edge(8) = frend;              %The last sample of the frame
edge(9) = frend + 1;          %The first sample after the frame
edge(10)= frend + slidhalf;   %The last sample of the right side slide

Nframes = size(swb_frames, 1);         %The number of frames
leftslid = 1/(slidlen+1) * [1:slidlen]; %Averaging slides for left...
rightslid = rot90(rot90(leftslid));    %...and right end of a frame

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               SOFTENING THE EDGES OF FRAMES LINEARLY
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
swb_frames_ave = swb_frames;

%The first frame
swb_frames_ave(1, edge(7):edge(10)) = rightslid .*
    swb_frames(1, edge(7):edge(10));

for i=2:Nframes-1,           % Intermediate frames
    swb_frames_ave(i, edge(1):edge(4)) = leftslid .*
        swb_frames(i, edge(1):edge(4));
    swb_frames_ave(i, edge(7):edge(10)) = rightslid .*
        swb_frames(i, edge(7):edge(10));
end

% The last frame
swb_frames_ave(Nframes, edge(1):edge(4)) = leftslid .*
    swb_frames(Nframes, edge(1):edge(4));

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               CASCADING THE FRAMES ADDING THE EDGES TOGETHER
%                               %The first frame
endave(1,:) = swb_frames_ave(2, edge(1):edge(2)) +
              swb_frames_ave(1, edge(7):edge(8));
swb = [swb_ frames(1,edge(3):edge(6))  endave(1,:)];

for i=2:Nframes-1,                               %Intermediate frames
    startave(i,:)= swb_frames_ave(i-1,edge(9):edge(10)) +
                  swb_frames_ave(i, edge(3):edge(4));
    endave(i,:)= swb_frames_ave(i+1,edge(1):edge(2)) +
                 swb_frames_ave(i, edge(7):edge(8));
    swb =[swb startave(i,:) swb_frames(i,edge(5):edge(6)) endave(i,:)];
end

%The last frame
startave(Nframes,:) = swb_frames_ave(Nframes-1, edge(9):edge(10)) +
                      swb_frames_ave(Nframes, edge(3):edge(4));
swb = [swb  startave(i,:)  swb_frames(Nframes,edge(5):edge(8))];

```