PUHEEN TUOTTAMISEN KUVAAMINEN PARAMETROIMALLA KÄÄNTEISSUODATUKSELLA ESTIMOITU GLOTTISHERÄTE

Paavo Alku
Akustiikan ja äänenkäsittelytekniikan laboratorio, Teknillinen korkeakoulu
Paavo.Alku@hut.fi

Soinnillisen äänteen herätesignaali, värähtelevien äänihuulten välistä purkautuva glottisherätе, voidaan estimoida käyttämällä ns. käänteissuodatusmenetel mää. Puheen tuottamisen analyysi muodostuu tällöin tyypillisesti kahdesta vaiheesta: (a) glottisherätteen laskennasta käänteissuodatuksella ja (b) saatujen virtauspulssijonojen parametroinnista. Jälkimmäisen vaiheen tarkoitus on kuvata puheen tuoton herätesignaalin oleellisin informaatio numeerisessa muodossa. Tässä artikkelissa tarkastellaan niitä menetelmiä, joita on kehitetty glottisherätteen parametrointiin. Menetelmät kuvataan jakamalla ne aika- ja taajuusalueen tekniikoihin, ja jokaisen parametrin kohdalla on koostettu tietoa niiden käyttösovellutuksista ja tyypillisistä arvoista. Lopuksi vertaillaan tunnetuimpien tekniikoiden käytettävyyttä äänitutkimuksessa.

Avainsanat: *puheen tuottaminen, käänteissuodatus, glottisherätе, parametrointi*

1. JOHDANTO

Käänteissuodatus on laajasti sovellettu menetelmä soinnillisen äänteen herätesignaalin, äänihuulten välistä purkautuvan glottisherätteen, estimoimiseen (Miller, 1959; Fant, 1960). Käänteissuodatuksessa muodostetaan ensin malli glottisherätettä suodattaneelle ääniväylälle. Suodattamalla puhesignaali ääniväylän käänteiskuvauksella voidaan ääniväylän vaikutus kumota, ja tuloksena saadaan estimaatti äänen alkuperälle, ääniväylän herätesignaalille, glottispulssimuodolle. Käänteissuodatuksen idean puheen tuottamisen tutkimuksessa esitti ensimmäistä kertaa Miller (1959), jonka jälkeen on syntynyt lukuisia samaan ideaan perustuvia tekniikoita. Nämä voidaan jakaa kahteen ryhmään riippuen siitä, mitä informaatiosignaalia käänteissuodatuksessa käytetään lähtökohtana. Ensimmäisen käänteissuodatusalgoritmien ryhmän muodostavat menetelmät, joissa glottisherätе estimoidaan käyttämällä ns. Rothenbergin maskilla vastaanotettua suuaukon tilavuusnopeussignaalia (Rothenberg, 1973). Toisen ryhmän menetelmät käyttävät input-signaalinaan vapaasta kentästä mikrofonilla äänitettyä paineaaltoa (Wong ym., 1979; Alku, 1992; Gobl & NiChasaide, 2003). Tähän ryhmään kuuluvissa menetelmissä on käänteissuodatuksen huomioitava ns. huulisäteily, eli suuaukon tilavuusnopeussignaalin

KIITOKSET
Artikkeli liittyy Suomen Akatemian rahoittamaan projektiin (nr. 200859) "Multidisciplinary research project in the expression of emotion in spoken Finnish – Methodology for acoustical analysis of emotion in speech". Kiitos myös kahdelle nimettömälle arvioijalle, erityisesti sille tamperelaiselle, artikkelin kieliasua parantavista kommenteista.

muuttuminen paineaalloksi tietyn etäisyyden päässä puhujasta olevassa mikrofonissa. Voidaan osoittaa, että tämä vaikutus vastaa derivaattaa, siis aikasignaalin muutosnopeutta, puhetaajuuksilla (Flanagan, 1972). Tällöin on ymmärrettävää, että tietoa glottisvirtauksen tasakomponentista (DC-komponentti) ei voida saada, mikäli käänteissuodatus perustuu vapaan kentän paineaallon hyväksikäyttämiseen. Sen sijaan, mikäli lähtöinformaationa on suuaukolta mitattu virtaussignaali, voidaan kalibroidun Rothenbergin maskin avulla estimoida äänilähteestä paitsi signaalin aaltomuoto myös sen todelliset amplitudiarvot sekä vaihto- (AC) että tasakomponentista (DC).

Puheen tuottamisen analysoiminen käänteissuodatuksella toteutetaan tyypillisesti kahdessa vaiheessa. Ensimmäinen vaihe on varsinainen käänteissuodatus, joka tuottaa tuloksena glottisherätteen (tai sen derivaatan) aika-alueen signaalina. Toinen analyysin vaihe, pulssimuodon parametrointi, lähtee lasketusta, tyypillisesti useita satoja tai tuhansia diskreettejä näytteitä sisältävästä aikasignaalista, ja kuvaa tämän aaltomuodon oleellisimman informaation muutamalla numeerisella tunnusluvulla. Glottisherätteen parametrointi on tärkeä äänilähdetutkimuksen vaihe, koska siinä tehdyt päätökset esimerkiksi parametrin valinnan suhteen vaikuttavat siihen, miten käänteissuodatuksen antama informaatio näkyy tutkijalle. Parametrivalinnassa äänen tuottoa analysoivan tutkijan tulisi tuntea käytettävissä olevat parametrit ja ennen kaikkea se, mitä glottisherätteen ominaisuutta kukin tunnusluku mittaa. Tällainen glottisherätteen parametrien yleistuntemus auttaa valitsemaan kuhunkin käyttötarkoitukseen parhaiten sopivan tunnusluvun, joka edesauttaa tutkittavan ilmiön siirtymistä käänteissuodatuksen antamasta virtaussignaalista esimerkiksi datan tilastolliseen käsittelyyn ja lopulta tutkimuksen päätelmien tekoon.

Glottisherätteen parametrointimenetelmiä käytetään eniten puheen tuottamisen perustutkimuksessa. Tavoitteena on analysoida ja kuvata sitä suurta vaihtelua, mikä glottisherätteellä on, kun ihminen tuottaa erityyppistä soinnillista äännettä. Esimerkkejä tällaisista glottisherätteen ominaisuuksista, joita ihminen hyödyntää arkipäiväisessä puhekommunikaatiossaan, ovat äänen intensiteetin (Monsen & Engebretson, 1977; Holmberg ym., 1988; Gauffin & Sundberg, 1989; Dromey ym., 1992; Sundberg ym., 1993; Sulter & Wit, 1996) ja emootiosisällön (Laukkanen ym., 1996, 1997; Cummings & Clements, 1995; Gobl & Ni Chasaide, 2003) säätäminen. Vaikka näihin liittyvä puheen tutkimus on perinteisesti ollut perustutkimusta, on äänilähteen toiminnan kuvaamisella myös tällä alueella uusia sovellutuksia esimerkiksi rikostutkinnallisessa äänitutkimuksessa liittyen puhujan tunnistamiseen (Plumpe ym., 1999). Toinen glottisherätteen parametroinnin sovellutusalue on lääketieteellinen äänihäiriöiden (Hillman ym., 1989, 1990) ja äänen kuormittumisen tutkimus (Lauri ym., 1997; Vilkman ym., 1997). Kolmas sovellutusalue parametrointitekniikoille on puhesynteesi ja jossain määrin myös puheenkoodaus (Carlson ym., 1991; Childers & Hu, 1994; Childers, 1995). Näissä puheteknologian sovellutuksissa on glottisherätteen parametrointi noussut viime aikoina suuren kiinnostuksen kohteeksi, sillä parametroinnin tiedetään antavan tietoa, jota voidaan hyödyntää esimerkiksi syntetisoitaessa puhetta eri emootiokategorioissa (Campbell, 2003; Campbell & Mokhtari, 2003).

Tämä artikkeli on yleiskatsaus glottisherätteen parametrointiin. Tavoitteena on tutustuttaa lukija näihin puheentuottamisen kuvaamiseen kehitettyihin tekniikoihin, ja kuvata sitä, millaisia arvoja parametrit tyypillisesti saavat erilaisissa puheentuottotapahtumissa. Artikkeli ei kuvaa varsinaista käänteissuodatusta, vaan lähtöoletuksena pidetään tilannetta, jossa kään-

teissuodatus on tehty käyttäen joko suuaukon virtaussignaalia tai vapaan kentän paineaaltoa. Katsauksessa rajoitutaan lisäksi niihin menetelmiin, joissa parametroinnin lähtöinformaationa on ainoastaan käänteissuodatuksen antama glottisheräte (tai sen derivaatta). Sen sijaan sellaisia äänilähdettä kuvaavia menetelmiä, joissa yhdistetään käänteissuodatuksen tulos toiseen informaatiosignaaliin, esimerkiksi subglottaaliseen paineeseen (Sundberg ym., 1993), ei käsitellä. Mikäli jossakin parametrointimenetelmässä käytetään glottisherätteen derivaattaa, oletetaan artikkelissa, että se lasketaan aina diskreetissä aika-alueessa kahden peräkkäisen näytteen erotuksena (ts. digitaalisena suodatuksena FIR-suodattimella, jonka siirtofunktio on $1-z^{-1}$).

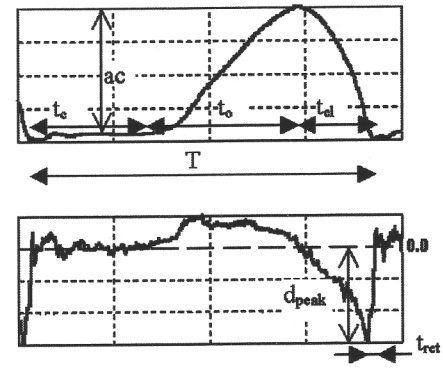
2. GLOTTISHERÄTTEEN PARAMETROINTIMENETELMIÄ

Käänteissuodatuksella estimoidun glottisherätteen parametrointiin on kehitetty useita eri tekniikoita. Seuraavassa näitä käsitellään kahdessa pääryhmässä riippuen siitä, tehdäänkö glottispulssijonon esittäminen numeerisessa muodossa aika- vai taajuusalueessa. Aika-alueen parametrit jaetaan lisäksi kahteen ryhmään: (1) aika-parametreihin, joissa käytettään pelkästään glottisherätteen (tai sen derivaatan) eri vaiheen aikakestoja, ja (2) amplitudiparametreihin, joissa aika-alueen signaalia kuvataan glottisherätteen (tai sen derivaatan) virtausarvoilla.

2.1 Aika-alueen menetelmät

Luonnollisin tapa kuvata glottisherätettä on analysoida pulssijonoa aika-alueen signaalina etsimällä siitä tietyt kriittiset ajanhetket kuten esimerkiksi äänihuulten avautumis- ja sulkeutumishetket ja näihin liittyvät virtausarvot. Tätä laskentaa on havainnollistettu kuvassa 1, jossa ylempi käyrä esittää vokaalista [a:] (miespuhuja, normaali ääntö) käänteissuodattamalla laskettua glottisherätettä ja alempi käyrä tämän derivaattaa. Käyttäen kuvan merkintöjä saadaan alla olevat glottisherätteen parametrit:

Suhteellinen aukioloaika (Open quotient, OQ) = $(t_o + t_{cl}) / T$ (Holmberg ym., 1988; 1989; Alku & Vilkman, 1996; Scherer ym., 1998)
(OQ korvataan joskus suhteellisella kiinnioloajalla (Closed quotient, CQ) = t_c / T = 1 - OQ (Sulter & Witt, 1996; Price , 1989; Iwarsson ym., 1998))
Pulssimuodon kallistuma (Speed quotient, SQ) = t_o / t_{cl} (Holmberg ym., 1988; 1989; Alku & Vilkman, 1996; Sulter & Witt, 1996)
Suhteellinen sulkeutumisaika (Closing quotient, ClQ) = t_{cl} / T (Holmberg ym., 1988; 1989; Alku & Vilkman, 1996; Sulter & Witt, 1996)
Suhteellinen palautumisaika (Return quotient, RQ) = t_{ret} / T (Price, 1989)
Normalisoitu amplitudisuhde (Normalized amplitude quotient, NAQ) = $ac / (d_{peak} T)$ (Alku ym., 2002; Bäckström ym., 2002)



Kuva 1. Yksi jakso käänteissuodattamalla estimoidusta glottisherätteestä (ylempi kuva) ja sen derivaatta (alempi kuva). Lähtökohtana oleva puhesignaali oli [a]-vokaali, jonka tuotti miespuhuja normaalilla ääntötavalla. Aikakestot: jakson pituus (T), sulkuvaihe (t_c), avautumisvaihe (t_o), sulkeutumisvaihe (t_{cl}), paluuvaihe (t_{ret}) (ts. aika, joka kuluu, kun derivaatta palaa negatiivisen maksimin jälkeen takaisin nollatasoon). Amplitudiarvot: AC-taso (ac), derivaatan negatiivinen maksimiarvo (d_{peak}).

2.1.1. Aikaparametrit

Aikaparametreilla tarkoitetaan tunnuslukuja, jotka kuvaavat glottisherätettä (tai sen derivaattaa) käyttämällä pelkästään aikakestoja (siis ei esimerkiksi virtausarvoja) mittaavia lukuja. Aikaparametrien etu on tällöin siinä, että niillä voidaan karakterisoida glottisheräte riippumatta siitä, millä amplitudiasteikolla signaali on esitetty. Tällöin aikaparametrejä käytettäessä voidaan glottisheräte parametroida välittämättä esimerkiksi siitä, onko verrattavat puhenäytteet äänitetty vakioetäisyydellä. Puheen tuottamisen tutkimuksessa käytetyimmät glottisherätteen parametrit ovat OQ, SQ ja ClQ. Holmberg ym. (1988) tarkastelivat näitä parametreja laajahkossa tutkimuksessa, jonka kohteena olivat äänilähteen ominaisuudet puhujan tuottaessa ääntä kolmessa eri äänekkyysmoodissa (hiljainen, normaali ja voimakas puhe). Holmberg ym. käyttivät Rothenbergin maskiin perustuvaa käänteissuodatusta glottisherätteen analysoimiseksi. Heidän tutkimuksessaan oli 45 koehenkilöä (25 miestä, 20 naista). Aika-parametrit osoittivat, että glottispulssin muoto muuttuu särmikkäämmäksi (ts. OQ laskee ja SQ kasvaa), kun puhuja voimistaa ääntään. Tästä trendistä poikkeuksen muodosti ne naisten tuottamat äänet, joissa äännön voimakkuus muuttui normaalista voimakkaaksi. Sulter ja Wit (1996) analysoivat äänilähteen peräti 224 eri puhujalta. Puhujat, joiden tehtävänä tässäkin tutkimuksessa oli tuottaa soinnillista äännettä kolmessa eri äänekkyysluokassa, jaettiin kahteen ryhmään sen mukaan, olivatko he saaneet äänikoulutusta. Sulterin ja Witin (1996) saamien tulosten mukaan äänikoulutuksen vaikutus ei juurikaan näkynyt glottisherätteen parametreissa. Tilastollisesti merkitsevä efekti näkyi ainoastaan naispuhujien ClQ:ssa, joka oli suurempi niillä puhujilla, jotka olivat saaneet äänikoulutusta, sekä miespuhujien SQ-arvossa, joka myös oli suurempi koulutusta saaneilla. Puhujan sukupuolella havaittiin sen sijaan olevan vaikutusta useisiin parametreihin. Esimerkiksi miesäänten glottisherätteen suhteellinen kiinnioloaika (ts. parametri CQ) oli suurempi kuin naispuhujilla, kun taas suhteellinen sulkeutumisaika (ts. parametri ClQ) oli miehillä pienempi kuin naisilla. Aikaparametrien OQ, SQ and ClQ saamat tyypilliset arvot voidaan koostaa em. kahdesta tutkimuksesta (arvot on tuotettu normaalilla äänen voimakkuudella): Holmberg ym:n (1988) mukaan miespuhujilla keskimääräinen OQ oli 0,60, keskimääräinen SQ 1,82 ja keskimääräinen ClQ 0,22. Naispuhujille nämä kolme parametria olivat keskimäärin arvoissa 0,76, 1,65 ja 0,29. Sulter ja Wit (1996) saivat hieman poikkeavia parametriarvoja: miespuhujilla OQ oli keskimäärin 0,49, SQ 1,52 ja ClQ 0,20. Naispuhujille vastaavat parametrit olivat Sulterin ja Witin (1996) mukaan 0,55, 1,36 ja 0,2.

Dromey ym. (1992) tutkivat glottisherätteen aikaparametrien muutosta kokeessa, jossa puhujan tehtävä oli voimistaa äänen voimakkuutta 5 dB:n askelin. Tutkimuksessa analysoitiin kymmenen naispuhujaa ja todettiin, että OQ-parametri pieneni äänen voimakkuuden kasvaessa. Sen sijaan SQ:n arvo ei seurannut monotonisesti äänenpainetasoa (Sound Pressure Level, SPL), vaan sen arvo ensiksi kasvoi äänen voimakkuuden lisääntyessä, mutta suurimmilla intensiteettiarvoilla alkoi jälleen laskea. Price (1989) tutki eroja mies- ja naispuhujien glottisherätteissä käyttämällä parametreja CQ ja RQ. Hänen saamiensa tulosten mukaan naispuhujilla oli lyhempi suhteellinen kiinnioloaika (siis pienempi CQ-arvo) kuin miespuhujilla. Lisäksi glottispulssin sulkuvaiheen suhteellista pituus mitattuna ääniväylän pääherätteen aikahetkestä glottiksen sulkuhetkeen oli naisilla suurempi. Glottisherätteen aikaparametrejä on lisäksi käytetty tutkittaessa muu-

toksia äänellisen kuormituksen aikana (Lauri ym., 1997; Vilkman ym., 1997). Näissä tutkimuksissa on havaittu mm. naispuhujien äänentuoton muuttuminen hyperfunktionaaliseen suuntaan, mitä ilmensivät kasvanut SQ:n arvo ja pienentynyt ClQ:n arvo. Puheäänten lisäksi glottisherätteen aikaparametrejä on käytetty lauluäänen tutkimisessa (Iwarsson ym., 1998; Sundberg, Andersson & Hultqvist, 1999; Sundberg, Cleveland, Stone & Iwarsson, 1999). Aikaparametrien antaman äänilähteen objektiivisen kuvauksen mukaan on esimerkiksi havaittu, että country-laulajat tuottavat lähes saman typpisen glottisvirtauksen sekä puhuessaan että laulaessaan (Sundberg, Cleveland, Stone & Iwarsson, 1999).

Kaikkien edellä esitettyjen aikaparametrien laskenta edellyttää kriittisten aikahetkien kuten glottiksen avautumishetken, maksimaalisen virtauksen hetken ja glottiksen sulkeutumishetken etsimistä käänteissuodatuksen antamasta glottisherätteen estimaatista. Estimoitu glottisherätteen aaltomuoto on usein kohinainen johtuen ennen kaikkea käänteissuodatuksessa tapahtuneesta epätäydellisestä ääniväylän kumoamisesta. Tällöin kriittisten aikahetkien määrittäminen on hankalaa, ja lasketut arvot saattavat vaihdella jaksosta toiseen. Vaikka estimoitu glottisheräte olisi kohinaton, ovat tietyt ajanhetket, erityisesti glottiksen avautumishetki, vaikeita määrittää, sillä aaltomuodossa ei välttämättä tapahdu hetkellistä, selvästi määriteltävää muutosta. Näistä ongelmista johtuen on aikaparametrit joissain yhteyksissä laskettu käyttäen kriittisten aikahetkien määritelmiä, joita ei voida suoraan liittää äänihuuliväräh-telyn fysiologisiin tapahtumiin kuten avautumis- ja sulkeutumishetkeen. Tällöin voidaan esimerkiksi OQ-parametrin laskennan tarvitsema glottiksen aukiolovaiheen pituus korvata keinotekoisella aikakestolla, joka määräytyy siitä ajasta, jonka virtaussignaali on tietyn ennalta määrätyn suhteen (esim. 50 %) verran virtausminimin yläpuolella (Dromey ym., 1992; Sapienza ym., 1998).

Glottisherätteen aikaparametroinnin helpottamiseksi voidaan käyttää muitakin menetelmiä, kuin edellä mainittua keinotekoisten kriittisten aikahetkien määrittämistä. Eräs tällainen menetelmä on vast'ikään esitetty aikaparametri NAQ (Normalized Amplitude Quotient) (Alku ym., 2002; Bäckström ym., 2002). NAQ:n erikoisuus on siinä, että se tuottaa glottisherätteen aikaparametrin, mutta laskenta tehdään ilman kriittisten aikapisteiden etsintää. Voidaan osoittaa (Fant, 1997), että ottamalla suhde kahdesta amplitudiarvosta, virtauksen AC-arvosta ja virtausderivaatan negatiivisesta maksimista, saadaan aikakesto, joka on glottiksen sulkeutumisvaiheen osa (ks. Kuva 1). Siinä missä ClQ mittaa koko sulkeutumisvaiheen pituutta, keskittyy NAQ kyseisen osan energeettisimpään alueeseen. NAQ on siis läheistä sukua ClQ:lle ja näiden välillä on suuri korrelaatio (Alku ym., 2002). Koska NAQ:n laskennan käyttämät molemmat amplitudiarvot (sekä virtauksesta että sen derivaatasta) ovat jakson maksimeja, on niiden määrittäminen helppoa ja niiden arvot eivät ole kovin häiriöalttiita käänteissuodatuksen artefaktoille.

2.1.2. Amplitudiparametrit

Kun käänteissuodatus tehdään suuaukon virtaussignaalista käyttämällä kalibroitua Rothenbergin maskia, voidaan saatavaan glottisherätteen estimaattiin liittää relevantti amplitudi-informaation (Rothenberg, 1973). Kuvassa 2 on esitetty käytetyimmät amplitudiparametrit sekä virtaussignaalista (ylempi kuva) että sen derivaatasta (alempi kuva). Tavallisimmin käytetyt kolme amplitudiarvoa ovat virtaussignaalin minimiarvo eli tasakomponentti (DC-komponentti), tämän erotus virtauksen huippuarvosta (AC-

komponentti) sekä derivaatan negatiivinen huipputaso. Nollasta poikkeavan DC-komponentin tavallisin fysiologinen selitys on se, että äänihuulet eivät ole sulkeutuneet täydellisesti. Vuotoisan (engl. breathy) äännön tapauksessa on tyypillistä, että glottis ei sulkeudu täysin edes äänihuulten keskikohdasta. Normaalissa ääntötavassa on mahdollista, että sulkeutuminen on jokseenkin täydellistä äänihuulten lähes koko pituudelta lukuunottamatta niiden takaosaa, jonne jää virtausta läpi päästävä aukko. Toinen tekijä, joka tuottaa nollasta poikkeavan DC-komponentin on äänihuulten vertikaalinen liike. DC-komponentti esiintyy glottisherätteessä varsin usein: Holmberg ym:n (1988) tutkimuksessa todettiin, että lähes kaikissa analysoiduissa glottisherätteissä oli nollasta poikkeava DC-komponentti. Heidän tutkimuksessaan todettiin lisäksi, että DC-virtaus kasvoi merkitsevästi siirryttäessä normaalilla voimakkuudella tuotetuista äänistä hiljaisiin ääniin. Mies- ja naispuhujien kesken ei Holmberg ym:n (1988) tutkimuksessa löydetty eroa DC-komponentin esiintymisessä. DC-komponentin keskiarvo normaalilla äänen voimakkuudella on todettu olevan noin 0,10 l/s (Holmberg ym, 1988; Sulter & Witt, 1996).

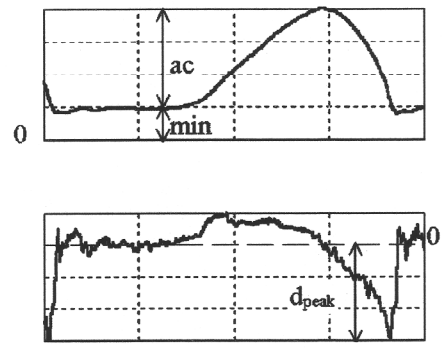
Glottisvirtauksen AC-amplitudin on todettu korreloivan syntyvän puhesignaalin äänenpainetason, SPL:n, kanssa: mitä voimakkaampi ääni, sitä suurempi virtauksen AC-amplitudi on (Hertegård & Gauffin, 1995). Virtauksen AC-amplitudin on lisäksi osoitettu korreloivan lähdesignaalin spektrin F0:n amplitudin kanssa (Gauffin & Sundberg, 1989). Miespuhujilla AC-amplitudi on tyypillisesti suurempi kuin naispuhujilla johtuen siitä, että miesten äänihuulet ovat pitemmät ja siksi värähtelevien äänihuulten väliin jäävän glottiksen maksimipinta-ala on myös suurempi (Holmberg ym., 1988; Hertegård, 1994). Tyypillisiä AC-amplitudille saatuja arvoja on seuraavat: Holmberg ym. (1988) ilmoittivat normaalilla voimakkuudella tuotettujen äänten AC-parametrin keskiarvoksi 0,26 l/s miespuhujilla ja 0,14 l/s naispuhujilla. Sulterin ja Witin (1996) mittauksissa raportoitiin selvästi suuremmat arvot: keskiarvo miesäänten AC-amplitudille oli 0,57 l/s ja naispuhujille 0,26 l/s.

On myös mahdollista yhdistää glottisvirtauksen AC-amplitudi ja DC-taso yhteen amplituditason parametriin (Isshiki, 1981). On osoitettu, että DC-arvon ja AC-amplitudin suhde korreloi äänen subjektiivisesti havaitun vuotoisuuden kanssa (Frizell ym., 1986).

Ääniväylän tärkein akustinen heräte syntyy glottiksen sulkeutumisvaiheessa virtauksen hidastuman saavuttaessa hetkellisen maksimiarvonsa (Fant, 1960). Koska tämä hetki on energeettisesti tärkein puheen tuottotapahtuman hetki, on luonnollista käyttää kyseistä aikahetkeä parametroinnissa. Käytetyin amplituditason parametri, joka keskittyy tähän ääniväylän pääeksitaation hetkeen, on glottisvirtauksen derivaatan negatiivinen maksimiamplitudi (e.g., Holmberg ym, 1988; Gauffin & Sundberg, 1989; Sundberg ym., 1993; Sulter & Witt, 1996; Fant, 1997). Tämän amplitudiarvon tiedetään korreloivan vahvasti syntyvän äänen SPL:n kanssa (Gauffin & Sundberg, 1989). Derivaatan negatiiviselle maksimille on saatu seuraavia tyypillisiä arvoja: Holmberg ym. (1988) raportoivat ko. amplitudiarvon olevan normaalivoimakkuudella äännettäessä keskimäärin 280 l/s^2 miespuhujilla ja 164 l/s^2 naispuhujilla, kun taas vastaavat arvot Sulterin ja Witin (1996) tutkimuksessa olivat 1026 l/s^2 ja 504 l/s^2.

Kaikissa edellä käsitellyissä menetelmissä äänen tuottamisen parametrointi perustuu tiettyjen aika- tai amplitudiarvojen erottamiseen glottisherätteen estimaatista. On myös mahdollista parametroida käänteissuodat-

tamalla saatu glottisheräte tai sen derivaatta käyttämällä ennakolta valittua, tietyistä matemaattisista funktioista määräytyvää aaltomuotoa, joka sovitetaan alkuperäiseen glottisherätteeseen. Toisin sanoen tällaisilla parametreillä pyritään mallintamaan koko aaltomuoto eikä sen yksittäisiä merkittäviä näytteitä. Käytetyimpiä tähän tarkoitukseen käytettyjä synteettisiä pulssimuotoja on ns. Liljencrants-Fant-malli (LF-malli), jossa glottisherätteen derivaattaa kuvataan yhden jakson aikana kosini- ja eksponentiaalifunktioilla, jotka määräytyvät viidestä numeerisesta arvosta (Fant ym., 1985). LF-mallia on käytetty puheen tuottamisen parametroinnissa yhdessä automaattisen käänteissuodatuksen kanssa parametroimalla äänen tuottoa esimerkiksi eri ääntötyypeissä (Strik & Boves, 1992; Fröhlich ym., 2001). On myös mahdollista käyttää polynomia glottisvirtauksen mallintamisessa (Childers & Ahn, 1995).



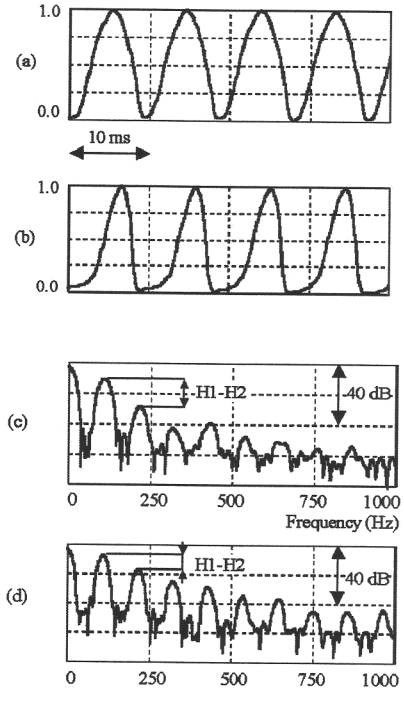
Kuva 2. Yksi jakso käänteissuodatuksella estimoidusta glottisherätteestä (ylempi kuva) ja sen derivaatta (alempi kuva). Lähtökohtana oleva puhesignaali oli [a]-vokaali, jonka tuotti miespuhuja normaalilla ääntötavalla. Amplitudiarvot: Virtauksen minimitaso (min), AC-taso (ac), derivaatan negatiivinen maksimiarvo (d_{peak}).

2.2. *Taajuusalueen menetelmät*

Kun ihminen muuttaa äänentuottotapaa, on tästä usein seurauksena glottisherätteen spektrin kaltevuuden (engl. spectral decay, spectral tilt) muutos. Näin tapahtuu esimerkiksi silloin, kun äänen tuotossa muutetaan äänen voimakkuutta: hiljainen ääni merkitsee tavallisimmin sitä, että glottisvirtaus on yleismuodoltaan pyöreä, jolloin sen spektri vaimenee jyrkästi taajuuden kasvaessa. Voimakkaamman äänen synnyttäminen edellyttää tavallisesti särmikkäämpää muotoa glottisherätteen aika-alueen signaalissa, mikä taajuusalueessa merkitsee hitaammin vaimenevaa spektrin verhokäyrää. Tästä syystä glottisherätteen parametrointi tehdään usein taajuusalueessa mittaamalla käänteissuodatuksella saadun glottisherätteen spektrin verhokäyrän vaimenemista taajuuden funktiona. Spektri lasketaan tavallisesti perinteisellä FFT-muunnoksella. Se voi olla joko ns. pitch-asynkroninen spektri, jolloin informaation tarkastelu taajuusalueessa tehdään tyypillisesti käyttäen harmonisia komponentteja. Joissain tekniikoissa taajuusmuunnos tehdään yksittäiselle glottispulssijonon jaksolla (ns. pitch-synkroninen spektri), jolloin luonnollisesti käsittely ei voi perustua harmonisiin komponentteihin. FFT:n asemesta on myös mahdollista sovittaa glottissignaaliin parametrinen spektri käyttäen esimerkiksi all-pole-tyyppistä spektriä.

Childers & Lee (1988) ovat parametroineet glottisherätettä taajuusalueessa käyttäen pitch-asynkronisen spektrin harmonisia komponentteja. Heidän kehittämänsä menetelmä, Harmonic Richness Factor (HRF), määritetään harmonisten amplitudisuhteena, jossa osoittajaan summataan perustaajuutta (F0) suurempien harmonisten tasot ja nimittäjässä on F0:n taso. Tämän parametrin on osoitettu heijastavan ääntötyyppiä: narinaääntä kuvaa suuri arvo, normaaliääntöä keskisuuri kun taas vuotoisaa ääntöä vastaa kaikkein pienin HRF:n arvo. Samantyyppistä periaatetta käyttivät myös Howell & Williams (1988; 1992), jotka mittasivat glottisherätteen spektrin vaimenemista sovittamalla alimpiin harmonisiin

regressiosuoran ja mittaamalla spektrin kallistumaa regressiosuoran kulmakertoimella. Paljon käytetty glottisherätteen taajuusalueen parametrointitekniikka on myös ns. H1-H2-arvo, jossa spektrin vaimenemista mitataan F0:n ja toisen harmonisen tasoerolla (ks. kuva 3). Tämän parametrin on osoitettu mm. korreloivan positiivisesti lauluäänissä aikaparametrin CQ kanssa (Titze & Sundberg, 1992). Alku ym. (1997) esittivät menetelmän, joka perustuu pitch-synkronisen glottisherätteen spektrin käyttöä äänen tuottamisen parametroinnissa. Menetelmä, Parabolic Spectral Parameter (PSP), käyttää glottisspektrin vaimenemisen kvantifiointiin toisen asteen polynomia, joka



Kuva 3. Miespuhujan [e]-vokaalista käänteissuodatuksella estimoidut glottisherätteet vuotoisan ääntötavan (kuva 3a) ja puristeisen ääntötavan (kuva 3b) tapauksessa (y-akseli arbitaarinen). Vuotoisan ääntötavan glottisherätteen spektri on esitetty kuvassa 3c ja puristeisen ääntötavan glottisherätteen spektri kuvassa 3d. Glottisspektrin kaltevuutta on parametroitu H1-H2-tunnusluvulla (Sundberg ym., 1993), jonka arvo kuvassa 3c on 18,4 dB ja kuvassa 3d 9,6 dB.

sovitetaan optimaalisesti pitch-synkroniseen spektriin käyttäen neliösummakriteeriä. Tällöin spektrin vaimeneminen kuvautuu toisen asteen polynomin yhteen kertoimeen. PSP-laskentaan kuuluu, tiettävästi ainoana menetelmänä, normalisointi, jolla spektrin vaimeneminen suhteutetaan F0:sta riippuvaan teoreettiseen maksimaaliseen spektrin kaltevuuteen. PSP:n on osoitettu pystyvän tehokkaasti erottelemaan eri ääntötyypeillä tuotetuista äänistä lasketut glottispulssijonot.

3. JOHTOPÄÄTÖKSET

Soinnillisen äänteen herätteenä toimivan glottispulssijonon laskentaan käänteissuodatuksella liittyy lähes aina saatujen pulssimuotojen parametrointi eli signaalin kuvaaminen tunnusluvuin. Puheen tuottamisen tutkimuksessa, varsinkin suunniteltaessa isohkoja mittausasetelmia, on syytä tarkoin miettiä, mitä parametrointimenetelmää tulisi käyttää, mikäli tutkimuksessa hyödynnetään käänteissuodatusta. Oli valittu parametrointimenetelmä mikä tahansa, jää osa alkuperäisen glottisherätteen informaatiosta sitoutumatta kyseiseen tunnuslukuun. Valitsemalla kyseessä olevaan tutkimuskohteeseen parhaiten soveltuva parametri voidaan vähentää tällaista "hukkaan valuvaa" informaatiota kuvattaessa käänteissuodatuksen antamia pulssimuotoja.

Useat piirteet puhesignaalissa, esimerkkinä äänen intensiteetin säätö, määräytyvät glottisherätteen kannalta etupäässä virtauspulssin ominaisuuksista ääniraon sulkeutumisvaiheen aikana. Aikaparametrien suhteen tämä tarkoittaa sitä, että käytettävien parametrien joukkoon tulisi valita joko ClQ:n tai NAQ:n. Mikäli käsiteltävää dataa on paljon ja on olemassa riski, että pulssimuodot ovat käänteissuodatuksessa vääristyneitä, on perusteltua käyttää NAQ-parametria. Sulkeutumis-

vaiheen tärkeys glottisvirtauksessa merkitsee myös sitä, että glottisvirtauksen derivaatan negatiivista maksimiarvoa tulisi hyödyntää valittaessa amplitudiparametria. Tämän arvon luotettava laskenta olisi tehtävä käyttäen kyllin laajaa puhesignaalin kaistaleveyttä: varsinkin silloin, kun käsiteltävä materiaali sisältää puristeisella ääntötavalla tai suurella SPL-tasolla tuotettua puhetta, olisi kaistaleveyden oltava vähintään 4 kHz (Alku & Vilkman, 1995). Kaistaleveyden valinta päätetään tavallisesti siinä vaiheessa, kun äänitettyjä puhenäytteitä aletaan siirtää tietokoneelle käänteissuodatusta varten. Useimmissa automaattisissa käänteissuodatusmenetelmissä (esim. Alku, 1992) ei ole rajoituksia kaistaleveydelle, jolloin kaistaleveys on tutkijan itsensä asetettavissa ja tällöin on syytä valita tarpeeksi laaja taajuuskaista.

Kun viime vuosikymmeninä tehtyjä puheen tuottamisen tutkimuksia tarkastellaan, on hieman yllättävää todeta, että glottisherätteen parametrointi tehdään useimmiten käyttäen pelkkiä aika-alueen menetelmiä. Outoa on se, että monet tehdyistä tutkimuksista koskevat sellaisia puheen tuottamisen ilmiöitä, joissa erojen voitaisiin olettaa näkyvän juuri glottisherätteen spektrin kaltevuudessa. Olisi siis perusteltua hyödyntää enemmän taajuusalueen tekniikoita sen sijaan, että käytetään glottisvirtauksen aika-alueen muodon kuvaamiseen useita rinnakkaisia aikaparametrejä.

Mikäli puheen tuottamisen tutkimukseen voidaan liittää virtausmaskia käyttävä käänteissuodatus, saadaan äänilähteestä arvokasta amplituditason informaatiota (ts. kuvan 2 mukaiset minimiarvo ja AC-taso). On kuitenkin syytä muistaa, että maskin käyttö rajoittaa luonnollista puheen tuottamista varsinkin koehenkilöillä, joilla ei ole aikaisempaa kokemusta virtausmaskin käytöstä. Joissain mittauksissa, esimerkkinä äänen kuormituksen tarkastelu realistisessa ympäristössä henkilön suorittaessa työtehtäviään, on maskin käyttö täysin poissuljettu. Toinen maskin käyttöä rajoittava tekijä on sen vaikutus käänteissuodatuksessa käytettävän virtaussignaalin kaistaleveyteen: Hertegård & Gauffin (1992) ovat osoittaneet, että maskin tasainen amplitudivaste ulottuu vain 1.5 kHz:iin. Tällainen kaistarajoitus on parametrointia vahvasti vääristävä tekijä varsinkin silloin, kun äänimateriaalissa on samanaikaisesti mukana vähän korkeita taajuuksia sisältävää puhetta (esim. hiljaiset äänet tai vuotoisalla ääntötavalla tuotettu puhe) sekä ääniä, joissa korkeiden taajuuksien osuus on merkittävä (esim. voimakkaat äänet ja puristeisella ääntötavalla tuotettu puhe). Mikäli maskin käytön rajoitukset koetaan vakavina, on puheen tuottamisen tutkimus tehtävä vapaan kentän paineaaltoa hyödyntäen, jolloin voidaan tuottaa täysin luonnollista puhetta eikä kaistaleveys rajoitu. Vaikka tällöin menetetään tieto todellisista glottisvirtauksen amplitudiarvoista, saadaan kuitenkin herätesignaalin oleellisin aika- ja taajuusalueen informaatio kuvattua käyttämällä hyväksi aikaparametreja (erityisesti ClQ tai NAQ) tai taajuusalueen lähdespektrin kaltevuutta mittaavia tunnuslukuja (esimerkiksi H1-H2 tai PSP).

VIITTEET

Alku, P. (1992). Glottal wave analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering. Speech Communication, 11, 109-119.

Alku, P., Bäckström, T. & Vilkman, E. (2002). Normalized amplitude quotient for parameterization of the glottal flow. Journal of the Acoustical Society of America, 112, 701-710.

Alku, P., Strik, H. & Vilkman, E. (1997). Parabolic Spectral Parameter - A new method for quantification of the glottal flow. Speech Communication, 22, 67-79.

Alku, P. & Vilkman, E. (1995). Effects of bandwidth on glottal airflow waveforms estimated by inverse filtering. Journal of the Acoustical Society of America, 98, 763-767.

Alku, P. & Vilkman, E. (1996). A comparison of

glottal voice source quantification parameters in breathy, normal and pressed phonation of female and male speakers. Folia Phoniatrica et Logopaedica, 48, 240-254.

Bäckström, T., Alku, P. & Vilkman, E. (2002). Time-domain parameterization of the closing phase of glottal airflow waveform from voices over a large intensity range. IEEE Transactions on Speech and Audio Processing, 10, 186-192.

Campbell, N. & Mokhtari, P. (2003). Voice quality: the 4th prosodic dimension. Teoksessa Proceedings of the 15th International Congress of Phonetic Sciences, 2417-2420.

Campbell, N. (2003) Towards synthesising expressive speech; Designing and collecting expressive speech data. Teoksessa Proceedings of the European Speech Processing Conference, 1637-1640.

Carlson, R., Granström, B. & Karlsson, I. (1991). Experiments with voice modelling in speech synthesis. Speech Communication, 10, 481-489.

Childers, D.G. (1995). Glottal source modeling for voice conversion. Speech Communication, 16, 127-138.

Childers, D.G. & Ahn, C. (1995). Modeling the glottal volume-velocity waveform for three voice types. Journal of the Acoustical Society of America, 97, 505-519.

Childers, D.G. & Hu, H.T. (1994). Speech synthesis by glottal excited linear prediction. Journal of the Acoustical Society of America, 96, 2026-2036.

Childers, D.G. & Lee, C.K. (1988). Vocal quality factors: Analysis, synthesis, and perception. Journal of the Acoustical Society of America, 90, 2394-2410.

Cummings, K. & Clements, M.A. (1995). Analysis of the glottal excitation of emotionally styled and stressed speech. Journal of the Acoustical Society of America, 98, 88-98.

Dromey, C., Stathopoulos, E.T. & Sapienza, C.M. (1992). Glottal airflow and electroglottographic measures of vocal function at multiple intensities. Journal of Voice, 6, 44-54.

Fant, G. (1960). Acoustic Theory of Speech Production. The Hague: Mouton.

Fant, G. (1997). The voice source in connected speech. Speech Communication, 22, 125-139.

Fant, G., Liljencrants, J. & Lin, Q. (1985). A four-parameter model of glottal flow. Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Sweden, 4, 1-13.

Flanagan, J.L. (1972). Speech Analysis, Synthesis, and Perception, New York: Springer.

Fritzell, B., Hammarberg, B., Gauffin, J., Karlsson, I. & Sundberg, J. (1986). Breathiness and insufficient vocal fold closure. Journal of Phonetics, 14, 549-553.

Fröhlich, M., Michaelis, D. & Strube, H.W. (2001). SIM-Simultaneous inverse filtering and matching of a glottal flow model for acoustic speech signals. Journal of the Acoustical Society of America, 110, 479-488.

Gauffin, J. & Sundberg, J. (1989). Spectral correlates of glottal voice source waveform characteristics. Journal of Speech and Hearing Research, 32, 556-565.

Gobl, C. & Ni Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. Speech Communication, 40, 189-212.

Hertegård, S. (1994). Vocal fold vibration as studied with flow inverse filtering. Academic Dissertation, Dept. of Logopedics and Phoniatrics, Karolinska Institutet, Huddinge University Hospital, Sweden.

Hertegård, S. & Gauffin, J. (1992). Acoustic properties of the Rothenberg mask. Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Sweden, 2-3, 9-18.

Hertegård, S. & Gauffin, J. (1995). Glottal area and vibratory patterns studied with simultaneous stroboscopy, flow glottography and electroglottography. Journal of Speech and Hearing Research, 38, 85-100.

Hillman, R.E., Holmberg, E.B., Perkell, J.S., Walsh, M. & Vaughan, C. (1989). Objective assessment of vocal hyperfunction: An experimental framework and initial results. Journal of Speech and Hearing Research, 32, 373-392.

Hillman, R.E., Holmberg, E., Perkell, J.S., Walsh, M. & Vaughan, C. (1990). Phonatory function associated with hyperfunctionally related vocal fold lesions. Journal of Voice, 4, 52-63.

Holmberg, E.B., Hillman, R.E. & Perkell, J.S. (1988). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. Jour-

nal of the Acoustical Society of America, 84, 511-529.

Holmberg, E., Hillman, R.E. & Perkell, J.S. (1989). Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch. Journal of Voice, 3, 294-305.

Howell, P. & Williams, M. (1988). The contribution of the excitatory source to the perception of neutral vowels in stuttered speech. Journal of the Acoustical Society of America, 84, 80-89.

Howell, P. & Williams, M. (1992). Acoustic analysis and perception of vowels in children's and teenagers' stuttered speech. Journal of the Acoustical Society of America, 91, 1697-1706.

Isshiki, N. (1981). Vocal efficiency index. In K.N. Stevens and M. Hirano (Eds.), Vocal Fold Physiology, Tokio: University of Tokyo Press, 193-203.

Iwarsson, J., Thomasson, M. & Sundberg, J. (1998). Effects of lung volume on glottal voice source. Journal of Voice, 12, 424-433.

Laukkanen, A-M., Vilkman, E., & Alku, P. (1996). Physical variations related to stress and emotional state: A preliminary study. Journal of Phonetics, 24, 313-335.

Laukkanen, A-M., Vilkman, E., & Alku, P. (1997). On the perception of emotions in speech: the role of voice quality. Logopedics Phoniatrics Vocology, 22, 157-168.

Lauri, E-R., Alku, P., Vilkman, E., Sala, E. & Sihvo, M. (1997). Effects of prolonged oral reading on time-based glottal flow waveform parameters with special reference to gender difference. Folia Phoniatrica et Logopaedica, 49, 234-246.

Miller, R.L. (1959). Nature of the vocal cord wave. Journal of the Acoustical Society of America, 31, 667-677.

Monsen, R.B. & Engebretson, A.M. (1977). Study of variations in the male and female glottal wave. Journal of the Acoustical Society of America, 62, 981-993.

Plumpe, M.D., Quatieri, T.F. & Reynolds, D.A. (1999). Modeling of the glottal flow derivative waveform with application to speaker identification. IEEE Transactions on Speech and Audio Processing, 7, 569-586.

Price, P.J. (1989). Male and female voice source characteristics: Inverse filtering results. Speech Communication, 8, 261-277.

Rothenberg, M. (1973). A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. Journal of the Acoustical Society of America, 53, 1632-1645.

Sapienza, C.M., Stathopoulos, E.T. & Dromey, C. (1998). Approximations of open quotient and speed quotient from glottal airflow and EGG waveforms: Effects of measurement criteria and sound pressure level. Journal of Voice, 12, 31-43.

Scherer, R.C., Arehart, K.H., Guo, C.G., Milstein, C.F. & Horii, Y. (1998). Just noticeable differences for glottal flow waveform characteristics. Journal of Voice, 12, 21-30.

Strik, H. & Boves, L. (1992). On the relation between voice source parameters and prosodic features in connected speech. Speech Communication, 11, 167-174.

Sulter, A.R. & Wit, H.P. (1996). Glottal volume velocity waveform characteristics in subjects with and without vocal training, related to gender, sound intensity, fundamental frequency, and age. Journal of the Acoustical Society of America, 100, 3360-3373.

Sundberg, J., Andersson, M. & Hultqvist, C. (1999). Effects of subglottal pressure on professional baritone singers' voice sources. Journal of the Acoustical Society of America, 105, 1965-1971.

Sundberg, J., Cleveland, T.F., Stone, R.E, Jr. & Iwarsson, J. (1999). Voice source characteristics in six premier country singers. Journal of Voice, 13, 168-183.

Sundberg, J., Titze, I. & Scherer, R. (1993). Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. Journal of Voice, 7, 15-29.

Titze, I. & Sundberg, J. (1992). Vocal intensity in speakers and singers. Journal of the Acoustical Society of America, 91, 2936-2946.

Vilkman, E., Lauri, E-R., Alku, P., Sala, E. & Sihvo, M. (1997). Loading changes in time-based parameters of glottal flow waveforms in different ergonomic conditions. Folia Phoniatrica et Logopaedica, 49, 247-263.

Wong, D.Y., Markel, J.D. & Gray, A.H. Jr. (1979). Least squares glottal inverse filtering from acoustic speech waveforms. IEEE Transactions on Acoustics, Speech, and Signal Processing, 27, 350-355.

SPEECH PRODUCTION AND THE PARAMETERISATION OF THE GLOTTAL VOLUME VELOCITY WAVEFORM ESTIMATED BY INVERSE FILTERING

Paavo Alku, Acoustics Laboratory, Helsinki University of Technology, Finland

Estimation of the source of voiced speech, the glottal volume velocity waveform, with inverse filtering involves usually a parameterisation stage, where the obtained flow waveforms are expressed in numerical form. This stage of the voice source analysis, the parameterisation of the glottal flow, is discussed in the present paper. The paper aims to give a review of the different methods developed for the parameterisation and it discusses how these parameters have reflected the function of the voice source in various voice production studies.

Keywords: speech production, inverse filtering, glottal excitation, parameterisation