



Aalto-yliopisto

Puheteknologian perusteet

Luento 1 – Puheääni

Kurssilla “Informaatioteknologian perusteet”

Tom Bäckström

8.3.2021

Puhe ja kieli

Johdanto

- Ei ole sattumaa että tämä luento esitetään *puhutussa* muodossa.
 - Miksi?
- Mikä media sopii parhaiten elämäsi tärkeimpiin hetkiin?
 - Kun pyydät rakkaintasi naimisiin (kun kosit)?
 - Työhaastatteluun?
- Puheella on hyvin erityinen rooli ihmisten kanssakäymisessä ja kommunikaatioissa.

⇒ Puhe on meidän *ensisijainen kommunikaatiomoodi*.

Puhe ja kieli

Johdanto

- Toisaalta “*yksi kuva kertoo enemmän kuin tuhat sanaa*”
 - Näköaisti on tehokas työkalu *tiedonsiirtoon* (download).
- Puheääni on parhaimmillaan *vuorovaikutustilanteissa*.



Oppimistavoitteet

Johdanto

- Tutustua puheäänien peruskäsitteisiin
 - Puheentuotto ja lingvistiikka
 - Akustisen puhesignaalin informaatio sisältö
- Tuntustutaan puheteknologian osa-alueisiin ja perusteisiin
 - Puheen siirto ja koodaus
 - Puheen ehostus (kohinanpoisto, lähde-erotus)
 - Puheen tunnistus ja synteesi
- Tällä kurssilla tarjotaan yleiskatsaus
 - Johdanto jossa ei mennä syvemmälle teoriaan

Sisältö

Johdanto

Puheääni

Puhetyylit

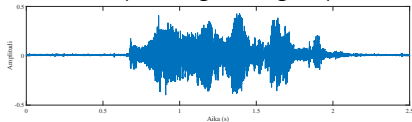
Puheentuotto ja kieli



Paineaalto

Puheääni

- Ääni on ilmassa kulkeva paineaalto.
- Puheäänien taajuusalue noin 80 Hz – 20 kHz.
- Ääntä nauhoitetaan mikrofoniin, joka muuntaa paineaallon sähköiseksi signaaliksi.
- Nykyään tämä signaali muunnetaan yleensä vielä digitaaliseen muotoon AD-muuntimen (analog-to-digital) avulla.



Soita ääniesimerkki

Näytteenotto

Puheääni

- Jos oletetaan että
 - näytteenottotaajuus on 16 kHz:iä
(=16000 mittausta paineaallosta per sekunti) ja
 - jokainen näyte tallennetaan 16 bittisinä numeroina, niin
- Puheäänien tallentamiseen tarvitaan noin 256 kbit/s.
 - ⇒ Tunnin tallenteelle tarvitaan siis lähes 100 MB tallennustilaa.
- Pakkaamattoman äänen tallennus vie kohtuuttomasti tilaa, vai viekö?
 - Kuinka paljon informaatiota puheäänessä onkaan?
 - Minkä tyyppistä informaatioita puheäänessä on?

Informaation tyypit

Puheääni

- Lingvistinen informaatio (puhuttu teksti)
 - Foneemit (äänteet)
 - Tavut ja sanat
 - Lauseet jne.
 - Kieli, murre ja aksentti
- Paralingvistinen (muu kuin lingvistinen) informaatio
 - Tila: Henkilöllisyys, tunnetila, terveydentila, ikä, sukupuoli
 - Kommunikaatiotyylit ja oheisviestintä: Puhetyyli, sosiaalinen konteksti ja -hierarkia, painotus
 - Fysiologiset parametrit: Kielen, huulten, lihasten yms. asento sekä niiden koko ja jännitys, ilmavirtauksen määrä
 - Akustiset kuvaukset: Äänenkorkeus, kireys, huokoisuus, puhenopeus
- Ei ole itsestäänselvää mitä informaatiota puheäänessä on.



Informaation määrä

Puheääni

- Foneemit/äänteet
 - Kielessä on tyypillisesti noin 25–32 kirjainta eli yhdessä äänneessä on noin 5 bittiä informaatiota.
 - Puhenopeus on tyypillisesti välillä 5–15 äännettä per sekunti.
 - Lingvistinen informaatiotiheys on siksi enimmillään noin $5 \times 15 = 75$ bit/s.
 - Paralingvistisen informaation määrää on vaikea määrittää.
 - Valistunut arvaus on että sitä ei ole enempää kuin lingvististä informaatiota, eli < 75 bit/s.
 - Yhteensä informaatiotiheys on siis kokoluokka 150 bit/s.
 - Digitaalisen puhesignaalin informaatiotiheys oli 256 kbit/s eli noin 2000 kertaa suurempi!
 - Puheen käsittelyn keskeisiä tavoitteita on kaivaa oleellinen informaatio (150 bit/s) esiin informaatiotulvasta (128 kbit/s).
-

Puhetyylit

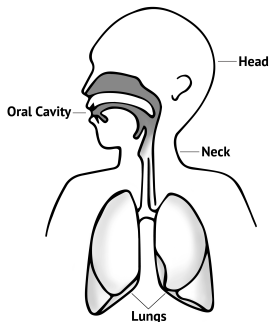
Mitä eri puhetyylejä voit keksiä? Mitä eri puhetyyliä kategorioita on olemassa?

- Tunnetila; innostunut, vihainen, pettynyt, lempeä, pelokas ...
- Terveystila; terve, nuha, väsynyt, humalainen, patologiat kuten syöpä äänihuulissa ...
- Ilmaisun keinoja; tavallinen puheääni, julkinen esiintyminen, saarna, monologi, kuiskaus, imitointi ...

Ajatuksesta ilmavirtaan

Puheentuotto ja kieli

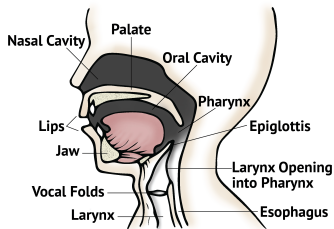
- Puhe alkaa aivoissa, jossa muodostetaan haluttu viesti, luodaan lauserakenne, sekä lähetetään lihaksille tarpeelliset ohjausimpulssit.
- Mekaaninen puheentuotto alkaa keuhkoista, jotka supistuvat siten että ilmanpaine niiden sisällä nousee ja ilma alkaa virrata ulospäin.



Äänilähde

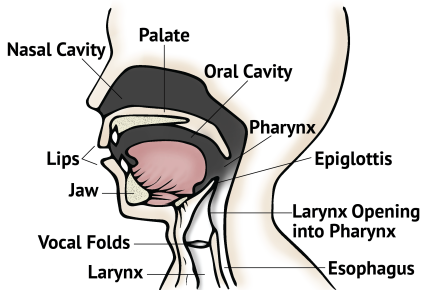
Puheentuotto ja kieli

- Puheäänen muodostavat kaksi mekanismia:
 - *Soinnilliset äänneet*: Äänihuulia sopivasti jännittämällä ne alkavat värähdellä ilmavirrassa tuottaen jaksollisen aaltomuodon.
 - *Soinnittomat äänneet*: Kun huulia vie lähelle toisiaan, kieltä lähelle hampaita tai kitalakea, tai jos ääniväylään aiheuttaa muun ahtauman, syntyy ahtaumassa kohinaa (turbulenssia).



Fysiologian sanastoa

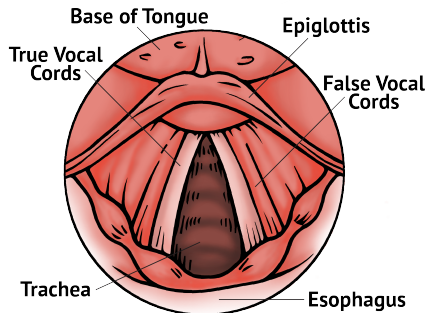
Puheentuotto ja kieli



English	Suomi
nasal cavity	nenäväylä
palate	kitalaki
oral cavity	suuväylä
lips	huulet
tongue	kieli
jaw	leuka
pharynx	nielu
epiglottis	kurkunkansi
glottis	äänirako
vocal folds	äänihuulet
larynx	kurkunpää
esophagus	ruokatorvi
vocal tract	ääniväylä

Soinnilliset äänteet

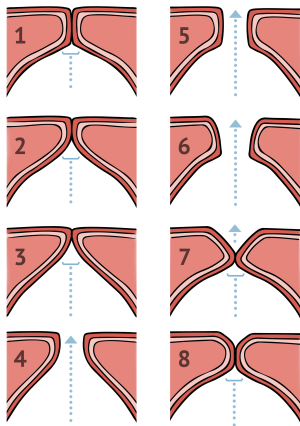
Puheentuotto ja kieli



English	Suomi
vocal fold	äänihuuli
trachea	henkitorvi
epiglottis	kurkunkansi

Äänihuulet ylhäältäpäin kuvattuna.

Äänihuulten toiminta



https://youtu.be/mJedwz_r2Pc?t=17

<https://www.youtube.com/watch?v=W-nS9fgs7Ro>

Ääniväylän vaikutus

Puheentuotto ja kieli

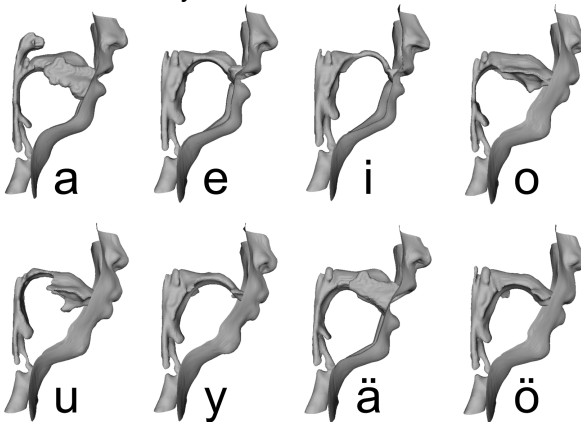
Ääniväyläksi kutsutaan “kurkkua”, eli sitä putkea joka alkaa äänihuulilta ja päättyy huuliin.

- Erillaisia äänteitä toutetaan muuttamalla ääniväylän muotoa.
 - Tarkkaile miten suusi asento muuttuu kun sanot ääneen aeaeaeaeaea, eieieieie, äyäyäyää.
- Ääniväylällä on resonansseja, jotka korostavat tiettyjä taajuuksia äänessä.
 - Resonanssit muuttuvat kun ääniväylän muotoa muuttaa.
 - Vokaalit erotetaan toisistaan resonanssien (=formanttien) taajuuksien perusteella

Ääniväylän vaikutus

Puheentuotto ja kieli

Ääniväylä suomen eri vokaaleilla

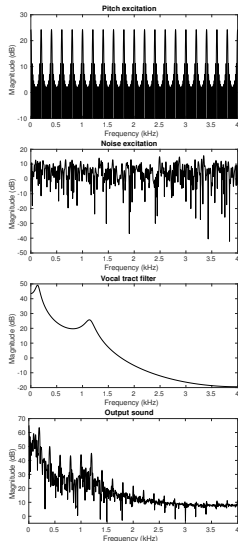
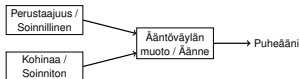


©Jarmo Malinen, Aalto-yliopisto

Äänilähteet ja ääniväylä

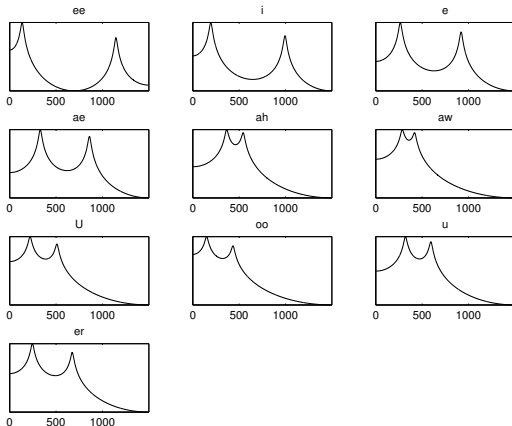
Puheentuotto ja kieli

- Äänihuulet tuottavat harmonisen (soinnillisen) signaalin.
- Soinnittomat äänteet ovat luonteeltaan kohinaa.
- Ääniväylän resonanssit (formantit) muovaavat spektriä.
- Soinnilliset ja soinnittomat äänteet summautuvat ja syötetään ääniväylään (esimerkki).



Ääniväylän formantit eri vokaaleilla

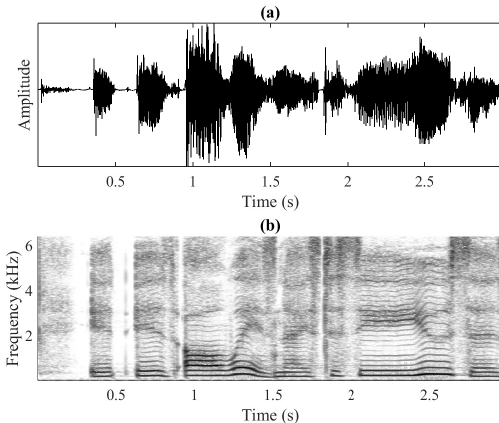
Englannin vokaalien resonanssit taajuusakselilla



Spektrogrammi

Puheentuotto ja kieli

Signaalin spektri ajan funktiona, eli spektrogrammi



Foneemi

Puheentuotto ja kieli

- Kirjoitetun kielen perusyksikkö: kirjain
- Puhutun kielen perusyksikkö: *foneemi*
 - Pienin äänen osanen jolla on merkitys
 - Kääntäen, muutos foneemissa (voi) aiheuttaa muutoksen sanan/lauseen merkityksessä
 - Kissa/kassa, Timo/Timi, pieni/sieni, kehu/rehu
- Puhutun kielen foneemit on kuvattu standardissa
 - International Phonetic Alphabet (IPA) – “äänneaakkoset”

IPA Esimerkkejä – Vokaaleja

Puheentuotto ja kieli

IPA	Esimerkki	IPA	Esimerkki
i	city, see, meat	ʏ	German: müssen
y	German: über, Rübe	o	German: Ofen, Roman
ɨ	rose's	ɛ	bed
ʉ	rude	œ	German: Hölle, göttlich
ʋ	Irish: caol	ɜ	bird
u	ough, you, threw	ɞ	Irish English: but
ɪ	sit	ʌ	run, won, flood
Y	German: füllt	ɔ	law, caught, all
ʊ	put, hood	æ	cat, bad
e	German: Genom, Methan	ɐ	German: oder
ø	French: peu	a	hat
ə	about, arena	œ	Swedish: hört
ɘ	Dutch: ik	ɑ	father
ɚ	Australian English: bird	ɒ	not, long, talk

IPA Esimerkkejä – Konsonantteja

Puheentuotto ja kieli

IPA	Esimerkki	IPA	Esimerkki
b	buy, cab	θ	thigh, math
d	dye, cad, do	p	pie, spy, cap
ð	thy, breathe, father	r	rye, try, very (trill)
dʒ	giant, badge, jam	ɹ	rye, try, very (approximant)
f	phi, caff, fan	s	sigh, mass
g	guy, bag	ʃ	shy, cash, emotion
h	high, ahead	t	tie, sty, cat, atom
j	yes, yacht	tʃ	China, catch
k	sky, crack	v	vie, have
l	lie, sly, gal	w	wye, swine
m	my, smile, cam	z	zoo, has
n	nigh, snide, can	ʒ	equation, pleasure, vision, beige
ŋ	sang, sink, singer		

Yksityisyys puhekommunikaatiossa

Puheentuotto ja kieli

- Kuiskauksella on erityinen merkitys.
 - Kuiskauksella kerrotaan *salaisuuksia* rajatulle kuulijakunnalle.
 - Kääntäen, kuiskattu viesti on tärkeämpää tietoa kuin tavallinen puhe.
- Ihmisillä on yleensä hyvin tarkka käsitys yksityisyyden tasosta;
 - Kun kuiskaan, antaa se heti viestille tietyn merkityksen.
 - Kun haluat kertoa mehukkaan salaisuuden kahvilassa ystävällesi, olet automaattisesti varovainen.
 - Kun kuvaillet eilisen bileitä, on tarina aivan erilainen riippuen siitä onko kuulija läheisin ystäväsi vai äitisi.

Yhteenveto

- Puheääni on ihmisen tärkein kommunikaatio- ja vuorovaikutusmuoto.
 - Puheääni on paineaalto joka
 - muodostuu äänihuulten oskilaatioista (soinnilliset äänteet) sekä
 - turbulensseista ääniväylässä (soinnittomat äänteet) ja jota
 - muokataan muuttamalla ääniväylän muotoa eli sen resonansseja=formantteja.
 - Puheääni on “kantoaalto”, joka välittää lingvististä ja paralingvististä informaatiota.
 - Foneemi on puheäänen perusyksikkö joka kuvaa pienintä elementtiä joka kantaa jotain merkitystä.
 - Foneemit on kuvattu standardoidussa aakkosissa, jotka tunnetaan nimellä International Phonetic Alphabet (IPA).
-