

# Puheteknologian perusteet

## Luento 1 – Puheääni

Kurssilla “Informaatioteknologian perusteet”

Tom Bäckström

26.2.2018

# Puhe ja kieli

## Johdanto

- ▶ Ei ole lainkaan sattumaa että tämä luento esitetään *puhutussa* muodossa.
- ▶ 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*.

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

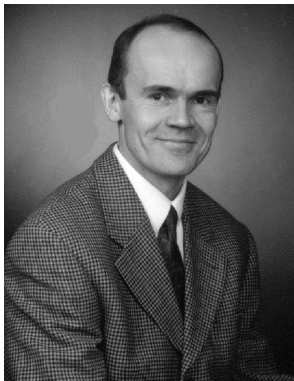


- ▶ Professor of practice Tom Bäckström
- ▶ Väitellyt Otaniemessä 2004, töissä Saksassa 2008-2016 (Fraunhofer IIS ja Friedrich-Alexander University), takaisin Otaniemessä 2016-
- ▶ Tutkimusaiheet
  - ▶ Puheensiirto (puhelimesta puhelimeen ja puhelimesta pilveen)
  - ▶ Yksityisyys puhepalveluissa
  - ▶ Puheäänien matemaattinen mallinnus
  - ▶ Ehostus (kohinanpoisto, monimikrofonimenetelmät), signaalinkäsittely, lineaarialgebra jne.
- ▶ Opetus: ELEC-E550 Puheen käsittely (5 ECTS)



# Puheteknologian muut proffat

## Johdanto



Akatemiaprofessori Paavo Alku  
analyysi, koodaus, synteesi



Professori Mikko Kurimo  
tunnistus, synteesi

# Orientaatio

## Johdanto

Paritehtävä:

- ▶ Mitä puheteknologian sovelluksia olette käyttäneet? (3min)

# Oppimistavoitteet

## Johdanto

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

# Sisältö

Johdanto

Puheääni

Puheentuotto ja kieli

Puhetyylit

Puhetyylit

Puheteknologian sovelluksia

- Puheenkoodaus

- Puheensiistaus/-ehostus

- Puhekäyttöliittymiä

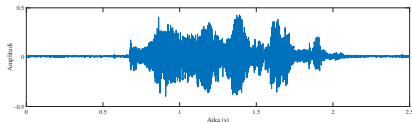
- Puheentunnistus

- Puhesynteesi

# Paineaalto

## Puheääni

- ▶ Ääni on ilmassa kulkeva paineaalto.
- ▶ Puheäänen taajuusalue noin 80 Hz – 20 kHz.
- ▶ Ääntä nauhoitetaan mikrofoniolla, 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 8 kHz:iä  
(=8000 mittausta paineaallosta per sekunti) ja
  - ▶ jokainen näyte tallennetaan 16 bittisinä numeroina, niin
- ▶ Puheäänen tallentamiseen tarvitaan noin 128 kbit/s.
  - ⇒ Tunnin tallenteelle tarvitaan lähes 500 megatavua!
- ▶ Pakkaamattoman äänen tallennus vie kohtuuttomasti tilaa, vai viekö?
  - ▶ Kuinka paljon informaatiota puheäänessä onkaan?
  - ▶ Mitä tyyppistä informaatioita puheäänessä on?

# Informaation tyypit

## Puheääni

- ▶ Lingvistinen informaatio (puhuttu teksti/kieli)
  - ▶ Foneemit (äänteet)
  - ▶ Tavut ja sanat
  - ▶ Lauseet jne.
  - ▶ Kieli, murre ja aksentti
- ▶ Paralingvistinen (muu kuin lingvistinen) informaatio
  - ▶ Tila: Henkilöllisyys, tunnetila, terveydentila, ikä, sukupuoli
  - ▶ Kommunikaatiotyyli 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 äänteessä 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 128 kbit/s eli noin 1000 kertaa suurempi!
- ▶ Puheen käsittelyn keskeisiä ongelmia/tavoitteita on kaivaa se oleellinen ja haluttu informaatio (150 bit/s) esiin informaatiotulvasta (128 kbit/s).

# Harjoitus

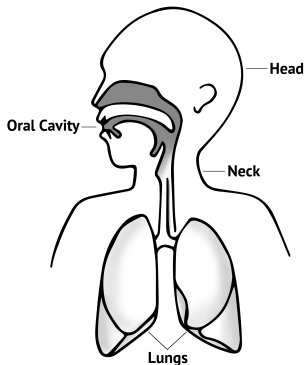
## Puheääni

1. Kuunnelkaa seuraavia ääniesimerkkejä ja analysokaa niiden lingvistinen ja erityisesti paralingvistinen sisältö. Kuka puhuu? Ikä, sukupuoli, koko? Tunnetila? Puhetyyli? Sosiaalinen konteksti?
2. Äänittäkää omaa ääntänne (esim. puhelimen videokameralla):
  - 2.1 Vertaa ääntäsi ääniesimerkkeihin. Kuuletko paralingvistisen sisällön erot?
  - 2.2 Kuulostaako oma äänesi siltä kuin uskoit sen kuulostavan?

# Ajatuksesta ilmapirtaan

## 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.



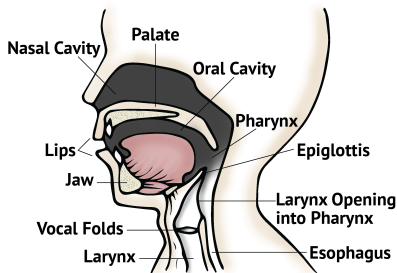
# Äänilähde

## Puheentuotto ja kieli

- Puheäänen muodostavat kaksi mekanismia:

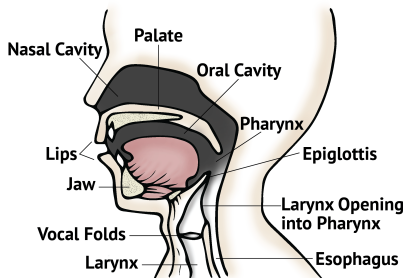
**Soinnilliset äänteet** Äänihuulia sopivasti jännittämällä ne alkavat värähdellä ilmapirrassa tuottaen jaksollisen aaltomuodon.

**Soinnittomat äänteet** Kun huulia vie lähelle toisiaan, kieltä lähelle hampaita tai kitalakea, tai jos ääniväylään aiheuttaa muun ahtauman, syntyy ahtaumassa kohinaa (turbulenssia).



# Fysiologian sanasto

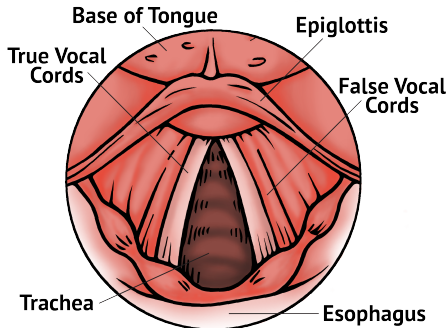
## Puheentuotto ja kieli



| 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

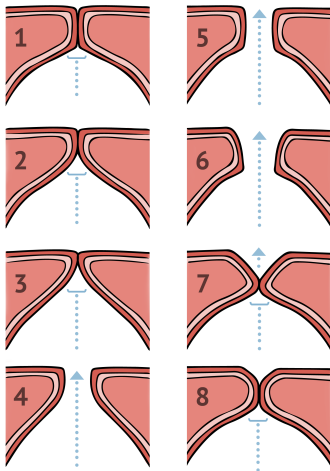


Äänihuulet ylhäältäpäin kuvattuna.

| English    | Suomi       |
|------------|-------------|
| vocal fold | äänihuuli   |
| trachea    | henkitorvi  |
| epiglottis | kurkunkansi |

# Äänihuulten toiminta

Puheentuotto ja kieli



[https://youtu.be/mJedwz\\_r2Pc?t=17](https://youtu.be/mJedwz_r2Pc?t=17)

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

# Ääniväylän vaikutus

## Puheentuotto ja kieli

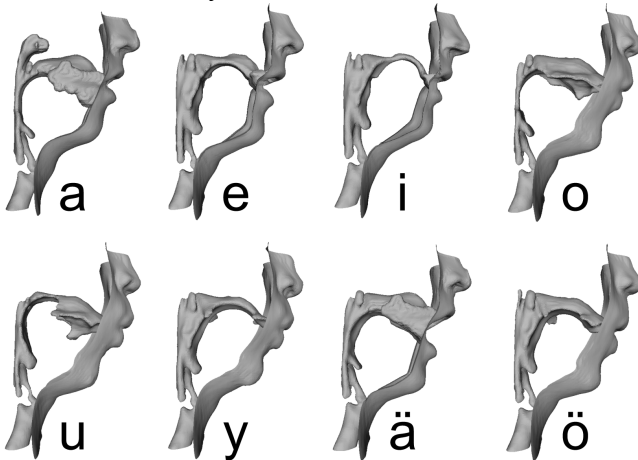
Ääniväyläksi kutsutaan “kurkkua”, eli sitä putkea joka alkaa äänihuulilta ja päättyy huuliin.

- ▶ Erillaisia äänteitä voidaan tuottaa muuttamalla ääniväylän muotoa.
  - ▶ Kokeile sanoa 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

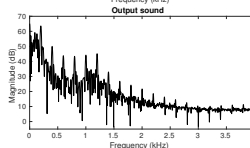
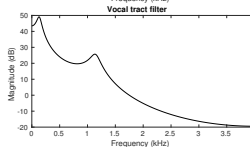
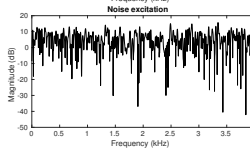
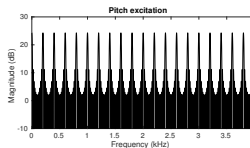
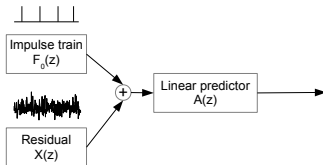


(C) Jarmo Malinen, Aalto-yliopisto

# 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).



# Lineaarinen ennustaminen

## Puheentuotto ja kieli

- ▶ Usein käytetty ääniväylä-malli on lineaarinen ennustin.
  - ▶ Jos puhesignaali on  $s_n$ , voidaan muodostaa ennustin

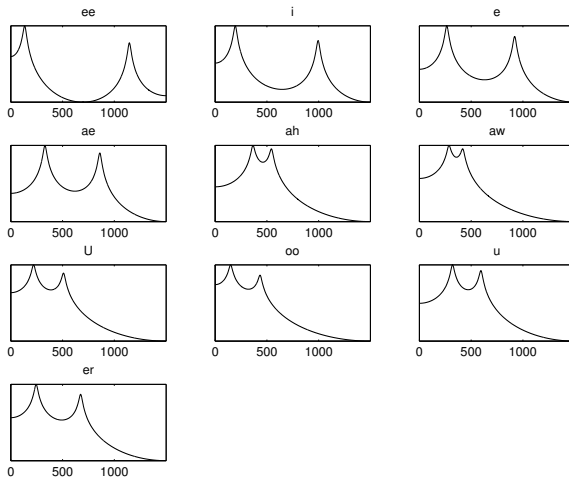
$$\hat{s}_n = - \sum_{k=1}^m a_k s_{n-k}.$$

- ▶ Arvaamme arvon  $\hat{s}_n$  painotettuna summana aiemmista näytteistä  $s_{n-k}$ .
  - ▶ Skalaarit  $a_k$  ovat painokertoimia. Eri äänteillä käytetään eri painokertoimia.
- ▶ Lineaarinen ennustin tunnetaan myös nimillä
  - ▶ Infinite impulse response (IIR) -suodin.
  - ▶ Autoregressive (AR) -suodin.
- ▶ Suotimen Z-muunnoksessa  $A^{-1}(z)$ , eli sen spektrissä, näkyvät ääniväylän resonanssit huippuina.

# Ääniväylän formantit/resonanssit eri vokaaleilla

Puheentuotto ja kieli

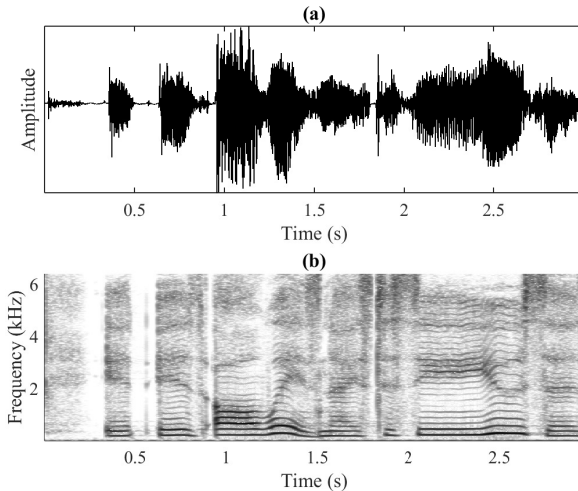
## Englannin vokaalien resonanssit taajuusakselilla



# 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 (aakkoset) standardissa
  - ▶ International Phonetic Alphabet (IPA)

# IPA Esimerkkejä – Vokaaleja

Puheentuotto ja kieli

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

# IPA Esimerkkejä – Konsonantteja

Puheentuotto ja kieli

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

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 ...

# Yksityisyys puhekommunikaatiossa

## Puhetyylit

- ▶ 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 puheäänien perusyksikkö joka kuvaa pienintä elementtiä joka kantaa jotain merkitystä.
- ▶ Foneemit on kuvattu standardoidussa aakkosissa, jotka tunnetaan nimellä International Phonetic Alphabet (IPA).

Lisää ensi kerralla!