



HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

# Ilmausta pitempien kokonaisuuksien tuottaminen puhesyntetisaattorilla

Antti Suni, Martti Vainio

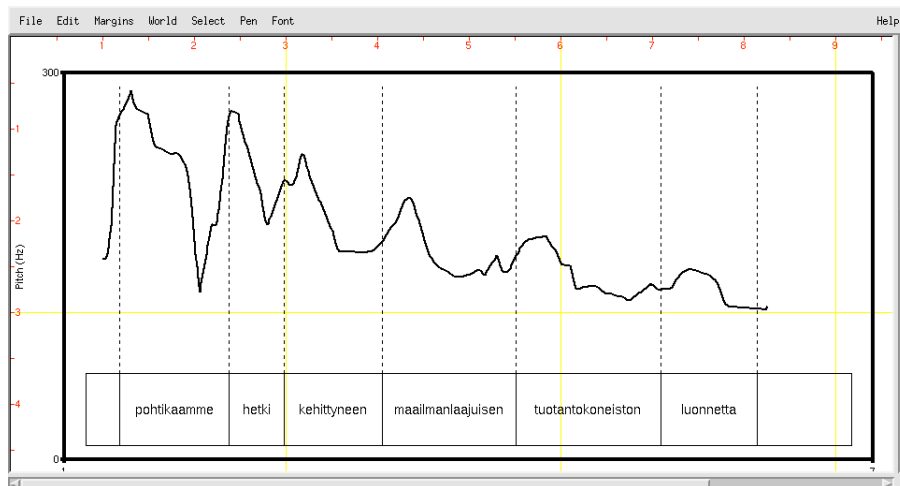
Helsingin yliopiston puhtetieteiden laitos

Fontetiikan Päivät 2006



## Taustaa – suomen lauseintonaatio

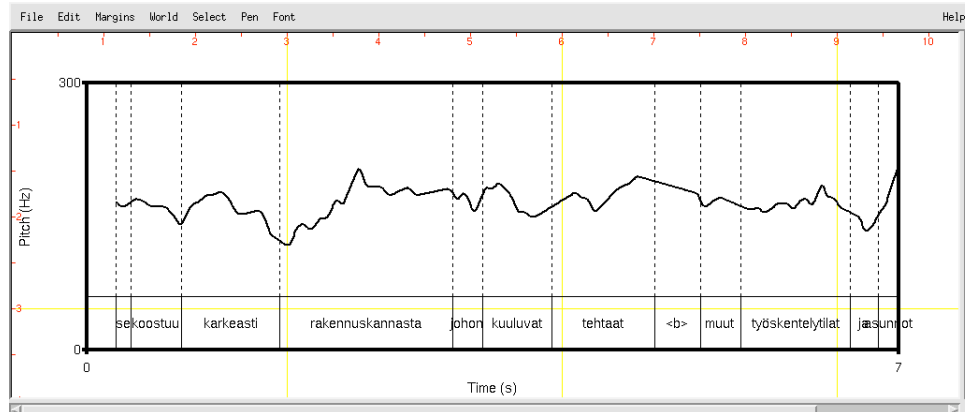
- “Normaalipainotteisen väitelauseen intonaatio koostuu hitaasti laskevien painollisten tavujen jonosta...”





## Suomen lauseintonaatio?

- Kaikki lauseet eivät ole normaalipainotteisia väitelauseita



## HMM-synteesi

- Markovin piilomalleihin perustuva järjestelmä, joka opetetaan puheentunnistimien tapaan.
- Prosodian mallintamiseen käytetään kymmeniä kontekstuaalisia piirteitä
  - Mallien parametrit estimoidaan useiden esiintymien perusteella, joten pienistä epätarkkuuksista opetusdatassa ei ole haittaa
  - viereiset äänteet, kvantiteetti, äänteen positio tavussa, tavun positio sanassa, tavujen määrä sanassa, lauseen positio virkkeessä, paino, aksentti..
  - piirreavaruus valtava – mallit klusteroidaan päätöspuilla.
  - Spektri, perustaajuus ja kesto klusteroidaan toisistaan riippumattomasti – opetusdatan tarjoamat mahdollisuudet tulee hyödynnettyä kattavasti
- Puhe generoidaan suoraan opetetuista kepstri-, F0- ja kestopalleista
  - Mallien deltoja käytetään rajakohtien tasoitukseen – ei havaittavia epäjatkuvuuskohtia
  - Puhe tuotetaan MLSA-vokooderilla



## HMM-synteesi – kontekstuaaliset piirteet

- Trifoni, vasen ja oikea äänne
- Trifoni + tavun positio lauseessa
- Trifoni+ tavu lauseessa + tavun sanassa
- Nykyinen järjestelmä, n. 60 piirrettä



## Opetusmateriaali

- Konkatenatiivisessa synteesissä: mahdollisimman tasaista ja neutraalia, jotta liitoskohdat saadaan sujuviksi, tyypillisesti lyhyitä yksittäisiä lauseita
- HMM-synteesissä: prosodisesti rikkaan, luonnollisen puheen käyttö mahdollista
- Käytettävissä oleva materiaali ei sovellu kappaleiden lukemiseen
- Äänikirjat
  - Runsaasti olemassa olevaa digitaalista materiaalia
  - Teksti valmiina
  - Hyvät puhujat



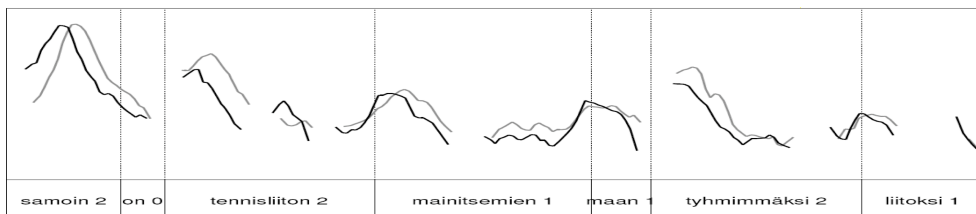
## Opetusdatan prosodian annotointi

- Tavoitteena automaattinen opetusprosessi
- Tauot
  - automaattinen taukojen vahvuuden (0-3) annotointi alkuperäisen puhujan ja yksinkertaisen referenssisyntetisaattorin tuottamien taukojen kestoja ja loppuhidastuksia vertailemalla
- Topiikki
  - automaattisesti vertailemalla puhunnosten alkujen perustaajuutta referenssisyntetisaattorin ja puhujan välillä, (ei siis tekstistä)
- Sanojen prominenssi, havaittu "painokkuus" (0-3)



## Prominenssin annotointi

- Kuulonvaraisesti osa lauseista (n.80% yhteisymmärrys)
- Regressiomalli
  - Mallia varten laskettiin useita intensiteettiä, keston ja perustaajuuteen liittyviä piirteitä, perustaajuuteen liittyvät piirteet osoittautuivat merkittävimmiksi
  - $R^2 = 79.4$
- Prominenssi- ja piirteillä opetettu synteesi
  - Vanhalla materiaalilla perustaajuuden RMSE alle 2 st. Kertoo kuitenkin lähinnä opetusmateriaalin säännöllisyydestä
  - Äänikirjasynteesi: 103 122 131 301





## Konseptista puheeksi

### ■ Puhuttava teksti

generoidaan automaattisesti

Esim. "Minne täältä sairaalalta oikein pääsee??"

### ■ Luotettava tieto

informaatorakenteesta;  
aksentuaation ennustaminen  
helppoa?

### ■ Vähän tarvetta, tai "liian yksinkertaisia" sovelluksia (limited domain -synteesi)

```

<message>
....
<utterance>

<phrase>
<token type="word"><word ort="Tänään" pos="ADV"/></token>
<token type="word"><word ort="seuraavat" pos="A" ppos="mod"/></token>
<token type="word"><word ort="vuorot" pos="N" ppos="head"/></token>
<token type="word"><word ort="pysäkillä" given="1" pos="N"/></token>
<token type="word"><word ort="Hatanpään" pos="N" given="1" ppos="mod"/></token>
<token type="word"><word ort="sairaala" pos="N" given="1" ppos="head"/></token>
<token type="word"><word ort="ovat" pos="V"/></token>
</phrase>
<phrase>
<token type="word"><word ort="Kaupin" pos="N" ppos="mod" focus="1"/></token>
<token type="word"><word ort="sairaalalle" pos="N" given="1" ppos="head"/></token>
<token type="word"><word ort="linja" pos="N"/></token>
<token type="word"><word ort="21" pos="NUM"/></token>
<token type="word"><word ort="neljän" pos="NUM" ppos="mod" focus="1"/></token>
<token type="word"><word ort="minuutin" pos="N" ppos="head"/></token>
<token type="word"><word ort="kuluttua" pos="V"/></token>
</phrase>
...

```

Tänään	1	Kaupin	3	kuluttua	0
seuraavat	2	Sairaalalle	0		
vuorot	1	linja	0		
pysäkillä	0	kaksikymmentä	1		
Hatanpään	1	yksi	2		
sairaala	1	neljän	3		
ovat	1	minuutin	1		



## Tekstistä puheeksi – prominenssin ennustaminen

### ■ Mahdotonta?

- Painotuksien valintaan vaikuttavat mm. syntaksi, semantiikka, diskurssi- ja informaatorakenne, rytmi ja puhujan mieliala – kaikki automaattisen analyysin suhteen hataralla pohjalla
- Huoliteltu lukupuhe ehkä säännöllisempää
- Vaihtoehtoisia painotuksia paljon, virheet ennustuksessa eivät välttämättä kuulosta virheiltä

### ■ Käytössä olevat vihjeet

- Syntaktinen analyysi(FDG)
  - Sanaluokka, perusmuoto, NP:t, morfologinen analyysi, yhdyssanat
  - Tarvetta muillakin osa-alueilla, esim. rajageminaatio ja sivupainot
- Informaatorakenne, uudet ja mainitut sanat
- Semanttinen paino, unigrammit
- Sanan positio fraasissa, virkkeessä ja kappaleessa



## Prominenssin ennustaminen säännöillä

```

foreach w in phrase:

    if PoS(w) in [N, ADJ, ACC_ADV, NUM]: p = 2
    elif PoS(w) in [V, PRN, SOME_ADV]: p = 1
    else: p = 0

    if given(w): dec(p)
    if modifier(w):
        if first_new_in_NP(w) and given(head(w)):
            inc(p)
        elif PoS(w) = PRN and rest_NP_given(w):
            inc(p)

    if head(w) and generic(w): dec(p)
    if focus_particle(w): inc(p)
    if emph_next(w): inc_p(next_content(p))

    if starts_sent(w):
        if starts_paragraph(w): inc(p)
        if question(w): inc(p)

    p = many_other_rules(w, p)

    if p > 2: decr_p(next(w))
    w(p) = p

foreach w:
    if last_verb_in_focus(w): inc(w(p))

rhythmical_adjustments()

```

Lapin pelastuslaitos on lähettänyt *Venäjälle sammutusapua* Kantalahdessa riehuvan *metsäpalon sammutukseen*. Maaherra Hannele Pokka sai virka-apupyynnön eilen Murmanskin maaherralta. Paikalle lähti kahdeksan palomiestä sekä sammutusauto, miehistönkuljetusauto ja pumppuja, kertoo Lapin lääninhallitus. *Suomen* valtio korvaa kaikki sammutusavun kustannukset.

**Suomi on tarjonnut apua myös etelämpänä riehuvien metsäpalojen sammutukseen, mutta Venäjä ei ole katsonut sitä tarvitsevana.**

	Suomi	on	tarjonnut	apua	myös	etelämpänä	riehuvien	metsäpalojen	sammutukseen	mutta	Venäjä	ei	ole	katsonut	sitä	tarvitsevana
function-content	2	0	2	2	2	2	2	2	2	0	2	2	0	2	2	2
PoS	2	0	1	2	0	2	1	2	2	0	2	2	0	1	1	1
given	1	0	1	1	0	2	1	1	1	0	1	2	0	1	1	1
emphasize_next	1	0	1	1	0	3	1	1	1	0	1	2	0	1	1	1
after_emphasis	1	0	1	1	0	3	0	1	1	0	3	0	0	1	1	1



## Loppusanat

- Kelvollisen lausetta pidempiä kokonaisuuksia puhuvan syntetisaattorin vaatimukset:
  - Kohdetyyliä vastaava opetusmateriaali
  - HMM-synteesi
  - Relevanttien prosodisten piirteiden annotointi
  - Riittävän hyvä prosodisten piirteiden ennustaminen