

# HMM-POHJAISEN PUHESYNTTEESIN LAADUN PARANTAMINEN GLOTTISPULSSIKIRJASTON AVULLA

Tuomo Raitio (et al.)  
Signaalinkäsittelyn ja akustiikan laitos  
Aalto-yliopisto

Akustiikkapäivät, 11.5.2011

- I. Taustaa
- II. Ihmisen puheentuotto
- III. Puhesynteesijärjestelmä
- IV. Tuloksia ja ääninäytteitä

- Puhesynteesin tavoite on tuottaa mistä tahansa tekstistä hyvälaatuista ja luonnollista puhetta eri puhetyyleillä ja puhujilla
- Kaksi keskeistä puhesynteesimetodia:

## **Konkatenatiivinen synteesi (eng. unit selection)**

- Perustuu valmiiksi äänitettyjen akustisten yksiköiden peräkkäin liittämiseen
- Tuottaa parhaimmillaan erittäin luonnollisen kuuloista puhetta
- Työläs rakentaa, huono adaptoitavuus eri puhetyyleihin ja puhujiin

## **Tilastollinen synteesi**

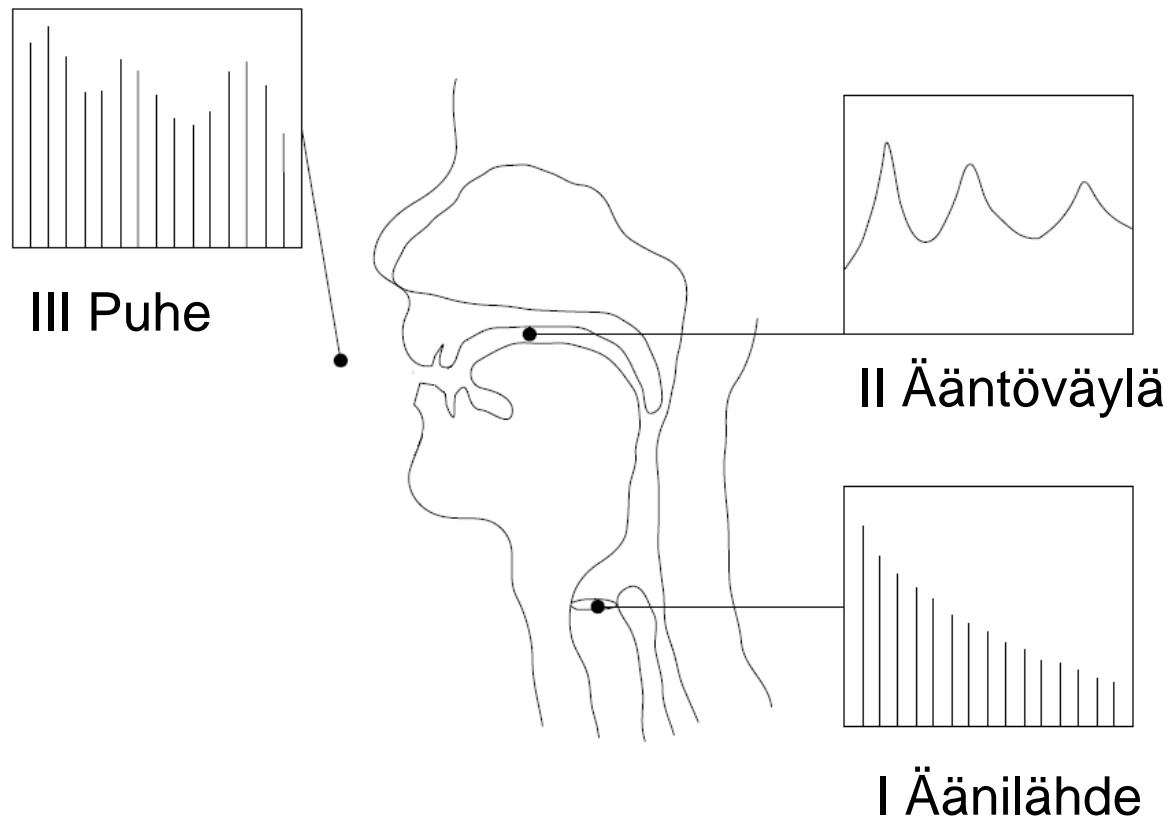
- Perustuu puheen parametrien tilastolliseen mallintamiseen Markovin piilomallien avulla (eng. Hidden Markov Models, HMMs)
- Parempi adaptoitavuus!

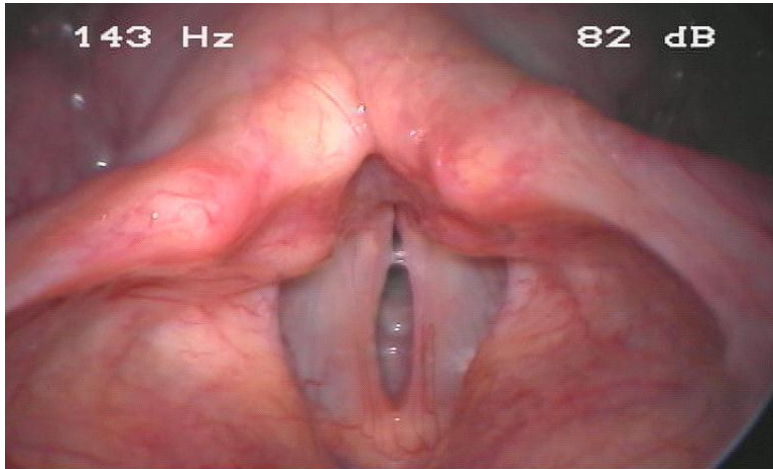
**Ongelma:** Nykyiset HMM-pohjaiset syntetisaattorit kärsivät puheen heikosta laadusta ja luonnollisuudesta

**Ehdotus ratkaisuksi:**

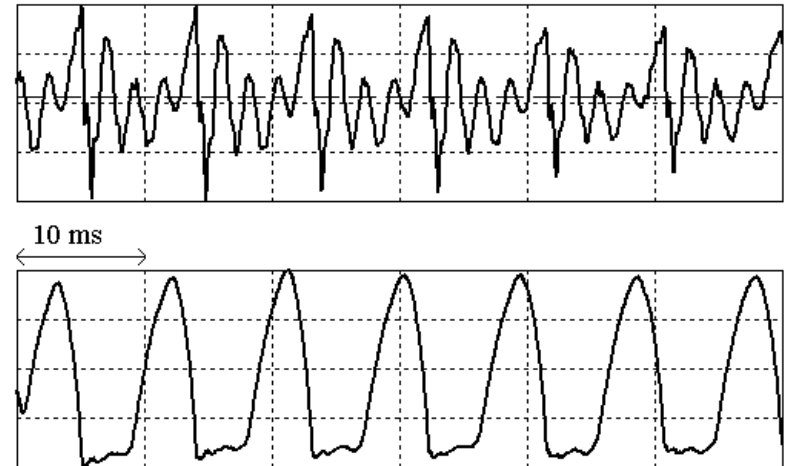
- (1) Puhe hajotetaan ääniväylään ja äänilähteeseen, jotka parametrisoidaan erikseen. Äänilähteestä rakennetaan glottispulssikirjasto.
- (2) Mallinnetaan edellämainitulla tavalla laskettuja puheen parametrejä tilastollisesti (HMM)
- (3) Synteesissä puhe rekonstruoidaan glottispulssikirjastosta valituista pulsseista ja puheen parametreista

## II. Puheentuottomekanismi





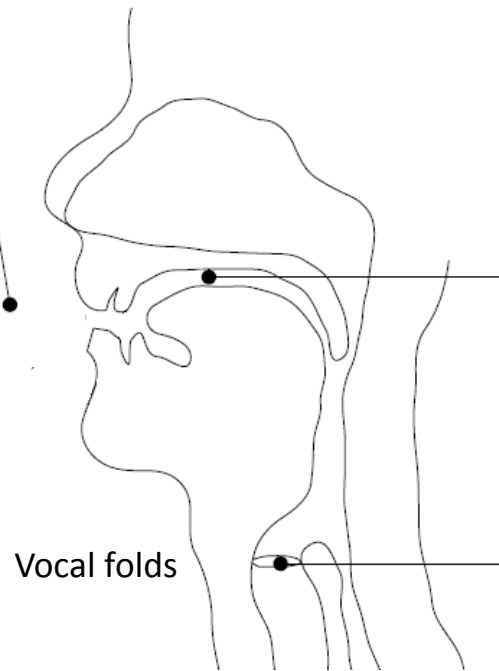
Pysäytyskuva värähtelevistä  
äänihuulista.



Äänipainesignaali (ylempi) ja  
estimoitu glottisheräte (alempi)

Äänilähteen  
käänteissuodatuksen  
avulla voidaan  
puhesignaalista estimoida  
glottisheräte ja  
ääntöväylä

Puhe



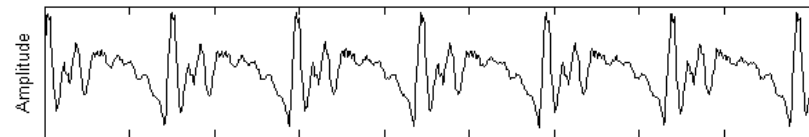
Ääntöväylä

Vocal folds

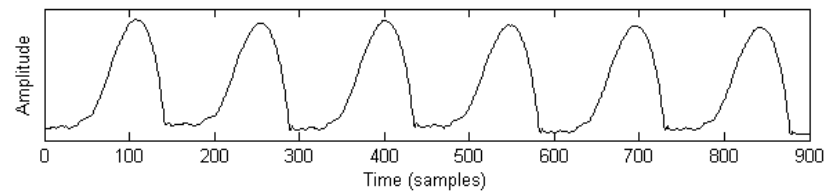


Glottisheräte

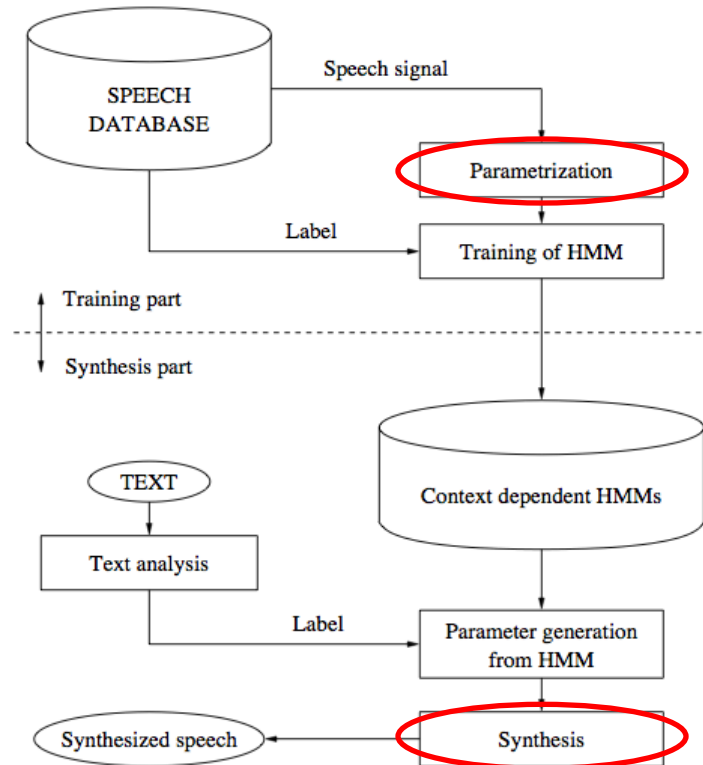
Puhesignaali



Estimoitu  
glottisheräte

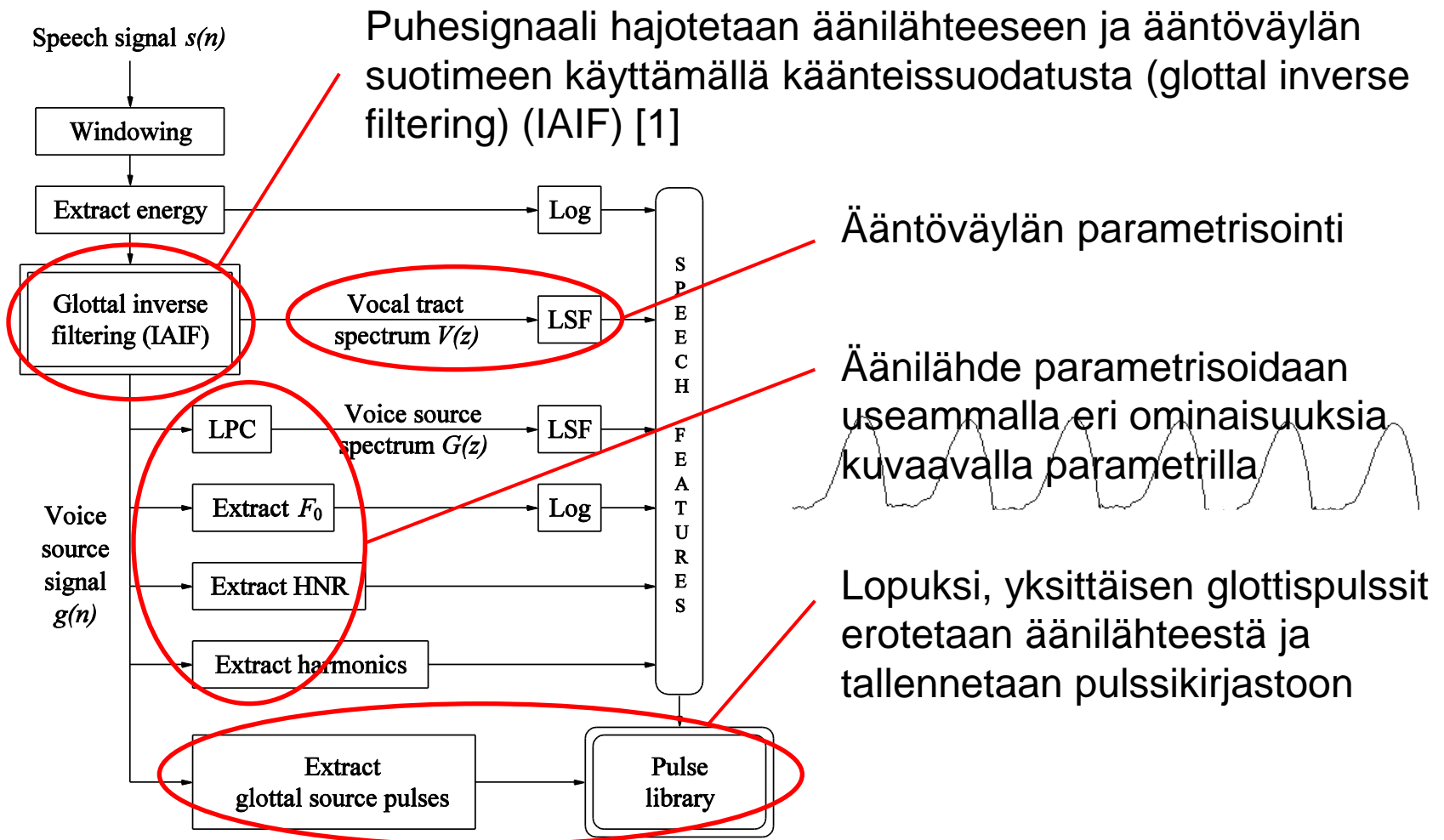


# III. Puhesynteesijärjestelmä



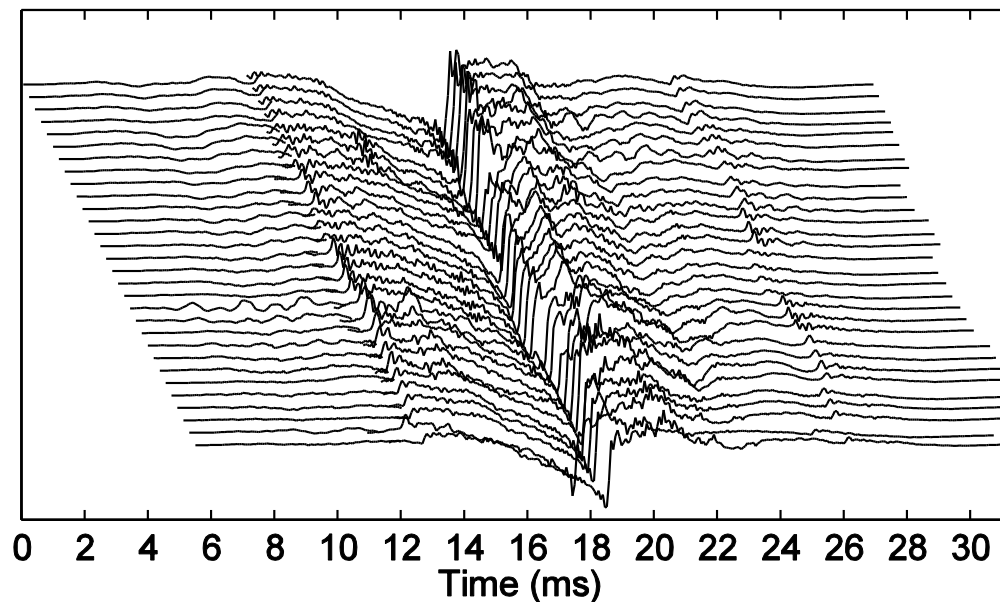


# Puheen parametrisointi



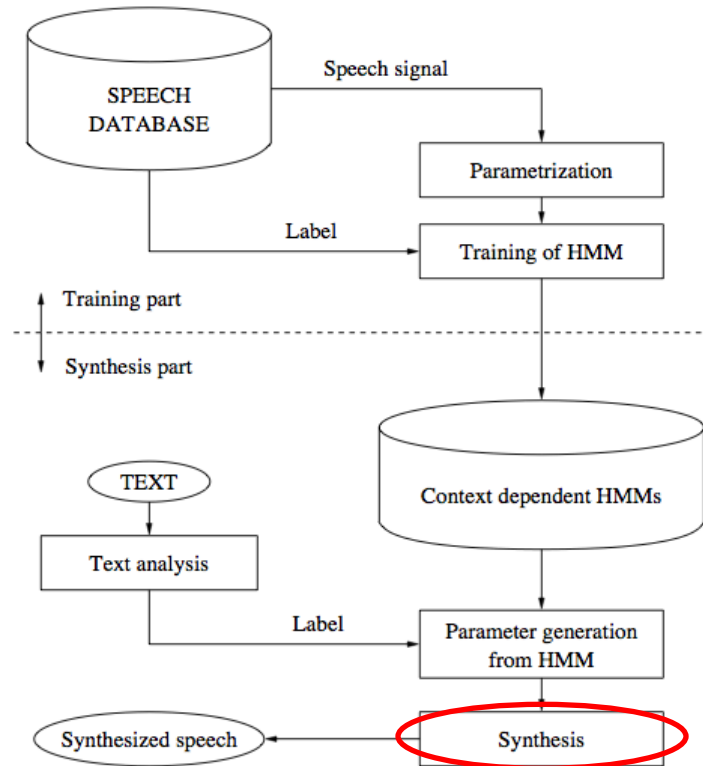
# Pulssikirjasto

Pulssikirjasto sisältää satoja tai jopa tuhansia glottispulsseja sekä niitä vastaavat äänilähdeparametrit.



Ikkunoituja ja differentioituja glottispulsseja miespuhujan glottispulssikirjastosta.

# III. Puhesynteesijärjestelmä



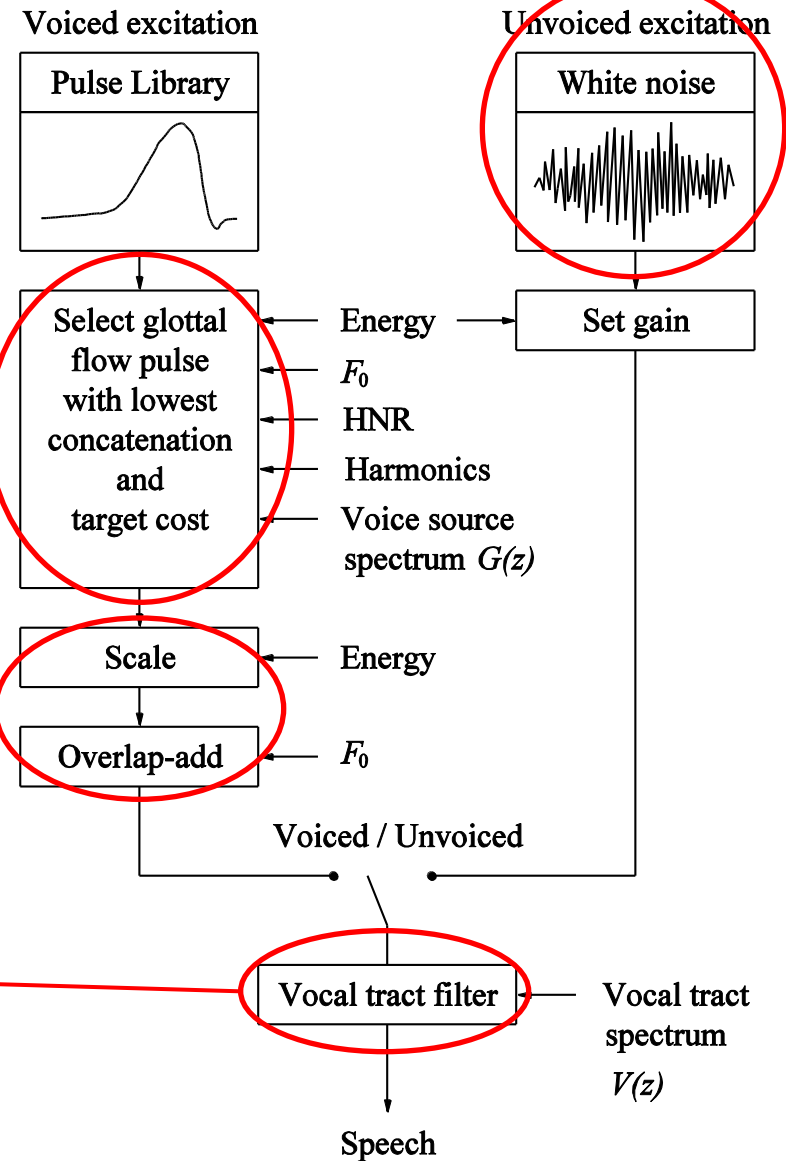
# Synteesi

Synteesivaiheessa herätesignaali generoidaan etsimällä parhaiten lähdeparametreihin sopivat glottispulssit pulssikirjastosta

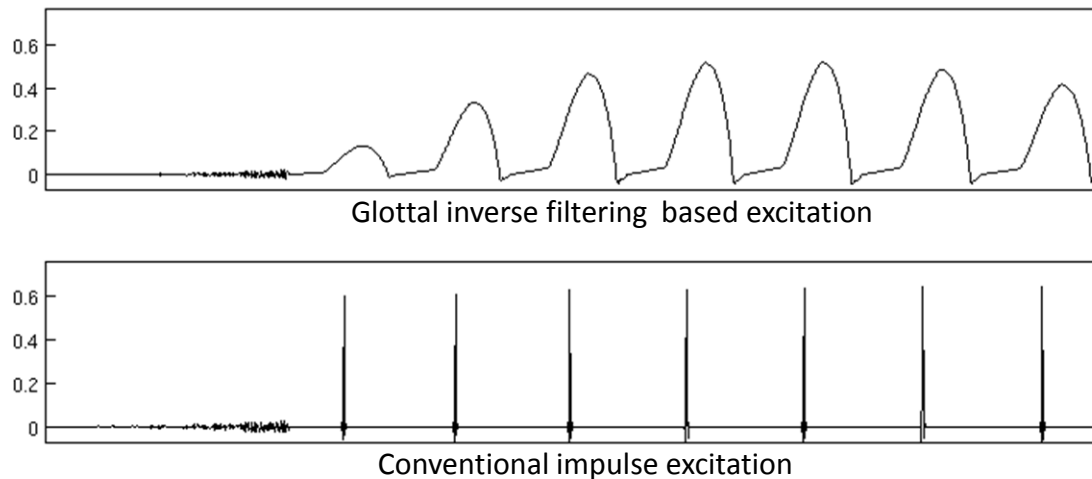
Ainoastaan pulssien amplitudia skaalataan, jonka jälkeen pulssit liitetään toisiinsa (overlap-add)

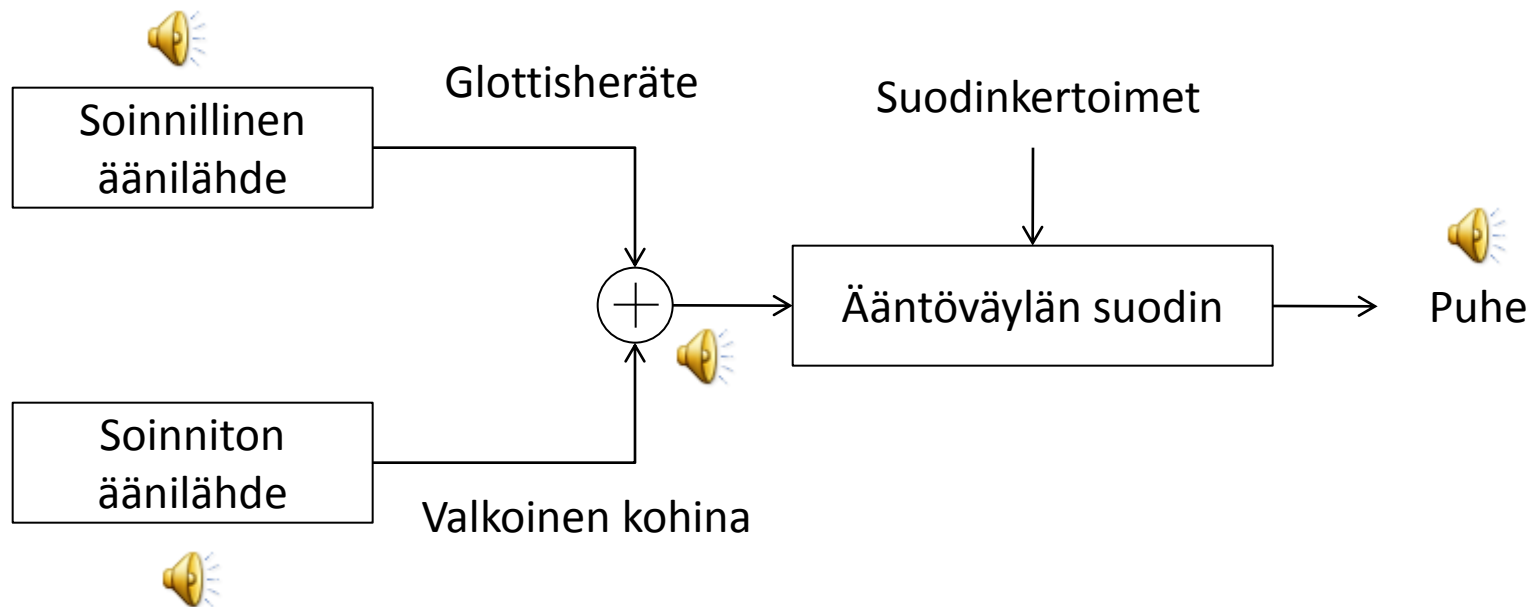
Valkoista kohinaa käytetään soinnittomien äänien herätteenä

Lopuksi yhdistetty heräte suodatetaan ääniväylän suotimella, jolloin saadaan aikaan puhetta.



Äänilähteen yksityiskohtainen parametrisointi yhdistettynä tallennettuun glottispulssikirjastoon mahdollistaa luonnollisenkaltaisen äänilähteen kontrolloidun generoimisen, ja siten luonnollisen prosodian, puhujien ja puhetyyliä toistamisen.

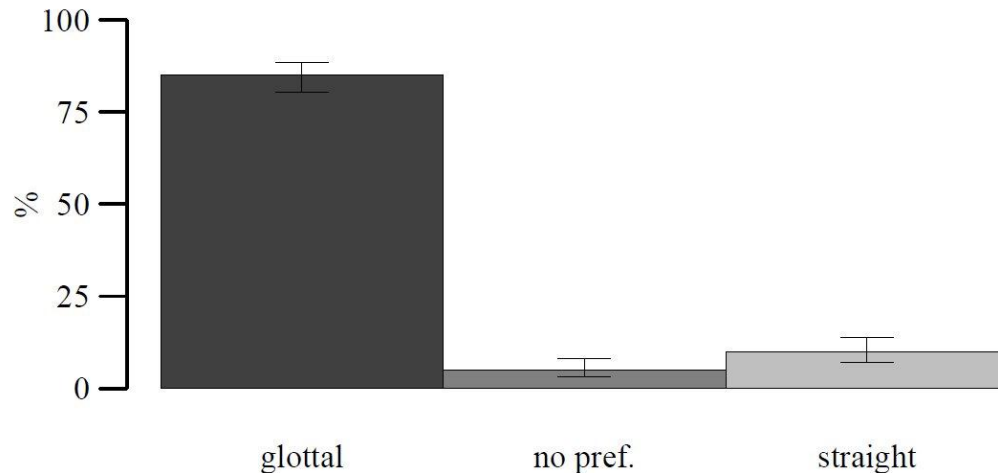




## II. Tuloksia ja ääninäytteitä



Aikaisemmin käytössä oli vain yksi pulssi jota muokattiin äänilähteen parametrien mukaan.

Alla kuuntelukokeen tulos, jossa verrataan yhden pulssin menetelmää eniten käytettyyn hyvälaatuiseen vokooderitekniikkaan STRAIGHT [2].



# Yhden pulssi tekniikka







GlottHMM ääninäytteitä (2008)




Mies (fin)			
			

Ääninäytteet (2009)

Eng	Mies	Nainen
		
		

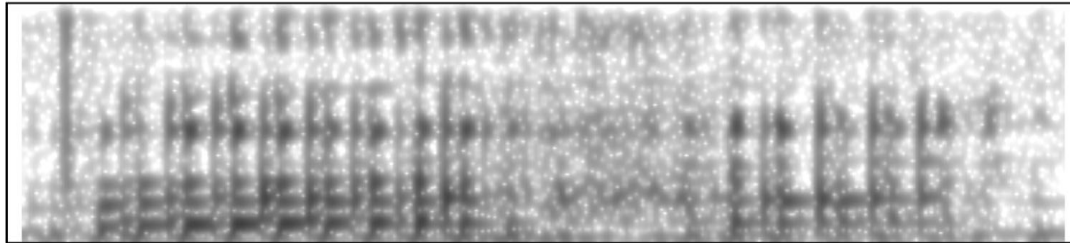
Synteesikisan ääninäytteitä (2010)

Englanti			
Mandariinikiina			

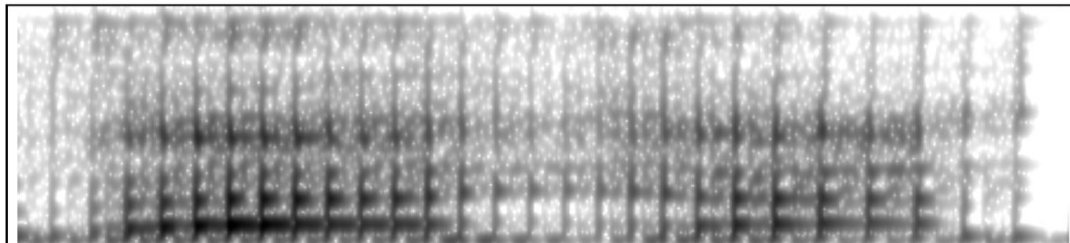
Huutosynteesi	Spontaani synteesi	Laulusynteesi
		



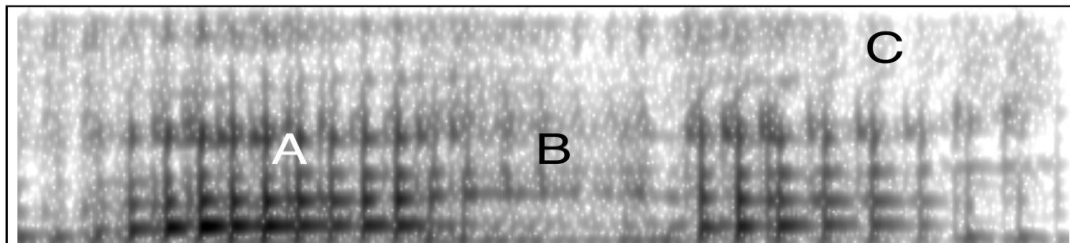
# Pulssikirjastotekniikka



Alkuperäinen



Yhden pulssin tekniikka

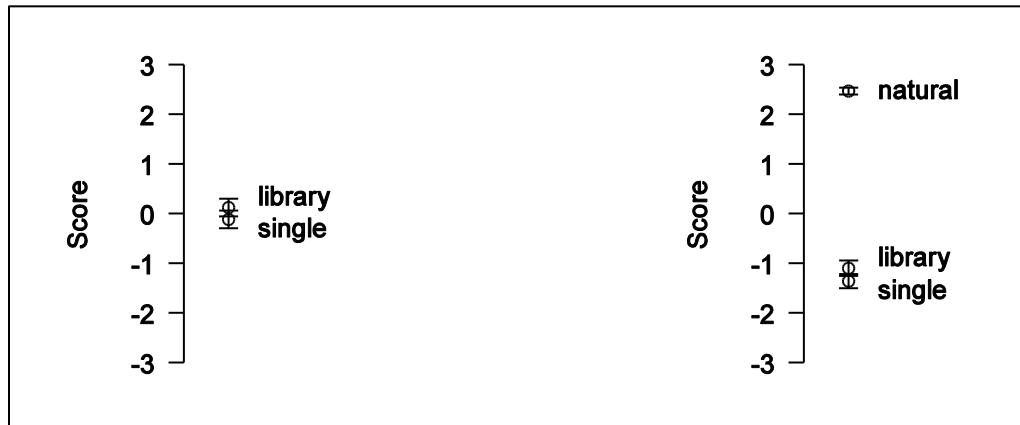


Pulssikirjastotekniikka

Spektrogrammi (0–8000 Hz) sanasta “vähän”. Huomaa parantunut  
A) diplofonian B) soinnillisten frikatiivien C) suurien taajuuksien mallintuminen.

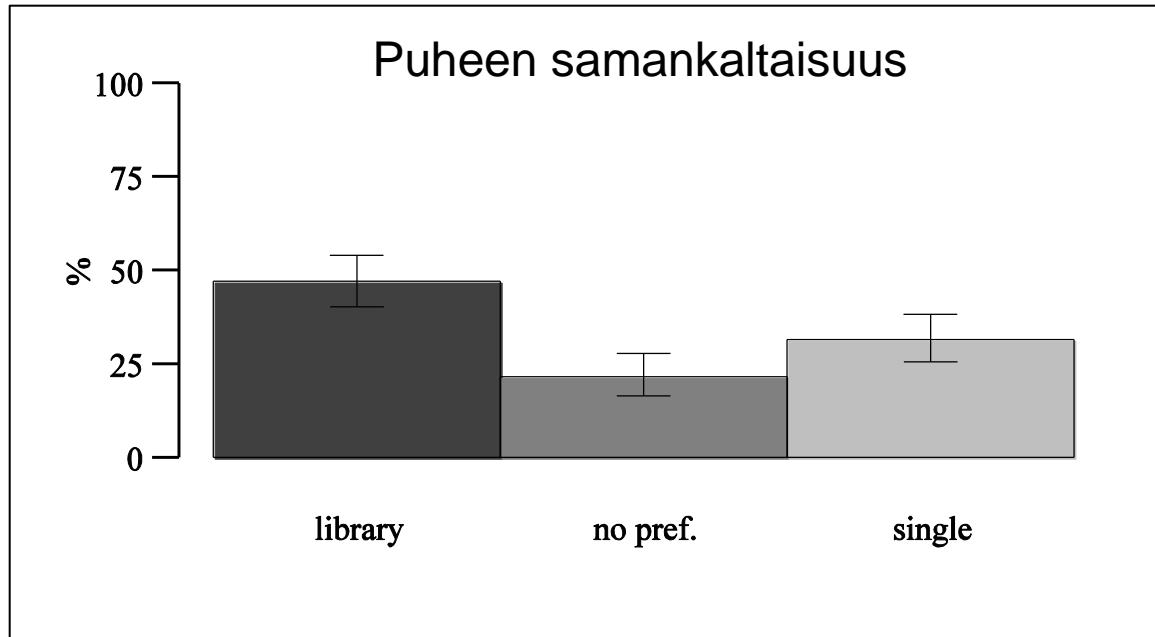
# Pulssikirjasto vs. yhden pulssin tekniikka

Puheen laatu











Pulssikirjastotekniikkaa (library) pidetään hieman yhden pulssin tekniikkaa (single) parempana [3]

# Pulssikirjasto vs. yhden pulssin tekniikka



Pulssikirjastotekniikalla (library) tuotettu puhe on enemmän alkuperäisen äänen kaltainen kuin yhden pulssin tekniikalla (single) tuotettu puhe [3]

# Pulssikirjastotekniikka

Pulssikirjasto (ICASSP'11)	1pulssi	Pulssikirj.
Suomi		
Suomi		
Englanti		
Englanti		

Pulssikirjasto (ICASSP'11)	Pulssikirj.
Lisää näytteitä	  

- ❑ Uusi ihmisen puheentuottoon perustuva laadukas puhesyntetisaattori
- ❑ Mahdollistaa puheen eri piirteiden paremman mallintamisen ja toistamisen
- ❑ Pulssikirjastotekniikka tuottaa luonnollisemman äänilähteen ja siten mahdollistaa erittäin luonnolliselta kuulostavan puheen tuottamisen

- [1] P. Alku, “Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering,” *Speech Communication*, vol. 11, no. 2–3, pp. 109–118, 1992.
- [2] T. Raitio, A. Suni, J. Yamagishi, H. Pulakka, J. Nurminen, M. Vainio, and P. Alku, “HMM-based speech synthesis utilizing glottal inverse filtering,” *IEEE Trans. on Audio, Speech, and Lang. Proc.*, vol. 19, no. 1, pp. 153–165, Jan. 2011.
- [3] T. Raitio, A. Suni, H. Pulakka, M. Vainio, and P. Alku, “Utilizing Glottal Source Pulse Library for Generating Improved Excitation Signal for HMM-Based Speech Synthesis”, accepted for publication in the Intl. Conf. on Acoust., Speech and Signal Processing (ICASSP), 2011.