

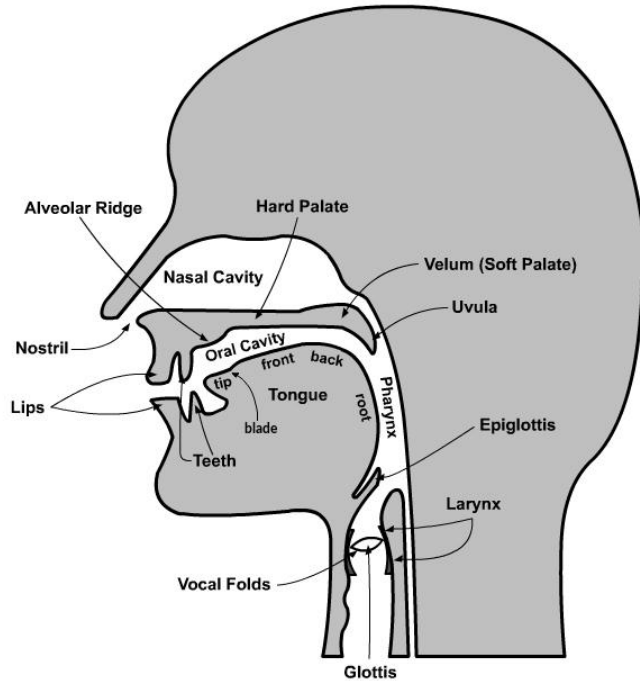
Introduction to real-time speech processing

DPlug meeting

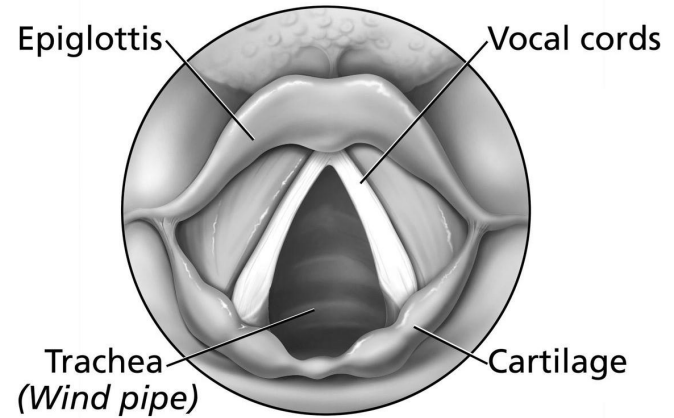
23/08/2022

Principles of speech articulation

Anatomy of speech organs



adapted from MadBeppo.com, 2021



from National Cancer Institute

Phonation

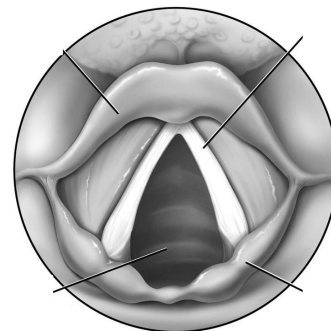
Voiced

- modal
- breathy
- creaky

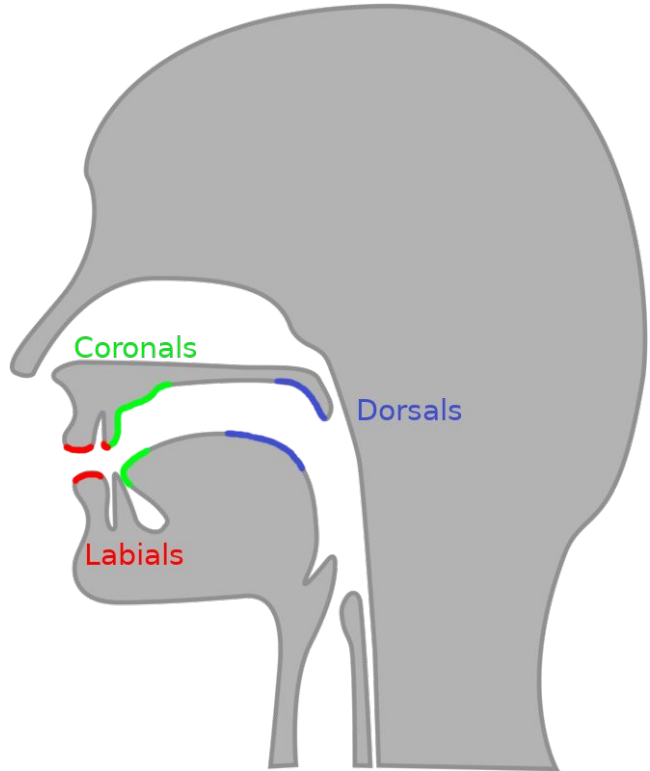
'vain', 'zen', 'game'

Voiceless

'fame', 'sane', 'came'



Places of articulation



Labials

'pie', 'buy', 'my'
'fee', 'vie'

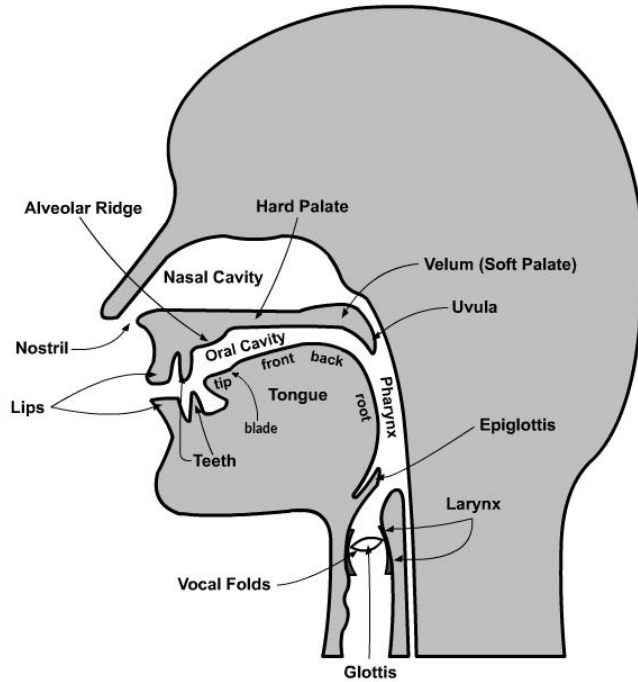
Coronals

'thigh', 'thy'
'tie', 'die', 'nigh', 'sigh', 'zeal', 'lie'
'shy', 'she'

Dorsals

'hack', 'hang', 'hag'

Nasality



Nasal *'ram', 'ran', 'rang'*

Nasalized *'on', 'in' in french*

Manners of articulation

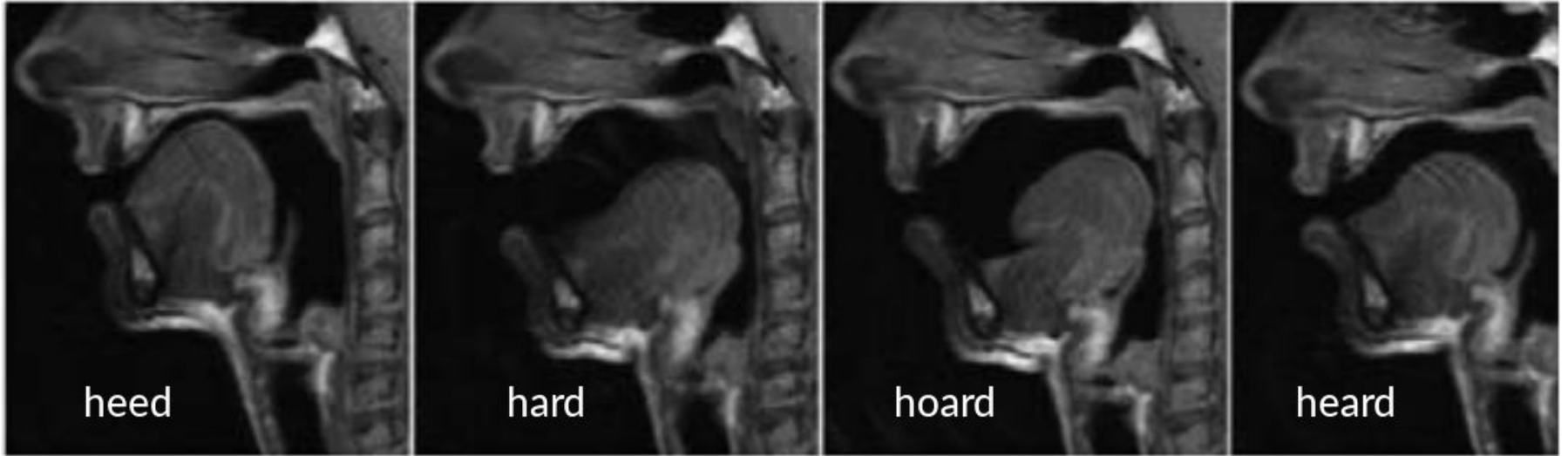
Plosives *'pie', 'buy'*
 'tie', 'die'
 'kye', 'guy'

Fricatives *'fie', 'vie'*
 'thigh', 'thy'
 'sigh', 'zeal'
 'shy'

Approximants *'lie'*
 'we'



Vowels



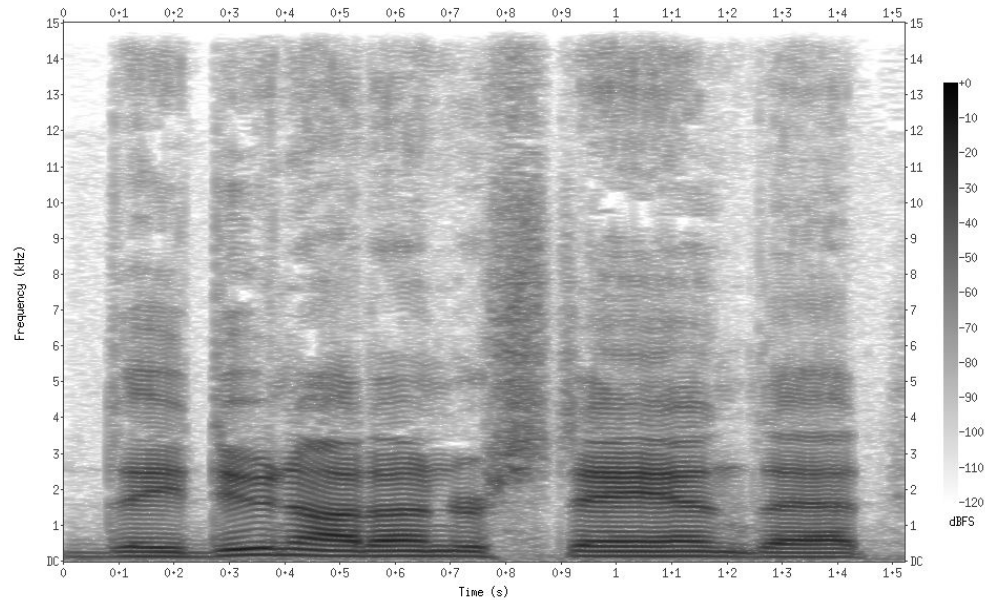
MRI of 4 New Zealand English vowels
from Watson et al., 2009

Speech acoustics

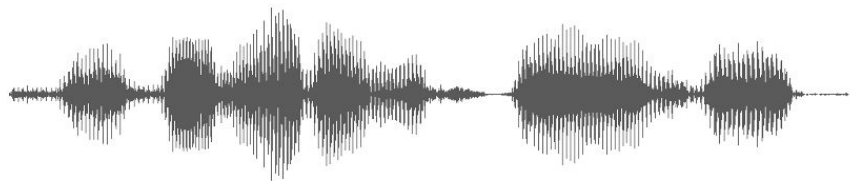
Example spectrogram of speech



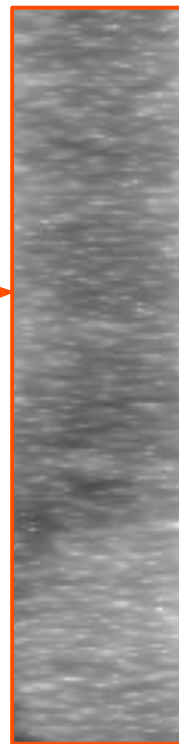
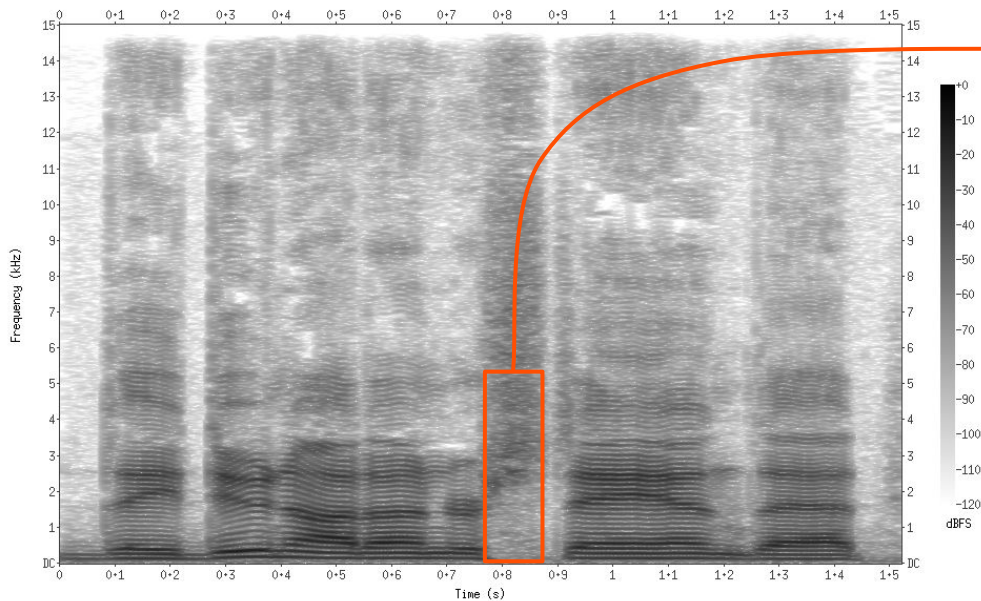
they do not understand that



Phonation - F0

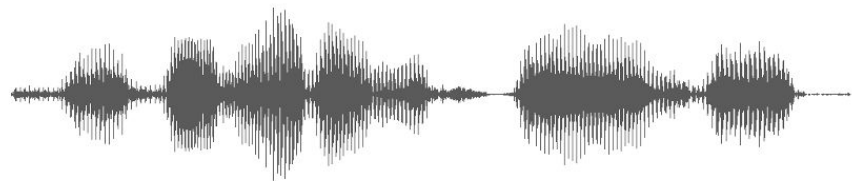


they don't understand that

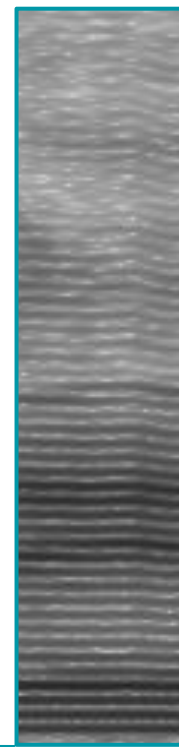
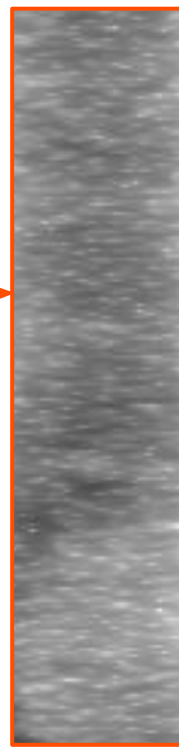
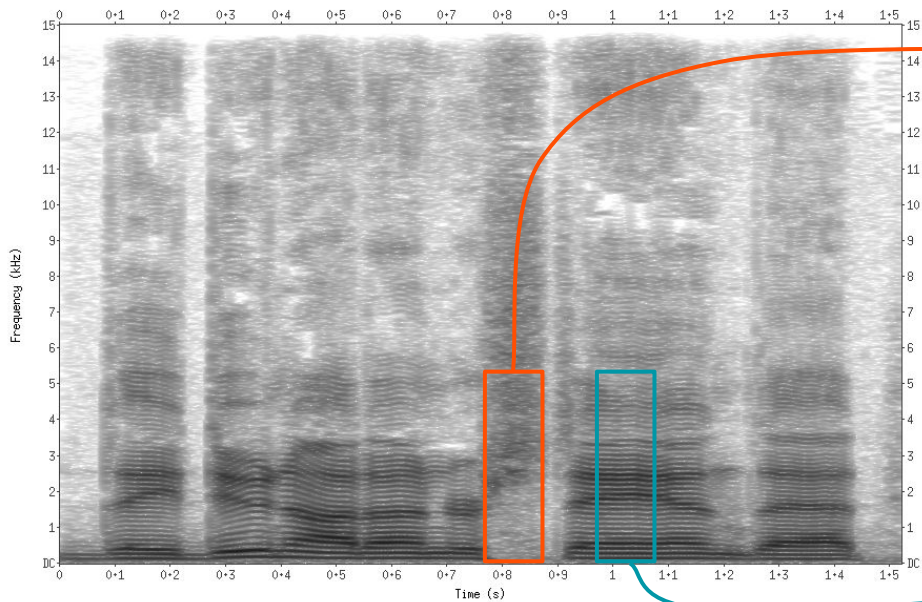


Voiceless

Phonation - F0

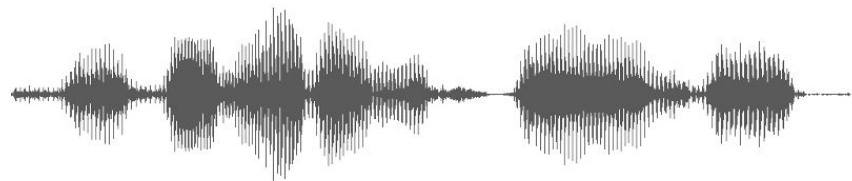


they don't understand that

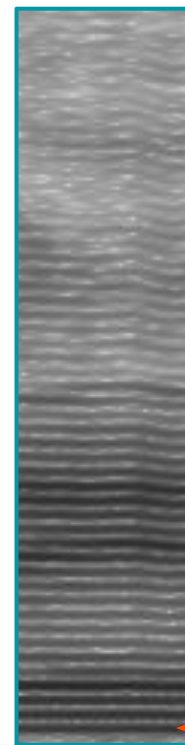
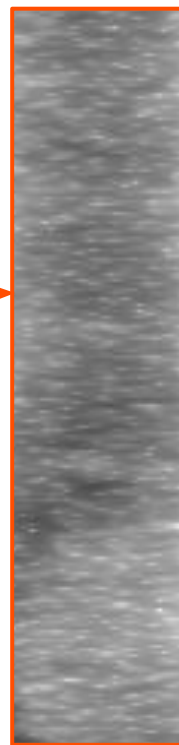
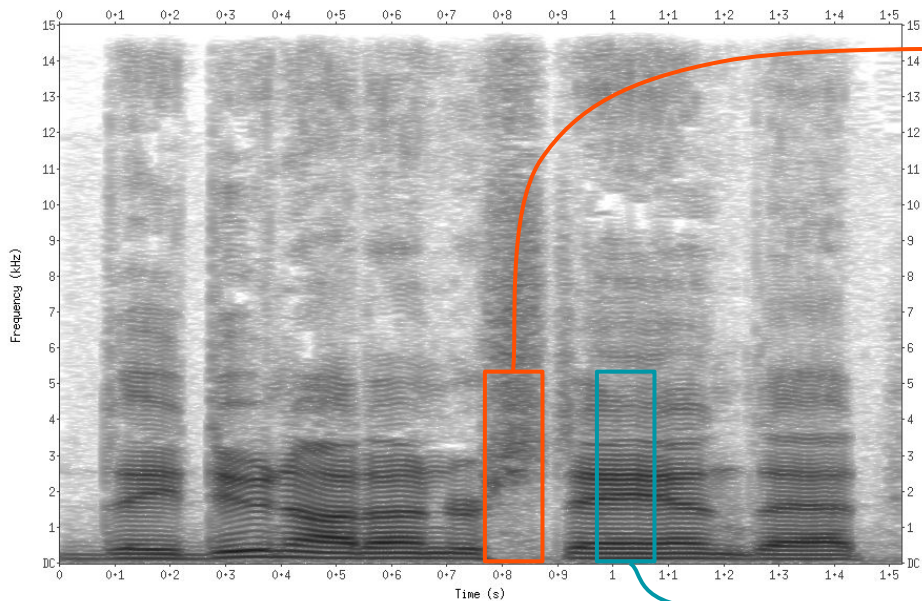


Voiced

Phonation - F0

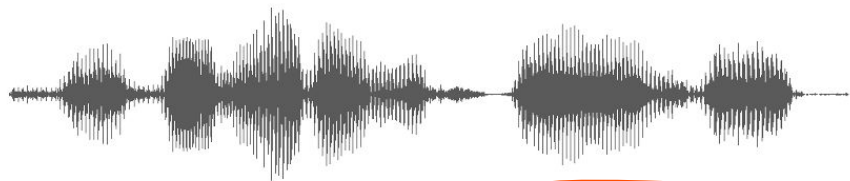


they do not understand that

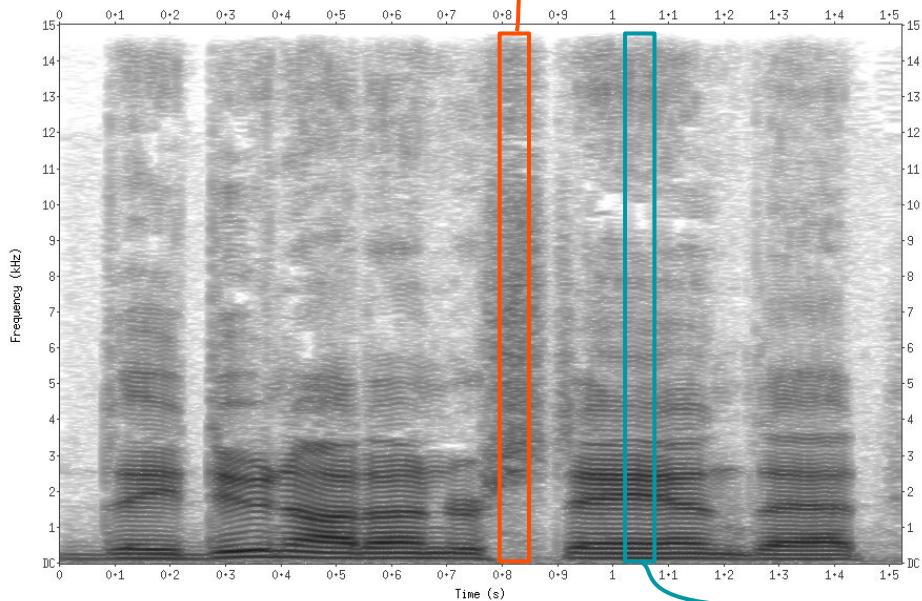


F0

Mel cepstrum



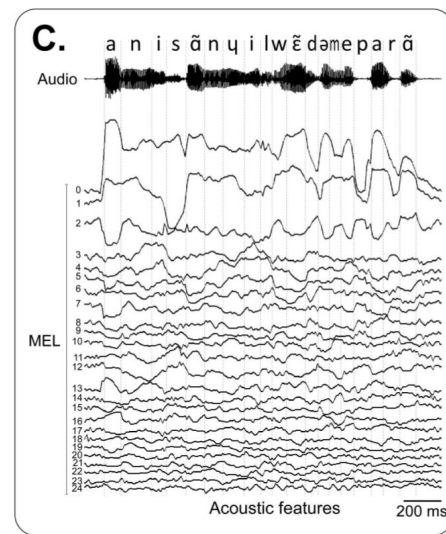
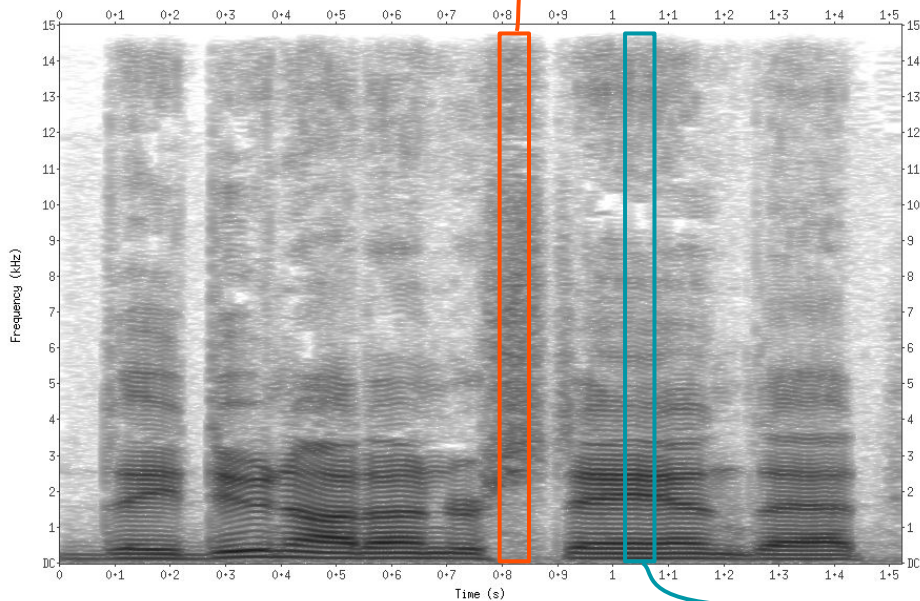
they do not understand that



Mel cepstrum

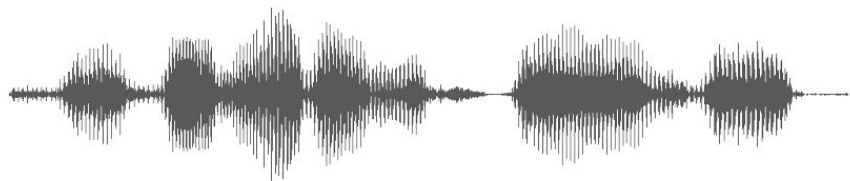


they do not understand that

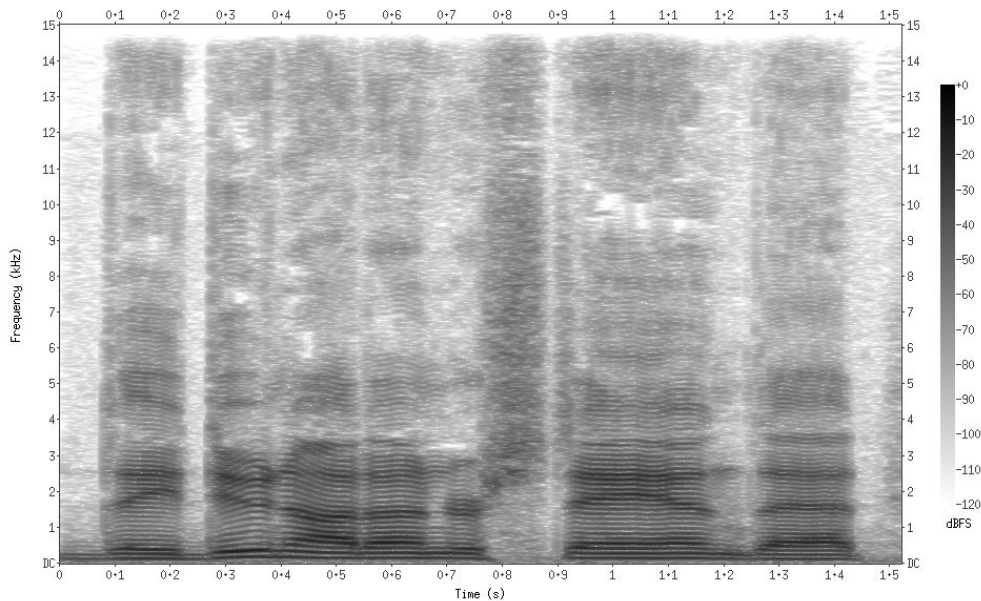


25 Mel cepstral coefficients
adapted from Bocquelet et al., 2016c

Formants F1-F2



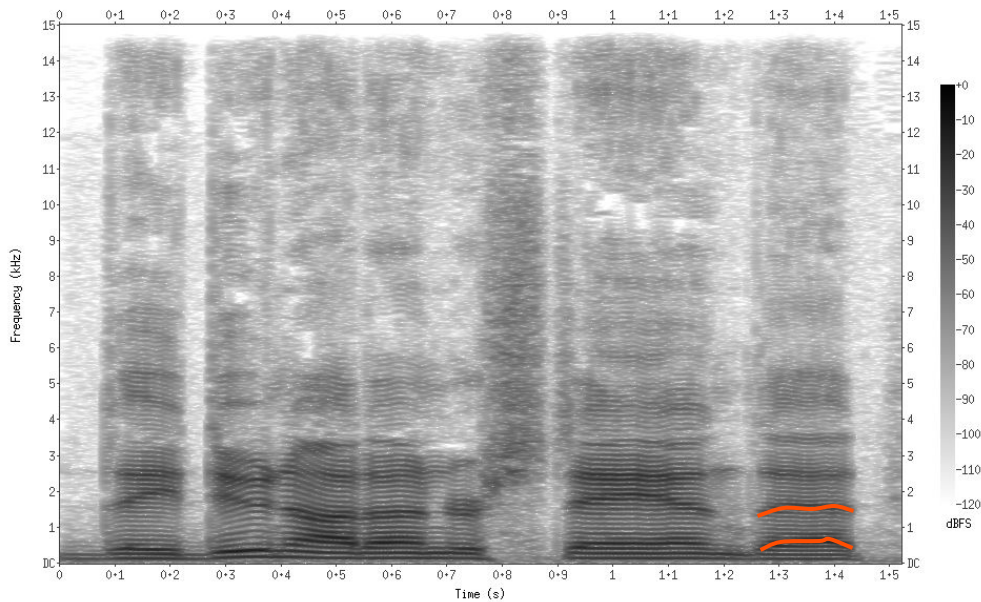
they do not understand that



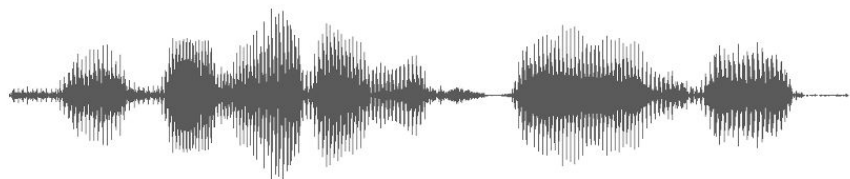
Formants F1-F2



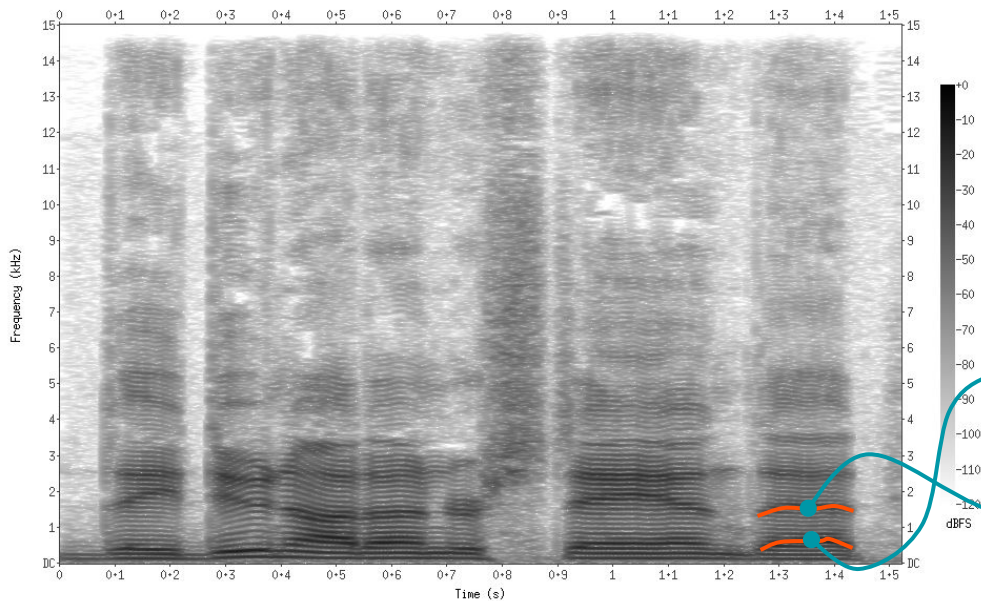
they do not understand that



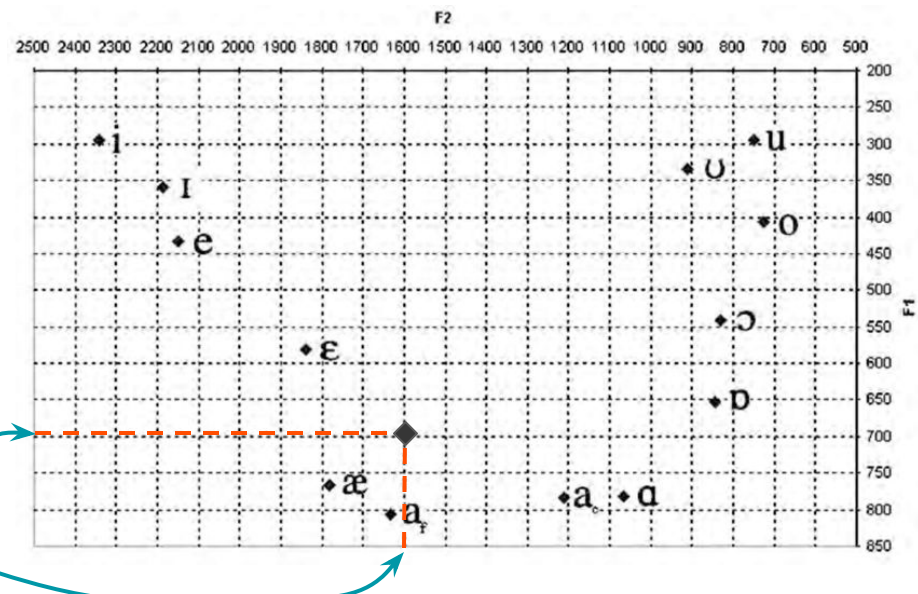
Formants F1-F2



they d o n o t u n d e r s t a n d t h a t

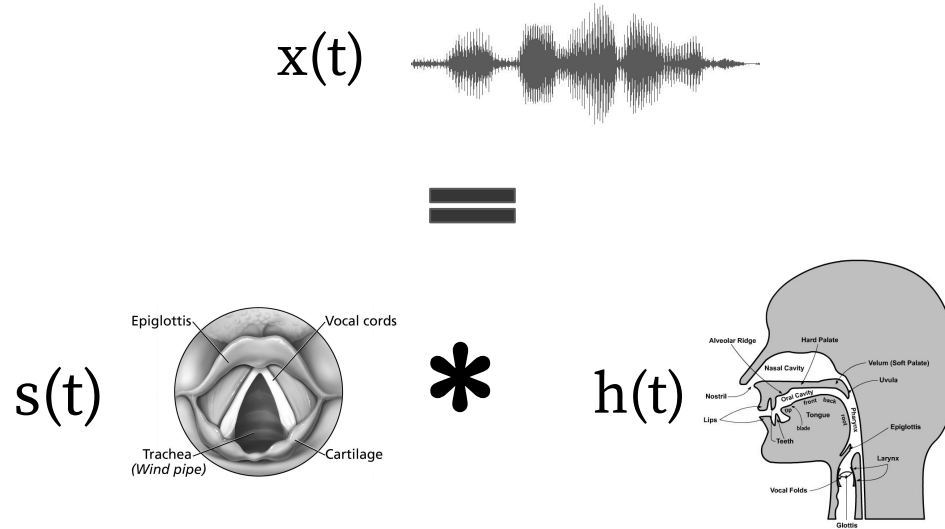


F1 F2 diagram of some IPA vowels (*Hitch, 2017*)

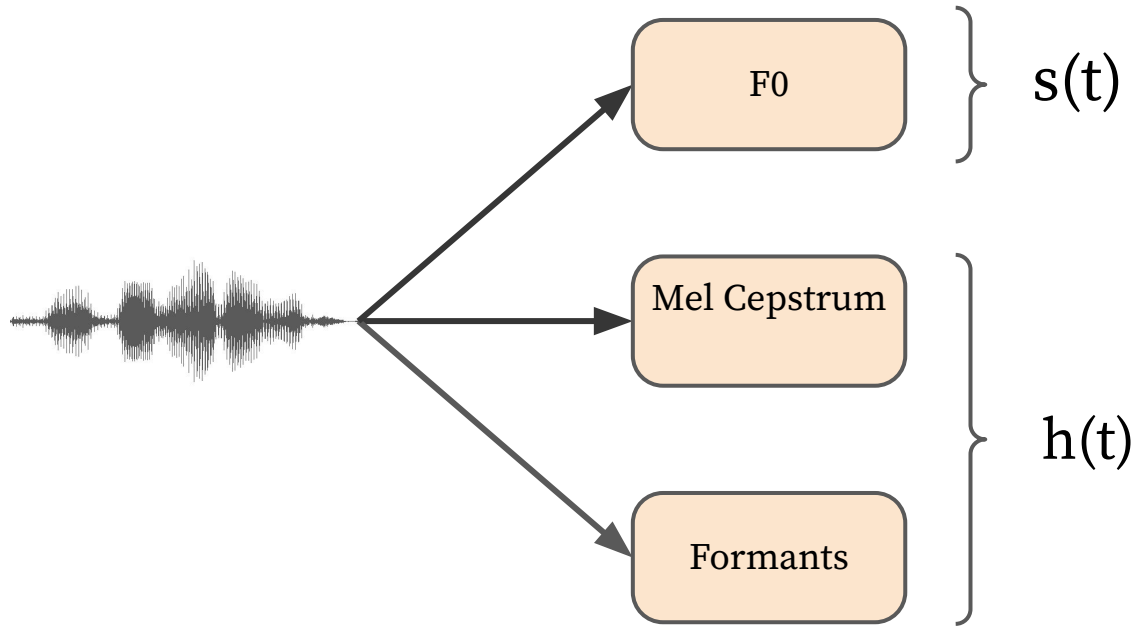


Speech synthesis

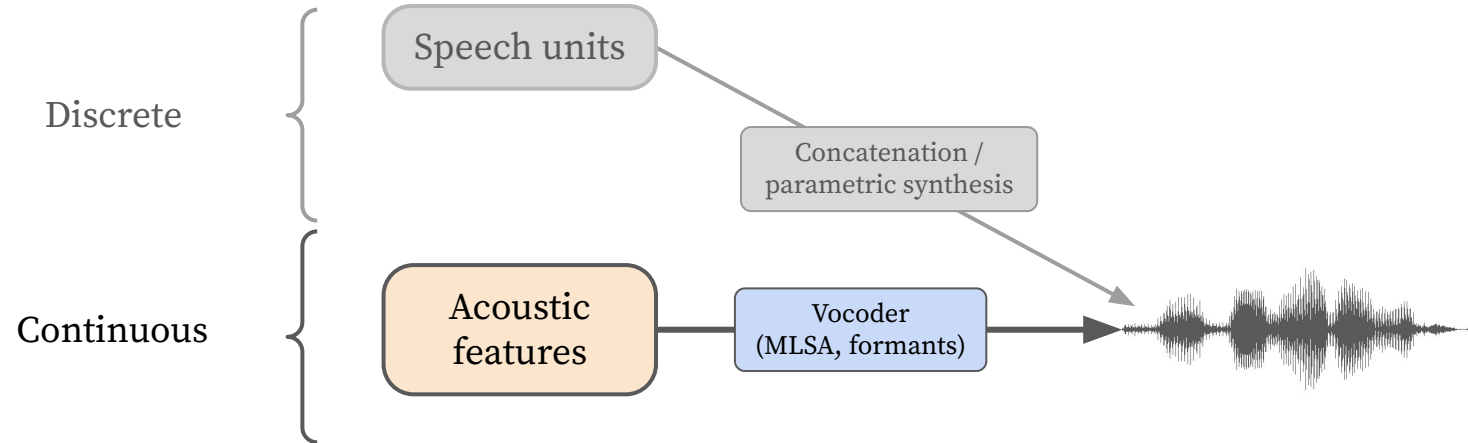
Source-filter model of speech production



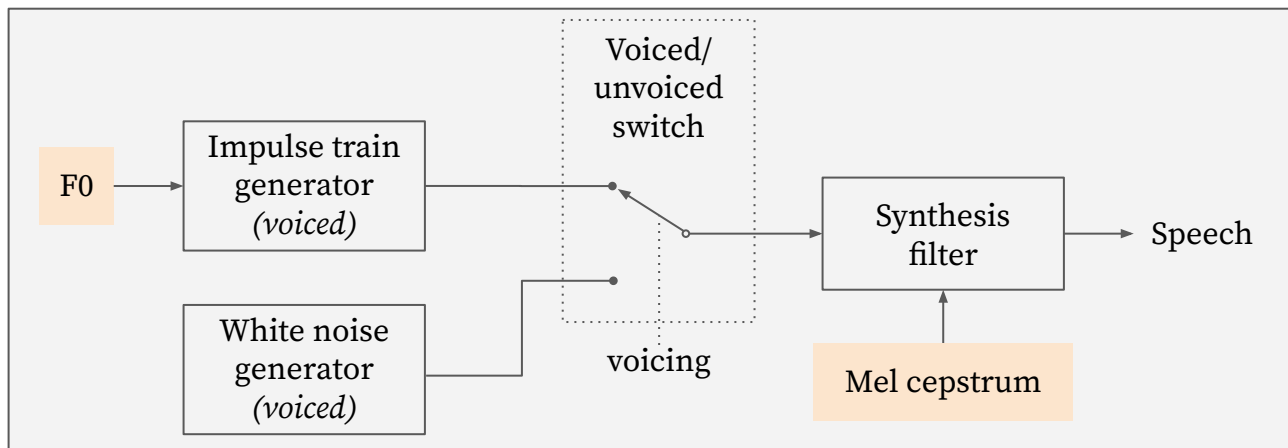
Analysis



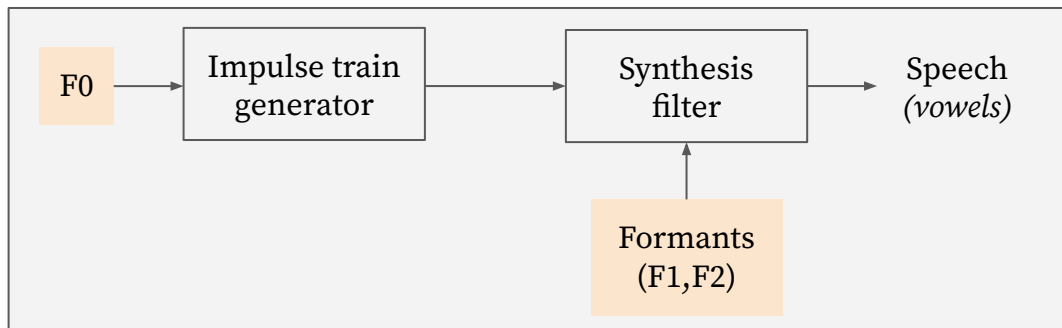
Synthesis



Synthesis



MLSA filter

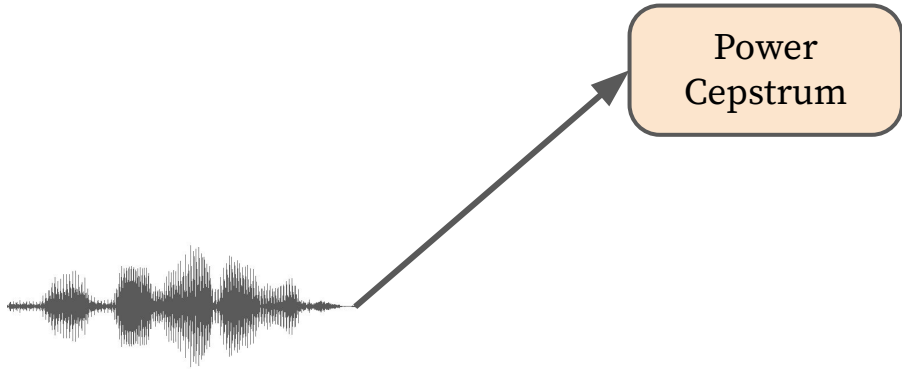


Klatt synthesizer
(simplified for vowels)

Thank you for your attention

Appendix

Cepstral analysis

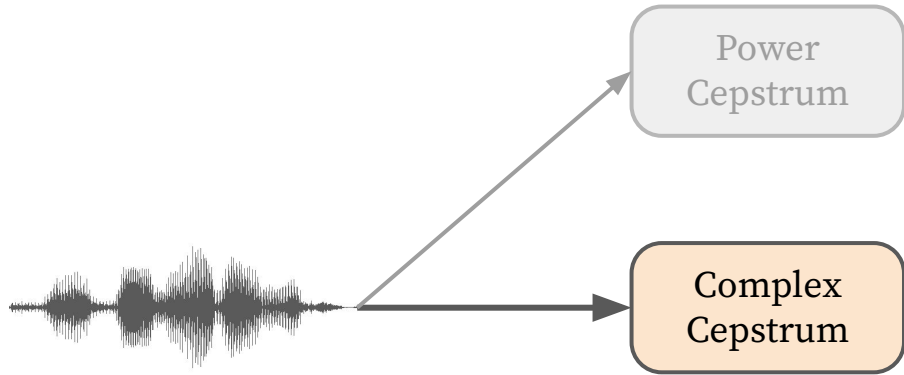


$$C(\tau) \triangleq \left| \mathcal{F}^{-1} \left\{ \log \left(|\mathcal{F} \{x(t)\}|^2 \right) \right\} \right|^2$$

Source-filter separation with the power cepstrum

$$\begin{aligned}C(\tau) &= \left| \mathcal{F}^{-1} \left\{ \log \left(|\mathcal{F} \{s(t) * h(t)\}|^2 \right) \right\} \right|^2 \\&= \left| \mathcal{F}^{-1} \left\{ \log \left(|S(f)|^2 \cdot |H(f)|^2 \right) \right\} \right|^2 \\&= \left| \mathcal{F}^{-1} \left\{ \log \left(|S(f)|^2 \right) + \log \left(|H(f)|^2 \right) \right\} \right|^2 \\&= \left| \mathcal{F}^{-1} \left\{ \log \left(|S(f)|^2 \right) \right\} + \mathcal{F}^{-1} \left\{ \log \left(|H(f)|^2 \right) \right\} \right|^2\end{aligned}$$

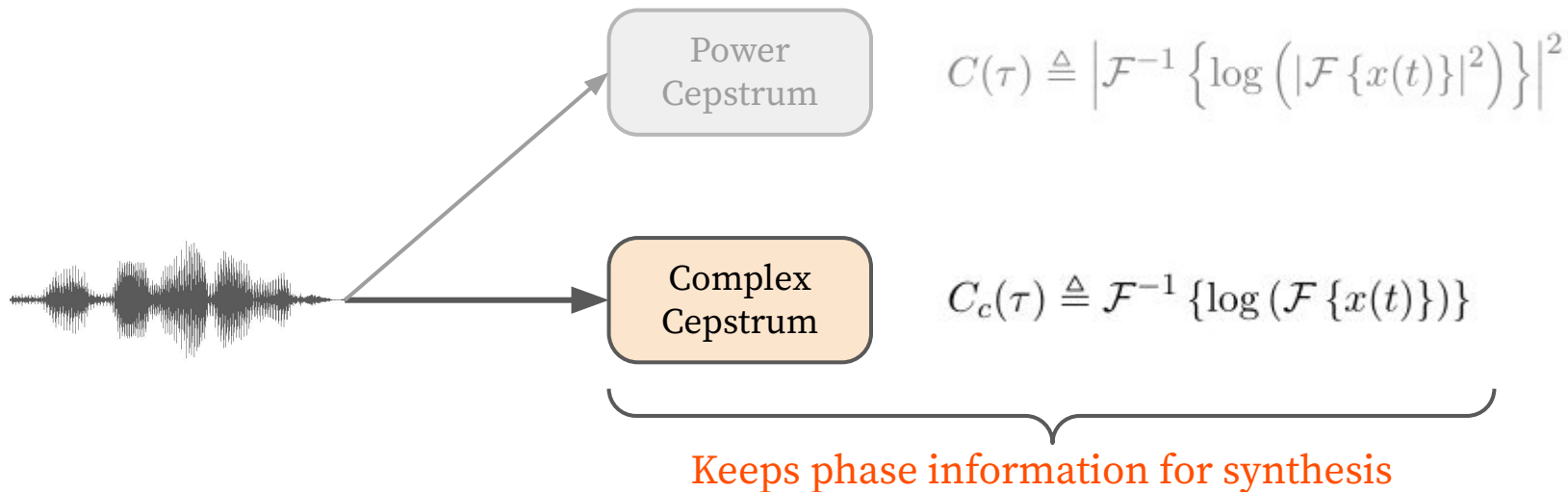
Cepstral analysis



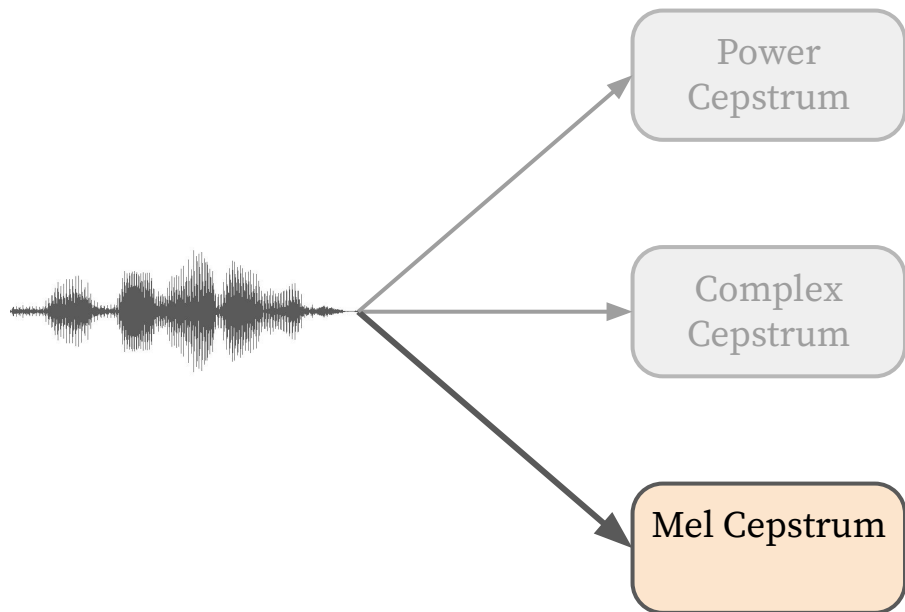
$$C(\tau) \triangleq \left| \mathcal{F}^{-1} \left\{ \log \left(|\mathcal{F} \{x(t)\}|^2 \right) \right\} \right|^2$$

$$C_c(\tau) \triangleq \mathcal{F}^{-1} \{ \log (\mathcal{F} \{x(t)\}) \}$$

Cepstral analysis



Cepstral analysis

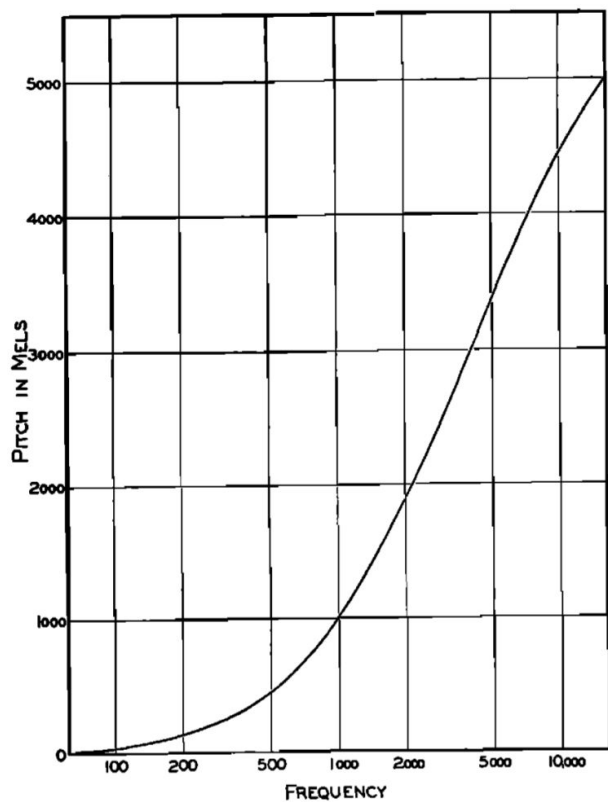


$$C(\tau) \triangleq \left| \mathcal{F}^{-1} \left\{ \log \left(|\mathcal{F} \{x(t)\}|^2 \right) \right\} \right|^2$$

$$C_c(\tau) \triangleq \mathcal{F}^{-1} \{ \log (\mathcal{F} \{x(t)\}) \}$$

Complex cepstrum with warping of the spectrum on the mel scale

Mel scale



Original mel scale.

Subjective relation between pitch and frequency
as measured by Stevens et al., 1937