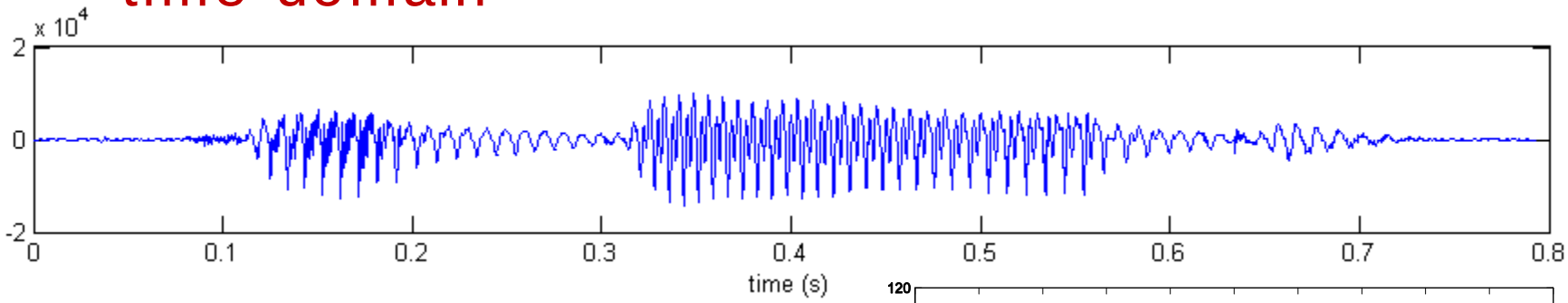
05 Spectrogram Analysis of Thai Speech I

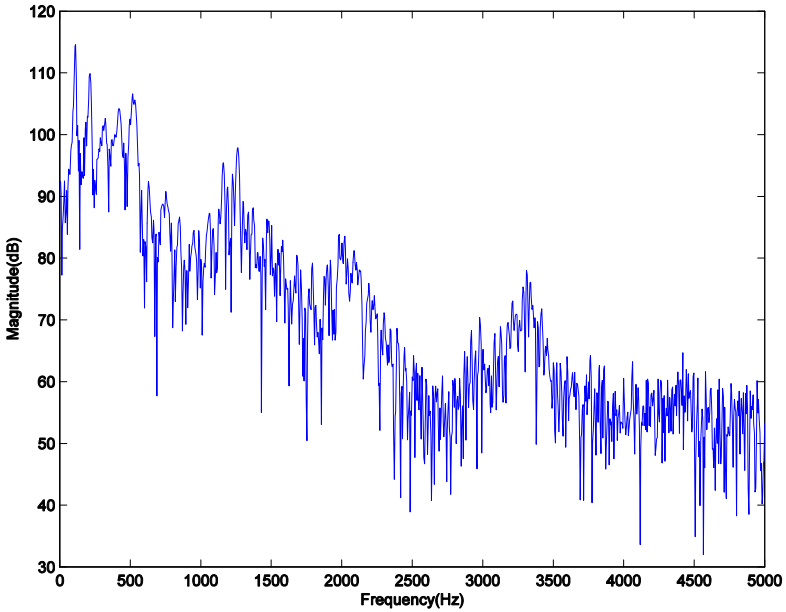
Sound Analysis

- time-domain



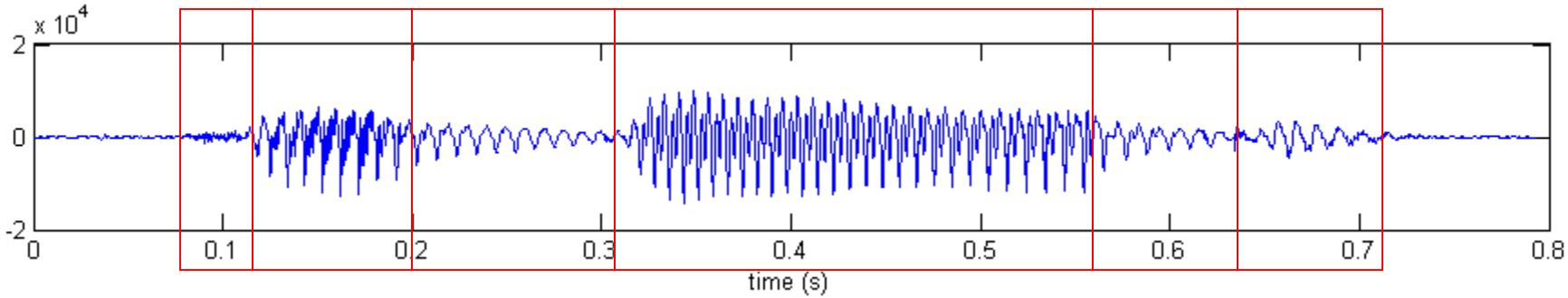
- frequency-domain

  -Fourier transform (discrete-time)



Speech is non-stationary and time-varying signal. So, analyzing speech by lumping all time points together is not very useful.

Time-varying characteristic



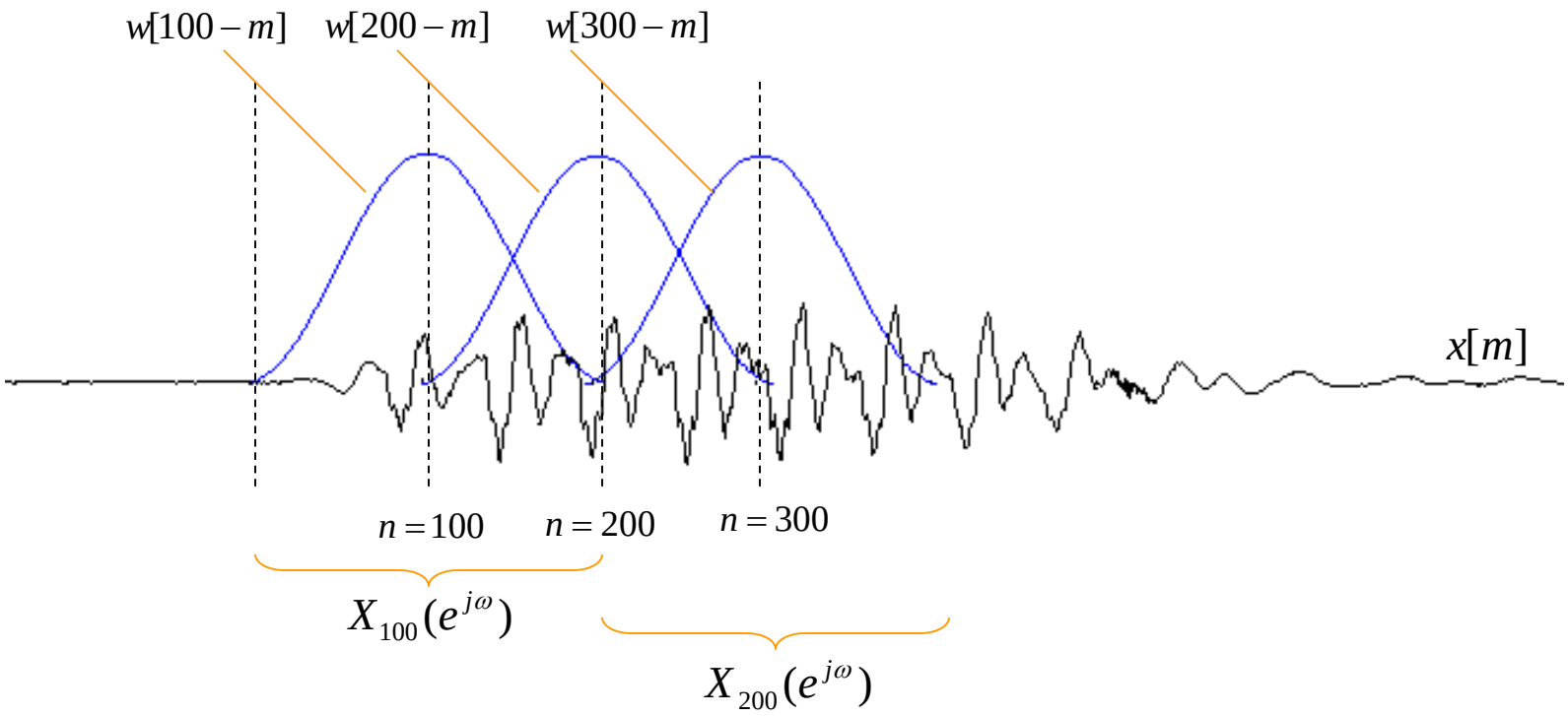
If we find the Fourier transform of the whole signal, we will see the frequency-domain characteristics of the whole signal.

However, do we expect the characteristics of the speech signal in each box to be the same?

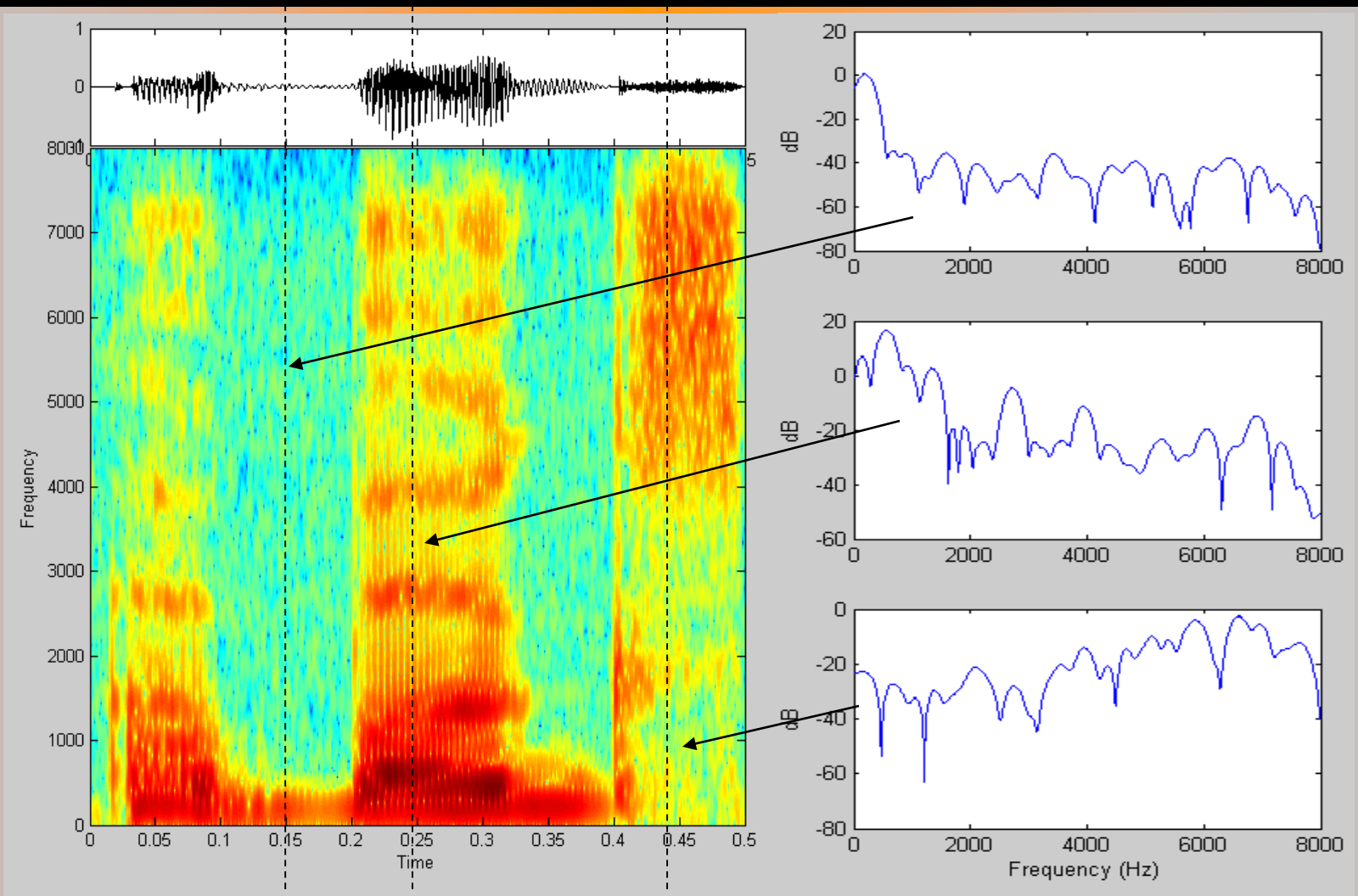
Sound Analysis

- Short-time Fourier Analysis

$$X_n(e^{j\omega}) = \sum_{m=-\infty}^{+\infty} w[n-m]x[m]e^{-j\omega m}$$



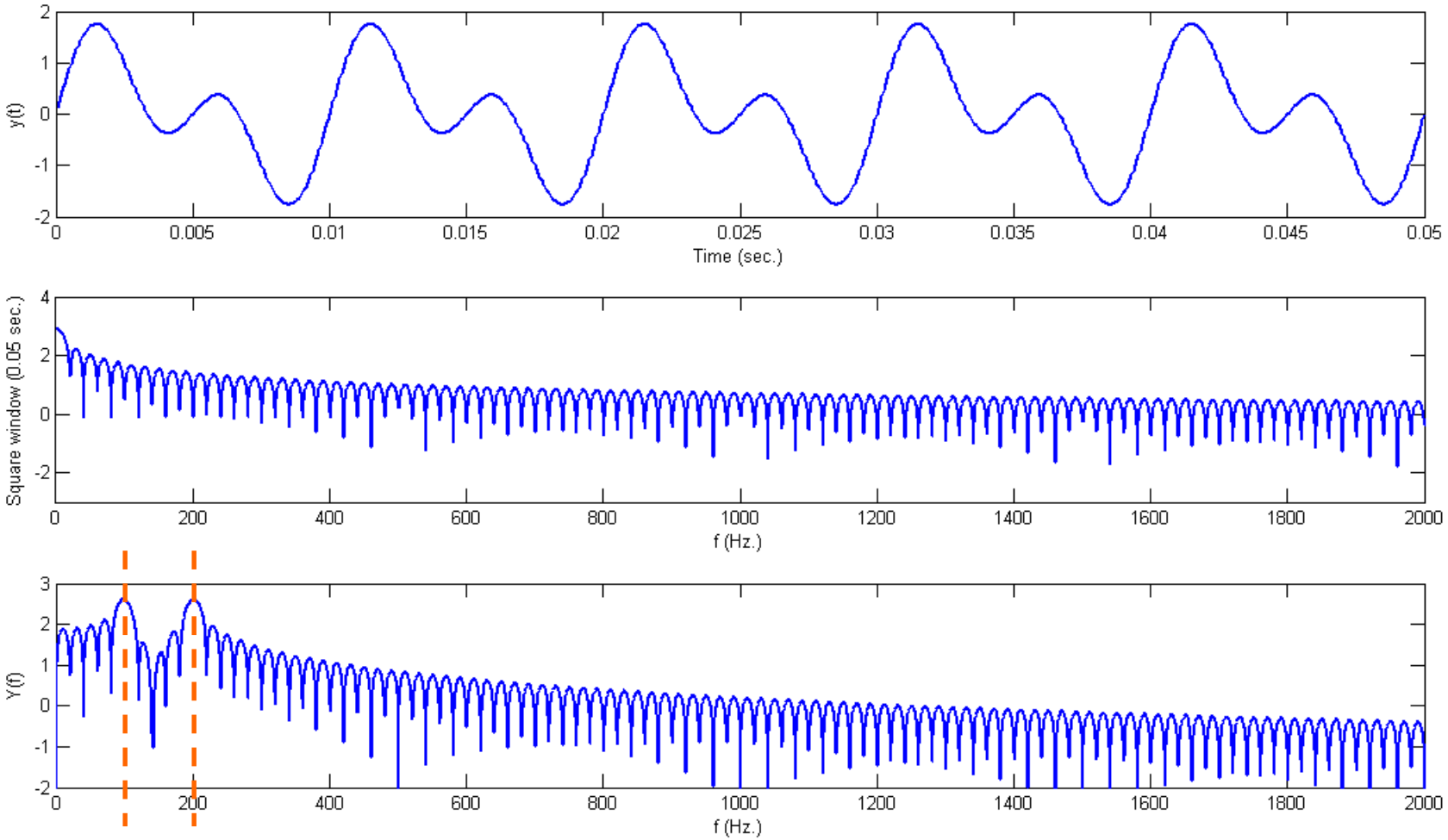
Spectrogram



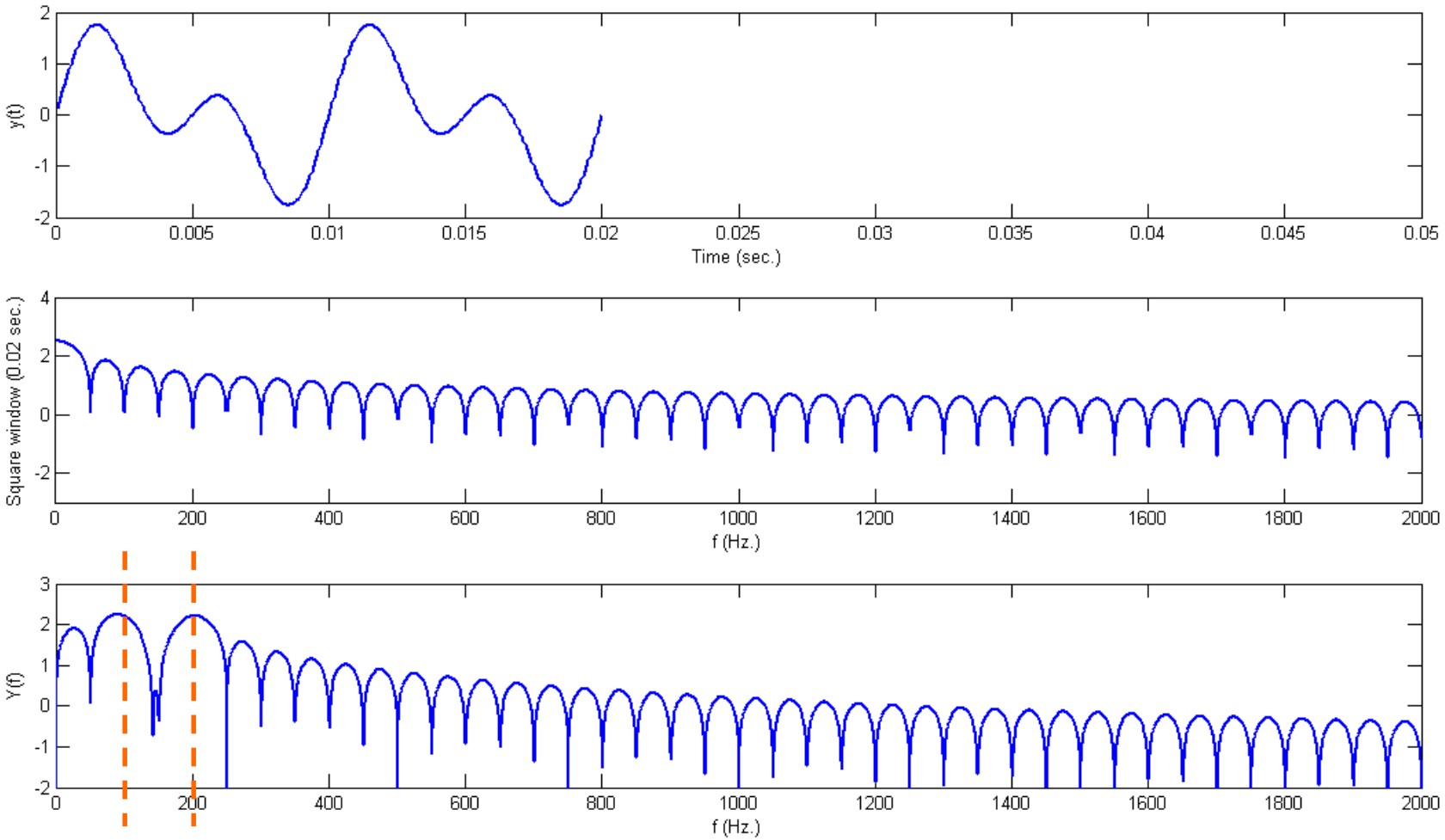
Time vs. Frequency Resolution

- use short time window
  - can capture abrupt acoustic events
  - but Fourier Transform of a short window is long in frequency domain
    - $x[k]w[k] \leftrightarrow X(e^{j\omega}) * W(e^{j\omega})$
    - So, we lose frequency resolution
- use long time window
  - several short acoustic events got mixed in to the same frame
  - a long time window has good frequency resolution

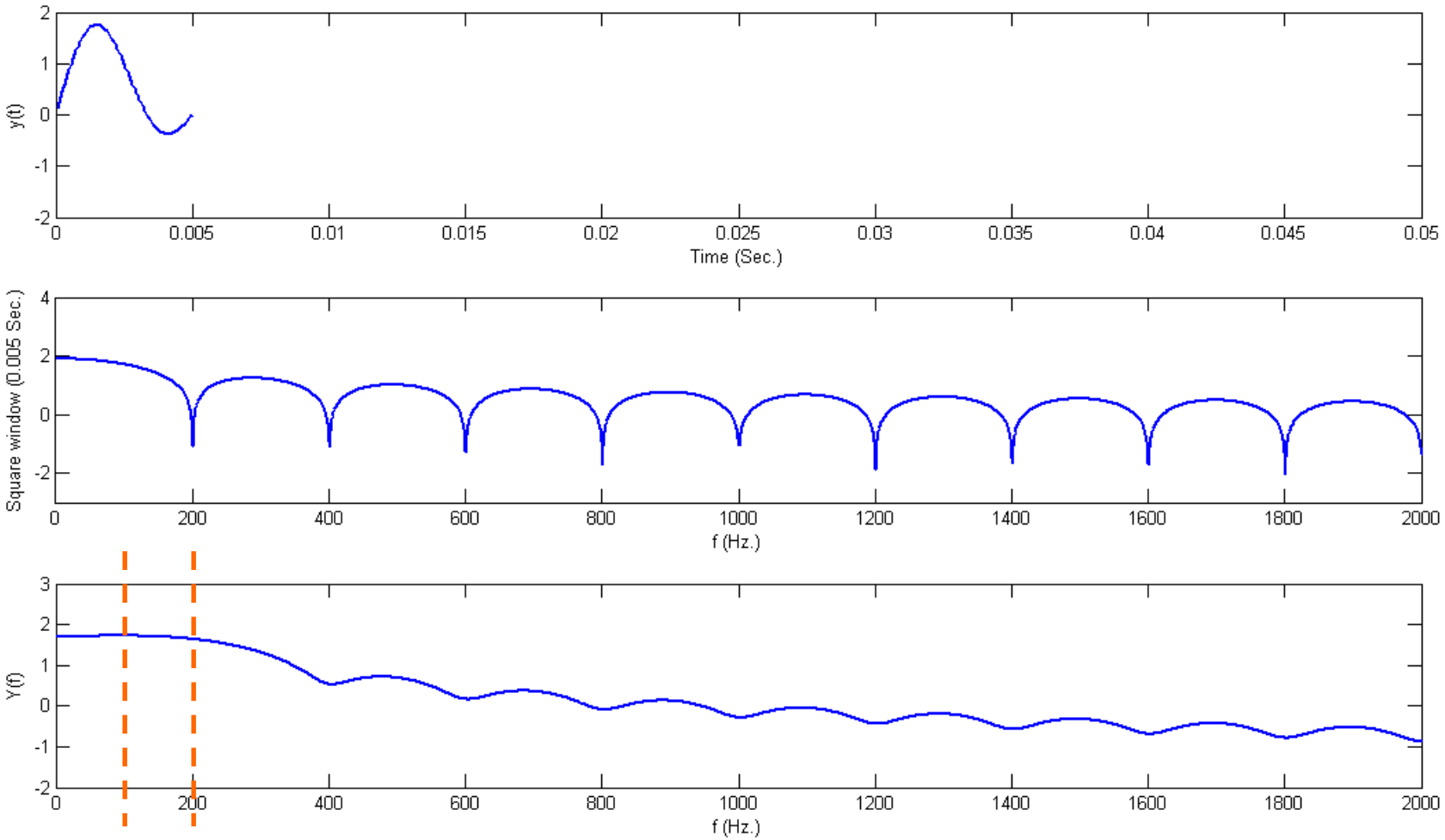
Window Length: Revisited



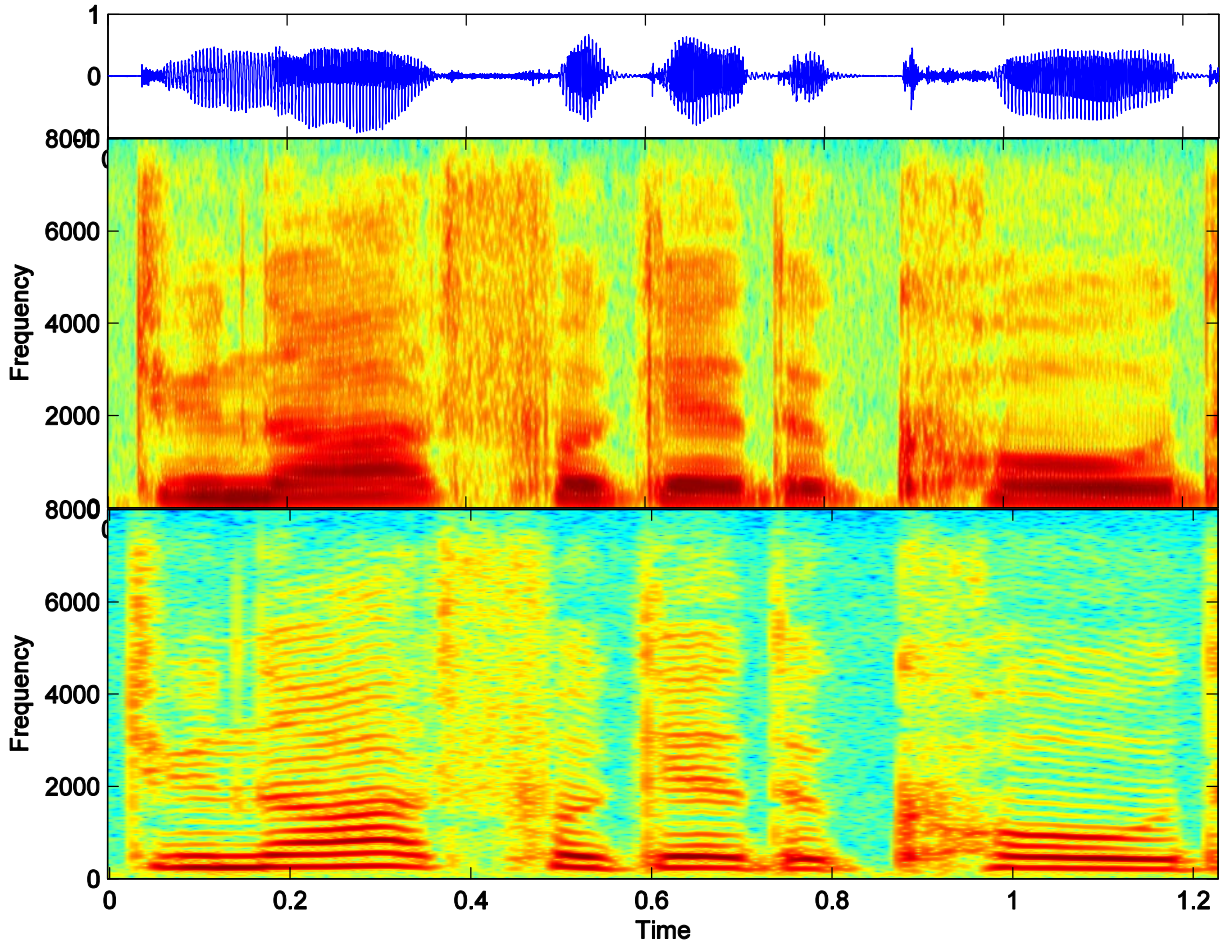
Window Length: Revisited



Window Length: Revisited



Wide-band / Narrow-band Spectrogram



Wide-band

- good time resolution
- can see pitch structure
- can see abrupt acoustic event

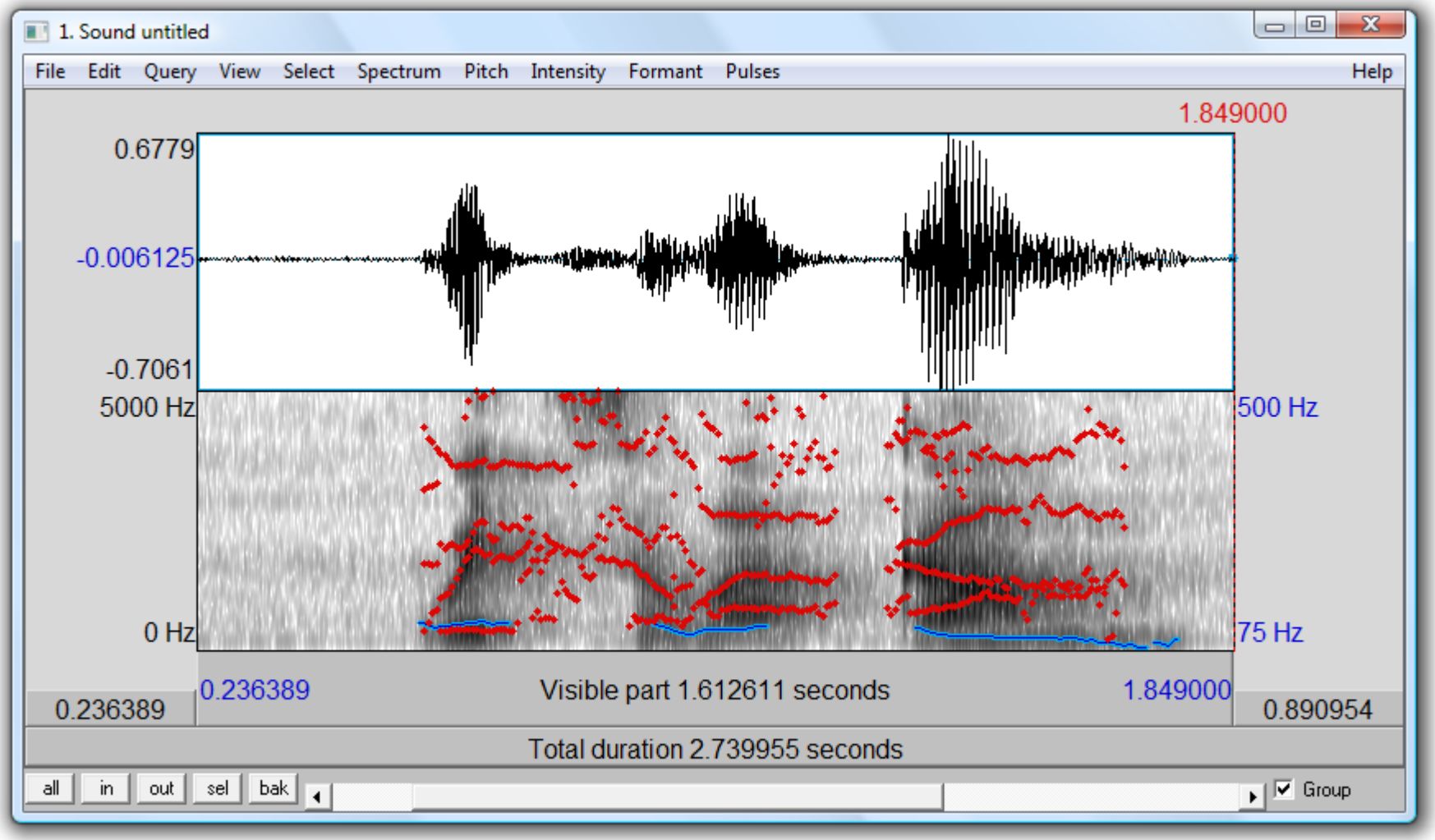
Narrow-band

- good frequency resolution
- can see harmonic structure
- fine frequency discrimination

Praat: Sound Analysis Tool

http://www.praat.org

Praat: Sound Analysis Tool



Praat Functionalities

Speech analysis:
spectral analysis (spectrograms)
pitch analysis
formant analysis
intensity analysis
jitter, shimmer, voice breaks
cochleagram
excitation pattern
Speech synthesis:
from pitch, formant, and intensity
articulatory synthesis
Listening experiments:
identification and discrimination tests
Labelling and segmentation:
label intervals and time points on multiple tiers
use phonetic alphabet
use sound files up to 2 gigabytes (3 hours)
Speech manipulation:
change pitch and duration contours
Filtering

Learning algorithms:
feedforward neural networks
discrete and stochastic Optimality Theory
Statistics:
multidimensional scaling
principal component analysis
discriminant analysis
Graphics:
high quality for your articles and thesis
produce Encapsulated PostScript files
integrated mathematical and phonetic symbols
Programmability:
easy programmable scripting language
communicate with other programs
(the **sendpraat** source code)
create hypertext manuals with sound I/O

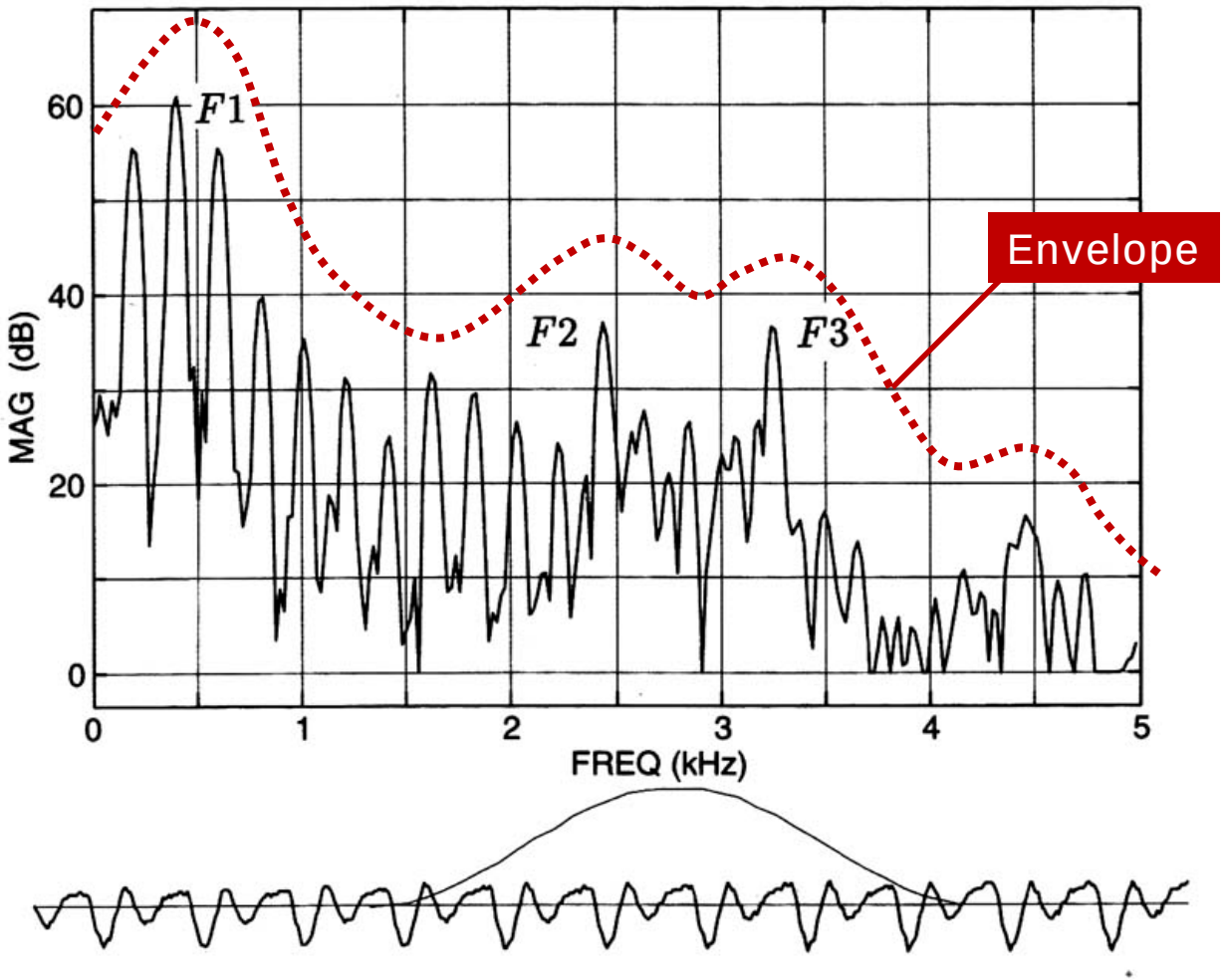
Classes of Sounds

- classify by degree of constriction
  - Consonant
    - obstruction in the path of air flow
      - complete closure
      - narrow constriction
    - abrupt change in vocal tract configuration
    - abrupt change in the signal
  - Semivowel
    - between vowel and consonant
    - non-abrupt change / slightly constricted
  - Vowel
    - relatively no obstruction of the air flow
    - smooth gradual change in the signal

Vowels

- pressure in the vocal tract and subglottal pressure are different enough for outward airflow
- the vocal folds are slack enough to vibrate
- no obstruction in the vocal tract
- different position of the tongue → different vowels
- rounding of the lips

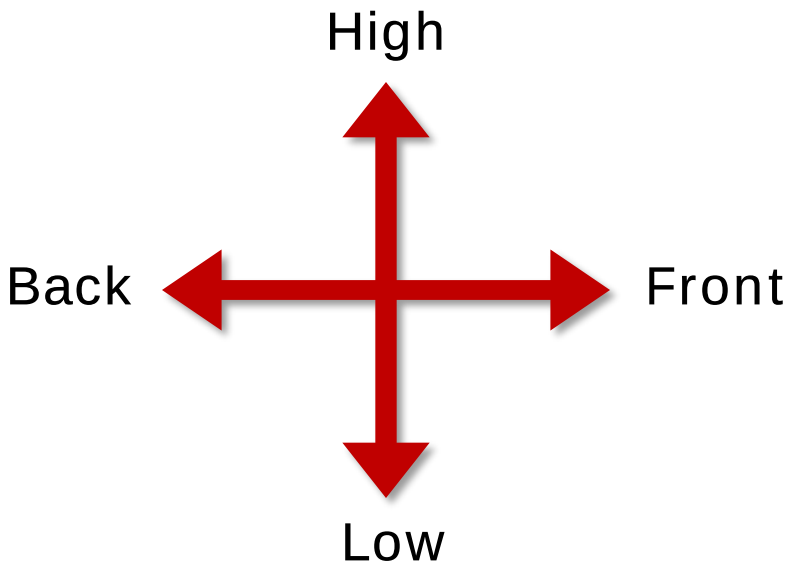
Vowel Spectrum



Picture from
Stevens 1999

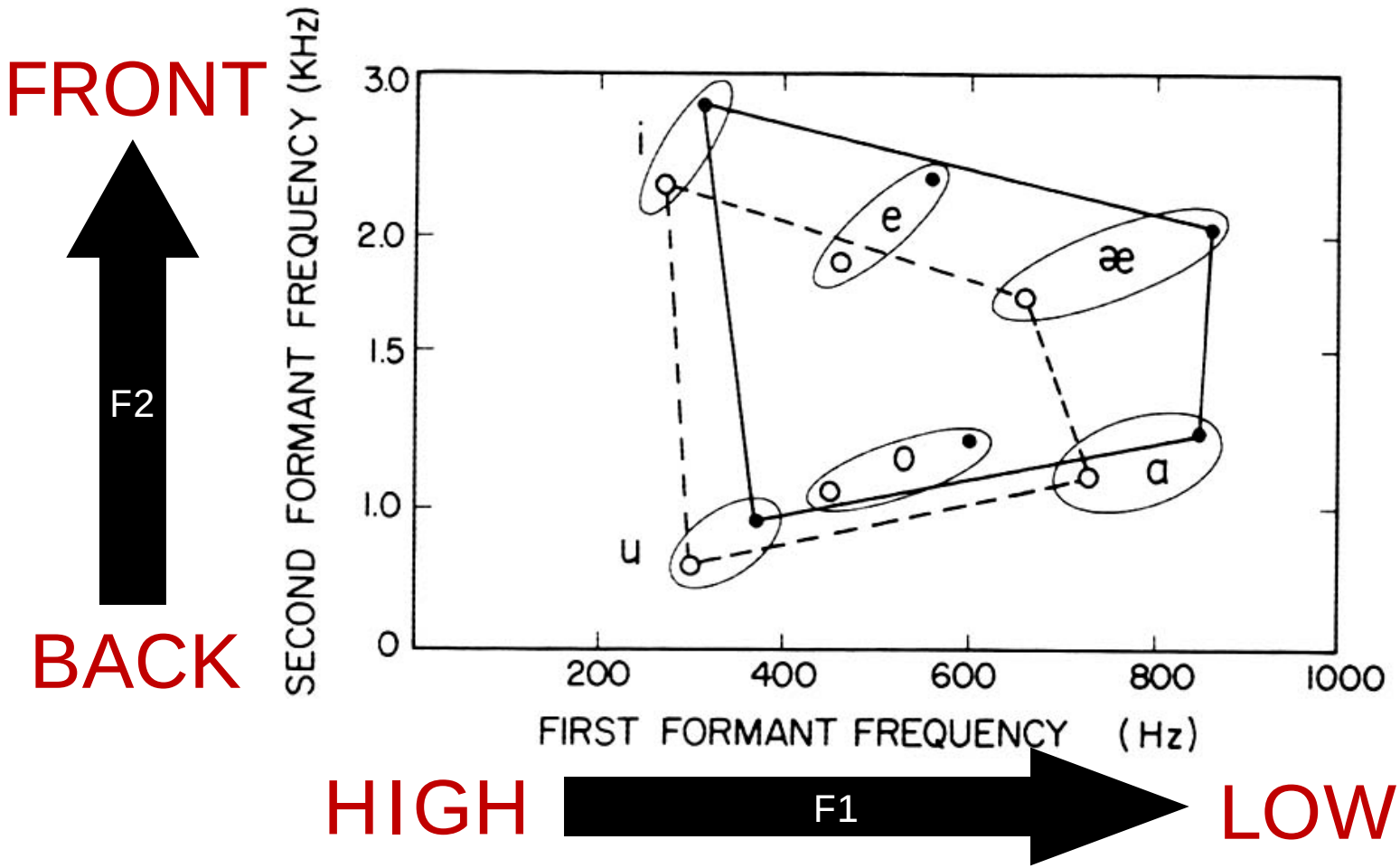
Tongue Position

- move in 2-D



- results in deviation of F1 and F2 from the neutral position

Vowel Charts



Picture from
Stevens 1999

Average Values for Basic American English Vowels

Male

Vowel	F1 (Hz)	F2 (Hz)
i	270	2290
e	460	1890
x	660	1720
a	730	1090
o	450	1050
u	300	870

Female

Vowel	F1 (Hz)	F2 (Hz)
i	310	2790
e	560	2320
x	860	2050
a	850	1220
o	600	1200
u	370	950

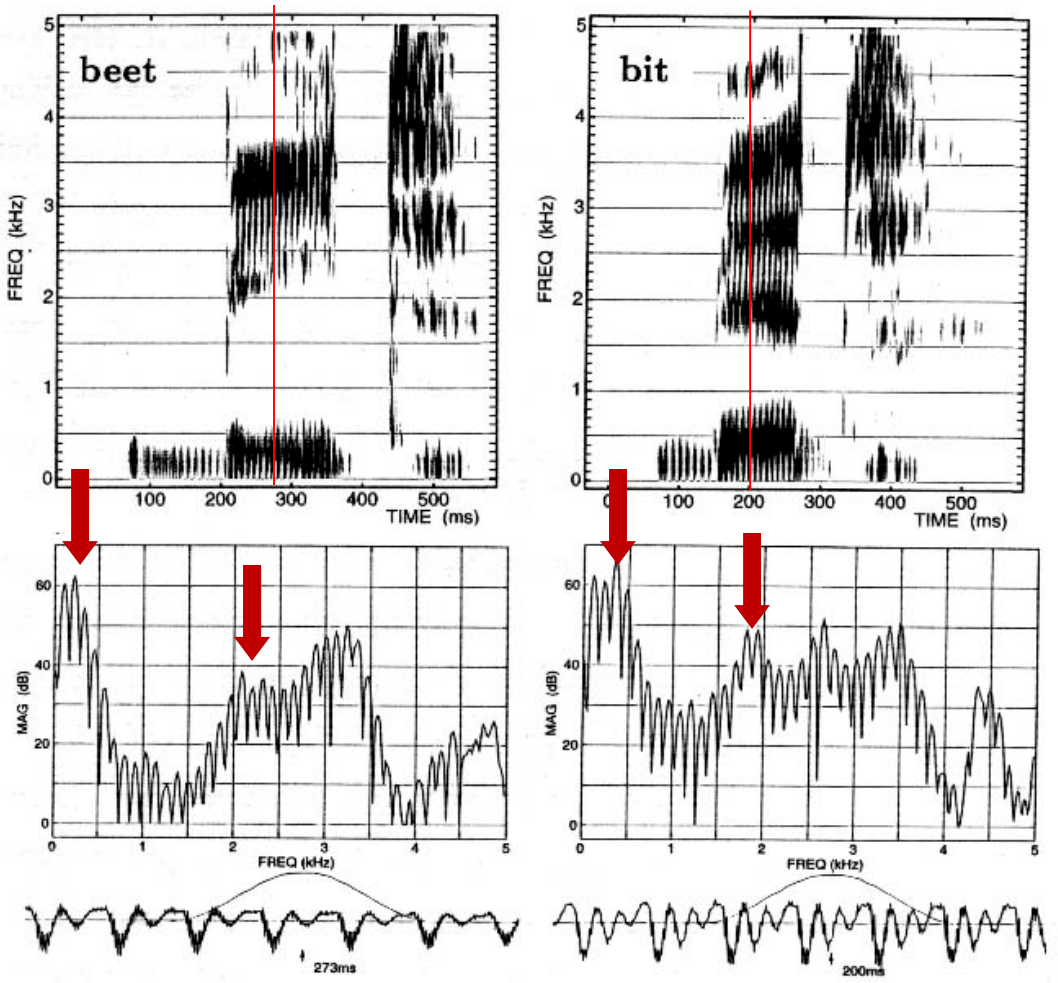
After Stevens 1999

Tense-Lax

- lax → smaller degree of constriction
- lax → shorter in duration
- F1, F2 move away from the corners closer to the neutral position

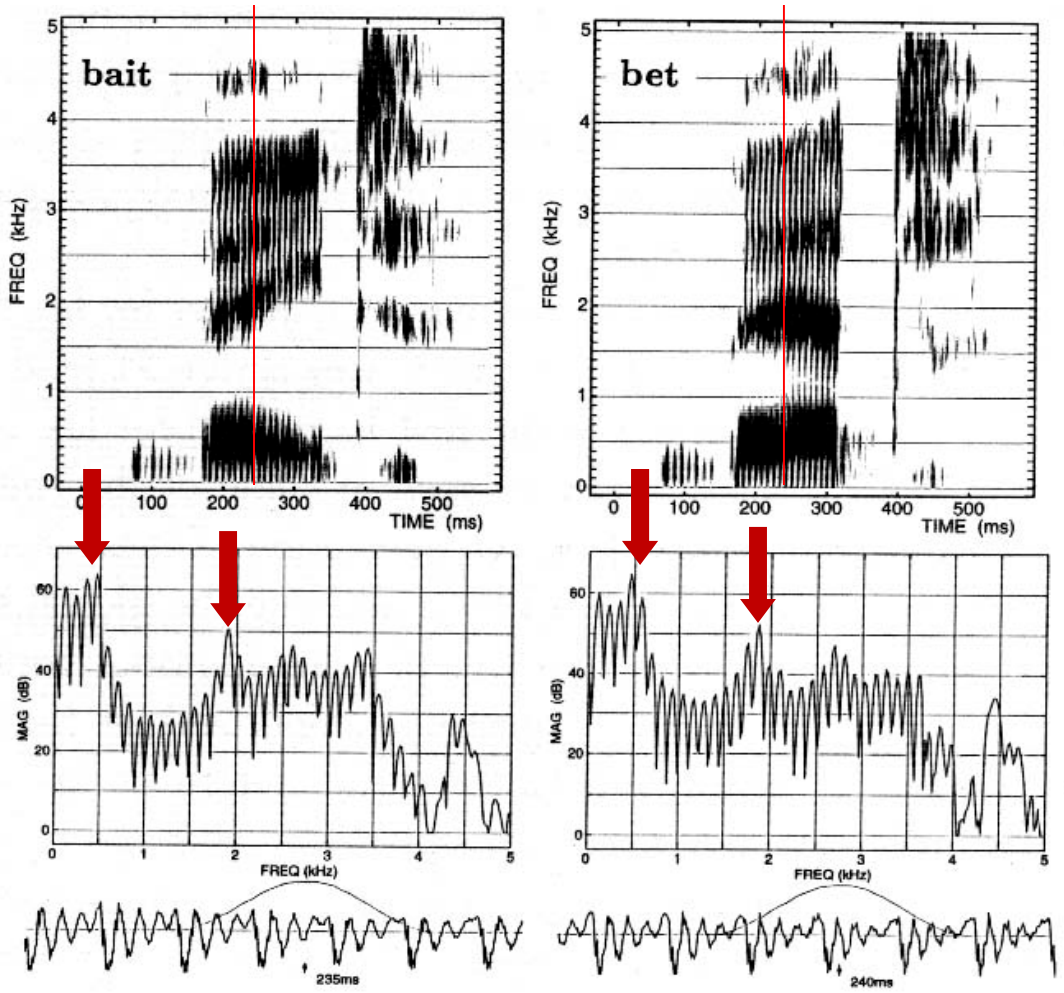
- <u>examples of tense-lax vowel pair</u>
  - beet (b-ii-t^) vs. bit (b-i-t^)
  - bait (b-ee-t^) vs. bet (b-e-t^)
  - คาด (kh-aa-t^-2) vs. คัด (kh-a-t^-3)
  - รูด (r-uu-t^-2) vs. รุด (r-u-t^-3)

beet vs. bit



Pictures from
Stevens 1999

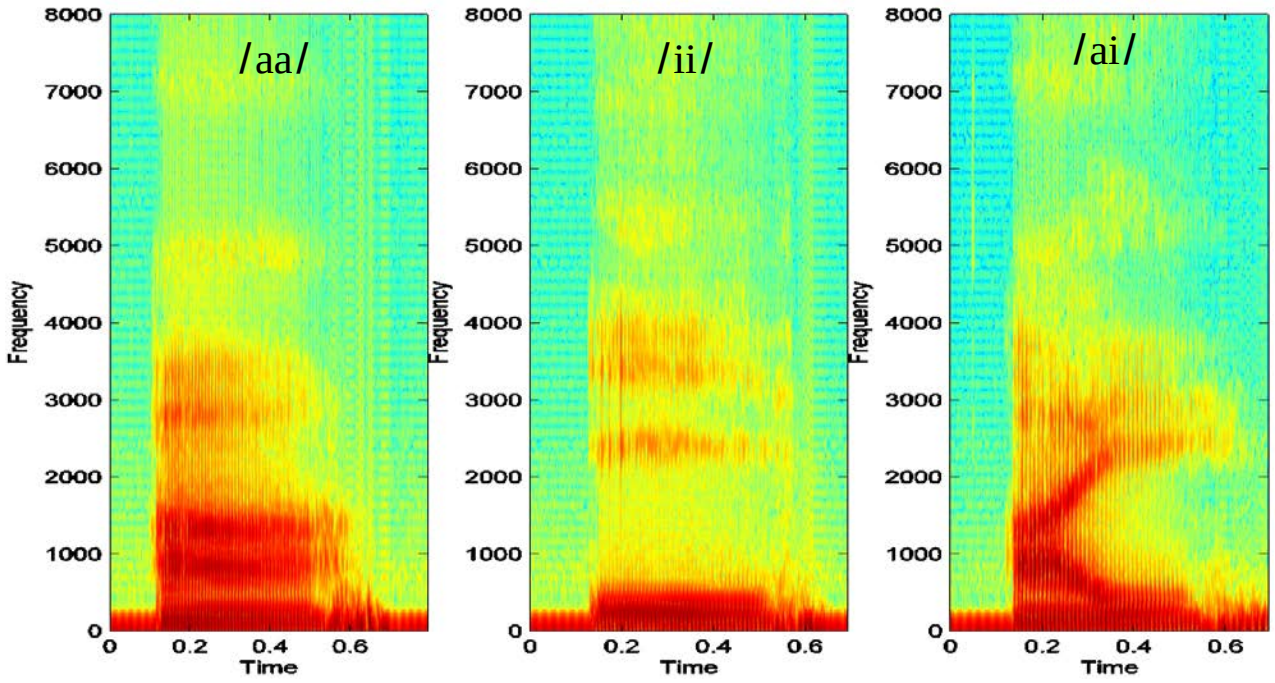
bait vs. bet



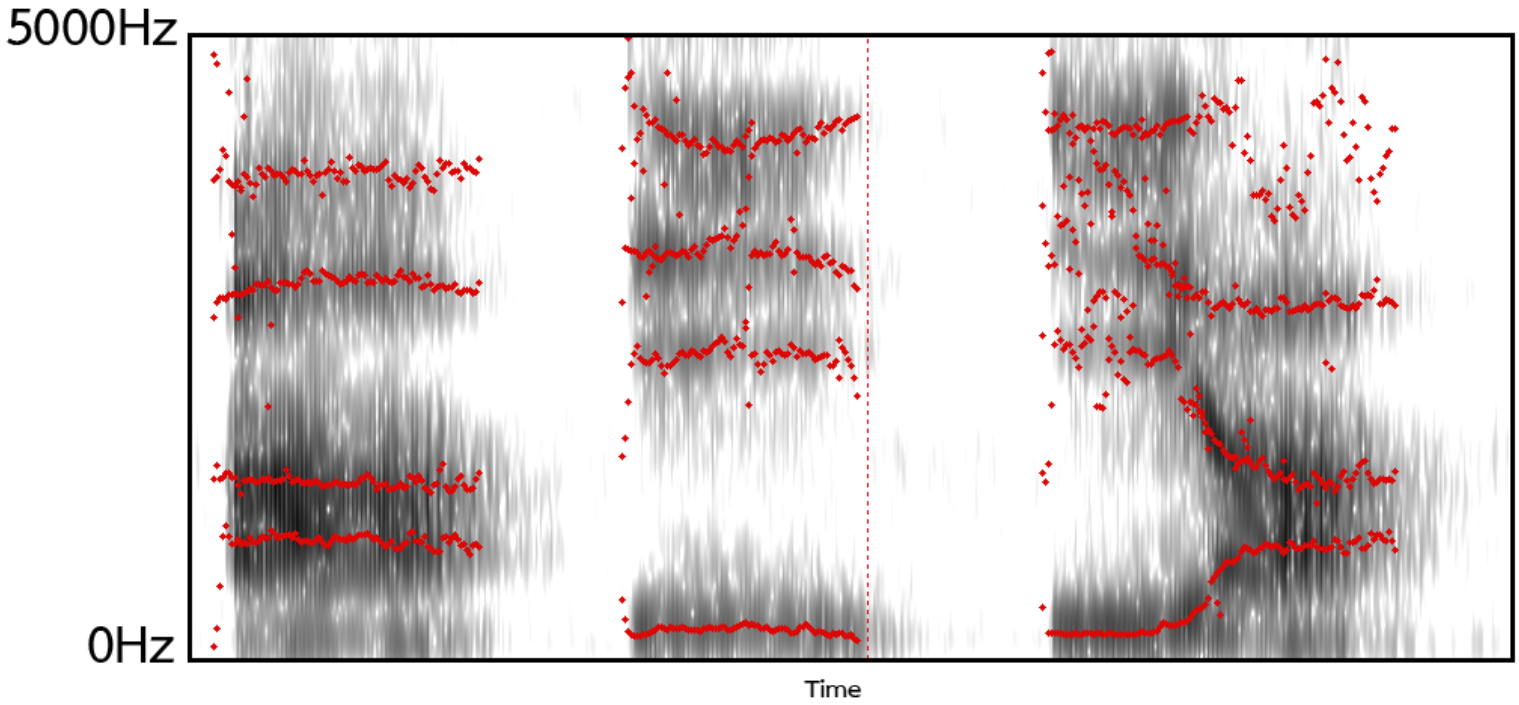
Pictures from
Stevens 1999

Diphthongs

- combination of 2 single vowels
- smooth movement in formant frequencies from one vowel to the other



Dipthongs



Thai Vowels

Single Vowels

a	อะ
aa	อา
i	อิ
ii	อี
v	อึ
vv	อื
U	อุ
uu	อู
e	เอะ
ee	เอ
x	แอะ
xx	แอ

o	โอะ
oo	โอ
@	เอาะ
@@	ออ
q	เออะ
qq	เออ

Dipthongs

ia	เอียะ
iia	เอีย
va	เอือะ
vva	เอือ
ua	อัวะ
uua	อัว

Consonants

- classified based on “manner of articulation”
  - fricative consonants (eg. /f/ ฟาง, fan)
  - stop consonants (eg. /b/ บาง, ban)
  - affricates (eg. /c/ จาง, January)
  - nasal consonants (eg. /m/ มา, man)

Consonants

Fricatives

f	ฝน, ฟัน
s	สาย, ศิลา,รักษา, ซ่อน

Affricates

ch	ชอบ, เฌอ

Stops

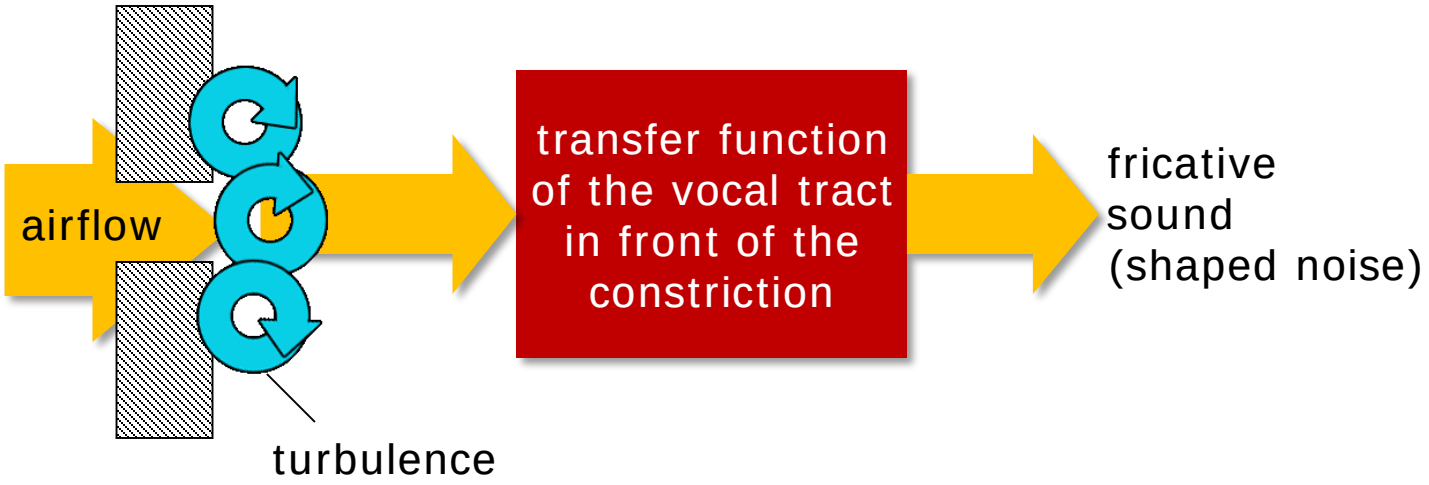
p	ปาก
t	เต้น, กุฎิ
c	จะ
k	ก่อน
ph	พบ, ภัย, ผ่าน
th	ทิ้ง, ธง, เฒ่า, ฐาน, มณโฑ
kh	คน, เขิน, ฆ่า
b	บอก
d	ด้าน, ชฎา

Nasals

m	ไม่
n	นาน, เณร
ng	เงิน

Fricative Consonants

- Narrow constriction at some point along the vocal tract
- generate turbulence noise in the vicinity of the constriction



(radiation characteristic does not depend on the vocal tract shape, so it can be included into the source)

Labial Fricatives

- constriction at the lips
- virtually no tube in front of the turbulence noise
- output signal is approximately the turbulence noise
- usually weaker than other fricatives
- Thai Labial fricatives sounds
  - /f/ in fan,ฟาง

Voiced-unvoiced fricative

- The vocal folds are vibrating (slack), while the air flow through the narrow constriction is maintained → "voiced"
- The vocal folds are not vibrating (strict), while the air flow is maintain → "unvoiced" or "voiceless"

voiced labial fricative → van

voiceless labial fricative → fan, ฟัน

Alveolar Fricatives

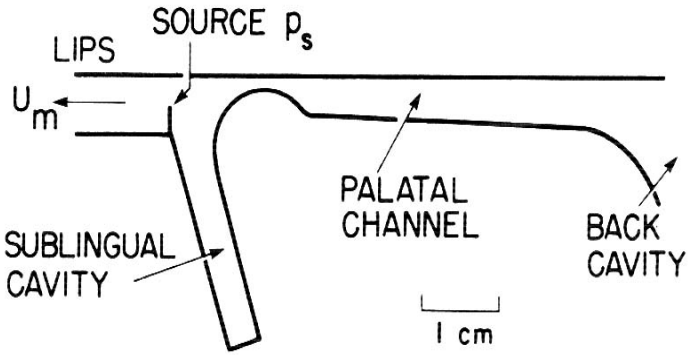
- constriction between the tongue blade and the roof of the mouth

voiced alveolar fricative → zip

voiceless alveolar fricative → sip, สาน

Palatal Fricatives

- Point of constriction is a few mm. posterior to the alveolar ridge
- The tongue blade is shaped in such a way as to produce a long and narrow channel behind the point of maximum constriction.



voiced palatal fricative → Gigi

voiceless palatal fricative → shy

Dental Fricatives

- placing tongue between upper and lower teeth

voiced dental fricative → <u>th</u>at

voiceless dental fricative → <u>th</u>ief

Experiments

- Waveform of Fricatives
- Spectrograms of
  - f-aa-z^-0
  - s-aa-z^-0
  - sh-aa-z^-0 (English /sh/)
  - s-ii-z^-0
  - f-ii-z^-0
  - sh-ii-z^-0 (English /sh/)

Thai Fricatives

f	ฝน, ฟัน
s	สาย, ศิลา,รักษา, ซ่อน

Stop Consonants

- make complete closure in the oral cavity while maintaining the air flow from the lungs
- pressure behind the closure increases
- promptly release the closure (might generate the turbulence noise at the just-released closure → release burst)
- during the beginning of the closure phrase,
  - vocal folds vibrate → voiced
  - vocal folds do not vibrate → voiceless

Place of Articulation

- closure can be made at:
  - labial → Labial stop consonant
  - alveolar ridge+tongue tip → Alveolar stop consonant
  - hard palate+tongue body → Velar stop consonant
- The spectral shape of the release burst for labial and alveolar can be explained in the same way as the spectral shape of the fricative consonant.
  - labial fricative → labial stop release burst
  - alveolar fricative → alveolar stop release burst
- For Velar , the portion of the vocal tract in front of the closure gives mid-freq. resonance.

Aspiration

- After the release of the closure of a voiceless stop, if the glottis is widely spread, the air flow rush through the glottis will cause turbulence noise at the glottis.
- spread glottis → aspirated stop consonant
- otherwise → unaspirated stop consonant

Stop Consonants

voiced labial stop → /b/ in bus, เบา

voiceless unaspirated labial stop → /p/ in spin, ปืน

voiceless aspirated labial stop → /ph/ in pen, พาน

voiced alveolar stop → /d/ in den, เดา

voiceless unaspirated alveolar stop → /t/ in star, ตอน

voiceless aspirated alveolar stop → /th/ in ten, ทาน

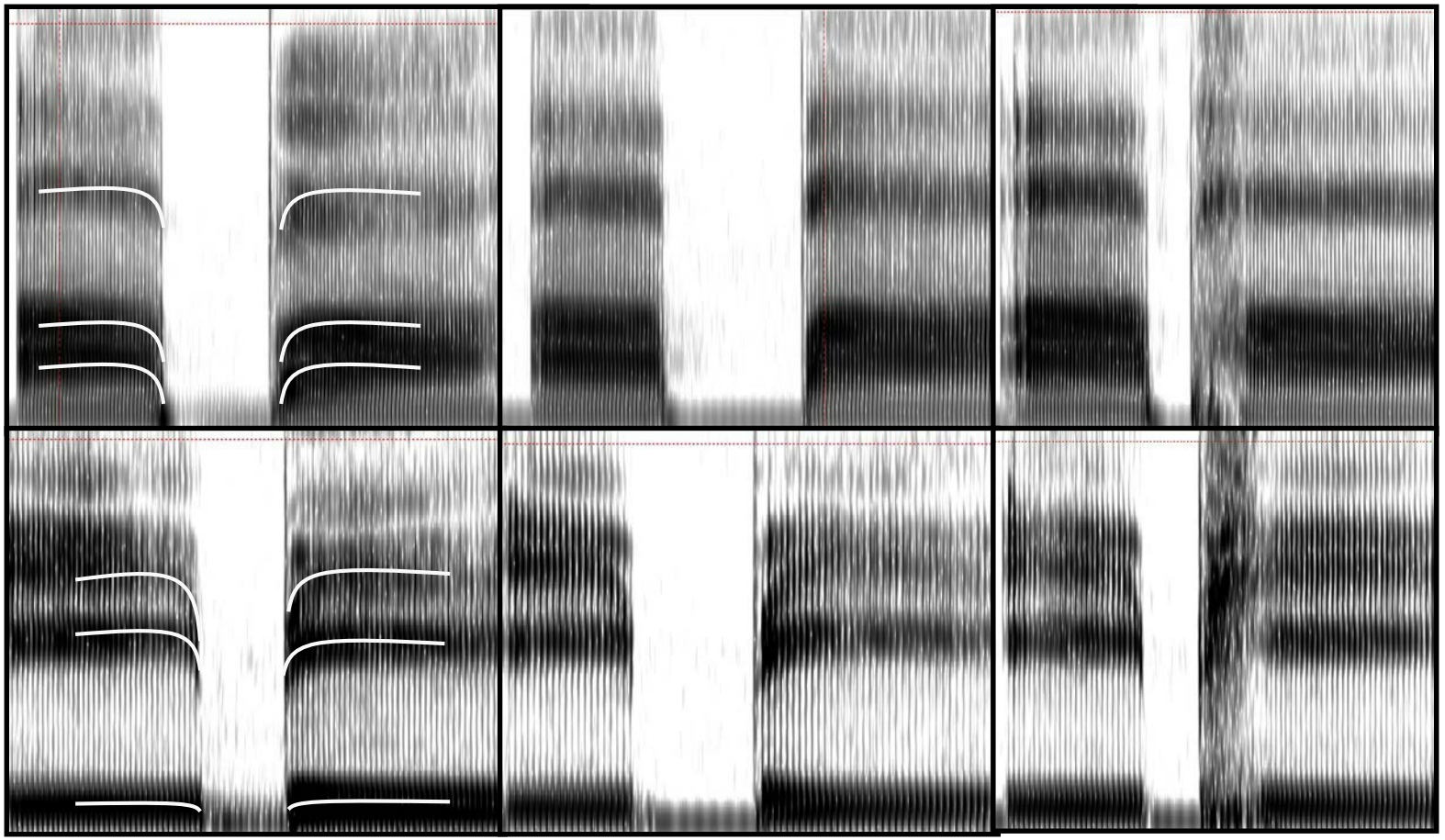
voiced velar stop → /g/ in gun, เกา

voiceless unaspirated velar stop → /k/ in scar

voiceless aspirated velar stop → /kh/ in keep, คาน

Labial Stop Consonants

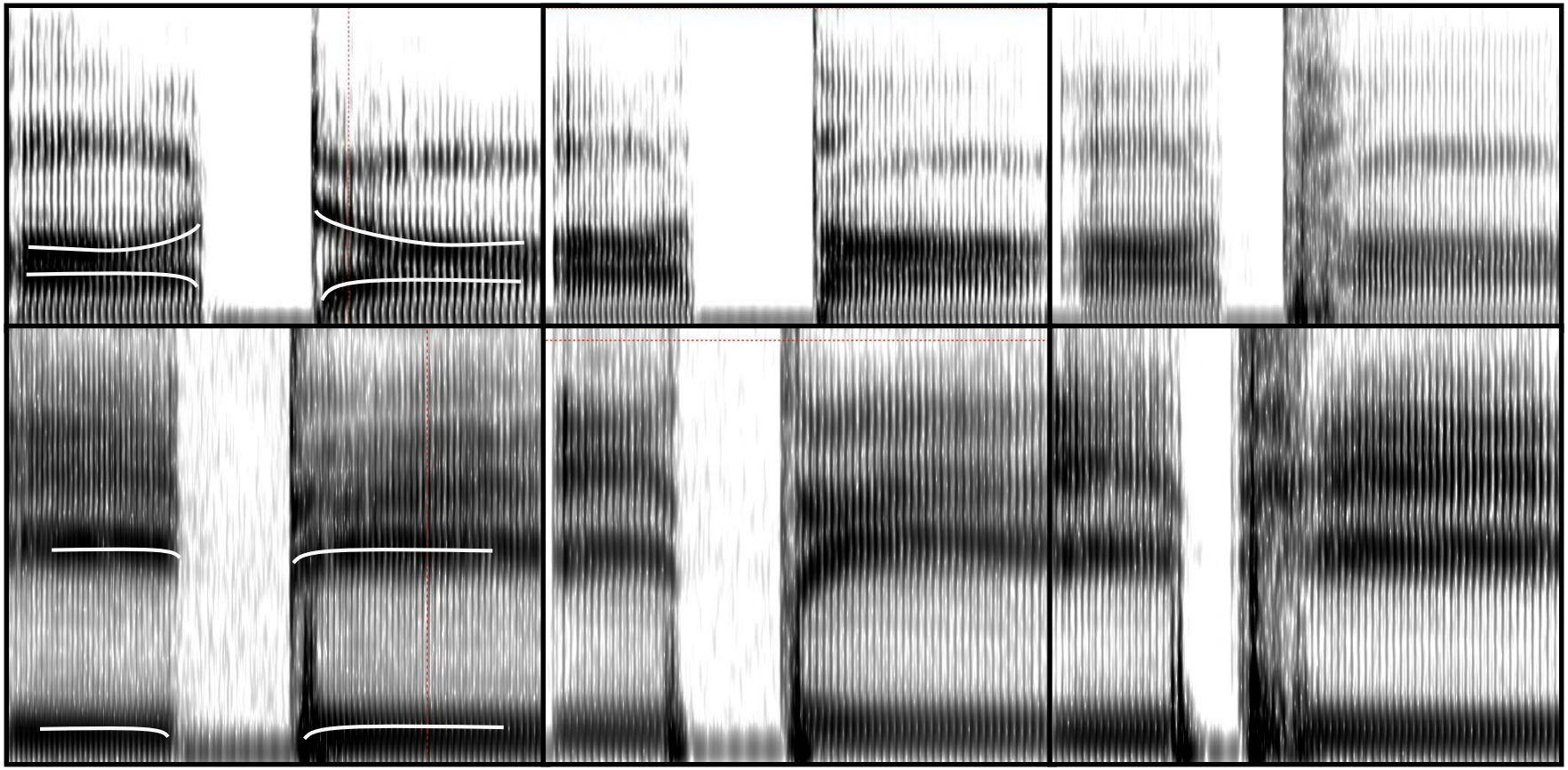
z-aa-z^ **b**-aa-z^ | z-aa-z^ **p**-aa-z^ | z-aa-z^ **ph**-aa-z^



z-ii-z^ **b**-ii-z^ | z-ii-z^ **p**-ii-z^ | z-ii-z^ **ph**-ii-z^

Alveolar Stop Consonants

z-aa-z^ **d**-aa-z^ | z-aa-z^ **t**-aa-z^ | z-aa-z^ **th**-aa-z^

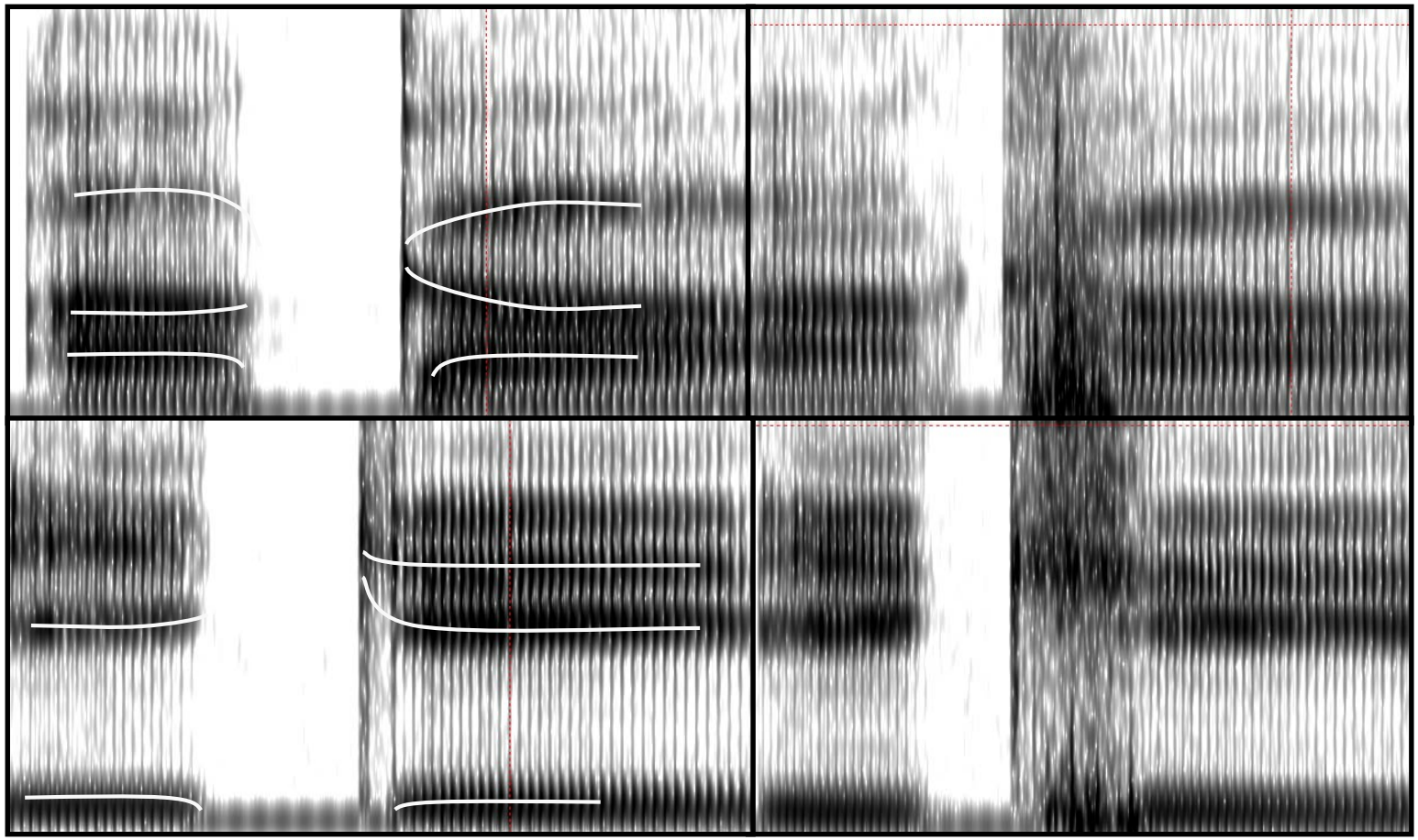


z-ii-z^ **d**-ii-z^ | z-ii-z^ **t**-ii-z^ | z-ii-z^ **th**-ii-z^

Alveolar Stop Consonants

z-aa-z^ **k**-aa-z^

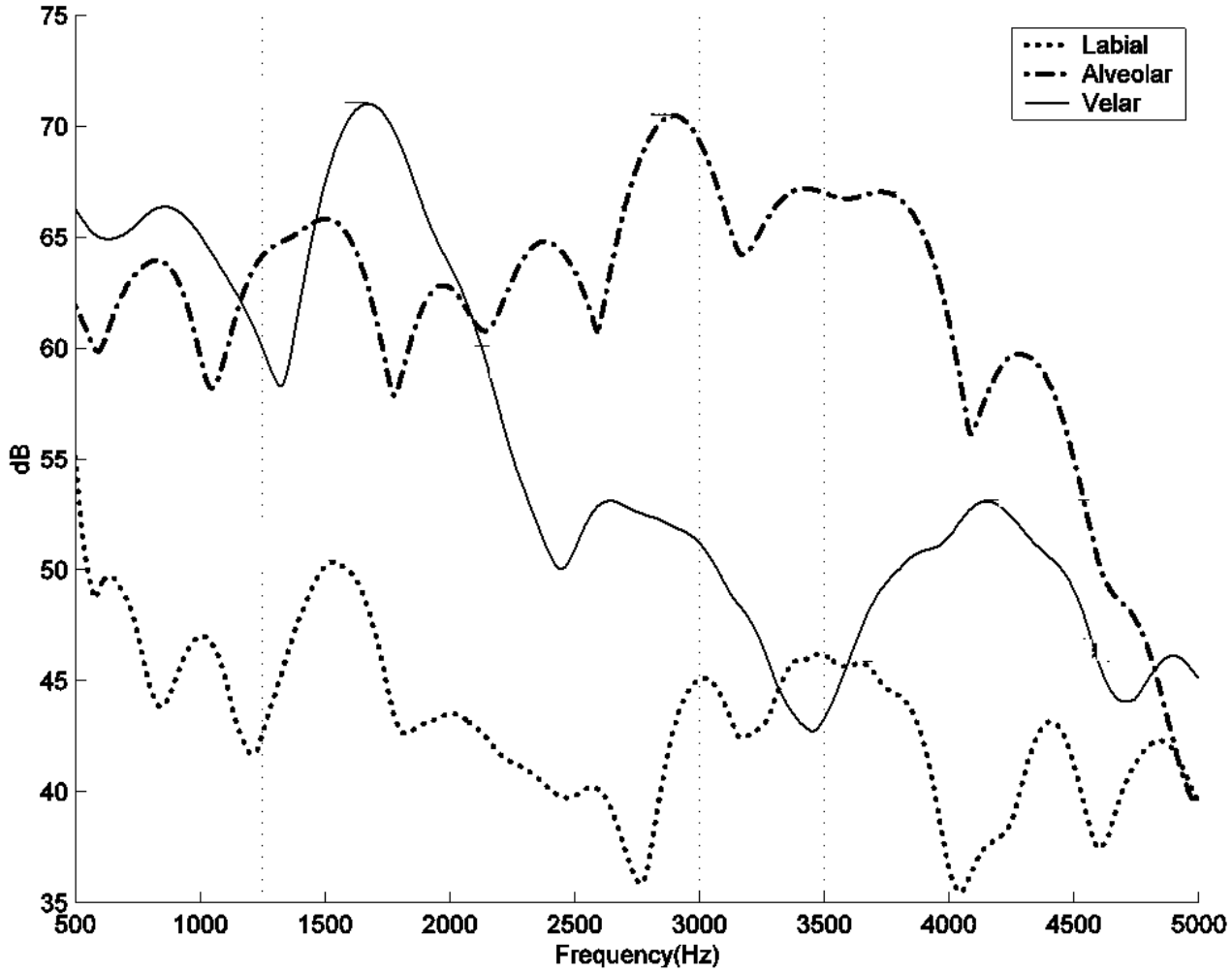
z-aa-z^ **kh**-aa-z^



z-ii-z^ **k**-ii-z^

z-ii-z^ **kh**-ii-z^

Burst Spectra



Picture from Suchato 2004

Affricates

- make complete closure like stop consonants
- release the closure and generate the turbulence noise like fricatives

voiceless palatoalveolar affricate →/ch/ in church, ชาน

voiced palatoalveolar affricate → /c/ in judge , จิก