AUTOMATIC SPEECH RECOGNITION

Lecture 5

Sounds in Languages

Last Lecture

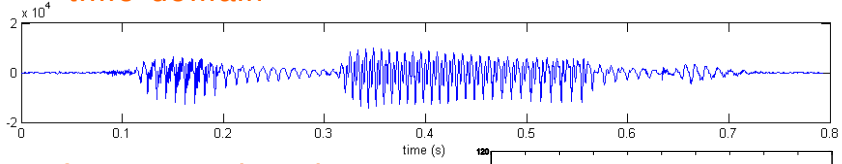
- Physiology of human speech production
- Source-filter model
- glottal source / noise source
- Radiation characteristic
- Vocal tract transfer function
  - Uniform tube
  - Concatenated tube
  - Perturbation theory

This Lecture

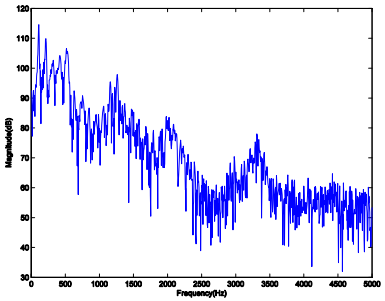
- Spectrogram
  - helps visualizing sounds
- Classes of sound (vowel, consonant)
  - production
  - types
  - acoustic evidence

Sound Analysis

- time-domain



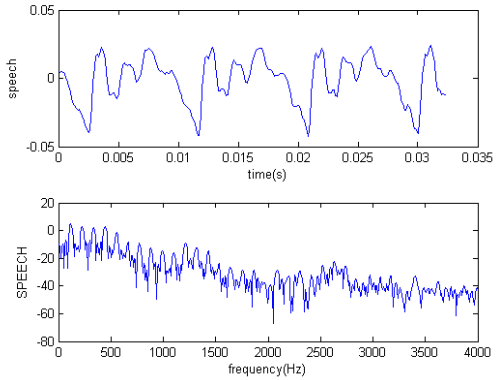
- frequency-domain
  -Fourier transform (discrete-time)



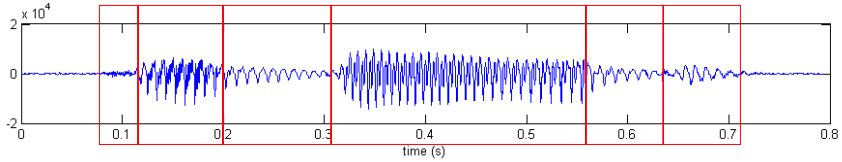
Speech is non-stationary and time-varying signal. So, analyzing speech by lumping all time points together is not very useful.

Understanding Frequency Components

Speech signal has a lot more frequency components



Time-varying characteristic



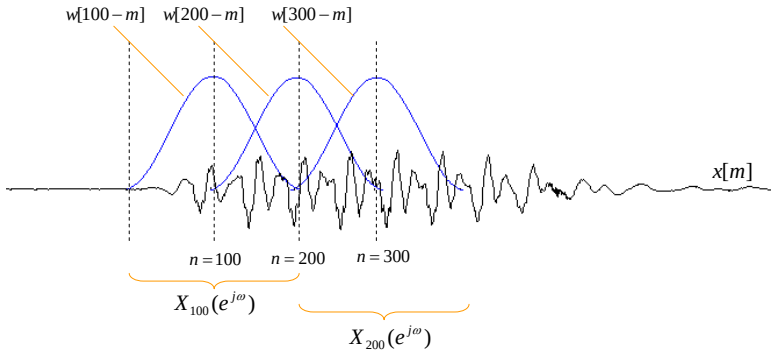
If we find the Fourier transform of the whole signal, we will see the frequency-domain characteristics of the whole signal.

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

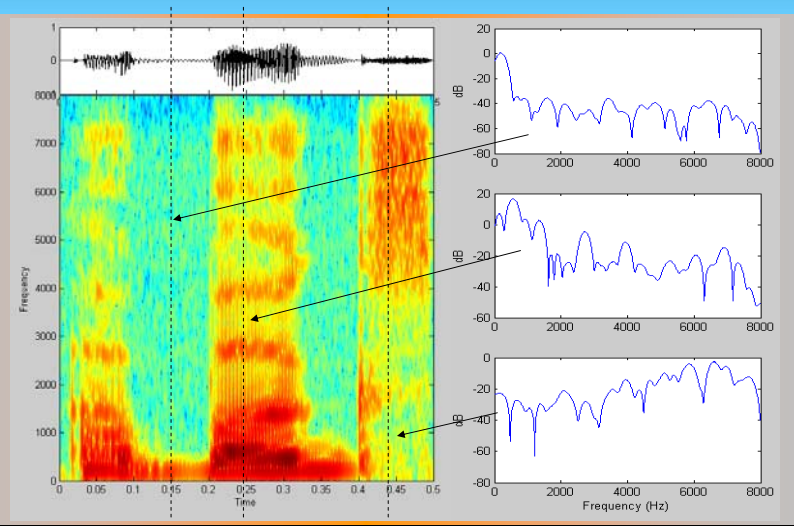
Sound Analysis

- Short-time Fourier Analysis

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



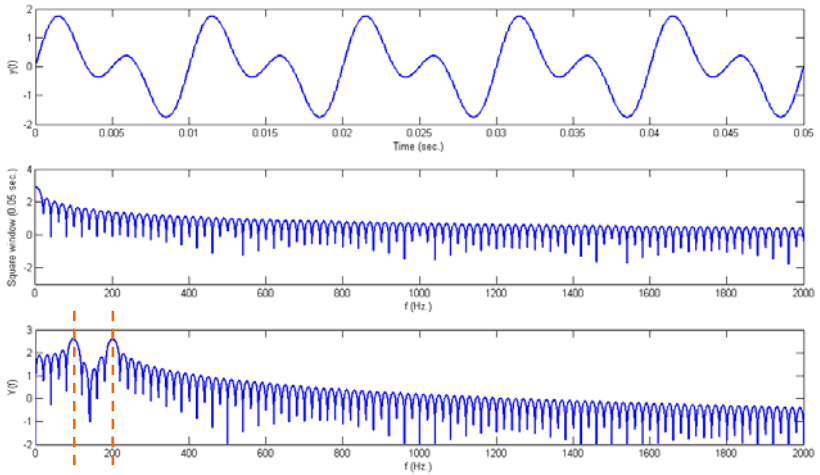
Spectrogram



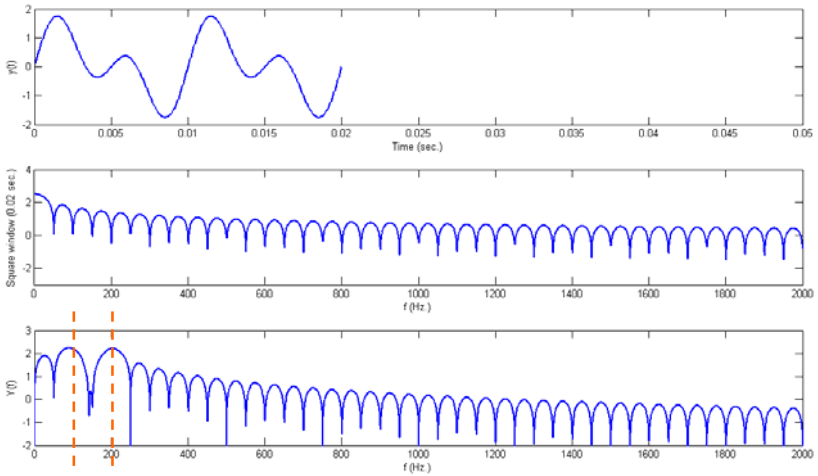
Time vs. Frequency Resolution

- 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

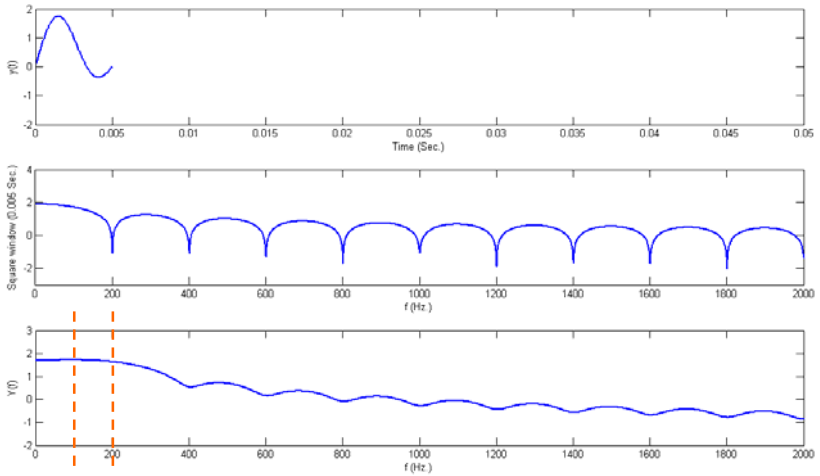
Window Length: Revisited



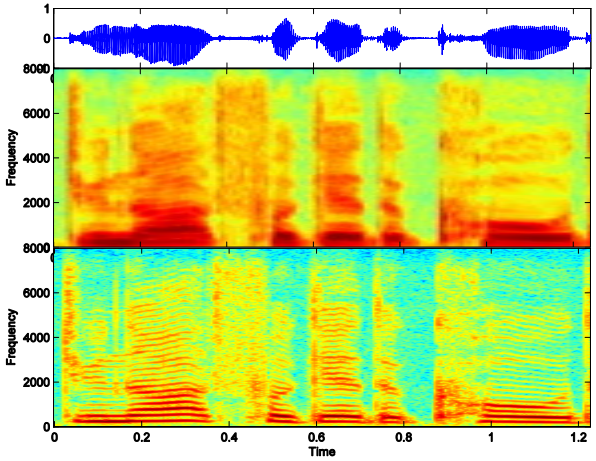
Window Length: Revisited



Window Length: Revisited



Wide-band / Narrow-band Spectrogram



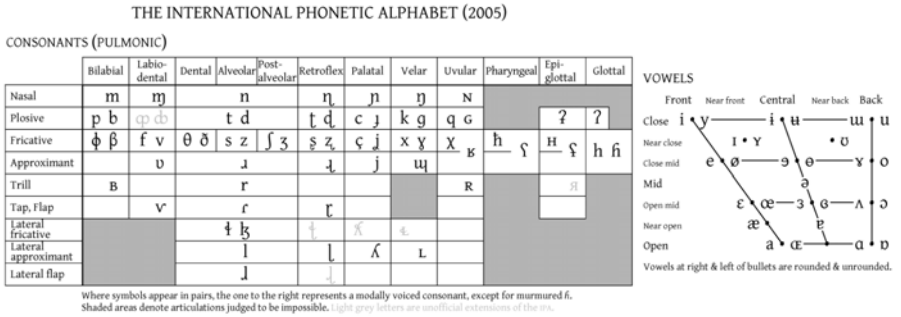
Wide-band
- good time resolution
- can see pitch structure
- can see abrupt acoustic event

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

Classes of Sound

- 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

IPA


THE INTERNATIONAL PHONETIC ALPHABET (2005)

CONSONANTS (PULMONIC)

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Epi-glottal	Glottal
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	N			
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʡ	ʔ
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	ʜ ʢ	h ɦ
Approximant		ʋ		ɹ		ɻ	j	ɰ				
Trill	B			r					R		ʀ	
Tap, Flap		ⱱ		ɾ		ɽ						
Lateral fricative				ɬ ɮ								
Lateral approximant				l		ɭ	ʎ	ʟ				
Lateral flap				ɺ								

Where symbols appear in pairs, the one to the right represents a modally voiced consonant, except for murmured ɦ.
Shaded areas denote articulations judged to be impossible. Light grey letters are unofficial extensions of the IPA.

VOWELS

A Note on Phonetic Symbols

- For our convenience in this class (avoiding problems with fonts), we will not use the standard International Phonetic Alphabets (IPA) as the default representation of sounds.
- We will use a set of notation presented in the next slides to represent Thai sounds.
- Note that notations used for some English sounds discussed in this lecture are not standard. (not IPA)

Sounds in Thai

p	ปาก
t	เต้น, กฏิ
c	จะ
k	ก่อน
z	อาน
ph	พบ, ภัย, ผ่าน
th	ทิ้ง, ธง, เต่า, ฐาน, มณโฑ
ch	ชอบ, เฌอ
kh	คน, เขิน, ฆ่า
b	บอก
d	ด้าน, ชฎา

m	ไม่
n	นาน, เณร
ng	เงิน
l	เล่น, กีฬา
r	รอ, ฤทัย
f	ฝน, ฟัน
s	สาย, ศิลา,รักษา, ซ่อน
h	โหน, เฮฮา
w	ว่า
j	ย้อน, หญิง

pr	ประสาน
phr	พราน
tr	เตรียม
kr	กราบ
khr	คร่า
pl	ปลา
phl	พลาด
thr	จันทรา
kl	เกลอ
khl	เคลื่อน
kw	กวาง
khw	ขวา

br	เบรน
bl	บลู
fr	ฟราย
fl	เฟลม
dr	ดรากอน

Sounds in Thai

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

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

p^	พบ
t^	เกร็ด
k^	ปาก
n^	หาร
m^	ลม
ng^	ฟาง
j^	ยาย
w^	กาว

f^	กราฟ
l^	แอล
s^	เอส
ch^	คลัช

Sounds in Thai

- กนกวรรณกร
- วชิรญาณวโรรส
- ฮกเฮงล้ง
- แป๊ะเจี๊ยะ
- แอสเซมบลี
- โปรเทคเตอร์

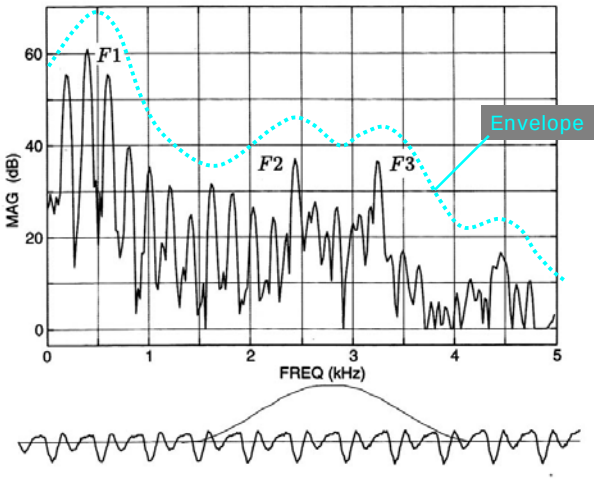
Classes of Sound

- Vowels
- Consonants
- Semi-vowels

Vowel

- 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

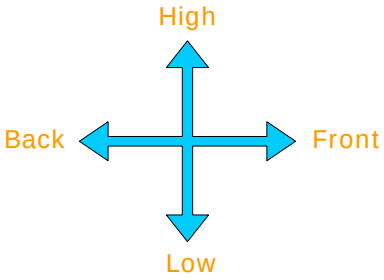
Vowel Spectrum



Picture from Stevens 1999

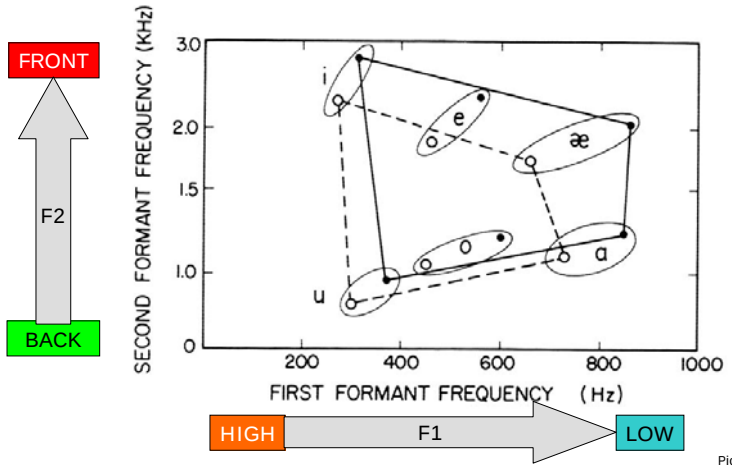
Tongue Position

- move in 2-D



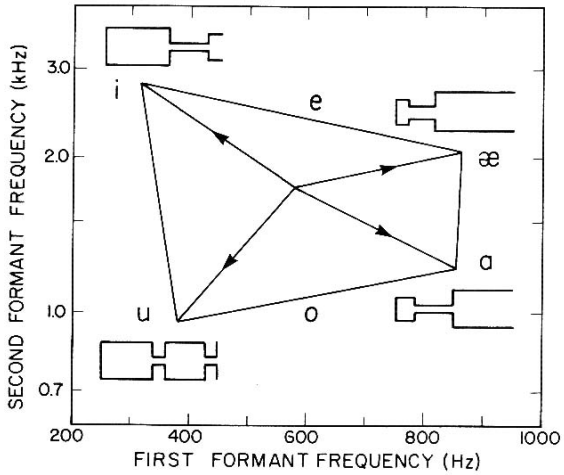
- results in deviation of F1 and F2 from the neutral position

Vowel Chart



Picture from Stevens 1999

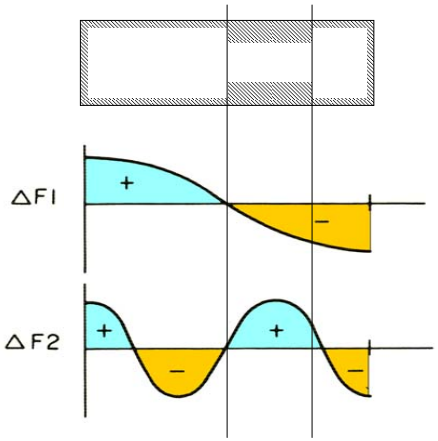
Perturbation for the Cardinal Vowels



Picture from Stevens 1999

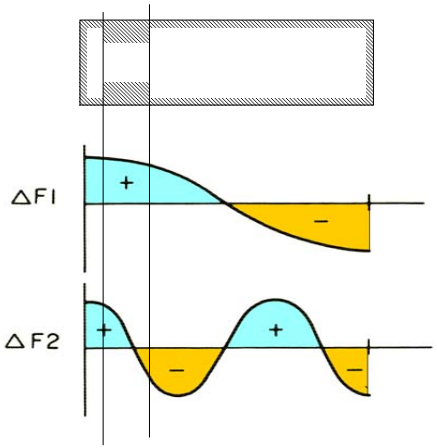
Perturbation in /ii/

- /ii/
- Constriction in the palatal region
- Result:
  - Lo F1
  - Hi F2



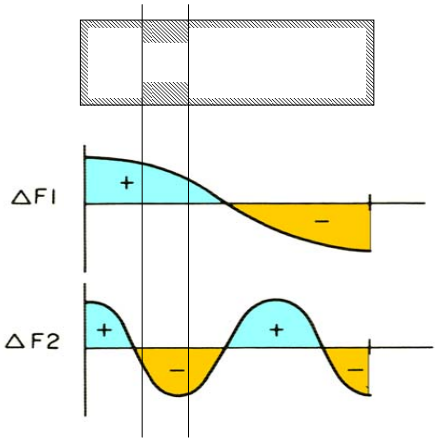
Perturbation in /xx/

- /xx/
- Constriction in the lower pharyngeal region
- Result:
  - Hi F1
  - Hi F2



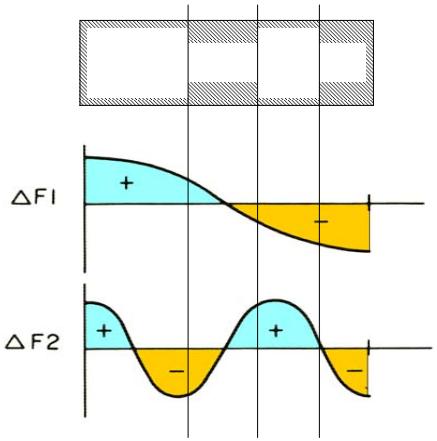
Perturbation in /aa/

- /aa/
- Constriction in the lower pharyngeal region (above /xx/)
- Result:
  - Hi F1
  - Lo F2



Perturbation in /uu/

- /uu/
- Constriction in the vicinity of the soft palate
- Rounding of the lips
- Result:
  - Lo F1
  - Lo F2



Average Values for Basic American English Vowels

Male			Female		
Vowel	F1 (Hz)	F2 (Hz)	Vowel	F1 (Hz)	F2 (Hz)
i	270	2290	i	310	2790
e	460	1890	e	560	2320
x	660	1720	x	860	2050
a	730	1090	a	850	1220
o	450	1050	o	600	1200
u	300	870	u	370	950

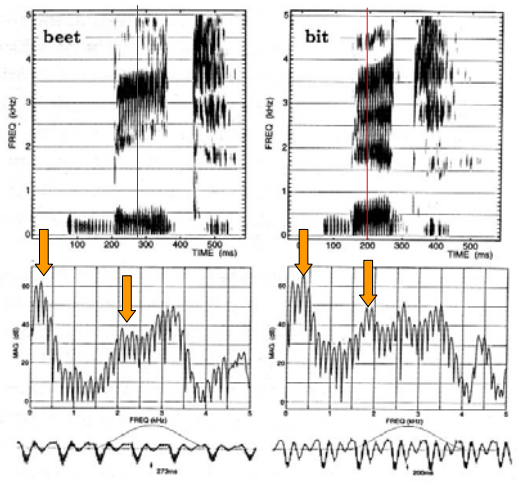
After Stevens 1999

Tense-Lax

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

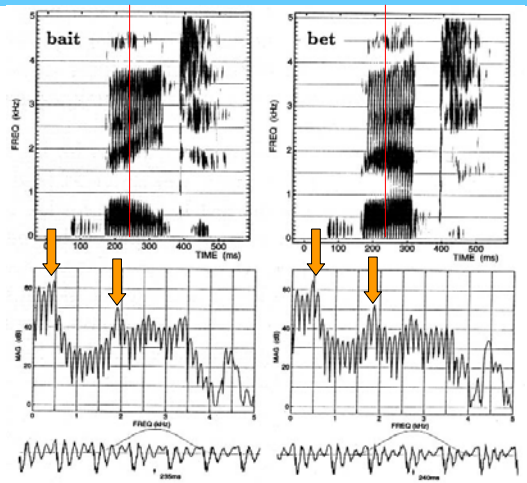
- examples of tense-lax vowel pair
  - beet (/b ii t^/) vs. bit (/b i t^/)
  - bait (/b ee t^/) vs. bet (/b e t^/)
  - คาด (/kh aa t^/) vs. คัด (/kh a t^/)
  - รูด (/r uu t^) vs. รุด (/r u t^/)

beet vs. bit



Pictures from Stevens 1999

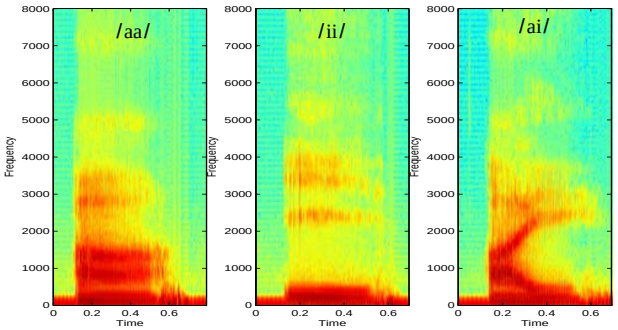
bait vs. bet



Pictures from
Stevens 1999

Diphthongs

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



Vowels in Thai Language

- 18 single vowel sounds (9 tense + 9 lax)

lax	tense
a (อะ)	aa (อา)
i (อิ)	ii (อี)
v (อึ)	vv (อื)
u (อุ)	uu (อู)
x (แอะ)	xx (แอ)
e (เอะ)	ee (เอ)
o (โอะ)	oo (โอ)
@ (เอาะ)	@@ (ออ)
q (เออะ)	qq (เออ)

Vowels in Thai Language

- 9 diphthongs

lax	Tense
ia (เอียะ)	iia (เอีย)
va (เอือะ)	vva (เอือ)
ua (อัวะ)	uua (อัว)
	ai (ไอ) (a j^)
	au (เอา) (a w^)
	qqi (เอย) (qq j^)

Classes of Sound

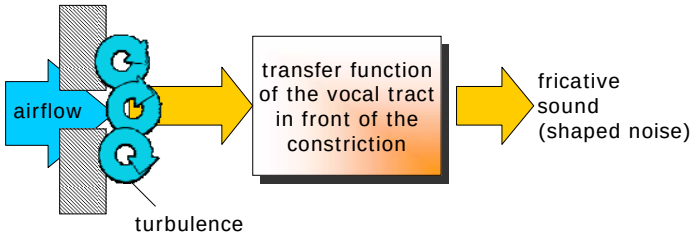
- Vowels
- Consonants
- Semi-vowels

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)

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
- Labial fricatives sounds
  - /f/ in fan,ฟาง
  - /v/ in van

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"

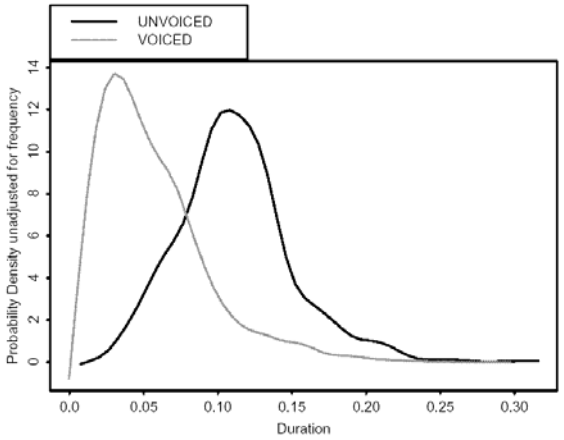
voiced labial fricative → /v*/ in van

voiceless labial fricative → /f/ in fan, ฟัน

Department of Computer Engineering, Chulalongkorn University
Spoken Language Systems Research Group

Voiced-unvoiced fricative

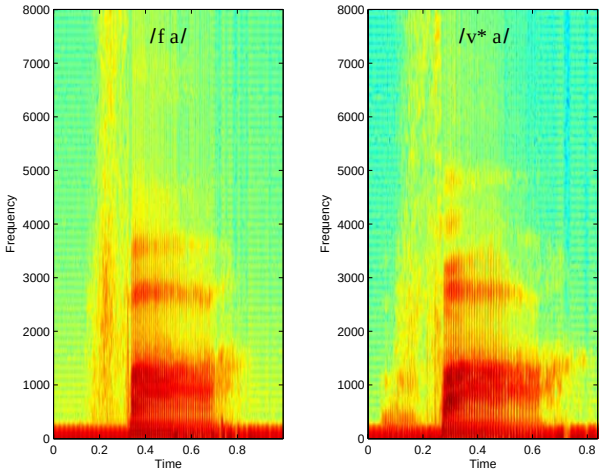
- A voiced fricative tends to be shorter than an unvoiced one.



Special Topics in Computer Science | First Semester 2008 | Lecture 5
ATIWONG SUCHATO

Department of Computer Engineering, Chulalongkorn University
Spoken Language Systems Research Group

Labial Fricatives



Special Topics in Computer Science | First Semester 2008 | Lecture 5
ATIWONG SUCHATO

Department of Computer Engineering, Chulalongkorn University
Spoken Language Systems Research Group

Alveolar Fricatives

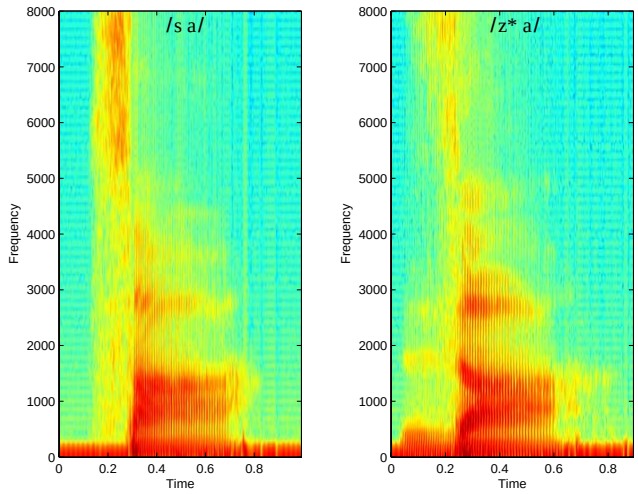
- constriction between the tongue blade and the roof of the mouth

voiced alveolar fricative → /z*/ in zip

voiceless alveolar fricative → /s/ in sip, สาน

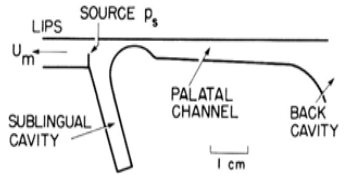
Special Topics in Computer Science | First Semester 2008 | Lecture 5
ATIWONG SUCHATO

Alveolar Fricatives



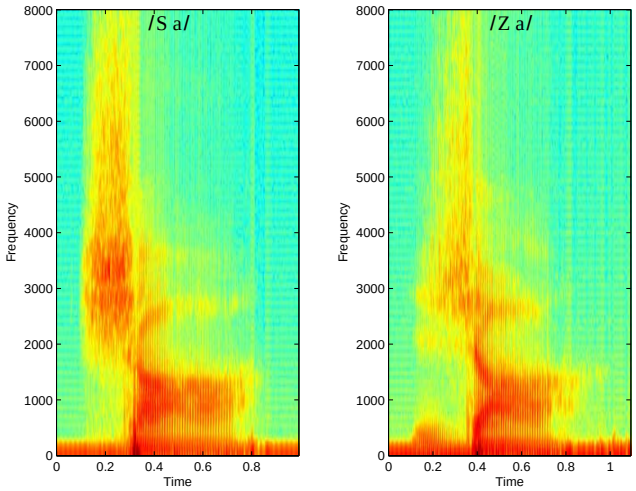
Palatal Fricatives

- 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 → /Z/ in Gigi
voiceless palatal fricative → /S/ in shy

Palatal Fricatives

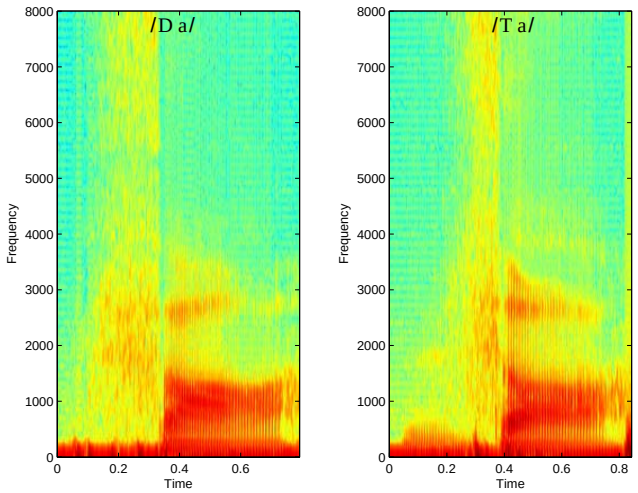


Dental Fricatives

- placing tongue between upper and lower teeth

voiced dental fricative → /D/ in that
voiceless dental fricative → /T/ in thief

Dental Fricatives



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 → /p^h/ in pen, พาน

voiced alveolar stop → /d/ in den, เดา

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

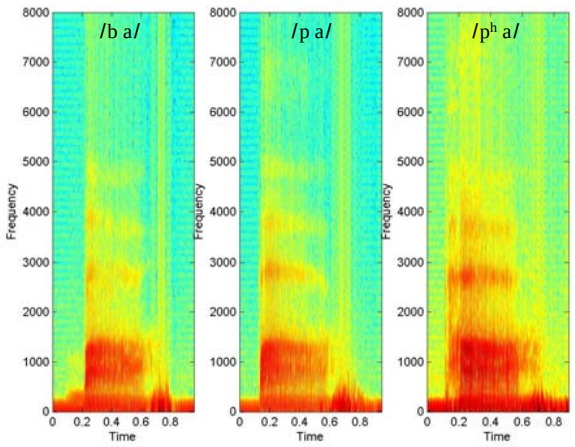
voiceless aspirated alveolar stop → /p^h/ in pen, พาน

voiced velar stop → /g/ in gun, เกา

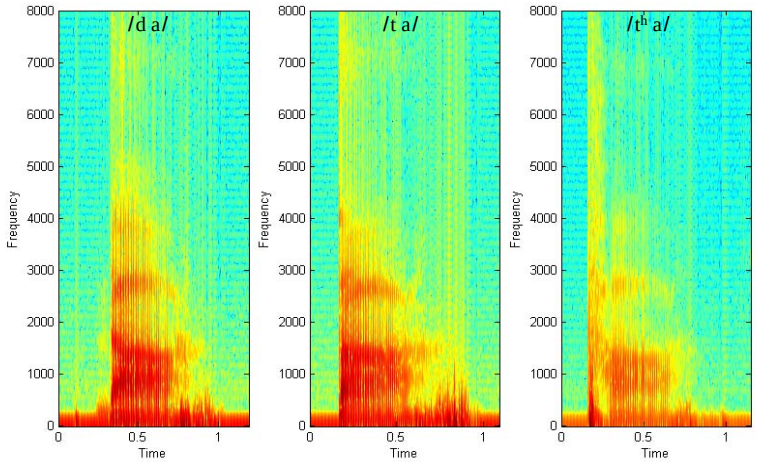
voiceless unaspirated velar stop → /k/ in scar

voiceless aspirated velar stop → /k^h/ in keep, คาน

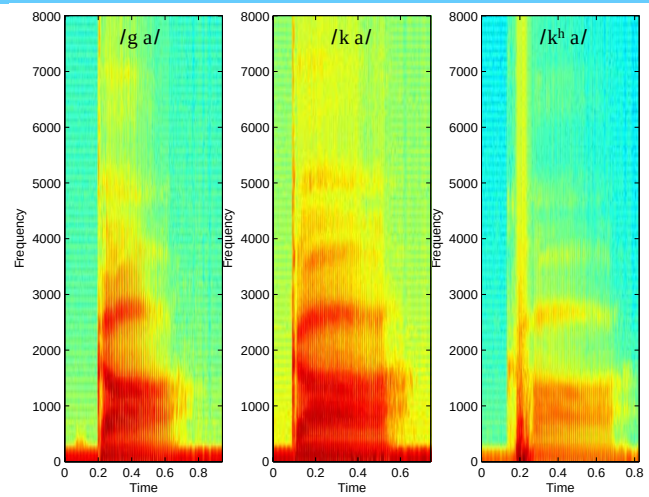
Labial Stop Consonants



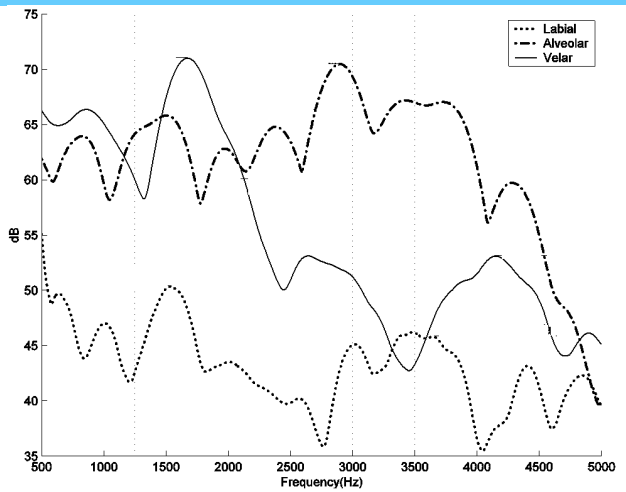
Alveolar Stop Consonants



Velar Stop Consonants



Burst Spectra



Picture from Suchato 2004

Affricates

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

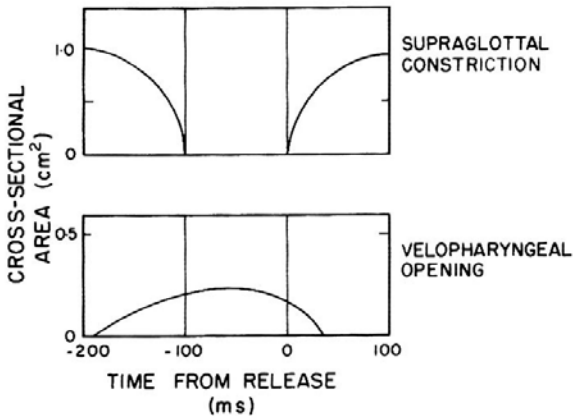
voiceless palatoalveolar affricate →/ch/ in church, ชาน

voiced palatoalveolar affricate → /c/ in judge , จิก

Nasal Consonants

- form a complete closure at some point along the oral cavity
- velopharyngeal port is open during the closure
- no pressure increase behind the constriction
- side branch introduce a pole/zero pair in the spectrum
- decrease in signal amplitude due to loss in the nasal cavity

Velopharyngeal port opening



Pictures from Stevens 1999

Nasal Consonants

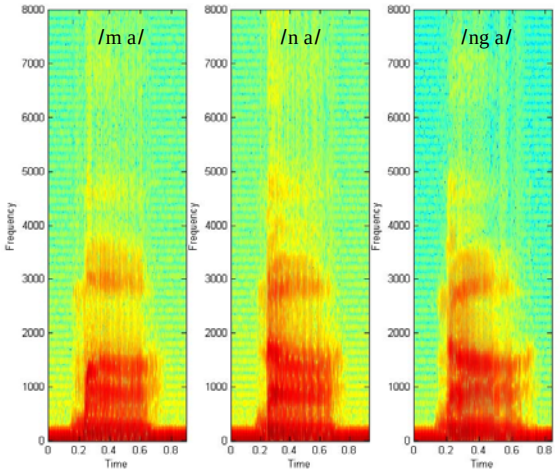
- Three places of articulation like stops

labial nasal → /m/ in man, มาก

alveolar nasal → /n/ in not, นาน

velar nasal → /ng/ in sing, เงา

Nasal Consonants



Consonant Summary

Place of Articulation

Manner of Articulation

	Labial	Dental	Alveolar	Palatal	Velar
Stop	p b		t d		k g
Fricative	f v	T D	s z	S Z	
Nasal	m		n		ng

voiceless voiced

Next Lecture

- Semi-vowel
- Spectrogram Reading