Chapter 3

Acoustic Theory of Speech Production
语音产生的声学理论

Outline

- Speech production mechanism
- Speech signal: waveforms and spectra
- Sounds of language => phonemes(音素)
- English speech sounds
- Initials(声母) and finals(韵母) of Mandarin(中文普通话)

Basic Speech Processes

- idea→sentences → words → sounds → waveform
  - Idea: it's getting late, I should go to lunch, I should call Al and see if he wants to join me for lunch today
  - Sentences/Words: Hi Al, did you eat yet?
  - Sounds: /h/ /ay/-/ae/ /l/-/d/ /ih/ /d/-/y/ /u/-/iy/ /t/-/y/ /ε/ /t/
  - Coarticulated Sounds: /h- ay-l/-/d-ih-j-uh/-/iy-t-j-ε-t/ (hial-dija-eajet)

Basic Speech Processes

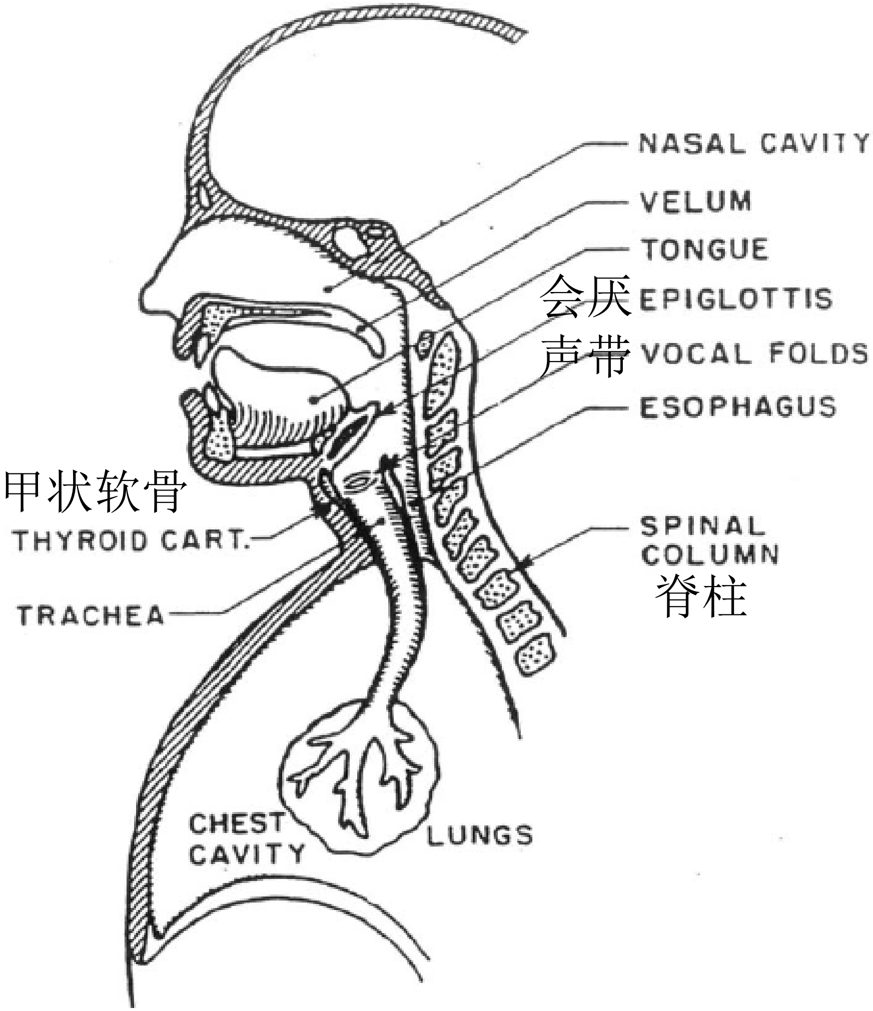
- remarkably, humans can decode these sounds and determine the meaning that was intended—at least at the idea/concept level (perhaps not completely at the word or sound level)
- often machines can also do the same task
  - speech coding: waveform →(model) → waveform
  - speech synthesis: words → waveform
  - speech recognition: waveform → words/sentences
  - speech understanding: waveform → idea

Basics

- speech is composed of a sequence of sounds
- sounds (and transitions between them) serve as a symbolic representation of information to be shared between humans (or humans and machines)
- arrangement of sounds is governed by rules of language (constraints on sound sequences, word sequences, etc)-- /spl/ exists, /sbk/ doesn’t exist
- linguistics(语言学) is the study of the rules of language
- phonetics(语音学) is the study of the sounds of speech

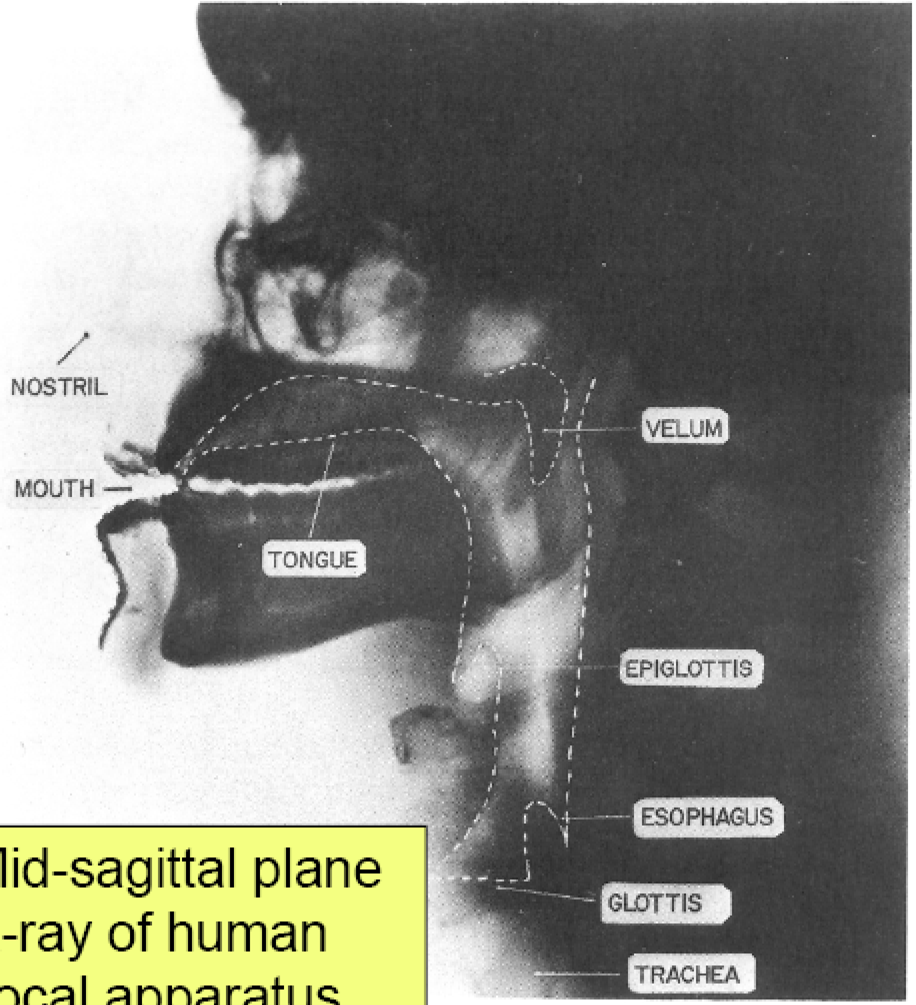
Speech Production Mechanism

Speech Production Mechanism



- air enters the lungs via normal breathing and no speech is produced (generally) on in-take
- as air is expelled from the lungs, via the trachea 气管 or windpipe, the tensed vocal cords within the larynx 喉 are caused to vibrate (Bernoulli oscillation) by the air flow
- air is chopped up into quasi-periodic pulses which are modulated in frequency (spectrally shaped) in passing through the pharynx (the throat cavity), the mouth cavity, and possibly the nasal cavity; the positions of the various articulators (jaw, tongue, velum, lips, mouth) determine the sound that is produced

Human Vocal Apparatus(器官)

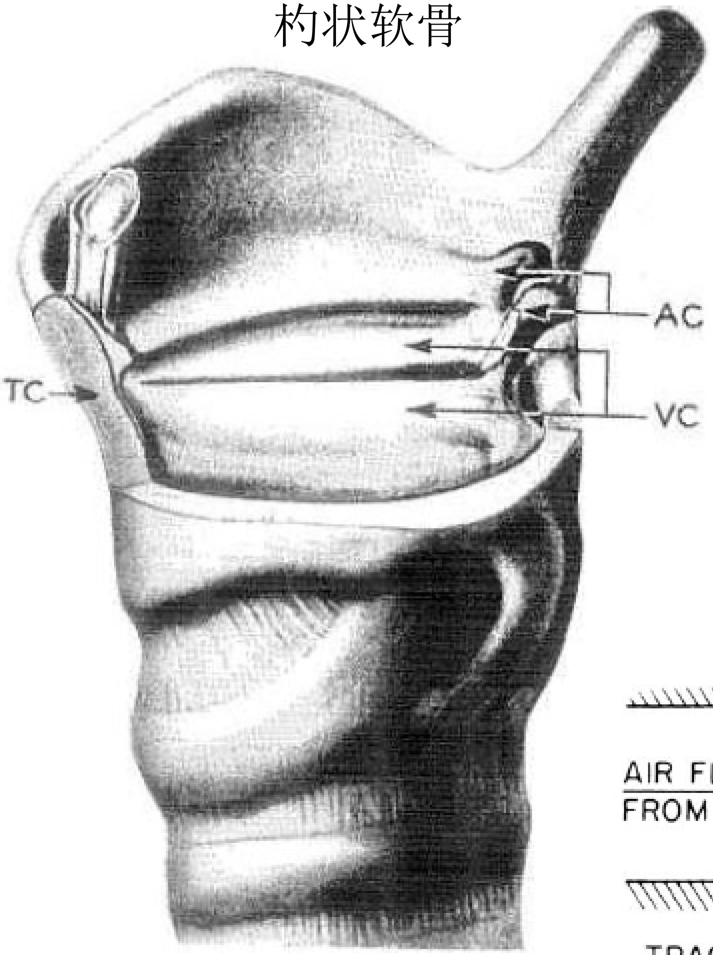


Mid-sagittal plane X-ray of human vocal apparatus

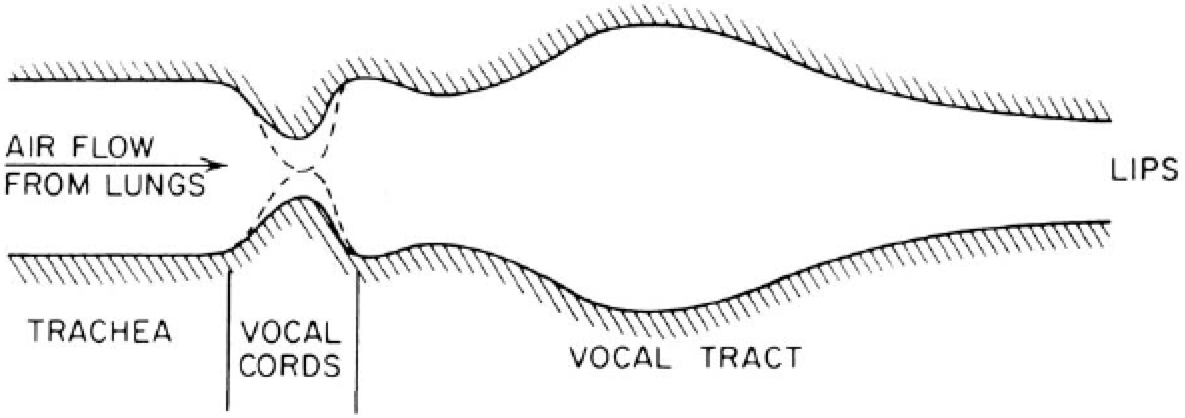
- vocal tract(声道) —dotted lines in figure; begins at the glottis(声门) (the vocal cords 声带) and ends at the lips
  - consists of the pharynx(咽) (the connection from the esophagus 食道 to the mouth) and the mouth itself (the oral cavity)
  - average male vocal tract length is 17.5 cm
  - cross sectional area (横截面积), determined by positions of the tongue, lips, jaw and velum, varies from zero (complete closure) to 20 sq cm
- nasal tract(鼻腔) —begins at the velum and ends at the nostrils
- Velum(软腭) —a trapdoor-like mechanism at the back of the mouth cavity; lowers to couple the nasal tract to the vocal tract to produce the nasal sounds like /m/ (mom), /n/ (night), /ng/ (sing)

Vocal Cords

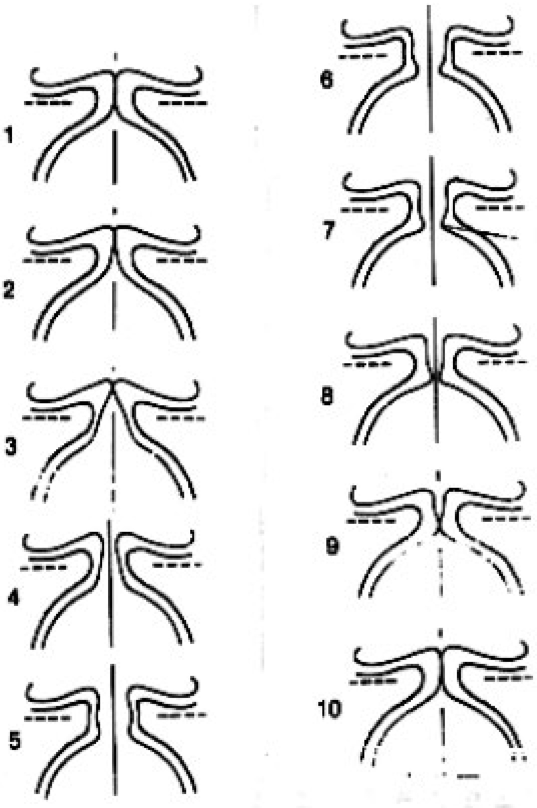
arytenoid cartilage
杓状软骨



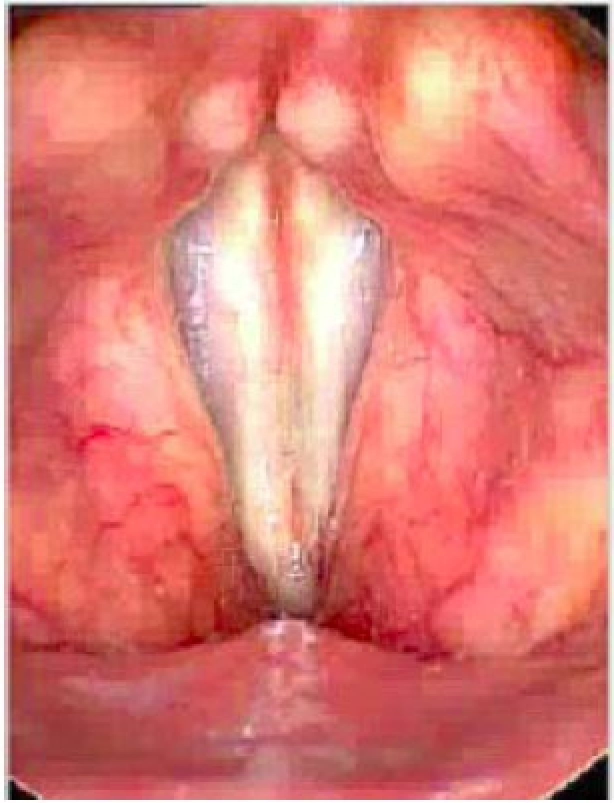
The vocal cords (folds) form a relaxation oscillator. Air pressure builds up and blows them apart. Air flows through the orifice and pressure drops allowing the vocal cords to close. Then the cycle is repeated.



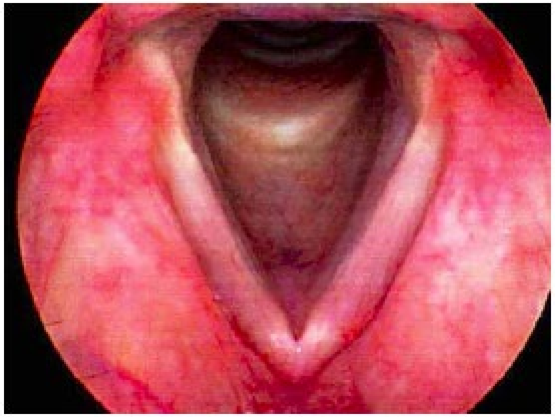
Vocal Cord Views and Operations



Bernoulli Oscillation

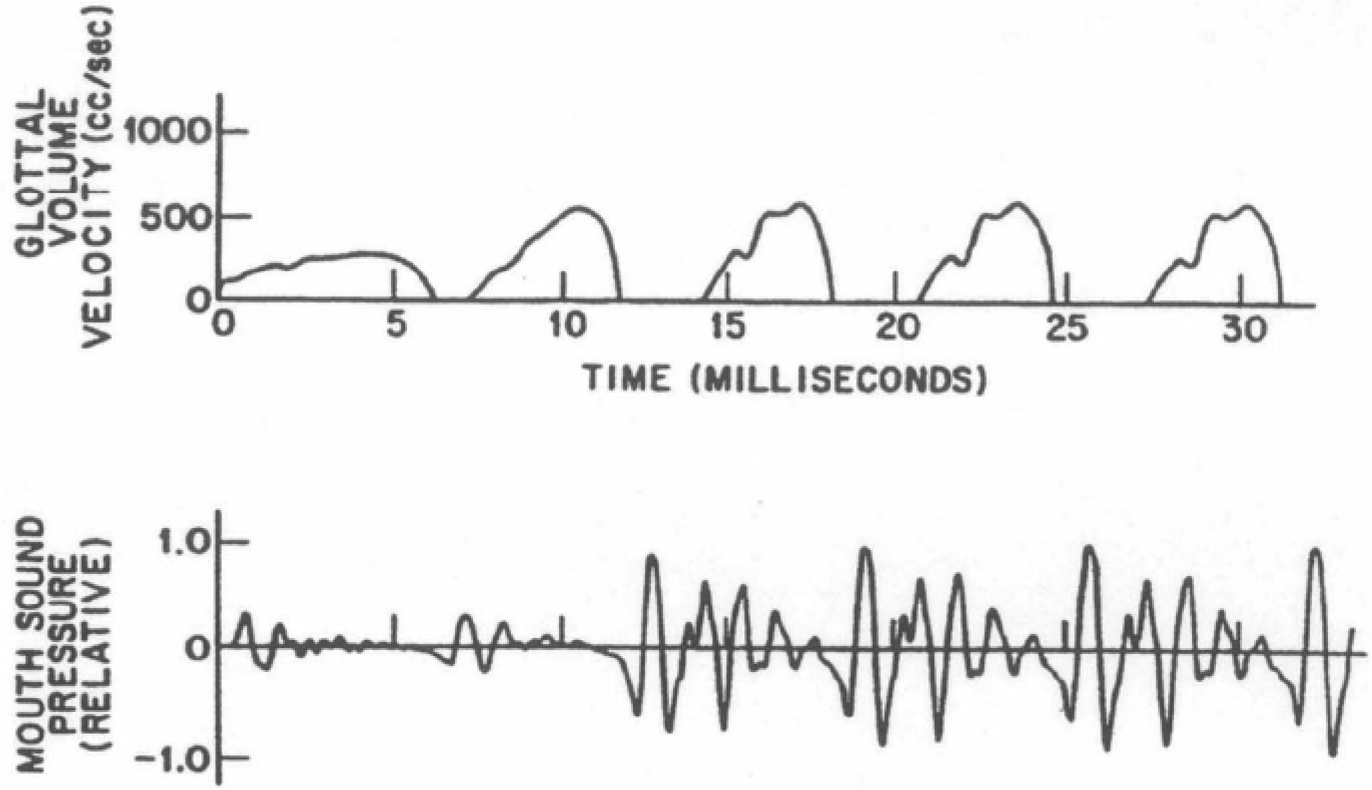


Tensed Vocal Cords – Ready to Vibrate



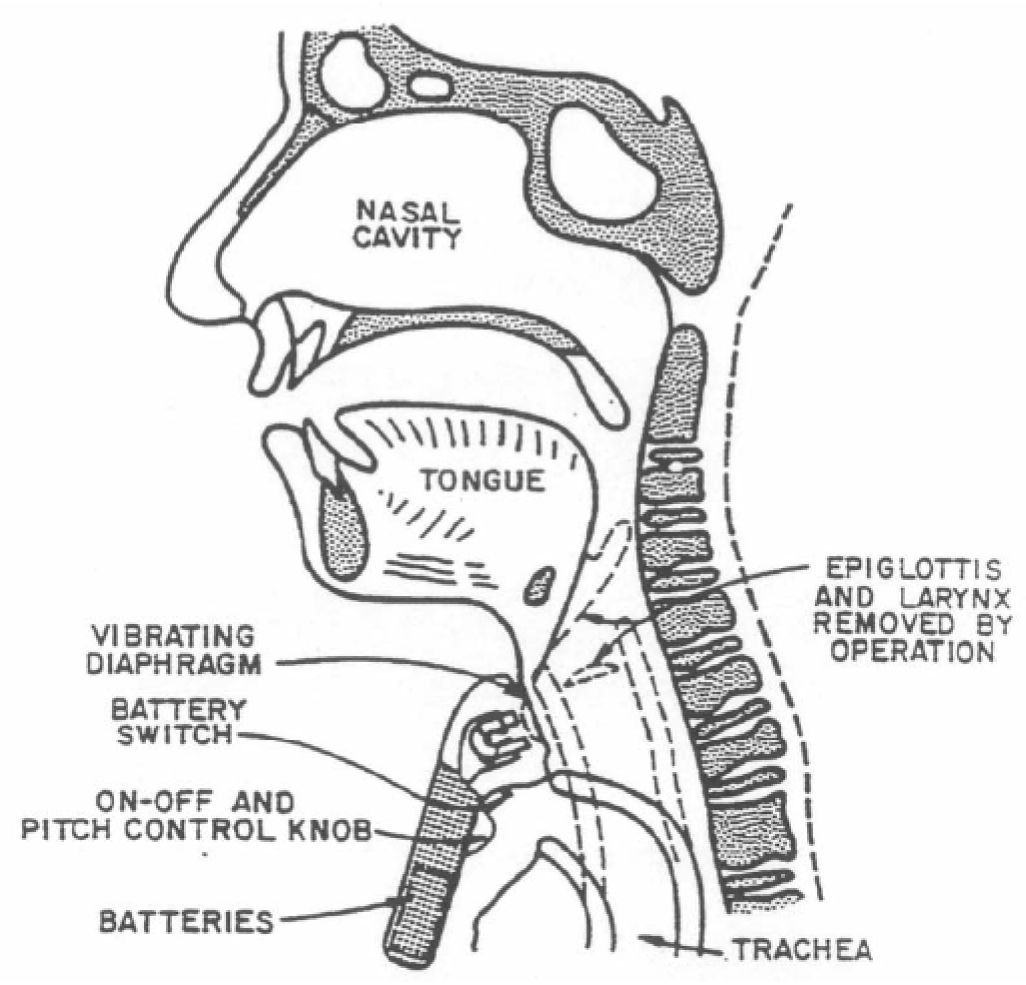
Lax Vocal Cords – Open for Breathing

Glottal Flow

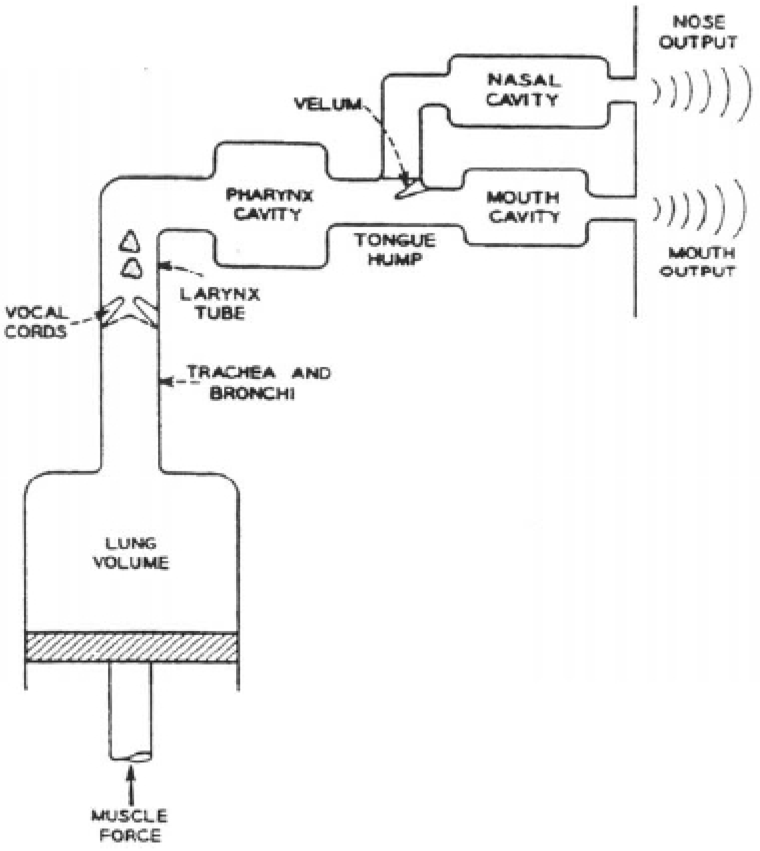


- Glottal volume velocity and resulting sound pressure at the mouth for the first 30 msec of a voiced sound
  - 15 msec buildup to periodicity => pitch detection issues at beginning and end of voicing; also voiced-unvoiced uncertainty for 15 msec

Artificial Larynx


NASAL CAVITY
TONGUE
EPIGLOTTIS AND LARYNX REMOVED BY OPERATION
VIBRATING DIAPHRAGM
BATTERY SWITCH
ON-OFF AND PITCH CONTROL KNOB
BATTERIES
TRACHEA

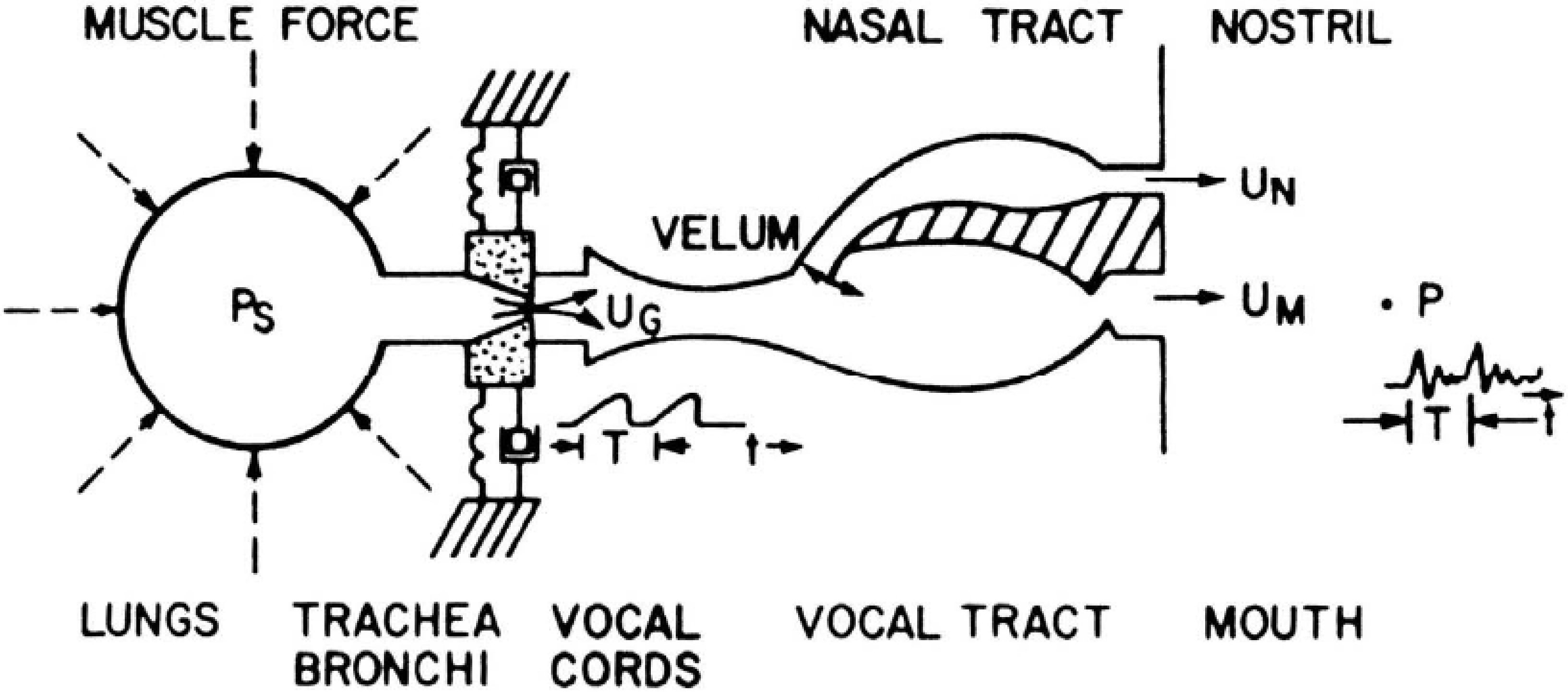
Schematic Production Mechanism

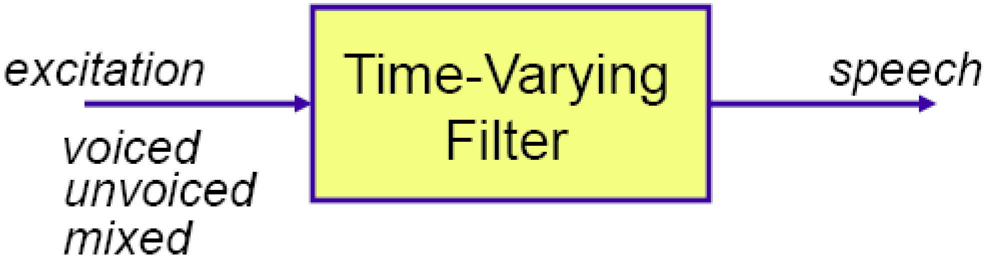


Schematic representation of physiological mechanisms of speech production

- lungs and associated muscles act as the source of air for exciting the vocal mechanism
- muscle force pushes air out of the lungs (like a piston pushing air up within a cylinder) through bronchi and trachea
- if vocal cords are tensed, air flow causes them to vibrate, producing voiced or quasi-periodic speech sounds (musical notes)
- if vocal cords are relaxed, air flow continues through vocal tract until it hits a constriction in the tract, causing it to become turbulent, thereby producing unvoiced sounds (like /s/, /sh/), or it hits a point of total closure in the vocal tract, building up pressure until the closure is opened and the pressure is suddenly and abruptly release, causing a brief transient sound, like at the beginning of /p/, /t/, or /k/

Abstractions of Physical Model





excitation

voiced
unvoiced
mixed

Time-Varying Filter

speech

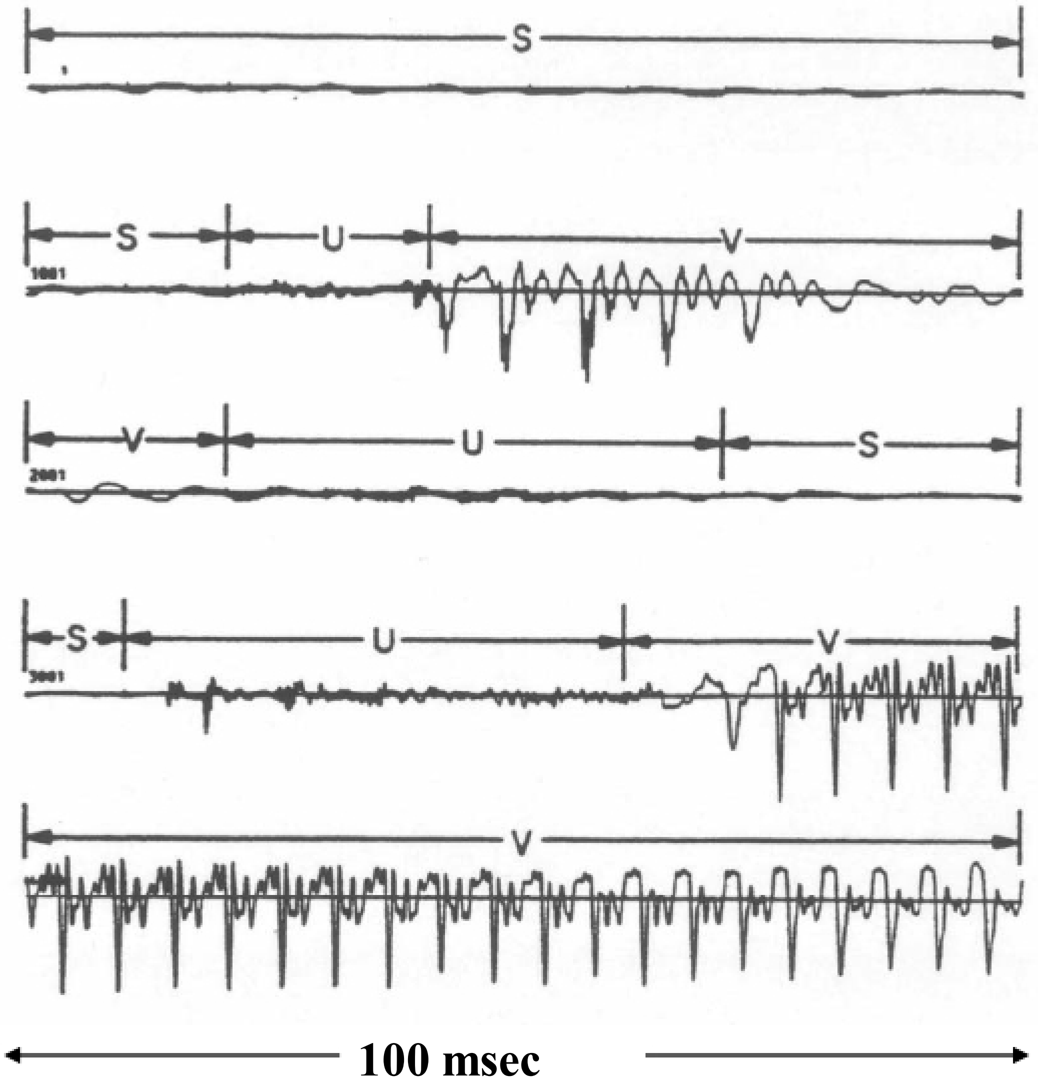
The Speech Signal

The Speech Signal

- speech is a sequence of ever changing sounds
- sound properties are highly dependent on context(语境) (i.e., the sounds which occur before and after the current sound)
- the state of the vocal cords, the positions, shapes and sizes of the various articulators—all change slowly over time, thereby producing the desired speech sounds

  ⇒need to determine the physical properties of speech by observing and measuring the speech waveform (as well as signals derived from the speech waveform—e.g., the signal spectrum)

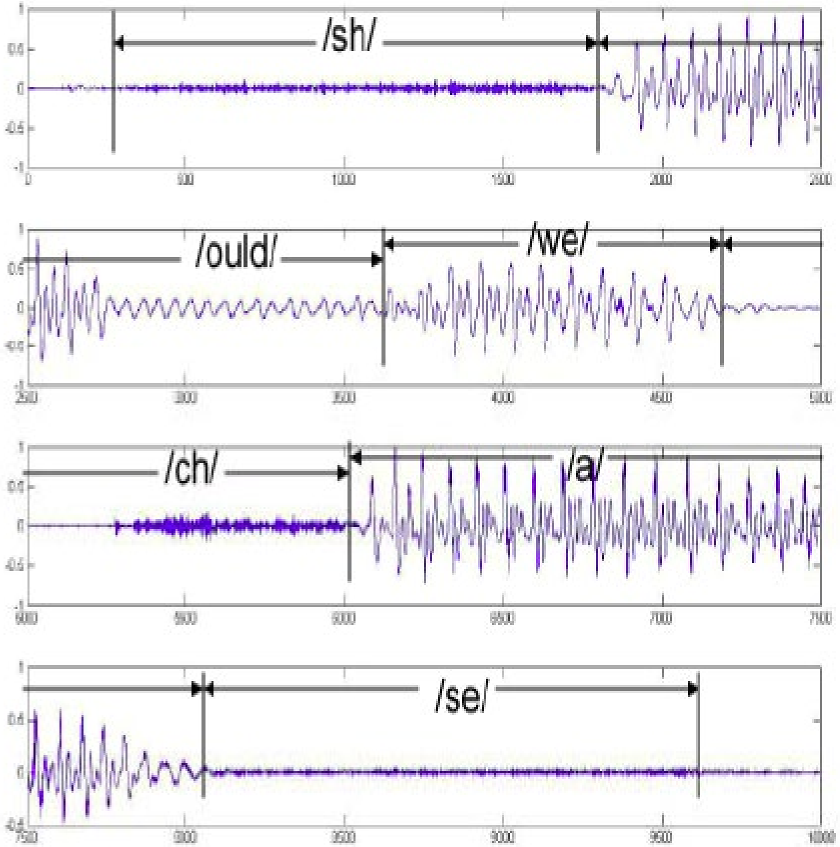
Speech Waveforms and Spectra



- 100 msec/line; 0.5 sec for utterance
- S-silence-background: no speech
- U-unvoiced: no vocal cord vibration
- V-voiced: quasi-periodic speech
- speech is a slowly time varying signal over 5-100 msec intervals
- over longer intervals (100 msec-5 sec), the speech characteristics change as rapidly as 10-20times/second
- no well-defined or exact regions where individuals sounds begin and end

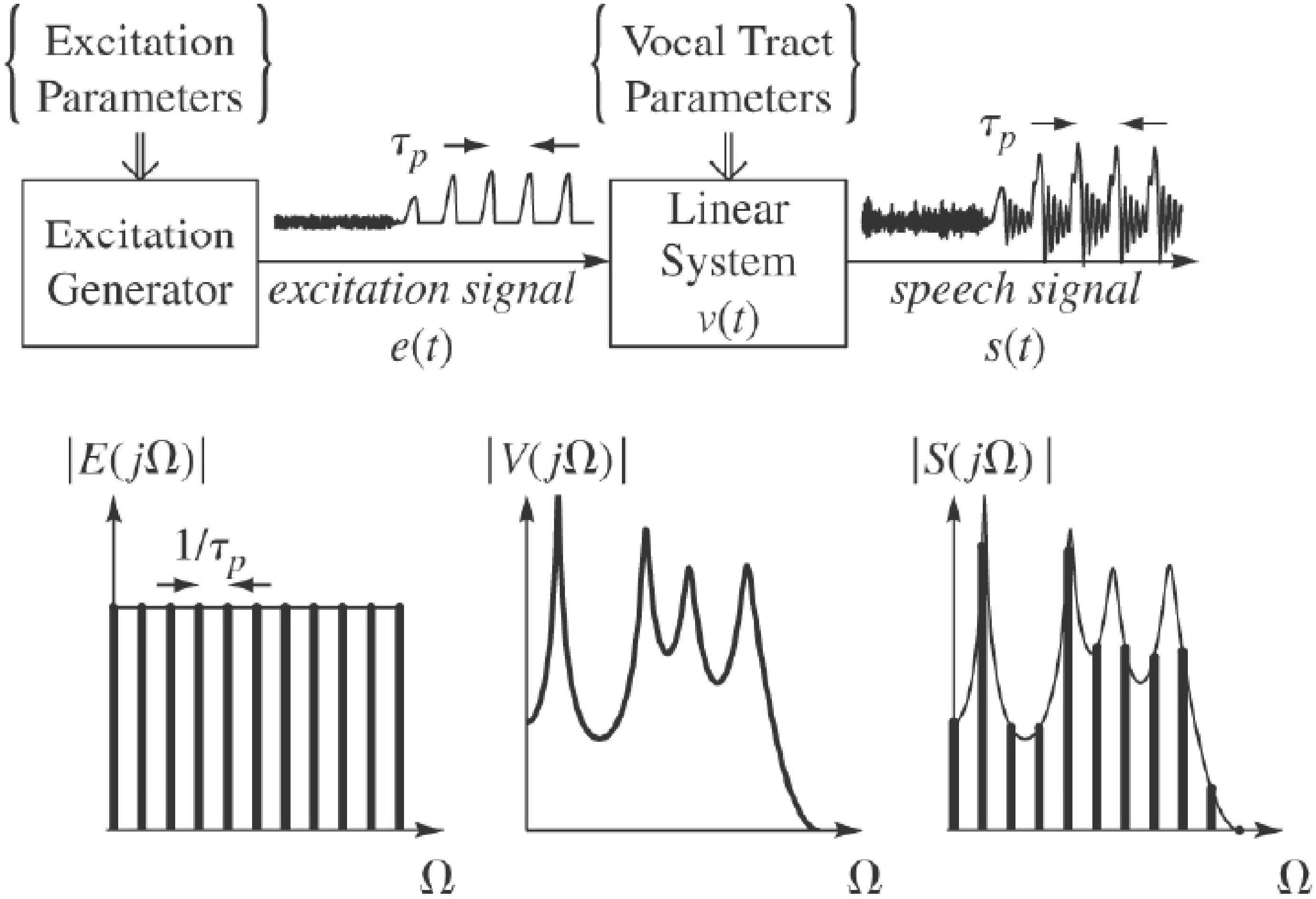
Speech Sounds

- “Should we chase”
  - (Praat demo)

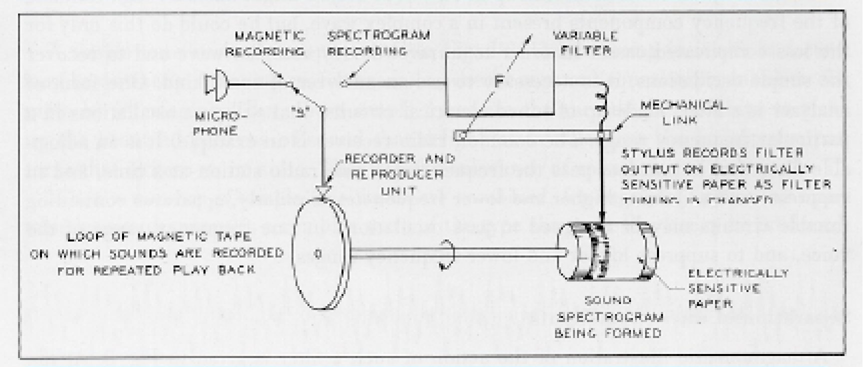


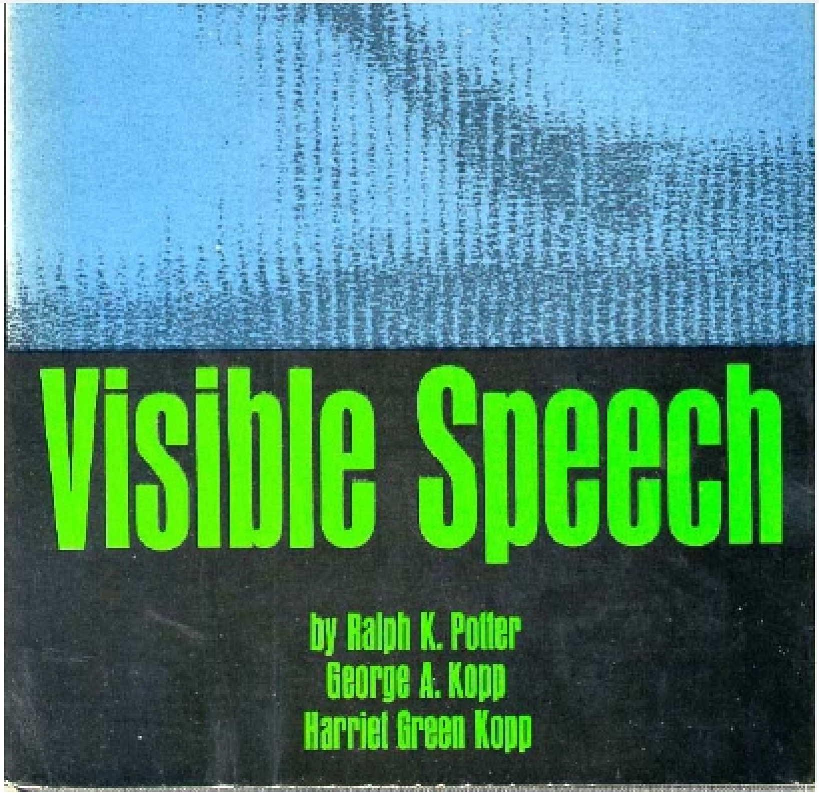
  - hard to distinguish weak sounds from silence
  - Hard to segment with high precision

Source-System Model of Speech Production



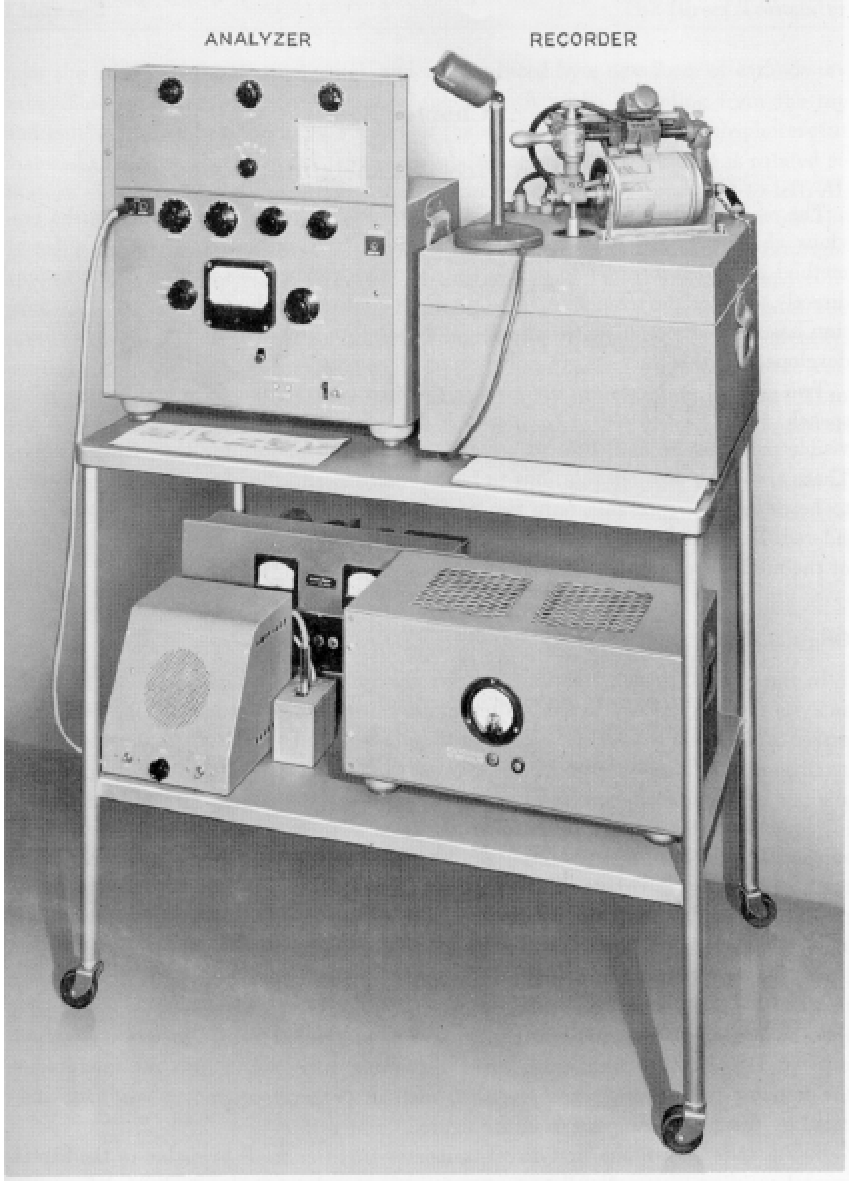
Making Speech "Visible" in 1947


MAGNETIC RECORDING
SPECTROGRAM RECORDING
VARIABLE FILTER
F
MICRO-PHONE
"S"
MECHANICAL LINK
RECORDER AND REPRODUCER UNIT
STYLUS RECORDS FILTER OUTPUT ON ELECTRICALLY SENSITIVE PAPER AS FILTER TUNING IS CHANGED
LOOP OF MAGNETIC TAPE ON WHICH SOUNDS ARE RECORDED FOR REPEATED PLAY BACK
0
ELECTRICALLY SENSITIVE PAPER
SOUND SPECTROGRAM BEING FORMED



Visible Speech

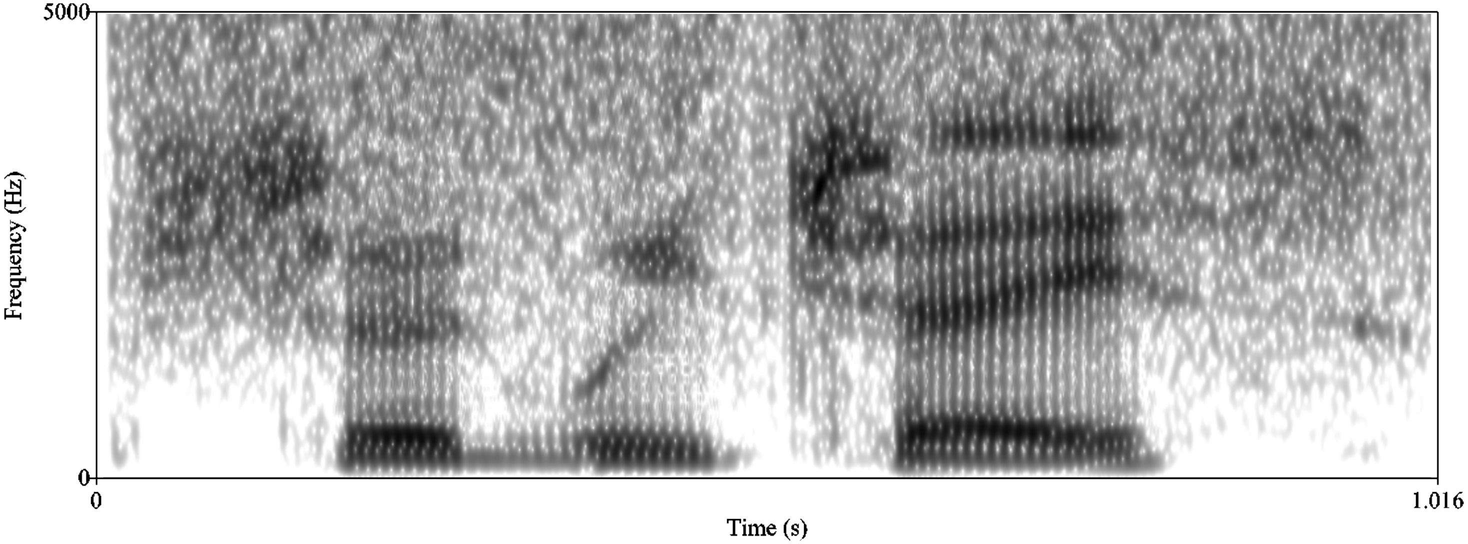
by Ralph K. Potter
George A. Kopp
Harriet Green Kopp



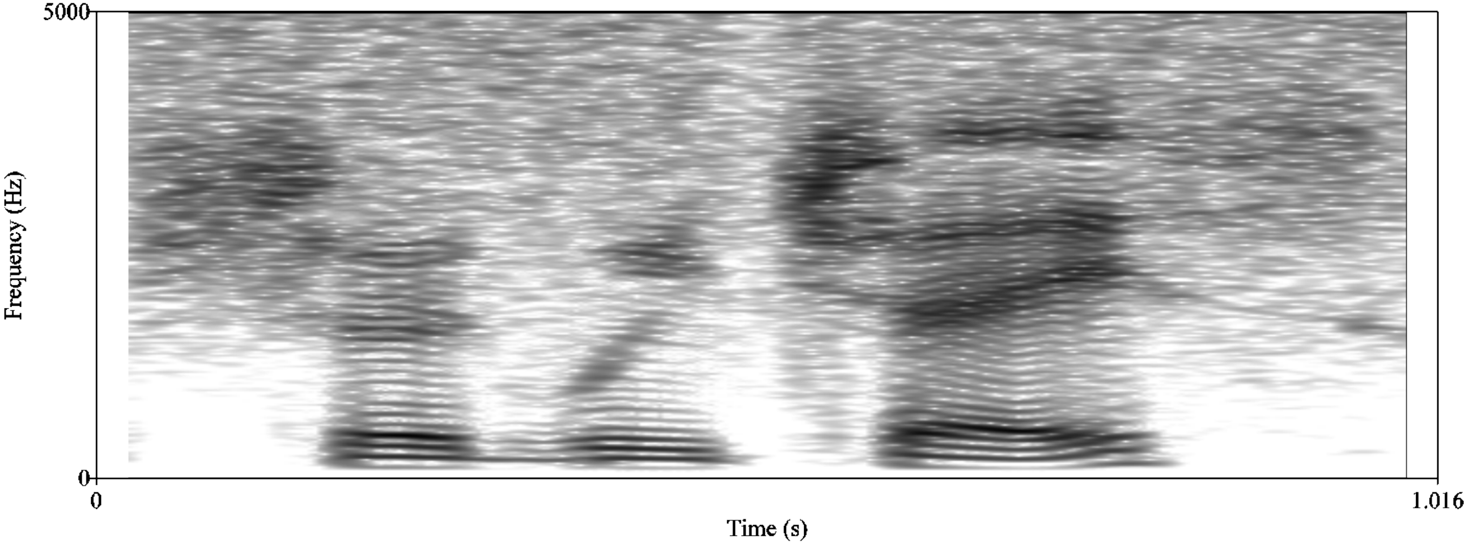
Spectrogram Properties

- speech spectrogram
  - sound intensity versus time and frequency
- wideband spectrogram
  - spectral analysis on 16 msec sections of waveform using a broad (125 Hz) bandwidth analysis filter, with new analyzes every 1 msec
  - spectral intensity resolves individual periods of the speech and shows vertical striations(条纹) during voiced regions
- narrowband spectrogram
  - spectral analysis on 50 msec sections of waveform using a narrow (40 Hz) bandwidth analysis filter, with new analyzes every 1 msec
  - narrowband spectrogram resolves individual pitch harmonics and shows horizontal striations during voiced regions

Wideband and Narrowband Spectrograms

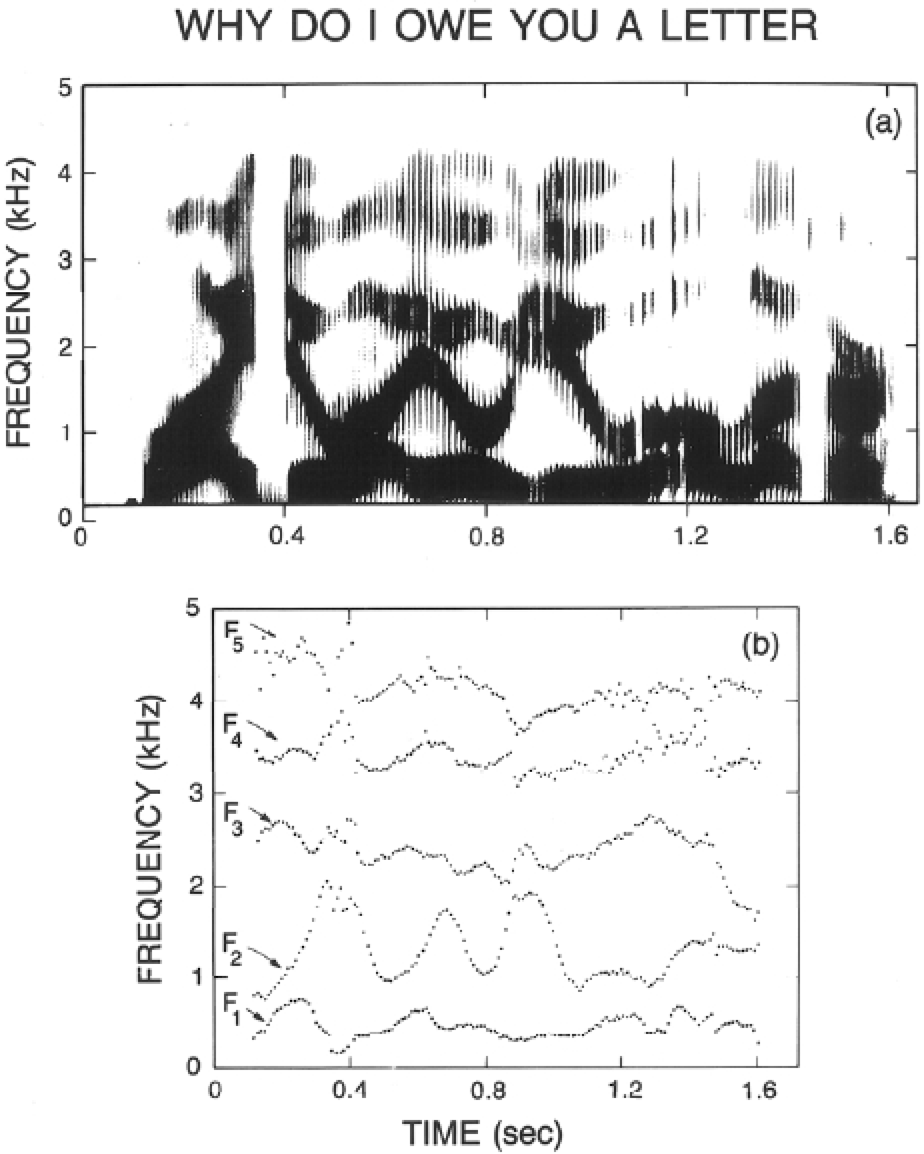


10ms windows



50ms windows

Spectrogram and Formants



<u>Key Issue</u>

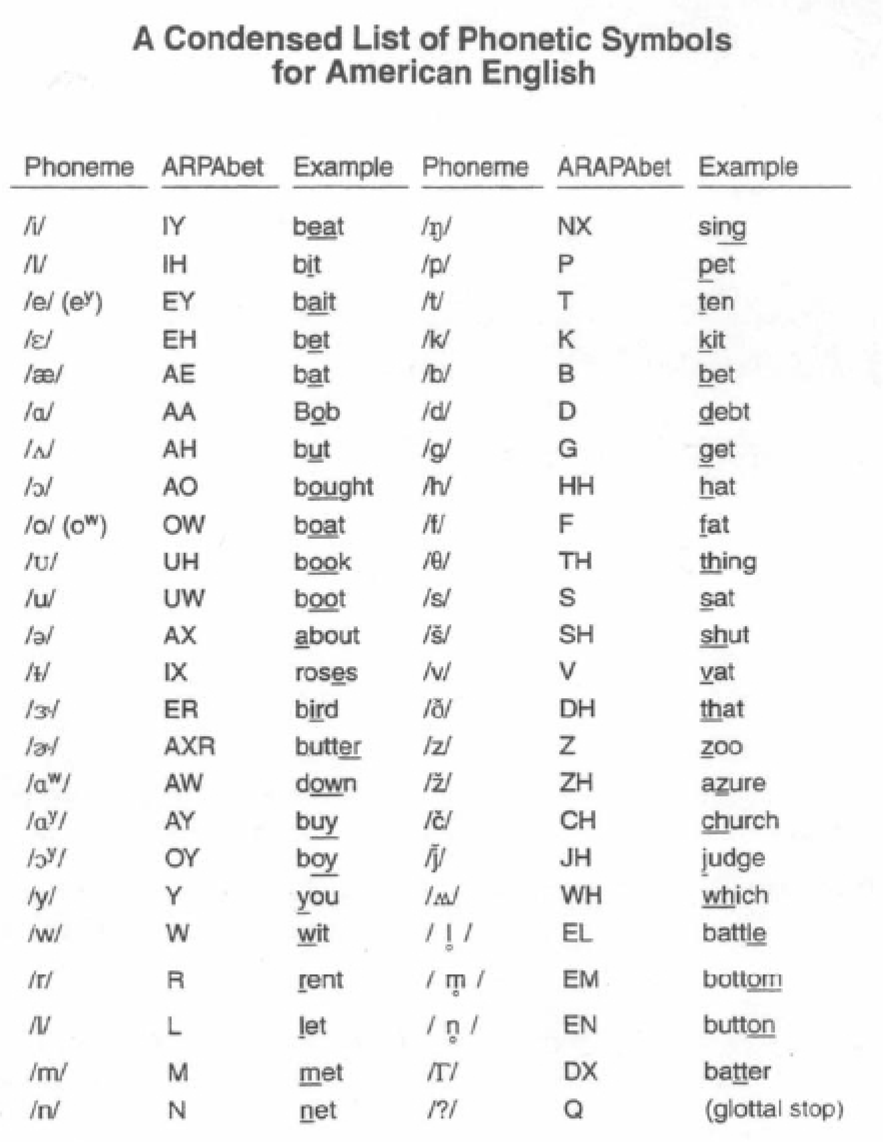
reliability in estimating formants from spectral data

Summary

- basic speech processes — from ideas to speech (production), from speech to ideas (perception)
- basic vocal production mechanisms — vocal tract, nasal tract, velum
- source of sound flow at the glottis; output of sound flow at the lips and nose
- speech waveforms and properties — voiced, unvoiced, silence, pitch
- speech spectrograms and properties —wideband spectrograms, narrowband spectrograms, formants

Sounds of Language: Phonemes

English Speech Sound



A Condensed List of Phonetic Symbols
for American English

Phoneme	ARPAbet	Example	Phoneme	ARAPAbet	Example
/i/	IY	beat	/ŋ/	NX	sing
/ɪ/	IH	bit	/p/	P	pet
/e/ (e^y)	EY	bait	/t/	T	ten
/ɛ/	EH	bet	/k/	K	kit
/æ/	AE	bat	/b/	B	bet
/ɑ/	AA	Bob	/d/	D	debt
/ʌ/	AH	but	/g/	G	get
/ɔ/	AO	bought	/h/	HH	hat
/o/ (o^w)	OW	boat	/f/	F	fat
/ʊ/	UH	book	/θ/	TH	thing
/u/	UW	boot	/s/	S	sat
/ə/	AX	about	/š/	SH	shut
/ɨ/	IX	roses	/v/	V	vat
/ɝ/	ER	bird	/ð/	DH	that
/ɚ/	AXR	butter	/z/	Z	zoo
/ɑ^w/	AW	down	/ž/	ZH	azure
/ɑ^y/	AY	buy	/č/	CH	church
/ɔ^y/	OY	boy	/ǰ/	JH	judge
/y/	Y	you	/ʍ/	WH	which
/w/	W	wit	/ l̥ /	EL	battle
/r/	R	rent	/ m̥ /	EM	bottom
/l/	L	let	/ n̥ /	EN	button
/m/	M	met	/ɾ/	DX	batter
/n/	N	net	/ʔ/	Q	(glottal stop)

- ARPABET representation
- 48 sounds
  - – 18 vowels(元音)/diphthongs(复合元音)
  - – 4 vowel-like consonants(辅音)
  - – 21 standard consonants
  - – 4 syllabic sounds(成音节辅音)
  - – 1 glottal stop(喉塞音)

Phonemes—Link Between Orthography(拼写) and Speech

- Orthography→sequence of sounds
  - Larry → /L/ /AE/ /R/ /IY/
- Speech waveform → sequence of sounds
  - based on acoustic properties (temporal) of phonemes
- Spectrogram → sequence of sounds
  - based on acoustic properties (spectral) of phonemes

We use the phonetic code as an intermediate representation of language and therefore it is essential to understand the acoustic and articulatory properties of all of the sounds (phonemes) of a language in order to design the best speech processing systems (especially for speech synthesis and speech recognition applications)

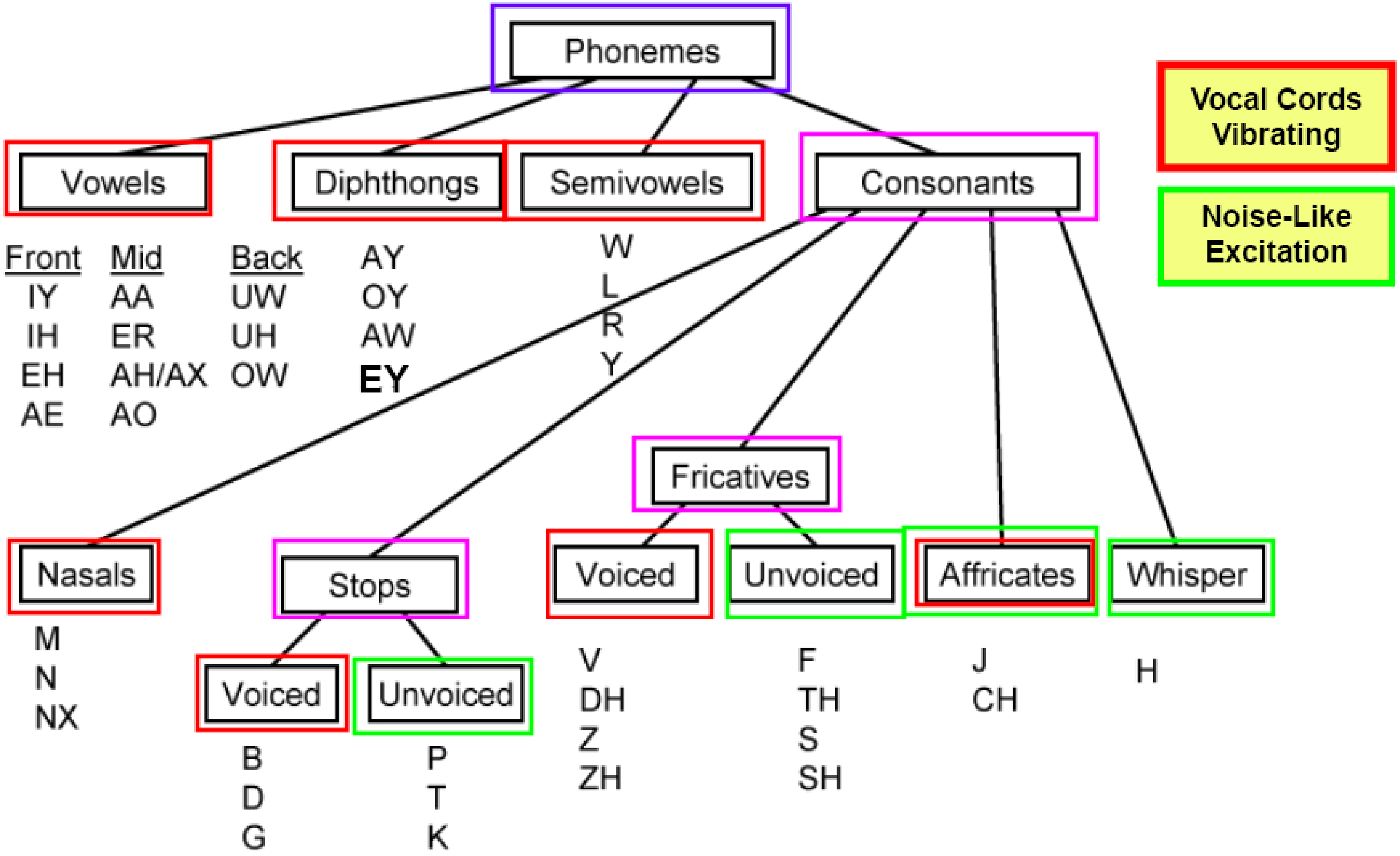
Phonetic Transcription

- based on ideal (dictionary-based) pronunciations of all words in sentence
  - ‘My name is Larry’-/M/ /AY/-/N/ /EY/ /M/-/IH/ /Z/-/L/ /AE/ /R/ /IY/
  - ‘How old are you’-/H/ /AW/-/OW/ /L/ /D/-/AA/ /R/-/Y/ /UW/
  - ‘Speech processing is fun’-/S/ /P/ /IY/ /CH/-/P/ /R/ /AH/ /S/ /EH/ /S/ /IH/ /NG/-/IH/ /Z/-/F/ /AH/ /N/
- word ambiguity abounds
  - ‘lives’-/L/ /IH/ /V/ /Z/ (he lives here) versus /L/ /AY/ /V/ /Z/ (a cat has nine lives)
  - ‘record’-/R/ /EH/ /K/ /ER/ /D/ (he holds the world record) versus /R/ /IY/ /K/ /AW/ /D/ (please record my favorite show tonight)

Reduced Set of American English Sounds

- 39 sounds
  - 11 vowels (front, mid, back) classification based on tongue hump position
  - 4 diphthongs (vowel-like combinations)
  - 4 semi-vowels 半元音 (liquids边音/流音 and glides滑音)
  - 3 nasal consonants
  - 6 voiced浊 and unvoiced清 stop consonants塞音
  - 8 voiced and unvoiced fricative consonants擦音
  - 2 affricate consonants赛擦音
  - 1 whispered sound
- look at each class of sounds to characterize their acoustic and spectral properties

Phoneme Classification Chart



Phonemes

Vowels

Front	Mid	Back
IY	AA	UW
IH	ER	UH
EH	AH/AX	OW
AE	AO	

Diphthongs

AY
OY
AW
EY

Semivowels

W
L
R
Y

Consonants

Nasals

M
N
NX

Stops

Voiced

B
D
G

Unvoiced

P
T
K

Fricatives

Voiced

V
DH
Z
ZH

Unvoiced

F
TH
S
SH

Affricates

J
CH

Whisper

H

Vocal Cords Vibrating

Noise-Like Excitation

Vowels

- longest duration sounds – least context sensitive
- can be held indefinitely in singing and other musical works (opera)
- carry very little linguistic information (some languages don’t display vowels in text- e.g. Hebrew 希伯来语, Arabic阿拉伯语)

Vowels and Consonants

- ***Text 1: all vowels deleted***

  Th_y n_t_d s_gn_f_c_nt _mpr_v_m_nts _n th_ c_mp_ny’s_m_g_, s_p_rv_s__n _nd m_n_g_m_nt.

  (They noted significant improvements in the company’s image, supervision and management.)

- ***Text 2: all consonants deleted***

  A__i_u_e_ _o_a__ _a_ __a_e_ e__e__ia___ __e _a_e, _i__ __e __i_e_ o_ o__u_a_io_a_ e___o_ee_ __i_____ _e__ea_i__.

  (Attitudes pay stayed toward essentially the same, with the scores of occupational employees slightly decreasing)

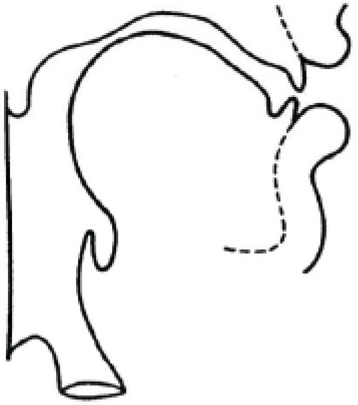
Vowels

- produced using fixed vocal tract shape
- sustained sounds
- vocal cords are vibrating ⇒ voiced sounds
- cross-sectional area of vocal tract determines vowel resonance frequencies and vowel sound quality
- tongue position (height, forward/back position) most important in determining vowel sound
- usually relatively long in duration (can be held during singing) and are spectrally well formed

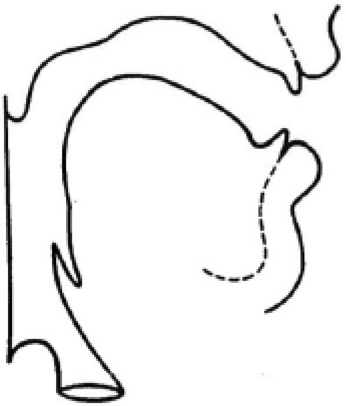
Vowel Production

- No significant constriction (阻塞) in the vocal tract
- Usually produced with periodic excitation
- Acoustic characteristics depend on the position of the jaw, tongue, and lips

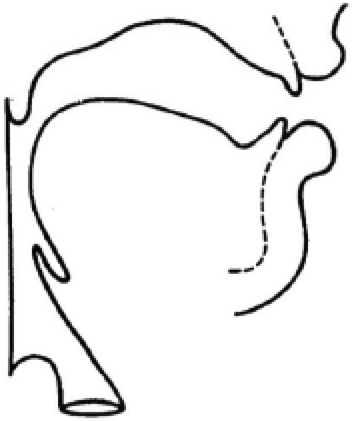
[i]



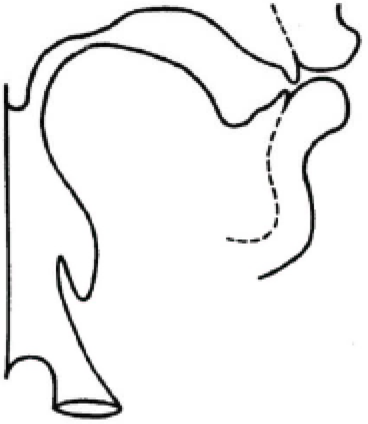
[æ]



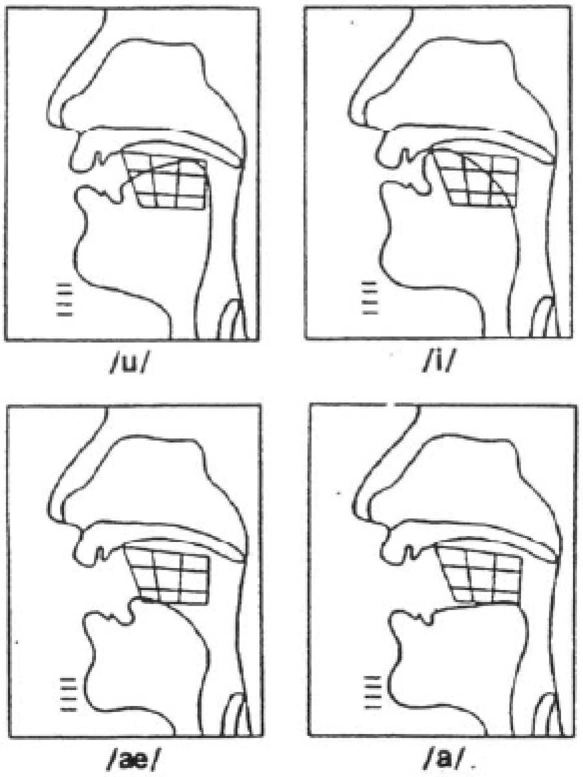
[ɑ]



[u]

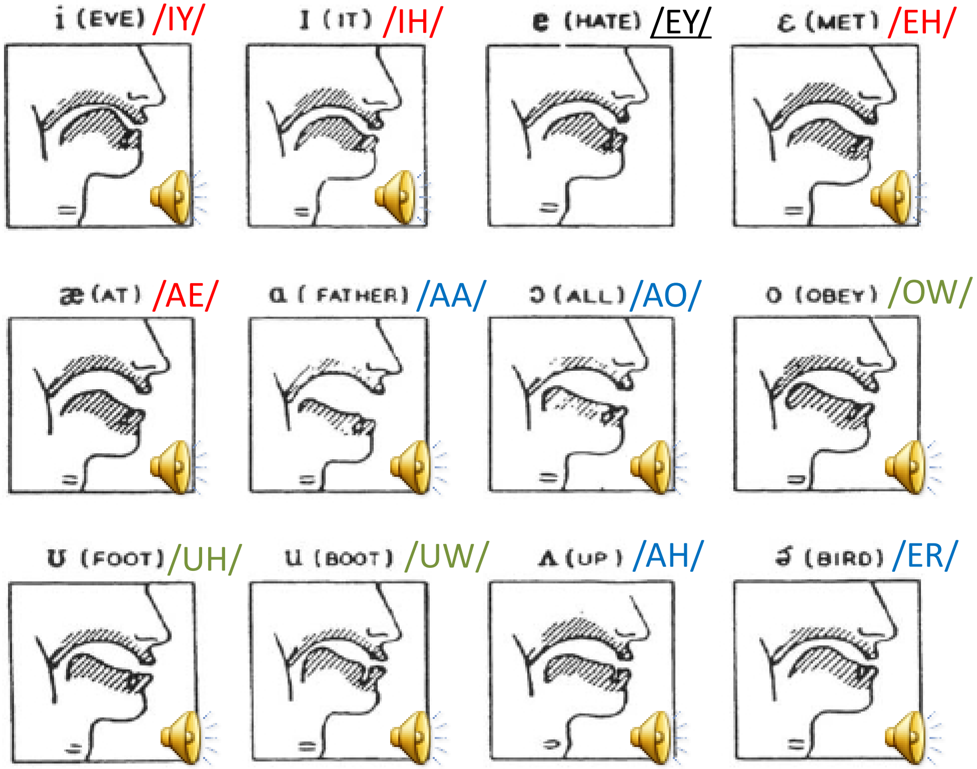


Vowel Articulatory Shapes



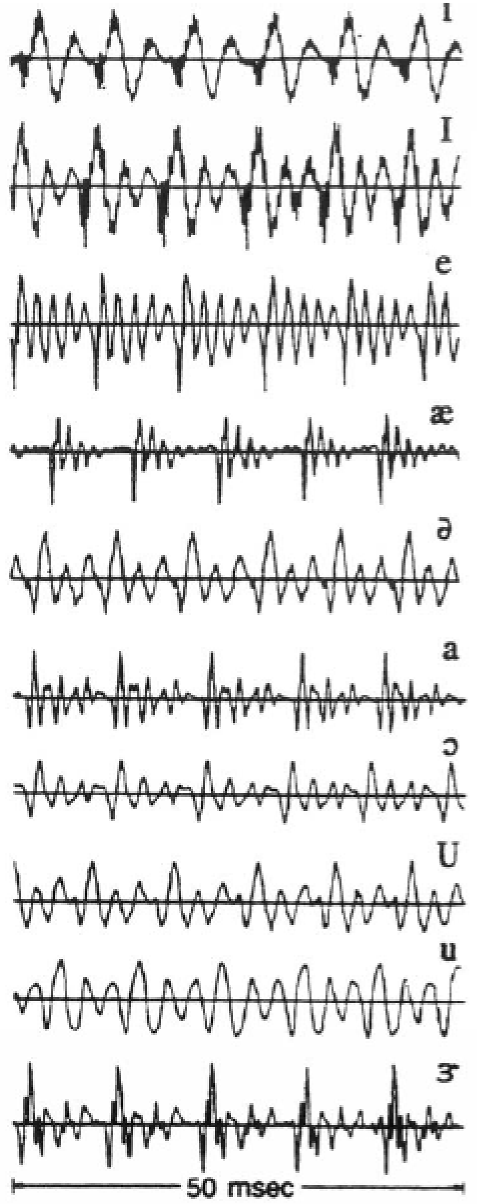
TONGUE POSITION

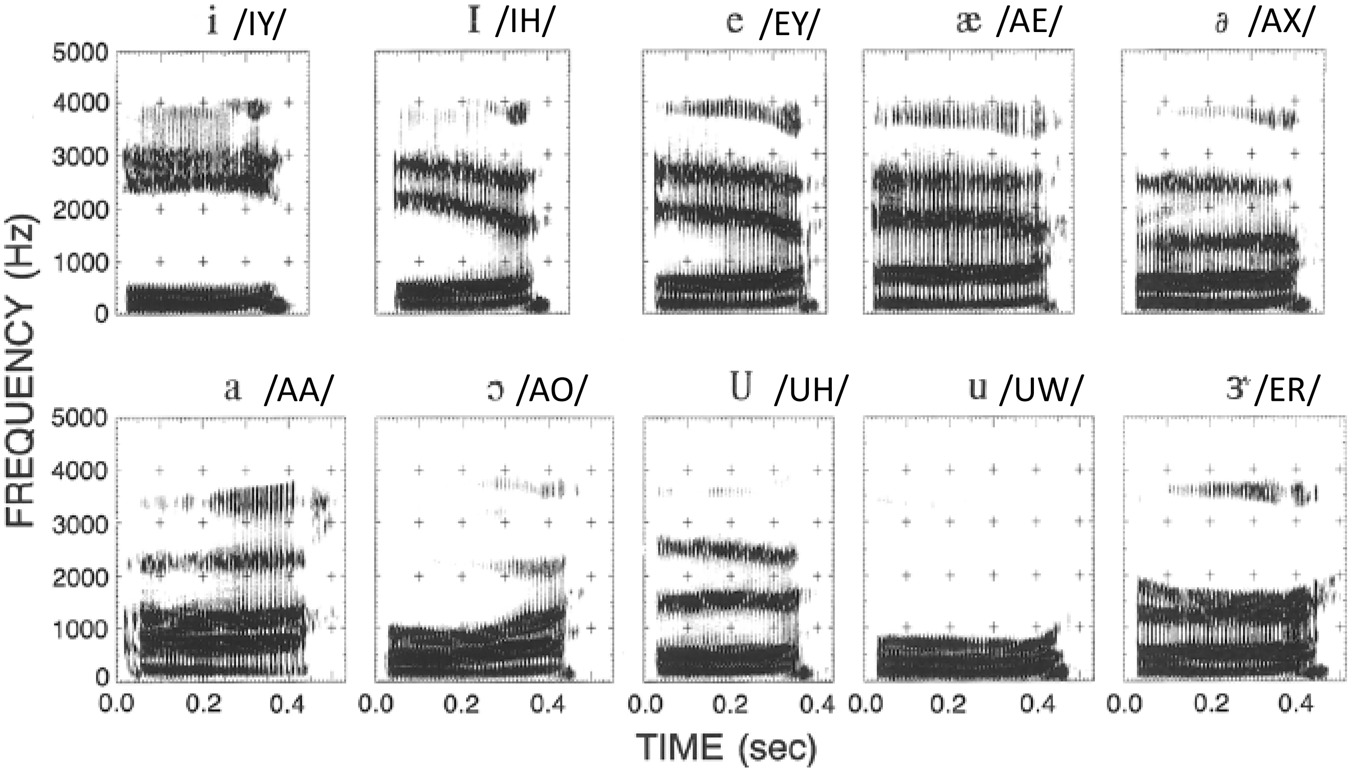
		FRONT	BACK
TONGUE HEIGHT	HIGH	1. i	
	MID	2. I 3. ɛ	.7 u .6 U
	LOW	4. ae	.5 a



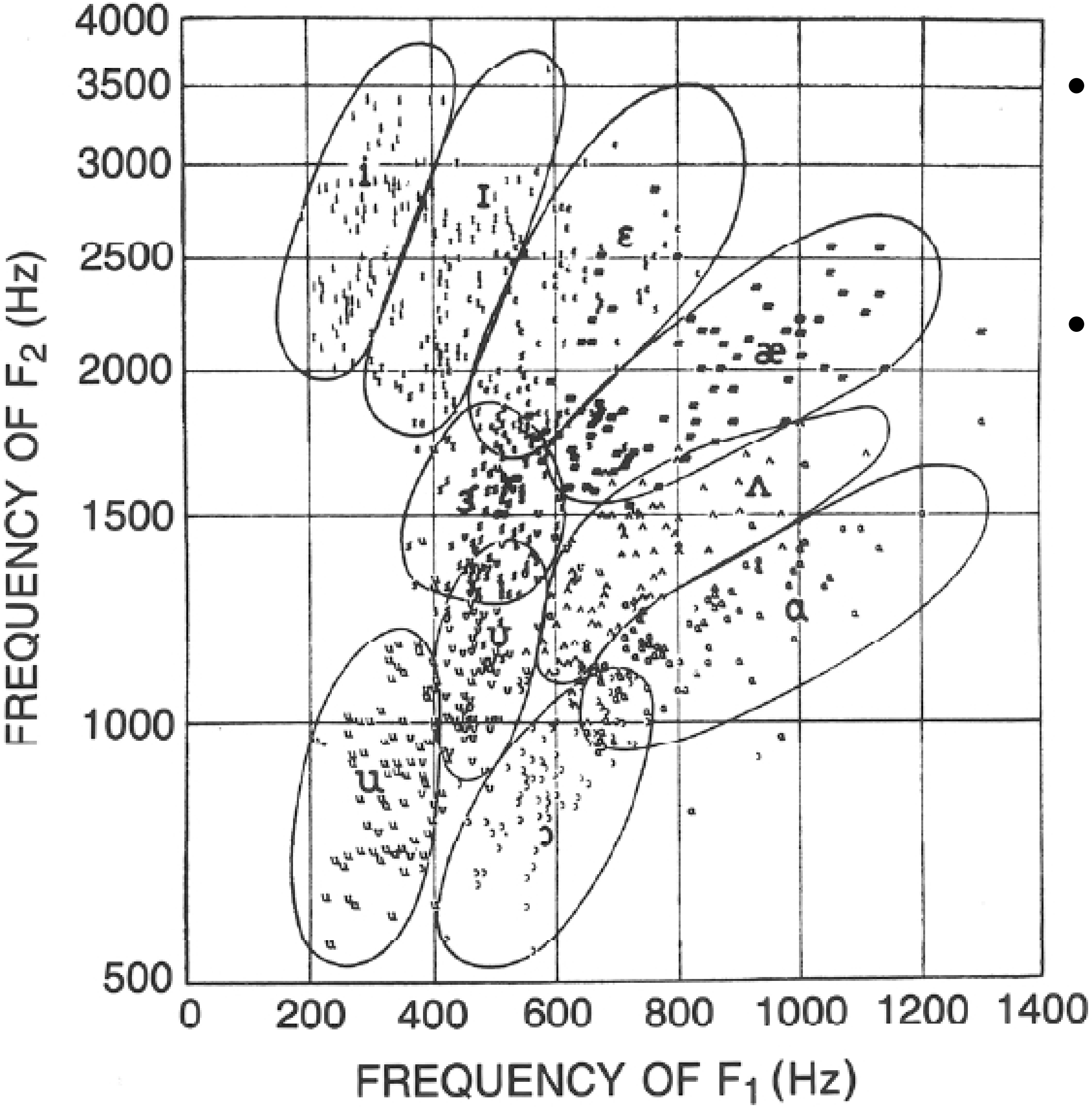
- tongue hump position (front, mid, back)
- tongue hump height (high, mid, low)
- /IY/, /IH/, /EH/,/AE/ => front => high resonances
- /AA/, /AO/, /AH/, /ER/ => mid => energy balance
- /UH/, /UW/, /OW/ => back => low resonances

Vowel Waveforms & Spectrograms





Vowel Formants

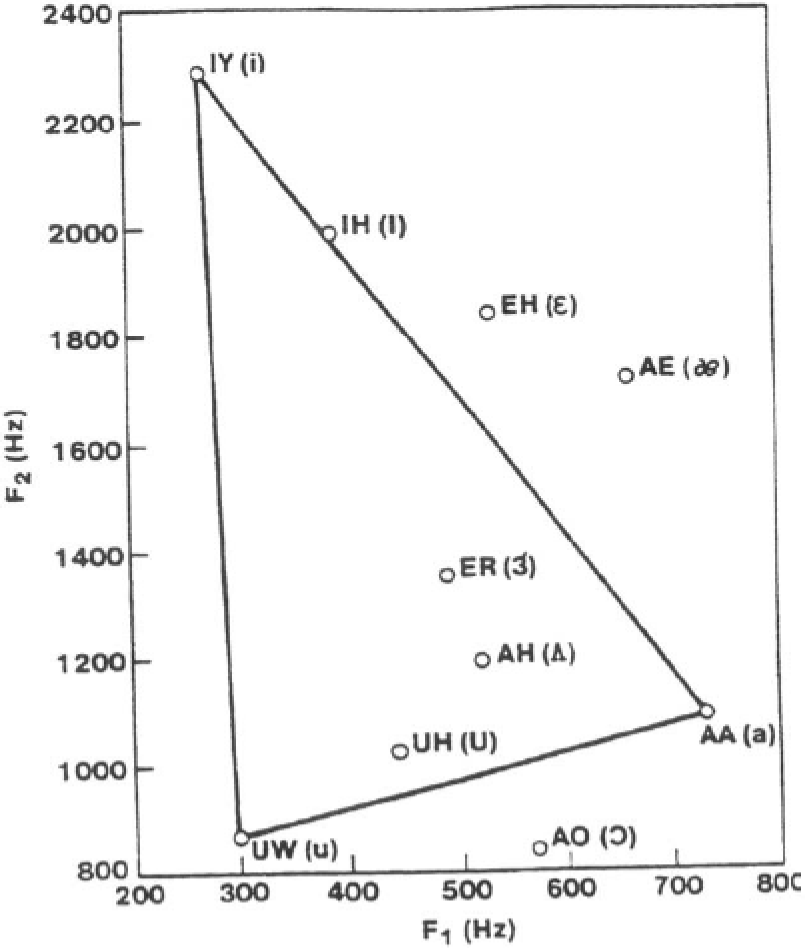


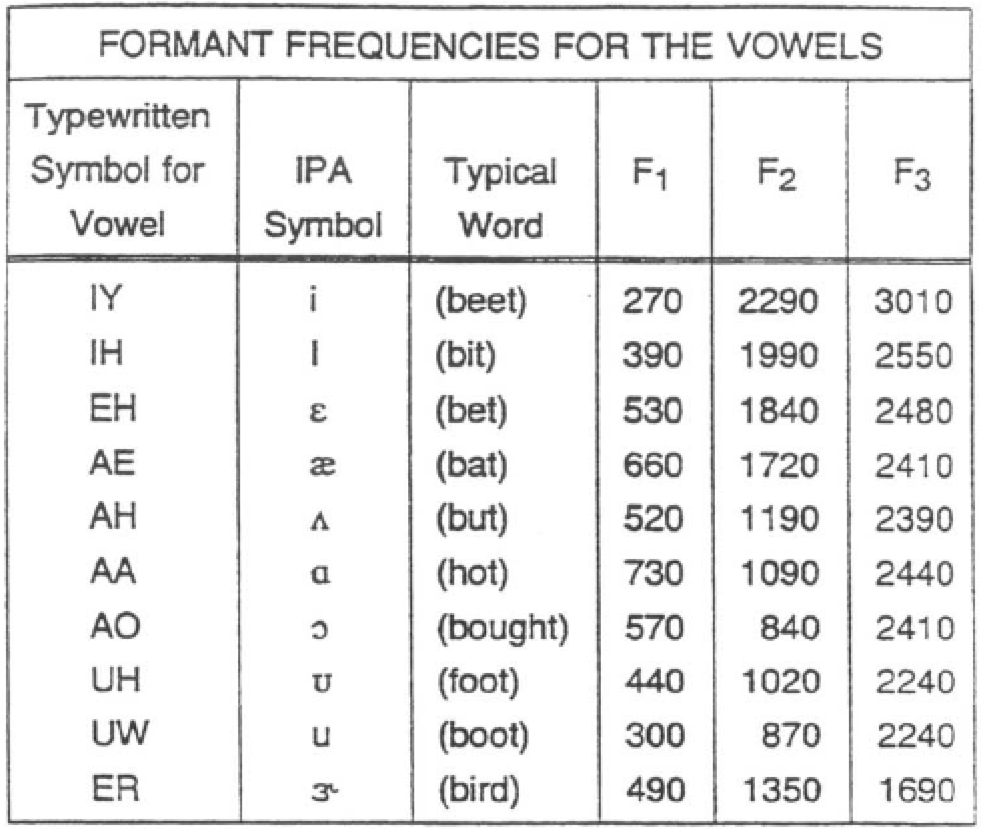
- Clear pattern of variability of vowel pronunciation among men, women and children
- Strong overlap for different vowel sounds by different talkers

  => no unique identification of vowel strictly from resonances

  => need context to define vowel sound

The Vowel Triangle





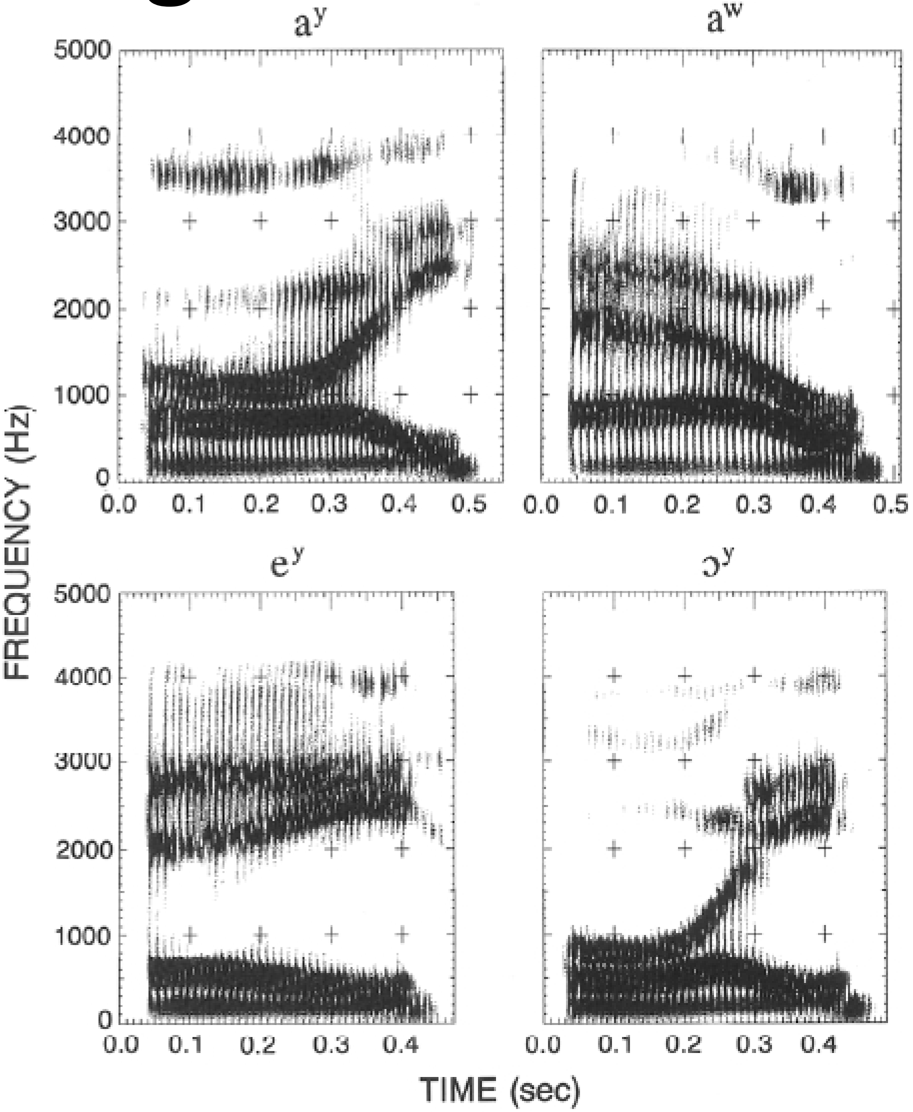
FORMANT FREQUENCIES FOR THE VOWELS

Typewritten Symbol for Vowel	IPA Symbol	Typical Word	F_1	F_2	F_3
IY	i	(beet)	270	2290	3010
IH	ɪ	(bit)	390	1990	2550
EH	ε	(bet)	530	1840	2480
AE	æ	(bat)	660	1720	2410
AH	ʌ	(but)	520	1190	2390
AA	ɑ	(hot)	730	1090	2440
AO	ɔ	(bought)	570	840	2410
UH	ʊ	(foot)	440	1020	2240
UW	u	(boot)	300	870	2240
ER	ɝ	(bird)	490	1350	1690

Centroids of common vowels form clear triangular pattern in F1-F2 space

Diphthongs

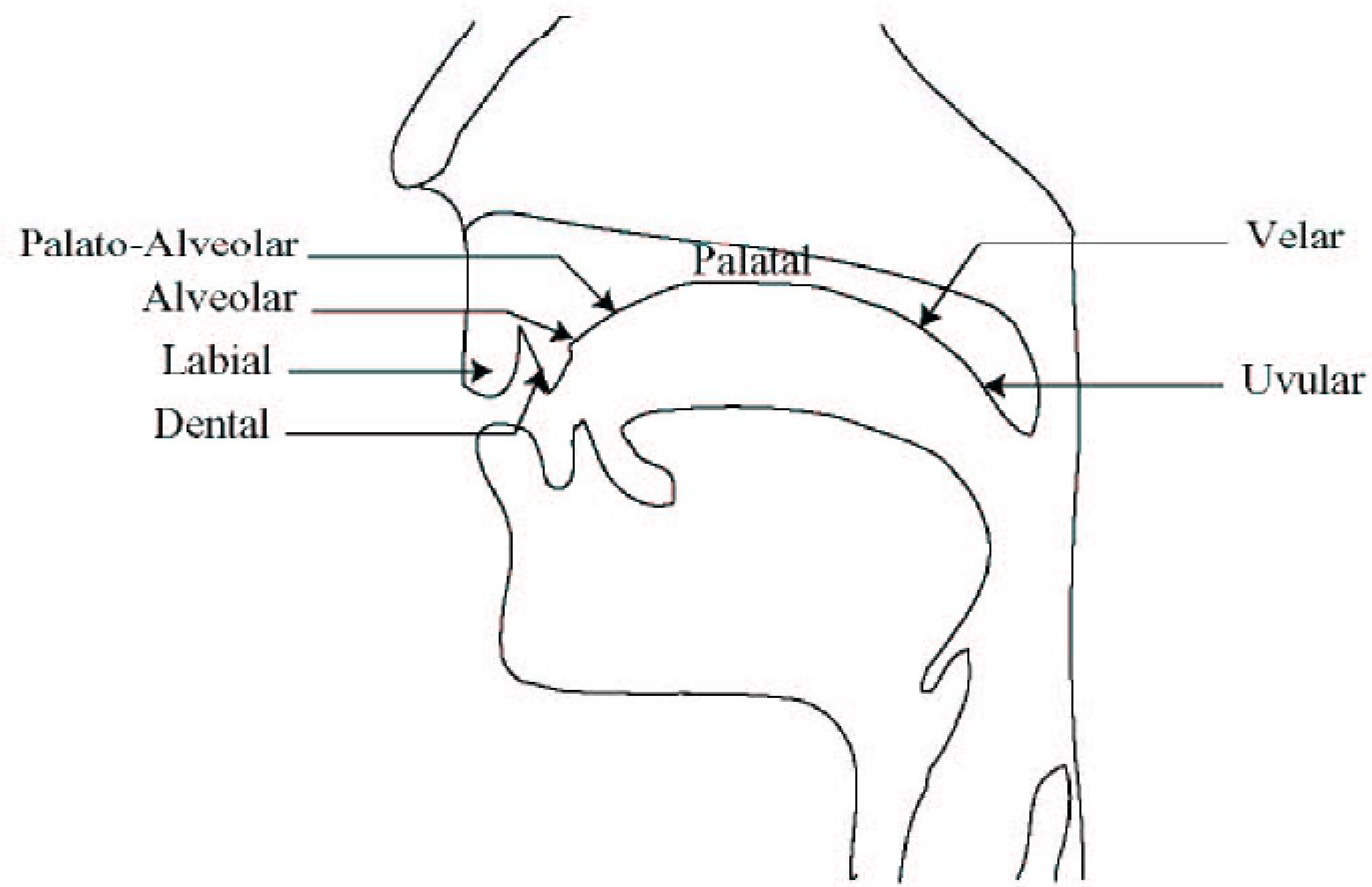
- Gliding speech sound that starts at or near the articulatory position for one vowel and moves to or toward the position for another vowel
  - /AY/ in buy
  - /AW/ in down
  - /EY/ in bait
  - /OY/ in boy
  - /OW/ in boat (usually classified as vowel, not diphthong)
  - /Y/ in you (usually classified as glide)



Distinctive Features

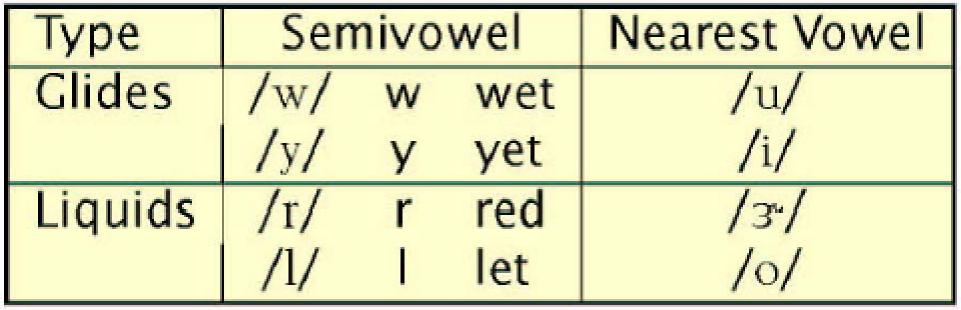
- Classify non-vowel/non-diphthong sounds in terms of distinctive features 区别性特征
  - place of articulation 发音部位
    - Bilabial 双唇音(lips)—p,b,m,w
    - Labiodental 唇齿音(between lips and front of teeth)-f,v
    - Dental 齿音(teeth)-th,dh
    - Alveolar 齿龈音 (front of palate)-t,d,s,z,n,l
    - Palatal 硬腭音(middle of palate)-sh,zh,r
    - Velar 软腭音(at velum)-k,g,ng
    - Pharyngeal 咽音(at end of pharynx)-h
  - manner of articulation 发音方式
    - Glide/Liquid—smooth motion-w,l,r,y
    - Nasal—lowered velum-m,n,ng
    - Stop—constricted vocal tract-p,t,k,b,d,g
    - Fricative—turbulent source-f,th,s,sh,v,dh,z,zh,h
    - Voicing—voiced source-b,d,g,v,dh,z,zh,m,n,ng,w,l,r
    - Mixed source—both voicing and unvoiced-j,ch
    - Whispered--h

Place of Articulation



Semivowels (Liquids and Glides)

- vowel-like in nature (called semivowels for this reason)
- voiced sounds (w-l-r-y)



Type	Semivowel			Nearest Vowel
Glides	/w/	w	wet	/u/
	/y/	y	yet	/i/
Liquids	/r/	r	red	/ɝ/
	/l/	l	let	/o/

- acoustic characteristics of these sounds are strongly influenced by context—unlike most vowel sounds which are much less influenced by context

uh-{w,l,r,y}-a

<u>Manner</u>: glides/liquids
<u>Place</u>: bilabial (w), alveolar (l),palatal (r)

Nasal Consonants

- The nasal consonants consist of /M/, /N/, and /NG/
  - nasals produced using glottal excitation => voiced sound
  - vocal tract totally constricted at some point along the tract
  - velum lowered so sound is radiated at nostrils鼻孔
  - constricted oral cavity serves as a resonant cavity that traps acoustic energy at certain natural frequencies (anti-resonances or zeros of transmission)
  - /M/ is produced with a constriction at the lips => low frequency zero
  - /N/ is produced with a constriction just behind the teeth => higher frequency zero
  - /NG/ is produced with a constriction just forward of the velum => even higher frequency zero

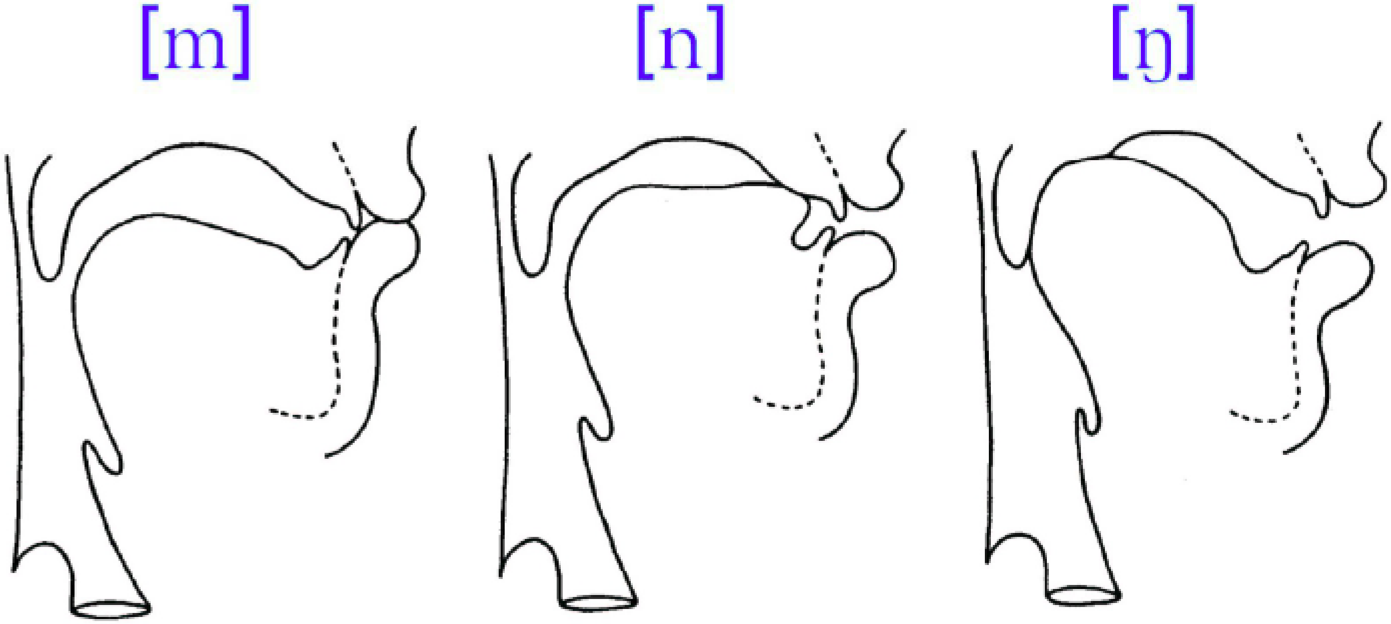
🔈 🔈 🔈

uh-{m,n,ng}-a

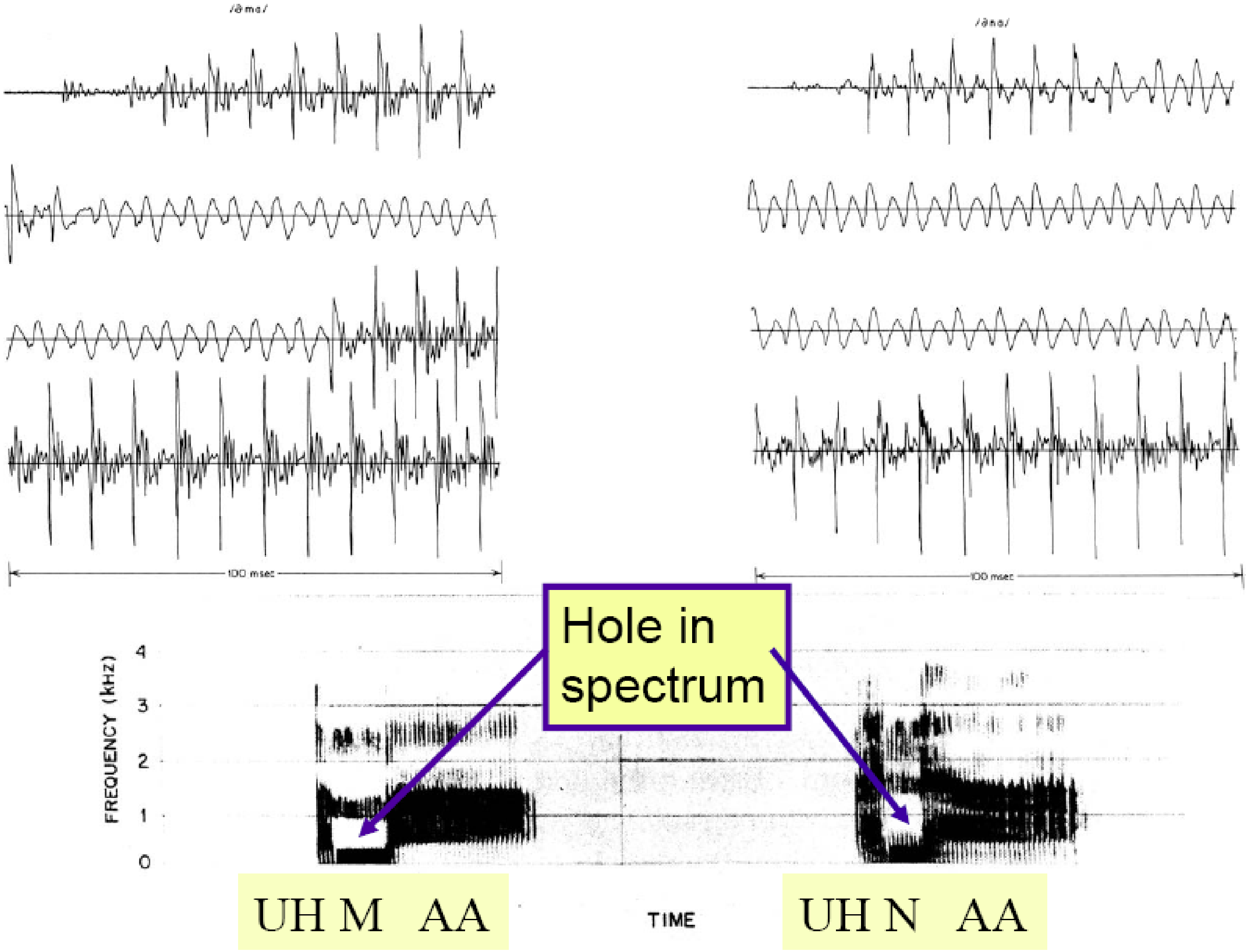
Manner: nasal
Place: bilabial (m), alveolar (n), velar(ng)

Nasal Production

- Velum lowering results in airflow through nasal cavity
- Consonants produced with closure in oral cavity
- Nasal murmurs have similar spectral characteristics

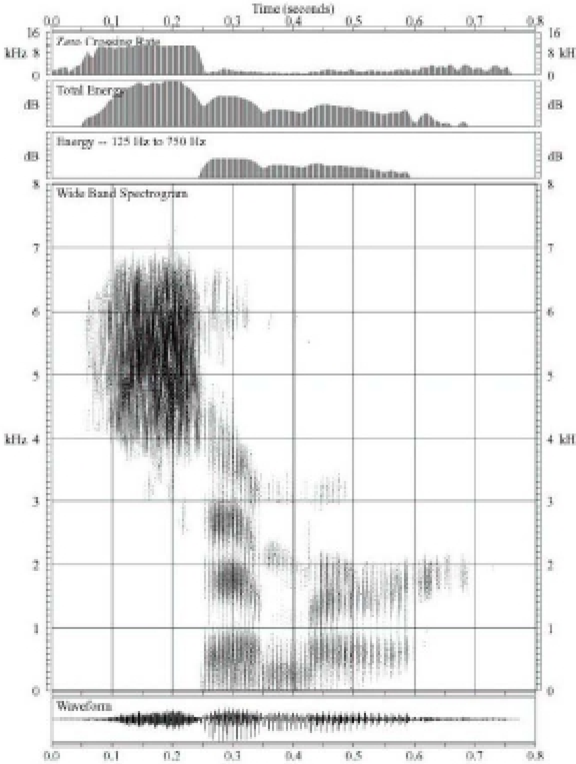


Nasal Sounds

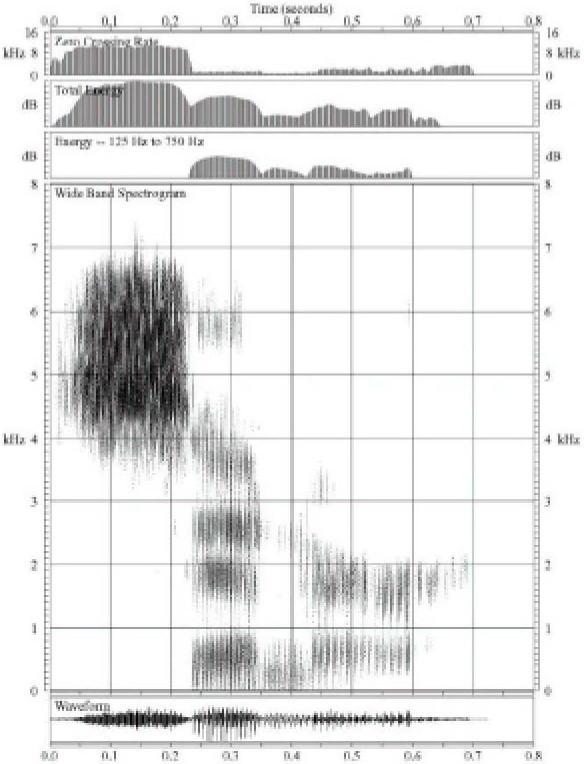


Hole in spectrum

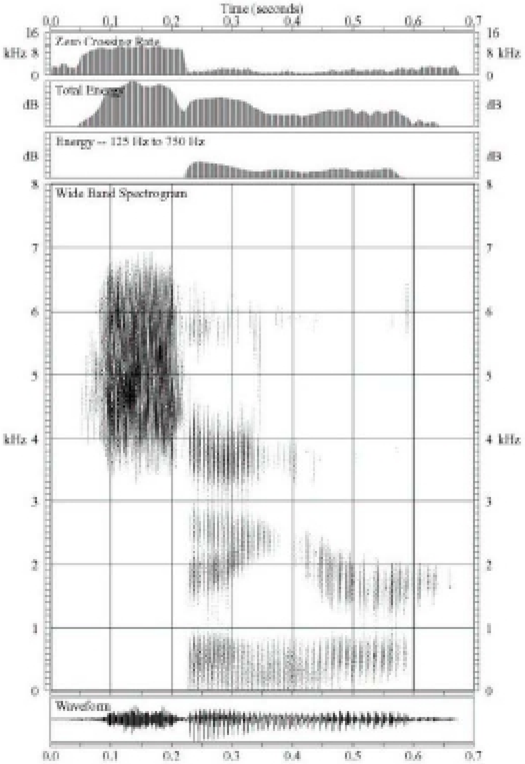
Nasal Spectrogram



simmer
/sɪmɝ/



sinner
/sɪnɝ/



singer
/sɪŋɝ/

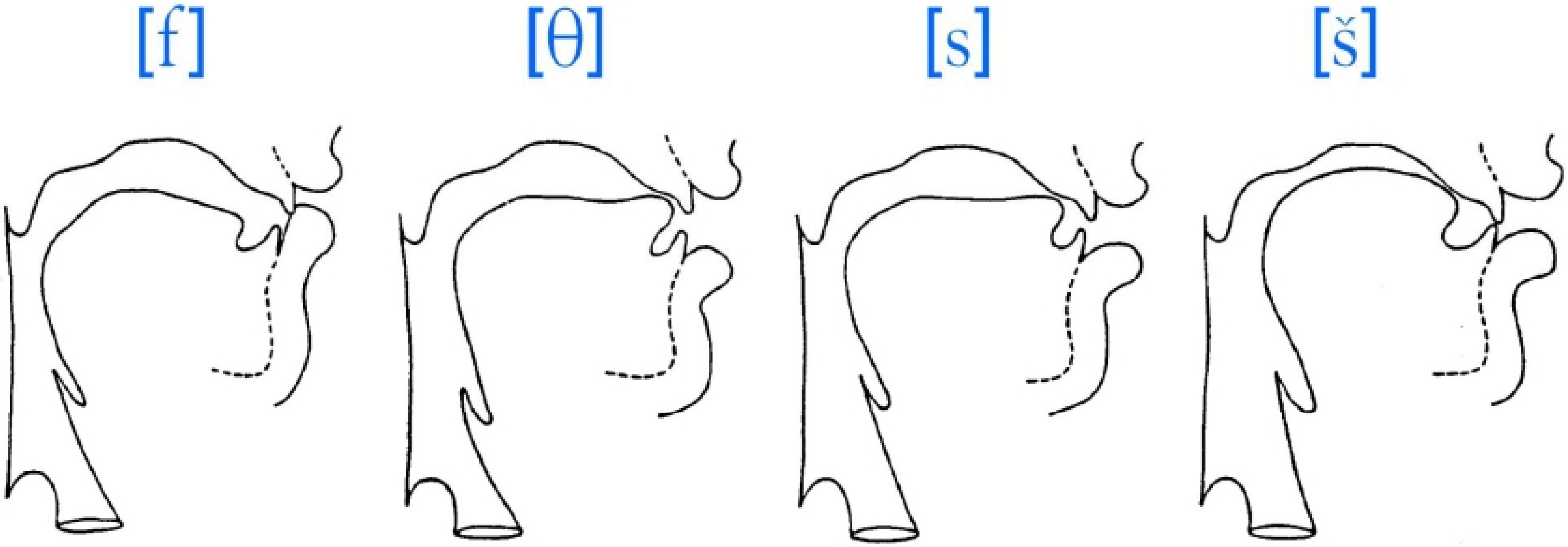
Unvoiced Fricatives

- Consonant sounds /F/, /TH/, /S/, /SH/
  - produced by exciting vocal tract by steady air flow which becomes turbulent in region of a constriction in the vocal tract
    - /F/ constriction near the lips
    - /TH/ constriction near the teeth
    - /S/ constriction near the middle of the vocal tract
    - /SH/ constriction near the back of the vocal tract
  - noise source at constriction => vocal tract is separated into two cavities
  - sound radiated from lips – front cavity
  - back cavity traps energy and produces antiresonances (zeros of transmission)

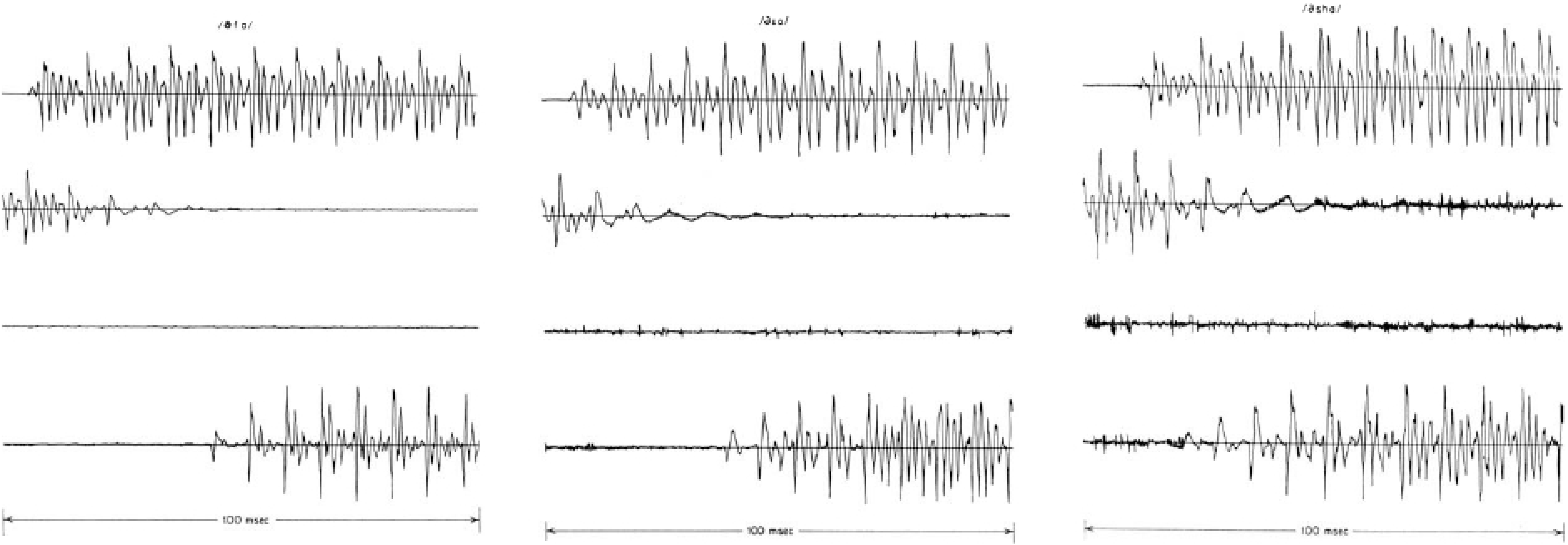
uh-{f,th,s,sh}-a

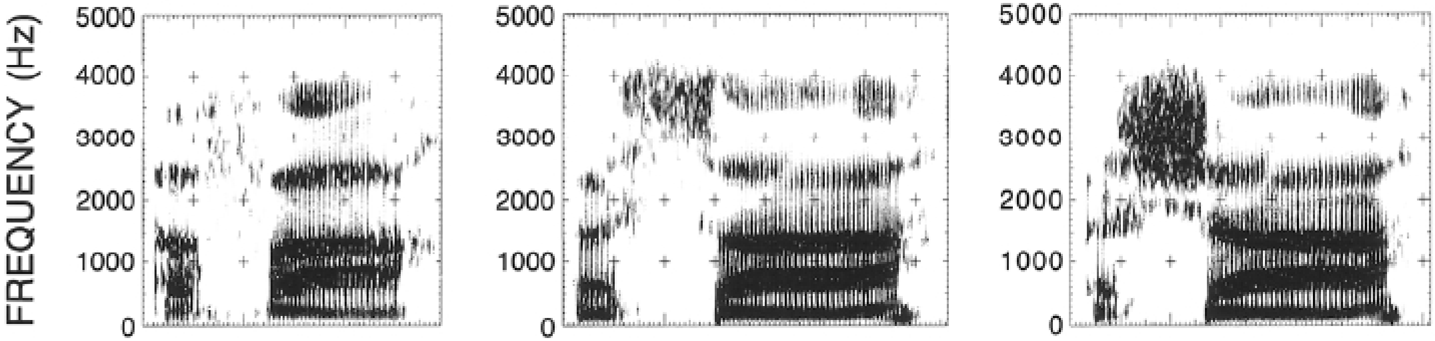
Manner: fricative
Place: labiodental (f), dental (th), alveolar (s), palatal (sh)

Unvoiced Fricative Production



Unvoiced Fricatives



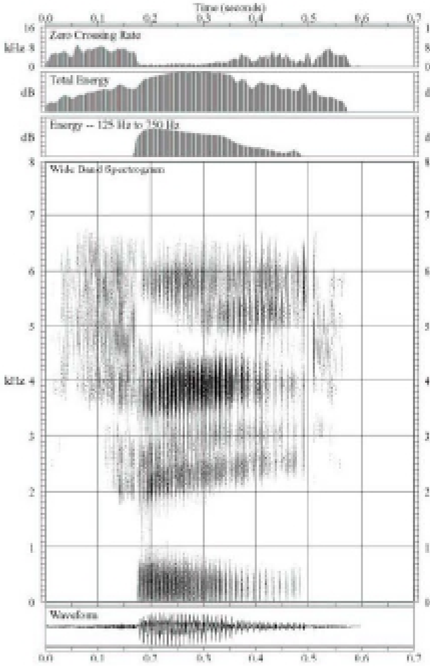


UH F AA

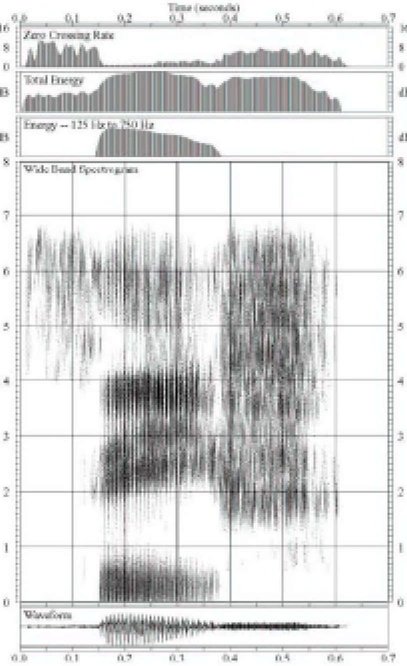
UH S AA

UH SH AA

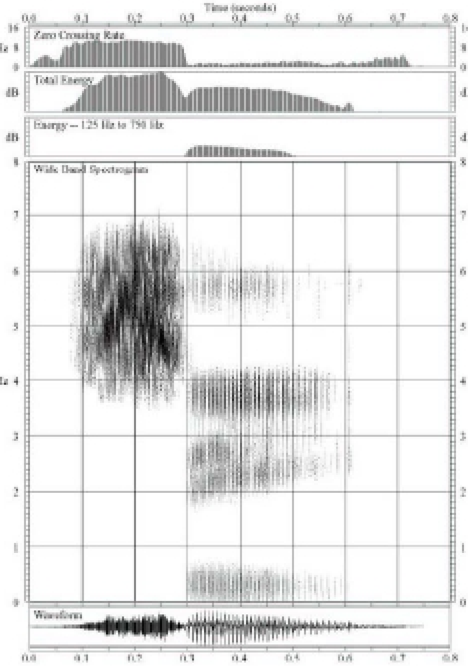
Unvoiced Fricative Spectrograms



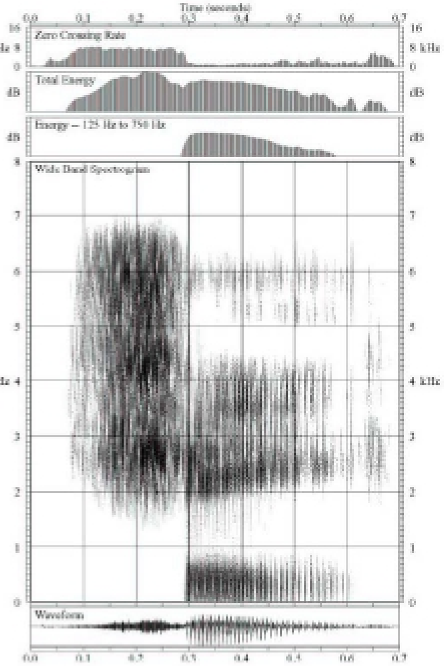
fee
/fi^y/



thief
/θi^{y}f/



see
/si^y/



she
/ši^y/

Voiced Fricatives

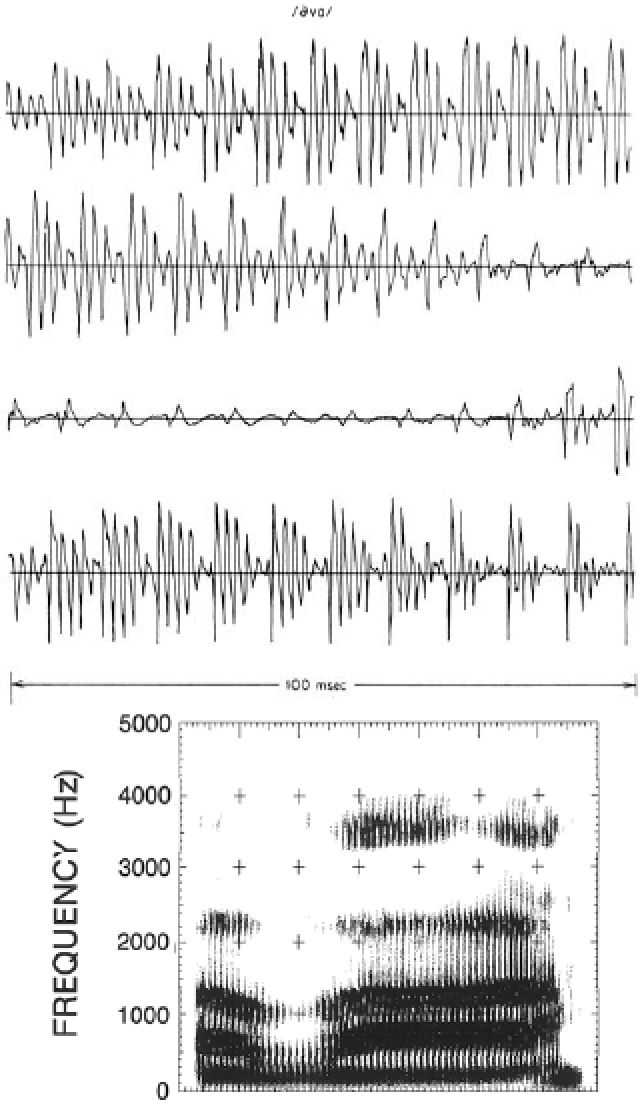
- Sounds /V/,/DH/, /Z/, /ZH/
  - place of constriction same as for unvoiced counterparts
  - two sources of excitation; vocal cords vibrating producing semi-periodic puffs of air to excite the tract; the resulting air flow becomes turbulent at the constriction giving a noise-like component in addition to the voiced-like component

🔊 🔊 🔊 🔊

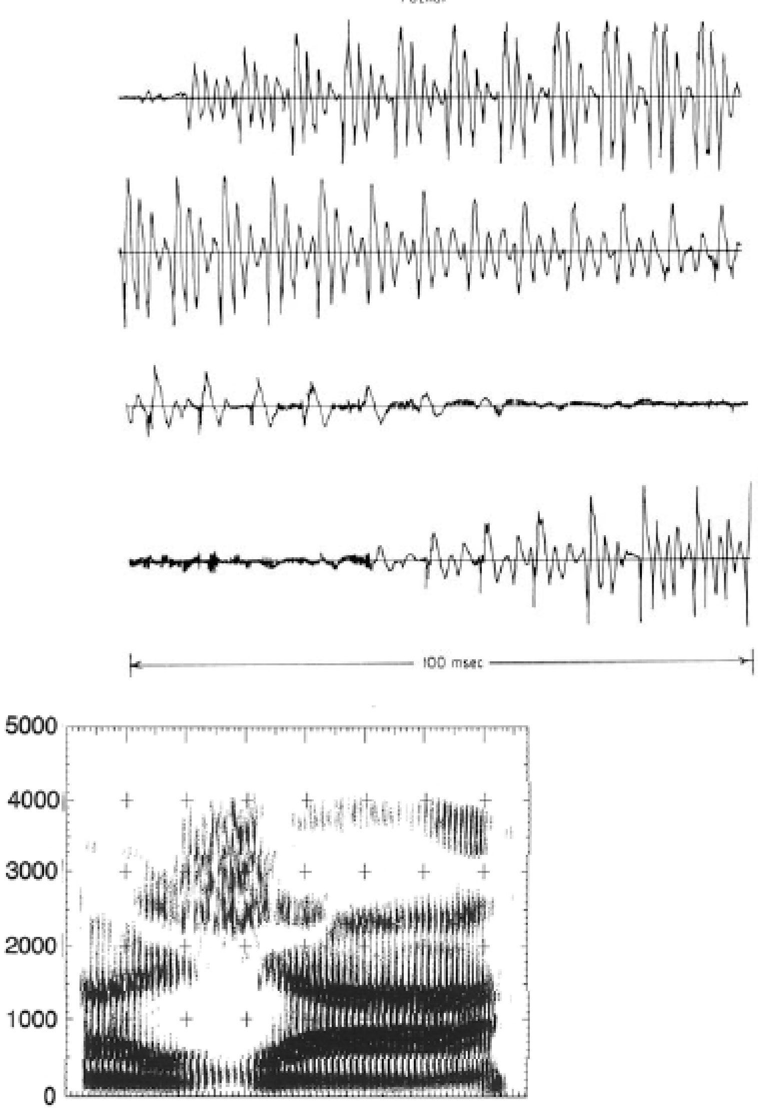
uh-{v,dh,z,zh}-a

Manner: fricative
Place: labiodental (v), dental (dh), alveolar (z), palatal (zh)

Voiced Fricatives



UH V AA



UH ZH AA

Voiced and Unvoiced Stop Consonants

- sounds-/B/, /D/, /G/ (voiced stop consonants) and /P/, /T/ /K/ (unvoiced stop consonants)
  - voiced stops are transient sounds produced by building up pressure behind a total constriction in the oral tract and then suddenly releasing the pressure, resulting in a pop-like sound
    - /B/ constriction at lips
    - /D/ constriction at back of teeth
    - /G/ constriction at velum
  - no sound is radiated from the lips during constriction => sometimes sound is radiated from the throat during constriction (leakage through tract walls) allowing vocal cords to vibrate in spite of total constriction
  - stop sounds strongly influenced by surrounding sounds
  - unvoiced stops have no vocal cord vibration during period of closure => brief period of frication (due to sudden turbulence of escaping air) and aspiration (steady air flow from the glottis) before voiced excitation begins

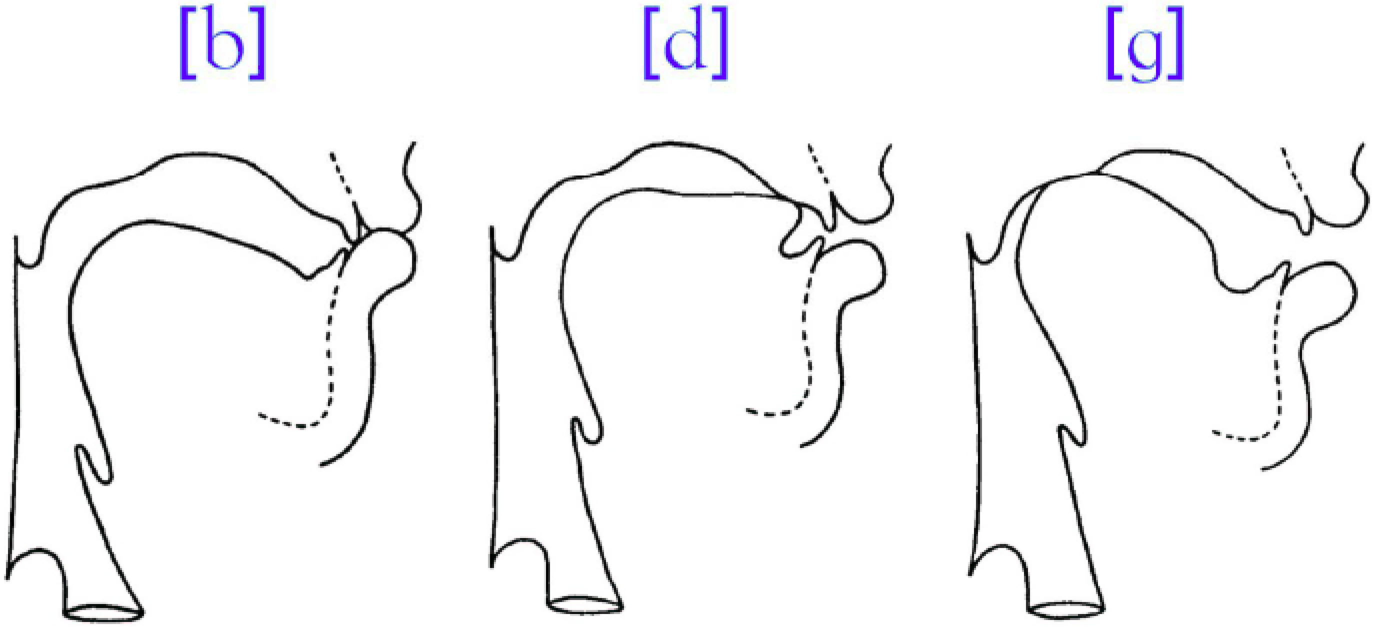
uh-{b,d,g}-a

uh-{p,t,k}-a

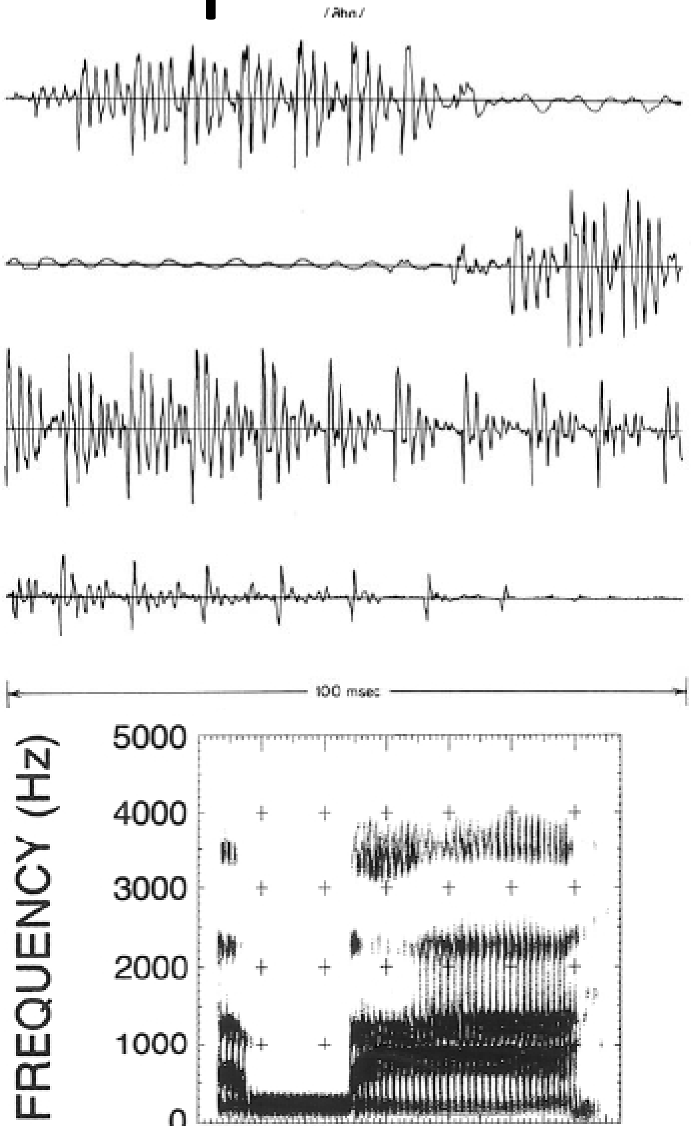
Manner: stop
Place: bilabial (b,p), alveolar (d,t), velar (g, k)

Stop Consonant Production

- Complete closure in the vocal tract, pressure build up
- Sudden release of the constriction, turbulence noise
- Can have periodic excitation during closure

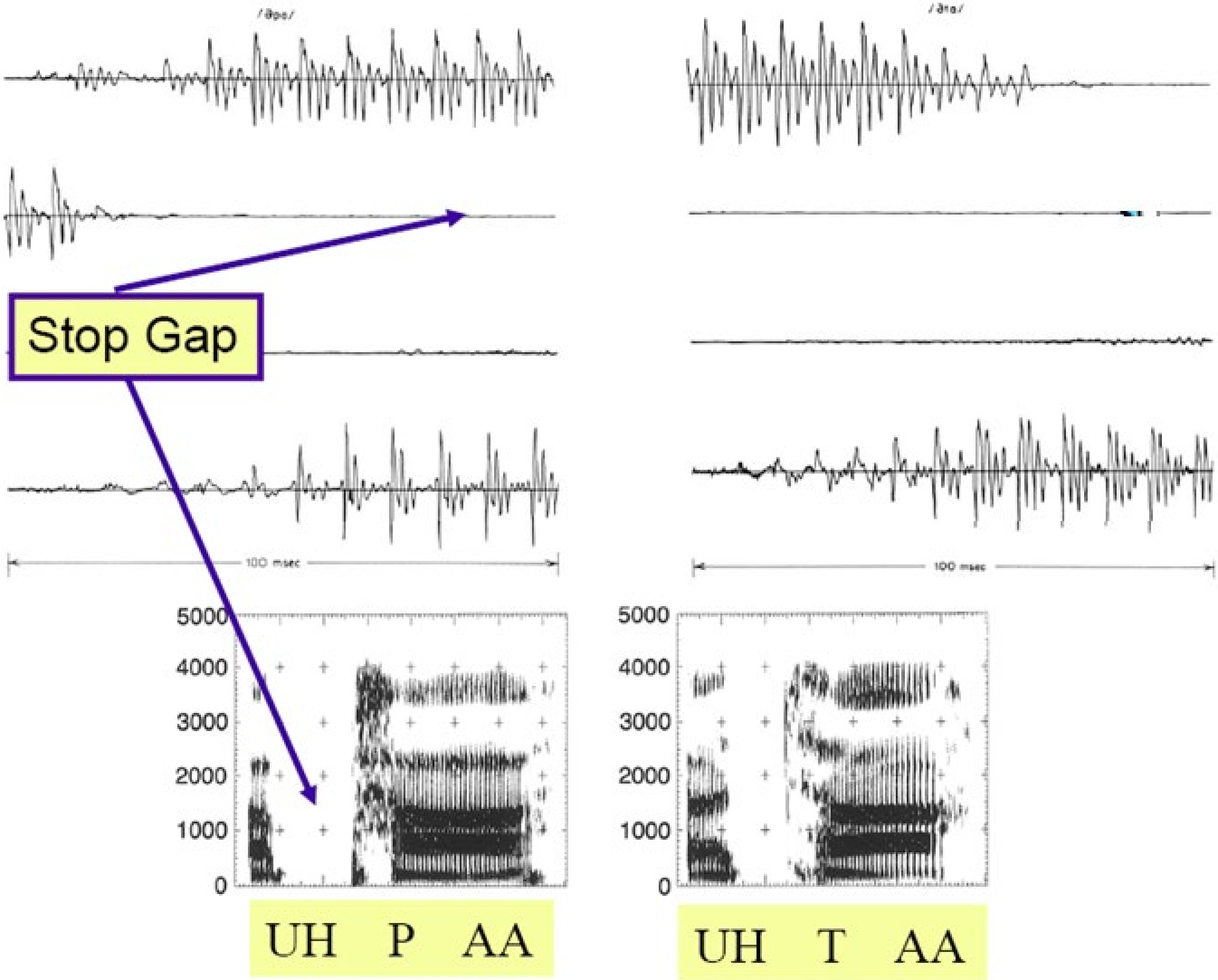


Voiced Stop Consonant



UH B AA

Unvoiced Stop Consonants



Stop Gap

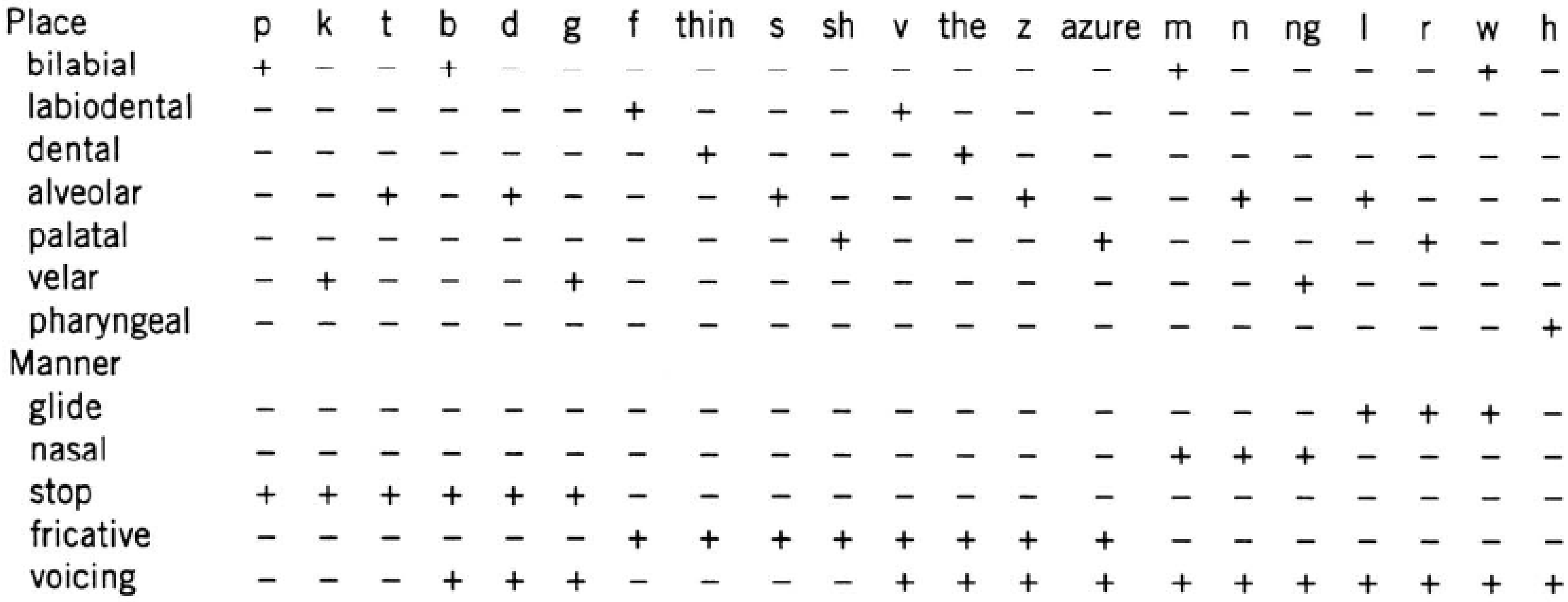
Affricates and Whisper

- Affricates
  - Dynamical sound
  - Can be modeled as the concatenation of a stop and a fricative
  - /CH/ = /T/ + /SH/
  - /JH/ = /D/ + /ZH/
- Whisper /H/
  - Produced by exciting the vocal tract by a steady airflow
  - Without the vocal cords vibrating, but with turbulent flow being produced at the glottis
  - The characteristics of /H/ are invariably those of the vowel that follows /H/

🔊 🔊 🔊

uh-{ch,jh,h}-a

Distinctive Phoneme Features

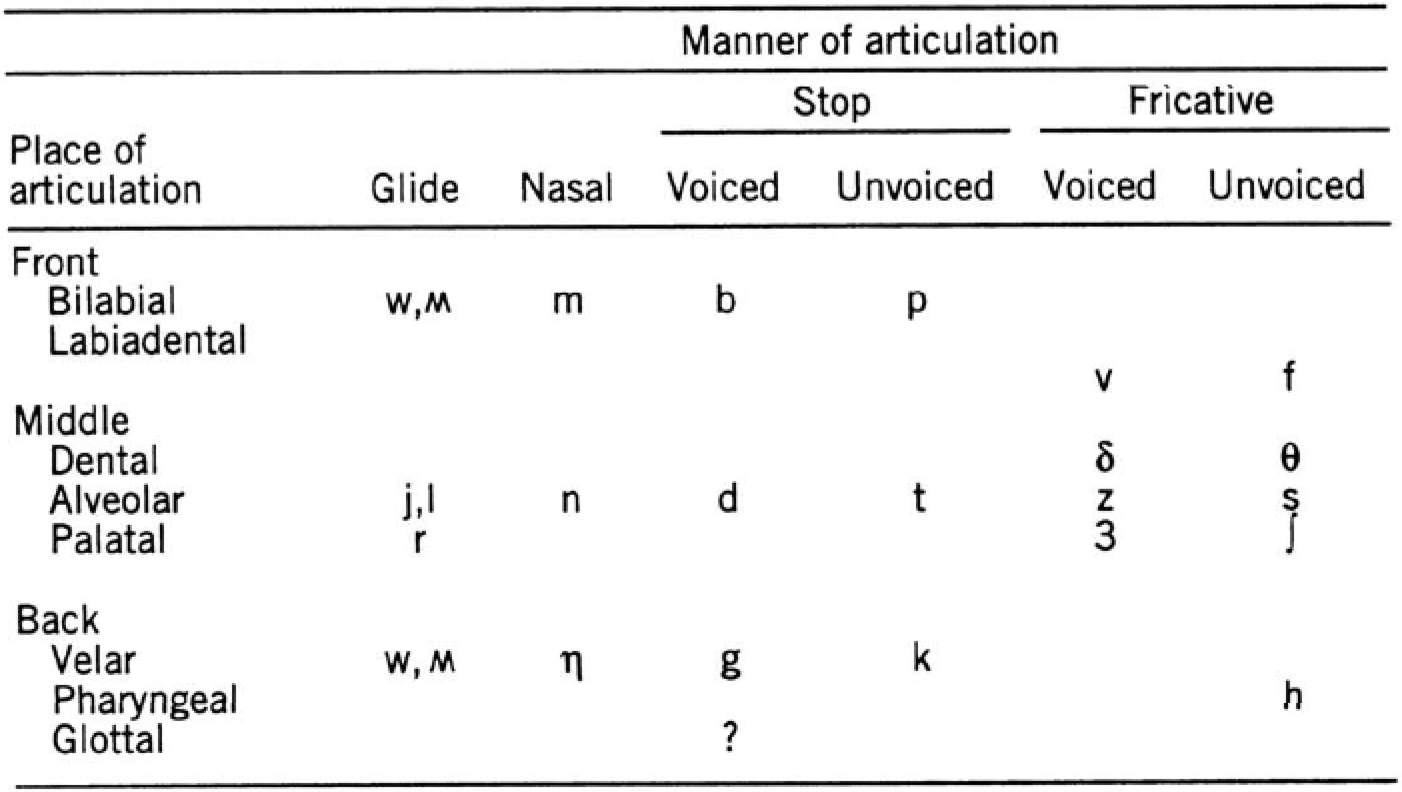


Place	p	k	t	b	d	g	f	thin	s	sh	v	the	z	azure	m	n	ng	l	r	w	h
bilabial	+	–	–	+	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	+	–
labiodental	–	–	–	–	–	–	+	–	–	–	+	–	–	–	–	–	–	–	–	–	–
dental	–	–	–	–	–	–	–	+	–	–	–	+	–	–	–	–	–	–	–	–	–
alveolar	–	–	+	–	+	–	–	–	+	–	–	–	+	–	–	+	–	+	–	–	–
palatal	–	–	–	–	–	–	–	–	–	+	–	–	–	+	–	–	–	–	+	–	–
velar	–	+	–	–	–	+	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–
pharyngeal	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+
Manner																					
glide	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	–
nasal	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	–	–	–	–
stop	+	+	+	+	+	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
fricative	–	–	–	–	–	–	+	+	+	+	+	+	+	+	–	–	–	–	–	–	–
voicing	–	–	–	+	+	+	–	–	–	–	+	+	+	+	+	+	+	+	+	+	+

FIGURE 17.7 Binary distinctive feature set of Jakobson et al. From [10].

- the brain recognizes sounds by doing a distinctive feature analysis from the information going to the brain
- the distinctive features are somewhat insensitive to noise, background, reverberation => they are robust and reliable

Distinctive Features



			Manner of articulation			
			Stop		Fricative	
Place of articulation	Glide	Nasal	Voiced	Unvoiced	Voiced	Unvoiced
Front						
Bilabial	w,ʍ	m	b	p		
Labiadental					v	f
Middle						
Dental					δ	θ
Alveolar	j,l	n	d	t	z	s
Palatal	r				3	ʃ
Back						
Velar	w, ʍ	η	g	k		
Pharyngeal						h
Glottal			?			

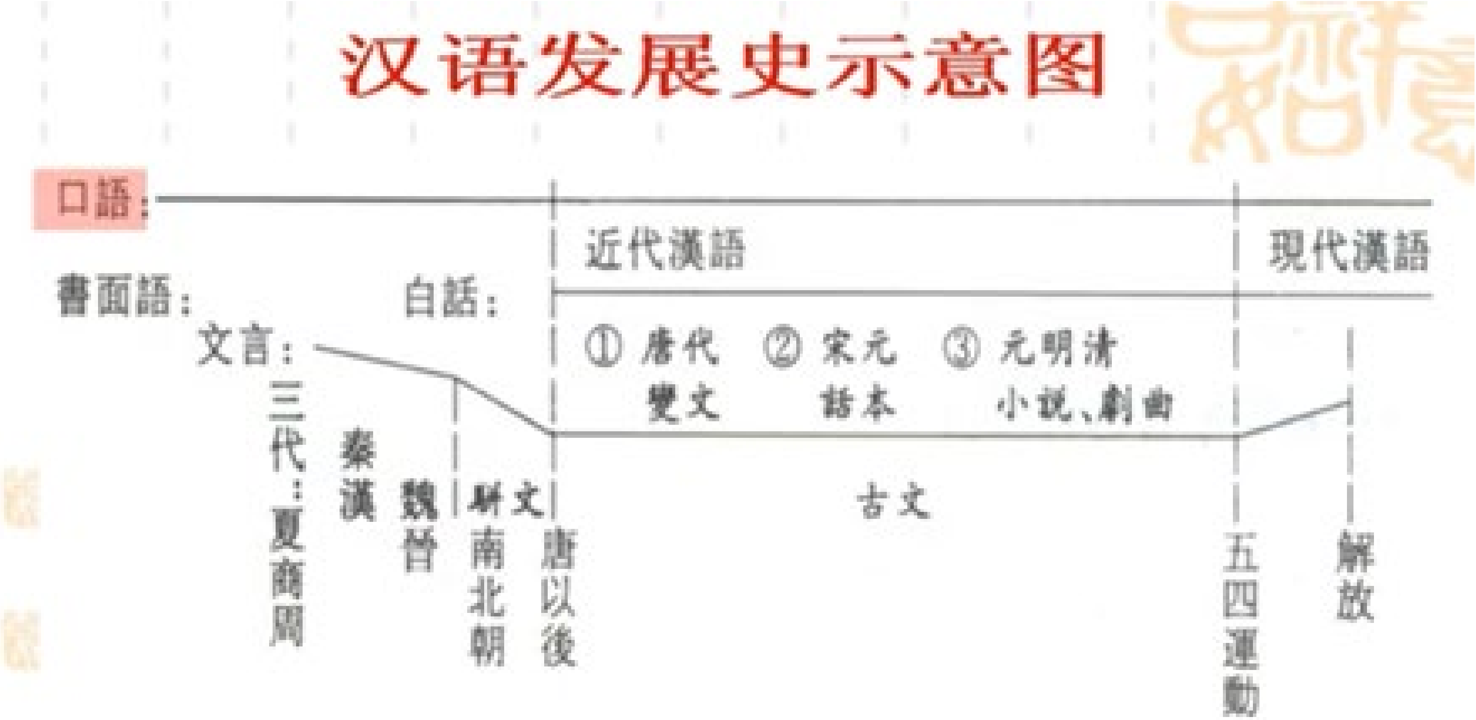
FIGURE 17.8 Articulatory classification of consonants. From [15].

- place and manner of articulation completely define the consonant sounds, making speech perception robust to a range of external factors

中文普通话的韵母与声母

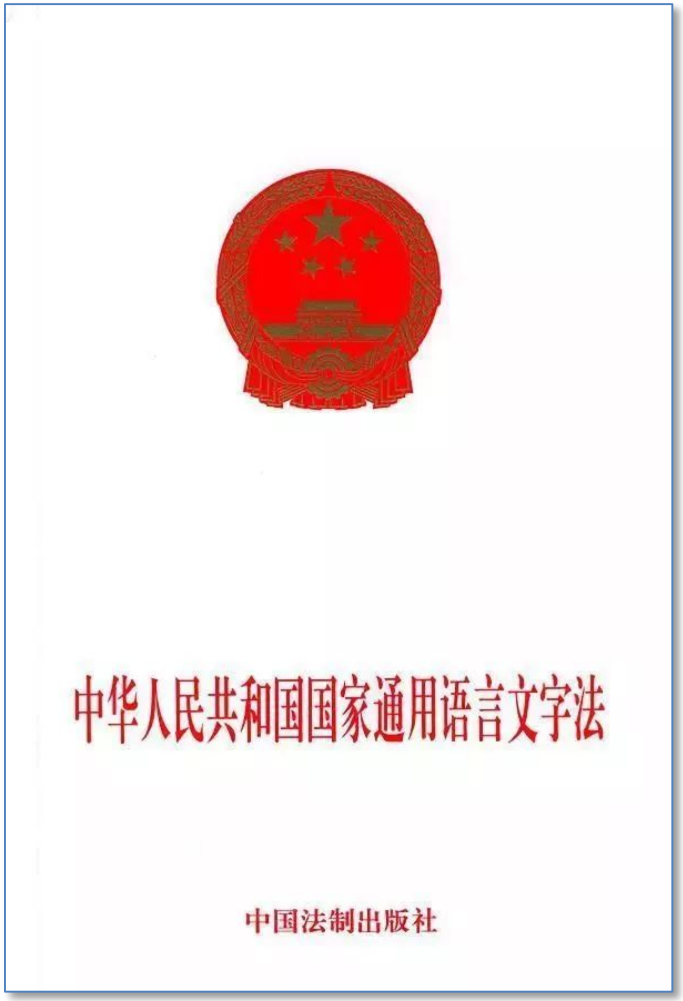
普通话与方言

- 汉语是我国使用人数最多的语言，也是世界上使用人数最多的语言，是联合国六种正式工作语言之一
- 汉语是我国汉民族的共同语，我国除占总人口91.51%的汉族使用汉语外，有些少数民族也转用或兼用汉语



普通话与方言

- 现代汉语有标准语（普通话）和方言之分
- 普通话以北京语音为标准音、以北方话为基础方言、以典范的现代白话文著作为语法规范
- 2000年10月31日颁布的《中华人民共和国国家通用语言文字法》确定普通话为国家通用语言

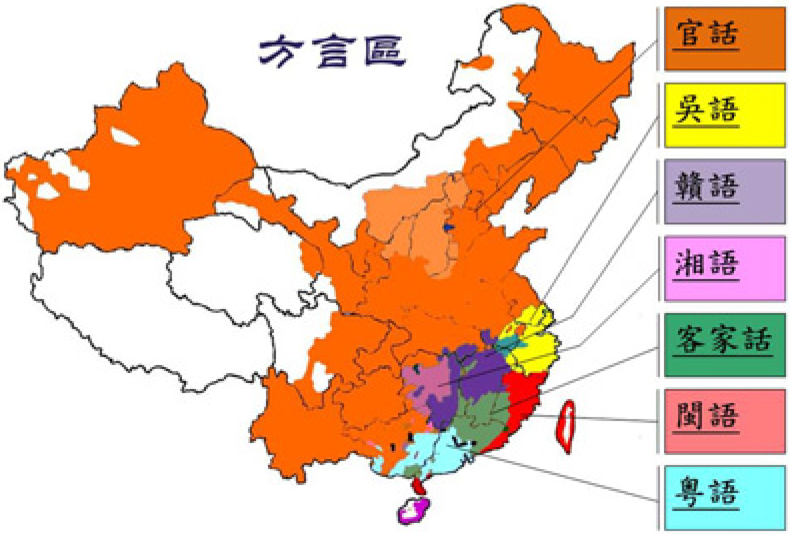


中华人民共和国国家通用语言文字法

中国法制出版社

普通话与方言

- 言汉语方言通常分为七大方言：北方方言、吴方言、湘方言、赣方言、客家方言、粤方言、闽方言。各方言区内又分布着若干次方言和许多种土语。其中使用人数最多的北方方言分为北方官话、西北官话、西南官话、下江官话四个次方言。



https://humanum.arts.cuhk.edu.hk/Lexis/lexi-mf/dialect.php

韵母和声母

- 汉字音节中开头的辅音音素叫声母；韵母是声母后面的音素部分。

- 元音和辅音：对音素自身性质的分析结果
- 声母和韵母：对汉语音节结构的分析结果

韵母

- 汉语普通话中，每个音节都必须有韵母

- 韵母共有38个
  - 8个单韵母
  - 14个复韵母
  - 16个鼻韵母

- 单韵母
  - /a/ /i/ /u/ /v/ /ii/ /iii/ /e/ /o/
  - 单韵母在单独发音时，发音器官的形状基本保持不变
- 复韵母
  - /ai/ /ei/ /au/ /ou/ /ia/ /ie/ /ua/ /uo/ /ve/ /er/
  - /iao/ /iou/ /uai/ /uei/
  - 在发音过程中存在频谱特征的动态变化

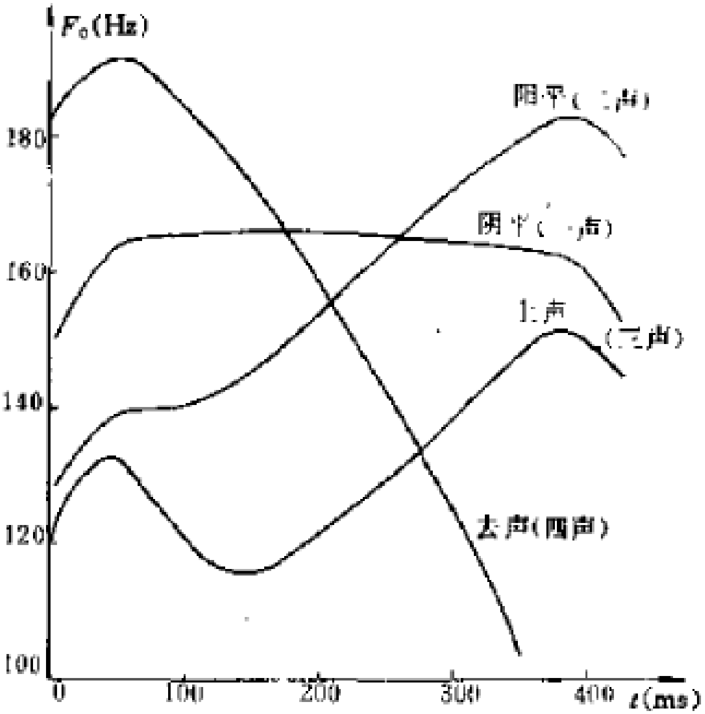
- 鼻韵母
  - 以/n/ 或 /ng/ 结尾的韵母
  - /an/ /ian/ /uan/ /van/ /en/ /in/ /un/ /vn/
  - /ang/ /iang/ /uang/ /eng/ /ing/ /eng/ /ong/ /iong/
  - 发音时存在鼻腔和口腔的耦合，对于主要元音的发音特征有较大影响

声母

- 21个
- 发音时器官的状态变化较大，动态特性很强
- 依据阻挡的具体情况对声母进行分类
  - 塞音：声道完全阻塞 /b/ /d/ /g/ /p/ /t/ /k/
  - 擦音：声道阻碍的缝隙面积很小 /s/ /f/ /x/
  - 通音：声道阻碍的缝隙面积大一些 /l/
  - 鼻音：浊辅音 /m/ /n/

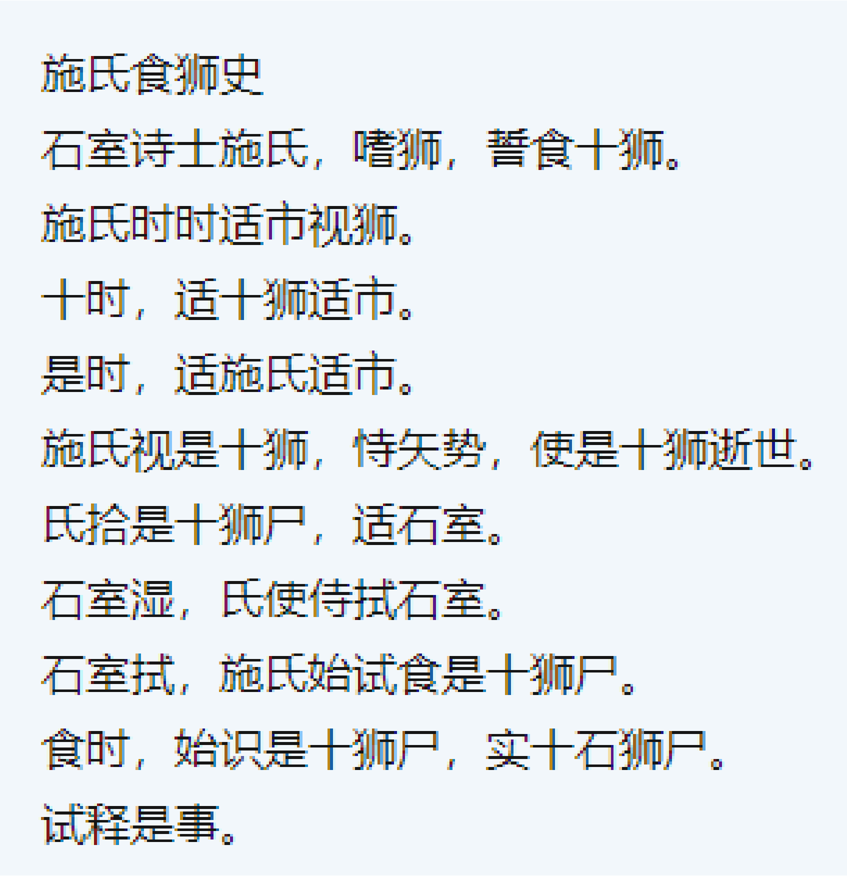
声调

- 汉语普通话中有5种声调
  - 阴平、阳平、上声、去声、轻声



- 上声变调
  - “555”

声调



施氏食狮史

石室诗士施氏，嗜狮，誓食十狮。

施氏时时适市视狮。

十时，适十狮适市。

是时，适施氏适市。

施氏视是十狮，恃矢势，使是十狮逝世。

氏拾是十狮尸，适石室。

石室湿，氏使侍拭石室。

石室拭，施氏始试食是十狮尸。

食时，始识是十狮尸，实十石狮尸。

试释是事。

赵元任（1892年11月3日—1982年2月24日），现代著名学者、语言学家、音乐家，中国现代语言学先驱，被誉为“中国现代语言学之父”

Summary

- sounds of the English language—phonemes, syllables, words
- phonetic transcriptions of words and sentences — coarticulation across word boundaries
- vowels and consonants — their roles, articulatory shapes, waveforms, spectrograms, formants
- distinctive feature representations of speech