

Digital Audio and Speech Processing

(Sayısal Ses ve Konuşma İşleme)

Prof. Dr. Nizamettin AYDIN

naydin@yildiz.edu.tr

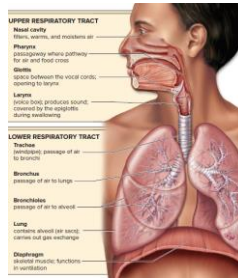
nizamettinaydin@gmail.com

<http://www3.yildiz.edu.tr/~naydin>

Speech Production

Anatomy of the speech organs

- The speech organs can be broadly divided into three groups:



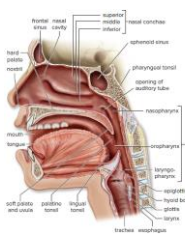
- Lungs:**
 - serve as a power supply
 - provides airflow to the larynx
- Larynx (Vocal chords):**
 - modulate the airflow into either a periodic sequence of puffs or a noisy airflow source
 - A third type of source is impulsive
- Vocal tract:**
 - converts modulated airflow into spectrally colored signal

1

2

The vocal tract

- The vocal tract can be divided into:



- Velum (soft palate):**
 - controls airflow through the nasal cavity
 - In its open position is used for nasals
- Hard palate:**
 - hard surface at the roof of the mouth.
 - When tongue is pressed against it, leads to consonants
- Tongue:**
 - Away from the palate produces vowels;
 - close to or pressing the palate leads to consonants
- Teeth:**
 - used to brace the tongue for certain consonants

Sasannah Nelson Longenbaker - Mader's understanding human anatomy & physiology (2020)

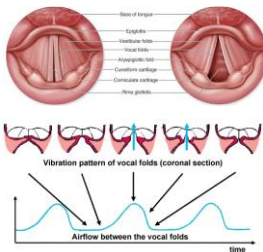
- Lips:**
 - can be rounded or spread to shape consonant quality, or closed completely to produce certain consonants

3

Vocal Cords

Tensed Vocal Cords – Lax Vocal Cords

Ready to Vibrate Open for Breathing



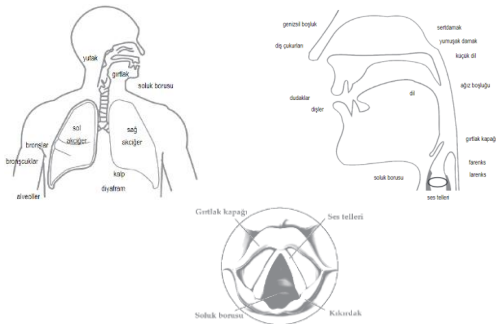
- Two masses of **flesh, ligament and muscle** across the **larynx**
- Fixed at the front of the larynx but free to move at the back and sides

4

Vocal Cords

- Can be in one of three primary states

- Breathing:**
 - Glottis is wide, muscles are relaxed, and air flows with minimal obstruction
- Voicing:**
 - vocal folds are tense and are brought up together.
 - Pressure builds up behind, leading to an oscillatory opening of the folds ([video](#))
- Unvoiced:**
 - similar to breathing state, but folds are closer, which leads to turbulences

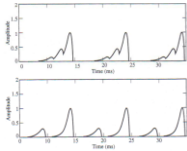


5

6

Vocal Cords

- Other (minor) forms of voicing include
 - Hoarse voice:**
 - voicing period (pitch) jitters, as what results from laryngitis or a cold
 - Breathy voice:**
 - aspiration occurs simultaneously while voicing (audio)
 - Creaky voice:**
 - vocal folds are very tense and only a portion oscillates.
 - Result is a harsh sounding voice (audio)

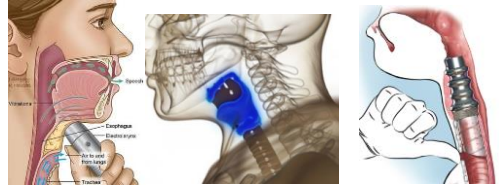


- Vocal fry:**
 - folds are very relaxed, which leads to secondary glottal pulses (video)
- Diplophonic:**
 - secondary pulses occur, but during the closed phase

7

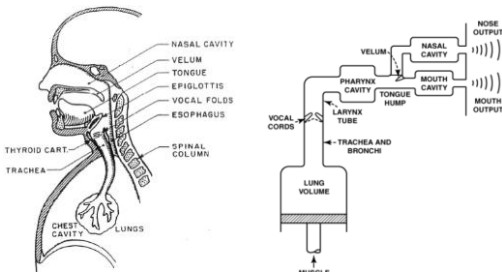
Artificial Larynx

- An electromechanical device that enables a person after laryngectomy to produce speech.
 - When the device is placed against the region of the laryngectomy a buzzing sound is made that can be converted into simulated speech by movements of the lips, tongue, and glottis. (video)
 - Called also **electrolarynx**



8

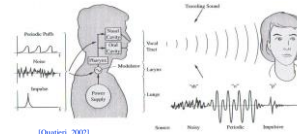
Schematic View and Model of Vocal Organs



9

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 (windpipe), the tensed vocal cords within the larynx 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

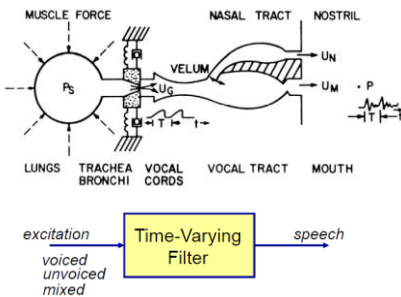


(Quatieri, 2002)

10

Models of speech production

- Abstractions of Physical Model



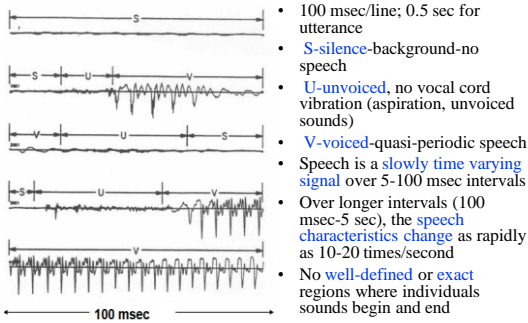
11

The Speech Signal

- Speech is a **sequence** of ever changing sounds
- Sound properties are highly dependent on **context** (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)

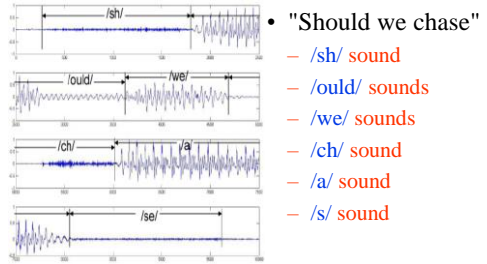
12

Speech Waveforms and Spectra



13

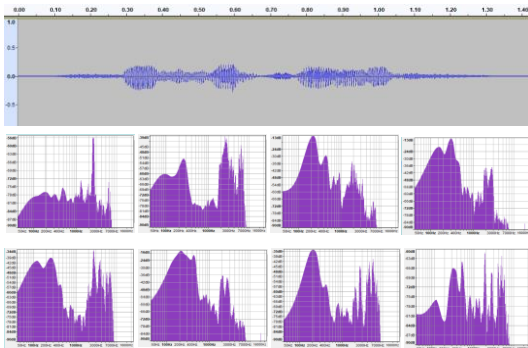
Speech Sounds



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

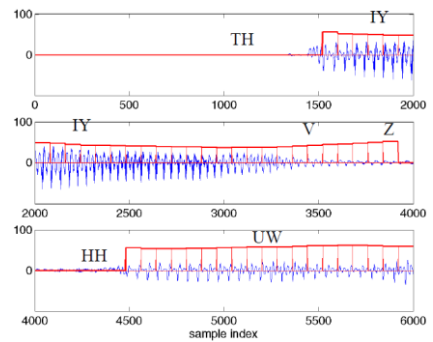
14

Speech Sounds



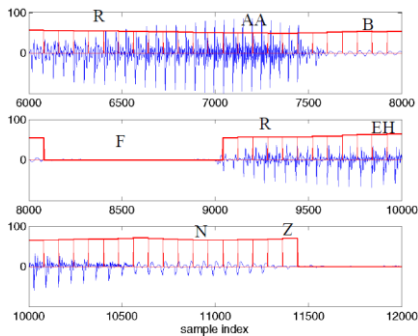
15

Estimate of Pitch Period



16

Estimate of Pitch Period



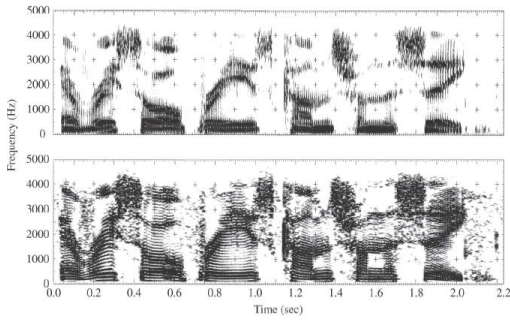
17

Spectrogram Properties

- **Speech Spectrogram**
 - sound intensity versus time and frequency
 - **wideband spectrogram**
 - spectral analysis on 15 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 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

18

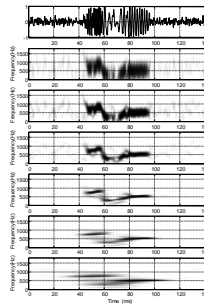
Wideband and Narrowband Spectrograms



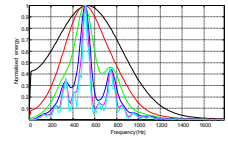
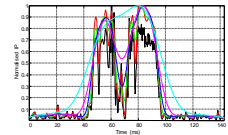
19

Wideband and Narrowband Spectrograms

TFDs with 16, 32, 64, 128, 256, 512 points windowing



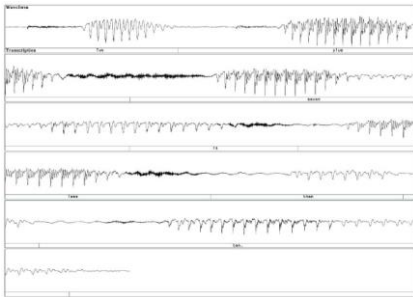
Normalised IP and energy with 16(black), 32(red), 64(green), 128(blue), 256(magenta), 512(cyan) points windowing



20

Speech Sentence Waveform

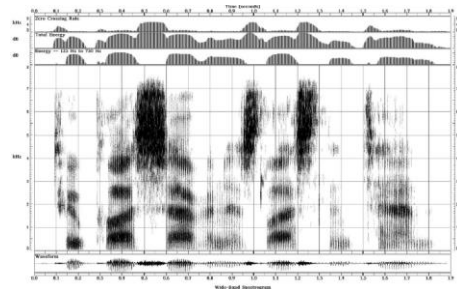
- "Two plus seven is less than ten"



21

Speech Wideband Waveform

- "Two plus seven is less than ten"



22

Acoustic theory of speech production

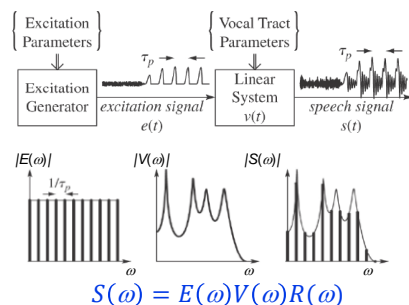
- Speech occurs when a source signal passing through the **glottis** is modified by the **vocal tract** acting as a **filter**
- Models of this kind are generally known as **source-filter** models
- Using the theory of linear time invariant (LTI) systems, the overall process can be modeled in the **z-domain** as

$$S(z) = E(z)P(z)O(z)R(z)$$
 - where $E(z)$ is the glottal source, and $P(z)$, $O(z)$, $R(z)$ are the transfer functions at the **pharynx**, **oral cavity** and **lips**
- This model can be simplified as

$$S(z) = E(z)V(z)R(z)$$
 - $P(z)$ and $O(z)$ are combined into a single **vocal-tract transfer function**, which represents the **filter** component of the model

23

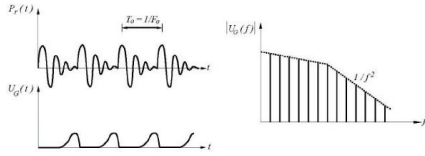
Source-System Model of Speech Production



24

Sound Source for Voiced Sounds

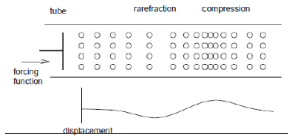
- Modelled as a volume velocity source at glottis



	F_0 ave (Hz)	F_0 min (Hz)	F_0 max (Hz)
Men	125	80	200
Women	225	150	350
Children	300	200	500

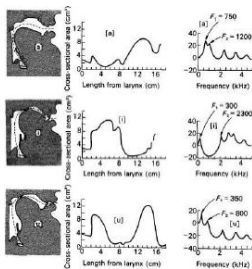
Acoustic waves

- The effect of a sound source causes air particles to move back and forth, so the wave spreads from the source
 - In some areas, particles come close together (compression) whereas in others they move further apart (rarefaction).
- Sound waves are longitudinal waves



Acoustic reflection and tube models

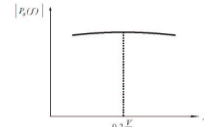
- If the area of the tube remains constant, the wave propagates through the tube



- However, if the area changes, then the impedance changes, which causes reflection, which in turn leads to standing waves, which then cause resonances
- Thus, the impedance pattern on the tube determines the resonance properties of the model

Sound Source for Unvoiced Sounds

- Turbulence noise is produced at a constriction in the vocal tract
 - Aspiration noise is produced at glottis
 - Frication noise is produced above the glottis
- Modelled as series pressure source at constriction, $P_s(\omega)$



V : Velocity at constriction D : Critical dimension = $\sqrt{\frac{4A}{\pi}} \approx \sqrt{A}$

Acoustic reflection and tube models

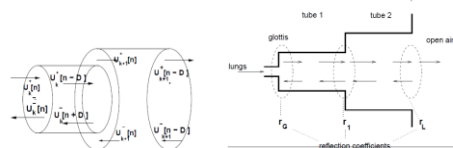
- Boundary conditions in the tube will determine how acoustic waves are reflected at the end of the tube
 - At certain frequencies, determined by the length of the tube and the speed of sound, the backward and forward waves will reinforce each other and cause resonances
- The volume velocity can be modelled at position x and time t as:

$$u(x, t) = u^+(t - x/c) - u^-(t + x/c)$$
 - u^+ , u^- , c : forward-backward-travelling wave, backward-travelling wave, the speed of sound
- The pressure becomes

$$p(x, t) = \frac{\rho c}{A} (u^+(t - x/c) + u^-(t + x/c))$$
 - $\frac{\rho c}{A}$: characteristic impedance of the tube
 - Notice how in this case the two waves add up as they meet at point x

Acoustic reflection and tube models

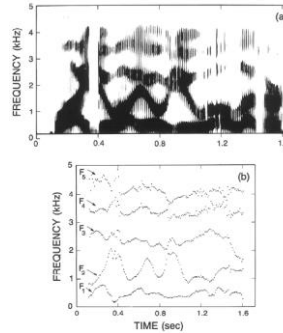
- The vocal tract can then be modeled as a series of short uniform tubes connected in series



Parametrization of Spectra

- Acoustic theory shows that the transfer function of energy from the excitation source to the output can be described in terms of the **natural frequencies** or **resonances** of the tube
- Resonances known as **formants** or **formant frequencies** for speech and they represent the frequencies that pass the most acoustic energy from the source to the output
 - Typically there are 3 significant formants below about 3500 Hz
 - Formants are a highly efficient compact representation of speech

Spectrogram and Formants



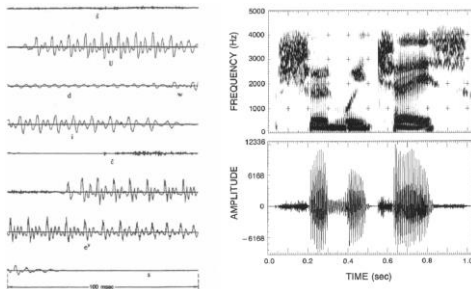
- "Why do i owe you a letter"
- Key Issue here is the reliability in estimating formants from the spectral data

31

32

Waveform and Spectrogram

- "Should we chase"



33

Acoustic Theory 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

34

English Speech Sounds

A Condensed List of Phonetic Symbols for American English

Phoneme	ARPAbet	Example	Phoneme	ARPAbet	Example
/ɪ/	IY	bit	/ɔ/	OX	slight
/ʊ/	UH	bit	/ɒ/	P	pot
/eɪ/	EY	bat	/t/	T	top
/e/	EH	bat	/k/	K	kit
/æ/	AE	bat	/b/	B	bat
/ɑ/	AA	bat	/d/	D	dot
/ʌ/	AH	but	/g/	G	get
/ɔ/	AO	bought	/h/	HH	hat
/oʊ/	OW	bought	/f/	F	fat
/ʊ/	UH	book	/θ/	TH	thing
/u/	UW	boot	/s/	S	sat
/ə/	AX	about	/ʃ/	SH	ship
/ɪ/	IC	moses	/v/	V	vet
/ɛr/	ER	bird	/dʒ/	DJ	jet
/ɛr/	AXR	buried	/z/	Z	zip
/ɪ/	AW	open	/ʒ/	ZH	azure
/ɪ/	AY	day	/tʃ/	CH	church
/ɪ/	OY	boy	/dʒ/	JH	judge
/ɪ/	Y	you	/w/	WH	which
/ɪ/	W	we	/l/	L	leaf
/ɪ/	R	rent	/ɛm/	EM	bottom
/ɪ/	L	let	/ɛn/	EN	bottom
/ɪ/	M	met	/ɛk/	DK	batper
/ɪ/	N	net	/ɔ/	O	(glottal stop)

- ARPABET representation

- 48 sounds
- 18 vowels/diphthongs
- 4 vowel-like consonants
- 21 standard consonants
- 4 syllabic sounds
- 1 glottal stop

- Pulmoner ünsüzlerin uluslararası sesbilim abecesindeki (IPA, 2005) gösterimi

	Çift Dudak	Dudak-dış	Dişil	Diş-yuvaneli	Artdiş-yuvaneli	Üstdamak	Damaksil	Artdamak	Küçük dil	Boğazsil	Gırtlak
Genizsil	m	nj		n		ɲ	ɟ	ŋ	ɳ		
Patlama	p b			t d		ʈ ɟ	ɕ ɟ	k ɡ	q ɢ		ʔ
Sürtünücü	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	ħ	ʕ	ɦ
Yanal		ɸ		ɹ		ɻ	ʝ	ɰ			
Akıcı				r							
Çarpmalı				ɾ							
Yanal sürtünücü				ʃ ʒ							
Yan daralmalı				l		ɭ	ɮ	ɮ			
Yanal çarpmalı				ɹ							

35

36

- Pulmoner olmayan ünsüzlerin uluslararası sesbilim abecesindeki (IPA, 2005) gösterimi

Şaklamlar	Ötümü	İçe Patlamalar	Dişe Patlamalar
Ø Çift Dudak	b Çift Dudak	p* Çift Dudak	
l Dişil	d Dişil/Diyuvasal	t* Dişil/Diyuvasal	
l Artdiyuvasal	ʃ Damaksil	k* Artdamaksil	
# Damak-diyuvasal	ʒ Artdamaksil	s* Sürünücü-diyuvasal	
ll Yanal-diyuvasal	ç Küçük Dil	* Ve benzeri	

37

- Diğer imlerin uluslararası sesbilim abecesindeki (IPA, 2005) gösterimi

ʍ Ötümü, dudak-artdamaksil, sürünücü
w Ötümü, dudak-artdamaksil, yanıl
ɥ Ötümü, dudak-damaksil, yanıl
ɰ Ötümü, gırtlak kapağı, sürünücü
ç Ötümü, gırtlak kapağı, sürünücü
ʔ Gırtlak kapağı, patlamalı
ɶ, ɷ Diyuvasil-damaksil, sürünücü
ɰ Ötümü, diyuvasil-yanıl, çarpmalı
ɸ Eşzamanlı f ve x

38

- Ayrıncı imlerin uluslararası sesbilim abecesindeki (IPA, 2005) gösterimi

ɰ Ötümü ɶ, ɷ	ɰ Hırıltılı b, ɸ	ɰ Dişil t, ʈ
ɰ Ötümü ɶ, ɷ	ɰ Boğuk b, ɸ	ɰ Dilucu t, ʈ
ɰ Soluklu ɰ, ɰ	ɰ Dil-damaksil t, ʈ	ɰ Dil palası t, ʈ
ɰ Daha yuvartak ɰ	ɰ Dudaksil ɰ, ɰ	ɰ Genizsil e
ɰ Az yuvartak ɰ	ɰ Damaksil ɰ, ɰ	ɰ* Geniz çıkışı ɰ
ɰ İleride ɰ	ɰ Artdamaksil ɰ, ɰ	ɰ Yanal çıkışı ɰ
ɰ Geride ɰ	ɰ Boğazsil ɰ, ɰ	ɰ Akustik olmayan çıkış ɰ
ɰ Merkezde ɰ	ɰ Artdamaksil ya da Boğazsil ɰ	
ɰ Orta-merkezde ɰ	ɰ Yüksek ɰ (ɰ = ötümü, diş yuvasal, sürünücü)	
ɰ Seslemel ɰ	ɰ Alpak ɰ (ɰ = ötümü, çift-dudak, yanıl)	
ɰ Seslemel olmayan ɰ	ɰ İleri Dil Kökü (ATR) ɰ	
ɰ r-söyleyişli ɰ, ɰ	ɰ Geri Dil Kökü (RTR) ɰ	

39

- Türkçenin Ses Dizgesi
- October 2017
- Edition: 2. Baskı-2020
- Publisher: Seçkin Yayınevi
- ISBN: 9789750245220
- <https://www.turcademy.com/tr/kitap/turkceenin-ses-dizgesi-ses-unluler-unsuzler-konusma-organlari-9789750245220>
- Kuramsal ve Uygulamalı Sesbilim
- May 2020
- Edition: 1. Baskı
- Publisher: Seçkin Yayınevi
- Editor: İpek Pınar Uzun
- ISBN: 9789750260124

40

Phonemes—Link Between Orthography and Speech

- Orthography
 - sequence of sounds
 - Larry → /l/ /æ/ /r/ /i/ /y/ (/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
- The bottom line is that we use a **phonetic code** as an intermediate representation of language, from either orthography or from waveforms or spectrograms

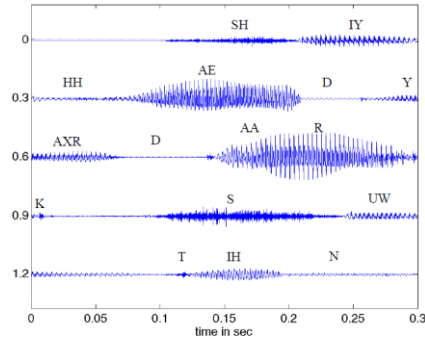
41

Phonetic Transcriptions (Orthography)

- 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/ /N/ /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)

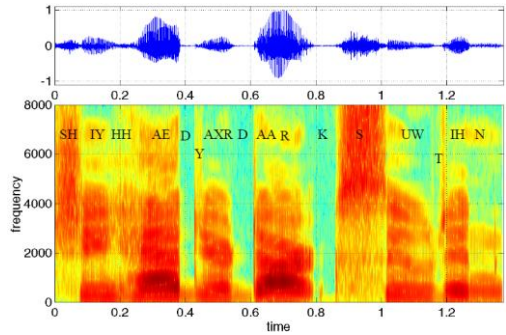
42

She had your dark suit in...



43

Wideband Spectrogram



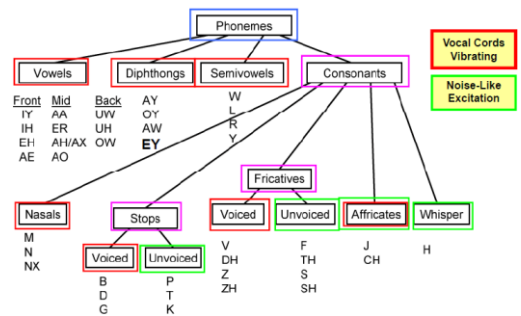
44

Reduced Set of English Sounds

- 39 sounds
 - 11 vowels (front, mid, back) classification based on tongue hump position
 - 4 diphthongs (vowel-like combinations)
 - 4 semi-vowels (liquids and glides)
 - 3 nasal consonants
 - 6 voiced and unvoiced stop consonants
 - 8 voiced and unvoiced fricative consonants
 - 2 affricate consonants
 - 1 whispered sound
- Look at each class of sounds to characterize their acoustic and spectral properties

45

Phoneme Classification Chart



46

Vowels and Consonants

- Vowels are longest duration sounds
 - letters in the alphabet such as A, E, I, O, and U
 - least context sensitive
 - Can be held indefinitely in singing and other musical works (opera)
 - Carry very little linguistic information
 - some languages do not display vowels in text (Hebrew, Arabic)
- Consonant are letters of the alphabet that represents a basic speech sound produced by obstructing the breath in the vocal tract
 - For example:
 - T is pronounced using the tongue (front part)
 - K is pronounced using the tongue (back part)
 - B is pronounced with the lips
 - H is pronounced in the throat
 - F is pronounced by forcing air through a narrow gap
 - M is pronounced using the nasal passage

47

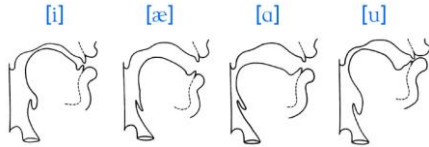
Vowels and Consonants

- 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_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 toward pay stayed essentially the same, with the scores of occupational employees slightly decreasing)

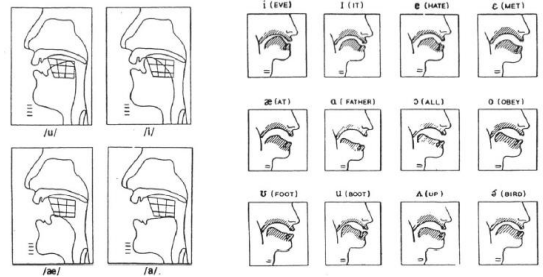
48

Vowel Production

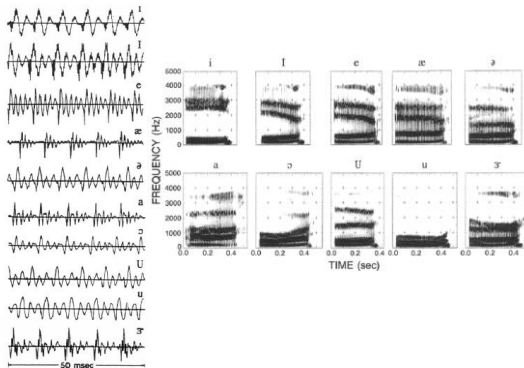
- No significant constriction in the vocal tract
- Usually produced with periodic excitation
- Acoustic characteristics depend on the position of the jaw, tongue, and lips



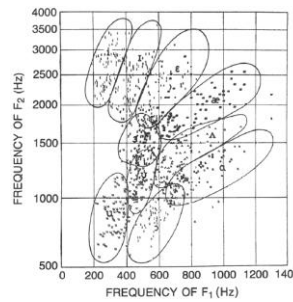
Vowel Articulatory Shapes



Vowel Waveforms & Spectrograms

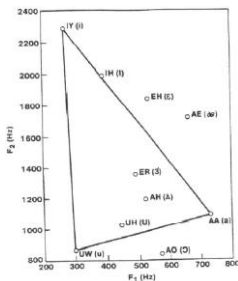


Vowel Formants



- 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

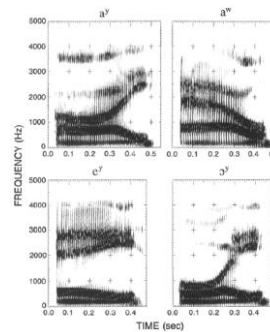
The Vowel Triangle



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

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	e	(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	ɜ	(bird)	490	1350	1690	

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)

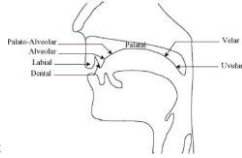
Distinctive Features

- Classify non-vowel/non-diphthong sounds in terms of distinctive features

– place of articulation

- Bilabial (lips)—p,b,m,w
- Labiodental (between lips and front of teeth)—f,v
- Dental (teeth)—th,dh
- Alveolar (front of palate)—t,d,s,z,n,l
- Palatal (middle of palate)—sh,zh,r
- Velar (at velum)—k,g,ŋ
- Pharyngeal (at end of pharynx)—h

Places of Articulation



– manner of articulation

- Glide—smooth motion—w,l,r,y
- Nasal—lowered velum—m,n,ŋ
- Stop—constricted vocal tract—p,t,k,h,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,ŋ,w,l,r
- Mixed source—both voicing and unvoiced—j,ç
- Whispered—h

55

Semivowels (Liquids and Glides)

- vowel-like in nature
 - called semivowels for this reason
- voiced sounds
 - w-l-r-y
- acoustic characteristics of these sounds are strongly influenced by context
 - unlike most vowel sounds which are much less influenced by context

- Manner:
 - glides
- Place:
 - bilabial (w), alveolar (l), palatal (r)

56

Nasal Consonants

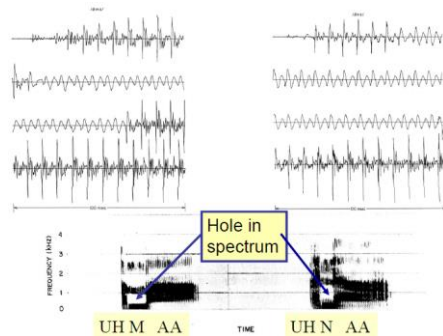
- The nasal consonants consist of /M/, /N/, and /NG/
 - nasals produced using glottal excitation => voiced sounds
 - vocal tract totally constricted at some point along the tract
 - velum lowered so sound is radiated at nostrils
 - constricted oral cavity serves as a resonant cavity that traps acoustic 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

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



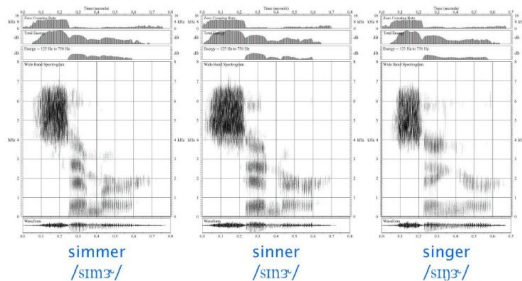
57

Nasal Sounds



58

Nasal Spectrograms

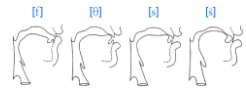


59

Unvoiced Fricatives

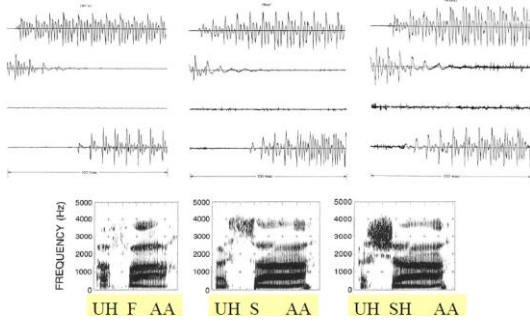
- Consonant sounds /F/, /TH/, /S/, /SH/
 - produced by exciting vocal tract by steady air flow which becomes turbulent in region of a constriction in the vocal tract
 - /F/ constriction near the lips
 - /TH/ constriction near the teeth
 - /S/ constriction near the middle of the vocal tract
 - /SH/ constriction 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)

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



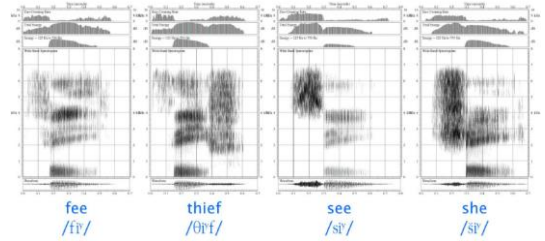
60

Unvoiced Fricative Sounds



61

Unvoiced Fricative Spectrograms



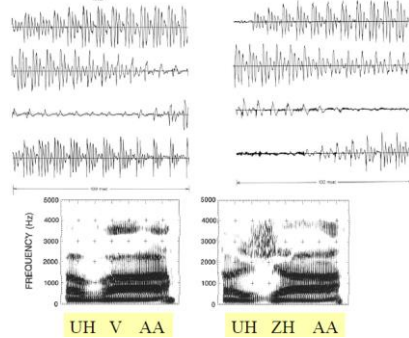
62

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

63

Voiced Fricatives



64

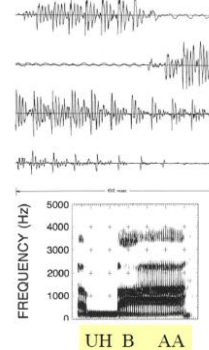
Voiced and Unvoiced Stop Consonants

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



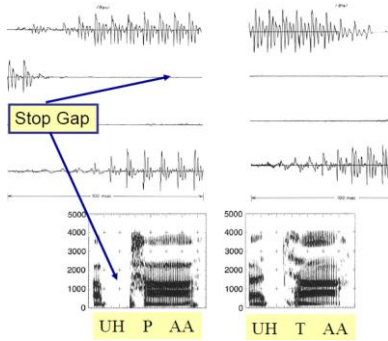
65

Voiced Stop Consonant



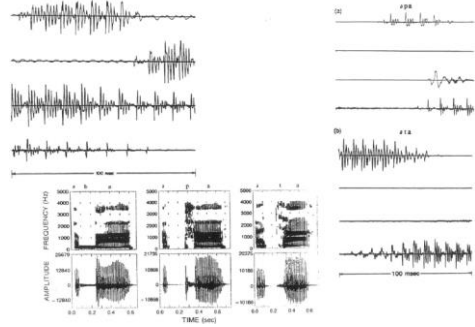
66

Unvoiced Stop Consonants



67

Stop Consonant Waveforms and Spectrograms



68

Distinctive Phoneme Features

Place	p	k	t	b	d	g	f	thin	s	sh	v	the	z	azure	m	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

69

Distinctive Features

Place of articulation	Manner of articulation					
	Glide	Nasal	Stop		Fricative	
			Voiced	Unvoiced	Voiced	Unvoiced
Front						
Labial	w, m	m	b	p		
Labiodental					v	f
Middle						
Dental					ð	θ
Alveolar	j, l	n	d	t	ʒ	ʃ
Palatal	r				ç	ç
Back						
Velar	w, m	ŋ	g	k		
Pharyngeal						ħ
Glottal			ʔ			

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

70

2.2.1. Perde, Şiddet, Kalite

Konuşma sesleri, *perde* (Ing. *pitch*), *şiddet* (Ing. *loudness*) ve *kalite* (Ing. *quality*) açısından üç alt ulama ayrılmaktadır. Perde frekansla, şiddet genlikle, kalite ise tını ile ilişkilidir. Aynı perde ve şiddeti taşıyan bir ses dalgası, kalite açısından /sal/ ve /sel/ sözcüklerindeki ünlüler gibi farklılık gösterebilmekte ya da kalite olarak aynı özelliği göstermesine karşın perde açısından değişebilmektedir.

2.2.2. Formant

Ses üretimi sırasında basıncı hava akımı, ses tellerinin titreşimine neden olmakta ve bu titreşim sonucu oluşan ses sinyali, gırtlak, ağız ve kimi durumlarda geniz boşluklarını da titreşime geçirmektedir. Bunun sonucunda, tınlayıcı niteliği taşıyan bu organlardan geçen ses, değişerek konuşma sesine dönüşmektedir.

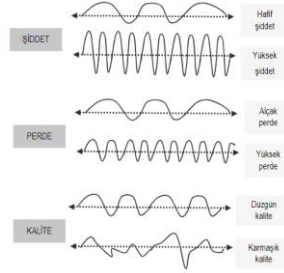
Ses yolundaki boşlukların biçimi ve hacmi, ses üretiminin titreşimine neden olan hareketleriyle birlikte değişim gösterdiği için, bu boşlukların tınlama sıklıkları da farklılaşmakta ve bu tınlama sıklıkları *formant* ya da *biçimleyici* (Ing. *formant*) olarak tanımlanmaktadır. Akustik Sesbilgisi alanının öncülerinden biri olan Gunmar Fant tarafından ses taşıyıcısındaki spektral tepe noktaları olarak adlandırılan formantlar, ses yolundaki akustik titreşimler olarak da yorumlanmaktadır.

71

72

Ses yolundaki akustik kayıplarla belirlenen formant bant aralıkları 0-5000 Hz arasında değişim gösteren dört temel formant değerinden oluşmaktadır: Birinci formant (F1/500-1000 Hz), ikinci formant (F2/1000-2000 Hz), üçüncü formant (F3/2000-3000 Hz), dördüncü formant (F4/3000-4000 Hz). F1 ve F2 formant değerleri ünlülerin üy boyutluluk özelliklerini belirlemede kullanılırken, F3 ve F4 formant değerleri ise cinsiyete dayalı değişimleri belirlemede rol oynamaktadır.

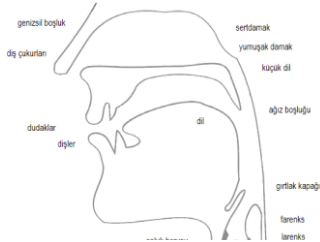
• Sesin perdesi, şiddeti ve kalitesi



73

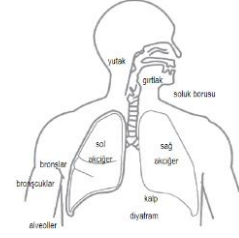
74

- <https://ws1.turcademy.com/ww/webviewer.php?doc=20800#page=41&zoom=auto,-47,544>
- Konuşma organları



75

- Solunum sistemi



76

DURAL HAVA AKIMI VE AERODİNAMİK SÜREÇLER

Bernoulli Etkisi (İng. *Bernoulli Effect*) adı altında da incelenen aerodinamik süreçlerde, ses tellerinin titreşmesi, akciğerlerden gelen havanın etkisiyle gırtlak kaslarının gerilmesi ve esnekliğinin değişmesiyle ilişkilendirilmektedir.

Akciğerlerdeki havanın dışarı atılması sonucu çıkış havası (İng. *egressive air*) ortaya çıkar. Egresif hava aracılığıyla konuşma seslerinin büyük bir kısmı üretilir. Egresif havanın tersine, havanın akciğerlere akması sonucu konuşma sesiyle ilgili havayla geçekleştirilebilir. Ancak gırtlak konuşmada bu havayla üretilen sesler pek fazla kullanılmamakta ve bu tür seslerin kalitesi düşük, boğuk ve belirsiz olmaktadır.



77

3.2. Gırtlak

Seslerin üretimi ve biçimlenmesinde etkin rol oynayan gırtlak (İng. *larynx/glottis*), nefes borusunun üstü, ses yolunun ise alt kısmında bulunur. Gırtlığın iç kısmında, konuşma seslerinin çıkarılmasında edici rolü olan ses telleriyle onların titreşimini denetleyen üç kıkırdak yapı yer almaktadır: *Kalkans Kıkırdak (Çıstığıkıkırdak)*, *Halkansı Kıkırdak (Gırtlaksal)*, *Yüzüksü Kıkırdak*. Gırtlaktaki ses tellerinin uzamasına ya da kısalmasına yarayan döndü çift, biri tek olmak üzere de beş temel kas bulunmaktadır.

Cinsiyete göre biçimlenen gırtlak yapısında, kadınların ses tellerinin kıvrımları kısa olduğu için temel sıklık değerleri 150-220 Hz aralığında ve gırtlak çevresi 110-112 milimetre arasındadır. Erkeklerde ise, ses tellerinin kıvrımları kadınlara göre daha uzun olduğu için temel sıklık değerleri 90-160 Hz aralığında ve gırtlak çevresi de 135-136 mm civarındadır.

3.3. Ses Telleri

Seslerin çıkarılmasında çok önemli bir rol üstlenmiş olan ses telleri (İng. *vocal cords*), gırtlığın işlevlerini kontrol eden kasların hareketlerine

78

göre gerilim, esneklik, yükseklik, genişlik, uzunluk ve kalınlık gibi fiziksel özelliklerinde değişim gösterir. Seslerin çıkarılması sırasında ses tellerinin titreşmesi, **örüm** (Ing. *voice*) adı verilen temel bir özelliğin ortaya çıkmasını sağlamaktadır. Ünlülerin tümü ve ünsüzlerin bir bölümü ötümlü olma özelliği taşırlar, yani bu sesler üretilirken ses telleri titreşir. Ses tellerindeki her bir titreşim, o bölgede bir açılma-kapanma hareketine karşılık gelmektedir.

Cinsiyete göre değişiklik gösteren bu titreşim hareketi, erkeklerde saniyede ortalama 120 titreşim döngüsü olarak tekrarlanırken, kadınlarda bu döngü ortalama 200 titreşim biçimindedir. Ses perdesi yükseldikçe ses tellerindeki titreşim sayısı da artmaktadır. Örneğin, yeni doğan bir bebeğin ağlama sesinde saniyede ortalama 400 titreşim hareketi bulunmaktadır. Ses tellerinin arasında bulunan ve hançere olarak da adlandırılan boşluğun iç kavrımları, erkeklerde 17-24 mm arasında, kadınlarda ise 13-17 mm arasında değişmektedir.



GIRTLAK ÜNSÜZÜ [ʔ]

Nefes tutulurken ses telleri tam bir şekilde kapalı tutulabilir, bu durumda ses telleri sıvaca kapalı tutulup açıldığında serbest kalan aközifer çöküşü hava **gırtlak ünsüzünün** (Ing. *glottal stop*) [ʔ] üretilmesini sağlar. Çoğu dil ve lehçe de seslerin olarak da kullanılan bu ses genellikle öksürük bağlanmasında duyulabilir. Ses telleri birbirinden ayrıık konumdayken basınçlı hava akımının ses telleri arasındaki boşluğa hücum etmesi sonucu, iptilebilir bir gırtlak ünsüzü üretilir, bu süreç sırasında üretilen ses genellikle [h] olarak kullanılmaktadır.

3.6. Yumuşak Damak ve Sert Damak

Ağız boşluğunun arka kısmında kash, geniş bir doku olarak yer alan **yumuşak damakın** (Ing. *velum*) en temel alt organı **küçük dil**dir (Ing. *uvula*). Küçük dil, ağızın arka tarafından sarkın ve aynı yardımcıyla da rahatlıkla görülebilen bir ektendir. Yumuşak damak (arddamak) nefes alıp verme sürecinde alçalarak burundan alınan havanın rahatlıkla içeri girmesini sağlar. Konuşma sırasında yumuşak damak, burun yutağının duvarına doğru yükseltildiğinde hava ağızdan dışarı atılır. Bu sesler, ağır sesleri olarak tanımlanmaktadır. Yumuşak damak, havanın ağızdan ve burundan dışarı atılmasına olanak sağlayacak şekilde aşağı doğru da indirilebilir, bu durumda çiğlerden gelen hava akımı burun boşluğunda dolayarak ağızdan çıkarılır. Örneğin [m] ve [n] geniz ünsüzleri gibi. Damakın, üst sıra dişlere kadar olan bölümü ise konuşma seslerinin oluşumunda önemli rol oynayan sert damak olarak tanımlanmaktadır (Bkz. Bölüm 5).

3.7.4. Dil

Dil (Ing. *tongue*), iç ve dış kısmında bulunan kaslar aracılığıyla yukarı, aşağı, ileri ve geri olmak üzere dört farklı yönde hareket edebilme ve şekil alabilme özelliğine sahiptir. **Dilucu** (Ing. *tongue apex*), **dil kenarı** (Ing. *rim*), **dil palası** (Ing. *blade, lamina*), **dil sırtı** (Ing. *dorsum*) ve **dil kökünden** (Ing. *root*) oluşan dilin iç kısmındaki kaslar, dilin en uç kısmını gösteren dilucunun yukarı-aşağı ya da sağa/sola oynatılmasını sağlar. Konuşma seslerinin çıkarılmasında önemli bir işlevi olan dil, ağız boşluğunun üst kısmındaki konuşma göre, ağız içinde ardamak/yumuşak damak ve öndamak/sertdamakla bütünlüştür.

3.4. Yutak

Gırtlığın içinde konumlanan ve ses tellerinin hareket etmesini sağlayan **yutak** ya da **farenks** (Ing. *pharynx*), 12-14 cm uzunluğunda kas ve mukozadan oluşan bir yapıdır. Yutağın kas yapısında üstkonstriktör, orta konstriktör, ön konstriktörle, yutağın aşağı doğru hareketini sağlayan stillo-farengeus, salpiro-farengeus, palato-farengeus gibi çeşitli kaslar bulunur. Yutak, yukarıdan aşağıya doğru burun yutağı (nazofarenks), ağız yutağı (orofarenks), gırtlak yutağı (larinofarenks) gibi konuşma fazları olarak adlandırılan alt bölümlerden oluşur. Burun yutağı burun boşluğunun arka kısmında, ağız yutağı ağız boşluğunun arka kısmında ve gırtlak yutağı ise boyun omurları hizasında konumlanır. Genizsil boşluk, timpan boşluk, ağız ve gırtlak da yutakla bağlantılı bölümlerdir.

3.5. Ses Yolu (Vokal Trakt)

Geniz yolu ve ağız yolu gibi bölümleri içeren **ses yolu** ya da **vokal trakt** (Ing. *vocal tract*), temel sıklık değerinin ve ses kalitesinin belirlenmesi açısından önemli bir işleve sahiptir. Cinsiyete göre farklılık gösteren ses yolu kalınlığı erkeklerde 16.9 cm, kadınlarda ise 14.1 cm; ses yolu uzunluğu ise erkeklerde 90-190 mm arasında, kadınlarda ise 160-300 mm kadardır.

3.7. Ağız Boşluğu

Dudaklar, alt ve üst dişler, damak, yanaklar, altçene kemiği ve dil den oluşan **ağız boşluğu** (Ing. *oral cavity*), dişlerin dizilişine göre iki bölüme ayrılır.

3.7.1. Dudaklar

Konuşma seslerinin üretilmesinde önemli bir yeri olan dudaklar, ağız bölgesinin çevresini kaplayan çeşitli kaslar tarafından yönetilir. Seslerin çıkarılış sırasında farklı konumlara getirilebilen **dudaklar** (Ing. *lips*) kapalı, açık, düz ya da yuvarkat biçim alabilir.

3.7.2. Dişler

Ağız boşluğu içinde konumlanan alt sıra dişler ve üst sıra dişler, kesici dişler, köpek dişleri, küçük azı dişleri ve büyük azı dişleri olmak üzere kendi içinde bölümlere ayrılır. Özellikle ünsüzlerin sesletiminde dişler önemli bir işleve sahiptir.

3.7.3. Çene Kemiği

Konuşma seslerinin üretilmesi sırasında aşağı ve yukarı doğru hareket edebilme özelliği taşıyan **çene kemiği** (Ing. *mandible*), ağız boşluğu içinde bulunan alt sıra dişler ve üst sıra dişler arasındaki açığı ayarlayarak, dudakların durumunun ve biçiminin belirlenmesinde etkin rol oynamaktadır.

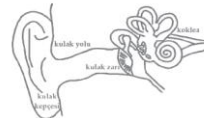
3.8. Kulak

Dış kulak (Ing. *external ear*), ses dalgalarının kulğa ulaşmasında etkin rol oynayan kulak kepçesi ve halkalardan oluşur. İşitme kanalı olarak da tanımlanabilen dış kulak yolu, 2.5 cm uzunluğundadır ve 3000-4000 Hz aralığındaki seslere iyileşmesinde büyük rol oynar.

Dış kulak ve orta kulak (Ing. *middle ear*), ses dalgalarının titreşmesini sağlayan daire biçimindeki kulak zarıyla birbirinden ayrılır. Orta kulakta titreşen ses dalgaları kulak zarında mekanik sese dönüşür ve bu ses oldukça ince yapıtlı üç kemikçik (çekiş, örs, üzengi) tarafından iş kulğa iletilir.

İç kulakta (Ing. *internal ear*) denge kontrolünü sağlayan varem dairesel kanallar ve 33 mm uzunluğundaki salyangoz adı verilen koklea, orta kulakta üretilen mekanik ses dalgalarını usur hücreleri aracılığıyla elektrik akımına dönüştürür. Bu akım işleme sanitleri aracılığıyla beyin optuna ve temporal loba iletilir.

Şekil 13. Kulak Yapısı





85

- Pulmoner ünsüzlerin uluslararası sesbilim abecesindeki (IPA, 2005) gösterimi

	Çift Dudak	Dudak-dış	Dışıl	Diş-yuvası	Artdiş-yuvası	Üstdamak	Damaklı	Artdamak	Küçük dil	Boğazlı	Gırtlak
Genzsil	m	n									
Patlamlı	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Sürtünücü	f β	v	θ ð	s z	ʃ ʒ	ʧ ʤ	x ɣ	χ ʁ	h	ʕ	ħ
Yanal		ɸ		ɹ		ɻ	ʎ	ɰ			
Akıcı				r					ʀ		
Çarpmalı				r		ɽ					
Yanal sürtünücü				ɬ ɮ							
Yan darımlı				l		ɭ	ʎ	ʟ			
Yanal çarpmalı				ɻ							

86

- Pulmoner olmayan ünsüzlerin uluslararası sesbilim abecesindeki (IPA, 2005) gösterimi

Şaklamalar	Ötümü	İçe Patlamalar	Dişe Patlamalar
o Çift Dudak	b	Çift Dudak	p
l Dışıl	d	Dışıl/Dişyuvası	t
l Artdişyuvası	ʃ	Damaklı	k
ɸ Damak-dışyuvası	ʒ	Artdamaklı	s
h Yanal-dışyuvası	ç	Küçük Dil	ʎ

87

- Diğer imlerin uluslararası sesbilim abecesindeki (IPA, 2005) gösterimi

ʌ	Ötünsüz, dudak-artdamaklı, sürtünücü
w	Ötümü, dudak-artdamaklı, yanal
ɥ	Ötümü, dudak-damaklı, yanal
ɰ	Ötümü, gırtlak kapağı, sürtünücü
ʕ	Ötümü, gırtlak kapağı, sürtünücü
ʔ	Gırtlak kapağı, patlamalı
ɛ̥, ɛ̜	Dışyuvası-damaklı, sürtünücü
ɰ	Ötümü, dışyuvası-yanal, çarpmalı
ɰ	Eğzamlı f ve x

88

- Ayrımcı imlerin uluslararası sesbilim abecesindeki (IPA, 2005) gösterimi

ɰ	Ötümü ɲ, ɳ	ɰ	Hırtlı ɸ, ɸ	ɰ	Dışıl ʈ, ɖ
ɰ	Ötümü ɶ, ɷ	ɰ	Boğuk ɸ, ɸ	ɰ	Dilucu ʈ, ɖ
ɰ	Soluklu ɸ, ɸ	ɰ	Dil-dudaklı ʈ, ɖ	ɰ	Dil palası ʈ, ɖ
ɰ	Daha yuvanlı ɰ	ɰ	Dudaklı ɸ, ɸ	ɰ	Genzsil ɛ
ɰ	Az yuvanlı ɰ	ɰ	Damaklı ɸ, ɸ	ɰ	Genz çıkışı ɸ
ɰ	İleride ɶ	ɰ	Artdamaklı ɸ, ɸ	ɰ	Yanal çıkışı ɸ
ɰ	Geride ɶ	ɰ	Boğazlı ɸ, ɸ	ɰ	Akustik olmayan çıkışı ɸ
ɰ	Merkezde ɸ	ɰ	Artdamaklı ya da Boğazlı ɸ		
ɰ	Orta-merkezde ɸ	ɰ	Yüksek ɶ (ɰ = ötümlü, dış yuvası, sürtünücü)		
ɰ	Sesizleşmiş ɶ	ɰ	Alçak ɶ (ɰ = ötümlü, çift-dudak, yanal)		
ɰ	Sesizleşmiş olmayan ɶ	ɰ	İlen Dil Kökü (ATR) ɶ		
ɰ	r-söyleyişli ɶ, ɶ	ɰ	Geni Dil Kökü (RTR) ɶ		

89