Cognitive Media Processing #5

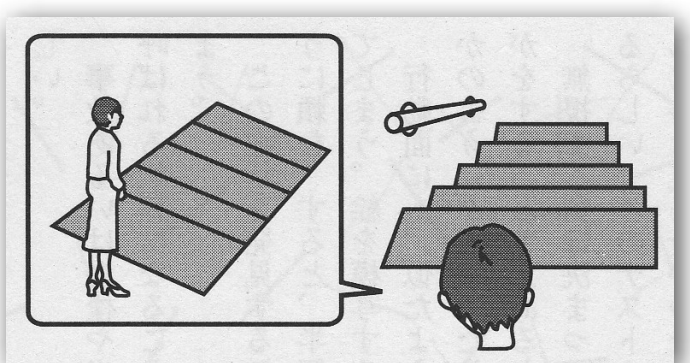
Nobuaki Minematsu

Menu of the last lecture

- Wonders of sensation that I've talked about so far.
  - Unconscious processing
  - Blind spot, blind sight, color illusion, size illusion, etc
- Other wonders of sensation
  - Visual sensation described by a doctor with brain damage.
  - Some peculiar behaviors of autistic individuals
  - A claim on brain info. processing from a brain scientist
- BBC documentary
  - "Derek Tastes of Earwax" ("共感覚の不思議")
  - "Seeing colors by hearing sounds"
  - **https://bit.ly/CMP-D4**
- The first assignment
- Summary

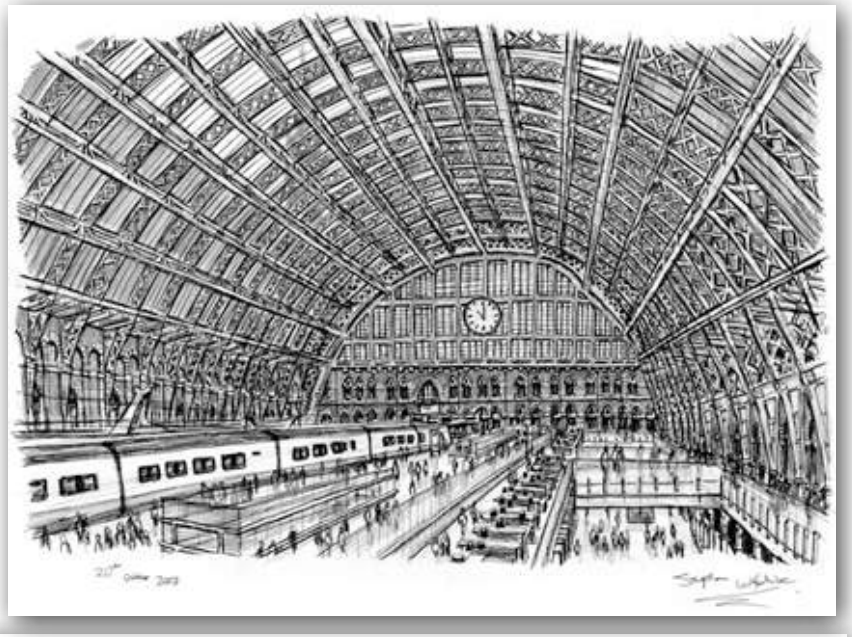
Some facts caused by brain damages

- "I'm living with a damaged brain." (Dr. Kikuko Yamada)
  - Higher-level brain dysfunction (高次脳機能障害)
    - A part of the brain does not function well and she can be aware of that.
    - A medical doctor herself describes what she can sense through her damaged brain.
  - Seeing = conversion of a 2D image into a 3D image
    - What happens if the visual region of the brain has some dysfunction.
    - Stairs = just a plane with some linear segments
      - Cannot tell whether the stairs go up or down.
    - Chopsticks partially hidden at the background of a rice bowl.
      - Two separate objects cannot be bound into one object.
    - Shadows cannot give depth perception.
      - No difference between the two images below.



Sensation by autistics

- Stephen Wiltshire as “human camera”
  - Extraordinary memory of visual stimuli, especially buildings in a landscape.
  - But poor at spoken language, environmental changes, etc.

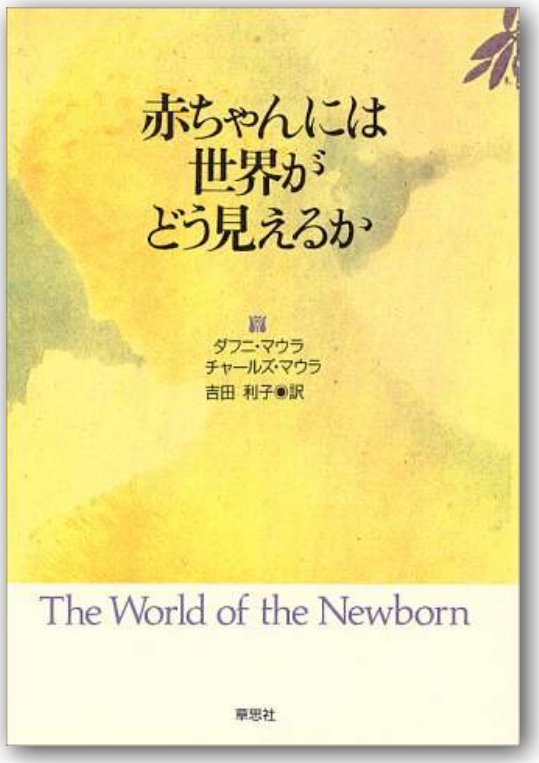


A report from CBS news

- The Tool Man

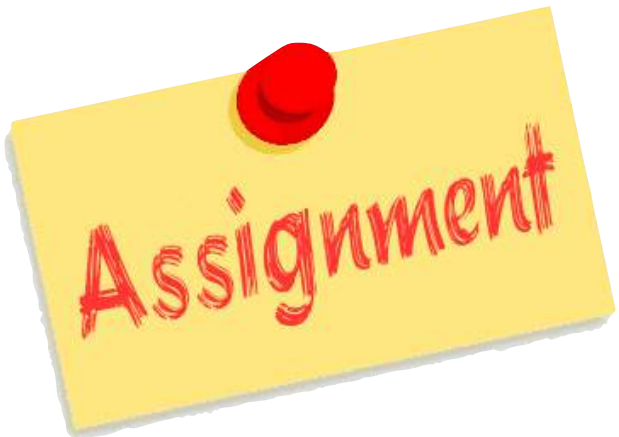
A wonder of sensation

- A 45-min documentary film on synesthesia made by BBC
  - Perceiving colors while seeing or hearing numbers
  - https://bit.ly/CMP-D4 (Two videos are available in English and in Japanese)
- Every baby is like that.
  - “The world of the newborn” (D. Maurer and C. Maurer, 1989)

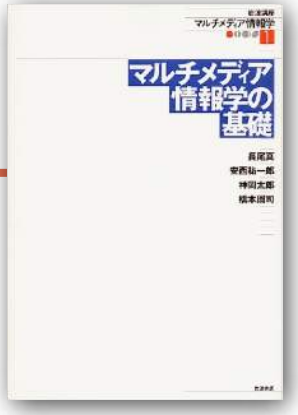


Assignment

- Assignment
  - Read a research paper related to the first four lectures of this class.
  - **Submit two PDF files: 1) the paper and 2) summarization of the paper and your comments on the paper**
  - All the materials used in the lectures are available at:
    - https://www.gavo.t.u-tokyo.ac.jp/~mine/japanese/CMP/class.html
    - Ramachandran's article on synesthesia is also found there.
- Length
  - Two or more pages of A4 size for 2)
- Submission
  - Your report should be submitted via. ITC-LMS.
  - **The filenames must be in the following format.**
    - **[student_id]_paper.pdf and [student_id]_[name].pdf**
    - **36-302439_paper.pdf (paper)**
    - **36-302439_NobuakiMinematsu.pdf (summary and comments)**
- Deadline = 23:59:59 on Nov. 7.



Title of each lecture



- Theme-1
  - ~~Multimedia information and humans~~
  - ~~Multimedia information and interaction between humans and machines~~
  - ~~Multimedia information used in expressive and emotional processing~~
  - ~~A wonder of sensation - synesthesia -~~
- Theme-2
  - Speech communication technology - articulatory & acoustic phonetics -
  - Speech communication technology - speech analysis -
  - Speech communication technology - speech recognition -
  - Speech communication technology - speech synthesis -
- Theme-3
  - A new framework for “human-like” speech machine #1
  - A new framework for “human-like” speech machine #2
  - A new framework for “human-like” speech machine #3
  - A new framework for “human-like” speech machine #4

Speech Communication Tech.

- Articulatory & acoustic phonetics -

Nobuaki Minematsu

Today's menu

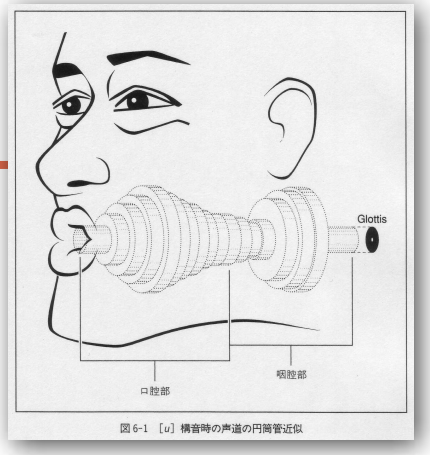
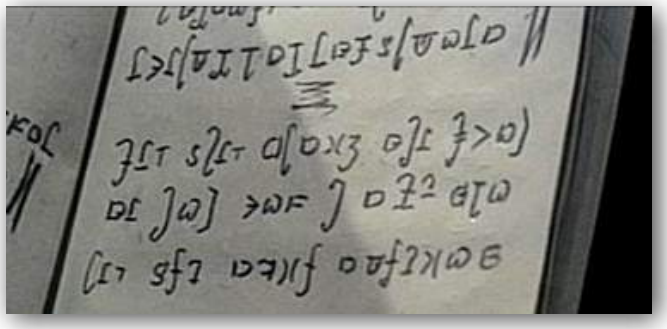
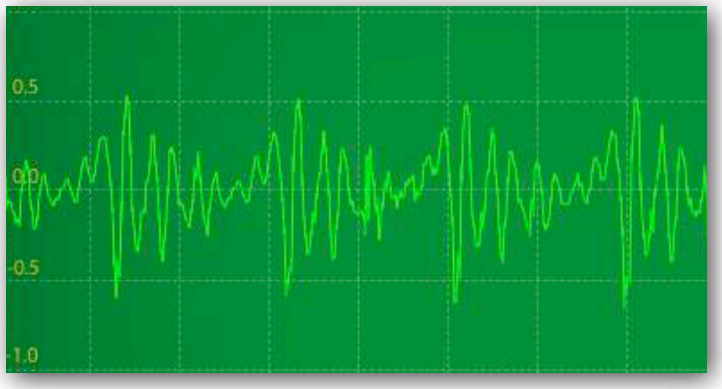


図 6-1 ［u］構音時の声道の円筒管近似

- Speech --> sounds --> vibrations (waves) of air particles
- Fundamentals of phonetics
  - How are vowel sounds produced?
  - Phonetics = articulatory phonetics + acoustic phon. + auditory phon.
- More on articulatory phonetics
  - Observation of speech organs



- More on general phonetics
  - General phonetics = language independent phonetics
  - How to symbolize language sounds found in any language?
- More on acoustic phonetics
  - Vowels as standing waves
    - Resonance frequency = formant frequency
  - Link between acoustic phon. and articulatory phon.
- Summary

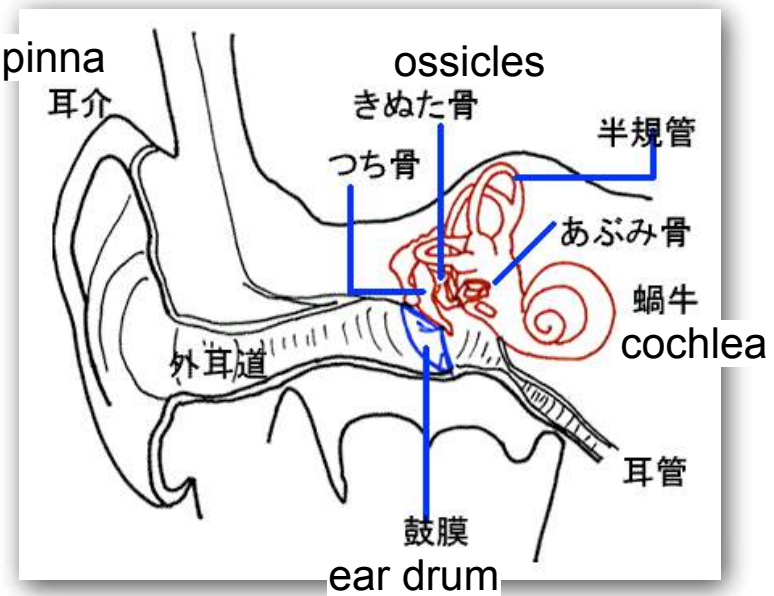


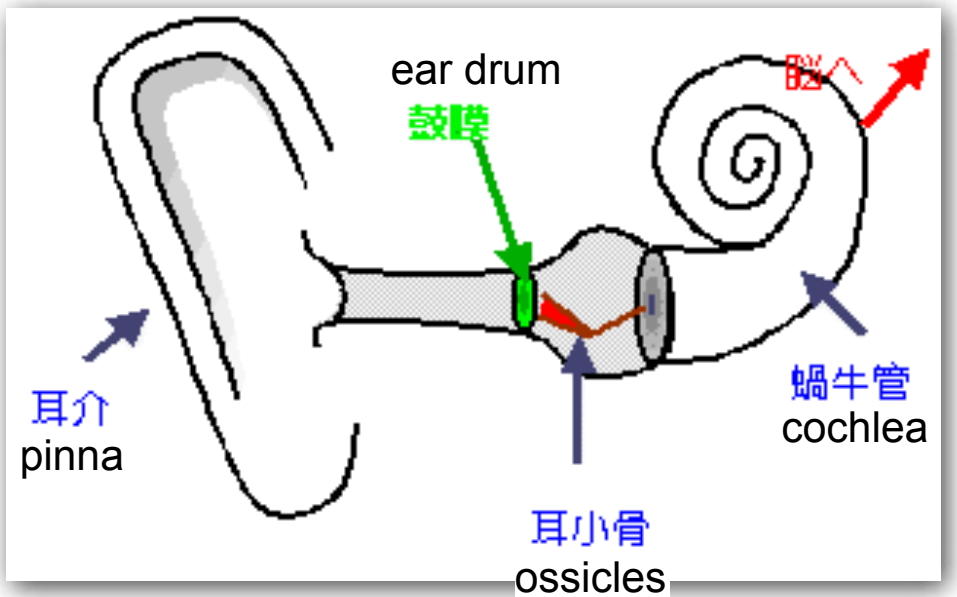
Speech = vibrations of air particles

- What is speech?
  - “AH!” generates vibrations of air particles such as O_2, CO_2, and N_2.
  - Each particle just vibrates but does not move from a place to another.
    - If particles travel from a place to another, they are called “wind”.
    - If particles just vibrate around a certain place, they are called “sound”.
    - And the vibration patters can be transmitted easily, i.e., “wave”.
      - The velocity of transmission of air particle vibrations (sounds) is about 330 m/sec.

蝸牛＝かぎゅう＝カタツムリ＝snail

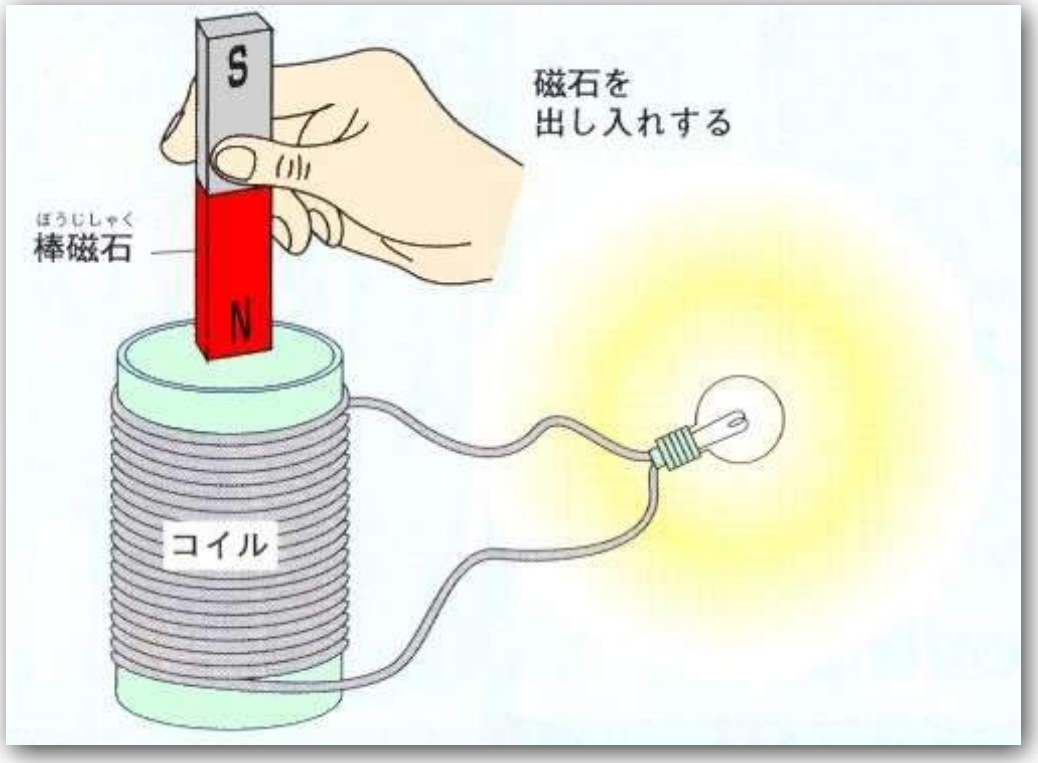
- A simple question.
  - Can air particle vibrations move or vibrate a thing?

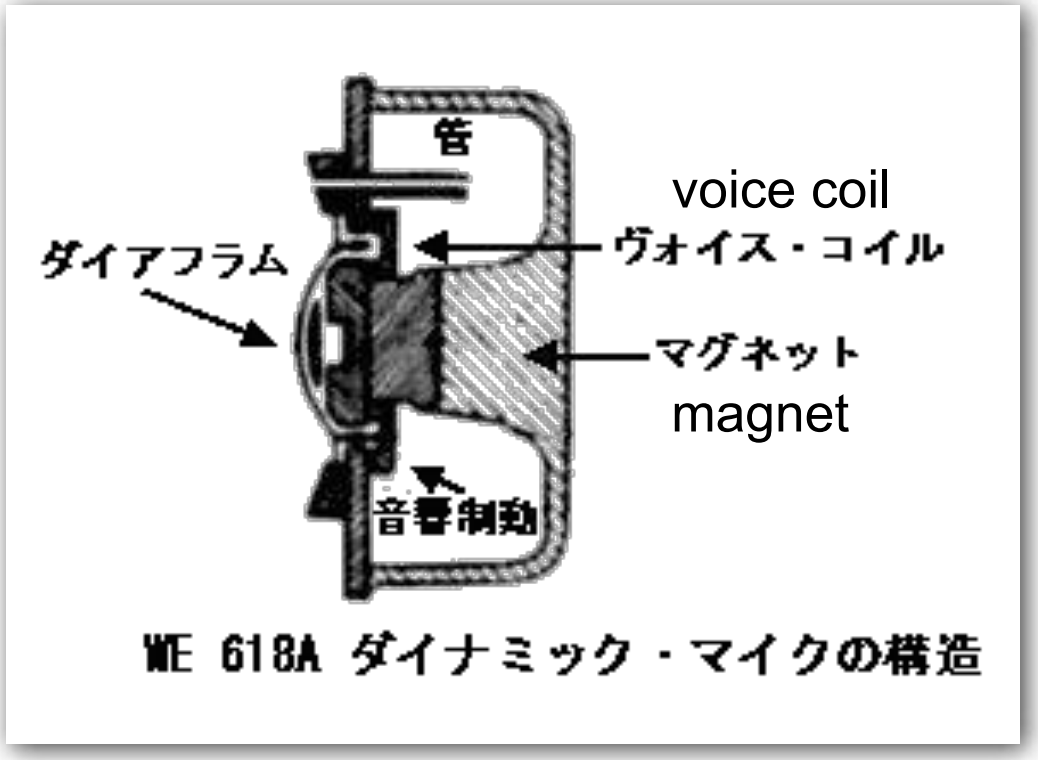




Speech = vibrations of air particles

- If air particles can vibrate a conductive device or material, and
- If vibration of the device is made in a magnetic field, what happens?

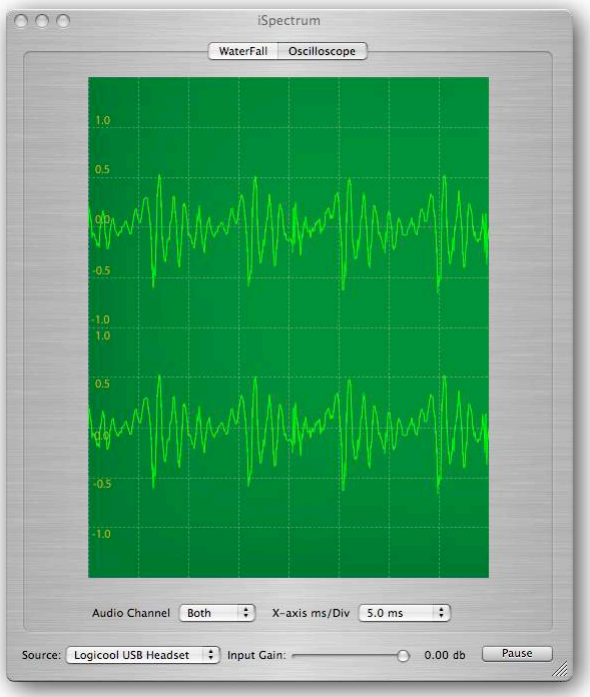


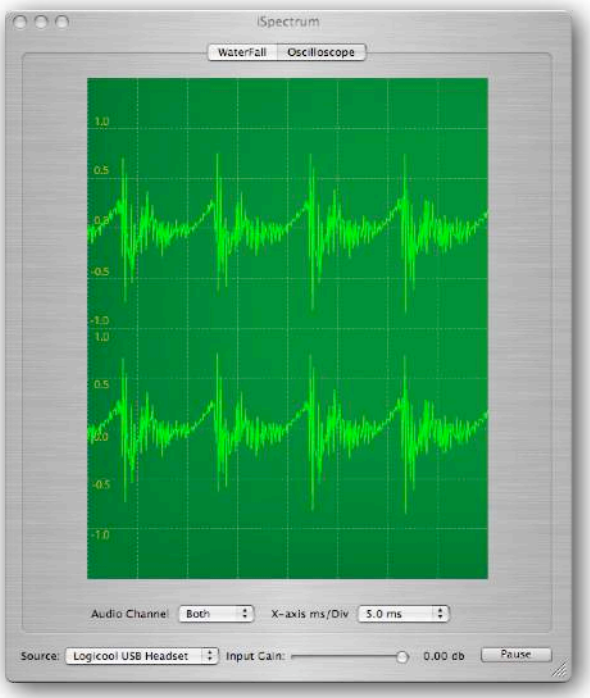


An electric current runs!!

Speech = vibrations of air particles

- Air particle vibrations = electricity vibrations
  - Can be observed using an oscilloscope.



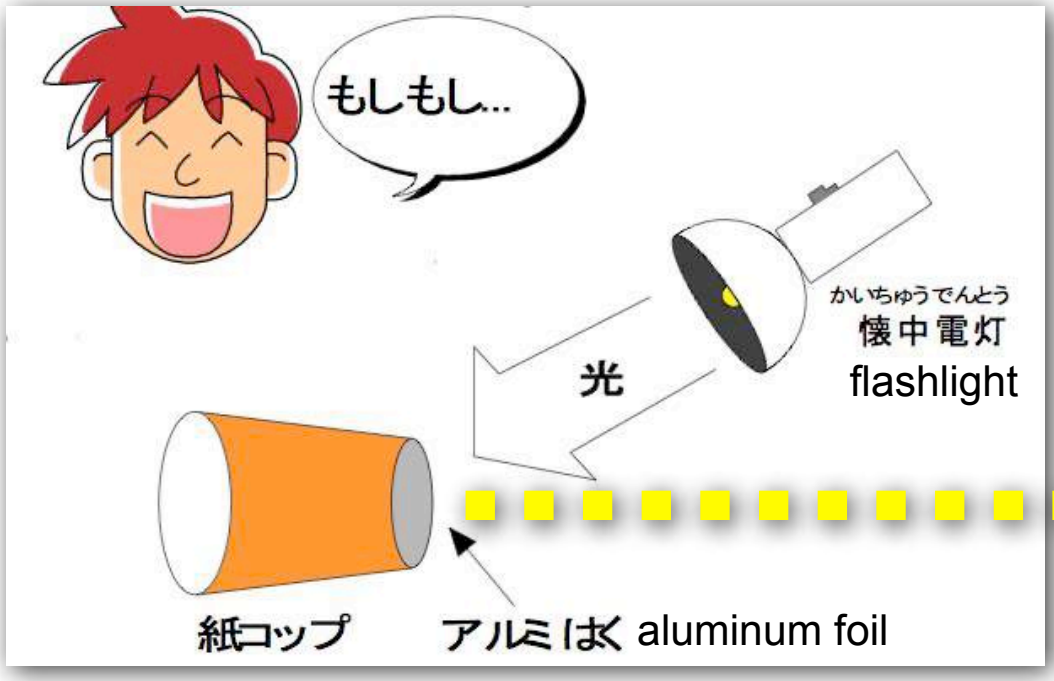


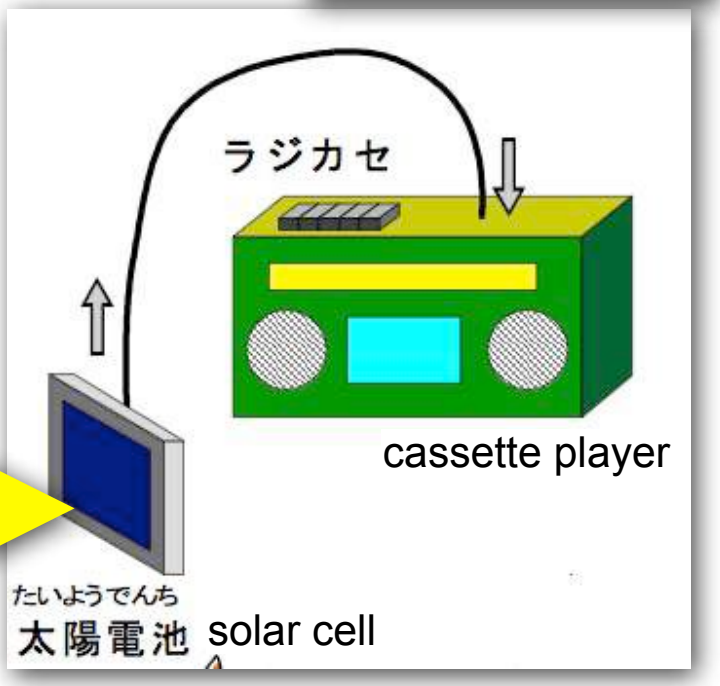
cone of a loud speaker

- Function of a loud speaker
  - Vibrations of electricity --> vibrations of a speaker cone --> vibration of air particles
  - Interesting youtube video : https://www.youtube.com/watch?v=cSLnD3XaVGI
- Function of a microphone
  - Vibrations of air particles --> vibrations of a voice coil --> vibrations of electricity

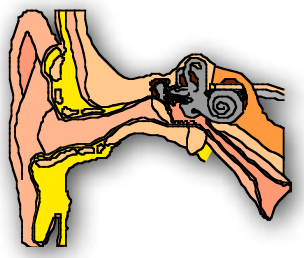
Speech = vibrations of air part

- The media of vibration can be different from air particles.

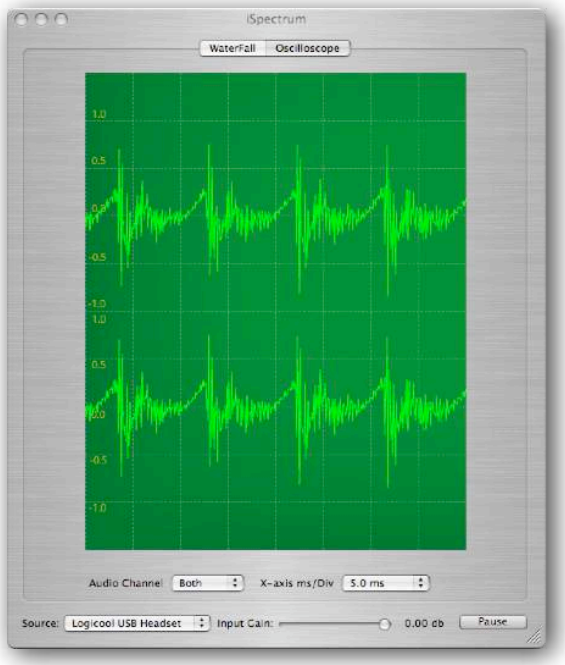




- What is needed is just vibration patterns of any medium.
  - Vibrations of air particles, foil, light, electricity, cone (paper), and air particles.
  - If fingers are linked to the language region of a brain, we can understand the message by touching the aluminum foil!!!
  - The vibrations have to be realized as “air particle” vibration for humans.
    - Because only ears are linked to the language region of a brain.



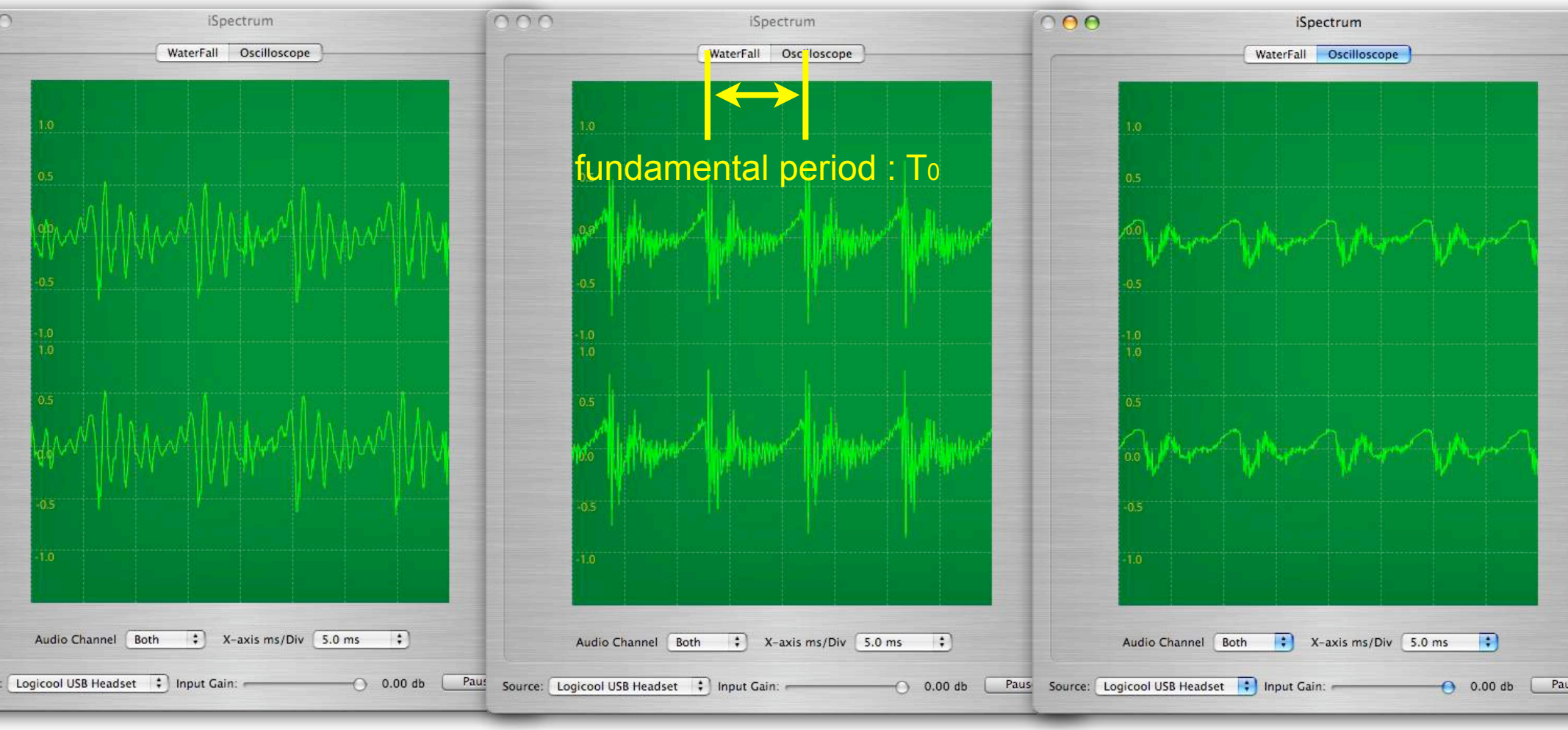
Speech = vibrations of air particles



- The four aspects of tones (sounds)
  - Height of tones (pitch of tones)
    - High tones and low tones
  - Loudness of tones
    - Loud tones and soft tones
  - Duration of tones
    - Long tones and short tones
  - Timbre of tones (color of tones, 音色, 声色)
    - ????
    - If two tones have the same height, the same loudness, and the same duration but the two tones are perceived as different tones, then, the two tones differ in their timbre.
    - /a/ and /i/ /a/ and /a/
      - difference in phoneme, difference in gender

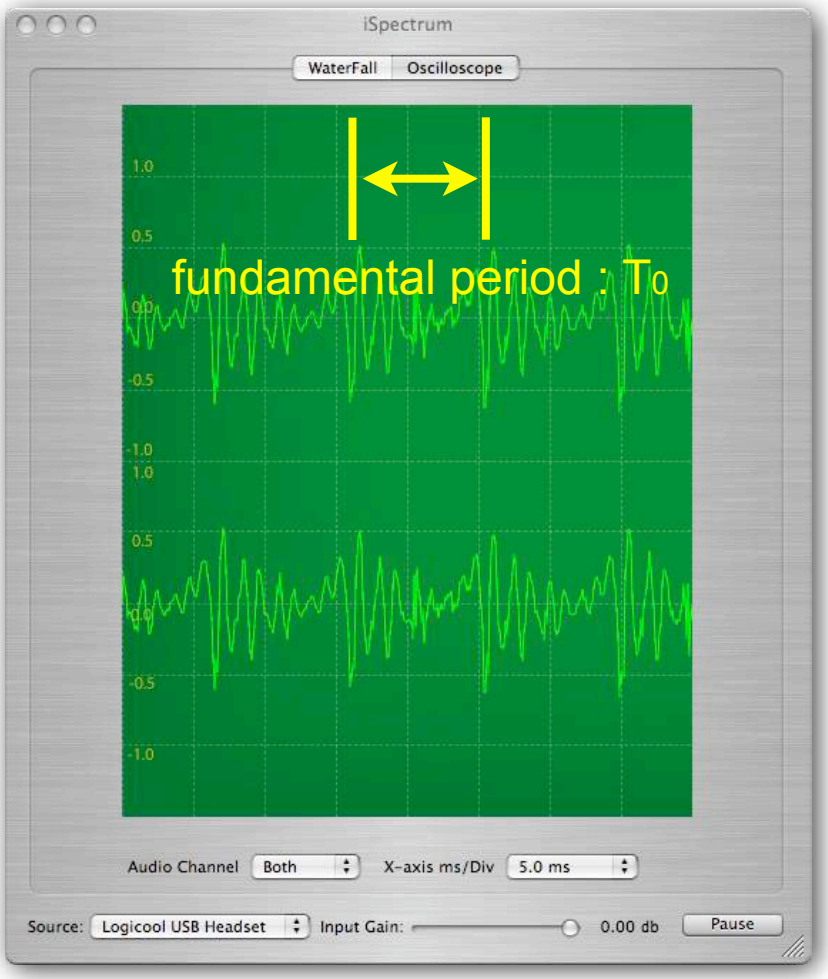
Speech = vibrations of air particles

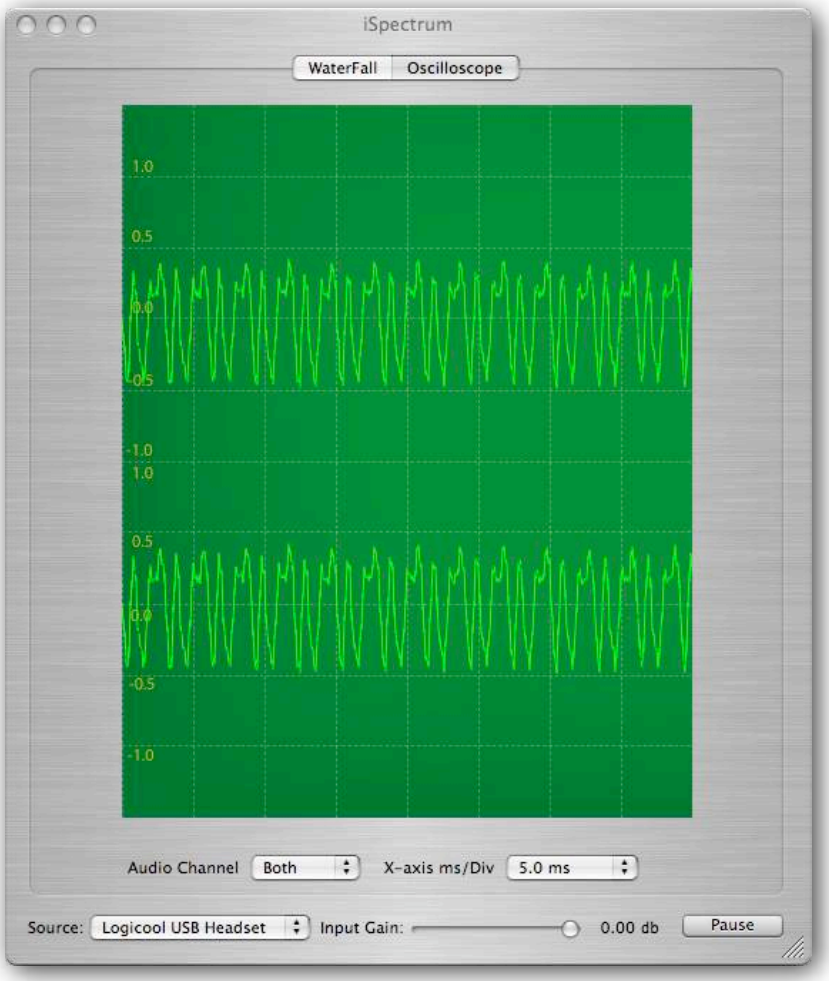
- Close observation of air particle vibration patterns.
  - /a/, /i/, and /u/ with the same height of tone.
  - They are periodic signals (waveforms).



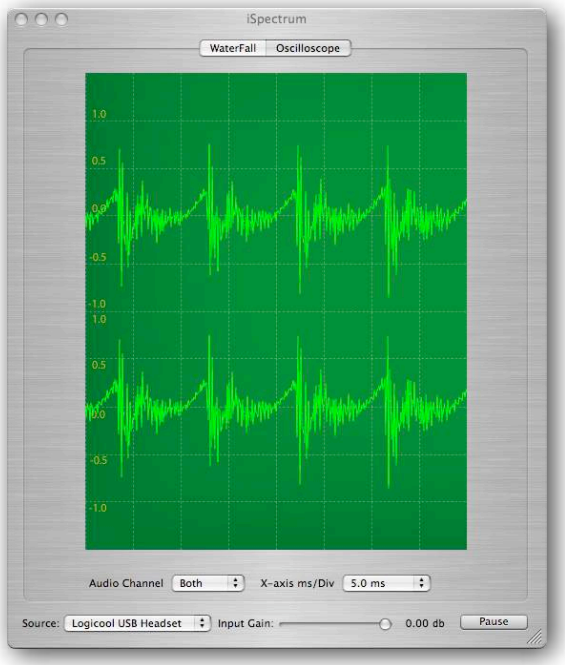
Speech = vibrations of air particles

- Close observation of air particle vibration patterns.
  - Low /a/ and high /a/ in pitch
  - F_0 : fundamental frequency (pitch) = $1/T_0$ = 1/fundamental period





Speech = vibrations of air particles



- The four aspects of tones (sounds)
  - Height of tones (pitch of tones)
    - High tones and low tones
  - Loudness of tones
    - Loud tones and soft tones
  - Duration of tones
    - Long tones and short tones
  - Timbre of tones (color of tones, 音色, 声色)
    - ????
    - If two tones have the same height, the same loudness, and the same duration but the two tones are perceived as different tones, then, the two tones differ in their timbre.
    - /a/ and /i/ /a/ and /a/
      - difference in phoneme, difference in gender

Today's menu

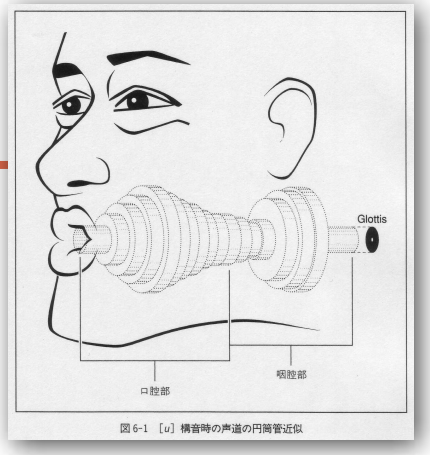
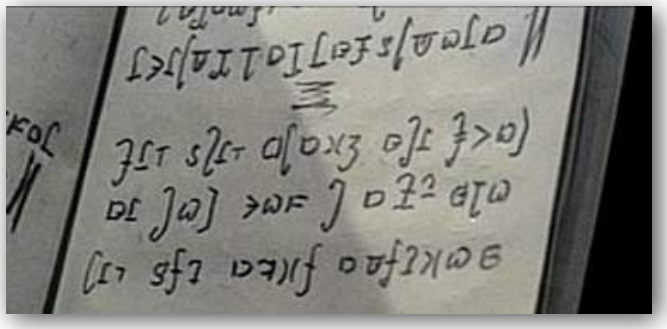
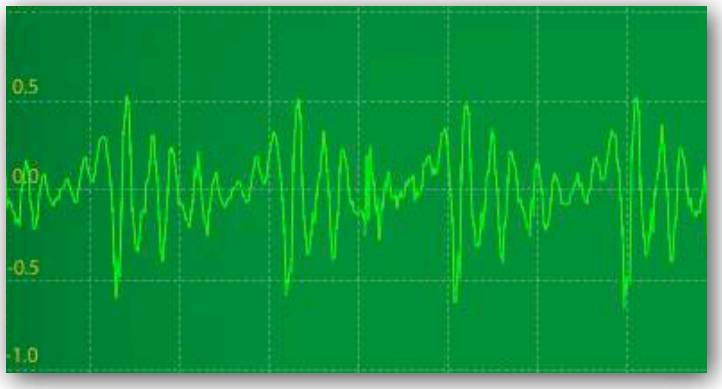


図 6-1 ［u］構音時の声道の円筒管近似

- Speech --> sounds --> vibrations (waves) of air particles
- Fundamentals of phonetics
  - How are vowel sounds produced?
  - Phonetics = articulatory phonetics + acoustic phon. + auditory phon.
- More on articulatory phonetics
  - Observation of speech organs

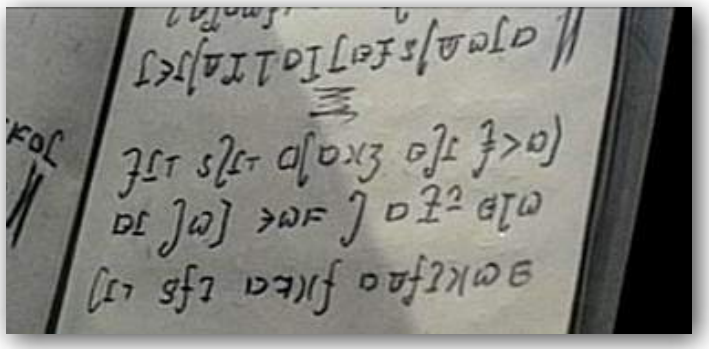


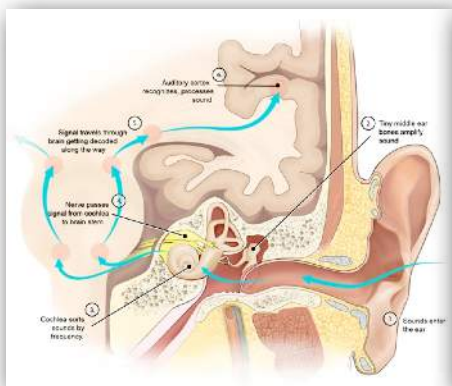
- More on general phonetics
  - General phonetics = language independent phonetics
  - How to symbolize language sounds found in any language?
- More on acoustic phonetics
  - Vowels as standing waves
    - Resonance frequency = formant frequency
  - Link between acoustic phon. and articulatory phon.
- Summary



What is phonetics?

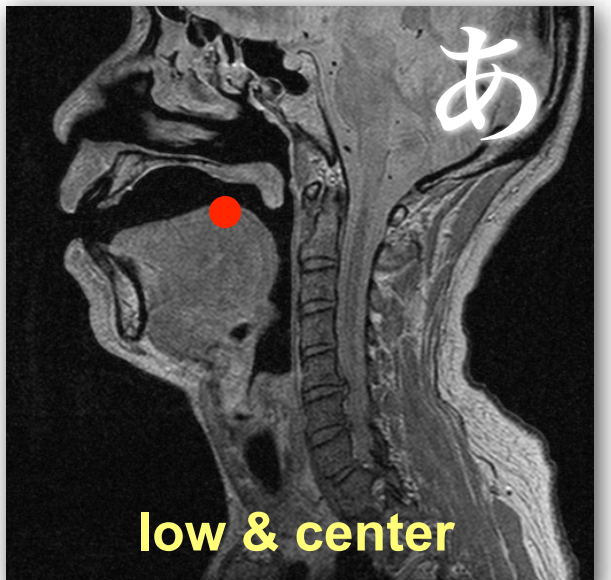
- Phonetics
  - Focus on sounds that can convey linguistic messages.
  - Describe or transcribe utterances independently of language.
    - IPA symbols (IPA = International Phonetic Alphabet)
    - If a new language is found and a new sound is found,
      - IPA (A=association) gives a new IPA symbol for that sound.
  - General phonetics and XXXX phonetics
- Three kinds of phonetics
  - Articulatory phon. + acoustic phon. + auditory phon.
    - Focus is put on articulatory, acoustic, or physiological phenomena.

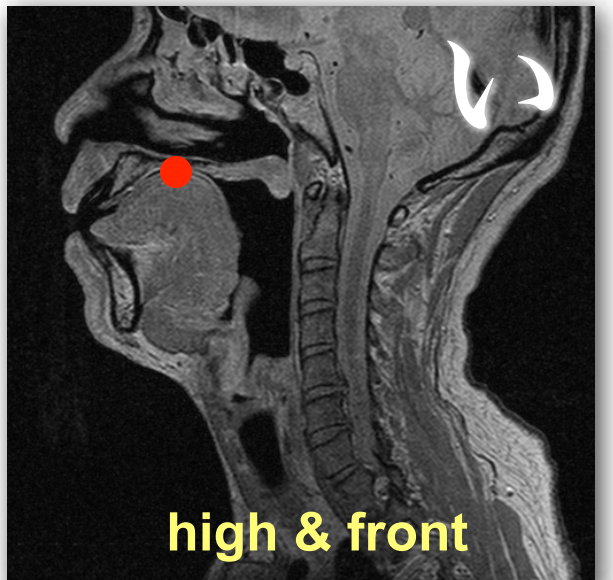


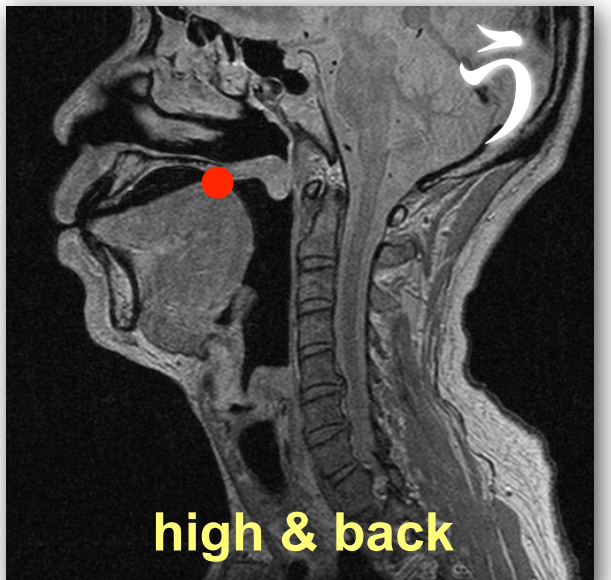


Fundamentals of phonetics

- How are vowels produced in the mouth?
  - Vowels : speech sounds produced with an open vocal tract (tube) so that there is no obstacle to air flow at any point above the glottis. (glottis = 声門)
  - Consonants : speech sounds that are articulated with complete or partial closure in the vocal tract.
- Classification of the vowels
  - In terms of deformation of the inner space in the vocal tract.
    - Vertical & horizontal position of the tongue
    - Lip rounding or not

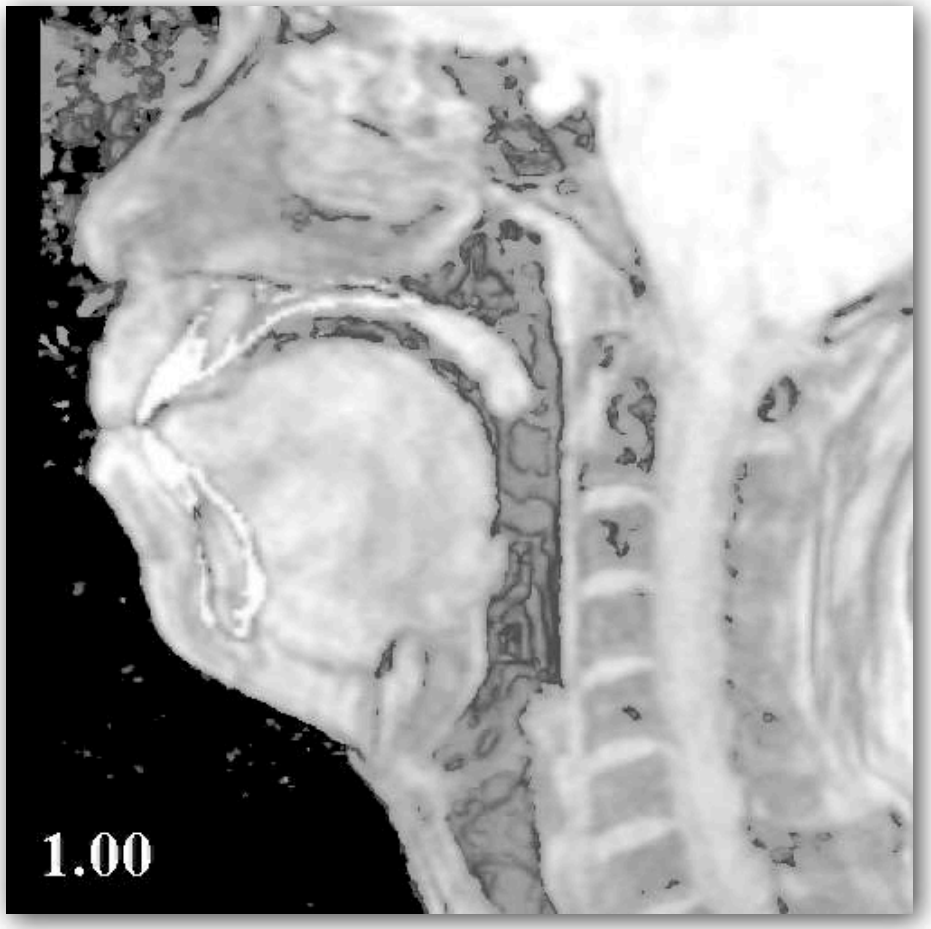


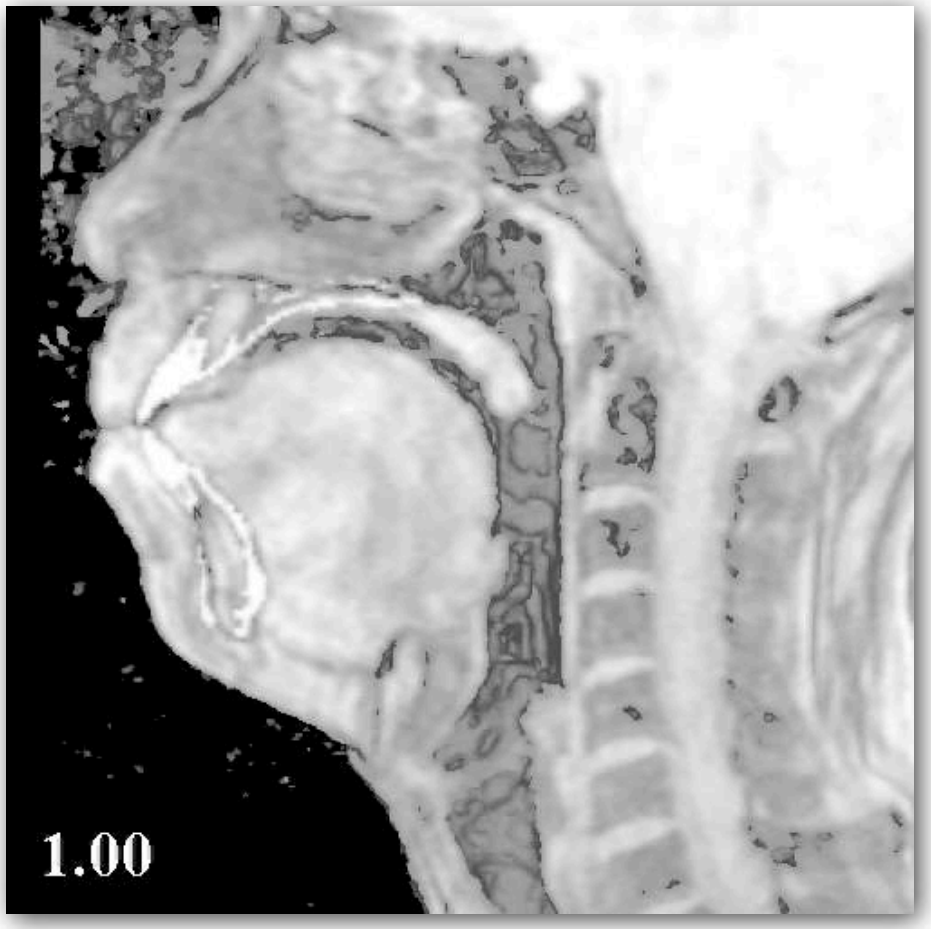




Fundamentals of phonetics

- Dynamic movement of the tongue



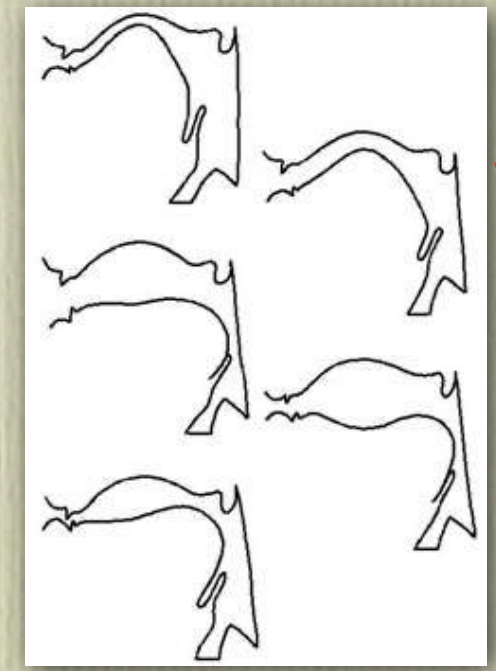


Provided by ATR Corp.

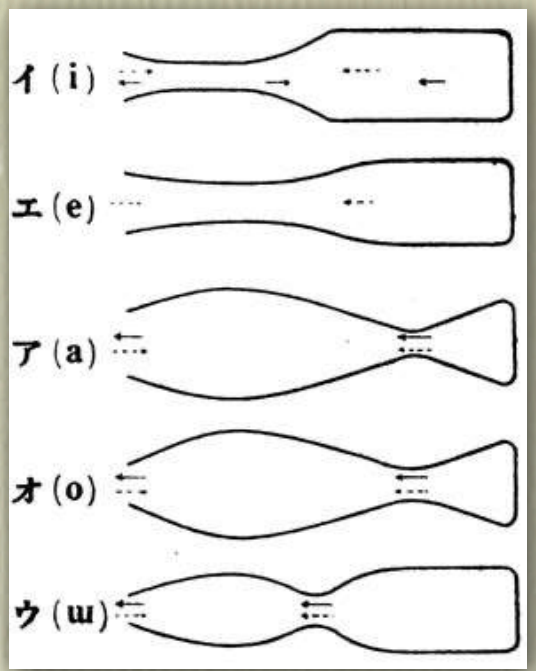
- Timbre difference = shape difference of the inner space

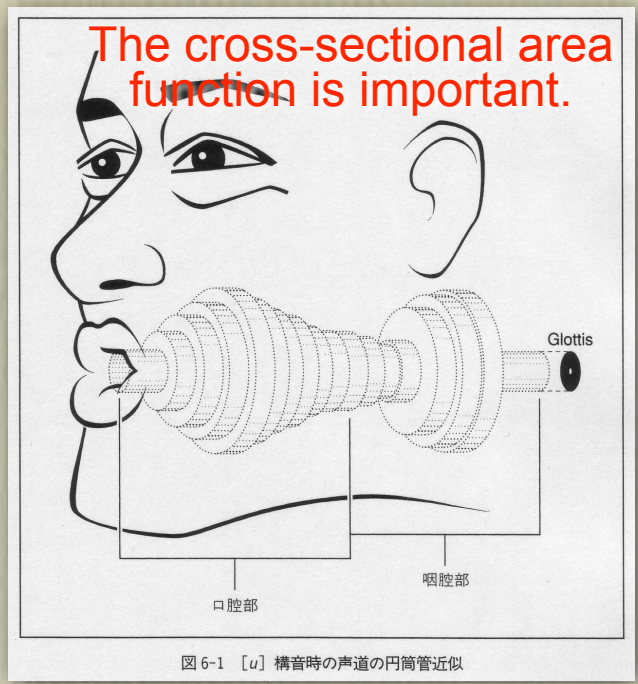
Fundamentals of phonetics

- Air flow --> buzzer sound --> variously shaped tubes --> various timbres



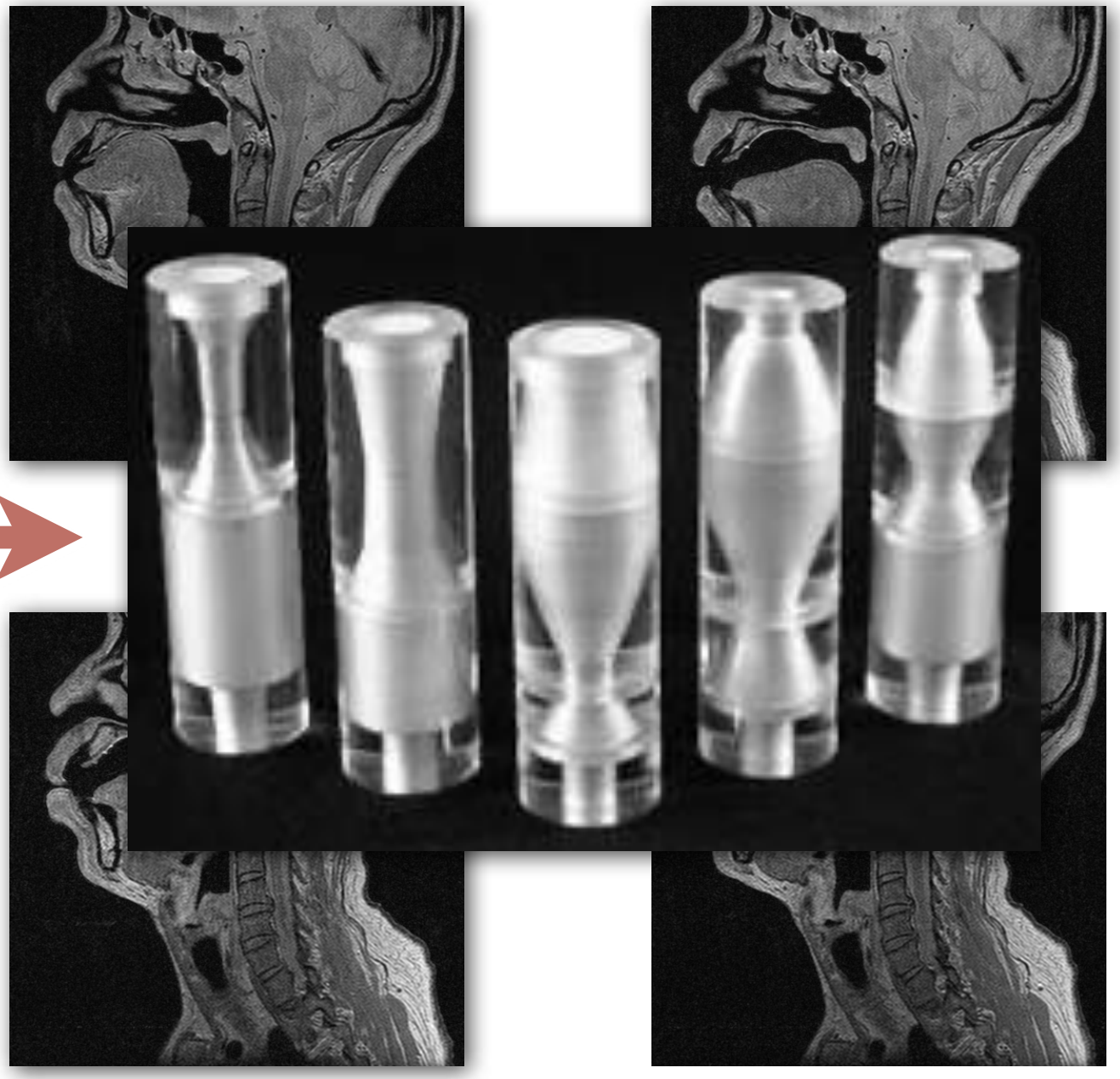
Stretching does not change the timbre.





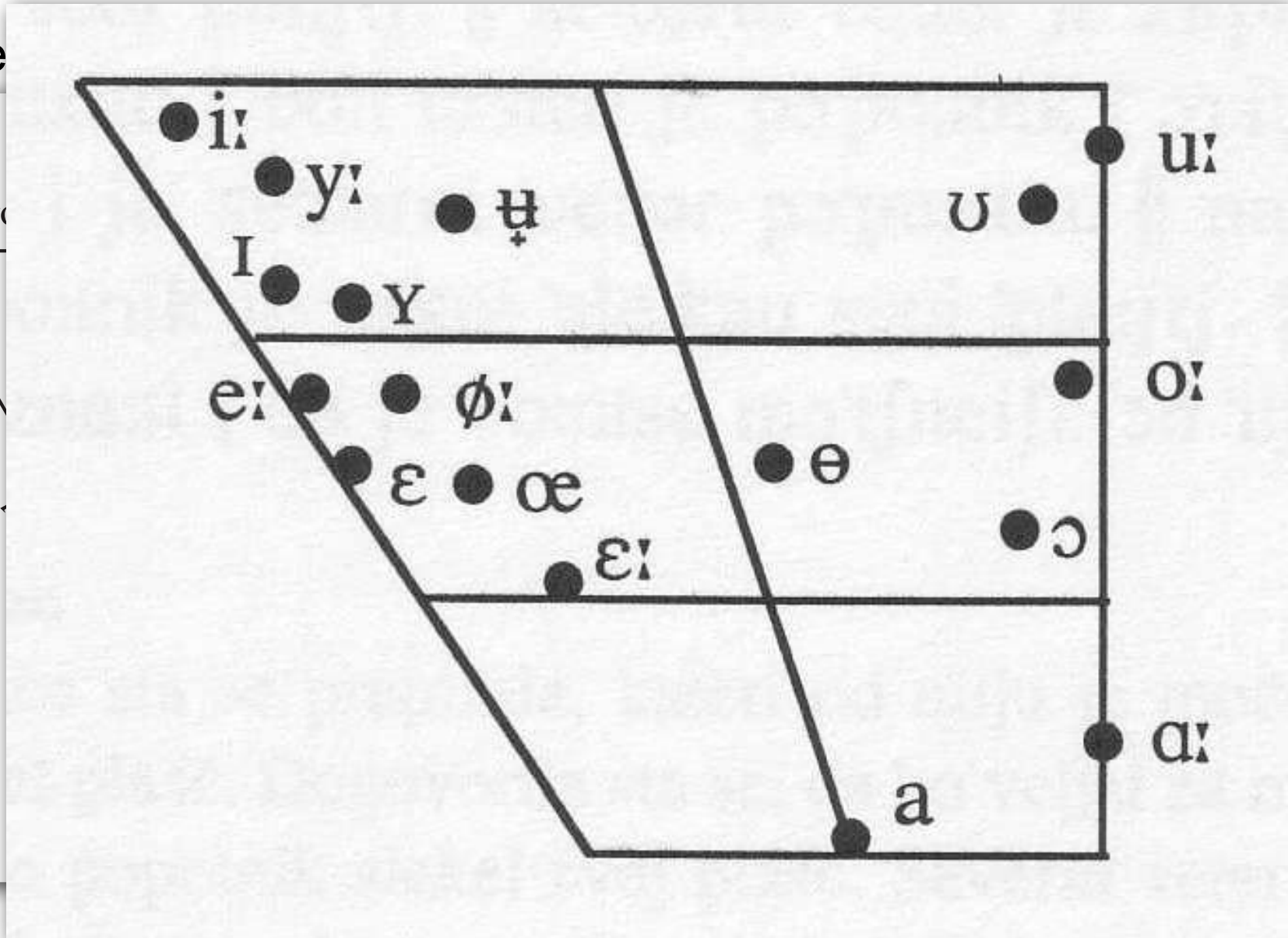
We're always breaking our instrument in vain.

Glottal source + throat = buzzer + tube



Fundamentals of phonetics

- Vowe



5 5 x 2 + 1

Today's menu

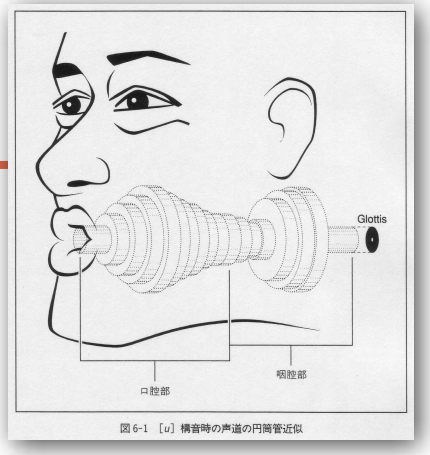
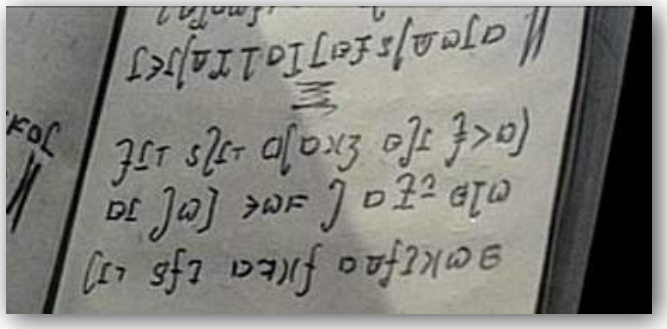
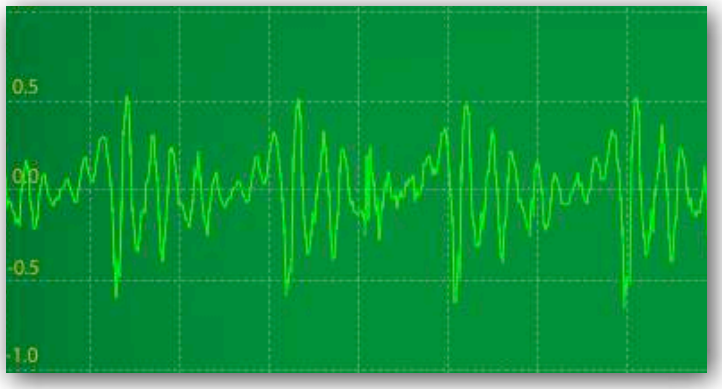


図 6-1 ［u］構音時の声道の円筒管近似

- Speech --> sounds --> vibrations (waves) of air particles
- Fundamentals of phonetics
  - How are vowel sounds produced?
  - Phonetics = articulatory phonetics + acoustic phon. + auditory phon.
- More on articulatory phonetics
  - Observation of speech organs

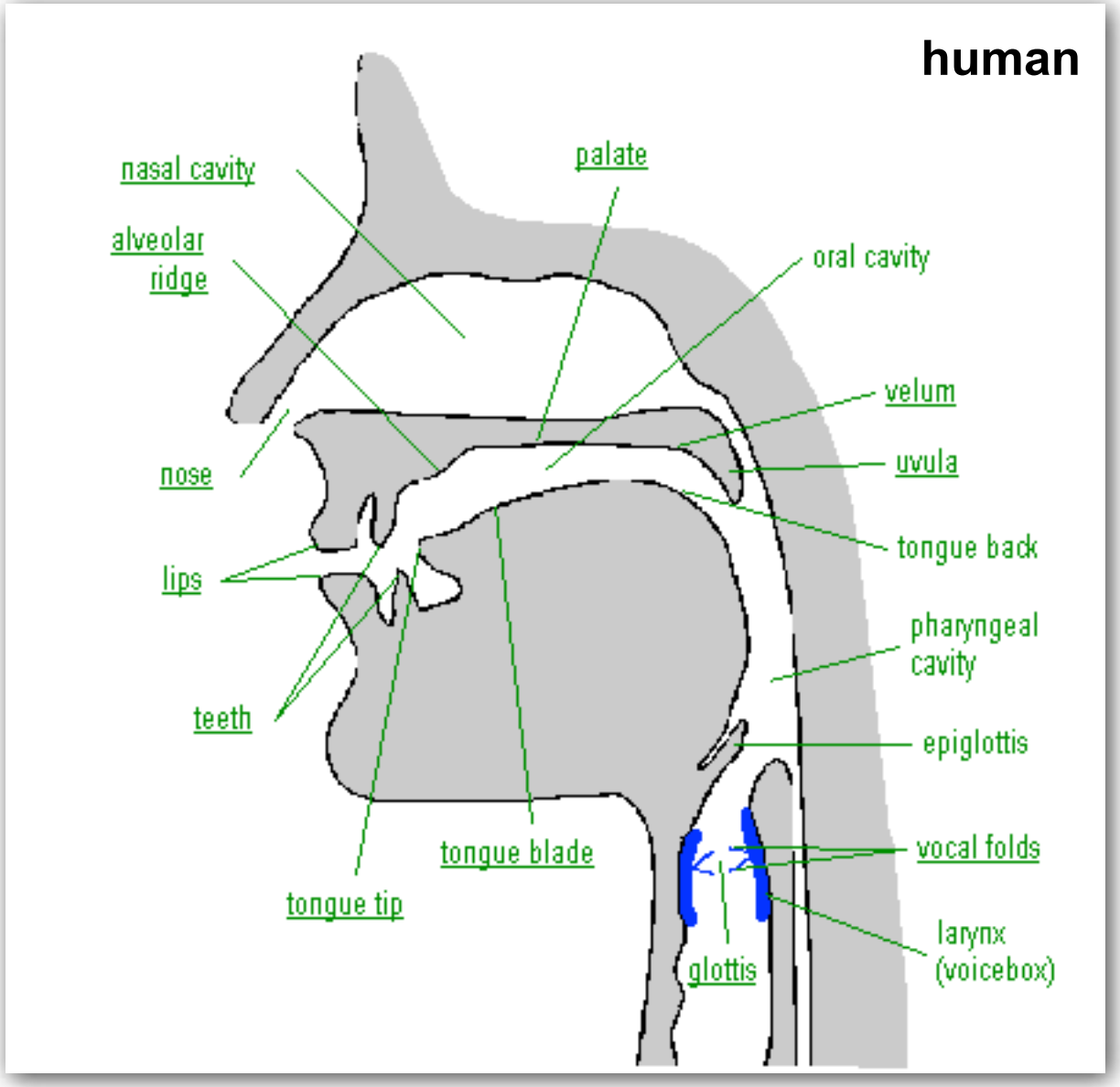


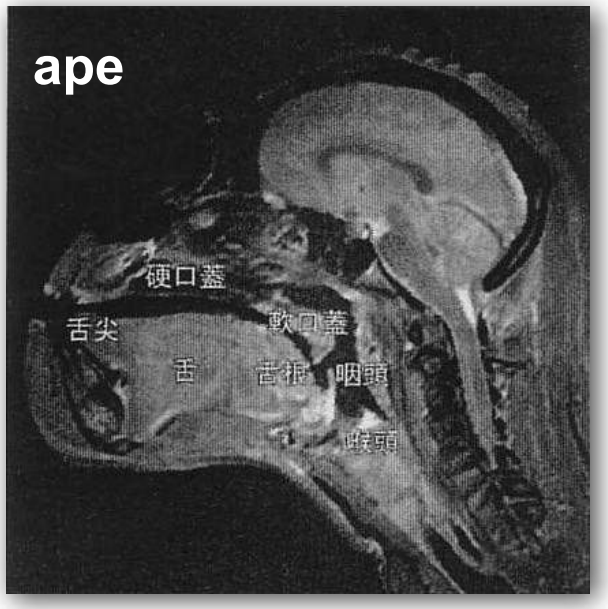
- More on general phonetics
  - General phonetics = language independent phonetics
  - How to symbolize language sounds found in any language?
- More on acoustic phonetics
  - Vowels as standing waves
    - Resonance frequency = formant frequency
  - Link between acoustic phon. and articulatory phon.
- Summary

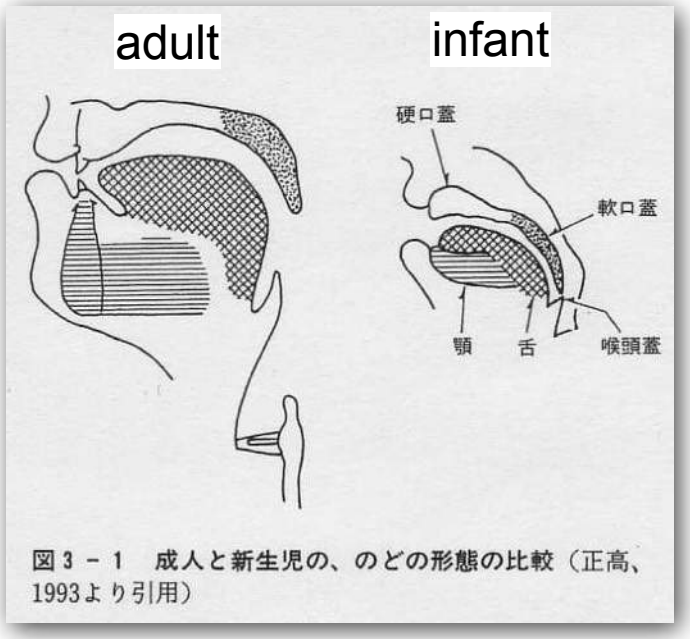


Articulatory phonetics

- Organs related to speech production

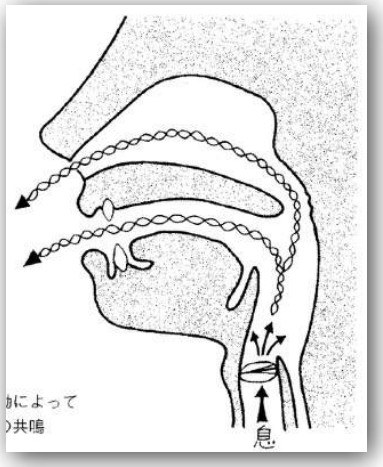


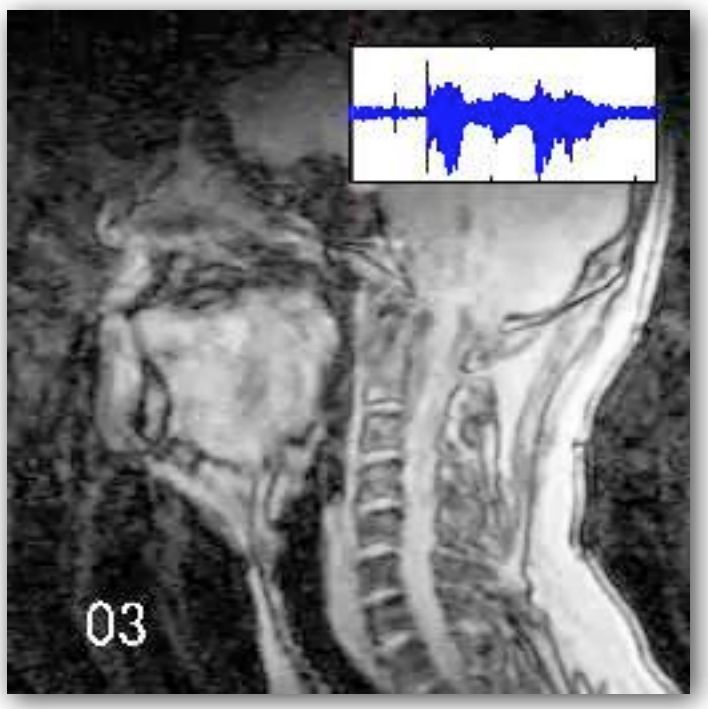


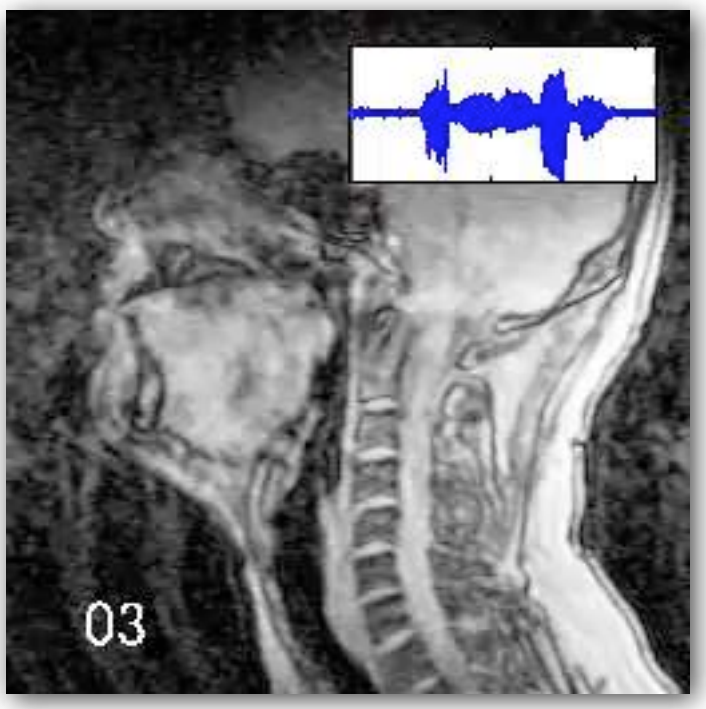


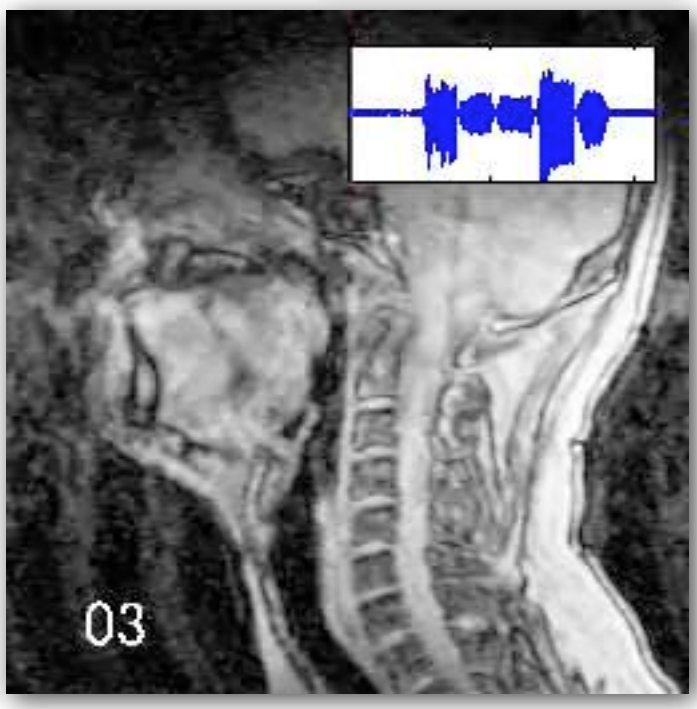
Articulatory phonetics

- Your nose (nasal cavity) can work as a special instrument.
  - Cannot produce /m/ or /n/ with your nose held closed.
  - A pathway into the nasal cavity is required to generate /n/ and /m/.





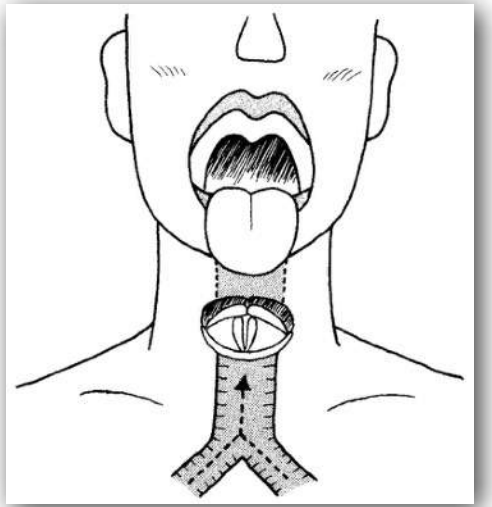




replay

Articulatory phonetics

- The glottis, generator of buzzer sounds



Today's menu

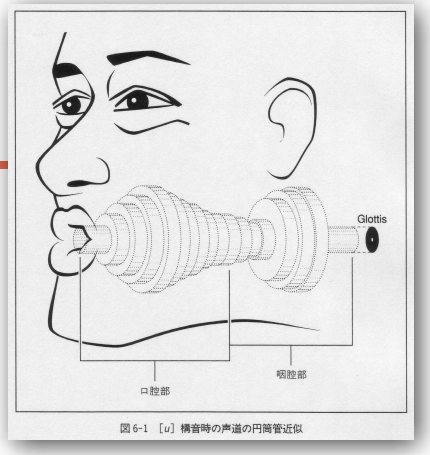
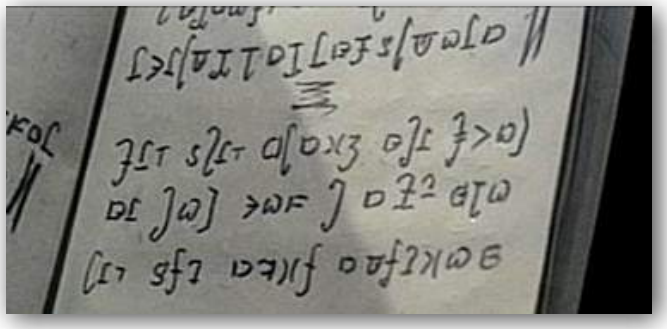
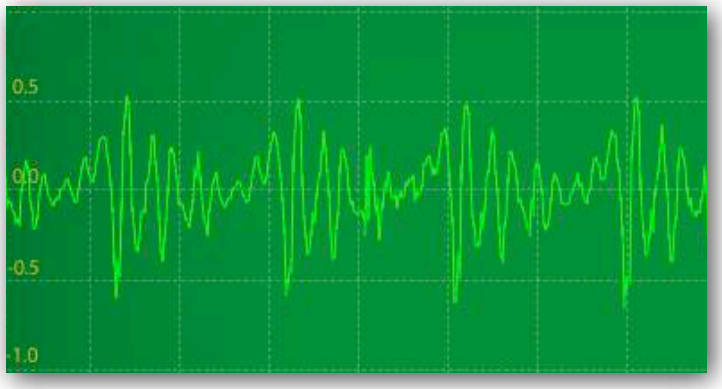


図 6-1 ［u］構音時の声道の円筒管近似

- Speech --> sounds --> vibrations (waves) of air particles
- Fundamentals of phonetics
  - How are vowel sounds produced?
  - Phonetics = articulatory phonetics + acoustic phon. + auditory phon.
- More on articulatory phonetics
  - Observation of speech organs

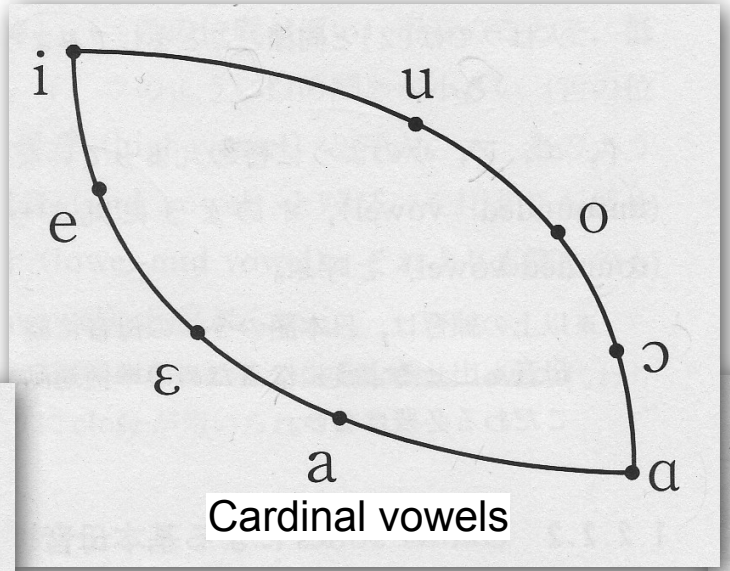


- More on general phonetics
  - General phonetics = language independent phonetics
  - How to symbolize language sounds found in any language?
- More on acoustic phonetics
  - Vowels as standing waves
    - Resonance frequency = formant frequency
  - Link between acoustic phon. and articulatory phon.
- Summary



General phonetics

- 18 fundamental and theoretical vowels -- cardinal vowels
  - Reference vowels used to describe the vowel sounds in a specific language.
    - Theoretically and artificially defined vowels
  - Position of the tongue x lip (un)rounding gives a set of 18 vowels.



Cardinal vowels

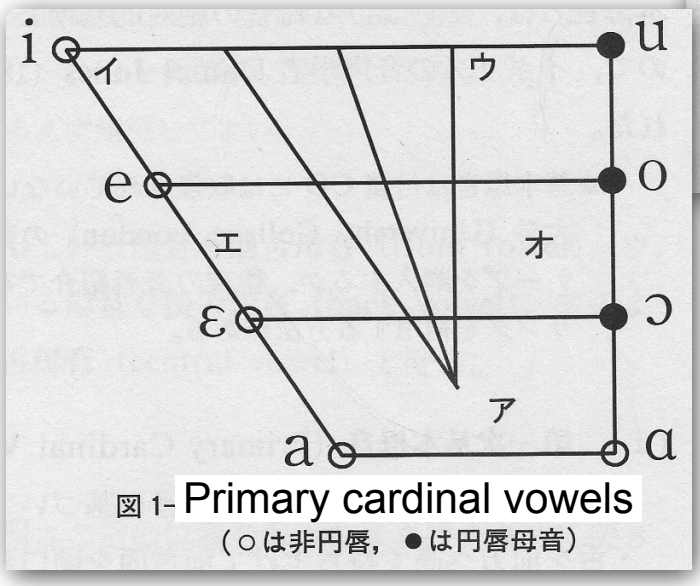
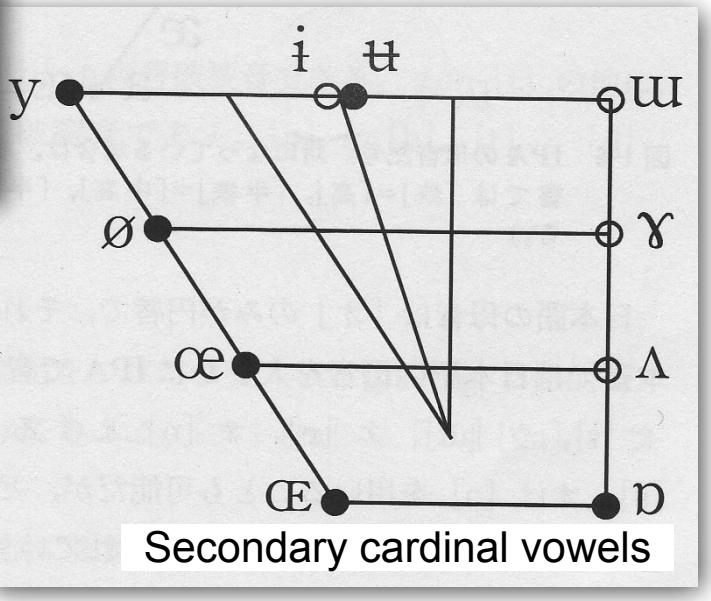


図 1- Primary cardinal vowels
（○は非円唇，●は円唇母音）

●: rounding

○: unrounding



Secondary cardinal vowels

General phonetics

- Classification of consonants
  - Complete or partial closure in the vocal tract.
  - Where and how closure happens in the vocal tract.
    - Where = place of articulation
    - How = manner of articulation
    - Condition of the vocal folds = voiced or unvoiced

place of articulation

manner of articulation

CONSONANTS (PULMONIC)

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

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

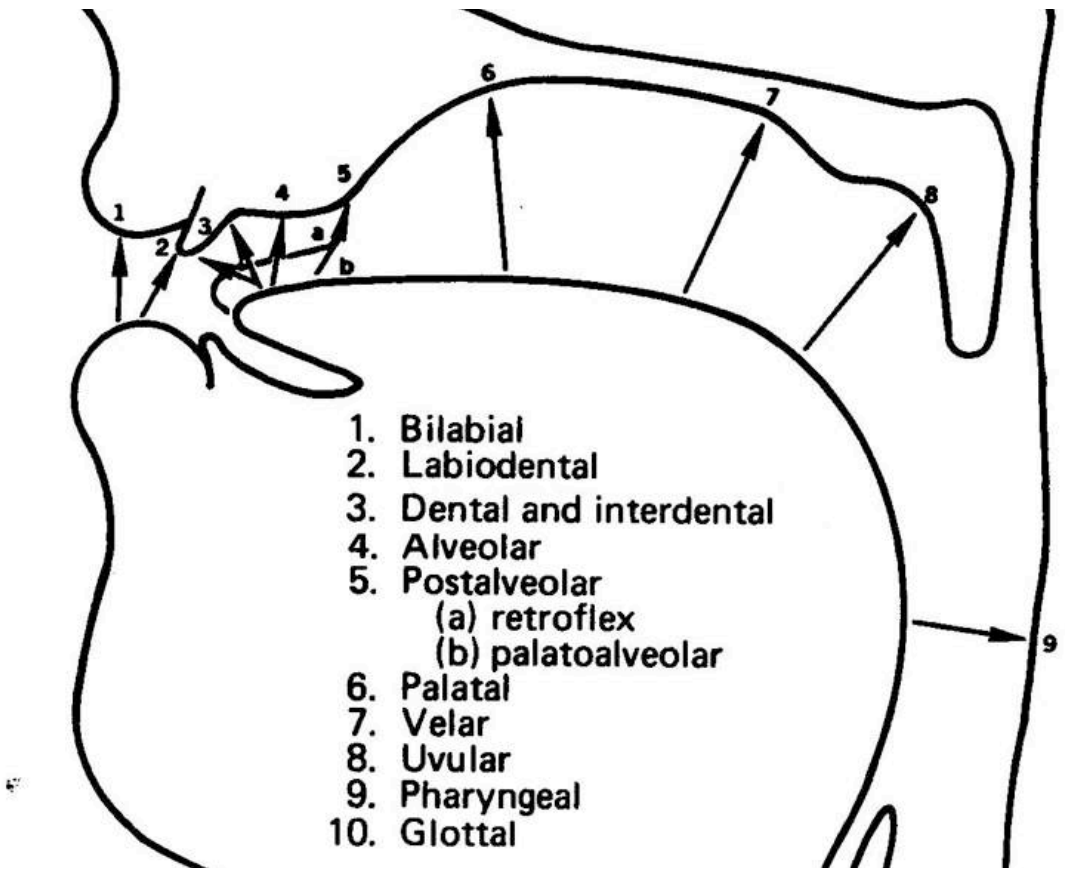
http://phonetics.ucla.edu/course/chapter1/flash.html

General phonetics

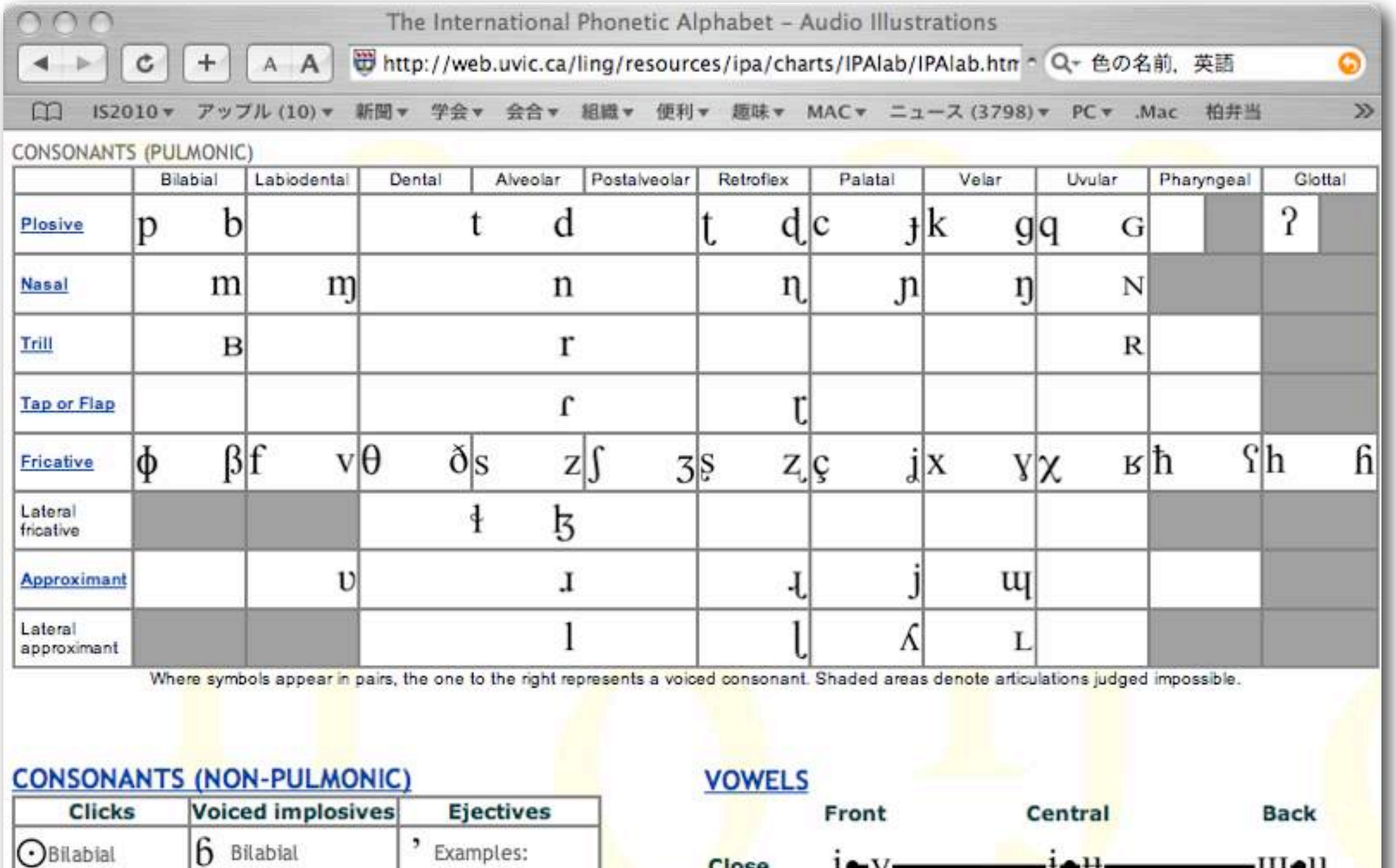
- Where complete or partial closure happens?

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Post-alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ



General phonetics



The International Phonetic Alphabet – Audio Illustrations

http://web.uvic.ca/ling/resources/ipa/charts/IPAlab/IPAlab.htm

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

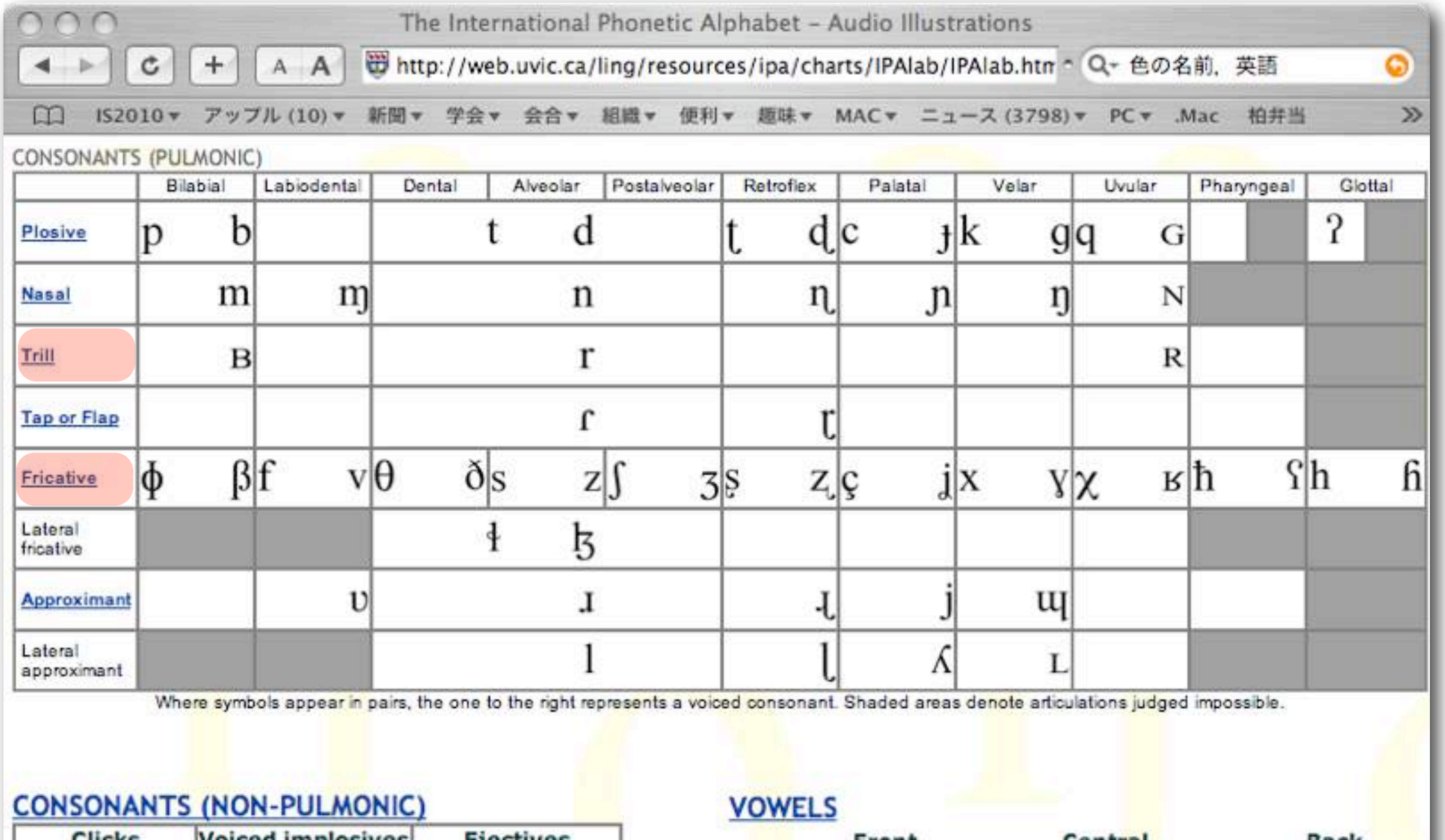
Clicks	Voiced implosives	Ejectives
ʘ Bilabial	ɓ Bilabial	’ Examples:

VOWELS

	Front	Central	Back
Close	i • y	ɨ • ʉ	ɯ • u

General phonetics

http://web.uvic.ca/ling/resources/ipa/charts/IPAlab/IPAlab.htm



The International Phonetic Alphabet – Audio Illustrations

http://web.uvic.ca/ling/resources/ipa/charts/IPAlab/IPAlab.htm

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q G		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	N		
Trill	B			r					R		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

VOWELS

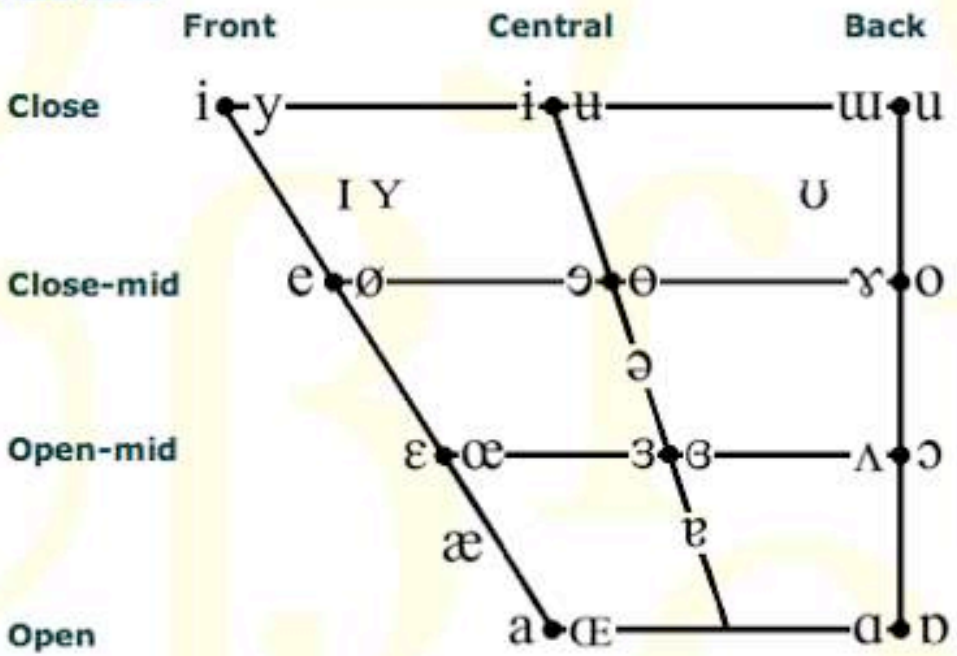
General phonetics

http://web.uvic.ca/ling/resources/ipa/charts/IPAlab/IPAlab.htm

CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
ʘ Bilabial	ɓ Bilabial	’ Examples:
ǀ Dental	ɗ Dental/alveolar	pʼ Bilabial
ǃ (Post)alveolar	ʄ Palatal	tʼ Dental/alveolar
ǂ Palatoalveolar	ɠ Velar	kʼ Velar
ǁ Alveolar lateral	ʛ Uvular	sʼ Alveolar fricative

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

OTHER SYMBOLS

ʍ Voiceless labial-velar fricative

w Voiced labial-velar approximant

ɥ Voiced labial-palatal approximant

ʜ Voiceless epiglottal fricative

ʢ Voiced epiglottal fricative

ʡ Epiglottal plosive

ɕ ʑ Alveolo-palatal fricatives

ɺ Alveolar lateral flap

ɧ Simultaneous ʃ and x

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary. k͡p t͡s

SUPRASEGMENTALS

ˈ Primary stress

ˌ Secondary stress

Today's menu

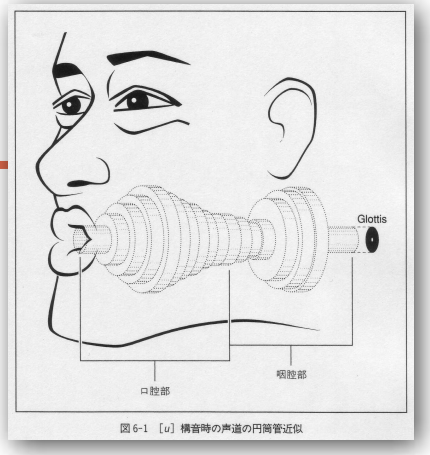
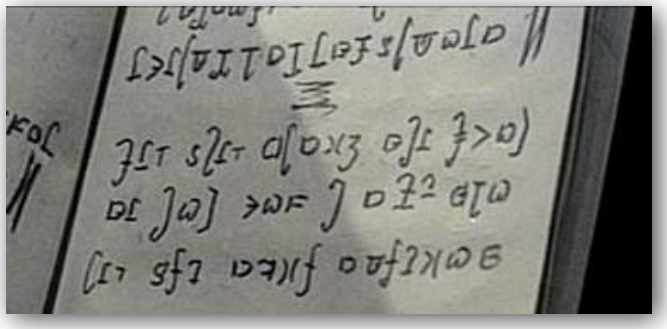
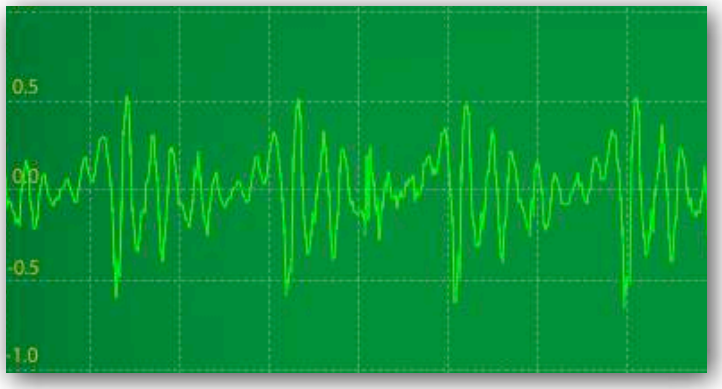


図 6-1 ［u］構音時の声道の円筒管近似

- Speech --> sounds --> vibrations (waves) of air particles
- Fundamentals of phonetics
  - How are vowel sounds produced?
  - Phonetics = articulatory phonetics + acoustic phon. + auditory phon.
- More on articulatory phonetics
  - Observation of speech organs

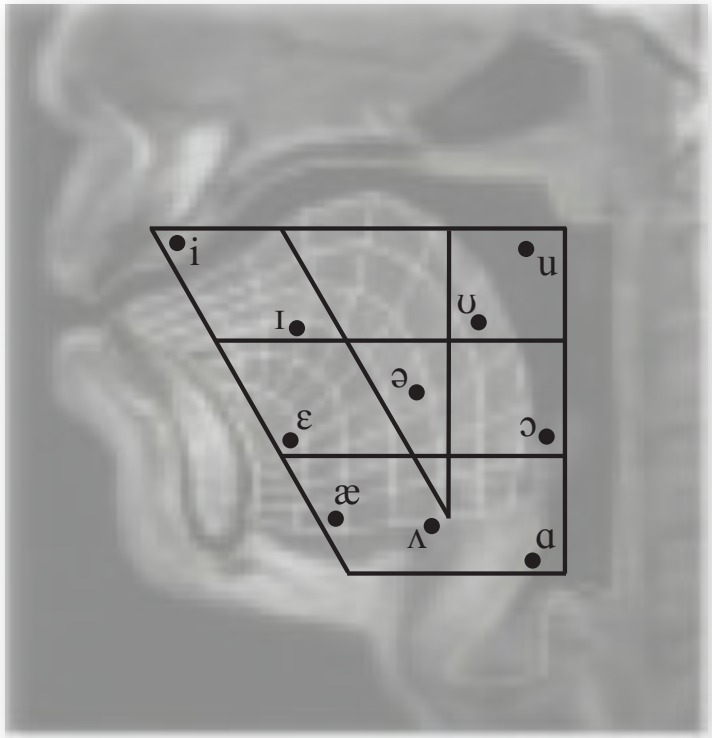


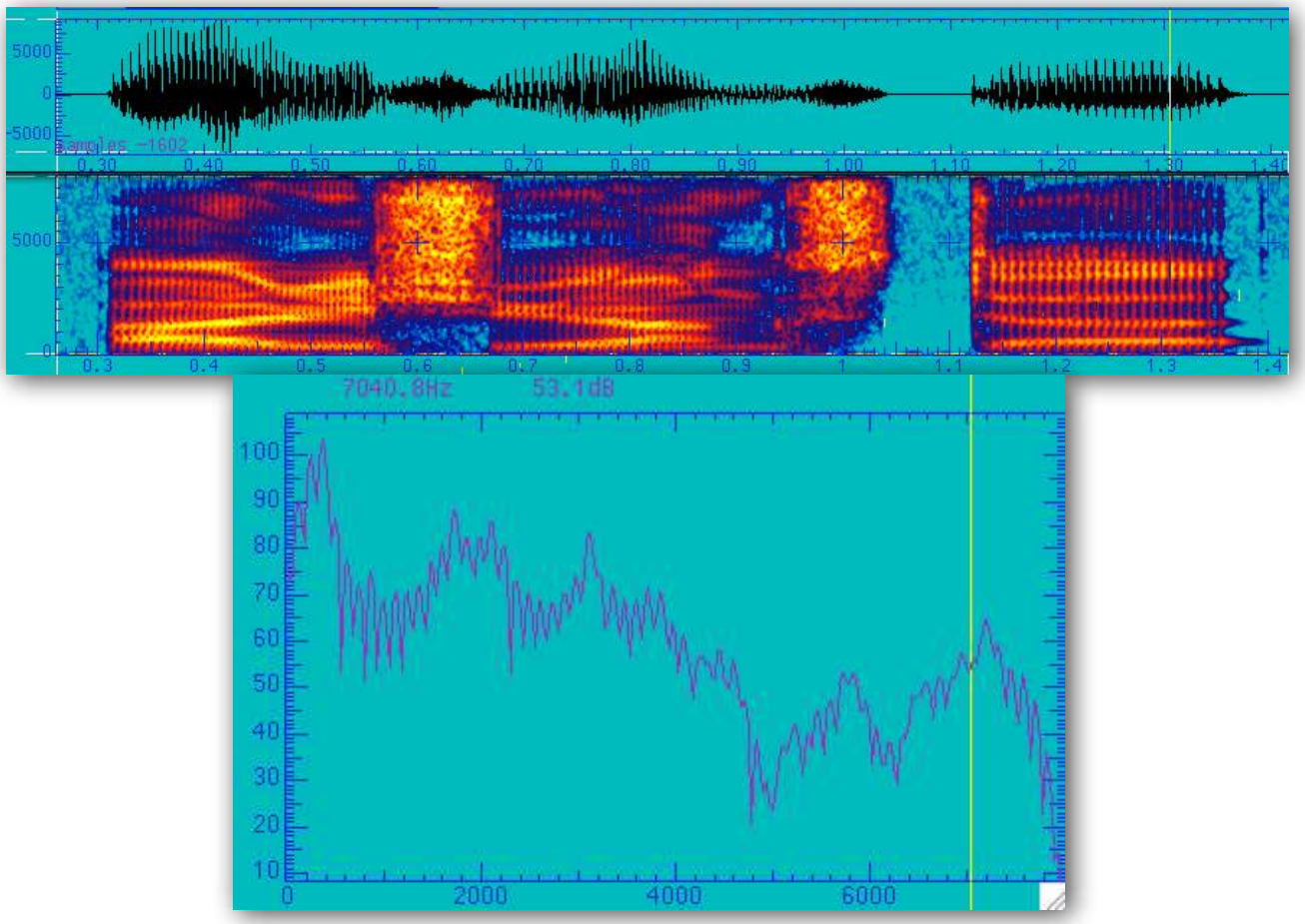
- More on general phonetics
  - General phonetics = language independent phonetics
  - How to symbolize language sounds found in any language?
- More on acoustic phonetics
  - Vowels as standing waves
    - Resonance frequency = formant frequency
  - Link between acoustic phon. and articulatory phon.
- Summary



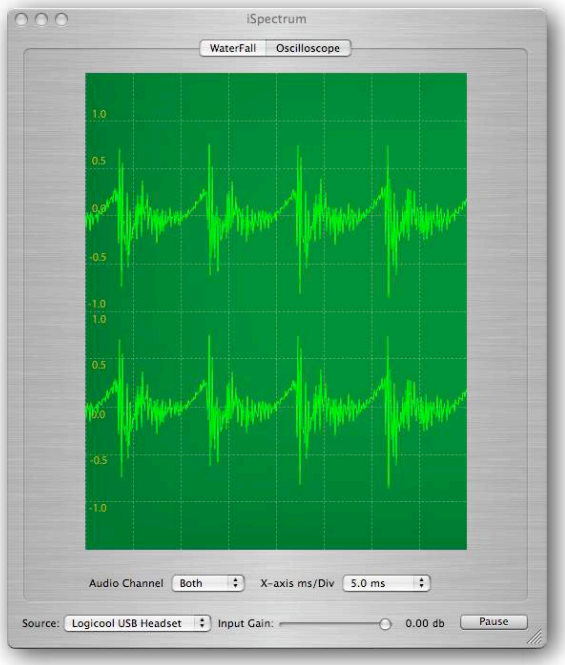
Acoustic phonetics

- Articulatory phonetics
  - Focus is on how speech organs generate individual language sounds.
- Acoustic phonetics
  - Focus is on what kind of acoustic characteristics are observed in individual sounds.





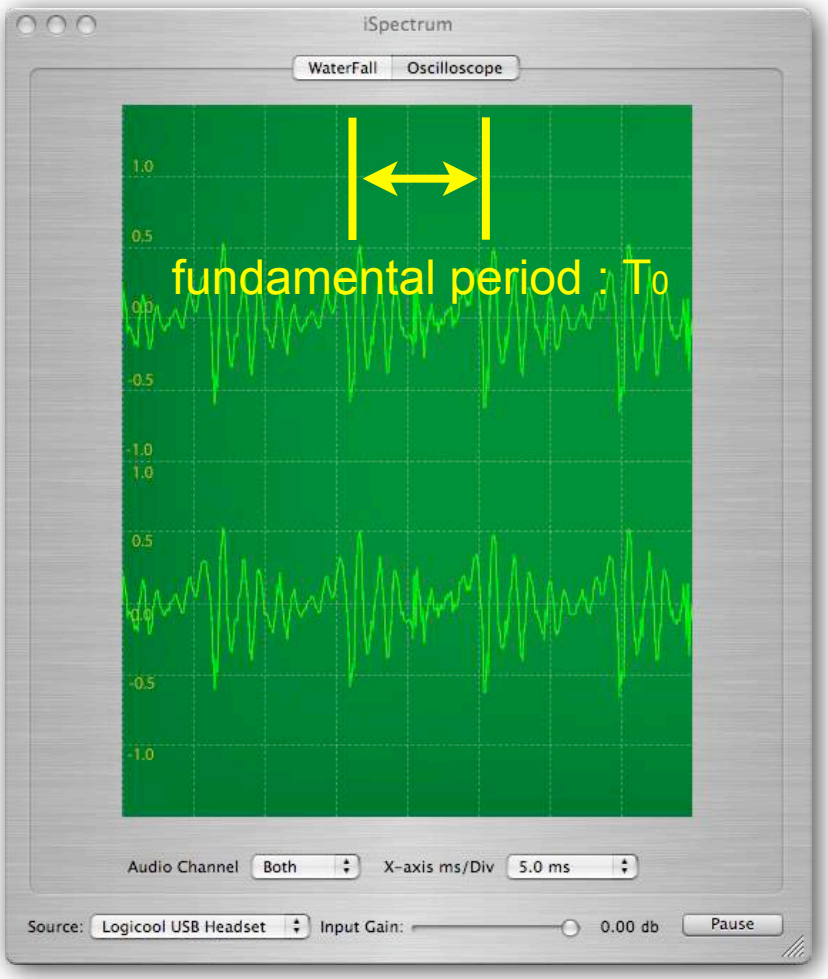
Speech = vibrations of air particles

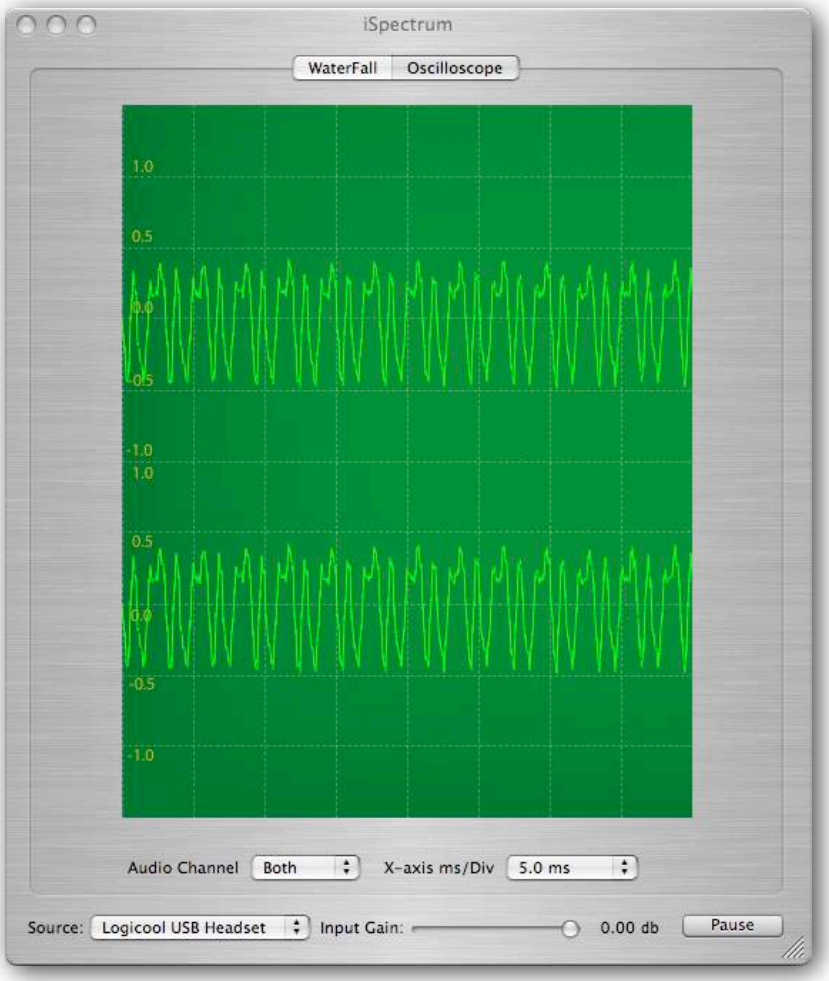


- The four aspects of tones (sounds)
  - Height of tones (pitch of tones)
    - High tones and low tones
  - Loudness of tones
    - Loud tones and soft tones
  - Duration of tones
    - Long tones and short tones
  - Timbre of tones (color of tones, 音色, 声色)
    - ????
    - If two tones have the same height, the same loudness, and the same duration but the two tones are perceived as different tones, then, the two tones differ in their timbre.
    - /a/ and /i/ /a/ and /a/
      - difference in phoneme, difference in gender

Speech = vibrations of air particles

- Close observation of air particle vibration patterns.
  - Low /a/ and high /a/ in pitch
  - F_0 : fundamental frequency (pitch) = $1/T_0$ = 1/fundamental period





Acoustic phonetics

- Vowel = a special kind of compression waves (longitudinal waves)

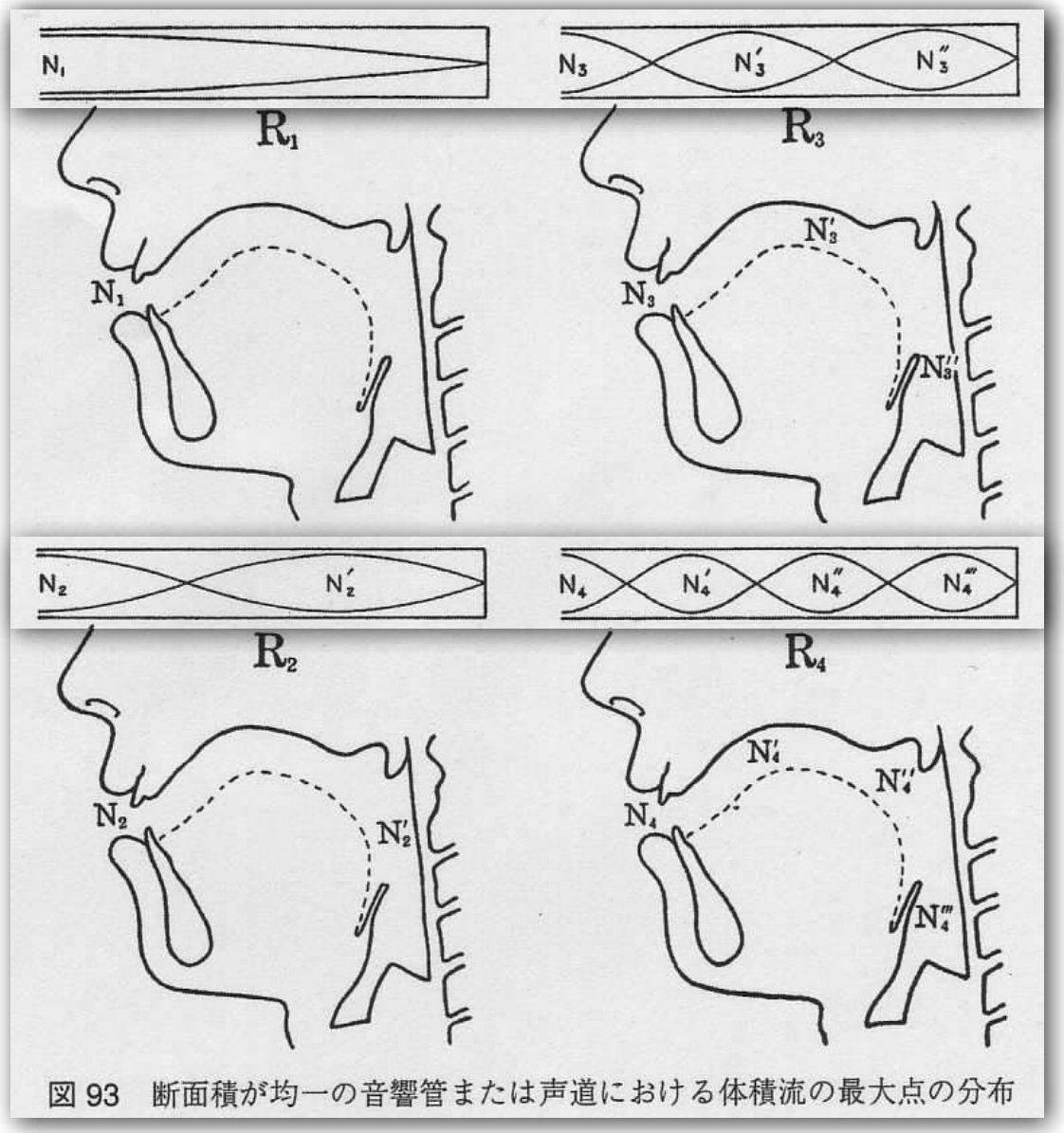
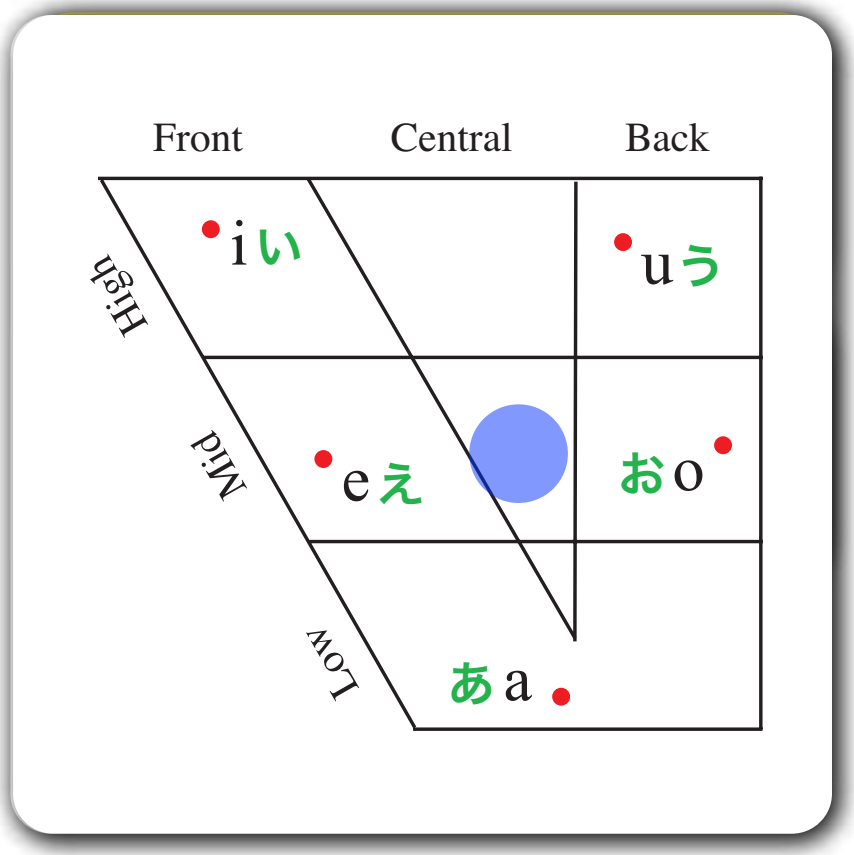
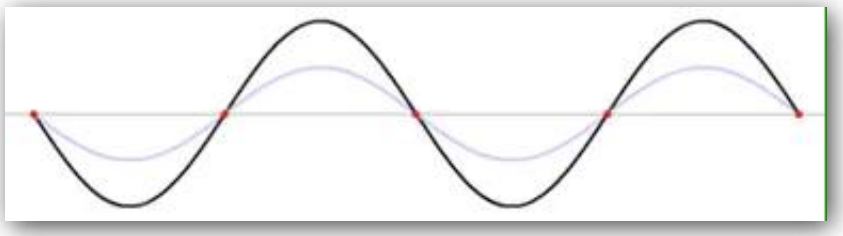


図 93　断面積が均一の音響管または声道における体積流の最大点の分布





Acoustic phonetics

- Other vowels = standing waves generated through a complicated tube

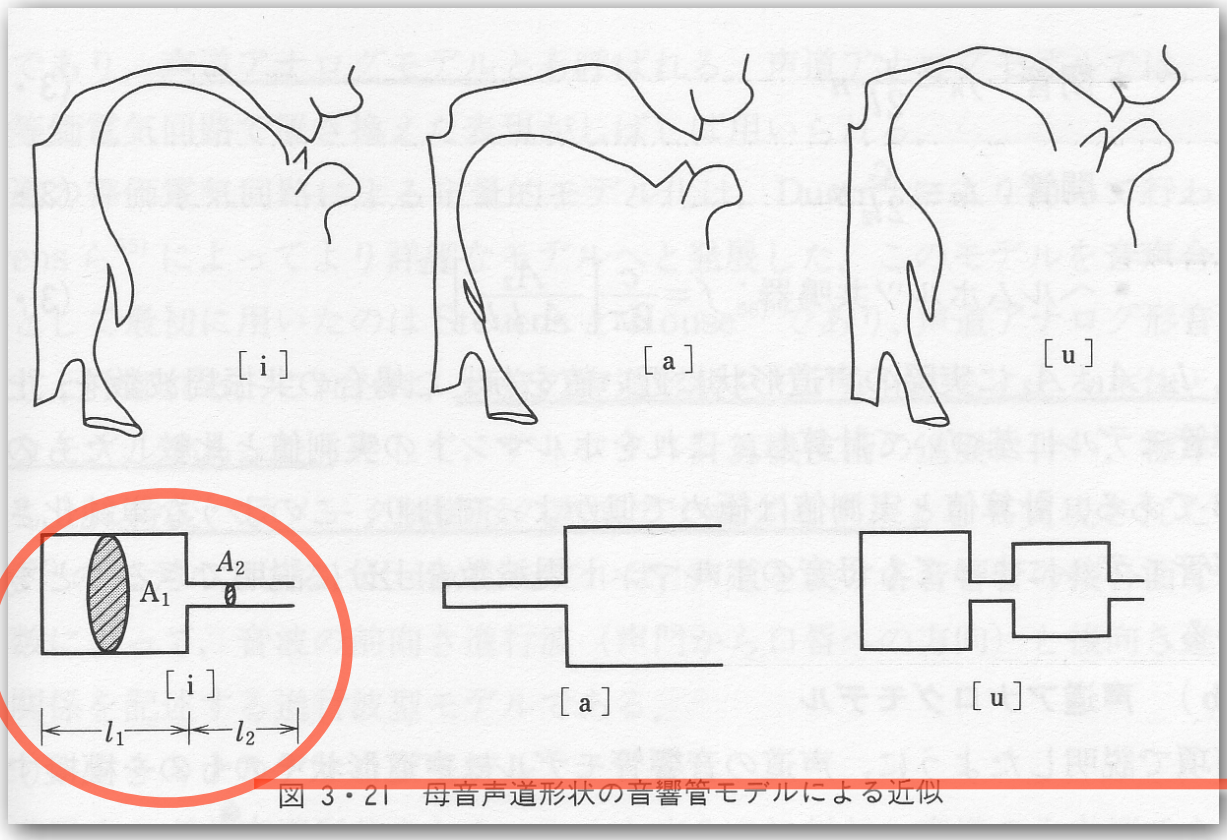


図 3・21　母音声道形状の音響管モデルによる近似

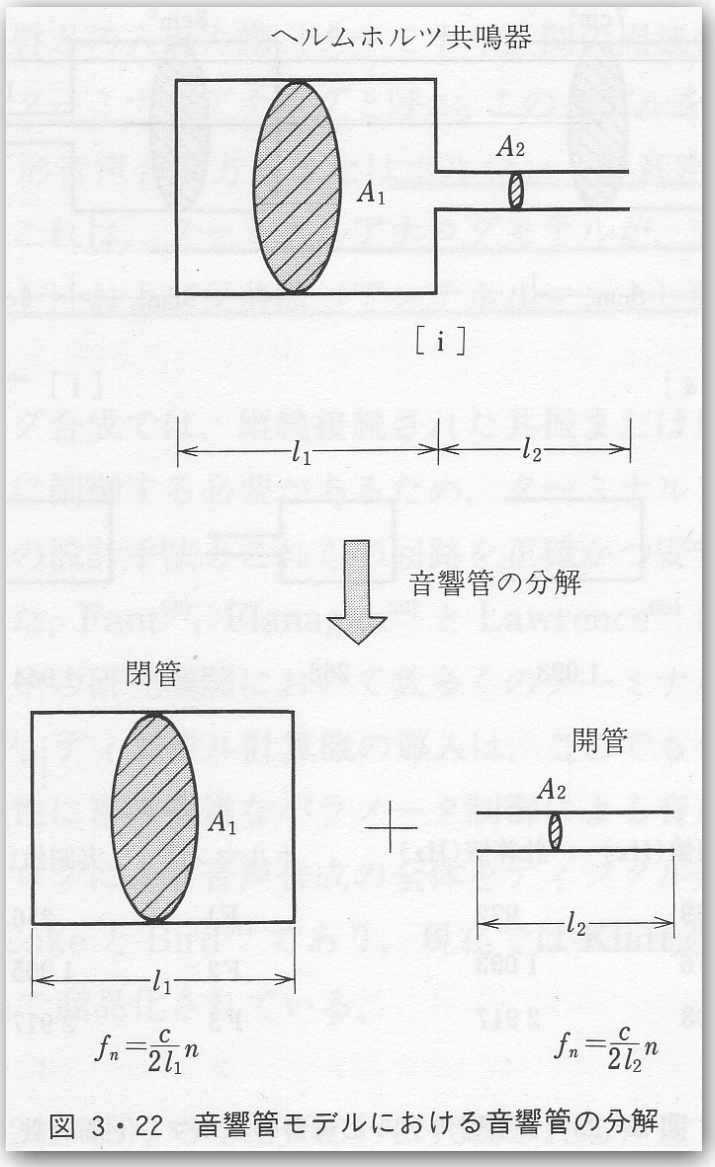
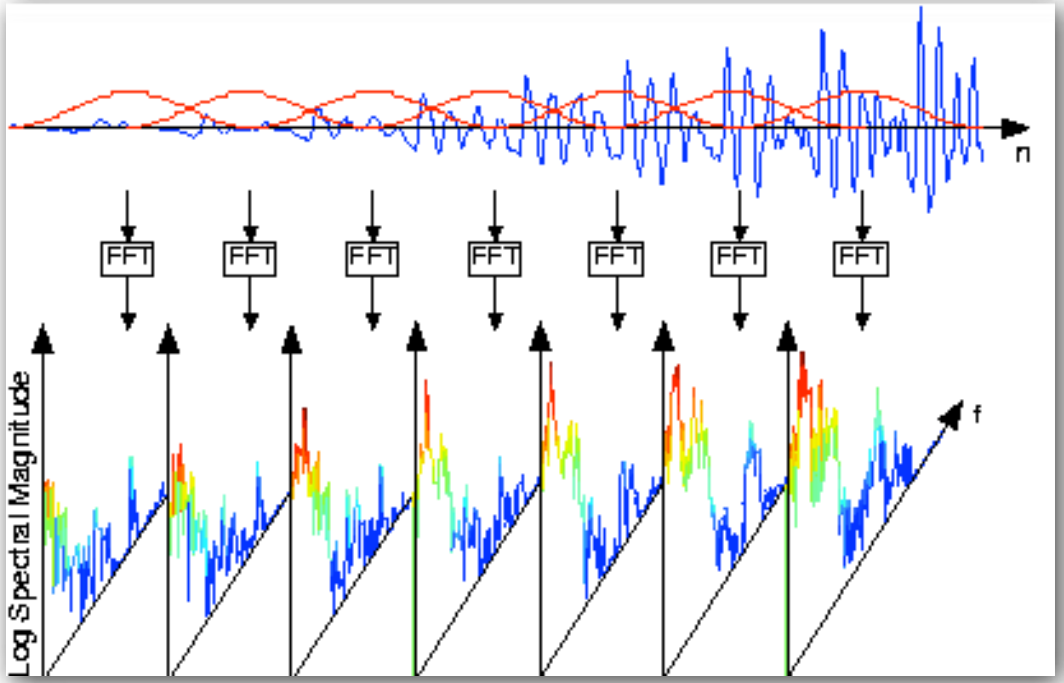


図 3・22　音響管モデルにおける音響管の分解

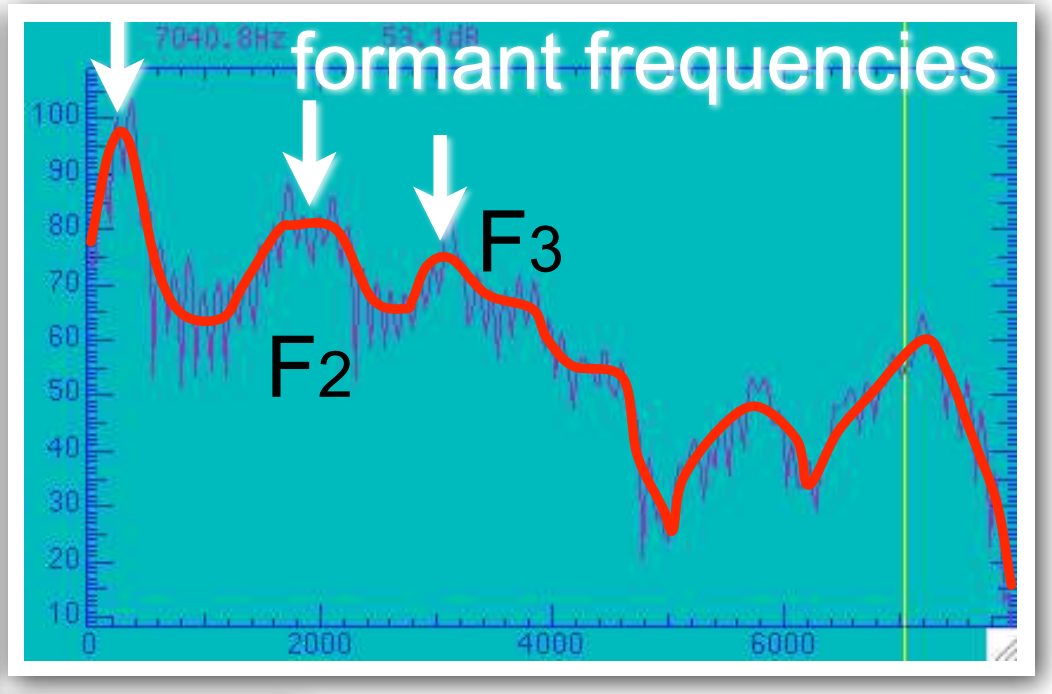
$$f_n = \frac{c}{2l_1}n \qquad f_n = \frac{c}{2l_2}n \qquad f = \frac{c}{2\pi}\left[\frac{A_2}{A_1 l_1 l_2}\right]^{1/2}$$

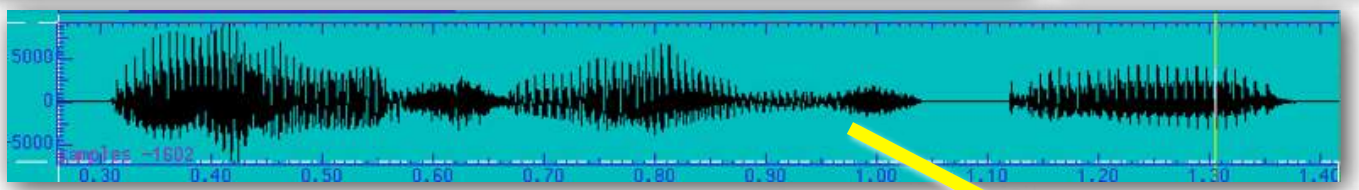
Acoustic phonetics

- From waveforms to spectrums
  - Windowing + FFT + log-amplitude

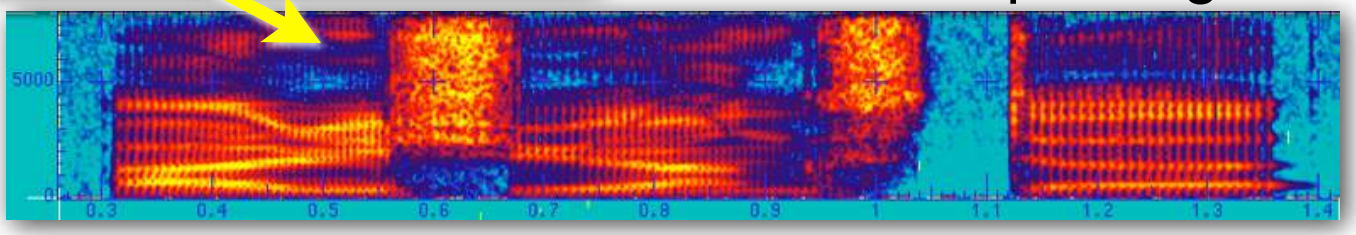


F_1



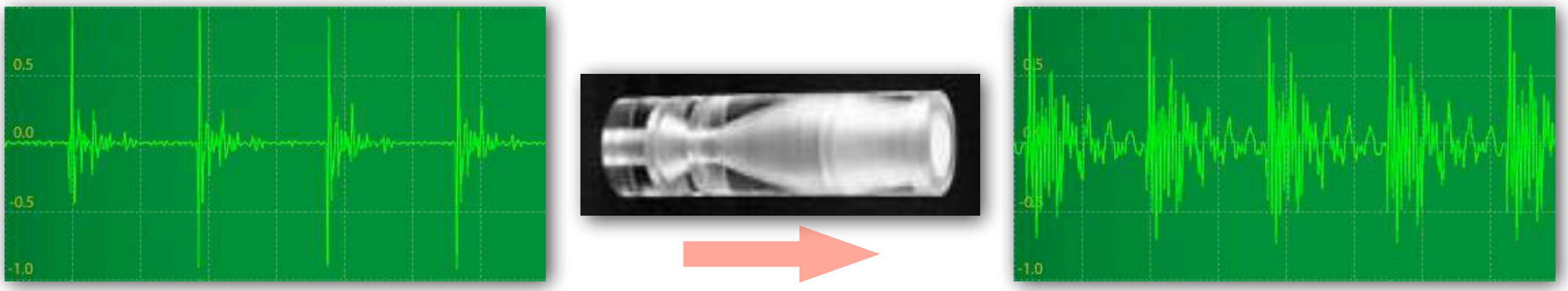


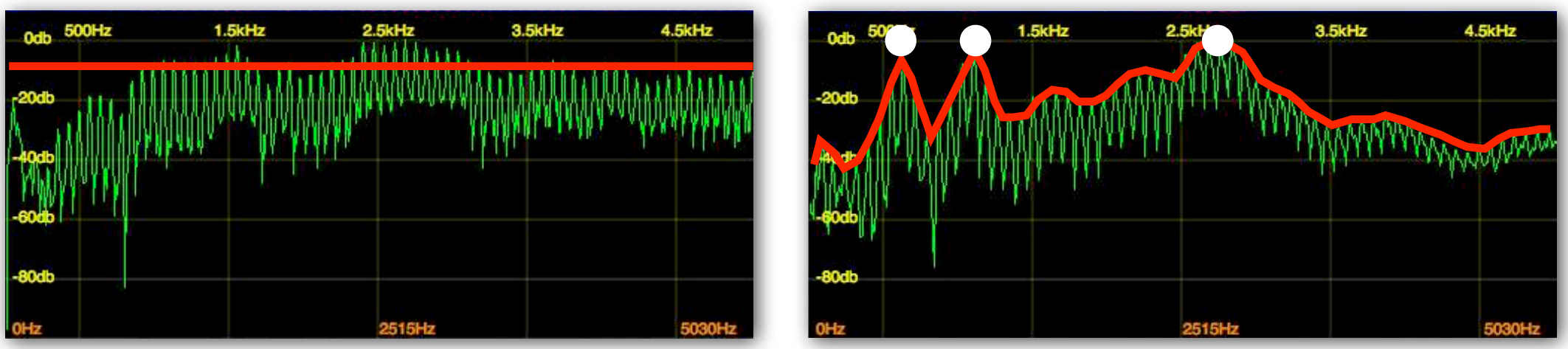
spectrogram



Acoustic phonetics

- Spectrum of a vowel sound





Resonance = concentration of the energy on specific bands that are determined only by the shape of a tube used for sound generation.

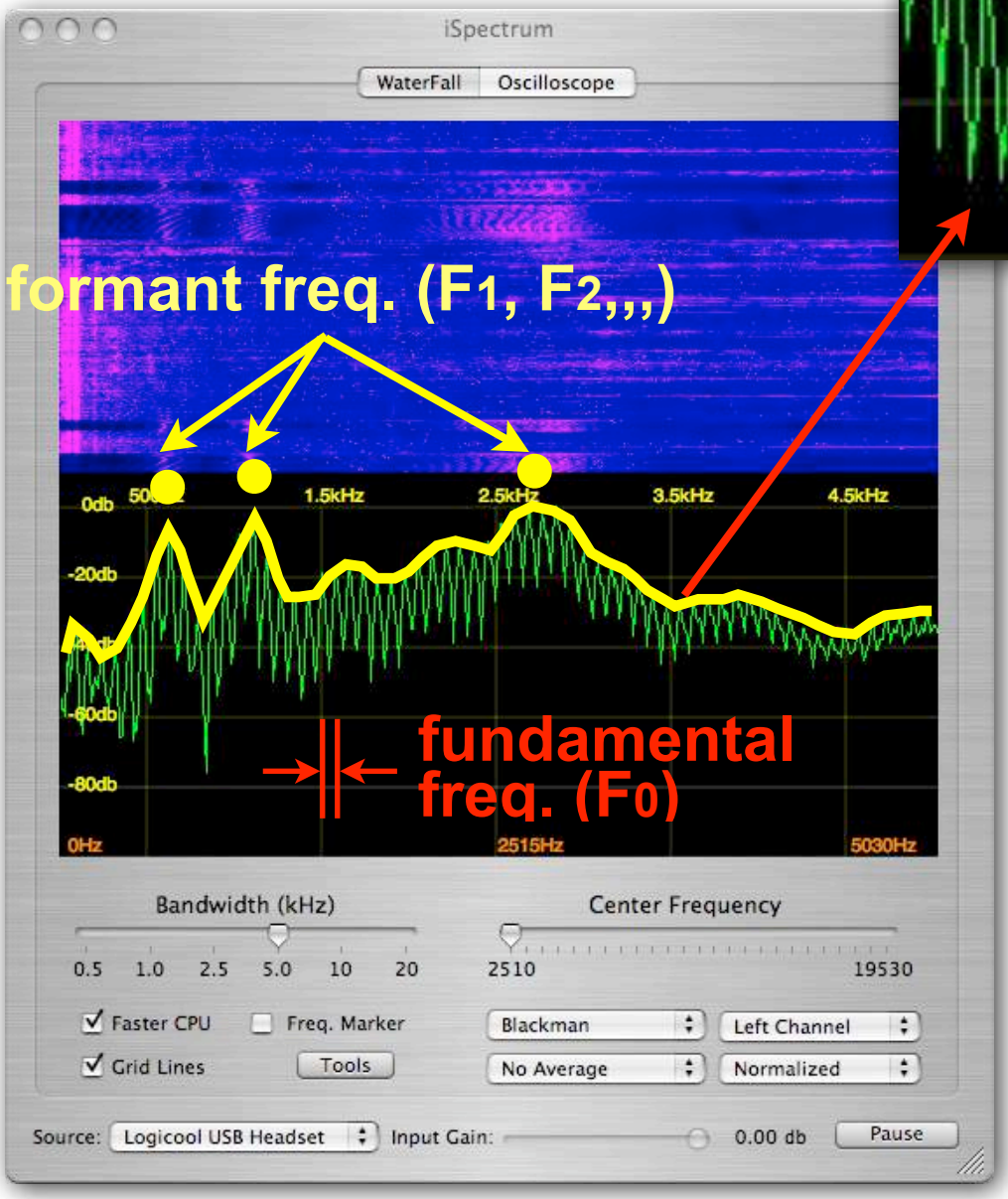
Timbre = energy distribution pattern over the frequency axis

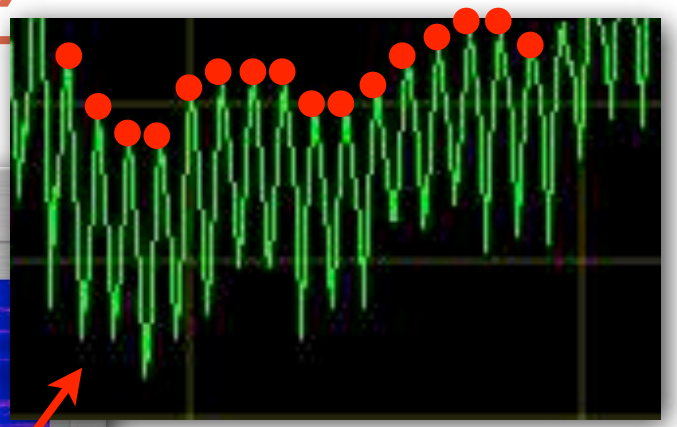
Fundamental frequency (F_0) and timbre

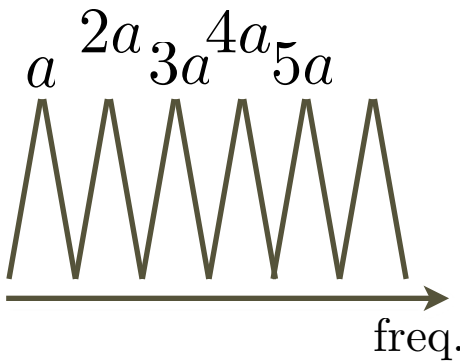
- F_0 and timbre observed in the spectrum

喉の形を変えると共振周波数が変わる。つまり，エネルギー分布の様子（パワースペクトル）が変わる。

これを，音響用語では音色と呼ぶ。楽器の違いは音色の違い，母音の違いも音色の違いである。話者の違いもまた，音色の違いである



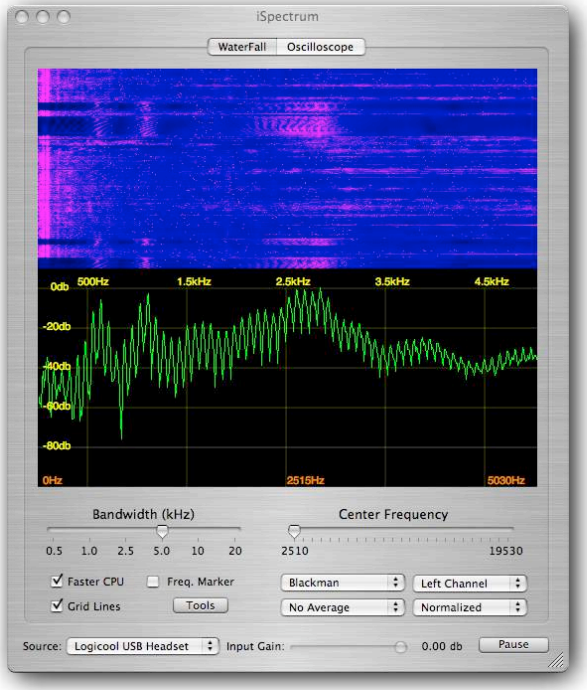




厳密には「音高＝a」であって，ピークの間隔ではない。調波構造が無くても音高は感覚できる。

Speech = vibrations of air particles

- The four aspects of tones (sounds)
  - Height of tones (pitch of tones)
    - High tones and low tones
  - Loudness of tones
    - Loud tones and soft tones
  - Duration of tones
    - Long tones and short tones
  - Timbre of tones (color of tones, 音色, 声色)
    - ????
    - If two tones have the same height, the same loudness, and the same duration but the two tones are perceived as different tones, then, the two tones differ in their timbre.
    - /a/ and /i/ /a/ and /a/
      - difference in phoneme, difference in gender



Timbre = energy distribution pattern over the frequency axis

Determined only by the shape of a tube used for sound generation

Acoustic phonetics

- Other vowels = standing waves generated through a complicated tube

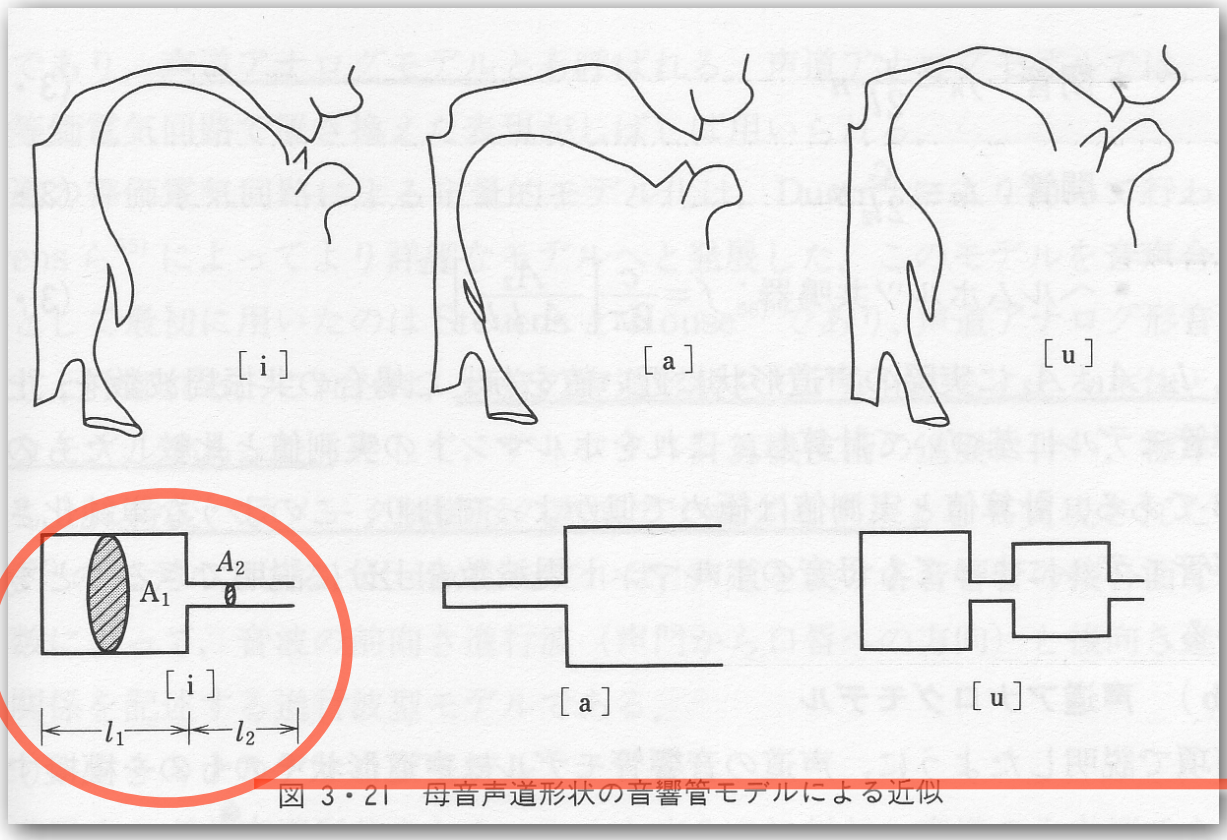


図 3・21　母音声道形状の音響管モデルによる近似

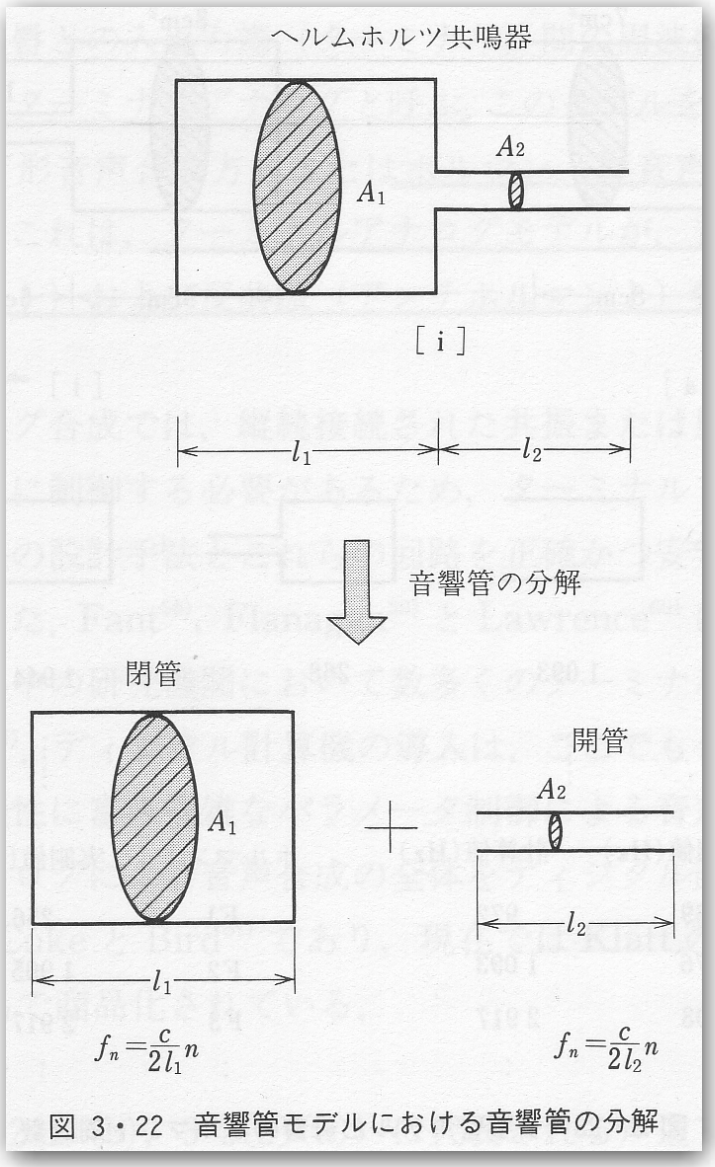
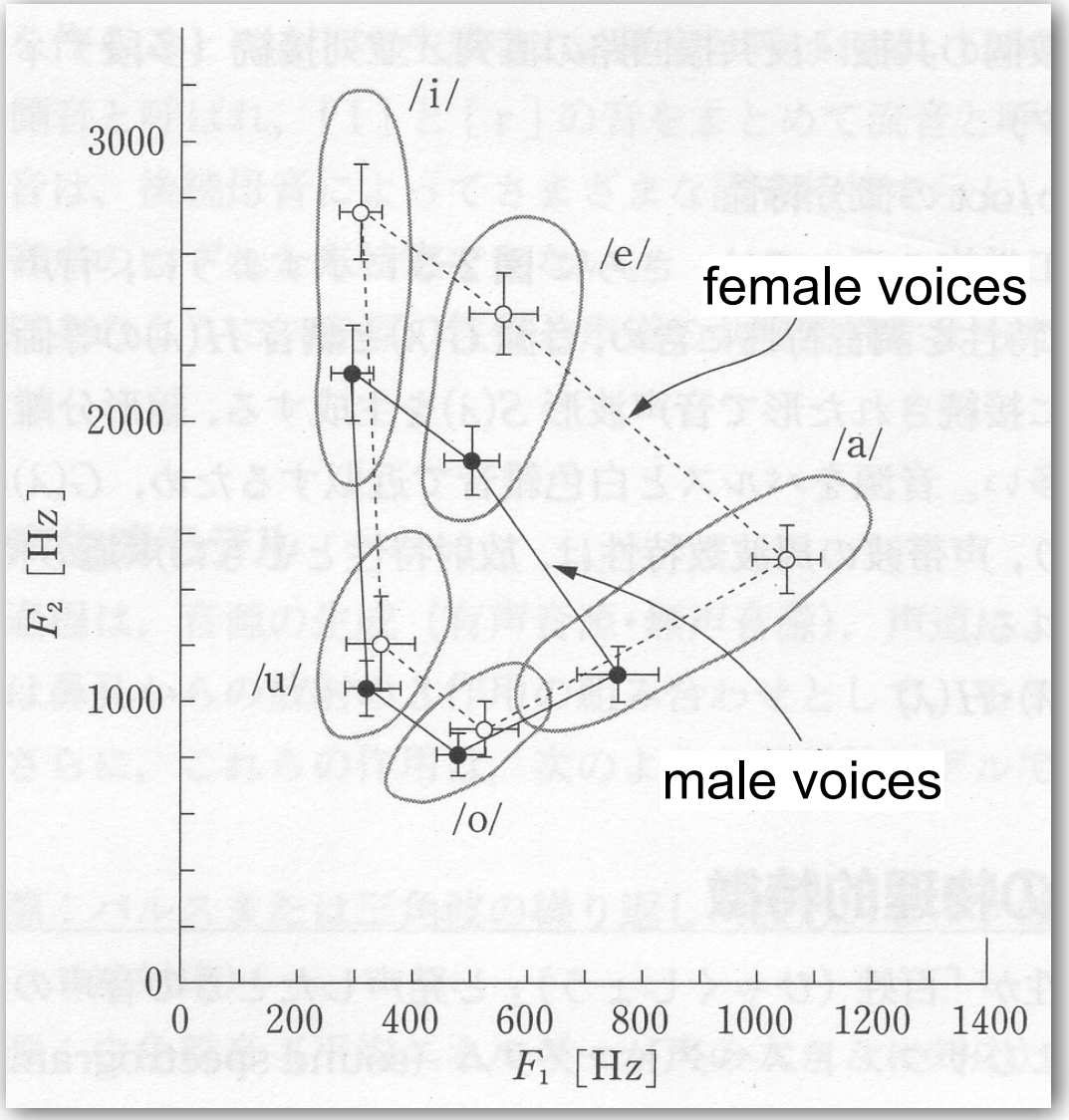


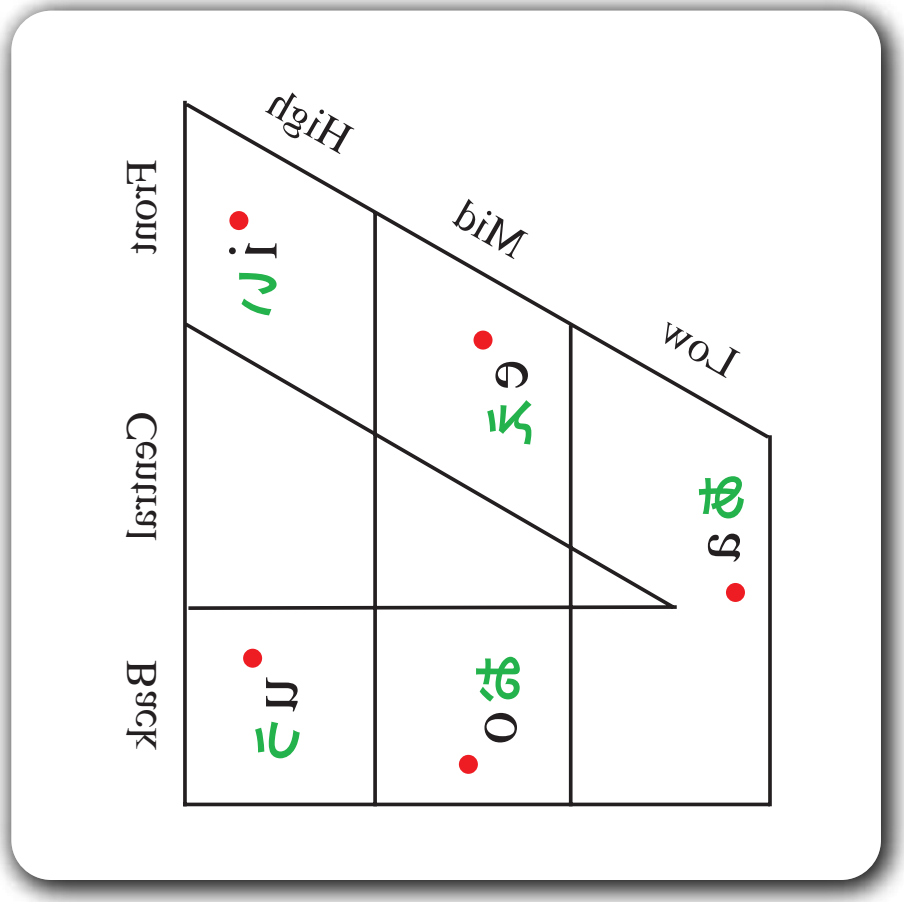
図 3・22　音響管モデルにおける音響管の分解

$$f_n = \frac{c}{2l_1}n \qquad f_n = \frac{c}{2l_2}n \qquad f = \frac{c}{2\pi}\left[\frac{A_2}{A_1 l_1 l_2}\right]^{1/2}$$

Acoustic and articulatory phonetics

- Shape difference = resonance frequency difference
  - /a/ and /i/ /a/ and /a/





Today's menu

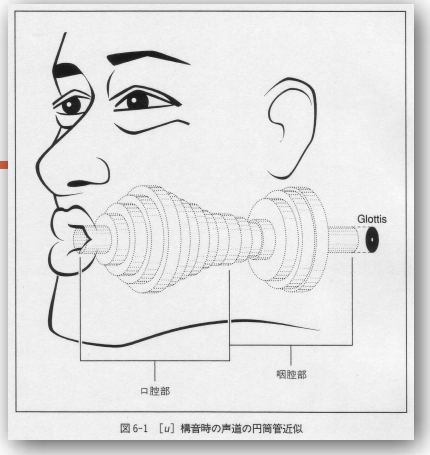
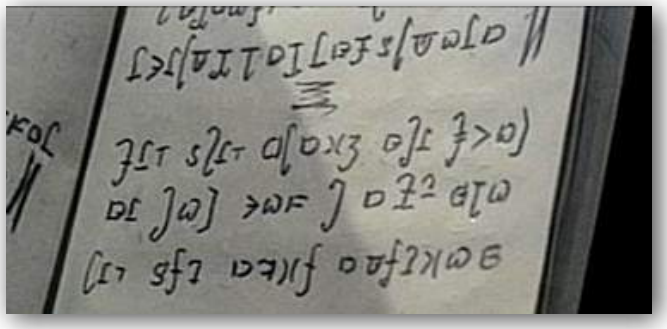
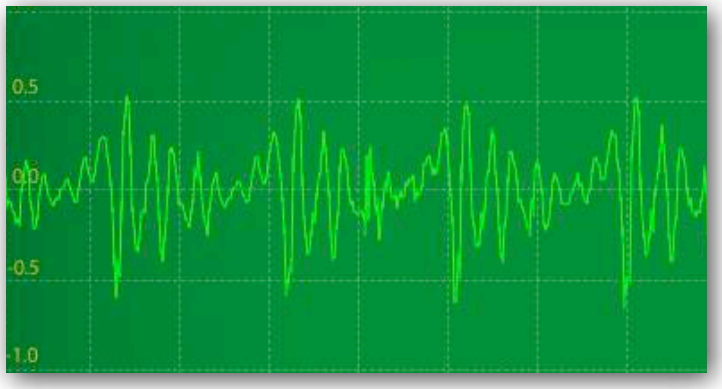


図 6-1 ［u］構音時の声道の円筒管近似

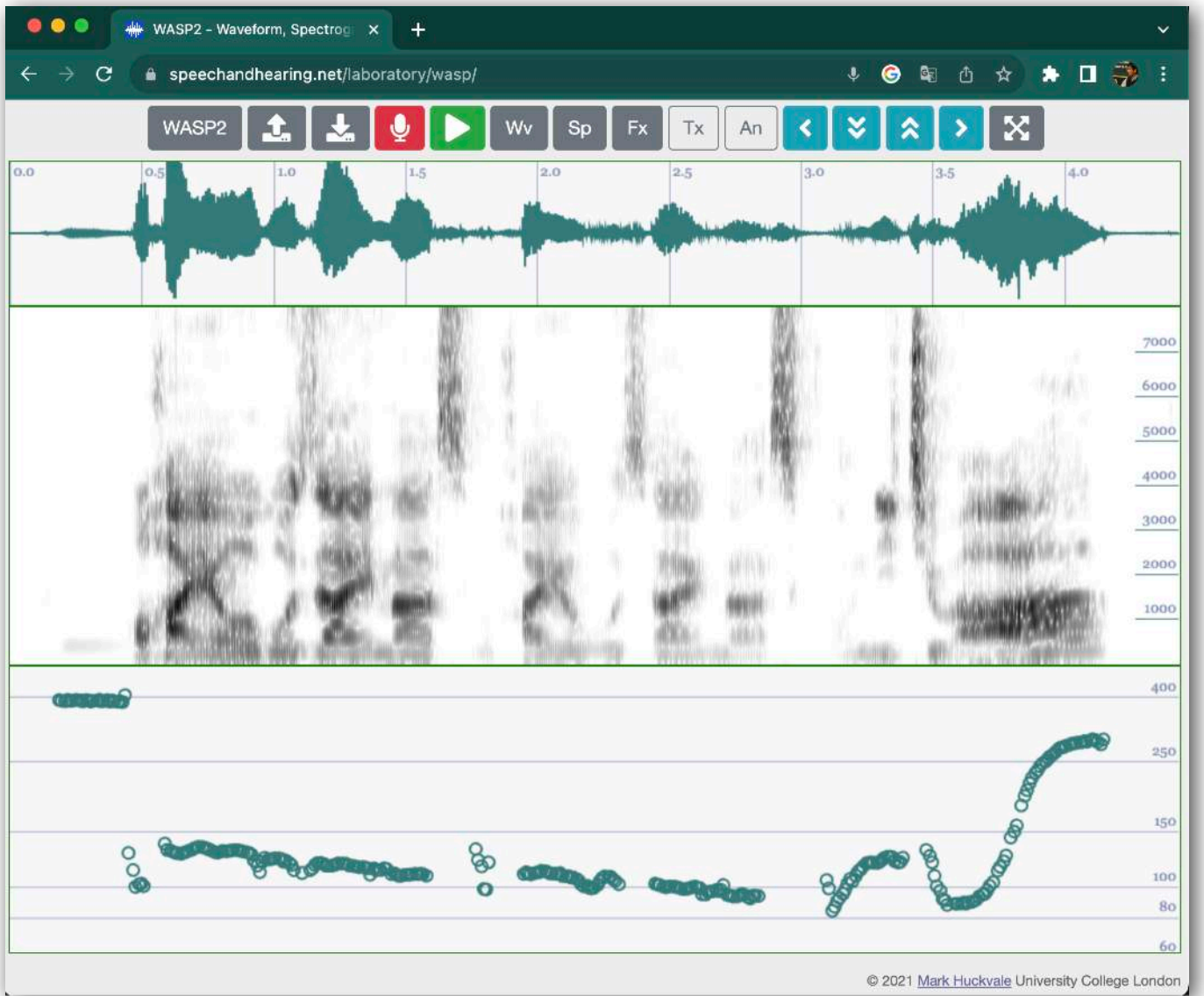
- Speech --> sounds --> vibrations (waves) of air particles
- Fundamentals of phonetics
  - How are vowel sounds produced?
  - Phonetics = articulatory phonetics + acoustic phon. + auditory phon.
- More on articulatory phonetics
  - Observation of speech organs



- More on general phonetics
  - General phonetics = language independent phonetics
  - How to symbolize language sounds found in any language?
- More on acoustic phonetics
  - Vowels as standing waves
    - Resonance frequency = formant frequency
  - Link between acoustic phon. and articulatory phon.
- Summary



Web-based speech analyzer



https://www.speechandhearing.net/laboratory/wasp/

Recommended books