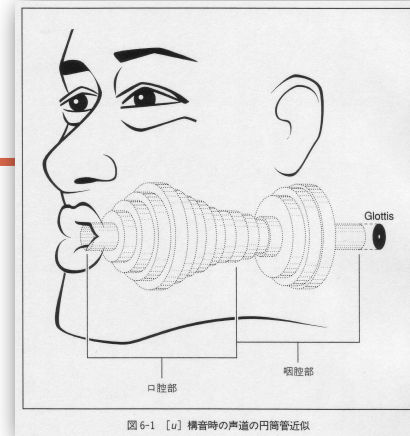


Cognitive Media Processing #6

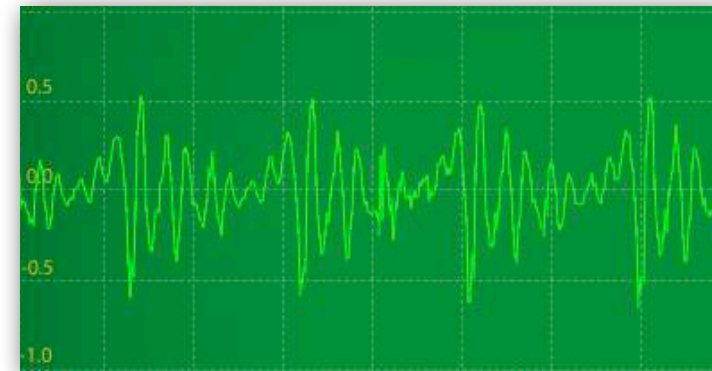
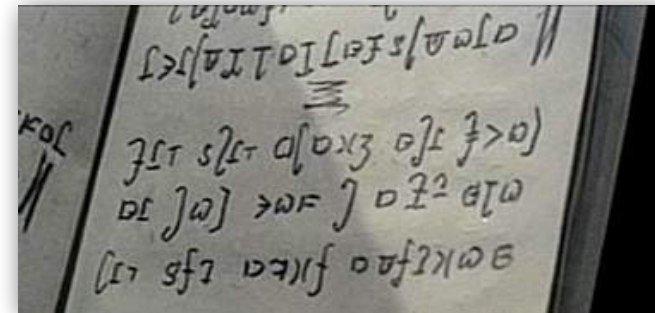
Nobuaki Minematsu



Today's menu

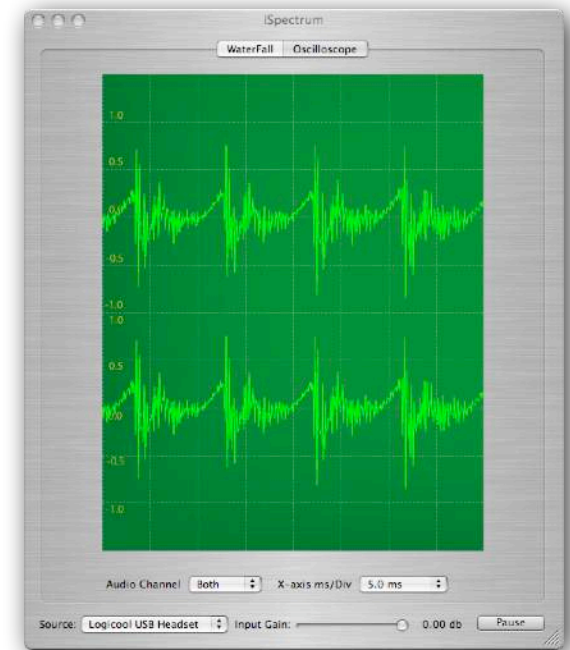


- Speech --> sounds --> vibrations (waves) of air particles
- Fundamentals of phonetics
 - How are vowel sounds produced?
 - Phonetics = **articulatory** phonetics + **acoustic** phon. + **auditory** phon.
- More on **articulatory** phonetics
 - Observation of speech organs
- More on **general** phonetics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



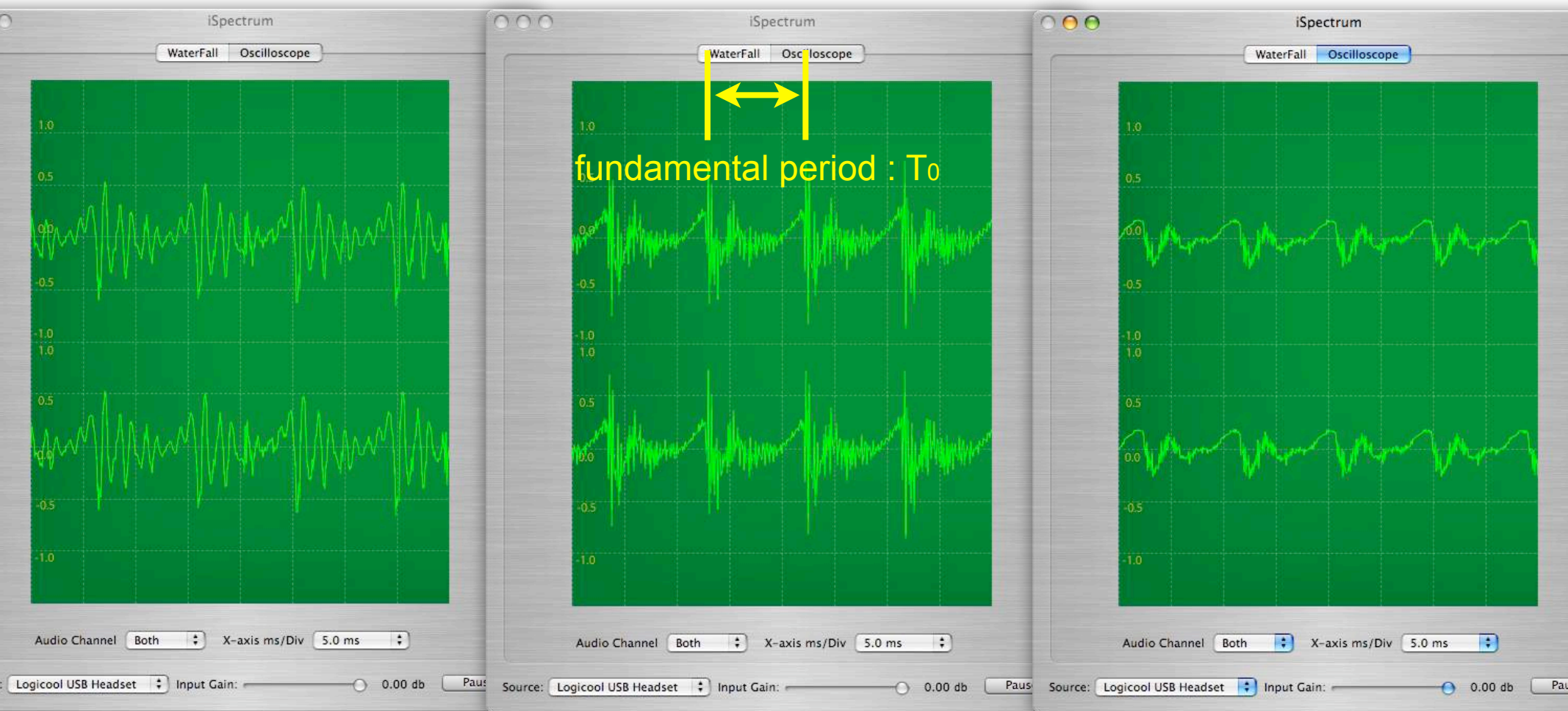
Speech = vibrations of air particles

- 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



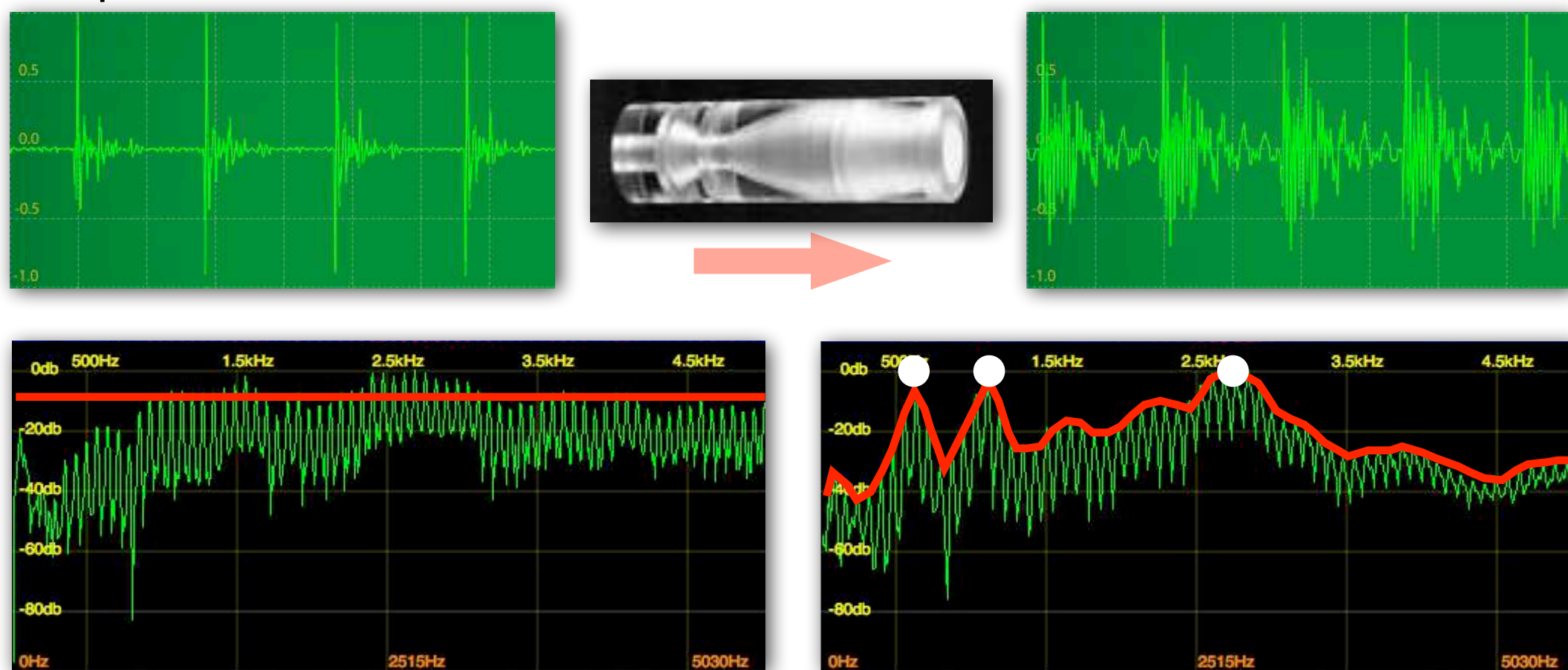
Speech = vibrations of air particles

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



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.

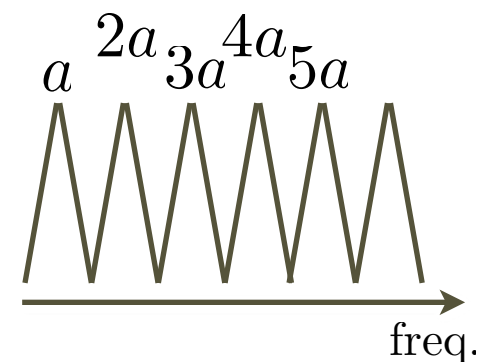
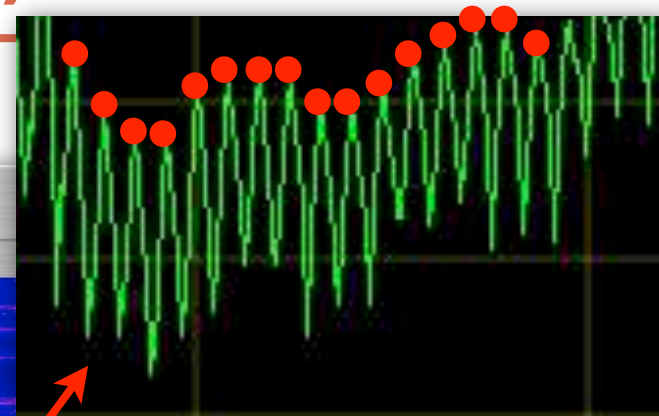
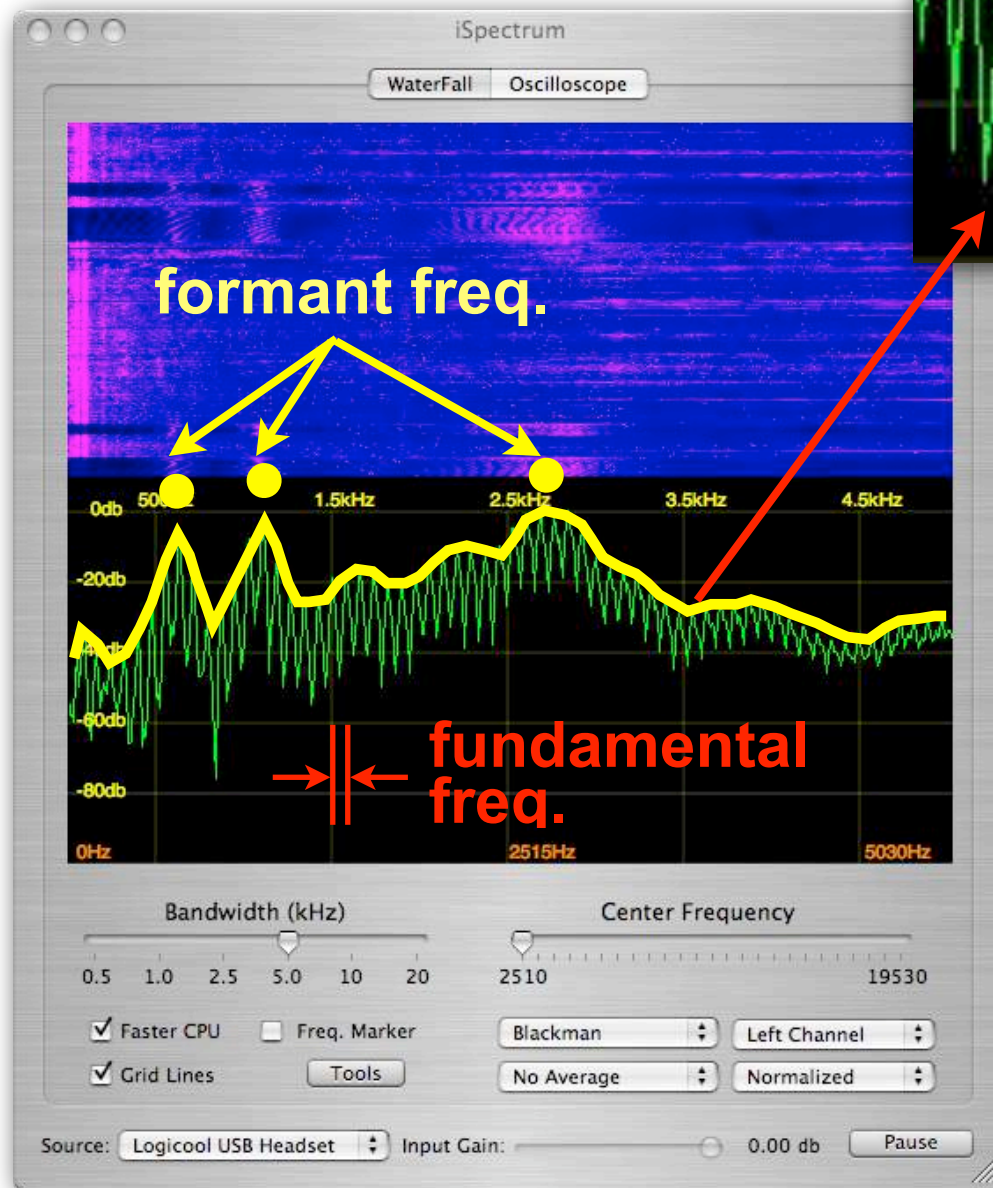
Timbre = energy distribution pattern over the frequency axis

Fundamental frequency (F0) and timbre

- F0 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

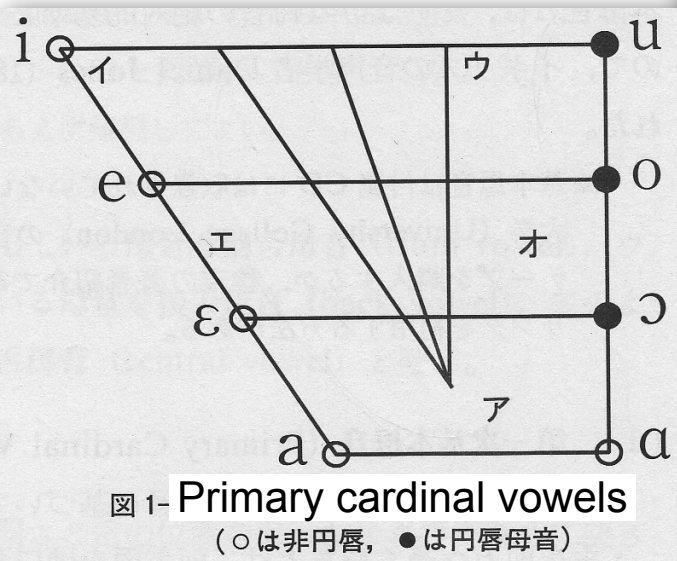
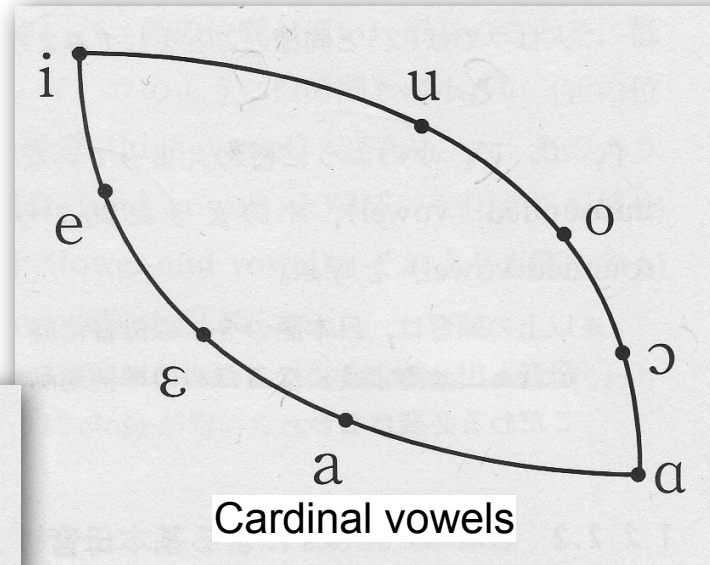


厳密には「音高 = a 」であって、ピークの間隔ではない。調波構造が無くても音高は感覚できる。

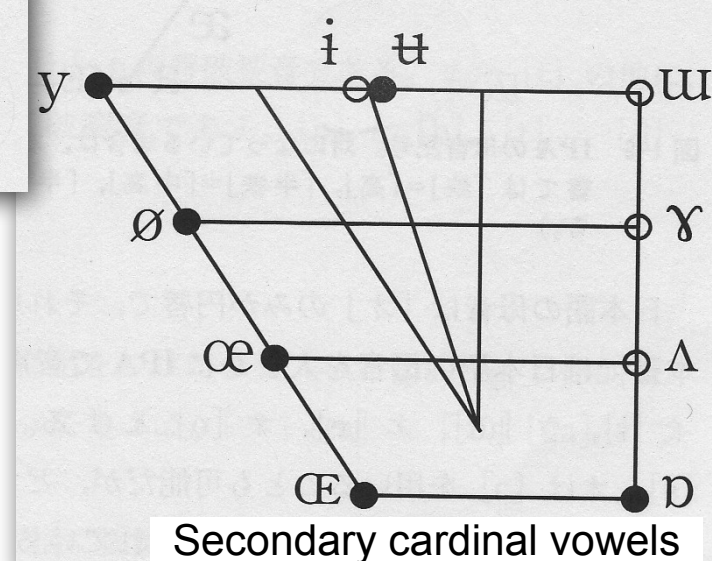


General phonetics

- 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.



●: rounding
○: unrounding



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

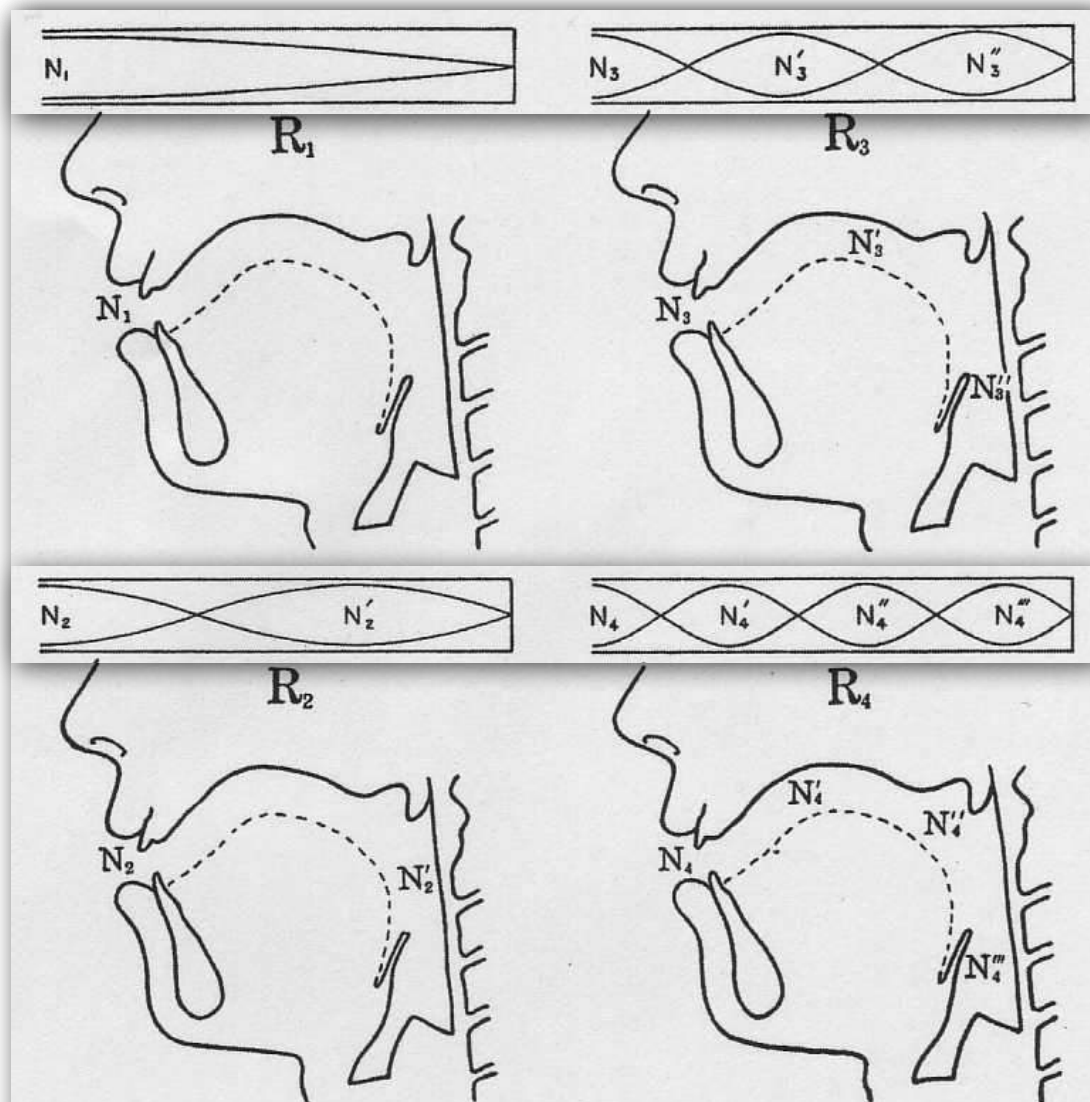
CONSONANTS (PULMONIC)

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

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

Acoustic phonetics

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



standing
waves

formant
frequency

$$F_n = \frac{c}{\underline{4}l} (2\underline{n} + 1)$$

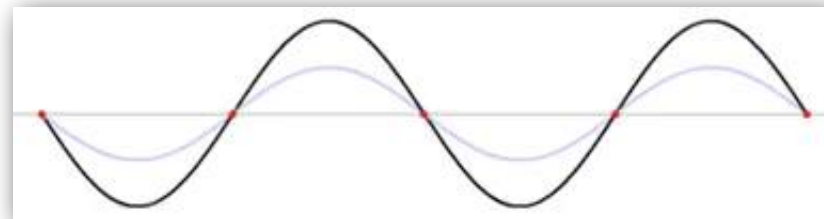
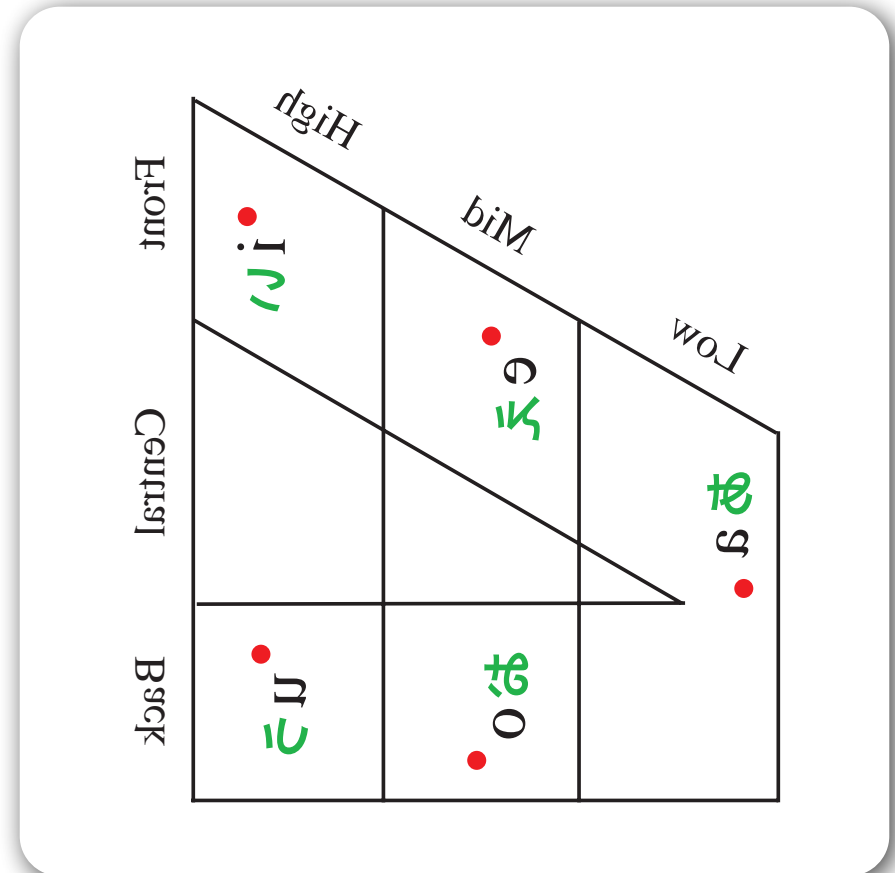
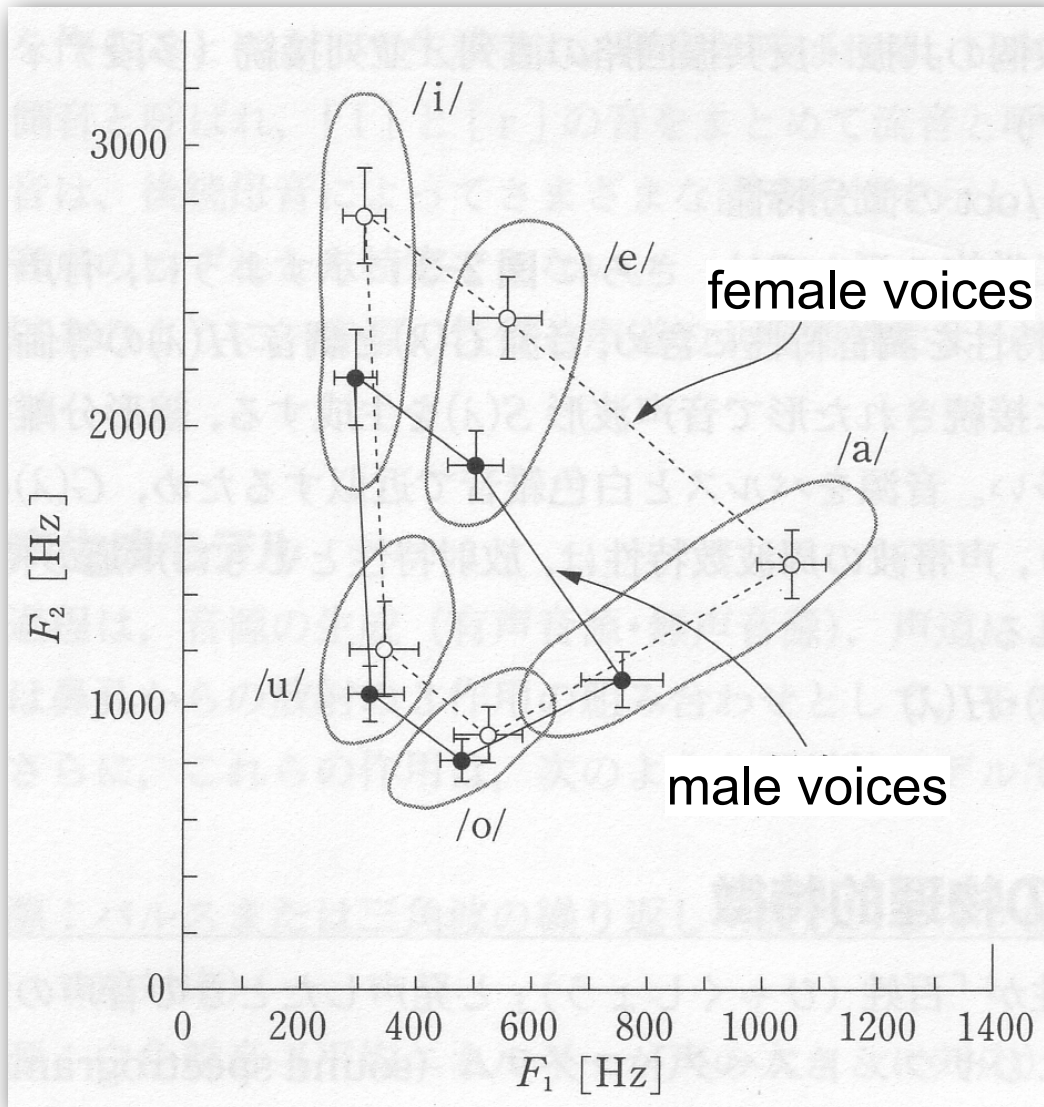


図 93 断面積が均一の音響管または声道における体積流の最大点の分布

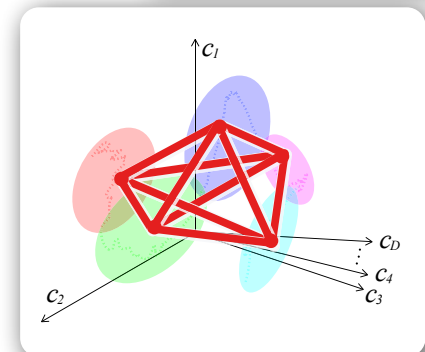
Acoustic and articulatory phonetics

- Shape difference = resonance frequency difference

- /a/ and /i/ /a/ and /a/



Title of each lecture

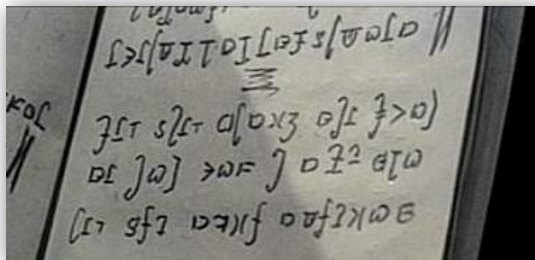


- Theme-1
 - ~~Multimedia information and humans~~
 - ~~Multimedia information and interaction between humans and machines~~
 - ~~Multimedia information used in expressive and emotional processing~~
 - ~~A wonder of sensation - synesthesia -~~
- Theme-2
 - ~~Speech communication technology - articulatory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

Phones and phonemes

• Phones

- A phone is the minimal unit of speech of any language.
- Phonetic symbols are **language-independent** and used by phoneticians to transcribe speech of any language. Defined by by Int. Phonetic Association.
- Should be used like [a b c d e f g].



Dental	Alveolar	Postalveolar	Retroflex
t	d		t̪ d̪
	n		ɳ

• Phonemes

- A phoneme is the minimal unit of speech of a specific language, perceived by native speakers of that language.
- Phonemic symbols are **language-dependent** and used by ordinary people to transcribe speech of that language. Can be defined by a user.
- Should be used like / a b c d e f g /.

/arajurugeNzituo/ → [ɐrɐjɪrɪgɐ̃ɳd͡zɪtsɪɔʔ]

Phones and phonemes

- Sounds of a class can be perceived as sounds of different classes.
 - と**ん**ぼ, と**ん**ねる, ど**ん**ぐり [m], [n], [ŋ]
 - ラ**い**ト, ラ**い**ト right, light
 - す**い**か, たべま**す**か? [su], [s]
 - For Japanese, they are of one class but for foreigners, they may be of different classes.
 - Japanese may have less capacity of phone discrimination but they may have better capacity of discriminating sounds by their duration.
 - おばさん, おばあさん, おかやま (岡山), おおかやま (大岡山), おおおかやま (大岡山)
- Phonemes and allophones
 - One phoneme is sometimes acoustically realized as different phones.
 - **ん** → [m], [n], [ŋ] [m], [n], and [ŋ] are allophones of **ん**.
- Phones are objective(acoustic) and phonemes are subjective(mental)?
 - #phones >> #phonemes (30 -- 40). Phones are finer units of speech.
 - Definition of phones is still based on listeners' abstraction of speech acoustics.
 - The definition does not discriminate between adults' [a] and kids' [a] although their spectral envelopes are very different between them.

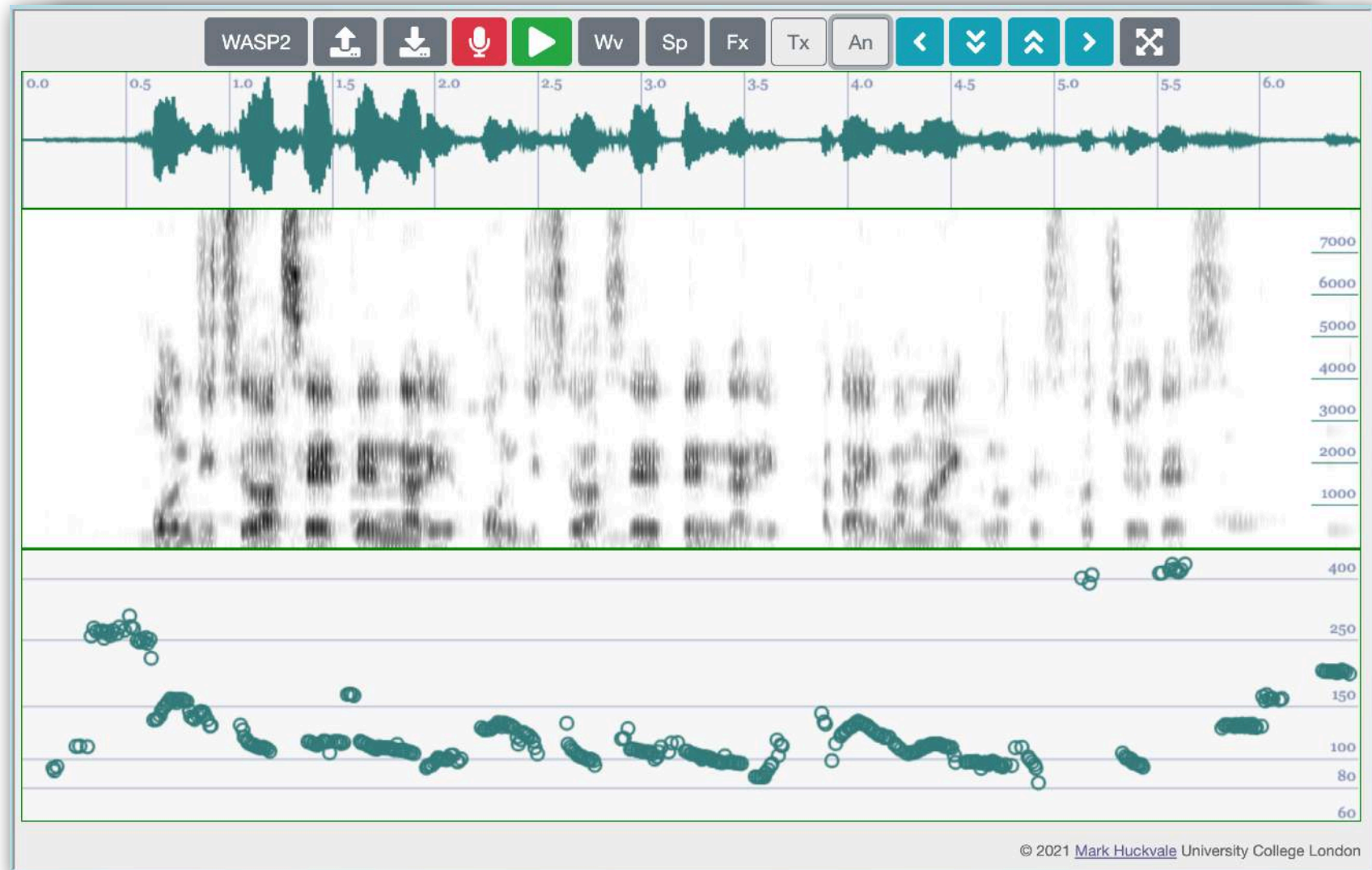
Speech Communication Tech.

- Acoustic analysis of speech -

Nobuaki Minematsu



Web-based speech analyzer



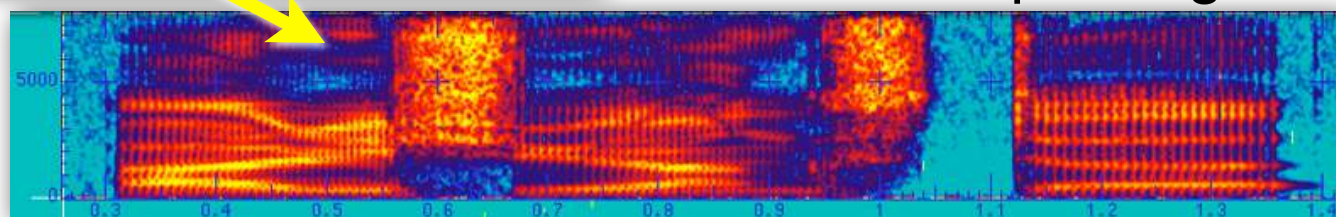
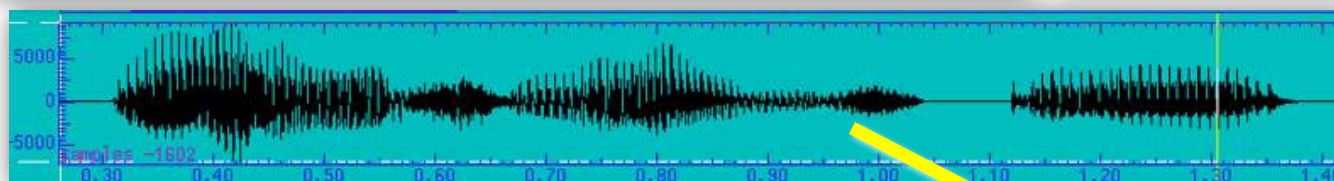
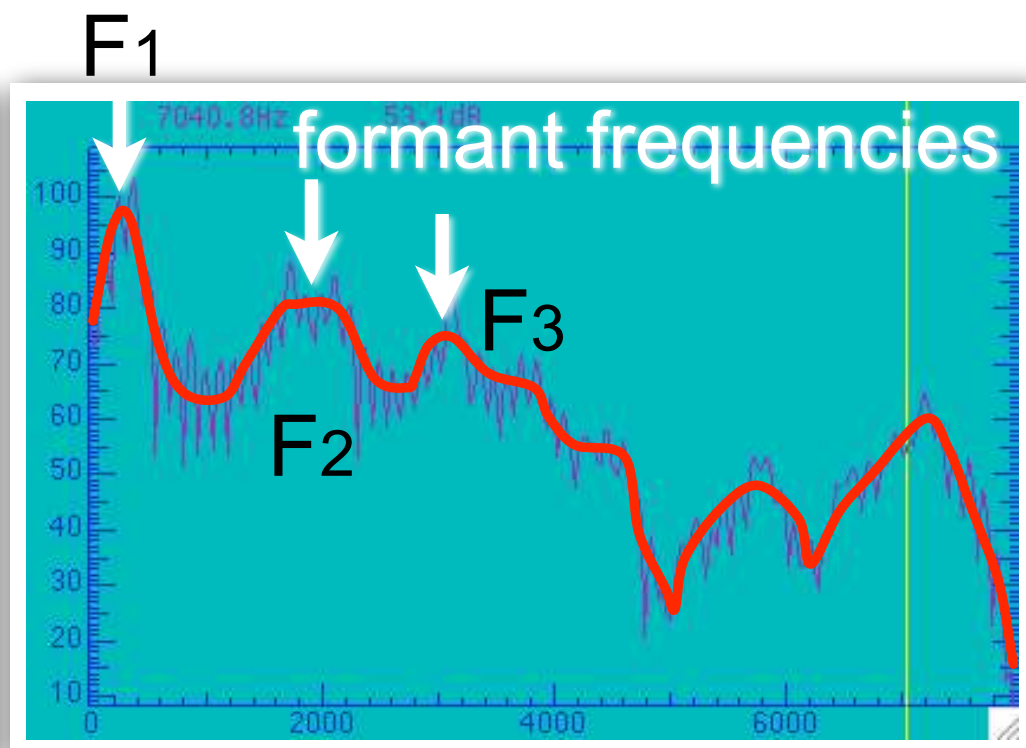
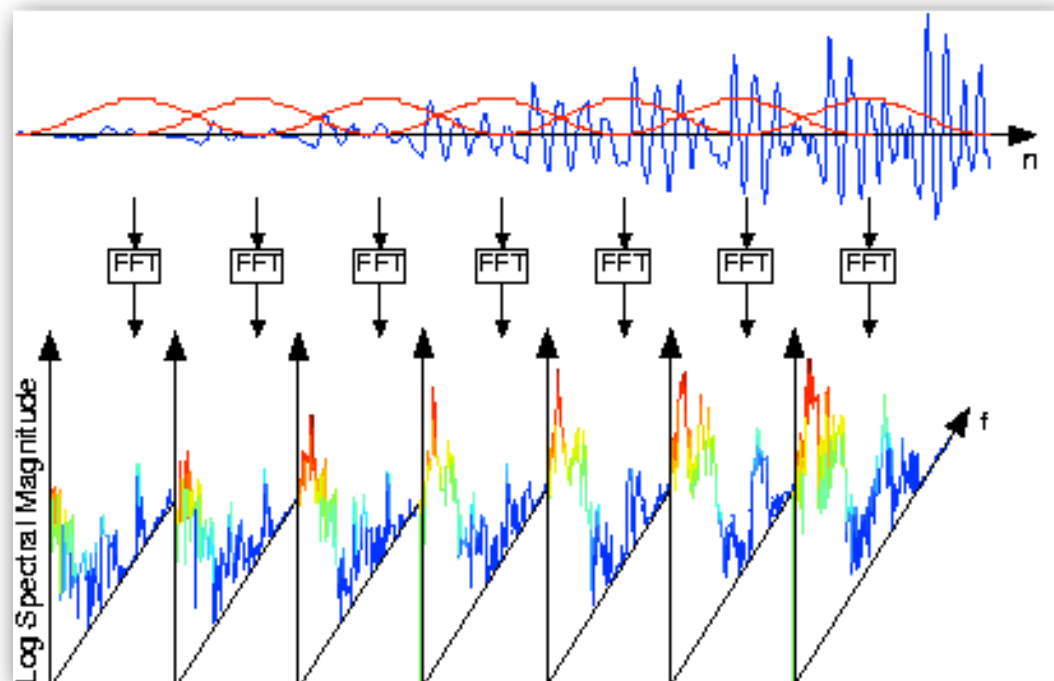
Today's menu

- More on details of acoustic phonetics (continued)
 - Characteristics of human hearing
 - Fundamental frequency and pitch again
 - Fourier analysis of speech signals
 - Simple hearing tests
- Technology for acoustic analysis of speech
 - Source-filter model of speech production $S(\omega) = G(\omega)H(\omega)R(\omega)$
 - Cepstrum method to separate source and filter
 - Advanced analysis tool of STRAIGHT
 - Some morphing examples
 - LPC, PARCOR, and the shape of a vocal tube
- Spectrums/waveforms of various language sounds
 - Vowels, semivowels, liquids, nasals, voiced fricatives, unvoiced fricatives, glottals,
 - Voiced plosives, unvoiced plosives, voiced affricatives, and unvoiced affricatives
 - Speech recognition as spectrum reading
- Summary



Acoustic phonetics

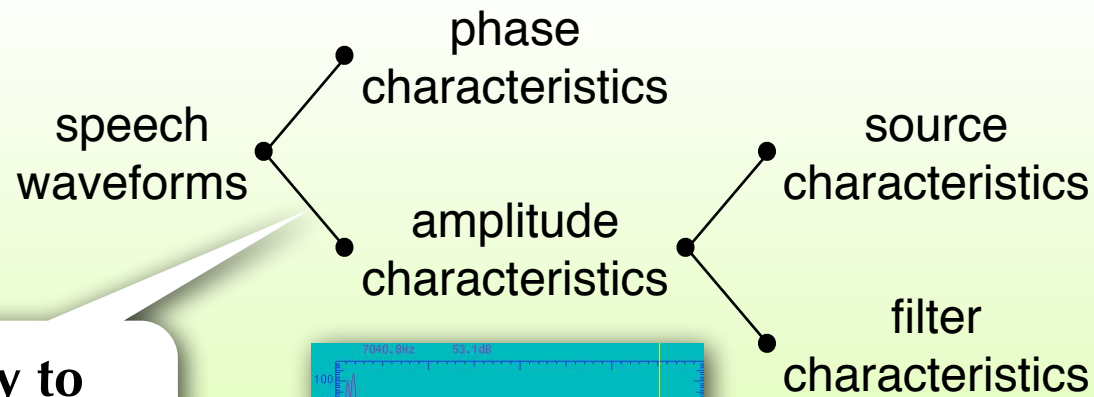
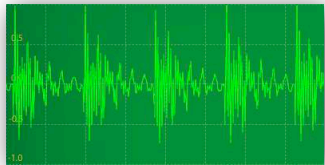
- From waveforms to spectrums
 - Windowing + FFT + log-amplitude



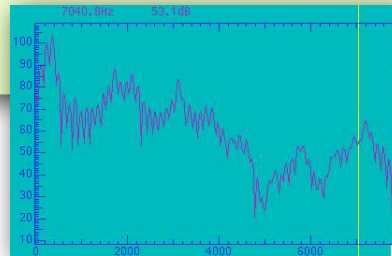
spectrogram

Waveform to spectrum

- From waveforms to spectrums
 - Windowing + FFT + **log-amplitude**
- Insensitivity of human ears to phase characteristics of speech
 - Human ears are basically “**deaf**” to phase differences in speech.
 - It is not impossible for us to discriminate two sounds **acoustically** with different phase characteristics but we don't discriminate them **linguistically**.
 - No language treats those two sounds as two different *phonemes*.

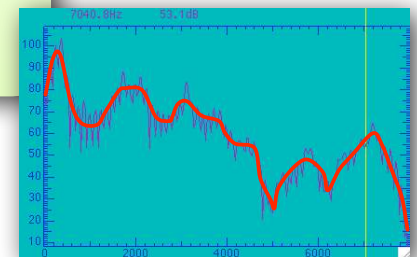
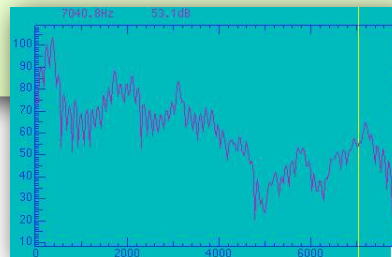
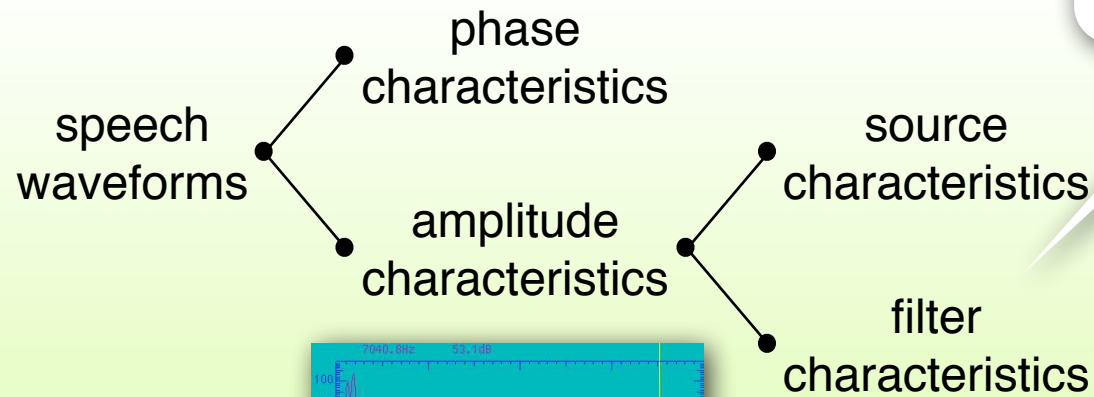
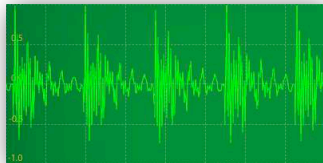


**Insensitivity to
phase differences**



Spectrum to spectrum envelope

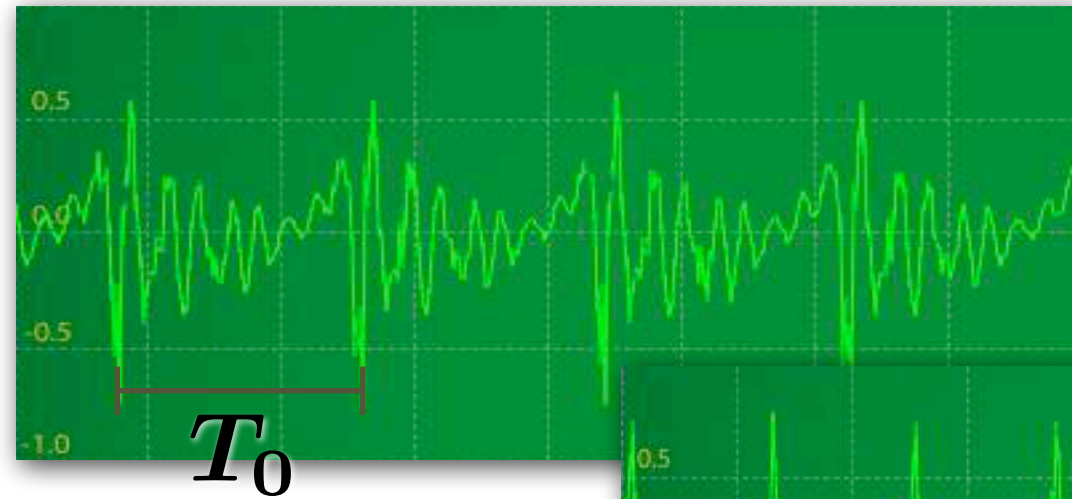
- From spectrums to spectrum envelopes
 - log-amplitude spectrum -> smoothing -> spectrum envelope
- Humans' insensitivity to pitch differences when perceiving phonemes.
 - /a/ with high tone and /a/ with low tone are perceived to be of the same class.
 - Separation of pitch (fundamental frequency) can be done by spectrum smoothing.



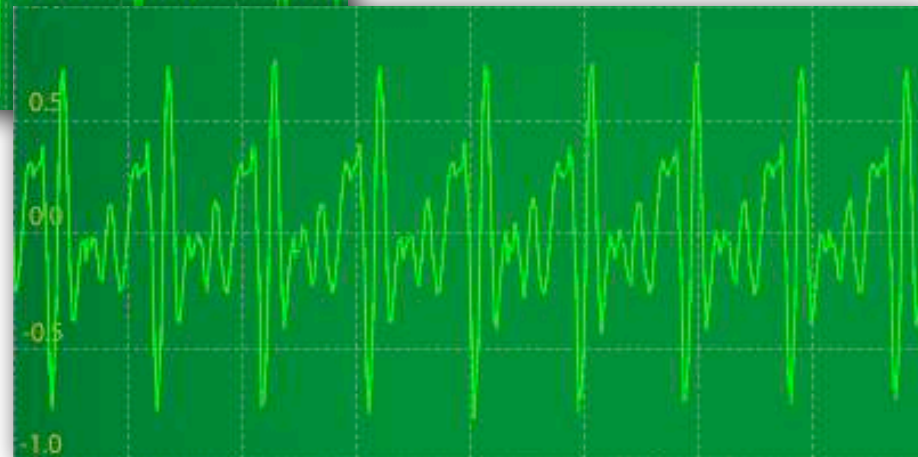
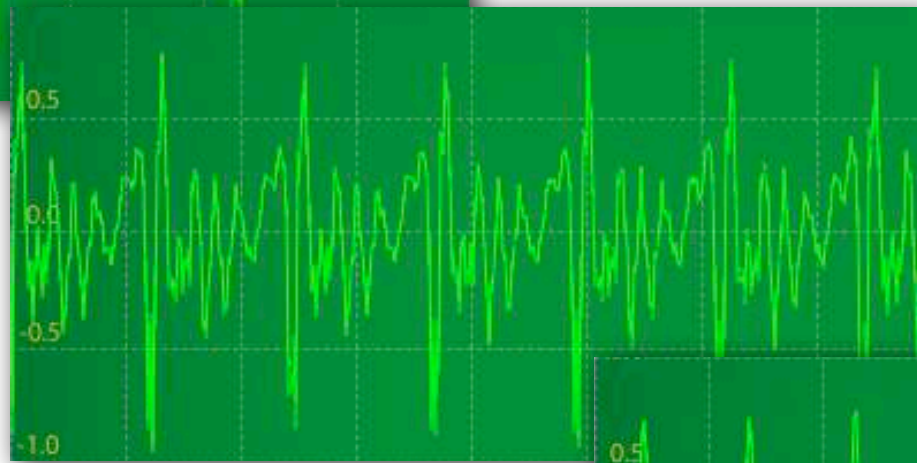
Insensitivity to pitch differences

Speech waveforms and pitch

- /a/ (low) --> /a/ (middle) --> /a/ (high)



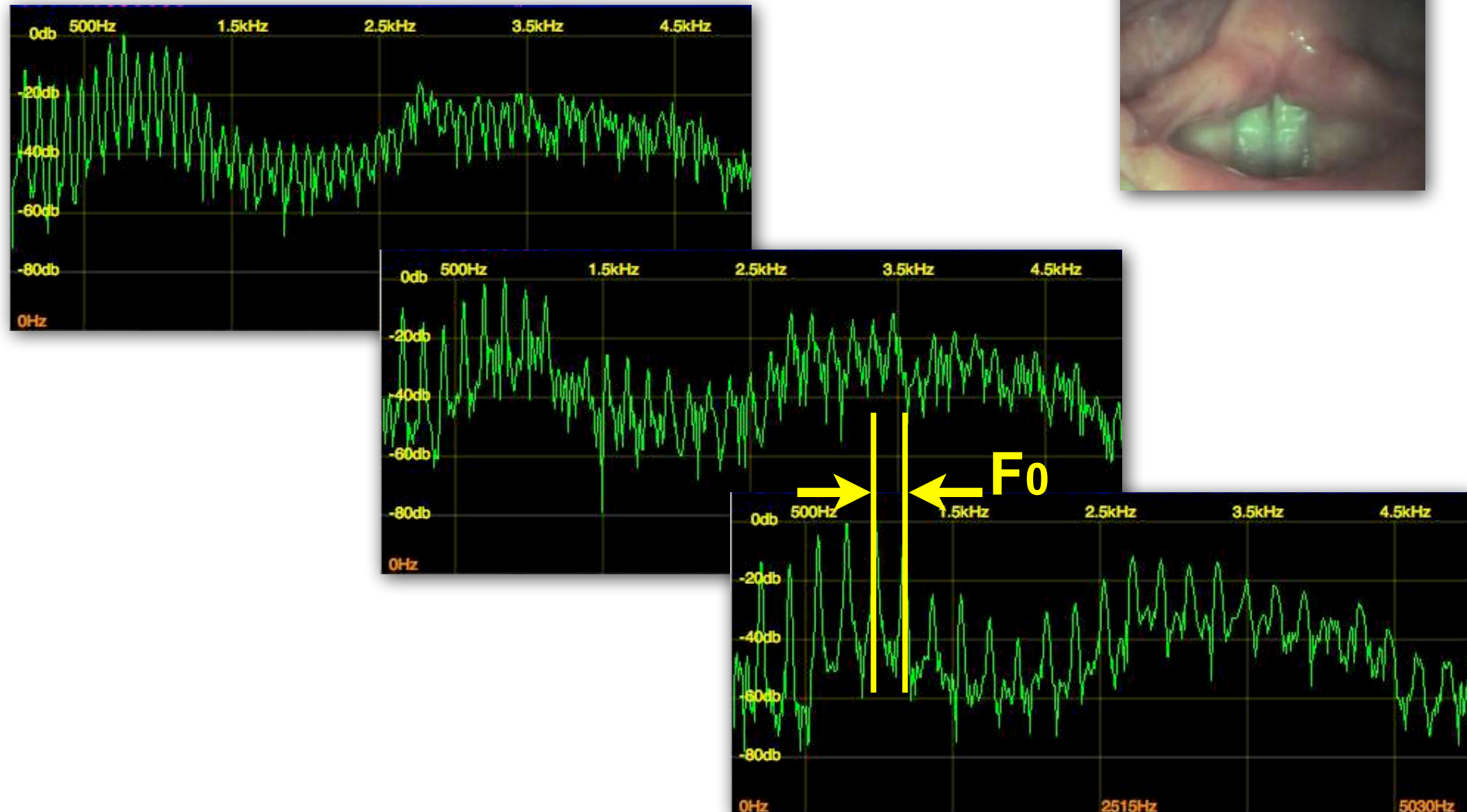
fundamental period



$F_0 = 1/T_0$ fundamental frequency

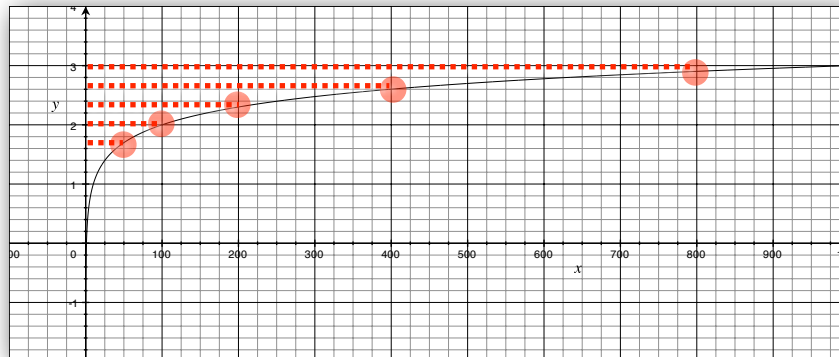
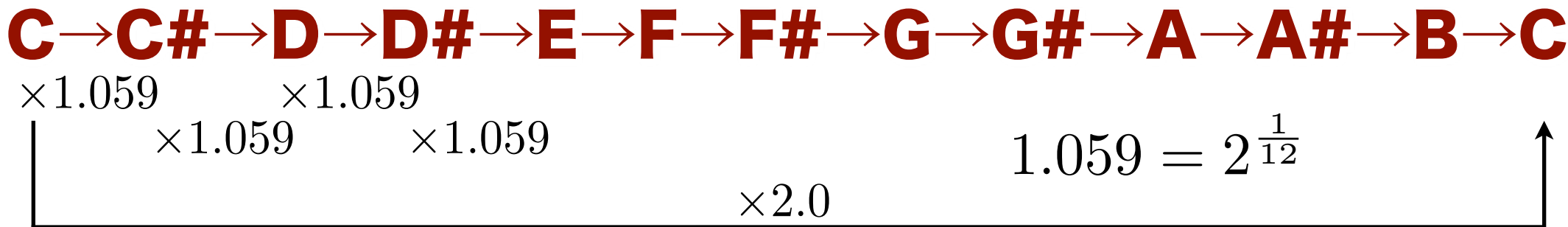
Speech spectrums and pitch

- /a/ (low) --> /a/ (middle) --> /a/ (high)



1 octave = doubled F_0

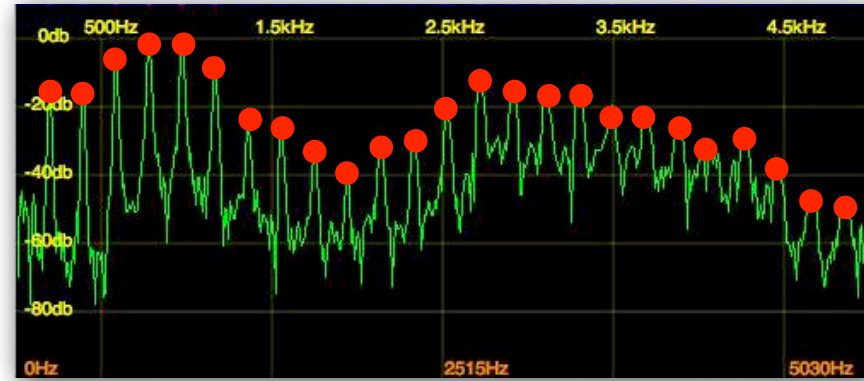
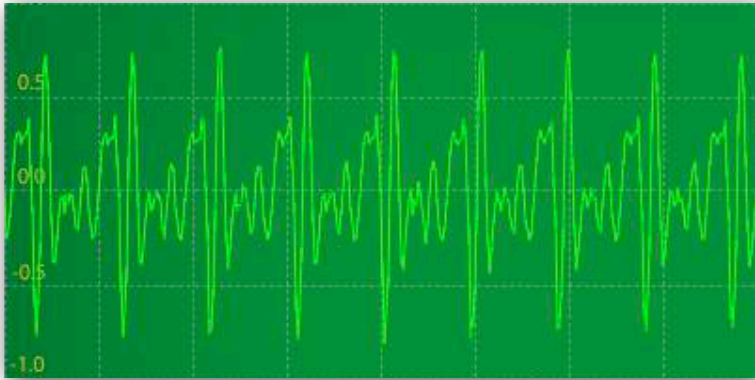
- Mathematical mechanism of music (scale)



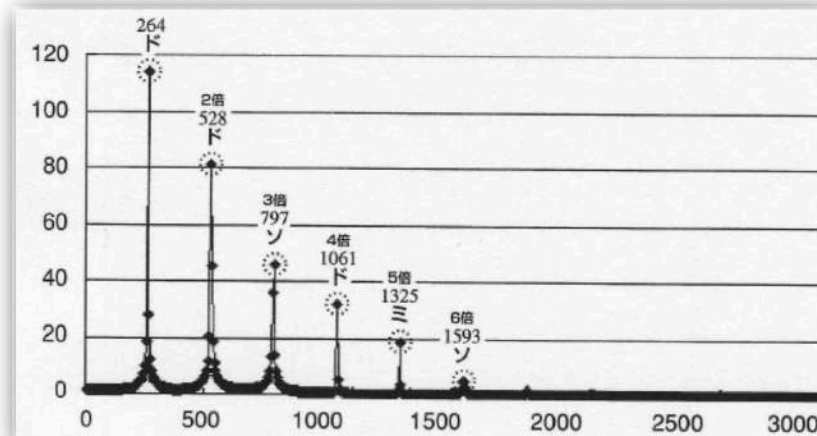
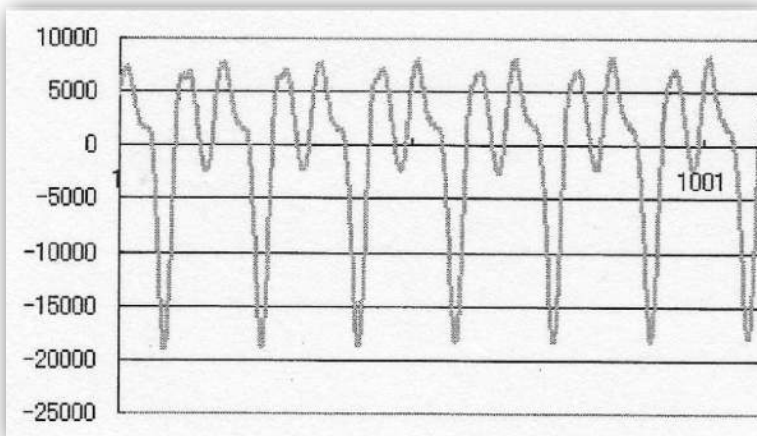
$$y = \log_{10}(x)$$

Harmonic structure

- Speech waveforms and their **log** power spectrum



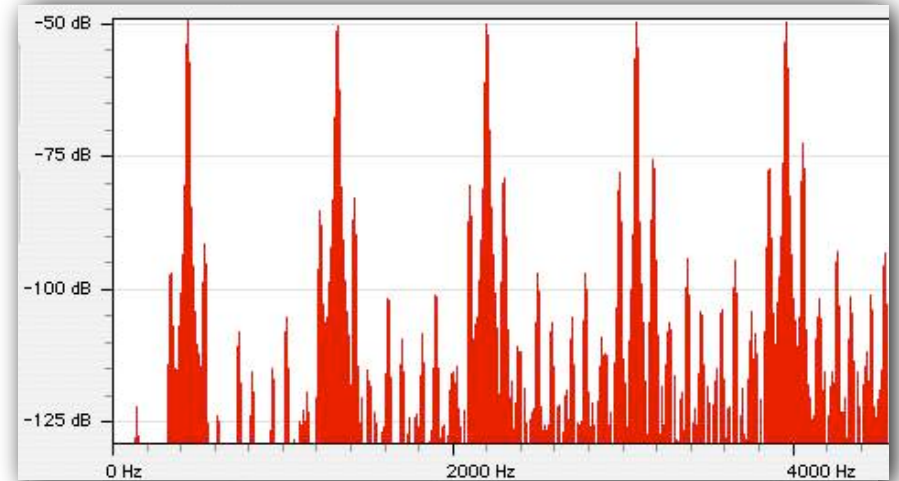
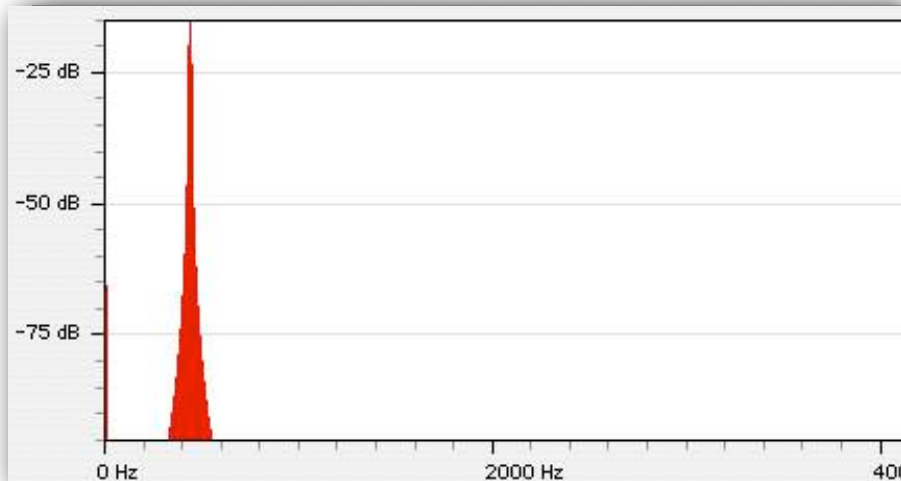
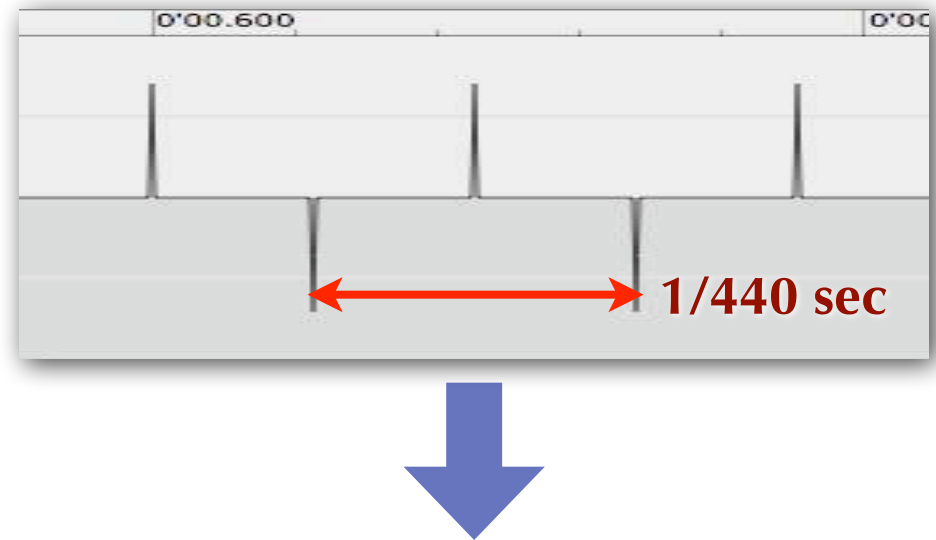
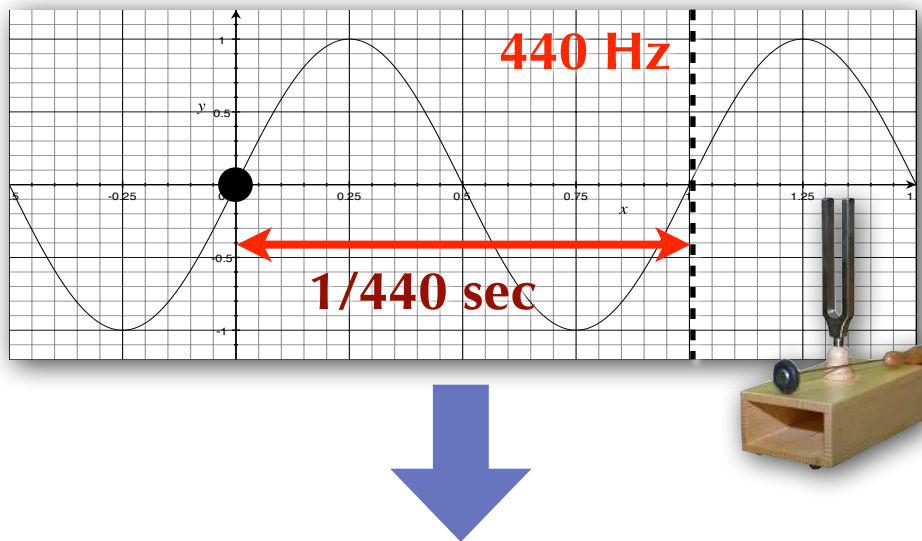
- Guitar sound waveforms and their **linear** power spectrum



- Fundamental tone + 2nd harmonic + 3rd harmonic +
 - Fourier series of periodic signals -> Results (spectrums) also become periodic.

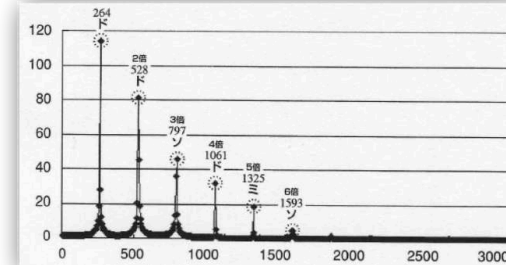
Harmonic structure

- Pure tone and complex tone
 - Waveforms and their linear/log power spectrum



Fourier series and speech production

- Periodic signals are decomposed into \sum sinusoidal waveforms
 - Periodic signals have a set of line spectrums.



- Fourier series of a train of impulses

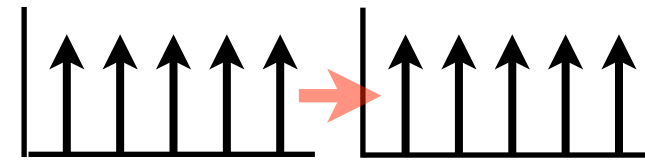
- A train of impulses

$$g(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT_0)$$

- Fourier series of a train of impulses

$$g(t) = \sum_{n=-\infty}^{\infty} \alpha_n e^{jn\omega_0 t}$$

$$\alpha_n = \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} g(t) e^{-jn\omega_0 t} dt = \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} \delta(t) e^{-jn\omega_0 t} dt = \frac{1}{T_0}$$



- Fourier transform of a train of impulses

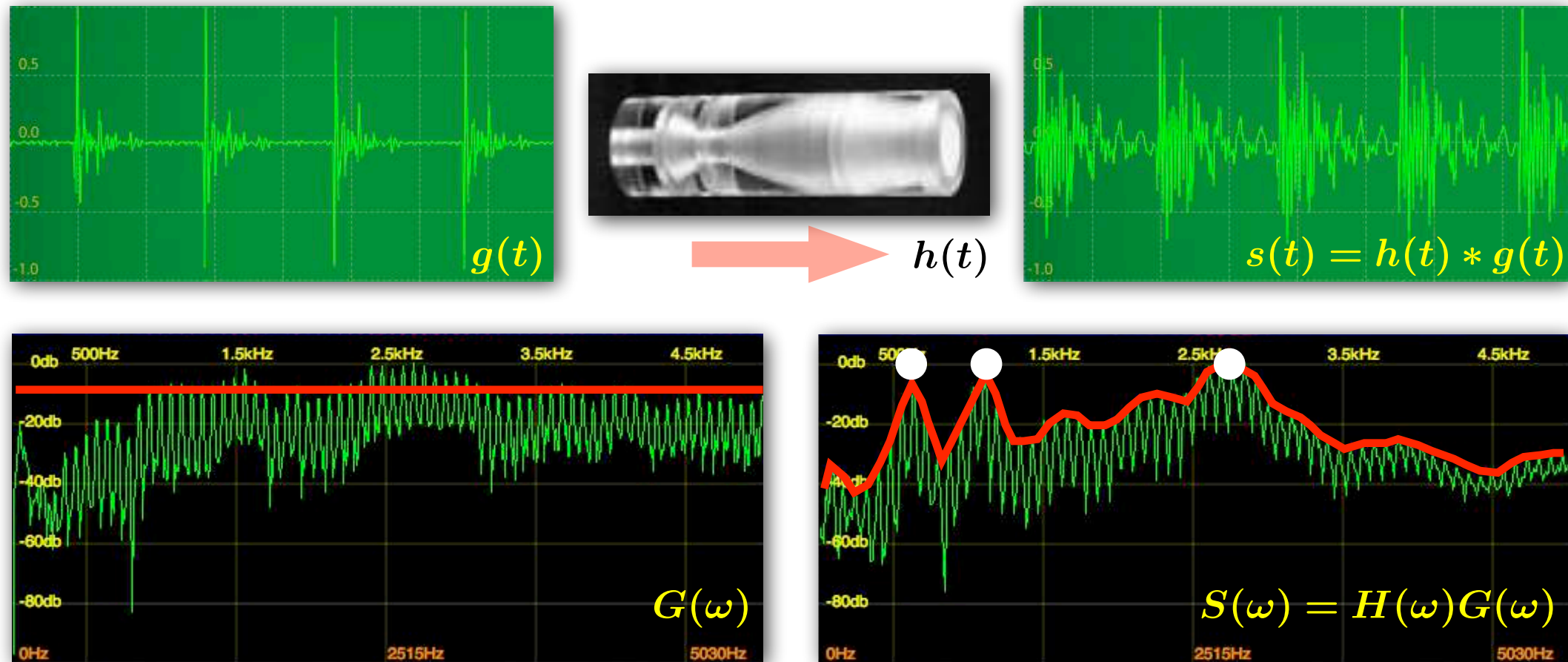
$$G(\omega) = \frac{1}{T_0} \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \{1 \times e^{jn\omega_0 t}\} e^{-j\omega t} dt = \frac{2\pi}{T_0} \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0)$$

- Vowel production as **convolution** of impulse response

- Vocal tract (tube) functions as a filter : impulse response of $h(t)$
- Glottal source waveform : $g(t)$, output waveform : $s(t)$
- $s(t) = h(t) * g(t)$

Acoustic phonetics

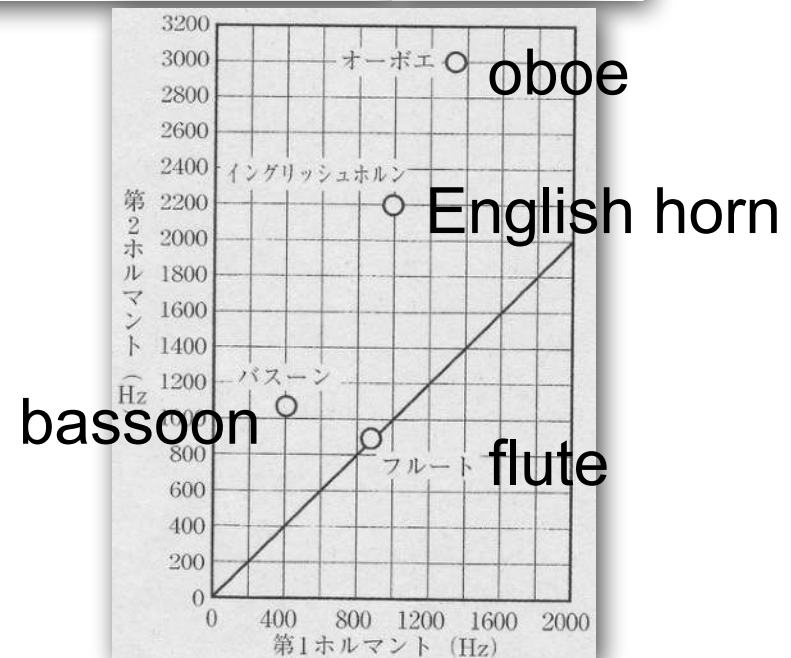
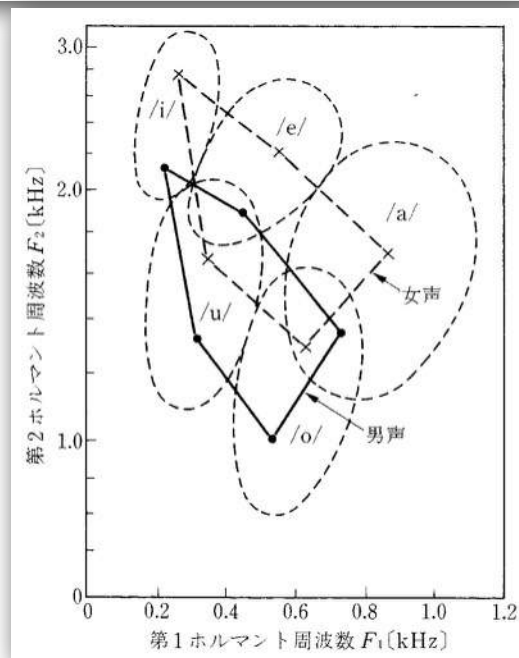
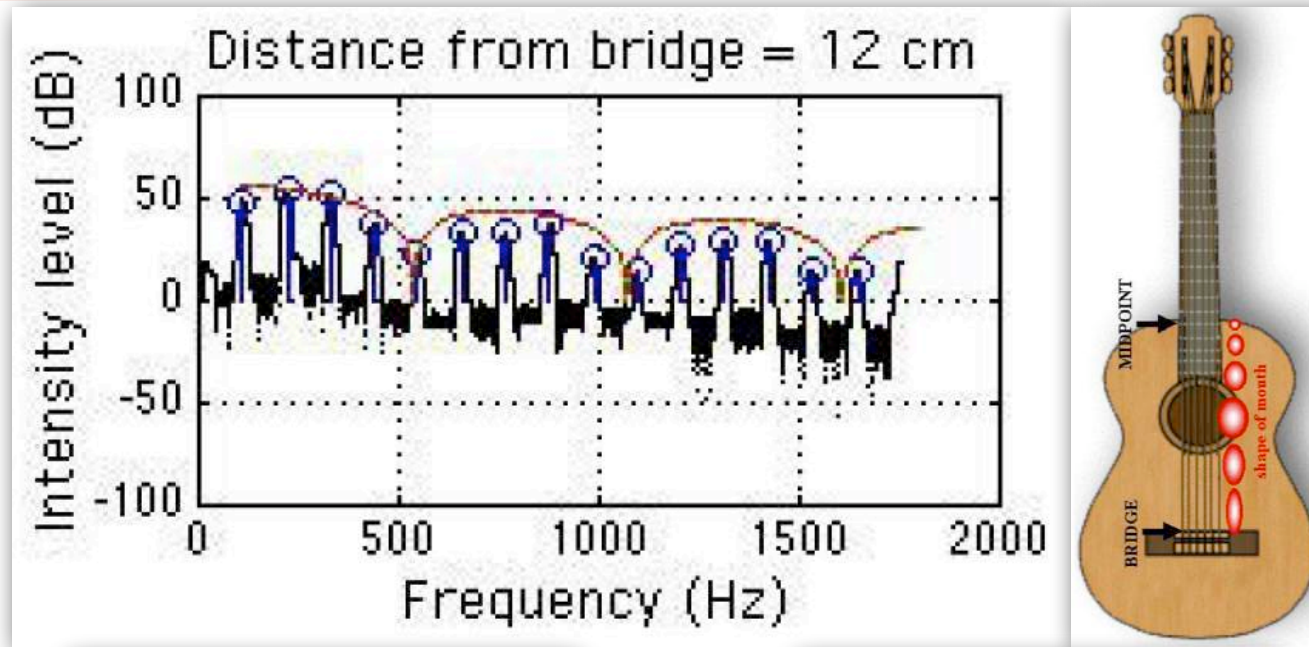
- Spectrum of a vowel sound



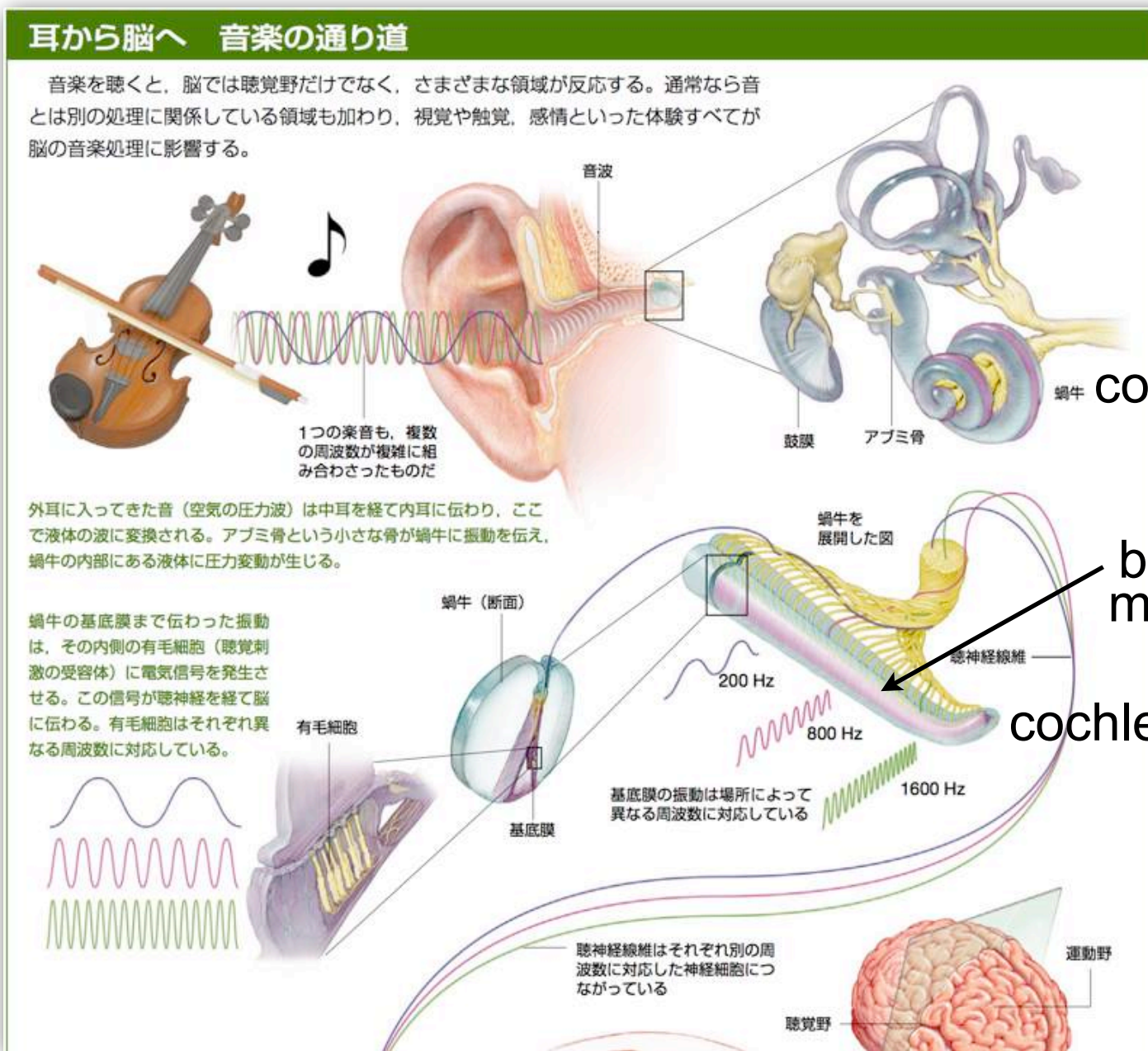
Resonance = concentration of the energy on specific bands that are determined only by the shape of a tube used for sound generation.

Timbre = energy distribution pattern over the frequency axis

Spectrum analysis of guitar sounds



How sounds are processed in the ears



Ear = Fourier analyzer

- Frequency of activation of each band of the cochlea

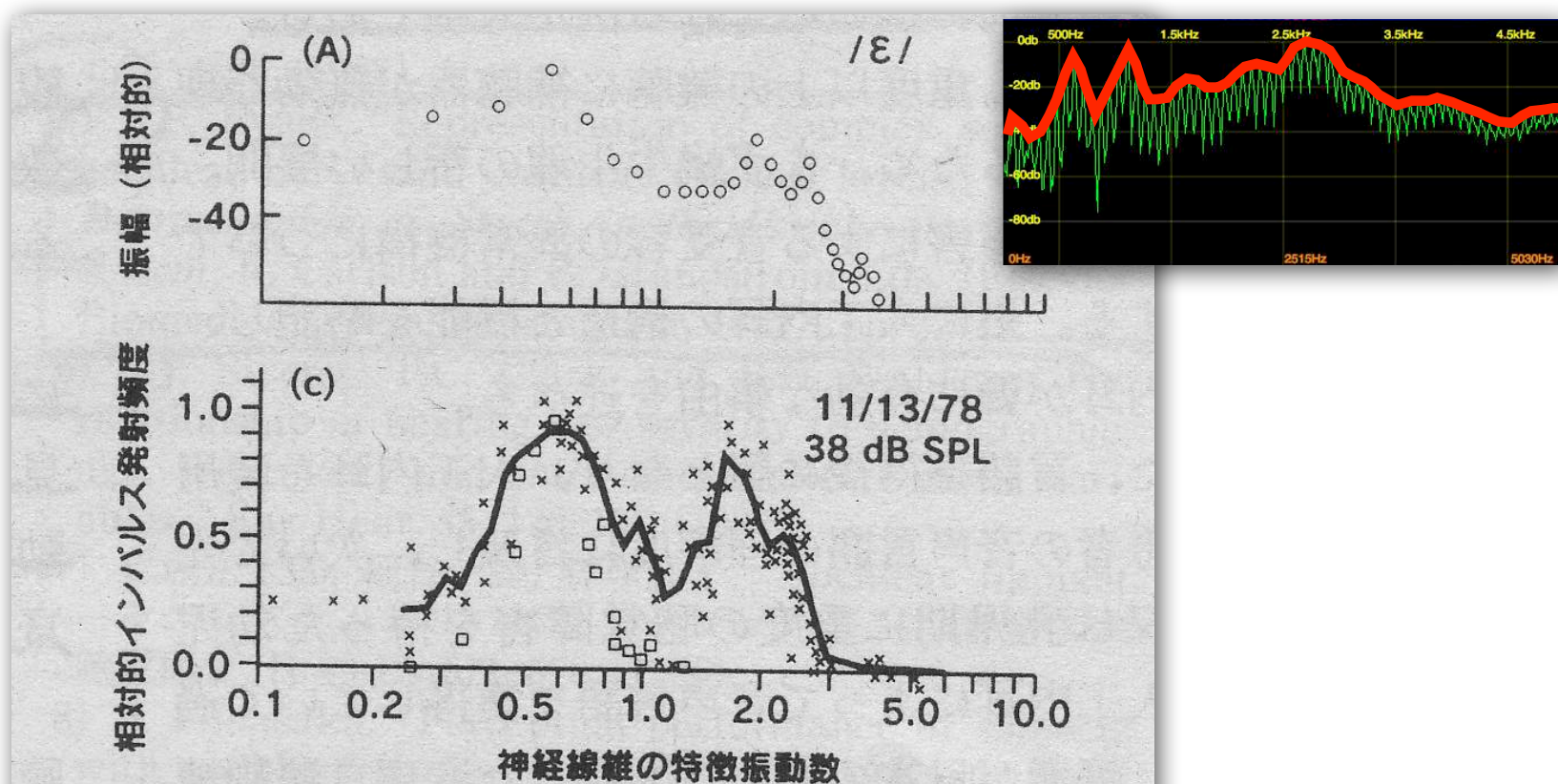
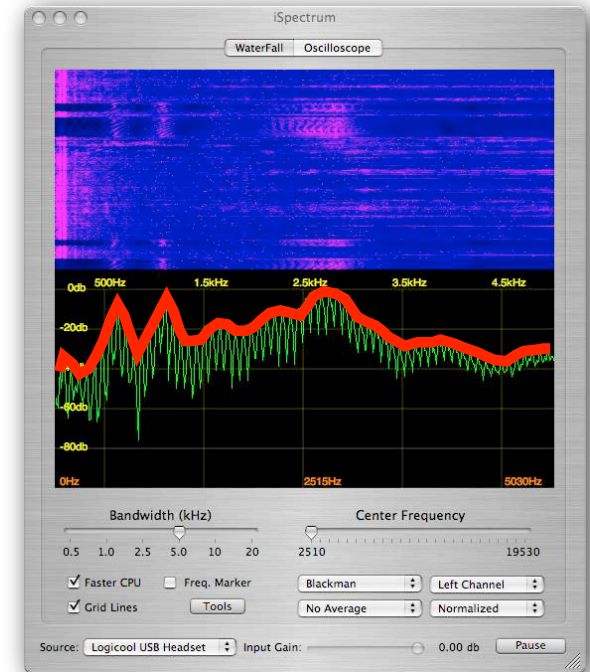


図-3 合成母音/ε/に対する聴神経の反応

正常蝸牛は時々刻々と入力音のフーリエ解析をしている。この機能は蝸牛障害で失われる。A：/ε/の振幅スペクトル，C：様々な特徴振動数の神経繊維の/ε/に対する相対的インパルス発射頻度（文献5）より引用）

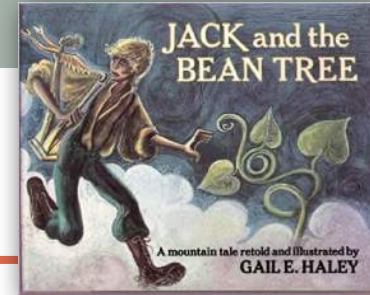
Speech = vibrations of air particles

- 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



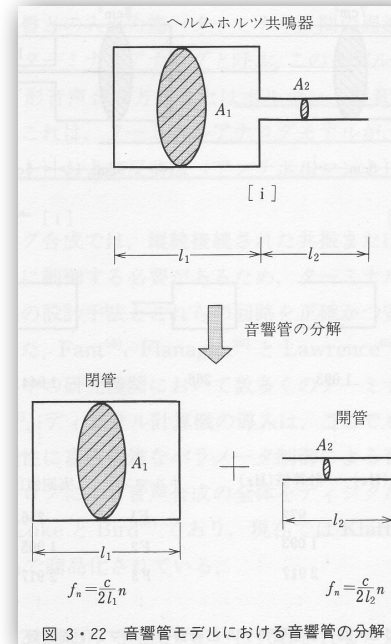
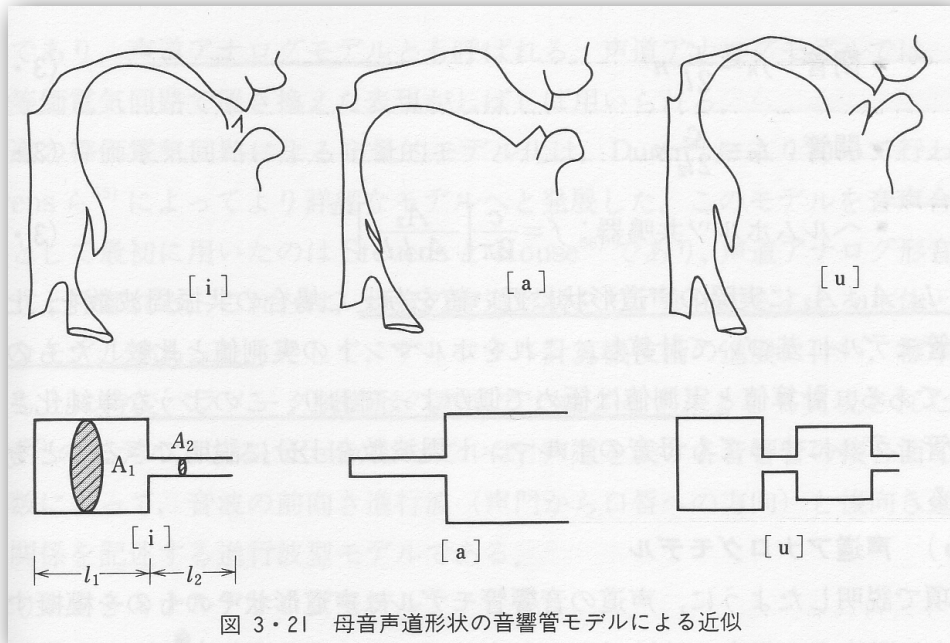
Timbre = energy distribution pattern over the frequency axis

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



Hearing test 1

- Q : what kind of *acoustic* change is expected when a vocal tube becomes shorter?



$$f_n = \frac{c}{2\underline{l_1}} n$$

$$f_n = \frac{c}{2\underline{l_2}} n$$

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

A → B → C

Hearing test 2

- Q : Guess what kind of *tube shape* change happened by hearing the sounds before and after the tube shape change.

A

=



A



B

B

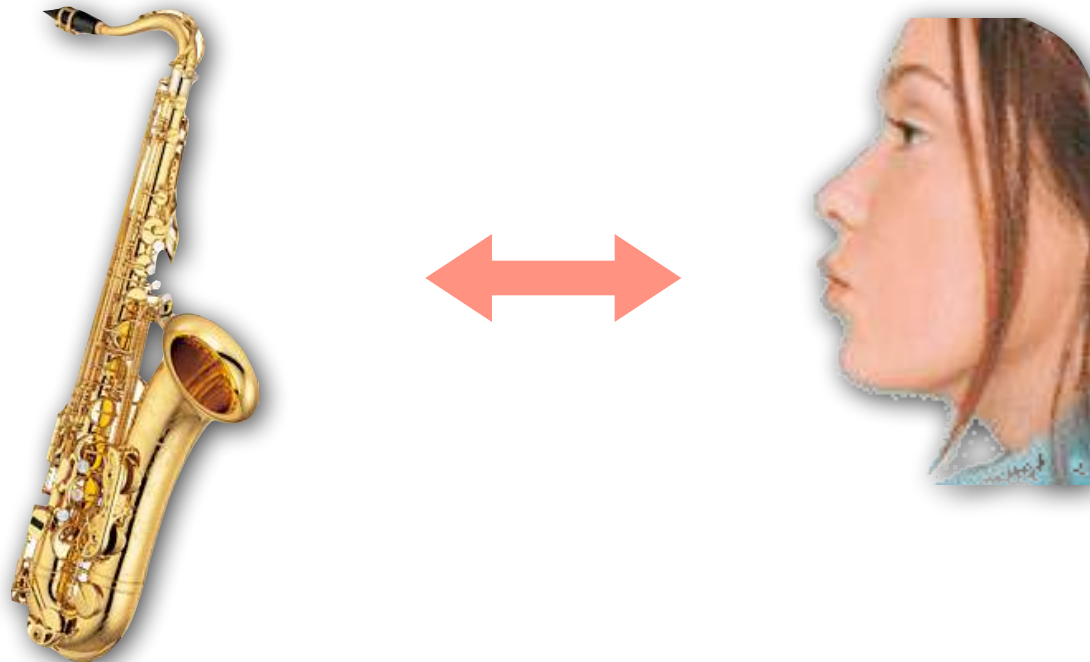


A

Replay

Hearing test 2

- Q : Guess what kind of tube shape change happened by hearing the sounds before and after the tube shape change.



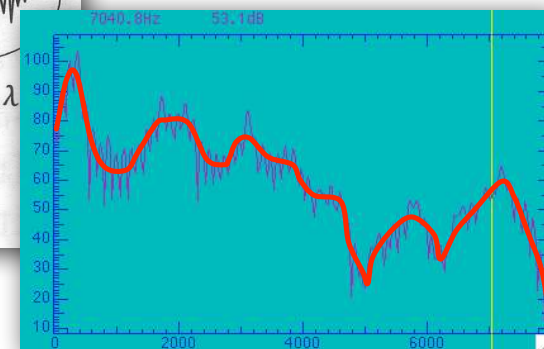
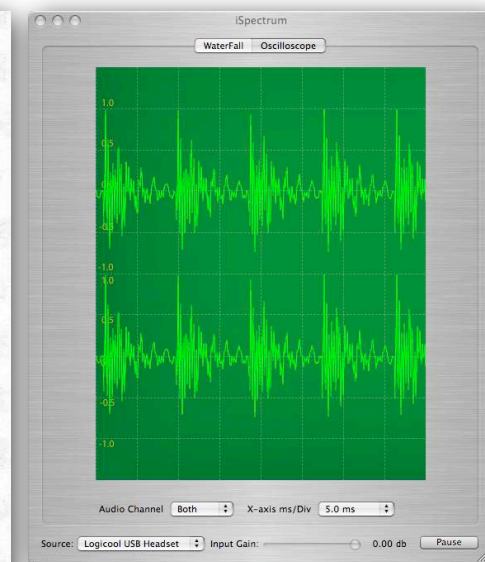
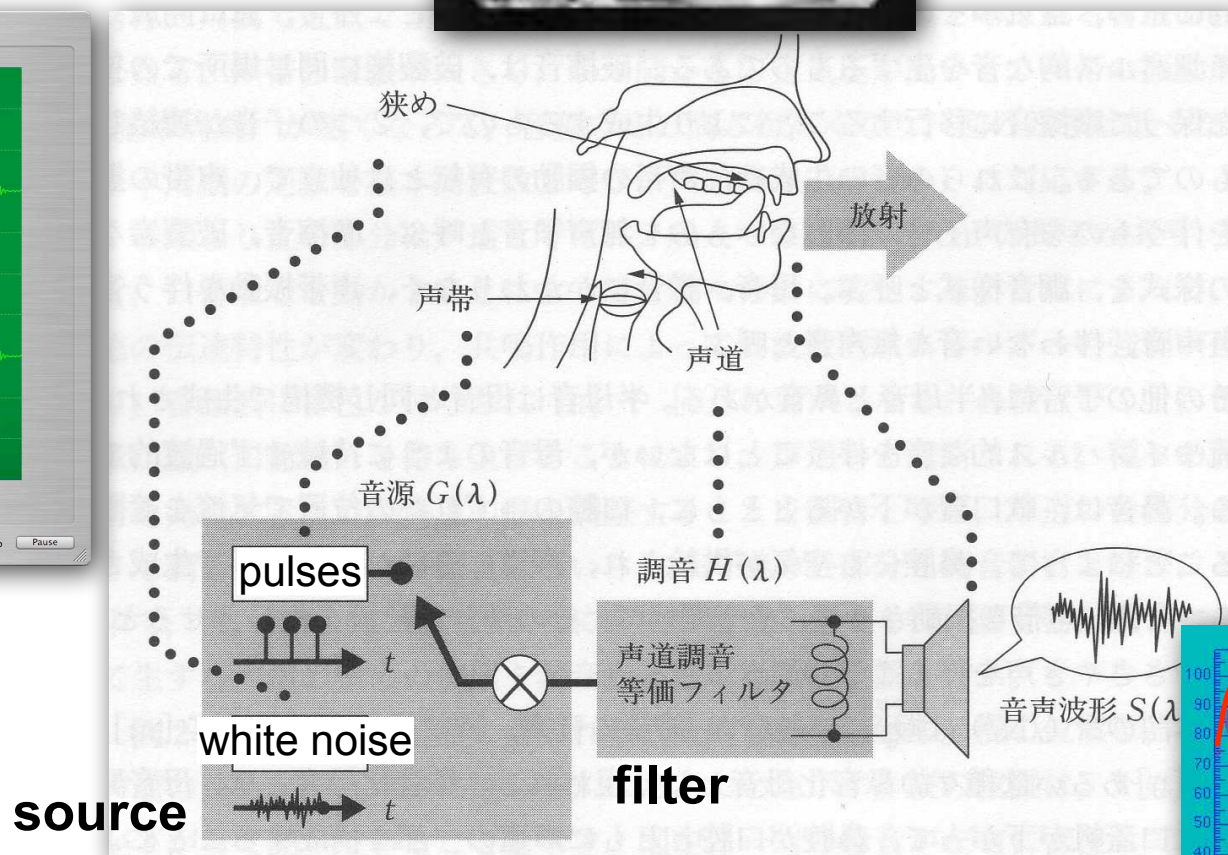
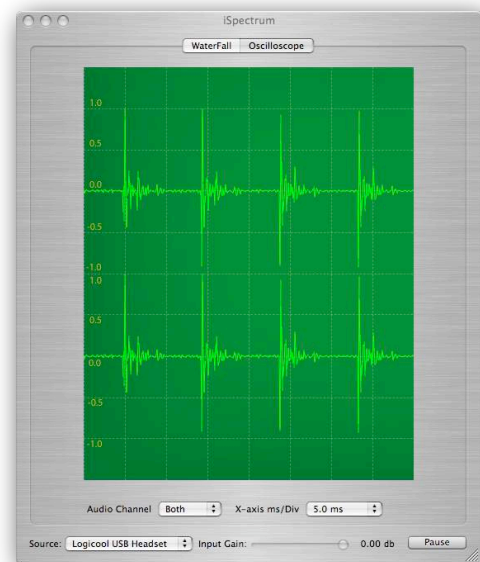
Today's menu

- More on details of acoustic phonetics (continued)
 - Characteristics of human hearing
 - Fundamental frequency and pitch again
 - Fourier analysis of speech signals
 - Simple hearing tests
- Technology for acoustic analysis of speech
 - Source-filter model of speech production $S(\omega) = G(\omega)H(\omega)R(\omega)$
 - Cepstrum method to separate source and filter
 - Advanced analysis tool of STRAIGHT
 - Some morphing examples
 - LPC, PARCOR, and the shape of a vocal tube
- Spectrums/waveforms of various language sounds
 - Vowels, semivowels, liquids, nasals, voiced fricatives, unvoiced fricatives, glottals,
 - Voiced plosives, unvoiced plosives, voiced affricatives, and unvoiced affricatives
 - Speech recognition as spectrum reading
- Summary



Modeling of speech production

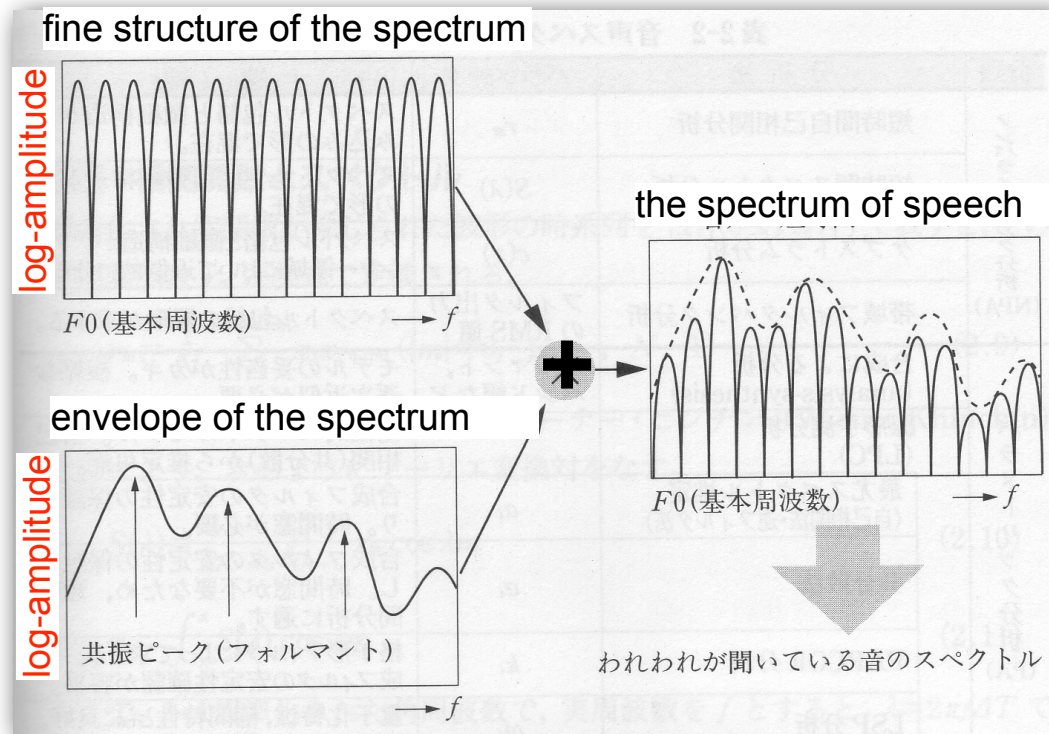
- Mathematical modeling of speech production -- source & filter model --
 - Linear independence between source and filter



$$S(\omega) = G(\omega)H(\omega)R(\omega)$$

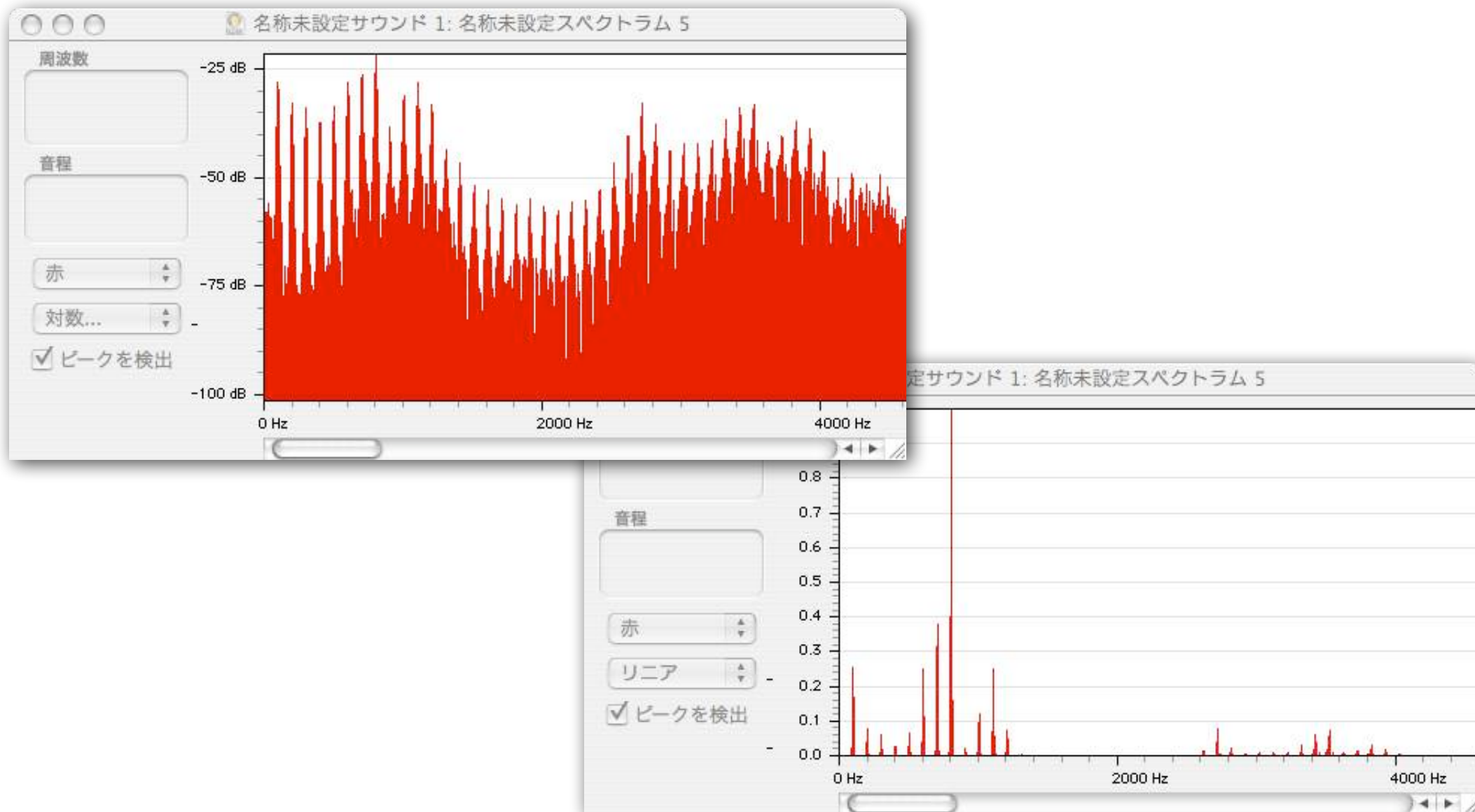
Modeling of vowel production

- Mathematical modeling of speech production -- source & filter model --
 - Separation between the spectrums of source and filter



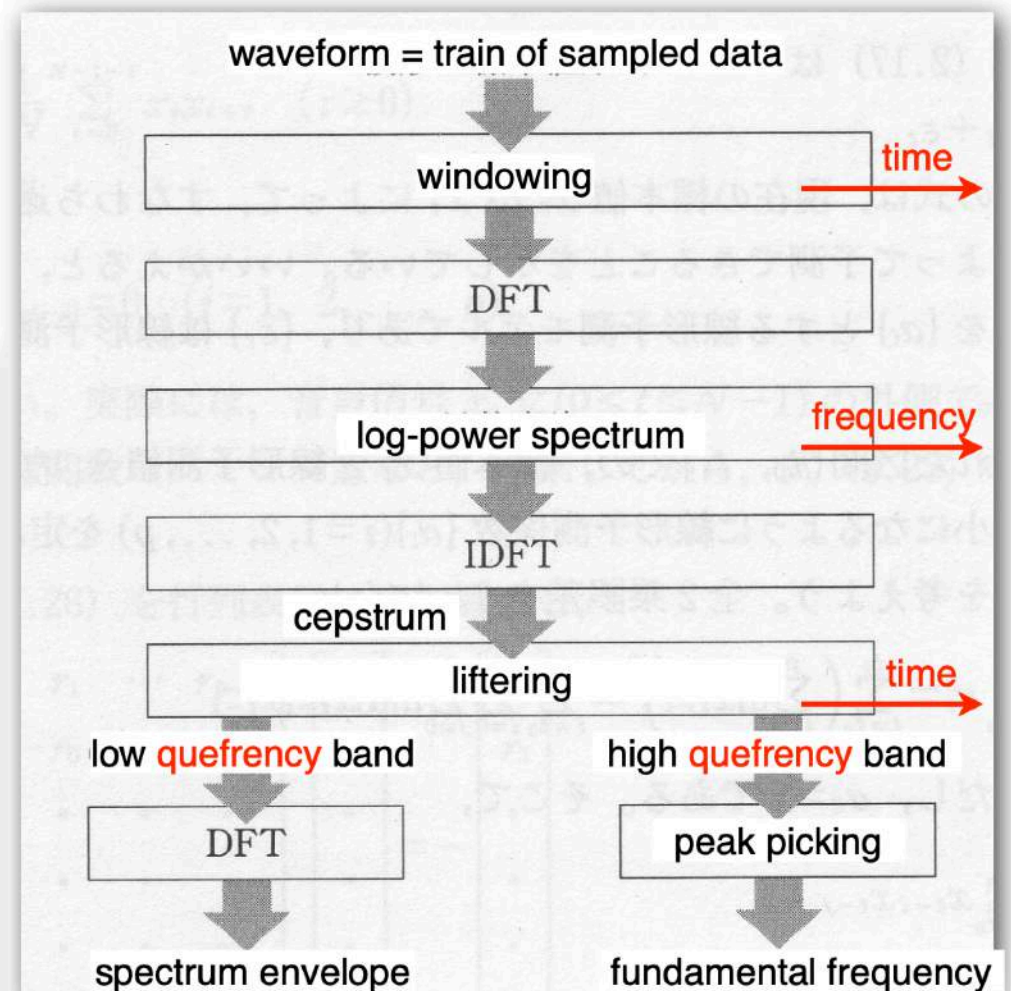
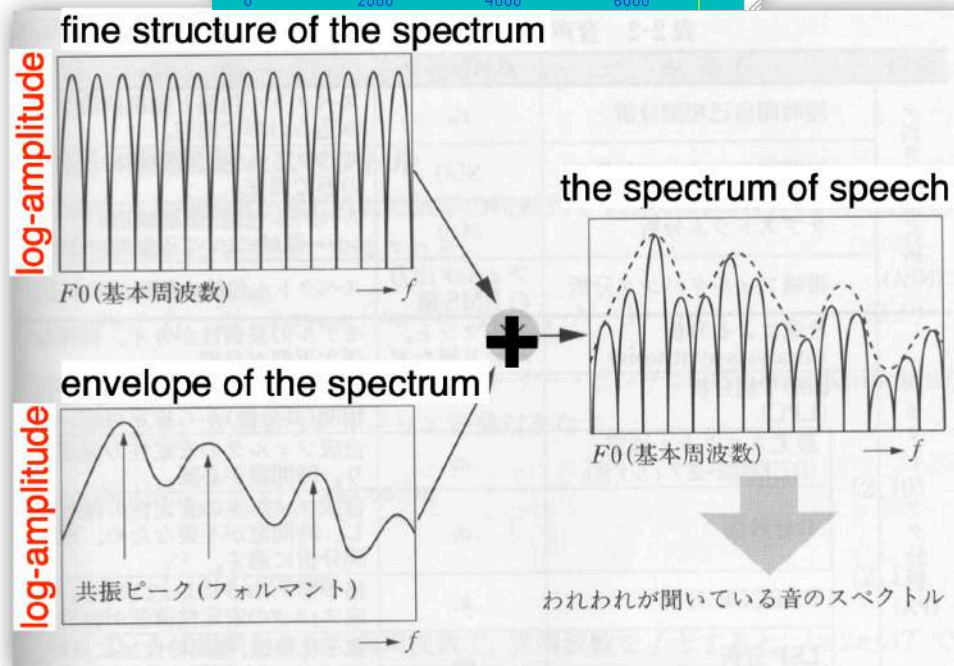
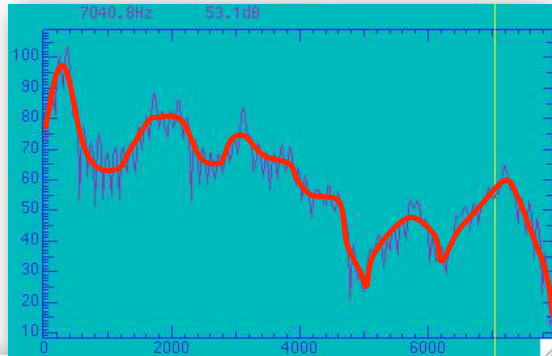
A comment on linear and log spectrums

- Their appearances are very different.



Extraction of spectrum envelopes

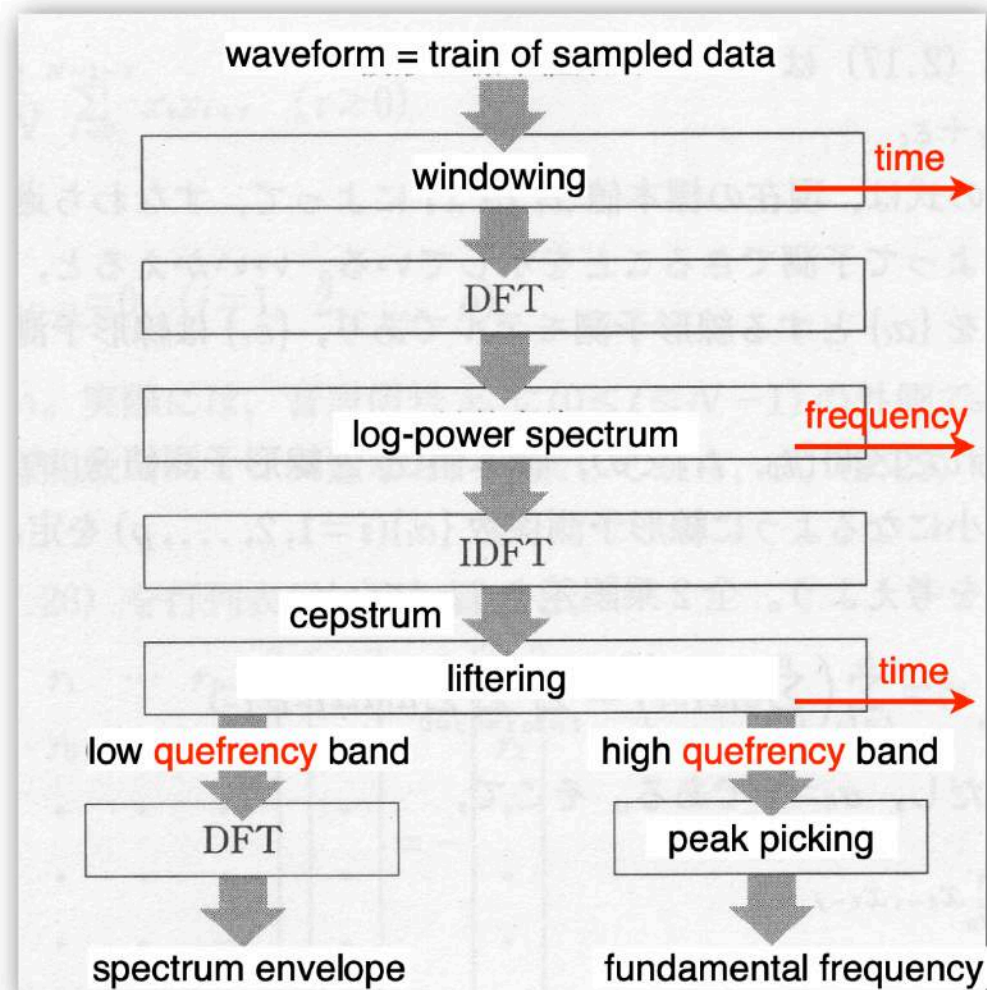
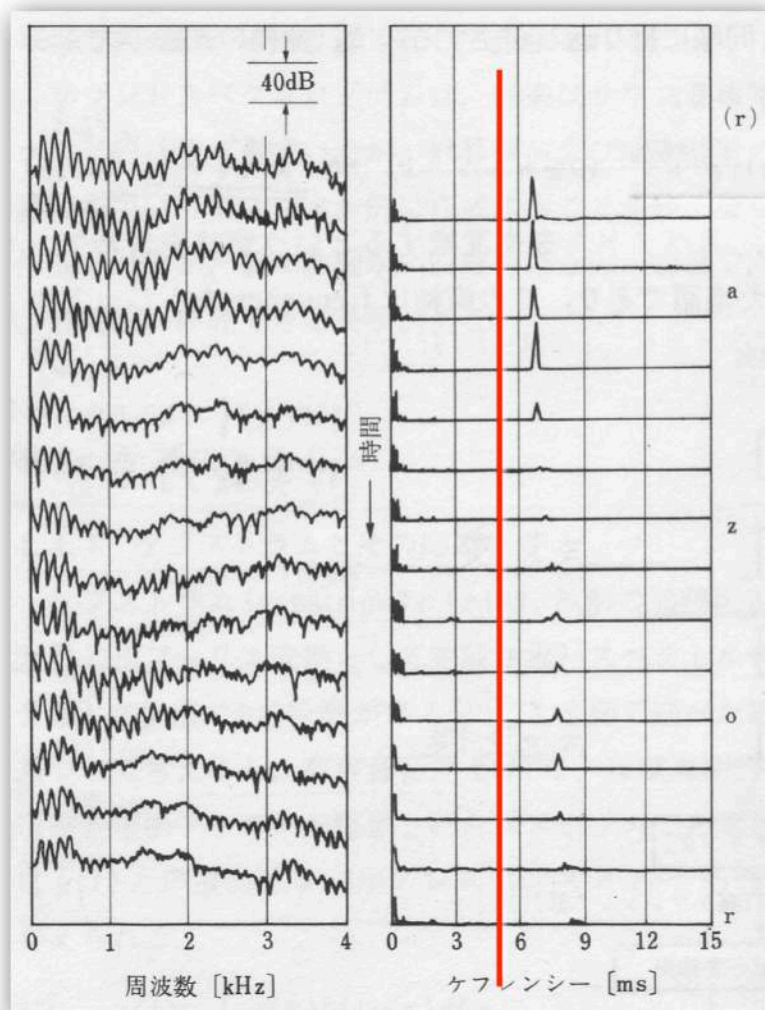
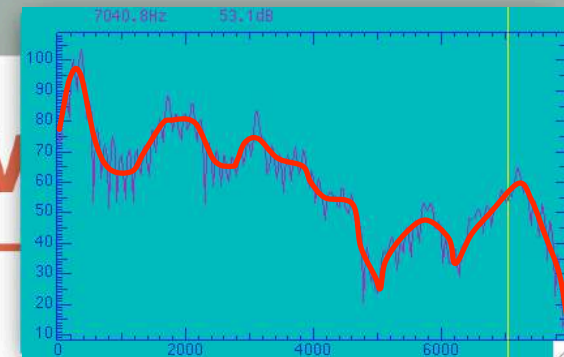
- Cepstrum method
 - Windowing + FFT + log-amplitude --> a spectrum with pitch harmonics
 - Smoothing (LPF) of the fine spectrum into its smoothed version



Extraction of spectrum env

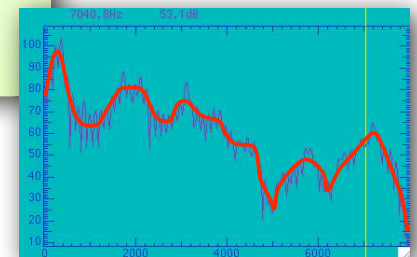
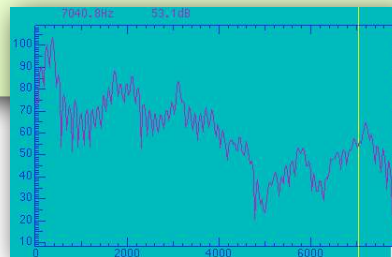
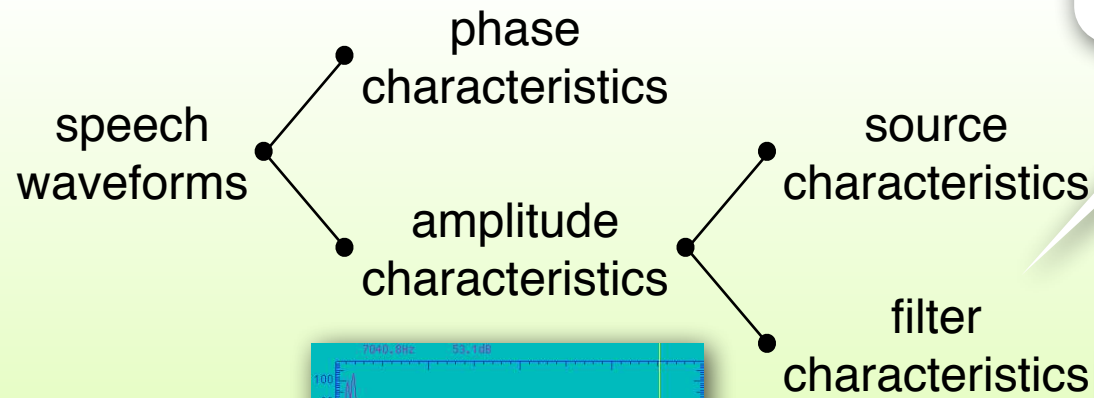
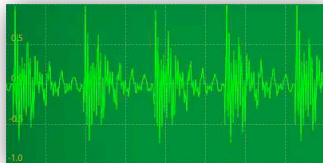
- Cepstrum method

- Windowing + FFT + log-amplitude --> a spectrum with pitch harmonics
- Smoothing (LPF) of the fine spectrum into its smoothed version



Spectrum to spectrum envelope

- From spectrums to spectrum envelopes
 - log-amplitude spectrum -> smoothing -> spectrum envelope
- Humans' insensitivity to pitch differences when perceiving phonemes.
 - /a/ with high tone and /a/ with low tone are perceived to be of the same class.
 - Separation of pitch (fundamental frequency) can be done by spectrum smoothing.



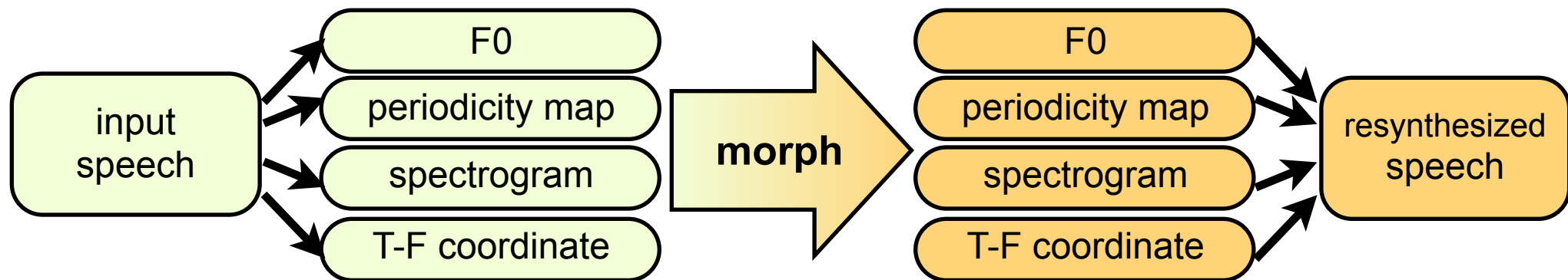
Insensitivity to pitch differences

Advanced technology for analysis

- STRAIGHT [Kawahara'06]

- High-quality analysis-resynthesis tool

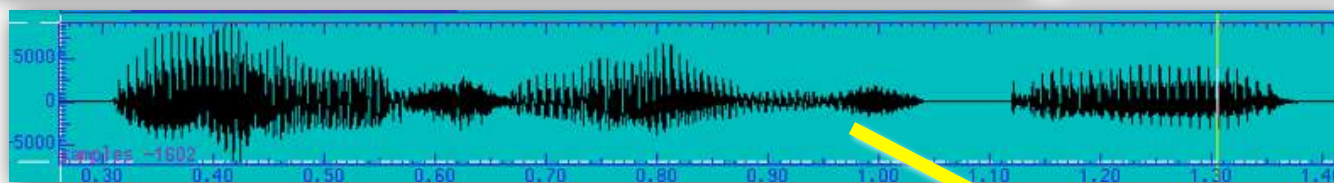
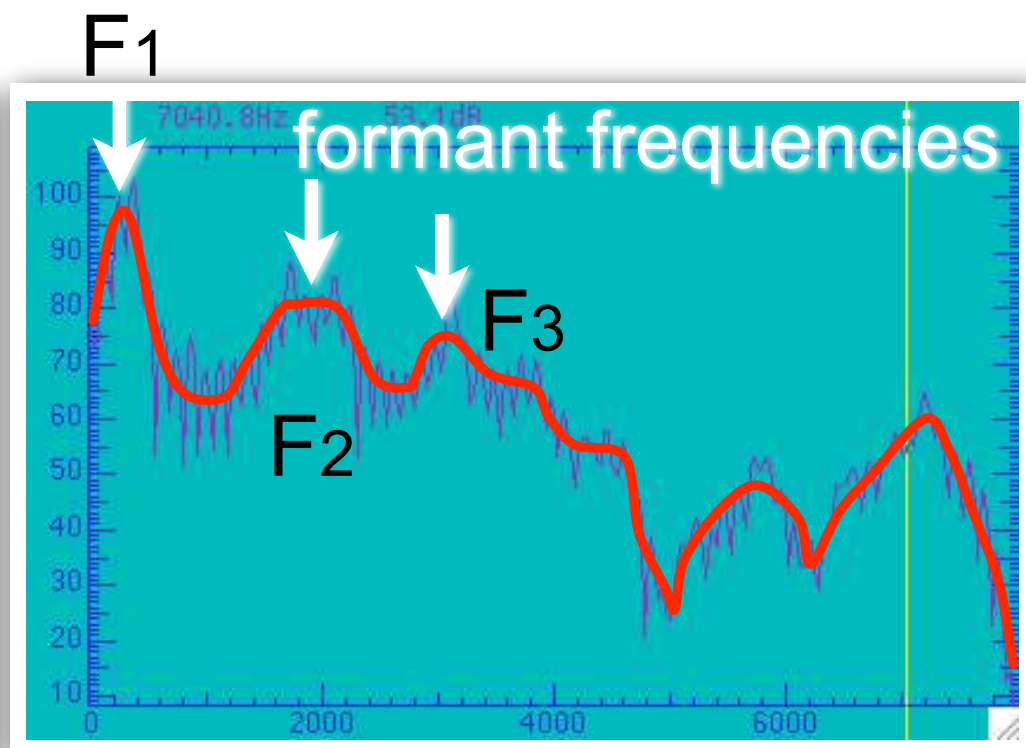
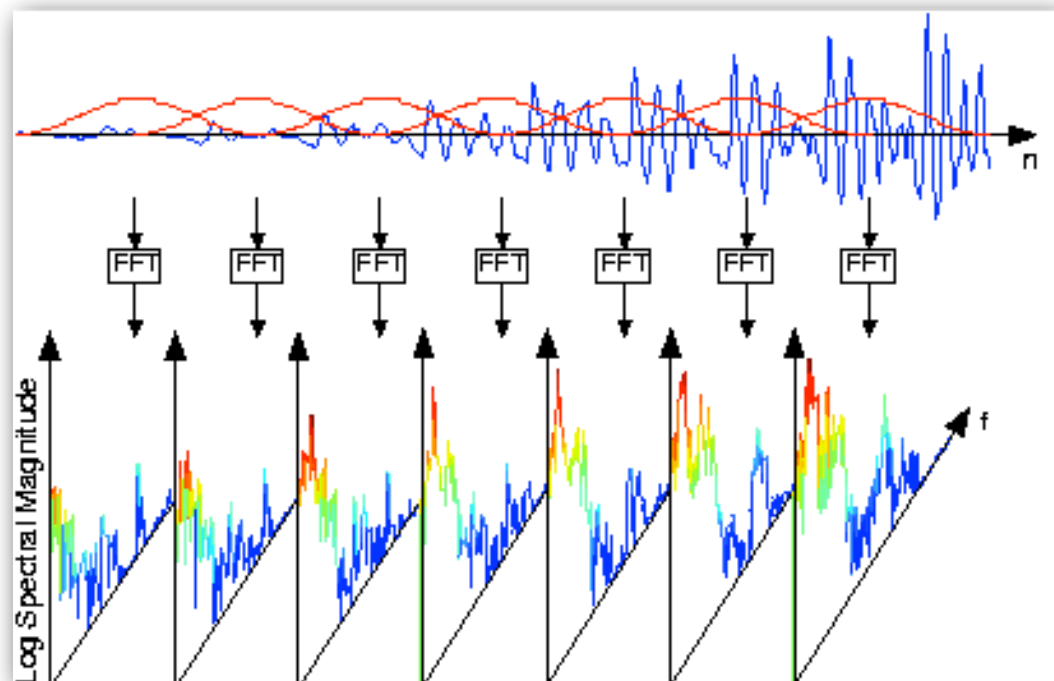
- Decomposition of speech into
 - Fundamental frequency, spectrographic representations of power, and that of periodicity
- High-quality speech morphing tool



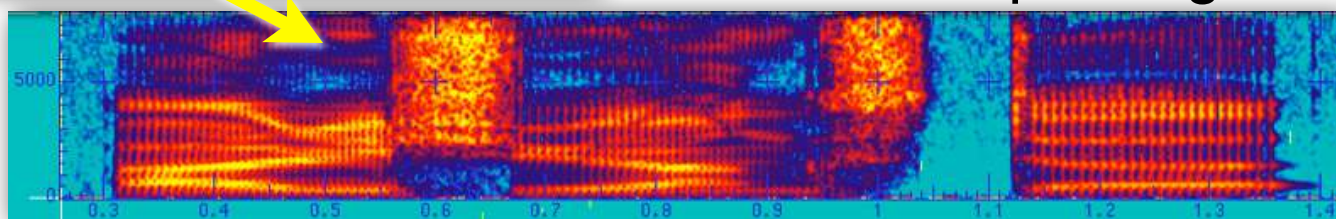
- Spectrographic representation of power
 - F0 adaptive complementary set of windows and spline based optimal smoothing
- Instantaneous frequency based F0 extraction
 - With correlation-based F0 extraction integrated
- Spectrographic representation of periodicity
 - Harmonic analysis based method

Acoustic phonetics

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

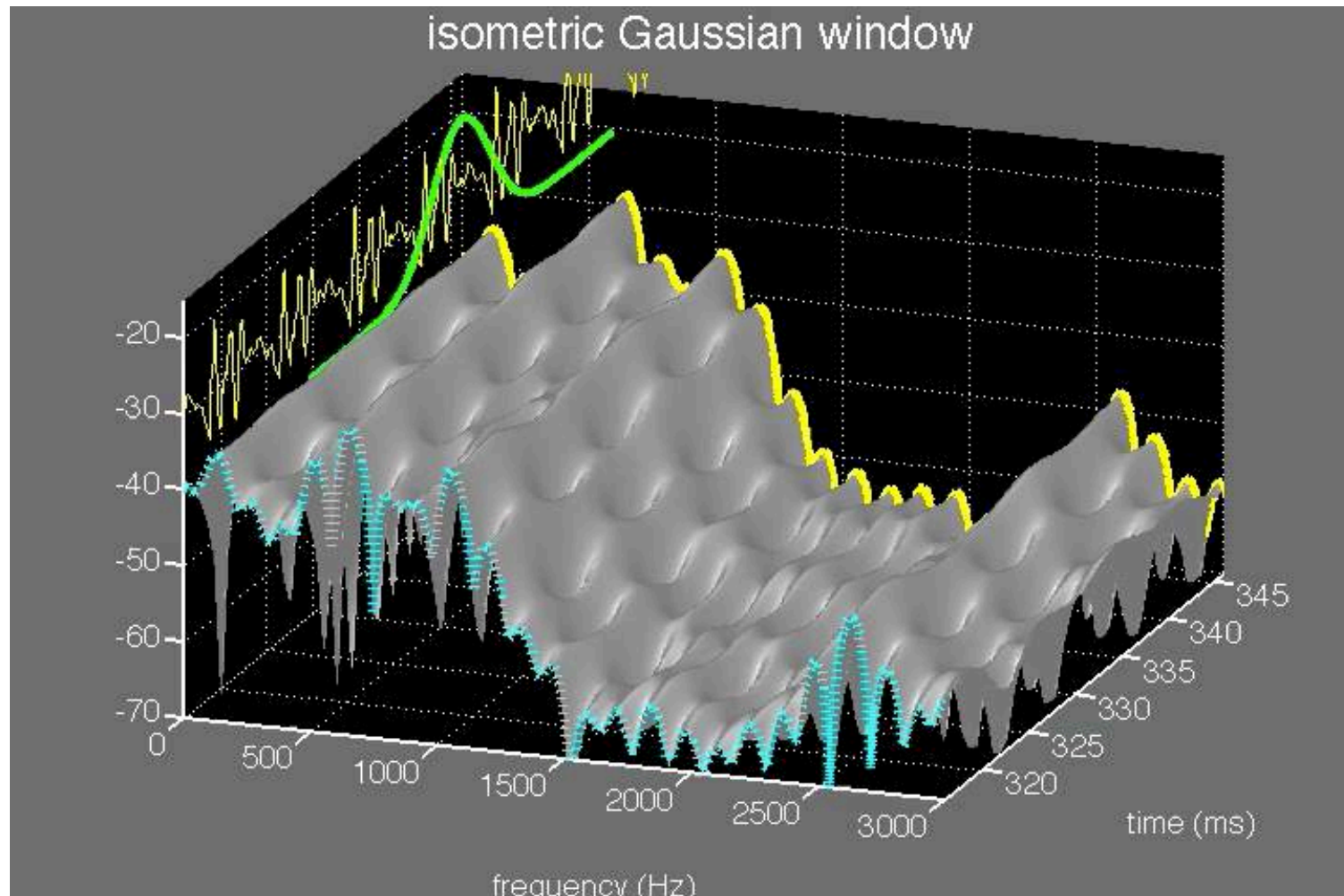


spectrogram



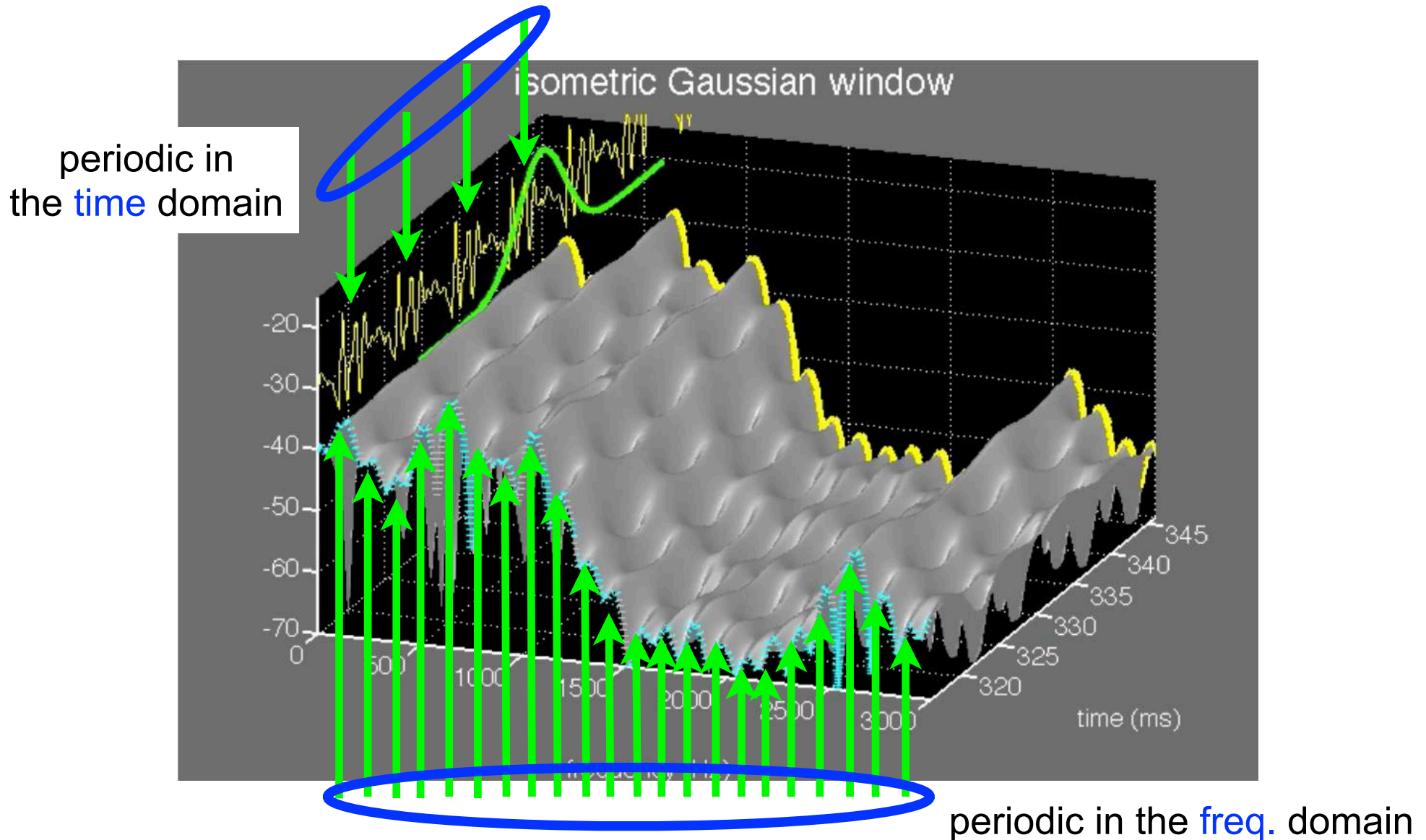
Advanced technology for analysis

- Short Time Fourier Transform (STFT)-based spectrogram



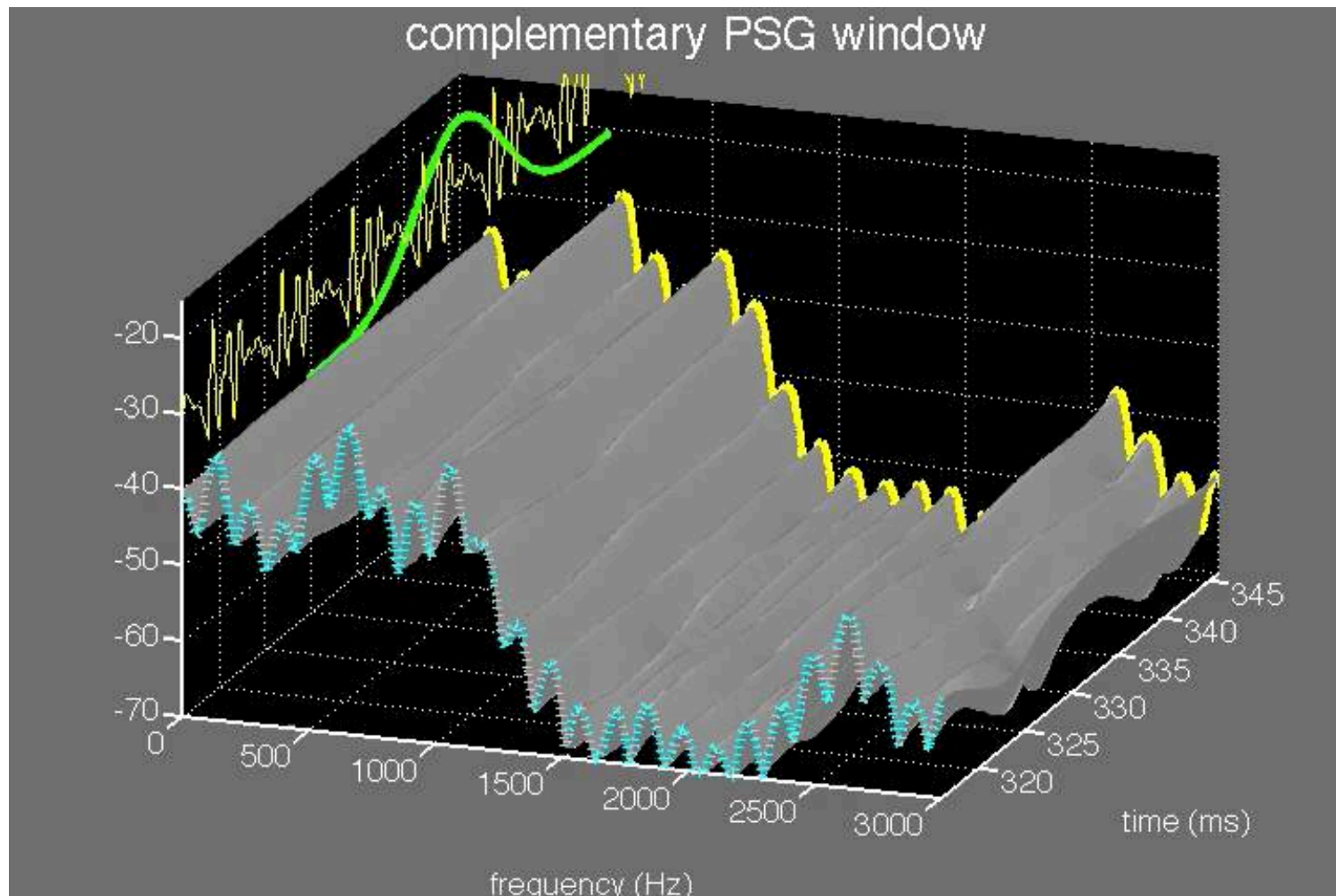
Advanced technology for analysis

- Short Time Fourier Transform (STFT)-based spectrogram



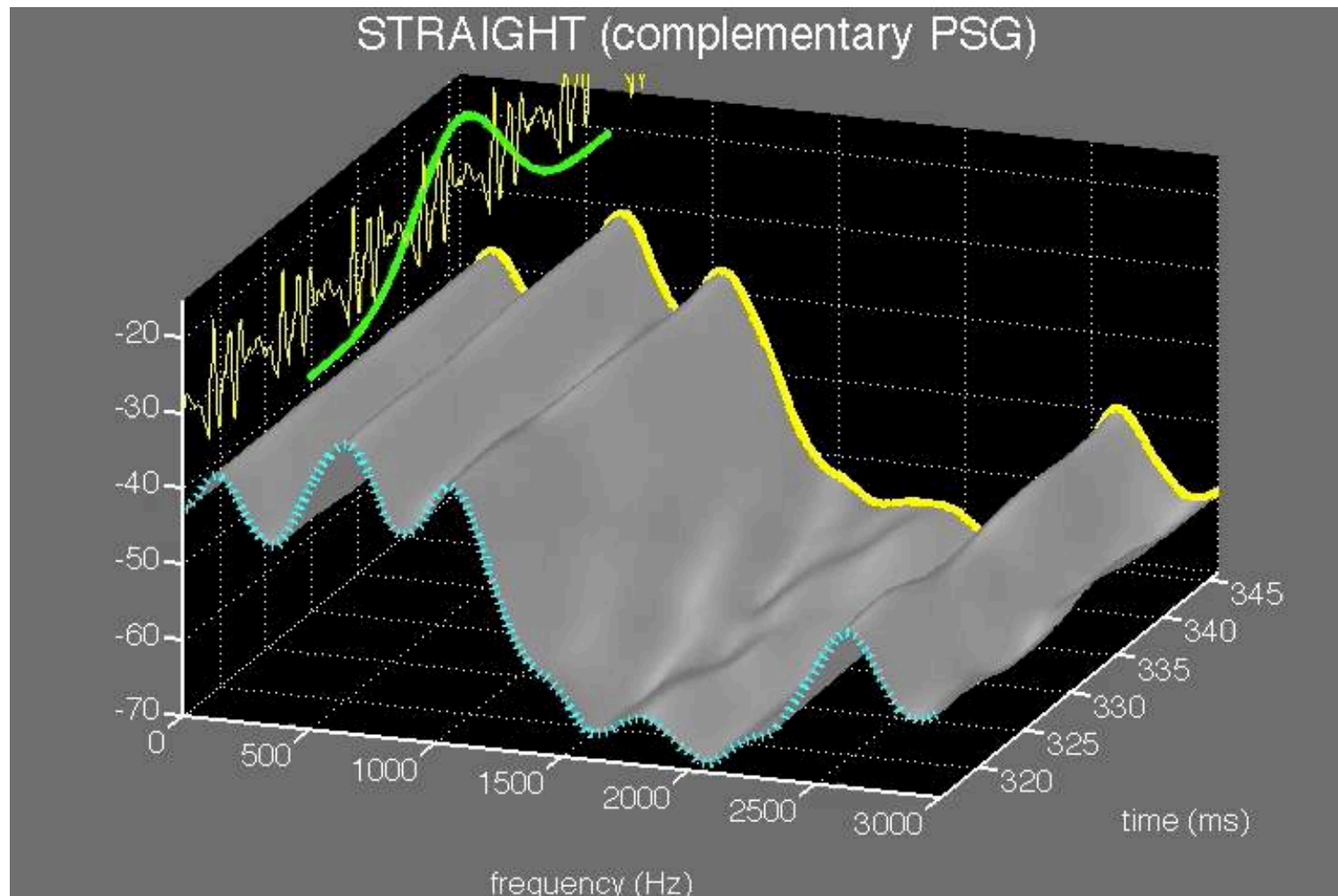
Advanced technology for analysis

- Complementary pitch-synchronous Gaussian window removes the repetitive structure in the time domain.

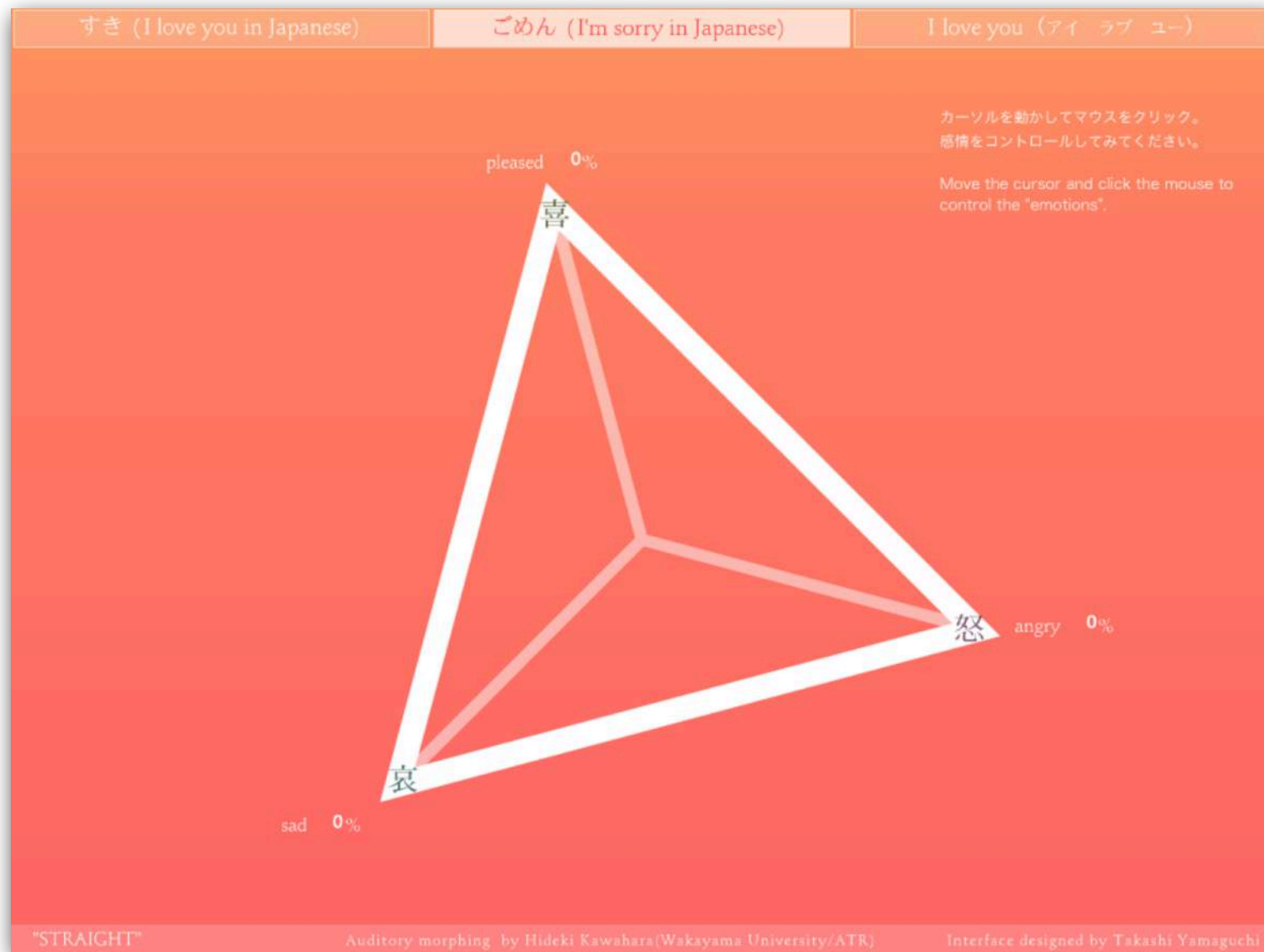


Advanced technology for analysis

- Spline-based optimum smoothing reconstructs the underlying smooth time-frequency representation.

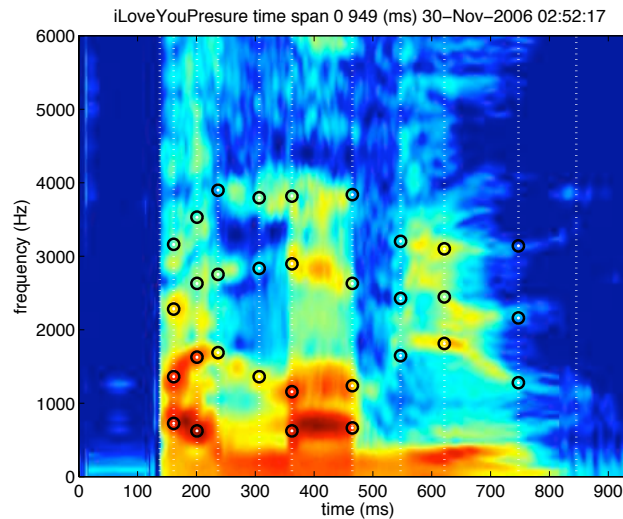


Examples of speech morphing



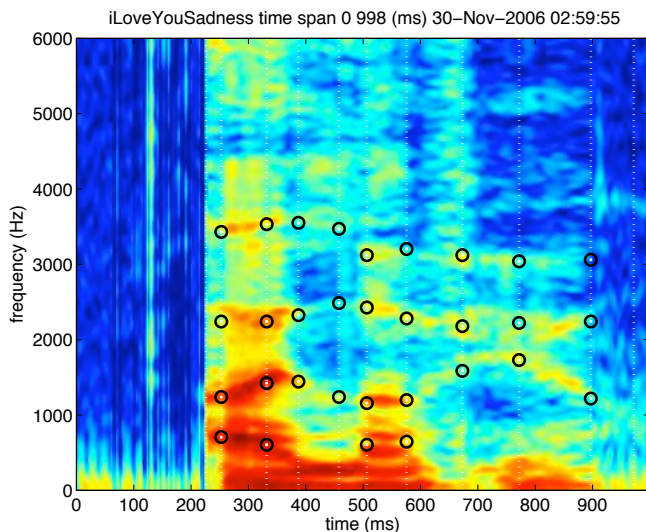
Examples of speech morphing

- Morphing with some anchor points

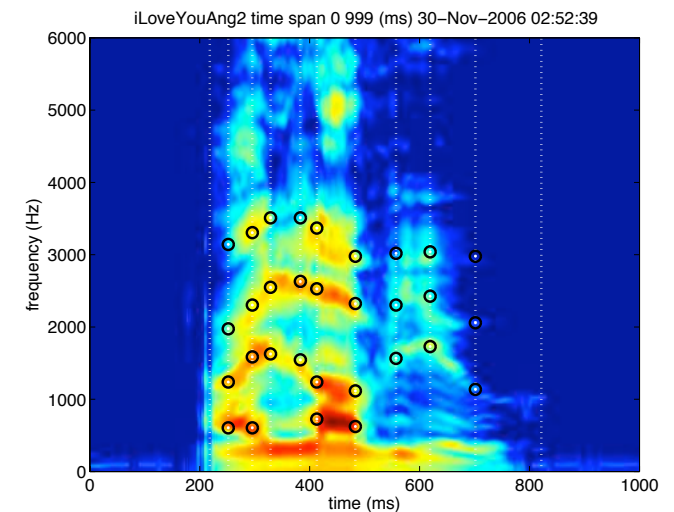


Pleasure

Sadness

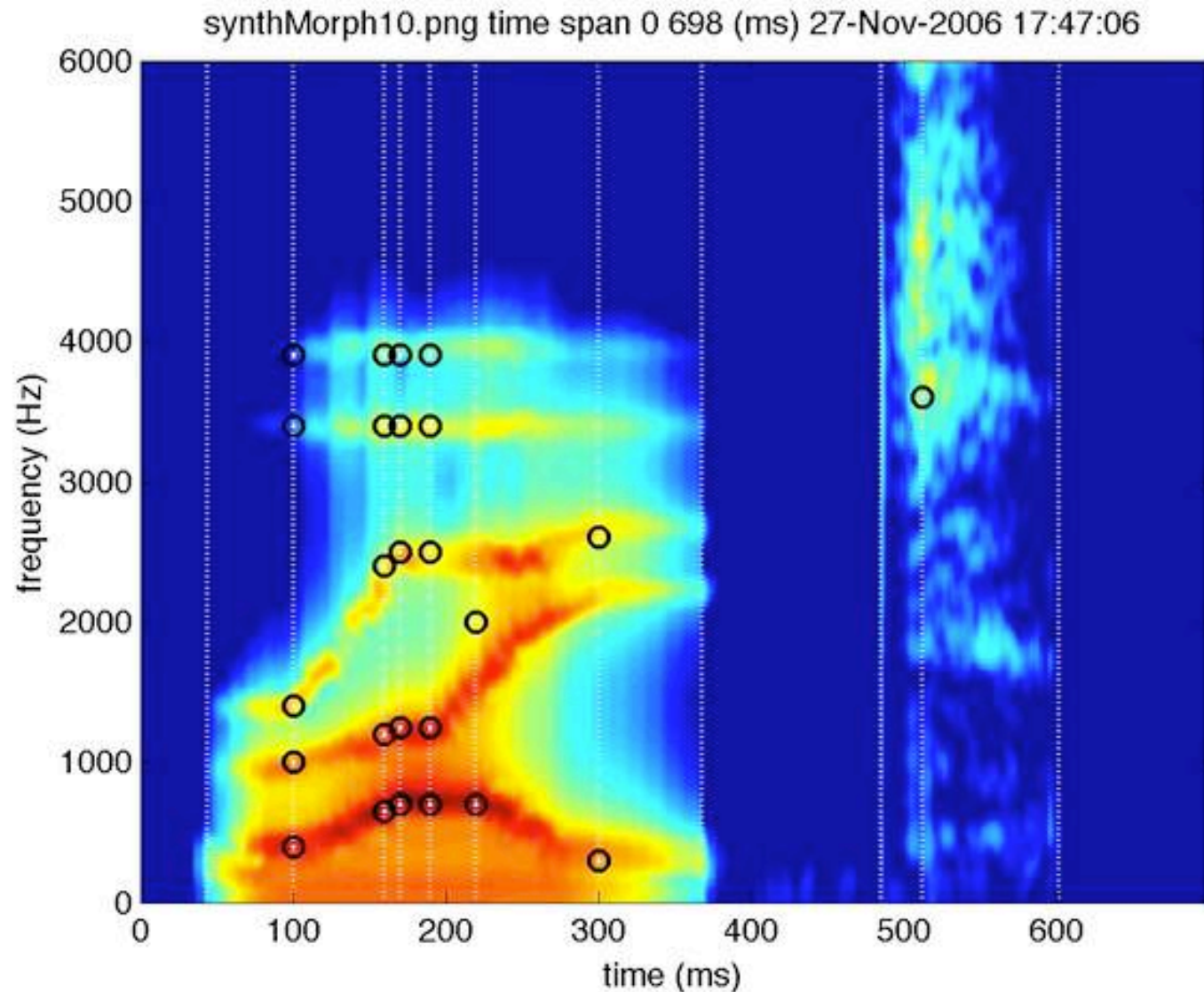


Anger



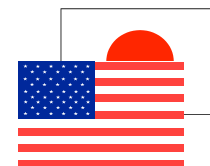
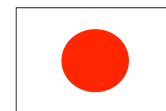
Examples of speech morphing

- R to L morphing bet. r/l-ight generated by Klatt synthesizer [Kubo+'98]

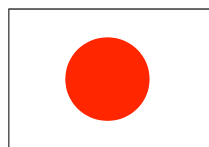


Examples of speech morphing

- Morphing of native utterance and its accented version [Kato+'11]
 - Use of a pair of word utterances spoken by a bilingual speaker
 - Normal Tokyo Japanese
 - Heavily American accented Japanese



igaku (medical science)



1.

fundamental frequency (F0)

3.

phonetic duration (dur)

4.

spectral envelope & aperiodicity (sp_ap)

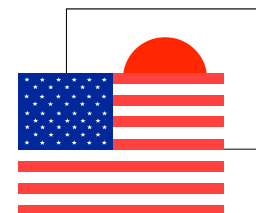
5.

F0 & dur (F0_dur)

6.

all the parameters (all)

7. = 2.



2.

0

0.25

0.5

0.75

1

morphing rate

Vocal tract shape and spectrum envelope

- Linear Predictive Coding (LPC)

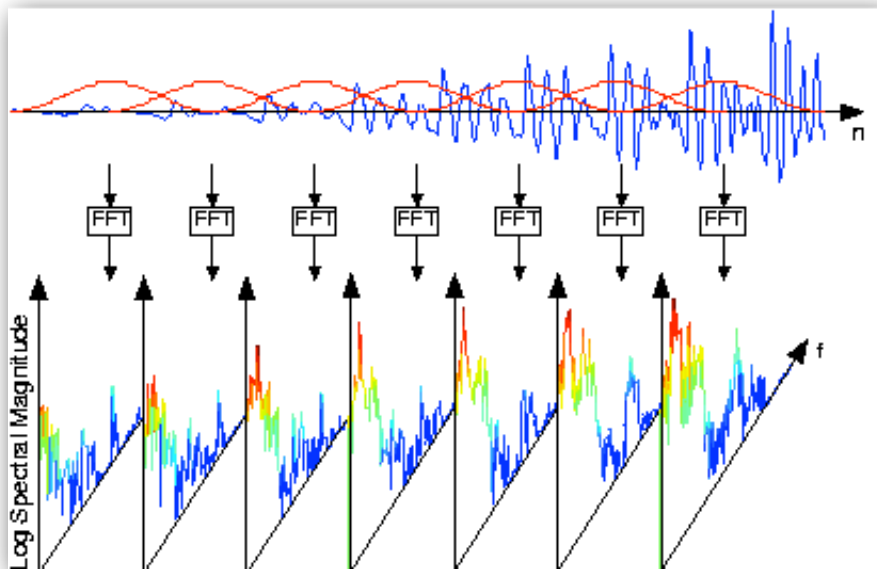
- Speech signal at time t , s_t , is predicted as weighted addition of some old signals.

$$s_t \leftarrow - \sum_{k=1}^p \alpha_k s_{t-k}$$

$$s_t + \sum_{k=1}^p \alpha_k s_{t-k} = \sum_{k=0}^p \alpha_k s_{t-k} = \varepsilon_t \quad (\alpha_0 = 1.0)$$

$$S(z) + \alpha_1 S(z)z^{-1} + \alpha_2 S(z)z^{-2} + \dots + \alpha_p S(z)z^{-p} = E(z)$$

$$S(z) = \frac{1}{1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \dots + \alpha_p z^{-p}} \quad E(z) = A(z)E(z)$$

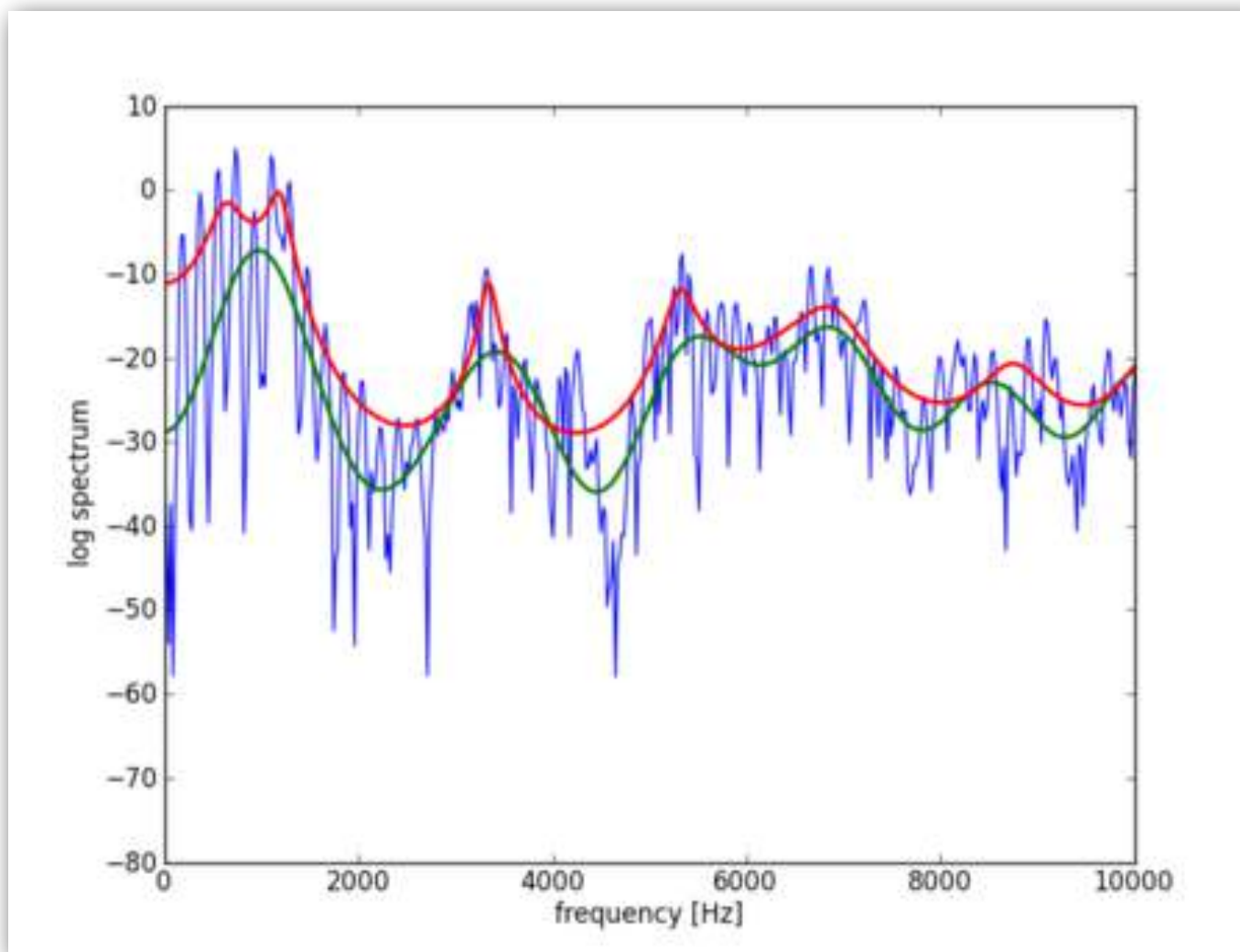


$\{\alpha_k\}$ are estimated for a frame in such a way that error term for that frame, ε_t , should be minimized.

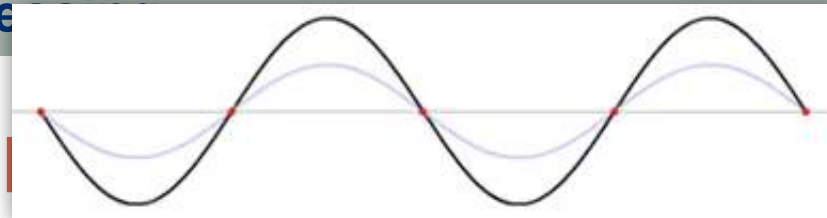
If error term is assumed to be white noise, then, the spectrum envelope shape is determined by $A(z)$.

Spectrum envelopes by CEP and LPC

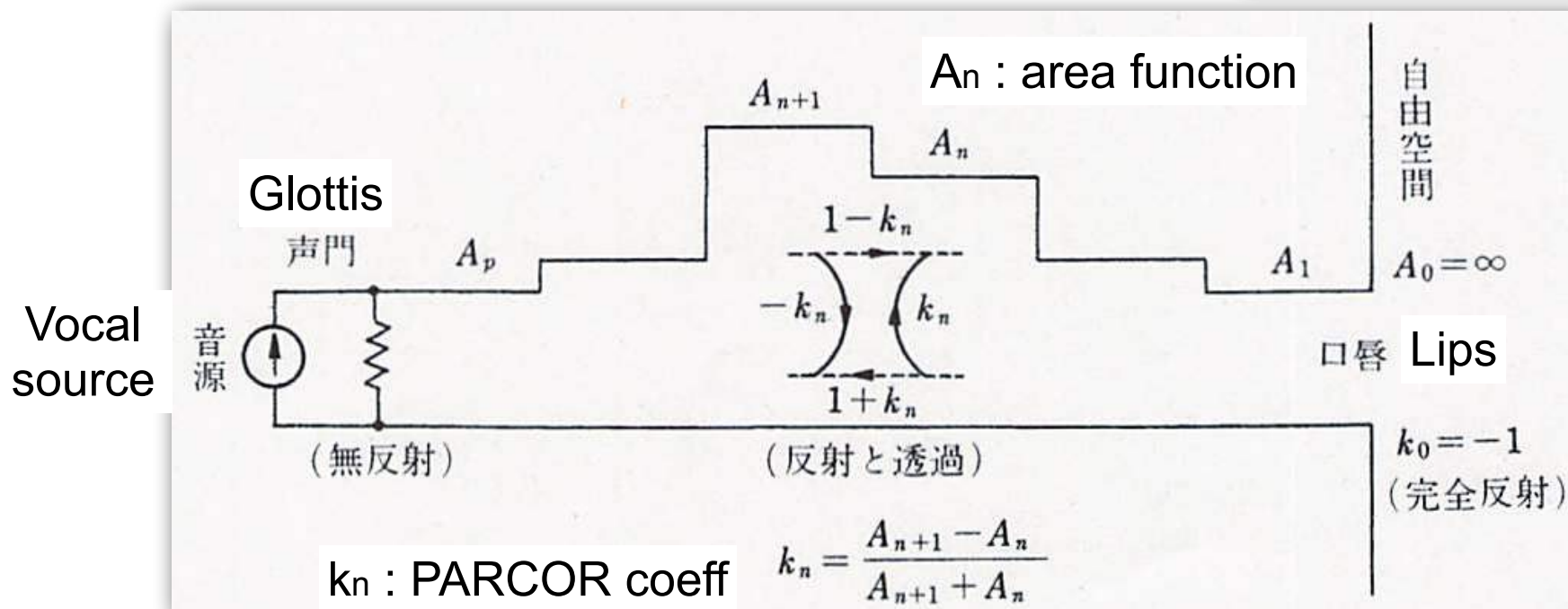
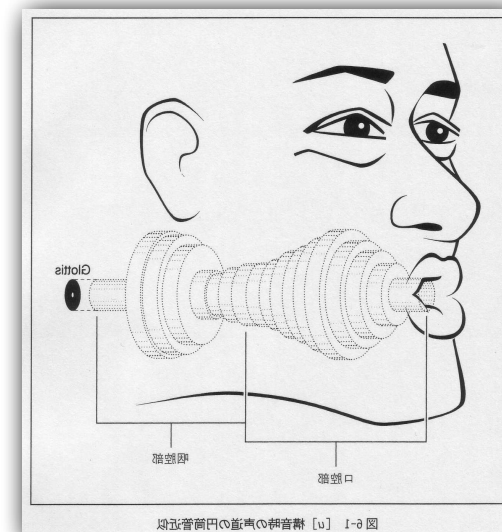
- Cepstrum-based envelope is just a smoothed spectrum.
- Spectral peaks are emphasized in LPC-based envelope.



LPC to vocal tract and



- $\{\alpha_k\}$ to the area function of the vocal tube.
- LPC coefficients are transformed into PARCOR (PARTial auto-CORrelation) coefficients.
- PARCOR coeff. are transformed to reflection coefficients between two consecutive short tubes.
- Finally PARCOR coefficients are related to the cross-sectional area of each short tube.



Today's menu

- More on details of acoustic phonetics (continued)
 - Characteristics of human hearing
 - Fundamental frequency and pitch again
 - Fourier analysis of speech signals
 - Simple hearing tests
- Technology for acoustic analysis of speech
 - Source-filter model of speech production $S(\omega) = G(\omega)H(\omega)R(\omega)$
 - Cepstrum method to separate source and filter
 - Advanced analysis tool of STRAIGHT
 - Some morphing examples
 - LPC, PARCOR, and the shape of a vocal tube
- Spectrums/waveforms of various language sounds
 - Vowels, semivowels, liquids, nasals, voiced fricatives, unvoiced fricatives, glottals,
 - Voiced plosives, unvoiced plosives, voiced affricatives, and unvoiced affricatives
 - Speech recognition as spectrum reading
- Summary

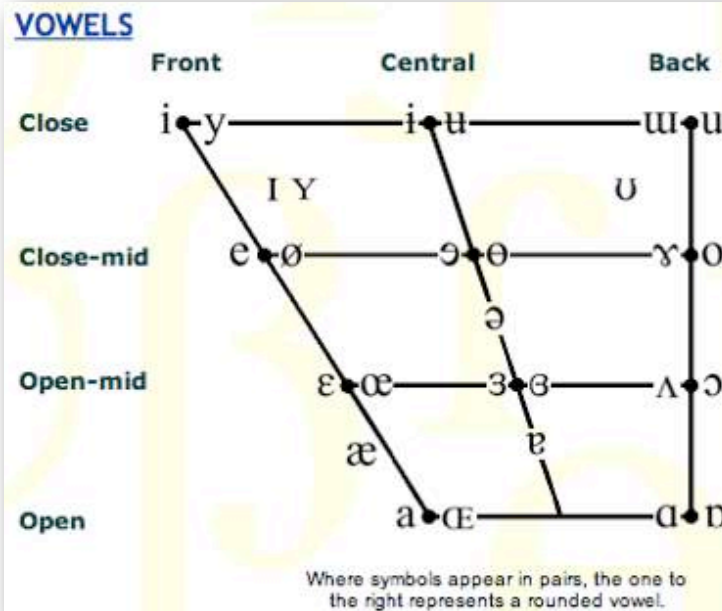


Various sounds in languages

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
<u>Plosive</u>	p b		t d			ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
<u>Nasal</u>		m ɱ		n ɳ		ɳ̠	ɲ	ŋ	ɴ		
<u>Trill</u>		ʙ		r					ʀ		
<u>Tap or Flap</u>				ɾ		ɽ					
<u>Fricative</u>	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							
<u>Approximant</u>			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.



Vowels

- Characteristics of vowels
 - Front vowels of /i/ and /e/: resonance at higher frequency bands
 - Middle vowels of /a/: energy distribution over a wide frequency range
 - Back vowels of /u/ and /o/: lower bands are dominant in energy distribution
 - Unvoiced vowels

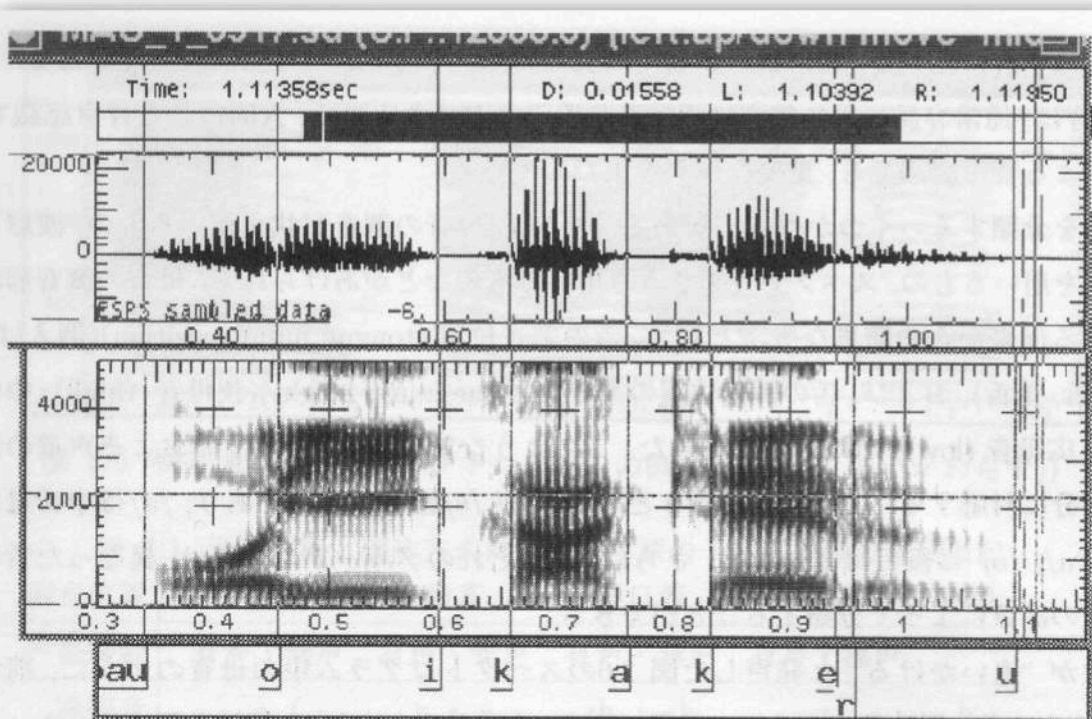


図 2.6 “追いかける/oikakeru/”と発声した音声の波形とスペクトログラム

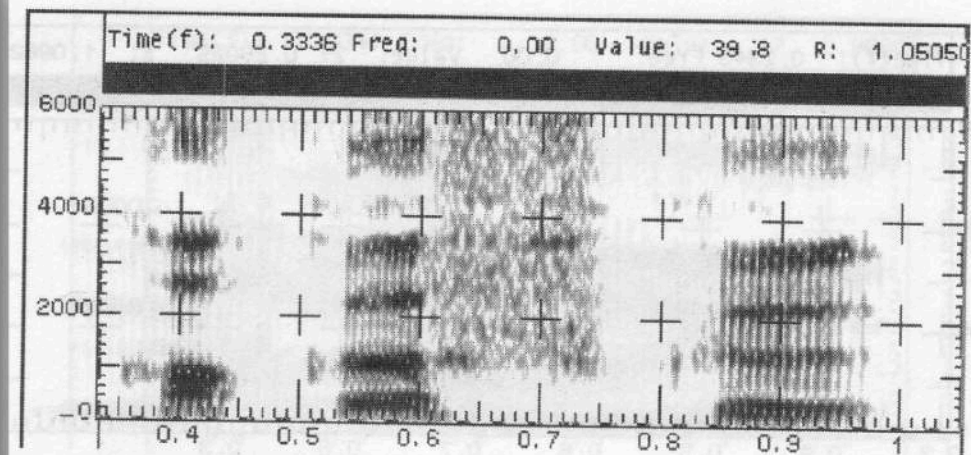
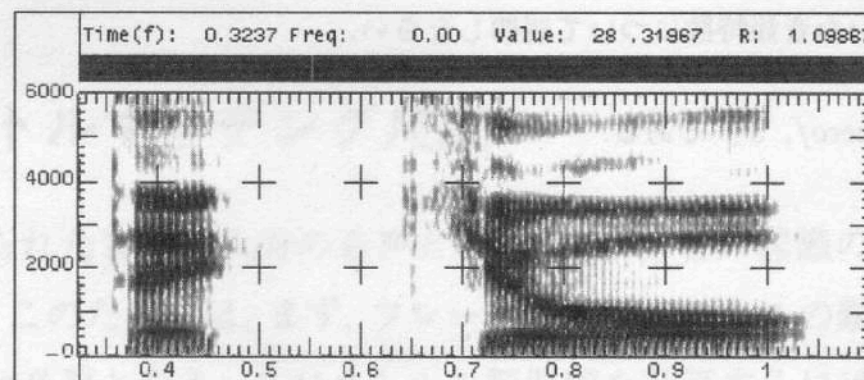
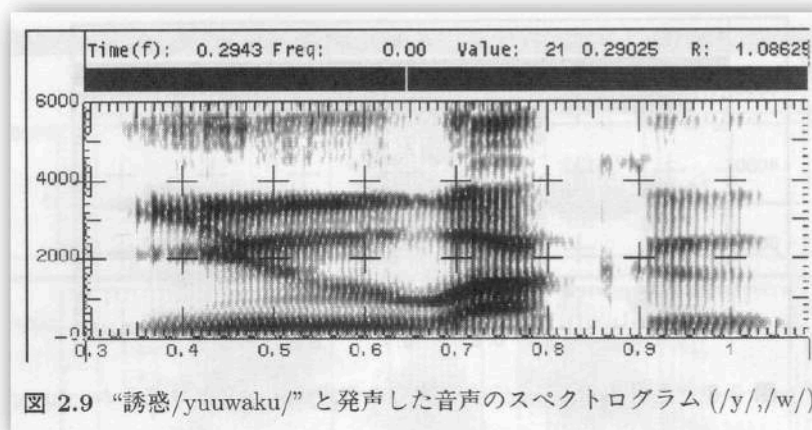


図 2.8 “克服/kokufuku/”と発声した音声のスペクトログラム

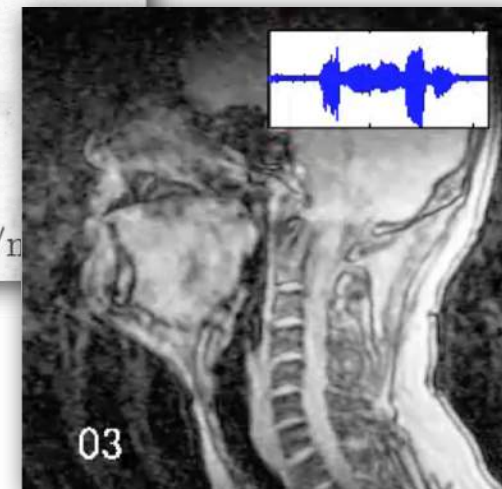
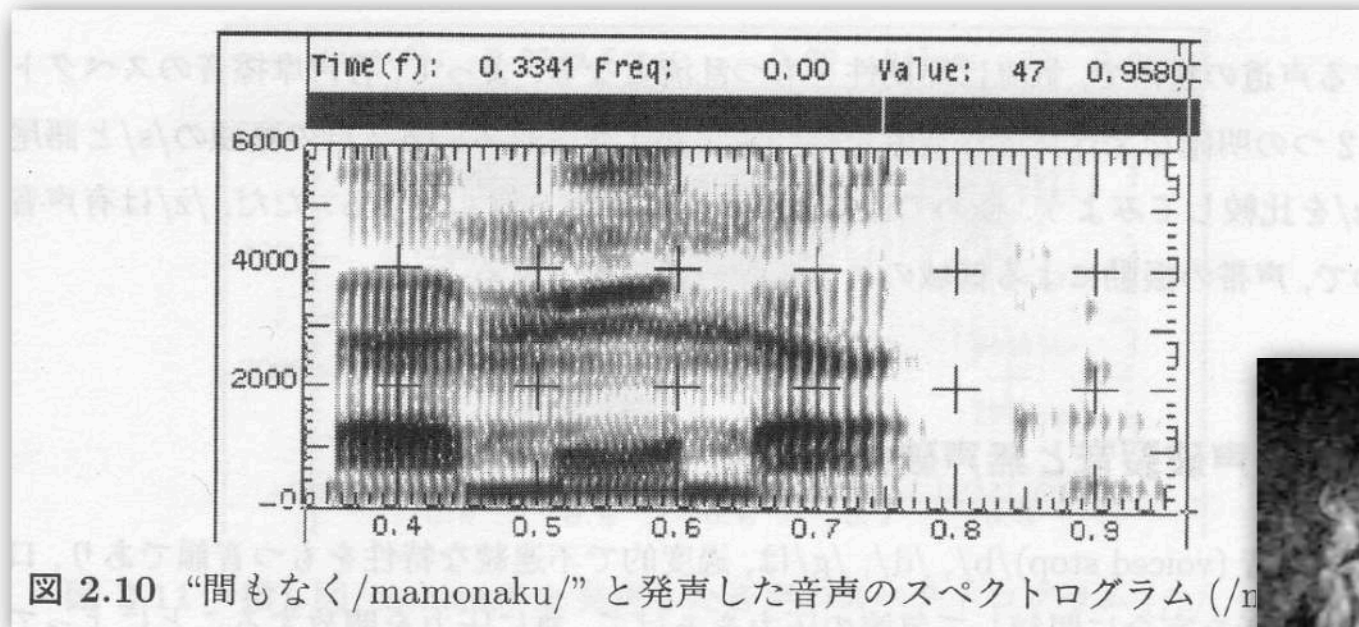
Semivowels and liquids

- Characteristics of semivowels and liquids
 - /w/, /y/, and /r/: characterized by their transitional parts from/to neighboring phones.
 - Large dependency on phonemic context.



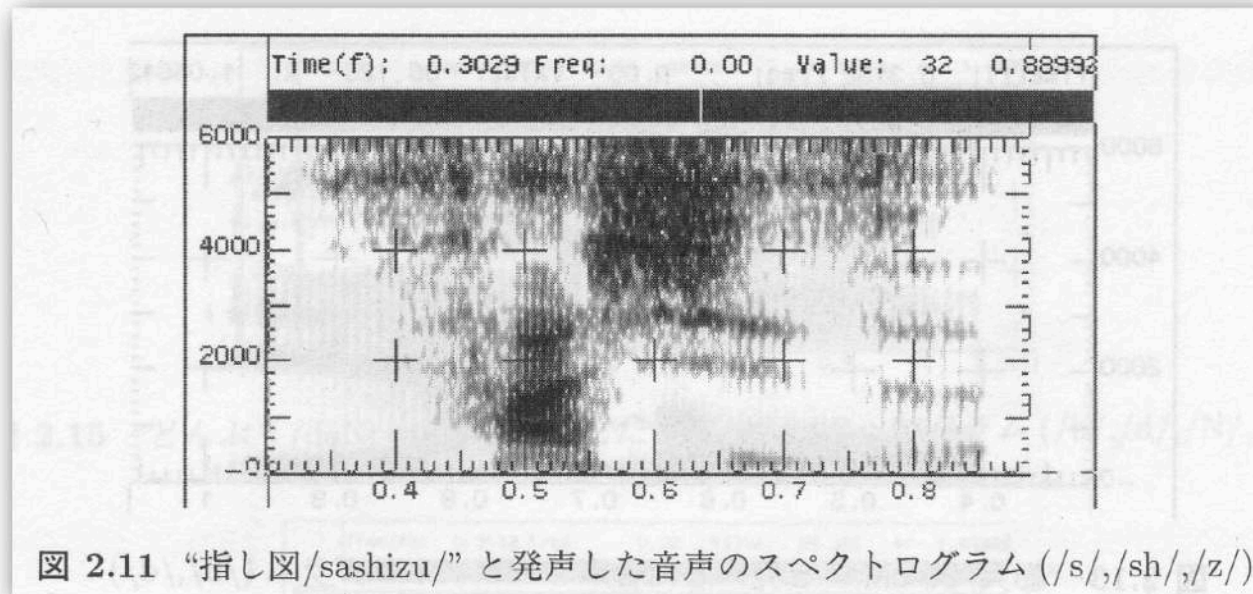
Nasals

- Characteristics of nasals
 - /m/, /n/, /ŋg/, /N/: a pathway to the nasal cavity shows its own acoustic features.
 - Closed vocal cavity and open nasal cavity cause antiresonance.
 - Transitional parts from/to neighboring vowels are useful in identifying nasal sounds.



Unvoiced fricatives and glottals

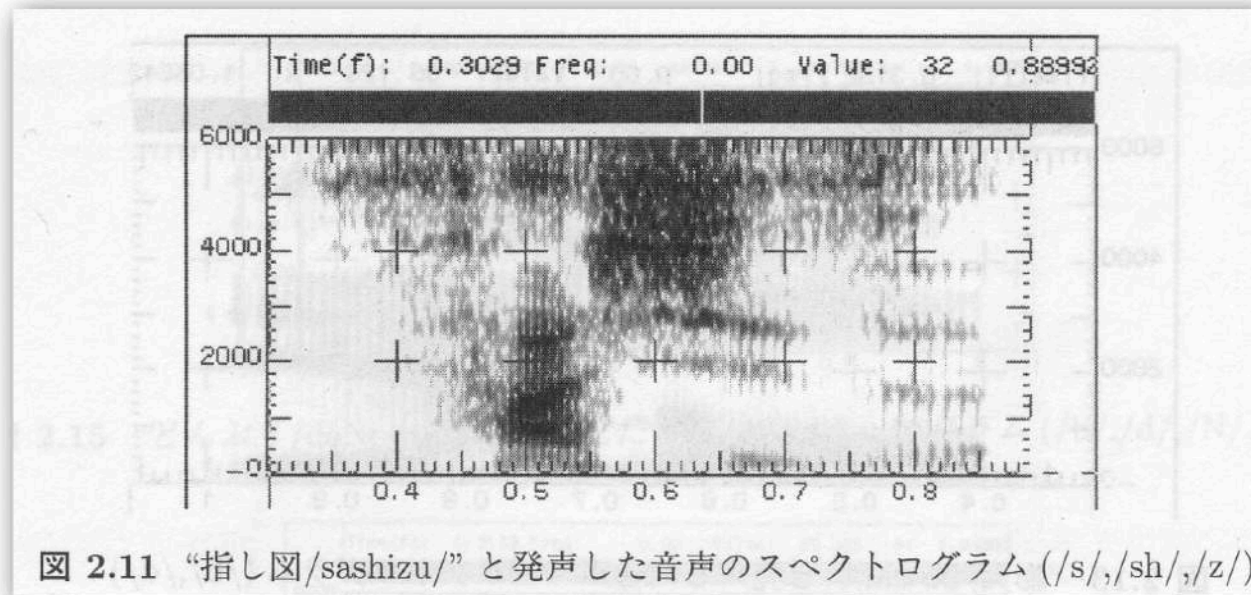
- Characteristics of unvoiced fricatives and glottals
 - /f/, /s/, /sh/: energy distribution at higher frequency bands
 - A vocal cavity from the (almost) closing point to the lung causes antiresonance.



- /h/: fricative at glottis
 - The shape of the vocal cavity is the same as that of the succeeding vowel.
 - No antiresonance

Voiced fricatives

- Characteristics of voiced fricatives
 - /v/, /z/, /zh/: source sounds are generated at two positions, glottal sources and fricative sources
 - Energy distribution is found at a very low frequency band due to the glottal source.



Voiced plosives and unvoiced plosives

- Characteristics of voiced plosives and unvoiced plosives
 - /b/, /d/, /g/ /p/, /t/, /k/
 - Complete closure in the vocal tract at a time and abrupt release of air flow
 - Buzz-bar: closed vocal tract + vocal fold vibration --> radiation from the skin
 - Transitional parts from/to neighboring vowels are useful in identifying nasal sounds.

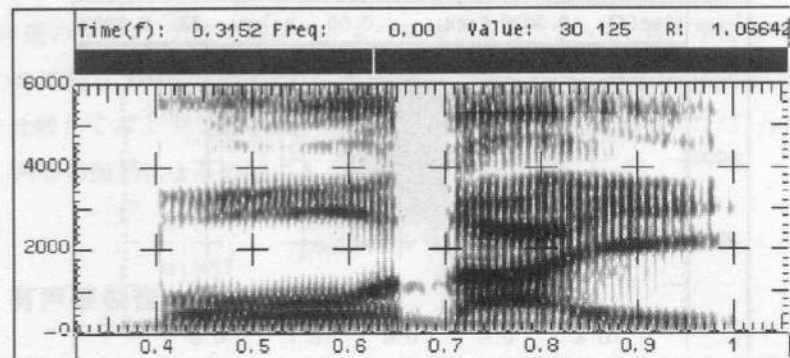


図 2.13 “膨大/bo-dai/”と発声した音声のスペクトログラム (/b/,/d/)

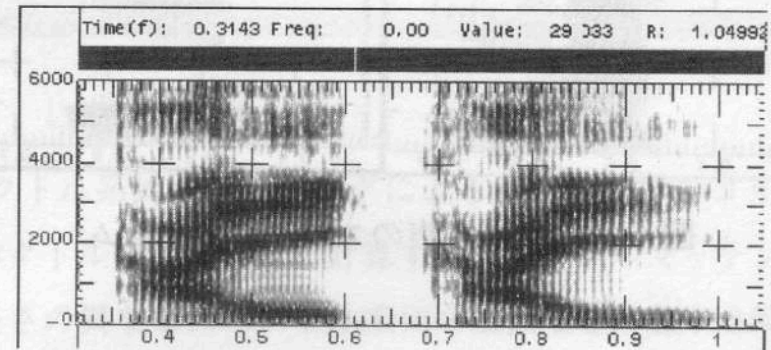


図 2.16 “大会/taikai/”と発声した音声のスペクトログラム (/t/,/k/)

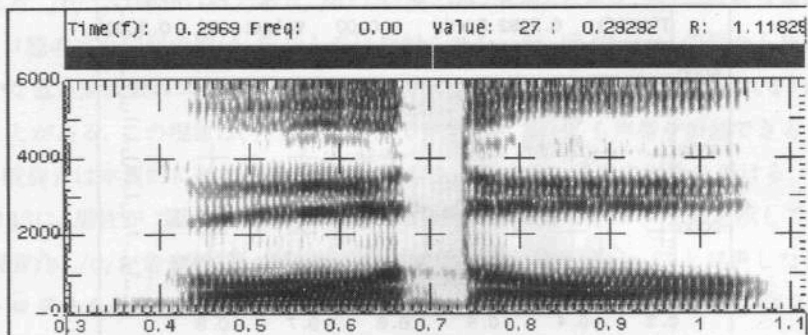


図 2.14 “合同/go-do-”と発声した音声のスペクトログラム (/g/,/d/)

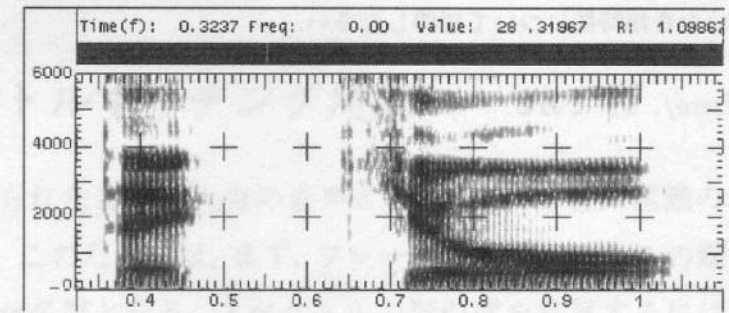
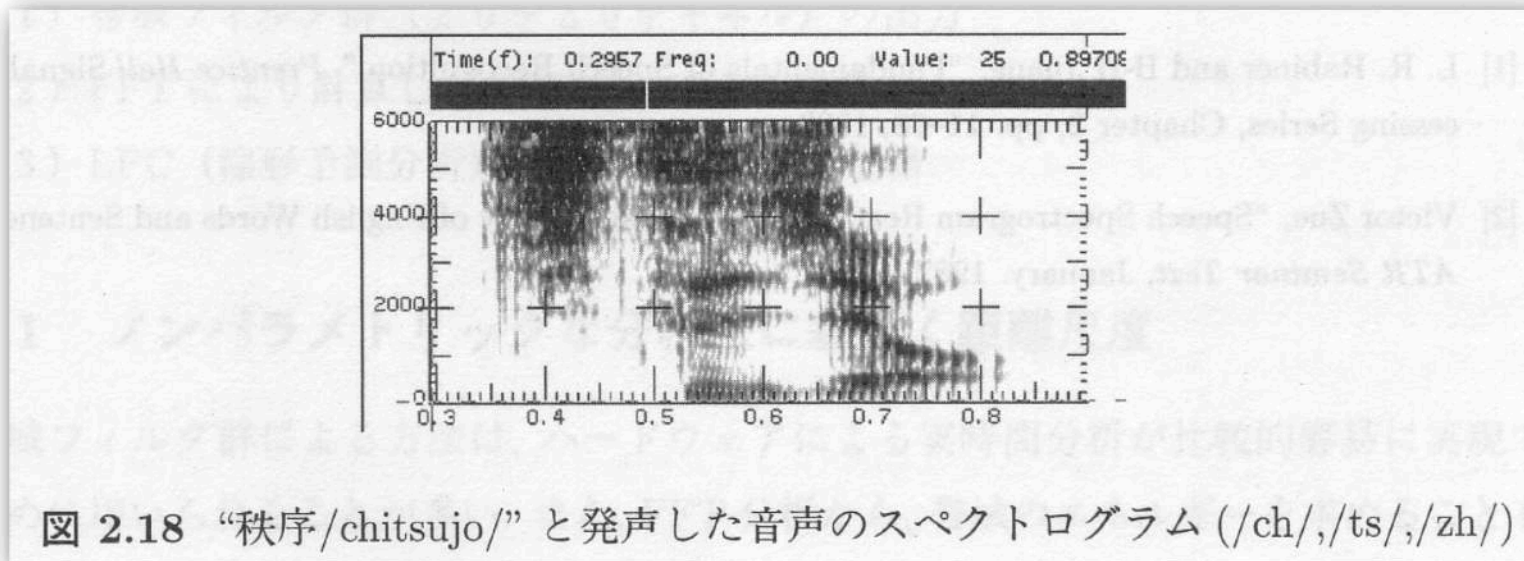


図 2.17 “鉄橋/teQkyo-”と発声した音声のスペクトログラム (/t/,/Q/,/ky/)

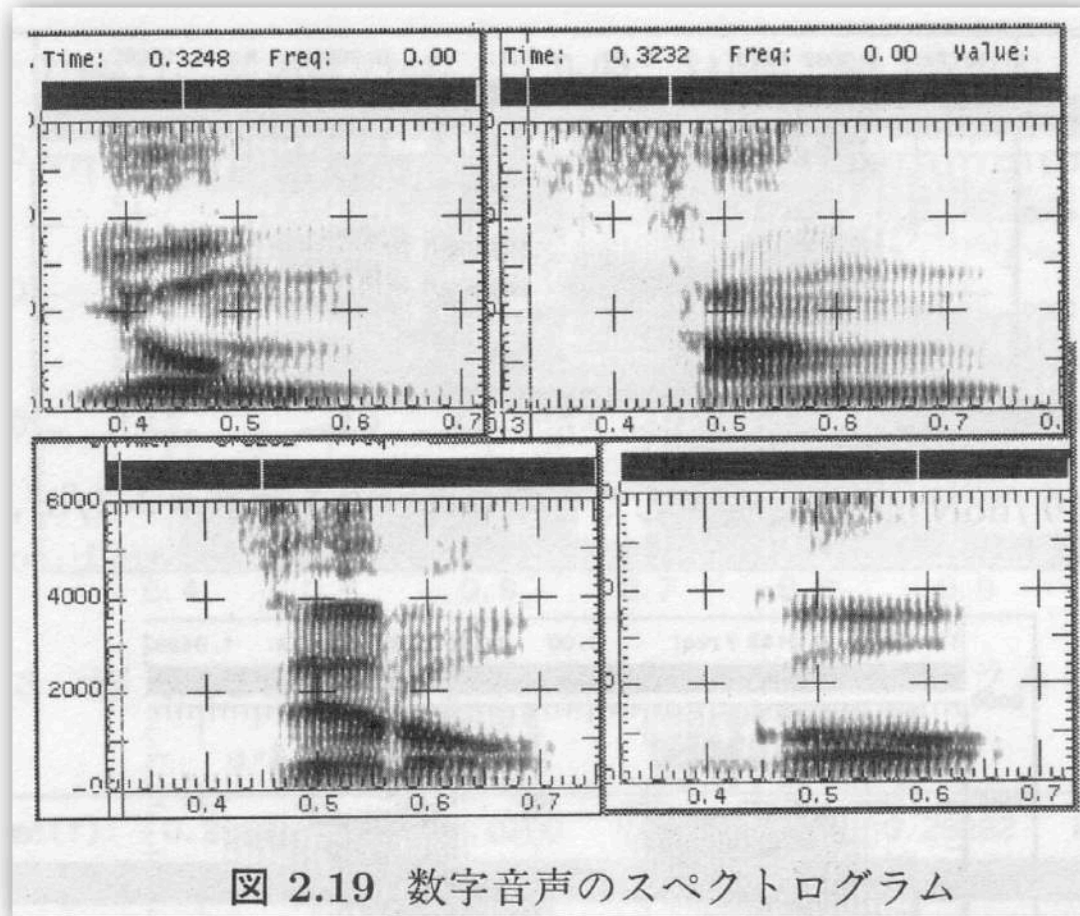
Voiced affricates and unvoiced affricates

- Characteristics of voiced affricates and unvoiced affricates
 - Affricate = plosive + fricative
 - /dz/, /dh/, /ts/, /ch/



Spectrum reading

- What are these?
 - Hint : they are numbers.



- This is the task that is done by a speech recognizer.

Today's menu

- More on details of acoustic phonetics (continued)
 - Characteristics of human hearing
 - Fundamental frequency and pitch again
 - Fourier analysis of speech signals
 - Simple hearing tests
- Technology for acoustic analysis of speech
 - Source-filter model of speech production $S(\omega) = G(\omega)H(\omega)R(\omega)$
 - Cepstrum method to separate source and filter
 - Advanced analysis tool of STRAIGHT
 - Some morphing examples
 - LPC, PARCOR, and the shape of a vocal tube
- Spectrums/waveforms of various language sounds
 - Vowels, semivowels, liquids, nasals, voiced fricatives, unvoiced fricatives, glottals,
 - Voiced plosives, unvoiced plosives, voiced affricatives, and unvoiced affricatives
 - Speech recognition as spectrum reading
- Summary



Recommended books

