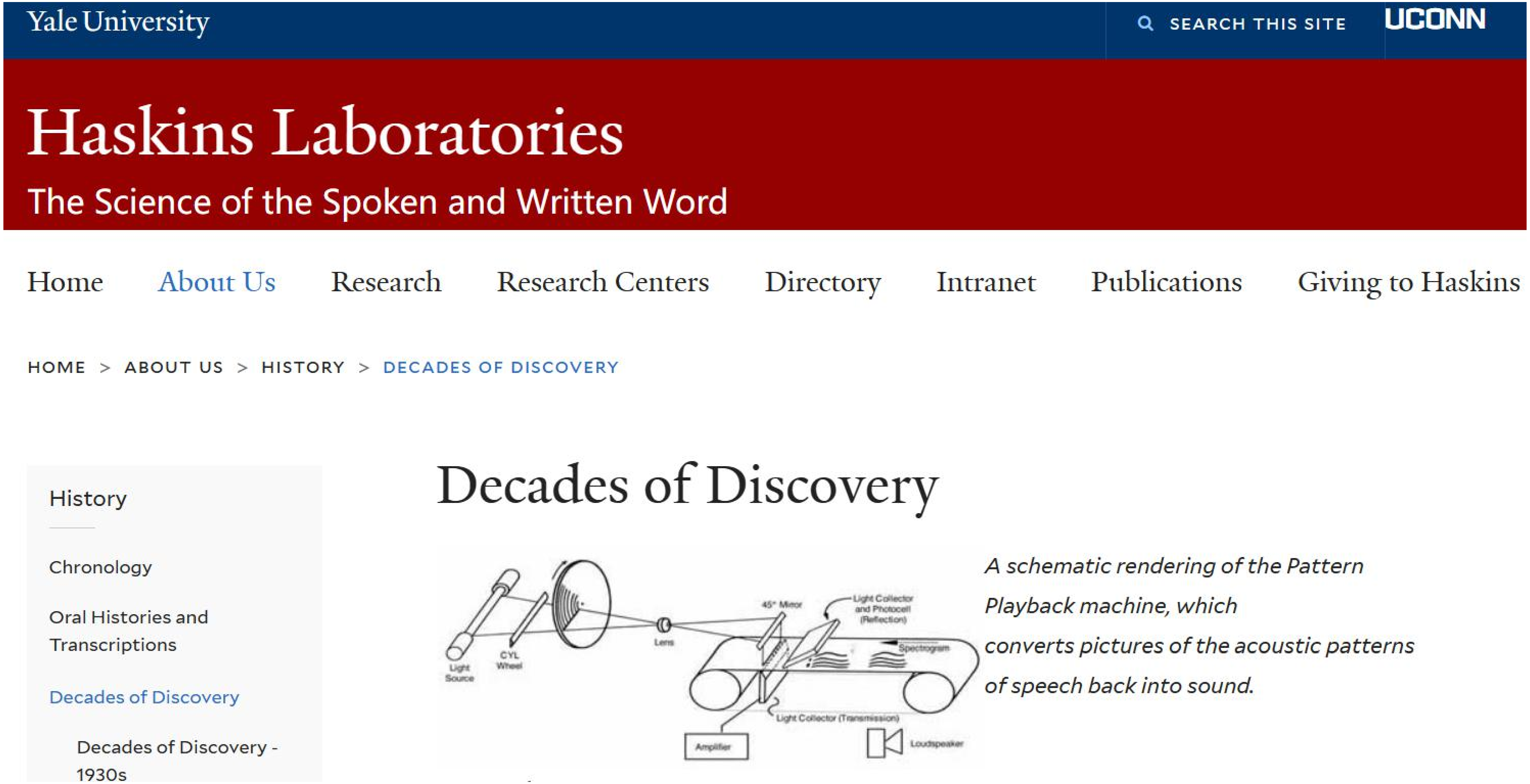
http://dkim.rbind.io/img/headers/perception-production-link.jpg

SPEECH PERCEPTION AND PRODUCTION

Chapter 3

QUESTIONS

- What is a theory? What is a hypothesis? What is a model?
- What are differences between the experimental condition and the control condition?
- What is priming? What's eye-tracking?
- Is human language a result of natural selection?



Yale University

SEARCH THIS SITE

UCONN

Haskins Laboratories

The Science of the Spoken and Written Word

Home About Us Research Research Centers Directory Intranet Publications Giving to Haskins

HOME > ABOUT US > HISTORY > DECADES OF DISCOVERY

History

Chronology

Oral Histories and Transcriptions

Decades of Discovery

Decades of Discovery - 1930s

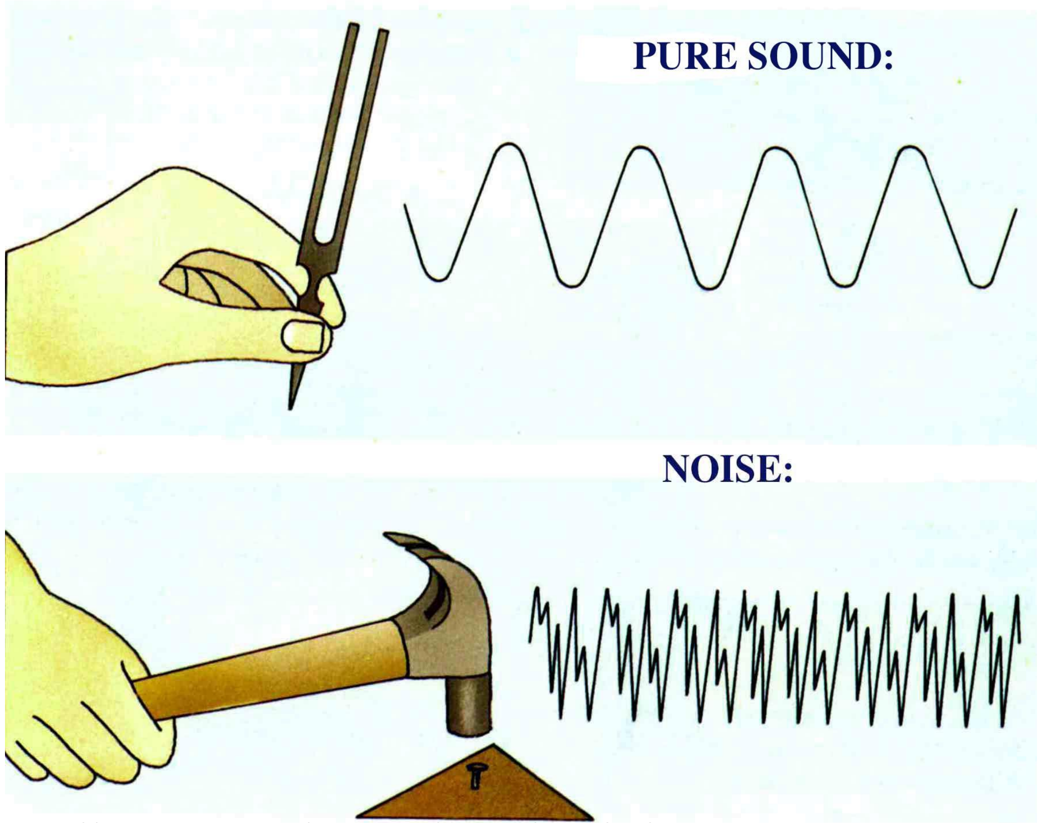
Decades of Discovery

Light Source
CYL Wheel
Lens
45° Mirror
Light Collector and Photocell (Reflection)
Spectrogram
Light Collector (Transmission)
Amplifier
Loudspeaker

A schematic rendering of the Pattern Playback machine, which converts pictures of the acoustic patterns of speech back into sound.

CHAPTER 3

1. Auditory perception
2. The speech stream
3. Development of speech perception
4. Theories of speech perception



http://www.pixelstalk.net/wp-content/uploads/2016/06/Noise-Sound-Wave-Picture.jpeg

SOUND PROPERTIES

Frequency (Hz)

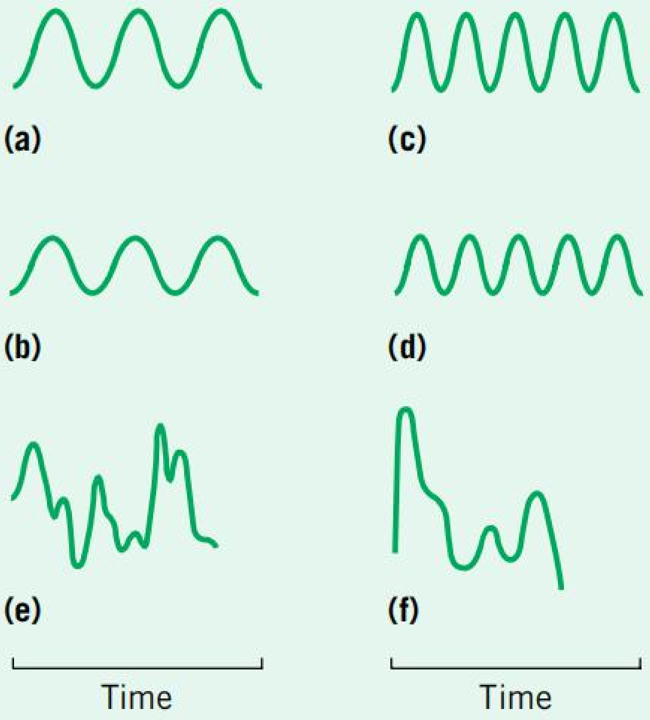
Number of wavelengths that pass by a given point in a given amount of time. (Pitch)

Amtplitude (Hz)

Amount of change that a wave undergoes during one cycle. (Loudness)

Figure 3.1 Examples of Sound Waves (G9.3)

Waveforms (a) and (b) have the same frequency but different amplitudes, as do waveforms (c) and (d). Waveforms (a) and (c) have the same amplitudes but different frequencies, as do waveforms (b) and (d). Waveform (e) is aperiodic noise, while waveform (f) is one cycle of a periodic musical note played on a clarinet.



SOUND PROPERTIES

Fundamental frequency:
Lowest frequency produced by a vibrating object.

Overtone
Frequencies higher than the fundamental that are also produced by a vibrating object.

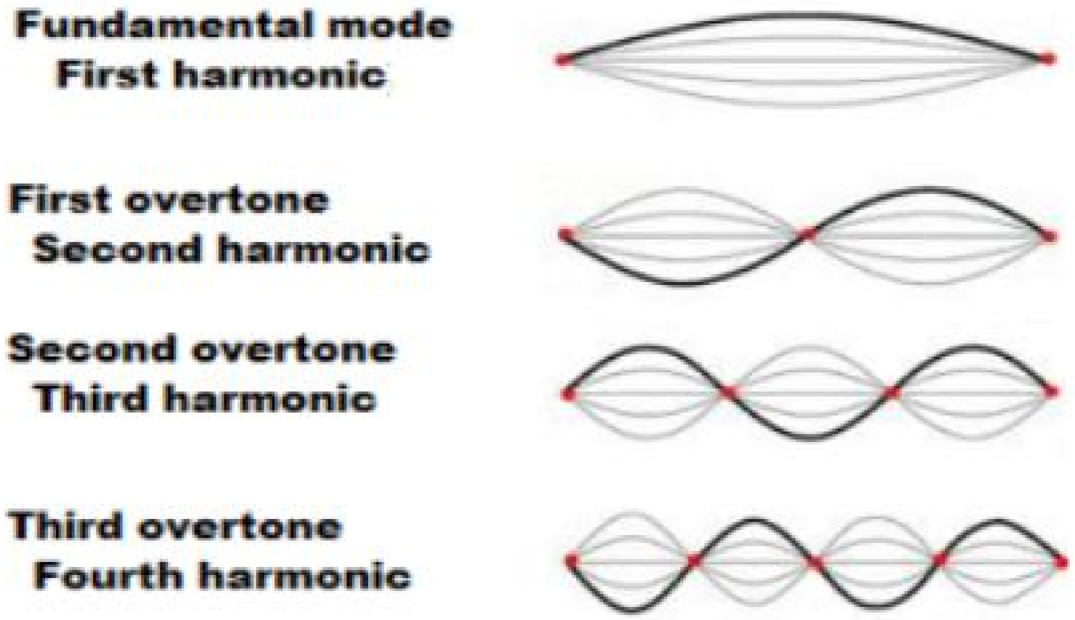


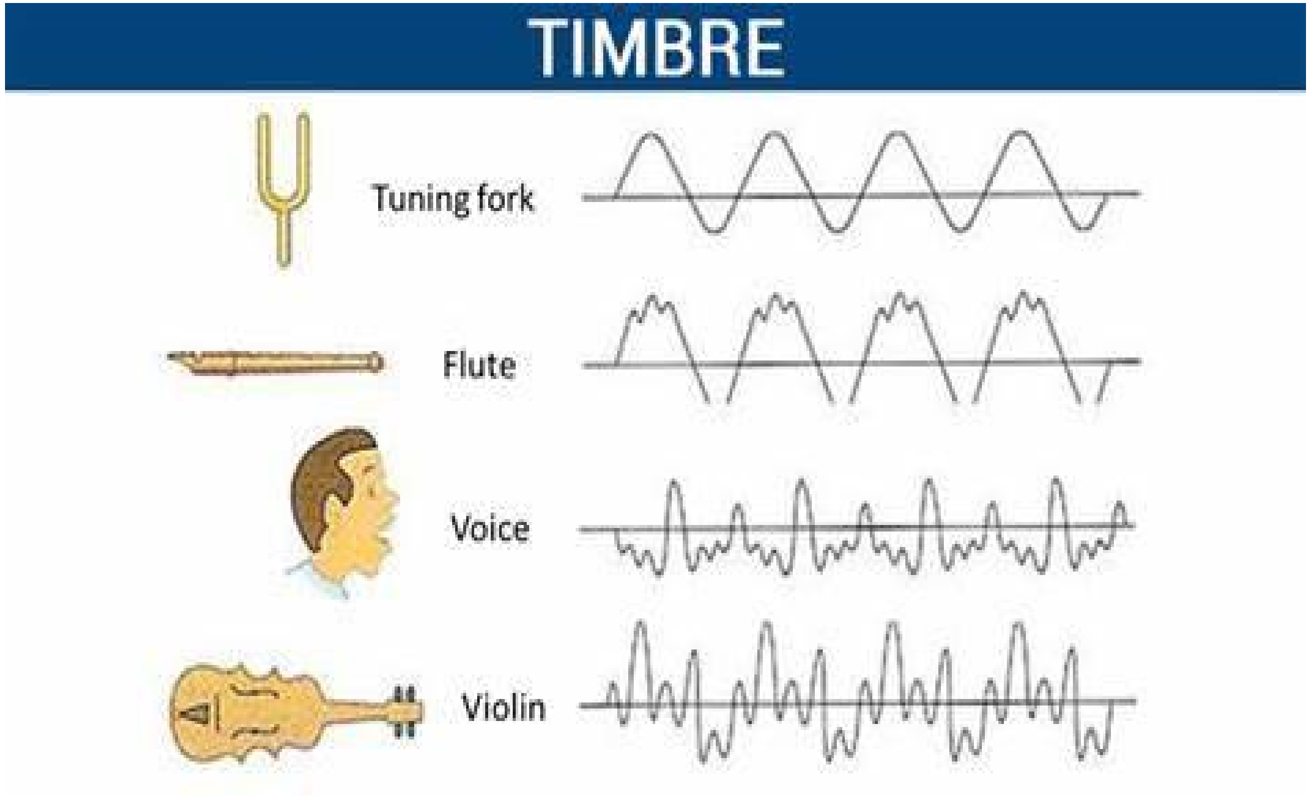
Fig. 1

SOUND PROPERTIES

Fundamental frequency:
Lowest frequency produced by a vibrating object.

+

Overtone
Frequencies higher than the fundamental that are also produced by a vibrating object.



= sound wave complexity (timbre,音色)

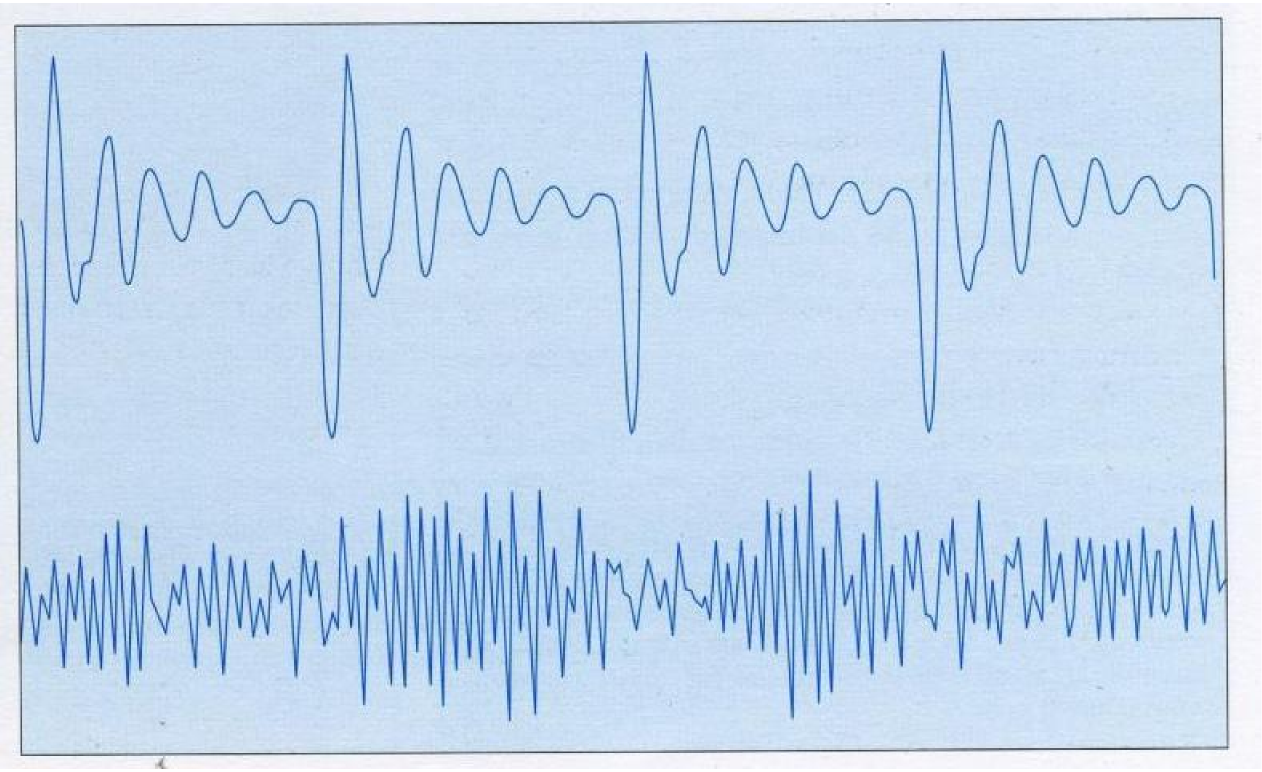
SOUND PROPERTIES

Aperiodic sound

A sound with no regular repeating pattern. (Consonants,sh,p)

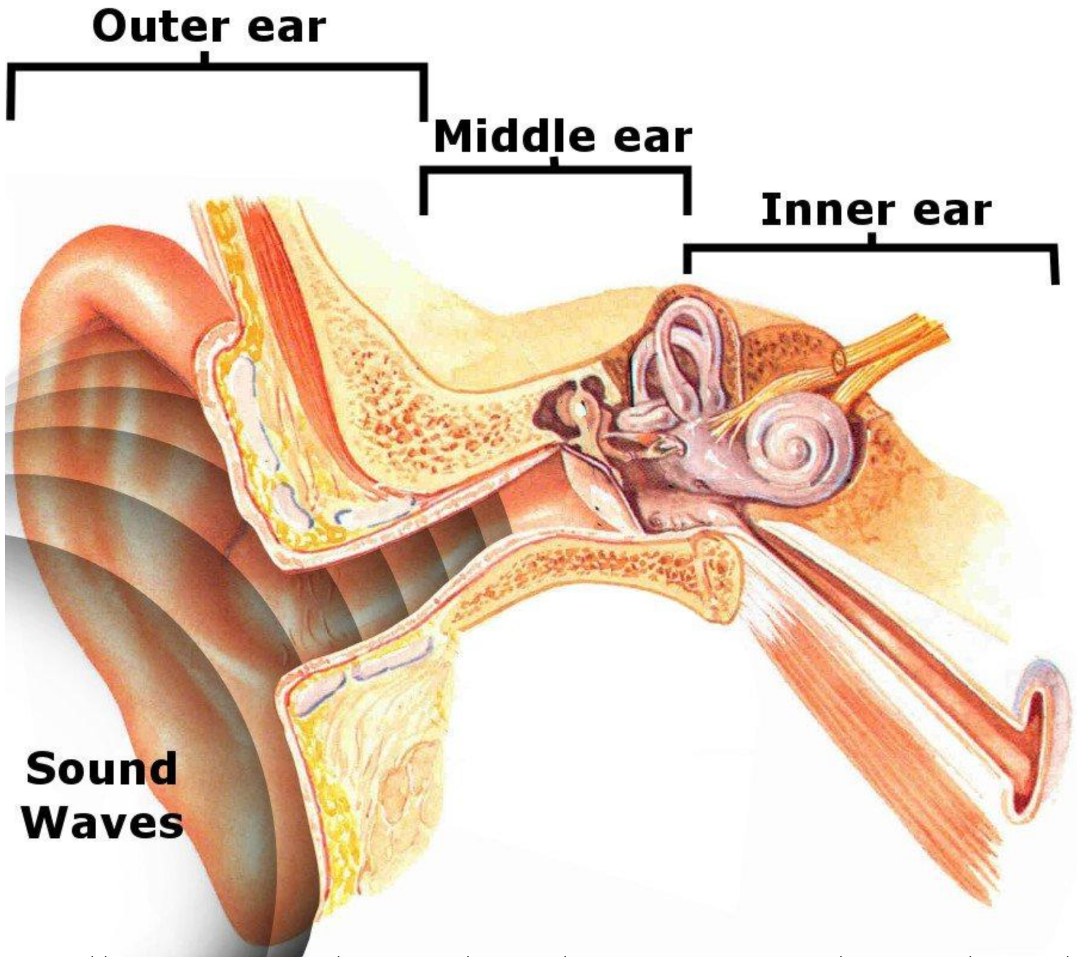
Periodic sound

A wound with regular repeating pattern. (Vowels,singing)



https://image2.slideserve.com/4466005/waveforms-of-the-vowel-a-and-the-consonant-s-l.jpg

FROM SOUND TO THOUGHT



https://www.scienceabc.com/wp-content/uploads/ext-www.scienceabc.com/wp-content/uploads/2017/08/Ear-compartments.jpg-.jpg

耳蜗

Figure 3.2 The Outer, Middle, and Inner Ear

The outer ear collects sound, the middle ear amplifies it, and the inner ear (cochlea) converts the sound waves into neural impulses that are sent to the brain via the auditory nerve.

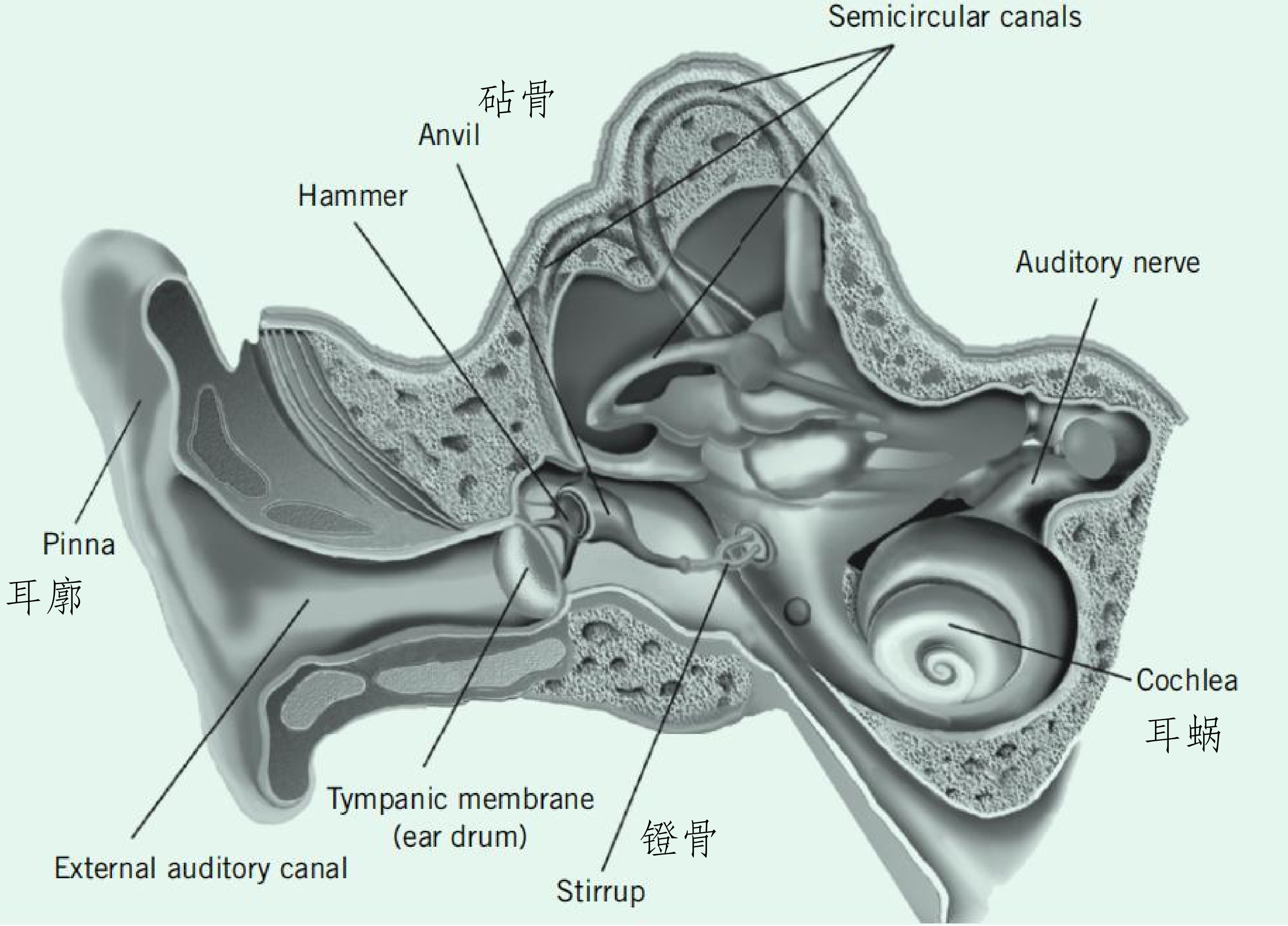
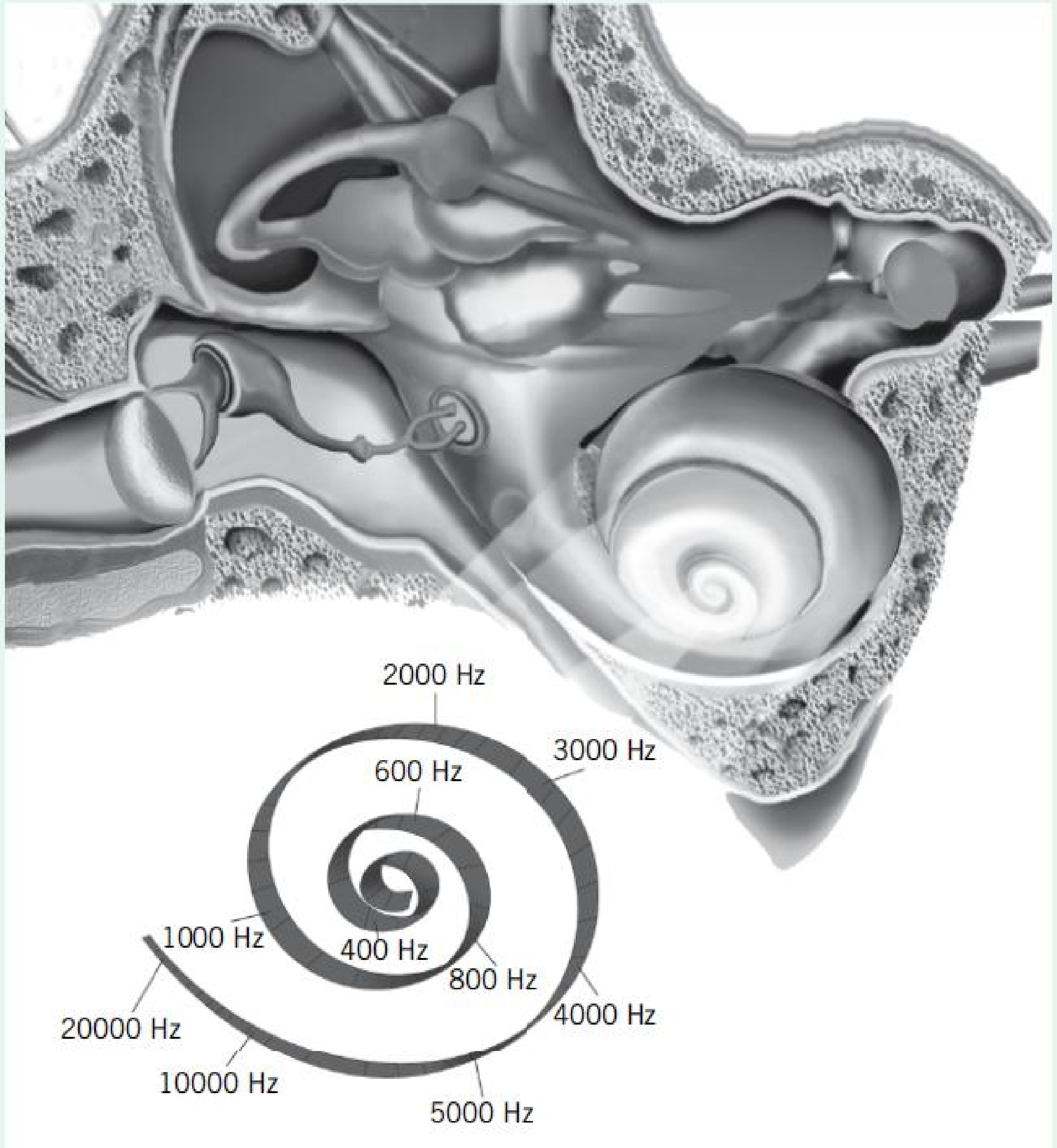
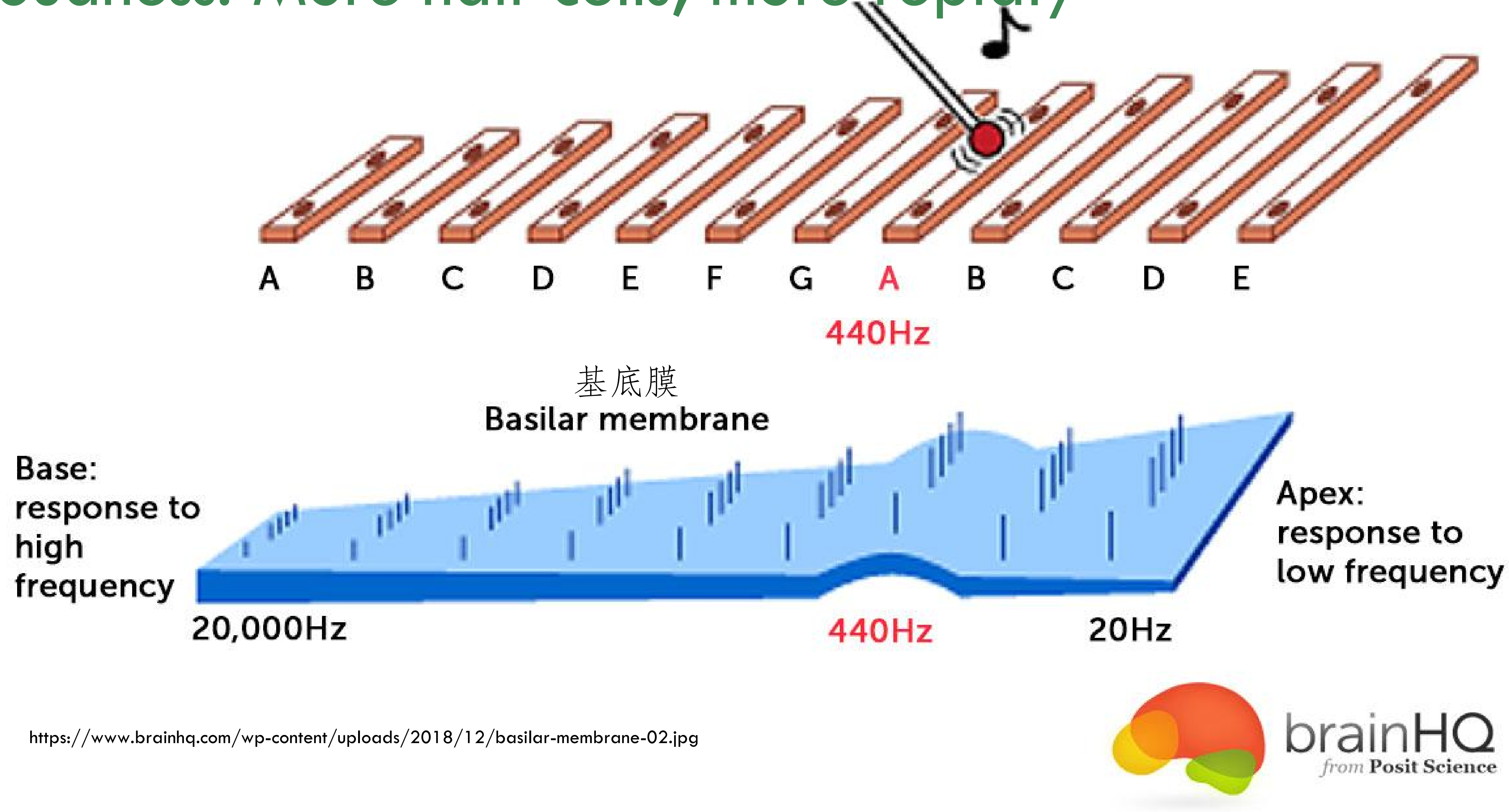


Figure 3.3 Tonotopic Organization of the Basilar Membrane

Hair cells along the basilar membrane are organized according to the frequency they are sensitive to, with the highest frequency positioned at the opening of the cochlea.

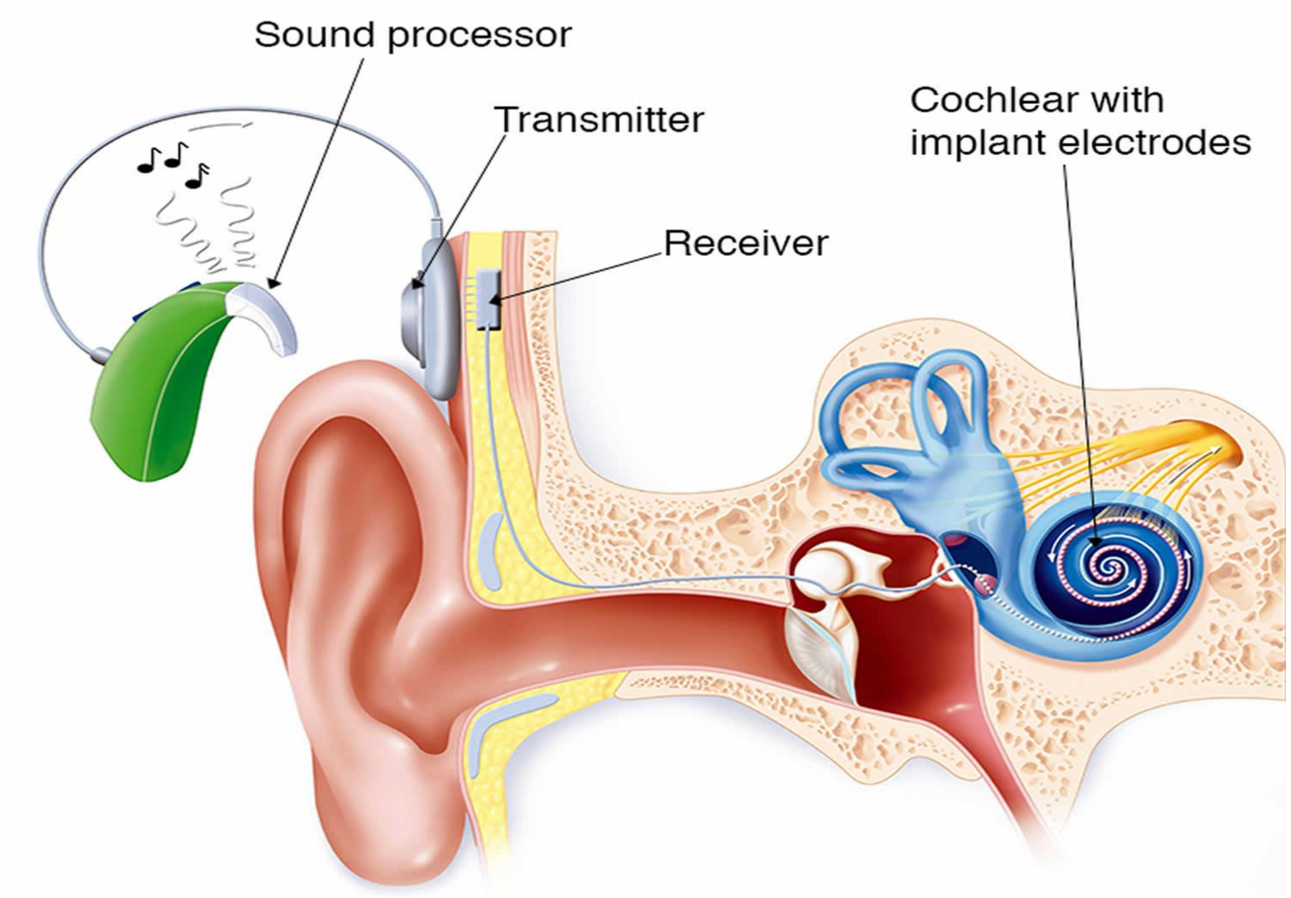


Loudness: More hair cells; more repidly



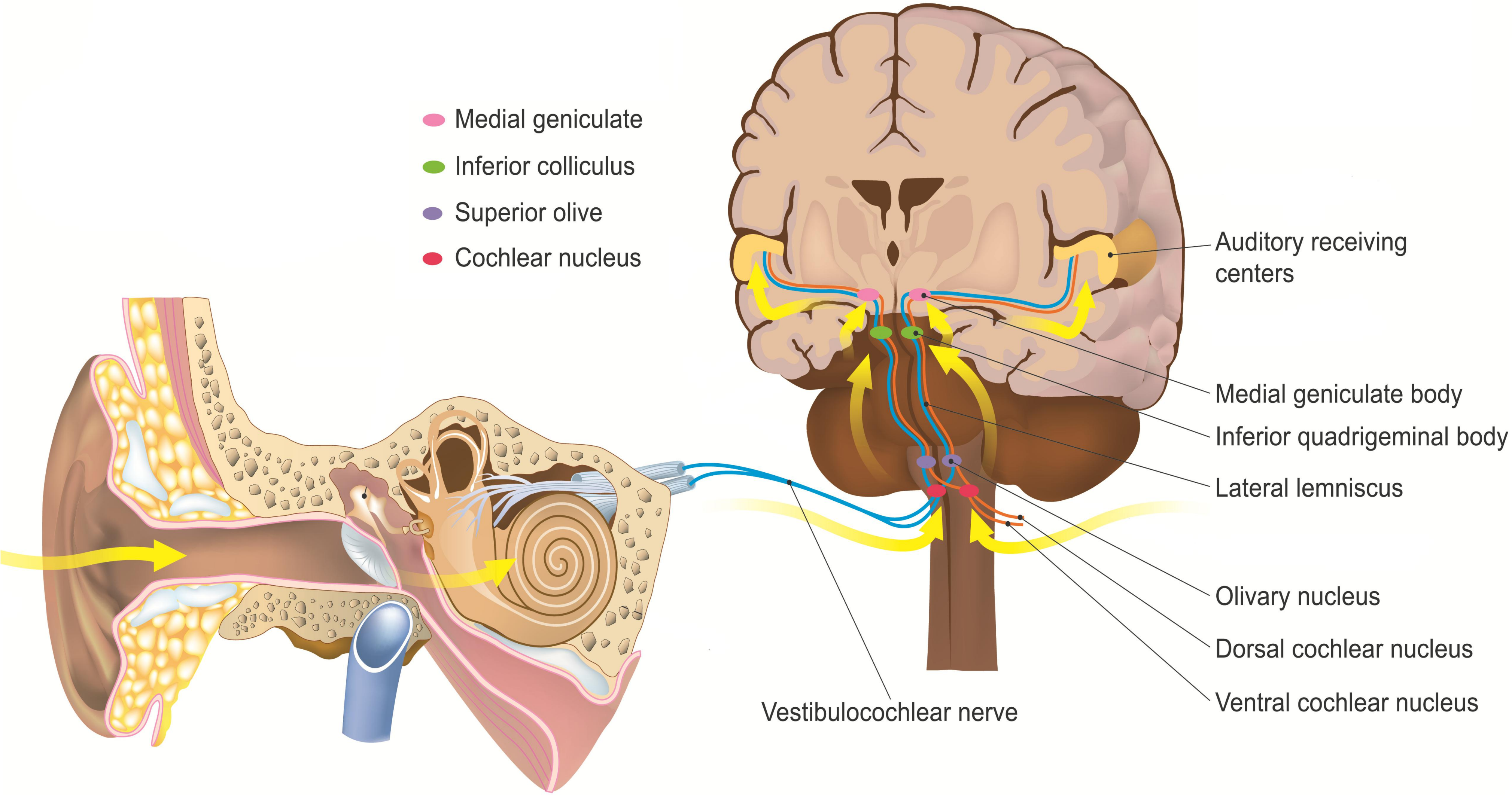
https://www.brainhq.com/wp-content/uploads/2018/12/basilar-membrane-02.jpg

COCHLEA IMPLANT



https://healthjade.com/wp-content/uploads/2018/09/cochlear-implant.jpg

AUDITORY PERCEPTION IN THE BRAIN



https://teachmeanatomy.info/wp-content/uploads/The-Cochlear-Nuclei-of-the-Auditory-Path

AUDITORY PERCEPTION IN THE BRAIN

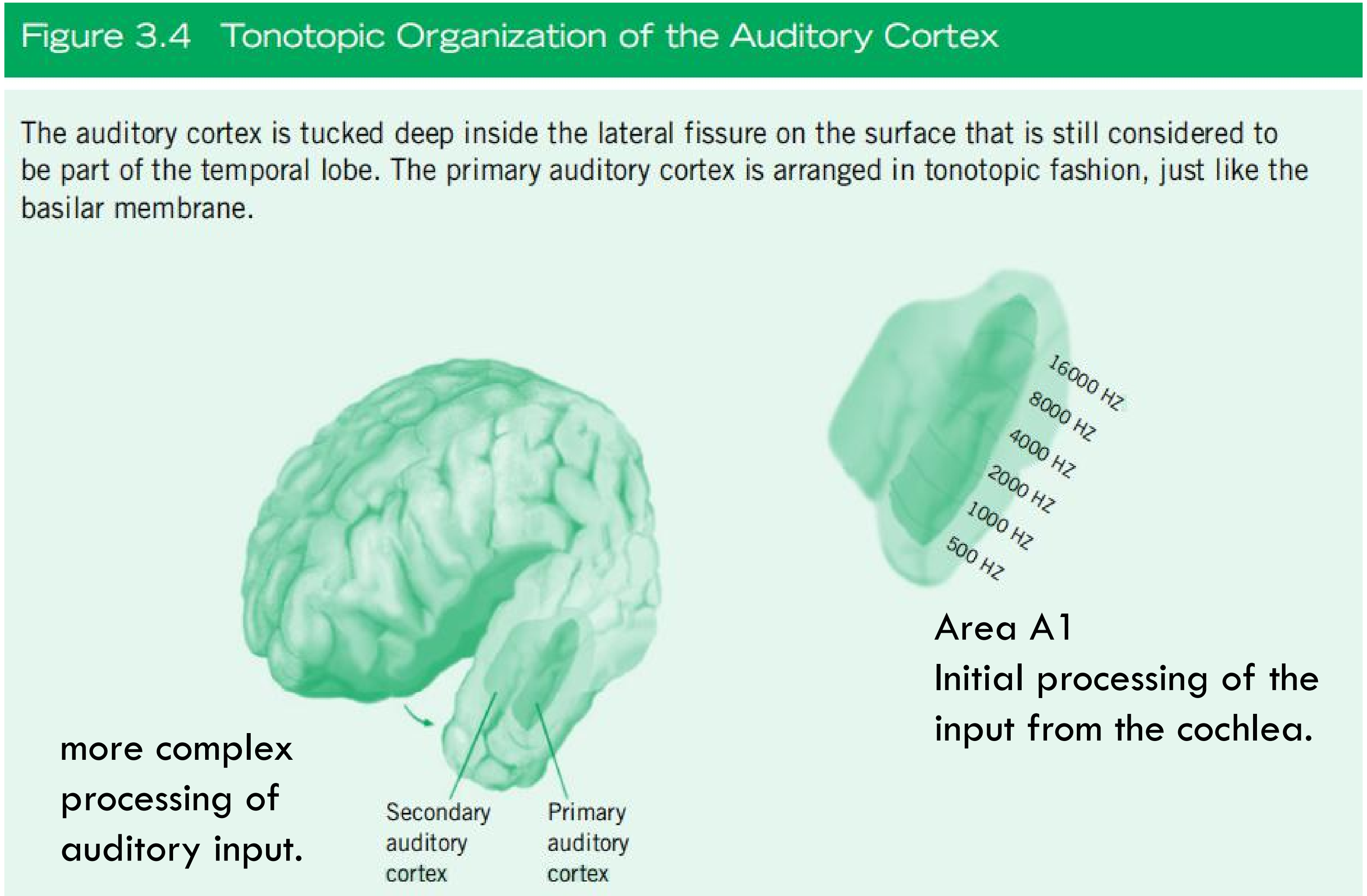
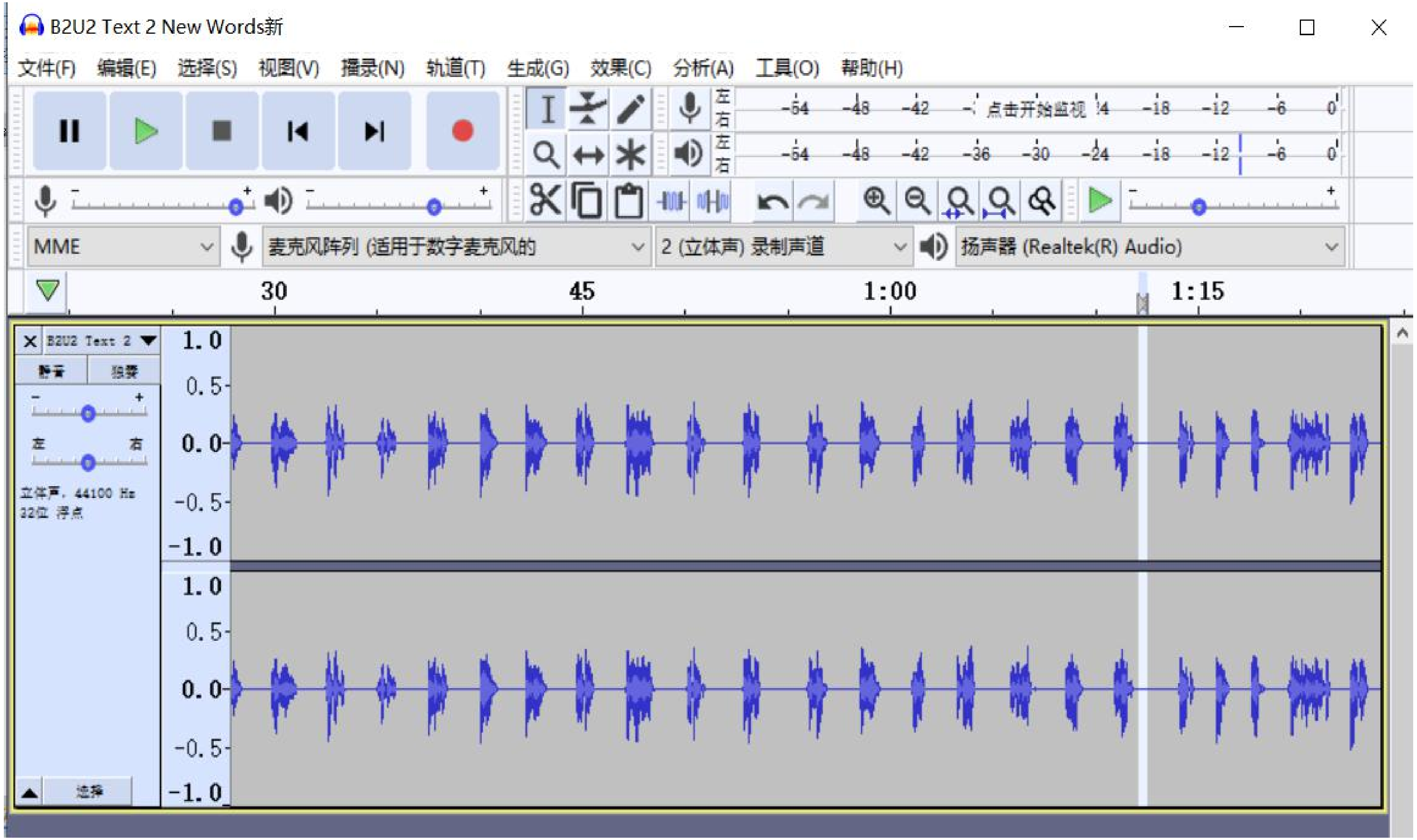


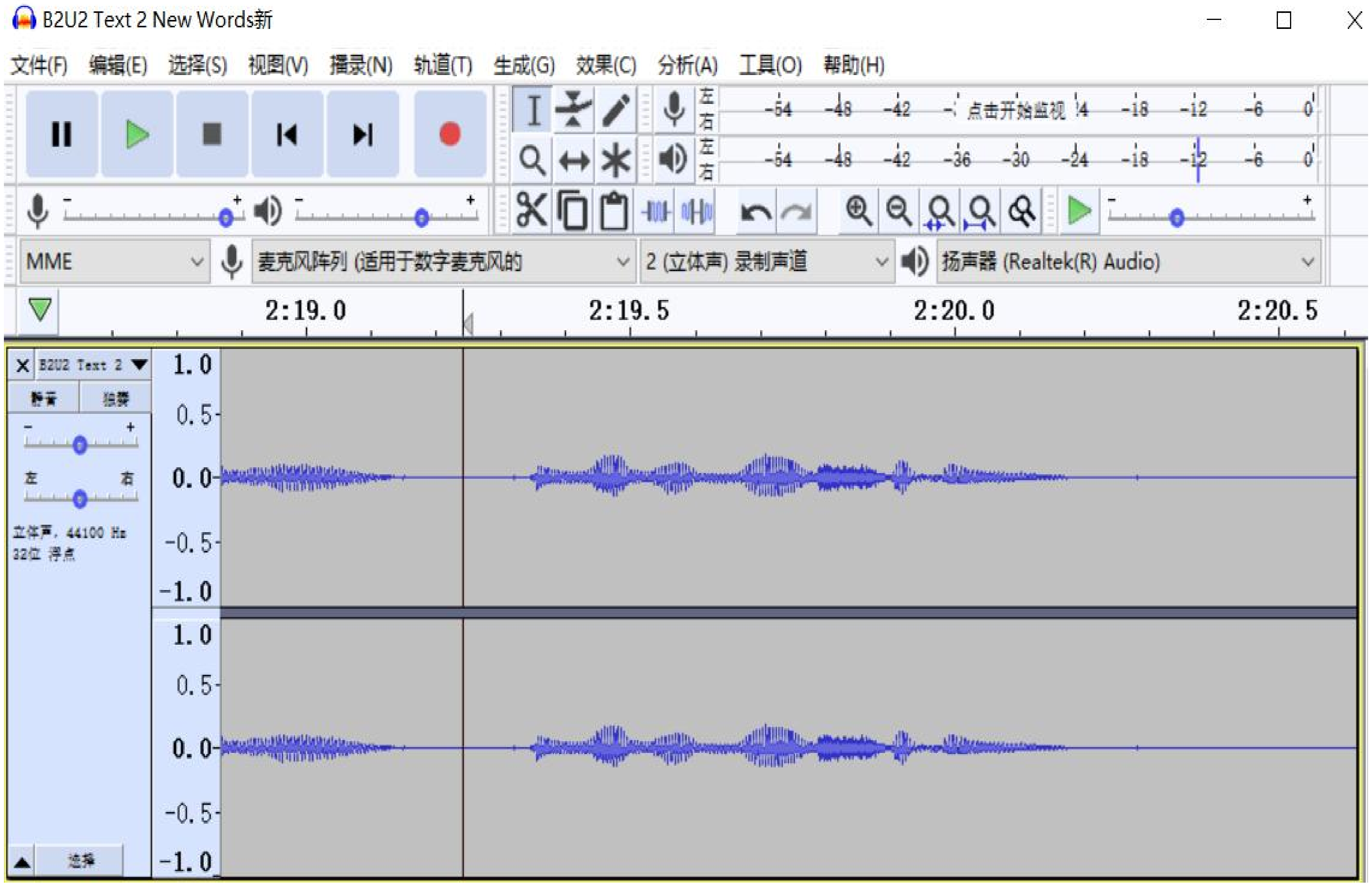
Figure 3.4 Tonotopic Organization of the Auditory Cortex

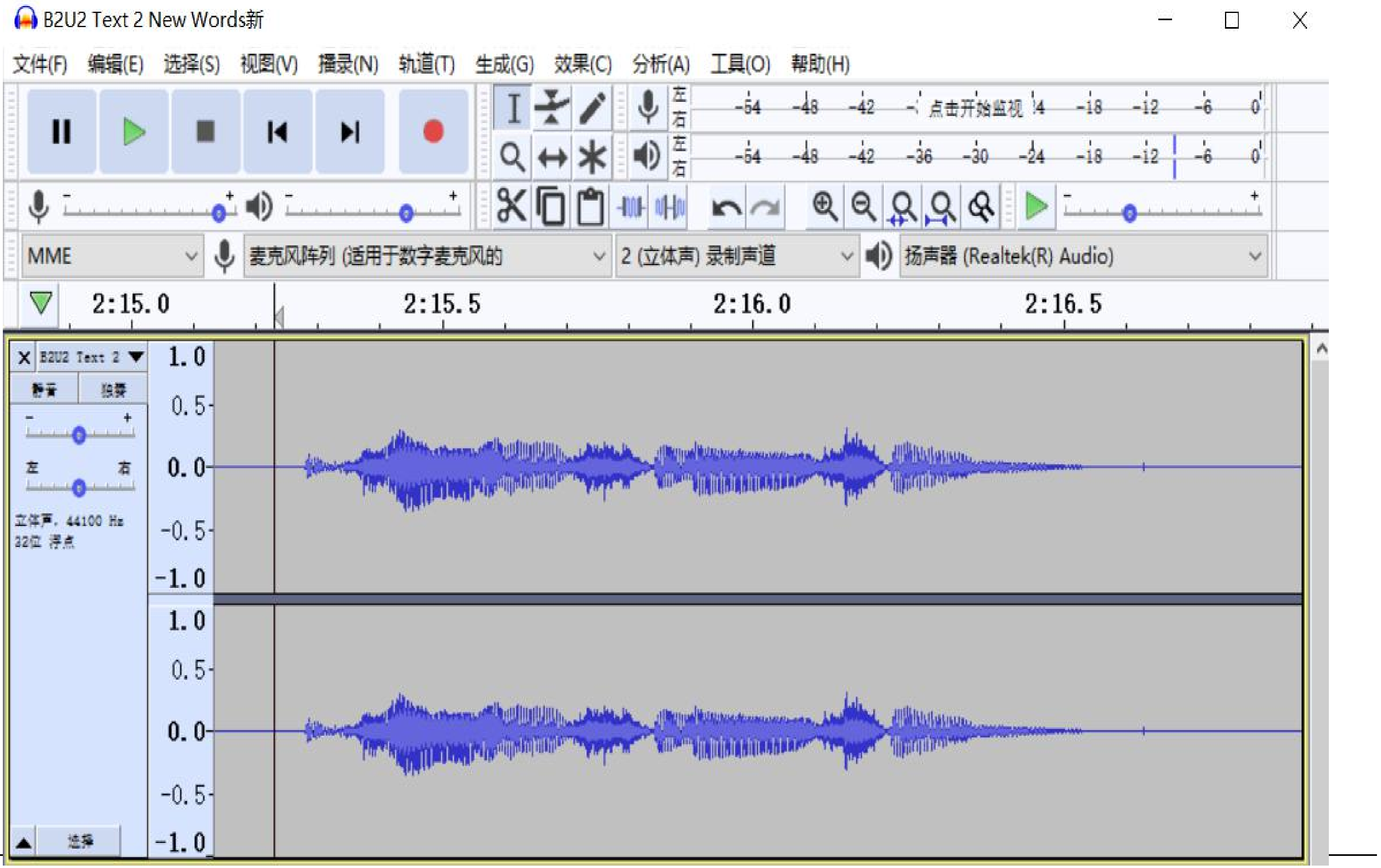
The auditory cortex is tucked deep inside the lateral fissure on the surface that is still considered to be part of the temporal lobe. The primary auditory cortex is arranged in tonotopic fashion, just like the basilar membrane.

Area A1
Initial processing of the input from the cochlea.

more complex processing of auditory input.

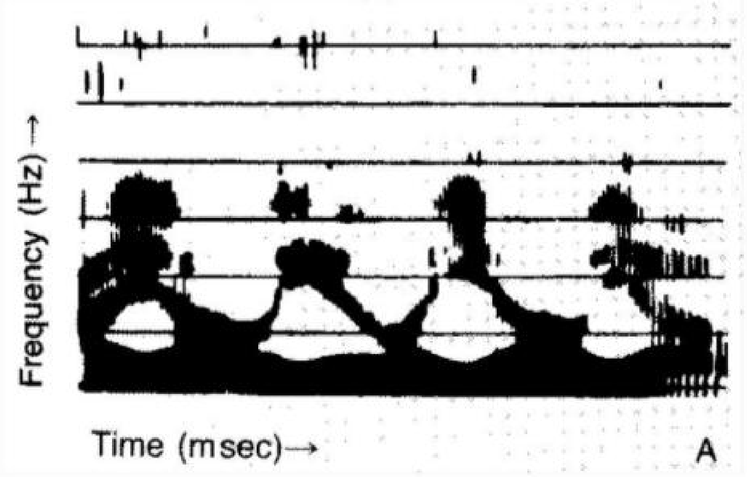

B2U2 Text 2 New Words新
文件(F) 编辑(E) 选择(S) 视图(V) 播录(N) 轨道(T) 生成(G) 效果(C) 分析(A) 工具(O) 帮助(H)
点击开始监视
MME
麦克风阵列 (适用于数字麦克风的
2 (立体声) 录制声道
扬声器 (Realtek(R) Audio)
30
45
1:00
1:15
B2U2 Text 2
静音
独奏
立体声, 44100 Hz
32位 浮点
选择
1.0
0.5
0.0
-0.5
-1.0


B2U2 Text 2 New Words新
文件(F) 编辑(E) 选择(S) 视图(V) 播录(N) 轨道(T) 生成(G) 效果(C) 分析(A) 工具(O) 帮助(H)
点击开始监视
MME
麦克风阵列 (适用于数字麦克风的
2 (立体声) 录制声道
扬声器 (Realtek(R) Audio)
2:19.0
2:19.5
2:20.0
2:20.5
B2U2 Text 2
静音
独奏
左
右
立体声, 44100 Hz
32位 浮点
选择
1.0
0.5
0.0
-0.5
-1.0


B2U2 Text 2 New Words新
文件(F) 编辑(E) 选择(S) 视图(V) 播录(N) 轨道(T) 生成(G) 效果(C) 分析(A) 工具(O) 帮助(H)
点击开始监视
MME
麦克风阵列 (适用于数字麦克风的
2 (立体声) 录制声道
扬声器 (Realtek(R) Audio)
2:15.0
2:15.5
2:16.0
2:16.5
B2U2 Text 2
静音
独奏
左
右
立体声, 44100 Hz
32位 浮点
选择
1.0
0.5
0.0
-0.5
-1.0

THE SPEECH STREAM

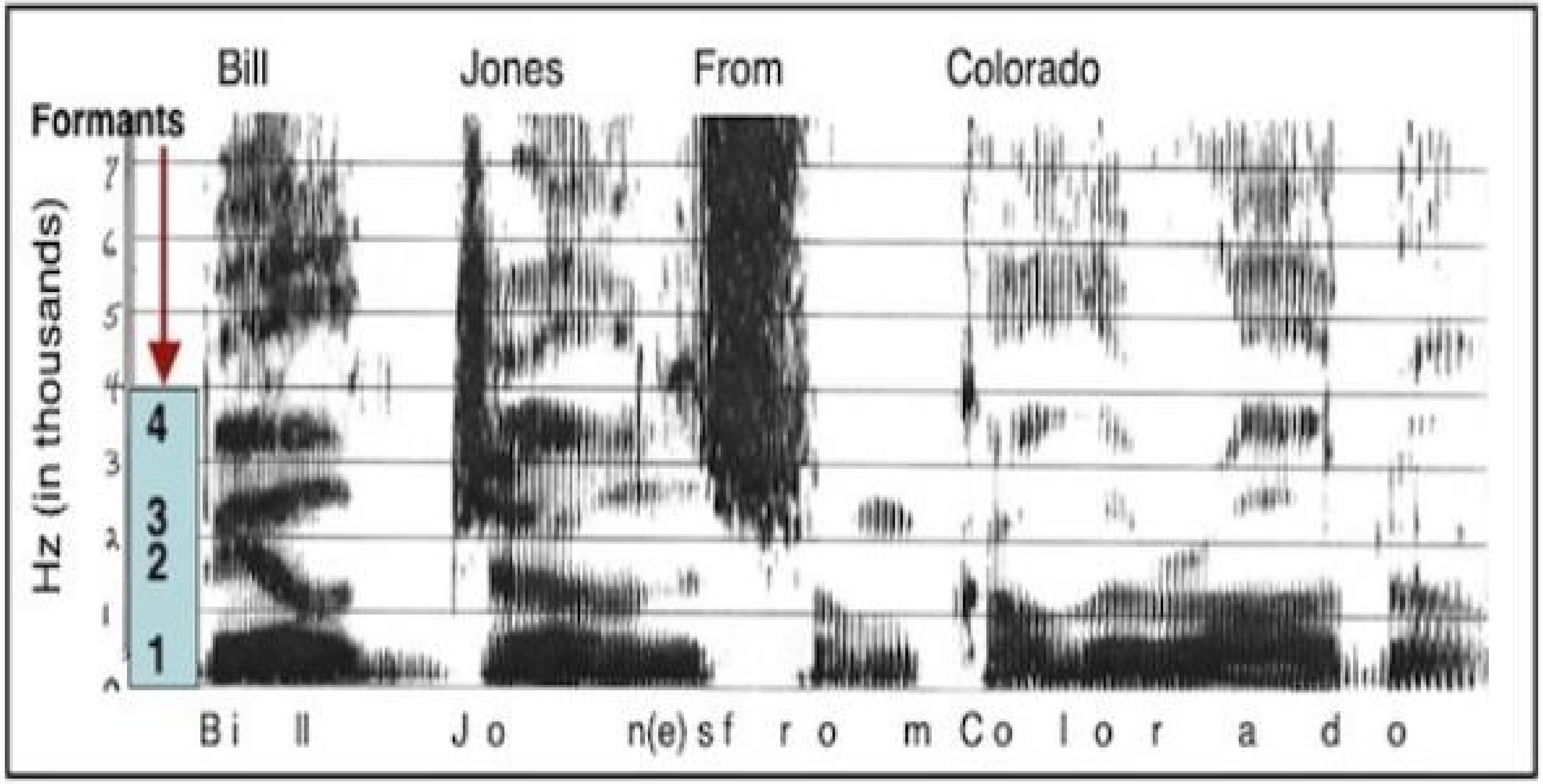
- Speech stream: a continuous flow of ever-changing frequencies and amplitudes
- Speech perception system: Infer intended phonemes and word boundaries on the basis of multiple cues within the speech stream.



I owe you a Yo-Yo.

https://image1.slideserve.com/2622360/perceiving-speech6-l.jpg

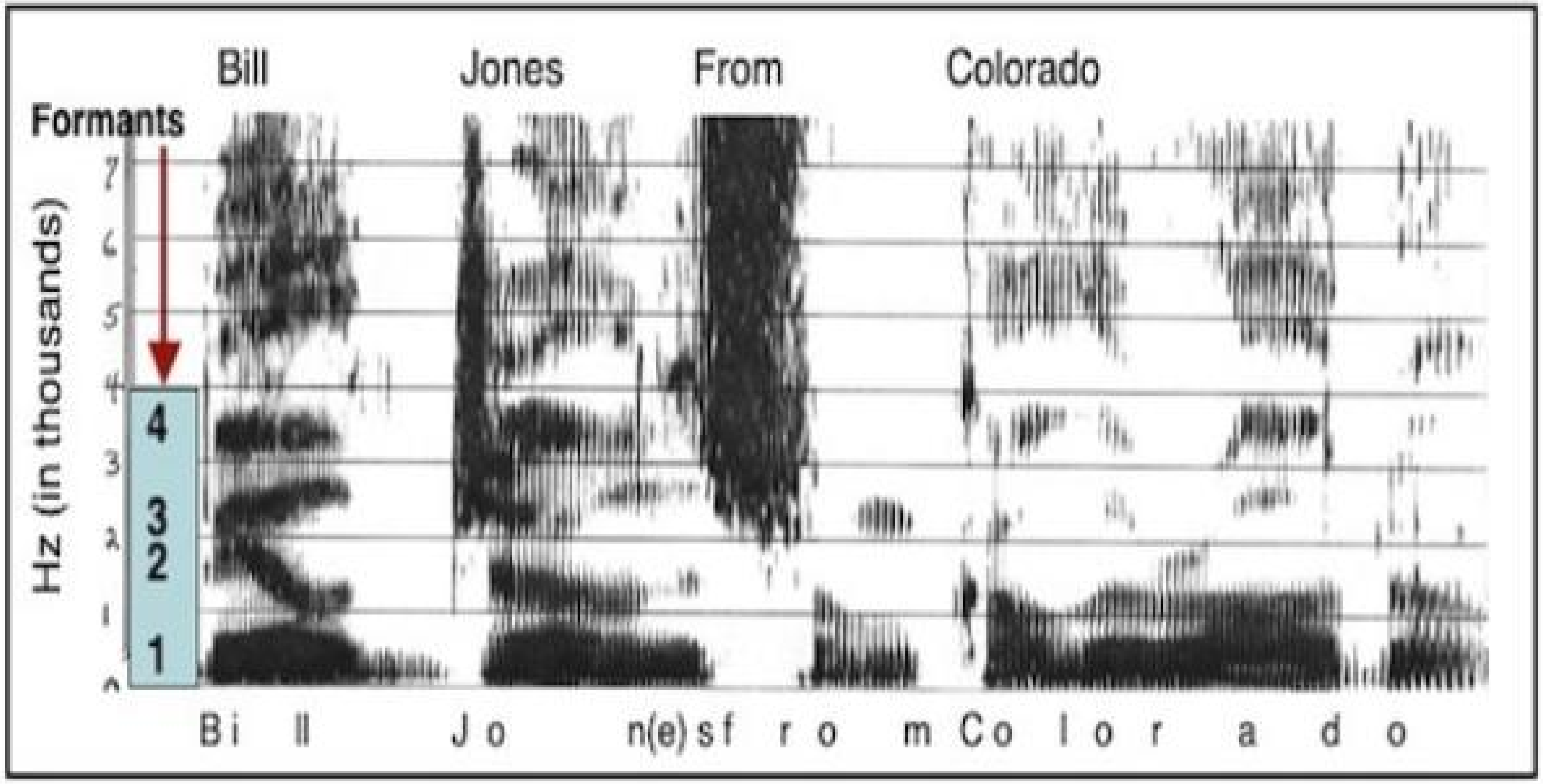
THE SPEECH STREAM



http://hearinghealthmatters.org/waynesworld/files/2014/08/Spectrogram.jpg

Spectrogram (语谱图): A chart displaying the pattern of frequencies in the speech stream and how those patterns change over time.

THE SPEECH STREAM



http://hearinghealthmatters.org/waynesworld/files/2014/08/Spectrogram.jpg

Fundamental frequency:

Male: 75-150 Hz

Female: 150-300 Hz

Prosody (韵律):

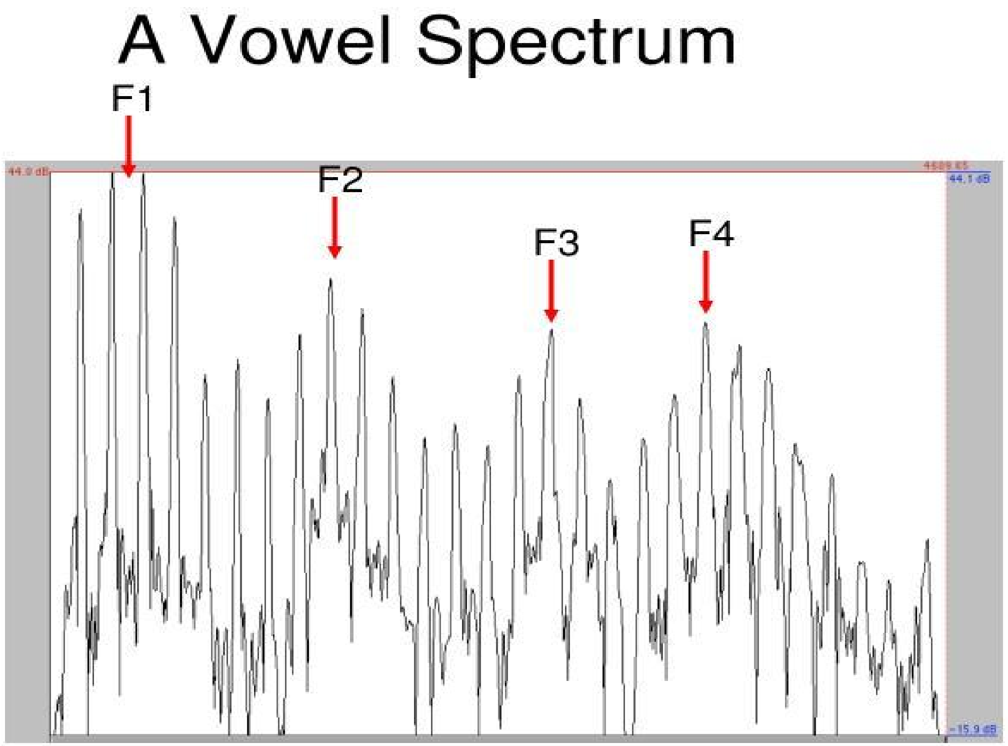
Fluctuations of the fundamental frequency during an utterance. (Variation in pitch)

Intonation (语调);

Rhythm (节奏);

Stress (中银)

THE SPEECH STREAM



Note:

F0 ≈ 160 Hz

https://image2.slideserve.com/3741088/a-vowel-spectrum-n.jpg

Fundamental frequency:

Male: 75-150 Hz

Female: 150-300 Hz

Formant (共振峰):

Bands of high-amplitude sound at certain frequencies above the fundamental frequency.

THE SPEECH STREAM

Periodic speech stream：

Vowels；

Sonorant (响音）： a speech sound that usually serves as a consonant but sometines as a vowel, like, r, n, l, and m (little).

Aperiodic speech stream：

Fricative(摩擦音): A consonant that is produced by consticting the airstream to creat friction (s, sh, f).

Plosive (爆破音): A consonant that is produced by momentarily blocking and the releasing the airstream (b, p, d, t, g, k).

THE SPEECH STREAM

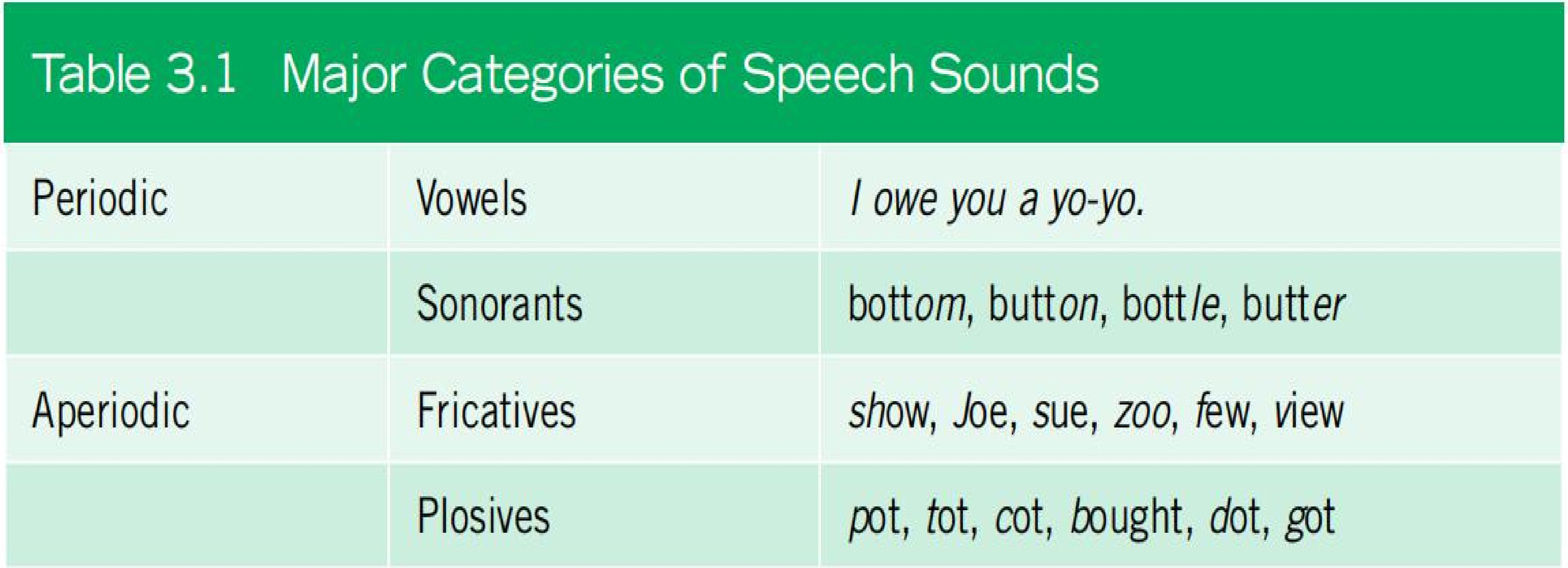


Table 3.1 Major Categories of Speech Sounds

Periodic	Vowels	*I owe you a yo-yo.*
	Sonorants	bott*om*, butt*on*, bott*le*, butt*er*
Aperiodic	Fricatives	*sh*ow, *J*oe, sue, *z*oo, *f*ew, *v*iew
	Plosives	*p*ot, *t*ot, cot, *b*ought, *d*ot, got

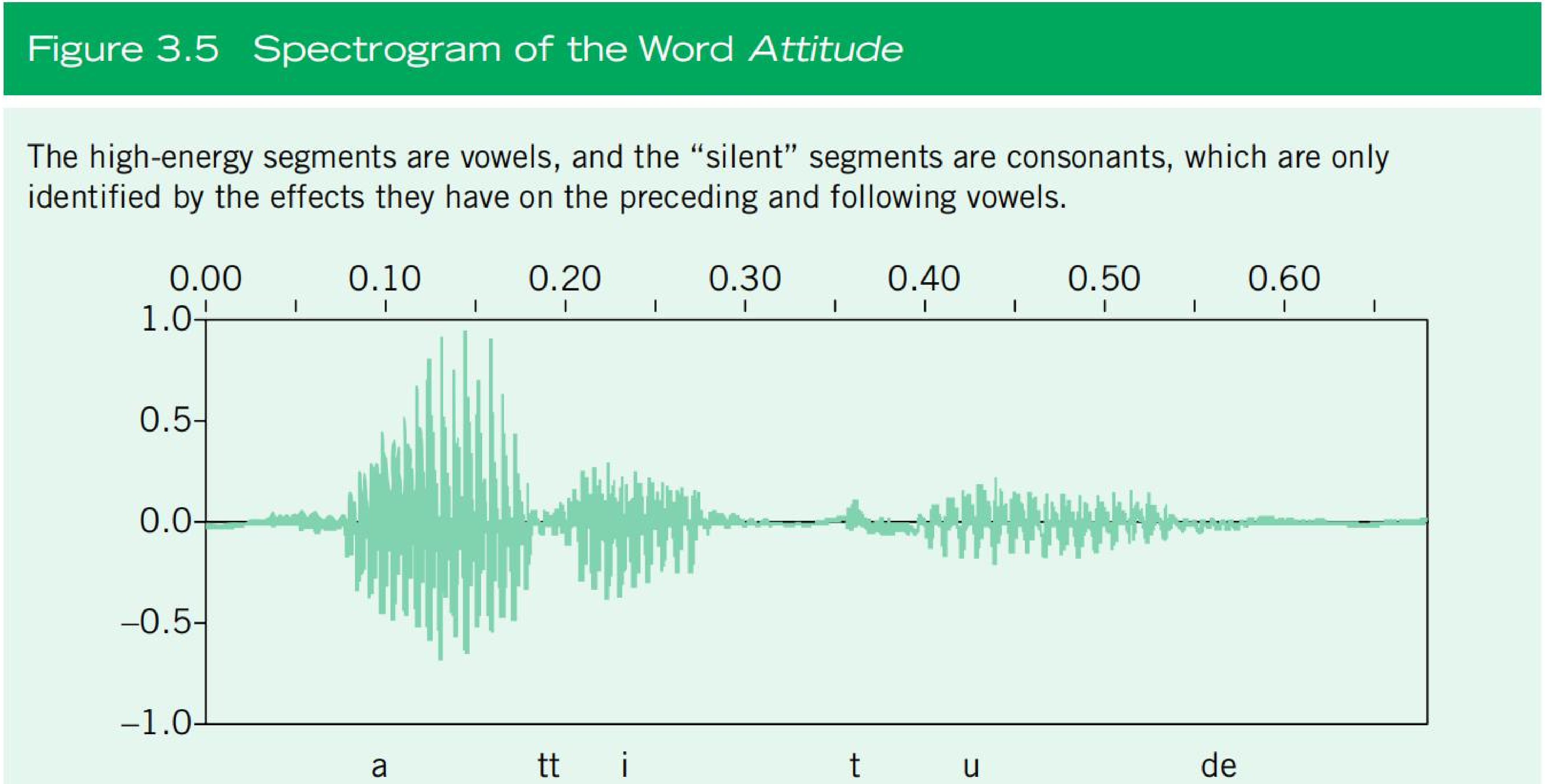


Figure 3.5 Spectrogram of the Word *Attitude*

The high-energy segments are vowels, and the “silent” segments are consonants, which are only identified by the effects they have on the preceding and following vowels.

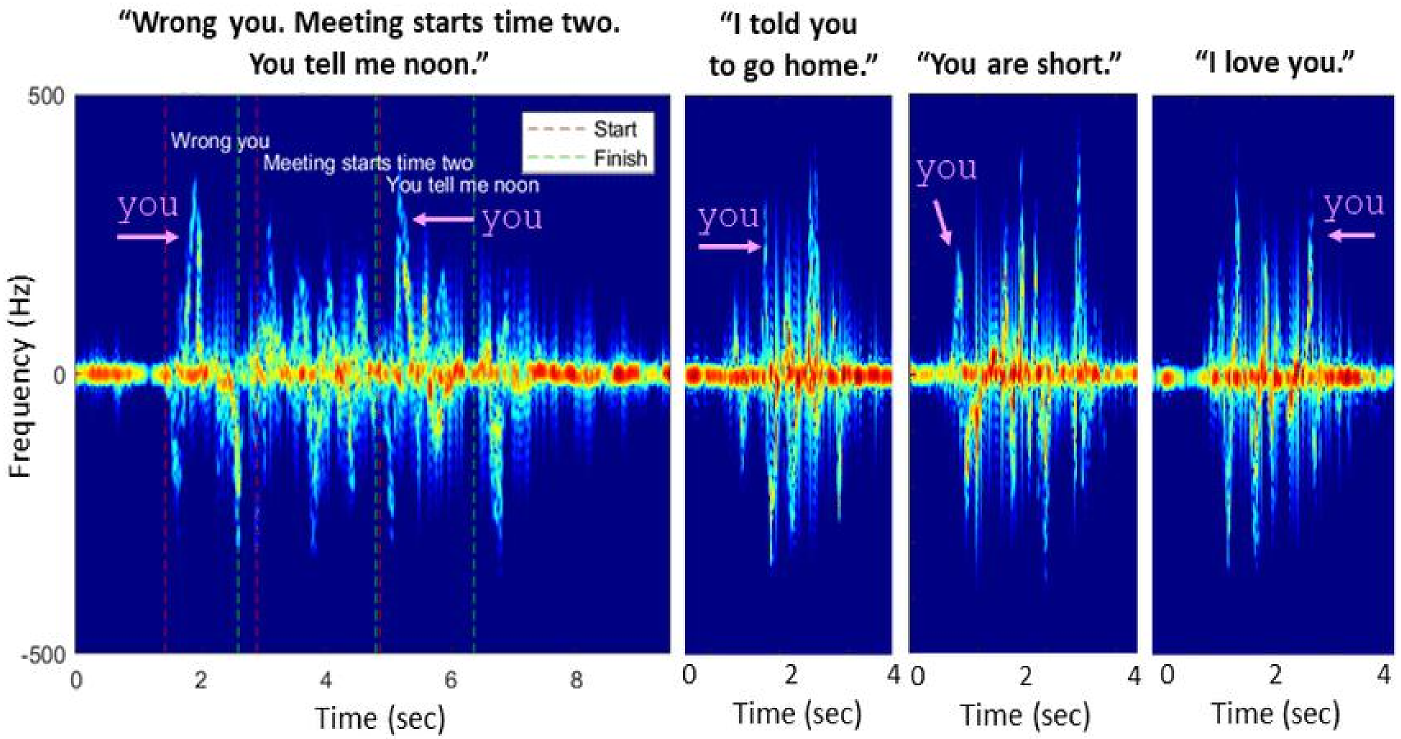
Formant transition （共振峰迁移）：

A modification of a formant due to a preceding or following consonant.

THE SOUND OF SILENCE

Coarticulation (协同发音):

The process of overlapping phonemes in the speech stream.



"Wrong you. Meeting starts time two. You tell me noon."

"I told you to go home."

"You are short."

"I love you."

THE SOUND OF SILENCE

Aspiration (送气)：
The puff of air accompanying the release of some plosives.
bah vs. pah
Phonation： bah
Aspiration: pah

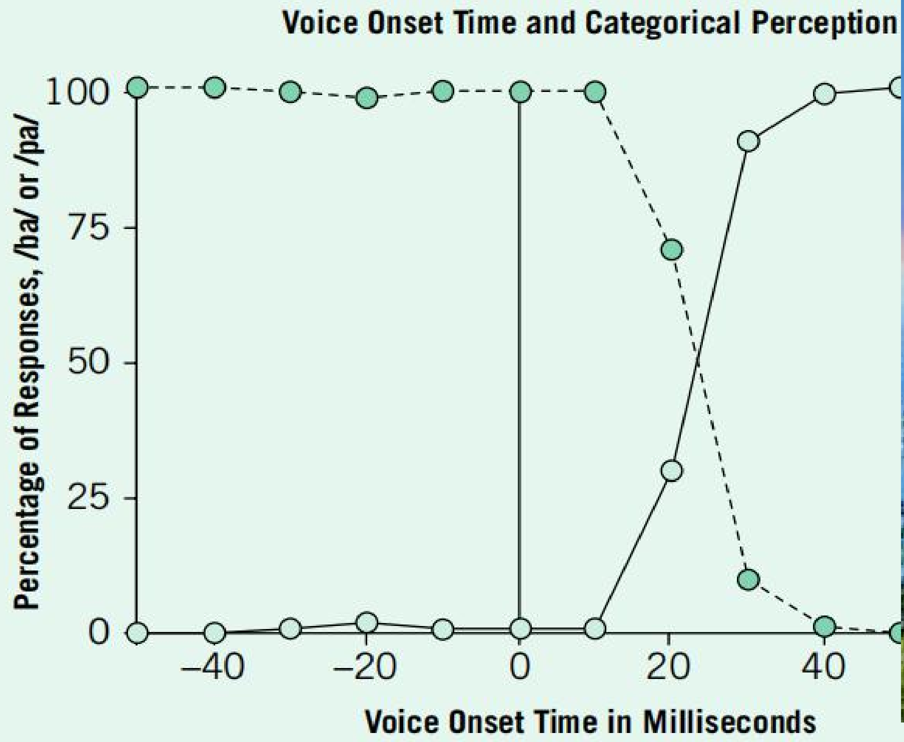
VOT (voice onset time): the differece in time between the release of a plosive consonant and the beginning of vocal fold vibration.

Categorical Perception:

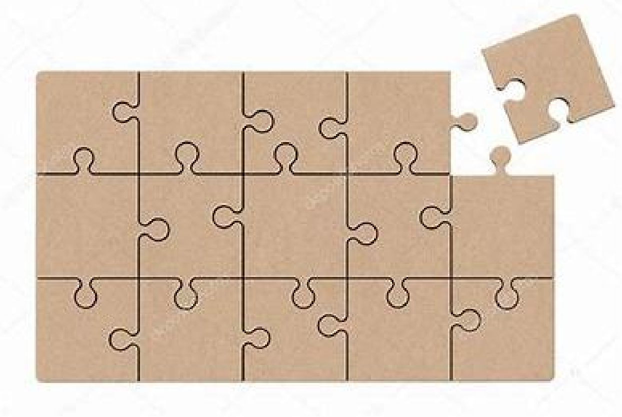
The process of experiencing continuous changing stimuli as belonging to two or more discrete sets.

Figure 3.6 Voice Onset Time

When synthetic speech is used to create a continuum of voice onset times ranging from 0 to 50 milliseconds, native speakers of English do not perceive a series of syllables from *pa* to *ba*. Instead, they perceive a series of *pa* syllables followed by a se

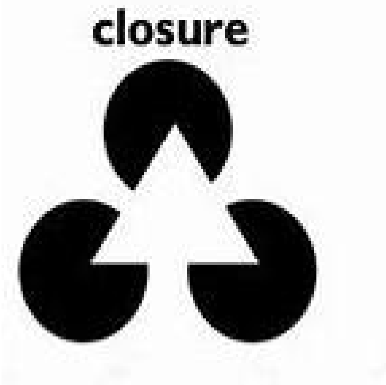


PHONEMIC RESTORATION



The process of filling in missing segments of the speech stream with contextually appropriate material.

Warren, 1970
The state governnors met with their respective legi*latures convening in the capital city.



Warren & Warren, 1970
It was found that the *eel was on the axle.
It was found that the *eel was on the shoe.

READ MY LIPS

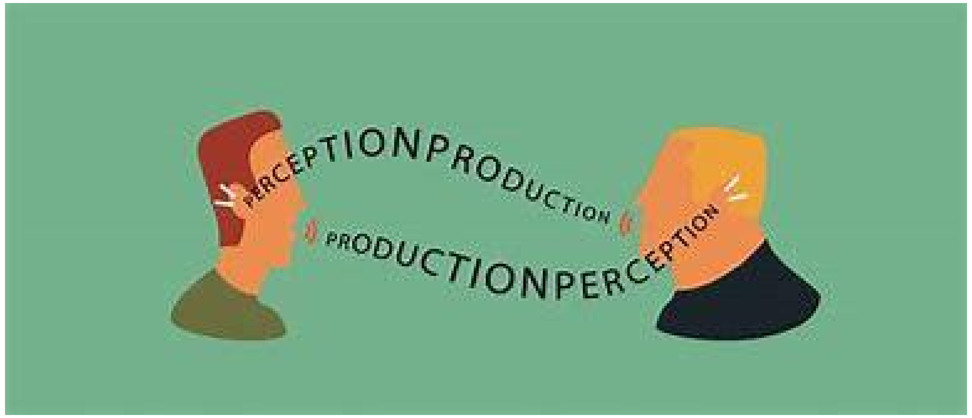
McGurk Effect (McGurk & MacDonald, 1976):
A perceptual phenomenon that demonstrates an interaction between hearing and vision in speech perception.

OPERATIONS IN SPEECH PERCEPTION

1. Analog acoutic patterns must be converted to digital codes at multiple levels of language-specific structure (phoneme, syllables, and words).
2. The categorization of speech signals should be sensitive to fine-grained cues and also flexible to accommodate the tremendous acoustic variability that exists across talkers.
3. The boundarities between words must be identified.
4. All the operations must be executed with breathtaking speed. (1--15 phonemes per second in casual speech)
5. Speech input must be routed to the grammatical and semantic systems and the motor system.

SPEECH PRODUCTION

- English speakers generate about two to three words per second.
- Average literate adults have a mental lexicon of 50,000-100,000 entries.
- Articulation involving 80 muscles.
- Few errors: once or twice every 1,000 words. Dimensions of vowel



THE LEMMA MODEL OF LEXICAL SELECTION AND FORM ENCODING

What to say

Willem (Pim) J. M. Levelt

The Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands

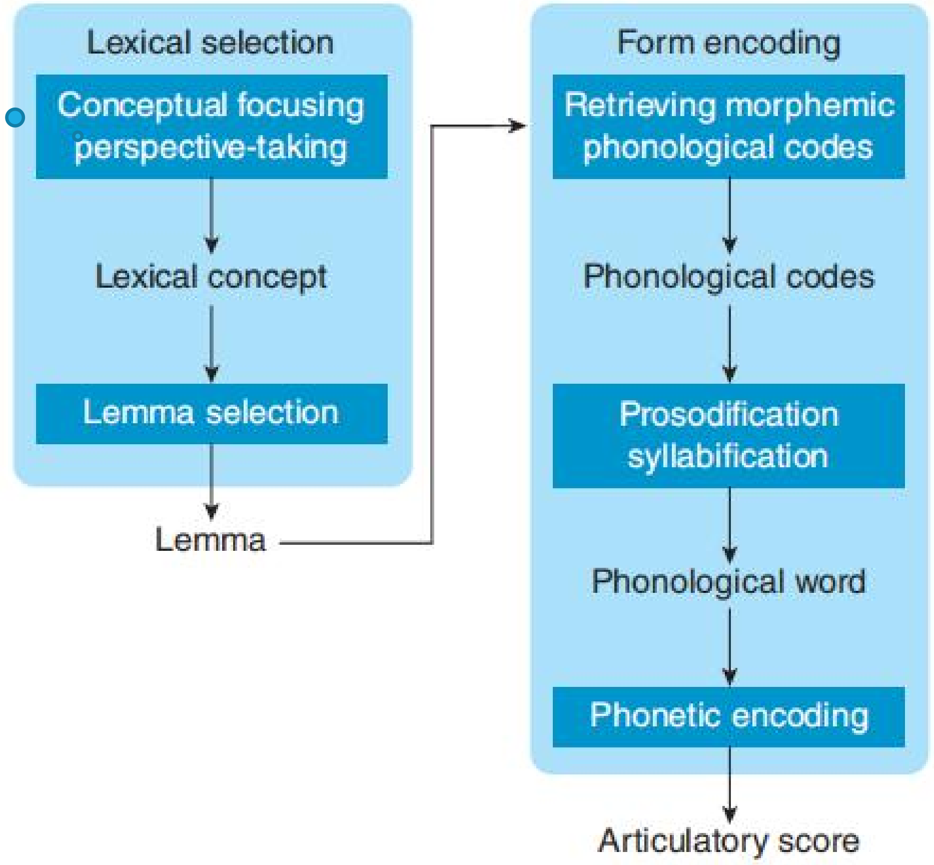
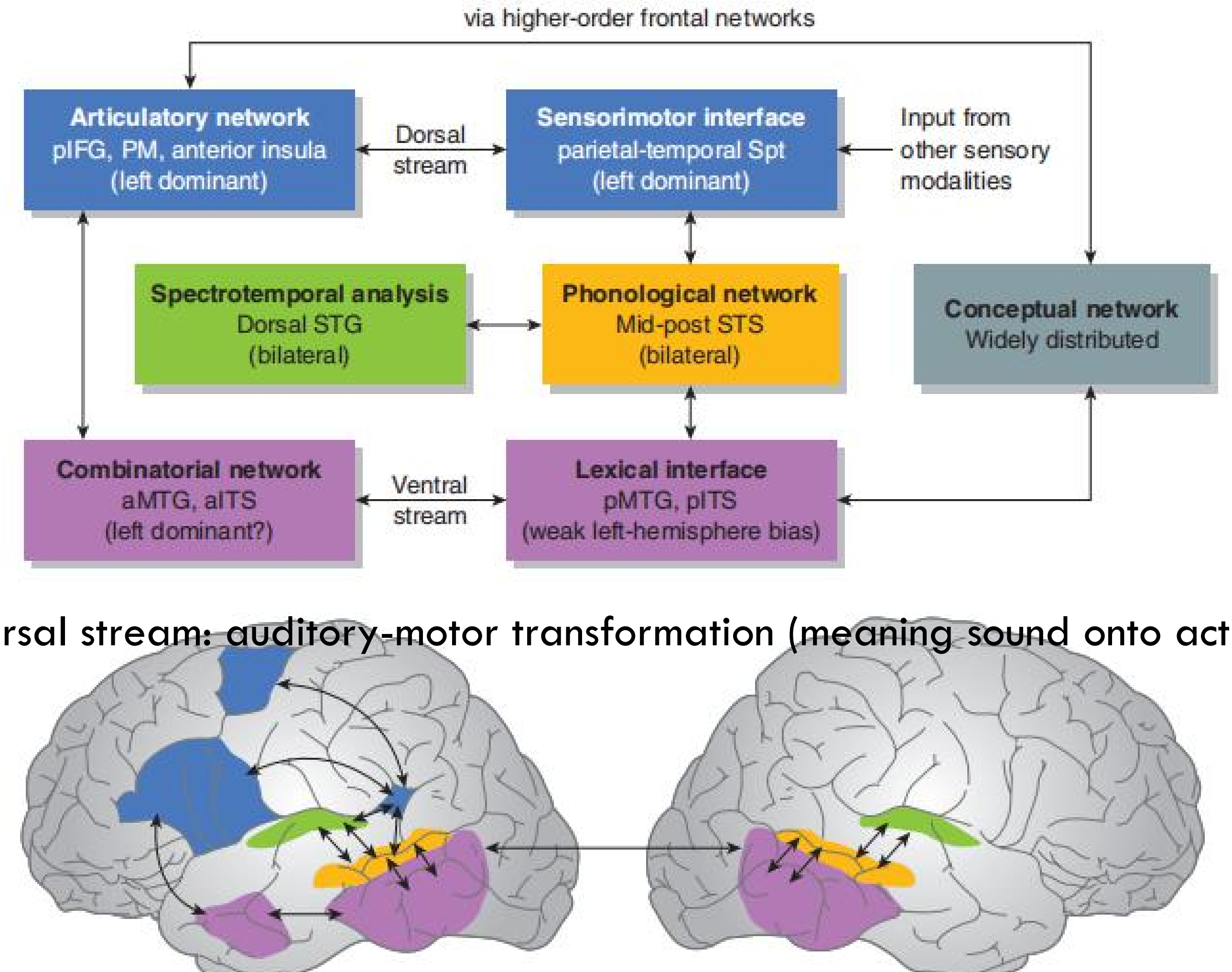


Figure 6.1 Serial two-system architecture of the Lemma Model: Two stages of lexical selection followed by three stages of form encoding. (From Levelt, 2001, p. 13465.) Copyright (2001) National Academy of Sciences, U.S.A.

DUAL STREAM MODEL (GREGORY KICKOK & DAVID POEPPEL)

Spectrotemporal analysis: The computation of which sound frequencies are modulated at which rates.



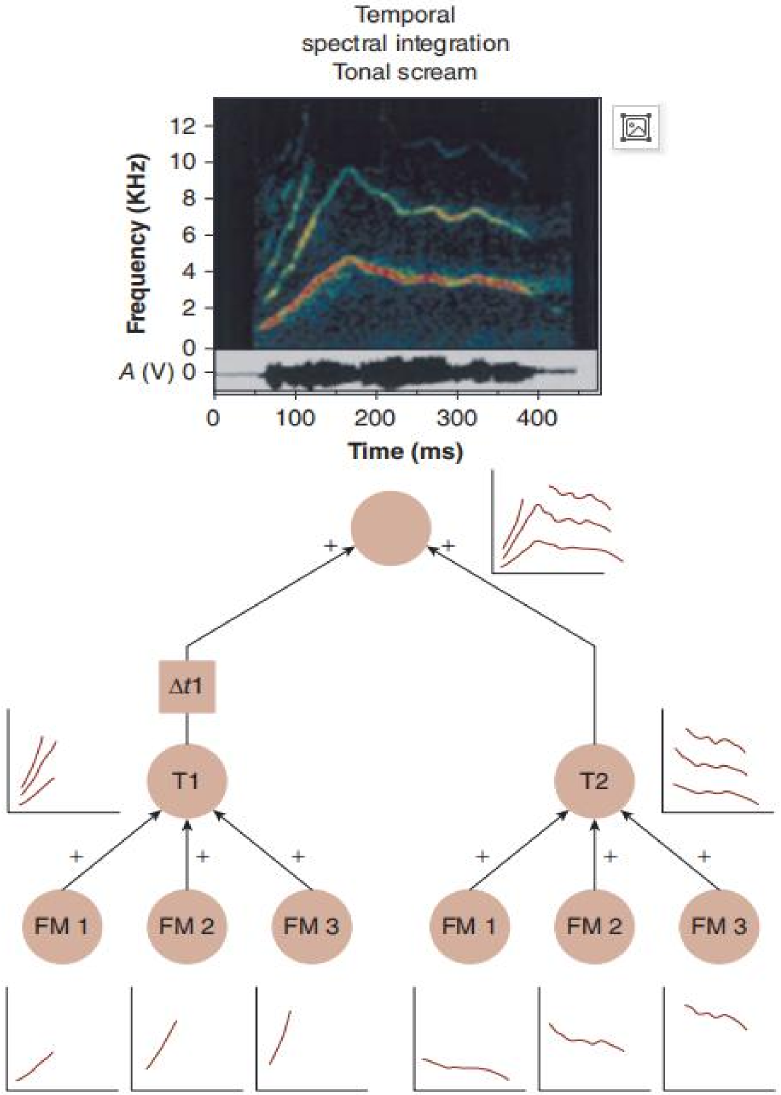
cat:/k/,/æ/,/t/
pat:/p/,/æ/,/t/

Dorsal stream: auditory-motor transformation (meaning sound onto action)

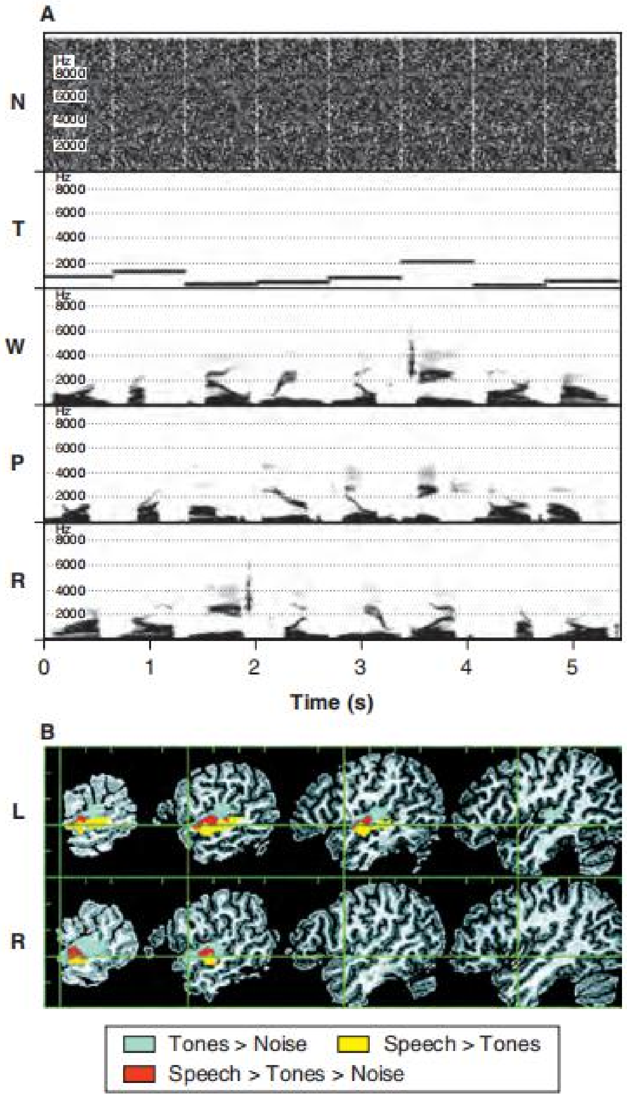
Ventral stream: comprehension (mapping sound onto meaning)

Figure 5.1 The Dual Stream Model of speech perception. (From Hickok & Poeppel, 2007, p. 395.)

HIERARCHICAL ORGANIZATION FOR SPEECH PERCEPTION



BOTH HEMISPHERES CONTRIBUTE TO SPEECH PERCEPTION



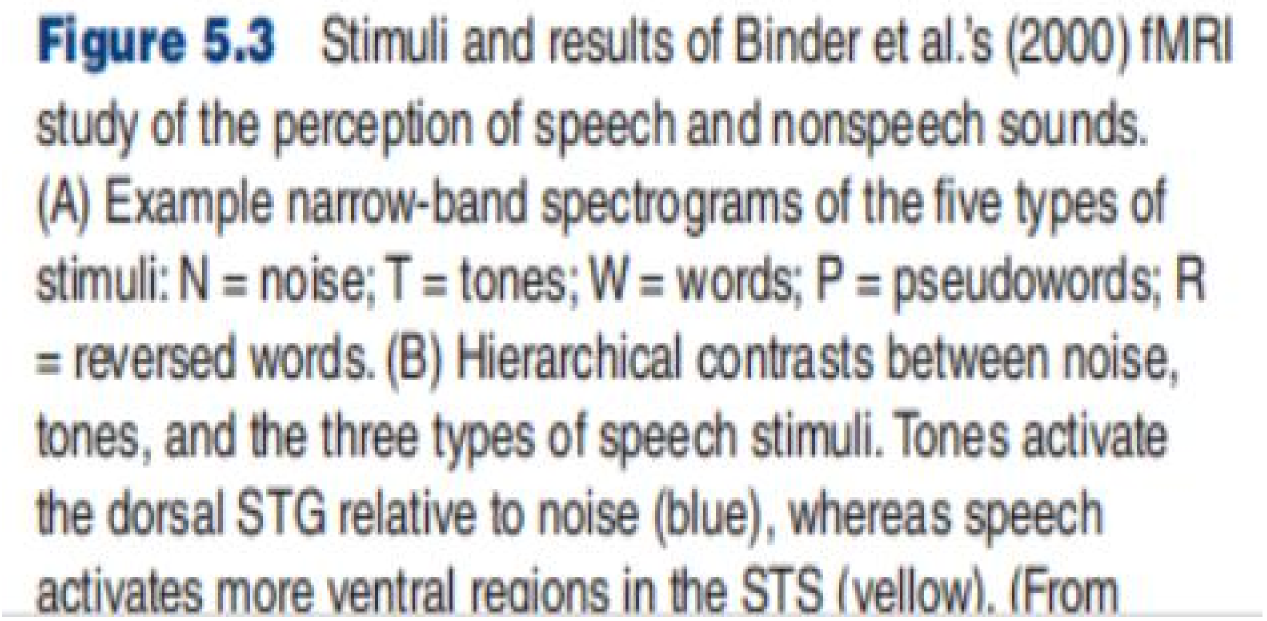
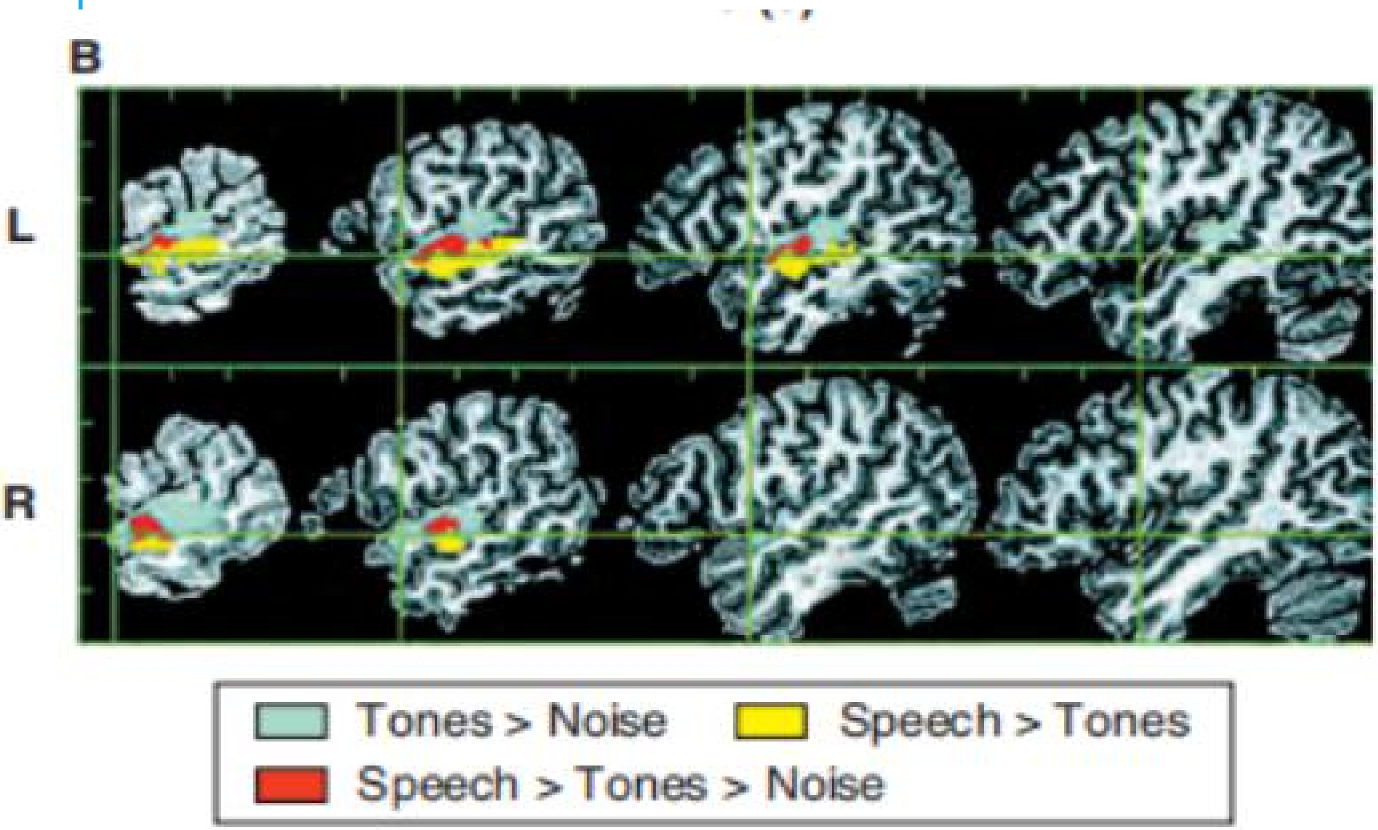


Figure 5.3 Stimuli and results of Binder et al.'s (2000) fMRI study of the perception of speech and nonspeech sounds. (A) Example narrow-band spectrograms of the five types of stimuli: N = noise; T = tones; W = words; P = pseudowords; R = reversed words. (B) Hierarchical contrasts between noise, tones, and the three types of speech stimuli. Tones activate the dorsal STG relative to noise (blue), whereas speech activates more ventral regions in the STS (yellow). (From

BOTH HEMISPHERES CONTRIBUTE TO SPEECH PERCEPTION



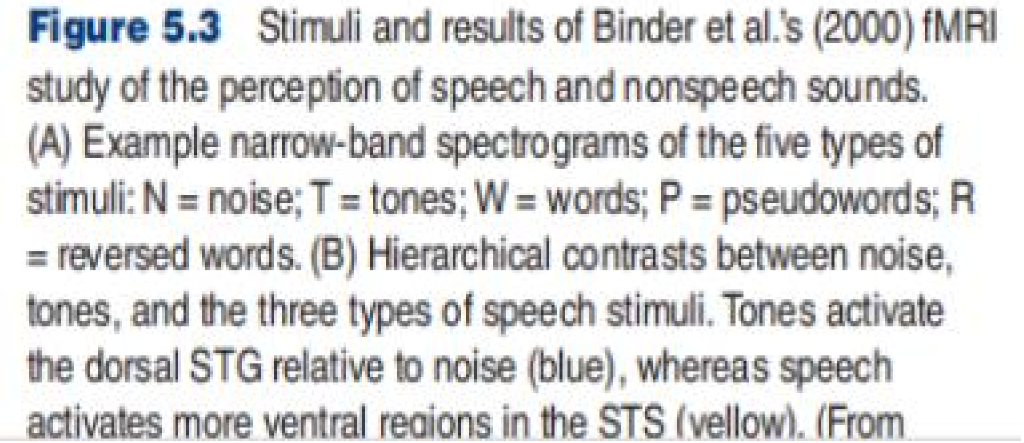
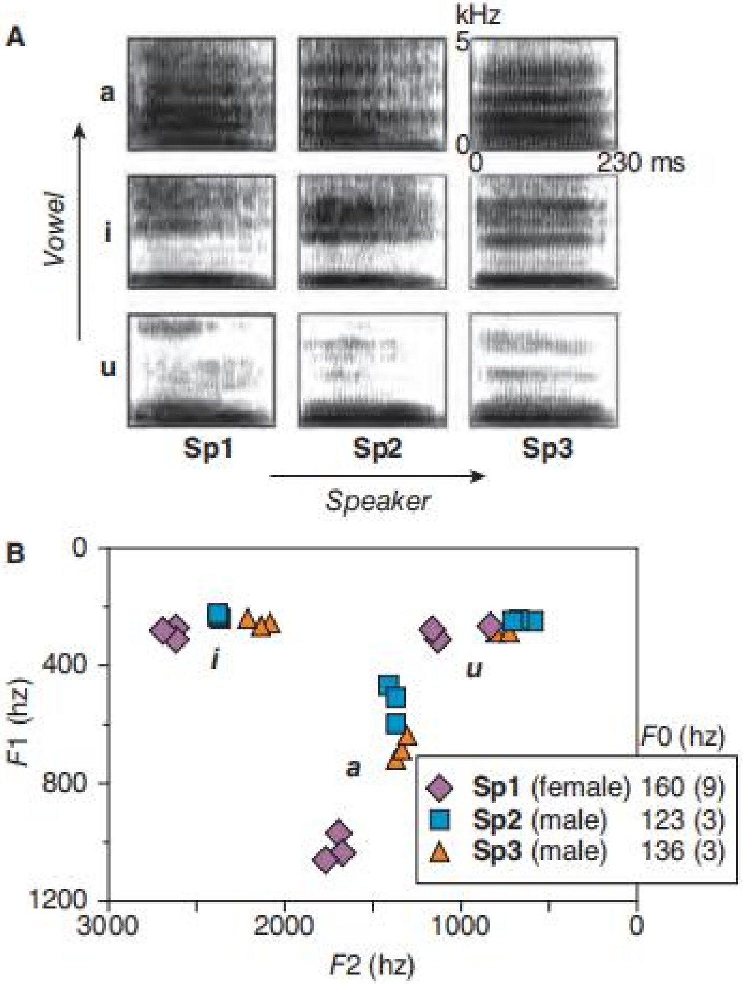


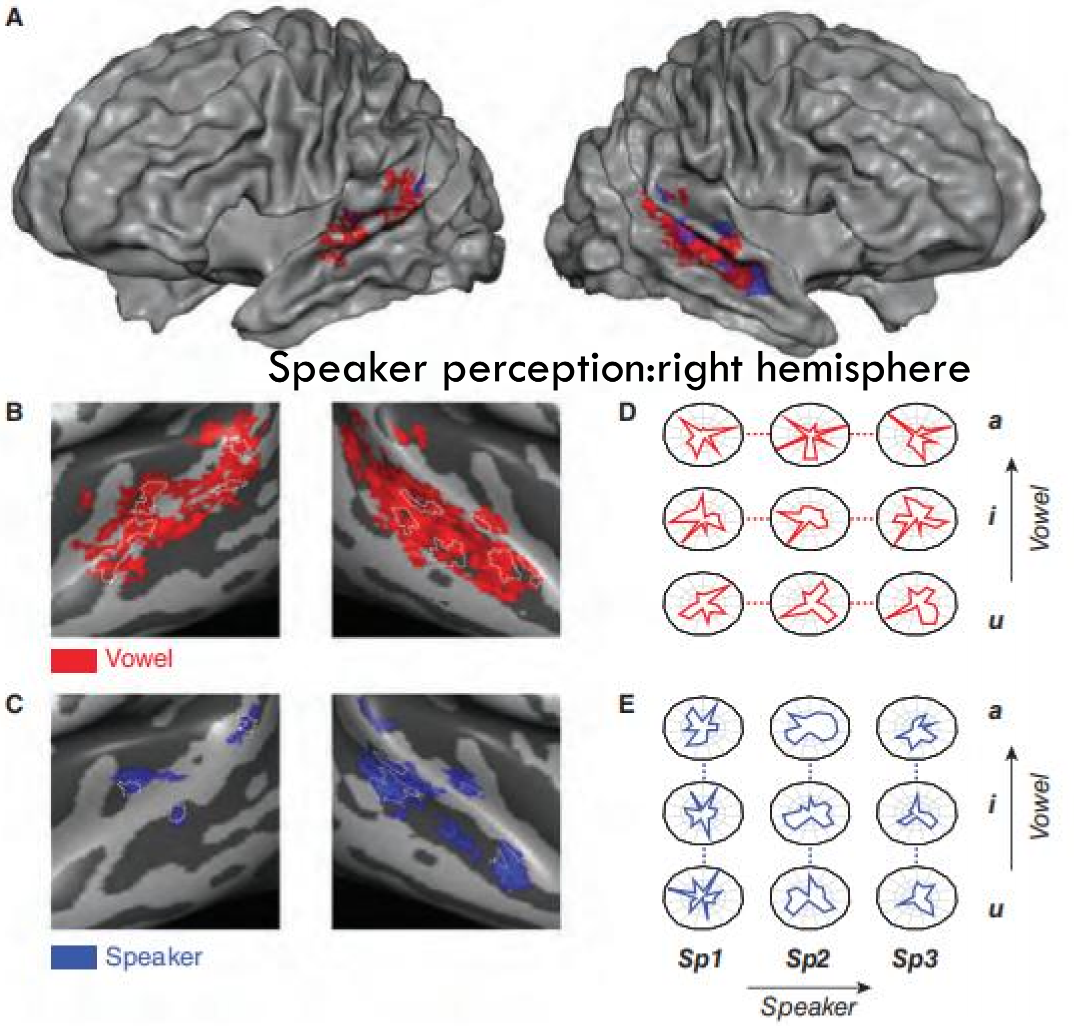
Figure 5.3 Stimuli and results of Binder et al.'s (2000) fMRI study of the perception of speech and nonspeech sounds. (A) Example narrow-band spectrograms of the five types of stimuli: N = noise; T = tones; W = words; P = pseudowords; R = reversed words. (B) Hierarchical contrasts between noise, tones, and the three types of speech stimuli. Tones activate the dorsal STG relative to noise (blue), whereas speech activates more ventral regions in the STS (yellow). (From

NEW: WHO SAYS WHAT



Vowel perception:bilateral STG and middle STS

Speaker perception:right hemisphere



DIFFERENT HEMISPHERES DIFFERENT SPEECH PERCEPTION (WADA PROCEDURE, WADA & RASMUSSEN, 1960)

Wada procedure:A way to temporarily shut down an entire hemisphere by injecting sodium amobarbitol(氨巴比妥钠) into either the left or the right carotid artery.

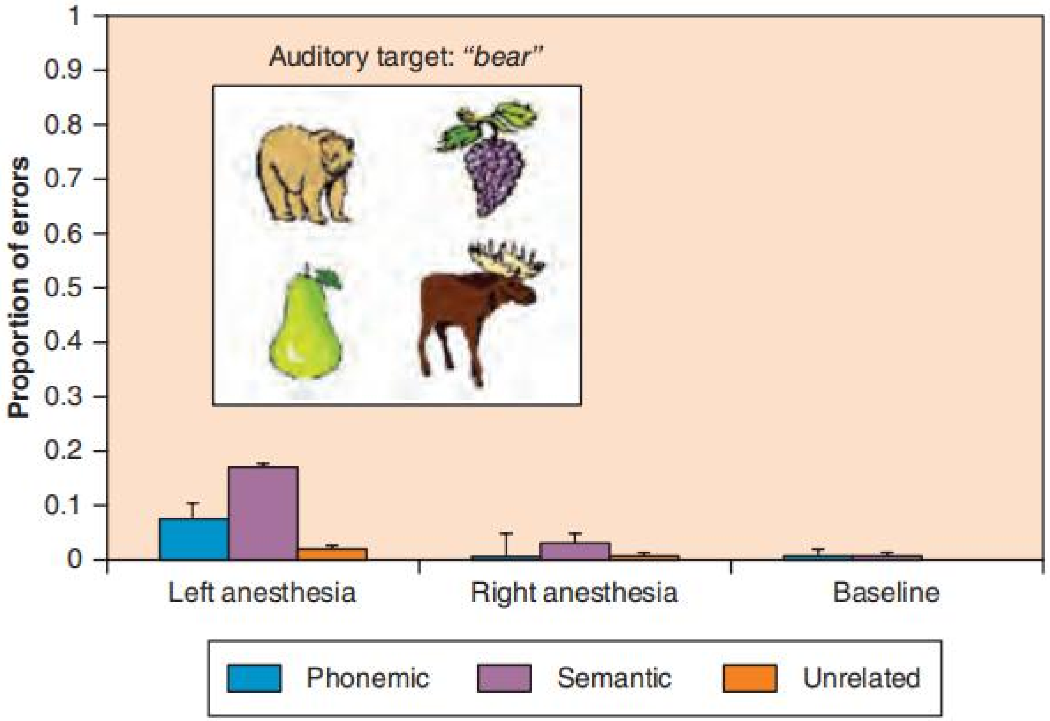


Figure 5.7 Speech recognition in patients undergoing Wada procedures. A sample stimulus card is presented along with average error rates of patients during left-hemisphere anesthesia, right-hemisphere anesthesia, or no anesthesia. Subjects were presented with a target word auditorily and asked to point to the matching picture. Note that overall performance is quite good and further that when patients make errors, they tend to be semantic in nature (selection of a semantically similar distractor picture) rather than a phonemic confusion (selection of a phonemically similar distractor picture). (From Hickok, 2009b, p. 124.)

HYPOTHESIS: THE EARLY CORTICAL STAGES OF SPEECH PERCEPTION ARE BILATERALLY ORGANIZED.

Wada procedure:A way to temporarily shut down an entire hemisphere by injecting sodium amobarbitol(氨巴比妥钠) into either the left or the right carotid artery.

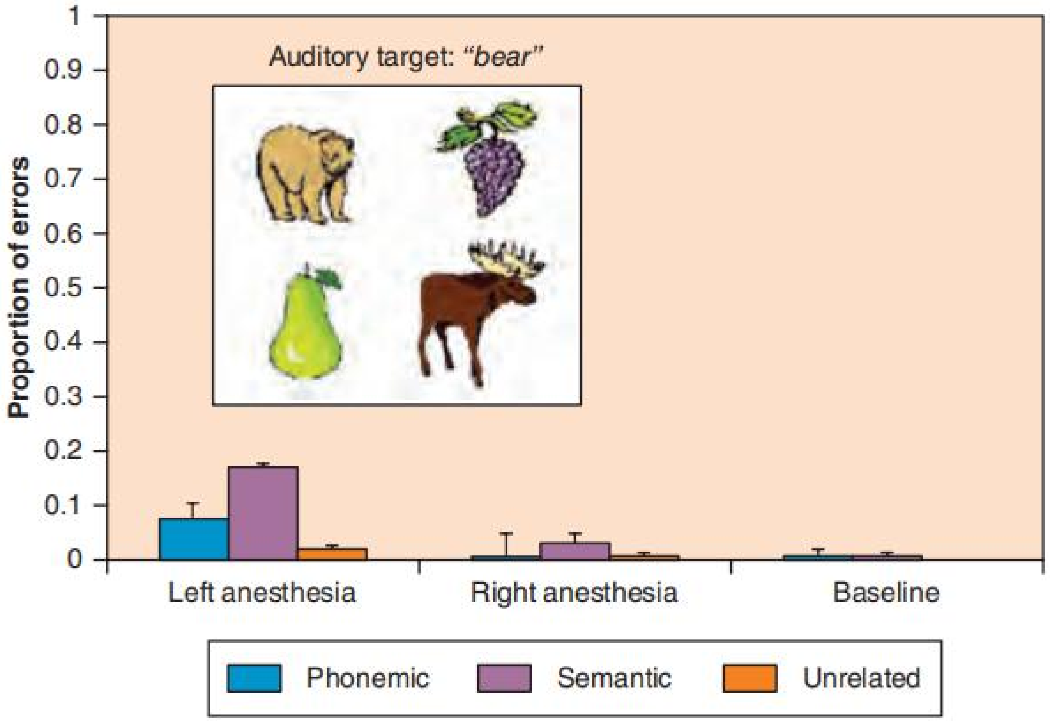


Figure 5.7 Speech recognition in patients undergoing Wada procedures. A sample stimulus card is presented along with average error rates of patients during left-hemisphere anesthesia, right-hemisphere anesthesia, or no anesthesia. Subjects were presented with a target word auditorily and asked to point to the matching picture. Note that overall performance is quite good and further that when patients make errors, they tend to be semantic in nature (selection of a semantically similar distractor picture) rather than a phonemic confusion (selection of a phonemically similar distractor picture). (From Hickok, 2009b, p. 124.)

Word deafness: A disorder in which speech perception is impaired, despite intact hearing and sometimes even intact recognition of nonspeech sounds.

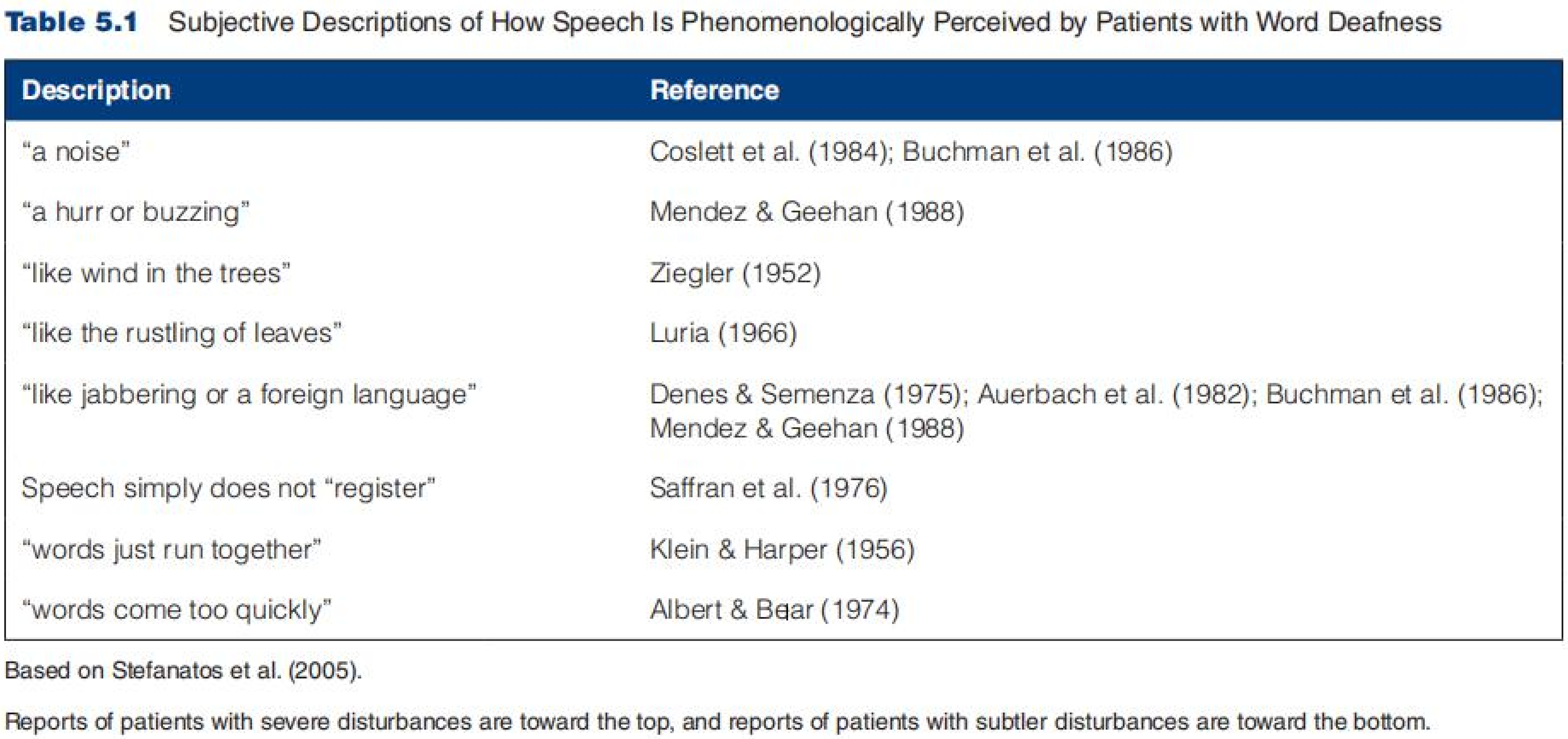


Table 5.1 Subjective Descriptions of How Speech Is Phenomenologically Perceived by Patients with Word Deafness

Description	Reference
“a noise”	Coslett et al. (1984); Buchman et al. (1986)
“a hurr or buzzing”	Mendez & Geehan (1988)
“like wind in the trees”	Ziegler (1952)
“like the rustling of leaves”	Luria (1966)
“like jabbering or a foreign language”	Denes & Semenza (1975); Auerbach et al. (1982); Buchman et al. (1986); Mendez & Geehan (1988)
Speech simply does not “register”	Saffran et al. (1976)
“words just run together”	Klein & Harper (1956)
“words come too quickly”	Albert & Bear (1974)

Based on Stefanatos et al. (2005).

Reports of patients with severe disturbances are toward the top, and reports of patients with subtler disturbances are toward the bottom.

72% (STG)

DIFFERENT HEMISPHERES DIFFERENT TEMPORAL WINDOWS FOR SPEECH PERCEPTION

Left hemisphere: integrating signals on the time scale of rapidly varying phonemes.

Right hemisphere: integrating signals on the time sclae of longer-duration syllables.

Voice-onset time between /k/ and /g/: 20-80 ms

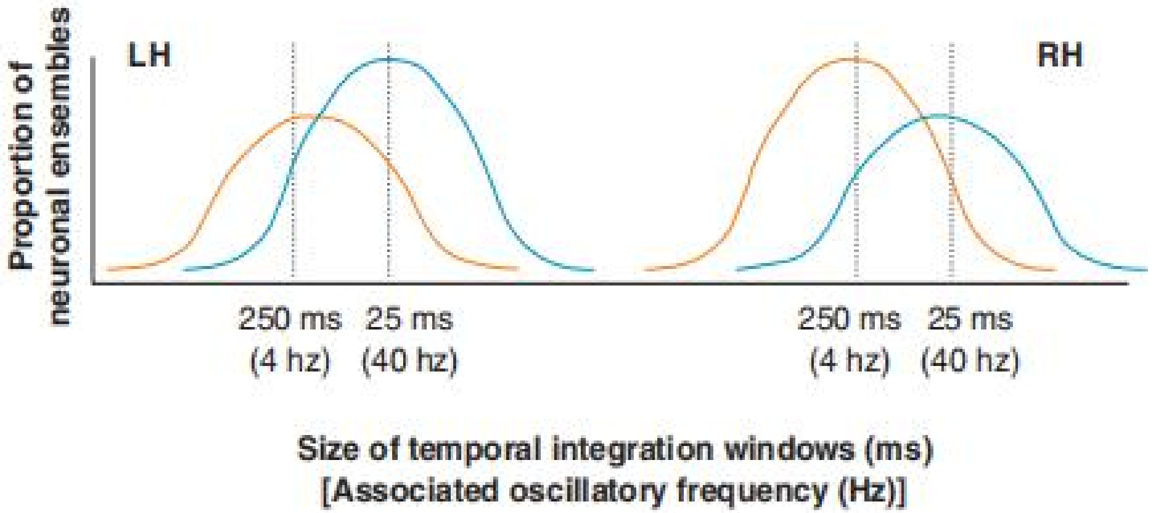
Syllabic stress: 150-300 ms.

DIFFERENT HEMISPHERES DIFFERENT TEMPORAL WINDOWS FOR SPEECH PERCEPTION

Voice-onset time between /k/ and /g/: 20-80 ms

Syllabic stress: 150-300 ms.

“Asymmetric sampling in time” hypothesis (Poeppel et al., 2008)

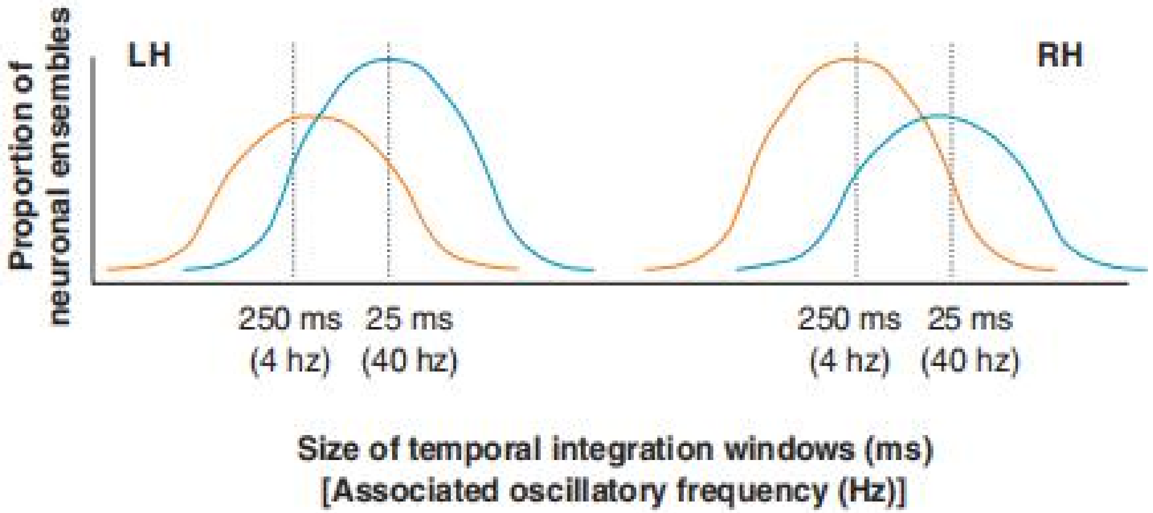


DIFFERENT HEMISPHERES DIFFERENT TEMPORAL WINDOWS FOR SPEECH PERCEPTION

Voice-onset time between /k/ and /g/: 20-80 ms

Syllabic stress: 150-300 ms.

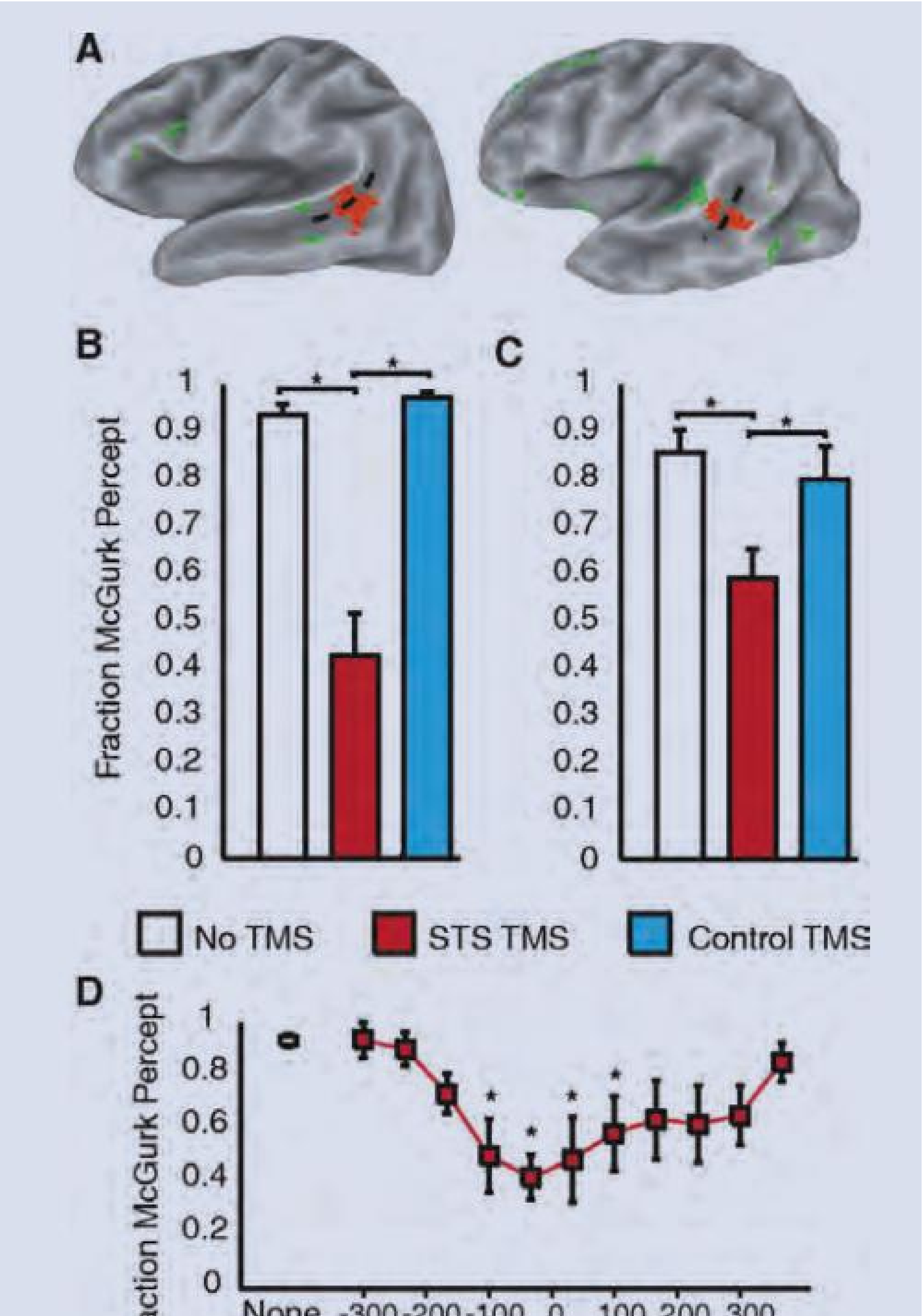
“Asymmetric sampling in time” hypothesis (Poeppel et al., 2008)



AUDITORY-VISUAL INTEGRATION DURING SPEECH PERCPETION

McGurk Effect: is an astonishing illusion which demonstrates that during normal face-to-face speech perception, the brain automatically fuses the simultaneously occurring auditory and visual signals (McGurk & MacDonald, 1976).

Left posterior STS plays a critical role in auditory–visual integration during speech perception


A
B
Fraction McGurk Percept
*
*
C
*
*
No TMS
STS TMS
Control TMS
D
Fraction McGurk Percept
*
*
*
*
None