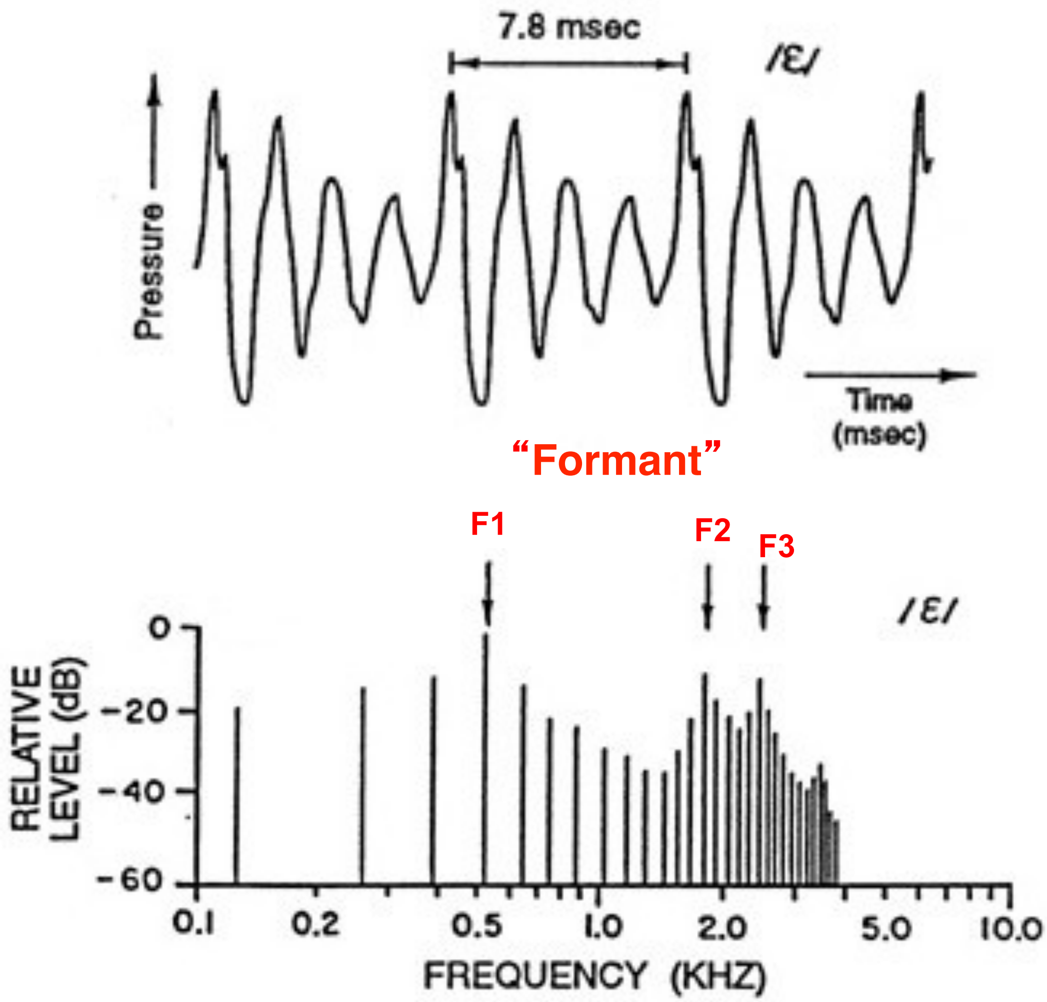
580.422

System Bioengineering II: Neurosciences

Central Auditory System (2)

Prof. Xiaoqin Wang

Spectral and Temporal Characteristics of Speech

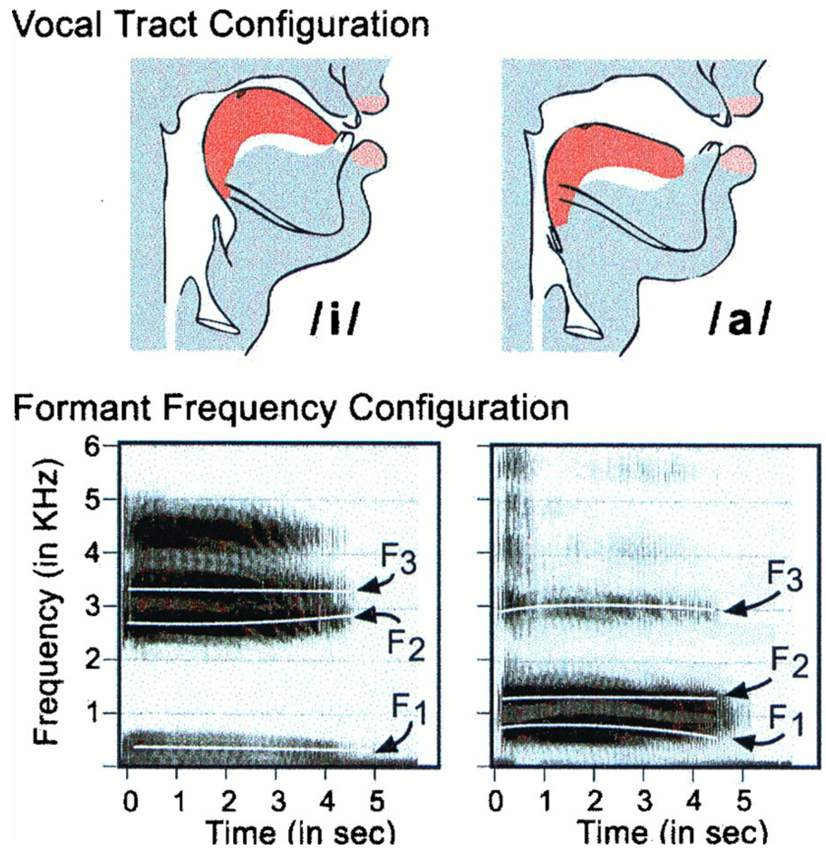


Amplitude-time Waveform

Spectrum

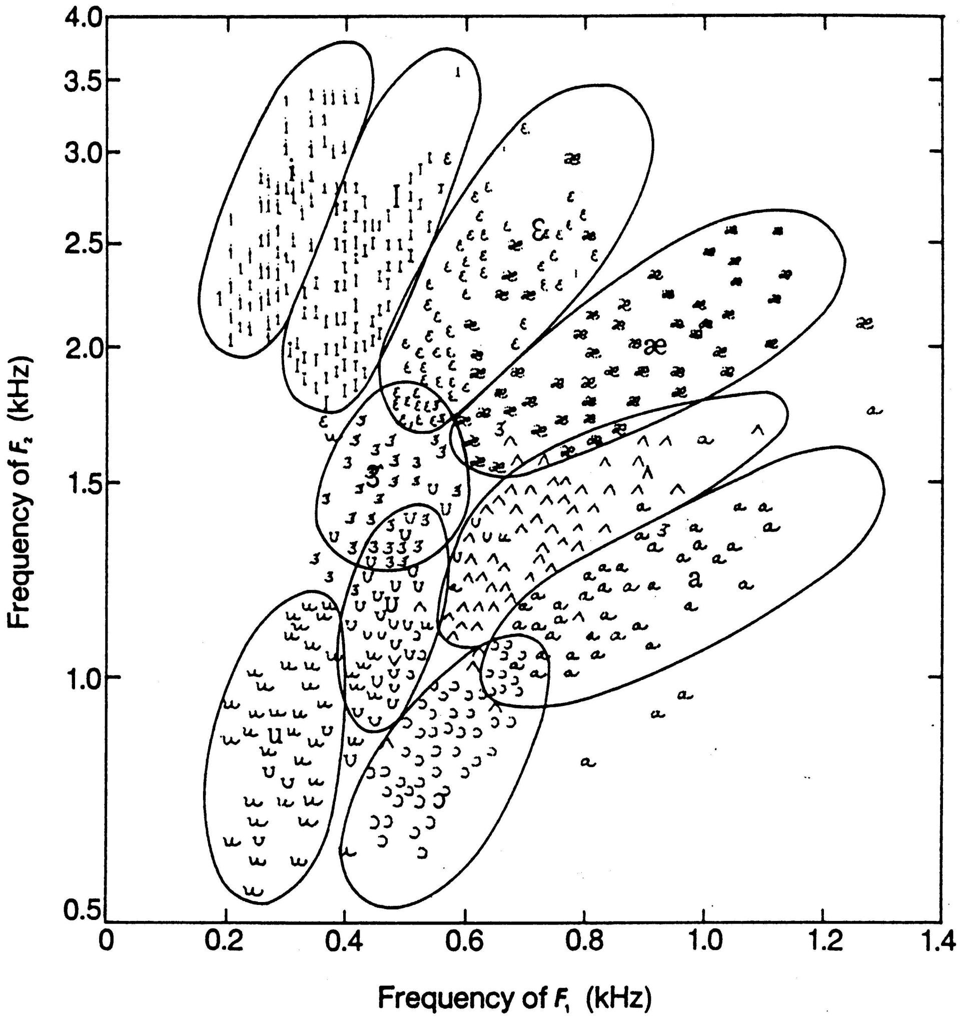
[From: Sachs and Young]

Speech Production Mechanisms



Vocal tract positions (*Upper*) and spectrographic displays (*Lower*) for the vowels/i/as in “heat” and/a/as in “hot.” Formant frequencies, regions of the frequency spectrum in which the concentration of energy is high, are marked for each vowel. [From: Kuhl P K, *Proc Natl Acad Sci U S A.* 97(22):11850-7 (2000)]

Formant Representation of English Vowels



Frequency of the second formant versus frequency of the first formant of English vowels

Peterson and Barney (*JASA*, 1952)

Spectrum of English Vowels

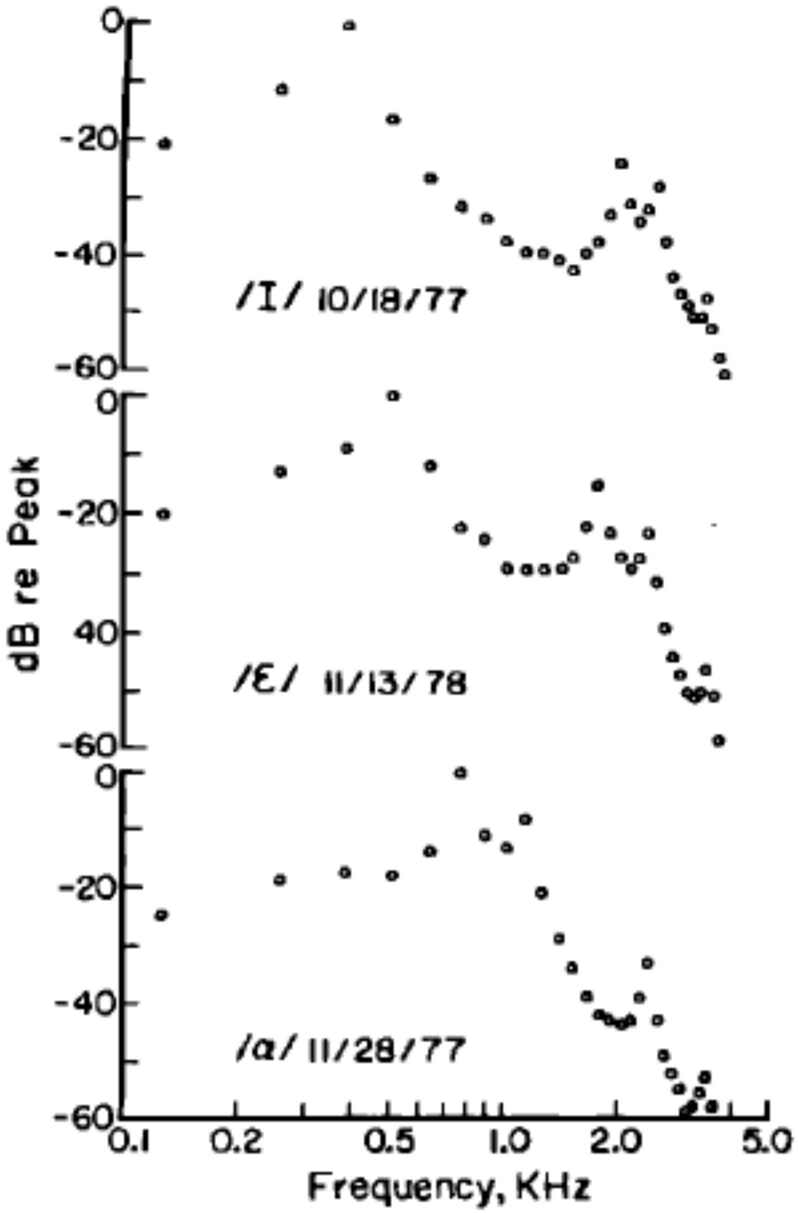


FIG. 1. Top plot shows sound pressure near the eardrum for constant amplitude electrical signal into the earspeaker. This was measured using the probe tube system described by Sokolich (1977) during experiment 10/18/77. Bottom three plots show amplitude spectra of the three vowels for which data are presented in this paper; these were measured near the cat's eardrum during the experiments. Each point is the amplitude of one harmonic of these steady-state vowels; the vowels are periodic with fundamental frequency 128 Hz.

Auditory nerve can represent speech sound (e.g., vowel) by:

1)Firing rate
2)Temporal discharge pattern

Representations of vowel spectrum by normalized firing rate of auditory nerve (AN)

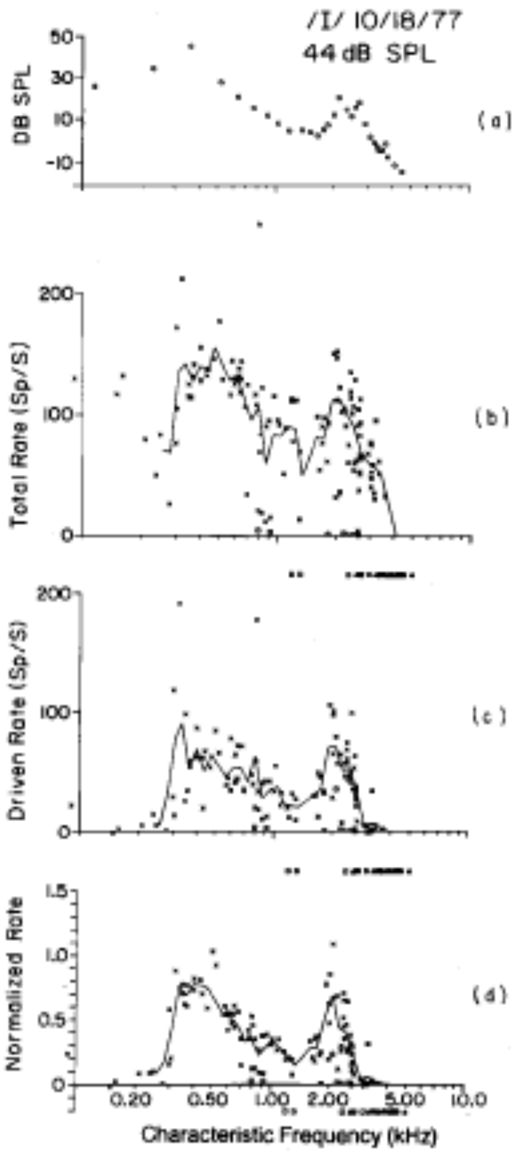


FIG. 2. (a) Spectrum of the synthesized /I/ measured near the eardrum in experiment 10/18/77; plotted as the sound-pressure level of each harmonic when the overall stimulus was at 44 dB SPL. (b) Discharge rate in response to /I/ at 44 dB SPL of each fiber studied at this sound level in experiment 10/18/77 (115 fibers). Each point is rate of one unit plotted at the unit's CF. Units with spontaneous rate less than 1/s plotted with squares; other units with X's. Line is average rate of units with spontaneous rate greater than 1/s, computed as described in the text. (c) Same data as in (b), except plotted as rate minus spontaneous rate. (d) Same data as in (b), except plotted as (rate minus spontaneous rate) divided by (saturation rate minus spontaneous rate).

Firing rate representation of AN degrades at high sound levels

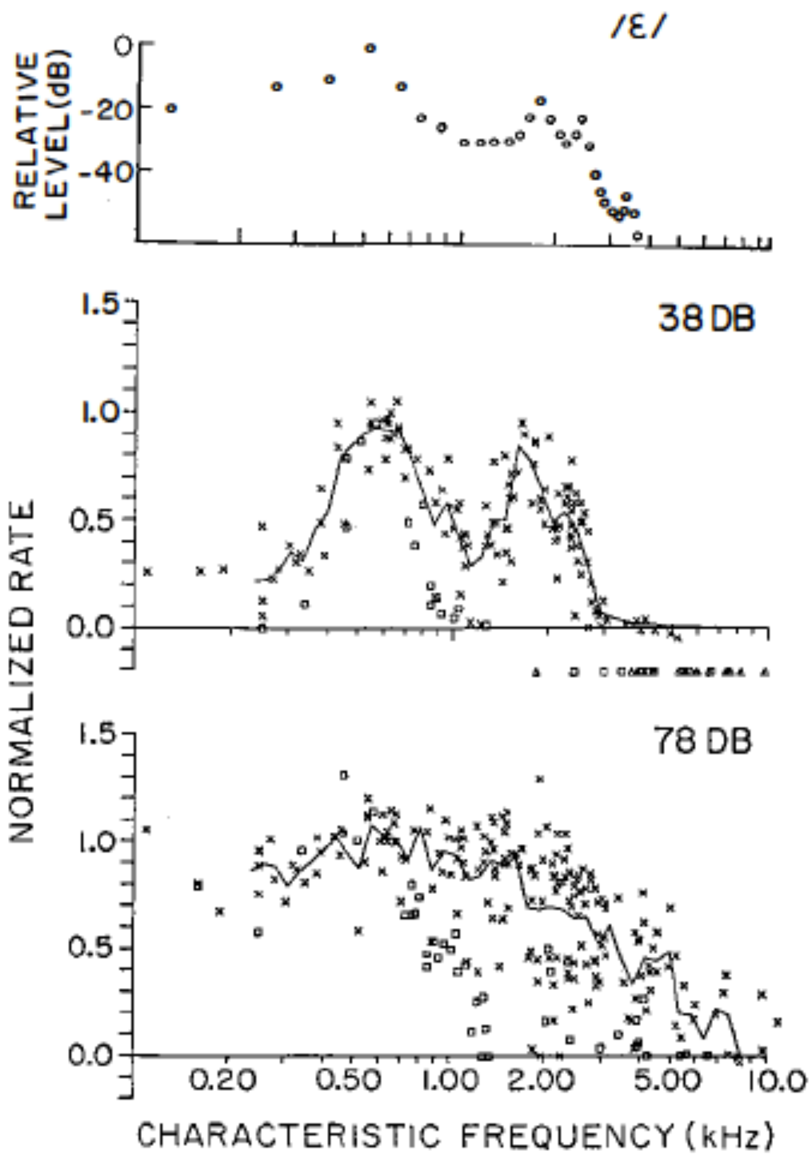


Figure 2 Top: Amplitude spectrum of /ε/. *Middle:* Plot of normalized rate vs characteristic frequency for units studied on 11/13/78 with /ε/ as the stimulus at 38 dB SPL. *Bottom:* Same as middle but at stimulus level of 78 dB SPL.

Sachs and Young 1979

Low-spontaneous AN can represent vowel spectrum by firing rate at high sound levels

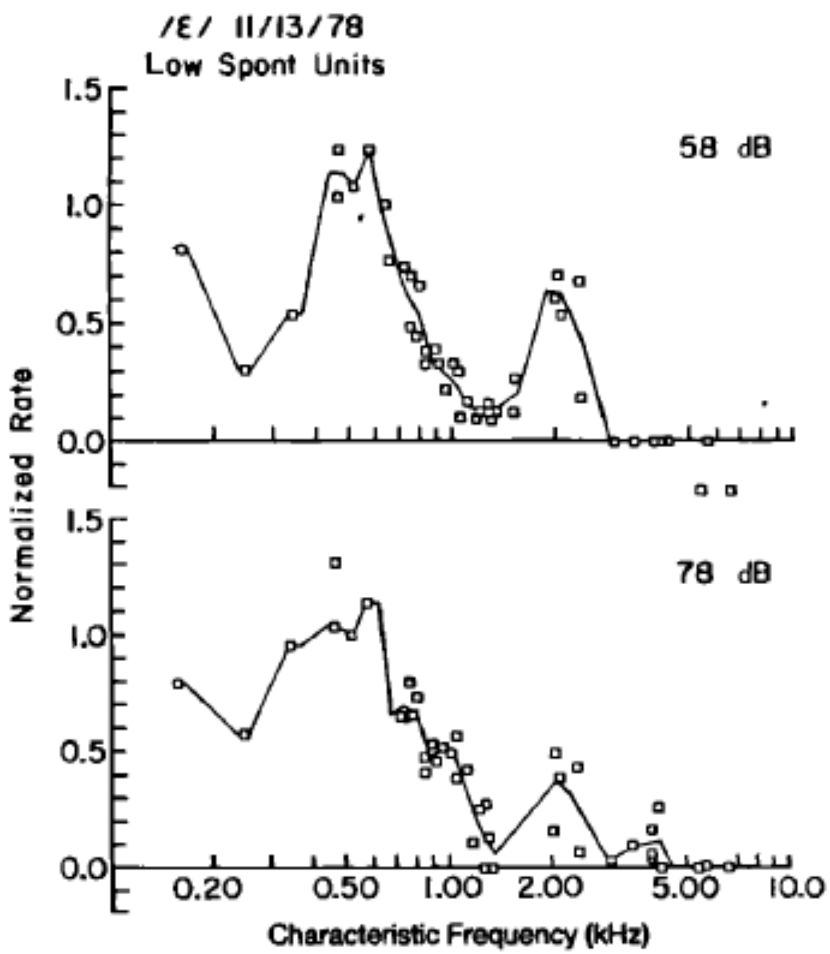


FIG. 12. Normalized rate vs CF for low spontaneous rate units (less than 1/s) studied on 11/13/78 with /ε/ as the stimulus. Data are replotted from Fig. 6. In this case, the average curves include all windows containing at least one unit instead of at least three as in other plots.

Temporal Responses to Vowel in AN

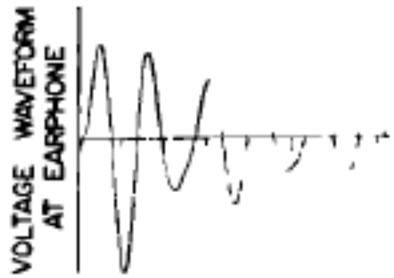
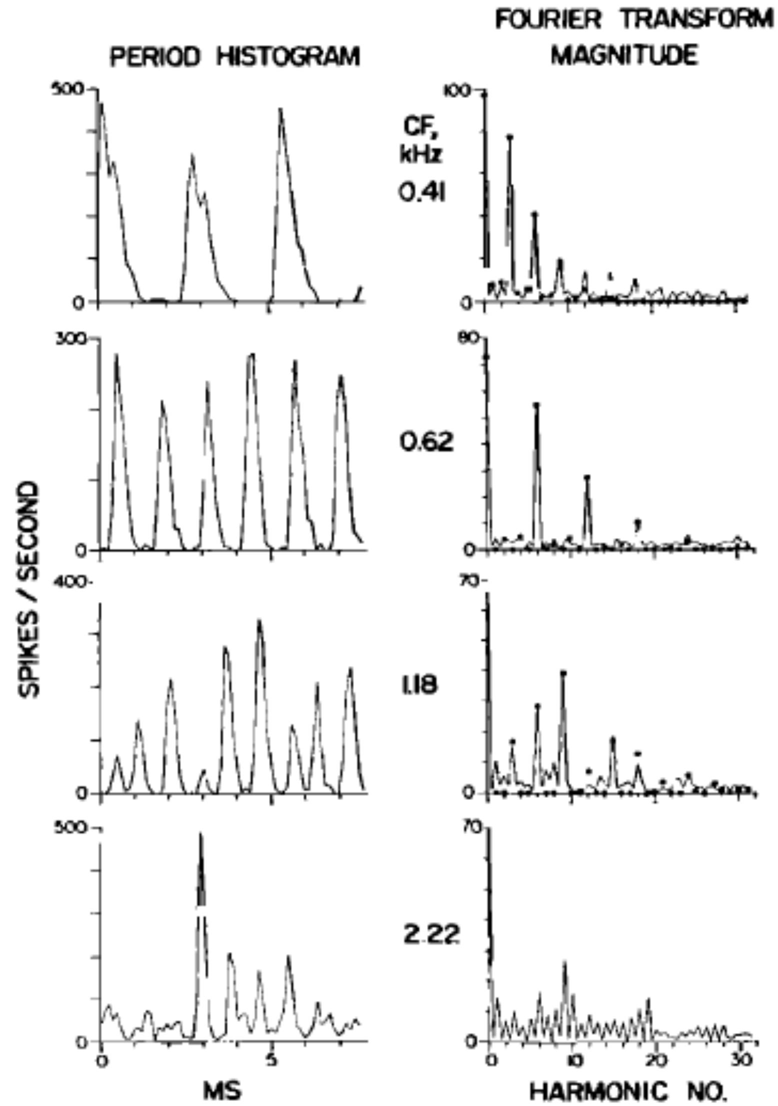


FIG. 2. Period histograms (left column) and amplitude spectra of Fourier transforms of the period histograms (right column) for four single fibers studied on 11/28/77 using /ɑ/ as the stimulus. CFs of the units are shown in the center column. Period histograms are estimates of instantaneous discharge rate as a function of time through one cycle of the vowel; computed with 64 bins per cycle. One pitch period of the vowel (electrical signal at earphone input) is shown at the top of the left column; time scale is the same as the period histograms. Fourier transforms computed from two cycle period histogram which introduces a "noise" point between each adjacent pair of stimulus harmonics. The dots on the top three Fourier transform plots show the spectra of model histograms fit to these period histograms using the method discussed in the text and in the Appendix.

$$r(n)=R_0+2\sum_{k=1}^{N/2-1} R_k\cos(2\pi kn/N+\theta_k)\,,$$

$$n=0,1,2,\ldots,N-1\,.$$

The magnitudes of the Fourier transforms (the R_k)



Representing Vowel Spectrum by Synchronized (phase-locked) Firing
"Average localized synchronized rate (ALSR)"

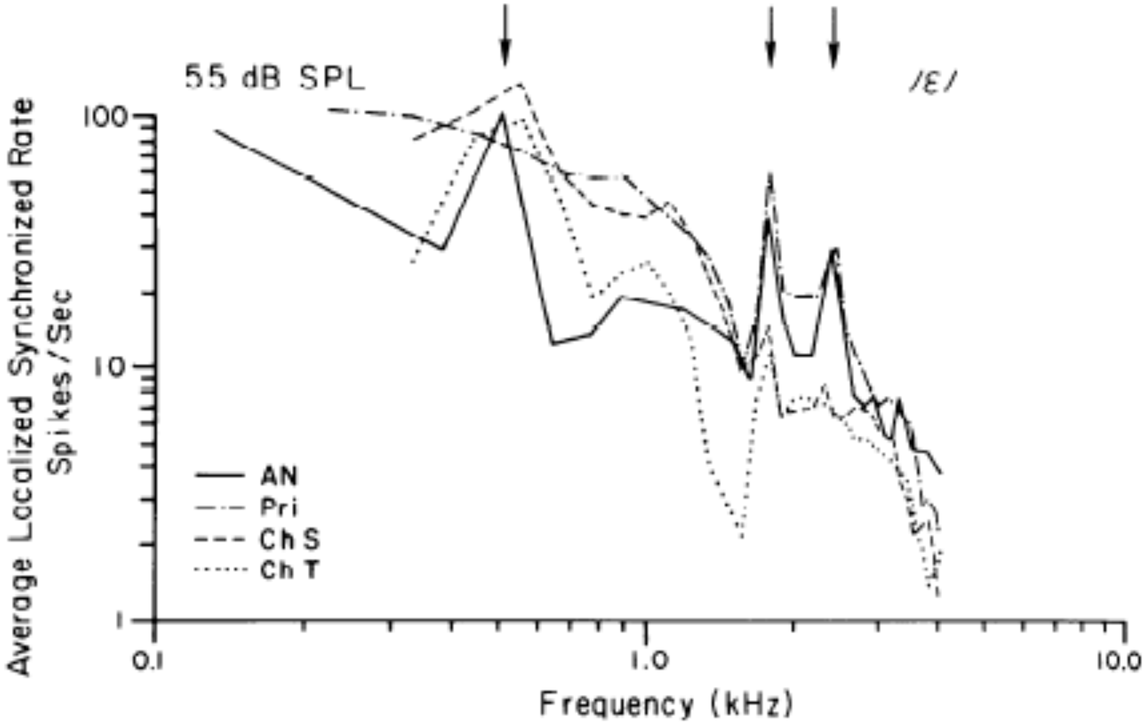


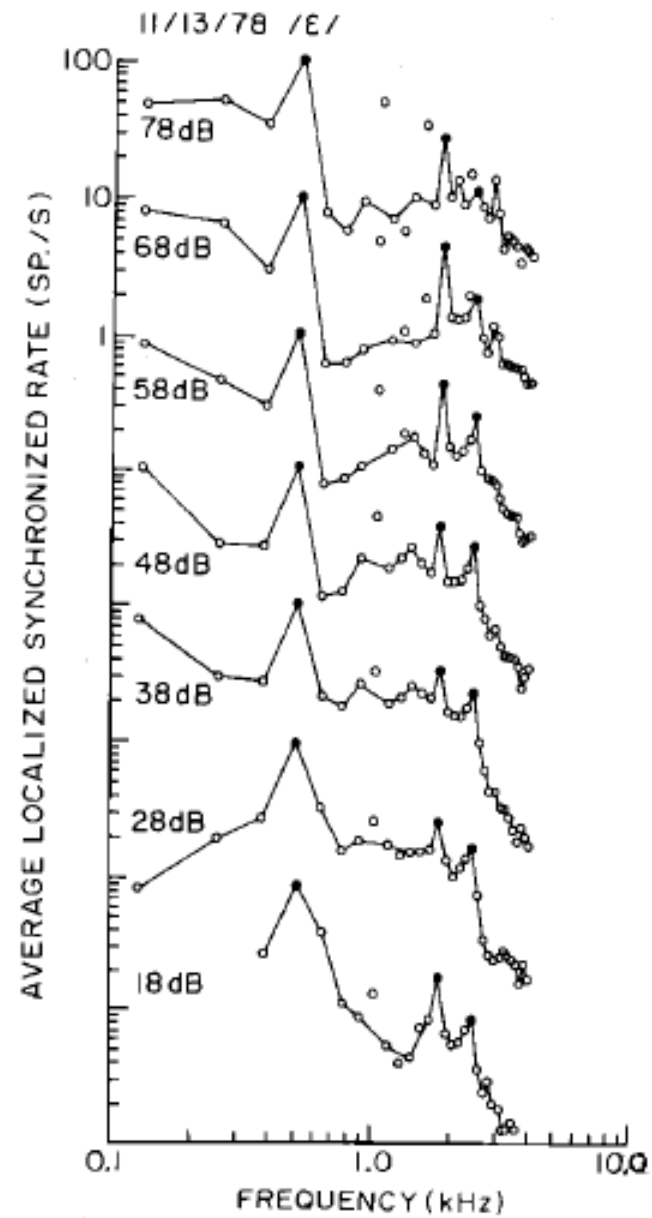
FIG. 18. Average localized synchronized rate profiles of responses of ANFs (solid line), Pri units (dot-dashed line), ChS units (dashed line), and ChT units (dotted line) to /ε/ at 55 dB SPL. All lines are plotted on the same scale. ANF data from Young and Sachs (1981).

$$\mathrm{ALSR}(k) = \frac{1}{M_k} \sum_{l \in C_k} R_{kl}$$

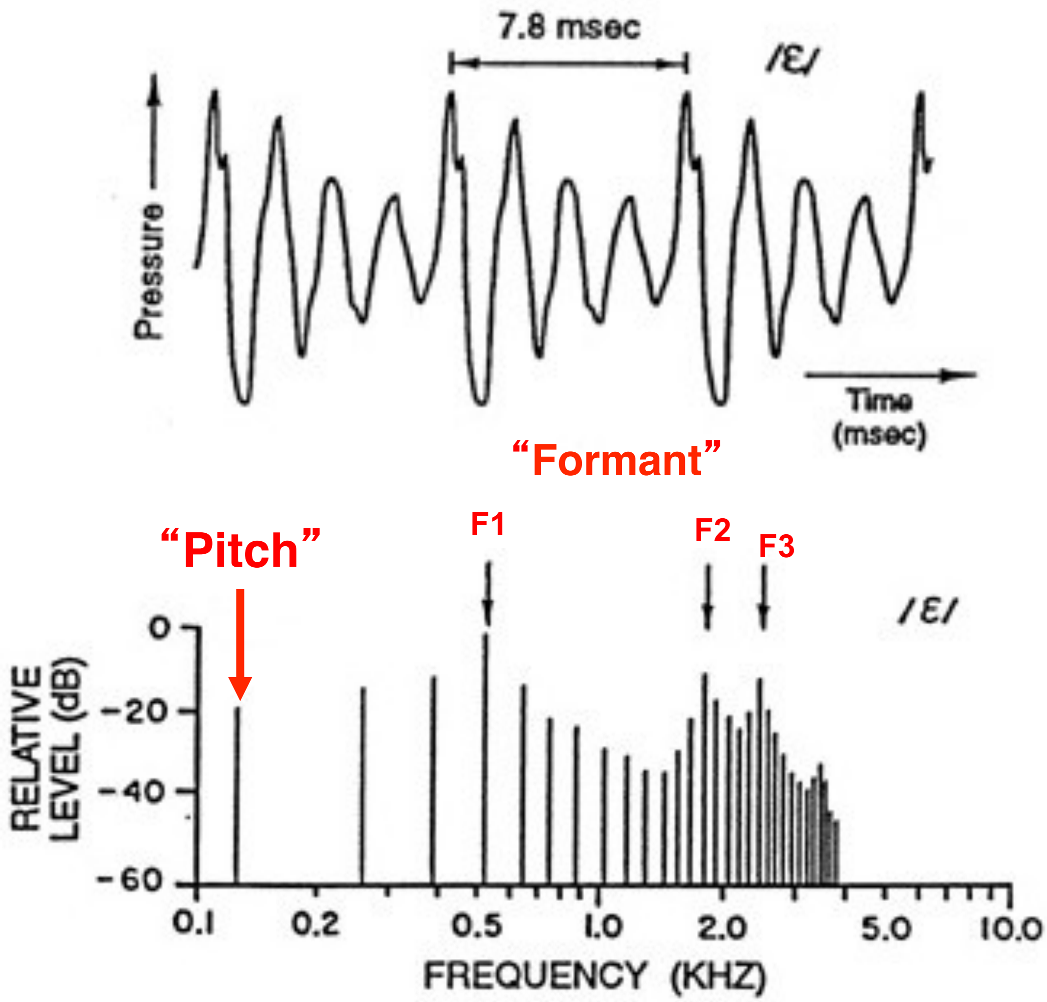
where C_k is the set of units with CF between $0.707kf_0$ and $1.414kf_0$; M_k is the number of units in C_k; and f_0 is the fundamental frequency of the vowel.

Firing Synchrony Representation Maintained at High Sound Levels

FIG. 6. The points show average localized synchronized rate computed according to Eq. (3) for responses to /ε/ at all sound levels used in experiment 11/13/78. There is one point for each harmonic up to the 32nd. Points corresponding to formant frequencies are plotted with filled circles. Ordinate is scaled logarithmically. Plots are shifted vertically from one another by one order of magnitude for clarity. Maximum response in each plot is about 100 spikes/s. The lines are drawn through all points except those corresponding to the 2nd and 3rd harmonics and the sum and difference tones of the first two formants. See rules given in text.



Spectral and Temporal Characteristics of Speech

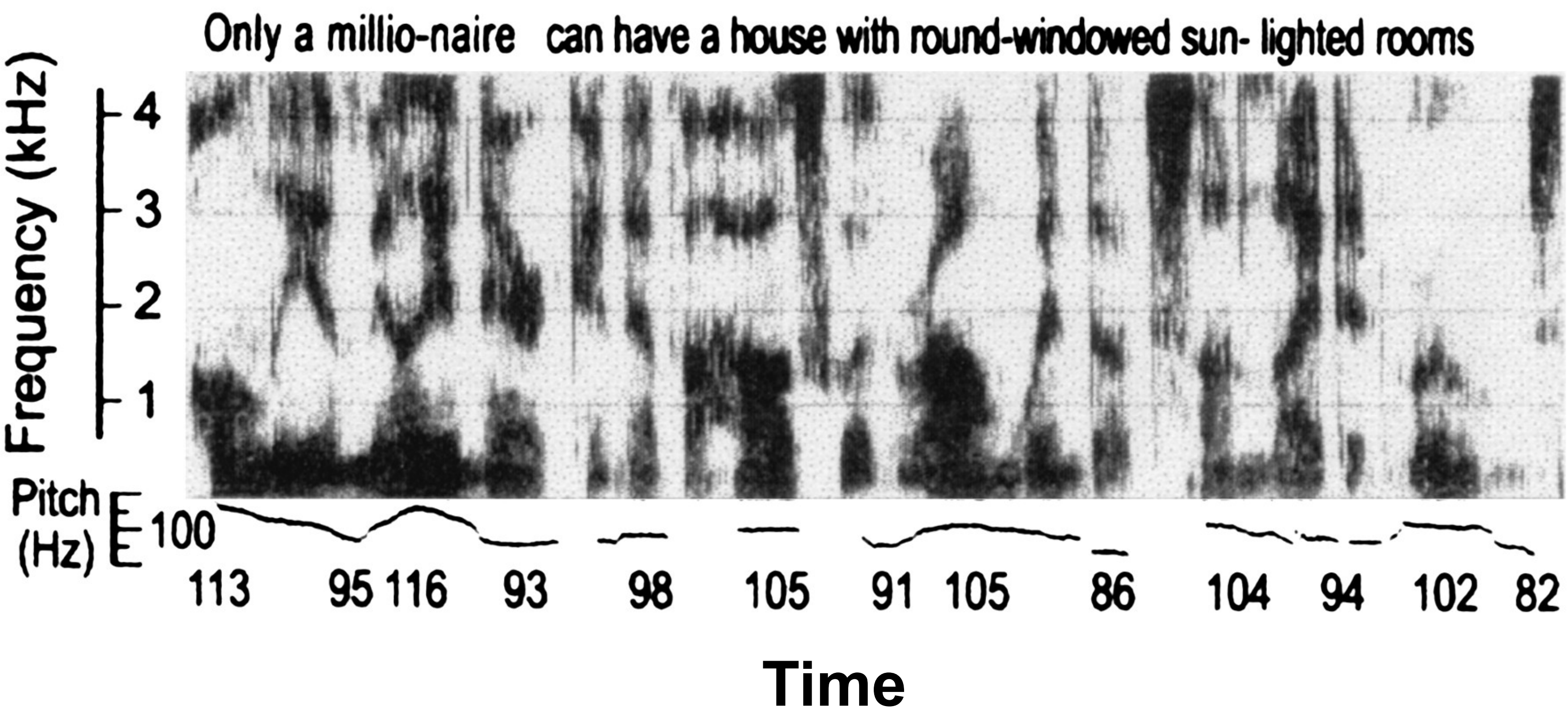


Amplitude-time Waveform

Spectrum

[From: Sachs and Young]

Pitch is a paralinguistic cue in English



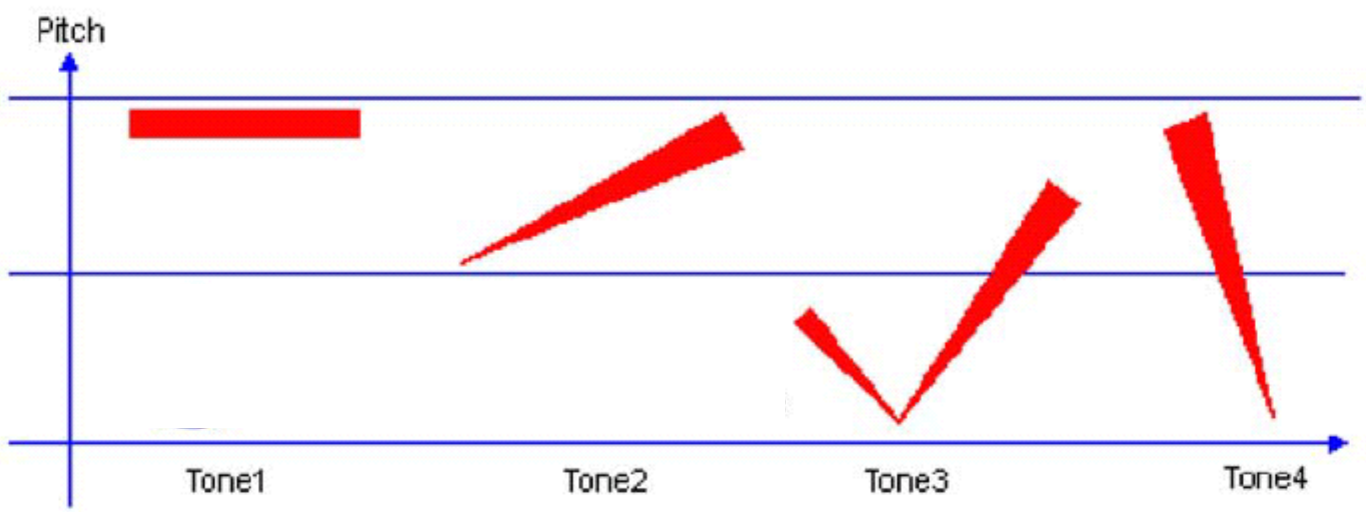
Spectrographic display of running speech showing the formant frequencies and the pitch (fundamental frequency) of the voice over time. Increases in pitch indicate primary stress in the utterance [From: Pickett, J M. *The Acoustics of Speech Communication.* Boston: Allyn and Bacon; 1999]

Pitch is a crucial linguistic cue in tonal languages

Four Chinese Tones (中文四声)

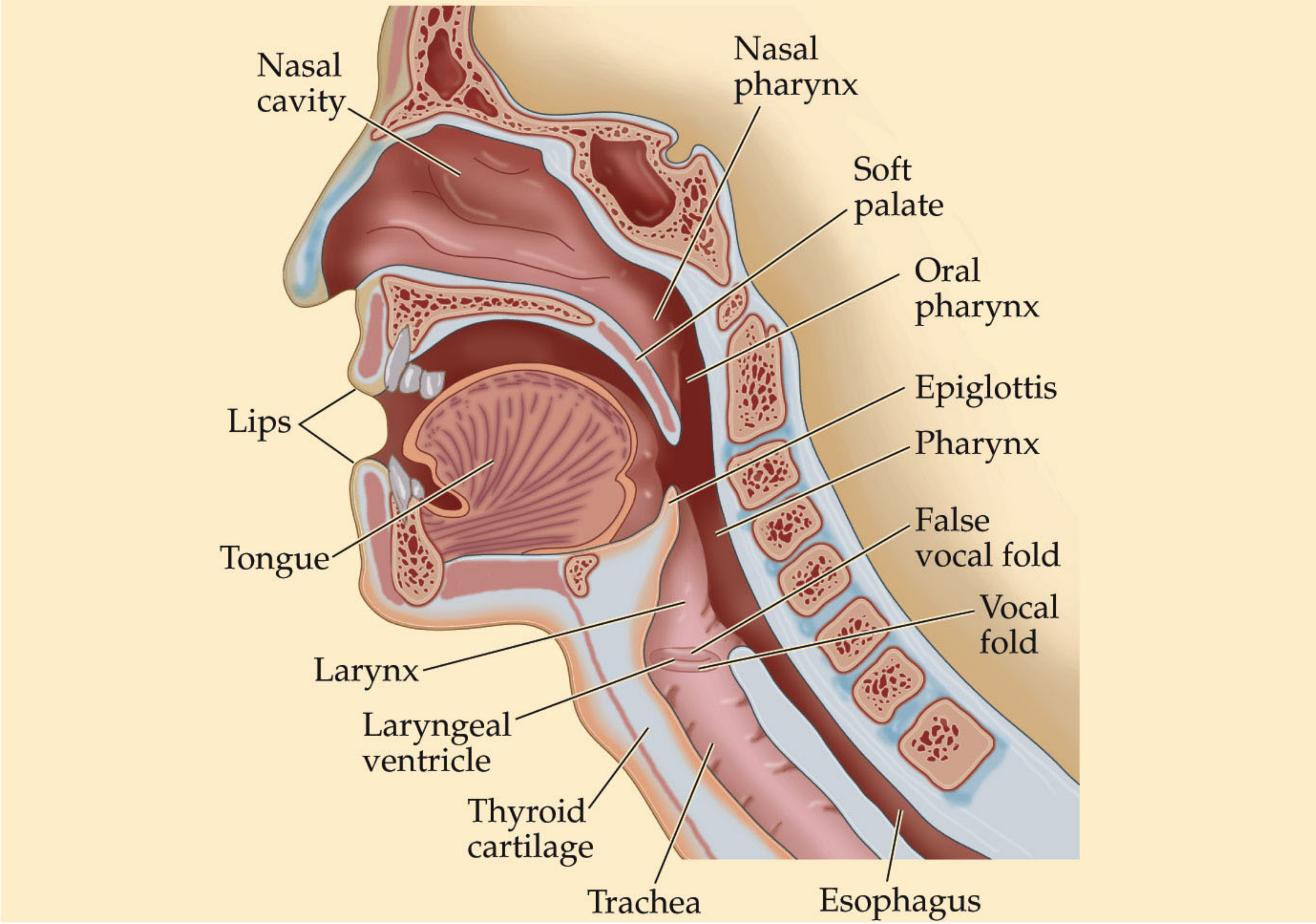
"Wang"

汪（vast)　王（King)　网（net)　忘（forget)



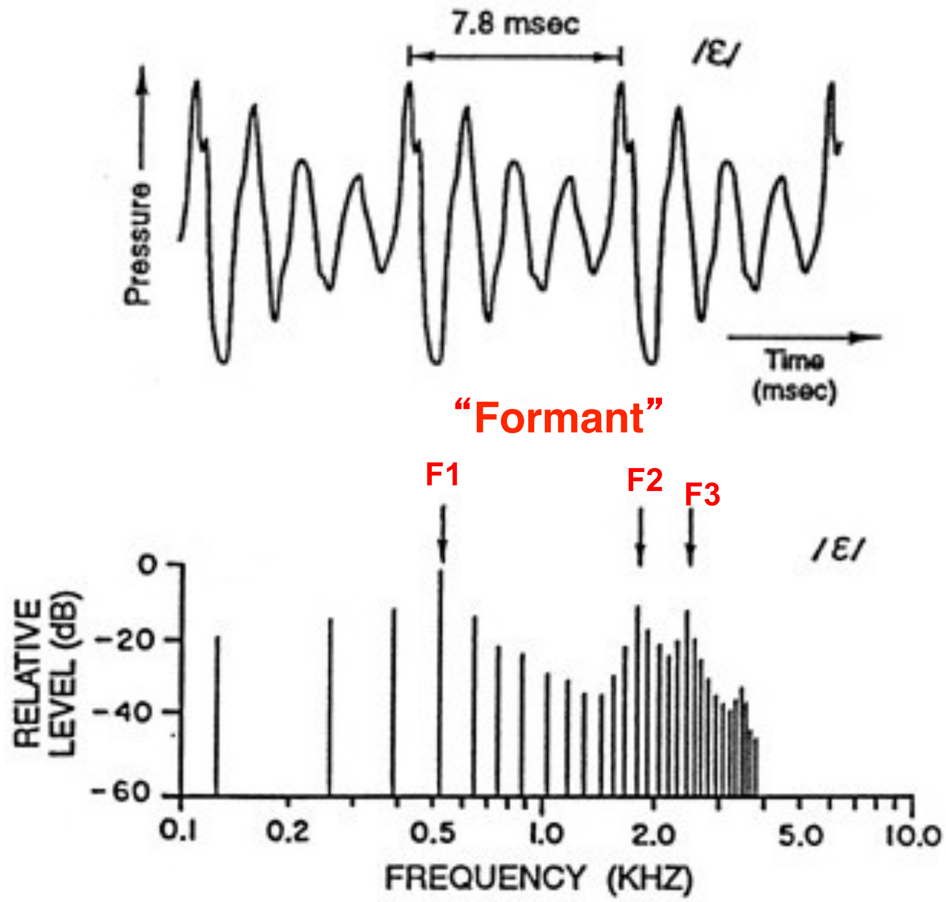
"Categorical Perception" (learned)

Box 27A Speech



NEUROSCIENCE, Fourth Edition, Box 27A

Spectral and Temporal Characteristics of Speech

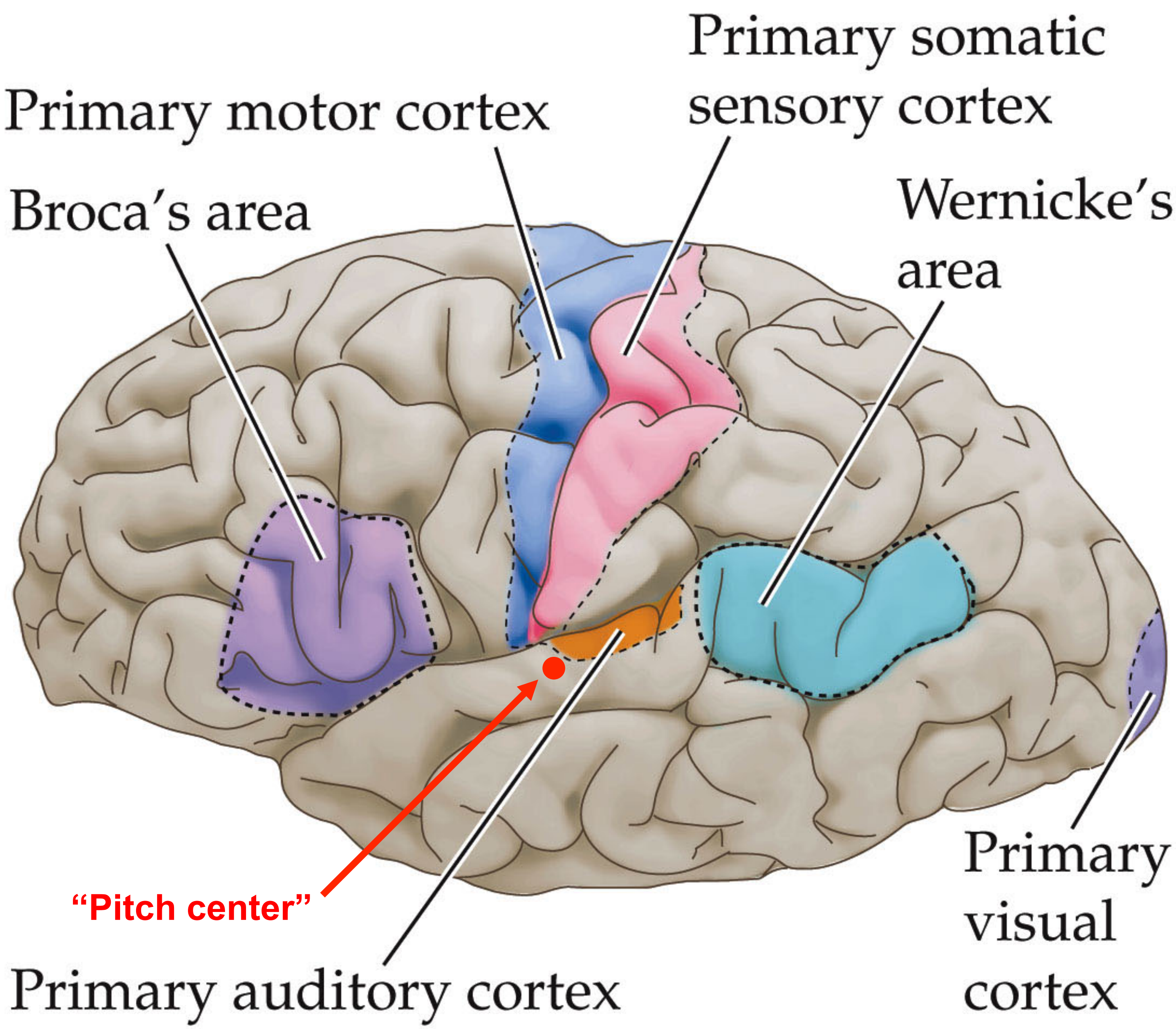


Amplitude-time Waveform

Spectrum

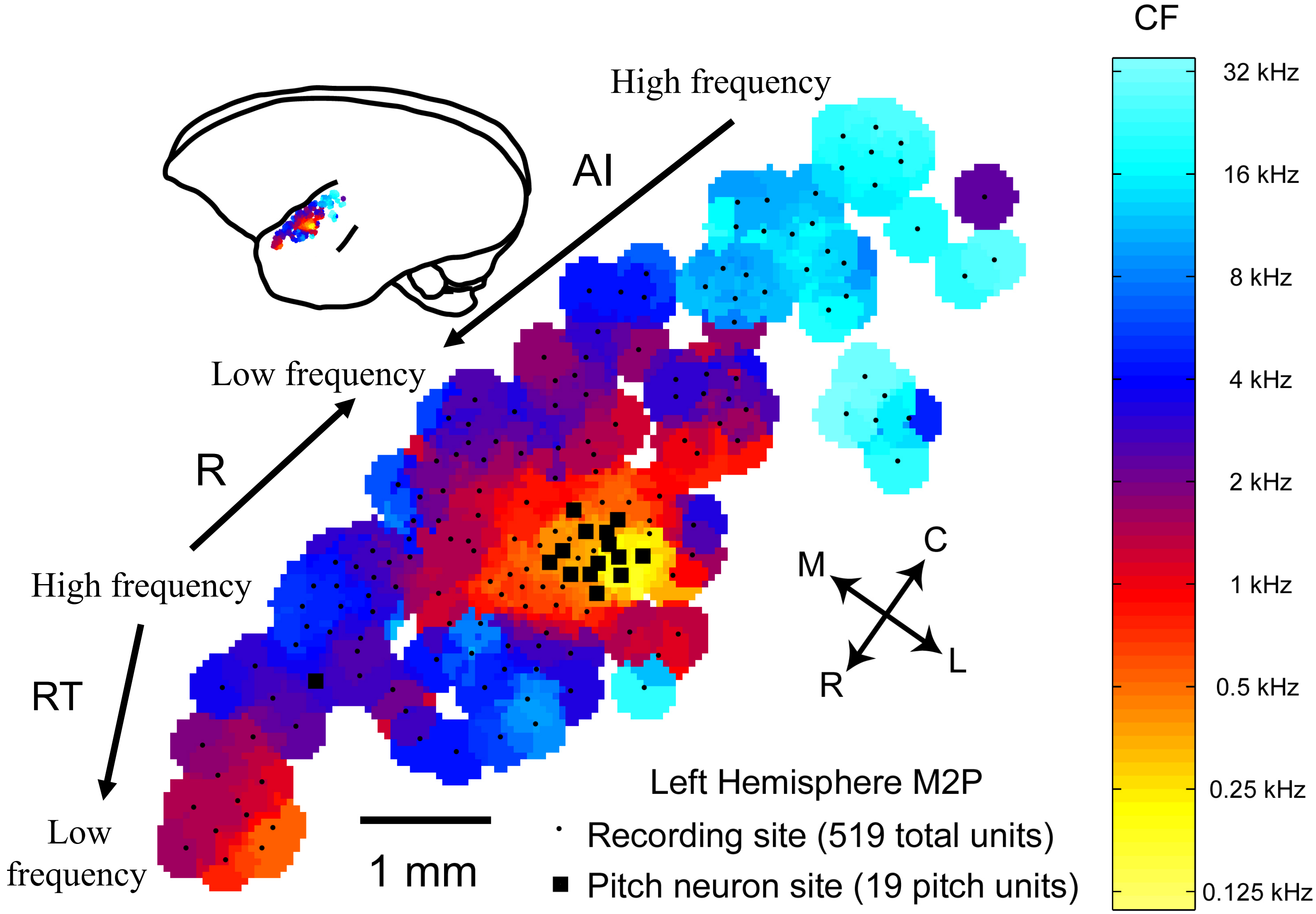
Amplitude waveform (*upper*) and spectrum (*lower*) of English vowel /ɛ/. Peaks of the spectrum are called "formants", referred to as first (F1), second (F2), etc., from low to high frequency. [From: Prof. Sachs and Young]

Figure 27.1 The major language areas of the brain



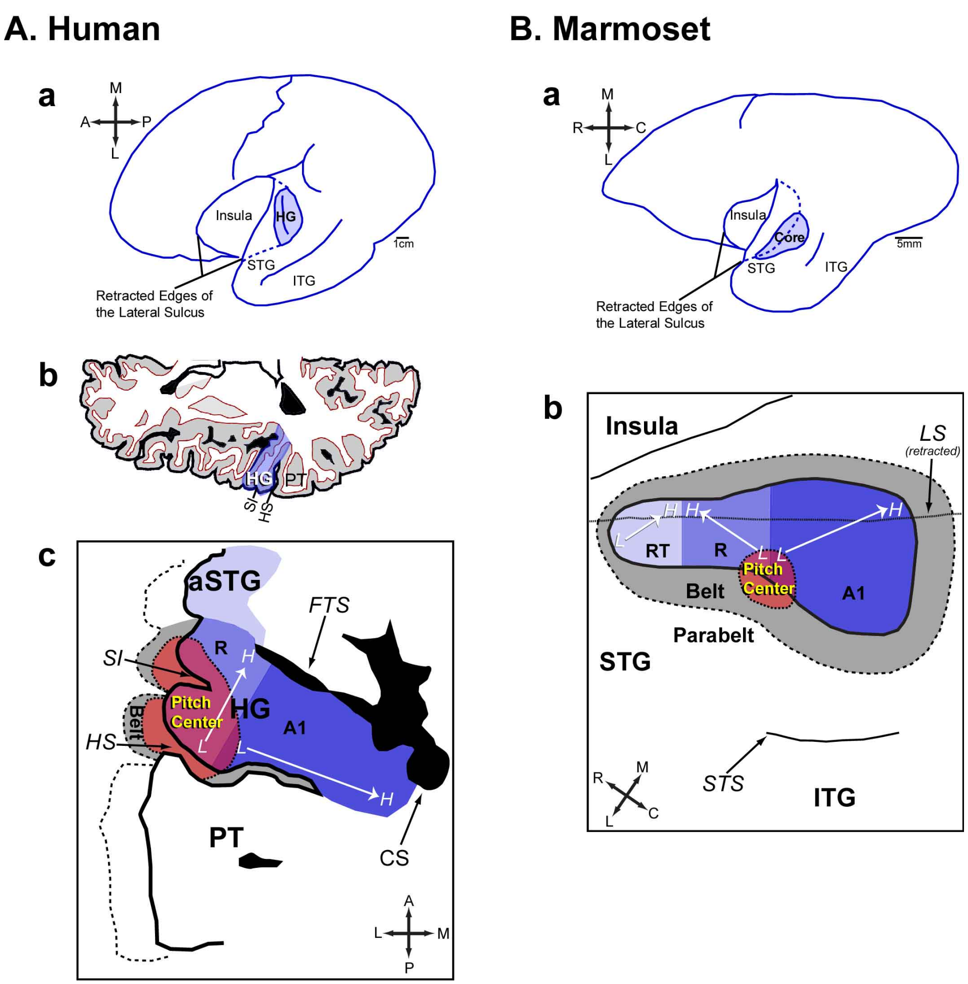
NEUROSCIENCE, Fourth Edition, Figure 27.1

A special “Pitch region” in marmoset auditory cortex



Bendor and Wang (*Nature*, 2005)

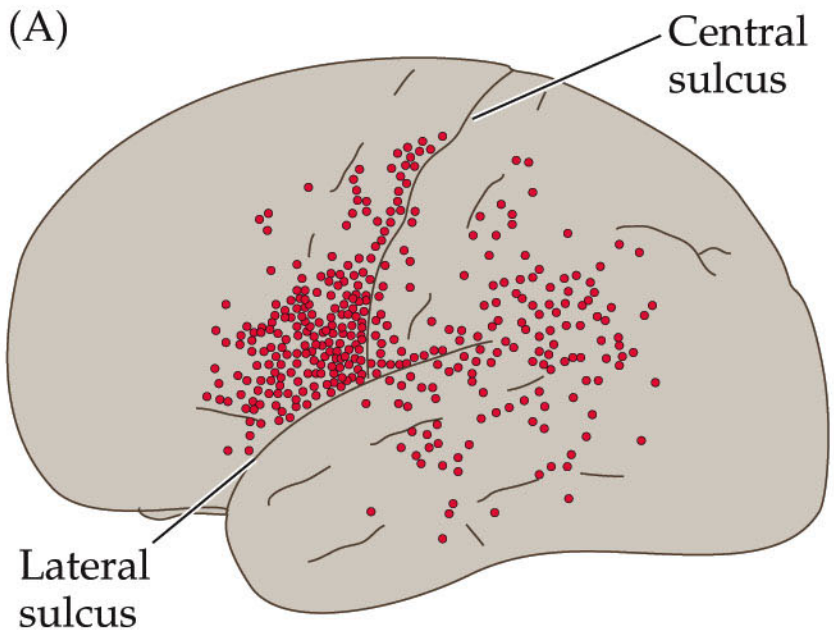
"Pitch center" in primate auditory cortex

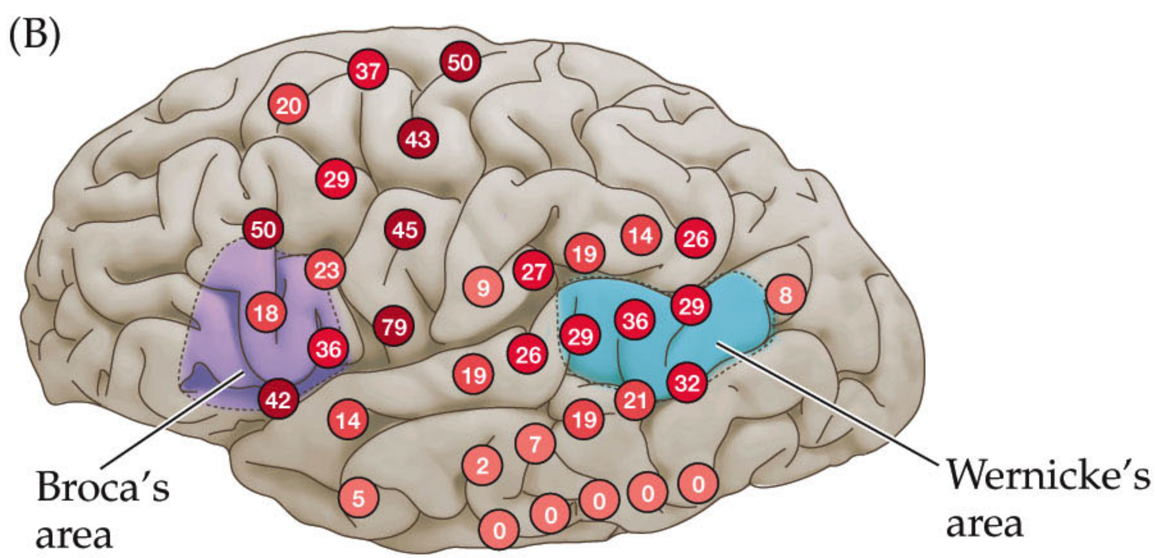


Location of the "Pitch center" in human and marmoset auditory cortex. (a) (i) Side view of a human brain, (ii) horizontal cross section of temporal lobe, and (iii) magnified view of Heschl's gyrus. Primary auditory cortex is presumed to occupy the medial portion of Heschl's gyrus (with variability between subjects). The location of neighboring areas (R, pitch center, lateral belt) is an approximation based on Schneider et al. [2005], Formisano et al. [2004], and Patterson et al. [2002]. (b) (i) Side view of the brain of a marmoset monkey and (ii) a magnified view of the temporal lobe, indicating core, belt, parabelt, and the pitch center. The borders between each auditory area are estimated on the basis of data from Bendor and Wang [2005], and Pistorio et al. [2004]. Abbreviations: AI, primary auditory cortex; aSTG, anterior superior temporal gyrus; CS, circular sulcus; FTS, first transverse sulcus; H, high frequency; HG, Heschl's gyrus; HS, Heschl's sulcus; ITG, inferior temporal gyrus; L, low frequency; LS, lateral sulcus; PT, planum temporale; R, area R (rostral auditory cortex); RT, area RT (rostrotemporal auditory cortex); SI, intermediate sulcus; STG, superior temporal gyrus; STS, superior temporal sulcus. [From: Bendor and Wang , *Curr Opin Neurobiol* , 2006]

Why Do Our Voices Sound Different On Tape?

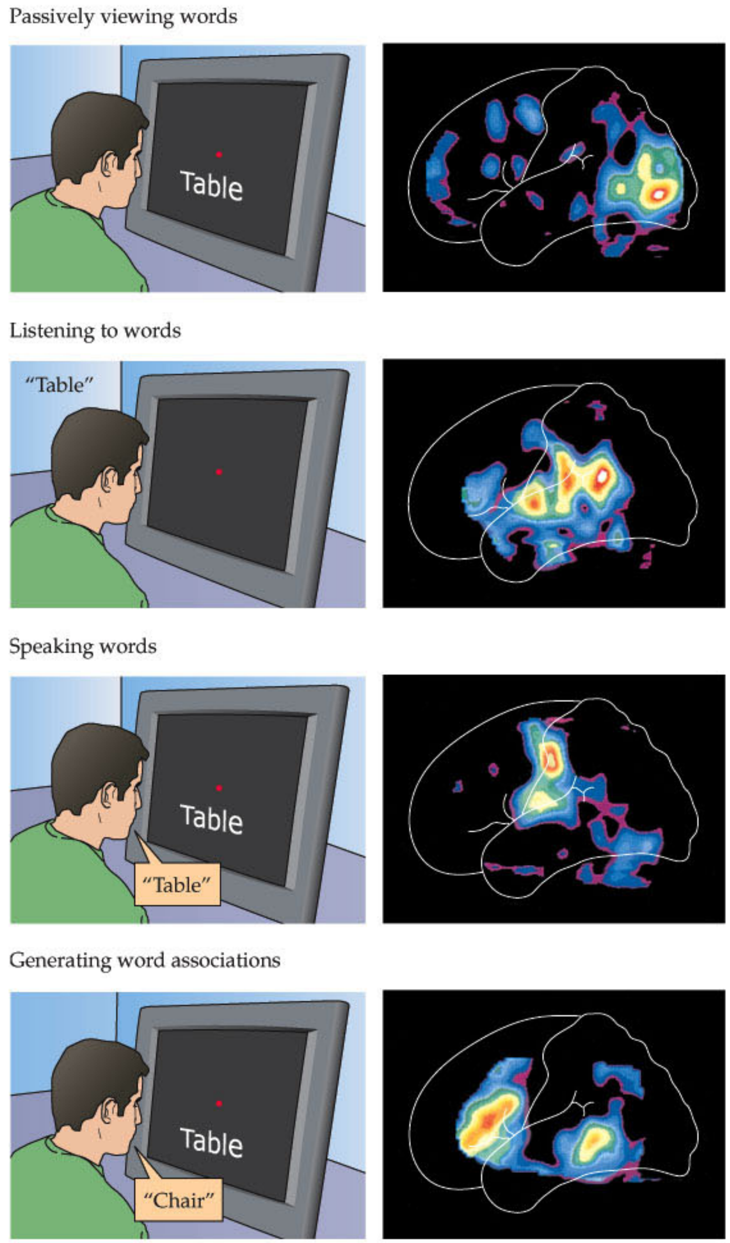
Figure 27.5 Variability of language representation among individuals





NEUROSCIENCE, Fourth Edition, Figure 27.5

Figure 27.6 Language-related regions of the left hemisphere mapped by PET



NEUROSCIENCE, *Fourth Edition*, Figure 27.6

Vocal feedback control:

Human listeners compensate for shifted formant frequency

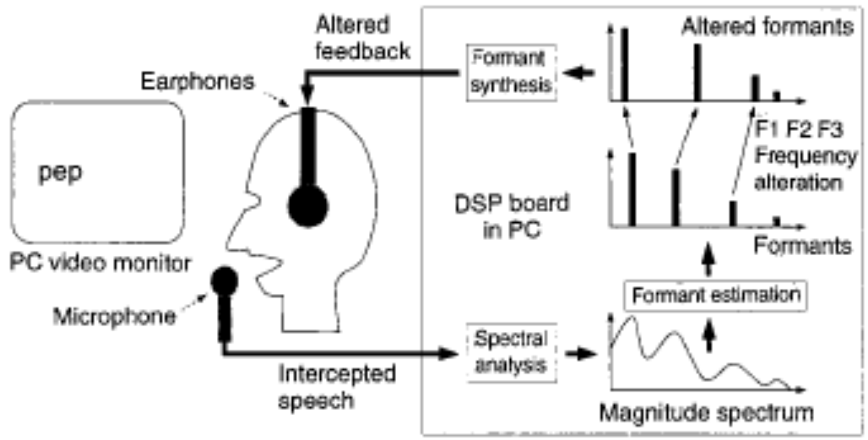


Fig. 1. The apparatus used in the experiments. CVC words were prompted on the personal computer (PC) video monitor. Subjects were instructed to whisper the word; we used whispered speech to minimize the effects of bone conduction which are strong in voiced speech. While the subject whispered, the speech signal was picked up by a microphone and sent to a digital signal processing (DSP) board in the PC. The DSP board processed successive intervals of the subject's speech into synthesized, formant-altered feedback with only a 16-ms processing delay [such a delay is nondisruptive; see reference to DAF in (2)]. Each interval was first analyzed into a 64-channel, 4 kHz–wide magnitude spectrum from which formants (which are generally peaks in the spectrum) were estimated (all graphs are schematic plots of magnitude versus frequency). The frequencies of the three lowest frequency formants (F1, F2, and F3) were then shifted to implement a desired feedback alteration (as explained below). The shifted formants were then used to synthesize formant-altered whispered speech. This synthesized speech was fed back to the subject via earphones at sufficient volume that he essentially heard only the synthesized feedback of his whispering.

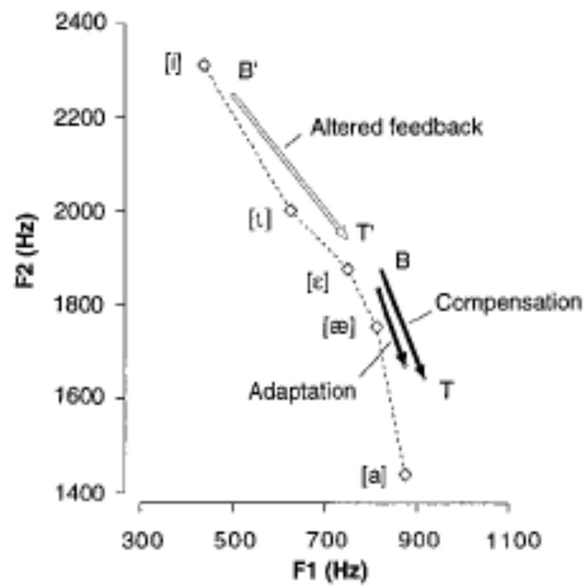


Fig. 2. Altered feedback and resulting compensation and adaptation for a single subject (subject OB).

Houde and Jordan (1998)

Reduced Response to One's Own Speech in Human Auditory Cortex

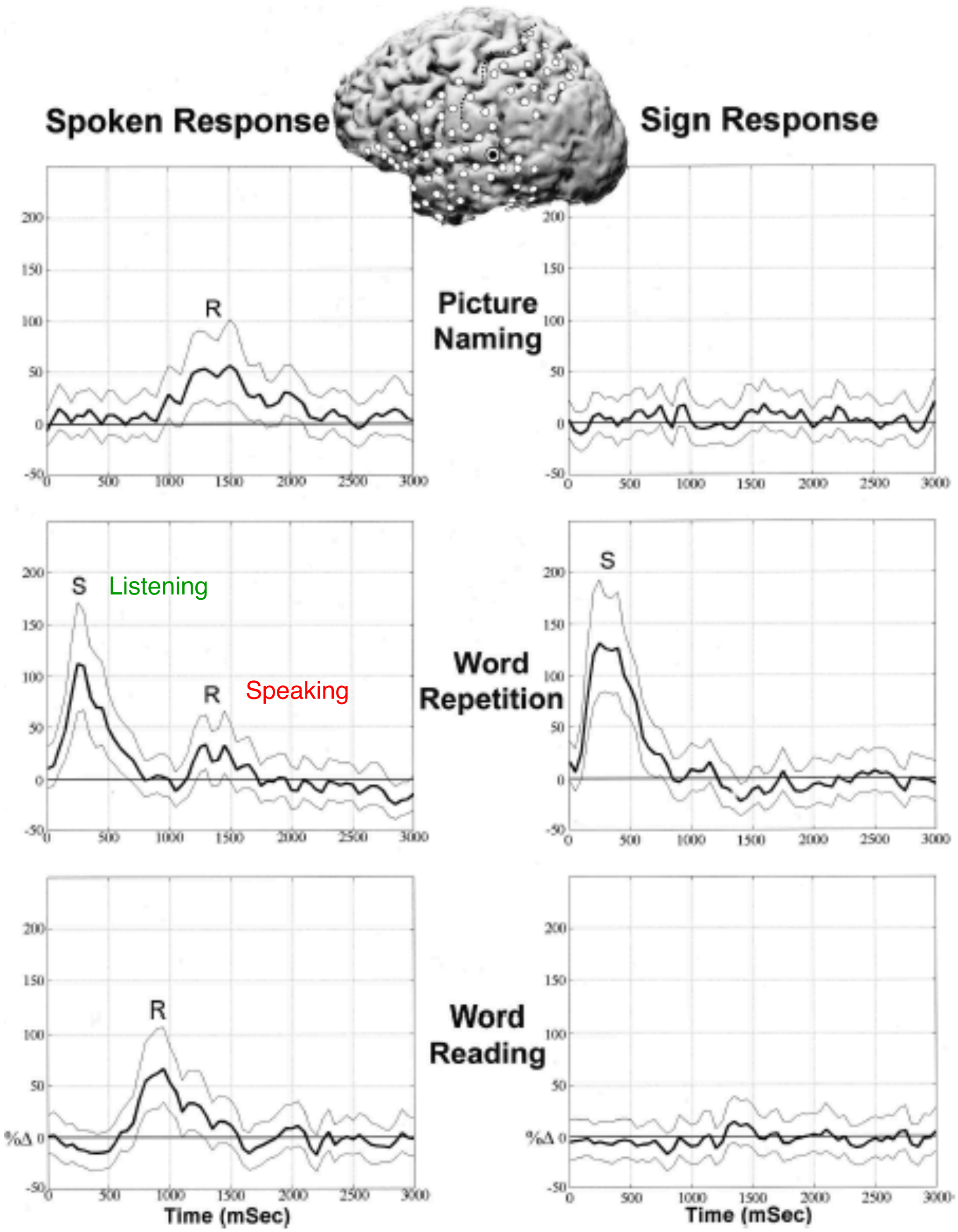


Figure 4. Plots of gamma (80- to-100-Hz) power augmentation in another electrode over posterior superior temporal gyrus, during three language tasks using either spoken or signed responses. Plotted units as in figure 4. Peaks of gamma augmentation during auditory stimuli are labeled "S," and peaks during verbal responses are labeled "R."

Crone et al. (2001)

Vocalization-induced inhibition in auditory cortex

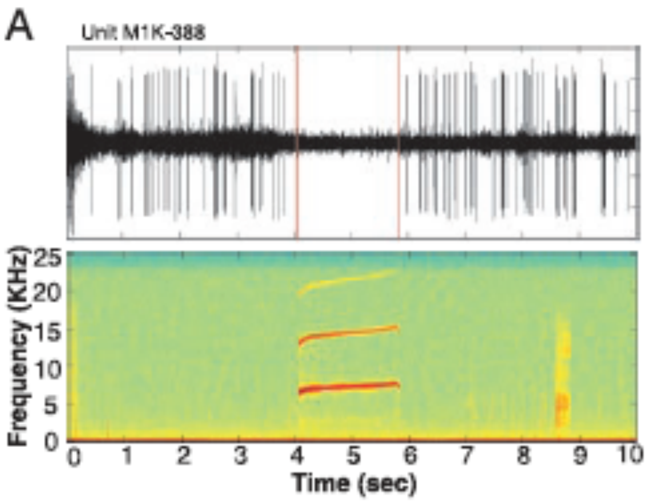


FIG. 1. Representative examples of the suppression of spontaneous neural activities in the auditory cortex by self-initiated vocalizations. *Top*: extracellular recording traces containing well-isolated single units; *bottom*: corresponding spectrograms of acoustic recording showing captured phee vocalizations. Two vertical red lines mark the onset and offset of each vocalization. *A*: a unit with high spontaneous activity is completely suppressed while the animal is vocalizing. The suppression appears to begin before the onset of the vocalization. *B*: a second example where 2 units are captured simultaneously

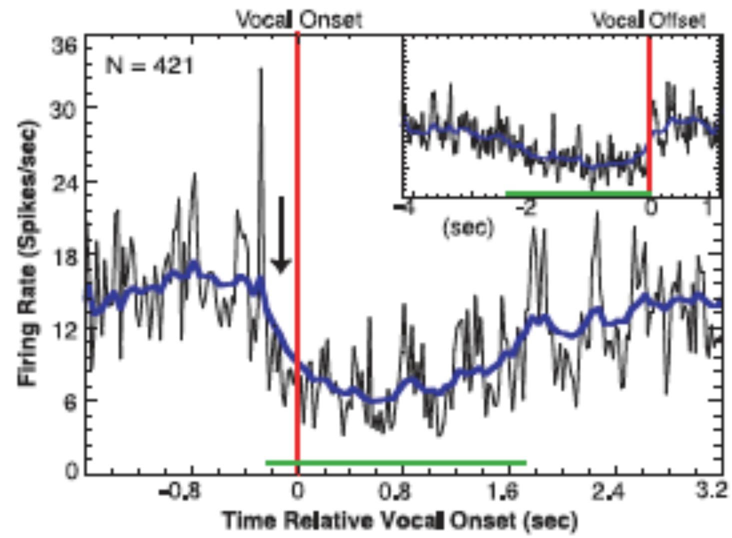
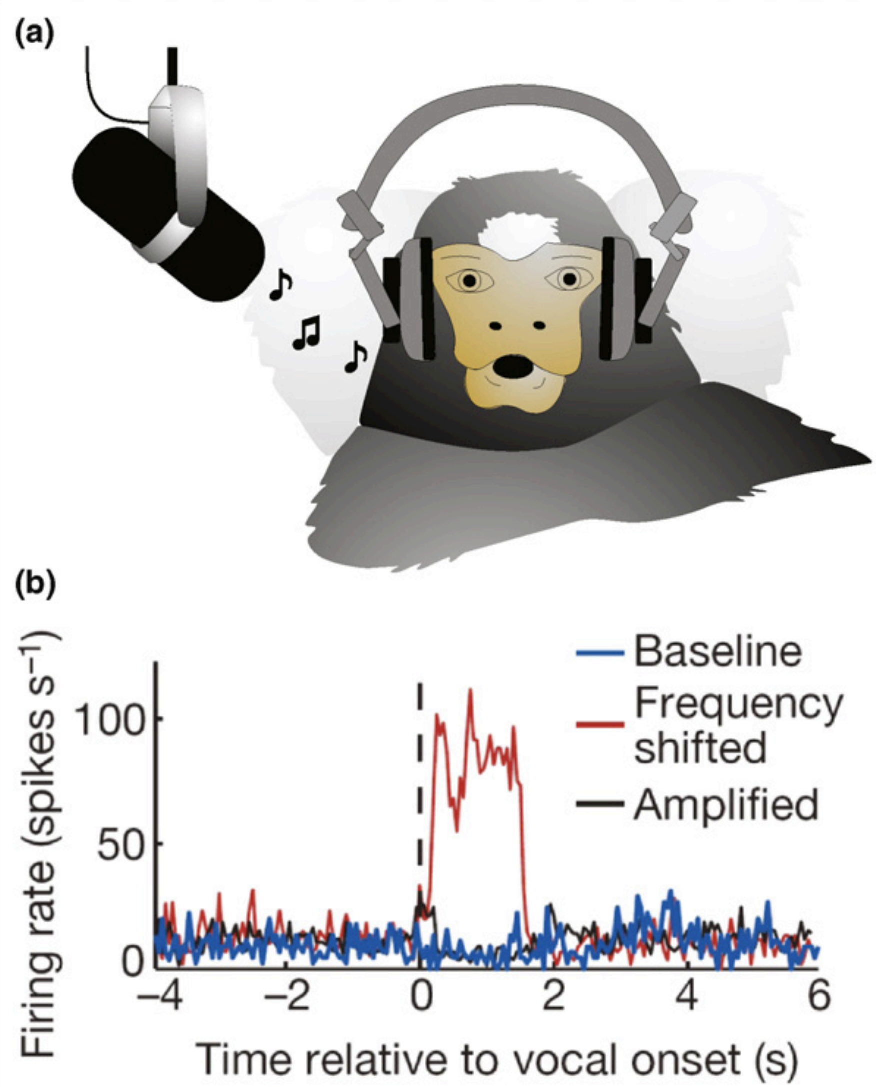


FIG. 4. Average suppressed vocalization response. A population histogram for all suppressed responses aligned by vocal onset is shown (binwidth = 20 ms). The vocal onset is indicated by a red line, and the time axis is referenced to this point. The blue line is a moving average (100-ms window) and shows that suppression begins prior to vocalization as indicated by an arrow. The green bar indicates the period over which suppression was continuously significant ($P < 0.05$). Suppression was significant starting 220 ms before vocal onset and remained until 1,730 ms after vocal onset. *Inset*: a population histogram aligned by vocal offset is shown. The green line indicates significant suppression ($P < 0.05$) that lasts until 20 ms after vocalization. The binwidth and the window size for the moving average (blue line) are the same in both plots.

Eliades and Wang (*J. Neurophysiology*, 2003)

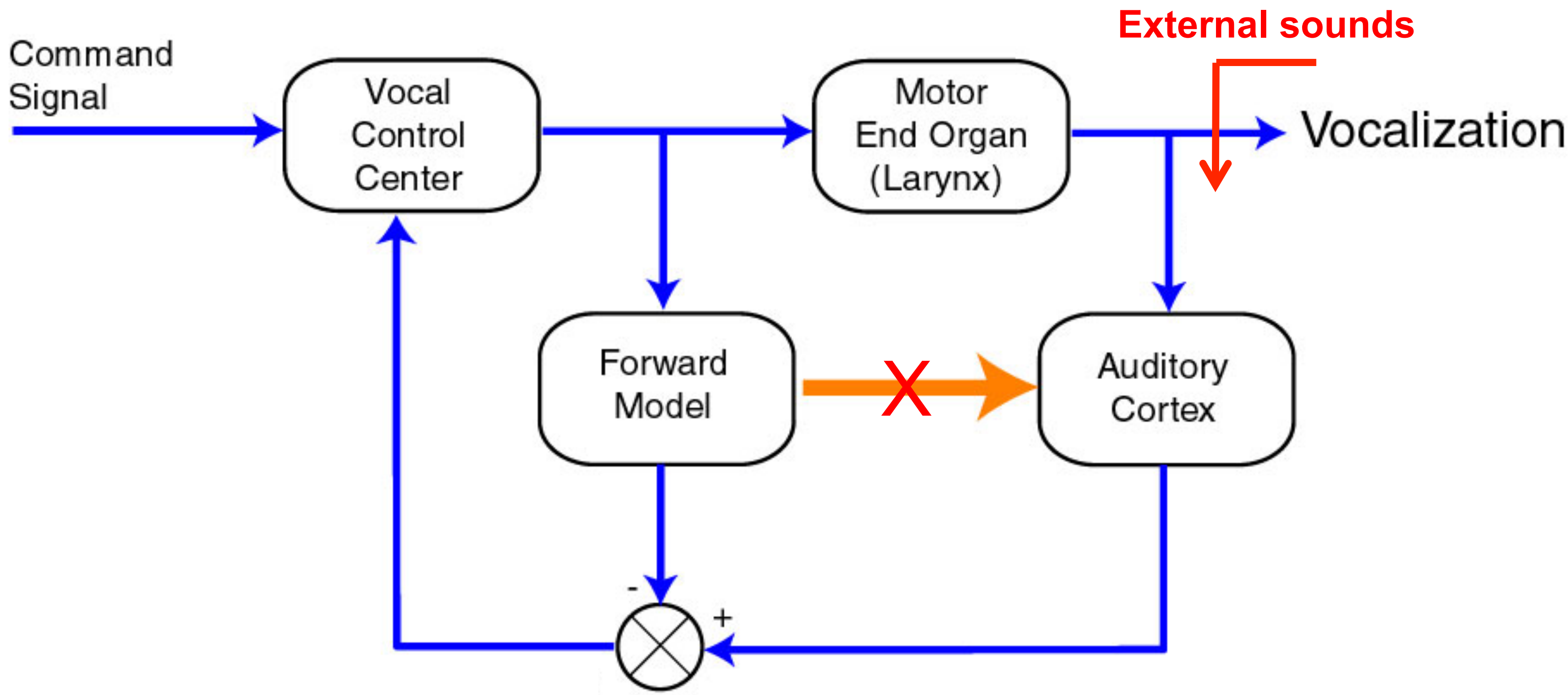
Altered Feedback Changes Vocal Suppression in Auditory Cortex



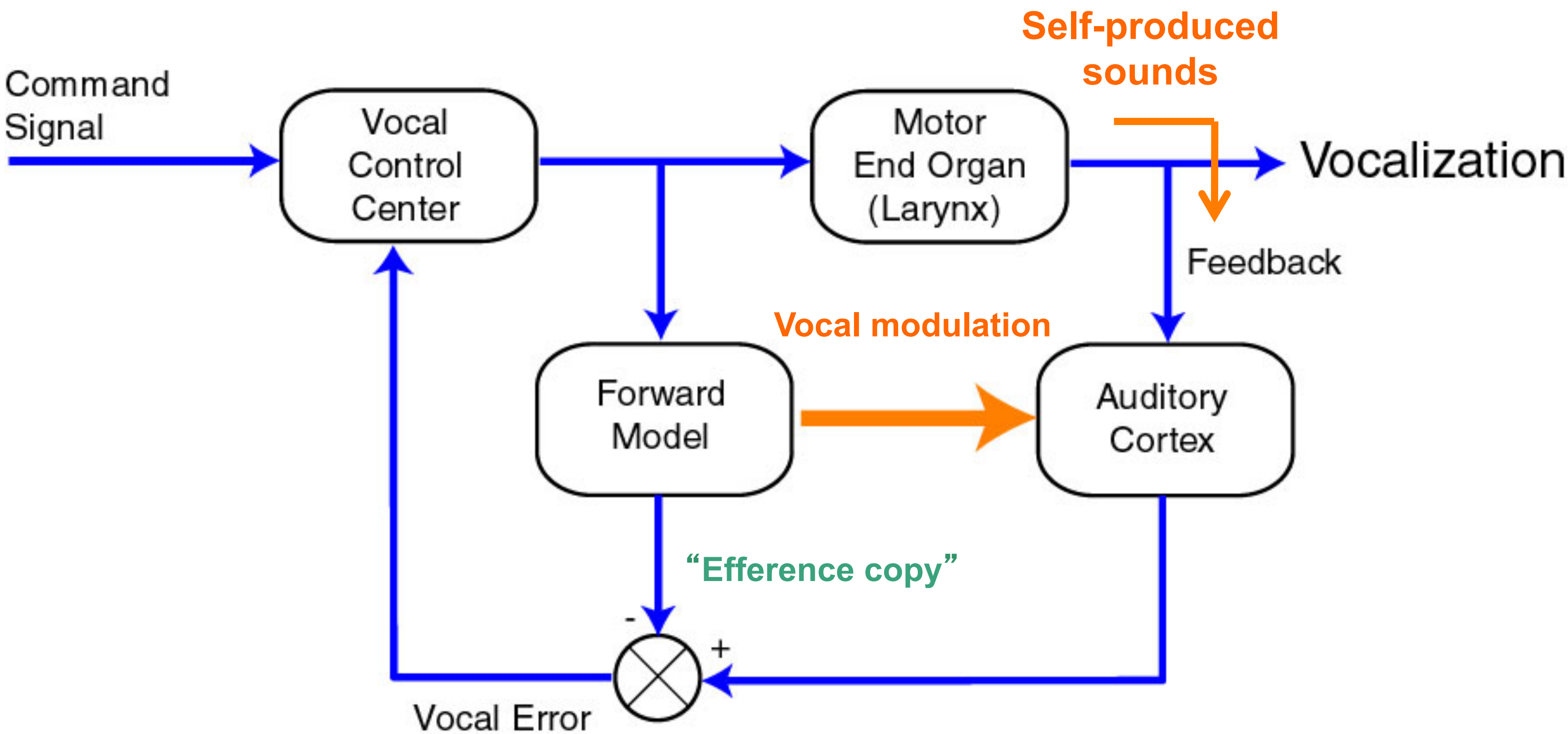
Effect of altered feedback on auditory cortex neurons. (a) 'Marmoset in a recording studio.' Investigators recorded from auditory cortical neurons as a marmoset vocalized. The auditory feedback from the vocalization was relayed to the marmoset without delay under both normal and altered conditions. (b) Peristimulus time histogram exhibiting the basic effect. This auditory cortical neuron was suppressed during normal vocalization (blue trace). However, when the auditory feedback of the vocalization was shifted in the frequency domain, the neuron exhibited a large increase in firing rate (red trace). The neuron's frequency tuning was transiently altered at the outset of the vocalization by a CD, rendering it sensitive to deviations from the expected auditory feedback, an example of a shifting RF in the auditory frequency domain.

Eliades and Wang (*Nature 2008*)

Auditory-vocal interactions in auditory cortex



Auditory-vocal interactions in auditory cortex



Why Do Voices Sound Different On Tape?

1) Only air conduction, no bone conduction
2) Vocal production system modulates auditory cortex during speaking

Auditory–motor interactions during musical performance

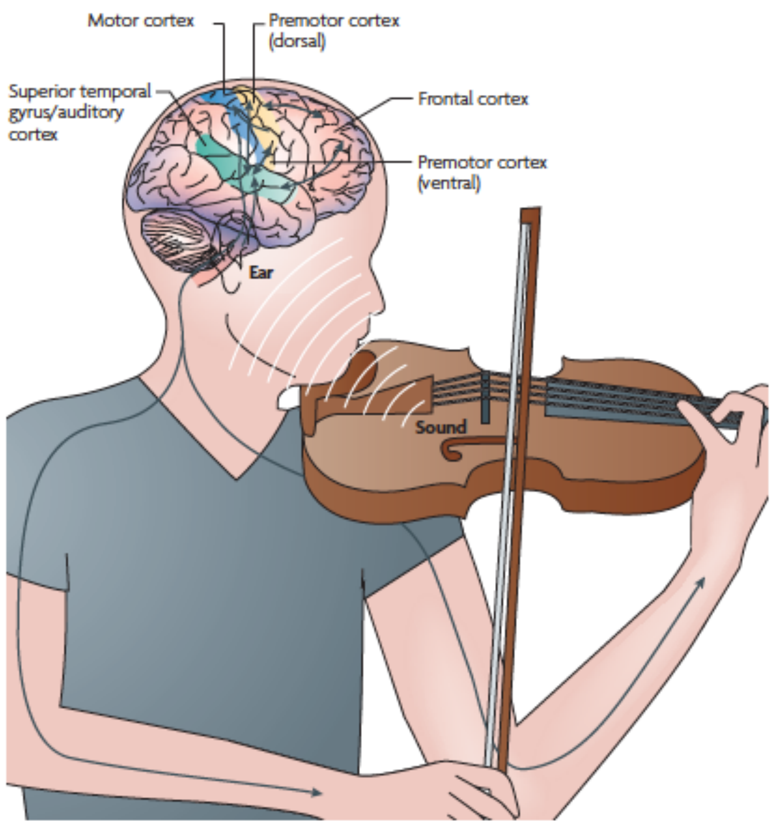


Figure 1 | **Auditory–motor interactions during musical performance.** This figure illustrates the feedback and feedforward interactions that occur during music performance. As a musician plays an instrument, motor systems control the fine movements needed to produce sound. The sound is processed by auditory circuitry, which in turn is used to adjust motor output to achieve the desired effect. Output signals from premotor cortices are also thought to influence responses within the auditory cortex, even in the absence of sound, or prior to sound; conversely, motor representations are thought to be active even in the absence of movement on hearing sound. There is therefore a tight linkage between sensory and production mechanisms.

Zatorre (Nature Review Neuroscience 2007)

Suggested readings:

1) “*Neuroscience*” textbook, Chapter 13: The auditory system

2) Lu, T., L. Liang and X. Wang. “Temporal and rate representations of time-varying signals in the auditory cortex of awake primates”. *Nature Neuroscience* 4:1131-1138 (2001)

3) Eliades, S.J. and X. Wang. Neural Substrates of Vocalization Feedback Monitoring in Primate Auditory Cortex. *Nature* 453: 1102-1106 (2008) .