11

Music and Speech Perception

Chapter 11 Music and Speech Perception

- Music
- Speech

Music

Music is a way to express thoughts and emotions.

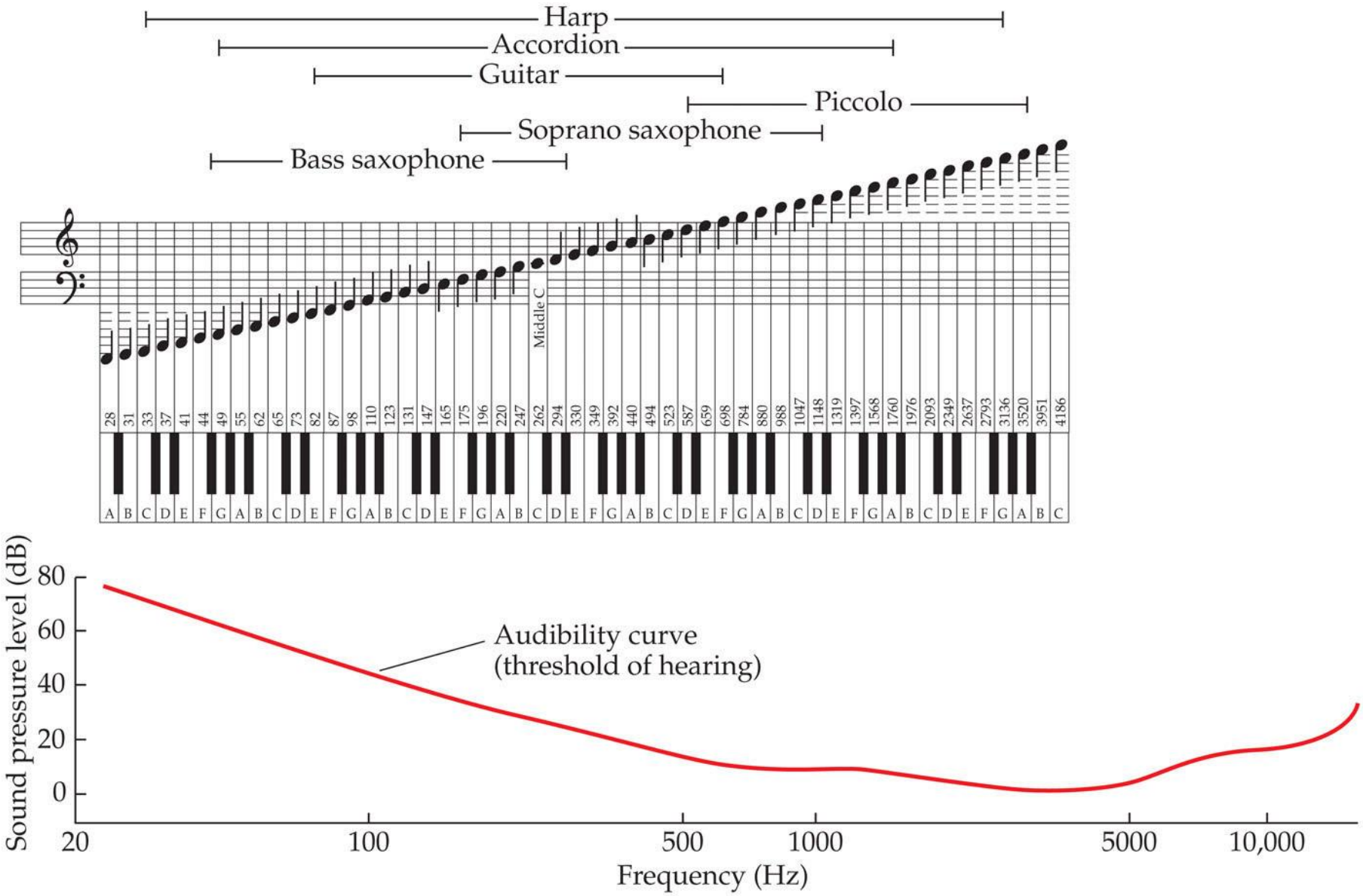
- Oldest known musical instruments are 30,000 year-old flutes carved from animal bones.
- Pythagoras was obsessed with numbers and music intervals.

Music

Musical notes

- Sounds of music extend across a frequency range from about 25 to 4500 Hz.
- Pitch: The psychological aspect of sounds related mainly to perceived frequency.

Figure 11.1 The sounds of music extend across a frequency range from about 25 to 4200 hertz



***SENSATION & PERCEPTION 4e*, Figure 11.1**

Music

Octave: The interval between two sound frequencies having a ratio of 2:1.

- Example: Middle C (C_4) has a fundamental frequency of 261.6 Hz; notes that are one octave from middle C are 130.8 Hz (C_3) and 523.2 Hz (C_5).
- C_3 (130.8 Hz) sounds more similar to C_4 (261.6 Hz) than to E_3 (164.8 Hz).
- There is more to musical pitch than just frequency!

Music

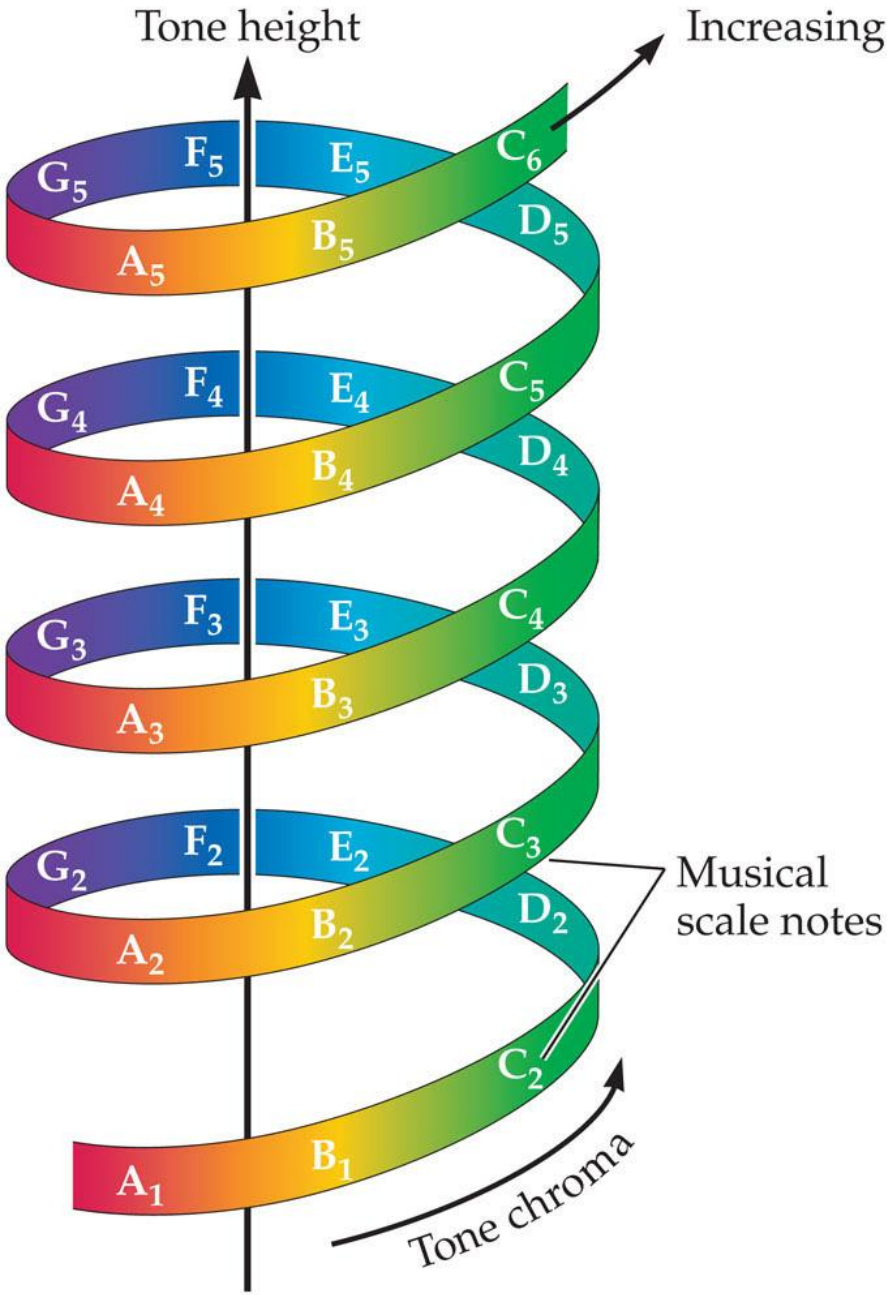
Tone height: A sound quality corresponding to the level of pitch. Tone height is monotonically related to frequency.

Tone chroma: A sound quality shared by tones that have the same octave interval.

- Each note on the musical scale (A–G) has a different chroma.

Musical helix—can help to visualize musical pitch.

Figure 11.2 A helix illustrating the two characteristics of musical pitch: tone height and tone chroma



***SENSATION & PERCEPTION 4e*, Figure 11.2**

Music

Musical instruments—most produce notes below 4000 Hz.

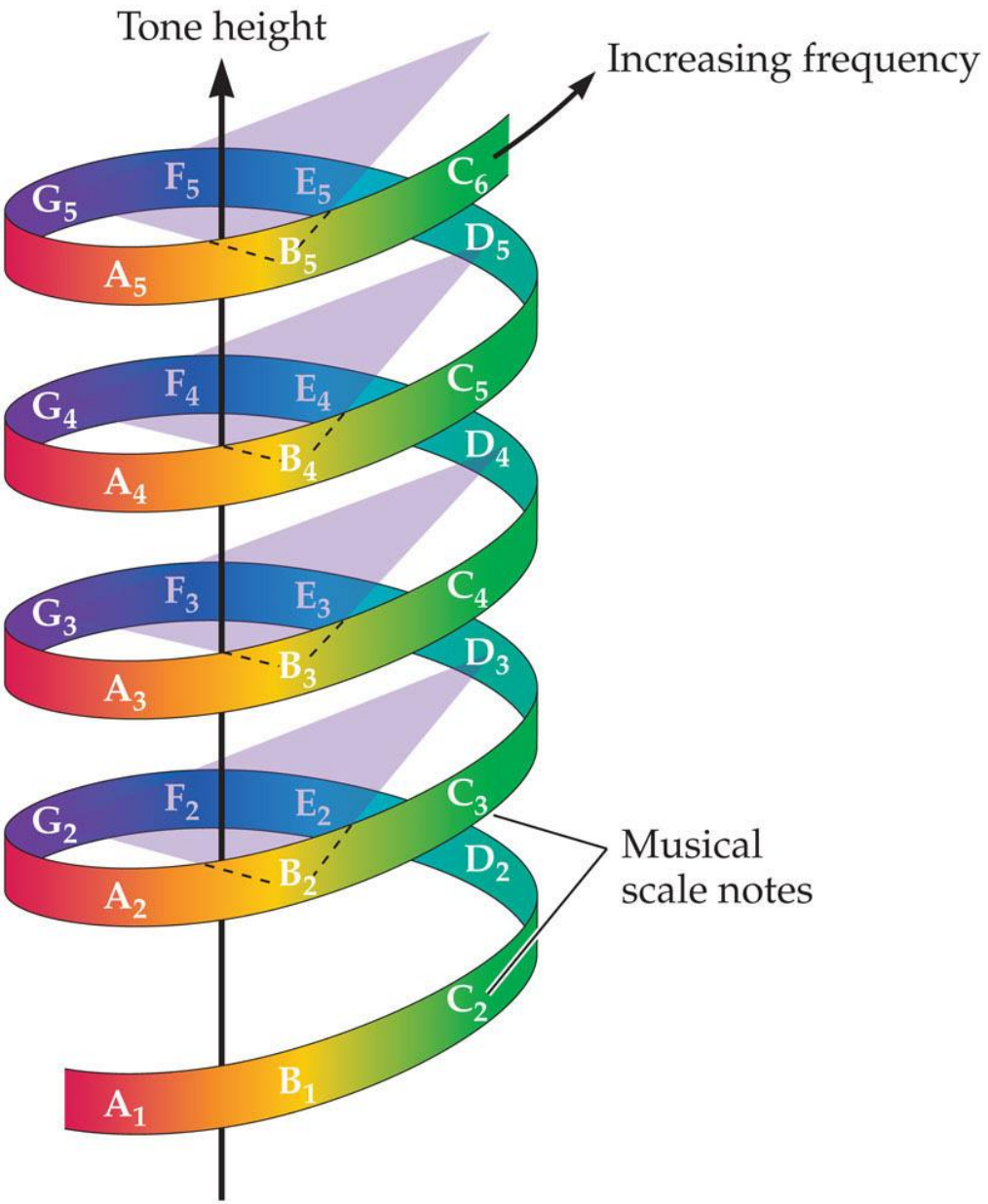
- Listeners have great difficulty perceiving octave relationships between tones when one or both tones are greater than 5000 Hz.

Music

Chords: Created when three or more notes with different pitches are played simultaneously.

- Can be consonant or dissonant
  - Consonant: chords with simple ratios of note frequencies.
  - Dissonant: chords with less elegant ratios of note frequencies.

Figure 11.3 Chords are made up of three or more notes and can be played with different tone heights while their chromatic relationships are maintained



***SENSATION & PERCEPTION 4e*, Figure 11.3**

Music

Cultural differences

- Some relationships between notes, such as octaves, are universal.
- Research on music perception: Western versus Javanese
  - Javanese culture
    - Fewer notes within an octave
    - Greater variation in note's acceptable frequencies

Music

- Western and Javanese musicians' estimates of intervals between notes correspond to the music scale from their culture.
- Six-month-old infants detect inappropriate notes in both scales but U.S. adults only detect deviations from the Western scale.

Music

Absolute pitch (AP): A rare ability whereby some people are able to very accurately name or produce notes without comparison to other notes.

- Highly prized skill among musicians
- Debate as to whether AP is due to nature or nurture
- More likely for people who begin musical training at a young age

Music and Emotion

- Music affects mood and emotions.
- Some clinical psychologists practice music therapy.
  - Example: Music can have a positive impact on pain, anxiety, mood, and overall quality of life for cancer patients.

Music

Melody: A sequence of notes or chords perceived as a single coherent structure.

- Examples: “Twinkle, Twinkle, Little Star” or “Baa Baa Black Sheep”
- Not a sequence of specific sounds but a relationship between successive notes
  - Melodies can change octaves or keys and still be the same melody even if they have completely different notes.

Music

Notes and chords vary in duration.

Tempo: The perceived speed of the presentation of sounds.

Music

Rhythm—not just in music!

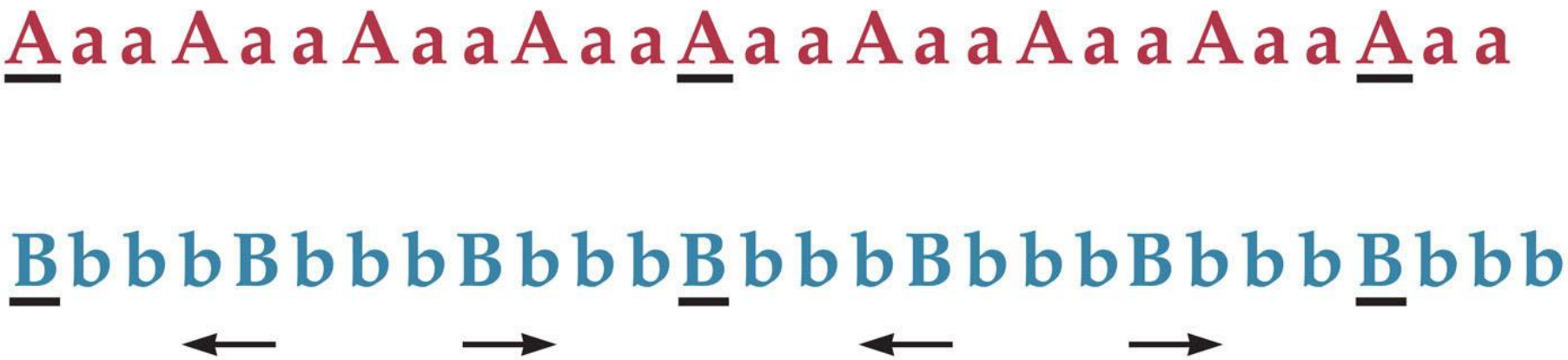
- Lots of activities have rhythm—walking, waving, swimming, finger tapping, etc.
- Bolton (1894)—experiments with identical sounds, perfectly spaced in time, but no rhythm
  - Listeners reported hearing first sound of group as “accented,” while the rest remained unaccented.
  - Examples of this phenomenon: Car and train rides

Music

Syncopation: Any deviation from a regular rhythm.

“Syncopated auditory polyrhythms”: When two rhythms are played together but slightly out of sync, one dominates.

Figure 11.5 When one rhythm is dominant, listeners tend to perceive the timing of beats in the nondominant rhythm adjusted to conform with the dominant rhythm



***SENSATION & PERCEPTION 4e*, Figure 11.5**

Melody development

- 8-month-olds—able to learn and recognize new melodies after hearing them for only 3 minutes
- 7-month-olds—able to learn and recognize differences between Mozart sonata movements after parents played the melodies at home every day for 2 weeks

Speech

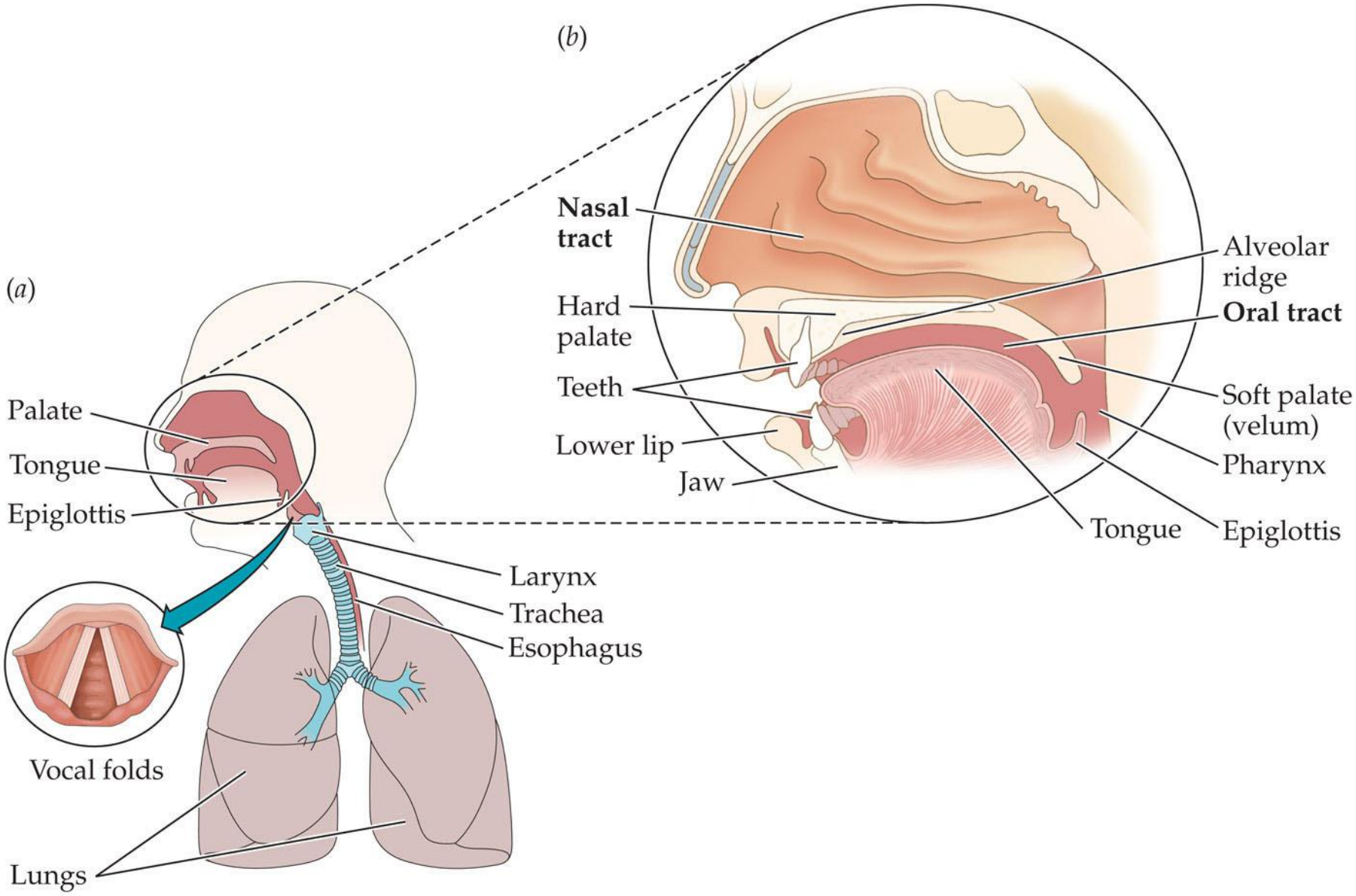
Humans are capable of producing many different speech sounds.

- About 5000 languages spoken today, utilizing over 850 different speech sounds

Vocal tract: The airway above the larynx used for the production of speech.

- Includes the oral tract and nasal tract
- Flexibility of vocal tract—important in speech production

Figure 11.6 The anatomy underlying speech production



***SENSATION & PERCEPTION 4e*, Figure 11.6**

Speech

Speech production

- Respiration (lungs)
- Phonation (vocal cords)
- Articulation (vocal tract)

Speech

Respiration and phonation

- Initiating speech—diaphragm pushes air out of lungs, through trachea, up to larynx.
- Phonation: The process through which vocal folds are made to vibrate when air pushes out of the lungs.

Respiration and phonation (*continued*)

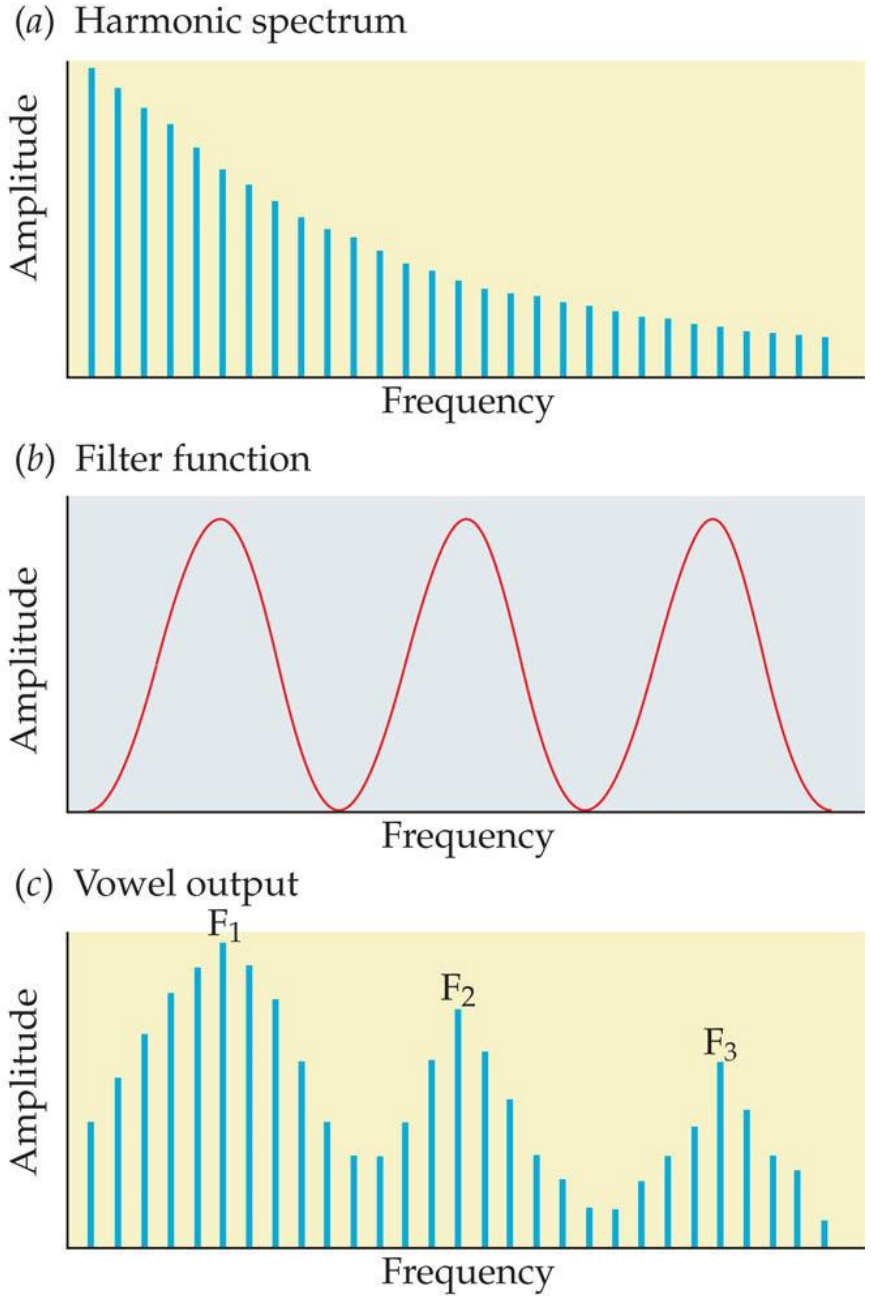
- At larynx—air must pass through two vocal folds.
  - Children: Small vocal folds, high-pitched voices
  - Adult men: Larger mass of vocal folds, low-pitched voices

Speech

Articulation: The act or manner of producing a speech sound using the vocal tract.

- Area above larynx: Vocal tract
- Humans can change the shape of their vocal tract by manipulating their jaws, lips, tongue body, tongue tip, and velum (soft palate).
  - These manipulations are articulation.
  - Resonance characteristics created by changing size and shape of vocal tracts to affect sound frequency distribution

Figure 11.7 Harmonic spectrum (a), filter function (b), and vowel output (c)



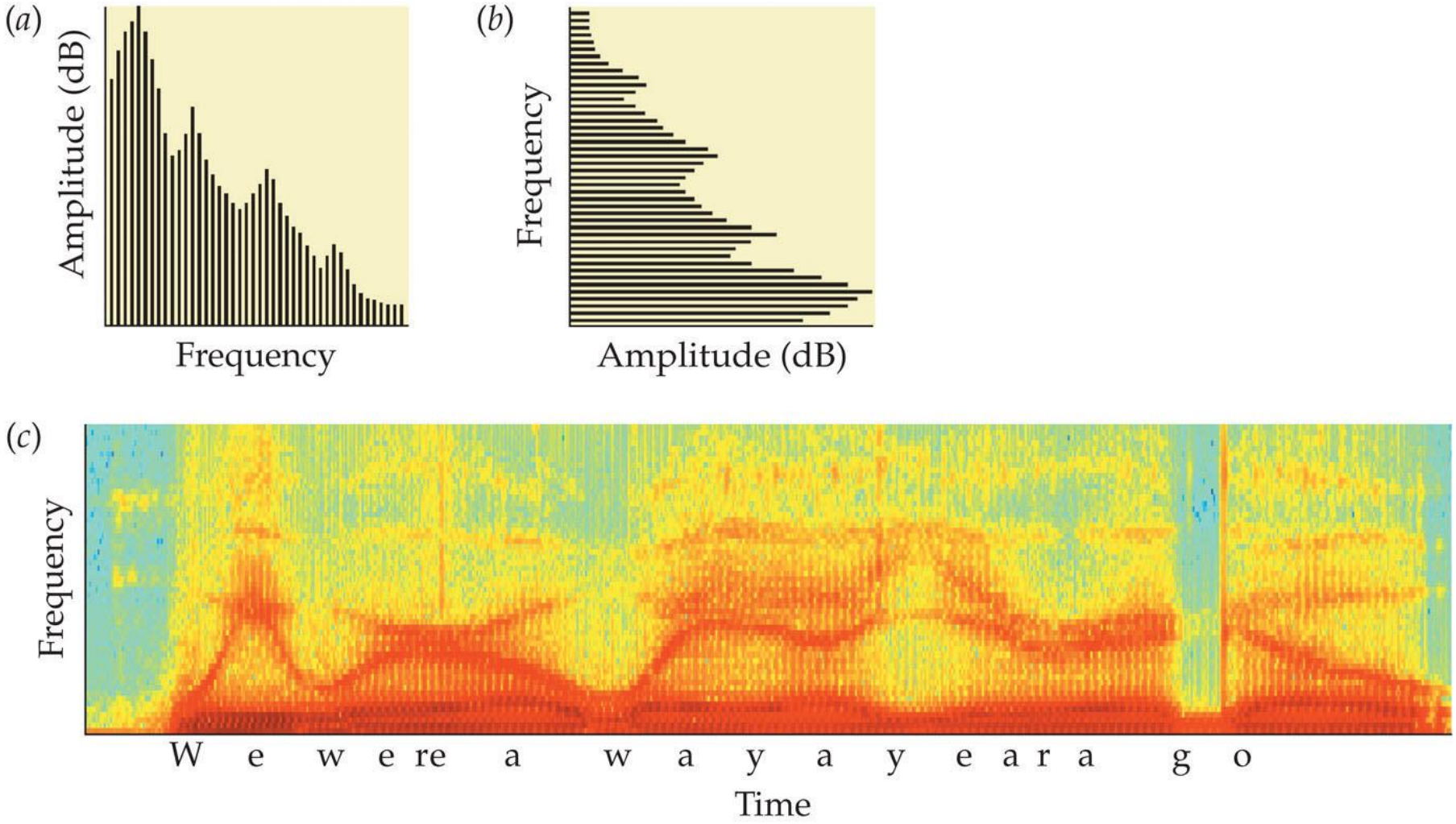
***SENSATION & PERCEPTION 4e*, Figure 11.7**

Speech

Formant: A resonance of the vocal tract that creates a peak in the speech spectrum.

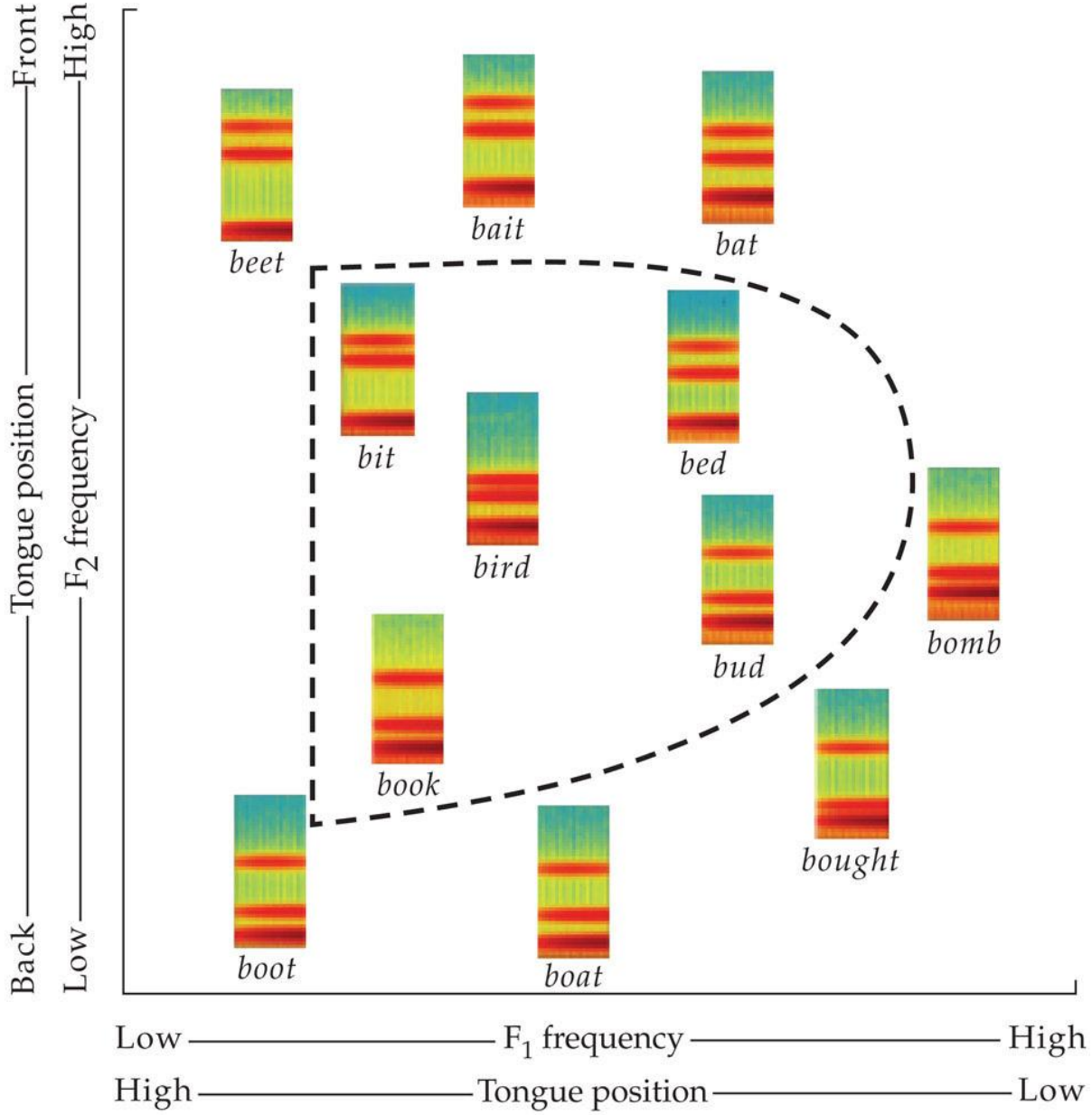
- Labeled by number, from lowest to highest (F_1, F_2, F_3)—concentrations in energy occur at different frequencies, depending on length of vocal tract
- Spectrogram: A pattern for sound analysis that provides a three-dimensional display plotting time on the horizontal axis, frequency on the vertical axis, and intensity in color or gray scale.

Figure 11.8 Sound spectrogram



***SENSATION & PERCEPTION 4e*, Figure 11.8**

Figure 11.9 Vowel sounds of English, and the frequencies of the first formant (F_1) and second formant (F_2) related to position of the tongue



***SENSATION & PERCEPTION 4e*, Figure 11.9**

Classifying speech sounds

- Speech sounds are most often described in terms of articulation.
  - Place of articulation (e.g., at lips, at alveolar ridge, etc.)
  - Manner of articulation (e.g., totally, partially, or slightly obstructed airflow)
  - Voicing: Whether the vocal cords are vibrating or not

Classifying speech sounds (*continued*)

- Some sounds are common across languages, and others, such as English ‘th’ and ‘r,’ are fairly uncommon.

Speech

Speech production of consonants

- Place of articulation. Airflow can be obstructed
  - At the lips (‘b,’ ‘p,’ ‘m’)
  - At the alveolar ridge (‘d,’ ‘t,’ ‘n’)
  - At the soft palate (‘g,’ ‘k,’ ‘ng’)

Speech production of consonants (*continued*)

- Manner of articulation. Airflow can be
  - Totally obstructed (‘b,’ ‘d,’ ‘g,’ ‘p,’ ‘t,’ ‘k’)
  - Partially obstructed (‘s,’ ‘z,’ ‘f,’ ‘v,’ ‘th,’ ‘sh’)
  - Only slightly obstructed (‘l,’ ‘r,’ ‘w,’ ‘y’)
  - First blocked, then open (‘ch,’ ‘j’)
  - Blocked at mouth but allowed to go through nasal passage (‘n,’ ‘m,’ ‘ng’)

Speech production of consonants (*continued*)

- Voicing. The vocal cords may be
  - Vibrating (‘b,’ ‘m,’ ‘z,’ ‘l,’ ‘r’)
  - Not vibrating (‘p,’ ‘s,’ ‘ch’)

Speech

Speech production—very fast

- 10–15 consonants and vowels per second
- Coarticulation: The phenomenon in speech whereby attributes of successive speech units overlap in articulatory or acoustic patterns.
  - Inertia prevents tongue, lips, jaw, etc. from moving too fast.
  - Experienced talkers position tongue, etc. in anticipation of next consonant or vowel, causing coarticulation.

Speech perception

- Computer programs—coarticulation presents a major challenge
  - How to recognize the ‘d’ sound in *deem, doom,* and *dam* as the same when it is different every time because of coarticulation?
  - Computers have to process all possibilities. Increased processing power improves speech recognition greatly (e.g., Siri on iPhone).

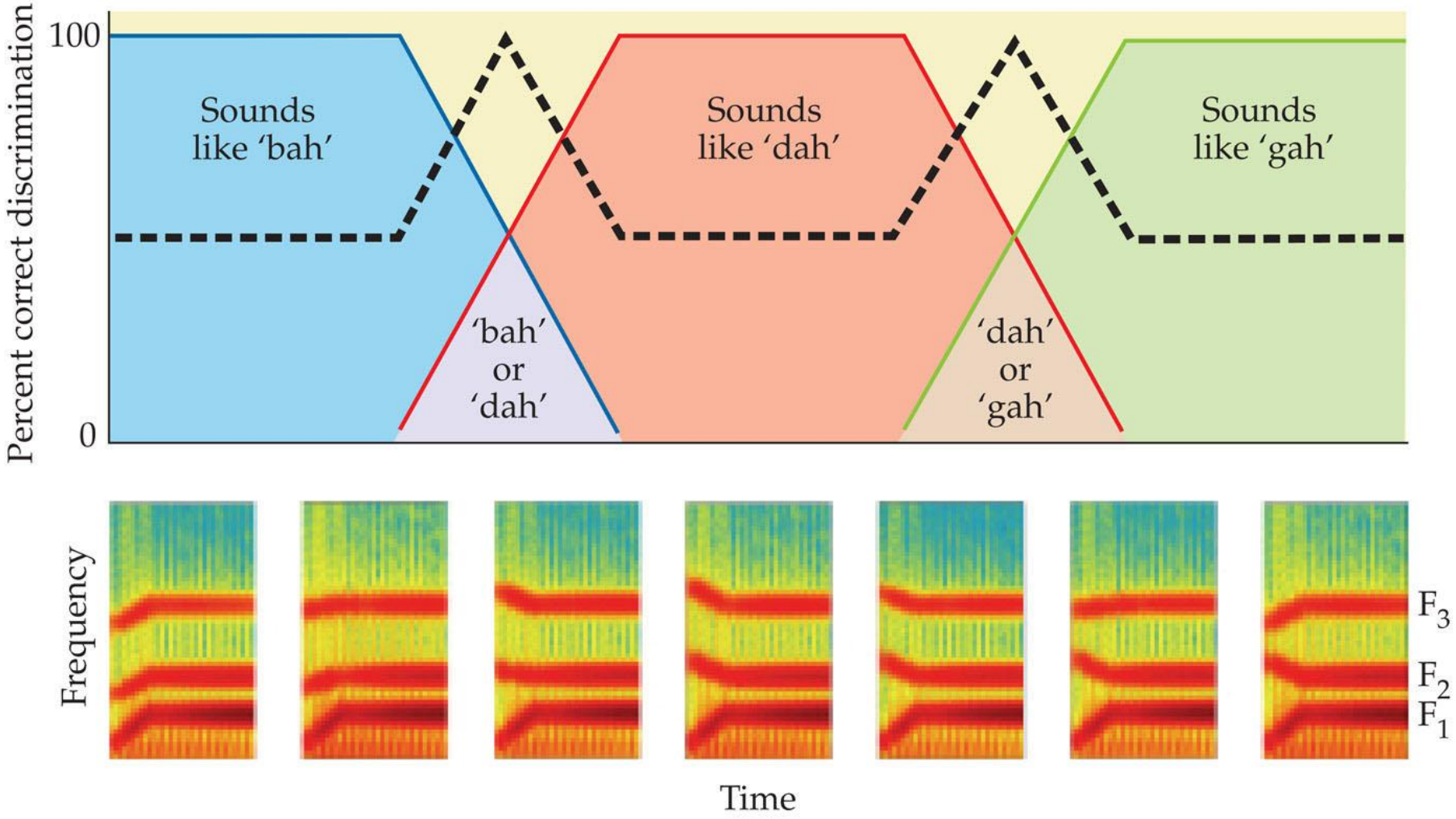
Speech perception (*continued*)

- How do humans recognize sounds despite coarticulation?

Categorical perception

- Researchers can manipulate sound stimuli to vary continuously from ‘bah’ to ‘dah’ to ‘gah.’
- However, people do not perceive the sounds as continuously varying.
- Instead, people perceive sharp categorical boundaries between the stimuli—categorical perception.

Figure 11.13 The sound spectrograms indicate auditory stimuli that change smoothly from a clear ‘bah’ on the left through ‘dah’ to a clear ‘gah’ on the right



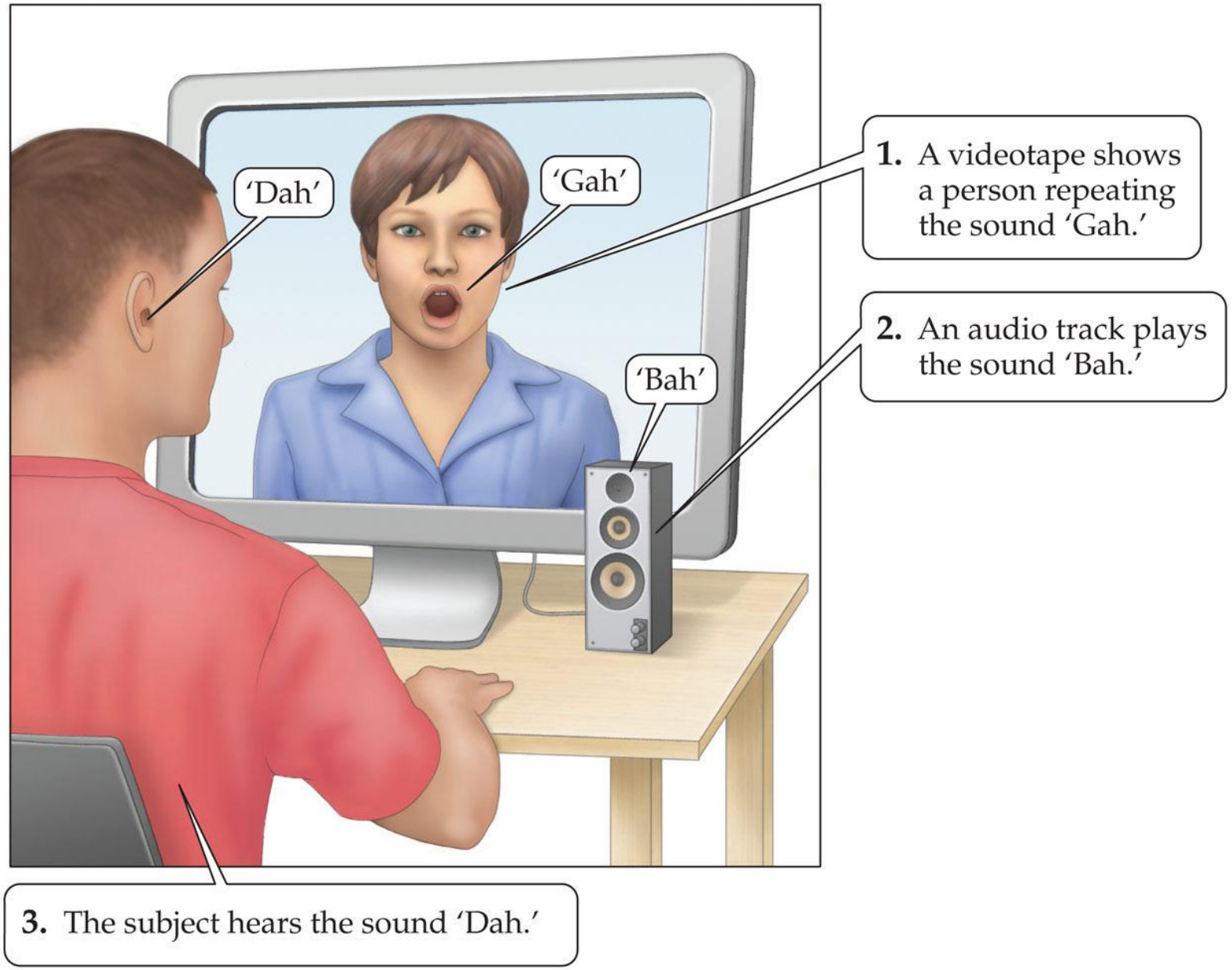
***SENSATION & PERCEPTION 4e*, Figure 11.13**

Speech

How special is speech?

- “Motor theory” of speech perception: Motor processes used to produce speech sounds are used in reverse to understand the acoustic speech signal.
- Supported by the McGurk Effect
  - McGurk and MacDonald (1976) showed that what someone sees can affect what they hear.

Figure 11.14 The motor theory was bolstered by, among other things, a somewhat peculiar finding by McGurk and MacDonald



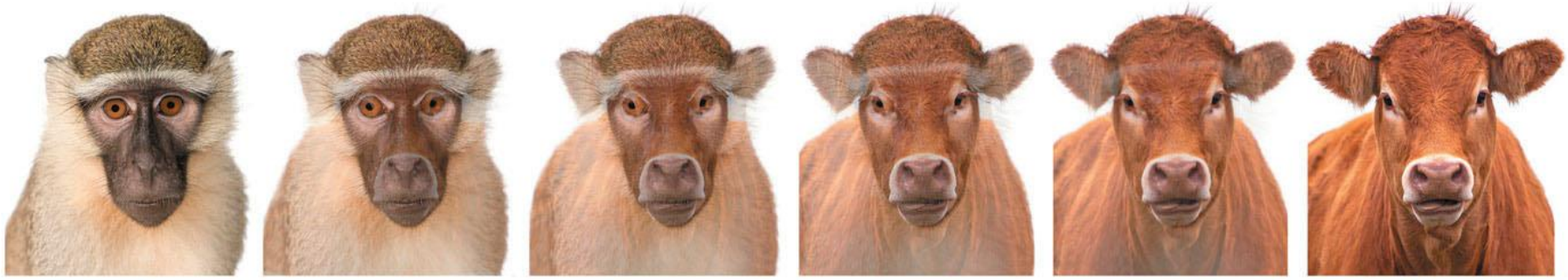
***SENSATION & PERCEPTION 4e*, Figure 11.14**

Speech

Problems for the motor theory of speech perception

- Speech production is just as complex, so appealing to production to understand perception doesn’t help much.
- Nonhuman animals can learn to respond to speech signals.
- Categorical perception also occurs in the perception of musical intervals, faces, and facial expressions.

Figure 11.16 People categorically perceive changes between images of familiar animals such as monkeys and cows



***SENSATION & PERCEPTION 4e*, Figure 11.16**

Speech

Coarticulation and spectral contrast

- Research has focused on understanding speech perception in terms of general ways that hearing and perception work.
- Example: Perception of coarticulated speech is explained by the fundamental ways the auditory system enhances contrast between successive sounds.
- Contrast enhancement is a general property of perception and occurs in many forms.

Using multiple acoustic cues

- Perception depends on experience
- Similar to face recognition
  - Similar features can be used in different combinations, and we can rely on multiple cues to recognize a face.

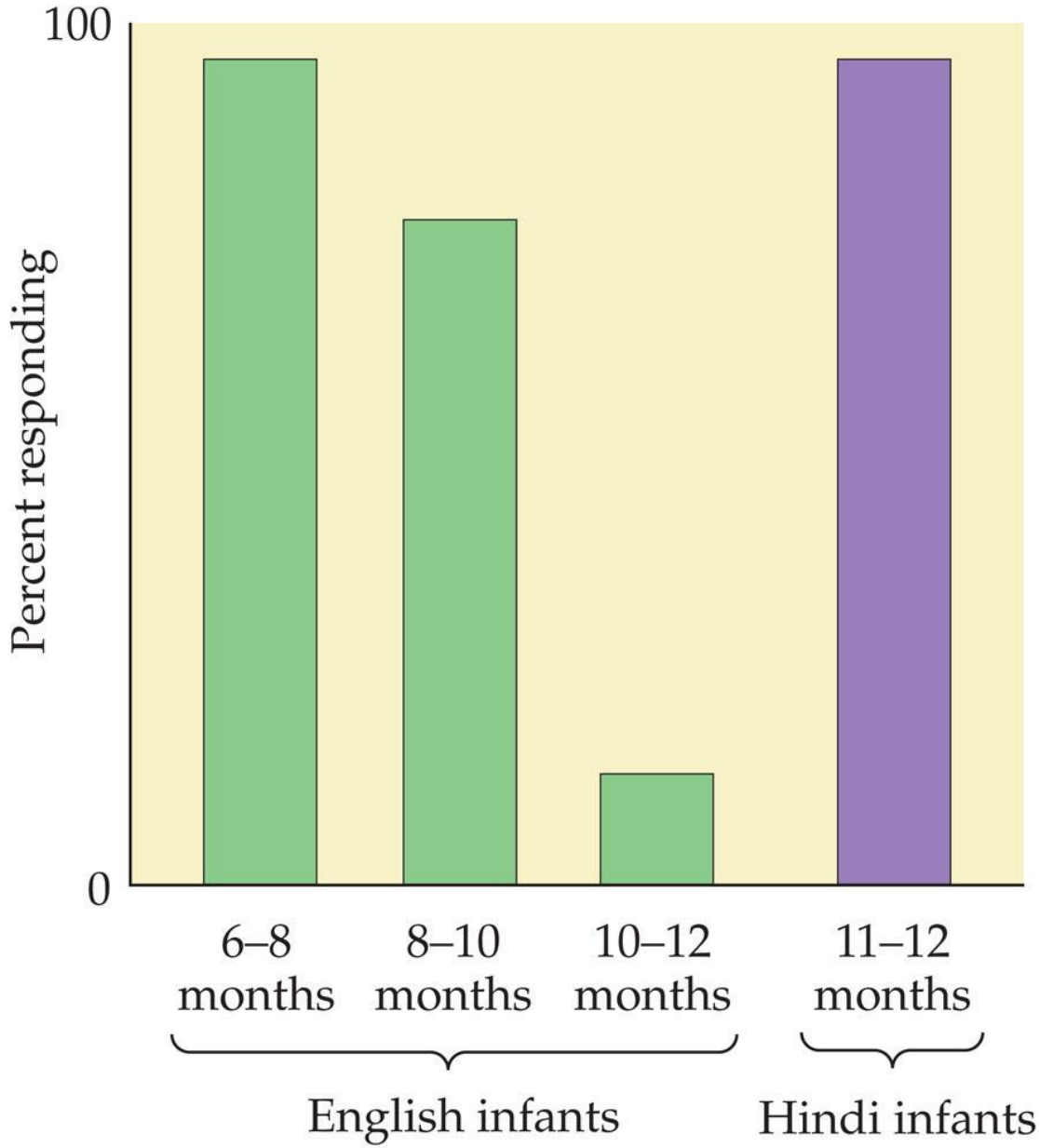
Learning to listen

- Babies learn to listen even before they are born!
- Prenatal experience: Newborns prefer hearing their mother's voice over other women's voices.
- Four-day-old French babies prefer hearing French over Russian.
- Newborns prefer hearing children's stories that were read aloud by their mothers during their third trimester of pregnancy.

Becoming a native listener

- Sound distinctions are specific to various languages.
- Example: ‘r’ and ‘l’ are not distinguished in Japanese.
- Infants begin filtering out irrelevant acoustics long before they start to say speech sounds.

Figure 11.20 Hindi has a dental stop consonant produced with the tongue tip touching the teeth and a retroflex stop consonant that requires the tongue to bend up and back in the mouth



***SENSATION & PERCEPTION 4e*, Figure 11.20**

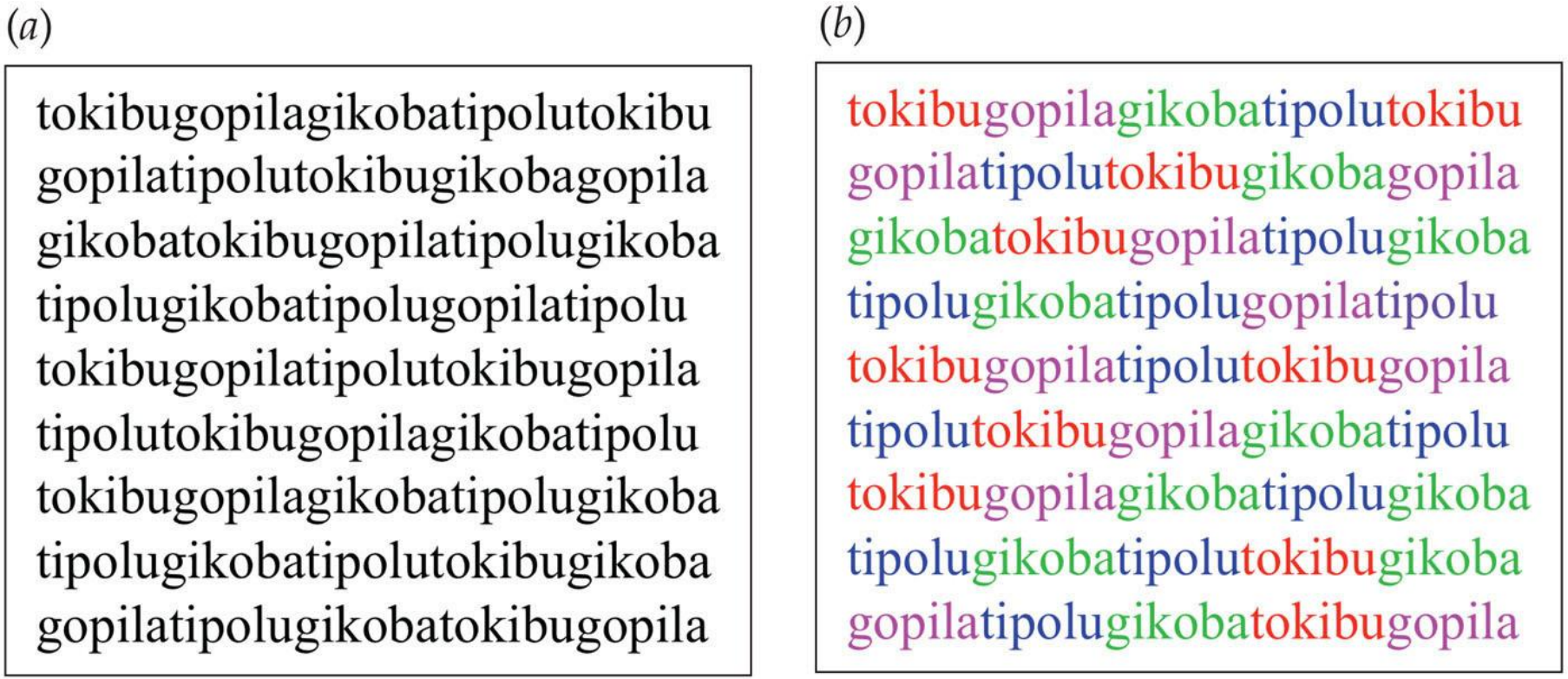
Learning words

- How do we know where one word ends and another begins?
- Research by Saffran, Aslin, and Newport (1996)
  - Created a novel language and infants listened to sentences for two minutes.
  - Afterwards, infants could already distinguish between words and non-words in the novel language.

Learning words (*continued*)

- Statistical learning: Certain sounds (making words) are more likely to occur together and babies are sensitive to those probabilities.

Figure 11.22 Eight-month-old infants can learn to pick out words from streams of continuous speech based on the extent to which successive syllables are predictable or unpredictable



***SENSATION & PERCEPTION 4e*, Figure 11.22**

Speech in the brain

- Brain damage follows patterns of blood vessels, not brain function, so is difficult to study.
- PET and fMRI studies help us learn about speech processing in the brain.

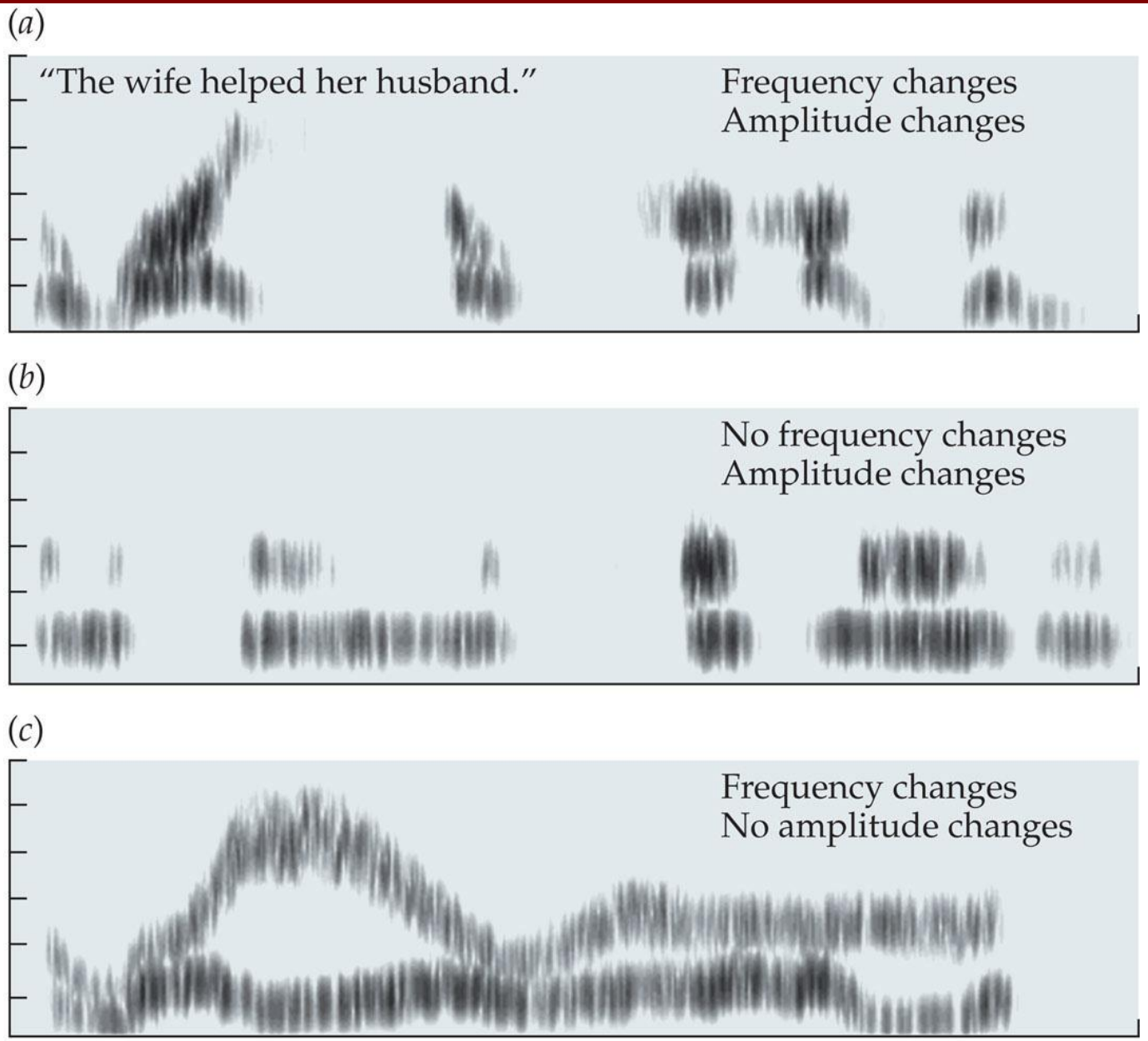
Speech in the brain (*continued*)

- Listening to speech: Left and right superior temporal lobes are activated more strongly in response to speech than to nonspeech sounds.
  - Hard to create well-controlled nonspeech stimuli because humans are so good at understanding even severely distorted speech.

Speech in the brain (*continued*)

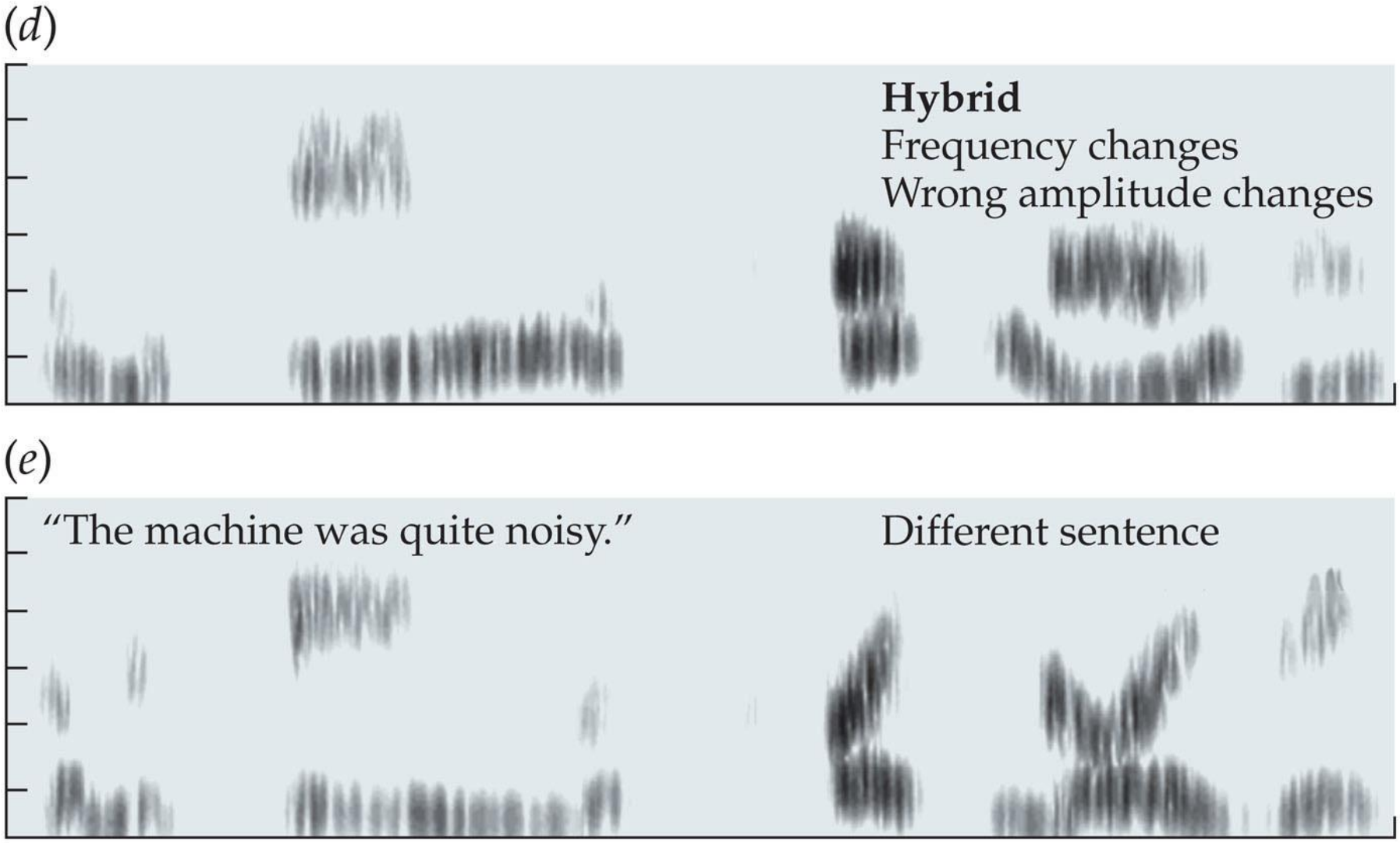
- Categorical perception tasks: Listeners attempt to discriminate sounds like ‘bah’ and ‘dah’ while having their brain scanned.
- As sounds become more complex, they are processed by more anterior and ventral regions of superior temporal cortex.
- Research indicates that some “speech” areas become active when lip-reading.

Figure 11.24 Stimuli created to measure cortical responses to acoustic complexity versus responses to speech (Part 1)



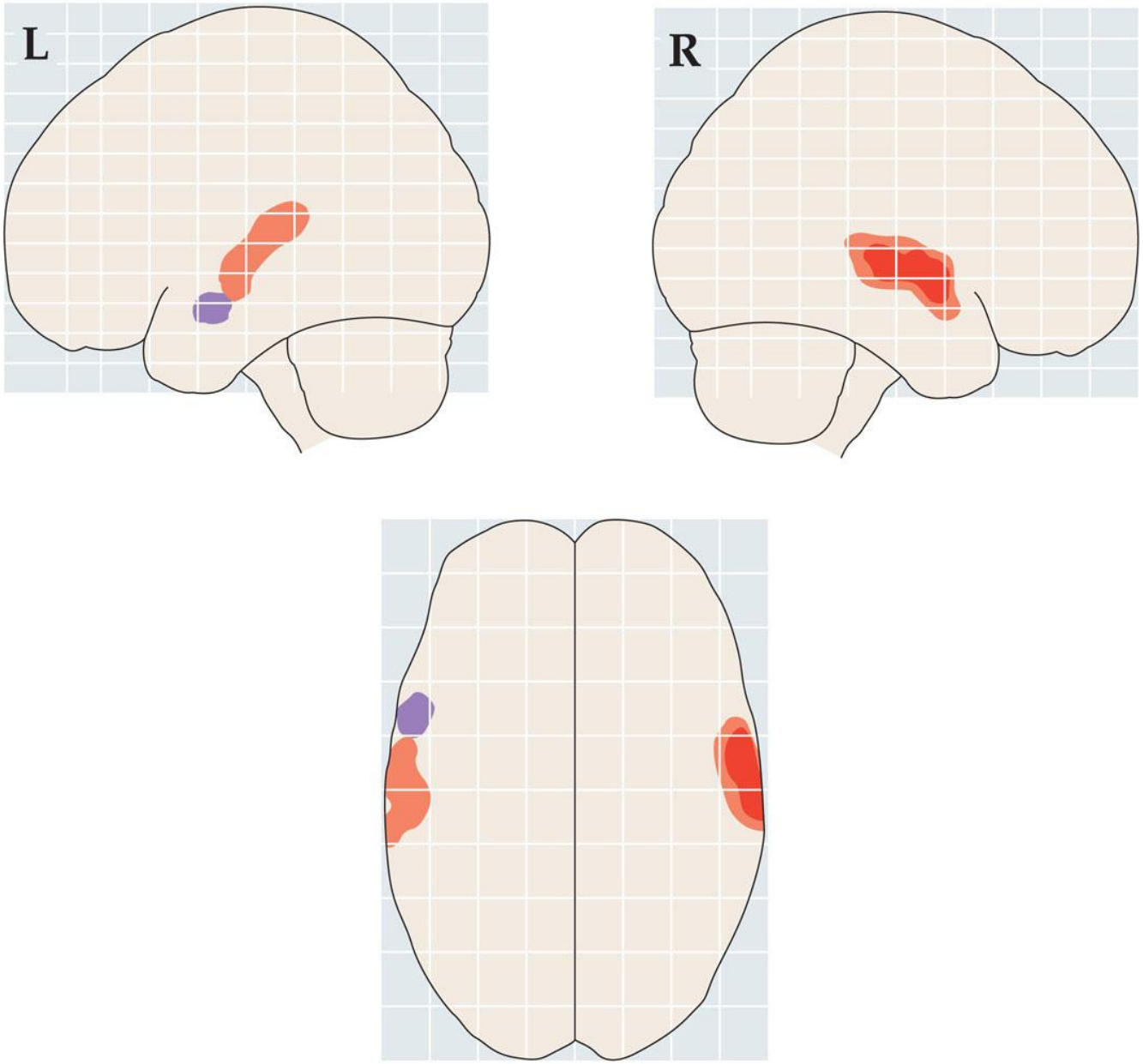
***SENSATION & PERCEPTION 4e*, Figure 11.24 (Part 1)**

Figure 11.24 Stimuli created to measure cortical responses to acoustic complexity versus responses to speech (Part 2)



***SENSATION & PERCEPTION 4e*, Figure 11.24 (Part 2)**

Figure 11.25 Substantial neural activity was observed in both left and right superior temporal lobes when listeners heard hybrid sentences



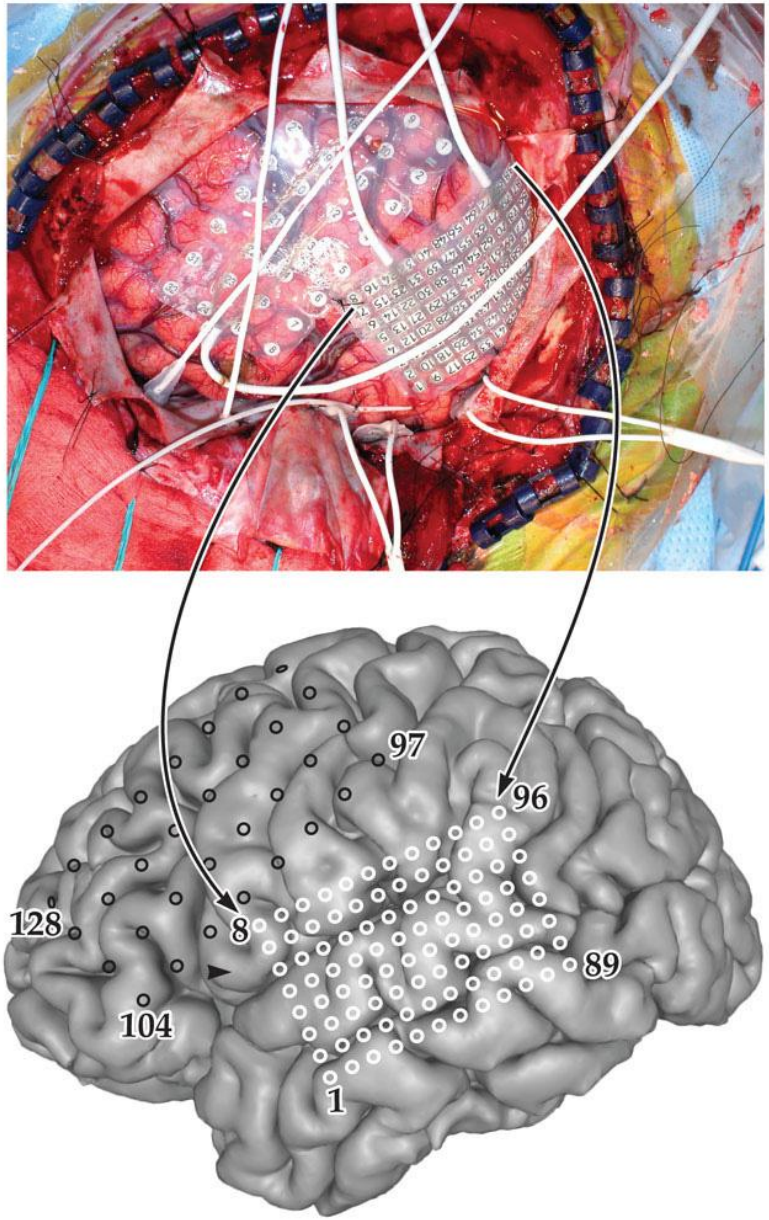
***SENSATION & PERCEPTION 4e*, Figure 11.25**

Speech

Sometimes, electrical recordings are taken directly from human brains prior to surgery.

- Electrodes are implanted in cortex to determine the function of certain areas before surgery.

Figure 11.26 Prior to performing brain surgery neurosurgeons often place electrodes directly on the surface of the brain to localize regions of neural activity



***SENSATION & PERCEPTION 4e*, Figure 11.26**

Speech

Neural responses in the brain matched behavioral responses by the subjects.

- Sounds that people labeled as the same had the same neural responses.
- Sounds that people labeled as different had different neural responses.