Quantitative approaches to the study of neural coding

©Bertrand Delgutte, 2003-2005
The Problem

• Neural and behavioral data have different forms that are not easily comparable
  – In some cases, data from different species and states (awake vs. anesthetized) are being compared
  – Even when neural and behavioral data have similar forms (e.g. a threshold), they both depend on arbitrary criteria

• Both neural and behavioral responses are probabilistic: They are not identical for different repetitions of the same stimulus. Therefore, they provide a noisy (imprecise) representation of the stimulus.

• Are there quantitative tools to describe the precision of stimulus representation in neural and behavioral data?
Questions that can be addressed by quantitative methods

- Does a particular class of neurons provide sufficient information to account for performance in a particular behavioral task?
- Does a particular neural code (e.g. rate-place vs. temporal) provide sufficient information for the task?
- Which stimulus feature is a particular neuron or neural population most sensitive to? Are there “feature detector” neurons?
- Is there stimulus information in the correlated firings of groups of neurons?
- How efficient is the neural code?
Outline

• Signal detection theory (a.k.a. ideal observer analysis)
  – Single neuron
  – Combined performance for neural population

• Shannon information theory
Neural variability limits detection performance

• When a sound stimulus is presented repeatedly, the number of spikes recorded from an auditory neuron differs on each trial. The blue and red surfaces are model spike count distributions for two pure tones differing in intensity by 3 dB.

• The overlap in the spike count distributions limits our accuracy in identifying which of the two stimuli was presented based on the responses of this neuron.

• A measure of the separation between the two distributions (and therefore of the neuron’s ability to discriminate) is the discriminability index $d'$, which is the difference in means of the two distributions divided by their standard deviation.

• The just noticeable difference (JND) or difference limen (DL) is often taken as the intensity increment for which $d' = 1$. This criterion corresponds to 76% correct in a two-interval, two-alternative psychophysical experiment.

\[
d' = \frac{\Delta N}{\sigma_N} \approx 1
\]
Conditional Probability and Bayes’ Rule

**Conditional Probability:**

\[ P(S \mid R) = \frac{P(S, R)}{P(R)} \]

**Statistically Independent Events:**

\[ P(S, R) = P(S) P(R) \]

\[ P(S \mid R) = P(S) \]

**Bayes’ Rule:**

\[ P(S \mid R) = \frac{P(R \mid S) P(S)}{P(R)} \]
Bayesian Optimal Decision and the Likelihood Ratio

The Problem:
- Choose between two alternatives (stimuli) S0 and S1 with prior probabilities \( P(S0) \) and \( P(S1) \) given the observation (neural response) \( R \) so as to minimize the probability of error.
- Conditional probabilities (stimulus-response relationships) \( P(R|S0) \) and \( P(R|S1) \) are known.

Bayes’ Optimal Decision Rule:
- Given a specific value of \( R \), choose the alternative which maximizes the posterior probability:

\[
P(S1 | R) > P(S0 | R)
\]

Equivalent Likelihood Ratio Test (LRT):

\[
LR = \frac{P(R | S1)}{P(R | S0)} \quad \text{when} \quad S1 \quad \text{is preferred}
\]

LRT separates priors (under experimenter control) from neuron’s stimulus-response characteristic.
Properties of the Likelihood Ratio

\[ LR = \frac{P(R \mid S1)}{P(S0)} > \frac{P(S0)}{P(S1)} \]

- Separates computation (left side) from biases and costs (prior probabilities, right)
- If the conditional probabilities \( P(R \mid Si) \) are either Poisson or Gaussian with equal variances, LRT reduces to comparing the response \( R \) with a threshold \( \gamma \):

\[
\begin{align*}
S1 \\
R > \gamma \\
S0
\end{align*}
\]

- LRT works not only for scalar observations (e.g. a spike count from a single neuron), but also for multidimensional observations (e.g. temporal discharge patterns and ensembles of neurons).
- Decision rule invariant to monotonic transformation (e.g. logarithm)
- For independent observations (e.g. simultaneous observations from multiple neurons or multiple observation intervals from same neuron), log\( LR \) is additive:

\[
\log LR(R1, R2, \ldots, RN) = \sum_{i=1}^{N} \log LR(Ri)
\]
Detection and False Alarm Probabilities

- The neural response (in general LR) is compared to a criterion to make a decision about which stimulus was presented.
- Two types of errors: “misses” and “false alarms”.
- $PF$ is the probability of false alarm.
- $PD$ is the probability of detection; $PM = 1-PD$ is the probability of a miss.
Receiver Operating Characteristic (ROC)

Conditional Probabilities (Poisson)

ROC Curves

- The area under the ROC curve gives a distribution-free measure of performance.
\( d' \) as a measure of performance

- If the conditional probabilities \( P(R|S0) \) and \( P(R|S1) \) are Gaussian with equal variances, then \( d' = \Delta r/\sigma \) completely determines the performance (the ROC curve):
  \[
  PC_{21} = \frac{1}{\sqrt{2\pi}} \int_{-d'/\sqrt{2}}^{\infty} e^{-z^2/2} dz
  \]

- Many probability distributions (including Poisson) approach a Gaussian when the mean response becomes moderately large.

- If so, performance in non-Gaussian cases can be approximated by \( d' \). If, as in the Poisson case, the variances are unequal under the two alternatives, they can be averaged.

- \( d' \) is a poor measure of performance when the number of spikes is very small or when the two stimuli are widely separated.
Visual discrimination of random-dot patterns by monkeys and cortical neurons

- Single-unit recordings from Area MT (middle temporal) of awake macaques
- Single-unit responses and behavioral performance were recorded simultaneously
- Performance of typical neuron matches behavioral performance

Figures removed due to copyright reasons.
Please see:
Forward masking in the auditory nerve?

Figures removed due to copyright reasons.

Please see:

Psychophysical masked thresholds grow much faster with masker level than neural thresholds. Maximum masking can reach 30-50 dB.
Performance based on spike count from single AN fiber severely deviates from Weber’s law

Discharge rate (sp/sec)

Intensity DL (dB)

Intensity (dB SPL)

Auditory-nerve fiber

Hypothetical neuron verifying Weber’s law

Poisson statistics assumed
The "lower envelope principle"

Somatosensory: Detection of sinusoidal vibration

Auditory: Pure tone detection

Mountcastle (1972)

Psychophysical performance is determined by the most sensitive neuron in the population.

Please see:
Optimal pooling of information across neurons

• Performance in discriminating two stimuli can be improved by combining information across neurons. Specifically, if the spike counts from $N$ neurons are either Poisson or Gaussian and statistically-independent, the optimum combination rule is to form a weighted sum of the spike counts. The discriminability index for this optimum combination is given by $d'^2 = \sum_i d'_i^2$.

• The structure of the optimum detector is identical to that of a single-layer perceptron in artificial neural networks. The weights can be interpreted as the strengths of synaptic connections and the threshold device as the threshold for all-or-none spike discharges in a postsynaptic neuron.

• When responses of the different neurons are not statistically independent, benefits of combining information may be better or worse than in independent case depending on the nature of the correlations.

Delgutte (unpublished)
Siebert (1968) developed the first model for predicting the performance of a sensory system based on the activity in primary neurons. His model incorporated all the key features of AN activity known at the time, including logarithmic cochlear frequency map, cochlear tuning, saturating rate-level functions and Poisson discharge statistics.

The model predicts constant performance in pure-tone intensity discrimination over a wide range of intensities (Weber’s law) by relying on the unsaturated fibers on the skirts of the cochlear excitation pattern.

Psychophysical experiments (Viemeister, 1983) have since ruled out this model because listeners show good performance even when masking noise restricts information to a narrow band around the tone frequency.
Spread of excitation to remote cochlear places makes model predict Weber’s Law…

...but performance severely degrades in band-reject noise.
Rate vs. Temporal Codes

- Temporal code can convey more information than rate code
- Distinction between rate and temporal codes not that easy:
  - Rate has to change when new stimulus occurs. How fast can it change and still be called "rate code"?
  - Poisson process completely specified by instantaneous firing rate. If instantaneous rate tracks stimulus waveform, is it a rate or a temporal code?

From:
Used with permission.
Rate-Place Code

- Narrow tuning functions
- Neurons are topographically organized to form map of stimulus parameter
- Each stimulus activates a restricted population of neurons
- Stimulus estimated from location of maximum activity along map
- Examples:
  - Retinal map of space
  - Tonotopic map
  - ITD map in barn owl (Jeffress model)
A distributed code: The population vector (Georgopoulos, 1986)

- Broad tuning functions around a preferred direction
- Topographic map not necessary
- Each stimulus activates entire neural population to a degree
- Each neuron’s response can be seen as a vector, with angle equal to the neuron’s best direction and magnitude equal to firing rate
- Vector sum across neural population accurately points to the stimulus direction
- Example: Direction of arm movement in motor cortex.
- Could apply to any angular dimension, e.g. sound azimuth.
Entropy of a Random Variable

Definition: \[ H(X) = -\sum_{x \notin X} p(x) \log_2 p(x) \]

- Entropy is expressed in *bits* (binary choices)
- For a uniform probability distribution \((p(x) = 1/N\) for all \(x)\), the entropy is the logarithm (base 2) of the number \(N\) of possible values of \(X\). All other distributions have a lower entropy.
- The entropy represents the number of independent binary choices required to uniquely specify the value of a random variable, i.e. it quantifies the “uncertainty” about that variable. It is also the average length of a binary string of 1’s and 0’s required to encode the variable.
Entropy rates in sounds

For a continuous signal source, the entropy rate (in bits/s) is

\[ R = \lim_{T \to \infty} \frac{1}{T} H(T) \]

- CD-quality audio: 44,100 Hz X 16 bits X 2 channels = 1.4 Mb/s
- MP3: 32-128 kb/s
- Telephone: 30 kb/s
- Vocoder: 2.4-9.6 kb/s
- Nonsense speech: 8 phonemes/sec X \(\log_2(32)\) phonemes = 40 b/s (overestimate because not equiprobable)
- This lecture: 2 bits/3600s = 0.0006 b/s
- What about neuron firings?
Entropy rate of spike trains

**Spike Times (binomial distribution, McKay & McCulloch 1952)**

\[ R_T \approx r \log_2 \left( \frac{e}{r \Delta t} \right) \]

\( r \) is the average firing rate, \( \Delta t \) the temporal resolution

*Example:* \( H_T = 950 \) bits/s (3.2 bits/spike)

for \( r = 300 \) spikes/s, \( \Delta t = 1 \) ms

**Spike Counts (exponential distribution - optimal)**

\[ R_R = \left[ \log_2 (1 + \bar{n}) + \bar{n} \log_2 (1 + 1/\bar{n}) \right] / T \]

\( n = rT \) is the mean spike count, \( T \) is the recording time

*Example:* \( H_R = 26 \) bits/s (0.1 bits/spike) for \( r = 300 \) spikes/s, \( T = 300 \) ms

The entropy rate of a spike train gives an upper bound on the information about the stimulus that can be transmitted by the spike train.
Conditional Entropy and Mutual Information

- The entropy \( H(S) \) represents the uncertainty about the stimulus in the absence of any other information.
- The conditional entropy \( H(S|R) \) represents the remaining stimulus uncertainty after the neural response has been measured.
- \( I(S,R) = H(S)−H(S|R) \) is the mutual information between \( S \) and \( R \); it represents the reduction in uncertainty achieved by measuring \( R \).
- If \( S \) and \( R \) are statistically independent, then \( I(S,R) = 0 \) and \( H(S,R) = H(S) + H(R) \).
- By symmetry, \( I(S,R) = I(R,S) = H(R)−H(R|S) \).
Direct method for computing mutual information between stimulus and neural response

- The most obvious method for measuring $I(S,R)$ would be to subtract $H(S|R)$ from $H(S)$. But $H(S|R)$ is hard to measure because it requires estimating the stimulus from neural response.

- **Trick:** By symmetry, $I(S,R)$ is also $H(R) - H(R|S)$. $H(R|S)$ is the entropy of the part of the neural response that is NOT predictable from the stimulus, i.e. the noise in the response.

**Method:**

- Present stimulus set $S$ with probability $P(S)$. Measure neural response to many presentations of each stimulus.
- Estimate $P(R)$ for entire stimulus set and compute $H(R)$ from $P(R)$.
- To estimate the noise in the response, average response to all presentations of same stimulus, and subtract average from response to each trial. Compute $H(R|S)$ from the estimated noise distribution.
- Subtract $H(R|S)$ from $H(R)$.
- Assumptions about noise distribution (e.g. Gaussian) can be made to simplify estimation of probabilities.
- In practice, this method becomes prohibitive in its data requirements for large dimensionality, e.g. when computing information for population of neurons.
Both the stimulus $S(t)$ and the additive noise $N(t)$ have Gaussian probability distributions which are fully characterized by the power spectra $S(f)$ and $N(f)$.

$$R_{INFO} = \int_{0}^{W} \log_2 \left( 1 + \frac{S(f)}{N(f)} \right) df$$

Information rate is entirely determined by available bandwidth $W$ and signal-to-noise ratio $S/N$.

Telephone channel: $S/N=30$ dB, $W=3000$ Hz => $I = 30$ kb/s
Natural and synthetic songs are coded more efficiently than noise in zebra finch auditory neurons.

3 types of stimuli:
- Song: 20 natural songs from male zebra finch
- Syn-song: Mimics spectral and temporal modulations in song
- ML-noise: Noise with limited range of modulations

Neural recording from 3 locations:
- MLd: Auditory midbrain
- Field L: Primary cortex
- CM: Secondary cortex involved in song processing

Using the stimulus reconstruction method to estimate stimulus information in spike train

- Computing $I(S,R)$ directly from the definition often requires too much data.

- **Data processing theorem:** If $Z=f(R)$, then $I(S,Z) \leq I(S,R)$. “Data processing cannot increase information”.
  - Special case: If $\hat{S}$ is an estimate of $S$ based on neural response $R$, then $I(S,\hat{S}) \leq I(S,R)$

- **Method:**
  - Use Gaussian stimulus $S(t)$ and compute linear estimate $\hat{S}(t)$ from neural response.
  - Define noise as $N(t) = S(t) - \hat{S}(t)$, assumed to be Gaussian.
  - Compute power spectra of $N(t)$ and $S(t)$.
  - Use formula for information rate of Gaussian channel to estimate $R_{INFO}$.

- The method is meant to give a lower bound on the information rate. The better the stimulus estimate and the Gaussian assumption, the more accurate the bound.

From:
Linear reconstruction of a stimulus from the spike train (Bialek et al., 1991)

- Given a stimulus waveform $s(t)$ and a spike train $r(t)$, what is the linear filter $h(t)$ which gives the least-squares estimate of $s(t)$ from $r(t)$?

$$\hat{s}(t) = h(t) * r(t) = \sum_i h(t - t_i)$$

Find $h(t)$ to minimize $E = \int_0^T (s(t) - \hat{s}(t))^2 \, dt$

- The solution is given (in the frequency domain) by the Wiener filter:

$$H(f) = \frac{S_{rs}(f)}{S_r(f)}$$

- $H(f)$ is the Fourier transform of $h(t)$, i.e. the filter frequency response, $S_r(f)$ is the power spectrum of the spike train, and $S_{rs}(f)$ is the cross-spectrum between $r(t)$ and $s(t)$.
Example of linear stimulus reconstruction from a spike train

Figures removed due to copyright reasons.

Please see:

- Fly visual system (“H1 neuron”), moving grating stimulus
- Reconstruction resembles lowpass filtered stimulus
Male bullfrogs produce an advertisement call to convey information about location and breeding readiness to both sexes.

From:
Used with permission.
Information rates in bullfrog primary auditory fibers are higher for call than for noise

- Linear stimulus reconstruction method was used to estimate (a lower bound on) information rate for both broadband Gaussian noise and call-shaped noise.
- Information rates are higher for call-shaped noise (40-250 bits/s) than for broadband noise (10-50 bits/s).

Figures removed due to copyright reasons.
Please see:
Coding efficiency in bullfrog primary auditory fibers

Coding Efficiency = \frac{\text{Information Rate}}{\text{Entropy Rate in Spike Train}}

• The coding efficiency is a measure of the fraction of the information contained in the spike train that is used to code the stimulus.

• Coding efficiency ranged from 0.05 to 0.2 for the broadband stimulus, 0.35 to 0.9 for the call-shaped noise.

• These high coding efficiencies were achieved despite the poor quality of the stimulus reconstructions.

Figures removed due to copyright reasons.
Please see:
**Information rates and coding efficiencies in sensory neurons**

<table>
<thead>
<tr>
<th>Neural System and Species</th>
<th>Information Rate (bits/spike)</th>
<th>Information Rate (bits/s)</th>
<th>Coding Efficiency</th>
<th>Best Timing Precision and Variance : Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant stimulus</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cat retinal ganglion cells</td>
<td>-0.04-0.10</td>
<td>0.4-0.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Primary visual cortex (VI) of rhesus monkeys</td>
<td>-</td>
<td>0.62</td>
<td>-</td>
<td>~1.3</td>
</tr>
<tr>
<td>Middle temporal area (MT) of rhesus monkeys</td>
<td>0.025</td>
<td>0.89 ± 0.29</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Inferior temporal area (IT) of rhesus monkeys</td>
<td>-0.13</td>
<td>0.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hippocampus of rhesus monkeys</td>
<td>-0.18</td>
<td>0.32 (max = 1.2)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Variable stimulus: reconstruction method</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HI motion-sensitive neuron of a fly</td>
<td>0.75</td>
<td>64</td>
<td>30%</td>
<td>~2 ms</td>
</tr>
<tr>
<td>Frog auditory afferents</td>
<td>0.66</td>
<td>23</td>
<td>11%</td>
<td>-</td>
</tr>
<tr>
<td>Vibratory receptors of the bullfrog sacculus</td>
<td>2.6</td>
<td>155</td>
<td>50-60%</td>
<td>~0.4 ms</td>
</tr>
<tr>
<td>Cricket mechanoreceptors</td>
<td>0.6-3.2</td>
<td>75-294</td>
<td>50-60%</td>
<td>-</td>
</tr>
<tr>
<td>Salamander retinal ganglion cells</td>
<td>1.9</td>
<td>3.7 (up to 10 for a population of &gt;10 cells)</td>
<td>26% (&gt;79% for &gt;10 cells)</td>
<td>-</td>
</tr>
<tr>
<td>The MT of anesthetized rhesus monkeys</td>
<td>-0.65</td>
<td>9.7 (max = 12.3)</td>
<td>-</td>
<td>2-4 ms</td>
</tr>
<tr>
<td>The MT of alert rhesus monkeys</td>
<td>0.56</td>
<td>5.5</td>
<td>&lt;30%</td>
<td>-</td>
</tr>
<tr>
<td><strong>Variable stimulus: direct method</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HI motion-sensitive neuron of a fly</td>
<td>2.43</td>
<td>80</td>
<td>50%</td>
<td>1.5-3 ms; ≤0.1</td>
</tr>
<tr>
<td>Salamander and rabbit retinal ganglion cells</td>
<td>3.7</td>
<td>16.3</td>
<td>59%</td>
<td>≥ 0.70 ms; ≥ 0.05</td>
</tr>
<tr>
<td>The MT of alert rhesus monkeys</td>
<td>1.5</td>
<td>12 (max = 29)</td>
<td>Up to 45%</td>
<td>&lt;2 ms; ~1.4</td>
</tr>
</tbody>
</table>

Table by MIT OCW.
Information Theory Pro’s and Con’s

Pros
• Does not assume any particular neural code
• Can be used to identify the stimulus features best encoded by neurons or to compare effectiveness of different putative neural codes
• One number summarizes how well stimulus set is coded in neural response

Cons
• Information estimate depends on stimulus set. Stimulus probabilities in environment hard to specify.
• Does not specify how to read out the code: the code might be unreadable by the rest of the nervous system.
• For all but simplest examples, estimation of mutual information requires huge amount of data. Methods that try to circumvent data limitations (e.g. stimulus reconstruction) make additional assumptions (e.g. linearity or Gaussian distribution) that are not always valid.
Slide 25:

Slide 35:
Motor Control of Speech:
Control Variables and Mechanisms

HST 722, Brain Mechanisms for Hearing and Speech

Joseph S. Perkell
MIT
Outline

• Introduction
• Measuring speech production
• What are the “controlled variables” for segmental (phonemic) speech movements?
• Segmental motor programming goals
• Producing speech sounds in sequences
• Experiments on feedback control
• Summary
Outline

• Introduction
  – Utterance planning
  – General physiological/neurophysiological features
  – The controlled systems
  – Example of movements of vocal-tract articulators

• Measuring speech production
• What are the “controlled variables” for segmental speech movements?
• Segmental motor programming goals
• Producing speech sounds in sequences
• Experiments on feedback control
• Summary
Utterance Planning

- Objective: generate an intelligible message while providing for “economy of effort” – stages:
  - Form the message (e.g. Feel hungry; smell pizza; together with a friend).
  - Select and sequence lexical items (words). “Do you want a pizza?”
  - Assign a syntactically-governed prosodic structure.
  - Determine “postural” parameters of overall rate, loudness and degree of **reduction** (and settings that convey emotional state, etc.)
    - Extreme reduction: “Dja wanna pizza?”
  - Determine temporal patterns: Sound segment durations depend on:
    - Phoneme length
    - Overall rate
    - Intrinsic characteristics of sounds
    - Position and number of syllables in word

- Result: an ordered sequence of goals for the production mechanism
Serial Ordering

• Evidence reflecting serial ordering in utterance planning: speech errors
  – Examples from Shattuck-Hufnagel (1979)
    • Substitution: *Any*may → *Any*way
    • Exchange: *em*eny → *en*emy
    • Shift: *bad high*way dri_ing → *high*way dri_ving
    • Addition: *the p/ublic*city would be → *p/ublic*ity
    • Omission: *sona*_umber ten → *sona*_umber ten
    • ?: *dig*ital processing → *dig*ital processing
  – See Averbeck et al. on neurophysiological evidence concerning serial ordering
General Physiological/Neurophysiological Features

- Muscles are under voluntary control
- Structures contain feedback receptors that supply sensory information to the CNS:
  - Surfaces: touch/pressure
  - Muscles:
    - length and length changes: spindles
    - Tension: tendon organs
  - Joints (TMJ): joint angle
- Reflex mechanisms:
  - Stretch
  - Laryngeal (coughing)
  - Startle
- Motor programs (low-level, “hard wired” neural pattern generators)
  - Breathing
  - Swallowing
  - Chewing
  - Sucking
- Low-level circuitry could be employed in speech motor control. The picture is complex, and a comprehensive account hasn’t emerged.

Figure by MIT OCW.
The controlled systems

• The respiratory system
  – most massive (slowly-moving structures)
  – Provides energy for sound production
    • Fluctuations to help signal emphasis
    • Relatively constant level of subglottal pressure
  – Different patterns of respiration: breathing, reading aloud, spontaneous, counting
  – Different muscles are active at different phases of the respiratory cycle – a complex, low-level motor program

• Larynx
  – Smallest structures, most rapidly contracting muscles
  – Voicing, turned on and off segment by segment
  – F0, breathiness – suprasegmental regulation

• Vocal tract
  – Intermediate-sized, slowly moving structures: tongue, lips, velum, mandible
  – Many muscles do not insert on hard structures
  – Can produce sounds at rates up to 15/sec
  – To do so, the movements are coarticulated

Please see:
Focus of lecture is on movements of vocal-tract articulators

• Consider the movements of each of these structures
• Approximate number of muscle pairs that move the
  – Tongue: 9
  – Velum: 3
  – Lips: 12
  – Mandible: 7
  – Hyoid bone: 10
  – Larynx: 8
  – Pharynx: 4
• Not including the respiratory system

• Observations:
  • A large number of degrees of freedom
  • A very complicated control problem
Outline

• Introduction

• Measuring speech production
  – Acoustics
  – Articulatory movement
  – Area functions

• What are the “controlled variables” for segmental speech movements?

• Segmental motor programming goals

• Producing speech sounds in sequences

• Experiments on feedback control

• Summary
• Acoustics – important for perception
  – Spectral, temporal and amplitude measures
• Vowels, liquids and glides:
  – Time varying patterns of formant frequencies
• Consonants:
  – Noise bursts
  – Silent intervals
  – Aspiration and frication noises
  – Rapid formant transitions

Measuring Speech Production

Figure removed due to copyright reasons.
Please see:

“The yacht was a heavy one”

• Movements
  – From x-ray tracings
  – With an Electro-Magnetic Midsagittal Articulometer (EMMA) System
    • Points on the tongue, lips, jaw, (velum)
  • Other parameters: air pressures and flows, muscle activity …
EMMA Data Collection

- Transducer coils are placed on subject’s articulators
- Subject reads text from an LCD screen
- Movement and audio signals are digitized and displayed in real time
- Signals are processed and data are extracted and analyzed
Analysis of EMMA data

- Algorithmic data extraction at time of minimum in absolute velocity during the vowel:
  - Vowel formants
  - Articulatory positions (x, y)
3-Dimensional Area Function Data

/i/ - 2 speakers

Transverse (horizontal)

/u/ - 2 speakers

Coronal (vertical)

- MR images of sustained vowels (Baer et al., JASA 90: 799-828)
  - Area functions are more complicated than they look in 2 dimensions
  - There are lateral asymmetries, but 2-D midsagittal (midline) movement data provide useful information

Reprinted with permission from
Outline

• Introduction
• Measuring speech production
• What are the “controlled variables” for segmental (phonemic) speech movements?
  – Possible controlled variables
  – Modeling to make the problem approachable: DIVA
  – A schematic view of speech movements
• Segmental motor programming goals
• Producing speech sounds in sequences
• Experiments on feedback control
• Summary
“What are the controlled variables?”

- The question has theoretical and practical implications
  - What are the fundamental motor programming units and most appropriate elements for phonological/phonetic theory?
  - What domains should be the main focus of research for diagnosis and treatment of speech disorders?

Objective of Speaker:
- To produce sounds strings with acoustic patterns that result in intelligible patterns of auditory sensations in the listener

Acoustic/auditory cues depend on type of sound segment:
- Vowels and glides: Time varying patterns of formant frequencies
- Consonants: Noise bursts, Silent intervals, Aspiration and frication noises, Rapid formant transitions

Please see:
Possible Motor Control Variables

Auditory characteristics of speech sounds are determined by:
1. Levels of muscle tension
2. Changing muscle lengths and movements of structures
3. The vocal-tract shape (area function)
4. Aerodynamic events and aeromechanical interactions
5. The acoustic properties of the radiated sound

Hypothetically, motor control variables could consist of feedback about any combination of the above parameters.
Modeling to make the problem approachable: DIVA

“Directions Into Velocities of Articulators” (Guenther and Colleagues – Next lecture)

- A neuro-computational model of relations among cortical activity, motor output, sensory consequences
- **Phonemic Goals**: Projections (mappings) from premotor to sensory cortex that encode *expected sensory consequences* of produced speech sounds
  - Correspond to *regions* in multidimensional auditory-temporal and somatosensory-temporal spaces
- Roles of feedforward and feedback subsystems will be discussed later.
A Schematic View of Speech Movements

• Planned and actual acoustic trajectories illustrate:
  – Auditory/acoustic goal *regions*  
  – Economy of effort (Lindblom)  
  – Coarticulation  
  – Motor equivalence  
  – Biomechanical saturation (quantal) effects

• When controlling an articulatory speech synthesizer, DIVA, accounts for the first four and
  – Aspects of acquisition  
  – Responses to perturbations

![Acoustic Space](image)

![Articulator Space](image)

Figure by MIT OCW.
Outline

• Introduction
• Measuring speech production
• What are the “controlled variables” for segmental speech movements?
• Segmental motor programming goals
  – Anatomical and acoustic constraints: Quantal effects
  – Individual differences – anatomy
  – Motor equivalence: A strategy to stabilize acoustic goals
  – Clarity vs. economy of effort
  – Relations between production and perception
    • Vowels
    • Sibilants
• Producing speech sounds in sequences
• Experiments on feedback control
• Summary
Anatomical and Acoustic Constraints on Articulatory Goals

• Properties of speakers’ production and perception mechanisms help to define goals for speech sounds that are used in speech motor planning.

• Some of these properties are characterized by quantal effects (Stevens), which can also be called “saturation effects.”

• Schematic example: A continuous change in an articulatory parameter produces two regions of acoustic stability, separated by a rapid transition.

• Hypothesis: some goals are auditory and can be characterized in terms of acoustic parameters: formant frequencies, relative sound level, etc.

• Languages “prefer” such stable regions
• The use of those regions by individual speakers helps to produce relatively robust acoustic cues with imprecise motor commands.
Goals for the vowel /i/ - A. An acoustic saturation (quantal) effect for constriction location (Stevens, 1989)

There is a range (in green) of back cavity lengths over which F1-F3 are relatively stable.

Many repetitions of /i/ in two subjects show a corresponding variation of constriction location.

However, as reflected in the articulatory data, the formants of /i/ are sensitive to variation in constriction degree.

Quantal and non-quantal articulatory-to-acoustic relations for /i/ and /a/

Figure 6. Formant values that result from manipulating the location and degree of constriction of an articulatory synthesizer for the vowel /i/ on the left, and /a/ on the right. The filled circles represent formant values at 1 mm increments of front-to-back tongue-body movement for /i/ and low-to-high tongue-body movement for /a/; and the tick marks show formant values that represent formant values at 1 mm increments of tongue-body movement from the least constricted to the most constricted configuration for each vowel.

Reprinted with permission from:
A biomechanical saturation effect for constriction degree for /i/

- Constriction degree and resulting formants can be stabilized
  - Stiffening the tongue blade (with intrinsic muscles)
  - Pressing the stiffened tongue blade against the sides of the hard palate through contraction of the posterior genioglossus (GGp) muscles

Reprinted with permission from:

- Constriction area (shaded) varies little, even with variation in GGp contraction (from a 3D tongue model by Fujimura & Kakita)
Tongue Contour Differences Among Four Speakers

• Note the tongue contour differences among the four speakers.

• An “auditory-motor theory of speech production” (Ladefoged, et al., 1972)
Effects of different palate shapes on vowel articulations

- Palatal shapes differ among individuals
- Palatal depth can influence
  - Spatial differences in vowel targets

Figures removed due to copyright reasons.
Please see:
Perkell, J. S. "On the nature of distinctive features: Implications of a preliminary vowel production study."
*Frontiers of Speech Communication Research*. Edited by B. Lindblom and S. Öhman.

Figures by MIT OCW.
Production of /u/

- Contractions of the styloglossus and posterior genioglossus
- Note: place of constriction & variation in constriction location

A schematic illustration of articulatory data for multiple repetitions of the vowels /a/ (tongue surface illustrated by ⋅⋅⋅⋅⋅⋅), /i/ (−−−−−) and /u/ (−−−−−−−), showing elliptical distribution of positioning of points on the tongue surface, with the long axes of the ellipses oriented parallel to the vocal-tract midline.

Figure by MIT OCW.

Figure by MIT OCW.
Stabilizing the sound output for the vowel /u/: Motor Equivalence

- **Hypothesis**: negative correlation between tongue-body raising and lip protrusion in multiple repetitions of the vowel

- **Hypothesis is supported in a number of subjects**
  - The goal for the articulatory movements for /u/ is in an acoustic/auditory frame of reference, not a spatial one
  - **Strategy**: Stay just within the acoustic goal region

Figures by MIT OCW.
Palatal Depth and Motor Equivalence

- Palatal depth can also influence variability in movement toward vowel targets
Motor Equivalence for /r/

- Speakers use similar articulatory trading relations when producing /r/ in different phonetic contexts (Guenther, Espy-Wilson, Boyce, Matthies, Perkell, and Zandipour, 1999, JASA).

- Acoustic effect of the longer front cavity of the blue outlines is compensated by the effect of the longer and narrower constriction of the red outlines (e.g., Stevens, 1998).

- F3 variability is greatly decreased by these articulatory trading relations.

- Conclusion: The movement goal for /r/ is a low value of F3 – an auditory/acoustic goal.
Clarity vs. Economy of Effort: Another principle (continuous, as opposed to quantal) that influences vowel categories (Lindblom, 1971)

- Used an articulatory synthesizer and heuristics to estimate the location of vowels in F1, F2 space, based on
  - A compromise between “perceptual differentiation” and “articulatory ease” and
  - The number of vowels in the language
- Approximated vowel distributions for languages containing up to about 7 vowels
- Later discussed in terms of a tradeoff between clarity and economy of effort, i.e., a relation between production and perception
Relations between Production and Perception

• Close linkage between production and perception:
  – Speech acquisition, with and without hearing
  – Speech of Cochlear Implant users
  – Second-language learning (e.g., Bradlow et al.)
  – Focused studies of production & perception (e.g., Newman)
  – Mirror neurons – a more general action-perception link (e.g., Fadiga et al.)

• Hypothesis:
  – Speakers who discriminate well between vowel sounds with subtle acoustic differences will produce more clear-cut sound contrasts
  – Speakers who are less able to discriminate the same sound stimuli will produce less clear-cut contrasts
Production Experiment

Data Collection
- Subjects: 19 young-adult speakers of American English
- For each subject:
  - Recorded articulatory movements and acoustic signal
  - Subject pronounced “Say___ hid it.”; ____ = cod, cud, who’d or hood
  - Clear, Normal and Fast conditions

Analysis
- Calculated contrast distance for each vowel pair:
  - Articulatory (TB) contrast distance: distance in mm between the centroids of the cod and cud TB distributions.
  - Acoustic contrast distance: distance in Hz between centroids of F1, F2 distributions for cod, cud

Reprinted with permission from
Perception Experiment

• Methods
  – Synthesized natural-sounding stimuli in 7-step continua – for cod-cud, who’d-hood
  – Each subject: Labeling and discrimination (ABX) tasks

• Results: ABX scores (2-step)
  – Ceiling effects: some 100% subjects probably had better discrimination than measured
  – For further analysis divide subjects into two groups:
    • HI discriminators - at 100% (above the median)
    • LO discriminators - (at median and below)

Reprinted with permission from:
Results & Conclusions

• HI discrimination subjects produced greater contrast distance than LO discrimination subjects (measured in articulation or acoustics)

• The more accurately a speaker discriminates a vowel contrast, the more distinctly the speaker produces the contrast

* Difference between HI and LO groups is significant at p < .001

Reprinted with permission from:
Copyright 2004, Acoustical Society America.
A Possible Explanation

• It is advantageous to be as intelligible as possible
• Children will acquire goal regions that are as distinct as possible
  – Speakers who can perceive fine acoustic details learn *auditory goal regions* that are *smaller* and *spaced further apart* than speakers with less acute perception, because
  – The speakers with more acute perception are more likely to reject poorly produced tokens when learning the goal regions
Consonants: A saturation effect for /s/ may help define the /s-j/ contrast

• Production of /j/ (as in “shed”)
  – Relatively long, narrow groove between tongue blade and palate
  – Sublingual space

• Production of /s/ (as in “said”)
  – Short narrow groove
  – No sublingual space

• Saturation effect for /s/:
  – As tongue moves forward from /j/, sublingual cavity volume decreases
  – When tongue contacts lower alveolar ridge, sublingual cavity is eliminated, resonant frequency of anterior cavity increases abruptly
  – After contact, muscle activity can increase further; output is unchanged

Figures removed due to copyright reasons.
Please see:
Relations Between Production and Perception of Sibilants

• Hypothesis: The sibilants, /s/ and /ʃ/, have two kinds of sensory goals:
  – Auditory: particular distribution of energy in the noise spectrum
  – Somatosensory: e.g., patterns of contact of the tongue blade with the palate and teeth

...Figures removed due to copyright reasons.
Please see:

• Speakers will vary in their ability to discriminate /s/ from /ʃ/
• Speakers use contact of the tongue tip with the lower alveolar ridge for /s/ to help differentiate /s/ from /ʃ/
  – This will also vary across speakers
• Across speakers, both factors, *ability to discriminate auditorily* between the two sounds and *use of contact* (a possible *somatosensory goal*), will predict the strength of the produced contrast

...
Methods (with the same 19 subjects as the vowel study)

• Production experiment – each subject:
  – Recorded: acoustic signal, and contact of the under side of the tongue tip with the lower alveolar ridge - with a custom-made sensor
  – Subject pronounced, “Say____ hid it.”; ____ = sod, shod, said or shed”
  – Clear, Normal and Fast conditions

• Analysis – calculated:
  – Proportion of time contact was made during the sibilant interval
  – Spectral median for /s/ and /ʃ/
  – Acoustic contrast distance:
    • Difference in spectral median between /s/ and /ʃ/  

• Perception experiment - each subject:
  – Labeled and discriminated (ABX) between synthesized stimuli from a seven-step said to shed continuum

Figures removed due to copyright reasons. Please see:
Results

Use of tongue-to-lower-ridge contact

Discrimination

Figures removed due to copyright reasons.
Please see:
"The Distinctness of Speakers' /s/-/ʃ/ Contrast is related to their auditory discrimination and use of an articulatory saturation effect."

- 12 subjects (left of vertical line) are classified as Strong (S) for use of contact difference (c) between /s/ and /ʃ/.
- The remaining subjects are classified Weak (W) for use of contact difference.
- Nine subjects had percent correct = 100; categorized as HI discriminators (right of line).
- 10 subjects had percent correct < 100; categorized as LO discriminators.
Produced contrast distance is related to

– Ability to discriminate the contrast

* Difference is significant, p < .01

– Use of contact difference

• Interactions
  – Speakers with good discrimination and use of contact difference: best contrasts
  – Speakers with one or the other factor: intermediate contrasts
  – Speakers with neither factor: poorest contrasts

Figures removed due to copyright reasons.
Please see:
"The Distinctness of Speakers' /s/-/ʃ/ Contrast is related to their auditory discrimination and use of an articulatory saturation effect.”
Outline (break time?)

- Introduction
- Measuring speech production
- What are the “controlled variables” for segmental speech movements?
- Segmental motor programming goals
- Producing speech sounds in sequences
  - An example utterance
  - Movements show context dependence
    - Velar movements
    - Lip rounding for /u/
  - Effects of speaking rate
  - Persistence of inaudible gestures at word boundaries
- Experiments on feedback control
- Summary
• An example utterance
• * indicates articulations that aren’t strongly constrained by communicative needs
• Articulations anticipate upcoming requirements: anticipatory coarticulation
• Coarticulation:
  – asynchronous movements of structures of differingsizes and movement time constants
  – a complicated motor coordination task

### Producing Sounds in Sequences

<table>
<thead>
<tr>
<th></th>
<th>k</th>
<th>æ</th>
<th>m</th>
<th>b</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tongue body</strong></td>
<td>Rise to contact roof of mouth to achieve closure &amp; silence</td>
<td>Release contact to generate a noise burst; move down to vowel position</td>
<td>Begin movement toward /r/ position</td>
<td>*</td>
<td>Maintain /r/ position</td>
</tr>
<tr>
<td><strong>Tongue blade</strong></td>
<td>*</td>
<td>Maintain contact with floor of mouth to stay out of the way</td>
<td>Begin retroflexion or bunching in anticipation of /r/</td>
<td>*</td>
<td>Maintain retroflexed or bunched configuration</td>
</tr>
<tr>
<td><strong>Lips</strong></td>
<td>Begin spreading for the vowel /æ/</td>
<td>Maintain position for vowel, then begin toward closure</td>
<td>Achieve &amp; maintain closure</td>
<td>Maintain closure</td>
<td>Release rapidly &amp; round somewhat</td>
</tr>
<tr>
<td><strong>Mandible</strong></td>
<td>Move upward to support tongue movement</td>
<td>Move downward to support tongue movement</td>
<td>Move upward to support lower lip movement</td>
<td>*</td>
<td>Move downward slightly to aid lip release</td>
</tr>
<tr>
<td><strong>Soft palate</strong></td>
<td>Maintain closure to contain pressure buildup</td>
<td>Begin downward movement to open velopharyngeal port for /m/</td>
<td>Begin closing movement toward onset of /b/</td>
<td>Reach closure at right instant to begin /b/; (move upward during /b/ to help expand v.t. walls - voicing)</td>
<td>*</td>
</tr>
<tr>
<td><strong>Vocal-tract walls</strong></td>
<td>Stiffen to contain air pressure buildup</td>
<td>*</td>
<td>*</td>
<td>Relax, perhaps expand actively to allow continuation of voicing for /b/</td>
<td>*</td>
</tr>
<tr>
<td><strong>Vocal-fold position</strong></td>
<td>Abduct maximally, with peak occurring at /k/ release</td>
<td>Adduct to position for voicing</td>
<td>Maintain position</td>
<td>Maintain position</td>
<td>Maintain position</td>
</tr>
<tr>
<td><strong>Tension on vocal folds</strong></td>
<td>Begin to raise tension to signal stress on following vowel</td>
<td>Achieve maximum tension for the F0 peak that signals stress</td>
<td>Lower tension to lower F0</td>
<td>Maintain tension</td>
<td>Maintain tension</td>
</tr>
<tr>
<td><strong>Respiratory system</strong></td>
<td>Increase subglottal air pressure to obtain a burst release for the /k/</td>
<td>Maintain subglottal air pressure for increased sound level to signal stress</td>
<td>Return to the previous value of subglottal pressure</td>
<td>Maintain pressure</td>
<td>Maintain pressure</td>
</tr>
</tbody>
</table>

Table by MIT OCW.
What happens when sounds are produced in sequences?

• When individuals speak to one another, additional forces are at play
  – Articulatory movements from one sound to another are influenced by dynamical factors: canonical targets are very rarely reached.
  – The speaker knows that the listener can fill in a great deal of missing information, so “reduction” takes place (see Introduction)
  – Speaking style (casual, clear, rapid, etc.) can vary
    • Amount of variation can depend on the situation and the interlocutor (a familiar speaker of the same language?)
Movements Show Context Dependence

• Coarticulation
  – At any moment in time, the current state of the vocal tract reflects the influence of preceding sounds (perservatory coarticulation) and upcoming sounds (anticipatory coarticulation)
  – Such coarticulation is a property of any kind of skilled movement (e.g., tennis, piano playing, etc.)
  – It makes it possible to produce sounds in rapid succession (up to about 15/sec), with smooth, economical movements of slowly-moving structures.

• During the /æ/ in “camping” (Kent)
Effects of Coarticulation and Speaking Rate on velar movements

- The velum has to be raised to contain the air pressure increase of obstruent consonants
  - Its height during the /t/ is context (vowel) dependent - coarticulation
  - In the context of a nasal consonant, vowels in American English can be nasalized due to coarticulation
  - This is possible because vowel nasalization isn't contrastive in American English
- The velum (like most other vocal-tract structures) is slowly-moving
  - At higher speaking rates, its movements become attenuated


Figure removed due to copyright reasons.
Coarticulation of lip rounding for /u/

- Lip rounding in production of the vowel /u/ in /hə’tu/
  - The first three sounds in the utterance are neutral with respect to lip rounding
  - The lips are fully protruded before the utterance begins
  - Coarticulation takes place whenever it doesn’t interfere with transmission of the message
  - It crosses syllable and word boundaries
  - Movements of different structures are asynchronous

Figures removed due to copyright reasons.

- Cineradiographic measurements – anticipatory and perseveratory coarticulation
  - Tongue movement from /i/ to /a/ can start during the /p/ because it can’t be heard and it isn’t constrained physically
  - The consonants have effects only on the pellet position for the /a/ (not /i/ or /u/).
  - The pellet is at an acoustically critical constriction in the vocal tract for /i/ and /u/, but not for /a/.
- Note the vertical variation for /a/ (possible for constriction location – QNS).
Cyclical movements:
- Higher rates show decreased movement durations, distances, increased speed (a measure of effort)

Speech vs. cyclical movements:
- Compared to cyclical, speech movements generally are faster, larger, shorter – perhaps because they have well-defined phonetic targets

Vowels produced in fast vs. clear speech:
- larger dispersions, goal-region edges that are closer together – less distinct from one another
Persistence of inaudible gestures at word boundaries

U. Tokyo X-ray μ-beam
(Fujimura et al. 1973)

**List Production**

“perfect, memory”

Figure removed due to copyright reasons.

**Phrasal Production**

“perfec(t) memory”

- Phrasal: /m/ closure overlaps /t/ release, making it inaudible; /t/ gesture is present nevertheless (c.f. Browman & Goldstein; Saltzman & Munhall)
- Findings replicated and expanded with 21 speakers
- Explanation (DIVA): Frequently used phonemes, syllables, words become encoded as feedforward command sequences
Outline

• Introduction
• Measuring speech production
• What are the “controlled variables” for segmental speech movements?
• Segmental motor programming goals
• Producing speech sounds in sequences
• Experiments on feedback control
  – DIVA: feedback & feedforward control
  – Long term effects: Hearing loss and restoration
  – An example of abrupt hearing and then motor loss
  – Responses to perturbations – auditory and articulatory
    • “Steady state” perturbations
    • Gradually increasing perturbations
    • Abrupt, unanticipated perturbations
  – Feedback vs. feedforward mechanisms in error correction
• Summary
Feedback and Feedforward Control in DIVA

- With acquisition, control becomes *predominantly feedforward*
- Feedback control – Uses *error detection and correction* - to teach, refine and update feedforward control mechanisms
- Experiments can shed light on
  - *sensory goals*
  - *error correction*
  - *mappings between motor/sensory and acoustic/auditory parameters*
Learning and maintaining phonemic goals: Use of Auditory Feedback

- Audition is crucial for normal speech acquisition
- Postlingual deafness: Intelligible speech, but with some abnormalities
- Regain some hearing with a Cochlear Implant (CI):
  - Usually show parallel improvements in perception, production and intelligibility

Acoustic measures of contrast between /l/ and /r/ 6 months after receiving a CI

Phonemic contrast is enhanced pre- to post-implant – typical for CI users, many of whom have somewhat diminished contrasts pre-implant

Figures by MIT OCW.
Long-term stability of auditory-phonemic goals for vowels

• Typical pre- (○) and post- (●) implant formant patterns: generally congruent with normative data (■)
  - FA: some irregularity of F2 pre-implant (18 years after onset of profound hearing loss)
    - One year post-implant: F2 values are more like normative ones
• Phonemic identity doesn’t change; degree of contrast can
• Goals and feedforward commands for vowels generally are stable
  - If they degrade from hearing loss, can be recalibrated with hearing from a CI

Data from 2 cochlear implant users

Reprinted with permission from:
Copyright 1992, Acoustical Society America.
Long-term stability of phonemic goals for sibilants in CI users

- Subjects 1 and 3: good distinctions between /s/ and /ʃ/ pre-implant –
  - Typical, decades following onset of hearing loss

- Subject 2: reversed values and distorted productions pre-implant
  - After about 6 months of implant use, sibilant productions improved

- These precisely differentiated articulations are usually maintained for years without hearing
  - Possibly because of the use of *somatosensory goals* – e.g. pattern of contact between tongue, teeth and palate

Reprinted with permission from:
Copyright 1994, Acoustical Society America.
Responses to abrupt changes in hearing and motor innervation

An NF2 patient with sudden hearing loss, followed by some motor loss

• Two surgical interventions
  – OHL: Onset of a significant hearing loss (especially spectral) from removal of an acoustic neuroma
  – Hypoglossal nerve transposition surgery → Some tongue weakness

• /s-/ʃ/ contrast: Good until second surgery, when contrast collapsed

• Hypothesis: Feedforward mappings invalidated by transposition surgery
  – Without spectral auditory feedback, compensatory adaptation (relearning) was impossible – as might be possible with hearing
  – Somatosensory goal deteriorated without auditory reinforcement

Spectral median for /s/ and /ʃ/ vs. weeks in an NF2 patient
Vowel Contrasts and Hearing Status

- Compare English with Spanish CI users, CI processor OFF and ON
- Previous findings: Contrasts increase with hearing, decrease without
- Hypothesis: Because of the more crowded vowel space in English, turning the CI processor OFF and ON will produce more consistent decreases and increases in vowel contrasts in English than in Spanish

Figure by MIT OCW.

- Average Vowel Spacing (AVS) – a measure of overall vowel contrast
- Change of AVS from processor ON to processor OFF (for 24 hours)
  - AVS: decreases for the English speaker, increases for the Spanish speaker
AVS – by subject

- Prediction: AVS increases with the CI processor (hearing)
- Changes follow the predicted pattern more consistently for English than Spanish speakers

![Graph showing changes in average vowel spacing for English and Spanish speakers with AVS on and off.](image)
Modeling Contrast Changes: Clarity vs. Economy of Effort

- DIVA contains a parameter that changes sizes of all goal regions simultaneously – to control speaking rate and clarity (e.g., AVS)

![Graph showing contrast changes over time]

- Shrinking goal region size – like what English speakers do with hearing
  - Produces *increased clarity* (contrast distance), *decreased dispersion*
  - Without hearing, *economy of effort* dominates

- With fewer vowels in Spanish, clarity demands aren’t as stringent
  - Acceptable contrasts may be produced regardless of hearing status, without changing goal region size
Effect of Varying S/N in Auditory Feedback

- Normal-hearing and CI subjects heard their own vowel productions mixed with increasing amounts of noise
- In general, AVS increased, then decreased with increasing N/S
- Possible explanation: With increasing N/S
  - If auditory feedback is sufficient, clarity is increased
  - As feedback becomes less useful, economy of effort predominates
- Similar result for /s-ʃ/ contrast, but with peak at lower NSR
Bite block experiments

Figures removed due to copyright reasons.

• Speakers compensate fairly well with the mandible held at unusual degrees of opening
• Compensations may be better for quantal vowels (with better-defined articulatory targets)
• Presumably, the speakers mappings are not as accurate for the perturbed condition
• Compensation continues to improve, possibly with the help of auditory feedback
Mappings can be Temporarily Modified: Auditory Feedback
(Houde and Jordan)

• Sensorimotor Adaptation

• Methods:
  - Feedback (whispered) vowel formants were gradually shifted
  - 16 msec delay
  - Subjects were unaware of shift

• Results:
  - Subjects adapted for shift by modifying productions in the opposite direction
  - Effect generalized to other consonant environments and to other vowels
  - Effect persisted in the presence of masking noise: “Adaptation”
  - Adaptation was exhibited later, simply by putting subjects in the apparatus (no shift)
  - Speakers use auditory goals and auditory-motor mappings.

Figures removed due to copyright reasons.
Please see:
Sensorimotor adaptation

• Subjects hear own vowels with F1 perturbed, are unaware of perturbation

Thesis project of Virgilio Villacorta
– Based on work of Houde & Jordan, but with voiced vowels

• Subjects partially compensate by shifting F1 in opposite direction; Shift is formant-specific
• Mismatch between expected and produced auditory sensations → Error correction
• 20 subjects: Varied in amount of compensation
• Is there a relation between perceptual acuity and amount of compensation?
Relation Between Adaptation and Auditory Discrimination

- DIVA and previous studies: Production goals for vowels are primarily regions in auditory space
  - Speakers with more acute auditory discrimination have smaller goal regions, spaced further apart
- Hypothesis:
  - More acute perceivers will adapt more to perturbation
- Measure of auditory acuity: JNDs on pairs of synthetic vowel stimuli
- Result: Hypothesis is supported

| jnd, ARI||F1 sep | F1 sep, ARI||jnd | jnd, F1 sep||ARI |
|---|---|---|---|---|---|
| $r$ | -0.717 | 0.661 | 0.656 |
| $r^2$ | 0.514 | 0.436 | 0.430 |
| $p$-score | 0.004 | 0.010 | 0.021* |
Modifying a Somatosensory-to-Motor Mapping

A “Force Field Adaptation” experiment (Ostry & colleagues)

• Methods:
  • Velocity-dependent forces applied (gradually) by a robotic device act to protrude the jaw: proportional to instantaneous jaw lowering or raising velocity
  • Jaw motion path over large number of repetitions (700) is used to assess adaptation, which may be evidence of:
    • Modification of somatosensory-motor mappings
    • Incorporation of information about dynamics in speech movement planning (Ostry’s interpretation)
Results

Summary and interpretation

- Subjects adapt to a motion dependent force field applied to the jaw during speech production.
- Kinesthetic feedback alone is not sufficient for adaptation; have to be in a “speech mode”.
- Control signals (mappings) are updated based on differences between expected and actual feedback.
- Information about dynamics is incorporated in speech motor planning.

Figures by MIT OCW.
Rapid drift of spectral median for /ʃ/  

- A CI “on-off” experiment

- Observations
  - Vowel SPL increased rapidly with CI processor off, decreased with processor on
  - Spectral median drifted upward toward /s/ during the 1000 seconds with processor off – Surprising, since the goals are usually stable
  - Hearing one aberrant utterance when the processor was turned on, speaker overcompensated to restore an appropriate /ʃ/

- Extremely narrow dental arches (and movement transducer coil on tongue) may have made it difficult for speaker to rely on somatosensory goal

- He may have had to rely predominantly on auditory feedback to maintain feedforward control on an utterance-to-utterance basis
Unanticipated Acoustic Perturbations (Tourville et al., 2005)

- Methods – like sensorimotor adaptation, but with sudden, unanticipated shift of F1
  - Subjects pronounced /CʌC/ words with auditory feedback through a DSP board
  - In 1 of 4 trials, F1 was shifted upward toward /æ/ or downward toward /ɜ/
- Results (averaged across 11 subjects)
  - Subjects produced compensatory modification of F1, in direction opposite to shift
  - Delay of about 150 ms. – compatible with other results, in which F0 was shifted.
- Result is compatible with error detection and correction mechanisms in DIVA
How long does it take for parameters to change when hearing is turned on or off?

- Subject pronounced a large number of repetitions of four 2-syllable utterances (e.g., done shed, don said; quasi-random order).
- CI processor state (hearing) was switched between on and off unexpectedly.

- Example results for one subject
  - Changes not evident until second vowel
  - Change may be more gradual for SPL than for F0

- Results varied among parameter and subject
  - Perhaps related to subject acuity?
Unanticipated Movement perturbation – Motor Responses

Figures removed due to copyright reasons.
Please see:

- Abbs et al.; others (1980s)
  - In response to downward perturbation of lower lip in closure toward a /p/,
  - Upper lip responds with increased downward displacement, accompanied by EMG and velocity increases
  - The response is phoneme-specific
Further observations and interpretation (Abbs et al.)

- Coordinated speech gestures are performed by “synergisms” –
  - Temporarily recruited combinations of neural and muscular elements that convert a simple input into a relatively complex set of motor commands
- There are alternative interpretations (Gomi, et al.)

Motor equivalence at the muscle and movement levels

Figures removed due to copyright reasons.
Motor and acoustic responses to unanticipated jaw perturbations

(In collaboration with David Ostry)

- Robot used to perturb jaw movements
  - Triggered by downward movement
- 50 repetitions/utterance e.g., “see red”
  - 5 perturbed upward (resistive)
  - 5 downward (assistive)

- Formants begin to recover 60-90 ms after perturbation; jaw does not
- Two other subjects were similar
- Evidence of within-movement, closed-loop error correction

Figure by MIT OCW.
Compensatory Responses to Unexpected Palatal Perturbation
(Honda, Fujino & Murano)

- A subject pronounced phrase:
  /ia ia ia ia ia ia ia ia ia/
- Movements and acoustic signal were recorded
- Palatal configuration was perturbed by inflation of a small balloon on 20% of trials (randomly determined)
- Feedback conditions:
  – Feedback not blocked
  – Auditory feedback blocked with masking noise
  – Tactile feedback blocked with topical anesthesia
  – Both types of feedback blocked
- Measures:
  – Articulatory compensations
  – Listener judgments of distorted sibilants

Figures removed due to copyright reasons.
Please see:
Results

• Perturbation caused distortions in /ʃ/ production

  – Compensation and feedback:
    • With feedback not blocked, speaker compensated within about 2 syllables
    • With auditory or tactile feedback blocked, speaker was much less able to compensate
    • With both forms of feedback blocked, compensation was worst

• Results are compatible with
  – Sensory goals as basic units
  – Use of mismatches between expected and actual sensory consequences to correct feedforward commands

<table>
<thead>
<tr>
<th>Mean error score for fricative consonant identification (%)</th>
<th>Syllable No. 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal auditory-feedback</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steady-state deflated</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Inflation</td>
<td>83</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Deflation</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Steady-state inflated</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Masked auditory-feedback</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steady-state deflated</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Inflation</td>
<td>72</td>
<td>39</td>
<td>39</td>
<td>44</td>
<td>28</td>
<td>42</td>
<td>33</td>
<td>50</td>
</tr>
<tr>
<td>Deflation</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>11</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>Steady-state inflated</td>
<td>47</td>
<td>39</td>
<td>44</td>
<td>50</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
</tr>
</tbody>
</table>
Feedforward vs. Feedback Control and Error Correction

- In DIVA, feedback and feedforward control operate simultaneously; feedforward usually predominates.
- Feedback control intervenes when there is a large enough mismatch between expected and produced sensory consequences (sensorimotor adaptation results).
- Timing of correction:
  - With a long enough movement, correction is expressed (closed loop) during the movement (e.g., “see red”).
  - Otherwise, correction is expressed in the feedforward control of following movements (e.g., /ʃ/ spectrum, vowel SPL, F0 when CI turned on).
  - Correction to an auditory perturbation takes longer than to a somatosensory perturbation (presumably due to different processing times).
- Additional Examples of error correction:
  - Closed-loop responses to perturbations (see Abbs, others).
  - Feedforward error correction with, e.g., dental appliances.
  - Responses to combined perturbations (cf. Honda & Murano).
  - All are compatible with DIVA’s use of feedback.
Outline

- Introduction
- Measuring speech production
- What are the “controlled variables” for segmental speech movements?
- Segmental motor programming goals
- Producing speech sounds in sequences
- Experiments on feedback control
- Summary
Summary of Main Points

• Highest level control variables for phonemic movements
  – Auditory-temporal and somatosensory-temporal goal regions

• Goal regions encoded in CNS
  – Projections (mappings) from premotor to sensory cortex:
    Expected sensory consequences of producing speech sounds

• Goal regions defined partly by articulatory and acoustic saturation effects that are properties of vocal-tract anatomy and acoustics
  – Most vowels: goals primarily auditory; saturation effects, acoustic
  – Consonants: both auditory and somatosensory goals; saturation effects, primarily articulatory (e.g., any consonant closure)

• Articulatory-to-acoustic motor equivalence (/u/, /r/)
  – Help stabilize output of certain acoustic cues
  – Evidence that goals are auditory
Summary (continued)

• Auditory feedback (CI users)
  – Used to acquire goals and feedforward commands
  – Needed to maintain appropriate motor commands with vocal-tract growth, perturbations
• Goals and feedforward commands are usually stable, even with hearing loss
• Clarity vs. economy of effort
  – Tradeoff evident when hearing (CI) is turned on, off, in presence of noise
• Relations between production and perception
  – Better discriminators produce more distinct sound contrasts
  – Better discriminators may learn smaller, more distinct goal regions
• Feedback and feedforward control
  – Frequently used sounds (syllables, words) are encoded as feedforward commands
  – Responses to perturbations: intra-gesture are closed loop; inter-gesture are via adjustments to subsequent feedforward commands
Components have hypothesized correlates in cortical activation

Hypotheses can be tested with brain imaging

Can quantify relations among phonemic specifications, cortical activity, movement and the speech sound output
SOME REFERENCES


Selected References

Slide 6 (right figure):

Slide 16 (left figure):

Slide 18, Slide 27 (right three), Slide 28, Slide 55:

Slide 20-21:

Slide 23 (left):

Slide 23 (mid/lower 2):

Slide 24:
Selected References (cont.)

Slides 26 (left set):

Slide 26 (top right):

Slide 27:

Slide 42:

Slide 56, Slide 57:

Slides 66:
Neuroimaging Correlates of Human Auditory Behavior
"The Problem"

Figure by MIT OCW.
Sound Perception

Noninvasive Physiologic Measures

Neural Activity
Figure above illustrates how auditory evoked potentials are measured. Potential shown was evoked by a click stimulus and was recorded in a human subject. Waveform is the averaged response to many click presentations. (AEP record from R.A. Levine)
(Courtesy of Robert Aaron Levine. Used with permission.)
General Picture of Auditory Evoked Potential Generation

1. \[ \text{AEP}(t) = V_1(t) + V_2(t) + \ldots + V_k(t) + \ldots \]
   - Potential produced by the kth cell population
   - Number of cells in the population

2. \[ V(t) = \sum_{i=1}^{N} v_i(t) \]
   - Potential produced by a particular population
   - Potential produced by one cell in that population

3. \[ v_i(t) = u_i(t) \cdot r_i(t) \]
   - \( u_i(t) \) = Unitary potential (potential produced each time cell i discharges)
   - \( r_i(t) \) = Immediate discharge rate (instantaneous rate at which cell i discharges following the AEP stimulus)

4. \[ V(t) = \sum_{i=1}^{N} u_i(t) \cdot r_i(t) = u(t) \cdot \sum_{i=1}^{N} r_i(t) = u(t) \cdot R(t) \]
   - Compound instantaneous discharge rate

5. Factors affecting the amplitude of V(t):
   - Number of cells in the population
   - Degree to which cells in the population discharge in synchrony with each other
   - Amplitude of u(t), r(t)

(see Goldstein and Kiang, 1958; Wang, 1979; Melcher and Kiang, 1996)
Furst et al. (1985) “Click lateralization is related to the $\beta$ component of the dichotic brainstem auditory evoked potentials of human subjects”

For binaural clicks with different ITDs and ILDs, quantified

- perception

- binaural difference potential

- Attributes of the binaural difference are correlated with the perception of binaural sound.
Binaural Difference Potential

Fig. 9. Binaural, sum of the monaurals, and binaural difference waveforms for both species. The binaural (solid lines) and sum of the monaural waveforms (dotted lines) are superimposed. The difference between these two waveforms, the binaural difference (BD) waveform, is plotted below. The recording electrodes were vertex to nape for both species. Stimuli were 10/sec; rarefaction clicks at 40 dB SL for the cat and 38 dB HL for the human. (from Fullerton et al., 1987)

- Binaural difference (BD) is derived from BAEPs evoked by monaural and binaural stimuli (above). BAEP: brainstem auditory evoked potential
- The BD reflects an interaction between converging signals from the two ears at the level of the brainstem.
Cellular Generators of the Binaural Difference Potential

Cellular generators of the binaural difference potential in cat. Diagonal line shadings indicate the generators of the first peak (‘b’; white on black) and possible generators of the second peak (‘d’; black on white). The schematic of the lower auditory system (at bottom) shows the generators’ relationship to other cells. NLL, nuclei of the lateral lemniscus; IC, inferior colliculus.

(From Melcher, 1996)
Discussion Questions:

• If the generator results are combined with the findings of Furst et al., what can be said about the neural processing underlying sound lateralization and binaural fusion?

• We generally think of the MSO as a coincidence detector. Are Furst et al.’s binaural difference data consistent with this idea?
Late Responses: dependence on attention and stimulus context

Idealized AEP evoked by transient stimuli (___) including components that are dependent on stimulus context and subject attention (….., -----). (from Hillyard and Kutas, 1983; also see Hillyard et al., 1973; Donchin et al., 1978).

Nd - or “processing negativity”
- produced when the subject attends to the stimuli
- visualized by taking the difference between responses to attended and unattended stimuli

N2 - or “N2000”, “mismatch negativity”
- occurs in response to “rare” stimulus (S2 below) in oddball paradigm
- can occur even when the subject is not attending to the stimuli
- dependent on stimulus modality (e.g. auditory vs. visual)

P3 - or “P300”
- occurs in response to “rare” stimuli (S2 below) in oddball paradigm when the subject is attending to the stimuli
- independent of stimulus modality

Figure 1-12. Schematic diagram of oddball stimulus presentation paradigm for P300 measurement (from Squires & Hecox, 1983). Selected measurement parameters are indicated. Responses are averaged separately for Stimulus Type 1 (i.e. the frequent stimulus) and Stimulus Type 2 (i.e. the rare or oddball stimulus). Note. From “Electrophysiological Evaluation of Higher Level Auditory Processing” by K.C. Squires and K.E. Hecox, 1983, Seminars in Hearing, 4 (4), p. 422. Reprinted by permission.
(from Hall, 1992)
Fig. 1. Mean for eight subjects of the non-signal (NS), signal (S) and difference (Δ) waveforms at each electrode site in the auditory condition. Isopotential topographic distributions are expressed as percentages of maximum response amplitude for the N1 and P2 components of the non-signal response (left) and the negative (N2 Δ) and positive (P3 Δ) components of the Δ waveform (right). Supraorbital (0) and vertex (electrode 3) traces from the 3 runs are superimposed. (From Simson et al., 1977)

NS - responses to standard stimuli (2000 Hz tone bursts)
S - responses to rare stimuli (1000 Hz tone bursts)
Δ - response to rare stimuli minus response to standard stimuli

Fig. 4. Frontal, vertex, and parietal (across-subjects averaged) difference waveforms obtained by subtracting the ERP to the 1000-Hz standard stimulus from that to the 1044-Hz deviant stimulus at different deviant-stimulus probabilities. The continuous line indicates the counting condition and the broken line the ignore condition. The amplitude of the fronto-centrally distributed MMN is decreased when the probability is increased from 2% to 10%. When the two stimuli are equiprobable, no MMN is seen. From Sams et al., 1985)

MMN - mismatch negativity

- Stimuli: Syllables, varied along two continua
- Subjects: Children with and without learning problems
- Measured discrimination and mismatch negativity
- The children with learning problems showed
  - deficits in their ability to discriminate syllables
  - abnormally small mismatch negativity

Conclusion
The behavioral deficits in the children with learning problems arose at a processing stage that precedes conscious perception.
PET
(radioactive tracer e.g., radio-labeled H$_2$O)

Brain Activity Increase $\rightarrow$ Metabolic Response $\rightarrow$ Blood Flow Increase $\rightarrow$ Tracer Increase

fMRI
(Blood Oxygenation Level-Dependent (BOLD))

Brain Activity Increase $\rightarrow$ Metabolic Response $\rightarrow$ Blood Flow Increase $\rightarrow$ Blood Oxygenation Increase $\rightarrow$ Image Signal Increase
Functional Magnetic Resonance Imaging (fMRI)

Above, left: fMRI activation in brainstem, thalamic and cortical auditory centers. Each panel shows a color activation map superimposed on an anatomical image intersecting one or more auditory structures. The presence of color indicates a statistically significant (p < 0.001) difference in image signal between stimulus "on" and "off" periods. Color indicates significance level. The slice thickness: 5-7 mm. In-plane resolution for functional imaging: 3x3 mm.

Above, right: Image signal vs. time in auditory cortical areas showing activation.

Paradigm commonly used to obtain fMRI activation maps. Stimulus "on" periods are alternated with stimulus "off" periods while MR images are acquired at regular intervals. Tic marks on time axis indicate image acquisition times. A sample image is shown at one interval.
Computational inflation of the cortical surface. In the inflated format, the cortex of sulci and gyri can be viewed simultaneously. See Fischl et al. (1999) *NeuroImage* 9: 195-207.

(Courtesy of Irina Sigalovsky. Used with permission.)
Scott et al. (2000) “Identification of a pathway for intelligible speech in the left temporal lobe”

- Four stimulus conditions that included speech and several forms of degraded speech:
  - speech
  - noise-vocoded speech
  - rotated speech
  - rotated, noise-vocoded speech

- The stimuli differed in:
  - intelligibility
  - presence of phonetic information
  - presence of pitch variations

- PET activity was compared between conditions.

**Conclusion**

Processing unique to intelligible speech is performed anteriorly in the left superior temporal sulcus, while lower-level processing is performed more posteriorly in the left STS and STG.
Discussion Questions:

- Scott et al. argue that their choice of stimuli may be better than previous ones for identifying sites of speech-specific processing. Do you agree?

- What assumptions have been made about the relationship between brain activity and the functional specificity of a brain region?

- Three stimulus conditions:
  - visual (videos of tools or faces)
  - auditory (sounds of tools or voices)
  - audio-visual (simultaneous images and sounds)

- High resolution fMRI of the superior temporal sulcus, a known region of multimodal convergence

- Three types of cortical patches were identified having:
  - auditory > visual response
  - visual > auditory response
  - auditory = visual

Conclusion
“...suggested by our data is that auditory and visual inputs arrive in the STS-MS in separate patches, followed by integration in the intervening cortex.”
Figure 2. Statistical parametric maps for contrasts of interest (group data). a, SPMs are shown as “glass brain” projections in sagittal, coronal, and axial planes. b, SPMs have been rendered on the group mean structural MRI brain image, normalized to the MNI standard stereotactic space (Evans et al., 1993). Tilted axial sections are shown at three levels parallel to the superior temporal plane: 0 mm (center), +2 mm, and -2 mm (insets). The 95% probability boundaries for left and right human PT are outlined (black) (Westbury et al., 1999). Sagittal sections of the left (x = -56 mm) and right (x = +62 mm) cerebral hemispheres are displayed below. All voxels shown are significant at the p < 0.05 level after false discovery rate correction for multiple comparisons; clusters less than eight voxels in size have been excluded. Broadband noise (without pitch) compared with silence activates extensive bilateral superior temporal areas including medial Heschl’s gyrus (HG) (b, center, yellow). In the contrasts between conditions with changing pitch and fixed pitch and between conditions with changing spatial location and fixed location, a masking procedure has been used to identify voxels activated only by pitch change (blue), only by spatial change (red), and by both types of change (magenta). The contrasts of interest activate distinct anatomical regions on the superior temporal plane. Pitch change (but not spatial location change) activates lateral HG, anterior PT, and planum polare (PP) anterior to HG, extending into superior temporal gyrus, whereas spatial change (but not pitch change) produces more restricted bilateral activation involving posterior PT. Within PT (b, axial sections), activation attributable to pitch change occurs anterolaterally whereas activation attributable to spatial change occurs posteromedially. Only a small number of voxels within PT are activated both by pitch change and by spatial change.


(Copyright 2003 Society for Neuroscience. Used with permission.)
Zimmer and Macaluso (2005) “High binaural coherence determines successful sound localization and increased activity in posterior auditory areas”

- **Main Experiment:**
  - fMRI and behavioral measurements during sound localization
  - manipulate sound location using ITD
  - manipulate ability to localize by manipulating binaural coherence
  - identify brain areas showing a correlation between activation and localization performance

- **Control Experiments:**
  - separated activation specifically correlated with localization performance from activation correlated with binaural coherence.

**Conclusion**
Within the superior temporal plane, only planum temporale showed activation specifically correlated with localization performance. It was concluded that binaural coherence cues are used by this region to successfully localize sound.
Do sound recognition and sound localization involve segregated networks (i.e., "what" and "where" pathways)? This question was addressed by Maeder and coworkers (2001). In an fMRI experiment, subjects were imaged in three conditions: (1) during a localization task, (2) during a recognition task, and (3) at rest (see right).

FIG. 1 Schematic representation of the experimental paradigm, the blocks and the temporal structure of the stimuli. L = localization task; R = recognition task; r = rest.

Ventral cortical areas showed greater activity during the recognition task (green, below), while dorsal areas showed greater activity during localization (red, below).

FIG.5 Active paradigm: 3-D projections of activation on smoothed normalized brain (group results). Areas more activated in recognition than localization are shown in green, areas more activated in localization than in recognition are shown in red. Adapted from Maeder et al. (2001) NeuroImage 14: 802-816.

Figures removed due to copyright reasons.
Please see:

- Subjects performed a sound lateralization task before and after cortical stimulation using transcranial magnetic stimulation (TMS)

- TMS: a noninvasive stimulation method that reversibly alters neuronal function

- Stimuli: Dichotic tones with various ITDs

- Task: indicate perceived location (left or right)

- Stimulation site: posterior parietal lobe

**Conclusion**

TMS produced a shift in sound lateralization, suggesting a role for posterior parietal cortex in spatial hearing.
“What” and “Where” Pathways of the Visual System

The “what” and “where” pathways in the visual system include areas specialized for processing depth perception (symbolized by a pair of spectacles), form (an angle), color, and direction (the curve ahead sign). The result is object recognition (the “what” pathway) or object location (the “where” pathway).

Figure removed due to copyright reasons.
Please see:

REFERENCES


fMRI: Mitigation of Scanner Acoustic Noise

Clustered Volume Acquisition: A method for removing the impact of scanner acoustic noise on auditory fMRI activation

Figure removed due to copyright reasons.
Please see:
Slide 7:

Slide 8:

Slide 10 (Top figure):

Slide 10 (Bottom figure):

Slide 11 (Top figure):

Slide 11 (Bottom figure):
HST.722 Brain Mechanisms of Speech and Hearing
Fall 2005

Dorsal Cochlear Nucleus
September 14, 2005

Ken Hancock
Dorsal Cochlear Nucleus (DCN)

- Overview of the cochlear nucleus and its subdivisions
- Anatomy of the DCN
- Physiology of the DCN
- Functional considerations
Dorsal Cochlear Nucleus (DCN)

- Overview of the cochlear nucleus and its subdivisions
- Anatomy of the DCN
- Physiology of the DCN
- Functional considerations
The cochlear nucleus

DCN

VCN

AN
AN fibers terminate in a “tonotopic” or “cochleotopic” pattern.
Major subdivisions of the cochlear nucleus

Figure by MIT OCW.
Summary of pathways originating in the cochlear nucleus

Inferior colliculus
Lateral lemniscus
Superior olive

Figure by MIT OCW.
Projections suggest DCN is a different animal than VCN

- (All roads lead to the inferior colliculus)
- VCN projects directly to structures dealing with binaural hearing and olivocochlear feedback
- DCN ???

Figure by MIT OCW.
Dorsal Cochlear Nucleus

- Overview of the cochlear nucleus and its subdivisions
  - DCN projections do not reveal its function
- Anatomy of the DCN
- Physiology of the DCN
- Functional considerations
Dorsal Cochlear Nucleus

- Overview of the cochlear nucleus and its subdivisions
  - DCN projections do not reveal its function
- Anatomy of the DCN
- Physiology of the DCN
- Functional considerations
Figure by MIT OCW.
Dorsal Cochlear Nucleus

- **Overview of the cochlear nucleus and its subdivisions**
  - DCN projections do not reveal its function
- **Anatomy of the DCN**
  - more complex than other CN subdivisions
  - nonauditory inputs
  - similar organization to cerebellar cortex
- **Physiology of the DCN**
- **Functional considerations**
Dorsal Cochlear Nucleus

- Overview of the cochlear nucleus and its subdivisions
  - DCN projections do not reveal its function
- Anatomy of the DCN
  - more complex than other CN subdivisions
  - nonauditory inputs
  - similar organization to cerebellar cortex
- Physiology of the DCN
- Functional considerations

Photograph of Eric Young removed due to copyright reasons. Please see:

Eric Young
Response Map classification scheme

~ Auditory Nerve

Type I

Type II

Type III

Type IV

Type V

Sound Level

Discharge Rate

FREQUENCY

Exhibitory Responses

Inhibitory Responses

BF Tone

Noise

Spont. Act

Figures by MIT OCW.
DCN: Vertical cells are type II and type III units

- Narrow V-shaped region of excitation
- No spontaneous activity
- Tone response >> noise response

Evidence: Antidromic stimulation (Young 1980)
Antidromic stimulation

1. Record from neuron
2. Shock its axon

Diagram showing:
- Recording electrode
- DCN
- Fusiform
- Giant
- Vertical
- Stimulating electrode
- DAS
- VCN

Figures by MIT OCW
DCN: "Principal" cells are type III and type IV units

- "Island of excitation" & "Sea of inhibition"
- BF rate-level curve inhibited at high levels
- Noise rate-level curve ≈ monotonic

<table>
<thead>
<tr>
<th>Rate, spikes/s</th>
<th>Rate, spikes/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.25</td>
<td>2</td>
</tr>
<tr>
<td>300</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency, kHz</th>
<th>5 10 20 25</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.25</td>
<td>2 4 6 8</td>
</tr>
<tr>
<td>2</td>
<td>4 8 16</td>
</tr>
<tr>
<td>5</td>
<td>10 20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sound Level, dB attn</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>75</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>15</td>
<td>150</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
</tr>
</tbody>
</table>

Evidence:
- Antidromic stimulation (Young 1980)
- Intracellular recording and labeling (Rhode et al. 1983, Ding et al. 1996)

Figures by MIT OCW.
Neural circuitry underlying DCN physiology:

type II units inhibit type IV units

Figures by MIT OCW
Classic experiment: type II units inhibit type IV units

- Multiunit recording (Voigt & Young, 1980, 1990)

Cross-correlogram

Type IV firing rate given type II fires at t=0

Inhibitory trough

Type II unit

Type IV unit

Figures by MIT OCW
DCN physiology so far...

- type II units inhibit type IV units
- **BUT** this analysis based on pure-tone responses

⇒ what happens with more general stimuli???
Inhibition from type II units doesn't account for everything

- (DCN responses to broadband stimuli cannot be predicted from responses to tones; *nonlinear*)
- Type II units do not respond to notch noise—whither the inhibition?
- Response map has two inhibitory regions?
DCN notch noise sensitivity due to **wideband inhibition**

**Broadband noise**
- AN input to type IV unit

**Notch noise**
- AN input to WBI

**Figure by MIT OCW.**

- **Strong AN input dominates**
- **Type IV response is excitatory**
  - Type IV loses greater portion of its excitatory input
  - WBI input dominates
  - Type IV response is inhibitory

*Nelken & Young 1994*
**PVCN: Is the D-stellate cell the wideband inhibitor?**

- such responses arise from radiate or stellate neurons (Smith & Rhode 1989)
- stellate cells send axons dorsally into the DCN, thus called “D-stellate cells” (Oertel et al. 1990)
- D-stellate cells are inhibitory (Doucet & Ryugo 1997)

- broadly-tuned, onset-chopper units are found in the PVCN (Winter & Palmer 1995)
- typically respond better to broadband noise than to tones

---

**Figure by MIT OCW.**

---

**Figure by MIT OCW.**
Summary: Circuitry of DCN deep layer

Type II inhibitory inputs (strong)

UIS from ? source (weak)

Excitatory input AN fibers?

Excitatory

Inhibitory

Spirou and Young 1991

Frequency

Best frequency

BF Tone

Noise

Level, dB attn

Rate, spike/s

W.B.I.

W.B.I.

“reciprocal” response properties

Figures by MIT OCW.
Summary of DCN anatomy and physiology

- Nonauditory
- Somatosensory
- Auditory
- Vestibular

- granule

- Nonauditory
  - Somatosensory
  - Auditory
  - Vestibular

- fusiform

- vertical cell
- type II unit
- narrowband inhibition

- D-stellate cell
- onset-chopper
- wideband inhibition

www.bsscommunitycollege.in   www.bssnewgeneration.in   www.bsslifeskillscollege.in

www.onlineeducation.bharatsevaksamaj.net   www.bssskillmission.in

WWW.BSSVE.IN
Dorsal Cochlear Nucleus

- **Overview of the cochlear nucleus and its subdivisions**
  - DCN projections do not reveal its function
- **Anatomy of the DCN**
  - more complex than other CN subdivisions
  - nonauditory inputs
  - similar organization to cerebellar cortex
- **Physiology of the DCN**
  - diverse response properties
  - complex interconnections
  - highly nonlinear
- **Functional considerations**
• Overview of the cochlear nucleus and its subdivisions
  - DCN projections do not reveal its function

• Anatomy of the DCN
  - more complex than other CN subdivisions
  - nonauditory inputs
  - similar organization to cerebellar cortex

• Physiology of the DCN
  - diverse response properties
  - complex interconnections
  - highly nonlinear

• Functional considerations
Filtering by the pinna provides cues to sound source location.

“Head Related Transfer Function” (HRTF)

Outer ear gain vs. Frequency (kHz)

Transformation (dB)

0°, +15°, -15°

“first notch” frequency changes with elevation

Figures by MIT OCW.
Type IV units are sensitive to HRTF first notch

- Type IV units are *inhibited* by notches centered on BF
- *Null* in DCN population response may code for sound source location

Reiss & Young 2005

May 2000

Figures by MIT OCW.
Dorsal Cochlear Nucleus

• Overview of the cochlear nucleus and its subdivisions
  - DCN projections do not reveal its function

• Anatomy of the DCN
  - more complex than other CN subdivisions
  - receives nonauditory inputs
  - has similar organization to cerebellar cortex

• Physiology of the DCN
  - diverse response properties
  - complex interconnections
  - highly nonlinear

• Functional considerations
  - coding sound source location based on pinna cues
DCN is a "cerebellum-like structure"

Generic cerebellum-like structure

- Predictive Inputs:
  - Corollary discharge signals
  - Higher levels of the same modality
  - Other sensory modalities (e.g. proprioception)
  - ?

- Input from a sensory surface

- Cochlea

- Granule layer

- Molecular layer

- Principal cell layer

- Sensory input layer

Figure by MIT OCW
Synaptic plasticity: Long-Term Potentiation (LTP)

• “Classical” LTP demonstration at the hippocampal CA3-CA1 synapse
• LTP evoked by *tetanic* stimulation (mechanism involves NMDA receptors)
Electric fish provide clues to cerebellum-like function

black ghost knifefish
(*Apteronotus albifrons*)

Photograph removed due to copyright reasons. Please see the Nelson Lab home page:
http://nelson.beckman.uiuc.edu

- electrical activity detected by *electric lateral line*
- afferent activity transmitted to electric lateral line lobe (ELL), analogous to DCN
Electric fields provide information about nearby objects

- **BUT** the fish generates its own electric fields:
  - tail movements
  - ventilation

⇒ cerebellum-like ELL helps solve this problem

Bell 2001
What do cerebellum-like structures do???

- Subtract the *expected* input pattern from the *actual* input pattern to reveal unexpected or *novel* features of a stimulus.
  
  - DCN: pinna movement is *expected* to shift the first notch, independent of what the sound source is doing.
Dorsal Cochlear Nucleus

• Overview of the cochlear nucleus and its subdivisions
  - DCN projections do not reveal its function
• Anatomy of the DCN
  - more complex than other CN subdivisions
  - nonauditory inputs
  - similar organization to cerebellar cortex
• Physiology of the DCN
  - diverse response properties
  - complex interconnections
  - highly nonlinear
• Functional considerations: the DCN may...
  - code sound source location based on pinna cues
  - extract novel components of response
DCN may play a role in tinnitus

- percept of noise, ringing, buzzing, etc.
- affects up to 80% of the population
- 1 in 200 are debilitated
- (not voices in the head)

So why DCN? Because tinnitus...
- involves plasticity
- may involve somatosensory effects

Levine 1999
Dorsal Cochlear Nucleus

• **Overview of the cochlear nucleus and its subdivisions**
  - DCN projections do not reveal its function

• **Anatomy of the DCN**
  - more complex than other CN subdivisions
  - nonauditory inputs
  - similar organization to cerebellar cortex

• **Physiology of the DCN**
  - diverse response properties
  - complex interconnections
  - highly nonlinear

• **Functional considerations: the DCN may...**
  - code sound source location based on pinna cues
  - extract novel components of response
  - contribute to tinnitus
Slide 5:

Slide 6:

Slide 7:

Slide 12:


Slide 15:

Slides 16, 18, 19:

Slide 20:


Slide 22:

Slide 23:

Slide 24:


Slide 25:


Slide 26:

Slide 29:

Slide 30:

Slide 32:

Slide 35:
Cortical correlates of audio-visual integration

Introduction

Animals have multiple sensory systems with which they can probe the external environment. Oftentimes, objects or events in the environment produce signals to which more than one sensory system responds, and it is the task of the cerebral cortex to integrate these separate modalities into a unified percept. We take this process for granted, but it is in fact a difficult problem, which becomes obvious when we start to ask ourselves how we would design a system to do the same. The current state of knowledge described in the literature is quite limited and is mostly phenomenological, i.e. describing how auditory and visual inputs can either modulate each other and/or activate association areas of the cortex. There does not seem to be a mechanistic understanding of how this works.

I have collected a number of papers that look at audio-visual integration at different levels, from single-unit studies to cortical activation patterns to behavioral. The methods used range from intracellular recordings to scalp-recorded ERP (event-related potentials) to fMRI / MEG / PET imaging. For your pleasure and convenience, most of these papers are quite short. I have chosen three of them to discuss in class (Discussion papers), three for background, and quite a few for ‘further reading’ (if you only want to read a few of these, the most interesting for me are marked by ♦).

Behavioral effects of audiovisual integration

Ultimately multi-modal (or multi-sensory) integration should lead to noticeable behavioral effects. For human, the most interesting case is speech perception, which is usually considered an auditory modality but is strongly influenced by visual input. Being able to see the person that is talking to you leads to an increased intelligibility, especially in adverse conditions (high noise, reverberation, competing talkers) as shown by Sumby and Pollack (1954). The visual input gives redundant cues to reinforce the auditory stimulus but also disambiguates some speech sounds which differ in place of articulation but sound similarly (such as /ba/ vs. /da/). Sumby and Pollack (1954) have shown that especially at low acoustic signal-to-noise ratios, the visual signal can dramatically increase word recognition, in some cases from near zero to 70% or 80%. In higher signal-to-noise ratio conditions, the visual signal still contributes, but the absolute effect is smaller because auditory performance alone is already high.

Beside this synergistic effect of auditory and visual input, there are a few other effects that clearly demonstrate the strong interaction of these two modalities. The first is the so-called ‘ventriloquist’ effect, where a synchronous yet spatially separate audio and visual signal is heard as originating from the visual location. Macaluso et al. (2004) were interested in finding brain areas that mediate this integration of spatially separate yet synchronous information into a single percept, through PET scans of human brains, and identified an area in the right inferior parietal lobule to be activated especially in this
condition. Bushara et al. (2001) used PET scans to identify the right insula being most strongly involved in audiovisual synchrony-asynchrony detection.

Another famous audiovisual ‘illusion’ is the ‘McGurk effect’ (McGurk and MacDonald, 1976), where the sound /ba/ is combined with the visual image of a talker articulating /ga/, leading to a robust percept of /da/. The usual explanation for this effect is that the brain tries to find the most likely stimulus given the (in this case conflicting) auditory and visual cues. Note that this effect is extremely robust and not susceptible to extinction, even after hundreds of repetitions. This effect has contributed to the notion that audiovisual integration occurs at a pre-lexical stage, early in the neural processing pathway.

**Neural correlates of audiovisual integration for ‘simple stimuli’**

Given the strong behavioral effects of audiovisual integration described above, it would be very interesting to explore the neural basis for these. However, we will first explore some more basic properties of audiovisual integration.

The canonical view of cortical sensory processing is that each sensory modality has a primary unimodal cortex, several higher-order unimodal association cortices, and that finally the various sensory modalities interact in multimodal association cortices. For the auditory case, the primary auditory cortex is located in the superior temporal gyrus or BA41 (see Fig. 1 for an overview of Brodmann areas). This area is surrounded by a belt and parabelt region, which are the auditory association areas. The middle and inferior temporal gyri (BA 20, 21, 37) are multisensory association areas, mainly auditory and visual. In fMRI studies of human brain activation, these multimodal areas activate uniformly in response to multimodal stimuli. Using high-resolution fMRI, it has recently been shown by Beauchamp et al. (2004) that in fact these multisensory areas (at least in the superior temporal sulcus) contain a patchwork of auditory, visual, and audiovisual areas, and each a few mm in size. It appears that the various unimodal areas send projections to small patches of multisensory cortex, after which the modalities are integrated in the intervening patches.

From human imaging and animal studies it is clear that there are special cortical areas which have multisensory responses. Komura et al. (2005) found that multisensory responses can also be found at lower levels, specifically in the auditory thalamus – the medial geniculate body (MGB). Traditionally, the thalamus is thought of as a relay station between brainstem/spinal cord and cortex, sending signals upward; but it is also known that it receives massive projections from the cortex itself. Komura and coworkers recorded from MGB shell neurons (which receive the cortical projections) in rat during a reward-based auditory spatial discrimination task, which was paired with an irrelevant yet variable light stimulus. Although MBG neurons did not respond to the light stimulus alone, the response was strongly modulated by the visual input, in particular the response was greater when auditory and visual stimuli were matched and smaller when they were conflicting. The matched and conflicting condition also led to a decrease vs. increase in reaction time, respectively, which again demonstrates the utility of integrating multiple modalities (which would usually be in agreement with one another, leading to a faster reaction). Interestingly, MGB shell neurons were also modulated by the amount of reward, although that is beyond the scope of our discussion.
A conceptual difficulty with studies of multisensory integration is that the paradigm does not always permit one to be sure that integration has in fact occurred. For example, in the mentioned study by Macaluso et al. (2004) where spatially separated audio and video was used, it is in principle possible that there was no integration, even though one would expect it based on prior psychophysical data. Also, conditions that are designed to produce multisensory integration often necessarily use somewhat different stimuli than conditions that are not aimed at producing integration. Bushara et al. (2003) devised an ingenious method to circumvent these problems, and were able to use exactly identical stimuli that sometimes produced integration and sometimes did not. By comparing brain activation (BOLD fMRI) from either category they were able to find a correlation between larger activation in specific areas and the (non-)occurrence of integration.

Because of these results, they propose that multisensory association areas work in parallel with primary sensory areas, instead of as a higher-level end-station, which is the usual view. This would agree with Komura et al. (2005) who showed that error rates were sensitive to matched or conflicting simultaneous audiovisual signals, which also had strong neural correlates (enhanced or depressed rate responses).

A completely different aspect of multisensory effects on sensory processing is attentional cueing, i.e. one modality biases the observer such that the other modality will respond preferentially to the cued object. As an example, consider hearing a familiar voice calling your name from a crowd of people; you will reflexively turn towards the sound location and focus your visual attention of the same location, leading you to find your friend more quickly than had he been silent. Such cueing effects are well documented in psychophysical experiments, and it has been proposed that by directing attention towards an object, the neural signal from that object propagates faster through the brain. This is the proposed neural correlate of judging attended objects to appear earlier than unattended objects, even if they appear simultaneously. McDonald et al. (2005) used a sound cue in a visual task where subjects were required to judge which of two lights switched on earlier. They were able to show by measuring event-related potentials (ERPs) from the scalp that the attended (cued) light yielded a larger ERP than the unattended light, even if they were simultaneous. Somehow this difference in magnitude is translated into a delay in subsequent stages of processing that lead to perception. These findings may again corroborate to some extent the findings of Komura et al. (2005) in that cueing by one modality can influence the response of neurons in the other modality, with subsequent clear-cut behavioral consequences.

**Neural correlates of audiovisual integration in speech perception**

We have already described how important visual signals are for speech perception, both in the synergistic sense of aiding speech intelligibility (Sumby and Pollack, 1954) as well as in creating illusions such as the McGurk effect (McGurk and MacDonald, 1976). In the previous section we explored how relatively simple audiovisual stimuli that may or may not be fused into single objects can influence neural responses, both for individual neurons as well as for the whole brain. How is this for speech? Does brain activity differ between purely listening to speech as compared to listening and seeing speech (beside predictable activation of purely visual sensory areas)? And what about purely seeing speech (no sound)? Does this activate any of the same cortical areas? The latter question
of as academic interest in understanding language processing, but also has a more practical importance in that this is how many deaf people ‘listen’ to others, i.e. by lipreading, also called speechreading. It would have great value to know what areas are the most important for speech reading, and whether differences in speechreading ability (which are large between people) have a identifiable neural basis.

The classic paper in this context is Calvert et al. (1997). They used fMRI to find brain areas of increased activity when either listening to speech or silent lipreading, and the surprising finding was that silent lipreading activates primary and association auditory cortices. This was interpreted to mean that audiovisual integration occurs at a very early level in the neural pathway, even before association areas are activated (although these may cooperate in parallel, instead of hierarchical, cf. Bushara, 2003). Auditory areas were not activated by closed-mouth, non-speech movements of the lower face. A series of subsequent experiments similar to these were conducted, and with improvements in fMRI technology it became controversial whether primary auditory cortex is indeed activated by silent lipreading. Most investigators failed to find group-averaged activation of primary auditory cortex, although auditory association areas (e.g. BA 42, 22) do reliable activate during silent lipreading. Hall et al. (2005) found such kind of result, with the exception that for some proficient lipreaders, superior temporal gyrus did activate. It seems that currently there is no unambiguous answer to the question whether silent lipreading activates primary auditory cortex; there seems to be evidence pro and contra. Interestingly, Hall et al. (2005) describes other cortical areas that vary in activity as a function of speechreading proficiency. For example, high activity in the left inferior frontal gyrus was associated with poor speechreading ability. The explanation is that the greater task difficulty requires more extensive use of cognitive processes, which are located in the frontal lobe. However, the left inferior frontal gyrus is Broca’s area (BA 44, 45), traditionally assumed to support articulatory-based mechanisms of speech production and executive aspects of semantic processing. From these and other recent results, it is becoming increasingly clear that Broca’s area is probably also involved in language comprehension, supporting the formation of syntactic and semantic structure and syntactic working memory.

Paulesu et al. (2003) also found activation the perisylvian language area and of Broca’s area (BA 44) in lipreading, also for non-lexical lipreading (NLLR, formed by playing video backwards). Lexical lipreading (LLR, forward video) differentially activated the more anterior part of Broca’s area (BA 45), and also the left inferior temporal cortex. Therefore, Paulesu and coworkers assumed that these two areas might be particularly important for lexical access in lipreading. As in other recent imaging studies of lipreading, they found activation of auditory association areas (but not primary auditory areas) for both LLR and NLLR.

Conversely, auditory input can also activate visual sensory areas, as studied by Giraud and Truy (2002). In both normal and cochlear-implant subjects the fusiform gyrus and early visual cortex (BA 18, 19) was activated by listening to speech (no visual signal). It is assumed that the expectancy of visual correlates of speech are responsible for this effect. In cochlear-implant patients, the visual area activation was much greater than for normal-hearing subjects, showing the greater reliance they presumably place on visual cues. This shows that multimodal integration is flexible and can be strengthened when one modality is degraded.
References

Background

  Nice short paper showing the (large) benefit (in terms of words correctly perceived) of seeing the face (mouth) that is talking to you, with an interesting interaction of the size of the vocabulary.

  The famous ‘McGurk effect’ described: an auditory /ba/ paired with a visual /ga/ leads to a perception of /da/.

  Classic paper on neuroimaging of speechreading; was the first to show auditory cortex activation from speechreading alone. Later studies have usually found that primary auditory cortex is not activated by speechreading alone, throwing some controversy on the issue.

Discussion papers

  Shows that specific cortical areas are more active (fMRI) when an audiovisual stimulus is perceived as one, compared to when exactly the same audiovisual stimulus is perceived as separate unimodal events.

  Visible speech aids speech reception both when moving images are presented as well as still images. In this study, fMRI was used to assess if different cortical areas subserved these two visual categories. They found that similar language-based cortical areas were activated in either case, although stronger activation was seen for the moving images.

  Single-unit responses in rat MGB (belt region) are modulated by visual stimulus and reward, in an audiovisual discrimination task. Shows that neuron response predicts behavioral response in hit/miss/reject/false alarm conditions; also shows neural and behavioral correlates of task difficulty.

Further reading (♦ most interesting)

  MEG imaging of human brain in response to visual, auditory, and audiovisual letters. It was found in the superior temporal sulcus audiovisual responses are suppressed relative to the unimodal responses, in contrast to single-unit responses, which are usually potentiated for multimodal stimuli.

The authors wanted to identify which brain areas are primarily involved in detection audio-visual asynchrony, and found the right insula to correlate best with the asynchrony detection effort as measured by response latency.

  Cochlear-implant patients (but not normal-hearing subjects) produce strong visual cortex activation when hearing speech.

  A PET-scanning study of lexical and non-lexical (reversed video) lipreading implicates Broca’s area in speech perception, as well as some other areas. Associative (non-primary) auditory cortex was activated for both stimulus conditions.

  Using PET scanning the authors identify cortical regions that are sensitive to the synchrony and/or spatial coincidence of visual and auditory speech.

  The authors show with high-resolution fMRI that audiovisual association cortex has patches of auditory, visual, and audiovisual cortex, instead of a homogenous audiovisual cortex.

  Includes a literature overview of neuroimaging studies of speech/lip-reading. Investigates with fMRI which cortical areas are active both in auditory speech and visual speech comprehension, and defines some cortical areas that correlate with individual speechreading proficiency.

  Measuring ERP (event-related potentials) from human scalp the authors show that time-order judgments correlate with the magnitude of ERP, not its timing. In particular, they find no evidence for the prior held belief that attended stimuli propagate faster through the CNS than unattended stimuli. Auditory cues were used to bias attention towards one of two visual cues, which could appear simultaneously or with a time delay.
Figure 1. Outline of Brodmann areas with functional attribution. From http://spot.colorado.edu/~dubin/talks/brodmann/brodmann.html.
Neural Centers and Perceptual Characteristics of Auditory Short-term Memory
Anna A. Dreyer

Studies of auditory perception rarely incorporate findings regarding storage of short-term auditory memory. Investigators rarely incorporate auditory working memory into the neural circuits and anatomical regions they hypothesize underlies perception. This review focuses on the brain regions, neural circuits and perceptual characteristics of human auditory working memory, the storage of auditory percepts necessary to perform short-term, trial-by-trial tasks, in contrast to global, permanent memory formation. The neural areas participating in auditory working memory greatly overlap with those implicated in auditory perception, suggesting the importance of incorporating working memory into our understanding of auditory perception.

Investigators interested in elucidating the mechanisms of auditory memory have focused on characterizing the perceptual limitations and neurophysiological responses. Psychophysical studies seek to understand human properties of the representation, single-cell recordings look for responses characteristic of short-term retention and imaging studies implicate the brain regions necessary to perform various short-term memory tasks. In addition, lesion studies have been important in examining the perceptual and neurophysiological defects when a key area implicated in short-term auditory memory has been removed. This review will describe how each methodology has shaped the current understanding of the mechanisms underlying auditory memory. Retention of non-verbal, perceptual information needed in pitch and lateralization discrimination is discussed to allow for comparisons between neurophysiological studies in non-human primates and human perceptual and imaging studies.

Psychophysical studies find parametric segregation of short-term memory stores

Psychophysical studies have explored system storage capacities in delayed performance of tasks with respect to variation in sensory parameters, such as duration, pitch, spatial location and memory load of auditory stimuli. In particular, investigators have discovered that auditory memory has separate stores for important auditory parameters. Anourova et al. (1999) conducted experiments testing whether pitch and location short-term memory can be disrupted with pitch and location-based distracters. Subjects were asked to identify whether two stimuli, separated by a delay, matched in frequency, in cases where either a stimulus of a different pitch or in a different spatial location served as a distracter during the delay. Anourova et al. found that only stimuli of the same modality (pitch of spatial location) degraded performance, concluding that pitch and location are contained in different auditory stores. Findings of Clarke et al. (1999) corroborate those findings. Semal and Demany (1991) suggest that pitch and timbre also have distinctly independent representations in auditory working memory.

Other studies characterize the pitch or location-based stimuli that serve as the most distracting. For instance Deutch (1972) also asked subjects to compare two test tones and played distracter tones between the two test tones. Subjects were told to ignore the distracter tones and indicate whether the two tones were similar or different in pitch. However, as also found by Anourova et al (1999) and Clarke et al. (1999), performance degraded with distracter tones having similar pitches to the test tones. Performance worsened dramatically when the distracter tones were 2/3 tone away from the test tone, regardless of the tone frequency, and improved for distracter tones higher and lower in frequency. These findings indicate that distracter tones are
most effective when they are close in frequency to the test tones. This study and another conducted by the author a year later (Deutsch, 1973) also adds that the auditory memory storage is logarithmically organized, since the errors scaled with tonal distance between the distracter and test tone.

As evidence of further parameter segregation in auditory working memory, Clement et al. (1999) suggested that pitch and loudness memory processing have separate representations and possibly different processing centers. This study measured performance during a frequency and an intensity discrimination task and found that the patterns of performance degradation as a function of delay between the two test stimuli differs substantially, as shown in Figure 1.

This finding is analogous to different working memory domains processing contrast and spatial frequency in the visual system, and provides further evidence of different memory stores for important auditory working memory parameters.

**Neurophysiological studies find neural correlates of short-term retention in medial geniculate body and auditory cortex**

Relatively few studies have directly recorded neural responses from auditory cortical areas participating in short-term memory tasks. Studies have implicated the auditory cortex in short-term memory processing in cats (Neff et al., 1975), dogs (Chorazyna and Stepien, 1961) and monkeys (Stepien et al., 1960). This relatively primitive stage of cortical processing was thoroughly investigated by Gottlieb et al., (1989) who concluded that the cortical neurons retain information about the first tone frequency during interstimulus intervals (ISI) using a rate code. Gottlieb et al. trained monkeys to discriminate two sounds with a delay in a performance and a passive listening condition, and found that the firing rate during the ISI correlates with the first tone in both conditions. In addition, they found that the firing rate of some units depended on whether the frequency of the first tone matched the frequency of the second tone. This study strongly implicates these neurons in the circuitry of short-term working memory (STWM). Participation of auditory cortex in working memory is also supported by magnetoencephalography study of tone loudness (Lu et al., 1992).
In a similar study, Sakurai (1990) implicated the medial geniculate body (MGB), a processing center downstream from the auditory cortex, in addition to A1 in STWM. In his experiments, rats made discriminatory same/different responses regarding the frequency of two test tones, while Sakurai recorded spiking patterns from the auditory cortex, medial geniculate body and the inferior colliculus. Although many of the cells in each area showed differences in firing rates during the presentations of the two stimuli, some cells in the auditory cortex and the medial geniculate body showed differential, stimulus-dependent activity during the delay between the two stimuli, indicating involvement in retention. Both this and the Gottlieb et al. (1989) study suggest a neural correlate of retention of the two stimuli with a rate code dependent on the first stimulus.

Figure 2: Responses of a medial geniculate body (MGB) unit to demonstrate the delay correlate between tone presentation periods and the delay period. Left: diagrams represent peri-stimulus time (PST) histograms during a sample tone (3 s), the interstimulus delay (3 s) and the test tone (1 s). Data are presented for six types of trials with tones identified as H=high, L=low, and responses as G=go and NG=no go. Mismatch conditions are H-L and L-H. Right: Bargrams of median spike frequencies (not Hz) during the sample tone, delay and test tone, and for the first and second halves of the delay. The asterisk shows statistically significant differences (p<.05) between high and low tones during the sample tone, delay and test tone. Differences in the number of spikes also exist in both halves of the delay for different stimuli. Adapted from Sakurai (1990).

Another non-cortical area neurophysiologically implicated in working-memory in the hippocampus. Although this area is often associated with long-term, reference memory, Sakai (1994) finds that some hippocampal neurons are solely involved in either working or reference memory, whereas other neurons are involved in both types of memories.

Imaging studies (see below) implicate the prefrontal cortex (PFC) in STWM (Romanski and Goldman-Rakic, 2002). The PFC consists of several cytoarchitectonically defined regions and there is evidence that these regions both distinct with respect to sensory modality and
stimulus features (Levy and Goldman-Rakic, 2000; Carmichael and Price, 1995) and integrate information across sensory modalities (Fuster et al. 1990; Bodner et al, 1996). Romanski et al. (1999) find distinct regions in the belt and parabelt regions that have neuronal targets in the frontal lobe of the prefrontal cortex in the rhesus monkey. Since the belt and parabelt regions of A1 have implications of SHTM involvement (Gottlieb et al., 1989; Sakurai, 1990), these histological studies give further evidence that the prefrontal cortex incorporates auditory STWM neural substrate downstream. Neurophysiological recordings of PFC activity in relation to auditory STWM tasks would shed light on the neural circuitry involving the MGB, A1 and PFC, but such studies are lacking.

**Imaging work implicates auditory cortical, frontal, parietal and cerebellar areas in STWM tasks**

Functional imaging studies have shed light on neural correlates of perceptual processing and implicate various cortical areas in auditory working memory. These studies suggest that the auditory cortex as well as higher cortical, as well as cerebellar, areas participate in working memory processing, but the degree of activation and the precise neural correlate depends on the type of memory task. Many imaging studies find activation of the prefrontal cortex in STWM tasks (e.g. Romanski and Goldman-Rakic, 2002). Levy and Goldman-Rakic (2000) find “domain-specific” organization of working memory function in the prefrontal cortex, and activation of one or more of several subdivisions of dorsolateral prefrontal cortex, depending on the type of information processed. Romanski and Goldman-Rakic (2002) specifically implicate the dorsolateral and ventrolateral prefrontal cortices in retention of auditory stimuli. Celsis et al. (1999) also suggested frontal and prefrontal regions play a role in auditory working memory in maintenance of tonal patterns.

For pitch comparison memory tasks, Zatorre et al. (1994) found greater activation of the right frontal and temporal lobes, and the parietal and insular cortex in high memory load conditions. Temporal lobe activation is expected, with its recognized role in musical processing, as is frontal activation, known for executive functions. Parietal and insular cortical activation probably played a role in attention. These findings are in line with those of Griffiths et al. (1999) who also found more extensive right lateralized network including the cerebellum, posterior temporal and inferior frontal regions when subjects were making “same/different” judgments while comparing the pitch of 6 tones. This finding may be in sharp contrast with a pitch matching study by Gaab et al. (2003) of greater working pitch memory activation of the left supramarginal gyrus (SMG) and the left dorsolateral cerebellum, as well as other structures in the temporal cortex.
Figure 3: Brain activation pattern (P<.05, corrected) during the initial imaging time points (0-2s after the ends of the auditory stimulation). The pattern is dominated by strong (left>right) activation of the superior temporal gyrus bilaterally including Heschl’s gyrus and auditory association cortex; anterior, lateral and posterior to Heschl’s gyrus including the planum temporale and the supramarginal gyrus. Activation also includes bilateral posterior dorsolateral frontal regions, superior parietal region on the right side, left>right pre-SMA region and lobules V and VI of the left cerebellum.

Celsis et al (1999) also implicated left SMG and right frontal and primary and secondary auditory cortices in the activation of a task requiring memory judgments between tones of different pitch height or spectral content. Thus, there seems to be some disagreement about the side of the asymmetry of activation patterns in pitch memory tasks.

That the cerebellum has non-motor functions is found in several studies, including auditory verbal memory function (Grasby et al, 1993), tone recognition (Holcomb et al, 1998) and musical duration discrimination (Parsons, 2001). Mathiak et al. (2004) also implicates cerebellar involvement in non-verbal auditory memory. Although the latter study suggests that auditory cortex activated early in the delay between the first and second test tones, and that the SMG and cerebellum and activated later, Gaab et al. (2003) finds that the cerebellum activates immediately and sees cerebellar activation through the task. However, both studies agree on cerebellar involvement in STWM tasks.

For audiospatial memory tasks, Martinkauppi et al. (2002) found bilateral activation in superior, middle and inferior frontal gyri and posterior parietal and middle temporal cortices. These areas were similarly found in the Zatorre et al. (1994) pitch study but here with bilateral activation. Martinkauppi et al. (2002) also revealed similar activation in the audio- and visio-spatial tasks, indicating that audio- and visio-spatial performance in STM tasks are located along a common neural pathway.
Lesion studies indicate debilitation in STWM performance with bilateral auditory association cortical lesions.

Lesion studies in monkeys implicate superior temporal cortex, a high-level auditory association area, in short-term memory tasks (Colombo et al., 1990; Colombo, 1996). However, Colombo (1996) suggests that only bilateral lesions impact auditory short-term memory performance significantly. This study used a delayed matching-to-sample task with high-frequency and low-frequency tones. Monkeys had to indicate whether two consecutive stimuli (with 2 s pause) were the same or different. Monkeys received lesions of association cortex with testing after each unilateral operation. The lesions corresponded to cytoarchitectonic area TA, sparing primary (area TC) and secondary (area TB) of auditory cortex. Visual performance after the first operation was unaffected, but auditory performance was affected in some monkeys after many sessions and some were unable to achieve preoperative levels. With bilateral lesions, monkeys were unable to reach preoperative levels with extensive testing, with worse performance deteriorating with delay length.

This study indicates that good performance is possible with one superior temporal cortex, corroborating findings of greater unilateral activation (Zatore, 1994; Gaab, 2003; Celsis, 1999), and clearly implicates this area in STWM processing.

Conclusion: complex interactions of STWM circuitry

Combined psychophysical, neurophysiological, imaging and lesion studies implicate a complex network of neuronal areas involved in STWM processing. Psychophysical work suggests that different memory representations and processing exists for important STWM parameters such as pitch, timbre, spatial location and loudness, with only slight parametric interaction in memory tasks. Neurophysiological studies find neural correlates of rate-based coding and retention of information about prior stimuli in MGB and A1, but further studies need to map out the connections and coding schemes involving these centers, as well as other center implicated using imaging work. Functional imaging studies suggest a complex asymmetry in pitch and spatial location memory tasks involving primary and secondary auditory cortices,
frontal and prefrontal areas, parietal areas as well as the cerebellum. Studies disagree about the temporal patterns of activation and about the right or left asymmetry of activation. Lesion studies in monkeys suggest that only bilateral auditory cortical lesions significantly affect performance, a finding that may corroborate some imaging studies with asymmetric activation patterns of activity. Because of the wide involvement of many cortical (and sub/supracortical) areas in STWM processing as well as disagreement regarding activated areas, further investigations of perceptual auditory phenomena should keep in mind possible involvement of STWM in designing tasks and evaluating results.
Suggested Papers

Background

Assigned Papers

Further reading
Works Cited


Brain Attending a Cocktail Party

Introduction

Imagine after our topic presentation, we go to a crowded pub to celebrate. You are trying hard to ignore those eighties tunes blaring out of the loudspeakers while you are chatting with your colleagues. Suddenly, this familiar melody from a Mozart’s piano concerto grabs your attention – yes, your cell phone alerts you that your partner just called for the third time…

This classic phenomenon, in which our brain analyzes a scene by perceptually organizing sensory data to form auditory objects (or auditory streams), is often referred to as the “Cocktail Party Effect” (Cherry, 1953). Many cues have been identified that influence perceptual organization, but only little is known about the actual brain mechanisms underlying this phenomenon. In this proposed topic, we look at the latest development in the quest to find the neural basis for auditory stream segregation.

Background – Auditory grouping mechanisms

The ability to form auditory objects is important in the natural environment where sounds arriving at our ears are a resultant of all spectro-temporal components that may have arisen from different auditory events. We constantly analyze an auditory scene by trying to group related components from one source, and segregating out other frequency components that are not in the attended object (Carlyon, 2004). This ability of bringing acoustical events to the attention foreground may increase the chance of survival for a species in the animal kingdom through auditory awareness to the movement of their predators. Humans also rely on auditory grouping mechanisms for daily communication, especially in the presence of noise and other competing sources, since we need to group simultaneous components originating from a single source across frequency, as well as grouping events across time, in order to hear whole words and messages. In his seminal book, Bregman (1990) provides us with working definitions on the terminologies used in the world of auditory scene analysis. He described auditory stream segregation as:

The general process of auditory scene analysis in which links are formed between parts of the sensory data. These links will affect what is included and excluded from our perceptual descriptions of distinct auditory events.
If these links, which are also commonly referred to as streaming cues, are correct sensory parts across time, we refer to such perceptual grouping as sequential. However, if these sensory data coexist in time, and the formation of multiple auditory objects are as a result of perceptually partitioning of the spectral contents into distinct objects, e.g., harmonics in a vowel, it is referred to as simultaneous grouping. Apparent spatial location, onset / offset synchrony, frequency proximity, and fundamental frequency are but some of the common acoustic cues that we employed in auditory scene analysis.

**Experimental paradigm – Buildup of Streaming**

How does one systematically investigate the process of stream segregation that occurs in a complex auditory environment? A popular paradigm is to use an ambiguous auditory figure that could either be heard as one stream with a galloping rhythm (commonly labeled as “Horse”) or as two concurrent streams with two different tempi (“Morse”) (See Figure 1). The basic stimulus consists of a high tone A alternating with a low tone B, in repeated ABA_ sequences. If the frequency difference ($\Delta f$) between the A and B tones is small, then neighboring tones tend to bind together perceptually, resulting in a “Horse” rhythm. Conversely, if $\Delta f$ is large, the A and B tones are no longer bind to each other, resulting in a “Morse” rhythm. The tone repetition rate also influences the percept: the faster the repetition rate, the more binding there is between the A and B tones (van Noorden, 1975).

At intermediate values of $\Delta f$ and tone repetition rate, the initial galloping “Horse” percept changes after a few seconds of listening to a “Morse” percept (Anstis and Saida, 1985; Bregman, 1978; Carlyon et al., 2001). With this systematic change in auditory percept over time, it is possible to record neural responses at various points during an ongoing sequence of sounds without any change in the evoking stimulus, and compare the neural responses associated with dramatically different percepts.
Figure 1 For the correct parameters, these sequences are ambiguous and can be heard with one or two perceptual organizations with different rhythms: (Left): a characteristic galloping rhythm (“Horse”); (Right): 2 isochronous streams, like Morse code (“Morse”). Colored regions correspond to perceptual streams. (Taken from Fig. 2, Cusack, 2005).

Key Areas of Research

Many investigations into the neural correlates of auditory streaming employ the aforementioned ABA- paradigm, or the variants thereof, but their generic approaches and the specific questions they address can be divided into several distinct classes. One class of approach is to use single-unit recordings in animals to infer perceptual effects of stream segregation. Fisherman et al (2001, 2004) showed that multiunit spiking responses to tone sequences in the primary auditory cortex of awake monkeys follow the pattern that one might expect on the basis of published psychophysical data from human subjects. Bee and Klump (2004) performed similar sequential streaming experiments and recorded neural responses in the auditory forebrain of awake starlings. They concluded that while there are preattentive auditory processes, such as frequency selectivity and forward masking, that contribute to the perceptual segregation of sequential acoustic events having different frequencies into separate auditory streams, there may be additional processes that are required to account for all known perceptual effects related to sequential stream segregation. The major confounding factor in interpretations of these conclusions is that the neural response patterns that are putatively associated with the one- and two-stream percepts were always induced using physically different stimuli.

Another class of approach combines human neuromagnetic and behavioral measures. Gutschalk et al (2005) used magnetoencephalography (MEG) to examine the neural bases of auditory stream formation. They concluded that there is a tight coupling between auditory cortical activity and streaming perception, suggesting that an explicit representation of auditory streams may be maintained within nonprimary auditory areas. They hypothesized that selective adaptation by feature-specific neurons might be a general neural mechanism subserving perceptual organization. Cusack (2005) combined functional magnetic resonance imaging
(fMRI) with psychophysical experiments to test for an effect of perceptual organization across the whole brain.

The other class of approach addresses the specific interaction between auditory streaming and attention. This remains a controversial topic. Due to time constraint, the scope of the proposed discussion will not include literature that explicitly addresses this issue, but some key results are summarized here for completeness. It has been suggested that streaming is a preattentive process based on recording of event-related potentials (ERP) using a mismatch negativity paradigm (MMN) (Sussman et al., 1999). It was argued by these authors that since MMN, which can occur in the absence of attention, is elicited only in the infrequent odd-ball presentation within-streams, the two potential streams of sound were segregated without attention being focused on the auditory stimuli. However, Carlyon et al. (2001) and Cusack et al. (2003) argue that the buildup of streaming is an attentive process, and previous experiments by van Noorden (1975) suggests that listeners have control over their perception in the ambiguous $\Delta f$ region. In other words, whether or not streaming is an attentive process is not conclusive, but the techniques used to probe such a high cognitive process are, nonetheless, interesting to follow.

The last class of active research area is computational modeling of stream segregation (e.g., Beauvois and Meddis, 1991; McCabe and Denham, 1997). An interesting neural network model, known as ARTSTREAM (Grossberg, 2004), uses adaptive resonance theory to propose how the brain achieves auditory scene analysis. Even though this neural network structure does not have clear neural correlates, its putative structure may inspire neurophysiologists to find neural units that have similar behaviors. This is perhaps an interesting paper to discuss if the class would like to be exposed to neural network modeling.

**Recent Developments and Proposed Papers for Discussion**

In this month's publication of *Neuron*, Micheyl and colleague (Micheyl, 2005) reported their comparative experiments on auditory stream segregation of human and awake monkeys. Unlike in previous studies, they used identical stimuli to conduct psychophysical experiments in humans and single-unit extracellular recordings in the primary auditory cortex of awake monkeys. Interestingly, using signal detection theory discussed in previous sessions, the authors can now compare quantitatively the stochastic neural responses with the probabilistic perceptual judgments.
Earlier this year, Cusack (2005) used fMRI to test for an effect of perceptual organization across the whole brain and suggested that the intraparietal sulcus (IPS) might play a role in perceptual organization. This is also exciting in light of the growing evidence in the literature suggesting that the IPS is also important for binding in vision, touch and cross-modally.

In a sequel to Wang’s (2005) paper, Bartlett et al (2005) extended the investigation on recordings of auditory cortical neurons in awake marmosets to 2-sound sequences. Not only do their results demonstrate that persistent modulations of the responses of an auditory cortical neuron to a given stimulus can be induced by preceding stimuli, they also find that decreases or increases of responses to the succeeding stimuli are dependent on the spectral, temporal, and intensity properties of the preceding stimulus. Such long-lasting modulation by stimulus context in the cortex suggests these response properties may be important for auditory streaming and segregation.

All the proposed papers for discussion highlight some new experimental approaches or latest findings on the neural correlates of auditory stream segregation. This is a very active area of research on many fronts, ranging from neurophysiology to psychoacoustics. However, we are still in the infancy stage in the quest of understanding the brain mechanism behind auditory scene analysis. Next time when you are in a pub drinking a beer, marvel at how the brain accomplishes such an amazing feat!

Papers for Discussion


Background


Further Reading


Reference


A Gene for Speech?
What the FOXP2 gene can tell us about
the neural control of speech and language

In contrast with all other animals, humans have evolved the ability to acquire spoken language. Humans’ capacity for speech must, in some respect, be derived from our genes. What is it in our genes that enables us to acquire speech? What would such a gene do, and what neural structures would depend on its expression? What would happen if it were disrupted? There has been speculation about a genetic substrate for speech ability for a long time, but it wasn’t until very recently that scientists were able to track down and identify a candidate gene, one that would be necessary for the acquisition of spoken language.

A speech-impaired family known as the KE family afforded geneticists a promising opportunity to look for the “speech gene.” Half the members of the KE family have severe impairments in speech and language, implying a hereditary mutation that is passed on in a simple manner. Those with the disorder exhibit deficits in speech motor control: they have trouble with fine movements in the lower half of the face, disrupting their ability to speak. However, their disorder is not entirely articulatory: those with the mutation also have defects unrelated to verbalization, for example, a deficit in making a written list of words starting with a particular letter.¹

In 2001, after years of research with the KE family and CS, an unrelated patient with a similar speech abnormality, geneticists Lai and Fisher alighted on the FOXP2 gene on chromosome 7q31, a region known to be related to brain function. (The name FOXP2 comes from forkhead box, for the forkhead box protein family involved in transcription.) FOXP2 is expressed in many regions in the developing
brain, including the cortical plate, basal ganglia, thalamus, inferior olive, and cerebellum. A single base-pair mutation caused the FOXP2 mutation resulting in speech defects in both CS and the KE family.\(^2\)

To determine FOXP2’s contributions to speech ability – to establish what it actually does when it is expressed – recent studies have looked for differences in brain structure between normal humans and those with mutated copies of the gene. If we can determine what cortical structures the gene is regulating, and correlate structural irregularity with behavioral impairments, we can learn more about the neural circuits for speech processing and articulation. The papers I have proposed elucidate the role of FOXP2 with data about its expression in speech and motor centers in the brain, as well as the anatomical, functional, and behavioral consequences of its disruption.

**PAPER 1: Liégeois et al., 2003**, *neuroimaging (fMRI)*. The functional abnormalities in members of the KE family can be revealed in functional imaging studies. Results by Liégeois show a large discrepancy in brain activation between subjects with normal and mutant copies of the FOXP2 allele. Specifically, affected family members showed an atypical distribution of activation during both speech (nonword repetition) and language (word generation) tasks. Particularly striking was an abnormally broad activation throughout cortex and a simultaneous underactivation in Broca’s area and other speech-related cortical and subcortical brain regions.\(^4,5\) The study also discusses morphological abnormality in these structures, principally the caudate nucleus and the putamen. Other related imaging studies\(^1\) use voxel-based morphometry to show bilateral reductions in gray matter in the ventral cerebellum. All this is strong evidence that FOXP2 mediates cortical processes related to language.

Figure removed due to copyright considerations.

Because FOXP2 is expressed in other animals, albeit in a slightly different form, we can induce mutations in the gene and examine the relations between gene irregularity and the resulting phenotypic differences. Shu et al. engineered two types of knockout mice, a type with two disrupted copies of FOXP2 (homozygous) and a type with one normal and one disrupted copy (heterozygous, the genotype of the affected members of the KE family). Heterozygous knockout mice made many fewer ultrasonic vocalizations in response to separation from the mother than did control mice, whereas homozygous knockout mice made no vocalizations at all. This result implies that FOXP2 plays a role in the development of a neural circuit employed in social communication, not uniquely related to human speech. In addition, heterozygous knockouts exhibited a delay in motor development, while homozygous knockout mice had a severe motor impairment, affirming the gene’s involvement in motor and sequencing areas of the brain.

Histological analysis of the mice illustrated irregular cerebellar sections in knockout mice, particularly in the Purkinje cells and in the external granular layer. FOXP2 therefore seems to be important for both cerebellar development and mechanisms of social communication.

PAPER 3: Teramitsu et al. 2004, histology & vocal learning in songbirds. FOXP2 is also expressed in songbirds, and current research seems to suggest it plays a role in birds’ vocal learning. According to Teramitsu et al., FOXP2 is localized to related subcortical structures in humans and birds (speech areas / song nuclei, cerebellar Purkinje cells). This colocalization suggests associated mechanisms underlying vocal learning in both humans (speech) and birds (song). FOXP2 and the related
gene FOXP1 show parallel expression in the human and songbird brain, predicting a possible role of FOXP1 in speech ability.

This topic is very relevant to studying the brain mechanisms for speech. The FOXP2 gene has the potential to be a powerful tool for exploring the neural circuits that allow us to communicate via spoken language. However, the field would certainly benefit from some spirited conversation about the limitations of these methods, as well as any unfounded conclusions that should be avoided. Furthermore, the function of the gene itself is an interesting topic of study. If FOXP2 is a “language” gene, what is it doing in mice, songbirds, and fungi? If it is merely a gene enabling fine motor movements such as those used in speech, what leads to defects in language development in individuals lacking the normal gene? What is the relationship between the motor mouth control FOXP2 may regulate and the language ability its disruption impairs? Does a hearing- or motor-feedback loop, such as that described in our speech motor control papers, play a part in any of this?

Finally, where did this gene come from? By comparing the gene’s expression in other animals, we can gain insight on when and how speech ability evolved in our species (and in no others).

The search for and evaluation of the speech gene, FOXP2, is an exciting and extremely recent issue, with broad-reaching implications for research on the evolution of human speech. Current research combines techniques from genetics, brain imaging, immunohistochemistry, and behavioral science to get at the function of FOXP2 in humans and animals. The FOXP2 gene is critical for the development of the neural substrates of speech and language, and will prove to be important to our understanding of circuits underlying speech and articulation.
References


* proposed for discussion articles
** proposed for review/background articles
Absolute Pitch

Jianwen Wendy Gu

October 26, 2005

HST.722 Brain Mechanisms for Hearing and Speech

Introduction

Pitch is a fundamental attribute of sound, which has led to extensive research on pitch processing, categorization, and memory with the goal of elucidating the complex workings of the auditory system. The phenomenon of absolute pitch (AP), the ability to identify or produce a specified pitch without external reference, provides a unique opportunity to study the perception and neural coding of pitch.

Background

Most people have at least some degree of AP, as it is fairly easy to identify a tone as belonging to the range of a piccolo rather than a tuba. Furthermore, even people with no musical training are good at remembering and singing the pitch of a favorite song. True AP possessors, however, can do these tasks with an order of magnitude better accuracy (Figure 1). From cognitive studies, AP seems to require two separable cognitive components: very narrow fixed pitch categories and the association of these categories with verbal labels. [4]

For people with AP ability, pitch categories are as fixed and familiar as color categories are for everyone else. However, this category structure for pitch does not affect their perception of pitch, only their ability to label. AP possessors and non-possessors have equivalent acuity and perceptual thresholds for pitch differences. [3]

AP is not an all-or-none ability; rather, it exists along a continuum. Some people with AP can label only tones produced by one particular instrument. Others have AP for only a single tone and fail to show the automatic and rapid identification found in true AP possessors. [3]
Figure removed due to copyright considerations.

Figure 1: Distributions of pitch-name responses to randomly presented tones, plotted as distance from correct response.

Key Issues

AP Acquisition

Musical training is essential for AP to manifest itself and most importantly, this training must happen early in life (Figure 2), which supports the hypothesis of a critical period during which AP can be acquired [3]. The type of training does have some influence, for

Figure removed due to copyright considerations.

Figure 2: Data from retrospective reports on age of acquisition of AP, as modeled by a Gamma function.

much musical training involves teaching of relative pitch, but it cannot account for all of
the variance in the population. In fact, AP ability often develops without explicit tutoring: mere exposure to tones and their labels can be sufficient. [4]

A fundamental question regarding AP acquisition remains to be answered: which neural events are responsible for this fairly fixed developmental time course? There are many neurodevelopmental processes with critical periods, such as language acquisition and the columnar segregation of neurons in the visual cortex. Understanding the molecular mechanisms that lead to AP acquisition could shed light on neural development in general. [4]

AP and Genetics

Since most people who receive musical training from a young age do not develop AP, early training can be only part of the story. Two recent findings point to a genetic component. First, there are significant associations between siblings who demonstrate AP even when environmental factors are controlled. Second, AP may be differentially distributed across different human populations. For example, people of Asian descent have a much greater incidence of AP than those of other backgrounds. Moreover, this higher rate of AP is found across several Asian ethnicities (Chinese, Korean, and Japanese) that are distinct from one another culturally and is unrelated to speaking tonal languages. Neither Korean nor Japanese is a tonal language, and the higher incidence has also been reported among Asian-Americans who speak only English. [4]

Although these observations support the idea of a genetic component, it is difficult to control all environmental factors. More rigorous methods and larger samples must be utilized to identify genetic influences of AP with certainty. [4]

Neural Correlates of AP

The development of structural and functional imaging techniques have enabled researchers to probe the neural correlates of AP. Anatomically, regions of the right superior temporal cortex and planum temporalis (PT) tend to be smaller in AP possessors than non-possessors. [3] Early studies using event-related potentials (ERPs) indicated that listeners with AP showed an absent or reduced electrical scalp component that is thought to index the updating of working memory. Listeners without AP had a physiological response to a certain pitch change, indicating that some type of online memory system had been refreshed. AP possessors did not show this response presumably because their memory representations consist of fixed, absolute values for each pitch. Therefore, pitch representation requires encoding into working memory but no updating once it’s in memory. [4]

A related phenomenon has been observed using measures of cerebral blood flow (CBF) as shown in Figure 3, specifically positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). An area of the right frontal cortex believed to be important for monitoring pitch information in working memory was more active among musicians lacking AP than in those with AP. AP possessors apparently used their categorical representation of tones in such tasks instead of requiring continuous maintenance of a sensory trace. [4]
Conversely, a different area of the frontal lobe, the posterior dorsolateral cortex, was more active in subjects with AP than in control musicians, when listening to tones. This area is known to be involved in establishing and maintaining conditional associations in memory. Thus, it is a logical candidate region for the link between a pitch and its label in AP possessors. Additional support for this idea comes from the finding that the same area is active in both subjects with AP and those without AP when asked to label pairs of tones that form musical intervals. [4]

These findings show that working memory and associative learning aspects of AP draw on neural resources that are reasonably well understood. However, this still does not explain the component of AP related to fixed pitch categories. Part of the answer to this issue may lie in the function of subcortical nuclei important for periodicity coding and coincidence detection, such as the inferior colliculus, which could provide input based on intrinsic oscillations to the auditory cortex of AP possessors. [4]
Topic Papers

People with absolute pitch process tones without producing a P300

Klein et al. measured the P300 in seven AP possessors and seven non-possessors in response to auditory and visual stimuli. The P300 is a positive-going component of the ERP, which is believed to be a manifestation of the processes of maintaining or updating working memory. It is obtained in the “oddball” procedure in which two discrete stimuli (one frequent and one rare) are presented in a Bernoulli sequence; the subject counts the rare stimulus, which elicits a large P300. The control subjects showed standard ERPs in both visual and auditory modalities. AP subjects, however, showed a small P300 for the rare auditory stimulus but a normal P300 for the rare visual stimulus. These results suggest that AP possessors have access to permanent representations of the tones so they do not need to utilize working memory to fetch and compare representations for novel stimuli. [2]

Functional anatomy of musical processing in listeners with absolute pitch and relative pitch

Zatorre et al. used structural and functional brain imaging techniques to investigate the neural basis of AP. Twenty right-handed musically trained volunteers were included in the study: 10 AP possessors and 10 subjects with good relative pitch (RP) but no AP. PET was used to measure CBF during the presentation of musical tones to the subjects. Similar patterns of increased CBF were seen in auditory cortical areas in both groups. The AP group also demonstrated activation of the left posterior frontal cortex, an area thought to be related to learning conditional associations. This activity was also observed in non-AP subjects when they made relative pitch judgments of intervals. Activity within the right inferior frontal cortex was observed in RP but not in AP subjects during the interval judgment task, suggesting that AP possessors need not access working memory mechanisms in this task. MRI measures of cortical volume indicated a larger left PT in the AP group. [5]

Absolute pitch and planum temporale

In this study by Keenan et al., anatomical MRI was used to measure PT size of a right-handed group of 27 AP musicians, 27 nonmusicians, and 22 non-AP musicians. The authors hoped to settle the debate over whether the increased leftward asymmetry seen in AP possessors was due to an increase in the left PT, a decrease in the right PT, or a redistribution and reallocation of resources between the two hemispheres. The AP musicians showed greater leftward asymmetry and smaller right absolute PT size compared to the other two groups. The increased PT asymmetry of AP musicians was not seen in the group of early beginning non-AP musicians, suggesting that this asymmetry might be genetically based. [1]
Conclusion

In summary, AP is an ability that involves both higher and lower levels of auditory processing. Its manifestation is influenced by the interaction of genes and environment. Thus, it is a neatly packaged phenomenon for studying many aspects of brain function.

References


