The Gestural Structure of Speech
Dance of the Vocal Organs
Secrets of the Dance...

- All the *information* in our messages (thoughts, ideas, etc.) of arbitrary complexity must ultimately be associated with unique simple dances.

- We can convey a potentially infinite number of different messages... infinite number of dances.

- Where in the dance is the *information***???
Words

• Words (or, morphemes) are elements of a language that have distinct meanings and which are also associated (arbitrarily) with distinct "dances" of the vocal tract organs.

• This is the informational or contrastive function of the dance.

• Despite differences among individual speakers, some essential properties of the mapping must be the shared for all speakers of a given language.

• Where in the dance do the essential properties reside?
• How do individuals perform the dance differently?
Gesture Hypothesis

• Dance can be decomposed into steps or gestures:

• A gesture is...

  • a constriction action of one of the vocal tract organs.

  • For example, the words "bad", "pad", "meter", and "Bingo" all begin with closure gesture of the lips organ.

  • We observe that in producing such words, speakers' upper and lower lips always come together to form tight seal.

  • This closure action is an essential property of the dance for these words.
Gestures as *units of information*

- Gestures are themselves meaningless, but they function to distinguish words (minimal messages units) from one another.

- Gestures of *distinct constricting organs* can be used to distinguish words from one another in all languages.

- For example:
  - "bad" begins closure gesture of the lips organ.
  - "dad" begins with a closure gesture of the tongue tip organ.
  - "gal" begins with a closure gesture of the tongue body organ.
Oral Constriction Gestures

LIPS  Tongue Tip  Tongue Body
Constriction Gestures: markers

LIPS  Tongue Tip  Tongue Body

apa  ata  aka
Gesture production

- Each gesture can be thought of as a (motor) “task” to achieve a goal for a particular task variable.
  - for example, the task in producing a consonants like /b, p, m/ is to reduce the distance between the lips (Lip Aperture, the task variable) to zero (the goal).

- What might be relevant goals and task variables for consonant and vowel gestures?

- The change over time of the state of the task variable can be controlled by a dynamical system with a target and a stiffness.

- The changing state causes changes in the articulators of the vocal tract that can produce those state changes.
### Gesture task variables

<table>
<thead>
<tr>
<th>Task</th>
<th>Articulators</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP</td>
<td>lip protrusion</td>
</tr>
<tr>
<td>LA</td>
<td>lip aperture</td>
</tr>
<tr>
<td>TTCL</td>
<td>tongue tip constrict location</td>
</tr>
<tr>
<td>TTCD</td>
<td>tongue tip constrict degree</td>
</tr>
<tr>
<td>TBCL</td>
<td>tongue body constrict location</td>
</tr>
<tr>
<td>TBCD</td>
<td>tongue body constrict degree</td>
</tr>
<tr>
<td>VEL</td>
<td>velic aperture</td>
</tr>
<tr>
<td>GLO</td>
<td>glottal aperture</td>
</tr>
<tr>
<td></td>
<td>upper &amp; lower lips, jaw</td>
</tr>
<tr>
<td></td>
<td>upper &amp; lower lips, jaw</td>
</tr>
<tr>
<td></td>
<td>tongue tip, tongue body, jaw</td>
</tr>
<tr>
<td></td>
<td>tongue tip, tongue body, jaw</td>
</tr>
<tr>
<td></td>
<td>tongue body, jaw</td>
</tr>
<tr>
<td></td>
<td>tongue body, jaw</td>
</tr>
<tr>
<td></td>
<td>velum</td>
</tr>
<tr>
<td></td>
<td>glottis</td>
</tr>
</tbody>
</table>

![Diagram of articulators and gestures](image)
Tasks, articulators, redundancy, synergy

- Performance of any skilled motor task requires cooperation of several independently moveable body parts, which we call articulators.
  - e.g., reaching for an object on a table

- There is large (possibly infinite) set of articulator postures that will achieve the task. This is sometimes called redundancy.

- When we learn to perform a task, we learn a pattern of dependency among the articulators specific to the task. This is called a synergy or a coordinative structure.

- The synergy allows the task to be performed in different ways in different environmental contexts.

- Different actors learn to “tune” the synergy differently, resulting in different articulator movements for the same functional task.
Harnessing Redundancy
Synergies in Speech

• Tasks in speech are gesture constrictions that form the consonants and vowels.

• Task that is in common to /p,b,m/ is the closure of the lips.

• What articulators are part of the synergy for a lip closure?
  • jaw
  • lower lip
  • upper lip

• Different people learn to tune the synergy differently: They employ different relative contributions of these articulators.

• Relative contributions differ when some perturbing events occur in the world.

• Relative contributions differ when the task is produced in the context of other tasks.
Speaker differences in Lip Closure synergies

“ship”

Speaker A

Speaker B

UL

LL

Jaw

LA

Time (ms)

Time (ms)
Speaker differences in Lip Closure synergies

“back”

Speaker A

Speaker B

UL

ULy

LL

LLy

Jaw

MNIy

LA

LA

Time (ms)

Time (ms)
Compensation for perturbation

- **Compensation**
  When the task is threatened by a perturbation of one articulator (e.g., yanking on the speaker's jaw as (s)he is about to produce a lip closure), other articulators, remote from the site of the perturbation, act to meet the challenge (e.g., increased displacement of upper lip) (Kelso et al, 1983).

- **Speed**
  Compensatory action is extremely fast (20 ms or so). This implicates direct inter-articulator cooperation. There is not enough time for an executive to "manage" responses to perturbation.

- **Task-specificity**
  Response to perturbation is task-specific, not hard-wired. If the subject is producing /z/, instead of /b/, response is not seen.
Lip Task performance in different contexts

- The relative contribution of the articulators in the synergy may differ when the task is produced in different contexts in which one of the articulators may be required for some other task.

- For example, lip closure in “back” vs. “been”.

- Jaw is recruited to be low in “back” because of the low vowel (/æ/) and high in “been” because of high vowel (/I/),

- More upper lip lowering emerges in “back” than in “been”.

![Graph showing time in milliseconds on the x-axis and upper lip height in millimeters on the y-axis, with two lines representing “back” and “been”.]
Tongue Tip Task performance in different contexts

- In the context of different vowels, the tongue tip closure task for (/t,d,n/) is produced with a different combination of articulators: tongue body, tongue tip.

- This is sometimes called “coarticulation”
Speech as *audible gesture*

- Gestures are analogous to the use of "signs" in languages such as American Sign Language (ASL).

- In sign, gestures of the arms, hands, and fingers are communicated optically to the visual system of the receiver.

- In speech, gestures of the tongue, lips, and larynx are (largely) invisible, but are communicated acoustically to the auditory system of the receiver.

- For gestures that are potentially visible, optic and auditory information are combined into a single gestural percept.
Gestures and sound production

- Two functions of sound production need to be distinguished:
  - Sound generation
  - Sound shaping

- Sound generation: causing air to vibrate at audible frequencies

- In the case of musical instruments, this is the function of lips against the mouthpiece in trumpet, or the air passing over the reed in a wind instrument.

- Device generating sound is the sound source.
Sound sources in speech

- vibration of the larynx
- turbulent ("jet") noise of air rushing thru narrow slit
- "pop" when built-up pressure is released
Sound shaping

- Generated sound is shaped, or filtered, by passing it through tubes of various lengths.

- Tube lengths determine the spectrum of the sound, the relative strength of the frequencies or overtones that compose it.

- In trumpets or wind instruments, different shaping is accomplished by fingering.

- In most mammals, the length of vocal tube is used as information about an animal's size.

- Filter functions in speech:
  - Constrictions of different organs produce changes in the effective lengths of vocal tract tubes.
  - Allowing air to pass through the nose or not.
Speech Gestures and sound

- Gestures of functionally distinct constricting organs can distinguish words:
  - Larynx (generates sound source)
  - Velum (shapes sound generated at larynx)
  - Oral constrictors: (shapes sound generated at larynx)
    - lips
    - tongue tip
    - tongue body
Types of Gestures

laryngeal
- opening (voicelessness) - "pad", "Sue"
- narrowing (voicing) - "bad", "zoo"

velic
- opening (nasal) - "mad"
- closing (oral) - "bad"

oral
Constricting Organs:
- Lips
- Labial - "bought"
- Tongue Tip
- Coronal - "dot"
- Tongue Body
- Dorsal - "got"
- Tongue Root
- Radical - "rot"
Laryngeal Gestures
Velic Gestures

“Jane may earn more money by working hard.”

Velic Open:
NASAL

Velic Closed:
ORAL

“Jane”

“hard”
Oral Constriction Gestures

- LIPS
- Tongue Tip
- Tongue Body
Gesture Combinations

• English words can begin with combinations of Oral, Laryngeal and Velic constriction gestures.

• The resulting combinations are usually analyzed as consonants or consonant segments.

• From the gestures we illustrated, we can form 9 combinations (consonants) in English.

<table>
<thead>
<tr>
<th>VELIC</th>
<th>closed</th>
<th>closed</th>
<th>open</th>
</tr>
</thead>
<tbody>
<tr>
<td>LARYNX</td>
<td>narrow</td>
<td>open</td>
<td>narrow</td>
</tr>
<tr>
<td>LIPS</td>
<td>“bought”</td>
<td>“pot”</td>
<td>“Mott”</td>
</tr>
<tr>
<td>TT</td>
<td>“dot”</td>
<td>“tot”</td>
<td>“not”</td>
</tr>
<tr>
<td>TB</td>
<td>“got”</td>
<td>“cot”</td>
<td>“pong”</td>
</tr>
</tbody>
</table>

• But there are more than 9 consonants in English. Where do the rest come from?
Differentiating oral constriction gestures

A given constrictor can produce several different distinctive gestures by varying:

- **Constriction Degrees** (how narrow is the constriction?)
  - stop ("dip, tip")
    complete obstruction of tube generates “pop” sound source
  - fricative ("zip, sip")
    narrowing to create jet noise source
  - approximant ("rip")
    narrowing with no source change

- **Constriction Locations** (exactly where is it?)
Constriction Locations for TT fricatives

dental
“thick”

alveolar
“sick”

palatoalveolar
“Shick”
Multiple oral constrictions

“lie”
- **Tongue Tip**
  CD: stop
- **Tongue Body**
  CD: approximant, CL: uvular

"rye"
- **Lips**
  CD: approximant
- **Tongue Tip/Body**
  CD: approximant, CL: palatal
- **Tongue Root**
  CD: approximant
Traditional (IPA) description of consonants and gestural analysis

(1) Laryngeal gesture results:
   - voiced (<laryngeal narrowing)
   - voiceless (<laryngeal opening)

(2) Location of oral constriction gesture
   - bilabial, labiodental
dental, alveolar, palato-alveolar
   palatal, velar, uvular, pharyngeal

(3) central or lateral

(4) Velic gesture results:
   - nasal (<velic opening)
   oral (<velic closure)

(5) Degree of oral constriction gesture
   - stop
dentritic approximate
Vowel and consonant gestures

- How do vowel gestures differ from consonant gestures?

(1) consonants are more constricted than vowels

  • exceptions?

(2) vowel gestures are formed more slowly and "last longer" than consonant gestures
Vowel gestures

“heed”

“hod”

“who’d”

3 basic vowels that occur in almost all languages

PALATAL

PHARYNGEAL

VELAR

CONSTRICION LOCATIONS
Fig. 1.4  The positions of the vocal tract in the author's pronunciation of the vowels in the words: (1) "heed", (2) "hid", (3) "head", (4) "had", (5) "hod", (6) "haved", (7) "hood", and (8) "who'd".
Systems for describing vowels

Location and degree of dorsal constriction

degrees: narrow mid wide (and intermediate)
locations: palatal velar uvular pharyngeal

Tongue Position system

high-low position of tongue body
front-back position of tongue body
round-unrounded

Formant (resonance) system
- high-low value of F1 (related to tongue body height)
- high-low value of F2 (related to tongue body front-back)