Quantal events generated by the structural and temporal variation of the vocal tract

Brad Story
Speech, Language, and Hearing Sciences
University of Arizona

Acknowledgements: Research supported by NIH R01-04789
Goal:

Develop a model of speech production that

- maintains physiologic relevance,
- is reasonably controllable and,
- to some degree, allows access to three levels:

Coordinated movement → Speech production → Speech perception

Implication: Model should produce intelligible speech at the syllable, word or sentence level

Parts of a VT area function model

- Trachea
- Voice source
- Vocal Tract
- 1D Tubular waveguide

- yielding walls
- \( \approx \) viscosity
- \( \approx \) heat conduction
- \( \approx \) lip radiation
- skin radiation
- nasal tract
- piriform sinuses

\[ r = \frac{A_1 - A_2}{A_1 + A_2} \]

\[ b_1 = f_1 r + b_2 (1 - r) \]

\[ f_2 = f_1 (1 + r) - b_2 r \]

Q. How do speakers choose specific spatial locations and temporal patterns with which to execute particular consonants?
Outline

1. Model of vocal tract shape

2. “Quantal” relation of constriction location and formant perturbation

1. Model of Vocal Tract Shape

Two Categories of Vocal Tract Movements

• **Shaping**: modulate the shape or geometry of the entire vocal tract - *vowels*

• **Valving**: produce and release constrictions - *consonants*

---


**Shaping:** Derived from principal component analysis of vocal tract shapes for vowels

\[
V(x) = \frac{\pi}{4} [\Omega(x) + q_1 \phi_1(x) + q_2 \phi_2(x)]^2
\]

Time-dependent variation of $q_1$ and $q_2$ to produce a vowel sequence

\[ V(x, t) = \frac{\pi}{4} [\Omega(x) + q_1(t)\phi_1(x) + q_2(t)\phi_2(x)]^2 \]

Postural component  Vocal tract deformation due to articulation

**Shaping**: modulate the shape or geometry of the vocal tract

**Composite**: time-varying vocal tract shape

**Consonant (valving):** requires specification of *constriction location, degree of closure, and temporal characteristics*

• **Tier I - Shaping**
  vowel shapes are imposed on neutral vocal tract

• **Tier II - Valving**
  consonant perturbations are imposed on the vowel substrate
Composite: time-varying vocal tract shape
Shaping + Valving = VCV:
Presence of a constriction perturbs the VV formant frequencies from their natural course.

Q. How do speakers choose specific spatial locations and temporal patterns with which to execute particular consonants?

2. “Quantal” relation of constriction location and formant perturbation – *for voiced stops*

How to relate changes in x-sect area to changes in formants?

\[ \frac{\Delta F_n}{F_n} \leftrightarrow \frac{\Delta a(i)}{a(i)} \]

Neutral vocal tract shape

Sensitivity function

\[ \frac{\Delta F_n}{F_n} = \sum_{i=1}^{N_{\text{areas}}} S_n(i) \frac{\Delta a(i)}{a(i)} \]

Acoustic Sensitivity Functions

At each resonance (formant) frequency, the difference of kinetic and potential energy along the vocal tract is calculated and divided by the total energy in the system.

\[ S_n(i) = \frac{KE_n(i) - PE_n(i)}{TE_n} \quad n = 1, 2, 3, \ldots \quad \text{and} \]

\[ i = [1, \ldots, N_{\text{areas}}], \]

\[ \frac{\Delta F_n}{F_n} = \sum_{i=1}^{N_{\text{areas}}} S_n(i) \frac{\Delta a(i)}{a(i)} \]

Schroeder, 1967;
Mermelstein, 1967;
Heinz, 1967;
Fant and Pauli, 1975
**Example:** Sensitivity of F1 to vocal tract shape change

Sensitivity function

\[
\frac{\Delta F_1}{F_1} = \frac{\sum_{i=1}^{N_{\text{areas}}} S_1(i) \frac{\Delta a(i)}{a(i)}}{N_{\text{areas}}}
\]

\(\Delta a(i) > 0\) and \(S_1(i) > 0\) \(\rightarrow\) \(\Delta F_1 > 0\)

\[F_1 \uparrow\]
Example: Sensitivity of $F_1$ to vocal tract shape change

\[
\frac{\Delta F_1}{F_1} = \sum_{i=1}^{N_{\text{areas}}} S_1(i) \frac{\Delta a(i)}{a(i)}
\]

$\Delta a(i) < 0$ and $S_1(i) > 0 \rightarrow \Delta F_1 < 0$

\[
\frac{\Delta F_1}{F_1} = \sum_{i=1}^{N_{\text{areas}}} S_1(i) \frac{\Delta a(i)}{a(i)}
\]

\[\Delta a(i) > 0 \text{ and } S_1(i) < 0 \rightarrow \Delta F_1 < 0\]
Sensitivity function

\[
\frac{\Delta F_1}{F_1} = \sum_{i=1}^{N_{\text{areas}}} S_1(i) \frac{\Delta a(i)}{a(i)}
\]

\[\Delta a(i) < 0 \text{ and } S_1(i) < 0 \rightarrow \Delta F_1 > 0\]

\[F_1 \uparrow\]
Sensitivity of F1, F2, & F3 to vocal tract shape change

\[
\frac{\Delta F_n}{F_n} = \sum_{i=1}^{N_{\text{areas}}} S_n(i) \frac{\Delta a(i)}{a(i)}
\]

Sensitivity functions divide the vocal tract into regions that, when constricted or expanded, will perturb F1, F2, and F3 in a unique pattern.

Upon **constriction**, formants are perturbed in the direction of the + or −.

Example: VCV
Constrict Region 4 (R4)
Occlusions - stops

**Probe** the vocal tract shape by generating VCVs with a continuum of constriction locations

Perceptual experiments (consonant identification)

Do these same regions exist with different vowel configurations?
Example: back to

i $\rightarrow$ a
Time-varying region plot

Region lengths and polarities change depending on vowel context

Time-varying region plot
Consonant activation

Need to activate the consonant during the time that the regions are organized as they are for the neutral tract shape

Consonant activation

Consonant activation

Consonant activation

Consonant activation

Goal is to use the time-varying region plot to guide modeling of word and sentence-level speech.

Example: 
“abracadabra”
Example of sentence-level speech:

“He had a rabbit and a boat”
Thank you