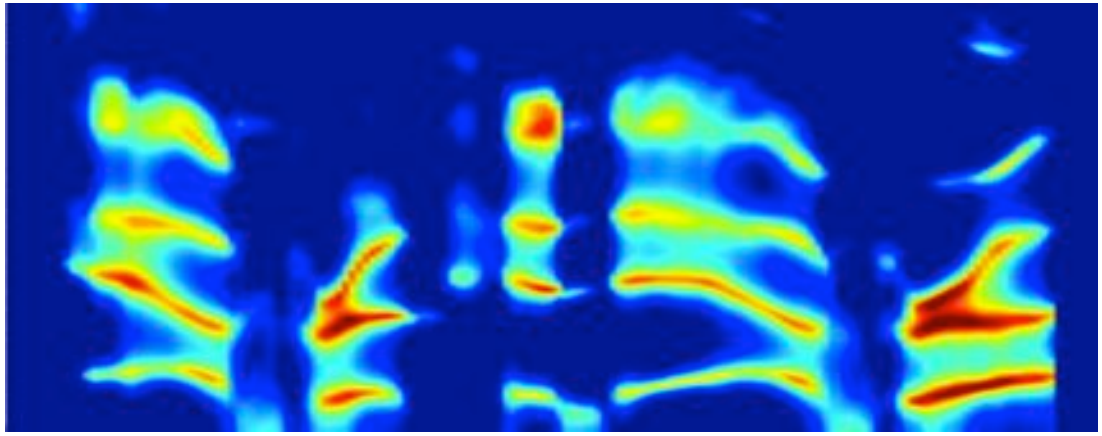
Willard R. Zemlin Memorial Lecture

Structure, Movement, Sound and Perception



Brad Story

Speech, Language, and Hearing Sciences

University of Arizona

November 15, 2013

ARIZONA

Disclosure Statements

- Much of the work reviewed, discussed, and demonstrated was funded by NIH R01 DC04789, R01 DC011275 and NSF BCS-1145011 B. Story, PI.

- A portion of the presenter's salary is funded by NIH DC011275.

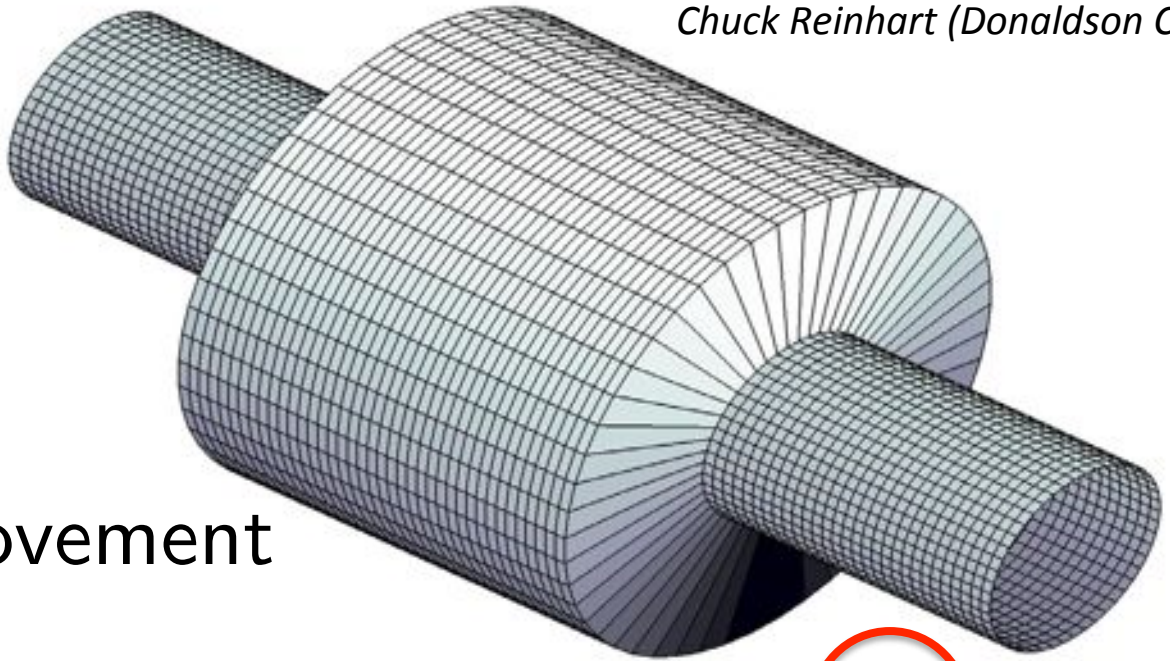
- The presenter has no financial interest related to the content of this presentation, but is receiving an honorarium.

- The presenter has no relevant nonfinancial relationships.

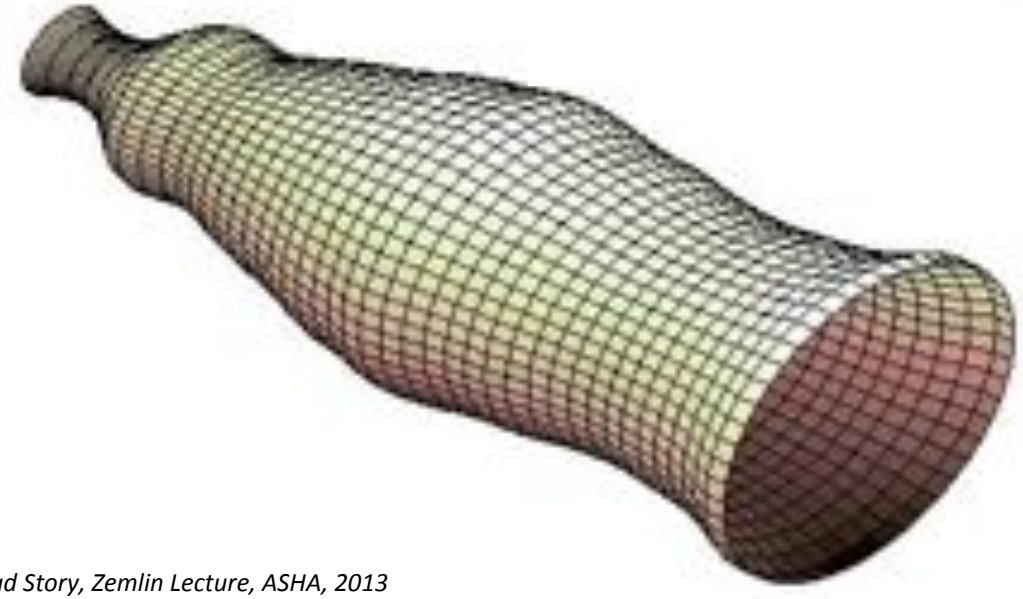
- The presenter's co-authors associated with some of the work presented here (Titze, Hoffman, Bunton) have no responsibility for the content of this presentation, and make no endorsement of it. The responsibilities of the content rest solely with B. Story.

Structure → Sound

Jim Rothman (Donaldson Co.)
Dr. Ron Anderson (Bemidji St. Univ.)
Chuck Reinhart (Donaldson Co.)



Structure → Movement



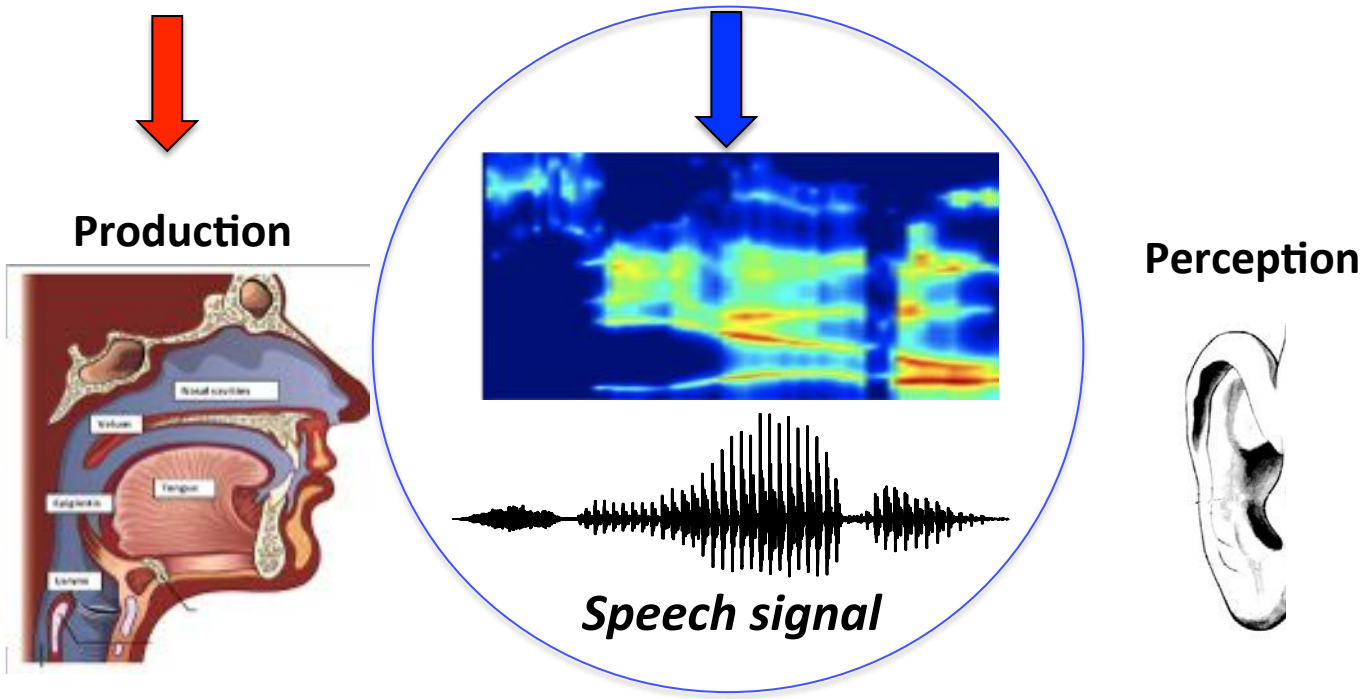
→ Sound

Brad Story, Zemlin Lecture, ASHA, 2013

Structure $\rightarrow$ Movement $\rightarrow$ Sound $\rightarrow$ Perception

Models of Speech Production:.

- Speech (or some aspect) produced by **mechanical**, **electronic**, or **digital** means that emulates the human **speech production system** ***and/or*** the **acoustic characteristics** of the speech signal.



Models of Speech Production:

Purpose?

"…indeed, the purpose of a model is to ***substitute simple structures for complex ones***."

"The great virtue of speech synthesizers [*models*] is that they help us make such simplifications."

-F. Cooper (1961). Speech synthesizers

Development of speech production models

pre -1600 **1600** **1700** **1800** **1900** **2000**

Joseph Faber (1844-)

German anatomist/mechanic

"The Amazing Talking Machine"

Perhaps the most well-designed and functional mechanical talking machine.

Represented the sound generating parts of the speech production system *(simulation)*

What was Faber's purpose in developing a talking machine?

It seems to have simply been a desire to create a machine that speaks like a human *(to simulate human speech production)*

Some observers noted that the machine had a "strong German accent" but in general spoke better English than Faber himself…

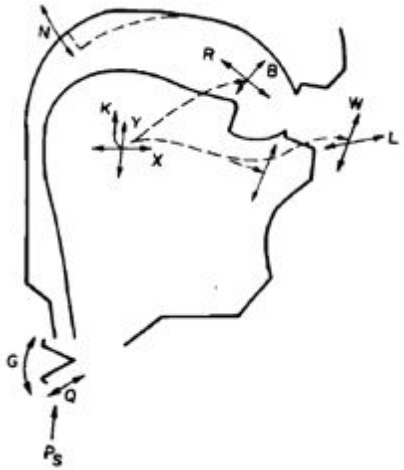
Faber's own speaking patterns were imposed on the hand and foot motions used to produce speech with the machine!

Articulatory Synthesis

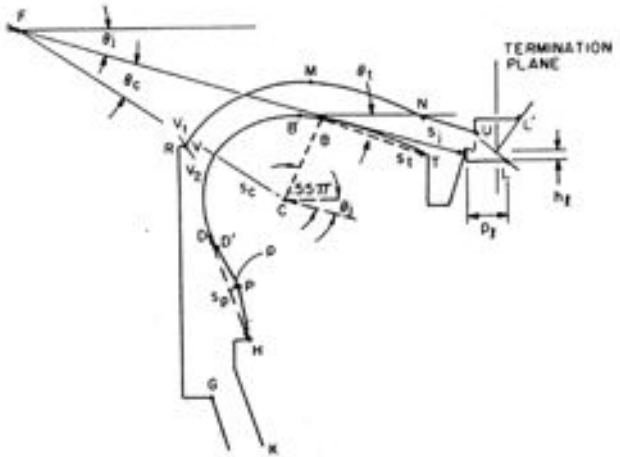
•Mathematical replication of the physics and physiology of the speech production system.

•Allows control of the positions and physical characteristics of the tongue, velum, jaw, lips, larynx/vocal folds.

Replication of observed articulatory movement to produce artificial speech



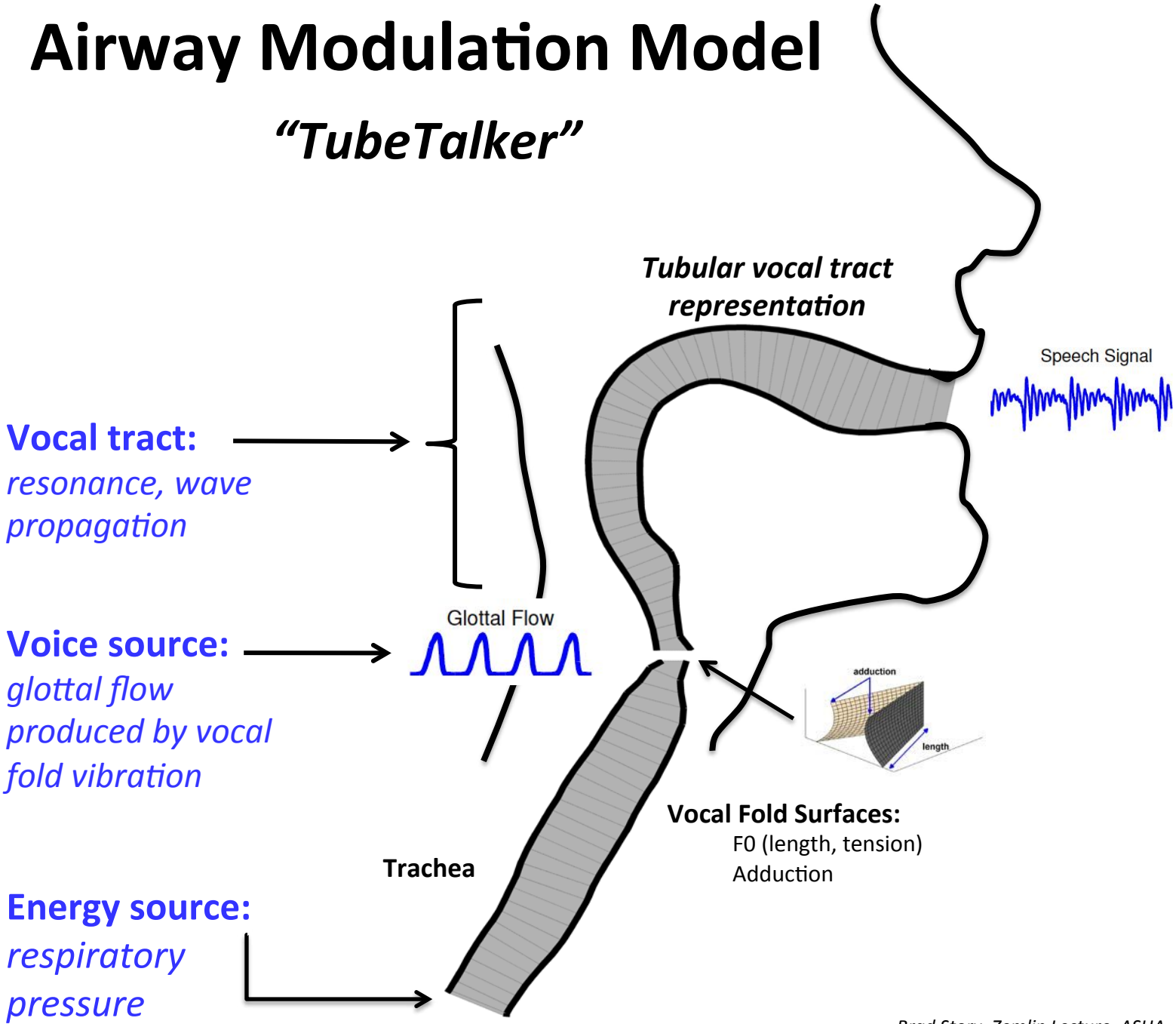
Coker, 1968; 1976



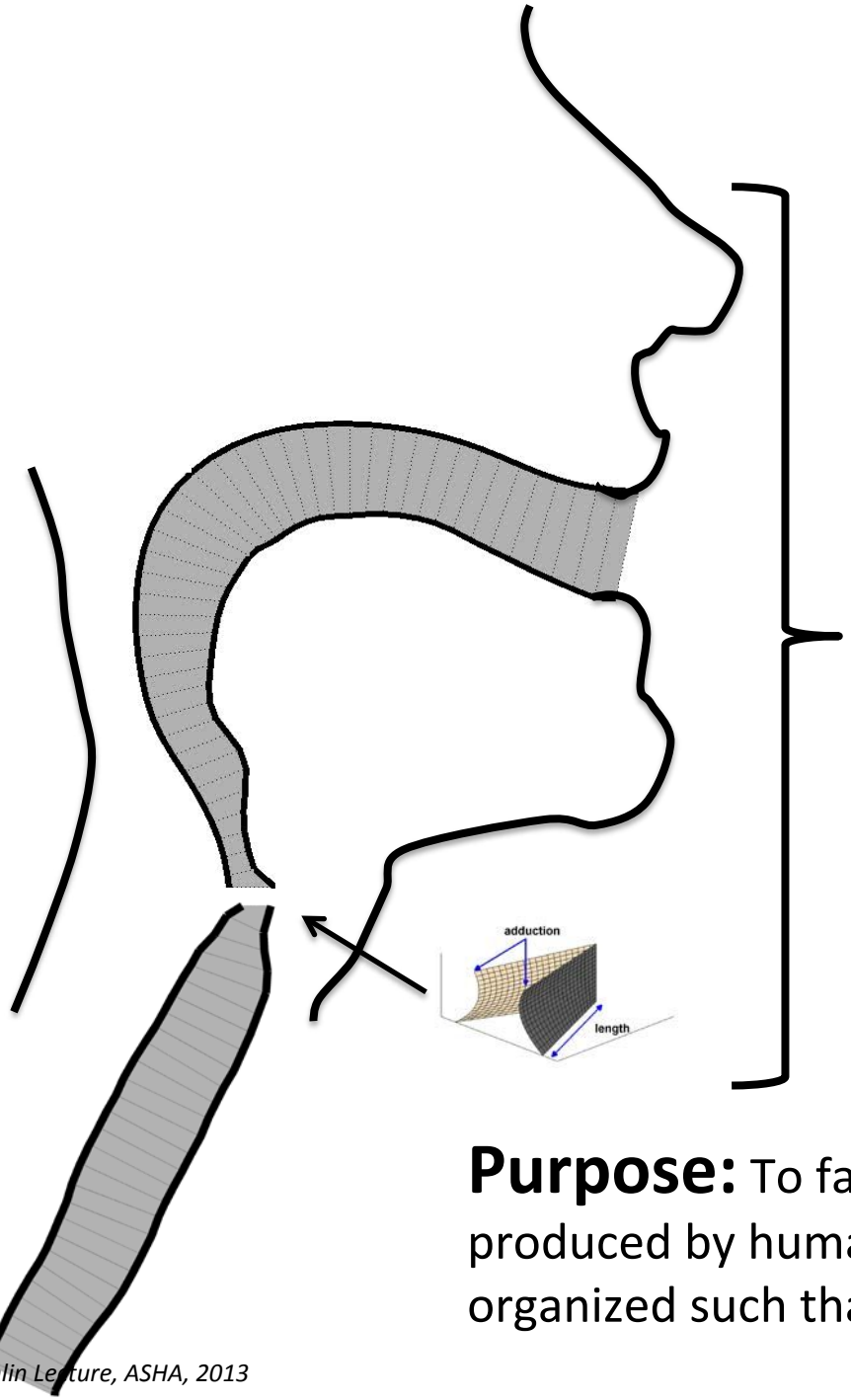
Mermelstein, 1973;
Rubin et al., 1981

Airway Modulation Model

"TubeTalker"



Brad Story, Zemlin Lecture, ASHA, 2013



Airway Structure

+

Modulation of the airway imposed by movement

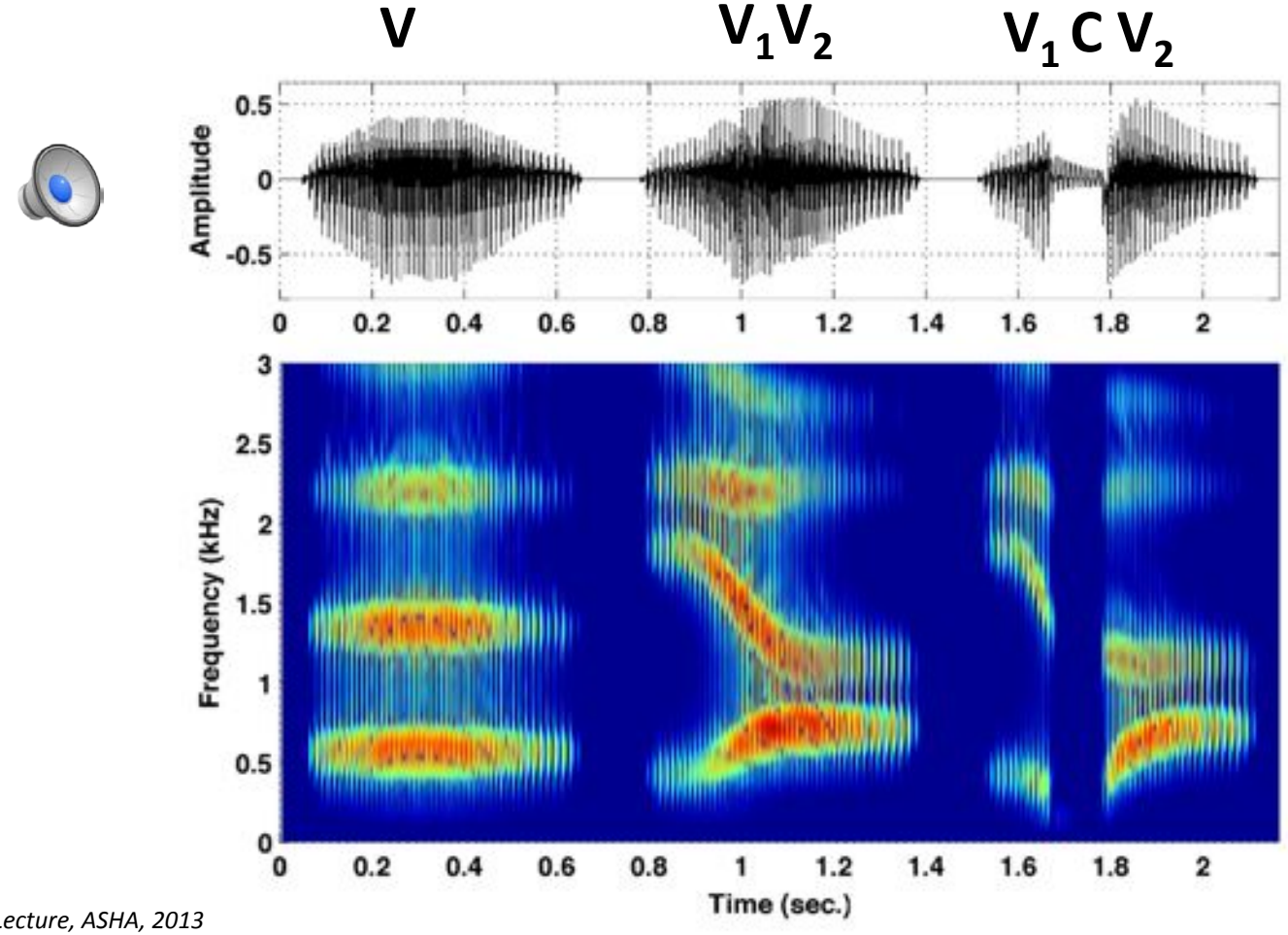
(articulation & vibration)

Purpose: To facilitate understanding how sound is produced by humans, and how that sound can be organized such that it is perceived as speech.

What can a model do for us?

– Gary Weismer, 2012 Zemlin Memorial Lecture

"Need to learn more about the ***flow of speech movements*** and the ***flow of acoustic 'hotspots'***, which presumably are locked onto by the ***speech perception*** mechanisms."



Brad Story, Zemlin Lecture, ASHA, 2013

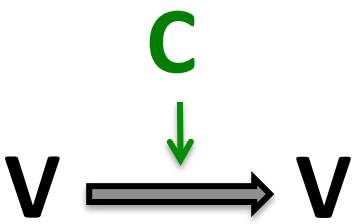
Two "Tiers" of Airway Modulation

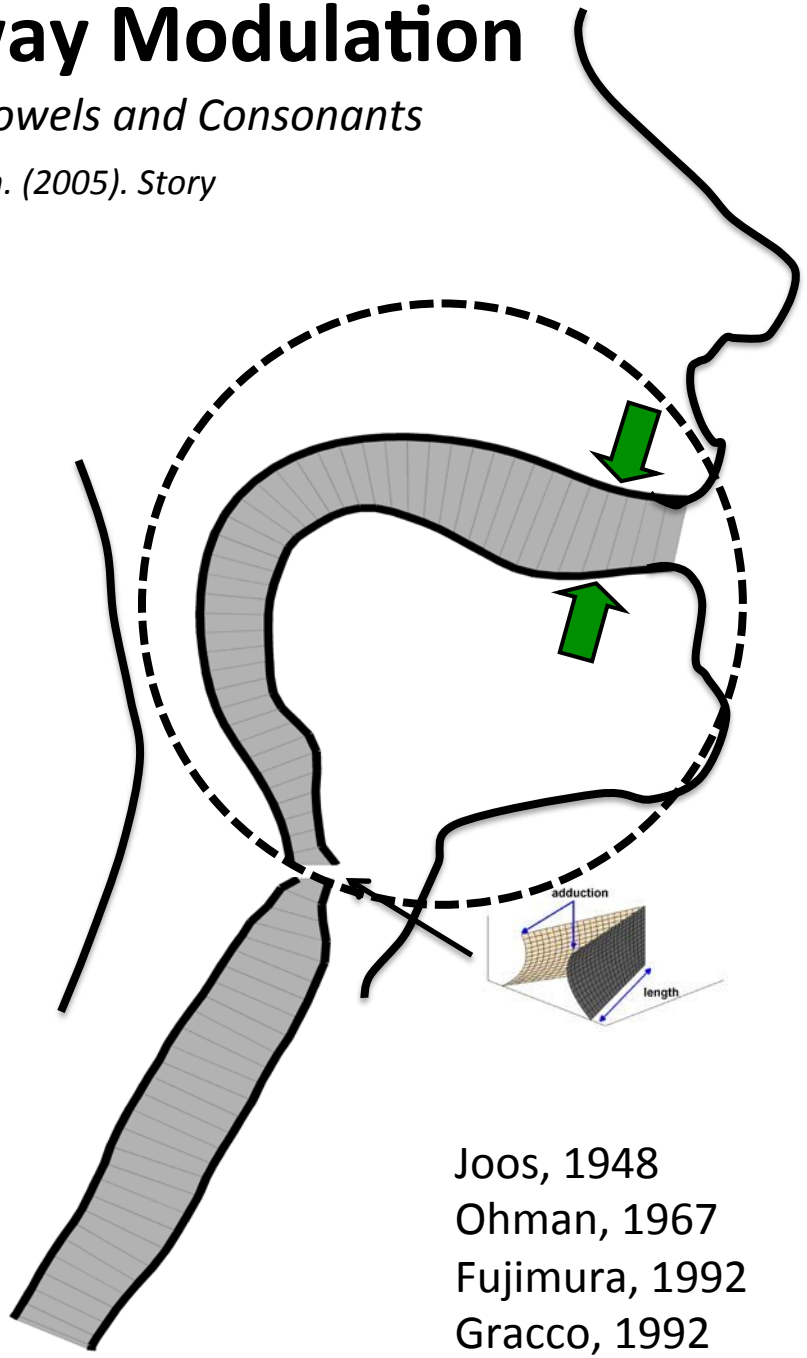
Simultaneous Production of Vowels and Consonants

J. Acoust. Soc. Am. (2005). Story

Tier 1 - Shaping: slowly-varying changes imposed on the shape of a neutral vocal tract - ***vowels***

- **Tier 2 - Valving**: modulate vowels with local and severe constrictions - ***consonants***



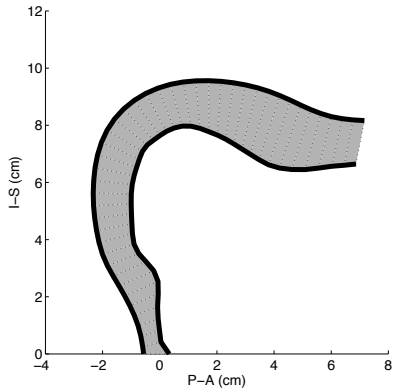


Joos, 1948
Ohman, 1967
Fujimura, 1992
Gracco, 1992

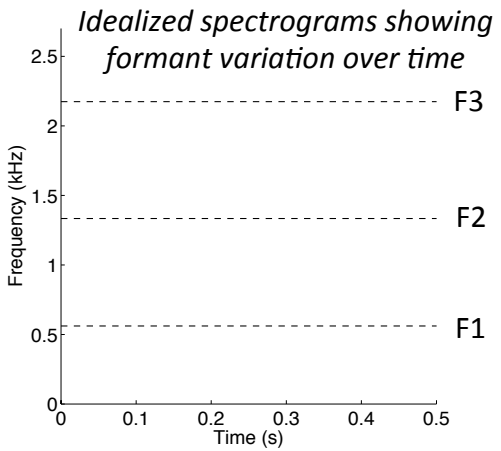
Two-Tier Model for Shaping and Valving the Vocal Tract

Base Structure:

"Neutral" vocal tract shape sets the acoustic "background"

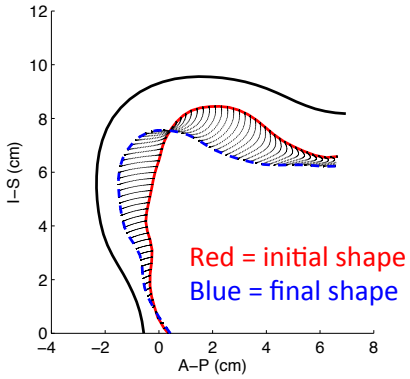


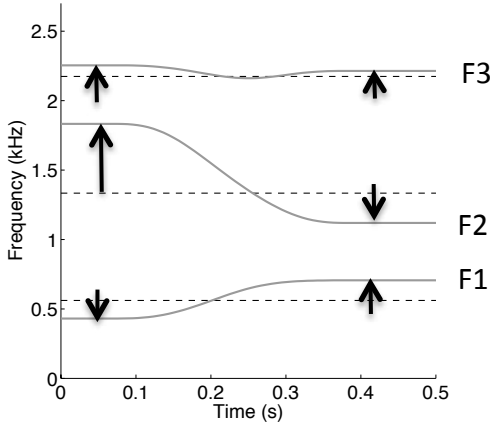
Idealized spectrograms showing formant variation over time



Tier 1: *Shaping*

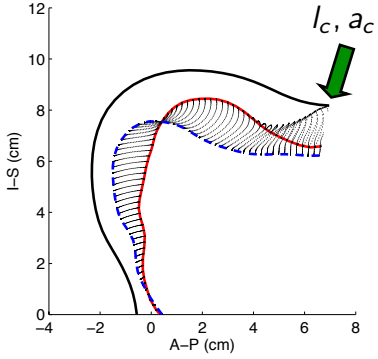
Global modulations of the neutral vocal tract shape perturb the formants over a "long" time scale

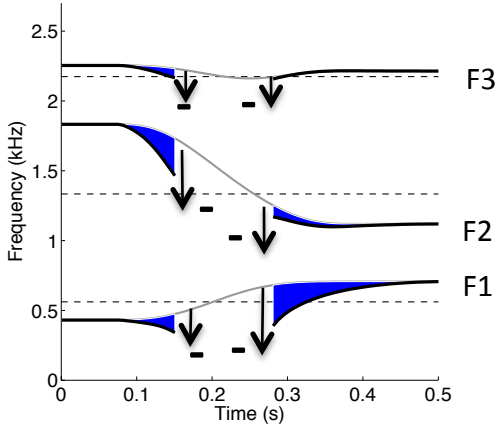




Tier 2: *Valving*

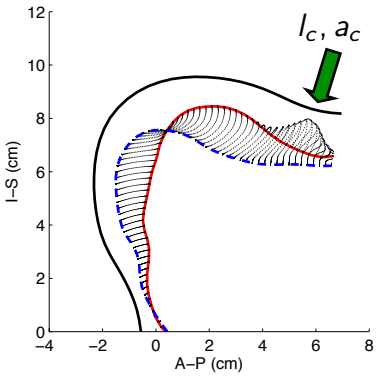
Local perturbation superimposed on underlying vocal tract shape - deflects the formants away from their path, but over a "short" time scale

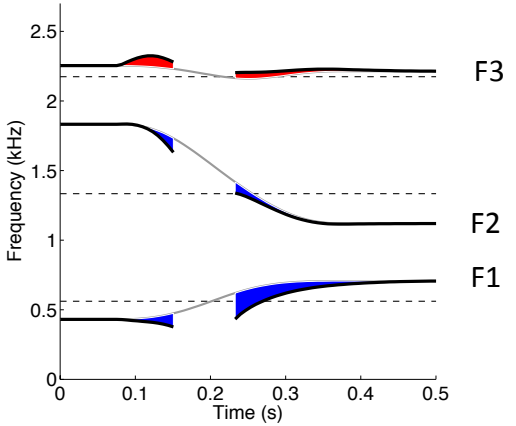




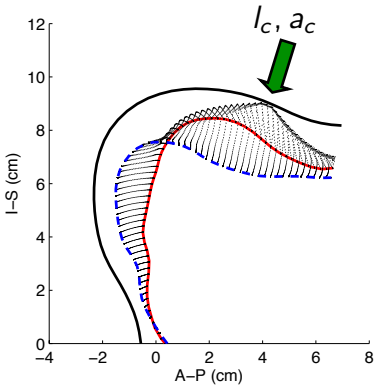
Modification of location and degree of constriction

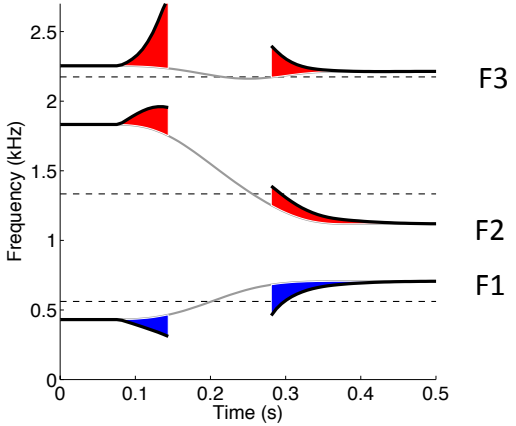
$l_c = -1.2$ cm from lips
$a_c = 0.05$ cm^2



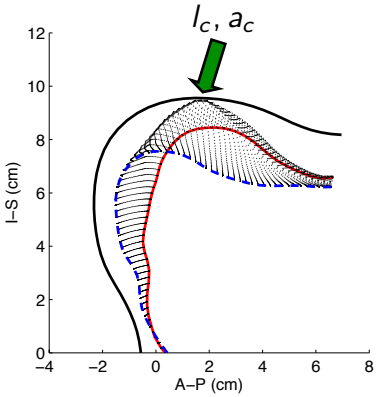


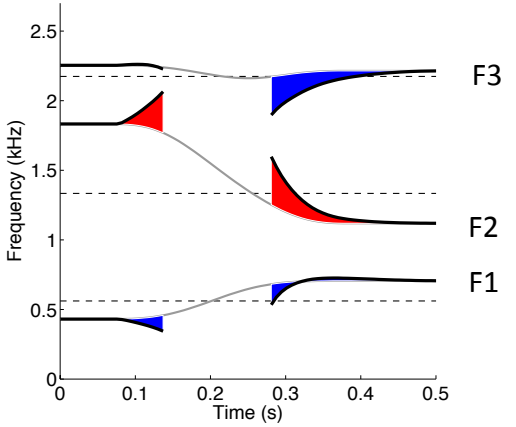
$l_c = -2.8$ cm from lips
$a_c = 0.0$ cm^2





$l_c = -5.6$ cm from lips
$a_c = 0.0$ cm^2





"The Relation of Vocal Tract Shape, Formant Transitions, and Stop Consonant Identification"

Story and Bunton, JSLHR, 53, 1514-1528 (2010)

Question: Are listeners sensitive the deflection patterns present in a VCV relative to the underlying VV?

Experiment:

- 20 VCVs representing change of constriction location were simulated with the airway modulation model.
- 3 VV contexts
- Forced choice (b,d,g) paradigm

Results

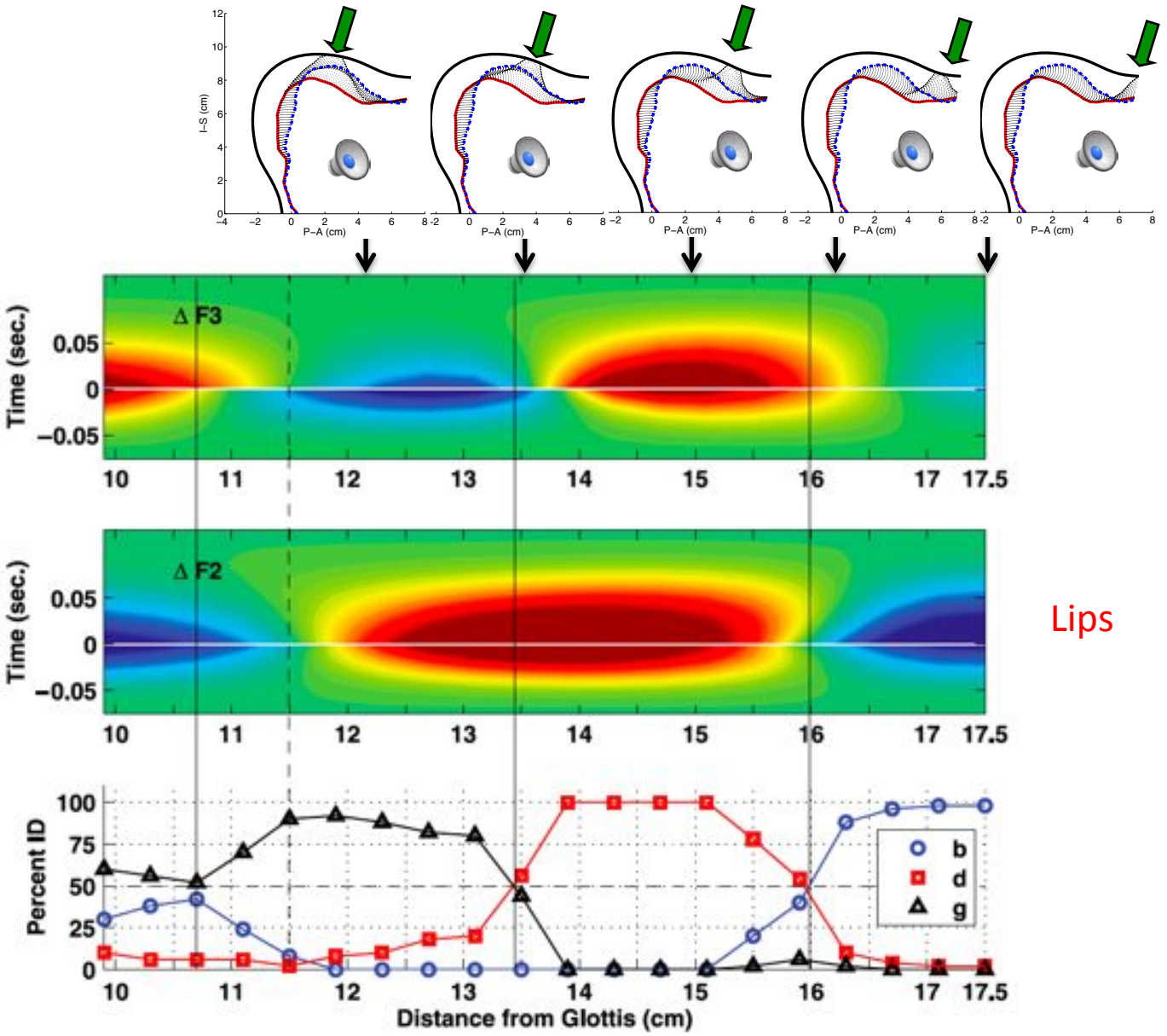
[əCi]

Formant Deflection Patterns

Velar region

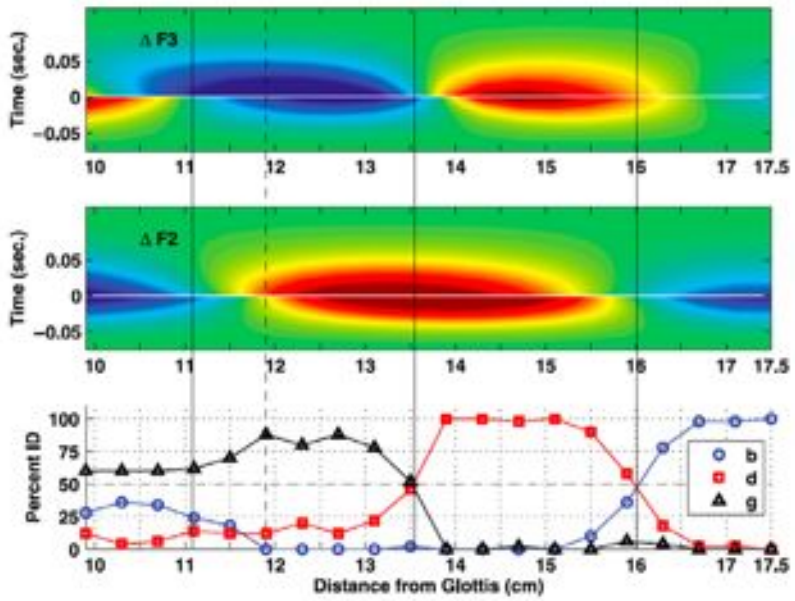
Lips

Listener identification of voiced stops

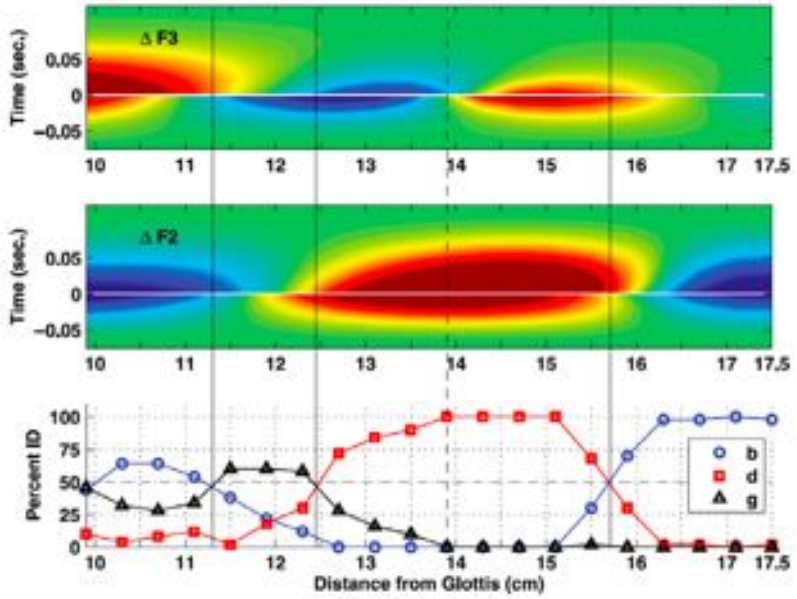


Length Dimension of Vocal Tract – Constriction Location

[əCɑ]



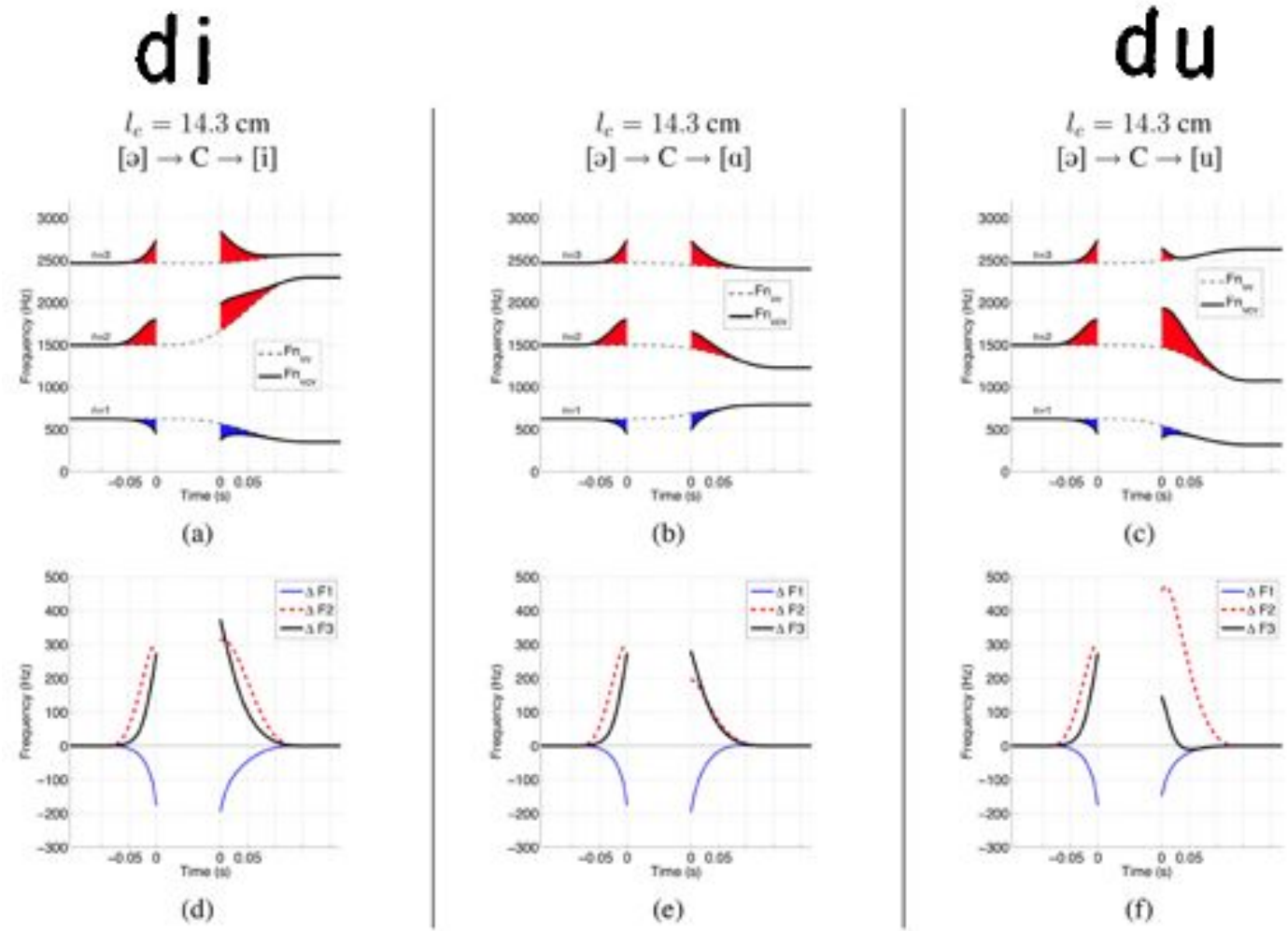
[əCu]



Brad Story, Zemlin Lecture, ASHA, 2013

Three different VV contexts – Same constriction location

Even though the absolute direction of formants may be different,
the deflection patterns all maintain the same polarity



Brad Story, Zemlin Lecture, ASHA, 2013

Structure $\rightarrow$ Movement $\rightarrow$ Sound $\rightarrow$ Perception

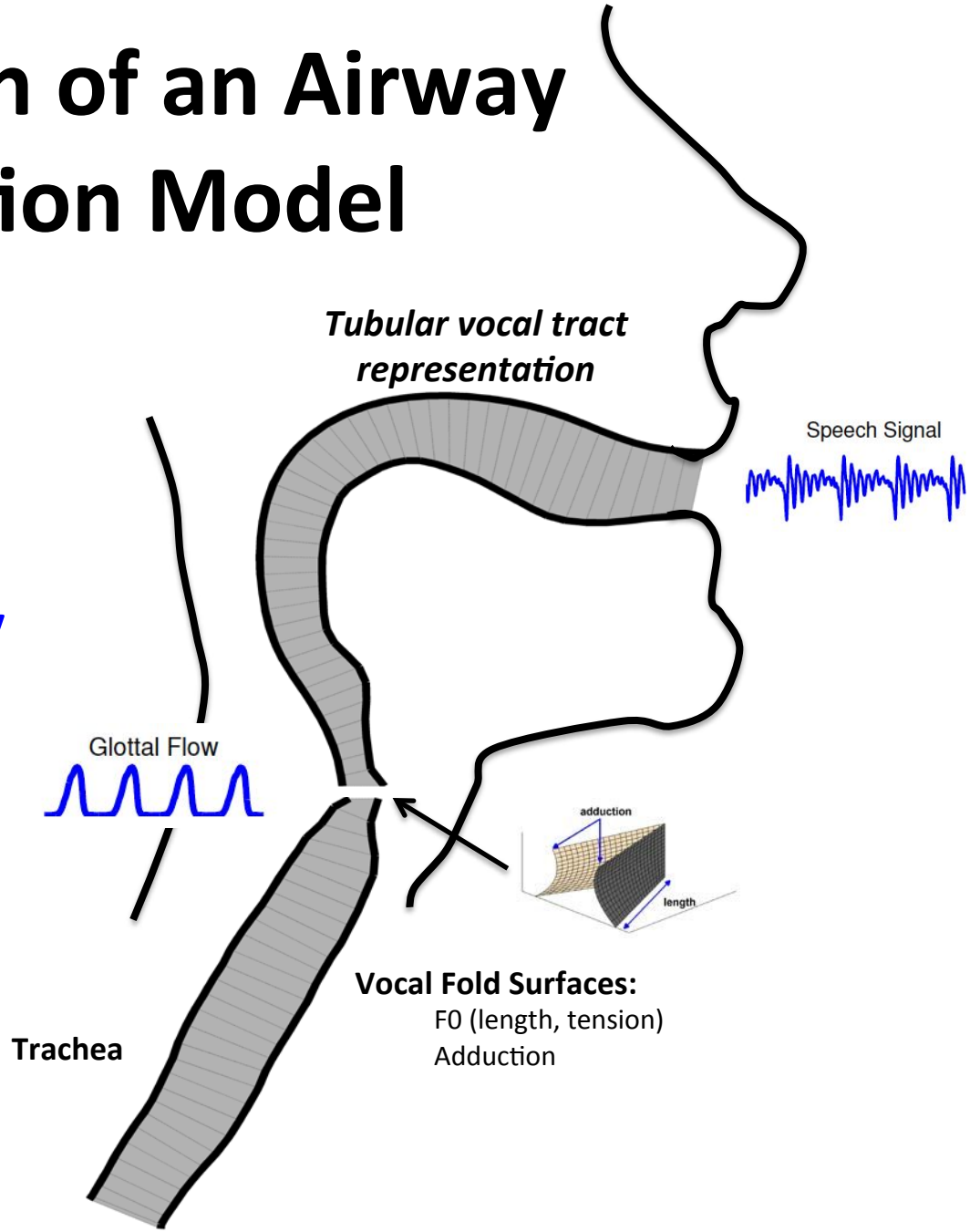
On the Origin of an Airway Modulation Model

Airway Structure

+

Modulation of the airway

(articulation & vibration)



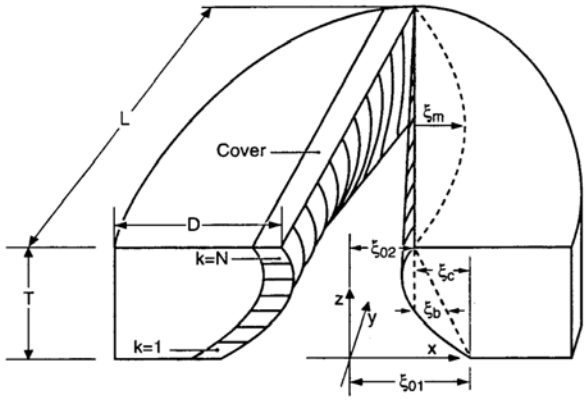
The Voice Source:
Models of Vocal Fold Vibration

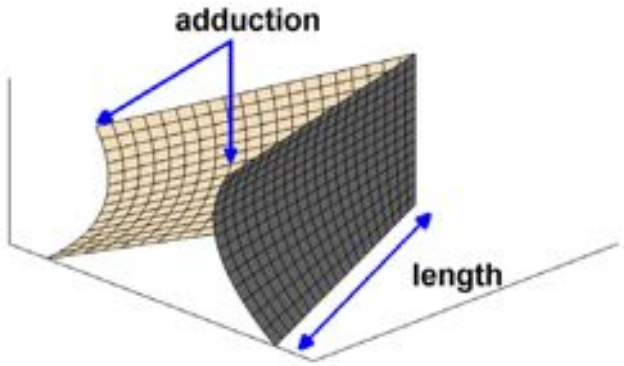
Parameterization of the glottal area, glottal flow, and vocal fold contact area

Ingo R. Titze
University of Iowa, Iowa City, Iowa 52242

J. Acoust. Soc. Am. 75 (2), February 1984

Kinematic model





Glottal width defined by two components:

$$g(y,z,t) = 2[\xi_0(y,z,t) + \xi(y,z,t)]$$

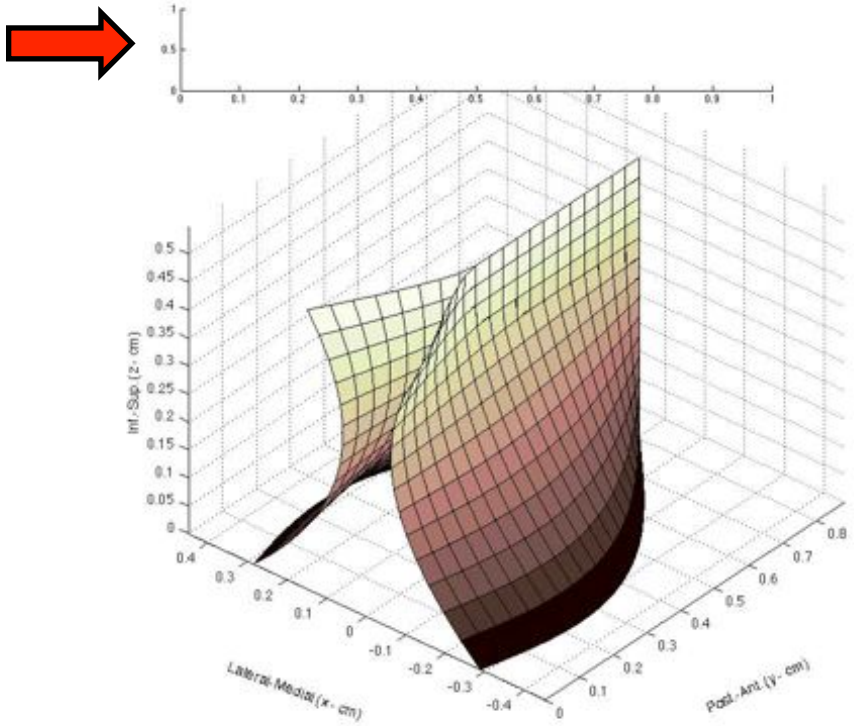
Slowly-varying postural component

Vibrational displacement

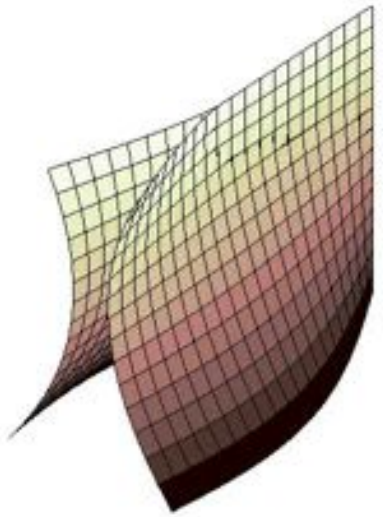
Titze, I.R. (2006). *The myoelastic aerodynamic theory of phonation*, NCVS, pp. 197-214.

Vocal fold motion is wave-like – and modulates the airspace between them

Glottal area



Vibration, abduction, and adduction



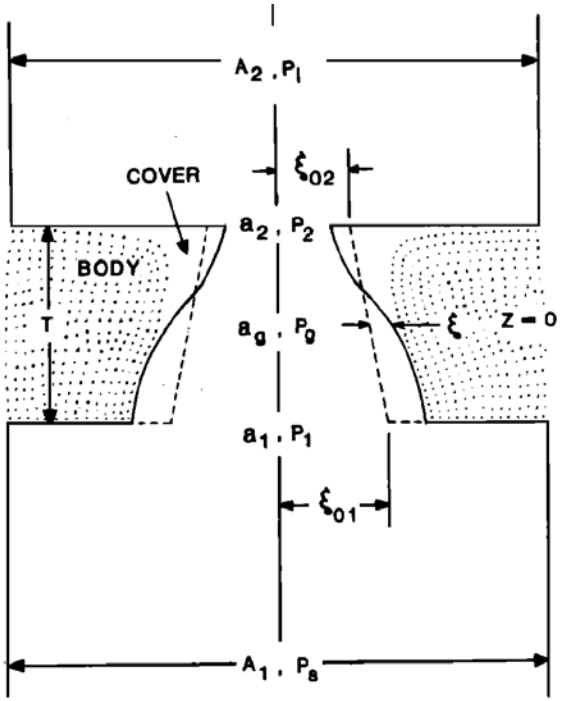
Brad Story, Zemlin Lecture, ASHA, 2013

Mechanisms of ***sustained oscillation*** of vocal fold vibration

The physics of small-amplitude oscillation of the vocal folds

Ingo R. Titze
Voice Acoustics and Biomechanics Laboratory, Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242

J. Acoust. Soc. Am. **83** (4), April 1988



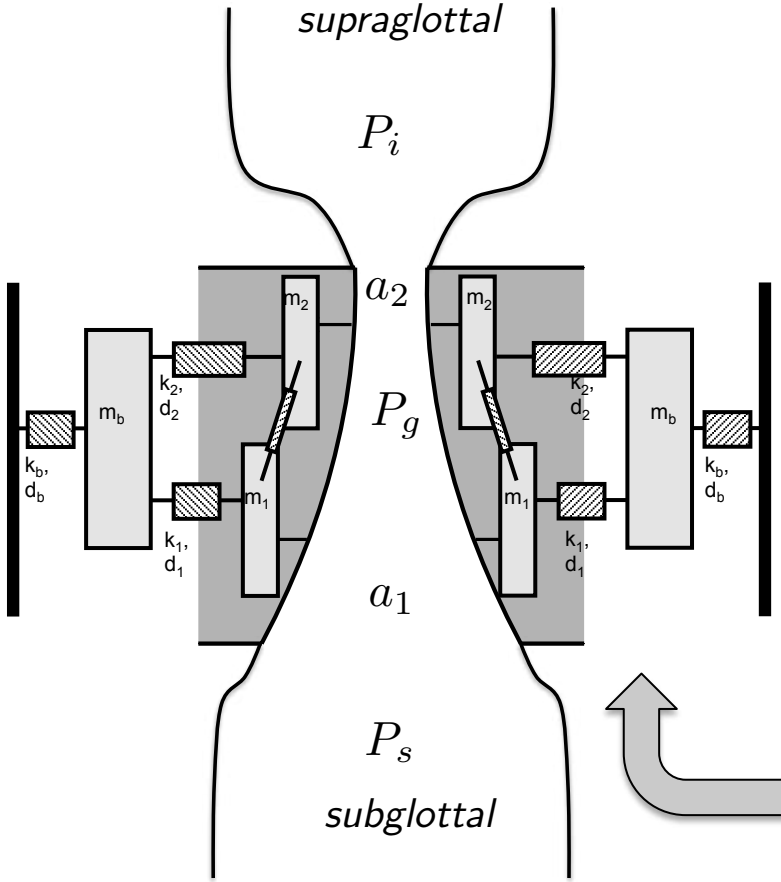
A main point…

For oscillation to be sustained, the intraglottal (driving) pressure must be *in phase* with the tissue velocity

Three-mass model of the vocal folds (1995)

Lumped-element, self-oscillating model that simulated the body-cover vocal fold structure

Story and Titze



Titze and Story (2002)

Muscle activation levels

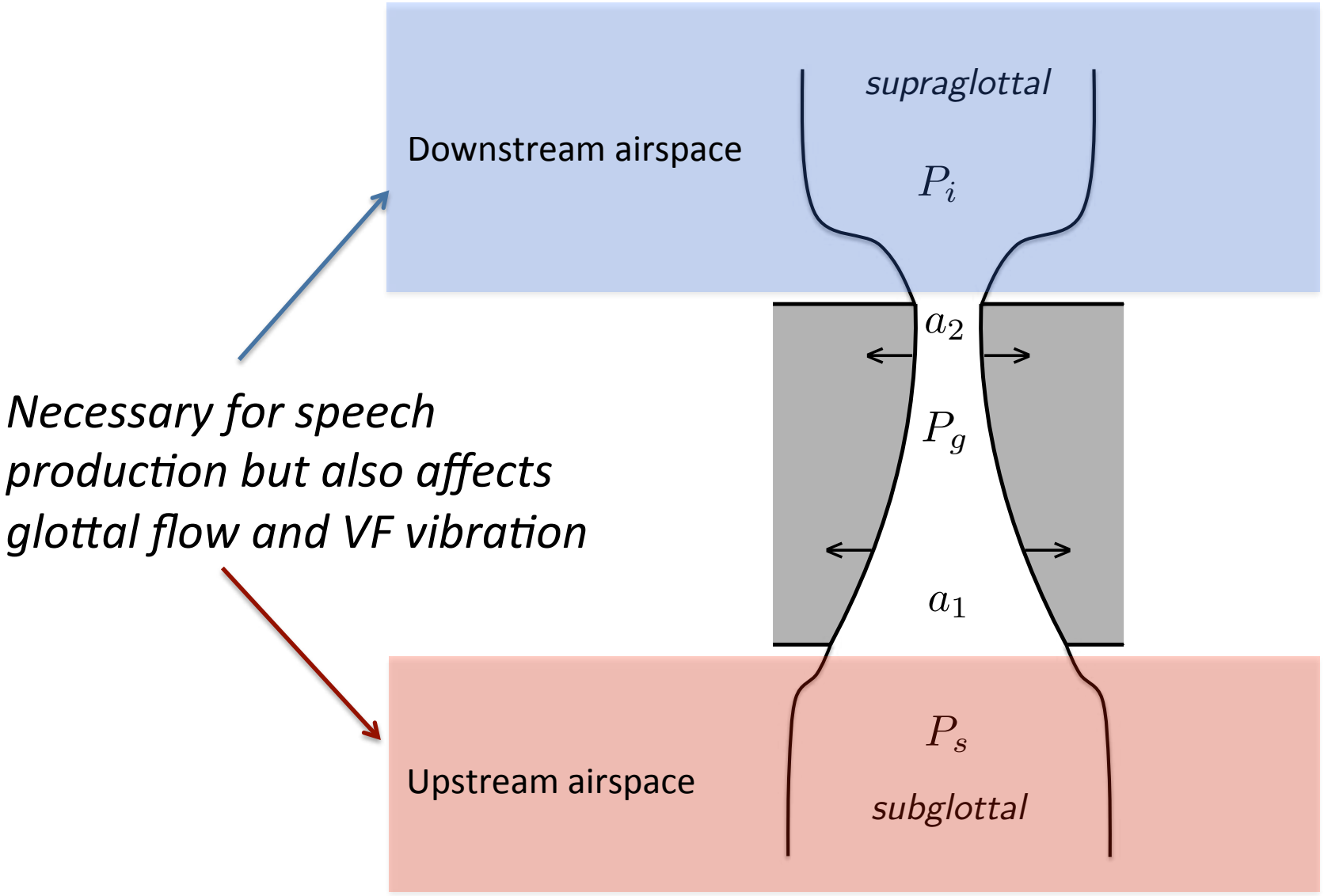
a_{CT}
a_{TA}

Transformation "rules"

m_1, m_2, m_b
k_1, k_2, k_b, k_c
L_o, T_o

Mechanical parameter values

But the larynx is coupled to upstream and downstream airways…



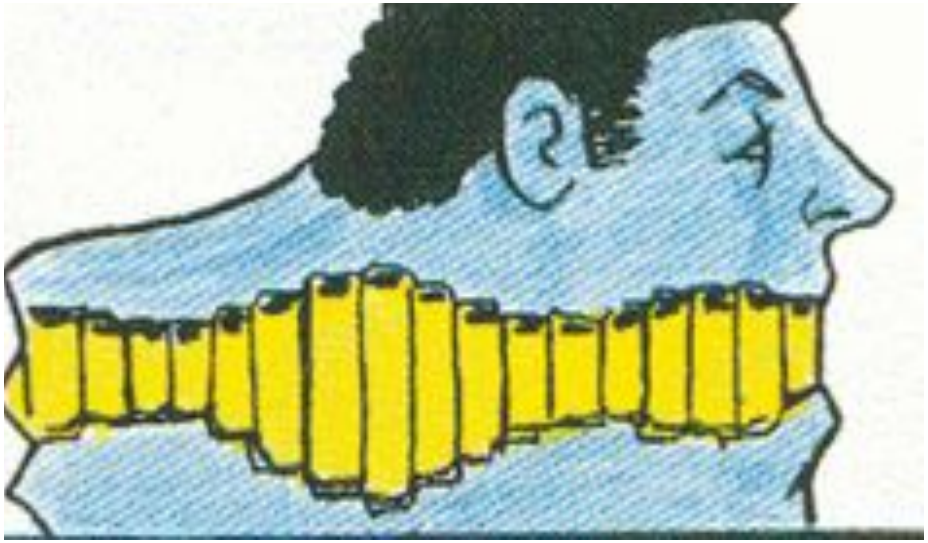
Necessary for speech production but also affects glottal flow and VF vibration

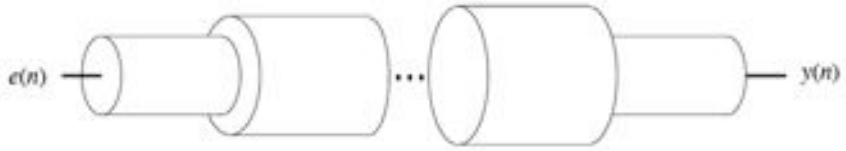
Models of Wave Propagation in the Vocal Tract

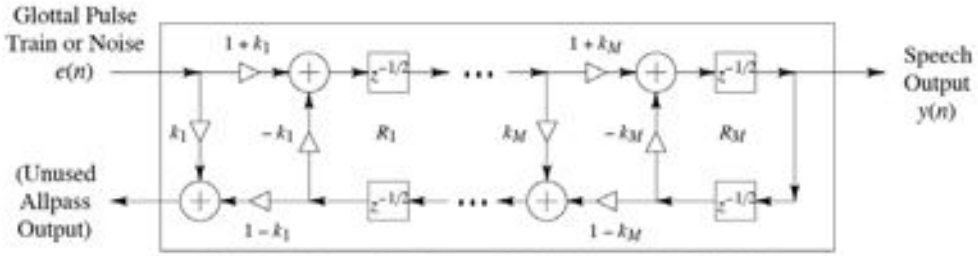
…and trachea and nasal passages

1D Waves: Wave Reflection Approach

Kelly and Lochbaum, 1962
Strube, 1982
Liljencrants, 1985
Story, 1995 (diss.)
Milenkovic





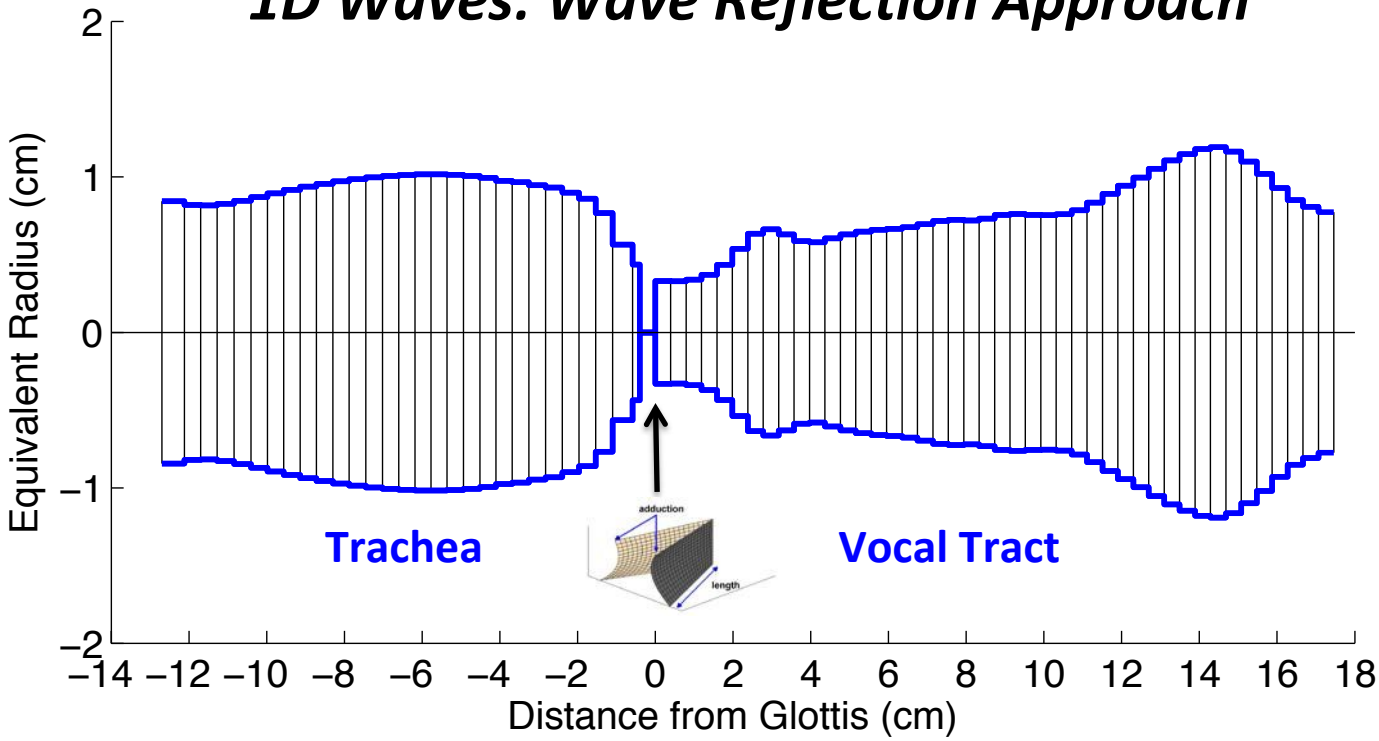


Kelly-Lochbaum Vocal Tract Model (Piecewise Cylindrical)

1D Waves: Wave Reflection Approach

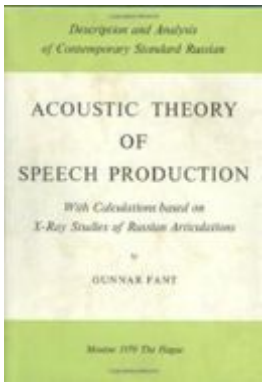
Energy Losses

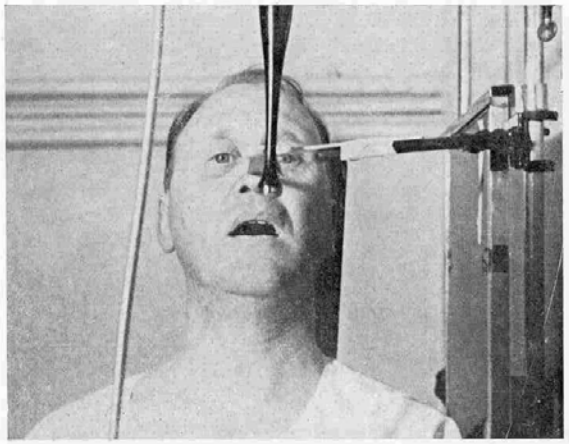
- yielding walls
- ≈viscosity
- ≈heat conduction
- ≈lip radiation
- skin radiation



This model is only useful if the ***shape of the vocal tract*** can be quantified as a variation in cross-sectional area

"Acoustic Theory of Speech Production" – Gunnar Fant (1960)

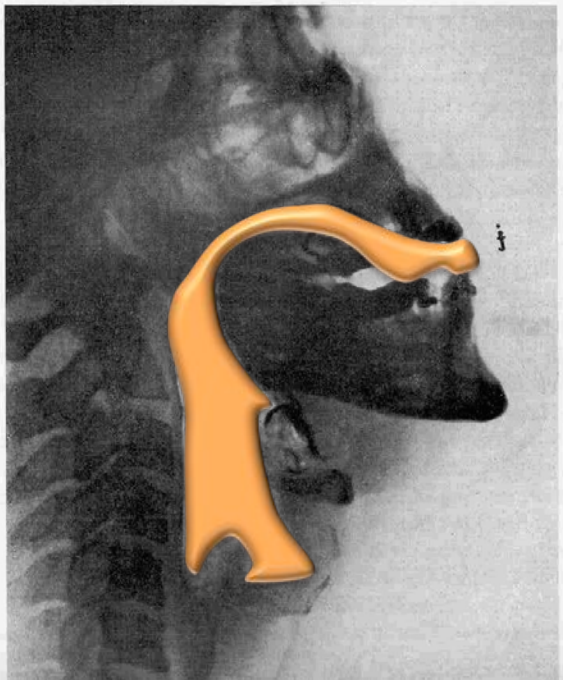




•Goal was to measure the cross-sectional area of the vocal tract from glottis to lips.

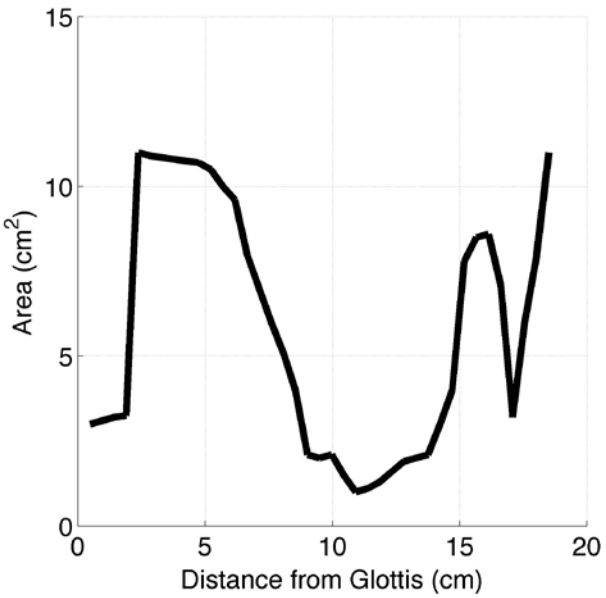
•Vocal tract shapes for Russian vowels and consonants (for one speaker).

•Shapes were obtained via X-ray projection and plaster casts of the oral cavity.

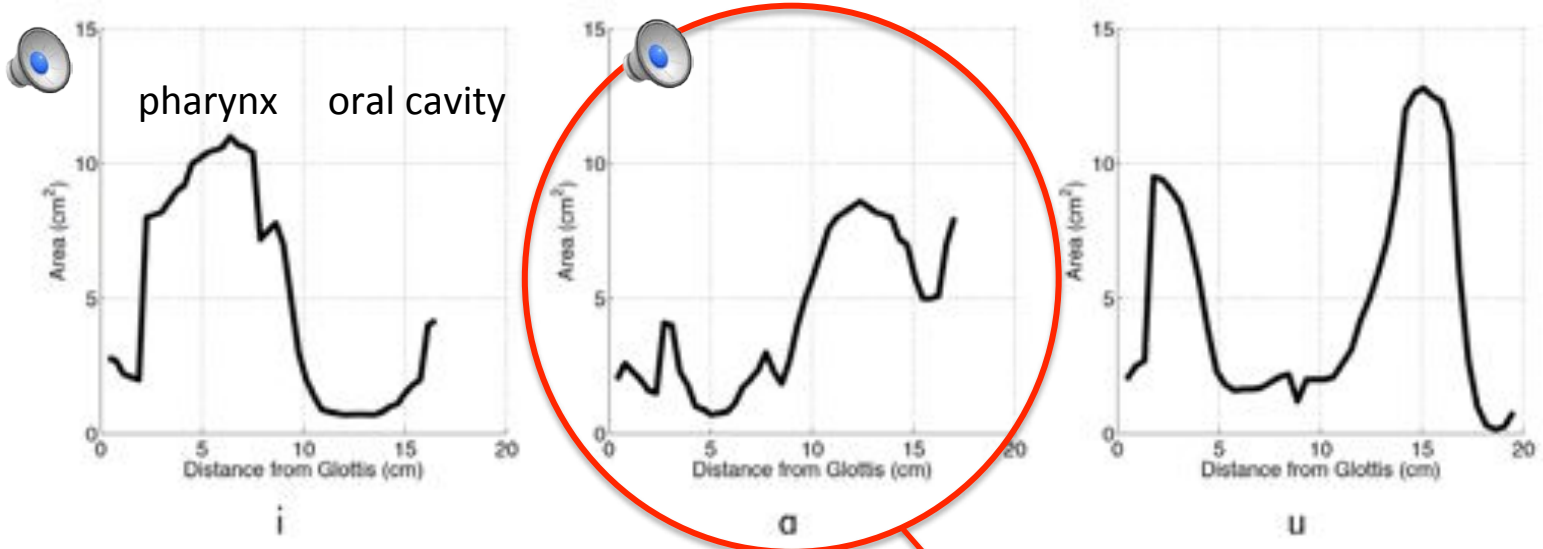


From Fant (1960), p. 95

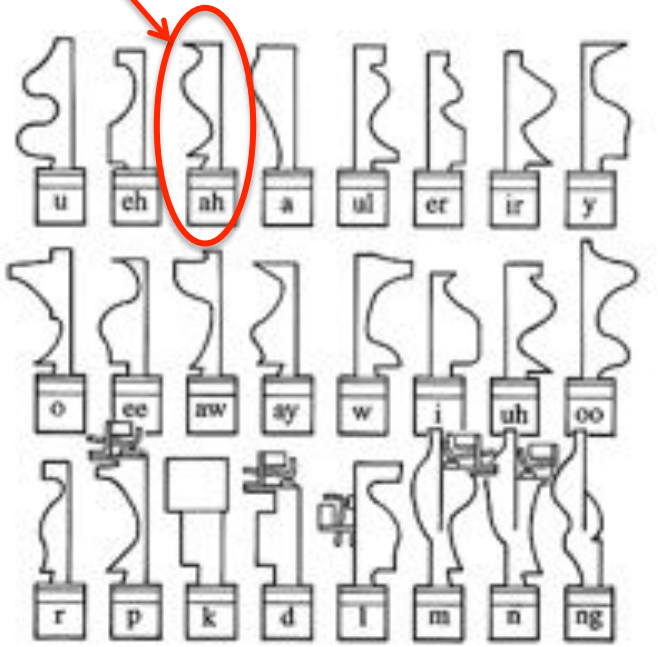
Vocal tract ***area function***



Fant's area functions for [i, ɑ, u]



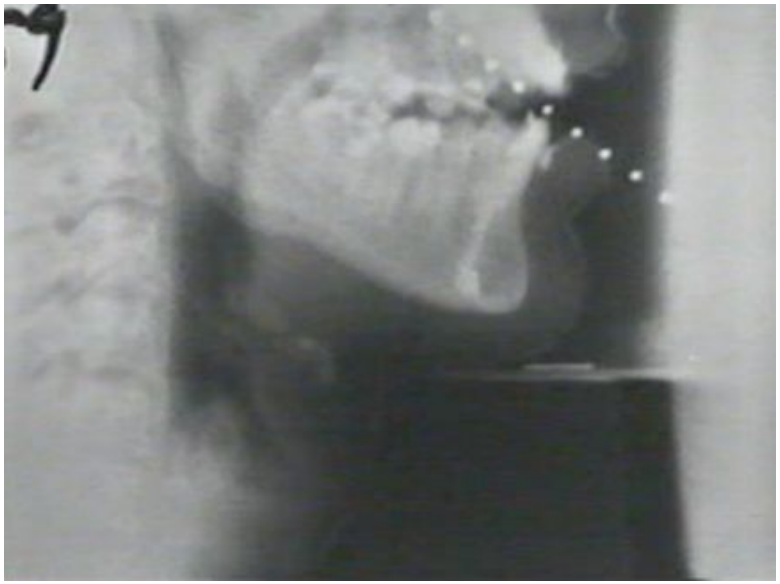
Wooden carvings of Fant (1960) area functions by Martin Riches



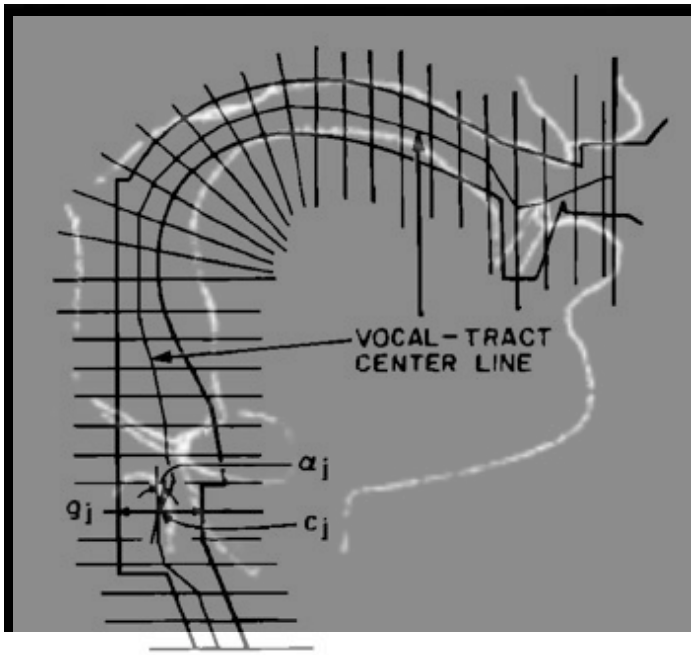
http://www.floraberlin.de/soundbag/index53.html

X-ray motion pictures (1960s)

X-ray projection images from cineradiography



Midsagittal tracings



Midsagittal to area conversion

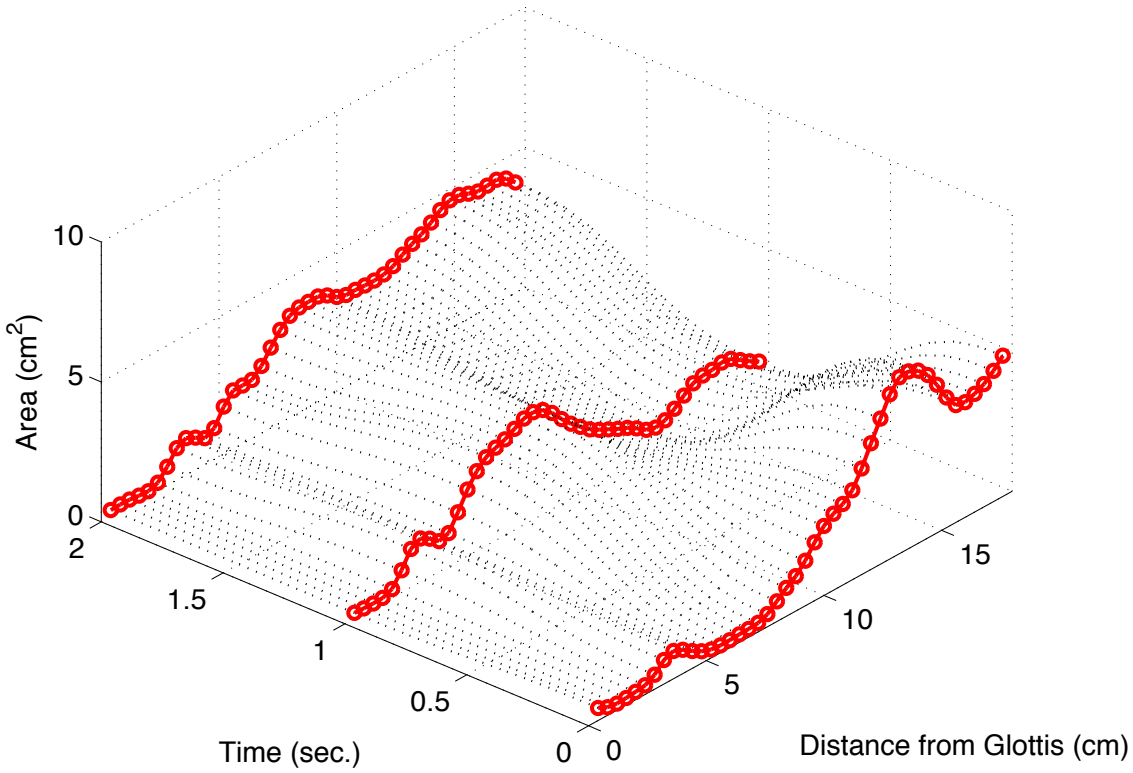
$d =$ cross-distance

$$A = kd^{\alpha}$$

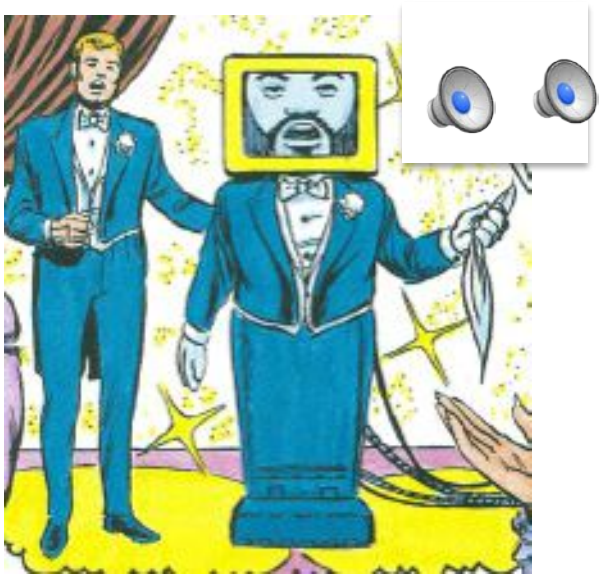
Mermelstein (1973). Articulatory model for the study of speech production, JASA

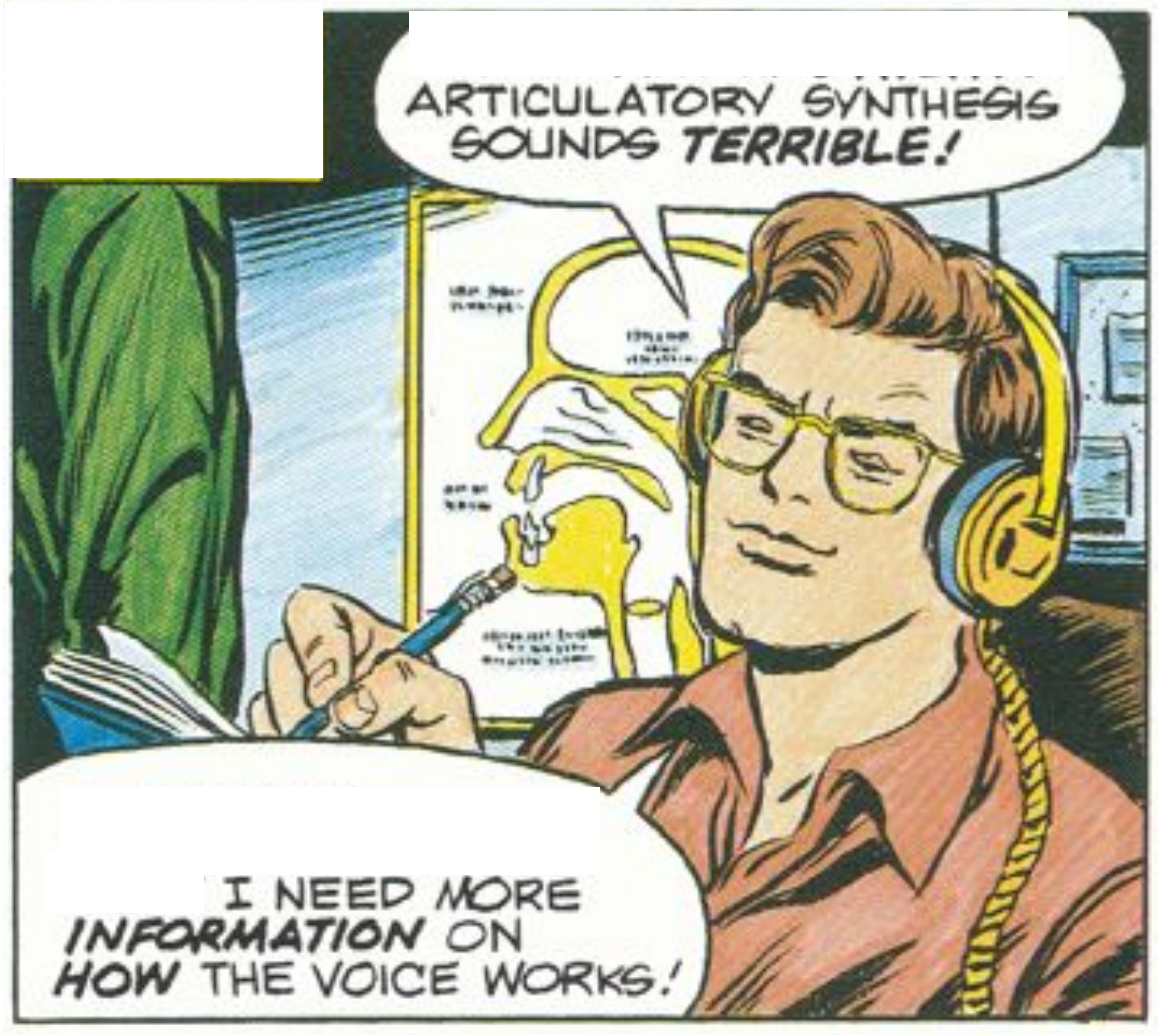
- To produce simulated/synthetic speech, each measured vocal tract shape was considered a “target”;

- Intermediate shapes were interpolated between the targets



Early speech simulation (1993):
“I enjoy the simple life as long as there’s plenty of comfort”





The Recording and Research Center, Denver Center for the Performing Arts, NCVS (1992)

Vocal tract imaging in the 1990's

*MRI and X-ray CT + **image processing***

MR Imaging of the Vocal Tract during Vowel Production[1]

A. V. Lakshminarayanan, PhD • Sungbok Lee, MS
Martin J. McCutcheon, PhD

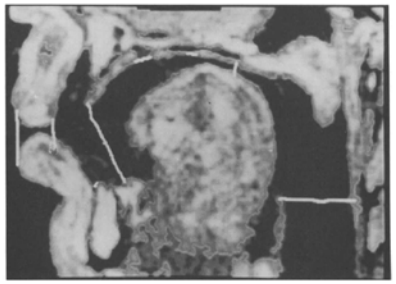
JMRI 1991; 1:71–76

J. Magnetic Resonance Imaging (1991)

The Correspondence of Vocal Tract Resonance With Volumes Obtained From Magnetic Resonance Images

Christopher A. Moore
University of Pittsburgh
Pittsburgh, PA

Journal of Speech and Hearing Research, Volume 35, 1009–1023, October 1992



Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels

T. Baer
Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, United Kingdom

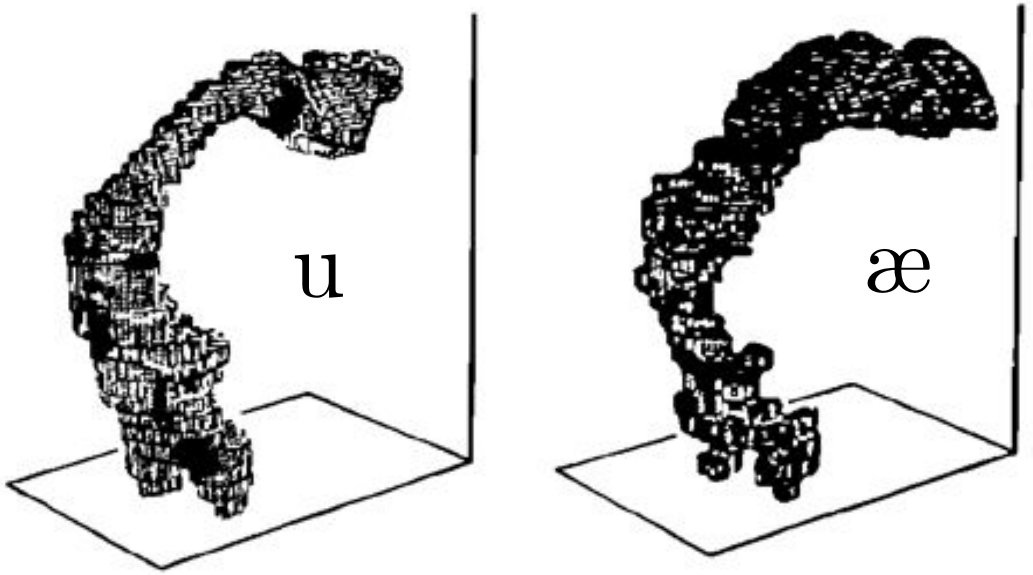
J. C. Gore
Department of Diagnostic Imaging, Yale School of Medicine, New Haven, Connecticut 06501

L. C. Gracco and P. W. Nye
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

J. Acoust. Soc. Am. **90** (2), Pt. 1, August 1991

2 talkers
4 vowels each

3D reconstructions of the vocal tract!

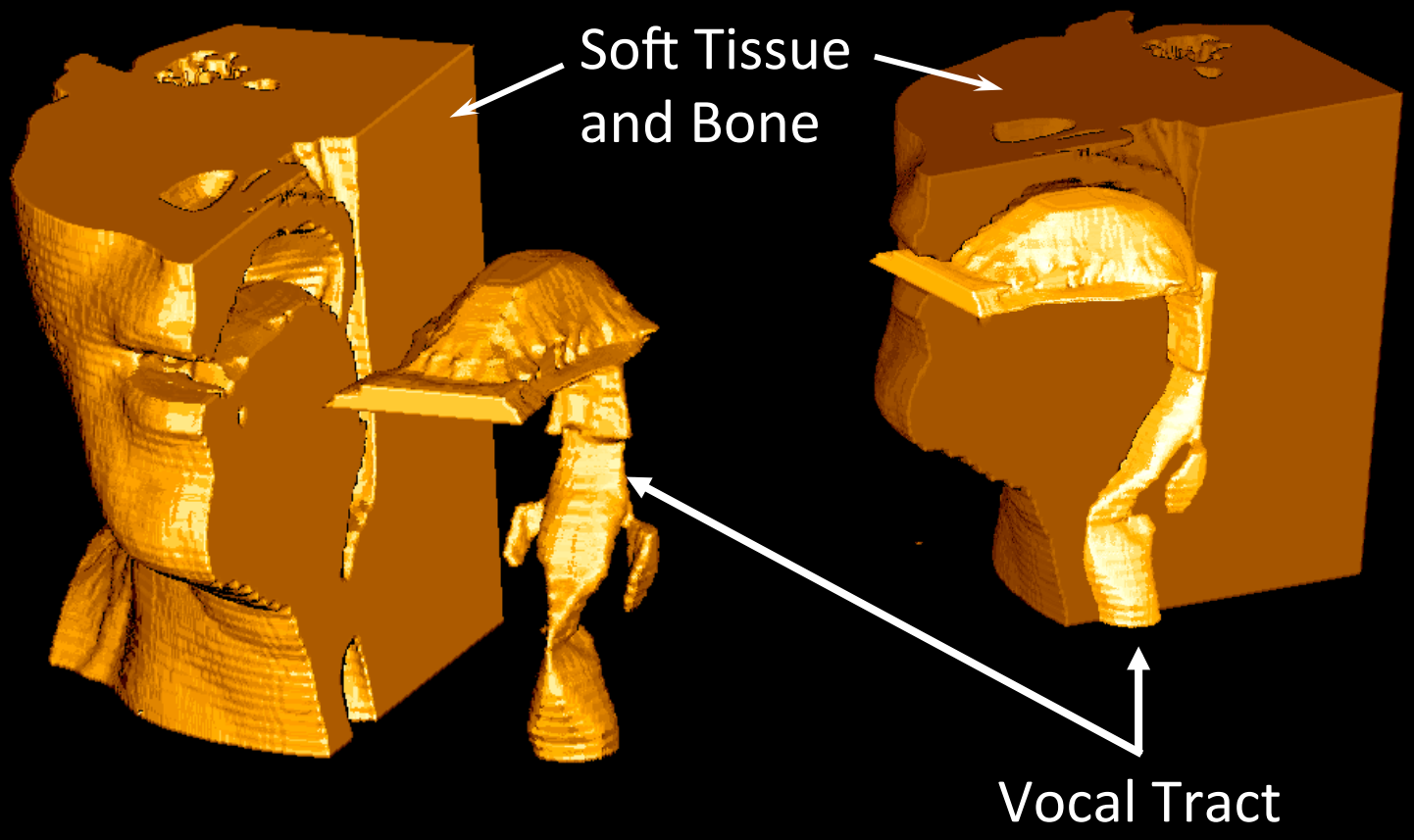


Brad Story, Zemlin Lecture, ASHA, 2013

Building a vocal tract model:

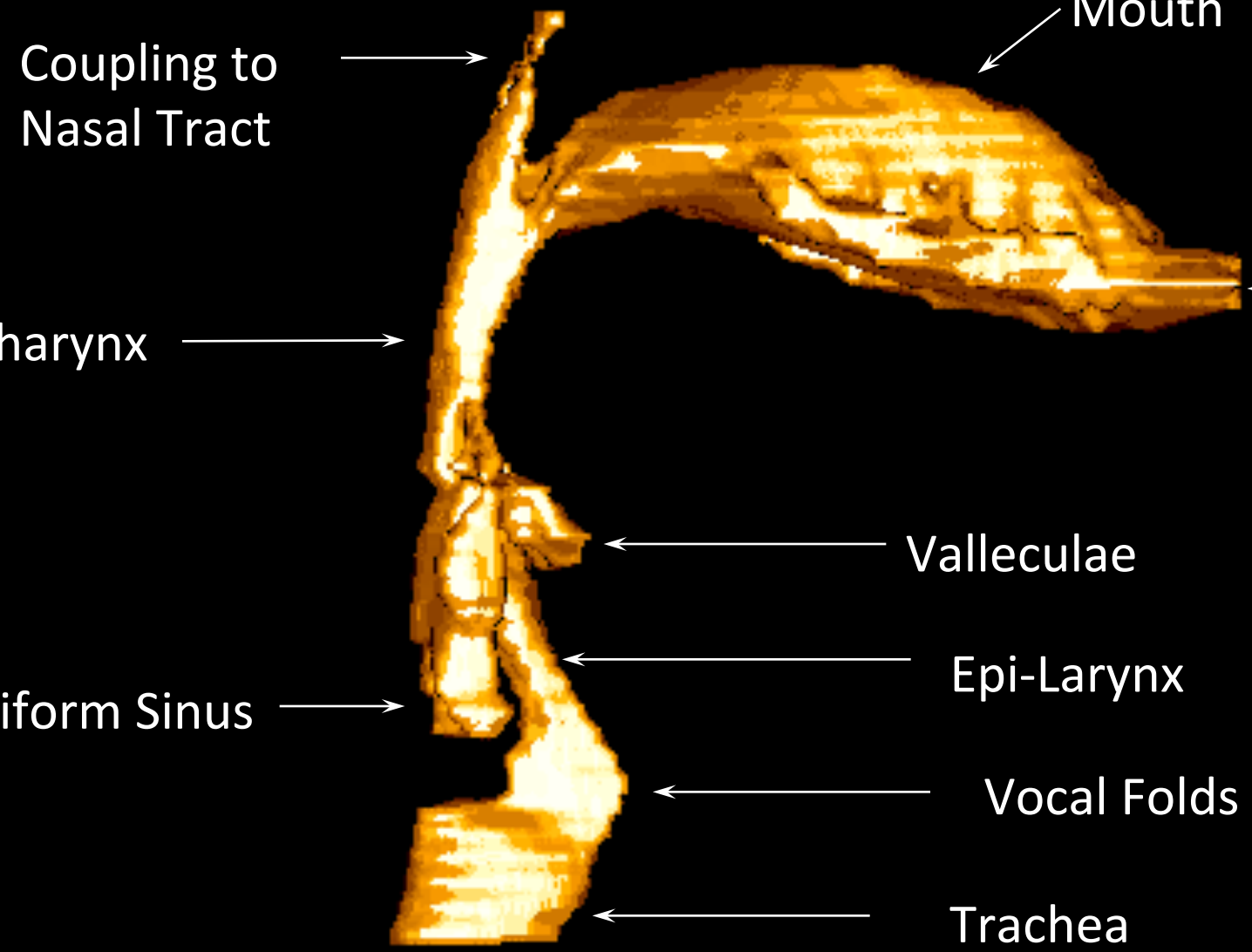
Reconstruction of the vocal tract

Imaging: MRI, X-ray CT, etc



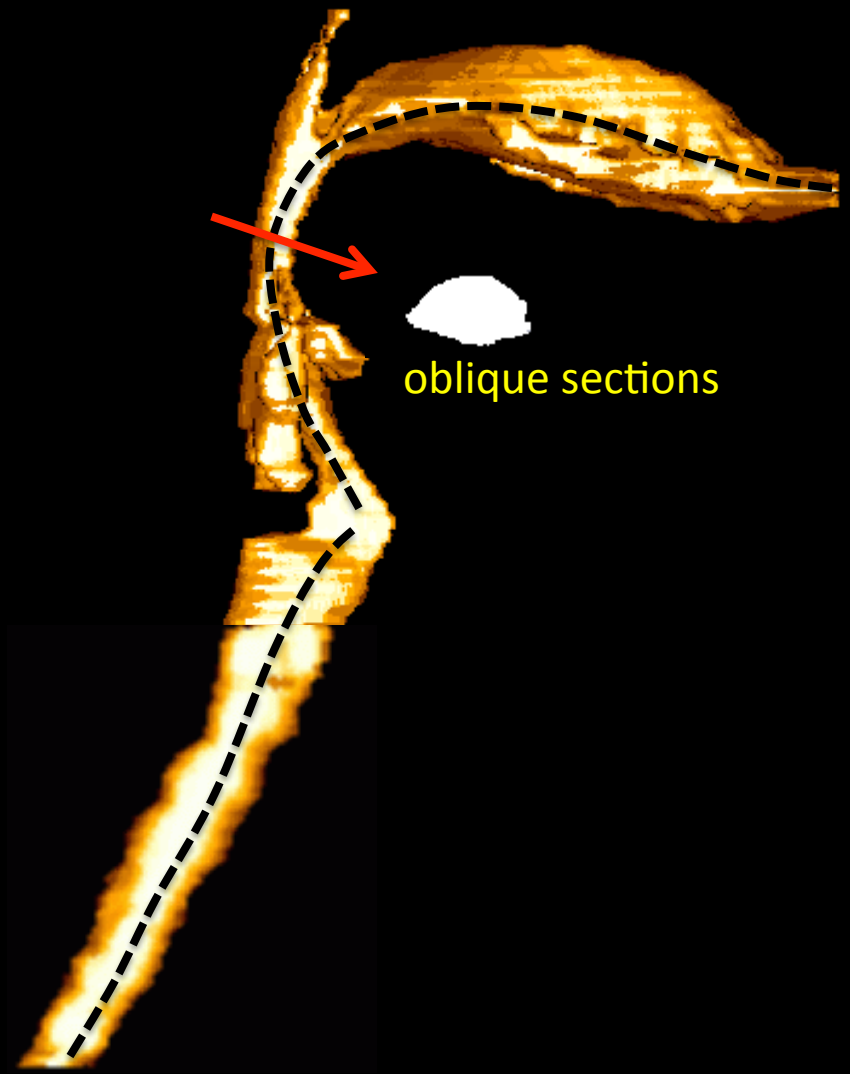
CT images used for reconstructions shown

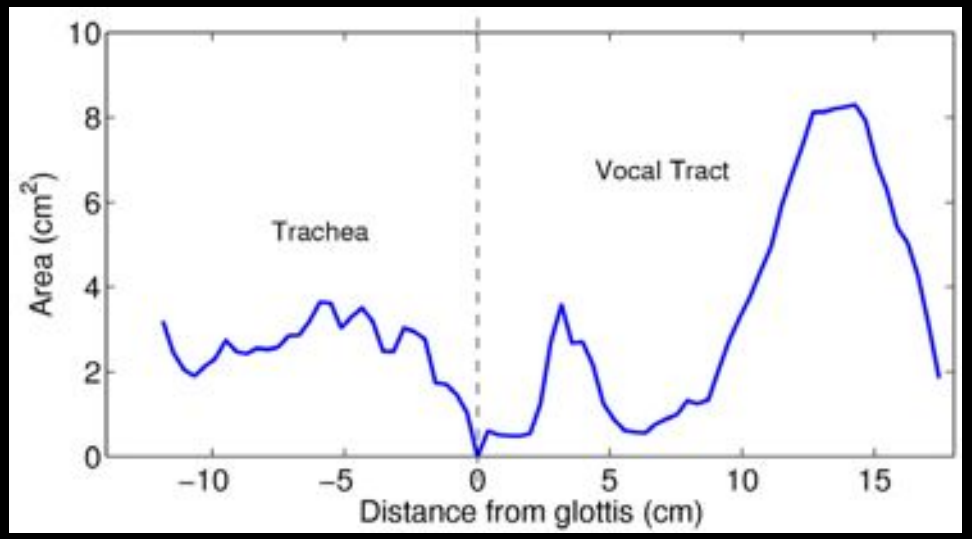
CT: Vowel [a] – *male*



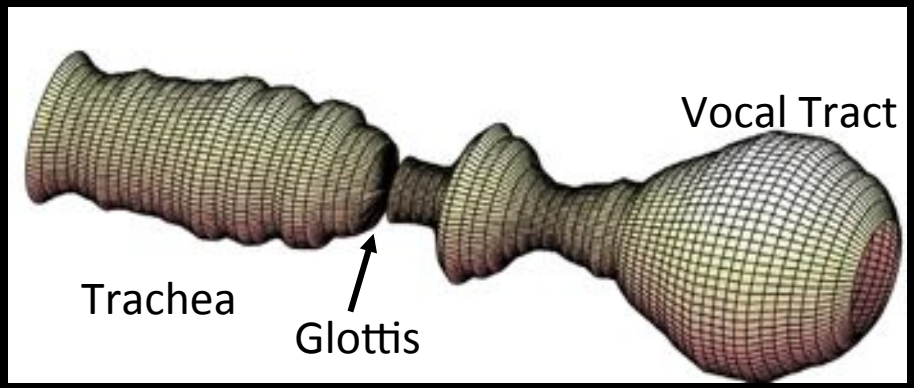
Measurement of the "Area Function"

3-D shape → Area Function





Equivalent Tubular Representation



Vocal tract area functions from magnetic resonance imaging

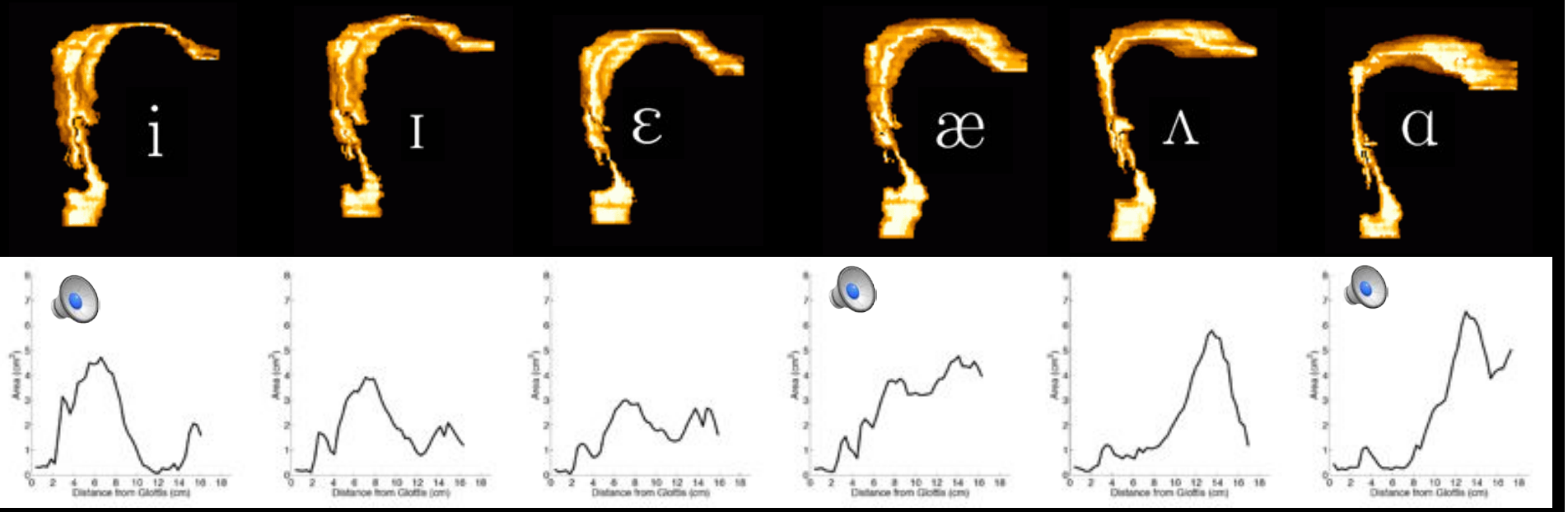
Brad H. Story and Ingo R. Titze
Department of Speech Pathology and Audiology, National Center for Voice and Speech, University of Iowa, Iowa City, Iowa 52242

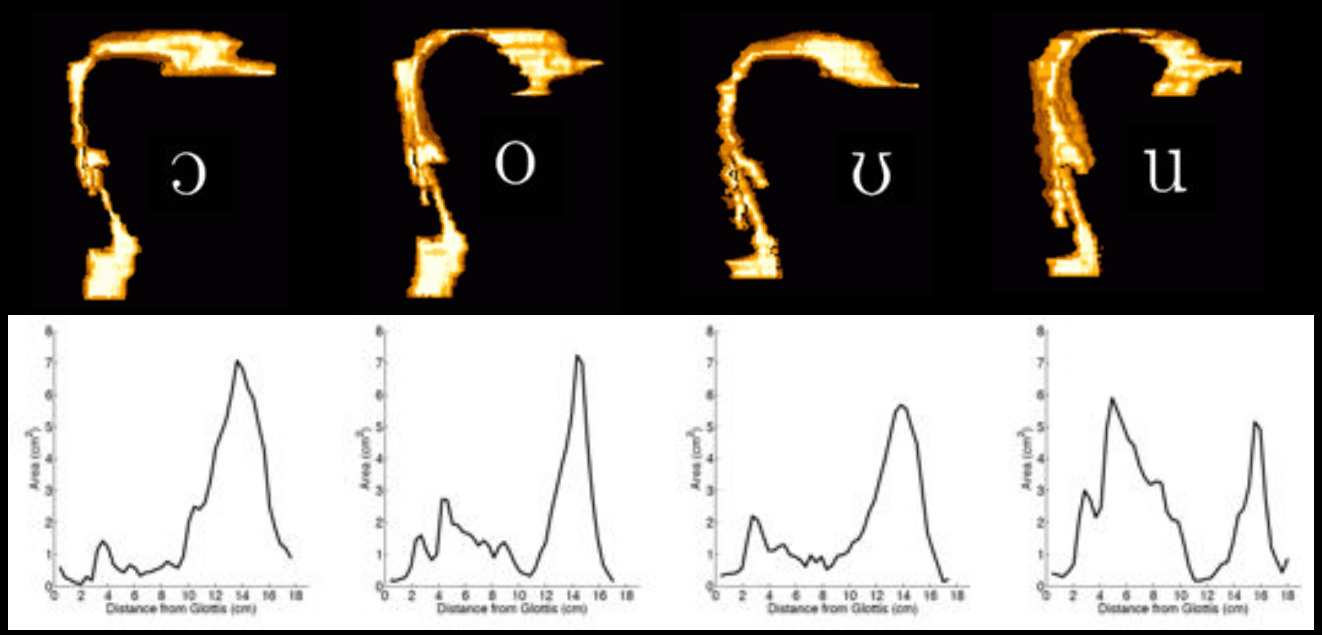
Eric A. Hoffman
Division of Physiologic Imaging, Department of Radiology, University of Iowa College Medicine, Iowa City, Iowa 52242

J. Acoust. Soc. Am. **100** (1), July 1996

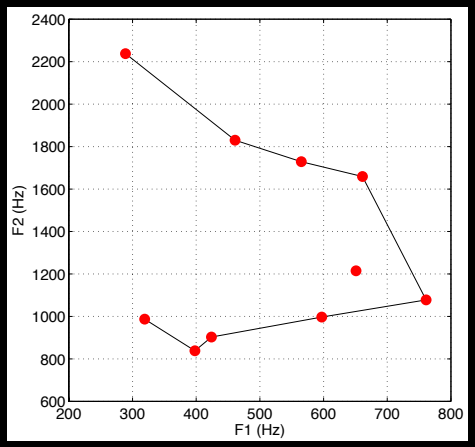
- **One male talker:**
  - 18 vocal tract shapes – ***"static"*** (no movement)
  - Each image set = 26 slices (24 cm FOV), 5 mm thick
  - Required about ***10-15 minutes*** to acquire one image set (i.e., 1 vowel or consonant)
  - MR environment requires ***high intensity (loud) production*** of vowels

3D reconstructions and area functions for 10 vowels





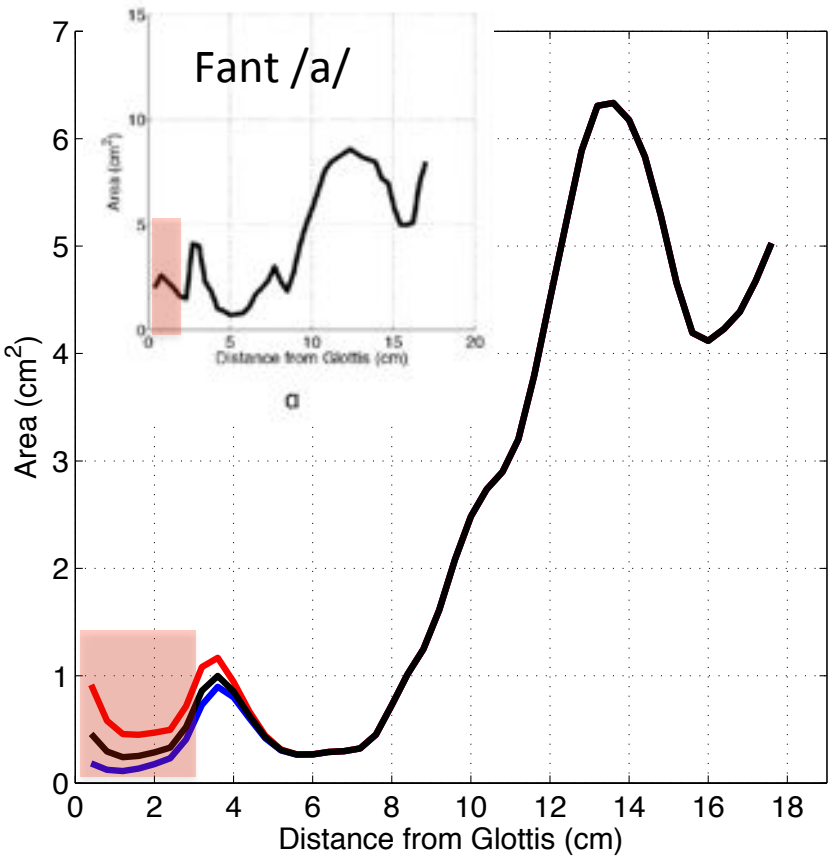
[F1,F2] vowel space

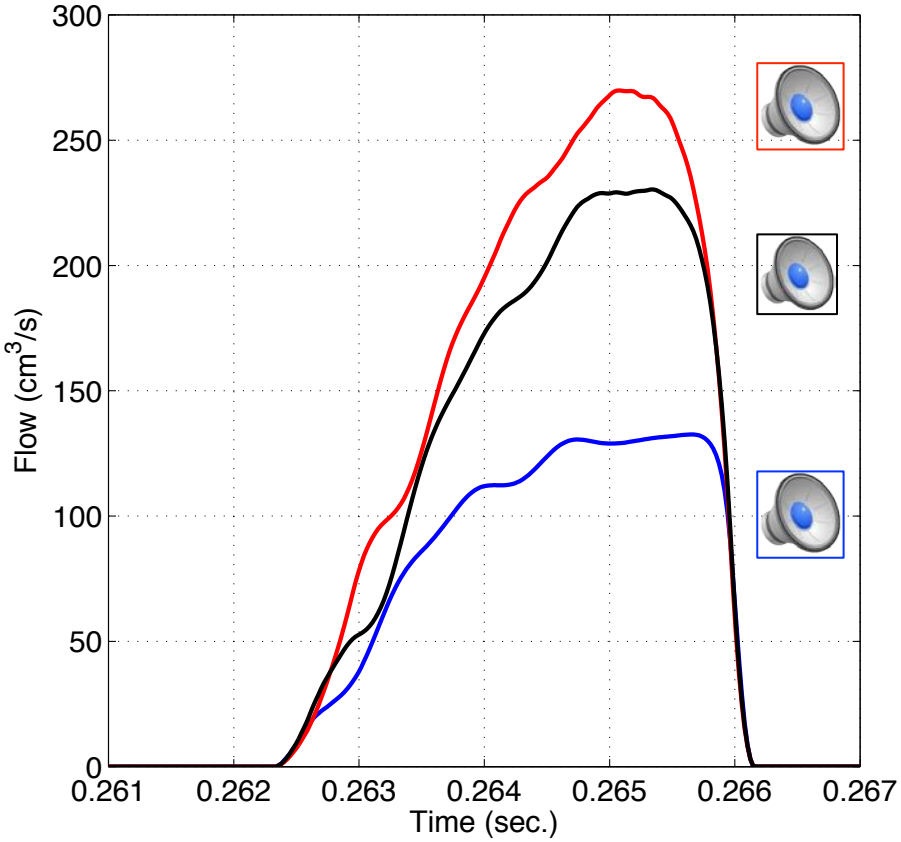


Brad Story, Zemlin Lecture, ASHA, 2013

Cross-sectional area of the epilaryngeal tube

Smaller than in the Fant (1960) area function set





Modification of the epilaryngeal configuration can affect vocal fold vibration, glottal flow pulse shape, and vocal tract resonance frequencies … Voice Quality

Brad Story, Zemlin Lecture, ASHA, 2013

An articulatory study of fricative consonants using magnetic resonance imaging

Shrikanth S. Narayanan and Abeer A. Alwan[a)]
Speech Processing and Auditory Perception Laboratory, Department of Electrical Engineering, UCLA, 405 Hilgard Avenue, Los Angeles, California 90024

Katherine Haker[b)]
Imaging Medical Group, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Los Angeles, California 90048

1995

(Received 6 September 1994; accepted for publication 3 April 1995)

Vocal tract area functions from magnetic resonance imaging

Brad H. Story and Ingo R. Titze
Department of Speech Pathology and Audiology, National Center for Voice and Speech, University of Iowa, Iowa City, Iowa 52242

Eric A. Hoffman
Division of Physiologic Imaging, Department of Radiology, University of Iowa College Medicine, Iowa City, Iowa 52242

1996

(Received 30 October 1995; accepted for publication 20 February 1996)

Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics

Abeer Alwan[a)] and Shrikanth Narayanan[b)]
Speech Processing and Auditory Perception Laboratory, Department of Electrical Engineering, School of Engineering and Applied Sciences, UCLA, 405 Hilgard Avenue, Los Angeles, California 90095

Katherine Haker
Imaging Medical Group, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Los Angeles, California 90048

1997

Vocal tract area functions for an adult female speaker based on volumetric imaging

1998

Brad H. Story
National Center for Voice and Speech, WJ Gould Voice Research Center, Denver Center for the Performing Arts, Denver, Colorado 80209

Ingo R. Titze
National Center for Voice and Speech, WJ Gould Voice Research Center, Denver Center for the Performing Arts, Denver, Colorado 80209 and Department of Speech Pathology and Audiology, University of Iowa, Iowa City, Iowa 52242

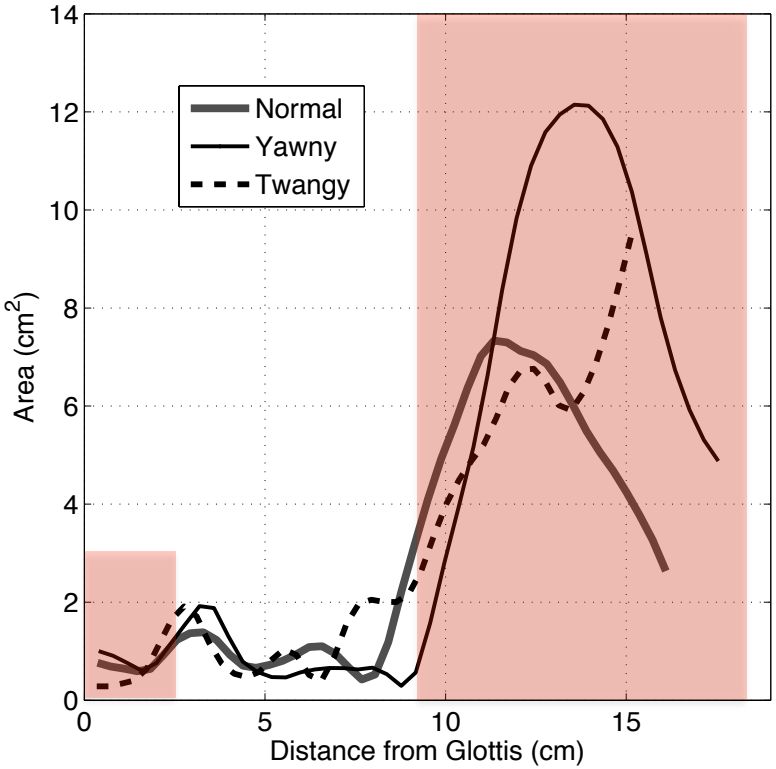
Eric A. Hoffman
Division of Physiologic Imaging, Department of Radiology, University of Iowa College of Medicine, Iowa City, Iowa 52242

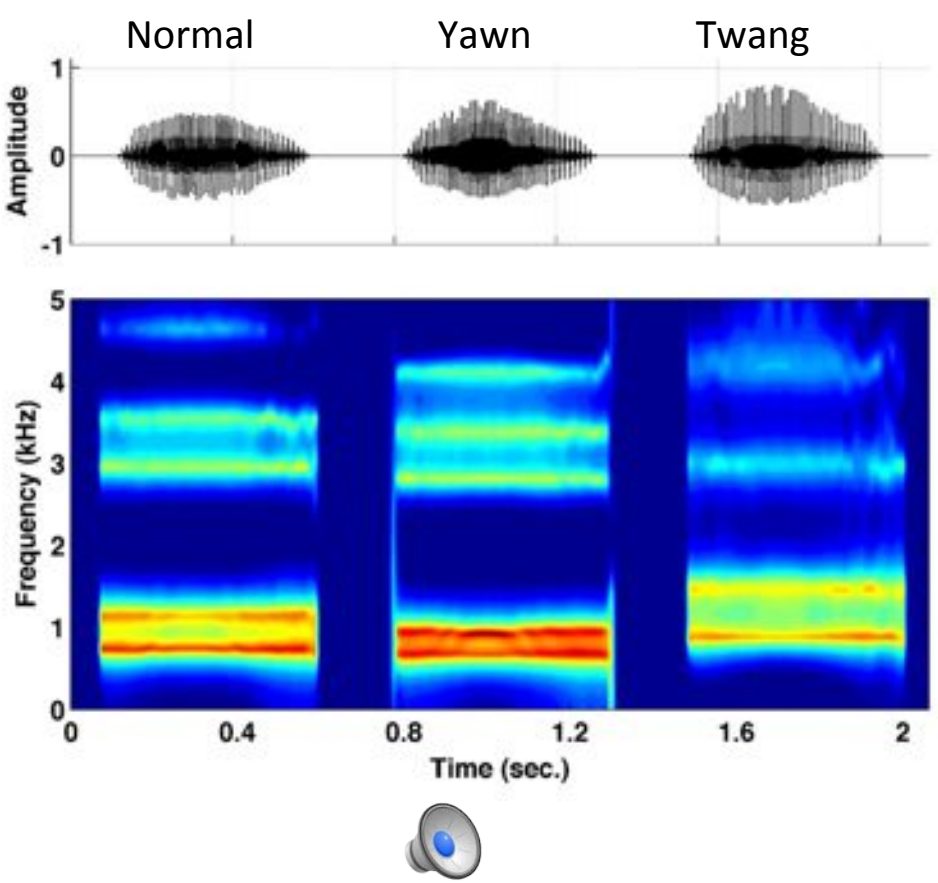
(Received 25 September 1997; accepted for publication 7 April 1998)

The relation of vocal tract shape to three voice qualities (2001)

J. Acoust. Soc. Am. , Story, Titze, and Hoffman

One male talker: /a/ vowel produced in three different voice qualities

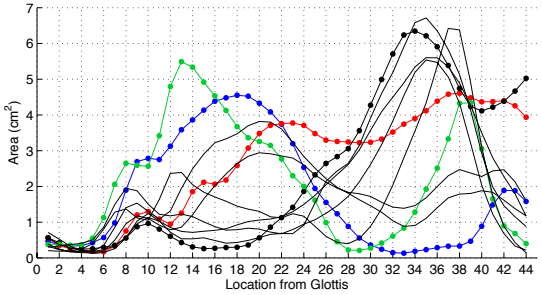




Vocal Tract

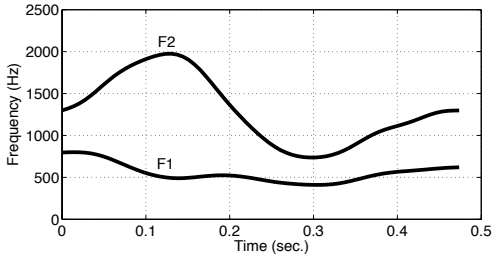
Structure → Movement

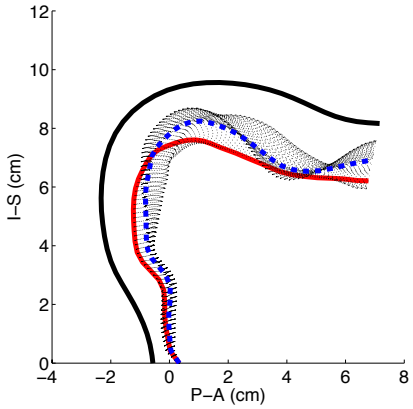
1. Statistical Analysis of Vocal Tract Area Functions



↔

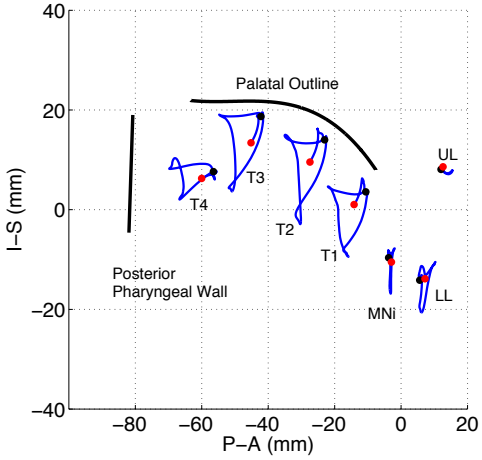
2. Mapping of Formants to Vocal Tract Shape





↔

3. Analysis of Articulatory Kinematic Data



John Westbury
Gary Weismer
U. Wisc-Madison – XRMB

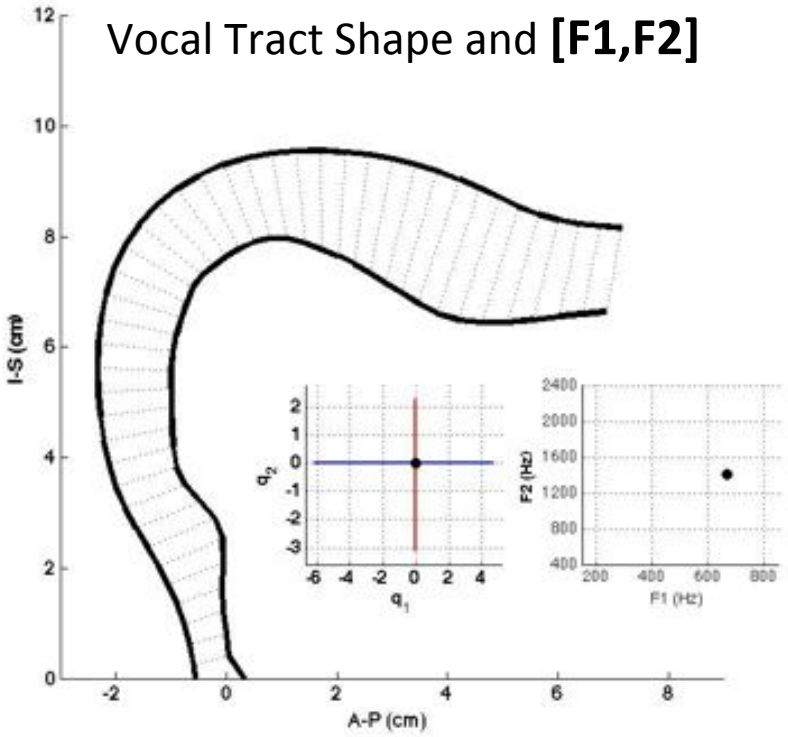
1. Principal Component Analysis (PCA) of Area Functions

J. Phonetics (1998). Story and Titze

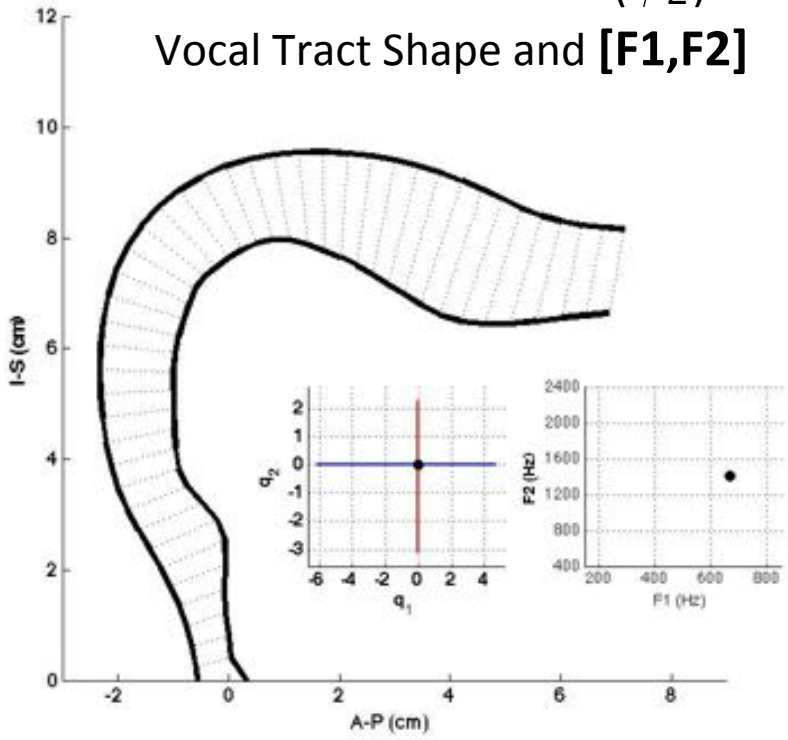
$$V(x) = \frac{\pi}{4}[\Omega(x) + q_1\phi_1(x) + q_2\phi_2(x)]^2$$

Mean VT shape ***Modes (components)***

Effect of the **1st** Mode (ϕ_1) on Vocal Tract Shape and **[F1,F2]**



Effect of the **2nd** Mode (ϕ_2) on Vocal Tract Shape and **[F1,F2]**



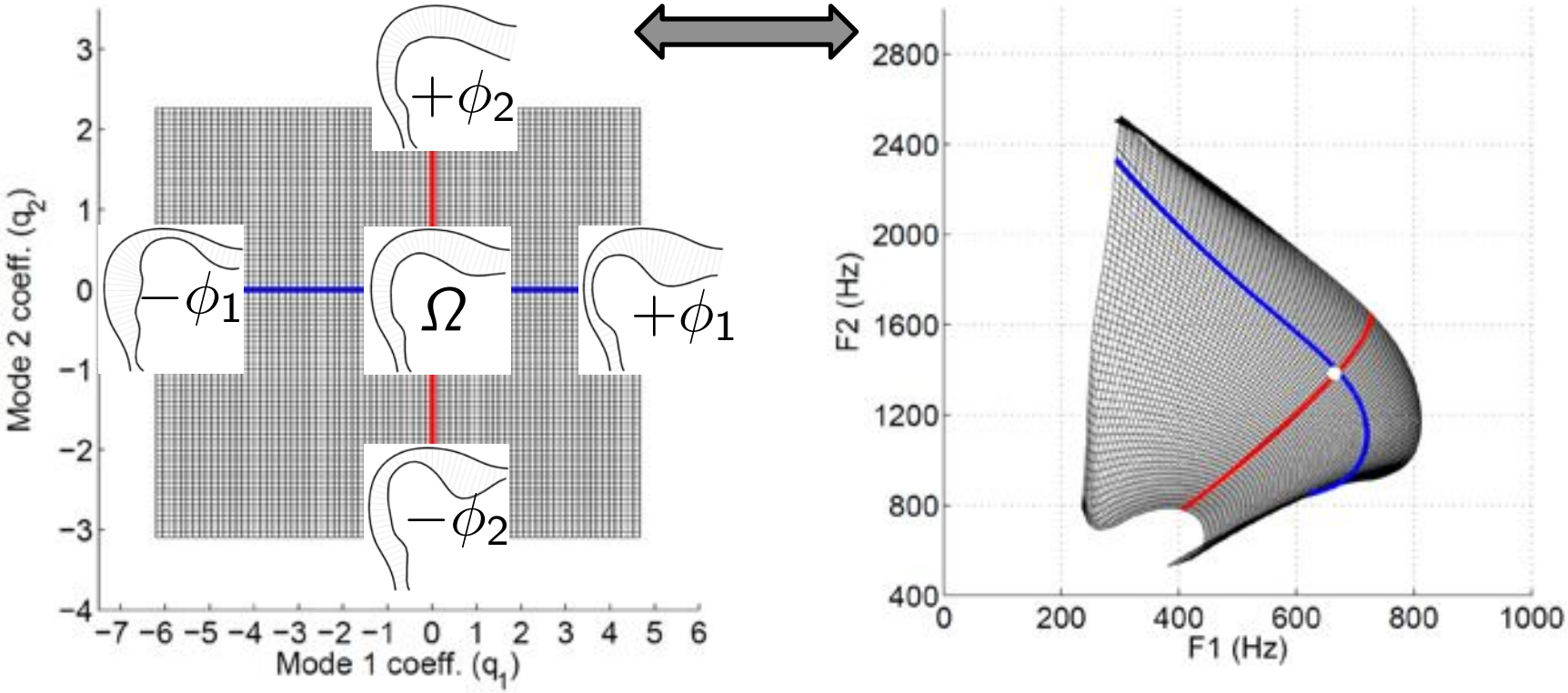
Brad Story, Zemlin Lecture, ASHA, 2013

One-to-one mapping from vocal tract shape to vowel space

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

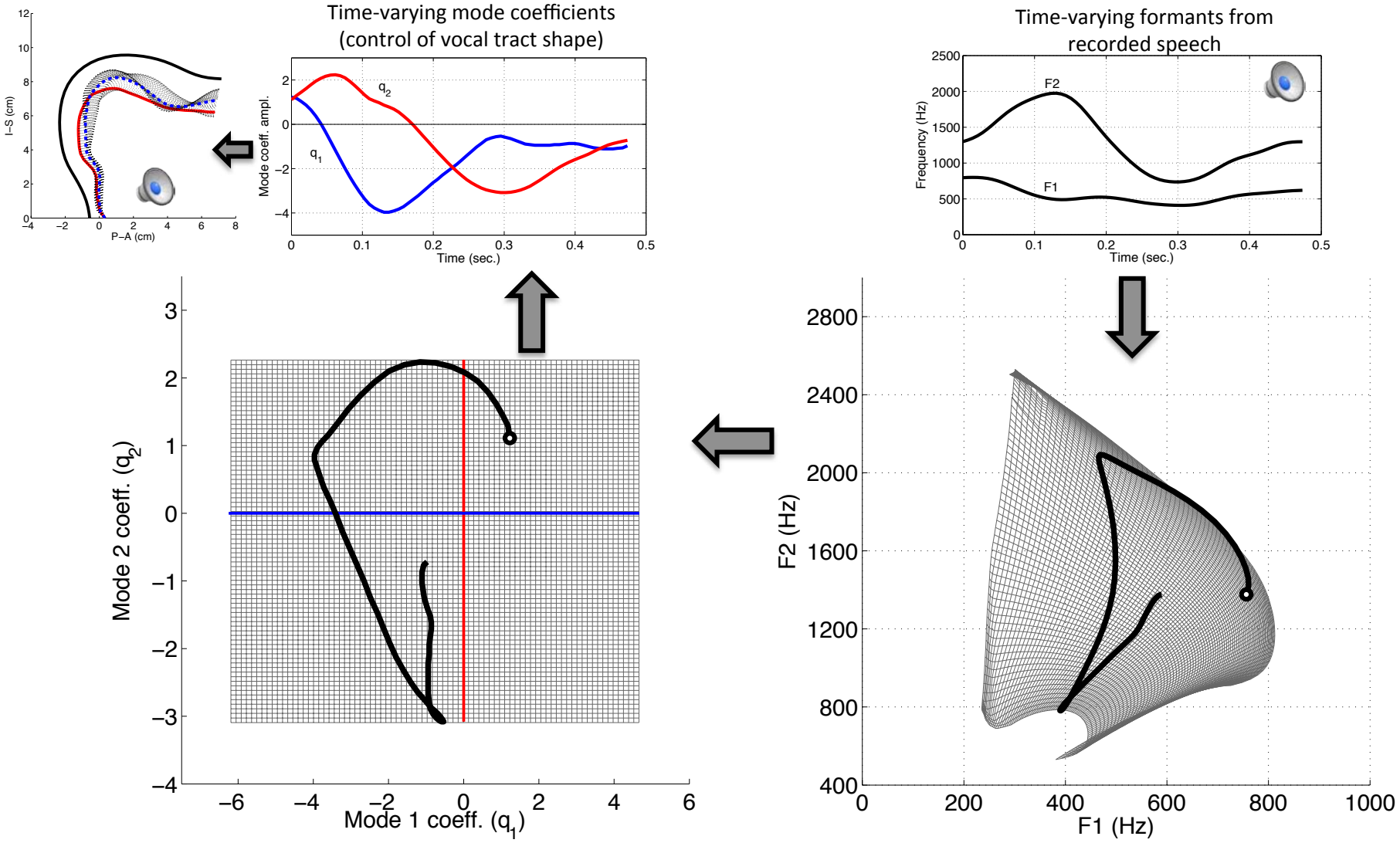
Average (mean)

[F1, F2] vowel space



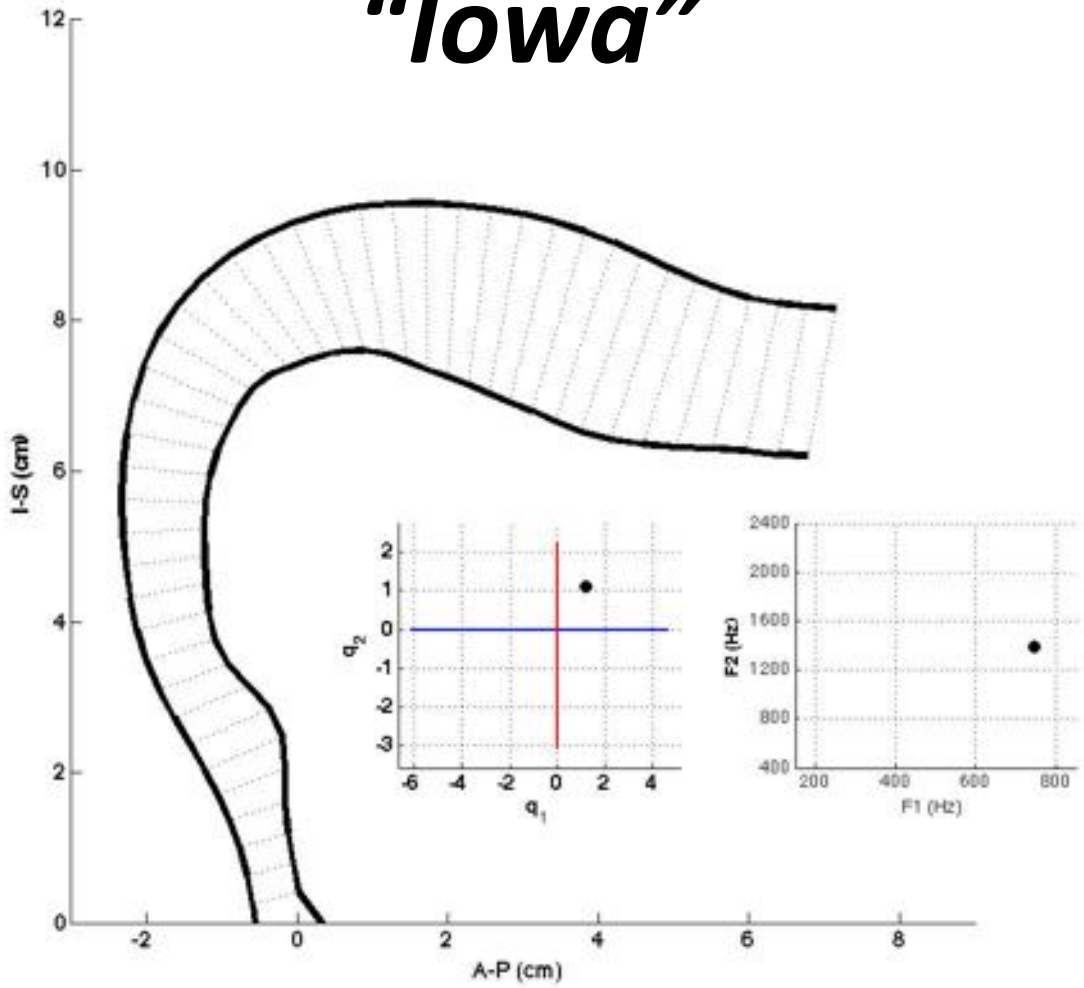
Brad Story, Zemlin Lecture, ASHA, 2013

2. Mapping from formants to vocal tract shape



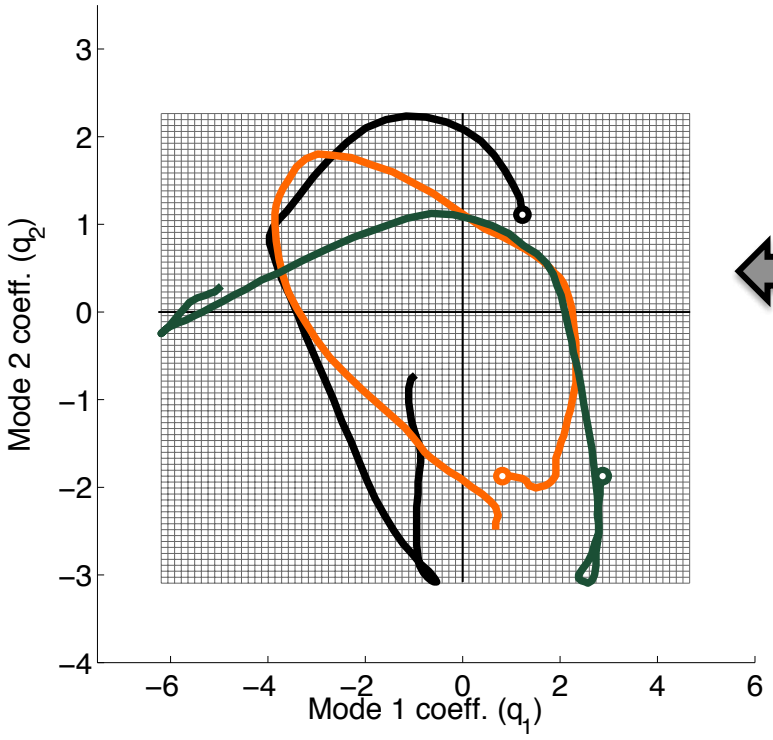
Mapping between vocal tract shape and [F1,F2] trajectory

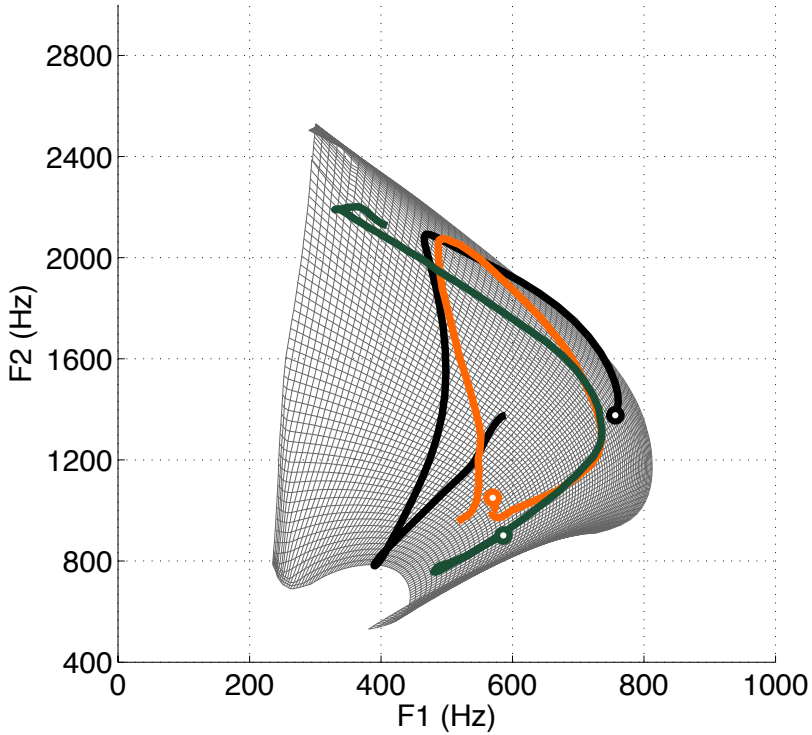
"Iowa"



Brad Story, Zemlin Lecture, ASHA, 2013

Maps of Three States

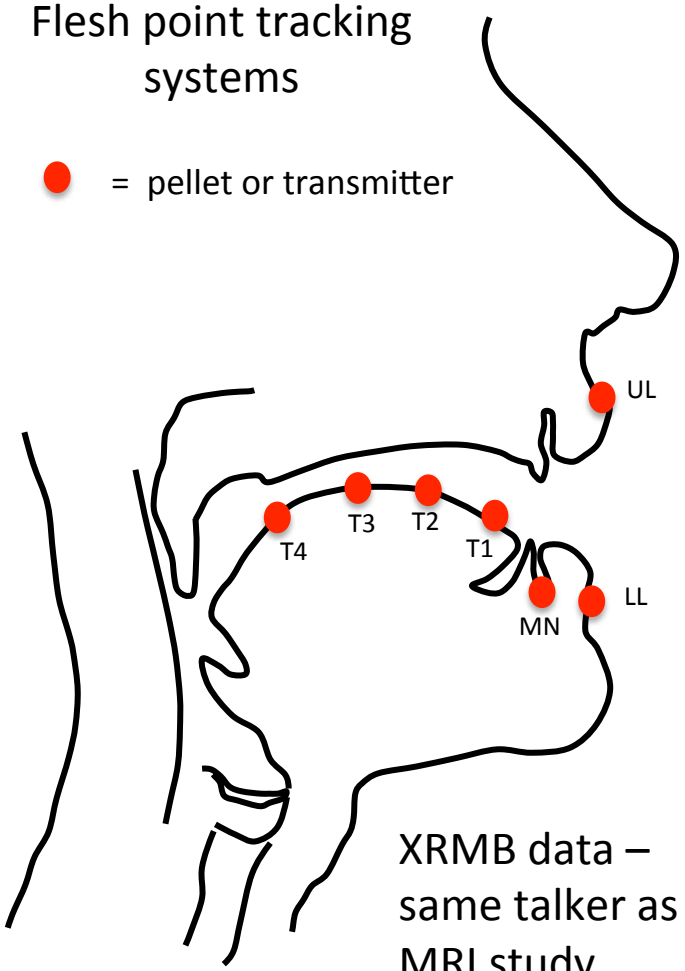




3. Compare to Articulatory Kinematic Data

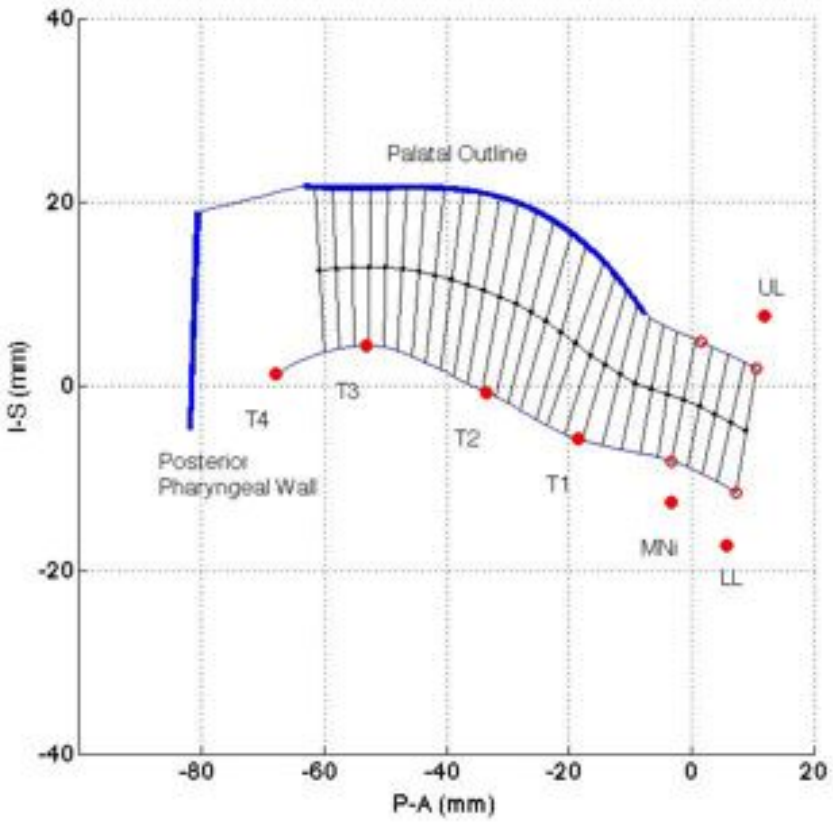
Flesh point tracking systems

= pellet or transmitter



XRMB data – same talker as MRI study

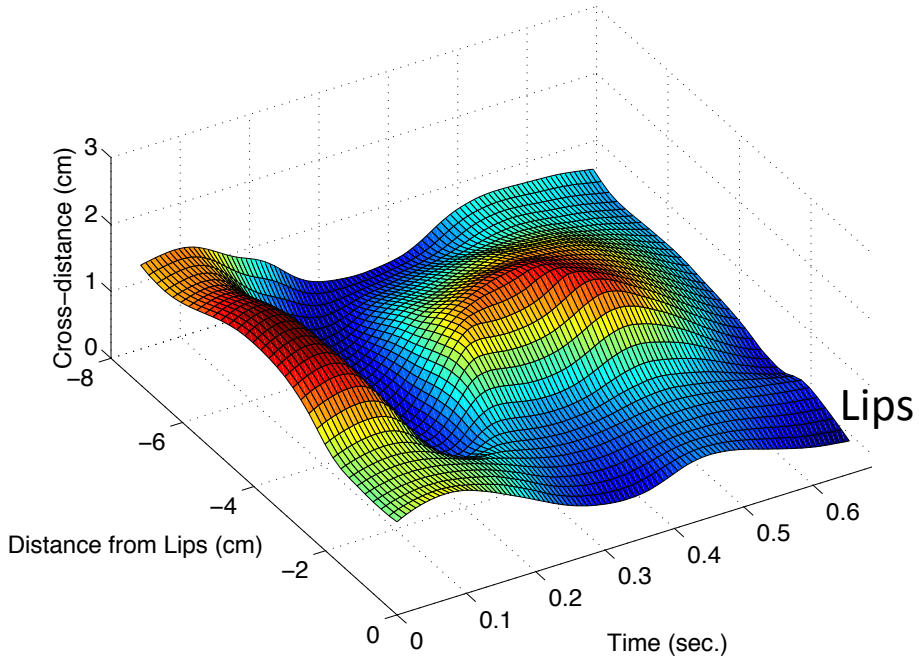
"Iowa"



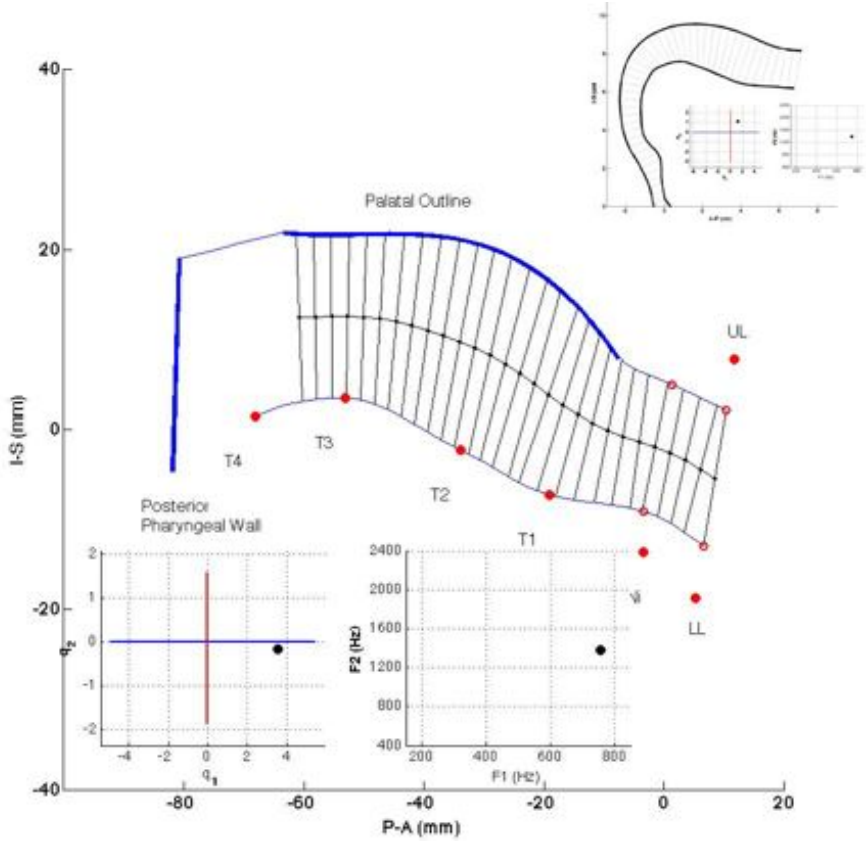
3. Compare to Articulatory Kinematic Data

“Iowa”

Time-varying cross-distance (XRMB)



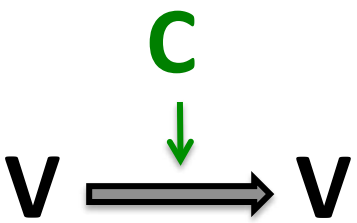
XRMB mapping between vocal tract shape and [F1,F2] trajectory

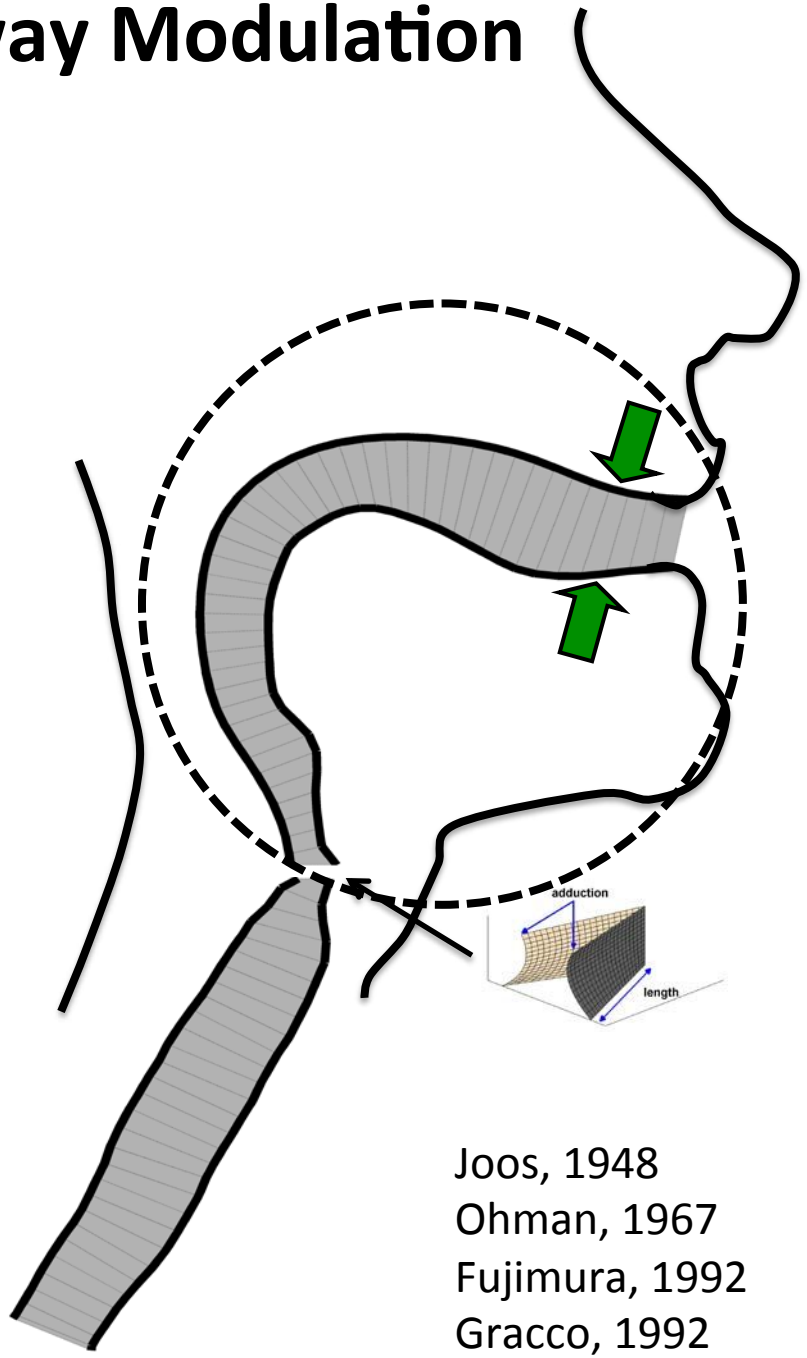


Two "Tiers" of Airway Modulation

Tier 1 - Shaping: slowly-varying changes imposed on the shape of a neutral vocal tract - ***vowels***

- **Tier 2 - Valving**: modulate vowels with constrictions - ***consonants***

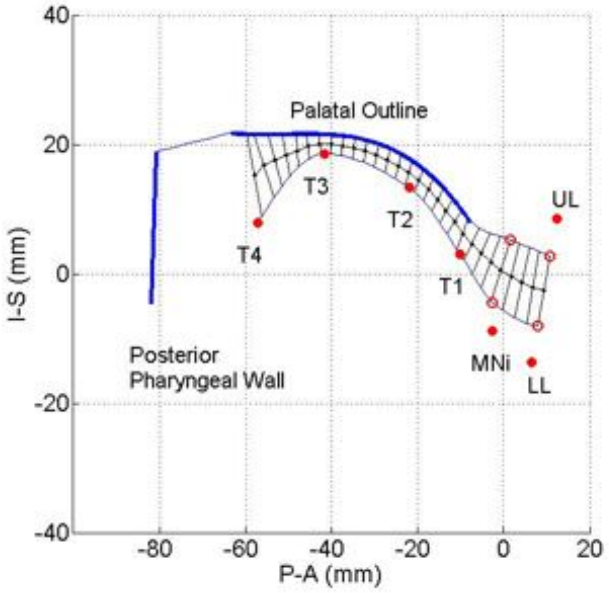




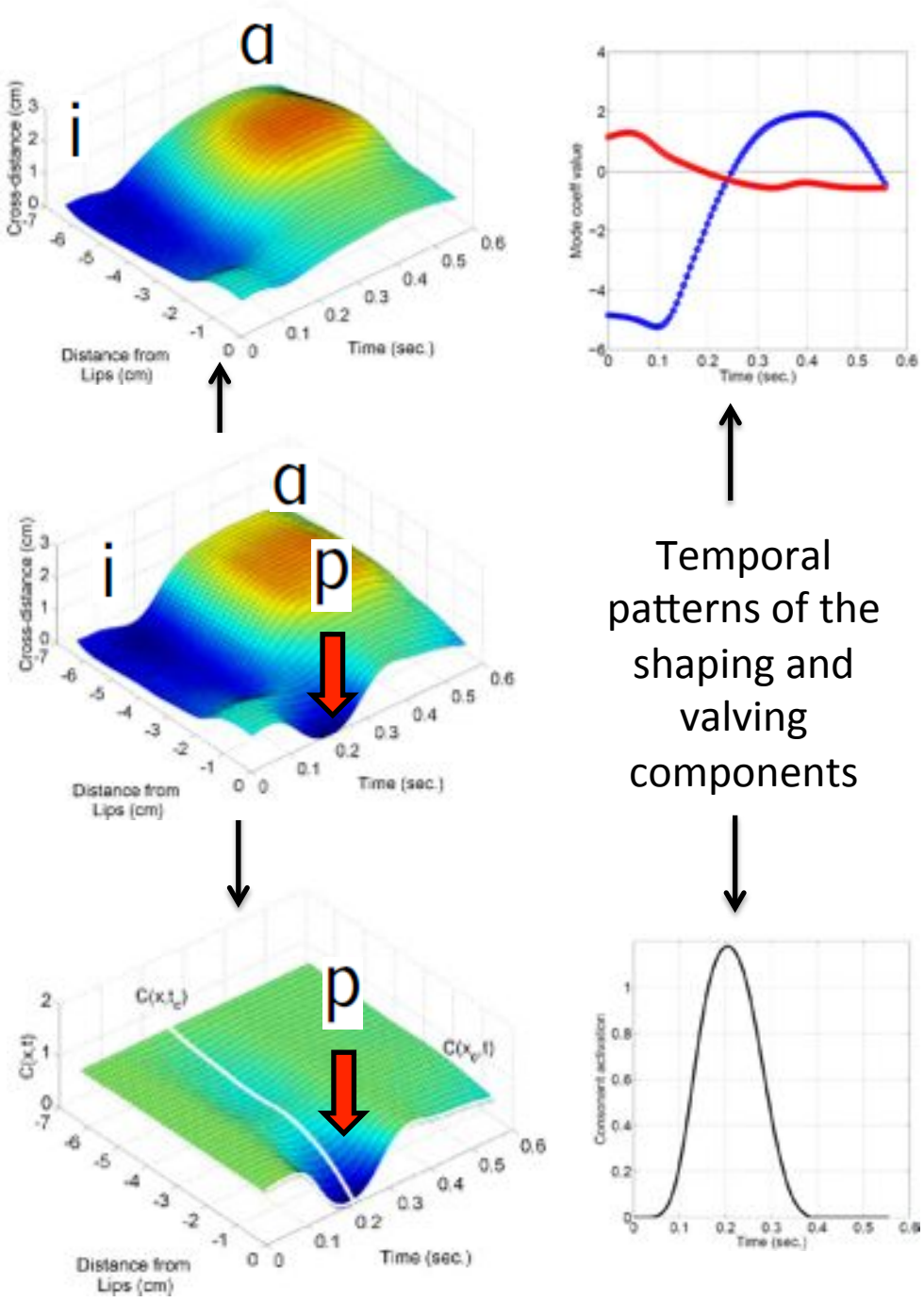
Joos, 1948
Ohman, 1967
Fujimura, 1992
Gracco, 1992

Articulatory Kinematic Data:

Separate the shaping and valving components



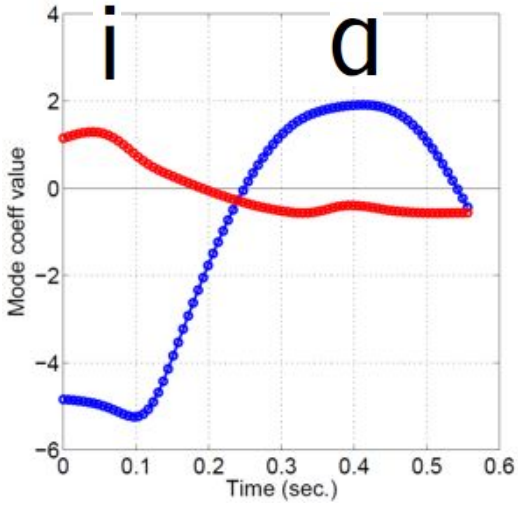
ipa

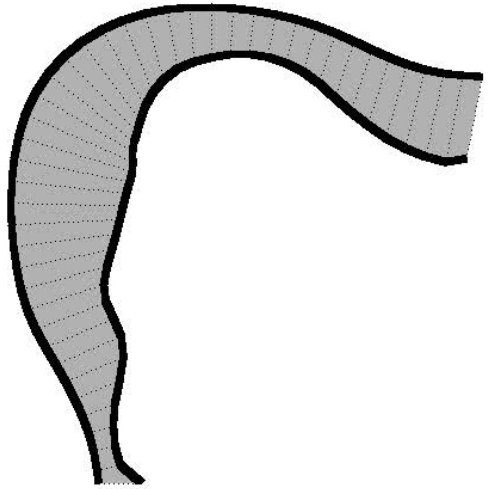


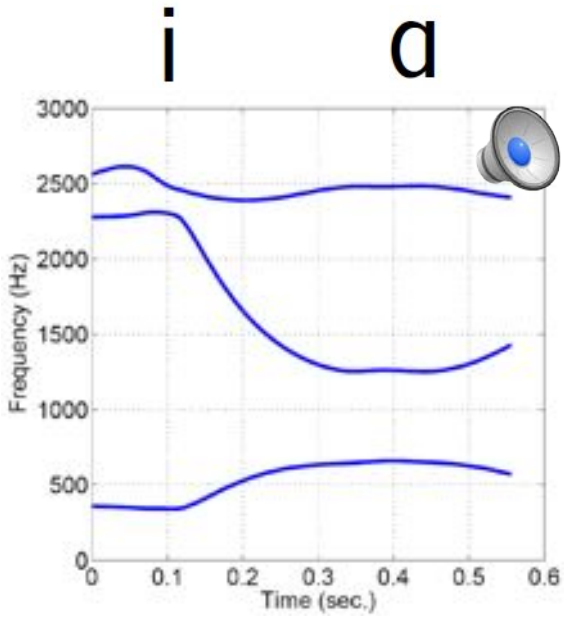
Temporal patterns of the shaping and valving components

Vowel and consonant contributions to vocal tract shape J. Acoust. Soc. Am. (2009). Story

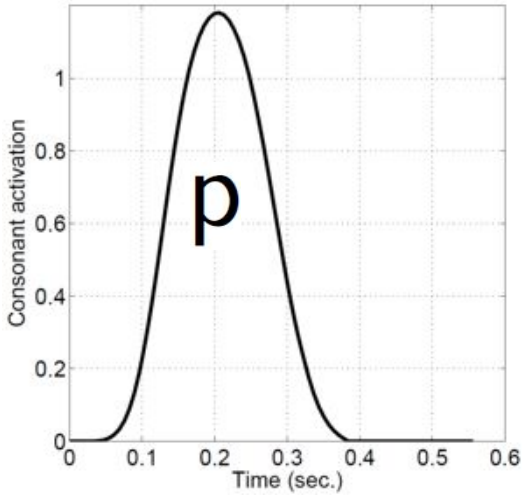
Mode coeffs

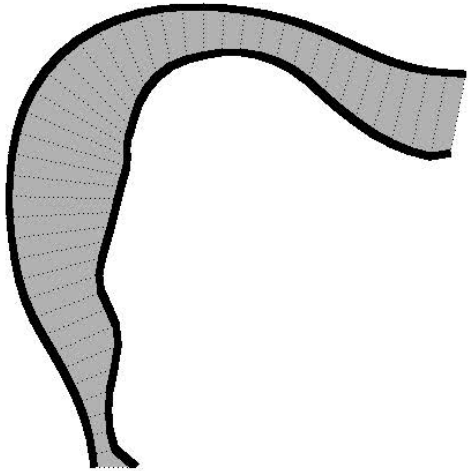


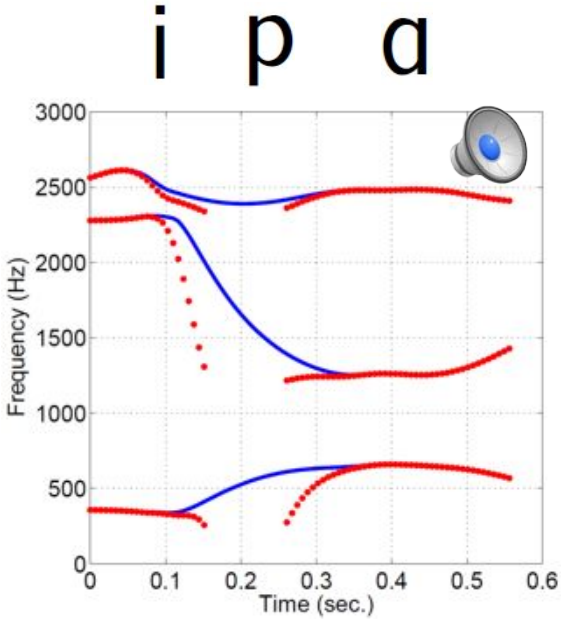




+Constriction





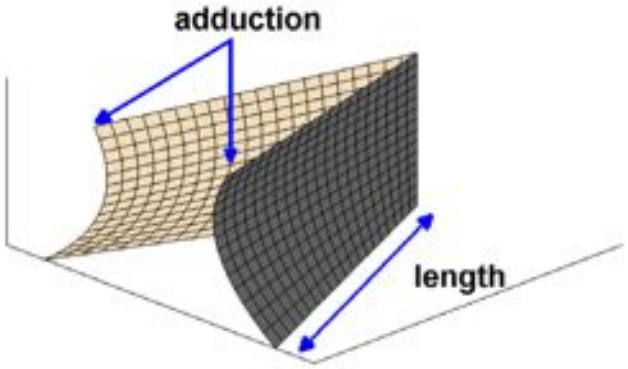


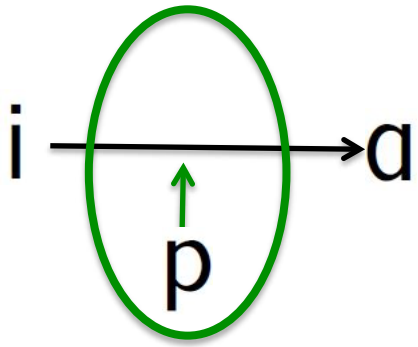
Brad Story, Zemlin Lecture, ASHA, 2013

Simultaneous Laryngeal Actions

- Must change ***fundamental frequency F0*** *(vibrational frequency of vocal folds).*

- ***Abduct*** vocal folds for voiceless consonants (and respiration), and ***adduct*** for voiced consonants and vowels





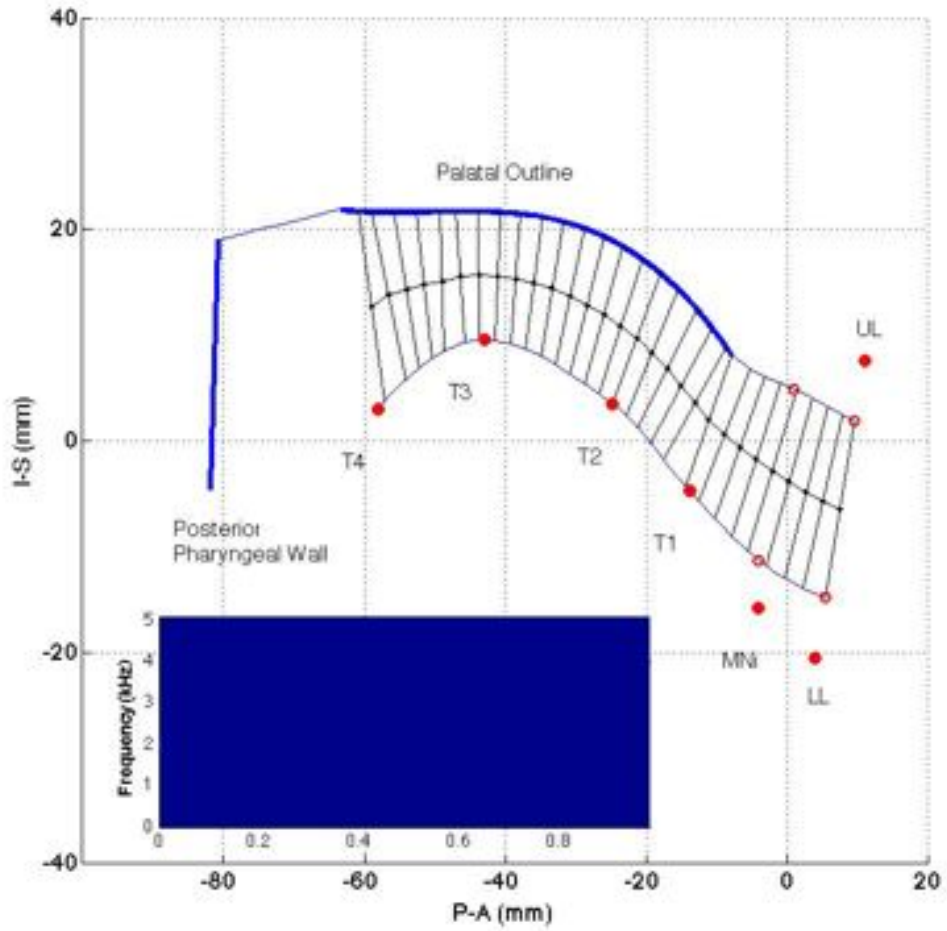
Brad Story, Zemlin Lecture, ASHA, 2013

Apply to word and phrase level speech: **“abracadabra”**

C… → b r k d b r

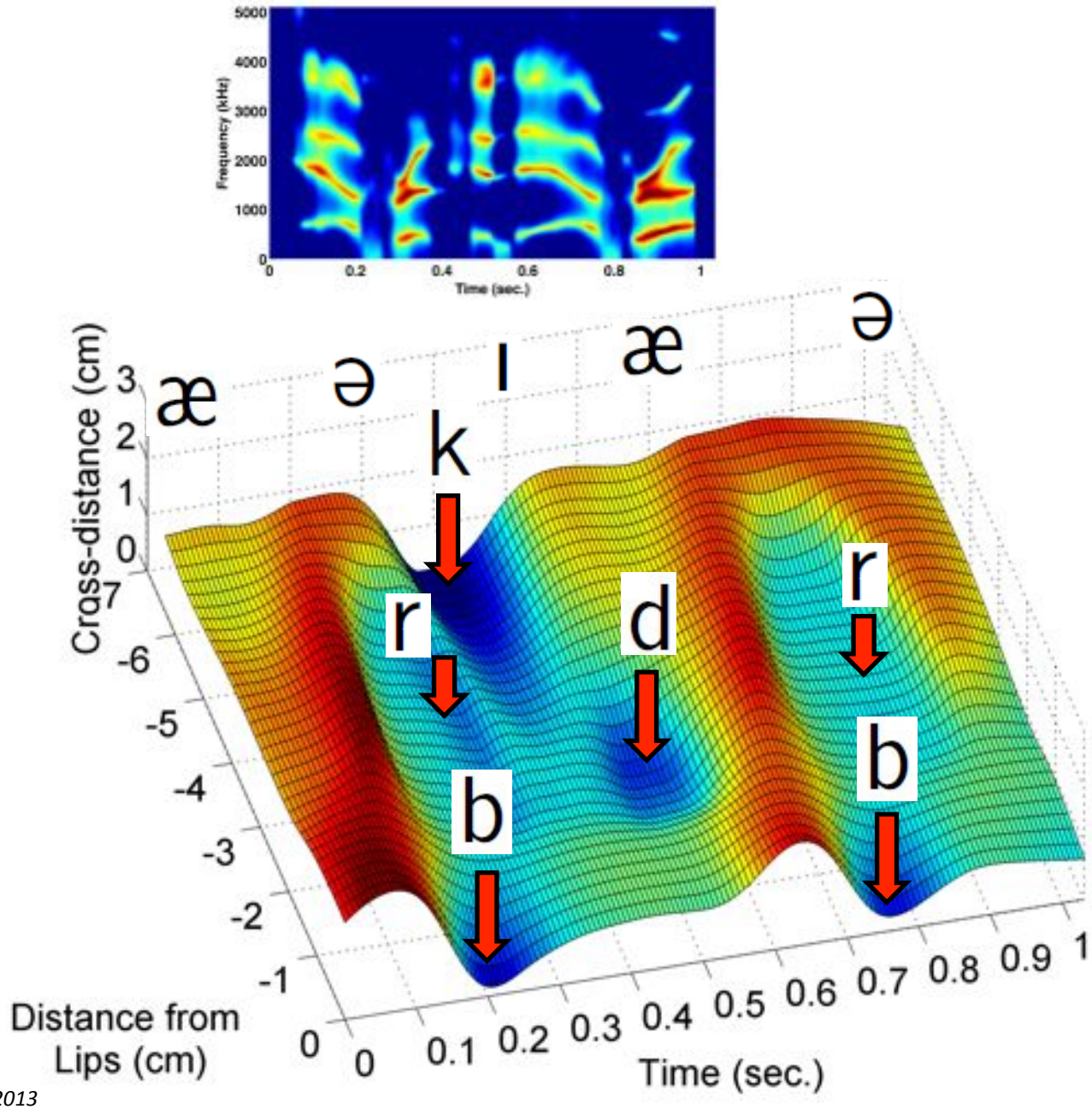
VV… → æ ə ɪ æ ə

- Use both kinematic and acoustic analysis, along with models, to simulate word and phrase-level speech



Brad Story, Zemlin Lecture, ASHA, 2013

Apply to word and phrase level speech: **“abracadabra”**

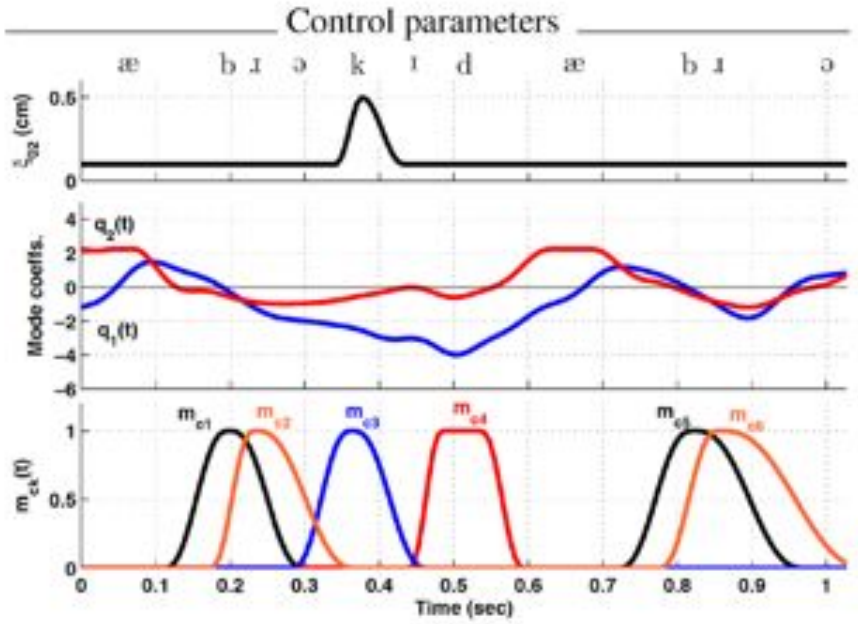


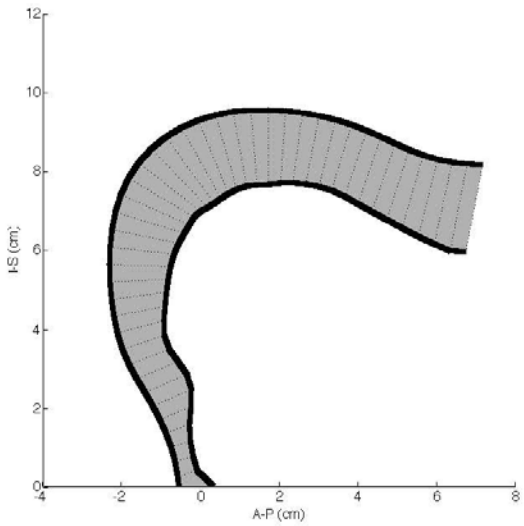
Brad Story, Zemlin Lecture, ASHA, 2013

VF adduction/ abduction

Shaping

Valving

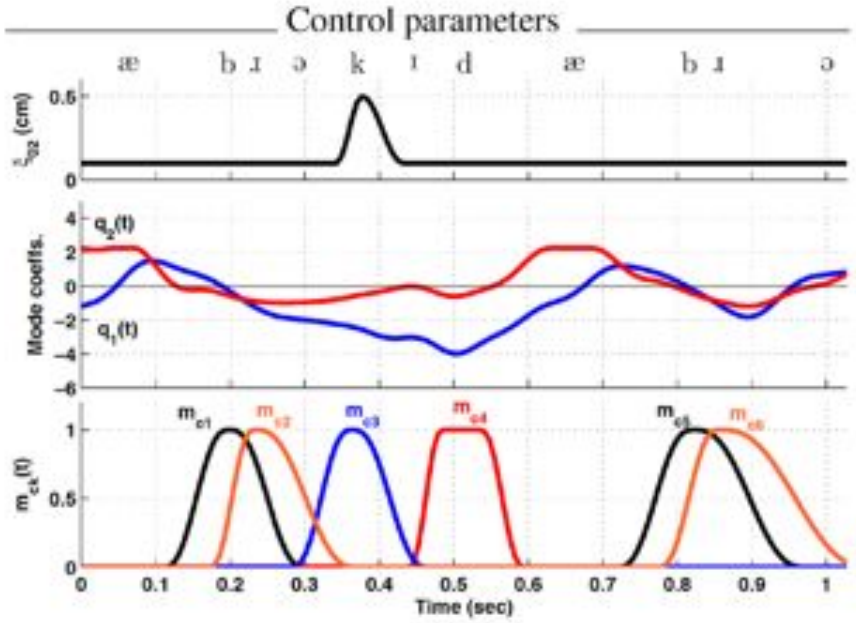


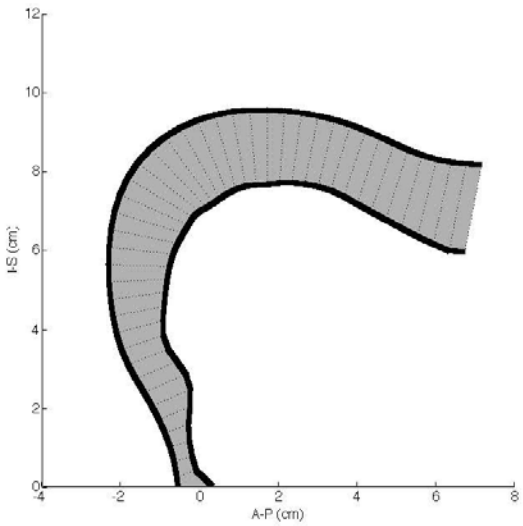


VF adduction/ abduction

Shaping

Valving

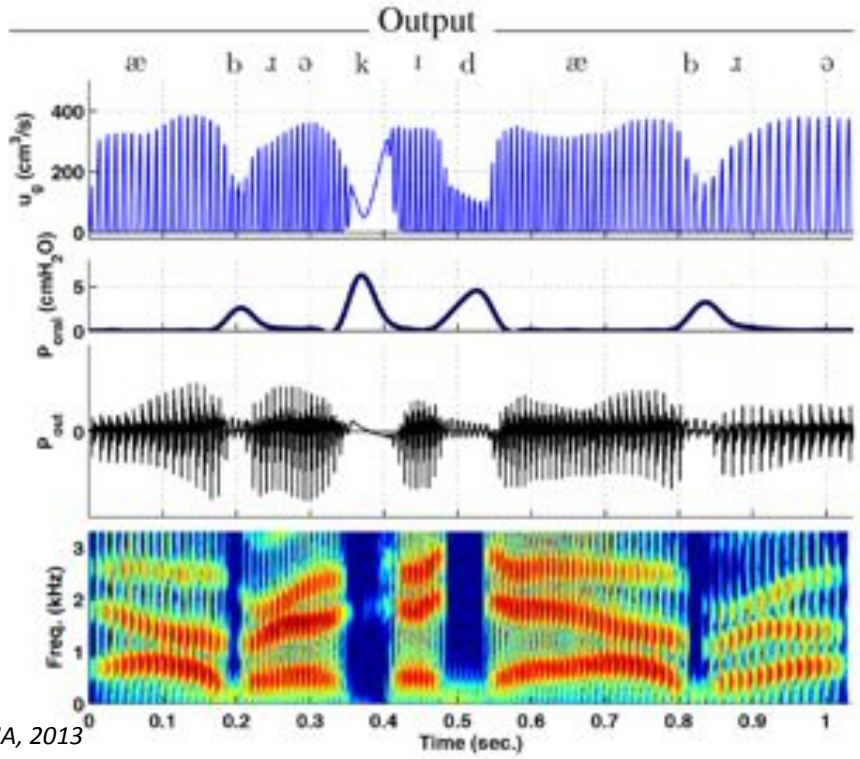




Glottal flow

Intraoral pressure

Radiated acoustic pressure



"Phrase-level speech simulation with an airway modulation model of speech production"

Computer Speech and Language, Story, (2013)

Modifications

Temporal Patterns:

Other Phrases:

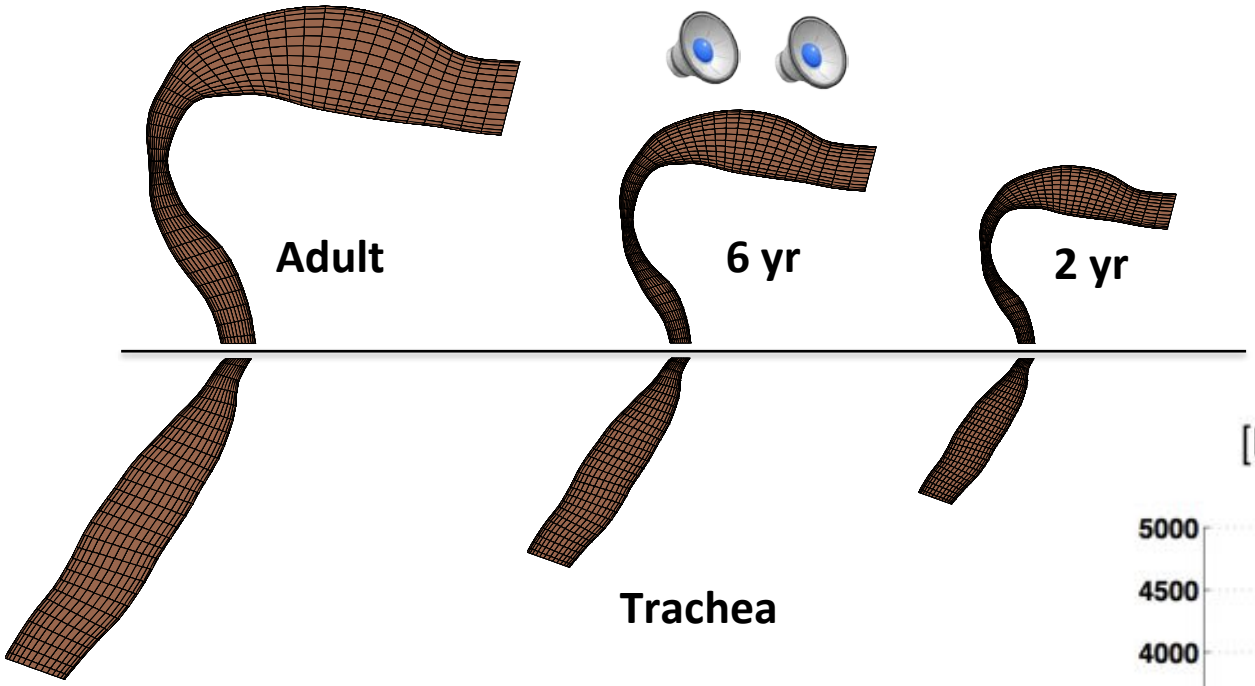
Speech

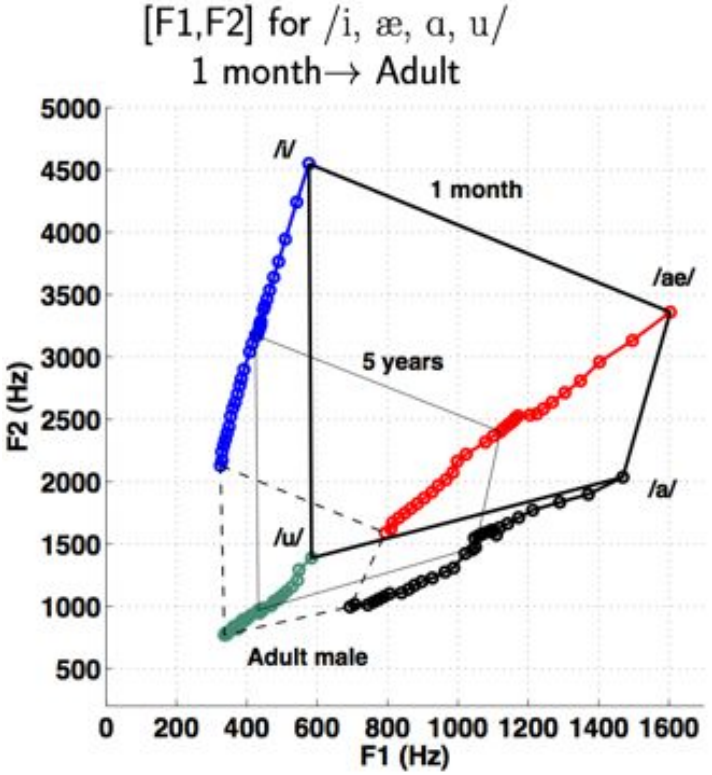
Song 1

Song 2

Song 3

Vocal tract changes during development





In collaboration with Kate Bunton (U. Arizona) and Houri Vorperian(Waisman Ctr., UW-Madison)

Willard R. Zemlin
Speech and Hearing Science
Anatomy and Physiology
Preface to the 4th Edition

"Communication is an incredibly complex process, by no means fully understood.

We, as humans, are complex and a description of the structure and function of the speech and hearing mechanism must of necessity also be complex, if it is to be at all complete"

…and we must use models to simplify both structure and function if it is to be at all understood.

Thank you!