USC University of Southern California

SAIL

Bridging Speech Science and Technology –
Now and Into the Future

Shrikanth (Shri) Narayanan

University Professor and Nikias Chair in Engineering, **USC**
Signal Analysis and Interpretation Laboratory (**SAIL)**
Information Sciences Institute (**ISI**)

Visiting Faculty Researcher, **Google**

Co-founder/Chief Scientist, Behavioral Signals
Co-founder/Chief Science Officer, Lyssn

Interspeech 2023
Dublin, Ireland

USC School of Engineering

http://sail.usc.edu
University of Southern California

Immensely grateful to ISCA

for giving a **home**—*a warm and collegial interdisciplinary forum for fellowship*—to grow and contribute and to develop enduring friendships

(*since my first meeting attendance in Berlin'93*)

Deepest gratitude to all my incredible, impressive, inspiring, intellectually generous and indulgent

- *Teachers*
- *Mentors*
- *Colleagues*
- *Collaborators*
- *Students*

(highly intersecting sets!)

So many to name, but indebted to all

And the institutions that have educated, shaped and supported me

CEG

AT&T Shannon Labs

Signal Analysis and Interpretation Laboratory

....technologies to understand the human condition
and to support and enhance human capabilities and experiences

SAiL

creating inclusive technologies and technologies for inclusion

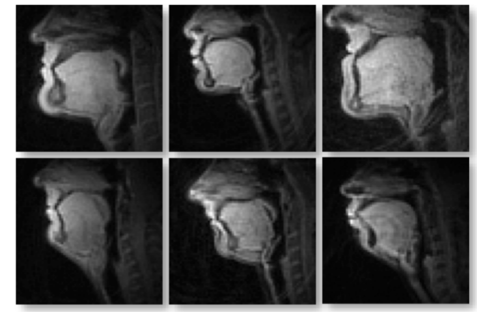
EST. 2000

USC
School of Engineering

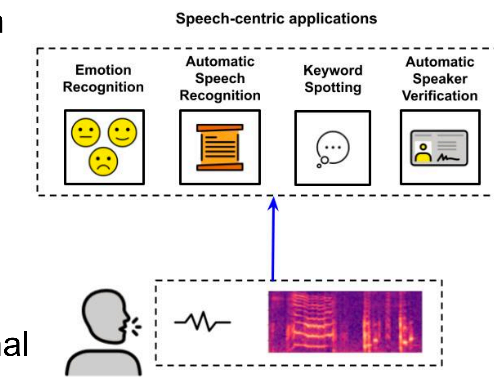
http://sail.usc.edu

Various research threads

- ***Speech Science and Linguistics*** ✅
  - investigating speech and language production: from its cognitive conception, to its biomechanical execution, to its signal properties
  - diagnostic and therapeutic applications in cancer, neurological disorders
  - Supported by: Dynamic Imaging Science Center, Brain & Creativity Institute

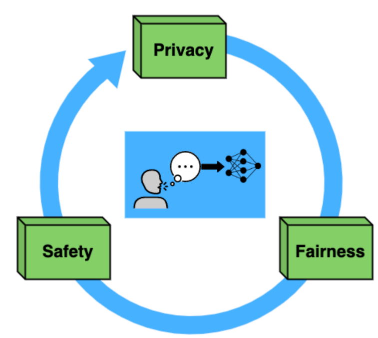


- ***Speech, Audio and Language Processing***
  - Speech activity detection, Speaker diarization, Automatic Speaker and Speech recognition, Prosody modeling, Nonverbal vocalization/disfluency modeling, Speech synthesis, Speech translation, Dialog/Conversational agents
  - Inclusive and robust speech processing: children speech, pathological speech
  - Multilingual spoken language processing, Cross-cultural interactions
  - Audio processing, sound event modeling, music information retrieval



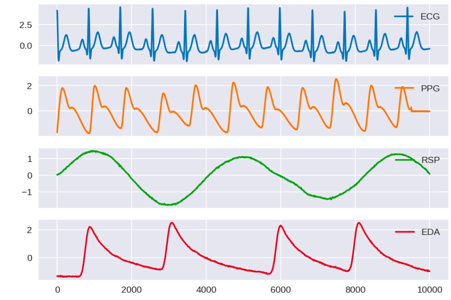
- ***Emotions Research and Affective Computing***
  - Speech emotion expression, perception and recognition: scientific inquiry and computational modeling
  - Language expressions of emotion, social media, evaluation of foundation models
  - Multimodal affective computing: from human speech, language, interaction, and other biobehavioral signals

- ***Trustworthy Machine Intelligence***
  - Federated learning, applications in audio, multimodal data, biosignal/medical imaging
  - Adversarial attacks in human centered realms: incl. speech/speaker recognition
  - Ethics in human subjects research, privacy constraints of human-centered signals

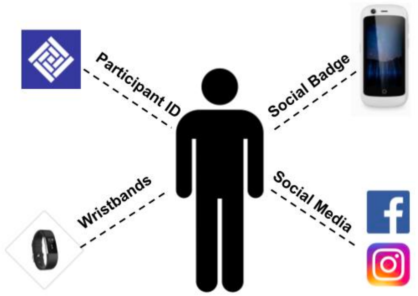


Various SAIL research threads

- ***Biosignal Sensing, Imaging and Modeling***
  - Signal processing of wearable and biomedical signals: ECG, EDA, EEG, Eye Tracking, Accelerometry, … ; environmental signals
  - Medical Imaging (esp. head/neck, airway/vocal tract, brain MRI, ophthalmology)
  - Experimentation in ecologically valid, natural settings



- ***Behavioral Signal Processing and Behavioral Machine Intelligence*** ✅
  - Engineering approaches from wearable & environmental sensing to AI methods — to illuminate human trait, state and behavior
  - Create tools for screening, diagnostic, intervention support
  - Application domains across the life span: developmental disorders (notably, Autism), anxiety, depression, OCD, suicide, relationships, dementia; health and wellbeing in workplace, home

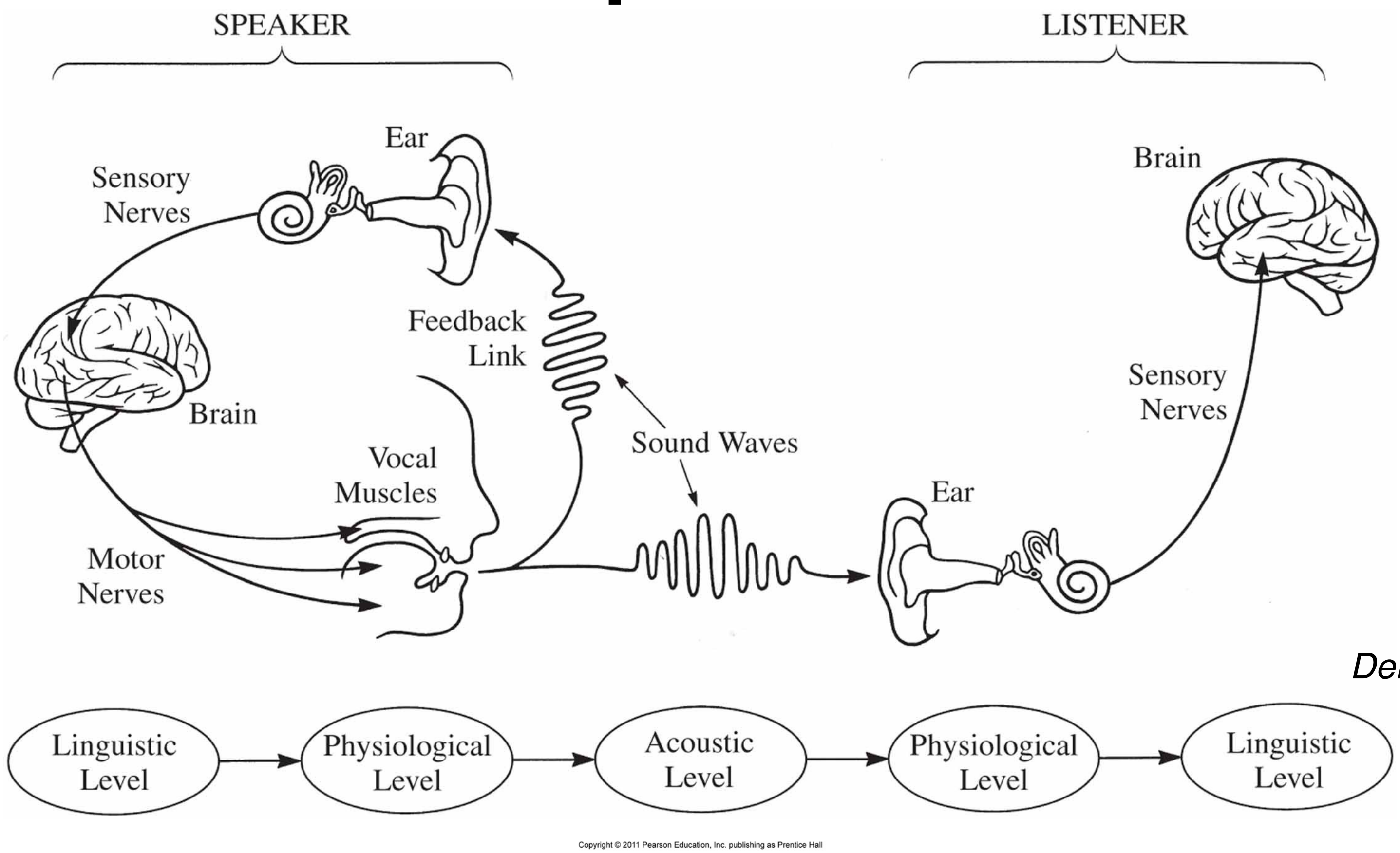


- ***Computational Media Intelligence*** ✅
  - Understanding media stories, and their impact on human experiences, behavior and action: from individual to socio-cultural scale, including affective aspects, aesthetics
  - Creating AI tools (video, audio, language analysis) for understanding representations and portrayals, including stereotypes, in media such as TV, movies, ads, news, …
  - Music Information Processing and Retrieval: film music, music videos, music generation

7

The fascinating universe of speech communication

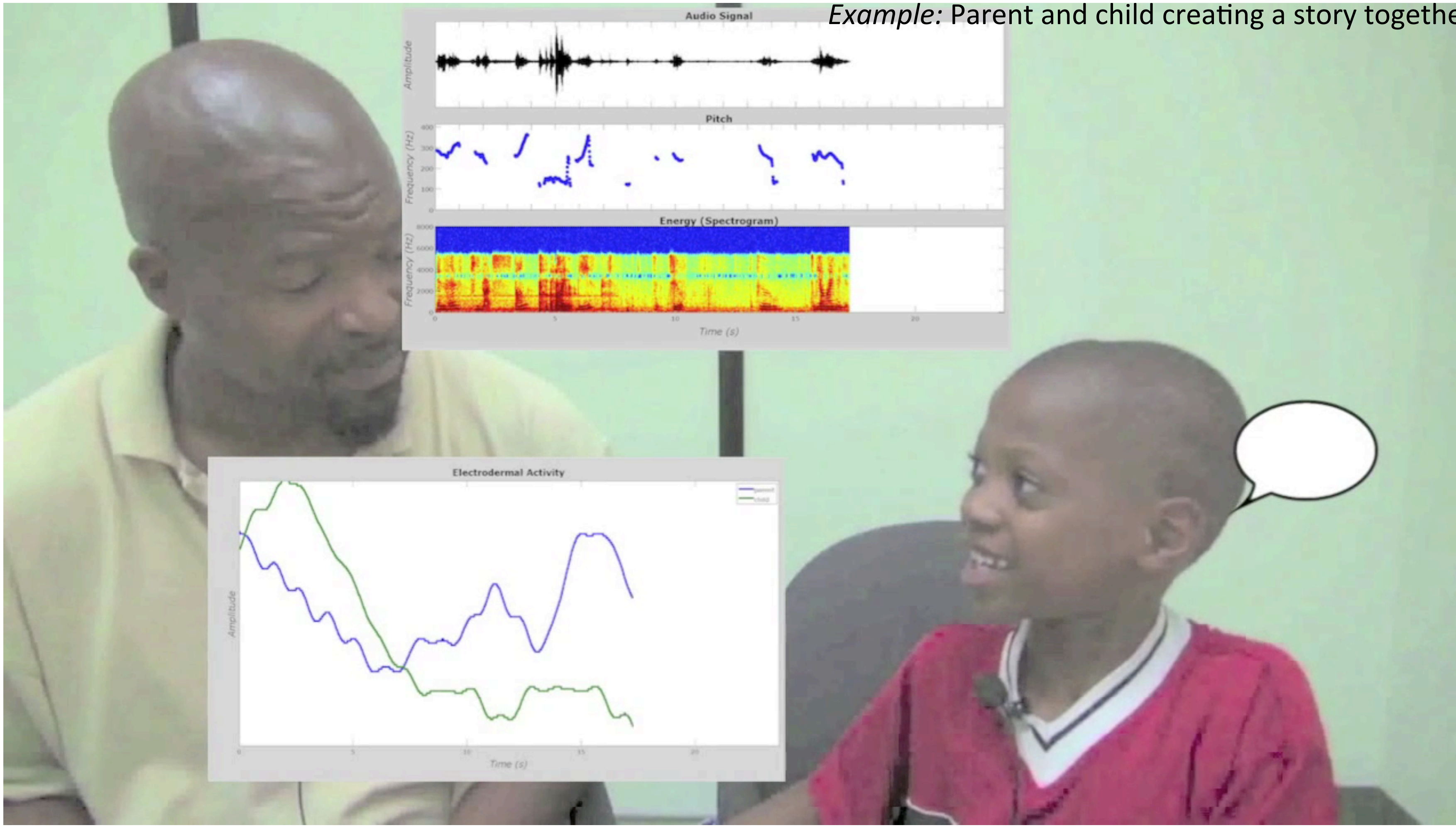
The Speech Chain



Denes & Pinson, 1963

- **complex orchestration of mental, physiological, physical, social processes**
- **information encoding and processing at multiple levels:**
  - *neuro-cognitive, motoric, sensory, socio-behavioral*
- **a signal *from, for—and about—*people**

Speech Chain in Action



- *speech and language encode and provide access to* ***intent, emotions****, and a variety of information about* ***demographic traits*** *(age, gender, size…),* ***physical/psychological/health state,*** *and* ***interaction context***
- *these attributes/constructs are often intricately related*

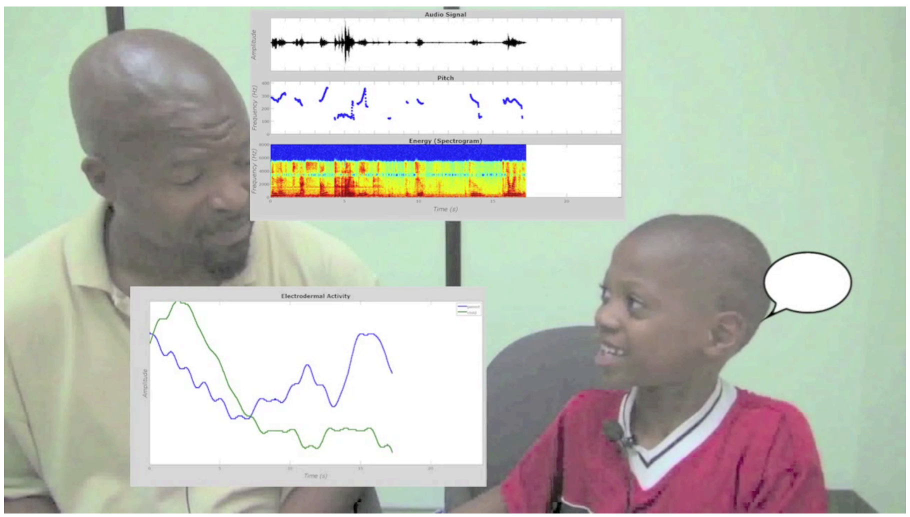
a “feature” than “bug”

Rich diversity and variability along many dimensions

within and across people and their contexts

- **Individual demographic differences**
  - age, gender, socio-cognitive levels, language background
  - ability: neuro-cognitive diversity, e.g., verbal, minimally verbal, non verbal ability
- **Interaction details**
  - dyadic, triadic, small group,…; structured, semi structured, free unrestricted
- **Interlocutors involved and socio-cultural context**
  - siblings, peers, parents, clinicians, teachers, therapists, unfamiliar people
- **Environment and ambient context**
  - speech, non speech human sounds
  - environmental sounds of home, school, clinic, playground, …
  - outdoor, indoor, and variability over time therein
- **Sensing technology possibilities**
  - on-person/environment, close-talking/far-field, accompanying video, meta data
- **Processing/modeling goals and purpose**
  - local details (e.g. amount of speech), global behavioral details (affect)
  - handle varying types of abstraction in data and desired description

Speech Science and Technology Research: Twin intertwined goals



Analyzing, understanding, characterizing *variability*

Addressing the influence of the multiple interacting sources of *variability*

E.g., disassociating neurocognitive differences in the presence of factors related to age-dependent physical development and biological sex ("gender") differences in children in learning context

A perennial endeavor

- Technologies that work for <u>everyone</u> and in <u>all</u> contexts: understand and create experiences consistent with the rich variety in *who, what, where, how, when*,…

Inclusive technologies essential for equitable experiences

Shrikanth S. Narayanan and Asad M Madni. Inclusive Human centered Machine Intelligence. The Bridge, 50:113–116, National Academy of Engineering, 2020.

Thank you organizers!

INTERSPEECH 2023
20th-24th August 2023 - Dublin, Ireland
https://interspeech2023.org/

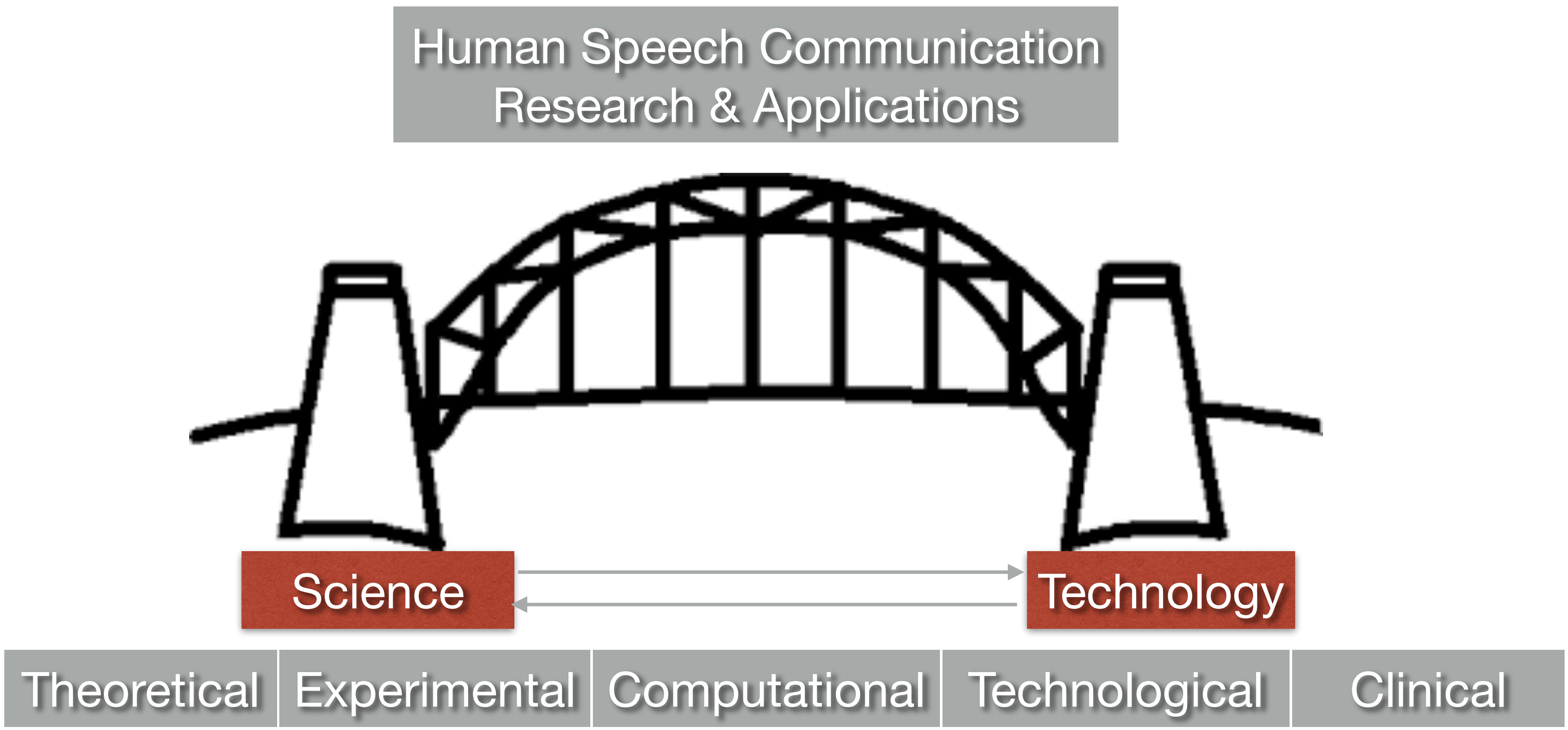
Interspeech 2023

Inclusive Spoken Language Science and Technology – Breaking Down Barriers

- Understanding the rich diversity and variability in human speech communication
- Creating technologies that are trustworthy and be trusted: robust, inclusive and equitable, safe and secure, …

An interdisciplinary expedition

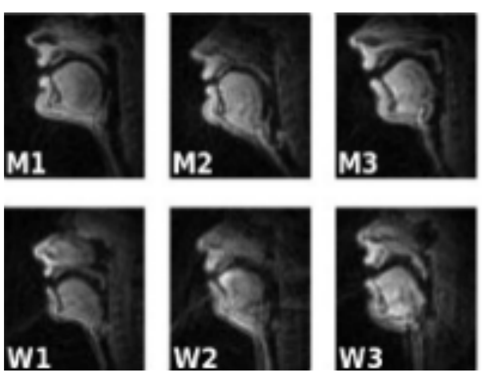
bridging science and technology—one constantly driving the other—leading to novel insights and innovative applications

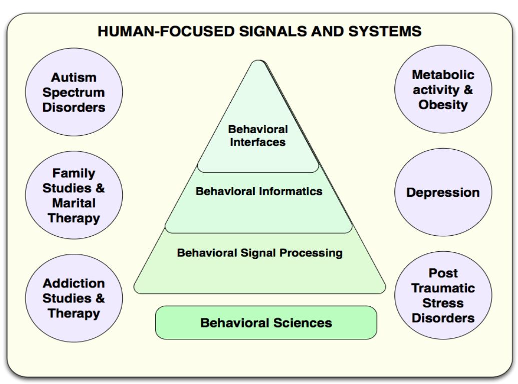


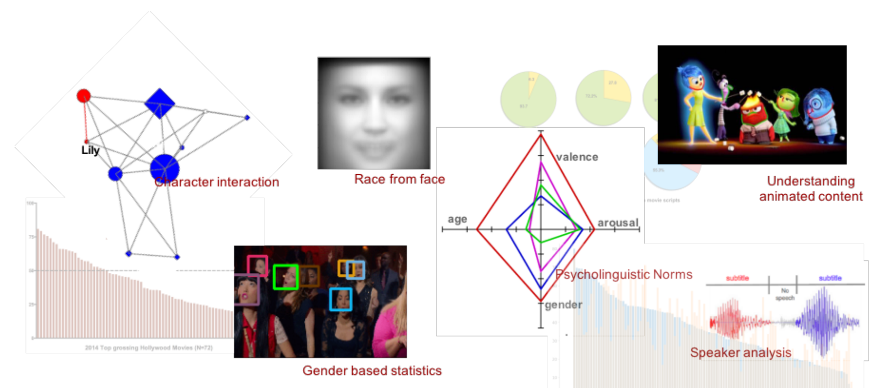
Highlights sampled from three research threads

- **Speech Science**
  - *the most fascinating human vocal instrument*
- **Rich speech processing**
  - *behavioral machine intelligence for health*
- **Speech processing supporting media intelligence**
  - *inclusive media representations and portrayals*

THEN — NOW — NEXT







Highlight 1

Speech Science

From multimodal data and models to

scientific discovery and clinical advances

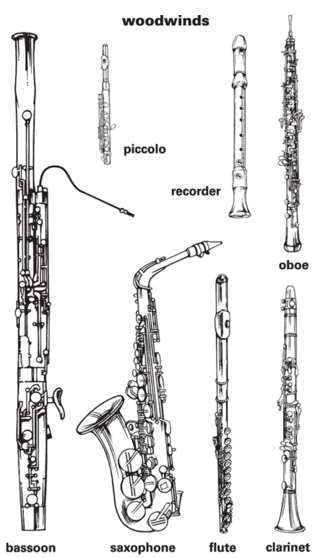
- investigating speech and language production: from its cognitive conception, to its bio-mechanical execution, to its signal properties
- technology applications in speech recognition, biometrics, synthesis
- diagnostic and therapeutic applications in cancer, neurological disorders

Christina Hagedorn, Tanner Sorensen, Adam Lammert, Asterios Toutios, Louis Goldstein, Dani Byrd, Shrikanth Narayanan. Engineering Innovation in Speech Science: Data and Technologies. *SIG 19 Speech Science Perspectives of ASHA*. 4(2): 411-420, 2019

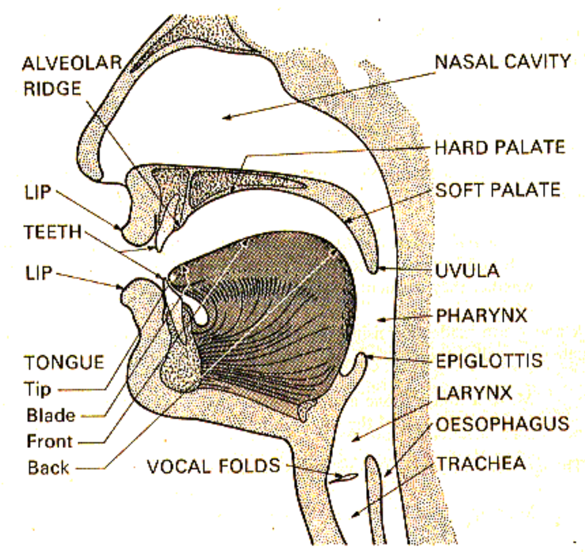
Focus on: “how do we play the vocal instrument?”

Multimodal methods in Speech Communication Science

understanding the structure and function of the human vocal instrument



https://gabbygomez.weebly.com/woodwind-istruments.html



Wells and Colson, Practical Phonetics (1971)

Speech communication science

A long and rich history in the use of technologies for

- *Acquiring the right data*
  - from recording speech audio in a variety of environments to using advanced instrumental techniques for observing speech production
- *Analyzing and modeling data*
  - from spectral analysis and linear prediction to novel machine learning methods for analysis, data visualization and theory building
- *Applying and using data*
  - from facilitating screening and diagnostics to supporting interventions and tracking outcomes

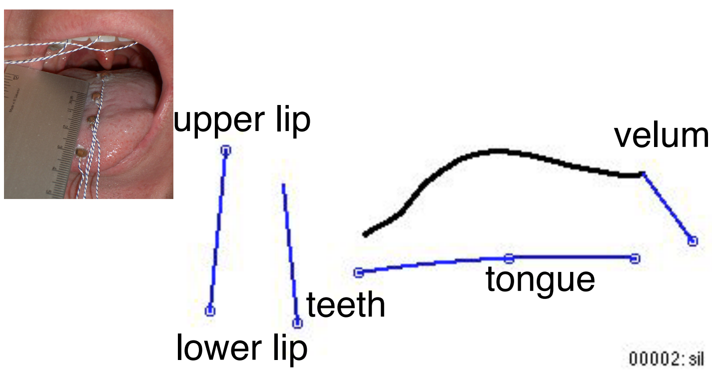
Technologies for illuminating speech production

X-ray (Stevens, 1962)

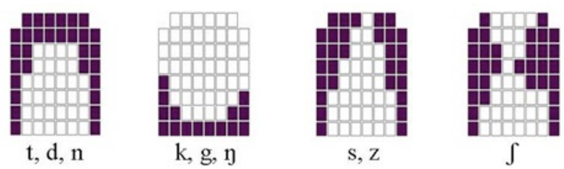
http://psyc.queensu.ca/~munhallk/05_database.htm

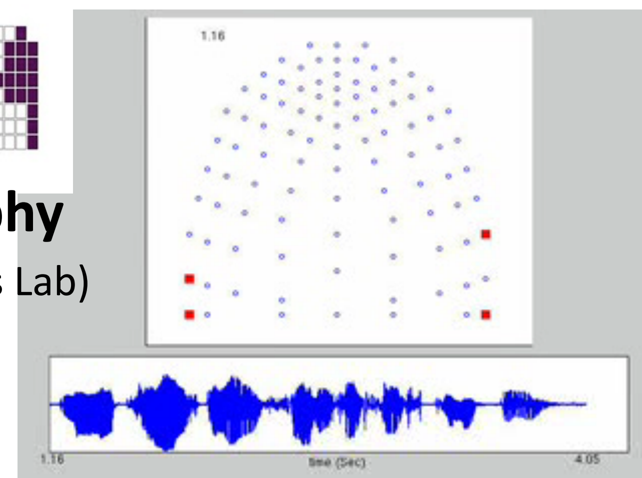
Ultrasound (Stone, 1980)

http://www.speech.umaryland.edu



Electromagnetometry





Electropalatography

(courtesy: UCLA Phonetics Lab)

(1990s)

Newer Possibilities Emerged: Structural magnetic resonance imaging (MRI)

circa 90s

Capable of 3D imaging of the hydrogen concentration in human body

Number of advantages:

- Non-invasive, no ionizing radiation
- Arbitrary scan plane: Information on complete vocal tract geometry
- Excellent, flexible structural differentiation: Good soft tissue contrast, SNR
- Amenable to computerized 3D modeling: reconstruction and visualization
- Quantitative information: area function and acoustic relations
- Variability analyses

Limitations/Challenges

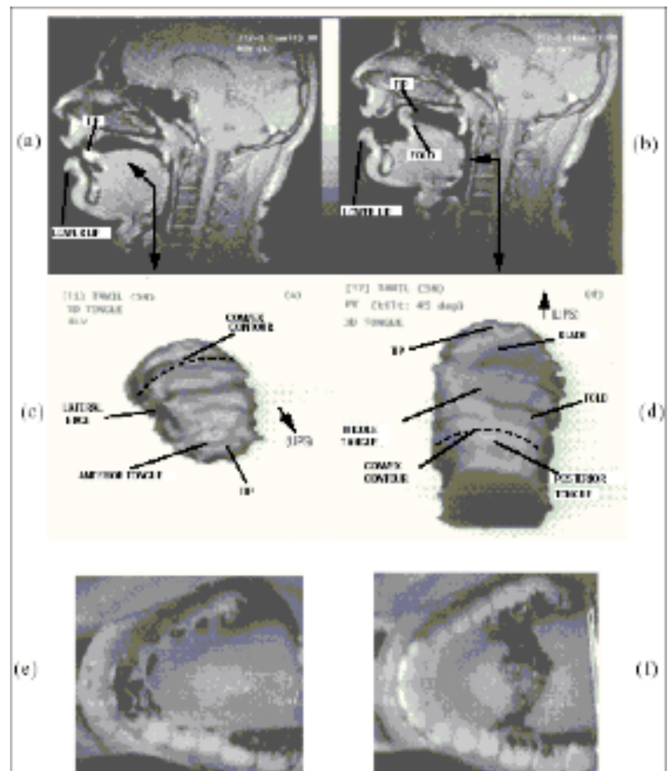
- Slow: Spatial & Temporal resolution tradeoffs, optimizing to a given application
- Noisy images: Susceptibility, blurring artifacts
- Imaging teeth
- Interaction with other physiological activities: respiration, swallowing, ..
- Clean, Synchronized audio (and other modalities, as needed)
- Ease of experimentation, including cost and portability

Early use of MRI: Static Vocal tract Information

- Information on 3D vocal tract structure/shape, inter-speaker variations
- Accurate vocal tract measurements: area functions, length
- Detailed studies on vowels and a number of continuant speech sounds
- Facilitated new speech modeling studies
  - *Vowels, Nasals, Fricatives, Liquids: in English and other languages*

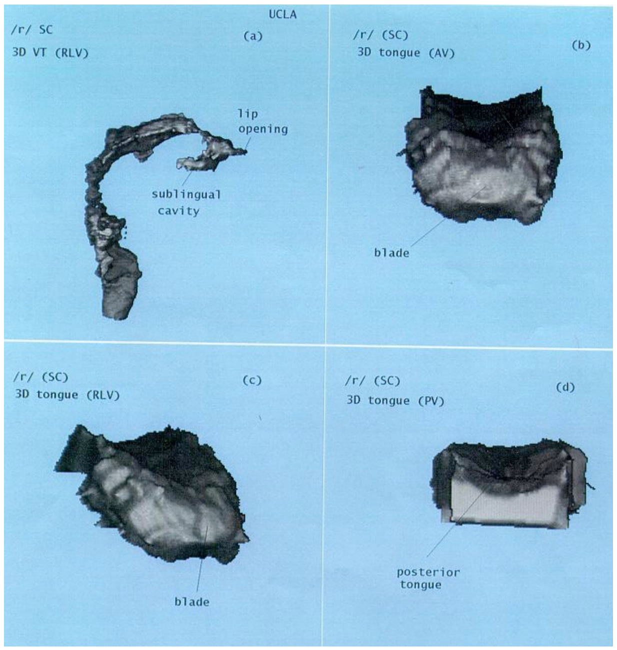
Midsagittal vowel images from Haskins (from Goldstein)

3D airway reconstruction for vowel /a/, Univ. Iowa (Story) http:everest.radiology.uiowa.edu/nlm/app/vocal/vocal.html

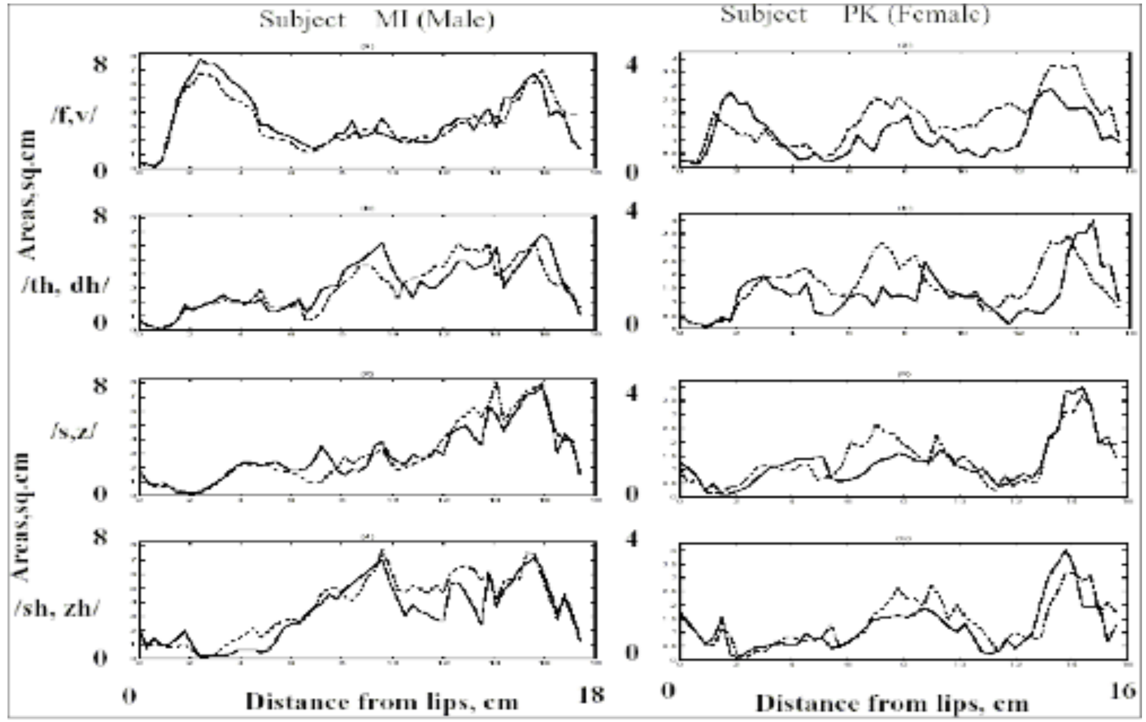


Narayanan, S., Byrd, D., and Kaun, A. **Geometry, kinematics, and acoustics of Tamil liquid consonants.** *J. Acoust. Soc. Am.,* pp. 1993—2007. 1999

Example: modeling fricatives



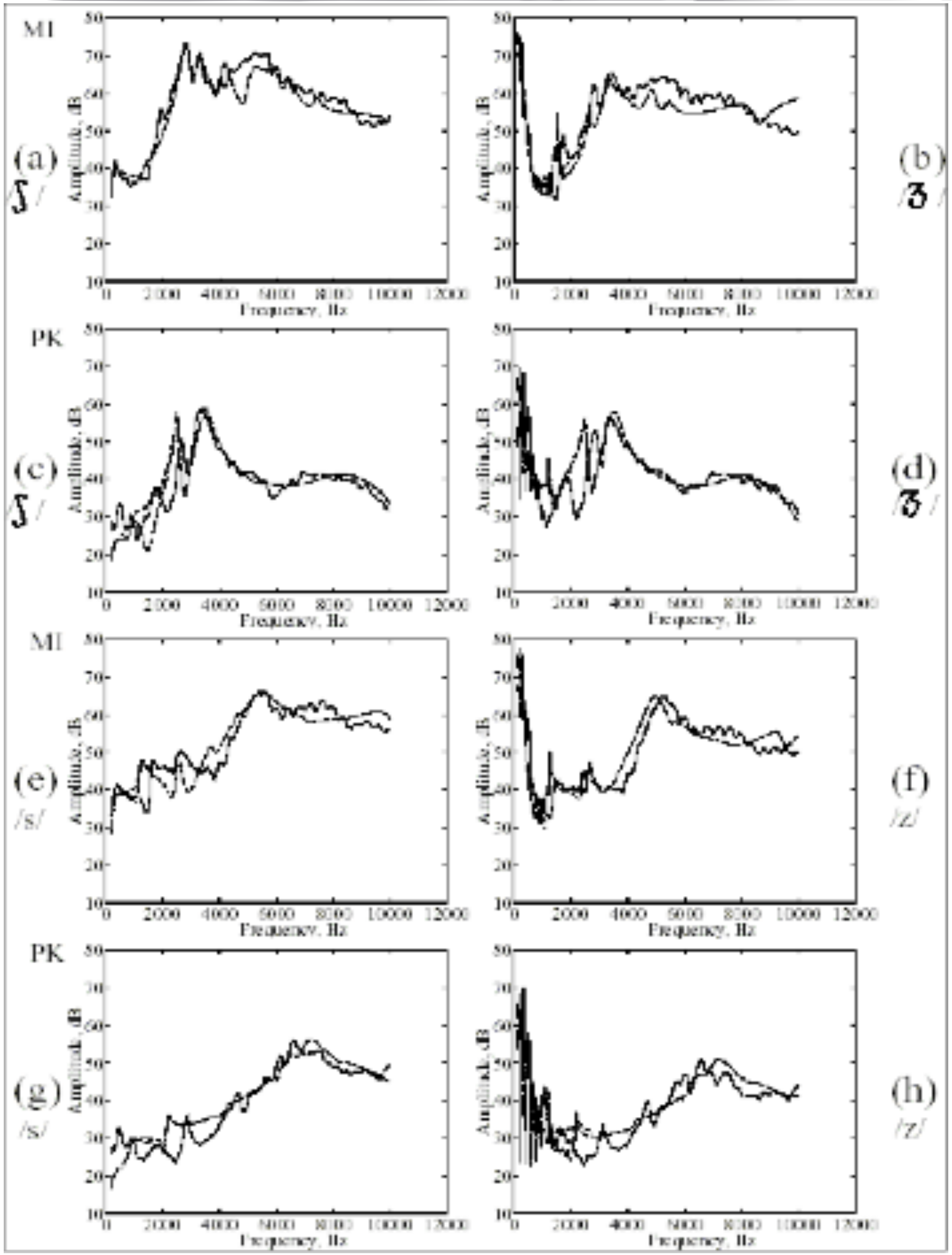
3D Vocal tract and tongue shapes for /sh/



MRI-derived area functions for fricatives

Narayanan, S., Alwan, A., and Haker, K. (**1995**), “**An articulatory study of fricative consonants using magnetic resonance imaging**,” *J. Acoust. Soc. Am.*, vol. 98, pp. 1325—1347.

Support modeling studies: e.g. Fricatives



Strident fricative spectra derived from hybrid source model inputs using the parametric dipole spectra: dashed (model); solid (natural speech)

Narayanan, S., and Alwan, A. **Noise source models for fricative consonants.** *IEEE Trans. Speech and Audio Processing*, vol. 8, no. 3, pp. 328—344. 2000

BUT:

Limited to snapshot postures of a dynamic phenomenon.

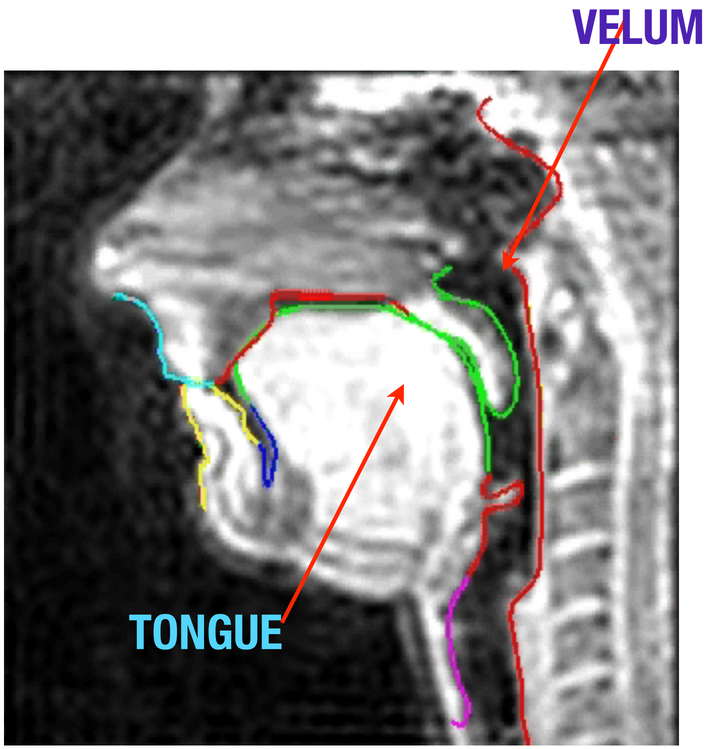
PS: It took 10-15 seconds to acquire a single image slice!

MRI: Toward real time acquisition for speech (2003)

Improving MRI temporal resolution

- A non 2D-FFT acquisition strategy (*spiral k-space trajectory*) on a GE Signa 1.5T CV/i scanner with a low-flip angle spiral gradient echo, 9-10 images/second
- Adapted pulse sequence originally developed for cardiac imaging
- Effective reconstruction rates of 24-35 frames/second
  - sliding window reconstruction technique

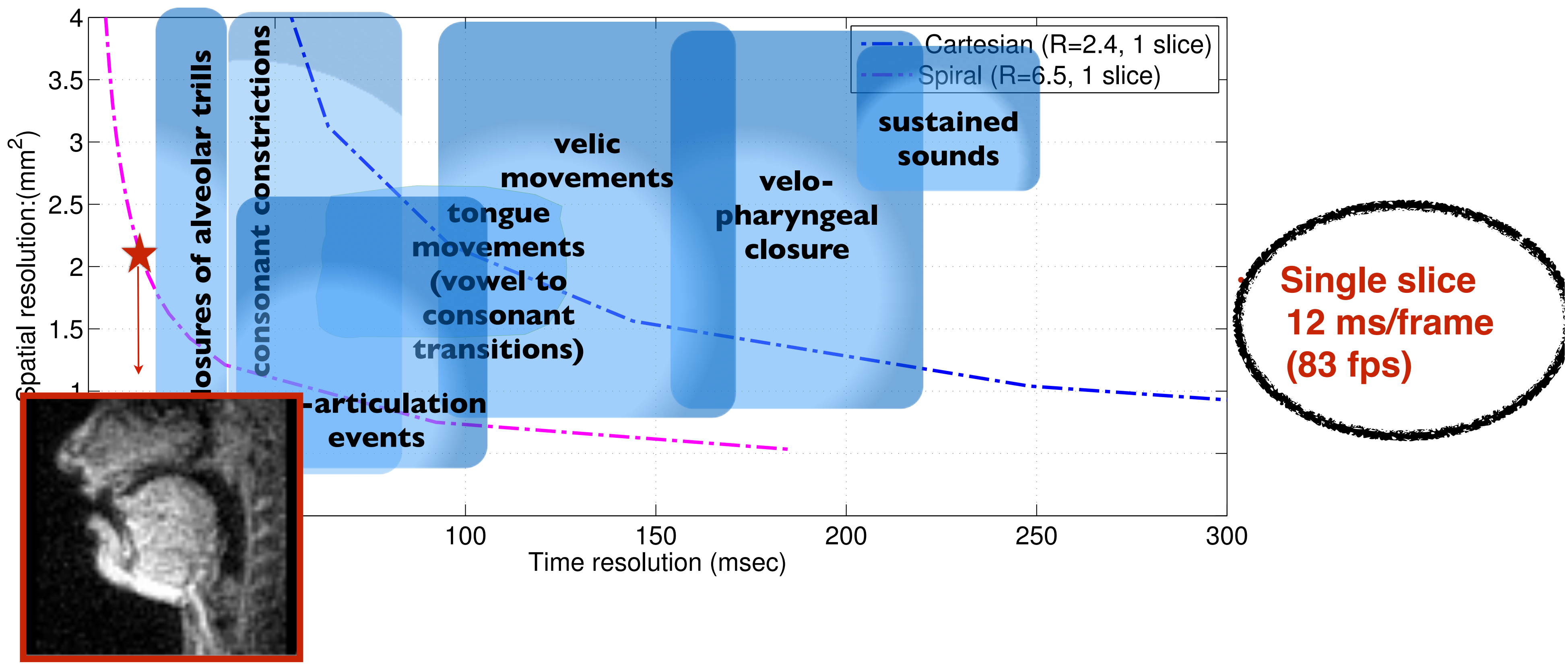
First to use real-time MRI and synchronous noise-cancelled audio to understand vocal tract movements during natural speech production.



S. Narayanan, K. Nayak, S. Lee, A. Sethy, and D. Byrd. **An approach to real-time magnetic resonance imaging for speech production**. *J. Acoust. Soc. Am.*, 115:1771-1776, 2004.

Spatial vs. Time resolution: speech MRI

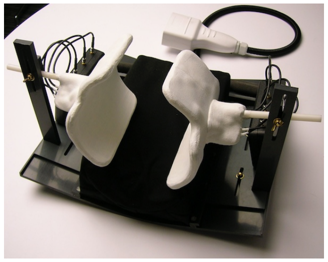
- Our newer system (**circa 2015**) enables visualization of all speech tasks

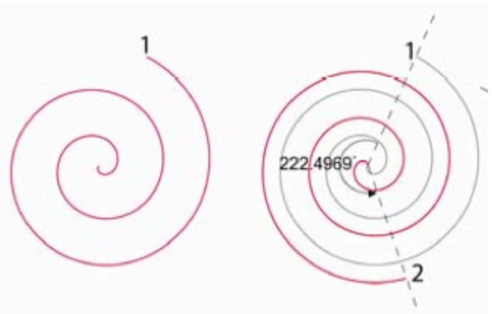


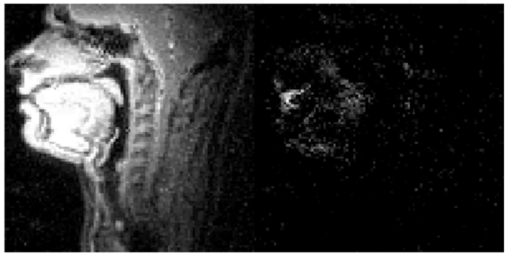
S. Lingala, Y. Zhu, Y-C. Kim, A. Toutios, S. Narayanan, K. Nayak. A fast and flexible MRI system for the study of dynamic vocal tract shaping. *Magnetic Resonance in Medicine*. 77(1): 112-125, 2017

How?

- **Highly accelerated RT-MRI of speech is achieved by synergistic engineering advances**
  - Novel custom upper-airway coil design
  - Parallel imaging
  - Fast spiral readouts
  - Constrained reconstruction/TT-GRAPPA
    - compressed sensing ideas







S. Lingala, Y. Zhu, Y-C. Kim, A. Toutios, S. Narayanan, K. Nayak. A fast and flexible MRI system for the study of dynamic vocal tract shaping. Magnetic Resonance in Medicine. 77(1): 112-125, 2017

Real-time MRI at 83 fps, 2.4 mm/pixel

A child speaker

International Phonetic Alphabet (IPA) database

http://sail.usc.edu/span/rtmri_ipa

the rtMRI IPA chart

Click on any of the red-colored speech sounds or utterances below to see their production captured with real-time MRI.

Consonants (Pulmonic)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Consonants (Non-Pulmonic)

Clicks

ʘ Bilabial
ǀ Dental
ǃ (Post)alveolar
ǂ Palatoalveolar
ǁ Alveolar Lateral

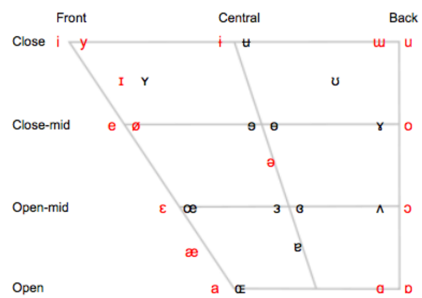
Voiced Implosives

ɓ Bilabial
ɗ Dental/Alveolar
ʄ Palatal
ɠ Velar
ʛ Uvular

Ejectives

p' Bilabial
t' Dental/Alveolar
k' Velar
s' Alveolar Fricative

Vowels



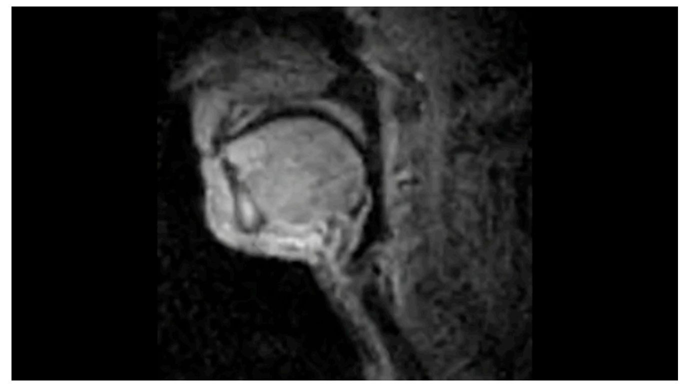
Other Symbols

ʍ Voiceless labial-velar fricative
w Voiced labial-velar approximant
ɥ Voiced labial-palatal approximant
ʜ Voiceless epiglottal fricative
ʢ Voiced epiglottal fricative
ʡ Epiglottal plosive
ɕ ʑ Alveolo-palatal fricatives
ɺ Voiced alveolar lateral flap
ɧ Simultaneous ʃ and x
ŋ! Alveolar nasal click
ts dʒ Affricates
k͡p (double articulation)

Words, Sentences and Passages

heed, hid, hayed, head, had, hod, howed, hood, hoed, who'd, hud, hide, how'd, hoy'd, hued
bead, bid, bayed, bed, bad, bod, bawed, bode, booed, bud, bide, bowed, Boyd, byued
beet, bit, bait, bet, bat, pot, but, bought, boat, boot, put, bite, bird, abbot, bute

"She had your dark suit...", "Don't ask me to carry...", "The girl was thirsty...", "Your good pants..."
Rainbow Passage, Grandfather Passage

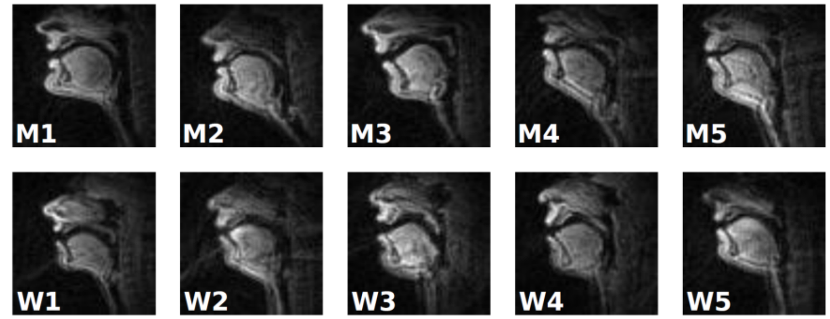


fricatives

A. Toutios, S. Lingala, C. Vaz, J. Kim, J. Esling, P. Keating, M. Gordon, D. Byrd, L. Goldstein, K. Nayak, and S. Narayanan, "Illustrating the Production of the International Phonetic Alphabet Sounds using Fast Real-Time Magnetic Resonance Imaging," in *Proc. Interspeech*, 2016.

USC-TIMIT: A MULTIMODAL ARTICULATORY DATA CORPUS

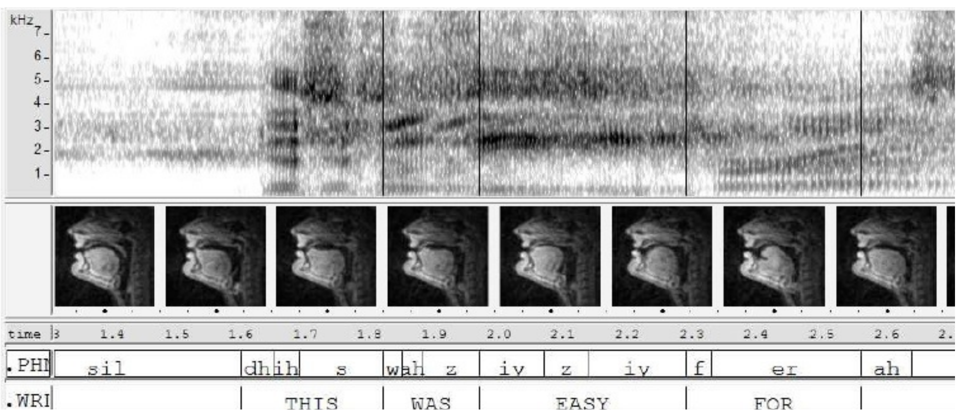
- 10 American English talkers (5M, 5F).
- Real time MRI (5 speakers also with EMA) and synchronized audio.
- 460 sentences each (>20 minutes)
- Freely available for speech research.

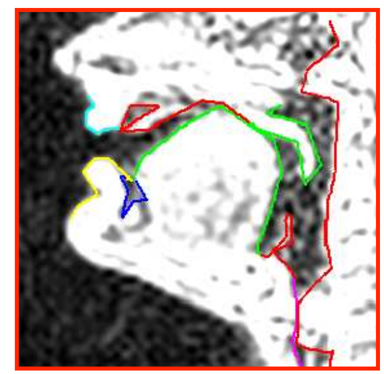


WEB-LINK (with download info):

http://sail.usc.edu/span/usc-timit/

SAIL homepage: http://sail.usc.edu



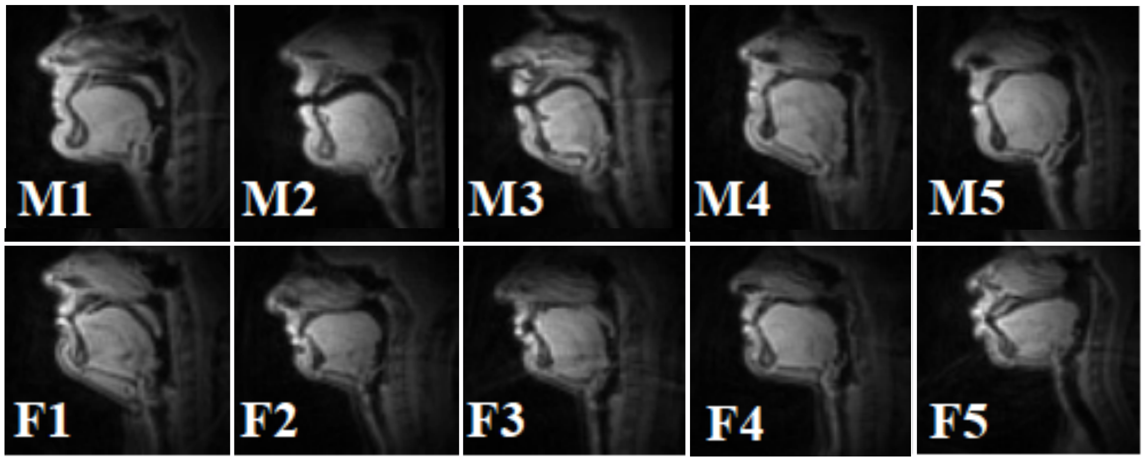


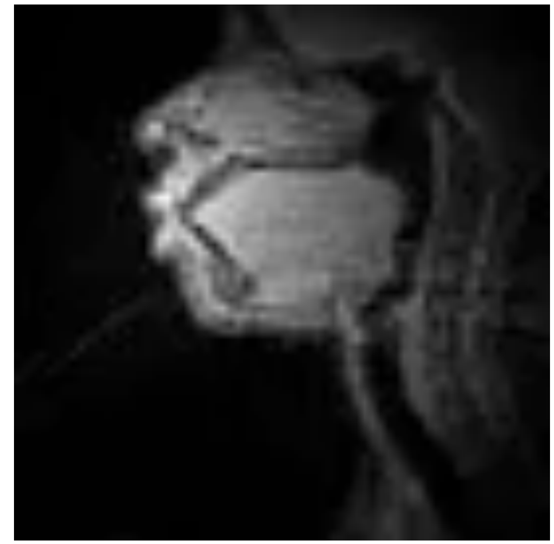
- Narayanan et al. A Multimodal Real-Time MRI Articulatory Corpus for Speech Research. *InterSpeech*. 2011
- Narayanan et al. Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research. *J. Acoust. Soc. Am*. 136(3): 1307-1311. 2014

USC-EMO-MRI corpus

https://sail.usc.edu/span/usc-emo-mri/index.html

A multimodal dataset for emotional speech production

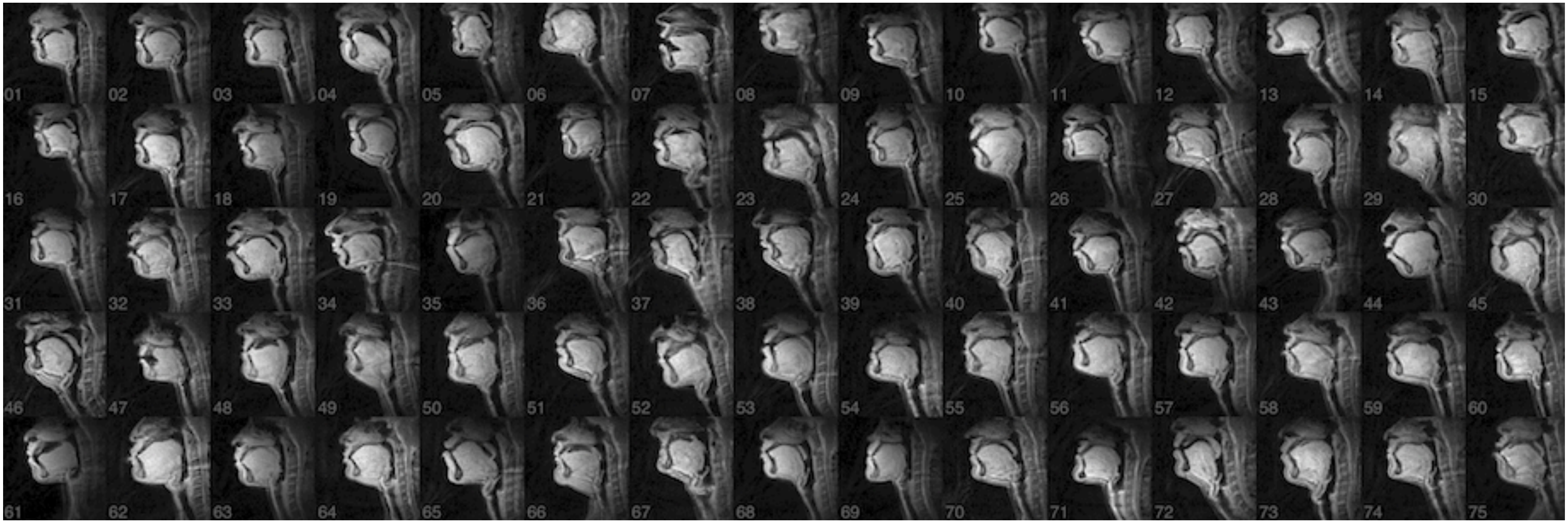




- MRI video (23.180 fps) + speech audio (20kHz)
- The "Grandfather passage" and 6-7 sentences
- 4 acted emotions (neutral, angry, happy and sad)
- Emotion quality evaluation (at least 11 evaluators)

Jangwon Kim et al., "USC-EMO-MRI corpus: An emotional speech production database recorded by real-time magnetic resonance imaging", in 10-th ISSP, 2014
Jangwon Kim, Asterios Toutios, Sungbok Lee, Shrikanth Narayanan. Vocal tract shaping of emotional speech. Computer Speech & Language. 64: 101100, 2020

75-Speaker Speech MRI Database



2D realtime MRI videos with audio for variety of speech tasks, 3D volumetric data for sustained sounds, high-resolution static anatomical T2-weighted upper airway MRI images includes Raw RT-MRI data

Yongwan Lim, Asterios Toutios, Yannick Bliesener, Ye Tian, Sajan Goud Lingala, Colin Vaz, Tanner Sorensen, Miran Oh, Sarah Harper, Weiyi Chen, Yoonjeong Lee, Johannes Töger, Mairym Lloréns Montesserin, Caitlin Smith, Bianca Godinez, Louis Goldstein, Dani Byrd, Krishna S. Nayak, Shrikanth S. Narayanan. **A multispeaker dataset of raw and reconstructed speech production real-time MRI video and 3D volumetric images**. *Scientific Data* 8, 187, 2021

Access datasets/tools: https://sail.usc.edu/span/resources.html

USC Dynamic Imaging Science Center (DISC)

circa ~2021

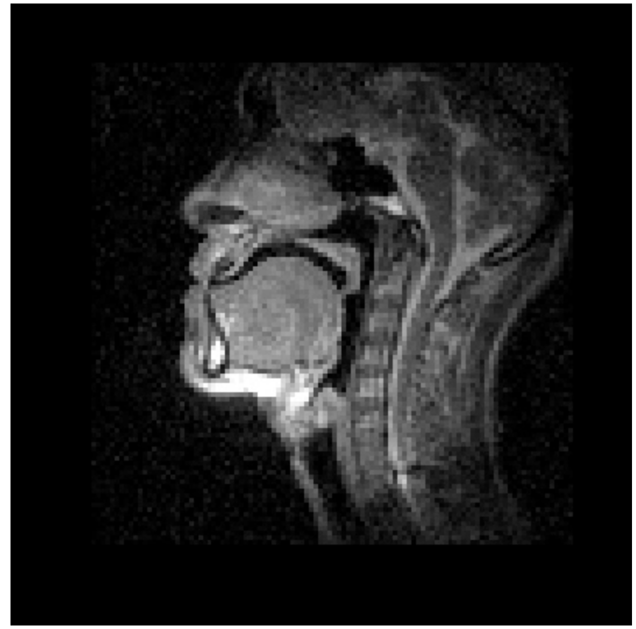
End-to-End Pipeline: from multimodal imaging to informatics through advances in signal processing and AI

DYNAMIC
IMAGING
SCIENCE
CENTER
disc.usc.edu
ESTABLISHED
2021

Developing imaging technology to better understand human health and disease especially during movement

disc.usc.edu

Custom High Performance Low Field System



- ~100 frames/second
- fewer imaging artifacts at air-tissue interfaces
- quieter audio environment (=> better quality speech recording)
- relaxed off-resonance constraints of HPLF enables prolongation of repetition time (TR) and improved sampling efficiency
- SAR (tissue heating) constraints present at ≥1.5T are virtually eliminated

Analysis and modeling of data

inspired new analytical and modeling approaches,
leading to new scientific and theoretical insights

- **Image analysis**
- **Deriving**
  - morphological (structural) details, and
  - linguistically meaningful articulatory features

- **Some case studies**
  - Characterizing vocal tract morphology
  - Example linguistic and paralinguistic analyses
  - Relation between articulatory & acoustic representations
  - Automatic Speech/Speaker/Emotion recognition
  - Articulatory strategies

Vikram Ramanarayanan, Sam Tilsen, Michael Proctor, Johannes Töger, Louis Goldstein, Krishna Nayak, Shrikanth Narayanan. Analysis of Speech Production Real-Time MRI. *Computer Speech & Language*. 52:-1-22, 2018

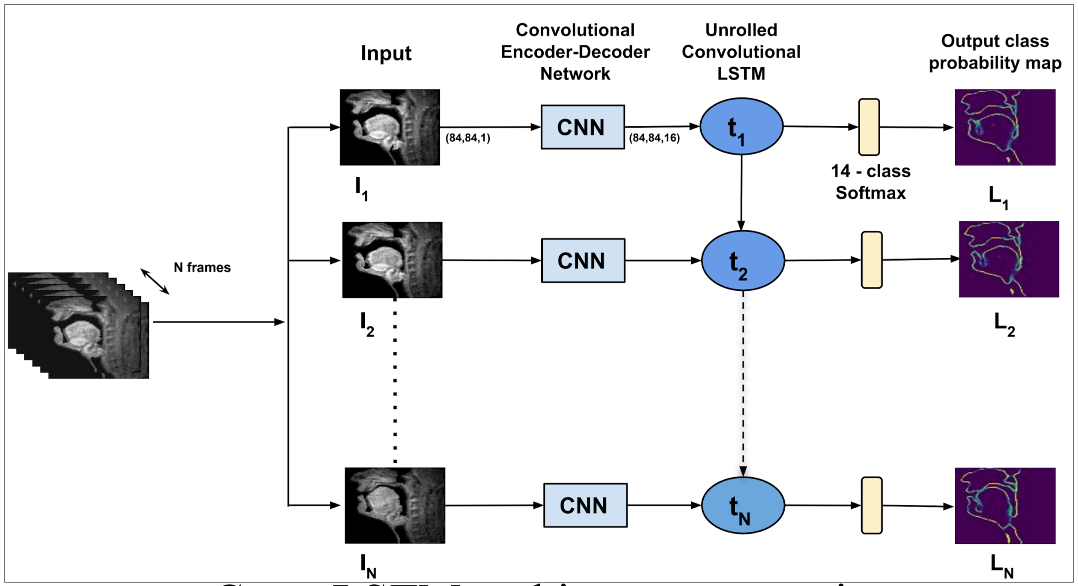
Articulator Tracking: Segmenting Vocal tract contours

Model-Based Image Segmentation In The Fourier Domain

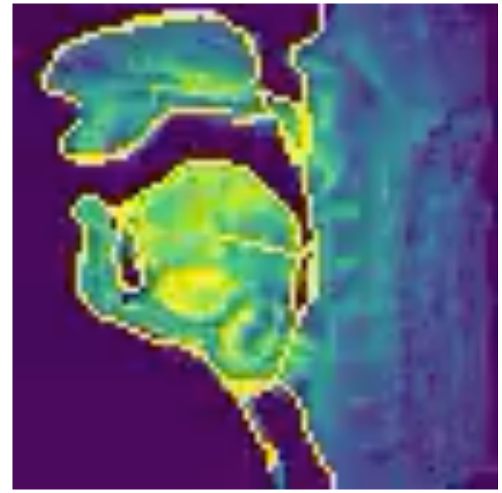
hierarchically optimize to model fit to the image in the Fourier domain using **gradient descent**

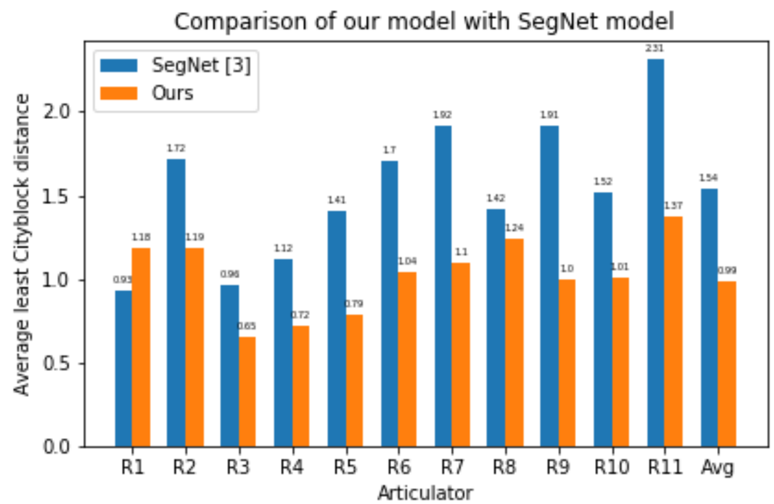
Erik Bresch and Shrikanth Narayanan. **Region segmentation in the frequency domain applied to upper airway real-time magnetic resonance images**. *IEEE Transactions on Medical Imaging*. 28(3): 323--338, March 2009.

Vocal tract articulatory contour detection using spatio temporal context



Conv-LSTM architecture capturing spatio-temporal context

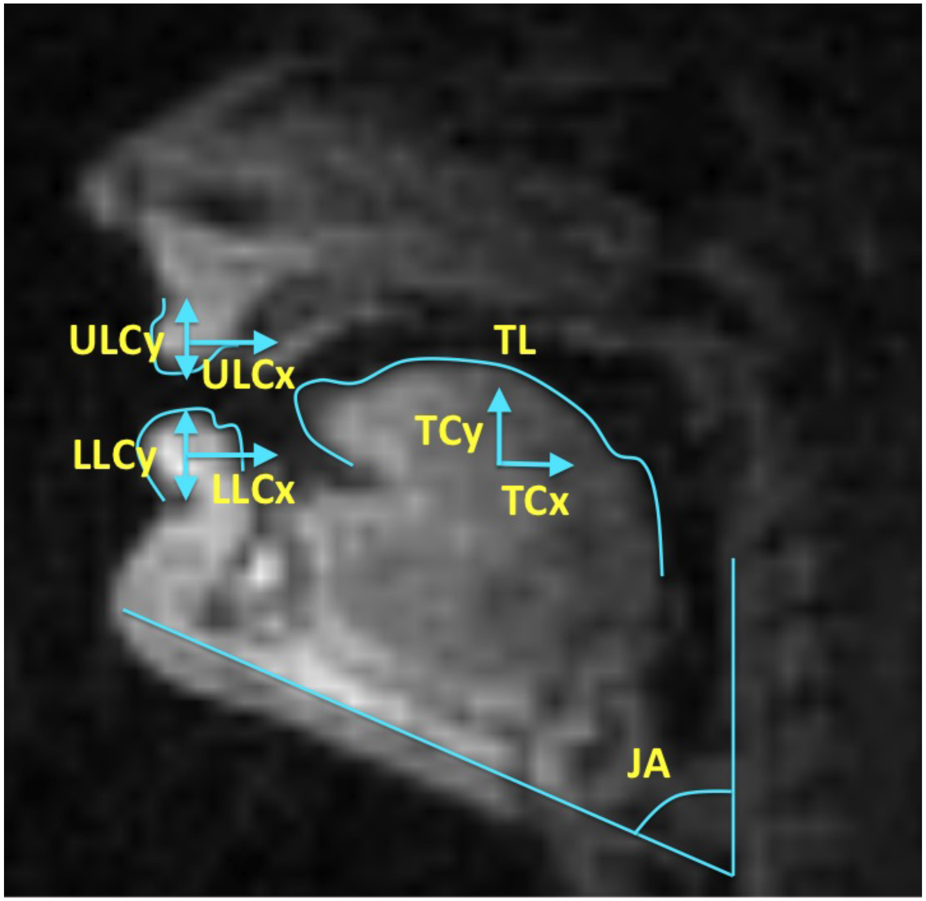


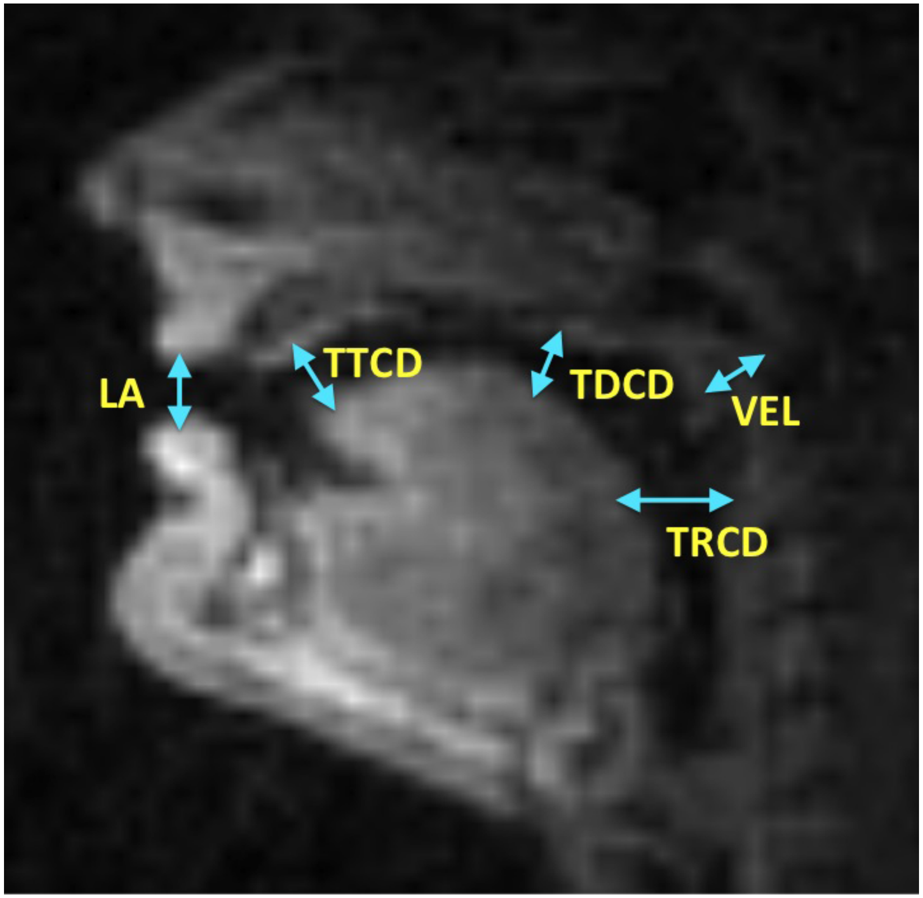


Performance comparison against the image based model

S Ashwin Hebbar, Rahul Sharma, Krishna Somandepalli, Asterios Toutios, Shrikanth Narayanan. VOCAL TRACT ARTICULATORY CONTOUR DETECTION IN REAL-TIME MAGNETIC RESONANCE IMAGES USING SPATIO-TEMPORAL CONTEXT ICASSP, 2020

Articulatory Posture & Constriction Task Variables

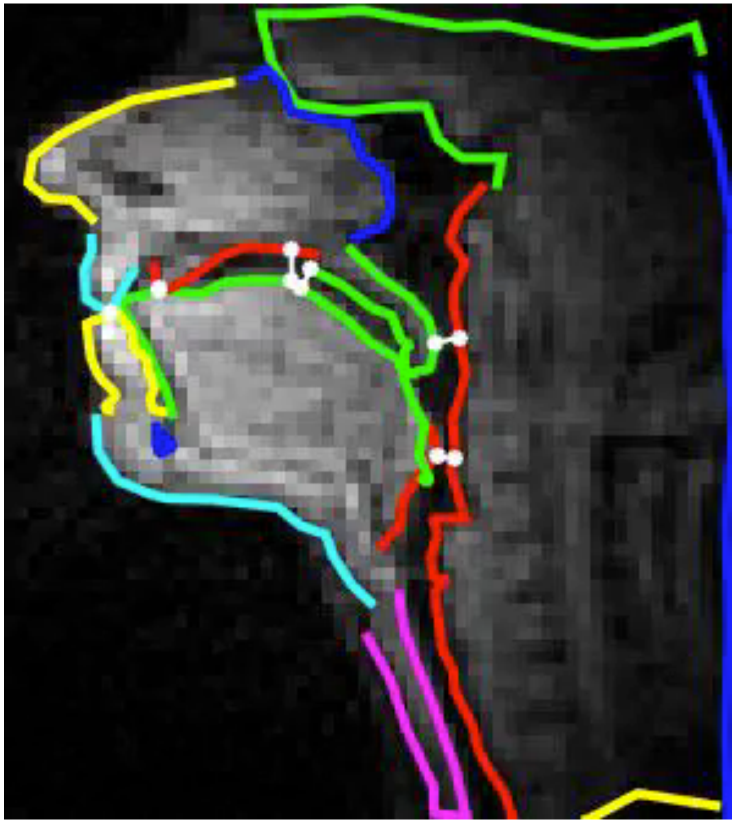




These feature sets are useful for modeling speech production dynamics

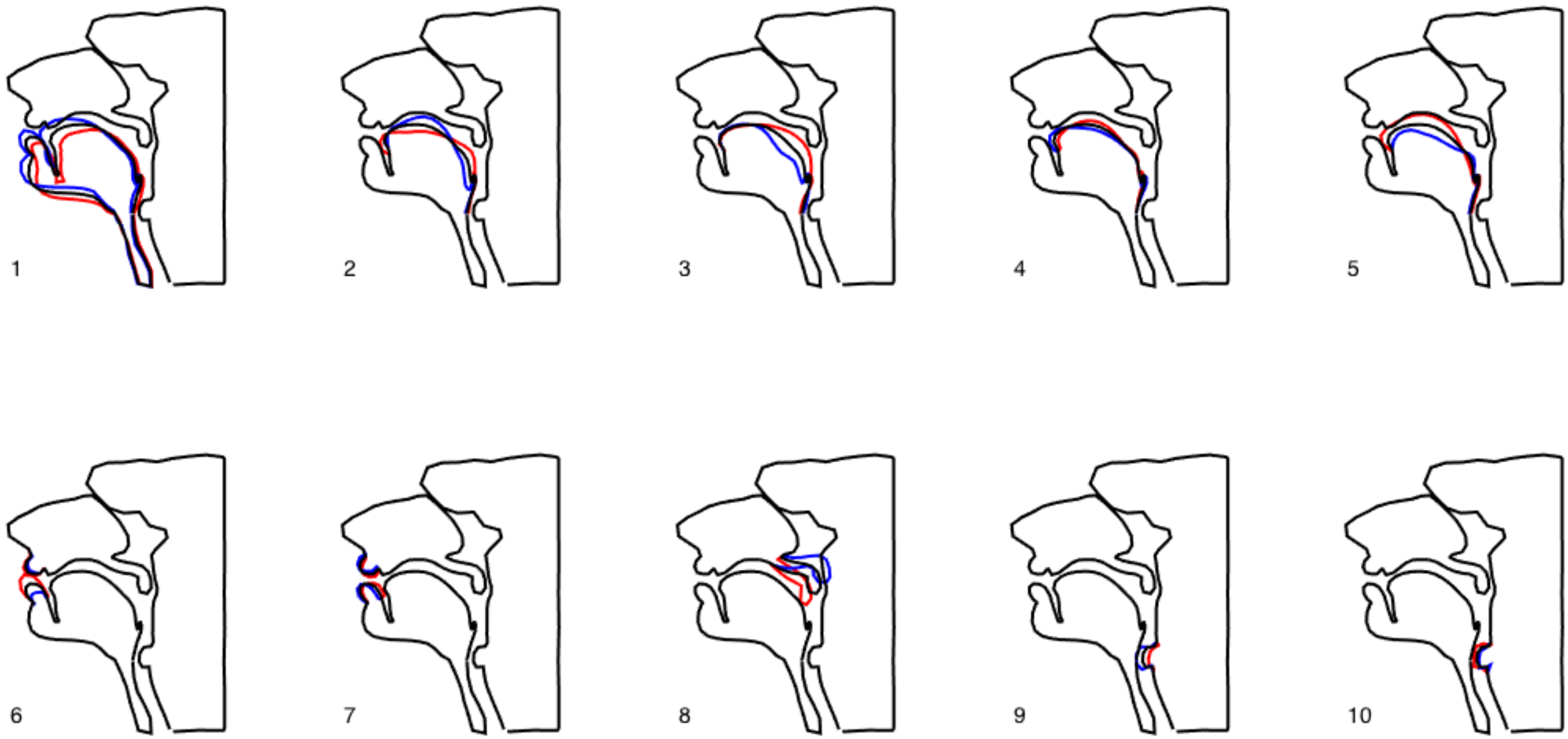
- Adam Lammert, Louis Goldstein, Shrikanth Narayanan and Khalil Iskarous. **Statistical Methods for Estimation of Direct and Differential Kinematics of the Vocal Tract**. *Speech Communication*. 55: 147–161, 2013.
- Vikram Ramanarayanan, Adam Lammert, Louis Goldstein, Shrikanth Narayanan, **Are Articulatory Settings Mechanically Advantageous for Speech Motor Control?**, *PLoS ONE*, vol. 9, no. 8, pp. e104168, 2014.

Tracking Constriction Variables



- Sorensen, T., Toutios, A., Goldstein, L., & Narayanan, S. **Characterizing vocal tract dynamics with real-time MRI**, ***Conference on Laboratory Phonology***, Ithaca, NY. 2016
- Vikram Ramanarayanan, Louis Goldstein, Dani Byrd and Shrikanth S. Narayanan, **An investigation of articulatory setting using real-time magnetic resonance imaging** *J. Acoust. Soc. Am.*, 134:1(510-519), 2013

Speaker-Specific Articulatory Models

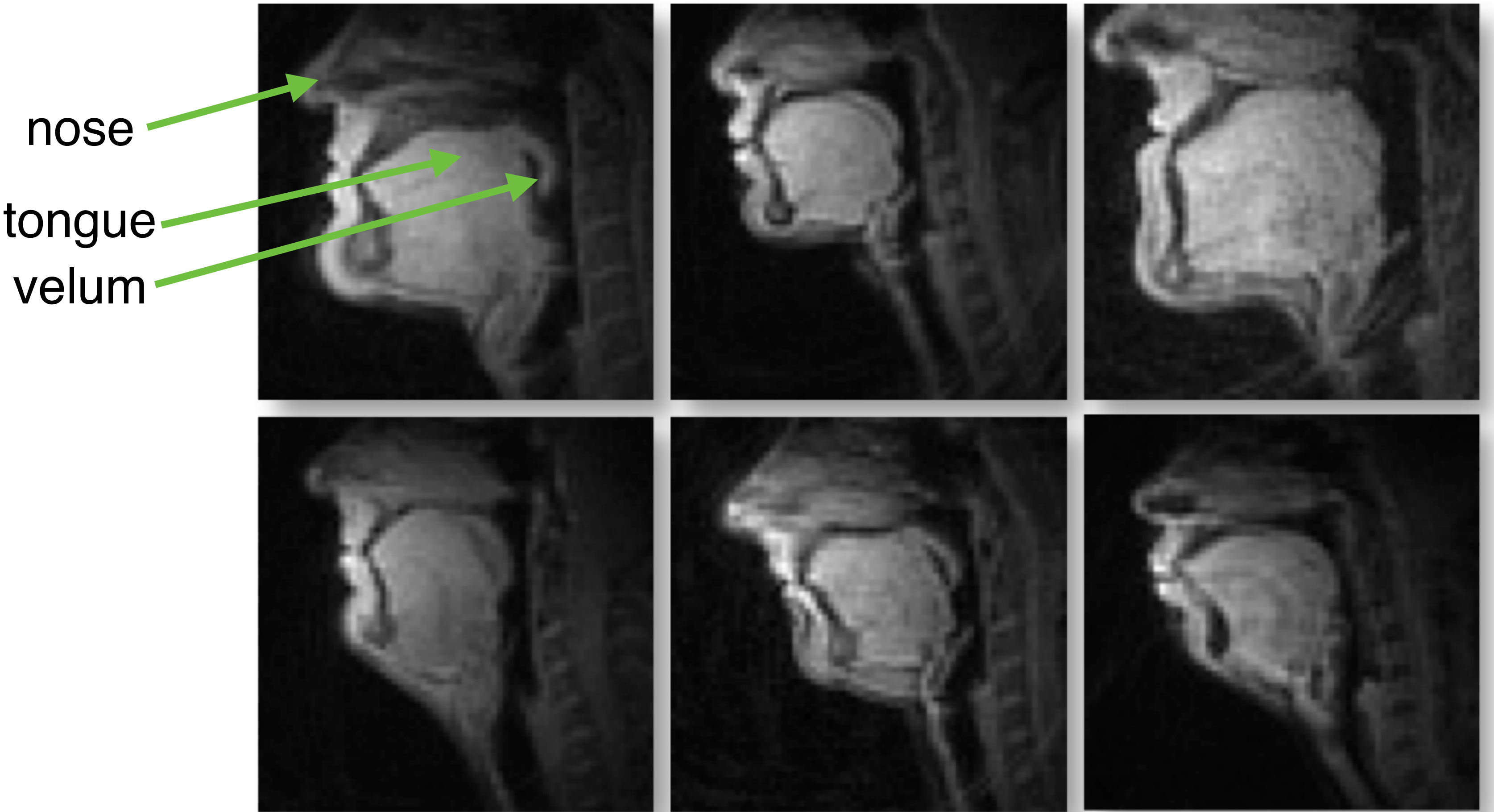


- Toutios, A., & Narayanan, S. S. **Factor analysis of vocal-tract outlines derived from real-time magnetic resonance imaging data**. Proc. International Congress of Phonetic Sciences, 2015

Analysis of data

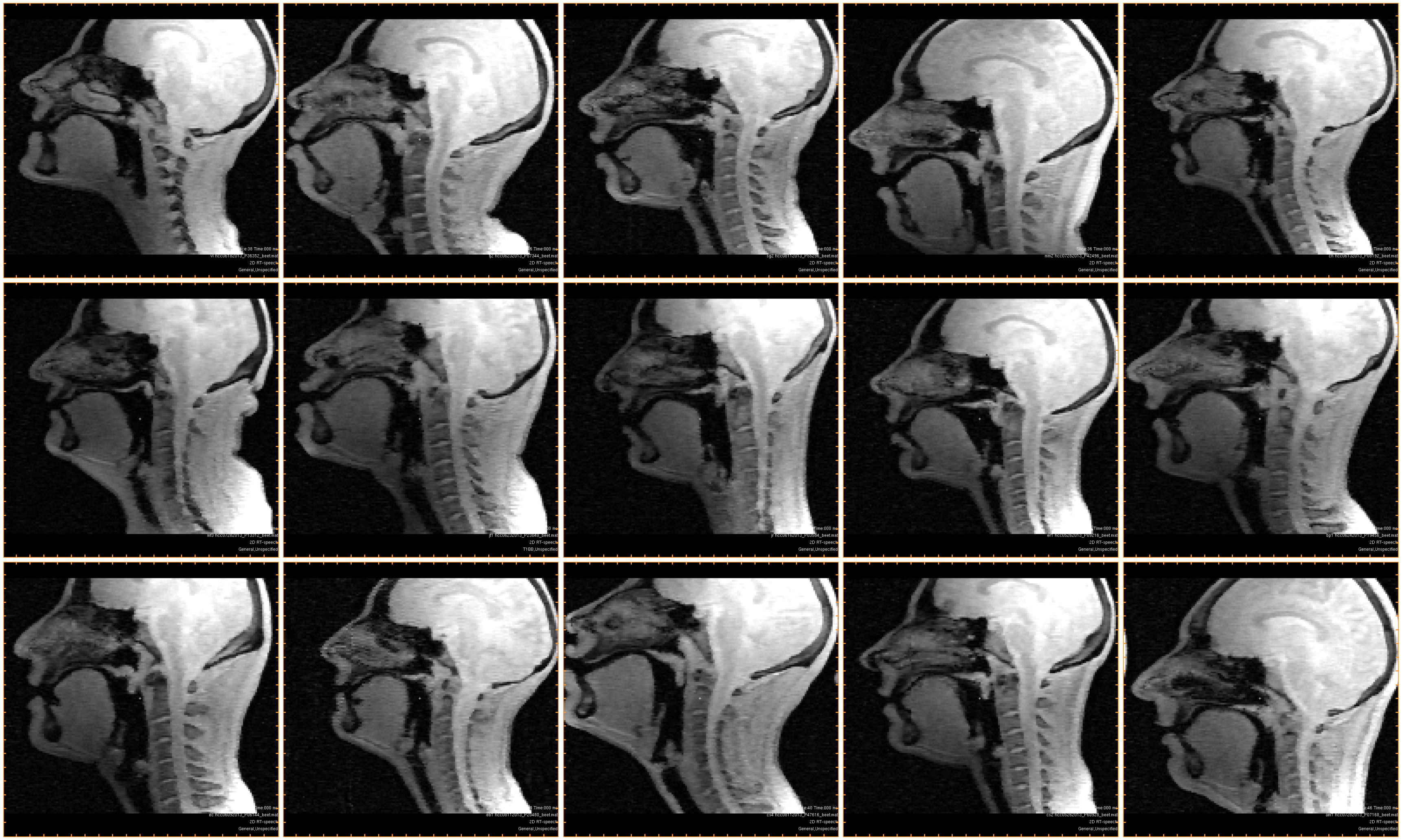
- Image analysis
- Deriving
  - morphological (structural) details, and
  - linguistically meaningful articulatory features
- Some case studies
  - Vocal tract morphology
  - Example linguistic and paralinguistic analyses
  - Relation between articulatory & acoustic representations
  - ASR and Speaker Verification
  - Articulatory strategies

Different individuals….



…each with a uniquely shaped vocal instrument

And with differing articulatory strategies during speech



Fifteen different individuals producing vowel /i/

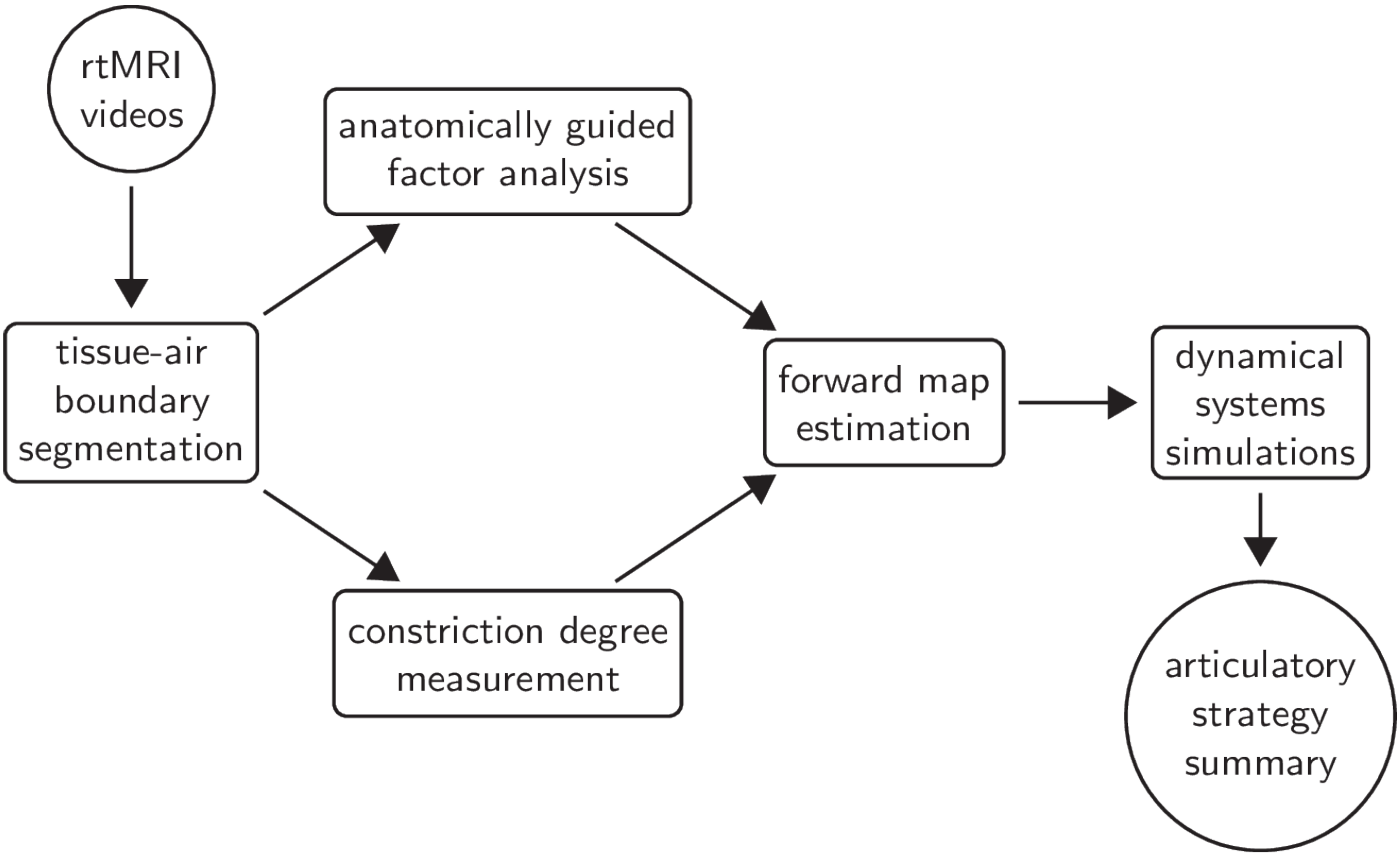
Modeling production mechanisms

Articulatory synergies

how talkers use their articulatory organs

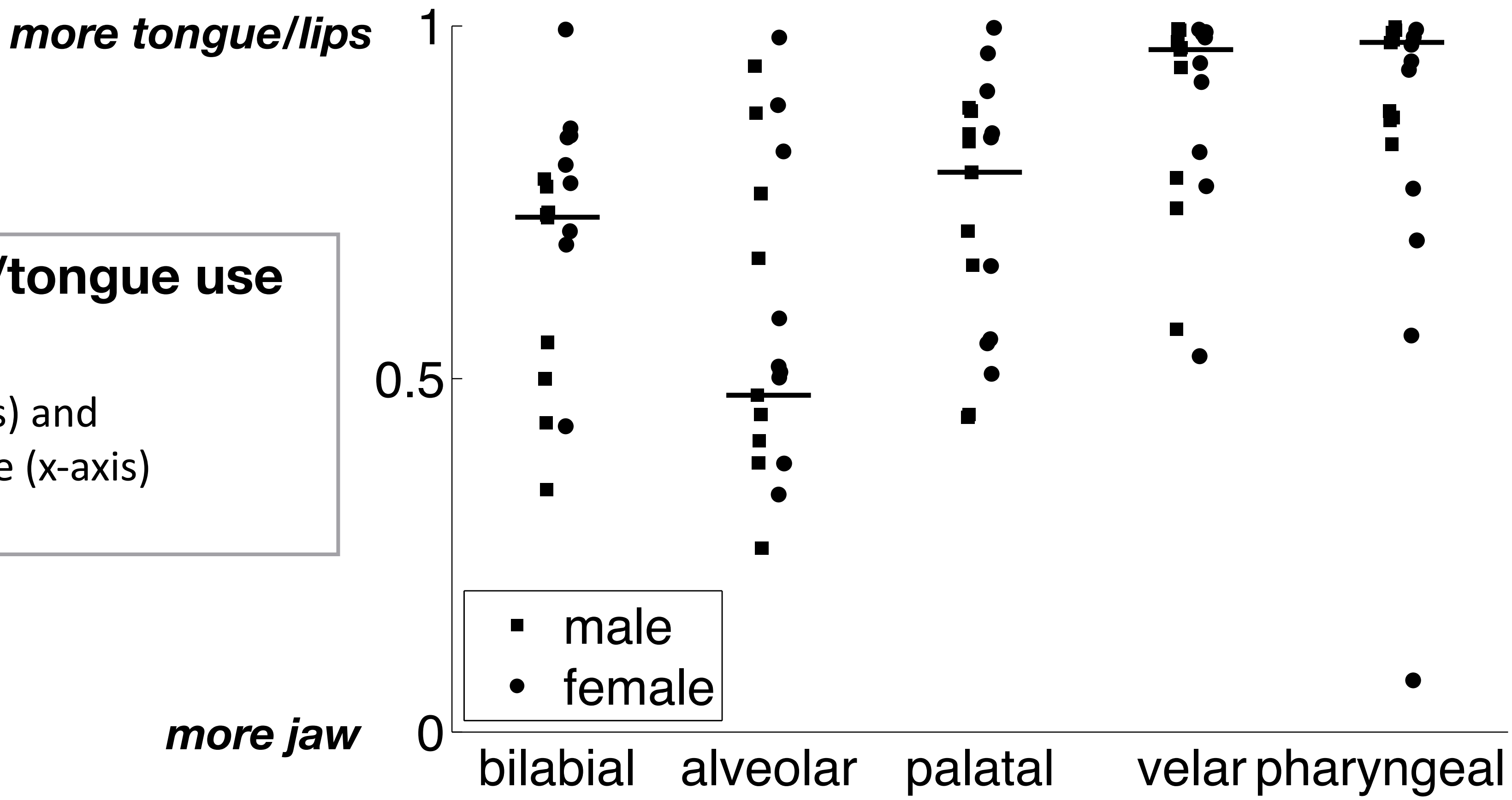
- Vocal tract is a redundant system
- Articulators have overlapping functions
  - e.g., *both* jaw and lips contribute to bilabial constrictions
- Speakers have several ways to change airway shape to make a constriction
- We call these **articulatory synergies**

Quantifying Individual Articulatory Synergy



44

Articulatory strategies across speakers



ratio of lips/tongue use to jaw use

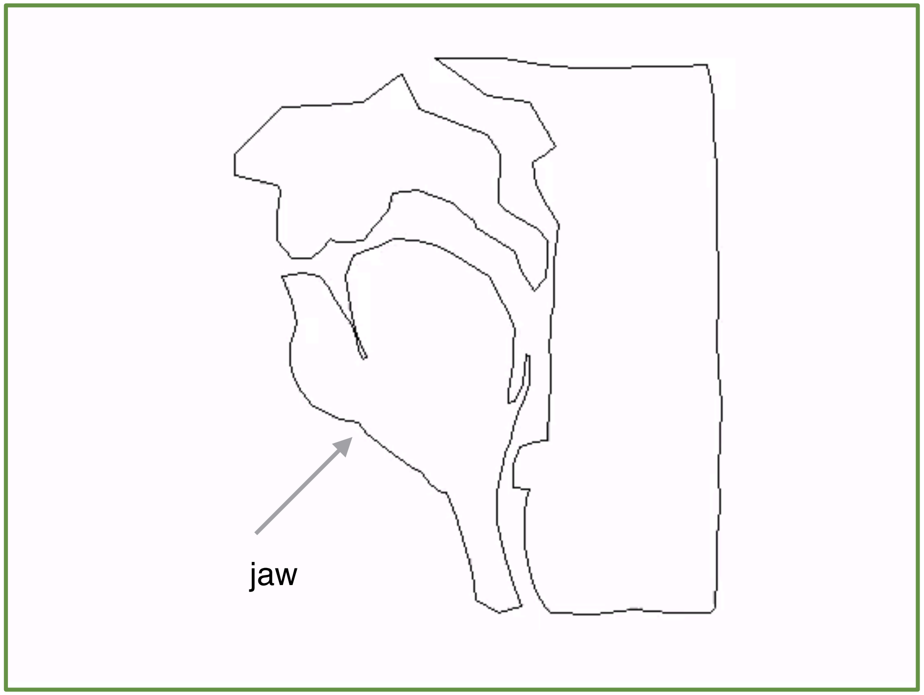
by speaker (dots) and constriction type (x-axis)

insights using data from 18 speakers (9F, 9M) in task dynamic simulation

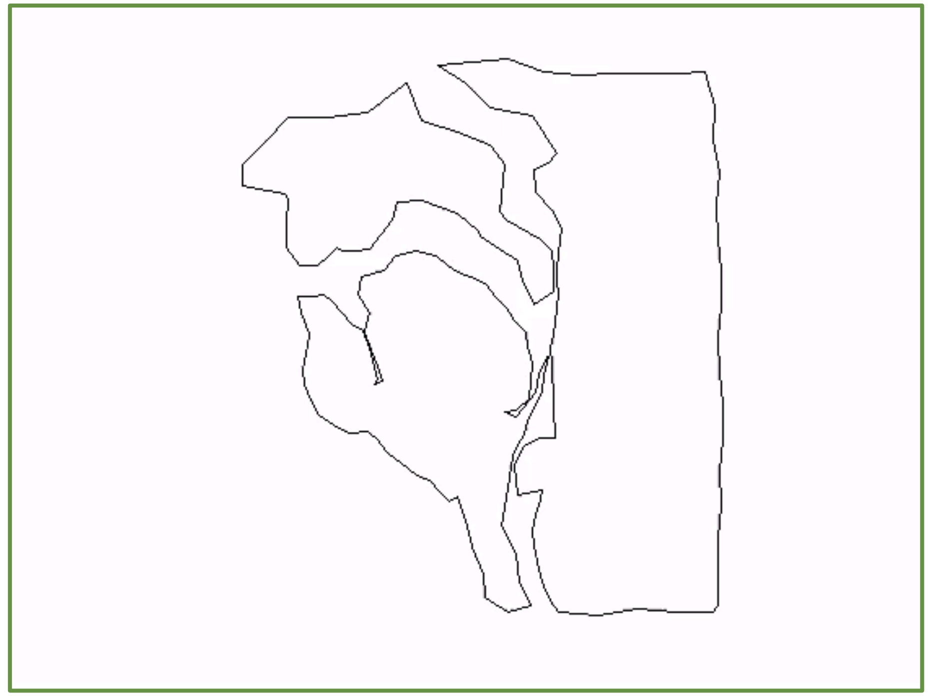
Tanner Sorensen, Asterios Toutios, Louis Goldstein, Shrikanth Narayanan. **Characterizing vocal tract dynamics across speakers using real-time MRI.** Proc. Interspeech, 2016 [BEST STUDENT PAPER!]

Alveolar closure

small jaw movement, speaker M3



prediction

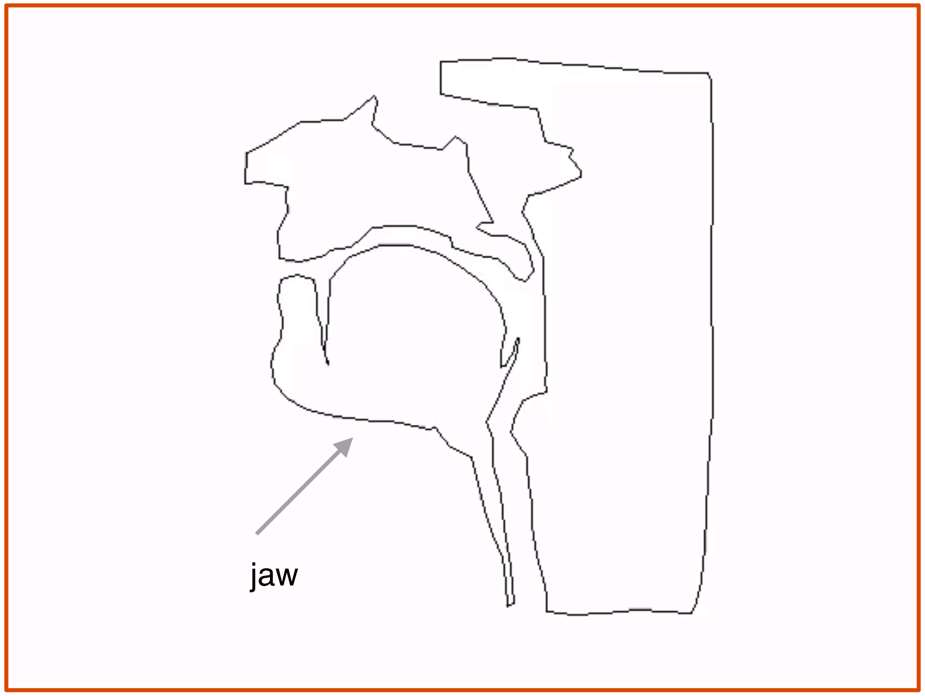


observation

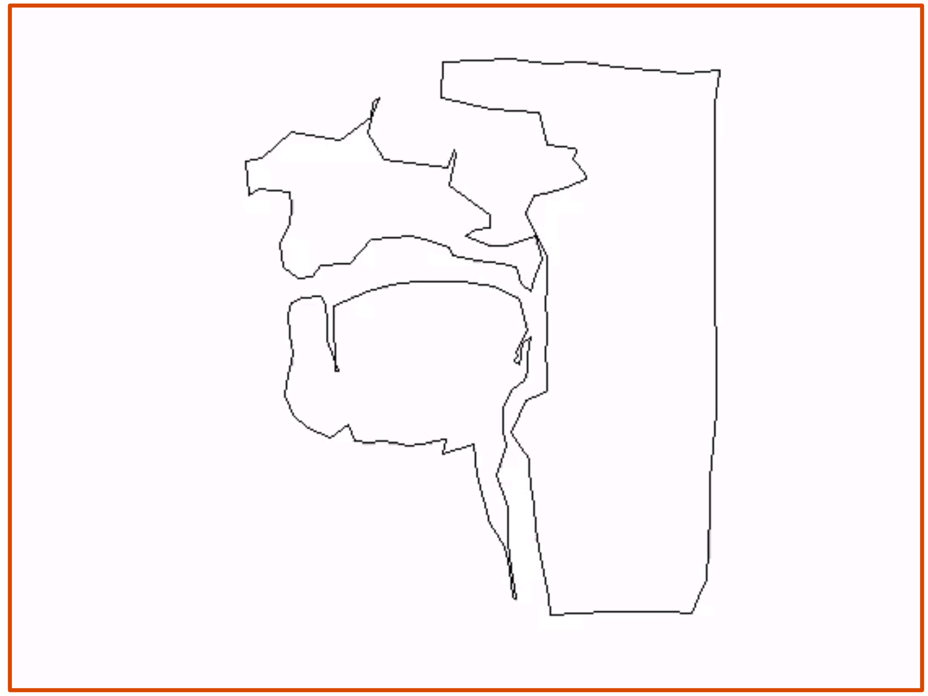
46

Alveolar closure

large jaw movement, speaker F9



prediction



observation

Quantifying Individual Articulatory Synergy

- real time MRI of the vocal tract can be used to estimate forward kinematic map
- forward kinematic map differs by speaker according to vocal tract geometry
- articulatory strategies predicted on the basis of vocal tract geometry can be compared against observed strategies

→ **tool for relating vocal tract structure and function**

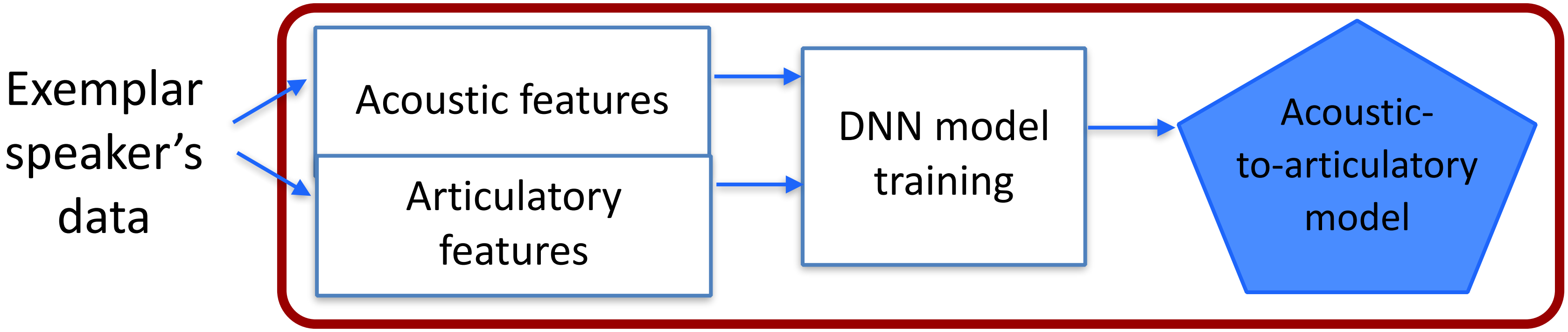
Tanner Sorensen, Asterios Toutios, Louis Goldstein, and Shrikanth Narayanan. **Task-dependence of articulator synergies.** 145(3): 1504-1520, *J. Acoust. Soc. Am.* 2019

production information for automatic speech/speaker/emotion recognition?

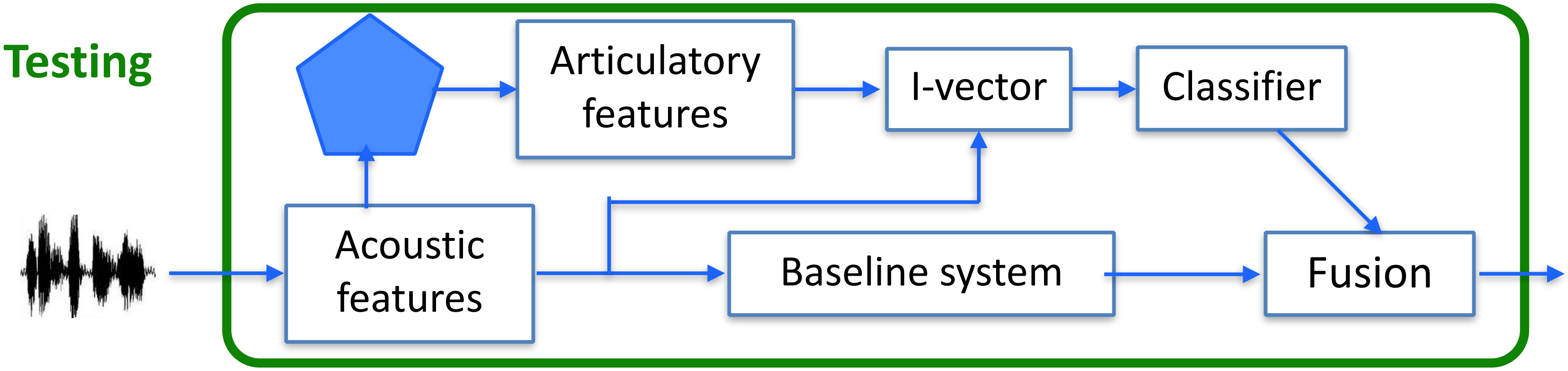
- Prasanta Ghosh and Shrikanth Narayanan. **Automatic Speech recognition using articulatory features from subject-independent acoustic-to-articulatory inversion**. *J. Acoust. Soc. Am.* 130 (4): EL251-EL257, 2011.
- Prasanta Ghosh, Louis Goldstein and Shrikanth Narayanan. **Processing speech signal using auditory-like filterbank provides least uncertainty about articulatory gestures.** *J. Acoust. Soc. Am.* 129(6): 4014-4022, 2011.
- Prasanta Ghosh, and Shrikanth Narayanan. **A generalized smoothness criterion for acoustic-to-articulatory inversion.** *J. Acoust. Soc. Am.* 128(4):2162-2172, 2010.
- Ming Li, Jangwon Kim, Adam Lammert, Prasanta Ghosh, Vikram Ramanarayanan and Shrikanth Narayanan. **Speaker verification based on the fusion of speech acoustics and inverted articulatory signals.** *Computer, Speech, and Language*. 36: 196-211, March 2016
- Jangwon Kim, Asterios Toutios, Sungbok Lee, and Shrikanth Narayanan. **Vocal tract shaping of emotional speech.** Computer, Speech and Language, 64, 2020.

Application of inversion

Inversion model training



Testing



Improving modeling in ASR, emotion recognition and speaker ID tasks

Ming Li et al., *“Speaker verification based on the fusion of speech acoustics and inverted articulatory signals"* in Computer Speech and Language, 36: 196-211, March 2016

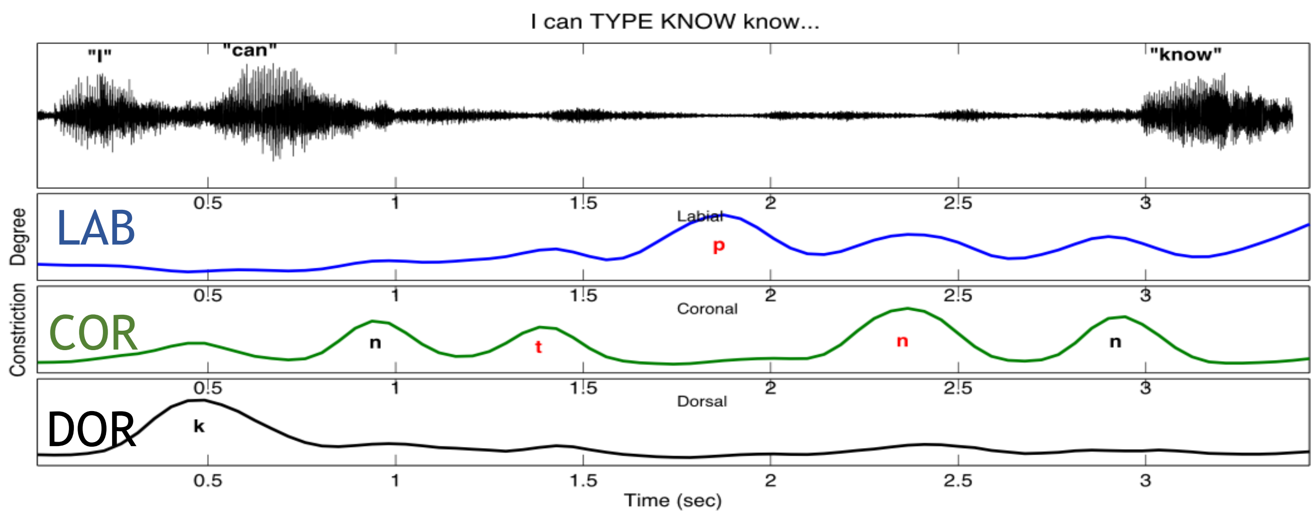
Jangwon Kim et al., *“A study of emotional information present in articulatory movements estimated using acoustic-to-articulatory inversion“* in Proceedings of APSIPA, 2012

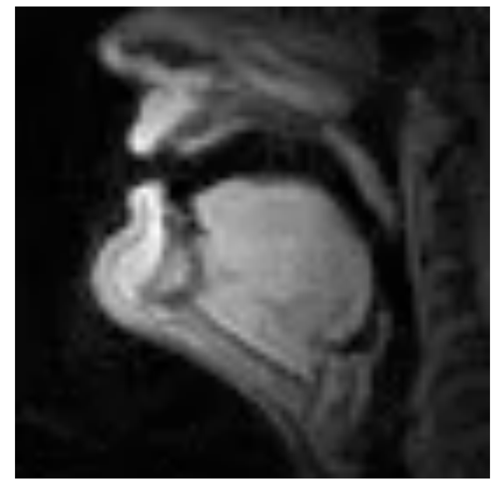
Clinical Applications: Use cases

Characterizing Articulation in Apraxic Speech

rtMRI reveals covert (unphonated) articulation of entire words

"I can TYPE KNOW know…"





Patient with primary progressive aphasia showing apraxia of speech

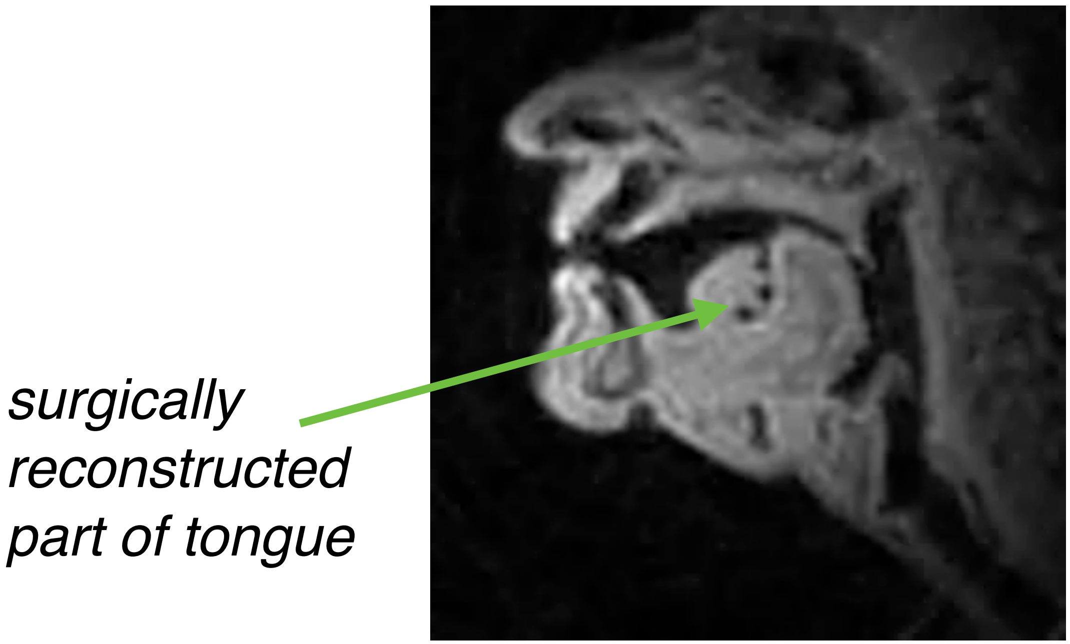
Clinical Takeaway: Apraxia of speech affects ability to select appropriate vocal tract movements for a target word/phrase and coordinate them in time, suppressing other movements. *Errors may not always be auditorily perceptible*

Christina Hagedorn, Michael Proctor, Louis Goldstein, Stephen Wilson, Bruce Miller, Maria Luisa Gorno-Tempini, and Shrikanth S. Narayanan. **"Characterizing Articulation in Apraxic Speech Using Real-time Magnetic Resonance Imaging."** *Journal of Speech, Language and Hearing Research* **(2017)**

Head and Neck Cancer

Head and neck cancer impairs speech and swallowing

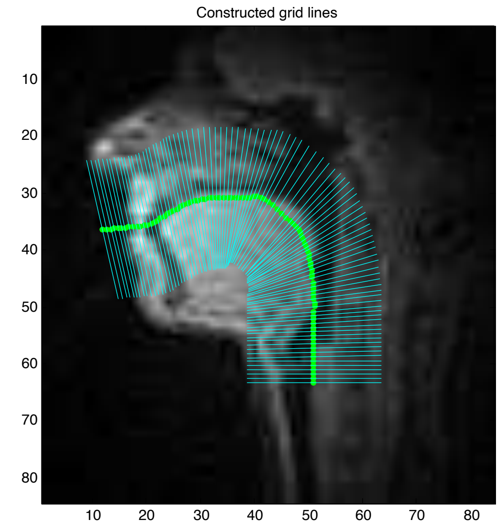
- Cancer-associated cachexia, Peripheral nerve damage
- Radiation-induced fibrosis
- Surgical treatment (*glossectomy*) effects



Christina Hagedorn, Jangwon Kim, Uttam Sinha, Louis Goldstein, Shrikanth Narayanan. **Complexity of Vocal Tract Shaping in Glossectomy Patients and Typical Speakers: A Principal Component Analysis.** *J. Acoust. Soc. Am.* 149 (6): 4437–4449, 2021

Christina Hagedorn, Yijing Lu, Asterios Toutios, Uttam Sinha, Louis Goldstein, and Shrikanth Narayanan. **Variation in compensatory strategies as a function of target constriction degree in post-glossectomy speech**. *J. Acoust. Soc. Am. Express Letters,* 2(4): 045205, 2022

Quantitatively Indexing Lingual Flexibility



How freely does the tongue move within the vocal tract?

- PCA analysis of cross-distance airway data identifies relatively few components explaining patterns of lingual displacement in patient data
- Loading plots, displaying positive and negative values of eigenvalue-scaled loading coefficients, reflect range of tongue mobility

Patient data require fewer components than do typical speakers' data to capture the same amount of variance

→ **Patients use fewer distinct vocal tract shaping patterns**

54

Hagedorn, C., Kim, J., Zu, Y., Sinha, U., Goldstein, L. & Narayanan, S. **Complexity of Vocal Tract Shaping in Glossectomy Patients and Typical Speakers: A Principal Component Analysis**, *J. Acoust. Soc. of Am*, 149(6), 4437-4449. 2021

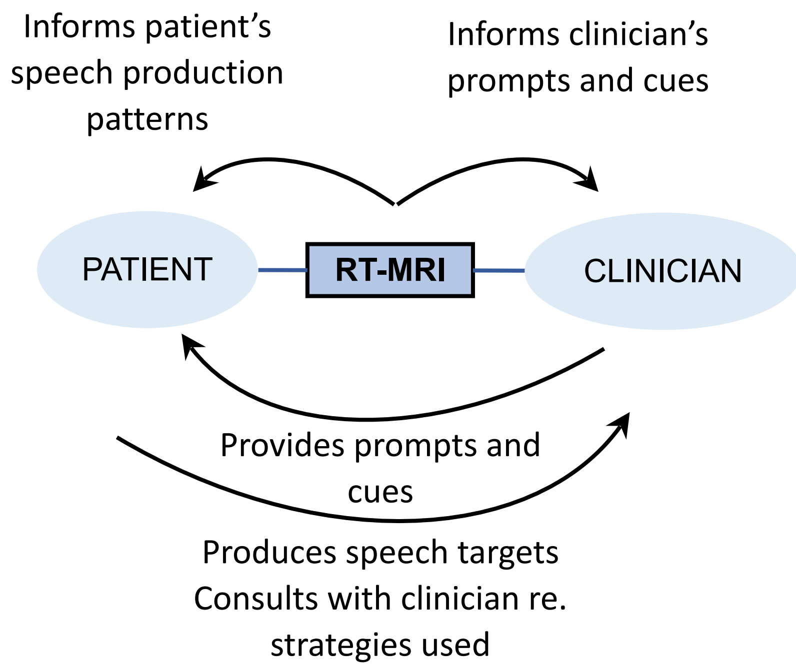
NEXT

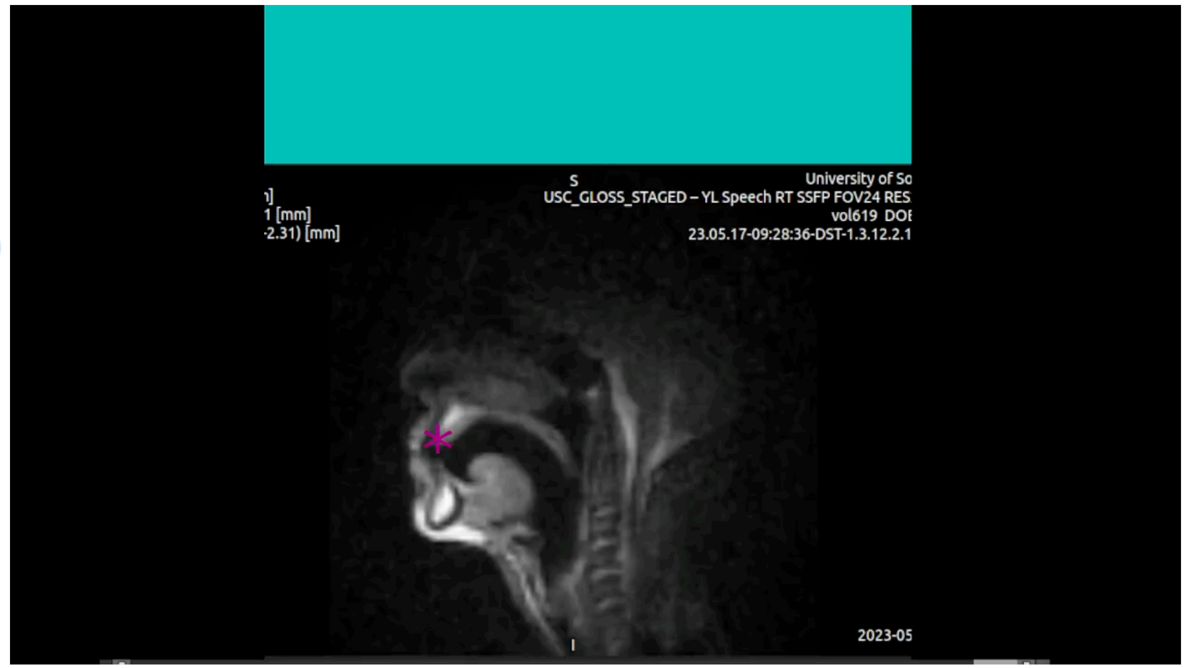
things just underway

Developing first Ever RT-MRI Biofeedback for Speech Rehabilitation

Christina Hagedorn
Krishna Nayak
Uttam Sinha

- **Improving the efficacy of post-operative speech rehabilitation**
  - *imaging as the basis of the biofeedback tool, allowing **both** the clinician and patient to see the entire vocal tract and hear the resulting acoustics in real time*





Hagedorn, C., Kumar, P., Villegas, B., Ouyoung, M., Cui, S., Sheth, M., Narayanan, S., Nayak, K., & Sinha, U.. The Role of High-Performance Low Field Magnetic Resonance Imaging in the Management of Tongue Cancer [Podium Presentation]. The American Head and Neck Society (AHNS) 11th International Conference on Head & Neck Cancer, Montreal, Canada. 2023

Mapping the dynamics of production in stuttering

Yijing Lu
Louis Goldstein

New insights possible

- initial consonant repetition is widely hypothesized as difficulty in *planning* the following vowel (Wingate, 1988; Howell, 2004; Postma & Kolk, 1993; Guenther, 2016)

'p-p-p-p-p-p-people'

'p-p-p-pot'

Our data however suggest that while speaker is repeating the initial consonant the tongue posture for the following vowel is *already* taking place

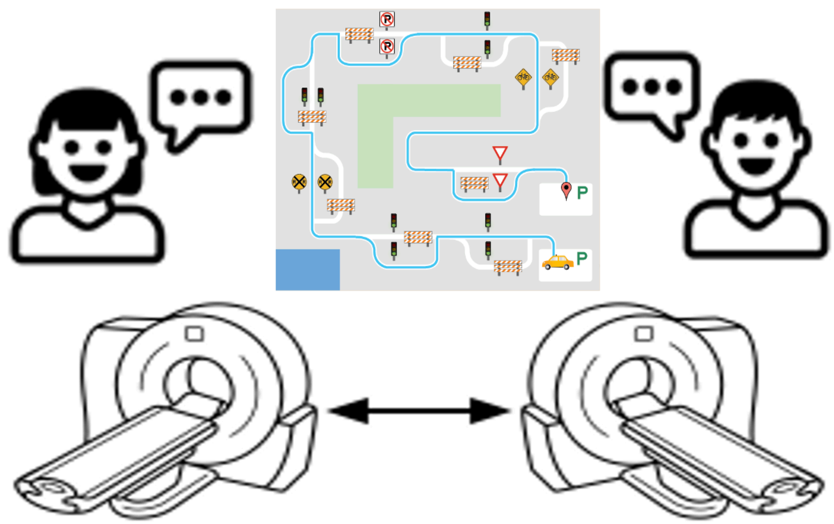
Yijing Lu, Louis Goldstein, Shrikanth Narayanan. Upcoming vowel gestures are articulated during initial consonant dysfluencies. 13th Oxford Dysfluency Conference. Oxford, September 2023

Structured Variability in Vocal Tract Articulation Dynamics

Dani Byrd
Louis Goldstein
Krishna Nayak

Understand and address the rich and pervasive variability in speech production, both within and across individuals and for varied interactional contexts

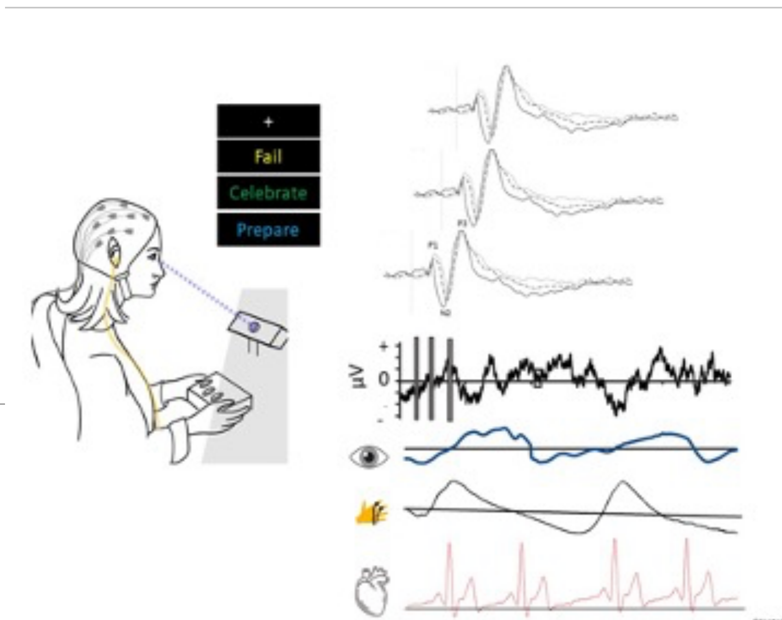
- *over tasks and over time*: *days, weeks, years*
- *in communicative interaction to observe how humans plan and produce speech collaboratively with one another at a high level of spatiotemporal detail*

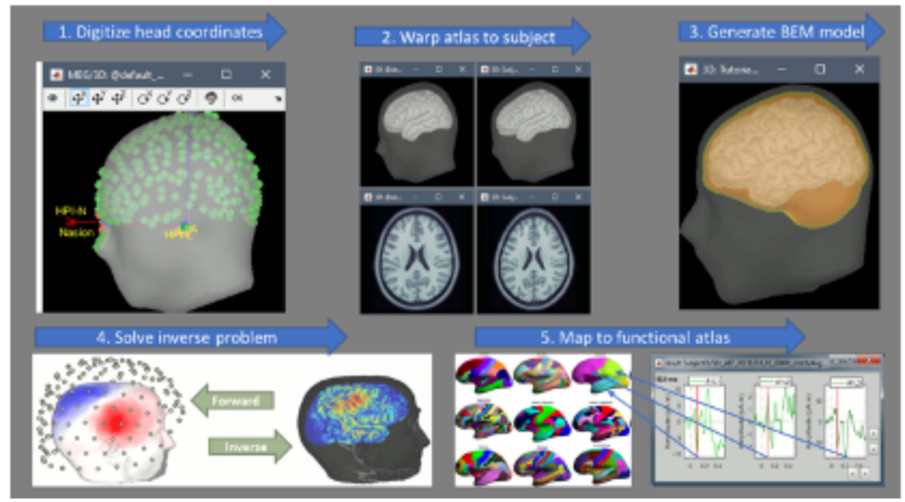


SCHEMATIC OF VISUAL PRESENTATION FOR A JOINT MAZE NAVIGATION TASK WITH TWO INTERCONNECTED RTMRI MACHINES

Multimodal integration of neural and biobehavioral signals for mapping preconscious and conscious processing

Rich multimodal sensing of brain-body response to affectively-salient linguistic stimuli





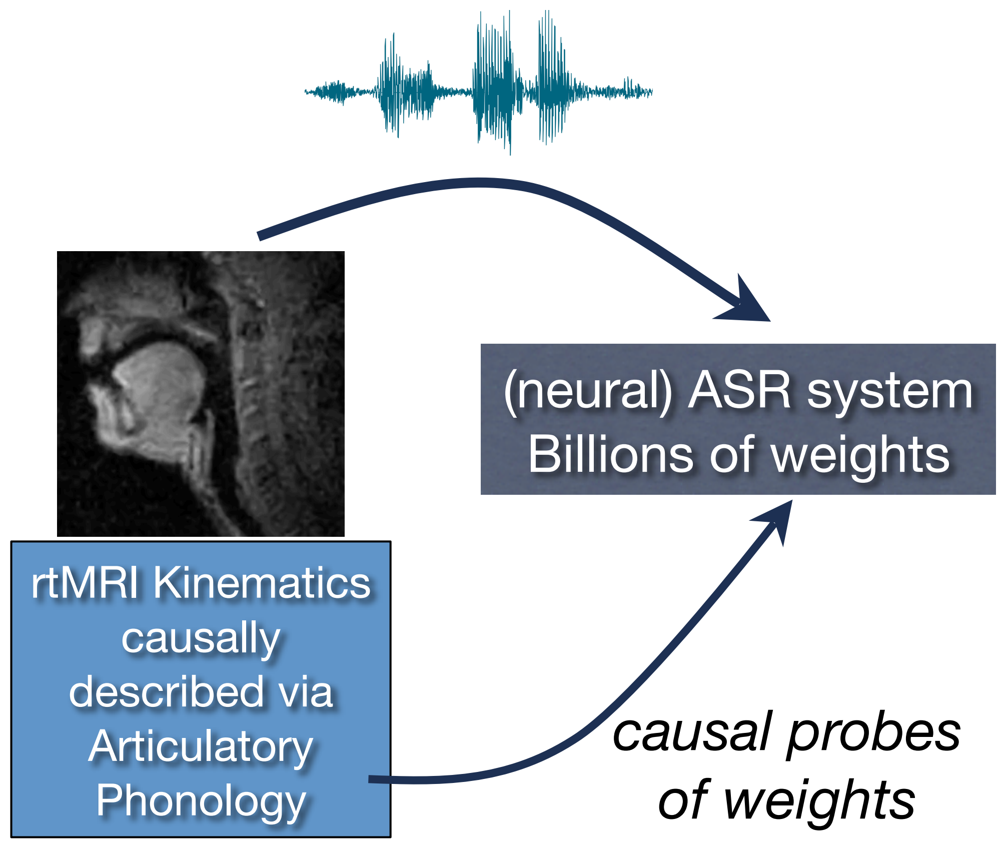
Robust signal processing, machine learning and prediction

With — **USC:** Bogdan, Byrd, Cahn, Damasio, Ferrara, Habibi, Leahy, Lerman; **UCLA:** Blank

Khalil Iskarous
Dani Byrd

Use of rtMRI to open the speech AI blackbox

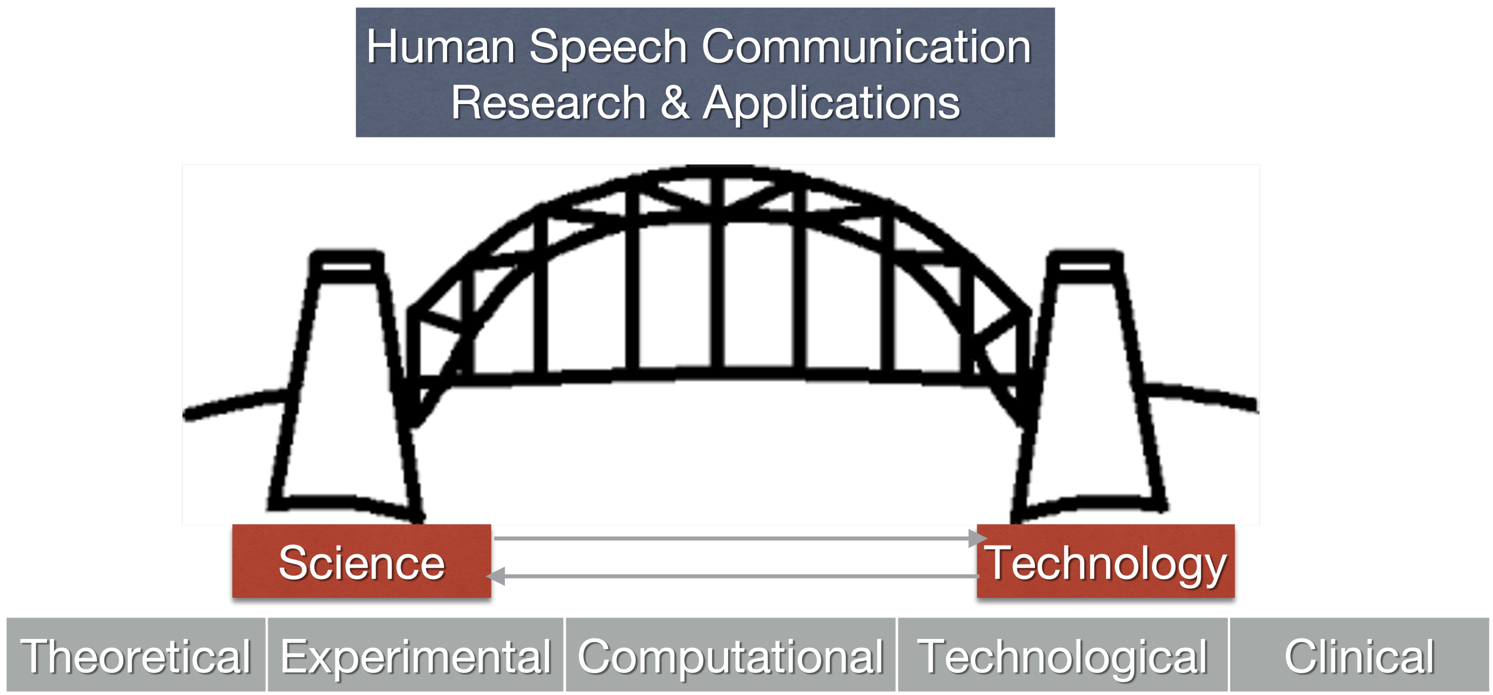
- Speech AI systems have powerful algorithms and architectures: but we still don't understand what these systems *know* in their parameters
- Technology allows us to observe and measure the data generation process for speech in detail not possible before
- **We are developing techniques to co-register rtMRI-measured speech and corresponding ASR parameters to probe how the latter capture the former**
    - how neural model architectures extract causal information from data?



"Analysis-by-Generation"

Connecting it all together with data and models

Planning, Processing and Production in Speech Interaction



Highlight 2

Rich Speech Processing and Behavioral Machine Intelligence

- engineering approaches to illuminate human trait and mental state
- screening, diagnostic, intervention support in mental and behavioral health

USC School of Engineering

SUPPORT FROM NIH, NSF, DoD, IARPA,
Simons Foundation, Guggenheim Foundation,
Toyota, Apple, Amazon

University of Southern California

PREVALENCE OF SELECTED HEALTH CONDITIONS (IN THE US)

All impact the production, processing and use of speech and language

Condition	Ages	Prevalence*
Autism spectrum disorder	Children (typically diagnosed as children, but persist over lifetime)	1.5% (lifetime)
Posttraumatic stress disorder	Adults	3.5% (one year)
Mood disorders (e.g., depression)	Adults	9.5% (one year)
Alcohol addiction/abuse	All	6.6% (one year)
Illicit drug use (nonmarijuana)	All	2.5% (one year)
Parkinson's disease	> 80 years old	1.9% (lifetime)
Dementia (e.g., Alzheimer's disease)	> 60 years old	6.5% (lifetime)

***Sources listed in:**
Daniel Bone, Chi-Chun Lee, Theodora Chaspari, James Gibson, and Shrikanth Narayanan. Signal Processing and Machine Learning for Mental Health Research and Clinical Applications. IEEE Signal Processing Magazine. 34(5): 189-196, September 2017

Autism Spectrum Disorder

- **1 in 36 US children diagnosed with ASD (CDC, 2022)**
- **ASD characterized by difficulties in social communication, reciprocity; presence of repetitive or stereotyped behaviors and interests**

Diagnostic targets include
Prosody
Turn-taking
Affective expressions
Shared enjoyment
….

CREDIT: WPS/ADOS TRAINING VIDEO

CDC: https://www.cdc.gov/ncbddd/autism/data.html

PREVALENCE OF SELECT HEALTH CONDITIONS (IN THE US)

Condition	Ages	Prevalence*
Autism spectrum disorder	Children (typically diagnosed as children, but persist over lifetime)	1.[illegible]etime)
Posttraumatic stress disorder	Adults	[illegible].5% (one year)
Mood disorders (e.g., depression)	Ad[illegible]	9.5% (one year)
Alcohol addiction/abuse	[illegible]ll	6.6% (one year)
Illicit drug use ([illegible]a)	All	2.5% (one year)
Parkinson's [illegible]ease	> 60 years old	1.9% (lifetime)
Dementia (e.g., Alzheimer's disease)	> 65 years old	6.5% (lifetime)

speech & language as biomarkers

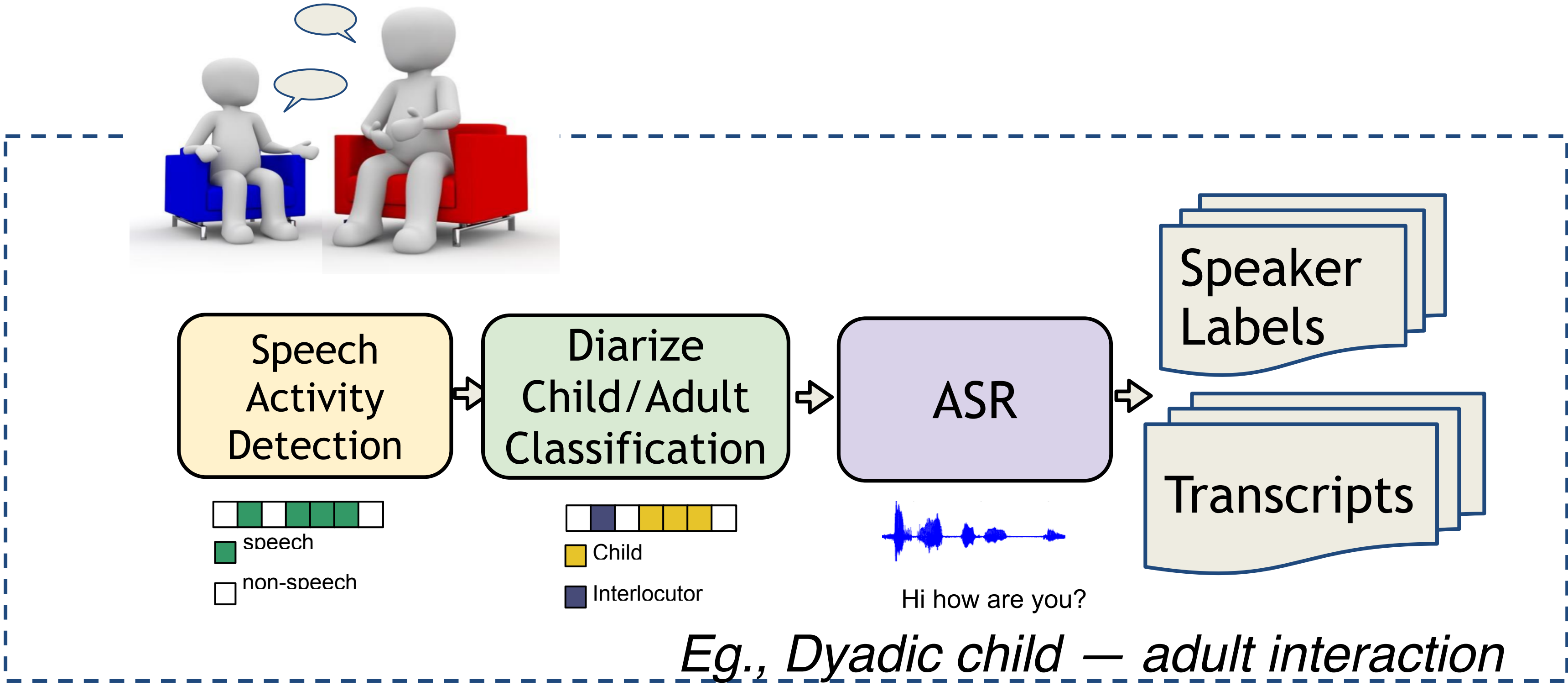
***Sources listed in:**
Daniel Bone, Chi-Chun Lee, Theodora Chaspari, James Gibson, and Shrikanth Narayanan. Signal Processing and Machine Learning for Mental Health Research and Clinical Applications. IEEE Signal Processing Magazine. 34(5): 189-196, September 2017

Tremendous advances in speech and language technologies

- Voice Activity Detection
- Speaker diarization
- Alignment
- Transcription
- Keyword spotting
- Prosody Modeling: Intonation, Phrasing, Prominence
- Voice Quality
- Synthesis
- Enhancement
- Dialog Act Tagging
- Interaction modeling: Turn taking dynamics, Entrainment
- Speaker/Verification Identification
- Affective Computing from Speech and Language
- Speaker State and Trait Characterization
- Joint speech and visual cue processing
- …

offer foundation for many downstream inquiry/applications

Latency as a biomarker of behavioral change in ASD



nature › scientific reports › articles › article

Article | Open Access | Published: 28 February 2022

Intra-topic latency as an automated behavioral marker of treatment response in autism spectrum disorder

Elizabeth P. McKernan, Manoj Kumar, Adriana Di Martino, Lisa Shulman, Alexander Kolevzon, Catherine Lord, Shrikanth Narayanan & So Hyun Kim

Scientific Reports **12**, Article number: 3255 (2022) | Cite this article

67

Lexical and acoustic features can be used to track changes in mental health states over time

Longitudinal study of speech samples from adults with serious mental illness answering open ended prompts (2-3 mins phone call)

- ***Acoustic: pitch, intonation, inter-word pause***
- ***Lexical: emotion, complexity, affect, concreteness,..***

PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Clinical state tracking in serious mental illness through computational analysis of speech

Armen C. Arevian, Daniel Bone, Nikolaos Malandrakis, Victor R. Martinez, Kenneth B. Wells, David J. Miklowitz, Shrikanth Narayanan

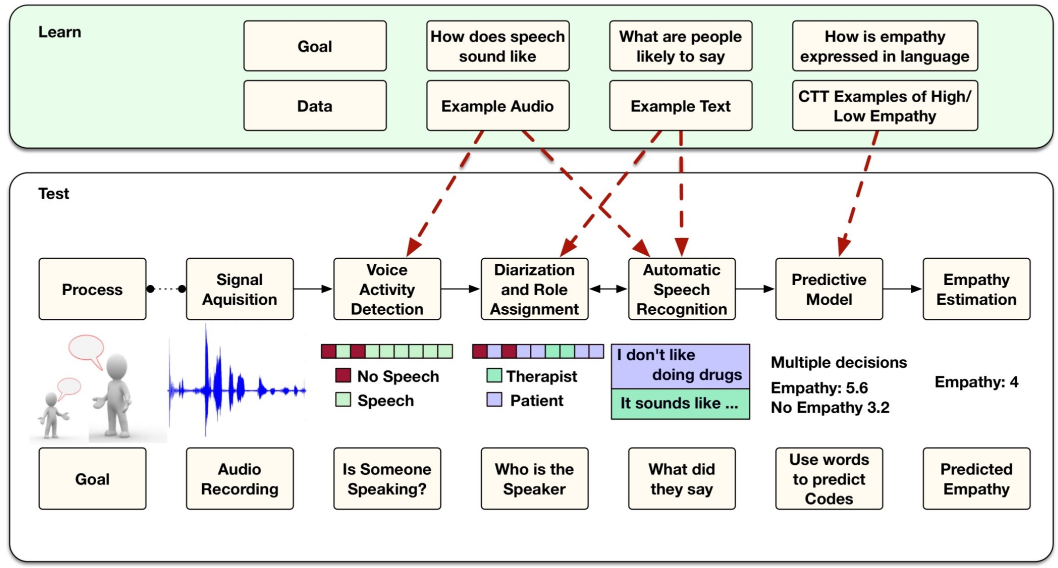
Published: January 15, 2020 • https://doi.org/10.1371/journal.pone.0225695

Psychotherapy: conversation based support

Illuminating what works, for whom, how and why

Motivational Interviewing

https://www.youtube.com/watch?v=EvLquWI8aqc



Expressed empathy as a marker of therapy efficacy

PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

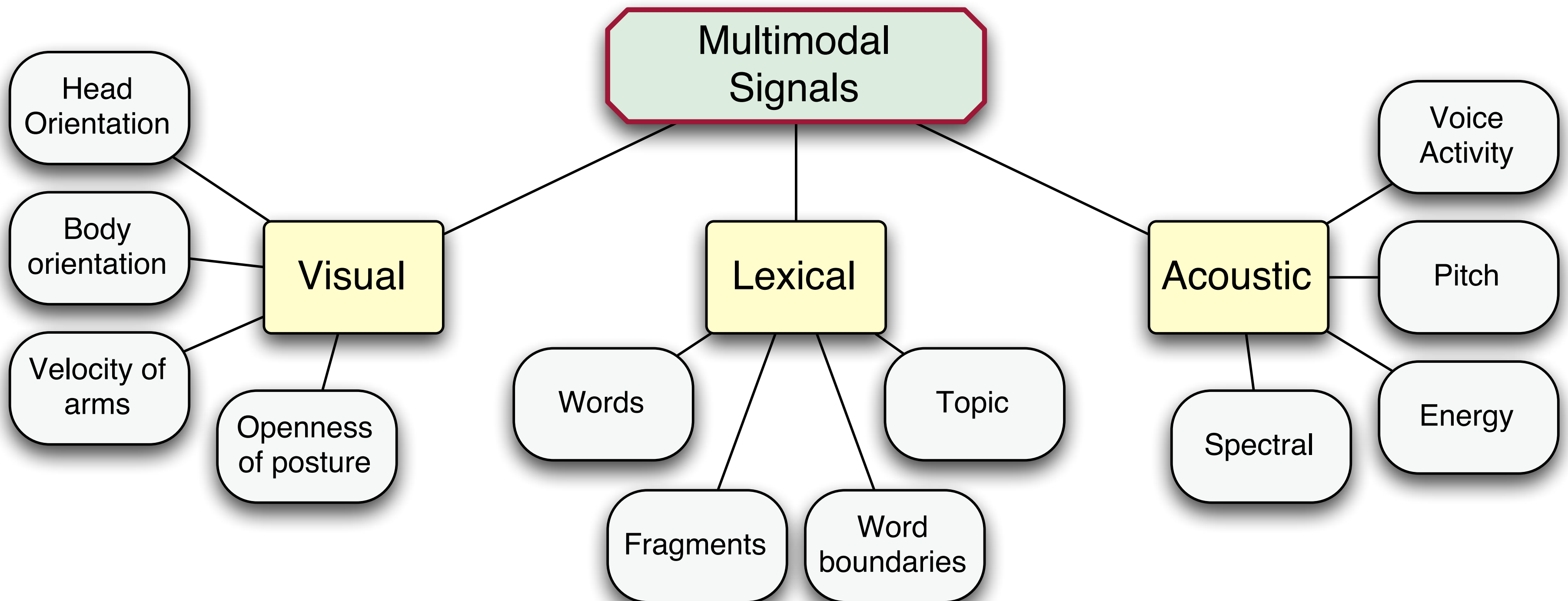
"Rate My Therapist": Automated Detection of Empathy in Drug and Alcohol Counseling via Speech and Language Processing

Bo Xiao, Zac E. Imel, Panayiotis G. Georgiou, David C. Atkins, Shrikanth S. Narayanan

Published: December 2, 2015 • https://doi.org/10.1371/journal.pone.0143055

Automatic prediction of constructs e.g., affect

—*a multimodal machine intelligence exercise*

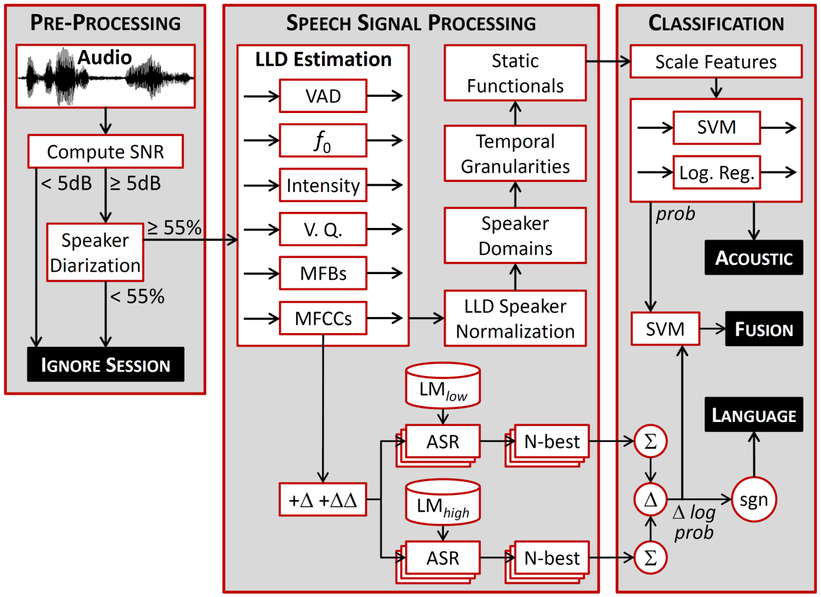


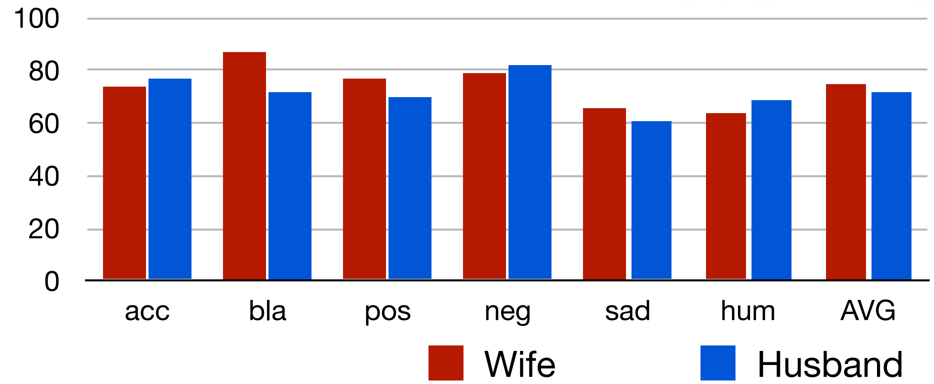
Dyadic Interactions of Couples in Relationship distress

Characterizing affective dynamics, humor, blame patterns as behavioral coding

(circa 2008)

Speech and Language Processing





Binary classification, linear SVM: Prosody (pitch, energy), spectral (MFCCs), voice quality (jitter, shimmer) features

Classifier Type	Accuracy
Baseline Chance	50%
Language	75.4%
Acoustic	79.6%
Fusion	**82.1%**

Most blaming words in terms of discriminative contribution — Least blaming words in terms of discriminative contribution

High Blame		Low Blame	
word	*Δ log prob*	*word*	*Δ log pro*
YOU	-9.61	**UM**	6.01
YOUR	-4.06	**THAT**	2.67
ME	-2.53	**I**	2.57
TELL	-1.51	**WE**	2.36
ACCEPT	-1.45	**THINK**	2.07

M. BLACK, ET AL "AUTOMATIC CLASSIFICATION OF MARRIED COUPLES' BEHAVIOR USING AUDIO FEATURES" - INTERSPEECH 2010
M. BLACK, ET AL TOWARD AUTOMATING A HUMAN BEHAVIORAL CODING SYSTEM FOR MARRIED COUPLES' INTERACTIONS USING SPEECH ACOUSTIC FEATURES. SPEECH COMMUNICATION. 55(1):1-21, 2013
GEORGIOU, BLACK, LAMMERT, BAUCOM AND NARAYANAN. "THAT'S AGGRAVATING, VERY AGGRAVATING": IS IT POSSIBLE TO CLASSIFY BEHAVIORS IN COUPLE INTERACTIONS USING AUTOMATICALLY DERIVED LEXICAL FEATURES? PROCEEDINGS ACII, 2011

Models of Interaction Mechanisms

Interaction Synchrony / Entrainment [Kimura 2006]

Mutual adaptation of verbal/nonverbal behaviors in dyadic interactions

Positive vs. Negative valence in interactions

Higher degree of entrainment in positive interactions [Kimura 2006, Warner 1987]

Entrainment measures as features for automatic classification [Margolin 1998]

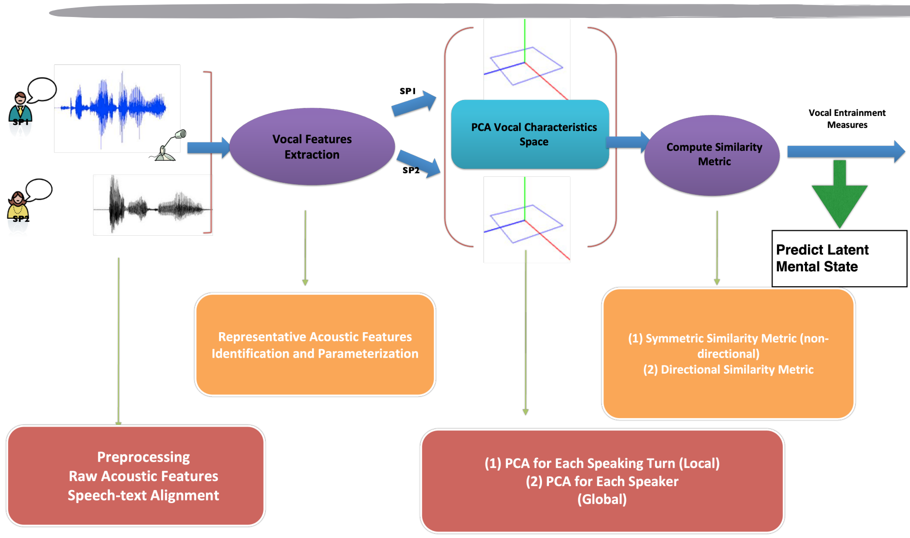
Quantification of Prosodic Entrainment

Signal-derived quantitative measure

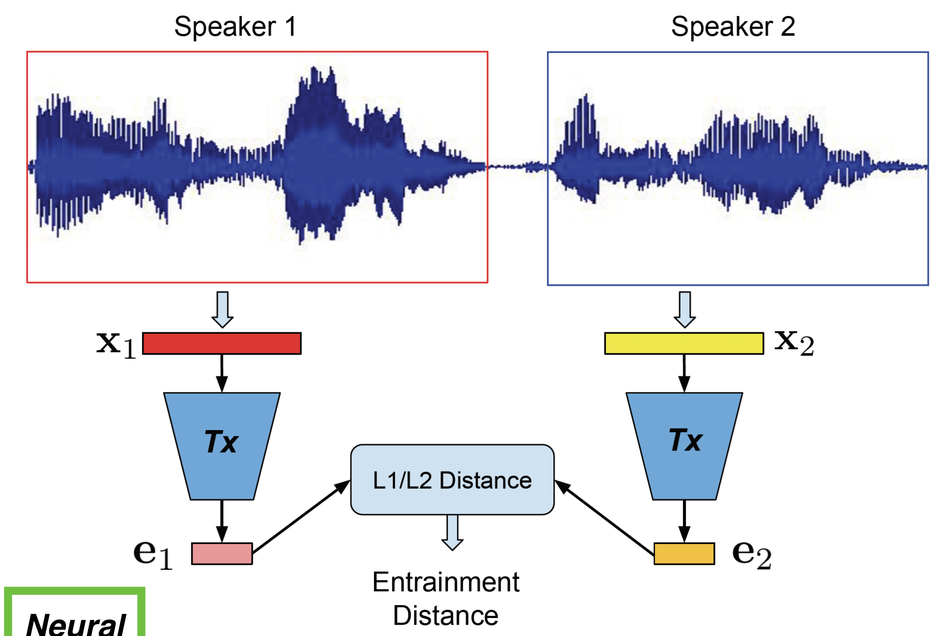
"HOW DO TWO PEOPLE SOUND ALIKE AS THEY INTERACT IN A CONVERSATION?"

Computing Vocal Entrainment

“HOW MUCH DO TWO PEOPLE *SYNCHRONIZE* IN A CONVERSATION?”



Classic



Neural

Computational entrainment measures useful in predicting

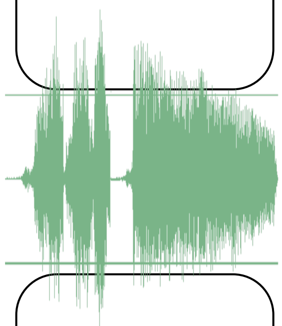
- ***couple therapy codes (agreement and blame)***
- ***couples therapy outcome***
- ***emotional bond in Suicide risk assessment interviews***

CHI-CHUN LEE, ET AL. COMPUTING VOCAL ENTRAINMENT: A SIGNAL-DERIVED PCA-BASED QUANTIFICATION SCHEME WITH APPLICATION TO AFFECT ANALYSIS IN MARRIED COUPLE INTERACTIONS. COMPUTER, SPEECH, AND LANGUAGE. 28(2): 518-539, MARCH 2014

MD NASIR, BRIAN BAUCOM, SHRIKANTH NARAYANAN, PANAYIOTIS GEORGIOU. MODELING VOCAL ENTRAINMENT IN CONVERSATIONAL SPEECH USING DEEP UNSUPERVISED LEARNING. IEEE TRANSACTIONS ON AFFECTIVE COMPUTING. 13(3): 1651–1663, 2022

Engineering a technology pipeline:
from speech to target constructs

Processing Steps	Recording and Pre-Processing	Voice Activity Detection	Automatic Speaker Segmentation	Automatic Speech Recognition	Automatic Acoustic Feature Recognition	Computational Behavioral Modeling
Goal of Each Step	Capture Raw Signal Data	Is someone speaking?	Who is speaking?	What did each speaker say?	What are the non-verbal cues?	Feedback to experts & New Features of behavioral change



speech
non-speech

interlocutor
child

I like the globe

Do you like this puzzle?

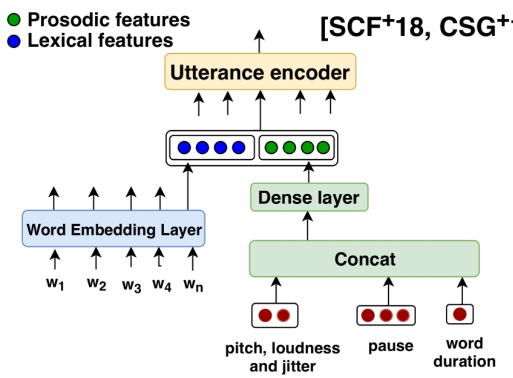
✓ vocal intensity
✓ pitch-matching
✓ turn-taking
✓ voice quality

74

Multimodal Behavior Understanding

Speech + Text

[SCF+18, CSG+19]

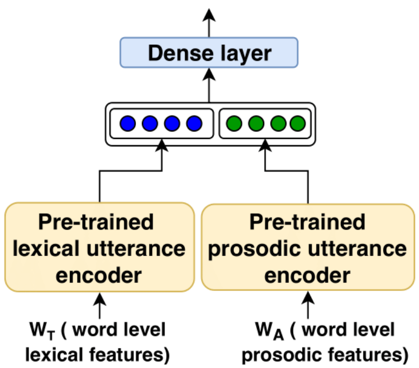


FUSION: Level of

Words

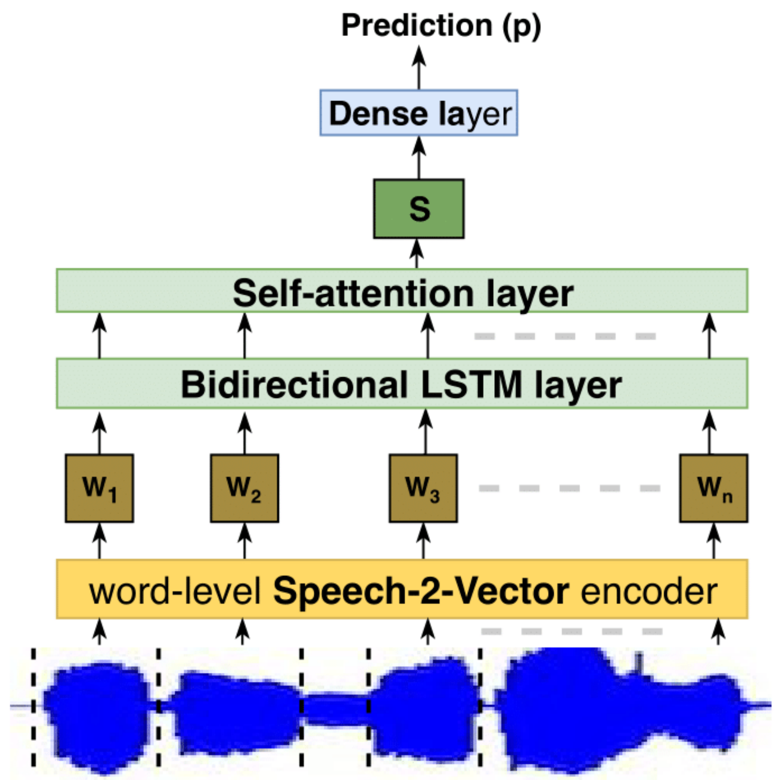
Or

Utterances



acoustic-prosodic information complements lexical information (Singla et al., 2018)

"end to end" Transcription-free prediction



speech features and word segmentation information for predicting spoken utterance-level target labels (Singla et al., 2020)

- Karan Singla, Zhuohao Chen, David Atkins, and Shrikanth Narayanan. Towards end-2-end learning for predicting behavior codes from spoken utterances in psychotherapy conversations. In Proceedings of ACL pp. 3797–3803, 2020.
- Karan Singla, Zhuohao Chen, Nikolaos Flemotomos, James Gibson, Dogan Can, David Atkins, and Shrikanth S. Narayanan. Using Prosodic and Lexical Information for Learning Utterance-level Behaviors in Psychotherapy. In Proceedings of InterSpeech, 2018.

Multi-label Multi-task Modeling: Psychotherapy Behaviors

Generalizing

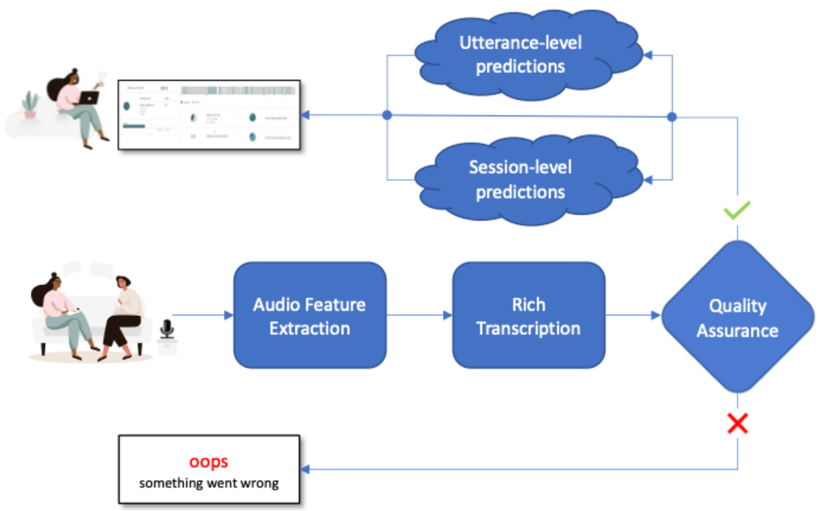
across domains: Motivational Interviewing, Cognitive Behavioral Therapy,…

- **Multi-label learning**
  - benefits prediction of less frequently occurring behaviors by leveraging modeling of more frequent behaviors
- **Multi-task learning**
  - benefits prediction of behaviors across domains by modeling common behaviors
- **Modeling user-turn context useful**
- **Evaluation on two psychotherapy approaches**
  - *Motivational Interviewing* (11 aggregate MISC codes; 345 sessions)
  - *Cognitive Behavioral Therapy* (11 CTRS codes; 92 sessions)
  - Deep Multi label Multi task Context aware learning: >5% absolute improvement in code prediction for both domains

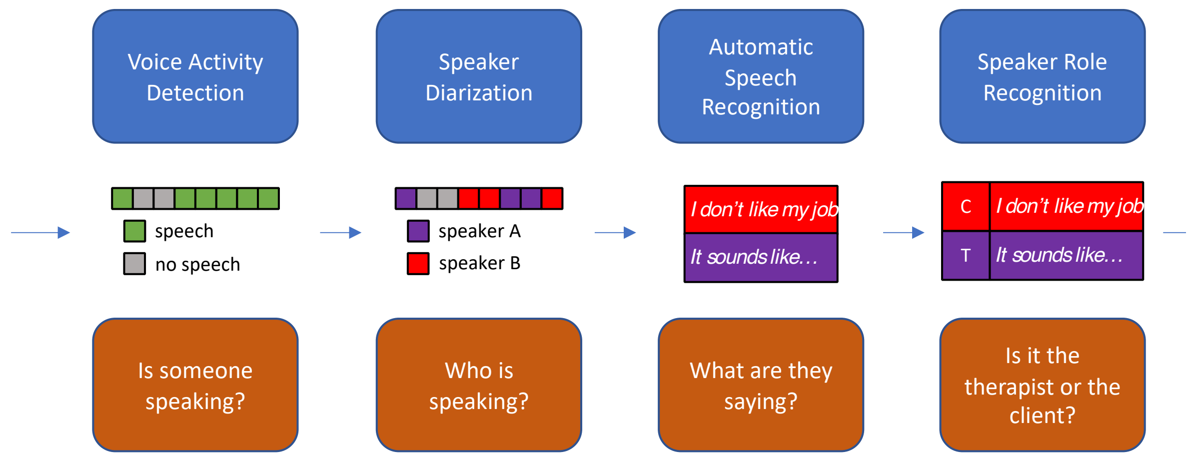
J. Gibson, D. Atkins, T. Creed, Z. Imel, P. Georgiou and S. Narayanan, "Multi-label Multi-task Deep Learning for Behavioral Coding," in IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2019.2952113. 13(1): 508-518, 2022

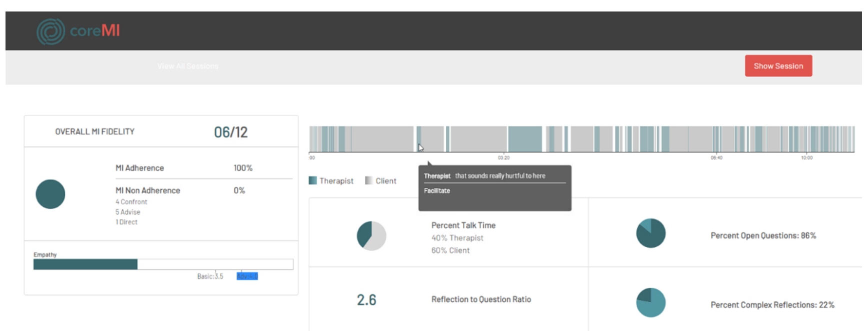
Dissemination in mental health clinics:
Automated therapy evaluation

Processing workflow



Rich Transcription Pipeline





key results provided to the user:

- session timeline with utterance-level codes
- session-level codes
- summary indicators and session dynamics
- overall fidelity to the therapeutic approach

N. Flemotomos, V.R. Martinez, Z. Chen, K. Singla, V. Ardulov, R. Peri, D. Caperton, J. Gibson, M.J. Tanana, P. Georgiou, J. Van Epps, S.P. Lord, T. Hirsch, Z.E. Imel, D.C. Atkins, and S. Narayanan **Automated Evaluation of Psychotherapy Skills Using Speech and Language Technologies**, *Behavior Research Methods.* (2021).

Speech as biomarker:
interplay between factors?

*Speech and language encode and provide access to **intent, emotions**, and a variety of information about **demographic traits** (age, gender, size…), **physical/ psychological health state,** and **interaction context.***

These attributes/constructs are often intricately related.

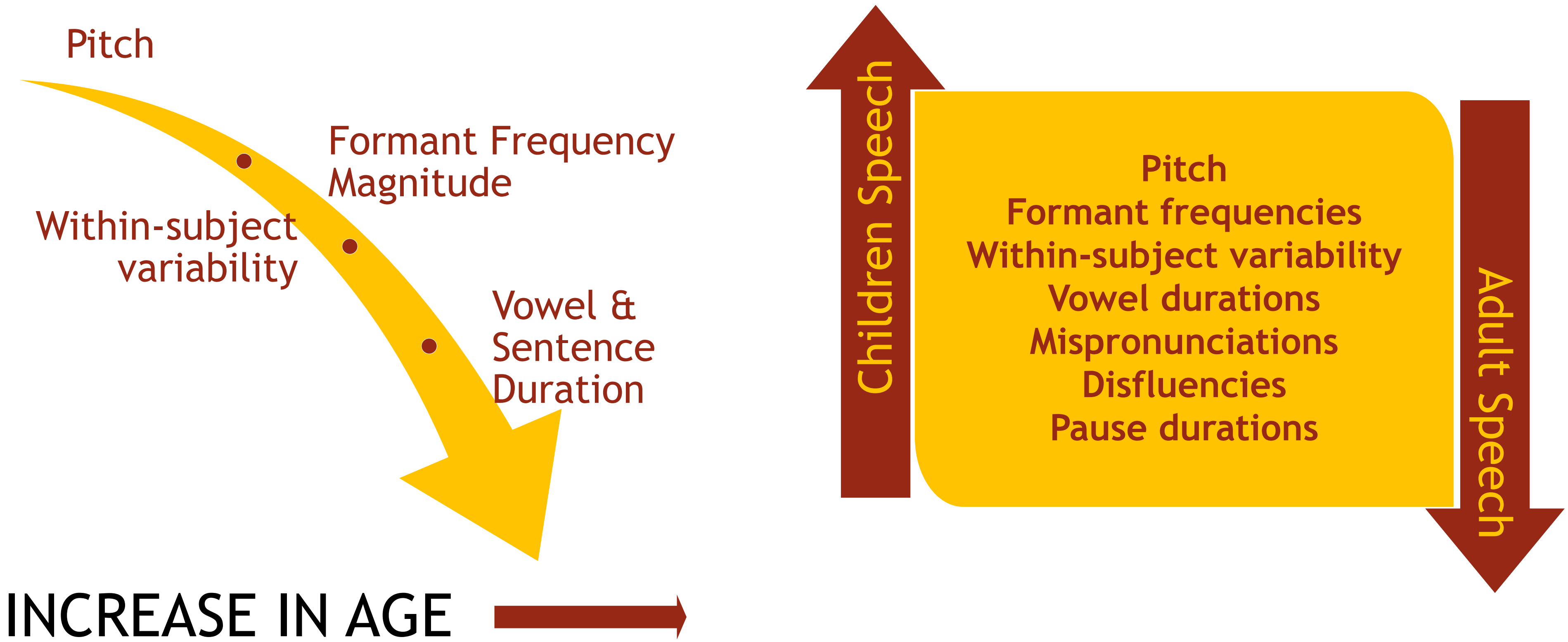
Health + Age related changes?
across life span

A spotlight on
Processing Children's Speech

- **What is special about it?**
  - **Review acoustic properties**
- **Robust speech processing techniques**

Developmental changes revealed in speech

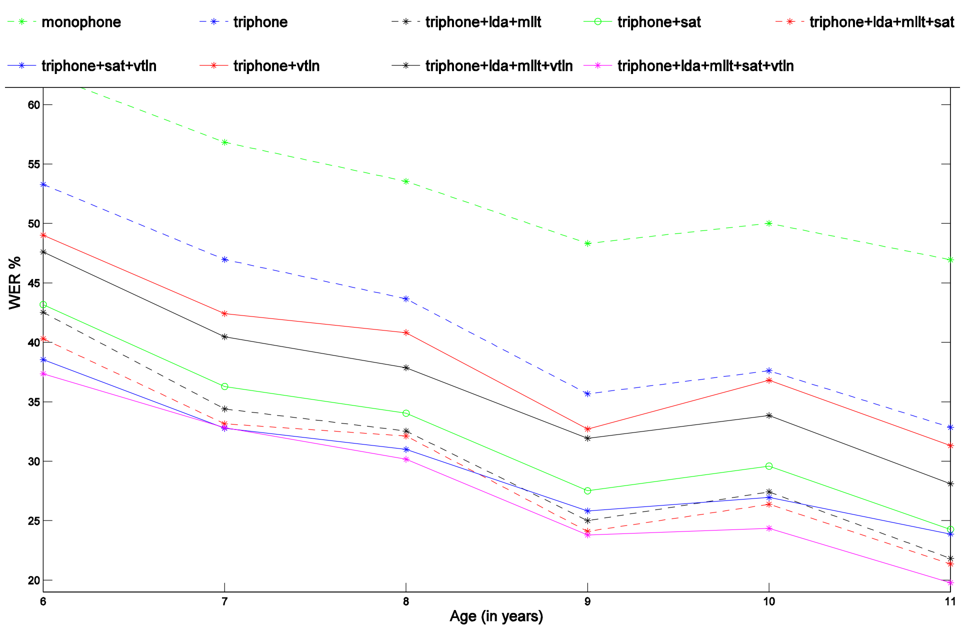
Reduction in speech parameter values as a function of age as children grow



- Sungbok Lee, Alexandros Potamianos, and Shrikanth Narayanan. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. J. Acoust. Soc. Am., 105:1455-1468, Mar. 1999 (Selected Research Article)
- Sungbok Lee, Alexandros Potamianos, and Shrikanth S. Narayanan. Developmental acoustic study of American English diphthongs. J. Acoust. Soc. Am., 136(4):1880–1894, oct 2014.

Early work in ASR of Children Speech

- **Performance varies with age: 2-5 times error than adult speech**
  - 50% relative error reduction due to frequency warping and model adaptation, larger for speakers under 12 years
  - Despite improvement relative error rate is at least 30% higher for 6-9 year olds
  - Age-dependent models provide an additional 10% relative error rate reduction
  - Front end vocal tract normalization (especially when training-testing age mismatch), speaker normalization, other spectral adaptation techniques



WER decreases almost linearly with increase in age

- Shrikanth Narayanan and Alexandros Potamianos. Creating conversational interfaces for children. IEEE Trans. Speech and Audio Processing, 10(2):65-78, 2002.
- Alexandros Potamianos and Shrikanth Narayanan. Robust recognition of children's speech. IEEE Trans. Speech and Audio Processing, 11:603-616, Nov. 2003.
- Prashanth Gurunath Shivakumar, Alexandros Potamianos, Sungbok Lee and Shrikanth Narayanan. Improving Speech Recognition for Children using Acoustic Adaptation and Pronunciation Modeling. In Proceedings of Workshop on Child Computer Interaction (WOCCI 2014), Singapore, September, 2014

Improving Speaker Detection & Diarization In Interactions Involving Children

- **Large within-class variability** especially for child from age, gender, clinical symptom severity (Lee 1999, 2014; Gerosa 2009)
- **Lack of sufficient & balanced** training data covering different factors/conditions

novel machine learning, formulating solutions in a "case specific" manner, leveraging interaction context

- Rimita Lahiri, Manoj Kumar, Somer Bishop, and Shrikanth Narayanan. Learning domain invariant representations for child-adult classification from speech. In Proceedings of ICASSP, May 2020.
- Nithin Rao, Manoj Kumar, So Hyun Kim, Catherine Lord, and Shrikanth Narayanan. Meta-learning for robust child-adult classification from speech. In Proceedings of ICASSP, May 2020.
- Manoj Kumar, So Hyun Kim, Catherine Lord, and Shrikanth Narayanan. Speaker Diarization for Naturalistic Child-Adult Conversational Interactions using Contextual Information.. J. Acoust. Soc. Am., 147(2):EL196–EL200, February 2020.
- Monisankha Pal, Manoj Kumar, Raghuveer Peri, Tae Jin Park, So Hyun Kim, Catherine Lord, Somer Bishop, and Shrikanth Narayanan. Meta-learning with Latent Space Clustering in Generative Adversarial Network for Speaker Diarization. IEEE/ACM Transactions on Audio, Speech and Language Processing. 29: 1204-1219, 2021

Improving Speech Recognition For Children

creating and bringing contemporary ML advances

- **Manoj Kumar, So Hyun Kim, Catherine Lord, Thomas Lyon, and Shrikanth Narayanan. Leveraging Linguistic Context in Dyadic Interactions to Improve Automatic Speech Recognition for Children. Computer, Speech and Language, 63, 2020.**

- **Prashanth Gurunath Shivakumar, Shrikanth Narayanan. End-to-End Neural Systems for Automatic Children Speech Recognition: An Empirical Study. Computer Speech & Language. 72:101289, 2022**

End to End Child-centric ASR Summary

- End-to-end systems provide near constant improvements over all age categories after adaptation on child speech
- Absolute WER with the end-to-end systems better than DNN-HMM
  - *Gap in performance between adult and children wider for end-to-end systems compared to DNN-HMM ASR*
- Addition of large amounts of adult speech is found to be beneficial (more benefits for ASR for younger children)
- Transformer network architectures are the best performing models when the train–test mismatch is low, however they do not generalize well
- CTC loss based models are robust to children speech recognition; Sequence-to-sequence models can breakdown during high mismatch conditions
- Better performance with greedy decoding without language model
- Benefits established with end-to-end ASR for adult speech still **do not** translate completely to children speech
  - *ASR for children is 10 – 19 times worse than Adults and 6 times worse despite adaptation on children speech*

Prashanth Gurunath Shivakumar, Shrikanth Narayanan. End-to-End Neural Systems for Automatic Children Speech Recognition: An Empirical Study. Computer Speech & Language. 72:101289, 2022

Diverse Applications

- **HELP US DO THINGS WE KNOW TO DO BUT MORE EFFICIENTLY, CONSISTENTLY**
  - » READING ASSESSMENT

- MATTHEW BLACK, JOSEPH TEPPERMAN AND SHRIKANTH NARAYANAN. AUTOMATIC PREDICTION OF CHILDREN'S READING ABILITY FOR HIGH-LEVEL LITERACY ASSESSMENT. IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING. 19(4): 1015 - 1028, 2011.
- JOSEPH TEPPERMAN, SUNGBOK LEE, SHRIKANTH NARAYANAN AND ABEER ALWAN. A GENERATIVE STUDENT MODEL FOR SCORING WORD READING SKILLS. IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING. 19(2): 348-360, 2011.

- **HELP HANDLE NEW DATA, CREATE NEW MODELS TO OFFER NEW INSIGHTS**
  ➡CREATE TOOLS FOR SCIENTIFIC DISCOVERY
  - » CAUSAL INDICATORS OF TRUTHFULNESS IN FORENSIC INTERVIEWS

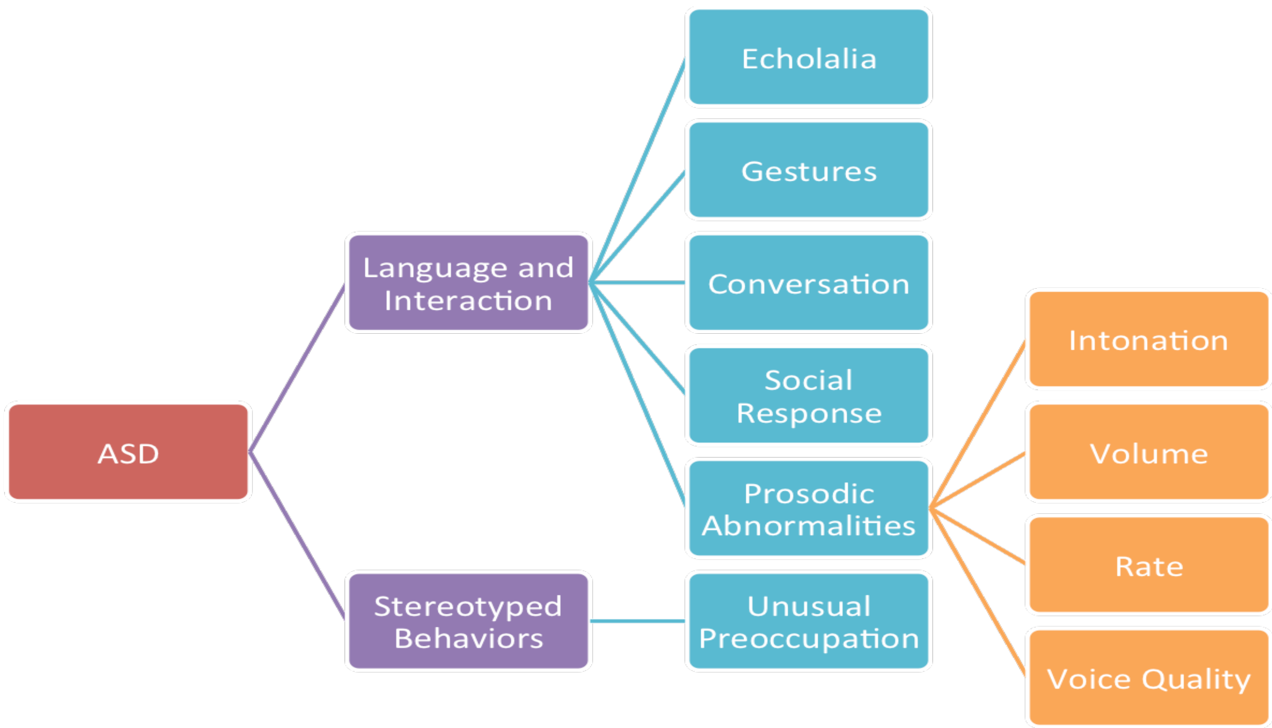
- ZANE DURANTE, VICTOR ARDULOV, MANOJ KUMAR, JENNIFER GONGOLA, THOMAS LYON, SHRIKANTH NARAYANAN. CAUSAL INDICATORS FOR ASSESSING THE TRUTHFULNESS OF CHILD SPEECH IN FORENSIC INTERVIEWS. COMPUTER SPEECH & LANGUAGE. 71:101263, 2022

✓HELP CREATE TOOLS TO SUPPORT DIAGNOSTICS, PERSONALIZED INTERVENTION, AND TRACKING ITS RESPONSE TO TREATMENT
  - » SCREENING AND DIAGNOSIS IN AUTISM SPECTRUM DISORDER

Autism Spectrum Disorder (ASD)

- **1 in 36 US children diagnosed with ASD (CDC, 2022)**
  - 1% prevalence in Asia, Europe, North America, 2.6% in S. Korea

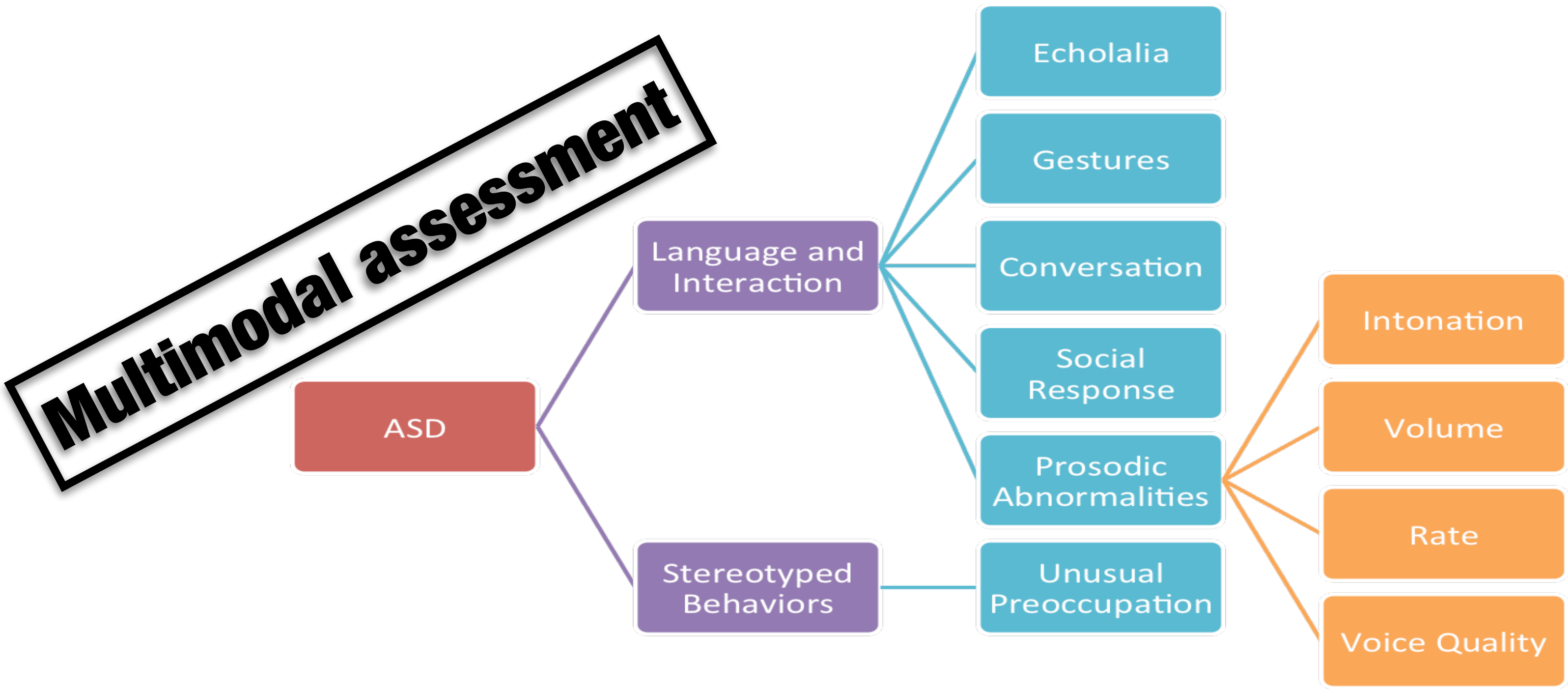
- **Difficulties in social communication, reciprocity; Repetitive or stereotyped behaviors and interests**
  - heterogeneous across individuals and contexts



86

Opportunities for rich multimodal approaches in Autism Spectrum Disorder (ASD)

- Better understand communication and social patterns of children
- Stratify behavioral phenotyping with quantifiable and adaptable metrics
- Track, quantify children's progress during interventions



D. Bone, M. Goodwin, M. Black, C-C.Lee, K. Audhkhasi, and S. Narayanan. Applying Machine Learning to Facilitate Autism Diagnostics: Pitfalls and promises. Journal of Autism and Developmental Disorders. 45(5), 1121-1136, 2015

Daniel Bone, Somer Bishop, Matthew P. Black, Matthew S. Goodwin, Catherine Lord, Shrikanth S. Narayanan. Use of Machine Learning to Improve Autism Screening and Diagnostic Instruments: Effectiveness, Efficiency, and Multi-Instrument Fusion. Journal of Child Psychology and Psychiatry. 57(8): 927-937, August 2016

Quantifying Atypical Prosody

Qualitative clinical descriptions are general and contrasting

ADOS*

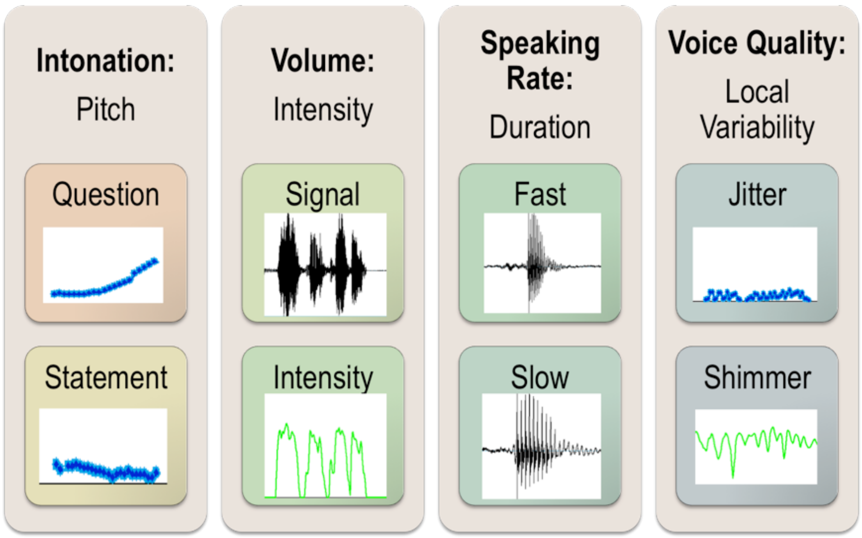
Module 3

"slow, rapid, jerky and irregular in rhythm, odd intonation or inappropriate pitch and stress, markedly flat and toneless, or consistently abnormal volume"

Structured assessment may not capture how atypical prosody affects social functioning apart from pragmatics

*Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Jr, Leventhal, B.L., DiLavore, P.C., … & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, **30**, 205–223.

Operationalizing: Acoustic, Language and Turn taking features



- **Acoustic features**: pitch (6), volume (6), rate (4), and voice quality (8)
  - Intonation: F0 curvature, slope, center
  - Volume: Intensity curvature, slope, center
  - Rate: Boundary (turn end word), Non boundary
  - Voice Quality: Jitter, Shimmer, CPP, HNR

- **Global Turn-taking Measures:** ***speech %***, ***silence %***, ***overlap %*** (*interruption %*), and ***median latency*** (time between turn exchanges)
- **Rate:** ***speaking rate*** (*SR*, #-words/utt. dur.; includes pausing), *per-word* **articulation *rate*** (*AR*, syl/word dur.), ***intra-utterance pausing duration***
- **Language:** *features from LIWC normalized by the total number of words*

(1) words per sentence (WPS)– a rough approximation of mean-length-of-utterance (MLU); (2) first-person, singular pronouns (I, me, mine); (3-5) positive emotion, negative emotion, and affect (positive or negative) language; (6-8) assents (OK, yes), non-fluencies (hm, umm), and fillers (I mean, you know).

DANIEL BONE, CHI-CHUN LEE, MATTHEW P. BLACK, MARIAN E. WILLIAMS, SUNGBOK LEE, PAT LEVITT, AND SHRIKANTH NARAYANAN, "THE PSYCHOLOGIST AS AN INTERLOCUTOR IN AUTISM SPECTRUM DISORDER ASSESSMENT: INSIGHTS FROM A STUDY OF SPONTANEOUS PROSODY", JOURNAL OF SPEECH, LANGUAGE, AND HEARING RESEARCH, 57:1162–1177, AUGUST 2014.

Analysis summary: child-psychologist interaction during ADOS administration

Objective insights from computational processing: mutual influence

- Prosodic, turn-taking, and language features of the interacting ***psychologist*** and ***child*** indicate conversational quality degrades for children with greater ASD severity:
  - psychologist and child speak with more intonational variability
  - psychologists vary their strategies to engage, reacting to the child's behavior
    - talk more when child does not; wait more when child takes more time
    - but also *evidence for entrainment e.g., voice quality matching*
    - child may be reluctant to discuss themselves, and may not follow up on prompts to engage psychologist; child uses less personal pronouns, esp. "I"
    - psychologist back-channels less, Child uses less fillers
  - psychologist's speech also shows evidence of vocal entrainment e.g., matching voice quality
- Interacting psychologist's speech features predict symptom severity of the child
  - ***Modeling Interaction Dynamics is Critical***

DANIEL BONE, CHI-CHUN LEE, THEODORA CHASPARI, MATTHEW P. BLACK, MARIAN E. WILLIAMS, SUNGBOK LEE, PAT LEVITT AND SHRIKANTH NARAYANAN, ACOUSTIC-PROSODIC, TURN-TAKING, AND LANGUAGE CUES IN CHILD-PSYCHOLOGIST INTERACTIONS FOR VARYING SOCIAL DEMAND, INTERSPEECH, 2013.

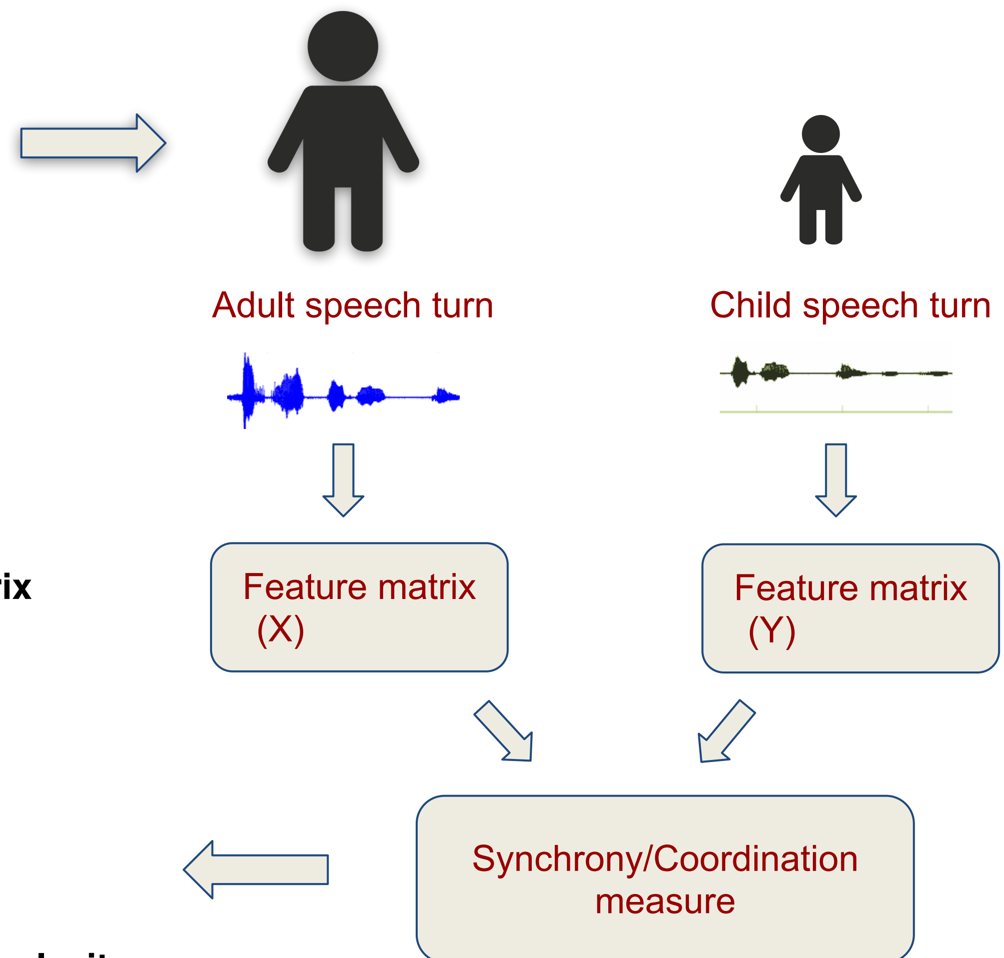
DANIEL BONE, CHI-CHUN LEE, MATTHEW P. BLACK, MARIAN E. WILLIAMS, SUNGBOK LEE, PAT LEVITT, AND SHRIKANTH NARAYANAN, "THE PSYCHOLOGIST AS AN INTERLOCUTOR IN AUTISM SPECTRUM DISORDER ASSESSMENT: INSIGHTS FROM A STUDY OF SPONTANEOUS PROSODY", JOURNAL OF SPEECH, LANGUAGE, AND HEARING RESEARCH, 57:1162–1177, AUGUST 2014.

YOUNG KYUNG KIM, RIMITA LAHIRI, MD NASIR, SO HYUN KIM, SOMER BISHOP, CATHERINE LORD AND SHRIKANTH NARAYANAN. ANALYZING SHORT TERM DYNAMIC SPEECH FEATURES FOR UNDERSTANDING BEHAVIORAL TRAITS OF CHILDREN WITH AUTISM SPECTRUM DISORDER. PROCEEDINGS OF INTERSPEECH, 2021

Quantifying synchrony

For every consecutive turn pair, coordination measures are computed, and they are averaged across all such turn pairs to get a session level measure

Child adult interaction



Acoustic: **Prosodic/Spectral Feature matrix**
Lexical**: BERT embeddings matrix**

Acoustic: **Squared cosine distance of complexity measure;**
Dynamic Time Warping distance
Lexical**: Word Movers Distance**

Quantifying interpersonal synchrony: insights

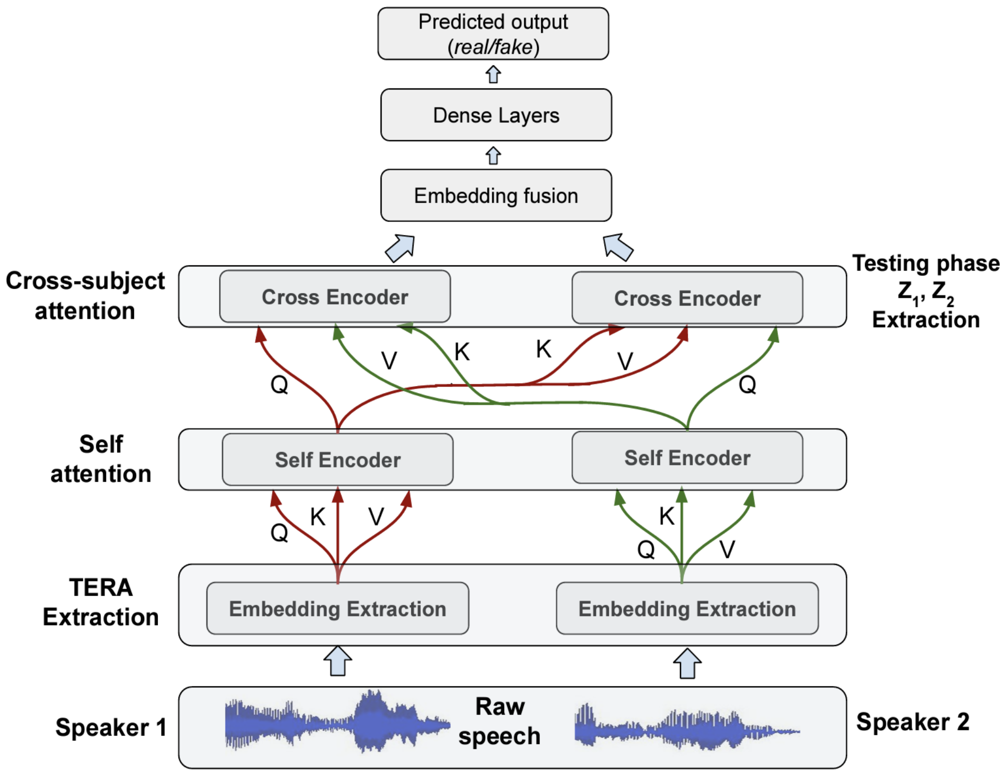
Child adult interaction

- Children with ASD diagnosis show less synchrony in both social and emotional subtasks in terms of the absolute values of the introduced measures

- Measures are ***complementary***: improved distinction between ASD and non-ASD groups when vocal and lexical synchrony measures are fused: 40% relative improvement in F1 score

Rimita Lahiri, Md Nasir, Manoj Kumar, SoHyun Kim, Somer Bishop, Cathy Lord, Shrikanth Narayanan. Interpersonal synchrony across vocal and lexical modalities in interactions involving children with Autism Spectrum Disorder. J. Acoust. Soc. Am. Express Letters 9(2): 095202, 2022

Interpersonal synchrony: towards more data-driven approaches



- Context-aware modeling using data-driven frameworks
- Conformer based framework to model context
- Cross-subject attention

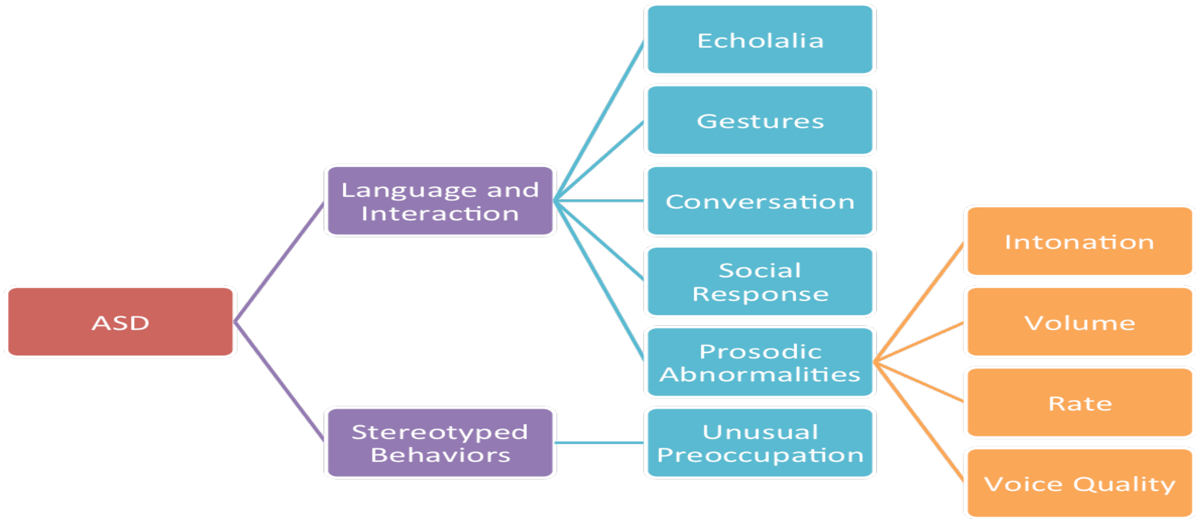
- **Both local and global context help**
- **Joint modeling of interlocutors help capture nuances of synchrony**

93

Rimita Lahiri, Md Nasir, Catherine Lord, So Hyun Kim, Shrikanth Narayanan. A Context-Aware Computational Approach for Measuring Vocal Entrainment in Dyadic Conversations. Proceedings of ICASSP, 2023

ASD: Opportunities for rich multimodal learning approaches

- Better understand communication and social patterns of children
- Stratify behavioral phenotyping with quantifiable and adaptable metrics
- Track, quantify children's progress during interventions



1. Daniel Bone, Matthew S. Goodwin, Matthew P. Black, Chi-Chun Lee, Kartik Audhkhasi, and Shrikanth Narayanan. Applying Machine Learning to Facilitate Autism Diagnostics: Pitfalls and promises. Journal of Autism and Developmental Disorders. 45(5), 1121-1136, 2015
2. Daniel Bone, Somer Bishop, Matthew P. Black, Matthew S. Goodwin, Catherine Lord, Shrikanth S. Narayanan. Use of Machine Learning to Improve Autism Screening and Diagnostic Instruments: Effectiveness, Efficiency, and Multi-Instrument Fusion. Journal of Child Psychology and Psychiatry. 57(8): 927-937, August 2016
3. Victor Ardulov, Victor R Martinez, Krishna Somandepalli, Shuting Zheng, Emma Salzman, Catherine Lord, Somer Bishop, Shrikanth Narayanan. Robust Diagnostic Classification and Policies via Q-Learning. Scientific Reports 11, 11730. 2021

Child-centric Conversational Systems: An ongoing endeavor

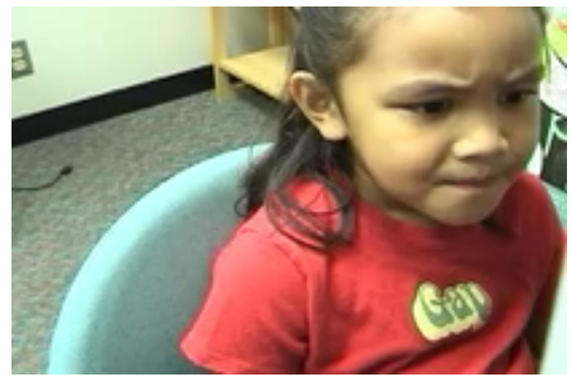
AT&T Bell Labs, 1996

8-14 YEARS

FRUSTRATION

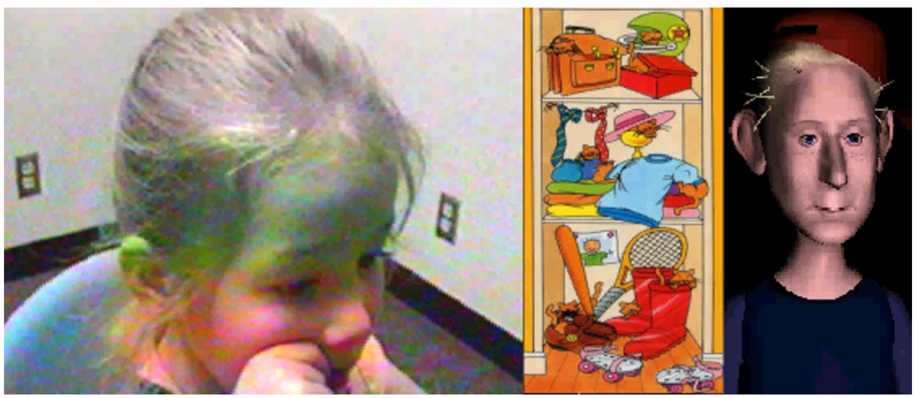
POLITENESS

CONFIDENT VS. UNCERTAIN



USC 2002

PRE-K

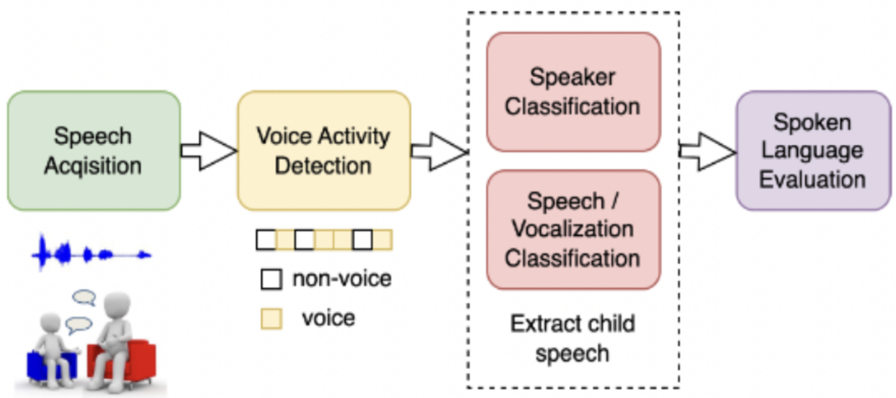


- S. NARAYANAN AND A. POTAMIANOS. CREATING CONVERSATIONAL INTERFACES FOR CHILDREN. IEEE TRANS. SPEECH AND AUDIO PROCESSING, 10(2):65-78, 2002.
- S. YILDIRIM, S. NARAYANAN AND A. POTAMIANOS. DETECTING EMOTIONAL STATE OF A CHILD IN A CONVERSATIONAL COMPUTER GAME. COMPUTER, SPEECH, AND LANGUAGE. SPECIAL ISSUE ON AFFECTIVE SPEECH, 2010.
- E. MOWER, C-C. LEE, J. GIBSON, T. CHASPARI, M. WILLIAMS, S. NARAYANAN. ANALYZING THE NATURE OF ECA INTERACTIONS IN CHILDREN WITH AUTISM. INTERSPEECH, 2011.

NEXT

Assessing language levels in ASD

- **Language is the single best predictor of long-term outcomes in ASD, hence critical to assess and address with interventions**
  - *optimal methods for assessing spoken language in ASD based on **natural spoken language samples***
  - *obtainable during conversational interactions, including elicited using conversational agents and analyzed using speech processing*
- **Scalable deployment leveraging full stack HLT technologies**
  - *stratified and personalized, and across multiple languages*



[1] Amount of intelligible speech offers a strong indicator of children's language capabilities

[2] Requires robust child-centric speech processing for accurately discerning timing information — self supervised methods useful

1. Anfeng Xu, Rajat Hebbar, Rimita Lahiri, Tiantian Feng, Lindsay Butler, Lue Shen, Helen Tager-Flusberg, Shrikanth Narayanan. Understanding Spoken Language Development of Children with ASD Using Pre-trained Speech Embeddings. Proceedings of Interspeech, 2023
2. Rimita Lahiri, Tiantian Feng, Rajat Hebbar, Catherine Lord, So Hyun Kim, Shrikanth Narayanan. Robust Self Supervised Speech Embeddings for Child-Adult Classification in Interactions involving Children with Autism. Proceedings of Interspeech, 2023

Aging related changes in voice and speech characteristics: Elderly speech?

Human voice and speech features manifest information about

- cognitive deficit and slower brain processing
- certain mood states often observed in dementia
- impairments of the neuro-motoric mechanisms of speech production

Aging is associated with several changes in voice patterns

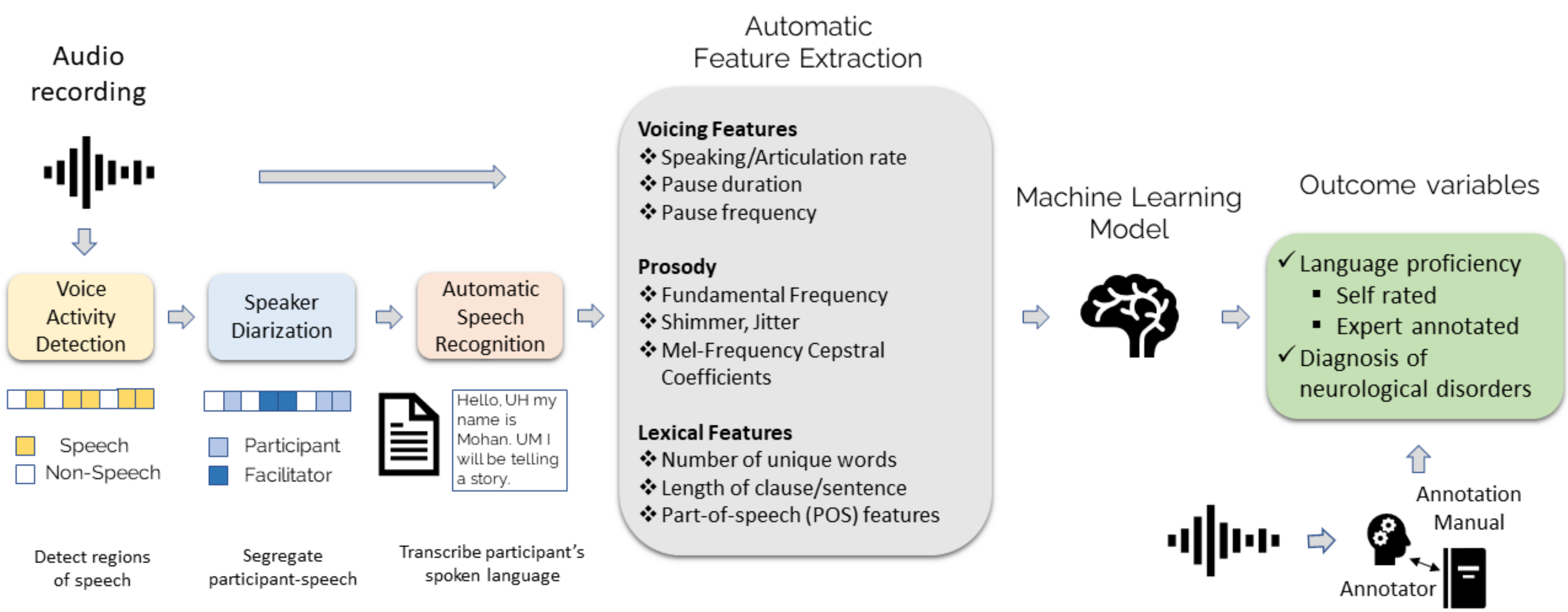
- reduction in vocal range
- reduction in fundamental frequency in females
- increase in fundamental frequency in males
- increased vocal jitter (variation in fundamental frequency) and shimmer (variation in amplitude)

Phonetic and phonological changes in speech

- lowered speech and articulation rate
- increased pause duration between syllables, words and sentences
- increased hesitation and repetition of words/phrases

Tracking symptom severity in dementia

- **Speech and language biomarkers offer a window into (physical) aging *and* cognitive decline**
  - *vocal speech fluency effects, pause duration/frequency, prosody, vocal affective expressions, anomia, or impaired word finding*
  - *role of language background differences e.g., bilingual/multilingual abilities*
- **Scalable, longitudinal deployment through speech technologies**
- **Open questions: interplay with language background including bi/multilingualism, educational and socio-cultural context factors**



Help Fill Gaps

Help Connect Dots

Source; ThinkStock

Twin goals: **Understanding <u>and</u> addressing variability**

- S. NARAYANAN AND P. GEORGIOU. BEHAVIORAL SIGNAL PROCESSING: DERIVING HUMAN BEHAVIORAL INFORMATICS FROM SPEECH AND LANGUAGE. PROCEEDINGS OF THE IEEE. 101(5): 1203 - 1233, 2013.
- D. BONE, C.-C. LEE, T. CHASPARI, J. GIBSON, AND S. NARAYANAN. SIGNAL PROCESSING AND MACHINE LEARNING FOR MENTAL HEALTH RESEARCH AND CLINICAL APPLICATIONS. IEEE SIGNAL PROCESSING MAGAZINE. 34(5): 189-196, SEPTEMBER 2017
- CHI-CHUN LEE, THEODORA CHASPARI, EMILY MOWER PROVOST, SHRIKANTH S. NARAYANAN. AN ENGINEERING VIEW ON EMOTIONS AND SPEECH: FROM ANALYSIS AND PREDICTIVE MODELS TO RESPONSIBLE HUMAN-CENTERED APPLICATIONS. PROCEEDINGS OF IEEE. 2023

Highlight 3

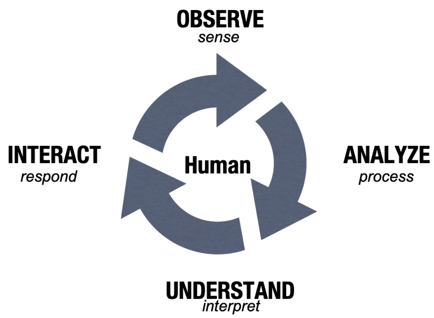
Trustworthy speech processing and creating trusted technologies

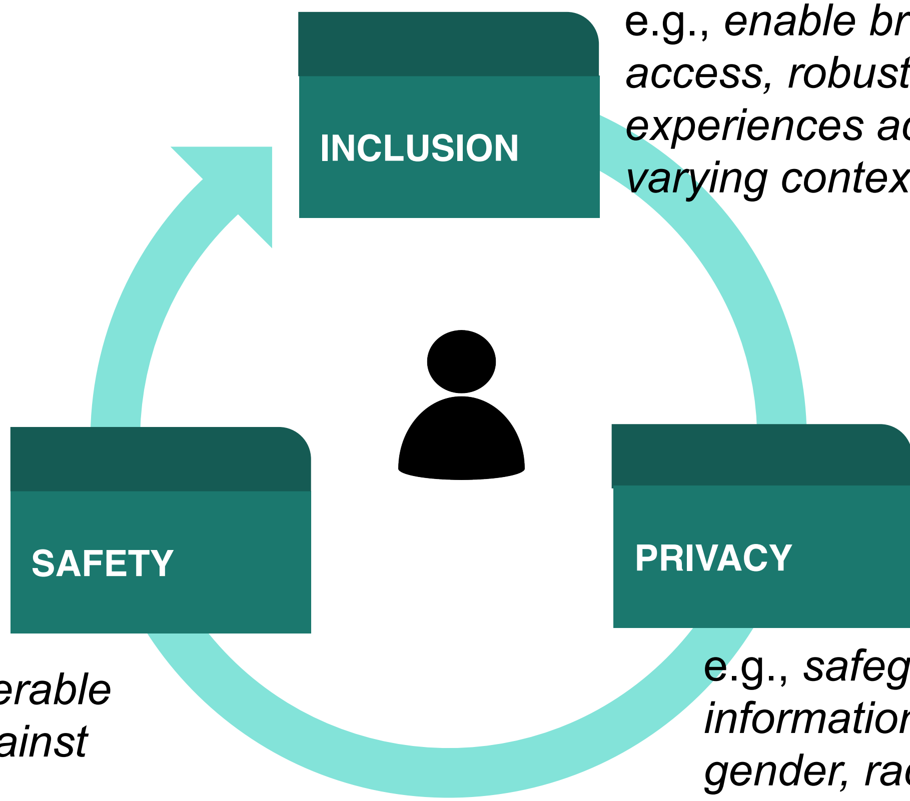
- rich diversity and heterogeneity: people and their contexts and needs
  - broadening access ***and*** presence while personalizing experiences
  - information useful in some contexts may ***not*** be so in others; may be vulnerable to attacks and misuse

- **Arindam Jati, Chin-Cheng Hsu, Monisankha Pal, Raghuveer Peri, Wael AbdAlmageed, Shrikanth Narayanan. Adversarial Attack and Defense Strategies for Deep Speaker Recognition Systems. Computer Speech & Language. 68: 101199, 2021**
- **Raghuveer Peri, Krishna Somandepalli, Shrikanth Narayanan. A study of bias mitigation strategies for speaker recognition. Computer Speech & Language. 79:101481, 2023.**
- **Nicholas Mehlman, Anirudh Sreeram, Raghuveer Peri, Shrikanth Narayanan. Mel frequency spectral domain defenses against adversarial attacks on speech recognition systems. 3(3): 035208. J. Acoust. Soc. Am. Express Letters, 2023**
- **Tiantian Feng, Rajat Hebbar, Nicholas Mehlman, Xuan Shi, Aditya Kommineni, and Shrikanth Narayanan. A Review of Speech-centric Trustworthy Machine Learning: Privacy, Safety, and Fairness. APSIPA Transactions on Signal and Information Processing. 12(3), 2023**

Some elements toward enabling

Trustworthy Human-centered Machine Intelligence





e.g., *enable broad access, robust experiences across varying contexts*

e.g., *safeguard protected information (PII: age, gender, race, etc.)*

e.g., *protect vulnerable people; guard against malicious attacks*

others
explainability
interpretability
transparency
resilience
flexibility
…

102

Credit: DeAndreBush

Highlight 3
Computational Media Intelligence

- understanding media stories, and their impact on human experiences, behavior and action: from individual to socio-cultural scale
- support diversity and inclusion:
  - ***tools for awareness, tools for change***

From inclusive technologies ➡ To technologies for inclusion

- KRISHNA SOMANDEPALLI, TANAYA GUHA, VICTOR MARTINEZ, NAVEEN KUMAR, HARTWIG ADAM AND SHRIKANTH NARAYANAN, COMPUTATIONAL MEDIA INTELLIGENCE: HUMAN-CENTERED MACHINE ANALYSIS OF MEDIA, PROCEEDINGS OF THE IEEE. 109(5): 891-910, 2021

Case study: Quantifying Media Portrayals

- **Understand gender, age, race, appearance, ability representations and portrayals**
  - on screen *and* behind the scenes

- **But can go beyond measuring (unconscious) bias and stereotypes…**
  - Provide insights into positive, societally meaningful portrayals e.g., of STEM
  - Assist creators with analytical tools during the creative process
  - Enable quantitative causal models for decision making

In collaboration with
Geena Davis Institute on Gender in Media
If she can see it, she can be it.™

With support from
Google

Case study: Quantifying Media Portrayals

how is someone seen and heard?
how are they portrayed in the stories we tell?

- **Understand gender, age, r ons**
  - on screen *and* behind t
- **But can go b g (unconscious) bias and stereotypes..**
  - Provid e, societally meaningful portrayals e.g., of STEM
  - A alytical tools during the creative process
  - tive causal models for decision making

In collaboration with

Geena Davis Institute on Gender in Media
If she can see it, she can be it.™

105

Illustration: On-Screen Time, Speaking Time

Screen-time

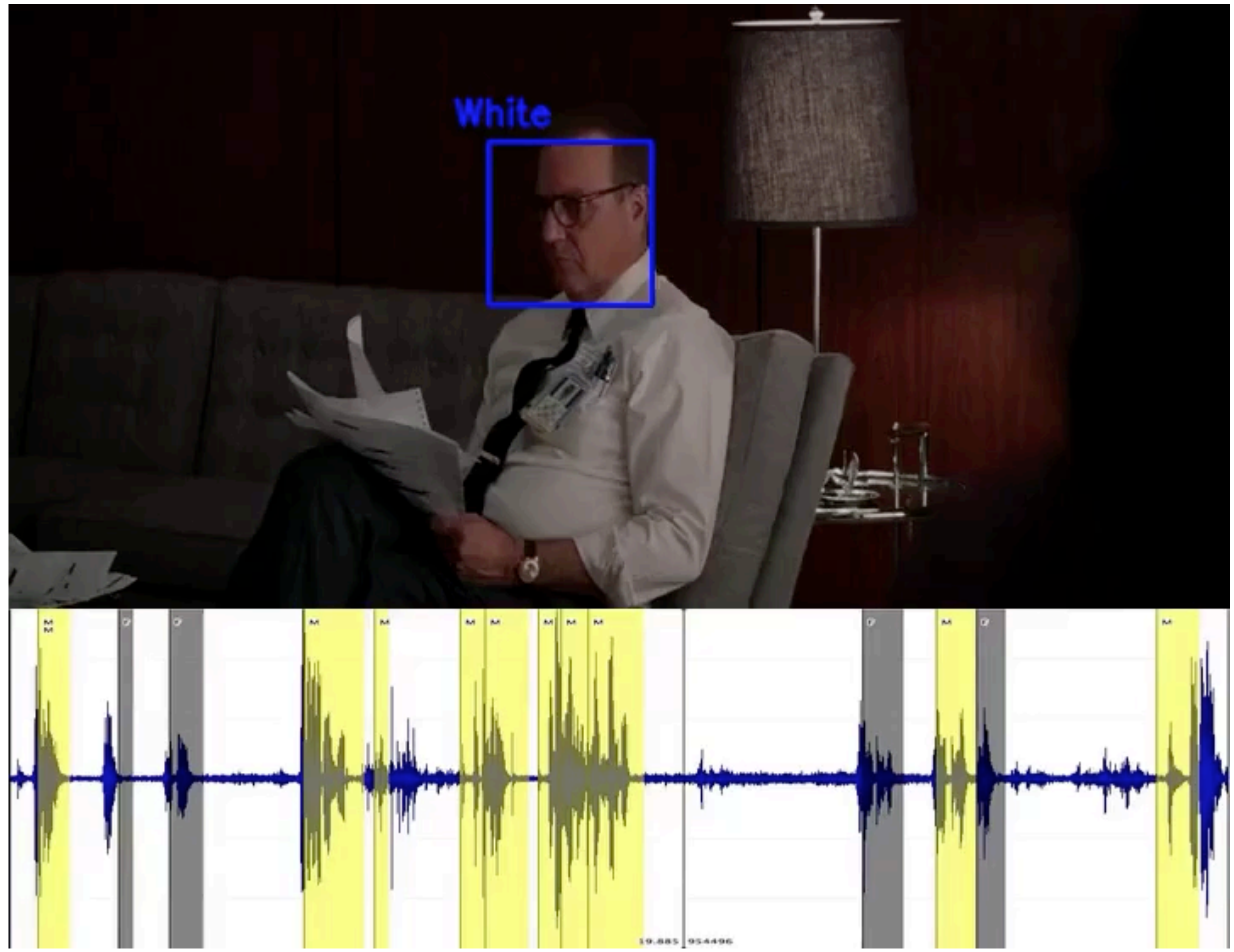
Female face

Male face

Speaking-time

Female speaker

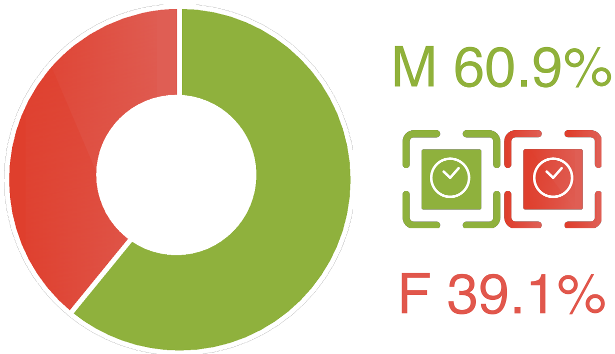
Male speaker



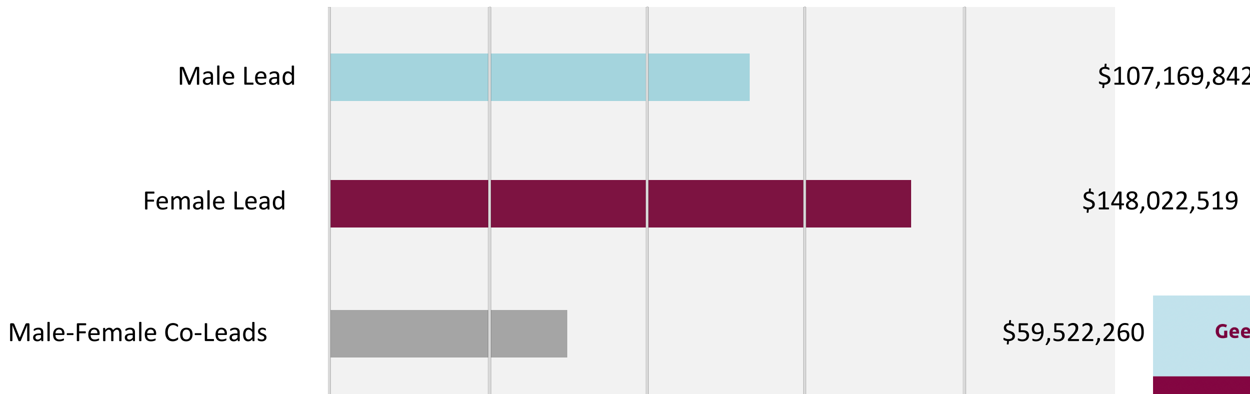
Geena Davis Institute on Gender in Media
If she can see it, she can be it.™

On top grossing live action US "Hollywood" Films for 2017, 2018

2017 Screen Time by Gender (N = 94)



2017 Speaking Time by Gender (N = 94)

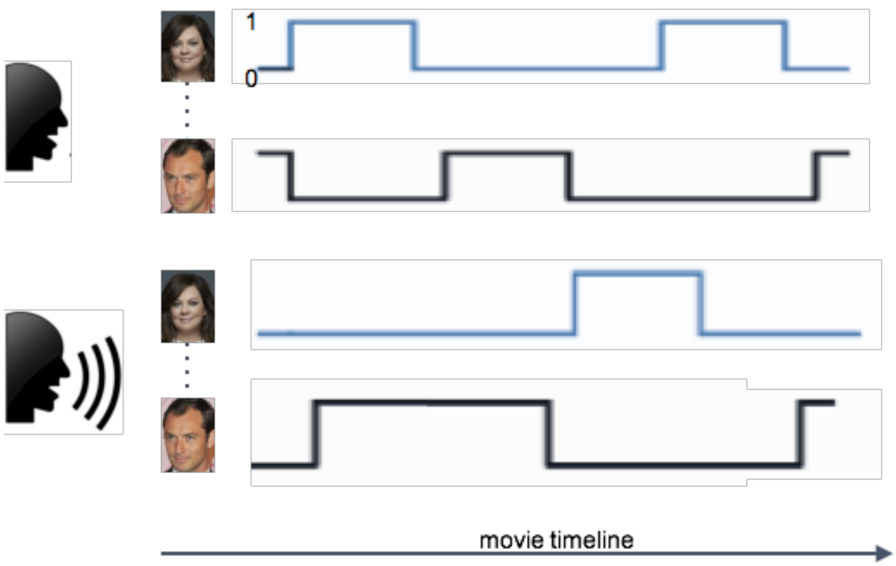


RECENT

2022: Using AI to measure representation, inclusivity in the most popular TV shows of U.S over the past 12 years.

Joint Audio-visual Analysis: Sample insights

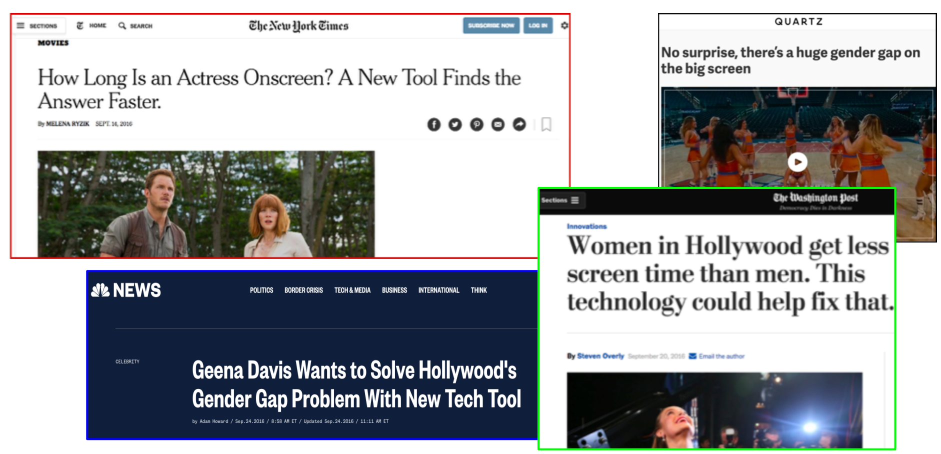
representational disparity



Data from 17 Hollywood blockbusters..

	No face	male face	female face
No speech	26.5%	49.7%	24.8%
male voice	20.9%	51.1%	28.0%
female voice	16.6%	50.4%	33.0%

… seen less even while speaking



108

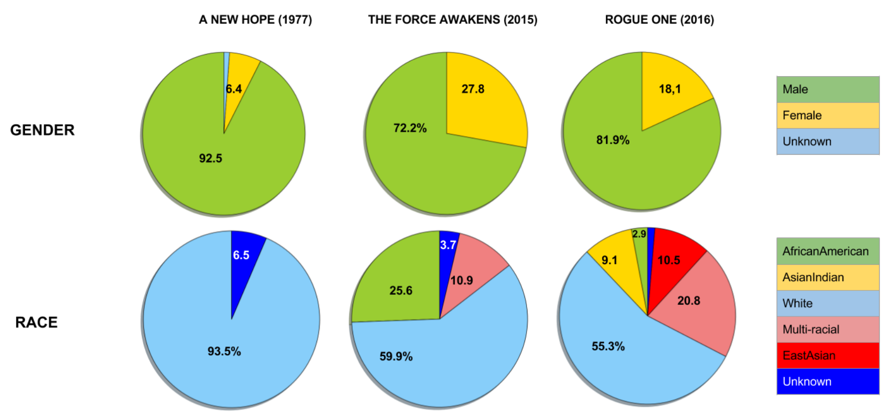
USC University of Southern California

Geena Davis Institute on Gender in Media
If she can see it, she can be it.™

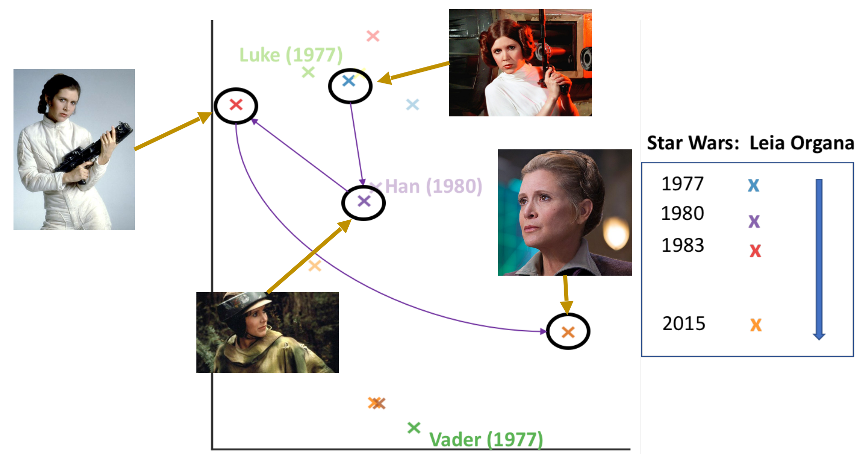
Dialog and Interaction Analytics

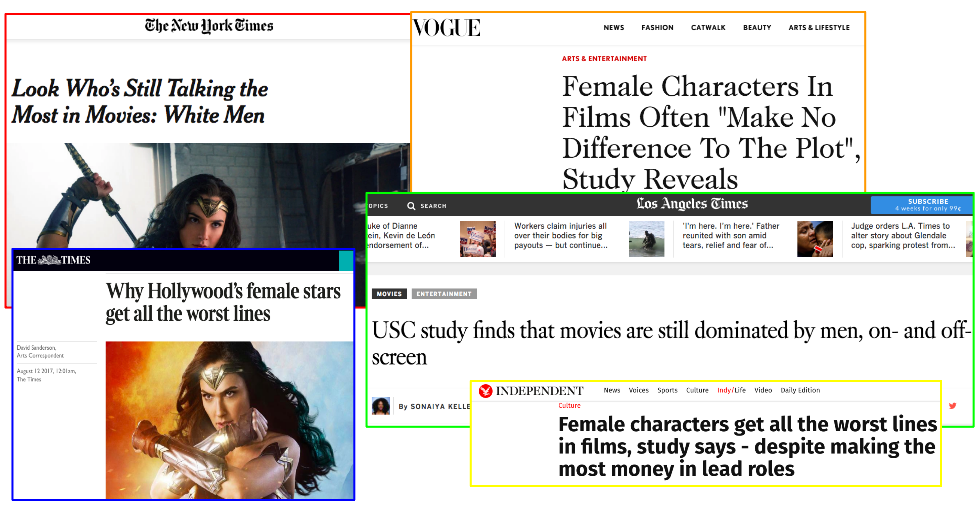
Dialog and interaction language analytics from text documents e.g., scripts, books, subtitles: *who is saying what to whom and how*

Representations over time: A case study of Star Wars trilogy



Distribution of dialogues for race and gender from movie scripts





THE Hollywood REPORTER

MOVIES

Universal Teams With Geena Davis Institute, USC for Software to Increase Latinx Representation

9:00 AM PST 2/19/2020 by Rebecca Sun

109

To probe further — media application references sail.usc.edu/~ccmi

RECENT…

- DIGBALAY BOSE, RAJAT HEBBAR, TIANTIAN FENG, KRISHNA SOMANDEPALLI, ANFENG XU, SHRIKANTH NARAYANAN. MM-AU:TOWARDS MULTIMODAL UNDERSTANDING OF ADVERTISEMENT VIDEOS. PROCEEDINGS OF THE 31ST ACM CONFERENCE ON MULTIMEDIA, 2023
- RAJAT HEBBAR, DIGBALAY BOSE, SHRIKANTH NARAYANAN. SEAR: SEMANTICALLY-GROUNDED AUDIO REPRESENTATIONS. PROCEEDINGS OF THE 31ST ACM CONFERENCE ON MULTIMEDIA (MM'23), 2023
- SABYASACHEE BARUAH, S. NARAYANAN. CHARACTER COREFERENCE RESOLUTION IN MOVIE SCREENPLAYS. PROC. OF FINDINGS OF ACL, 2023
- RAHUL SHARMA, SHRIKANTH NARAYANAN. AUDIO-VISUAL ACTIVITY GUIDED CROSS-MODAL IDENTITY ASSOCIATION FOR ACTIVE SPEAKER DETECTION. IEEE OPEN JOURNAL OF SIGNAL PROCESSING. 4: 225-232, 2023
- DIGBALAY BOSE, RAJAT HEBBAR, KRISHNA SOMANDEPALLI, HAOYANG ZHANG, YIN CUI, KREE COLE-MCLAUGHLIN, HUISHENG WANG, SHRIKANTH NARAYANAN. MOVIECLIP: VISUAL SCENE RECOGNITION IN MOVIES. PROCEEDINGS OF 2023 IEEE/CVF WACV. 2023
- VICTOR MARTINEZ, KRISHNA SOMANDEPALLI, SHRIKANTH NARAYANAN. BOYS DON'T CRY (OR KISS OR DANCE): A COMPUTATIONAL LINGUISTIC LENS INTO GENDERED ACTIONS IN FILM. PLOS ONE. 17(12):1–23, 2022
- RAHUL SHARMA, KRISHNA SOMANDEPALLI, SHRIKANTH NARAYANAN. CROSS MODAL VIDEO REPRESENTATIONS FOR WEAKLY SUPERVISED ACTIVE SPEAKER LOCALIZATION. IEEE TRANSACTIONS ON MULTIMEDIA. 2022
- SABYASACHEE BARUAH, KRISHNA SOMANDEPALLI, SHRIKANTH NARAYANAN. REPRESENTATION OF PROFESSIONS IN ENTERTAINMENT MEDIA: INSIGHTS INT FREQUENCY AND SENTIMENT TRENDS THROUGH COMPUTATIONAL TEXT ANALYSIS. PLOS ONE. 17(5): E0267812. 2022
- KRISHNA SOMANDEPALLI, RAJAT HEBBAR, SHRIKANTH NARAYANAN. ROBUST CHARACTER LABELING IN MOVIE VIDEOS: DATA RESOURCES AND SELF-SUPERVISED FEATURE ADAPTATION. IEEE TRANSACTIONS ON MULTIMEDIA. 24: 3355-3368, 2022

…

FOUNDATIONAL…

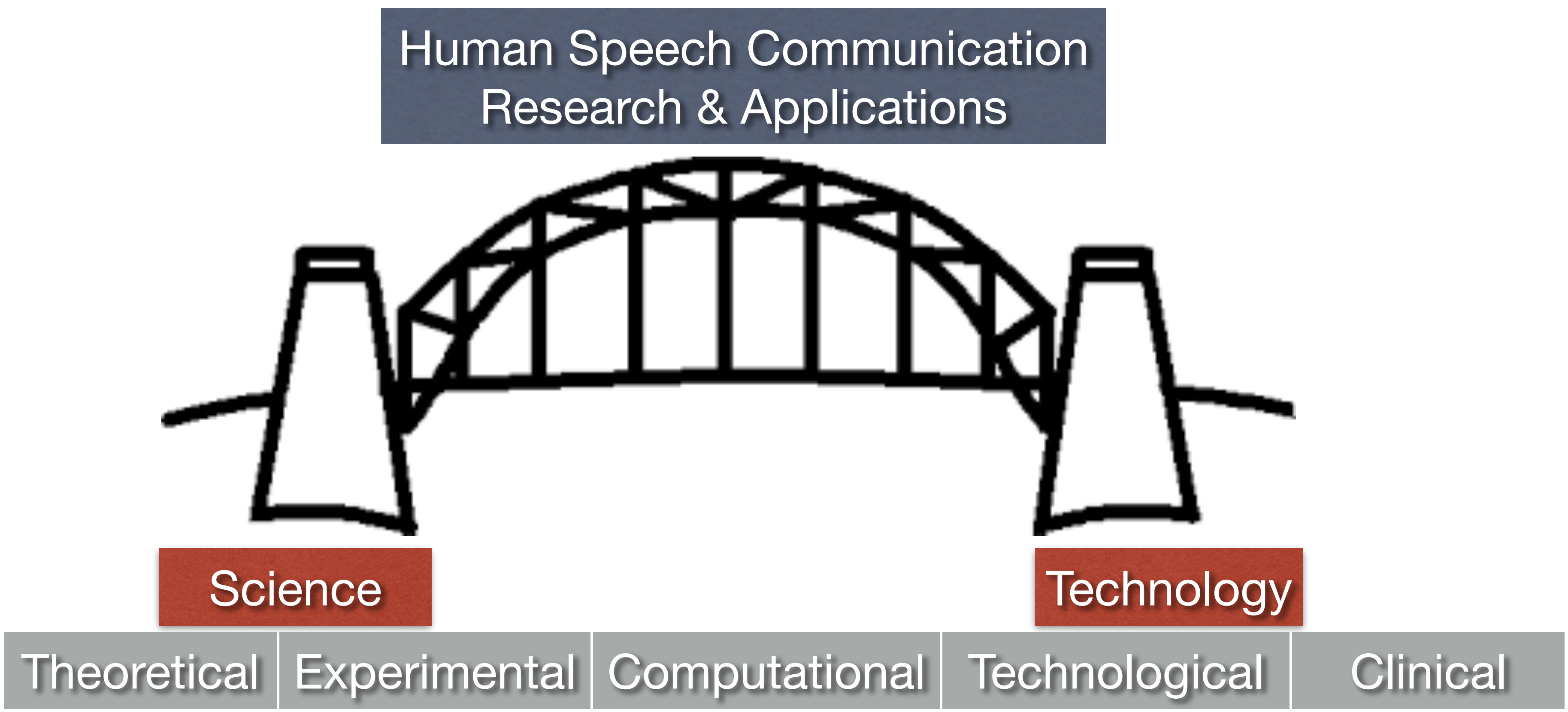
- TANAYA GUHA, CHE-WEI HUANG, NAVEEN KUMAR, YAN ZHU, SHRIKANTH S. NARAYANAN. GENDER REPRESENTATION IN CINEMATIC CONTENT: A MULTIMODAL APPROACH. IN PROCEEDINGS OF 17TH ACM INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION(ICMI), 2015
- ANIL RAMAKRISHNA, NIKOLAOS MALANDRAKIS, ELIZABETH STARUK AND SHRIKANTH NARAYANAN. A QUANTITATIVE ANALYSIS OF GENDER DIFFERENCES MOVIES USING PSYCHOLINGUISTIC NORMATIVES. EMNLP 2015.
- ANIL RAMAKRISHNA, VICTOR R. MARTÍNEZ, NIKOLAOS MALANDRAKIS, KARAN SINGLA AND SHRIKANTH NARAYANAN. LINGUISTIC ANALYSIS OF DIFFERENCE IN PORTRAYAL OF MOVIE CHARACTERS. PROCEEDINGS OF ACL, 2017
- VICTOR MARTINEZ, KRISHNA SOMANDEPALLI, KARAN SINGLA, ANIL RAMAKRISHNA, YALDA UHLS, SHRIKANTH NARAYANAN. VIOLENCE RATING PREDICTION FROM MOVIE SCRIPTS. PROCEEDINGS OF AAAI, 2019
- VICTOR MARTINEZ, KRISHNA SOMANDEPALLI, YALDA TEHRANIAN-UHLS AND SHRIKANTH NARAYANAN. JOINT ESTIMATION AND ANALYSIS OF RISK BEHAVIO RATINGS IN MOVIE SCRIPTS. EMNLP 2020

Summary

✓ENGINEERING INNOVATIONS CAN PROVIDE CRUCIAL ADVANCES IN SPEECH SCIENCE

✓SCIENTIFIC KNOWLEDGE ABOUT SPEECH AND LANGUAGE PROCESSES CAN ENABLE RICH HUMAN CENTERED TECHNOLOGIES THAT CAN IMPACT NUMEROUS SOCIETAL REALMS

✓CREATING TRUSTWORTHY SPEECH PROCESSING ESSENTIAL FOR BROAD TRUSTED USE



Closing thoughts

Speech research is a continuing journey filled with rich and meaningful societal possibilities

- the field rejuvenates perpetually, just getting better in each step: continually leading to new discoveries, technological innovations and an ever expanding reach of applications
- will keep our community ***relevant*** and ***vibrant*** in the global stage for years to come

SUPPORTED BY:
NSF, NIH, DARPA, IARPA, USG AGENCIES, ONR, ARMY,
SIMONS FOUNDATION, GUGGENHEIM FOUNDATION, GOOGLE, APPLE, AMAZON, DISNEY, TOYOTA

Work reported represents efforts of
<u>numerous</u>
colleagues and collaborators
deeply grateful to all of them