The Role of Source and Filter Characteristics in Human Talker Identification: Experiments with Laryngeal and Electrolarynx Speech

Tyler K. Perrachione[1], Cara E. Stepp[2,3,5], Robert E. Hillman[2,3,4], Patrick C.M. Wong[6,7,8]

[1]Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology, [2]Harvard-MIT Division of Health Science & Technology, [3]Massachusetts General Hospital Center for Laryngeal Surgery & Voice Rehabilitation, [4]Department of Surgery, Harvard Medical School, [5]Departments of Computer Science and Engineering & Rehabilitation Medicine, University of Washington, [6]Roxelyn & Richard Pepper Department of Communication Sciences & Disorders, Northwestern University, [7]Northwestern Interdepartmental Neuroscience Program, Northwestern University, [8]Department of Otolaryngology – Head and Neck Surgery, Northwestern University

Abstract

Differences in individuals' vocal anatomy and physiology result in unique acoustic features of their vocalizations. Humans are exceptionally attuned to these variations and use them to identify familiar individuals. Although these abilities are often called "voice recognition", talker identity cues actually arise through interactions between acoustic excitation produced at the source (typically, the larynx) and both static and dynamic properties of the filter (vocal tract, articulators, and their manipulations during speech). We investigated the differential contributions of source- and filter-related information to talker identification through four experiments using laryngeal (typical) and electrolarynx speech from 5 talkers. Using an electrolarynx energy source removed individual differences in vocal anatomy, leaving only unique filter properties for talker identification.

Listeners learned talker identity best from typical, laryngeal speech, which contained both unique source and filter cues. Listeners were also able to learn talker identity from electrolarynx speech, which homogenized talker source characteristics. Curiously, listeners did not generalize talker identity across source mechanisms: Training on laryngeal or electrolarynx speech resulted in chance performance identifying the same talkers using the other source mechanism. We consider the implications of these results for models of talker identification and articulatory compensation during electrolarynx use.

Background

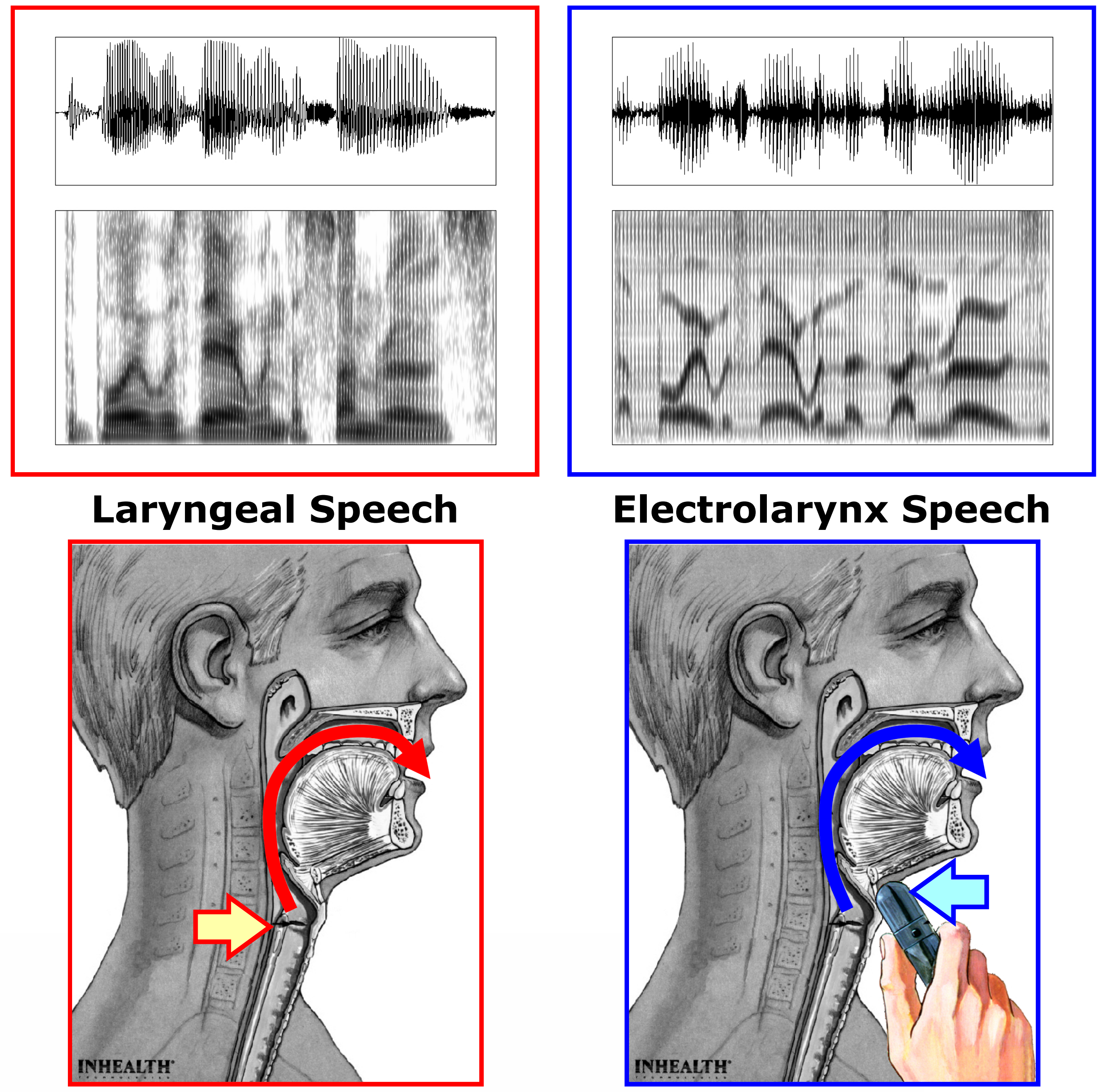
Talker identity is the product of interacting acoustic cues:

- **Source characteristics** of the voicing mechanism
  - Laryngeal anatomy and physiology
  - F_0 and its dynamics, glottal waveform, etc.
- **Filter characteristics** of the vocal tract
  - Anatomy and physiology of the pharyngeal, oral, and nasal cavities
  - Constrain the range of resonance (F1, F2, etc.)
- **Dynamic manipulation** of articulators
  - Socioculturally acquired phonetic features
  - Variation due to language, dialect, idiolect

Carrell, (1984); Perrachione & Wong (2007);

How do source and filter characteristics differentially contribute to perception of talker identity?

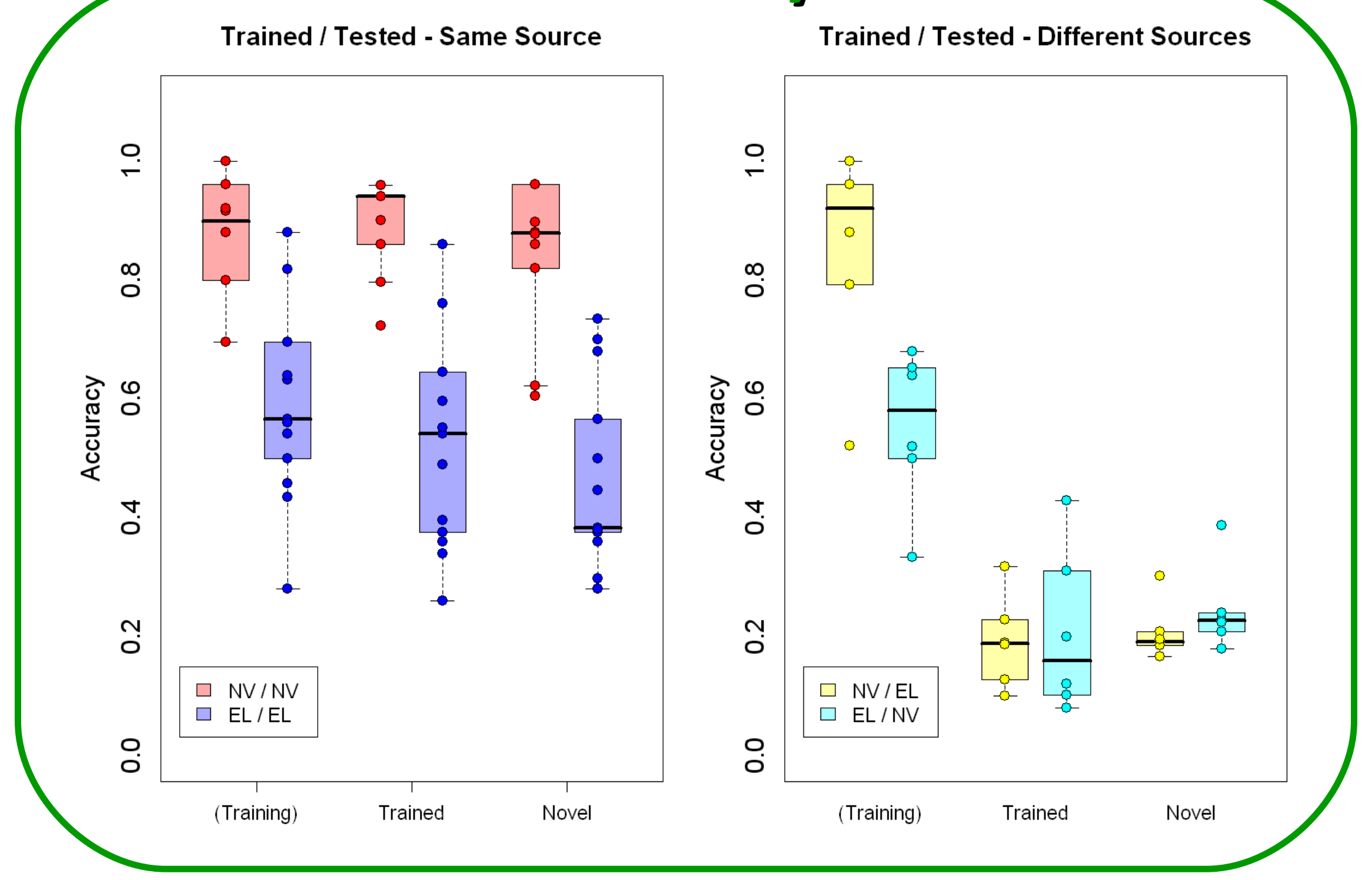
- **Experimental challenge:** how to separate the 1-to-1 correspondence between unique vocal sources and filter characteristics (i.e. within a single talker)
- **Electrolarynx:** a battery-powered device that provides a mechanical voice source through the tissues of the neck
  - Replacement voice source for total laryngectomy patients
  - **Homogenizes source characteristics while preserving individual differences in filter characteristics**
  - e.g., Listeners can still distinguish male/female EL users based on formant spacing (Brown & Feinstein, 1997)


Laryngeal Speech

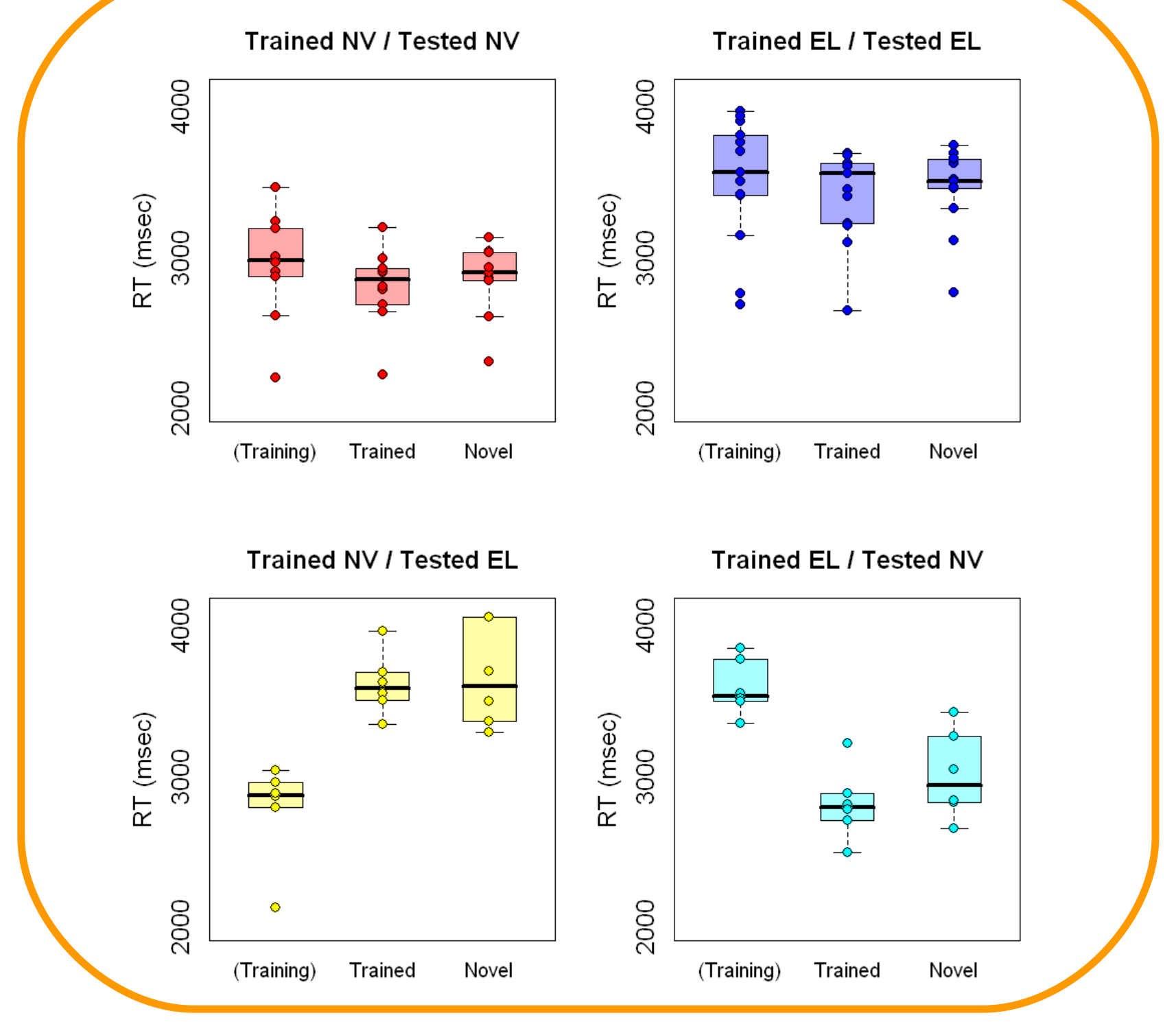
Electrolarynx Speech

Results

Accuracy



Reaction Time



Error Analysis

Train NV / Test NV

Talker	1	2	3	4
5	7	6	10	14
4	12	5	43	
3	13	4		
2	12			

Train EL / Test EL

Talker	1	2	3	4
5	39	25	46	72
4	42	49	117	
3	74	45		
2	111			

Source Mechanism: *"(Training)"*

- Better talker identification from NV training
  - $t(33) = 5.875$, $p < 1.4\times10^{-6}$
- Faster reaction time to NV stimuli
  - $t(33) = -5.444$, $p < 5\times10^{-6}$
  - Evitts & Searle (2006)

Sentence Content: *"Trained"* vs. *"Novel"*

- More accurate ID from trained sentences
  - $F(1,31) = 10.250$, $p < 0.0032$
- Faster RT to trained sentences
  - $F(1,31) = 4711.653$, $p < 2\times10^{-35}$
- No interaction with source mechanism.

Filter-Characteristics Generalization:

- Listeners accuracy was at chance for testing on the untrained source mechanism:
  - NV-EL: $t(5) = -0.009$, $p = 0.993$
  - EL-NV: $t(5) = 0.644$, $p = 0.548$

Patterns of Correct Identification and Errors:

- Across source mechanisms, listeners differed in the talkers they found most identifiable:
  - $\chi^2(4) = 16.703$, $p < 0.0025$
- Patterns of errors across source mechanism showed some similarities, but differed significantly overall
  - $\chi^2(9) = 27.875$, $p < 0.001$

Conclusions:

- **Individuals are uniquely identifiable even from EL speech.**
- **Listeners' perceptual space for talkers differs across source mechanisms (based on identification errors)**
- **Likewise, Training talker identification on one source mechanism does not generalize to speech from the other mechanism.**
  - **Talker identity may emerge as a "gestalt" from a number of underlying cues; is more than the sum of its acoustic parts.**
  - **Talkers may employ different articulatory strategies to increase intelligibility during naïve use of an EL device.**

Methods

Recordings

- 15 male native American English-speakers with no discernable accent
  - Ages 20-38 years, mean = 26.6 years
- 14 sentences (IEEE, 1969) recorded at 50 kHz
  - **Normal laryngeal voice (NV)** and using an **electrolarynx (EL)**
  - TrueTone™ electrolarynx (Griffin Labs), fixed F_0 of 109Hz

Intelligibility Assessment & Stimulus Selection

- 8 native English-speaking listeners judged pairs of EL recordings
  - "Which recording is more intelligible?"
- 210 stimulus pairs (15 talkers × 14 sentences)
  - Recordings in a pair were of the same sentence
  - Each talker was paired against every other talker equally
- Intelligibility rankings determined following Meltzner & Hillman (2005)
  - Most intelligible talker used as an example stimulus
  - Next 5 most intelligible talkers used as stimuli for identification

Talker Identification Training & Testing

Conditions: Parametric training-testing paradigms investigated generalization of talker-identification abilities across source mechanisms

- 🟥 Train NV — Test NV (N = 10)
- 🟦 Train EL — Test EL (N = 13)
- 🟨 Train NV — Test EL (N = 6)
- ▢ Train EL — Test NV (N = 6)

Subjects: Undergraduate students, native speakers of American English, normal speech and hearing, N = 35

Paradigm:

- **Training:** Learn to identify (with feedback) 5 talkers from 5 training sentences with training source mechanism (NV or EL)
- **Generalization & Post-test:** Identify those talkers from both trained and novel sentences with testing source mechanism (NV or EL)

References

Perrachione, T.K. & Wong, P.C.M. (2007) "Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex." *Neuropsychologia*, 45, 1899-1910.

Brown Jr., W.S. & Feinstein, S.H. (1997) "Speaker sex identification utilizing a constant laryngeal source." *Folia Phoniatrica*, 29, 240-248.

Meltzner, G.S. & Hillman, R.E. (2005) "Impact of aberrant acoustic properties on the perception of sound quality in electrolarynx speech." *Journal of Speech, Language, and Hearing Research*, 48, 766-779.

Evitts, P.M. & Searl, J. (2006) "Reaction times of normal listeners to laryngeal, alaryngeal, and synthetic speech." *Journal of Speech, Language, and Hearing Research*, 49, 1380-1390.

Carrell, T.D. (1984) *Contributions of fundamental frequency, formant spacing, and glottal waveform to talker identification.* (Doctoral dissertation). Indiana University, Bloomington, Indiana.

Acknowledgments

The authors wish to thank Yoomin Ahn, Allison Barr, Louisa Ha, Katie Brag, and James Heaton for their assistance in this project.

This research was funded by Massachusetts General Hospital (R.H.) and NIH (USA) grants R01DC008333 & R21DC009652 to P.W.

Contact

Tyler K. Perrachione tkp@mit.edu
http://web.mit.edu/tkp/www/

Cara E. Stepp cstepp@alum.mit.edu
http://faculty.washington.edu/cstepp/