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

Tyler K. Perrachione1, Cara E. Stepp2,3,5, Robert E. Hillman3,4, Patrick C.M. Wong2,6,7
1Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology, 2Harvard-MIT Division of Health Science & Technology, 3Massachusetts General Hospital Center for Laryngeal Surgery & Voice Rehabilitation, 4Department of Surgery, Harvard Medical School, 5Department of Computer and Engineering, Rehabilitation Medicine, University of Washington, 6Rowan & Richard Pepper Department of Communication Sciences & Disorders, Northwestern University, 7Northwestern Interdepartmental Neuroscience Program, Northwestern University, 8Department 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 differences and are able 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 less able to generalize talker identity across source mechanisms: Training on laryngeal or electrolarynx speech resulted in chance performance on talker identification, but 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 voice mechanism (vocal anatomy and physiology, F0, glottal waveform, etc.)
- Filter characteristics of the vocal tract
- Articulatory and phonetic features
- Variations due to individual differences

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

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 F0 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 x 14 sentences)
- Recording in pair was the same talker
- Each talker was paired against every other talker equally
- Intelligibility rankings determined following Mehlman & Hillman (2005)
- Most intelligible talker used as an example stimulus
- Next 5 most intelligible talkers used as stimuli for identification

Source Mechanism: 

T("Training")
Better talker identification from NV tranings

- t(33) = 5.875, p < 1.4 x 10^-10
- Faster reaction time to NV stimuli
- t(33) = -5.446, p < 10^-10
- Evans & Gareau (2008)

Sentence Content:

T("Trained") vs. "Novel"
More accurate ID from trained sentences
FL(1) = 4711.653, p = 2 x 10^-35
Faster RT to trained sentences
FL(1) = 4711.603, p = 2 x 10^-30
No interaction with source mechanism.

Error Analysis

Train EL / Test EL
42
49
13
12
7
6
4
3
2
1
4
3
2
1
1

Pattens of Correct Identification and Errors:

- Across source mechanisms, listeners differed in the talkers they found most identifiable:

- Patterns of errors across source mechanism showed some similarities, but differed significantly overall:

- t(4) = 16.703, p < 0.0025

- Across source mechanisms, errors differed in the talkers they found most identifiable:

- t(4) = 27.857, p < 0.001

Results

Accuracy

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 EV – Test EL (N = 13)
- Train EL – Test NV (N = 6)

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

Paradigms:

- 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)

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 “guardian” 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 naive use of an EL device.

References

Brown Jr., W.S. & Feinstein, S.H. (1997) “Speaker sex identification on one source mechanism does not generalize to speech from the other mechanism.”


Evitts & Searle (2006)


Evitts & Searle (2006)

Acknowledgments

The authors would like to express their appreciation to Cara E. Stepp, Robert E. Hillman, and James Heaton for their assistance in this project.

Contact

Tyler K. Perrachione
tkper@mit.edu

Cara E. Stepp
castepp@alum.mit.edu

NIH (USA) grants R01DC008333 & R21DC009652 to P.W.

Cara E. Stepp

http://faculty.washington.edu/castepp/

http://faculty.washington.edu/castepp/

http://faculty.washington.edu/castepp/

http://faculty.washington.edu/castepp/

http://faculty.washington.edu/castepp/