HHS Public Access

Author manuscript

J Speech Lang Hear Res. Author manuscript; available in PMC 2015 November 23.

Published in final edited form as:

J Speech Lang Hear Res. 2014 October ; 57(5): 1651–1665. doi:10.1044/2014_JSLHR-S-13-0161.

Talker identification across source mechanisms: Experiments with laryngeal and electrolarynx speech

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

[1]Boston University

[2]Massachusetts Institute of Technology

[3]Harvard University

[4]Massachusetts General Hospital

[5]The Chinese University of Hong Kong

[6]Northwestern University

Abstract

Purpose—To determine listeners' ability to learn talker identity from speech produced with an electrolarynx, explore source and filter differentiation in talker identification, and describe acoustic-phonetic changes associated with electrolarynx use.

Method—Healthy adult control listeners learned to identify talkers from speech recordings produced using talkers' normal laryngeal vocal source or an electrolarynx. Listeners' abilities to identify talkers from the trained vocal source (Experiment 1) and generalize this knowledge to the untrained source (Experiment 2) were assessed. Acoustic-phonetic measurements of spectral differences between source mechanisms were performed. Additional listeners attempted to match recordings from different source mechanisms to a single talker (Experiment 3).

Results—Listeners successfully learned talker identity from electrolarynx speech, but less accurately than from laryngeal speech. Listeners were unable to generalize talker identity to the untrained source mechanism. Electrolarynx use resulted in vowels with higher F1 frequencies compared to laryngeal speech. Listeners matched recordings from different sources to a single talker better than chance.

Conclusions—Electrolarynx speech, though lacking individual differences in voice quality, nevertheless conveys sufficient indexical information related to the vocal filter and articulation for listeners to identify individual talkers. Psychologically, perception of talker identity arises from a "gestalt" of the vocal source and filter.

INTRODUCTION

The voice plays many important roles in human social behavior, beyond being the principal acoustic source for speech communication. In particular, listeners are sensitive to the individually distinctive properties of a talker's voice and speech, and we use this information to discern who is speaking in a process called *talker identification*. Scientifically, much remains unknown about the perceptual and psychological representation of vocal identity,

such as how it is robust or susceptible to manipulations of the acoustic signal, and whether it can be decomposed into individually distinctive constituent parts. Clinically, despite the profound social importance of vocal identity, much is still unknown about how vocal identity is preserved or compromised in talkers who rely on other acoustic sources besides the larynx to produce speech.

Certain medical conditions (usually advanced laryngeal cancer) may necessitate that individuals undergo a laryngectomy, in which the larynx is removed and the vocal filter is separated from the pulmonary source of acoustic energy. Despite removal of the endogenous vocal source, speech production ability can be restored following laryngectomy by use of an electrolarynx – an external device that can be placed against the soft tissue of the neck and vibrates as an alternative source of vocal energy during articulation. This alaryngeal voice source is a typical modality for individuals immediately after surgery, and continues to be a viable option in the long-term for up to 50% of laryngectomees (Hillman et al., 1998; Koike et al., 2002; Lee, Gibson, & Hilari, 2010; Robertson et al., 2012; Eadie et al., 2013; Varghese et al., 2013), particularly among those who do not adopt esophageal or tracheoesophageal speech. Although electrolarynx users have lowered intelligibility, recent work indicates that global quality of life does not significantly differ as a function of alaryngeal speech mode (Eadie et al., 2013; cf. Moukarbel et al., 2011). Healthy individuals can likewise learn to produce speech using an electrolarynx by foregoing typical laryngeal phonation during articulation.

Comparatively little is known about the perception of electrolarynx speech, including whether listeners are sensitive to subtle phonetic information in this modality (Weiss & Basili, 1985), and whether electrolarynx speech produced by different individuals contains sufficient cues to convey those talkers' unique vocal identity. There is only one previous report investigating the indexical information conveyed by electrolarynx speech produced by different individuals (Coleman, 1973). In that study, listeners' performed with a high degree of accuracy at discriminating whether pairs of electrolarynx speech recordings were produced by the same or different talkers, suggesting that some indexical information may be present to facilitate talker identification. However, there has heretofore been no investigation of listeners' ability to identify individual talkers from their electrolarynx speech – much less whether knowledge of talker identity can generalize across a change in source mechanism.

In the present report, we investigate the ability of naïve listeners to learn to identify talkers based on speech produced either with an electrolarynx or with the laryngeal vocal source. We also examined whether listeners' knowledge of talker identity based on speech from one vocal source could generalize to speech produced with the other source mechanism. This line of research was undertaken with dual goals in mind. First, we wished to ascertain the extent to which individuals who use an electrolarynx to produce speech evince a unique vocal identity. Second, we sought to better understand the perceptual processes involved in talker identification. In particular, electrolarynx speech allows us to explore the unique contributions of the vocal source and filter to talker identity, including whether they are perceptually separable from the combined representation of the source and filter together. If listeners are able to learn to recognize talkers from electrolarynx speech, this would not only

demonstrate that an important social expression of individuality was maintained following laryngectomy, but would also reveal that listeners are sensitive to individually distinguishing features of the vocal filter – independent of homogeneity in the vocal source. Whether listeners can generalize talker identity across changes in source mechanism will additionally reveal the extent to which source and filter contributions to talker identity are perceptually separable when both are present in the speech signal.

Previous research investigating which acoustic features contribute to the perception of talker identity has described a wide range of possible, disparate cues, including properties of the vocal source mechanism (Carrel, 1984; Walton & Orlikoff, 1994), structural properties of pharyngeal, oral, and nasal cavities comprising the vocal filter (Baumann & Belin, 2008; Lavner, Gath, & Rosenhouse, 2000), and dynamic, learned manipulations of the source and filter involved in producing speech (Remez, Fellowes, & Rubin, 1997; Perrachione, Chiao, & Wong, 2010). Listeners have variously demonstrated the ability to identify talkers from either source information (Carrel, 1984) or filter information (Remez, Fellowes, & Rubin, 1997) presented in isolation. However, it is tenuous to assert the ecological significance of any of these features for two principal reasons. First, outside the laboratory, a listener very rarely encounters a vocal filter separate from its source, or *vice versa*, and it remains largely unknown how talker identification involves the perceptual integration of multiple cues – *i.e.*, whether the perceptual distinctions based on one acoustic property are retained following changes to other properties (Lavner, Gath, & Rosenhouse, 2000). Second, studies of voice perception often employ only short samples of isolated vowels as stimuli (*e.g.*, Baumann & Belin, 2008; Latinus & Belin, 2011a), presenting voices in an impoverished form compared to how they are typically encountered (by reducing dynamic properties of running speech) and potentially leading listeners to adopt acoustically-general (as opposed to voice-specific) perceptual strategies.

The question of whether talker identification is possible from electrolarynx speech allows the parallel investigation of the extent to which contemporaneous acoustic cues to talker identity from the source and filter are perceptually separable. By training healthy adult controls to produce speech with an electrolarynx, we were able to generate speech samples in which information relating to the vocal source differed, but information relating to the filter was presumably held constant. This design not only allowed us to answer the clinical and scientific questions posed above, it additionally afforded the opportunity to investigate what acoustic-phonetic changes are associated with electrolarynx use – an opportunity not usually possible with electrolarynx use in a clinical setting – and how these changes might affect listeners' ability to perceive speech and identify talkers. All together, these questions help inform how electrolarynx use impacts speech phonetics, as well as the ways in which perception of electrolarynx speech is fundamentally similar to the perception of laryngeal speech for both linguistic and paralinguistic information, with social and quality of life implications for laryngectomee users of electrolarynges.

EXPERIMENT 1: TALKER IDENTIFICATION FROM LARYNGEAL AND ELECTROLARYNX SOURCES

Background

There is considerable reason to believe that electrolarynx speech is amenable to conveying talker identity. Although the properties of the vocal source are relatively homogenized across talkers using an electrolarynx, this mode of speech preserves a wide range of filter-related and dynamic articulatory cues known to vary between talkers, including the acoustic characteristics of vowels (Hillenbrand et al., 1995), speech rate (Voiers, 1964), pronunciation, and dialect (Perrachione, Chiao, & Wong, 2010). Evidence that listeners can make use of specifically filter-related cues in determining talker identity comes from work showing accurate categorization of talkers' sex from electrolarynx speech (Brown & Feinstein, 1997). Moreover, listeners have been shown to demonstrate a high degree of accuracy when discriminating two talkers when both produce speech using an electrolarynx (Coleman, 1973). However, it is so far unknown whether sufficient cues are available in the filter-only information of electrolarynx speech for listeners to learn individual talker identity (*e.g.*, formant frequencies and bandwidths corresponding to talkers' oral and pharyngeal volumes, time-varying spectral modulations corresponding to talkers' idiosyncratic patterns of articulation, etc.), and whether they can generalize this identity to novel utterances produced by the same talker. In Experiment 1 we investigated the questions: (1) whether listeners can learn talker identity from speech produced using an electrolarynx, (2) whether learned talker identity generalizes across changes in the linguistic content of utterances, and (3) how learning talker identity from electrolarynx speech compares to learning talker identity from speech produced with the laryngeal vocal source.

Methods

Stimuli—A set of 20 sentences was recorded in this experiment (lists 2 and 8 from the "Harvard sentences"; IEEE, 1969, see Appendix A). These sentences were read by 17 males who were recruited for this study as healthy, native-speakers of standard American English (aged 20–38 years, mean = 26.6 years). One participant was unable to produce speech with the electrolarynx without simultaneous aspiration and was excluded from the study. Another participant’s recording was marred by technical difficulties. Consequently, samples from a total of 15 talkers were acquired for this study. All talkers reported having no history of speech, hearing, or language disorder, and talkers were homogeneous with respect to regional accent. Participants gave informed, written consent to provide recordings for use in these experiments as approved by the Institutional Review Board at Massachusetts General Hospital.

Acoustic signals were collected from a headset microphone (AKG Acoustics C 420 PP) and recorded digitally (50 kHz sampling rate) with Axon Instruments hardware (Cyberamp 380, Digidata 1200) and software (Axoscope) while the speakers were seated in a sound-treated room. Recordings made with Axoscope were converted to ".wav" format, parsed, and amplitude-normalized (waveform peak) by token using Adobe Audition® software.

Talkers were first asked to read the sentences aloud with their laryngeal voice to gain familiarity with the reading material. The participants were then shown a TruTone™ electrolarynx (Griffin Labs) that had its pitch-modulation capabilities disengaged. The fundamental frequency of this device was fixed at a constant value of 109 Hz. The talkers were instructed on the use of this device and were allowed to practice using it to produce speech of their choosing until they were comfortable with its use (5–10 minutes). General instruction on the use of the electrolarynx included helping the individual find the best neck placement location and explaining how to maintain a closed glottis during speech production. Although it is possible to produce EL speech without a closed glottis, doing so improves intelligibility in speakers with intact anatomy. Two recordings were then made of the participants reading the set of sentences: 1) using their laryngeal voice and 2) using the electrolarynx. Talkers were instructed in both cases to read the sentences "as naturally as possible."

Stimulus Selection—From among the 15 talkers, those who made the most intelligible recordings with the electrolarynx were selected for use in the perceptual experiments. A high-intelligibility criterion was used for selecting talkers in order to facilitate listeners' ability to understand the speech content and thereby use speech-based cues to talker identity, as well as to reduce the use of global between-talker differences in intelligibility as a cue to identity. In a brief listening experiment, 8 additional participants (age 18–29, M = 21.1 years, 6 female), who were drawn from the same population and met the same inclusion criteria as listener participants in Experiments 1–3, were recruited to judge the intelligibility of the samples. In a self-paced, computer-based paradigm, listeners heard pairs of electrolarynx recordings and indicated which of the two recordings was more intelligible. Each pair consisted of the same sentence produced by two different talkers. Talkers were paired equally often with every other talker, and listeners heard 14 sentences produced by each of the 15 talkers an equal number of times, resulting in 210 stimulus pairs. Stimulus presentation and listening environment were as in Experiment 1, below.

The relative intelligibility of each talker (Fig. 1) was determined based on the scaled probability of his being judged as more intelligible than each of the other talkers, following the procedures described in Meltzner & Hillman (2005). This value reflects the normalized probability (z) across listeners that a given talker will be selected as more intelligible than another talker, averaged over all possible talker pairs, with the scale then shifted so that the least intelligible talker has a value of zero. The ordering of talkers resulting from this procedure is equivalent to using the raw number of times each talker was selected as the “more intelligible” across all listeners. However, in addition to producing a rank-order of intelligibility, this procedure also has the advantage of revealing the magnitude of those intelligibility differences. Correspondingly, one talker had a markedly higher intelligibility ranking than the others, so his stimuli were reserved for use in familiarizing naïve listeners with electrolarynx speech. The next five most highly-ranked talkers had similar scaled intelligibility ratings and were selected for use as stimuli in the perceptual experiments. Recordings from the remaining nine talkers were not used in any of the subsequent perceptual experiments.

Participants—Young adult native speakers of American English (*N* = 25, age 18–21, M = 19.6 years, 15 female) provided informed written consent to participate in this study, as approved by the Northwestern University Institutional Review Board. All participants reported having normal speech and hearing, being free from psychological or neurological disorders, and having no prior experience with electrolarynx speech.

Procedure—Participants in the listening experiment were assigned to one of two experimental conditions, in which they learned to identify the 5 talkers based on recordings of their speech produced either using their laryngeal voice (*N* = 12), or the electrolarynx (*N* = 13). The perceptual experiments were conducted in a sound-attenuated chamber using a self-paced, computer-based paradigm. Stimulus presentation was managed using E-Prime 1.2 (Psychology Software Tools, Inc.) via a SoundBlaster Audigy NX USB sound card and Sennheiser HD250-II circumaural headphones. An experimental session was divided into training, training assessment, and testing phases. This design has been shown to produce effective learning of talker identity in a short experimental session (Perrachione & Wong, 2007). The Electrolarynx condition was preceded by an additional familiarization phase, in which listeners had the opportunity to gain exposure to electrolarynx speech and learned to recognize speech from this device. In total, familiarization, training, and testing were completed in approximately 45 minutes.

Electrolarynx speech familiarization: Before beginning to learn talker identity, participants in the Electrolarynx condition were first familiarized with the nature of electrolarynx speech. Listeners heard 5 sentences produced by the talker with the highest intelligibility rating (see "Stimulus Selection", above), while the text of that sentence was displayed on the screen. After hearing each recording, participants were given the option to repeat the same sentence (if they had difficulty discerning its content) or to proceed to the next sentence. Neither the talker, nor the content of any of the familiarization sentences, appeared in the training experiment.

Training phase: In both the Electrolarynx and Laryngeal Voice conditions, listeners learned to recognize the talkers through paired blocks of passive and active training. Each passive training block consisted of ten trials: Listeners heard two repetitions of each of the 5 talkers' recordings of a single sentence while a number (1–5) designating that talker appeared on the screen. In the subsequent active training block, listeners heard the recordings again and indicated which of the 5 talkers they were hearing by button press. Participants received corrective feedback, indicating whether they had identified the correct talker or, if incorrect, what the correct response should have been. Training blocks were repeated for 5 training sentences, for a total of 50 passive and 50 active training trials.

Training assessment phase: Following the five pairs of passive-active training blocks, participants underwent a 'practice test', in which they identified each of the five talkers saying each of the five training sentences, for a total of 25 trials presented in random order. Participants continued to receive corrective feedback as in the Training phase, and overall performance during the Training Assessment phase was used as an index of training attainment.

Test phase: Following the Training Assessment phase, listeners underwent a final talker identification test. The test consisted of the five talkers' recordings of the five training sentences, as well as an additional five novel test sentences, each with two repetitions presented in a random order, for a total of 100 trials (50 trained, 50 novel trials). Participants proceeded from trial to trial without receiving any feedback during the Test phase.

Results

Participants' ability to learn talker identity from either an electrolarynx or laryngeal vocal source was assessed by their accuracy on the Training Assessment and Test phases of each condition (Fig 2). In the Training Assessment phase, listeners were successful at learning talker identity in both the Laryngeal Voice (mean accuracy = 87.6%, *s.d.* = 10.7%) and Electrolarynx (M = 57.9%, *s.d.* = 15.9%) conditions. Participants successfully learned talker identity from electrolarynx speech, performing significantly better than chance in this condition [one-sample t-test vs. chance (20%); $t_{12} = 8.58$, $p < 2\times10^{-6}$, two-tailed; Cohen's d = 2.38]. Nonetheless, the Laryngeal Voice condition produced significantly more accurate learning of talker identity than the Electrolarynx condition [independent-sample t-test; t_{23} = 5.42, $p < 0.00002$, two-tailed; d = 2.26].

Performance in the Test phase was analyzed using a 2×2 repeated-measures analysis of variance with *condition* (Laryngeal Voice vs. Electrolarynx) as the between-participant factor and *content familiarity* (Trained vs. Novel) as the within-participant factor. As in the Training Assessment, participants were significantly more accurate when identifying talkers' from their laryngeal voice (M = 87.8%, *s.d.* = 9.8%) than the electrolarynx (M = 50.3%, *s.d.* = 16.7%) [main effect of condition; $F_{1,23} = 45.35$, $p < 7.2\times10^{-7}$, $\eta^2 = 0.646$]. Participants were also more accurate identifying the talkers from the trained (M_{LV} = 89.6%, *s.d.* = 7.6%; M_{EL} = 53.0%, *s.d.* = 18.6%), as opposed to novel (M_{LV} = 85.8%, *s.d.* = 12.6%; M_{EL} = 47.6%, *s.d.* = 16.0%), sentences [main effect of content familiarity; $F_{1,23} = 13.31$, $p < 0.0081$, $\eta^2 = 0.027$]. Novel sentence content did not disproportionately affect talker identification in one source mechanism compared to the other [no condition × content interaction; $F_{1,23} = 0.25$, $p = 0.62$].

Discussion

Listeners are able to accurately learn talker identity from electrolarynx speech, despite homogeneity in the source mechanism (i.e., when talkers lack individuating characteristics of the vocal source: breathy/creaky voice quality, fundamental frequency, pitch dynamics, etc.) Additionally, they are able to generalize their knowledge of talker identity to novel electrolarynx speech produced by the same talker. This result suggests that the idiosyncratic phonetics of individuals who chronically use an electrolarynx may be sufficient to distinguish them from other electrolarynx users. Moreover, we intentionally limited the range of variation in the electrolarynx condition to a single source mechanism with a fixed fundamental frequency across all talkers, which severely limited variations in the source between speakers. Source characteristics were highly homogeneous, although not necessarily identical due to subtle differences in the harmonic structure of the source resulting from potential differences in neck anatomy (e.g., Meltzner, Kobler, & Hillman, 2003), the location and pressure of the electrolarynx head against speakers' skin, and

whether or not speakers were able to maintain a closed glottis throughout productions. (Moreover, such variability was likely attested not only between talkers, but also among the various recordings of each individual). Listeners' ability to learn talker identity under this conservative design demonstrates that talkers do exhibit a unique vocal identity when producing speech with an electrolarynx, even when using the same device model and configuration. Given the range of electrolarynx technologies available today, additional degrees of freedom, including pitch range and dynamic pitch, will certainly afford enhanced vocal individuality among clinical users (Liu & Ng, 2007).

Talker identification from electrolarynx speech was less accurate than from typical, laryngeal speech – a predictable result given the reduced degrees of freedom (no differences in pitch or pitch dynamics, voice quality, *etc*.) available in that condition. Listeners have a lifetime of experience recognizing talkers from the rich acoustic complexity of speech, and reducing the available cues by removing variability due to differences in the source mechanism incurs a corresponding decrement in identification ability. Additionally, the talker identification abilities we measured here in a five-alternative forced-choice paradigm (58% correct identification) were not as accurate as those measured previously in a simple paired-sample discrimination paradigm (90% correct discrimination, Coleman, 1973). This difference in performance may be the result of differing mnemonic demands of these two tasks, with the present study requiring listeners to learn and maintain the identity of five different talkers throughout the experiment, whereas the Coleman (1973) task required participants to make discrimination judgments *de novo* on each trial. Previous work has studied the effects of mnemonic demands (Legge, Grosmann, & Pieper, 1984) on talker identification, and how talker identification vs. discrimination may differ in their cognitive and perceptual demands (Van Lancker & Kreiman, 1987). Chance performance between the two task designs likewise differs (20% vs. 50%). However, the level of accuracy observed in Experiments 1 likely does not represent the best possible performance we might expect from listeners when identifying talkers from electrolarynx speech. Additional familiarity with electrolarynx speech may further enhance listeners' ability to recognize individual electrolarynx users (*e.g.*, Knox & Anneberg, 1973). Indeed, participants in the present study had only a modest amount of time to learn to recognize these previously unknown talkers from an unfamiliar source mechanism. Previous research with unfamiliar foreign-accent and unfamiliar foreign-language speech has shown that talker identification in these situations improved with increasing exposure to the target talkers (Perrachione & Wong, 2007; Winters, Levi, & Pisoni, 2008).

It is also interesting to note that there was no source by content interaction in the data. This raises the possibility that a reduction in listener accuracy to novel sentence content arises due to similar cognitive processing of both electrolarynx and laryngeal speech: namely, that listeners use the consistent patterns of filter-related phonetics associated with the speech of different individuals to construct representations of vocal identity (Perrachione & Wong, 2007; Perrachione, Del Tufo, & Gabrieli, 2011).

EXPERIMENT 2: TALKER IDENTIFICATION ACROSS SOURCE MECHANISMS

Background

The ability of listeners to successfully learn talker identity from electrolarynx speech, presumably by learning consistent differences between talkers in the dynamic and structural properties of the vocal filter, raises the possibility that listeners may be able to generalize this knowledge to speech produced by the same talkers using their laryngeal source. Likewise, the same distinctive filter information is presumably learned during talker identification from laryngeal speech, and listeners who learn talker identity in this way may also be able to generalize their knowledge of talkers' filters to electrolarynx speech, even though it does not preserve vocal source characteristics. Alternately, listeners may not encode talker identity as a collection of separate, transferable features, but may instead form a "vocal gestalt" which depends on a complete, undifferentiated set of features for establishing vocal identity (Kreiman & Sidtis, 2011). In this experiment, we assess whether listeners are able to generalize knowledge of talker identity learned from one source mechanism to the same talkers' speech from the untrained mechanism, as well as whether such generalization depends on which source mechanism was originally used for training.

Methods

Stimuli—The stimuli used in Experiment 2 are the same as those from the perceptual experiment in Experiment 1.

Participants—A new group of young adult native speakers of American English (N = 24, age 18–38 years, M = 21.3, 15 female) provided informed, written consent to participate in this study, as approved by the Northwestern University Institutional Review Board. All participants reported having normal speech and hearing, being free from psychological or neurological disorders, and having no prior experience with electrolarynx speech.

Procedure—Participants were assigned to one of two experimental conditions (both N = 12), in which they learned to identify the talkers based on recordings of their speech produced using one of the source mechanisms and were then tested on their ability to generalize talker identity to speech produced by the other mechanism. The listening environment and stimulus presentation parameters were the same as Experiment 1. Like Experiment 1, the experimental session was divided into Training, Training Assessment, and Testing phases. Participants either (**1**) learned to identify talkers from recordings of their laryngeal voice (training and training assessment phases) and were then tested on their electrolarynx recordings (test phase), or (**2**) learned to identify talkers from recordings produced using the electrolarynx (training and training assessment phases) and were then tested on recordings of their laryngeal voice (test phase). Unlike Experiment 1, because listeners in both conditions of Experiment 2 would identify talkers from electrolarynx speech, both conditions were preceded by the additional familiarization phase, during which listeners gained exposure to electrolarynx speech. The details of each phase (training, training assessment, and testing) were identical between Experiments 1 and 2, with the

exception of the switch in source mechanisms between the Training Assessment and the Test phase.

Results

Participants' ability to learn talker identity from either an electrolarynx or laryngeal source was assessed by their accuracy on the Training Assessment, and their ability to generalize to the other source mechanism was assessed by their accuracy on the Test phases of each condition (Fig 3). Like Experiment 1, listeners successfully learned identity both from talkers' laryngeal voice (training assessment mean accuracy = 87.7%, s.d. = 13.5%) or electrolarynx recordings (M = 53.8%, s.d. = 15.5%), and participants learned to identify talkers from electrolarynx speech significantly better than chance [one-sample t-test vs. chance (20%); $t_{11} = 7.53$, $p < 2\times10^{-5}$, two-tailed; $d = 2.17$]. Learning talker identity from recordings of laryngeal speech again resulted in significantly more accurate identification than the electrolarynx recordings [independent-sample t-test; $t_{23} = 5.69$, $p < 0.00002$, two-tailed; $d = 2.43$].

Performance in the Test phase (ability to generalize to the novel source mechanism) was analyzed using a 2×2 repeated-measures analysis of variance with *training source mechanism* (Laryngeal Voice vs. Electrolarynx) as the between-participant factor and *content familiarity* (Trained vs. Novel) as the within-participant factor. Unlike Experiment 1, there was no effect of source mechanism [$F_{1,22} = 0.19$, $p = 0.66$], no effect of content familiarity [$F_{1,22} = 0.12$, $p = 0.73$], and no interaction [$F_{1,22} = 1.65$, $p = 0.21$].

In fact, neither training condition nor sentence familiarity had any effect in this experiment because, as shown in Fig. 3, in all permutations listeners were unable to generalize talker identity from the trained source mechanism to the untrained mechanism. Listeners who learned to identify talkers from their laryngeal voice were no better than chance at identifying the same talkers from their electrolarynx recordings, whether for familiar [$t_{11} = 0.77$, $p = 0.46$, two-tailed] or novel sentences [$t_{11} = 0.43$, $p = 0.67$, two-tailed]. Likewise, listeners who learned to identify talkers from their electrolarynx recordings were no better than chance at identifying the same talkers from their laryngeal voice, both for familiar [$t_{11} = -0.43$, $p = 0.67$, two-tailed] or novel sentences [$t_{11} = 0.73$, $p = 0.48$, two-tailed]. (Preceding t-tests are all one-sample vs. chance (20%).)

Discussion

The results of Experiment 2 unambiguously demonstrate that, after learning the identity of talkers from one source mechanism, listeners are, on average, unable to generalize that knowledge to the other source mechanism, irrespective of which mechanism was originally learned. It is perhaps unsurprising that listeners who learned talker identity from laryngeal speech – where differences in source information such as pitch and voice quality may have constituted the primary cues to talker identity – were unable to generalize this knowledge to the electrolarynx source mechanism. However, it is more surprising that listeners who learned talker identity from the electrolarynx source – where there were no differences in source information, and talker identity was signaled by differences in filter information

alone – were unable to use their knowledge of differences in filter features to identify the same talkers from laryngeal speech.

It would be incorrect to infer from the present results that structural information about the vocal filter does not contribute to talker identity. Previous work has shown that listeners are able to use filter-only information, as encoded in sinusoidal analogs of peak formant frequencies, to identify familiar talkers (Remez, Fellowes, & Rubin, 1997). Listeners can also be trained to recognize unfamiliar talkers from only their format-based vocal structure and speech dynamics (Sheffert et al., 2002). Moreover, listeners' subjective impressions of similar-sounding talkers are preserved across the presence or absence of vocal source information (Remez, Fellowes, & Nagel, 2007). Instead, this result may suggest that the acoustic-phonetic features to talker identity related to the vocal filter may not be learned independently and (at least initially) cannot be differentiated from the vocal source during talker identification, even when the properties of that source are homogeneous across different talkers. It may alternately be the case that, in order to accommodate the electrolarynx when producing fluent speech, talkers in this study made substantial changes to the static or dynamic features of their vocal tracts such that filter information between voicing mechanisms was, in fact, not reliably consistent for listeners. In fact, despite the fact that speakers were instructed to produce speech "as naturally as possible", it was observed that attempts to produce intelligible speech using the EL may have caused speakers to be more deliberate in their articulations. This observation motivated our investigation into potential acoustic-phonetic differences between laryngeal and electrolarynx speech – particularly those related to vocal tract resonance.

ACOUSTIC-PHONETIC DIFFERENCES BETWEEN LARYNGEAL AND ELECTROLARYNX SPEECH

Background

We sought to quantify differences in acoustic-phonetic properties of the vocal filter between talkers' use of the two voicing mechanisms in Experiment 2. If the filter properties varied widely and inconsistently between the two voicing mechanisms, this may suggest listeners' failure to generalize talker identity was due to insufficiently consistent information when changing source mechanism. Alternately, if the differences in filter characteristics between the two mechanisms were within the range of intra-talker variation, or the changes introduced by the electrolarynx were stereotyped and predictable, a failure to generalize is more likely related the inability of listeners to perceptually disentangle source and filter features of talker identity.

Methods

From the recordings of each source mechanism used in the perceptual experiments, 30 vowel tokens, selected for their clarity across talkers, were measured for position in F1 × F2 space ("vowel space"). The following tokens from each talker were used: /ɑ/: *pot*, *soft*, *rod*, *follow*, *across*; /i/: *heat*, *breeze*, *beat*, *feet*, *tea*, *sea*; /u/: *booth*, *two*, *fruit*, *used*, *huge*, *through*; /o/: *rose*, *source*, *follow*, *smoky*; /æ/: *back*, *pass*, *lack*, *sand*; /e/: *came*, *playing*, *stain*, *straight*, *page*. Vowels were measured through their longest steady-state portion

beginning at least two periods of phonation (or two electrolarynx pulses) after the preceding phoneme (to minimize coarticulation effects) and stopping at least two periods before either the following consonant, the onset of creaky voice, the end of phonation, or the transition to the second phone in a diphthong. If a viable vowel region could not be identified based on these criteria, measurement of that token was excluded for the talker. The mean values of F1 and F2 in the region described above were determined using the formant tracker implemented in Praat (Boersma & Weenink, 2009).

The precision of the acoustic measurements was assessed through intra-rater reliability on a re-measured subset (30%) of the tokens across both source mechanisms. Pearson's product-moment correlation coefficients revealed a high rate of intra-rater reliability on these re-measurements: 99% for vowels produced using the laryngeal source (median F1 discrepancy = 11 Hz; median F2 discrepancy = 65 Hz) and 91% for vowels produced using the electrolarynx (median F1 discrepancy = 35 Hz; median F2 discrepancy = 68 Hz). In addition, the accuracy of the acoustic measurements was assessed through inter-rater reliability. A second individual simultaneously measured the formants of a subset (80%) of the tokens across both source mechanisms. A correspondingly high rate of inter-rater reliability was observed between these two individuals: 98% for laryngeal vowels (median F1 discrepancy = 15 Hz; median F2 discrepancy = 63 Hz) and 96% for electrolarynx vowels (median F1 discrepancy = 30 Hz; median F2 discrepancy = 48 Hz).

Results

The vowel spaces measured from each source mechanism are illustrated in Fig. 4, and the mean formant frequency values for each vowel (across talkers and tokens) are listed in Table 1. These measurements for each formant (F1 and F2) were analyzed in separate repeated measures analyses of variance, with *source mechanism* (Laryngeal Voice vs. Electrolarynx) and *vowel* (/i/, /e/, /æ/, /a/, /o/, /u/) as within-subject factors. Vowels produced using the electrolarynx had a consistently higher F1 than those produced using talkers' laryngeal voice [main effect of source mechanism; $F_{1,4} = 64.14$, $p < 0.0014$, $\eta^2 = 0.676$]. The difference in F1 between source mechanisms did not vary differentially across vowel qualities [no source by vowel interaction; $F_{5,20} = 1.25$, $p = 0.323$]. Use of the electrolarynx was not associated with differences in F2 [no main effect of source; $F_{1,4} = 0.90$, $p = 0.397$]. There was a significant source by vowel interaction [$F_{5,20} = 3.97$, $p < 0.012$, $\eta^2 = 0.172$], such that the vowel /u/ had a significantly lower F2 in talkers' electrolarynx speech than with their laryngeal voice [paired *t*-test, $t_{29} = 3.76$, $p < 0.0008$, $d = 0.64$]. No other vowel's F2 differed significantly between the two source mechanisms. An example from a single talker's speech across the two source mechanisms is illustrated in Fig. 5.

Discussion

The change from talkers' own laryngeal source to the electrolarynx source was associated with at least two specific changes in the acoustic-phonetic properties of vowel quality. First, there was an overall raising of the frequencies in the first formant, which was evinced by all vowel categories. Higher F1 frequencies are predicted by two-tube models of the vocal tract as pharyngeal cavity length shortens (Johnson, 2003). Indeed, shortening of the pharyngeal cavity is effected by canonical placement of the electrolarynx against the soft tissue of the

neck beneath the mandible – several centimeters superior to the position of the closed glottis – creating a side branch in the vocal tract and introducing an antiresonance affecting the first formant (Meltzner, Kobler, & Hillman, 2003). This pattern is consistent with previous observation of low-frequency energy reduction in electrolarynx speech (Qi & Weinberg, 1991), including higher F1 frequencies. Other studies of alaryngeal speech, including esophageal (Sisty & Weinberg, 1972) and whispered (Jovičić, 1998) speech, have also reported raised formant frequencies; however, such increases were observed in both F1 and F2, whereas the present study found increases in only F1 for electrolarynx speech produced with an intact vocal tract.

Second, there was a significant backing (F2 reduction) of the vowel /u/ in electrolarynx speech, but no change in any other vowel. To understand this difference, it is important to note that, during laryngeal speech, our talkers exhibited significant fronting of the /u/ vowel compared to its more canonical position (Hillenbrand et al., 1995). Fronting of the vowel /u/ in connected speech is a common feature of the varieties of American English represented in our sample of talkers (Perrachione, Chiao, & Wong, 2010; Hall-Lew, 2011). It is likely that increased attentional control of speech articulation during electrolarynx use resulted in a comparatively "clear speech", canonical production of /u/, despite talkers' tendency to raise its F2 during conversational laryngeal speech.

Overall, the consistency of acoustic-phonetic filter features of talkers' speech between the two source mechanisms was mixed: F1 changed significantly, but predictably, across all items, whereas F2 was largely unaffected. The overall shape of the vowel space and relative position of vowels to one another did not change, and the idiosyncratic shape of each talker's vowel space likewise appeared consistent between source mechanisms. These results suggest that, although laryngeal and electrolarynx speech differ in some ways, many features remain reliably consistent between them and should have been available for cross-mechanism talker identification. That participants did not utilize these consistent features to generalize talker identification across source mechanisms raises the question of whether listeners are ever sensitive to consistency in certain acoustic-phonetic features independent of changes to others.

EXPERIMENT 3: MATCHING TALKER IDENTITY ACROSS SOURCE MECHANISMS

Background

To determine whether listeners are sensitive to any of the consistent features that reliably distinguish talkers between the two source mechanisms, we conducted a final, follow-up experiment designed to maximize the detection of similarities across source mechanisms. Instead of making a decision among five talkers based on memories of their identity, we instead asked participants to match utterances produced by the same talker across source mechanisms. On every trial of this experiment, all the information participants needed to make a decision about talker identity was available to them without having to learn and retain features to talker identity, or to compute how these features might predictably change following a change in vocal source.

Methods

Stimuli—The stimuli used in Experiment 3 consisted of eight sentences from both source modalities; these eight sentences were comprised of all five of the training sentences and three of the test sentences from Experiments 1 and 2 (see Appendix A).

Participants—A third group of young adult native speakers of American English (N = 12, age 18–22 years, M = 20.8, 10 female) provided informed written consent to participate in this study, as approved by the Northwestern University Institutional Review Board. All participants reported having normal speech and hearing, being free from psychological or neurological disorders, and having no prior experience with electrolarynx speech.

Procedure—Participants in Experiment 3 attempted to match talkers' electrolarynx recordings to recordings of their laryngeal voice and vice-versa. In two AXB discrimination conditions, participants heard a target recording from one source mechanism ("X") and attempted to match it to the recording produced by the same talker in the other source mechanism ("A" or "B").

In the LV-EL-LV condition, participants heard a sentence recorded by one talker using his laryngeal voice while "#1" appeared on the screen, then the same sentence recorded by either the first or second talker using the electrolarynx while "X" appeared on the screen, and finally the same sentence recorded by a second talker using his laryngeal voice while "#2" appeared on the screen. Participants were instructed to decide whether the recording labeled "X" was produced by the first or second talker and respond by button press. After responding, participants received feedback indicating whether they had chosen correctly.

In the EL-LV-EL condition, participants heard a sentence recorded by one talker using the electrolarynx while "#1" appeared on the screen, then the same sentence recorded by either the first or second talker using his laryngeal voice while "X" appeared on the screen, and finally the same sentence recorded by a second talker using the electrolarynx while "#2" appeared on the screen. Participants were instructed to decide whether the talker labeled "X" produced recording #1 or recording #2 and respond by button press. After responding, participants received feedback indicating whether they had chosen correctly.

All participants completed both experimental conditions separately, and the order was counterbalanced across participants. Each condition consisted of 80 trials, in which each talker was paired with every other talker an equal number of times, every sentence was heard an equal number of times, each talker was equally probable as the correct response, and the first or second talker were equally probable as the correct response. The environment and stimulus presentation were the same as Experiments 1 and 2.

Results

Participants' ability to correctly discern which of two talkers made an electrolarynx recording, and which of two electrolarynx recordings were made by a given talker, is illustrated in Fig. 6. Listeners were able to correctly match an electrolarynx recording to the corresponding talker's laryngeal voice 58.0% of the time, on average – significantly better than chance [LV-EL-LV condition; one-sample t-test vs. chance (50%); $t_{11} = 3.99$, $p <$

0.0025, $d = 1.15$, two-tailed]. Similarly, listeners were able to correctly match talkers' laryngeal voice to their electrolarynx recordings 58.6% of the time – again significantly better than chance [EL-LV-EL condition; one-sample *t*-test vs. chance (50%); $t_{11} = 3.12$, $p < 0.01$, $d = 0.90$, two-tailed]. Listeners' accuracy did not differ between the two AXB discrimination conditions [paired-sample *t*-test; $t_{11} = 0.31$, $p = 0.76$]. When listeners' accuracy on this task was calculated separately for each quarter of the task (*i.e.*, the first 20 trials, the second 20, etc.), there was no evidence of performance improvement (learning) across either session [no effect of quarter in a repeated-measures analysis of variance of participants' accuracy by *quarter* (1, 2, 3, 4) and *task* (EL-LV-EL, LV-EL-LV); $F_{1,11} = 0.86$, $p = 0.37$], despite receiving corrective feedback on every trial.

Discussion

In Experiment 3, listeners demonstrated that they are able to successfully determine which of two talkers produced an utterance using the other source mechanism. This result suggests that there are cues to talker identity that are preserved between source mechanisms (*e.g.*, potentially, vocal tract resonance, speech rate, patterns of pronunciation, duration-based lexical stress, nasality, etc.), and that listeners are sensitive to these cues for matching utterances. It is worth noting that, although reliable across participants and significantly better than chance, participants' ability to match talkers' speech across source mechanism was modest – averaging only about 58% correct in both conditions – and lower than listeners' ability to discriminate talker pairs in a related study using only electrolarynx speech (Coleman, 1973). This result may suggest that, although filter-based features of talker identity are preserved across source mechanisms and occasionally perceptually salient to listeners, access to such similarity may be obfuscated in the presence of conflicting information about the vocal source.

The qualitative disconnect between the results of the present experiment, in which listeners were successfully able to discern a consistent talker across source mechanisms, from Experiment 2, in which they were not, also suggests that caution is necessary when considering the extent to which experiments involving the *discrimination* of voices are able to accurately ascertain the acoustic features used by listeners during the *identification* of voices (cf., Baumann & Belin, 2008). That is, although some acoustic features may facilitate behavior in one experimental design, the same features may not be used (or used in the same way) under differing task demands, or even by different listeners (Lavner, Rosenhouse, & Gath, 2001).

GENERAL DISCUSSION

Across three perceptual experiments, we observed that listeners are able to accurately learn talker identity from speech produced using an electrolarynx. These data demonstrate that the speech of electrolarynx users is able to manifest an individual vocal identity that may be sufficient to distinguish them from other electrolarynx users, even under rigid laboratory constraints where the operational characteristics of a single device were fixed across talkers.

Listeners have the ability to take advantage of individual differences in pronunciation and other dynamic cues of electrolarynx speech for the purposes of talker identification,

demonstrating not only the range of phonetic information available in electrolarynx speech, but also a perceptual sensitivity to it. Listeners' ability to identify talkers from electrolarynx speech further resembled their ability to identify talkers from laryngeal speech in its robustness to novel sentential content. This reifies the idea that, when listening to speech, listeners not only encode the meaningful content of an utterance but also learn the consistent acoustic-phonetic nuances of individual talkers (Palmeri, Goldinger, & Pisoni, 1993; Theodore & Miller, 2010). Such phonetic consistency not only supports talker identification, but also facilitates the recognition of novel speech by familiar talkers – even when the source characteristics of that speech are unfamiliar. It is worth noting that the successful application of knowledge about the idiosyncratic nature of a talker's phonetics may depend on listeners' expectations as to whether they are hearing that talker (Johnson, 1997; Magnuson & Nusbaum, 2007) – a distinction that may be related to listeners' failure to generalize talker identity across source mechanisms in Experiment 2 where expectations about vocal identity were overwhelmed by the novel source mechanism.

However, listeners were still more accurate at learning to identify talkers from laryngeal speech than electrolarynx speech, presumably because of the increased information afforded by additional distinctive features such as pitch and voice quality. A lifetime of reliance on both source and filter cues results in a significant decrement of listeners' talker identification accuracy when only filter-related cues are available to distinguish talkers. It is worth pointing out that, in the present study, the magnitude of this difference was assessed after only a very brief laboratory training exercise. It is reasonable to suppose that individuals with more exposure to electrolarynx speech – such as clinicians and friends and family of electrolarynx users – would exhibit even better talker identification abilities than those of naïve participants (*cf.* Knox & Anneberg, 1973).

A psychologically interesting aspect of these results is that listeners did not demonstrate an ability to generalize talker identity learned from one source mechanism to speech produced using the other mechanism. Acoustic analysis of speech from the two mechanisms revealed phonetic features that were reliably consistent across source mechanisms, as well as features that differed significantly but predictably. This result suggests that structural information about talker identity is not an immutable property of anatomy available to listeners independent of speech content (Latinus & Belin, 2011b), but rather comprises a system that listeners employ dynamically in the service of producing speech (Perrachione, Chiao, & Wong, 2010). Listeners did, however, demonstrate an ability to match which of two speech samples produced by different source mechanisms came from the same talker. This result affirms the idea that some information about talker identity is retained across changes in source mechanism, even if listeners appear unable to use that information in the task of explicitly identifying a talker. Together, these results endorse the idea advanced by Kreiman and Sidtis (2011) that, rather than making talker identity judgments based on a collection of independent, transferable features, listeners form a holistic perceptual representation of talker identity – a sort of "vocal gestalt" – that synthesizes source, filter, and linguistic information. For instance, despite the relative homogeneity of the source mechanism in electrolarynx speech (i.e., no differences voice quality, pitch, dynamic amplitude modulation, etc.), sufficiently distinctive indexical cues to talker identity remain in the

acoustic signature of the vocal filter (e.g., oral and pharyngeal volume, formant dispersion) and its dynamics during speech articulation (e.g., distinctive vowel space, speech rate, accent, patterns of duration-based syllabic or lexical stress, etc., (Hewlett, Cohen & MacIntyre, 1997)) to facilitate reliable percepts of talker identity from electrolarynx speech.

It is worth pointing out that the observed successes and limitations at identifying talkers from electrolarynx speech, and generalizing talker identity across changes in source mechanism, were the product of only a short, laboratory training paradigm. It remains possible – and even likely, based on anecdotal reports from clinicians working extensively with electrolarynx users – that increased familiarity with the speech of electrolarynx users may facilitate talker identification in this medium. For instance, previous work has indicated that additional laboratory training can improve talker identification from unfamiliar speech environments, such as accented or foreign language speech (Perrachione & Wong, 2007; Winters, Levi, & Pisoni, 2008). Additionally, there is evidence from studies of both brain and behavior that the representation and processing of familiar voices, such as those of friends, family, and close associates, differ from the representations of new or unfamiliar voices, such as those used in laboratory training exercises like the present study (Van Lancker & Kreiman, 1987; Nakamura et al., 2001; Beauchemin et al., 2006; Johnsrude et al., 2013).

These findings also suggest that everyday electrolarynx users may be able to maintain a unique, individual vocal identity, possession of which is a key psychosocial feature (Sidtis & Kreiman, 2011). However, given the widespread use of electrolarynx devices for speech restoration (Hillman et al., 1998; Koike et al., 2002), additional clinical and technical research into improving the communicative efficacy of electrolarynx speech remains needed. Advances that facilitate the quality of electrolarynx speech, including greater capacity for prosodic flexibility, a spectrum more closely resembling laryngeal speech, and reduction of noncommunicative device noise, are also likely to increase its capacity to convey distinctive indexical information. Correspondingly, it also remains the purview of future work to determine whether the amount of acoustic-phonetic variation among clinical electrolarynx users is more or less conducive to talker identification by both naive and expert listeners compared to what was observed in the present experiments.

CONCLUSIONS

Although electrolarynx speech may lack individual differences in voice quality related to properties of the vocal source, it nevertheless conveys sufficient indexical information about individual variability in the vocal filter and dynamic speech articulations for listeners to learn to distinguish and identify individual electrolarynx users based on short recordings of their speech. Talker identification, therefore, can be successful even in the absence of differences in the vocal source. However, listeners were not able to generalize their knowledge of talker identity following a change in the source mechanism, regardless of whether talker identity was originally learned from laryngeal or electrolarynx speech. Similarly, although individuals were more accurate than chance at matching talkers based on their laryngeal and electrolarynx speech, their ability to match talkers across source mechanisms was considerably less than previous studies of within-source-mechanism

discrimination. Taken together, these results suggest that, when listeners learn to identify talkers from speech, they form gestalt perceptions of talker identity that do not dissociate information separately attributable to the acoustics of the vocal source or vocal filter.

ACKNOWLEDGMENTS

We thank James Heaton, Satrajit Ghosh, Yoomin Ahn, Louisa Ha, Allison Barr, Nicole Lomotan, Rachel Gould, and James Sugarman for their contributions. This work was supported by the Massachusetts General Hospital (R.H.), grants 5T32DC000038-17 and 5T32HD007424 from the National Institutes of Health (C.S.), and grants from the National Science Foundation (BCS-0719666 & BCS-1125144) and National Institutes of Health (R01DC008333 & K02AG035382) awarded to PW. TP was supported by a NSF Graduate Research Fellowship.

REFERENCES

Baumann O, Belin P. Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. Psychological Research. 2008; 74:110–120. [PubMed: 19034504]

Beauchemin M, De Beaumont L, Vannasing P, Turcotte A, Arcand C, Belin P, Lassonde M. Electrophysiological markers of voice familiarity. European Journal of Neuroscience. 2006; 23:3081–3086. [PubMed: 16819998]

Boersma P, Weenink D. Praat: Doing phonetics by computer. 2009 Available online: http://www.fon.hum.uva.nl/praat/.

Brown WS Jr, Feinstein SH. Speaker sex identification utilizing a constant laryngeal source. Folia Phoniatrica. 1997; 29:240–248. [PubMed: 924312]

Carrell, TD. Doctoral dissertation. Bloomington, Indiana: Indiana University; 1984. Contributions of fundamental frequency, formant spacing, and glottal waveform to talker identification.

Clements KS, Rassekh CH, Seikaly H, Hokanson JA, Calhoun KH. Communication after laryngectomy: An assessment of patient satisfaction. Archives of Otolaryngology – Head and Neck Surgery. 1997; 123:493–496. [PubMed: 9158395]

Coleman RO. Speaker identification in the absence of inter-subject differences in glottal source characteristics. Journal of the Acoustical Society of America. 1973; 53:1741–1743. [PubMed: 4719259]

Eadie TL, Day AMB, Sawin DE, Lamvik K, Doyle PC. Auditory-perceptual speech outcomes and quality of life after total laryngectomy. Otolaryngology – Head and Neck Surgery. 2013; 148:82–88. [PubMed: 23008330]

Finzia C, Bergman B. Health-related quality of life in patients with laryngeal cancer: A post-treatment comparison of different modes of communication. The Laryngoscope. 2001; 111:918–923. [PubMed: 11359178]

Hall-Lew, L. The completion of a sound change in California English; Proceedings of the 17th International Congress of Phonetic Sciences; 2011. p. 807-810.

Hewlett N, Cohen W, MacIntyre C. Perception and production of voiceless plosives in electronic larynx speech. Clinical Linguistics & Phonetics. 1997; 11:1–22.

Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America. 1995; 97:3099–3111. [PubMed: 7759650]

Hillman RE, Walsh MJ, Wolf GT, Fisher SG, Hong WK. Functional outcomes following treatment for advanced laryngeal cancer. Part I--Voice preservation in advanced laryngeal cancer. Part II--Laryngectomy rehabilitation: The state of the art in the VA System. Annals of Otology, Rhinology, & Laryngology. 1998; 107:1–27.

Institute of Electrical and Electronics Engineers. IEEE recommended practices for speech quality measurements. IEEE Transactions on Audio and Electroacoustics. 1969; 17:225–246.

Johnson, K. Acoustic and Auditory Phonetics. 2nd ed.. Malden, MA: Blackwell; 2003.

Johnson K. The role of perceived speaker identity to F0 normalization of vowels. Journal of the Acoustical Society of America. 1997; 88:642–654. [PubMed: 2212287]

Johnsrude IS, Mackey A, Hakyemez H, Alexander E, Trang HP, Carlyon RP. Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice. Psychological Science. 2013; 24:1995–2004. [PubMed: 23985575]

Jovičić ST. Formant feature differences between whispered and voiced sustained vowels. Acustica. 1998; 84:739–743.

Knox AW, Anneberg M. The effects of training in comprehension of electrolaryngeal speech. Journal of Communication Disorders. 1973; 6:110–120. [PubMed: 4776546]

Koike M, Kobayashi N, Hirose H, Hara Y. Speech rehabilitation after total laryngectomy. Acta Otolaryngolgy. 2002; (Suppl. 547):107–112.

Kreiman, J.; Sidtis, D. Foundations of Voice Studies. Malden, MA: Wiley-Blackwell; 2011.

Lavner Y, Gath I, Rosenhouse J. The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. Speech Communication. 2000; 30:9–26.

Latinus M, Belin P. Anti-voice adaptation suggests prototype-based coding of voice identity. Frontiers in Psychology. 2011a; 2:175. 1–12. [PubMed: 21847384]

Latinus M, Belin P. Human voice perception. Current Biology. 2011b; 21:R143–R145. [PubMed: 21334289]

Lee MT, Gibson S, Hilari K. Gender differences in health-related quality of life following total laryngectomy. International Journal of Language & Communication Disorders. 2010; 45:287–294. [PubMed: 20131961]

Legge GE, Grosmann C, Pieper CM. Learning unfamiliar voices. Journal of Experimental Psychology – Learning, Memory, and Cognition. 1984; 10:298–303.

Liu H, Ng ML. Electrolarynx in voice rehabilitation. Auris Nasus Larynx. 2007; 34:327–332. [PubMed: 17239553]

Magnuson JS, Nusbaum HC. Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. Journal of Experimental Psychology – Human Perception and Performance. 2007; 33:391–409. [PubMed: 17469975]

Meltzner GS, Hillman RE. Impact of aberrant acoustic properties on the perception of sound quality in electrolarynx speech. Journal of Speech, Language, and Hearing Research. 2005; 48:766–779.

Meltzner GS, Kobler JB, Hillman RE. Measuring the neck frequency response function of laryngectomy patients: Implications for the design of electrolarynx devices. Journal of the Acoustical Society of America. 2003; 114:1035–1047. [PubMed: 12942982]

Moukarbel RV, Doyle PC, Yoo JH, Franklin JH, Day AMB, Fung K. Voice-related quality of life (V-RQOL) outcomes in laryngectomees. Head and Neck. 2011; 33:31–36. [PubMed: 20848430]

Nakamura K, Kawashima R, Sugiura M, Kato T, Nakamura A, Hatano K, Nagumo S, Kubota K, Hrioshi F, Ito K, Kojima S. Neural substrates for recognition of familiar voices: a PET study. Neuropsychologia. 2001; 39:1047–1054. [PubMed: 11440757]

Palmeri TJ, Goldinger SD, Pisoni DB. Episodic encoding of voice attributes and recognition memory for spoken words. Journal of Experimental Psychology – Learning, Memory and Cognition. 1993; 19:309–328.

Perrachione TK, Wong PCM. Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. Neuropsychologia. 2007; 45:1899–1910. [PubMed: 17258240]

Perrachione TK, Chiao JY, Wong PCM. Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. Cognition. 2010; 114:42–55. [PubMed: 19782970]

Perrachione TK, Del Tufo SN, Gabrieli JDE. Human voice recognition depends on language ability. Science. 2011; 333:595. [PubMed: 21798942]

Qi Y, Weinberg B. Low-frequency energy deficit in electrolaryngeal speech. Journal of Speech and Hearing Research. 1991; 34:1250–1256. [PubMed: 1787706]

Remez RE, Fellowes JM, Rubin PE. Talker identification based on phonetic information. Journal of Experimental Psychology: Human Perception and Performance. 1997; 23:651–666. [PubMed: 9180039]

Remez RE, Fellowes JM, Nagel DS. On the perception of similarity among talkers. Journal of the Acoustical Society of America. 2007; 122:3688–3696. [PubMed: 18247776]

Robertson SM, Yeo JCL, Dunnet C, Young D, MacKenzie K. Voice, swallowing, and quality of life after total laryngectomy – esults of the west of Scotland laryngectomy audit. Head & Neck. 2012; 34:59–65. [PubMed: 21416548]

Sheffert SM, Pisoni DB, Fellowes JM, Remez RE. Learning to recognize talkers from natural, sinewave, and reversed speech samples. Journal of Experimental Psychology – Human Perception and Performance. 2002; 28:1447–1469. [PubMed: 12542137]

Sidtis D, Kreiman J. In the beginning was the familiar voice: Personally familiar voices in the evolutionary and contemporary biology of communication. Integrative Psychological and Behavioral Science. 2011; 46:146–159. [PubMed: 21710374]

Sisty NL, Weinberg B. Formant frequency characteristics of esophageal speech. Journal of Speech and Hearing Research. 1972; 15:439–448. [PubMed: 5047882]

Theodore RM, Miller JL. Characteristics of listener sensitivity to talker specific phonetic detail. Journal of the Acoustical Society of America. 2010; 128:2090–2099. [PubMed: 20968380]

Van Lancker D, Kreiman J. Voice discrimination and recognition are separate abilities. Neuropsychologia. 1987; 25:829–834. [PubMed: 3431677]

Varghese BT, Mathew A, Sebastian S, Iype EM, Sebastian P, Rajan B. Cbjective and perceptual analysis of outcome of voice rehabilitation after laryngectomy in an Indian tertiary referral cancer centre. Indian Journal of Otolaryngology and Head & Neck Surgery. 2013; 65:S150–S154.

Voiers WD. Perceptual bases of speaker identity. Journal of the Acoustical Society of America. 1964; 36:1065–1073.

Walton JH, Orlikoff RF. Speaker race identification from acoustic cues in the vocal signal. Journal of Speech, Language, and Hearing Research. 1994; 37:738–745.

Weiss MS, Basili AG. Electrolaryngeal speech produced by laryngectomized subjects: Perceptual characteristics. Journal of Speech and Hearing Research. 1985; 28:294–300. [PubMed: 4010259]

Winters SJ, Levi SV, Pisoni DB. Identification and discrimination of bilingual talkers across languages. Journal of the Acoustical Society of America. 2008; 123:4524–4538. [PubMed: 18537401]

APPENDIX A

The content of the sentence stimuli for these experiments was taken from the “Harvard Sentences” (IEEE, 1969). Listeners assessed talker intelligibility from sentences #1–14. Experiments 1 & 2 utilized sentences #1–5 for the familiarization phase, and sentences #1–10 for the test phase. Experiment 3 utilized sentences #1–8 for all parts. The syllables from which acoustic measurements of vowel formants were made are underlined.

List 2 of the “Harvard Sentences” (IEEE, 1969)

1 The boy was there when the sun rose.

2 A rod is used to catch pink salmon.

3 The source of the huge river is the clear spring.

4 Kick the ball straight and follow through.

5 Help the woman get back to her feet.

6 A pot of tea helps to pass the evening.

7 Smoky fires lack flame and heat.

8 The soft cushion broke the man's fall.

9 The salt breeze came across from the sea.

10 The girl at the booth sold fifty bonds.

List 8 of the “Harvard Sentences” (IEEE, 1969).

11 A yacht slid around the point into the bay.

12 The two met while playing on the sand.

13 The ink stain dried on the finished page.

14 The walled town was seized without a fight.

15 The lease ran out in sixteen weeks.

16 A tame squirrel makes a nice pet.

17 The horn of the car woke the sleeping cop.

18 The heart beat strongly and with firm strokes.

19 The pearl was worn in a thin silver ring.

20 The fruit peel was cut in thick slices.

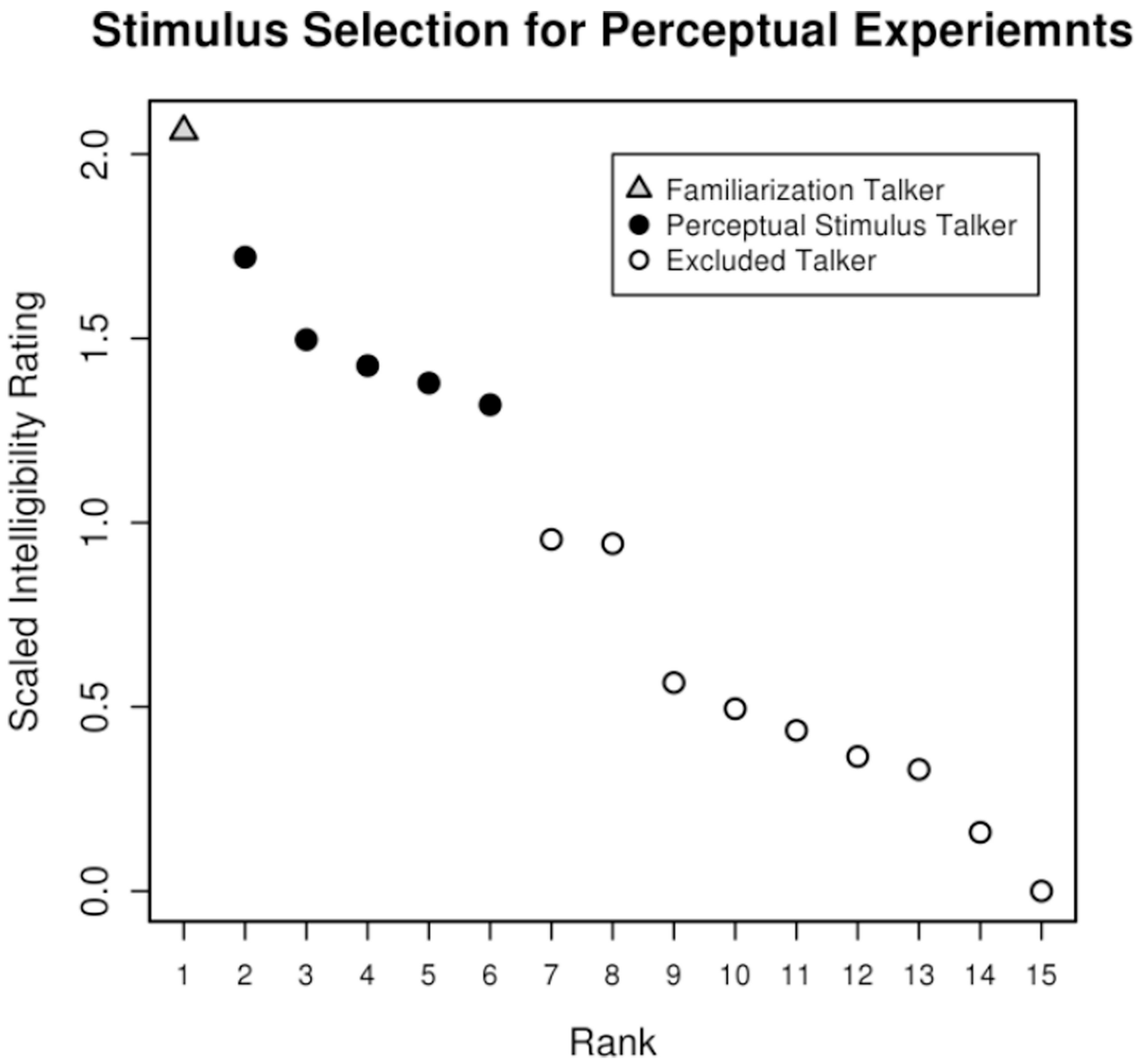

Fig. 1. Selection of talkers for perceptual experiments

Talkers whose electrolarynx recordings were rated as most intelligible by naïve listeners were used in the perceptual experiments. The talker with the highest intelligibility rating was reserved for familiarizing listeners with electrolarynx speech.

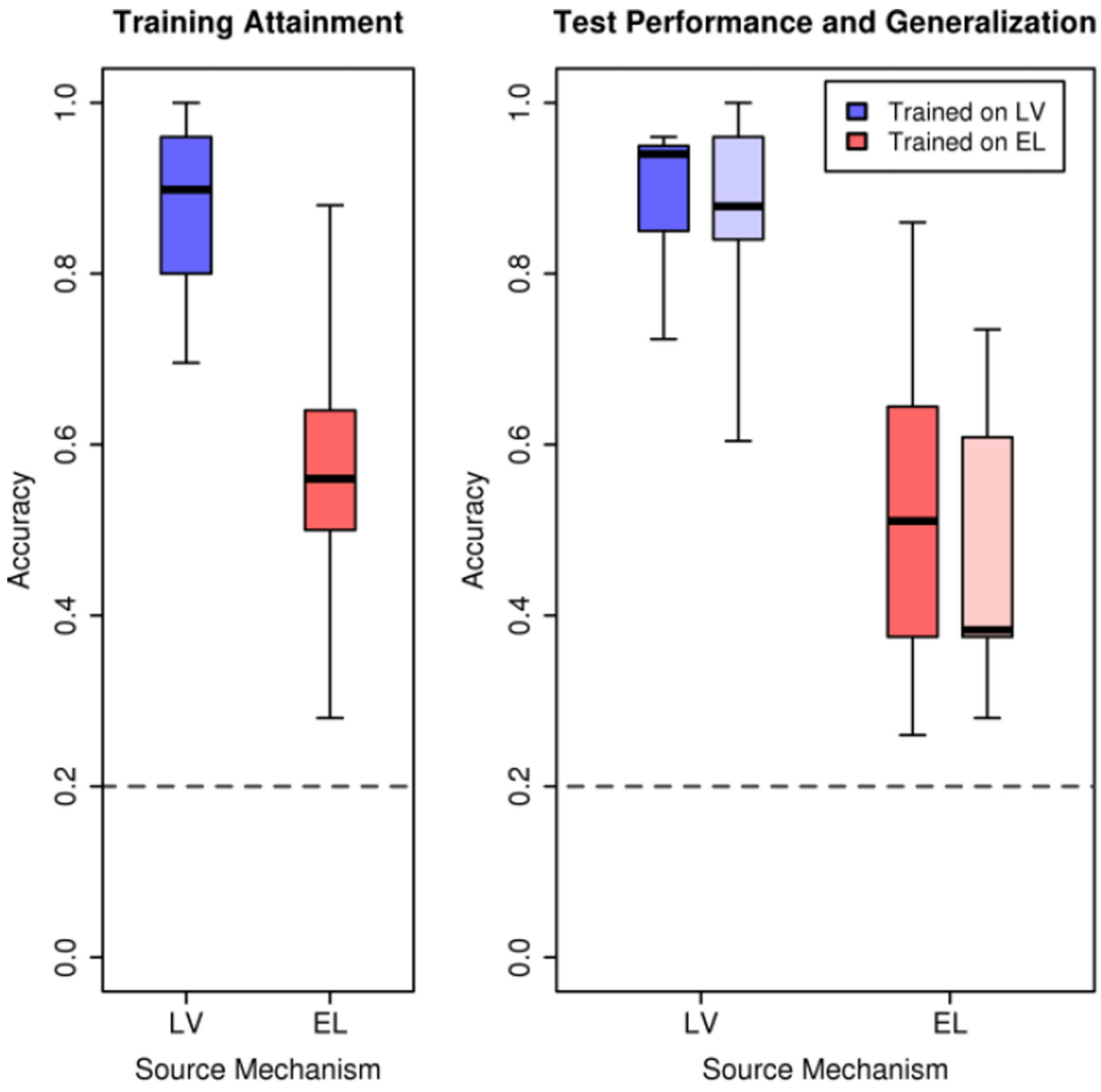

Fig 2. Learning talker identity from laryngeal and electrolarynx sources

Listeners learned talker identity successfully from both laryngeal (LV) and electrolarynx (EL) vocal sources, but were significantly more accurate at learning talker identity from a laryngeal source. For both vocal sources, listeners are significantly more accurate at identifying talkers from familiar sentences (darker boxes) than from novel ones (lighter boxes). Horizontal dashed line indicates chance (20%). *Boxplots:* Solid horizontal bar represents the median; colored area: interquartile range; dashed whiskers: maximum and minimum values.

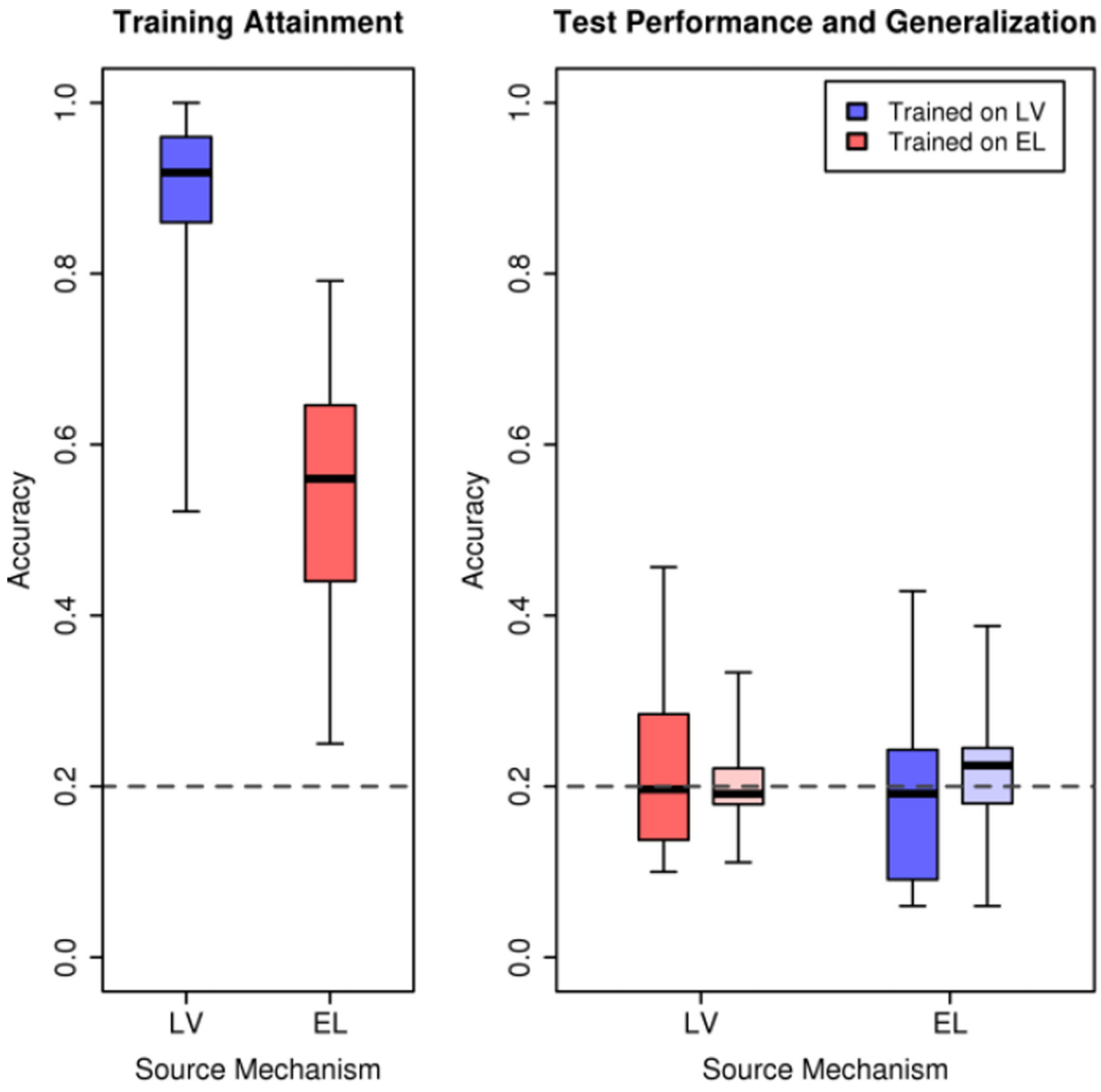

Fig 3. Generalizing talker identity across source mechanisms

Listeners learned talker identity successfully from both laryngeal (LV) and electrolarynx (EL) vocal sources, but were unable to generalize talker identity to the untrained mechanism. After learning talker identity from one source mechanism, listeners performed no better than chance at identifying the same talkers using the other mechanism, regardless of whether sentence content was familiar (darker boxes) or novel (lighter boxes). Figure conventions as in Figure 2.

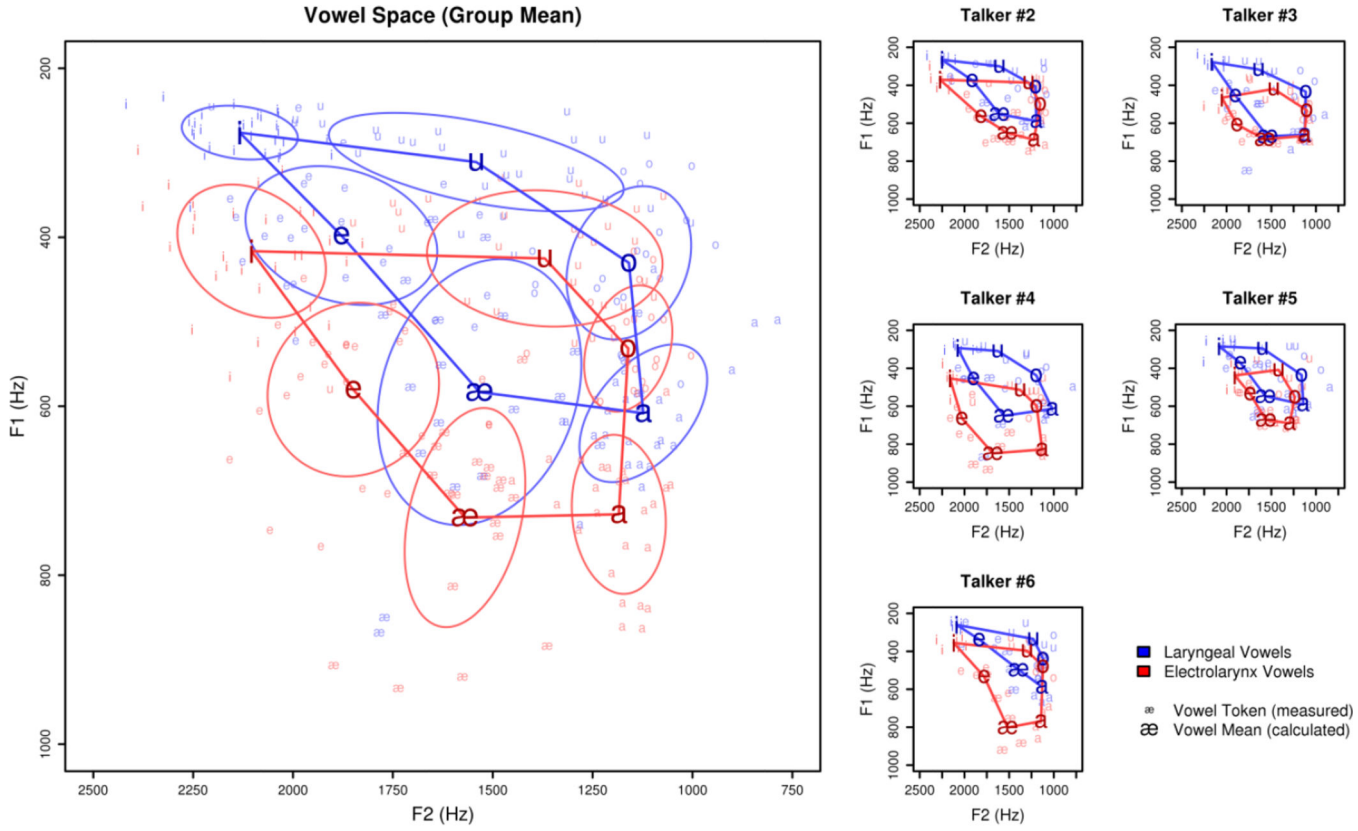

Fig 4. Phonetic differences across source mechanisms

Participants' use of the electrolarynx was associated with a significantly higher F1 across all vowels. The direction of the ordinate and abscissa depict articulatory orientation. Large symbols designate mean locus of each vowel, and solid lines demarcate mean vowel space, for each source mechanism. Small symbols designate values of individual tokens. Ellipses indicate the area within one standard deviation of the vowels' mean. The large panel at left illustrates the mean vowel spaces for the two source mechanisms across all five talkers; the smaller panels at right depict the vowel spaces for each talker individually. Note the consistent displacement of F1 from Electrolarynx speech across talkers, as well as the within-talker similarities in vowel-space shape between source mechanisms.

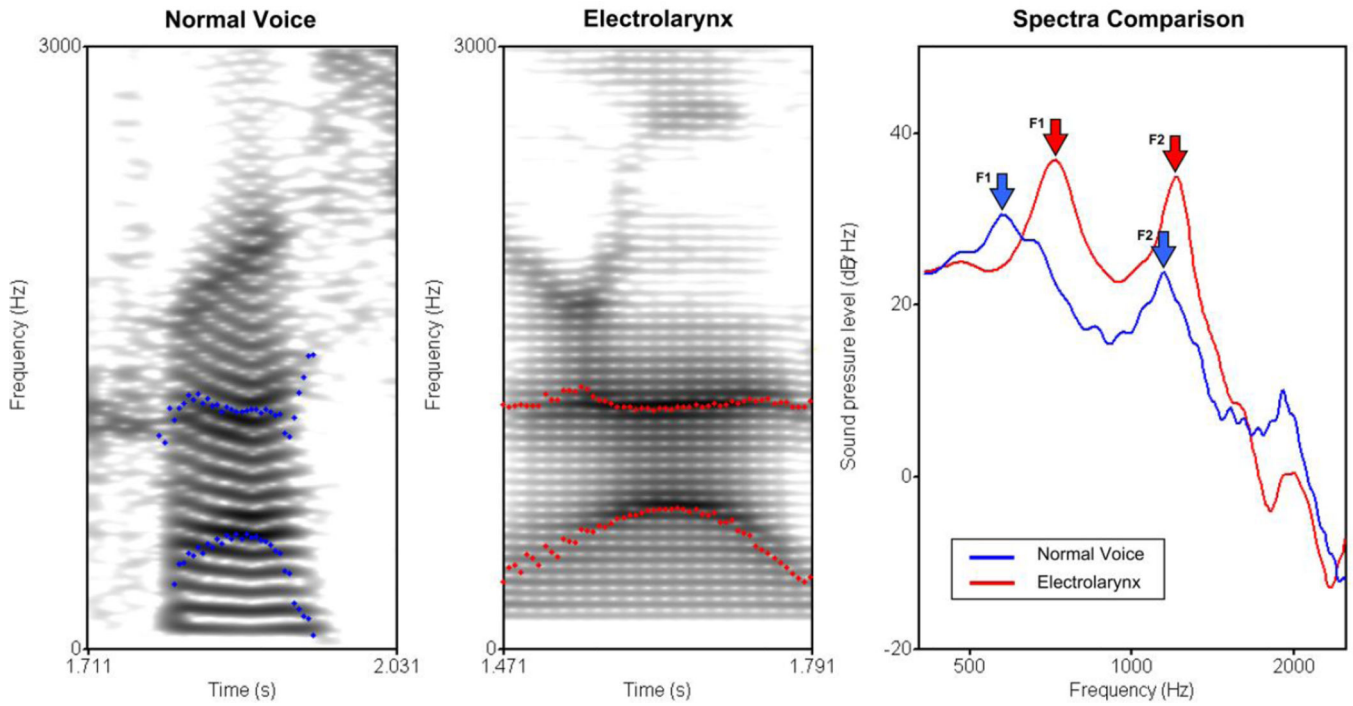

Fig 5. Comparison of within-subject phonetic differences

A single talker's production of the syllable [kɹɑːs] in the word *across* using either his laryngeal voice (left) or the electrolarynx (middle) are shown. Point contours trace the first and second formants. Average spectra, extracted from the middle of the vowel are shown (right), demonstrating the higher F1 associated with electrolarynx speech. The spectrograms also illustrate other differences between laryngeal and electrolarynx speech, including changes to F0, pitch dynamics, and aspiration.

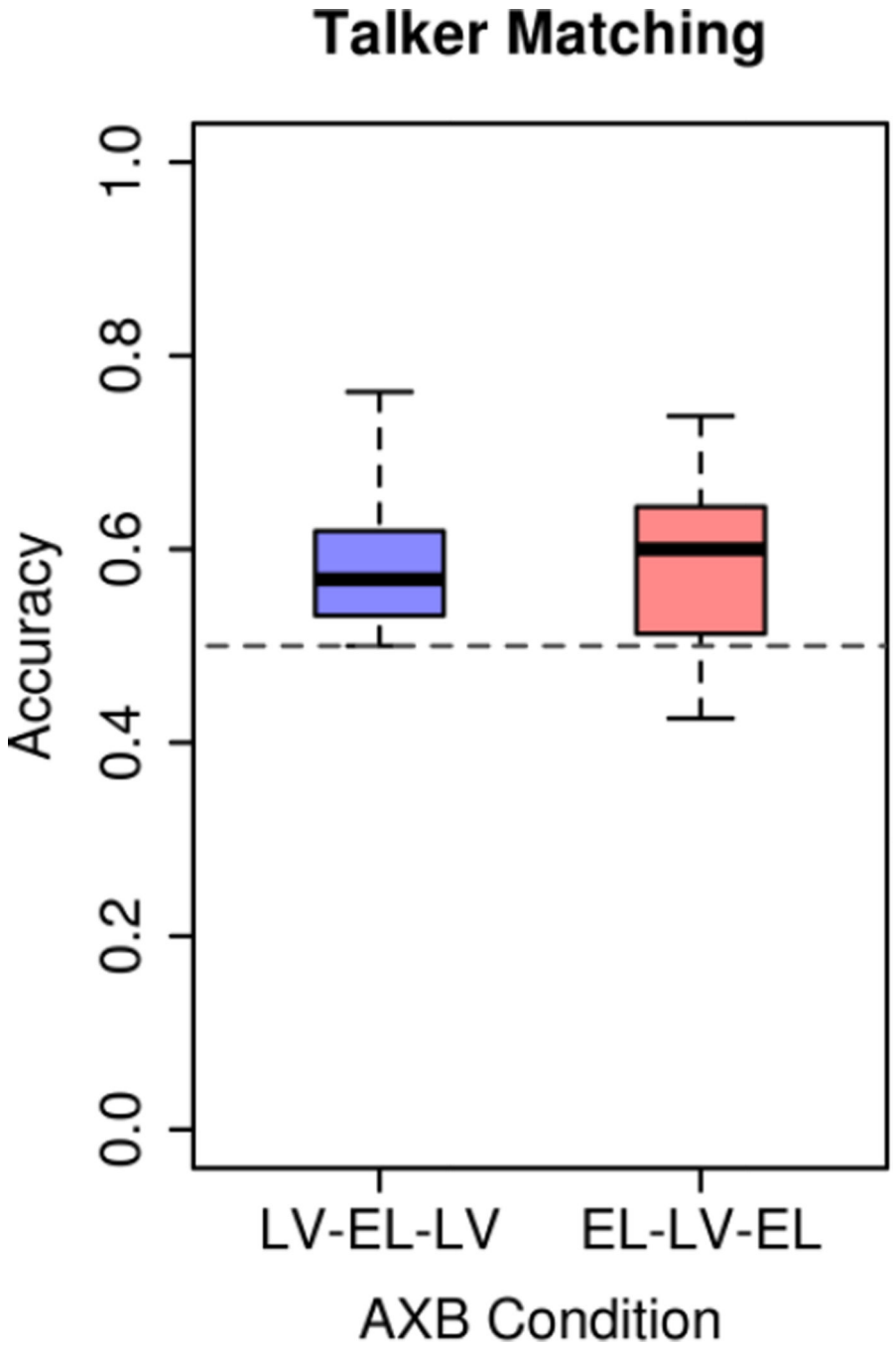

Fig 6. Matching talkers across source mechanisms

Listeners were significantly more accurate than chance at matching talker identity across source mechanisms in both conditions. Performance between the two conditions did not differ. Horizontal dashed line indicates chance (50%). Other figure conventions as in Figure 2.

Table 1

Mean (± s.d.) formant frequencies (Hz) for vowels measured in continuous speech from laryngeal and electrolarynx sources.

		Vowels										
Source		**/ɑ/**		**/æ/**		**/e/**		**/i/**		**/u/**		**/o/**	
Laryngeal	F1	608	(±67)	583	(±129)	398	(±67)	276	(±26)	425	(±48)	532	(±74)
	F2	1123	(±131)	1533	(±209)	1879	(±196)	2133	(±119)	1541	(±302)	1159	(±127)
Electrolarynx	F1	728	(±77)	732	(±106)	580	(±84)	416	(±64)	425	(±66)	532	(±61)
	F2	1185	(±96)	1570	(±123)	1849	(±175)	2104	(±152)	1370	(±241)	1161	(±90)

J Speech Lang Hear Res. Author manuscript; available in PMC 2015 November 23.