Annual Review of Applied Linguistics, **35** (2015), pp. 173–194.
© Cambridge University Press, 2015
doi: 10.1017/S0267190514000270

Voice Quality and Identity

Robert J. Podesva and Patrick Callier
Stanford University
podesva@stanford.edu; pcallier@stanford.edu

ABSTRACT

Variation in voice quality has long been recognized to have functions beyond the grammatically distinctive or phonetically useful roles it plays in many languages, indexing information about the speaker, participating in the construction of stance in interaction, or serving to identify the speaker as a unique individual. Though the links between voice quality and identity have been studied in phonetics, sociolinguistics, linguistic anthropology, forensic linguistics, and speech technology, considerable work remains to be done to problematize the ways in which the voice is taken as covering privileged, immediate meanings about the speaker's body and to break apart the ideologies that construct it as an inalienable, unitary, and invariant facet of a speaker's identity. We point out promising directions in recent research on the voice and bring up ideas for where this important area of research should be taken.

INTRODUCTION

Voice quality can be defined as "the extragrammatical suprasegmental properties of speech resulting from the configuration of the vocal apparatus" (Podesva, 2007, p. 478). Some scholars have described voice quality as a "quasi-permanent" characteristic of individual speakers (Abercrombie, 1967, p. 91), due partly to the physiology of the vocal apparatus and partly to the sedimentation of vocal behavior over the lifecourse (Laver, 1968, p. 44), the latter of which is strongly influenced by sociocultural convention. Others have emphasized that voice quality can change from moment to moment. Modifications of voice are mandated by linguistic structure in languages that contrast sounds in terms of breathiness, creakiness, or nasality (DiCanio, 2009; Garellek & Keating, 2011; Gordon, 2001; Gordon & Ladefoged, 2001); in other cases, they emerge as statistical tendencies dependent on phrase position, such as the prevalence of creaky voice in phrase-final position in English (Henton & Bladon, 1988) and Finnish (Ogden, 2001). But it is necessary to also keep sight of the fact that in all languages, regardless of linguistic structure, voice quality functions as a rich phonetic resource through which speakers can display affect and take stances in interaction, a point emphasized in recent research in the fields of linguistic anthropology and sociolinguistics (Podesva, 2007; Sicoli, 2010; Zimman, 2012).

Much of the early literature on voice quality, particularly from the perspective of phonetics, aimed to provide accurate terminology and reproducible methods for describing distinctive qualities (Catford, 1964; Esling & Harris, 2004; Ladefoged,

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

1988; Laver, 1980), and for quantifying them acoustically (Kuwabara & Ohgushi, 1984). As Laver (1968, p. 43) noted, many of the descriptors used to characterize voice qualities—labels like "husky," "plummy," "thin," "rich," "velvety," and "reedy"—were too vague for widespread use in phonetic description. Yet these descriptors illustrate that voices, contrary to other elements of phonetic material, are described in many of the same ways in which people (or their possessions) are described. Moreover, listeners infer a great deal about speakers from their voices alone, such as physical chracteristics (e.g., height, health status), behavioral traits (e.g., whether they smoke), and psychological states (e.g., mood, stress level; Kreiman, Vanlancker-Sidtis, & Gerratt, 2005). The additional observation that voice quality patterns vary widely across individuals and communities suggests that voice quality can serve as an important resource in the construction of identity. Indeed, one of Laver's (1968) goals was "to facilitate interdisciplinary discussion of the indexical function of voice quality" (p. 43). This article represents another step toward that end, one which benefits from zeroing in on kinds of identities that can be signaled through variation in the voice and how such identities are indexed.

We review in this article a number of ways that voice quality can serve as an index of identity. We begin first with a discussion of identity categories, particularly those corresponding to gender, race, class, and linguistic identities, as well as the kinds of ideologies about these categories that circulate in media representations. The importance of categories notwithstanding, we follow Bucholtz and Hall (2005) in treating categories as only one of several subject positions relevant to the study of identity. As they asserted, "identities encompass (a) macro-level demographic categories; (b) local, ethnographically specific cultural positions; and (c) temporary and interactionally specific stances and participant roles" (Bucholtz & Hall, 2005, p. 592). Given the capacity for voice quality modifications to enable speakers to enact shifts in identity as interaction unfolds, we turn in the next section to our second way of looking at voice quality: as a means of stancetaking. Our focus is on affective stance, as fluctuations in voice quality can be viewed iconically as changes in emotional and physical states. We nevertheless emphasize the ideological nature of such readings, pointing out several cases of stancetaking in the context of voice quality modifications where iconic readings are not possible. A third way in which voice quality participates in identity construction is in its potential to distinguish individual speakers. This topic has garnered attention in a number of disparate fields, including forensics, speech perception, and speech technology. Although voice quality has attracted attention from scholars in several other disciplines—such as music (Dong, Sundberg, & Kong, 2014; Heard, 2012; Johnson & Kempster, 2011) and discourse organization (Gorisch, Wells, & Brown, 2012; R. Ogden, 2004; Sikveland, 2012; Zellers & Post, 2012)—we do not cover this work here, due to the more peripheral relevance of identity in its study; it is nevertheless clear that matters of identity surface in each of these areas. We conclude by identifying a few directions in which we think the study of voice quality and identity might fruitfully proceed.

Throughout this article, we draw a variety of connections between linguistic signs (i.e., particular voice qualities) and identity. Before describing the kinds of

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

identities that the voice can signal, we review our semiotic framework for conceptualizing these connections. The voice has been called a bridge between the body and language, or between physicality and subjectivity (Cameron, 2001; Dolar, 2006). A person's voice is often understood to be a unique, inalienable property (Eidsheim, 2012). In televised genres such as Japanese animation, it is portrayed as a vehicle for the fixed, inner moral orientation of heroes and villains, beyond all possibility of disguise (Starr & Greene, 2006). As Cameron (2001) has pointed out, however, voices are also the object of design, curation, and manipulation across realms of social activity—not just in the media, but in speech technology and, of course, projects of identity construction, the voice is actively crafted to achieve various social meanings.

Cameron (2001) gave multiple examples of situations in which voices are actively "designed" or manipulated for profit motive: in south Indian call centers where telephone operators take on middle American accents, by "fantasy makers" at a phone sex hotline (Hall, 1995), at talking ATM machines, and so on. In this and the more mundane cases where an individual speaker seeks to manipulate his or her own voice for one reason or another, they are adjusting "something that has always been taken to be 'plastic,' easier to alter than almost any other physical characteristic" but that "is a bodily phenomenon, its unique qualities determined by unique characteristics of an individual body" (Cameron, 2001, p. 83). This tension between the voice's malleability and its closeness to the particularity of our own bodies is part of what bolsters the common belief "that voices are the willed and authentic expression of an individual's 'true' identity" (Cameron, 2001, p. 83).

That the body plays a direct role in determining parameters of the voice may tempt researchers to adopt an essentialist viewpoint. For example, Ohala (1983) argued that low pitch (and low formant frequencies) signals largeness and power, which are both associated with masculinity, given men's tendency to have larger larynxes (and larger vocal tracts). Subsequent work (Puts, Hodges, Cárdenas, & Gaulin, 2007) picks up this line of argumentation in accounting for the finding that lower-pitched men's voices are evaluated as sounding more dominant. However, the large larynxes required to produce low pitch sounds are frequently called upon to produce high pitch to convey emphasis and displays of affect, including anger, for example (Gussenhoven, 2004). The body is thus not a bedrock that determines the form of speech. Work on the body since Butler (1990) has refused to conceive of the body as a unitary phenomenon governed entirely by the laws of nature, focusing on the ways in which bodies are subject to historical processes. This allows us to put the voice under the lens of sociological analysis, while at the same time not denying its manifold connections with the physical realities of the body.

The ideological connection between the voice and the body means that the semiotic profile of voice quality is somewhat different from even other kinds of phonetic variation. Early work on the voice, recognized the dually social and individual nature of the voice, and compared its expressive functions to that of gesture (Sapir, 1927), or has understood the voice primarily as a repository of information "given off" by the speaker (Laver, 1968). In the Peircean framework (Laver & Trudgill, 1979; Parmentier, 1987; Silverstein, 1976) for understanding

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

how such meanings are conveyed, the voice is best taken as "indexical" of various sets of facts. For a sign to be indexical means that it represents its meaning by being somehow connected to it—by causality, proximity in space or time, or even through some kind of pointing gesture. The voice is impacted by a number of physical exigencies which leave behind unmistakable indexical footprints. For example, the amount of fluid in the nasal cavity, the presence or absence of inflammation in the vocal folds, or the degree of physical control a speaker can exert over the muscles controlling phonation—all possibly impacted by medical conditions—are some of the causative factors behind the voice's capacity to index what Laver (1968) captioned as "medical" information. Hearing the voice of someone who is affected by one of these conditions, hearers can (correctly) hear their illness.

Another way in which the sound of the voice can be linked to the meanings it conveys is through iconicity, or relations of resemblance. Stross (2013) understood the meanings conveyed by falsetto voice (and high pitch in general), for instance, as involving iconic resemblances. Though the voice is rarely purely iconic (except, for instance, where one voice is understood as resembling another), it is very common for the voice to index relations of resemblance, a phenomenon called iconic indexicality or rhematicity. For example, in Lachixío Zapotec, high pitch and falsetto phonation are used to index the metaphorical size of participants' social positions (Sicoli, 2007). Lower-ranked participants use higher pitch, indexically associated with small size, to iconically index their social smallness relative to their interlocutor. In other situations, speakers of Lachixío Zapotec can creatively index authority through low pitch and breathy phonation, which may also have a rhematic relationship with large body size. Importantly, however, the kinds of physical realities that language users understand the voice as indexing may themselves be culturally constrained. Shayan, Ozturk, & Sicoli (2011) explored two major ways that languages organize the means of describing pitch variability: one, the *high-low* axis most often used in English, and the other, a *thin-thick* axis often deployed in Turkish, Zapotec, and Farsi, noting that these axes appear to be grounded in different aspects of subjective experience both offering their affordances for iconic indexicality.

At the same time, immense amounts of theoretical work have been assembled on the topic of how *multiple* voices can be assembled in the speech of one talker (Tannen, 2007) or *animator* in Goffman's typology of participant roles (Goffman, 1974, 1979). The term *heteroglossia* (Bakhtin, 1982, 1984) is a popular caption for this set of phenomena. Work on heteroglossia and voices has generally defined the voice in terms of formal features far broader than those traditionally studied under the rubric of voice quality, with Agha (2005) expanding the definition of a contrast between voices to potentially encompass any metrical contrast (patterned difference on some formal dimension) between two segments of signifying material. While Agha's definition is far too broad for us to accept here, we do recognize that individual speakers can and do use the formal features ideologically tied to the "voice" in a narrower sense to perform "voicing contrasts" of the sort Agha describes. The performance of constructed dialogue (sometimes called reported speech) is one such arena, which we review in detail below. There is also

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

emerging research on "voice registers," mostly from Sicoli (2007, 2010), whose work we also review below, but we must pass over other key work in the study of heteroglossia (e.g., Keane, 2011), in some cases despite their brief treatments of formally relevant phenomena, such as the affectively valent "intonational shadows" in Don Gabriel's narrative of his son's death (Hill, 1995).

From the authors' perspective, the mere existence of intraspeaker variation in the formal features that fall under our rubric for "voice quality" is evidence enough to falsify ideologies that tie the voice inalienably to an embodied, essential identity unique to the voice's owner. But iconoclasm of this sort is not our central task, for it does not yield an account of how such essentialisms arise, or whether and how they come to order the ways in which the voice mediates social relations and is understood by social actors. It is clear that such an account must confront the voice as a carrier of meanings often grounded not in the linguistic system or the observable social universe, but in the human body and the basic acoustic properties of the world. At the same time, we must of course honor the complex role which human societies have played in regimenting not just the systems of meaning themselves but also the materials out of which signs are made—the speakers' bodies, vocal practices, and ways of hearing that walk before any instance of meaning construction.

VOICE QUALITY AS AN INDEX OF GROUP MEMBERSHIP

Community Patterns

As a launching point, we begin with the observation that speakers' voices have the potential to signal their membership in particular groups. Studies have examined both differences in patterns of voice quality production between social groups as well as the inferences drawn about speakers' identity category affiliations on the basis of their voices. In this section, we focus on the macro-social categories of gender, race and ethnicity, and class while highlighting the problems associated with viewing identity strictly in terms of categories. We also consider potential differences in prototypical voice quality patterns across languages, which can perhaps most clearly be observed in the case of bilinguals.

Sexual dimorphism in the bodies, and more specifically vocal tracts, of women and men yield sex-based vocal differences in pitch and perhaps phonation (e.g., Simpson, 2009). Given the tremendous variation within sex classes, cross-cultural differences, and longitudinal change, however, it is clear that physical differences merely define the envelope of vocal possibilities rather than determine where within the possible range women and men will fall. Nevertheless, listeners can determine speaker sex reliably on the basis of vowels in isolation (Coleman, 1971) and connected speech (Gelfer & Bennett, 2013). While voice quality plays some role discriminating sex class, listeners can perform above the level of chance even when voice quality cues are not present (Fellowes, Remez, & Rubin, 1997). In fact, listeners can accurately judge speaker sex even when speech is played

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

backwards or sped up (Lass, Mertz, & Kimmel, 1978), which is perhaps indicative of the primacy of sex class identification. Interactional context plays an important mediating role, as illustrated by the finding that after listeners have been listening to male voices, subsequent voices that they hear will sound more female (Zäske, Fritz, & Schweinberger, 2013).

Numerous studies have noted the tendency for women to produce breathier phonation than men, perhaps due to sex differences in laryngeal physiology (Henton & Bladon, 1985). However, recent work by Simpson (2012) calls into question the acoustic methods used to substantiate this trend, as the most frequently used metric, amplitude of the first harmonic minus the amplitude of the second harmonic (H1 − H2), artificially inflates this value for female speakers. In a similar vein, creaky voice has been found to predominate in the speech of men, perhaps because its characteristically low pitch levels are more compatible with men's relatively lower baseline F0, though this tendency is certainly subject to sociocultural influence. As Henton & Bladon (1988) reported, even though men are creakier than women in their study, men who speak a variety of received pronunciation exhibit less creak than men who originally hail from the north of England. More recently, studies in communities in the United States have found the opposite gender trend, with women in California (Yuasa, 2010) and Washington, DC (Podesva, 2010) producing higher rates of creaky voice than their male counterparts. Yuasa (2010) further reported intragender difference, such that Japanese women produce lower rates of creaky voice than their American female counterparts. Podesva (2010) also reported intragender differences in the rate with which women in Washington, DC use falsetto. While African American women use high rates of falsetto, both African American men and white women and men use falsetto at low rates. These striking intragender differences bring to light the important point that even though language users orient to gender, gender should not be viewed as an axis of identity that operates independently of other dimensions of identity, such as race or geographic region.

At the same time, even though gender norms cannot be identified on the basis of production patterns, such norms may serve as powerful objects of ideology. For example, in the case of transgender women, stereotypical notions of the female voice play an important role in the transition process. A. B. Hancock, Krissinger, & Owen (2011) reported that trans women's quality of life directly correlates with how their voice is perceived by others. While much work on the linguistic practices of transgender speakers focuses on F0 (McNeill, Wilson, Clark, & Deakin, 2008), as has also been the case for gay men (Gaudio, 1994; Smyth, Jacobs, & Rogers, 2003) and lesbians (Moonwomon-Baird, 1997; Van Borsel, Vandaele, & Corthals, 2013), many other linguistic features including phonation (Gorham-Rowan & Morris, 2006) come into play. For example, a case study of a 15-year-old trans female adolescent reports that voice therapy (which included targeted instruction on producing breathy voice) proved effective on the basis of acoustic and perceptual evidence as well as client reports (Hancock & Helenius, 2012). While this strand of research conducted by speech clinicians highlights the essentializing forces of hegemonic gender ideology, ethnographic approaches reveal that individuals may

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

have more complex relationships to dominant ideologies. For example, Zimman (2012) made the important point that even though the trans men in his yearlong ethnography show a longitudinal trend toward a more hegemonically masculine linguistic style, results vary widely from one linguistic feature to another, and more strikingly across individuals, some of whom identify as trans men or FTM (female to male) and others of whom identify as genderqueer.

The foregoing discussion of gendered voice quality uncovers some of the difficulties associated with conceptualizing gender as an identity category marker, in part because what is often indexed is not a disembodied gender identity, but a gendered social type. Paddock (1977), for example, described a low "growly" register used by men in Newfoundland, known as "roach" (p. 318); crucially, the term refers not to the speech style of all men, but a rougher type that might engage in seal hunting. Voice quality variation is a resource for managing the variety of personas a speaker might inhabit. Breathy voice might enable a phone sex operator to take on the persona of an individual who is perpetually aroused (Hall, 1995), while falsetto might enable a gay man to project a "diva" persona to escape an interactionally vulnerable position (Podesva, 2007). Creaky voice, meanwhile, has been used in Chinese TV drama to index a woman's dangerous sexuality (Callier, 2010) or the performance of a "sexy" persona by adult video actresses (Kajino & Moon, 2011).

It is likely because of these characterological indexes of voice quality that class differences have been observed in a number of communities, particularly in the United Kingdom (J. Esling, 1978; Henton & Bladon, 1988; Knowles, 1978; Stuart-Smith, 1999; Trudgill, 1974), where class is a particularly salient axis of social distinction. Working class speakers have been reported to produce more harsh voice in Edinburgh (J. Esling, 1978), as well as greater pharyngealization in both Edinburgh and Glasgow (Stuart-Smith, 1999). Descriptions of voice quality in these studies are extremely detailed. Consider Knowles's (1978) description of the voice quality characterizing the (working class) accent of Liverpool:

> In Scouse, the centre of gravity of the tongue is brought backwards and upwards, the pillars of the fauces are narrowed, the pharynx is tightened, and the larynx is displaced upwards. The lower jaw is typically held close to the upper jaw, and this position is maintained even for "open" vowels. The main auditory effect of this setting is the "adenoidal" quality of Scouse, which is produced even if the speaker's nasal passages are unobstructed. (Knowles, 1978, p. 89)

Importantly, while this characterization describes the phonetic content of the Scouse accent, so too does it include how the accent is subjectively evaluated and a description of its typical speaker's bodily hexis, both key components of a distinctive social type.

Character types have also been influential in theorizing the relationship between voice quality and race. While the discussion on creaky voice above focused primarily on European and European American speech, creaky voice has also been described as a prevalent feature of Chicano English. Mendoza-Denton (2011),

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

for example, discussed creaky voice's stylistic role in a variety of Chicano media sources. She argued that creaky voice constitutes one of many stylistic components of a "hardcore Chicano gangster" and can be harnessed to construct a "hardcore" persona in nonmediated contexts, as well. Creaky voice has also been reported as a salient feature of Maori English in New Zealand, as Szakay (2012) has found both that Maori English speakers are creakier (on the basis of the acoustic measure H1 − H2) than Pakeha English speakers, and that listeners utilize these phonation differences in the process of dialect identification.

The majority of work examining the voice quality–race connection has focused on African American identity. Studies have shown that listeners can correctly distinguish white from African American speakers on the basis of speech samples, including those lacking grammatical and lexical features of African American English (Irwin, 1977; Thomas & Reaser, 2004), suggesting that voice quality could play an important cue. While studies have not endeavored to isolate the specific components of voice quality that might participate in racial identification, studies have nonetheless found that nonmodal voice qualities are used in a variety of African American styles, such as a falsetto "battlin" style commonly used among users of Hip Hop Language (Alim, 2004, pp. 65–73) and a preacher style characterized by a "gravelly" and "strained" quality (Britt, 2011, pp. 216 and 223, respectively). While the latter has been characterized as a resource for garnering the respect typically granted to African American preachers, the related harsh voice quality examined in Moisik (2013) occurs primarily with the portrayal of stereotypes of African Americans in popular media. In other words, racialized voice qualities can at once serve as badges of cultural pride and vehicles of stereotype perpetuation.

Stereotypes linking kinds of voices with racialized others can have harmful consequences for housing and employment. As Purnell, Idsardi, & Baugh (1999) discovered, listeners can identify nonstandard, ethnically marked dialects on the basis of very short segments of speech (such as the word "hello"), and landlords use this information in discriminating against potential tenants. There is likely a racial component to this discrimination, as differences in evaluation have been observed for nonstandard accents associated with ethnic minorities compared to those associated with the racial majority; French-accented speakers fared more favorably in securing employment than Japanese-accented speakers (Hosoda & Stone-Romero, 2010). Although American listeners appear to be very sensitive to race-based distinctions in voice quality, different patterns are evident in other communities. In Singapore, for example, even though older listeners can discern ethnicity on the basis of speech, younger listeners cannot, a finding the author attributes to the combined effects of policy and individual ethnic consciousness (Tan, 2012).

In addition to gender, class, and race, voice quality can index linguistic identity. Languages (and dialects, as pointed out by Pittam, 1987)—most likely due to sociocultural influences—can differ from one another in their "phonetic setting," or language-specific long-term vocal configuration. Identifying cross-linguistic differences in phonetic setting is a project undertaken by Mennen, Scobbie, de

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

Leeuw, Schaeffler, and Schaeffler (2010) to aid in second language acquisition, under the rationale that these phonetic settings are particularly challenging to learn. This argument is supported by a study of 12 Castillian Spanish-Catalan bilinguals showing that speakers exhibited less vocal variability in their nondominant language (Bruyninckx, Harmegnies, Llisterri, & Poch-Olive, 1994), suggesting that cultural influences are only part of the story, and that language proficiency likely plays a role as well. Nevertheless, examinations of intraspeaker differences among bilinguals provide striking evidence for language-specific voice quality. In an examination of code-switching among two Finnish-English children, changes in voice quality often accompanied switches between play and negotiation of play, which were themselves typically accompanied by code-switching (Halmari & Smith, 1994). More recently, Cantonese-English bilinguals were reported to exhibit systematic differences in voice quality depending on language, with mean spectral energy higher and spectral tilt lower in Cantonese than in English (Ng, Chen, & Chan, 2012).

Media Representations

The sociolinguistics of mass-mediated communication has gained a fair amount of theoretical attention of late (Agha, 2007; Bell & Gibson, 2011; Coupland, 2009), and the questions mediated linguistic performance raise for our accounts of identity and social meaning are numerous. Certainly, voice quality is often an object of manipulation in the representation of identity in mass-mediated sources. Japanese Anime has been a particularly rich source of data on voice quality variation in mass media, and also serves as the springboard for Starr and Greene's (2006) exploration of "sweet" voice, a kind of "head voice" associated with mature, traditionally feminine characters. Harsh voice quality is sometimes used in the portrayal of anime villains (Teshigawara, 2003), and likewise for villains in Mandarin Chinese TV dramas (Callier, 2012).

Harsh voice quality is also used in a wide variety of mediated sources in the United States to introduce a stereotypical African American voice or induce a contrast between two racially different voices being animated by a single speaker (Moisik, 2013), such as when African American comedian Steve Harvey animates the voice of a Black employee being fired from his job (p. 206). In these circumstances, harsh voice is sometimes also indexical of aggression, which hints at the affordances of voice quality for circulating harmful stereotypes of different racial and ethnic groups, as discussed above. Among other features, stereotyped representations of American Indians in white U.S. mass media have often involved deep or gravelly voices (Meek, 2006). Creaky voice has been wrapped into portrayals of Chicana/os in the United States, either to index (and effect the circulation of) character traits like "hardcore" in the voice of a 1990s Chicano rapper or to conjure up the stereotype of the volatile, dangerous "cholo" character type in white-marketed video games (Mendoza-Denton, 2011). Mediated performances can also reveal stereotypes about the regional distribution of voice quality, as Pennock-Speck

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

(2005) argues in his examination of three American actresses who use far less creak when portraying British women than when portraying American women. In both Moisik's (2013) and Mendoza-Denton's (2011) work, the same nonmodal voice quality is used in media produced within marginalized ethnic/racial communities, for circulation within those same groups, but also appears in portrayals of members of those communities packaged for circulation to a wider, largely White audience.

Broadcast advertising is another rich source for voice quality variability, and the choices producers make about what kind of voice-over or announcer voice to use can impact listeners' response to the advertising (Gélinas-Chebat & Chebat, 1999). Pennock-Speck and Rubio (2009) have paid particular attention to voice quality in advertising, finding a lopsided distribution not just to the kinds of voice qualities present in voice-overs but to the identities of the speakers employed to do them. Men outnumber women in Pennock-Speck and Rubio's bilingual corpus and are heard to be more authoritative and assertive, evaluations possibly linked to their being heard as deeper, breathier, and huskier in English advertisements. This is an excellent example of Cameron's (2001) argument about the proliferation of vocal design as a cultural phenomenon.

VOICE QUALITY AS AN INDEX OF STANCE

In the preceding section, we noted that while voice quality's role in indexing group identity categories cannot be denied, voice quality indexes a great deal more than categories alone. Voice quality is used to create gendered, racialized, and classed personas. These personas gain substance, and circulability, not only from their component and intersectional identity categories, but also from the ways individuals embodying these categories position themselves in interaction—in other words, through stancetaking. Stancetaking can be understood as

> a public act by a social actor, achieved dialogically through overt communicative means (language, gesture, and other symbolic forms), through which social actors simultaneously evaluate objects, position subjects (themselves and others), and align with other subjects, with respect to any salient dimensions of the sociocultural field. (Du Bois, 2007, p. 163)

Although much of the previous literature treats stancetaking as a mostly discourse-level strategy (Englebretson, 2007; Jaffe, 2009), we suggest that it is also useful to think of the (phonetic) vocal dimensions of stancetaking, particularly in considerations of affective stance.

As we discussed in the introduction, voice quality can be seen as iconic, or rhematic, of particular physical embodiments of affect. This is perhaps most evident in the case of nonverbal vocalizations like laughter, sobbing, and screams (Lima, Castro, & Scott, 2013). Particular voice qualities are also closely linked to

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

numerous affective states, such as growls to anger and excitement (Tsai et al., 2010); smile prosody to amusement (Aubergé & Cathiard, 2003); low harmonic energy to tenderness, sadness, and eroticism, and higher harmonic energy to joy, anger, and fear (Guzman, Correa, Muñoz, & Mayerhoff, 2013); and breathiness to emphasis in Javanese (Thurgood, 2004). While it is tempting to directly associate voice qualities with particular affective displays, researchers in this area stress that there is a one-to-many mapping from voice quality to affect (Gobl & Ní Chasaide, 2003). Vocal configurations characterized by tense voice, for example, have been shown to correlate with perceptions of stress, anger, confidence, and formality, while lax (and creaky) states have been connected to sounding relaxed, bored, or intimate (Ni Chasaide, Gobl, Hardcastle, & Beck, 2004), and these associations may be further mediated by environmental factors like intonational context (Callier, 2013). Given the substantial affective indeterminacy of voice quality variation (Chun & Podesva, 2010), it perhaps goes without saying that the indexical meaning of particular voice qualities is complex—even if we consider only the affective dimensions of meaning. While it has been a common practice in the analysis of discourse to mark particular voice qualities and analyze them as contextualization cues (e.g., Gumperz, 1982), it seems that the communicative intent underlying modulations in voice quality are far from transparent. We advocate turning the analytical focus to the various indexical possibilities for particular, situated productions of marked voice qualities.

Taking an approach that both assumes multiple indexicalities of voice and situates its production in a specific sociocultural context (meant here to encompass interactants' ideologies about the situation in which talk is produced as well as the positions utterances occupy in a sequence of turns) facilitates an understanding of how voice quality variation serves multiple functions at once, where affective stances give rise to interactional moves that are not reducible to the display of emotion. For example, stage actors might employ affect-laden voices that additionally serve the function of projecting the voice through a large auditorium (Raphael & Scherer, 1987). If we take the perspective of the breathy-voiced fantasy line operator discussed in Hall (1995), the voice does not merely index the speaker's (imagined) aroused state, but enables its producer to gain the upper hand in the verbal transaction of the call. Similarly, falsetto enables the gay man discussed in Podesva (2007) not simply to produce expressiveness, but to channel expressiveness into the construction of a diva persona.

Little research has investigated the presumably innumerable ways in which affective indexicalities of voice quality differ cross-culturally (cf. Stross, 2013), though cases of language-specific difference point in their direction. In their comparison of French and Japanese, Rilliard, Shochi, Martin, Erickson, and Auberge (2009) drew a distinction between emotions, which they view as involuntary, and attitudinal expressions, which they treat as relatively more subject to cultural convention. Examining the acoustic correlates of sarcasm, Cheang and Pell (2008) observed the importance of a lowered mean F0, while in Cantonese, the F0 level is raised (Cheang & Pell, 2009). That cultures show systematic, conventionalized differences in how affective stances are expressed through the

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

voice is evidence that voice qualities gain social meaning through their *resemblance* to affective states (rather than directly through the physical experience of emotion).

Although voice quality is often marshalled into the interactional project of affective stancework, a great many other kinds of stance can be constructed from vocal material. As discussed in the introduction, Sicoli (2007, 2010) illustrated that speakers of Lachixío Zapotec use shifts in voice quality to manage interpersonal stance, with breathy voice indexing authority and falsetto indexing respect. Grivicic and Nilep (2004) argued that in English creaky voice, when appearing on the word "yeah," signals passive recipiency, indicating either a dispreference for the topic or disalignment between interactants.

Similar shifts in alignment are commonly evident in cases of reported speech, or constructed dialogue. Following Tannen (2007), we prefer the term "constructed dialogue" over "reported speech" since these representations of language are very often not verbatim but are instead approximations of speech or thought. Vocally, constructed dialogue may be characterized by an "expanded [pitch] range accompanied by a raised baseline" (Chafe, 2002, p. 298), or a shift from modal to nonmodal phonation. As Podesva (2010) reported, constructed dialogue contains higher incidence of falsetto, as well as breathy, whispery, and harsh voice, though it is noteworthy that creaky voice (which was much more common than other voice qualities in general) did not occur at greater rates in constructed dialogue. Marked voice qualities may enable speakers to distinguish between their voices in the present interaction with those of other speakers, or past or future selves. This "othering" function of voice quality shifts is evident in a recent study by Carroll (2013), who reported that a disc jockey uses a "bundle of interactional and acoustic resources to design his talk as belonging to someone else" (p. 259). Intriguingly, the same disc jockey uses similar vocal modifications when stylizing ethnically marked accents. In these cases of stylization, the disc jockey is not merely distinguishing his voice, but taking a stance toward what the ethnically marked accents represent. Similar stancetaking practices are evident in Günthner's (1999) study on constructed dialogue in German, where speakers modified voice quality to communicate their "perspective toward the quoted utterances as concordant or discordant" (p. 685) and Sicoli's (2007) analysis of constructed dialogue in a Zapotec narrative, where low-pitched breathy phonation was used to evaluate contrasting voices in the narrative (pp. 181–185). Chafe (2002) suggested that in cases of negative evaluation, constructed dialogue may be acoustically characterized by a lowered pitch baseline and compressed range, though other work on mocking suggests that pitch levels can also be elevated and ranges, expanded in cases of negative evaluation, as well (p. 298). Whatever the phonetic content of constructed speech, deviations from modal phonation serve as significant interactional resources, and the stances that they enact have the potential to reshape the social meanings of particular voice qualities (Chun & Podesva, 2010, who showed how the indexical field for nonmodal phonation expands in unfolding interaction).

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

VOICE QUALITY AND INDIVIDUAL SPEAKER IDENTITY

While in the previous section we emphasized the ways in which an individual's voice may vary, many research areas touching on issues around the voice and identity have focused on the ways in which the voice is fixed to a particular speaker. Research in forensic voice identification and voice comparison seeks to validate or elaborate on the ways in which the justice system associates voices with individuals or compares two or more voice samples to each other. Ordinary language users frequently perform very similar speaker identification tasks, if with lower stakes—identifying talkers by their voice or using speaker-specific voice information to assist in language processing. There is a good deal of promising work in this area dealing with phonetic features that fall under our definition of phonetic voice quality. There is, however, a good deal of room for these literatures to mature, particularly in the emerging research on the shared encoding and processing of voices and faces.

Forensics

The role of voice and voice quality in the law often involves tying a particular voice, either recorded or qualitatively described, to a particular person. Research into "earwitness" testimony (Jessen, 2008; Laub, Wylie, & Bornstein, 2013) and vocal disguise (Andruski, Brugnone, & Meyers, 2007; Neuhauser, 2008; Zhang & Tan, 2008) have obvious forensic implications, as audio recordings are often used as legal evidence and must be able to be reliably associated with the correct speaking body. Although progress has been made both in validating existing legal practices with regard to speaker identification, as well as developing new techniques for doing so, for instance in exceptional vocal circumstances such as shouting (Blatchford & Foulkes, 2006), a consensus appears to be emerging that earwitness testimony is less reliable than eyewitness testimony (Hollien, Bennett, & Gelfer, 1983; Laub et al., 2013; Read & Craik, 1995), which has itself been the target of much critical scrutiny. Some forensic phoneticians have disavowed the task of speaker identification as part of their role altogether, instead preferring to characterize it as the comparison of discrete speech samples (French & Harrison, 2007). The role of technology in assisting these forensic aims appears to be increasing, though manual speaker recognition still outperforms automatic speaker recognition (Zhang & Tan, 2008).

Speech Perception

An important dimension of the relation between the voice and identity is how listeners encode and process voices. A couple of recent introductions to talker identity recognition are given by Creel & Bregman (2011) and Schweinberger, Kawahara, Simpson, Skuk, & Zäske (2014). Information about the talker is useful in identifying linguistic information (Nygaard & Pisoni, 1998; Sumner, Kim, King, & McGowan, 2014). Cues to speaker sex are especially useful, as listeners ordinarily use speaker sex in speaker recognition (Zäske et al., 2013), although they

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

are still able to do speaker recognition without obvious sex-related information (Fellowes et al., 1997).

Intonational cues may play a particularly important role in talker recognition (Abberton & Fourcin, 1978), in addition to other properties of fundamental frequency (F0) (Dommelen, 1987). Listeners' perception of the voice depends on expectations about how the speaker talks in general; identically high F0 peaks are judged less prominent in the speech of women than in that of men (Gussenhoven & Rietveld, 1998).

Voice quality is an important part of how people recognize speakers, though it is unclear just how important it is. Pittam (1987) reviewed early literature on speaker identification with spectral measures, especially the long-term average spectrum. Voice quality is not necessary to do speaker recognition. Listeners are very good at using phonetic information, even when the laryngeal voice source is replaced with a sine wave complex, to recognize familiar talkers (Remez, Fellowes, & Rubin, 1997). However, people may use variation in utterance-final phonation type to do speaker recognition (Bőhm & Shattuck-Hufnagel, 2009).

The kinds of cues people use in recognizing voices are encoded in memory along with other information identifying the talker. Latinus and Belin (2011) trained listeners to recognize a few different speakers, then used artificially modified versions of the speakers' voices and asked listeners to do a forced-choice speaker identification task. The modified voices improved listeners' propensity to categorize a voice averaged between all the speakers as belonging to the particular speaker whose voice had been so modified. According to the authors this provides evidence that listeners encode the identity of speakers in terms of vocal prototypes—as opposed to episodic exemplars or clusters of acoustic values. Vocal recognition appears to be similar to face recognition in this way. It has recently come to researchers' attention, in fact, that voice-related information may be encoded quite closely alongside information related to speakers' faces. Kamachi, Hill, Lander, and Vatikiotis-Bateson (2003) provided evidence for the shared encoding of faces and voices. González et al. (2011) and Föcker, et al. (2011) used EEG (electroencephalography) to study face and voice processing, and found that interactions between face- and voice-related activation occurs both in auditory processing (early in the processing timeline) as well as conceptual processing (late in the processing timeline).

Speech Technology

Speaker recognition is a task which speech scientists are actively trying to make computers better at performing (Altincay & Demirekler, 2000), and have been pursuing for some time (Hargreaves & Starkweather, 1963). Some such research has sought to conduct speaker identification under a variety of vocal conditions, including shouting (Shahin, 2006), but it has been rare to address the utility of voice quality in particular for these tasks. Nevertheless, speech technology has gotten better both at recognizing speaker identity (Zhang & Tan, 2008) and at conforming to the designs of speech designers (Cameron, 2001).

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

Though much research relating the voice to individual speaker identity presumes or searches for a stable, invariant link between the speaking body and at least some part of the speech signal, work on speaker recognition in various speaking conditions and forensic work critiquing the utility of earwitness testimony destabilizes such links, recognizing that both voices and speakers are complex, variable phenomena. Work linking the processing of voice identity to the processing of face identity, meanwhile, indicates that the contributions of voice quality to how we identify speakers are wrapped up with cues across signifying channels and sensory modalities. Though voice is enough, under the right circumstances, to identify a speaker, under ordinary circumstances speaker identity is an inference made from a combination of visual, auditory, and top-down cues.

CONCLUSION

The diversity of ways in which settings of the larynx and the nearby articulatory apparatus can relate to the construction of speaker identity is astonishing, but a recurring theme is that the styling of the voice "is a form of artifice whose most valued products are 'naturalness' and 'authenticity'" (Cameron, 2001, p. 83). The tie between the voice and the uniqueness of the speaker's body is substantiated by ideologies which at the same time erase the mutability and inherent fluctuation of the voice. In popular media, voice quality is an easy mix-in for indexing a character's (fixed) place in the moral universe (Callier, 2012; Starr & Greene, 2006; Teshigawara, 2003).

Treating the voice as inalienable from its speaker is an act of essentialism rooted in a view of the voice as a part of the body. While we do not advocate essentializing the body's role in determining the social meaning of voice qualities, the body is nevertheless a powerful medium through which the meaning potential of the voice is refracted. Among the various resources used for speech, voice quality is perhaps "closest" to the body and furthest from the abstractions of linguistic structure (Karpiński, 2012), at least in languages where voice quality is not phonologically contrastive. The work which is done by this closeness of voice and body has not been adequately explored. The ways in which the body itself is ideologically constituted—in classifications of sex, gender, and sexuality, in processes of racialization, in the valorization of "health" and over illness and disability—are all likely to impact the relationship between speakers, their voices, and their bodies. Work in this area by Zimman (Zimman, 2013; Zimman & Hall, 2009) is pioneering in this regard. Similarly, the discursive constitution of the voice as indexical of the personal and the individual, rather than the collective and the social, is likely to be a fruitful source of inquiry, as indicated by the tone of panic in some media coverage of vocal fry among American females, which lament the idea that women might be picking up "vocal patterns" (Quenqua, 2012).

The extent to which the voice indeed *is* linked to other ways in which the body acts as a medium for signifying behavior is indicated by the recent explosion of work in multimodal communication, encompassing voice, prosody, and gesture,

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

(Karpiński, 2012), as well as facial affect (Aubergé & Cathiard, 2003). In particular, the link between the face and the voice appears to be particularly intimate (Föcker et al., 2011; González et al., 2011; Kamachi et al., 2003; Lander, Hill, Kamachi, & Vatikiotis-Bateson, 2007; Rilliard et al., 2009), and yet there has not been a substantial effort on the part of researchers in language and social interaction to investigate how the face and voice operate as semiotic channels, in tandem and separately. Are certain facial expressions quantitatively paired with certain voice qualities? Do mismatches between facial expression and voice quality have particular interactional consequences?

The theoretical contributions of such work could be immense, not least as a corrective against how popular ideologies have constrained the ways in which the voice is understood by linguists and ordinary language users alike. As much as the voice has been called upon to act as a totalizing index of a speaker's whole personhood, its meanings are often indeterminate (Chun & Podesva, 2010), and projects of identity construction are always partial (Bucholtz & Hall, 2005), whether from the standpoint of the speaker or of the listener. We call for concerted effort from researchers at all intersections of language and social processes to address voice quality as a site for identity work. Given our present-day understanding of how ideology shoots through not just the social meaning of language but also that of the body and the individual, we can and should seek to better understand links between bodies, speech, and identities without fear of shoring up potentially damaging essentialisms and inequalities.

ANNOTATED BIBLIOGRAPHY

Cameron, D. (2001). Language: Designer voices. *Critical Quarterly*, *43*(4), 81–85. doi:10.1111/1467-8705.00392

This brief critical piece situates the voice as a product of both physical and cultural processes.

Laver, J. D. M. (1968). Voice quality and indexical information. *International Journal of Language & Communication Disorders*, *3*(1), 43–54.

Laver, J. D. M. (1980). *The phonetic description of voice quality*. Cambridge, UK: Cambridge University Press.

These foundational works catalogue the range of communicative functions served by voice quality and provide a comprehensive framework for phonetically characterizing voice quality.

Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics*, *11*(4), 478–504.

Sicoli, M. A. (2010). Shifting voices with participant roles: Voice qualities and speech registers in Mesoamerica. *Language in Society*, *39*(04), 521–553. doi:10.1017/S0047404510000436

These two articles represent recent moves toward conceptualizing voice quality as an interactional resource.

Thomas, E. R. (2011). Voice quality. In *Sociophonetics: An introduction* (pp. 224–250). London, UK: Palgrave Macmillan.

This chapter provides a comprehensive overview on up-to-date methods for the acoustic analysis of voice quality.

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

Zimman, L. (2012). *Voices in transition: Testosterone, transmasculinity, and the gendered voice among female-to-male transgender people*. (Unpublished doctoral dissertation). Boulder, CO: University of Colorado.

This dissertation locates the voice in the body, treated both as a physiological and social construct.

REFERENCES

Abberton, E., & Fourcin, A. J. (1978). Intonation and speaker identification. *Language and Speech*, *21*(4), 305–318.

Abercrombie, D. (1967). *Elements of general phonetics* (Vol. 203). Edinburgh, Scotland: Edinburgh University Press.

Agha, A. (2005). Voice, footing, enregisterment. *Journal of Linguistic Anthropology*, *15*(1), 38–59.

Agha, A. (2007). Recombinant selves in mass-mediated spacetime. *Language and Communication*, *27*(3), 320–335.

Alim, H. S. (2004). *You know my steez: An ethnographic and sociolinguistic study of styleshifting in a Black American speech community*. Durham, NC: Duke University Press.

Altincay, H., & Demirekler, M. (2000). An information theoretic framework for weight estimation in the combination of probabilistic classifiers for speaker identification. *Speech Communication*, *30*(4), 255–272.

Andruski, J. E., Brugnone, N., & Meyers, A. (2007). The value of F0, F3, and F4 in identifying disguised voices. *The Journal of the Acoustical Society of America*, *121*(5), 3188.

Aubergé, V., & Cathiard, M. (2003). Can we hear the prosody of smile? *Speech Communication*, *40*(1–2), 87–97. doi:10.1016/S0167-6393(02)00077-8

Bakhtin, M. M. (1982). Discourse in the novel. In M. M. Bakhtin, *The dialogic imagination: Four essays* (M. Holquist, ed. and trans.) (pp. 259–422). Austin: University of Texas Press. (Original work published 1935)

Bakhtin, M. M. (1984). *Problems of Dostoevsky's poetics*. (C. Emerson, ed. and trans.) Minneapolis: University of Minnesota Press. (Original work published 1929)

Bell, A., & Gibson, A. (2011). Staging language: An introduction to the sociolinguistics of performance. *Journal of Sociolinguistics*, *15*(5), 555–572. doi:10.1111/j.1467-9841.2011.00517.x

Blatchford, H., & Foulkes, P. (2006). Identification of voices in shouting. *The International Journal of Speech, Language and the Law*, *13*(2), 241–254.

Bőhm, T., & Shattuck-Hufnagel, S. (2009). Do listeners store in memory a speaker's habitual utterance–final phonation type? *Phonetica*, *66*(3), 150–168. doi:10.1159/000235658

Britt, E. (2011). "Can the church say amen": Strategic uses of black preaching style at the State of the Black Union. *Language in Society*, *40*(2), 211–233. doi:http://dx.doi.org/10.1017/S0047404511000042

Bruyninckx, M., Harmegnies, B., Llisterri, J., & Poch-Olive, D. (1994). Language-induced voice quality variability in bilinguals. *Journal of Phonetics*, *22*(1), 19–31.

Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, *7*(4–5), 585–614.

Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. New York, NY: Routledge.

Callier, P. (2010). *Voice quality, rhythm and valorized femininities*. Presented at Sociolinguistics Symposium 18, Southampton, UK.

Callier, P. (2012). *Variation in phonation type. Distributions and meanings in a mass-mediated context.* Presented at New Ways of Analyzing Variation 41, Bloomington, IN.

Callier, P. (2013). *Linguistic context and the social meaning of voice quality variation* (Unpublished doctoral dissertation). Washington, DC: Georgetown University.

Cameron, D. (2001). Language: Designer voices. *Critical Quarterly*, *43*(4), 81–85. doi:10.1111/1467-8705.00392

Carroll, H. B. (2013). Identifying stylizations in ethnically salient talk among disc jockeys. *Language in Society*, *42*(3), 259–286. doi:http://dx.doi.org/10.1017/S0047404513000237

Catford, J. C. (1964). Phonation types: The classification of some laryngeal components of speech production. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, & J. L. L. Trim (eds.), *In honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday, 12 September 1961* (pp. 26–37). London, UK: Longmans, Green.

Chafe, W. (2002). *Prosody and emotion in a sample of real speech*. Retrieved from http://site.ebrary.com/lib/stanford/Doc?id=10224765

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

Cheang, H. S., & Pell, M. D. (2008). The sound of sarcasm. *Speech Communication*, *50*(5), 366–381.
Cheang, H. S., & Pell, M. D. (2009). Acoustic markers of sarcasm in Cantonese and English. *The Journal of the Acoustical Society of America*, *126*(3), 1394–1405.
Chun, E., & Podesva, R. J. (2010). *Voice quality and indeterminacies of social meaning in constructed dialogue*. Paper presented at Sociolinguistics Symposium 18, Southampton, UK.
Coleman, R. O. (1971). Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech Language and Hearing Research*, *14*(3), 565. doi:10.1044/jshr.1403.565
Coupland, N. (2009). The mediated performance of vernaculars. *Journal of English Linguistics*, *37*(3), 284–300. doi:10.1177/0075424209341188
Creel, S. C., & Bregman, M. R. (2011). How talker identity relates to language processing. *Linguistics and Language Compass*, *5*(5), 190–204. doi:10.1111/j.1749-818X.2011.00276.x
DiCanio, C. (2009). The phonetics of register in Takhian Thong Chong. *Journal of the International Phonetic Association*, *39*(2), 162–188.
Dolar, M. (2006). *A voice and nothing more*. Cambridge, MA: MIT Press.
Dommelen, W. A. van. (1987). The contribution of speech rhythm and pitch to speaker recognition. *Language and Speech*, *30*(4), 325–338.
Dong, L., Sundberg, J., & Kong, J. (2014). Loudness and pitch of Kunqu Opera. *Journal of Voice*, *28*(1), 14–19. doi:10.1016/j.jvoice.2013.07.012
Du Bois, J. W. (2007). The stance triangle. In R. Englebretson (ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction* (pp. 139–182). Amsterdam, The Netherlands: John Benjamins.
Eidsheim, N. (2012). Voice as action: Toward a model for analyzing the dynamic construction of racialized voice. *Current Musicology*, (93), 9–33,152.
Englebretson, R. (ed.). (2007). *Stancetaking in discourse: Subjectivity, evaluation, interaction* (vol. 164). Amsterdam, The Netherlands: John Benjamins.
Esling, J. (1978). The identification of features of voice quality in social groups. *Journal of the International Phonetic Association*, *8*(1–2), 18–23.
Esling, J. H., & Harris, J. G. (2004). States of the glottis: An articulatory phonetic model based on laryngoscopic observations. In W. J. Hardcastle & J. M. Beck (eds.), *A figure of speech: a Festschrift for John Laver* (pp. 347–83). Mahwah, NJ: Routledge.
Fellowes, J. M., Remez, R. E., & Rubin, P. E. (1997). Perceiving the sex and identity of a talker without natural vocal timbre. *Perception & Psychophysics*, *59*(6), 839–849.
Föcker, J., Hölig, C., Best, A., & Röder, B. (2011). Crossmodal interaction of facial and vocal person identity information: An event-related potential study. *Brain Research*, *1385*, 229–245. doi:10.1016/j.brainres.2011.02.021
French, P., & Harrison, P. (2007). Position statement concerning use of impressionistic likelihood terms in forensic speaker comparison cases. *International Journal of Speech, Language and the Law*, *14*(1), 137–144.
Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association*, *41*(2), 185–205. doi:10.1017/S0025100311000193
Gaudio, R. P. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech*, *69*(1), 30–57. doi:10.2307/455948
Gelfer, M. P., & Bennett, Q. E. (2013). Speaking fundamental frequency and vowel formant frequencies: Effects on perception of gender. *Journal of Voice*, *27*(5), 556–566. doi:10.1016/j.jvoice.2012.11.008
Gélinas-Chebat, C., & Chebat, J.-C. (1999). Impact of voice on source credibility in advertising: A self-monitoring approach. *North American Journal of Psychology*, *1*(2), 323–342.
Gobl, C., & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, *40*(1–2), 189–212.
Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Cambridge, MA: Harvard University Press.
Goffman, E. (1979). Footing. *Semiotica*, *25*(1–2), 1–30. doi:10.1515/semi.1979.25.1–2.1
González, I. Q., León, M. A. B., Belin, P., Martínez-Quintana, Y., García, L. G., & Castillo, M. S. (2011). Person identification through faces and voices: An ERP study. *Brain Research*, *1407*, 13–26. doi:10.1016/j.brainres.2011.03.029
Gordon, M. (2001). Linguistic aspects of voice quality with special reference to Athabaskan. In S. Tuttle & G. Holton (eds.), *Proceedings of the 2001 Athabaskan Languages Conference. Fairbanks: Alaska Native Language Center* (pp. 163–178). Los Angeles, CA.

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

Gordon, M., & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, *29*(4), 383–406.
Gorham-Rowan, M., & Morris, R. (2006). Aerodynamic analysis of male-to-female transgender voice. *Journal of Voice*, *20*(2), 251–262. doi:10.1016/j.jvoice.2005.03.004
Gorisch, J., Wells, B., & Brown, G. J. (2012). Pitch contour matching and interactional alignment across turns: An acoustic investigation. *Language and Speech*, *55*, 57–76.
Grivicic, T., & Nilep, C. (2004). When phonation matters: The use and function of yeah and creaky voice. *Colorado Research in Linguistics*, *17*, 1–11.
Gumperz, J. J. (1982). *Discourse strategies* (vol. 1). Cambridge, UK: Cambridge University Press.
Günthner, S. (1999). Polyphony and the "layering of voices" in reported dialogues: An analysis of the use of prosodic devices in everyday reported speech. *Journal of Pragmatics*, *31*(5), 685–708. doi:10.1016/S0378-2166(98)00093-9
Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.
Gussenhoven, C., & Rietveld, T. (1998). On the speaker-dependence of the perceived prominence of F0 peaks. *Journal of Phonetics*, *26*(4), 371–380. doi:10.1006/jpho.1998.0080
Guzman, M., Correa, S., Muñoz, D., & Mayerhoff, R. (2013). Influence on spectral energy distribution of emotional expression. *Journal of Voice*, *27*(1), 129.e1–129.e10. doi:10.1016/j.jvoice.2012.08.008
Hall, K. (1995). Lip Service on the Fantasy Lines. In M. Bucholtz & K. Hall (eds.), *Gender articulated: Language and the socially constructed self*. New York, NY: Routledge.
Halmari, H., & Smith, W. (1994). Code-switching and register shift: Evidence from Finnish-English child bilingual conversation. *Journal of Pragmatics*, *21*(4), 427–445. doi:10.1016/0378-2166(94)90013-2
Hancock, A. B., Krissinger, J., & Owen, K. (2011). Voice perceptions and quality of life of transgender people. *Journal of Voice*, *25*(5), 553–558.
Hancock, A., & Helenius, L. (2012). Adolescent male-to-female transgender voice and communication therapy. *Journal of Communication Disorders*, *45*(5), 313–324. doi:10.1016/j.jcomdis.2012.06.008
Hargreaves, W. A., & Starkweather, J. A. (1963). Recognition of speaker identity. *Language and Speech*, *6*(2). Retrieved from https://search.proquest.com/docview/1299110668/B1943AB88D9467APQ/88?accountid=14026
Heard, D. C. (2012). "DON'T LET ME BE MISUNDERSTOOD": Nina Simone's Theater of Invisibility. *Callaloo*, *35*(4), 1056–1084,1124.
Henton, C. G., & Bladon, A. (1988). Creak as a sociophonetic marker. In L. Hyman, V. Fromkin & C. N. Li (eds.), *Language, speech and mind: Studies in honour of Victoria A. Fromkin* (pp. 3–29). London, UK: Routledge.
Henton, C. G., & Bladon, R. A. W. (1985). Breathiness in normal female speech: Inefficiency versus desirability. *Language & Communication*, *5*(3), 221–227. doi:10.1016/0271-5309(85)90012-6
Hill, J. (1995). The voices of Don Gabriel: Responsibility and self in a Modern Mexicano narrative. In D. Tedlock & B. Mannheim (eds.), *The dialogic emergence of culture* (pp. 97–147). Champaign: University of Illinois Press.
Hollien, H., Bennett, G., & Gelfer, M. P. (1983). Criminal identification comparison: aural versus visual identifications resulting from a simulated crime. *Journal of Forensic Sciences*, *28*(1), 208–221.
Hosoda, M., & Stone-Romero, E. (2010). The effects of foreign accents on employment-related decisions. *Journal of Managerial Psychology*, *25*(2), 113–132. doi:10.1108/02683941011019339
Irwin, R. B. (1977). Judgments of vocal quality, speech fluency, and confidence of southern Black and White speakers. *Language and Speech*, *20*(3), 261–266.
Jaffe, A. (2009). *Stance: sociolinguistic perspectives*. New York, NY: Oxford University Press.
Jessen, M. (2008). Forensic phonetics. *Language and Linguistics Compass*, *2*(4), 671–711. doi:10.1111/j.1749-818X.2008.00066.x
Johnson, A. M., & Kempster, G. B. (2011). Classification of the classical male singing voice using long-term average spectrum. *Journal of Voice*, *25*(5), 538–543. doi:10.1016/j.jvoice.2010.05.009
Kajino, S., & Moon, K. (2011). *Voice quality variation of Japanese porn actresses: The stylistic construction of sexual sweetness*. Paper presented at New Ways of Analyzing Variation 40, Washington, DC: Georgetown University.
Kamachi, M., Hill, H., Lander, K., & Vatikiotis-Bateson, E. (2003). "Putting the face to the voice": Matching identity across modality. *Current Biology*, *13*(19), 1709–1714. doi:10.1016/j.cub.2003.09.005
Karpiński, M. (2012). The boundaries of language: Dealing with paralinguistic features. *Lingua Posnaniensis*, *54*(2), 37–54. doi:10.2478/v10122–012–0013–1

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

Keane, W. (2011). Indexing voice: A morality tale. *Journal of Linguistic Anthropology*, *21*(2), 166–178. doi:10.1111/j.1548-1395.2011.01104.x

Knowles, G. (1978). The nature of phonological variables in Scouse. In P. Trudgill (ed.), *Sociolinguistic patterns in British English*. London, UK: Edward Arnold.

Kreiman, J., Vanlancker-Sidtis, D., & Gerratt, B. R. (2005). Perception of voice quality. In D. Pisoni & R. Remez (eds.), *Handbook of speech perception* (pp. 338–362). Malden, MA: Blackwell.

Kuwabara, H., & Ohgushi, K. (1984). Experiments on voice qualities of vowels in males and females and correlation with acoustic features. *Language and Speech*, *27*(2), 135–45. Retrieved from https://search.proquest.com/docview/1299120486/citation/B1943AB88D9467APQ/123?accountid=14026

Ladefoged, P. (1988). Discussion of phonetics: A note on some terms for phonation types. In O. Fujimura (ed.), *Vocal physiology: Voice production, mechanisms and functions* (pp. 297–317). New York, NY: Raven Press.

Lander, K., Hill, H., Kamachi, M., & Vatikiotis-Bateson, E. (2007). It's not what you say but the way you say it: Matching faces and voices. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 905–914. doi:10.1037/0096–1523.33.4.905

Lass, N. J., Mertz, P. J., & Kimmel, K. L. (1978). The effect of temporal speech alterations on speaker race and sex identifications. *Language and Speech*, *21*(3), 279–290.

Latinus, M., & Belin, P. (2011). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, *2*(175), 1–12. doi:10.3389/fpsyg.2011.00175

Laub, C. E., Wylie, L. E., & Bornstein, B. H. (2013). Can the courts tell an ear from an eye? Legal approaches to voice identification evidence. *Law & Psychology Review*, *37*, 119–158.

Laver, J. D. M. (1968). Voice quality and indexical information. *International Journal of Language & Communication Disorders*, *3*(1), 43–54.

Laver, J. D. M. (1980). *The phonetic description of voice quality*. Cambridge, UK: Cambridge University Press.

Laver, J. D. M., & Trudgill, P. (1979). Phonetic and linguistic markers in speech. In K. R. Scherer & H. Giles (eds.), *Social markers in speech* (Vol. 1). Cambridge, UK: Cambridge University Press.

Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behavior Research Methods*, *45*(4), 1234–1245.

McNeill, E. J. M., Wilson, J. A., Clark, S., & Deakin, J. (2008). Perception of voice in the transgender client. *Journal of Voice*, *22*(6), 727–733. doi:10.1016/j.jvoice.2006.12.010

Meek, B. A. (2006). And the Injun goes "How!" Representations of American Indian English in white public space. *Language in Society*, *35*(1), 93–128.

Mendoza-Denton, N. (2011). The Semiotic hitchhiker's guide to creaky voice: Circulation and gendered hardcore in a Chicana/o gang persona. *Journal of Linguistic Anthropology*, *21*(2), 261–280.

Mennen, I., Scobbie, J. M., de Leeuw, E., Schaeffler, S., & Schaeffler, F. (2010). Measuring language-specific phonetic settings. *Second Language Research*, *26*(1), 13–41. doi:10.1177/0267658309337617

Moisik, S. R. (2013). Harsh voice quality and its association with blackness in popular American media. *Phonetica*, *69*(4), 193–215.

Moonwomon-Baird, B. (1997). Toward a study of lesbian speech. In A. Livia & K. Hall (eds.), *Queerly Phrased: Language, Gender, and Sexuality* (pp. 202–213). New York, NY: Oxford University Press.

Neuhauser, S. (2008). Voice disguise using a foreign accent: Phonetic and linguistic variation. *The International Journal of Speech, Language and the Law*, *15*(2), 131–159.

Ng, M. L., Chen, Y., & Chan, E. Y. K. (2012). Differences in vocal characteristics between Cantonese and English produced by proficient Cantonese-English Bilingual speakers—A long-term average spectral analysis. *Journal of Voice*, *26*(4), e171–e176. doi:10.1016/j.jvoice.2011.07.013

Ni Chasaide, A., Gobl, C., Hardcastle, W. J., & Beck, J. M. (2004). Phonatory quality and affect. In W. J. Hardcastle & J. M. Beck (eds.), *A figure of speech: A Festschrift for John Laver* (pp. 323–346). Mahwah, NJ: Routledge.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics*, *60*(3), 355–376.

Ogden, R. (2001). Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association*, *31*(1), 139–152.

Ogden, R. (2004). Non-modal voice quality and turn-taking in Finnish. In E. Couper-Kuhlen & C. E. Ford (eds.), *Sound Patterns in Interaction* (pp. 29–62). Amsterdam, The Netherlands: John Benjamins.

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica*, *40*, 1–18.

Paddock, H. (1977). On roachness in English. *Phonetica*, *34*(4), 318–318. doi:10.1159/000259894

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

Parmentier, R. J. (1987). Peirce divested for non-intimates. *Semantic Inquiry*, *7*, 19–39.
Pennock-Speck, B. (2005). The changing voice of women. In J. J. C. García de Leonardo, J. T. Pérez, M. del Saz Rubio, C. M. Cuenca, B. Pennock Speck & M. J. Coperías Aguilar (eds.), *Actas: XXVIII Congreso Internacional, AEDEAN*. Valencia, Spain
Pennock-Speck, B., & Rubio, M. M. del S. (2009). Voice-overs in standardized English and Spanish television commercials. *Atlantis*, *31*(1), 111–127.
Pittam, J. (1987). The long-term spectral measurement of voice quality as a social and personality marker: A review. *Language and Speech*, *30*(1).
Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics*, *11*(4), 478–504.
Podesva, R. J. (2010). *Gender and the social meaning of non-modal phonation types*. Paper presented at the Berkeley Linguistics Society, Berkeley, CA.
Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*, *18*(1), 10–30. doi:10.1177/0261927´99018001002
Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. (2007). Men's voices as dominance signals: Vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, *28*(5), 340–344.
Quenqua, D. (2012). Young women often trendsetters in vocal patterns. *New York Times*, February 27, 2012. Retrieved from http://www.nytimes.com/2012/02/28/science/young-women-often-trendsetters-in-vocal-patterns.html
Raphael, B. N., & Scherer, R. C. (1987). Voice modifications of stage actors: Acoustic analyses. *Journal of Voice*, *1*(1), 83–87. doi:10.1016/S0892-1997(87)80029-2
Read, D., & Craik, F. I. (1995). Earwitness identification: Some influences on voice recognition. *Journal of Experimental Psychology: Applied*, *1*(1), 6.
Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 651–666. doi:10.1037/0096–1523.23.3.651
Rilliard, A., Shochi, T., Martin, J.-C., Erickson, D., & Auberge, V. (2009). Multimodal indices to Japanese and French prosodically expressed social affects. *Language and Speech*, *52*(2–3), 223–243. doi:10.1177/0023830909103171
Sapir, E. (1927). Speech as a personality trait. *American Journal of Sociology*, *32*(6), 892–905.
Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G., & Zäske, R. (2014). Speaker perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(1), 15–25. doi:10.1002/wcs.1261
Shahin, I. (2006). Enhancing speaker identification performance under the shouted talking condition using second-order circular hidden Markov models. *Speech Communication*, *48*(8), 1047–1055. doi:10.1016/j.specom.2006.01.005
Shayan, S., Ozturk, O., & Sicoli, M. A. (2011). The thickness of pitch: Crossmodal metaphors in Farsi, Turkish, and Zapotec. *The Senses and Society*, *6*(1), 96–105. doi:10.2752/174589311X12893982233911
Sicoli, M. A. (2007). *Tono: A linguistic ethnography of tone and voice in a Zapotec region (Mexico)* (Unpublished doctoral dissertation). Ann Arbor, MI: University of Michigan.
Sicoli, M. A. (2010). Shifting voices with participant roles: Voice qualities and speech registers in Mesoamerica. *Language in Society*, *39*(4), 521–553. doi:10.1017/S0047404510000436
Sikveland, R. O. (2012). Negotiating towards a next turn: Phonetic resources for "doing the same." *Language and Speech*, *55*, 77–98.
Silverstein, M. (1976). Shifters, linguistic categories and cultural description. In *Meaning in anthropology*. Santa Fe, NM: University of New Mexico Press.
Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*(2), 621–640.
Simpson, A. P. (2012). The first and second harmonics should not be used to measure breathiness in male and female voices. *Journal of Phonetics*, *40*, 477–490.
Smyth, R., Jacobs, G., & Rogers, H. (2003). Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society*, *32*(03), 329–350.
Starr, R., & Greene, R. (2006). *Beyond cuteness: The role of voice quality in performing stylized femininities in Japanese*. Presented at New Ways of Analyzing Variation 36, Columbus, OH.
Stross, B. (2013). Falsetto voice and observational logic: Motivated meanings. *Language in Society*, *42*(2), 139–162. doi:10.1017/S004740451300002X

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press

Stuart-Smith, J. (1999). Glasgow: Accent and voice quality. In P. Foulkes & G. Docherty (eds.), *Urban voices: Accent studies in the British Isles* (pp. 201–222). Leeds, UK: Arnold.

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Language Sciences*, *4*, 1015. doi:10.3389/fpsyg.2013.01015

Szakay, A. (2012). Voice quality as a marker of ethnicity in New Zealand: From acoustics to perception. *Journal of Sociolinguistics*, *16*(3), 382–397. doi:10.1111/j.1467-9841.2012.00537.x

Tannen, D. (2007). *Talking voices: Repetition, dialogue, and imagery in conversational discourse* (2nd ed.). Cambridge, UK: Cambridge University Press.

Tan, Y. Y. (2012). Age as a factor in ethnic accent identification in Singapore. *Journal of Multilingual and Multicultural Development*, *33*(6), 569–587.

Teshigawara, M. (2003). *Voices in Japanese animation: A phonetic study of vocal stereotypes of heroes and villains in Japanese culture*. (Unpublished doctoral dissertation). Victoria, BC: University of Victoria.

Thomas, E. R. (2011). Voice quality. In *Sociophonetics: An introduction* (pp. 224-250). London, UK: Palgrave Macmillan.

Thomas, E. R., & Reaser, J. (2004). Delimiting perceptual cues used for the ethnic labeling of African American and European American voices. *Journal of Sociolinguistics*, *8*(1), 54–87.

Thurgood, E. (2004). Phonation types in Javanese. *Oceanic Linguistics*, *43*(2), 277–295.

Trudgill, P. (1974). *The social differentiation of English in Norwich*. Cambridge, UK: Cambridge University Press.

Tsai, C.-G., Wang, L.-C., Wang, S.-F., Shau, Y.-W., Hsiao, T.-Y., & Auhagen, W. (2010). Aggressiveness of the growl-like timbre: Acoustic characteristics, musical implications, and biomechanical mechanisms. *Music Perception*, *27*(3), 209–222. doi:10.1525/mp.2010.27.3.209

Van Borsel, J., Vandaele, J., & Corthals, P. (2013). Pitch and pitch variation in lesbian women. *Journal of Voice*, *27*(5), 656.e13–656.e16. doi:10.1016/j.jvoice.2013.04.008

Yuasa, I. P. (2010). Creaky voice: A new feminine voice quality for young urban-oriented upwardly mobile American women? *American Speech*, *85*(3), 315–337. doi:10.1215/00031283-010-018

Zäske, R., Fritz, C., & Schweinberger, S. R. (2013). Spatial inattention abolishes voice adaptation. *Attention, Perception and Psychophysics*, *75*(3), 603–613.

Zellers, M., & Post, B. (2012). Combining formal and functional approaches to topic structure. *Language and Speech*, *55*, 119–39.

Zhang, C., & Tan, T. (2008). Voice disguise and automatic speaker recognition. *Forensic Science International*, *175*(2), 118–122. doi:10.1016/j.forsciint.2007.05.019

Zimman, L. (2012). *Voices in transition: Testosterone, transmasculinity, and the gendered voice among female-to-male transgender people*. (Unpublished doctoral dissertation). Boulder, CO: University of Colorado.

Zimman, L. (2013). Hegemonic masculinity and the variability of gay-sounding speech: The perceived sexuality of transgender men. *Journal of Language and Sexuality*, *2*(1), 1–39. doi:10.1075/jls.2.1.01zim

Zimman, L., & Hall, K. (2009). Language, embodiment, and the "third sex." In D. Watt & C. Llamas (eds.), *Language and identities* (pp. 166–178). Edinburgh, Scotland: Edinburgh University Press.

https://doi.org/10.1017/S0267190514000270 Published online by Cambridge University Press