Original Paper

Phonetica 2012;69:193–215
DOI: 10.1159/000351059

Received: August 2, 2012
Accepted after revision: March 5, 2013

Harsh Voice Quality and Its Association with Blackness in Popular American Media

Scott Reid Moisik

University of Victoria, Victoria, B.C., Canada

Abstract

Performers use various laryngeal settings to create voices for characters and personas they portray. Although some research demonstrates the sociophonetic associations of laryngeal voice quality, few studies have documented or examined the role of harsh voice quality, particularly with vibration of the epilaryngeal structures (growling). This article qualitatively examines phonetic properties of vocal performances in a corpus of popular American media and evaluates the association of voice qualities in these performances with representations of social identity and stereotype. In several cases, contrasting laryngeal states create sociophonetic contrast, and harsh voice quality is paired with the portrayal of racial stereotypes of black people. These cases indicate exaggerated emotional states and are associated with yelling/shouting modes of expression. Overall, however, the functioning of harsh voice quality as it occurs in the data is broader and may involve aggressive posturing, comedic inversion of aggressiveness, vocal pathology, and vocal homage.

1 Introduction

The modulation of voice quality as a means to signal a shift in attitude, ownership, or orientation towards linguistic content is an everyday occurrence but remains a relatively underexplored area of sociophonetic research [Foulkes and Docherty, 2006; Podesva, 2007]. Most of the work on voice quality conducted in the past 40 years [Trudgill, 1974; Esling, 1978; Knowles, 1978; Henton and Bladon, 1985, 1988; Di Paolo and Faber, 1990; Gaudio, 1994; Stuart-Smith, 1999; Sicoli, 2010; Yuasa, 2010] adopts the traditional Labovian approach of correlating linguistic variables of authentic speech behaviour with social categories, such as socioeconomic class, gender, and so forth. Even fewer sociophonetic studies deal with harsh voice quality (henceforth, HVQ): Paddock [1977] briefly describes Newfoundlander 'roachness', a 'low, growly register' used exclusively by men; Esling [1978] reports HVQ (primarily tense/pressed voice) as a correlate of working-class speech in Edinburgh, and Rose [1989] claimed that 'growling' occurs in the Zhenhai dialect of Wu Chinese spoken in the Ningpo region. Research

KARGER

E-Mail karger@karger.com
www.karger.com/pho

0031–8388/12/0694–0193
$38.00/0

Scott Reid Moisik
Department of Linguistics
University of Victoria
PO Box 3045, Victoria BC V8W 3P4 (Canada)
E-Mail srmoisik@uvic.ca

on voice quality in the context of linguistic performance is also uncommon [e.g. Gaudio, 1994; Podesva, 2007], although speech in performance is receiving increased attention from sociolinguists [see Bell and Gibson, 2011] because it is regarded as fertile ground for examination of the subtle, and sometimes not so subtle, ways that language is used to create and disseminate sociocultural ideologies and values.

The present paper represents the intersection of sociophonetic research on voice quality, the representation of ethnicity and race in the popular media, and speech in performance. Specifically, this study examines how some performers (mainly comedians and actors) use HVQ (viz. tense/pressed phonation often featuring growling and/or concomitant raised larynx voice) in the portrayal of black characters, personas, and stereotypes[1]. In these cases, HVQ is associated with exaggerated emotion and yelling/shouting and may index aggressive or transgressive attributes [Henry, 2002; Tsai et al., 2010]. However, several peripheral cases show that HVQ is not necessarily associated with these properties, but rather has broader associations and applications.

1.1 Voice Quality, Harshness and Growling

The central phonetic focus of this work is the use of HVQ. I adopt Abercrombie's [1967, p. 91] view of voice quality as 'a quasi-permanent quality running through all the sound that issues from [the speaker's] mouth'. Laver [1980, p. 2] observes that voice quality setting fulfils an important semiotic role in spoken communication by indexing physical, psychological, and social characteristics. For example, it may be possible to deduce from an individual's voice quality their physical size, the health of their vocal folds, whether they are angry, whether they originate from a particular speech community, and so forth.

Laver [1980, pp. 126–132, 141–156; also see Scherer, 1986, p. 152] defines HVQ by high variation in the duration (jitter) and intensity (shimmer) of the pulse cycle associated with the voice source and says it may be characterized by acoustic noise and a general increase in overall intensity. He notes that HVQ physiologically correlates with increased tension in the laryngeal and pharyngeal parts of the vocal tract, increased medial compression of the vocal folds, and possible damping of vocal fold motion caused by the ventricular folds. This description corresponds approximately with Catford's [1977, pp. 102–103] *anterior voice*, which implies that only the ligamentous (anterior) vocal folds oscillate due to high degrees of adductory force on the arytenoid cartilages and 'general sphincteric constriction of the (upper) larynx'. In describing the glottal level of this voice quality, the term *tense* [Laver, 1980, pp. 144–145] has been used to denote the increased muscle activity required to strongly adduct the vocal folds. Stevens [1998, pp. 82–85] uses the term *pressed* in the sense that the vocal folds are pressed together medially due to increased adductory force. Both Laver and Stevens note that tense/pressed voice increases acoustic energy in the upper harmonics [also see Zemlin, 1998, pp. 173–174].

[1] In this article, I use the terms *black* and *white* to refer to the phenotypic identity of performers, characters, and associated stereotypes. For the majority of sources considered, performers use these labels in their material. Based on consultations made on the issue, applying terms such as *African-American* or *European-American* makes an assumption of ethnic identity which may not agree with that of the performer [Rebecca Childs, pers. commun.]. For similar reasons, *Black Vernacular English* (*BVE*), rather than *African American Vernacular English*, is used in discussing the sociolinguistic variety associated with black communities in the United States.

Gerratt and Kreiman [2001] identify the following acoustic properties for supraperiodic non-modal phonation that characterize HVQ: (a) period-doubling, (b) amplitude modulation, and (c) low harmonics-to-noise ratio. Properties (a) and (b) imply quasi-periodic damping of the glottal source causing subharmonic structure in the source spectrum. They argue, based on evidence from their perceptual study, that HVQ and creakiness form a subcategory of non-modal phonation with distinct perceptual boundaries separating it from modal phonation. They observe that HVQ and creakiness are acoustically and perceptually distinct from each other, such that creakiness is distinguished from HVQ by 'extremely low' fundamental frequency (F0) and increased between-pulse vocal tract damping [Gerratt and Kreiman, 2001, pp. 375–376]. The other class of non-modal phonation is breathiness, which lacks distinct perceptual boundaries separating it from modal phonation. They suggest that the ventricular folds may play a role in the production of HVQ and creakiness.

Esling and Harris [2005] offer an elaborated interpretation of the articulatory mechanism responsible for HVQ. In their model, phonation types are divided into two major categories, constricted and unconstricted, which respectively denote the physiological engagement or disengagement of the laryngeal constrictor mechanism. The laryngeal constrictor operates to seal the larynx for the biological purpose of swallowing and other life-supporting functions [Fink, 1974]; it also plays a role in the production of many types of linguistic contrasts [Edmondson and Esling, 2006]. Laryngeal constriction functions to narrow and ultimately to close the epilaryngeal tube by means of contraction of the thyroarytenoid, thyroepiglottic, and aryepiglottic (AE) muscles, retraction of the tongue and, concomitantly, the epiglottis, and raising of the larynx [Esling et al., 2007]. Constricting the larynx in this manner causes the soft internal structures of the larynx to fold and buckle [or 'plicate'; Fink, 1974]. A key laryngeal state that arises during this compaction is ventricular incursion: the ventricular folds become pressed into the vocal folds, thereby changing the oscillatory behaviour of the vocal folds (by increasing damping, adding effective mass, interfering with the mucosal wave, and so forth), resulting in patterns of vibration associated with creaky or harsh voice [see Lindqvist-Gauffin, 1972; Laver, 1980, pp. 130, 144; Edmondson and Esling, 2006]. Since a raised larynx setting facilitates laryngeal constriction, constricted voice qualities such as HVQ may involve concomitant raised larynx voice quality, which is characterized by a rise in the frequency of vowel formants, as a consequence of the shortened vocal tract, and elevated F0 [Sundberg and Nordström, 1976; Laver, 1980, pp. 24–28; cf. Nolan, 1983, pp. 182–187].

Provided sufficient airflow is present, laryngeal constriction can result in vibration of the epilarynx (I use the impressionistic label *growling* here to remain neutral regarding what exactly is vibrating, since it is impossible to know for certain without visual evidence). Traill [1986] claimed that the so-called sphincteric phonation in !Xóõ is produced by vibration of the epilaryngeal structures; similarly, the 'raucous' and 'deep' pharyngeals of Agul [Catford, 1983, p. 347] probably involve epilaryngeal vibration, as Esling [1999, p. 364] suggests. Miller [2007] and Miller et al. [2009] posit an AE phonatory mechanism for distinctive 'epiglottalized' vowels in Ju|'hoansi and N|uu. The vocal style of the Jing role in Chinese Opera ostensibly epilaryngeal vibration, which would account for its growled quality [Tsai et al., 2010]. Some visual evidence suggests that growling involves vibration of the AE folds [Sakakibara et al., 2004; Moisik et al., 2010].

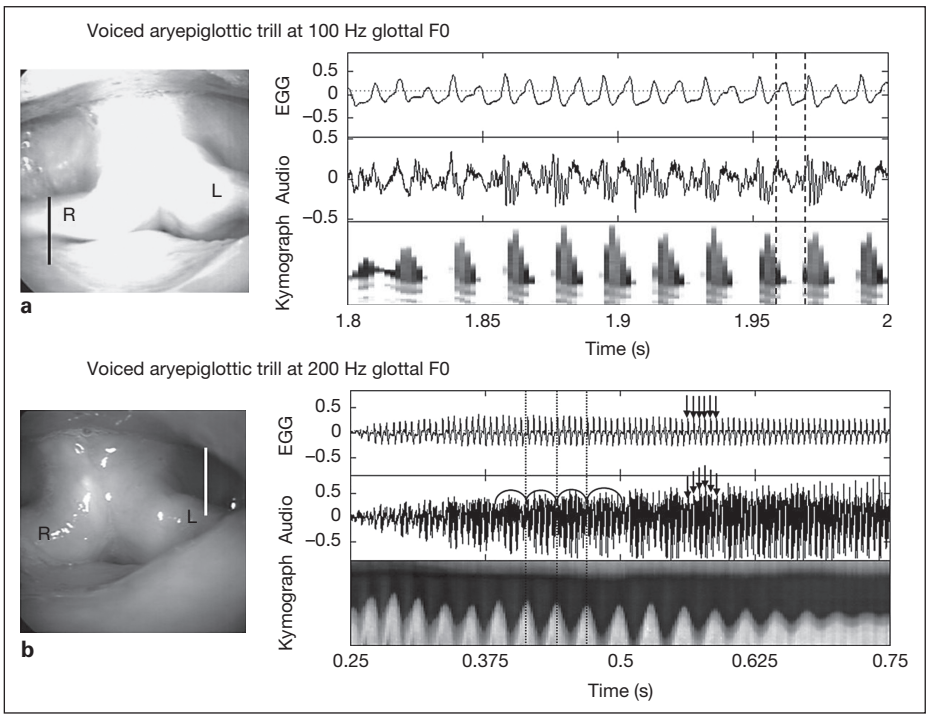

Fig. 1. High-speed laryngoscopy data (500 fps) of voiced AE trilling (growling) at 100 Hz (**a**) and 200 Hz (**b**) glottal F0 [data prepared uniquely for this article based on Moisik et al., 2010]. Laryngoscopic still frame images are on the left (R = right AE fold; L = left AE fold); solid lines in laryngoscopic frames indicate location of kymographic pixel strip. Dashed lines in **a** illustrate correspondence among alternating types of glottal closure, the period-doubled acoustic output, and the phases of the AE aperture (closing and opening). Arrows in **b** identify some glottal pulses in the EGG and acoustic signals; arcs iconically illustrate the amplitude modulation of the glottal pulse occurring throughout; dotted lines mark left AE aperture opening. Note that the kymograph in **b** only shows AE fold displacement (the AE aperture was not visible); the aperture is open when the fold displacement forms a peak in the kymographic image.

High-speed laryngoscopic video data[2] in figure 1 illustrate AE trills at glottal F0s of (a) 100 Hz and (b) 200 Hz, respectively. Each figure contains three time series plots: the electroglottograph (EGG) signal (top), the acoustic waveform (middle), and a kymographic image of AE fold displacement (bottom). These figures illustrate the correlates of AE vibration: the arcs indicate the imprint of the AE pulse event in the waveform and identify the amplitude modulation of the glottal pulses that results from the AE pulse (note: only a few pulses are indicated with arcs, but the effect extends throughout the figure).

Since the vocal folds and the AE folds share the same airflow, these structures tend to entrain each other. Entrainment means that the oscillatory behaviour of both sets of folds will tend to synchronize, even though the biomechanical properties governing

[2] This data was obtained by directly following the methodology in Moisik et al. [2010].

both systems (i.e. those of the AE and vocal folds) differ. Entrainment establishes an aryepiglottal-to-glottal (AE:G) pulse ratio, which varies depending on the specific configuration of the larynx (e.g. the amount of longitudinal tension on the vocal folds). Acoustically, entrainment results in amplitude modulation of the voice source, with period-doubled phonation occurring if the AE:G pulse ratio is approximately 1:2 (fig. 1a). Smaller pulse ratios yield variations of this (e.g. in fig. 1b, the AE:G pulse ratio is 1:5). One can therefore generalize that AE vibration causes amplitude modulation of the glottal source; thus, in the frequency domain, growling is associated with the presence of subharmonic structure stemming from the low-frequency AE periodicity relative to the periodicity of the glottal pulse.

While the oscillations that occur during growling tend to exhibit entrainment, the anatomical structure and configuration of the AE folds exhibit a strong tendency towards oscillatory irregularity [Moisik et al., 2010][3]. Furthermore, the AE folds form a narrow stricture above the vocal folds (sometimes furcated into multiple channels), which, combined with the high volume velocity used to drive their vibration, can produce turbulent flow. Consequently, AE vibration tends to have a noisy acoustic signature and instability in the frequency of the amplitude modulation of the voice source, often resulting in poorly defined harmonic structure.

The ventricular folds can also vibrate, producing the growl associated with the throat singing traditions of Tibetan, Tuvan, and Mongolian cultures [Catford, 1977, p. 102; Fuks, 1998; Levin and Edgerton, 1999; Lindestad et al., 2001; Sakakibara et al., 2004; Bailly et al., 2010]. Similar vibrations occur in heavy rock music [Borch et al., 2004]. However, compared with the vibration of the AE folds, which vibrate with more irregularity, ventricular phonation seems to be more regular, as suggested by Edmondson et al. [2001], possibly because it involves stronger glottal entrainment.

Without visual evidence (such as laryngoscopic imaging), one can only infer the specific physiological mechanism responsible for a given voice quality. Nonetheless, it remains productive to distinguish auditorily between the different varieties of HVQ, viz. those that involve purely vocal fold vibration (possibly with ventricular incursion), those with simultaneous vocal and ventricular fold vibration, those with simultaneous vocal and AE fold vibration, and those which exhibit a combination of these and possibly vibrations from other structures, such as the epiglottis, the arytenoids themselves, and so forth. For this work, the principle of auditory classification (with supporting acoustic evidence) is applied in distinguishing the various voice qualities found in the data. (The term *epilaryngeal* is used to discuss the physiological mechanism of growling in the data so as to remain somewhat neutral as to what exactly is causing it.)

1.2 Sociolinguistic Assessment of Performance

This study examined the voice quality of various performers in performance context. A wealth of recent literature focuses on speech in performance context, ranging from spontaneous conversation to staged contexts (e.g. television shows). Taken together, this work addresses sociolinguistic variables that convey regional and ethnic identity [e.g. Rampton, 1999; Coupland, 2001; Schilling-Estes, 2004; Remlinger, 2009; Benor, 2010; Bell and Gibson, 2011].

Speech behaviour in performance can function to create personas, characters, and parodies. Thus, performance does not necessarily represent authentic speech so

[3] For example, some variation in the AE pulse is evident in figure 1a.

much as it represents stylized speech, which often employs sociolinguistic stereotypes for various effects. Sociolinguistic stereotyping can be thought of as a process of enregisterment, in which distinct, often salient forms of speech become associated with ideologies of a particular social group [Remlinger, 2009, p. 119; Benor, 2010, p. 161]. For example, the process of enregisterment has produced the stereotypical notion that speakers of RP/Standard British English are socially refined [Agha, 2003, p. 254]. Like stereotyping in general, sociolinguistic stereotyping is not strictly bound to objective facts about social groups [Labov, 1978, p. 314].

Sociolinguistic stereotypes are among the many sociolinguistic resources that comedians, actors, musicians, amateur performers, and ordinary speakers use when performing. The performer's goal may be an accurate portrayal of a particular sociolinguistic identity, but such performances, especially if done ineptly, can propagate ethnic and racial stereotypes [Coupland, 2001; Chun, 2004; Pao, 2004; Meek, 2006]. Conversely, the comedian or actor may choose to emphasize stereotypes (in general) to reflect, repossess, subvert, or challenge the ideologies these express [Zolten, 1993, p. 65].

Voice quality serves as one of the sociolinguistic resources that performers can manipulate. Podesva [2007, p. 497] observes that voice quality is a tool for moderating social orientation and group affiliation. Speakers may creatively change voice quality according to pragmatic context [Sicoli, 2010] or use it to project a persona associated with certain social values [Yuasa, 2010]. Coupland [2001, p. 350] observes that speakers emphasize stylized utterances to contrast or to underscore incompatible or dichotomized sociosemiotic content. Correspondingly, we might expect performers to use exaggerated voice quality to enhance the intended effect of a portrayal of social identity or to enhance a contrast in the portrayal of two or more social identities in a given performance. HVQ, especially with growling, is arguably an excellent way to increase the impact of a stylized utterance, particularly when this voice quality is contrasted with unconstricted laryngeal qualities such as modal or breathy voice.

On a final and important note, sociolinguistic stereotyping and parody can be the tools of prejudice and discrimination. In this context, the analysis of stereotypes can unintentionally propagate the destructive and hurtful messages embodied in such productions. Thus, the researcher must necessarily proceed with caution and awareness of possible interpretations of such analyses. Ronkin and Karn [1999] contend, however, that the identification and study of linguistic parody or mocking serves to increase or enhance awareness of how racist or prejudicial ideologies are operationalized. Discussing such data in an academic context can serve to raise awareness of the vehicles used to express these ideologies and to invite discussion of racism and prejudice in a public forum for scrutiny and understanding.

2 Methodology

The present research involves a qualitative auditory survey of HVQ in (mainly) popular media from the United States dealing with blackness and several peripheral cases, which do not (overtly or necessarily) pertain to blackness (Part I), and a series of illustrative case studies conducted with qualitative auditory and acoustic analyses (Part II). Part I was conducted to locate relevant examples; Part II demonstrates the relationship between speech stylization using HVQ and the portrayal of blackness in the media.

The data are from YouTube (www.youtube.com), a popular online repository of video content. This website serves as a resource for material that would be difficult and time-consuming to assemble through other means. However, YouTube video data are subjected to digital compression techniques,

which can alter the frequency content of the associated audio signal. A further problem is that YouTube contributors use digitizing and recording equipment of variable quality, creating potential for further audio signal degradation. Because of the potential for distortion, poor quality, and the general variability of audio in YouTube data sources, the examination and comparisons of voice quality were made only within, as opposed to across, data sources, and no attempt was made to quantify features of the voice source. It was assumed that signal quality was uniform within a single data source.

2.1 Data Collection

The search for data began with the author's familiarity with a small set of cases of HVQ in connection with the representation of blackness; the goal for data acquisition was simply to find additional examples of this pattern. Thus, the data were collected using a non-random sampling approach. Since the data under consideration are sparsely distributed in a very large data pool, random sampling would have been insufficient to locate a reasonable quantity of relevant data.

Liberal definitions for *performer* and *performance* were used: professional and amateur mediatized data [Bell and Gibson, 2011] and spoken and sung data were accepted into the corpus. Also, no restriction was placed on the time period of a performance.

The portrayal of blackness was defined using the following criteria: a performance by an individual who is phenotypically black or represented as black (in the case of cartoons), or a performance that invokes blackness prominently in the content of a performer's speech or character. In several cases, performers explicitly state that the performance is about race (for example, comedian Steve Harvey's 'White and Black People Being Fired' routine). In many other cases, however, performances contain only indirect indications of race, either by virtue of the performer's phenotypic identity, or by a performer's associations with racialized content. For example, comedian Chris Rock explicitly racializes his work through the labeling of the content he produces, such as his comedy album *Bigger & Blacker* [Rock, 1999].

Searching was not limited to phenotypically black performers. In an attempt to balance the data pool, phenotypically non-black (mainly white) performers were also included in the search. Furthermore, a number of examples involving puppets were also admitted to the data set. Many of these peripheral cases provide fascinating illustrations of the diversity of applications for laryngeal qualities in performance; they are also valuable as a point of comparison to the data drawn from black performers and as a way to provide a more complete understanding of the role of HVQ in performance.

After identifying a potential source of data, YouTube was searched for examples of the source. The overwhelming majority of the data comes from popular American media, which, in this context, refers to music, radio and TV (both live action and cartoons), and stand-up comedy acts. Altogether, 34 data sources were identified. In many cases, additional data sources and peripheral cases were located in the search, especially for musicians; much of these data were omitted to constrain the scope of the analysis. Audio was extracted from the video for the purpose of the analysis.

2.2 Method for Part I: Qualitative Analysis of the Data Survey

The following details are noted for each data source: (a) the dates of birth and death (if appropriate) of the individual or date of broadcast (in the case of radio, TV, and movie characters), (b) the birth name of the individual, (c) the stage or character name (if any), (d) indications of how (laryngeal) voice quality is phonetically realized (the general glottal-F0 range and the presence of tense/pressed phonation, vocal-AE or vocal-ventricular phonation, or raised larynx voice; see table 1 for more details), (e) an impression of the frequency of occurrence of HVQ (occasional, frequent, or continuous), (f) whether the voice quality is likely to be pathological or not, and (g) any additional phonetic observations of interest. All phonetic observations in Part I are auditory impressions of the author and an additional trained phonetician. Links to specific YouTube videos are not provided because these resources are impermanent and cannot be guaranteed to be available. In most cases, to find examples on YouTube it should be sufficient to simply search for the performer or performance using the information (such as the performer's name) provided in the summary tables.

2.3 Method for Part II: Qualitative Acoustic and Auditory Analysis of Case Studies

For each case study, a multiline transcription of the data is provided that includes the following information: text transcript, IPA segmental transcription, intonation-prosody and voice quality transcription, and manually evaluated, syllable-by-syllable indication of glottal F0 and, where appropriate,

Table 1. Voice quality symbols based on the VoQS system used in the analysis

Symbol	Name	Constricted	Description
V	modal voice	N	- regular glottal cycle with complete closure - low variation in pulse period and intensity
V̰	creaky voice	Y	- irregular vocal fold vibration possibly with ventricular incursion - low intensity and low F0
V̤	tense or pressed harsh voice	Y	- irregular and intense vocal fold vibration possibly with ventricular incursion - high jitter (period variation) and high shimmer (amplitude variation) in glottal pulse - no epilaryngeal vibration
V!	growled harsh voice	Y	- epilaryngeal vibration (most likely AE) - glottal-pulse amplitude modulation and low harmonics-to-noise ratio
V̰!	whispery growled harsh voice	Y	- variant of growled harsh voice with vocal fold abduction - increased airflow yields increased noise - vocal fold configuration similar to that in breathy voice
V!!	ventricular voice	Y	- simultaneous oscillation of the vocal and ventricular folds (the physiological basis of throat singing
L̝	raised larynx voice	Y	- shortened vocal tract due to larynx raising - upward shift in formant frequencies - some pharyngeal constriction possible

All of these terms imply the presence of vocal fold vibration (hence *voice*); *constricted* refers to general epilaryngeal constriction (see section 1.1). Y = Yes; N = no.

epilaryngeal F0. Select syllables are illustrated using 100 ms of the audio waveform retrieved from the mid-point of the vowel and a Hamming-windowed narrow-band spectrum of this signal. The analysis focuses on broad patterns such as amplitude modulation of the voice source in the time domain or the presence or absence of subharmonic or noisy harmonic structure in the frequency domain. To mitigate the possibility that identified features of the audio are simply due to digital compression or poor quality recording, the harsh examples were compared with non-harsh qualities obtained from the same audio signal (although this is only demonstrated here in case studies 3.1.2 and 3.1.3).

All acoustic analyses were conducted using Praat [Boersma and Weenink, 2010]. In the time domain (waveform), the presence or absence of amplitude modulation is taken as the primary indicator of HVQ with growling (see section 1.1). Pressed and tense phonation are not demonstrated acoustically because these have features (jitter and shimmer) that must be compared to an established baseline for a speaker, something which cannot be obtained given the limited nature of the data.

Auditory analyses performed by the author and an additional trained phonetician identify stretches of the speech signal using a slightly modified version of the Voice Quality Symbols (VoQS) system [Ball et al., 1995]. The reason for this modification is as follows: in the VoQS system, harsh voice is represented as {V!}, regardless of whether there is epilaryngeal vibration (growling) or not. HVQ is a general voice quality category: it implies voice source aperiodicity and generally greater intensity than creaky phonation [Laver, 1980]. Thus, the term applies both to vocal fold vibration alone (tense/pressed voice) and to vocal fold vibration with concomitant epilaryngeal vibration. In this study, two symbols for HVQ were used to signify whether growling occurs or not: {V!} signifies HVQ with growling and {V̰} signifies HVQ without growling. The VoQS system includes {V!!} for transcribing simultaneous vocal and ventricular fold vibration (called ventricular phonation). Finally, {L̝} denotes raised larynx voice quality. Table 1 provides a summary of the symbols used in the present study.

Table 2. Phenotypically black performers and characters who use HVQ

Dates	Name/title/stage name	Occupation/description	G-F0	V̰	V!	L̰	V!!	Notes	P?
1894–1975	Jackie 'Moms' Mabley	comedian	L-M	✓	✓		✓	C	Y
1901–1971	Louis Armstrong 'Satchmo'	jazz musician	L-M		✓	✓		F	Y
1904–1981	Dewey 'Pigmeat' Markham	comedian, actor	L-M		✓	✓		F	N
1910–1976	Chester Burnett 'Howlin Wolf'	American blues musician	L	✓	✓		✓	C	Y
1910–1986	Benjamin Crothers 'Scatman'	actor, dancer, singer	L-H	✓	✓			O	N
1913–1983	McKinley Morganfield 'Muddy Waters'	American blues musician			✓			O	Y
1922–1991	John Sanford 'Redd Foxx'	comedian, actor	L	✓	✓		✓	C	Y
1928–1943	**Amos** from *Amos & Andy* (Freeman Gosden)	radio character	L-M	✓	✓	✓		C	?
1928–2009	Koko Taylor	American blues singer	L-H		✓	✓		C	N
1932–	Little Richard	rock 'n' roll Musician	L-H	✓	✓			O	N
1941	'Bugle Boy of Company "B"'	Walter Lantz cartoon	L-M	✓	✓		✓	O V̰ V̰	-
1941	'Scrub Me Mama with a Boogie Beat'	Walter Lantz cartoon	L-M	✓	✓	✓		O	-
1952–	Laurence Tureaud 'Mr. T'	wrestler, actor	L-M	✓	✓			F	N
1957–	Broderick 'Steve' Harvey	comedian, actor	L-H	✓	✓			F	N
1965–	Martin Lawrence	comedian	L-H	✓	✓			O	N
1965–	Chris Rock	comedian	L-H	✓	✓			F	N
1971–	Johnathan Smith 'Lil Jon'	rap musician	L-M	✓	✓			C	N
1972–	Trevor Smith Jr. 'Busta Rhymes'	rap musician	L-M	✓	✓			O	N
1972–1985	Fat Albert (Bill Cosby)	cartoon character	L		✓	✓		C V̰!	N
1972–	Chris Tucker	comedian, actor	M-H	✓				C	N
1973–	Dave Chappelle	comedian		✓	✓	✓		O	N
2008	**Kirk Lazarus** (Robert Downey Jr.), Tropic Thunder [Stiller, 2008]	movie character	L-M	✓	✓	✓		O	N

G-F0 = Glottal F0 range where HVQ occurs (tendency); L, M, H = low, mid, high; C, F, O = constant, frequent, occasional; P? = pathological?; Y, N = yes, no. Names in bold indicate black characters that are played/voiced by white performers.

3 Phonetic Observations and Illustrative Case Studies

Table 2 contains a list of phenotypically black performers and characters; table 3 lists peripheral cases involving non-phenotypically black performers and characters. The range of phonetic detail in the implementation of HVQ varies along several basic phonetic parameters: (a) glottal F0; (b) glottal configuration (abduction or adduction);

Table 3. Phenotypically non-black performers and characters who use HVQ

Dates	Name/title/stage name	Occupation/ description	G-F0	V̤	V!	L̰	V!!	Notes	P?
1933–1935	Pop-Eye the Sailor Man (William Costello)	cartoon character	L-H				✓	C	?
1937–2008	George Carlin	comedian	L-M	✓	✓			O	Y
1941–2010	Don Van Vliet 'Captain Beefheart'	psychedelic musician	L-H	✓	✓		✓	C	N
1949–	Tom Waits	experimental musician	L-H	✓	✓		✓	F	Y
1951–	Robin Williams	comedian	L-H	✓	✓	✓		O	N
1952–	Randall Poffo 'Macho Man'	professional wrestler	L-H	✓	✓			F	N
1952–	Paul Reubens 'Pee Wee Herman'	actor	L-H		✓	✓		F	N
1953–	Terry Bollea 'Hulk Hogan'	professional wrestler	L-M	✓	✓			F	N
1969–2002	Cookie Monster (Frank Oz)	TV puppet	L-H		✓	✓		C	N
1969–	George Fisher 'Corpsegrinder'	death metal singer	L-H	✓	✓			C $V^{œ}$	N
1980–	Yoda (Frank Oz)	movie puppet	L-H	✓	✓	✓		C	N
1983–1986	Dr. Claw from *Inspector Gadget* (Frank Welker)	cartoon character	L		✓			C V̤!	N

G-F0 = Glottal F0 range where HVQ occurs (tendency); L, M, H = low, mid, high; C, F, O = constant, frequent, occasional; P? = pathological?; Y, N = yes, no; $V^{œ}$ = labialized voice (open round); V̤ = breathy voice; V̰ = whispery voice.

(c) the absence or presence of epilaryngeal vibration; (d) in growling, the (suspected) specific epilaryngeal structure that vibrates; (e) the extent of larynx raising; (f) the prosodic context (e.g. stressed or emphatically stressed vs. unstressed syllables), and (g) the vocalic context (close vs. open vowels).

Glottal F0 (a) does not strongly restrict HVQ (with or without growling), and there are numerous examples of growling at both ends of this scale. Amplitude modulation ratios (presumably AE:G) from 1:2 to 1:6 can be identified in the data (see section 3.1), but there is variability in the relationship from one syllable to the next because of the irregular nature of the epilaryngeal pulse (which suggests AE vibration). These data show that growling can indeed occur across a wide range of glottal F0 levels.

According to Laver [1980, p. 13], glottal configuration (b) in HVQ tends towards adduction. Examples of an abducted glottal state are the voices for Frank Welker's Doctor Claw and Bill Cosby's Fat Albert. Both of these voices constitute whispery growled HVQ, but they vary in the degree of whisperiness and occasionally exhibit purely voiceless growling (i.e. epilaryngeal vibration without any concomitant vocal fold vibration).

HVQ does not always involve growling (c); many of the voices observed are cases of pressed or tense phonation, probably involving compression of the ventricular folds against the vocal folds (i.e. ventricular incursion) in addition to the expected increase in adductory force and medial compression of the vocal folds (factors which should cause perturbed vocal fold dynamics and corresponding acoustic effects). For voices that do employ growling, the question of which physiological mechanism causes the growl (d) remains open (since there is no visual evidence for these data). However, we may reasonable infer that most examples of growling in the data involve AE oscillation because of their auditory correspondence with careful phonetic productions observed laryngoscopically (illustrated in section 1.1). The acoustic evidence helps to substantiate this auditory impression because all the examples auditorily judged to involve AE vibration show irregular amplitude modulation, which is more indicative of AE than ventricular fold vibration [as suggested in Edmondson et al., 2001]. Some examples in the data may feature ventricular phonation (e.g. the voice of Popeye the Sailor Man), but these were in the minority according to the auditory evaluation.

Based on the case study measurements, epilaryngeal F0 values were observed to range between 60 and 160 Hz, with a mean of 106 Hz (30 Hz SD) and concomitant glottal F0 values averaging at 330 Hz (125 Hz SD). These values for epilaryngeal F0 are generally higher than those in Moisik et al. [2010], where AE fold frequencies between 40 and 100 Hz are reported, with higher values occurring in the increased tension contexts. However, they did not evaluate cases with glottal F0s above 200 Hz. The higher values of epilaryngeal F0 observed here suggest elevated levels of laryngeal tension in connection with the higher average glottal F0 values.

Voice quality effects associated with larynx raising (e) were salient in several voices, especially the cartoon and puppet characters (such as those of Jim Henson and Frank Oz's Muppets and the voice of Paul Reubens' Pee Wee Herman).

Concerning the prosodic parameter (f), the prosodic context, it is evident in the data that HVQ, and growling in particular, tends to correlate with stressed syllables, syllables receiving focal or tonic stress, or syllables occurring in emphatic contexts (such as yelling). However, this tendency is not absolute: growling and HVQ can be heard in unstressed or non-emphasized contexts as well, or they can be continuously employed throughout a performance [also see Sakakibara et al., 2004].

Finally, with regard to (g), the vocalic distribution of HVQ, growling and HVQ tend to occur with vowels that are relatively open, such as [ɛ ɑ ɔ]. This observation is consistent with those made in the literature [Laver, 1980, p. 128; Rose, 1989]. Following the articulatory model put forth by Esling [2005] the 'retracted' vowels (i.e. [ɑ ɒ ʌ ɔ]), all relatively open, are predicted to bias the engagement of laryngeal constriction more so than more close vowels. The data seem to corroborate this prediction, but the flexibility and relative independence of the laryngeal constrictor mechanism from the lingual articulator mechanism allow HVQ and growling to occur with any vocalic setting, as illustrated in the case studies below.

3.1 Case Studies

Three case studies are examined which illustrate the acoustic properties and distribution of HVQ for the performance-mode speech associated with the representation of blackness and its stereotypes. In each case, a transcription of a selected performance segment is provided alongside qualitative acoustic analyses in the time and frequency domains. These cases will be discussed further in section 4.

3.1.1 Chris Rock on Rap Music

The segment in example 1 is part of a routine by Chris Rock (a black American comedian and actor) on the subject of rap music from *Chris Rock: Never Scared* [Gallen, 2004] and illustrates the typical voice he uses for his stage persona. Growling is most commonly heard with emphatic syllables, but longer stretches do occur. Also, growling is not restricted by glottal F0 and frequently occurs at the upper extent of the glottal F0 range (~300 to 400 Hz). Raised larynx voice quality is not heavily employed, but it does occasionally occur. The example below illustrates these observations: in line 1, Chris Rock is speaking in his stage persona voice, which is produced with pressed phonation throughout and growling over emphatic (heavily stressed) syllables. The beginning of line 2 is modal; it marks a transition into the imaginary quoted speech found in the remainder of this sample (lines 3–5), in which features of HVQ return. Thus, the use of HVQ does not strictly correlate with moments of (imagined) quoted speech.

Often Chris Rock produces phonatory register jumps (sudden switching from one mode of phonation to another) during a single syllable: usually the first half of the syllable will be produced with pressed phonation and then growling will be engaged during the second half, generally corresponding to a 20- to 200-Hz drop in glottal F0; examples are in lines 3, 4, and 5.

Example 1. Transcription of Chris Rock's monologue on rap music from *Chris Rock: Never Scared* [Gallen, 2004]; from top to bottom: text, IPA, voice quality and prosody/intonation, glottal F0 (GL), and epilaryngeal F0 (EPL). YouTube Search: 'Chris Rock on rap music'.

Now, I'm thirty-nine right? And I still love rap music. I LOVE rap music (applause).
You know? I LOVE IT! But I'm tired of defending it.

1	Cause	You	gotta	de	fend	rap	music	
[	kʰʌdʒ	ju	ɣʌ ɾə	ɾəʔ	fɛ̃n	ɹæpʔ	mju zɨk	]
{**V!**	M	H	M		**H**	M	L	}
GL	255	345	370 250	215	310	200	170 100	
EPL			80		95	100		

2	man	cause	people	always	go		'That's	not		music!		
[	mæ̃:ŋ	kʰə	pi pɻ	ɔz	go		ðæts	nɑtʔ		mju	zɨk	]
{**V**	L		**H**	M	L	}	{ **Ṿ** M	L	}{**Ṿ!**	M	H	}
GL	160	170	315 260	250	160		GL 255	155		220	264	
EPL										100		

3	That's	not		**art!**		That's	garbage!		
[	ðæs	nɑtʔ		ʔɑ:t		ðæts	gɑ:	bɪdʒ	]
{ **Ṿ̩**	M	L	}{**V!Ḷ**	M-H	}{**V!**		H-L		}
GL	255	160		195-220		190	440-290	175	
EPL				60-Ø			Ø-140	75	

4	How	can	you	listen	to		that	garbage?!		
[	haw	kʰn̩	ju	lɪ sn̩	tʰu		ðæt͡kˇ	gɑ:	bɪdʒ	]
{ **Ṿ**	H			M		}{ **V!**		H-L		}
GL	400	305	265	340 220	275		215	390-250	200	
EPL							100	Ø-140		

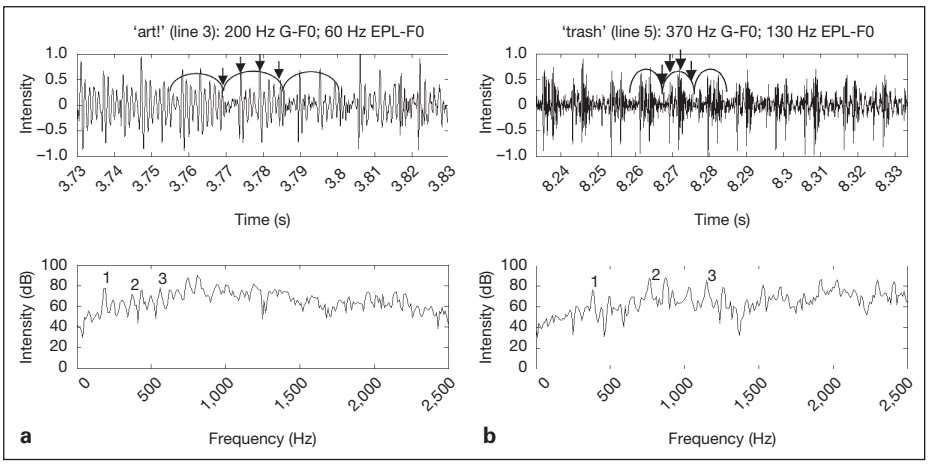

Fig. 2. Waveforms and spectra of black American comedian Chris Rock's on-stage persona. In the waveforms, arcs illustrate some of the suspected individual epilaryngeal pulses associated with amplitude modulation, and arrows mark some of the individual glottal pulses; in the spectra, the first three glottal harmonics are numbered.

5	How	can	you	listen		to		that	**trash?!**"
[	haw	kɜ̃	ju	lɪ	sn̩	t^h		æt^ʔ	ʧ:ɹæ:::ʃ::]
{ **V!**	H			M			} { **V!**	M	M-H }
GL	330	300	260	340	235			280-255	370
EPL								Ø-135	130

In figure 2, two acoustic illustrations of growled syllables from example 1 are provided. Arcs are used to indicate some of the individual epilaryngeal pulses associated with the growling quality that serves to modulate the amplitude of the glottal source (arrows mark some of the individual glottal pulses). The spectra show the locations of the first three glottal harmonics in relationship to additional spectral content in the form of subharmonics, which were probably generated by the epilaryngeal amplitude modulation. On account of aperiodicity in the epilaryngeal pulse, the subharmonics do not form narrow spectral peaks; rather, they are distributed across a broader spectral range. Both spectra also exhibit high levels of interharmonic noise, which probably corresponds with turbulence generated by the narrow stricture formed by the epilaryngeal tube.

3.1.2 Steve Harvey's 'White and Black People Being Fired'

In this example, black American comedian Steve Harvey juxtaposes the reactions that a white male employee character ('Bob') and a black male employee character ('Willie') have to being fired. The racial identity of these characters is overtly stated. For example, Steve Harvey introduces the routine by saying: 'Black people handle getting fired different from white folk'. The contrast is partly based on Harvey's 'putting on a voice' [see Bakhtin, 1986; Coupland, 2001, p. 346] to manage character identities in the performance, and, to this end, voice quality plays an important role in complementing other linguistic and textual factors. For the white

employee, Harvey primarily uses modal phonation at a relatively high F0 with occasional breathiness and vocal tremor (both in frequency and amplitude), all possible indices of fear or nervousness [Scherer, 1986]; the effect suggests a character who is dismayed at the news of being fired. In stark contrast, the voice used for the black employee has an exaggerated glottal F0 range and vocal intensity and HVQ with growling on emphatic syllables at the high and low ends of the glottal F0 range: the voice used could be interpreted as yelling/shouting. The characterization is of someone who anticipates being fired (which is explicitly stated by Steve Harvey) and who, upon hearing that he is indeed fired, is outraged and expresses this reaction in vocal paroxysms and profanity. Apart from the voice quality and lexical choices, the two characters are also distinguished by segmental variation, as the IPA transcription shows. For the black employee, Steve Harvey uses phonological features associated with BVE, such as post-vocalic r-lessness, monophthongization of /aw/ to [a:], fortition of the interdental fricative, and so forth, all of which are absent in the white character's speech.

Example 2. Transcription of Steve Harvey's monologue on the firing of white and black employees from *One Man* [Vinson, 1997]; from top to bottom: text, IPA, voice quality and prosody/intonation, glottal F0 (GL), and epilaryngeal F0 (EPL). YouTube search terms: 'Steve Harvey white and black people fired'.

in the voice of Tom, the white employer
'Listen to me. Bob you're making this so difficult.
I know you're going to have a tough time explaining this to Becky (audience laughs), but we're gonna have to let you go.'
in the voice of Bob, the white employee
'Oh Jesus! Oh Tom, what am I going to do?'

1	What	about		the	mortgage?			What	**about**		the	children's		college		fund?							
[	wʌt	ə	bawt	ðə	mɔːɹ	gɪʤ̚		wʌt	ə	bawt	ðə	ʧɪɬ	dɹĩnz̚	kʰaɬ	ɪʤ	fɅ̃nd	]						
{**V**	H		M		**M**			H		M		**H**		**M**		**M**	}						
GL	400	375	330	265	300	215		405	385	350	270	360	250	275	195	230							

2	Oh		father					God!				(audience laughs)												
[	oːːw		fɑ	ðɻ				gɑd		]														
{**V**	**H**		M			}	{**V̲**	M		}														
GL	395		280	230				330	260															

(Lines 1 and 2 produced with F0 and amplitude tremor.)

(... skit continues to the firing of Willie, the black employee)

in the voice of Tom, the white employer
'Jesus! Willie! I know you're going to have a tough time explaining this to Willamina...'
in the voice of Willie, the black employee

3	Oh	this	ain't	**bout**			no	Willa-			mi-			na,										
[	ʔɔ:::	dɪs	ẽ	baːt			now	wɪ	ɬə		mi			nəʔ	]									
{**V!**	H	H	M	**M**		}	{**V**	M	M		H		}	{**V!**		}								
GL	480	440	360	420				305	280	230	340			140										
EPL	150		?	160										75										

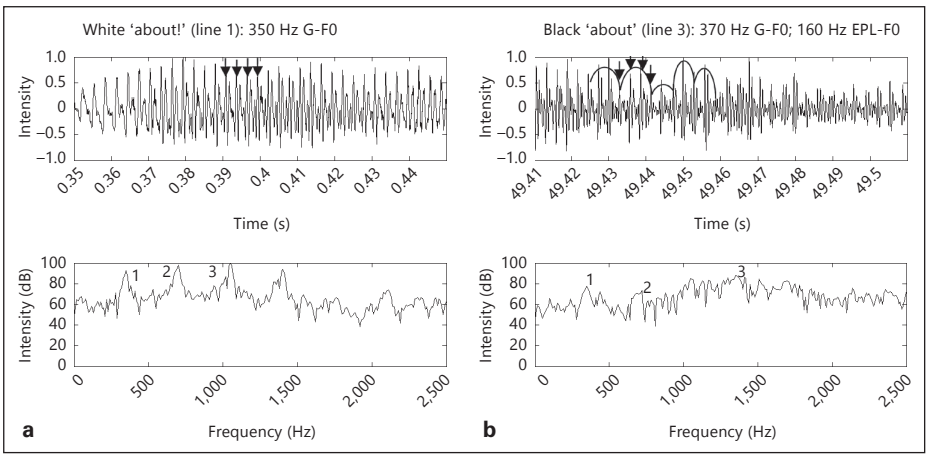

Fig. 3. Waveforms and spectra of black American comedian Steve Harvey portraying the white employee character (**a**) and the black employee character (**b**). (See fig. 2 for more details.)

4	this	about	Willie	Turner!	
[	dɪs	ə baːt	wi ɬi	$t^hɹ$::	nə]
{ **Ɣ!**	H	M	M	**H**	}
GL	445	175 360	330 260	285 150	
EPL	150	125	80	95 85	

The acoustic samples illustrate the difference in voice quality used for the white (fig. 3a) and black (fig. 3b) characters for the stressed syllable of the word *about,* produced at roughly the same glottal F0. The waveform for the black employee provides evidence of irregular amplitude modulation, indicating epilaryngeal vibration. The corresponding spectral profile shows weak glottal harmonics with high levels of interharmonic noise, the presence of subharmonics, and increased noisy energy above 1,000 Hz: characteristics that are all consistent with HVQ. All of these features are absent in the acoustic plots associated with the voice of the white employee. The glottal harmonics form sharp spectral peaks, indicating strong periodicity of the glottal pulse (low jitter); there is no sign of the type of amplitude modulation associated with growling, which auditory impression confirms.

3.1.3 A White Actor on *Scrubs* 'Doing' Black

This case study (example 3) features a white character, John Dorian (played by white male actor, Zach Braff) from the sitcom *Scrubs*, explicitly performing a stereotypical black caricature while talking to his black friend, Christopher Turk (played by black male actor, Donald Faison). The setup (before line 2) for the performance of the black caricature is produced in what is presumably the John Dorian character's neutral voice, distinguished by its moderately low F0 values and modal phonation, indicating an unconstricted laryngeal setting. Just prior to line 2 ('and you be yellin' like'), a switch occurs from this neutral voice to the stereotype voice, which exhibits HVQ characterized by tense/pressed voice, raised larynx voice quality in conjunction with elevated glottal F0 (roughly 200 Hz above the neutral voice level), and

growling, mainly on emphatic syllables. The stereotype voice is qualified as a 'yelling' voice in the dialogue, which may partly explain the occurrence of HVQ, but this does not give us a satisfying explanation for the use of phonetic variables such as extremely increased glottal F0 level and the features associated with laryngeal constriction, i.e. raised larynx voice quality and growling. None of these qualities are requirements for yelling; they are merely potential correlates. (In relation to the data survey, see table 2; this voice is strikingly similar to the voice that black American comedian Chris Tucker uses for his stage persona.)

The segmental level also reflects the vocal switching that occurs in the skit, and the (apparently) careful management of ethnolinguistic resources [e.g. Benor, 2010]. Several segmental variables associated with BVE appear in connection with the performance of the black stereotype (again, monophthongization, non-rhotic codas, fortition of the interdental fricative, and so forth). None of these features are used in the neutral speech sections (before line 2). In the morphosyntactic domain, there is deletion of the copula in line 2 ('he behind…'). Another example is the inappropriate use of habitual *be* in 'you be yellin', a highly frequent feature of Mock Ebonics [Ronkin and Karn, 1999, p. 366].

Example 3. Transcription of part of a dialogue from *Scrubs* (Season 2, Episode 12, 'My New Old Friend', aired January 9th, 2003) on the subject of a black stereotype; from top to bottom: text, IPA, voice quality and prosody/intonation, glottal F0 (GL), and epilaryngeal F0 (EPL). YouTube search terms: 'Scrubs my new old friend black people stereotype'

John Dorian:

1	I	hate	that	stereotype		that	**all**	black	people		yell	at	movie		screens.	
[	ʔaj	ɦet	ðɪt	ste-^{i}	ɾɪp	ðət	ɑɬ	blæk	p^{i}	bɬ	jɛɬ	ət˺	mu	vi	skɻinz̥	]
{**V**		H		M			**M**	M			L		L		L	}
GL	170	285	315	208	160	140	145	135	120	110	120	120	115	115	100	

(modal voice) You know, like you go to see some horror flick and you be yellin' like

2	'Don't	go	in	there	girl,	he	be	**hind**	the	door!'	
[	dõŋ	gow	ĩn	dɛ-	gɻ:ɬ	hi	bi	ʜã::	dɪ	do:w	]
{ **Ļ V!**	H	H	H	H	**M**	H	H	**H**		**H**	}
GL	380	510	540	485	455	475	485	550	271	550	
EPL		90			150			90		110	

(modal voice) You know (chuckle)? It's like… it's offensive.

Christopher Turk:
You wish you were allowed to yell at the screen, doncha? (pause)

John Dorian:

3	'Why	did	she	go	in	there?			I	mean,	he's	be	hind	the	door!'	
[	waj	ɾɪd	ʃi	gow	ĩn	ðɛɪ			ʔaj	mĩn	hiz	ba	ɦã::n	dɪ	do:w	]
{ **Ṿ**	**L-H**	M	M	M	**H**	H-L	}{	**Ļ V!**	M	M	**H**	H	H		**H-L**	}
GL	280-420	340	300	300	425	485-270			260	300	400	360	400	400	500-200	
EPL		90							83	140	72				81	

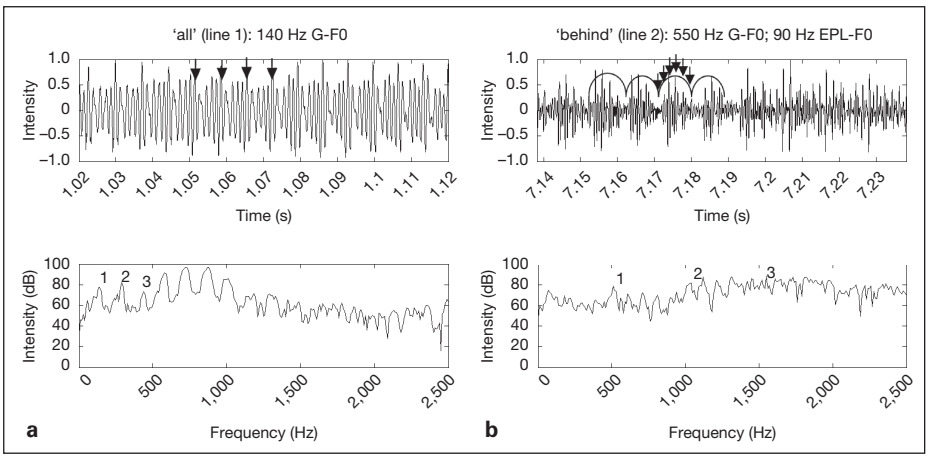

Fig. 4. Waveforms and spectra of *Scrubs* character John Dorian (white American actor Zach Braff) in his neutral voice (**a**) and 'doing' black (**b**). (See fig. 2 for more details.)

The acoustic data chosen for this example illustrate the difference between the neutral voice of the John Dorian character and the stereotype voice. The example in figure 4a is spoken in modal voice, and the corresponding waveform and spectrum support this observation: glottal pulses are regular in time and show minimal amplitude variation; in correspondence with this characteristic, the spectral profile contains well-defined harmonic structure and interharmonic noise is low. The syllable associated with the black stereotype voice in figure 4b stands in contrast: there is amplitude modulation, probably associated with an epilaryngeal vibration, and the glottal harmonics have nearly been lost in the spectrum, which is rich in additional subharmonic content, interharmonic noise, and increased acoustic energy above 1,000 Hz: all features indicative of HVQ with growling.

4 Discussion: Associations and Functioning of HVQ in the Data

In this section, the associations and functioning of HVQ are considered. Section 4.1 examines the use of HVQ in connection with portrayals of stereotypical black characters. Sections 4.2 and 4.3 illustrate the broader nature of HVQ in the data set.

4.1 Using Voice Quality in Contrasting Portrayals of Racial Stereotypes

Some of the performers under consideration manipulate laryngeal properties of their voice quality to correlate with the racial identity of the caricatures or characters they portray. Detailed illustrations were given in the Steve Harvey (3.1.2) and *Scrubs* (3.1.3) case studies; another example can be heard in black comedian Dave Chappelle's 'White and Black Peoples' Food' routine [Lathan, 2004]. In such cases, the performers overtly connect their performance with the theme of racial identity (e.g. through denotative content) and, in several cases, explicitly state that they are contrasting racial identities, as in the Steve Harvey case [for discussion on how black comedians explore issues of race through comedy, see Zolten, 1993].

The voices used for black characters[4] feature HVQ with moderate use of growling (particularly on emphatically stressed syllables), expanded glottal F0 range, and increased intensity; the characters typically express heightened emotional states (anger, outrage, surprise or exasperation). Stereotypical portrayals of white characters, when they occur, can involve the use of neutral-to-lowered larynx voice quality, mid-to-low glottal F0, and modal voice, although occasionally, slight breathiness, increased nasality, and elevated glottal F0 do occur (the examples given above provide illustrations of some of these features). The white characters are portrayed as being in neutral or non-aggressive emotional states [for example, the white employee character in the Steve Harvey (3.1.2) example is dismayed at being fired].

A possible, general analysis is that the voice quality mechanism distinguishing the black and white characters centres on whether or not the laryngeal constrictor mechanism (increased vocal-ventricular fold contact, aryepiglotto-epiglottal narrowing, possible epilarynx vibration, and a possible raised larynx setting, see section 1.1) is engaged. The black characters are performed using a constricted laryngeal state, exhibiting varying degrees of HVQ (with or without growling), and the white characters are performed with an unconstricted laryngeal state. It may be that performers are exploiting the salient auditory difference between these two basic settings of the larynx to help convey the racial juxtaposition forming the basis of these comedy routines. HVQ, particularly with growling, arguably has strong perceptual salience [see Gerratt and Kreiman, 2001]. Thus, it is a strong candidate for being encoded as part of a sociophonetic stereotype, and exaggerated contrast in voice quality has been established as a technique people use to signal a shift in the identity of the character/persona (or stereotype) being portrayed [Coupland, 2001; Podesva, 2007].

The phonetic grounding of these stereotypes is probably unrelated to authentic linguistic behaviour. Consider, for comparison, the fact that the speakers of the Zhenhai dialect of the Ningpo region, who can exhibit growling as part of the realization of their tone system, have been stereotyped by neighbouring groups as being confrontational or angry-sounding: 'better to argue with a person from Suchow than converse with someone from Ningpo' [Rose, 1989, p. 230]. A key difference between these cases is that, while the Zhenhai stereotype is an example of the value-encoding (or enregisterment) of an actual sociophonetic property of the language, HVQ is not a sociophonetic feature of any variety of BVE. Yet, in the performances under consideration here, other linguistic variables that do correlate with, most apparently, BVE and its stereotypes [Ronkin and Karn, 1999] are used, as was noted in sections 3.1.2 and 3.1.3. However, in comparison to segmental or morphosyntactic variables, voice quality, especially phonatory voice quality, arguably has a stronger association with affective state, of which it is reportedly a primary index [Laver, 1980, 1994; Nolan, 1983, p. 62; Scherer, 1986; Teshigawara, 2003]. Thus, the occurrence of voice quality in these cases is more likely associated with the portrayal of an affective stereotype than a sociolinguistic one.

4.2 Affective Associations of HVQ: Aggression and Its Comedic Inversion

In the cases addressed in the above section (4.1), HVQ is possibly associated with yelling/shouting voices (e.g. the black stereotype voice in case study 3.1.3 is identified

[4] HVQ appears in connection with black stereotypes portrayed in popular American media of the early 1900s. Examples include the Walter Lantz cartoons of the 1940s and Freeman Gosden's Amos character from Amos 'n' Andy (table 2).

as 'yelling' by the John Dorian character) and with angry or transgressive personalities, commonly associated with (particularly masculine) black stereotypes [Henry, 2002]. In general, HVQ is associated with extreme emotional states, such as anger or rage [Laver, 1980, p. 130; Scherer, 1986, p. 158] and aggressive posturing (the voices of professional wrestlers' personas, death metal vocalists[5], and some villain cartoon characters, such as Dr. Claw, support this assertion) [Ohala, 1996; Teshigawara, 2003; Tsai et al., 2010], but the relationship is not one of mutual entailment (i.e. it is not the case that HVQ must signify aggression and that aggression always correlates with HVQ), but rather, one of association. The same is true for yelling/shouting: the fact that someone yells does not necessarily mean they will be using HVQ (and vice versa). In this data, HVQ (with or without growling) is heard in non-emphatic and non-emotionally heightened speech, in speech where joy or other non-angry emotions are being expressed, and in cases without evident aggressive posturing (for example, Bill Cosby's Fat Albert is usually depicted as calm, happy, and friendly, but the character growls almost continuously).

A curious finding is that growling occurs throughout the glottal F0 range, and in several cases, it is paired with very high F0 levels. This is particularly true of the *Scrubs* stereotype voice (section 3.1.3) and Chris Tucker's on-stage voice (table 2): the combination of high glottal F0 with growling may serve to counteract the aggressiveness associated with HVQ for the sake of (what could be called) comedic inversion, generating laughter through a violation of listener expectations [Berger, 1976; Meyer, 2000].

A similar effect might be achieved by the use of exaggerated raised larynx voice quality with growling. This effect possibly explains the tendency for the use of extreme raised larynx voice quality with persistent growling in the voices of cartoon and puppet characters, such as Bill Cosby's Fat Albert, Frank Oz and Jim Henson's Muppets (notably Cookie Monster, Miss Piggy, Animal, Fozzie Bear, and Kermit the Frog), Frank Oz's Yoda from *Star Wars: The Empire Strikes Back* [Kershner, 1980], and Paul Reuben's Pee Wee Herman character (table 3). While HVQ with growling can be used to create a threatening persona, raised larynx voice seems to have the opposite effect by raising formant frequency (at least F1) and glottal F0[6], both of which are associated with small size and less imposing individuals or submissiveness [Ohala, 1996, p. 1813]. It is maybe not surprising then, that, in the survey, raised larynx voice tends to occur in media content intended for children.

4.3 Vocal Stylization: Imitating the Voice of Smokers?

HVQ is a possible feature of voice pathology caused by smoking. This analysis could explain the occurrence of HVQ in the voice of some performers reviewed here (e.g. Redd Foxx and Moms Mabley), though it cannot be offered as an explanation in every case. For example, the voice of Chris Rock's stage persona (discussed in section 3.1.1) prominently features HVQ. However, in interviews, Chris Rock tends not to use

[5] I would suggest that the baseline vocalization technique of death metal conveys an overall auditory impression of extreme low frequency through growling (which is associated with low frequency subharmonic structure) and low frequency resonance (which is attained through extreme lip protrusion and is possibly a means to compensate for the elevated larynx posture required to produce the growling).

[6] However, Nolan [1983, pp. 182–187] claims that raised larynx voice quality can involve a decrease in F2 and F3, rather than an increase, contra Sundberg and Nordström [1976]. For an argument that raised larynx is the high glottal F0 equivalent of pharyngealized voice, see Esling [1999] and Edmondson and Esling [2006].

his stage persona voice; rather, he uses a voice characterized by modal phonation and generally does not give any impression of being afflicted by vocal pathology.

It is possible that the HVQ of Chris Rock's stage persona voice emulates the voices of influential comedians, such as Redd Foxx, Dewey 'Pigmeat' Markham, Moms Mabley (table 2) and George Carlin (table 3). If this is correct, then this means that comedians like Chris Rock may have repurposed HVQ to serve as an index of the stage voices of famous American comedians, which happen to exhibit HVQ because of smoking. The same may be true of musicians who seem to emulate the harsh vocal style of influential singers: this pattern is particularly notable for Tom Waits and Captain Beefheart, who both closely echo the singing and musical styles heard in jazz, rhythm and blues, and rock and roll[7] as performed by musicians such as Howlin' Wolf, who probably had vocal pathology from smoking. Complicating this explanation, however, is the fact that most of the performers with suspected vocal pathology still demonstrate the ability to modulate the degree of HVQ in their performances. This capacity undermines the notion that vocal pathology is the ultimate source of the vocal styles associated with these genres of comedic and musical performance in the first place.

5 Conclusion

This paper has focused on performers' use of laryngeal phonetic properties of voice quality, particularly those associated with HVQ, in the portrayal of racial identity and stereotype in the context of popular media of the United States. A corpus of select data was qualitatively analyzed to evaluate the occurrence of HVQ in connection with the portrayal of black characters and stereotypes. Several examples were found in which the performer (or the performer's character) explicitly states the racial identity of the characters being portrayed (e.g. the Steve Harvey, section 3.1.2, and *Scrubs,* section 3.1.3, case studies). It was claimed that several performers overtly contrast racial identity partially through contrasting settings of the laryngeal constrictor mechanism. In these specific data, the constricted state was used for black characters, while the unconstricted state was used for white characters (as in the Steve Harvey case study, section 3.1.2). In these cases, the black stereotype exhibits HVQ in association with the depiction of exaggerated emotions, an aggressive stance, and yelling/shouting modes of speaking.

In the remaining data, HVQ is not strictly associated with sociophonetic stereotyping but rather functions more broadly. HVQ occurs as part of vocal stylization associated with various forms of music, as part of general aggressive posturing, and in connection with comedic functions. Yelling/shouting or vocal pathology explanations are ruled out as general root causes of the occurrence of HVQ, although they are probably explanatory in a subset of cases or as partial accounts. Sections 4.2 and 4.3 suggest that the use of HVQ is more nuanced and may serve comedic or even homage functions.

[7] An interesting connection not explored here is the fact that many of these styles originated in and were originally only accessible to black communities, jazz being a particularly prominent example [Anderson, 2004].

Taken together, the data considered here represent only a small fragment of the complex way that race (and ethnicity more generally) is depicted in the popular media of the United States. A number of issues can be explored in future research. First, this study focuses on the performers' representation of black characters and stereotypes using HVQ. One might ask follow-up questions, such as whether other racial or ethnic identities are ever (stereotypically) portrayed with HVQ, how blackness is portrayed in popular media from other geographic locations, and what other voice qualities correlate with representations of black social identity in popular media. Second, since most of the examples reviewed here are associated with males, one might wonder whether female performers employ HVQ, and, if so, what its function might be in such cases. Third, since, in this data, comedians most often overtly identified the racial identity of the characters they performed, an interview-based study of such performers could help obtain insight into whether they are consciously manipulating voice quality to contrast the racial or ethnic identity of characters they portray, and what their perspective is on the function of these voice qualities in sociophonetic stereotyping.

Acknowledgements

I wish to acknowledge Alexandra D'Arcy for her encouragement and support with the research. I am also grateful (if not deeply indebted) to the advice, input, and support offered by John Esling, Allison Benner, and Becky Childs. Laryngoscopic images were obtained with the generous support of John Esling and Lise Crevier-Buchman. I also appreciate the comments and corrections provided by my reviewers and the editorial team. Remaining blunders remain my own.

References

Abercrombie, D.: Elements of general phonetics (Edinburgh University Press, Edinburgh 1967).
Agha, A.: The social life of cultural value. Lang. Commun. *23:* 231–273 (2003).
Anderson, M.: The white reception of Jazz in America. Afr. Am. Rev. *38:* 135–152 (2004).
Bailly, L.; Henrich, N.; Pelorson, X.: Vocal fold and ventricular fold vibration in period-doubling phonation: physiological description and aerodynamic modeling. J. acoust. Soc. Am. *127:* 3212–3222 (2010).
Bakhtin, M.M.: Speech genres and other late essays (University of Texas Press, Austin 1986).
Ball, M.J.; Esling, J.H.; Dickson, B.C.: The VoQS system for the transcription of voice quality. J. int. phonet. Ass. *25:* 71–80 (1995).
Bell, A.; Gibson, A.: Staging language: an introduction to the sociolinguistics of performance. J. Socioling. *15:* 555–572 (2011).
Benor, S.B.: Ethnolinguistic repertoire: shifting the analytic focus in language and ethnicity. J. Socioling. *14:* 159–183 (2010).
Berger, A.A.: Anatomy of the joke. J. Commun. *26:* 113–115 (1976).
Boersma, P.; Weenink, D.: Praat: doing phonetics by computer (2010).
Borch, D.Z.; Sundberg, J.; Lindestad, P.-Å.; Thalén, M.: Vocal fold vibration and voice source aperiodicity in 'dist' tones: a study of a timbral ornament in rock singing. Logop. Phoniat. Vocol. *29:* 147–153 (2004).
Catford, J.C.: Fundamental problems in phonetics (Edinburgh University Press, Edinburgh 1977).
Catford, J.C.: Pharyngeal and laryngeal sounds in Caucasian languages; in Bless, Abbs, Vocal fold physiology: contemporary research and clinical issues, pp. 344–350 (College Hill Press, San Diego 1983).
Chun, E.W.: Ideologies of legitimate mockery: Margaret Cho's revoicings of Mock Asian. Pragmatics *14:* 263–289 (2004).
Coupland, N.: Dialect stylization in radio talk. Lang. Soc. *30:* 345–376 (2001).
Di Paolo, M.; Faber, A.: Phonation differences and the phonetic content of the tense-lax contrast in Utah English. Lang. Var. Change *2:* 155–204 (1990).
Edmondson, J.A.; Esling, J.H.: The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. Phonology *23:* 157–191 (2006).
Edmondson, J.A.; Esling, J.H.; Harris, J.G.; Li, S.; Lama, Z.: The aryepiglottic folds and voice quality in the Yi and Bai languages: laryngoscopic case studies. Mon-Khmer Stud. J. *31:* 83–100 (2001).

Esling, J.H.: Voice quality in Edinburgh: a sociolinguistic and phonetic study; doct. diss. University of Edinburgh, Edinburgh (1978).
Esling, J.H.: The IPA categories 'pharyngeal' and 'epiglottal': laryngoscopic observations of pharyngeal articulations and larynx height. Lang. Speech *42:* 349–372 (1999).
Esling, J.H.: There are no back vowels: the laryngeal articulator model. Can. J. Ling. *50:* 13–44 (2005).
Esling, J.H.; Harris, J.G.: States of the glottis: an articulatory phonetic model based on laryngoscopic observations; in Hardcastle, Beck, A figure of speech: a Festschrift for John Laver, pp. 347–383 (Erlbaum, Mahwah 2005).
Esling, J.H.; Zeroual, C.; Crevier-Buchman, L.: A study of muscular synergies at the glottal, ventricular and aryepiglottic levels. Proc. 16th Int. Congr. Phonet. Sci., Saarbrücken 2007, pp. 585–588.
Fink, B.R.: Folding mechanism of the human larynx. Acta oto-lar. *78:* 124–128 (1974).
Foulkes, P.; Docherty, G.: The social life of phonetics and phonology. J. Phonet. *34:* 409–438 (2006).
Fuks, L.: From air to music: acoustical, physiological and perceptual aspects of reed wind instrument playing and vocal-ventricular fold phonation; doct. diss. Royal Institute of Technology, Stockholm (1998).
Gallen, J.: Chris Rock: Never Scared (HBO 2004).
Gaudio, R.P.: Sounding gay: pitch properties in the speech of gay and straight men. Am. Speech *69*: 30–57 (1994).
Gerratt, B.R.; Kreiman, J.: Toward a taxonomy of nonmodal phonation. J. Phonet. *29:* 365–381 (2001).
Henry, M.: He *is* a 'bad mother*$%@!#': *Shaft* and contemporary black masculinity. J. pop. Film TV *30:* 114–119 (2002).
Henton, C.G.; Bladon, A.W.: Breathiness in normal female speech: inefficiency versus desirability. Lang. Commun. *5:* 221–227 (1985).
Henton, C.G.; Bladon, A.W.: Creak as a sociophonetic marker; in Hyman, Li, Language, speech and mind: studies in honor of Victoria A. Fromkin, pp. 3–29 (Routledge, London 1988).
Kershner, I.: Star Wars: Episode V – The Empire Strikes Back (Twentieth Century Fox Film Corporation 1980).
Knowles, G.: The nature of phonological variables in Scouse; in Trudgill, Sociolinguistic patterns in British English, pp. 80–90 (Arnold, London 1978).
Labov, W.: Sociolinguistic patterns (Blackwell, Oxford 1978).
Lathan, S.: Dave Chappelle: For What It's Worth (Pilot Boy Productions 2004).
Laver, J.: The phonetic description of voice quality (Cambridge University Press, Cambridge 1980).
Laver, J.: Principles of phonetics (Cambridge University Press, Cambridge 1994).
Levin, T.; Edgerton, M.: The throat-singers of Tuva. Scient. Am. *281:* 80–87 (1999).
Lindestad, P.-Å.; Södersten, M.; Merker, B.; Granqvist, S.: Voice source characteristics in Mongolian 'throat singing' studied with high-speed imaging technique, acoustic spectra, and inverse filtering. J. Voice *15:* 78–85 (2001).
Lindqvist-Gauffin, J.: A descriptive model of laryngeal articulation in speech. Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol., Stockh., No. 13, pp. 1–9 (1972).
Meek, B.A.: And the Injun goes 'How!': representations of American Indian English in white public space. Lang. Soc. *35:* 93–128 (2006).
Meyer, J.C.: Humor as a double-edged sword: four functions of humor in communication. Commun. Theor. *10:* 310–331 (2000).
Miller, A.L.: Guttural vowels and guttural co-articulation in Jul'hoansi. J. Phonet. *35:* 56–84 (2007).
Miller, A.L.; Brugman, J.; Sands, B.; Namaseb, L.; Exter, M.; Collins, C.: Differences in airstream and posterior place of articulation among N|uu clicks. J. int. phonet. Ass. *39:* 129–161 (2009).
Moisik, S.R.; Esling, J.H.; Crevier-Buchman, L.: A high-speed laryngoscopic investigation of aryepiglottic trilling. J. acoust. Soc. Am. *127:* 1548–1558 (2010).
Nolan, F.J.: The phonetic bases of speaker recognition (Cambridge University Press, Cambridge 1983).
Ohala, J.J.: Ethological theory and the expression of emotion in the voice. Proc. 4th Int. Conf. on Spoken Lang. Processing, Philadelphia 1996, vol. 3, pp. 1812–1815.
Paddock, H.: On roachness in English. Phonetica *34:* 318 (1977).
Pao, A.C.: False accents: embodied dialects and the characterization of ethnicity and nationality. Theatre Topics *14:* 353–372 (2004).
Podesva, R.J.: Phonation type as a stylistic variable: the use of falsetto in constructing a persona. J. Socioling. *11:* 478–504 (2007).
Rampton, B.: Styling the other: introduction. J. Socioling. *3:* 421–427 (1999).
Remlinger, K.: Everyone up here: enregisterment and identity in Michigan's Keweenaw peninsula. Am. Speech *84:* 118–137 (2009).
Rock, C.: Bigger & Blacker (Dreamworks 1999).
Ronkin, M.; Karn, H.E.: Mock Ebonics: linguistic racism in parodies of Ebonics on the Internet. J. Socioling. *3:* 360–380 (1999).
Rose, P.: Phonetics and phonology of yang tone phonation types in Zhenhai. Cah. Ling. Asie Or. *18:* 229–245 (1989).
Sakakibara, K.I.; Fuks, L.; Imagawa, H.; Tayama, N.: Growl voice in ethnic and pop styles. Proc. Int. Symp. on Musical Acoustics, Nara 2004.
Scherer, K.R.: Vocal affect expression: a review and a model for future research. Psychol. Bull. *99:* 143–165 (1986).
Schilling-Estes, N.: Constructing ethnicity in interaction. J. Socioling. *8:* 163–195 (2004).

Sicoli, M.A.: Shifting voices with participant roles: voice qualities and speech registers in Mesoamerica. Lang. Soc. *39:* 521–553 (2010).
Stevens, K.N.: Acoustic phonetics (MIT Press, Cambridge 1998).
Stiller, B.: Tropic Thunder (Comedy, DreamWorks 2008).
Stuart-Smith, J.: Voice quality in Glaswegian. Proc. 14th Int. Congr. Phonet. Sci., San Francisco 1999, pp. 2553–2556.
Sundberg, J.; Nordström, P.E.: Raised and lowered larynx – the effect on vowel formant frequencies. Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol., Stockh., No. 17, pp. 35–39 (1976).
Teshigawara, M.: Voices in Japanese animation: a phonetic study of vocal stereotypes of heroes and villains in Japanese culture; doct. diss. University of Victoria, Victoria (2003).
Traill, A.: The laryngeal sphincter as a phonatory mechanism in !Xóõ (Bushman); in Singer, Lundy, Variation, culture and evolution in African populations. Papers in honour of Dr. Hertha de Villiers, pp. 123–131 (Witwatersrand University Press, Johannesburg 1986).
Trudgill, P.: The sociolinguistic differentiation of English in Norwich (Cambridge University Press, Cambridge 1974).
Tsai, C.-G.; Wang, L.-C.; Wang, S.-F.; Shau, Y.-W.; Hsiao, T.-Y.; Auhagen, W.: Aggressiveness of the growl-like timbre: acoustic characteristics, musical implications, and biomechanical mechanisms. Music Percept. *27:* 209–222 (2010).
Vinson, C.: Steve Harvey: One Man (HBO 1997).
Yuasa, I.P.: Creaky voice: a new feminine voice quality for young urban-oriented upwardly mobile American women? Am. Speech *85:* 315–337 (2010).
Zemlin, W.R.: Speech and hearing science: anatomy and physiology; 4th ed. (Allyn & Bacon, Boston 1998).
Zolten, J.J.: Black comedians: forging an ethnic image. J. Am. Culture *16:* 65–76 (1993).