



# Phrase-Final Voice Quality Variation Among Black and Latinx Southern California Youth

Nicole R. Holliday<sup>1</sup>

<sup>1</sup>Department of Linguistics, University of California, Berkeley, United States of America  
nicole.holliday@berkeley.edu

## Abstract

In the U.S., differences between ethnolectal varieties such as Latinx English and African American English have been well-described for decades, especially in the realm of segmental phonological variation. However, few studies have examined patterns of voice quality features across ethnolectal varieties of U.S. English. The current study examines acoustic correlates of creakiness, modalness and breathiness (H1-H2, HNR, and CPP values) in the speech of 24 Southern California high school students: 17 Latinx and 7 Black American. Results of multiple regression analyses reveal significant differences in phrase-final voice quality, with Latinx students displaying patterns that are breathier and Black students displaying ones that are more modal or creakier. These results may be due to different strategies across ethnolects for accomplishing acoustic cues to phrase-final position. They can also reflect different levels of participation in the ongoing increase in the use of creaky voice among young speakers of Mainstream American English. This study acts as a first step towards understanding how racialized communities in the U.S. may participate in sociophonetic changes in progress that interact with the making of prosodic meaning.

**Index Terms:** voice quality, breathiness, sociophonetics, African American English, Latinx English, ethnolinguistic variation

## 1. Introduction

Ethnolinguistic differences, especially those between groups of white and Black speakers, have been a major area of interest in sociophonetic research over the last several decades. Within and between communities, systematic sociolinguistic differences between groups have also been well-described. Much of this research has focused on segmental phonological differences between ethnolects, with few studies focusing on intonational and prosodic differences [1,2]. However, such differences are frequently identified as salient between other types of groups; for example, studies on phonation type (especially creaky voice) have found perceptually salient differences between speaker by gender and age [3]. Additionally, recent work on the pragmatic and social meaning of voice quality has revealed that listeners attach different types of social stereotypes to speakers that employ different types of creaky, modal and breathy voice qualities [4,5]. As a result, understanding how different groups may variably employ various voice quality features is an important question for sociophonetic research going forward.

A number of studies have found that American English speakers routinely employ creaky voice not only as a stylistic feature that carries social meaning, but also as a cue to end

Intonational Phrases (IPs) or speaking turns [6]. Researchers who have focused on the social meaning of creaky voice have largely investigated its socially-salient gender stereotypes but less work had been done on how such voice qualities may index ethnic differences. Greer and Winters [3] find that listeners associate creaky voice with female gender and youth, but also with an attempt at authoritativeness. Other studies, such as Yuasa's [7] have found an increase in the use of creaky voice across domains, with an overall rise in the use of and perception of use of this phonation type over the past few decades. Pragmatic studies on creaky voice have focused on its function to indicate phrase-finality in American English [8]. Cross-linguistically, speakers have a number of cues that they may employ to indicate phrase-finality. In Spanish, creaky voice has been less well-documented than in English, but studies have found that speakers can indicate the end of a turn using modal voice as well as non-modal strategies like breathiness [9], while heritage speakers are more likely to use creakiness similar to their monolingual anglophone counterparts [10]. With respect to breathy voice, though it has been less well-described than creakiness across varieties of American English, recent studies have found that listeners generally associate it with a persona that is friendly and attractive [11,12].

Despite these recent findings about how voice quality variation may be both produced by speaker and perceived by listeners, nearly all of the aforementioned studies have focused primarily or even exclusively on white speakers of Mainstream American English (MAE). This represents a serious lacuna, due to the changing demographics of the U.S., especially in multiethnic and multi-lingual urban centers. The current study aims to build on earlier research on the pragmatic meaning of variable phrase final phonation types, as well as studies on how voice quality may be employed in ethnolinguistic variation in order to better understand how a wide variety of American English speakers may employ such differences for social and pragmatic purposes. The research questions are as follows: (1) Are there differences between Latinx and Black speakers with respect to voice quality in phrase-final vowels, as measured by H1-H2, Cepstral Peak Prominence, and Harmonics to Noise Ratio? (2) Does the comparison of the two groups' values for each measure indicate different sociopragmatic strategies for the realization of phrase-finality? (3) Do Black and Latinx students employ phrase-final voice quality differences in ways that parallel the findings of previous work on young, white MAE speakers?

It is important to note that perceptual studies about the social and pragmatic meanings of creaky and breathy voice, have generally not differentiated between the different types of creaky and breathy voice quality types that have been detailed in phonetic work. While authors such as Keating and Esposito

[13] have made clear that much of what we refer to as creaky voice maybe actually be caused by several types of physical properties of the voice. For this reason, the current study will focus on a comparative analysis of several of the most frequently documented acoustic properties that have been shown to distinguish creaky, breathy, and modal voice, and will largely limit its comparisons to the two groups analyzed here. The measures of interest (H1-H2, CPP, and HNR) will be discussed individually ahead of the results obtained for each metric.

## 2. Methodology

### 2.1 Participants and Data Collection

Data for this study comes from a large-scale sociolinguistic project focused on ethnolinguistic variation among teens, aged 14-17, in Southern California. Sociolinguistic interviews were collected in 2019 by the author and a research team of two undergraduate students and six high school seniors involved in a Summer Enrichment Program (SEP) at a local college. 24 peer sociolinguistic interviews of ~20 minutes in duration were recorded using Zoom H1 recorders and Audio Technica AT831 lapel microphones. 17 speakers identified as Latinx and 7 identified as Black American. 14 speakers identified as female and 10 identified as male. The SEP program had no participants who identified as white or Native American, and very few who identified as Asian, which is why only Black and Latinx students are considered here. Participants came from communities within Los Angeles and San Bernardino counties, both of which have a plurality of Latinx residents. All participants reported themselves as lower or middle class, and all came from communities with a median income below the average for LA County. All 17 Latinx participants reported speaking both English and Spanish at home on a daily basis. None of the Black participants reported speaking Spanish or any language other than English at home.

### 2.2. Data Processing and VQ Analysis

Following the interviews, the research team transcribed and annotated all interview data for each participant using the software program ELAN [14]. The interview data was then force-aligned using DARLA's semi-automated system [15]. The present analysis focuses on the final vowel in each Intonational Phrase of each interview in order to address questions related to possible sociopragmatic differences in the realization of phrase-finality. For each IP, the final vowel was extracted and voice quality (VQ) parameter extraction was performed using the PraatSauce [16] suite of scripts in Praat [17]. This package measures 34 harmonic amplitude components from the low-, mid-, and high-frequency regions of the signal (e.g. H1, H2, A1, A2, A3, H2k, H5k), cepstral peak prominence (CPP), and harmonic and amplitude differences (H1-H2c, H1-A1) [18, 19]. Following the methods in these earlier studies, each vowel was divided into five parts of equal duration. Measurements were made for each vowel by averaging the value for each measure of that segment. The mean number of tokens available for analysis for each speaker was 139.

### 2.3 Statistical Analysis

Following extraction of the parameters of interest, I ran several mixed-effects regression models in R [20] that were designed to test for differences between the Black and Latinx speaker groups for several correlates of breathiness, creakiness, and modality (H1-H2, HNR, and CPP), given that a previous LASSO regression model indicated that these features were most likely to contribute to the observed differences between the two groups. Unless otherwise indicated, all measures are corrected for the effects of formants. All models are linear mixed effects models taking ethnicity, gender, and their interaction as fixed effects and speaker as a random effect unless otherwise noted. All models were implemented in R [20] and data visualizations were created using the ggplot2 package [21].

## 3. Results

Results of the models for H1-H2, HNR, and CPP will each be presented in turn to better articulate the differences between the two groups for each parameter. However, it is important to note that these voice quality (VQ) measures are complex and overlapping, and the relationships between the three parameters of interest must be taken together in order to fully address the research questions at hand. Overall, results indicate substantial differences in phrase-final phonation between the Black and Latinx groups, with Black speakers overall demonstrating patterns that are more modal or creaky, and Latinx speakers demonstrating a more breathy pattern. The results the analyses for each parameter will be discussed individually and then they will be interpreted and discussed together.

### 3.1 H1-H2c

H1-H2c (corrected for formants) is a measure of spectral tilt that is designed to capture the amplitude difference between the first two harmonics in a given speech signal. H1-H2c generally reflects glottal constriction, with a lower value indicating greater constriction [22, 23]. In general, phonation that is perceived as creaky voice has lower values of H1-H2c, because the glottis is usually constricted. However, the current study is primarily focused on creakiness, modalness, and breathiness on a continuum between the two speaker groups, and will use to variety of methods to address potential differences. Results of a mixed-effects regression model for H1-H2c with ethnicity and gender as fixed effects and speaker as a random effect do indeed show significant differences between the Black and Latinx groups, with the Latinx group demonstrating overall higher values for H1-H2c (Est.= 5.907, SE= 0.120,  $p < .01$ ). Significant differences were also observed between gender groups (Est.=0.98, SE= 0.178,  $p < .01$ ), and for the interaction between race and gender (Est.= 1.321, SE= 0.216,  $p < .01$ ), indicating that the female Latinx speakers have the highest H1-H2c values.

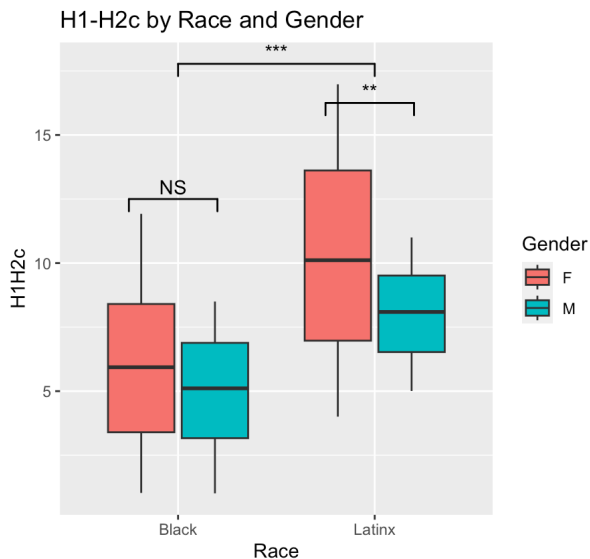


Figure 1: Mean H1-H2c values for Latinx and Black speaker groups, separated by speaker gender.

In isolation, given that H1-H2c values can be indicative of differences between modal and non-modal voicing, these results indicate that the Latinx speakers are likely to be using phonation that is less modal, and possibly less creaky or breathier. The overall higher values for the Latinx speakers may be interpreted as evidence of less constricted phonation generally. Alternatively, it could indicate less phrase-final creakiness in the speech of the Latinx participants than in the speech of the Black participants, since previous work has found that speech with a higher H1-H2c is typically less creaky [19].

### 3.2 HNR

Harmonics to Noise Ratio (HNR) is a measure primarily used to quantify additive noise in the vocal signal. In general, non-modal speech (breathy, creaky, etc.) will have higher values for HNR, given that incomplete glottal closures, such as those observed in creaky or breathy phonation, will typically result in more noise in the signal [19]. HNR measurements, taken using PraatSauce [16] are extracted separately for 0-500Hz, 0-1500Hz, 0-2500Hz, and 0-3500Hz. Researchers such as Xu et al [24] claim that low HNR below 500 (HNR05) is more likely to be associated with a creakier voice quality, while low HNR below 3500 (HNR35) is more likely to be associated with breathy voice quality. As the current analysis is interested primarily in phrase-final phonation and non-modal speech, the analysis of HNR will be limited to the lowest and highest parts of the signal, HNR05 and HNR35. These analyses were again conducted via mixed effects regression models testing for differences in HNR by ethnicity and gender and its interaction, with a random effect of speaker.

With respect to HNR05, there appear to be no significant differences between the two race groups, nor are there significant differences by gender or the race/gender interaction (all  $p > .05$ ). However, the Latinx group shows slightly higher mean HNR values than the Black speaker group. For HNR35, however, the model results indicate significant differences by race (Est.= 15.943, SE=0.374,  $p < .01$ ). Overall, Black speakers have lower HNR values than Latinx speakers (Est.= 5.226,

SE=0.443,  $p < .01$ ), though this differences appears to be somewhat driven by values for Black female speakers, though the interaction is not significant.

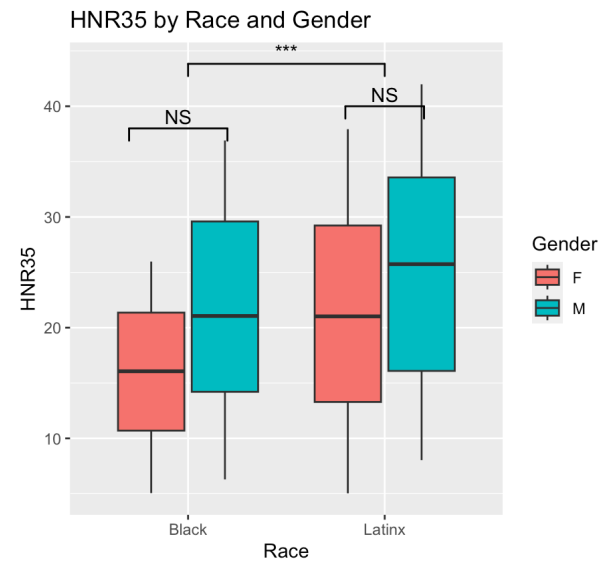


Figure 2: Mean HNR35 values for Latinx and Black speaker groups, separated by speaker gender.

These results indicate that while both speaker groups demonstrate substantial noise in the speech signal below 3500Hz, the Latinx speakers have values that are more likely to be associated with breathy voice, while the Black speakers have values that are more likely to be associated with modal or creakier phonation. This parallels the results that we observed for the differences between H1-H2c values for the two speaker groups.

### 3.3 CPP

Cepstral Peak Prominence (CPP), like HNR, is another measure of harmonics to noise, however, it is measured by taking “the difference in amplitude of the cepstral peak and a regression line to normalize for window size and energy” [25]. One straightforward description of CPP is that greater periodicity in the signal shows stronger harmonics resulting from a more prominent cepstral peak, thus CPP is a useful metric for examining voice quality, especially breathiness [26]. Previous studies have found that relative CPP values can distinguish breathy from modal voice quality in both production and perception [27, 28]. To test for differences in values between the groups, I conducted another mixed effects regression model for CPP by ethnicity and gender and the interaction, again with a random effect of speaker. Results indicate that the Black speakers in this sample have overall lower CPP values than Latinx speakers, patterning similarly to the HNR35 results discussed above. Of particular interest is the fact that the Black male speakers have the lowest CPP values of all groups (mean =6.831), indicating that they are most likely to be participating in phrase-final modal or even creaky phonation. In contrast, the Latinx females have the highest CPP values, which provides another indication that they are using breathier phrase-final phonation than the other groups (mean=13.854), particularly when these results are combined with the findings above.

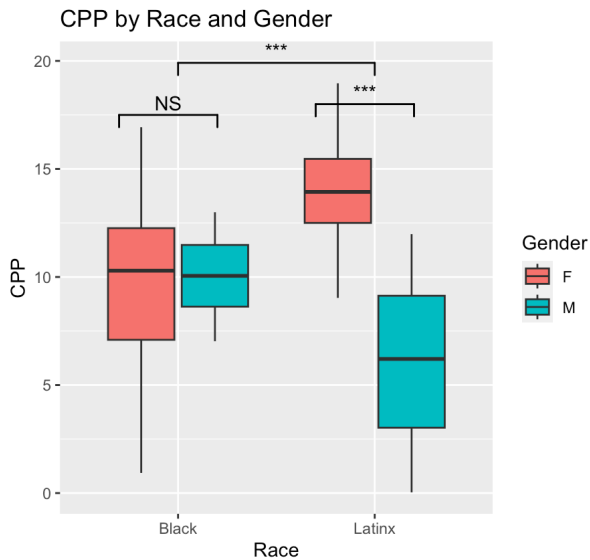


Figure 3: Mean CPP values for Latinx and Black speaker groups, separated by speaker gender.

Regression model results indicate significant differences by race (Est.= 4.644, SE=0.179,  $p < .01$ ) and by gender (Est.= 0.616, SE=0.228,  $p < .01$ ). There is also a significant result for race\*gender interaction (Est.= 8.579, SE=0.273,  $p < .01$ ). One other result of interest is that the speakers in Latinx male group and those in Black female speaker group do seem to demonstrate more within-group variation than their other counterparts. This may be due to more consistent voice quality differences between speakers within each of the groups, but more data from each speaker would be necessary to confirm this hypothesis.

#### 4. Discussion and Conclusion

The results of multiple mixed effects regression models indicate significant differences between how the Latinx and Black teenagers in this sample employ phrase-final non-modal phonation in naturalistic interview speech. Overall, the Latinx group demonstrates higher H1-H2c values and higher CPP values than the African American group. With respect to HNR, Latinx speakers demonstrate higher values than Black speakers in frequencies below 3500m Hz, though there is no significant difference between the groups when only frequencies between 0-500Hz are considered.

Taken together, these results indicate that Latinx speakers are more likely to use phrase-final breathy phonation, while Black speakers are more likely to use phrase final creaky or, in some cases, modal phonation. The results for H1-H2c and HNR primarily show that the Latinx speakers are breathier in phrase-final position than Black speakers, and the results for CPP support the conclusion that Black speakers may also be creakier, especially in light of previous studies on the use of creaky voice in American English more broadly [3,5]. However, it is important to note that since these earlier studies have focused primarily on white speakers, there is no clear result with which to compare the Black or Latinx speakers in this sample. Going forward, phonetic and sociolinguistic work

should address the lacuna related to voice quality research on Black American speakers, especially across age and regional groups.

One plausible explanation for this pattern is that the Latinx speakers are participating in language-contact phenomena resulting from community and/or individual level bilingualism. Previous studies have indicated that monolingual Mexican Spanish speakers generally do not employ phrase-final creaky voice, while heritage speakers of Spanish have been found to do so [9,10]. However, these earlier works primarily focused on the difference between creaky and modal-phrase final-phonation, without particular attention to possible effects of breathiness. The current study has demonstrated that young, bilingual Spanish-English speakers in California may participate in a different type of strategy for marking phrase-finality using breathy voice quality.

Future studies should investigate how listeners perceive these acoustic differences between groups of young speakers. For example, it would be useful to understand whether American listeners perceive phrase-final breathiness as a unique feature of Latinx English, and whether they can successfully perform ethnic identification of speakers based on voice quality properties alone. Additionally, while the various social meanings of creaky voice in American English are generally better understood than the social meanings of breathy voice, researchers have not yet specifically investigated how listeners perceive non-white speakers who employ non-modal voice quality, or how they may be subject to the negative social stereotypes that are often attached to users of creaky voice. A better understanding both the socio-pragmatic and indexical meanings of variable voice quality across ethnolects will be especially important for systems that rely on speech recognition or that aim to perform sociolinguistically accurate speech synthesis.

#### 5. Acknowledgements

The author would like to thank research assistants Emily Vomvass and Jade Hill for their help with data collection and transcription, and Franny Brogan for assistance with early data processing. She would also like to thank Li-Fang Li and Yoonjeong (Cynthia) Lee for early feedback.

#### 6. References

- [1] Thomas, E. R. Prosodic features of African American English. In S. Lanehart (Ed.), *The Oxford handbook of African American Language*, pp. 420-438. Oxford: Oxford University Press. 2015
- [2] Holliday, N. "Prosody and sociolinguistic variation in American Englishes." *Annual Review of Linguistics* vol 7, pp. 55-68. 2021
- [3] Greer, S.D. and Winters, S.J., 2015. The perception of coolness: Differences in evaluating voice quality in male and female speakers. In *ICPhS*.
- [4] Davidson, L., "The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages". *Wiley Interdisciplinary Reviews: Cognitive Science*, 12(3). 2021
- [5] Podesva, R.J., and P. Callier. "Voice quality and identity." *Annual review of applied Linguistics* 35 pp. 173-194. 2015
- [6] Garellek, M. "Perception of glottalization and phrase-final creak." *The Journal of the Acoustical Society of America* 137, no. 2. pp. 822-831. 2015
- [7] Yuasa, I.P.. "Creaky voice: A new feminine voice quality for young urban-oriented upwardly mobile American women?." *American Speech* 85, no. 3, pp 315-337. 2010

- [8] Heldner, M., M. Włodarczak, Š. Beňuš, and A. Gravano. "Voice quality as a turn-taking cue." In *Interspeech 2019, Graz, Austria, 15-19 September 2019*, pp. 4165-4169. 2019.
- [9] González, C., C. Weissglass, and D. Bates. "Creaky voice and prosodic boundaries in Spanish: An acoustic study." *Studies in Hispanic and Lusophone Linguistics* 15, no. 1: 33-65. 2022
- [10] Kim, JY.. "Heritage speakers' use of prosodic strategies in focus marking in Spanish." *International Journal of Bilingualism* 23, no. 5: pp. 986-1004. 2019.
- [11] Babel, M., J. King, G. McGuire, and T. Miller. "Acoustic determiners of vocal attractiveness go well beyond apparent talker size." *Laboratory Report: University of British Columbia and University of California, Santa Cruz*. 2011.
- [12] Holliday, N. "Siri, you've changed! Acoustic properties and racialized judgments of voice assistants." *Frontiers in Communication* 8. 2023.
- [13] Keating, P., and C. Esposito. "Linguistic voice quality." *UCLA Working Papers in Phonetics* 105, pp 85-91. 2007
- [14] "ELAN (Version 6.7) [Computer software]. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>". 2023.
- [15] Reddy, S. and J. Stanford. "A Web Application for Automated Dialect Analysis". In *Proceedings of NAACL-HLT 2015*. 2015
- [16] Kirby, J.. *Praatsauce: Praat-Based Tools for Spectral Analysis*. 2018.
- [17] P. Boersma and D. Weenink, "Praat: doing Phonetics by Computer [Computer program version 6.2.14]," <https://www.fon.hum.uva.nl/praat/>. 2023.
- [18] DiCanio, C. "The phonetics of register in Takhian Thong Chong." *Journal of the International Phonetic Association* 39, no. 2, pp 162-188. 2009.
- [19] Garellek, M., W.F. Katz, and P.F. Assmann. "The phonetics of voice." *The Routledge handbook of phonetics* pp: 75-106. 2019.
- [20] R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>, 2019.
- [21] H. Wickham, *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag, 2016.
- [22] Keating, P., M. Garellek, and J. Kreiman. "Acoustic properties of different kinds of creaky voice." In *ICPhS*, vol. 2015, no. 1, pp. 2-7. 2015.
- [23] Lee, Y., P. Keating, and J. Kreiman. "Acoustic voice variation within and between speakers." *The Journal of the Acoustical Society of America* 146, no. 3, pp 1568-1579. 2019.
- [24] Xu, C. P. Foulkes, P. Harrison, V. Hughes, and J. H. Wormald. "Contributions of acoustic measures to the classification of laryngeal voice quality in continuous English speech." In *Proceedings of the International Congress of Phonetic Sciences (ICPhS)*. ASSTA, 2023.
- [25] Garellek, M. and P. Keating. "The acoustic consequences of phonation and tone interactions in Jalapa Mazatec." *Journal of the International Phonetic Association* 41, no. 2 pp 185-205. 2011.
- [26] Hillenbrand, J., Cleveland, R.A. and Erickson, R.L., "Acoustic correlates of breathy vocal quality". *Journal of Speech, Language, and Hearing Research*, 37, no 4, pp.769-778. 1994.
- [27] Esposito, C.M., 2010. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics*, 38 no 2, pp.306-316. 2010
- [28] Esposito, Christina M., and Sameer ud Dowla Khan. "The cross-linguistic patterns of phonation types." *Language and Linguistics Compass* 14, no. 12. 2020