INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Acoustic differences between English /t/ glottalization and phrasal creak

Marc Garellek[1], *Scott Seyfarth*[1]

[1]Department of Linguistics, UC San Diego
mgarellek@ucsd.edu, sseyfarth@ucsd.edu

Abstract

In American English, the presence of creaky voice can derive from distinct linguistic processes, including phrasal creak (prolonged irregular voicing, often at edges of prosodic phrases) and coda /t/ glottalization (when the alveolar closure for syllable-final /t/ is replaced by or produced simultaneously with glottal constriction). Previous work has shown that listeners can differentiate words in phrasal creak from those with /t/ glottalization, which suggests that there are acoustic differences between the creaky voice derived from phrasal creak and /t/ glottalization. In this study, we analyzed vowels preceding syllable-final /t/ in the Buckeye Corpus, which includes audio recordings of spontaneous speech from 40 speakers of American English. Tokens were coded for presence of phrasal creak (prolonged irregular voicing extending beyond the target syllable) and /t/ glottalization (whether the /t/ was produced only with glottal constriction). Eleven spectral measures of voice quality, including both harmonic and noise measures, were extracted automatically and discriminant analyses were performed. The results indicate that the discriminant functions can classify these sources of creaky voice above chance, and that Cepstral Peak Prominence, a measure of harmonics-to-noise ratio, is important for distinguishing phrasal creak from glottalization.

Index Terms: creaky voice, phonation, glottalization, voice

1. Introduction

Creaky voice tends to have a low and irregular F0, as well as increased vocal fold constriction, relative to modal voice. However, previous studies have shown that numerous subtypes of creaky voice exist [1, 2, 3]: some forms of creaky voice are constricted but have a regular F0, whereas other types are not constricted but are irregular in F0 [3, 4]. In this study, we test whether different *linguistic* sources of creaky voice are produced with different acoustic characteristics.

In American English, creaky voice has several distinct linguistic origins. First, creaky voice may occur as a result of so-called '/t/ glottalization,' the phenomenon in which glottal constriction is produced simultaneously with or instead of /t/ [5,6,7,8]. For example, the word 'about' can be produced without glottalization as [əbaʊt], or as glottalized [əbaʊʔ]. Glottalized and non-glottalized tokens of 'about' are shown in the top row of Figure 1. The glottalized token (top-right, panel b) shows an increasing glottal period and greater F0 irregularity towards as the vowel progresses, with no identifiable stop closure or release burst associated with the coda /t/. In contrast, a clear stop closure and release burst can be seen in the non-glottalized token (top-left, panel a).

Another source of creaky voice in American English is phrase-final creak, the phenomenon in which ends of prosodic phrases are creaky [9, 10]. In contrast to /t/ glottalization, which affects only coda /t/ of a single word [11], phrase-final creak

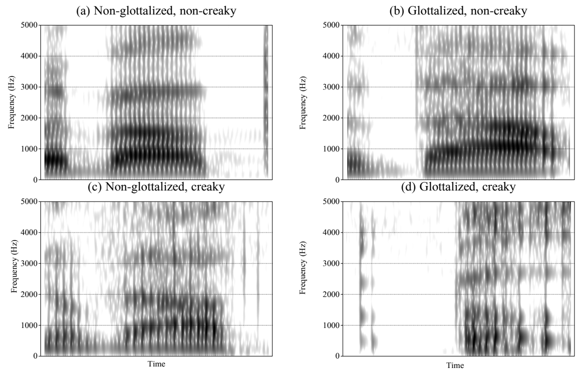

Figure 1: *Sample spectrograms of 'about,' with presence of glottalization crossed with phrasal creak. The creaky tokens (panels c and d) are differentiated from the non-creaky ones (panels a and b) by irregularity throughout the vowels and voiced portions of the word. In the glottalized productions (panels b and d), there is also a distinct region of strong glottal irregularity that is localized to the end of the syllable.*

may span several words and appears on any voiced sound. Indeed, some speakers produce creaky voice over large portions of phrases, and not just at the ends of prosodic domains. Such cases might be used to index speaker identity [12]. In our study, we refer to both phrase-final creak and longer durations of creaky voice as 'phrasal creak.'

Because /t/ glottalization and phrasal creak act on different domains (only on /t/ for glottalization; over long strings of segments for phrasal creak), these two sources of creaky voice can co-occur on a particular word. For example, the bottom-right panel of Figure 1 shows a token of 'about' with both /t/ glottalization (as seen by the lack of stop closure and [t] release burst) and phrasal creak (as seen by the irregular voicing over the entire word). Can listeners disambiguate between different linguistic sources of creaky voice? It has been shown that English listeners are able to disambiguate words with /t/ glottalization (e.g. 'atlas' [æʔləs]) from non-glottalized counterparts (e.g. 'Alice' [æləs]), even when the latter occur in phrasal creak [13]. This suggests that there are systematic acoustic differences between /t/ glottalization and phrasal creak. However, this previous study used stimuli from laboratory speech recorded by only one phonetically trained speaker. This leaves it unclear as to which acoustic features are perceptually relevant and robust for disambiguating /t/ glottalization from phrasal creak, and which may be idiosyncratic.

In this paper, we investigate whether English speakers produce /t/ glottalization and phrasal creak with different kinds of creaky voice in spontaneous speech that are acoustically well-

http://dx.doi.org/10.21437/Interspeech.2016-1472

discriminated. If indeed they do, this would suggest that different subtypes of creaky voice can be utilized for different linguistic purposes (e.g. /t/ glottalization vs. phrasal creak) – that is, subtypes of creaky voice can contrast within a particular language.

2. Corpus data

The recordings used in this study were obtained from the Buckeye Corpus of spontaneous American English [14], which contains speech of 40 adults (gender-balanced) from Ohio. The recordings took place in a quiet room and were digitized at 16 kHz with 16-bit resolution. The corpus has canonical phonemic transcriptions for each word (generated by automatic alignment software), as well as narrow phonetic transcriptions, which were made by corpus annotators who hand-corrected the labeling and segmentation of the canonical phonemic transcriptions.

We analyzed the following subset of the corpus, which is taken from the dataset constructed for and described in [15], summarized here with additional exclusions and hand-corrections. Words with syllable-final /t/ in the canonical form were extracted from the corpus, based on a syllabification algorithm adapted from [16]. We selected only tokens of coda /t/ that were preceded by a vowel at least 50 ms long (as short samples are problematic for voice analyses), were not part of a complex coda (e.g., excluding 'kept', 'rant'), and which were realized phonetically as a voiceless stop [t] or as a glottal stop [ʔ]. 11,947 tokens in the corpus met these initial criteria. We hand-inspected each token, and excluded 346 tokens that were disfluent, were misaligned, or had noisy or clipped recordings.

Glottal stops were identified based on the provided corpus annotations, and verified during hand-inspection. An instance of /t/ was considered to be realized as a glottal stop if it had irregular voicing localized to the onset and offset of the target coda, and no [t] release burst. The provided annotations were found to be largely accurate: of the 6,044 glottal stops meeting the criteria, only 462 (7.6%) were hand-corrected to a [t], and of the 5,557 [t] codas meeting the criteria, 1,399 (25.2%) were hand-corrected to a glottal stop. During hand-inspection, we also removed a net of 532 tokens that were found to be voiced, but which were originally annotated as voiceless in the corpus.

Phrasal creak was identified based on the annotations provided in the corpus log files combined with hand-inspection. Phrasal creak was defined as a period of irregular voicing that lasted for at least twice the duration of the target vowel preceding syllable-final /t/. Thus, a /Vt/ sequence with creaky voice localized only to the target vowel would not be considered to occur in phrasal creak, because the creaky voice does not extend beyond the target vowel. This criterion was used so as to avoid misidentifying glottalization as phrasal creak.

3. Acoustic analyses

3.1. Acoustic measures

Eleven acoustic measures, shown in Table 1, were obtained automatically using VoiceSauce [17]. For tokens of [t], the measures were extracted over the preceding vowel; for tokens of [ʔ], the measures were extracted over the the vowel plus any portion of voiced glottalization.

The 11 measures included fundamental frequency (F0), measured using the STRAIGHT algorithm [18], plus three noise measures: cepstral peak prominence (CPP), a normalized measure of noise [19], harmonics-to-noise ratio [20] below 500 Hz (HNR05), and subharmonics-to-harmonics ratio (SHR), which measures the amplitude difference between the harmonics and subharmonics in the signal [21]. Because F0 tends to be irregular during prototypical creaky voice, we expect both CPP and HNR05 to be lower in creakier voice qualities [3, 22]. Moreover, the frequent occurrence of period-doubled phonation during creaky voice should increase the SHR [3]. The seven remaining measures were estimates of spectral tilt in various frequency bands. All of these seven measures are expected to be lower in creakier voice qualities, because of the increased glottal constriction and/or abrupt glottal closure [3,22,23]. Because these measures also vary as a function of vowel quality, they were corrected for vowel formants, following [24], which allows for cross-vowel comparisons in voice quality. Corrected measures are shown with asterisks.

Due to the importance of correct F0 tracking for many of these measures, we took two steps to exclude tokens that were likely mistracked. First, we excluded 414 tokens which contained at least one analysis frame in which F0 was either half or double the F0 value in the previous frame (within a 10% margin). While F0 may drop rapidly during creaky voice, an octave jump is most likely to be a tracking error. Second, we standardized F0 within two groups within each speaker: modal [t] tokens, and tokens with phrasal creak and/or glottalization. We excluded tokens with a mean F0 > 2.5 standard deviations from each speaker's modal or creaky mean, as appropriate.

Finally, outliers > 2.5 standard deviations from each speaker's mean on any of the acoustic measures were excluded (excluding F0; 11.1% of the total data). Overall, 8,936 vowels were analyzed, whose distributions into the four groups (non-glottalized non-creaky, non-glottalized creaky, glottalized non-creaky, and glottalized creaky) are shown in Table 3.

Table 1: *Acoustic measures used in the discriminant analysis. Included in the analysis were averages of each measure, as well as changes between the first and final third of the target vowel. Asterisks indicate measures corrected for vowel formants.*

Measure	Explanation
H1*-H2*	Difference in amplitude between H1 & H2
H2*-H4*	Difference in amplitude between H2 & H4
H1*-A1*	Difference in amplitude between H1 & harmonic nearest F1
H1*-A2*	Difference in amplitude between H1 & harmonic nearest F2
H1*-A3*	Difference in amplitude between H1 & harmonic nearest F3
H4*-2K*	Difference in amplitude between H4 & harmonic nearest 2000 Hz
2K*-5K*	Difference in amplitude between Harmonic & nearest 2000 Hz harmonic nearest 5000 Hz
F0	Fundamental frequency
CPP	Cepstral peak prominence
HNR05	Harmonics-to-noise ratio $<$500 Hz
SHR	Subharmonics-to-harmonics ratio

3.2. Linear discriminant analysis

A linear discriminant analysis (LDA), implemented in R [25], was used to investigate the contribution of the acoustic measures to the identification of glottal stops and phrasal creak. Linear discriminant analysis is used to find a linear combination of predictors that results in the best separation of two or more groups.

For each of the eleven acoustic measures, two variables were entered into the analysis as predictors. First, we included the average value of the measure over the vowel for each token. For vowels that were followed by a glottal stop, this value was averaged over the portion of glottalization as well. Second, we included the change in each measure from the first third to the final third of the vowel (or vowel plus glottalization). We expect that the presence of a glottal stop might be better discriminated by the change in acoustic measures over time, rather than the average value over the vowel. All variables were standardized within speaker before being entered into the analysis.

3.2.1. *Classification results*

The LDA produced three functions to discriminate among the four groups (the four combinations of creak and glottalization). The first discriminant function (LD1) accounted for 86.3% of the explained variance; the second (LD2) accounted for 11.4%; and the third accounted for 2.3%. In the following discussion, we ignore the third function, as it contributed relatively little to the analysis.

For each of the four groups, we calculated the mean predicted value of the first two discriminant functions. Figure 2 shows confidence ellipses around these group means. The first discriminant function, on the x-axis, is mainly used to separate [t] from [ʔ]. The second function, on the y-axis, is mainly used to separate creaky from non-creaky tokens. Thus, the LDA was able to identify two orthogonal dimensions within the data: glottalization and phrasal creak. Additionally, the non-creaky tokens (the upper ellipses) were better separated than the creaky tokens (the lower ellipses). This indicates that glottalization was harder to identify in the presence of phrasal creak. The confusion matrix from the LDA is shown in Table 2, and classification scores for each group are shown in Table 3. In general, the model was reasonably successful at identifying glottalization, but was strongly influenced by the prior probabilities of each group.

Table 2: *Confusion matrix from LDA analysis. 'Modal' refers to non-creaky tokens.*

Actual → Predicted ↓	Modal [t]	Creaky [t]	Modal [ʔ]	Creaky [ʔ]
Modal [t]	1820	138	635	88
Creaky [t]	9	4	10	4
Modal [ʔ]	1077	251	4184	613
Creaky [ʔ]	8	13	42	40

Table 3: *Number of tokens per glottalization-creak group, with classification scores for the LDA model based on leave-one-out cross-validation.*

	Modal [t]	Creaky [t]	Modal [ʔ]	Creaky [ʔ]
Count	2914	406	4871	745
Sensitivity	62%	1%	86%	5%
Specificity	86%	99%	52%	99%

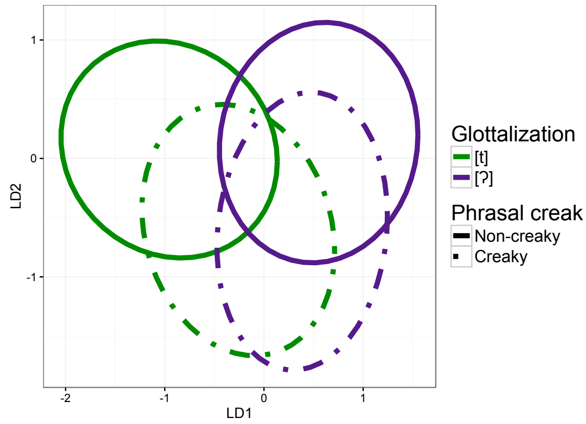

Figure 2: *LD1-LD2 space. The ellipses represent 50% confidence intervals around the mean of each group.*

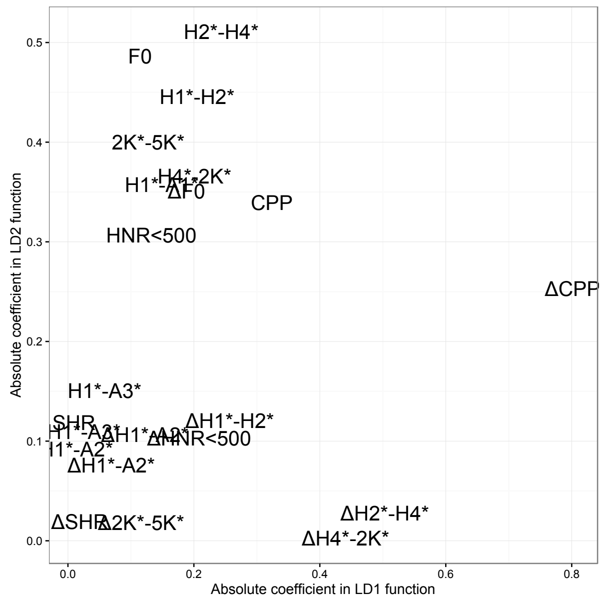

Figure 3: *Absolute coefficients of the acoustic predictors in the first two discriminant functions. Predictors farther to the right side of the chart are more important in LD1, which identifies glottalization (see Figure 2). Predictors farther to the top of the chart are more important in LD2, which identifies phrasal creak.*

3.2.2. *Acoustic measures contributing to discrimination*

Figure 3 shows the coefficient of each acoustic predictor in the first and second discriminant functions. The first discriminant function, which primarily identifies glottalized stops, is most strongly associated with the change in cepstral peak prominence over the duration of the vowel plus glottal stop (ΔCPP). Cepstral peak prominence decreases near the end of the vowel – indicating lower periodicity or greater noise – but especially in the context of a glottal stop, as seen in the left panel of Figure 4. In general, the variables that measured a change over time were more strongly associated with the first discriminant function than the variables that measured average noise or spectral tilt.

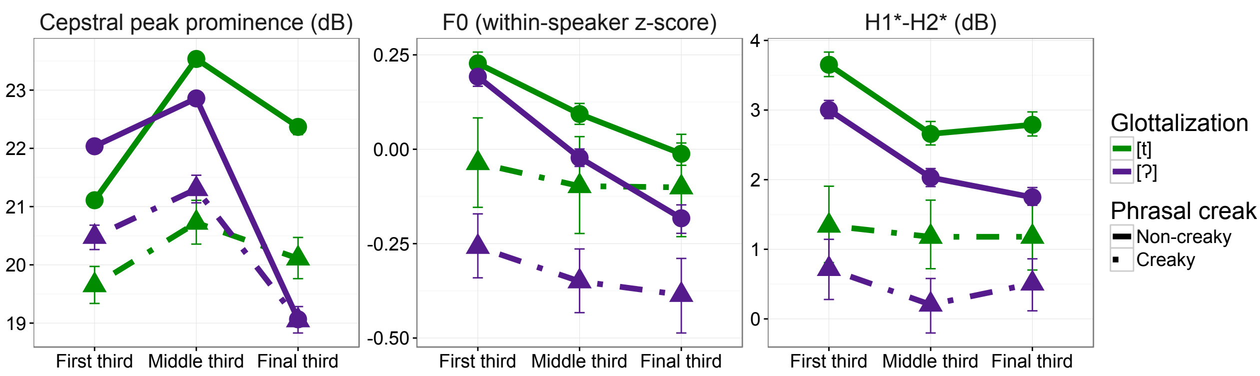

Figure 4: *Changes in cepstral peak prominence (CPP, left panel), F0 (center panel), and H1*-H2* (right panel) over course of vowel, as a function of the four groups. Raw measures are shown for CPP and H1*-H2*, but all measures were standardized within-speaker before being entered into the analysis. Error bars show bootstrapped 95% confidence intervals of the means.*

The second discriminant function, which primarily identifies phrasal creak, is most strongly associated with F0 and spectral tilt measures averaged over the entire vocalic portion. Creaky vowels have overall lower F0 and spectral tilt (e.g., H1*-H2* and H2*-H4*), regardless of whether the /t/ is glottalized to [ʔ] (see center and right panels in Figure 4). While the largest difference in these measures is between tokens with phrasal creak (broken lines) and those without phrasal creak (solid lines), it can also be seen glottalized tokens tend to have lower F0 and spectral tilt relative to non-glottalized tokens. However, the average differences in these measures were not a strong contributor to the discrimination of glottalization (see Figure 3).

4. Discussion

The goal of this study was to determine what factors can be used to reliably identify /t/ glottalization (producing /t/ as [ʔ]) and phrasal creak (prolonged irregular voicing often occurring at edges of prosodic phrases). A discriminant analysis showed that phrasal creak is most robustly identified on the basis of lower F0 and spectral tilt, consistent with previous descriptions [3, 11]. While creaky vs. non-creaky tokens were not well-discriminated in absolute terms, this was most likely due to the relatively smaller number of tokens with phrasal creak in the sample.

Using linear discriminant analysis, we also found glottalized tokens are differentiated by a *rapid change* in noise over time. Previous studies have shown that listeners are sensitive to the average noise characteristics of the voice [26, 27]. Thus, future work should investigate if English listeners use rapid changes in noise to identify glottalization.

Also worth noting is the fact that neither glottalization nor phrasal creak is strongly associated with the subharmonics-to-harmonics ratio (SHR), even though creaky voice is often double-pulses or multiply-pulsed [3, 28]. However, this result does not imply that glottalization and phrasal creak do *not* have multiply-pulsed voicing; rather, the SHR measure is not as well associated with a particular source of creaky voice compared to other noise or the spectral tilt measures.

Lastly, we find that glottalization and phrasal creak differ acoustically in two respects. First, they are primarily (but not solely) associated with different acoustic properties: noise for glottalization, F0 and spectral tilt for creak. Second, although creaky vowels differ from non-creaky ones in terms of average acoustic measures, glottalized vowels differ from non-glottalized ones mainly in terms of a change in acoustic measures. Thus, creaky vowels are overall lower in pitch and spectral tilt, whereas glottalized vowels are characterized by increased noise over time.

As for whether these two linguistic sources of creaky voice differ in their production, the results of this study suggest that both glottalization and phrasal creak are prototypically creaky (Figure 4): both types probably have increased constriction (as seen by their lower spectral tilt; right panel), and both have lower F0 (center panel) and irregular voicing (lower periodicity; left panel) relative to non-glottalized and non-creaky vowels. Therefore, differences in production between glottalization and phrasal creak likely have more to do with timing than with categorically distinct articulations.

5. Conclusions

American English has different linguistics sources of creaky voice. In this study, we investigated how two such sources – /t/ glottalization and phrasal creak – differ acoustically. We do not find evidence that speakers use different sub-types of creaky voice for /t/ glottalization vs. phrasal creak. However, phrasal creak is on average noisier than /t/ glottalization, which in turn shows sharp increases in noise towards the vowel's end. Altogether, these findings suggest that, when distinguishing between different sources of creaky voice, English listeners exploit spectral tilt as well as changes in noise over the vowel's time course.

6. Acknowledgements

For help with corpus annotation, we would like to thank Meagan Rose Baron, Julia Dinwiddie, Hilda Parra, Alexander Wang, and Yushi Zhao. Part of this project was presented at the Spring 2016 meeting of the Acoustical Society of America.

7. References

[1] A. Batliner, S. Burger, B. Johne, and A. Kießling, "MÜSLI: A classifcation scheme for laryngealizations," in *Proceedings of ESCA workshop on prosody*, Lund, 1993, pp. 176–179.

[2] J. Kane, T. Drugman, and C. Gobl, "Improved automatic detection of creak," *Computer Speech and Language*, vol. 27, pp. 1028–1047, 2013.

[3] P. Keating, M. Garellek, and J. Kreiman, "Acoustic properties of different kinds of creaky voice," in *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, 2015.

[4] J. Slifka, "Some physiological correlates to regular and irregular phonation at the end of an utterance," *Journal of Voice*, vol. 20, pp. 171–186, 2006.

[5] J. Pierrehumbert, "Prosodic effects on glottal allophones," in *Vocal fold physiology: voice quality control*, O. Fujimura and M. Hirano, Eds. San Diego: Singular Publishing Group, 1995, pp. 39–60.

[6] M. K. Huffman, "Segmental and prosodic effects on coda glottalization," *Journal of Phonetics*, vol. 33, pp. 335–362, 2005.

[7] M. Sumner and A. G. Samuel, "Perception and representation of regular variation: The case of final /t/," *Journal of Memory and Language*, vol. 52, pp. 322–338, 2005.

[8] D. Eddington and M. Savage, "Where are the moun[ʔə]ns in Utah?" *American Speech*, vol. 87, pp. 336–349, 2012.

[9] J. Kreiman, "Perception of sentence and paragraph boundaries in natural conversation." *Journal of Phonetics*, vol. 10, pp. 163–175, 1982.

[10] L. Redi and S. Shattuck-Hufnagel, "Variation in the realization of glottalization in normal speakers," *Journal of Phonetics*, vol. 29, pp. 407–429, 2001.

[11] M. Garellek, "Voice quality strengthening and glottalization," *Journal of Phonetics*, vol. 45, pp. 106–113, 2014.

[12] R. J. Podesva and P. Callier, "Voice quality and identity," *Annual Review of Applied Linguistics*, vol. 35, pp. 173–194, 2015.

[13] M. Garellek, "Perception of glottalization and phrase-final creak," *Journal of the Acoustical Society of America*, vol. 137, pp. 822–831, 2015.

[14] M. A. Pitt, L. Dilley, , K. Johnson, S. Kiesling, W. Raymond, E. Hume, and E. Fosler-Lussier, *Buckeye Corpus of Conversational Speech (2nd release)*, Department of Psychology, Ohio State University, Columbus, OH, 2007.

[15] S. Seyfarth and M. Garellek, "Coda glottalization in American English," in *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, 2015.

[16] K. Gorman, "Generative phonotactics," Ph.D. dissertation, University of Pennsylvania, 2013.

[17] Y.-L. Shue, P. A. Keating, C. Vicenik, and K. Yu, "VoiceSauce: A program for voice analysis," in *Proceedings of the International Congress of Phonetic Sciences*, Hong Kong, 2011, pp. 1846–1849.

[18] H. Kawahara, A. de Cheveigné, and R. D. Patterson, "An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: revised TEMPO in the STRAIGHT-suite," in *ICSLP-1998*, 1998, p. 0659.

[19] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson, "Acoustic correlates of breathy voice quality," *Journal of Speech and Hearing Research*, vol. 37, pp. 769–778, 1994.

[20] G. de Krom, "A cepstrum-based technique for determining harmonics-to-noise ratio in speech signals," *Journal of Speech and Hearing Research*, vol. 36, pp. 254–266, 1993.

[21] X. Sun, "Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio," in *Proceedings of Acoustics, Speech, and Signal Processing (ICASSP)*, Orlando, FL, 2002, pp. I–333–I–336.

[22] M. Garellek and P. Keating, "The acoustic consequences of phonation and tone interactions in Jalapa Mazatec," *Journal of the International Phonetic Association*, vol. 41, pp. 185–205, 2011.

[23] H. M. Hanson, K. N. Stevens, H.-K. J. Kuo, M. Y. Chen, and J. Slifka, "Towards models of phonation," *Journal of Phonetics*, vol. 29, pp. 451–480, 2001.

[24] M. Iseli, Y.-L. Shue, and A. Alwan, "Age, sex, and vowel dependencies of acoustic measures related to the voice source," *Journal of the Acoustical Society of America*, vol. 121, pp. 2283–2295, 2007.

[25] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, available at http://www.R-project.org 2014.

[26] J. Kreiman and B. R. Gerratt, "Perceptual interaction of the harmonic source and noise in voice," *Journal of the Acoustical Society of America*, vol. 131, pp. 492–500, 2012.

[27] M. Garellek, R. Samlan, B. R. Gerratt, and J. Kreiman, "Modeling the voice source in terms of spectral slopes," *Journal of the Acoustical Society of America*, vol. 139, pp. 1404–1410, 2016.

[28] B. R. Gerratt and J. Kreiman, "Toward a taxonomy of nonmodal phonation," *Journal of Phonetics*, vol. 29, pp. 365–381, 2001.