Voice Quality and Prosody in English

Melissa A. Epstein
University of Maryland Dental School
E-mail: mae001@dental.umaryland.edu

ABSTRACT

Understanding prosodic voice quality variations in English requires both taking into account the various linguistic sources for changes in voice quality and adequately tracking and measuring the range of naturally occurring voice qualities in English. In this study, voice quality variations were tracked using both quantitative measurements and qualitative assessments. Samples were taken from 3 speakers producing variants of a single sentence. Results show there are consistent effects of prominence and phrase boundaries on voice quality. Both prominent words and phrase-initial words display a "tenser" voice quality than their non-prominent and phrase-final counterparts. Phonetic pitch and phonological tone do not have an effect on modal voice quality. There is, however, an increase in non-modal phonation associated with Low boundary tones, but not with low phonetic pitch.

1. INTRODUCTION

Segmental variation in speech is partially attributable to the prosodic location of the segments, such as prosodic domain-final, prosodic domain-initial and accented positions (see [2] and [7] for reviews). Non-segmental voice quality variations in English have also been found in these positions. For example, creak is used to mark the ends of paragraphs, sentences and words [8, 9, 10, 13].

The purpose of this study is to investigate the effects of prosody on voice quality in American English. This study will address the following questions:

1. Is there a correlation between phonetic pitch (F0) and voice quality? And is there an effect of phonological tone on voice quality?
2. Is there an effect of prominence on voice quality?
3. Is there an effect of phrase position on voice quality?

2. METHODS

There are a number of methodological limitations to previous research on prosodic voice quality variations. First, previous studies often only reported on creakiness, and usually tracked creak by taking qualitative assessments of the time amplitude waveform, i.e. by labeling certain types of waveforms as creaky or subtypes of creaky. The use of this technique will not capture small, relative changes in voice quality, resulting in an undersampling of phenomena. Second, many studies did not track and control for all of the factors that could contribute to changes in voice quality, such as allophonic and personal variations. Finally, most studies have focused on declarative sentences. As a result, one cannot deduce if changes in voice quality are due to the presence of a phrase accent or boundary or to the phonological pitch of the accent or boundary.

Consequently, procedures for this study were designed to account for the effects of syntactic variations, personal and segmental variations, and to track small, relative changes in voice quality. Three native speakers of English were analyzed for this study, two women (S1 and S3) and one man (S2). Since this study is exploratory in nature the subject pool was constrained, but still yielded over 1000 samples for analysis.

2.1 Speech materials

This study uses two corpora of sentences to allow for both identifying prosodic voice quality variations and factoring out personal and segmental voice quality variations. First, the "test" corpus identifies the effects of prosodic variations on voice quality. The test corpus sentences are identical except for intonational tune (declarative or interrogative) and location of the accented word. The sentences are:

- **Dagada** gave Bobby doodads.
- Dagada gave Bobby **doodads**.
- **Dagada** gave Bobby doodads?
- Dagada gave Bobby **doodads**?

Words in bold are accented and receive narrow focus. The test sentence is designed so all vowels are surrounded by voiced consonants and so the sentence begins and ends with unstressed syllables.

Second, the "base" corpus supplies a baseline value for each speaker's general pronunciation tendencies for each word in the test corpus. To factor out the segmental and personal voice quality variations within the test sentence each word in the test sentence is normalized against its pronunciation in the base sentence. In the normalized measurements, the value of the difference between the test and base words is divided that by the value of base word. The base corpus consists of the following sentences:

- **Mary** said "Dagada" today.
- **Dana** said "gave" today.
- **Nancy** said "Bobby" today.
- **Peter** said "doodads" today.

Words in bold are accented and receive narrow focus. The base sentences are designed so that each base sentence and each base word are said as identically as possible. Speakers are asked to accent the bolded first name, so the test word of interest has a weaker accent and does not have an extreme

 ISBN 1-876346-48-5

pitch excursion. Note that there is no claim being made for any special status of the voice quality of the baseline values – they only provide a reference value against which the test values can be compared.

2.2 Evaluation of voice quality

This study focuses on the continuum of modal voice qualities that is characterized by changes in the tensions and compressions of the vocal folds. The endpoints of this continuum are tense voice and lax voice. *Tense voice* is associated with high values of adductive tension, medial compression and longitudinal compression of the vocal folds; *lax voice* is associated with the opposite. Tense and lax voice have different characteristic voice source pulse shapes. Tense voice is characterized by skewing and abrupt changes in the shape of the glottal pulse, and is correlated with an increase in the amplitude of the high frequency harmonics in the source and speech spectra. Lax voice is characterized by a sinusoidal voice source pulse, and is associated with attenuation of the high frequency harmonics in the source and speech spectra [11]. Thus, in this study, changes in modal phonation are evaluated by assessing the shape of glottal flow. The terms *creaky* and *breathy* are reserved to describe the non-modal endpoints of the voice quality continuum, and those voice qualities that can be observed directly from the waveforms or spectrogram.

2.3 Recording procedures

Signals were transduced with a high-quality 1.0" Bruel & Kjaer condenser microphone placed 5 cm from the subjects' lips. Signals were sampled at 20 kHz and downsampled to 10 kHz. All data from a subject were collected during a single session. First, the base sentences were recorded. To encourage resetting of the subjects' pitch range, the base sentences were interspersed with random, unrelated sentences that also had a word boldfaced for narrow focus. Next, the test sentences were recorded in 10 blocks of six sentences, with the order of the six sentences randomized within each block. Each test sentence was preceded by a short scenario. To encourage resetting of the subjects' pitch ranges, subjects read aloud one or more short poems following each block of test sentences.

2.4 Assessment of glottal flow

Differentiated glottal flow was obtained by inverse filtering the speech pressure signal. The differentiated glottal flow was then fit with a revised version of the LF (Liljencrants/Fant) model of the glottal pulse, and measurements of the glottal pulse were taken from the model (see Figure 1) [5, 6].

Individual glottal pulses were selected for inverse-filtering and LF-fitting using signal analysis software developed at UCLA's Bureau of Glottal Affairs. Vowel-medial glottal cycles were selected from the vowel of each syllable of each word in the corpus. Vowel-edge cycles were also selected from the sentence-final syllable and from each narrowly-focused syllable to better assess the effects of sentence-final intonation contours and prominence. In the statistical analysis, measurements were averaged across samples from a single syllable. Individual cycles that could not be fit with the LF model or cycles from syllables that did not contain a single characteristic glottal cycle were set aside and labeled as "not LF-fittable". Approximately 10% of the cycles in the corpus could not be fit with the LF model. For details on the data analysis see [4].

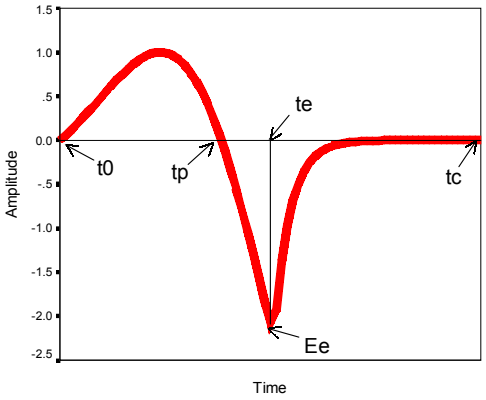

Figure 1. Revised LF model of differentiated glottal flow.

The shape of the glottal pulse was then quantified using three traditional LF measurements and a fourth new measurement of spectral shape. The traditional measurements are (adapted from [11]):

- **EE** (spectral intensity): the amplitude of the negative peak of the differentiated glottal pulse. High values are correlated with an increase in amplitude of all the harmonics in the glottal source spectrum and with a tenser voice quality.
- **RK** (glottal symmetry/skew) $\frac{((te-t0)-(tp-t0))}{(tp-t0)}$ Low values are correlated acoustically with fewer spectral notches and higher amplitude high frequency harmonics, and thus a tenser voice quality.
- **OQ** (open quotient) $\frac{(te-t0)}{(tc-t0)}$ Low values of OQ are correlated acoustically with higher amplitude high frequency harmonics, and thus a tenser voice quality.

The fourth measurement is "**Spectral Linearity**" (Lin). Spectral linearity is calculated in the following manner. First, a line is created by connecting the peaks (dB) across the entire FFT spectrum of a glottal pulse. Then, the value of the squared correlation, r^2, for the regression is calculated. Regression analysis measures how close the line fit to the spectrum is to a straight line; a straight line implies that the glottal spectrum decreases gradually without any strongly attenuated frequencies. The value of spectral linearity is small for lax sinusoidal pulses, because they have very strong low frequency components, but very weak high frequency components. The value of spectral linearity is large for tense glottal pulses that are skewed and/or have a closed phase, because these features amplify high frequencies in the spectrum.

ISBN 1-876346-48-5

2.5 Prosodic labeling

The corpus was prosodically labeled by two trained labelers. The labeling system was closely based on the ToBI (Tones and Break Indices) transcription standard [1] and is called Simple Tones. In Simple Tones there are two tonal categories that are assigned on the basis of the tonal goal of the pattern: Low (low and fall) and High (high and rise). There are also two types of accents. First, "prominent" accents are defined as the "most prominent" pitch accent(s) in a phrase. Second, "boundary" phrase tones collapse together the ToBI intermediate phrase accent and the intonation phrase boundary tone into a single boundary tone defined by the tonal goal of the pattern. Simple Tones was purposely designed for this small, highly constrained corpus and does not adequately handle all intonation features of English. See [4] for details.

2.6 Statistical methods

Statistical analyses were performed to assess the effects of prosody on both modal and non-modal phonation using the non-parametric Kruskal-Wallis test. In review, all modal phonation has been normalized and is fit with the LF model and assessed with LF measurements. Due to the exploratory nature of this study, a high alpha level of 0.2 was chosen for the analysis of modal phonation. Since there were either four or five dependent variables for each analysis, for any one variable significance was reached at $p \leq 0.05$ or $p \leq 0.04$. Results were considered a "trend" if $0.05 < p \leq 0.075$. Correlations were judged to be strong for $r \geq 0.65$. In a separate evaluation, chi-square tests for the effects of prosody on non-modal phonation were considered significant at $p \leq 0.05$. The data used for each factor are:

- Prominent Tone (High/Low) – vowel-medial & -edge for stressed syllables; vowel-medial for unstressed syllables
- Boundary Tone (High/Low) – vowel-medial & edge samples
- Prominence (Prominent/Not) – vowel medial samples
- Phrase Position (Initial/Final) – vowel-medial & -edge for stressed syllables; vowel-medial for unstressed syllables

3. RESULTS

3.1 Phonetic pitch and phonological tone

For phonetic pitch, it was found that normalized F0 does not correlate strongly with any of the normalized LF measurements (see Figure 2 for an example). In other words, as F0 increases or decreases, no LF measurement correspondingly increases or decreases. For the individual speakers, two speakers have the same result. One speaker, on the other hand, has two measurements that correlate strongly with F0: Lin ($r = 0.683$, $p \leq 0.001$) and RK ($r = -0.663$, $p \leq 0.001$).

For phonological tones, it was found that there is not a consistent effect on voice quality, across measurements, across speakers, or when comparing prominent and boundary tones. For example, for prominent accents S1 indicates Low tones have a tenser voice quality, S3 indicates High tones have a tenser voice quality and S2 is inconsistent across measurements.

For non-modal phonation, different effects were found for prominent and boundary tones. For prominent tones, there is no effect of tone type on non-modal phonation (for all speakers: chi-square = 1.603, $p \leq 0.205$). Boundary tones, however, show an effect: Low boundary tones exhibit more non-modal phonation than High boundary tones (for all speakers: chi-square = 45.761, $p \leq 0.001$). This effect also holds for the three speakers individually. In summary, non-modal phonation is associated with Low boundary tones but not with a general lowering of pitch.

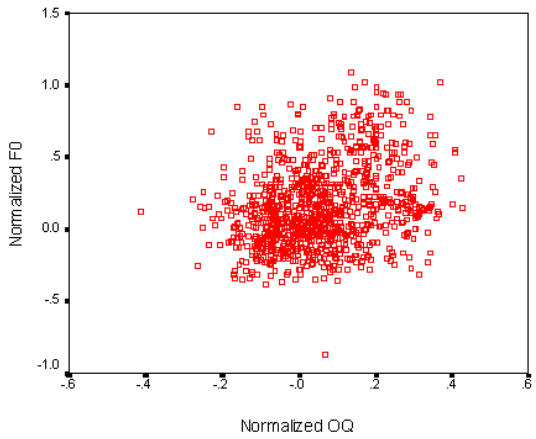

Figure 2. Scatterplot for correlation of normalized F0 and normalizedOQ.

3.2 Prominence

Speakers do use voice quality to distinguish between prominent and non-prominent words. Prominent words have a tenser voice quality than non-prominent words. For all sentence types combined, all speakers show an effect of prominence on voice quality. When looking at interrogatives and declaratives separately, though, between speaker differences are found. For interrogatives, all speakers show this effect to some extent across measurements; for declaratives only two out of three speakers show a clear effect, the third speaker is not consistent across measurements.

Furthermore, in interrogatives, but not declaratives, prominent words are characterized by low values of the parameter OQ. OQ is also used to indicate that Low tones have a tenser voice quality, and interrogatives are characterized by Low prominent tones. This combination of results indicates that a small value of OQ is a marker for Low prominent tones.

Regarding the proportion of not LF-fittable waveforms, the majority occurred in non-prominent words (for all speakers: chi-square = 16.636, $p \leq 0.001$). This distribution indicates that non-vowel-initial creaky waveforms are associated with a lax voice quality, since non-prominent words were found to have laxer LF parameters during modal phonation.

ISBN 1-876346-48-5

3.3 Phrase position

Speakers do use voice quality to distinguish between phrase-initial and phrase-final prominent words. Phrase-initial words have a tenser voice quality than phrase-final words. For all sentence types combined, all speakers show that phrase-initial words have a tenser voice quality. When looking at interrogatives and declaratives separately, though, between speaker differences were found. For declaratives, all speakers show this effect to some extent; for interrogatives only two out of three speakers show a clear effect. These results suggest a weakening of voice quality across the sentence.

Regarding the proportion of not LF-fittable waveforms, the largest proportions occurred at the phrase edges (for all speakers: chi-square = 86.689, $p \leq 0.001$). For all speakers, the phrase-final syllable has increased levels of non-modal phonation and for two speakers the phrase-initial syllable has increased levels of non-modal phonation.

4. DISCUSSION

There is a trend across subjects and across parameters to use differences in modal voice quality to distinguish between prominent and non-prominent words and to distinguish between phrase-initial prominent words and phrase-final prominent words. Subjects are also using non-modal phonation to distinguish between Low boundary tones and High boundary tones.

So, why do speakers use tense voice quality in prominent and phrase-initial positions? The answer may be that both these positions are subject to articulatory and acoustic strengthening. Strengthening can be viewed as an increase in articulatory effort for *particular* articulators. Thus, when a segment is strengthened, not all the speech articulators have an increase in effort, just the ones involved in that particular articulation [7]. So, an increase in effort for the muscles of phonation – the vocalis muscles, the interarytenoid muscles and the lateral cricoarytenoid muscles – would adduct the vocal fold and create a tense voice quality.

Strengthening prominent and phrase-initial vowels with a tense voice quality has a number of advantages. First, an increase in strength aids in lexical access [3]. Vowels with tenser voice qualities acoustically have greater amplitudes of all harmonics in their source spectra, and the high frequency harmonics in particular, causing greater amplification of the formant frequencies. Second, a tense voice quality allows for better pitch discrimination by listeners [13]. Consequently, it is sensible for speakers to use a tense instead of lax voice quality on prominent words in English, because this will better enable listeners to perceive the phonological tonal contrasts that also appear on these words. Prominent words having a laxer voice quality at the end of the sentence can be seen as a form of "weakening".

In conclusion, using the term "non-modal" for variations in voice quality can lead one to believe that the use of these variations is random and undesirable. Instead, it has been found that changes in voice quality occur in predictable positions – at phrase boundaries and on accented words.

ACKNOWLEDGMENTS

I would like to thank Pat Keating, Jody Kreiman, Sun-Ah Jun, Peter Ladefoged and Maureen Stone for their comments on this work. This research was funded in part by NIH grants DC01797, DC01758 and DE07309.

REFERENCES

[1] M. Beckman and G. Ayers, "Guidelines for ToBI labelling," version 3, Ms. The Ohio State University, 1997.

[2] T. Cho, *Effects of Prosody on Articulation*, PhD dissertation, UCLA, 2001.

[3] K.J. de Jong, "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," *JASA*, vol. 97 (1), pp. 491-502, 1995.

[4] M.A. Epstein, *Voice Quality and Prosody in English*, PhD dissertation, UCLA, 2002.

[5] M. Epstein, B. Gabelman, N. Antoñanzas-Barroso, B. Gerratt and J. Kreiman, "Source model adequacy for pathological voice synthesis," *Proceedings of the XIVth International Congress of Phonetic Sciences*, vol. 99 (3), pp. 2049-2052, 1999.

[6] G. Fant, J. Liljencrants and Q. Lin, "A four-parameter model of glottal flow," *Paper presented at the French-Swedish Symposium, Grenoble, France*, 1985.

[7] C. Fougeron, "Prosodically conditioned articulatory variation: A review," *UCLA Working Papers in Phonetics*, vol. 97, pp. 1-74, 1999.

[8] A. Hagen, *Linguistic Functions of Glottalizations and their Lanugage Specific Use in English and German*, PhD dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg and MIT, 1997.

[9] C. Henton and A. Bladon, "Creak as a sociophonetic marker," in *Language, Speech and Mind: Studies in Honor of Victoria A. Fromkin*, L.M. Hyman and C.N. Li, Eds., pp. 3-29. London: Routledge. 1988.

[10] I. Lehiste, "The phonetic structure of paragraphs," in *Structure and Process in Speech Perception*, A. Cohen and S.G. Nooteboom, Eds., pp. 195-203. New York: Springer-Verlag. 1975.

[11] A. Ní Chasaide and C. Gobl, "Voice source variation," in *The Handbook of Phonetic Sciences*, W.J. Hardcastle and J. Laver, Eds., pp. 427-461. Oxford: Blackwell. 1997.

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

[13] D. Silverman, "Pitch discrimination during breathy versus modal phonation," in *Papers in Laboratory Phonology VI*, J. Local, R. Ogden and R. Temple, Eds. Cambridge: Cambridge University Press. 2003.