Rising pitch and quoted speech in everyday American English

Joseph Tyler[1]

[1] Department of English Literature and Linguistics, Qatar University, Doha, Qatar
josephctyler@gmail.com

Abstract

Phonetic variation in rising pitch has been analyzed for how it correlates with contextual factors like speaker gender, utterance type (questions vs. statements) and turn position (turn-medial vs. turn-final). This paper analyzes variation in terminal rising pitch between quoted and non-quoted speech, using data from the Santa Barbara Corpus of Spoken American English. Results show rises in quoted speech start and end higher, rise more overall, have steeper slopes but are no different in duration. These results are gender-dependent, however, for while women produce 65% of all rises in the corpus sample, they produce 100% (n=23) of the quoted speech rises.

Index Terms: rising pitch, intonation, quoted speech, reported speech, gender, uptalk, HRT, corpus analysis

10.21437/SpeechProsody.2014-98

1. Introduction

Terminal rising pitch, also known as uptalk or the High Rising Terminal, has received attention for its semantic and pragmatic meanings [1, 2], its social meanings [3, 4], its perception effects [5, 6] and its production patterns [4, 7-10]. Production studies have used corpora to explore phonetic variation in rises and how such variation correlates with contextual factors like gender, utterance type (questions vs. statements) and turn position (turn-medial vs. turn-final) [11, 12]. Most of these studies used speech from a corpus of map task data [13]. This study makes two novel contributions: first, it uses everyday conversational American English; second, it looks at how rises in quoted speech differ from rises in other contexts.

1.1. Literature Review

Corpus studies of rising pitch in speech production have shown systematic correlations between phonetic dimensions of those rises and contextual factors. The rises have been analyzed phonetically for their starting pitch, ending pitch, rise onset position, rise span, and pitch dynamism [4, 11, 12]. Some studies code phonological representations of a rising pitch contour [14], using a transcription system like ToBI [15], and other studies combine both phonetic and phonological analyses.

This variation in rise form has been correlated with a range of contextual factors, e.g. speaker gender, utterance type (statement vs. question) and turn position (turn-medial vs. turn-final). Results show systematic differences in rise production between men and women: women have been found to produce more rises [4, 9], to start their rises later [3, 4], and to use more pitch dynamism [12]. In addition, results for speakers in Southern California, London and New Zealand suggest that women produce bigger rises than men [3, 4, 9, 12]. This result persists even when using the ERB scale, a scale designed to control for physiologically-determined differences [12].

In addition to gender effects, rises have been found to vary by utterance type, turn position, dialect, age, and a rise's position in the discourse [4]. Other corpus work using different measures of intonation have examined pitch variation by speech register [14] or speech situation [16]. What has not received as much attention is how rises vary by the stylistic implementation of the speech by the speaker. Rampton [17] describes stylization as "reflexive communicative action in which speakers produce specially marked and often exaggerated representations of languages, dialects, and styles that lie outside their own habitual repertoire (at least as this is perceived within the situation at hand)". One such type of stylization may be the performance of quoted speech, especially in the recounting of a narrative.

Quoted speech, also known as direct reported speech, has been studied from a number of perspectives. The intonation of quoted speech was analyzed many years ago by Bolinger [18], who argued that the goal of the speaker in such situations is to "re-enact the original intonation," though the achievement of this goal may be more or less successful. Klewitz and Couper-Kuhlen [19] discuss how quoted speech is sometimes set apart intonationally from the discourse around it, e.g. with prosodic changes at the boundaries between quoted and non-quoted speech. But more often, they report, intonation may mark reported speech but not exactly at the boundaries, serving as a more indirect cue to what is quoted. In addition, a variety of discourse types and data sources have been used to show quoted speech is produced with wider pitch range than non-quoted speech [20-22].

From this work on intonation and quoted speech, it remains unclear how terminal rises are produced in quoted speech compared to rises in other contexts, though the fact that quoted speech often has wider pitch range suggests rises may be larger in quoted speech than elsewhere. The data for many of the corpus studies discussed above have been elicited through map tasks [4, 9, 11, 23, 24], leaving open the question of how such behavior generalizes to more everyday speech contexts. In this study, phonetic features of rises in everyday conversational English are examined and compared in both quoted and non-quoted speech. The results can tell us something about how the quoted speech is realized in rises and how those rises compare to rises more generally.

2. A Corpus Analysis

The spoken data analyzed in this study come from Disc 1 of the Santa Barbara Corpus of Spoken American Spoken English (henceforth SBC). The SBC includes conversations from different regions of the United States, discussing a variety of topics in everyday settings. Some conversations are dyadic and some are multi-party (up to 9 speakers). This corpus sample contains 51 unique speakers across 14 different conversations; except for two speakers who were in two conversations, all speakers were in only one conversation. Recordings lasted an average of 67 minutes, with the shortest and longest being 28 minutes and 120 minutes respectively. They took place in a range of settings, including private settings (living room, kitchen, bedroom, dining room) but also professional contexts (an office, a classroom or a bank).

Speakers' home states were around the US, with more common states being California, Indiana and Illinois, but also Alabama, Montana and Texas. Thirty-one speakers lived in their home state, ten were listed as living in a different state than their home state, and ten had this information missing.

In addition to the naturalness of the data in this corpus, a second important benefit of the SBC is the transcription system used. All conversations have been transcribed into time-stamped intonation units with terminal pitch marked as either rising (marked with a "?"), flat (marked with a ",") or falling (marked with a "."). A simple search for intonation units marked with a "?" reveals hundreds of rises. In addition to the relative ease of identifying the rises, it also supplies a sample of *perceived* rises, regardless of their specific phonetic characteristics.

All rises that the transcript suggested were phonetically analyzable were annotated to a TextGrid in Praat [25]. Cases where the transcript indicated that environmental noise or overlapping speech masked the rising pitch were excluded. This resulted in a set of 636 tokens from Disc 1 of the SBC that were then subjected to phonetic analysis.

The rises were analyzed phonetically using a pitch window of 50-500Hz. The rise domain was annotated from the start of the rise to the end. The rise start was determined impressionistically, though guided by the phonetic trough and with a slight buffer after the syllable onset to reduce segmental effects. The rise start usually occurred on the last stressed syllable of the intonation unit. The rise end was the phonetic peak, i.e. the point at the end of the rise with the highest Hz value.

All tokens were analyzed with a Praat script that extracted the rise-start pitch (in Hz) and the rise-end pitch (in Hz), as well as rise duration information. During the original TextGrid annotation, some tokens did not exhibit a pitch value or showed an unreliable pitch value (e.g. apparent halving or doubling errors). Because these instances were going to result in unreliable output from the Praat script, they were marked for follow-up manual analysis.

The tokens marked for manual measurement were then analyzed individually. During manual measurement, the pitch settings were adjusted according to the author's discretion to control for halving or doubling errors. For some rises, the script outputted Hz values of zero; these tokens were analyzed individually. Outliers were also examined manually; outliers were identified as values at the extremes (<75Hz, >400Hz, and with a span that was >200Hz or negative) or those that were more than two standard deviations away from a speaker's mean. Following the manual analyses, (5%) of the 636 total tokens ended up being excluded because they were unmeasurable. The remaining 603 tokens served as the basis of the corpus analysis. These tokens were produced by 39 unique speakers (see Table 1 for the distributions across speakers). In addition to rise-start and rise-end, variables were also defined for rise-span, defined as rise-end minus rise-start, and rise-slope, defined as rise-span(Hz) divided by duration(seconds).

While the original phonetic measurements were taken on the linear Hz scale, they were converted in two ways for analysis: 1) log-transformation, and 2) conversion to the ERB (Equivalent Rectangular Bandwidth) scale. The log transformations resulted in a more normal distribution. The ERB scale was developed using psychoacoustic tests to track changes in Hz to perceived changes in prominence, and has been argued to be more appropriate for the analysis of intonation [26]. It has been argued to better correspond to listeners' perceived pitch scale. It has also been used in other research on rising pitch [3]. Duration measures were log-transformed to increase the normality of the distribution.

In addition to the phonetic measurements, each rise was coded for various contextual factors, including quoted speech, speaker gender, utterance type (question vs. statement) and turn position (turn-medial vs. turn-final). While part of a larger project, this paper focuses on rises in quoted and non-quoted speech. For this study, quoted speech was coded exclusively as direct reported speech, i.e. where the speech was presented as said exactly as it was said in the past. In each of the following examples, the reported speech ("oh you mean adults" and "oh so you're going to host them are you") is presented as a word-for-word representation of what was said in the original speech context. The original transcription format of the SBC is maintained.

(1) PAMELA: (H) she said,
oh you mean,
... adults=?

(2) ALINA: And I said,
oh.
.. So you're going to host them are you?

As is visible in these two examples, the quoted speech is introduced with the quotative "to say". In this corpus sample, all 23 tokens of quoted speech were introduced with a quotative, 17 with the verb to say (e.g. (1) and (2)), 6 with "to go" (e.g. (3)).

(3) CAROLYN: And they look at] you and they go,
... <Q the what Q>?

While it is not the focus of this study, quotative use has been examined for its systematic variability, e.g. in the context of narrating a life story [27] as well as in language change [28].

Table 1: *Distribution of rises by speaker*

Speaker	Gender	#Rises	#Quoted	Speaker	Gender	#Rises	#Quoted
1	f	65		22	f	2	
2	f	9		23	f	1	
3	f	1		24	m	51	
4	f	19		25	f	13	
5	m	1		26	m	6	
6	m	14		27	f	12	
7	m	12		28	f	12	1
8	f	22	1	29	f	4	
9	m	13		30	m	62	
10	m	17		31	f	1	
11	f	9		32	f	2	
12	f	16	2	33	f	18	1
13	m	1		34	f	12	
14	m	20		35	f	13	
15	f	27	3	36	m	4	
16	f	39	12	37	m	8	
17	f	8		38	m	11	
18	f	33	1	39	m	4	
19	f	35	1	40	m	1	
20	f	19		41	m	3	
21	f	16					

Table 1 lists the number of rises and quoted rises by speaker, showing that all quoted speech tokens were produced by women. Therefore, all results are gender-dependent, for while women produce 65% of all rises in the corpus sample, they produce 100% (n=23) of the quoted speech rises. Moreover, while eight different women in the corpus sample produced measurable rises in quoted speech contexts, a single woman (speaker 16) produced a majority (52%).

2.1. Results

Table 2 shows the distribution of untransformed values for rise-start, rise-end, rise-span, rise-duration, and rise-slope (risespan/duration) in quoted and non-quoted speech.

Table 2. *Descriptive statistics for rise measurements.*

Quoted	N	Rise-start (Hz)	Rise-end (Hz)	Rise-span (Hz)	Rise-dur (sec)	Rise-slope
Min	23	121	160	6	0.105	27
Max	23	361	745	425	0.634	1139
Mean	23	215	320	105	0.290	404
SD	23	56	127	101	0.145	358
Non-Quoted						
Min	580	79	91	-24	0.047	-113
Max	580	350	636	352	1.716	1724
Mean	580	174	224	50	0.333	189
SD	580	48	79	51	0.228	204

The distribution suggests quoted speech rises tend to start higher, end higher, rise more and be steeper. In this table, minimum and maximum rise-span may not correspond to the difference between rise-end and rise-start because they could be the values for different tokens.

In order to test whether quoted speech rises are produced differently from non-quoted rises, mixed models were used, which have been found to have advantages over other models [29]. Random effects for speaker and conversation (the sound file from which each speaker's data are drawn) were included in order to control for variation due to speaker- or conversation-specific differences. The predictor variable was a binary variable for quoted vs. non-quoted speech. The dependent variables were the rise-start, rise-end, rise-span, and rise-slope values on both log(Hz) and ERB scales. Also included was a dependent variable for rise duration (in seconds, log-transformed). These models were implemented using the `lmer` function in R [30]. Because the `lmer` function does not output a p-value, p-values were retrieved with the `pvals.fnc` function.

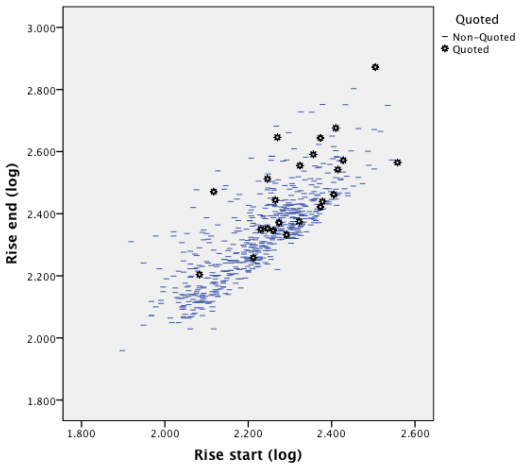
Results in Table 3 show that quoted speech has higher rise-start pitch, rise-end pitch, rise-span, and rise-slope in both Hz and ERBs. No difference was found for rise durations. These results indicate that rises in quoted speech start higher, end higher, rise more and more steeply in this corpus sample. They do not, however, rise for a longer duration. In order to test whether the pitch results depend on duration differences, the same models were run for rise-start, rise-end, and rise-span but with log(rise-duration) as a covariate. In these models, quoted speech remained a significant predictor for all pitch measures, with rise-duration also significantly predicting the rise-end and rise-span measures, but not the rise-start measures. And when data from male speakers are excluded, significant effects for the quoted speech variable remain.

Table 3. *Quoted vs. non-quoted speech rises.*

	Quoted vs. non-quoted speech
Rise-start (log(Hz))	ß=.074, SE=.016, t=4.56, pMCMC<.001
Rise-start (ERB)	ß=.003, SE=.001, t=4.56, pMCMC<.001
Rise-end (log(Hz))	ß=.117, SE=.024, t=4.93, pMCMC<.001
Rise-end (ERB)	ß=.005, SE=.001, t=4.93, pMCMC<.001
Rise-span (log(Hz))	ß=.087, SE=.026, t=3.33, pMCMC=.001
Rise-span (ERB)	ß=.002, SE=.001, t=2.382, pMCMC=.021
Rise-slope (risespan(log) /duration)	ß=.234, SE=.097, t=2.43, pMCMC=.014
Rise-slope (risespanERB /duration)	ß=.007, SE=.003, t=2.073, pMCMC=.042
Rise duration (log(seconds))	ß=-.024, SE=.059, t=-.40, pMCMC=.690

Figure 1 presents results for quoted and non-quoted rise-starts and rise-ends.

Figure 1: *Scatterplot for quoted and non-quoted rises*



While there is overlap in the distributions, this plot shows that quoted speech rises tends to start and end higher in pitch.

Among the rises in quoted speech, some tokens reach the extremes of pitch excursion. For example, on the word "balcony" in the excerpt in (4), the speaker reaches almost 800Hz. The pitch contour for this rise is plotted in Figure 2.

(4) ALINA: of course I said,
.. <VOX ~Cassandra,
... you wanna play one [bounce off the] balcony VOX>?

The domain of the rise is marked in Figure 2 as the portion between the two vertical dotted lines and corresponds to the nucleus of the first syllable to the nucleus of the last syllable

of the word "balcony". This remarkable rise is in the context of narrating a story about someone who was jumping. The narrator is annoyed, and quotes her own speech in a dramatic and excited manner. The speaker's excitement in the story may have resulted in extremely high pitch on the final rise.

Figure 2: *Pitch contour for the utterance "wanna play one bounce off the balcony?" produced by speaker 16 (Alina), with the rise domain indicated by vertical dotted lines.*

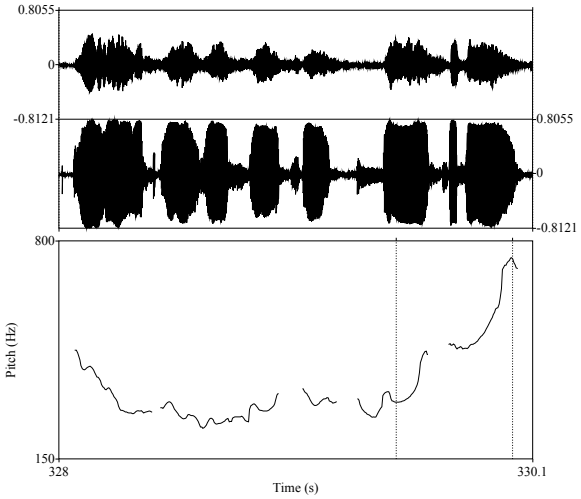
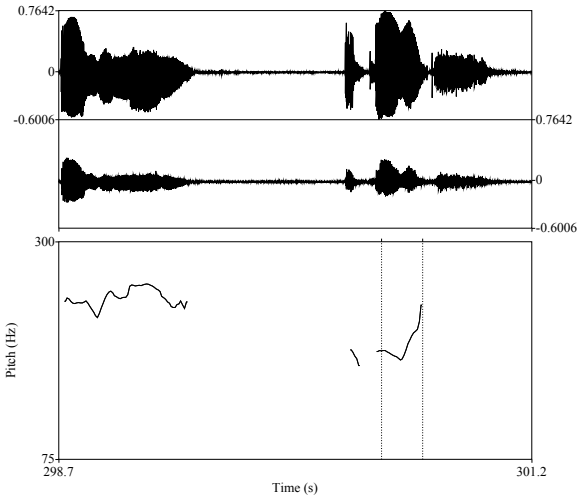


Figure 3: *Pitch contour for the utterance "oh you mean… adults?" produced by speaker 15 (Pamela), with the rise domain indicated by vertical dotted lines.*



Another speaker seemed to exploit pre-rise pauses as part of her repertoire. For example,

(5) PAMELA: (H) she said,
oh you mean,
... adults=?

The domain of the rise covers the onset through the end of the word "adults". As is visible in Figure 3, there is a sizeable pause (794ms) between the words "mean" and "adults". The pitch contour also shows that the speaker drops in pitch before starting the rise. This study has not explored the relationship between the pitch of terminal rises and pre-rise pitch, but (5) suggests that that relationship may display meaningful variation.

While this paper has not specifically addressed distinctions between types of rises other than being quoted or not, one popular contrast is between rises in polar questions and uptalk (often defined as rises on declarative syntax that are not questioning [31]). Uptalk has been derided in popular media [32] and studied in scholarly work [5]. Statistical models were run again with variables for rise duration, questioning speech act and interrogative syntax as covariates (to control for the factors of uptalk). In all of these models, the effects of quoted speech on the pitch measures remain significant. Therefore, the findings reported here for quoted speech are not reducible to an uptalk effect.

3. Discussion

The results discussed above suggest that in the context of quoted speech, (female) speakers produce rises that start higher, end higher, rise more overall and have a steeper rise. These results complement existing research that has determined quoted speech in general to be characterized by higher than average pitch [20-22]. One possible explanation for this is that the act of quoting, i.e. of representing speech from another time, is generally dramatic. Rendering speech from outside the immediate context is a kind of performance. The nature of that performance includes higher, larger and steeper than average rises.

The findings that quoted speech rises are produced with different intonation from non-quoted speech rises may contribute to ongoing discussions on the nature of speech stylization. Given Rampton's [17] definition of stylization referenced above, quoted speech may itself be a case of stylization, where the speech is marked and exaggerated, lying outside the speaker's habitual repertoire. Notably, the speaker could be quoting themselves, as in (4) above. While the original production that the speaker is quoting is not available for comparison, it seems likely the original did not end in a rise that reached almost 800 Hz (if, in fact, she said these words at all). Therefore, her goal may not necessarily be to most accurately reproduce the original speech, as Bolinger [18] had claimed. Instead, the special circumstances of quoting, even when the one being quoted is oneself, can lead a speaker to mark the quoted speech as other. One of the mechanisms by which speakers can mark quoted speech as other is through distinctive intonation. The results discussed above suggest that the size and slope of a rise, in addition to wider pitch range [20-22], can help stylize quoted speech as marked speech.

4. Acknowledgements

The author would like to thank Paul Warren for sharing references and Irene Theodoropoulou for input.

5. References

[1] Lai, C., "Rises all the way up: The interpretation of prosody, discourse attitudes and dialogue structure," Linguistics, University of Pennsylvania, 2012.

[2] Gunlogson, C., "A Question of Commitment," *Belgian Journal of Linguistics,* vol. 22, pp. 101-136, 2008.

[3] Warren, P. and Daly, N., "Sex as a factor in rises in New Zealand English," in *Gendered speech in social context: Perspectives from gown and town*, J. Holmes, Ed., ed Wellington: Victoria University Press, 2000, pp. 99-115.

[4] Ritchart, A. and Arvaniti, A., "Do We All Speak Like Valley Girls? Uptalk in Southern Californian English," presented at the Acoustical Society of America, San Francisco, 2013.

[5] Tomlinson Jr, J. M. and Fox Tree, J. E., "Listeners' comprehension of uptalk in spontaneous speech," *Cognition,* vol. 119, pp. 58-69, 2011.

[6] Tyler, J., "Discourse Prosody in Production and Perception," PhD, Linguistics, University of Michigan, Ann Arbor, 2012.

[7] Fletcher, J., Grabe, E., and Warren, P., "Intonational variation in four dialects of {E}nglish: the high rising tune," in *Prosodic typology: The Phonology of Intonation and Phrasing*, S. Jun, Ed., ed: Oxford University Press, 2005.

[8] Warren, P., "Patterns of late rising in New Zealand English," *Language Variation and Change,* vol. 17, 2005.

[9] Barry, A. S. and Arvaniti, A., ""Uptalk" in Southern Californian and London English," in *BAAP 2006 Colloquium*, ed. Queen Margaret University College, Edinburgh, 2006.

[10] Guy, G., Horvath, B., Vonwiller, J., Daisley, E., and Rogers, I., "An Intonational Change in Progress in Australian English," *Language in Society,* vol. 15, pp. 23-51, 1986.

[11] Fletcher, J. and Harrington, J., "High-Rising Terminals and Fall-Rise Tunes in Australian English," *Phonetica,* vol. 58, pp. 215-229, 2001.

[12] Daly, N. and Warren, P., "Pitching it differently in New Zealand English: Speaker sex and intonation patterns," *Journal of Sociolinguistics,* vol. 5, pp. 85-96, 2001.

[13] Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., *et al.*, "The Hcrc Map Task Corpus," *Language and Speech,* vol. 34, pp. 351-366, 10 1991.

[14] Carmichael, L. M., "Situation-Based Intonation Pattern Distribution in a Corpus of American English," PhD, Linguistics, University of Washington, 2005.

[15] Silverman, K. E. A., Beckman, M., Pierrehumbert, J., Ostendorf, M., Wightman, C. W. S., and Price, P., "ToBI: A standard scheme for labeling prosody," in *Proceedings of the 2nd International Conference on Spoken Language Processing*, Banff, Canada, 1992, pp. 867-879.

[16] Cheng, W. and Warren, M., "// / CAN i help you //: The use of rise and rise-fall tones in the Hong Kong Corpus of Spoken English," *International Journal of Corpus Linguistics,* vol. 10, pp. 85-107, 2005.

[17] Rampton, B., "Interaction ritual and not just artful performance in crossing and stylization," *Language in Society,* p. 149, 2009.

[18] Bolinger, D. L., "The intonation of quoted questions," *Quarterly Journal of Speech,* vol. 32, pp. 197-202, 1946/04/01 1946.

[19] Klewitz, G. and Couper-Kuhlen, E., "Quote-unquote. The role of prosody in the contextualization of reported speech sequences," *Pragmatics,* vol. 9, pp. 459-485, 1999.

[20] Wennerstrom, A., *The Music of Everyday Speech:Prosody and Discourse Analysis*: Oxford University Press, 2001.

[21] Jansen, W., Gregory, M. L., and Brenier, J. M., "Prosodic correlates of directly reported speech: Evidence from conversational speech," presented at the ITRW on Prosody in Speech Recognition and Understanding, Molly Pitcher Inn, Red Bank, NJ, USA, 2001.

[22] Hirschberg, J. and Grosz, B., "Intonational Features of Local and Global Discourse Structure," presented at the Proceedings of the Speech and Natural Language Workshop, 1992.

[23] Fletcher, J., Wales, R. J., Stirling, L. F., and Mushin, I. M., "A dialogue act analysis of rises in Australian English map task dialogues," in *Speech prosody 2002: Proceedings of the 1st international conference on speech prosody*, ed Aix-en-Provence: Universite de Provence, 2002, pp. 299-302.

[24] Shobbrook, K. and House, J., "High Rising Tones in Southern British English," in *15th International Congress of Phonetic Sciences*, Barcelona, 2003, pp. 1273 - 1276.

[25] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer," ed, 2009.

[26] Hermes, D. J. and van Gestel, J. C., "The frequency scale of speech intonation," *J Acoust Soc Am,* vol. 90, pp. 97-102, Jul 1991.

[27] Keller-Cohen, D. and Gordon, C., ""On Trial": Metaphor in Telling the Life Story," *Narrative Inquiry,* vol. 13, pp. 1-40, 2003.

[28] Cukor-Avila, P., "She say, She go, She be like: Verbs of Quotation over Time in African American Vernacular English," *American Speech,* vol. 77, pp. 3-31, 2002.

[29] Baayen, R. H., Davidson, D. J., and Bates, D. M., "Mixed-effects modeling with crossed random effects for subjects and items," *Journal of Memory and Language,* vol. 59, pp. 390-412, 2008.

[30] R Development Core Team, "R: A language and environment for statistical computing," ed. Vienna, Austria: R Foundation for Statistical Computing, 2012.

[31] Cruttenden, A., *Intonation*. Cambridge [U.K.] ; New York, NY, USA: Cambridge University Press, 1997.

[32] Davis, H. (2010, October 6) The Uptalk Epidemic: Can you say something without turning it into a question? *Psychology Today*. Available: http://www.psychologytoday.com/blog/caveman-logic/201010/the-uptalk-epidemic