THE PHONETICS AND DISTRIBUTION OF NON-QUESTION RISES IN TWO VARIETIES OF AMERICAN ENGLISH

Meghan E. Armstrong[1], Page E. Piccinini[2], Amanda Ritchart[2]

[1]UMass Amherst, [2]UC San Diego
armstrong@umass.edu, ppiccinini@ucsd.edu, aritchart@ucsd.edu

ABSTRACT

Utterance-final rises that do not convey questions have been of interest to linguists for years, for many varieties of English. Though a popular stigma exists in the U.S. that this phenomenon is typical of Southern Californian (SoCal) speakers, no study has compared IP-final rise use in U.S. varieties in terms of frequency or phonetic realization. We examine IP-final rises in two dialects of American English (SoCal and Massachusetts) produced in narratives. While variation in phonetic realization was found, there were no differences in the frequency of rises across region or gender. SoCal females produced the longest rises, while females from both regions produced rises with steeper slopes than their male counterparts. Thus while two geographically distant varieties are similar with respect to frequency, sociolinguistic variation is still found. We speculate that phonetic differences in rises produced by SoCal females may contribute to the stigma that *uptalk* is "Valley Girl speak".

Keywords: intonation; prosody; variation; sociophonetics; high rising terminals

1. INTRODUCTION

The use of rising pitch movement at the end of intonational phrases (IPs) on declaratives is strikingly common across varieties of English (i.e., in Australia, New Zealand, Falkland Islands, Canada, the U.K. and U.S.) [16]. The actual shape of these IP-final rises varies from dialect to dialect, as do the pragmatic functions of the rises [7, 11, 16, 17]. Within American English (henceforth AmEng), the variety spoken in Southern California (henceforth SoCal), is stereotypically known to exhibit this phenomenon [13]. Researchers have referred to these rises as high rising terminals (HRTs), but for AmEng the phenomenon is commonly known as *uptalk* [6]. Using a map task as well as a retelling of a short video clip, [13] investigated the use of rises in SoCal English, finding an effect for gender such that female speakers used more IP-final rises than male speakers overall. An interaction for discourse function was also found: while there was no gender effect found for statements, females used IP-final rises more for holding the floor than male speakers did. Still earlier work has documented both production and perception of IP-final rises in other U.S. varieties (see [1] for SoCal AmEng and [4] for Southern and Midland AmEng), but to date there has been no systematic comparison of regional varieties of AmEng. Thus, although the phenomenon is stereotypically associated with SoCal speech, it is not clear whether SoCal speakers actually produce more IP-final rises than speakers of other regional varieties of AmEng.

In order to work towards establishing a clearer picture of IP-final rise use in AmEng, we compared speakers from the Southwestern corner of the U.S. (SoCal) with speakers from the Northeastern corner (Massachusetts, henceforth Mass). These regional varieties are known to differ substantially in terms of segmental features [10], which gives us reason to hypothesize that we might find intonational differences between the two groups as well; in this case, in their use of IP-final rises. Here we compare the speech of 16 speakers (8 SoCal, 8 Mass) as they told a short story of a personal experience, i.e., a narrative [9]. In a comparison of IP-final rises comparing four different speech styles in Australian English, [8] found that rises were overwhelmingly favored in narratives. They suggest that this is because listener involvement and plot developing are vital in narratives. Through the IP-final rise, the authors argue that Australian English speakers are able to check for listener understanding about what is being said during the narrative, and thus verify their interlocutor's comprehension. Since speakers are not guided by specific instructions in narratives the way they are in map tasks, for example, we might consider narrative speech to be more natural than other tasks used to examine uptalk in AmEng to date. We therefore saw narratives as a fruitful discourse style to operationalize in order to compare the two dialects in question.

Thus, given prior findings on IP-final rise use in other English varieties, we ask: (1) In narrative speech, are there effects of region and gender in terms of the frequency of IP-final rises in utterances that are not questions? (2) Are there effects of region and gender in terms of the phonetic realization of the rises?

2. METHODS

2.1. Speakers

Sixteen total speakers were recorded: 8 SoCal speakers (4 female, 4 male) and 8 Mass speakers (4 female, 4 male). All speakers were native speakers from their respective region.

2.2. Tasks and Procedures

For all speakers, recordings took place in a soundproof booth using a 48 kHz sampling rate and 16-bit quantization. SoCal recordings took place at the University of California, San Diego, and Mass recordings took place at the University of Massachusetts, Amherst. Two tasks were recorded for each speaker: a narrative and giving directions. Only data from the narratives are reported here. In the narrative task, speakers were asked to tell the experimenter a story about something that had happened to the speaker, for example a childhood story or something that had happened to the speaker over the weekend.

The mean duration of the narratives for SoCal speakers was 1.67 minutes (SD: 0.74 minutes); the mean duration for the narratives for Mass speakers was 1.46 minutes (SD: 0.73 minutes). The experimenter was always female.

2.3. Annotations and Measurements

Using Praat's TextGrid annotation tool [3], IPs were annotated by two transcribers for each recording. Utterances roughly correlated with break index 4 according to the MAE_ToBI labeling conventions for break strength [2]. For each utterance the transcriber also noted the utterance contour shape (rising, level, or falling), the utterance discourse type (non-question, question, or non-question filler), and the "story mode" of the speaker (in-story mode or out-of-story mode). Contour types were labeled based on F0 movement between the nuclear stressed syllable and the end of the utterance: rising F0 between these two points was labeled a "rise", level/sustained pitch was labeled "level", and falling pitch between the two points was labeled "falling". Both pitch tracking in Praat and auditory perception were used to assign the categories for contour shape. Utterances were considered questions if they could plausibly be answered with a "yes" or "no" response, or if they were *wh*-questions, syntactically marked with the relevant *wh*-word. All other utterances were considered non-questions. We also included information about whether or not the utterance was a discourse filler. Non-question fillers were IPs that were either strictly a filler word (e.g.,

um, uh), a conjunction (e.g., *and, but*) or the combination of the two (e.g., *but um…*). Out-of-story mode was defined as any instance in which the speaker was not telling the narrative, e.g., if the speaker was asking the experimenter a question about the task or answering brief follow-up questions by the experimenter. Any conflicts in annotation were resolved by a third transcriber. Additionally, for any utterance ending in a rise, the third transcriber marked the beginning of the rise (minimum F0 of rise) to the peak of the rise (maximum F0 of rise).

For each IP-final rise, several measurements were taken: (1) the duration of the rise from the onset of the rise until the peak, (2) the F0 at the onset of the rise (the F0 minimum), and (3) the F0 at the peak of the rise (the F0 maximum). All F0 measurements were transformed to ERB (ERB = 21.4 * log10(Hz * 0.00437 + 1)) to account for gender differences [5]. Based on these measurements, two additional variables were computed: (1) pitch excursion (defined as the difference between the F0 maximum and F0 minimum in ERB), and (2) slope of the rise (defined as pitch excursion divided by the duration of the rise from minimum F0 to rise peak). Examples of the labeled rises are given in Figures 1 and 2. The onset of the rise and the rise peak are marked as *onset* and *peak,* respectively.

2.4. Statistical analyses

To test for significant effects, logistic and linear mixed effects models (LMEMs) were run. Due to data scarcity, questions were not included in any of the models (see Table 1 for a summary of number of tokens), nor were utterances produced in out-of-story mode, since they were not considered part of the narrative. The logistic regressions were used to test whether there was a difference in frequency of rises based on region and gender. Two models were run, one with the dependent variable comparing usage of final rises versus final level pitch, and one with the dependent variable comparing usage of final rises versus final falls. For both models the fixed effects were: (1) region (SoCal, Mass), (2) gender (female, male), and (3) utterance type (non-filler, filler). An interaction of region and gender was included in the model. Speaker was included as a random intercept. This was the maximal, uncorrelated random-effects structure that converged. The models did not converge with random slopes.

Linear regressions were used to test whether there were differences in duration and slope for rises based on region and gender. Two models were run, one with the dependent variable as rise duration in

milliseconds, and one with the dependent variable as rise slope in ERB per millisecond. Due to sparseness of data, only non-filler rises were examined (see Table 1). For both models the fixed effects were: (1) region (SoCal, Mass), and (2) gender (female, male). An interaction of region and gender was included in the models. Speaker was included as a random intercept. This was the maximal, uncorrelated random-effects structure that converged.

For all models, significance of fixed effects was assessed using model comparison. Alpha was set at $p < 0.05$.

Table 1: Number of tokens per each variable. Percentages per group are in parentheses.

Utterance Type	Rise	Level	Fall	TOTAL
Non-question	249 (32%)	290 (37%)	243 (31%)	782 (100%)
Filler	14 (7%)	152 (73%)	41 (20%)	207 (100%)
Question	4 (40%)	2 (20%)	4 (40%)	10 (100%)
Sex	Rise	Level	Fall	TOTAL
Female	181 (30%)	258 (42%)	171 (28%)	610 (100%)
Male	86 (22%)	186 (48%)	117 (30%)	389 (100%)
Region	Rise	Level	Fall	TOTAL
SoCal	143 (27%)	233 (43%)	159 (30%)	535 (100%)
Mass	124 (27%)	211 (45%)	129 (28%)	464 (100%)
Story Mode	Rise	Level	Fall	TOTAL
In story	245 (28%)	403 (46%)	228 (26%)	876 (100%)
Out of story	22 (18%)	41 (33%)	60 (49%)	123 (100%)

3. RESULTS

3.1. Frequency of rises

For the model examining use of rising versus level IP-final pitch movements, level IP-final pitch movements were significantly more frequent than rising ones, as found by a significant intercept [β = -1.28, $SE = 0.22$, $\chi^2(1) = 20.84$, $p < 0.001$]. The only fixed effect that was significant was utterance type [$\beta = -2.32$, $SE = 0.32$, $\chi^2(1) = 78.03$, $p < 0.001$]. In this model, level IP-final pitch movements were used for fillers more often than rises. For the model examining use of rising versus falling IP-final pitch movements, only utterance type was significant [β =

-1.33, $SE = 0.39$, $\chi^2(1) = 12.67$, $p < 0.001$]. For this model, falls were used more frequently than rises to end filler utterances. A summary of tokens per fixed effect is given in Table 1.

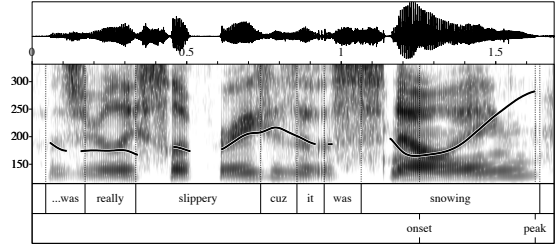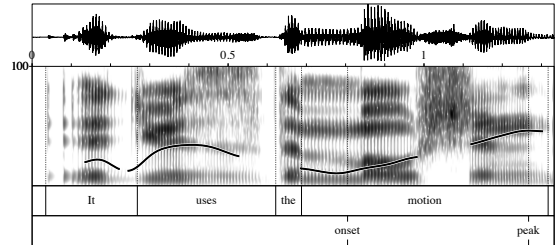**Figure 1**: Non-question rise produced by a Mass female speaker.

Figure 2: Non-question rise produced by a SoCal male speaker.

3.2. Duration and slope of rises

For the model examining rise duration, the fixed effects alone were not significant, but there was a significant interaction of region and gender [β = 172.93, $SE = 73.95$, $\chi^2(1) = 4.76$, $p < 0.05$]. Follow-up simple regressions found that, for SoCal speakers, females had significantly longer rises than males [$r = -0.30$, $p < 0.001$], but there was no gender difference for Mass speakers [$r = 0.07$, $n.s.$]. See Figure 3 for summary of results.

Figure 3: The durations of rises in milliseconds separated by region (SoCal, Mass) and gender (female, male).
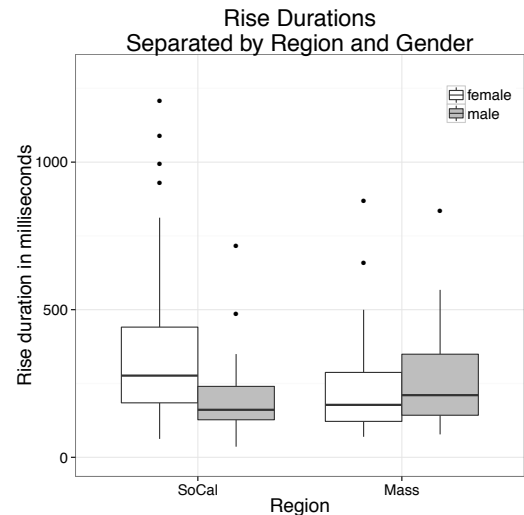
For the model examining rise slope, females had significantly steeper rises than males [β = -0.003, *SE* = 0.0008, $\chi^2(1)$ = 9.32, *p* < 0.01]. Region and the interaction of region and gender were not significant.

Additionally we tested how correlated rise duration and rise slope were in order to test whether our two quantitative dependent variables were truly measuring different aspects of rises. A simple linear regression found the two variables to be highly negatively correlated, with a longer duration resulting in a shallower slope [*r* = -0.30, *p* < 0.001]. Closer inspection of the data reveals that the steepest slopes had shorter durations, and shallow rises had very long durations. Of these, SoCal females produced the steepest (short) rises as well as the longest (shallow) rises. We therefore believe both measures are important for understanding rises as produced by males and females in different varieties of American English.

4. GENERAL DISCUSSION

Contrary to the popular stereotype that females use rises more frequently than males, our data show no significant effect of gender for rise use. In fact, the most frequent contour shape was a level contour, which emerged as the most frequent shape for both males and females. Further research is needed to assess whether this finding is task-specific. While rises were relatively frequent for non-questions, they were infrequently used for fillers. Fillers were most often produced with a level contour. Thus, it is possible that the pragmatic information contributed through rises is not congruent with the discourse markers or connectives that we observed in these narratives.

Our results are interesting in light of the stereotypes and assumptions made about IP-final rises, especially the folk ideas about *uptalk* that often surface in the (American) media. Not only are IP-final rises not restricted to females from Southern California, but there are no significant differences in the frequency of non-question rises between regions. Our results suggest, at least for narratives, that college students in the Southwest and Northeast pockets of the U.S. use IP-final rises at similar rates. Future work might explore whether this holds true across discourse styles. However, our results confirm [8]'s findings that non-question rises are common in narratives, this time for two varieties of AmEng. In any case, our results do not provide evidence for IP-final rises as a phenomenon restricted to "Valley Girls".

With respect to our phonetic analysis, we find a gender difference in SoCal English not found in Mass: the duration of rises produced by females is longer than that of men. [14] have argued that IP-final rises with "prolonged" pitch tend to be processed with a forward-looking discourse function, while "unprolonged" rises have a backward-looking function. The listeners in [14]'s study were students at a university in Northern California, but no information is given on the listeners' origins. It is not clear whether the differences we found for rise duration is related to differences in pragmatic choices made by SoCal females (i.e., more forward-looking utterances), or whether this is simply a phonetic innovation. Perception experiments would be necessary to test how listeners perceive the longer versus shorter rises in our data in terms of their pragmatic functions, and whether duration would influence their perception of rise meaning.

We also speculate that the stereotype regarding the prevalence of rises in "Valley Girl" speech may be related to the fact that SoCal females appear to exploit two phonetic aspects of rises: SoCal females produce rises that are very short and steep, but they also produce the longest rises, and when they do, these are quite shallow in slope. Though further work is necessary to assess and define an "intonational gesture space", one hypothesis is that SoCal females make greater use of this space. Our results suggest that both slope and duration should be considered when modeling this space. We also propose that the general stereotype about females and IP-final rise use could be related to the distinct phonetic implementations we have reported here. Previous literature has shown a positive correlation between high pitch and fast speaking rate. [15] found that faster speech rate seems to be a secondary cue to interrogativity in rises for some languages. [12] also found that utterances manipulated for pitch but not duration were rated as sounding faster when listeners heard higher pitch. Since SoCal females had steeper slopes and in turn higher rises, it is possible they might be perceived as "talking in questions" more frequently. Further tests are necessary to observe speech rate and perceived speech rate.

While we have not yet characterized differences between rises under a phonological framework, our data show that phonetic variation in rises is present in AmEng, just as it is in other macrodialects of English. Unlike other English varieties, however, there is a great paucity in research on dialectal intonational variation in AmEng. Thus, our methods and findings postulate new questions and provide directions for research in this area.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] Barry, A. S. 2007. *The form, function and distribution of high rising intonation in Southern California and Southern British English*. Ph.D. thesis, University of Sheffield, Sheffield, UK.

[2] Beckman, M., Hirchberg, J., Shattuck-Hufnagel, S. 2006. The original ToBI system and the evolution of the ToBI framework In: Jun, S. (ed), *Prosodic Typology: The Phonology of Intonation and Phrasing*. New York: Oxford University Press, 9–54.

[3] Boersma, P., Weenink, D. 2009. Praat: doing phonetics by computer [Computer program]. Version 5.4.04, retrieved 28 December 2014 from http://www.praat.org/.

[4] Clopper, C. G., Smiljanic, R. 2005. Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics* 39(20), 237–245.

[5] Glasberg, B. R., Moore, B. 1990. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1-2), 103–138.

[6] Gorman, J. 1993. Like, Uptalk? *New York Times*.

[7] Gussenhoven, C. 1983. *A semantic analysis of the nuclear tones of English*. Bloomington: Indiana University Linguistics Club.

[8] Guy, G., Horvath, B., Vonwiller, J., Daisley, E., Rogers, I. 1986. An intonational change in Australian English. *Language in Society* 15(1), 23–51.

[9] Labov, W., Waletzky, J. 1967. Narrative analysis. In: Helm, J. (ed), *Essays on the Verbal and Visual Arts*. Seattle & London: University of Washington Press, 12–44.

[10] Labov, W., Ash, S., Boberg, C. 2005. *Atlas of North American English*. New York: Mouton de Gruyter.

[11] de Pijper, J. R. 1983. *Modelling British English intonation: An analysis by resynthesis of British English intonation*. The Hague: Mouton.

[12] Rietveld, A.C.M., Gussenhoven, C. 1987. Perceived speech rate and intonation. *Journal of Phonetics* 15, 273-285.

[13] Ritchart, A., Arvaniti, A. 2014. The form and use of uptalk in Southern Californian English. *Proc. of Speech Prosody 7* Dublin, 331–335.

[14] Tomlinson, J. M. Jr., Fox Tree, J. E. 2011. Listeners' comprehension of uptalk in spontaneous speech. *Cognition* 119(1), 58–69.

[15] van Heuven, H., van Zanten, E. 2005. Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication* 47, 87-99.

[16] Warren, P. 2005. Patterns of late rising in New Zealand: Intonational variation or intonational change? *Language Variation and Change*, 17(2), 209–230.

[17] Watt, D. L. E. 1994. *The phonology and semology of intonation in English: An instrumental and systemic perspective*. Bloomington: Indiana University Linguistics Club.