THE IMPLEMENTATION OF PHRASAL PROSODY BY NATIVE AND NON-NATIVE SPEAKERS OF ENGLISH: SS ANOVA FOR MULTI-SYLLABIC INTONATION CONTOURS

Tuuli H. Morrill

George Mason University, Fairfax, Virginia, U.S.A
tmorrill@gmu.edu

ABSTRACT

Non-native speakers often have difficulties with prosody; stress and intonation patterns that differ from those of native speakers can contribute to "foreign accent," even at high proficiency levels. Although effects of a listener's native language on the perception of prosody are well established, few studies have examined non-native prosody production. In particular, it is not known whether the placement of prominence in words and phrases differs between native and non-native speakers in reliable ways. The current study examines a corpus of native and non-native English speech, to look for group differences in the production of intonation contours.

Using smoothing spline analysis of variance (SS ANOVA), the phrasal pitch contours of identical speech passages produced by 197 speakers of four different native languages are compared. Results suggest that a speaker's native language has predictable effects on the production of phrasal intonation patterns.

Keywords: intonation, prosody, stress, non-native production, ssanova

1. INTRODUCTION

Prosodic differences across languages pose problems for second language learners and non-native speakers. In languages like English, stress patterns can be used by native speakers for word segmentation and lexical access; non-native speakers may mis-perceive the locations of word boundaries, affecting processing. Speakers of languages which do not have lexical stress may even be unable to perceive the phonetic cues of stress [1]. Non-native speakers are also described as *producing* stress patterns incorrectly, or using atypical phrasal intonation contours. However, there have been few investigations of non-native prosody production, and it is not clear whether word- and phrase-level prosodic patterning differs between native and non-native speakers in reliable ways.

There has been more recent work on perception, where research has shown the effects of native language are somewhat predictable. For example, native speakers of English expect to hear stressed syllables at the beginnings of words; thus, they parse words in a foreign language according to these patterns [2]. In addition, the global pitch and timing patterns of utterances affect perception – phrase-final lengthening and breaks in intonation contours are perceived as demarcating words. Whereas a French speaker might associate a rising pitch with a phrase-final syllable, an English speaker could interpret this as a continuing utterance. These effects of prosodic transfer suggest that speakers of languages that differ from English in their word- or phrase-level prosody may implement prosodic patterns that reflect those of their native language.

The complex nature of prosodic analysis has no doubt contributed to the scarcity of non-native prosody production studies. Prosodic labeling systems, such as the Tones and Break Indices system (ToBI) [3], can be used to transcribe pitch accents and intonation contours; however, transcription is labor-intensive, and conventions necessarily differ from language to language. Meanwhile, the acoustic correlates of stress and accent are highly variable, making objective or automatic extraction of prosodic information difficult. Recently, researchers have begun to develop novel approaches to this problem, employing statistical techniques to quantify differences in prosodic patterns. For example, generalized additive models have been used to examine the "branching" of pitch contours in compounds [4] and smoothing spline analysis of variance (SS ANOVA) has been used to examine pitch contours of syllables and words in tone languages [5,6]. SS ANOVA is used for the comparison of curves along multiple reference points [7]; thus, it may also prove an effective method for examining contours over longer time-scales, such as phrases.

2. CURRENT STUDY

This study examines intonation contours produced by native and non-native speakers of English. English intonation contours are typically shaped by tonal targets on accented syllables and at phrase boundaries, as well as the interpolated pitch changes

between them [8]. Thus, intonation contours produced by non-native speakers may be shaped differently if (1) accented syllables occur at different times for native and non-native speakers, and (2) high and low boundary tones are used differently by native and non-native speakers. If accent timing and boundary tones are influenced by native language, intonation contours produced by speakers of the same native language should exhibit regularities. Examining pitch contours with SS ANOVA is expected to reveal significant differences in the shapes of the contours produced by the different language groups.

3. METHODS

3.1. Materials

The corpus consists of 591 utterances taken from the publically available George Mason University Speech Accent Archive [9]. All speakers (n = 197) read an identical passage (the "Stella" passage) designed to elicit the phoneme inventory of English. The passage contains declarative sentences, sentences containing clauses, and lists of items.

The current study includes 55 native speakers of American English, 50 native speakers of Mandarin, 48 native speakers of Arabic, and 44 native speakers of Korean. These languages were selected because they are prosodically distinct from English, and contain large speaker populations who study and speak English.

3.2. Analysis

Three phrases from the Stella passage are included in this analysis. The pitch (F0) contour was extracted from each phrases using the Praat auto-correlation algorithm [10]; the range for extraction was set at 50-450Hz. Gaps in the contour were interpolated from the points on either side of the gap, and artifacts were removed by smoothing with a bandwith of 5Hz. The contour was then divided into 1000 equally spaced time points from which an F0 value was extracted. F0 values were transformed to semitones relative to 1Hz and normalized by speaker to allow for comparison across speakers. SS ANOVA was implemented with the gss package in R [11] and F0 contours were modeled with 95% Bayesian confidence intervals.

4. RESULTS

Figure 1-3 show modeled pitch contours of three phrases. Only tokens in which speakers produced the entire phrase without pauses or disfluencies are included in the analysis. In the figures, where there is no overlap between two contours (places where the white or gray ribbons do not touch), the language groups have produced distinct patterns. Alignment with the plot text is intended to be only approximate.

4.1. List intonation: "six spoons of fresh snow peas"

This phrase occurs at the beginning of a list of items. The modeled F0 contours contain four notable patterns (Figure 1).

Figure 1: **Modeled F0 contours with 95% confidence intervals for the phrase *six spoons of fresh snow peas***

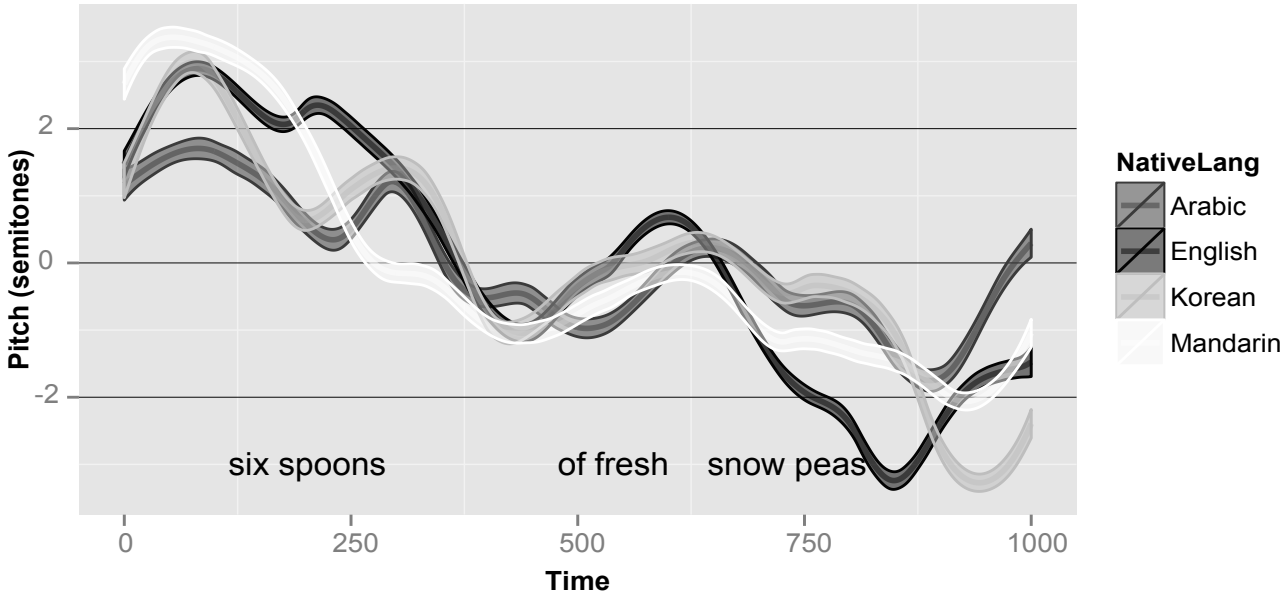
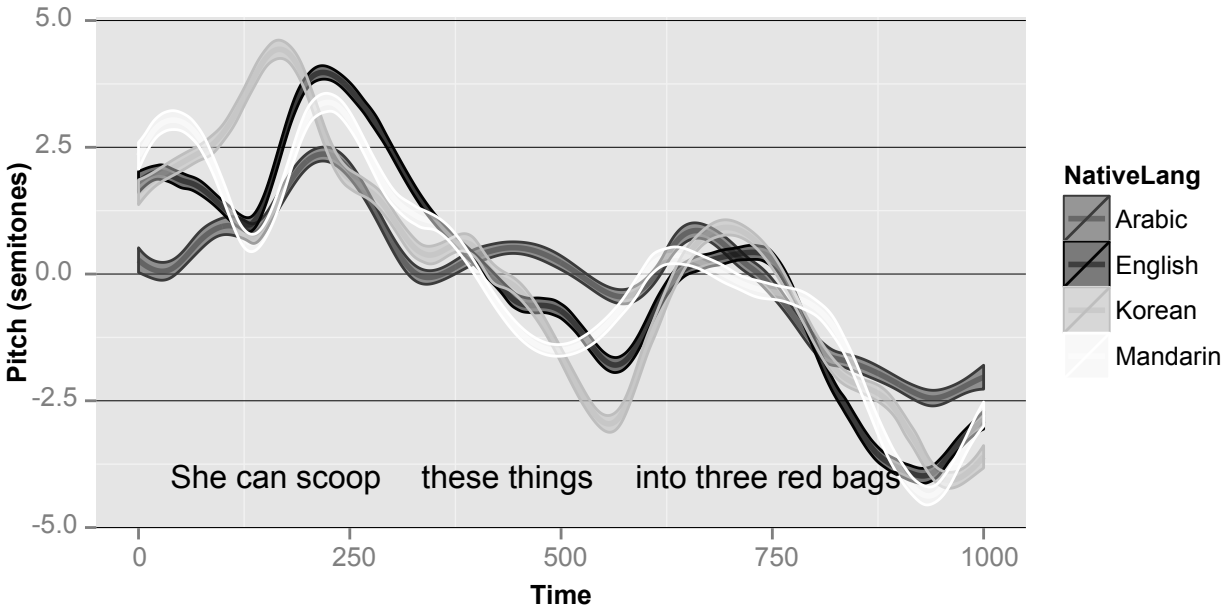


Figure 2: **Modeled pitch contours with 95% confidence intervals for the phrase *She can scoop these things into three red bags***



Two of these consist of similarities across native language groups: (1) All of the groups divide the phrase into two smaller intonational groupings, with a low F0 occurring at the final boundary of *six spoons.* This lowered F0 is followed by a uniform rise to a pitch accent on *fresh* in *fresh snow peas.* (2) All of the groups exhibit a rising F0 at the end of the phrase, typical of list intonation in English.

In the first intonational grouping (*six spoons),* the English, Arabic, and Korean groups show two pitch peaks – however, the native English speakers maintain a higher F0 peak on *spoons* than the other language groups. The Korean and Arabic groups pattern closely in the placement of multiple F0 peaks. The second intonational grouping (*of fresh snow peas)* contains an F0 lowering for all groups. Here, the native English speakers exhibit the fastest F0 decrease. This pattern matches the realization of typical English list intonation (transcribed in ToBI as L-H%), containing a low intermediate boundary tone followed by a rising tone [3].

4.2. Utterance-initial phrase: "She can scoop these things into three red bags..."

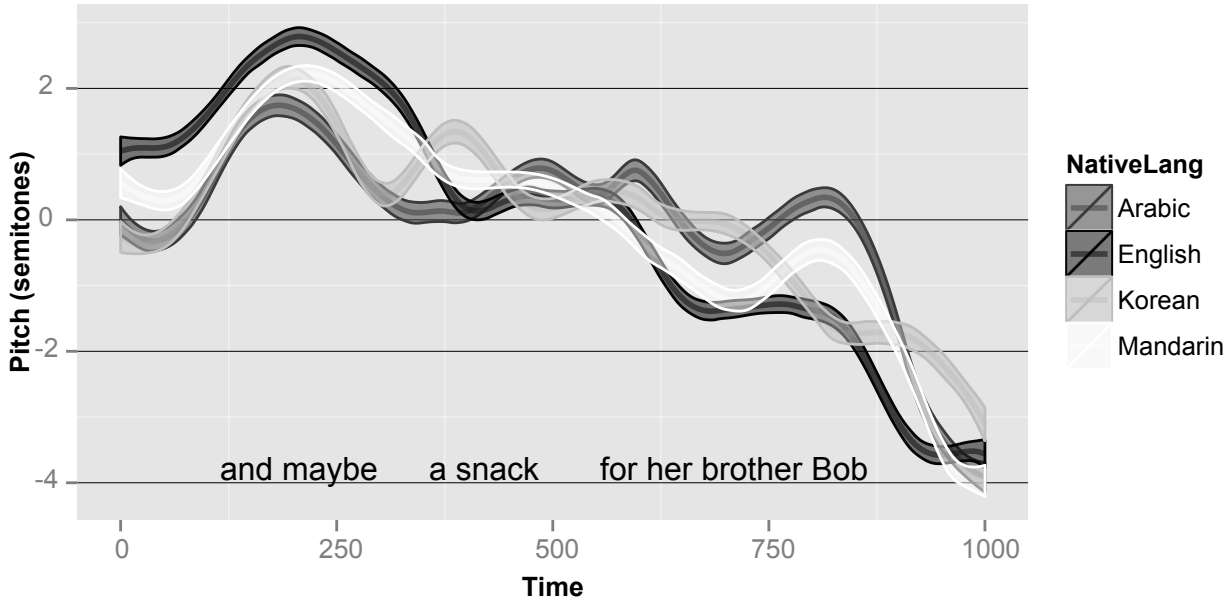
The modeled F0 contours for this phrase again demonstrate the presence of two intonational groupings (Figure 2). This result in two smaller phrases, *She can scoop these things,* and *into three red bags,* that are separated by a dip in F0 for the native English, Korean and Mandarin groups. The native Arabic speakers instead appear to maintain a higher tone for the boundary at *things,* which was confirmed by auditory and visual inspection of multiple tokens. The beginning of this utterance, *She can scoop...,* contains notable language-specific patterns. The Korean speakers show a much earlier F0 peak than the other language groups. Whereas the other groups reach an F0 peak at the location corresponding to a prominence on *scoop,* the Korean speakers reach the F0 peak on the preceding syllable, *can.* This is most likely a prosodic transfer effect from the common Korean phrase-initial intonation pattern consisting of Low-High tones across the first two syllables (in this case, *she can*) [12].

4.3. Utterance-final phrase: "…and maybe a snack for her brother Bob."

This phrase contains two main patterns (Figure 3). All of the groups have an F0 rise in the initial section (…*and maybe*). In the final section (*brother Bob*) a single tonal grouping can be seen in the Arabic, Mandarin, and native English groups.

The Korean and Arabic speakers exhibit multiple smaller F0 rises whereas the native English speakers instead produce two lowered F0 plateaus after the initial peak in the utterance. The Korean speakers place an apparent F0 peak where no other language group does (at time point ~350), reflecting the frequently produced pattern of an accent in *a snack.* As with the example in 4.2, this could be an effect of prosodic transfer from Korean phrasal intonation patterns.

Figure 3: **Modeled pitch contours with 95% confidence intervals for the phrase *...and maybe a snack for her brother Bob***



5. DISCUSSION

Speakers within each native language group produce intonational contours with a high degree of consistency; this was true for utterance-initial, utterance-final, and clausal (list intonation) phrases. There are also patterns that are present across all native language groups. Speakers generally agree on the presence of a high F0 peak towards the beginning of the phrase, with lowering F0 until a low or high phrase-final boundary tone. In addition, there are instances of higher-level tonal groupings within each phrase. In most cases, these groupings divide the longer intonational phrase into two smaller phrases, centered around the production of a high-tone pitch accent. This similarity across groups could indicate that the non-native speakers have acquired some aspects of English global prosodic patterning, or it could reflect a universal preference for higher-level tonal groupings within a phrase.

As hypothesized, the non-native language groups exhibit regularities in the way their F0 contours differ from those of native speakers. Korean speakers showed transfer from native-language phrasal intonation patterns (Figures 2 & 3); they showed frequent F0 rises and falls corresponding to prosodic word-sized tonal groupings (Figures 1 & 3). The native Arabic speakers also showed frequent F0 rises and falls (Figures 1 & 3); this is consistent with higher frequency of repeated pitch accents in some dialects of Arabic compared to English [13]. Although the native Mandarin speakers produced tonal groupings and boundary tones similarly to native English speakers, they exhibited fewer clear F0 peaks corresponding to pitch accents.

These results support a recent proposal by Jun [14,13] for the inclusion of "macro-rhythm" in prosodic typology. Languages with strong macro-rhythm, such as Korean and some dialects of Arabic, exhibit regular repeated tonal groupings at the prosodic word and intermediate phrasal levels, while languages with weak macro-rhythm, such as Mandarin, do not. The current results suggest that macro-rhythm could influence non-native speech production – in a medium-strength macro-rhythm language like English, non-native speakers from both ends of the scale differ from native speakers in ways consistent with their own native language.

6. CONCLUSION

The current findings suggest that this contour modeling approach to examining phrasal intonation can contribute to our understanding of cross-linguistic prosodic patterning. By using statistical techniques to examine many speakers producing the same utterance, it is possible to overcome some of the problems associated with F0 contour analysis. Additional work is needed to quantify the number and type of observed pitch patterns, as well as to systematically compare the current analyses to traditional prosodic labeling systems (e.g., ToBI).

7. REFERENCES

[1] Dupoux, Peperkamp, S. & Sebastián-Gallés, N. 2010. Limits on bilingualism: Stress 'deafness' in simultaneous French-Spanish bilinguals. *Cognition* 114, 266–275

[2] Cutler, A., Dahan, D. & van Donselaar, W. 1997. Prosody in the comprehension of spoken language: A literature review. *Lang. Speech* 40, 141–201

[3] Beckman, M. & Hirschberg, J. 1994. The ToBI annotation conventions. The Ohio State University and AT&T Bell Laboratories, unpublished manuscript. Available at ftp://ftp.ling.ohio-state.edu/pub/phonetics/TOBI/ToBI/ToBI.6.html.

[4] Kösling, K., Kunter, G., Baayen, H. & Plag, I. 2013. Prominence in Triconstituent Compounds: Pitch Contours and Linguistic Theory. *Lang. Speech*

[5] Moisik, S. R., Lin, H. & Esling, J. H. 2014. A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *J. Int. Phon. Assoc.* 44, 21–58

[6] Yiu, S. 2014. in *Proceedings of the 4th International Symposium on Tonal Aspects of Languages (TAL 2014)* 143–146 (ISCA Archive).

[7] Gu, C. *Smoothing Spline ANOVA Models*. 2002. New York: Springer

[8] Beckman, M. & Pierrehumbert, J. 1986. Intonational structure in Japanese and English. *Phonol. Yearb.* 3, 255–309

[9] Weinberger, S. George Mason University Speech Accent Archive. at <http://accent.gmu.edu/>

[10] Boersma, P. & Weenink, D. *Praat: Doing phonetics by computer [Computer program]*. (Software and manual available online at http://www.praat.org, 2012).

[11] Gu, C. 2014. Smoothing Spline ANOVA Models: R Package gss. *J. Stat. Softw.* 58

[12] Jun, Sun-Ah. 1996. *The phonetics and phonology of Korean prosody: intonational phonology and prosodic structure*. Taylor & Francis.

[13] Jun, Sun-Ah. 2014. in *Prosody typology II: the phonology of intonation and phrasing*. Oxford University Press

[14] Jun, S.-A. 2011. Tone-based macro-rhythm from the perspective of prosodic typology. *J. Acoust. Soc. Am.* 130, 2471–2471