F_0 Declination in English and Mandarin Broadcast News Speech

Jiahong Yuan, Mark Liberman

University of Pennsylvania

Abstract

This study investigates F_0 declination in broadcast news speech in English and Mandarin Chinese. The results demonstrate a strong relationship between utterance length and declination slope. Shorter utterances have steeper declination, even after excluding the initial rising and final lowering effects. Initial F_0 tends to be higher when the utterance is longer, whereas the low bound of final F_0 is independent of the utterance length. Both top line and baseline show declination. The top line and baseline have different patterns in Mandarin Chinese, whereas in English their patterns are similar. Mandarin Chinese has more and steeper declination than English, as well as wider pitch range and more F_0 fluctuations. Our results suggest that F_0 declination is linguistically controlled, not just a by-product of the physics and physiology of talking.

Index Terms: declination; F_0; regression; convex-hull

1. Introduction

It has been observed in many languages that the pitch contour over the course of an utterance has a downward trend, normally called F_0 declination in the literature (e.g., Cohen et al., 1982; Ladd 1984). F_0 declination is expected and used for normalization by listeners, e.g., when two stressed syllables sounded equal in pitch, the second was actually lower (Pierrehumbert, 1979; Terken, 1991, 1994). In the last fifty years, F_0 declination has been extensively investigated. Debates persist, however, over two related questions: Is F_0 declination just a by-product of the physics and physiology of talking, or is it also linguistically controlled? And is declination only the result of local F_0 events, or does it require phrase-scale pre-planning?

1.1 Is F_0 declination linguistically controlled?

F_0, or the fundamental frequency of speech, is the rate of vibration of the vocal folds during voice production. F_0 is determined by the stiffness and effective mass of the vocal folds and the subglottal air pressure (Baer, 1979; Hollien, 1983; Titze, 1988; Stevens, 2000). Intrinsic laryngeal muscles, especially the cricothyroid muscle (CT), are the main contributor to the adjustment of the stiffness and effective mass of the vocal folds. The contraction of CT raises F_0; the relaxation of CT, along with the activity of other laryngeal muscles, lowers F_0 (Collier, 1975; Atkinson, 1978). Extrinsic laryngeal muscles, which suspend and support the larynx, can also change the states of the vocal folds through vertical larynx movement (Ohala, 1978; Honda et al., 1999; Hirose, 2010), and F_0 falls as the larynx moves down.

The causes of F_0 declination over an utterance have been under debate for a long time. Some researchers claimed that F_0 declination is due to a drop in subglottal air pressure (Lieberman, 1967; Collier, 1975; Gelfer et al., 1983). This view was challenged by others who argued that the subglottal

pressure fall cannot be responsible for all, or even most, of the observed F_0 drop, and hence other factors must be involved (Maeda, 1976; Ohala, 1978). Maeda (1976) proposed that F_0 declination is caused by "tracheal pull," which gradually lowers the sternum and the larynx (as a result of its linkage to the sternum) due to decreasing lung volume. An increase in tracheal pull will cause a gradual rotation of the cricoid cartilage and decrease of the vocal fold tension, and thus a decrease in F_0. Both the drop in the subglottal air pressure and "tracheal pull" were seen as an automatic by-product of respiratory activities. Breckenbridge (1977) and Ohala (1978), however, argued that declination is part of the linguistic code, and is therefore purposeful and must be controlled by laryngeal muscles. Strik and Boves (1995) examined the two arguments against a major role for subglottal pressure in F_0 declination: 1. The lowering in the subglottal pressure cannot explain all of the decrease in F_0; and 2. F_0 declination is part of the linguistic code and must be controlled by laryngeal muscles. They claimed that both F_0 and the subglottal pressure can be decomposed into a local and a global component whereas laryngeal muscles only have a local component; and it is the global downtrend of the subglottal pressure that determines the global downtrend of F_0, i.e., F_0 declination. They also claimed that the downtrend in the subglottal pressure is actively controlled by respiratory muscles; it is not a passive process but part of the linguistic code. However, whether subglottal pressure is actively controlled in speech production is subject to debate (Ladefoged, 1967; Ladefoged and Loeb, 2009; Ohala, 1990). Vaissière (1983) proposed that F_0 declination could be due to a "laziness principle", because F_0 rises are harder to produce than F_0 falls (Sundberg, 1979; Xu and Sun, 2002**).**

1.2 Does F_0 declination require phrase-scale preplanning?

The most common way to characterize F_0 declination is in terms of straight lines fitted to F_0 contours. A line fitted to local F_0 valleys in an utterance is called the "baseline", and a line fitted to

local F_0 peaks is called the "top line". It has been debated whether one of the two lines is more fundamental than the other, or whether the two lines should be independently described (e.g., Maeda, 1976; O'Shaughnessy, 1976; Gårding, 1979; Cooper and Sorensen, 1981; see discussion in Cohen et al., 1982). Lieberman et al. (1984) argued that the linear regression line fitted to all F_0 points was a better descriptor of the F_0 contour than the baseline and top line. In contrast to the "overt decline" approach in which declination lines are fitted to empirical F_0s, the "implicit decline" approach assumes more abstract baselines that are not directly observable from pitch contours (e.g., Pierrehumbert, 1980; Fujisaki and Hirose, 1984; see discussion in Ladd, 1993).

There are two types of arguments for the existence of phrase-scale preplanning in F_0 declination. The first argument is that utterance duration affects declination slope and the height of the initial F_0 peak. A correlation between utterance duration and declination slope, i.e., longer utterances have less steep slope, has been found for the baseline (Maeda, 1976; 't Hart, 1979), the top line (Sorensen and Cooper, 1980; Cooper and Sorensen, 1981), and the overall regression line (Swerts, 1996). Liberman and Pierrehumbert (1984) argued that the correlation between the top line declination rate and utterance duration did not necessarily support the existence of phrase-scale preplanning. They found that the top line declination (and the correlation between declination rate and utterance duration) could be explained by the exponential decay of the peaks of high pitch accents ("downsteps"), i.e., each peak is a constant fraction of the previous one. The relationship between utterance duration and the height of the first F_0 peak was unclear. Some researchers reported that the first F_0 peak increases in longer utterances (Cooper and Sorensen, 1981; Shih, 2000; Rialland, 2001; Prieto et al., 2006) whereas others reported that the first F_0 peak remains more or less constant regardless of utterance duration (Sternberg et al, 1980; Liberman and Pierrehumbert, 1984; Prieto et al., 1996). van Heuven (2004) reported that it

was the size of the first downstep, not the first F_0 peak, that was proportional to the number of items in enumerations.

The second type of argument is that the F_0 declination rate is determined by speaking style and sentence type. It was reported that read speech had steeper and more frequent declination than spontaneous speech (Laan, 1997; Lieberman et al., 1984), and that the declination slope was controlled in read speech but not in spontaneous speech (Tøndering, 2011). Thorsen (1980) reported that declarative sentences have the most steeply falling contours, syntactically unmarked questions have the least falling contours, and in between these two extremes are other types of questions and non-terminal declaratives. Umeda (1982) claimed that F_0 declination is situation dependent. In her study, F_0 declination appears when all content words in a sentence are spoken as equally important or equally unimportant and when the speaker makes a gradual preparation for sentence termination.

Finally, Liberman and Pierrehumbert (1984) proposed to distinguish between "hard" and "soft" preplanning. In their proposal, "hard" preplanning is an essential part of intending to say something whereas "soft" preplanning is a preparation that a speaker may or may not choose to make. Prieto et al. (2006) studied ten speakers of Romance languages, and found that the majority of speakers, but not all of them, had a higher initial F_0 peak in longer utterances. They interpreted the result as evidence for "soft" preplanning in sentence production.

1.3 Aim of the study

In this study, we investigate F_0 declination in two languages, American English and Mandarin Chinese. Although relatively less studied than in English, F_0 declination in Mandarin Chinese has been consistently observed in a number of studies (Xu, 1999; Shih, 2000; Yuan, 2004), and also has been incorporated into various intonation models of the language (Yuan et al, 2002; Ni and Hirose, 2006).

Cross-linguistic comparison of F_0 declination will help to reveal whether it is linguistically controlled: If different languages have different F_0 declination patterns, it will be most likely that F_0 declination is linguistically controlled, not just a by-product of the physics and physiology of talking.

Unlike most previous studies of F_0 declination that relied on controlled experiments with small amounts of laboratory speech, this study uses existing large speech corpora. The F_0 contour of a sentence is affected by many linguistic and situational factors, as well as speaker characteristics. For example, a sentence may be uttered differently by different speakers, or by the same speaker in different moods. It may bear a focus, may be of a particular intonation type, and may contain different numbers of intonation or intermediate phrases in the prosodic domain. We recognize that it would be extremely difficult to investigate the effects of all possible factors on F_0 declination using large speech corpora. Some of these factors, such as phrasing and intonation type, require manual or automatic annotation. Other factors, such as speaker characteristics, cannot be meaningfully studied if the materials of the speakers are not balanced and comparable (for example, in a situation that speaker A has 100 sentences and speaker B only has 10, and we don't have prosodic annotation of these sentences). In this study we focus on two factors in F_0 declination – utterance length and language. We argue that in a large collection of natural speech corpora, many factors will tend to balance and cancel each other, and the effect of utterance length on declination will be revealed. Moreover, the two corpora used in this study (one in American English and one in Mandarin Chinese) are comparable in terms of topics, style, and size. Therefore, the difference between the two corpora will most likely represent the difference between the two languages. The study will contribute to the literature by examining the questions about F_0 declination from a new angle, through analysis of large speech corpora.

2. Data and Method

Two broadcast news speech corpora were used - the 1997 English Broadcast News Speech (LDC98S71) and the 1997 Mandarin Broadcast News Speech (LDC98S73). We extracted the "utterances", the between-pause units that are time-stamped in the transcripts from the corpora. The utterances were aligned with the transcripts using the Penn Phonetics Lab Forced Aligner (http://www.ling.upenn.edu/phonetics/p2fa/), and those containing a pause longer than 50 ms were excluded. Utterances from unknown speakers and reporters, i.e. those whose names were not tagged in the corpora, were also excluded. Finally, the utterances between one and four seconds long were selected for analysis in this study. The data set includes 5,652 English utterances from 73 speakers and 8,383 Mandarin utterances from 27 speakers. The distribution of utterances per speaker is shown in Figure 1.

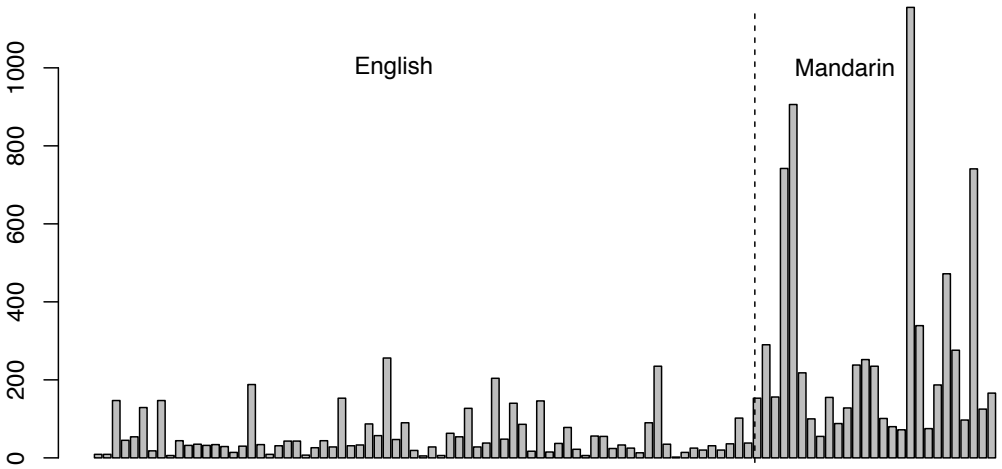

Figure 1. The number of utterances per speaker in the data.

The F_0 contours of the utterances were extracted using *esps/get_f0* with a 10 ms frame rate (Talkin and Lin, 1996). The contours were linearly interpolated to be continuous over the unvoiced segments, and smoothed by passing them (both forward and reverse to avoid phase distortion, *filtfilt*) through a Butterworth low-pass filter with normalized cutoff frequency at 0.1. The F_0 values were then converted to semitones according to equation (1). The base frequency used for calculating semitones, F_0_base in the formula, was speaker dependent: the 5th percentile of all F_0 values of a given speaker.

$$Semitone = 12 * \log_2(\frac{F_0}{F_0\_base}) \qquad (1)$$

As reviewed in the introduction, it has been debated in the literature how to measure F_0 declination and whether the top line and baseline of the F_0 contour function differently. In this study, we applied two methods to measure F_0 declination. First, a linear regression line was fitted to each F_0 contour using the least-squares method. The slopes of the fitted lines were used to represent declination slopes. Secondly, we used a "convex hull" algorithm to identify local F_0 peaks and valleys (to ignore small F_0 perturbations, as illustrated in Figure 2). The algorithm finds the peak F_0 in an F_0 contour and constructs a convex envelop over the contour, which is monotonically non-decreasing from the starting point to the peak point and monotonically non-increasing from the peak point to the end point. On each side of the F_0 peak point, the differences between the convex envelope and the F_0 data at all time points are computed and the time point that has the maximal difference is selected as a boundary, if the difference is larger than a pre-determined threshold value. Each side of the F_0 peak point is then divided into two subsegments at the new boundary, a convex envelope is constructed for each subsegment, and the procedure repeats recursively until no new boundaries can be found given the threshold value. The F_0 peaks located by the algorithm form the top line of the F_0 contour, and the boundaries located by the algorithm are F_0 valleys, which form the baseline. The threshold value was manually tuned by observing the detected F_0 peaks and valleys: one semitone was found to work well

for both English and Mandarin Chinese. (We note that the algorithm was used in a syllable segmentation task by Mermelstein, 1975). Figure 2 is an example from our data showing the peaks and valleys detected by the convex-hull algorithm.

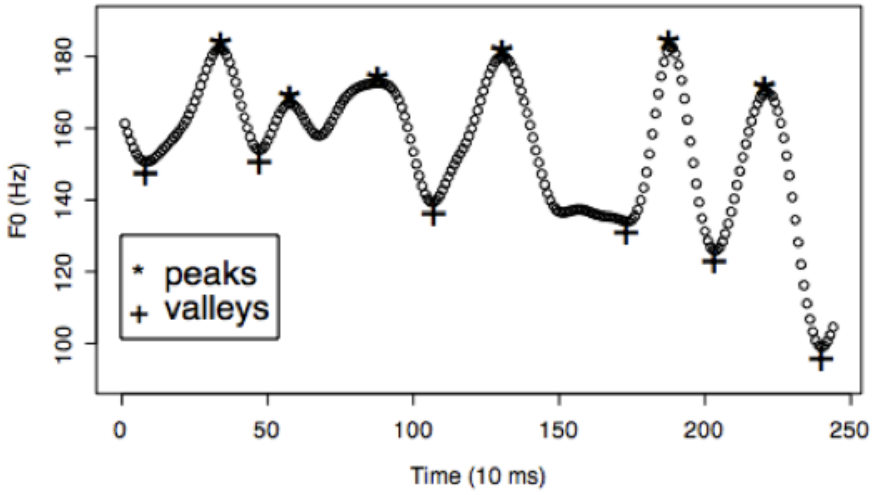

Figure 2. F_0 peaks and valleys detected by using the convex-hull algorithm.

3. Analysis

3.1. Regression lines

With regard to the linear regression lines fitted to the overall phrasal F_0 contours, 90.7% of the phrases in Mandarin Chinese have a negative slope whereas only 71.5% of the phrases in English have a negative slope. The positive slopes may come from a variety of sources such as question intonation and speaker characteristics. Because this study is to investigate the nature of F_0 declination, the utterances with a positive regression slope (i.e., a rising regression line) will be excluded from the analysis below.

Figure 3 (left panel) shows the boxplots of the negative slopes only. A t-test shows that the negative slopes in Mandarin Chinese are lower (steeper) than those in English ($t = -24.97$, $p < 0.01$). It is unclear why Mandarin Chinese has steeper downward F_0 contours than English. One possible explanation is

that Mandarin Chinese broadcast new speech has wider pitch range than English broadcast news speech, as shown in Figure 3 (right panel). Previous studies have shown that listeners expected more declination in wide pitch range utterances than in narrow pitch range utterances (Pierrehumbert, 1979).

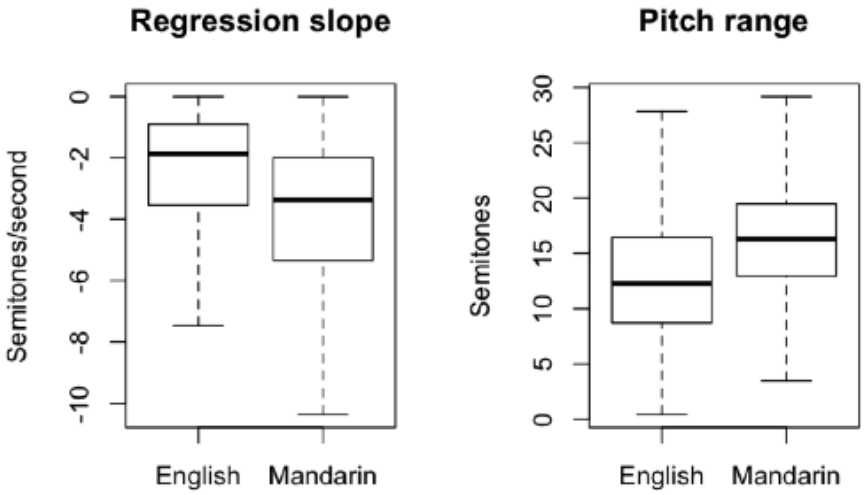

Figure 3. Boxplots of regression slope and pitch range in English and Mandarin Chinese.

The relationship between utterance duration and declination slope is shown in Figure 4. The figure shows that there exists a correlation between utterance length and declination slope: The shorter the utterance, the steeper the slope (i.e., the absolute value of the declination slope is higher). A linear mixed-effects model is used to test the effects of language and utterance duration, as fixed factors, on the declination slope, in which speaker is treated as a random effect factor. The result shows that utterance duration is a significant factor in predicting declination slope ($t = -30.53$, $p < 0.01$), the longer the utterance, the smaller the (absolute) slope. Language is also a significant factor. Mandarin has steeper slopes than English ($t = 8.46$, $p < 0.01$). And the interaction between language and utterance duration is also significant ($t = -8.30$, $p < 0.01$).

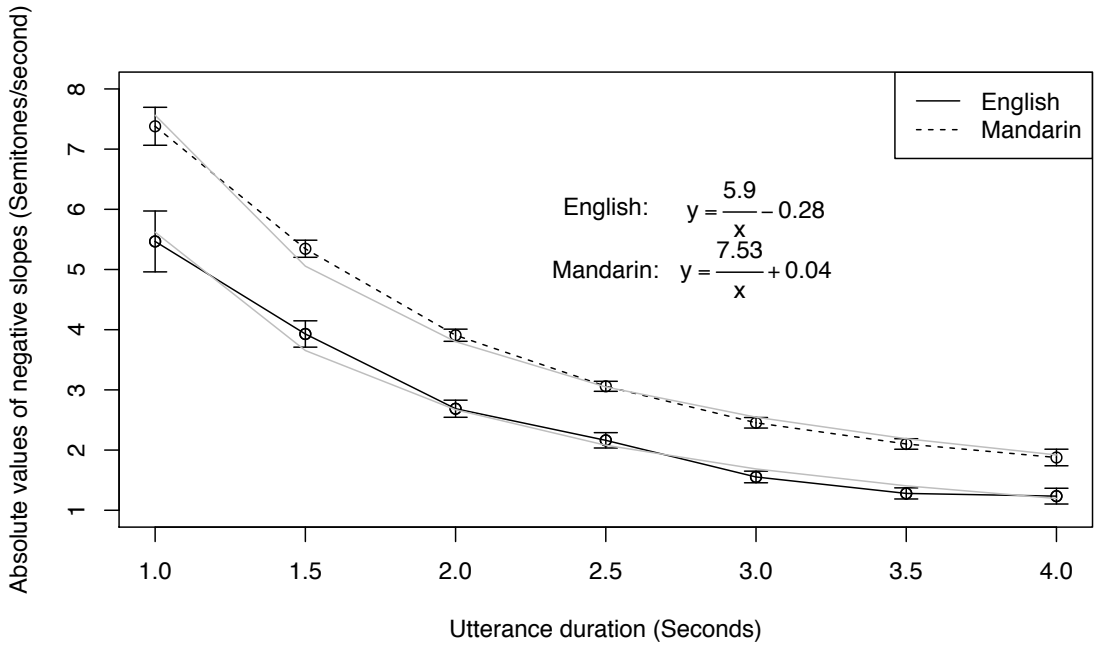

Figure 4. Mean absolute declination slope vs. utterance length. The durations were rounded to the nearest number shown on the x-axis. The grey curves are the fitted functions representing the relationship between slope (y) and duration (x).

From the fitted functions that model the relationship between declination slope and utterance duration shown in Figure 4, we can see that declination slope increases approximately linearly with the reciprocal of utterance duration. The relationship between time and slope is, however, complicated by the fact that the F_0 contour is not a straight line (Gelfer et al., 1983). Typically, F_0 reaches its first peak quickly, then fluctuates and gradually declines, and finally ends with a final lowering (Liberman and Pierrehumbert, 1984; Vaissière, 2005). To study whether the time-slope relationship holds without the effect of initial raising and final lowering, we fitted a regression line to the middle points of an utterance only, excluding the points in the first and last 500 ms of

the utterance. The slopes from using only the middle points are compared with those using all points in Table 1. Because the middle-points regression excluded one second of points from an utterance, only the utterances longer than two seconds were used in the comparison. As we can expect the slopes of the middle points only are less steep than the slopes of all points. What is interesting is that in both English and Mandarin Chinese the time-slope relationship still holds after excluding the initial and final 500 ms, i.e., shorter utterances have steeper declination within the middle part of the utterance.

Table 1. The slopes (mean ± 2sd, in semitones/second) from using the full F_0 contours and using the middle part of the F_0 contours only (excluding the initial and final 500 ms).

Duration (seconds)	English (full contour)	English (middle part)	Mandarin (full contour)	Mandarin (middle part)
2.0	-2.68 ± 0.07	-2.04 ± 0.17	-3.90 ± 0.05	-3.00 ± 0.14
2.5	-2.16 ± 0.06	-1.83 ± 0.11	-3.06 ± 0.04	-2.38 ± 0.09
3.0	-1.55 ± 0.05	-1.36 ± 0.08	-2.45 ± 0.04	-1.81 ± 0.07
3.5	-1.28 ± 0.05	-0.97 ± 0.07	-2.10 ± 0.04	-1.60 ± 0.06
4.0	-1.23 ± 0.07	-1.04 ± 0.09	-1.88 ± 0.07	-1.39 ± 0.10

3.2. Initial height and final lowness

Figure 5 and 6 show the initial height and final lowness of F_0 contours in English and Mandarin Chinese as a function of utterance length. The initial height is the highest F_0 value in the first 500 milliseconds of a F_0 contour and the final lowness is the lowest F_0 value in the final 500 milliseconds. We can see from Figure 5 that the initial height increases with utterance length. A linear regression shows that the increase rate is 0.51 semitones per second for English ($t = 5.66$, $p < 0.01$) and 0.69 semitones per second for Mandarin Chinese ($t = 9.38$, $p < 0.01$). Both of them are statistically significant. The final

lowness is, as shown in Figure 6, independent of the utterance length. A linear regression shows that the rate of final lowness over utterance length is not significantly different from 0 for both English ($t = 1.34$, $p > 0.1$) and Mandarin Chinese ($t = 1.23$, $p > 0.1$). These results suggest that the speakers preplan the initial F_0 height based on the length of the utterance to speak whereas they simply drop the F_0 to the lower bound of their pitch range, i.e., "pitch floor", at the end of the utterance. Liberman and Pierrehumbert (1984) demonstrated that the final low F_0 is "a relatively invariant characteristics of a speaker's voice".

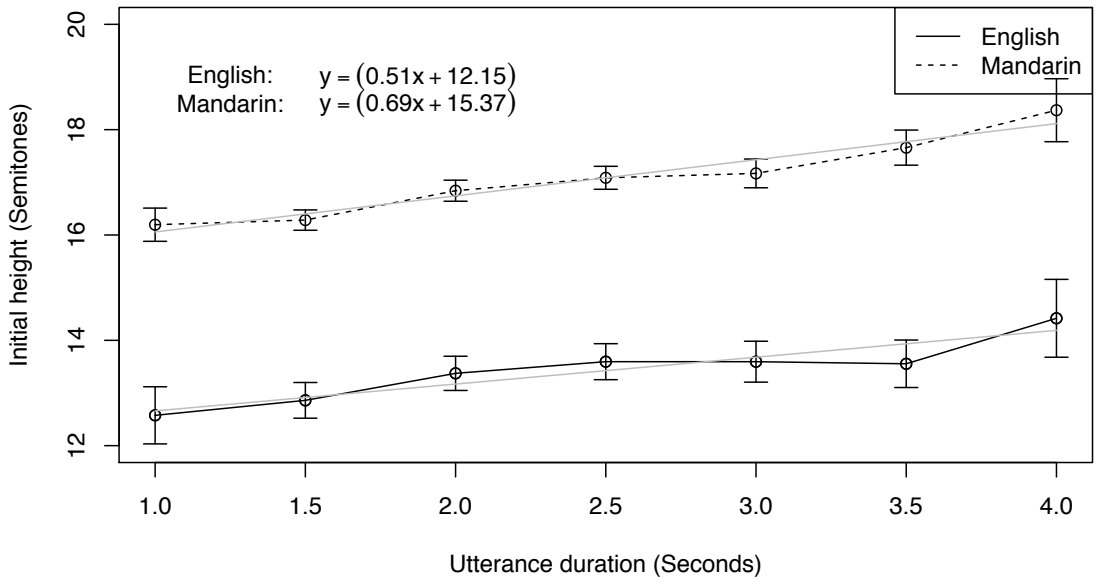

Figure 5. The highest F_0 in the first 500 ms of F_0 contours. The grey lines are linear regression lines.

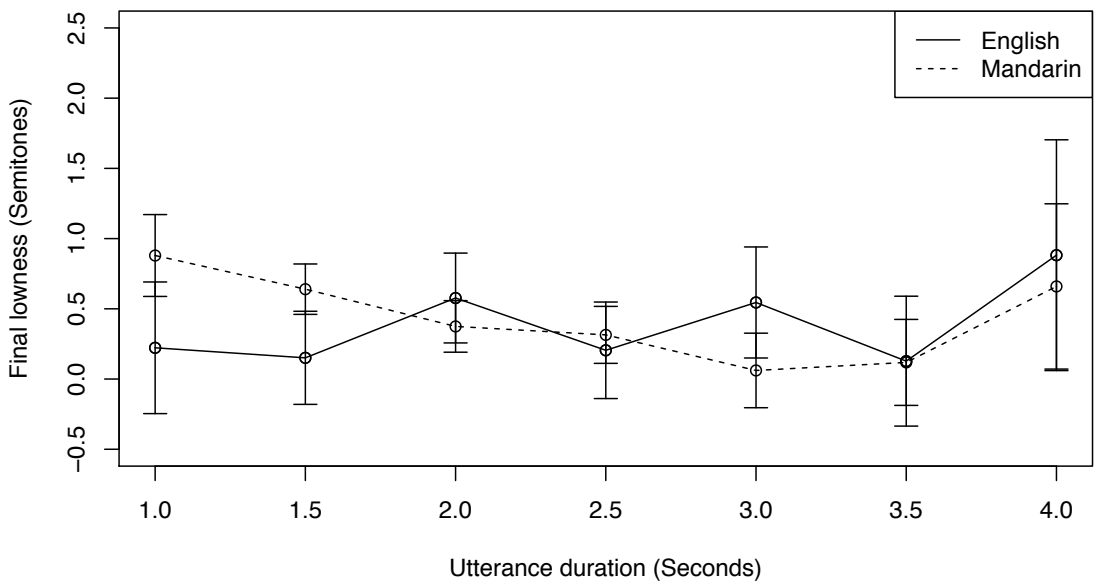

Figure 6. The lowest F_0 in the final 500 ms of F_0 contours.

3.3. Top and bottom lines

Figure 7 shows the top line and baseline patterns in English and Mandarin Chinese. The lines were drawn by taking average of the F_0 peaks and valleys (detected by the convex-hull algorithm) at relative utterance positions. The point on the top line at the relative position of .2, for example, represents the F_0 peaks that appeared between 10 and 30 percent of the utterance duration.

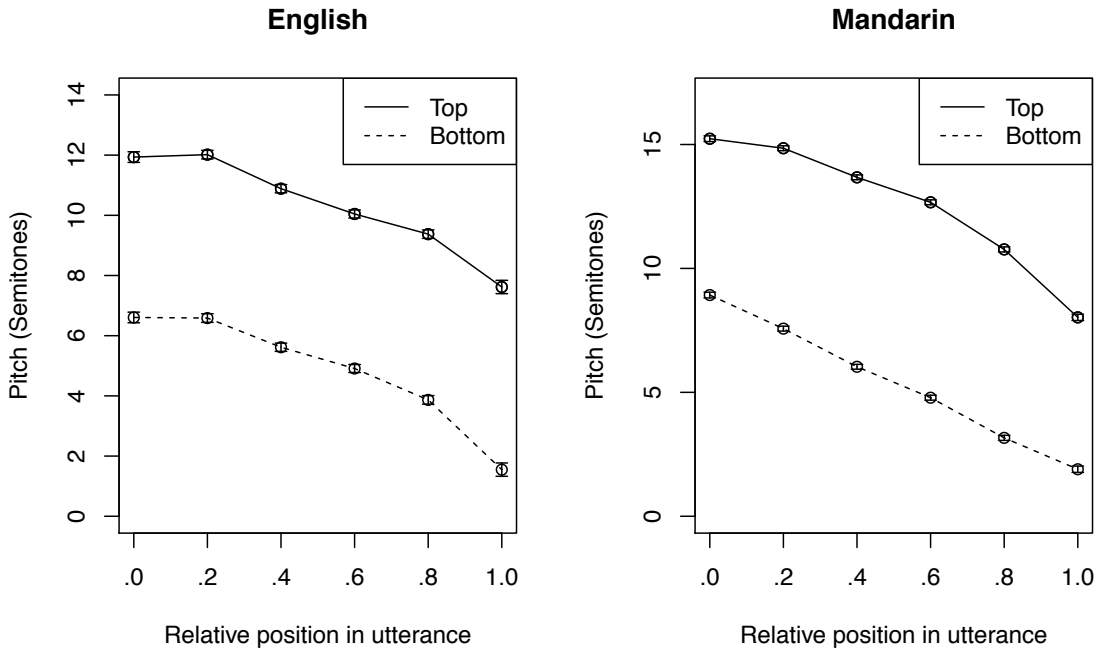

Figure 7. Top and bottom lines in English and Mandarin Chinese. The peaks and valleys were grouped based on their sequence positions in the utterance.

From the figure we can see that both the top line and baseline show declination, in both English and Mandarin Chinese. Also, the top line has final lowering in both languages. The baseline of Mandarin Chinese is close to a straight line, which is consistent with the observation reported in Yuan (2004). The top line and baseline have different patterns in Mandarin Chinese, whereas in English they are very similar, both consisting of three parts: initial plateau, middle declination, and final lowering.

Finally, Figure 8 shows the average total number of F_0 peaks and valleys per second in the F_0 contours of English and Mandarin Chinese. Because the number of F_0 peaks and valleys found by the convex-hull algorithm depends on the threshold used in the algorithm, results from a range of threshold values are presented in the figure. We can see

that Mandarin Chinese has more F_0 peaks and valleys in every second of speech than English, and that the difference becomes larger when the threshold value is larger, i.e., when only larger peaks and valleys are counted. This is probably due to the effect of lexical tones in Mandarin Chinese.

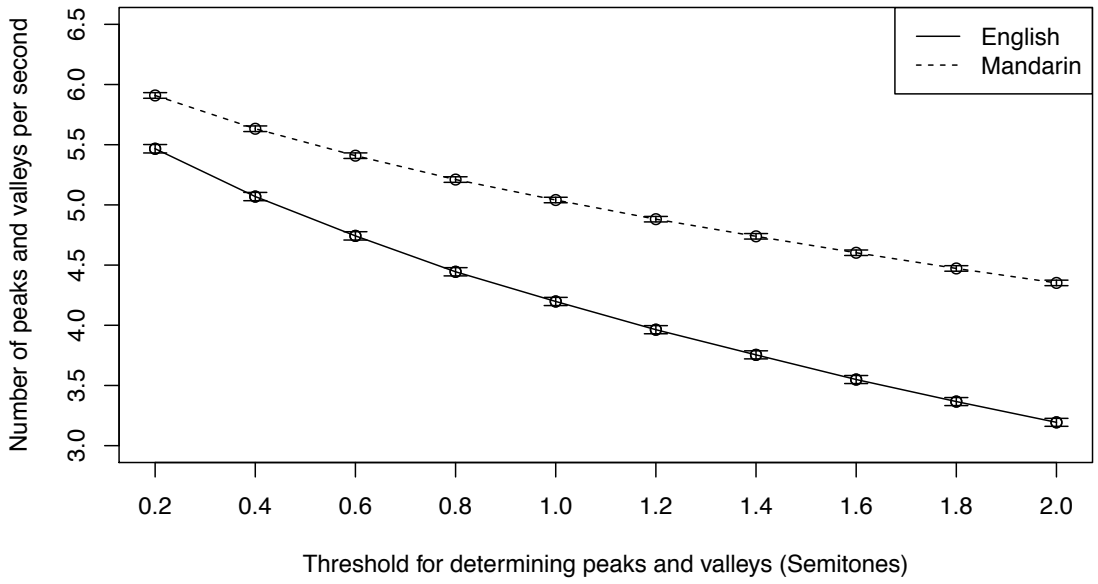

Figure 8. The average total number of F_0 peaks and valleys per second in English and Mandarin Chinese. F_0 peaks and values were determined by the convex-hull algorithm with threshold values from 0.2 to 2.0 semitones.

4. Discussion and Conclusions

In this study we investigated F_0 declination in large broadcast news speech corpora in English and Mandarin Chinese. We applied two methods, linear regression and convex-hull analysis. The former is used to measure declination slope, and the latter is used to

extract F_0 peaks and valleys in an utterance for depicting its top line and baseline patterns.

Analysis of the data demonstrated a strong correlation between declination slope and utterance length: the shorter the utterance, the steeper the declination is. This relationship holds when excluding the initial and final 500 ms, i.e., using only the middle points of an utterance to fit a line. Moreover, the initial F_0 peak is higher when the utterance is longer. These results suggest that F_0 declination is controlled by speakers, and that there is preplanning or preprogramming of declination in speech production at the phrasal level. Besides F_0 declination, preplanning in speech production has been investigated and attested from a variety of perspectives (Nooteboom, 1995), for example, anticipatory speech errors (Nooteboom and Cohen, 1975), voice initiation time (Holmes, 1984), and inspiration in speech (Whalen and Kinsella-Shaw, 1997). Our results support the existence of phrase-scale preplanning in broadcast news speech. A similar investigation of F_0 declination in large-scale collections of spontaneous speech is needed to understand the size or amount of pitch-related look-ahead in spontaneous speech production. Further investigation is also needed to establish to what extent F_0 declination and other types of speech preplanning are speaker-dependent characteristics ("soft" preplanning).

In our data, both the top line and baseline show declination, and the top line has final lowering in both languages. In our English data, the baseline and top line are similar, both consisting of three parts: initial plateau, middle declination, and final lowering. In Mandarin Chinese, the top line is similar to the top line in English, while the baseline is close to a straight line. This cross-linguistic difference may suggest that top line and baseline declinations are independent phenomena; they are not automatic by-products of

some physiological process, but linguistically controlled. Future studies, in particular cross-linguistic studies of more languages, are needed to understand the mechanisms that underlie the top line and baseline patterns in F_0 declination.

Finally, our results showed that Mandarin Chinese has wider pitch range and more F_0 fluctuations than English, probably due to the effect of lexical tones. Previous studies have investigated this topic using controlled speech materials. Eady (1982) demonstrated that Mandarin Chinese had a greater average rate of F_0 change and more F_0 peaks and valleys than English. Keating and Kuo (2012), however, showed a different result. They found Mandarin had a wider pitch range only in single word utterances but not in prose passage or story voices. And they suggested that the pitch range difference between Mandarin Chinese and English was not due to some generic property of tone languages, but specifically because of Mandarin's high falling tone (Tone 4). Our result from analyzing large speech corpora is consistent with Eady (1982) but not Keating and Kuo (2012). Interestingly, Han et al. (2011) found that in both speech and traditional music, tonal languages have more pitch direction changes and larger pitch intervals than non-tonal languages, suggesting a link between the tonal characteristics of a culture's music and its speech.

5. Acknowledgements

An earlier version of this paper was presented at Interspeech 2010. This work was supported in part by NSF grants 0964556.

References

Atkinson, J. E., 1978. Correlation analysis of the physiological factors controlling fundamental voice frequency. J. Acoust. Soc. Am. 63, 211-222.

Baer, T., 1979. Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. J. Acoust. Soc. Am. 65, 1271-1275.

Breckenridge, J., 1977. Declination as a phonological process. Bell Laboratories Technical Memorandum, Murray Hill, NJ.

Cohen, A., Collier, R., 't Hart J., 1982. Declination: construct or intrinsic feature of speech pitch? Phonetica 39, 254-273.

Collier, R., 1975. Physiological correlates of intonation patterns. J. Acoust. Soc. Am. 58, 249-255.

Cooper, W. E., Sorensen, J. M., 1981. Fundamental Frequency in Sentence Production. Springer-Verlag.

Eady, S. J., 1982. Differences in the F_0 patterns of speech: Tone language versus stress language. Language and Speech 25, 29-42.

Fujisaki, H., Hirose, K., 1984. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. J. Acoust. Soc. Jpn. (E) 5, 233–242.

Gårding, E., 1979. Sentence intonation in Swedish. Phonetica 36, 207-215.

Gelfer, C. E., Harris, K. S., Collier, R., Baer, T., 1983. Is declination actively controlled? In: I. R. Titze, R. C. Scherer (Eds.), Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control, The Denver Center for the Performing Arts, Inc. pp. 113-126.

Han, S., Sundararajan, J., Bowling, D. L., Lake, J., Purves, D., 2011. Co-variation of tonality in the music and speech of different cultures. PLoS ONE 6, e20160, 1-5.

’t Hart, J., 1979. Exploration in automatic stylization of F_0 contours. IPO APR 14, 61-65.

van Heuven, V. J., 2004. Planning in speech melody: Production and perception of downstep in Dutch. In: H. Quené, V. J., van Heuven (Eds.), On Speech and Languages: Studies for Sieb G. Nooteboom, Netherlands Graduate School of Linguistics, pp. 83-93.

Hirose, H., 2010. Investigating the physiology of laryngeal structures. In: W. J. Hardcastle, J. L. Laver, F. E. Gibbon (Eds.), The Handbook of Speech Sciences (second edition), Wiley-Blackwell, pp. 130-152.

Hollien, H., 1983. In search of vocal frequency control mechanisms. In: D. M. Bless and J. H. Abbs (Eds.), Vocal Fold Physiology: Contemporary Research and Clinical Issues, College-Hill Press, pp. 361-367.

Honda, K., Hirai, H., Masaki, S., Shimada, Y. 1999. Role of vertical larynx movement and cervical lordosis in F0 control. Language and Speech 42, 401-411.

Holmes, V. M., 1984. Sentence planning in a story continuation task. Language and Speech 27, 115-134.

Keating, P., Kuo, G., 2012. Comparison of speaking fundamental frequency in English and Mandarin. J. Acoust. Soc. Am. 132, 1050-1060.

Laan, G. P. M., 1997. The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. Speech Communication 22, 43-65.

Ladd, D. R., 1984. Declination: a review and some hypotheses. Phonology yearbook 1, 53-74.

Ladd, D. R., 1993. On the theoretical status of “the baseline” in modelling intonation. Language and Speech 36, 435-451.

Ladefoged, P. 1967. Three areas of experimental phonetics. Oxford University Press.

Ladefoged, P., Loeb, G., 2009. Preliminary studies on respiratory activity in speech. http://www.linguistics.ucla.edu/people/ladefoge/LadefogedAndLoebRespiration.pdf.

Liberman, M., Pierrehumbert, J., 1984. Intonational invariance under changes in pitch range and length. In: M. Aronoff, R. Oerhle (Eds.), Language Sound Structure, MIT Press, pp. 157-233.

Lieberman, P., 1966. Intonation, perception and language. Ph.D. thesis, MIT.

Lieberman, P., Katz, W., Jongman, A., Zimmerman, R., Miller, M., 1984. Measures of the sentence intonation of read and spontaneous speech in American English. J. Acoust. Soc. Am. 77, 649-657.

Maeda, S., 1976. A characterization of American English intonation. Ph.D. thesis, MIT.

Mermelstein. P., 1975. Automatic segmentation of speech into syllabic units. J. Acoust. Soc. Am. 58, 880-883.

Ni, J., Hirose, K., 2006. Quantitative and structural modeling of voice fundamental frequency contours of speech in Mandarin. Speech Communication 48, 989-1008.

Nooteboom, S. G., 1995. Limited lookahead in speech production. In: F. Bell-Berti, L. J. Raphael (Eds.), Producing Speech: Contemporary Issues, American Institute of Physics, pp. 3-18.

Nooteboom, S. G., Cohen, A., 1975. Anticipation in speech production and its implication for perception. In: A. Cohen, S. G. Nooteboom (Eds.), Structure and process in speech perception, Springer, pp. 124-142.

Ohala, J. J., 1978. The production of tone. In: V. A. Fromkin (Ed.), Tone: a linguistic survey, Academic Press, pp. 5-39.

Ohala, J. J., 1990. Respiratory activity in speech. In: W. J. Hardcastle, A. Marchal (Eds.), Speech production and speech modeling, Kluwer Academic Publishers, pp. 23-53.

O'Shaughnessy, 1976. Modelling fundamental frequency and its relationship to syntax, semantics, and phonetics. Ph.D. thesis, MIT.

Pierrehumbert, J., 1979. The perception of fundamental frequency declination. J. Acoust. Soc. Am. 66, 363-369.

Pierrehumbert, J., 1980. The Phonology and Phonetics of English Intonation. Ph.D. thesis, MIT.

Prieto, P., D'Imperio, M., Elordieta, G., Frota, S., Vigário, M., 2006. Evidence for soft preplanning in tonal production: Initial scaling in Romance. Proceedings of Speech Prosody 2006, pp. 803-806.

Prieto, P., Shih, C., Nibert, H., 1996. Pitch downtrend in Spanish. Journal of Phonetics 24, 445-473.

Rialland, A., 2001. Anticipatory raising in downstep realization: Evidence for preplanning in tone production. Proceedings of Symposium Cross-Linguistic Studies of Tonal Phenomena: Tonogenesis, Japanese Accentology, and Other Topics, pp. 301-321.

Shih, C., 2000. A Declination Model of Mandarin Chinese. In A. Botinis (Ed.), Intonation: Analysis, Modelling and Technology, Kluwer Academic Publishers, pp. 243-268.

Sorensen, J. M., Cooper, W. E., 1980. Syntactic coding of fundamental frequency in speech production. In: R. A. Cole (Ed.), Production and Perception of Fluent Speech, Lawrence Erlbaum Associates, pp. 399-440.

Stevens, K. N., 2000. Acoustic Phonetics, MIT Press, pp. 55-126.

Sternberg, S., Wright, C. E., Knoll, R. L., Monsell, S. 1980. Motor programs in rapid speech: additional evidence. In: R. A. Cole (Ed.), Production and Perception of Fluent Speech, Lawrence Erlbaum Associates, pp. 507-534.

Strik, H., Boves, L., 1995. Downtrend in F_0 and P_{sb}. Journal of Phonetics 23, 203-220.

Sundberg, J., 1979. Maximum speed of pitch changes in singers and untrained subjects. Journal of Phonetics 7, 71–79.

Swerts, M., Strangert, E., Heldner, M., 1996. F_0 declination in read-aloud and spontaneous speech. Proceedings of ICSLP '96, pp. 1501-1504.

Talkin, D., Lin, D., 1996. Get_f0 online documentation. ESPS/Waves, Entropic Research Laboratory.

Terken, J., 1991. Fundamental frequency and perceived prominence of accented syllables. J. Acoust. Soc. Am. 89, 1768-1776.

Terken, J., 1994. Fundamental frequency and perceived prominence of accented syllables. II. Nonfinal accents. J. Acoust. Soc. Am. 95, 3662-3665.

Thorsen, N. G., 1980. A study of the perception of sentence intonation - evidence from Danish. J. Acoust. Soc. Am. 67, 1014-1030.

Titze, I. R., 1988. The physics of small-amplitude oscillation of the vocal folds. J. Acoust. Soc. Am. 83, 1536-1552.

Tøndering, J., 2011. Preplanning of intonation in spontaneous versus read aloud speech: evidence from Sanish. Proceedings of ICPhS XVII, pp. 2010-2013.

Umeda, N., 1982. "F_0 declination" is situation dependent. Journal of Phonetics 10, 279-290.

Vaissière, J., 1983. Language-independent prosodic features. In: A. Cutler, D. R. Ladd (Eds.) Prosody: Models and Measurements, Springer-Verlag, pp. 53-66.

Vaissière, J., 2005. Perception of Intonation. In: D. B. Pisoni, R. E. Remez (Eds.), The Handbook of Speech Perception, Blackwell Publishing, pp. 236-263.

Whalen, D. H., Kinsella-Shaw, J. M., 1997. Exploring the relationship of inspiration duration to utterance duration. Phonetica 54, 138-152.

Xu, Y., 1999. Effects of tone and focus on the formation and alignment of F0 contours. Journal of Phonetics 27, 55-105.

Xu, Y., Sun X., 2002. Maximum speed of pitch change and how it may relate to speech. J. Acoust. Soc. Am. 111, 1399-1413.

Yuan, J., 2004. Intonation in Mandarin Chinese: Acoustics, Perception, and Computational Modeling. Ph.D. thesis, Cornell University.

Yuan, J., Shih, C., Kochanski, G. P., 2002. Comparison of declarative and interrogative intonation in Chinese. Proceedings of Speech Prosody 2002, pp. 711-714.