F_0 Declination in English and Mandarin Broadcast News Speech

Jiahong Yuan, Mark Liberman

University of Pennsylvania, USA

jiahong@ling.upenn.edu, myl@cis.upenn.edu

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. Both topline and baseline show declination, but they are independent. The topline 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.

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

1. Introduction

Declination refers to the downward trend of F_0 over the course of an utterance [1, 2]. It was found that declination was expected and used for normalization by listeners, i.e., when two stressed syllables sounded equal in pitch, the second was actually lower [3]. The physiological causes for F_0 declination include the tracheal pull [4], the downtrend of subglottal pressure [5, 6], and the activity of laryngeal muscles [7]. It has been unclear, however, whether declination is part of the linguistic code and speaker-controlled, or an automatic by-product of some physiological process [2, 8].

Another, but related debate about declination is whether it is a global attribute that requires whole-phrase planning, or a concatenation of local events. Studies have found that the F_0 topline, i.e., the line connecting F_0 peaks, was steeper in short sentences than in long ones [9], and the initial F_0 peak increased with sentence length [10]. [11] proposed that the slope of the declination depends on the sentence type: it is steepest for terminal declaratives and least steep for interrogatives. In [12], however, it was argued that declination can be explained by downstep plus final lowering effects. Downstep is the lowering effect triggered by a low tone that is between two high tones; final lowering refers to the final F_0 peak being lower than expected from the downstep effect alone. [13] suggested that final lowering is grammaticalized and independent of declination.

F_0 declination has also been found in Mandarin Chinese, a tonal language [14, 15]. [16] demonstrated that there is declination on high tone sequences, in which there are no low tones to trigger downsteps. [17] showed a linear baseline declination effect in Mandarin Chinese, as illustrated by Figure 1 below. [18] studied dialectal variation in declination pattern of Taiwan Mandarin.

Besides declination, the F_0 contour of a sentence is affected by many linguistic and situational factors. Most previous studies of declination used controlled experiments to minimize the effects of the other factors. In this study, we investigate F_0 declination in large broadcast news speech corpora. We argue that by using large and natural speech corpora the other factors will balance and cancel each other, and the "true" declination effect will be revealed. We will also compare the declination effect between English and Mandarin Chinese. Cross-linguistic comparison of declination has rarely been undertaken in previous studies. Such a comparison will help us better understand the nature of declination.

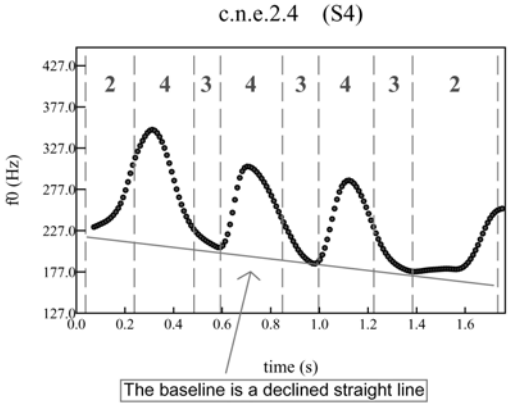

Figure 1: *Baseline declination in Mandarin Chinese.*

2. Data and Method

Two broadcast news speech corpora were used for this study, 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 *LDC98T22* and *LDC98T24*, from the corpora. The utterances were forced aligned using the PPL Forced Aligner [19] and those containing a pause longer than 50 ms were excluded. The utterances from unknown speakers and reporters, i.e., whose names were not tagged in the corpora, were also excluded. Finally, the utterances between one and four seconds long were selected and utilized for this study. The data include 5,652 English utterances and 8,383 Mandarin utterances.

The F_0 contours of the utterances were extracted using esps/get_f0 with a 10 ms frame rate [20]. The contours were linearly interpolated to be continuous over the unvoiced segments, and smoothed by passing them through a Butterworth low-pass filter with normalized cutoff frequency at 0.1. The F_0s were then converted to semitones by using the formula below:

$$Semitone = 12 * \log_2(\frac{F_0}{F_{0\_base}})$$

The base frequency used for calculating semitones was speaker dependent, defined as the 5th percentile of all F_0 values for that speaker.

Two methods were applied 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 then analyzed.

Secondly, we used convex-hull, a peak detection algorithm that has been successfully applied in syllable segmentation tasks [21], to identify local F_0 valleys and peaks. Figure 2 shows an example of the peaks and valleys detected by convex-hull in our data.

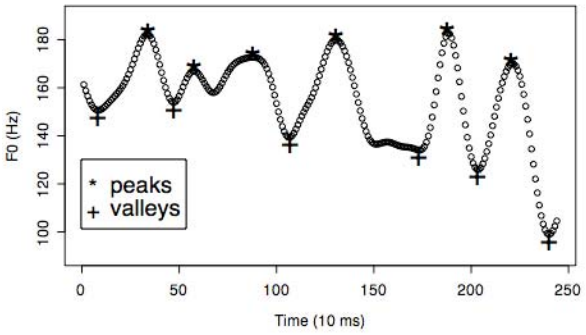

Figure 2: *F_0 peaks and valleys detected by using Convex-hull.*

In the analysis below, the F_0 peaks were connected to form the topline and the F_0 valleys were connected to form the baseline.

3. Results

3.1. Regression lines

The fitted lines in Mandarin Chinese have a higher percentage of negative slopes than those in English: 90.7% Mandarin vs. 71.5% English. The mean slopes of English and Mandarin utterances are shown in Figure 3.

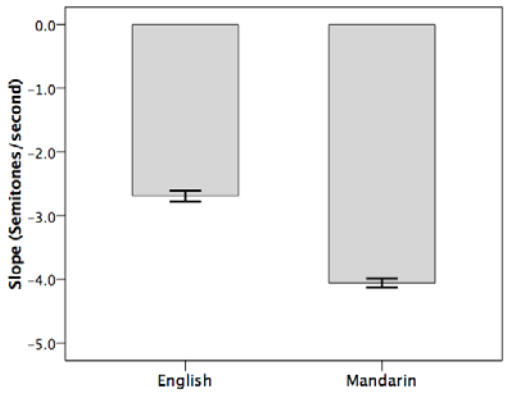

Figure 3: *Mean slopes of English and Mandarin utterances.*

The negative slopes in Mandarin Chinese are also steeper than those in English. Figure 4 shows the mean values of the negative slopes only (excluding the positive ones). We can see that Mandarin has a steeper slope than English when the utterance length is the same.

It is not obvious why Mandarin Chinese has more and steeper declination than English. One possible explanation is that Mandarin Chinese has wider pitch range than English in broadcast news speech, which is shown in Figure 5. Previous studies have found that listeners expected more declination in wide pitch range utterances than in narrow pitch range utterances [3].

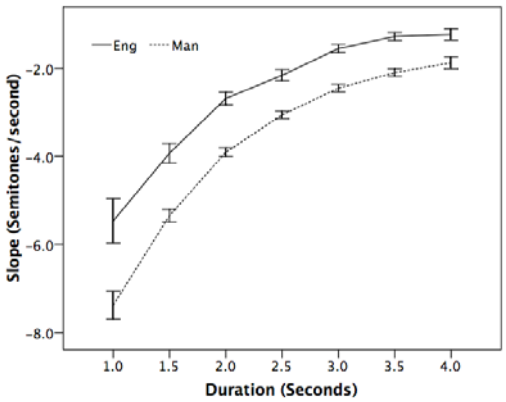

Figure 4: *Declination slope vs. utterance length. The durations were rounded to the nearest number shown on the x-axis.*

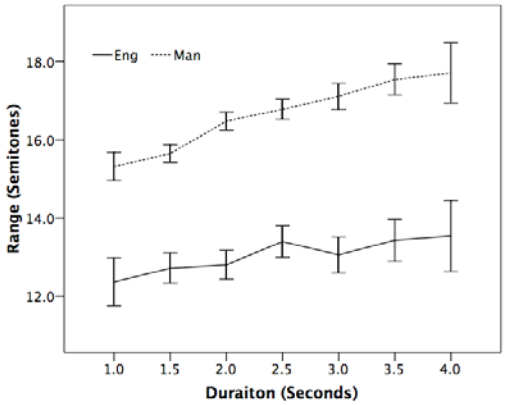

Figure 5: *Pitch range vs. utterance length. The durations were rounded to the nearest number shown on the x-axis.*

From Figure 4 we can clearly see a correlation between utterance length and declination slope. The shorter the utterance, the steeper the slope is. The relationship between time and slope is, however, complicated by the fact that the F_0 contour is not a straight line. Typically, F_0 reaches its first peak quickly, then fluctuates and gradually declines, and finally ends with a final lowering.

To study whether the time-slope relationship holds without the effect of initial rising 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 Figure 6 and 7. 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.

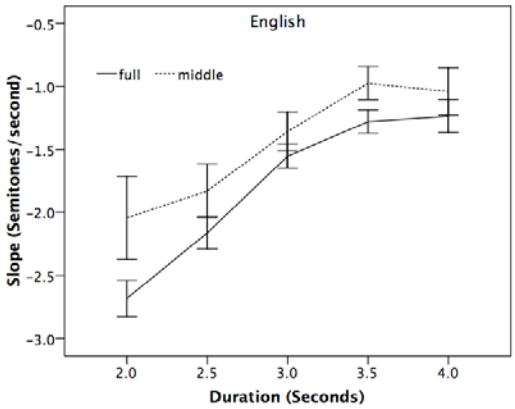

Figure 6: *Regression over all points vs. the middle points of English utterances (excluding the initial and final 500 ms).*

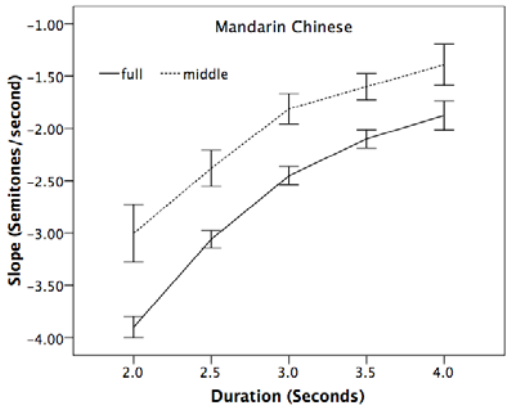

Figure 7: *Regression over all points vs. the middle points of Mandarin utterances (excluding the initial and final 500 ms).*

3.2. Top and bottom lines

Figures 8 and 9 show the topline and baseline patterns in English and Mandarin Chinese respectively. The lines were drawn by taking average of the F_0 peaks and valleys at relative utterance positions. The point on the topline at the relative position of .2, for example, represents the F_0 peaks that appeared between 10 and 30 percent of the utterance duration.

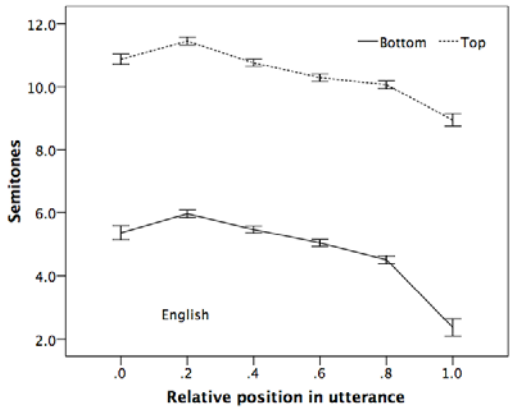

Figure 8: *Top and bottom lines in English. The peaks and valleys were grouped based on their relative positions in the utterance.*

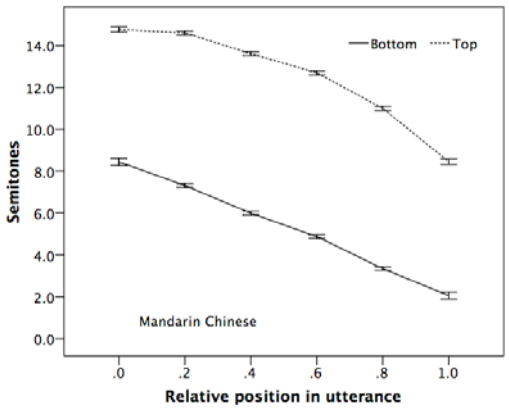

Figure 9: *Top and bottom lines in Mandarin. The peaks and valleys were grouped based on their relative positions in the utterance.*

From the figures we can see that both the topline and baseline show declination, in both English and Mandarin Chinese. Also, the topline has final lowering in both languages. The baseline of Mandarin Chinese is close to a straight line. This is consistent with the observation reported in [17]. The topline and baseline have different patterns in Mandarin Chinese, whereas in English they are very similar, both consisting of three parts: initial rising, middle declination, and final lowering.

Finally, Figure 10 shows the total number of F_0 peaks and valleys in an utterance in English and Mandarin Chinese. We can see that Mandarin Chinese has more F_0 peaks and valleys, i.e., more F_0 fluctuations, than English. This is probably the effect of lexical tones in Mandarin Chinese.

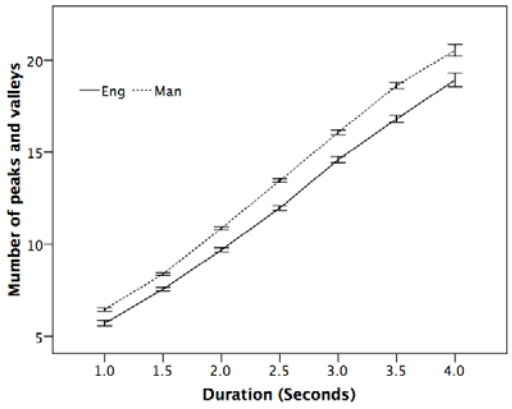

Figure 10: *The total number of F_0 peaks and valleys in an utterance in English and Mandarin Chinese.*

4. Conclusions and discussion

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. 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 topline 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. This result may

suggest that the declination slope is controlled by speakers, and that there is preplanning on declination in speech production.

Both the topline and baseline show declination, and the topline has final lowering in both languages. In Mandarin Chinese, the baseline is close to a straight line, which is different from its topline. Menwhile in English, the baseline and topline are similar, both consisting of three parts: initial rising, middle declination, and final lowering. This cross-linguistic difference may suggest that topline and baseline declinations are independent phenomena; they are not automatic by-products of some physiological process, but linguistically controlled.

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.

5. References

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

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

[3] Pierrehumbert, J., "The perception of fundamental frequency declination", JASA 66, 363-369, 1979.

[4] Maeda, S., A characterization of American English intonation, PhD dissertation, MIT, 1976.

[5] Lieberman, P., Intonation, perception, and language, Cambridge: MIT Press, 1967.

[6] Collier, R., "Physiological correlates of intonation patterns", JASA 58, 249-255, 1975.

[7] Ohala, J. J., "Respiratory activity in speech", In W. J. Hardcastle and A. Marchal (eds), *Speech production and speech modeling*, pp. 23-53. Netherlands: Kluwer Academic Publishers, 1990.

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

[9] Swerts M., Strangert E., and Heldner M., "F_0 declination in spontaneous and read-aloud speech", Proceedings of ICSLP, Philadelphia, 3, 1501-1504, 1996.

[10] Cooper. W. E. and Sorenscn, J. M., *Fundamental frequency in sentence production*. New York: Springcr-Verlag, 1981.

[11] Thorsen N., "A Study of the Perception of Sentence Intonation – Evidence from Danish", JASA 67, 1014-1030, 1980.

[12] Liberman, M. and Pierrehumbert J., "Intonational invariance under changes in pitch range and length", In M. Aronoff and R. Oehrle (eds), *Language sound structure*, pp. 157-233, Cambridge: MIT Press, 1984.

[13] Arvaniti, A. (to appear), "On the presence of final lowering in British and American English", In C. Gussenhoven and T. Riad (eds), *Tones and Tunes, volume II, Phonetic and Behavioural Studies in Word and Sentence Prosody*, Berlin, New York: Mouton de Gruyter.

[14] Shen, J., "Beijinghua Shengdiao de Yinyu he Diaoyu (Pitch range and intonation of the tones of Beijing Mandarin)", In Beijing Yuyin Shiyan Lu (Acoustic studies of Beijing Mandarin), pp. 73-130, Beijing University Press, 1985.

[15] Garding, E., "Speech act and tonal pattern in Standard Chinese: constancy and variation," Phonetica 44, 13-29, 1987.

[16] Shih, C., "A declination model of Mandarin Chinese," In A. Botinis (ed.) Intonation: Analysis, Modelling and Technology, Dordrecht/Boston/London: Kluwe, Academic Publishers. Pp. 243-268, 2003.

[17] Yuan, J., *Intonation in Mandarin Chinese: Acoustics, perception, and computational modeling*, Ph.D. dissertation, Cornell University, 2004.

[18] Huang, Y. H. and Fon, J., "Dialectal Variations in Tonal Register and Declination Pattern of Taiwan Mandarin", Proceedings of Speech Prosody 2008, pp. 605-608, 2008.

[19] http://www.ling.upenn.edu/phonetics/p2fa/

[20] Talkin, D.; Lin, D., Get_f0 online documentation. ESPS/Waves, Entropic Research Laboratory, 1996.

[21] Mermelstein. P., "Automatic segmentation of speech into syllabic units," JASA, 58(4), 880-883, 1975.