
F₀ Declination in English and Mandarin Broadcast News Speech

Jiahong Yuan and Mark Liberman

University of Pennsylvania

Sep. 27, 2010

Now with extra Spanish!

After submitting the paper to this conference,
we added data from Spanish news broadcasts.

We'll report (some of) the results in this presentation.

Introduction (1)

- Declination refers to the downward trend of F_0 over the course of an utterance (Cohen, Collier and 't Hart, 1982).
- Possible physiological causes for F_0 declination:
 - Downtrend of subglottal pressure (Lieberman, 1967)
 - Tracheal pull (Maeda, 1976)
 - Activity of laryngeal muscles (Ohala, 1990).
- Debates about F_0 declination:
 - Is it (only) a by-product of the physics and physiology of talking?
Or is it (also) a linguistically-controlled feature?
(Ladd 1984, Strik and Boves 1995).
 - Is it (only) the result of local events and fixed-rate processes?
Or does it require phrase-scale pre-planning?

Introduction (2)

Findings and theories about F_0 declination:

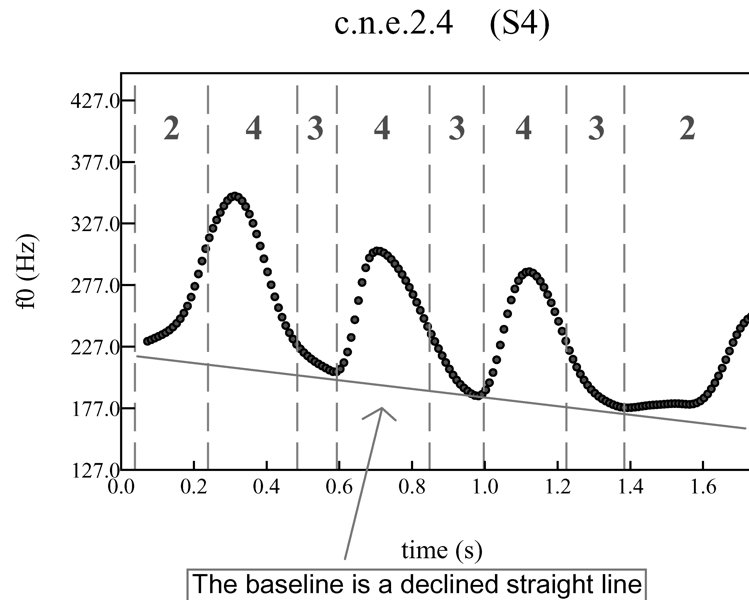
- The slope of the declination depends on the sentence type: it is steepest for terminal declaratives and least steep for interrogatives (Thorsen, 1980).
- The initial F_0 peak increased with sentence length (Cooper and Sorensen, 1981).
- Declination could be explained by (local) downstep plus (local) final lowering effects (Lieberman and Pierrehumbert, 1984).
- The F_0 topline, i.e., the line connecting F_0 peaks, was steeper in short sentences than in long ones (Swerts *et al.*, 1996).
- Final lowering is grammaticalized and independent of declination (Arvaniti, 2009).

Introduction (3)

- 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 hypothesize that in large natural speech corpora, the other factors will balance and cancel each other, and the “true” declination effect will be revealed.
- In any case, this source of evidence allows us to examine the debates about declination from another angle.

Introduction (4)

- F_0 declination is also found in Mandarin Chinese (Xu, 1999):
 - Declination on high tone sequences (Shih 2003)
 - Linear baseline declination (Yuan 2004)



- We will compare the declination effect in English and in Mandarin Chinese.

Data

- Three broadcast news speech corpora were used:
 - 1997 English Broadcast News Speech (*LDC98S71*)
 - 1997 Mandarin Broadcast News Speech (*LDC98S73*)
 - 1997 Spanish Broadcast News Speech (*LDC98S74*)
- The “utterances” (between-pause units in the transcripts) were aligned with the transcripts using the PPL Forced Aligner:
 - Units containing a pause longer than 50 ms were excluded.
 - Units from unknown speakers were also excluded.
- Utterances 1-4 seconds long were selected for this study:
 - 5,652 in English; 8,383 in Mandarin; 3,904 in Spanish

(All data and scripts will be published in association with the paper)

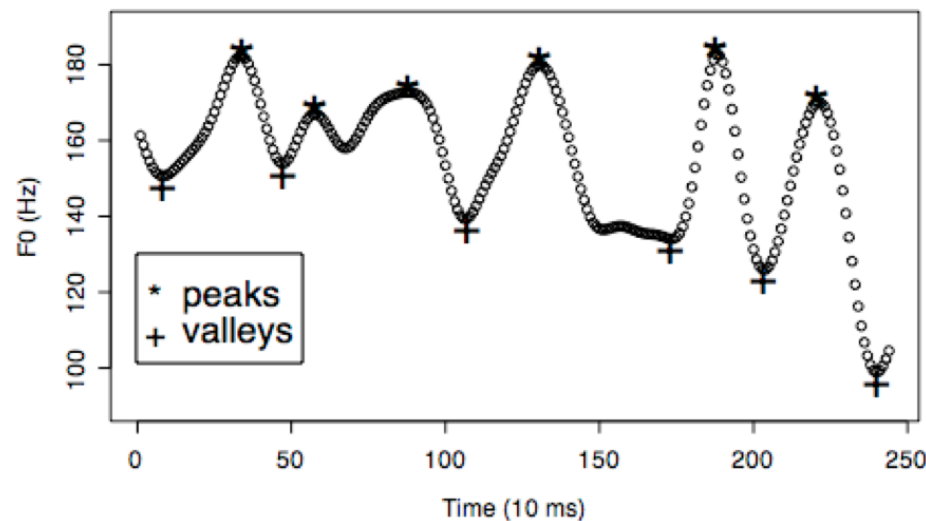
Methods (1)

- F_0 contours were extracted using `esps/get_f0` with a 10 ms frame rate.
- 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.
- F_0 values were converted to semitones. The base frequency used for calculating semitones was speaker dependent, defined as the 5th percentile of all F_0 values for that speaker.

$$\text{Semitone} = 12 * \log_2\left(\frac{F_0}{F_0_base}\right)$$

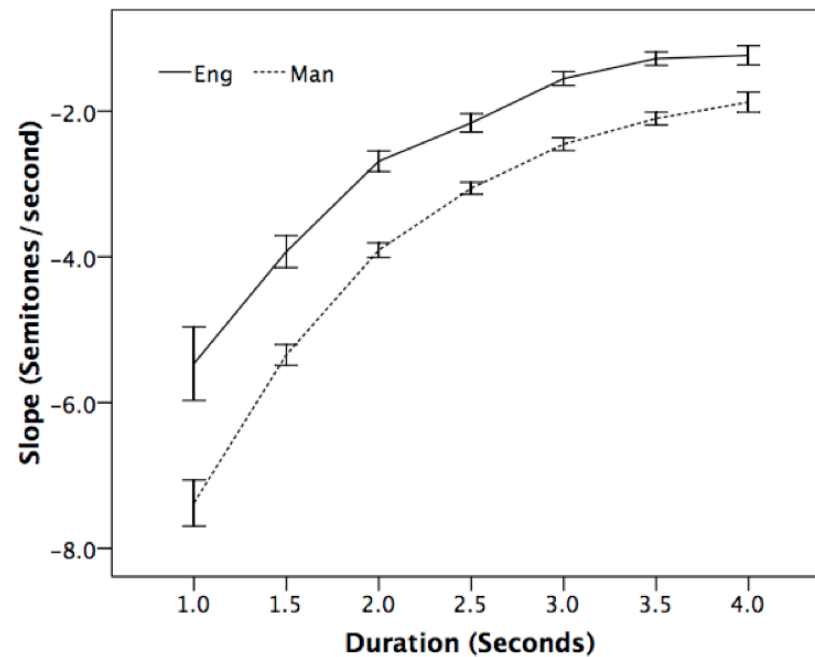
Method

- Two methods were applied to measure F_0 declination.
 - A linear **regression line** was fitted to each contour using the least-squares method.
 - The convex-hull algorithm (Mermelstein, 1975) was used to identify local F_0 valleys and peaks. The F_0 peaks were connected to form the **topline** and the F_0 valleys were connected to form the **baseline**.



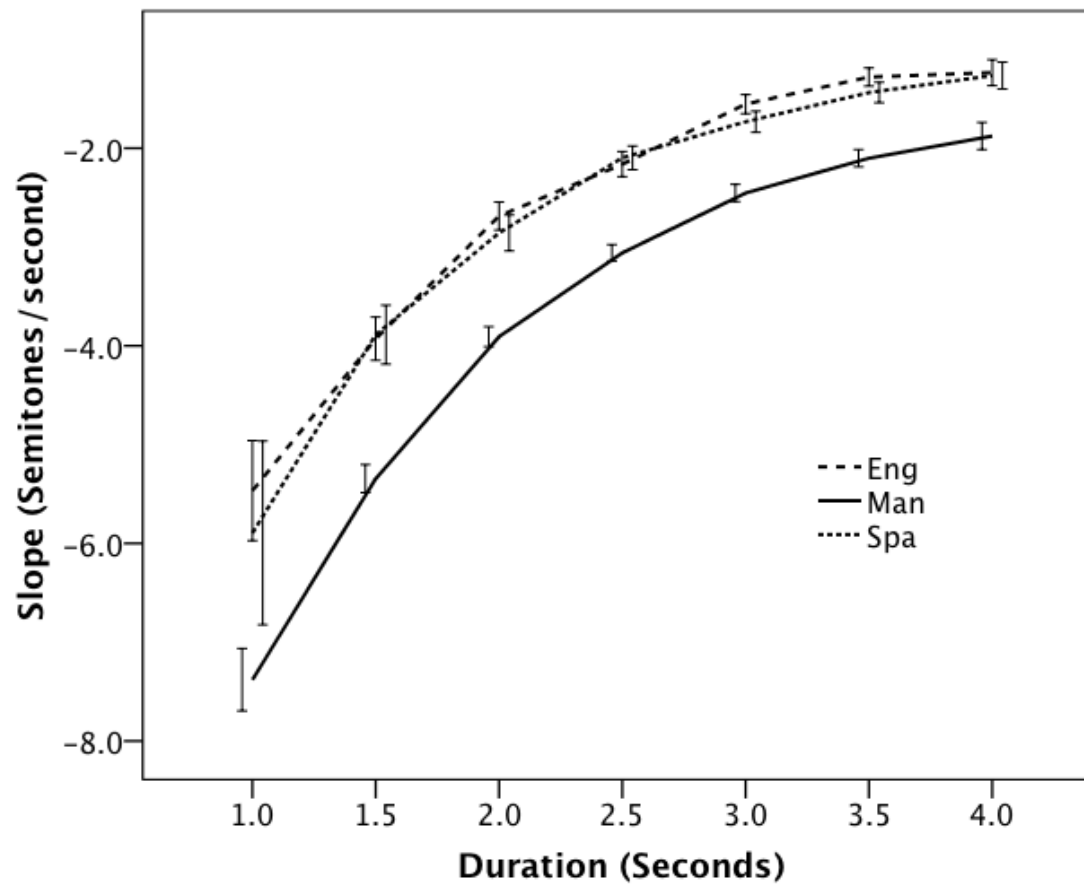
Results: Length and Slope

- A strong correlation between utterance length and declination slope: The shorter the utterance, the steeper the regression line is.



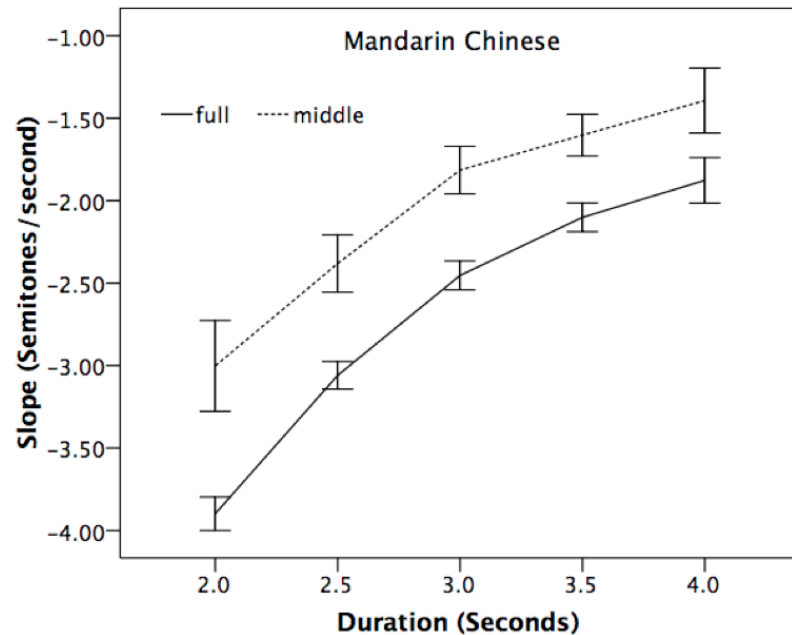
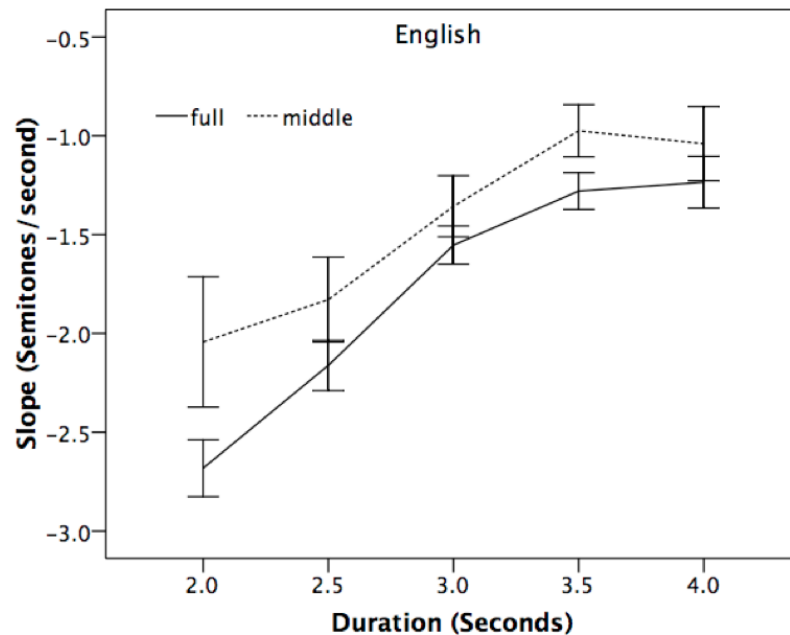
And also in Spanish!

Our Spanish data were very close to English:



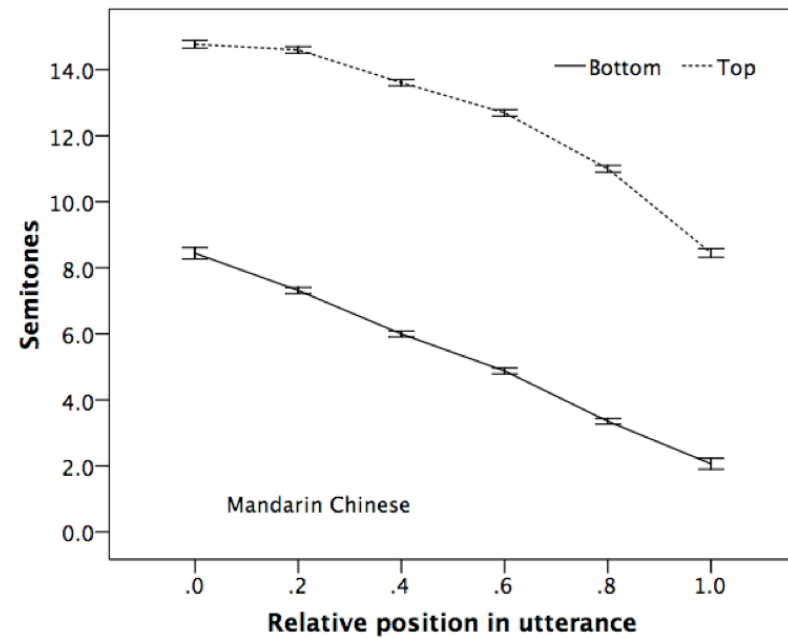
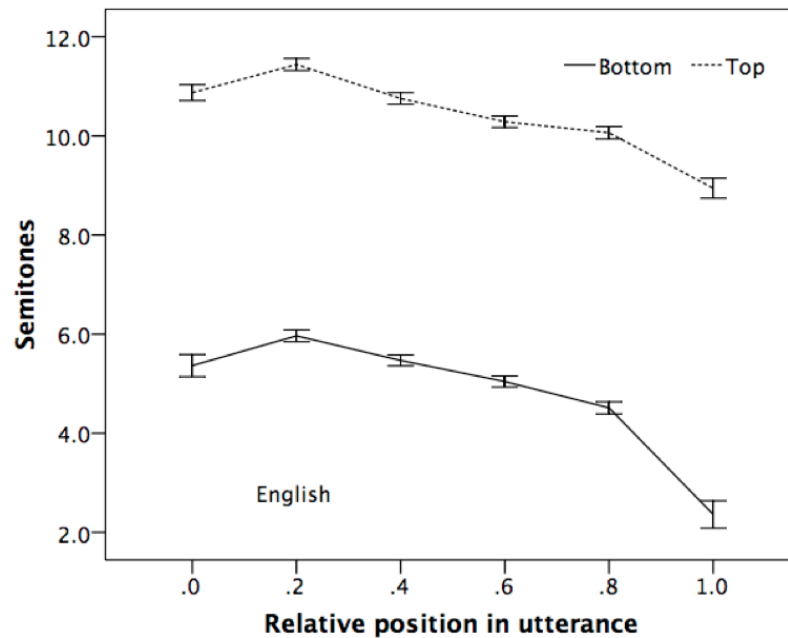
Results: Length and Slope, Excluding Edges

- The time-slope relationship still holds after excluding the initial and final 500 ms. That is, shorter utterances have steeper declination within the middle region, from which likely “edge effects” are excluded.



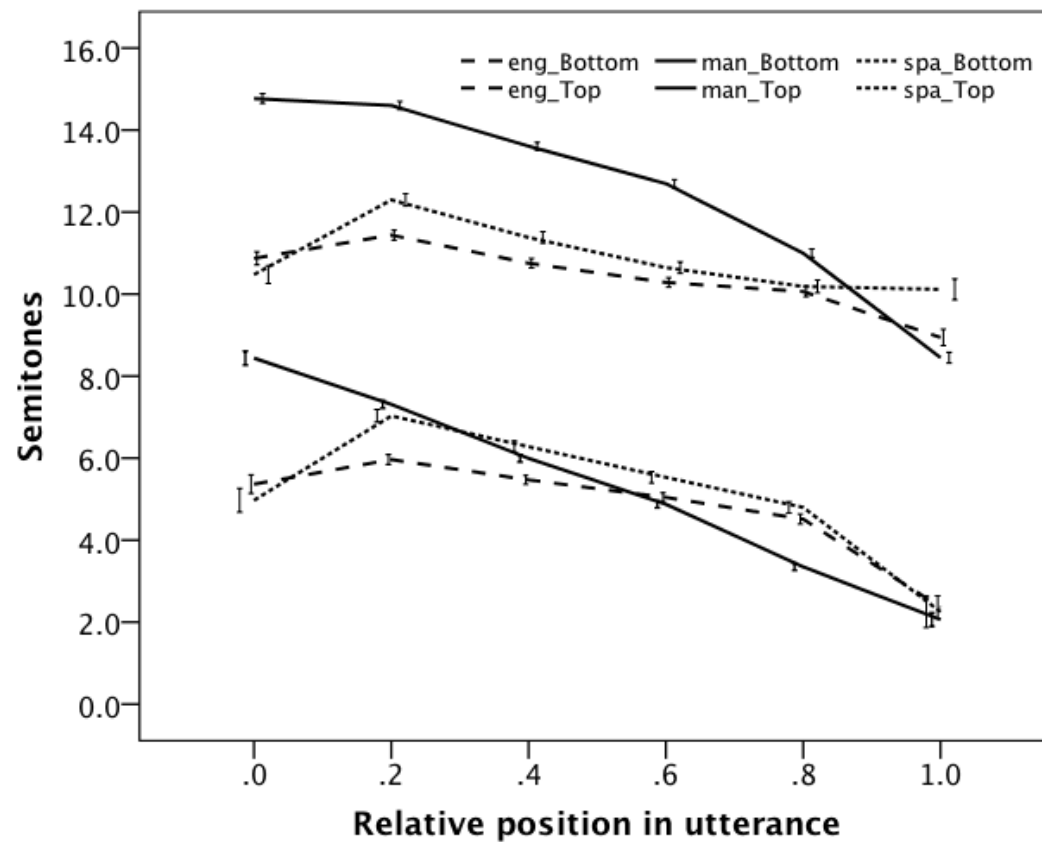
Results: Topline and Baseline

- Both the topline and baseline show declination, and the topline has final lowering in both languages.
- The baseline of Mandarin Chinese is close to a straight line.
- In English the topline and baseline are very similar: initial rising, middle declination, and final lowering.



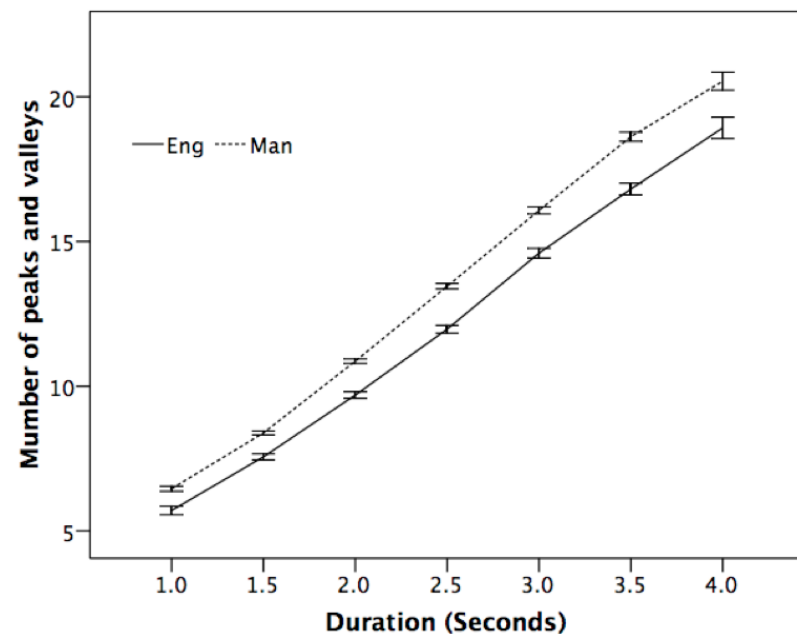
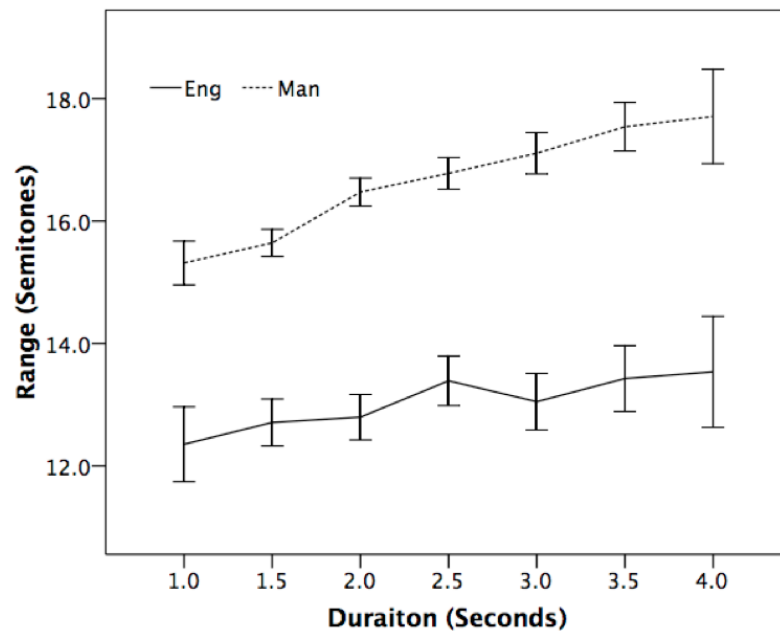
And also in Spanish!

- Again, Spanish is very much like English (in this measure, on this data):



Results: Pitch Range, Number of Extrema

- In this sample, Mandarin Chinese has wider pitch range than English.
- Mandarin Chinese also has more F_0 peaks and valleys than English, i.e. more F_0 fluctuations



Conclusions and discussion (1)

- We investigated F_0 declination in large broadcast news speech corpora in English and Mandarin Chinese (... and Spanish).
- We applied two methods, linear regression and convex hull. Regression was used to measure overall declination slope, and the convex hull was used to extract F_0 peaks and valleys for representing the topline and baseline patterns.
- There was a strong relation between declination slope and utterance length: the shorter the utterance, the steeper the declination. This relationship still holds when the first and last 0.5 sec. are excluded.
- This result suggests that speakers may control declination slope (directly or indirectly) in a way that requires phrase-level planning.

Conclusions and discussion (2)

- Both the topline and baseline show declination, and the topline has final lowering -- in all three languages.
- In Mandarin Chinese, the baseline is close to a straight line, which is different from the topline.
- In English and Spanish, the baseline and topline are similar, both consisting of three parts: initial rising, middle declination, and final lowering.
- This cross-linguistic difference suggests that topline and baseline declinations are partly independent phenomena, and that they are not entirely automatic by-products of some physiological process, but are linguistically controlled at least in part.
- Finally, our results showed that Mandarin Chinese has wider pitch range and more F_0 fluctuations than English.

This is probably due to the effect of lexical tone, but may also represent a difference in broadcast styles.

Thank you!

More on Sources

- 1997 English Broadcast News Speech (*LDC98S71*):
ABC World News Tonight, CNN Headline News, CNN Early Prime,
CNN Prime News, CNN The World Today, PRI The World,
CSPAN Public Policy, CSPAN Washington Journal
- 1997 Mandarin Broadcast News Speech (*LDC98S73*):
CC-TV, KAZN-AM, VOA
- 1997 Spanish Broadcast News Speech (*LDC98S74*):
ECO, Univision, VOA