

# DECLINATION OF READ SPEECH IN L1 AND L2 ENGLISH

Sally Chen and Janice Fon

National Taiwan University  
[sallychentw@gmail.com](mailto:sallychentw@gmail.com)

## ABSTRACT

In this study, declination of native and non-native English was investigated based on their prosodic grouping. Read speech were obtained from a national standardized English test in Taiwan. Mandarin L2 learners of English who obtained the highest score were categorized as High-L2, and those who barely passed as Low-L2. Two additional native groups of English and Mandarin were recruited to record the same/translated materials under similar settings. Their recordings were compared with the test-takers'.

Data labelling followed the convention of Tone and Break Indices for English. Results showed that L2 learners in general assigned more intonational phrases to the same texts than the native speakers. Greater degrees of declination were observed in native Mandarin than English, showing language-specific preferences. In addition, a developmental trend was found in the degree of declination for learners at different proficiency levels, and it could not be completely explained away by the effect of duration.

**Keywords:** declination, English, L2 prosody, read speech, ToBI.

## 1. INTRODUCTION

Declination was first realized as a result of physiological constraints [2], and it was not until later that the issue has been investigated in the realm of speech. The downward trend is commonly observed in speech production, whereas in speech perception, listeners tend to automatically compensate for such a downtrend [6]. Empirical studies further supported that F0 declination is not just a by-product of the physiology, but is also linguistically controlled [9].

In fact, the issue of declination is extremely complicated [6]. For instance, it is believed to be relevant to mental programming, as speakers make use of this cue in differentiating parenthetical clauses from main clauses [4]. In addition, the gradual decrease of F0 serves as a major cue not only in signalling utterance ends, but in decoding discourse structures [7]. In this way, and perhaps with the additive cue of reset, the presence of declination

facilitates communication. Speakers use declination to demarcate boundaries of various linguistic levels, and the melodic patterns can be effectively perceived by listeners.

It is indicated that the domain of declination is worth noticing, as declining F0 trends are likely to occur on units of different sizes, varying from lower-level phrases to discourse structures [7,8]. Time-dependency has also been found in previous studies. For example, the initial height was found positively correlated with utterance length, and this holds true cross-linguistically, as both Mandarin and English native data showed a similar pattern [9].

Other influential factors include genre and language backgrounds. With similar characteristics of declination patterns, Swerts et al. found that read speech was more time-dependent and showed larger reset than spontaneous speech [8]. As for language backgrounds, since learners of a second or foreign language usually speak more slowly, the slopes of the best-fit trend-lines of their intonational phrases are flatter than those of native speakers. For example, the slope of Mandarin learners' English utterances was found consistently flatter than that of native American English speakers', which had been speculated as the effect of L1 transfer [5].

However, since the declination slope of Mandarin is steeper than that of English [9], the flatter slope in Mandarin learner's English is not likely due to L1 transfer. Given research has also been scant with regards to the effect of L2 proficiency level over declination, this study aims to investigate whether duration of intonational phrase is directly related to the extent of declination, and whether L2 proficiency level plays a role in affecting the declination patterns.

## 2. METHOD

### 2.1. Subjects

Both native and non-native data were included in this study. The L2 English data were selected from a learner corpus, while native English and Mandarin data were recorded in a sound-attenuated phonetic lab under a similar test condition. The details of these data are described as follows.

### 2.1.1. The L2 data

The LTTC English Learner Corpus is a test-based corpus with data collected from the General English Proficiency Test (GEPT), a national standardized English proficiency test regularly held in Taiwan. A set of 16 audio recordings was selected from the intermediate level of the oral test. The assessment of GEPT oral tests uses a general score of a 5-point scale (1 as the lowest; 3 as the passing score). Half of the recordings selected received a grade of 3 (LL2,  $N=8$ ), and the other half received a grade of 5, which is the highest score (HL2,  $N=8$ ).

### 2.1.2. Native English speakers

As American accent is the mainstream for English education in Taiwan, a group of eight native speakers of American English was recruited for recording to serve as the native reference. They were given the same materials and instructions as the test takers, but the recording took place in a phonetics lab.

### 2.1.3. Native Mandarin speakers

To examine possible transferring effect from one's native language, another group of eight native Taiwan Mandarin speakers were also recruited, serving as the Mandarin reference. These speakers shared the same native language background as the HL2 and LL2 groups of the corpus data; furthermore, they had all been English majors back in college and were graduate students of linguistics at the time of recording. Both their production of English and Mandarin texts were recorded. These speakers were categorized as advanced learners of English (Adv-L2), whose data filled the existing gap between native English and the HL2, given that the corpus data was selected from the intermediate level of the GEPT, and there is plenty of room between the highest score and the native norm of English.

## 2.2. Materials and procedure

### 2.2.1. The English texts

The materials used in the read-aloud session of the test were two English texts consisting of 165 words in total. Each sentence was a statement, with punctuations clearly marked on the test sheet. The test sheets were given to the test takers one minute prior to recording. The same criteria held true for data collection conducted in the lab. All participants were instructed to read the two texts at a comfortable speed within two minutes.

### 2.2.2. The Mandarin texts

The two texts were translated into Mandarin following a comparable scheme: All sentences were statements with punctuations clearly marked, and alternations between short and long phrases of the original text were maintained. Instead of using transliteration, the English personal or place names in the original text were replaced by common local Mandarin names and place names in Taiwan. This was to ensure that participants would stay in the same language mode while being recorded. Moreover, to accommodate the fact that Mandarin is a tone language, and T2 and T3 in Mandarin do not contain a high tone [3], all the initial and final characters of each potential chunking site used words of either a T1 or a T4, both of which contained a high tone. This makes cross-language top-line comparison possible. The same materials and instructions were given in Mandarin during the recordings.

## 2.3. Data collection and processing

### 2.3.1. Equipment

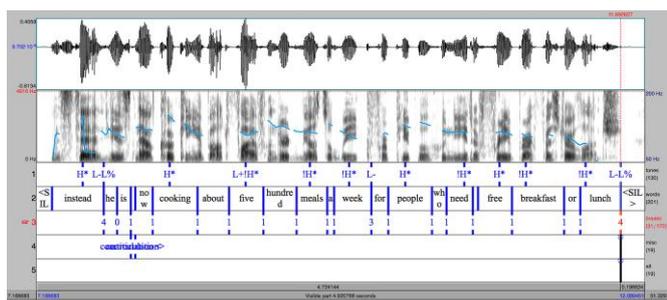
A SHURE SM10A head-mounted microphone connected to a KORG MR-1000 mobile recorder was used for data collection for both native language groups. The sampling rate was set at 48kHz and later downsampled to 22.5 kHz for analyses.

### 2.3.2. Data labelling

Following the English ToBI convention [1], two trained phoneticians independently labelled the data, and agreed on the final labels.

In addition to the *words* tier, core tiers of a standard ToBI transcription include the *tones* and the *break index* tiers. In the *tones* tier, syllables received prosodic salience (pitch accents) and utterance ends (phrase accents and boundary tones) were denoted in association with the height of the tonal targets. In this study, single high tones or bi-tonal combinations with an H element were marked for pitch accents, and L-L% for the phrase accent and boundary tones of the final fall in a declarative utterance. In the *breaks* tier, levels of break indices (BIs) were marked according to the corresponding size of disjuncture between two adjacent words. BI4 represents the maximal level of disjuncture, and always appear at the boundary of intonational phrases (IPs) in the *tones* tier. An example of the labelling is shown in Figure 1.

**Figure 1:** Demonstration of ToBI labelling.



For data analyses, the percentage of BI4s over all BIs included was calculated, and so was the duration of each IP and the maximal value of its first and last accents to extract the top-lines.

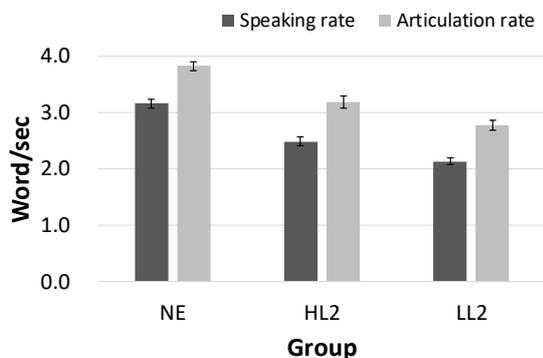
### 3. RESULTS AND DISCUSSION

Three main parts were reported in this section. First, speech rates and the number of intonational phrases of the L2 groups were compared with the native norm. Declination levels of the English and Mandarin native norms as well as those obtained from the learners' data were then compared. Finally, the effect of duration was also examined for native and non-native English production.

#### 3.1. Speaking and articulation rates

Speaking rate was calculated by dividing the total number of words over the duration spent (in seconds) for each participant, and articulation rates was calculated with all pauses within the recording excluded. The group averages of the L2 and native English data were shown in Figure 2.

**Figure 2:** Speaking and articulation rates of the native and non-native English read-aloud data (Error bars stand for standard errors).

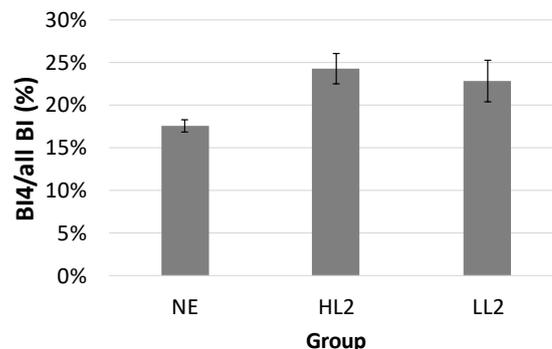


As expected, for reading the same texts, native speakers were faster than high-proficiency learners, who in turn, were faster than their lower-proficiency counterparts. Two separate one-way ANOVAs were conducted to compare the effect of language proficiency, and both reached significance [speaking rate:  $F(2, 21) = 47.85, p < .0001$ ; articulation rate:  $F(2, 21) = 31.33, p < .0001$ ]. This validates that the data of the two L2 groups belong to different populations that contrast each other in fluency.

#### 3.2. Intonational phrases

Now that the same texts were read by all participants, a higher percentage of BI4 indicates more intonational phrases were assigned. As shown in Figure 3, more IPs were assigned by the two L2 groups, as compared to the native English speakers. A one-way ANOVA examining the effect of proficiency on BI4 ratio reached marginal significance [ $F(2, 21) = 3.24, p = .06$ ], which validated the observation. However, the pairwise comparison showed that the difference between HL2 and LL2 was not significant.

**Figure 3:** The assignment of BI4/all BIs (error bars stand for standard errors).



#### 3.3. Declination

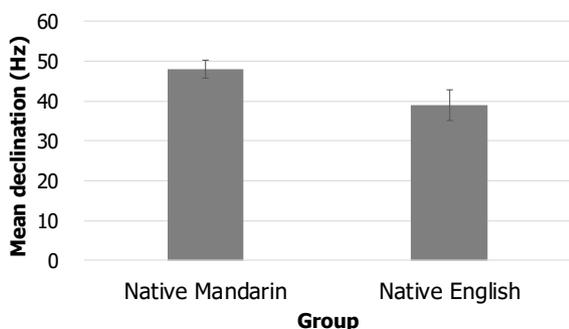
##### 3.3.1. Language-intrinsic difference

The first analysis with regards to declination was the language-intrinsic preference of the declination level for native Mandarin and English speakers. The degree of declination was calculated as the group average of the F0 difference between the first and the last peak of each IP. The results would also serve as a basis for later comparisons with L2 learners at different English proficiency levels. Declination level was defined as the difference between the maximal values of the first and the last peaks of each IP.

As shown in Figure 4, compared with English, a larger declination level was observed for Mandarin,

and a one-way ANOVA on this language-intrinsic difference proved the significance [ $t(317) = 2.04, p < .05$ ]. This supports previous findings [9].

**Figure 4:** Declination level of the two groups of native speakers (error bars stand for standard errors).



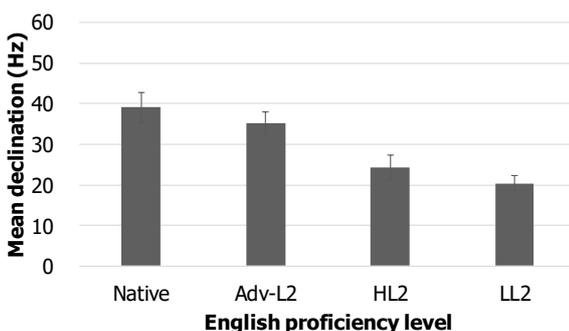
### 3.3.2. Effect of language proficiency

Based on the native data, the degree of declination was then examined for all groups at different English levels, which included the two L2 groups from the corpus, and the native English and Adv-L2 groups from the recorded data. The effect of proficiency was clearly observed.

As shown in Figure 5, the extent of declination exhibits a likely developmental path, with native English speakers employing the largest difference between the first and last peaks of the IPs, followed by the advanced learners, which was in turn followed by the HL2 and LL2 groups of the learner corpus.

A one-way ANOVA test comparing English proficiency levels revealed significance [ $F(3,659) = 10.20, p < .0001$ ]. Post-hoc Bonferroni tests showed all pairwise comparisons were significant except for the one between Native and Adv-L2, and that between HL2 and LL2.

**Figure 5:** Declination of speakers of different proficiency levels (error bars stand for standard errors).



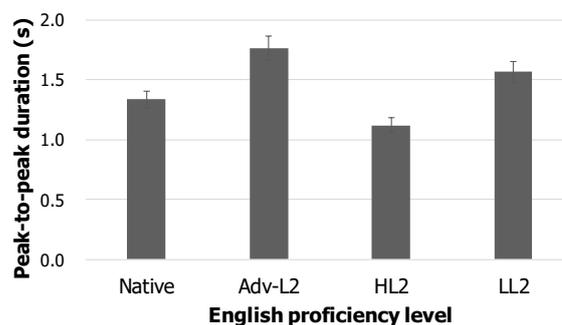
### 3.3.3. Declination vs. duration of the unit

Since proportionally more IPs were assigned by the two L2 groups from the learner corpus (Figure 3), chances are the extent of declination is affected by the mean duration of the declination units. This is because fewer chunks imply longer duration, which may lead to larger declination levels. To examine for this possibility, the mean peak-to-peak duration of the IPs was calculated for the four groups.

As shown in Figure 6, the developmental trend observed for the declination level did not appear in duration. The difference of declination level across groups with various English proficiency levels seemed not directly resulting from the duration factor.

A one-way ANOVA test on the first-to-last peak duration for each declination was significant [ $F(3,659) = 10.41, p < .0001$ ]. Post-hoc Bonferroni tests showed that the pairwise comparison between native and Adv-L2 was significant, and so did the difference between Adv-L2 and HL2, as well as that between HL2 and LL2.

**Figure 6:** Duration between first- and last peaks of the declination unit (error bars stand for standard errors).



Previous studies showed smaller amount of declination in Mandarin learners of English, as compared to native English speakers' production [5]. In this study, we further investigated learners of different English proficiency levels, and based on the results obtained, we found the extent of declination positively correlates with the learners' English proficiency level; the higher the proficiency, the greater the extent. Moreover, given declination level of Mandarin was in fact larger than that of English, as revealed by our data, the developmental trend on declination of L2 English should not be a direct result of L1 transfer, which has been the speculation of previous research [5].

#### 4. CONCLUSION

Previous studies had shown that the slope of the learners' English utterances was flatter than that of the natives [5]. This study further showed that the degree of declination positively reflects L2 proficiency. However, unlike what was previously conjectured (cf. [5]), such a result could not have come from negative L1 transfer, as Mandarin in general showed a greater extent of declination than English (also cf. [9]). More studies are needed to tease apart possible interactions between declination level and unit duration.

#### 5. REFERENCES

- [1] Beckman, Mary E., Elam, Gayle Ayers. 1993. Guidelines for ToBI Labelling. The Ohio State University Research Foundation.
- [2] Cohen, Antonie, t'Hart, J. 1967. On the anatomy of intonation. *Lingua* 19, 177-192.
- [3] Fon, J., Chiang, W.-Y., Cheung, H. 2004. Production and perception of two dipping tones (T2 and T3) in Taiwan Mandarin. *Journal of Chinese Linguistics* 32(2), 249-280.
- [4] Kutik, Elanah J., Cooper, William E., Boyce, Suzanne. 1983. Declination of fundamental frequency in speakers' production of parenthetical and main clauses. *J. Acoust. Soc. Am.* 73(5), 1731-1738.
- [5] Liu, Ching-Kang. 2006. MAE\_ToBI reflects where EFL students fail in acquiring the intonation of English. *Languages, Literary Studies, and International Studies: An International Journal* 3, 69-84.
- [6] Shih, Chilin. 2000. *A declination model of Mandarin Chinese*.
- [7] Swerts, Marc, Geluykens, Ronald. 1994. Prosody as a marker of information flow in spoken discourse. *Language and Speech* 37(1), 21-43.
- [8] Swerts, Marc, Strangert, E., Heldner, M. 1996. F0 declination in spontaneous and read-aloud speech. *Proc. 15<sup>th</sup> International Conference on Spoken Language Processing*, Philadelphia.
- [9] Yuan, J., & Liberman, M. (2014). F0 declination in English and Mandarin broadcast news speech. *Speech Communication* 65, 67-74.