Declination: A Mismatch between Taiwanese EFL Students and Native Speakers of American English in Oral Performance

Ching-Kang Liu
Department of Foreign Languages and Applied Linguistics
National Taipei University
Taipei, Taiwan 237, R. O. C.

Key Words: declination, pitch accent, intonation

Abstract

This study focuses on how the oral performance of the Mandarin-speaking EFL students in Taiwan (MSs) does not match that of native speakers of American English (AESs) in \( F_0 \) declination and locations of the pitch accents.

Seven female EFL students and seven female American English speakers who grew up in California produced, respectively, 50 expressions in context or in separate sentences. The sound recording was analyzed through PRAAT and PitchWorks for phonetic data.

The results show that the EFL students produced two specific patterns deviating from those produced by the AESs. First, the pitch accent was located at a different part of the contour, or even if it was located at the same segment, both the pattern and length of the intonation phrase differed from those produced by the AESs. Second, the AESs had consistent gradual declination in certain intonation phrases while the EFL subjects failed to produce the same declination in the same expressions. In addition, the data indicate that the EFL students and the AESs differ significantly in producing the \( F_0 \) declination when uttering the same statements.

I. Introduction

Declination is the tendency of the fundamental frequency to gradually fall over the course of an utterance. It has been treated as a common tendency in many expressions of English (Ladd, 1996). This topic, however, has rarely been included in EFL/ESL textbooks of pronunciation or intonation. Most textbooks focus on the patterns of stress, rhythm, and various types of pitch contours matching different feelings and attitudes in all possible kinds of phonological environments. Some might include the contrastive examples of how non-native speakers of English might deviate while pronouncing English vowels or consonants (Celce-Murcia, 1996). Very few chapters have involved contrastive studies between EFL/ESL and native speakers of English in the area of intonation. For instance, Celce-Murcia et. al. (1996) in Teaching Pronunciation provide detailed descriptions of different features of English intonation, but the issue of declination was not discussed in the book. Lane (2005) in Focus on Pronunciation 2 also provided as many important features of English intonation as possible in the book, but declination was not included in the contents. The aim of this study, therefore, is to examine two things: (1) whether the declination of the fundamental frequency \( (F_0) \) produced by Taiwanese Mandarin speakers (MSs) significantly differs from that produced by American English speakers (AESs), and (2) whether the differences can be located and interpreted if significant differences are found.

The data and findings presented in this study are part of a continuing study regarding the features of \( F_0 \) declination trends. In this report, only the declination of fundamental frequencies related to 15 declarative expressions will be
discussed and the discussion will be based on the
the phonetic data of short expressions (See
Appendix). While the differences between the $F_0$
declination trendlines were expected, the detailed
information concerning whether these differences
could be located and how, if possible, they could
be interpreted for the EFL students in Taiwan will
also be discussed.

II. Literature Review

Since the concept of the decrease of the $F_0$
through time was first noticed by Pike (1945) and
since this concept was introduced using the term
"declination" by Cohen & t'Hart (1967), it has
been treated as a universal principle of speech
intonation. Although some (e.g., Xu, 1997) argue
that "the existence of declination as an
independent $F_0$-determining mechanism is far
from being firmly established," most intonation
models adopt declination as a baseline upon
which more local $F_0$ patterns reside. Some
researchers have even noticed that the $F_0$
declination is acquired when the speakers are still
children (Dodane, Konopczynski, Santi, &
Espesser, 1999).

The concept of the declination of the $F_0$
in English was well elaborated by Pierrehumbert
(1980), who verified how the declination of the
$F_0$ could be accounted for as the result of
downstep—the stepwise lowering of pitch at
specific pitch accents. This line of research
continued to the conclusion that the value of each
accent peak in a downstep series is a constant
proportion of the previous peak. (Pierrehumbert
and Beckman, 1988). Ladd (1996) had a whole
section (pp. 71-78) discussing how declination in
English had been observed in earlier studies. No
matter how this topic should be presented or
interpreted, the declination of the $F_0$ is obviously
an important, or even indispensable, phonological feature in English declarative
expressions and it is worth serious research in the
EFL field in anticipation of finding information conducive to helping pronunciation instruction.

Among the previous studies, Gussenhoven
(2002) had one interesting discussion regarding
how the declination of the $F_0$ can be affected by
the way speakers' efforts are used. This seems to
have potential for the interpretation of how
different language speakers might create
different types of declination. Or more
specifically, how EFL learners might fail to
produce the $F_0$ declination of English the way
AESs do. Gussenhoven (2002) stated, "The
process of energy generation relies on the fact
that speakers appear to spend more effort on the
beginning of utterances than on the ends," so "a
natural consequence of the fall-off in energy is a
gradual drop in intensity, and a weak, gradual
lowering of the fundamental frequency." Given
that a statement is pronounced without starting
another "new generating point of energy," the
declination will occur. It sounds like a natural,
biological constraint that might directly affect
how a language user controls the airstream and
the energy while uttering a statement. For
second language learners, it can be assumed that
all the mechanisms employed in the process of
using the target language are more
"effort-taking" than when using the native
tongue. If this assumption is true, the second
language user may have to break an expression
into more junctures, resulting in the phenomenon
that it takes the EFL student a significantly
longer time to finish a statement than a native
speaker does. If the lengthened uttering time of a
statement were a possible factor highly related to
the slope of the $F_0$ declination, it would also
make sense to accept the concept that declination
would be less steep in longer utterances than in
shorter ones (Swertsy, Strangertz, & Heldnerz,
1996).

Another concept to be mentioned here is that
"the significance of declination does not lie
in its slope. Rather, it is variation at the edges that
is interpreted in terms of initiation and finality.
Thus, high beginnings signal new topics; low
beginnings, continuations of topics. A reverse
relation holds for the end of the utterance: high
endings signal continuation, low endings finality
and end of turn" (Gussenhoven, 2002). If the
"edges" do not correspond to the initiation or
finality of a sentence, they might also not
correspond to the initiation or finality of an
intonation unit or a thought group. When all the
linguistic factors like "sentence focus, topic
initiation, word stress, and lexical tone" are
involved in the forming of declination, it can be
a complicated process, with the result that,
"rather than being a basic principle of intonation,
declination is more likely to be an artifact of
various linguistic factors plus certain local
physiological constraints" (Xu, 1997).

In this study, Gussenhoven's theory was
taken as the basic support for interpreting the
declination of the $F_0$ in English. The study,
therefore, will not deal with the physiological
issues, but only with the acoustic and statistical
data of declination. The data will be analyzed
and measured according to the steepness of the
slope or the trendline of the best fit with the
pitch contour. The statement may be divided, based on phonological features, into different junctures which will be analyzed separately.

III. Method

1. Subjects
Seven female Mandarin-speaking college students and seven female American English speakers from California produced the language data, respectively. The MSs were all English majors at National Taipei University and the AESs were all college students at UC Santa Barbara. The MSs had some time to practice what they were asked to read and could repeat reading the materials if they were not satisfied with their performance. Instructions for the reading suggested that the subjects of both groups produce the utterances as naturally as possible, as if they were doing a casual talk. If unsatisfactory situations occurred, the subjects were asked to redo the recording.

2. Materials
The recorded materials included 15 separate utterances of declarative utterances (See Appendix I). These utterances were composed of simple syntactic structures (e.g., S + Vi + modifiers). These expressions were all short utterances, ranging from four words to nine words, and the situations related to these utterances were considered neutral, so that unexpected emotional tones could be avoided. In other words, the participants were told to avoid reading these expressions in an excited voice. In addition, these utterances were chosen because they could be uttered in one breath in most simple sentences. With only one breath reflecting most likely one intonation unit (IU), we can focus on other phonological features that might be attributed to the phenomenon of declination. These examples may also help interpret why the American English speakers produced the utterances with steeper best-fit trendlines than the Mandarin-speaking EFL students.

3. Measurements
Since “there are as yet no un-criticized methods available to quantitatively determine the slope and the domain of declination” (Swertsy, Strangertz, & Heldnerz, 1996), the current study will employ a process that was argued to be one of the most reliable descriptors of sentence $F_0$ contours. According to a study done by Lieberman et al. (1985), among the three different ways of calculating declination measures, the all-points line is a better descriptor of sentence $F_0$ contours, compared with the fitting of linear regression lines to local peaks (topline) or to local valleys (baseline).

In this study, a modified approach used by Swerts et al. (1996) was employed. Estimation of the slope of declination was conducted by fitting an all-points regression line to the $F_0$ points with semitone values extracted with the PRAAT phonetic analysis software. The estimated values of slopes were calculated through Microsoft Excel’s built-in statistical program. Declination was observed on the basis of the absolute value of the slope of a best-fit trendline through Excel’s built-in statistical function. Given a set of data, $(x_i, y_i)$, where $x$ refers to time and $y$ refers to the frequencies of pitch in every 10 msecs, we have reason to believe that there exists a linear relationship between the variables $x$ and $y$, so we can plot the data and draw a “best-fit” straight line through the data (trendline). Then this relationship is governed by the equation $y = mx + b$. We can then find the slope, $m$, and $y$-intercept, $b$, for the data. The R-squared value is the square of the correlation coefficient between the values of $x$ and $y$.

To illustrate what a best-fit trendline looks like, one of the seven subjects in each group was randomly selected and the trendlines of sentence (1) read by the two subjects were compared. The results show (Figure 1) that the AES formed a declination with the $m$ value of -14.586 while the MS formed another $m$ value of -2.6482, indicating that the AES produced an utterance with a steeper best-fit trendline than that produced by the Mandarin speaker. The values of $m$’s produced by the Excel statistical program are quite reliable because they matched very well with the values of slopes calculated by means of SPSS. The $m$ values, therefore, can be obtained either with Microsoft Excel or the statistical package SPSS.
The temperature is very low today.

![Graph](image)

Figure 1. The example of the best-fitting line to estimate the rate of declination in one of the utterances

IV. Results and Discussion

The data were examined in two formats. First, all the individual slope (m) values of the 15 utterances produced by the seven participants in both groups were compared using a paired-samples t test (df = 104) in SPSS. Then we did the same t test on the averaged slope values of each utterance (df = 14) produced by the seven subjects of each group (see Table 1). The results of the t test based on individual slope values show that the MSs and AESs displayed a significant difference in their oral production \((t = 5.772, p < 0.01, 2\text{-tailed})\). The results of a t test to compare the means of slopes of the two groups, again, showed that these two groups differ significantly \((t = 4.453, p < 0.01, 2\text{-tailed})\). These results certainly verify that the declination of the fundamental frequency \((F_o)\) produced by MSs significantly differs from that produced by AESs.

<table>
<thead>
<tr>
<th>Sentences</th>
<th>MSs</th>
<th>AESs</th>
<th>Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>-1.88</td>
<td>-6.40</td>
<td>4.52</td>
</tr>
<tr>
<td>(2)</td>
<td>-2.88</td>
<td>-8.46</td>
<td>5.58</td>
</tr>
<tr>
<td>(3)</td>
<td>-2.20</td>
<td>-5.66</td>
<td>3.46</td>
</tr>
<tr>
<td>(4)</td>
<td>-4.09</td>
<td>-11.32</td>
<td>7.23</td>
</tr>
<tr>
<td>(5)</td>
<td>-6.33</td>
<td>-6.77</td>
<td>0.44</td>
</tr>
<tr>
<td>(6)</td>
<td>-2.05</td>
<td>-10.83</td>
<td>8.78</td>
</tr>
<tr>
<td>(7)</td>
<td>-4.58</td>
<td>-5.11</td>
<td>0.53</td>
</tr>
<tr>
<td>(8)</td>
<td>-2.44</td>
<td>-7.08</td>
<td>4.65</td>
</tr>
<tr>
<td>(9)</td>
<td>-1.00</td>
<td>-1.15</td>
<td>0.15</td>
</tr>
<tr>
<td>(10)</td>
<td>-2.06</td>
<td>-3.29</td>
<td>1.24</td>
</tr>
<tr>
<td>(11)</td>
<td>-1.55</td>
<td>-3.55</td>
<td>2.00</td>
</tr>
</tbody>
</table>

Table 1. The Average m Values and the Differences between the Two Sets of Data (D = AES-MSs)
1. The Duration of the Utterance

When observing the averaged \( m \) values for the 15 utterances in Table 1, we not only noticed that the MSs consistently produced oral utterances with less steep declination lines than did the AESs, but also noticed that some utterances produced by the AESs have steeper \( F_0 \) trendlines than those of the MSs, while others don’t. For instance, sentences (4) and (6) reflect sharper differences between the two groups (with the mean differences of \( m \) values: 7.33 and 8.78, respectively), but sentences (5), (7), (9) and (13) show very little difference between the two groups (with the mean differences of \( m \) values: 0.44, 0.53, 0.15, and 0.36, respectively).

By closely observing the data for sentence (7), we noticed that the length of the utterance might contribute to the discrepancies in the duration of the utterance. To figure out whether duration could be a possible factor that might affect the slope values of the best-fit trendlines, the data on \( F_0 \) frequencies were rearranged based on the syllables of the utterance instead of on the utterance’s intonation contour. For each syllable, the initial and final \( F_0 \) values were extracted from the pitch listing values provided by PRAAT. In Figure 2, the slope values of the best-fit trendlines based on the \( F_0 \) frequencies of the syllables are almost identical (-10.794 (MS) vs. -10.753 (AES)) if the trendlines were drawn on the basis of the pitch values extracted from the initial and ending points of each syllable. In so doing, the factor of time was eliminated and only the regression lines of the scatterplots of the pitch values served as the data.

![Figure 2](image-url)

**Figure 2.** The best-fit lines based on the normalized \( F_0 \) frequencies of the initial and ending points of each syllable of sentence (7) *They show me everything* for both Mandarin speakers (above) and American English speakers (below)
When we looked at the best-fit trendlines based on the $F_0$ frequencies (semitones) in Figure 3, we noticed that the AES produced a steeper trendline than did the MS. The reason is straightforward: The duration of the utterance produced by the MS is longer (1.51 seconds) than that produced by the AES (1.06 seconds). The mathematical concept underlined could be interpreted by examining the two buildings (Buildings A and B) in Figure 4: These two buildings with the same height will require different distances to reach the top if different lengths of the ground spaces are given. In other words, the shorter the distance to reach the top, the steeper the line will be. That is to say, the shorter line $b$ is, the steeper line $a$ will be.

In this study, we have noticed that AESs consistently produced 15 utterances with steeper declination lines than did MSs. At the same time, the AESs also consistently finished these utterances in a significantly shorter time than the MSs ($t = 11.407, p < 0.001, 2$-tailed) (see Figure 3). When examining the correlation between the length of the utterances and the differences between the slope values of the best-fit trendlines, we noticed that the correlation is also significant ($r = 0.609, p = .000$). In other words, we can say that the length of MSs’ utterances can be a very good predictor of the slope values of the best-fit trendlines.

![Figure 4. The two graphs show that $a_1$ is steeper than $a_2$ because $b_1$ is shorter than $b_2$.](image)
The average length of each utterance

![The average length of each utterance](image)

Figure 5. The average duration (of the seven subjects in each group) of the 15 utterances

2. The Location of the Pitch Accent

The second possible predictor for values of the slope might be the differences between the average F0 frequencies of the first half and those of the second half of each utterance. The significant correlation coefficient between the average slope values of the 15 utterances and the

![Graph showing the collective F0 contours of sentence (4)](image)

Figure 6. The collective F0 contours of sentence (4) *Give it to her now* produced by the group of seven MSs

When examining Figures 6 and 7 closely and splitting each of the graphs into two parts—the left half and the right half—we noticed that almost all the AESs tended to decline their utterance at about one-third of the utterance, making each utterance move down sharply in the second half, while the MSs did not show such a clear tendency.

![Graph showing the collective F0 contours of sentence (4)](image)

Figure 7. The collective F0 contours of sentence (4) produced by the group of seven MSs
There are some exceptions. In utterance (4) one Mandarin-speaking subject produced the utterance with a larger ratio between the 1st and 2nd half of the utterance than one American English speaking subject. But the slope of the best-fit trendline produced by the AES is still larger than that of the MS. Figure 8 shows two possible contributing factors to this result. First, the AES spent a shorter time (0.8 second) than the MS did (1.14 seconds) completing the utterance. As we have discussed in the previous section, shorter duration could be a factor that affects the slope value of the best-fit trendline.

![Figure 8. The best-fit trendlines of sentence (4) with the MS's 1st-2nd differences larger than the AES's](image)

In addition, when we observed the contours of sentence (4) produced by any randomly selected subject from each group, we noticed that the American English speakers uniformly put their high pitch accents (H*) on “to” or “it” while five of the seven Mandarin-speaking subjects put the high pitch accent on the word “give” and only two put it on “to” or “it” (see Table 2).

Table 2. The AES-MS differences between the two sets of average m values: The pitch accent set (H*) and that of the end of the intonational phrase

<table>
<thead>
<tr>
<th>ID</th>
<th>L- (MSs) F₀ (Hz)</th>
<th>L- (AESs) F₀ (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>190.34</td>
<td>81.09</td>
</tr>
<tr>
<td>2</td>
<td>162.66</td>
<td>175.53</td>
</tr>
<tr>
<td>3</td>
<td>77.27</td>
<td>153.8</td>
</tr>
<tr>
<td>4</td>
<td>179.16</td>
<td>146.43</td>
</tr>
<tr>
<td>5</td>
<td>206.09</td>
<td>170.51</td>
</tr>
<tr>
<td>6</td>
<td>77.99</td>
<td>169</td>
</tr>
<tr>
<td>7</td>
<td>165.92</td>
<td>75.31</td>
</tr>
<tr>
<td>Mean</td>
<td>151.35</td>
<td>138.81</td>
</tr>
</tbody>
</table>

Table 2 also shows that the averaged F₀ frequency value of the high pitch accents produced by the AESs is larger than that produced by the MSs, and the averaged F₀ frequency value of the lowest ending syllables produced by the AESs is smaller than that produced by the MSs. Figures 9 and 10 provide a visualized illustration of the locations of high
pitch accents and the slope values of the best-fit trendlines of the utterance.

Figure 9. The $F_0$ contours of sentence (4) produced by two subjects randomly selected from each of the two groups

Figure 10. The best-fit trendlines of sentence (4) produced by the same speakers as in Figure 9

V. Conclusion

This study has so far provided strong support for two positive answers to the research questions. First, the declination of $F_0$ produced by the MSs significantly differs from that produced by the AESs, and the $F_0$ declination line created by the AESs is consistently steeper than that created by the MSs in every expression. Second, the differences can be located and interpreted. For the second research question, the findings of this study lead to the following conclusions.

The MSs took a longer time to finish reading each of the 15 expressions. This obviously resulted from the phenomenon that the AESs tended to use more clustering/liaison while the MSs tended to keep every phoneme clearly pronounced.

As for the location of the pitch accent, in each chunk or intonation unit, the AESs tended to place the pitch accent on a syllable different from where the MSs tended to place it. For instance, in the chunk “the temperature is very low...”, all the AESs put their pitch accent on the syllable “-pera-” in the word “temperature” whereas all the MSs put their pitch accent on the syllable “tem-” in “temperature.”

When the utterances were split into two halves or into different intonation units, the AESs tended to have a steeper $F_0$ declination line compared with that produced by the MSs in each chunk or intonation unit.

The above findings will hopefully provide English instructors or EFL students with information on subtle qualities of English prosodic features. These findings may also help create intonation models for different types of pitch contour for expressions carrying different meanings semantically and pragmatically. Language instruction in the field of intonation can, accordingly, focus more on the length of the uttering time and where and how high to put the high and low pitch accents. Paying attention to
the specific location might help language learners put the right pitch at the right places. Declination is only the natural result of an utterance; however, declination itself is not the target of language teaching. It should be obtained if other things are done. And this study provides three possible factors that might contribute to the result of declination: the duration time of an utterance, the ratio of the first half and the second half of the $F_0$, and the location of the high pitch accents.

Notes

1 Given a set of data $(x_i, y_i)$ with $n$ data points, the slope, $y$-intercept and correlation coefficient, $r$, can be determined using the following:

$$m = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

2 In order to make sure that the slope values ($m$) of the best-fit trendlines provided by the statistical function of the Excel, twenty sets of listed semitone values of $F_0$ were randomly selected and retested through the Statistical Package SPSS. The slope values calculated by the two packages are identical.

References


下傾現象：台灣學生與英語母語人士的差異

劉慶剛
國立台北大學應用外語學系

摘要

本研究主要觀察比較台灣學習英語的學生和美國英語母語人士，在口語表現方面，是否有語調線(best-fit trendlines)傾斜度及重音落點上的差異。七位台灣的大學生和七位在美國加州土生土長的大學生分別錄製五十個單句或篇章的句子，錄製的語料則藉 PRAAT 和 PitchWorks 兩個語音分析軟體擷取相關數據。結果顯示台灣學生有兩項特質與英語母語人士有顯著不同的表現。第一個不同點是重音的落點，就變落點相同，其他與語調落點相關的語調單位長度及分割點都有明顯差異。第二個不同點是英語母語人士之語調線斜率均大於台灣學生之語調線斜率，且兩者之間有顯著差異。

關鍵字：語調線斜率、重音落點、語調。