Learning L2 Prosody Is More Difficult than You Realize–

F0 Characteristics and Chunking Size of L1 English, TW L2 English and TW L1 Mandarin

Chiu-yu Tseng[1] *& Chao-yu Su* [1,2]

[1] Phonetics Lab, Institute of Linguistics, Academia Sinica, Taipei, Taiwan
[2] Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan
cytling@sinica.edu.tw

Abstract

We compare the F0 output and chunking size of speech units of L1 English, TW L2 English and L1 Mandarin to see what some of the major intrinsic prosodic difference could be, what prosodic features could account for TW L2 English and in what way prosodic transfer occurs. Results show that the fundamental prosodic difference of the two languages is how in the pitch domain English requires sharper high/low contrast by higher-level prosodic units but less such contrast in accentuating lower-level prosodic units whereas Mandarin patterns are the exact opposite. Explanations are provided regarding how TW L2 English differs from L1, why prosodic transfer merits detailed analysis, and why mastering English prosody is especially difficult.

Index Terms: F0 contrast, contrast degree, prosodic units, prosodic differentiation and chunking size.

1. Introduction

There has long been ample research attention on comparative analysis of L1/L2 characteristics, though mainly from the pedagogical perspective of how to learn and master the target L2. The focus had been mainly on segmental features. However, user-oriented innovations of speech technology in recent years have led to the realization that L2 English speakers whose population outnumbers those of L1 make up a larger market than learners of L2 English; their needs require separate and different tailoring. Development of computer aided language learning (CALL) tools and language identification became issues of attention. To this end, more understanding of the major linguistic and speech characteristics of L2 English is needed; these features are by no means limited to segmental features only [1, 2].

Among the existing L2 English population, Asian English makes up a community that distinguishes itself from the other L2 Englishes in at least the following three aspects: (1) Phonologically, most of Asian languages are tonal or pitch accent, either syllable- or mora-timed; among them the best known examples include Chinese, Japanese, Vietnamese and Thai. (2) Syntactically, Asian languages are highly varied, for example the word order of Chinese is SVO (subject-verb-object) while Japanese and Korean are SOV. (3) Orthographically, Asian languages are more varied than the other languages in the world. For instance, Indonesian and Malaysian adopted the Roman alphabet, Korean, Thai and Vietnamese use non-Roman alphabet, Chinese is well known for its logographic system, and Japanese uses a mixture of non-Roman alphabet with the Chinese logographs. Phonological awareness of the tonal and rhythmic aspects shared among the majority of Asian population spells out clearly that any speech technology or CALL developed for Asian L2 Englishes at large will have to address features that are common to Asian English in general, and characteristics that are specific to each and every Asian language in particular. In other words, the prosodic aspects are by no means to be overlooked since their phonological role is as important as the segmental ones. As a matter of fact, more understanding of the prosodic properties of Asian L2 English should help facilitate technology innovations significantly. The awareness has motivated a consortium of the region that aimed at collecting speech corpora and consolidating research results that would hopefully facilitate some of the goals over time; the AESOP (Asian English Speech cOrpus Project) team is by far the largest action force [3]. Our group is a member the AESOP consortium concentrating on Taiwan (TW) L2 English features; our research focus has been on comparative analysis of prosodic properties of TW L2 English with L1 American English and L1 Mandarin in the hope to tease apart distinct differences between L1 and TW L2 English on the one hand, and explains in what way TW Mandarin triggers and/or transfers to TW L2 English on the other.

We have studied the acoustic characteristics of on-focus/no focus contrasts of narrow focus [4] as well as English lexical (word) stress [5] and found that in both cases similar F0 and duration patterns were found between TW L2 English and L1 American English, but the difference was mainly on the degree of contrasts in that TW L2 English always exhibited smaller degrees of contrasts than those produced by L1, especially with respect to F0 contrasts. We believe this is a major reason of why TW L2 English sounds flatter overall; and asserts that prosodic under-differentiation is a major characteristic of TW L2 that may be useful to language differentiation and CALL development. Interestingly, in a previous study of lexical stress where speech data of both TW and Beijing (BJ) L2 English were compared to L1 English, we found that similar patterns of F0 and duration were shared by all three populations; both TW and BJ L2 English exhibited less degree of contrasts; the least contrast was found in BJ L2 English [6]. In other words, prosodic differentiation of both TW and BJ L2 English is not as distinct as L1. More interestingly, the less differentiated duration patterns found in TW and BJ L2 English are not shared by Vietnamese (VN) Australian L2 English who exhibited longer syllable duration and higher F0 [7]. The results suggests that similar patterns shared by TW and BJ L2 English may largely be attributed to dialectal L1 difference between TW Mandarin and BJ Putonghua; while the difference between the two Chinese dialects and VN cautions overgeneralization of effects or

transfer caused by tones and syllable-timed rhythmic patterns which all three languages share.

Since lack of sufficient F0 contrasts is by far the most significant difference that distinguishes TW L2 English from L1, we are interested to know what causes the TW speakers to do so. Moreover, how much of the F0 realization by TW speakers can be attributed to tones and intonation patterns from TW Mandarin [8]. In the present study we will compare the F0 output and chunking size of speech units of L1 English, TW L2 English and L1 Mandarin. Specific research questions are (1) what may be some of the major intrinsic prosodic difference between Mandarin and English in addition to Mandarin tones, (2) in what additional and specific ways TW L2 English sounds different from L1 English and (3) whether indeed prosodic transfer from L1 Mandarin occurs.

Speech Materials and Annotation

1.1. Speech data

Read speech of L1 English, L2 English and L1 Mandarin are used. The materials of English speech are 2 reading tasks of the AESOP-ILAS (Asian English Speech cOrpus Project—Institute of Linguistics Academia Sinica) corpus [3]: (1) reading of the passage of "The North Wind and the Sun" at normal speech rate and volume. The passage contains a total of 3 paragraphs which can be broken down to 5 sentences consisting of 5 dependent clauses and 8 independent clauses, or 113 words (144 syllables). (2) Reading sentences that consist of question elicited focuses at specific phrasal and sentential positions corresponding to broad and narrow focus, respectively. For example, *Context: Will 3-day delivery be fast enough?* Reply: "No. We need OVERNIGHT delivery" where the provided context requires the answer to disambiguate. Speech data of 10 L1 North American English speakers (5 male and 5 female) and 10 Taiwan L2 speakers (5 male and 5 female) were analyzed. The materials of Mandarin speech are also 2 reading task: (1) reading of 26 discourse pieces coded CNA in the COSPRO database [9] (approximately 55 min/11600 syllables/85MB). (2) Simulating broadcast of weather forecast coded WB (approximately 45 min/7070 syllables/50MB). Speech data of 1 male and 1 female native speaker of Taiwan Mandarin were analyzed.

1.2. Preprocessing and annotation

The speech data of L1 English, L2 English and L1 Mandarin were tagged in layers. The preprocessing layer is force-aligned segments by the HTK Toolkit followed by manual spot-checking by trained transcribers. Discourse units were manually tagged independently for 5 levels of perceived discourse prosodic boundaries B1 through B5; and 5 levels of prosodic units the syllable (SYL), the prosodic word (PW), the prosodic phrase (PPh), the breath group (BG, a physio-linguistic unit constrained by change of breath while speaking continuously) and the multiple phrase speech paragraph PG. By default the boundary breaks, prosodic units and their relationship are SYL/B1<PW/B2<PPh/B3<BG/B4<PG/B5.

2. Method

The Command response model [10], commonly known as the Fujisaki Model, is used for comparison of the magnitude, variety and contrast degree of F0 characteristics. The model assumed that surface F0 is superimposed by 3 major components: baseline, phrase component and accent component which represent register of the speaker, local hump (Aa) and global declination (Ap) of F0 contours, respectively. Ap and Aa denote magnitude/strength of phrase and accent component. As a result, the model made possible normalizing speaker register and decomposing output F0 trajectory into a combination of super-positioning of the global contour of a larger unit onto the specific accentuation of a local unit. In particular, the tone component Aa has been widely adopted to represent syllabic tones of tone language thereby separating the effects from tones from intonation and provides evidence of why tones and intonation of tone languages could not be interpreted as F0 contours at face value. Ap and Aa are derived by our self-developed Ap/Aa extraction software for F0 contour. For Mandarin, Mixdorff's filter-based method [11] is adapted for our extraction where Alpha is set to 2. For English, our system extracts phrase components by approaching bottom line of F0 contour with positive Aa values. Alpha is set to 3 for Ap extraction of English. Magnitude, variety and contrast of extracted Aps and Aas are calculated. Figure1 is an example of Ap extracted in Mandarin. We calculate mean/ standard deviation of Aps to represent magnitude and variety of Aps. Contrast is defined as average of pairwise variability between neighborhood Ap pair (neighborhood arrow pairs in Figure1).

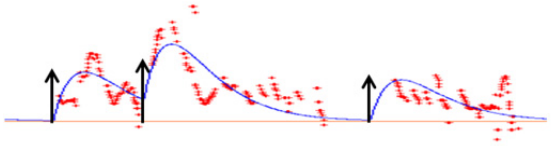

Figure1. An example of Ap extraction by an speech paragraph of Mandarin. Red curve, blue curve and black arrows denote original F0 contour, optimized phrase components and Aps respectively.

The same rational is also applied on Aas that magnitude, variety; and contrast degree of Aa within each prosodic phrase are derived for comparison.

Chunking size is also examined among L1 English, L2 English and L1 Mandarin for comparison. Speech chunk is defined by annotated prosodic units the word (PW), the prosodic phrase (PPh), the breath group (BG) and the multiple phrase speech paragraph (PG), whereas the size of chunk is defined by the number of syllables of each corresponding prosodic unit.

3. Results

In order to examine (1) Mandarin-English Prosodic difference, (2) source of L2 accent and (3) possible prosodic transfer, L1 English, L2 English and L1 Mandarin are compared by magnitude/variety/contrast of commands and chunking size as follows.

3.1. Magnitude, Variety and Contrast of Command

3.1.1. Ap

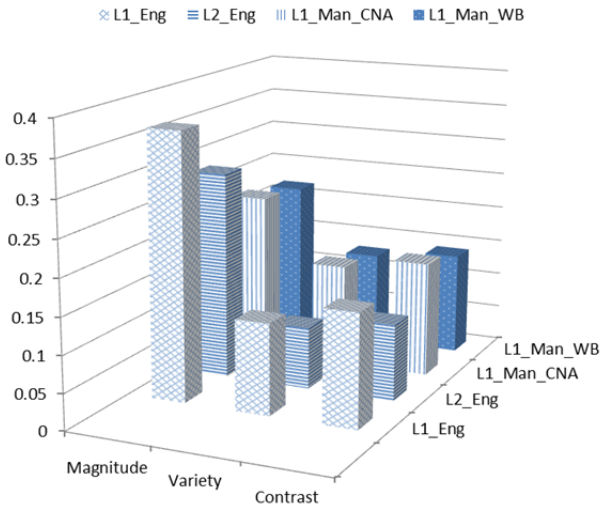

Figure2. Diagrammatic comparison among L1 English, L2 English and L1 Mandarin by magnitude, variety and contrast of Ap

Table1. List of magnitude/variety/contrast value by Ap among L1 English, L2 English and L1 Mandarin

Lan/Type / Feature	English		Mandarin	
	L1	L2	CAN	WB
Magnitude	0.365	0.285	0.227	0.217
Variety	0.125	0.085	0.140	0.128
Contrast	0.155	0.105	0.158	0.141

By examining the magnitude/variety/contrast of Ap that represents overall global declination of each speaking chunk (prosodic unit), results show that the greatest and lowest magnitude occurs in L1_Eng and L1_Man_WB; the greatest and lowest variety occurs in L1_Man_CNA and L2_Eng; the greatest and lowest Contrast occurs in L1_Man_CNA and L2_Eng. By comparing magnitude/variety/contras of Ap by language pair, L1 English is 1.64/0.93/1.04 times to L1 Mandarin; L1 English is 1.28/1.47/1.48 times to L2 English. L1 Mandarin is 0.78/1.58/1.42 times to L2 English.

3.1.1.1 Discussion

The L1_Man/L1_Eng comparison shows the effect of global declination of English is greater than Mandarin; however, similar degree of variety/contrast is found between English and Mandarin. Even though absolute values (magnitude) differ by language, relative values (variety/contrast) show common properties of higher-level F0 components across languages. In other words, high degree of contrast by higher-level F0 components is language-independent which according to the command response model is attributed to physiological factor of breathing and release allocation of overall source of energy. While the L1 data showed how physiology caused effects do not differ, the L1_Eng/L2_Eng comparison shows insufficient degree of L2 speech by magnitude/variety/contrast of higher-level F0 components and appears to be one source of distinct TW L2 accent of overall flatness. In addition, L2_Eng/L1_Man comparison shows variety/contrast is not as great as mother tongue and magnitude is slightly more salient than mother tongue by higher-level F0 components. In general, phrase/utterance-level prosody of TW L2 speech is not equal to mother tongue Mandarin, and it means no significant prosodic transfer was found by higher-level F0 components.

3.1.2. Aa

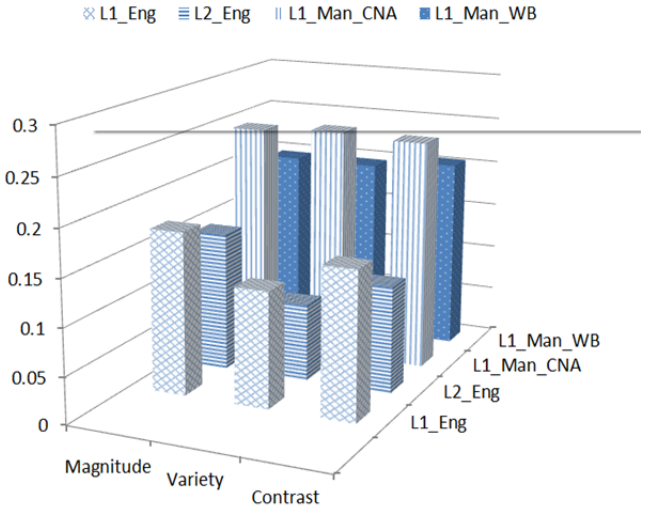

Figure3. Diagrammatic comparison among L1 English, L2 English and L1 Mandarin by magnitude, variety and contrast of Aa

Table2. List of magnitude/variety/contrast value by Aa among L1 English, L2 English and L1 Mandarin

Lan/Type / Feature	English		Mandarin	
	L1	L2	CNA	WB
Magnitude	0.174	0.149	0.246	0.196
Variety	0.124	0.083	0.251	0.196
Contrast degree	0.158	0.115	0.247	0.205

By examining the magnitude/variety/contras of Aa that represents local emphasizing, results show that the greatest and lowest magnitude occurs in L1_Man_CNA and L2_Eng; the greatest and lowest variety occurs in L1_Man_CNA and L2_Eng; the greatest and lowest contrast occurs in L1_Mandarin and L2_English. By comparing the magnitude/variety/contras of Aa by language pair, L1 English is 1.17/1.49/1.37 times to L2 English; L1 English is 0.79/0.55/0.70 times to L1 Mandarin. L1 Mandarin is 1.48/2.69/1.97 times to L2 English.

3.1.2.1 Discussion

The L1_Man/L1_Eng comparison shows Mandarin with greater magnitude and degree of variety/contrast than English. In other words, Mandarin exhibits more salient effect by local F0 components than English which could be attributed to manipulation to produce tones. The L1_Eng/L2_Eng comparison shows insufficient degree of L2 speech by magnitude/variety/contrast of local F0 components and it appears to be another source of TW L2 accent of insufficient at the word level as well. The L2_Eng/L1_Man comparison shows weaker magnitude/variety/contrast of TW L2 English speech than mother tongue Mandarin by word/syllable-level

F0 movement. In general, local F0 variation of TW L2 English is not equal to their mother tongue Mandarin suggesting no significant prosodic transfer is found in the local F0 component which corresponds to tones.

3.2. Chunking Size

Table 2 summarizes size of chunking across language and speech type.

Lan/Type / Prosodic unit	English		Mandarin	
	L1	L2	CNA	WB
PW	3.5	3.0	2.3	2.3
PPh	8.3	5.0	7.4	9.6
BG	18	21	40.9	38.8
PG	38	38	81.0	97.3

By examining the size of prosodic chunking across language/speech type ("type" here refers to speech genre, format and/or format) the largest size by order of prosodic units PW, PPh, BG and PG appears in L1_Eng, L1_Man_WB, L1_Man_WB and L1_Man_WB; the smallest size by PW, PPh, BG and PG appears in L1_Man_CNA/WB, L2_Eng, L1_Eng and L1/L2_Eng. By cross-type examination within Mandarin, the size of higher-level prosodic units (BG/PG) appears to be more type-specific than lower level units (PW/PPh). The type-specific chunk size echoes our previous results of L1 Mandarin university classroom lecture where the size of PG/BG is 7.8/2.5 times to the Mandarin CNA data used for the present investigation [12]. Comparison of the unit PW between English and Mandarin reveals PW in Mandarin is slightly smaller than English. Comparison between L1 English and L2 English reveals L2 speakers tend to speak in shorter PPh, a result of smaller planning unit during speech production [13].

3.2.1. Discussion

Comparison between L1_Man and L1_Eng shows difference of chunking size in PW, BG and PG with PPh being the only exception. Further, between-type comparison in the two types of read Mandarin speech demonstrates that the size of higher-level prosodic units (BG/PG) are type-related than lower-level units (PW/PPh). The present results echo our previous findings that distinguish read speech from spontaneous university classroom lecture [9], and provide further evidence of the significance of speech type with respect to higher level prosodic units in speech planning. Excluding type- related chunking/planning in higher-level (BG/PG), the major difference between Mandarin and English is shorter prosodic word in L1 Mandarin. PW/PPh comparison between L1 and L2 English shows difference is shorter phrases in TW L2 English. These results demonstrate that one possible source of TW L2 accent might be caused by relatively shorter planning of PPh. By comparison between L1_Man and L2_Eng by PW/PPh, no similar chunking size was found. In summary, given the simple structure of the English data, the chunking size of L1 speakers is still larger than that of L2. On the other hand, the L1 Mandarin data of chunking size provides evidence of how much reduced the planning unit becomes when producing L2.

4. General Discussion

Through comparison between L1 Mandarin and L1 English, we found that some of the major underlying prosodic difference by F0 patterns between Mandarin and English could be attributed to (1) the intrinsic global slope of higher-level (larger) prosodic units in English requires sharper high-to-low (declination) contrast than Mandarin while the lesser degree of F0 raising is required for lower-level (smaller) prosodic units. (2) Reversed patterns are found in Mandarin. Contrary to English, while higher degree of F0 raising is required for lower-level (smaller) prosodic units, the global contour of higher-level (larger) units only requires a much flatter slope than English. However, we would like to point out here that though more manipulation of local F0 contours is needed to produce Mandarin, it does not imply easier facilitation of learning English lexical stress because once again stress-induced F0 differentiation requires much higher degree of high/low contrasts [5] while tone-induced F0 differentiation is mutually exclusive discrimination patterns instead [14]. In addition, another major prosodic difference between Mandarin and English by planning unit appears to be shorter prosodic word in Mandarin which could be attributed to the role of the syllable in Mandarin morphology and word formation. We believe the above results provide both the reason and explanation of why English prosodic properties are especially difficult for Mandarin learners to acquire.

Comparison between L1 English and TW L2 English shows how under-differentiation of L2 speech was found not only in higher-level F0 components in terms of F0 height and slope but also local F0 hump. Relatively shorter planning size of the prosodic phrase also contributes to TW L2 English. By comparison between L1 Mandarin and L2 English, no significant prosodic transfer was found by higher-level prosodic components, local F0 hump and size of planning unit.

5. Conclusion

In the present study, we performed cross-linguistic/L1-L2 comparisons of English and Mandarin by F0 contours across different size of prosodic unit as well as by local accentuation triggered F0 raising on lower-level smaller-sized prosodic units in order to attribute more accurately their respective contribution to output F0 patterns. The results showed how output F0 contours taken at face value may be misleading; why the assumption of tone transfer merits further investigation. Overall English requires sharper high/low contrasts by phrase and by words while TW Mandarin, in spite of syllabic tones, does not rely on contrast degree to realize phonological as well as prosodic differentiation. The current study also explains why and how TW L2 English sounds flatter than L1 English and how it can be used to distinguish the two kinds of English. In short, we believe our results have shed some lights on some of the intrinsic prosodic differences between Mandarin and English; prosodic transfer is more complex than transplanting any independent feature at face value and learning L2 prosody can be more taxing than assumed. We would like to continue further investigations on L2 prosody along the same vein, and implement our findings in technology development in the future.

6. References

[1] Magen, H.S., "The perception of foreign-accented speech", Journal of Phonetics, vol. 26, 381-400, 1998.

[2] Anderson-Hisieh, J., Johnson, R. and Koehler, K. "The relationship between native speakers judgments of nonnative pronunciation and deviance in segmentals, prosody and syllable structure", Language Learning 42: 4 529-555, 1992.

[3] Visceglia, T. Tseng, C-Y. Kondo, M. Meng, H. and Sagisaki, Y. "Phonetic aspects of content design in AESOP (Asian English Speech cOrpus Project)", Oriental COCOSDA 2009 6 pages. Beijing, China, 2009.

[4] Visceglia, T. Su, C-Y. and Tseng, C-Y. "Comparison of English Narrow Focus Production by L1 English, Beijing and Taiwan Mandarin Speakers", Oriental COCOSDA 2012 47-51. Macau, China, 2012.

[5] Tseng, C-Y. Su, C-Y. and Visceglia, T. "Underdifferentiation of English Lexical Stress Contrasts by L2 Taiwan Speakers", Slate 2013 164-167. Grenoble, France, 2013.

[6] Visceglia, T. Tseng, C-Y. Su, C-Y. and Huang, C-F. "Realization of English Narrow Focus by L1 English and L1 Taiwan Mandarin Speakers", The 7th International Congress of Phonetic Sciences. Hong Kong, China, 2011.

[7] Mixdorff, H. and Ingram, J. "Prosodic analysis of foreign-accented English", Proc. Interspeech 2009, 6-10 Sep. Brighton UK, 2009.

[8] Tseng, C-Y and Su, Z-Y. "What's in the F0 of Mandarin Speech – Tone, Intonation and beyond", ISCSLP 2008 45-48. Kunming, China, 2008.

[9] Tseng, C-Y, Cheng, Y-C and Chang, C-H. "Sinica COSPRO and Toolkit ― Corpora and Platform of Mandarin Chinese Fluent Speech", Oriental COCOSDA 2005 23-28. Jakarata, Indonesia, 2005.

[10] Hirose, K. Fujisaki, H. and Yamaguchi, M. "Synthesis by rule of voice fundamental frequency contours of spoken Japanese from linguistic information", IEEE, 1984.

[11] Mixdorff, H. "A Novel Approach to the Fully Automatic Extraction of Fujisaki Model Parameters", ICASSP 2000, vol. 3, pages 1281-1284, Istanbul, Turkey, 2000.

[12] Tseng, C-Y. Su, Z-Y. and Lee, L-S. "Mandarin Spontaneous Narrative Planning—Prosodic Evidence from National Taiwan University Lecture Corpus", Interspeech 2009. 2943-2946. Brighton. 2009.

[13] Tseng, C-Y, Su, Z-Y, Huang, C-F and Visceglia, T. "An Initial Investigation of L1 and L2 Discourse Speech Planning in English", ISCSLP2010, 55-59. Tainan/Sun Moon Lake, Taiwan, 2010.

[14] Tseng, C-Y. "Beyond Sentence Prosody", Interspeech2010. Makuhari, Japan, 2010.