Where and How to Make an Emphasis? –

L2 Distinct Prosody and Why

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

[1] Phonetics Lab, Institute of Linguistics, Academia Sinica, Taipei
[2] Taiwan International Graduate Program, Academia Sinica, Taipei

cytling@sinica.edu.tw

Abstract

It has been reported in the literature that L2 English prosody differs from L1 at the lexical, syntactic and discourse levels; characterized by under-differentiated word stress and narrow focus as well as smaller discourse units, respectively. Using continuous speech data of L1 TW Mandarin, L1 English and TW L2 English, the present study compares information-structure related L2 prosody through emphasis placement in phrases and in multi-phrase paragraphs. Results show that TW L2 English is marked by fewer and less varied emphasis patterns while emphasis placement is only related to one specific genre of L1 Mandarin. Acoustic analyses also reveal that the contrast strength of L2 produced emphases is again less robust and therefore under-differentiated than L1. These findings suggest that sources of L2 prosody are multi-fold, contributions from constraints of speech planning at lexical, sentential and discourse and information levels may all take part. We believe our findings can be applied to L2 English teaching as well as CALL technology development.

Index Terms: information structure, emphasis, focus, L1 English prosody, L2 English prosody, lexical prosody, syntactic prosody, discourse prosody, information prosody, contrast patterns, contrast strength, robustness, differentiation, under-differentiation

1. Introduction

Learning more about the phonetic aspects of Asian English, not exclusively the segmental aspects, have drawn more attention in recent years [1]. The assumption is that since the phonology of a large number of Asian languages includes lexical tones and syllable timing, it is of interest how these features may interact with L2 English produced by speakers of these languages, and how we could achieve better understanding of some of the less known prosodic features lf L2 English produced by Asians as well, especially with respect to discourse and pragmatic properties. Among these L2 Englishes, Taiwan L2 English should serve as a good example since the most used language in Taiwan is Mandarin Chinese. Reported studies on Taiwan (TW) L2 English discourse prosody have addressed issues related to discourse organization [2, 3], sentential prosody from syntax elicited narrow focus [2, 3] and lexical prosody from word stress [4]. Since The major distinction of TW L2 discourse prosody is smaller unit size and flatter intonation contour; the former a common cross-linguistic L2 feature but the latter may be TW specific [2]. Interestingly, the major finding of TW L2 narrow focus is not whether the contrastive stress is correctly placed and produced, but rather less robust realization of on-focus/no-focus contrast shown most significantly in F0 high vs. low (H/L) difference in both directions. We argued that the under-differentiated contrast patterns contribute significantly to why TW L2 English may sound flat and less expressive. The results also suggest that prosodic differentiation may be one specific difficulty to TW learners and may surface in other respects of TW L2 English as well. Hence we believe the issue merits further investigation. [2, 3] and subsequently investigated lexical prosody of English polysyllabic word stress which requires F0 and duration contrasts to systematically alter, and found that contrary expectation, syllable-timed Mandarin contributed little to patterns of duration contrasts. However, both L1 and TW L2 speech showed under-differentiation of F0 contrast patterns in different ways and for different reasons [4]. The majority L1 speakers showed only 2-way (binary) contrast between the stressed/unstressed syllables instead of the 3-way primary/secondary/tertiary contrasts; the secondary stress was the cause. At the surface, some of the secondary stress cannot be differentiated from the primary counterpart while in other cases it is realized as the tertiary stress. More fine-grained analyses revealed that fluctuations were correlated to sequential position anchored by the primary stress. The F0 of secondary stress preceding the primary stress is often elevated to similar level of the primary stress, as in "C*a*li*f*o*r*nia, *i*nform*a*tion" whereas the F0 of secondary stress following the primary stress is often lowered to the level of tertiary stress as in "elev*a*t*o*r, Janu*a*r*y*" [5]. The most robust contrast always exists between the primary stress, followed by lowering and compression. Assuming that in physical terms the primary stress must stand out in the speech signal, post-stress lowering is more significant to constitute the H/L contrast. Thus we argued that the inconsistency or secondary stress production was caused by assimilation due to forward planning and pre-planned post-primary compression. Robust F0 H/L contrast only occurs after the primary stress when mandatory differentiation is required while under-differentiation is accommodated when little difference of meaning would occur. As expected, the above stress-related differentiation patterns were highly varied in TW L2 data; under-differentiated lexical stress was indeed a major TW L2 feature.

The above findings suggest that information related focus and/or emphasis that involve planning the allocation and placement of where to emphasis, followed by correlating production of the necessary contrast patterns may contribute to TW L2 prosody. When an L2 speaker is not certain where to emphasize the focal point(s) in a phrase or sentence, whether and how the multiple phrases in a speech paragraph should be differentiated from one another to express information weighting, and finally how to produce the differentiation contrasts to correctly signal the information structure, then his/her output would definitely become less intelligible. [6, 7, 8]

In order to further understanding of the prosodic aspects of TW L2 English that makes it distinct, in the following study

978-1-4799-4219-0/14/$31.00

we will present comparisons of focus/emphasis patterns of TW L2 English to L1 American English together with focus/emphasis from three genres of L1 TW Mandarin speech by individual phrases from continuous and by discourse structure. Following the results of our previous findings, the only acoustic correlate used for comparison is the F0.

2. Speech Materials and Annotation

2.1. Speech data

The materials of English speech are data of 2 reading tasks of the AESOP-ILAS (Asian English Speech cOrpus Project—Institute of Linguistics Academia Sinica) corpus [1]: (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, and a total of 113 words (144 syllables). Speech data of 10 gender balanced L1 North American English speakers and 10 gender balanced Taiwan L2 speakers were analyzed. The materials of Mandarin speech used are (1) 1 male and 1 female reading of 26 discourse pieces coded CNA in the COSPRO database [10] (approximately 55 min/11600 syllables/85MB). (2) 1 male and 1 female reading of simulating broadcast of weather forecast coded WB (approximately 45 min/7070 syllables/50MB). (3) 1 male speech of spontaneous classroom lecture coded LEC (approximately 26 min/7660 syllables/49 MB).

2.2. Preprocessing and annotation

The speech data of L1 English, TW 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 and emphases were manually tagged independently.

2.2.1. Tagging discourse units by perceived boundaries and breaks

Discourse units were manually tagged by 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.

2.2.2. Tagging emphases by perceived degree of prominence

The same speech data are further manually tagged by trained transcribers into a string of emphasis/non-emphasis tokens (ETs) for 4 degrees of perceived strength of prominence defined as follows:

- E0-- reduced pitch, lowered volume, and/or contracted segments
- E1--normal pitch, normal volume and clearly produced segments
- E2--raised pitch, louder volume and irrespective of the speaker's tone of voice
- E3--higher raised pitch, louder volume and with the speaker's change of tone of voice

2.3. Methods

Emphasis allocation is examined in two larger discourse prosodic units the PPh and breath-group (BG). Patterns of emphasis are derived by the following procedures.

2.3.1. Deriving patterns of perceived emphasis in PPh

We assume that (1) placement of emphases in a PPh is pre-planned to reflect where key information is located and (2) placement patterns are limited. ET sequences (see 2.2.2) within each PPh are used to represent emphasis patterns. The same sequence patterns are merged into a unique type. The merged types of emphasis patterns are then calculated for respective frequency and ranked. Emphasis patterns whose frequency ranked lower than 2% are collapsed and classified as 'others'

2.3.2. Deriving emphasis allocation in BG by F0 features

We assume that similar rationale is also used the next larger prosodic layer BG to reflect discourse information weighting, and templates of emphasis allocation could also be derived from the speech signal. Since pervious findings [4] revealed how lexical stress as well as phrase accentuation is most saliently realized by F0, instead of using perceptual tagging of relative emphases by PPh, the sub-unit of BG, we further assume that emphasis allocation in BG could be found in F0 patterns. By adopting the high/low (H/L) concept in phonology [11] to larger units as well, we compared the overall F0 H/L difference of PPh as a representation of higher-level emphasis. To facilitate the validity of relative H/L difference by phrase, we used the command response model [12] to filter the F0 into overall contour Ap and locally accentuated sections Aa, thus by default Ap is a more accurate representation of the overall F0 contour of the chosen unit. We then compare the H/L difference of a succession of PPh in a BG as follows: Each extracted phrase Ap contour is labeled as 1 if its peak is higher than that of the previous Ap; and 0 if the peak is lower than previous Ap (the default value of first Ap without relative value is set as 1). Interestingly, we quickly note that in our data, successive directional H/L sequence is only limited to 4 phrases at most, alternating between H-to-L or L-to-H patterns. Such alternation echoes the up- vs. down-stepping concept in phonetics [13], and could readily be extended to describe similar effects by larger units. As a result, the pattern of up- or down-stepping and the number of steps could be further merged into limited patterns representing allocation of even higher-level emphasis. Figure1 is an illustration of Ap sequences and their respective H/L assignments, whereas the red and blue bars indicate the direction of up- and down-stepping, respectively.

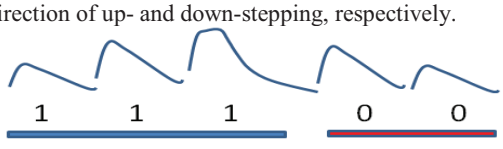

Figure1: An example of a sequence of Ap in a BG. 1 and 0 represent the relative H or L of each PPh in relation to its preceding PPh. The blue and red bar represents the state of up-stepping and down-stepping by PPh, respectively.

3. Results and Discussion

Emphasis patterns are examined to see (1) in what ways TW L2 English differs from L1 English (2) whether TW specific patterns could be attribute to Mandarin and (3) whether TW patterns are general Mandarin features or genre relate, and if so why.

3.1. Emphasis allocation in PPh

Overall patterns of emphasis allocation by PPh are collapsed and shown in Figure 2. Looking at the pie graphs from the left, the results reveal that across the 3 genres of TW Mandarin data, about 70-77% of PPhs could be accounted for by 6 most frequent emphasis patterns (E1, E2 E1, E1 E2 E1, E1 E2, E2 and E2 E1 E2). However, note that the distribution of the same patterns differs by speech genre. For passive prose reading, the top two most frequently pattern are E2 E1 (one phrase initial emphasis, 30%) and E1 (no emphasis 10.74%). For simulating weather reporting the two favored patterns are E1 E2 (one phrase final emphasis, 25%) and E1 E2 E1 (one phrase medial emphasis 13.02%). For university classroom lecture the two patterns are LEC E1 (no emphasis, 39%) and E2 E1 (one phrase initial emphasis, 11.84%)

However, for L2 English about 75 % of L2 English could be accounted for by only by 4 most frequent emphasis patterns (E1, E2 E1, E2 and E1 E2); whereas about 80 % of L1 English requires 7 emphasis patterns instead (E1, E2 E1, E1 E2 E1, E1 E2, E2, E2 E1 E2 and E1 E2 E1 E2). Note that the top two patterns for L2 English are E2 (emphasizing the entire phrase, 38.67%) and E1 E2 (one phrase final emphasis, 22.36%); and for L1 English E2 (emphasizing the entire phrase, 25.76%) and E1 E2 (one phrase final emphasis, 20.20%).

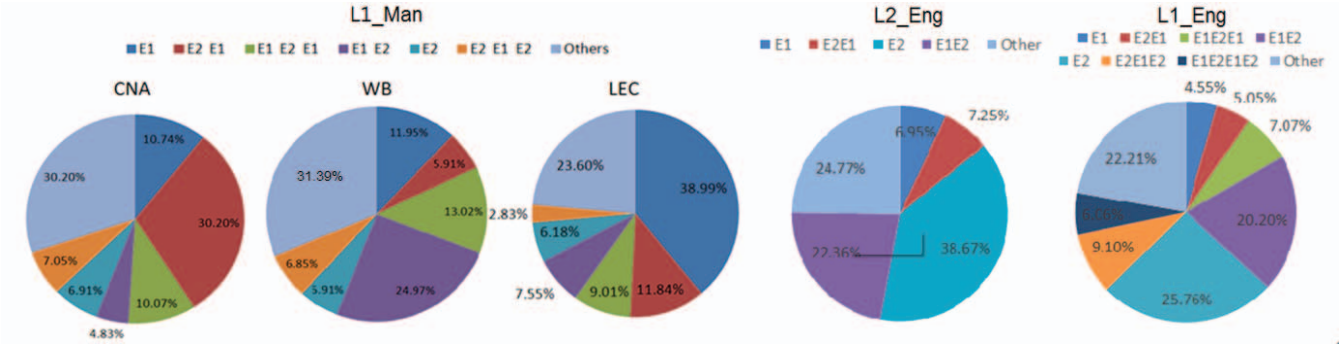

Figure2: The distribution of emphasis patterns (ET type and sequence) in PPh by language and genre.

3.1.1. Discussion

Using L1 Mandarin speech as reference, TW speakers are as capable as the L1 English speakers, each using more diversified emphasis patterns (7 for three L1 Mandarin genres and 8 for L1 English). Comparing the top two favored Mandarin emphasis patterns revealed that aside from 'others', the top choice of CNA (prose reading) is the phrase-initial (or default) pattern while WB (simulating weather report) is the reverse phrase-final pattern instead. We believe this is due largely to information content. The second choice showed 10% of the phrases of CNA are produced with no emphasis and hence little information content while 13% pf WB used a much more elaborate E1 E2 E1 pattern, suggesting how WB is loaded with more complicated information content than prose. Moreover, classroom lecture LEC is distinctly different: where phrases with no emphasis take up nearly 40%; default phrase-initial emphasis takes up another 10%. Note these two most favored patterns make up half of the lecture data. Assuming they represent less information content, it becomes clear how filler phrases are as important as significant information content in continuous speech; its role mainly to provide reference context that requires less cognitive resource and efforts to process. As a result, content information would properly surface through contrast. In summary, TW speakers employ genre specific focus patterns in L1 in relation to information structure.

The data of L1 and L2 English reading the same story reflect a different picture. L2 speakers resort to fewer (5 tin total) and less elaborate patterns, either by emphasizing the entire phrase E2 (39%) or by emphasizing the phrase end E2 E1(22%). These two patterns (61%) are not their top choices when producing L1 Mandarin, suggesting they become much flat sounding when producing L2 English. As for L1 English, note that although the top two choices are identical with L2, the distribution of both patterns is less (26% and 20%, respectively, totaling 46%). As mentioned before, the rest of L1 English data were distributed by 5 additional and more varied patterns, showing more elaborate manipulation of information and expression. In general, L2 speakers produce too many accentuated phrases E2 (39%) making and sound more emphatic than necessary. Emphasizing the phrase end E1 E2 (22%) is similar to simulating weather report WB in L1 Mandarin (25%) suggesting placement of content information in phrase final at phrase final position. The above phrase level characteristics suggest that L2 speakers may adopt different information disposition and contribute to L2 accent.

3.2. Emphasis allocation in BG

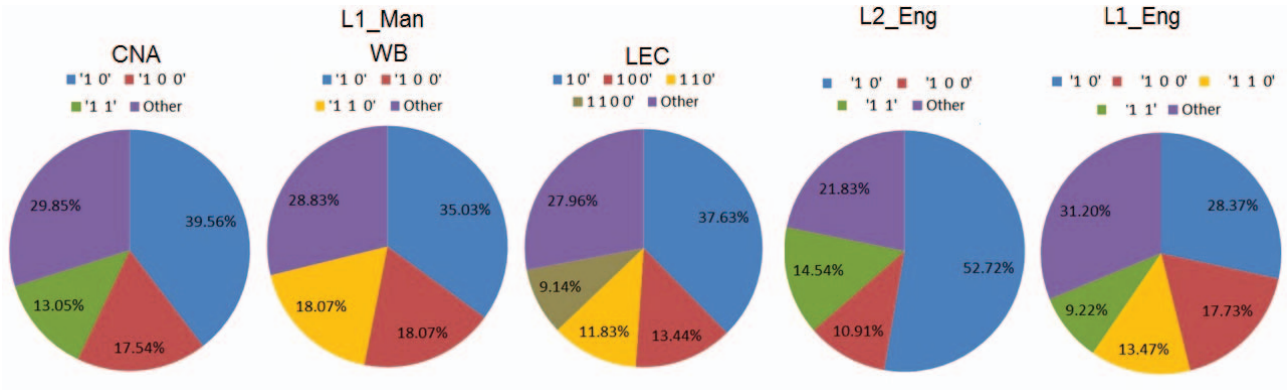

Figure 3: The distribution of emphasis patterns in BG by language and genre.

The allocation of relative PPh emphasis in BG is shown in Figure3. Results reveal about 70% of PPhs in CNA_L1_Man could be accounted for by 3 most frequent emphasis patterns (1 0, 1 0 0, 1 1 or 1 step down, 2 steps down, 1 step up). About 70 % of PPhs in WB_L1_Man could be accounted for by 3 most frequent emphasis patterns (1 0, 1 0 0 0, 1 1 0 or 1 step down, 3 steps down, 1 step up then down 1 step). About 73% of PPhs in LEC_L1_Man could be accounted for by 4 most frequent emphasis patterns (1 0, 1 0 0 0, 1 1 0, 1 1 0 0, or 1 step down, 3 steps down, 1 step up then down 1 step, 1 step up then down 2 step). About 80% of PPhs in L2_Eng could be accounted for by 3 most frequent emphasis patterns (1 0, 1 0 0 0, 1 1 or 1 step down, 3 steps down, 1 step up). About 80% of PPhs in L1_Eng could be accounted for by 3 most frequent emphasis patterns (1 0, 1 0 0, 1 1 0, 1 1 or 1 step down, 2 steps down, 1 step up then down 1 step, 1 step up).

3.2.1. Discussion

Comparing paragraphs formed by multiple phrases and their overall F0 height manipulation by steps, we found similar distribution between CNA_L1_Man and WB_L1_Man, whereas '1 step up' in CNA is replaced by a more complex pattern '1 step up followed by 1 step down' in WB, again suggesting how information content is reflected in discourse structure as well. As expected, spontaneous lecture further demonstrates more complex information allocation than WB, whereas the '1 step up followed by 1 step down' pattern in WB is replaced by more complicated '1 step up followed by 2 steps down' pattern in LEC, suggesting that university lecture is laden with much more intricate information content. In summary, of the three Mandarin speech genres examined, the relationship by information complexity can be expresses as spontaneous speech>simulating weather broadcast>prose reading.

The results also demonstrate that the most distinct feature between L1_Eng and L2_Eng is the much higher ratio of the declination pattern '1 step down' in L2_Eng. In other words, L2 English showed overuse of simple planning at the higher discourse level (paragraph production), i.e., planning by individual phrases than by the paragraph, a result that echoes our previous finding of using smaller discourse unit [2, 3]. Moreover, information disposition of L2_Eng is most similar to prose reading in Mandarin.

4. General Discussion & Conclusion

The above study of TW L2 English showed when speaking continuously, prosodic difficulties are multi-folded. In addition to mastering word-level lexical prosody, sentential prosody such as narrow focus due to structural specifications, and paragraph association due to discourse organization, there is also information structure to master through where to make the emphases and how to produce it properly. As evidenced in seemingly ambiguous information allocation shown in emphasis placements and patterns, TW L2 speakers had little choice but resort to what appears to be simpler planning strategies available; their choice different from their L1 norms both at the phrase level (3.1.1.) as well as paragraph level (3.2.1). Together with difficulties and uncertainties from lexical (word stress) and syntactic (narrow focus) levels that contribute to prosody realization, their interactions further contributes to why the overall prosody of TW L2 English departs from L1 English. We believe the above study of information related higher level prosodic features not only sheds light on better understanding of TW L2 English, but may also be applicable to L2 prosody in general. Moreover, all of these findings could be used in teaching and CALL applications of English as L2, especially with respect to how to help learners to acquire the perceptual sensitivity for differentiating contrast strength and the production robustness of correlating contrast patterns.

5. Reference

[1] 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.

[2] Visceglia, T., Tseng, C. Y., Su, Z. 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.

[3] 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

[4] 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.

[5] Tseng, C-Y. and Su, C-Y. “Prosodic Differences between Taiwanese L2 and North American L1 speakers—Under-differentiation of Lexical Stress”, Speech Prosody 2014, Dublin, Ireland, 2014.
[6] Kruijuff-Korbayova, I. & Steedman, M. “Discourse and Information Structure. Journal of Logic, Language and Information”, 12:249-259, 2003.
[7] Van Donel, M. “The relation between textual information structure and perceived prominence in discourse. Prosodic Aspects of Information Structure in Discourse, the Netherlands”, IFOTT, 9-40, 1999.
[8] Van Donel, M. “Prosodic characteristics of focal structure. Prosodic Aspects of Information Structure in Discourse”, the Netherlands: IFOTT, 115-144, 1999.
[9] Tseng, C-Y. and Su, C-Y. “Discourse Prosody and Context – Global F0 and Tempo Modulations”, Interspeech 2008 1200-1203. Brisbane, Australia, 2008.
[10] 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.
[11] http://www.ling.ohio-state.edu/~tobi/
[12] Hirose, K. Fujisaki, H. and Yamaguchi, M. “Synthesis by rule of voice fundamental frequency contours of spoken Japanese from linguistic information”, IEEE, 1984.
[13] Connell, B. “Downdrift, Downstep, and Declination”, In Proceedings of the TAPS (Typology of African Prosodic Systems Workshop), 2001.