

## L2 DISCOURSE AND INFORMATION PLANNING AND THEIR PROSODIC IMPLICATIONS

*Chiu-yu Tseng<sup>1</sup> & Chao-yu Su<sup>1,2</sup>*

<sup>1</sup>Phonetics Lab, Institute of Linguistics, Academia Sinica, Taipei, Taiwan

<sup>2</sup>Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan  
cytling@sinica.edu.tw

### ABSTRACT

In this paper we provided a concise literature review on the prosody of Taiwan (TW) L2 English that featured by under-differentiated contrast patterns exemplified in lexical stress from word stress and to syntax induced narrow focus. We therefore hypothesized that higher-level discourse planning through patterns of chunking and phrasing and information structure through placements of emphases by degree may further cause prosodic deviations. We analyzed patterns of chunking size and consistency by boundary breaks of a total of 18 L2 TW speakers to the same number of L1 speakers and found higher variation of L2 speech with reference to L1 speech and among L2 population as well. We analyzed emphasis allocation patterns and found L2 speakers used less varied patterns and less type of variation. Comparison of L1 read English and L1 spontaneous Mandarin showed that emphasis patterns are shared regardless of language and speech type; the difference lies only in their distribution. We believe these findings could aid CALL development of prosody and expression.

*Index Terms*— discourse structure, chunking and phrasing, higher-level planning, information structure, emphasis, emphasis allocation patterns, prosody, under-differentiation

### 1. INTRODUCTION

English is the lingua franca of our world, and Asia is home to the largest number of English learners and speakers. It has been claimed that combining native and non-native speakers, India now has more people who speak or understand English than any other country in the world. Following India is the People's Republic of China [1]. Thus, research in Asian English dialects from a multidisciplinary perspective with tailored prospects of application to develop computer aided language learning (CALL) tools has become more imperative than ever. Consequently, considerable research interest has emerged in studying topics of L2 accent, language identification and recognition [2, 3 for example]. A comprehensive understanding of the features and variation present in spoken L2 English, often referred

generically as foreign accent [4, 5, 6], is a fundamental issue for the development of English language education as well as spoken language science and technology in Asia. The present study focuses on features specific to Taiwan (TW) L2 accent, which we believe would contribute to understanding features and accent of Asian L2.

Compared with L2 segmental features, much less is known about the nature of L2 prosody features and their implications in communication and pedagogy. Our aim is to find and catalogue features that may impede intelligibility from corpus analysis, look for systematic patterns and hopefully derive technologically applicable prosody models. However, studies of suprasegmental features are particularly difficult because of the following challenges due to high variability in acoustic features and lack of consensus about communicative significance of suprasegmental features [7]

In our previous prosodic investigations of L2 TW English, by far the most interesting findings that we would catalogue as TW L2 accent are prosodic under-differentiation of syntax-elicited narrow focus and lexicon-defined word stress [10]. Acoustic analysis of by sentential prosody with narrow focus showed that TW L2's production of narrow focus is less robust in F0 and amplitude than their L1 counterparts, resulting in less degree of contrast and therefore making TW L2 English sounding less differentiable than intended [8, 9]. Further investigations of lexical prosody due to word stress are even more complex. While initial findings from analyzing word stress patterns showed that though TW L2 speakers were able to maintain similar degree of stress-induced duration contrast as L1 speakers, the degree of contrast in F0 and amplitude is again less robust, making word stress in TW L2 English less differentiable due to lack of pitch and loudness contrasts [10, 11]. However, more fine-tuned analysis of poly-syllabic words further revealed features that may be specific to TW L2 English. While L1 speakers' preferred realization of word stress turned out to be through robust binary stress/no-stress contrast anchored by the primary stress, in that pre-primary syllables are elevated to near-primary magnitude whereas post-primary syllables are suppressed in to near-tertiary stress in F0, duration and intensity, thereby creating robust contrast between the primary stress and its following

syllables [11]. The 3-way primary/secondary/tertiary contrast is merged into a binary stress/no-stress contrast with the secondary stress merged either to the primary or tertiary stress pending its sequential position. In other words, the fluctuation of the secondary stress is position conditioned and completely predictable, resulting in more robust binary (stressed/unstressed) contrast with a high/low (H/L) pattern. The results not only enabled us to explain the variation of secondary stress at face value, but also led us to examine the stress output of TW L2 English from a different perspective. As expected, the position-dependent merge of the secondary word stress is difficult for TW L2 speakers. Contrary to L1 speakers, TW L2's production of word stress is to reduce all non-primary stresses to the status of tertiary regardless of position, thereby producing a low/high/low (L/H/L) pattern accompanied once again with less degree of contrast.

Along the same vein of cataloguing prosodic under-differentiation as a TW L2 feature, we would like to further investigate two more issues that also involve differentiation in the prosodic domain: (1) overall higher level discourse planning through chunking and breaking patterns and (2) information-structure induced weighting arrangements through focusing and emphasizing, both of which in continuous speech.

In our previous prosodic investigations of acoustic analyses, we have found that overall TW L2 speakers exhibited more frequent chunking and breaking than L1 speakers, making their discourse/paragraph units smaller in size. In addition, their allocation of discourse chunking is also highly inconsistent [12]. The present study presented below is to further investigate discourse chunking and information arrangement, the former through chunking and breaking patterns, the latter through analysis of perceived degrees of prominences. We will analyze placements discourse boundaries and perceived degrees of emphases as a direct reference to allocation of information weighting. In addition to comparable analysis of L1 and L2 TW English, we will include data and analysis of TW Mandarin spontaneous speech to provide possible factors of L1 similarity in addition to L1/L2 difference. Due to space limit, we will only be able discuss the prosodic implications of the findings may without accompanying acoustic analyses.

## 2. SPEECH MATERIALS AND ANNOTATION

### 2.1 Speech data

Read speech of L1 English, L2 English and L1 Mandarin spontaneous speech are used. The materials of English speech are two passage reading tasks from the AESOP-ILAS and AESOP2-ILAS, two corpora of TW L2 English under the protocols of AESOP (Asian English Speech

cOrpus Project) collected by the ILAS (Institute of Linguistics Academia Sinica) group [13, 14]. Data from AESOP-ILAS is reading of "The North Wind and the Sun" (henceforth NW&S) passage at normal speech rate and volume. The passage contains a total of 113 words (144 syllables) in 3 paragraphs, 5 sentences with 8 independent clauses and 5 dependent clauses. Data from AESOP2-ILAS is reading of "The Cinderella Fairy Tale" (henceforth Cinder) passage which contains a total of 759 words (1,000 syllables) in 14 paragraphs; 82 sentences with 93 independent clauses and 49 dependent clauses. English speech data is balanced by gender and number of speakers. Data of "The North Wind and the Sun" passage includes speech from 10 L1 North American English speakers and 10 TW L2 speakers. Data of "The Cinderella Fairy Tale" passage includes speech from 8 L1 North American English speakers and 8 TW L2 speakers. Data of L1 Mandarin spontaneous speech is recording of university classroom lecture (henceforth LEC) by 1 male native speaker on digital signal processing. Approximately 26 minutes of speech totaling 7660 syllables in 49 MB.

### 2.2 Processing and annotation

The speech data of L1 English, TW L2 English and L1 Mandarin were tagged in layers for discourse as well as information structure. The preprocessing layer is force-aligned segments by the HTK Toolkit followed by manual spot-checking by trained transcribers. Discourse units and information status 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 information structure by perceived degree of prominence

The same speech data are also manually tagged by trained transcribers into a string of perceived emphasis/non-emphasis tokens (ETs) by degrees of strength as a reference of information weighting. We defined 4 degrees of perceived prominence 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

### 3. METHODS OF ANALYSIS

We assume that change of breath (boundary B4) and change of paragraph (boundary B5) reflects different levels and units of higher level discourse/paragraph planning. Hence among-speaker consistency or agreement of tagged discourse boundaries is derived as a reference of discourse planning. Methods used to derive boundaries B4/B5 consistency/agreement are described in Sec. 3.1; methods used to analyzed and derive patterns of emphasis allocation by discourse units PPh is described in Sec. 3.2. The derived results also enabled us to make L1/L2 difference both by English and by English/Mandarin comparison.

#### 3.1 Among-speaker B4/B5 consistency/agreement

Two indicators are used to quantify among-speaker consistency of discourse boundaries B4/B5: (1) average consistency and (2) distribution by speaker overlap. Average consistency (AC) is defined as follows.

$$AC = \frac{1}{BN} \sum_{B_i=1}^{BN} \frac{1}{SN} \sum_{S_j=1}^{SN} C_{S_j B_i}$$

Where  $BN$  and  $SN$  represent number of B4/B5 and speaker  $S_j$  and  $B_i$  denote index of B4/B5 and speaker  $C_{S_j B_i}=1$  when  $j$ th speaker with  $i$ th B4/B5  
 $C_{S_j B_i}=0$  when  $j$ th speaker without  $i$ th B4/B5

B4/B5 distribution is also plotted by speaker agreement to show more details of among-speaker consistency/agreement

#### 3.2 Patterns of perceived emphasis by PPh

We assume that in each PPh (1) placement of emphases is pre-planned to reflect allocation of key information and (2) patterns emphasis (E) alternation by linear sequence are limited. Hence ET sequences (see 2.2.2) by PPh are used to represent sequential patterns of emphasis allocation. The same sequence patterns are merged into a unique type. The merged types of emphasis patterns are then calculated for respective frequency and ranked by distribution. Emphasis patterns whose frequency rank is lower than 2% are collapsed into a single category under the term 'others'.

## 4. RESULTS AND DISCUSSION

### 4.1. Discourse organization

In order to investigate L1/L2 difference and distinct L1/L2 features, distribution of discourse boundaries and size (length) of discourse units are compared between L1 and L2 speakers. In addition, L1/L2 consistency by paragraph boundaries (B4/B5) is also compared to see how L2 speakers' discourse planning may differ from L1.

#### 4.1.1 Discourse planning by distribution of respective boundaries

Figure 1 and Table1 show the distribution of discourse boundaries by L1 and L2 English speech. Results show that more percentage of B3 breaks and less percentage of B5 breaks are found in L2 speech across the two datasets NW&S and Cinder. The B3 percentage of L2 is 1.39 and 1.07 times to L1, respectively; whereas the B5 percentage of L2 is 0.58 and 0.6 times to L1. However, B4 percentage of L2 is 0.39 times to L1 in NW&S but not differentiable in Cinder.

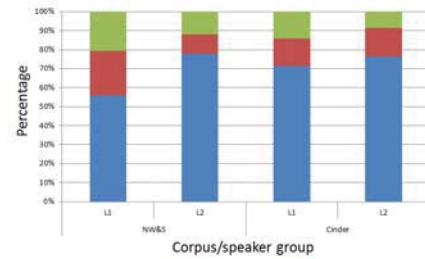


Figure1. Distribution of discourse boundaries by corpus/speaker group

Corpus/speaker group		Break		
		B3	B4	B5
NW&S	L1	55.88%	23.53%	20.59%
	L2	78.00%	10.00%	12.00%
Cinder	L1	71.26%	14.78%	13.96%
	L2	76.26%	15.33%	8.40%

Table1. Distribution of discourse boundaries by corpus/speaker group

#### 4.1.1.1 Discussion

The above results suggest that L1 and TW L2 speakers use different strategies to plan discourse units. The L2 results are similar to an earlier study of data of NW&S by TW speakers [12], thus showing that TW speakers are not sensitive to the Cinder passage which is not only longer but also with much more variation of utterance type and duration. In general though L1 speakers are sensitive to chunking by smaller lower-level units; they are able to maintain higher-level discourse coherence as shown in their consistencies within speaker group. In addition, note that L1 speech varies by narratives type. However, TW L2 speakers exhibit opposite patterns regarding higher-level units.

#### 4.1.2 Scale of discourse planning by size of discourse units

Figure2 and Table2 show the size (by number of words) of discourse units PPh/BG/PG by speaker group and narrative type. Results show that compared with L1 speech, L2 speakers use shorter PPh but longer BG and PG overall. The L2/L1 ratio values by size of PPh/BG/PG for NW&S are 0.66, 1.18 and 1.02, respectively; and for Cinder are 0.92, 1.11 and 1.52, respectively.

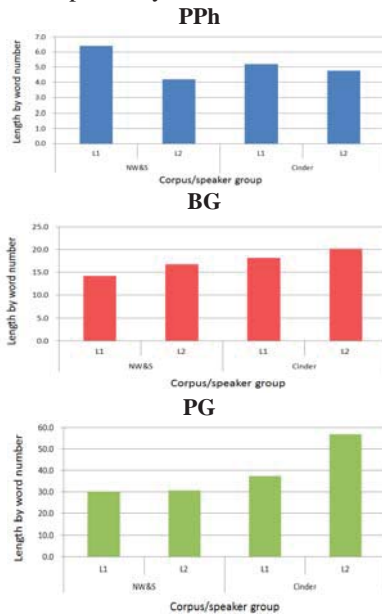


Figure2. Length of discourse units by prosodic layer and corpus/speaker group

Corpus/Speakers	Prosodic unit			
	PPh	BG	PG	
NW&S	L1	6.4	14.1	30
	L2	4.2	16.7	30.6
Cinder	L1	5.2	18.1	37.3
	L2	4.8	20.1	56.8

Table2. Length of discourse units by prosodic layer and corpus/speaker group

#### 4.1.2.1 Discussion

The above results appear to suggest that L2 speakers employ shorter planning scale for lower-level units and longer planning scale for higher-level units across narrative types. The shorter planning scale at lower discourse level in L2 speech confirms our previous study [12]; however, the longer higher-level planning scale found is against our hypothesis for L2 speech. Therefore, we further investigated among-speaker consistency/agreement with higher level units by boundary breaks B4/B5 of L2 speech; the results are presented in 4.1.3 below.

#### 4.1.3 Among-speaker consistency/agreement by B4/B5

Among-speaker consistency/agreement by B4/B5 is derived by L1 and L2 speech respectively and compared to see if L2

speakers could organize discourse as consistently and systematically as L1. Table3 shows average among-speaker consistency by B4/B5 by L1 and L2. Higher average consistency is found in L1 than L2 by 10% (61.82 vs. 51.82%) and 5.2% (35.71 vs. 30.41%) for NW&S and Cinder, respectively.

Corpus	Speaker group	
	L1	L2
NW&S	61.82%	51.82%
Cinder	35.71%	30.41%

Table.3 Among-speaker consistency/agreement by B4/B5 by corpus/speaker group

We further compared speaker overlap of boundaries B4/B5. Figure3 shows B4/B5 distribution by speaker overlap and datasets NW&S and Cinder, respectively. By NW&S, the most distinct difference between L1 and L2 appears in 100%, 60%, 30% and 10% overlap rate among speakers. Higher B4/B5 percentage by L1 than by L2 is found in 100% and 10% overlap and vice versa in 60% and 30% overlap. By Cinder, the most distinct difference between L1 and L2 appears in 62.5%, 37.5%, 25% and 12.5% respectively. Higher B4/B5 percentage by L1 than by L2 is found in 62.5% and 37.5% overlap rate and vice versa in 25% and 12.5% overlap rate.

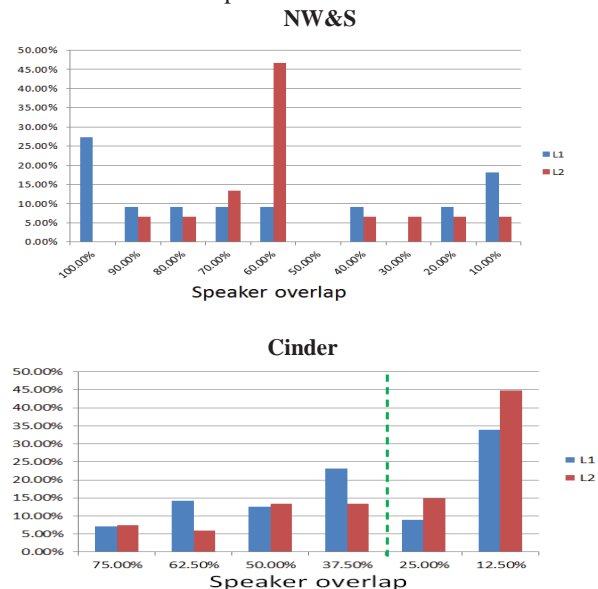


Figure3. B4/B5 distribution by speaker overlap and corpus/speaker group

#### 4.1.3.1 Discussion

The results show higher average among-speaker consistency of B4/B5 by L1 than L2 for both datasets NW&S and Cinder. However, the more varied structure and overall longer size of the Cinder passage with more varied sentence

types may be the reason for both L1 and L2 speakers vary in planning, and the within-group consistency is lower for both groups than patterns found for NW&S. The B4/B5 distribution by speaker overlap in Cinder shows about 58% of B4/B5 in L1 speech (adding all blue bars left of green dotted line in Figure3, i.e., 7.14%+14.29%+12.50%+23.21%.) could achieve 37.5% agreement; whereas only 40 % of B4/B5 in L2 speech (adding all red bars left of green dotted line in Figure3, i.e. 7.46%+5.97%+13.43%+14.93%.) could do the same. The results at face value may seem to suggest that L2 speakers were able to plan longer paragraphs than L1. However, note that they some discourse boundaries were skipped by L2 speakers, implying their planning of paragraph units is more inconsistent than L1 speakers. In other words, taking L1 speakers as the norm, the L1/L2 difference by B4/B5 should be interpreted as L2 speakers' inconsistency to realize higher-level discourse constrained chunking, continuation or termination among paragraphs as systematically and clearly as L1speakers. As a result, the inconsistency may also be considered as a higher level feature that may impede to overall intelligibility of L2 speech.

#### 4.2. Information allocation

In this section we add L1 Mandarin data in order to test L1/L2 difference and features that may be distinct L1 in general. The distribution patterns of perceived emphases in two English read-speech datasets NW&S, Cinder and one L1 Mandarin spontaneous speech dataset LEC were

calculated as a reference information planning and allocation. The results are presented in Figure 4. In addition, the Mandarin LEC data also made possible comparison between read and spontaneous speech.

The L1 English data shows that between datasets NW&S and Cinder, 7 most frequent patterns are shared, i.e., 'E1' 'E2E1' 'E1E2E1' 'E1' 'E1E2' 'E2' 'E2E1E2' 'E1E2E1E2', but their distribution differs by dataset. The L2 English data, in contrast, shows that only 4 most frequent patterns are shared, i.e., 'E1' 'E2E1' 'E1E2'. The results suggest that TW L2 speakers use less emphasis patterns than their L1 counterparts; their speech may sound less expressive overall. Comparison of L1 speech between English and Mandarin revealed very interesting results: of the 7 most common emphasis patterns used in read L1 English, the spontaneous L1 Mandarin data shares 6 patterns, thereby showing how native competence differs little regardless of speech format. We believe the L1 similarity can also be interpreted as a reference of the high demand of cognitive constraints needed to plan and produce L2 speech, even for proficient speakers. Another interesting feature of L1 spontaneous Mandarin is the higher percentage (40%) of 'E1' (no emphasis) patterns used in comparison with read speech. Since phrases with no emphasis carry less information load, it may suggest that more filler phrases are used in spontaneous speech, but requires further investigation in the future.

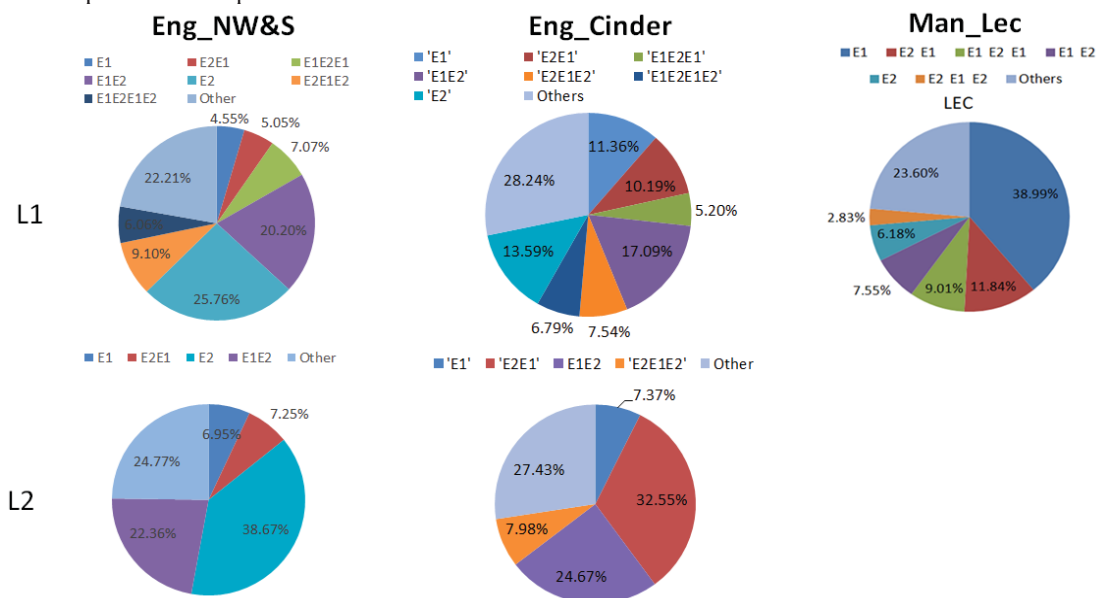


Figure4. Emphasis distribution by corpus/speaker group

##### 4.2.1 Discussion

Results of emphasis patterns in L1 speech show that both English and Mandarin speakers are able to use versatile



emphasis patterns to arrange and express information structure; the patterns used are in fact common. However, we also note that distribution of patterns varies by language which in part may be due to language-specific phonological and syntactic features. In contrast to L1 English speakers, TW L2 speakers use fewer and less complex emphasis patterns to arrange information weighting. In addition, over use of two patterns are found in TW L2 English, namely, 'E2' (38.67%) in NW&S where an entire phrase is emphasized; and 'E2E1' (32.55%) in Cinder where the phrase-initial is emphasized. It demonstrates that L2 speakers tend to over-apply simple patterns than L1 speakers. We therefore suspect that two features may be common across L2 speech in general, one is less complex information structure; another is over use of simple patterns.

## 5. DISCUSSION AND CONCLUSION

We have discussed briefly our previous studies on syntax-induced narrow focus and lexical stress of word stress that one of the major prosodic features of TW L2 English is under-differentiation (Sec. 1). In the present extended study to reading of longer passages that involves discourse and information planning. The above L1/TW L2 findings from English on discourse planning suggest that when reading out loud familiarized text pieces, TW L2 speakers tend to be more conservative in discourse planning, using more lower-level discourse units, skipping discourse units, and less higher-level units. There is less intra-L2 consistency of higher-level units than intra-L1; and less the inter-L1/L2 consistency as well. Overall, the L1 consistent higher-level units are larger in size than L2; thereby further substantiating L2 difficulty of higher-level discourse and paragraph planning [12]. We believe this is a feature is more general in nature rather than specific to TW speakers only. Further analysis of allocation patterns by perceived emphasis as a reference of information structure also show that overall L2 speakers' arrangements of information structure is less complex and less varied than L1. Cross-linguistic L1 English/Mandarin comparison show how in fact L1 complexity is similar, making the L1/L2 difference more pronounced. Information allocation is also found less systematic and less expressive in TW L2 speech. We believe this is another feature that is not TW specific, but L2 in general.

Since discourse chunking and phrasing as well as emphasis allocation are expressed through prosodic means, namely, alternation patterns of pitch H/L variation, rhythmic patterns and loudness variation, the present findings on discourse structuring would result in less distinct paragraph association; on less complex emphasis allocation patterns would result in less explicit expression of information weighting. We therefore predict that the combined

contribution may further cause the prosody of TW L2 English to be further under-differentiated and further substantiate under-differentiation as a TW English accent. We believe the prosodic manifestations may differ more from language to language. In our future work, we will continue our acoustic analysis in the same direction. Finally, we believe our study would aid CALL development on paragraph chunking and emphasis placements in general, with tailored tools on prosodic expression in particular.

## 7. REFERENCES

- [1] Zhao, Y. and K.P. Campbell. 1995. "English in China". *World Englishes* 14 (3): 377-390.
- [2] Nariai, T. & Tanaka, K. 2008. "A study of pitch patterns of Japanese English analyzed via comparative linguistic features of Japanese and English" (pp. 776-779) *Proceedings of Interspeech 2008* September 22-26, 2008, Brisbane, Australia
- [3] Crystal, D. "Subcontinent Raises Its Voice" *Guardian Weekly*: Friday 19 November 2004.
- [4] Magen, H.S., "The perception of foreign-accented speech", *Journal of Phonetics*, vol. 26, 381-400, 1998.
- [5] Anderson-Hsieh, 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.
- [6] Mixdorff, H. and Ingram, J. "Prosodic analysis of foreign-accented English", *Proc. Interspeech 2009*, 6-10 Sep. Brighton UK, 2009.
- [7] Jilka, M., "Different Manifestations and Perceptions of Foreign Accent in Intonation," in *Non-native Prosody: Phonetic Description and Teaching Practice*, Trouvain, J. and Gut, U. (eds), pp. 77-96, Mouton de Gruyter, 2007.
- [8] 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.
- [9] 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
- [10] 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.
- [11] 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.
- [12] Visceglia, T., Tseng, Chiu-yu, Su, Zhao-yu, and Huang, Chi-Feng. 2010. *Discourse Prosody Planning in L1 and L2 English*. Oriental-COCOSDA. Kathmandu, Nepal.
- [13] Tseng, C-Y. 2011. *Phonotactic and Discourse Aspects of Content Design in AESOP (Asian English Speech corpus Project)*. Oriental COCOSDA 2011. Hsinchu, Taiwan.
- [14] Visceglia, T., Tseng, C-Y. Kondo, M., Meng, H. and Sagisaka, Y. 2009. "Phonetic aspects of content design in AESOP (Asian English Speech corpus Project)", *Oriental COCOSDA 2009*. Beijing, China.