Computational Linguistics and Chinese Language Processing
Vol. 19, No. 4, December 2014, pp. 61-76

Some Prosodic Characteristics of Taiwan English Accent

Chao-yu Su[∗+], Chiu-yu Tseng[+] and Jyh-Shing Roger Jang[#]

Abstract

The present study examines prosodic characteristics of Taiwan (TW) English in relation to native (L1) English and TW speakers' mother tongue, Mandarin. The aim is to investigate 1) how TW second-language (L2) English is different from L1 English by integrated prosodic features 2) if any transfer effect from L2s' mother tongue contributes to L2 accent and 3) What is the similarity/difference between L1 and L2 by prosodic patterns of word/sentence. Results show the prosody of TW L2 English is distinct from L1 English; however, TW L2 English and TW Mandarin share common prosodic characteristics which differentiate from L1 English. Analysis by individual prosodic feature shows distinct L2 features of TW English which might attribute to prosodic transfer of Mandarin. One feature is less tempo contrast in sentence that contributes to different rhythm; another is narrower loudness range of word stress that contributes to less strong/weak distinction. By examining prosodic patterns of word/sentence, similarity analysis suggests L1 and L2 speakers produce prosodic patterns with great within-group consistency respectively but their within-group patterns are distinct to counterpart group. One pattern is loudness of sentence and another one is timing/pitch patterns of word. The above prosodic transfer effect and distinct TW L2 patterns of prosody are found in relation to syntax-induced narrow focus and lexicon-defined word stress which echo our previous studies of TW L2 English and could be implemented to CALL development.

Keywords: Prosody, L1, L2, Mandarin, English, Contrast, Lexical Prosody, Narrow Focus.

[∗] Institute of Information System & Application, National Tsing Hua University, Taiwan
E-mail: morison@gate.sinica.edu.tw

[+] Institute of Linguistics, Academia Sinica, Taipei, Taiwan
E-mail: cytling@sinica.edu.tw

[#] Department of Computer Science & Information Engineering, National Taiwan University, Taiwan
E-mail: jang@mirlab.org

1. Introduction

Computer assistant language learning (CALL) offers many advantages which differ from a traditional classroom setting where one teacher is responsible for a group of students. CALL allows learners to decide and adjust the level and pace of learning individually by. Another advantage that the classroom setting could not provide is unlimited access of on-line high-quality comparison between speech produced by a learner and a native speaker. By far the most popular CALL systems are computer-assisted pronunciation teaching (CAPT) system based on automatic speech recognition (ASR) outcome. The goals of CAPT are automatic diagnosis of pronunciation including specific or global error (Witt & Young, 2000; Coniam, 1999; Moustroufas & Digalakis, 2007), but the focus has been on segmental errors. However, in recent years studies focusing on suprasegmentals have shown that in addition to segmental information, prosodic information is in fact indispensable. Specifically, when detailed information of the consonant and vowel segments in the speech signal is removed, results show how listeners pay attention to prosodic features such as the pitch variation, rhythm alternation, loudness change as well as intonation. The resulting speech without any segmental and lexical content suggests that listeners are also sensitive to prosodic information (Scruton, 1996; Trofimovich & Baker, 2006; Munro, 1995). This has led to more research attention to investigate prosody in relation to comprehensibility and accent of native vs. non-native speech; and a more balanced understanding regarding the contribution from both the segmental and suprasegmental aspects of language (Derwing & Munro, 1997; Anderson-Hsieh *et al.*, 1992; Munro & Derwing, 1999, Celce-Murcia *et al.*, 1996; Derwing *et al.*, 1998). Reported studies that applied prosodic training for second-language (L2) learners have demonstrated that computer-assisted prosody training systems did improve the overall comprehensibility of L2 speech (Hardison, 2004; Hirata, 2004). These studies showed prosody training with a real-time pitch display could improve both prosody and segmental accuracy, as judged by native speaker raters, while similar effect is found for English-speaking learners of Japanese. Another study demonstrated that aligning Mandarin English duration patterns with native English using resynthesis technology and dynamic time warping also brought significant increase in intelligibility (Tajima *et al.*, 1997). Complementary findings are studies that showed how incorrect timing and stress patterns are often cited as major contributors to intelligibility deficit (Benrabah, 1997; Anderson-Hsieh *et al.*, 1992). However, it appears that considerable gap does exist between research findings and software development. CALL systems are usually criticized as not necessarily “linguistically and pedagogically sound” (Derwing & Munro, 2005; Neri *et al.*, 2002). For example, a study specifically states that most CALL programs were developed with little understanding of phonology and how to apply phonological knowledge to teaching (Pennington, 1999). In short, there is less understanding of L2 prosody, and even less CALL systems that have applied features of L2 prosody into the system.

The present study is developed from the above discussed background and aims to analyze prosodic characteristics of TW L2 English accent supported by linguistic knowledge. The speech data used in the present study is AESOP-ILAS (Asian English Speech cOrpus Project collected by the Institute of Linguistics, Academia Sinica) representing accent of Taiwan L2 English, which is part of AESOP that was designed and constructed to represent to include various kinds of L2 English spoken in Asia (Visceglia *et al*., 2009) with built-in linguistic knowledge (Anderson-Hsieh *et al*., 1992). Built-in linguistic knowledge in the corpus design is to elicit characteristics which are predicted to be present in L2 English speech. Our previous studies have catalogued a series of TW L2 features that may impede intelligibility. The series of studies to TW L2 accent started from prosodic under-differentiation which is not only found in syntax-elicited narrow focus but also in lexicon-defined word stress. Acoustic analysis of syntax-elicited narrow focus also showed that TW L2's production of narrow focus is less robust in F0 and amplitude than L1 (Visceglia *et al.*, 2011; Visceglia *et al*., 2012). Further investigations of lexical-stress prosody showed the degree of contrast in F0 and amplitude is again less robust, making word stress in TW L2 English less differentiable (Tseng *et al*., 2012). The above two studies showed that lack of pitch and loudness contrasts is one of major feature of TW L2 accent in both word and sentence prosody. Further analysis revealed more complex L1s' features in words that may be difficult for TW L2 speakers (Tseng & Su, 2014). Native (L1) speakers may choose to realize word stress through binary stress/no-stress contrast anchored by the position of primary stress. Post-primary syllables are reduced to near-tertiary stress while pre-primary syllables are elevated to near-primary magnitude in F0. The 3-way primary/secondary/tertiary contrast is merged into a binary stress/no-stress contrast with robust prosodic contrast between the primary stress and its following syllable(s). As expected, the position-related merge of the secondary word stress is difficult for TW L2 speakers.

In addition to the above prosodic difference found between L1and TW L2 English, we also compared TW L2 accent and TW Mandarin, the target L2 speakers' mother tongue, and found in what ways TW L2 accent could be attributed to their L1 Mandarin features (Nguyen *et al*., 2008). Following this line of research, TW Mandarin is also included in the present study to further examine if and how some TW L2 English accent can further be attributed to Mandarin.

The present study aims to incorporate prosodic features found to contribute to TW L2 accent, and try to conduct prosody classification among L1 English, L2 English and Ll Mandarin by machine learning technology. The aim is to test if L1 English, L2 English and Ll Mandarin could be discriminated from each other by integrated prosodic features elicited by syntax-induced narrow focus and lexicon-defined word stress. Further discrimination analysis compares distinct prosodic characteristics of TW L2_Eng and TW L2_Eng-L1_Man shared characteristics of prosody to verify if prosodic features of TW L2_Eng are in relation to Mandarin. In addition, speaker-pair similarity by prosodic patterns is computed to test (1) difference between L1 English and TW L2 English groups and (2) cohesion within L1 English/TW L2 English group.

2. Speech Data

Read speech of Native English (L1_Eng), Taiwan L2 English (L2_Eng), Taiwan Mandarin (L1_Man) are used in present analysis. The materials of English speech are 5 reading tasks from the AESOP-ILAS recoded by 9 L1 (4M&5F) and 9 L2 (5M&4F) speakers. These 5 tasks are designed to elicit production of English segmental and suprasegmental characteristics including: (1) word-level features such as segmental by target words in carrier sentence; (2) phrase boundary phenomena such as declarative falls and interrogative rises by target words at phrase boundaries (3) form, timing and location of pitch accents, which are used to create phrasal and sentential prominence (broad and narrow focus) by target words in narrow focus position. 20 target words with 2-, 3- and 4-syllable of all possible stress patterns (Appendix A) are embedded in Task1 to Task 3. (4) function words in stressed and unstressed positions and (5) prosodic disambiguation of syntactic structures.

In section 3.1 and 3.2, the sentences in task 1 to task 5 are used for prosody classification among L1_Eng, L2_Eng and Ll_Man. In section 3.3, lexicon-defined prosodic similarity among speakers is computed by 20 stress-balanced target words in carrier sentence, Task1, to eliminate effect from higher level. An example of target word marked in boldface in carrier sentence is as follow.

- I said **SUPERMARKET** five times.

The sentences with broad and narrow focus in task 3 are used to test syntax-elicited prosodic similarity among speakers. An example of sentence in which broad and narrow focus are embedded is as follow. Narrow focus and broad focus are marked in boldface and italic respectively.

Context: Do you buy fruit at the farmer's market?

- No. I *usually* buy *fruit* at the **SUPERMARKET** because they stay *open later*.

After selecting sentences with acceptable F0 extraction, 369 L1_Eng and 434 L2_Eng sentences are used in present analysis.

The material of L1_Man is intonation balanced speech corpus (3441MB, 31:10) in SINICA COSPRO (Tseng *et al.*, 2003) which aims to examine role of intonation with respect to prosodic grouping in Mandarin speech. 3 types of sentences including declarative, interrogative and exclamatory with balanced POS combination are designed and collected in this corpus. In order to compare with English materials (task1 and task3 in AESOP-ILAS) in which all sentences are declarative, only declarative sentences are included in present analysis. Speech of one male and one female with good recording quality are chosen for analysis. After further selecting sentences with acceptable F0 tracking, 288 L1_Man declarative sentences are used in present analysis. Prosodic words in Mandarin are adopted as units of word-layer segmentation and corresponding feature extraction.

2.1 Annotation

All data were pre-processed automatically for segmental alignment using the HTK Toolkit, which was then manually spot-checked by trained transcribers for accuracy. F0 values were extracted and measured using a semitone scale.

3. Feature Extraction & Classification

3.1 Feature Extraction

Prosodic features used in present study are F0, duration, intensity. Each feature is z-normalized by sentence first then each sentence is encoded as a feature vector representing prosodic characteristics with hierarchical structure by sentence and word layer. The higher-level features, namely sentence-level features are derived by average of features in subsidiary units, namely word while word-level features are computed by subsidiary phoneme. In addition to conventional 6 types of general feature representation including mean, standard deviation, maximum, minimum, range and pairwise contrast referring to PVI (Grabe & Low, 2002) by each feature and each layer, histogram representation is also adopted to show more detailed properties of feature distribution. The adoption of histogram representation also could overcome inconsistent dimension among sentences which derived from varied number of words and phonemes thus requirement of consistent dimension could be fulfilled for classifier input. Two prosodic features encoded by histogram representation are mean and pairwise contrast by subsidiary units in sentence and word layer. Present histogram representation encodes prosodic features with 7 bins in which distribution of units is normalized to 100%. Normalized duration and F0 values were further refined to remove intrinsic physical properties based on previous knowledge. The intrinsic physical property for duration denotes segmental duration of each phoneme and intrinsic physical property for F0 denotes intonation of each sentence. 200 prosodic features in total are used in the present study.

3.2 Classification

Two popular classifiers for prosody classification among L1_Eng, L2_Eng and Ll_Man used are introduced as follows.

3.2.1 KNNC

The principle of k-nearest-neighbor classifier coded as KNNC (Cortes & Vapnik, 1995) is based on concept that data instances of the same class should be nearer in the feature space. As a result, for a given unknown data point x, the class is determined by K nearest points of x. The principles compute the distance between x and all the data points in the training space to decide K which is used for assign/predict class of unknown data point x.

3.2.2 SVM

Given a set of data with each example in data marked by binary categories, a support vector machine (SVM) (Coomans & Massart, 1982) training algorithm builds a model that assigns examples into one category or the other as accurate as possible while examples of the separate categories are divided by a clear gap that is as wide as possible. Unknown data points are then predicted to belong to a category based on which side of the gap they fall on.

3.3 Discrimination Analysis by Prosodic Features

Discrimination analysis is conducted between pair of speaker group by 200 prosodic features described in section 3.1. P value (Lehmann, 1997) is adopted as discriminative indicator between pair of speaker group. In a statistical test, sample results are compared to likely population conditions by way of two competing hypotheses: the "null hypothesis" is a neutral statement about "no difference" between two groups; the other, the "alternative hypothesis" is the statement that the person performing the test would like to conclude if the data will allow it. The *p*-value is the probability of obtaining the observed sample results when the null hypothesis is actually true. It could be quantified by the conditional probability $\Pr(X|H)$ (X is a random variable representing the observed data and H is the statistical hypothesis under consideration) which gives the likelihood of the observation if the hypothesis is assumed to be correct. If this *p*-value is very small, it suggests that the observed data is different from the assumption that the null hypothesis is true, and thus that hypothesis must be rejected and the other hypothesis accepted as true.

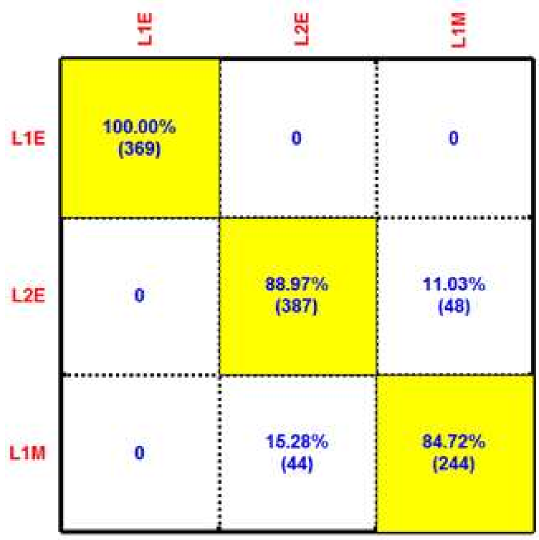
3.4 Similarity Comparison by Prosodic Patterns

The similarity is defined by cosine measure between any two of L1/L2 speakers by prosodic patterns of word/sentence. The value of point (i, j) in the matrix denotes cosine distance between speaker i and speaker j. In following section, the matrix is represented by a plot with i×j grids in which shading value of each grid denotes value of point (i, j). The darker the color is, the more similar between speakers i and j.

4. Results

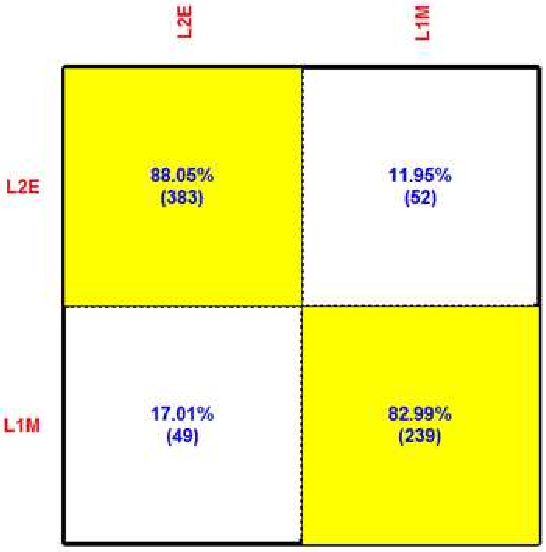
4.1 Prosody Classification among L1_Eng, L2_Eng and Ll_Man

In order to test if L1 English, TW L2 English and TW L1 Mandarin could be identified from each other by prosody, classification is conducted and performance is computed by 2 classifiers, SVM/KNNC. Average recognition rate is 91.57% by SVM and 81.86% by KNNC respectively. Figure 1 shows recognition rate in form of confusion matrix by best classifier, SVM and results suggest L1_Eng with most distinct characteristic with the others, L2_Eng and L1_Man. L1_Eng could be 100% identified from L2_Eng and L1_Man; however, only 88.97% of L2_Eng and 84.74% of L1_Man could be recognised from the others. Further binary classification is conducted between L2_Eng and L1_Man and shows best recognition rate 86.03% by SVM. Figure 2 shows confusion matrix which demonstrates only 88.05% of L2_Eng and 82.99% of L1_Man could be identified from each other.



	L1E	L2E	L1M
L1E	100.00% (369)	0	0
L2E	0	88.97% (387)	11.03% (48)
L1M	0	15.28% (44)	84.72% (244)

Figure 1. The recognition rate among L1_Eng, L2_Eng and Ll_Man by prosodic features and SVM



	L2E	L1M
L2E	88.05% (383)	11.95% (52)
L1M	17.01% (49)	82.99% (239)

Figure 2. The recognition rate between L2_Eng and Ll_Man by prosodic features and SVM

4.1.1 Discussion

The above results suggest that L1_Eng could be differentiated from L2_Eng and L1_Man; however, confusion is found between L2_Eng and L1_Man. In other words, L1_Eng is distinct from L2_Eng and L1_Man prosodically; on the other hand, L2_Eng and L1_Man share some common prosodic characteristics which differentiate from L1_Eng. In the following section, discrimination analysis is conducted by prosodic features to show distinct prosodic characteristics of L2_Eng from L1_Eng and common prosodic characteristics between L2_Eng and L1_Man.

4.2 Discrimination Analysis by Prosodic Features

Table 1 shows most distinct prosodic characteristics between L2_Eng and L1_Eng. After pairwise discrimination analysis between L2_Eng and L1_Man is conducted by each prosodic feature, the most discriminative features are computed and listed in Table1. Results show most discriminative prosodic features by lowest 5 p-values in L2_Eng vs. L2_Eng are 'mean by normalized F0', 'minimum by normalized F0', 'mean by normalized volume', 'maximum by normalized volume' and 'stand deviation by normalized duration' in sentence layer and maximum/PC/stand deviation/range/histogram_dimension#3 by normalized volume in word layer.

Table 1. The most distinct prosodic characteristics between L2_Eng and L1_Eng by p-value

Speech Pair Layer	L2_Eng vs. L1_Eng
Sentence Layer	'NorF0_Mean'
	'NorF0_Min'
	'NorVol_Mean'
	'NorVol_Max'
	'NorDur_STD'
Word Layer	'NorVol_Min'
	'NorVol_PC'
	'NorVol_STD'
	'NorVol_Range'
	NorVol_hisBySubMean_D3'

Table 2. The most similar prosodic characteristics between L2_Eng and L1_Man by p-value

Speech Pair Layer	L2_Eng vs. L1_Man
Sentence Layer	'NorVol_DisBySubPC_D5'
	'NorDur_DisBySubPC_D1'
	'NorDurWOIntri_DisBySubMean_D5'
	'NorDur_DisBySubPC_D3'
	'NorF0_PC'
Word Layer	'NorF0_Mean'
	'NorVol_Range'
	'NorF0Res_DisBySubMean_D2'
	'NorF0_DisBySubPC_D6'
	'NorVol_DisBySubPC_D7'

Table 2 shows common prosodic characteristics between L2_Eng and L1_Man. Pairwise discrimination between L2_Eng and L1_Man is conducted by prosodic feature and most similar features are listed in Table 2. Results show most similar prosodic features by highest 5 p-values by L2_Eng vs. L1_Man are 'histogram_dimension#5 by pairwise contrast of normalized volume', 'histogram_dimension#1&3 by pairwise contrast of normalized duration', ' histogram_dimension#5 by normalized duration without intrinsic properties' and 'pairwise contrast by normalized F0' in sentence layer and 'mean by normalized F0', 'range by normalized volume', 'histogram_dimension#2 by f0 without intonation effect', 'histogram_dimension#6 by normalized F0'and 'histogram_dimension#7 by normalized volume in word layer.

4.2.1 Discussion

The results show F0/duration/volume in sentence layer and volume in word layer contribute to TW L2 accent. By discrimination analysis between L2_Eng and L1_Man, results demonstrate F0/duration/volume in sentence layer and F0/volume in word layer are shared L2_Eng-L1_Man prosodic properties. We further assume that distinct features of L2 accent might attribute to prosodic characteristics borrowed from their mother tongue, namely L1_Man thus distinct features of L2Eng are compared with L2Eng-L1Man shared features. The results show distinct L2_Eng features do overlap with L2Eng-L1Man common features. Comparison by sentence layer shows similar features found coexisting in L2Eng-L1Eng distinct features and L2Eng-L1Man common features (green in Table 1 and Table 2) are stand deviation by normalized duration in L1Eng-L2Eng distinct features and histogram_dimension#1&3 by pairwise contrast of normalized duration in L2Eng-L1Man common features. Pairwise contrast is defined by between-phone variation and the property is similar to stand deviation representing global variation; thus we could regard them as overlap. In summary, the results suggest tempo contrast by syntax-elicited narrow focus in sentence layer and loudness range by lexicon-defined word stress in word layer are distinct L2 features of TW English which might attribute to prosodic transfer of Mandarin, namely L2s’ mother tongue.

4.3 Similarity Comparison by Prosodic Patterns

In addition to analysis by individual prosodic feature in section 3.2, similarity is computed between any two of L1/L2 speakers by prosodic patterns of word/sentence. After between-speaker similarity is derived, we examine if between-speaker similarity is greater when they are in the same speaker group. The aim is to test if consistency within each speaker group (L1/L2) and discrimination between L1 and L2 could be found.

4.3.1 Similarity in Word Prosody

Figure 3, 4 and 5 show similarity matrix between any two of L1/L2 speakers by prosodic patterns of word. First row by normalized duration in Figure 3 demonstrates by color lightness, first L1 speaker is more similar with speaker 1 to speaker 9 than speaker 10 to speaker 18 which represent L1 speakers and L2 speakers respectively. In addition, the left-top block by green dotted cross demonstrates L1 speakers with more consistency within group than the other blocks. It suggests L1 with greater cohesion/consistency than right-top (L1 vs. L2), left-bottom (L1 vs. L2) and right-bottom (L2 vs. L2). Right-bottom (L2 vs. L2) block also shows secondary consistent which is darker than right-top (L1 vs. L2), left-bottom (L1 vs. L2). It suggests L2s' prosodic patterns are consistent as well. Normalized duration without intrinsic properties in Figure3 further shows that removing intrinsic duration could further help to discriminate L1 and L2.

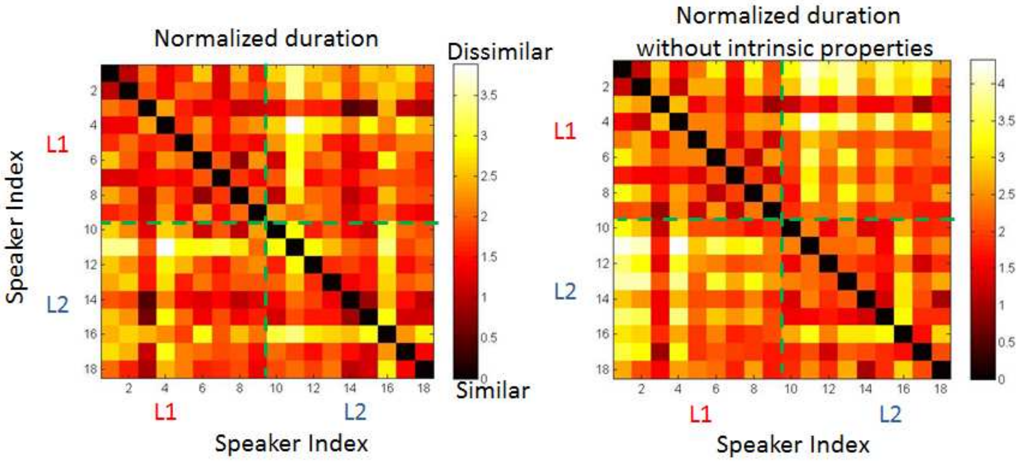


Figure 3. The similarity between any two of L1/L2 speakers by duration patterns in word layer. Color bars show the more dark the color, the more similar between two speakers. The value of point (i,j) in the matrix represents cosine distance between i and j that diagonal indicates self-similarity with darkest color. The green dotted cross represents boundary between L1 and L2 speakers.

Figure 4 also shows great cohesion within speaker group (L1&L2) respectively and great difference between speaker group (L1 vs. L2) by normalized F0 and normalized F0 without intonation effect; however, removing intonation appears not to improve L1-L2 discrimination significantly.

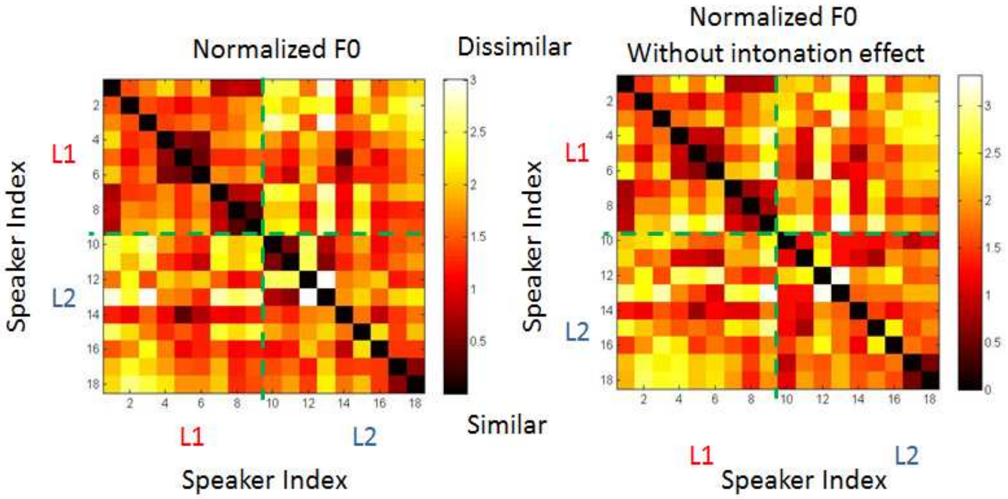


Figure 4. The similarity between any two of L1/L2 speakers by F0 patterns in word layer.

Figure 5 shows similarity matrix by normalized intensity. Results show no significant discrimination found between L1 and L2.

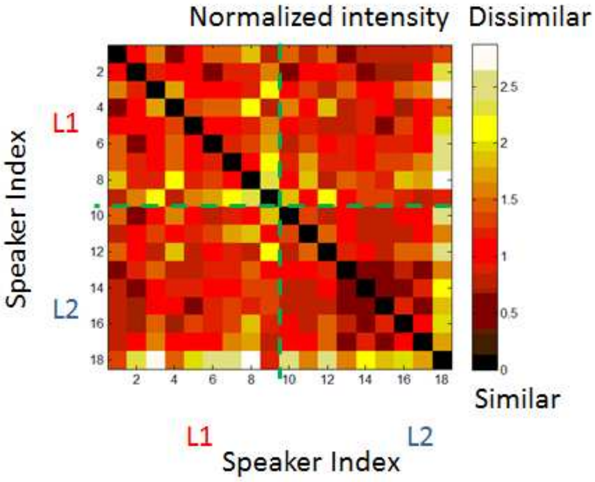


Figure 5. The similarity between any two of L1/L2 speakers by intensity patterns in word layer.

4.3.1.1 Discussion

By between-speaker similarity of word by duration/F0, the two distinct blocks by shading value representing L1s' and L2s' patterns are found. It suggests between-speaker similarity by word layer is greater when they are in the same speaker group. In other words, L1 and L2 produce respective timing/pitch patterns of word with great within-group consistency but within-group features are distinct from counterpart group. Between-group discrimination and within-group consistency is not found by loudness patterns. The results suggests timing/pitch patterns elicited by lexicon-defined word stress in word layer are distinct L2 features of TW English.

4.3.2 Similarity in Sentence Prosody

Figure 6, 7 and 8 show similarity matrix between any two of L1/L2 speakers by prosodic patterns of sentence. By Figure 6 and 7, no significant discrimination between L1 and L2 is found by normalized duration, normalized duration without intrinsic properties, normalized F0 and normalized F0 without intonation.

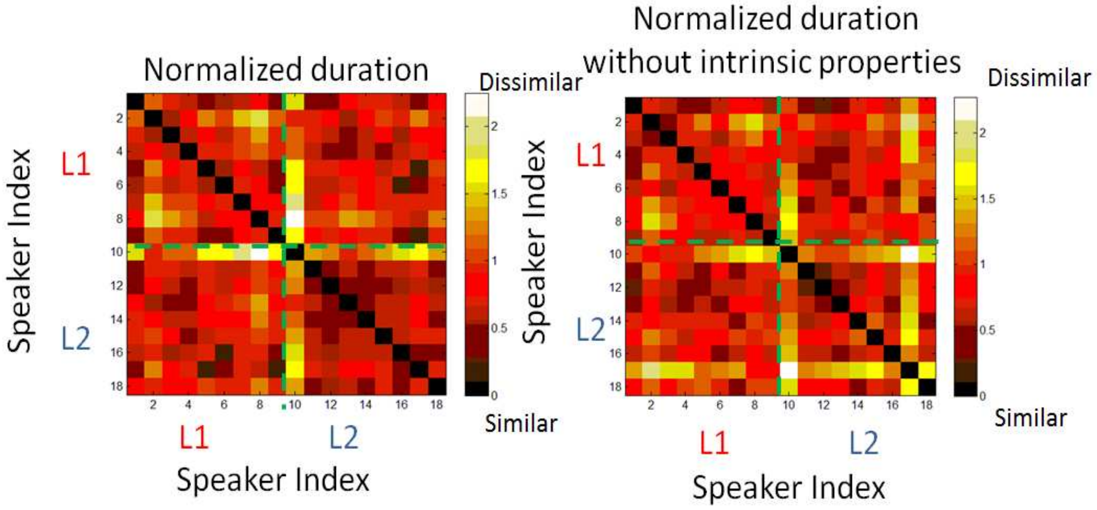


Figure 6. The similarity between any two of L1/L2 speakers by duration patterns in sentence layer.

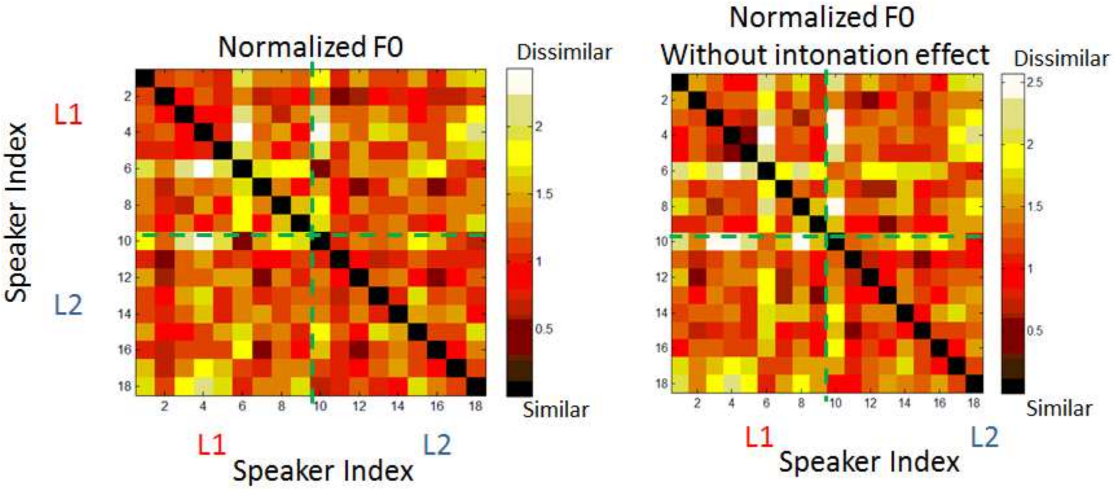


Figure 7. The similarity between any two of L1/L2 speakers by F0 patterns in sentence layer.

Figure 8 shows intensity patterns of sentence with great within-group cohesion and great between-group difference in both L1 and L2.

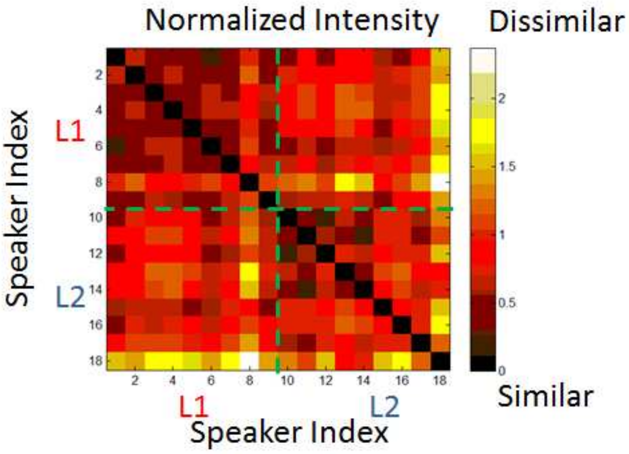


Figure 8. Similarity between any two of L1/L2 speakers by intensity patterns in sentence layer.

4.3.2.1 Discussion

By intensity similarity of sentence, the two distinct blocks by shading value representing L1s' and L2s' patterns are found. It suggests between-speaker similarity by intensity of sentence is greater when they are in the same speaker group. In other words, L1 and L2 produce respective prosodic patterns with great within-group consistency but within-group features are discriminative to counterpart group. Between-group discrimination and within-group consistency is not found by timing/pitch patterns. The results suggest loudness patterns elicited by syntax-induced narrow focus in sentence layer are distinct L2 feature of TW English.

5. Discussion and Conclusion

The present study examines prosodic characteristics of Taiwan English in relation to native English and Mandarin, mother tongue of TW speakers. Prosody classification among native English, TW L2 English and TW Mandarin is conducted by machine learning technology and results show Taiwan L2 English is found to be distinct from L1 English in prosody. However, TW L2 English and Taiwan Mandarin share some common prosodic characteristics which differentiate them from L1_Eng. Further comparison by each prosodic feature shows distinct L2 features of TW English can be attributed to prosodic transfer of Mandarin is tempo contrast elicited by syntax-induced narrow focus in sentence layer and loudness range by lexicon-defined stress in word layer. By examining prosodic patterns of word/sentence, similarity analysis suggests that between-speaker similarity is greater when they are in the same speaker group in both word and sentence layer. In other words, L1 and L2 speakers produce respective prosodic patterns with great within-group consistency but their within-group patterns are discriminative to counterpart group by loudness patterns in sentence layer and timing/pitch patterns in word layer. We believe the above study with incorporated linguistic knowledge not only sheds light on better understanding of TW L2 English, but can also be applied CALL system implementation. Future works will include providing prosody evaluation matrix of L2 by word and by sentence with degree measures of similarity and improvement scoring so that L2 learners will become more sensitive to prosody features.

Reference

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of non-native pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*, 529-555.

Benrabah, M. (1997). Word-stress: A source of unintelligibility in English. *IRAL*, XXXV(3), 157-165.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge, England: Cambridge University Press.

Coniam, D. (1999). Voice Recognition Software Accuracy with Second Language Speakers of English. *System*, *27*(1), 49-64.

Coomans, D., & Massart, D. L. (1982). Alternative k-nearest neighbour rules in supervised pattern recognition : Part 1. k-Nearest neighbour classification by using alternative voting rules. *Analytica Chimica Acta*, *136*, 15-27.

Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *19*(1), 1-16.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*, 379-397.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*(3), 393-410.

Grabe, E. & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In Gussenhoven, C. & Warner, N. (eds.) *Papers in Laboratory Phonology 7*, Berlin, Mouton de Gruyter, 515-546.

Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology*, *8*, 34-52. http://llt.msu.edu Retrieved from

Hirata, Y. (2004). Computer-assisted pronunciation training for native English speakers learning Japanese pitch and duration contrasts. *Computer Assisted Language Learning*, *17*, 357-376.

Lehmann, E. L. (1997). Testing Statistical Hypotheses: The Story of a Book. *Statistical Science*, *12*(1), 48-52.

Moustroufas, N., & Digalakis, V. (2007). Automatic pronunciation evaluation of foreign speakers using unknown text. *Computer Speech and Language*, *21*(1), 219-230.

Munro, M. J. (1995). Nonsegmental factors in foreign accent: ratings of filtered speech. *Studies in Second Language Acquisition*, *17*, 17-34.

Munro, M. J., & Derwing, T. M. (1999), Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *49*(Supp. 1), 285-310.

Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, *15*(5), 441-467.

Nguyen, T. A. T., Ingram, J., & Pensalfini, R. (2008). Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics*, *36*(1), 158-190.

Pennington, M. (1999). Computer-aided pronunciation pedagogy: Promise,limitations, directions. *Computer Assisted Language Learning*, *12*(5), 427-440.

Scruton, R. (1996). The eclipse of listening. *The New Criterion*, *15*(30), 5-13.

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, *25*, 1-24.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*, 1-30.

Tseng, C.-Y., & Su, C.-Y. (2014). Prosodic Differences between Taiwanese L2 and North American L1 speakers—Under-differentiation of Lexical Stress. *Speech Prosody 2014*, Dublin, Ireland.

Tseng, C-Y., Su, C.-Y., & Visceglia, T. (2013). Underdifferentiation of English Lexical Stress Contrasts by L2 Taiwan Speakers. *Slate 2013*, 164-167. Grenoble, France.

Tseng, C.-Y., Cheng, Y.-C., Lee, W.-S. & Huang, F.-L. (2003). Collecting Mandarin speech databases for prosody investigations. *Oriented COCOSDA 2003*. Sentosa, Singapore.

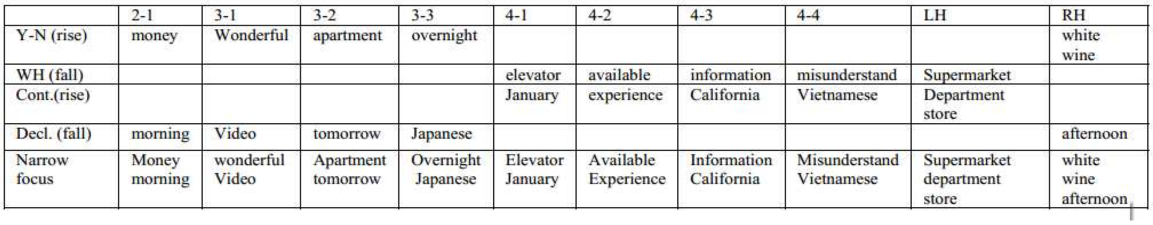
Visceglia, T., Tseng, C. Y., Kondo, M., Meng, H., & Sagisaki, Y. (2009). Phonetic aspects of content design in AESOP (Asian English Speech cOrpus Project). *Oriental COCOSDA 2009*. Beijing, China.

Visceglia, T., Tseng, C. Y., Su, Z. Y. & Huang, C. F. (2011). Realization of English Narrow Focus by L1 English and L1 Taiwan Mandarin Speakers. *the 7th International Congress of Phonetic Sciences*. Hong Kong, China.

Visceglia, T., Su, C. Y., & Tseng, C. Y. (2012). Comparison of English Narrow Focus Production by L1 English, Beijing and Taiwan Mandarin Speakers. *Oriental COCOSDA 2012*, 47-51. Macau, China.

Witt, S. M., & Young, S. J. (2000). Phone-level pronunciation scoring and assessment for interactive language learning. *Speech Communication*, *30*(2-3), 95-108.

Appendix A. Target words by syllabicity, stress type and experimental condition



	2-1	3-1	3-2	3-3	4-1	4-2	4-3	4-4	LH	RH
Y-N (rise)	money	Wonderful	apartment	overnight						white wine
WH (fall)					elevator	available	information	misunderstand	Supermarket	
Cont.(rise)					January	experience	California	Vietnamese	Department store	
Decl. (fall)	morning	Video	tomorrow	Japanese						afternoon
Narrow focus	Money morning	wonderful Video	Apartment tomorrow	Overnight Japanese	Elevator January	Available Experience	Information California	Misunderstand Vietnamese	Supermarket department store	white wine afternoon