L1/L2 Difference in Phonological Sensitivity and Information Planning - Evidence from F0 Patterns

Chao-yu Su [1, 2, 3] *& Chiu-yu Tseng* [1]

[1] Institute of Lingustic, Academia Sinica, Taiwan
[2] Taiwan International Graduate Program, Academia Sinica, Taiwan
[3] Institute of Information Systems and Application, National Tsing Hua University, Taiwan
cytling@sinica.edu.tw

Abstract

Assuming that linguistic specifications and information planning contribute to different levels of prosodic organization that cumulatively constitute output prosody, quantitative analysis of respective contributions can be derived through normalization procedures that remove levels of interactions involved. The current study attempts to account for how L2 prosody departs from the L1 norm in the two levels mentioned and whether an account can be offered. F0 patterns of word English stress categories (primary, secondary and tertiary) and emphases in controlled conditions (narrow-, broad- and non-focus) are compared using speech data from English L1 and Mandarin L2 speakers. L1 speech exhibits similar F0 patterns of binary high-low contrasts in both stress/non-stress as well as focus/non-focus categories, suggesting comparable planning are used to express phonological and information planning. However, L2's primary stress and emphasis exhibited less degree of F0 high-low contrast, coupled with reversed F0 patterns in both the secondary and tertiary categories as well as non-emphases conditions. The results demonstrate that being less sensitive to phonological categories may also affect information planning in similar ways. We believe the results explain how stress and focus interact to cause L2 accent and unintelligibility, help understand stress and focus composition of L1-and-L2 speech, ,and are readily applicable to CALL.

Index Terms: L2 English, communicative function, lexical stress, focus structure, F0 constitution

1. Introduction

Foreign accent is found related to effectiveness of speech communication. Previous studies show that presence of foreign accent leads to communication breakdown and compromised intelligibility [1]. Nonnative accented utterances require more time for native speakers to process than native-produced speech [2], thus illustrating why processing foreign accent is more demanding cognitively for both native and non-native listeners. However, since the majority of reported works focused on segmental variations, how foreign accent degrades communication in the prosodic domain remains less known in spite of some recent reports. For example, though studies did demonstrate that multiple communicative functions are conveyed through prosody, including at least lexical, syntactic, semantic, information planning…etc. [3, 4, 5], the prosodic aspects of foreign accent still merit more investigation due largely to its close correlation with linguistic structures, information planning as well as communicative intents.

On the L1 side, it is now well accepted that communicative functions are reflected in various prosodic levels and jointly attribute to prosody output to achieve the intended communication goal [3, 4, 5, 6]. For example, it is reported that at least four prosody-related linguistic specifications have been found to relate to communicative functions, namely, lexical stress, focus structure, sentence modality (question vs. non-question) and segmentation [7]; the prosodic levels involved are phonological (stress), information planning (focus structure), syntactic (sentence modality) and phrasing (segmentation). Another study compared misplaced stress patterns with mispronounced phonemes and found that misplaced stress is three times more likely to break down communication than phoneme [8]. Later acoustic studies on word level lexical stress revealed that both Japanese [9] and Taiwan (TW) Mandarin L2 [10] English speakers do not produce sufficient acoustic contrast between stressed and unstressed syllables as native speakers do, and such under-differentiated contrast inhibits intelligibility. A significant correlation is also found between misplacement or absence of prosodic sentence stress cues and L2 speakers' level of comprehensibility for both naïve and expert listener groups [11]. Furthermore, similar patterns of under-differentiated contrast are found in narrow focus produced by Mandarin L2 speakers [12]. Yet the collective effects of stress and focus status still remain unknown. In other words, the majority of reported L2 studies concentrated more on how each contributing prosodic factor may differ from L1, but less on how these contributing factors interact. Since output prosody is the integrated outcome of complex interactions of multiple contributions, it is reasonable to assume that L2 learners would have difficulty to pinpoint exactly why their prosodic modulation is different from the L1 norm and how to make improvement.

In contrast to previous studies of single linguistic specification and its corresponding prosodic level and effect, the current study thus attempts to concentrate on interactions by studying how lexical stress and focus structure may collectively contribute to F0 contour as a first step towards better understanding foreign accent. The selected tool of analysis is the command-response model [4, 13] which by definition categorizes F0 contour into long-term/global tendency (phrase command), short-term/local humps (accent commands) and a constant (base frequency). The global tendency is found associated with paragraph association [6] and will not be addressed in the present study. The accent command, by definition, is in relation to local F0 trend; however, how accent commands may interact with linguistic specifications is still not clear. A preliminary study on German shows the relation between accent command and focus structure and provided

some examples that showed how narrow focus would significantly boost the magnitude of accent command [14]. Along the same vein, the present study further assumes that accent commands are related to lexical stress at the lower prosodic level while focus structure would superimpose from a higher prosodic level to collaboratively contribute to accent commands. Thus patterns of lexical stress (primary, secondary and tertiary) and superimposing effect from focus structure (narrow focus, broad focus and non-focus) will be derived from accent commands for both L1 and L2. Our goal is two-fold. One is compare the L1-L2 difference at levels of lexical stress and focus structure while considering their interaction at the same time. Another is to further model L1's accent commands by lexical stress and focus structure jointly to test the predictability. We will show in the end how lexical stress and focus structure composite and account for allocation and magnitude of accent commands of L1.

2. Speech Materials and Annotation

Subsets of the AESOP-ILAS [15] speech database are used for the present study. AESOP (Asian English Speech cOrpus Project) is a multinational collaboration of data collection specifically designed to elicit features of L2 English across Asia. The materials used include Task1 and Task 3 which were designed to elicit lexical stress and contrastive focus, respectively. A total of 20 frequency- controlled and stress balanced (2-4 syllables) target words were embedded in carrier sentences (Appendix A) and in sentences controlled for broad and narrow focus (Appendix B). Speech data of a total of 41 speakers are analyzed 11 L1 North American L1 speakers (5M 6F) and 30 TW L2 speakers (15M and 15F). 40 sentences by each speaker are adopted and thus total 440 sentences of native speech and 1200 sentences of Taiwan Mandarin English are used for analysis.

2.1. Processing & Annotation

The speech data of L1 English, TW L2 English were tagged by multiple layers of linguistic specifications. The preprocessing layer is force-aligned segments by the HTK Toolkit followed by manual spot-checking by trained transcribers. Following the tagging of segment, lexical stress (primary, secondary and tertiary) is labelled manually in syllable unit by dictionary transcription. Focus status (narrow focus broad focus and non-focus) is tagged in word unit by an English native speaker and further aligned into corresponding syllable units. By aligning with syllable, accent commands at different level could be analyzed by consistent unit. An example is as follows.

Table 1. *An L1-annotated example by focus status and lexical stress aligned with syllable for "No. I usually buy fruit at the SUPERMARKET because they stay open later". 3, 2, 1 by focus status represent narrow focus broad focus and non-focus and 2, 1, 0 by stress type represent primary, secondary and tertiary stress.*

Text	No	I	usually				buy	fruit	a	the	supermarket			
Focus	1	1	2	2	2	2	1	2	1	1	3	3	3	3
Stress	2	2	2	0	0	0	2	2	2	2	2	0	1	0

Text	because		the	stay	open		later	
Focus	1	1	1	1	2	2	2	2
Stress	0	2	2	2	2	0	2	0

3. Method

Since that linguistic/phonological variables can be attributed to various prosodic levels while jointly contribute to output prosody, relative acoustic as well as linguistic/phonological variables are postulated and described in 3.1.1 and 3.1.2, respectively, in order to compare L1-L2 acoustic differences at levels of lexical stress and focus structure. At the same time, a subtraction procedure separating higher –level focus effect form lower –level lexical stress for acoustic variables is proposed and described in 3.2.

3.1. Variables for modeling accent command

3.1.1. Acoustic variables

The major acoustic variable F0 is extracted by the command-response model [4, 13]. The model, by definition, decomposes three contributing components as long-term/global tendency (phrase component), short-term/local humps (accent component) and a constant (base frequency). The 3 components are represented by (1). A previous method based on filter is adopted for parameter extraction [16]. Two major acoustic variables, namely position and magnitude of accent command are used for analysis and modeling in the present study.

$$F_0(\mathrm{t}) = \ln(F_b) + \sum_{i=1}^{I} A_{pi} G_p(t - T_{0i}) + \sum_{j=1}^{J} A_{aj}[G_a(t - T_{1j}) - G_a(t - T_{2j})] \quad (1)$$

where $G_p(t) = \alpha^2 t \exp(-\alpha t), for\ t \geq 0$

$G_a(t) = min[1 - (1 + \beta t)\, exp(-\beta t), \gamma], for\ t \geq 0$

$\alpha = 3,\ \beta = 20$

3.1.2. Linguistic/phonological variables

L2 linguistic, in this case phonological, variables are directly derived by annotation (X1-X12 in Table2). The present study further assumes information context also plays an important role in planning of accent command and information density is defined as average neighborhood amount by information content conveyed by focus degree at current position and surrounding context. The scale setup for surrounding context in the present study is pre- and post- 2 syllables/words. Table2 lists total 14 linguistic/phonological variables

$ID_i = \frac{1}{2n+1}\sum_{i=-n}^{n} X4_i$ (2) where *i* is the positon index of current word

Table 2. *A summary of linguistic/phonological variables*

Code	Feature	Code	Feature
X1	Level of pre boundary break	X8	Stress type
X2	Level of post boundary break	X9	Contrast with previous
X2	Focus index by phrase position	X10	Contrast with post stress
X4	Focus degree	X11	Relative position by
X5	Contrast with previous focus	X12	Type of boundary effect
X6	Contrast with post focus	X13	Infor Density By Syllable
X7	Relative position by narrow focus	X14	Infor Density By Word

3.2. Separating higher –level focus effect form lower –level lexical stress

Focus structure is assumed as superimposing its effect onto lexical stress. An additional subtraction procedure is listed as follows for deriving separated focus contribution from lexical stress.

$$X_F = X_o - X_s,\ X_o = \begin{bmatrix} X_{s_1} \\ X_{s_2} \\ \cdots \\ X_{s_i} \end{bmatrix},\quad X_s = \begin{bmatrix} M_{k_1} \\ M_{k_2} \\ \cdots \\ M_{k_i} \end{bmatrix} \qquad (3)$$

$$k_i = ST_J\ ,\quad J = \delta_{s_i=ST_J}(s_i),\quad M_{k_i} = M_{ST_J} = \frac{1}{N_{ST_J}} \sum_{s_i \in ST_J} X_{S_i}$$

where X_F, X_O, Xs, s, ST, i, j, represent focus effect, original patterns by magnitude of accent command, lexical stress effect, stress variable, stress type, index of accent command and index of stress type respectively.

3.3. Modeling procedure for accent command

Position and magnitude of accent command are two major observations to model in the present study. Modeling is through a two-step procedure. The position model marks each syllable for presence/absence of accent command. After position of accent command marked, magnitude of the accent commands is further modeled.

3.3.1. Modeling position of accent command

In order to classify each syllable into presence/absence of accent command, a decision tree is adopted. The decision tree is a tree-like model which predicts response variable by decision rules from the root node down to a leaf node in the tree data [17]. Split criterion in the present study is Gini's diversity index

3.3.2. Modeling magnitude of accent command

Three Regression techniques are adopted for approximating magnitude of accent command by refined linguistic variables. 3 Regression techniques are multivariable linear regression, robust regression and neural network. Multivariable linear regression (MLR) approximates the relationship between a response variable and linear combination of explanatory variables [18]. In the present study, the response variable is the magnitude of accent command and the explanatory variables are linguistic variables by focus structure, lexical stress and articulatory continuity. Robust regression (RoFit) is an extension of multivariable linear regression and it creates a model that is less sensitive by outliers by iteratively reweighted least squares [19]. A feedforward neural network (FNN) is a modeling technique for approximating response variable by non-linear functions which contains sets of adaptive weights learned from explanatory variables [20]. The layer number used here is 30.

4. RESULTS

4.1. Magnitude of accent command by stress type, focus structure and L1/L2

4.1.1. L1-L2 difference at stress level

Figure 1 shows magnitude patterns of accent command by stress type for L1 and L2. Mandarin L2 English shows less degree of contrast among primary, secondary and tertiary stresses than L1.

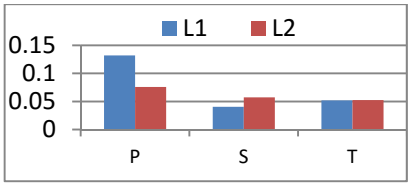

Figure 2. *Magnitude patterns of accent command by stress type and L1/L2. P, S and T represent primary, secondary and Tertiary stress.*

4.1.1.1 Discussion

The results show under-differentiation of Mandarin L2 English at the level of lexical stress, especially how L1 realized the three-way stress contrast in an optimal binary (stress/un-stress) ways by merging the unstressed categories. TW L2 speech exhibits a general pattern of less robust degree of contrast. It is therefore reasonable to assume that the phonological contrast required to represent lexical stress is difficult for nonnative speakers through their under-differentiated realization of F0. When lexical stress is not realized as robustly as do native speakers, part of the communicative functions would be hampered.

4.1.2. L1-L2 difference at focus level

Figure 2 and Figure 3 show magnitude patterns of accent command by lexical stress and focus type for L1 and L2 respectively. Figure 4 shows separated narrow focus effect by removing lexical stress.

4.1.2.1 L1-L2 difference by lexical stress and focus type

Comparison between L1 (Figure 2) and L2 (Figure 3) shows magnitude difference by lexical stress and focus type. In general, TW L2 English shows less contrast degree among narrow focus, broad focus and non-focus than L1. L1's accent commands significantly boost the primary stressed syllable of words that are narrow-focused.

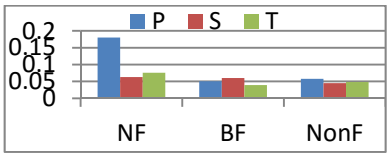

Figure 2. *Magnitude patterns of accent command by lexical stress and focus type for L1. P, S and T represent primary, secondary and Tertiary stress. NF, BF, NonF represent narrow focus, broad focus and non focus.*

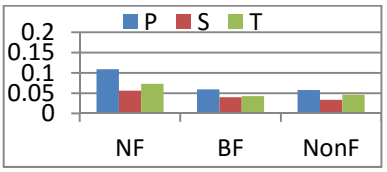

Figure 3. *Magnitude patterns of accent command by lexical stress and focus type for L2.*

4.1.2.2 L1-L2 difference by narrow focus effect separated from lexical stress

Narrow focus effect extracted by removing lexical stress for magnitude of accent command is further examined. The superimposed narrow focus is aligned with lower-level stress type to examine interaction between two levels. Patterns

shown in Figure 4 present the L1-L2 comparison. In general, the results indicate that TW L2 English at the focus level is less contrastive than L1. L1 patterns show how discrimination between lexical stresses is not only retained but also boosted at when focused. However, TW L2 English demonstrates less differentiated primary and tertiary stresses while secondary stress departs from the L1 norm even more. It is only reasonable that these distinct patterns further reduced lexical stress discrimination.

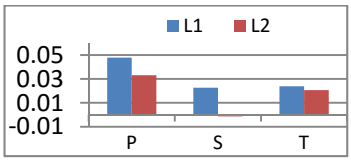

Figure 4. *Narrow focus effect separated from lexical stress by magnitude of accent command for L1 and L2. P, S and T represent primary, secondary and Tertiary stress.*

4.1.2.3 Discussion

The results at focus level without lexical stress show (1) under-differentiation of TW L2 English, and (2) superimposing focus effect with lexical stress discrimination is also difficult for TW L2 speakers. When focus structure is not realized fully, communicative functions are further hampered in addition to lexical under-differentiation at the lower-level. Note also how L1 speakers only maintained the NF/others binary contrast as they do with stress merging.

4.2. Modeling accent command by lexical stress and focus structure

4.2.1. Modeling position of accent command

The position of accent command is modeled by decision tree and overall accuracy for prediction is 93.66%. Most important factors found are 'contrast with previous focus (X5)', 'focus degree (X4)' related to focus structure and 'information density by syllable (X13)', 'information density by word (X14)' related to information density.

4.2.1.1 Discussion

The high accuracy (93.66%) of prediction suggests that L1's positions of accent commands are systematic and jointly predictable by lexical stress, focus structure and articulatory continuality. However, the respective weight of contribution with respect to predicting the presence/absence of accent command differs. 3 proposed factors differs Respective weighting of contribution from higher to lower ranks is derived. Contributions from focus structure, lexical stress and articulatory continuity are now known; their interaction can also be accounted for. Thus, the contribution spectrum by the 3 factors accounts for coarse grained planning of accent command

4.2.2. Modeling magnitude of accent command

The magnitude of accent command is modeled by different regression methods and the error of root mean square by each method is listed in Table 3. It turns out that 30-layer FNN performs the best. However, the difference among the 3 methods is not significant. Following the modeling, the contribution weight by two types of linear regression is further analyzed and listed in Table 4. The top-3 contributing weights are identical cross regression types and they are 'focus degree' related to focus structure, 'stress type' and 'relative position by primary stress' related to lexical stress

Table 3. *Root mean square error by multivariable linear regression (MLR), Robust regression (RoFit) and feedforward neural network (FNN)*

Regression method	MLR	RoFit	FNN
RMSE	0.1166	0.1176	0.1102

Table 4. *Contributing weight by MLR and RoFit.*

Regression / Refined linguistic feature	MLR	RoFit	Regression / Refined linguistic feature	MLR	RoFit
Level of pre boundary break	1.38	1.74	Stress type	2.36	2.2
Level of post boundary break	-0.89	-0.92	Contrast with previous	-1.89	-1.74
Focus index by phrase position	-1.03	-0.74	Contrast with post stress	0.09	0.15
Focus degree	2.88	2.71	Relative position by	2.41	2.3
Contrast with previous focus	0.09	0.02	Type of boundary effect	-1.56	-1.97
Contrast with post focus	0.58	0.63	Infor Density By Syllable	-1.03	-0.48
Relative position by narrow focus	0.99	0.36	Infor Density By Word	1.61	1.52

4.2.2.1 Discussion

The results showed that L1's magnitude of accent commands is systematic and predictable by lexical stress, focus structure and articulatory continuity jointly. The analysis show top-3 contributing weights by MLR and RoFit are identical. The contributing weights are focus structure>lexical stress> articulatory continuality; showing how higher level contribution outweighs the lower counterparts. The spectrum by the 3 factors thus accounts for finer grained planning of accent command.

5. Discussion

The above results reveal how the local F0 patterns of TW L2 English are different from L1 by both lexical stress and focus structure. L2 patterns of lexical stress show under-differentiation at lower level while superimposed focus structure is also found to be less contrastive than L1 patterns. The overall cumulative effects of under-differentiation from both levels jointly contribute to L2 accent, with a clearer account of the constitution of unintelligibility of TW L2 English. Following the L2 patterns, L1's local F0 is further modeled by lexical stress and focus structure. The results suggest the L1's F0 planning is systematic and closely related to information allocation. Articulatory continuity is also found as an important factor involving with the information planning.

6. Conclusions

The present study successfully teased apart the F0 constitution by lexical stress, focus structure, articulation continuity and their interaction to account for unintelligibility of TW L2 English. While all three factors examined are systematic and predictable for L1 prosody, it is not the same for L2. We learned now that while L1 merge multiple contrasts into binary opposition for both phonological contrast and focus structure, L2's under-differentiated phonetic realization of lexical-phonological contrast as well as indistinct realization of higher level information-focus structure jointly attribute to reduced intelligibility and foreign accent. We believe these results help explain how stress and focus interact to cause L2 accent and unintelligibility, help understand stress and focus composition of L1-and-L2 speech and are readily applicable to CALL. Future work will center on global F0 constitution associated with speech paragraph and its interaction with local features.

7. References

[1] Cheng, L. L., "Moving beyond Accent: Social and Cultural Realities of Living with Many Tongues", TOPICS IN LANGUAGE DISORDERS 19(4) · AUGUST 1999

[2] Munro, M. J. and Derwing, T. M., "Processing Time, Accent, and Comprehensibility in the Perception of Native and Foreign-Accented Speech", Language and Speech, July/September 1995 vol. 38 no. 3289-306, 1995.

[3] Bailly, G., Holm, B., "SFC: a trainable prosodic model", Speech Communication 46: 348-364, 2005.

[4] Fujisaki, H., Wang, C., Ohno, S., Gu, W., "Analysis and synthesis of fundamental frequency contours of Standard Chinese using the command–response model", Speech communication 47: 59-70, 2005.

[5] Xu, Y. "Speech melody as articulatorily implemented communicative functions", Speech Communication. 46, 220–251, 2005.

[6] Tseng, C. Y., Pin, S. H., Lee, Y. L., Wang, H. M. and Chen Y.C., "Fluent speech prosody: framework and modeling", Speech Communication, Special Issue on Quantitative Prosody Modelling for Natural Speech Description and Generation 46(34): 284-309, 2005.

[7] Mixdorff, H, "Speech Technology, ToBI, and Making Sense of Prosody", In Bel, Bernard & Marlien, Isabelle (Eds.) Speech Prosody 2002. Proceedings, Aix-en-Provence, France, 2002.

[8] Bond, Z., and Small, L. H., "Voicing, vowel and stress mispronunciations in continuous speech", Perception and Psychophysics, 34, 470–474, 1983.

[9] Nakamura, S. "Analysis of Relationship between Duration Characteristics and Subjective Evaluation of English Speech by Japanese learners with regard to Contrast of the Stressed to the Unstressed", Journal of Pan-Pacific Association of Applied Linguistics, 14(1), 1-14, 2010.

[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] Warren, P., Elgort, I., and Crabbe, D. "Comprehensibility and prosody ratings for pronunciation software development", Language Learning & Technology, 13(3), 87-102, 2009.

[12] 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.

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

[14] Mixdorff, H, "An Integrated Approach to Modeling German Prosody". Volume 25, Studientexte zur Sprachkommunikation, Dresden, 2002.

[15] 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. Beijing, China, 2009.

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

[17] Utgoff, P. E. "Incremental induction of decision trees", Machine learning, 4(2), 161-186, 1989.

[18] Pedhazur, E J., "Multiple regression in behavioral research: Explanation and prediction" (2nd ed.). New York: Holt, Rinehart and Winston, 1982

[19] Andersen, R., "Modern Methods for Robust Regression", Sage University Paper Series on Quantitative Applications in the Social Sciences, 2008.

[20] Auer, P., Harald, B., Wolfgang, M., "A learning rule for very simple universal approximators consisting of a single layer of perceptrons", Neural Networks 21, 2008.

8. Appendix

Appendix A

Task 1:

Carrier sentence: "I said TARGET WORD five/ten times." 20 Target words by syllabicity (2-4) and stress type (syllable number/primary stress position): money, morning, wonderful, video, apartment, tomorrow, overnight, Japanese, elevator, January, available, experience, information, California, misunderstand, Vietnamese, supermarket, department store, white wine, afternoon.

Appendix B

Examples of Task 3:

Target words in narrow focus:

1. Context: Are we allowed to make audio and video recordings?
   Answer: No. VIDEO recordings are not allowed.
2. Context: Have you been trained to do this job?
   Answer: No. But I think EXPERIENCE is more important than training.