L2 Stress Perception: The reliance on different acoustic cues

Qian Wang

Department of Linguistics
University of Victoria
wangqian@uvic.ca

Abstract

Second language learners of English often experience difficulties in English lexical stress perception because of prosodic transfer of their first language (L1). It has thus been hypothesized that the problem of Chinese learners with English stress arises from tonal transfer. However, little research has been devoted to the investigation of the phonetic details of second language (L2) stress perception. The present research focused on Chinese L2 learners of English. Their reliance on the three acoustic cues, F0, intensity and duration, is compared against a baseline of native English speakers, in terms of English stress perception of manipulated nonsense tokens. The results show that while all three cues have significant effects on stress perception for native English speakers, only F0 has a decisive effect on Chinese learners' stress judgments. This study suggests that there is transfer of reliance on F0 in the acquisition of L2 English stress, rather than transfer of tone at the phonological level.

1. Introduction

Lexical stress plays an important role in native speakers' perceptions and processing of speech [1][2]. However, researchers have found that learners from a non-stress language background may not possess a system of stress in the same way as native speakers do [3][4]. For example, French learners of English, with a fixed-stress background, are found to be 'stress deaf' [3]. In other words, they have difficulties perceiving stress contrasts.

Similarly, other researchers have shown that Chinese speakers, with a tonal background, also have problems in acquiring English lexical stress. This has been attributed to prosodic transfer of tone. Juff [5] found that Chinese speakers used Tone 1 with an inordinate degree of length to indicate lexical stress. For Cantonese speakers, Chao [6] indicated that they associated high and low tones with stressed and unstressed syllables. The Chinese subjects in Archibald's study [4] did not seem to have acquired principles of English stress assignment. They appeared to treat stress as a purely lexical phenomenon in the same way they treat tone. Archibald [4] suggested that this may result from their incapability "to utilize the cues for stress (vowel quality, heavy syllables, etc.)".

Despite the recognition of possible prosodic transfer from Chinese speakers' tonal backgrounds in the acquisition of English lexical stress, few acoustic analyses have been devoted to the investigation of the phonetic details of Chinese learners' acquisition of English stress, especially in terms of perception. As Flege [7] pointed out in his research on L2 speech development at the segmental level, it is important to take phonetic details into account in order to gain a better understanding of the possible transfer of L1. The same is true for studies at the prosodic level [8][9]. In order to expand the understanding of the acquisition of English stress by Chinese learners and offer insights into the more general area of L2 prosodic acquisition, detailed phonetic and instrumental studies are necessary.

This study focuses on the acoustic cues used by Chinese learners of English in stress perception. It may be possible that the observed L1 transfer of tone is, in fact, a result of a heavy reliance on F0, rather than intensity or duration, in the acquisition of English stress. To compare the use of acoustic cues in English stress perception, Chinese learners of English (CE) and native English speakers (NE) were asked to make stress judgments for disyllabic tokens with manipulated F0, duration and intensity values. The goals of this research are to answer three questions. First, are Chinese learners 'deaf' to English stress? Second, if they can perceive English stress, how are their perceptions affected by the three cues? Finally, do they rely heavily on F0 in English stress perception?

2. Perception Experiment

Previous research has used systematically manipulated lexical tokens to study the weight of different cues in L1 stress perception [10][11][12][13]. No similar studies have been reported in the L2 literature. The present study is an attempt to use the same technique for the investigation of L2 learners' perception of English stress.

2.1. Materials

Three two-syllable nonsense words were created, *latmab, nizdit,* and *tetsep*. These words conform to the English phonotactic structure of syllables. Three vowels were chosen, *[æ, ɪ, ɛ]*, and each word contained the same vowel in both syllables. The use of nonsense words helps to reduce performance differences between NE and CE that may arise based on their knowledge and familiarity with test tokens. Furthermore, using a same vowel in the two syllables allowed the researcher to investigate the effects of acoustic cues without confounding the intrinsic difference between vowels in terms of F0, duration and intensity. The three nonsense words were produced with two stress patterns by a trained phonetician. Four repetitions of each token were recorded. The most clear repetition was used as the basis for manipulation.

The manipulations of F0, duration and intensity were applied to the nucleus of each syllable in each nonsense word (see Table 1). There were five levels of manipulations for each cue. Each nonsense word was first manipulated to Level Three. In this level of manipulation, the two syllables had the same values of F0, duration or intensity. The values were achieved by averaging F0, duration and intensity contours of the two syllables in original production. Manipulation Level One, Two, Four and Five were based on Level Three. For example, Level One of F0 manipulation made the first syllable 50 Hz lower than the second. This was achieved by taking a Level Three

token, and reducing the first syllable by 25 Hz and raising the second syllable by 25Hz. In a similar way, Level Two of F0 manipulation brought the difference between the two syllables to 25 Hz. Level Four and Level Five were mirror images of Level two and Level One, respectively. In other words, in Level Four and Five, the first syllable was higher in F0 than the second syllable. The same five level procedure was used for the manipulations of duration and intensity. Table 1 represents the manipulation levels mentioned above.

Table 1: Manipulation of F0, Duration and Intensity

	Levels				
Correlates	1	2	3	4	5
F0	-50	-25	0	25	50
Duration (D)	0.5	0.75	1	1.25	1.5
Intensity (I)	-9	-4.5	1	4.5	9

For each word form, *tetsep, nizdit,* and *latmab,* 125 tokens were created by using all possible combinations of the F0, duration (D) and intensity (I) manipulation levels (5*5*5). Therefore, the three experimental word forms yielded a total of 375 tokens (3*125). In addition to these nonsense tokens, 100 real English words recorded by the same phonetician were also used in the perception experiment as foils. The responses to these tokens were used in the screening process.

2.2. Participants and Procedure

Two groups of listeners participated in the study. Sixty-eight university level Chinese learners of English (CE) were tested in a language lab in Nanjing University of Science and Technology, China. Thirty-eight university level native English speakers (NE) were tested in the phonetics lab in the Department of Linguistics at the University of Victoria, Canada. Participants in both groups listened to the test tokens over headphones connected to individual listening stations. In response to the audio presentation of the stimuli, the participants indicated the stress position by clicking one of the two choices on the screen, either *1st Syllable* or *2nd Syllable*. The next stimuli were presented automatically after a 2s interval, regardless of whether a choice was made or not. Each individual's responses were recorded and saved to a log file, where *A* represented the choice of first syllable stress and *B* represented the choice of second syllable stress and *0* represented timed-out responses.

The presentation of the stimuli was controlled by a *C++* program written for this experiment. The 475 tokens (375 nonsense tokens and 100 real English words) were presented in 25 blocks. There were 19 tokens in each block. Each block was introduced with a beep followed by 15 nonsense words and four real words. In each block, the 15 nonsense tokens were divided into five sets, each set containing one token of each of the three word forms with a real word between every set of nonsense tokens. The order of presentation for the 125 tokens of each word form was randomized for each participant. After each block, there was a 30s break. Proceeding the test, there were two practice blocks with exactly the same format but with tokens that were not used in the real test.

2.3. Analysis

The responses to the 100 real English words were used to screen for participants who may not have understood the perception task. Only participants who achieved 80% or higher success rate for the real words were used for further analysis. Thirty-four CE and 25 NE satisfied this criterion.

The number of Initial Stress (IS) judgments (i.e. stress on the first syllable) was determined for each token and the Initial Stress Percentage (ISP) was calculated for each token for the two groups separately. For CE, an ISP of 100% for a token means that all the 34 CE participants perceived this token to have IS. Similarly, for NE, an ISP of 100% means that all 25 NE participants agreed on the IS judgment. Statistical analysis was conducted on the two sets of ISPs to determine whether 1) all three cues are correlates for NE English stress perception, 2) the same is true for CE, 3) F0 is a more important cue for CE than for NE.

Two four-way univariate ANOVAs were conducted in SPSS, one for NE and one for CE. The dependent variable was ISP, the fixed factors being F0, intensity and duration, each with 5 different steps. Furthermore, the variable for Word Form was included as a covariate. Partial Eta squared (η^2) was calculated to determine the effect size of the different cues.

2.4. Results

A total of 22,125 responses were analyzed from the perception test (59 subjects * 375 tokens). Overall, 44.21% of the responses favored IS and 54.99% favored Final Stress (FS) and 0.8% of the results were timed-out responses. No IS preference was observed in the data.

The responses are summarized in the following two figures, Figure 1 for NE and Figure 2 for CE. The X-axis represents the five different manipulation levels. The Y-axis was the percentage of IS responses. The percentage of IS responses at each level of F0 manipulation in combination with all levels of duration and intensity manipulation is indicated by the square and similarly the ISP at each level of intensity manipulation is indicated by the diamond and duration by the triangle. Three lines were plotted to represent the change in the percentage of IS responses as the manipulation level changes from one to five for each of the three cues.

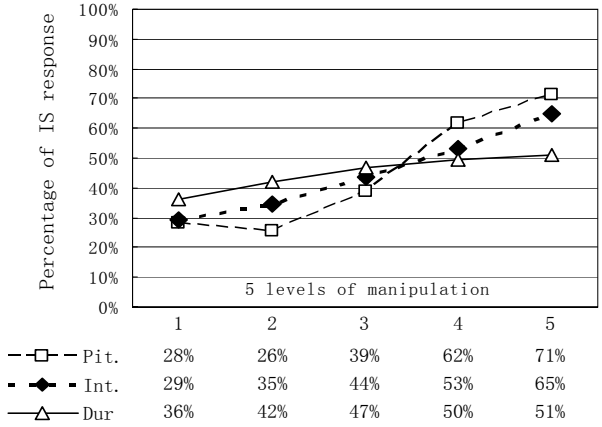

Figure 1: ISP of NE for the 375 words as a function of a) F0; b) intensity; and c) duration manipulation

Figure 1 shows that for NE, F0 change from level one to five resulted in an increase of IS responses, from 28% to 71%,

intensity from 29% to 65% and duration from 36% to 51%. An ANOVA showed that F0 had a significant effect on NE stress judgment, F (4, 249) = 87.947, p< 0.001 and so did intensity, F (4, 249) = 43.000 , p< 0.001 and duration. F (4, 249) = 7.459, p< 0.001. F0, alone, can explain 58.6% (η^2 = 0.586) of the variance in ISP, which was larger than the effect size of intensity (η^2 = 0.409) and duration (η^2 = 0.107). The results are consistent with previous studies in that all three cues are relevant correlates for stress perception for NE. Furthermore, the current research showed that the effect of F0 was primary and the effects of intensity and duration were secondary.

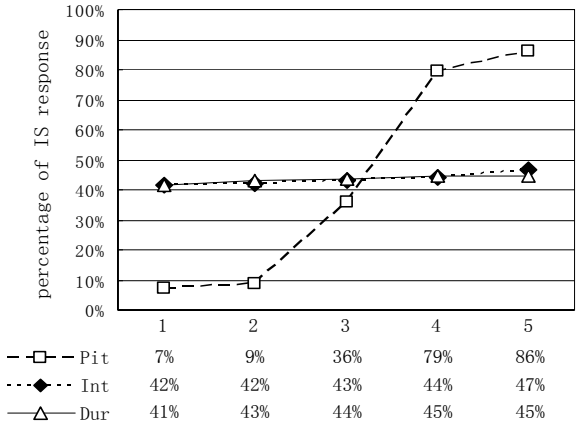

Figure 2: ISP of CE group for the 375 words as a function of a) F0; b) intensity; and c) duration manipulation

As can be seen from Figure 2, for the CE group, when F0 manipulation changed from Level One to Level Five, IS responses increased from 7% to 86%. The change in intensity from Level One to Five caused an increase of ISP from 41.57% to 46.63% and similarly changes in duration resulted in an increase of ISP from 41.37% to 44.82%. An ANOVA showed that the effect of F0 was significant in CE stress perception, F (4, 249) = 842.758, p< 0.001. Intensity, F (4, 249) = 1.788, p = 0.132 and duration. F (4, 249) = 0.866, p = 0.485, on the other hand, did not show significant effects for CE judgments of stress on the manipulated tokens. F0, alone, accounted for 91.2% (η^2 = 0.912) of the variance in ISP for CE. The effect size of intensity (η^2 = 0.028) and duration (η^2 = 0.014) were minimal.

Comparing the two groups in terms of ISP change as a result of F0 manipulation, it can be observed that the F0 manipulation resulted in an increase of 79% of ISP in CE (86%-7%) and only resulted in an increase of 43% in NE (71%-28%). Furthermore, if we compare the effect size of F0 in the two groups, we can see that F0 can explain 58.6% of the variance observed in NE group and can explain as much as 91.2% of the variance in CE group. Therefore, F0 has a stronger effect on the ISP of CE than on the ISP of NE.

3. Discussion

Based on the results presented above, the three questions raised in the introduction can be addressed. First, are Chinese learners 'deaf' to English stress? Second, if they can perceive English stress, how are their perceptions affected by the three cues. Finally, do learners rely heavily on F0 in English stress perception?

3.1. Deaf or Not Deaf

First, Chinese learners can indeed perceive the position of English stress. Different from French learners of English, Chinese learners in this study are not 'deaf' to stress in English. The results indicate that systematic changes in CEs' stress judgments can be observed for manipulated tokens. Given the appropriate cues, they can perceive the difference between stressed and unstressed syllables. When the first syllable is lower in F0, the chance of it being perceived as stressed is significantly lower than when it is higher in F0. This is consistent with previous studies on native and non-native speakers' perceptions of stress position, which have showed that non-native speakers are not very different from native speakers in their ability to correctly identify the stress position [14][15]. However, these studies also pointed out that it cannot be assumed that the two groups use the same types of information in perceiving stress, given the multiple cues associated with stress.

3.2. Primary reliance on F0

In stress perception, while NE used all three cues, not all three cues are relevant for CE stress judgment. The difference between CE and NE in stress perception lies in the weight of F0. While F0 contributes a great deal to NE stress perception, its effect in CE stress perception is decisive. The results also show that Chinese speakers rely solely on F0 information in stress judgment and they are indeed 'deaf' to duration and intensity contrasts.

Previous discussions on L1 prosodic transfer of the phonological unit, tone, fail to offer phonetic details in L2 stress perception. It is unclear why there is selective reliance on certain cues and deafness to other cues. In order to explain such differences in terms of cue reliance, the phonetic realization of tone and stress must be taken into consideration. In terms of a phonological point of view, Chinese and English belong to two different prosodic systems, with Chinese as a typical tonal language [16] and English as a typical stress language [17]. In terms of phonetic point of view, tone and stress differ from each other in their acoustic realizations. In the investigation of tonal perception in Chinese, it was found that the most important acoustic correlate for tone in Chinese is pitch. Through an experimental study, Howie [18] demonstrated that Mandarin speakers "could apparently make little use of any features other than pitch as cues for the perception of tonal distinctions." Using synthesized speech, Lin [19] also obtained similar results to Howie. F0 was proved to be the most discriminatory cue. If F0 contour was changed into a constant F0 value, any change in duration or amplitude parameters would not lead listeners to perceive different tones. On the other hand, various investigations of the acoustic correlates of stress have shown that stress is not the result of a single mechanism [10][11][12][13]. Despite the lack of agreement on the relative weight of each cue, researchers generally recognize stress as a result of a composite of factors such as F0, duration and intensity and the perception and production of stress can not be attributed to any one single cue.

Given the earlier findings, it appears that what is being transferred in L2 stress perception is not the phonological tonal unit *per se*, but the heavy reliance on phonetic F0 cue. The acoustic correlate that is most heavily relied on in an L1 prosodic structure tends to be borrowed in the acquisition of an L2 prosodic structure at the same level (lexical tone to

lexical stress). L2 learners may have a perceptual bias toward the familiar cues in L1. A similar conclusion has been drawn from a study with Vietnamese speakers' acquisition of English stress [8]. Both pitch and intensity, which were important correlates for tones in Vietnamese, are used as correlates for production and perception of English stress. In another study, Chinese speakers were found to use F0 instead of duration in signaling stress in English production [20]. Evidence can also be found in English learners of Chinese' acquisition of tone. Gandour [21] found that English speakers rely on pitch height, which is an active acoustic correlate for stress in English, rather than pitch contour in tone perception. On the other hand, Lee [22] found that learners with a tonal language background were much better in tone discrimination, even if the tones are different. In other words, it appears that the linguistic function of pitch in learners' native tonal language background can help them to perceive tones in an L2 tonal language [23].

The investigation of the acoustic details of Chinese learners' English stress perception in this study and the general agreements made by other researchers in L2 perception and production above suggest that it may be inadequate to remain restricted to the phonological level in the discussion of L1 prosodic transfer to L2 prosodic acquisition. The prosody transfer does not operate at the level of phonology, but rather through phonetic details.

3.3. Implications for future studies

In order to offer more evidence of the prosodic transfer at the phonetic level, studies with L2 speakers with other language backgrounds need to be conducted. For example, Ueyama [9] found that Japanese learners of English at the beginning level do not use duration difference in English stress production and argued that it is because durational difference is not used at the word accent level in Japanese. It would be interesting to see if Japanese speakers can make use of duration in response to the stimuli used in the present study.

4. Conclusions

In the present paper, an experiment was conducted with Chinese learners of English to study their reliance on the three acoustic correlates in stress perception. While the manipulation of F0, intensity and duration all result in considerable change in the percentage of Initial Stress Judgment by native speakers of English, only F0 manipulation has an effect on Chinese learners' stress judgments. The manipulation of intensity and duration is largely irrelevant to their perception. Given the results, it is argued that phonetic details must be taken into consideration in explaining prosodic transfer of first language (L1) to the target language.

5. References

[1] Field, J., 2005. Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly,* 39(3), 399-423.

[2] Cutler, A., Clifton, C.F., 1984. The use of prosodic information in word recognition. In *Attention and Performance x: Control of Language Processes,* H. Bouma, D.G. Bouwhuis (eds.). Hillsdale, NJ: Lawrence Erlbaum, 183-196.

[3] Peperkamp, S., Dupoux, E., 2002. A typological study of stress 'deafness'. In *Laboratory Phonology 7,* C. Gussenhoven, N.Warner (eds.). Berlin: Mouton de Gruyter, 203-240.

[4] Archibald, J., 1997. The acquisition of English stress by speakers of nonaccentual languages: Lexical storage versus computation of stress. *Linguistics,* 35, 167-181.

[5] Juffs, A., 1990. Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *International Review of Applied Linguistics, 28*(2), 99-118.

[6] Chao, Y.R., 1980. Chinese tone and English stress. In *The Melody of Language,* L.R. Waugh, C.H. VanSchooneveld (eds.). Baltimore, MD: University Park Press, 41-44

[7] Flege, J.E., 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics,* 15, 47-65.

[8] Nguyen, T.A.T., 2003. *Prosodic Transfer: The Tonal Constraints on Vietnamese Acquisition of English Stress and Rhythm.* Unpublished dissertation, University of Queensland, Queensland, Australia.

[9] Ueyama, M., *Prosodic transfer: An acoustic study of L2 English vs. L2 Japanese*. 2000, University of California, Los Angeles.

[10] Fry, D.B., 1955. Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America,* 27, 765-768.

[11] Fry, D.B., 1958. Experiments in the perception of stress. *Language and Speech,* 1, 126-152.

[12] Sluijter, A.M.C., van Heuven, V.J., 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100(4), 2471-2485.

[13] Sluijter, A.M.C., van Heuven, V.J., Pacilly, J.J.A., 1997. Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustical Society of America,* 101(1), 503-513.

[14] Davis, S.M., Kelly, M.H., 1997. Knowledge of the English Noun-Verb Stress Difference by Native and Nonnative Speakers *Journal of Memory and Language* 36(3), 445-460

[15] Adams, C.; Munro, R.R., 1978. In search of the acoustic correlates of stress: Fundamental frequency, amplitude, and duration in the connected utterances of some native and non-native speakers of English. *Phonetica* 35(3), 125-156.

[16] Gandour. J.T., 1978. The perception of tone. In *Tone: A Linguistic Survey*, V.A. Fromkin (ed.) New York: Academic Press, 41-72.

[17] Beckman, M.E., 1986. *Stress and non-stress accent*. Dordrecht, Holland: Foris.

[18] Howie, J.M., 1976. *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge: Cambridge University Press.

[19] Lin, M., 1988. Acoustic properties and perceptual cues of tone in Putonghua. *Zhongguo Yuwen*, 3, 182-193.

[20] Chen, Y., Robb, M.P., Gilbert, H.R., Lerman, J.W., 2001. A Study of Sentence Stress Production in Mandarin Speakers of American English. *Journal of acoustic society of America,* 109(4), 1681-1690.

[21] Gandour, J.T., 1983. Tone perception in Far Eastern languages. *Journal of Phonetics,* 11, 149-175.

[22] Lee, Y., Vakoch, D., Wurm, L., 1996. Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 25(5), 527-544.

[23] Jongman, A., Wang, Y., Moore, C., Sereno, J.A. (in press). Perception and production of Mandarin Chinese tones. In *Handbook of Chinese Psycholinguistics,* E. Bates, L.H. Tan, O. Tseng (eds.) Cambridge University Press.