Perception of L2Mandarin tones by L1Swedish learners at three proficiency levels

Yasuko Nagano-Madsen[1], *Xinzheng Wan*[2]

Dept. of Languages & Literatures, University of Gothenburg, Sweden

Yasuko.nagano-madsen@sprak.gu.se, xinzheng.wan@sprak.gu.se

Abstract

Cross-linguistic studies on the perception of Mandarin tones are many. However, how speakers of lexical pitch accent language such as Swedish perceive Mandarin tones is still a widely open question. In this study, the perception of Mandarin tones in mono- and di-syllabic words by Swedish learners of Chinese at three proficiency levels was examined systematically. Results showed that the major difficulties in tonal perception disappear by the time of intermediate level (18 months) except for the confusion between T2 and T3. The effect of voice quality in identifying T3 was found to override a pitch cue for mono-syllabic words but not for di-syllabic words. Confusion between a falling tone (T4) with a rising tone (T2 or T3) was common at the beginner level but not at the intermediate level. Some of these findings were predictable from the analysis of the Swedish prosodic system as well as from the results of the previous study on the perception of Japanese by Swedish learners.

Index Terms: L2 acquisition, perception, Mandarin tones, L1 Swedish, monosyllabic, disyllabic, proficiency levels,

1. Introduction

It has been known that the neural basis of first language (L1) processing differs from that of second language (L2) processing [1]. This is true even for neural process underlying perceptual identification of L1 and L2 [2]. Although studies of L2 acquisition are many, those for prosodic acquisition are still limited. As the uniqueness of prosodic acquisition in comparison with segmental acquisition is discussed previously [3][4][5], more cross-linguistic studies are needed in order to develop how L2 prosody is acquired. Much of today's research on L2 acquisition adopts the concept of 'interlanguge' postulated by Selinker as a basic principle of discipline [6]. Interlanguage is a linguistic system that is different from both the learner's L1 and the L2 and changes continuously during the acquisition. The present study also takes this approach by examining how learners' ability of L2 tonal perception as interlanguage changes depending on the level of proficiency.

This study examines the acquisition process of L2 Mandarin tonal perception by L1 Swedish learners for three proficiency levels. Since both Chinese and Swedish are relatively well studied for their prosodic features, it will enable us to examine how and when the L1 prosodic transfer takes place.

2. Previous studies

It is well known that the perception of L2 prosody is strongly influenced by the learner's L1 prosody. There are three main language groups that differ in the use of pitch for linguistic purpose: (a) 'true' tone languages such as Chinese, Thai, Vietnamese, and Yoruba; (b) lexical pitch accent languages such as Japanese, Swedish, and Norwegian; and (c) languages such as English, German, French and Inuit in which pitch does not bear lexical function. Although cross-linguistic studies of Mandarin tone perception are many, we are not aware of any systematic studies in which the perception of L2Mandarin tones was examined by L1Swedish learners.

2.1. Lexical pitch accent in Swedish

In order to examine the effect of L1 prosodic transfer, the two lexical pitch accents in Swedish are shown in Figure 1 below. Both words have F0 pattern of HLH but the two words differ in the exact timing of F0 Hs and Ls with segments.

It has been claimed that the critical manifestation of the lexical pitch accents in Swedish and Japanese is the timing of F0 gesture with segments while pitch range is only relevant for intonation [7] Therefore, it is expected that the Swedish learners have more difficulty in identifying pitch patterns that are differentiated by pitch range such as T2 and T3. See Figure 2 in which the F0 contour for T2+T2 is compared with that for T2+T3. The identification of tone in the second syllable may be difficult as it involves changes in both a pitch range and timing with the segment.
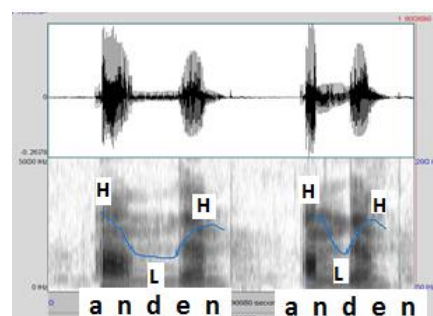

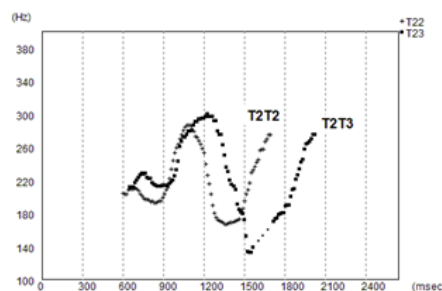Figure1: *The two types of lexical pitch accents in Swedish.*

Figure 2: *Comparison of T2+T2 with T2+T3).*

2.2. Perception of Japanese pitch accent

In the previous study of how Swedish learners perceive lexical pitch accent in Japanese, it was shown that they perform best in differentiating a high level pitch (unaccented) from a falling pitch (accented) even though Swedish does not have such a contrast [8]. On the other hand, to judge the timing of a pitch fall was found to be more difficult, particularly when it appears towards the end of a word. There was a strong tendency that the Swedish listeners identify the place of accent as one more/syllable delayed. This is attributed to the different phonetic manifestation of Swedish and Japanese accent. In Japanese, a sharp F0 fall is observed after the accented syllable, whereas in Swedish, it occurs on the accented syllable. Furthermore, listeners were confused about the location of accent as it moves towards the end of the word. It is conjectured that Swedish learners interpret the F0 fall of accent as an utterance final lowering for sentence intonation.

3. Experiment

3.1. Informant

Thirty-eight Swedish learners who are studying Chinese at three proficiency levels at a Swedish University participated in the present experiment. Level one (25 hours, 14 students), level two (230 hours, 12 students) and level three (500 hours, 12 students). Between each group, there are approximately six months intervals.

3.2. Material and procedure

The stimuli used in this experiment include two parts. The first part consists of 24 monosyllables equally designated to the four Chinese tones. In other words, each tone has 6 tokens. The second part is 16 disyllables representing all the 16 possible different combinations of the four tones in Mandarin. One female professional Chinese native speaker read the stimuli. The experiments were conducted in normal classrooms. The subjects were asked to complete an identification task. The subjects first listened to the 24 monosyllables, each of which was read twice. The subjects were provided with a response sheet that listed the four possible interpretations for each syllable and they were asked to select the interpretation that most closely matched the syllable that they heard. Secondly, the subjects listened to the 16 disyllables, each of which was also read twice. The subjects were then asked to mark the tones to the disyllables in accordance with what they heard. Both the mono- and disyllables were played in random order.

3.2.1. Monosyllabic words

Six sets of existing monosyllabic words differing in initial consonant and tone such as /mā, má, mǎ, mà/, and /lā, lá, lǎ, là/were recorded and randomized in the stimuli. A total of twenty-four tokens were obtained for each student.

The F0 configurations for /mā, má, mǎ, mà/ are shown in Figure 3. Note that the four tones differ not only in the pitch patterns but also in other features. T4 is the shortest while the duration of T3 is twice as long as that of T4. Also T4 regularly contained a creaky voice (dotted line in the figure) at a low pitch region.
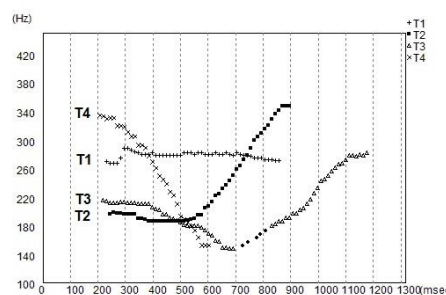

Figure 3: *F0 configurations for /mā, má, mǎ, mà/ from the stimuli.*

3.2.2. Disyllabic words

A set of sixteen disyllabic words representing all the possible tone combinations was recorded by the same speaker as above. They are all existing words having nasal, lateral, or voiced plosive as a constituent consonant. Because of the phonological process known as tone sandhi, T3 changes its realization when combined with another tone. The F0 value for T3 is usually 214 in a pitch scale when it stands alone. But the value will be changed when it is combined with other tones. The basic rules are the following: the value for the first T3 changes to 35, which is exactly the same as T2, when the disyllable is T3+T3. The value for the first T3 changes to 211, when the disyllable is T3+T1, or T3+T2, or T3+T4. A combination of two falling tones (T4+T4) as well as two rising tones (T2+T2) is slightly downstepped so that the F0 peak and valley in the second syllable is lower than that in the first syllable. A combination of two high level tones (T1+T1) is least affected by tonal environment and shows very little or no declination.

4. Results

For the perception experiment with monosyllabic stimuli, there are 24 stimuli for 38 listeners, resulting in a total of 912 tokens. As for the disyllabic stimuli, there are 16 stimuli for 38 listeners, resulting in a total 608 tokens.

4.1. Tones in monosyllabic stimuli

4.1.1. Misperceived tones (errors)

The mean percentage error scores for the four tones by three proficiency levels are shown in Figure 4. It is shown that the perception of T3 is easy and stable at all levels of proficiency while other tones are perceived differently between the three levels. The misperception rates are high for T2 and T4 but it

decreases thereafter. As for T2, a sharp decrease is observed between level 1 (beginner-6 months) and level 2 (beginner-12 months) while it is between level 2 (beginner-12 months) and level 3 (intermediate-18 months) for T4. A one-way ANOVA with Level (1,2,3) as between subject factor revealed a significant effect of Level in perceiving T1, T2, and T4 but not T3. Feedbacks from the listeners showed that the glottalization during T3 stimuli gives a strong cue for discrimination. The Figure shows how the perceptual pattern of L2 Mandarin tones by L1 Swedish learners as interlanguage shifts from the proficiency level 1 to 3. At level 3, all the tones except for T2 are perceived correctly.
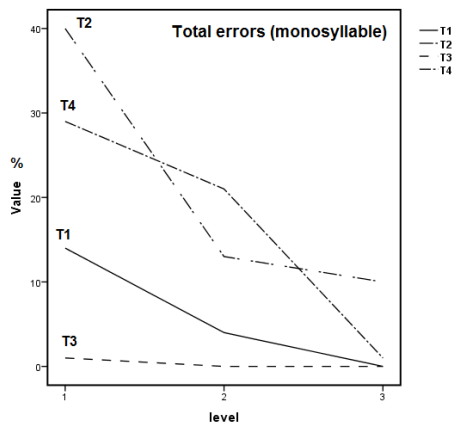

Figure 4: *The rate of tonal misperception for monosyllabic stimuli.*

4.1.2. Analysis of errors

The misperceived tones were further analyzed as to which tone it was mistaken for. The most frequent error types were T2->T3 and T4->T2. See Figure 5 below.
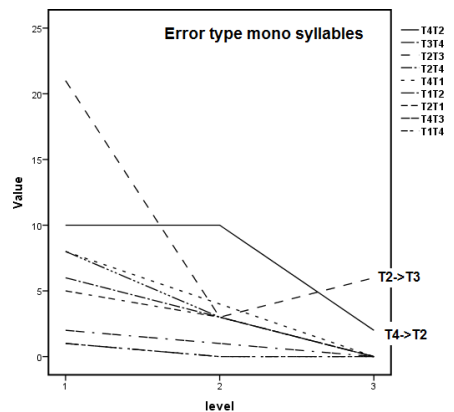

Figure 5: *The score of error types for monosyllabic stimuli.*

4.2. Tones in disyllabic stimuli

4.2.1. Misperceived tones (errors) in the first syllable

The mean percentage error scores for the four tones in the first syllable of disyllabic stimuli are presented in Figure 6 for the three proficiency levels. A one-way ANOVA with Level (1,2,3) as between subject factor revealed a significant effect of Level in perceiving all the four tones. The error pattern

observed in the first syllable of disyllabic stimuli can be characterized as T1 being easy while the other three tones are more difficult. T3, which shows a steady correct judgement for monosyllabic stimuli, is not observed in this position. Instead, T1 is the easiest throughout the three proficiency levels while T4 is difficult at level 1 but improves thereafter. In other words, the perceptual pattern in the first syllable differs from that of the mono-syllabic words.
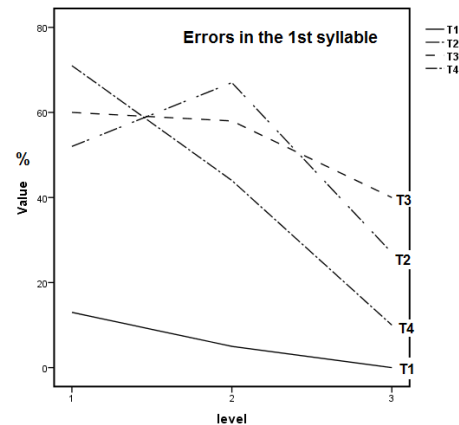

Figure 6: *Errors in the first syllable.*

4.2.2. Analysis of errors - first syllable

The analysis of misperceived tone types in the first syllable is presented in Figure 7 below. The three most typical error patters at the beginner levels are: T3>T4, T4>T2, and T2>T4. Note that these types of errors disappear completely or drastically at the proficiency level 3 (intermediate). Another error type T2>T3, on the other hand, is not a dominant error type from the beginning but persistently present even at the level 3.
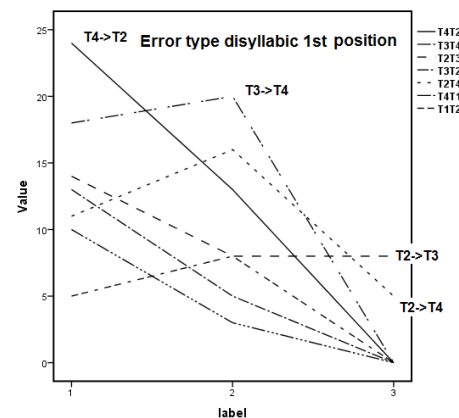

Figure 7: *Error types in the first syllable.*

4.2.3. Misperceived tones (errors) in the second syllable

The mean percentage error scores for the four tones in the second syllable of disyllabic stimuli are presented in Figure 8 for the three proficiency levels. A one-way ANOVA with Level (1,2,3) as between subject factor revealed a significant effect of Level in perceiving all the tones except for T3 and T4 at levels 1 and 2. The tonal misperception pattern for this

position is characterized as T2 being most difficult while other tones are less difficult.
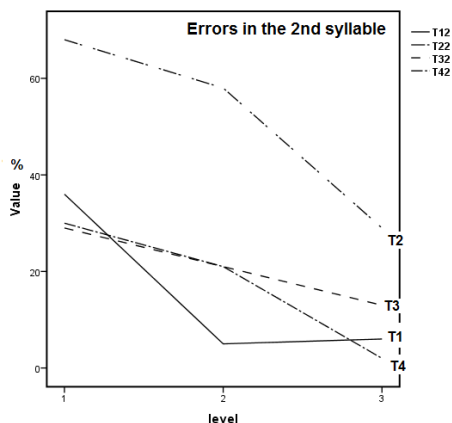

Figure 8: *Schematic diagram of speech production.*

4.2.4. *Analysis of errors - second syllable*

The results of error types for the second position are presented in Figure 9. The misperception of T2 as either T3 or T4 should be noted.
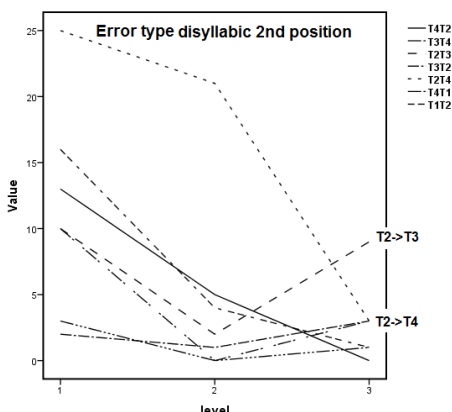

Figure 9: *The score of error types for monosyllabic stimuli.*

5. Discussion

In this section, the results of our study are discussed in relation of L1 prosodic transfer and universal. The results of our study show that the tonal perception of Mandarin tones differs greatly depending on whether the word in question is monosyllabic or disyllabic. Furthermore, for a disyllabic word, whether it is the first syllable or the second. It also reveals that the L2 acquisition process of tonal perception is not a quantitative change but a qualitative one. The most misperceived tone in the first position of disyllabic stimuli is T4 for the proficiency level 1 but not for level 2 and 3 students.

As for the monosyllabic stimuli, the order of perception (from easy to difficult) is T3>T1>T4>T2 (level 1), T3>T1>T2>T4> (level 2), and T3>T1>T4>T2 (level 3). The most consistent and easy perception for monosyllabic stimuli

is T3 for which a regular glottalization is present in voice quality. Similar results are reported for L1 German speakers and the authors conjecture glottalization as a perceptual cue [9].

In contrast to studies with monosyllabic stimuli, those with disyllabic stimuli conducted from a cross-linguistic perspective appear to be very limited. Some of the previous studies with L1 Spanish, Hungarian, and Korean are compared with the present study [10][11][12][13]. While they agree that the tonal perception of Mandarin varies depending on the position of the syllable in question, there seems to be little agreement in the results for tones in the first syllable. The results, however, were more consistent for the second syllable in which the tonal perception is (from easy to difficult) T4>T3 or T1>T2. In our study, the same order was present for the proficiency level 3 but not for levels 1 and 2. All the three previous studies as well as the present study agree that the most difficult tone to perceive in the second position is T2.

The prediction based on the previous study of Swedish learners as well as from the Swedish prosodic structure (2.2.) was that Swedish learners would perform best in differentiating the level tone (T1) from other three tones (level vs. tones with pitch change). For mono-syllabic words, this prediction was not born out because T3 with glottalization performed better. It can be hypothesized that the voice quality cue overrides the pitch cue even though it is not the feature of L1 prosody since similar result was found even for German learners. However, the matter is complicated since the identification score was better for T1 than for T3 for di-syllabic stimuli even though T3 retains (when it appears in the second syllable). The reason for poorer identification for T3 in di-syllabic stimuli is conjectured to be the contextual effect that changes the phonetic realization of T3 in the first position.

Another major error was confusion between a falling tone (T4) with a rising tone (T2 or T3) which may appear less common cross-linguistically. This can be considered as negative L1 transfer since A rising pitch is not common in the Swedish prosodic system, be that a manifestation of the lexical pitch accent or a manifestation of question intonation. This confusion, however, disappears drastically by the time the learners reach level 3 (intermediate - 18 months).

The most persistent error which was present even at level 3 was to identify T2 as T3. This error was observed regardless of the word type or position of the syllable in a word. This finding agrees with the major previous studies that conclude the most difficult tone for L2 learners to perceive is T2 (summarized in [14]), which can be hypothesized as universal. It should also be noted that T2/T3 contrast before T3, which was found to be particularly difficult for Swedish learners may also be rooted from a non-L1 feature. It should be noted, however, it has been an issue whether the T2/sandhi-T3 merger is partial or categorical is actually dependent on the dialect of Mandarin in question [15].

6. Implication to teaching, learning and technology

As we live in a globalized and multilingual society, we should be aware that many of us are speaking L2 in our everyday communication. And the importance of L2 will constantly increase in the future as more and more communities become integrated. In this section, we will discuss how the findings of our paper can contribute to the classroom teaching.

It has been known that the processing of L2 speech differs from that of L1. It has also been well known that L2 acquisition is influenced both by the speaker's L1 and his age. The present paper focuses on the perceptual acquisition of L2 Mandarin tones by Swedish university students around the age of 20.

In teaching the perception and production of Mandarin tones to the Swedish learners at the new beginner level, it is desirable to start with the difference between pitch rise and fall since many students found it difficult to discriminate the two. Students should be given a focused instruction for combination of tones that are difficult to them. In doing so, it would be desirable to demonstrate the F0 configurations to point out that the difference are mainly in the timing of F0 peak and valley as well as pitch register.

It was already shown earlier (Figure 2) how the F0 configuration of T2+T2 is compared with that for T2+T3. The two differ both in the timing of F0 peak and valley as well as the pitch range/register. In Figures 10, 11, and 12, further examples are shown to demonstrate the F0 configurations of some disyllabic words used in the present study. They are all 'easy to confuse' patterns for L1 Swedish listeners. T4T2 pattern differs from T1T2 pattern mainly in pitch rage (Figure 10) while T2T2 differs from T2T3 in the timing of F0 fall and rise (Figure 11), the latter is similar to the manifestation of Swedish word accents. The two patterns shown in Figure 10 differ both in the pitch range and timing of F0 fall and rise.

The critical manifestation of the lexical pitch accents in Swedish and Japanese is known to be the timing rather than the pitch range as described in section 2. Therefore, it is expected that the Swedish learners have more difficulty in identifying pitch patterns that are differentiated pitch range. It is interesting to note that both patterns shown in Figures 11 and 12 are similar in F0 configurations, but differ in the timing as well as in pitch register and glottalization. The confusion of T4+T2 with T1+T2 (Figure 10) may look strange in the first place, but a close examination reveals that the two combinations are similar in F0 configurations. As for T2 after T4, it is produced at lower pitch register than usual T2 and involves slight glottalization. A confusion of T2 with T3 occurs in such a position with such a tonal combination.
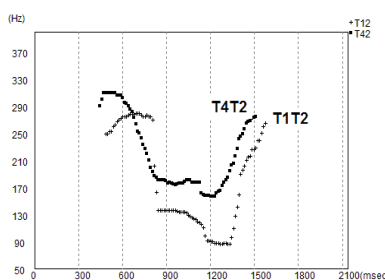

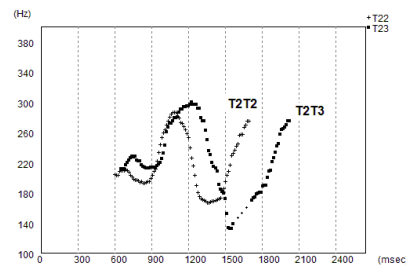Figure 10: *Comparison of T4+T2 with T1+T2*.
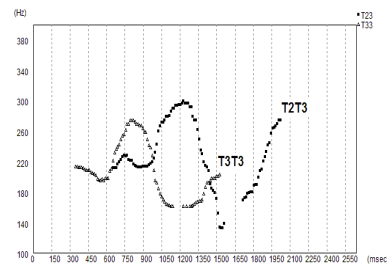

Figure 11: *Comparison of T2+T2 with T2+T3*.

Figure 12: *Comparison of T3+T3 with T2+T3*.

There has been a computer-assisted instruction system for self-teaching of discriminating Mandarin tones [16]. Such a system can be combined with more specific knowledge of learners' L1 for more efficient learning.

7. Conclusion

One of the major issues in L2 acquisition is the role of L1 transfer and the universals. How much of the perceptual errors found in the present data can be attributable to the prosodic structure of Swedish? What are the universals in L2 prosodic acquisition? The current theory of L2 acquisition predicts that the learners' errors become more and more similar to those by native speakers, i.e. errors found in the L1 acquisition by children. Does it mean that the error pattern found for the higher proficiency level becomes similar to that for L1 acquisition by Mandarin speakers? Although the current state of research on the L2 prosodic acquisition is still far from being able to answer all these questions. Despite a large body of research on L2 prosody, little attention has been paid to how the acquisition pattern, be that a perception or a production, changes as 'interlanguage' as a function of the proficiency level. Furthermore, most cross-linguistic studies have not given enough attention in describing exactly how L1 transfer takes place for prosodic acquisition. More studies from typologically diverse languages are needed in order to advance the theories of L2 prosodic acquisition as well as to apply such theories to teaching and technology.

8. References

[1] D. Perani and J. Abutalebi, "The neural basis of first and second language processing," *Cognitive neuroscience,* vol.15, no. 2, pp.202–206, 2005.

[2] D. Callan, J. A. Jonesa, A. . Callanc, and R. Akahane-Yamada "Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory–auditory/orosensory internal models," *NeuroImage,* vol. 22, no. 3, pp .1182–1194, 2004.

[3] Y. Nagano-Madsen, "Acquisition of L2 Japanese intonation -data from Swedish learners (in Japanese)," *in Japanese Speech Communication* 2, pp.1-27. Tokyo: Hitsuji Shobou, 2014.

[4] Yasuko Nagano-Madsen (2015) "Acquisition process of L2 Japanese intonation by Swedish learners – Interlanguage or prosodic transfer?" In The Scottish Consortium for ICPhS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 0820.1-4 retrieved from http://www.icphs2015.info/pdfs/Papers/ICPHS1041.pdf

[5] Y. Nagano-Madsen, "Prosodic phrasing unique to the acquisition of L2 intonation- analysis of L2 Japanese intonation by L1 Swedish learners," in *INTERSPEECH* 2015-16th Annual Conference of the International Speech Communication Association, September 8–12, Dresden, Germany, Proceedings, pp. 100–104, 2015.

[6] Selinker, "Interlanguage." *IRAL* 10, pp. 209-231, 1972.

[7] G. Bruce, *Swedish word accents in sentence perspective*, Lund University, 1977.

[8] Y. Nagano-*Madsen, "Perception of Japanese accent by Swedish learners," Proceedings Fonetik 2000*, pp.97-100, 2000.

[9] H. Ding, O. Jokisch and R. Hoffmann, "Perception and Production of Mandarin Tones by German Speakers," in *SPEECH PROSODY*, Illinois 2010, papers 100153.pdf, 2010.

[10] J. Chen and Sh. Sun "On the Perception of Tones in disyllables of Mandarin Chinese by South Korean students (in Chinese),"*Yangtze River Academic*, no.2, pp.114-121, 2012.

[11] M. Chen and J. Wang "On the Development of Chinese Disyllable Tone of CSL learners,(in Chinese)" *Journal of Yunan Normal University* (Teaching and Research on Chinese as A Foreign Language Edition), vol.8, no.4, pp.24-32, 2010.

[12] G. Wang, "On Perception of Tones in Double Syllables of Mandarin Chinese for Foreign Students of Spanish-speakers,(in Chinese)"*TCSOL Studies*, vol. 59, no.3, pp.29-40, 2015.

[13] Y. Wang, "A quantitative study of the tone-marking for disyllabic Chinese words by Hungarian Students, (in Chinese)"*Chinese teaching in the world*, vol.44, no.2, pp.90-97, 1998.

[14] B. Yang, Perception and Production of Mandarin Tones by Native Speakers and L2 Learners, Berlin, Heidelberg : Springer Berlin Heidelberg, 2015.

[15] J. Myers and J. Tsay " Investigating the phonetics of Mandarin tone sandhi" Taiwan Journal of Linguistics , pp. 29-68, 2003.

[16] Q. Sun, S. Hiki, and K. Sunaoka "A computer-assisted instruction system for self-teaching of discriminating Chinese four tones" The Journal of Modernization of Chinese Language Education, pp.52-59, 2012