

# Perception and Production of L2 Mandarin Tones by Swedish Learners

Yasuko Nagano-Madsen  
Dept. of Languages & Literatures,  
University of Gothenburg,  
Gothenburg, Sweden  
Yasuko.nagano-madsen@sprak.gu.se

Xinzheng Wan  
Dept. of Linguistics and Philology  
Uppsala University,  
Uppsala, Sweden  
Xinzheng.Wan@lingfil.uu.se

**Abstract**— This study presents the results of perception and production of L2 Mandarin tones in mono- and di-syllabic words by Swedish learners at the beginner level. Although studies of perception and production on Mandarin tones are many, those by speakers of lexical-pitch accent language such as Swedish are still very limited. The result reveals both discrepancy and agreement between perception and production. Swedish learners perform best in discriminating a level tone (T1) from contour tones (T2, T3, T4) both in perception and production. Discrepancy between perception and production was noted for T3. In perception, the identification of T3 was second best after the level tone (T1), but the production of T3 was found to be difficult.

**Keywords**— *perception; production; F0 manifestatio; L2 Mandarin tones; lexical-pitch accent*

## I. INTRODUCTION

This study examines the perception of production of Mandarin tones by Swedish learners at the acoustic level. It is well known that the neural basis of first language (L1) processing differs from that of second language (L2) processing [1]. It is hypothesized that the processing of L2 becomes similar to that of L1 as the proficiency level becomes higher. This is true even for neural processes underlying perception and production of L1 and L2 [2]. Although perception and production of L2 Mandarin tones are many, most studies use monosyllabic words and studies that examine perception and production simultaneously are still limited. Furthermore, seen from a prosodic typology, studies from the speakers of a lexical pitch-accent language such as Swedish is rare. Most European languages such as English, German, and Dutch do not have a lexical pitch-accent or a tone. However the large majority of languages in Africa and Asia are either lexical pitch-accent languages or tone languages. Swedish, Norwegian and Serbo-Croatian are the few lexical pitch-accent languages found in Europe.

The uniqueness of prosodic acquisition in comparison with segmental acquisition is discussed previously in [3], [4], [5], and the need of more cross-linguistic is urged in order to develop how L2 prosody is acquired and processed. Exactly what kind of L1 transfer takes place and at which stage? Are there universals that override such L1 interference? The present study is directed towards answering such questions. It

examines the acquisition and processing of L2 Mandarin tonal perception and production by L1 Swedish learners at the beginner level (five – seven months of learning). 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.

## II. SWEDISH PITCH ACCENT

It is known that the perception and production of L2 prosody is strongly influenced by the learner's L1 prosody. Since pitch register is only relevant for discourse function, i.e. for the manifestation of focus [6] in Swedish, 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. Fig. 1 below shows the F0 contours of accent 1 and accent 2 in Swedish (West coast dialect).

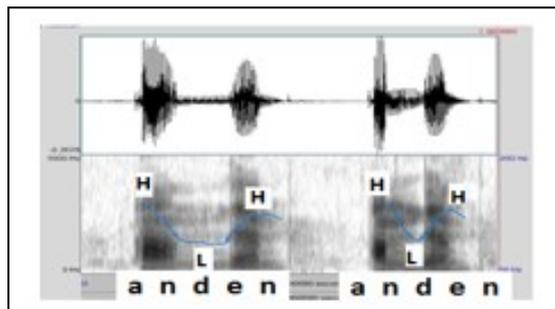


Fig. 1 Two types of pitch accents in Swedish.

## III. EXPERIMENT

### A. Informants

The informants who participated in the perception experiment was 14 Swedish students who are studying Chinese at the beginner level in a Swedish University. Of these, 9 students participated in the production experiment as well.

### B. Stimuli and procedure for perception experiment

The stimuli used in this experiment include two parts. The first part consists of 24 monosyllabic words that are equally distributed across the four Mandarin tones. The second part is

16 disyllables representing all the 16 possible different combinations of the four tones in Mandarin. One female native speaker of Chinese read the stimuli twice and the subjects were asked to differentiate the tones after listening to the stimuli. After the monosyllabic words, the subjects listened to the 16 disyllabic words, each of which was also read twice. The present paper uses part of the experiment reported earlier in [7].

C. Stimuli and procedure for the production experiment

Nine students were asked to produce the mono- and disyllabic words that differ in tones. The reading list consists of /ma/ with the four lexical tones as well as 16 disyllabic words of all the possible tone combinations. Each student repeated the monosyllabic word five times and disyllabic words twice. Altogether, 468 tokens were obtained for the production data. The students did not listen to or imitate the stimuli beforehand. The recorded data was judged by two native Chinese speakers who are also teachers of Chinese. The acoustic analysis was conducted by using PRAAT and SUGI on the PC and pitch range was exhibited in logarithmic scale.

IV. RESULTS

A. Perception

Altogether, 560 tokens were obtained from 14 listeners for the perception experiment. The results are presented in Table 1 below. As for the monosyllabic words, the correct identification score was in the order of T3 > T1 > T4 > T2. The error rates are high for T2 and T4 but it decreases thereafter. 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. Listeners' comments showed that they had relied on the presence of glottalization in discriminating T3.

TABLE I.

Tone	IDENTIFICATION OF TONES IN MONOSYLLABIC WORDS (%)			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	89	9	0	2
Tone 2	7	62	23	8
Tone 3	0	0	98	2
Tone 4	10	12	2	76

The results of the perception experiment for the first syllable in disyllabic words are presented in Table 2 below. The results differ from those of the mono-syllabic words. (cf. Table I). Here it is T1 that shows the best identification score while other tones are considerably more difficult to be identified. The misperception found for the monosyllabic words (T2 > T3) was also present in this position. Other frequent misperception were T3 > T4 and T4 > T1/2.

The results of the perception experiment for the second position in disyllabic words are presented in Table 3 below.

Here, the perception pattern is similar to that of the monosyllabic words except that T3 does not score high in this position. In this position, misperception as T4 (falling) increases.

TABLE II.

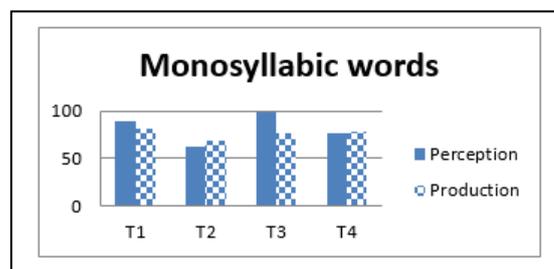
Tone	IDENTIFICATION OF TONES 1 <sup>st</sup> SYLLABLE OF DYSYLLABIC WORDS (%)			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	80	17	0	3
Tone 2	8	48	34	10
Tone 3	0	19	55	26
Tone 4	20	25	5	50

TABLE III.

Tone	IDENTIFICATION OF TONES 2 <sup>nd</sup> SYLLABLE OF DYSYLLABIC WORDS (%)			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	68	32	0	0
Tone 2	0	45	16	39
Tone 3	0	5	70	25
Tone 4	11	46	0	43

B. Production and perception

The number of the correct production is summarized and presented together with the results of the perceptual identification score in percent below. As expected, the results are better for monosyllabic words than for disyllabic words. In some categories, the production result is better than the perception, which may look unusual. This might be because of the time lag between the perception experiment and production experiment. Since the perception experiment was conducted anonymously, it was not possible to examine a one to one



correlation for each participant.

Fig.2: Correct perception and production in percent.

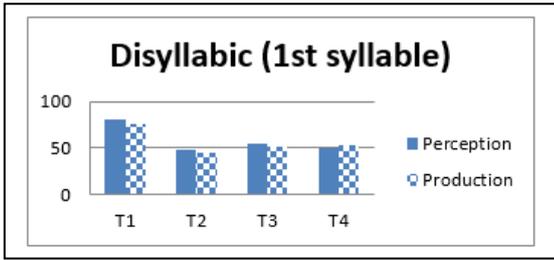


Fig.3: Correct perception and production in percent.

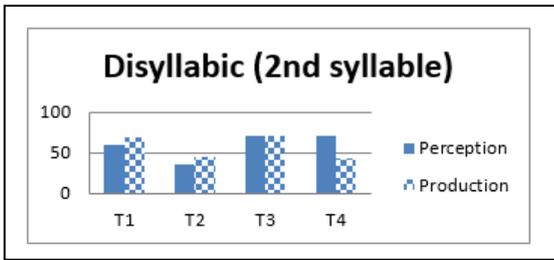


Fig.4: Correct perception and production in percent.

V. ACOUSTIC CHARACTERISTICS

A. F0 manifestation

Since the production data of disyllabic words are more complicated, only the production data of monosyllabic words are shown and discussed in this session. As for the production data of L2 Mandarin tones, pitch register and contour type are the two basic parameters. Fig. 5 shows the F0 configurations of the Mandarin tones produced by a native speaker. Figs. 6 to 14 show the corresponding F0 configurations by 9 Swedish students.

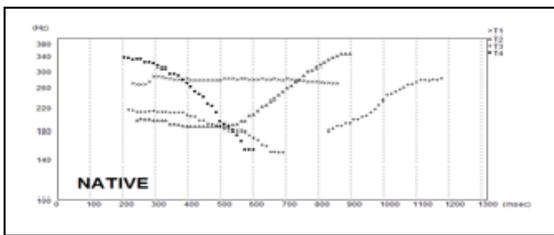


Fig. 5. F0 contours for Mandarin tones by a native speaker.

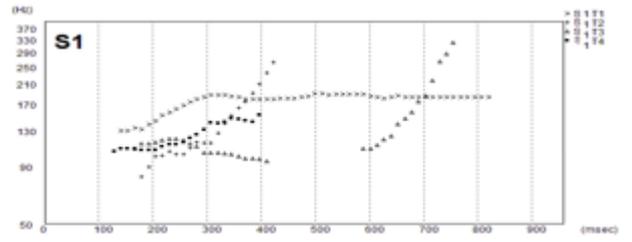


Fig.6.

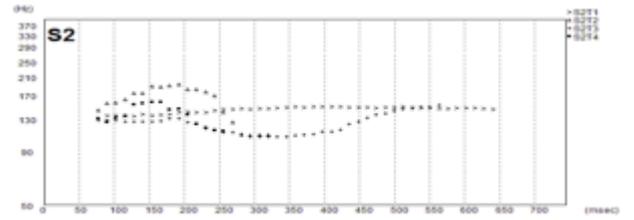


Fig.7.

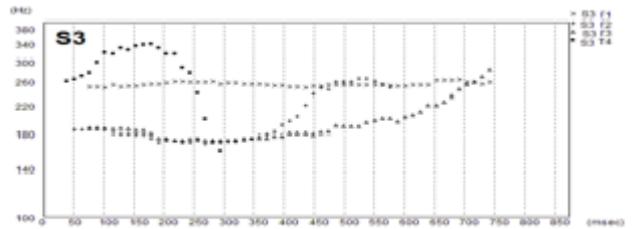


Fig.8.

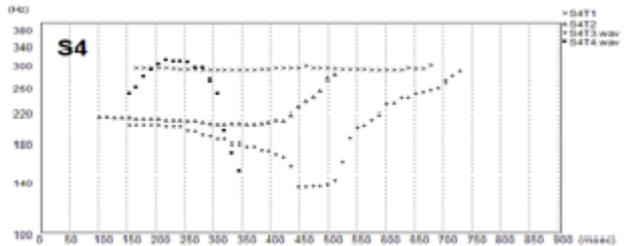


Fig.9.

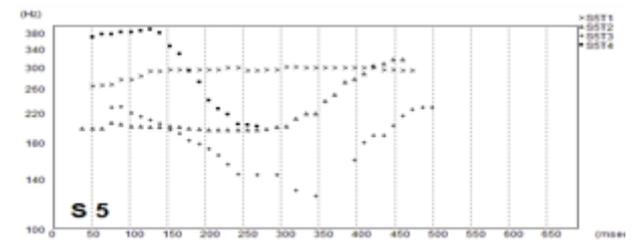


Fig.10.

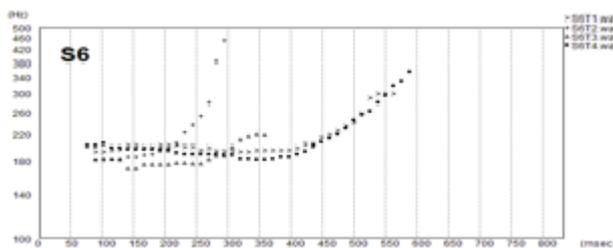


Fig.11.

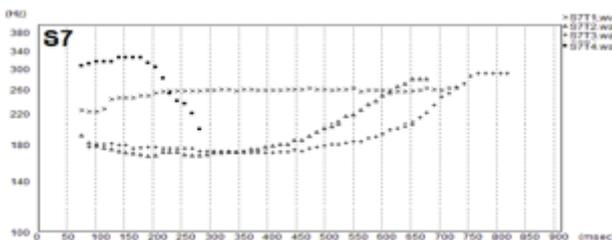


Fig.12.

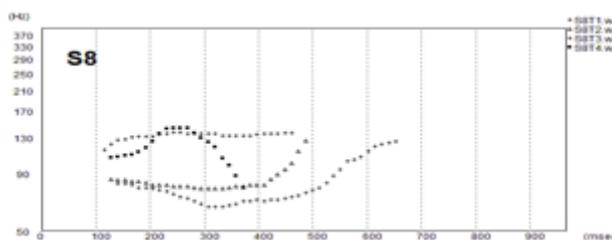


Fig.13.



Fig.14.

### B. Summary of the F0 characteristics

The tonal spaces for the 9 students were compared and the following points are noted.

(1) All except one student (S6) differentiated the level tone (Tone 1) from other contour tones. Student 6 used a rising F0 contour for all the four tones, some differing in duration. The contrast between a level tone vs. contour tones can be hypothesized as a first step of Mandarin tone acquisition by Swedish learners.

(2) Two students (S2 and S9) differentiated a level tone from contour tones with extremely narrow pitch range. In their production, contour tones are not correctly differentiated.

(3) Other six students (S1, S3, S4, S5, S7, S8) differentiated all the four tones correctly with slightly different ways with regard to the use of pitch register, duration, and glottalization.

(4) There are at least three different ways of manifestation in differentiating T2 and T3. The first is to utilize both pitch register/glottalization and timing of the F0 rise like a native speaker (S4, S8). The second is to use only the durational difference (S1, S3, S7), which is a strategy reported for English learners [8]. The third is to use only the difference in pitch register (S5). Although glottalization gives a powerful cue in perception, its use is limited in L2 production since only two speakers (S4, S5) used it.

## VI. DISCUSSION

Swedish learners perform best in discriminating high level pitch (T1) from contour tones both in perception and production despite that Swedish prosodic phonology does not have such a contrast. A similar result was reported even for the perception of L2 Japanese (lexical) pitch accents by Swedish learners. Swedish learners of Japanese perform best in differentiating a high level pitch (unaccented) from a falling pitch (accented) [9]. Hao's experiment with English speakers also report that the most accurately perceived and produced tones were a level tone (T1) and a falling tone (T4) [10]. Assuming that the processing of L2 prosody involves both universal and language specific features, the discriminating ability of (high) level F0 from a contour F0 in both perception and production, can be hypothesized as one of the universals. However, the present study did not agree with Hao's results for English speakers with regard to a falling tone (T4).

Tone 3 shows a discrepancy between perception and production. It has received the highest rate of correctness in the perception task because of the glottalization indicating non-tonal cue plays an important role in tone identification. However, in production, the use of glottalization was found to be difficult since only two out of nine students used it. Others have differentiated T3 from T2 by other manifestation strategies such as different duration or different pitch register.

As expected, the dimension of pitch register was found to be difficult both in perception and production for Swedish learners at the beginner level. This point can be attributed to the L1 influence since pitch register is not part of the lexical contrast in the Swedish prosodic structure.

The results of the present study from Swedish learners are compared with some other studies with different linguistic background as cited in [11], [12]. 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.

Cross-linguistic comparison of perception and production of L2 Mandarin tones shows diverse results. Many of these results are not comparable straightforwardly since the material, method, and procedure are often different. Further cross-linguistic studies that examine perception and production of Mandarin tones simultaneously are desirable. Studies with disyllabic stimuli are also needed.

## REFERENCES

- [1] D. Perani and J. Abutalebi, "The neural basis of first and second language processing," *Cognitive neuroscience*, vol.15, no. 2, 2005, pp.202–206.
- [2] D. Callan, J. A. Jones, A. Callan, 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, 2004, pp .1182–1194.
- [3] Y. Nagano-Madsen, "Acquisition of L2 Japanese intonation -data from Swedish learners (=in Japanese)," in *Japanese Speech Communication 2*, Tokyo: Hitsuji Shobou, 2014, pp.1-27.
- [4] Yasuko Nagano-Madsen, "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> , 2015.
- [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, 2015, pp. 100–104, 2015.
- [6] G. Bruce, *Swedish word accents in sentence perspective*, Lund University, 1977.
- [7] Y. Nagano-Madsen and X. Wang, "Perception of L2 Mandarin Tones by Swedish Learners at Three Proficiency Levels," in *Proceedings of the International Workshop on Language Teaching, Learning, and Technology*, San Francisco, USA, 2016, pp23-28
- [8] Y-H. S. Chang, Distinction between Mandarin Tones 2 and 3 for L1 and L2 Listeners, on Z. Jing-Schmidt (ed.) *Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23)*, Vol, 2011, pp.84-96
- [9] Y. Nagano-Madsen, "Perception of Japanese accent by Swedish learners," *Proceedings Fonetik 2000,2000*, pp.97-100.
- [10] Hao, Y-C. "Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers", *Journal of Phonetics* 40, 2012, pp269-279.
- [11] 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.
- [12] B. Yang, *Perception and Production of Mandarin Tones by Native Speakers and L2 Learners*, Berlin, Heidelberg : Springer Berlin Heidelberg, 2015.