Perception of Lexical Tones by Swedish Learners of Mandarin

Man Gao
Dalarna University
mao@du.se

Abstract

Models of cross-language perception suggest that listeners' native language plays a significant role in perceiving another language, and propose that listeners assimilate non-native speech sounds to similar sounds in their native language. In this study, the effect of native language on the perception of Mandarin tones by Swedish learners is examined. Swedish learners participated in an identification task, and their performance was analyzed in terms of accuracy percentages and error patterns. The ranking of difficulty level among the four lexical tones by Swedish listeners differs from that found among English native listeners in previous studies. The error patterns also reveal that Swedish listeners confuse Tone 1 and 2, Tone 3 and 4, and Tone 2 and 4, the first two pairs rarely being confused by English listeners. These findings may be explained with the assimilation account: Swedish learners assimilate Tone 3 and 4 to Swedish pitch accents, thus they exhibit a unique pattern when perceiving the tones in Mandarin.

1 Introduction

Recent years have witnessed a rapid growth of Chinese as a foreign language in Sweden, with an increasing number of high school students choosing to study Chinese. Two major hurdles for Swedish students wishing to learn Mandarin Chinese are the tonal system and the orthographic system, i.e. the Chinese characters. Previous studies suggest that it is difficult for learners who are not from a tone language background to acquire tones (Kiriloff 1969, Shen 1989). However, Swedish, along with a few other languages such as Japanese, have what are known as pitch accents. Pitch differences are used in more restrictive ways to contrast meaning among certain sets of words, as between *anden* 'the duck' and *anden* 'the spirit' in Swedish. Pitch accent languages are therefore often treated as being typologically intermediate between tone languages (e.g. Chinese) and non-tonal languages (e.g. English), at times even described loosely as having "another type of tone system" (McGregor 2015:346). The question whether Swedish pitch accents exert any influence on learning Mandarin tones is thus significant both from a theoretical and practical perspective (for Swedish teachers and learners). Therefore a pilot experiment was conducted to examine how Swedish learners perceive Mandarin tones and the possible influence of Swedish pitch accents.

Cross-language perception research tends to be somewhat complicated, however. There are typically multiple related factors that come into play during the process of perceiving non-native speech, and many of these additionally often interact to a great extent (Jenkins & Yeni-Komshian, 1995). For example, the listener's age, experience with the non-native target language, amount of exposure to the target language, and the degree of similarity with the native language, among many other factors, may affect how sound contrast of a non-native language is perceived (Best, 1995; Flege, 1995; Yamada, 1995). Influence of the native language on the perception of a foreign language has consistently proven to be significant across a wide range of studies (Wenk, 1986; Odlin 1989; Jenkins & Yeni-Komshian, 1995). Among the theoretical models in this field is the Perceptual Assimilation Model (PAM) (Best, 1995). In brief, it focuses on the per-

ception of non-native speech contrasts, and hypothesizes that listeners tend to assimilate non-native speech segments to the most similar ones among their native phonetic categories. Another influential model is the Speech Learning Model (SLM) (Flege, 1995). It focuses on the learning of second languages (thus for both perception and production), and one of its suggestions is that learners relate perceptually the sounds in a second language (L2) to the most similar sounds in their first language (L1). A deeper comparison between the models is beyond the present scope. However, while PAM and SLM differ in many aspects, they make similar hypothesis in certain regards. In non-technical terms, if two languages have certain speech sounds that are highly similar, the listener will assimilate the non-native sound to the native sound category in perception. In other words, learners will make reference to the native sounds when interpreting the corresponding non-native ones if these are similar enough. This hypothesis has been tested on segmental categories across a multitude of languages (Best et al., 1988; Best & Strange, 1992; Flege, 1988; 1991; 1993; Guion et al., 2000; Polka, 1992; etc.). Thus far, however, a few studies have put them to test on prosodic categories (Hao, 2014; Reid, et al., 2015; Alexander and Wang, 2016).

A considerable amount of research has been devoted to the perception of Mandarin tones by non-native speakers. To begin with, results suggest that discriminating and identifying Mandarin tones may generally be regarded as fairly challenging for listeners from a non-tonal language background (e.g. Kiriloff, 1969; Broselow, Hurtig, Ringen, 1987; Shen, 1989; Chen, 1997; Wang et al., 1999). However, the relative degree of perceptual difficulty appears to vary across the four Mandarin tones. Several studies on speakers from a non-tonal language background have found that Tone 4 tends to be the easiest among the four to perceive correctly, and Tone 2 and Tone 3 often considerably more difficult (Kiriloff, 1969; Broselow, Hurtig & Ringen, 1987; Hao 2012). English speakers tend mainly to confuse the Tone 2-Tone 3 pair, as well as the Tone 1-Tone 4 pair (Kiriloff, 1969; Chen, 1997; So & Best, 2010; Hao, 2012). As mentioned previously, research has commonly found learners with a tonal language background to out-perform those with a non-tonal language background in various perception tasks (Lee et al., 1996; Liang & van Heuven, 2007). Unfortunately, so far few studies have examined the performance of pitch accent language speakers on the perception of Mandarin. So and Best (2010) investigated the perception of Mandarin tones by naive listeners (having had no previous training in Mandarin) from three language backgrounds: Hong Kong Cantonese (tonal), Japanese (pitch accent) and Canadian English (non-tonal). They found that listeners with the tone language and pitch accent language backgrounds (Cantonese, Japanese) outperformed those with a non-tonal background (English). However, the predicted assimilation between certain Japanese pitch accents and Mandarin tones sharing pitch contours was not found in this study. The authors' explanation is that such mapping of similar pitch patterns across the two languages has not yet been established owing to the limited previous exposure to Mandarin tones. The current paper will report a small-scale study aiming to examine whether Swedish learners assimilate prosodic categories with similar pitch contours in Mandarin and Swedish in perception tasks.

2 Prosodic categories in Mandarin and Swedish

2.1 Mandarin tones

Mandarin Chinese has four lexical tones and an additional 'neutral' tone. Each lexical tone is carried by a monosyllable and is used to contrast lexical meaning. Table 1 below summarizes the tone system in Mandarin. Conventionally, the four tones are named Tone 1, 2, 3 and 4 respectively. Alternatively, they may be referred to by their descriptive labels, corresponding to the overall shape of their pitch contour. Tone 1 (high level) is the only level (static) tone in Mandarin; the rest are contour (dynamic) tones. The neutral tone (Table 1, bottom row), sometimes referred to as 'Tone 5', only occurs in unstressed short syllables. It does not have a stable pitch height or contour but is dependent on its tonal environment.

Tone	Example	Pinyin	Translation	Descriptive name
Tone 1	妈	mā	'mother'	(high) level
Tone 2	麻	má	'hemp'	rising
Tone 3	马	mǎ	'horse'	low/low falling-rising
Tone 4	骂	mà	'scold'	falling
Tone 5	吗	ma	'question particle'	n.a.

Table 1: Lexical tones in Mandarin (Chinese)

Figure 1 displays the pitch contours of the four tones over time, extracted from four Mandarin female speakers' production of the syllable *ma* in isolation (Burnham et al., 2015:1461). The four tones span over the pitch range from low to high. Tone 1 is a level tone that stays in the high pitch range. Tone 2 is a rising tone that rises from the mid pitch range to the high pitch target. Tone 3 is a low falling tone that reaches its low target and rises slightly at the end. Tone 4 is a falling tone that first rises to the high pitch range and then drops dramatically toward the pitch target in the lowest pitch range.

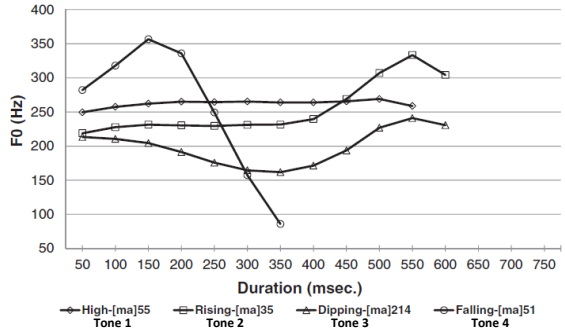

Figure 1: Pitch contours of four Mandarin tones (Burnham et al., 2015:1461).

A system based on the Scale of Five Pitch Levels (Chao 1968) is often used to represent the Mandarin tones. The pitch range in divided into five levels of relative pitch height (from 1 to 5, low to high). Tone 1 is represented as 55, which means that it starts at the highest level (5) and ends there (5). Tone 2 is characterized as 35, Tone 3 as 214 and Tone 4 as 51. An alternative characterization is offered within the framework of Autosegmental Phonology (Goldsmith, 1976). The phonological representations for Mandarin tones contain two parts: register and Tone (Yip, 1980, 1989). In simplified terms, register in this system refers to the pitch range where a tone is realized ([-upper] for Tone 3 and [+upper] for the other three tones), and Tone specifies the direction of pitch change: H or L ([+raised] or [-raised]). Tone 1 is then represented as H, Tone 2 is LH, Tone 3 is L and Tone 4 is HL.

2.2 Swedish pitch accents

Along with Norwegian and Japanese, Swedish is labelled as a pitch accent language, or alternately a word accent language. Swedish has two pitch accents: Accent 1 ('acute') and Accent 2 ('grave'). Like lexical tones in Mandarin, they may contrast the meaning between words containing the same segmental string. Unlike lexical tones in Mandarin, however, the Swedish accents do not contrast monosyllabic words in Swedish, and Accent 2 is only seen in words with more than one syllable (Elert 1981). Various analyses argue that the distribution of the two accents may be accounted for with phonological and morphological rules (Bruce 1977, Gårding 1977, Riad 1996). There are only about 350 (Elert 1971) to 500 minimal pairs (Clark & Yallop, 1990) relying on the pitch accent contrast.

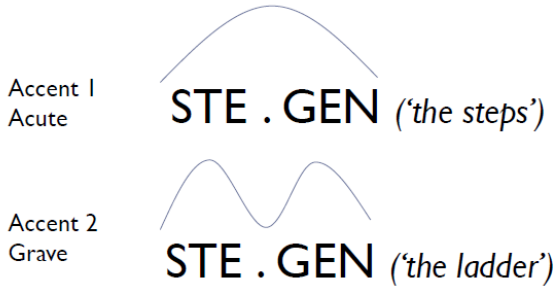

Figure 2: Schematic representation of pitch accents in Central Standard Swedish (Adopted from Engstrand 1997:62).

Figure 2 illustrates a schematic representation of the two accents in Central Standard Swedish. Accent 1 is described as a 'single falling' tone that has only one peak. Accent 2 has two peaks, reflecting primary and secondary stress respectively (Malmberg 1963).

Figure 3 displays the pitch contours of the two Swedish pitch accents over time, extracted from three female Swedish speakers' productions of disyllabic words (Burnham et al., 2015:1462). The

single falling pitch contour of Accent 1 is most easily seen when the entire word is considered. For the disyllabic word carrying Accent 2, the two separate peaks are discernible by a falling contour on each of the syllables.

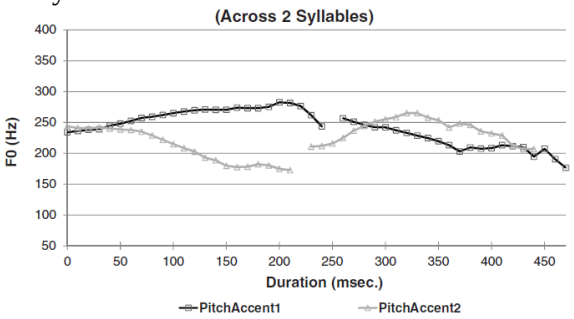

Figure 3: Pitch contours of two Swedish pitch accents (Burnham et al., 2015:1462).

Bruce (1986) proposed a phonological representation for the two Swedish pitch accents within the framework of Autosegmental Phonology (Goldsmith, 1976). He used a star notation (*) to represent the association between tone and the stressed syllable. Accent 1 is represented as HL* and Accent 2 as H*L. Bailey (1988) adopted a similar representation: HL is used to represent the single falling accent (Accent 1), and HLHL for the double peaks of Accent 2. However, he proposed that the underlying representation of the two pitch accents is the same: HL.

2.3 A comparison

Having touched upon the basic phonetic features and the phonological treatment of the prosodic systems in Mandarin and Swedish, a brief comparison between Chinese tones and Swedish pitch accents will be offered. To begin with, both languages use tonal (pitch) variations to contrast meaning, although only a subset of Swedish words rely on such contrast. Secondly, the falling pitch contour is found in both languages, as evidenced by pitch contours extracted from empirical data and phonological representations. In Swedish, disyllabic words carrying Accent 1 have the falling contour (mainly) on the second syllable, whereas disyllabic words carrying Accent 2 display two consecutive falling contours, one on each syllable. In Mandarin Chinese, the falling contour is seen in (single) syllables carrying Tone 3 or Tone 4. Thirdly, Tone 3 and 4 differ from Swedish pitch accents in manner. For Tone 3, the falling contour is followed by a slight rise at the end, when pronounced in isolation. Tone 4 displays a very sharp fall from the highest to the lowest pitch level. The falling contour is not quite as dramatic for Tone 3, in line with its labelling as '214' in Chao's (1968) Scale of Five Pitch Levels and as 'L' in Yip's framework (1980, 1989). However, empirical data of the Swedish pitch accents (Burnham et al., 2015) clearly reveals that neither accent is associated with a dramatically falling contour when compared to the Chinese tones. Finally, Mandarin Tone 1 is a level (static) tone, a type not found in Swedish. Tone 2 has a rising tone, displaying a pitch rise throughout the carrier syllable, thus also not resembling any of the Swedish pitch accents in terms of associated pitch contour.

3 Method

16 high school students (10 Male and 6 Female) who have studied Mandarin Chinese as modern language for 3 to 4 terms participated in an identification task. They were recruited from two high schools located in Jönköping and Västra Götaland Counties respectively. Twelve (10M, 2F) students are from the school in Västra Götaland and study Chinese for 120 minutes per week; four are from the school in Jönköping and study Chinese for 180 minutes weekly. All students are native speakers of Swedish[1], and except Mandarin Chinese they all have knowledge of one or two non-tonal languages as foreign or second language (e.g. English, German). All of them were very used to reading and writing Pinyin (the Chinese phonetic alphabet).

The listening material included a total of 40 tokens of 10 different syllables: *ba, pao, fa, ge, mo, pi, tan, wan, ya* and *yi*. These syllables were chosen because they can carry all four lexical tones in Mandarin. They furthermore consist of consonants and vowels which are commonly found in other languages and were considered less challenging for the

[1] Though these students may speak a form of Swedish that is different from the Swedish presented in Figure 3, it has been reported in the literature (Gårding, 1977; Riad, 1996) that Swedish dialects vary mainly in terms of the timing pattern between the peak and the segmental string for both accents, and the number of peaks for Accent 2. As to the shape of corresponding pitch contours, the falling contour is seen in nearly all dialects for both Accent 1 and Accent 2 (Gårding, 1977).

students, who would thus be in a better position to focus only on the perception of tones. The tokens were presented to the Swedish learners in random order. Five monosyllables were added before these to act as fillers and warm-up items; responses to these five items were excluded from the analysis.

A female native Beijing Mandarin speaker was recruited to produce the speech material for the perception task. She was instructed to read a list of monosyllables in random order, and was recorded in an anechoic chamber. Acoustic data were collected at a sampling rate of 16 kHz with a Brüel & Kjær microphone. The distance between the speaker's mouth and the microphone was adjusted for optimal output (about 30 cm). The speech output was captured on the hard drive. The recorded speech material was subsequently reorganized using Praat (Boersma and Weenink, 2005) and presented to the Swedish learners in the form of wav format. In the perception experiment, every Swedish learner listened to the prepared listening material on a laptop using a headset and they were asked to complete a four-alternative identification task; after hearing each monosyllable, they had to select the corresponding tone on a response sheet. Participants were given six seconds to respond before being presented with the next token. The decision time for a similar task was reported to be less than four seconds (So and Best, 2010).

4 Results

The students' performance in the perception task is summarized in figure 4. Accuracy percentages were calculated separately for the two schools. Students from the school in Jönköping overall performed significantly better than those from Västra Götaland, with the former scoring between 65% and 95% and the latter 37.5% to 80%. Significantly, however, both groups presented the same pattern in terms of the relative level of difficult among the four Mandarin tones. For that reason the results for both groups will be pooled in the following analysis.

The students' accuracy rates indicate that Tone 3 is the easiest to identify, followed by Tone 4, then Tone 1, and, finally, with Tone 2 being the most difficult. Paired samples t-tests show that only the accuracy rates for Tone 3 and Tone 4 are not significantly different (p=0.315); all the other combinations are different at the level of p=0.020 (for Tone 1 and 3 pair, and Tone 1 and 4 pair) or p=0.000.

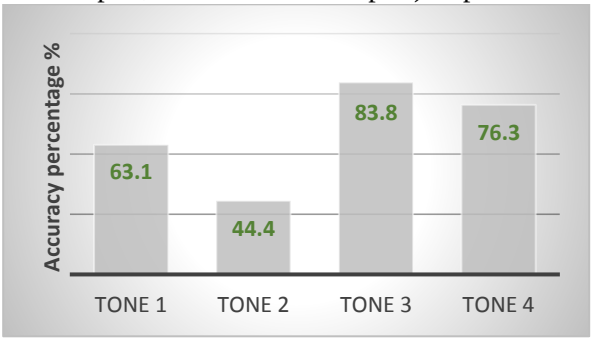

Figure 4: Accuracy percentage of identification task.

The error patterns for the identification task were also analyzed and compared. Table 2 displays the error matrix, in which the top row represents listeners' answers and the first column corresponds to the actual tone. Tone 1 is most likely to be misperceived as Tone 2 and vice versa. Tone 3 is most likely to be misperceived as Tone 4, but not vice versa; rather Tone 4 tokens, like Tone 1, are most likely to be misidentified as Tone 2. Some listeners were unable to identify a small number of tokens, most of which being Tone 2 syllables.

Response / Target	Tone 1	Tone 2	Tone 3	Tone 4	unable to identify
Tone 1	**63.1%**	27.5%	0.6%	8.1%	0.6%
Tone 2	23.8%	**44.4%**	13.1%	16.3%	2.5%
Tone 3	3.8%	4.4%	**83.8%**	7.5%	0.6%
Tone 4	3.8%	19.4%	0.6%	**76.3%**	0

Table 2: Lexical tones in Mandarin (Chinese)

In sum, Swedish students mainly display confusion among the following tone pairs: Tone 1 and 2, Tone 3 and 4, and Tone 2 and 4.

5 Discussion

Results of the identification task show that the relatively level of difficulty among the four lexical tones for Swedish learners is: Tone 2, Tone 1, Tone 4 and Tone 3, from most to least difficult. In comparison, for English learners Tone 3 and Tone 2 are more difficult to perceive than Tone 1 and Tone 4,

according to previous results[2] (Kiriloff, 1969; Broselow, Hurtig & Ringen, 1987; Hao 2012, 2014). The assimilation hypothesis from both PAM (Best, 1995) and SLM (Flege, 1995) may be used to account for this finding. Comparison of the Mandarin and Swedish prosody systems in section 2 above suggests that Tone 3 and 4 both exhibit falling contours that resemble Swedish Accent 1 and 2. It is thus possible that Swedish learners assimilate Tone 3 and Tone 4 in Mandarin to Swedish Accent 1 and 2. Although Tone 3 receives the highest accuracy rate among the four tones, it is not significantly higher than the accuracy rate for Tone 4. The accuracy rates for the other two tones, Tone 1 and Tone 2, on the other hand, are found to be significantly lower than Tone 3 and Tone 4. Following the same line of reasoning, this may be because they cannot map them onto any prosodic categories in Swedish. This is consistent with findings from Hao's (2014) study: native English learners considered Tone 3 the least English-like lexical tone through an English-likeness rating task, and also perceived Tone 3 with the lowest accuracy rate in the identification task.

Analysis of Swedish learners' error patterns lends additional support to the assimilation account. The most commonly reported confusions for native English speakers, namely those among Tone 2 and 3 and between Tone 1 and 4 (Kiriloff, 1969; Chen, 1997; So & Best, 2010; Hao, 2012), were not found for the Swedish students. Since Swedish as a pitch accent language differs from English as a non-tonal language, it is maybe the case that Swedish influences the perception of Mandarin tones in a different manner. For Swedish speakers, Tone 3 and Tone 4 are more similar to the pitch accents and to each other, and thus not difficult to differentiate from Tone 1 and Tone 2 (this is especially true to Tone 3). The error patterns found in this study reveal that Tone 1 and Tone 2 pair is most problematic for Swedish learners. Tone 1 is nearly exclusively misperceived as Tone 2, and majority of misperceived Tone 2 tokens were labelled as Tone 1. However, there is certain proportion of Tone 2 tokens misperceived as Tone 3 or Tone 4. In combined with the accuracy rates reported in Figure 4, we may conclude here that the rising tone (Tone 2) is most challenging for Swedish learners. Another source of confusion for Swedes is the Tone 2 and Tone 4 pair, which is rarely found among native English speakers in the literature (Kiriloff, 1969; Chen, 1997; So & Best, 2010; Hao, 2012). Maybe this is because Tone 2 and Tone 4 share some similarities in terms of pitch type (both are contour tones) and pitch height (Tone 2 and part of Tone 4 are active in the mid to high pitch range), which confuse Swedish listeners. But in order to verify this explanation, further research that examines Swedish speakers' strategy (i.e. perceptual cues) when perceiving tones is recommended.

The two groups of learners, both of whom had studied Mandarin for 3 or 4 terms prior to the investigation, performed quite differently in this study. Several factors may be contributing to this. One concerns the amount of exposure to the target language (Mandarin). The Jönköping students studied approximately 60 minutes more each week than those students from Västra Götaland whom they outperformed. Second, brief and informal interviews with the two instructors indicate that they may have quite different teaching style. The teacher from Jönköping stated that she tried to speak as much Mandarin as possible in class, and put a lot of emphasis on improving students' spoken proficiency. The second teacher seemed to place a lot of emphasis on vocabulary and grammar, and to be speaking mostly Swedish to his students. Therefore, the differences between the two school students in the perception task may stem from their different proficiency level in Mandarin, especially regarding spoken proficiency

6 Conclusion

The current study is the first attempt to investigate whether Swedish learners assimilate Mandarin tones to Swedish pitch accents in perception. It first provided a brief overview of the prosody systems in Mandarin Chinese and Swedish. Tone 3 and 4 in Mandarin along with Accent 1 and 2 in Swedish

[2] A direct comparison between the Swedish learners' and English learners' performance cannot be made in this study for two reasons. One is that the raw data from previous studies is not available, thus it is impossible to apply any statistical test to verify any observed differences. Secondly, though the current study employed similar experimental design as the previous research, other factors may differ, such as participants' Chinese proficiency level and the speech material used in the listening tasks, etc. It is highly recommend to conduct a future study that compares the error patterns from matched groups of Swedish and English speakers using the same listening materials.

have a falling contour; they also receive similar phonological representation. According to the two cross-language perception models PAM and SLM, the non-native listeners are expected to assimilate the non-native sound to the closest native sound category in perception. Two groups of Swedish learners participated into a Mandarin tone identification task. The results showed that they could identify Tone 3 and 4 with a higher degree of accuracy than the other two tones, which is consistent with the assimilation hypotheses in both PAM and SLM models. Furthermore, analysis of the error patterns provided additional support for the assimilations hypothesis. This study therefore constitutes an attempt to extend the PAM and SLM models to prosodic categories, and also revealed possible clues regarding the effect of Swedish pitch accents on learners' perception of Mandarin tones. Further research is clearly needed, however, especially into (1) assessing the perceptual similarity between Mandarin tones and Swedish pitch accents; and (2) analyzing Swedish learners' production of Mandarin tones for a complete understanding of the acquisition of Mandarin tones by Swedish learners.

Acknowledgments

I would like to thank the Swedish students who participated in this study and their teachers who helped me to collect the data.

References

Jennifer Alexander and Yue Wang. 2016. Cross-language Lexical-tone Identification. In *Proceedings of the 5th International Symposium on Tonal Aspects of Languages (TAL 2016),* 28-32. New York.

Leslie M. Bailey. 1988. A Non-linear analysis of Pitch Accent in Swedish. *Linguist*, 75:103-204.

Catherine T. Best. 1995. A direct realist view of cross-language speech perception. In Winifred Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press. pp.171–204.

Catherine T. Best, Gerald W. McRoberts, and Nomathemba M. Sithole. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 4:45-60.

Catherine T. Best and Winifreed Strange. 1992. Effects of phonological and phonetics factors on cross language speech perception on approximants. *Journal of phonetics*, 20:305-330.

Paul Boersma and David Weenink. 2005. Praat: doing phonetics by computer (Version 5.1.15). Retrieved from http://www.praat.org/.

Ellen Broselow, Richard R. Hurtig, and Catherine Ringen. 1987. The perception of second language prosody. In G. Ioup and S.H. Weinberger (eds.), *Interlanguage Phonology: The Acquisition of Second Language Sound System. Cambridge*, Newbury House Publishers. pp. 350-361.

Gösta Bruce. 1977, *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

Gösta Bruce. 1986. How floating is focal accent? In K. Gregersen and H. Bosbøll (eds.), *Nordic prosody IV*. Odense University Press. pp. 41–49.

Denis Burnham, Benjawan Kasisopa, Amanda Reid, Sudaporn Luksaneeyanawin, Francisco Lacerda, Virginia Attina, Nan Xu Rattanasone, Iris-Corinna Schwarz and Diane Webster. 2015. Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholingustics*, 366: 1450-1491.

Yuen Ren Chao. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Qinghai Chen. 1997. Toward a sequential approach for tonal error analysis. *Journal of the Chinese Language Teachers Association*, 32:21–39.

John Clark and Colin Yallop. 1990. *An introduction to phonetics and Phonology*. Oxford: Basil Blackwell.

Claes-Christian Elert. 1971. *Tonality in Swedish: Rules and a List of Minimal Pairs*. Umeå: Umeå University Department of Phonetics.

Claes-Christian Elert. 1981. Ljud och ord i svenskan 2. Acta Universitatis Umensis, *Umeå Studies in the Humanities*, vol. 40. Stockholm: Almqvist and Wiksell International.

Olle Engstrand. 1997. Phonetic interpretation of the word accent contrast in Swedish: evidence from spontaneous speech. *Phonetica*, 54:61–75.

James Emil Flege. 1988. Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, 84:70-79.

James Emil Flege. 1991. The interlingual identification of Spanish and English vowels: orthographic evidence. *The Quarterly Journal of Experimental Psychology*, 43A(3):701-731.

James Emil Flege. 1993. Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, 93:1589-1608.

James Emil Flege. 1995. Second language speech learning: Theory, findings, and problems. In Winifred Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium, MD: York Press. pp.233–277.

John Goldsmith. 1976. *Autosegmental Phonology*. MIT dissertation.

Susan G. Guion, James Emil Flege, Rieko Akahane Yamada, and Jesica C. Pruitt. 2000. An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, 1075:2711–2724.

Eva Gårding. 1977. *The Scandinavian word accents*. Lund: Gleerup.

Yen-Chen Hao. 2012. Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, 40:269–279.

Yen-Chen Hao. 2014. The Application of the Speech Learning Model to the L2 Acquisition of Mandarin Tones. In *Proceedings of the 4th International Symposium on Tonal Aspects of Languages (TAL 2014),* 67-70. Nijmegen.

James J. Jenkins and Grace H. Yeni-Komshian. 1995. Cross-language speech perception: perspective and promise. In Winifred Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press. pp.433–463.

Constantine Kiriloff. 1969. On the auditory perception of tones in Mandarin. *Phonetica*, 20:63–67.

Yuh-Shiow Lee, Douglas A. Vakoch, and Lee H. Wurm. 1996. Tone perception in Cantonese and Mandarin: A cross linguistic comparison. *Journal of Psycholinguistic Research*, 25:527–542.

Jie Liang and Vincent J. van Heuven. 2007. Chinese tone and intonation perceived by L1 and L2 listeners. In Carlos Gussenhoven and Tomas Riad (eds.), *Tones and tunes. Volume 2: Experimental studies in word and sentence prosody*. Berlin: Mouton de Gruyter. pp.27-61.

William Bill McGregor. 2015. *Linguistics: An Introduction*. New York: Bloomsbury Academic.

Bertil Malmberg. 1963. *Phonetics*. New Yorker: Dover Publications.

Terence Odlin. 1989. *Language Transfer: Cross-linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.

Linda Polka. 1992. Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics*, 521:37–52.

Amanda Reid, Denis Burnham, Benjawan Kasisopa, et al. 2015. Perceptual assimilation of lexical tone: the roles of language experience and visual information. Atten Percept Psychophys. 77(2):571-591.

Tomas Riad. 1996. Remarks on the Scandinavian Tone Accent Typology. *Nordlyd: Tromsø University Working Papers on Language and Linguistics*, 24:129–156.

Xiaonan Susan Shen. 1989. Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association*, 243:27–48.

Connie K. So and Catherine T. Best. 2010. Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Language and Speech*, 532:273-293.

Yue Wang, Michelle M. Spence, Allard Jongman, and Joan A. Sereno. 1999. Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, 1066:3649–3658.

Brian J. Wenk. 1986. Crosslinguistic influence in second language phonology: Speech rhythms. In Kellerman, E. and M. Sharwood Smith (eds.), *Crosslinguistic Influence in Second Language Acquisition*. New York: Pergamon Press. pp.120–133.

Reiko Akahane Yamada. 1995. Age and acquisition of second language speech sounds: perception of American English /r/ and /l/ by native speakers of Japanese. In Winifred Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press. pp.305–320.

Moira Yip. 1980. *The Tonal Phonology of Chinese*. Cambridge, MA: The MIT Press.

Moira Yip. 1989. Contour tones. *Phonology*, 6:149-174.