

*Proceedings of the first workshop*

## **Chinese Accents and Accented Chinese**

**CAAC2014**

*at the Nordic Center, Fudan University, Shanghai, 9-10 October 2014.*

Eds. Marjoleine Sloos & Jeroen van de Weijer.

Publication date: October 2015

## Contents

### **Jeroen van de Weijer**

The role of Universal Grammar in second language acquisition .....1

### **Mingqiong Luo**

Perception and production of Mandarin nasal codas by Shanghainese speakers.....16

### **Xinlu Yang**

Oral fluency in the Mandarin Chinese speech of Uygur learners.....38

### **Bijun Ling & Jie Liang**

Focus and tonal implementation in Shanghai Chinese..... 56

### **Wenjun Chen**

A Study of Pitch Pattern of Declarative Sentences Uttered by Chinese speakers of English in Comparison with Native Speakers of English.....70

# The role of Universal Grammar in second language acquisition

Jeroen van de Weijer

*Shanghai International Studies University*

## Abstract

This paper discusses the question of whether Universal Grammar is relevant to the study of second language acquisition—in particular second language phonology. It will be shown that the notion of Universal Grammar has been hard to define coherently, which makes it difficult to implement it in second language acquisition research. One framework which offers a possibly explicit definition of Universal Grammar is Optimality Theory, in which phonological constraints are proposed to be universal. This makes it possible to assess the evidence for Universal Grammar directly. However, cases where Universal Grammar has been invoked to explain certain aspects of second language acquisition (in particular, errors made by second language learners) turn out to be largely spurious. The tentative conclusion must be that there is no role for Universal Grammar in second language acquisition. Arguably, this is also a better research strategy methodologically.

**Keywords:** Universal Grammar, second language acquisition, error analysis, Optimality Theory.

## 1 Introduction

Second language acquisition is a vibrant area of linguistic investigation, with many important applications. Of great importance is why second language learners make the errors that they do. Of course, error analysis is not the only or ultimate objective of second language acquisition research. For instance, some learners avoid using difficult patterns, so that some errors do not show up because they have not been made, but as a result such learners sound (or write) less natural than native speakers. Learners also show differences in speed acquiring target structures either in syntax, semantics or phonology. Still, error analysis provides insight

*Proceedings of the first workshop “Chinese Accents and Accented Chinese” (1<sup>st</sup> CAAC) 2015,  
at the Nordic Center, Fudan University, Shanghai, 9-10 October 2014.*

Eds. Marjoleine Sloos & Jeroen van de Weijer.

into what L2 learners do wrong, a necessary step to develop exercises to improve. Thus, a traditional goal of second language acquisition research is to identify what factors are involved in mistakes (e.g. Corder (1967, 1981)). Some of these factors are obviously related to the learner's first language (L1), but the literature has shown that others are not (see e.g. White (2003)). Sometimes such errors are attributed to universal laws or tendencies across languages, referred to as "Universal Grammar". There have been some studies in second language syntax that claim an effect of UG in this area (e.g. White (2003: Chapter 2)). In this paper we investigate the explanatory potential of Universal Grammar in second language phonology.

This paper is organised as follows: in section 0 we present a short discussion of error analysis in second language acquisition (henceforth SLA). Errors in SLA are typically attributed to a number of factors, of which Universal Grammar (henceforth UG) is one. Section 0 then tries to define UG, illustrating the difficulties this concept presents. We will offer a tentative working hypothesis, based on Optimality Theory, to make the investigation of SLA more realistic. Section 1 presents a number of case studies in which UG has been invoked as a way of explaining typical SLA patterns for a number of languages. The final section points out the limitations of this study and briefly concludes.

## 2 Errors in second language acquisition

Second language acquisition is a path toward (near)native competence in a second (or foreign, or third) language which typically involves at least some errors. It is important to understand what causes these errors, so that adequate practice and testing material can be developed. It will also advance our understanding of the way language works, with implications for multilingualism, language loss and other areas.

In this section we will discuss the causes of errors. A number of caveats should be made. We sidestep a whole debate on what exactly constitutes an error, and we also recognise the fact that errors are not the only source of information about a learner's progress or failures. For instance, avoidance of difficult sounds or complicated syntactic patterns also clearly plays a role in SLA (see e.g. Färch & Kasper (1984)). A learner who avoids such patterns may not

show errors, but may still sound less-native because they lack patterns that are found in L1 speakers. Finally, we focus on pronunciation errors, i.e. phonetic or phonological errors. The questions with respect to UG that are raised here can also be pursued for syntax or other linguistic subfields. We will return to this in the final section.

A first source of errors by L2 learners is clearly the interference of L1, also referred to as (negative) transfer, although the precise cognitive mechanisms this involves remain ill understood. The learner will use the sounds, the segmental and suprasegmental patterns (or rules), and the structures of their native language (or another previously learned language) in the L2 (or a language learned later). For instance, a speaker whose L1 does not have an aspiration contrast (such as Dutch or French) will have difficulties with aspiration languages such as English or Chinese and typically make mistakes, for instance by not aspirating stops that should be aspirated in the L2. The literature on interference is vast (starting from Lado (1957) and Weinreich (1953) through Flege (1995) and Best (1995) up to the present day) and we will not add to this here.

A second source of errors is often identified as “Universal Grammar” in the literature (see e.g. Cook (1985), Schachter (1988), Finer (1991), Felix (1995), Thomas (2004), Liceras (2010), and, very recently, Ambridge, Pine & Lieven (2014) for support, criticism or discussion of this idea). For instance, sometimes languages lack certain structures or sounds so that they cannot be transferred to the L2 that is being learned. In this case, certain tendencies may recur in different L2s. Such cases are the focus of this paper. Consider, for instance, the L2 acquisition of a lexical tone language (such as Chinese) by a speaker whose L2 does not have lexical tone. Since there is (supposedly) nothing in the native language to compare the tones with, the learner may be subject to difficulties, and make errors, that are suffered by other L2 learners of non-tone languages too. If errors in cases like this are recurrent, they might be explained by invoking “universal” principles, and if we understand such errors better, we may again develop adequate material to avoid or repair them. Consider another example: both English and (Mandarin) Chinese have a contrast between aspirated (voiceless, or “fortis”) and non-aspirated (voiced, or “lenis”) stops. In English this contrast appears both word-initially and word-finally, but in Chinese it only occurs initially. Thus, in final position, Chinese

learners of English are predicted to face difficulties and make mistakes, which is borne out by the literature (see below). These mistakes take on a number of shapes, and since they cannot be attributed to interference from the native language (since Chinese lacks word-final obstruents completely), in cases like this Universal Grammar is invoked. In other words, the fact that Chinese learners face difficulties and make errors is a fact predicted by interference, but the exact shape of the errors is not predicted by interference. If it is a consistent error, it might be predicted by a factor we refer to as Universal Grammar. We return to this case below (section 4.2).

A third source of errors include hypercorrections, e.g. the overapplication of phonological or morphological rules in the target language (see e.g. Pinker (1995)). These clearly cannot be attributed to either interference or to Universal Grammar, but rather seem to be due to general learner strategies (rule formation), quite similar to what happens in first language acquisition.

### 3 What is Universal Grammar?

In this section we will discuss possible definitions of Universal Grammar. In this section we will first discuss the more general notion of “universals”, which, purportedly, includes UG. After that we will examine some of the definitions that have been proposed in the literature and discuss this with a focus on their relevance to SLA. Finally, we will adopt a working definition by identifying UG with Optimality Theory constraints (Prince & Smolensky 1993 [2004]).

Major (2008: 76) gives a useful overview what, in his view, should be included in language universals (which, by the way, is also not included as a separate entry in Brown (2006)). The list Major proposes is given in (1):

## (1) Universals:

- UG
- learning principles
- markedness considerations (Greenberg 1966, 1978)
- rules
- processes
- constraints (Prince & Smolensky 1993 [2004])
- stylistic universals

This is an interesting list which merits some discussion. Note that the constraints in Optimality Theory (OT; Prince & Smolensky (1993 [2004])) are supposed to supersede the “rules” and “processes” of earlier generative grammar, so, if OT is on the right track, these could be struck off the list in (1). Secondly, the content of OT’s constraints, or at least of its markedness constraints, has been identified as precisely the “markedness considerations” that were uncovered by the earlier work of Greenberg and others. The central insight of OT was to promote these “considerations” to fully-fledged principles of the grammar (see e.g. Kager (1999), who considers constraints to be part of UG, and van de Weijer (2014) for general discussion). Thus, the content of UG in the list in (1), according to Major, excludes OT constraints, although these are also universal. The question is what Major considers part of UG (he does not give a definition). [Note: The *Encyclopedia of Language and Linguistics* (Brown 2006) does not, among its 3,000+ entries, have an entry for UG, although many articles in this encyclopedia refer to the concept.] Below we will return to this issue and identify OT constraints with UG. First, we will look at a number of definitions from the literature and discuss their relevance to SLA.

An important aspect of UG appears to be its relation with typology. In a very useful exposition of the history of the concept of UG and its relation to second language acquisition, Thomas (2004) proposes the definition in (2):

(2) Thomas (2004: 2, 25)

“[U]niversal grammar consists of the linguistic properties shared by all languages”

At first glance, this definition seems largely irrelevant to SLA. Examples of linguistic properties shared by all languages might be very general properties, such that all languages have a lexicon, or a distinction between nouns and verbs, or syllable structure (however any of these terms are defined). If these are shared by all languages, then they will, by definition, be equal for an L<sub>1</sub> or L<sub>2</sub> and not affect language acquisition. Of course, difficulties arise when the syllable structure of L<sub>1</sub> is not the same as that of a target L<sub>2</sub>--but in this case, the differences are not part of UG and therefore irrelevant to our question if UG is involved in SLA. It may of course be the case that there are deeper, perhaps unsuspected, linguistic properties that are shared by all languages, such as syntactic constituency, the Subjacency condition in syntax (Chomsky (1981), see discussion in White (2003: 121ff.)) or (some version of) the Obligatory Contour Principle (e.g. McCarthy (1986)) in phonology. Again, if all languages share these properties then no special predictions are made with respect to SLA. The most interesting hypothesis is the one inherent in Optimality Theory: the properties of UG are identified with OT constraints. These constraints are universal, thus exist in all languages (conforming to the definition in (2)), but they are violable: individual languages rank the constraints differently and outputs are computed on the basis of satisfaction of the constraints by different candidate forms. We will return to this possibility below.

It should be pointed out that if all languages share some particular property, this is not necessarily an argument for UG, in the sense that it must somehow be intrinsic to language (or innate). It may also be explained as a property that must emerge as a result of the fact that different languages are acquired under similar circumstances and used for similar purposes.

A second, and perhaps even more basic, aspect of UG that is often mentioned is its relation to language acquisition. In Chomsky's original work, it is observed that a first language is acquired so fast and on the basis of such poor data, that a “language acquisition device” was posited (Chomsky 1965), which, in the early proposals, contained principles and parameters:

principles being inviolable principles across languages, and parameters that could differ (within boundaries) between languages (such as word order, or syllable structure). This language acquisition device was also identified with UG. The definition in (3) is from Dresher & Hornstein (1976):

(3) Dresher & Hornstein (1976: 329)

UG = “the principles according to which languages are organized and learned”

This rather opaque definition clearly applies to first language acquisition rather than second language acquisition. It is therefore not surprising that an explicit definition of UG has not been agreed upon. Chomsky quotes: “The concept of UG remains highly enigmatic, and appears to lie beyond the reach of contemporary inquiry” (Chomsky 2007: 24). The explanatory adequacy of the hole approach has also been questioned: “The impression that there is poverty of the stimulus leads to the conclusion that the child must be born with the constraints of UG already encoded in its brain. But UG does not explain why some violations are not found in child language: it just lists the cases as a taxonomy of principles, constraints, and parameters” (Bouchard 2013: 231). This is also a point made by Wechsler (2010: 341ff), in an investigation of a morphosemantic universal.

Returning to Major’s list of universals in (1), above we observed that it is possible to narrow down the list of universals to OT constraints, learning strategies and stylistic universals. With respect to constraints, a fundamental question is whether constraints are innate or acquired on the basis of data the child is exposed to. This question will not be discussed here but see van de Weijer (2012), van de Weijer & Sloos (2013), and van de Weijer (2014), among many others, for arguments to regard constraints as being acquired on the basis of first-language data. The fact that constraints could play a role in second language acquisition is very interesting, especially if these constraints could not have been acquired on the basis of first-language data. This would constitute strong evidence for the idea that constraints (as part of UG) are innate. Consider a language which freely allows final voiced obstruents: language learners of such a language will not posit a generalisation against final voiced obstruents following the logic of constraint acquisition (see references above). If

speakers of such a language devoiced final obstruents in a second language, this could be taken as evidence for the action of a constraint against final voiced obstruents – a constraint that could not have been learned on the basis of the L<sub>1</sub> data. In the following section, we will look at three case studies that are variations on the same theme, where evidence for a particular pattern is absent in the native language, or where second language learners even *avoid* pronunciations in their L<sub>2</sub> that are permitted in their native language.

## 4 Case studies invoking UG

This section presents a number of case studies in which UG has been invoked to explain certain facts about SLA.

### 4.1 Chinese tones

In section 2 I noted that interference might not play a role if a language (L<sub>1</sub>) lacks certain sounds or categories of grammar which do appear in the L<sub>2</sub> that is being acquired. An example might be lexical tone in an L<sub>2</sub> (e.g. Chinese) where the learner's L<sub>1</sub> does not have lexical tone (e.g. English). [Note that errors in tonal L<sub>2</sub> acquisition usually take the form of confusion matrices (e.g. Tones 2 and 3 are often confused by L<sub>2</sub> learners of Chinese; see e.g. Zhu (2002: Chs. 3, 4)), but it is also possible to analyse general errors in perception or in production.] However, of course, the phonetic property that is responsible for lexical tone in tone languages, i.e. pitch, is also used in stress languages, albeit with a different function, for instance to signal stress, intonation, and focus. These functions of pitch are expected to show lots of interference when non-tone language speakers acquire a tone language as L<sub>2</sub>. This is indeed what is found (see e.g. Wang et al. (1999)). Moreover, length and loudness are also phonetic factors that occur in both languages. In Chinese, for instance, Tone 3 is generally produced with longer duration than other tones, while Tone 4 is relatively short (Xu 1997). Speakers of English will associate these length contrasts with long and short vowels, because this is a feature they are exposed to in their native language; longer duration is associated with Tone 3 (Blicher, Diehl & Cohen 1990). Thus knowledge of English (a non-tonal language) can

still interfere with tonal acquisition in Chinese. There is no clear evidence if and how UG would play a role in a case like this.

#### 4.2 Voiced-voiceless contrast for Chinese learners of English

As is well known, standard Mandarin Chinese permits obstruents in initial position but not in final position. Wang (1995) found that in an experiment with nonsense words, Chinese learners of English replace some of the voiced obstruents with voiceless ones, as in (4):

(4)      vɪg →vɪk

(This result was replicated for existing (non-nonsense) words by Qibi (2014)).

These results are interpreted by Broselow, Chen & Wang (1998) as “emergence of the unmarked” (TETU, see Kager (1999: Ch. 5)), since voiceless stops are cross-linguistically preferred over voiced ones in a typical neutralization position (end of the word). Chinese native phonology does not have final obstruents, and therefore not a rule of final devoicing either, so by definition the native phonology has no bearing on the question why this particular error is made. Thus, TETU refers to the (*a priori* unexpected) effect of a constraint that could not have been learned on the basis of ambient data, which is used in Optimality Theory as potential evidence for the universality (and perhaps innateness) of OT constraints.

This is an interesting case which raises the important question if the “interlanguage” (the stage at which these particular Chinese learners make these particular errors) is rule-governed (or constraint-governed) or not. If we accept the idea that all errors are outputs of actual grammars (where ‘grammars’ correspond to the mental organizations that learners have imposed on the language data to which they are exposed), then this seems a fair case of an effect of Universal Grammar, specifically a constraint like \*VOICE (“No (final) voiced obstruents”). However, two facts should be noted: the error of final devoicing in the data by Wang (1995) only occurred rather infrequently: only 19% of the errors for final voiced stops involved final devoicing (Broselow, Chen & Wang 1998: 264). Errors like final epenthesis (pronouncing *bag* like a disyllable - 36%) and deletion of the final voiced obstruent (43%) were

much more common. (Both of these errors can be explained by the native language phonology, since Chinese does of course allow CV syllables.) This seems rather a poor result for a putatively universal constraint; at any case the interference solutions are more common. A second option is to impute these errors to a more general factor like ease of articulation. Final voiceless stops are easier to pronounce than final voiced stops (which is, no doubt, the phonetic basis of the \*VOICE constraint), so Chinese learners find these difficult and therefore make mistakes. Under this second approach, no appeal to UG is necessary and therefore seems preferable.

### 4.3 Korean final vowel insertion

Our final case study comes from Korean, where some loanwords from English (variably) show final vowel insertion, which is not warranted by the native Korean phonological system. This is discussed by a number of authors, among whom Rhee (2002), Kang (2003), Rhee & van de Weijer (2003) and Boersma & Hamann (2009). Data from Kang are given in (5), indicating syllable boundaries:

- (5)      bat → pæ.t<sup>h</sup>i              pad → pæ.t<sup>h</sup>i  
           deck → tɛ.k<sup>h</sup>i              gag → kæ.k<sup>h</sup>i  
           hip → hi.p<sup>h</sup>i              tube → t<sup>h</sup>ju.pi

Other words (not given in (5)) take no final vowel, or do so variably. Final vowel insertion is not motivated by native phonology, because final [t] is allowed in native Korean phonology (see any of the sources mentioned above). Final vowel insertion could therefore be explained by a (universal) preference for CV syllables (in OT terms, this could be captured by the constraint NOCODA). An analysis along these lines is proposed by Goad (2002) (not employing the constraint NOCODA but another OT constraint). This case is interesting and clearly different from the previous one, because Korean, unlike Mandarin Chinese, does allow final voiceless obstruents. Although these data predominantly concern loanword incorporation, they have also been reported in second language acquisition (Broselow & Kang 2013).

Rhee (2002) suggests that final vowel insertion in these data is a way for the Korean learners to preserve the contrast that is known or perceived to exist in English (*bag* vs. *back*) but which the learners cannot (yet) produce faithfully. This is also essentially the explanation offered by Kang (2003). Finally, it should, again, be pointed out, that vowel insertion does not take place consistently, so that reliance on a putative universal seems rather too strong in this case.

To conclude this section, we have briefly investigated some cases where first-language interference does not seem to play a role in the errors that second language learners make. This is not necessarily an argument for a role of universal grammar (however this is defined): in all cases simpler explanations, based on simple phonetics or functional considerations, are available.

## 5 Conclusion

The outlook for UG in general is bleak. There is no consensus about its definition or content, which makes a workable application in a field such as second language acquisition difficult.

An obvious limitation to the case studies examined in section 0 is that they were limited to phonetics/phonology. Cases have been described in the literature that are supposed to show the operation of syntactic principles of UG (e.g. White (1989, 2003)). This may mean that UG is limited to syntax, which would be an interesting result to contemplate further. After all, syntax is the truly generative component of language.

As to future research, it is necessary to examine more case studies where UG has been invoked to account for SLA errors (or other conspicuous patterns not identified as errors).

If UG plays no or only a minor role in SLA errors, as the discussion here and the case studies examined so far suggest, then the role of transfer is even larger than usually assumed. Research might focus on this source of errors, and underscores the importance of contrastive analysis. It is important, however, that not only transfer causes errors, but also general learning principles, e.g. those resulting in hypercorrection errors, while other factors (such as motivation) also play a role. Further study of these general cognitive principles is called for.

Finally, I wish to mention a methodological point which, in my view, is at least as important as the theoretical arguments probed above. This is that *not* relying on UG to explain errors is a sound research strategy: only if other explanations have been sought out and explored, resource could be taken to a powerful (and stipulative) mechanism such as UG. Explaining errors as a result of UG is explaining errors away, but is not really explaining.

### References

- Ambridge, Ben, Julian M. Pine & Elena V. M. Lieven. 2014. Child language acquisition: Why universal grammar doesn't help. *Language* 90.3, 53-90.
- Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In W. Strange (ed.), *Speech Perception and linguistic Experience: Issues in Cross-language Research* 171-204. Baltimore: York Press.
- Blicher, Deborah L., Randy L. Diehl & Leslie B. Cohen. 1990. Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*,
- Boersma, Paul & Silke Hamann. 2009. Loanword adaptation as first-language phonological perception. In Andrea Calabrese & W. Leo Wetzels (eds.), *Loan Phonology*, 11-58. Amsterdam and Philadelphia: John Benjamins.
- Bouchard, Denis. 2013. *The Nature and Origin of Language*. Oxford: Oxford University Press.
- Broselow, Ellen, Su-I Chen & Chilin Wang. 1998. The Emergence of the Unmarked in Second Language Acquisition. *Studies in Second Language Acquisition* 20, 261-80.
- Broselow, Ellen & Yoonjung Kang. 2013. Second Language Phonology and Speech. In J. Herschensohn & M. Young-Scholten (eds.), *The Cambridge Handbook of Second Language Acquisition*. Cambridge: Cambridge University Press.
- Brown, Keith. 2006. *Encyclopedia of Language and Linguistics*. London: Elsevier.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, Noam. 2007. Of minds and language. *Biolinguistics* 1.1, 9-27.
- Cook, Vivian. 1985. Chomsky's universal grammar and second language learning. *Applied Linguistics* 6, 19-44.
- Corder, Stephen Pit. 1967. The significance of learner's errors. *IRAL-International Review of Applied Linguistics in Language Teaching* 5.1-4, 161-70.
- Corder, Stephen Pit. 1981. *Error Analysis and Interlanguage*. Oxford: Oxford University Press.
- Dresher, B. Elan & Norbert Hornstein. 1976. On some supposed contributions of artificial intelligence to the scientific study of language. *Cognition* 4.4, 321-78.
- Färch, Claus & Gabi Kasper. 1984. Two ways of defining communication strategies. *Language Learning* 34, 45-63.
- Felix, Sascha. 1995. Universal Grammar in L2 Acquisition. In Lynn Eubank, Larry Selinker & Mike Sharwood-Smith (eds.), *The current state of interlanguage*, 139-51. Amsterdam and Philadelphia: John Benjamins.
- Finer, Dan. 1991. Binding Parameters in Second language acquisition. In Lynn Eubank (ed.), *Point/Counterpoint: Universal Grammar in the Second Language*, 351-74. Amsterdam and Philadelphia: John Benjamins.
- Flege, James E. 1995. *Second language speech learning: theory, findings, and problems*. Baltimore: York Press.
- Goad, Heather. 2002. Markedness in right-edge syllabification : parallels between L1 and L2. Paper presented at GLOW 2002, Amsterdam and Utrecht.
- Greenberg, Joseph H. (ed.) 1966. *Universals of language; report of a conference held at Dobbs Ferry, New York, April 13-15, 1961*, 2nd edn. Cambridge, Mass.: MIT Press.
- Greenberg, Joseph H. 1978. Some generalizations concerning initial and final consonant clusters. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language*, vol. 2, 243-79. Stanford: Stanford University Press.

- Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.
- Kang, Yoonjung. 2003. Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20, 219-73.
- Lado, Robert. 1957. *Linguistics across Cultures; Applied Linguistics for Language Teachers*. Ann Arbor, MI: University of Michigan Press.
- Liceras, Juana M. 2010. Second Language Acquisition and Syntactic Theory in the 21st Century. *Annual Review of Applied Linguistics* 30.1, 248-69.
- Major, Roy C. 2008. Transfer in second language phonology. In Jette G. Hansen Edwards & Mary L. Zampini (eds.), *Phonology and second language acquisition*, 63-94. Amsterdam and Philadelphia: John Benjamins.
- McCarthy, John J. 1986. OCP effects: gemination and antigemination. *Linguistic Inquiry* 17.2, 207-63.
- Pinker, Steven. 1995. Why the child holded the baby rabbits: A case study in language acquisition. In Daniel N. Osherson & Lila R. Gleitman (eds.), *An invitation to cognitive science*, vol. 1, 107-33. Cambridge, MA: MIT Press.
- Prince, Alan & Paul Smolensky. 1993 [2004]. *Optimality Theory - Constraint Interaction in Generative Grammar*. London: Blackwell.
- Qibi, Xinyue. 2014. *Aspiration errors in English by Chinese Learners*. MA dissertation, Shanghai International Studies University.
- Rhee, Sang Jik. 2002. *Empty nuclei in Korean*. PhD dissertation, Leiden University.
- Rhee, Sang Jik & Jeroen M. van de Weijer. 2003. Epenthesis of [I] in Native and Loanword Korean Phonology. *Proceedings of the 2nd Korean International Conference on Phonology: Optimality Theory in Retrospect and in Prospect*, 67-79. Seoul: The Phonology-Morphology Circle of Korea.
- Schachter, Jacquelyn. 1988. Second language acquisition and its relationship to universal grammar. *Applied Linguistics* 9, 219-35.
- Thomas, Margaret. 2004. *Universal Grammar in second language acquisition*. London and New York: Routledge.

- van de Weijer, Jeroen M. 2012. *Grammar as selection: Combining Optimality Theory and Exemplar Theory*. Nagoya: Kougaku Shuppan.
- van de Weijer, Jeroen M. 2014. The Origin of OT Constraints. *Lingua* 142, 66-75.
- van de Weijer, Jeroen M. & Marjoleine Sloos. 2013. Learning Markedness Constraints: The Case of French. In Suzanne Aalberse & Anita Auer (eds.), *Linguistics in the Netherlands 2013*, 188-200. Amsterdam and Philadelphia: John Benjamins.
- Wang, Chilin. 1995. *The acquisition of English word-final obstruents by Chinese speakers*. PhD dissertation, State University of New York at Stony Brook.
- Wang, Yue, Michelle M. Spence, Allard Jongman & Joan A. Sereno. 1999. Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America* 106, 3649-58.
- Wechsler, Stephen. 2010. What 'you' and 'I' mean to each other: Person indexicals, self-ascription, and theory of mind. *Language* 86.2, 332-65.
- Weinreich, Uriel. 1953. *Languages in Contact: Findings and Problems*. The Hague and Paris: Mouton.
- White, Lydia. 1989. *Universal Grammar and Second Language Acquisition*. Amsterdam and Philadelphia: John Benjamins.
- White, Lydia. 2003. *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.
- Xu, Yi. 1997. Contextual Tonal Variations in Mandarin. *Journal of Phonetics* 25, 61-84.
- Zhu, Hua. 2002. *Phonological development in specific contexts: Studies of Chinese-speaking children*. Clevedon: Multilingual Matters.

# Perception and Production of Mandarin Nasal Codas by Shanghainese Speakers

Mingqiong Luo

*Shanghai International Studies University*

## Abstract

Flege's (1995) Speech Learning Model (SLM) is one of the best-known models concerning second language sound acquisition. It hypothesizes that sounds in the L<sub>1</sub> and L<sub>2</sub> are related perceptually “at a position-sensitive allophonic level” in the form of “phonetic categories”, but it only predicts a moderate correlation between perception and production. This study intends to test the hypothesis and prediction in SLM by studying native Shanghainese speakers’ production and perception of syllable-final nasals in Standard (Beijing) Mandarin (SM). Shanghainese and SM are two major Chinese dialects with little mutual intelligibility. SM allows two nasal phonemes in syllable-final position: the coronal nasal /n/ and the velar nasal /ŋ/, while Shanghainese only has one nasal phoneme /N/, which has three allophones: [̃] (nasalization of the preceding vowel or rhyme), [ɲ] and [ŋ]. Based on the SLM, this study predicts that native Shanghainese speakers have difficulty in distinguishing between [n] and [ŋ] that follow [i], [ə] and [y] nuclei, but not when they follow [a] and [u] nuclei. This is tested in a production and a perception experiment. The results reveal a strong correlation between production and perception of the [an]-[aŋ] rhymes, but a modest correlation for the [in]-[iŋ] and [ən]-[əŋ] rhymes. This may be attributed to the L<sub>1</sub> interference of the L<sub>2</sub> as well as to the L<sub>2</sub> linguistic environment. The study shows that the SLM makes correct predictions for the study of sound production and perception across Chinese dialects, and may serve as a viable theoretical framework in similar studies.

**Keywords:** perception and production, Mandarin, Shanghainese, syllable-final nasals.

## 1 Introduction

In second language acquisition (SLA), knowledge of the learner's first language (L1) is believed to be one of the many factors that result in foreign accent in his/her second language (L2), other factors being age, length of residence, gender, motivation and instruction (Harnsberger 2001). The Speech Learning Model (Flege, 1995) postulates that the influence results from the interference between phonetic categories in the learner's two languages.

Phonetic categories can be defined as the "language-specific aspects of speech sounds" that "are specified in long-term memory representations" (Flege 1995: 239). The effect of the L1 phonetic categories on those of the L2 is mainly two-fold: (i) they serve as the "automatic interpretive schemata" which guide the learners to attend to those phonetic properties of the L2 sounds that are considered to be important in the L1 phonology; (ii) they filter out properties of L2 sounds that are generally considered redundant by the L1 phonology. For example, Flege (1995) found that speakers of many languages (e.g., Dutch, French, Italian, Portuguese) who are L2 learners of English tend to pronounce the voiceless stops /p, t, k/ as unaspirated. This is either because voiceless stops are always realized as unaspirated in these languages, or because [aspiration] is not a categorical feature in the phonological systems of these languages.

Therefore, in order to acquire an L2 sound, the learner has to establish a new phonetic category for it. Failure to establish an L2 phonetic category will result in a failure to perceive the L2 sound accurately. In this case, perceptual assimilation will occur, which means that the L2 sound is assimilated to the closest L1 sound, and that the learner perceives and produce the L2 sound according to the categorical features (or "contrastive phonetic properties", as Flege (1995: 238) puts it) of the L1 sound, which will result in a misinterpretation of the L2 sound. The establishment of a new L2 phonetic category, however, does not necessarily guarantee accurate perception, because the new category established by an L2 learner may differ from that of the L1 speaker. Flege (1995: 239) hypothesizes that the phonetic category established for L2 sounds by a bilingual may differ from a monolingual's if: (i) the bilingual's category is "deflected" away from the L1 category to maintain phonetic contrast between categories in a common L1-L2 phonological space; (ii) the bilingual's representation is based on different features, or feature weights than a monolingual's.

As to the analytical nature of the phonetic category, the SLM makes the hypothesis that sounds in the L1 and L2 are related perceptually to one another "at a position-sensitive allophonic level", rather than at a more abstract phonemic level (Flege 1995: 239). This hypothesis has at least two important

implications: (i) the L2 sound is perceived with different accuracy at different positions within the syllable. For example, native Japanese learners of English (NJ) perceive the English liquids [l] and [ɹ] more accurately in word-final than in word-initial position because allophones of English liquids are categorized differently at syllable-final position by native Japanese learners (Takagi 1993 cited by Flege 1995: 240); (ii) The perception of L2 sounds is influenced by the corresponding L1 sounds at the same position of the syllable. To be more specific, the perception of an L2 sound at the syllable-initial, syllable-medial and syllable-final positions is influenced by the closest L1 sounds in the corresponding positions.

Establishing an L2 phonetic category is important not only for the perception of the L2 sound, but also for the production of the L2 sound. According to the SLM, the production of a sound eventually corresponds to the properties represented in its phonetic category representation (Flege 1995: 239). That is to say, if the new L2 phonetic category is categorized according to the correct phonetic properties, the learner may pronounce it the way the native speakers of L2 pronounce it; but if the L2 phonetic category is categorized according to a deflected or even a wrong phonetic property, it is impossible for the learner to produce it the correct way.

However, although establishing L2 phonetic categories is perceptually-based, and tends to have a direct and decisive influence on the production of the L2 sound, no strong and direct correlation is necessarily observed between production and perception. The SLM predicts only a moderate correlation between production and perception of L2 sounds, even for highly experienced speakers of L2 (Flege 1999). To be more specific, it is possible for an L2 learner to be excellent in perception but poor in production of the L2 sounds. But if an L2 learner is poor in perception, it is not possible for him/her to be excellent in production. Good production must entail good perception.

This study intends to test whether the SLM framework is applicable to the study of Chinese speakers of two dialects. We choose to study the perception and production of the nasal codas by native Shanghainese speakers of SM for three reasons: (i) Shanghainese and SM are known for their significant phonological difference in terms of the nasal codas; (ii) SM nasal codas are known to be difficult for Shanghainese to discern, but the reasons are not entirely clear; (iii) the subjects are easy to access.

The rest of the paper is arranged as follows: Section 2 is a brief introduction of the SM and Shanghainese nasal coda inventories, and the specific hypotheses and predictions made according to the SLM. Section 3 is the description of two experiments conducted to test the hypotheses made in

Section 2, their objects of study, materials used, experiment procedures, and their results. Section 4 is the discussion; and Section 5 is the conclusion.

## 2 Background

Shanghainese and SM are two major Chinese dialects, which belong to the Wu Dialect Group and the Northern Dialect Group, respectively. It is well-known that SM has two nasal codas: /n/ and /ŋ/, which form minimally contrastive pairs with each other, and are thus considered as two independent nasal phonemes, like in (1).

(1)	看	vs.	康	宾	vs.	冰
	kan <sup>51</sup>		kang <sup>55</sup>	bin <sup>55</sup>		bing <sup>55</sup>
	/k <sup>h</sup> an <sup>51</sup> /		/k <sup>h</sup> aŋ <sup>55</sup> /	/pin <sup>55</sup> /		/piŋ <sup>55</sup> /
	‘see’		‘health’	‘guest’		‘ice’
	根	vs.	更	前	vs.	强
	gen <sup>55</sup>		geng <sup>51</sup>	qian <sup>35</sup>		qiang <sup>35</sup>
	/kən <sup>55</sup> /		/kəŋ <sup>51</sup> /	/tɕ <sup>h</sup> jan <sup>35</sup> /		/tɕ <sup>h</sup> jaŋ <sup>35</sup> /
	‘root’		‘more’	‘front’		‘strong’

The SM nasal-final rhymes in the citation form are listed in pairs as follows:

(2)	Pinyin:	an-ang	in-ing	un-ong	ün-iong	en-eng
		ian-iang				
		uan-uang				uen-ueng
		üan				
	IPA:	an-aŋ	in-iŋ	u(ə)n-uŋ	yn-yŋ	ən-əŋ
		ian-iaŋ				
		uan-uaŋ				uən-uəŋ
		yan <sup>2</sup>				

<sup>2</sup> This rhyme does not have a velar-nasal-coda counterpart. Besides, it is in complementary distribution with the [uan] rhyme except when it follows the zero onset.

In connected speech, the two nasal codas in SM are realized as their citation forms in Intonational Phrase (IP)-final position<sup>3</sup>, but in IP-medial position, i.e., in the VN.C context, they will be assimilated to the following onset consonant in place of articulation and surface as their medial variants. Since SM has 23 initials which can be grouped into seven types according to place of articulation (bilabial, labiodental, dental, velar, palatal, retroflex and glide) (Luo 2014, Yuan 1983), the SM nasal codas [n] and [ŋ] have seven variants in IP-medial position: bilabial [m], labiodental [m̥], coronal [n], velar [ŋ], palatal [ɲ], retroflex [ɳ], and one glide variant. For example:

(3)	i.	拼搏	pin.bo	/p <sup>h</sup> in.pwo/	[p <sup>h</sup> im.pwo]	‘strive’
		评判	ping.pan	/p <sup>h</sup> iŋ.pan/	[p <sup>h</sup> im.pan]	‘judge’
	ii.	办法	ban.fa	/ban.fa/	[bãŋ.fa]	‘approach’
		方法	fang.fa	/faŋ.fa/	fãŋ.fa]	‘method’
	iii.	反弹	fan.tan	/fan.tan/	[fan.tan]	‘rebound’
		访谈	fang.tan	/faŋ.tan/	[fan.tan]	‘interview’
	iv.	宾客	bin.ke	/pin.k <sup>h</sup> ɿ/	[biŋ.k <sup>h</sup> ɿ]	‘guest’
		饼干	bing.gan	/piŋ.kan/	[piŋ.kan]	‘biscuit’
	v.	传记	zhuan.ji	/tʂwan.tei/	[tʂwaŋ.teɿ]	‘biography’
		讲究	jiang.jiu	/tɕjaŋ.tejəu/	[tɕjaŋ.tejəu]	‘particular’
	vi.	真正	zhen.zheng	/tʂɿn.tʂɿŋ/	[tʂɿŋ.tʂɿŋ]	‘real’
		强壮	qiang.zhuang	/tɕ <sup>h</sup> jaŋ.tʂwaŋ/	[tɕ <sup>h</sup> ãŋ.tʂwãŋ]	‘strong’
	vii.	贪玩	tan.wan	/t <sup>h</sup> an.wan/	[thãN.wãŋ]	‘love to play’
		中央	zhong.yang	/tʂuŋ.jaŋ/	[tʂũN.jãŋ]	‘center’

As can be seen from the examples above, in connected speech, the SM nasal codas [n] and [ŋ] are likely to change their phonetic quality due to place assimilation to the following syllable onsets. Sometimes, [n] is realized as [ŋ], and [ŋ] as [n]. It is of great interest that native SM (Beijing) speakers seldom get confused by these phonetic changes in perception, i.e., they can accurately tell whether the nasal coda is /n/ or /ŋ/. This study intends to reveal whether native Shanghaiese speakers will

<sup>3</sup> The domain of the place assimilation rule concerning SM nasal codas is the intonational phrase, which means that the place assimilation rule takes effect as long as the nasal coda is followed by another syllable in an intonational phrase. At the end of an intonational phrase the nasal coda is realized as its citation form. This paper uses prosodic words instead of intonational phrases as the examples for ease of illustration.

get confused by these phonetic changes in perception. This result is revealing in terms of the level of analysis in perception, i.e., whether the analysis takes place at the allophonic level, or at a more specific phonetic level.

In contrast, Shanghainese only has one nasal coda phoneme. Opinions differ as to the quality of the phoneme and its allophones. Some scholars (e.g. Zhu (1996) and Yang (2002)) argue that there is only one nasal coda in Shanghainese, i.e. the velar nasal /ŋ/, without allophones. Other scholars (e.g. Qian (1987, 2007), Duanmu (2008), Gu (2007), Luo (2014)) assume three (non-contrasting) nasal codas in Shanghainese, all allophones of one nasal phoneme /N/, which is unspecified for place features. To be more specific, Qian (2007) and Duanmu (2008) claim that the three nasal codas (allophones) are alveolar [n], velar [ŋ] and the nasalized coda [ ~ ]<sup>4</sup>, while Gu (2007) claims that the three allophones are palatal [ɲ], velar [ŋ] and the nasalized coda [ ~ ]. These two opinions can be schematized as follows:

(3)	ã	əŋ	oŋ		ã <sup>5</sup>	əŋ	oŋ
	iã	in	ioŋ		iã	iŋ	ioŋ
	uã	uəŋ	uã		uəŋ		
	yn				ɲ		
	(Qian 2007: 386)				(Gu 2007: 11)		

As can be seen from (3), the low open vowel [a] always co-occurs with the nasalized allophone [ ~ ], and the rounded back vowel [o] always co-occurs with the velar nasal coda allophone [ŋ]. But the high front vowel [i] and the central vowel [ə] are problematic. Some scholars (e.g. Zhu (1996) and Yang (2002)) suggest that they are followed by the velar nasal [ŋ], others (e.g., Duanmu (2008) and Qian (1987)) suggest that they are followed by the coronal nasal [n], and still others (e.g. Qian (1987) and (Gu, 2007)) maintain that they are followed by the palatal nasal [ɲ]. The reasons for differences in opinion are three-fold: (i) diachronically speaking, Shanghainese has always been in a constant change in terms of its sound quality. According to Qian (1987: 111), the nasal rhymes in standard Shanghainese<sup>6</sup> have undergone two major changes in the past 30 years: [ã] and [ã̃] have merged into

<sup>4</sup> In Shanghainese, [ã] sounds similar to [aŋ] (Qian 1987), i.e., the vowel is nasalized by a velar nasal. The phonetic difference between [V] and [Vŋ] is that in the former, the nasal cavity is opened at the same time that the vowel starts being pronounced, while in the latter the nasal sonorant begins audibly later than the vowel.

<sup>5</sup> Gu (2007) uses [ ~ ] to stand for the whole nasal rhyme, but the present author thinks that the rhyme should include the nuclear vowel. According to Gu (2007), the front [a] and the back [ɑ] in the Shanghainese rhyme merge into the mid [A]. Therefore, I will use [ã] to stand for the nasalized rhyme.

<sup>6</sup> Standard Shanghainese refers to the dialect spoken by native Shanghainese speakers living in the city area, in contrast to other varieties which are spoken in suburban and rural areas in Shanghai.

[ã], and [əŋ], [iŋ] and [uəŋ] have changed into [əŋ], [iŋ] and [uəŋ], respectively. According to Qian (1987), [A] is the mid low open vowel, and if this vowel is nasalized it sounds similar to the velar nasal [ŋ]; in the palatal nasal [ɲ] the place of articulation is in between that of the coronal nasal [n] and the velar nasal [ŋ]. (ii) Synchronically speaking, Shanghainese has many varieties, such as the standard inner-city variety, which is distinct from varieties spoken in the suburban and rural areas of the city. These varieties differ in terms of the nasal codas. Some varieties, such as the Songjiang variety, have [ŋ] as the only nasal coda, while other varieties, such as the inner-city variety, have [ɲ] as one of the three position-sensitive allophones (You, 2013); (iii) Methodologically speaking, some scholars prefer to represent the sound with specific phonetic properties, while others prefer to keep the phonetic details to a minimum and use typologically more common symbols to highlight the phonological contrast. For example, [ɲ] is closer to the phonetic realization of the sound, but typologically less common; while [n] is not so faithful to the phonetic realization of the sound, but more common typologically. Despite these complications and differences in opinion, there is consensus that in all varieties of Shanghainese there is only one nasal coda phoneme, for which we will use the symbol /N/.

According to the SLM, phonetic properties are the key to perception, so we will adopt the more phonetically-friendly way of representation, and use [ ~ ], [ɲ], [ŋ] to represent the three allophones of the nasal coda phoneme /N/ in Shanghainese, which co-occur with [A], [i, ə, y] and [o], respectively. In contrast, SM has two nasal codas [n] and [ŋ], which follow each of the five nuclear vowels, i.e., [a], [i], [ə], [u], and [y]. That is to say, each of the nasal coda allophones in Shanghainese has two correspondents in SM. How will Shanghainese speakers perceive and produce the SM nasal codas?

According to the SLM hypothesis that L1 and L2 sounds are related perceptually at a position-sensitive allophonic level, we can formulate the following two hypotheses. Firstly, native Shanghainese speakers will have difficulty in distinguishing between the SM nasal codas. Specifically, they will find it difficult to distinguish between [n] and [ŋ] after the vowels [i], [ə] and [y] because in this position Shanghainese has [ɲ], which is in between [n] and [ŋ] in terms of place of articulation. On the other hand, they will find it relatively easier to distinguish between [n] and [ŋ] after the vowels [a] and [u], because in this position Shanghainese has [ ~ ] and [ŋ], respectively, which are both similar to SM [ŋ]. We envisage the following two possibilities: (i) the native Shanghainese speakers will perceive the Shanghainese nasal coda [ɲ] for both of the SM nasal codas. In this case, they cannot discern the phonetic difference between [ɲ], [n] and [ŋ] at all; (ii) the native Shanghainese speakers perceive one of the SM nasal codas as [ɲ], while they establish a new phonetic category for the other nasal coda. In this case, they might do well in perceiving the SM nasal coda for which they have

established a new category but fail to achieve the same accuracy for which they have not. Secondly, context-sensitive phonetic variations of SM nasal codas in IP-medial position due to place assimilation should not influence the overall performance of native Shanghainese speakers in perception.

As to production, based on the SLM hypothesis that the accuracy with which L2 segments are perceived limits how accurately they will typically be produced, we predict that the SM rhymes for which few or no errors are made in perception tend to be more accurately produced by native Shanghainese speakers. However, SLM does not predict a direct correlation between perception and production, i.e., native Shanghainese speakers who do well in perceiving the SM nasal codas will not necessarily do well in producing them, but native Shanghainese speakers who can produce the SM nasal codas accurately will tend to perceive them with greater accuracy.

In order to test these hypotheses, a production experiment and a perception experiment were conducted. Experiment I is a production test which intends to examine the accuracy with which native Shanghainese speakers are able to produce SM nasal codas. Experiment II is a perception test which intends to reveal the accuracy with which these speakers are able to perceive the SM nasal codas. The results of the two experiments are then compared to investigate the relationship between production and perception of SM nasal codas by Shanghainese speakers.

### **3 Description of the Experiments**

#### **3.1 Experiment I: Production**

##### *3.1.1 Participants*

Fourteen native Shanghainese speakers (7 females and 7 males) were tested, ranging in age from 19 to 36. All participants were born and raised in Shanghai and their parents were also both native Shanghainese. All subjects spoke Shanghainese with their parents, and were educated in SM since kindergarten. Thirteen subjects were college students and one was a university teacher at the time of the study. Their educational background ensures that they were bi-dialectal and that they were fully capable of recognizing the Pinyin transcriptions for the words they heard or read in the questionnaire.

### 3.1.2 Materials

The questionnaire consisted of two types of reading-aloud exercises: monosyllabic words and bi-syllabic or tri-syllabic words. The mono-syllabic words were represented in Chinese Pinyin instead of orthographic characters, so as to ensure that when errors happened, they should be attributed to the participants' inability to distinguish between two nasal coda phonemes instead of the participants' failure to recall the correct Pinyin representation of the word, or their lack of knowledge of the correct pronunciation of the orthographic character. For example:

- (4)      can          cang  
          man          mang<sup>7</sup>

The bi- and tri-syllabic words in the questionnaire were represented in orthographic characters, so that it would be easy for the participants to recognize and read them. Syllables with /in, iŋ, ən, əŋ/<sup>8</sup> rhymes are specified for corresponding orthographic characters so that when errors occur, they are attributed to the inability to pronounce the rhymes accurately, instead of their failure in recalling the correct Pinyin representations. Besides, the bi- and tri-syllabic words are grouped according to the place feature of the onset consonant of the syllable following the target coda nasal, which results in seven groups: the nasal + bilabial /p, p<sup>h</sup>, m/ group, the nasal + alveolar /t, t<sup>h</sup>, n, l/ group, the nasal + velar /k, k<sup>h</sup>, x/ group, the nasal + palatal /tɕ, tɕ<sup>h</sup>, ɕ/ group, the nasal + dental /ts, ts<sup>h</sup>, s/ group, the nasal + retroflex /tʂ, tʂ<sup>h</sup>, ʂ, z/ group, and the nasal + zero onset /ø, j, w/ group.<sup>9</sup> The nasal coda phonemes /n, ŋ/ are realized as [m], [n], [ŋ], [ɲ], [n]<sup>10</sup>, [ŋ], and a variation with a glide realization, respectively (Zheng & Liu, (2005), Zheng & Bao (2003)). For each of the seven groups, all eight pairs of MC nasal rhymes (see above) were included. But since /uən/ and /ən/ share the nuclear vowel and the nasal coda, i.e., the major part of the rhyme, and differ only in the absence of the glide or not, they are considered to be one instead of two rhymes. Thus, the number of MC nasal rhymes is actually 15. Therefore, the questionnaire contains a total of 7 \* 15 = 105 bi- and tri-syllabic words grouped according to the variations of the nasal codas.

<sup>7</sup> The corresponding IPA transcriptions for these Pinyin representations are /ts<sup>h</sup>an/, /ts<sup>h</sup>aŋ/, /man/, and /maŋ/ respectively. Their meanings vary with different tones.

<sup>8</sup> These rhymes are known to be difficult to distinguish for non-native speakers of SM in both perception and production.

<sup>9</sup> The labiodental /f/ is not included because it is not a common sound and overlaps with the bilabial consonantal onsets in terms of the place feature.

<sup>10</sup> The dental nasal and the alveolar nasal are represented by the same symbol in the IPA.

### 3.1.3 Procedures

The participants read the materials aloud clearly at a medium speed one by one and recordings were made for each of them. The recordings were made in Praat (Boersma & Weenink 2010) and a headphone with a microphone in a quiet office room in Shanghai International Studies University (SISU) and were later evaluated by two native SM speakers, including the one recorded for the dictation test. An error was scored when both of the native SM speakers agreed that the nasal coda was produced otherwise. In the case of disagreement, the author acted as the third party and made the decision as to whether there was an error.

### 3.1.4 Results and discussion

The results were presented as the number and percentage of errors that the participants made in the experiment in Table 1. In this table, the number in a cell indicates the number of errors made by all the 14 participants together concerning a specific nasal rhyme; shaded cells indicate that no errors were made. The percentages were obtained by dividing the number of errors by the total number of syllables containing a specific rhyme, multiplied by the total number of participants, i.e., 14. The percentage number can be understood as the average number of errors made by one participant for every 100 monosyllabic or multisyllabic words, and the numbers can be compared to see in which case the nasal coda is more difficult to produce. For example, Table 1 shows that altogether 4 errors were made by the participants for the /in/ rhyme in word-final position. Since there are 2 syllables containing [in] in word-final position, the error percentage is  $4/(2 \times 14) \times 100\% = 14.29\%$ . In the same vein, the total number of errors made by the participants concerning the same rhyme in word-medial position is 21, and there are 7 syllables containing [in]. Therefore, the error percentage in this case is  $21/(7 \times 14) \times 100\% = 21.43\%$ . Comparing the two percentages, we can draw the conclusion that the rhyme /in/ in word-medial position is slightly more difficult for native Shanghainese speakers than in word-final position.

Table 1 shows that the errors are highly regular: an overwhelming majority of errors were made for the nasal codas following the /i/ and /ə/ nuclear vowels. Sporadic errors were made when the nasal codas follow the nuclear vowel /y/. No errors were made when the nasal phonemes follow other nuclear vowels. To put it another way, MC nasal rhymes /in/-/iŋ/ and /ən/-/əŋ/ are difficult for native Shanghainese speakers to produce accurately, while for other rhymes the subject performed at ceiling level. It also shows that the alternation of the nasal codas into its contextual variants in word-medial position due to regressive place assimilation, i.e. /n/ and /ŋ/ to [m, ŋ, n, ŋ, ɲ, ɲ] triggered by the

following onset consonants, doesn't seem to affect the overall performance of the participants. This is supported by the following three facts: (i) for those rhymes where no errors were made when they are in word-final position, contextual variants of the nasal coda in word-medial position did not cause higher error percentages; (ii) for the /in/, /iŋ/ and /əŋ/ rhymes, more errors were made in word-medial position than in word-final position, while for the /ən/ rhymes, more errors were made in word-final position than in word-medial position; (iii) although there were remarkably fewer errors for /ən/ rhymes when they were followed by coronal consonants (when /n/ is realized exactly as [n]) than when they were followed by other consonants, no such regularity is found for other rhymes.

Table 1. Numbers of errors of nasal codas in word-final and word-medial positions in production.

Rhyme		a		ia		ua		ya		i			ə		uə		y		u	
		n	ŋ	n	ŋ	n	ŋ	n	n	ŋ	n	ŋ	n	ŋ	n	ŋ	n	ŋ	n	ŋ
<b>Speech-final</b>									4	16	13	8								
<b>Percentage</b>									<b>14.3</b>	<b>51.1</b>	<b>46.4</b>	<b>28.6</b>								
<b>Speech-medial</b>	<b>Total</b>								21	54	15	47								
									<b>21.4</b>	<b>55.1</b>	<b>15.3</b>	<b>48.0</b>								
	+labial								3	8		10								
	+alveolar								3	9		7								
	+velar								3	7	6	3								
	+palatal								2	8	4	11								
	+dental								2	8		12								
	+retroflex								4	7		3								
+zero								4	7	5	1									

(Note: In the questionnaire, there are 28 mono-syllabic words in which the nasal coda is in word-final position, and 105 bi- and tri-syllabic words in which the nasal coda is in word-medial position.)

Table 1 shows that the errors are highly regular: an overwhelming majority of errors were made for the nasal codas following the /i/ and /ə/ nuclear vowels. Sporadic errors were made when the nasal codas follow the nuclear vowel /y/. No errors were made when the nasal phonemes follow other nuclear vowels. To put it another way, MC nasal rhymes /in/-/iŋ/ and /ən/-/əŋ/ are difficult for native Shanghainese speakers to produce accurately, while for other rhymes the subject performed at ceiling level. It also shows that the alternation of the nasal codas into its contextual variants in word-medial position due to regressive place assimilation, i.e. /n/ and /ŋ/ to [m, ŋ, n, ŋ, ɲ, ɲ] triggered by the following onset consonants, doesn't seem to affect the overall performance of the participants. This is supported by the following three facts: (i) for those rhymes where no errors were made when they are

in word-final position, contextual variants of the nasal coda in word-medial position did not cause higher error percentages; (ii) for the /in/, /iŋ/ and /əŋ/ rhymes, more errors were made in word-medial position than in word-final position, while for the /ən/ rhymes, more errors were made in word-final position than in word-medial position; (iii) although there were remarkably fewer errors for /ən/ rhymes when they were followed by coronal consonants (when /n/ is realized exactly as [n]) than when they were followed by other consonants, no such regularity is found for other rhymes.

A closer look reveals that the position of the SM nasal codas may have an effect on their production accuracy. As can be seen from Table 1, the [in]-[iŋ] rhymes are slightly easier to produce accurately in word-final position than in word-medial position. But for the [ən]-[əŋ] rhymes, things are a bit complicated: in word-final position, [ən] seems to be significantly more difficult to produce accurately than [əŋ], while in word-medial position, the converse is true. This shows that in production the consonantal onsets that follow the SM nasal codas seem to have more influence on [ən]-[əŋ] rhymes than on [in]-[iŋ] rhymes.

### 3.2 Experiment II: Perception

The perception experiment intended to examine the characteristics of the native Shanghainese speakers' perception of SM nasal codas. It was carried out immediately after the production experiment for each participant. The participants taking part in the perception test were the same as those taking part in the production test, but the materials and procedures were different.

#### 3.2.2 Materials

The questionnaire was composed of two parts: the first one tested the perception of mono-syllabic words, and the second one perception of bi- and tri-syllabic words. All items were presented in multiple-choice form with only two choices: A or B. Both the mono-syllabic and the multi-syllabic words were presented in Chinese Pinyin instead of orthographic characters so as to ensure that an incorrect choice would be due to the fact that the participant was unable to distinguish between the nasal coda phonemes in perception, and not because the participant lacked the knowledge of the correct pronunciation of the orthographic character. For example:

- (5)      ban          bang  
          fang.men    fan.men

There were altogether 16 questions in the first part, covering the 16 rhymes in SM. The second part was similar to that in the production experiment in that the questions were grouped according to the

place feature of the onset consonant of the syllable following the target coda nasal, resulting in seven groups, repeated here as follows: the nasal + bilabial /p, p<sup>h</sup>, m/ group, the nasal + alveolar /t, t<sup>h</sup>, n, l/ group, the nasal + velar /k, k<sup>h</sup>, x/ group, the nasal + palatal /tɕ, tɕ<sup>h</sup>, ɕ/ group, the nasal + dental /ts, ts<sup>h</sup>, s/ group, the nasal + retroflex /tʂ, tʂ<sup>h</sup>, ʂ, z/ group, and the nasal + zero onset /ø, j, w/ group. Each group had 15 A/B choice questions, so the total number of the words was 7×15 = 105 (see also 3.1.2).

### 3.2.3 Procedure

Three native MC speakers were invited to record for the dictation prior to the test, and only one of the recordings was selected. The other two were deleted either because the speaker had a strong northern accent or because there was too much background noise in the recording. Every word was read only once, with one or two second intervals in between for the listeners to make choices. The participants were tested individually. They listened to the recording with a headphone connected to the computer. All recordings were made via Praat and a headphone with a microphone in a quiet office room in Shanghai International Studies University (SISU). All answers were forced-choice, either A or B.

### 3.2.4 Results and Discussion

The questionnaires were collected and evaluated by the present author alone. The results will be presented as numbers and percentages of incorrect answers that the participants gave in the experiments. In the table below, the number in the cell indicates the number of incorrect answers given by all 14 participants combined for a specific nasal rhyme, and the shaded cells indicate that all the answers were right. The percentage is obtained by dividing the number of incorrect answers by the total number of the syllables containing that specific rhyme, multiplied by the total number of the participants, i.e., 14. The percentage can be understood as the average number of incorrect answers given by one participant for every 100 mono-syllable words or multi-syllable words, and the numbers can be compared to see in which cases the nasal coda was more difficult to perceive accurately. For example, the table shows that there are altogether 2 incorrect answers given by the participants concerning the /in/ rhyme in word-final position, for one test syllable. Thus, the percentage is  $2/(1 \times 14) \times 100\% = 14.29\%$ . Similarly, the total number of errors made by the participants concerning the same rhyme in word-medial position is 30, and there are 7 syllables containing [in] in word-medial position. Thus, the error percentage in this case is  $26/(7 \times 14) \times 100\% = 26.53\%$ . Comparing the two percentage numbers, the conclusion can be drawn that the rhyme /in/ in word-medial position is significantly more difficult for native Shanghainese speakers to perceive accurately than it is in word-final position.

Table 2. Number of incorrect answers of nasal codas in word-final and word-medial position in perception.

Rhyme		a		ia		ua		ya		i		ə		uə		y		u	
		n	ŋ	n	ŋ	n	ŋ	n	ŋ	n	ŋ	n	ŋ	n	ŋ	n	ŋ	n	ŋ
<b>Speech-final (Percentage)</b>										2 <b>14.3</b>		2 <b>14.3</b>	4 <b>28.6</b>		2 <b>14.29</b>	1 <b>7.14</b>			
<b>Speech-medial</b>	<b>Total (Perc)</b>									26 <b>26.5</b>	12 <b>12.2</b>	18 <b>18.4</b>	11 <b>11.2</b>						1 <b>1.02</b>
	+labial									3	2	2	1						
	+alveolar									3		3	1						
	+velar									1		1	6					1	
	+palatal									2	2								
	+dental									9	1	4							
	+retroflex									6	4	2	2						
	+zero									2	3	6	1						

(Note: In the questionnaire, there are 16 multiple choice questions on mono-syllabic words in which the nasal coda is in word-final position, and 105 on bi- and tri-syllabic words in which the nasal coda is in word-medial position.)

The results are similar to those obtained in the production experiment in that among all the 8 pairs of SM nasal rhymes, the /in/-/iŋ/ rhymes and the /ən/-/əŋ/ rhymes are most difficult for the NSs to produce accurately, because most errors were made concerning these rhymes. They are also similar in that the position of the nasal coda in the speech does not influence the performance of the NSs in perception. This is supported by the following two facts: (i) for those rhymes where no incorrect answers were given when they are in word-final position, contextual variants of the nasal coda in word-medial position did not cause any more incorrect answers; (ii) although results show that there are fewer incorrect answers for the /in, iŋ, ən/ rhymes in word-final position than in word-medial position, there are significantly many more errors for the /əŋ/ rhyme in word-final position than in word-medial position.

The results obtained in the two experiments justifies the two hypotheses made in Section 2, which are repeated here as follows: (i) native Shanghainese speakers have difficulty in distinguishing between the SM nasal codas, but the difficulty is different for different nuclear vowels; (ii) Context-sensitive phonetic variations of SM nasal codas in word-medial position due to place assimilation do not influence the overall performance of native Shanghainese speakers in perception.

### 3.3 A comparison

In order to better understand the correlation between production and perception, let us compare the results obtained in the two experiments, summarized in the two tables below. Table 3 is a comparison of the results obtained from the production and perception experiments for the nasal codas in both word-final and word-medial position. The number in the cell indicates the numbers of errors made, or incorrect responses, by the 14 participants combined. The percentage is obtained by dividing the number in the corresponding cell by the total number of questions in the questionnaire, multiplied by 14. For example, for the /in/ rhyme, the participants made 25 errors in the production test in total. There are altogether  $2+7 = 9$  syllables containing this rhyme in the production questionnaire. Therefore, the percentage is  $25/(9 \times 14) \times 100\% = 19.84\%$ . That is to say, on average, every participant made nearly 20 errors in production for every 100 questions. In the perception experiment, the participants gave 28 incorrect answers concerning this rhyme. There are altogether  $1+7 = 8$  syllables that tested this rhyme in the perception questionnaire. Therefore, the percentage is  $28/(8 \times 14) \times 100\% = 25\%$ . This means that, on average, every participant gave 25 incorrect answers for every 100 syllables. Comparing the two percentages, we can draw the conclusion that the participants performed better in the perception of /in/ rhyme than they did in production. Table 4 is a comparison of the results obtained from the production and perception experiments for individual participants. The number in each cell indicates the percentage number of errors, or incorrect answers, made by the participant concerning a specific rhyme. The percentage is obtained by dividing the number of errors or incorrect answers by the total number of questions concerning a certain rhyme in the questionnaire. For example, Participant 1 made eight errors concerning [in] in production experiment, and the number of questions in the production questionnaire concerning [in] is  $2+7=9$ . Therefore, the percentage number of errors in the corresponding cell is  $8/9 \times 100\% = 88.9\%$ . In contrast, Participant 1 gave two incorrect answers concerning [in] in perception experiment, and the number of questions concerning [in] in perception questionnaire is  $1+7=8$ . Therefore, the percentage number of incorrect answers in the corresponding cell is  $2/8 \times 100\% = 25\%$ .

Table 3. Comparison between production and perception of SM nasal codas by all participants.

Rhyme	a		ia		ua		a	i		ə		uə		y		u	
	n	ŋ	n	ŋ	n	ŋ	n	n	ŋ	n	ŋ	n	ŋ	n	ŋ	n	ŋ
<b>Production</b>								25	70	28	55						
<b>(Percentage)</b>								<b>19.8</b>	<b>55.6</b>	<b>22.2</b>	<b>43.7</b>						
<b>(Percentage) Perception</b>								<b>25</b>	<b>10.7</b>	<b>17.9</b>	<b>13.4</b>	<b>1.79</b>	<b>.89</b>	<b>.89</b>			
								28	12	20	15	2	1	1			

As Table 3 shows, no mistakes were made for the /an, aŋ, ian, iaŋ, uan, uaŋ, yan, yŋ, un, uŋ/ rhymes either in the production or the perception experiment. For the /iŋ, əŋ, əŋ, yŋ/ rhymes, the percentages of incorrect answers in perception is lower than that for production, but for the /in, uən, uəŋ/ rhymes, the percentage of incorrect answers in perception is higher than that for production.

Table 4. Results for individual participants (in percentages only).

Rhyme		a		ia		ua		ya	i		ə		uə		y		u	
		n	ŋ	n	ŋ	n	ŋ	n	n	ŋ	n	ŋ	n	ŋ	n	ŋ	n	ŋ
1	Prod.								88.9		11.1	77.8						
	Perc.								25	37.5		12.5						
2	Prod.								22.2	66.7		11.1						
	Perc.																	
3	Prod.									88.9		55.6						
	Perc.									12.5	12.5	12.5						
4	Prod.									100	22.2	55.6						
	Perc.								37.5	25	37.5							
5	Prod.									100	33.3	33.3						
	Perc.								12.5	12.5	62.5	37.5	12.5					
6	Prod.										44.4	22.2						
	Perc.								50	25		25						
7	Prod.										33.3	22.2						
	Perc.																	
8	Prod.								22.2	77.7	33.3	66.7						
	Perc.								25	25								
9	Prod.								33.3	11.1	33.3	33.3						
	Perc.								25			37.5	12.5					
10	Prod.									88.9	11.1	77.8						
	Perc.								25	25		12.5						
11	Prod.								11.1		11.1	44.4						
	Perc.								37.5	12.5	50							
12	Prod.								66.7	11.1	22.2	22.2						
	Perc.								12.5		12.5	25						
13	Prod.								11.1	100	11.1	55.6						
	Perc.								25			12.5						
14	Prod.								33.3	100	44.4	33.3						
	Perc.								75		75	12.5			12.5		12.5	

Table 4 is a detailed record of the errors made and incorrect answers given by individual participants. It shows the same regularity as Table 3, i.e., incorrect answers or errors are concentrated in the areas for /in, iŋ, ən, əŋ/ rhymes, while for other rhymes, few incorrect answers or errors are found. It is interesting to see that some participants (e.g., participants 2 and 7) did all right in the perception test, but made some errors in production, while some participants did better in the production test than in perception, such as Participant 6 did for the /in, iŋ/ rhymes.

#### 4 General Discussion

The perception experiment results shows that native Shanghainese speakers can accurately perceive the SM nasal codas [n] and [ŋ] in [an, aŋ, ian, iaŋ, uan, uaŋ, yŋ, uŋ] rhymes, but they have difficulty in perceiving them in [in, iŋ, ən, əŋ, uən, uəŋ, yn] rhymes. This result confirms our first prediction made in Section 2. It means that the native Shanghainese speakers have successfully established new phonetic categories for [n] and [ŋ] in the coda position following the nuclear vowels [a] and [u], but fail to do so for [n] and [ŋ] in the coda position following the nuclear vowels [i] and [ə], while [y] falls somewhere in between.

The individual performances recorded in Table 4 shows that some participants, such as 1, 3, 4, 5, 8 and 10, confused SM [n] and [ŋ] in perception, and made roughly the same number of errors for the [in]-[iŋ] or [ən]-[əŋ] rhyme pair in production and perception, while some participants, such as 9, 11, 12, 13, and 14, did well in perceiving one of the two SM nasal codas but failed in doing so for the other. This result confirms the hypothesis we made about the specific types of difficulties the native Shanghainese speakers may come across in perceiving the two SM nasal codas (see section 2).

As to the relationship between production and perception, the results show that for the SM rhymes /an, aŋ, ian, iaŋ, uan, uaŋ, yan, yŋ, uŋ/, native Shanghainese speakers' production correlates well with perception, i.e., both production and perception are excellent. However, as Table 3 shows, for the SM rhymes /in, iŋ, ən, əŋ, uən, uəŋ, yn/, the relationship between production and perception varies. For the /iŋ, ən, əŋ/ rhymes, the percentage of errors in production is higher than that in perception. For the /u(ə)n, uəŋ/ rhymes, the percentage rate of incorrect answers in perception seems higher than in production, but if we regard these as part of the /ən, əŋ/ rhymes, the percentage of incorrect answers in perception is still lower than that in production. That is to say, in [iŋ], [ən] and [əŋ] rhymes, native Shanghainese speakers' performance in perception is better than it is in production, which naturally follows from the SLM's prediction that good production entails good perception. However, for the /in/ rhyme, Table 3 shows that the percentage rate of errors in production is lower than in perception,

which means that native Shanghainese speakers' production of the SM /in/ rhyme is better than their perception. Similar results were found for individual participants (see Table 4). For example, participant 6 did not make any errors for the SM rhymes /in/ and /iŋ/ in production, but gave 4 and 2 incorrect answers, respectively, in perception. Participant 11 gave 3 and 4 incorrect answers for the /in/ and /ən/ rhymes in perception, respectively, but only made one error for the two rhymes in production. These facts contradict the hypothesis we made in Section 2 that good production entails good perception. The reason may be that although the native Shanghainese speakers have established new phonetic categories for SM [n] and [ŋ] as codas of the [i] and [ə] nuclei, the categories are based on different features, or "feature weights" (Flege 1995:239) from those of the native SM speakers. In the case of Participant 6, one possibility for the mismatch is that she may have established new phonetic categories for SM [n] and [ŋ] as codas of the [i] and [ə] nuclei, and even categorized them for the accurate place features, but attached more undue weight to their place feature than the native SM speakers do, so that she is not able to discard redundant phonetic information in their surface forms and accurately tell the underlying forms. In this case, the native speakers can accurately produce the SM nasal codas because the nasal codas are categorized for the accurate feature, but they cannot perceive them accurately because they cannot discard redundant information. Participant 11 seems to have a similar problem as Participant 6. That is, the two participants processed the SM nasal codas on the more specific phonetic level, rather than the allophonic level.

As to Participants 2 and 7, who are perfect in perception but erroneous in production, it can be concluded that they have established accurate phonetic categories for the SM nasal codas following the [i] and [ə] nuclei. However, they cannot control their speech organs to produce them accurately, which may be attributed to a lack of practice or to a lack of care in accurate production.

Last but not the least, why is it that the nasal codas in /an, aŋ, ian, iaŋ, uan, uaŋ, yn<sup>11</sup>, yŋ, uŋ/ rhymes are so easy for native Shanghainese speakers to perceive and produce that their performances are immaculate? It cannot be explained by their L1 background alone. An L2 linguistic perspective can shed better light.

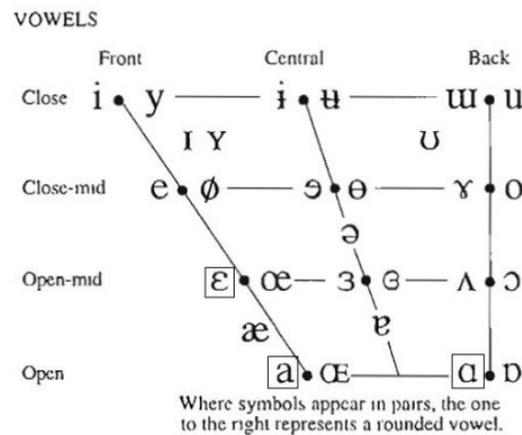
It is well-known that in SM, there are only six vowel phonemes: a, o, ɤ, i, u, y, (e.g. Yuan (1983) and Duanmu (2008))<sup>12</sup>, each of which has allophones. For example, the low vowel /a/ has at least three

<sup>11</sup> Only sporadic errors were found for the /yn/ rhyme, so we take it to be not difficult for the Shanghainese native speakers to perceive and produce it.

<sup>12</sup> According to Lee & Zee (2003), the vowel phonemes in Standard Chinese (SC) are /ɿ, y, a, ə, u, ʏ/. The issue is not relevant in this paper.

allophones<sup>13</sup>: [ɑ], [a] and [ɛ]. The surface forms of the underlying vowel /a/ are decided by the place features of the nasal codas that follow them. When it precedes the velar nasal /ŋ/, it surfaces as the back vowel [ɑ]; when it precedes the coronal nasal /n/, it surfaces as the front vowel [a]; when it is between the high front glide [j] and the coda /n/, it surfaces as the mid front vowel [ɛ]. The following vowel chart illustrates the three allophones:

(7)



The allophones of /a/ are circled in the squares in the vowel chart. The corresponding rewrite rules for these changes are indicated as follows:

- (8)
- i.  $a \rightarrow \alpha / \_ \eta$
  - ii.  $a \rightarrow a / \_ n$
  - iii.  $a \rightarrow \varepsilon / j\_n, y\_n$

The reason for the alternation of the vowel is place assimilation. For example, in (8i), the [+back] feature of the velar nasal [ŋ] is assimilated to the preceding vowel /a/, which causes it to change into the allophone [ɑ]; in (8ii), the [-back] feature of the coronal nasal [n] agrees with the preceding nuclear vowel [a]; and in (8iii), the [-back] feature of the pre-nuclear glide makes the [-back] feature of the nuclear vowel more prominent so that the underlying /a/ changes into [ɛ] on the surface. That is to say, assimilation makes MC nasal codas with the nuclear vowel /a/ phonetically easier to perceive and produce. This is also true with the /yn/-/yŋ/ rhyme pair and the /un/-/uŋ/ rhyme pair. In the former case, the /yŋ/ rhyme surfaces as [juŋ] because the [+back] feature of the nasal coda /ŋ/

<sup>13</sup> Some scholars suggest that /a/ has more than three allophones because they consider the /a/ in the /ia/, /ai/, and /au/ as different allophones of /a/. I also believe that the /a/ in the rhymes /uan/ and /uaŋ/ are different phonemes. Since this question doesn't concern us here, we ignore it in this paper.

assimilates to the nuclear vowel /y/, which is [-back]. This process results in the split of [y] into [iu], in which [i] is [-back] while [u] is [+back]. In the latter case, the /un/ rhyme surfaces as [uən], because the [-back] feature of the nasal coda /n/ assimilates to the nuclear vowel /u/, which is [+back]. This process results in the insertion of the schwa [ə], which is [-back], in between [u] and [n]. The rules involved here can be written as follows:

- (9) Vowel split:            y    →    iu / \_ŋ  
 Vowel epenthesis:        Ø    →    ə / u\_n

The nuclear vowels are the same whether the nasal codas are in word-medial position or in syllable- or word-final position. They serve as the cues for the native speakers to perceive and producing SM nasal codas. That is to say, the success of the speakers in perceiving and producing the nasal phonemes in SM may be partly attributed to their success in perceiving and producing the nuclear vowels that precede the nasal phonemes.

Things are a bit different when we look at the nasal codas in the /in/-/iŋ/ rhyme pair and the /ən/-/əŋ/ rhyme pair. There seems to be no change on the part of the nuclear vowels in these rhymes, although the /ə/ in the rhyme /əŋ/ is closer to [ɜ], which is lower and more back in the vowel chart than [ə]. But the feature matrix for [ɜ] is the same as that of [ə], both of which are [+back, -low, -round]. That is to say, the change of [ə] to [ɜ] is phonetically slight. This may cause difficulties for native Shanghainese speakers to perceive and produce the nasal codas after this vowel. As for the nasal codas in the /in/-/iŋ/ rhyme pair, place assimilation also works. According to the native SM speakers in this experiment, the /iŋ/ rhyme surfaces as [iəŋ].<sup>14</sup> The rule involved in this process can be written as:

- (10)    Ø    →    ə / i\_ŋ

The epenthetic [ə] makes the /iŋ/ rhyme easier to distinguish from the /in/ rhyme. The difficulty in perceiving and producing the nasal phonemes in the /in/-/iŋ/ rhyme pair by the NSs may be attributed to their failure to perceive the epenthetic [ə] in the /iŋ/ rhyme.

This explanation justifies the SLM hypothesis that the L<sub>1</sub> and L<sub>2</sub> sounds are perceptually related at a “position-sensitive” allophonic level. The reason is that the linguistic environment can provide cues for better perception.

<sup>14</sup> But in this case, the [i] is different from the pre-nuclear glide [j]. It is longer and more emphatic than the epenthetic [ə], which is short and fuzzy.

## 5 Conclusion

This study has found that the nasal codas in Shanghainese and SM are related to one another at the allophonic level, neither at the abstract phonemic level nor at the context-sensitive phonetic level. Besides, it also found that L2 linguistic environments can be helpful for the accurate perception of the SM nasal codas, i.e., the nucleus may serve as the phonetic cues for the native speakers to perceive the nasal codas. The two findings confirmed the SLM hypothesis that sounds in the L1 and L2 are related perceptually to one another “at a position-sensitive allophonic level”. This study also found that native Shanghainese speakers’ production of the SM nasal codas does not correlate with their perception very well. This fact confirms the SLM prediction that production and perception may not be perfectly aligned, even for highly experienced speakers of an L2 (Flege, 1999). In conclusion, Flege’s SLM, though a model on second language acquisition, can effectively explain the perception and production difficulties encountered by Chinese bi-dialectal speakers.

The advantage of this study is that the materials used in the experiments cover all the possible positions for SM nasal codas to appear, i.e. both word-finally and word-medially, and all the possible context-sensitive phonetic variations. The limitations of this study is that the participants of the experiments are mostly well-educated young Shanghainese speakers; if the subjects could be more varied in background, the results may be more revealing. Further studies can draw on the advantage of the present study while improving on its limitations.

## Acknowledgements

This study is subsidized by Chinese Scholarship Council (CSC) and the Young Scholars' Innovative Teamwork Project (No. 2011114025) of Shanghai International Studies University (SISU). I also want to thank Marjoleine Sloos for her gracious help to me and enlightening comments on this paper, and thank Jeroen van de Weijer for his review and correction of mistakes in this paper. My sincere thanks also go to Chen Wenjun for her review of this paper, and the anonymous reviewers for their helpful comments.

## References

- Boersma, Paul., & Weenink, David. 2010. *Praat: Doing phonetics by computer* [computer program]. Downloaded from [www.praat.org](http://www.praat.org).
- Duanmu, San. 2008. *Syllable structure: The limits of variation*. New York: Oxford University Press.

- Flege, James. E. 1995. Second language speech learning: Theory, findings, and problems. In William Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* 233-277, Timonium: York Press.
- Flege, James. E. 1999. The relation between L2 production and perception. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville and A. Bailey (Eds.), *XIVth International Congress of Phonetics Sciences*. Berkeley, CA: Department of Linguistics, Univ. of California at Berkeley, 1273-1276.
- Harnsberger, James. D. 2001. On the relationship between identification and discrimination of non-native nasal consonants. *The Journal of the Acoustical Society of America*, 1101, 489-503.
- Lee, Wai-Sum., & Zee, Eric. 2003. Standard Chinese (Beijing). *Journal of the International Phonetic Association*, 33.1, 109-112.
- Luo, Mingqiong. 2014. *Chinese syllable structure---an X-bar approach*. PhD dissertation, Shanghai International Studies University.
- 顾钦 [Gu, Qin]. 2007. 语言接触对上海市区方言语音演变的影响 [The influence of language contact on the phonetic evolution of the dialects in Shanghai urban districts]. Shanghai Normal University, PhD dissertation.
- 钱乃荣 [Qian, Nairong]. 1987. 上海方言音变的微观 [A closer look at sound changes in Shanghai dialect]. 语言研究 [Linguistic Research], 2, 104-115.
- 钱乃荣等 [Qian, Nairong]. 2007. 上海话大辞典 (辞海版) [Dictionary of Shanghai dialect]. Shanghai: 上海辞书出版社 [Shanghai Lexicographic Publishing House].
- Yuan, Jiahua. 1983. *A concise book on Chinese dialects*. Beijing: Beijing Word Reform Publishing House.
- 杨蓓 [Yang, Bei]. 2002. 上海儿童的上海话音系习得 [A study on the phonological acquisition of local Shanghainese children]. PhD dissertation, Fudan University.
- 游汝杰 [You, Rujie]. 2013. 上海地区方言调查研究 [An investigation of the sub-dialects in Shanghai]. Shanghai: 上海: 复旦大学出版社 [Fudan University Publishing House].
- 郑玉玲、鲍怀翘 [Zheng, Yuling & Bao, Huaiqiao]. 2003. 论普通话/N<sub>1</sub>C<sub>2</sub>/的协同发音 [On the co-articulation of N<sub>1</sub>C<sub>2</sub> sequences in Mandarin]. 《第六届全国现代语音学学术会议论文集》 [Anthology of the Sixth National Modern Phonetics Symposium], 37-43.
- 郑玉玲、刘佳 [Zheng, Yuling & Liu, Jia]. 2005. 普通话N<sub>1</sub>C<sub>2</sub>C#C协同发音的声学模式 [The phonetic features of N<sub>1</sub>C<sub>2</sub>C#C Co-articulation in Putonghua]. 《南京师范大学文学院学报》 [Journal of the Arts College in Nanjing Normal University], 3, 61-66.
- 朱晓农 [Zhu, Xiaonong]. 1996. 上海音系 [Shanghainese phonology]. 《国外语言学》 [Overseas Linguistics], 2, 29-37.

# Oral Fluency in the Mandarin Chinese Speech of Uygur Learners

Xinlu Yang

*College of Foreign Languages, Xinjiang Normal University*

*School of Foreign Languages, Tongji University*

## Abstract

The principal dimensions of L2 performance and L2 proficiency can be comprehensively captured by the notions of complexity, accuracy and fluency. This study investigates speech samples collected from 30 Uygur learners at two distinct level of proficiency (primary and intermediate). The two groups of students were compared with 10 native speakers of Mandarin Chinese. For all the native and nonnative speakers, speech rate, unfilled and filled pauses, the mean length of runs, repeats and self-correction, which were the best predictors of productive fluency, were investigated. The results indicate that there were significant differences of the temporal variables and speech management strategies in Mandarin spontaneous speech by Uygur learners and native speakers. In addition, gender and the task type affected the temporal variables and speech management strategies in Mandarin spontaneous speech by Uygur learners. The level of language proficiency only affected the repeat frequency by Uygur learners. These results provide important pedagogical implications for Uygur learners of Mandarin Chinese.

**Keywords:** Uygur, oral fluency, Mandarin Chinese.

## 1 Introduction

Speaking is considered to be one the most complex human faculties, as it involves motor, linguistic as well as cognitive skills (Levelt 1993). A speaker's general proficiency is normally referred to as "fluency". This is a commonly used notion in foreign language teaching with different definitions. Chambers (1997) considers qualitative and quantitative aspects of fluency. Qualitative aspects of fluency are associated with language mastery and native-like performance.

Quantitative aspects are empirically identifiable and quantifiable temporal variables. On the contrary, Brumfit (1984) defines fluency as “natural language use whether or not it results in native-speaker-like language comprehension or production”. Lennon (1990) and Schmidt (1992) emphasize that fluency is a performance phenomenon, a skill exercised in real time. Schmidt (1992) focuses on the processing and production of spoken language in real time and defines fluency as a “primarily temporal phenomenon”, stating that “fluency in speech production is an automatic procedural skill”. For foreign language learners, much of the grammatical encoding becomes automatic in a gradual process.

In linguistic research, the predominant approach to fluency is the measurement of temporal variables in speech. The description of these variables supports a ‘fluency equals automaticity’ approach, which include quantifiable markers being (a) speech rate: the speed at which an individual executes articulatory movements for speech production (e.g. Crystal & House (1990)). Two common measures are distinguished: speech rate and articulation rate. Speech rate is measured as the number of syllables produced in a speech sample divided by the time needed for completing the speech sample (e.g. Cotton (1936)). Articulation rate is measured as the number of syllables produced in a timed speech sample following removal of silent intervals from the samples (e.g. Crystal & House (1990)); (b) mean length of runs: the amount of speech they utter between pauses and can “reflect a word, a phrase, a sentence or a series of sentences depending on the task and the rate of output” (Grosjean 1980). It indicates the level of automaticity of a speaker, i.e. the “level of routinization of knowledge representation” in speech as well as the “level of access to all the syntax and lexis the speaker controls” (Towell 2002); (c) unfilled pauses: silence or the occurrence of non-speech acoustic events such as breathing and noise (Gut 2009). These variables occur in each speaker’s output and therefore count as primary variables in the description of productive fluency.

Apart from temporal variables, frequency of occurrence of filled pauses, repetitions and self-corrections are also regarded as key fluency markers (Lennon 1990). They are called speech management strategies, which are useful tools for productive fluency enhancement (Foster &

Tavakoli 2009). Filled pauses consist of “non-lexical fillers such as ‘uh’ and ‘erm’ and elongations of sounds” (Götz 2013). Repeats refer to a word or words uttered repetitively during speech (Siegman & Feldstein 1979). Self-correction is defined as “the speaker retraces (or notionally ‘erases’) what has just been said, and starts again, this time with a different word or sequence of words” (Biber et al. 1999).

Temporal variables and speech management strategies of Mandarin Chinese learners’ oral fluency have been investigated by many researchers (e.g. Chen (2008, 2012), Dong (2011)). The results indicate that temporal variables like unfilled pauses and the mean length of runs by second language learners of Chinese are significantly different from those found in native speakers. They also discover that the temporal variables, especially the duration of unfilled pauses, are affected by many factors like the level of language proficiency and gender. For some of these variables, like the mean length of runs, second language learners may reach a similar level as native speakers. In these studies, the subjects are limited to foreign learners of Mandarin Chinese. Much less research has been conducted on the oral fluency of China’s minority learners of Mandarin Chinese (e.g. Uygur speakers), especially speakers with elementary Mandarin Chinese proficiency. Uygur belongs to Turkic branch of Altaic language family. It is a non-tone language spoken in Xinjiang Uygur Autonomous Region of western China, which displays a fixed stress pattern (Zhao & Zhu 1985).

A model which relates to the temporal variables of fluency is the Adaptive Control of Thought Model (Anderson 1983). Anderson proposes two types of knowledge: declarative knowledge (i.e. knowledge of abstract facts and data) and procedural knowledge (i.e. knowledge of doing certain things repeatedly until the production becomes automatic. Fluent speech would reflect the use of procedural knowledge. Hesitant speech would be indicated by the use of declarative knowledge. In the present study, speech data is analyzed to see the relationship between the declarative knowledge and procedural knowledge.

To deal with the general concerns of fluency in the use of Mandarin Chinese as L1 and L2 by native and non-native speakers, the present study examines the temporal variables (speech rate,

mean length of runs, unfilled pauses) and speech management strategies (filled pauses, repetitions and self-corrections) in the spontaneous speech of 30 Uygur learners of Mandarin Chinese with two different levels of Mandarin proficiency (primary and intermediate). Ten Mandarin Chinese speakers are selected as a control group. Previous studies also indicate that the pause duration and frequency are affected by different speech tasks that differed in the degree of content familiarity (Tavakoli & Skehan 2005). Therefore, the present study used three different tasks to investigate oral fluency comprehensively.

## **2 Research design**

The experimental design allows comparative studies to be made in the following categories:

- i. Inter-group: between native and non-native speakers of Mandarin Chinese in tasks in Mandarin Chinese
- ii. Intra-group: between the non-native speakers of primary and intermediate language proficiency
- iii. Between tasks: compare performance of native and non-native speakers in three different tasks in Mandarin Chinese

The following research questions guide this study. First, are there significant differences with respect to the temporal variables and speech management strategies in Mandarin spontaneous speech between Uygur learners and Mandarin Chinese speakers in all task situations? The hypothesis is that this is the case. Second, which factor or factors affect the temporal variables and speech management strategies in Mandarin spontaneous speech by Uygur learners and Mandarin Chinese speakers in all task situations? We hypothesize that gender and task type both affects the temporal variables and speech management strategies in Mandarin Chinese spontaneous speech by Uygur learners and

Mandarin Chinese speakers in all task situations. The level of language proficiency is predicted to only affect the repeat frequency by Uygur learners.

### 3 Methodology

#### 3.1 Participants

Forty participants took part in this study. Two groups of subjects were recruited. Group I consisted of thirty native speakers (22 females, 8 males) (age: 29-35) of Uygur. They were students of Xinjiang Normal University in Xinjiang Uygur Autonomous region, China. They had 14 years training experience in Mandarin Chinese. Among the subjects, thirteen subjects had passed HSK (Chinese proficiency test) for Grade 4 to Grade 5, which indicates the primary level of Mandarin Chinese. The other seventeen participants had passed HSK test for Grade 6 to Grade 8, which indicates the secondary level of Mandarin Chinese. Group II consisted of ten subjects (five males, five females) who were students of Xinjiang Normal University (age: 22-24) served as control group. Their native language is Mandarin Chinese. All of them had passed the National Putonghua proficiency test for Grade 2A.<sup>1</sup> None of the participants reported any hearing, vision, or reading deficiencies.

#### 3.2 Task situations

As mentioned earlier, the tasks selected involved differing degrees of content complexity for the subjects. The tasks selected in the present research were: topic narrative, topic argumentative and storytelling. Materials to elicit spontaneous speech included questions on topics familiar to the subjects and four pictures. Topic narrative and topic argumentative were elicited by asking each subject two questions. Questions selected to elicit spontaneous responses were on topics familiar to the subjects in their textbooks. They were taken from *Conquering HSK in 21 days*.

---

<sup>1</sup> The National Putonghua Proficiency Test (PSC) is a language test system developed by the Institute of Applied Linguistics, the Ministry of Education, and The State Administration of Radio Film and Television in October 1994. The aim of the PSC is to test level of mastery and ability in using Putonghua for speakers whose native language is Chinese.

These questions were:

- i. Please tell me about your hometown.
- ii. Would you prefer travelling on your own or with other people? Why?

The story telling task consisted of a series of six pictures arranged in logical order. These pictures were selected from *Picture Description*. The subjects were asked to “Please tell the story illustrated by the pictures in detail.”

### 3.3 Procedure

The participants were tested individually in a quiet room using a personal computer. All audio recordings were made using the microphone of a Sennheiser PC 300 high-quality headset. Topic narrative and topic argumentative tasks took about 7-10 minutes, respectively. The subjects were fully prepared, because they were given five minutes to prepare for the tasks. When the subjects fulfilled the story telling task, they were given five minutes to look over the pictures before the recording. This task lasted for 10-12 minutes. The experimenter presented the tasks through Microsoft PowerPoint and instructed the subjects in each task.

### 3.4 Data analysis

All the speech samples were analysed by Praat software (Boersma & Weenink 2015). The following variables were analysed in the present study.

- i. Speech rate: this variable included both speech rate and articulation rate in all speech samples. Speech rate is defined as the total number of syllables/total speech time, i.e. syllables per minute. Articulation rate is defined as the total number of syllables/total articulation time (total speech time-total articulation time), i.e. syllables per minute.
- ii. Mean length of runs: average number of syllables between unfilled pauses (more than 3 seconds).

- iii. Unfilled pause duration: each pause of 3 seconds or more than 3 seconds in duration. According to Riegenbach (1991), unfilled pauses shorter than 3 seconds are generally regarded as articulation pauses.
- iv. Filled pause duration: the duration of non-lexical fillers and lengthening of sounds.
- v. Pause/Repeats/Self-correction frequency: the number of pauses/repeats/self-corrections that occur per second.

## 4 Results

### 4.1 Data analysis

#### 4.1.1 Speech rate

Means and standard deviations that illustrate the findings for both speaking rate and articulation rate for three tasks are reported in Tables 1 and 2. The results indicate that Uygur learners produced slower speech than Mandarin Chinese speakers. As can be observed from Table 2, the articulation rate increased with the development of language proficiency. Both two groups of speakers produced slower speech in the story telling task. Moreover, male Uygur learners produced slower speech than female Uygur speakers. We performed three-way (Gender  $\times$  Proficiency  $\times$  Task) ANOVAs on the speech rate and the articulation rate, respectively. The statistical results indicated significant effects for gender on speech rate ( $F(2,120) = 4.24$ ,  $p = 0.04$ ), and proficiency ( $F(2,120) = 65.22$ ,  $p < 0.001$ ). A post hoc Scheffé test confirmed that the speaking rate of Uygur learners was slower than Mandarin Chinese speakers ( $p < 0.001$ ). As for the articulation rate, the main effects of task and proficiency on articulation rate was significant ( $F(2,120) = 94.81$ ,  $p < 0.001$ / $F(2,120) = 49.40$ ,  $p < 0.001$ ). A post hoc Scheffé test confirmed that the articulation rate of storytelling task was slowest ( $p < 0.001$ ). Moreover, the articulation rate of Uygur learners of Chinese was slower than Mandarin Chinese speakers ( $p < 0.001$ ).

Table 1. Mean and standard deviation of speaking rate by Uygur learners of Chinese and Mandarin Chinese speakers (syllable per minute) (M=male speakers, F=female speakers).

Proficiency	Task A				Task B				Task C			
	Mean		SD		Mean		SD		Mean		SD	
	M	F	M	F	M	F	M	F	M	F	M	F
Primary	119	139	18.76	27.02	110	120	17.74	30.27	87	121	13.26	20.00
Intermediate	124	134	12.37	27.48	127	132	12.49	27.24	121	117	34.08	30.77
Mandarin Chinese	199	206	48.15	55.25	206	212	28.57	56.95	172	201	29.09	44.31

Table 2. Mean and standard deviation of articulation rate by Uygur learners of Chinese and Mandarin Chinese speakers (syllable per minute)(M=male speakers, F=female speakers).

Proficiency	Task A				Task B				Task C			
	Mean		SD		Mean		SD		Mean		SD	
	M	F	M	F	M	F	M	F	M	F	M	F
Primary	198	209	27.05	26.22	188	193	8.30	31.69	82	109	3.69	19.79
Intermediate	209	217	19.57	41.76	207	198	23.23	33.17	117	112	34.93	22.70
Mandarin Chinese	279	264	17.23	60.62	280	266	24.92	50.89	169	187	29.92	56.81

#### 4.1.2 Unfilled pause duration and frequency

Means and standard deviations that illustrate the findings for pause duration during three different tasks in Mandarin Chinese are reported in Table 3. We performed a three-way ANOVA (Gender × Proficiency × Task) on the unfilled pause duration. The ANOVA yielded a significant effect for gender,  $F(1,2558) = 4.28, p < 0.05$ , and proficiency  $F(2,2558) = 10.52, p < 0.001$ , as well as a significant Gender × Proficiency interaction  $F(2,2558) = 3.92, p = 0.039$ . A

post hoc Scheffé test indicated that the duration of unfilled pauses by Uygur speakers was longer than that of Mandarin Chinese speakers ( $p = 0.034$ ,  $p = 0.001$ ).

*Table 3. Mean and standard deviation of unfilled pause duration by Uygur learners of Chinese and Mandarin Chinese speakers(s) (M=male speakers, F=female speakers).*

Proficiency	Task A				Task B				Task C			
	Mean		SD		Mean		SD		Mean		SD	
	M	F	M	F	M	F	M	F	M	F	M	F
Primary	0.95	0.83	0.65	0.64	0.98	0.94	0.76	0.88	1.16	0.81	0.08	0.57
Intermediate	0.84	0.92	0.56	0.92	0.80	0.82	0.69	0.81	0.85	0.87	0.67	0.74
Mandarin Chinese	0.82	0.68	0.82	0.65	0.71	0.72	0.43	0.66	0.80	0.78	0.56	0.46

Unfilled pause frequency was also measured in the present study. The results are presented in Table 4. The results indicated that unfilled pauses occurred most frequently in the storytelling task. Moreover, Uygur learners produced more unfilled pauses than Mandarin Chinese speakers. We performed a three-way ANOVA (Gender  $\times$  Proficiency  $\times$  Task) on the unfilled pause frequency. The ANOVA yielded a significant effect for gender  $F(1,120) = 14.43$ ,  $p < 0.001$ , proficiency  $F(2,120) = 23.98$ ,  $p < 0.001$ , and task  $F(2,120) = 5.80$ ,  $p = 0.004$ . The results of a post hoc Scheffé test confirmed that two groups of subjects paused more frequently when they fulfilled the story telling task ( $p = 0.018$ ,  $p = 0.007$ ). Moreover, Uygur learners paused more frequently than Mandarin Chinese speakers ( $p < 0.001$ ).

Table 4. Mean and standard deviation of unfilled pause frequency by Uygur learners of Chinese and Mandarin Chinese speakers (number per second) (M=male speakers, F=female speakers).

Proficiency	Task A				Task B				Task C			
	Mean		SD		Mean		SD		Mean		SD	
	M	F	M	F	M	F	M	F	M	F	M	F
Primary	0.67	0.60	0.07	0.09	0.74	0.65	0.11	0.05	0.81	0.77	0.03	0.19
Intermediate	0.77	0.64	0.11	0.20	0.71	0.59	0.13	0.15	0.85	0.68	0.19	0.17
Mandarin Chinese	0.51	0.41	0.21	0.12	0.48	0.37	0.12	0.09	0.65	0.46	0.19	0.14

#### 4.1.2 Mean length of runs

Table 5 presents means and standard deviations of mean length of runs by two groups of speakers in three tasks. The results showed that there was a big difference between Uygur learners and native speakers. Uygur learners of intermediate language proficiency produced longer mean lengths of runs. All of the speakers produced longer mean lengths of runs in the first two tasks. We performed a three-way ANOVA (Gender × Proficiency × Task) on the mean length of runs. The ANOVA yielded a significant effect for gender  $F(1,2670) = 45.87, p < 0.001$ , task  $F(2,2670) = 10.95, p < 0.001$ , and proficiency  $F(2,2670) = 185.78, p < 0.001$ , as well as significant interactions Gender × Proficiency,  $F(4,2670) = 5.49, p < 0.001$ , Gender × Task, Proficiency × Task,  $F(2,2670) = 13.31, p < 0.001, F(2,2670) = 18.79, p < 0.001$  Gender × Proficiency × Task interactions  $F(4,2670) = 17.69, p < 0.001$ . A post hoc Scheffé test confirmed that the MLR by Mandarin Chinese speakers was longer than that by Uygur learners of Chinese ( $p < 0.001$ ).

Table 5. Mean and standard deviation of mean length of runs by Uygur learners of Chinese and Mandarin Chinese speakers (number of syllables) (M=male speakers, F=female speakers).

Proficiency	Task A				Task B				Task C			
	Mean		SD		Mean		SD		Mean		SD	
	M	F	M	F	M	F	M	F	M	F	M	F
Primary	4.80	5.17	3.22	3.18	4.02	5.09	2.21	3.36	3.57	4.32	1.79	2.87
Intermediate	4.50	5.12	3.27	3.55	4.89	5.19	3.13	3.36	4.39	4.67	2.72	3.12
Mandarin Chinese	6.71	8.45	5.41	5.54	6.68	12.28	4.73	8.34	5.83	9.25	4.23	5.53

## 4.2 Speech management strategies

### 4.2.1 Filled pause duration and frequency

The means and standard deviations of filled pauses duration are presented in Table 6. The results demonstrated that Uygur male learners produced longer filled pause duration than Mandarin Chinese speakers. We performed a three-way ANOVA (Gender  $\times$  Proficiency  $\times$  Task) on the filled pause duration. The ANOVA yielded a significant effect for gender  $F(2,876) = 8.99$ ,  $p = 0.003$  and proficiency  $F(2,876) = 3.59$ ,  $p = 0.03$ , as well as a significant Gender  $\times$  Proficiency interaction  $F(2,876) = 6.45$ ,  $p = 0.002$ .

Table 6. Mean and standard deviation of filled pause duration by Uygur learners of Chinese and Mandarin Chinese speakers (s) (M=male speakers, F=female speakers).

Proficiency	Task A				Task B				Task C			
	Mean		SD		Mean		SD		Mean		SD	
	M	F	M	F	M	F	M	F	M	F	M	F
Primary	0.39	0.38	0.19	0.18	0.43	0.40	0.20	0.16	0.39	0.36	0.21	0.15
Intermediate	0.33	0.36	0.13	0.22	0.37	0.34	0.25	0.13	0.33	0.41	0.11	0.17
Mandarin Chinese	0.30	0.63	0.08	0.31	0.37	0.49	0.22	0.22	0.29	0.36	0.09	0.19

The means and standard deviation of filled pause frequency are also shown in Table 7. The results revealed that Uygur learners produced more filled pauses than Mandarin Chinese speakers. We also performed a three-way ANOVA (Gender × Proficiency × Task) on the filled pause frequency. The results indicated that only the main effect of proficiency was significant  $F(2,120) = 15.55, p < 0.001$ . A post hoc Scheffé test confirmed that filled pauses occurred more frequently in Uygur learners’ production ( $p < 0.001, p = 0.009$ ).

Table 7. Mean and standard deviation of filled pause frequency by Uygur learners of Chinese and Mandarin Chinese speakers (number per second). (M=male speakers, F=female speakers).

Proficiency	Task A				Task B				Task C			
	Mean		SD		Mean		SD		Mean		SD	
	M	F	M	F	M	F	M	F	M	F	M	F
Primary	0.22	0.24	0.16	0.18	0.31	0.28	0.16	0.14	0.30	0.29	0.24	0.19
Intermediate	0.28	0.21	0.18	0.19	0.35	0.26	0.15	0.22	0.19	0.24	0.15	0.18
Mandarin Chinese	0.07	0.01	0.09	0.03	0.07	0.10	0.07	0.10	0.04	0.06	0.05	0.09

#### 4.2.2 Repeats

The means and standard deviation of repeat frequency are also shown in Table 8. As can be seen from Table 8, Uygur learners repeated more frequently than Mandarin Chinese speakers. Moreover, Uygur learners with primary language proficiency repeated more frequently than the learners with intermediate language proficiency. We performed a three-way ANOVA (Gender  $\times$  Proficiency  $\times$  Task) on the repeat frequency. The ANOVA yielded only a significant effect for proficiency  $F(2,120) = 17.75$ ,  $p < 0.001$ . A post hoc Scheffé test confirmed that there was a significant difference between Uygur learners and Mandarin Chinese speakers ( $p < 0.001$ ,  $p = 0.009$ ), Uygur learners with different levels of proficiency ( $p = 0.03$ ).

Table 8. Mean and standard deviation of repeat frequency by Uygur learners of Chinese and Mandarin Chinese speakers (number per second) (M=male speakers, F=female speakers).

Proficiency	Task A				Task B				Task C			
	Mean		SD		Mean		SD		Mean		SD	
	M	F	M	F	M	F	M	F	M	F	M	F
Primary	0.07	0.03	0.11	0.02	0.08	0.06	0.09	0.05	0.09	0.07	0.05	0.03
Intermediate	0.05	0.03	0.02	0.03	0.04	0.03	0.03	0.04	0.04	0.03	0.03	0.04
Mandarin Chinese	0.01	0.01	0.01	0.02	0.01	0.00	0.01	0.00	0.01	0.00	0.02	0

#### 4.3 Self-correction

Table 9 shows the means and standard deviations of self-correction frequency. The results revealed that Uygur learners self-corrected more frequently than Mandarin Chinese speakers. We performed a three-way ANOVA (Gender  $\times$  Proficiency  $\times$  Task). The ANOVA yielded a significant effect for proficiency  $F(2,120) = 9.55$ ,  $p < 0.001$ , but not for other factors and

interactions. A post hoc Scheffé test showed that self-correction occurred more frequently in Uygur learners’ productions ( $p < 0.001, p = 0.003$ ).

Table 9. Mean and standard deviation of self-correction frequency by Uygur learners of Chinese and Mandarin Chinese speakers (number per second) (M=male speakers, F=female speakers).

Proficiency	Task A				Task B				Task C			
	Mean		SD		Mean		SD		Mean		SD	
	M	F	M	F	M	F	M	F	M	F	M	F
Primary	0.06	0.05	0.04	0.05	0.03	0.03	0.05	0.02	0.05	0.04	0.01	0.03
Intermediate	0.02	0.06	0.02	0.03	0.06	0.04	0.02	0.04	0.07	0.03	0.03	0.03
Mandarin Chinese	0.01	0.003	0.02	0,01	0.02	0.009	0.02	0.01	0.02	0.02	0.02	0.02

## 5 Discussion and conclusions

The present study investigated the temporal variables (speech rate, mean length of runs, unfilled pauses) and speech management strategies (filled pauses, repetitions and self-corrections) between Uygur learners with different levels of language proficiency and Mandarin Chinese speakers. The results confirmed our hypothesis that there were significant differences with respect to the temporal variables and speech management strategies in Mandarin Chinese spontaneous speech between Uygur learners and Mandarin Chinese speakers. Such differences may reflect that learners (whose first language is well established) have to allocate greater processing resources to suppress their first language (Guion et al. 2000). It may also reflect learners’ difficulty in automatizing their second language (Towell 2002). They relied on declarative memory (i.e. memory for factual information) in their processing of second language input.

Moreover, there were many factors, including the level of language proficiency, gender and the task type that affected the temporal variables and speech management strategies by the two groups of speakers. The familiarity degree of the tasks affected the oral fluency as indicated in Tavakoli & Skehan's study (2005). The gender difference of Uygur learners may be influenced by the gender difference in their native language (Grosjean & Deschamps 1975) which needs to be investigated in future research. Among the three tasks, the picture description task would be more difficult for the speakers. This is probably because, in their textbooks, Uygur learners have been trained for the narrative and argumentative topics, but not for the storytelling task. Moreover, the narrative task topic would be much easier for the speakers, because it is a less structured task in that it follows more freedom of lexical and grammatical choice. On the other hand, the storytelling task is a highly structured task in that it requires the speaker to produce particular lexical and grammatical items.

However, Uygur learners with intermediate language proficiency did not show significant differences on these temporal variables and speech management strategies, except for the repeat frequency. In the language production model proposed by Anderson (1983), fluent speech production requires procedural knowledge, which underlies skilled behaviour. Thus, the temporal variables of Uygur learners of Chinese with primary and intermediate Chinese proficiency were not significant, which indicates the slower conversion of declarative knowledge, i.e. knowledge about the world, into procedural knowledge. The frequent occurrence of unfilled pauses and self-corrections by Uygur learners of Chinese indicated a high planning pressure in spontaneous speech. These speech planning strategies implied a self-monitoring ability in second language production. However, it develops slower for speakers with primary and intermediate Chinese proficiency.

The following pedagogical implications may be considered based on the above findings: first, teacher-training should emphasize the importance of speaking with normal native rhythms. The teachers should make sure the students are aware of such subtle aspects of speech which could minimize the perception of "foreignness" of their speech. Second, the

students should be provided with the language model appropriate for acquiring temporal features of the second language.

In order to support the results presented here, this research should be repeated with learners with advanced language proficiency and with second language learners of other languages.

### Acknowledgements

Funding was provided by the Ministry of Education in China Project of Humanities and Social Sciences 13YJC740124. Thanks to Marjoleine Sloos for organizing the workshop “Chinese accents and accented Chinese” (9-10 October, 2014). Thanks to the reviewers for their valuable comments.

### References

- Anderson, John R. 1983. *The Architecture of Cognition*. Cambridge, Massachusetts, London: Harvard University Press.
- Boersma, Paul & Weenink, David. 2010. Praat version 5.4.19, Netherlands, University of Amsterdam.
- Brumfit, Christopher. 1984. *Communicative Methodology in Language Teaching: The Roles of Fluency and Accuracy* 129-33. Cambridge: Cambridge University Press.
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward. 1999. *Longman grammar of spoken and written English*. London/New York: MIT Press.
- Chambers, Francine. 1997. What do we mean by fluency? *System*, 25.4, 535-544.
- Chen, Mo & Wang, Jian.Qin. 2008. A case study of prosodic boundary features of spoken Chinese output. *Chinese Language Learning*, 4, 102-108.
- Chen, Mo. 2012. Chinese oral fluency of CSL learners of American English speakers. *Language Teaching and Research*, 2, 17-24.
- Cotton, Jack. C. 1936. Syllabic rate: A new concept in the study of speech rate variation. *Communications Monographs*, 3.1, 112-117.

- Crystal, Thomas H., & House, Arthur S. 1990. Articulation rate and the duration of syllables and stress groups in connected speech. *The Journal of the Acoustical Society of America*, 88.1, 101-112.
- Dong, Jie. 2011. Analysis of oral fluency in Standard Mandarin speech by intermediate learners. MA thesis. Shanghai, Fudan University.
- Foster, Pauline, & Tavakoli, Parvane. 2009. Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning*, 59.4, 866-896.
- Götz, Sandra. 2013. *Fluency in native and nonnative English speech*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Grosjean, François, & Deschamps, Alain. 1975. Analyze the different temporal variables of English and French. *Phonetica*, 31.3-4, 144-184.
- Grosjean, François. 1980. Temporal variables within and between languages. In H. W. Dechert and M. Raupach (eds). *Towards a Cross-Linguistic Assessment of Speech Production*. *Language Frankfurt*, 39-53.
- Guion, Susan G., Flege, James E., Liu, Serena H., & Yeni-Komshian, Grace H. 2000. Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, 21.2, 205-228.
- Gut, Ulrike. 2009. *Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German*. In Thomas Kobnen & Joybrato Mukberjee (eds.), *English Corpus Linguistics*, Vol. 9. Peter Lang.
- Lennon, Paul. 1990. Investigating fluency in EFL: A quantitative approach. *Language learning*, 40.3., 387-417.
- Levelt, Willem. J. 1993. *Speaking: From intention to articulation*. Cambridge, MA: MIT press.
- Munro, Murry J., & Derwing, Trace M. 1998. The effects of speaking rate on listener evaluations of native and foreign-Accented Speech. *Language Learning*, 48.2, 159-182.
- Riggenbach, Heidi. 1991. Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, 14.4, 423-441.
- Schmidt, Richard. 1992. Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14.4, 357-385.

- Siegman, Aron W., & Feldstein, Stanley. 1979. *Of speech and time: Temporal Speech Patterns in Interpersonal Contexts*. Hillsdale, N.J.: Laurence Erlbaum.
- Tavakoli, Parvaneh, & Skehan, Peter. 2005. Strategic planning, task structure, and performance testing. *Planning and Task Performance in a Second Language*, 239-273.
- Towell, Richard. 2002. Relative degrees of fluency: A comparative case study of advanced learners of French. *International Review of Applied Linguistics in Language Teaching*, 40.2., 117-150.
- Ye, Ang Long. 1999. *Picture description*. Zhuhai: Zhuhai Press.
- Zhao, Xiang Ru., & Zhu, Zhi Ning. 1985. *A brief description of Uygur*. Beijing: Minzu Press.
- Zheng, Li Jie. 2006. *Conquering HSK in 21 days*. Beijing: Foreign Language Teaching and Research Press.

# Focus and tonal implementation in Shanghai Chinese

Bijun Ling    Jie Liang

*Tongji University*

## Abstract

This paper reports the results of an experiment, which aimed to get a better understanding of the acoustic realization of tones and focus in Shanghai Chinese. In Shanghainese Chinese, disyllabic words may form tone sandhi domains, within which the tone of first syllable represents the vocal register features (yin/ yang) and the second syllable realizes the dynamic targets (rising/ lowering) of mono-syllable. In experiment, disyllabic nouns were elicited as No-Focus and Focus of the utterance. The results indicate that: (a) Disyllabic word performs as a prosodic unit, within which each syllable is influenced by focus; (b) Analyses of the Fo patterns within disyllabic words revealed that, in focus, the Fo adjustment of each syllable is dependent on its tonal target: [H] is raised while [L] remains the same; and syllables with higher tonal targets are affected by focus most.

**Keywords:** Shanghai Chinese, Focus, Fo.

## 1 Introduction

Focus is a grammatical category that determines which part of the sentence contributes new, non-derivable, or contrastive information (Halliday 1967). Focus can be signalled prosodically or syntactically, or both, but in oral communication speakers normally use prosodic encoding to highlight focus constituents. Focus is reflected by Fo rising, longer duration and increased intensity. In stress languages, focus is indicated by the presence of pitch accents on focused constituents and absence of pitch accents on post-focal constituents; the Fo peak is located on the stressed syllable within the focus domain (Ladd 1996, Xu & Xu 2005), as illustrated in Figure 1. In tone languages, Fo contours (high/low, rise/fall) are employed to distinguish the

lexical meanings and pragmatic meanings (e.g. focus). How are lexical tones modified to convey pragmatic meanings? How do lexical tones interact with intonation to form the surface Fo contours?

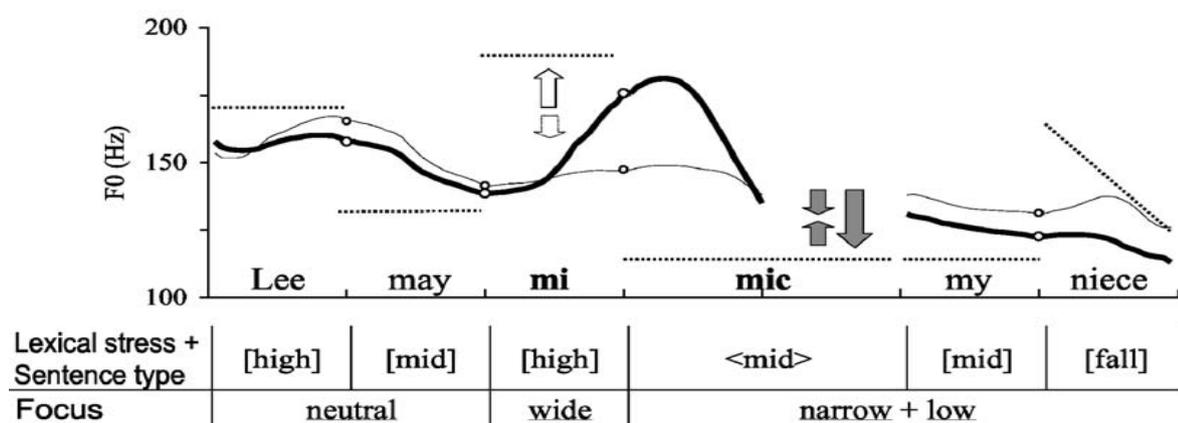


Figure 1. Fo contour of “Lee may mimic my niece” with neutral focus (grey line) and focus on “mimic” (black line) (cited from Xu & Xu 2005).

Recent research has greatly increased our knowledge of the interaction between the four lexical tones and focus in Mandarin Chinese, a typical tone language. Focus is reflected by expansion of the pitch range in focus position and compression of the pitch range in post focus position (Xu 1999); moreover, in focus position, lexical tones are produced with enhanced distinctiveness of their tone targets: [H] is higher, while [L] is lower than in non-focus position (Xu 1999, Chen 2008), as illustrated in Figure 2. We believe that new data, particularly from a different type of tone language, may shed further light on the acoustic realization of focus.

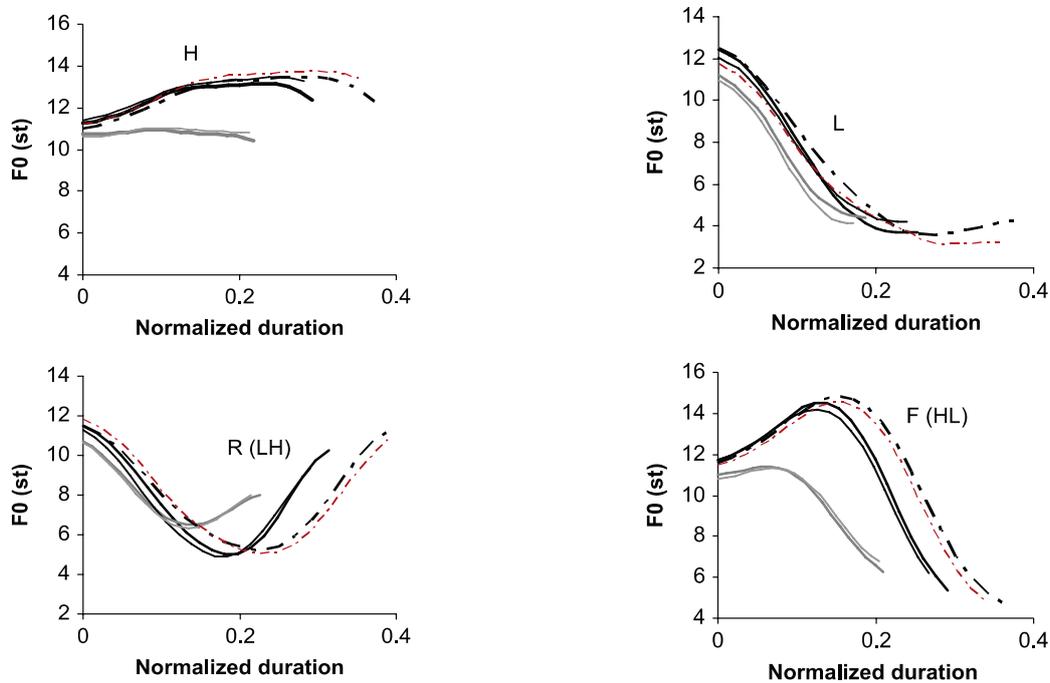


Figure 2. *F<sub>0</sub> contours of four tones in Mandarin Chinese (black-focus, grey-neutral) (cited from Chen 2008).*

Shanghai Chinese (SHC), a Wu dialect, is also a tone language. There are five citation tones in SHC, and its tonal system can be described by three dimensions (Zhu 2005): (1) *F<sub>0</sub> contour*: falling (T<sub>1</sub>) and rising (T<sub>2</sub>-T<sub>5</sub>); (2) *Register*: high register (T<sub>1</sub>, T<sub>2</sub>, T<sub>4</sub>) and low register (T<sub>3</sub>, T<sub>5</sub>) (caused by preceding breathy voiced plosives); (3) *Duration*: long (T<sub>1</sub>-T<sub>3</sub>) and short (T<sub>4</sub>, T<sub>5</sub>) (caused by following glottal codas). Different from Mandarin Chinese, a syllable-tone language, Shanghai Chinese is a so-called word-tone language (Zhang 1988, Shih 1989), because in SHC polysyllabic words form a tone sandhi domain with a fixed *F<sub>0</sub>* pattern, which depends on the underlying tone of the initial syllable. In other words, polysyllabic words retain only the tone target of the initial syllable, which spreads over the first two syllables, and any following syllable gets a default low tone (Yip 2002). The value of the five citation tones and their *F<sub>0</sub>*

patterns in disyllabic words are listed in Table 1 (using Chao's five-level numerical scale, which divides a speaker's pitch range into five steps with 5 indicating the highest end and 1 the lowest) (Chao 1930).

Table 1. Tone values (values cited from Qian 1988).

Tone	Citation	Disyllabic words
T <sub>1</sub>	52 (HL)	55+31 (H+L)
T <sub>2</sub>	34 (MH)	33+44 (M+H)
T <sub>3</sub>	23 (LH)	22+44 (L+H)
T <sub>4</sub>	5 (H)	33+44 (M+H)
T <sub>5</sub>	12 (LH)	11+23 (L+H)

In Mandarin Chinese, if a polysyllabic word is in focus, each syllable keeps its own tone target but in a more distinctive way, which means the  $F_0$  adjustment of each syllable is dependent on its tone targets, and the last syllable within the polysyllabic word is influenced most by focus. In SHC, on the other hand, polysyllabic words form a tone sandhi domain with a fixed  $F_0$  pattern (Zee 1988, Duanmu 1999). This leads to the following questions: (1) Does SHC have a different mechanism of encoding focus, compared to Mandarin Chinese? (2) Within the focus domain, does the  $F_0$  contour of each syllable undergo the same adjustment or are adjustment different depending on a syllable's tonal target (like in MC)? (3) Does each syllable of the disyllabic word in focus have the same adjustment range? If not, which of the two syllables is influenced most? To answer these questions, we compare the Max $F_0$ , Min $F_0$  and  $F_0$  range of five disyllabic words in focused and non-focused condition; and we compare the mean  $F_0$  of each syllable within disyllabic words in focused and non-focused condition.

## 2 Methodology

### 2.1 Stimuli

This production experiment consisted of two sets of dialogues which had to be read aloud: dialogue A contained a yes-or-no question (e.g. Did Manager Bao see Boss Bao at Suzhou?) and the answer confirmed the question by repeating it (e.g. Manager Bao saw Boss Bao at Suzhou.) Because there was no new information, we regarded the answer as a neutral focus statement; dialogue B contained an interrogative question (e.g. Who saw Boss Bao at Suzhou?) and the answer gave the required information as a narrow focus, which included initial, medial and final focus. Therefore, we had 20 dialogues (5 tones \* 4 focus conditions) in total. The target words are listed in Table 2.

Table 2. Stimuli.

Tone	Initial (Manager X)	Medial (place name)	Final (Boss X)
T <sub>1</sub>	包总 (pə tsoŋ)	苏州 (su tsɿ)	包董 (pə toŋ)
T <sub>2</sub>	郝总 (hə tsoŋ)	广州 (kuaŋ tsɿ)	郝董 (hə toŋ)
T <sub>3</sub>	毛总 (mə tsoŋ)	扬州 (iaŋ tsɿ)	毛董 (mə toŋ)
T <sub>4</sub>	柏总 (paʔ tsoŋ)	德州 (tɛʔ tsɿ)	柏董 (paʔ toŋ)
T <sub>5</sub>	白总 (baʔ tsoŋ)	绿州 (loʔ tsɿ)	白董 (baʔ toŋ)

Example dialogues are given below, in which the underlined word is the narrow focus.

Dialogue A:

包总 辣苏州 看到 包董 勒伐?

pətsəŋ ləʔ sutsɿ k'ə tə pətoŋ ləʔ vaʔ

Did Manager Bao see Boss Bao at Suzhou?

Neutral focus

包总 辣苏州 看到勒 包董.

pətsəŋ ləʔ sutsɿ k'ətələʔ pətoŋ

Manager Bao saw Boss Bao at Suzhou.

Dialogue B:

啥人辣苏州看到包董勒?

san'in ləʔ sutɕ k'ətɔ pətɔŋ ləʔ

Who saw Boss Bao at Suzhou?

包总辣啥地方看到包董勒?

pətɔŋ ləʔ sadifaŋ k'ətɔ pətɔŋ ləʔ

Where did Manager Bao see Boss Bao?

包总辣苏州看到啥人勒?

pətɔŋ ləʔ sutɕ k'ətɔ san'in ləʔ

Who did Manager Bao see at Suzhou?

Initial focus

包总辣苏州看到勒包董.

pətɔŋ ləʔ sutɕ k'ətɔləʔ pətɔŋ

**Manager Bao** saw Boss Bao at Suzhou.

Medial focus

包总辣苏州看到勒包董.

pətɔŋ ləʔ sutɕ k'ətɔləʔ pətɔŋ

Manager Bao saw Boss Bao at **Suzhou**.

Final focus

包总辣苏州看到勒包董.

pətɔŋ ləʔ sutɕ k'ətɔləʔ pətɔŋ

Manager Bao saw **Boss Bao** at Suzhou.

## 2.2 Subjects and Recording

Two male and two female speakers of SHC, 25-35 years old, participated in the experiment. All four speakers were born and grew up in Shanghai, mainly in Yangpu and Hongkou districts. All speakers were recorded in a sound-insulated booth at Tongji University. The speakers were asked to read the questions and answers from a reading list four times, at their normal speaking rate and as naturally as possible. All speakers were aware of the fact that the experiment involved a study of prosody in SHC, but were naïve as to the specific purpose.

## 2.3 Acoustic and Statistical Analyses

The rhyme of each syllable was manually labelled in Praat (Boersma & Weenink, 2010). Fo contours were obtained by taking 10 equidistant temporal points (in Hz) of the rhyme of each syllable by using the script “ProsodyPro” (Xu, 2013). These raw values in Hz were transformed into semitones (st; a psycho-acoustic scale equal to musical intervals) and averaged across speakers and repetitions of the same sentence for each discourse context separately. Formula (1) derives semitones (st) from frequency in Hz, F:

$$(1) \quad st = 12 * \text{Log}_2(F/50)$$

Subsequently, we computed the mean Fo of each syllable in each disyllabic word. We also measured the max Fo, min Fo and Fo range (the highest minus the lowest pitch) of each whole disyllabic word.

### 3 Results

Our goal was to investigate the acoustic realization of focus in a word-tone language. First, we examined the maxFo, minFo and Fo range of the whole disyllabic word. A Multivariate General Linear Model was conducted with Fo range, max Fo and min Fo of each disyllabic word as dependent factors, and Focus (two levels: neutral and focus), and Tone (five levels) as independent factors. Second, we examined the Fo contour of each syllable of each disyllabic word in focus, to see the trend and scope of adjustment of each syllable under focus. The mean Fo of each syllable was taken as a dependent factor, and Focus (two levels: neutral and focus), Tone (five levels) and Syllable (two levels: S<sub>1</sub> & S<sub>2</sub>, for first and second syllable in the word) were treated as independent factors.

#### 3.1 The maxFo, minFo and Fo range of disyllabic words

Comparing the Fo contours of disyllabic words in focus and in neutral condition, there are two things to be noted. First, in Shanghai Chinese, disyllabic words are influenced by focus as a whole. Second, there are three distinct focus-related pitch ranges: (1) expanded in focused words, (2) it is suppressed (lowered and compressed) in post-focus words, (3) neutral in pre-focus words. This is illustrated in Figure 3.

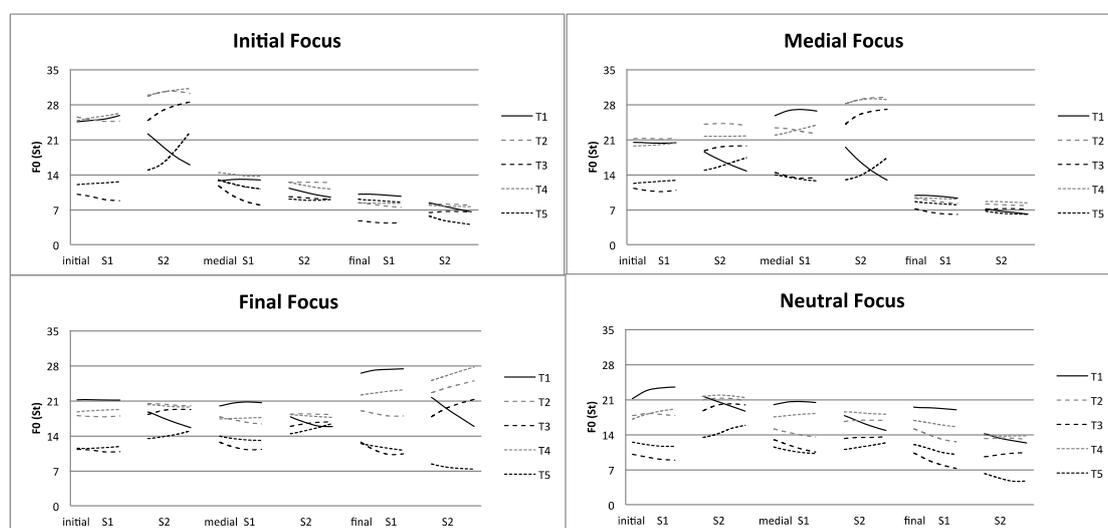


Figure 3. The mean F<sub>0</sub> contours of disyllabic words in initial, medial and final positions (averaged across speakers and repetitions).

The Multivariate General Linear Model indicated that focus and tone had main effects on MaxF<sub>0</sub> (focus:  $F = 15.274$ ,  $p < 0.001$ ; tone:  $F = 15.274$ ,  $p < 0.001$ ), MinF<sub>0</sub> (focus:  $F = 4.541$ ,  $p = 0.034$ ; tone:  $F = 40.135$ ,  $p < 0.001$ ) and F<sub>0</sub> range (focus:  $F = 126.430$ ,  $p < 0.001$ ; tone:  $F = 59.999$ ,  $p < 0.001$ ) of disyllabic words; focus \* tone had a significant influence on MinF<sub>0</sub> ( $F = 2.705$ ,  $p = 0.030$ ) and F<sub>0</sub> range ( $F = 6.699$ ,  $p < 0.001$ ).

The data were divided by tone, and Independent Sample T tests were conducted to see how focus influenced the maxF<sub>0</sub>, minF<sub>0</sub> and F<sub>0</sub> range of disyllabic words with different tones. The results indicated that focus made the maxF<sub>0</sub> and F<sub>0</sub> range of all disyllabic words increase significantly, but only the minF<sub>0</sub> of T<sub>2</sub> and T<sub>4</sub> disyllabic words were increased significantly by focus, while the minF<sub>0</sub> of T<sub>1</sub>, T<sub>3</sub> and T<sub>5</sub> disyllabic words didn't show significant change. The mean maxF<sub>0</sub>, minF<sub>0</sub> and F<sub>0</sub> range of disyllabic word of each tone was listed in Table 3.

Table 3. The maxFo, minFo and Fo range of disyllabic words

		T1	T2	T3	T4	T5
maxFo	Neutral	26.87	26.38	25.11	25.57	23.31
	Focus	31.79	30.23	29.56	30.18	27.07
		P<0.001	P<0.001	P<0.001	P<0.001	P=0.001
minFo	Neutral	18.23	20.74	16.24	20.85	16.41
	Focus	17.61	23.37	15.86	23.17	16.84
		P=0.576	P=0.004	P=0.715	P=0.003	P=0.668
Fo range	Neutral	8.64	5.64	8.87	4.71	6.90
	Focus	14.18	6.86	13.70	7.01	10.23
		P<0.001	P=0.046	P=0.002	P=0.001	P<0.001

To summarize, there are several things to be noted. First, in SHC, disyllabic words, as prosodic units, are influenced by focus as a whole. Secondly, there is a tri-zone focus adjustment: pitch range was expanded in focused words, suppressed in post-focus words, and neutral in pre-focus words, which is the same as in Mandarin Chinese (Xu 1999). Thirdly, focus induced an Fo range expansion mainly by raising the MaxFo of disyllabic words, while the change of MinFo depended on the tone of the initial syllable of the disyllabic word: only the MinFo of T2 and T4 disyllabic words were significantly raised, while the MinFo of T1, T3 and T5 disyllabic words didn't show any significant change. What caused the different adjustment of minFo of disyllabic words? In order to answer this question, we investigated the Fo pattern within each disyllabic word.

### 3.2 The Fo pattern of disyllabic words

Figure 4 displays the mean Fo contours of disyllabic words starting with each tone in initial, medial and final position, where the black line indicates neutral focus and the dotted line indicates narrow focus. Each tone exhibits a distinctive pattern of Fo movement characteristic of the tonal feature(s).

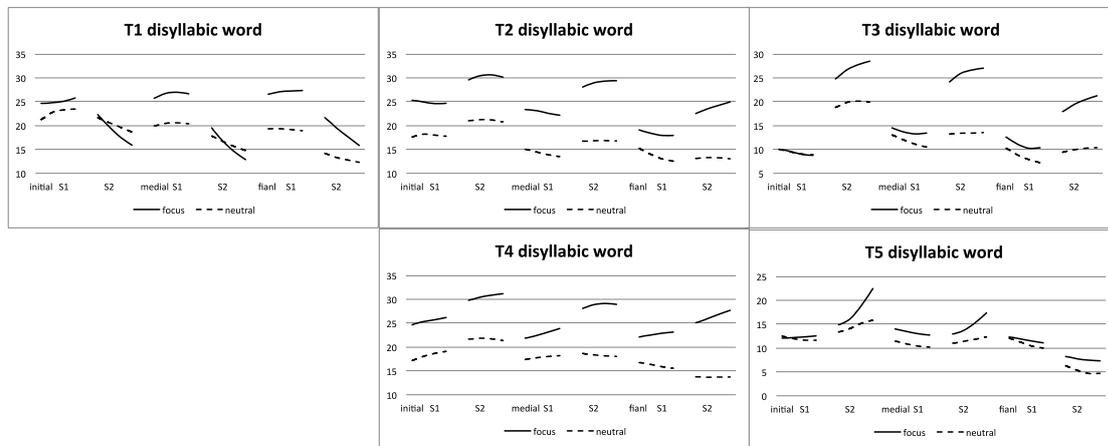


Figure 4. Mean  $F_0$  contours of disyllabic words starting with different tones in initial, medial and final positions.

In T<sub>1</sub> disyllabic words, the first syllable ( $\sigma_1$ ) is level high and the second one ( $\sigma_2$ ) has a falling tone, which means that the onset  $F_0$  of  $\sigma_2$  is dependent on the  $F_0$  of  $\sigma_1$ , and therefore we used the offset  $F_0$  of  $\sigma_2$  as an independent factor. In focus, the mean  $F_0$  of  $\sigma_1$  rose significantly ( $t = 8.502, p < 0.001$ ) by 5.14 st on average; the offset  $F_0$  of  $\sigma_2$  didn't show any significant change ( $t = 0.670, p = 0.504$ ).

In T<sub>2</sub> and T<sub>4</sub> disyllabic words, both syllables are level high and the  $F_0$  of the second syllable was higher than the first one. In focus, the mean  $F_0$  of  $\sigma_1$  (T<sub>2</sub>:  $t = 3.033, p = 0.003$ ; T<sub>4</sub>:  $t = 4.038, p < 0.001$ ) and  $\sigma_2$  (T<sub>2</sub>:  $t = 6.343, p < 0.001$ ; T<sub>4</sub>:  $t = 6.681, p < 0.001$ ) were both raised significantly, and the  $F_0$  of  $\sigma_2$  increased more than  $\sigma_1$  in both T<sub>2</sub> and T<sub>4</sub> disyllabic words.

In T<sub>3</sub> disyllabic words, the first syllable is a low level tone caused by a preceding breathy voiced plosive, and the second syllable is level high. In focus, the mean  $F_0$  of  $\sigma_1$  didn't show a significant change ( $t = 0.855, p = 0.395$ ), while the mean  $F_0$  of  $\sigma_2$  was significantly raised by 4.98 st on average ( $t = 6.606, p < 0.001$ ).

In T<sub>5</sub> disyllabic words, the first syllable is also a level low tone due to a preceding breathy voiced plosive, but the second syllable is low rising. So we used the offset  $F_0$  of  $\sigma_2$  as an independent factor, like what we did for T<sub>1</sub> disyllabic words. In focus, the mean  $F_0$  of  $\sigma_1$  didn't

show a significant change ( $t = 1.090$ ,  $p = 0.279$ ), but the offset  $F_0$  of  $\sigma_2$  rose significantly ( $t = 3.173$ ,  $p = 0.002$ ).

Table 4. *The mean  $F_0$  of first and second syllable within disyllabic words*

	(st)	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>
S <sub>1</sub>	Neutral	23.5	23.8	20.3	23.3	19.8
	Focus	30.7	26.1	21.2	26.1	20.8
		$p < 0.001$	$p = 0.003$	$p = 0.395$	$p < 0.001$	$p = 0.314$
S <sub>2</sub>	Neutral	18.5	24.2	22.8	23.6	22
	Focus	17.7	28.6	27.7	28.7	26.3
		$p = 0.504$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p = 0.001$

To summarize, disyllabic words in SHC have fixed  $F_0$  patterns, depending on the tone type of the initial syllable. In focus, both syllables within disyllabic words exhibited a distinctive pattern of  $F_0$  movement characteristic of its tonal feature(s): [L] tones didn't show any significant change while all other tones were significantly raised. This explains why only the Min $F_0$  of T<sub>2</sub> and T<sub>4</sub> disyllabic words were significantly raised, while those of other words remained the same. The Min $F_0$  of T<sub>2</sub> and T<sub>4</sub> disyllabic words turned out to depend on the  $F_0$  of  $\sigma_1$ , (with tone target [M]), which was raised significantly in focus. On the contrary, the Min $F_0$  of T<sub>1</sub> [HL] disyllabic word and T<sub>3</sub>/T<sub>5</sub> [LH] disyllabic words were realized by the offset  $F_0$  of  $\sigma_2$  and the  $F_0$  of  $\sigma_1$  respectively, which have [L] as the tone targets, and which weren't affected by focus.

Finally, within disyllabic words, the magnitude of adjustment of each syllable was not the same, and in fact syllables with higher targets changed more, i.e. the first syllable of T<sub>1</sub> disyllabic words and the second syllable of other words underwent significantly larger changes. The tone features and  $F_0$  movement patterns and range of each disyllabic word are summarized in Table 6.

Table 6. Fo movement pattern and range of each syllable (\*means most significant change).

Tone	First syllable		Second syllable	
T1 (HL)	H ↑	*(5.14 st)	L	(-0.76 st)
T2 (MH)	M ↑	(2.36 st)	H ↑	*(4.41 st)
T3 (LH)	L	(0.84 st)	H ↑	*(4.98 st)
T4 ( <u>H</u> )	M ↑	(2.72 st)	H ↑	*(5.12 st)
T5 ( <u>LH</u> )	L	(1.02 st)	H ↑	*(4.21 st)

#### 4 Discussion and conclusion

There have been many studies on the acoustic realization of focus in different languages. Although in different languages focus is generally reflected as the rise of MaxFo, the extension of duration and the enhancement of intensity, the Fo-segment alignment is different from language to language, reflecting the features of different languages. In stress languages, the MaxFo and the highest magnitude of Fo adjustment is located on the stressed syllable (Gussenhoven, 1983; Ladd, 1996; Xu & Xu, 2005), while in syllable-tone languages, the location of MaxFo depends on the tone type of the syllable: [H] is even more raised and [L] is even more lowered; and the highest magnitude of Fo adjustment is located on the last syllable of the prosodic unit (Xu, 1999; Yuan, 2004).

In this paper, we studied the acoustic realization of focus in Shanghai Chinese, a word-tone language, in order to shed some new light on the prosodic encoding of focus. By comparing the prosodic realization of disyllabic words in two discourse contexts, focus and neutral, we found out that speakers adjusted MaxFo and Fo range as a function of focus. Focus is realized with a magnified Fo contour characteristic of the domain-initial lexical tone, which reflects its characteristics as a tone language. Within the prosodic unit, the magnitude of adjustment of the syllable with the highest pitch target is most significant, which is, in turn, characteristic of a stress language.

Actually, our findings are in line with the acoustic realization of focus in Tokyo Japanese

(Lee & Xu 2012), a well-known pitch accent language. In both languages, there is consistent expansion of  $F_0$  range and raising of Max  $F_0$  and Mean  $F_0$  on the focused item, but focus does not affect Min  $F_0$  in any observable way; the accented syllable is phonetically realized as [H] and rises significantly on focus (Lee & Xu 2012). Is this just a coincidence or is word-tone languages belong to accent languages (Yip 2002)? More research work needs to be done to clarify this question by investigating more word-tone languages and comparing them to accent languages.

### References

- Boersma, Paul. & Weenink, David. 2010. *Praat: Doing phonetics by computer* [computer program]. Downloaded from [www.praat.org](http://www.praat.org).
- Chao, Yuen-Ren. 1930. *A system of tone letters*. *Le Maitre Phonétique*, 30, 24–27. [Reprinted in Fangyan (1980) 2, 81–83.].
- Chen, Yiya. & Gussenhoven, Carlos. 2008. *Emphasis and tonal implementation in Standard Chinese*, *Journal of Phonetics*, 36, 724–746
- Chen, Mathew. 2004. *Tone sandhi: patterns across Chinese dialects*, Cambridge University Press
- San, Duanmu. 1999. *Metrical Structure and Tone: Evidence from Mandarin and Shanghai*. *Journal of East Asian Linguistics* 8, 1–38.
- Gussenhoven, Carlos. 1983. *Testing the reality of focus domains*. *Language and Speech*, 26(1), 61–80.
- Halliday, Michael Alexander Kirkwood. 1967. *Intonation and Grammar in British English*. The Hague : Mouton
- Ladd, Robert. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
- Lee, Albert. & Xu, Yi. 2012. *Revisiting focus prosody in Japanese*. *Proceedings of Speech Prosody*, 2012
- Shih, Chi-lin. 1988. *Tone and intonation in Mandarin*. *Working Papers of the Cornell Phonetics Laboratory*, 3, 83–109.
- Xu, Yi. 1999. *Effects of tone and focus on the formation and alignment of  $F_0$  contours*. *Journal of Phonetics*, 27(1), 55–105.

- Xu, Yi., & Xu, X. 2005. *Phonetic realization of focus in English declarative intonation*. *Journal of Phonetics*, 33, 159-197.
- Xu, Yi. 2013. *ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis*. TRASP'2013, Aix-en-Provence, France.
- Yip, Morris. 2002. *Tone*, Cambridge University Press.
- Jia, Yuan. 2008. *Intonation in Mandarin Chinese: Acoustic, perception, and computational modeling*. Ph.D. dissertation, Cornell University, Ithaca, NY.
- Zhang, Zhengsheng. 1988. *Tone and Tone Sandhi in Chinese*. Ph.D. dissertation, The Ohio State University, Columbus, OH.
- 钱乃荣等 [Qian, Nairong]. 2007. *上海话大辞典(辞海版)*[Dictionary of Shanghai dialect]. Shanghai: 上海辞书出版社 [Shanghai Lexicographic Publishing House].
- 徐云扬[Xu, Yunyang]. 1988. *自主音段音韵学理论与上海声调变读*[Auto-segmental Phonology and the tone sandhi in Shanghai Chinese]. *中国语文*[Zhongguo Yuwen], 5, 331-349
- 朱晓农 [Zhu, Xiaonong]. 2005. *上海话声调实验录*[An Experimental Study in Shanghai Tones]. 上海 : 上海教育出版社 [Shanghai: Shanghai Educational Publishing House] .

# **A Study of Pitch Pattern of Declarative Sentences Uttered by Chinese speakers of English in Comparison with Native Speakers of English**

Wenjun Chen

*School of Humanities, Ningbo DaHongYing University, China*

## **Abstract**

Chinese-accented English appears to involve special features in its pitch patterns. This paper uses statistical analysis for identifying the pitch patterns which are characteristic for English sentences spoken by Chinese speakers. Statistical significance of pitch pattern differences between Chinese English and native English is evaluated depending on the variation (implemented as Standard Deviation) of pitch contours. While it is generally acknowledged that Chinese-accented English sounds flat in intonation, statistics in this paper suggest that pitch range in Chinese-accented English is actually wider and Chinese/accented English pitch contours are less flat than in English speakers, indicating that Chinese English speakers have enough (or even more) rising and falling contours in their English sentence intonation.

**Keywords:** Chinese-accented English, pitch contour, native language transfer.

## **1 Introduction**

Despite the fact that the downtrend of the pitch is the main pattern in both English and Chinese declarative sentences, the pitch contour of Chinese utterances is primarily dependent on the lexical tone of each word. In contrast, without the restriction of tone, the movement of the pitch contour in English declarative sentences mostly relies on the intonation of the sentences. Wang (1990) discovered that the speech rate for Chinese is higher than English, as there are 8.17 syllables per second for Chinese speakers and 7.6 for English speakers with 3.07



changes of the pitch contours in Chinese English. However, it is worth noting that this conclusion only applies to the word level with no further analysis of sentences.

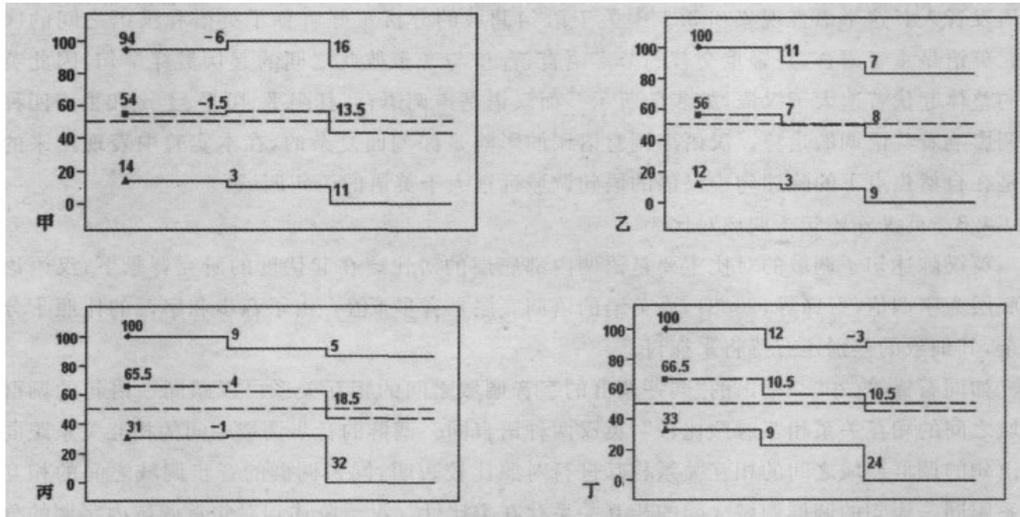


Figure 2. Pitch contours of Chinese declarative sentences (from Guo 2010).

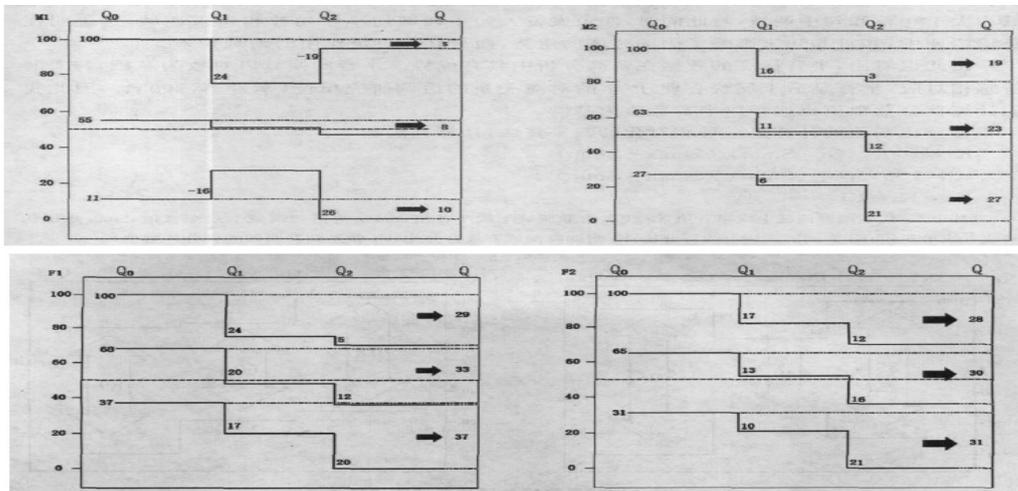


Figure 3. Pitch contours of English declarative sentences (from Guo 2010).

In the present study, the pitch movement contours as produced by Chinese speakers of English are statistically compared with those of native English speakers. Two parameters are used to examine if the pitch contours of Chinese English are indeed more flat than those of natives.

## 2 Background

### 2.1. Material

Three declarative sentences were chosen from the MOCHA-TIMIT corpus (Alan Wrench, 1999) (a database with 460 short sentences intended to allow for investigations of connected speech processes and intonation) with 6, 11, and 16 syllables, respectively:

- (1) This is easy for us.
- (2) She is thinner than I am.
- (3) I gave them several choices and let them set the priorities.

Subjects were divided into several groups according to their mother language; in addition, all the subjects who spoke Chinese English were further divided into two groups based on their English proficiency (subjects in the Chinese Teacher Group had learnt English for over 20 years, and all had a master degree in Language Teaching; the subjects in the Chinese Student Group are college freshman who had studied English for 6 years but still spoke English with an obvious Chinese accent) so as to know if a higher level of English leads to better intonation. Each speaker was required to read all these three sentences, resulting in a total of 33 syllables. Measuring MaxFo and MinFo of each syllable manually, 66 values were determined for each speaker. Thus, 1056 pitch values were obtained from 33 syllables.

### 2.2. Outline of the Analysis Method

Each sample utterance was analysed in Praat (Boersma&Weenink, 2015) to extract fundamental frequency patterns (the fundamental frequency pattern is considered as equivalent to the pitch pattern). The sentences were segmented into syllables manually. Then the maximum and minimum pitch values of each syllable (MaxFo( $\sigma$ ), MinFo( $\sigma$ )) were determined.

Pitch trend lines in the three types of sentences (from simple to compound sentences, with varying length of 6, 11, 16 syllables, respectively) were drawn based on the Mean Fo( $\sigma$ ) of each syllable.

If we regard  $\text{MaxFo}(\sigma)$  and  $\text{MinFo}(\sigma)$  of all syllables in a sentence as a set of data, the standard deviation is one of the ways to examine how spread out the numbers are: a large SD indicates relatively more variation in pitch movements.

### 2.3. Statistical Measure Used in the Analysis

#### 2.3.1. Mean $\text{Fo}(\sigma)$

The mean value of individual syllables from each group was calculated as follows:

$$(4) \quad \text{MeanFo}(\sigma) = \left\{ \sum_{i=1}^N [\text{MaxF0}(i) + \text{MinF0}(i)] / 2 \right\} / N$$

where  $N$  is the number of subjects from each group (see section 3.1)

and  $i$  is the certain syllable produced by each group member

#### 2.3.2. Pitch Range

The pitch range also helps us get an idea of whether the pitch contour of an utterance is flat or not. The pitch range of a sentence is measured as follows:

$$(5) \quad \text{Range} = \left\{ \sum_1^N [\text{MaxF0}(N) - \text{MinF0}(N)] \right\} / N$$

where  $\text{MaxFo}(N)$  is the maximum value of all pitch data per speaker for each sample sentence within a group (in Hz)

and  $\text{MinFo}(N)$  is the minimum value of all pitch data per speaker for each sample sentence within a group in (Hz)

and  $N$  is the number of subjects from each group

A minimum of 75Hz and a maximum of 300Hz was set for male voices, and a range from 100 to 500Hz for female voices to rule out creaky voice phonation.

### 3 Methodology

#### 3.1. Subjects

##### *Native Chinese speakers*

The group of native Chinese speakers consisted of 12 subjects aged between 20 to 35 (6 males, 6 females), and was divided into two subgroups:

Teacher Group (2 male, 2 female), hereafter referred to as *Chi(T)*

Student Group (4 male, 4 female), hereafter referred to *Chi(S)*

##### *Native English speakers*

The group of native English speakers (hereafter referred to as *Ntv*) consisted of 4 subjects aged between 20 to 25 (2 males, 2 females), all college students from the United States.

#### 3.2 Procedure

Each subject was given 5 minutes to practice reading the speech materials before the recording. They were also asked to articulate clearly and to repeat a sentence until the speech sample was satisfactory. No other specific instruction for the pronunciation of English was given to the subjects.

### 4 Results

The data are divided into 3 major parts in accordance with sentence length for analysis.

#### 4.1.1 6-Syllable Sentences

Of the three male groups, the pitch contour of *Chi(S)* group showed the greatest variation in pitch movement, with an SD value of 40.93. For all the female groups, the *Chi(T)* group exhibited the greatest pitch movement variation. There is no discernible pattern in the value of pitch ranges and SD values; however, it does show that Chinese English speakers don't necessary produce sentences with a flatter pitch contour than Native English speakers do. We also noticed some features influencing the trend of the pitch curves: unlike *Chi(T)* and *Ntv*, whose data demonstrate an initial rising pitch followed by a falling pitch at the end of sentences, the major peaks in *Chi(S)* were always near the end of the sentence, which is quite different from the pitch declination pattern for English declarative sentences.

As is indicated in the following figures, the degree of the pitch contours movement is demonstrated by the range bars in Figure 4.

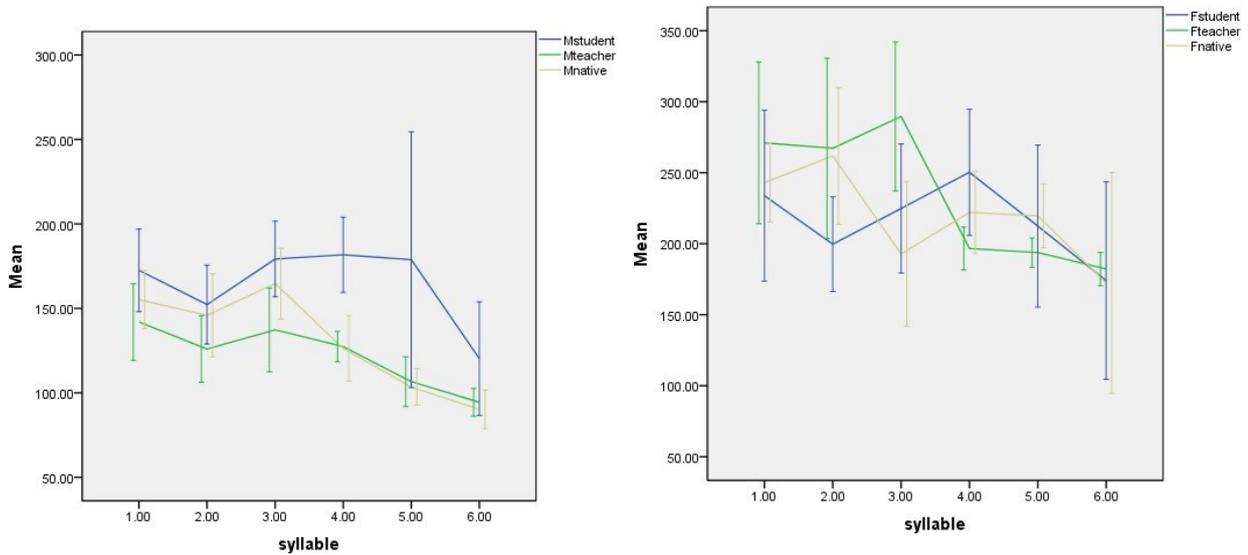


Figure 4. Pitch contours for sentence 1, each bar indicates the range of each syllable (Hz). The left pane represents the pitch contours for males and the right pane represents the pitch contours for females.

#### 4.2. 11-Syllable sentences

A sentence consisting of 11 syllables cannot be regarded as a long one in English, yet its prosody turns out to be easier to master for Chinese students than longer utterances. Despite the irregularity in the mean pitch range values of the 11-syllable sentences uttered by the six groups, the SD numbers do show a certain pattern: for both males and females,  $SD_{Chi}(T\&S) > SD_{Ntv}$ . (average SD value of Chinese male speakers 34.24 versus 22.83 for native male speakers, and Chinese female speakers 44.59 versus 34.33 for native female speakers), indicating that Chinese English is characterized by larger variations in pitch movement than native English.

The degree of the pitch contours movement is demonstrated by the range bars in Figure 5.

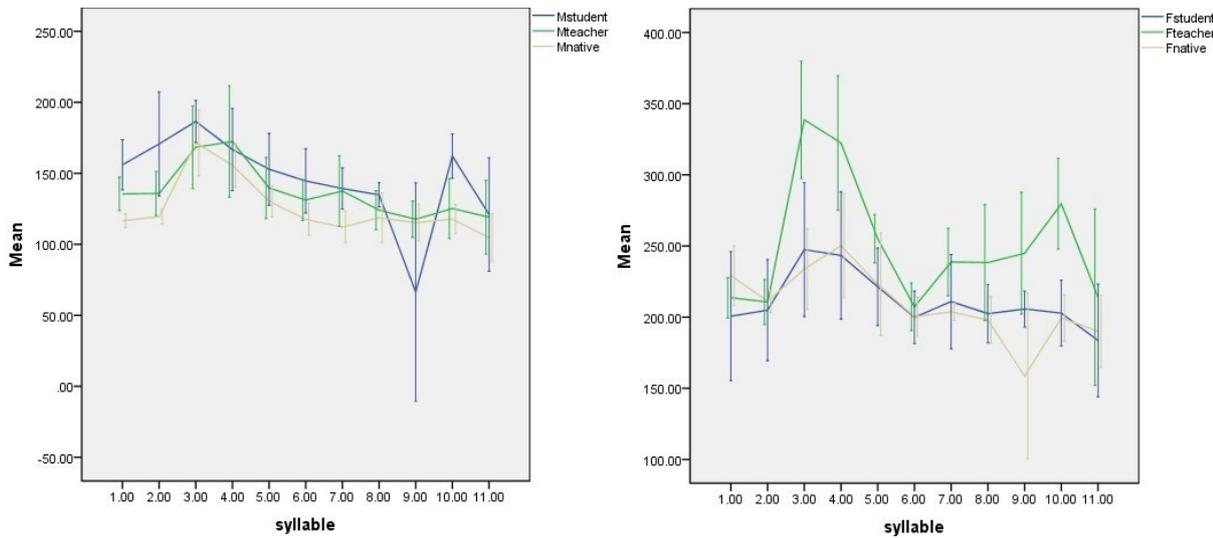


Figure 5. Pitch contours for sentence 2, each bar indicates the range of each syllable (Hz). The left pane represents the pitch contours for males and the right pane represents the pitch contours for females.

With a noticeable large amount of pitch movements indicated by major changes of the bar lengths, the Chinese female teachers exhibit their ability in making the intonation tuneful, with much rising and falling in cadence. In fact, Chinese Female English teachers are very good at producing great pitch variance in their utterance; this may be due to their unique gender, social and occupational characteristics.

#### 4.3 16-Syllable Sentences

The sentence like “I gave them several choices and let them set the priorities” is in fact of medium size in English; however, this compound sentence has two clauses. This prolonged length in prosody makes it difficult for Chinese English speakers to handle. For a sentence with 16 syllables, the pattern is that for both males and females,  $\text{Range } Chi(S\&T) > \text{Range } Ntv$  (average pitch range of Chinese male speakers 136.5 versus 105.93 for native male speakers, and Chinese female speakers 175.4 versus 105.5 for native female speakers), and  $\text{SD } Chi(S\&T) > \text{SD } Ntv$  (Chinese male speakers 33.54 in average versus 26.53 for native males speakers, the Chinese female 46.94 versus 23.07 for the native female). This shows that for this rather long sentence, Chinese English exhibits a greater degree of variability in pitch contours and thus actually sounds less flat than native English.

A comparison of the three situations reveals that Chinese English exhibits larger pitch contour movements than Native English does, and this pattern is more stable as the sentences get longer.

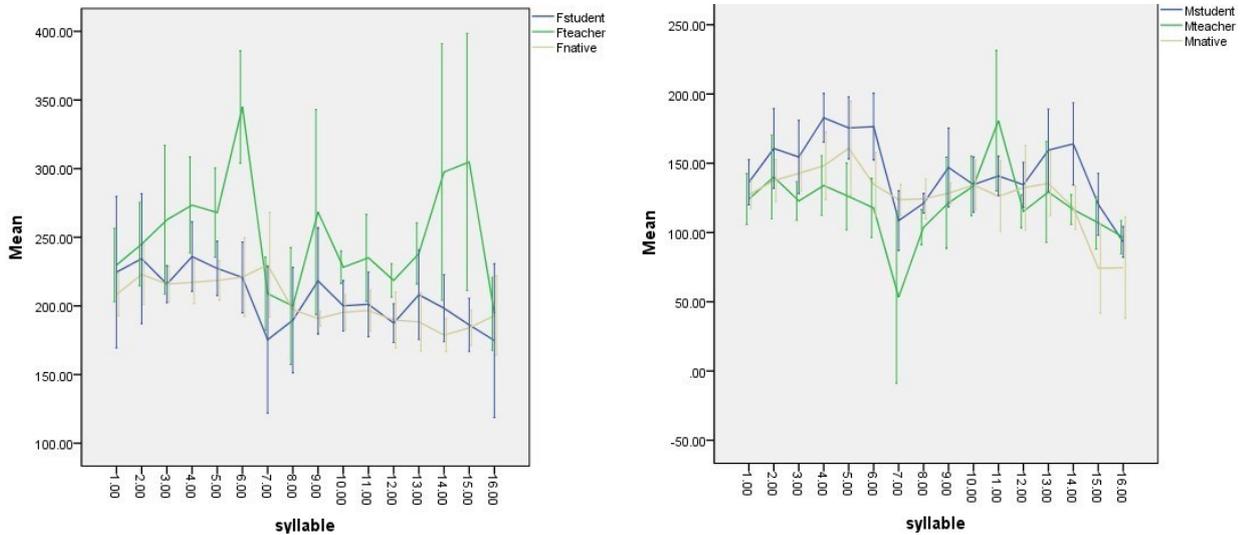


Figure 6. Pitch contours for sentence 3, each bar indicates the range of each syllable (Hz). The left pane represents the pitch contours for males and the right pane represents the pitch contours for females.

A comparison was made between the data produced by the three stimulus: for both Chinese males and females, a continuous increase in the average pitch ranges was observed along with the growth of number of syllables in a sentence; in contrast, pitch ranges in the Ntv group either decrease or remain in the same level for longer sentences; at the same time, average SD values in the Ntv group (in average, 41.59 for the 6-syllables, 28.58 for the 11-syllables and 24.8 for the 16-syllables) also show signs of decreasing, while average SD values in Chinese groups are very close to each other for the three stimulus (in average, 43.74 for the 6-syllables, 39.41 for the 11-syllables and 40.49 for the 16-syllables). For the Ntv groups, the general tendency is that the longer the sentence is, the smaller the SD values are, whereas such a pattern was not observed in either the Chi(T) or the Chi(S) group. We also found that in average, the SD value of Chinese group was close to that of the Ntv group in short sentence (43.74 vs 41.59 in the 6-syllables sentences); however, the gap becomes larger when sentences get longer (39.41 for the Chinese group vs 28.58 for the Ntv group in the 11-syllables and 40.49 vs 24.8 for the 16-syllables sentence). The increasing gaps suggest the difficulties of the learners handling longer sentences.

A Mann Whitney analysis was further carried out on all data from three sample sentences to examine the difference of SD values between Chi(T), Chi(S) and Ntv. The result demonstrates significant differences between female Chi(S), Chi(T) and Ntv ( $p = 0.024$ ,  $df = 1$ ); However, no significant difference was found between female Chi(S) and Chi(T), and neither is there any difference between the male groups. This might suggest that high-level English learners may have the same intonation problems as the beginners do, even if they are better at pronouncing individual phonemes, syllables or words than beginners or intermediate level learners. Hence, I tentatively suggest that Chinese tonal transfer is stronger than any other forms of transfer.

#### 4.4 Other Findings

Two other patterns were observed. First, two or more major peaks show up in the pitch contours of the Chi(T&S) group, in contrast to the Ntv's group, which had only a single major peak per sentence. This is especially evident in longer sentences, as shown in Figures 7 and 8.

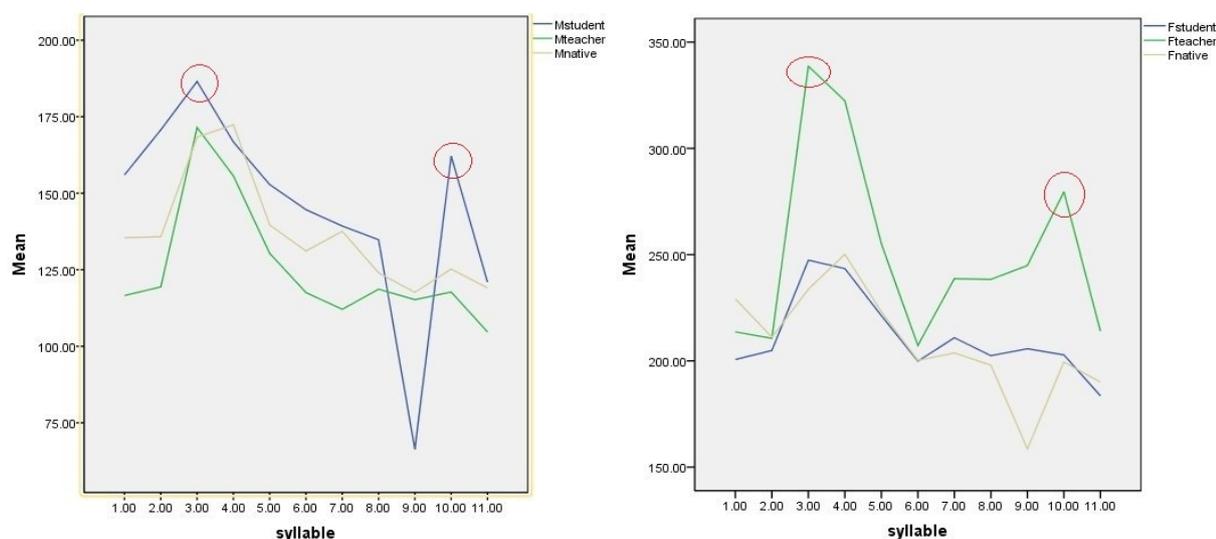


Figure 7. Major peaks shown

For both the male and the female Chinese English groups, an extra one or two major pitch movements appear close to the end of the sentence, while a one-major-peak pattern was observed in native speakers in the initial position. This typical phenomenon might not be caused directly by the transfer of Chinese, as there is not such pitch contours presentation in Chinese declarative sentences either. However, it may be as well caused by the confusion that

Chinese English Learners may have about “sentence flattening out” pattern. As tonal awareness is so deeply rooted that it stops the learners from detecting English pitch contour’s gradual, continuous and integrated change pattern, without knowing which, they would probably assume that there should be an emphasis at the end of a sentence.

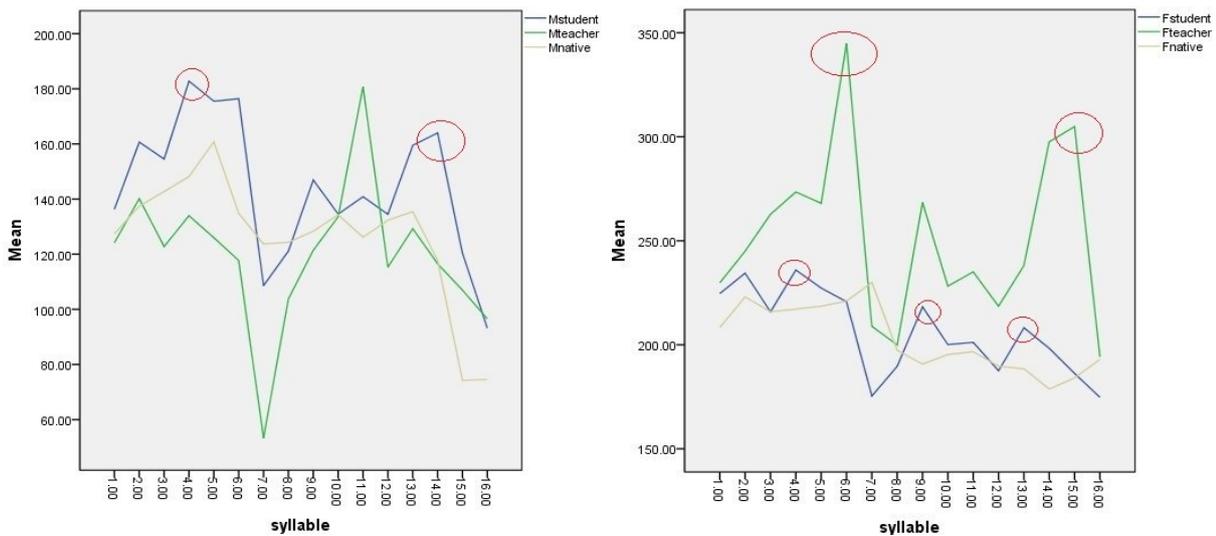


Figure 8. Major peaks shown in pitch contour of 16-syllable sentence (Hz).

Second, for the *Ntv* groups, the length of the variation bars gradually varies with each rising or falling of the contour, as is indicated in Figure 8.

The range bars give us a general idea of how the pitch range of each syllable varies within different groups. As is shown in the above two figures, for native English, we observe longer bars at each peak of the pitch contour whereas shorter bars occur at the valleys. In contrast, changing of the bars’ length shows irregularity for Chinese English intonation as follows.

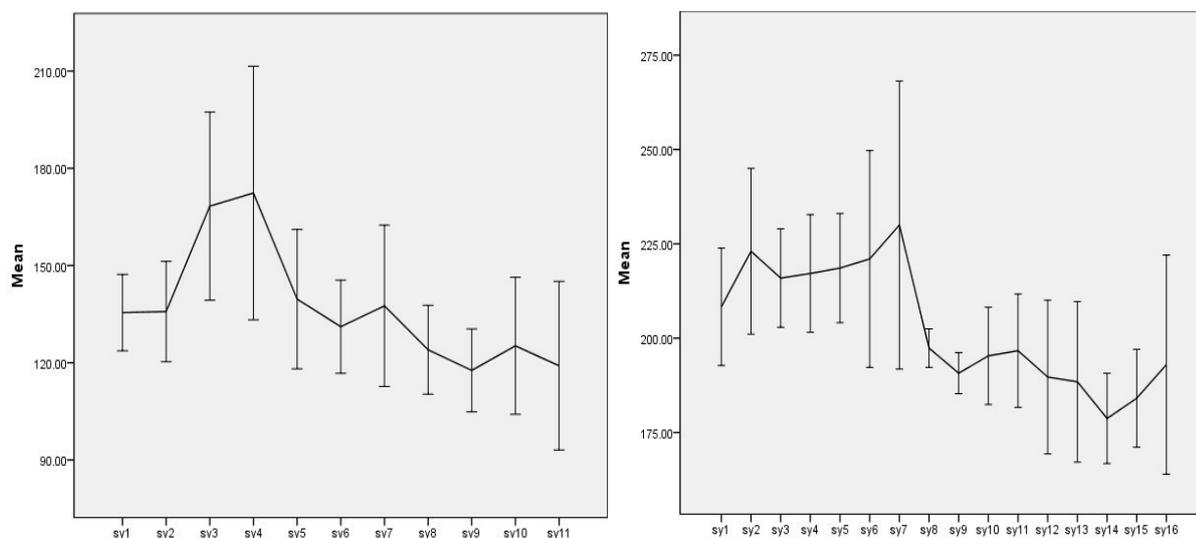


Figure 9. Examples of the native speakers' pitch contours for sentences 2 and 3. Each bar indicates the range of one syllable (Hz).

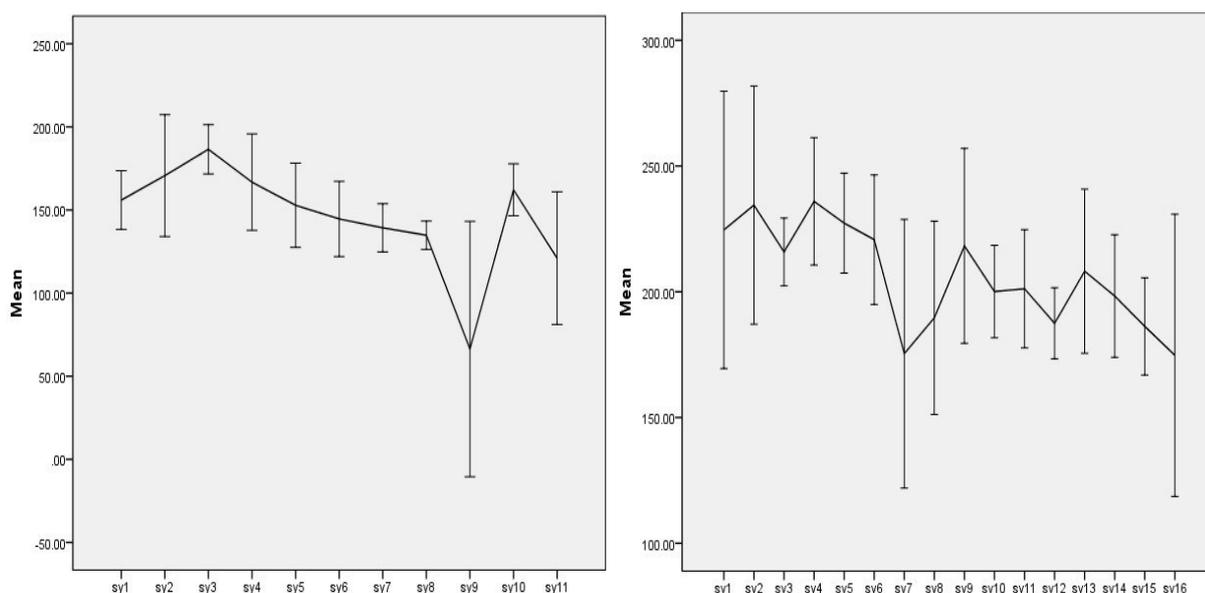


Figure 9. Examples of the Chinese-accented English pitch contours for sentences 2 and 3. Each bar indicates the range of one syllable (Hz).

As is shown in Figures 16 and 17, the bar length exhibits obvious irregularity with cases of bar prolonging at the valleys and shortening at the peaks of the contour. This might be an indication of language transfer of Chinese, a tonal language in which every syllable acquires a tone, in other words, a fixed pitch range. With the restriction of tonal awareness of their

mother language, Chinese English speakers would also put a pitch range on each syllable of an English sentence. Even if they cannot decide what tones go to a syllable, there always has to be a tone for each syllable. Chinese English learners would probably randomly choose some pitch range for each syllable and make them sound like they are also with lexical tones; therefore we won't observe a continuous and gradual variation of the "range bar" in Chinese English like that in native English. In the following post hoc experiment, two sentences with 16 words (syllables) each were uttered by the 12 Chinese subjects. All four tones were included and then organized differently in each sentence as follows:

(6) 我 昨 天 听 见 他 说 声 调 实 验 对 比 很 有 趣。

wǒ zuó tiāntīngjiàn tā shuō shēng diào shí yàn duì bǐ hěn yǒu qù

214 35 55 51 51 55 55 55 51 35 51 51 214 35 214 51

*'Yesterday I heard he said that the tonal experiment was very interesting'*

(7) 你 最 好 不 要 整 天 戴 这 支 黄 眼 镜 看 小 说。

nǐ zuì hǎo bú yào zhěng tiāndài zhe zhī huáng yǎn kàn jìng xiǎo shuō

214 51 214 35 51 21 55 53 51 33 35 21 51 51 21 55

*'You'd better not wear these yellow glasses when reading'*

Each word above was transcribed with the five degree notation (Chao, 1968) for its pitch variation. The following figures exhibit the pitch contours of the two Chinese sentences and the range bar at each syllable.

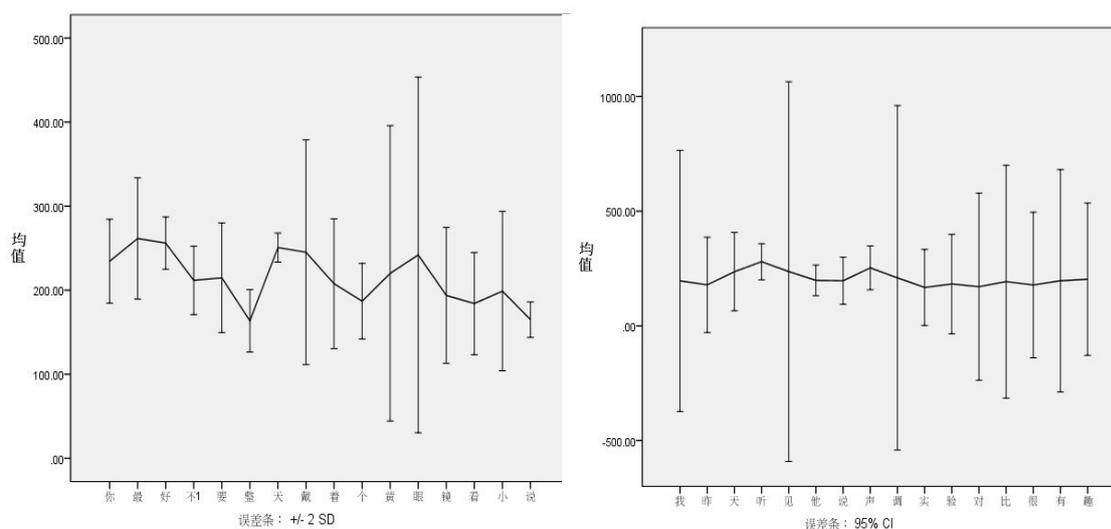


Figure 10. Examples of pitch contours for 16-syllable Chinese sentences. Each bar indicates the range of one syllable (Hz).

## 5. Conclusion

In the discussion of the intonation of declarative sentences, we found that the idea that Chinese English sounds flat might be very misleading; in fact, the longer the sentence is, the more likely it is that Chinese English speakers utter it with wider pitch ranges and greater degree in pitch movements than native speakers do. Chinese English teachers, especially the females, are likely to use more rising and falling contours in cadence for classroom language, which may contribute to the largest variability in pitch contours in the female teacher group in all three cases. In addition, unlike native English, the pitch contour of Chinese English shows no sign of flattening out when a sentence gets longer, and there is typically another major peak near the end of the sentence. In Chinese, every word in a sentence is pronounced with a set pitch value, which feature is likely transferred into Chinese English. In other words, with the restriction of tonal awareness, Chinese English speakers usually tend to put a random pitch range on each syllable of an English sentence and thus make it sound like it is also has a lexical tone. Groups of English learners with different levels showed no significant difference in terms of pitch contour handling; this might be suggesting that tonal transfer could be very strong that it happens on different levels of Chinese English learners.

## References

- Boersma, Paul & Weenink, David (2015). Praat: doing phonetics by computer [Computer program]. Version 5.2.02, retrieved 17 October 2010 from <http://www.praat.org/>
- Bolinger, Dwight (1958). A theory of pitch accent in English. *Word* 14(2/3):109-149.
- Chao, YuenRen.(1930) A system of tone letters. *Le Maître Phonétique* 45: 24-27.
- Guo, Jia. (2010). An analysis of the declarative intonation in English. *NanKai Linguistics*, 2:57-69.
- Gao, Wei. (2013). The realization of prominence in English learner intonation. *Foreign Languages and the Teaching* 1: 56-58.
- Ladefoged, Peter. (1984) 'Out of chaos comes order': Physiological, biological, and structural patterns in phonetics. In M.P.R. Van den Broeke and A. Cohen (eds.) *Proceedings of the Tenth International Congress of Phonetic Sciences*. Dordrecht: Foris Publications 83-95.
- Nariai, Tomoko & Tanaka, Kazuyo (2010). A study of pitch patterns of sentence utterances by Japanese speakers of English in comparison with native speakers of English. *L2WS-2010*.
- Stephen J, Eady (1982). Differences in the *F<sub>0</sub>* pattern of speech—tone language versus stress language. *Language and Speech*, 25:29-42.
- Wang, GuiZhen (1990). Characteristics of *F<sub>0</sub>* Pattern in the Language of Chinese and English. *Foreign Language* 1: 40-46.
- Wrench, Alan. The MOCHA-TIMIT Articulatory Database. <http://www.cstr.ed.ac.uk/artic/mocha.html>, Queen Margaret University College, 1999.
- Xu, Yi.(1999). Effects of tone and focus on the formation and alignment of *F<sub>0</sub>* contours. *Journal of Phonetics* 27:55-105.

**Appendix 1**

*Mean Fo-values of each of 6 syllables, pitch range and SD of Sentence 1 uttered by Male & Female Chi(S), Chi(T) and Ntv (Hz).*

	Syllables	Mean(M)	Mean(F)	Range(M/F)	SD(M/F)
Chi(S)	1	172.53	233.83	142(M)	40.93(M)
Chi(S)	2	152.25	199.73	142.15(F)	53.20(F)
Chi(S)	3	179.25	224.83		
Chi(S)	4	181.73	250.3		
Chi(S)	5	178.83	212.4		
Chi(S)	6	120.3	174		
Chi(T)	1	141.94	271.03	64.72(M)	23.16(M)
Chi(T)	2	125.96	267.2	175.75(F)	57.65(F)
Chi(T)	3	137.28	289.6		
Chi(T)	4	127.4	196.6		
Chi(T)	5	106.65	193.75		
Chi(T)	6	94.41	182.2		
Ntv.	1	155.18	243.03	84.19(M)	31.94(M)
Ntv.	2	145.9	261.78	151.15(F)	51.24(F)
Ntv.	3	164.75	192.8		
Ntv.	4	126.33	222.05		
Ntv.	5	103.45	219.6		
Ntv.	6	90.2	172.4		

**Appendix 2**

Table2: Mean Fo-values of each of 11 syllables, pitch range and SD of Sentence 2 uttered by Male& Female Chi(S), Chi(T) and Ntv (Hz).

	Syllables	Mean(M)	Mean(F)	Range(M/F)	SD(M/F)
Chi(S)	1	155.98	200.65	166.16(M)	42.21(M)
Chi(S)	2	170.7	204.88	107.65(F)	35.75(F)
Chi(S)	3	186.53	247.43		
Chi(S)	4	166.75	243.4		
Chi(S)	5	152.83	221.28		
Chi(S)	6	144.63	199.88		
Chi(S)	7	139.28	210.9		
Chi(S)	8	134.8	202.45		
Chi(S)	9	66.35	205.73		
Chi(S)	10	162.15	202.83		
Chi(S)	11	120.99	183.55		
Chi(T)	1	135.48	213.65	84.42(M)	26.27(M)
Chi(T)	2	135.8	210.6	221.95(F)	53.42(F)
Chi(T)	3	168.33	338.65		
Chi(T)	4	172.35	322.4		
Chi(T)	5	139.63	255.05		
Chi(T)	6	131.13	207.28		
Chi(T)	7	137.55	238.7		
Chi(T)	8	124	238.35		
Chi(T)	9	117.63	244.9		
Chi(T)	10	125.23	279.63		
Chi(T)	11	119.07	214		
Ntv.	1	116.58	229.18	105.4(M)	22.83(M)
Ntv.	2	119.4	211.2	131.4(F)	34.33(F)
Ntv.	3	171.48	233.73		
Ntv.	4	155.68	250.2		
Ntv.	5	130.35	223.13		
Ntv.	6	117.55	200.25		
Ntv.	7	112.07	203.75		
Ntv.	8	118.63	197.98		
Ntv.	9	115.22	158.65		
Ntv.	10	117.73	199.48		
Ntv.	11	104.69	190		

## Appendix 3

Table 3: Mean *F<sub>0</sub>*-values of each of 16 syllables, pitch range and SD of Sentence 1 uttered by Male & Female Chi(S), Chi(T) and Ntv(Hz).

	Syllable	Mean(M)	Mean(F)	Range(M/F)	SD(M/F)
Chi(S)	1	136.28	224.6	114(M)	31.45(M)
Chi(S)	2	160.68	234.43	130(F)	36.04(F)
Chi(S)	3	154.5	215.85		
Chi(S)	4	182.78	235.95		
Chi(S)	5	175.5	227.28		
Chi(S)	6	176.43	220.73		
Chi(S)	7	108.56	175.38		
Chi(S)	8	121.13	189.65		
Chi(S)	9	146.95	218.25		
Chi(S)	10	134.55	200.1		
Chi(S)	11	140.85	201.18		
Chi(S)	12	134.45	187.48		
Chi(S)	13	159.48	208.18		
Chi(S)	14	164.03	198.3		
Chi(S)	15	120.4	186.23		
Chi(S)	16	93.2	174.73		
Chi(T)	1	124.15	229.78	159(M)	35.62(M)
Chi(T)	2	140.13	245	220.8(F)	57.84(F)
Chi(T)	3	122.78	262.75		
Chi(T)	4	133.98	273.48		
Chi(T)	5	126	268		
Chi(T)	6	117.7	344.9		
Chi(T)	7	53.3	208.9		
Chi(T)	8	103.79	199.93		
Chi(T)	9	121.43	268.4		
Chi(T)	10	133.58	228.2		
Chi(T)	11	180.75	235.05		
Chi(T)	12	115.28	218.55		
Chi(T)	13	129.26	238.1		
Chi(T)	14	116.55	297.58		
Chi(T)	15	106.95	304.88		
Chi(T)	16	96.58	194.25		
Ntv.	1	127.33	208.3	105.93(M)	26.53(M)
Ntv.	2	137.45	223.03	105.5(F)	23.07(F)
Ntv.	3	142.73	215.93		
Ntv.	4	148.18	217.15		
Ntv.	5	160.78	218.58		
Ntv.	6	134.83	221		
Ntv.	7	123.7	229.98		
Ntv.	8	124.33	197.35		
Ntv.	9	128.33	190.73		
Ntv.	10	134.3	195.33		
Ntv.	11	126.15	196.68		
Ntv.	12	132.33	189.7		
Ntv.	13	135.45	188.43		
Ntv.	14	117.85	178.73		
Ntv.	15	74.21	184.08		
Ntv.	16	74.59	192.98		