

The University of Arizona Electronic Theses and Dissertations  
Reproduction and Distribution Rights Form

Name (Last, First, Middle) Woods, Anna, Christina	
Degree title (eg BA, BS, BSE, BSB, BFA): BA	
Honors area (eg Molecular and Cellular Biology, English, Studio Art): Linguistics	
Date thesis submitted to Honors College: May 5, 2010	
Title of Honors thesis: Perception of an Allophonic Distinction from Conversational Speech.	
:The University of Arizona Library Release	I hereby grant to the University of Arizona Library the nonexclusive worldwide right to reproduce and distribute my dissertation or thesis and abstract (herein, the "licensed materials"), in whole or in part, in any and all media of distribution and in any format in existence now or developed in the future. I represent and warrant to the University of Arizona that the licensed materials are my original work, that I am the sole owner of all rights in and to the licensed materials, and that none of the licensed materials infringe or violate the rights of others. I further represent that I have obtained all necessary rights to permit the University of Arizona Library to reproduce and distribute any nonpublic third party software necessary to access, display, run or print my dissertation or thesis. I acknowledge that University of Arizona Library may elect not to distribute my dissertation or thesis in digital format if, in its reasonable judgment, it believes all such rights have not been secured. Signed: <u>Anna Woods</u> Date: <u>5-5-10</u>

PERCEPTION OF AN ALLOPHONIC DISTINCTION  
FROM CONVERSATIONAL SPEECH

By

ANNA CHRISTINA WOODS

---

A Thesis Submitted to The Honors College

In Partial Fulfillment of the Bachelors degree  
With Honors in

Linguistics

THE UNIVERSITY OF ARIZONA

May 2010

Approved by:



Dr. Natasha Warner

Department of Linguistics

## I. INTRODUCTION

Mandarin Chinese has two affricates that can be confusing to English learners. The first is represented in the Pinyin romanization system as *j* and the second as *zh*. In the International Phonetic Alphabet (IPA,) they are represented as respectively [tɕ] and [tʂ] (Lee 2003). [tɕ], henceforth referred to as *j*, is a laminal alveolo-palatal affricate, pronounced with the front part of the tongue obstructing the flow of air at the anterior portion of the hard palate (Dow 1972). [tʂ], henceforth referred to as *zh*, is an apical post-alveolar affricate, made by raising the tip of the tongue against the anterior part of the hard palate behind the alveolar ridge (Dow 1972).

### A. REASONS FOR CONFUSION BETWEEN J AND ZH

#### i. SIMILARITY TO ENGLISH SOUNDS

Some English learners have trouble with these sounds because they are so similar in articulation to English affricates. Learners often will substitute their own English *dr* for Mandarin's *zh* (Dow 1972). This confusion was traced by Dow to the tongue and lips. The cavity between the palate and the tongue must be more hollow and the tip of the tongue slightly more retroflexed for *zh* than for *dr* or the Mandarin *j*. For *j*, the tip of the tongue should be behind the lower front teeth, not up by the alveolar ridge as is done with English affricates (Dow 1972). For both *j* and *zh*, the lips should, of course be rounded when the following vowel is rounded, but the lips should be in a natural or slightly spread position with all other vowels. It is common to pronounce English affricates like *dr* or *tr* with some amount of lip rounding in every case (Dow 1972).

#### ii. ROMANIZATION OF STANDARD MANDARIN CHINESE

Another evidence of confusion between *j* and *zh* may be found in the romanization systems of Mandarin Chinese that were created for English-speaking students. Often, these systems do not distinguish between palatal consonants (which includes *j*) and the retroflex consonants (which

includes *zh*) (Norman 1988). For example, in the Wade-Giles romanization system, developed in 1859 by Thomas Francis Wade, *zh* and *j* are represented with the same letters: *ch*. The Wade-Giles system was designed by a Cambridge professor of Chinese and native speaker of English, and the choice of symbols used was thus influenced by English pronunciation (Tao 1991.) In the Wade-Giles system, the only difference made between *j* and *zh* is the context in which they appear. *j* only appears before high front vowels (like [i y]), whereas *zh* can appear before a variety of other sounds, but never high front vowels. Although to English speakers, the two sounds may seem the same, they are different, and learners who use more ambiguous romanization systems such as Wade-Giles may find difficulty in learning to distinguish between the *j* and *zh*.

Native speakers of Mandarin are told to learn the distinction between *j* and *zh*, as demonstrated by two different phonetic systems created by native speakers. One system is called Zhuyin fuhao, or BoPoMoFo, and uses pictorial symbols to represent each sound. *j* and *zh* are given different symbols in this system. Hanyu Pinyin is the system used in the People's Republic of China's elementary schools to help teach pronunciation (Tao 1991.) This system also delineates between *j* and *zh*, and is the system used in this paper to represent all tokens.

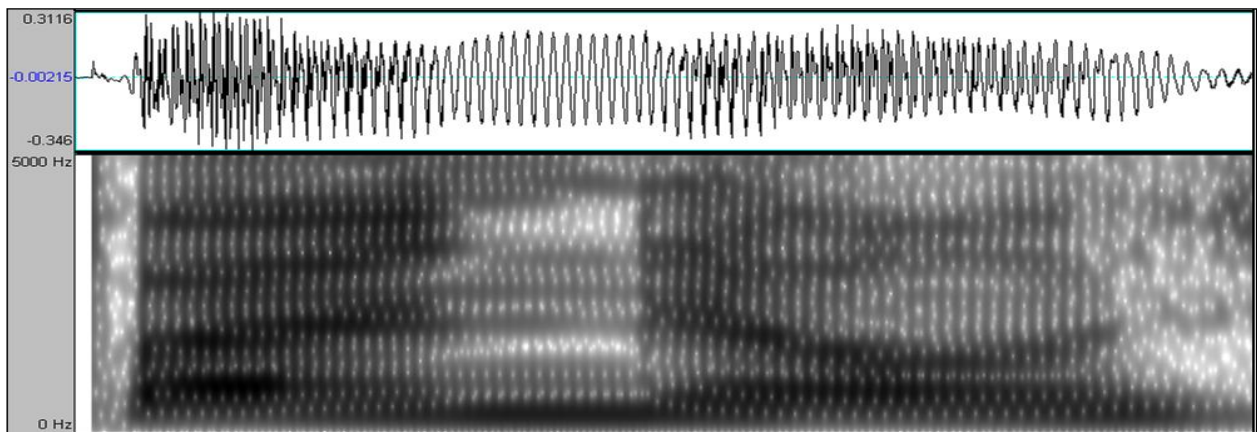
### iii. PHONOLOGICAL STATUS OF J AND ZH

The palatal series of consonants in Mandarin (Pinyin *j q x*) are in complementary distribution with three different series of consonants: the dental sibilants (*z c s*), the retroflexes (*zh ch sh*) and the velars (*g k h*). Each consonant in those three groups never occurs before high front vowels, and the palatals only occur before high front vowels. Many scholars present arguments for which two groups actually form an allophonic pair, but no definitive decision can be made because of a lack of proof in the phonemic analysis of the four groups (Norman 1988). Although *j* and *zh* are taught as different sounds to native speakers and learners of Mandarin alike, the apparent allophony raises many questions about the underlying representation of the two sounds. One might wonder if they are

underlyingly the same sound, but are pronounced differently based on their conditioning environment, or if they indeed need to be taught as distinct sounds in order to reduce confusion.

#### iv. PERCEPTION OF CONVERSATIONAL SPEECH

Conversational speech may also have an effect on listeners' ability to correctly perceive the difference between *j* and *zh*. When speech is produced spontaneously in conversation, sounds can be lost or changed to very different ones (Johnson 2004). In this case, the affricate *zh* is sometimes even articulated like a nasal or glide (see Fig. 1).



/ b          a                  n          tʂ          a                                  ŋ          /  
 [ b          a                  n          j                  ã          ]

Fig. 1 Spectrogram of Pinyin: *bān zhǎng* (English: The section leader).

The conditioning high-front vowel environment of *j* may also become very unclear or deleted in spontaneous speech, making correct perception of allophones *j* and *zh* seem unlikely. Without cues from the conditioning environment, these allophones may seem nearly identical, especially to English speakers who are also influenced by *j* and *zh*'s closeness of articulation, and any romanization systems that do not clearly distinguish between the two sounds.

## B. STUDY OF MANDARIN NATIVE SPEAKERS' PERCEPTION OF J AND ZH

In order to further investigate the effects of spontaneous speech production in perception of *j* and *zh*, native speakers of Mandarin were invited to participate in a study. With the combination of reduction in the conditioning environment and of the affricates themselves, even Native speakers of were expected to have trouble perceiving the difference between the two. The extent to which this is true is explored in the following experiment.

## II. METHOD

### A. SUBJECTS

5 native speakers of Standard Mandarin were recruited from the University of Arizona student body. The speakers stated they spoke Standard Mandarin (not another specific dialect) predominately in the home growing up and at school. Subjects were recorded in a sound-proof booth while having a conversation with a friend in another location over the phone. Phone conversations were used to elicit natural, spontaneously produced speech.

10 female and 7 male subjects participated in a perception experiment following the collection of the phone conversations. Each subject reported having grown up in China or Taiwan<sup>1</sup> and are currently studying in the United States. All participants either predominately spoke standard Mandarin in the home and/or at school, or speak it regularly now in daily life.

### B. STIMULI

112 *j*-initial and 112 *zh*-initial syllables were extracted from the native speakers' phone conversations. The syllables were either monosyllabic words or a syllable from a bi- or tri-syllabic word in Mandarin. Each *j*- syllable chosen had a possible counterpart in Mandarin that started with

---

<sup>1</sup> Subjects reported their hometowns to be in locations such as Hubei, Southwest China, Anhui, Nanjing, Qingdao, Zhejiang, Sichuan, Shanghai, Hunan, Fujian, Heilong jiang, and southern Taiwan. The Zhuyin fuhao (commonly known as BoPoMoFo) phonetic system was used to administer the experiment for the Taiwanese speaker, because pinyin was not standard during the time of this speaker's education.

*zh* and had the same vowel coda except for the high front vowel. This was done to examine whether the high front vowel's reduction or deletion would make the *j*- and *zh*-syllables sound very similar.

See table 1 for a list of syllables used in the experiment. Acoustic measurements will also be done on each syllable's vowel coda to further examine the extent of the effect of reduction on the presence of the high-front vowel in *j*-syllables. See Figures 2-4 for examples of this reduction.

Table 1: *j*- and *zh*-syllables used in perception experiment matched side by side with corresponding vowel coda and tone. The second column from the left and the last column show how many different instances of each token were found and used. Some syllables were used despite not having a syllable with corresponding tone, but there are equal amounts of *j*- and *zh*-syllables overall.

jiang3	27	zhang3	11
jiao4	16	zhao4	4
jiao1	8	zhao1	2
ju4	5	zhu4	6
jiao3	4	zhao3	14
jiong3	4	zhong3	27
jiang1	3	zhang1	12
jian3	2	zhan3	2
jiu4	29	zhou4	0
jia1	9	zha1	0
jia4	3	zha4	0
jiong1	0	zhong1	15
jiong4	0	zhong4	3
jiao2	0	zhao2	2
jun3	0	zhun3	1
jian1	0	zhan1	1

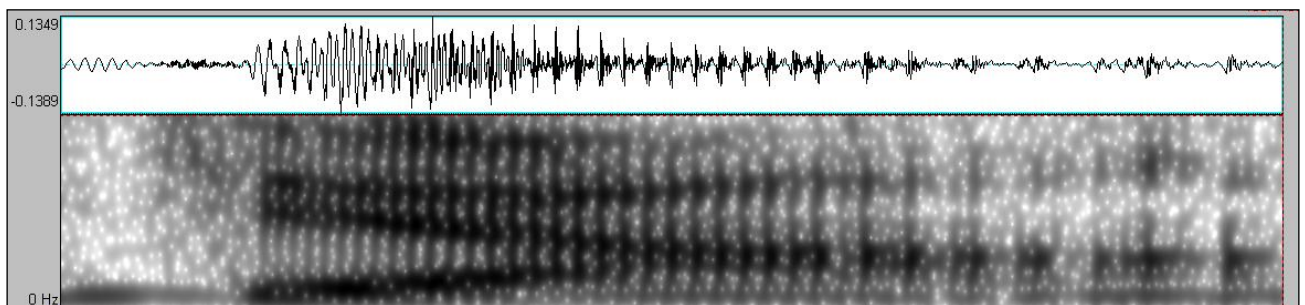


Fig. 2 Spectrogram of more carefully pronounced jiang token taken from conversation.

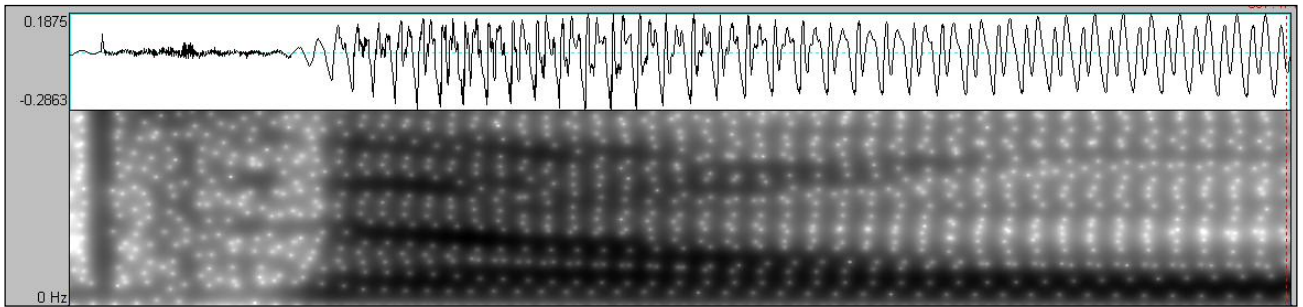


Fig. 3 Spectrogram of more reduced jiang token taken from conversation.

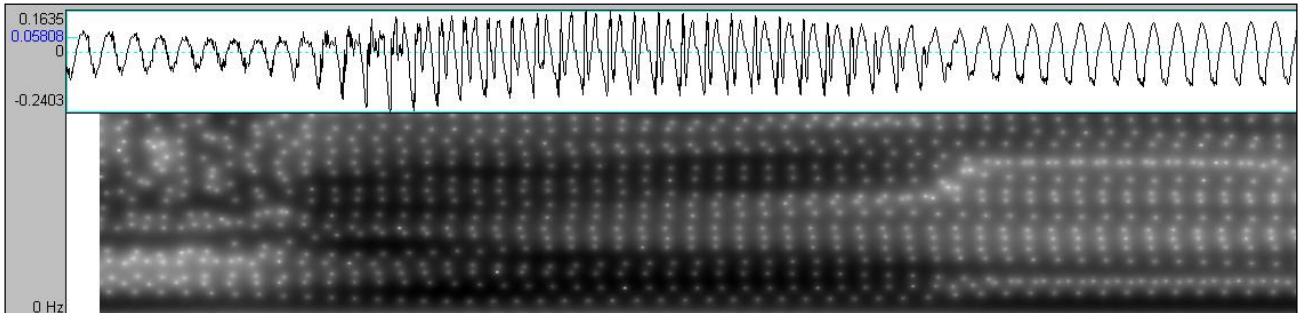


Fig. 4 Spectrogram of reduced zhang token taken from conversation.

The beginning boundary of each token was set at the offset of strong second formant of the previous word, at a drastic change in amplitude, or halfway through the rise of the second formant into the next word. The end boundary for the syllable tokens were set at the end of strong second formant, or for syllables ending in a nasal consonant, the end of the nasal structure in the first formant.

112 *j*-consonants and 112 *zh*-consonants were also extracted from the conversations. For the consonants, the beginning boundary was set with the same criteria as the syllable boundaries, and the end boundary was set at the onset of clear second formant of the syllable, or at any change in the amplitude or formant structure. This was done in an attempt to ensure that any trace of the conditioning high front vowel would not be present in the stimulus. Some examples of boundary placement can be seen below.

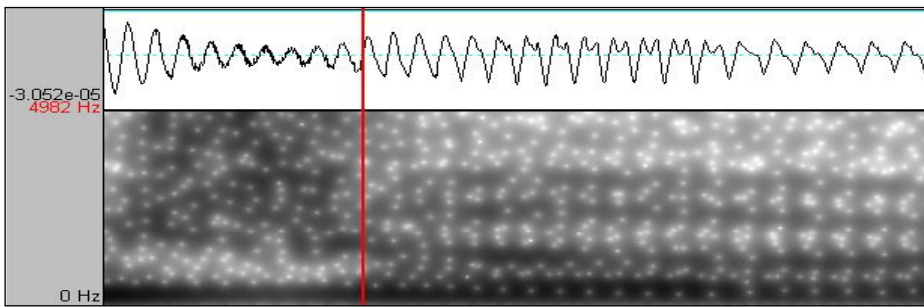


Fig. 5 *zhu* token. The red line shows where the end of the consonant boundary was placed. Note the sudden change in amplitude and stronger formant structure. The beginning of the consonant boundary was placed at the offset of strong second formant.

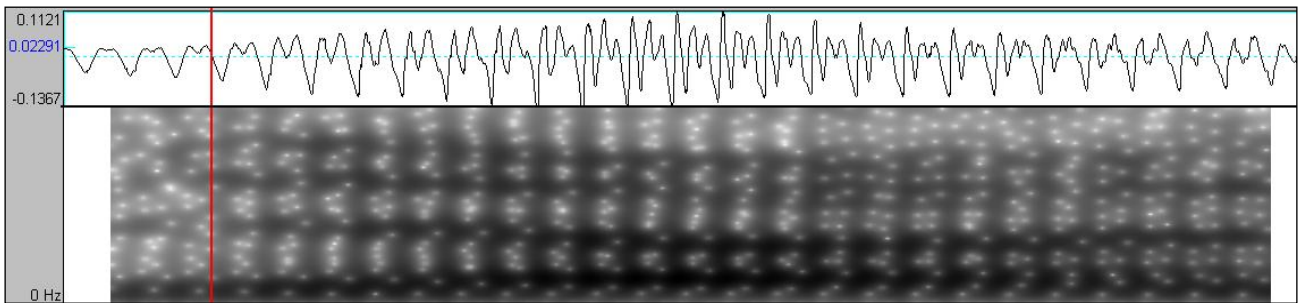


Fig. 6 *jiao* token. The red line shows where the end of the consonant boundary was placed. Again, note the sudden change in amplitude and stronger formant structure. The beginning of the consonant boundary was placed at the offset of strong second formant.

### C. PROCEDURE

Native speakers of Mandarin participated in an experiment modeled after Logan et. al.'s 1991 training experiment for native speakers of Japanese learning to distinguish between English *r* and *l*. There were three blocks in the experiment: pretest, training with feedback, and post-test. This model was used in order to compare the native speakers' results with a future similar training experiment for native English speakers.

Participants were seated in a sound-proof booth and listened to each stimulus, then responded whether they heard *j* or *zh* by pressing a button. They responded to consonant-only stimuli in the first block of the experiment, and then were played whole-syllable stimuli in the second block. During the syllable block, participants responded whether they thought the syllable started with a *j* or *zh*.

Each block consisted of a pretest, training session and post-test. During each training session, participants were shown a screen that said either “correct” or “incorrect,” giving them feedback after their response. The pretest and post-test did not include such feedback, but simply proceeded to the next stimulus after visual confirmation was given that their answer was recorded. Participants' answers and reaction times were recorded for each block.

The pretest and training sessions both consisted of 30 randomized stimuli, and the post-test was given in two parts, each part containing 82 randomized stimuli. Each experiment block had the same amount of stimuli from each of the five talkers, and each block had equal amounts of *j*- and *zh*-tokens. The experiment lasted about 20 minutes, and subjects were paid five dollars as compensation.

### III. RESULTS

Figure 7 shows the proportion *j*-responses to the context in which the syllable appeared. A two-factor within-subjects ANOVA was conducted, with the factors context (consonant only, syllable) and consonant (*j*, *zh*). The dependent variable was the proportion *j*-responses (whether to *j* or *zh* stimuli). The main effect of consonant was significant ( $F(1,16)=159.61, p<.001$ ), as was the interaction ( $F(1,16)=215.31, p<.001$ ). There was no main effect of context ( $F(1,16)=1.82, p>.1$ ). Because of the significant interaction, the simple effect of consonant was tested for each context separately. Listeners responded significantly to *j* and *zh* when they heard the syllable context ( $F(1,16)=582.16, p<.001$ ). However, they could not distinguish the difference if they heard only the consonant ( $F(1,16)=1.72, p>.1$ ).

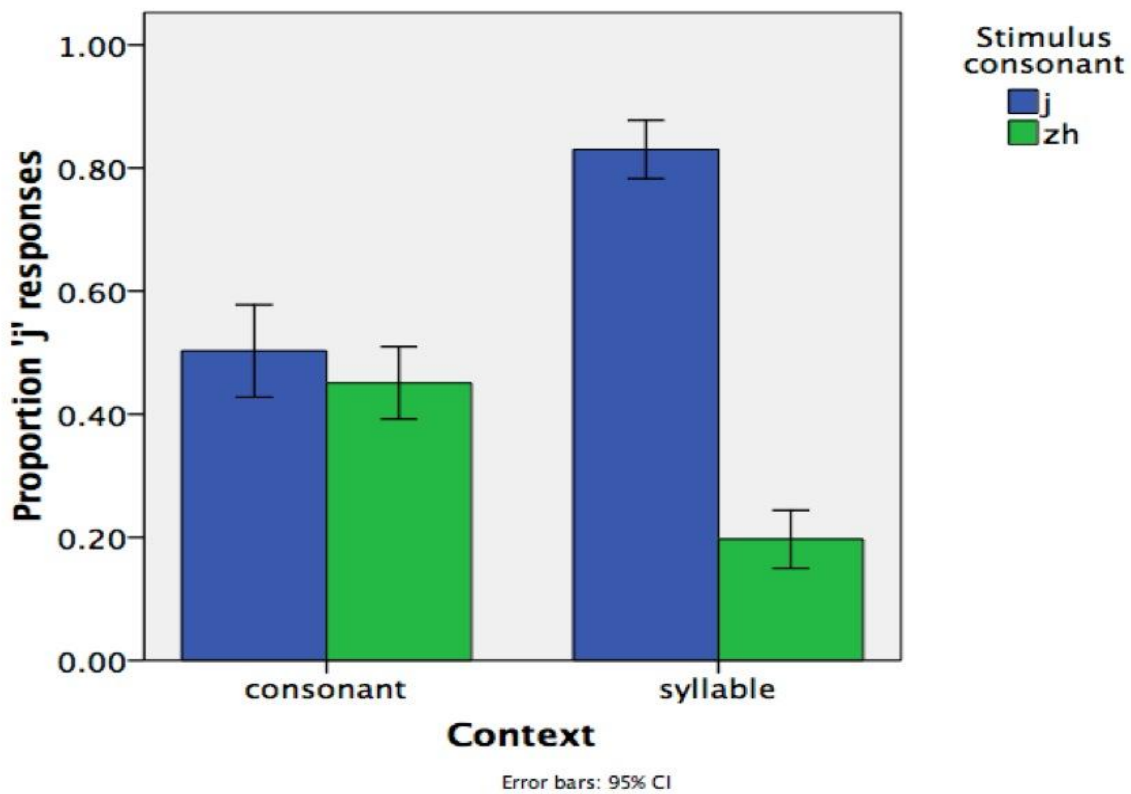


Fig. 7 Proportion *j*- responses to *j* and *zh* stimuli, for C-only vs. whole syllable stimuli.

Figure 8 shows the proportion of correct responses in the syllable context to the log of word frequency. This graph shows that for syllables starting with *j*, a higher word frequency in the stimulus corresponded to lower accuracy in determining the difference between *j* and *zh*. This may have been due to deletion or reduction of the conditioning high front vowel or of the consonant itself in high-frequency *j*-tokens. There was a significant correlation between the accuracy of response and log of word frequency ( $r = -.54$ ,  $p < .001$ ) for *j*-tokens. There was no significant correlation between the accuracy of response and log of word frequency ( $r = -.14$ ,  $p > .1$ ) for *zh*-tokens.

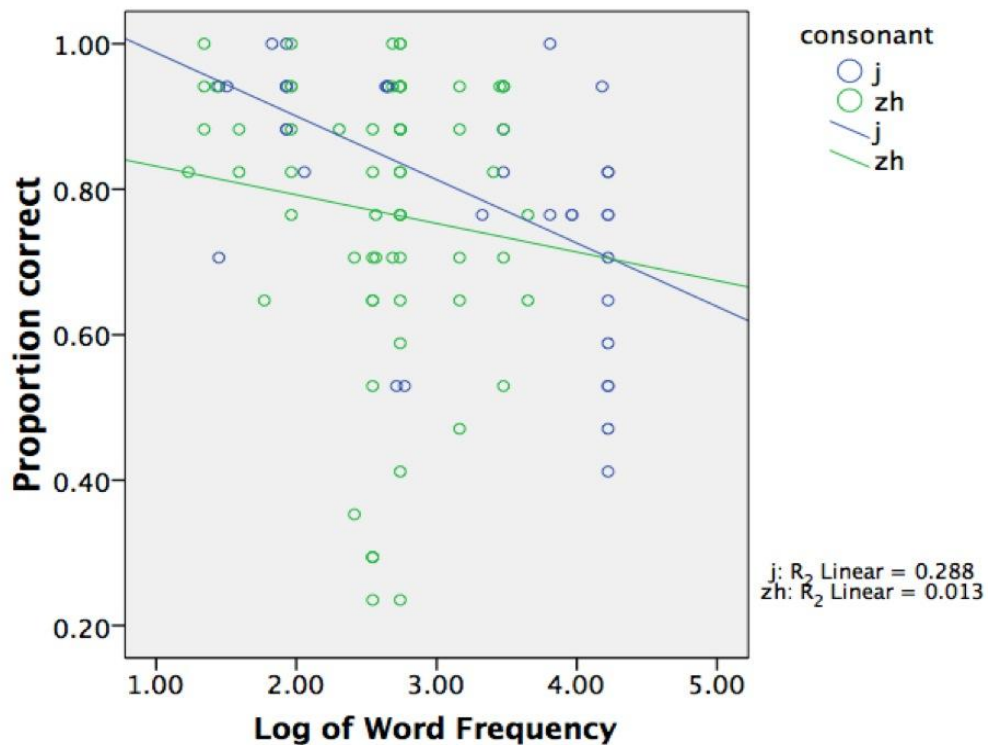


Fig. 8 Proportion correct by log of word frequency (Chinese Community Information Center)

Figure 9 shows the proportion of syllable duration in milliseconds to the log of word frequency. For *j*-tokens, there was a significant correlation between syllable duration and log of word frequency ( $r=-.48$ ,  $p<.005$ ). This may have been due to shortening of the syllable in reduced speech. Spontaneously produced tokens are often shorter and contain less clear articulatory information, supporting the explanation that they are more reduced (cf. Pluymaekers et al. 2005). For *zh*-tokens, there was no significant correlation between syllable duration and log of word frequency ( $r=-.11$ ,  $p>.1$ ).

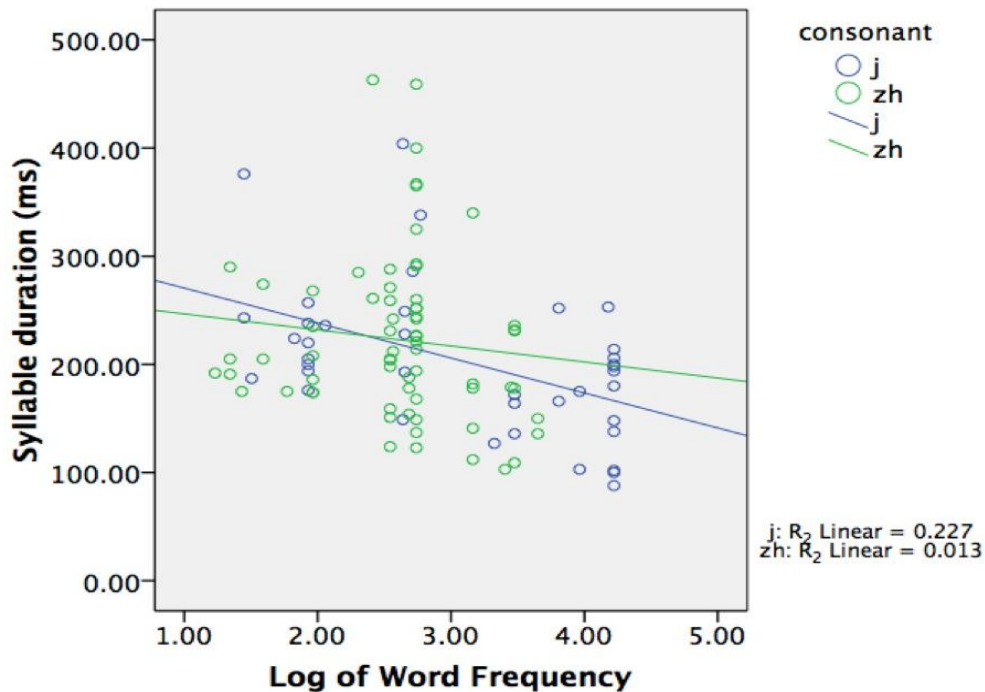


Fig. 9 Word frequency by syllable duration.

Figure 10 shows the proportion *j*-responses to the author's rating of perceptibility of the conditioning environment. A one-factor ANOVA with degree of perceived deletion of conditioning environment (clear, less clear, little perception or deleted) as the within-subjects factor was performed. Only *j*-words were included in this test because the conditioning environment only existed for them. This graph shows that when the conditioning high front vowel was more clear, subjects were able to correctly identify *j*-tokens almost 90% of the time. When the conditioning high front vowel was deleted or only slightly perceptible, subjects were more likely to misperceive *j*-tokens as *zh*. The statistical difference between clear and deleted tokens was significant ( $F(3,48)=10.39, p<.001$ ).

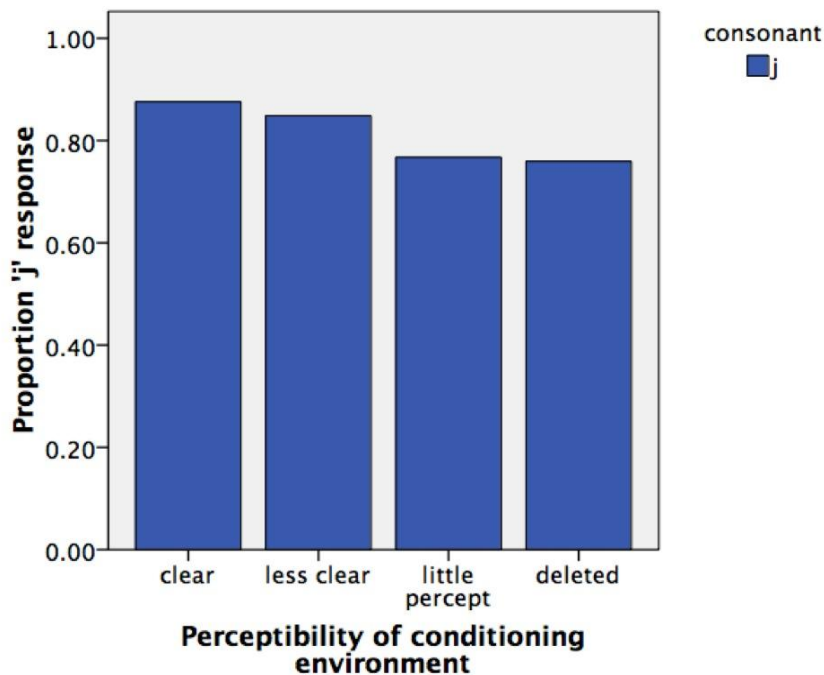


Fig. 10 Effect of reduction of conditioning [i y] on perception of *j* and *zh*.

#### IV. DISCUSSION

In this study, high word frequency and short duration, two possible indicators of reduced speech were shown to correlate with decreased ability to correctly perceive the consonants *j* and *zh*. Significant differences in perception ability were also recorded when listeners were presented with the consonant alone, and when they were given the whole syllable. Other factors affecting perception were also investigated, such as the perceived presence or absence of the high-front vowel following *j*-tokens.

From the information gathered, it may be the case that native Mandarin speakers use the following vowel more than the consonant to distinguish *j* and *zh*. It might also be the case that they used a combination of vowel and consonant to make their decision. Even when listening to highly reduced syllable tokens, speakers were still able to more accurately distinguish between *j* and *zh* than they were when listening to the consonant alone. This may support Steven's 2002 claim that

speech is perceived by extracting information from the regions around abrupt changes in speech sounds, and not just from the perception of individual sounds.

Another factor that might have influenced listeners' decision is tone. Mandarin has lexical tone information that may have helped listeners determine what syllable or word they heard, and at the same time, which consonant was at the beginning of that syllable. The extent to which tone had an effect on their decision in this case is not known at this point. It may be the case that tone, like other speech sounds is reduced in conversational speech, further decreasing listeners' ability to make accurate decisions about what they heard.

In conversational speech, allophonic distinctions can become ambiguous because of alteration or deletion of the allophones' conditioning environment, or of the consonant itself. In the case of the affricate *j*, the deletion or reduction of the conditioning high-front vowel may have contributed to the decrease in listeners' ability to correctly identify the consonant.

Perception was also less accurate for higher frequency words, because in conversational speech, high-frequency words often have deletion or reduction of the conditioning environment for the allophone. High-frequency words also often have altered consonants, further contributing to the misperception of the words (Pluymaekers et.al. 2005, Johnson 2004).

## V. CONCLUSION

Conversational speech and its effect on speech perception is very important to study in this case. Listeners, whether native or non-native all frequently encounter both carefully and spontaneously produced syllables that start with *j* or *zh*, so in order to understand how to better teach learners to hear these non-English consonants, one should study all possible manifestations of this sound.

The native speakers' data in this study showed that because correct perception of *j* and *zh* was not just based on hearing the consonant alone, it may be true that importance should be placed on both the differences in articulation of the two consonants, and also their surrounding environments to help learners of Mandarin improve their perception of *j* and *zh*. This study lays the groundwork for a future study about learners' ability to distinguish between *j* and *zh*. Non-native subjects will be given the same experiment as was administered in this study, but they will also hear carefully produced tokens. They will either be told to listen for the high front vowel to make their decision, or else not given any information about the two sounds, and their accuracy will be compared. This may further contribute to the study of language perception and acquisition, and show that variance between careful and conversational speech has a place in the study of language acquisition.

## References

- Dow, F. D. M. (1972). *An Outline of Mandarin Phonetics*. (Australian National University Press, Canberra), pp. 37-50.
- Duanmu, S. (2007). *The Phonology of Standard Chinese: Second Edition*. (Oxford University Press, New York), pp. 26-34.
- Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (eds.) *Spontaneous Speech: Data and Analysis*. Proceedings of the 1st Session of the 10th International Symposium. Tokyo, Japan: The National International Institute for Japanese Language, 29-54.
- Lee, W. S. and Zee, E. (2003). Illustrations of the IPA: Standard Chinese (Beijing). *Journal of the International Phonetic Association*, 33(1), 109-112.
- Lin, Y. H. (2007). *The Sounds of Chinese*. (Cambridge University Press, Cambridge), pp. 45-49.
- Lively, S., Logan J., and Pisoni D. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, 89(2), 874-886.
- Norman, J. (1988). *Chinese*. (Cambridge University Press, Cambridge), pp. 139-141.
- Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*, 118(4), 2561-2569.
- Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, 111, 1872-1891.
- Tao, H. and Cole, C. (1991). Wade-Giles or Hanyu Pinyin. *Cataloging & Classification Quarterly*, 12(2) 105 — 124.