PROCESSING SPEAKER VARIABILITY IN MANDARIN SPOKEN WORD RECOGNITION: A CLINICAL EXPLORATION

Yu Zhang[1], Suju Wang[2], Yingying Shang[2]

[1]Oklahoma State University, [2]Peking Union Medical College Hospital
yu.zhang10@okstate.edu, wangsuju@pumch.cn, yyingshang@aliyun.com

ABSTRACT

Speaker variability has been found to affect speech perception and spoken word recognition task performance in people with normal hearing. In this study, we tested Mandarin Chinese spoken word recognition in people with mild to moderate hearing loss. Five participants listened to pairs of disyllabic Chinese such as 国王–皇后 *king–queen*, which varied in semantic relationship and speaker identity, and made lexical decisions on the second item in a pair. With these limited number of clinical cases, we found evidence for semantic priming, although the effect of speaker variability on word recognition task was not observed. This result corroborated with prior auditory priming studies with normal hearing listeners that found no evidence of speaker variability influencing the deeper level of processing, such as lexical semantics.

Keywords: speaker variability, spoken word recognition, tone language, short-term priming.

1. INTRODUCTION

Speaker variability refers to variation in acoustic stimuli incurred by changes in talker identity. Since no two people have exactly the same vocal tract, speaker variability is an integral aspect in natural speech communication. It is known to provide indexical information about the speaker's background [1]. Spoken words are arguably the smallest meaningful unit in speech communication and speaker variability has been found to affect spoken word recognition performance in various laboratory tasks, e.g. [5, 10]. These tasks typically involve participants listening to speech stimuli produced by more than one speaker and making judgement on lexical form and/or meaning, reflecting two different levels of processing, i.e., phonetics and semantics.

In a landmark study, Andruski, Blumstein and Burton [4], using short-term semantic priming, found that sub-phonemic variation in the acoustic stimuli can affect word recognition performance or lexical access, leading to the question whether speaker variability can equally influence spoken word recognition. Limited evidence has been found with speaker variability influencing lexical semantics, although the priming paradigm has yielded some evidence of speaker variability negatively affecting processing of spoken word form [6, 7, 8].

A central theme from these studies revolves around the representation of word form and meaning in the mental lexicon and to what extent it is episodic such that acoustic variability may influence spoken word recognition performance. While listeners seem to be more prone to sub-phonemic, such as voice onset time (VOT), variability during spoken word recognition, the majority of the evidence was based on people with normal hearing from a non-tone language background.

In this study, we report an experiment involving people with mild to moderate sensorineural hearing loss from a tone language background to explore how speaker variability might affect this clinical population with Mandarin Chinese as their native tongue.

2. METHOD

Following prior studies investigating word processing in real time, the experiment presented here uses the short-term semantic priming paradigm, in which participants listen to pairs of stimuli and make lexical decisions on the second item ("target") of a pair. The first item of a pair is considered "prime". There is an inter-stimulus interval of 50 ms between prime and target. Some primes are semantically associated with the targets and the same targets are also paired with semantically unrelated primes to allow for the measurement of priming in reaction time difference. Such paradigm has been used in visual word recognition of Chinese characters, e.g. [12], to suggest that lexical meaning is activated as much and early as phonological form in Mandarin. Clinically, it has not been widely adopted to test word recognition.

2.1. Materials

The speech materials were Chinese disyllabic words and nonword items. All stimuli are disyllabic words because most modern Chinese words are disyllabic [3, 4]. The average frequency of occurrence for the semantically associated primes of targets was 82.25 counts per million; the average frequency of

occurrence for the corresponding unrelated primes was 83.46 counts per million, based on a speech corpus on Mandarin Chinese [2]. These two frequency counts are not significantly different from each other statistically, $p = .962$.

In constructing the Mandarin Chinese disyllabic stimuli, all phonemes and tone pairs were included, following the phonological representations specified in [11]. The prime and target items were controlled not only for semantic associations, but also with respect to phonological representations. There is no overlap between each type of prime word and its target in terms of initials, finals, and tone pairings, the three necessary components to specify a Chinese syllable. An example of semantically related prime, unrelated prime, and real-word target are: 国王 *king*, 事实 *fact*, 皇后 *queen*. A total of 52 pairs of prime and target were created, with non-word target fillers. The complete list is given in Appendix 1.

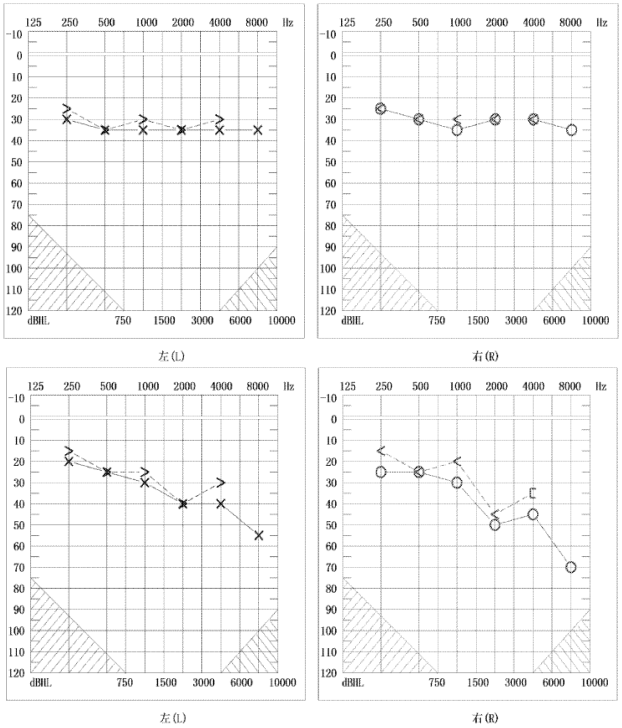
Two male native speakers of Mandarin Chinese were recruited to record the stimuli in a sound-attenuated booth, using an Audio-technica AT825 microphone connected to a personal computer (Dell OptiPlex 980) through a USBPre microphone interface. The recordings were sampled using the Brown Lab Interactive Speech System (BLISS) [9] at 22,050 Hz with 14-bit quantization. Stimulus items were then identified from the waveform display using Mev, the waveform editor of BLISS, and saved as individual sound files. BLISS was used to normalize the peak amplitude of all the individual audio files. The average F0 and duration were measured to evaluate their acoustic differences. The F0 between the two speakers was not significantly different, $t(51) = .42$, $p = .675$. The duration of target words was significantly different between the two speakers, $t(51) = 8.32$, $p < .001$.

2.2. Participants and procedure

Five females with sensorineural hearing loss participated in the experiment. Their age ranged from 24 to 62 years old. They all speak Mandarin Chinese as their native language with no history of other communicative or cognitive disorders. Fig. 1 shows the audiograms for participant XH-2, a mild hearing loss case and XH-3, a moderate hearing loss case. The complete hearing screening result for all participants is given in Appendix 2.

The participants were instructed in Mandarin Chinese to listen to pairs of items delivered through a headset and then make decisions on the second item. If they believed that the second item was a real Chinese word, then they should press the Chinese labelled button WORD on a computer keyboard. Otherwise, they should press the NONWORD label in Chinese characters on the same keyboard. The participants were tested individually in a quiet room, wearing a pair of headphones (Sennheiser HD 380 Pro) at their comfortable volume adjusted before each experiment session started.

Figure 1: The audiograms for participants XH-2 (upper) and XH-3 (lower).



All participants received all possible combinations of the semantic and speaker relations: (1) related, same speaker; (2) unrelated, same speaker; (3) related, different speaker, and (4) unrelated, different speaker. Therefore, semantic relation (related vs. unrelated) and speaker relation (same vs. different) are two within-subjects independent variables. Participants were given ten pairs of practice items to get familiarized with the task and were told to respond as soon as possible without sacrificing accuracy. The reaction time and lexical decision accuracy were measured as dependent variables. The entire experiment was run on a Dell Latitude 5480 laptop computer equipped with BLISS.

2.3. Results

The reaction time and accuracy data were acquired by BLISS and analysed in SPSS 17.0. Table 1 shows the average response accuracy and the average reaction time data across four conditions. Analysis of Variance (ANOVA) was performed on the arcsine transformed accuracy data. The overall accuracy of the lexical decision task was 87% based on real-word targets. The main effect of word relation was

significant [$F(1, 4) = 8.38$, $p = .044$, $\eta^2_p = .68$]. No other effects were statistically significant.

The overall reaction time was 1147 ms. The main effect of word relation was significant [$F(1, 4) = 8.08$, $p = .047$, $\eta^2_p = .67$]. No other effects were statistically significant.

Table 1: Mean reaction time (RT, in ms with SD) and percentage correct (PC, in % with SD) of the lexical decision task responses, as a function of speaker relation (SR) and word relation (WR).

SR	Same		Different	
WR	Related	Unrelated	Related	Unrelated
RT	1070 (289)	1240 (317)	1058 (205)	1221 (291)
PC	98 (3)	88 (12)	92 (9)	71 (34)

3. DISCUSSION

A semantic priming experiment was conducted on a group of five clinical hearing-loss participants, in order to explore the effect of talker variability on processing lexical semantics, from a tone language perspective and in people with hearing loss. On average, responses to real word targets preceded by related primes were 16% more accurate and 166 ms faster than those preceded by unrelated primes, as demonstrated by the significant main effect of word relation. However, because neither the main effect of speaker relation nor the interaction between the two independent variables were statistically significant, speaker variability was not found to influence the semantic processing of the Mandarin spoken words. This corroborates with the results from [8], where talker variability was not observed to influence the magnitude of semantic priming in normal hearing participants whose native language was not tonal. The results from this study also are consistent with [6], where talker variability, as well as VOT variability, was not found to influence processing of lexical semantics.

The implication of this result to the lexical activation process is twofold. In light of [6, 8], the evidence that speaker variability does not impact lexical semantics could suggest that the lexicon is devoid of surface variability as the level of processing deepens, for both tone and non-tone language users. On the other hand, results from this study could also be interpreted as that people with hearing loss are not as sensitive to voice differences as to word differences.

This study is limited in that only five clinical cases were involved, which restricted the power of the analysis and restricted a strong generalization. Nevertheless, the robust priming effect from these clinical participants suggests that the psycholinguistic paradigm of auditory priming may be extended to clinical populations with sensorineural hearing loss.

4. FUTURE DIRECTION

Future work may compare normal and hearing-impaired populations using the priming paradigm to explore how voice processing and word recognition interacts in a tone language, where listeners perceive fundamental frequency contours to mark meaningful contrasts.

Recognizing the need for more phonetically balanced and psychometrically reliable and valid instruments in the clinical setting, the current word list warrants more clinical validation.

5. ACKNOWLEDGEMENT

Deep gratitude is due to Chao-Yang Lee for the guidance in designing an auditory priming study on spoken word recognition, in addition to the discussions on the composition of the Mandarin stimuli used in the experiment. The authors also thank three anonymous reviewers for their constructive feedback and suggestions to make our work more impactful.

6. REFERENCES

[1] Abercrombie, D. 1967. *Elements of General Phonetics*. Chicago: Aldine.

[2] Cai, Q., Brysbaert, M. 2010. SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, 5, 1–8.

[3] Duanmu, S. 1998. Wordhood in Chinese. In: J. L. Packard (ed.), *New Approaches to Chinese Word Formation: Morphology, Phonology and the Lexicon in Modern and Ancient Chinese*. Berlin, Germany: Mouton de Gruyter, 135–196.

[4] Duanmu, S. 1999. Stress and the development of disyllabic vocabulary in Chinese. *Diachronica*, 16, 1–35.

[5] Goldinger, S. D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *J. of Exp. Psychol: Learn, Mem, and Cogn*, 22, 1166–1183.

[6] Lee, J., Lee, C.-Y. 2016. Effects of voice-onset time and talker variability on lexical access. *J. Acoust. Soc. Am.* 139, 2018.

[7] Lee, C.-Y., Zhang, Y. 2015. Processing speaker variability in repetition and semantic/associative priming. *J. of Psycholinguistic Research,* 44, 237–250.

[8] Lee, C.-Y., Zhang, Y. 2018. Processing lexical and speaker information in repetition and semantic/associative priming. *J. of Psycholinguistic Research*, 47, 65–78.

[9] Mertus, J. A. 2000. *The Brown Lab Interactive Speech System*. Providence, RI: Brown University.
[10] Schacter, D. L., Church, B. A. 1992. Auditory priming: Implicit and explicit memory for words and voices. *J. of Exp. Psychol: Learn, Mem, and Cogn*, 18, 915–930.
[11] Zhao, X., Li, P. 2009. An online database of phonological representations for Mandarin Chinese. *Behavior Research Methods*, 41, 575–583.
[12] Zhou, X., Marslen-Wilson, W. 2000. The relative time course of semantic and phonological activation in reading Chinese. *J. of Exp. Psychol: Learn, Mem, and Cogn*, 26, 1245–1265.

Appendix 1. Disyllabic Mandarin word stimuli used in the experiment, listed in simplified Chinese characters and corresponding English translation.

Related Prime	Unrelated Prime	Target
警察 police	长度 length	小偷 thief
塑料 plastic	年纪 age	回收 recycle
声音 sound	礼物 gift	画面 image
乌鸦 crow	花费 cost	喜鹊 magpie
池塘 pond	公路 highway	荷花 lotus
机智 witty	铃铛 bell	勇敢 brave
画报 pictorial	平稳 steady	杂志 magazine
地点 place	花生 peanut	人物 person
安静 quiet	律师 lawyer	活泼 lively
信息 info	西瓜 watermelon	来源 source
公司 company	频道 channel	企业 factory
钓鱼 fishing	生气 angry	划船 boating
饼干 cookie	消息 message	糕点 pastry
马路 road	特点 feature	逛街 stroll
空调 AC	红色 red	冷气 cool air
发现 discover	人口 population	察觉 detect
问题 problem	足球 soccer	答案 answer
蝴蝶 butterfly	头盔 helmet	蜻蜓 dragonfly
冠军 champion	数学 math	奖励 reward
睡觉 sleep	交代 tell	休息 rest
锻炼 exercise	天使 angel	游泳 swim
皮球 ball	厨房 kitchen	玩耍 play
食堂 cafeteria	手机 cell phone	宿舍 dorm
飞机 aircraft	父亲 father	大炮 cannon
电池 battery	世界 world	能量 energy
老鹰 eagle	组合 combo	小鸡 chicks
公主 princess	衬衫 shirt	王子 prince
再见 goodbye	科学 science	道别 farewell
蔬菜 vegetable	音乐 music	水果 fruit
颜色 color	管理 manage	图画 picture
现代 modern	帮忙 help	古老 ancient
同学 classmate	公园 park	老师 teacher
国王 king	事实 fact	皇后 queen
战争 war	草原 grassland	和平 peace
敌人 enemy	房间 room	朋友 friend
面包 bread	电梯 elevator	牛奶 milk
太阳 sun	汽车 car	月亮 moon
树木 tree	记者 reporter	森林 forest
困难 difficult	长城 the Great Wall	容易 easy
失败 failure	邮票 stamp	成功 success
文字 character	天堂 heaven	语言 language
日本 Japan	意思 meaning	中国 China
握手 handshake	钢琴 piano	你好 hello
球队 team	房产 estate	教练 coach
黑夜 night	护照 passport	白天 day
北京 Beijing	决定 decision	上海 Shanghai
富有 rich	演员 actor	贫穷 poor
报纸 newspaper	季节 season	新闻 news
鼓励 encourage	城市 city	加油 root for
愉快 happy	法官 judge	悲伤 sad
乞丐 beggar	啤酒 beer	富翁 the rich
锋利 sharp	跳舞 dance	迟钝 obtuse

Appendix 2. Hearing screening results for the participants in this study.

Participant Code	Age	Ear Tested	Pure-tone threshold (dB HL)					
			250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
XH-1	62	L	25	20	20	25	20	40
		R	60	40	35	25	35	40
XH-2	52	L	30	35	35	35	35	35
		R	25	30	35	30	30	35
XH-3	53	L	20	25	30	40	40	55
		R	25	25	30	50	45	70
XH-4	24	L	15	30	45	40	15	20
		R	15	25	40	45	15	20
XH-5	62	L	45	45	55	45	25	50
		R	40	45	50	45	30	50

L = left; R = right.