INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Cross-Linguistic Perception of Utterances with Willingness and Reluctance in Mandarin by Korean L2 Learners

Wenqian Li[1], *Jung-Yueh Tu*[2]

[1]Shanghai Jiao Tong University, China
[2]National Chengchi University, Taiwan
wenhsi@sjtu.edu.cn, jytu@nccu.edu.tw

Abstract

This study investigated the cross-linguistic perception of attitudinal intonation with willingness and reluctance in Mandarin by Korean L2 learners. In the current study, 20 Korean L2 learners of Mandarin (KL2) and 20 native Mandarin listeners (CL1) were instructed to rate perceived degree of willingness (1-5 Likert scale) from the utterances (with willingness, reluctance, and neutrality) produced by 2 native Mandarin speakers (one male and one female). The rating results showed that 1) the rating scores of willing attitude were significantly higher than those of reluctant attitude by KL2; 2) utterances of willingness and neutrality tend to be perceived less willing by KL2 than by CL1; 3) KL2 had a narrower rating range on the perception of attitudinal intonation than CL1. Specifically, Korean females had a wider rating range than Korean males. The findings indicated that 1) utterances of willingness, neutrality, and reluctance in Mandarin were accurately perceived by KL2; 2) willingness carried by attitudinal intonation was weakened through L2 pragmatic comprehension by KL2; 3) Korean females were more sensitive than Korean males on the perception of attitudinal intonation. The overall results suggest significant effects of language experience and gender difference on the perception of Chinese utterances with willingness and reluctance.

Index Terms: speech perception, attitudinal speech, Korean L2 learners, language experience, gender differences

1. Introduction

Imagine that you are having a conversation with a friend, and you are planning to go to the restaurant you always want to go to. When she/he asks you what time you are free, you answer her/him with a feeling of eagerness. This example exemplifies the feeling of willingness in social situations. Feelings of (un)willingness embedded in prosody are paralinguistic and non-linguistic information which is "added by the speaker to modify and supplement the linguistic information" and conveys "the age, gender, idiosyncrasy, and physical and emotional states of the speaker, and so on, and cannot generally be controlled by the speaker" [1, p.1]. It is essential to perceive emotional and attitudinal utterances, as information beyond the literal, semantic and symbolic content in speech can lead to successful communication or breakdown between interlocutors. Accordingly, studies of perception on emotional and attitudinal speech are of high importance.

1.1. Classification of emotion and attitude

Couper-Kuhlen (1986) characterized emotion as the state of the speaker while attitude as a kind of behavior [2]. Wichmann (2000) subdivided attitude into 'propositional attitude' (conveying the speaker's opinion, belief, knowledge on a person or an issue) and 'behavioral attitude' (reflecting the intended or perceived behavior of the speaker) [3]. Gu and Fujisaki (2013) made a more clear-cut distinction between emotion and attitude [4]. Emotion is, as they claimed, relevant to the inner and physical state of the speaker which is usually expressed involuntarily. On the contrary, attitude relates to external state and can be controlled consciously by the speaker.

According to Gu and Fujisaki's (2013) distinction [4], this paper will classify the two chosen parameters, 'willingness' and 'reluctance', as 'attitude'. Specifically, this study focuses on how language experience and gender influence the perception of Mandarin utterances with a dichotomy between attitudes: willingness and reluctance.

1.2. Previous research on emotional and attitudinal speech

Due to the subtle differences between 'emotion' and 'attitude', previous studies on emotional and attitudinal speech always share the same experimental paradigm, which can be roughly divided into two types. The first type intended to describe the acoustic characteristics of a particular emotion/attitude or diverse emotions/attitudes, such as anger, happiness, and surprise [6, 7, 8, 9, 10]. The second type of studies examined the perception of emotional/attitudinal speech and usually involved several variables, such as age and gender [11, 12, 13, 14, 15, 16, 17, 18]. In some studies [6, 8, 12, 15, 18], neutral utterances were produced and analyzed as a control group to examine the differences between emotional/attitudinal utterances.

Among the perceptual studies, the variables investigated included age [11, 12], gender [12, 13, 18], audiovisual cues [14, 15], and language [5, 14, 15, 16]. For the effect of language experience on the perception of emotional/attitudinal speech, Mac et al. (2010) did perception experiments with French and Vietnamese participants based on an audiovisual corpus of 16 attitudes of Vietnamese [14]. They suggested that attitudes of 'declaration', 'exclamation of positive surprise', 'doubt-incredulity', 'authority', 'irritation', and 'seduction' were well perceived by both native listeners and non-native listeners. Lu et al. (2012) further investigated the prosodic perception of social affects in Mandarin Chinese with French and Vietnamese [5]. And they suggested that the majority of their chosen attitudes were well recognized except 'contempt', 'irony' and 'confidence'. Barkhuysen et al. (2010) asked Czech participants to rate their perceived emotional state of

http://dx.doi.org/10.21437/Interspeech.2020-1640

Dutch speakers [15]. They found that Czech participants could recognize emotions without relying on lexical cues. Brown et al. (2014) carried out two perception experiments with Korean and English participants without knowledge of Korean [16]. They suggested that some acoustic features correlated with politeness may be understood in similar ways across cultures. The above studies investigated perception of listeners across languages with no L2 experience and found that participants generally perceived emotions/attitudes well to a great extent. Thus, how L2 learners, with the target language experience, perceive attitudes across languages aroused our interest.

As for gender differences, on the perception of emotional/attitudinal speech, Kaya et al. (2017) suggested that "recognition of male speakers was better than female speakers for all emotional states (p. 268)", which included discomfort, neutral, and comfort utterances of children in their experiment [12]. For male productions, male listeners were better at perceiving utterances than female listeners. For female productions, female listeners were better at perceiving utterances than male listeners. Waaramaa et al. (2018) indicated that gender differences were statistically significant for emotion identification among children, with girls demonstrating a higher level of accuracy [13]. In addition, Li et al. (2020) investigated gender differences in Mandarin attitudinal speech (willingness and reluctance) and found that female listeners were generally better at perceiving willingness and reluctance than males and that female speakers' utterances tended to be perceived as more willing [18]. The current study plans to explore how perception on attitudinal speech (willingness and reluctance) in Mandarin Chinese is influenced by participant's language experience (L1 vs. L2) and gender.

2. Method

2.1. Participants

There were 20 Korean L2 learners (10 females and 10 males; age range: 17-34 years old; mean age: 25.3 years old) and 20 native Mandarin listeners (10 females and 10 males; age range: 22-28 years old; mean age: 23.7 years old) participating in this study. The Korean L2 learners who participated in the current study have learned Chinese for 5-10 years (M = 6.4 yrs, SD =1.87) and all of them have stayed in China for 3 years and above. None had difficulty in hearing and speaking. None had long experience (3 years and above) of working in service industries. All participants received a small payment for their participation.

2.2. Stimuli

The stimuli were 20 sentences (each sentence contains 6 to 10 words) including statements and questions, as given in Table 1.

We designed two scenarios for each target sentence in order to create a sense of naturalness for the speakers. In the first scenario, the speaker was induced to use willing attitude to finish a conversation with the experimenter. In the second scenario, the speaker was induced to use reluctant attitude to finish a conversation. Situational prompts including the background information of the scenarios were given and two to four turns of conversation were elicited, as illustrated in Table 2.

Table 1: *Stimuli used in Experiment 1.*

Stimuli	Meaning
1. 我在楼下等你一会儿吧。	'I'll wait for you downstairs for a while.'
2. 我们今天就玩一局游戏。	'We'll play a round of game today.'
3. 再来一瓶可乐。	'Give me one more coke.'
4. 周六我有时间。	'I'm free on Saturday.'
5. 我明天来体育馆找你。	'I'll see you at the gym tomorrow.'
6. 我在这里住了一个月了。	'I've stayed here for a month.'
7. 我要寄一个包裹。	'I want to send a parcel.'
8. 我自己去找他吧。	'I'll meet with him on my own.'
9. 我帮你查一下。	'Let me check it for you.'
10. 我今晚要熬夜看书了。	'I'm going to stay up late tonight to read.'
11. 我们在哪个教室比赛？	'Which classroom are we having a quiz?'
12. 明天中午可以吗？	'How about tomorrow at noon?'
13. 我要留下来陪你吗？	'Should I stay with you?'
14. 我看能不能早点去。	'I'll see if I can come earlier.'
15. 我要分给室友吗？	'Do I have to share it with my roommate?'
16. 我也要跟你去美国吗？	'Do I have to go to the United States with you?'
17. 我们去吃日料吗？	'Are we going to have a Japanese meal?'
18. 要不打电话说？	'How about talking on the phone?'
19. 还要在这里住一夜吗？	'Do you want to stay here for another night?'
20. 要不你五点来吧？	'Perhaps you can come at five?'

Two native Mandarin speakers (one male and one female) from northern China did the recording. The speakers recorded in a quiet and comfortable room with the experimenter. First, the speakers were asked to rehearse the materials until they were acclimated to the scenarios and sentences. They were then requested to produce and record the conversations. The speaker read all target sentences in a neutral tone without any emotions before recording attitudinal utterances. After recording neutral utterances, they were asked to finish 60 rounds of conversations with the experimenter. The recording process took around 1 hour for each speaker. In total, 120 stimuli (2 speakers * 3 attitudes * 20 sentences) were

Table 2: *Example of a target sentence with two scenarios.*

Target sentence:	
明天中午可以吗	'How about tomorrow at noon?'

Two scenarios:

Willing attitude		**Reluctant attitude**	
Hint	**Dialogue**	**Hint**	**Dialogue**
小李和小王是好友，他们计划约会。	'Li and Wang are good friends, and they are going to have a date.' 小李：我们什么时候一起去吃饭？ 小王（**情愿地**）：*明天中午可以吗？* 'Li: When do we go out for a meal?' 'Wang (**willingly**): *How about tomorrow at noon?'*	小李和小王是同事，小李无法完成任务希望小王可以帮忙。小王不太情愿，但还是同意了。	'Li and Wang are colleagues. Li can't finish her task and wants Wang's help. Wang is not willing but agrees to help her eventually.' 小李：你现在有空吗？你可以帮我整理这些文件吗？ 小王：对不起，但是我今天有点忙。 小李：明天呢？我实在做不完了。 小王（**不情愿地**）：*明天中午可以吗？* 'Li: Are you free now? Can you help me with these documents?' 'Wang: Sorry, but I am quite busy today.' 'Li: What about tomorrow? I can't deal with them.' 'Wang: (**reluctantly**) *How about tomorrow at noon?'*

generated. Individual sound files were imported and segmented using the software Praat [19].

2.3. Procedure

Participants did the perceptual experiment via an online questionnaire with a recording file attached. The stimuli were presented to the participants in a randomized order. The participants were asked to listen to each sentence carefully and then rate it from 1 to 5, according to their perceived degree of willingness (1 represents particularly reluctant, 2 represents slightly reluctant, 3 represents neutral, 4 represents slightly willing, and 5 represents particularly willing). The experiment took about 15 minutes for each listener. The data were processed and visualized using the software SPSS 25 [20].

3. Results

3.1. Effect of language experience

A one-way analysis of variance (ANOVA) was conducted for the average rating scores of all stimuli on perception of native Chinese listeners (CL1). A Shapiro-Wilk test showed the scores of CL1 were normally distributed ($p > 0.05$). A Levene's test showed the equality of variances for a variable ($p = .064$). The main effect of attitude was significant [$F = 369.103$, $p < .001$]. A Tukey's HSD test showed that the average rating scores of willingness by CL1 (M = 3.85, SD = .506) were significantly higher than those of neutrality (M = 2.79, SD = .311) [95%CI: 0.89 ~ 1.22 , $p < 0.001$]. And the average rating scores of neutrality by CL1 were significantly higher than those of reluctance (M = 1.99, SD = .458) [95%CI: 0.64 ~ 0.96 , $p < 0.001$].

A Shapiro-Wilk test showed the scores of Korean L2 listeners (KL2) were normally distributed ($p > 0.05$). A Levene's test showed the equality of variances for a variable ($p = .081$). The main effect of attitude was significant [$F = 196.542$, $p < .001$]. A Tukey's HSD test showed that the average rating scores of willingness by KL2 (M = 3.44, SD = .520) were significantly higher than those of neutrality (M = 2.68, SD = .402) [95%CI: 0.59 ~ 0.94 , $p < 0.001$]. And the average rating scores of neutrality by KL2 were significantly higher than those of reluctance (M = 2.00, SD = .465) [95%CI: 0.52 ~ 0.87 , p < .001] (see Table 3).

Table 3: *The average rating scores, number, and standard deviation of Korean listeners for perceiving all utterances.*

Attitude	Group	Average rating scores	Number	Standard deviation
Willing	CL1	3.85	80	.506
	KL2	3.44	80	.520
Neutral	CL1	2.79	80	.311
	KL2	2.68	80	.402
Reluctant	CL1	1.99	80	.458
	KL2	2.00	80	.465

An independent samples t-test was made to compare the average rating scores of CL1 and KL2 for listening to different attitudes. The average rating scores of CL1 for listening to willing attitude (M = 3.85, SD = .506) were significantly higher than those of KL2 (M = 3.44, SD = .520) [$t = 4.992$, $p < .001$]. The average rating scores of CL1 for listening to neutral attitude (M = 2.79, SD = .311) were significantly higher than those of KL2 (M = 2.68, SD = .402) [$t = 2.000$, $p < .05$] (see Figure 1).

Based on the t-tests, Figures 1-5 show a descriptive result with significant differences between average rating scores of participants being marked by asterisks (<*> = p < 0.05; <**> = p < 0.005; <***> = p < 0.001).

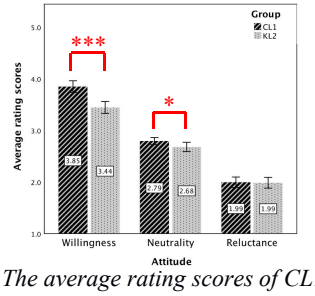

Figure 1: *The average rating scores of CL1 and KL2 for listening to utterances with 'willingness', 'neutrality', and 'reluctance'.*

3.2. Effect of gender

An independent samples t-test was made to compare the average rating scores of CL1 and KL2 listeners of different genders for listening to different attitudes. The average rating scores of female listeners of KL2 for listening to neutral attitude (M = 2.57, SD = .432) were significantly lower than those of male listeners of KL2 (M = 2.78, SD = .344) [$t = -2.434$, $p < .05$]. The average rating scores of Korean females for listening to reluctant attitude (M = 1.85, SD = .518) were significantly lower than those of Korean males (M = 2.12, SD = .365) [$t = -2.696$, $p < .05$] (see Figure 2).

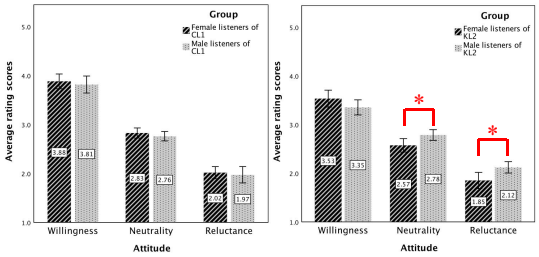

Figure 2: *The average rating scores of CL1 and KL2 listeners of different genders for utterances with 'willingness', 'neutrality', and 'reluctance'.*

For the perception by CL1, the average rating scores of female utterances perceived as reluctant attitude (M = 2.12, SD = .499) were significantly higher than those of male utterances (M = 1.87, SD = .377) [$t = 2.552$, $p < .05$] (see Figure 3).

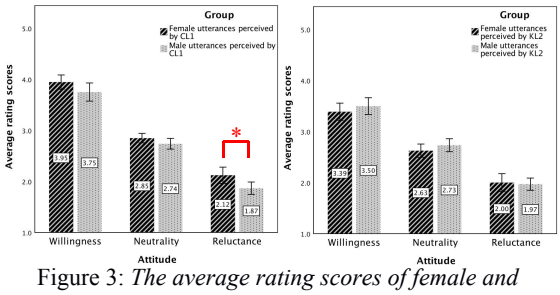

Figure 3: *The average rating scores of female and male utterances perceived as 'willingness', 'neutrality', and 'reluctance' by CL1 and KL2*

An independent samples t-test was made to compare the average rating scores by different genders for CL1 and KL2 when listening to different attitudes produced by females and males. The average rating scores by Chinese males when listening to reluctant attitude (M=2.15, SD=.566) produced by females were significantly higher than the average rating scores of reluctant utterances produced by males (M=1.80, SD=.415) [$t = 2.262$, $p < .05$] (see Figure 4).

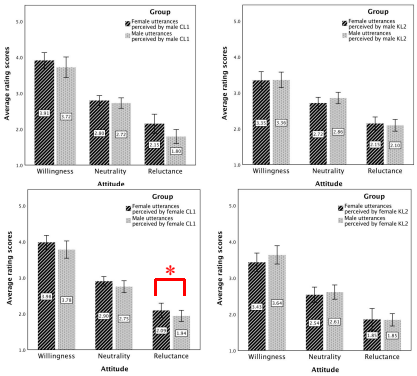

Figure 4: *The average rating scores by CL1 and KL2 when listening to female and male utterances with 'willingness', 'neutrality', and 'reluctance'.*

An independent samples t-test was made to compare the average rating scores by female and male listeners in CL1 and KL2 when listening to different attitudes. The average rating scores by female CL1 listeners when listening to willing attitude (M = 3.88, SD = .470) were significantly higher than those of female KL2 listeners (M = 3.53, SD = .545) [t = 3.054, $p < .005$]. The average rating scores by female CL1 listeners when listening to neutral attitude (M = 2.83, SD = .323) were significantly higher than those of female KL2 listeners (M = 2.57, SD = .432) [t = 2.960, $p < .005$]. The average rating scores by male CL1 listeners when listening to willing attitude (M = 3.81, SD = .544) were significantly higher than those by male KL2 listeners (M = 3.35, SD = .484) [$t = 4.108$, $p < .001$] (see Figure 5).

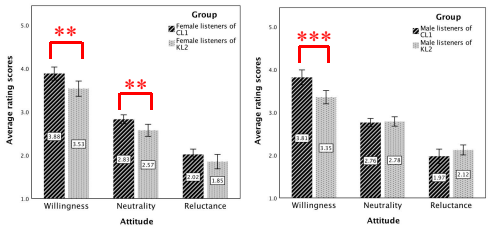

Figure 5: *The average rating scores by female and male listeners for CL1 and KL2 when listening to utterances with 'willingness', 'neutrality', and 'reluctance'.*

4. Discussion

4.1. Effect of language experience

For the influence of language experience, the ratings of CL1 and KL2 showed that utterances with willingness were the highest, followed by neutrality, and then reluctance, which indicated that attitudinal speech (willingness, neutrality, and reluctance) were accurately perceived by both native speakers and non-native speakers. The results also revealed that CL1 was generally better at perceiving willing and neutral attitudes in Mandarin than KL2 and that CL1 had a wider rating range on the perception of attitudinal intonation than KL2. The average rating scores on different attitudes by CL1 were generally higher than those by KL2, indicating that KL2 tend to perceive the attitudinal utterances as unwilling. It is speculated that willingness carried by attitudinal intonation was weakened through L2 pragmatic comprehension, which needs further investigation.

Utterances with neutrality were mostly perceived as reluctance in the present study. Jiang and Pell (2017) found that neutral utterances typically lead to comparatively higher ratings of speaker's confidence [6]. This discrepancy is probably because their study and the current study differ in two aspects. Their study investigated expressed confidence and doubt in spoken language while our study targeted attitudes with willingness and reluctance. Also, the participants in their study were native Canadian English speakers whereas our participants were native Mandarin speakers and Korean L2 learners. The specific attitudes and cultural differences may lead to distinct ratings. Further research is required to provide more evidence of the perception on different attitudes across languages and cultures.

4.2. Effect of gender

With regard to gender differences, female KL2 listeners were found to be better than male KL2 listeners in perceiving reluctant attitude while CL1 showed no significant gender differences. The ratings of female KL2 listeners on willing attitude were higher than those of male KL2 listeners and the ratings of female KL2 listeners on reluctant attitude were lower than KL2 males. KL2 females had a wider rating range of attitudinal speech than KL2 males, which implied that KL2 females were more sensitive when listening to attitudinal speech. According to some sociolinguistic studies [21, 22, 23], we speculate that females are highly sensitive to the social context and the better perception by females is due to the social conditions among which females appear more accommodating and perform more emotional labor.

As for CL1, the ratings of reluctant utterances produced by females were higher than those of utterances produced by males. Further calculation showed that the significant effect came from the ratings by CL1 males. Reluctant utterances produced by females were rated higher than reluctant utterances produced by males in the process of CL1 males' rating. Our results suggest that males tend to perceive females' reluctance as a more willing attitude. Baumeister et al.(2002) found that narcissism caused males to aggress against the females who has refused them [24]. Narcissistic tendencies could lessen their perception of refusal, such as reluctant utterances.

For female listeners, CL1 females were better than KL2 females on perception of willing and neutral attitudes. For male listeners, CL1 males were better than KL2 males on perception of willing attitude.

5. Conclusion

In conclusion, the findings indicated that willingness, neutrality, and reluctance in Mandarin speech were accurately perceived but weakened through L2 pragmatic comprehension by Korean L2 learners. As L2 learners, Korean females were more sensitive than Korean males on the perception of attitudinal intonation.

6. References

[1] H. Fujisaki, "Information, prosody, and modeling - with emphasis on tonal features of speech - ," *Proc. Speech Prosody*, Nara, pp. 1-10, 2004.

[2] E. Couper-Kuhlen, *An Introduction to English Prosody*. London: Edward arnold, 1986.

[3] A. Wichmann, "The attitudinal effects of prosody, and how they relate to emotion," in *Speech and Emotion, September 5-7, Newcastle, Northern Ireland, UK, Proceedings, 2*000, pp. 143-148

[4] W.-T. Gu, and H. Fujisaki, "Data acquisition and prosodic analysis for Mandarin attitudinal speech," in *Eastward Flows the Great River: Festschrift in Honor of Professor William*, 2013.

[5] Y. Lu, V. Auberge, and A. Rilliard, "Tonal influences on the prosodic cross-linguistic perception of Mandarin social affects by French and Vietnamese listeners," in *TAL 2012 - 3th International Symposium on Tonal Aspects of Languages, May, Nanjing, China, Proceedings*, 2012.

[6] X. Jiang, and Marc D. Pell, "The sound of confidence and doubt," *Speech Communication*. vol. 88, pp. 106-126, 2017.

[7] D. Jiang, and L. Cai, "Chinese emotional speech classification based on prosodic features," In *ACII 2003 - The first Chinese Conference on Affective Computing and Intelligent Interaction*, *December, Beijing, China, Proceedings,* 2003.

[8] A. Li, "Acoustic analysis of friendly attitude speech," *Studies of the Chinese Language*, vol. 05, pp. 418-431, 2005.

[9] S. Zheng, *Prosodic Analysis on Polite and Friendly Attitudes of Mandarin Chinese as a Second Language*. Nanjing: Nanjing Normal University, 2014.

[10] L. Zhao, X. Qian, C. Zou, and Z. Wu, "A study on emotional feature analysis and recognition in speech signal," *Journal of China Institute of Communications*, vol. 21, pp. 18-24, 2000.

[11] B. M. Ben-David, S. Gal-Rosenblum, P. H. H. M. van Lieshout, and V. Shakuf, "Age-related differences in the perception of emotion in spoken language - the relative roles of prosody and semantics," *Journal of Speech, Language, and Hearing Research*, vol. 62, pp. 1188-1202, 2019.

[12] H. Kaya, A. A. Salah, A. Karpov, O. Frolova, A. Grigorev, and E. Lyakso, "Emotion, age, and gender classification in children's speech by humans and machines," *Computer Speech & Language*, vol. 46, pp. 268-283, 2017.

[13] T. Waaramaa, T. Kukkonen, S. Mykkänen, and A. Geneid, "Vocal emotion identification by children using cochlear Implants, relations to voice quality, and musical interests," *Journal of Speech, Language and Hearing Research (Online)*, vol. 61, pp. 973-985, 2018.

[14] D. Mac, V. Auberge, A. Rilliard, and E. Castelli, "Cross-cultural perception of Vietnamese audio-visual prosodic attitudes," in *SP 2010 - 5th International Conference of Speech Prosody, May 10-14, Chicago, USA, Proceedings*, 2010, paper 356

[15] P. Barkhuysen, E. Krahmer, and M. Swerts, "Crossmodal and incremental perception of audiovisual cues to emotional speech," *Language and Speech*, vol. 53, pp. 3-30, 2010.

[16] L. Brown, B. Winter, K. Idemaru, and S. Grawunder, "Phonetics and politeness: Perceiving Korean honorific and non-honorific speech through phonetic cues," *Journal of Pragmatics*, vol. 66, pp. 45-60, 2014.

[17] B. M. Ben-David, N. Multani, V. Shakuf, F. Rudzicz, and P. H. H. M. van Lieshout, "Prosody and semantics are separate but not separable channels in the perception of emotional speech- test for rating of emotions in speech," *Journal of Speech, Language and Hearing Research (Online)*, vol. 59, pp. 72-89, 2016.

[18] W. Li, J. W. Wong, and J. Tu, "Gender differences on the perception and production of utterances with willingness and reluctance in Chinese," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 4-8 Barcelona, Spain, Proceedings,* 2020, pp. 7154-7158.

[19] P. Boersma, and D. Weenink, Praat: Doing phonetics by computer. Version 6.0.14, http://www.praat.org/. 2018.

[20] IBM Corp. Released 2017. IBM SPSS Statistics for Mac OS, Version 25.0. Armonk, NY: IBM Corp.

[21] J. Bernard, *The Sex Game*. New York: Atheneum, 1973.

[22] P. Trudgill and S. Trudgill, *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press, 1974.

[23] R. Lakoff, *Language and Woman's Place*. New York: Harper and Row, 1975.

[24] R. F. Baumeister, K. R. Catanese, and H. M. Wallace, "Conquest by force: A narcissistic reactance theory of rape and sexual coercion," *Review of general psychology.* vol. 6, pp. 92-135, 2002.