Prosodic realization of politeness in the presence of non-prosodic cues in Mandarin Chinese

Jiayong He$^{1,2}$, Jing Tang$^3$, Stella Gryllia$^4$, Aoju Chen$^5$

$^1$Tongji University, Shanghai, China
$^2$University of Electronic Science and Technology of China, Chengdu, China
$^3$Sichuan University of Arts and Science, Dazhou, China
$^4$Webster University Leiden Campus, Leiden, The Netherlands
$^5$Utrecht University, Utrecht, The Netherlands

82239926@qq.com, tangjing5352@163.com, stella.gryllia94@webster.edu, aoju.chen@uu.nl

Abstract

Recent years have seen a growing interest in the prosodic realization of politeness. But little is still known on the interplay between prosody and non-prosodic cues to politeness, such as the use of a hesitation marker and choice of utterance types. This study examines whether prosody is used to a lesser extent in the presence of the hesitation marker (en ‘uhm’) and whether this pattern is strengthened in yes/no questions than in imperatives in Mandarin Chinese polite speech, as suggested by Haan’s Functional Hypothesis. Neutral and polite speech was elicited from 22 native speakers of Mandarin Chinese via a dialogue game in imperatives, yes/no questions and statements. Speech rate, mean pitch, pitch span and breathiness were measured at the sentence level. The results show that across utterance types, the speakers used speech rate, mean pitch and breathiness to express politeness, as in speech without a hesitation marker. But the speakers also increased pitch span to express politeness, different from speech without a hesitation marker. These findings provide no evidence for the Functional Hypothesis and suggest that prosody is not necessarily used for different communicative functions to a lesser extent in the presence of non-prosodic cues.

Index Terms: politeness, Mandarin Chinese, hesitation marker, speech rate, mean pitch, pitch span, breathiness

1. Introduction

As a frequently used strategy in interpersonal communication, politeness has received substantial attention over the past decades [1]. Previous research has mainly addressed the morpho-syntactic and lexical aspects of politeness. The prosodic aspect of politeness has been less studied. Investigations into the prosodic aspects of politeness have shown that speakers use a number of cues, for instance, breathiness, change in pitch and speech rate, to express politeness across languages, although the exact use of prosodic parameters varies cross-linguistically. For instance, speakers of Mandarin Chinese use breathier voice at the beginning of a conversation to sound more friendly (an attitude close to politeness)[2], similar to speakers of Japanese [3,4] when expressing politeness. Moreover, speakers of Mandarin also raise the pitch to sound more friendly or polite. Specifically, speakers of Mandarin increase the mean pitch at the utterance level or they raise the mean pitch at an utterance initial or medial position more than at an utterance final position to sound more friendly or polite [2,5], similar to raising pitch in the utterance level by speakers of English [6] and in the final position by speakers of Japanese [7], but opposite to lowering overall utterance-level pitch by speakers of Korean [8] and Catalan [9]. Moreover, speakers of Mandarin also increase speech rate to express politeness [5,10,11,12,13], different from speakers of Korean [8], varieties of German [14], Catalan [9], Japanese [7] and similar to speakers of English [6]. Finally, speakers of Mandarin do not appear to vary pitch span to convey politeness [5,10], unlike speakers of Korean [8], but similar to speakers of Greek [15].

Like prosodic research on politeness in other languages, previous studies on prosodic realization of politeness in Mandarin are primarily concerned with the use of prosody in the absence of non-prosodic cues, such as hesitation markers and choice of utterance types. As some studies have found that hesitation markers play a role in expressing politeness [8,14,16,17], the question arises as to whether prosody is used differently in the presence of a hesitation marker. Furthermore, utterance types seem to affect perceived politeness, with requests made via interrogatives sounding more polite than those made via imperatives [11,18]. However, the function of prosody in polite expression has not been studied systematically in different utterance types. This brought up the question of how prosody interacts with morpho-syntactic cues such as utterance types. As postulated in the Functional Hypothesis [19] (stemming from research on question intonation in Dutch), prosody tend to be used to a lesser extent to convey a meaning in the presence of (stronger) morpho-syntactic cues (e.g. final rise in the absence of morpho-syntactic cues in declarative questions vs. absence of final rise in the presence of wh-word and inversion in wh-questions). The question pertinent to politeness research is whether there is a functional trade-off between prosody and utterance types in Mandarin Chinese polite speech.

The current study has set out to address the above-mentioned questions by examining the use of speech rate, mean pitch, pitch span and breathiness when expressing politeness in the presence of the hesitation marker en ‘uhm’ in different utterance types (i.e. statement, yes/no question, imperative) in Mandarin Chinese. Extending the Functional Hypothesis to non-lexical cues such as hesitation markers, we hypothesize that prosody will be used to a lesser extent in the presence of a hesitation marker and this pattern will be more observable in yes/no questions and least observable in imperatives.
2. The production experiment

2.1. Participants

Twenty-two native speakers of Mandarin Chinese (12M, 10F, \(\bar{x} = 20\) years, SD = 1.81) participated in this study. They were university undergraduates at the time of testing. None of them had a musical background or reported deficits in vision or hearing. They were paid a small fee for their participation.

2.2. Stimuli

We constructed a total of 42 stimuli: 7 scenarios \(\times\) 3 utterance types (statement, yes/no question, imperative) \(\times\) 2 attitudes (neutral, polite). The statements used the hedged performatives \(\text{wò xǐng} \ ‘\text{I was hoping}', the yes/no questions the indirect directive beginning \(\text{nǐn néng} \ ‘\text{could you}', and the imperatives the politeness marker \(\text{qǐng} \ ‘\text{please}'. Examples (1)-(3) illustrate stimuli from one scenario in the neutral condition, while (4)-(6) illustrate stimuli from the same scenario in the polite condition. The stimuli in the two attitude conditions differed in that the stimuli in the polite condition contained the hesitation marker \(\text{èn} \ ‘\text{uhm}’\) and were preceded by a background sentence (B). The position of the hesitation marker does not change. Though speakers may articulate \(\text{èn} \ ‘\text{uhm}'\) in difference places in an utterance, it is common to be at the beginning.

(1) 我想请您帮我买一下票。 [statement]

I was hoping you could help me buy the ticket.

(2) 您能帮我买一下票吗? [yes/no question]

Could you help me buy the ticket?

(3) 请帮我买一下票。 [imperative]

Please help me buy the ticket.

B: At a metro station, you ask a stranger to help you buy a ticket.

(4) 嗯…我想请您帮我买一下票。 [statement]

Um… I was hoping you could help me buy the ticket.

(5) 嗯…您能帮我买一下票吗？ [yes/no question]

Um… Could you help me buy the ticket?

(6) 嗯…请帮我买一下票。 [imperative]

Um… Please help me buy the ticket.

The stimuli were put in two sets (neutral and polite) and within each set they were distributed in three blocks in accordance with the utterance type. The blocks were arranged according to a balanced Latin Square, while the stimuli within the blocks remained in a fixed order.

2.3. Task and procedure

The recordings were made in two sound-attenuated booths at two universities in Chengdu, China. The data were recorded with a Shure SM10A head-mounted microphone using xRecorder [20] and were directly saved on a PC via a M-Audio 410 FireWire external sound card (sampling rate 16kHz, 16-bit, mono).

The recording session had two parts. In the first part (neutral condition), participants were instructed to utter the sentence that appeared on the screen as naturally as possible without expressing any emotion, while in the second part (polite condition), participants were instructed to imagine that they were talking to a stranger, to read the context that appeared on the screen and to utter the sentence with appropriate politeness.

The task was self-paced. Participants had the option to record a target sentence again if they felt they had not uttered it in an appropriate way.

2.4. Acoustic measurements

A total of 924 utterances (42 stimuli \(\times\) 22 participants) were first auto-segmented using xSegmenter [20]. Then the onsets and offsets of the segments were manually adjusted in Praat [21]. The following four measurements were obtained at the utterance level: speech rate (number of syllables per utterance divided by the utterance duration in seconds, not including the filled pause between the hesitation marker and the utterance); pitch span (difference between the highest and lowest pitch); mean pitch; and breathiness. To measure P0/pitch, a praat script [22] was used to extract 10 F0 points from each vowel and nasal.

For breathiness, we used a praat script [22] to measure the H1-H2 of the vowels at three portions of each phone and only the data in the middle portion were used to ensure liability [23]. Pitch span and mean pitch were calculated in semitones (st) with a reference of 50 Hz.

2.5. Statistical analysis

We ran a series of linear mixed-effect models using the lmer function of the lme4 package [24] in R [25]. SPEECH RATE, PITCH SPAN, MEAN PITCH or BREATHTHNESS were the dependent variables. All models started with a full model, including ATTITUDE (polite, neutral) and UTTERANCE TYPE (statement, yes/no question, imperative) as within-subject fixed factors, and GENDER as a between-subject fixed factors, random intercepts for PARTICIPANTS and ITEMS, and by-PARTICIPANTS and by-ITEMS random slopes for the effect of politeness [26]. GENDER was included as a fixed factor in order to explore whether it might interact with ATTITUDE and UTTERANCE TYPE. Little is known on gender-related differences in prosodic realization of politeness. But gender might play a role considering that fact that male speakers typically have a narrower pitch range and a lower mean pitch than female speakers. This may in turn influence to what extent male and female speakers use pitch and other prosodic parameters to express politeness.

An example of the structure of the full model is given in (7).

(7) lmer (Speech Rate ~ Utterance Type \(\times\) Attitude \(\times\) Gender + Utterance Type \(\times\) Attitude + Utterance Type \(\times\) Gender + (1 + Attitude|Item) + (1 + Attitude |Participant), data=data, REML=F).

In our modelling, predictors were removed from the model sequentially, one by one, starting with the least significant predictor according to the anova function in R, until all the remaining predictors were statistically significant. P-values were obtained by likelihood ratio tests during model comparisons. If there was a significant interaction effect, emmeans and contrast function in "emmeans" package in R [27] were used to examine pairwise differences.

The best-fit lmer model for each of the prosodic measurements are given in (8) to (11).

(8) SpeechRate ~ Gender \(\times\) Attitude \(\times\) Utterance Type + (1 + Attitude|Item) + (1 + Attitude |Participant), data=data, REML=F, \(\chi^2(6) = 237, p<.0001\)

(9) PitchSpan ~ Utterance Type + Attitude + Gender + (1|Item) + (1+Attitude|Participant), data=data, REML=F, \(\chi^2(8)=82.1, p<.0001\)
3. Results

3.1. Speech rate

Speech rate was consistently increased from neutral to polite condition across gender and utterance type \([\beta=0.368, \text{SE}=0.141, t=2.609, p=0.014]\) (Figure 1). We also found a significant interaction of attitude, utterance type and gender. For both genders, the increase of speech rate in statements were the largest, but the rank of the increase across utterance type was not consistent. For the female speakers, the increase of speech rate in statements was slightly larger than that in yes/no questions (but not significantly), and the increase in yes/no questions was significantly larger than that in imperatives \([\beta=0.1974, \text{SE}=0.080, t=2.446, p=0.0146]\). For the male speakers, the increase of speech rate was the largest in statements, less large in imperatives, and the least in yes/no questions. Only the difference of speech rate increase in statements and yes/no questions were significant \([\beta=0.2286, \text{SE}=0.080, t=2.712, p=0.0068]\).

![Figure 1: Mean speech rate for polite and neutral speech](image)

The rank of speech rate increase from neutral to polite condition for female speakers: St> YN > Im; The rank for male speakers: St> Im > YN (St> YN). * is the significant code at 0.05 level.

3.2. Mean pitch

As shown in Figure 2, the mean pitch in the polite speech condition was significantly higher than the neutral speech condition \([\beta=0.108, \text{SE}=0.032, t=3.34, p<0.05]\). Furthermore, there was a main effect of utterance type. That is, the mean pitch of yes/no questions was significantly higher than the mean pitch of imperatives \([\beta=0.569, \text{SE}=0.078, t=7.27, p<0.001]\); the mean pitch of imperatives was also significantly higher than the mean pitch of statements across different genders and attitudes \([\beta=0.189, \text{SE}=0.078, t=2.42, p<0.05]\). The interactions gender \times\ attitude and utterance type \times\ attitude were not significant.

![Figure 2: Mean pitch for neutral and polite speech](image)

3.3. Pitch span

As shown in Figure 3, both male and female speakers significantly increased their pitch span in the polite condition \([\beta=0.244, \text{SE}=0.080, t=3.05, p<0.01]\). There was also a main effect of gender: the male speakers used significantly larger pitch span than the female speakers in both neutral and polite condition across different utterance types \([\beta=0.303, \text{SE}=0.131, t=2.32, p<0.05]\). Moreover, we found a significant main effect of utterance type; the pitch span was wider in yes/no questions than in statements \([\beta=0.601, \text{SE}=0.132, t=4.550, p<0.001]\), and was the largest in imperatives \([\beta=0.493, \text{SE}=0.132, t=3.73, p<0.001]\) (Figure 4). The interaction between attitude and utterance type as well as the interaction between attitude and gender were not significant.

![Figure 3: Mean F0 span of female and male speakers for polite and neutral speech](image)

3.4. Breathiness

Speakers used significantly breathier voice to express polite speech than neutral speech across different utterance types \([\beta=2.305, \text{SE}=2.75, t=2.75, p<0.001]\) (Figure 5). Moreover, female speakers used significantly breathier voice than male speakers \([\beta=0.314, \text{SE}=0.146, t=2.14, p<0.05]\) (Figure 6). The interaction between attitude and gender was not significant.

There was also a significant main effect of utterance type on breathiness. Speakers used significantly breathier voice in yes/no questions than in imperatives and statements \([\beta=0.557, \text{SE}=0.096, t=5.32, p<0.001]\; [\beta=0.694, \text{SE}=0.096, t=7.29, p<0.001]\), but there was no significant difference in their use of breathy voice between imperatives and statements \([\beta=0.137, \text{SE}=0.096, t=1.44, p>0.05]\) (Figure 5). The interaction between utterance type and gender was not significant.

![Figure 4: Mean F0 span per utterance type for neutral and polite speech](image)

![Figure 5: Mean F0 span per utterance type for neutral and polite speech](image)
4. Discussion and conclusions

The present study examined whether the prosody of politeness is realized differently in the presence of the hesitation marker en ‘uhm’ in different utterance types in Mandarin Chinese. We have found both similarities and differences in the use of prosody between speech with and without a hesitation marker. Specifically, speakers use a faster speech rate, a higher mean pitch and a breathier voice in polite speech than in neutral speech, in line with the findings from speech without a hesitation marker [2,5,10,11,12,13]. But speakers also use pitch span to express politeness in the presence of a hesitation marker contra the finding from earlier studies on polite speech that pitch span is not distinctively different in polite and neutral Mandarin speech [5,10]. In addition, speakers use speech rate, mean pitch, pitch span and breathiness in different utterance types to the same degree in order to distinguish polite speech from neutral speech.

To conclude, prosody is not used to a lesser extent in politeness speech with a hesitation marker and the morphosyntactic cues do not offset the function of prosody as we hypothesized. Our study suggests that a functional trade-off between prosody and non-prosodic cues does not necessarily occur for different meaning attributes.

5. Acknowledgements

We appreciate the help of Can Hu, Guiyue Lu, Jiawei Song, Yutong Wang, Yuzhen Zhang, Siran Liu, Yichen Liu, Junhong Zhang, Wanjun Meng, Junyi Mao, and Yanlei Su.

6. References


