Same Same But Different: The Influence of Ambiguity Awareness on Speech and Gesture Production

Jiajun Gao (hvyjg1@nottingham.edu.cn)
School of Education and English, 199 Taikang E Rd, Yinzhou, Ningbo, Zhejiang, China

Yan Gu (yan.gu@essex.ac.uk)
Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK
Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, UK

Abstract

We explored (1) the differences in prosody and gesture when speakers were aware and unaware of ambiguities, and (2) the insight of multimodal ambiguity resolution on communication efficiency. Thirty-two Mandarin speakers articulated twenty-two ambiguous Mandarin sentences. Half could be disambiguated using prosody (half couldn't). First, participants articulated each sentence and explained its meaning to a confederate, revealing their dominant interpretation and ambiguity awareness. Second, participants articulated the same ambiguous sentences twice according to hints indicating two meanings. Results showed participants hardly realised ambiguities. Speakers produced mostly more prominent prosody and more gestures when recognising ambiguities. When ambiguity was aware, prosodically unambiguous sentences were produced with various prosodic cues, with referential and non-referential gestures. However, prosodically ambiguous sentences were produced with more referential but hardly any non-referential gestures. In conclusion, speakers adopt multimodal strategies to achieve communication efficiency with a trade-off between modalities, depending on their ambiguity awareness.

Keywords: prosody and gesture; Chinese; ambiguity awareness; multimodal ambiguity resolution; communicative efficiency and effort; trade-off hypothesis

Introduction

Inherent in language communication is the challenge of linguistic ambiguity, in which a single expression may give rise to multiple interpretations (Biau et al., 2018), potentially leading to misunderstandings (Harley, 2013; Harley, 2017; Warren, 2013). While there are many ways to disambiguate, such as sentence processing (Chernova, & Chernigovskaya, 2015) and using contextual cues (Molet et al., 2010), in face-to-face communication, prosody and gestures are crucial means.

Prosodic cues, including pause, stress, rhythm, and intonation of language (Allbritton et al., 1996; Tseng et al., 2005; Wennerstrom, 2001; Xu, 2001), are crucial for disambiguation (Fodor, 2002; Schafer et al., 2005; Snedeker & Trueswell, 2003). For instance, different word duration (Wiener et al., 2012), pause duration (Gollrad et al., 2010), and prosodic contour duration (Lamekina & Meyer, 2023) can all have a positive impact on listeners' interpretation of ambiguous sentences. Furthermore, the role of prosodic cues is not limited to the sentence being communicated at the moment. It can also assist in predicting forthcoming ambiguous structures (Lamekina & Meyer, 2023; Snedeker & Trueswell, 2003), prompting listeners to attend to the prosodic cues in the subsequent segments, thus achieving efficient communication. Interestingly, prosodic cues employed by Chinese speakers in disambiguation rely on the ambiguity type, such as using stress to resolve structural ambiguities (Zhou et al., 2012). In addition, pauses can also be seen as effective disambiguation cues (Jun, 2005; Nespor & Vogel, 2007; Shen, 1993). For instance, in the sentence '他让赵先生本月 15 日前去汇报' (tā ràng zhào-xiān-sheng běn-yuè shí-wǔ-rì qián qù huì-bào), pausing before the target character '前' means 'he tells Mr. Zhao to report on the 15th of this month' while pausing after '前' means 'he tells Mr. Zhao to report by the 15th of this month'.

Nevertheless, prosodic cues alone are sometimes insufficient to fully resolve ambiguities. This may be due to various reasons such as speakers' incompetence with prosodic cues, marked by reduced sensitivity to speech prosody as they age (Keller, 2006), coupled with cognitive impairments (Diehl et al., 2008) or auditory deficiencies (Hopyan-Misakyan et al., 2009). Another factor is that the ambiguity itself cannot be addressed by prosodic cues, owing to the diversity of linguistic features, which becomes more salient within ambiguous Chinese sentences. Chinese displays special phonetic features that are less common in Indo-European languages. For instance, homophonic words, sharing identical pronunciations while conveying disparate meanings, are pervasive in Chinese (Grzybek, 2009). For example, the Chinese sentence "他倒了一杯水" (tā dào-le yì-bēi-shuǐ) can convey either the meaning "He fills the cup with water." or "He empties the cup.". The character "倒" can mean either pour *into* or pour *out*, leading to a lexical ambiguity. Compared to seeking a prosodic resolution, this ambiguous instance is easier to resolve if the speaker gestures "pour into" or "pour out", because "倒" sounds nearly identical when it refers to "pour into" or "pour out".

The above example shows that gestures can resolve ambiguities and facilitate communication efficiency. Communication efficiency pertains to effectively transmitting information between communicators with minimal effort (Grzyb et al., 2022; Rasenberg et al., 2022). In this context, effort encompasses the cognitive

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

and physical resources expended by both the speaker and listener during communication (Rasenberg et al., 2022). While prosody and gestures are vital in disambiguating sentences, their separate or combined effects on communication efficiency and effort remain unclear.

Furthermore, somewhat in line with the communication efforts, previous studies argue that whether or not being aware of the existence of ambiguity influences speakers' use of audio and visual cues to disambiguate (Zabotkina et al., 2020). For example, adults use prosodic cues for disambiguation *only* when they recognize ambiguities (Fox Tree & Meijer, 2000; Kraljic & Brennan, 2005; Snedeker & Trueswell, 2003), and speakers employ gestures to aid communication when they are aware of possible verbal ambiguities (Holler & Beattie, 2003). However, we have limited knowledge of how the awareness of ambiguity affects both prosody and gesture, but such research can better understand speakers' multimodal disambiguation, as well as the trade-off between information modalities.

While previous studies on disambiguation through gestures covered various ambiguity types and different age cohorts, they have excessively centred around Indo-European languages (Henrich et al., 2010). Regardless of whether in children or adults, gestures consistently demonstrate their value in facilitating communication (Biau et al., 2018; Brown & Kamiya, 2016; Holle et al., 2012; Kidd & Holler, 2009; Kita, 2014; Okahisa & Shirose, 2018; Smith & Kam, 2015; Yow, 2015). Furthermore, by addressing ambiguities, gestures improve robots' ability to comprehend human instructions more precisely (Botting et al., 2010; Scholl & McRoy, 2019; Weerakoon et al., 2020). Nevertheless, little research is on the role of gestures in non-Indo-European languages (Vigliocco et al., 2014), while Chinese possess special linguistic features that generate ambiguities that are less common in Indo-European languages. Particularly, different interpretations of some ambiguous Chinese sentences cannot be phonetically distinguished due to the absence of discernible phonetic differences. In addition, the Chinese language relies on a higher degree of contextuality than English (Watkins & Biggs, 1996), and an analysis of Chinese may shed light on different patterns of gestural resolution of ambiguity.

Thus, this study examined the effectiveness of audio and visual resolution in clarifying ambiguous Chinese sentences when speakers were aware and unaware of ambiguities, and when prosody can and cannot mark ambiguities. By addressing these aspects, we aimed to better understand ambiguity resolution, the interplay between prosody and gestures, and their facilitation of communication. We asked two research questions:

RQ1: What are the differences in employing prosody and gesture when speakers are aware and unaware of ambiguity? We hypothesize that being aware of ambiguities leads to an increased saliency of prosody (Snedeker & Trueswell, 2003) and gesture frequency in comparison with failing to recognise ambiguity.

RQ2: What are the differences in multimodal resolutions for prosodically ambiguous sentences compared to prosodically unambiguous sentences? According to the communication efficiency hypothesis, participants are less likely to gesture when speech prosody alone is sufficient to disambiguate. However, when prosodic differences cannot address ambiguities, participants may make efforts to produce more gestures.

Methodology

Participants

Thirty-two Chinese-native students (5 males, 27 females) (Mean age = 20.97 years, range 19 - 23 years) from the University of Nottingham Ningbo China participated in this study for course credit. The number of participants was decided based on a power analysis using the G*Power (version 3.1) with 0.8 power and a medium effect size of 0.5 (Field et al., 2012). In addition, the researcher appointed one confederate per participant to stimulate participants' communicative intent. These confederates were 32 additional recruits or participants who had previously completed the experiment. All participants exhibited no hearing or speech impairments. Participants signed an informed consent. The study obtained ethical approval from the University of Nottingham Ningbo China.

Apparatus and stimuli

The stimuli comprised 22 ambiguous Chinese sentences adapted from Huang and Li (2012), each having two different interpretations (see the full stimuli in the OSF file: https://osf.io/8djwm). The sentences were divided equally into two groups of disambiguation types. The first group (N = 11) could be disambiguated through prosodic cues, such as a pause and stress. Conversely, the second group (N = 11) is challenging to disambiguate solely with prosodic cues.

There was a consistency of the disambiguation types for these ambiguous sentences between the two Chinese-native authors. Additionally, 15 naive raters, blind to the research purpose, independently judged if prosodic cues could assist in clarifying the ambiguity of these 22 Chinese sentences. The overall consensus reached 96.97% for prosodically unambiguous sentences (N = 11) and 90.91% for prosodically ambiguous sentences (N = 11).

Procedures

First, in Exp1, participants saw each ambiguous sentence on a computer screen without hints, requiring them to read the sentence to the confederate and then explain it in their own words. Each sentence was independently presented on a PowerPoint slide. As participants interpreted the 22 sentences intuitively and spontaneously, their explanations revealed participants' dominant interpretations and awareness of ambiguity. If they were aware of the ambiguity, they should provide two interpretations of the same sentence. If not, their explanation would reflect their dominant interpretation. In addition to collecting spontaneous awareness, the purpose of obtaining the dominant interpretation was to control for the possible effect of a non-dominant interpretation (less predictable) on prosodic production in Exp2. During the experiment, confederates were not encouraged to give feedback, although nods and headshakes were allowed.

In Exp 2, participants came across the same sentences as Exp 1 again, but this time they saw each of the two hints of the same sentence (suggesting two meanings) on two slides. For instance, the sentence "王先生借了李先生一本书" (wáng-xiān-shēng jiè-le lǐ-xiān-shēng yì-běn-shū) can be either interpreted as "Mr. Wang lent a book to Mr. Li" or "Mr. Wang borrowed a book from Mr. Li". On one slide, participants saw the target sentence with the hint "借出" (lend) underneath it, and on another slide, they saw the same sentence with a different hint "借入" (borrow). For each slide, participants solely articulated the target ambiguous sentence (but not the hint) according to the hint information. Confederates were not encouraged to give feedback but solely indicated whether they understood the participant's meaning with nods or headshakes. To motivate the communicative intent of speakers, they were told and could see that the confederate would guess and mark down what interpretation the sentence referred to (mean accuracy rate = 94.18%). Speakers were not told to use prosodic or gesture cues. All participants first took part in Exp1, followed by the Exp2. The sequence of these 22 Chinese sentences was randomised, creating two counterbalanced versions. The order of the two hints was also counterbalanced.

All stimuli were displayed on a computer screen (MacBook Pro, resolution 2560×1600). Participants completed experiments in a spacious, lit, and quiet room. They were audiovisually recorded using Audacity 3.3.2 (16 bit; 44.1 kHz) and a phone camera (4K; 30 fps).

Annotations

Speech articulations were annotated in Praat 6.3.10 (Boersma & Van Heuven, 2001). From Exp1, participants' dominant interpretation of ambiguous sentences and their ambiguity awareness were coded. Most participants had a similar dominant interpretation for 22 ambiguous sentences (M = 82.52%, range 53.12% - 100%). Only one participant recognised the ambiguity of the same sentences, interpreting two meanings in 10 out of 22 ambiguous sentences (45.45%). Specifically, this participant recognised significantly more ambiguous instances when interpreting prosodically unambiguous sentences (N = 7 out of 11, 63.64%) than those in prosodically ambiguous sentences (N = 3 out of 11, 27.27%), binominal sign test, $p = .013$, 95% CI [0.35, 1.0]. Conversely, the remaining participants articulated only one interpretation of the ambiguous sentences, indicating a general lack of awareness of the semantic ambiguity.

For both experiments, first, utterance boundaries were automatically detected and manually checked in Praat. Second, each sentence was given an ID indicating its meaning (according to participants' interpretation in Exp1 or hints provided in Exp2). Third, we coded whether the hint of the sentence aligned with the participant's dominant interpretation in Exp2 (according to information from Exp1). Fourth, prosodic cues (pausing; lengthening; different pronunciations; accented) were indicated. Repetition, errors, or disfluencies were noted, and the final best production was used.

For sentences that can be marked by prosody, there were four types of coding: (1) Binary encoding to indicate pause positions in sentences (N = 4) that could be disambiguated by pauses. Pauses were labeled as 'before' if they appeared at position A, or as 'nonbefore' (no pause at this position). (2) One sentence '他好说话' (tā hao shuō-huà) can be disambiguated by two distinct tones ('hǎo 3', or 'hào 4'). Participants' production of the third or fourth tone was coded, respectively. (3) Five sentences used stress as prosodic cues, where 'stressed' or 'unstressed' were labeled for the target characters. (4) The last sentence '他们多半是大学生' (tā-men duō-bàn shì dà-xué-shēng) used speech rate for disambiguation, where the target word '多半' could either be articulated with a longer or shorter duration, meaning 'majority' or 'probability'. A Praat script was used to automatically extract the pause, tone, intensity, pitch, and duration of target sentences or items. In addition, the descriptive data of participants' overall speech rates for each sentence were calculated by dividing the number of characters by the duration of that sentence (sec).

Gestures were coded in ELAN 6.5 (Wittenburg et al., 2006). The type of gesture was coded in Exps1 and 2 according to iconic, metaphoric, point, beat, and pragmatics (McNeill, 1992). Furthermore, iconic, pointing, and metaphorical gestures were categorized as referential gestures whereas beats and pragmatic gestures were categorised as non-referential gestures (Graziano et al., 2020; Vila-Gimenez & Prieto, 2021). The descriptive data of referential and non-referential gesture rates were calculated from the binary coding of gesture presence ('1') or absence ('0') in each sentence. Two additional raters, blinded to the aim of this study, independently coded 15% of the participants' (N = 5) gesture performance in Exp2, including the presence of a gesture and the functional distinctions between referential and non-referential gestures. The consistency of identifying whether there was a gesture or not was 98.32%, and the overall agreement of gesture functions was 90.05%.

Statistical analysis

Linear Mixed-Effects models (linear DVs) and GLMM models (binary DVs) in R were used for data analysis (Brown, 2021). First, to examine the effect of awareness of ambiguity on prosody, we compared the saliency of prosodic cues for prosodically unambiguous sentences (e.g., pause length, tone pitch, stressed word duration, and speech rates) between Exp 1 and Exp2.

For gesture analysis, we first compared the frequency of gesture use between participants who were aware and unaware of ambiguities within the Exp1. Then we compared the overall and referential gesture frequency between Exp1 and Exp2.

Furthermore, focusing on Exp2, we investigated whether dominance (hint aligns with the dominant interpretation), prosodic ambiguity (whether prosody can make the distinction), and the interaction between dominance and prosodic ambiguity (IVs) influenced prosodic (e.g., speech rate (words per sec), mean pitch (semitone), mean intensity, intensity maximum, and intensity range and gestural production (referential; nonreferential). Participants and ambiguous sentences

were set as random intercepts and prosodically ambiguity was determined as the random slope to the participant.

Results

Effects of ambiguity awareness

Table 1 presents the descriptive statistics of the prosodic and gestural measures of the sentences produced as dominant interpretations in prosodically unambiguous conditions in Exp1 and Exp2.

Table 1: The mean (M) and standard deviation (SD) for prosodic and gestural features of prosodically unambiguous sentences in dominant interpretations (unaware of ambiguity) in Exp1 (E1) and Exp2 (E2).

Measures	**E1 prosodically unambiguous sentences**	**E2 prosodically unambiguous sentences**
Speech Rate	3.49 (3.07)	3.86 (1.06)
Mean Pitch (ST)	25.82 (5.32)	25.62 (6.36)
Mean Intensity (dB)	60.5 (5.23)	57.52 (7.41)
Max Intensity (dB)	66.3 (6.36)	62.93 (8.51)
Intensity Range (dB)	12.76 (6.64)	11.55 (6.67)
Ref Gesture (%)	0.85 (0.09)	51.42 (0.46)
Non-Ref Gesture (%)	2.41 (0.15)	17.05 (0.29)

Note: Values of mean pitch have been converted to semitone.

Prosody

In participants' dominant interpretations of prosodically unambiguous sentences (N = 11), when pause can be used as disambiguation cues (N = 4), participants employed significantly longer pause duration when they were aware of ambiguities in Exp2 (M = 0.63 sec, SD = 0.89) compared to failing to recognise ambiguities in Exp1 (M = 0.15 sec, SD = 0.13) (β = 0.04, p < .001). Concerning one sentence using tone variations of the target character '好' ('hǎo 3' or 'hào 4') to disambiguate (N = 1), although speakers in Exp1 who did not recognise ambiguities articulated both the third tone (M = 25.93 ST, SD = 6.51) and the fourth tone (29.37 ST, SD = 7.2) in a higher mean pitch than those in Exp2 (M = 23.71 ST, SD = 3.95 for hao 3, M = 28.5 ST, SD = 7.64 for hao 4), no significant difference was observed between the Exp 1 and the Exp 2 regarding either the third tone (β = 1.09, p = .20) or the fourth tone (β = 1.03, p = .88). In the case of sentences disambiguated by stress (N = 5), longer duration of the stressed characters was found when speakers successfully received the disambiguation cues (M = 0.3 sec, SD = 0.61) compared to when they were unaware of ambiguities (M = 0.26 sec, SD = 1.05) (β = 1.16, p = .007). In addition, neither the mean pitch (M_{Exp1} = 27.67 ST, SD = 4.91; M_{Exp2}= 27.36 ST, SD = 6.03, β = 0.22, p = .85) nor the maximum intensity (M_{Exp1} = 70.25 dB, SD = 4.39; M_{Exp2} = 69.53 dB, SD = 8.24, β = 0.55, p = .65) of the stressed characters were significantly different between Exp1 and Exp2. For the sentence disambiguated by articulating different speaking rates of the target characters '多半' (N = 1), participants spoke significantly slower to express both the meaning of 'majority' (M = 0.58 sec, SD = 0.16) and 'possibility' (M = 0.49 sec, SD = 0.12) when they were aware of the ambiguous situations than when failing to realize the existence of ambiguities (M = 0.43 sec, SD = 0.06, β = - 0.01, p < .001 for 'majority', M = 0.38 sec, SD = 0.07, β = - 0.49, p = .003 for 'possibility').

Gestures

In Exp1 the very speaker who recognised ambiguities exhibited higher gesture rates (M = 31.82%, SD = 0.48, N = 7) than those who failed to be aware of ambiguities (M = 2.35%, SD = 0.15, N = 16).

Additionally, comparing gesture production in Exp1 (unaware; dominant interpretation) with the hinted counterpart in Exp2, participants who were unaware of the ambiguity made significantly fewer gestures (N = 16, M = 2.35%, SD = 0.15) than those in Exp2 where disambiguation hints were provided (N = 564, M = 81.01%, SD = 0.39) (β = - 0.38, p < .001), regardless of disambiguation types. Specifically, speakers in Exp1 gestured less frequently either in facing prosodically unambiguous (N = 7, M = 2.05%, SD = 0.14) or prosodically ambiguous sentences (N = 9, M = 2.64%, SD = 0.16) compared to those in Exp2 (N = 341, M = 69.38%, SD = 0.46 for prosodically unambiguous sentences; N = 341, M = 99.31%, SD = 0.08 for prosodically ambiguous sentences) (Figure 1).

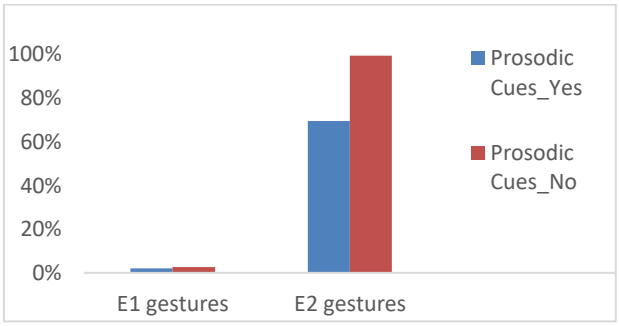

Figure 1: *Participants' gesture performance in Exp1 and Exp2 when articulating ambiguous sentences*.

Moreover, participants who failed to recognise the ambiguity in Exp1 used significantly fewer referential gestures (N = 4) (M = 0.59%, SD = 0.09) than those in Exp2 (N = 511) (M = 74.93%, SD = 0.46) (β = - 0.74, p < .001), irrespective of whether prosodic cues could resolve ambiguities.

Prosodic and gestural resolution in ambiguity

Table 2 presents the descriptive statistics of the prosodic and gestural measures of the sentences produced with two hints in Exp2. Participants were 0.31 syllable faster per second at articulating sentences aligned with their dominant interpretations in comparison to non-dominant interpretations (β = 0.083, p < .001), regardless of whether prosodic cues could resolve ambiguities. Non-dominant interpretations had 8.09 dB higher mean intensity (β = 0.243, p = .027) and 0.5 dB higher maximum intensity (β = 0.439, p = .004) compared to dominant interpretations

for sentences that could use prosodic cues to mark ambiguity. However, neither mean intensity (β = 0.101, p = .514) nor maximum intensity (β = -0.167, p = .437) was significant when ambiguous sentences remained undistinguished by prosodic cues, demonstrating that intensity did not contribute to addressing ambiguities in such instances. There was no significant difference in the mean pitch between dominant and non-dominant interpretations (β = 0.054, p = .589), irrespective of whether prosody could disambiguate.

Table 2: The mean (M) and standard deviation (SD) for prosodic and gestural features of sentences elicited by two different hints in Exp2.

Measures	Hint aligns with the dominant interpretations	Hint does not align with the dominant interpretation
Speech Rate	3.89 (1.07)	3.58 (1.01)
Mean Pitch (ST)	25.64 (4.15)	25.61 (4.22)
Mean Intensity (dB)	50.80 (4.66)	58.89 (4.81)
Max Intensity (dB)	70.74 (4.90)	71.24 (4.96)
Intensity Range (dB)	23.97 (6.20)	24.57 (5.98)
Ref Gesture (%)	75 (0.43)	74.15 (0.44)
Non-Ref Gesture (%)	9.09 (0.29)	9.52 (0.29)

Prosodic resolution

Controlling for participants' dominant interpretations, we focused on sentences that can use prosody to mark ambiguity in Exp2. First, for those ambiguous sentences disambiguated by the pause, a significant influence of two pause positions on disambiguation was identified when participants used pauses as prosodic cues (48.06% for pausing at position A, 51.94% for 'no pauses at this position', β = 7.23, p < .001). For instance, when disambiguating "这种糖果五块五十粒" (zhè-zhǒng táng-guǒ wǔ-kuài-wǔ-shí-lì), pausing before the second "五" meant "5¥ buys 50 candies", while pausing after it meant "5.5¥ buys 10 candies". Second, in the case of ambiguous sentences resolved through two distinct tones, a significant difference emerged in the mean pitch of the two tones (β = 0.29, p < .001). When "好" was pronounced in the third tone (M = 23.05 ST, SD = 4.39) in "他好说话" (tā hǎo-shuō-huà), the sentence meant "he tends to be flexible". However, when "好" was in the fourth tone (M = 28.99 ST, SD = 4.69), the sentence's meaning changed to "he likes talking". Third, when participants resolved ambiguities by stressing characters, compared to unstressed characters, they articulated stressed characters with longer duration (β = 0.092, p < .001), wider intensity range (β = 2.912, p < .001), higher maximum intensity (β = 2.151, p < .001), and higher mean pitch (β = 0.29, p = .009) (see Table 3). For instance, when participants stressed "起来" in "我想起来了" (wǒ xiǎng qǐ-lái-le), the sentence meant "I want to stand up" while unstressing "起来" changed the meaning to "it comes to my mind". Finally, for the sentence "他们多半是大学生" (tā-men duō-bàn shì dà-xué-shēng) where the speech rate of target words "多半" aided in disambiguating, its meaning of "majority" had a longer duration (M = 0.58, SD = 0.15) than its meaning of "probability" (M = 0.47, SD = 0.12), β = 0.094, p = .0002.

Table 3: The mean (M) and standard deviation (SD) of prosodic features for sentences disambiguating by stress.

Measures	Stressed characters	Unstressed characters
Duration (sec)	0.37 (0.15)	0.29 (0.14)
Intensity Range (dB)	14.56 (4.81)	11.95 (5.45)
Max Intensity (dB)	67.99 (5.80)	66.21 (5.07)
Mean Pitch (ST)	27.36 (6.03)	25.72 (5.04)

Gestural resolution

There was no significant difference in gesture production between sentences aligned and misaligned with the dominant interpretation (p > 0.05) when participants were aware of ambiguities in Exp2. Importantly, controlling for participants' dominant interpretation, participants were significantly more inclined to gesture when confronted with ambiguous sentences that could not be disambiguated through prosodic cues (M = 98.15%, SD = 0.13, N = 704) compared to sentences that could be disambiguated using prosody (M = 67.05%, SD = 0.47, N = 704) (β = 3.429, p < 0.001). A further analysis according to the referentiality of gestures revealed that such differences were mainly driven by referential gestures such that they were more often in the prosodically ambiguous condition (M = 97.30%, SD = 0.16) than in the prosodically non-ambiguous condition (M = 51.85%, SD = 0.5) (β = 4.352, p < 0.001) (Figure 2). For instance, participants were highly likely to produce gestures for the sentence "王先生借了李先生一本书" (wáng-xiān-shēng jiè-le lǐ-xiān-shēng yì-běn-shū) that prosody cannot mark distinctions for different meanings ("Mr. Wang lent a book to Mr. Li" and "Mr. Wang borrowed a book from Mr. Li" (Figure 3).

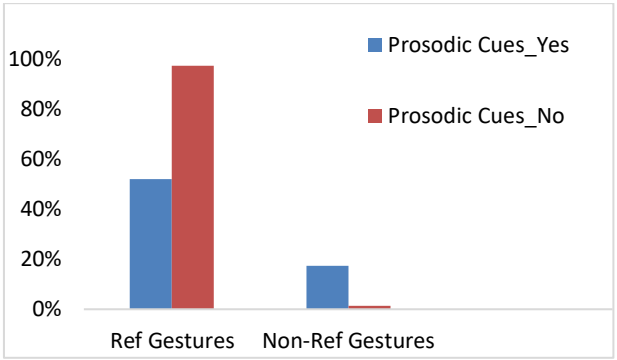

Figure 2: *Participants' gesture performance when articulating ambiguous sentences in Exp2.*

Furthermore, the proportion of non-referential gestures (M = 17.33%, SD = 0.39) was higher when prosodic cues effectively resolved ambiguities than when prosody could not resolve ambiguities (M = 1.28%, SD = 0.11), β = 5.029, p = .003. Specifically, there were more beats (M = 0.056, SD = 0.231, β = 2.613, p < 0.001) and pragmatic gestures (M = 0.116, SD = 0.321, β = 3.282, p < 0.001) in the prosodically unambiguous sentences compared to the

prosodically ambiguous sentences (M_{beats} = 0.005, SD = 0.075, $M_{pragmatic}$= 0.007, SD = 0.084).

Figure 3: *Gestures in two interpretations of "王先生借了李先生一本书": (a)"Mr. Wang lent a book to Mr. Li"; (b)"Mr. Wang borrowed a book from Mr. Li".*

Discussions

This study examined speakers' prosodic and gestural resolution of ambiguities (being aware or unaware) in Chinese sentences and explored the implications of audiovisual means for efficient communication. The findings revealed that very few people were aware of ambiguities in spontaneous speech without a hint (Foerst, 2017) (only one participant recognised ambiguities). Speakers also hardly made any gestures when they were unaware of an ambiguity. There were both effects of ambiguity awareness and capacity of prosodic disambiguation on multimodal production.

Past research showed that when recognising an ambiguity, speakers employed more prominent prosodic cues such as longer pauses, stressed characters, and slower speaking rates to disambiguate (Diehl et al., 2008; Snedeker & Trueswell, 2003; Zabotkina et al., 2020; Wiener et al., 2012). Participants in our research used pausing, stressed characters with longer pauses, higher mean pitch, and maximum intensity to mark ambiguities, despite no significant difference in the mean pitch and maximum intensity of stressed characters between those recognising and unrecognising the ambiguity. Given the tonal systems of Mandarin (Huang & Li, 2012; Jongman et al., 2006), participants employed two distinct tones to resolve ambiguities, although there was no significant difference in the mean pitch between aware and unaware of ambiguities. This could be explained by the more effective way of using a tonal contrast between 'hao 3' and 'hao 4' rather than solely relying on acoustic differences to resolve ambiguities, thus highlighting the dynamic nature of speech production.

In addition, when speakers were aware of ambiguities in Exp2, non-dominant interpretations had higher mean and maximum intensity than dominant ones when prosodic cues effectively disambiguated sentences, indicating that speakers still made efforts to highlight the unmarked interpretation but only when such information was possible to be informative in prosody. However, the speech rates of participants' dominant interpretations were faster than those of non-dominant interpretations, irrespective of whether prosodic cues could resolve ambiguities. This is because the duration of words and sentences was longer when speakers articulated less predictable non-dominant meanings (Seyfarth, 2014).

Furthermore, ambiguity awareness impacts gesture production. In Exp1 the only participant being aware of ambiguities gestured more frequently than that of other participants who were unaware of ambiguities. Similarly, participants gestured significantly more (e.g., increased number of referential gestures) when gaining the disambiguation hints in Exp2 than without having any ambiguity awareness (Exp1).

Thus, speakers indeed used multimodal marking of ambiguities in Chinese sentences. Particularly, even when prosodic cues alone were sufficient for disambiguation, participants still exhibited a high proportion of referential gestures (51.85%). Additionally, participants also produced a rather high proportion of non-referential gestures (17.33%). Interestingly, such non-referential gestures (e.g., beats and pragmatic gestures), were accompanied by prosodic prominence, as supported by (Krahmer & Swerts, 2007). 79.21% of the beats were accompanied by a prosodic accentuation.

By contrast, participants exhibited an extremely higher rate of gestures (98.15%) when prosodic cues were insufficient in resolving ambiguities than when prosody successfully functioned. This suggests a stronger tendency for a multimodal approach in disambiguation (Holler & Beattie, 2003; Khalili et al., 2014; Kidd & Holler, 2009), and communication in general (Higham & Hebets, 2013; Holler & Levinson, 2019; Vigliocco et al., 2014). These could be explained by the trade-off hypothesis between resolving ambiguities and achieving communication efficiency (Grzyb et al., 2022; Rasenberg et al., 2022), indicating a balance between competing goals in communication and manual efforts.

Additionally, there was a reduced occurrence of non-referential gestures when prosodic cues were ineffective in disambiguating. This could be due to the fact that different types of gestures also compete in gesture production, and non-referential gestures were not prioritized in resolving ambiguity. They were more often produced in prosodically unambiguous sentences where the coupling use of prosodic prominence and beat gestures demonstrated a parallel in prosody and gesture, employing both channels at the same time.

Conclusion

This first study on the audiovisual resolution of ambiguities in Chinese sentences revealed that speakers employed various strategies in disambiguation for effective communication, which became more prominent when they were aware of ambiguities. They used pauses, tone contrasts, stressed characters, and speech rates, along with gestures to disambiguate prosodically unambiguous sentences, but more referential gestures to clarify prosodically ambiguous ones. In addition, speakers employed more salient prosody (longer pauses; slower speech rates, etc.) coupled with a higher gesture rate when they recognised ambiguities. In sum, speakers' ambiguity awareness influences their prosodic and gestural production and they adopt a multimodal approach to achieve communication efficiency with a trade-off between modalities. Future should investigate how multimodal disambiguation facilitates comprehension.

Acknowledgments

We thank all participants. Preliminary results are reported in Speech Prosody 2024. The work is supported by The National Social Science Fund of China (20BYY179).

References

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition*, *22*(3), 714.

Biau, E., Fromont, L. A., & Soto-Faraco, S. (2018). Beat gestures and syntactic parsing: an ERP study. *Language Learning*, *68*, 102-126.

Boersma, P., & Van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glot International*, *5*(9/10), 341-347.

Botting, N., Riches, N., Gaynor, M., & Morgan, G. (2010). Gesture production and comprehension in children with specific language impairment. *British Journal of Developmental Psychology*, *28*(1), 51-69.

Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920960351.

Brown, A., & Kamiya, M. (2016). Gesture in the resolution of syntactic ambiguity: Negation and quantification in English. *UK-CLC*, 18.

Chen, T., & Rao, R. R. (1998). Audio-visual integration in multimodal communication. *Proceedings of the IEEE*, *86*(5), 837-852.

Chernova, D., & Chernigovskaya, T. V. (2015). Syntactic Ambiguity Resolution in Sentence Processing: New Evidence from a Morphologically Rich Language. *EAPCogSci*, *1419*, 129-133.

Diehl, J. J., Bennetto, L., Watson, D., Gunlogson, C., & McDonough, J. (2008). Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain and Language*, *106*(2), 144-152.

Field, Z., Miles, J., & Field, A. (2012). Discovering statistics using R. *Discovering Statistics Using R*, 1-992.

Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. In *Speech Prosody 2002, International Conference*.

Foerst, A. (2017). Ambiguity. *CrossCurrents*, *67*(4), 653-665

Fox Tree, J. E., & Meijer, P. J. (2000). Untrained speakers' use of prosody in syntactic disambiguation and listeners' interpretations. *Psychological Research*, *63*(1), 1-13.

Gollrad, A., Sommerfeld, E., & Kügler, F. (2010). Prosodic cue weighting in disambiguation: Case ambiguity in German. In *Speech Prosody 2010-Fifth International Conference*.

Graziano, M., Nicoladis, E., & Marentette, P. (2020). How referential gestures align with speech: Evidence from monolingual and bilingual speakers. *Language Learning*, *70*(1), 266-304.

Grzyb, B., Frank, S. L., & Vigliocco, G. (2022). Communicative efficiency in multimodal language. https://doi.org/10.31234/osf.io/a9wt3

Grzybek, J. (2009). Polysemy, homonymy and other sources of ambiguity in the language of Chinese contracts. *Comparative Legilinguistics*, *1*, 207–215. https://doi.org/10.14746/cl.2009.01.15

Harley, T. A. (2013). *The psychology of language: From data to theory*. Psychology Press.

Harley, T. A. (2017). *Talking the talk: Language, psychology and science*. Psychology Press.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and Brain Sciences*, *33*(2-3), 61-83.

Higham, J. P., & Hebets, E. A. (2013). An introduction to multimodal communication. *Behavioral Ecology and Sociobiology*, *67*, 1381-1388.

Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A. D., Ward, J., & Gunter, T. C. (2012). Gesture facilitates the syntactic analysis of speech. *Frontiers in Psychology*, *3*, 74.

Holler, J., & Beattie, G. (2003). Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener?. *Gesture*, *3*(2), 127-154.

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639-652.

Hopyan-Misakyan, T. M., Gordon, K. A., Dennis, M., & Papsin, B. C. (2009). Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. *Child Neuropsychology*, *15*(2), 136-146.

Huang, B. R., & Li, W. (2012). Xiandai hanyu [Modern Chinese], *Beijing Book Co. Inc. 2012 (Vol. 2).*

Jongman, A., Wang, Y., Moore, C., and Sereno, J. (2006). "Perception and production of mandarin tone," in Handbook of East Asian Psycholinguistics. Vol. 1. eds. P. Li, L. H. Tan, E. Bates, and O. J. L. Tzeng (England: Cambridge University Press), 209–217.

Jun, S. A. (Ed.). (2005). *Prosodic typology: The phonology of intonation and phrasing*. OUP Oxford.

Keller, B. Z. (2006). Ageing and speech prosody. In *Speech Prosody* (Vol. 2006, pp. 1-5).

Khalili, M., Rahmany, R., & Zarei, A. A. (2014). The Effect of Using Gesture on Resolving Lexical Ambiguity in L2. *Journal of Language Teaching & Research*, *5*(5).

Kidd, E., & Holler, J. (2009). Children's use of gesture to resolve lexical ambiguity. *Developmental Science*, *12*(6), 903-913.

Kita, S. (2003). Pointing: A foundational building block of human communication. In *Pointing* (pp. 9-16). Psychology Press.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, *57*(3), 396-414.

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee?. *Cognitive Psychology*, *50*(2), 194-231.

Lamekina, Y., & Meyer, L. (2023). Entrainment to speech prosody influences subsequent sentence

comprehension. *Language, Cognition and Neuroscience*, *38*(3), 263-276.
McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.
Molet, M., Urcelay, G. P., Miguez, G., & Miller, R. R. (2010). Using context to resolve temporal ambiguity. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(1), 126.
Nespor, M., & Vogel, I. (2007). 1 986, Prosodic phonology. *Dordrecht: Foris*.
Okahisa, T., & Shirose, A. (2018). Influence of hand gestures on prosodic disambiguation of syntactically ambiguous phrases. *Acoustical Science and Technology*, *39*(2), 171-174.
Rasenberg, M., Pouw, W., Özyürek, A., & Dingemanse, M. (2022). The multimodal nature of communicative efficiency in social interaction. *Scientific Reports*, *12*(1), 19111.
Schafer, A. J., Speer, S. R., & Warren, P. (2005). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. *Approaches to Studying World-situated Language Use*, 209-225.
Scholl, C., & McRoy, S. (2019). Using gestures to resolve lexical ambiguity in storytelling with humanoid robots. *Dialogue & Discourse*, *10*(1), 20-33.
Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, *133*(1), 140-155.
Shen, X. S. (1993). The use of prosody in disambiguation in Mandarin. *Phonetica*, *50*(4), 261-271.
Smith, W. G., & Kam, C. L. H. (2015). Children's use of gesture in ambiguous pronoun interpretation. *Journal of Child Language*, *42*(3), 591-617.
Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*(1), 103-130.
Tseng, C. Y., Pin, S. H., Lee, Y., Wang, H. M., & Chen, Y. C. (2005). Fluent speech prosody: Framework and modeling. *Speech Communication*, *46*(3-4), 284-309.
Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130292.
Vila-Gimenez, I., & Prieto, P. (2021). The value of non-referential gestures: A systematic review of their cognitive and linguistic effects in children's language development. *Children*, *8*(2), 148.
Warren, P. (2013). *Introducing psycholinguistics*. Cambridge University Press.
Watkins, D. A., & Biggs, J. B. (1996). *The Chinese learner: Cultural, psychological, and contextual influences*. Comparative Education Research Centre, Faculty of Education, University of Hong Kong, Pokfulam Road, Hong Kong; The Australian Council for Educational Research, Ltd., 19 Prospect Hill Road, Camberwell, Melbourne, Victoria 3124, Australia..
Weerakoon, D., Subbaraju, V., Karumpulli, N., Tran, T., Xu, Q., Tan, U. X., ... & Misra, A. (2020). Gesture enhanced comprehension of ambiguous human-to-robot instructions. In *Proceedings of the 2020 International Conference on Multimodal Interaction* (pp. 251-259).
Wennerstrom, A. (2001). *The music of everyday speech: Prosody and discourse analysis*. Oxford University Press.
Wiener, S., Speer, S. R., & Shank, C. (2012). Effects of frequency, repetition and prosodic location on ambiguous Mandarin word production. In *Speech Prosody 2012*.
Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *5th International Conference on Language Resources and Evaluation (LREC 2006)* (pp. 1556-1559).
Xu, Y. (2011). Speech prosody: A methodological review. *Journal of Speech Sciences*, *1*(1), 85-115.
Yow, W. Q. (2015). Monolingual and bilingual preschoolers' use of gestures to interpret ambiguous pronouns. *Journal of Child Language*, *42*(6), 1394-1407.
Zabotkina, V., Bottineau, D., & Boyarskaya, E. (2020). Cognitive mechanisms of ambiguity resolution. In *International Conference on Cognitive Sciences* (pp. 201-212). Cham: Springer International Publishing.
Zhou, P., Su, Y. E., Crain, S., Gao, L., & Zhan, L. (2012). Children's use of phonological information in ambiguity resolution: a view from Mandarin Chinese. *Journal of Child Language*, *39*(4), 687-730.