

## Multimodal Resources in Turn-Taking in Semi-Institutional Mandarin Multiparty Interactions

Jiixin Tian

### Abstract

This study investigates the utilization of multimodal resources in organizing turn-taking during multiparty interactions in a Mandarin talk show. By applying multimodal conversation analysis and interactional linguistics to 5.8 hours of impromptu talk show data, the study reveals that the chair and the other participants orient to their dual roles, both institutional and real-life, to configure a semi-institutional setting. Besides, the multimodal resources can be effectively used by the participants, i.e., the host and the guests, to manage contingencies during turn-taking, including visible cues, embodied movements, and pragmatic (in)completion. The findings contribute to our understanding of the dynamics of turn-taking in semi-institutional settings and shed light on the interplay of multimodal resources in larger group conversations. The research expands the existing literature on multiparty institutional conversations in Mandarin.

**Keywords:** multimodal conversation analysis, turn-taking, multiparty, semi-institution, Mandarin

### 1 Introduction

Turn-taking is a fundamental principle of social interaction, which is not predetermined but is interactively organized between the participants in ordinary conversation (Couper-Kuhlen & Selting, 2018). Sacks et al. (1974) observed that the first possible completion point is projectable so that turn-taking is possible. Since then, various studies have shown how syntactic, intonational, pragmatic, and embodied resources (gaze and body movements) contribute to the projection of possible turn completion points (Ford & Thompson, 1996; Ford et al., 1996; Tanaka, 1999; Li, 2014). In institutional interactions, however, turn-taking systems are characterized by varying restrictions on the speaker, timing, and content of the next turn (Heritage & Clayman, 2010). Though topics and actions have been organized from the outset, the “chair” and participants can locally manage their turns via visible resources (e.g., gaze, gestures, body torso) with verbal conduct, which are crucial for interactions in large groups (Markaki & Mondada, 2012; Mondada, 2013a).

Research on how various resources contribute to the turn-taking system in Mandarin began four decades ago (Thompson, 2019), and many studies have observed that the organization of turn-taking in Mandarin is related to syntax (Li, 2014; Gao & Ren, 2021), prosody (Li, 2014; Lim, 2019), gestures (Yang, 2011; Tao, 2019), gaze (Yang, 2011; Li, 2014; Endo, 2019) and pragmatic actions (Li, 2014). However, many of these previous studies involving Mandarin focused on ordinary conversations. How turn-taking organizations are managed with multimodal resources in multiparty institutional interactions (i.e., meetings) is underexplored in Mandarin Chinese.

This study utilizes multimodal conversation analysis (Mondada, 2016) and interactional linguistics to explore how multimodal resources are jointly used to organize turns in a semi-institutional, multiparty talk. The data was collected from a Mandarin TV talk show, and the focus of this study is on how the host and guests on the show use multimodal resources to manage turn-taking organization and fulfill their responsibilities within this semi-institutional setting. Through

an analysis of these interactions, we aim to gain a deeper understanding of the dynamics of institutional interactions and how participants use multimodal resources to manage turn-taking in such contexts.

## 2 Turn-taking

### 2.1 Turn-taking in institutional interaction

In a conversation, a turn refers to an utterance produced by a speaker when they occupy the floor (Couper-Kuhlen & Selting, 2018). Upon the completion of the first speaker's turn, the next speaker, based on their mutually shared knowledge, can orient to take the floor. Turn-taking in ordinary conversations is not predetermined but instead is interactively and locally organized between the participants through the turn-constructive component and the turn-allocation component (Sacks et al., 1974, p.727). The basic unit of the turn-constructive component is known as "turn-constructive units (TCUs)" (Sacks et al., 1974, p.702). TCUs are projectable, indicating that the recipient can anticipate the first possible completion point in advance of its occurrence (Couper-Kuhlen & Selting, 2018). This point is referred to as a "transition-relevance place (TRP)" (Sacks et al., 1974, p.703), and it is at this point that the turn-allocation component becomes pertinent, determining who can occupy the floor next (Couper-Kuhlen & Selting, 2018).

Sacks et al. (1974) proposed that the techniques of turn-allocation can be divided into two groups: (a) the current speaker selects the next speaker; and (b) the current speaker self-selects. Additionally, they suggested that there is a basic set of rules that dictate how turns are constructed.

(1a) if the turn is constructed in such a way that the current speaker selects the next, the speaker selected has the right and obligation to take the next turn to speak. If, however, (1b) the turn is not constructed in such a way that the current speaker selects the next, next speakers may, but need not, select themselves, with the first starter acquiring rights to the turn. If (1c) no next speaker self-selects, the current speaker may, but need not, continue. If the rules of turn-allocation (1a-c) at the TRP of an initial TCU are applied in such a way that the current speaker continues, they (2) re-apply recursively at each following TRP until a transfer of speakership is ensured. (Couper-Kuhlen & Selting, 2018, p.32)

However, turn-taking procedures in institutional interactions, where the participants are obliged to fulfill a specific agenda, differ from those in ordinary conversation. The topics, actions, and the sequence of speakership are predetermined and follow a predictable pattern from the outset in some forms of institutional interaction (Heritage & Clayman, 2010). Heritage and Clayman (2010) summarized the institutional turn-taking system in three groups: turn-type pre-allocation, mediated, and mixed pre-allocated / mediated, as shown in Table 1. The three groups of the institutional turn-taking system fall along a continuum defined by the level of restriction on "when and which persons may speak (the addressee of the question) and the type of contribution they may make (responding to the question)" (Heritage & Clayman, 2010, pp.37-38). The turn-type pre-allocation is on the one end of the continuum, which requires strict restrictions to enable the institutional representative to maintain control over the content and the overall structure of the interaction (Heritage & Clayman, 2010). The mediated turn-taking system is located at the other end and is typical of meetings, where a "chair" is responsible for mediating turns in a large group. The more formal the institutional interaction is, the greater the chair's responsibility to allocate turns and control the topical progression is. Conversely, non-formal institutional interactions do

not involve the restriction of turn types or the pre-allocation of turns (Hauser, 2008), and therefore, more self-selection and next-turn allocation by the current speaker are observed, and more digressions may be permitted (Asmuß & Svennevig, 2009). Ilie (2004) proposed semi-institutional discourse to describe the feature of talk shows, exhibiting a hybrid of institutional characteristics, i.e., a particular participation framework and turn-taking system, along with elements of everyday conversation such as the local management of turn-taking.

**Table 1**

*Types of institutional turn-taking systems (Heritage & Clayman, 2010, p.38)*

Turn-taking system	Location	Broad function
Turn-type pre-allocation	News interviews, courts, formal classroom	Management of talk for overhearers, normally in large-scale settings
Mediated: a “chair” mediates turns as the addressee and initiator of talk by other participants	Meetings	Management of turn-taking traffic in a large group
Mixed pre-allocated/mediated	Specialized counseling techniques, dispute mediation	Management of conflictual or sensitive interactions

## 2.2 Multimodal resources for turn-taking

The field of conversation analysis has first examined syntax and phonetics/prosody in managing local contingencies. Then, the interplay of syntax and prosody has been found to play a role in the given action and sequential context of turn organization in languages with early projection, such as English and German (Couper-Kuhlen & Selting, 2018). However, the fundamental dimension of human action lies in its multimodality (Mondada, 2018). Hayashi (2005) once described turns at talk as “multimodal packages for the production of action (and collaborative action) that make use of a range of different modalities, e.g., grammatical structure, sequential organization, organization of gaze and gesture, spatial-orientational frameworks, etc., in conjunction with each other” (p.22). Since Kendon (1967) explored the relationship between gaze and turn-taking, with the advancement of technology, a growing body of research has been focused on multimodality in interactions, specifically on how co-participants orient to each other’s multimodal conduct and assemble it in meaningful ways, moment by moment (Mondada, 2016).

Ford et al. (1996) and Ford and Thompson (1996) have shown that multimodal resources, including syntactic, intonational, pragmatic, and embodied (gaze and body movement) resources, are usually convergent to project the possible turn completion in English. However, Tanaka (1999) observed that pragmatic resources are more oriented than other resources in Japanese because their syntactic (in)completion largely depends on their terminal boundary. Pragmatic resources are defined as the pragmatic completion points in which nothing is being projected “beyond itself in a way of a longer story, account, or other agenda” (Ford & Thompson, 1996, p.150). Similarly, Li (2014) identified that pragmatic resources could override other resources in turn-taking in Mandarin Chinese.

How multimodality works for multiparty interactions is also gaining attention from the conversation analytic perspective. Markaki and Mondada (2012) investigated how co-participants manage turn-taking in corporate meetings of a multinational company by orienting to both verbal (i.e., mentioning the specific country) and embodied conduct (e.g., gaze, point, body posture, etc.). They found that the current speaker moved their gaze and pointed at the representative of a specific country when mentioning that country verbally, while the other participants also turned their gaze to the designated speaker. Additionally, Mondada (2013a) explored how multimodal resources are utilized to manage interactional orders and allocate turns in French political meetings, where the chairman and participants use embodied conduct (e.g., gesture, gaze, body posture, and walking) in coordination with the specific spatial arrangement to sequence the audiences to the debate. Participants orient toward both the incremental organization of the turn and the position of the chairman within the room to be selected as the next speaker. Mondada (2013a) argued that visible resources, such as pointing, gesturing, turning toward, back, and away, body torque, and gaze, are critical for identifying and pre-selecting future speakers and progressively emerging and making the next candidates visible without disrupting the ongoing talk.

Previous research has demonstrated the use of multimodal resources in both everyday conversations and institutional settings. This study seeks to expand on that research by analyzing the use of multimodal resources in semi-institutional meetings, specifically focusing on how the chair and other participants manage the turn-taking and thereby fulfill their obligations.

### 2.3 Turn-taking in Mandarin Chinese

Scholars have increasingly investigated the use of different interactional resources for turn-taking in Mandarin Chinese. Syntactically, the topic-comment structure can inhibit the co-participants from taking their turn at the juncture (Li, 2014), while syntactic parallelism, such as copula-complement structures, adjectival predicate, conditional constructions, and complex sentences featuring a clause of purpose, may prompt the second speaker to come in early and jointly complete the syntax-in-progress (Gao & Ren, 2021). Additionally, run-on sentences (called *liushuiju* ‘flowing water sentences’) -- a series of phrasal constructions or elliptical clauses sharing a topic -- create fluid and flexible speech units, where prosody plays a significant role in signaling possible turn completion (Li & Ono, 2019, p.4). Additionally, prosodic and phonetic parameters such as pitch register, pitch range, duration, and pause may be used to project possible turn completion (Li, 2014). In particular, Lim (2019) demonstrated that turn continuation in Chinese is organized through the interplay of syntax and prosody cooperatively. Discontinuous constituents after possible completion are marked as “continuing” by the combination of prosodic cues such as rush-throughs, lack of pitch reset, and declining intonation contour. Furthermore, some grammatical structures, including serial verb construction, unexpressed nominal argument structure, topic-comment structure, and verb-resultative complement, are also oriented as a resource for turn-continuation despite not being produced prosodically as continuous.

Non-verbal conduct has also been studied in Mandarin interactions. Hand drops and self-grooming in the face and head area can signal turn-yielding, while gaze and touch are used to claim the floor. Non-gaze, gesture holding, thinking face, and finger count may be used to maintain turns (Yang, 2011; Li, 2014). Tao (2019) investigated the use of gestures in constructing list sequences and found that reiterative gestures, e.g., the intermittent repetition of like strokes, can be used to maintain the conversation floor and serve other discourse functions. Endo (2019) focused on how the gaze shift and the epistemic marker *wo juede* ‘I think/feel’ facilitate (dis)affiliation and

participation. Specifically, the speakers were observed to look away from participants with conflicting opinions and toward those whom they wanted to select as the next speaker.

Wu (1997) demonstrated how marginal roles can actively or passively claim their speakership via *A*-prefaced and *EI*-prefaced turns accompanied by different patterning of body movements in multiparty conversation. She also noted that *EI*-prefaced utterances in an independent intonation contour could claim speakership at a TRP, while they tended to be latched onto the turn component if attempted at a non-TRP, albeit orthographically *ei* in Mandarin is always represented as a stand-alone unit in writing (Wu, 2014).

Working on Mandarin Chinese everyday face-to-face conversation, Li (2014) has been the first to systematically address the individual relevance of syntax, prosody, and body movements (hand and torso) to turn organization. In addition, she also explored the interaction of syntax, prosody, body movements, and pragmatic resources in the projection of possible turn completion. Li argued that the constellation of linguistic and embodied resources made the turn completion recognizable, but they functioned differently. Although linguistic resources are essential, they may (or may not) be overridden by co-occurring body movements. Moreover, pragmatic completion can override linguistic and embodied resources in turn projection when they conflict.

Li (2016) then focused on how visual resources assist in managing the possible completion of syntactically incomplete turns and discovered that verbal and visual resources work together in two ways. One way is that the turn is initiated by an incomplete syntactic structure and completed by bodily-visual conduct. The other way is that the turn is constituted by an incomplete syntactic structure, but the bodily-visual conduct may indicate possible turn completion.

Previous studies have shown that linguistic resources (particle, syntax, prosody), non-verbal conduct (gestures, gaze, facial expression), and pragmatic resources play roles in Mandarin turn management in both dyad and multiparty interactions. However, their data were collected from ordinary conversations. The way multiparty interactions unfold in a semi-institutional setting (Ilie, 2001) remains underexplored in Mandarin Chinese. Therefore, this study aims to investigate how multimodal resources are employed in turn management within an institutional multiparty interaction.

### 3 Data and Methodology

The data for this study is collected from 圆桌派 *Yuan Zhuo Pai* ‘Roundtable π,’ a Mandarin talk show on social issues that has been on the air since 2016 and completed six seasons. The show is aired on Youku, a Chinese movie and video streaming website, and can also be viewed on the Youku official channel on YouTube. The show is selected for two primary reasons. First, the goal of this show is to facilitate impromptu talk without any predetermined scripts, allowing all participants involved to share their views and opinions equally. This aspect was highlighted by the production team (*douban*, n.d.). Second, the face-to-face conversation recordings of the show enable the analysis of embodied movements and other paralinguistic resources.

The show consists of one male host, 窦文涛 *Wentao Dou*, with three guests who sit around a table and discuss their opinions and experiences on a specific topic raised by the host while having tea or snacks. After watching the fourth season of the show, I noticed a combination of characteristics pertaining to how the host mediates the turn with various guests. The guests in this show vary along a continuum based on the frequency of participation. When a new, “special” guest is involved, the talk show is much more like a formal meeting in which the host discharges his responsibility for managing the turn-taking, either by nominating the special guest(s) to respond

or by designing particular questions that can only be answered by the special guest(s) due to their epistemic primacy. On the other hand, when regular guests are involved, the talk show is more akin to an informal meeting, and the host has less control over the conversation. This was also made explicitly by the host in Season 4 Episode 26, where he stated that “the host is not working as a host, and the guests are not like guests.”

This study aims to investigate how multiparty participants (the host and the guests) in a semi-institutional setup manage their interactions through multimodal resources. The data was collected from seven episodes of the show that I described above (Appendix A), which has a total duration of about 350 minutes. All the selected episodes feature three long-time friends – 周轶君 *Yijun Zhou* (female), 马家辉 *Jiahui Ma* (male), and 梁文道 *Wendao Liang* (male) – as participants. They have known each other for many years, as the host mentioned that “staying with friends, like staying with you guys, I really felt very safe” in Season 2, Episode 23. While the settings of different seasons were slightly different, the general setup is depicted in Figure 1. The four participants sit around a table (Fig.1 – A, B, C, D), and there is a screen behind the host's right side (Fig.1 - E)

**Fig. 1**

*The setting of Roundtable π*

A.



B.



C.



D.



E.



After collecting the episodes, I transcribed the verbal conduct using Jeffersonian conventions (2004) with slight modifications to accommodate Mandarin Chinese. This included capturing the final pitch direction at each TCU boundary due to the tonal nature of the language (Li, 2019). The transcriptions were organized into three tiers: the first tier is the original (Mandarin) utterance transcribed using *Pinyin* (the Chinese Phonetic Alphabet); the second tier consists of morpheme-by-morpheme glosses in English; the last tier presents idiomatic English translations of the Mandarin utterances. To present the details of embodied conduct, Burch's (2014) conventions were adopted and modified with a plus sign symbol and aligned above each verbal

utterance to show co-occurrence with the talk. Frame grabs were also used to provide visual context for the analysis. Appendix B provides more detailed explanations of the abbreviations and terms used.

Multimodal conversation analysis and interactional linguistics are employed in this study. Multimodal conversation analysis views the sequentiality, temporality, projectability, and progressivity of action as mobilized through a variety of resources working holistically beyond linguistic materials alone (Mondada, 2014). This perspective allows for an analysis of how social interactions unfold through various resources among participants. Meanwhile, interactional linguistics provides a perspective on language structure and uses in interaction (Li, 2014), providing insight into how linguistic structures shape interactions in this study.

## 4 Analysis

The present analysis examines the use of multimodal resources by the participants, i.e., the host and three guests, in managing their turns and thus fulfilling their obligations on the show. The analysis is divided into two parts, each of which highlights specific instances of multimodal resource use. The first part focuses on the way in which the host employs multimodal resources to allocate turns (Excerpts 1 and 2) and regulate the overall flow of conversation (Excerpt 3). The second part of the analysis explores how the guests utilize multimodal resources to self-select as the next speaker (Excerpts 4, 5, and 6) and how they also engage in selecting each other instead of the host (Excerpt 7).

### 4.1 Host

This section will first present an examination of the host's recurring practices in turn allocation, including (1) the host's creation of a shared interactional space for all participants to self-select through sweeping gaze (Excerpt 1); and (2) the host's nomination of the next speaker through gaze and address term (Excerpt 2). The subsequent part will focus on the host's self-section as the next speaker, which enables him to regulate the topic of discussion (Excerpt 3).

#### 4.1.1 Allocating the turns

Excerpt 1 is taken from Season 4, Episode 23, titled "Misreading," and the seating arrangement is shown in Figure 1A. Prior to the beginning of Excerpt 1, the host (Mr. Dou) expresses frustration about some misunderstandings among his colleagues but asserts that there is no need to specify everything in communication overly. However, as the excerpt begins, he finishes his complaint that it may not work sometimes.

#### Excerpt 1: "Reading air" (S4E23)

```

dou      -->gz>zhou(>line 02))
01 DOU   sihu ni jiu ganjue (.)
         seem you then feel

         dou          +slightly shakes head
02      ni yao bu +shuo mingbai le (0.6)
         you if NEG speak clear LE

         dou      +gz>ma +gz>zhou          +gz>ma
```

dou +slightly shakes head  
 03 +hai +zhen: ting bu dong +ei;  
 still true listen NEG understand PRT  
**'It seems that you feel like if you don't say it clearly, (they) really won't understand.'**

dou +gz>sweeps from liang to ma slowly ((>line 06))  
 zhou +gz>liang ((>line 07))  
 04 +(0.6)

05 ZHOU dui  
 right  
**'Yeah.'**

06 (0.2)

dou +gz>zhou((>line 07))  
 zhou +gz>ma +gz>liang  
 07 jiu riben +ren +jingchang +shuo  
 just Japan people usually speak

zhou +gz>dou((>line 10))  
 zhou +raise eyebrows  
 08 du +kongqi: ma  
 read air PRT  
**'It's just like what Japanese people often say 'read the air,' you know.'**

From line 01 to line 03, there is a complicated syntactic structure in which a conditional structure is embedded within an object clause. Initially, Dou produces the principal clause that *sihu ni jiu ganjue* 'it seems that you feel' (line 01) to project more to come. He then utters the content of the "feeling" in a conditional structure, describing the condition that *ni yao bu shuo mingbai le* 'if you don't say something clearly' (line 02) and then formulating the contingent on the condition that *hai zhen ting bu mingbai* '(they) cannot understand' (line 03). Dou emphasizes the negation in both of these clauses by slightly shaking his head twice while speaking. Despite a micropause at the end of line 01 and a noticeable silence at the end of line 02, the incomplete syntactic structure helps the current speaker hold his floor and other participants cannot take the floor during those moments.

The first possible completion point occurs at the end of line 03, where Dou completes the syntactic structure. Prosodically, he drops the pitch of the final syllable to indicate turn-yielding (Duncan, 1972), signaling a potential speakership transfer. Dou could have continued his speakership after the silence (line 04) because no one tried to take the floor at that moment. However, he has designed his turn as completed at this point, and nothing further to be projected in terms of his complaining action. Meanwhile, Dou does not fix his gaze on a specific guest at the TRP but instead sweeps his gaze over all the other participants, creating a shared interactional space (Mondada, 2013b) that enables all guests to equally self-select as the next speaker.

Notably, during the silence (line 04), Zhou gazes at Liang, indicating that she orients herself to the turn-yielding signals and other participants' responses before taking the floor. Zhou then produces *dui* 'right,' to agree with Dou's preceding opinion. The other two participants do not take turns during the gap of silence in line 06. Zhou then self-selects to echo Dou's previous complaint



by introducing the Japanese phrase *du kongqi* ‘read the air,’ and ends the current TCU with the particle *ma* to indicate that it may be a familiar concept to others (Lv, 1999).

One of the primary responsibilities of the host in this show is to facilitate the discussion. To this end, the host recurrently employs a sweeping gaze to direct attention toward all guests, thereby creating a shared interactional space in which all participants are equally able to self-select themselves. Correspondingly, the guests orient themselves towards the cue of the sweeping gaze as a signal for self-selection, enabling them to fulfill their obligations as guests in contributing to the ongoing topic.

However, the host’s facilitation of turn-taking is not limited to the creation of a shared interactional space. In some instances, the host will also nominate the next speaker, as illustrated in Excerpt 2. Specifically, we turn to an excerpt from Season 4 Episode 26, “Family Affection: Our Parents,” in which two participants (Dou and Liang) sit facing each other (Figure 1B). Prior to the excerpt, Ma has been arguing that there is not much that parents can do to influence their children in the age of the internet, while Dou expresses some disagreement. Seeking to persuade others of his viewpoint, Dou poses a polar question, and we join the excerpt at this juncture.

### Excerpt 2: It’s nothing to do with my father (S4E26)

- dou -->+gz>liang(>line 03))
- dou -->+#fig2.1 +moves LH>chest, makes a fist(>line 03))
- 01 DOU +nimen +dao zhe ge suishu  
you(plural) to this CL age
- 02 you mei you: ganjuedao (0.4)  
have NEG have feel
- dou +gz>right mid-air +gz>liang
- dou +turns head slightly right +LH extends 5F
- 03 jiushi +(0.6) +e +ni:: (0.2)  
just COP eh you
- dou +gz>slightly left +gz>liang(>line 09))
- dou +LH moves up & down twice
- 04 +yuelai +zai ni shen shang  
more on you body on
- dou +pinches 5F (#fig2.2)
- 05 +yue faxian ni fuqin  
more find you father
- ‘Have you ever felt or not that when you are at this age, you find that you are more alike to your father?’**
- 06 (1.0)
- dou +lean back, LH palm facing right > liang
- 07 +we- wendao ni juede ni  
wendao you think you
- dou +withdraws LH
- 08 +de xingcheng  
DE form
- dou +LH IF points upward +withdraws LH, makes a fist
- 09 gen +ni de fuqin +you shenme guanxi=



#fig2.1



#fig2.2

and you DE father have what relationship  
**‘Wendao, do you think your development has anything to do with your father?’**

liang +gz>mid-air((thinking face)) +gz>dou  
 liang +slightly shakes head, moves RH away from mouth  
 10 LIANG =+mei you +shenme guanxi ba  
 NEG have what relationship PRT  
**‘(I guess) it has nothing to do with my father.’**

11 DOU hhhhh

The host employs an A-not-A structure (*you mei you* ‘have-NEG-have,’ line 02) to formulate a polar question (lines 01 to 05) to further address his opinion. He initiates the question specifying the principal clause regarding whether the guests, who are all considered to be middle-aged, have a certain feeling (*nimen dao zhe ge suishu you mei you ganjue dao*, line 01 to line 02) while simultaneously gesturing with his left hand by making a fist in front of his chest to embody said “feeling.”

Although a 0.4-second pause occurs at the end of line 02, the incomplete syntactic structure assists Dou in retaining his turn, and he continues. However, he does not immediately articulate the primary element of his question, suggesting that he may be experiencing difficulty constructing the sentence, as evidenced in line 03. Initially, he utters *jiushi* ‘just be,’ a pause filler to hold his turn (Biq, 2001) and displays his trouble verbalizing the specific feeling (Zhang & Gao, 2012). Subsequently, during a 0.6-second pause, Dou withdraws his gaze from the addressed recipient, Liang, by turning his head slightly to the right, which may indicate that he is engaged in a “solitary word search” (Goodwin & Goodwin, 1986). Dou returns his gaze to Liang when he utters another filler word *e* ‘eh,’ and then prolongs the pronoun *ni* ‘you,’ followed by another micropause, both of which suggest that he is still in the process of searching for the right words and patterns, as he withdraws his gaze again at the beginning of line 04. Finally, Dou transitions to the primary aspect of his question, utilizing a double comparative structure *yue...yue...* ‘the more..., the more...’ (lines 04 and 05). Notably, this sentence contains grammatical errors, though understandable, suggesting that Dou has not yet succeeded in finding the appropriate phrasing.

It is noteworthy that the host employs a plural pronoun *nimen* ‘you (plural)’ (line 01) to address all three participants, thus indicating them as the intended recipients of the question, which verbally creates a shared interactional space. However, as the question progresses, Dou switches to the singular pronoun *ni* ‘you’ (line 03 and line 04) and maintains his gaze on Liang, except for brief moments when he is searching for words, thereby designating Liang as the next speaker (Kalma, 1992). Despite this, Liang seems not to orient to those turn-yielding cues, and consequently, he does not take the floor right away, as a one-second pause ensues after the question. This could be attributed to the fact that Liang was not looking at Dou at the time, therefore, may have missed the contextual cues for taking turns. Additionally, the grammatically incorrect sentence may have led other participants to wait for a self-repair before taking the turn themselves.

Failing to address the recipient via gaze, Dou upgrades his selection by nominating the addressed recipient’s name, Wendao, and pointing with his left hand (line 07). He then reformulates his question in a wh-format (*shenme*, ‘what,’ line 09). Liang immediately latches onto the question and provides his answer, though slightly dispreferred, stating that he does not believe his development has anything to do with his father while shaking his head simultaneously (line 10) in a *ba*-marked turn to downgrade his epistemic position (Kendrick, 2018). This latching behavior

demonstrates that Liang, as the addressed recipient, is orienting to the previous misarticulated question by the host.

In summary, this excerpt shows that the host attempts to create a shared interactional space by using plural pronouns, but also discharges his ability to allocate turns via gaze direction (Sacks et al., 1974). However, if this strategy fails due to the absence of mutual gaze, the host resorts to an upgraded measure, e.g., the use of address terms paired with gaze and gestures to select the next speaker.

#### 4.1.2 Controlling the discussion

In addition to allocating turns, the host is seen exercising his role in controlling the topic of discussion. The following excerpt is from Season 4, Episode 28, “Living Alone,” with the seat arrangement shown in Figure 1C. The discussion centered on the reasons for the growing trend of young people living alone. However, before this excerpt, they went off-topic and discussed the differences between “loneliness” and “solitariness.” Zhou then took the floor and presented her argument that the primary reason for young people living alone was their departure from their hometown. The excerpt begins at the end of Zhou’s statement.

##### Excerpt 3: Thank you for helping me drag the topic back (S4E28)

```

zhou      +gz>ma      +gz>dou((>line 03))
zhou      +moves BH SPV near to chest (#fig3.1)
dou      -->gz>zhou((>line 14))
dou      -->head rests on LH
01 ZHOU   .h +ta mei you +jiaren
          3SG NEG have family members

          zhou +puts BH down +nods
02      +zhu zaiyiqi +a
          live together PRT
          ‘They do not have family members to live with.’

          dou +raises RH SPU>zhou (#fig3.2)
03 DOU   .h +>ei<=
          PRT

          zhou +gz>liang
          zhou +raises LH SPU and puts it down;RH holds face
04 ZHOU  =ta shi +bei piao ye hao
          3SG COP Beijing drift also good

          zhou +gz>dou ((>line 08))
          zhou +raises RH SPU and puts it down
05      +shenme(.) [hu piao ye hao
          what Shanghai drift also good
          ‘Either they are Beijing-drifters or Shanghai-drifters,’

          +moves LH up and down
06 DOU   [+>dui dui dui dui dui<
          right right right right right
          ‘Yeah, yeah, yeah, yeah, yeah,’

```



zhou +widens eyes, extends BA(#fig3.3) ((>line 09))  
 07 ZHOU [+ta qishi shi [MEI YOU  
 3SG actually COP NEG have  
 ‘**Actually, he/she doesn’t have.**’



#fig3.3

dou +withdraws LH  
 08 DOU [>ni +ni ni< [>dui dui dui<  
 you you you right right right  
 ‘**You, you, you, yeah, yeah, yeah.**’

zhou +gz>liang  
 zhou +withdraws BA, LH under chin  
 09 ZHOU +ta [zhouwei +mei you +ren  
 3SG around NEG have people  
 ‘**There are no people around them.**’

10 LIANG [dui  
 right  
 ‘**Yeah!**’

dou +LH points>zhou ((>line 12))  
 11 DOU +ni ni ni ni zhe tebie hao  
 you you you you this very good  
 ‘**You, you, you, you. This is so good.**’

dou +gz>his RH IF((>line 14))  
 dou +#fig3.4a +#fig3.4b  
 12 +ni bang wo +ba zhe ge [+huati  
 you help I BA this CL topic



#fig.3.4

13 ZHOU [+hahahaha

dou +#fig3.4c  
 14 DOU you +zhuai huilai le  
 again drag back LE  
 ‘**You helped me drag back the topic.**’

Zhou highlights that young people are living alone in that they do not have family members to live with, which is embodied by her gesture of both hands moving towards her chest to signify “living together” (lines 01 and 02). This TCU ends with particle *a*, indicating that Zhou assumes this situation should have been known among the recipients (Wu, 2000), but they have ignored it and analyzed more abstract reasons. Throughout her turn, Zhou maintains her gaze on Dou, starting from the middle of her utterance in line 01 until line 03, and withdraws her hands (line 02), projecting a possible completion of the current turn. Dou, interpreting the gaze as a selection of the next speaker (Sacks et al., 1974), claims his speakership with *ei* (Wu, 2014) in line 03 and palms up his left hand towards Zhou, as shown in Fig 3.2. However, Zhou herself does not design it as a complete action. She might realize that other multimodal resources have projected the possible completion point. Hence, she immediately latches onto the syllable of Dou and redirects her gaze to another participant to secure her floor (line 04).

Zhou provides two specific examples, namely Beijing drifters (line 04) and Shanghai drifters (line 05), to illustrate individuals who live apart from their family members. She employs her left and right hands to visually represent the two distinct groups, moving them up and down once, respectively. Meanwhile, Dou maintains his left-hand gesture from line 03 and accentuates

it by employing upward and downward movements in an effort to assert his speakership (line 06). Simultaneously, Dou overlaps with Zhou's second example by repeatedly uttering the word *dui* 'yeah' at an accelerated pace, indicating his understanding of the gist of her message (Jefferson, 1984; Yeh & Huang, 2016). This vocalization, along with his accompanying gestures, suggests that he desires to take control of the conversation as the host.

Despite Dou's attempt to interject, Zhou persists in her turn, continuing to elaborate on her argument that young individuals lack the company of their family members, with her widened eyes and heightened vocal volume (line 07). Zhou's syntactic utterance is incomplete as the object of the transitive verb *meiyou* 'no have' is omitted. Her turn is likely interrupted by Dou's speedy repetition of the syllables *ni* 'you' and *dui* 'yeah' (line 08) as he seeks to claim the floor. Interestingly, Zhou supplements her incomplete utterance with a visual demonstration (Li, 2019). As previously observed, she moves both her hands towards her chest to represent "living with family members together" (line 01), in which her two hands signify the boundary of the family. Employing it as a substrate (Goodwin, 2018), Zhou extends her arms (line 07) to indicate that no family members are present.

Zhou's continuation on her turn prompts Dou to withdraw his left hand (line 08). Zhou, on the other hand, turns her gaze away from Dou and directs it to Liang as she proceeds to reformulate the previously abandoned TCU and completes it by stating that "there are no people around them" (line 09). Liang, the addressed recipient at this moment, acknowledges her opinion by *dui* 'yeah' (line 10). Zhou then withdraws her arms and puts her left hand under her chin, indicating her readiness to listen to the next speaker and projecting the possible completion point at the end of line 09. After several attempts to claim his speakership, Dou finally takes his turn. He begins by repeating the pronoun *ni* 'you' and positively assessing Zhou's previous talk (line 11). He further specifies that the positive assessment is made on her talk's relevance of the topic rather than her opinions by saying that *ni bang wo ba zhe ge huati you zhuai huilai le* 'you helped me drag back the topic' (line 12 and line 13). In the meantime, he uses consecutive gestures to connect Zhou's analysis with the original topic, as shown in Fig 3.4. Using his right index finger, he initially points at Zhou and then traces a curve while moving his finger down, as if he is connecting something on Zhou's side to his own side.

In contrast to the previous two excerpts, which mainly focus on how the host allocates the turn, this excerpt demonstrates how the host fulfills his other responsibility of controlling the track of the topic through self-selection via linguistic (prosody, repetition) and embodied resources (gaze, gestures). In addition, this excerpt also shows that pragmatic resources can override other multimodal resources in turn-yielding (Li, 2014), as Zhou maintains her speakership in lines 04 and 07. Though she completes a sentence syntactically, withdraws her gestures, and fixes her gaze on a particular recipient, she still does not design the turn-so-far as complete because she has more new information to project in terms of the pragmatic resource.

## 4.2 Guests

The preceding section illustrated the ways in which the host employs multimodal resources to manage turn-taking and regulate the topic. In this section, I will examine how the guests utilize such resources in managing their turns and thereby fulfill their responsibilities. First, I will discuss how the guests use coordinated verbal and embodied conduct (Excerpts 4 and 5), followed by an analysis of consecutive embodied and verbal conduct (Excerpt 6) to self-select themselves as the

next speaker in a timely manner. Finally, I will discuss how the guests select one another to speak rather than return the floor to the host (Excerpt 7).

#### 4.2.1 Self-selecting

The forthcoming excerpt is from Season 4 Episode 4, “996 Working Overtime,” with the seating arrangement illustrated in Figure 1D. Preceding the excerpt, Dou discusses the striving efforts of certain individuals to become billionaires and how they may feel perplexed if their descendants do not exhibit the same level of dedication as they did in the past.

##### Excerpt 4: Father (S4E4)

dou +gz>liang +gz>zhou  
dou +turns head to left  
dou +higher RH(>line 02))  
dou +leans forward  
01 DOU +zhe +jiu xiang youxie fuqin +(0.4)  
this just like some father

dou +gz>liang>ma +gz>between ma & liang  
dou +turns head to right  
dou +pinches RH 5F slightly  
dou +withdraw RH  
02 hen +nan +e:: lijie +zhe ge haizi  
very difficult eh:: understand this CL child

dou +gz>liang(>line 04))  
dou +frowns, tilts head, leans back slightly  
03 .h shuo +ni zenme xue +yishu  
speak you why study art

04 (0.5)

dou +gz>between liang & zhou  
05 +shi ba  
COP PRT  
**‘It is just like some fathers find it difficult to understand their children, saying that why you study art, right?’**

dou +gz>zhou  
dou +turns head to left  
06 +na wanyir jianglai +neng dang fan chi ma  
that thing future can serve as food eat Q

dou +turns his head from left to right  
07 +(0.3)

dou +gz>ma  
ma -->gz>dou  
08 +dui ba=  
right PRT  
**‘(Those fathers) saying that you cannot make a living with that thing, right?’**

ma +gz>lower left

- ma +lowers head, tilt  
 ma +extends LH IF> mid-air  
 09 MA =na +limian keneng  
 that inside probably
- ma +gz>dou  
 ma +LH IF moves downward  
 ma +withdraws LH IF  
 10 +wentao you ge +xing+bie wenti=  
 Wentao have CL gender problem  
 'It might have to do with the gender issue, Wentao.'
- ma +gz>lower left  
 ma +widens eyes  
 11 =>+yinwei< ni +gang ju li shuo fuqin ha  
 because you just give example speak father PRT  
 'Because you just mentioned father in your example.'
- ((word-searching sequence is deleted))
- ma +gz>zhou  
 ma +raises eyebrows, widens eyes  
 ma +extends RH IF  
 ma +higher RHIF  
 ma +withdraws RHIF  
 18 wo +jiu+shi na+zhong fuqin  
 I just COP that type father  
 'I am exactly that kind of father.'
- ma +gz>table +gz>zhou  
 ma +frowns, shakes head  
 ma +lowers head  
 ma +BH facing each other as a "praying" gesture  
 19 wo +yi ting wo nver +shuo zhao gongzuo  
 I as soon as listen I daughter speak find job  
 'As soon as I heard that my daughter is finding a job, ...'

The initial portion of the excerpt features a multiunit turn by Dou, where he describes and enacts the responses of certain fathers toward their children's career choices (lines 01 to 06). Dou first states that those fathers find it challenging to comprehend their children's decisions (line 02), following a 0.4-second pause (line 01). At this juncture, Dou averts his gaze, i.e., gazing somewhere between Ma and Liang, which indicates his intention to continue his turn (Kendon, 1967). He then enacts the fathers' responses, preceded by *shuo* 'to say, to speak' in questioning the utility of studying art, using the mirative term *zenme* 'why' in a rhetorical question to form a negative meaning (Liu, 2016) and thus displaying a disaffiliative stance, coordinating with frowning, tilting his head, and slightly leaning back (line 03). After a 0.5-second pause, Dou uses *shi ba* 'COP PRT' (line 04) to elicit agreement from his audience (Li & Thompson, 1989, p. 307). Yet, he averts his gaze again, i.e., gazing somewhere between Liang and Zhou (line 05), suggesting that his turn is still ongoing and preventing others from taking the floor (Yang, 2011; Li, 2014). Dou then makes a second enactment by posing another question that employs the derogatory term *wanyir* 'thing' in a rhetorical question to further illustrate those fathers' disaffiliation that they do not think their children can make a living with art. During a brief pause, he turns his head from

left to right (line 07) and then solicits approval again with *dui ba* ‘right PRT’ (line 08). Notably, Ma seizes the opportunity to self-select as the next speaker as soon as Dou turns his gaze in his direction. Similar to Excerpt 1, it is clear to see that Dou, the host, sweeps his gaze while talking to create the shared interactional space to discharge his responsibility to make the participants equally able to select themselves as the next speaker.

Ma designs his turn to latch onto Dou’s last syllable by suggesting that there may be a gender issue (lines 09 and 10) while also inserting the addressed recipient’s first name, Wentao, into his response (line 10). He tilts his head down to the left and points his left index finger at mid-air to actively secure his turn. He withdraws his left index finger at the end of his current TCU, possibly indicating turn-yielding (Yang, 2011). However, he rushes into the next TCU with the conjunction *yinwei* ‘because’ to explain why he has self-selected as the next speaker in that Dou has just (*gang*, just now) mentioned “father” in the example (line 11). Despite searching for words omitted in this excerpt, Ma further elaborates on his reason for self-selection, revealing that he is exactly the kind of father Dou has mentioned. Although both his verbal (i.e., syntactically completed sentence) and embodied conduct (i.e., gesture withdrawal, fixed gaze) project the end of the current turn, the pragmatic resources suggest that he is expected to provide additional information. As a result, he proceeds to share a story between his daughter and himself (line 19).

This excerpt illustrates the guest’s ability to orient to the ongoing talk and visual cues to timely self-select. By justifying his reason for taking the floor as a father, he establishes epistemic primacy within the interaction. Furthermore, he also employs embodied resources such as averted gaze and pointing gestures, coordinating with his verbal conduct, to claim and secure his speakership. Through his gradual explanation, he makes his contribution relevant to the topic.

The following excerpt (taken from the same episode as Excerpt 3) is going to display how the guest takes advantage of multimodal resources to claim speakership a little bit earlier in the current turn. Prior to this excerpt, Zhou and Dou discussed the observation that many young females prefer traveling alone, prompting Ma to recount his own experience of going to Disneyland Park as a teenager. He emphasizes his miserable experience due to his inability to share his feelings with anyone, particularly in a tour group where everyone else had a family with them, resulting in him feeling lonely. The excerpt begins at the end of Ma’s story.

### Excerpt 5: Travel Alone (S4E28)

ma -->gz>front left on the table  
 ma +slightly shakes head +RH PRV (#fig5.1)((>line 02))  
 01 MA wo +jiu: jianjue .h +bu keneng  
     I then resolute NEG possibility  
  
 ma +gz>dou  
 ma +withdraws RH  
 02 +yi ge ren qu:: you: .h you+leyuan=  
     one CL person go amuse amusement park  
     ‘I am, then, absolutely resolute that it is impossible for me to go to any amusement park alone.’  
  
 ma +gz>zhou ((>line 04))  
 ma +RH PRV +slightly shakes head  
 03 =+lian +yi ge ren +lvxing wo dou  
     even one CL person travel I all  
  
 ma +RH PRV shakes slightly +withdraws RH  
 04 +jinliang +bi: [kai  
     try my best avoid



#fig5.1



**'I even try my best to avoid traveling alone as much as possible.'**

liang -->gz>ma((>line 07))  
 liang +leans forward  
 05 LIANG [+dan  
 but

liang +extends RH IF>ma +withdraws RH  
 06 +ni shuo de zhe +ge  
 you speak DE this CL

liang +leans backward  
 07 +qingkuang wo juede b- tebie weishenme;  
 situation I think b- special why  
**'But I think the situation that you mentioned is special. Why is that?'**

Ma concludes his story by declaring that he will never visit an amusement park alone again, using the word *jianjue* 'resolute' with an accent to emphasize his firm attitude. He also shakes his head a little bit preceding the verbal conduct *bu keneng* 'impossible,' and uses a "stop" gesture towards Dou to emphasize the impossibility of the situation (line 01). Ma then generalizes his experience, searching for the appropriate term *youleyuan* 'amusement park' through the prolongation of *qu* 'go' and *you* 'amuse' to refer to amusement parks in general (line 02). Ma designs his turn as completed by withdrawing his right hand (Yang, 2011) and looking towards Dou as the next speaker (Sacks et al., 1974). However, Ma may feel that his generalization is not relevant enough to the topic of traveling alone. So, he quickly rushes into the next turn by extending his intonation contour. He also retrieves his previous "stop" gesture and turns his gaze to another recipient, Zhou (Yang, 2011; Kendon, 1967), to project that he has more to say and prevent others from taking the floor.

Ma continues to generalize his experience by upgrading his statement from avoiding visiting amusement parks alone to avoiding traveling alone, which is more relevant to the ongoing topic (lines 03 and 04). He emphasizes his determination by using the "even" structure with *lian ... dou ...* and displays his disaffiliation towards traveling alone by shaking his head and making the "stop" gesture again, even shaking it several times to emphasize his point. As he approaches the end of his statement, he withdraws his right hand to project the end of this turn (Yang, 2011). Pragmatically, the action is also possibly complete at this moment, as he has made a relevant generalization to the topic. Throughout his final resolution, he looks at Zhou (lines 03 and 04), which could indicate that he has selected Zhou as the next speaker through gaze (Sacks et al., 1974).

However, Liang self-selects as the next speaker by initiating a terminal overlap (line 05), where the first syllable *dan* 'but' overlaps with the last syllable of Ma's previous turn (line 04). *Bikai* 'avoid' is a verb-complement structure word, with the major meaning on the first syllable *bi*, which Ma elongates. Though the previous utterance has not been completed, what is said within and through it has been perfectly available to the recipient. Liang likely initiates a terminal overlap due to the gaze, as Ma does not intend to select him as the next speaker when he is projecting the TRP by withdrawing his hand (line 04). Liang leans forward as he begins speaking, co-claiming his turn. *Dan* 'but' projects a dispreferred response coming up (line 05). He first employs anaphora *zhe* 'this' to connect his following talk to Ma's previous story. *Zhe* 'this,' as a demonstrative pronoun in Mandarin, is often used to refer to the past speech uttered by any participants in a continuous conversation (Han & Shao, 2007). Therefore, *ni shuo de zhe ge qingkuang* 'the

situation that you mentioned,' in the lexical sense, refers to the situation of visiting an amusement park alone. He then frames his further assessment as a personal opinion using *wo juede* 'I think' that what Ma mentioned is just a particular case and then asks a question not for soliciting answers from others but for securing his floor for further explanations.

The preceding excerpts demonstrate the guests' ability to strategically employ multimodal resources such as syntax, prosody, gesture, gaze, and torso movement to effectively self-select as the next speaker within ongoing interactions. These resources are utilized holistically to navigate the contingencies of the situation. The following excerpt, taken from the same episode as Excerpt 1, illustrates another instance where a guest employs these multimodal resources linearly to self-select as the next speaker. Prior to this excerpt, Liang, a philosopher, introduces the works of cultural critics on misreading and misunderstanding to the other three participants, concluding that these phenomena are common occurrences in human communication. In this excerpt, we join the conversation at the end of Liang's talk.

### Excerpt 6: I am going to report to you good news (S4E23)

```

liang -->gz>zhou +gz>dou
liang -->BH SPV ((>line 08))
01 LIANG mou ge chengdu shang+mian
      some CL extent on

      liang +gz>zhou
      liang +BHs up and down ((>line 04))
02 +wudu huo +wujie
      misread or      misunderstand

      liang +gz>between dou&zhou +gz>dou
03 sh- +qishi shi +yi ge zhengchang de
      actually COP      one CL normal DE

      liang +gz>zhou +gz>dou
      liang +lowers BH
      zhou -->gz>liang((line 11))
04 .h +renlei +goutong xianxiang
      human communication phenomenon
      'To some extent, misreading or misunderstanding is actually a normal phenomenon of human communication.'

      liang +gz>zhou
      zhou +RH moves away from chin, extends RH IF +nods, withdraws RH
05 +(0.6) + (0.3)

06 ZHOU dui=
      right
      'yeah'

      liang +gz>dou
      liang +higher BH SPV
07 LIANG =+>erqie< shi
      moreover COP

      liang +gz>zhou ((>line 15))
      liang +withdraws LH, puts RH under chin, leans back ((>line 13))
08 bixu +de=

```

		necessary DE	
		<b>'And it is necessary.'</b>	
09	ZHOU	=shi de COP DE	
		<b>'(You are) right.'</b>	
10	DOU	°en° uhm	
		<b>'uhm'</b>	
	zhou	+moves RH away from chin, extends Fs	
11	ZHOU	+shuodao zhe ge speaking of this CL	
	zhou	+gz>dou	
	zhou	+RH PRD +extends BA	
12		+wo yao baogao +zhu wei I want report all CL	
	dou	+gz>zhou(>line 15))	
	ma	+gz>zhou(>line 15))	
	zhou	+gz>table	
	zhou	+withdraw BA, LH on table, RH under chin	
13		yi ge <u>da:</u> hao +xiao[xi] one CL big good news	
		<b>'Speaking of this, I want to report to you all a piece of good great news.'</b>	
	liang	+lean forwards	
14	LIANG	[+o[::, oh::,	
	dou	+lean forwards	
15	DOU	[+a:, ah:,	

This excerpt begins with Liang concluding that misreading or misunderstanding is normal in human communication, which is a consensus among philosophers and cultural critics (lines 01-04). During his speech, Liang frequently switches his gaze between Dou on his right and Zhou across from him. He also keeps both hands facing each other and moves them up and down in unison. Liang then lowers his hands slightly (line 04), which can be interpreted as a sign of yielding the turn (Li, 2014). Zhou orients to these multimodal resources and bodily claims her speakership during the 0.9-sec pause ensuing (line 05). She first moves her right hand away from her chin and extends her right index finger towards Liang to claim her speakership, followed by head nods to show her agreement or acknowledgment of Liang's conclusion, but simultaneously withdraws her hand. She then produces an agreement token to verbally acknowledge it (line 06). Meanwhile, Liang maintains his gaze on Zhou since the lapse (line 05), so he is likely to orient to Zhou's embodied claim of her turn; thereby, he employs a series of multimodal resources to secure his turn: he initiates *ergie* 'moreover' – prefaced turn to project more to come, latches onto the agreement token, speeds up the first word, switches his gaze to another participant, and raises his gesture again to reclaim his speakership (line 07). All those multimodal resources work holistically to project the non-completion of his current turn. Finally, Liang designs his turn as completed by withdrawing his left hand, putting his right hand under his chin at the last syllable (line 08), and leaning back to his home position. At the end of the current TCU, Liang gazes back to Zhou to

select her as the next speaker as she has self-selected herself in the previous sequence. Zhou displays her orientation to this and acknowledges Liang's upgraded assessment by uttering another agreement token *shi de* 'right.'

After Dou's *sotto voce* agreement token *en* 'uhm,' Zhou initiates a new TCU of reporting. To do this, she retrieves her previous embodied actions by moving her right hand away from her chin and extending all her fingers (lines 11 to 12). Zhou also uses the universal quantifier *zhuwei* 'everyone' to ratify all the participants as addressed recipients of her news, which piques the interest of the others. Liang and Dou align themselves as recipients through the elongated *ou* 'oh' and *a* 'ah' in a rising intonation, assembling their body movements (leaning forward) in anticipation of the news Zhou is about to report (lines 14 and 15).

This excerpt shows us that the guest gradually self-selects as the next speaker via multimodal resources, from embodied conduct to verbal conduct consecutively. In summary, Excerpts 5 to 7 demonstrate that the guests self-select themselves as the next speaker and claim their speakership via a different constellation of multimodal resources, through which they fulfill their obligations as a guest to make a contribution to the discussion in the show.

#### 4.2.2 Selecting other guests

The final excerpt is from the same episode of Excerpt 2. Before this excerpt, Zhou shared her personal experience of how her parents have dealt with her in the past and how she deals with them now. Zhou's parents were tolerant of any choices she made, and now she is trying to bear with their actions when they are childish. She acknowledges that she cannot discipline her parents and must show patience and understanding instead, and we join the excerpt at this juncture.

##### Excerpt 7: to bear with my parents (S4E26)

```

zhou  -->gz>ma +gz>dou
zhou                                     +RH IF point to herself +retract IF
01 ZHOU  suoyi  +xianzai shi +lundao wo:          +frenzhef
        so    now    COP  turn  I                    tolerate
        'So now it's my turn to bear with them.'

        zhou      +gz>ma
        zhou      +raise RA, RH>ma
02          .h +hhh +wo bu zhidao nimen
                I  NEG know  you(plural)

03          you mei you zhe yang [jingya-
        have NEG have this type experien-
        'hhh, I don't know whether or not you guys have this similar experien-?'

        ma          +gz>table>(>line 06))
        ma          +tilts head
04 MA          [>+youshihou<
                sometimes

05          ni shi (0.2)
        you COP
        'Sometimes you are...'

06          you: (0.3) yi ge xinlixuejia shuo ha (0.2)
        have      one CL psychologist speak PRT

```

**‘One psychologist has mentioned that...’**

This excerpt starts with Zhou offering her solution for dealing with her parents: to bear with them (line 01). She emphasizes that it is her turn to do so by pointing to herself with her right index finger and using an accented intonation on the word *wo* ‘I.’ When she mentions the idea of tolerating her parents, she delivers the word *renzhe* ‘to tolerate’ with a smiling voice and a burst of hearable laughter in the subsequent line (line 02) to convey a sense of the ridiculousness of the situation. She then asks the other participants if they share the same experience (lines 02 and 03), using *nimen* ‘you (plural)’ to ratify all the other participants as addressed recipients and make it possible for everyone to self-select when the question is completely formulated.

However, she maintains her gaze on Ma and raises her right arm, moving her right hand toward him to select him as the next speaker till the end of line 03. Ma demonstrates his orientation to her selection by responding immediately, even overlapping with the last two syllables of Zhou’s previous turn (line 04). As soon as he takes his turn to speak, he directs his gaze to the table and tilts his head. During the embodied conduct, he speeds up the time frame by saying *youshihou* ‘sometimes,’ then continues producing an incomplete sentence. The incomplete syntactic structure prevents others from taking the floor during the ensuing mini-pause. However, Ma abandons this structure for some reason and produces another syntactic structure in introducing a quotation in the subsequent TCU (line 06).

In summary, this excerpt demonstrates the guests’ proactive role in allocating turns in the show. Rather than relying on the host to allocate turns, they utilize multimodal resources to select the next speaker among themselves, which highlights their agency and involvement in this institutional setting. Furthermore, all participants, including the host and other guests, treat it as a recurring circumstance by seamlessly taking the floor when appropriate.

## 5 Discussion and Conclusion

By applying the framework of multimodal conversation analysis and interactional linguistics to an impromptu talk show, this study examines the multimodal resources in the organization of turn-taking in semi-institutional Mandarin multiparty interactions. Specifically, it focuses on how the host and guests locally manage their turns through linguistic resources (e.g., syntax, prosody), embodied resources (gaze, gestures, torso movements), and pragmatic resources to fulfill their respective roles within an institutional framework.

The first three excerpts demonstrate the host’s turn-allocation management via verbal and embodied conduct. The recurring practice in this show is that the host creates a shared interactional space (Mondada, 2013b) through sweeping gaze to enable guest self-selection as the next speaker (Excerpt 1). Sometimes the host can allocate the turn through fixed gaze at the TRP and can upgrade the resources into a more salient one, such as address term if the gaze selection fails due to the lack of mutual gaze (Excerpt 2). Additionally, the host can claim his floor via repetition, gazes, and gestures to control the topic track (Excerpt 3). By allocating turns and claiming the floor timely, the host fulfills his responsibilities in creating an equal discussion space and steering the topic.

The subsequent four excerpts showcase the guests’ use of multimodal resources to select themselves or others as the next speaker and contribute to ongoing discussions. Guests can timely self-select themselves as the next speaker due to their epistemic primacy as orienting to the ongoing talk and visual cues. They then secure their floor through holding gestures and averted gazes (Excerpt 4). They can initiate a terminal overlap by coordinating extending gestures and leaning forward to claim the floor verbally and embodiedly, as the meaning of the previous speaker

is clear (Excerpt 5). The next speaker can signal their eagerness to claim the floor through gestures, nods, and then verbal conduct to gradually self-select. The current speaker orients to these resources to fix their gaze on the speaker claiming the floor at TRP, facilitating speakership transfer (Excerpt 6). Furthermore, guests can proactively allocate turns to anyone in the show instead of returning the floor to the host (Excerpt 7). Through their active involvement of self-selecting and taking initiatives in allocating turns, guests demonstrate their agency and participation in the show.

Furthermore, multimodal resources do not always converge at the possible completion point, but the participants can locally manage the contingencies through the use of different multimodal resources (Li, 2014). First, visible cues, i.e., (mutual) gaze and gestures, are critical in identifying and pre-selecting future speakers (Mondada, 2013a). This can occur when the current speaker either opens the floor to all participants (Excerpt 1) or selects the next speaker (Excerpt 2) or when the next speaker pre-claims the floor (Excerpt 6). Second, prosodic resources (latching) and embodied action, i.e., averted gaze (Excerpts 3 and 4), and intensified gestures/facial expressions (Excerpts 3 and 6), which may be employed by either the same or the next speaker, can help them promptly secure their turns when they start a new TCU (Excerpts 3, 4, and 6). Lastly, if multimodal resources conflict at the TRP, then pragmatic (in)completion plays a pivotal role in turn-taking (Li, 2014). If the current speaker designs his/her turn as pragmatically complete, meaning that no new information can be projected, then the next speaker can take the floor (Excerpt 5). Otherwise, the current speaker can maintain their turn and extend it (Excerpts 3 and 4). It should be noted that pragmatic (in)completion is dependent on whether a new action is added or not but not on whether a new syntactic structure is employed.

In addition, the individuals engaged in the talk show are found to perform a dual role, both institutional and ordinary conversation, which leads to the configuration of a semi-institutional environment (Ilie, 2001) characterized by a specific participation framework. Besides fulfilling their obligations as the host and guests to accomplish a specific agenda, they also engage in casual conversation as friends through utilizing various multimodal resources to manage turn-taking.

On the one hand, the host and guests collaboratively contribute to the talk show's agenda, creating an institutional conversation (Ilie, 2001), evident in their recurring practices. The host allocates turns by addressing the recipient's name and coordinating with gaze and gestures (Excerpt 2). The host also self-selects to regulate the discussion topic to fulfill particular institutional goals (Excerpt 3). Both the host and guests also employ various semiotic resources to maintain the relevance of their utterances to the ongoing topic. Those semiotic resources include linguistic resources, such as anaphora *zhe* 'this' (Excerpts 3, 5, and 6) and *na* 'that' (Excerpt 4), and the temporal noun *gang* 'just now' (Excerpt 4), and embodied resources, such as pointing gestures to index the previous topic (Excerpts 4 and 5) and moving gestures from the previous speaker to the current one to establish a connection visually (Excerpt 3).

On the other hand, this talk show does not always take place under the control of the host but is also contingent on the participants' interaction, indicating co-participants' inclination towards framing casual conversations. One observable practice is the way participants allocate turns. Both the host and guests use sweeping gaze (Excerpts 1 and 4) and inclusive terms, such as *nimen* 'you (plural)' (Excerpts 1 and 7) and *zhuwei* 'everyone' (Excerpt 6) to establish a shared interactional space (Mondada, 2013b) among the four participants. This allows them to equally self-select as the next speaker at the TRPs, rather than relying solely on the host to allocate turns. Another practice is that the guests can self-select to (earlier) initiate a turn without necessarily being prompted by the host (Excerpts 5 and 6), coordinating with other embodied resources, such as gestures, torso movements, and gaze.

The shift between the two roles is monitored by the host, but it can also be negotiated between the host and guests (Ilie, 2001). However, it is essential to note that the participation framework of this show diverges from the semi-institutional talk shows analyzed by Ilie (2001). The participants in this show only need to design their turns based on the other three recipients and orient to resources produced by their co-participants, as there is no onlooking audience in the studio. Ilie's study primarily concentrated on the discursive and linguistic aspects of the talk show, whereas this study takes a step further by incorporating multimodal resources. For instance, the establishment of a shared interactional space through sweeping gaze and inclusive terms represents a distinctive semi-institutional characteristic in this multiparty conversation.

Altogether, the data analyzed in this study provides insight into the dynamics of Mandarin talk show. The findings illustrate how individuals manage turn-taking in a Mandarin talk show through multimodal resources and how participants orient to the dual roles to configure a semi-institutional setting. These findings deepen our understanding of the interplay of multimodal resources in the organization of turn-taking in larger groups and contribute to the current body of literature on multiparty institutional conversations in Mandarin.

## References

- Asmuß, B., & Svennevig, J. (2009). Meeting Talk: An Introduction. *The Journal of Business Communication (1973)*, 46(1), 3–22. <https://doi.org/10.1177/0021943608326761>
- Biq, Y.-O. (2001). The Grammaticalization of “Jiushi” and “Jiushishuo” in Mandarin Chinese. *Concentric: Studies in English Literature and Linguistics*, 27(2), 103–124. [https://doi.org/10.6241/concentric.ling.200106\\_27\(2\).0003](https://doi.org/10.6241/concentric.ling.200106_27(2).0003)
- Burch, A. R. (2014). Pursuing information: A conversation analytic perspective on communication strategies. *Language Learning*, 64(3), 651–684.
- Couper-Kuhlen, E., & Selting, M. (2018). *Interactional linguistics: studying language in social interaction*. Cambridge University Press.
- Duncan, S.D. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23, 283–292.
- Douban. (n.d.). *Roundtable π Season 5*. <https://movie.douban.com/subject/35529419/>
- Endo, T. (2019). Embodying stance: *wojue* ‘I feel/think’ and gaze. In X. Li., & T. Ono., (Eds). *Multimodality in Chinese interaction* (pp. 148–180). De Gruyter.
- Ford, C. E., Fox, B. A., & Thompson, S. A. (1996). Practices in the Construction of Turns: The “TCU” Revisited. *Pragmatics: Quarterly Publication of the International Pragmatics Association*, 6(3), 427–454. <https://doi.org/10.1075/prag.6.3.07for>
- Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: syntactic, intonational and pragmatic resources for the management of turns. In *INTERACTION AND GRAMMAR: Ochs, Elinor, Schegloff, Emanuel A., & Thompson, Sandra A. [Eds], England: Cambridge U Press, 1996*, pp 134–184 (pp. 134–184). Cambridge University Press. <https://doi.org/10.1017/CBO9780511620874.003>
- Gao, Y., & Ren, X. (2021). Syntactic parallelism and the co-production of syntactic units in Mandarin Chinese. *Chinese Language and Discourse*. 12(1), 109–134. <https://doi.org/10.1075/cld.00036.gao>
- Goodwin, C. (2018). *Co-Operative Action*. Cambridge University Press.
- Goodwin, M., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62(1-2), 51–76. <https://doi.org/10.1515/semi.1986.62.1-2.51>
- Han, H., & Shao, Z. (2007). A three dimensional cognitive study on demonstrative pronouns in English and Chinese. *Journal of East China University of Science and Technology (Social Science Edition)*. 22(3), 105–109. <https://doi.org/10.3969/j.issn.1008-7672.2007.03.022>
- Hauser, E. (2008). Non-formal institutional interaction in a conversation club: Conversation partners’ questions. *Journal of Applied Linguistics*, 5(3), 275–295.
- Hayashi, M. (2005). Joint turn construction through language and the body: Notes on embodiment in coordinated participation in situated activities. *Semiotica*, 156(156), 21–53. <https://doi.org/10.1515/semi.2005.2005.156.21>
- Heritage, J., & Clayman, S. (2010). *Talk in action: interactions, identities, and institutions*. Wiley-Blackwell.
- Ilie, C. (2001). Semi-institutional discourse: The case of talk shows. *Journal of Pragmatics*, 33(2), 209–254. [https://doi.org/10.1016/S0378-2166\(99\)00133-2](https://doi.org/10.1016/S0378-2166(99)00133-2)
- Jefferson, G. (1984). Notes on a systematic deployment of the acknowledgment tokens “Yeah”; and “Mm Hm.” *Papers in Linguistics (Edmonton)*, 17(2), 197–216. <https://doi.org/10.1080/08351818409389201>
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed).



- Conversation Analysis: Studies from the First Generation* (pp. 13-31). John Benjamins.
- Kalma, A. (1992). Gazing in triads: A powerful signal in floor apportionment. *British Journal of Social Psychology*, 31(1), 21–39. <https://doi.org/10.1111/j.2044-8309.1992.tb00953.x>
- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26(1), 22–63. [https://doi.org/10.1016/0001-6918\(67\)90005-4](https://doi.org/10.1016/0001-6918(67)90005-4)
- Kendrick, K. H. (2018). Adjusting epistemic gradients: The final particle ba in Mandarin Chinese conversation. *East Asian Pragmatics*, 3(1), 5–26. <https://doi.org/10.1558/eap.36120>
- Li, X. (2014). *Multimodality, Interaction and Turn-taking in Mandarin Conversation*. John Benjamins Publishing Company.
- Li, X. (2016). Some Interactional Uses of Syntactically Incomplete Turns in Mandarin Conversation. *Chinese Language and Discourse*, 7(2), 237–271. <https://doi.org/10.1075/cld.7.2.03li>
- Li, X., & Ono, T. (2019). Introduction: A multimodal approach to Chinese interaction. In X. Li., & T. Ono., (Eds). *Multimodality in Chinese interaction* (pp. 1-9). De Gruyter.
- Li, C., & Thompson, S. A. (1989). *Mandarin Chinese: a functional reference grammar*. University of California Press.
- Liu, B. (2016). The Mirative Use of “怎么[Zěnmě]” in Mandarin Chinese and Its Semantic Consequence. In *Chinese Lexical Semantics* (pp. 74–82). Springer International Publishing. [https://doi.org/10.1007/978-3-319-27194-1\\_8](https://doi.org/10.1007/978-3-319-27194-1_8)
- Lv, S. (1999). *Xiandai hanyu ba bai ci* [800 words in Modern Mandarin Chinese]. The Commercial Press.
- Lim, N. (2019). On co-operative modalities in the formulation of Mandarin Chinese turn-continuations. In X. Li., & T. Ono. (Eds). *Multimodality in Chinese interaction* (pp. 213-254). De Gruyter.
- Markaki, V., & Mondada, L. (2012). Embodied orientations towards co-participants in multinational meetings. *Discourse Studies*, 14(1), 31–52. <https://doi.org/10.1177/1461445611427210>
- Mondada, L. (2013a). Embodied and spatial resources for turn-taking in institutional multi-party interactions: Participatory democracy debates. *Journal of Pragmatics*, 46(1), 39–68. <https://doi.org/10.1016/j.pragma.2012.03.010>
- Mondada, L. (2013b). Interactional space and the study of embodied talk-in-interaction. In P. Auer, M. Hilpert, A. Stukenbrock & B. Szmrecsanyi (Eds.), *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives* (pp. 247-275). De Gruyter.
- Mondada, L. (2014). The local constitution of multimodal resources for social interaction. *Journal of Pragmatics*, 65, 137–156. <https://doi.org/10.1016/j.pragma.2014.04.004>
- Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, 20(3), 336-366.
- Mondada, L. (2018). Multiple temporalities of language and body in interaction: Challenges for transcribing multimodality. *Research on Language and Social Interaction*, 51(1), 85–106.
- Sacks, H., Schegloff, E., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4), 696-735. <https://doi.org/10.2307/412243>
- Tanaka, H. (1999). *Turn-Taking in Japanese Conversation: A Study in Grammar and Interaction*. John Benjamins Publishing Company.
- Tao, H. (2019). List gestures in Mandarin conversation and their implications for understanding multimodal interaction In X. Li., & T. Ono., (Eds). *Multimodality in Chinese interaction*

- (pp. 65-98). De Gruyter.
- Thompson, S. (2019). Multimodality and the study of Chinese talk-in-interaction. In X. Li, & T. Ono, (Eds). *Multimodality in Chinese interaction* (pp. 13–19). De Gruyter.
- Wu, R. (1997). Transforming Participation Frameworks in Multi-Party Mandarin Conversation: The Use of Discourse Particles and Body Behavior. *Issues in Applied Linguistics*, 8(2), 97–117. <https://doi.org/10.5070/L482005263>
- Wu, R. (2000). Final Particles in Mandarin Chinese: A Conversation Analysis of *A* and *OU* [Doctoral Dissertation, University of California Los Angeles].
- Wu, R. (2014). Managing turn entry: The design of EI-prefaced turns in Mandarin conversation. *Journal of Pragmatics*, 66, 139–161. <https://doi.org/10.1016/j.pragma.2014.03.003>
- Yang, P. (2011). Nonverbal Aspects of Turn Taking in Mandarin Chinese Interaction. *Chinese Language and Discourse: An International and Interdisciplinary Journal*, 2(1), 99–130. <https://doi.org/10.1075/cld.2.1.09yan>
- Yeh, K., & Huang, C. (2016). Mandarin-speaking children’s use of the discourse markers hao “okay” and dui “right” in peer interaction. *Language Sciences (Oxford)*, 57, 1–20. <https://doi.org/10.1016/j.langsci.2016.04.004>
- YOUKU. (2019). Roundtable π Season Four [Video]. YouTube. <https://www.youtube.com/watch?v=dUG258fIGZ8&list=PLATwx1z00HseewVmomMiP0IfNfT91D4ek>
- Zhang, W., & Gao. H. (2012). Ziran huihua zhong “jiushi” de huayu gongneng yu yufahua yanjiu [Functions of “Jiushi” in Everyday Conversation and Its Grammaticalization.]. *Yuyan jiaoxue yu yanjiu [Language Teaching and Research]* 2012(1).91–98.

### Author’s information

Jiixin Tian  
 University of Hawaii at Manoa  
 jiixin24@hawaii.edu

### Appendices

#### Appendix A: Database

Episode	Topic	Length
S2E23	Peace of mind: Does your sense of security make you safe?	44-min
S3E14	Rules: How do you understand rules?	52-min
S4E04	Working: Who wants 996? Who is forced to do 966?	62-min
S4E09	Shopping: purchasing, refunding, exchanging, troubling	56-min
S4E23	Misreading: Do you often encounter misunderstanding?	58-min
S4E26	Family affection: Our parents	62-min
S4E28	Living alone: Is it good to live alone?	53-min

#### Appendix B: Transcription conventions:

Talk (adopted from Jefferson 2004 & Li 2019)

[overlapping talk	
=	latching
wor:d	stretched sounds
wor-	cut-off sounds
°word °	talk with a reduced volume
<u>word</u>	emphasis/accent
WORD	loud talk
>word<	speed up (quicker tempo than the surrounding talk)
.h	hearable breathing in
(.)	micropause, shorter than 0.2 seconds
(0.7)	pause (timed in tenths of seconds)
,	mid-rise to high
;	low fall to low
hhh	laughter
£	smiling voice

### Grammatical gloss

COP	copular (e.g., shi)
3SG	3 <sup>rd</sup> singular pronoun
Q	question particle (e.g., ma)
CL	classifier (e.g., ge)
PRT	particle (e.g., a, ha)
NEG	negation (e.g., bu, mei)
LE	le 了
DE	de 的
BA	ba 把
ASP	aspect marker (e.g., le)

### Embodied actions (adopted from Burch 2014)

gz	gaze
BH	both hands
LH	left hand
RH	right hand
BA	both arm
RA	right arm
F	fingers
IF	index finger
SPU	supine up (palm facing up)
SPV	supine vertical (palm facing in)
PRD	prone down (palm facing ground)
PRV	prone vertical (palm facing out, i.e., “stop”)
+	onset of change
>	“to” – direction of gaze shift or movement
((>line XX))	continued movement or hold of gesture
-->	continuation of the action