Chapter 5. Dynamicity and Stativity of the Mandarin Chinese Imperfective Aspect: A Multimodal Perspective

5.1 Introduction

Aspect is concerned with “the way the grammar marks the duration of type of temporal activity denoted by the verb” (Crystal 1991: 27). It relates to “the internal temporary constituency” (Comrie 1976: 5), or “the internal temporal structure” (Trask 1993: 21), of a situation. Grammatical aspect allows for varying perspectives on the very same events (Givón 2001; Langacker 2001; Smith 1994). There are two basic grammatical aspect distinctions found in many languages, that is, perfective aspect vs. imperfective aspect. The perfective aspect views an event as a whole from an external perspective, whereas the imperfective aspect is characterized by “explicit reference to the internal temporal structure of a situation, viewing a situation from within” (Comrie 1976: 24). Consequently, the imperfective aspect registers finer-grained details and increases imagined image accuracy (Tversky & Paul 1998). Furthermore, the imperfective aspect is divided into two categories: habitual aspect and continuous aspect. The latter is subdivided into progressive aspect and non-progressive aspect (Comrie 1976: 28).

The imperfective aspect can be realized in one or several grammatical forms in different languages. For example, English only has one imperfective aspect marker – *be*...*ing* to encode events as ongoing. Langacker (2001: 259) treats the imperfective aspect in English (i.e. progressive aspect *be*...*ing*) as a device of “zooming in” and “taking an internal perspective” of an event. However, Mandarin Chinese has two main separate imperfective aspect markers – *zài* and *zhe*. Historically, *zhe* has long been considered as the single progressive aspect marker in Chinese. With the grammaticalization of *zài*, it is gradually taking the place of *zhe* as the commonly
acknowledged progressive aspect marker. By contrast, zhe is seen as the durative aspect marker. Syntactically, zài is used before the verb, whereas -zhe follows the verb. The former is autonomous, while the latter is dependent on other elements to form a sentence. For example, tā zài chí fàn (He is having dinner) is grammatical, but *tā chí zhe fàn (*He is having dinner) is ungrammatical. In order to be acceptable, it needs some other elements to be added like ne (a particle). Semantically, there has been a long debate about the differences between zài and zhe, and particularly, about zhe among the linguists. Some researchers proposed that the dualism of dynamicity vs. stativity is the basic semantic distinction between zài and zhe (e.g., Comrie 1976; Li & Thompson 1981; Smith 1991; Teng 1979). They concluded that zài dominantly co-occurs with active verb predicates, representing dynamicity. Zhe has been taken to usually be attached to stative verb predicates to express stativity (e.g., Dai 1991, 1997; Li & Thompson 1981; Smith 1991; Teng 1979). Smith (1991) even termed it the “stative imperfective zhe”. Some scholars argue that the core grammatical meaning of zhe is “durativity”, only focusing on the continuation of events or states (Dai 1991, 1997; Xiao & McEnery 2004). Yang (2013) suggests that the construction meaning of “zài + V” is “event is being on”, and that of “V + zhe” is “constant state”. However, others argued that there were two kinds of zhe in Mandarin Chinese: zhe₁ is the durative aspect marker, and zhe₂ is the continuous or progressive aspect marker, but they experienced different grammaticalization pathways (Liang 2010). While zhe₁ encodes the duration of static events, zhe₂ can also focus on the ongoingness of dynamic events, as zài does when the verb is an action verb (e.g., Lű 1980; Lu 1999; Wang 2012; Zhu 1982). Some hold that there is only one zhe in Chinese, and that it has two grammatical meanings, that is, the ongoingness of dynamic action and durativity of states (Lű 1980; Zhu 1982). Both Jin (1991) and Qian (2000) claimed that zhe cannot denote “ongoingness/progressive”, and “the basic grammatical meaning of zhe is ‘state’, signaling that the event is in a state and in a relatively stable and homogenous developmental stage” (Jin 1991: 26). Jin (1991) further argued that the differences of stativity and dynamicity resulted from the related verbs themselves rather than the grammatical meanings of zhe. Wang (2012) explains that the semantic contrast of dynamicity/stativity can be only applicable to zài,
which co-occurs with dynamic and durative verb predicates. She argues that the “homogeneity” of verb predicates (VPs) plays a central role in the semantic interpretation of zhe, which only co-occurs with homogeneous VPs. The overlapping use of zài and zhe lies in encoding the ongoingness of “dynamic and homogeneous events” (Wang 2012: 243). The debate on the aspectual construal of progressive and durative aspects in Mandarin Chinese is still ongoing among Chinese linguists. However, focusing only on the linguistic evidence, as the previous studies have done, easily leads to cyclic argumentation in reasoning about linguistic forms and their meaning. Co-speech gestures have been discussed as a possible means of indicating aspectual distinctions in event representations. Gestures with the imperfective aspect categories in both English and Chinese are longer and more complex than those with the perfective aspect (Duncan 2002; McNeill 2003, Matlock et al. 2012; Parrill et al. 2013). Duncan (2002: 197) has also noted that gestures with durative aspect are more complex in form and represent more semantic content than those with progressive aspect. It is still unclear whether and how the widely discussed dynamic-static distinction of progressive aspect and durative aspect in linguistics can also be represented in gestures.

Cognitive Linguistics claims that “(grammatical) meaning is identified with conceptualization” (Langacker 2008a: 30). Meaning as conceptualization has the potential to be realized not only in language, but also in co-speech gestures as well (Langacker 2008a, b; McNeill 1992). As simulated action (Hostetter & Alibali 2008), gesture is a window to “part of the speaker’s current cognitive being, her very mental existence, at the moment it occurs” (McNeill 2005:92). According to the thinking for speaking and gesturing hypothesis (Cienki 2008; Slobin 1987, 1996), differences in the grammatical options provided in a language or across languages can influence the speaker’s thinking in the moment-by-moment processes of talking and gesturing. Therefore, it is of great significance to study whether and how the two Mandarin imperfective aspect categories relate to Chinese speakers’ speaking and gesturing. This chapter intends to investigate whether the insight into the conceptualization of events provided by gestural data in natural conversations supports a particular semantic model.
of aspect for the imperfective in Chinese: that of the imperfective as one coherent whole category; that of \( zài \) and \( zhe \) as each expressing their own internally consistent semantic category; or that of \( zhe \) as comprising two subcategories, one of which functions more like \( zài \). This will be investigated by finding the answers to two empirical questions. First, do gestures accompanying imperfective utterances represent the semantic differences of dynamic vs. static distinctions between the progressive aspect and the durative aspect? Second, in what contexts do gestures with dynamic or static qualities accompany the imperfective constructions?

### 5.2 Methods

#### 5.2.1 Corpus description

Our data were extracted from 32 videos from *Yang Lan One on One*. Each video lasts for about 20 minutes (excluding the advertisements and inserted video clips), so the multimodal corpus is 640 minutes in total. I went through each video carefully, and sampled out all instances of the two kinds of imperfective aspect-marked utterances with iconic (concrete or metaphoric) gestures, which are described below. In total, there were 48 progressive \( zài \) and 40 durative \( zhe \) marked-utterances with such explicit concrete or metaphoric iconic gestures. The 88 imperfective aspect-marked utterances with iconic gestures constitute a sub-corpus of multimodal data marked with the imperfective aspect.

#### 5.2.2 Data coding

All the data were coded in ELAN, video annotation software developed by the Max Planck Institute for Psycholinguistics (see Lausberg & Sloetjes (2009) and [http://tla.mpi.nl/tools/tla-tools/elan/](http://tla.mpi.nl/tools/tla-tools/elan/) for details).
5.2.2.1 Speech

All the utterances marked by the imperfective aspect markers were annotated as the progressive aspect or the durative aspect, but only the utterances marked by progressive aspect marker zài and durative aspect marker zhe with iconic gestures were further analyzed. Two coders coded all the 88 imperfective aspect-marked utterances. The agreement was 98%.

5.2.2.2 Gesture

Every gesture phrase, i.e. a whole gestural movement, consists of several possible “phases”: preparation, stroke, stroke hold, hold, and retraction (Kendon 2004; McNeill 2005; Kita 1990). The preparation phase refers to the phase when the hand rises from the rest position to the position for starting the stroke. The stroke is essential in terms of the movement structure of a gesture, involving the peak of the effort exerted (McNeill 1992). It “displays semantic content related to that of the speech with which the gesture co-occurs” (Duncan 2002: 193). It may be synchronous with co-expressive speech or slightly precede the speech with which it links semantically. Stroke holds are “independent holds” and relatively static, which stand on their own. The hold is the phase in which the hands maintain the stroke’s final position and form motionlessly. However, sometimes beats rhythmically oscillate the hold with the prosody of the speech that was co-expressive with the gesture stroke or stroke hold. Retraction is the phase in which the hands return to rest at the place where they started or some other relaxed position. Only the stroke or stroke hold phase is obligatory to constitute a gesture, and the other phases are all optional.

The iconic (concrete or metaphoric) gestures accompanying the progressive zài or the durative zhe marked utterances were annotated in this study. In concrete iconic gestures, the physical entities, actions, and spatial properties being mentioned in speech are schematically mapped out by the shape and/or motion of the hands with some degree of iconicity. For example, while uttering She walked along the path, the speaker may move her right extended finger from left to right in front as if drawing out the path
she actually took. A metaphoric iconic gesture represents an abstract concept (i.e., the abstract entities, processes, and properties) by iconically depicting some physical referent in terms of which the abstract concept can be understood (Cienki 1998a, b, 2008a, b). For example, while talking about honesty or honest people with a verbal metaphoric expression (e.g., *she’s straight as an arrow*) or even without it, the speaker may make a gesture with a tense, open, vertically flat (straight) hand, thumb at the top to iconically depict the abstract concept HONEST as STRAIGHT (Cienki 1998a, b).

Such iconic gestures were annotated in detail in terms of gesture phases. Then we coded the property of the gesture stroke or stroke hold as dynamic or static. When the gesture stroke was in continuous motion, representing the action (concrete or abstract) involved in an event, it was coded as “dynamic”. When the gesture stroke or stroke hold was sustained, and even when there were biphasic beats added, it was coded as “static”. This was done because the beats do not represent action-related meaning in speech, but they are only important for the timing of vocalizations (Bernard et al. 2015). Two coders coded all 88 gestures. The agreement for gesture feature (dynamic versus static) was 97%.

### 5.3 Results

According to Table 5.1 below, you can see that almost all of the strokes of gestures accompanying the progressive *zài* are dynamic, which is far more frequently dynamic than those with the durative *zhe* utterances (93.75% vs. 55%). The strokes of gestures co-occurring with the durative *zhe* tend to be a little more dynamic than static (55% vs. 45%). The differences are significant ($\chi^2(1) = 18.032$, $p < 0.05$). The results suggest that gestures accompanying the progressive aspect do represent its aspectual meaning, that is, dynamicity, and those co-occurring with the durative aspect encode either dynamicity or stativity. The gestural data are consistent with the linguistic evidence that the progressive aspect *zài* encodes “progressivity” or “ongoingness” (e.g., Comrie 1976), and “dynamicity” of dynamic events (e.g., Smith 1991; Xiao & McEnery 2004;
Dai 1997), and that the durative aspect zhe can encode not only “stativity” of static events but also “dynamicity” of dynamic events (e.g., Lu 1999; Wang 2012). However, the gestural result also shows that there is no predominant use of durative zhe to encode the durative states rather than the dynamic events, which the previous linguistic studies emphasized.

Table 5.1: The frequencies of gesture feature with the imperfective aspect-marked utterances

<table>
<thead>
<tr>
<th>Imperfective aspect</th>
<th>Gesture stroke feature</th>
<th>Dynamic</th>
<th>Static</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Progressive zài</td>
<td>45</td>
<td>93.75</td>
<td>3</td>
<td>6.25</td>
</tr>
<tr>
<td>Durative -zhe</td>
<td>22</td>
<td>55</td>
<td>18</td>
<td>45</td>
</tr>
<tr>
<td>Total</td>
<td>67</td>
<td>76.14</td>
<td>21</td>
<td>23.86</td>
</tr>
</tbody>
</table>

Keeping the quantitative results into mind, I will now analyze the qualitative data to see the detailed contexts in which the dynamic and static gestures occurred for two imperfective aspect-marked utterances in the following sections. In addition, I also discuss some revealing examples about how co-speech gestures contribute to sentence meaning.

5.3.1 Dynamicity and the gestures with progressive aspect-marked utterances

The strokes of gestures accompanying progressive zài-framed utterances were prototypically dynamic, which explicitly represented the ongoingness of the relevant progressive aspect-framed actions. Although the handshape, orientation, movement quality etc. were varied, the gesture strokes shared the schematic feature of dynamicity with energetic motion for either concrete or abstract events.

The iconic gestures are argued to be constructed via different modes of representation or techniques of representation (Kendon 2004; Müller 1998, 2014, see Chapter 2 for detailed review). For concrete events, there are a lot of progressive aspect-
framed utterances, allowing for layered aspectual expression in natural conversations. That is, while the verb predicate is used with progressive aspect, other elements, adverbials in particular, coerce the whole utterance to be imperfective aspect, but not progressive aspect any more. In such cases, the most salient features of actions encoded by the verb predicates were still profiled by the co-speech gestures, such as in example (1a) and (1b). In (1a), the speaker recounted that his son worked very hard until 2 a.m. or 3 a.m. every night, writing and drawing. In speech, the verb predicate zài xiě a huà a (was writing and drawing) expresses the ongoingness of the dynamic writing and drawing events. However, the adverbial méi tiān wǎn shàng (every night) indicates that the whole utterance denotes habitual aspect. It means that the event of his son writing and drawing until 2 a.m. or 3 a.m. occurred repeatedly during the film making. Such dynamicity and iterativity have been mapped onto the gesture. As presented in Fig. 5.1a, his index finger touching upon the thumb of his right hand moved horizontally from the left to the right in a zigzag way as if holding a pen or pencil and reenacting writing or drawing with it three times while uttering dōu zài xiě a huà a (he was even writing or drawing).

(1) a. [[[wǒ méi tiān wǎn shàng kàn dào][tā liǎng][sān diǎn zhōng]
   tā] [dōu zài <0.5s> dōu zài xiě a huà a]]]
   he even PROG even PROG write PRT draw PRT
   “I saw that he was even writing and drawing every night at 2 a.m. or 3 a.m.”

b. [[[tā hái zài pá, cái][qī gè duō yuè], [tā zài pá]]]
   He still PROG crawl, still seven-CLF more month, he PROG crawl
   “He was still in the crawling phase. Because he was only more than seven months old, he was crawling.”
Before the utterance in (1b), the speaker was narrating the situation where her son was playing near a large pond in their garden. His son was too little to walk. It was dangerous that he crawled to the side of the pond. The verb predicate of the first sentence 在 到 (be crawling) in (1b) encodes progressive aspect, but the adverbial 还 (still) shows that the aspectual type of the whole sentence is imperfective, i.e., state (non-progressive) instead of progressive. As Fig. 5.1b shows, the speaker alternately moved her two fists back and forth once while uttering 他还 在 到 (he was still in the phase of crawling), and then held for a moment; then she made the same gesture again while uttering the second utterance even without a lexical affiliate for 到 (crawl), and she repeated the gesture again for the third progressive utterance 他在 到 (he was crawling). Hence, the gesture illustrates the crawling behavior that is typical of this phase of a seven-month-old child, but not the aspectual type of being in this phase, which is static. The actual action of writing or drawing on an imagined piece of paper with a pen, and the characteristics of crawling on the ground, were presumably simulated in the speakers’ mind and was reenacted vividly by the hand gestures. The dynamic and iterated gesture motion visually demonstrated not only that such actions were ongoing, but also how the actions were unfolding with time. They can be seen not only as a reflection of the mental image of the speakers’ immediate thinking, but also as a stimulus for activating the listeners’ mental image of such actions.

A concrete gestural image can provide the substrate for metaphorical depiction of an abstract event (Cienki 1998a, b; Cienki & Müller 2008a, b). Thus, for abstract events
framed by the progressive aspect, co-speech gestures occurring with them were iconic for physical actions being used to describe metaphoric events, as in example (2a), or iconic for the schematic process of ongoing inherently abstract events, as in example (2b).

(2) a. [zhèyàng de  hàngyě yě  zài  zhújiàn de  kāifāng].

such  GEN  field  also  PROG  gradual  PRT  open

“Such fields are also gradually opening to the public.”

b. [tā jiù gēn wǒ yǐzhí zài. yǐzhí zài  gā gā gā gā]

he  then  with I all the time  PROG  all the time  PROG  (onomatopoeia)

“Then he was talking to me all the time.”

Fig. 5.2a Gesture with “PROG open”  Fig. 5.2b Gesture with “PROG (talk)”

In (2a), the speaker moved her two hands with palm towards center from the center to the two sides as if opening something concrete (see Fig. 5.2a), which was synchronous with the progressive aspect construction zài zhújiàn de kāifāng (are gradually opening). The opening of the previous monopoly industries that she was discussing was simulated as if embodying something opening outward to the sides, like curtains or sliding doors. It demonstrates that the actions of abstract entities can be rooted in everyday actions.

In (2b), the speaker was recounting that the director was persuading her to do something she was not willing to do. She enacted out the embarrassing situation as presented in Fig. 5.2b. She thrusts her right hand with palm towards body between the addressee and herself alternately four times through the whole utterance. She used the
onomatopoeia $g\tilde{a}$ (the sound of the duck quacking to depict talking quickly and a lot) in speech, but her co-speech gestures metaphorically represented the process of talking: the opening and closing of her fingers can be interpreted as acting out the opening and closing of a mouth while talking. The schematic duration of the process of persuasion is embodied in the iterated back and forth movement of her hand. The metaphor is expressed in her gesture: COMMUNICATION IS LIKE TRANSFERING SOMETHING IN THE AIR (the CONDUIT metaphor of communication, known from Reddy [1979/1993]). The reduplication use of $g\tilde{a}$ in speech was represented by the dynamic, cyclic and iterated gestures, co-expressing the continuous and ongoing events of persuasion. In addition, her closed eyes and lowered head implied that she refused to communicate with the director.

The progressive aspect-marked utterances with generic verbs such as zuò (do), jinxing (proceed), etc. refer to abstract conceptual actions that do not themselves physically exist in the world. Consequently, they have no concrete images in the speakers’ mind. Instead, they were prototypically accompanied by the back and forth gestures and occasionally by cyclic gestures (Ladewig 2011, 2014), which are known to characterize the ongoingness and duration of the inherently abstract events, as example (3a) and (3b) demonstrate, respectively.

(3) a. [shìchăng lāobāixīng yě zài tiáozhèng]
   market ordinary people too PROG adjust
   “The ordinary people are adjusting themselves to markets too”

b. [wǒ jué de shì zài jinxīng yī-gē xīlī].
   I think PRT COP PROG doing one-CLF baptism.
   “I think we were going through the baptism (of tough life)”
In (3a), the speaker used the generic verb tiáozhěng (adjust) to express the abstract action people took towards the change of market. As Fig. 5.3a shows, he made a “back and forth” gesture by alternating the intense open right hand away and towards his body while uttering lǎobǎixìng yě zài tiáozhěng (the ordinary people also are adjusting). The “back and forth” gesture may stem from the BALANCE schema. The oscillation of hand(s) embodies the imbalance of an ongoing process with a changing state. In (3b), the actress employed the generic verb jìnxing (proceed). She made a two-handed mirroring cyclic gesture with the whole predicate by moving the hands downward-outward-inward vertically and then ending with an open hand palm up gesture with a hold for 10 seconds before the next utterance (see Fig. 5.3b). The highly schematic gestural form can be seen as abstracted from the perception of dynamic events in our lives that are cyclic in nature, which form the CYCLE schema in our minds.

In addition, there were also four gestures iconic for abstract entities (TIME) mentioned in progressive aspect utterances, as in example (4a) and (4b). Three were static and one was dynamic. In (4a), the speaker held her two hands, palms towards center in front, as if molding the length of something, as Fig. 5.4a presents. The static stroke was synchronous with the time adverb xiànzài (now), and it was held with energetic beats for emphasis when the progressive construction zài zuò (be doing) occurred. Such a gesture is very popular in the corpus when time is referred to. Thus, I consider it as a metaphoric gesture (TIME IS A CONTAINER or A LIMITED AMOUNT OF TIME IS A BOUNDED SPACE), which profiled the duration of the
general event “doing this thing” to some extent. Such a gesture has also been found to be used to teach one of the characteristics of English progressive aspect, i.e., duration, in second language classroom (Matsumoto & Dobs 2016).

(4) a. [wò xiànzái zài zuò zhègè shì].
   I now PROG do this thing
   “I’m doing this thing now.”

b. [tā yīzhí zài <0.2s> zài zhè sān sì tiān tā zài nèijiù]
   he all the time in in this three four day he PROG guilty
   “He felt guilty throughout the three or four days.”

![Fig. 5.4a Gesture with “now PROG do”](image1)

![Fig. 5.4b Gesture of “PROG feel guilty”](image2)

Furthermore, the mental activity verb “nèijiù a” (feel guilty) is also not affordable for gestural representation. As shown in (4b) and Fig. 5.4b, the speaker moved his right hand to the upper center with the extended thumb and index finger and the other fingers curled into the palm as preparation during the ending of the last utterance and starting of this utterance, then he moved it quickly from the center to the right periphery as he uttered tā yīzhí zài (he all the time), and then the gesture held with beats for the lexical affiliates sān sì tiān (three or four days) but kept static for the progressive construction zài nèijiù (feel guilty) until the end of this utterance. Such a vertically paralleling thumb and extended index finger representing a scale can also metaphorically stand for the abstract concept PERIOD OF TIME (three or four days), as the diametrically opposed and separated pads of the thumb and index finger are often used to indicate an extent
of something (Hassemer 2015). Although this gesture stroke was also dynamic, it is
different from the other dynamic gestures accompanying the progressive aspect
utterances discussed above. This dynamic gesture may imply the continuous flowing
of time or maybe it just helped bring the idea forward to the addressee. In a word, the
schematic verbs used in speech were correlated with the speakers’ way of construing
events in gestures; the metaphoric gestures for time can be seen as reifying the abstract
actions or just the period of time the actions take.

According to the above qualitative analyses, I can conclude that the prototypical
gestures accompanying the progressive zài-framed utterances dynamically reenact out
the physical actions of concrete or metaphoric events, even when the layered aspectual
type is imperfective like habitual or state rather than progressive aspect any longer. The
“back and forth” gestures and cyclic gestures represent the schematic process of
abstract events. Moreover, the peripheral gestures accompanying the progressive zài-
framed utterances statically mold the temporal period for which the abstract event lasted.
Thus, the ongoingness and duration of the progressive aspect-framed events are
embodied by their dynamic co-speech gestures.

5.3.2 Stativity/Dynamicity and the gestures with durative aspect-marked
utterances

Now I will turn to the features of gesture strokes of the durative aspect-marked
utterances. The qualitative analyses will present a panoramic picture of the linguistic
contexts in which the static and the dynamic gestures are accompanying the durative
aspect-marked utterances.

5.3.2.1 Static durativity

Among the 40 instances, there were 18 cases with a static stroke hold. The verbs
inherently involve long duration: cáng (hide), chéngzài (carry), ná (pick up, hold, carry),
bào (hold or carry), tí (carry), zhà (open), chān (support), kàn (watch), and two
adjectives indicated the acquired states: pàng (fat) and zhōng (swollen). The manner
verbs have ambiguous uses. On the one hand, they can denote movements leading to a certain configuration. On the other hand, they may denote just the resulting state. For example, the verb *ná* either denotes a dynamic action in *tā ná qǐ yī běn shū* (She picks up a book), or denotes the resulting state *tā ná zhe yī běn shū* (She is holding a book in her hand). When such manner verbs co-occur with the durative *zhe*, the co-speech gestures are characterized by static stroke hold with long duration.

Most of the static stroke holds enacted out the maintenance of one portion of the actual actions as in examples (5a) and (5b), some added complementary action as in example (5c), and those with adjectives demonstrated the resultant state as in example (5d). When the speaker was recounting that his daughter was crying at the beginning of a live TV show, he needed to hold her in his arms at that time in (5a). As shown in Fig. 5.5a, he kept his two crossed hands in front of his chest as if holding something in his arms, which was synchronous with the durative construction *bào zhe tā* (holding her).

(5) a. [qíshí wǒ jiù //bào zhe tā]  
Actually I then hold DUR her  
"Actually, I was holding her in my arms at that time."

b. [Yìmóu tā //yòngyuăn shì bào zhe jīqì]  
Yimou he forever be hold DUR machine  
"Yimou was holding the machine forever."

---

**Fig. 5.5a** Gesture with “hold DUR her”  
**Fig. 5.5b** Gesture with “hold DUR machine”
In (5b) and Fig. 5.5b, the speaker recounted that the director Yìmóu was afraid of breaking the recording machine, so he was always holding it by himself even when climbing mountains. The speaker held each hand with palm lateral towards center at his side, which is the apex of the holding action and synchronous with yōngyuăn shì bào zhe jīqi (was holding machine forever). The objects for holding in (5a) and (5b) were different, so the locations, handshapes of gestures differed from each other. This is also the case for other frequently used verbs like ná (hold, carry). However, the common characteristics for them are that the stroke hold was relatively static (except on some occasions when there were beats added to it for emphasis) and often maintained for the whole verb predicates, several utterances following or even until the ending of the whole turn.

(5) c. [wǒ zhè /yī biàn kàn zhe diànhishì, yībiān gěi tā hǎn zhe wǔ sì]
   I this one side watch DUR TV, one side help her shout DUR five four]
   “While I was watching TV, I was counting down ‘five, four’ with her.”

   d. [zhēng tiān lài yǎnpāo zhōng zhe]
   whole day big upper eyelid swell DUR
   “My upper eyelids were swollen in the whole day.”

In (5c), the speaker was narrating the situation when he was waiting for his wife to give birth to his son in the hospital. He moved his right fist from the rest position over his shoulder, which lacked a lexical affiliate in speech. As Fig. 5.5c displays, this
stroke hold was actually iconic for holding a TV remote control in his hand. His eye gaze moved away from the addressee upwards and stared at an imagined TV like one may find overhead in public places, which was synchronous with the hand gesture and also with the yī biān kàn zhe diànsī (was watching TV). Then he lowered his head and counted down “five, four” when speaking the next utterance with durative aspect construction yībiān gěi tā hàn zhe wǔ sì (was counting down: five, four…). During the whole process, the right hand gesture stayed in place, encoding the static duration of the action of holding TV remote control. The hand gesture, eye gaze and body movement encoded things like the visual and spatial perspective/orientation of the speaker (e.g., looking upward) and added a lot of information, which could not be interpreted from the verbal information alone. The total meaning of such a multimodal utterance is of great potential significance for communication. The static gesture starting before the durative aspect-marked utterances and being held through the whole situation instantiates the long maintenance and duration of the event.

In (5d), the speaker was narrating the results of his weight loss by drinking a lot of water without eating something else. The resultative state is demonstrated by the gesture in Fig. 5.5d. He put his open right hand with palm towards body over his right eye to mold (as if touching) the big and swollen eye, which was synchronous with the whole durative aspect-marked speech. Such a static stroke hold does not represent the nucleus of an action as those in (5a) to (5c), but profiles the duration of the resultant state. This is consistent with previous linguistic results indicating that the durative aspect zhe has a resultative state reading (e.g., Smith 1991).

When taking a closer look at the static gestures with the durative aspect-marked utterances, I found that the stroke hold usually preceded the lexical affiliate and occurred with the emergence of the topic, and that stroke holds tended to last over at least the clause the lexical affiliate was embedded in and even over several clauses until the turn ended. The long static duration of gestures with spoken durative aspect would seem to characterize the focus on the nucleus of an action in an event, perhaps “to facilitate the greater complexity of depiction” of some simultaneous events (Duncan 2002: 199).
5.3.2.2 Dynamic durativity

As Table 1 shows, 55% of the gestures accompanying durative aspect speech exhibited the kind of dynamicity found with the progressive aspect. The dynamic gestures accompanying imperfective durative aspect speech mainly occurred in the following three typical linguistic contexts.

First, they occurred with verbs that inherently express strong dynamicity. When the verbs in speech were highly dynamic, the gestures with the durative aspect were always as dynamic as those with the progressive aspect in speech. Consider (6a) and (6b), for instance. The same verb *tuīdòng* (push) was used with the two different aspect categories in metaphoric events with similar semantic purpose. Both of their co-speech gestures dynamically reenacted the action of pushing something as if in actual situations. In (6a) and Fig. 5.6a, the speaker in the same interview also moved his left hand with palm towards body away and towards his body alternately three times to reenact pushing the abstract “e-commerce” as pushing some concrete things. Similarly, in (6b) and Fig. 5.6b, the other guest moved his palm-down right hand downward with diminishing repetitions to act out pushing the abstract “e-commerce” as if pushing some physical object. Therefore, the gestures revealed that the durative aspect-marked event of pushing encoded the “ongoingness” and “dynamicity” as the progressive aspect-marked event of pushing did. In addition, the interlocutors demonstrated their “interactive alignment” both lexically (with the verb *tuīdòng*) and gesturally (with acting mode of repeated gestures) (Oben & Brône 2016: 32).

(6) a. [zhèyàng de shìqíng shì tuīdòng zhe zhōngguó de zhègè diànzǐ shāngwù.] such GEN thing COP push DUR China GEN this electronic commerce

“Such things were pushing the Chinese e-commerce.”

b. [[yǒu hěnduō yě yǒu zhèngfǔ…(0.6s)] [zài tuīdòng nèixiē diàn]]zì shāngwù. have many also have government PROG push those electronic commerce.

“There were many..also governments which were pushing those e-commerc...”
Second, when the monosyllabic action verbs such as bào (hold), chāo (copy), chī (eat), hē (drink) are followed by zhe and reduplicated as “V zhe V zhe”, the construction conveys that while the event denoted by the verb was ongoing, another new event encoded by the following clause or utterance occurs. In example (7a), the speaker was narrating the situation where he needed to throw a baby onto the ground for a film scene. Of course, he used a doll instead, but he was still very nervous. At one point, he thought he felt the doll move, as if it had been a real baby. Thus, he made a gesture in which his two hands with palm towards body shook in front of him, as Fig. 5.7a presents. The dynamic gesture was synchronous with the whole utterance with reduplicated durative aspect and the following clause. In speech, the durative aspect-framed event bào zhe (holding the doll in his arms) is static, and the second event dòng (move) is dynamic. The dynamicity of the new event was integrated into the gesture accompanying the durative utterance here. Hence, the co-speech gesture combined the duration of the holding action and the moving action, profiling the simultaneous occurrence of the two actions.

(7) a. [wō bào zhe bào zhe, jiā de jiù jué de dòng]
I hold DUR hold DUR, fake PRT then feel PRT move.

“When I was holding the doll, I felt it moved.”
(7) b. [wō ná zhe máojīn níng.]
I hold DUR towel twist.
“While holding the towel, I twisted it.”

c. [wōmen jiù ná zhe gē zhǐ bèi]
we then hold DUR CLF paper cup
dào le hěnduō xiāngbīn, bā bā bā bā]
pour ACTL much Champagne, [onomatopoeia].
“Then while holding a paper cup, we poured much Champagne.”

Third, the ná zhe (NP)V (hold and do something) construction was also usually accompanied by dynamic gestures to profile the duration of the second verb action while the action of “holding” was backgrounded. When two events occur
simultaneously, the durative aspect is likely to occur, because it functions to background events in discourse (e.g., Duncan 2002). Co-speech gestures with such utterances also represented this characteristic, as in example (7c) and (7d). As example (7c) and Fig. 5.7c displays, the speaker rotated his two fists simultaneously three times as if performing the action of holding something in his hands and twisting it, which was synchronous with the whole durative aspectual utterance. The salient motion in the gesture foregrounded the twisting event, and the fists backgrounded the event of holding a towel horizontally. In contrast, in example (7d), the speaker made a very complex gesture in contrasting handshape, space and movement quality for representing the two simultaneous events. The gesture involves a combination of stativity in one hand and dynamicity in the other hand. As Fig. 5.7d shows, his left hand with the index finger touching the thumb and the other fingers slightly curled represented the “paper cup” and its static maintenance backgrounded the long duration of the events of nà zhe gè zhī běi (was holding a paper cup). In comparison, his right fist, which enacted the holding of a round wine bottle, moved cyclically while he uttered dào le hěnduō xiāngbīn (poured much Champagne) and changed to iterated back and forth motion while he uttered bā bā bā bā (using the onomatopoeia to simulate the sound of pouring wine). The dynamic right hand gesture foregrounded the event of pouring Champagne, which revealed that the speaker poured much Champagne into many different paper cups for a lot of people there, rather than poured a lot of Champagne into a single paper cup. Apparently, the speaker provided gestural repair of a performance error in the speech. The speech and gesture form a composite (compare Enfield 2009 on composite utterances) to express the total meaning of the events, which go far beyond the speech alone.

To recapitulate: when a durative aspect-framed utterance expressing long duration or resultant state is framed by the durative aspect, its co-speech gesture tends to keep static and be held in place to represent stativity of a durative event; when the durative aspect frames such constructions as V zhe NP (verbs with high dynamicity), V zhe V zhe (V is reduplicated), and ná zhe V (hold/carry zhe V), the co-speech gestures are prototypically dynamic, like those in the progressive aspect-framed constructions.
5.4 Discussion and conclusions

When describing events with progressive aspect, almost all native Chinese speakers in the present dataset produced iconic gestures that were highly dynamic. They represented the finer-grained details of the actions in concrete and metaphoric events, or the schematic progress and temporal duration of the actions in inherently abstract events. When describing events with durative aspect, the Chinese speakers made static gestures with long duration to profile the maintenance of certain core features of actions in certain situations or the resultant state of certain actions. However, they tended to make dynamic gestures accompanying such durative aspect constructions as \( V \) zhe \( NP \) (verbs with high dynamicity), \( V \) zhe \( V \) zhe (\( V \) is reduplicated), and ná zhe \( V \) (hold/carry zhe \( V \)). Most of the gestures accompanying either the progressive aspect or the durative aspect speech focused on the internal core structures of events. The multimodal results provide new evidence for prototypical meanings of the two different imperfective aspect markers: the progressive aspect zài focuses on the internal stages of non-static situations and has a dynamic conceptual meaning, whereas the durative aspect zhe can focus not only on the internal stages of non-static situations but also on the external stages, like the resultant stages of static situations. It thus has dynamic as well as static conceptual meaning potentials, depending on the linguistic contexts in which it occurs. The findings are worth discussing in terms of the following two points.

5.4.1 Pure function of “zài \( V \)” construction vs. Multi-functions of “\( V \) zhe”

A lot of grammar book and other researchers have pointed out that the functions of two morphological forms –zài and zhe– are the same when it comes to expressing progressive aspect. Both of them denote that actions are ongoing (e.g. Chao 1979; Wang 2011, 2012; Zhu 1982). However, our multimodal data reveal that there are clear divisions of labor between zài and zhe, though they also have some overlapping functions. Both the linguistic and gestural evidence showed that the progressive aspect construction zài \( V \) has a single function of denoting an ongoing event. It profiles the
dynamic internal structure of an event. Such high dynamicity in conceptualization requires the use of activity verbs expressing strong action in speech and motivates the continuously dynamic movement in the co-speech gestures. In contrast, the $V \text{ zhe}$ construction has multiple functions, resulting in the complexity of its linguistic and gestural behaviors.

Firstly, when it co-occurs with the action verbs with strong dynamicity, the function of $V \text{ zhe}$ overlaps with that of $\text{zài} \ V$, focusing on the dynamic internal structure of an event and denoting its ongoingness without interruption. Accordingly, co-speech gesture produced with it also displays high dynamicity.

Secondly, the $V \text{ zhe}$ construction has the function of “zooming in” and focusing on the “frozen part” of the internal structure of an event, such as $\text{bào zhe}$ (keep holding/carrying something). Meanwhile, the static gestures accompanying the $\text{zhe}$-framed utterances in such circumstances were usually sustained for a very long time, even extending over several following utterances. The stativity of such gestures may provide a possible explanation for the reason why Chinese linguists termed $\text{zhe}$ a durative aspect particle, though both the $\text{zài}$- and $\text{zhe}$-framed utterances encode durative situations.

Thirdly, the $V \text{ zhe}$ construction has the function of focusing on the external stage of an event – the resultant state beyond the completion of an event, such as the final state of being fat or swollen. This is consistent with the recent empirical results of research on mental simulation in language comprehension that English progressive aspect not only highlights the internal structure of an event, but also the final physical state (Liu & Bergen 2016: 193). Such static gestures co-occurring with the durative $\text{zhe}$ aspect in such situations may be motivated by the mental simulation of resultant states acquired from a dynamic event. This would be consistent with the attention phenomenon of “endpoint focus”, known from the early work in cognitive semantics (e.g., Lakoff 1988).

Fourthly, the $V \text{ zhe}$ construction has a backgrounding function, so the maintenance of action is usually embedded in the simultaneous ongoing event. Accordingly, the gestures accompanying $\text{zhe}$ aspect-marked utterances in such cases had the dynamic
features found also in those with zài aspect-marked utterances. However, they had some different motivations behind them. In most cases, the non-stativity (i.e. dynamicity) and continuousness of the second event rather than the zhe-framed event in the serial verb constructions as nà zhe V motivated the dynamic expression of such complex events, realized in the gestures. The durativity of the zhe-framed event was backgrounded in the face of dynamicity of the following event. This revealed that when two events occurred simultaneously, the function of V zhe, like ná zhe V, was to provide the background information to facilitate foregrounding of the second event (Duncan 2002; Li 2014).

Smith (1991: 363) proposed that “zhe is changing from a stative imperfective to a general imperfective, gradually taking over the function of zài”. We do not agree with this proposition. It seems to be more reasonable to infer that the expansion of zhe’s functions promoted the occurrence of zài as a progressive aspect marker. Diachronically, only zhe has long been grammatically used to represent the events that were in progress (Zuo 2007); gradually, its semantic function extended to focus on the resultant state, and even the modal domain such as the use in imperative contexts (Chao 1979). Thus, the functional indeterminacy of the V zhe construction may have triggered the development of the zài V construction to focus on the pure progressivity of an event.

5.4.2 Embodied grammatical meaning

As an “aspect-prominent” language, Mandarin Chinese has grammaticalized aspeficial distinctions to mark dynamic durative events and static durative events. This study revealed that the grammatical meanings of the imperfective aspeficial constructions are also externally embodied in co-speech gestures. The gestures accompanying the progressive aspect-marked utterances dynamically enacted out the durative actions of the concrete and metaphoric events. In addition, the “back and forth” or cyclic gestures accompanying the progressive aspect-marked utterances also dynamically enacted out the schematic feature of an inherently abstract event in progress. The static durative events were grammatically marked by the imperfective aspect marker zhe and
accompanied by the static gestures that were maintained for a long time. When *zhe* encoded dynamic durative events, the accompanying gestures also displayed high dynamicity. The strong gestural sensitivity to the different aspectual categories and their different uses should result from the underlying mechanisms by which such grammatical categories affect our thinking for speaking and gesturing (Cienki & Müller 2008b; Slobin 1996). The schematic meanings of dynamicity and stativity manufactured by the gestures are an externalization or visualization of the speakers’ conceptualizations (i.e. mental structures and processes) of events (McNeill 1992). Moreover, dynamicity and stativity represented by the gestures derive from the ways in which hands make practical actions while experiencing concrete events in the world, as Streeck (2009) has claimed. The proposal here is that the schematic meaning (dynamicity or stativity) residing in the construal of events affects what grammatical aspectual perspective the speakers take and how the co-speech gestures are produced, though what the gestures act out is determined by the specific conceptual content inherent in the content verbs. This is also the case in the mental simulation in language comprehension: the lexicon determines what to mentally simulate, whereas grammatical categories, such as aspect, modulate how one simulates events that were described verbally (Bergen & Wheeler 2010).

Our findings provide empirical evidence from the perspective of grammatical conceptualization for the proposal that speech and gesture production are coordinated, forming an integrated system (e.g., McNeill 1992, 2005; Kendon 2004). The aspect-marked linguistic forms and their co-speech gestures were co-equal and co-expressive of the same aspectual information in different events. Temporally, all the strokes or stroke holds of the co-speech gestures overlapped with the imperfective aspect constructions, though some of them occurred before them. Semantically, the gestures accompanying the imperfective aspect markers *zài* also presented the same or closely related aspectual meanings in speech. For the same morphological item *zhe*, when it occurred in different linguistic contexts expressing different aspectual meanings, the gestural behavior adjusted with it accordingly. The same underlying conceptual potentials of the imperfective constructions (i.e. dynamicity or stativity) can be argued
to have directed the speakers’ speaking and gesturing. The semantic synchronization between the two forms of imperfective aspect-marked speech and their co-occurring gestures can be seen as orchestrated by the single integrated speech-gesture system.