Journal of Neurolinguistics 47 (2018) 105–122

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling

Semantic processing of self-adaptors, emblems, and iconic gestures: An ERP study

Kawai Chui[a,b,c], Chia-Ying Lee[b,d,e,f,*], Kanyu Yeh[g], Pei-Chun Chao[d,e]

[a] Department of English, National Chengchi University, Taipei, Taiwan, ROC
[b] Research Center for Mind, Brain and Learning. National Chengchi University, Taipei, Taiwan, ROC
[c] Research Center for Chinese Cultural Subjectivity, National Chengchi University, Taipei, Taiwan, ROC
[d] Institute of Linguistics, Academia Sinica, Taipei, Taiwan, ROC
[e] Institute of Neuroscience, National Yang-Ming University, Taipei, Taiwan, ROC
[f] Institute of Cognitive Neuroscience, National Central University, Taipei, Taiwan, ROC
[g] Graduate Institute of Linguistics, National Chengchi University, Taipei, Taiwan, ROC

ABSTRACT

The study investigates how the brain processes self-adaptors, semantically-unrelated emblems, and iconic gestures along with speech. The three types of gestures give rise to a continuum of semantic distinctions in relation to the accompanying speech. The overall N400 component occurred between 500 and 800 msec after the simultaneous gesture and speech onsets. In comparison to the speech-only condition, the reduced N400 evidenced the facilitation effect of iconic gestures at the centro-parietal sites. The meaningful yet non-speech-related emblems elicited enhanced N400 s at the left frontal-parietal sites; the meaningless self-adaptors produced the largest N400 effect over the scalp at the frontal-parietal sites. Self-adaptors had produced a larger negativity of N400 than emblems did at the centro-parietal regions. The results evidence the automatic integration of gesture and speech, and the diverse influence of gesture on processing. Only iconic gestures facilitate the semantic integration with speech. For a linguistic meaning to integrate with a semantically-unrelated emblem is less effortful than with a self-adaptor, suggesting that the processing of meanings proceeds more readily than the processing of a meaningless gesture occurring at the same time with speech.

1. Introduction

When people engage in a conversational talk, the use of hands and arms along with speech is prevalent and indispensable (Clark, 1996; Kendon, 2004; McNeill, 1992, 2005). Various types of hand-and-arm movements can co-occur with speech in daily communication, among which the current study focuses on self-adaptors, emblems, and iconic gestures. First, self-adaptors are self-touching movements such as scratching oneself on one's arm or removing something that got into one's eyes. In Example 1, the participants are talking about Koreans being good at marketing their tourism business. The female speaker sitting in the middle of the trio first raises her right hand to head level (the left frame in Fig. 1). While uttering the agreement marker *duì-a* 'right', the first person pronominal *wǒ*, the adverbial *jiù*, and the mental predicate *xiǎng* 'think' in Line 1, the speaker sweeps her bangs sideward to the left, as shown in the middle frame. Afterward, the hand returns onto the right knee (the right frame). The sweeping gesture does not convey any information in the interaction. In some situations, self-adaptors can be meaningful. For instance, a simple touch of the nose may mean

* Corresponding author. No.128, Sec. 2, Academia Rd, Nangang Dist, Taipei City 115, Taiwan, ROC.
E-mail address: chiaying@gate.sinica.edu.tw (C.-Y. Lee).

https://doi.org/10.1016/j.jneuroling.2018.04.004
Received 11 September 2017; Received in revised form 17 April 2018; Accepted 18 April 2018
Available online 10 May 2018
0911-6044/

An instance of a self-adaptor – the speaker in the middle sweeping her bangs (middle frame).

Fig. 1.

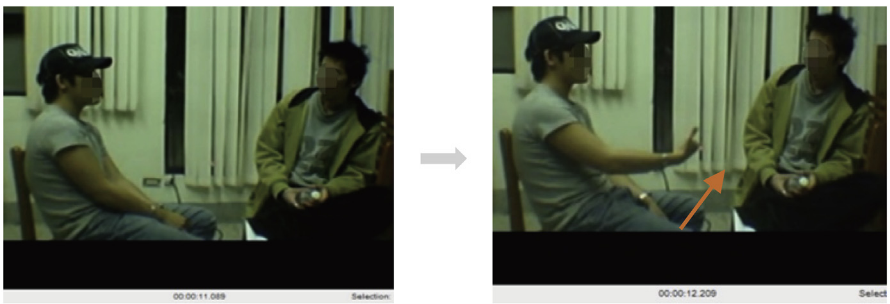
(1) 1 M: *...**duì** **ā...** **wǒ** **jiù** **xiǎng** shuō... mh...tamen zhēnde hěn huì*
right PRT 1SG then think COMPL PF 3PL really very be.good at
'Right, I then think that…mh…they are really very good at'

2 *xíngxiāo tāmen de nàge...guānguāng yǎnyì zhèyang*
market 3PL POSS that tourism entertainment like this
'marketing their tourism and entertainments.'

that someone is not being truthful. Such reliance on context for processing would require the availability of pragmatic or world knowledge. In the ERP studies of gesture, self-adaptors have rarely been of main concern, but they are a focus of interest in the present study, representing a type of gesture that does not bear any semantic relationship with the clausal utterances in the experimental trials. The study of self-touching acts which lack meaning is particularly pertinent to understand whether the integration of gesture and speech is automatic, and whether and to what extent gestures without meaning affect integration.

The second type of manual movement is that of emblematic gesture. Emblems have socially agreed-upon standards of well-formedness. There can be more than one gestural configuration for a particular emblem within a culture, yet each configuration still has a standard of well-formedness, and it is produced in much the same way across the users of a particular language, such as raising a hand and moving it from side to side for 'goodbye'. Gestures of this type can be produced without speech and convey meanings which may or may not have a semantic relationship with speech (Ekman & Friesen, 1969; Goldin-Meadow, 1999; Kendon, 1984, 1995, 2004; McNeill, 1992, 2005). In Example 2 about having one's hair dyed, at the time that the speaker on the left produces the conjunctive *suǒyǐ* 'so', his right hand starts to rise from the right thigh. During the uttering of the clause about the addressee on the right being dark-haired – *nǐ* 'you' *hái* 'still' *shì* 'copular' *hēifǎ* 'dark-haired', his right thumb extends and forms the thumbs-up gesture. See Fig. 2. The emblem expresses the speaker's unuttered positive evaluation of the content of the entire utterance. This thumbs-up gesture provides an example of a semantically-related emblem, the meaning of which can readily be integrated with the meaning of the sentence. This type of stimulus is not employed in the current study. The emblematic gestures presented to the participants in the experiment do not have a semantic relationship with the target words in the sentences. They are termed as 'semantically-unrelated emblems'.

The speech-related iconic gesture is the third type of manual movement for the study. Different from the case with the semantically-unrelated emblems, this type of gestures holds a semantic relationship with the target words in the sentences. Very often, speech and gesture present different pictures of the same event. McNeill's example is: At the moment the speaker utters 'she chases him out again', "speech conveys the ideas of pursuit and recurrence while gesture conveys the weapon used (an umbrella)" (McNeill, 1992, p. 13). In other words, the form of the speaker's hand in a shape as though to grip something is not linguistically expressed; the



Thumbs-up emblem (right frame).

Fig. 2. (2) M: *..suǒyǐ wǒ juédé **nǐ** **hái** **shì** **hēifà***
so ISG think 2SG still COP dark-haired
'So, I think you are still dark-haired.'

An iconic gesture depicting playing the music (right frame).

Fig. 3. (3) 1 M: *..jiùshì zài... pub lǐmiàn.. ránhòu yǒu nà zhǒng* ***xiànchǎng-yuèduì***

that is in pub inside then there.be that kind live-band

'That is, in the pub, there is a live band.'

utterance is structurally and semantically complete without the gesture. For gesture and speech to be complementary in conveying related, but non-redundant semantic information about the same referent can be illustrated in Example 3. The speaker on the left is talking about a pub scene in a movie, where there is a live band. At the time he produces the demonstrative *nà-zhǒng* 'that-kind', his hands come together in front of the chest. While uttering the nominal referent *xiànchǎng* 'live' *yuèduì* 'band', the fingers of his hands move up and down for the depiction of pressing the keys on a musical instrument. See Fig. 3. In this instance, the clausal utterance states the existence of a live band in the pub, whereas the accompanying iconic gesture depicts the unuttered meaning about what the live band does - the playing of a keyboard musical instrument. While the gestural information is associated with the lexical concept 'live band', both the gestural and linguistic knowledge are stored in the same semantic memory (Kutas & Federmeier, 2000). The present study investigates the complementary type of iconic gestures conveying semantic information which is related to the associated lexical referents but not overtly expressed in words in the experimental trials.

Self-adaptors, emblems, and iconic gestures are all common manual movements in interaction (Goldin-Meadow, 1999; Kendon, 2004; McNeill, 1992, 2000). In Chinese discourse, among 15 conversational excerpts from the NCCU (National Chengchi University) Corpus of Spoken Mandarin (Chui & Lai, 2008; Chui, Lai, & Chan, 2017),[1] totaling about 300 min of talk produced by 22 female and 11 male speakers (age range = 21–29; mean age = 24.5 years), 1,980 instances of gesture were analyzed. Aside from the pointing gestures, metaphoric gestures, and object-adaptors, 806 instances are self-adaptors, emblems, and iconic gestures. Out of the 806 instances, 425 iconic gestures constitute the majority (52.7%), while the 187 emblems (23.2%) and 194 self-adaptors (24.1%) still occupy substantial proportions of the total. The manual acts are mostly conspicuous and noticeable, and produced in the center space (see the Gesture Space in McNeill, 1992). Despite the pervasiveness of all three types of gesture, the past electrophysiological studies of the processing of gesture have scarcely considered the use of emblems and self-adaptors. Much of the ERP research has focused on iconic gestures (Holle & Gunter, 2007; Kelly, Kravitz, & Hopkins, 2004; Kelly, Ward, Creigh, & Bartolotti, 2007; Kelly, Creigh, & Bartolotti, 2010; Özyürek, Willems, Kita, & Hagoort, 2007; Wu & Coulson, 2005, 2007, 2011; Yap, So, Melvin; Yap, Tan, & Teoh, 2011; also see the reviews in; Özyürek, 2014 and in; Marstaller & Burianová, 2014). Here, self-adaptors, semantically-unrelated emblems, and iconic gestures are considered together on an equal basis for the understanding of the semantic processing of the three types of gesture along with speech.

1.1. ERP studies of the semantic processing of gesture and speech

The previous ERP studies of the neural processing of iconic gestures along with speech have mainly been concerned with semantic (in)congruence or violation of meaning. In Kelly et al. (2004), the study considered gesture and speech expressing identical meanings under a matching condition, gesture and speech conveying opposite meanings under a mismatching condition, and gesture and speech holding different but not contradictory meanings under a complementary condition. In another study, Kelly et al. (2010) investigated the extent of the speech-gesture integration in a Stroop-like task by manipulating the gender of the speaker and that of the gesturer, and the congruence between gesture and speech. Wu and Coulson (2005) examined the speech-gesture integration by taking context into account. The semantic relation between a gesture and a preceding action in a cartoon video clip was either congruous or incongruous. In another study, rather than separating speech and gesture, Wu and Coulson (2007) synchronized the two modalities for the investigation of the complementary type of semantic relation. In Wu and Coulson (2011), the former cartoon paradigm was used to examine static and dynamic gestures by pairing cartoon contexts with the following congruent and incongruent

[1] Permissions were obtained from all of the participants to use the audio-visual data from The NCCU Corpus of Spoken Mandarin for research. The data can be accessed online at TalkBank http://talkbank.org/access/CABank/TaiwanMandarin.html.

gestures of either type. Özyürek et al.'s (2007) study involved simultaneous gesture and speech in a sentence. They manipulated linguistic and gestural meanings which were either matched or mismatched within the sentence, and with either the linguistic or gestural meaning corresponding or not corresponding to the preceding context. Holle and Gunter (2007), for the purpose of investigating the role of gesture in speech disambiguation, focused on a moderate degree of semantic overlap between the two modalities. The complementary relation was paired with the dominant or subordinate meaning of the target word in the following sentence.

In all of the past studies, the negative-going N400 component was generally observed in cases of mismatched or incongruent conditions with semantic violation. Nonetheless, there was divergence regarding the detailed neural processing of iconic gestures. In Kelly et al. (2004), the semantic mismatch between the two modalities produced early P1−N1 and P2 effects followed by a larger negativity at 400 msec. They thus claimed that the semantic integration might start at the early sensory or phonological stages, and later at the semantic processing stage. However, while early effects were found for the complementary gestures, the N400 effect was only elicited in cases of semantic violation. Wu and Coulson (2005, 2011) found that incongruous gestures without speech showed an ERP effect at N450. In their later study (Wu & Coulson, 2007), congruous complementary gestures elicited smaller N300 and N400 than the amplitude in the speech-only trials. In Özyürek et al. (2007), both a speech-gesture mismatch and a mismatch with the preceding sentence produced a negative-going peak around 480 msec. The epoch between 350 and 550 msec was thus considered as the latency window for the cross-modal integration. Holle and Gunter (2007) found that a gesture that occurred in a preceding context could affect the processing of a target word in the following sentence. Incongruent gestures, again, elicited an N400 response.

In the literature, other factors have also been proposed to affect the linguistic-gestural integration besides semantic (in)congruence, including the perceived intentional coupling of gesture and speech (Kelly et al., 2007), the direction of the speaker's gaze (Straube, Green, Jansen, Chatterjee, & Kircher, 2010), and social eye gaze (Holler, Kelly, Hagoort, & Özyürek, 2012; Holler et al., 2014; see the recent reviews in; Obermeier & Gunter, 2015). In brief, these findings demonstrate that the integration of iconic gesture and speech can be modulated. Thus far, the early effects for automatic integration were only found in Kelly et al. (2004). Whether there might be different levels of semantic processing awaits further verification.

For the other two types of gesture – self-adaptors and emblems, while these two non-speech-related types of gesture co-occur with speech as commonly as iconic gestures do in everyday conversation, only a small amount of attention has been paid to understanding the neural processing of these gestures through the use of electrophysiological methodology in research on gesture. Gunter and Bach's (2004) study consisted of emblematic gestures and meaningless hand postures in pictures without speech. The ERPs elicited by the meaningless hand postures showed a right-frontal N300 and a centro-parietal N400, whereas the meaningful emblems had an earlier N300 effect and a reduced N400 with a posterior scalp distribution. The remaining questions are whether dynamic emblems would be processed similarly, and whether the processing of emblems would differ from that of the other gestural types, such as iconic gestures which also produced a reduced N400 under the congruent condition. In Proverbio, Gabaro, Orlandi, and Zani (2015), the stimuli were pictures of actors showing deictic, transformative, metaphoric, emblematic, iconic, and motor gestures along with facial expressions and verbal descriptions of the gestures, each preceded by a congruent or incongruent verbal description. Incongruent gestures, including emblems, elicited an anterior negativity N400 effect. Again, whether and to what extent the various types of incongruent gestures differed from each other in processing is still little known. These two studies examined emblems by the use of static pictures for semantic judgement, but there were confounding results with regard to the scalp distribution.

In regard to self-adaptors, they were included in one of the experiments on speech disambiguation in Holle and Gunter (2007). The results showed that both self-adaptors and iconic gestures conveying dominant meanings produced a larger N400 effect while processing a target word that expressed subordinate meaning. The study also found that the self-touching gestures weakened the impact of gesture on speech disambiguation, suggesting that the cross-modal integration can be modulated by the proportion of meaningful and meaningless gestures in the experiment. The results are further attested in Obermeier, Kelly, and Gunter (2015) which employed the same experimental materials as in Holle and Gunter (2007) to investigate the effect of a speaker's gesture style on speech disambiguation. Still, the nature of processing self-adaptors in a non-task-oriented situation is not well understood.

In summary, the former ERP studies of gesture have mainly focused on the (in)congruence between gesture and speech, the influence of gesture in disambiguation, and the effect of context on the processing of gesture and speech. How gestures themselves are processed naturally without a task to perform is little known, and still less is known as to what extent the different types of gesture bearing various kinds of semantic relationship with speech influence the linguistic-gestural integration.

1.2. The present study

For the uncovering of the neural mechanisms of language and gesture, it is foremost to first understand how the brain works to process different types of gesture occurring at the same time with a linguistic meaning. The present study thus aims to carry out an ERP experiment on the semantic processing of self-adaptors, semantically-unrelated emblems, and iconic gestures. Iconic gestures have a semantic relationship with target words, but self-adaptors and semantically-unrelated emblems do not. The latter differ from the former in conveying meanings. Moreover, whether a gestural instance is an iconic gesture, a semantically-unrelated emblem, or a self-adaptor is not determined in isolation, but in relation to the target word which bears the semantic load of the clausal utterance in the experiment (see Section 2.2. for details).

The three types of gesture are optimal to form a continuum of semantic distinctions. At one end are the self-adaptors lack of meaning in the clause; at the other end are the iconic gestures bearing speech-related, complementary meanings about the usual activities that people do with the concrete entities linguistically expressed by the target words; in between are semantically-unrelated emblems carrying non-speech-related meanings. The naturally occurring emblems, like the thumbs-up gesture in Fig. 2, are related to

Table 1
Types of gesture and semantic distinctions.

	Meaningful	Speech-related
Self-adaptors	–	–
Semantically-unrelated emblems	+	–
Iconic gestures	+	+

the speech. This is not the type of stimulus used in the current study, in that emblems are not related to the target words in the experiment. The semantically-unrelated emblems represent a type of cross-modal semantic distinction, as different from the self-adaptors and iconic gestures, based on whether a gesture carries a meaning vis-à-vis a target word, and whether the gestural meaning is related to the meaning of the target word. See Table 1.

Self-adaptors are particularly advantageous "to test whether gesture–speech integration is a primarily automatic process" (Holle & Gunter, 2007, p. 1186). In that the N400 component has been widely recognized as an index of the semantic processing of linguistic and non-linguistic stimuli, and also as an index of gesture-speech integration, if the self-adaptor condition produces an N400 amplitude different from that elicited by the speech-only trials, then the self-touching movements are automatically integrated with speech for semantic processing.

In order to obtain the natural neural responses of the participants, the present study diverges from the past studies in several ways. First, to compare the ERPs elicited by self-adaptors, semantically-unrelated emblems, and iconic gestures, the linguistic-gestural stimuli are presented in the same type of clausal structure, consisting of a light subject and a light verb. The last target word bears the semantic load. Second, to minimize the participants' recognition of the occurrence of a target word and gesture in a particular type of clause and to ascertain that similar target gesture–speech combinations are not presented in close succession, the experiment employs six other types of clauses varying in transitivity and animacy of the subject, each clause having a lexical constituent accompanied by a gesture in the beginning, the middle, or the end of the clause. They are used as fillers which create a natural communicative situation where a wide variety of clauses and various kinds of gesture can be produced in any clausal positions. All of the linguistic-gestural materials are presented in a setting of face-to-face communication, as gesture and speech are addressed to a person who provides acknowledgement upon receiving a message. Finally, the participants are not requested to perform any specific task. With no instruction to pay attention to gesture, if the participants still incorporate gesture with speech, the neural manifestations can be considered as evidence for the automatic integration of gesture and speech (Kelly et al., 2010).

Concerning the temporal synchrony between the two modalities, Habets, Kita, Shao, Özyurek, and Hagoort (2011) have investigated how the synchronization of speech and gesture influenced semantic integration. They found that gesture onset synchronizing with or preceding speech by 160 msec was optimal for the integration, "because the interpretation of the gesture could be, to some extent, guided by the speech to increase the semantic congruity between the two modalities" (Habets et al., 2011, p. 1852). Similar results were found in Obermeier, Holle, and Gunter (2011) in that the interaction was automatic when speech and the gesture stroke, the phase that bears the main meaning part of the gesture (McNeill, 1992, 2005), were in synchrony. Later, Obermeier and Gunter (2015) proposed another range for automatic integration - the stroke preceding or following speech within ± 120 msec. In short, "the closer speech and gesture are temporally to each other …, the more likely they are to be integrated with each other" (Özyürek, 2014, p. 6). In this study, the stroke onset and the speech onset are simultaneous, as they often coincide in natural discourse. Among all the 425 iconic gestures in our Chinese conversational data, in 75.1% of the cases the stroke onset and the speech onset occur at the same time.

The ERPs are time-locked to the onset of the target word denoting a concrete entity at the end of a statement. The speech onset is also simultaneous with the onset of the gesture stroke, preceded by the preparation phase during which the hand(s) naturally move(s) from a resting position to the beginning location of the gesture stroke. After the stroke is the retraction phase during which the hand(s) return(s) to the original starting position. The EEG is recorded as the participants watch videos of a speaker making an utterance to an addressee face to face, or uttering and gesturing simultaneously.

The research questions of the study are: Does the neural processing of information from the single modality of speech vary with that from the two modalities? Is the integration of gesture and speech automatic? Does the semantic processing differ across the three types of gesture? To what extent, if any, do the three types of gesture affect the integration? In reference to the past studies involving iconic gestures that conveyed complementary information not linguistically expressed, the N400 component was found to be either similar to or smaller than the amplitude in the speech-only trials (Kelly et al., 2004; Wu & Coulson, 2007). It is thus hypothesized that gestures which contain information associated with speech will be processed in a way different from gestures not semantically related to speech. A similar or reduced N400 component will be expected in cases of complementary iconic gestures where there is no semantic incongruence or violation between gesture and speech. Semantically-unrelated emblems and self-adaptors, on the other hand, will be expected to produce enhanced N400 effects, because the presence or absence of information in the gesture will not align with the linguistic information. Moreover, in consideration of the amount of information for integration, it is hypothesized that the semantically-unrelated emblems will produce a larger N400 effect than self-adaptors, since the processing of two meanings, one from the emblem and one from speech, is expected to be more effortful than the processing of a linguistic meaning and a self-touching gesture that does not provide semantic information. By taking into account three frequently-occurring types of gesture in the study, the ERP measurements will help in the understanding of how gestures per se are processed in a natural communicative setting, whether the cross-modal integration is automatic or not, and whether and to what extent different types of gesture giving rise to various cross-modal semantic relations would influence the integration of gesture and speech. The findings, as a whole, provide details about the linguistic-gestural integration in semantic processing.

2. Methods

2.1. Participants

Twenty undergraduate students (10 males, 10 females) were recruited for the experiment at National Chengchi University, Taipei. Their ages ranged from 20 to 24, with a mean age of 21 years. They were all healthy right-handed Mandarin-speaking adults (mean laterality coefficient = 91.4; Oldfield, 1971) with no reported history of neurological impairment. All of them had normal or corrected-to-normal vision. The experiment was conducted with the understanding and written consent of each participant in accordance with the approval of the Ethical Committee of National Taiwan University. The participants were paid for their participation.

2.2. Materials

2.2.1. Linguistic materials

Short videos displaying statements with or without co-occurring gestures were presented in the experiment. The linguistic materials were 84 common Chinese nominals as target words. All of the target words were disyllabic and referred to concrete entities such as *jiētī* 'stairs', *dìngzǐ* 'nail', *bàozhǐ* 'newspaper', *xiàngjī* 'camera', *chuānghù* 'window', or *diànhuà* 'telephone'. Each target word was the last word of a simple statement consisting of a demonstrative *zhè* 'this', *nà* 'that', *zhèlǐ* 'here' or *nàlǐ* 'there' as subject and a presentative verb *yoǔ* '(there) be' or a copular verb *shì* 'is' as predicate. Each of the statements was of three words and consisted of four to five syllables in total (see Table 2). This type of clausal statement is the one best suited for the investigation of how gestures themselves are processed in that it provides a neutral context in which the target words bear the semantic load solely without any potential context effects. Last, each of the 84 nominals was paired with a self-adaptor, a semantically-unrelated emblem, and an iconic gesture.

2.2.2. Fillers

The same number of filler statements was used in the experiment to minimize the participants' recognition of the occurrence of the gesture-speech combinations in a particular type of clause, and to avoid a close succession of similar target gesture-speech combinations. Eighty-four filler statements were constructed, varying by the type of grammatical pattern, transitivity, and animacy of the subject. As shown in Table 3, there were six structural types, with 13–15 different statements for each type. Every statement contained three words.

Just as for the target statements, the filler statements were also paired with the three types of gesture. The occurrence of gesture was more diverse, being coupled with the subject, the adverbial, the main verb, or the complement at random. Thus, a close-to-real-life setting was created in that a wide variety of clauses is used, and various kinds of gesture are produced in various clausal positions for the elicitation of the natural processing of the speech and gesture by the participants.

2.2.3. Gestural materials

Three types of gesture were used to be paired with the target nominals. First, complementary iconic gestures expressed meanings that were related to the lexical referents, depicting what people usually do with the referents, such as the use of the hammering gesture for the object *dìngzǐ* 'nail' and the press-the-shutter gesture for *xiàngjī* 'camera' in Table 4. Although the gestural meanings were not linguistically encoded, it is part of the knowledge about the target lexical concepts in semantic memory. Furthermore, there is no fixed way for an iconic gesture to depict a meaning. For instance, to configure the meaning of the pressing of the camera shutter, the speaker in Table 4 has both hands form a shape as if holding a camera, and her right index finger goes down as if pressing the shutter. The same meaning can just as well be depicted by one hand with the extended index finger going up and then down one time. Further, the same gesture of moving the index finger up and down is not necessarily produced just for 'press the shutter'; it can also be used to depict 'the clicking of a computer mouse'. Thus, a gestural form that is used as an iconic gesture does not carry one fixed meaning. In the study, whether a manual configuration was an iconic gesture or not was determined vis-à-vis the meaning of a target word which had the semantic load of the sentence.

Second, the semantically-unrelated emblems conveyed meanings that were not related to the target words. The emblems used in

Table 2
The structural patterns of experimental statements.

Subject	Verb	Target word
zhè 'this'	*shì* 'be'	*diànhuà* 'telephone'
nà 'that'	*yoǔ* '(there) be'	*dīngzǐ* 'nail'
zhèlǐ 'here'		*bàozhǐ* 'newspaper'
nàlǐ 'there'		*xiàngjī* 'camera'

Table 3
The structural patterns of filler statements.

The structural patterns of filler statements.

Structure*	Example
$S_{animate}$ V O	*mìshū ānpái xíngchéng* secretary arrange itinerary '(The) secretary arranges (the) itinerary.'
$S_{inanimate}$ V O	*dànǎo jiēshōu xùnxí* brain receive signal '(The) brain receives signals.'
$S_{animate}$ V V	*zuòzhě chǎngshì gǎibiàn* author try change '(The) author tries to change.'
$S_{inanimate}$ V V	*diànchí náqù huíshōu* battery take recycle 'Batteries are taken for recycling.'
$S_{animate}$ Adv V	*yóukè suíyì pāizhào* tourist freely take a photo 'Tourists take photos freely.'
$S_{inanimate}$ Adv V	*ménpiào jíjiāng màiwán* ticket soon sell out '(The) tickets were soon sold-out.'

*'S' as 'subject'; 'V' as 'verb'; 'O' as 'object'; 'Adv' as 'adverbial'

*'S' as 'subject'; 'V' as 'verb'; 'O' as 'object'; 'Adv' as 'adverbial'.

the study included gestures for the ideas of OKAY, THUMBS UP, THUMBS DOWN, VICTORY, NO, DON'T KNOW, ENUMERATION, MAKE A CALL, MONEY COUNT, QUOTES, WORSHIP, and CONGRATULATION. See the list in Table 5. The semantically-unrelated emblems here and the mismatched iconic gestures in other studies are alike in that the gestural meanings cannot be easily integrated with the speech. The differences between the two lie in whether there are socially agreed-upon standards of well-formedness and whether the gestures can be produced without speech. Iconic gestures are speech-related, and can be configured in one way or in another for a meaning. For instance, to depict the meaning of the playing of a keyboard musical instrument, the speaker in Fig. 3 has the fingers of his hands move up and down as though along a horizontal plane at waist height. The same meaning can as well be depicted with the palms of the hands facing each other, each slightly to the front of the body at each side, and the fingers moving up and down as if pressing the keys on a concertina. Thus, there is no one fixed way for an iconic gesture to depict a meaning. Further, the same gesture of moving the fingers up and down can mean 'key in information' for a computer, rather than 'play a keyboard musical instrument' in the case of the live band in Example 3. Thus, an iconic gesture does not carry a fixed meaning. In contrast, the semantically-unrelated emblems have socially agreed-upon standards of well-formedness, and can be produced without speech. Again, whether a gestural instance was a semantically-unrelated emblem or not was not determined in isolation, but in relation to a target word bearing the semantic load of the clausal statement.

Self-adaptors are hand movements without meaning in the clause. In this study, for the hand to touch one's hair, eyebrow, eye, cheek, nose, mouth, chin, earlobe, neck, shoulder, arm, and the clothes was paired with the targets at random. See the list of self-adaptors in Table 6, with the hand touching different parts of the body. Finally, for a gestural instance to be used as a self-adaptor was also determined vis-à-vis the meaning of a target word.

The semantic relationships between the target words and the three types of accompanying gestures were rated by 57 undergraduate students on a five-point Likert scale (0 = unrelated, 5 = most related). The ratings across gestural conditions showed significant differences (mean score 4.5 for iconic gestures, 1.1 for semantically-unrelated emblems, 1.1 for self-adaptors, $F_{2,249} = 7124$, $p < 0.001$).

Table 4
Examples of target statements and iconic gestures.
Examples of target statements and iconic gestures.

Target statement	Iconic gesture
nàlǐ yǒu dìngzǐ there be nail 'There is (a) nail.'	Hammer-something gesture  Left hand is in fist as if holding something; right hand in fist goes up and then down as if hammering something
zhè shì xiàngjī this COP camera 'This is (a) camera.'	Press-the-shutter gesture  Both hands configure a shape as if holding a camera; right index finger goes down as if pressing the shutter

2.2.4. Video clips

In this study, the production of gesture and speech was situated in a setting that very much resembled that of daily face-to-face interaction: A speaker and an addressee were seated face to face, with the speaker facing the video camera whereas the addressee's face was not seen in the clips. While the speaker was making an utterance, the addressee provided acknowledgement with a head nod. An actress in the role of the speaker was instructed to sit on a chair, utter verbal statements, and gesture with the hands and arms only without additional movements. She maintained a constant speech rhythm and hand-and-arm movements, suppressed salient facial expressions, and maintained eye-contact with the addressee. The addressee sat still across from the speaker and only provided acknowledgement with a head nod.

For the production of gesture, three phases were distinguished (McNeill, 1992, 2005): During the first 'preparation phase', the hand(s) move from a resting position to the beginning location of the gesture stroke. In the 'stroke phase', the main meaning part of the iconic and emblematic gestures is conveyed. After the stroke is the 'retraction phase' as the hand(s) return to the original starting position. In the study, full-fledged gestures were produced. The hands of the actress first left the resting position to go naturally to the location of the beginning of the gesture stroke. Then, the beginning of the stroke phase was synchronized with the onset of the target word. Finally, the hand(s) immediately returned onto the thigh(s) after the stroke phase. The addressee provided acknowledgement by the use of a head nod after the first syllable of a target word was produced for each of the statements.

There were four conditions with regard to the occurrence of gesture: 'speech-only', 'self-adaptor', 'semantically-unrelated emblem', and 'iconic gesture'.

1. Speech-only condition: Each statement was produced without the use of gesture. The actress put her hands on her thighs while uttering the whole clause *zhè* 'this' *shì* 'be' *xiàngliàn* 'necklace'. See (1) in Table 7.
2. Self-adaptor condition: Each statement was produced with the use of a self-adaptor. At the time speech began, the actress's hand(s) moved to the location of the beginning of the gesture stroke; when the last target word started to be uttered, the stroke was performed simultaneously, after which the hand(s) returned onto the thigh. Take *nàlǐ* 'there' *yoǔ* '(there) be' *jiētī* 'stairs' for instance (See (2) in Table 7). In the preparation phase, as the actress uttered the demonstrative *nàlǐ*, her right hand started to rise from the thigh to nose level. The stroke phase began with the production of the target word *jiētī*, as the actress's right fingers pinched the nose. Afterward, the hand returned onto the thigh.
3. Semantically-unrelated emblem condition: Each statement was produced with the use of an emblematic gesture. For the utterance *nà* 'that' *shì* 'be' *shuǐhú* 'water bottle', both hands started to separate in the preparation phase. Then, in the stroke phase, with the palms facing up, the hands moved to both sides of the body for the depiction of the idea 'I don't know'. Afterward, the hands returned onto the thighs. See (3) in Table 7.
4. Iconic-gesture condition: Each statement was produced with the use of an iconic gesture. Consider *zhèxiē* 'these' *shì* 'be' *lèsè* 'garbage'. In the preparation phase, the actress's right hand in fist moved up to chest level. As the target word *lèsè* was uttered, the hand swept rightward and downward as if throwing the garbage away. Then, the hand returned onto the thigh. See (4) in Table 7.

Table 5
The list of semantically-unrelated emblems.

Type	Example	Type	Example
OKAY		ENUMERATION	
THUMBS UP		MAKE A CALL	
THUMBS DOWN		MONEY COUNT	
VICTORY		QUOTES	
NO		WORSHIP	
DON'T KNOW		CONGRATULATION	

The study consisted of 84 statements, each containing a target word. Each statement was either not accompanied by a gesture or was accompanied by one type of gesture. The statements were counter-balanced across the participants so that no participant saw the same statement twice. The pairing of the 84 statements with each of the four conditions yielded a total number of 336 video clips. They were then divided into four lists (see Table 8); each list consisted of the same 84 statements with the same number of each type of condition in each of the lists, but no statement appeared with the same type of condition as in another list. In the experiment, each participant saw all of the clips from one list, within which each target sentence appeared only once.[2]

The mean length of all of the video clips was 2110 msec. The mean lengths of the video clips for each of the conditions were 1405 msec for the speech-only condition, 2314 msec for the self-adaptor condition, 2259 msec for the semantically-unrelated emblem condition, and 2461 msec for the iconic-gesture condition. The speech-only videos were shorter in duration than the videos of the other conditions because it usually took more time for the speaker to depict an intended action that resembled the same action in daily life, either whether the speaker swept her bangs to the left side (Example 1), or the speaker extended his right thumb for configuring the thumbs-up gesture (Example 2), or the speaker moved the fingers of his hands up and down for the playing of a keyboard musical instrument (Example 3).

As to the three types of gesture conditions, the mean durations of the gestural strokes were 693 msec for self-adaptors, 684 msec

[2] Table 7 shows the various structural patterns of the sentences in the experiment, as indicted in Table 2. For each sentence, it was either not accompanied by a gesture or was accompanied by one type of gesture, and no participant saw the same sentence twice.

Table 6
The list of self-adaptors.

Type	Example	Type	Example
Eyebrow		Chin	
Eye		Hair	
Nose		Neck	
Mouth		Shoulder	
Face		Arm	
Earlobe		Clothes	

for semantically-unrelated emblems, and 863 msec for iconic gestures. The iconic gestures were longer in duration than the other two types of gesture because 30 out of the 84 iconic gestures required the repetition of a hand movement so as to resemble the actions being depicted. For example, to depict *chuānglián* 'blind', the speaker's hands went up and down alternately for two times to look like the action of the pulling of the strings of a blind in daily life. Only two instances of the self-adaptors involved repetition, with the scratching of the shoulder or the upper arm two times. Finally, 42 out of the total 84 emblems involved repetition, but 22 out of the 42, such as MONEY COUNT, involved finger rather than arm movements within a gestural space smaller than those for iconic gestures, thus yielding shorter time spans. In the study, the instances of gestural repetition were performed two times at most.

Regarding the mean lengths of the video clips, the durations across the four conditions reached the significant level ($F_{3,332} = 465.5, p < 0.001$). The post hoc analysis showed that the speech-only condition was different from the gesture conditions and that the iconic-gesture condition was different from the other conditions, but that there was no difference between the emblem and the self-adaptor conditions. As to the gestural strokes, the durations across the three gesture conditions also reached the significant level ($F_{2,249} = 24.922, p < 0.001$), and the post hoc analysis showed that the iconic-gesture condition was different from the other two conditions.

Table 7
Four experimental conditions with examples.

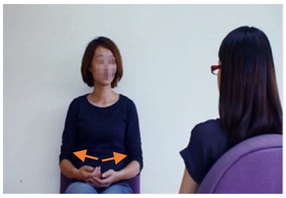
	Preparation	Stroke	Retraction
(1) Speech-only condition	 *zhè* 'this' *shì* 'be': No hand movement	 *xiàngliàn* 'necklace': No hand movement	 No hand movement
(2) Self-adaptor condition	 *nàlǐ* 'there' *yoǔ* '(there) be': Right hand rises from thigh to nose level.	 *jiētī* 'stairs': Right hand pinches the nose.	 Right hand returns onto thigh.
(3)Semantically-unrelated emblem condition	 *nà* 'that' *shì* 'be': Both hands move slightly sideward.	 *shuǐhú* 'water bottle': Hands move to sides to signify 'I don't know'.	 Hands return onto thigh.
(4) Iconic-gesture condition	 *zhèxiē* 'these' *shì* 'be': Right hand rises from thigh.	 *lèsè* 'garbage': Right hand sweeps rightward and downward as if throwing away garbage.	 Right hand returns onto thigh.

Table 8
The distribution of target sentences across the conditions in four lists.

	Speech-only condition	Self-adaptor condition	Semantically-unrelated emblem condition	Iconic-gesture condition	Total
List 1	21	21	21	21	84
List 2	21	21	21	21	84
List 3	21	21	21	21	84
List 4	21	21	21	21	84
Total	84	84	84	84	336

2.3. Experimental design and procedure

The materials were made into a within-subject design in which each participant watched videos of all four gestural conditions. Each video was presented one time. There were four lists of experimental trials, each including 21 trials for each of the four conditions. Eighty-four fillers were used in each list, resulting in a total of 168 trials per list. The trials in every list were pseudo-randomly divided into four blocks. Five comprehension questions were constructed for each block to establish the participants' concentration. In total, there were twenty questions, eight of which were constructed for the target trials. Six out of the eight questions were related to the accompanying iconic gestures. For example, the iconic gesture for this statement *zhèlǐ yǒu diànshì* 'Here is a TV set' was to depict selecting a TV channel by pressing a button on a remote control. The speaker in the video thus had the right hand closed and the extended thumb go down one time while uttering the last target word *diànshì*. After the video was played, a question with three choices appeared on the screen. It included the subject *dàjiā* 'everyone', the adverbial *jīngcháng* 'often', and a blank. The three choices

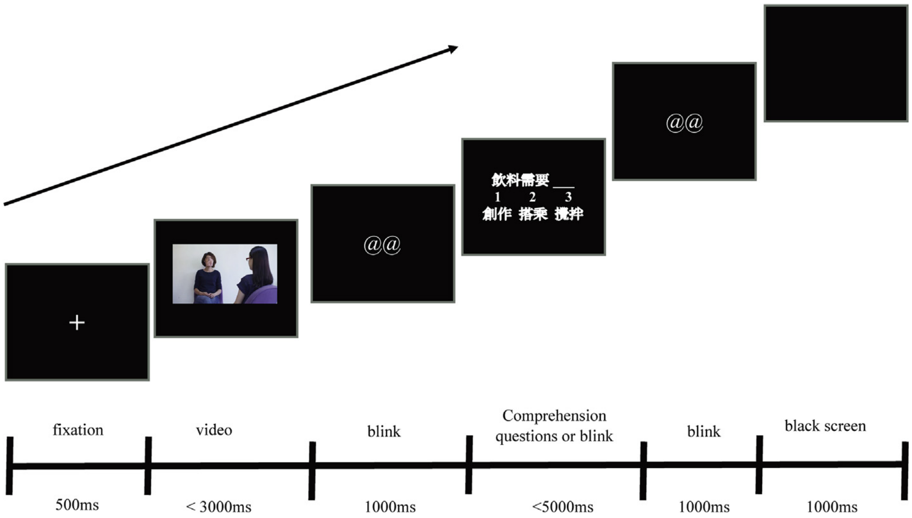


Fig. 4. Experimental procedure.

for the blank were: *zhuăntái* 'go to another channel', *chénmò* 'keep silent', and *mòshì* 'treat with indifference'. The first choice *zhuăntái*, being a lexical paraphrase of the select-a-TV-channel gesture, was considered as the best answer. When the same question with the same choices appeared in the case of the other three conditions, the target word *diànshì* was either not accompanied by a gesture or was accompanied by a self-adaptor or a semantically-unrelated emblem. These pairings did not suggest the best answer explicitly. The remaining two questions were related to the meaning of the target words. The best answers thus showed the semantic relationship between a target word and one of the three lexical choices. Before the experiment, the participants were informed that there would be questions about some clips in the form of a statement with a blank. The participant was required to select the best answer among three choices, one of which was a lexical paraphrase of the iconic gesture or showed a semantic relationship with the target word, by pressing a number on a keypad to fill in the blank.

The trials were programmed using the E-prime 2.0 software (Psychology Software Tools, Pittsburgh, PA, 2012). The stimuli were presented via a 17-inch computer monitor at a distance of 80 cm in front of the participants. The video clips were centered on a black background with a width of 24 cm and a height of 13 cm and played at 30 frames per second. Each trial began with a fixation cross for 500 msec, followed by the clips (mean length = 2110 msec). After the offset of the clips, an eye symbol appeared for 1000 msec. Then, either the eye symbol continued or a comprehension question was presented for 5 s. The participants pressed a button on a keyboard to answer the questions. A black screen appeared for 1000 msec after the participants had pressed the button. The average length of each trial was 1150 msec. See Fig. 4. The ERPs were time-locked to the onsets of the gestural stroke and the target word.

The participants took the experiment individually. They were led to an acoustically- and electrically-shielded room where the experimenters assisted them in putting on the EEG caps, after which an experimenter explained the procedure. The participants were asked to watch video clips, some of which were accompanied by questions about the content of the clips they had just watched without specific reference to gesture. The questions were answered by pressing one of the three buttons on a keypad. The participants were requested to stay still and only to blink during the interstimulus intervals when an eye symbol appeared on the screen or when responding to the questions.

The test session started with a block of 12 practice trials to familiarize the participants with the procedure, followed by 168 experimental trials in four blocks. After the practice block, the experimenter checked with the participants about the procedure to make sure they understood it clearly. Then, the participants started the experiments and finished the four blocks of experimental trials, with a 5-min break between blocks. The whole test lasted about 40 min.

2.4. EEG recording and data analysis

The electroencephalograms (EEG) were recorded using a 64-channel system from NeuroScan Inc. All electrodes were referenced online by default to the Reference electrode, and impedances maintained below 5 kΩ. The recording was also made at the right and left mastoids, and the EEG was re-referenced to the average of both mastoids off-line. Vertical and horizontal ocular movements (EOG) were monitored with electrodes placed below and above the left eye and at the right and left outer canthi, respectively. The EEG and the EOG recordings were sampled at 1000 Hz using a band pass filter of 0.15–30Hz.

For off-line analysis, the EEG signals were segmented into epochs which were time-locked to the onset of the target word from -100 to 1000 msec. Previous studies have used the empirical mode decomposition (EMD) or the ensemble empirical mode decomposition (EEMD) in ERP data analysis and have suggested that it can give better estimates of ERP latency and amplitude, increase the effect size, and improve the signal-to-noise ratio (SNR) between conditions (2015b, Al-Subari et al., 2015a; Chen, Chao, Chang, Hsu, & Lee, 2016; Cong et al., 2009; Hsu, Lee, & Liang, 2016; Kuo, Lin, Dressel, & Chiu, 2011; Lee, Chang, Hsieh, Deng, & Sun, 2012;

Williams, Nasuto, & Saddy, 2011; Wu et al., 2012). Thus, this study applied the analytic procedure-based EEMD analysis as described by Hsu et al. (2016) for data analysis. The EEMD analysis was performed with 10 times of sifting and 40 ensembles. The amplitude of the Gaussian noises added to a signal of each EEG trial of each channel was 10% amplitude of the standard deviation of the EEG signal. Each EEG segment was decomposed into nine intrinsic mode functions (IMFs) and one residual trend. Following Chen et al. (2016), we performed a summation of IMF 6, IMF7, and IMF 8 with the Hilbert spectra distributed between 0.5 and 7 Hz to extract the N400 signal. It was averaged across all of the trials for each condition at each electrode to obtain the event-related modes (ERMs) and to construct the original ERPs (Al-Subari, Al-Baddai, Tome, Volberg, et al., 2015a,b).

3. Results

3.1. Behavioral results

Each participant responded to twenty randomly appearing comprehension questions during the experiment. Concerning the six questions asked in the iconic-gesture trials, the accuracy rates in the iconic-gesture condition, the semantically-unrelated emblem

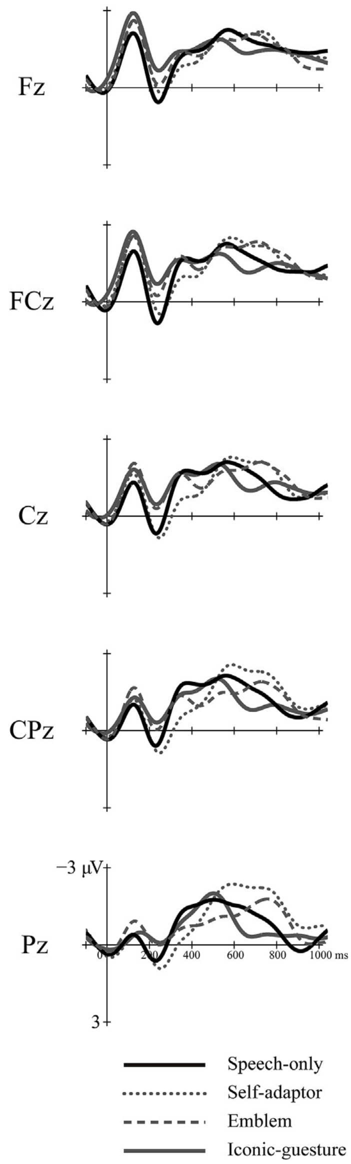


Fig. 5. The ERMs for the four conditions at central sites: Fz, FCz, Cz, CPz and Pz. Black solid lines represent the speech-only condition, grey dotted lines represent the self-adaptor condition, grey dashed lines represent the semantically-unrelated emblem condition, and grey solid lines represent the iconic-gesture condition.

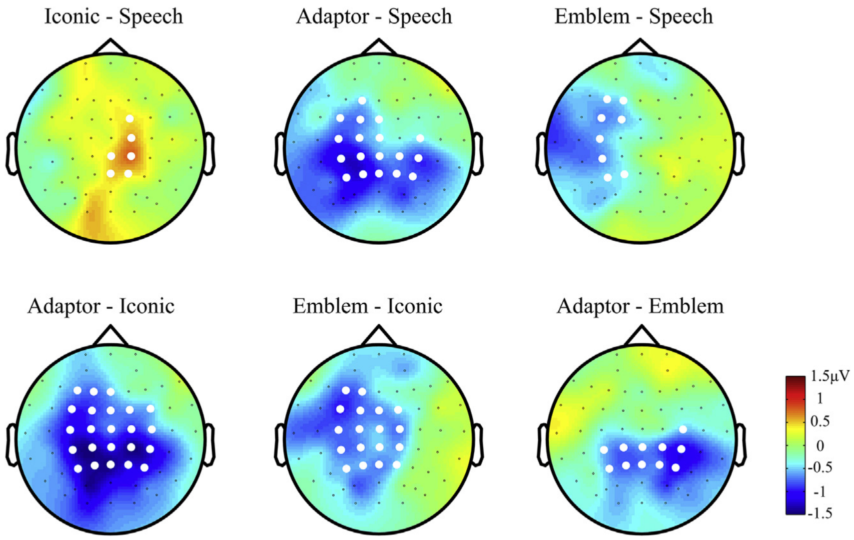


Fig. 6. Topographic maps for the differences of mean amplitude of N400 from 500 to 800 msec between gestures and speech (upper row) and among gestures (lower row). White dots represent the electrodes that showed significance for the contrasts ($p < 0.05$).

condition, the self-adaptor condition, and the speech-only condition were 82.8%, 76.7%, 60%, and 56.7% respectively. The behavioral performance of the participants indicates the attention of the participants to the presence and absence of hand movements.

3.2. ERP results

Fig. 5 presents the grand averaged ERMs of the four conditions at the midline electrodes. All of the conditions elicited typical brain responses for visual stimulation, including the N1, P2, and the negative-going wave (N400) from 400 to 800 msec. The effects of interests on N400 were measured in the time window from 500 to 800 msec over 25 electrodes (F3, F1, FZ, F2, F4, FC3, FC1, FCZ, FC2, FC4, C3, C1, CZ, C2, C4, CP3, CP1, CPZ, CP2, CP4, P3, P1, PZ, P2, P4). Repeated measures of ANOVA using the PROC GLM function in SAS were performed on the mean amplitudes of N400 with the gestural types (iconic gestures, semantically-unrelated emblems, self-adaptors, and speech-only) and electrode sites (25 electrodes) as the within-subject factors. For each ANOVA, the Geisser-Greenhouse adjustment to the degrees of freedom was applied to correct for violations of sphericity associated with repeated measures. For all F tests with more than one degree of freedom in the numerator, the corrected P-value is reported.

Planned comparisons using the SAS CONTRAST statement to reproduce the F-tests for effects were performed for the following hypotheses. First, to examine whether the information in gesture in alignment with the linguistic information would influence the integration of gesture and speech, the mean amplitudes of N400 of iconic-gestures, semantically-unrelated emblems, and self-adaptors were compared with those for the speech-only trials. Next, to investigate whether the semantic processing differs among the three types of gesture, comparisons were made among the three types of gestures. For each planned comparison, the F-value was adjusted and the corrected P-value was reported.

The ANOVA on N1 from 50 to 200 msec indicated no significant two-way interaction between the gestural types and the electrode sites ($F_{72,1368} = 1.73$, $p = 0.1002$) and no significant main effect for the gestural types ($F_{3,57} = 0.87$, $p = 0.4565$). The ANOVA on N400 revealed a significant two-way interaction between gestural types and electrode sites ($F_{72,1368} = 2.27$, $p = 0.0207$), while the main effect of gestural types was not significant ($F_{3,57} = 2.01$, $p = 0.1297$). Fig. 6 shows the results of the planned comparisons. For each planned comparison, the critical value of $F_{1,1368}$ at $p = 0.05$ is 3.85. The iconic-gesture condition elicited less negative N400 than the speech-only condition did on the central-parietal electrodes (the values of *F*s ranging from 4.16 to 15.58, ps $<$ .05). Both the self-adaptor and the semantically-unrelated emblem conditions elicited more negative N400 than the speech-only condition did on the frontal-parietal electrodes (the values of *F*s ranging from 3.87 to 36.01, ps $<$ .05). Additionally, both the self-adaptors and the semantically-unrelated emblems elicited more negative N400 than the iconic gestures did on almost every electrode (the values of *F*s ranging from 5.12 to 51.71, ps $<$ .05). Finally, the self-adaptors elicited a larger negativity of N400 than the semantically-unrelated emblems did on the central-parietal electrodes (the values of *F*s ranging from 6.33 to 29.32, ps $<$ .05).

4. Discussion

This study used event-related potentials to investigate the semantic processing of three types of frequently-occurring dynamic gestures on a continuum of semantic distinctions in relation to speech. A natural communicative setting was arranged, where the speaker and the addressee were seated face to face, acknowledgement was provided by the addressee while the speaker was making an utterance and gesturing, and the participants were not required to perform any particular task. The time-locked event was the speech onset of the target word in synchrony with the stroke onset of the accompanying gesture. Four conditions were involved, namely the speech-only condition, the self-adaptor condition, the semantically-unrelated emblem condition, and the iconic-gesture condition. N400 effects were found across the conditions within the latency window between 500 and 800 msec.

Broad temporal distributions of negativity for the N400 component have also been found in past studies. In Wu and Coulson (2005), the (in)congruency effects between gesture and speech were realized in the negative-going waveforms in the 600-900-msec window. It was 400–900 msec in Wu and Coulson (2011) with enhanced negativity for incongruent trials. In Habets et al. (2011), the broadly distributed negativity was between 300 and 900 msec. In the current study, the N400 amplitudes in the range of 500–800 msec differed across the various types of gesture. They also demonstrated a significant interaction effect between gestural types and electrode sites. As a recognized index of the semantic processing of linguistic and non-linguistic stimuli including gesture, the various N400 amplitudes provide evidence that the trials across the four conditions were processed variably in the brain. Specifically, the neural processing of information provided by way of a single modality was found to vary from that provided by way of two modalities. Since the participants were not required to take gesture into account, the electrophysiological responses distinguishing between the presence and absence of hand movements support the past findings that the two modalities are integrated, and that gesture affects language processing (Kelly et al., 2010; Kita & Özyürek, 2003; Özyürek & Kelly, 2007). Importantly, the current study provides neural evidence that the influence of gesture on the integration is diverse across types.

Dennett's (1987) notion of intentional stance can account for the diverseness of the cross-modal integration. According to this notion, humans naturally believe that other people's behavior is intentional and communicative, and such belief is used to understand and interpret language and behavior. The relationship between intention and the occurrence of gesture was attested in Melinger and Levelt (2004) in that speakers used gestures to convey part of their message on purpose. Kelly et al. (2007) manipulated the intentional relationship between gesture and speech, and found different N400 effects in the distinction between intentional and unintentional relations across modalities. "[I]t is possible that interlocutors in everyday discourse may naturally (and safely) assume that gestures that accompany speech are intended to go with that speech, and under normal circumstances, they may integrate speech and gesture by default" (Kelly et al., 2007, p. 231). In the present study, the setting in the video clips was established in a way very similar to that of real-life communication. Not only was the speaker talking and gesturing to an addressee, but the addressee also nodded for acknowledgement. The participants readily accepted that the setting represented a normal communicative situation, the person who uttered and produced a gesture in the video clips was a rational agent, and the speech and gesture were produced as intentional acts that conveyed meaning. In other words, gesture is intended to be integrated with the utterance, and this integration gives rise to the differences in ERP responses to self-adaptors, semantically-unrelated emblems, and iconic gestures with respect to magnitude and topographical distribution.

4.1. Iconic gestures

First, in accordance with our expectation, iconic gestures holding a high degree of semantic relatedness with the target words significantly elicited a smaller N400 component than the amplitude in the speech-only trials. The reduced amplitude supports the findings concerning the processing of complementary gestures in Kelly et al. (2004) and Wu and Coulson (2007). However, the early effects elicited by this type of iconic gestures in Kelly et al. (2004) were not attested here. Regarding the scalp topography, the reduced N400 amplitude in Wu and Coulson (2007) was largest over the anterior electrodes, but in the present study the effect was most prominent in the centro-parietal regions.

The lower amplitude of the N400 elicited by iconic gestures provides evidence that iconic gestures facilitate semantic integration. Applying the notion of intentional stance, the participants had expected that the production of gesture and speech was, by default, intentional and communicative, and that the gestural and linguistic entities were intended to be together. When they were produced together at the same time for semantic processing, a manual act that depicted an activity that people often perform with its associated target referent facilitated the integration with the linguistic information, since the gestural meaning was also about the same lexical concept and from the same semantic memory. The processing of information from both modalities hence proceeded readily.

4.2. Semantically-unrelated emblems

In contrast to the speech-related iconic gestures, the semantically-unrelated emblems, such as the gestures for victory, worship, and congratulation in Table 5, conveyed meanings that were not related to speech in the trials. The N400 amplitude for the emblems was more negative-going than the amplitude in the speech-only trials. This result does not align with Gunter and Bach (2004) investigation of emblems which were presented in pictures. The static emblems produced a prolonged N300 effect lasting through the N400 effect with a posterior scalp distribution, whereas in this study, the dynamic counterparts produced an N400 effect at the left frontal-parietal sites in comparison to the amplitude in the speech-only trials. The divergence in the results may be owing to the distinction between static and dynamic manual acts.

Contrary to the facilitation effect of the iconic gestures, the larger N400 amplitude for the semantically-unrelated emblems suggests effortful processing (Kutas & Federmeier, 2000; Obermeier & Gunter, 2015; Obermeier et al., 2011). Although the use of emblematic gestures was in alignment with the general expectation that they were intentional acts and communicative, the processing also took into account the semantic relationship between gesture and speech. In this study, the emblems held a low degree of semantic relatedness with speech, which was attested by the individuals' low ratings of the semantic relationship between the emblems and the target words. The integration of two unrelated meanings, one from an emblem and one from speech, was thus effortful.

4.3. Self-adaptors

Different from the meaningful iconic gestures and emblems, self-adaptors, such as the rubbing of one's eye or the scratching of oneself on the arm in Table 6, did not convey any meaning related to the clausal utterances in the trials. The self-touching movements elicited the largest amplitude of the N400 component, when compared to the amplitude in the speech-only trials. Similar results were found in Holle and Gunter (2007) and Obermeier et al. (2015). Though the dynamic self-adaptors in their studies did not co-occur with the target words, an enhanced negativity at N400 was elicited when a self-adaptor occurred in the preceding context and was followed by a target word conveying a subordinate meaning. Both the previous studies and this study provide ERP evidence that the processing of self-adaptors is effortful.

Concerning the scalp topography, the enlarged N400 component was distributed over the scalp at the frontal-parietal sites in this study, as compared to the amplitude for the speech-only condition. The result differs from Gunter and Bach's (2004) finding that the meaningless hand postures presented in pictures elicited a right frontal N300 and a centro-parietal N400. The inconsistent results, again, suggest that static and dynamic gestures can be distinct types of gesture.

The investigation of self-adaptors is useful for testing whether the integration of linguistic and gestural information is automatic. If this type of hand movement had been ignored because it did not convey meaning, the processing would not have been different from that for the speech-only trials at the least. Our finding that the ERP responses to self-adaptors were distinct shows that the neural processing had taken the gesture into account, regardless of its type and meaning. Applying Dennett's (1987) notion of intentional stance, the production of a self-adaptor was, by default, communicative and intended to go with speech. During semantic processing, the communicativeness of a self-touching act would need to rest upon the retrieval of other types of information such as pragmatic or world knowledge, hence resulting in effortful linguistic-gestural integration.

Finally, the expectation of this study that self-adaptors and semantically-unrelated emblems would produce enhanced N400 effects was supported. However, the fact that the N400 amplitude of self-adaptors was larger in the centro-parietal regions, as compared to emblems (Fig. 6), was not expected. The unexpectedness suggests that the processing of meanings, regardless of the semantic relatedness between the two modalities and the amount of information, proceeds more readily than that of a meaningless gesture occurring at the same time with a linguistic meaning.

All the findings together, the neural integration of gesture and speech is sensitive to the distinctions across self-adaptors, semantically-unrelated emblems, and iconic gestures. The lengths of the video clips and of the strokes should not have influenced the semantic processing since the differences in the duration between the self-adaptors and the semantically-unrelated emblems were not significant, but the differences in the ERP responses to these two types of gestures were significant. The larger and smaller N400 effects for the integration shed light on how the gesture-speech integration proceeds in semantic processing. That is, when a gesture occurs, regardless of type and meaning, it is considered as an intentional act and communicative, and integrates with speech automatically. During semantic processing, the brain is sensitive to the presence and absence of gesture. The influence of gesture varies across types and cross-modal semantic relationships. Only the speech-related iconic gestures facilitate the semantic integration with speech. For a linguistic meaning to integrate with a semantically-unrelated emblem is effortful. The integration with a gesture that lacks in meaning is even more effortful, as information other than that from the lexical constituent would need to be retrieved for integration.

5. Conclusion

The present study has investigated the neural processing of self-adaptors, semantically-unrelated emblems, and iconic gestures along with speech on an equal basis. The electrophysiological responses to the various types of semantic distinctions advance the current neuro-cognitive research on gesture and speech, and reveal more details about the influence of the under-studied types of gesture on the linguistic-gestural integration. The findings are a baseline for further investigation of other types of gestural relationship with speech which will be important to have more neural evidence for understanding the default integration of gesture and speech, the semantic effects on the integration, and the uncovering of the neural mechanisms of language and gesture.

Acknowledgements

This research was supported by grants from the Ministry of Science and Technology, Taiwan, ROC (MOST 103-2410-H-004 -180 -MY3).

We are very grateful to the anonymous reviewers for their helpful comments and suggestions on this article.

Appendix A. Speech transcription convention

Transcription of speech

[ ]	speech overlap
...(N)	long pause
...	medium pause
..	short pause

Appendix B. Abbreviations of linguistic terms

1SG	first person singular
2SG	second person singular
3PL	third person plural
COMPL	complementizer
COP	copular
POSS	possessive morpheme
PF	pause filler
PRT	particle

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jneuroling.2018.04.004.

References

Al-Subari, K., Al-Baddai, S., Tome, A. M., Goldhacker, M., Faltermeier, R., & Lang, E. W. (2015a). Emdlab: A toolbox for analysis of single-trial EEG dynamics using empirical mode decomposition. *Journal of Neuroscience Methods, 253*, 193–205.

Al-Subari, K., Al-Baddai, S., Tome, A. M., Volberg, G., Hammwohner, R., & Lang, E. W. (2015b). Ensemble empirical mode decomposition analysis of EEG data collected during a contour integration task. *PLoS One, 10*(4), e0119489.

Chen, W. F., Chao, P. C., Chang, Y. N., Hsu, C. H., & Lee, C. Y. (2016). Effects of orthographic consistency and homophone density on Chinese spoken word recognition. *Brain and Language, 157–158*, 51–62.

Chui, K., & Lai, H. L. (2008). The NCCU corpus of spoken Chinese: Mandarin, Hakka, and southern .in. *Taiwan Journal of Linguistics, 6*(2), 119–144.

Chui, K., Lai, H. L., & Chan, H. C. (2017). Taiwan spoken Chinese corpus. In R. Sybesma (Ed.). *Encyclopedia of Chinese language and linguistics* (pp. 257–259). Boston, MA: Brill.

Clark, H. H. (1996). *Using language.* Cambridge: Cambridge University Press.

Cong, F., Sipola, T., Huttunen-Scott, T., Xu, X., Ristaniemi, T., & Lyytinen, H. (2009). Hilbert-Huang versus Morlet wavelet transformation on mismatch negativity of children in uninterrupted sound paradigm. *Nonlinear Biomedical Physics, 3*(1), 1.

Dennett, D. (1987). *The intentional stance.* Cambridge, MA: MIT Press.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*(1), 49–98.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences, 3*(11), 419–429.

Gunter, T. C., & Bach, P. (2004). Communicating hands: ERPs elicited by meaningful symbolic hand postures. *Neuroscience Letters, 372*(1–2), 52–56.

Habets, B., Kita, S., Shao, Z., Özyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech–gesture integration during comprehension. *Journal of Cognitive Neuroscience, 23*(8), 1845–1854.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience, 19*(7), 1175–1192.

Holler, J., Kelly, S., Hagoort, P., & Özyürek, A. (2012). When gestures catch the eye: The influence of gaze direction on co-speech gesture comprehension in triadic communication. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.). *Proceedings of the 34th annual meeting of the cognitive science society (CogSci 2012* (pp. 467–472). Austin, TX: Cognitive Science Society.

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., & Özyürek, A. (2014). Social eye gaze modulates processing of speech and co-speech gesture. *Cognition, 133*(3), 692–697.

Hsu, C. H., Lee, C. Y., & Liang, W. K. (2016). An improved method for measuring mismatch negativity using ensemble empirical mode decomposition. *Journal of Neuroscience Methods, 264*, 78–85.

Kelly, S. D., Creigh, P., & Bartolotti, J. (2010). Integrating speech and iconic gestures in a stroop-like task: Evidence for automatic processing. *Journal of Cognitive Neuroscience, 22*(4), 683–694.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*(1), 253–260.

Kelly, S. D., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language, 101*(3), 222–233.

Kendon, A. (1984). Did gesture have the happiness to escape the curse at the confusion of babel? In A. Wolfgang (Ed.). *Nonverbal behavior: Perspectives, applications, intercultural insights* (pp. 75–114). Lewiston, NY: C.J. Hogrefe.

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of Pragmatics, 23*(3), 247–279.

Kendon, A. (2004). *Gesture: Visible action as utterance.* Cambridge: Cambridge University Press.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*(1), 16–32.

Kuo, C. C., Lin, W. S., Dressel, C. A., & Chiu, A. W. (2011). Classification of intended motor movement using surface EEG ensemble empirical mode decomposition. *Proceedings of the 2011 annual international conference of the IEEE engineering in medicine and biology society* (pp. 6281–6284). .

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463–470.

Lee, P. L., Chang, H. C., Hsieh, T. Y., Deng, H. T., & Sun, C. W. (2012). A brain-wave-actuated small robot car using ensemble empirical mode decomposition-based approach. *IEEE Transactions on Systems, Man, and Cybernetics - Part a: Systems and Humans, 42*(5), 1053–1064.

Marstaller, L., & Burianová, H. (2014). The multisensory perception of co-speech gestures – a review and meta-analysis of neuroimaging studies. *Journal of Neurolinguistics, 30*, 69–77.

McNeill, D. (1992). *Hand and mind-What gestures reveal about thought.* Chicago: University of Chicago Press.

McNeill, D. (2000). *Language and gesture.* Cambridge: Cambridge University Press.

McNeill, D. (2005). *Gesture and thought.* Chicago: University of Chicago Press.

Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture, 4*(2), 119–141.

Obermeier, C., & Gunter, T. C. (2015). Multisensory integration: The case of a time window of gesture-speech integration. *Journal of Cognitive Neuroscience, 27*(2), 292–307.

Obermeier, C., Holle, H., & Gunter, T. C. (2011). What iconic gesture fragments reveal about gesture-speech integration: When synchrony is lost, memory can help. *Journal of Cognitive Neuroscience, 23*(7), 1648–1663.

Obermeier, C., Kelly, S. D., & Gunter, T. C. (2015). A speaker's gesture style can affect language comprehension: ERP evidence from gesture-speech integration. *Social Cognitive and Affective Neuroscience, 10*(9), 1236–1243.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The edinburgh inventory. *Neuropsychologia, 9*(1), 97–113.

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651).

Özyürek, A., & Kelly, S. D. (2007). Gesture, brain, and language. *Brain and Language, 101*(3), 181–184.
Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience, 19*(4), 605–616.
Proverbio, A. M., Gabaro, V., Orlandi, A., & Zani, A. (2015). Semantic brain areas are involved in gesture comprehension: An electrical neuroimaging study. *Brain and Language, 147*, 30–40.
Psychology Software Tools, Inc [E-Prime 2.0]. (2012). Retrieved from http://www.pstnet.com.
Straube, B., Green, A., Jansen, A., Chatterjee, A., & Kircher, T. (2010). Social cues, mentalizing and the neural processing of speech accompanied by gestures. *Neuropsychologia, 48*(2), 382–393.
Williams, N., Nasuto, S. J., & Saddy, J. D. (2011). Evaluation of empirical mode decomposition for event-related potential analysis. *EURASIP Journal on Advances in Signal Processing, 2011*(1), 965237.
Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology, 42*(6), 654–667.
Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain and Language, 101*(3), 234–245.
Wu, Y. C., & Coulson, S. (2011). Are depictive gestures like pictures? Commonalities and differences in semantic processing. *Brain and Language, 119*(3), 184–195.
Wu, C. H., Lee, P. L., Shu, C. H., Yang, C. Y., Lo, M. T., Chang, C. Y., et al. (2012). Empirical mode decomposition-based approach for intertrial analysis of olfactory event-related potential features. *Chemosensory Perception, 5*(3–4), 280–291.
Yap, D.-F., So, W.-C., Melvin Yap, J.-M., Tan, Y.-Q., & Teoh, R.-L. S. (2011). Iconic gestures prime words. *Cognitive Science, 35*(1), 171–183.