International Journal of English Linguistics; Vol. 6, No. 2; 2016
ISSN 1923-869X E-ISSN 1923-8703
Published by Canadian Center of Science and Education

Production and Interaction between Gesture and Speech: A Review

Ying Gao[1], Yuqin Liu[1] & Chunyue Zhou[1]

[1] School of Software, Dalian University of Technology, Dalian, China

Correspondence: Ying Gao, School of Software, Dalian University of Technology, Tuqiang Street, No. 321, The Economic and Technological Development Zone, Dalian, China. Tel: 86-411-6227-4467. E-mail: gaoying004@dlut.edu.cn

Received: February 29, 2016 Accepted: March 20, 2016 Online Published: March 28, 2016

doi:10.5539/ijel.v6n2p131 URL: http://dx.doi.org/10.5539/ijel.v6n2p131

Abstract

Gesture in multimodal researches has been studied widely recently, and how gesture interacts with speech in communication is the focus in most researches. Some hypotheses or models about production and interaction between gesture and speech are introduced and compared in this paper. We find that it is generally agreed that speech production mechanism can be explained based on Levelt's Model; while there is no consistency about gesture production and the interaction between gesture and speech. Most of theories argue that gesture stems from the visual-spatial images in working memory; some models approve of the interactive relationship while others consider no interaction between gesture and speech. Further research will be made in the areas of theoretical and applicative aspects.

Keywords: gesture, speech, interaction, framework, models

1. Introduction

Messages can be expressed verbally or nonverbally. The most linguistic papers are about speech research generally. In recent years, there has been more and more papers about multi-modal messages, and which become the interest in some international conferences, as the GESPIN (gesture and speech in interaction) in France in 2015, GW (gesture workshop series) in UK in 2015, etc. The reasons for the increasing studies about gesture are because language or word information itself cannot deliver the entire information in communication. Gesture sometimes can represent certain information better than other ways, for example, hand gestures perform better than speech in terms of shapes. When speakers describe a grass, they can use hands to describe its shape and size and explain it is "like this big and like this in shape", avoiding the concrete spatial and dimensional words description. Speakers need use gesture to coordinate with speech in communication in most scenarios, for example, when hands are busy with some affairs, speech can express instead; while in some noisy situations, hand gestures can replace speech to deliver messages. One modal can help to eliminate the ambiguity of another modal or to emphasize a certain message. In this case, they have complementary functions in communication; while in other cases, they express the same meanings, and that means one modal is redundant. The interaction among modals is complex, the full explanation about communication need explicit illustration of these modals and their relationships. So the interaction between gesture and speech becomes the focus in this field.

Gestures are defined as kind of visible physical actions in communication (Kendon, 2004). Gesture is considered as an inseparable part in language system (McNeill, 2005), or as a multi-modal phenomenon (Cienki & Muller, 2008). Once identified, gestures can be classified along a number of dimensions, and these taxonomies are important in understanding the relationship between gesture and speech (Natasha Abner et al., 2015). McNeill (1992) uses the Kendon's continuum to differentiate them. With the degree of conventionalization, people accept sign language more because sign language has independent semantic contents, which consist of the communicative messages between or with deaf people. Sign language can develop without affiliating with speech. While some symbolized gestures (such as "thumb-up") are considered to have special language meanings (i.e., this gesture in most American and European cultures means approval of something, while it is a rude and offensive gesture in Islamic countries). Other gestures do not have standardized modes, and these gestures need coordinate with speech to complete communication. As for this kind of gestures, McNeill (2005) further suggests a complex of several continua to show their features (P Wagner et al., 2014):

(a) Continuum 1: relationship to speech (obligatory presence or obligatory absence of speech)

(b) Continuum2: relationship to linguistic properties (linguistic properties absent or linguistic properties present)

(c) Continuum 3: relationship to conventions (not conventionalized or fully conventionalized)

(d) Continuum 4: character of semiotics (global & synthetic or segmented & analytic)

The gestures studied in this paper are the ones which are produced during the course of spoken language production, that is co-speech gesture which are placed on the left in brackets, which have the features of obligatory presence, no linguistic properties, not conventionalized and having global meanings.

As for the production and interaction of gesture and speech, The influential hypotheses or models at present are *Information Packaging Hypothesis*, *Cross-modal Complementary Hypothesis, Sketch Model*, *Lexical Retrieval Hypothesis*, *GLD (Gesture of learning and Development) Model*, *GSA (Gestures as Simulated Action) Model*, *Interface Model* and *Growth Point Theory*. Nobe (2000) assumes that a comprehensive model of production and interaction of gesture and speech should explain the majority of gestures. However, there is no universal theoretical model which can meet the needs to date, while the current hypotheses can still provide effective explanations for our research to some extent. Many researchers make a series of studies about the production and interaction of gesture and speech, such as Autumn Hostetter et al. (2008) or Wagner (2014), etc. They have made good overviews on the interaction models between speech and gesture, but Autumn Hostetter introduced more ideas of gestures as simulated action by comparing with other models and Wagner provides an overview covering wide range of topics about gesture and speech, and the interaction occupies a small part of this paper. This paper tries to place the current theories and hypotheses (models) in the equal positions, clarify their main contents and analyze the differences among them. In addition, we provide some references for further research in this field.

2. The Studies of Gesture and Speech Interaction

Generally, there are mainly 2 views: one is *ballistic view*, and the other is interactive (Chu, M. & Hagoort, P. 2014). The first view regards that gesture and speech are the different parts in one comprehensive system (i.e., *GSA Model GLD*, *GLD Theory*, *GPT Theory*), they interact each other only in planning phases (i.e. preparation for what they want to say next). While the second view approves of their complementary relationships in communication (i.e., *Interface Model*, *LGP Model*), and proposes that speech and gesture interact not only in planning phases but also during their execution phases (De Ruiter, 2000; Melinger & Levelt, 2004; Nobe, 2000). *Sketch Model* (de Ruiter, 2000) focuses on how gesture is produced in communication and thinking process, as shown in Figure 1. When the message is conceptualized into communicative purposes in Conceptualizer, it will select the characteristics of visual-spatial images from working memory, and then the characteristics information will be sent into Gesture Planner to be converted into concrete actions, that are gestures. In other words, gesture is produced in visual-spatial images and generated in Gesture Planner; there is no need of the input of linguistic elements, simply and fast. The production of speech mechanisms is similar to Levelt's Model. The production of gesture and speech take place in 2 different systems, only over lapping each other in conceptualization processes.

Krauss (1999) postulates the *Lexical retrieval Hypothesis*, and constructs the Lexical Gesture Production model in 2000, as shown in Figure 2. This Hypothesis studies how gestures stimulate the production of speech. It assumes that gesture is produced in working memory, and has nothing to do with communication. Gestures are generated based on some basic spatial characteristics, no connection with visual-spatial images nor with speech production. Butterworth & Hadar (1989) regard that iconic gestures are produced based on the semantic features of vocabulary. These relevant features are selected from working memory and sent into Motor Planner, in other words, gestures stem from the basic images features. Speech and gesture are produced in different sections in memory (visual-spatial and propositional). Their semantic features are activated and dealt with in their own respective systems, and the interaction only takes place in the latter stage---- lexical retrieval process, in which gestures have impacts on the production of speech (Hostetter & Alibali, 2008), inducing the cross-modal interaction.

Kita & Ozyürek (2003) postulate *Interface Hypothesis* and divide the Conceptualizer into 2 parts: Communication Planner and Message Generator. Communication Planner is used to generate communicative proposes, having the similar functions of Conceptualizer (such as ascertaining communicative information, arranging the sequences of contents in messages and selecting modals, etc.). The Message Generator conceptualizes speech regarding the conversational contents and propositional representations, and the Formulator focuses on the selection of vocabulary and pronunciation. This model of speech production is similar to Levelt's model as well. Interface Hypothesis also considers gestures stem form visual-spatial images in working memory. When the communication purposes are identified in Communication Planner, the gesture will be produced in Action Generator. Unlike the former 2 hypotheses, this hypothesis assumes that speech has an effect on the production of gesture. Action Generator interacts with Message Generator bi-directionally and Formulator also interacts with Message Generator bi-directionally. Under these relationships, the production of gesture is both influenced by visual-spatial images in working memory and constrained by linguistic elements. The planning of gesture has also an impact on the planning of speech, likely, *Information Packaging Hypothesis* assumes that gestures stimulate the production of speech through some organizational spatial information.

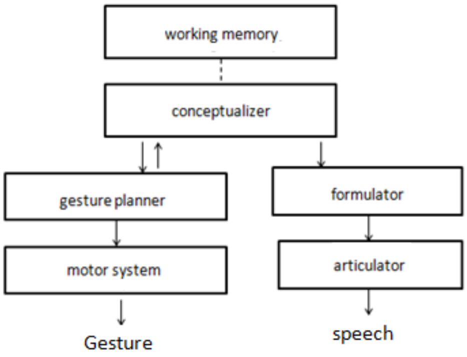

Figure 1. Simplified diagram for *Sketch Mode*

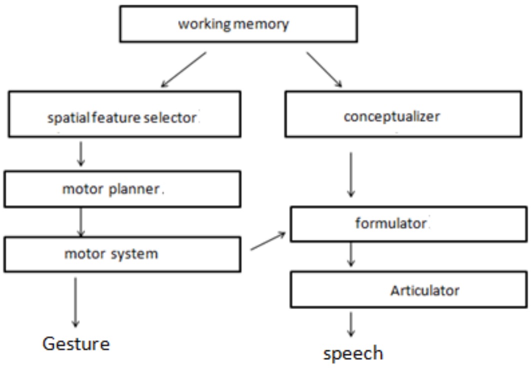

Figure 2. Simplified diagram for Lexical Retrieval Hypothesis

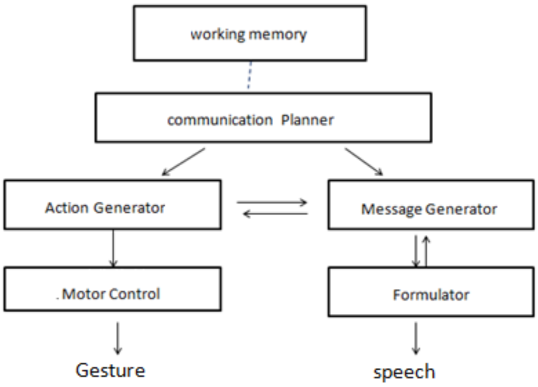

Figure 3. Simplified diagram for *Interface Hypothesis*

Overall Production Architecture (OPA) *Model* (Kopp et al., 2013; Bergmann et al., 2013) is established based on the conceptualization of multimodal information formation in working memory. It studies the semantic coordination cognition of gesture and speech, as shown in Figure 4. Other hypotheses give no further explanations in this field. The coordination formation takes place in working memory: the communicative propose induces multimodal activations and the dynamic activations transmit among visual-spatial representations, symbol-propositional representations and super-modal concepts. For example, “Round” this concept is connected with the visual-spatial features and the features are transmitted to the corresponding linguistic propositional denotation. When the multimodal information is activated, visual-spatial representations and propositional representations will be activated as well, then the activated images messages and symbol-propositional representations will select their corresponding modal messages independently in Message Generator and Image Generator respectively, finally the modal messages will be transmitted into Formulators, and then form speech and gesture.

Gestures as Simulated Action model, GSA (Hostetter & Alibali, 2008) postulates gestures stem from the spatial representations and mental images. Based on an embodied cognitive view, gesture is explained how to be produced from mental images, embodied simulation and speech. Embodied cognition believes that most cognition is related to physical bodies, and speakers will rely on their physical perceptions when they think and express, in other words, they will use gestures. The meanings of the linguistic objects (such as vocabulary, phrases and sentences) are all from physical perceptions not the abstract formalized symbols (Barsalou, 1999).

The speech information is dealt with through denoting the concrete things in real world by lexical concepts and the physical perceptions then simulate actions features. Mental images are also decided by the embodiment of perceptions and action simulation, in other words, images are expressed in physical perceptions and actions, reserving their spatial and physical characteristics. Action images are simulations of physical actions and visual images are simulations of visual perception. When language conceptions are activated, essentially, the perception and movement information are activated. The physical perception and movement information will be expressed in action images and visual images, finally, the information will be expressed in external formations-gestures. So gestures are the natural by-products of cognitive process of language expressions, it is hard to separate them. This point is similar to *Growth Point theory.* GSA model assumes that action planning activation always takes place in working memory, but whether it can produce gesture or not is also influenced by the speakers' gesture threshold and the aggregation of action activations, etc. Gesture threshold is decided by other elements, such as the extent of speakers' willingness to communicate, whether they believe gestures contribute to communication, and so on. If speakers approve that gestures do improve communication, the gesture threshold will be set in a low level, and more gestures will be produced, as shown in Figure 5.

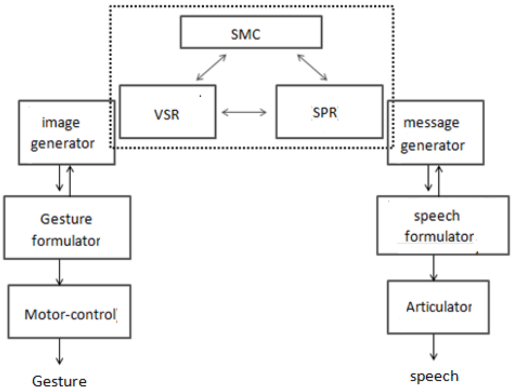

Figure 4. Simplified diagram for *OPA Model*

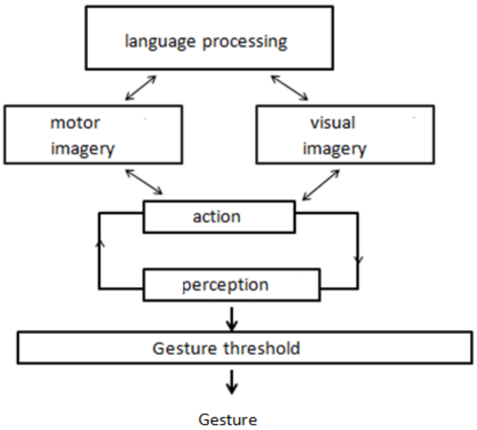

Figure 5. Simplified diagram for *GSA Model*

Unlike other models, *Growth Point theory* (GPT, McNeill, 2005) makes research in how gestures are produced in speech and thinking process. This theory assumes that a growth point is the basic unit in dialectical relationship between images and speech. It is also the starting unit of utterance theoretically. The organizational dynamic process will occur in it, so gestures are produced in growth points. Growth points connect images with linguistic contents; these connections will induce cognitive events. Note that the images here are different from the images in other models, is "linguistic domain….. is not the simple visual-spatial images" (McNeill, 2005). Gestures highlight the vital features of speakers' information by expressing new ideas or growth points, thus the whole conversation will be put forward. This theory believes that gesture and speech are the 2 parts in one system and they are not independent but indispensable when expressing a growth point.

Gesture of Learning and Development Theory by Goldin-Meadow (2003) stresses the mismatching between gesture and speech. This mismatching indicates that gesture production is not influenced by linguistic elements but independent. Butcher & Goldin-Meadow (2000) illustrate that gesture and speech are two independent systems initially, and integrate gradually into one system in growing development of children. When gesture and speech mismatch (that is, the two messages do not match), the communicative proposes will be better delivered by using gesture and speech simultaneously. Gesture can alleviate the cognitive loads and contribute to the production of speech. Gesture is beneficial not only to speech production, but also to the entire cognitive system (Goldin-Meadow, 2003).

Alibali (2000) proves in his *the Information Packaging Hypothesis* (IPH）that gesture production is related to the conceptual planning of speech by descriptive and explanatory experiments. Gesture and speech interact in the initial stage where the information is packaged, organized and allocated into different modals, thus the spatial-movement information to be expressed is packaged into chunks delivered by speech. The more the mission of conceptualizing speech, or the stronger visual-spatial techniques or the weaker language techniques the speakers have (Hostetter & Alibali, 2007), or the more information are introduced in their conversations (Bergmann & Kopp, 2006), the more gestures are produced.

3. Discussion

As for the mechanism of speech production, the *Sketch Model*, *LGP model* and *Interface Model* are all established based on Levelt's model, so they hold the similar ideas in terms of speech production, but they have different interpretations of gesture production and the interaction between gesture and speech. As shown in Table 1, these 3 models agree that gestures stem from the visual-spatial images in working memory, but they disagree with each other in terms of whether there are connections between gesture and speech production. These 3 models believe that speech production and gesture production belong to 2 different systems; they only interact in conceptualizing process in working memory. Specifically, *Sketch Model* assumes that gesture production has no connection with speech production; *LGP model* argues that in the latter production process, gesture will stimulate lexical search and generate speech; while the *Interface Model* considers they interact not only in conceptualizing process, but also in process of production.

Table 1. Comparative analyses among several models about gesture and speech production

	Sketch Model	**LGP Model**	**Interface Model**	**OPA Model**	**GSA Model**	**GLD Theory**	**GPT Theory**
Research focus	speech and gesture production & interaction	speech and gesture production & interaction	speech and gesture production & interaction	semantic coordination cognition between gesture and speech in working memory	Gesture production with embodiment cognition	The impact of gestures on speech	Production of gesture
Production of gesture	Visual-spatial images	Visual-spatial images	Visual-spatial images	No mention	Visual-spatial images	No mention	Visual-spatial images in linguistic domain
Production of speech	Based on Levelt's model	Based on Levelt's model	Based on Levelt's model	No mention	From embodied perception (physical experience)	No mention	No mention

Findings About interaction	Belong to 2 systems, overlap in conceptualizing process	Belong to 2 systems, gesture simulate the production of speech	Belong to 2 system, Interact each other in production process	Dynamic activation transmit among SMC, VSR and SPR.	Belong to 1 system, Gesture production constraint by speech production, gesture stimulate speech production	Develop into one system, gesture contributes to the speech production	Gesture production is influenced by speech elements

Note. **VSR**: visual-spatial representation; **SPR**: symbol-propositional representation; **SMC**: Super-modal Conceptualization.

In comparison, *GSA Model* studies the interaction between speech and gesture from the perspective of embodiment cognition. Unlike other models emphasizing the communicative functions of gesture, *GSA Model* focuses on the production of gesture. It analyzes gesture in cognitive system with a more dynamic view (Hostetter & Alibali, 2008). *GSA Model* does not clarify the bi-directional function between gesture and speech, but it indicates that there is a kind of bi-directional relationship between them theoretically. This model believes that gesture and speech are in one system, and the actions or movements in gesture both contribute to the identification of images characteristics speakers want to express (this point is similar to *Information Packaging Hypothesis*) and contribute to the selections of lexicons for those images characteristics (this point is similar to *Lexical Retrieval Hypothesis*).

The other 3 models: *OPA Model*, *GLD Model* and *GPT Model* stress one aspect of interaction between speech and gesture respectively. *OPA Model* does not explain clearly the production of gesture and speech, and its emphasis is the activation of speech and gesture and their coordination relationship in multimodal conceptualizing process in working memory. Other models do not make further explanation on this issue. *OPA model* believes that when modal conception is activated, these conceptions will activate their corresponding visual-spatial characteristics, which will induce the corresponding language information activation. These activations transmit among these three continuously, which stimulate the formation of image information and symbol propositional denotation, then the production of gesture and speech. This model is exemplified as activation transmission, and it explains successfully some findings about speech and gesture information packaging and information allocation under the cognitive and linguistic constraint (Wagner, 2014). Different from *OPA Model*, *GLD model* studies more about the impact of gesture on speech, and it argues that these two grow into one system during the growing development of children. It believes that gesture helps the production of speech. While GPT Model makes more research in production of gesture, postulating that gesture stems from visual-spatial images which is not the simple images but the images in the linguistic domain. GPT Model approves that effects on gesture production from speech and the whole conversation is put forward by gesture with the growth points. GLD Model and GPT Model prove strongly the importance of gesture in communication, and they point out that the gesture used conventionally in conversations can improve the effectiveness of communication.

4. Implication

The above models focus on the production and interaction of gestures and speech, it is imperative to note that mechanisms of gesture and speech production are still not well understood, especially the linguistic reasons for the production of gesture and speech as well as how they interact (Wagner, 2014). The intimate relationship between speech and gesture covers mainly 2 dimensions: timing and meaning. As for timing, the precise temporal coordination stills cannot reach the general agreement. It is accepted that the onset of a gesture phrase precedes the onset of speech, while this idea need further support. There is no agreement as to the time when gesture and speech come out; the temporal relation of these two is the key to judge whether these two belong to the same system and to understand how they interact. Another area that remains to be studied is whether speakers have firm intuitions about the production of gestures and the interactive action between gesture and speech. What's more, the issue that gesture derives from images or it images characteristics when speakers begin to speak needs further study.

As for meaning, most models support that gesture and gesture share underlying conceptual message and will express the same message, and is redundant with one another (de Ruiter et al., 2012); while others consider the supplementary or compensatory messages they deliver. It is insufficient of the lexical interaction between the gesture and speech in communication. For instance, the communicative functions of gestures are only explicitly illustrated in limited theories, and other theoretical models fail in this aspect. In terms of lexical retrieval process, the function of gestures for lexical retrievals, for instance, the more gesture will occur during pauses or inability to gesture can inhibit the fluency of output of speech, this issue needs further study.

Except the speech gesture relationship, the cognitive and communicative functions of gestures are still unknown well. Ferre (2014) suggests that pragmatic factors be added to broaden the production model of gestures. Other aspects in this field need further proving and exploring, such as the communicative strategy of gesture use and its pragmatic functions in different cultures, the relations between gestures and rhymes and the recognition and transmission of the multimodal information of gestures in the human-computer interaction etc. Deep study of these aspects will contribute to the improvement in a universal theory of production and interaction between gesture and speech.

Acknowledgments

The authors acknowledge the support of Project of the NO.7 China Foreign Language Education Fund (Grant No: ZGWYJYJJ2014A56), the support of the Fundamental Research Funds for the Central Universities (Grant No: DUT15RW135) and the support of Education & Technology Administration Center of China National Education Ministryfund (Grant No: EIJYB2015-066).

References

Abner, N., Cooperrider, K., & Goldin-Meadow, S. (2015). Gesture for Linguists: A Handy Primer. *Language and Linguistics Compass, 9*(11), 437-449. http://dx.doi.org/10.1111/lnc3.12168

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: we think, therefore we gesture. *Language and cognitive Processes*. http://dx.doi.org/10.1080/016909600750040571

Barsalou, W. L. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*. http://dx.doi.org/10.1017/S0140525X99532147

Bergmann, K., & Kopp, S. (2006). Verbal or visual? How information is distributed across speech and gesture in spatial dialogue. In D. Schlangen & R. Fernandez (Eds.), *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue*. Potsdam, Germany: Universitätsverlag.

Bergmann, K., Kahl, S., & Kopp, S. (2013). Modeling the semantic coordination of speech and gesture under cognitive and linguistic constraints. In *Proceedings of the International Conference on Intelligent Virtual Agents (IVA 2013)*. http://dx.doi.org/10.1007/978-3-642-40415-3_18

Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review*. http://dx.doi.org/10.1037/0033-295X.96.1.168

Cienki, A., & Müller, C. (2008). Metaphor, gesture, and thought. In R. W. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought*. Cambridge University Press. http://dx.doi.org/10.1037/0033-295X.96.1.168

GaëlleFerré. (2014). A multimodal approach to markedness in spoken French. *Speech Communication*. http://dx.doi.org/10.1016/j.specom.2013.06.002

Goldin-Meadow, S. (2001). Beyond words: the importance of gesture to Researchers and learners. *Child Development*. http://dx.doi.org/10.1111/1467-8624.00138

Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: relations between verbal and spatial skills and gestures production. *Gesture*. http://dx.doi.org/10.1075/gest.7.1.05hos

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: gestures as simulated action. *Psychonomic Bulletin and Review*. http://dx.doi.org/10.3758/PBR.15.3.495

Kendon, A. (2004). *Gesture-Visible Action as Utterance*. Cambridge University Press. http://dx.doi.org/10.1017/cbo9780511807572

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*. http://dx.doi.org/10.1016/S0749-596X(02)00505-3

Kopp, S., Bergmann, K., & Kahl, S. (2013). A spreading-activation model of the semantic coordination of speech and gesture. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society.*

McNeill, D. (2005). *Gesture and Thought*. University of Chicago Press. http://dx.doi.org/10.7208/chicago/9780226514642.001.0001

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

http://dx.doi.org/10.1075/gest.4.2.02mel

Nobe, S. (2000). Where to most spontaneous representational gestures actually occur with respect to speech? In D. McNeill (Ed.), *Language and Gesture*. Cambridge University Press. http://dx.doi.org/10.1017/cbo9780511620850.012

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).