The Relation of Speech and Gestures: Temporal Synchrony Follows Semantic Synchrony

Kirsten Bergmann, Volkan Aksu, and Stefan Kopp

SFB 673, Bielefeld University, Bielefeld, Germany
{kbergman, vaksu, skopp}@techfak.uni-bielefeld.de

Abstract

The close relationship of speech and gestures becomes conspicuously obvious in the temporal coordination of both modalities. In this paper we investigate in how far temporal synchrony is affected by the semantic relationship of gestures and their lexical affiliates. The results showed that when both modalities redundantly express the same information, the gesture's onset is closer to that of the accompanying lexical affiliate than when gestures convey complementary information: the closer speech and gestures are related semantically, the closer is their temporal relation. This novel finding is discussed with respect to implications for the production process of speech and gestures.

Index Terms: gesture, speech, production, semantic relation, temporal relation

1. Introduction

Co-speech gestures are characterized by being temporally and semantically synchronized with their accompanying speech. An example is illustrated in Figure 1. The speaker is describing a circular church window using the words "such a round window". The gesture is a circular drawing movement. That is, both modalities express more or less the same information about the window's shape redundantly and in temporal synchrony, whereby the gesture stroke starts slightly *before* the lexical affiliate 'round window'.

This multimodal utterance exemplifies two of the most conspicious features of the speech-gesture relationship: (1) the *semantic* and (2) the *temporal* coordination of both modalities. McNeill introduced the notion of 'semantic and phonological synchrony' for this phenomenon [1].

1.1. The semantic relation of speech and gestures

McNeill & Duncan [2] claim that speech and gestures are systematically organized in relation to one another in that they express the same underlying idea, but not necessarily express identical aspects of it. In many cases, the two modalities serve to reinforce one another, as in the introductory example where the circular movement of the hand conveyed gesturally what the word 'round' expressed in the accompanying speech. In other cases, the information to be expressed is distributed across the modalities such that the full communicative intentions of the speaker are interpreted by combining verbal and gestural information.

The semantic synchrony of both modalities can be thought of as a *continuum* of co-expressivity, with gestures encoding completely the same aspects of meaning as speech on one extreme. Although both modalities express information in their specific way, we refer to this as redundancy. Figure 2(a) gives an example for redundant meaning in speech and gesture. While the participant's utterance describes the position of the church ('left'), she also expresses the same information gesturally by positioning the church with her left hand. The gesture does not contain additional information and is, therefore, redundant in relation to speech.

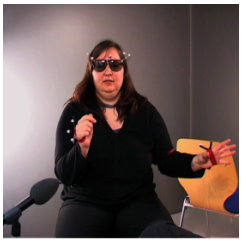
Figure 1: Annotation of a co-speech gesture accompanying "round church window": both modalities convey shape information redundantly and in temporal synchrony with the gesture stroke onset preceding the onset of the lexical affiliate (blue highlighting).

At the opposite extreme of the continuum there are gestures encoding aspects that are not uttered verbally, in other words these gestures complement speech. Figure 2(b) is an example for the complementarity of a gesture. The participant describes the position of a church clock. Without the accompanying gesture the recipient's mental representation of the clock could take different shapes, but the speaker depicts its shape it with her finger. So the specification of the clock's shape is a complementary feature of the speech-accompanying gesture.

Depending on the domain of investigation there seems to be either a 50:50 distribution of redundant and complementary gestures [3, 4] or a tendency of towards more redundant gestures [5].

1.2. The temporal relation of speech and gestures

Regarding the temporal relation of speech and gestures it is mostly uncontroversial that in naturally occurring discourse,



(a) Gesture accompanying the utterance "the church is on he left" as an example for gestural redundancy.

(b) Gesture accompanying the utterance "in the middle there's a clock" as an example for gestural complementarity

Figure 2: Examples for redundancy and complementarity in gestures.

gestures do either precede or synchronize with those words they are affiliated with [6, 7, 8, 9, 10, 11, 5, 12].

There are, however, different explanations for this phenomenon. Some researchers argue that the "gap" between speech and gesture onset results from difficulties in retrieving lexical items [6, 9, 13]. This idea is based on empirical evidence that the restriction of gesturing adversely affects speech (see [14] for a review). According to this view, gestures provide input for the speech production process where they assist in assessing words via cross-modal priming. In other words, there is an interaction between modality-specific formulation processes at a relatively *late* stage: when the gesture is readily planned and already in execution.

An alternative way of thinking has been proposed by de Ruiter [15, 16, 17] who assumes that "gesture and speech are planned together at an *early* state in utterance production" [17, p. 26]. In this view, a common process takes charge of distributing information across modalities. Apart from this process, speech and gesture are processed independently. De Ruiter gives two different explanations for the fact that gestures precede their verbal affiliates [15]. First, gestures do not have the complicated syntactic properties that spoken language has and, therefore, need less production time. And second, in utterances where an iconic gesture is made, the communicative intention often involves imagery. For speech, this imagery has to be translated into a propositional format which might require extra processing time.

That is, the point of contention in the literature is at which level of computation in the production of gestures the temporal gap between speech and gesture onset is caused. To shed some more light onto this point of discussion we investigate the role of semantics in the temporal coordination of both modalities in this paper. More specifically, we elucidate if the asynchrony of speech and gesture onset is influenced by meaning: are speech and gestures temporally closer aligned in the case of semantic redundancy? Or in other words, does a close temporal relation of speech and gestures reflect a close semantic relation of both modalities? To test this hypothesis, we employed an analysis a corpus of natural speech and gesture use engaged in a spatial description task. In Section 2 the corpus, its annotation, and reliability issues are described. Section 3 presents our results from data analysis which are discussed with regard to their relevance for production models in Section 4.

2. Corpus

The data used for this paper has been taken from the Bielefeld Speech and Gesture Alignment (SaGA) corpus which consists of a total of 25 route-description dialogues between native speakers of German. After getting a "bus ride" through a virtual reality environment, participants described the driven route including five major sights to an addressee in a face-to-face situation. The data was annotated in several steps. First, the audio tracks were annotated for speech using Praat [1], without reference to the video data. Then the video clips were annotated using the annotation software Elan[2] to identify single gesture occurrences structured into the following gesture phases: preparation, pre-stroke hold, stroke, post-stroke-hold, and retraction [1, 18]. After completion, both speech and gesture coding was merged in Elan.

Based on these fundamental segmentations, the data was further annotated: words were tagged with part-of-speech information, parsed for their syntactical structure, and coded for their dialogue context; gestures were classified (including gesture representation techniques) and coded for their gesture features; a subpart of the corpus (only sight descriptions) has been further coded for the gestures' referent objects and their spatio-geometrical properties (dimensionality, extents, symmetries, profiles, etc.). In total, the SAGA corpus consists of 280 minutes of video material containing 4961 iconic/deictic gestures, approximately 1000 discourse gestures and 39,435 words. For details see [19].

For current analysis, a sub-corpus of object descriptions of four sights (town hall, chapel, church square, fountain) was employed. This data (973 gestures) has been annotated for semantic information encoded in both speech and gestures as described in the following.

2.1. Lexical affiliates

In a first step, the lexical affiliate of each gesture stroke has been determined on the speech level. According to Schegloff [8] a lexical affiliate is the word(s) deemed to correspond most closely to a gesture in meaning. Therefore it serves as an explicit temporal link between gesture and speech. In our annotations we constrained the lexical affiliates to a minimum of words. These were typically nouns or adjectives as the speakers were engaged in object descriptions. In some cases it was not possible to choose a specific lexical affiliate. This is because some gesture annotations are not accompanied by a lexical affiliate on the speech level, for example when using comparisons ("it looks like") or colors ("a red church"). Because of these exceptions determining the lexical affiliate of the gesture is not an easy task. Therefore, we created an annotation list of rules to minimize error rates and to guarantee quality and reliability of annotation data. Some specific examples are listed below:

- For isolating the lexical affiliate as far as possible prepositions are omitted, if no relevant information is lost like: "middle" in place of "in the middle" or "left" in place of "on the left"
- Indefinite and definite article are not a part of the lexical affiliate: "round window" in place of "the round window"

[1] http://www.fon.hum.uva.nl/praat/

[2] http://www.lat-mpi.eu/tools/elan/

- Colors are ignored, because they cannot be gesticulated, unless the colors are inseparable like: “two blues spiral staircases”
- The lexical affiliate doesn’t involve the amount of entities on the speech level: “building” in place of “each building” or “street lights” in place of “two streetlights”

2.2. Semantic features

Both the speech and gestures were analysed with respect to the semantic information they represented, based on an established micro-analytic coding method using a range of semantic features [20, 21, 22, 23, 24, 25, 26]. The set of semantic features included in this analysis was considered to capture the kind of semantic information contained in our object description data. Our analysis focused on the amount of information represented, regardless of whether the information was complementary or redundant with regard to the information in the respective other modality. Hence, verbal utterance, as well as each iconic and deictic gestures were analyzed for the semantic information they contained, based on the following semantic categories with the rules used to annotate speech semantics:

- *Entity*: Commonly known as objects like “streets”, “landmarks” or “bridges” regarding spatial dialogs.
- *Relative Position (RelPos)*: Spatially distributed entities have got relative positions to one another. A possible example is “The tree is in front of the church” or “the hedge is near the tree”.
- *Shape*: This category determines the shape of an entity. In a sentence like “the clock is round” the adjective “round” describes the shape of the clock.
- *Amount*: This category describes particular number of entities, but also by words like “several” or ”many”.
- *Size*: It defines the size of an entity like “big” or “small”.
- *Property*: Other properties of entities like colors or materials are annotated as property.

Concerning the meaning of gestures the same categories are used. The first decision to be made is applied to the dynamics of each gesture. A gesture can be either dynamic or static. Dynamic gestures include a trajectory between starting point and target point, while static gestures only consist of a posture at a target position. In the latter case either *RelPos*, *Size* or *Amount* are taken into consideration. Typically, positioning gestures are done with one hand, while sizes are visualized with both hands, but in case of doubt the (verbal) context is decisive. If two entities are localized, *Amount* is annotated additionally. For dynamic gestures there is a wider range of possibilities. In a first step one has to distinguish gestures referring to actions and gestures referring to entities. For the latter ones the SFs *Shape*, *Size*, and *Amount* are considered. Supportive for the coder is a look at the gesture morphology where gesture shapes may be found. If the gesture conveys a *Shape*, typically the trajectory or the inner sides of the hands form it. *Size* can be found in a dynamic gesture as well, because sometimes a ‘scaling’ movement refers to the size of entities. Moreover, the morphology clearly contains information about the extent. Typically, *Amount* is assigned to a gesture if it refers to more than two entities. In these cases *RelPos* is annotated as well.

2.3. Reliability

Annotation-based data might be problematic as they are based on subjective judgements of the coders. Of vital importance for the significance of results is, therefore, the reliability of the annotation. It has to be shown that different annotators agree with respect to the coding judgements on which statistical analyses are based to make research results replicable. The standard method for gauging reliability of annotations are chance-corrected assessments of the agreement between multiple annotations of the same material. Accordingly, 13.5% of the data has been annotated by two annotators to investigate the degree of reliability.

For semantic feature coding, Cohen’s Kappa [27] was employed, as a metric to evaluate data on a nominal scale. We reached Kappa values of κ=0.76 for gestures and κ=0.86 for speech. Following [28], these values can be interpreted as substantial agreement. For the identification of lexically affiliated word’s onsets in speech, a metric variable, we employed a product-moment correlation (Pearson’s r), resulting in a value of r = 0.98. According to Diaz-Bone [29] this has to be interpreted as a strong correlation.

3. Results

3.1. Temporal relation of speech and gestures

In order to analyze the data a variable called *speech-gesture asynchrony* was calculated: the difference between the onset times of the lexical affiliate and the gesture stroke. Asynchrony values above 0 indicate that the gesture onset precedes its lexical affiliate, while values below 0 indicate that the gesture onset follows the lexical affiliate. This enables us to calculate the arithmetic mean of the speech-gesture asynchrony. The mean asynchrony between gesture and speech onset based on all 973 annotations amounts to 127.89 msec. That is, on average the gesture onset precedes its lexical affiliate. The modal value of the distribution is 0, meaning that in most annotations speech and gesture onsets begin simultaneously. The minimal value amounts to 3102 msec and the maximal value is -2300 msec.

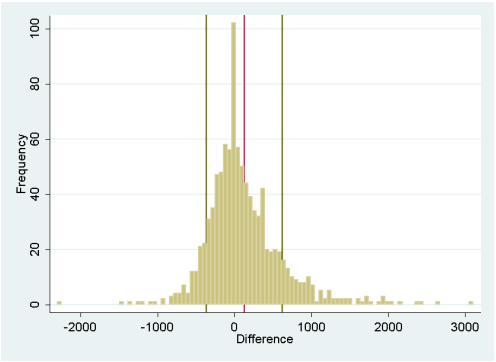


Figure 3: Distribution of gesture-speech asynchrony with the mean visualized in red and standard deviations visualized in green.

Because of these high variations it is useful to calculate a measure of dispersion. The standard deviation of the temporal differences amounts to 495.40 msec, which is quite high con-

sidering the general distribution of the differences. These values can be seen in Figure 3 which also shows an approximate Gaussian distribution of the differences. Given a higher number of cases an even stronger Gaussian distribution can be assumed.

3.2. Semantics of speech and gestures

In the course of judging the gesture semantics, each gesture got assigned between one and three SFs: 80.4% of the gestures have one SF, 16.6% of them have two SFs, and 3.0% have three SFs. Figure 4 summarizes the distribution of SFs and their combinations in gestures and their verbal affiliates. The categories *RelPos* (37.41%) and *Shape* (36.79%) are prevalent for both gestures with one SF as well as gestures in which two or three SFs are combined. This distribution reflects the fact that the data was elicited in a spatial communication task. The most frequent combination of SFs is *Amount+RelPos*. This combination typically occurs when several entities are depicted in relation to each other.

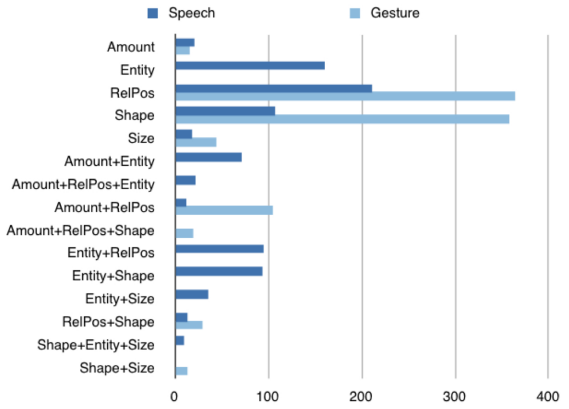


Figure 4: Absolute frequencies of SFs and their combinations in gestures and their verbal affiliates (only combinations with >10 cases considered).

With regard to the semantics of speech, there were between one and three SFs encoded in the lexical affiliates. 53.3% of the gestures have one SF, 32.9% of them have two SFs, and 5.8% have three SFs. The most often occurring categories are *RelPos* (21.7%), *Entity* (16.5%) and *Shape* (11.0%). Frequent combinations of SFs are *Entity+X*, e.g., *Entity+RelPos* (9.8%), *Entity+Shape* (9.7%), or *Entity+Amount* (7.3%). So the lexical affiliates typically express exactly one SF, or they name an entity in combination with further characterizing information.

3.3. Semantic relation of speech and gestures

Among all SFs, 63.1% are redundant while 36.9% are complementary to the accompanying speech. This distribution is reasonably in line with earlier findings on a level of semantic features [4, 3, 23, 5]. In terms of gesture-wise consideration, one finds 58.4% of the gestures being completely redundant, that is they do not have any complementary SFs. Another 28.8% of the gestures do not have any redundant SFs and therefore are exclusively complementary. Finally 12.8% of the gestures do have both redundant and complementary parts. Figure 5 summarizes the number of times that different types of SFs occur in gestures.

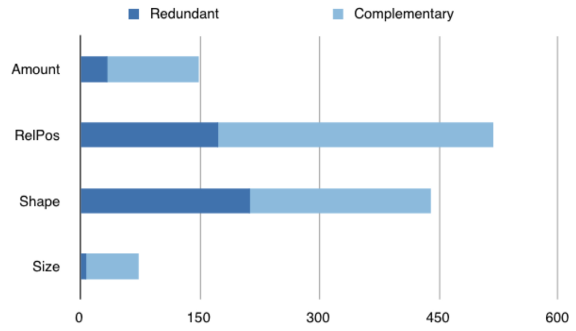


Figure 5: Distribution of the different kinds of redundant/complementary SFs.

3.4. Temporal relation vis-a-vis semantic relation

To test the hypothesis that the temporal relation of speech and gestures is related to the distribution of semantic information across modalities, we first compared redundant gestures (i.e., gestures without any complementary features) and complementary gestures (i.e., gestures without any redundant features) with respect to their gesture-speech asynchrony.

We began by calculating the arithmetic mean and standard deviation for complementary and redundant gestures. A total of 566 gestures is redundant, i.e., these gestures do not have any complementary semantic features. For these gestures the gesture-speech asynchrony amounts to 107.35 msec. The mean value of the 279 purely complementary gestures (i.e., no redundant semantic features) amounts to 251.05 msec (see Figure 6). A *t*-test comparing the asynchrony means for redundant and complementary gestures, revealed a significant difference across the two types of gestures ($T(542)=3.93$, $p<.001$): for redundant gestures the gesture stroke's onset is closer to its lexical affiliate than for complementary gestures. The standard deviations are both quite high (481.95 msec for redundant and 496.52 msec for complementary gestures), but do not vary relevantly from each others. It can be concluded, that distinguishing complementary and redundant gestures has an influence on the temporal asynchrony of speech and gesture onset.

Additionally, we also considered the onset of the gesture's preparation phase and its relation with the lexical affiliate's onset. Since not every gesture necessarily has a preparation phase, 492 redundant and 255 complementary gestures are taken into account here. The mean duration of preparations is 700.10 msec (SD=431.06 msec) for redundant gestures, and 751.59 msec (546.93 msec) for complementary gestures. This difference across gesture types not significant.

So far, we analyzed the poles of the expressivity continuum regarding their gesture-speech asynchrony: purely redundant or complementary gestures, only. But what about gestures between those extremes in which both redundant and complementary features are present? For these gestures (N=125) the mean gesture-speech asynchrony is -62.46 msec (SD=483.40 msec). That is, these gestures are atypical in the sense that the speech onset precedes the gesture stroke onset. A closer analysis revealed that a majority of these gestures convey the semantic feature *Amount* as in the utterance "two towers" accompanied by a two-handed gesture in which each hand represents one of the towers by a vertical trajectory. That is, the typical

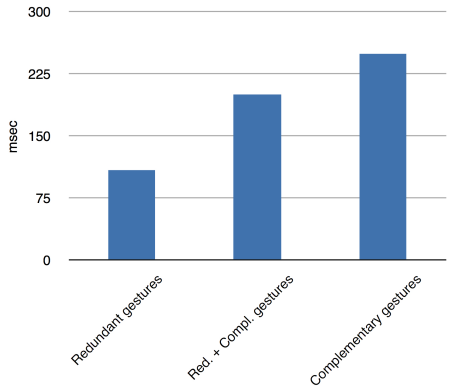


Figure 6: Mean values of gesture-speech asynchrony.

case of a lexical affiliate consisting only of a single noun or adjective are elongated by a cardinal. To investigate whether these cardinals were responsible for the negative gesture-speech asynchrony, we modified the coding of lexical affiliates by excluding cardinals. An analysis of the abbreviated affiliates revealed a mean gesture-speech asynchrony of 201.24 msec (SD=562.94 msec) which lies well between the asynchrony for redundant-only gestures and complementary-only gestures. That is, the typical situation of gesture onset preceding speech onset relates to the narrow notion of lexical affiliates without modifiers such as cardinals.

A one-way ANOVA taking all three gesture types into account revealed a significant main effect ($F(2,970)=6.06$, $p=.002$). Pairwise *t*-tests showed that this is due to the significant differences between redundant-only and complementary-only gestures as described in the previous section.

4. Discussion and conclusions

The aim of our corpus analysis was to shed light on the mechanisms and causes of the well-known temporal synchrony between speech and gestures. In particular, we have studied for the first time the relation between temporal synchrony and semantic synchrony. Our corpus analysis revealed that the temporal coordination of speech and gesture onset is actually sensitive to the semantic relation of both modalities. In particular, we found that the gesture-speech asynchrony is decreased when both modalities express the same content, while its is increased when gestures complement speech.

Regarding this finding there are generally two different explanations conceivable. The first one is that for redundant speech and gestures, the speech production process is faster. What might cause such a mechanism? A possible reason would be that the transformation from imagistic information into a propositional representation is faster because the redundant aspects of meaning are already activated for the purpose of gesture generation. In the case of non-redundancy, on the contrary, the activation of semantic features for speech production should take longer because there is no facilitating effect from gesture generation. This argument of an accelerated activation process on the content planning level is in line with an *early* interaction in the production process.

The alternative explanation would be that gesture adapts more strongly to the flow of speech to create the finer temporal synchrony. This means that the stronger semantic coupling between the two modalities leads to an increase of coordination at later stages of the production process. Given the commonly acknowledged hypothesis that gesture production is faster and generally ahead of the speech production process, under this explanation, the gesture production process delays gesture execution until a better prediction of the timing of the lexical affiliate is possible. This is consistent with the later onset of the gesture preparation phase in redundant gestures, but would require a relatively *late* interaction after speech formulation because timing information about surface elements of speech are passed on. Alternatively, the later onset of gesture preparation could be due to a more complex and more time-consuming gesture planning process caused by the fact that redundant gestures have to be coordinated more finely with speech. Assuming that formation of a gesture is still ongoing during its preparation phase (cf. [1]), in this case one would expect an increased duration of the preparation phase. This does not show up in our data, in which by trend the preparations of redundant gestures are even slightly shorter than in complementary ones.

Given our corpus data it is not possible to exclude one alternative or the other. Further research is definitely necessary to elucidate which mechanisms in the production process of speech and gestures actually result in the finding of synchronized semantic and temporal coordination. On the one hand, psycholinguistic experiments should be carried out to investigate the production process of multimodal utterances by manipulating the two mechanisms separately. On the other hand, computational simulations provide an alternative method to test whether both kinds of interactions result in the effects we observed empirically. We already developed such a computational model in earlier work [30, 31], see Figure 7. It consists of four processing modules to be involved in content planning and micro-planning of speech and gesture: *Image Generator, Preverbal Message Generator, Speech Formulator*, and *Gesture Formulator*. In addition, two dedicated modules (Motor Control and Phonation) are concerned with the realization of synchronized speech and gesture movements for virtual agents. All modules are modeled as software agents that operate concurrently and proactively on a central working memory, realized as a globally accessible, structured blackboard. This enables for any kind of interaction between the different modules and we are, thus, able to implement the possible interaction mechanisms as discussed above. The resulting speech-gestural behavior can then be evaluated in comparison with the empirical results reported in this paper.

5. Acknowledgements

This research is partially supported by the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center 673 “Alignment in Communication” and the Center of Excellence in “Cognitive Interaction Technology” (CITEC).

6. References

[1] D. McNeill, *Hand and Mind—What Gestures Reveal about Thought*. Chicago: University of Chicago Press, 1992.

[2] D. McNeill and S. Duncan, “Growth points in thinking-for-speaking,” in *Language and gesture*, D. McNeill, Ed. Cambridge, UK: Cambridge University Press, 2000, pp. 141–161.

[3] J. Cassell and S. Prevost, “Embodied natural language generation: A framework for generating speech and gesture,” 1997.

[4] H. Yan, “Paired speech and gesture generation in embodied conversational agents,” Master’s thesis, MIT, School

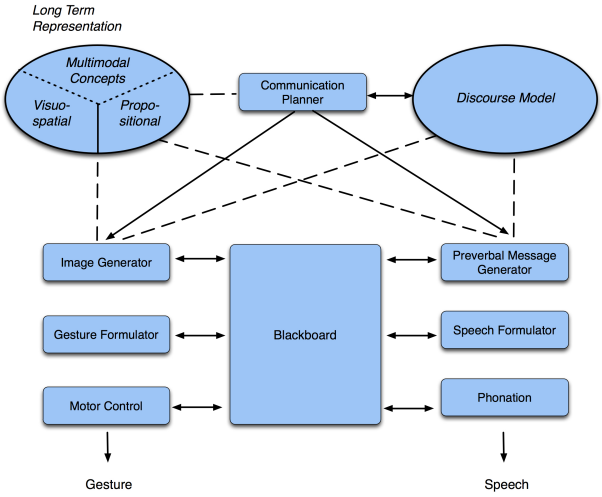


Figure 7: Architecture of our computational speech and gesture production model [30, 31] which allows to simulate interaction between modality-specific planning modules.

of Architecture and Planning, 2000. [Online]. Available: http://alumni.media.mit.edu/ȳanhao/hao_yan_thesis.pdf

[5] K. Pine, N. Lufkin, and D. Kirk, E. andMesser, "A microgenetic analysis of the relationship between speech and gesture in children: evidence for semantic and temporal asynchrony," *Language and Cognitive Processes*, vol. 22, pp. 234—246, 2007.

[6] B. Butterworth and G. Beattie, "Gesture and silence as indicators of planning in speech," in *Recent advances in the psychology of language: Formal and experimental approaches*, R. Campbell and G. Smith, Eds. New York: Plenum Press, 1978, pp. 347–360.

[7] J. Ragsdale and C. Silvia, "Distribution of kinesic hesitation phenomena in spontaneous speech," *Language and Speech*, vol. 25, pp. 185–190, 1982.

[8] E. Schegloff, *On Some Gestures' Relation to Talk*. Cambridge University Press, 1984, pp. 266–298.

[9] P. Morrel-Samuels and R. Krauss, "Word familiarity predicts temporal asynchrony of hand gestures and speech," *Journal of Experimental Psychology: Learning, Memory, & Cognition*, vol. 18, pp. 615–622, 1992.

[10] K. Chui, "Temporal patterning of speech and iconic gestures in conversational discourse," *Journal of Pragmatics*, vol. 37, pp. 871–887, 2005.

[11] P. Bernardis and M. Gentilucci, "Speech and gesture share the same communication system," *Neuropsychologia*, vol. 44, pp. 178–190, 2006.

[12] J. Blake, D. Myszczyszyn, A. Jokel, and N. Bebiroglu, "Gestures accompanying speech in specifically language-impaired children and their timing with speech," *First Language*, vol. 28, pp. 237–253, 2008.

[13] R. Krauss, Y. Chen, and R. Gottesman, "Lexical gestures and lexical access: A process model," in *Language and gesture*, D. McNeill, Ed. Cambridge, UK: Cambridge University Press, 2000, pp. 261–283.

[14] F. Rauscher, R. Krauss, and Y. Chen, "Gesture, speech, and lexical access: The role of lexical movements in speech production," *Psychological Science*, vol. 7, pp. 226–231, 1996.

[15] J. de Ruiter, "Gesture and speech production," Ph.D. dissertation, University of Nijmegen, 1998.

[16] ——, "The production of gesture and speech," in *Language and gesture*, D. McNeill, Ed. Cambridge, UK: Cambridge University Press, 2000, pp. 284–311.

[17] ——, "Postcards from the mind: The relationship between speech, imagistic gesture, and thought," *Gesture*, vol. 7, no. 1, pp. 21–38, 2007.

[18] S. Kita, I. van Gijn, and H. van der Hulst, "Movement phases in signs and co-speech gestures, and their transcription by human coders," in *Gesture and Sign Language in Human-Computer Interaction*, I. Wachsmuth and M. Fröhlich, Eds. Berlin/Heidelberg: Springer, 1998, pp. 23–25.

[19] A. Lücking, K. Bergmann, F. Hahn, S. Kopp, and H. Rieser, "The Bielefeld speech and gesture alignment corpus (SaGA)," in *LREC 2010 Workshop: Multimodal Corpora—Advances in Capturing, Coding and Analyzing Multimodality*, M. Kipp, J.-P. Martin, P. Paggio, and D. Heylen, Eds., 2010.

[20] G. Beattie and H. Shovelton, "Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation," *Semiotica*, vol. 123, pp. 1–30, 1999.

[21] ——, "Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech," *Journal of Language and Social Psychology*, vol. 18, pp. 438–462, 1999.

[22] ——, "An experimental investigation of the role of different types of iconic gesture in communication: A semantic feature approach," *Gesture*, vol. 1, pp. 129–149, 2001.

[23] K. Bergmann and S. Kopp, "Verbal or visual: How information is distributed across speech and gesture in spatial dialog," in *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue*, D. Schlangen and R. Fernandez, Eds., 2006, pp. 90–97.

[24] J. Holler and G. Beattie, "A micro-analytic investigation of how iconic gesture and speech represent core semantic features in talk," *Semiotica*, vol. 142, pp. 31–69, 2002.

[25] ——, "How iconic gestures and speech interact in the representation of meaning: Are both aspects really integral to the process?" *Semiotica*, vol. 146/1, pp. 81–116, 2003.

[26] ——, "The interaction of iconic gesture and speech," in *Proceedings of the 5th International Gesture Workshop*, A. Camurri and G. Volpe, Eds. Berlin/Heidelberg: Springer, 2004, pp. 63–69.

[27] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, pp. 37–46, 1960.

[28] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, pp. 159–174, 1977.

[29] R. Diaz-Bone, *Statistik für Soziologen*. Stuttgart: UVK Vertragsgesellschaft, 2006.

[30] S. Kopp, K. Bergmann, and I. Wachsmuth, "Multimodal communication from multimodal thinking—towards an integrated model of speech and gesture production," *Semantic Computing*, vol. 2, no. 1, pp. 115–136, 2008.

[31] K. Bergmann and S. Kopp, "Increasing expressiveness for virtual agents–Autonomous generation of speech and gesture in spatial description tasks," in *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems*, Budapest, Hungary, 2009, pp. 361–368.