Freigang F, Kopp S. Exploring the speech-gesture semantic continuum. Presented at the 1st European Symposium on Multimodal Communication, 2013, Valletta, Malta.

Exploring the speech-gesture semantic continuum

Farina Freigang and Stefan Kopp
farina.freigang@uni-bielefeld.de, skopp@techfak.uni-bielefeld.de
Faculty of Technology, Center of Excellence "Cognitive Interaction Technology" (CITEC)
Collaborative Research Center "Alignment in Communication" (SFB 673)
Bielefeld University, P.O. Box 100 131, D-33501 Bielefeld, Germany

In natural conversation, speech and gesture are usually one unit that is either produced or received by a communication partner. However, the relationship between the meaning of speech and the meaning of gesture can differ. Several terms have been used to specify these different relationships, ranging from "redundant" over "supplemental" to "mismatching" information. No consensus about the exact definition of these terms or the appearing variety in how speech meaning and gesture meaning relate to each other has been reached. We argue that this confusion is due to the fact that these terms address different dimensions of the speech-gesture semantic relationship, and therefore can hardly be related directly with each other. In the following, we discuss the terminology and related studies with regard to production and comprehension.

On the side of language *production*, McNeill (1992) already discussed semantic synchrony in general without going further into detail. Alibali and Goldin-Meadow (1993) were the first to report"mismatches" produced by children learning the concept of mathematical equivalence. This term is not completely agreed on by Willems, Özyürek, and Hagoort (2007) who found that the term "mismatch" should be used with an "incongruent" speech-gesture pair and not when gesture conveys "additional" but not contradicting information as speech. They referred to the mismatch phenomenon as speech-gesture "incongruence". Furthermore, Kelly, Özyürek, and Maris (2010) accepted both terms "mismatch" and "incongruence". In this context, other terms have been mentioned, e.g., speech-gesture "concordance", "concurrent" speech-gesture pairs, "redundant" gestures, and "semantic coordination" of speech-gesture pairs. A detailed definition of these terms and a comparison between them is as of yet still missing.

On the side of language *perception*, McGurk and MacDonald (1976) showed that speech perception is not a purely auditory process but that mouth gestures can influence the recipient's interpretation of what has been said by the message giver. Sometimes this interpretation results in a third meaning, different from the speech or mouth gesture own their own. Similar to the McGurk-MacDonald effect, one can assume that observed speech-gesture mismatches or incongruences may lead to a third interpretation by a subject. Habets, Kita, Shao, Özyurek, and Hagoort (2011) looked at semantic congruent and incongruent combinations ("matches" and "mismatches") or "semantic integration" of speech and gesture during comprehension in an EEG study and found that "mismatching gesture-speech combinations lead to a greater negativity on the N400 component in comparison with matching combinations" (p. 1852). This suggests a cognitive basis for what counts as mismatching in terms of whether speech and gesture can be integrated.

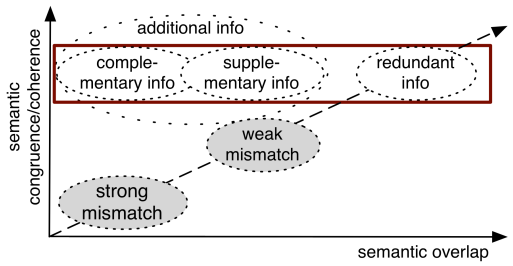

Figure 1: Two-dimensional space of semantic coordination.

From our point of view, the appearances and the understanding of the speech-gesture semantic relationship has a lot more depth to it than sketched so far. In figure 1, we propose a two dimensional space that separates the semantic overlap from the semantic congruence/coherence of speech and gesture. A gesture can convey complementary (*different but necessary*), supplementary (*additional*), or redundant (*corresponding, matching*) information in relation to speech (and vice versa). While this level of semantic overlap has been studied throughout (box in figure 1), it implicitly assumes a high level of coherence between speech and gesture meaning (in the sense of being integrable into a coherent unified interpretation). This congruence, we argue, makes for a second dimension. If a gesture is produced or received with neither semantic overlap nor congruence with speech meaning, we define this as a strong **semantic mismatch**, or hereafter just mismatch. A weaker mismatch is produced or received, if moderate overlap and intermediary congruence between speech and gesture meaning is given.

With this in mind, we define a continuum of mismatches between speech and gesture (dashed arrow in figure 1). In figure 2, three examples along the

continuum are illlustrated in more detail, depending on whether the concepts expressed in speech and gesture are totally different, whether they are derived from the same concept field or whether the concepts are the same. An incongruent speech-gesture pair is a strong mismatch, whereas there are weaker forms up to redundant information in either speech or gesture. Examples of produced or received messages are "high", "wide" and "round", and corresponding gestures which can be used interchangable (cf. figure 2).

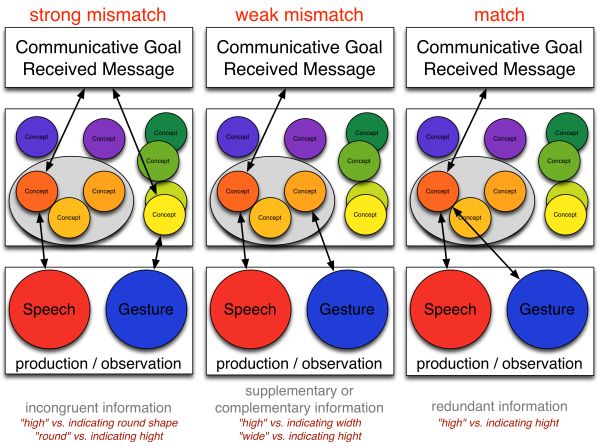

Figure 2: Concepts linked to speech and gesture.

On the basis of this theoretical discussion, how can we explore this phenomenon empirically? The research question on mismatch *production* is, to what extent are speech-gesture mismatches actually produced in natural communication and, since they are hardly found, when and how do they occur in artificial settings? The research question tackling the *comprehension* part of speech-gesture mismatches is, how do subjects cope with conflicting information of the two modalities speech and gesture, do they sometimes interpret a third meaning and are there different levels of impact in each modality?

In preliminary work, we investigated mismatches in natural language production and comprehension. In our first pilot study, the focus laid on mismatches in *production* and we learned that it is difficult to elicit speech-gesture mismatches from adult subjects. The subjects were shown pictures with optical illusions and had to describe the error in the picture either from memory or while looking at it (hands-free), which has been taped on video. We did neither control for the time a mismatch occurs nor for the type, we just created a condition of high cognitive load on the subject. The result was that an adult subject rather interrupts herself during the description process than accepting a semantic mismatch.

In our second pilot study we concentrated on *comprehension* mismatches, similar to Habets et al. (2011). We were able to confirm the tendency of a third meaning emerging from a speech-gesture mismatch (cf. McGurk and MacDonald, 1976) in some cases. In the experiment, we combined conventional gestures like clear pointing gestures up to vague open-arm gestures with spoken words like 'there' and 'everything'/'no clue' and jumbled them. The 64 video snippets (no filler gestures and words) where rather unnatural as they consisted of a single word aligned to one gesture. We decided for the gesture and the word not to appear in some context, since this may influence the subjects interpretation of the related meaning and we were hoping for yet a different meaning than the word and gesture meanings on their own. The results of a subsequently completed questionnaire showed a notable visual impact on the subjects, which may be due to the poor audio quality of the video or to the fact that the visual modality, being quiet dominant, acts as a modifier to the spoken words.

In order to investigate these research questions further, we are about to conduct further experiments. In a first experiment, subjects may have to determine the meaning of certain gestures and certain words independently. We expect clear, vague, and ambiguous meanings in both speech and gesture. Furthermore, we expect some gestures to act as a modifier to the spoken words. Subsequently, we will cluster similar meanings. These clusters will again be checked for their combined meaning by another set of subjects. We considered using predefined gesture lexicons like the "Berliner Lexikon der Alltagsgesten" (BLAG) (Posner, Noll, Krüger, & Serenari, 1999), however, the gesture performance is only shown imprecisely on pictures. Interestingly, the gesture meanings are about the same we have investigated so far. In a second experiment, we may use these predefined gesture and word meanings to conduct the perception experiment again with more attention to detail.

References

Alibali, M. W., & Goldin-Meadow, S. (1993). Modeling Learning Using Evidence from Speech and Gesture. In *Proceedings of the Annual Conference of the Cognitive Science Society.*

Habets, B., Kita, S., Shao, Z., Özyurek, A., & Hagoort, P. (2011). The Role of Synchrony and Ambiguity in Speech–Gesture Integration during Comprehension. *J Cogn Neuroscience*, *23*(8), 1845–1854.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two Sides of the Same Coin Speech and Gesture Mutually Interact to Enhance Comprehension. *Psychological Science*, *21*(2), 260–267.

McGurk, H., & MacDonald, J. (1976). Hearing Lips and Seeing Voices.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought.* University of Chicago Press.

Posner, R., Noll, T., Krüger, R., & Serenari, M. (1999). Berliner Lexikon der Alltagsgesten. *Berlin.*

Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When Language Meets Action: The Neural Integration of Gesture and Speech. *Cerebral Cortex*, *17*(10), 2322–2333.