Effects of mismatched multimodal communication on perception, integration and comprehension by different populations of learners/addressees.

The traditional debate about the relation between thought and language has been revised by embodied views of cognition. Whether cognition is proposed as grounded [1] or directly situated [2,3] in physical (multimodal) interaction with the world, the idea that cognition is best described as abstract symbol manipulation is challenged. The argumentation is partly based on evidence that gestures accompanying speech form part of a unified multimodal message, affecting the cognitive states of both sender and receiver [4].

These findings have also reverberated in the area of learning and education [5], even in the most abstract domains [6]. Much focus has been on production of gestures. As an example it is worth mentioning a finding that children who use gesture in conversation with a teacher during instruction in mathematical equivalence perform better when later tested [7]. It is argued that gesturing here functions as visuospatial reasoning, aiding the formation of conceptual knowledge. Gestures also play a role in language development: children often gesture before they use words [8,9].

Other studies have focused on how addressees perceive and understand gestures. Apart from aiding the speakers' mental processes, there is evidence that nonverbal channels at least sometimes are performed to communicate [10] and influence addressees' comprehension of a message [11]. A common method to study the integration of a multimodal message is to manipulate the way that different modalities are combined. Examples include spatially amplifying [12] or eliminating [11] gestures accompanying speech. Another approach is to introduce a mismatch between gesture and speech. Introducing different types of mismatch allows testing specific hypotheses about effects on semantic integration and comprehension of the message, as well as perception of the mismatch itself (noticing “something is wrong”).

Gestures (and other communicative behaviors) can be categorized as either propositional (representational of content) or interactional (non-representational, e.g. “beats”) [13]. An analogous dichotomy can be applied to mismatches. An example of a propositional mismatch is provided by a study where teachers expressed different or identical valid problem-solving strategies with their speech and gesture. It was found that the participating students’ comprehension was improved by the observation of gestures, but only when they were mismatched with the verbal channel (i.e. non-redundant). Mismatches involving or replacing specific subclasses of propositional gestures (iconic, metaphoric, deictic) can be further distinguished as of the corresponding type. An example of an interactional mismatch is to make gestures asynchronous to speech which has been shown to negatively affect recall [14]. Note that this type of mismatch can be applied to both propositional and interactional gestures, and have effects on perception as well as semantic integration or comprehension. Further variations of mismatch arise from manipulating the gaze of a speaker. Such mismatches can be purely interactional (disrupting eye-contact mediated turn-taking [15]) or deictic (directed to an “unexpected” object. It might however be necessary to limit the scope of the project to
gesture-speech mismatch.

The purpose of the proposed project is to explore how different types of mismatch affect perception and comprehension by children with communication impairments, such as specific language impairment (SLI) and autism spectrum disorder (ASD), in comparison to normally developing (ND) children. Finding amplified or diminished effects by mismatch related to different mechanisms of multimodal integration, or a stronger reliance on specific modalities, can guide the design of diagnostic tools as well as specific interventions in educational technology.

Overlap and specifics in symptoms of ASD and SLI with respect to multimodal communication has been studied, however not focused on mismatch. ASD children were found impaired in spontaneous gesture- as well as joint attention- behavior while performing tasks and playing with an examiner, compared to SLI children [16]. There are indications that SLI children rely more on gestures than ND children [17], possibly to compensate for a specific impairment of phonological working memory. ASD adolescents have been found to be impaired in their integration of speech with iconic gestures compared to a TD group [18].

There are also examples of structured study of different types of mismatches in multimodal communication. These have however been limited in scope and not concerned with communication disorders. One study varied both “semantic congruency” (same/different action expressed) and (a)synchronicity of speech and gestures and found an effect of congruency on the N400 event-related potential (indicative of semantic integration) as long as the time delay was brief (0-160 ms) [19]. Another study examined the effect of different types of propositional mismatch: “anaphor” (abstract deictic gestures in contradictory directions), “origo” (wrong direction towards/away from self in iconic gestures expressing actions) and “manner” (modulating, rather than mismatching, the manner of an action) [20]. All three types affected subsequent retelling, however no significant differences between the strength of the effects of the types of mismatch were found. Both these studies used video recordings of speakers intentionally performing a mismatched gesture, the former with the picture cropped to conceal the mouth of the speaker.

I intend to instead use (and manipulate) recorded motion-capture and/or eye-tracking data to animate an embodied conversational agent [12,13]. This will enable definition of a wider range of more precisely defined multimodal stimuli, for example application of a delay to hand gestures without delaying facial animation. Comprehension can be measured by performance on a task dependent on multimodal instruction, or on a post-test. (See suggestions for experimental paradigms below.) Effects on semantic integration of different modalities can be detected by modulations of the N400 signal. Perception of the mismatch itself can be self-reported, or possibly measured using eye-tracking. While addressees rarely fixate on gestures [21], there is reason to believe that gaze is drawn to movement that violates prediction. In unpublished previous work by the present author, an experiment revealed that subjects tended to spend more time looking at a virtual hand they controlled when a movement mapping they had previously adapted to was disrupted (see attached presentation slides
During the first year of the project, I plan to specify the types of mismatches to focus on, what predictions can be made and how they can be tested experimentally. This will practically mean literature studies and familiarization with the technical facilities for experimentation. I also believe I can benefit greatly from theoretical knowledge and practical experience of CCL affiliates, as the proposed project aligns with existing projects.

I do however have preliminary ideas for hypotheses to test: First, I would like to test if SLI children are more strongly affected by asynchronicity between speech and gestures. The result could reveal if and how they rely on observation of gestures to compensate for an impaired phonological working memory. Second, ASD children should be less sensitive to mismatched gestures that require “imagining” mental states in the speaker (e.g. abstract deictic gestures [3]) than those that do not (e.g. deictic gestures referring to external objects). ASD children have previously been shown to benefit from concretization when performing tasks requiring “theory of mind” [22].

I plan to adapt an existing experimental paradigm for testing comprehension by integrating mentioned matched/mismatched motion-capture animated embodied conversational agents [12], and measuring semantic integration and perception of the mismatch as described above. One option is to use a "picture matching test", where subjects have to select which one out of a set of pictures best matches a verbal description (in this case to accompanied by gestures). This test has been used to formally assess grammatical comprehension [23] as well as to investigate speech-gesture integration [18]. Another possibility is to have subjects retell a narrative or explanation by the agent, and code their verbal explanations as well as gesture use for inconsistencies with the original message [20]. Yet another possibility is to use a pre- and post-tests to evaluate comprehension of a conceptual explanation [24].

The plan for the second and third years is to define, pilot and execute experiments to test the hypotheses established during the first year, and then to analyze and publish the results. The fourth year will be dedicated to thesis writing, and possible follow-up experiments based on what I learn from the previous two years. Ideally a follow-up experiment could be a validation of a principle learned from the previous experiments that can been integrated into a diagnostic tool or educational application.
10. Alibali, Martha W, Dana C Heath, and Heather J Myers. "Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen." Journal of Memory and Language 44.2 (2001): 169-188.