Speech perception integrates signal from ear and eye. This is witnessed by a wide range of audiovisual integration effects, such as ventriloquism and the McGurk illusion. Some behavioral evidence suggest that audiovisual integration of specific aspects is special for speech perception. However, our knowledge of such bimodal integration would be strengthened if the phenomena could be investigated by objective, neutrally based methods. One key question of the present work is if perceptual processing of audiovisual speech can be gauged with a specific signature of neurophysiological activity, the mismatch negativity response (MMN).

MMN has the property of being evoked when an acoustic stimulus deviates from a learned pattern of stimuli. In three experimental studies, this effect is utilized to track when a coinciding visual signal alters auditory speech perception. Visual speech emanates from the face of the talker. Perception of faces and of speech shares the trait, that they are learned from infancy and seemingly specialized behaviorally and neurally. Due to this, speech and face encoding functions quasi-automatically and with high efficiency. However, perhaps owing to our long experience with human faces, which all are variations on a relatively constrained space of features, face perception is sensitive to manipulations of the structure of the face, the relation between its segments, and the properties of the segments. Does this sensitivity alter the influence of visual speech on the auditory speech percept? In two experiments, which both combine behavioral and neurophysiological measures, an uncovering of the relation between perception of faces and of audiovisual integration is attempted. Behavioral findings suggest a strong effect of face perception, whereas the MMN results are less clear.

Another interesting property of speech perception is that it is relatively tolerant towards temporal shifts between acoustic and visual speech signals. Here, behavioral studies report that perception of speech exhibits far greater temporal tolerance than towards non-speech stimuli. Current findings on neural correlates of this tolerance, however, are few and limited. Here, a novel experimental MMN paradigm is used in effort to shed light on integration asynchronous audiovisual speech. Based on individual behavioral estimates of temporal windows of tolerance, we ask if the MMN signal can be evoked at different points within and outside this window. Behavioral findings match earlier behavioral studies, whereas the MMN findings are ambiguous.

In conclusion, the work presented here sheds light onto two important aspects of speech perception. It also presents important methodological conclusions on the use of MMN as a neural marker of audiovisual integration.