Face configuration affects speech perception: Evidence from a McGurk mismatch negativity study

We perceive identity, expression and speech from faces. While perception of identity and expression depends crucially on the configuration of facial features it is less clear whether this holds for visual speech perception.

Facial configuration is poorly perceived for upside-down faces as demonstrated by the Thatcher illusion in which the orientation of the eyes and mouth with respect to the face is inverted (Thatcherization). This gives the face a grotesque appearance but this is only seen when the face is upright.

Thatcherization can likewise disrupt visual speech perception but only when the face is upright indicating that facial configuration can be important for visual speech perception. This effect can propagate to auditory speech perception through audiovisual integration so that Thatcherization disrupts the McGurk illusion in which visual speech perception alters perception of an incongruent acoustic phoneme. This is known as the McThatcher effect.

Here we show that the McThatcher effect is reflected in the McGurk mismatch negativity (MMN). The MMN is an event-related potential elicited by a change in auditory perception. The McGurk-MMN can be elicited by a change in auditory perception due to the McGurk illusion without any change in the acoustic stimulus.

We found that Thatcherization disrupted a strong McGurk illusion and a correspondingly strong McGurk-MMN only for upright faces. This confirms that facial configuration can be important for audiovisual speech perception. For inverted faces we found a weaker McGurk illusion but, surprisingly, no MMN. We also found no correlation between the strength of the McGurk illusion and the amplitude of the McGurk-MMN. We suggest that this may be due to a threshold effect so that a strong McGurk illusion is required to elicit the McGurk-MMN.
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.

**General information**

State: Published
Organisations: Department of Applied Mathematics and Computer Science, Cognitive Systems, Hearing Systems, Department of Electrical Engineering
Authors: Eskelund, K. (Intern), Andersen, T. (Intern), MacDonald, E. (Intern), Dau, T. (Intern)
Number of pages: 148
Publication date: 2014

**Publication information**

Place of publication: Kgs. Lyngby
Publisher: Technical University of Denmark (DTU)
Original language: English

Series: DTU Compute PHD-2014
Number: 341
ISSN: 0909-3192
Main Research Area: Technical/natural sciences

Electronic versions: phd341_Eskelund_K_2.pdf
Publication: Research › Ph.D. thesis – Annual report year: 2014

Mouth reversal extinguishes mismatch negativity induced by the McGurk illusion

The sight of articulatory mouth movements (visual speech) influences auditory speech perception. This is demonstrated by the McGurk illusion in which incongruent visual speech alters the auditory phonetic percept. In behavioral studies, reversal of the vertical mouth direction has been reported to greatly reduce the McGurk illusion (Rosenblum et al., 2000). Such findings support the idea that audiovisual integration in speech to some extent relies on information regarding facial configuration. Here we ask whether this behavioral effect is reflected in a difference in neural activity in the auditory cortex. Mismatch Negativity (MMN) is a component in the auditory Event-Related Potential (ERP) that is elicited by a change in the auditory percept. It has been shown that the McGurk illusion can induce a MMN. We conducted an experiment in which the MMN could be elicited by the McGurk illusion induced by visual speech with either upright (unaltered) or vertically reversed mouth area. In a preliminary analysis, we found a Mismatch Negativity component induced by the McGurk illusion for 6 of 17 participants at electrode Cz when the mouth area was upright. In comparison, these participants produced no Mismatch Negativity when the mouth was reversed. These findings mirrored behavioral findings in the same subjects of a strong McGurk response for normal audiovisual speech, which was greatly reduced for stimuli with reversed mouth area.

**General information**

State: Published
Organisations: Department of Applied Mathematics and Computer Science, Cognitive Systems, Hearing Systems, Centre for Applied Hearing Research, Department of Electrical Engineering
Authors: Eskelund, K. (Intern), Andersen, T. (Intern)
Pages: 133–134
Publication date: 2013
Conference: 14th International Multisensory Research Forum (IMRF 2013), Jerusalem, Israel, 03/06/2013 - 03/06/2013
Main Research Area: Technical/natural sciences

**Publication information**

Journal: Multisensory Research
Volume: 26
Issue number: 0
Audiovisual integration in speech perception: a multi-stage process
Integration of speech signals from ear and eye is a well-known feature of speech perception. This is evidenced by the McGurk illusion in which visual speech alters auditory speech perception and by the advantage observed in auditory speech detection when a visual signal is present. Here we investigate whether the integration of auditory and visual speech observed in these two audiovisual integration effects are specific traits of speech perception. We further ask whether audiovisual integration is undertaken in a single processing stage or multiple processing stages.

General information
State: Published
Organisations: Department of Informatics and Mathematical Modeling, Cognitive Systems, University College London
Authors: Eskelund, K. (Intern), Tuomainen, J. (Ekster), Andersen, T. (Intern)
Publication date: 2011

Host publication information
Title of host publication: Proceedings of the International Symposium on Auditory and Audiological Research 2011
Main Research Area: Technical/natural sciences
Electronic versions:
ISAAR_EskelundTuomainenAndersen.pdf
Links:
http://www.isaar.eu/
Source: orbit
Source-ID: 284470
Publication: Research - peer-review › Article in proceedings – Annual report year: 2011

Multistage audiovisual integration of speech: dissociating identification and detection
Speech perception integrates auditory and visual information. This is evidenced by the McGurk illusion where seeing the talking face influences the auditory phonetic percept and by the audiovisual detection advantage where seeing the talking face influences the detectability of the acoustic speech signal. Here we show that identification of phonetic content and detection can be dissociated as speech-specific and non-specific audiovisual integration effects. To this end, we employed synthetically modified stimuli, sine wave speech (SWS), which is an impoverished speech signal that only observers
informed of its speech-like nature recognize as speech. While the McGurk illusion only occurred for informed observers
the audiovisual detection advantage occurred for naïve observers as well. This finding supports a multi-stage account of
audiovisual integration of speech in which the many attributes of the audiovisual speech signal are integrated by separate
integration processes.

General information
State: Published
Organisations: Cognitive Systems, Department of Informatics and Mathematical Modeling
Authors: Eskelund, K. (Intern), Tuomainen, J. (Ekstern), Andersen, T. (Intern)
Pages: 447-457
Publication date: 2011
Main Research Area: Technical/natural sciences

Publication information
Journal: Experimental Brain Research
Volume: 208
ISSN (Print): 0014-4819
Ratings:
BFI (2018): BFI-level 1
BFI (2017): BFI-level 1
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 1.97 SJR 0.964 SNIP 0.823
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.157 SNIP 0.878 CiteScore 2.08
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.263 SNIP 1.009 CiteScore 2.27
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.496 SNIP 1.068 CiteScore 2.55
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.386 SNIP 1.048 CiteScore 2.48
ISI indexed (2012): ISI indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.455 SNIP 1.126 CiteScore 2.61
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 1.311 SNIP 1.03
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 1.388 SNIP 1.139
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 1.336 SNIP 0.992
Scopus rating (2007): SJR 1.254 SNIP 1.023
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 1.153 SNIP 1.125
Scopus rating (2005): SJR 1.354 SNIP 1.126
Scopus rating (2004): SJR 1.483 SNIP 1.075
Scopus rating (2003): SJR 1.531 SNIP 1.113
Scopus rating (2002): SJR 1.533 SNIP 1.167
Scopus rating (2001): SJR 1.326 SNIP 1.091
Scopus rating (2000): SJR 1.333 SNIP 1.138
Scopus rating (1999): SJR 1.305 SNIP 1.071
Original language: English
Speech-specificity of two audiovisual integration effects

Seeing the talker's articulatory mouth movements can influence the auditory speech percept both in speech identification and detection tasks. Here we show that these audiovisual integration effects also occur for sine wave speech (SWS), which is an impoverished speech signal that naïve observers often fail to perceive as speech. While audiovisual integration in the identification task only occurred when observers were informed of the speech-like nature of SWS, integration occurred in the detection task both for informed and naïve observers. This shows that both speech-specific and general mechanisms underlie audiovisual integration of speech.

General information
State: Published
Organisations: Cognitive Systems, Department of Informatics and Mathematical Modeling, University College London
Authors: Eskelund, K. (Intern), Tuomainen, J. (Ekstern), Andersen, T. (Intern)
Publication date: 2010

Host publication information
Title of host publication: Proceedings of AVSP
ISBN (Print): 978-4-9905475-0-9
Main Research Area: Technical/natural sciences
Conference: International Conference on Audio-Visual Speech Perception (AVSP) : Tuomainen, Hakone, Kanagawa, Japan, 01/01/2010
Source: orbit
Source-ID: 275433
Publication: Research - peer-review › Article in proceedings – Annual report year: 2010

Specialization in audiovisual speech perception: a replication study
Speech perception is audiovisual as evidenced by bimodal integration in the McGurk effect. This integration effect may be specific to speech or be applied to all stimuli in general. To investigate this, Tuomainen et al. (2005) used sine-wave speech, which naïve observers may perceive as non-speech, but hear as speech once informed of the linguistic origin of the signal. Combinations of sine-wave speech and incongruent video of the talker elicited a McGurk effect only for informed observers. This indicates that the audiovisual integration effect is specific to speech perception. However, observers might only have been motivated to look at the face when informed and audio and video thus seemed related. Since Tuomainen et al. did not control for this, the influence of motivation is unknown. The current experiment repeated the original methods while controlling eye movements. 4 observers participated in the experiment, which consisted of 3 conditions. In the non-speech condition, observers were trained and tested in their ability to categorize sine wave speech tokens in arbitrary categories. The natural speech condition was similar but used natural speech signals and observers categorized phonetic content. The speech-mode condition again used sine-wave speech stimuli but observers were informed of the speech-like nature of the stimuli and classified them according to the phonetic content. In all conditions, a white square which in some trials dimmed briefly was overlaid on the nose of the speaker. Observers were required to report this after primary target categorization. We found a significant McGurk effect only in the natural speech and speech mode conditions supporting the finding of Tuomainen et al. Performance in the secondary task was similar in all conditions indicating that observers did look near the mouth. We conclude that eye-movements did not influence the results of Tuomainen et al. and that their results thus can be taken as evidence of a speech specific mode of audiovisual integration underlying the McGurk illusion.

General information
State: Published
Organisations: Cognitive Systems, Department of Informatics and Mathematical Modeling
Authors: Eskelund, K. (Intern), Andersen, T. (Intern)
Publication date: 2009
Event: Poster session presented at International Multinational Research Forum, New York, USA, .
Main Research Area: Technical/natural sciences
Source: orbit
Speech-specific audiovisual perception affects identification but not detection of speech

Speech perception is audiovisual as evidenced by the McGurk effect in which watching incongruent articulatory mouth movements can change the phonetic auditory speech percept. This type of audiovisual integration may be specific to speech or be applied to all stimuli in general. To investigate this issue, Tuomainen et al. (2005) used sine-wave speech stimuli created from three time-varying sine waves tracking the formants of a natural speech signal. Naïve observers tend not to recognize sine wave speech as speech but become able to decode its phonetic content when informed of the speech-like nature of the signal. The sine-wave speech was dubbed onto congruent and incongruent video of a talking face. Tuomainen et al. found that the McGurk effect did not occur for naïve observers, but did occur when observers were informed. This indicates that the McGurk illusion is due to a mechanism of audiovisual integration specific to speech perception. However, the results of Tuomainen et al. might have been influenced by another effect. When observers were naïve, they had little motivation to look at the face. When informed, they knew that the face was relevant for the task and this could increase their motivation for looking at the face. Since Tuomainen et al. did not monitor eye-movements in their experiments the magnitude of the effect of motivation is unknown. The purpose of our first experiment was to replicate Tuomainen et al.’s findings while controlling observers’ eye movements using a secondary visual detection task. In our first experiment, observers presented with congruent and incongruent audiovisual sine-wave speech stimuli did only show a McGurk effect when informed of the speech nature of the stimulus. Performance on the secondary visual task was very good, thus supporting the finding of Tuomainen et al. In a second experiment, we further investigated speech-specific audiovisual integration by testing if the speech-specific audiovisual perceptual mode found in experiment 1 would be advantageous in detecting a speech signal in noise. Thresholds for detecting sine-waves speech in noise were measured for naïve and informed participants. We found that the threshold for detecting speech in audiovisual stimuli was lower than for auditory-only stimuli. But there was no detection advantage for observers informed of the speech nature of the auditory signal. This may indicate that identification and detection of audiovisual speech draw on separate processes. Reference: Tuomainen, J., Andersen, T., Tiippana, K., & Sams, M. (2005). Audio-visual speech perception is special. Cognition, 96(1), B13-B22.