View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Online Research Database In Technology

Technical University of Denmark

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

Eskelund, Kasper; MacDonald, Ewen; Andersen, Tobias

Published in:
Neuropsychologia

Link to article, DOI:
10.1016/j.neuropsychologia.2014.10.021

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Eskelund, K., MacDonald, E., & Andersen, T. (2015). Face configuration affects speech perception: Evidence from a McGurk mismatch negativity study. Neuropsychologia, 66, 48-54. DOI: 10.1016/j.neuropsychologia.2014.10.021

DTU Library
Technical Information Center of Denmark

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Neuropsychologia 66 (2015) 48–54

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Research Report

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

Kasper Eskelund [a,b,*,1], Ewen N. MacDonald [b,c], Tobias S. Andersen [a,b]

[a] Section for Cognitive Systems, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark
[b] CHeSS, Oticon Centre of Excellence for Hearing and Speech Sciences, Technical University of Denmark, Denmark
[c] Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, Denmark

ARTICLE INFO

Article history:
Received 25 May 2014
Received in revised form
23 September 2014
Accepted 14 October 2014
Available online 24 October 2014

Keywords:
Multisensory
Audiovisual
Speech perception
Face perception
Mismatch negativity
EEG

ABSTRACT

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.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

1. Introduction

Face perception has three important functions: face recognition, perception of facial expression and visual speech perception (cf. Bruce and Young, 2012). Face perception is special, differing from perception of other objects in a number of ways. Perhaps the most notable of these is the strong dependence of face recognition and perception of facial expression not only on features such as the mouth, eyes and nose but also, to a larger degree, on their configuration (Farah et al., 1998; Valentine, 1988).

Whether visual speech perception, as the third major function of face perception, is also dependent on configuration information is less clear. Understanding visual speech perception is particularly interesting because of the effect that automatic, subconscious speech reading has on auditory speech perception in face-to-face conversation. Evidence for this effect comes from studies showing that seeing the interlocutor's face facilitates speech perception (Sumby and Pollack, 1954) and from studies of the McGurk illusion. In the McGurk illusion (McGurk and MacDonald, 1976), an auditory phonetic percept is altered by seeing an incongruent visual phoneme. The resulting, illusory auditory percept may represent a combination of the incongruent acoustic and visual stimuli (e.g. acoustic /ga/+visual /ba/ producing an illusory percept /bga/). Or, it may produce a fusion percept, a third phoneme absent in either stimulus (e.g. acoustic /ba/ +visual /ga/ producing an illusory percept /da/). Finally, the visual phoneme may dominate the auditory percept (e.g. acoustic /ba/ +visual /ga/ producing an illusory percept /ga/). The automaticity and robustness of the McGurk effect is in stark contrast to the difficulty with which untrained observers speech read (Walden et al., 1977). This indicates that audiovisual speech perception can

* Corresponding author.
E-mail address: kaspereskelund@gmail.com (K. Eskelund).
[1] Kasper Eskelund was funded by the Oticon Foundation.

http://dx.doi.org/10.1016/j.neuropsychologia.2014.10.021
0028-3932/© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

be based on visual cues that are not directly accessible to most observers. Therefore the strength of the McGurk illusion is a good measure for the accuracy of perception of visual speech—perhaps even better than direct measures of speech reading ability. This has been the reason for several studies of configuration information in speech reading to study audiovisual, in addition to, visual speech perception (e.g. Rosenblum et al., 2000).

It is clear that visual and audiovisual speech perception rely heavily on feature information mainly from the lips, tongue and teeth as seeing only the mouth area is sufficient for speech reading and for eliciting the McGurk illusion (Hietanen et al., 2001; Jordan and Thomas, 2011; Rosenblum et al., 2000). Nevertheless, somewhat surprisingly, speech can also be read from faces even when the mouth area is entirely occluded and this can influence audiovisual speech perception (Jordan and Thomas, 2011). This effect is due to the fact that movements of extraoral face areas are correlated with movements of the mouth and articulators (Jordan and Thomas, 2011). Thus, the spatial relationship of these oral and extraoral features is a candidate for configuration information that may carry visual speech information.

Hietanen et al. (2001) examined the effect of configurational information in a very direct manner. They created visual stimuli consisting only of the eyes, nose and mouth by masking the rest of the face. The location of these facial features was either in their natural position or scrambled. While some effects of feature scrambling on the strength of the McGurk illusion were found, the effects were weak and dependent on speaker identity. Still, the study supports the notion that feature configuration can influence audiovisual speech perception.

Facial configuration has been shown to be difficult to perceive in inverted faces. Hence, face recognition (Farah et al., 1998; Valentine, 1988) and perception of facial expression (Prkachin, 2003) is impaired for inverted faces. Several studies have found face inversion effects for visual and audiovisual speech perception (Jordan and Bevan, 1997; Massaro and Cohen, 1996; Rosenblum et al., 2000). Some of these studies found strong effects and others none. The overall conclusion seems to be that the face inversion effect depends greatly on the visual stimulus as it can vary across speakers even when they articulate the same speech sounds. Thomas and Jordan (2002) extended this approach by examining the effect of different levels of visual blurring. They hypothesized that since feature information depends on higher resolution than configurational information (Goffaux and Rossion, 2007) observers must rely more on configuration information when the face is blurred. Thus, blurring should lead to a greater effect of inverting the orientation of the face. Their findings confirmed this hypothesis for speech reading, as well as for congruent and incongruent audiovisual speech.

Thompson (1980) devised a striking demonstration of our inability to perceive facial configuration in inverted faces, using a photograph of Margaret Thatcher. Misconfiguration, by vertical inversion of the mouth and eye segments (so-called Thatcherization), renders the face strikingly grotesque but this is only perceived when the face is upright and not when it is inverted (cf. Fig. 1). Thus the Thatcher illusion shows that configuration information is less effective when the face is presented upside down (Bartlett and Searcy, 1993; Bruce and Young, 2012; Carbon et al., 2005). Rosenblum et al. (2000) found that misconfiguration by Thatcherization could greatly reduce the strength of the McGurk illusion but only when the face was upright. However, this effect was not driven by inversion of the mouth segment, as it did not occur when the mouth segment was presented in isolation. These findings form strong support for configuration information being important for visual and audiovisual speech perception. Rosenblum and colleagues named this striking effect of face configuration on speech perception the McThatcher effect (Rosenblum, 2001).

In Rosenblum et al. (2000), the McThatcher effect was specific to certain phonemes just as the face inversion effect has been in most studies. For audiovisual stimuli, it was only for the visual dominance illusion of hearing acoustic /ba/+visual /va/ as /va/ that the full effect occurred. This indicates that facial configuration is more important for some phonemes than others. Thomas and Jordan (2002) came to the same conclusion noticing that the difference between visual /ga/ and /da/ is mostly visible in the oral cavity. Accordingly, this contrast seems less influenced by the face inversion effect and the McThatcher effect.

To summarize previous findings, we find that, on one side, many of them suggest an effect of facial configuration on speech perception but on the other, that the effects are highly variable and sensitive to details in the stimuli. Although deterred by this variability, we found the motivation for the current study in the power and usefulness of the McThatcher effect for investigating the relation between encoding of facial configuration and perception of audiovisual speech.

In the current study, we seek to find neural correlates of the McThatcher effect. If facial configuration truly influences audiovisual speech perception then it should be reflected in auditory evoked potentials such as the mismatch negativity (MMN, Näätänen et al., 1978). In its most basic form, the MMN is elicited by a deviant stimulus (e.g. a 1200 Hz tone) after a sequence of standard stimuli (e.g. 1000 Hz tones). Average ERPs due to deviant stimuli exhibit a negative deflection in the interval 100–250 ms covering a wide area of fronto-central electrodes. An MMN response can be produced by a noticeable deviance in a wide variety of acoustic features (pitch, intensity, duration, modulation or phoneme), and the magnitude of the negative deflection varies with the magnitude of the perceived difference (Näätänen and Alho, 1995; Näätänen et al., 2004). Although the MMN reflects early pre-attentive auditory perception, it is also evoked by visually induced auditory illusions, such as ventriloquism (Stekelenburg et al., 2004) and the McGurk illusion (Colin, 2002; Ponton et al., 2009; Saint-Amour et al., 2007; Sams et al., 1991; Stekelenburg and Vroomen, 2012). In typical McGurk-MMN paradigms, congruent audiovisual syllables (e.g. auditory /ba/+visual /ba/) are presented as standards, whereas incongruent (McGurk type) stimuli are deviants (e.g. auditory /ba/+visual /va/) (Colin, 2002; Stekelenburg and Vroomen, 2012; for a different method cf. Kislyuk et al., 2008). In such McGurk-MMN paradigms, stimulus deviance is only present in the visual signal. Thus, it is an auditory differential response evoked by the incongruent visual speech signal (i.e. the McGurk illusion), which produces the audiovisual McGurk-MMN response.

In the current study, we measured the McGurk-MMN for normal and Thatcherized faces with either upright or inverted orientation. We used the congruent audiovisual syllable /ba/ as the standard stimulus and the incongruent audiovisual combination of acoustic /ba/+visual /va/ as deviant stimulus as these were the phonemes for which Rosenblum et al. (2000) found the effect to be the strongest. To ensure that the McThatcher effect occurs for these specific stimuli, we also replicate Rosenblum et al.'s behavioral paradigm. Our hypothesis is that the McGurk-MMN will mirror behavioral findings and confirm the effect as being a truly perceptual effect. As the amplitude of the MMN is known to increase with perceived stimulus difference (Garrido et al., 2009; May and Tiitinen, 2010; Näätänen et al., 1978, 2004) we expect MMN amplitudes to be correlated with levels of behavioral McGurk responses.

2. Methods

2.1. Subjects

19 subjects (11 females) with a mean age of 24 years (range 18–38) participated in the experiment. MMN is known to show high inter-individual variability (Lang et al., 1995). Therefore, as the present study targets differences in McGurk-MMN with manipulated visual speech, we defined an exclusion criterion on basis of a recording of

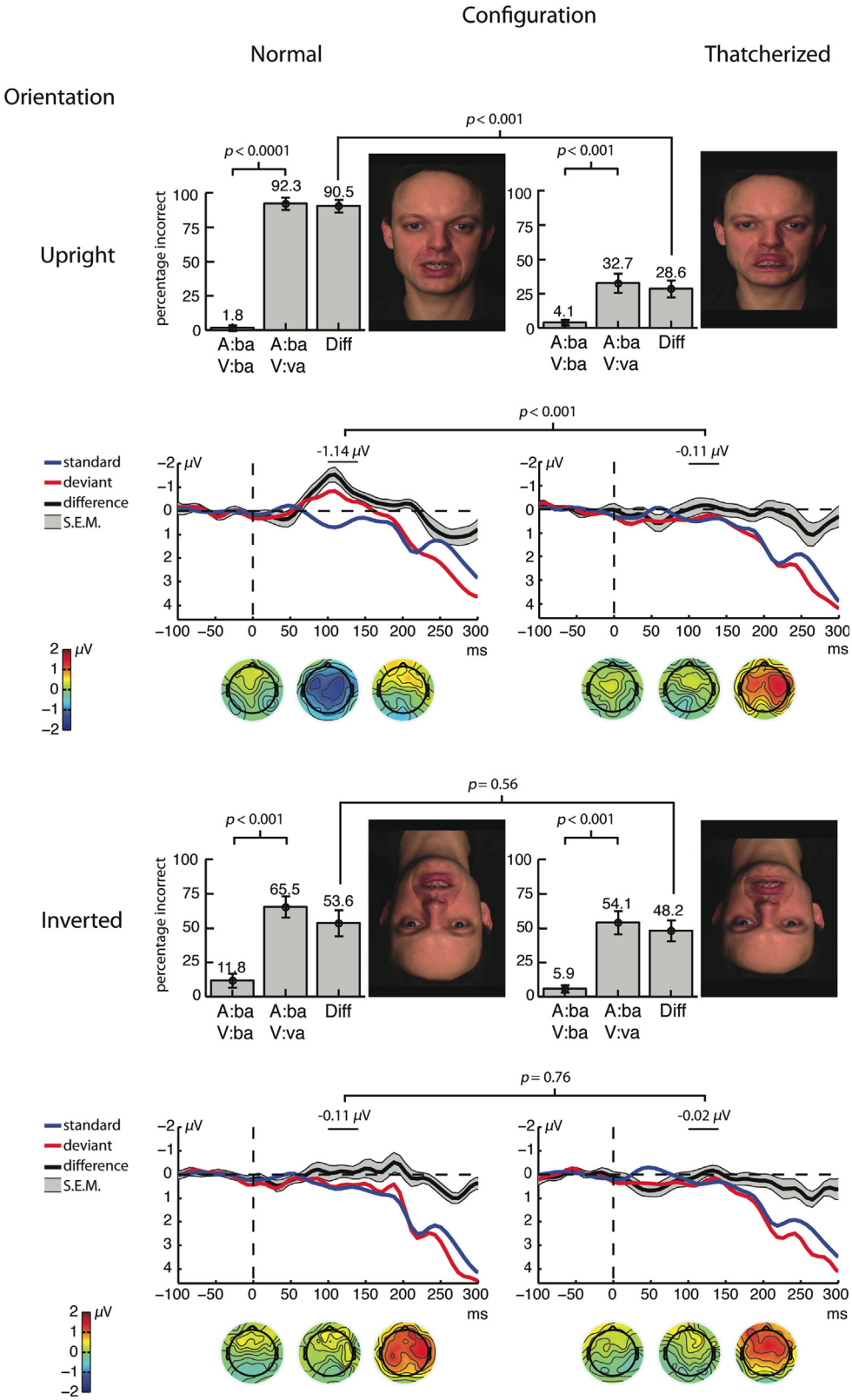

Fig. 1. Stimulus, response percentages, ERPs and scalp-maps for each of the four stimulus conditions. Stimulus: still frame from video showing maximal mouth opening. Response percentages: bar plot showing the percentage of incorrect responses as a measure of the strength of the McGurk illusion for congruent (A:ba/V:ba) and incongruent (A:ba/V:va) stimuli and the difference between them (Diff). Error-bars indicate standard error of mean. Shown are also *P*-values from statistical tests described in main text. ERPs: ERPs for standard and deviant stimuli and their difference, recorded at electrode Fz. Time zero is fixed at the voicing onset in the acoustic stimulus /ba/. Shaded area indicates standard error of the mean. The horizontal floating bar marks the interval 100–140 ms post-stimulus and the value above it indicate the average amplitude in that interval. Shown are also *P*-values from statistical tests described in main text. Scalp-maps show interpolated mean MMN amplitude at 0, 100 and 200 ms.

pure-tone MMN (Näätänen et al., 1978), as to reduce noise in our dataset by excluding subjects with a generally weak MMN response. For all subjects, acoustic MMN was recorded for 1000 Hz (standard) and 1200 Hz (deviant) tones of 100 ms duration with a SOA of 500 ms presented at 60 dB(A) SPL. The rate of deviant stimuli was 15% and a total of 1200 trials were presented. Subjects whose pure-tone MMN did not exceed −1 μV in the 100–200 ms interval were excluded. On basis of this, 8 subjects (5 female) were excluded.

2.2. Stimuli

Stimuli were generated from a video recording of syllables /ba/ and /va/. Each video was recorded at 30 fps and lasted 30 frames. Sound was recorded at 22.05 kHz sampling rate and 16-bit depth. The two visual speech tokens were edited in Adobe Premiere Elements 10 to produce the following visual manipulations of each: normal configuration, upright orientation; Thatcherized configuration, upright orientation; normal configuration, inverted orientation; Thatcherized configuration, inverted orientation (see Fig. 1). These eight visual speech tokens (/ba/ and /va/ × four visual manipulations) were combined with the acoustic /ba/ in Adobe Premiere Elements 10 to produce four congruent and four incongruent audiovisual speech stimuli.

McGurk-driven MMN responses may be confounded by purely visual responses to the visual speech signal. This is a problem in particular when studying perception of audiovisual speech compared to unimodal speech. In such studies, it is common practice to record ERPs due to the visual speech stimulus alone, and subsequently correct the audiovisual ERPs with these (cf. e.g. Colin, 2002; Möttönen et al., 2002; Sams et al., 1991). In contrast, the current experiment compares changes in the McGurk-driven MMN across four audiovisual conditions. Thus, in these audiovisual conditions, the visual response should be equal, eliminating the necessity of a correction for visual activation.

Subjects were seated in a comfortable armchair in a dimly lit, shielded EEG booth at a distance of 1.2 m from the visual display. Visual stimuli were presented on a 19 in. ViewSonic G90F CRT screen at a 60 Hz refresh rate. Sound was presented with a single Genelec 6010B monitor speaker positioned directly beneath the visual display, at an intensity of 60 dB (A) SPL measured at the head position of the subject. Stimulus presentation was controlled with Psychophysics Toolbox 3.0 (Kleiner et al., 2007).

2.3. Behavioral experiment

The behavioral task consisted of a random presentation of 20 repetitions of each of the eight audiovisual stimuli. After each trial, subjects were prompted to identify what they just heard in response categories “ba”, “da”, “fa”, or, “va”.

2.4. EEG experiment

A BioSemi ActiveTwo 64-channel EEG system referenced to the mean of two mastoid electrodes was used for recording EEG. Data were sampled at 512 Hz. EEG measurements were recorded in four conditions, each employing one of the four manipulations of the visual stimulus (cf. Fig. 1). Each condition was split into two blocks each containing an oddball sequence of 600 trials for a total of 1200 trials in each condition. The oddball sequence consisted of 85% standards, which were the audiovisual congruent syllable /ba/+/ba/, and 15% deviants, which were the incongruent audiovisual syllable /ba/+/va/. Stimuli were presented with a constant intertrial interval of 100 ms, during which there was a crossfading between the last frame of the preceding stimulus and the first frame of the following. The sequence was randomized with the condition that at least two standards succeeded each deviant. Each block was preceded by 30 presentations of the standard stimulus. No data collected during those trials were used in the analysis. The sequence of blocks was randomized with the constraint that blocks presenting every condition were presented once, before any block was presented a second time.

2.5. EEG preprocessing

Analysis of EEG data was performed within the EEGLAB toolbox (Delorme and Makeig, 2004). First, continuous EEG data were bandpass filtered between 1 and 30 Hz (for similar filtering choices, cf. e.g. Möttönen et al., 2002; Näätänen et al., 2004; Sams et al., 1991; Stekelenburg and Vroomen, 2012), before downsampling to 128 Hz. After filtering, data were epoched in the interval −100 to 600 ms with auditory onset at 0. Epochs were baselined to the 100 ms preceding auditory onset. Electrodes dominated by unusual, non-biological waveforms were selected by a measure of kurtosis and data in these channels was interpolated from surrounding electrodes. An ICA algorithm (runica) was used (not including interpolated channels) to prepare data for rejection of independent components generated by eye artifacts, by means of the EyeCatch algorithm (Bigdely-Shamlo et al., 2013). Epochs were finally thresholded at ±100 μV to remove remaining artifacts. The proportion of epochs removed from any subject's dataset during preprocessing did not exceed 2%. ERPs were generated by averaging the preprocessed data epochs. Individual MMN waveforms were computed by subtracting average ERPs due to standard stimuli from mean ERPs due to deviant stimuli (cf. Fig. 1).

3. Results

3.1. Behavioral experiment

Responses from the behavioral task were re-categorized as correct (“ba”) and incorrect (all other responses). Among the incorrect responses to the acoustic /ba/, categories “fa” and “va”, which are visually indistinguishable, were clearly dominant, while the response category “da” only accounted for 1.2% of all incorrect auditory identifications. We use the percentage of incorrect responses as the independent variable (cf. Fig. 1).

Responses were arcsine-transformed to correct for heterogeneity of variances and subjected to a three-way, repeated-measures ANOVA with factors Orientation × Thatcherization × Congruence. Factor Orientation had two levels (Normal and Inverted), factor Configuration had two levels (Normal and Thatcherized), and factor Congruence had two levels (Congruent and Incongruent). The analysis revealed that the interaction between the three factors was significant ($F(1,10)=40.7$, $P<0.0001$).

We proceeded to perform two-way, repeated-measures ANOVAs with factors Configuration × Congruence on data from the two Orientation conditions. For upright stimuli, the interaction between Configuration and Congruence was significant ($F(1,10)=137.2$, $P<0.0001$), indicating that the conflicting direction of the mouth segment did reduce audiovisual integration. For inverted stimuli, however, the interaction between Configuration and Congruence was not significant ($F(1,10)=0.58$, $P=0.56$). This suggests that Thatcherization did not alter audiovisual integration when in the context of an inverted face.

Subsequently, we performed two-way, repeated-measures ANOVAs with factors Orientation × Congruence on data from the two Thatcherization conditions. Here, normally configured stimuli revealed a significant interaction between Orientation and Congruence ($F(1,10)=20.5$, $P<0.01$), as did Thatcherized stimuli ($F(1,10)=7.9$, $P<0.05$). This indicates that Orientation influenced audiovisual integration both when the face was Thatcherized and when it was not.

Interestingly, while inverting the orientation of the face reduced audiovisual integration for stimuli with normal configuration, it improved audiovisual integration for Thatcherized stimuli. This could indicate a role for the orientation of the mouth segment. To investigate this, we conducted a separate two-way, repeated-measures ANOVA on normal, upright and Thatcherized, inverted stimuli, which share direction of the mouth segment but within either a matching or conflicting facial orientation. The difference in the strength of the McGurk illusion was significant ($F(1,10)=31.36$, $P<0.001$) suggesting that even with a shared mouth segment direction, the two facial contexts still influenced audiovisual speech perception differently.

Individual repeated-measures ANOVAs on Congruence revealed that the difference between performance with congruent and incongruent stimuli was significant for all visual stimulus types (normal configuration, upright $F(1,10)=480.6$, $P<0.0001$; Thatcherized, upright $F(1,10)=25.0$, $P<0.001$; normal configuration, inverted $F(1,10)=35.5$, $P<0.001$; Thatcherized, inverted $F(1,10)=40.1$, $P<0.001$). Thus, we found a significant McGurk illusion in all four stimulus conditions.

3.2. Mismatch negativity experiment

We subjected the 0–200 ms interval of the difference wave to a repeated-measures, one-tailed clustered permutation test with 2500

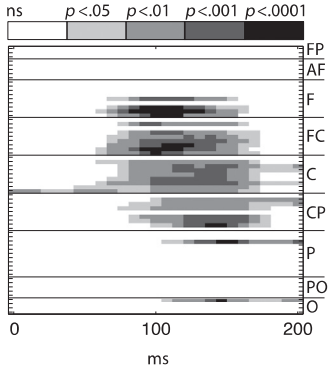

Fig. 2. Plot of significance of the McGurk-MMN for upright stimuli with normal facial configuration based on a repeated-measures, one-tailed clustered permutation test. Labels on the *y*-axis indicate frontal to anterior electrode clusters ranging from prefrontal (FP) to occipital (O). Electrodes are arranged from left (top) to right (bottom) within clusters.

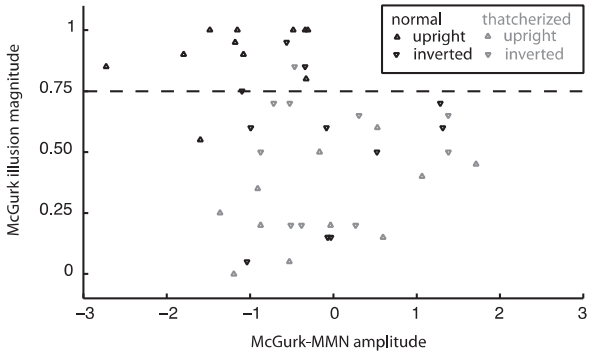

Fig. 3. Correlation of audiovisual MMN amplitude and behavioral audiovisual integration response for individual subjects. *X*-axis represents mean amplitude of the McGurk-MMN (difference wave) at electrode Fz. *Y*-axis represents the strength of the McGurk illusion measured as the percentage of incorrect responses in the incongruent condition minus the percentage of incorrect responses in the congruent condition. The dashed line indicates a threshold for the strength of the McGurk illusion. The McGurk-MMN is consistently negative above this threshold while very variable below the threshold.

permutations (for a detailed description cf. Groppe et al., 2011). We used a family-wise alpha level of 0.05 for determining statistical significance. Also, a one-tailed test was used, as any effect due to mismatch negativity would only be on the negative tail.

Only the upright face with normal configuration produced a difference wave that was significantly less than 0 at any time-point in any channel. In this condition, the difference wave recorded at a large ensemble of centro-parietal, central and fronto-central electrodes exceeded a *P*-value of 0.05 (cf. Fig. 2) during extended contiguous periods. Notably, the topographical distribution of the difference wave is centered over frontal and central electrodes. This is typical for auditory MMN (cf. e.g. Garrido et al., 2009; Näätänen and Alho, 1995), whereas differential potentials produced by visual-only MMN paradigms are centered at occipital and parietal sites (cf. e.g. Czigler, 2007; Stefanics et al., 2011).

The remaining conditions did not produce any MMN as their difference waves were not significantly below 0 in the target interval (0 to 200 ms). This suggests that the audiovisual MMN response is highly sensitive to both orientation inversion and Thatcherization.

We proceeded to compare difference waves from the four stimulus conditions. For this comparison, we again chose electrode Fz, which is a commonly used site for location of MMN in both auditory and audiovisual paradigms (Colin, 2002; Garrido et al., 2009; Näätänen and Alho, 1995; Stekelenburg and Vroomen, 2012). We extracted the mean amplitude in the 100–140 ms interval as a measure of mismatch negativity (Fig. 1). These values were subjected to a Wilcoxon signed-rank test, to test for differences across the difference waves in the four stimulus conditions. Here, the normal, upright condition proved significantly different from the other three conditions ($P < 0.02$ for each comparison). Comparisons not including the normal, upright condition did not yield any significant difference ($P > 0.70$ for each comparison). This again suggests a high sensitivity of the McGurk-MMN to Thatcherization and inversion.

In order to investigate the apparent discrepancy between the MMN and behavioral data we calculated the correlation between the mean MMN amplitude in the 100–140 ms interval and the difference in incorrect identifications between incongruent and congruent stimuli conditions across all subjects and conditions (cf. Fig. 3). The correlation between MMN amplitude and this behavioral McGurk measure was not significant ($P > 0.2$). As the estimated correlation may depend on the behavioral measure being constrained to values between zero and one, we also calculated the correlation between the MMN amplitude and the *Z*-score ($P > 0.2$; only for values greater than zero and less than one); as well as between the MMN amplitude and the arcsine transformed behavioral measure ($P > 0.1$) but this only confirmed the lack of correlation.

When looking at correlations between behavioral responses and McGurk-MMN, the results are inconclusive. The lack of correlation may seem surprising as we did find that both measures were higher in the normal, upright condition where we expected integration to be maximal. One explanation, which can never be excluded for a negative finding, is lack of statistical power due to an insufficient amount data. Another explanation is that this could indicate a non-linear effect where the MMN response only occurs when the McGurk illusion is very strong. This is unlike findings from several auditory MMN paradigms showing that MMN amplitude correlates well with perceived difference (May and Tiitinen, 2010; Näätänen, 2003; Näätänen et al., 1978). This has, to our knowledge, not been investigated for the McGurk-MMN. To investigate if the relationship between the perceived difference and the McGurk-MMN amplitude could be nonlinear we calculated the minimal behavioral response (difference in percentage incorrect between congruent and incongruent conditions) that elicited MMN negativity for all subjects in all conditions. We found that for behavioral measures of 75 percent points and above there was a consistent MMN. For behavioral measures below 75 percent points we found that the MMN was much more variable. This indicates that the McGurk illusion needs to be very strong to elicit the MMN consistently.

4. Discussion

4.1. Behavioral experiment

The present behavioral findings replicate Rosenblum et al.'s (2000) primary finding: The McThatcher effect. Thatcherization greatly reduces the influence of vision upon the auditory speech percept for an upright, but not for an inverted face. As a secondary finding we found a stronger face inversion effect than Rosenblum et al. (2000) in that inverting the normal face reduced the McGurk illusion more than in their study. While others have found smaller effects (Bertelson et al., 1994; Jordan and Bevan, 1997; Thomas and Jordan, 2002), our results are similar to those of Massaro and Cohen (1996). Given that the magnitude of the inversion effect has varied substantially across previous reports, this is not surprising. Overall, the McThatcher effect replicated in the present study supports the hypothesis that audiovisual speech perception is based not only on facial features but also on facial configuration (Rosenblum et al., 2000).

4.2. Mismatch negativity experiment

The non-Thatcherized upright face produced a strong McGurk-MMN response. The amplitude, latency and scalp distribution of this MGurk-MMN response is comparable with those reported in previous studies (Colin, 2002; Ponton et al., 2009; Saint-Amour et al., 2007; Stekelenburg and Vroomen, 2012). This signifies that the McGurk illusion we found in the behavioral experiment influenced activity in auditory cortex confirming that the effect is truly perceptual.

The two elements of the McThatcher effect were also reflected in the McGurk-MMN. First, we found no McGurk-MMN for an upright Thatcherized face reflecting that Thatcherization disrupts the McGurk illusion for upright faces. Second, we found no effect of Thatcherization on the McGurk-MMN for inverted stimuli. This mirrors the lack of difference found in the two matching behavioral conditions. However, as none of the inverted faces produced an MMN irrespective of facial configuration, only limited conclusions can be drawn from this regarding the effect of facial configuration for inverted faces. Thus, surprisingly, we found a much stronger face inversion effect in the McGurk-MMN than in the behavioral data.

4.3. General discussion

Our primary question was directed towards the role of facial configuration information in perception of visual and audiovisual speech. Our findings answer this and related questions in multiple ways.

First, our main finding is that the McThatcher effect is reflected in both behavioral and MMN responses. We found a strong McGurk illusion and a corresponding MMN for a normal upright face. The McGurk illusion was strongly reduced and the MMN eliminated when the face was Thatcherized. This confirms that speech perception is affected by facial configuration when the facial orientation is upright. For inverted faces, we found no effect of Thatcherization on the McGurk illusion and the corresponding MMN. This is in agreement with the notion that facial configuration has little influence on visual and audiovisual speech perception for inverted faces.

As a secondary finding we found a discrepancy between our behavioral findings and the MMN for the face inversion effect. Although face inversion decreased the strength of the McGurk illusion, it did not eliminate it. While we found a moderate McGurk illusion for inverted faces, these stimuli did not elicit an MMN response. Unfortunately, this means that our MMN data tells us little about the effect of Thatcherization for inverted faces.

Investigating this discrepancy further we found no correlation between the magnitude of the McGurk illusion and the amplitude of the McGurk-MMN. We suggest three possible explanations of this. First, the statistical power of our MMN data is limited, and it may be insufficient for finding a McGurk-MMN of smaller effect size. Another possibility is that the McGurk illusion for inverted faces is not truly perceptual but based on changes in behavior at another stage of perceptual processing, e.g. in response selection. As we find the McGurk illusion perceptually convincing even for inverted faces we do not believe that this is the correct interpretation but admit that this remains to be tested formally. Here, McGurk responses to the specific incongruent syllable combination (acoustic /ba/+visual /va/) is dominated by the categories "va" and "fa". These responses could in principle be due to both audiovisual integration and to response bias towards the visual stimulus. In the latter case, no McGurk-driven MMN would be produced. Repeating the experiment using a discrimination task (Rosenblum et al., 2000) or sensitivity measures from signal detection theory (Kislyuk et al., 2008) could help elucidate this.

Finally, the McGurk-driven MMN may differ from auditory MMN in having a non-linear relation to the perceived difference. We find this a likely explanation as the McGurk-MMN was consistent only when the behavioral data indicated a strong McGurk illusion and highly variable for weaker McGurk illusions. Whereas the relation between MMN amplitude and stimulus deviation is well-described for acoustic stimuli, we are not aware of any study targeting the relation between McGurk illusion strength and amplitude of the McGurk-MMN response. However, studies using McGurk-MMN, which also report the level of behavioral McGurk response, report near 100% McGurk illusion with incongruent stimuli (Stekelenburg and Vroomen, 2012), or, "a strong McGurk illusion" (Saint-Amour et al., 2007). Kislyuk et al. (2008) exclude subjects with "a weak McGurk effect". From these reports of behavioral McGurk illusion strength it may be that a strong behavioral McGurk response is a prerequisite for evoking an audiovisual MMN response. If this is the case, McGurk-driven MMN differs from auditory MMN (Garrido et al., 2009; May and Tiitinen, 2010; Näätänen et al., 1978, 2004) in not being a graded response, proportional to the degree of stimulus deviance, but only being evoked by a strong audiovisual integration response. This warrants caution in basing conclusions about audiovisual speech perception on the McGurk-MMN.

References

Bartlett, J.C., Searcy, J., 1993. Inversion and configuration of faces. Cogn. Psychol. 25, 281–316.

Bertelson, P., Vroomen, J., Wiegeraad, G., de Gelder, B. Exploring the relation between McGurk interference and ventriloquism, Proceedings of ICSLP 1994, Yokohama, pp. 559–562.

Bigdely-Shamlo, Kreutz-Delgado, K., Kothe, C, Makeig, S., 2013. EyeCatch: data-mining over half a million EEG independent components to construct a fully-automated eye-component detector. In: Proceedings of the IEEE Engineering in Biology and Medicine Conference, Osaka.

Bruce, V., Young, A.W., 2012. Face Perception. Psychology Press, London; New York.

Carbon, C.-C., Schweinberger, S.R., Kaufmann, J.M., Leder, H., 2005. The Thatcher illusion seen by the brain: an event-related brain potentials study. Cogn. Brain Res. 24, 544–555.

Colin, C., 2002. Mismatch negativity evoked by the McGurk–MacDonald effect: a phonetic representation within short-term memory. Clin. Neurophysiol. 113, 495–506.

Czigler, I., 2007. Visual mismatch negativity: violation of nonattended environmental regularities. J. Psychophysiol. 21, 224–230.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–21.

Farah, M.J., Wilson, K.D., Drain, M., Tanaka, J.N., 1998. What is "special" about face perception? Psychol. Rev. 105, 482–498.

Garrido, M.I., Kilner, J.M., Stephan, K.E., Friston, K.J., 2009. The mismatch negativity: a review of underlying mechanisms. Clin. Neurophysiol. 120, 453–463.

Goffaux, V., Rossion, B., 2007. Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. J. Exp. Psychol. Hum. Percept. Perform. 33, 995–1002.

Groppe, D.M., Urbach, T.P., Kutas, M., 2011. Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. Psychophysiology 48, 1711–1725.

Hietanen, J.K., Manninen, P., Sams, M., Surakka, V., 2001. Does audiovisual speech perception use information about facial configuration? Eur. J. Cogn. Psychol. 13, 395–407.

Jordan, T.R., Bevan, K., 1997. Seeing and hearing rotated faces: influences of facial orientation on visual and audiovisual speech recognition. J. Exp. Psychol. Hum. Percept. Perform. *23*, 388–403.

Jordan, T.R., Thomas, S.M., 2011. When half a face is as good as a whole: effects of simple substantial occlusion on visual and audiovisual speech perception. Atten. Percept. Psychophys. 73, 2270–2285.

Kislyuk, D.S., Möttönen, R., Sams, M., 2008. Visual processing affects the neural basis of auditory discrimination. J. Cogn. Neurosci. 20, 2175–2184.

Kleiner M., Brainard D.H., Pelli D.G., What's New in Psychtoolbox-3?, *Perception*, 36, ECVP 2007 abstract supplement.

Lang, A.H., Eerola, O., Korpilahti, P., Holopainen, I., Salo, S., Aaltonen, O., 1995. Practical issues in the clinical application of mismatch negativity. Ear Hear. 16, 118–130.

Massaro, D.W., Cohen, M.M., 1996. Perceiving speech from inverted faces. Percept. Psychophys. 58, 1047–1065.

May, P.J.C., Tiitinen, H., 2010. Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. Psychophysiology 47, 66–122.

McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. Nature 264, 746–748.
Möttönen, R., Krause, C.M., Tiippana, K., Sams, M., 2002. Processing of changes in visual speech in the human auditory cortex. Brain Res. Cogn. Brain Res. 13, 417–425.
Näätänen, R., 2003. Mismatch negativity: clinical research and possible applications. Int. J. Psychophysiol. 48, 179–188.
Näätänen, R., Alho, K., 1995. Mismatch negativity—a unique measure of sensory processing in audition. Int. J. Neurosci. 80, 317–337.
Näätänen, R., Gaillard, A.W.K., Mäntysalo, S., 1978. Early selective-attention effect on evoked potential reinterpreted. Acta Psychol. 42, 313–329.
Näätänen, R., Pakarinen, S., Rinne, T., Takegata, R., 2004. The mismatch negativity (MMN): towards the optimal paradigm. Clin. Neurophysiol. 115, 140–144.
Ponton, C.W., Bernstein, L.E., Auer, E.T., 2009. Mismatch negativity with visual-only and audiovisual speech. Brain Topogr. 21, 207–215.
Prkachin, G.C., 2003. The effects of orientation on detection and identification of facial expressions of emotion. Br. J. Psychol. 94, 45–62.
Rosenblum, L.D., 2001. Reading upside-down lips, ⟨http://www.faculty.ucr.edu/~rosenblu/VSinvertedspeech.html⟩.
Rosenblum, L.D., Yakel, D.A., Green, K.P., 2000. Face and mouth inversion effects on visual and audiovisual speech perception. J. Exp. Psychol. Hum. Percept. Perform. 26, 806–819.
Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., Foxe, J.J., 2007. Seeing voices: high-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. Neuropsychologia 45, 587–597.
Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O.V., Lu, S.-T., Simola, J., 1991. Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. Neurosci. Lett. 127, 141–145.
Stefanics, G., Kimura, M., Czigler, I., 2011. Visual mismatch negativity reveals automatic detection of sequential regularity violation. Front. Hum. Neurosci. 5.
Stekelenburg, J.J., Vroomen, J., 2012. Electrophysiological evidence for a multisensory speech-specific mode of perception. Neuropsychologia, 1425–1431.
Stekelenburg, J.J., Vroomen, J., de Gelder, B., 2004. Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. Neurosci. Lett. 357, 163–166.
Sumby, W., Pollack, I., 1954. Visual contribution to speech intelligibility in noise. J. Acoust Soc Am. 26, 212–215.
Thomas, S.M., Jordan, T.R., 2002. Determining the influence of Gaussian blurring on inversion effects with talking faces. Percept. Psychophys. 64, 932–944.
Thompson, P., 1980. Margaret Thatcher: a new illusion. Perception 9, 483–484.
Valentine, T., 1988. Upside-down faces: a review of the effect of inversion upon face recognition. Br. J. Psychol. 79, 471–491.
Walden, B.E., Prosek, R.A., Montgomery, A.A., Scherr, C.K., Jones, C.J., 1977. Effects of training on the visual recognition of consonants. J. Speech Hear. Res. 20, 130–145.