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.

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

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