Atten Percept Psychophys
DOI 10.3758/s13414-014-0821-1

A link between individual differences in multisensory speech perception and eye movements

Demet Gurler · Nathan Doyle · Edgar Walker ·
John Magnotti · Michael Beauchamp

Abstract The McGurk effect is an illusion in which visual speech information dramatically alters the perception of auditory speech. However, there is a high degree of individual variability in how frequently the illusion is perceived: some individuals almost always perceive the McGurk effect, while others rarely do. Another axis of individual variability is the pattern of eye movements make while viewing a talking face: some individuals often fixate the mouth of the talker, while others rarely do. Since the talker's mouth carries the visual speech necessary information to induce the McGurk effect, we hypothesized that individuals who frequently perceive the McGurk effect should spend more time fixating the talker's mouth. We used infrared eye tracking to study eye movements as 40 participants viewed audiovisual speech. Frequent perceivers of the McGurk effect were more likely to fixate the mouth of the talker, and there was a significant correlation between McGurk frequency and mouth looking time. The noisy encoding of disparity model of McGurk perception showed that individuals who frequently fixated the mouth had lower sensory noise and higher disparity thresholds than those who rarely fixated the mouth. Differences in eye movements when viewing the talker's face may be an important contributor to interindividual differences in multisensory speech perception.

Keywords Eye movements · Cognitive · Speech perception · Multisensory processing

D. Gurler
Department of Radiology, University of Alabama at Birmingham,
Birmingham, AL, USA
e-mail: demetgurlerr@gmail.com

N. Doyle · J. Magnotti
Department of Neurobiology and Anatomy, University of Texas
Health Science Center at Houston, Houston, TX, USA

E. Walker · M. Beauchamp
Department of Neuroscience, Baylor College of Medicine,
One Baylor Plaza, Houston, TX 77030, USA

M. Beauchamp (✉)
Department of Neurosurgery, Baylor College of Medicine,
One Baylor Plaza, Houston, TX 77030, USA
e-mail: michael.beauchamp@bcm.edu

Introduction

The McGurk effect is an illusion that demonstrates the interaction between the visual and auditory modalities during speech perception (McGurk & MacDonald, 1976). Participants are presented with an auditory syllable (e.g., “ba”) paired with a different visual syllable (e.g., “ga”). This incongruent visual information leads to the perception of a third, completely different syllable (e.g., “da”). However, some participants do not experience the illusion and perceive only the auditory component of the stimulus (Nath & Beauchamp, 2012). The brain responses of these two groups, termed strong perceivers and weak perceivers of the illusion, are markedly different in both adults (Nath & Beauchamp, 2012) and children (Nath, Fava, & Beauchamp, 2011). In strong perceivers, McGurk stimuli evoke large amplitude responses in the left superior temporal sulcus (STS), a brain area known to be important for integration of auditory and visual information (Beauchamp, Lee, Argall, & Martin, 2004a) and perception of the McGurk effect (Beauchamp, Nath, & Pasalar, 2010).

Another axis of individual differences is the pattern of eye movements made when viewing a face. Recent studies of individuals viewing static faces show remarkable differences in patterns of eye movements. Some individuals often fixate the mouth when viewing a face while others rarely do (Mehoudar, Arizpe, Baker, & Yovel, 2014; Peterson & Eckstein, 2012, 2013).

Published online: 26 March 2015

We hypothesized the existence of a link between these two axes of individual differences. The mouth movements of a talker are known to be highly correlated with speech acoustics (Yehia, Rubin, & Vatikiotis-Bateson, 1998) and increased time spent fixating a location in the visual scene increases the amount of information that can be extracted from that location (Henderson, 2003). Therefore, participants who spend more time fixating the mouth of a talking face might be expected to receive more precise information about visual speech, resulting in increased perception of the McGurk effect. To test this hypothesis, we used infrared eye tracking to measure the eye movements made while 40 participants viewed brief video clips of audiovisual speech that included McGurk syllables.

Materials and methods

Participants

All participants (*n* = 40, 19 M, 21 F, mean age 25 years) gave informed consent and were compensated for their time as approved by the University of Texas Committee for the Protection of Human Participants.

Audiovisual speech stimuli and task

The stimuli consisted of six different audiovisual speech videos with duration of approximately 2 s. The stimuli subtended approximately 20° of visual angle on an LCD monitor (1,024 x 768 resolution) positioned at eye level 60 cm from the participants. The sound pressure level of the speech was approximately 60 dB. After the conclusion of each video clip, participants reported their percept. The different video clips were presented repeatedly in random order (10 repetitions of each video for 20 participants, 30 repetitions of each video for 20 participants).

The videos were recorded using a digital video camera and edited with digital video editing software. The clips were presented at 30 frames/s, with a mean of 52 frames in each clip. Each video started and ended with the mouth in a neutral, mouth-closed position. Averaged across clips, the mouth movement commenced at frame 10 and finished at frame 39, resulting in mouth movements occupying 65 % of the total clip time. The stimuli are freely available for download from http://openwetware.org/wiki/Beauchamp:Publications.

Four stimulus videos consisted of congruent syllables: AbaVba, AgaVga, ApaVpa, AkaVka. These stimuli always evoked the expected percept (ceiling accuracy, 100 %). Two stimulus videos consisted of the mismatched syllables described in the original report (McGurk & MacDonald, 1976) created by splicing the auditory and visual components of the congruent audiovisual stimuli. Auditory "ba" was combined with visual "ga" (AbaVga) and auditory "pa" was combined with visual "ka" (ApaVka). These stimuli evoked either an illusory McGurk percept ("da" for AbaVga, "ta" for ApaVka) or a percept of the auditory component of the stimulus ("ba" for AbaVga, "pa" for ApaVka). A report of any percept other than that of the auditory component of the stimulus was classified as a McGurk percept. We also tested a scoring scheme in which any percept other than the auditory or the visual component of the stimulus was classified as a McGurk percept (see Results: additional analyses).

Eye tracking

Eye tracking was performed using an EyeLink video-based eye tracker (SR Research, Ottawa, ON) The eye tracker was used in head-free binocular mode with a sampling rate of 500 Hz and a spatial resolution of 0.25°. At the beginning of each experimental session, calibration and verification were performed using 13 targets distributed over the entire screen. To ensure high-quality eye tracking throughout the session, each trial began with a single calibration target presented at one of four corners of the invisible bounding box in which the video clip would later appear. Poor correspondence between the measured eye location and the fixation target indicated that eye tracker drift had occurred. In this case, the 13-target calibration and verification were repeated before resuming the experiment. Otherwise, the trial proceeded, with the disappearance of the calibration target and the appearance of the video clip. While the video clip played, there was no fixation target and participants were not explicitly instructed to fixate on the face or any other location (free-viewing). Because the eye-tracker calibration target at the beginning of each trial was presented peripherally and the video clip was presented centrally, participants always made at least one eye movement from the peripheral calibration target to a central gaze location located within the stimulus. Fixations were measured only during stimulus presentation (beginning at stimulus onset and ending at stimulus offset).

Eye movement analysis

Blinks, saccades and fixation locations throughout each video clip were identified using the SR Research Data Viewer; heat maps were created using the duration density option. Additional analysis was conducted in Matlab (Mathworks, Boston, MA). Separate regions of interest (ROI) were created for each stimulus, consisting of rectangular bounding boxes for the left eye, the right eye, and the mouth of the talker (Fig. 2d). The ROIs were the same size in each stimulus, and were constructed so that the area of the left eye ROI plus the area of the right eye ROI equaled the area of the mouth ROI to allow direct comparison between the eye and mouth ROIs. Because the talker's face was stationary in the center of the video frame, the ROIs did not change over the course of each video. The mouth ROI was large enough to encompass the mouth even at maximal mouth opening. Raw measures of fixation duration (ms) were converted into percent of stimulus length.

Behavioral model

We modeled the behavioral data using the noisy encoding of disparity model of the McGurk effect (Magnotti and Beauchamp, 2014). The source code for the model is freely available from http://openwetware.org/wiki/Beauchamp:NED. The model assumes that the disparity (discrepancy) between the incongruent auditory and visual components of each McGurk stimulus can be arranged along a single axis. The present study used two stimuli that were tested in 165 participants in a model validation study (Magnotti and Beauchamp, 2014). We fixed the stimulus disparity estimates at the values from the validation study and applied the model to the behavioral data from the present study to estimate two parameters for each participant, a sensory noise term and a disparity threshold (the model fits were blind to the eye movement data).

The sensory noise term captured the precision of the participant's trial-to-trial estimates of a given stimulus' audiovisual disparity. The model assumes that on each trial, the participant makes a noisy measurement of the true stimulus disparity, with the standard deviation equal to sensory noise. This measured value of stimulus disparity is then compared to the participant's disparity threshold. If the measured disparity is below threshold, the auditory and visual speech cues are integrated and the participant perceives the McGurk effect. If the measured disparity is above threshold, the cues are not integrated and the participant perceives the auditory speech component of the stimulus.

Results

Eye movement behavior

There was significant variability in the eye movement behavior of the different participants as they viewed the short video clips of audiovisual speech (Fig. 1). Some participants most often fixated the mouth of the talker, while other participants fixated primarily the eyes of the talker. Still other participants showed a distribution of fixations that encompassed both the mouth and the eyes of the talker. The eye movement pattern of each individual was consistent across stimuli so stimuli were grouped for further analysis.

McGurk perception

There was also significant variability across participants in how frequently the McGurk effect was perceived (Fig. 2a). Some participants never perceived the illusion while others always perceived it. We classified individuals as strong perceivers if they reported the McGurk percept on 50 % or more of the trials (n = 23, mean 84 %±4 % SEM) or as weak perceivers if they reported the effect on less than 50 % of the trials (n = 17, mean 23 %±4 % SEM).

To determine if these two different measures of individual variability were linked, we created an average map of the eye movement behavior of strong and weak perceivers of the McGurk effect. For strong perceivers, the fixation duration was strongly localized to the mouth of the talker (Fig. 2b). For weak perceivers, the fixation duration was strongly localized to the eyes of the talker (Fig. 2c). Because there are many possible fixation locations, the group average fixation maps are both smoother and contain lower peak values than found in any individual fixation map (Fig. 1).

To quantify this difference, we created regions of interest (ROIs) around the mouth of the talker and the left and right eye of the talker (Fig. 2d). The area of the mouth ROI equaled the total area of the left and right eye ROIs, so that any difference in fixation duration is not due to differences in ROI area.

As shown in Fig. 2e, strong perceivers spent significantly more time fixating the mouth than weak perceivers did: 41 % of fixation time for SP vs 23 % of total fixation time for WP, two-sample t-test $t_{38} = 2.91$, $P = 0.006$. Conversely, there was a trend towards weak perceivers spending more time fixating the eyes of the talker than strong perceivers (31 % for WP vs. 20 % for SP, $t_{38} = 1.89$, $P = 0.07$; Fig. 2f).

Correlation analysis

Our initial analysis relied on grouping our participants as strong or weak perceivers of the McGurk effect. In order to analyze the relationship between eye movement behavior and McGurk perception without grouping, we calculated the correlation between the mouth looking time and the frequency of the McGurk effect (Fig. 3a). There was a positive correlation ($r = 0.34$, $P = 0.03$) with each 2 % increase in mouth looking time resulting in a 10 % increase in McGurk frequency.

Our first correlation analysis examined the total duration of fixations that fell within the mouth ROI, meaning that other fixations (whether just outside the mouth ROI or a long distance from it) were ignored. Therefore, we performed a second analysis by selecting the most frequently fixated (hottest) point in the fixation heat map, as shown for each participant in Fig. 1. We calculated the distance from this most frequently fixated point to the center of the mouth to provide an index of mouth fixation independent of drawn ROIs. Correlating this distance with McGurk frequency for each participant revealed

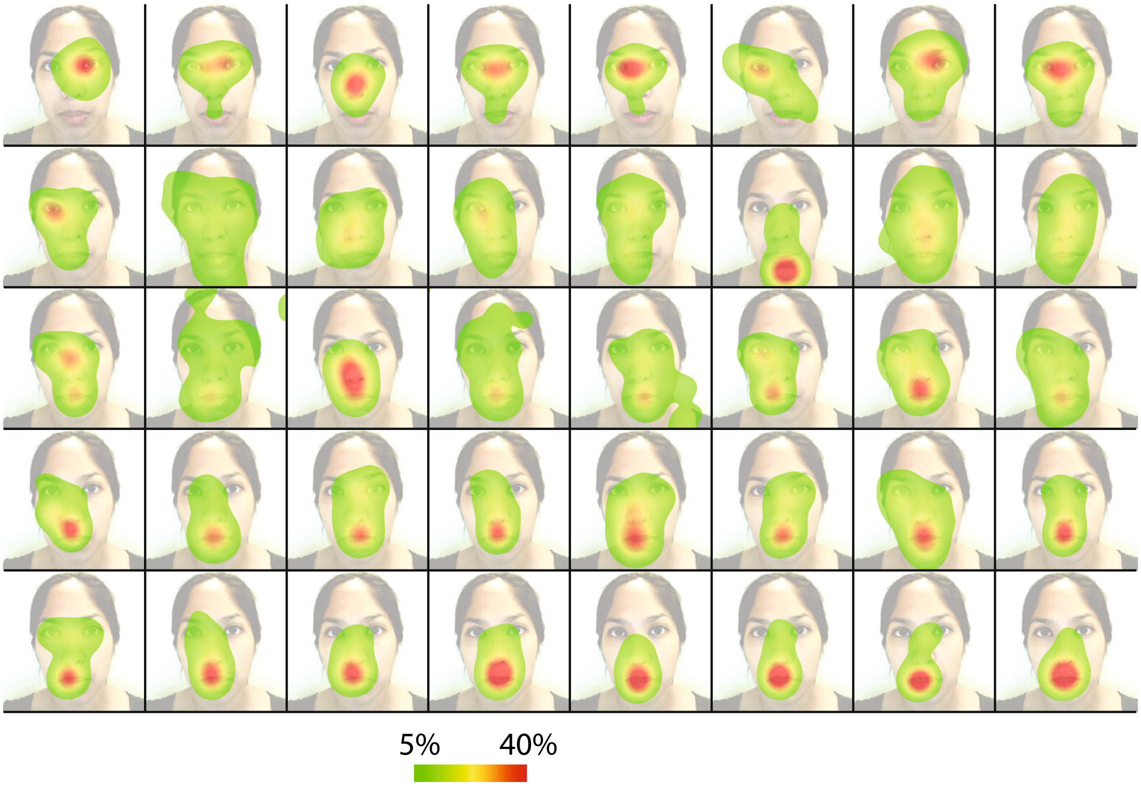

Fig. 1 Eye movement differences across participants. Each panel shows the eye movement behavior of a single participant while viewing audiovisual speech, averaged across presentations. The color scale shows the percentage of time spent fixating a given location as a percentage of total viewing time. The underlying face shows a still frame from the audiovisual speech video with contrast reduced to highlight the eye movement data. Participants are ordered by the amount of time spent fixating the mouth region (*top left* least amount of time, *bottom right* greatest amount of time)

a significant negative correlation (r = −0.35, P = 0.03, Fig. 3b): the further away from the mouth a participant fixated, the less likely they were to perceive the McGurk effect, with each degree further away from the mouth resulting in a 6 % decrease in McGurk frequency.

Additional analyses

Increased fixation of the mouth by strong perceivers of the McGurk effect could arise from an increased number of fixations on the mouth or by an increased duration of each mouth fixation. Strong perceivers made more fixations on the mouth ROI in each trial (1.1 for strong perceivers vs. 0.7 for weak perceivers, t_{38} = 2.54, P = 0.02) but the duration of each fixation was similar between groups (39 % of total trial time vs. 34 %, t_{38} = 1.46, P = 0.15).

In our initial analysis, any report that differed from the auditory component of the stimulus was scored as a McGurk percept. We also tested an alternative scoring scheme in which any report that differed from the auditory or the visual component of the stimulus was scored as a McGurk response. This scoring scheme did not change our findings. Strong perceivers spent significantly more time fixating the mouth than did weak perceivers (42 % vs. 23 %, two-sample t-test t_{38} = 3.18, P = 0.003) and there was a significant correlation between time spent fixating the mouth and the frequency of perceiving the McGurk effect (r = 0.37, P = 0.02).

Modeling of the behavioral data

To better understand the link between eye movements and speech perception, we applied the noisy encoding of disparity model of the McGurk effect (Magnotti and Beauchamp, 2014). The model uses behavioral data to estimate two parameters for each participant. The first is the participant's internal sensory noise. For instance, if during repeated presentations of the same McGurk stimulus a participant sometimes reports the McGurk percept and sometimes does not, the model infers that the participant has high internal sensory noise. The second parameter is the participant's sensitivity to the disparity (discrepancy) between the incongruent auditory and visual components of McGurk stimuli. Participants integrate the auditory and visual components of the stimulus (leading to the McGurk percept) if their estimate of the auditory-visual disparity falls below a fixed threshold. If it exceeds the threshold, the participant assumes that the auditory and visual components arise from different sources and should not be integrated (leading to a percept of the auditory component of the

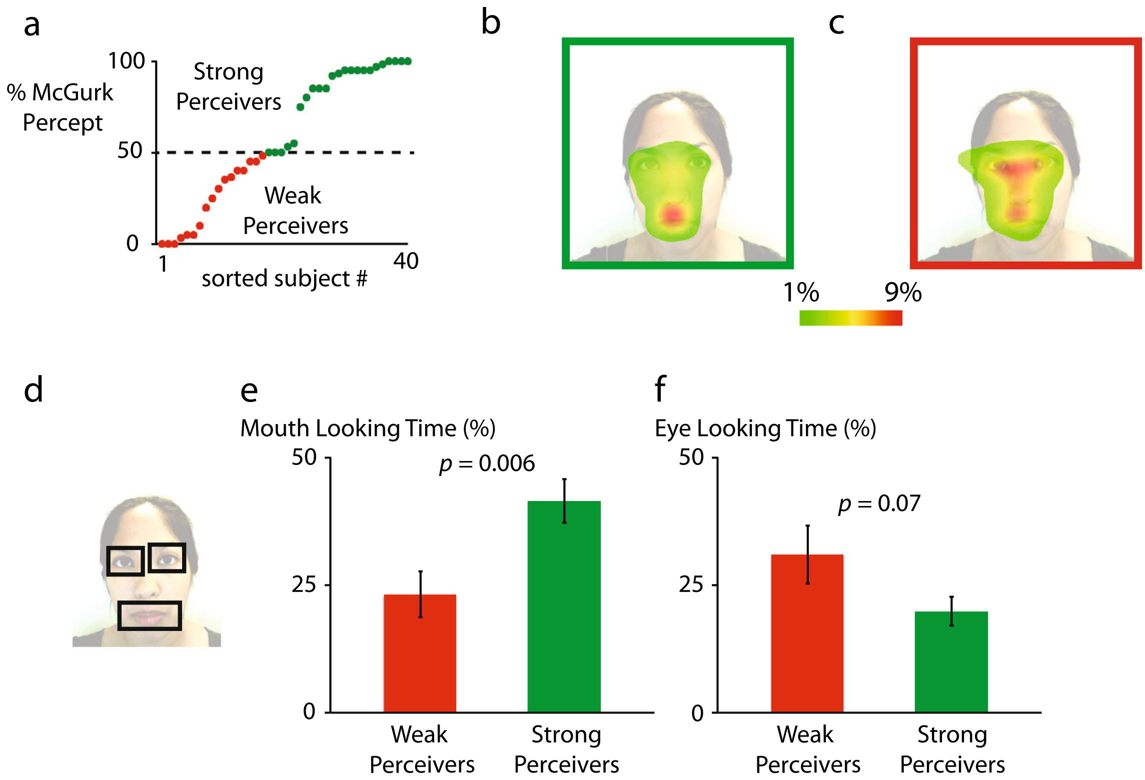

Fig. 2 **a**–**f** McGurk frequency differences across participants. **a** Each symbol shows a single participant, in increasing order by how frequently they perceived the McGurk effect. Participants were classified as strong (frequency > = 50 %; *green symbols*) or weak (*red symbols*) perceivers of the effect. **b** Average eye movement behavior of strong perceivers (illustrated by *green outline around still frame of video*). *Color scale* shows percentage of time spent fixating each location. **c** Average eye movement behavior of weak perceivers (*red frame*). **d** Still frame of video with *black outlines* showing the location of three regions-of-interest (ROIs) used for eye movement analysis (left eye, right eye, mouth). **e** The percentage of total stimulus duration spent fixating the mouth ROI (mouth looking time) for weak and strong perceivers. *Error bars* Standard error of the mean. **f** The percentage of total stimulus duration spent fixating the eye ROIs

stimulus). For instance, if a participant never reports the McGurk percept for any McGurk stimulus, the model infers that the participant has a low disparity threshold. The advantage of applying the model is that is provides a window into the internal processes that may underlie variable perception of the McGurk effect: the model parameters reflect cognitive processes rather than raw behavioral data. For instance, the model disentangles the effects of subject threshold and

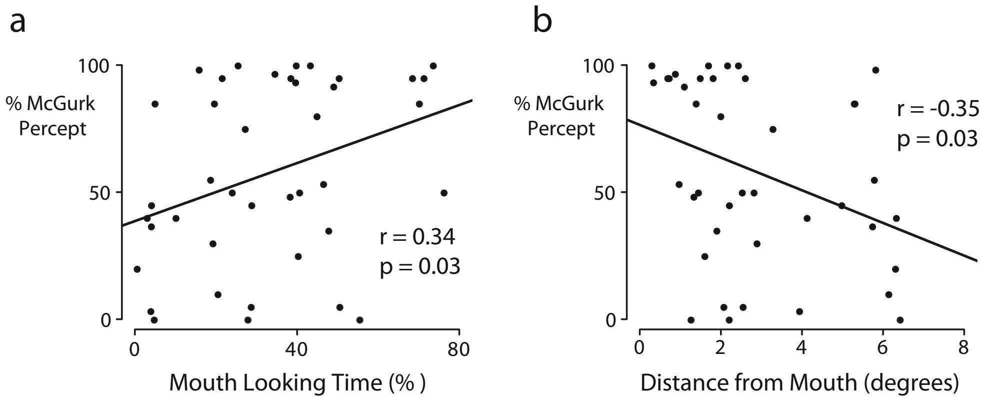

Fig. 3 **a**, **b** Correlations between McGurk frequency and eye movement measures. **a** The *x*-axis shows the participant's frequency of perceiving the McGurk effect . Each symbol represents an individual participant's percentage of total stimulus duration spent fixating the mouth ROI (mouth looking time). There is a positive correlation between McGurk frequency and mouth looking time. **b** The *y*-axis shows the distance from the mouth of the most frequently fixated point (hottest spot in the fixation heat map for that participant in Fig. 1). There is a negative correlation between McGurk frequency and fixation distance from the mouth

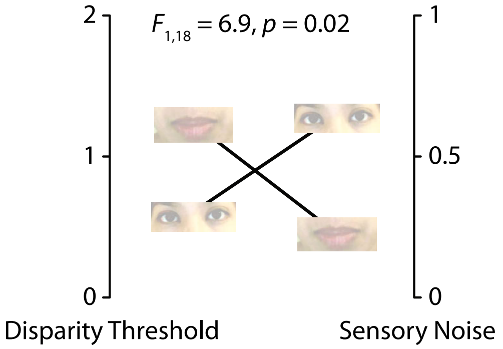

Fig. 4 Interaction between eye movements and model parameters. The noisy encoding of disparity model of McGurk effect was used to estimate the disparity threshold and sensory noise of participants who frequently fixated the mouth (highest quartile of mouth looking time, data point shown as an image of the talker's mouth) and participants who rarely fixated the mouth (lowest quartile of mouth looking time, data point shown as eye image since these participants fixated the talker's eyes). There was a significant interaction between the parameter values for each group, with mouth-fixaters showing higher disparity threshold but lower sensory noise

stimulus efficacy that are conflated in the raw behavioral score of percent of McGurk fusion percepts.

We divided participants into those who rarely fixated the mouth (lowest quartile of mouth looking time, time < 20 %) and participants who frequently fixated the mouth (highest quartile, time > 47 %) and examined the model's estimates of their sensory noise and disparity thresholds.

As shown in Fig. 4, there was a significant interaction ($F_{1, 18}$ = 6.9, P = 0.02) between participant group (low vs. high mouth looking) and the model parameters. The interaction was caused by lower sensory noise in participants who frequently fixated the mouth and lower disparity thresholds in participants who rarely fixated the mouth. To examine this relationship in more detail, we correlated mouth looking time with each model parameter. Across all participants, there was a negative correlation between mouth looking time and sensory noise ($r = -0.32$, $P = 0.047$), meaning that increased time spent fixating the mouth resulted in decreased sensory noise. There was a positive correlation between mouth ROI time and disparity threshold (r = 0.41, P = 0.008) meaning that participants who fixated the mouth were more likely to tolerate audiovisual disparity and make use of the visual speech information.

Discussion

We found large individual differences in the McGurk effect, with some participants never perceiving the illusion and others always perceiving it (Nath & Beauchamp, 2012) and large individual differences in eye movements, with some participants often fixating the mouth and others rarely fixating it (Mehoudar et al., 2014; Peterson & Eckstein, 2013). While previous studies of eye movement differences have presented static faces, our results show that individual differences in eye movements also occur while viewing dynamic talking faces. The main new result of our study is that of a relationship between these two axes of individual differences. Participants who perceived the McGurk effect more frequently were more likely to fixate the mouth of the talker. This relationship makes intuitive sense: perception of the McGurk effect requires integrating visual information from the mouth of the talker with auditory information from the voice of the talker to produce an illusory percept that differs from both the auditory and visual components of the stimulus (McGurk & MacDonald, 1976). Participants who fixated the mouth made more use of the visual information from the mouth of the talker.

The noisy encoding of disparity (NED) model of the McGurk effect provides a quantitative conceptual framework for this intuition (Magnotti and Beauchamp, 2014). The model uses behavioral data to characterize each participant by a threshold term (with a higher threshold indicating greater willingness to make use of visual speech information despite its incongruence with the auditory speech) and by a sensory noise term that measures the internal fidelity of their assessment of the discrepancy or disparity between the auditory and visual components of the stimulus (with higher sensory noise indicating a less precise representation).

The model revealed that participants who fixated the mouth had lower sensory noise in their internal representation of the stimulus. Movements of the jaw, cheeks and the mouth are highly correlated with vocal tract motion and speech acoustics (Yehia et al., 1998) and viewing only the lips of the talker improves perception of noisy audiovisual speech almost as much as viewing the entire face (Summerfield & McGrath, 1984; Thomas & Jordan, 2004). Increased time spent fixating a location in the visual scene increases the amount of information that can be extracted from that location (Henderson, 2003). Therefore, foveal viewing of the mouth region is likely to provide a more precise, less noisy representation of the talker's visual speech, and would be expected to result in a more consistent representation of the visual speech and a more consistent McGurk effect.

Participants who fixated the mouth had a higher disparity threshold as estimated by the model, meaning that they were more tolerant of the disparity between the incongruent auditory and visual components of the McGurk stimuli. Individuals who often fixated the mouth had lower sensory noise. Under the principles of Bayesian inference, this greater precision should lead them to weight visual speech more strongly, increasing their tolerance for audiovisual disparity and resulting in a stronger McGurk effect (Seilheimer, Rosenberg, & Angelaki, 2014). Conversely, participants who do not fixate the mouth would have greater sensory noise, leading them to weight visual speech information less strongly, decreasing their tolerance for audiovisual disparity and resulting in perception of the auditory component of the McGurk stimulus.

The NED model provides an explanation for a puzzling phenomenon observed in our data. On some trials, strong perceivers of the McGurk effect did not fixate the mouth but still reported the illusory fusion percept. We explain this by positing that their increased disparity threshold led them to incorporate the visual speech information even when it was only visible peripherally, resulting in the McGurk percept. Conversely, even on trials in which they fixated the mouth, weak perceivers did not report the illusory percept. Because of their lower disparity threshold and noisier internal representation of speech, weak perceivers did not use visual speech information even when present in their central vision.

The surprisingly high degree of individual variability in McGurk frequency, ranging from 0 % to 100 %, can be explained if patterns of eye movement differences are reinforced over time. In studies using static faces, participants who preferred to fixate the mouth showed this behavior consistently, in testing sessions conducted up to 18 months apart (Mehoudar et al., 2014). Over the thousands of exposures to talking faces experienced by all humans, individuals who consistently fixate the mouth could build up a more accurate and less variable internal representation of visual speech, as shown by the decreased sensory noise NED model parameter for participants who often fixated the mouth. Repeated exposure to high-quality visual speech information through mouth fixation would be expected to make internal representations of visual speech more accurate, resulting in a greater influence of visual speech on speech perception, and perhaps encouraging even more mouth fixations in a positive feedback loop.

The role of attention

The correlation with eye movements accounted for only about 12 % of the variability in individual differences in McGurk perception. One possible explanation for this relatively low value is visual attention. Withdrawing attention from an audiovisual McGurk stimulus by directing attention to a competing auditory or visual stimulus (Alsius et al., 2005), to a somatosensory stimulus (Alsius et al., 2007), or to a concurrent working memory task (Buchan and Munhall, 2012) reduces perception of the McGurk effect. Therefore, it may not be fixation location but rather the locus of visual attention that is the key determinant of whether participants perceive the McGurk effect. In this view, measurements of fixation location are an imperfect proxy for measurements of the location of the focus of spatial attention. If weak perceivers of the McGurk effect directed their spatial attention to the eyes (while fixating the mouth) or if strong perceivers of the McGurk effect directed their spatial attention to the mouth (while fixating elsewhere) the correlation between fixation location and McGurk perception would be much weaker than the correlation between attention and McGurk perception.

Implications for neural mechanisms

The left superior temporal sulcus (STS) plays an important role in integration of auditory and visual information and eye movements (Beauchamp et al., 2004a; Beauchamp, Lee, Argall, and Martin 2004b). When viewing McGurk stimuli, the left STS is more active in strong perceivers than weak perceivers in both adults (Nath & Beauchamp, 2012) and children (Nath et al., 2011) and disrupting activity in the left STS of strong perceivers using transcranial magnetic stimulation weakens McGurk perception (Beauchamp et al., 2010, Nath, & Pasalar, 2010). Fixating different regions of the face strongly modulates brain responses (Zerouali, Lina, & Jemel, 2013) and viewing mouth movements, even without sound, evokes activity in the STS (Puce, Allison, Bentin, Gore, & McCarthy, 1998). Therefore, the STS may be an important neural locus for the interaction between eye movements and multisensory speech perception.

Relationship to previous studies

Our demonstration of a link between the McGurk effect and eye movements and differs from the only previous study on the topic, that of Paré and colleagues (Paré et al., 2003) who concluded that “gaze fixations did not predict the likelihood of perceiving the McGurk effect”. There are many differences in methodology and analysis between the studies, such as the use of invasive scleral search coils in the Paré study and non-invasive infrared eye tracking in the present study, making a direct comparison of the results difficult. One key methodological difference may be the initial position of the eyes in the two studies. Participants in our study began each trial by fixating on a target presented at one of the four corners of the stimulus presentation location. After this peripheral fixation, the trial was initiated and the face video was presented centrally. All participants moved their eyes from the initial peripheral fixation location to a more centrally located fixation location once the face video appeared, forcing participants to choose a preferred fixation location within the face. In contrast, in the Paré study, participants did not have an initial peripheral fixation target; therefore, their eyes likely were located centrally at the beginning of the trial, making it unnecessary for them to change their eye position to foveate the centrally presented face. When viewing faces, a center start position results in longer saccade latencies and qualitative differences in eye movements compared with peripheral start positions (Arizpe, Kravitz, Yovel, & Baker, 2012). In future studies, it will be important to measure eye movements as participants view both dynamic faces (as in the present study) and static faces (as in previous studies). This would provide an answer to the question of whether individual differences in eye movement patterns are consistent across different facial motion conditions.

Participants are more likely to fixate the mouth of the talker when identifying words and more likely to fixate the eyes when identifying the talker's emotion, gender or intonation (Buchan, Paré, & Munhall, 2007; Lansing & McConkie, 1999, 2003; Peterson & Eckstein, 2012). It is interesting to speculate that although weak perceivers did not perceive the McGurk effect as often as strong perceivers, they might have been better at a task involving emotional evaluation.

The role of foveation

The most common fixation location besides the mouth was the talker's eyes, located 7° from the mouth. One possible explanation for our finding is that participants who fixated the eyes instead of the mouth were unable to extract visual speech information from the (peripherally viewed) talker's mouth. The "invisible unless fixated" explanation is argued against by a number of lines of evidence. The study of Paré and colleagues (Paré, Richler, ten Hove, & Munhall, 2003) found that participants experienced the McGurk effect even when fixating targets up to 40° from the mouth, much greater than the distance between the mouth and the eye ROIs in our experiment. Even low spatial acuity information about visual speech, as is available from the periphery, is sufficient to produce the McGurk effect (MacDonald et al., 2000; Summerfield & McGrath, 1984). Patients with central vision loss due to macular degeneration still report the McGurk effect (Wilson et al., 2008). Older adults with impaired visual acuity are still able to make use of visual speech cues (Hickson et al., 2004, Gagne and Wittich, 2009). As a direct test of the "invisible unless fixated" explanation, we performed a control experiment in which 16 participants viewed visual-only speech videos consisting of the syllables "ba" or "ga" while fixating a crosshairs overlaid on the right eye of the talker (7° from the center of the mouth). Fixation was verified using eye tracking; trials in which fixation was broken were discarded. Participants were able to accurately discriminate the mouth movements made by the talker (accuracy at ceiling, 99 %) despite not fixating the mouth region. Taken together, this body of evidence shows that the "invisible unless fixated" argument is not a tenable explanation for the observed link between eye movements and McGurk perception.

References

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology, 15,* 839–843.

Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research, 183,* 399–404.

Arizpe, J., Kravitz, D. J., Yovel, G., & Baker, C. I. (2012). Start position strongly influences fixation patterns during face processing: Difficulties with eye movements as a measure of information use. *PloS One, 7*(2), e31106. doi:10.1371/journal.pone.0031106

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004a). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nature Neuroscience, 7*(11), 1190–1192. doi:10.1038/nn1333

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004b). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron, 41*(5), 809–823. doi:10.1016/S0896-6273(04)00070-4

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *Journal of Neuroscience, 30*(7), 2414–2417. doi:10.1523/JNEUROSCI.4865-09.2010

Buchan, J. N., Paré, M., & Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Social Neuroscience, 2*(1), 1–13. doi:10.1080/17470910601043644

Buchan, J. N., & Munhall, K. G. (2012). The effect of a concurrent working memory task and temporal offsets on the integration of auditory and visual speech information. *Seeing and Perceiving, 25,* 87–106.

Gagne, J. P., & Wittich, W. (2009). *Visual impairment and audiovisual speech-perception in older adults with acquired hearing loss*. Chicago, IL: PHONAK Hearing Care for Adults.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*(11), 498–504. doi:10.1016/j.tics.2003.09.006

Hickson, L., Hollins, M., Lind, C., Worrall, L., & Lovie-Kitchin, J. (2004). Auditory-visual speech perception in older people: The effect of visual acuity. *The Australian and New Zealand Journal of Audiology, 26,* 3–11.

Lansing, C. R., & McConkie, G. W. (1999). Attention to facial regions in segmental and prosodic visual speech perception tasks. *Journal of Speech, Language, and Hearing Research, 42,* 526–539. doi:10.1044/jslhr.4203.526

Lansing, C. R., & McConkie, G. W. (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Percept Psychophys, 65*(4), 536–152.

MacDonald, J., Andersen, S., & Bachmann, T. (2000). Hearing by eye: How much spatial degradation can be tolerated? *Perception, 29*(10), 1155–1168. doi:10.1068/p3020

Magnotti, J. F., & Beauchamp, M. S. (2014). The noisy encoding of disparity model of the McGurk effect. *Psychonomic Bulletin & Review.* doi:10.3758/s13423-014-0722-2

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748. doi:10.1038/264746a0

Mehoudar, E., Arizpe, J., Baker, C. I., & Yovel, G. (2014). Faces in the eye of the beholder: Unique and stable eye scanning patterns of individual observers. *Journal of Vision, 14*(7), 6. doi:10.1167/14.7.6

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *NeuroImage, 59*(1), 781–787. doi:10.1016/j.neuroimage.2011.07.024

Nath, A. R., Fava, E. E., & Beauchamp, M. S. (2011). Neural correlates of interindividual differences in children's audiovisual speech perception. *Journal of Neuroscience, 31*(39), 13963–13971. doi:10.1523/JNEUROSCI. 2605-11.2011

Paré, M., Richler, R. C., ten Hove, M., & Munhall, K. G. (2003). Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the McGurk effect. *Perception & Psychophysics, 65*(4), 553–567. doi:10.3758/BF03194582

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences of the United States of America, 109*(48), E3314–E3323. doi:10.1073/pnas.1214269109

Peterson, M. F., & Eckstein, M. P. (2013). Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychological Science, 24*(7), 1216–1225. doi:10.1177/0956797612471684

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience, 18*(6), 2188–2199.

Seilheimer, R. L., Rosenberg, A., & Angelaki, D. E. (2014). Models and processes of multisensory cue combination. *Current Opinion in Neurobiology, 25,* 38–46. doi:10.1016/j.conb.2013.11.008

Summerfield, Q., & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 36*(1), 51–74. doi:10.1080/14640748408401503

Thomas, S. M., & Jordan, T. R. (2004). Contributions of oral and extraoral facial movement to visual and audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 30*(5), 873–888. doi:10.1037/0096-1523.30.5.873

Wilson, A., Wilson, A., Ten Hove, M. W., Pare, M., & Munhall, K. G. (2008). Loss of central vision and audiovisual speech perception. *Visual Impairment Research, 10,* 23–34.

Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication, 26,* 23–43. doi:10.1016/SO167-6393(98)00048-X

Zerouali, Y., Lina, J.-M., & Jemel, B. (2013). Optimal eye-gaze fixation position for face-related neural responses. *PloS One, 8*(6), e60128. doi:10.1371/journal.pone.0060128