Moreover, top-down characteristics (e.g., face familiarity) also affect the McGurk effect. Walker, Bruce, and O'Malley (1995) found that participants who are familiar with the face report less McGurk percepts than those who are unfamiliar with the face when the face and voice are from different persons. These studies imply that the McGurk effect can be modulated by either bottom-up or top-down attentional characteristics.

Value-driven attentional capture is a recently proposed mechanism of attention in addition to the salience-driven (bottom-up) and goal-driven (top-down) mechanisms (Anderson, 2013). Previous studies on value-driven attention, conducted in the visual domain, have shown that when stimuli are learned to predict reward, these stimuli would gain a competitive advantage that promotes attentional selection even when they are nonsalient and/or task irrelevant in perception (e.g., Anderson, 2013; Anderson, Laurent, & Yantis, 2011; Wang, Duan, Theeuwes, & Zhou, 2014; Wang, Yu, & Zhou, 2013). A few studies extended the concept to the cross-modal domain, showing that reward-associated sounds could affect the processing of visual stimuli. For example, auditory stimulus associated with high reward can increase the sensitivity of the perception of visual stimulus appearing simultaneously, even when sounds and reward associations are both irrelevant to the visual task (Pooresmaeili et al., 2014). Anderson (2016) demonstrated that relative to neutral sounds, previously reward-associated sounds capture attention, interfering more strongly with the performance of a visual task. However, it is currently unknown whether and how the value-driven mechanism of attention works in audiovisual speech perception in which the visual and auditory information is complex and highly relevant.

The face is the most important visual information in audiovisual speech perception. Raymond and O'Brien (2009) showed that a value-driven effect could be observed in the visual processing of faces. They trained participants to learn particular face–reward associations and then asked them to recognize whether a target face in an attentional blink (AB) task had been presented in the training phase. The authors found that the face recognition performance was higher for reward-associated faces compared with non-reward-associated faces. Moreover, while non-reward-associated-faces trials showed a typical AB effect, reward-associated-faces trials showed no AB effect, breaking through the constraints of AB on attentional selection. This study implies that value-associated faces would capture more attention and would be processed better than non-value-associated faces.

Considering that dynamic facial movements contain lots of visual information (e.g., mouth movements, other facial muscle movements, eye gaze), it is necessary to explore how people extract visual information from the dynamic talking faces for the purpose of audiovisual speech perception. By using the McGurk task and monitoring eye movements, previous studies have found that the mouth area of the talking face plays a critical role in the effect of visual information on audiovisual speech perception. In particular, perceivers show less time looking at the mouth area when the McGurk proportion decreases (i.e., when they make less use of visual information). For example, as the visual resolution of faces decreases, perceivers report fewer McGurk percepts and spend less time looking at the mouth area (Wilson, Alsius, Paré, & Munhall, 2016). Adding a concurrent cognitive task to the main McGurk task would decrease the McGurk proportion as well as the time looking at the mouth area (Buchan & Munhall, 2012). In addition, weak McGurk perceivers (i.e., perceivers who perceive the McGurk effect less frequently in general) fixate less on the talker's mouth area compared with strong McGurk perceivers (Gurler, Doyle, Walker, Magnotti, & Beauchamp, 2015; Hisanaga, Sekiyama, Igasaki, & Murayama, 2016).

The current study investigates whether and how the value-driven mechanism of attention works in audiovisual speech perception by using a training-test paradigm, which is often used in value-driven attention studies (e.g., Anderson, 2013, 2016; Anderson et al., 2011; Raymond & O'Brien, 2009; Wang et al., 2013). In the training phase, participants were asked to discriminate the gender of face pictures, in which;51

Committee for Protecting Human and Animal Subjects, School of Psychological and Cognitive Sciences, Peking University. Three participants were excluded, because they reported in the after-experiment interview that they concentrated on visual information in deciding the identity of spoken syllables during the task; another participant was excluded because of astigmatism that leads to poor quality of eye movement data. The remaining 28 participants were included in data analyses (17 females, mean age = 21.79, $SD = 2.10$). A power analysis was conducted by using G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009; Faul, Erdfelder, Lang, & Buchner, 2007). Since we did not find a previous study that is similar to the current investigation, we referred to a study concerning the modulation of attention on the McGurk effect (Alsius et al., 2005). This study showed a moderate effect size (Experiment 1, Cohen's $d = 0.694$). We estimated that we would need at least 19 participants, given Cohen's $d = 0.694$, $\alpha = 0.05$, and power = 80%. In the present study, the number of participants (28) is higher than the suggested number (19).

Apparatus and materials

Visual stimuli were presented on a 17-inch SONY CRT mon-

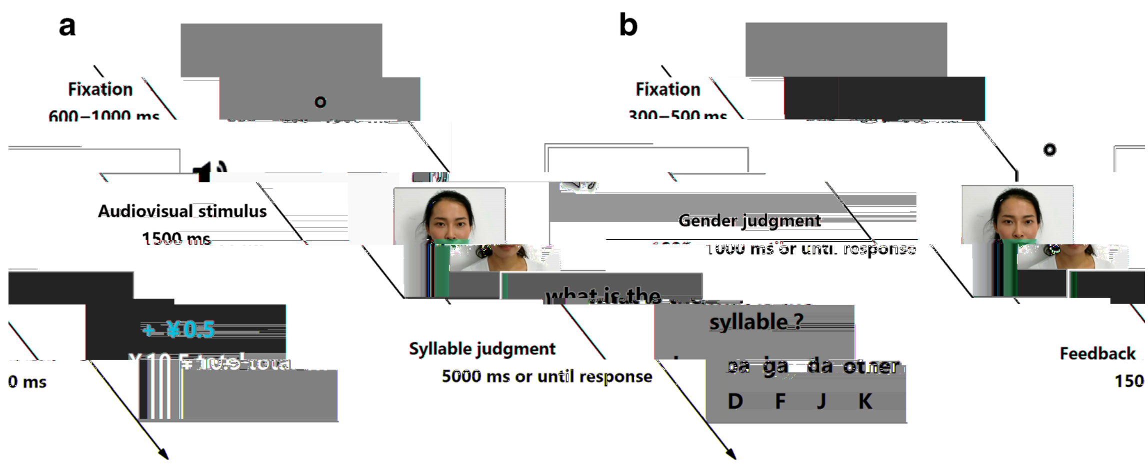

Fig. 1 Trial structure in the training-test paradigm. **a** The training phase. Participants were required to judge the gender of a face picture as quickly and as accurately as possible. A correct response for two faces (counterbalanced across participants) was associated with 80% probability of monetary reward and 20% probability of no reward; correct responses for the other two faces were not associated with reward. The feedback indicating the outcome of the current trial (correct and reward: "+ ¥ 0.5"; correct but no reward: "+0"; wrong: "Wrong!"; time out: "Too slow!"), and the total reward up to the current trial would be presented after response. **b** The test phase. Participants were instructed to identify which syllable ("ba," "ga," "da," and "other") was said by the talker in a video clip by pressing a corresponding button on the keyboard. A fixation randomly presented at one of the four corners of an invisible bounding box, followed by a video clip in which a talker said a syllable (McGurk stimuli, and audiovisual congruent/incongruent stimuli as fillers). No reward or feedback was delivered

on what they heard, and we excluded the participants who identified the syllable based on what they saw as we previously described. In addition, given that participants had to keep head static with a chin rest to ensure the collection of accurate eye movement data, they could not open their mouth to report what syllable they had identified. A button-pressing method was thus a viable way to record their responses (see also Fernández, Macaluso, & Soto-Faraco, 2017).

No feedback was provided in the test phase. The test phase consisted of 20 blocks with 24 trials each. Each type of audiovisual stimuli (i.e., congruent, incongruent, and McGurk stimuli, with congruent and incongruent stimuli acting as fillers) for each talker was equally presented in each block. The test phase consisted of two different types of faces across trials—that is, faces associated with reward in the training phase (i.e., reward-associated faces) and faces not associated with reward in the training phase (i.e., non-reward-associated faces).

Eye tracking was performed at a sampling rate of 1000 Hz during the training and test phases. Nine-point calibration and verification were performed at the beginning of each experimental phase; drift correction (and recalibration if necessary) was performed at the beginning of each block. Participants were required to look at the hollow circle at the beginning of each trial; if participants' fixations did not locate on the hollow circle within 5,000 ms, drift correction (and recalibration if necessary) would be performed again. During the display of video clips, no fixation sign was presented, and participants were not explicitly instructed to fixate on the face or any other location (i.e., free viewing).

Data analyses

Behavioral data

For the training phase, trials with reaction times (RTs) more than 2.5 standard deviations above or below the mean RT of each condition for each participant were excluded (2.4% of all the trials). A paired-samples t test was conducted on the mean RT and accuracy to examine the reward association effect.

For the test phase, we focused on responses to the McGurk stimuli. The proportion of each response category (i.e., "ba," "ga," "da," and "other") was calculated by dividing the number of responses for each category by the total number of McGurk trials (i.e., the congruent or incongruent trials were not included). For the McGurk effect, we first took a liberal definition—that is, a response of any percept (including "da," "ga," "other") other than the auditory target ("ba") was classified as a McGurk percept (e.g., Fernández et al., 2017; Gurler et al., 2015). We also tested the data based on a more conservative definition (that is, only a response of "da" was classified as a McGurk response) to evaluate the robustness of results. A paired-samples t test was conducted to compare the two conditions.

Signal detection analysis

We conducted a signal detection analysis for the behavior data of the test phase. The advantage of using a signal detection analysis is that the inclusion of filler trials could lead to improved estimates due to the additional data incorporated, and, more importantly, we could separately compare the response criterion (c) and the discriminability (d') for reward-associated and non-reward-associated faces. Specifically, a McGurk stimulus was regarded as a signal trial, and a filler stimulus (either the congruent or incongruent stimulus) was regarded as a noise trial. A response defined as a McGurk percept was

Results

Reaction time and accuracy in the training phase

Participants identified the gender of reward-associated faces significantly faster than non-reward-associated faces (446 vs. 451 ms), $t(27) = -2.239$, $p = .034$, $d = 0.423$, demonstrating that participants had learnt the face–reward association. There was no significant difference in terms of response accuracy between reward-associated and non-reward-associated faces (97.9% vs. 97.7%), $t\,(27) = 1.026$, $p = .314$.

McGurk effect in the test phase

The average accuracies in responding to the filler stimuli (i.e., congruent and incongruent stimuli) in different conditions were very high, ranging from 95.7% to 97.2%, indicating that participants performed the task carefully and effectively. For the McGurk stimuli, the proportion of each response category under each condition is shown in Table 1. According to the liberal definition of the McGurk percept (i.e., a response of any percept other than the auditory target was classified as a McGurk percept), the McGurk proportion was significantly higher for reward-associated faces than for non-reward-associated faces (60.1% vs. 49.9%), $t(27) = 2.438$, $p = .022$, $d = 0.461$, which was consistent with our hypothesis.

According to the conservative definition of the McGurk percept (i.e., only a response of "da" was classified as a McGurk percept), the McGurk proportion was marginally higher for reward-associated faces than for non-reward-associated faces (52.2% vs. 43.6%), $t(27) = 1.788$, $p = .085$, $d = 0.338$, which was consistent with the pattern reported above. In addition, the proportion of "ba" response (i.e., the true auditory target) was significantly lower for reward-associated faces than for non-reward-associated faces (39.9% vs. 50.1%), $t(27) = 2.438$, $p = .022$, $d = 0.461$,

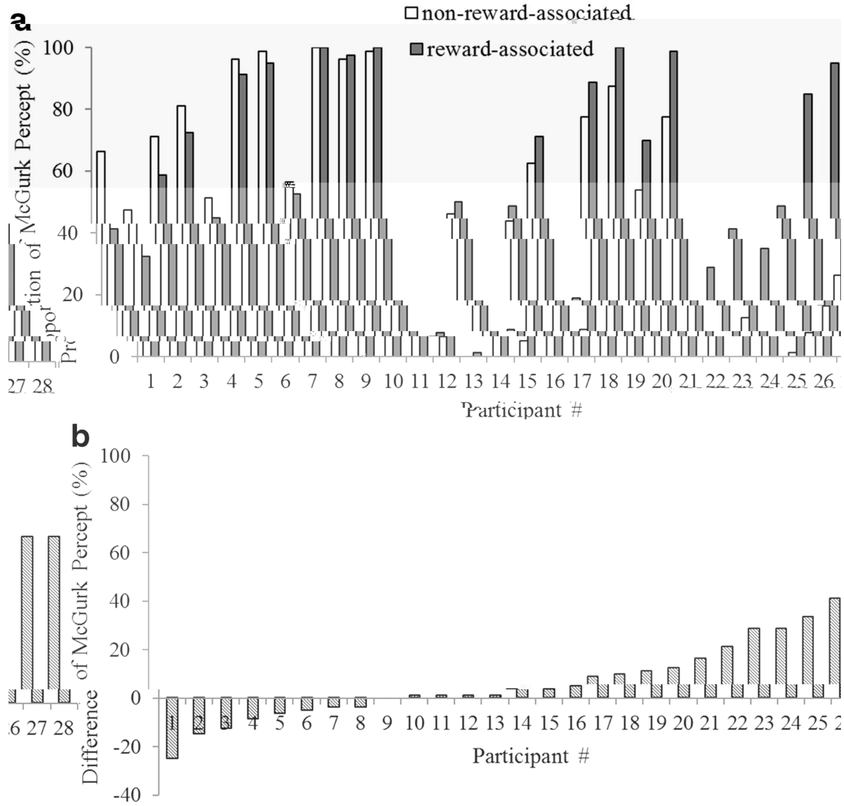

Fig. 3 Individual differences in the McGurk proportion based on the liberal definition of the McGurk percept. **a** Each participant's McGurk proportion for reward-associated and non-reward-associated faces. **b** The difference of the McGurk proportion between reward-associated and non-reward-associated faces for each participant

looking time on the mouth IA was marginally higher than on the eyes IA (p = .073), and there were no differences between other IAs (all ps > .269). The main effect of time period was also significant, $F(2, 54)$ = 126.548, p < .001, η_p^2 = .824. Planned comparisons showed that the proportion of looking time on the first time period was significantly lower than the other two time periods (all ps < .001), and there was no significant difference between the second and third periods (p = .293). The main effect of reward association was not significant, $F(1, 27)$ = 0.040, p = .843, η_p^2 = .001. The IA × Time Period interaction was significant, $F(6, 162)$ = 15.819, p < .001, η_p^2 = .369, so was the IA × Reward Association interaction, $F(3, 81)$ = 2.897, p = .040, η_p^2 = .097. The Time Period × Reward association interaction was not significant, $F(2, 54)$ = 0.558, p = .576, η_p^2 = .020. Importantly, the three-way interaction between IA, time period, and reward association was significant, $F(6, 162)$ = 2.373, p = .032, η_p^2 = .081, and we further explore this interaction below.

We conducted 4 (IA: mouth vs. eyes vs. nose/cheek vs. forehead) × 2 (reward association: reward-associated vs. non-reward-associated) repeated-measures ANOVA for the first, second, and third time periods, respectively. For the first time period (0 – 500 ms of the video clips; see Fig. 5, left panel), only the main effect of IA was significant, $F(3, 81)$ = 20.573, p < .001, η_p^2 = .432. Planned comparisons showed that the proportion of looking time on the forehead IA was significantly lower than other three IAs (all ps < .001), the proportion of looking time on the eyes IA was significantly lower than mouth IA (p = .029) and nose/cheek IA (p < .001), and there was no significant difference between mouth and nose/cheek IAs. The main effect of reward association and the interaction effect were not significant (all ps > .493).

For the second time period (500–1,100 ms of the video clips; see Fig. 5, middle panel), the main effect of IA was significant, $F(3, 81)$ = 22.510, p < .001, η_p^2 = .455. Planned comparisons showed that the proportion of looking time on the forehead IA was significantly lower than the other three IAs (all ps < .001), the proportion of looking time on the mouth IA was significantly higher than the other three IAs (all ps < .005), and there was no significant difference between eye and nose/cheek IA (p = .277). The main effect of reward association was not significant, $F(1, 27)$ = 0.388, p = .538, η_p^2 = .014. Importantly, the IA × Reward Association interaction was significant, $F(3, 81)$ = 2.908, p = .040, η_p^2 = .097. Planned t tests on simple effects showed that the proportion of looking time was significantly higher for reward-associated faces than for non-reward-associated faces (28.3% vs. 25.9%) on the nose/cheek IA, $t(27)$ = 2.328, p = .028, d = 0.440, although this effect did not reach significance if more stringent statistical tests were applied. This effect did not appear on

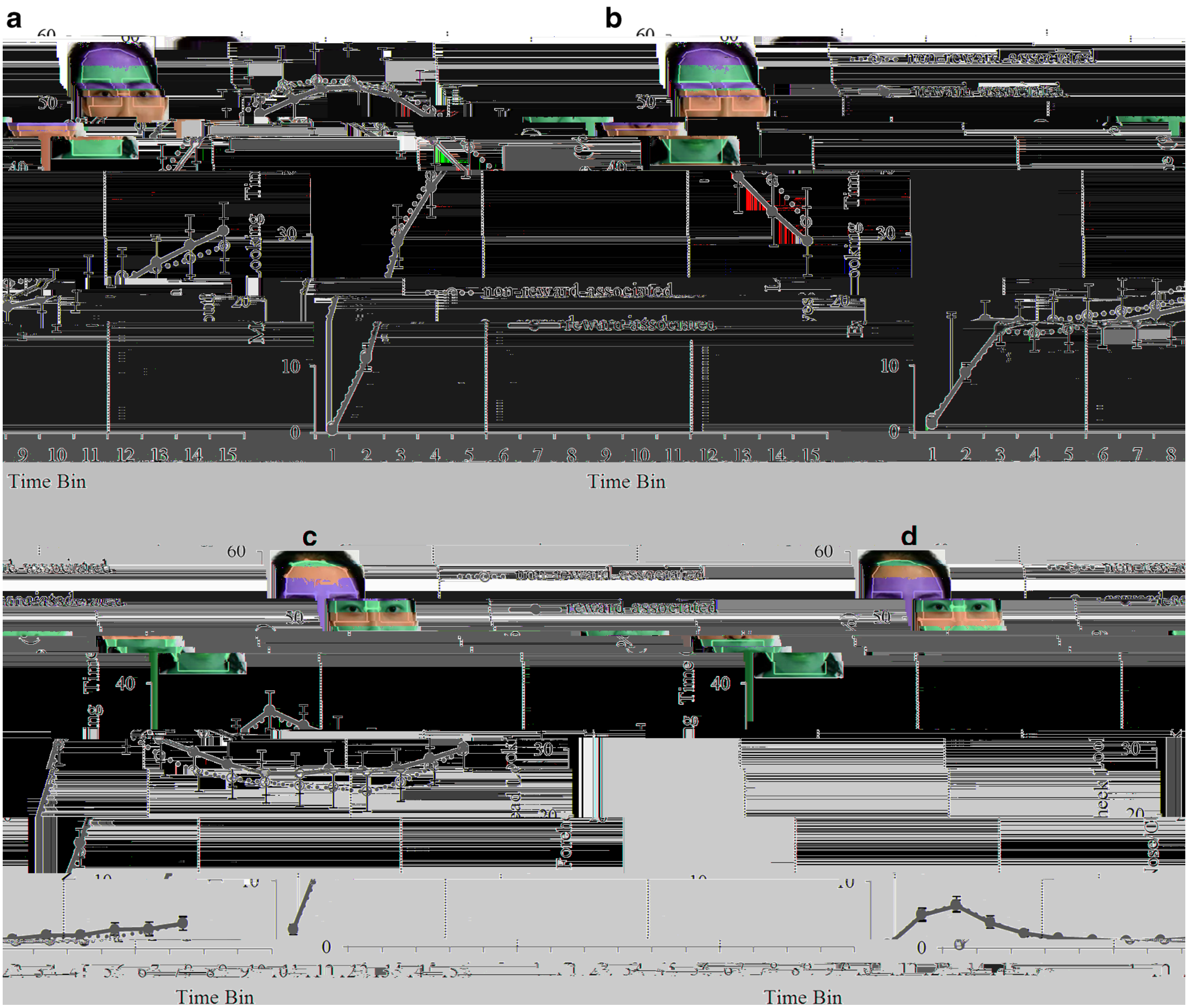

Fig. 4 Time course for the proportion of looking time on the interest area (IA) for (**a**) mouth, (**b**) eyes, (**c**) nose/cheek, and (**d**) forehead with standard errors. The whole McGurk stimulus presentation (1,500 ms) was divided into 15 time bins (100 ms for each) to further illustrate the change of the proportion of looking time on different IAs over time. The vertical lines separated time periods (i.e., 0–500 ms, 500–1,100 ms, and 1,100–1,500 ms of stimulus presentation) that we used in the statistical analyses

other IAs (all $ps > .101$), suggesting that compared with non-reward-associated faces, participants looked at reward-associated faces longer, but only on the extraoral facial area, which is somewhat inconsistent with our original hypothesis.

For the third time period (1,100–1,500 ms of the video clips; see Fig. 5, right panel), the main effect of IA was significant, $F(3, 81) = 12.763$, $p < .001$, $\eta_p^2 = .321$. Planned comparisons showed that the proportion of looking time on the forehead IA was significantly lower than the other three IAs (all $ps < .001$), and there were no significant differences between these three IAs (all $ps > .917$). The main effect of reward association was not significant, $F(1, 27) = 0.241$, $p = .627$, $\eta_p^2 = .009$. But the IA × Reward Association interaction was significant, $F(3, 81) = 3.408$, $p = .021$, $\eta_p^2 = .112$. Planned t tests on simple effects showed that the proportion of looking time was significantly lower for reward-associated faces than for non-reward-associated faces (36.0% vs. 40.0%) on the mouth IA, $t(27) = -2.122$, $p = .043$, $d = 0.401$, although this effect would not survive if more stringent statistical tests were applied. There were not reward association effects on other IAs (all $ps > .098$). The result here was surprising, as it indicated that participants were less likely to look at the mouth area of reward-associated faces, relatively to non-reward-associated faces, even though visual information in this area was thought to be a causer of McGurk effect. This is in contradictory to our original hypothesis.

We also collapsed data over the three time periods and conducted a 4 (IA: mouth vs. eyes vs. nose/cheek vs. forehead) × 2 (reward association: reward-associated vs. non-reward-associated) ANOVA. The IA × Reward Association interaction was significant (see Fig. 6a). Planned t tests on simple effects showed that the proportion of looking time on the nose/cheek IA was marginally higher for reward-associated faces than for non-reward-associated faces (26.9% vs. 25.1%), $t(27) = 2.034$, $p = .052$, $d = 0.384$, although this effect would not survive when more stringent statistical tests were applied. The pattern here again demonstrated the importance of extraoral facial areas in the value-driven McGurk effect.

Eye movements in the test phase: The proportion of fixation number

Figure 7 illustrates the change of the proportion of fixation number in different IAs over time. The 4 (IAs: mouth vs. eyes vs. nose/cheek vs. forehead) × 3 (time period: the first period vs. the second period vs. the third period) × 2 (reward association: reward-associated vs. non-reward-associated) repeated-measures ANOVA showed a pattern almost identical to what was reported for the proportion of looking time, above. Collapsing the data over the three time periods, we once again observed the interaction between IA and reward association, $F(3, 81) = 3.276$, $p = .025$, $\eta_p^2 = .108$. Planned t tests on simple effects showedowed a pa.39999961(ti)[(Co)16.5(ll)Gp of loOVtee

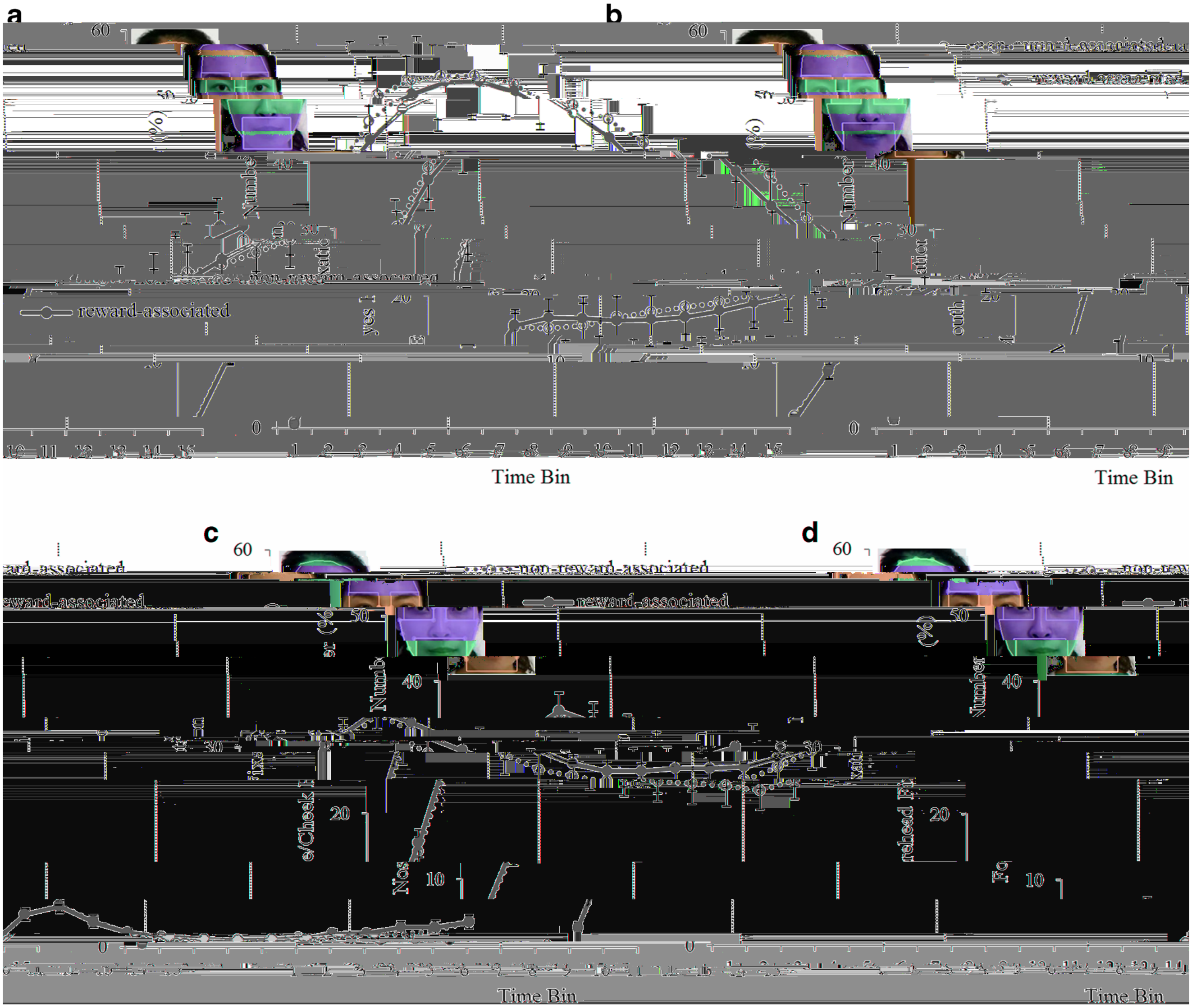

Fig. 7 Time course for the proportion of fixation number on the interest area (IA) for (**a**) mouth, (**b**) eyes, (**c**) nose/cheek, and (**d**) forehead, with standard errors. The whole McGurk stimulus presentation (1,500 ms) was divided into 15 time bins (100 ms for each) to further illustrate the change of the proportion of fixation number on different IAs over time. The vertical lines separated time periods (i.e., 0–500 ms, 500–1,100 ms, and 1,100–1,500 ms of stimulus presentation) that we used in the statistical analyses

central target face in each trial, we analyzed the number of first fixation in a particular IA, excluding the fixations outside of the face. We found only a significant main effect of IA, $F(3, 81) = 17.038$, $p < .001$, $\eta_p^2 = .387$, with more fixations on the nose/cheek IA (36.66%) and the mouth IA (35.31%) than on the eye IA (19.55%) and the forehead IA (7.90%).

Correlation analysis

Given that, in the training phase, participants spent the longest time looking at the nose/cheek area (73.91%) than any other areas, the correlation analysis was first conducted for this area. Over participants, the difference of the proportion of looking time at nose/cheek area between reward-associated and non-reward-associated faces in the training phase positively correlated with the difference of the McGurk proportion, either liberally or conservatively defined, between the two conditions in the test phase (see Fig. 8a), $r = .381$, $p = .045$; $r = .424$, $p = .024$. The same pattern was observed for the proportion of fixation number, $r = .350$, $p = .067$; $r = .397$, $p = .037$.

In addition, the difference of the proportion of looking time at the nose/cheek area between reward-associated and non-reward-associated faces in the training phase negatively correlated with the difference of the proportion of looking time at the mouth area between the two conditions in the test phase (see Fig. 8b), $r = -.333$, $p = .083$. The same pattern was obtained for the proportion of fixation number, $r = -.318$, $p = .099$.

Since that, in the test phase, the reward-association effect was observed in the nose/cheek area and the mouth area, the correlation analysis was conducted for these two areas. In the test phase, however, the difference of the proportion of looking time at the nose/cheek area between reward-associated faces and non-reward-associated faces did not correlate with the difference of McGurk proportion, either liberally or conservatively defined, between the two conditions, $r = -.158$, $p = .421$; $r = -.071$, $p = .719$. The null effect was also observed for the proportion of fixation number, $r = -.179$, $p = .362$; $r = -.121$, $p = .538$. Similarly, in the test phase, the difference of the proportion of looking time at the mouth area between reward-associated and non-reward-associated faces did not correlate with the difference of McGurk proportion, either liberally or conservatively defined, between the two conditions, $r = .030$, $p = .880$; $r = -.090$, $p = .648$. The null effect was also observed for the proportion of fixation number, $r = .045$, p

speech perception context, it is possible that multisensory integration processing was directly facilitated by reward association in the present study, resulting in more McGurk percepts for reward-associated faces. Nevertheless, it should be noticed that the McGurk effect cannot be equated with multisensory integration, because much more is involved with the McGurk effect than just multisensory integration, such as conflict resolution (e.g., Fernández et al., 2017; see also Alsius et al., 2018 for a review).

Furthermore, it should be mentioned that our results seem to show contrasts with a previous study in which Walker et al. (1995) investigated the influence of face familiarity (i.e., a form of value to some extent) on the McGurk effect, and found that participants who were familiar with the face reported less McGurk percepts than those who were unfamiliar with the face when the face and voice were from different persons. However, there are key differences between the studies. Participants in our study did not know the talkers before, and all the talkers' faces in the training phase were static pictures and appeared at the same frequency. That is, participants had the same familiarity of all the talkers' static faces, and had no prior knowledge of the talkers' dynamic facial movements. Walker et al. (1995) defined the familiarity in terms of participants having had face-to-face interactions with the talker in daily life, which means that participants were familiar not only with the talkers' static faces but also with the talkers' dynamic facial movements and voices. As the authors mentioned, participants were able to use their prior knowledge of those familiar faces (expectations of what speech events were likely and of how these events were realized through dynamic facial movements); the incongruence between the visual and auditory modality was thus easier to be detected, resulting in less report of McGurk percepts. The authors also found that when the face and voice were from the same person, there were no differences in McGurk percepts between the participants who were familiar with the faces and the participants who were unfamiliar with them, a pattern recently replicated (Magnotti et al., 2018

Munhall 2003), suggesting that information about mouth movements can be obtained from other areas in non-mouth-looking conditions.

Second, extraoral facial movements may provide useful visual information apart from the oral facial movements, which helps to elicit the McGurk effect. Thomas and Jordan (2004) manipulated the movements of the mouth and other facial areas independently, and found that the extraoral movements could promote the identification of audiovisual speech even when the mouth is kept static or removed from the face. Jordan and Thomas (2011) further found that the McGurk effect is observable even when the talker's face is occluded horizontally or diagonally (i.e., when the mouth area is occluded). In the present study, longer looking time and fixated more often on the extraoral area of reward-associated faces, compared with non-reward-associated faces, might help participants process the visual information provided by extraoral area, resulting in higher McGurk proportion.

To conclude, by associating faces with or without monetary reward in the training phase, we demonstrated that individuals could in the subsequent test phase report more McGurk percepts for reward-associated faces, relative to non-reward-associated faces, indicating that value-associated faces enhance the influence of visual information on audiovisual speech perception. The signal detection analysis revealed that participants have lower response criterion and higher sensory discriminability for reward-associated faces than for non-reward-associated faces, indicating that when the talking faces are associated with value, individuals tend to make more use of visual information in processing the McGurk stimuli. Surprisingly, we found that participants in the test phase had more looking time and number of fixations on the nose/cheek area of reward-associated faces than non-reward-associated faces; the opposite pattern was found for the mouth area. The correlation analysis revealed that the more participants looked at the nose/cheek area in the training phase due to reward, the more McGurk effect occurred in the test phase for reward-associated faces. These findings suggest that associating reward with a face may increase the attentional priority of the extraoral area, which contributes to the audiovisual speech perception.

Acknowledgement We thank Miss Yueyuan Zheng, Miss Jiayi Li, and Miss Muhang Li for proofreading, and Mrs. Haining Bi for allowing us to use her face picture in this article. This work was supported by the Natural Science Foundation of China (Grant No. 31630034) and the National Basic Research Program of China (973 Program: 2015CB856400).

Compliance with ethical standards

Conflict of Interest The authors declare no conflict of interest.

Open practices statement The data and materials are available on request from the corresponding author.

The experiment was not preregistered.

References

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. Current Biology, 15(9), 839–843. https://doi.org/10.1016/j.cub.2005.03.046

Alsius, A., Paré, M., & Munhall, K. G. (2018). Forty years after hearing lips and seeing voices: The McGurk effect revisited. Multisensory Research, 31(1), 111–144. https://doi.org/10.1163/22134808-00002565

Anderson, B. A. (2013). A value-driven mechanism of attentional selection. Journal of Vision, 13(3), 103–104. https://doi.org/10.1167/13.3.7

Anderson, B. A. (2016). Value-driven attentional capture in the auditory domain. Attention, Perception, & Psychophysics, 78(1), 242–250. https://doi.org/10.3758/s13414-015-1001-7

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. Proceedings of the National Academy of Sciences of the United States of America, 108(25), 10367–10371. https://doi.org/10.1073/pnas.1104047108

Brainard, D. H. (1997). The Psychophysics Toolbox. Spatial Vision, 10(4), 433–436. https://doi.org/10.1163/156856897X00357

Brancazio, L., Miller, J. L., & Paré, M. A. (2003). Visual influences on the internal structure of phonetic categories. Perception & Psychophysics, 65(4), 591–601. https://doi.org/10.3758/BF03194585

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 information. Seeing & Perceiving, 25(1), 87–106. https://doi.org/10.1163/187847611X620937

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. G. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191. https://doi.org/10.3758/BF03193146

Fernández, L. M., Macaluso, E., & Soto-Faraco, S. (2017). Audiovisual integration as conflict resolution: The conflict of the McGurk illusion. Human Brain Mapping, 11(38), 5691–5705. https://doi.org/10.1002/hbm.23758

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. Journal of the Acoustical Society of America, 108(1), 1197–1208. https://doi.org/10.1121/1.1288668

Gurler, D., Doyle, N., Walker, E., Magnotti, J., & Beauchamp, M. (2015). A link between individual differences in multisensory speech perception and eye movements. Attention, Perception, & Psychophysics, 77(4), 1333–1341. https://doi.org/10.3758/s13414-014-0821-1

Hisanaga, S., Sekiyama, K., Igasaki, T., & Murayama, N. (2016). Language/culture modulates brain and gaze processes in audiovisual speech perception. Scientific Reports, 6, 35265. https://doi.org/10.1038/srep35265

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. Attention, Perception, & Psychophysics, 73(7), 2270–2285. https://doi.org/10.3758/s13414-011-0152-4

Magnotti, J. F., Smith, K. B., Salinas, M., Mays, J., Zhu, L. L., & Beauchamp, M.S. (2018). A causal inference explanation for enhancement of multisensory integration by co-articulation. Scientific Reports, 8, 18032. https://doi.org/10.1038/s41598-018-36772-8

Mallick, D. B., Magnotti, J. F., & Beauchamp, M. S. (2015). Variability and stability in the McGurk effect: Contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review, 22*(5), 1299–1307. https://doi.org/10.3758/s13423-015-0817-4

Marques, L. M., Lapenta, O. M., Costa, T. L., & Boggio, P. S. (2016). Multisensory integration processes underlying speech perception as revealed by the McGurk illusion. *Language, Cognition and Neuroscience, 31*(9), 1115–1129. https://doi.org/10.1080/23273798.2016.1190023

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

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. https://doi.org/10.1073/pnas.1214269109

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. *Attention, Perception, & Psychophysics, 65*(4), 553–567. https://doi.org/10.3758/bf03194582

Pooresmaeili, A., FitzGerald, T. H. B., Bach, D. R., Toelch, U.,