Consciousness and Cognition 19 (2010) 135–143

Contents lists available at ScienceDirect

Consciousness and Cognition

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

# When one sees what the other hears: Crossmodal attentional modulation for gazed and non-gazed upon auditory targets

Pines Nuku, Harold Bekkering *

*Radboud University Nijmegen -Donders Institute for Brain, Cognition and Behaviour Centre for Cognition, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands*

ARTICLE INFO

*Article history:*
Received 1 May 2008
Available online 12 October 2009

*Keywords:*
Action–perception
Crossmodal attention
Inferential processes
Intention
Simulation
Rationalization

ABSTRACT

Three experiments investigated the nature of visuo-auditory crossmodal cueing in a triadic setting: participants had to detect an auditory signal while observing another agent's head facing one of the two laterally positioned auditory sources. Experiment 1 showed that when the agent's eyes were open, sounds originating on the side of the agent's gaze were detected faster than sounds originating on the side of the agent's visible ear; when the agent's eyes were closed this pat-tern of responses was reversed. Two additional experiments showed that the results were sensitive to whether participants could infer a hearing function on the part of the agent. When no ear was depicted on the agent, only a gaze-side advantage was observed (Experiment 2), but when the agent's ear was covered (Experiment 3), an ear side advantage was observed only when hearing could still be inferred (i.e., wearing the hat) but not when hearing was inferred to be diminished (i.e., wearing a helmet). The findings are discussed in the context of inferential and simulation processes and joint attention mechanisms.



## 1. Introduction

Imagine yourself joking with a colleague of yours about your new boss. While conversing, your colleague gazes to your right where the voice of your boss is, suddenly, clearly audible. In such a situation, you could orient yourself to what your colleague is saying, but more likely, you will try to hear the reaction of your boss and try to determine whether he was privy to these sensitive comments. This example illustrates one of the most ordinary and complex human activities: synthesizing information coming from auditory and visual channels with the aim of producing an appropriate response. The aim of the present study is to investigate the nature of visuo-auditory crossmodal orienting in joint attention (observer–object–agent) setting, trying to establish whether one's own auditory attention is facilitated by observing another's gaze direction.

Typically, crossmodal orienting requires participants to detect auditory and/or visual targets preceded by either visual and/or auditory primes. It is typically investigated by using a classical spatial-cueing paradigm (see Posner, 1980 for more on the spatial-cueing paradigm) where the onset of a prime in one modality affects the detection of a stimulus in another modality (McDonald, Teder-Salerjarvi, & Hillyard, 2000; Schmitt, Postma, & De Haan, 2000; for a review, see Spence & Driver, 2004). For example, in audio-visual cueing, occurring in both covert (Spence & Driver, 1997) and overt (Reisberg, Scheiber, & Potemken, 1981; but see Wolters & Schiano, 1989) orienting, the presentation of a sound does not only enhance the perceptual processing of subsequent visual stimuli, but it also facilitates the motor responses to visual stimuli appearing at the location nearby the auditory source (McDonald & Ward, 2000). Similarly, in visuo-auditory cueing, the presentation of a visual cue does also facilitate the detection of auditory targets (Gopher, 1973; Mazza, Turatto, Rossi, & Umiltà, 2006; Rorden & Driver, 1999).

* Corresponding author. Tel.: +31 24 3612632; fax: +31 24 3616066.
*E-mail address:* h.bekkering@donders.ru.nl (H. Bekkering).

1053-8100/$ - see front matter 
doi:10.1016/j.concog.2009.07.012

However, whereas there is little doubt that changes in luminance facilitate the detection of nearby auditory stimuli, it is still unclear whether observing a more complicated visual stimulus, such as another's gaze might facilitate the observer's attention toward the gazed auditory target location. Can one's gaze direction tell us what the other is listening, rather than seeing, and would this information help us in directing our own attention with the other's (visually) attended sound location?

Thanks to their peculiar morphology, that is the sclera-to-iris ratio, eyes seem to have evolved with the purpose of allowing easy discrimination of one's attended location (Kobayashi & Kohshima, 1997). That is why eyes are often considered to be a special reflexive attention triggering prime. Several studies have shown that gaze direction triggers the observer's gaze and visual attention in a reflexive manner. For example, infants do align preferentially with adult's eyes (Firestone, Turk-Browne, & Ryan, 2007; Henderson, Williams, & Falk, 2005; Hood, Willen, & Driver, 1998; Vecera & Johnson, 1995). Adults, on the other hand, orient with another's gaze direction even when they know that the observed gaze is both unpredictive (Emery, 2000; Friesen & Kingstone, 1998; Hietanen, 1999; Langton & Bruce, 1999; Langton, Watt, & Bruce, 2000) or counterpredictive (Downing, Dodds, & Bray, 2004; Friesen, Ristic, & Kingstone, 2004) to the observer's aim. Gazing, however, does not merely trigger the observer's visual attention: it plays a crucial role in human interaction and/or communication. The ability to share information and to comprehend the thoughts and intentions of others is a hallmark of human nature. Humans share one's experience (of observing an object or event) by following one's gaze and/or pointing gestures, and they use their eye contact and gestures to show or direct the attention of the people around them. The tendency to spontaneously direct attention to where someone else is looking, or joint attention, refers to a fundamental triadic (gazer–object–observer) process where another's gaze becomes the most reliable indicator of the another's intentions (Tomasello, Carpenter, Call, Behne, & Moll, 2005).

Considering the crucial role of another's gaze in both visuo-spatial orienting and social interaction the aim of the present study was two fold. First, it sheds some light on whether in a joint attention setting one's gaze cueing might affect one's audio-visual orienting. We predicted that the observation of an agent's gaze (visual prime) looking at the location of an upcoming sound (auditory target) would facilitate the observer's alignment of attention with the sound location. Second, since meaningful actions do automatically activate the observer's own cognitive repertoire (Gallese, Keysers, & Rizzolatti, 2004) we investigated whether the alignment of the observer's attention with the sound location might also occur when inferring the outcome of such crossmodal interactions, rather than merely following another's gaze direction.

## 2. Experiment 1

In a triadic version of a typical spatial-cueing paradigm we investigated whether the observation of another's gaze direction toward a sound location might facilitate the observer's attention toward that sound location. And this, in a joint attention setting where the participants represented the observer, the 3-D depiction of a static human-like head represented the agent, and the auditory targets represented the to-be-detected external target. The observers focused on the agent's head, and were required to report the sounds location propagating from either (by the agent) gazed upon location or from non-gazed upon location. We predicted that observing the agent gazing toward the upcoming target location would facilitate the observer's target detection. Observing the agent oriented away from the target location would not facilitate the target detection. In other words, we predicted that attentional facilitation would be the strongest when observing the agent gazing toward the auditory location. Agents with theirs eyes closed would not facilitate the observer's attention.

### *2.1. Method*

#### *2.1.1. Participants*

Sixteen participants (five males; mean age 22 years), all students from the University of Nijmegen, received either course credits or five Euro for their participation at the experiment. All had normal or corrected-to-normal vision and hearing, and all of them were right handed. They were naive as to the purpose of the experiment.

#### *2.1.2. Apparatus and Stimuli*

The prime, was presented on a flat-screen PC and consisted on 3-D grayscale representation of a human head, measuring approximately 10° (degree in visual angle) by 8°, which in half of the trials appeared as oriented 45° to the right, and in the remaining half of the trials it appeared as oriented 45° to the left side of the screen. In half of the trials were open and in the rest of the trials they were closed. The auditory target propagated from one of the two speakers positioned, at head level at an eccentricity of 56° to the right and 56° left of the central fixation. The sounds were created by a white-noise generator and lasted for 110 ms (see the upper panel in Fig. 1). In order to prevent observers from representing the space in terms of 'left-to-fixation' or 'right-to-fixation', we added at an eccentricity of 24° to the left and right of the fixation cross two additional non-functional speakers, leading the observers to believe that the target could propagate from any of four speakers, thus guessing the target location could not be derived by 'luck' alone.

#### *2.1.3. Design*

The combination of Head Orientation (head oriented 45° toward the speaker, head oriented 45° away from the speaker) and Gaze Type (eyes open, eyes closed) created 'head congruent' trials, where the target propagated from the head (with either open or closed eyes) primed location, and 'head incongruent' trials or trials where the target propagated from the

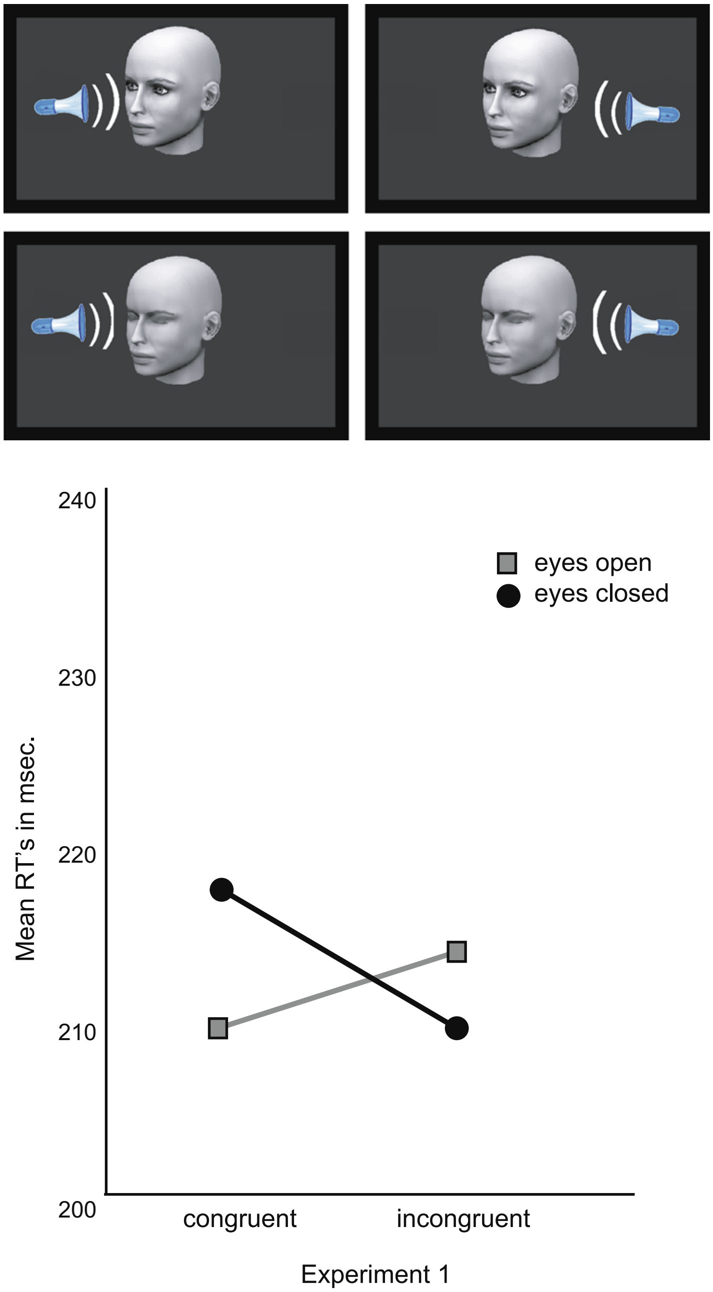


**Fig. 1.** The upper part of the figure illustrates the combination of head congruent/incongruent targets with closed/opens eyes. The lower part of the figure illustrates the mean reaction times, and shows that when the agent's eyes were open, auditory targets propagating from the eye side were detected faster the targets propagating from the ear side, whereas when they were closed the targets propagating from the eye side were detected slower than the targets propagating from the ear side.

unprimed location. Because the observers were instructed to ignore both eye and head orientation, the cueing effect would be measured by comparing the reactions to head congruent targets when the head appeared with the eyes open, with responses to head congruent targets when the eyes were closed.

Participants received 15 practice trials, followed by a block of 260 experimental trials including 40 catch trials (where no target was presented and no imperative response was expected) to control for random responses.

*2.1.4. Procedure*

A trial started with the fixation cross presented for 2000 ms and followed by the agent's head presentation oriented either to the left or to the right with the eyes either open or closed. After a variable interval of 200-, 300-, and 650-ms, used here to discourage random anticipatory responses, a 110-ms long sound was played from either the (56° to the) right speaker or from the (56° to the) left speaker, with equal probability. The prime remained visible on screen until either a response was given or 2000 ms had elapsed, whichever came first. Participants sat in a dimly lit, soundproof room, with their body midline 60 cm from the centre of the computer screen, and were reminded to constantly fixate the central fixation on the screen.[1] They were told that each trial would begin with the presentation of the agent trying to align its attention with either the left or right target location. Additionally, they were told that the agent's attempts were at chance level, since the target would appear with equal probability from the loudspeakers. Participants were required to press the response button as soon as they heard the sound propagating.

*2.1.5. Results*

Errors, like omissions, anticipations, and false alarms, accounting for 3, 8 of all trials, and too fast (RT < 100 ms) or too slow (RT > 800 ms) responses (1.4% of all trials) were removed from the analysis. The remaining data points, contributing to means for each participant were submitted to a repeated measure ANOVA, with Head Orientation and Gaze Type as within-subject factors. Although neither main effect reached the significance level (both $F$s < 1), Head Orientation and Gaze Type interacted with each other $F(1, 15)=6.4$, $p < .05$ (see lower panel in Fig. 1). This means that the response pattern in the eyes open condition [i.e., responses to 'head congruent target' trials (209 ms) being faster than responses to 'head incongruent target' trials (214 ms), $t(15) = 2.0$, $p = .056$] was significantly different from the response pattern in the eye closed condition [i.e., responses to 'head congruent' trials (219 ms) being slower than responses to 'head incongruent' trials (209 ms), $t(15) = 2.3$, $p < .05$]. No other factor, or combination of factors, reached the significance level.

*2.1.6. Discussion*

In line with our previous study (Nuku & Bekkering, 2008) showing that observing a laterally oriented head gazing toward a visual cue triggers the observer's attention, in this experiment we showed that observing the visual depiction of an agent with open eyes facilitates the detection of auditory targets gazed by agent's eyed side. Additionally we showed that cross-modal visuo-auditory orienting does occur in joint attention setting. Similarly to our earlier study, we also showed that observing laterally oriented head with closed eyes does not trigger the orienting of the observer's attention.

## 3. Experiment 2

In addition to demonstrating the existence of crossmodal orienting in a joint attention setting, Experiment 1 showed that when the agent's eyes were closed, the observer's responses were facilitated when the auditory sound propagated from the agent's ear side. Despite being a peculiar finding it confirms previous findings suggesting that visual cues other than human eyes alone do modulate the observer's attention (Tipples, 2002). Therefore, in the following experiment we investigated whether the presence of other face features, such as the agent's ear, might have caused the reversal in the cueing effect observed in Experiment 1. Thus, if the spatial contiguity between the ear and the sound were responsible for the faster reactions to auditory targets appearing at the ear side, by removing the agent's ear the 'ear-effect' would eventually disappear.

### 3.1. Methods

*3.1.1. Participants*

Fourteen participants (2 males; mean age 23 years), all students from the University of Nijmegen, received either course credits or five Euro for their participation at the experiment. All had normal (or corrected-to normal) vision and normal hearing, and were all right handed. Participants were naive as to the purpose of the experiment, and except for the missing ear (see upper panel in Fig. 2), this experiment was identical to the first experiment.

*3.1.2. Stimuli and procedure*

The rest of the experiment, including the procedure, the apparatuses and the materials used, was identical to Experiment 1.

*3.1.3. Results*

Errors as well as early and late responses accounted for 4.2% of all trials and were removed from the analysis. A repeated measures ANOVA with Head Orientation and Gaze Type as within-subjects factors, showed a significant effect for Head Orientation $F(1, 13) = 12.5$, $p < .05$, but no main effect for Gaze Type ($F < 1$, $p > .05$). Moreover, Head Orientation and Gaze Type did not interact ($F < 1$) indicating that the response pattern in the eyes open condition [i.e., responses to head congruent trials (272 ms) being faster than responses to head incongruent trials (281 ms), $t(13) = 3.20$, $p < .05$] was similar to the response pattern in the eyes closed condition [i.e., responses to head congruent trials (270 ms) were also faster than responses to head incongruent trials (280 ms), $t(13) = 2.75$, $p < .05$] (see lower panel in Fig. 2).

[1] Although the eye fixation was monitored by the experimenter only, we are confident that the participants complied with the instructions to fixate in the centre of the screen because in fast detection tasks like this one participants do spontaneously avoid making eye movements to target location (Posner, 1980). Furthermore, crossmodal capture of attention has been shown to occur even when controlling for gaze fixation (van der Lubbe & Postma, 2005). An eventual (if any) contribution of overt orienting in this study would not affect our hypothesis testing whether the observer's attention can be affected by auditory signals.

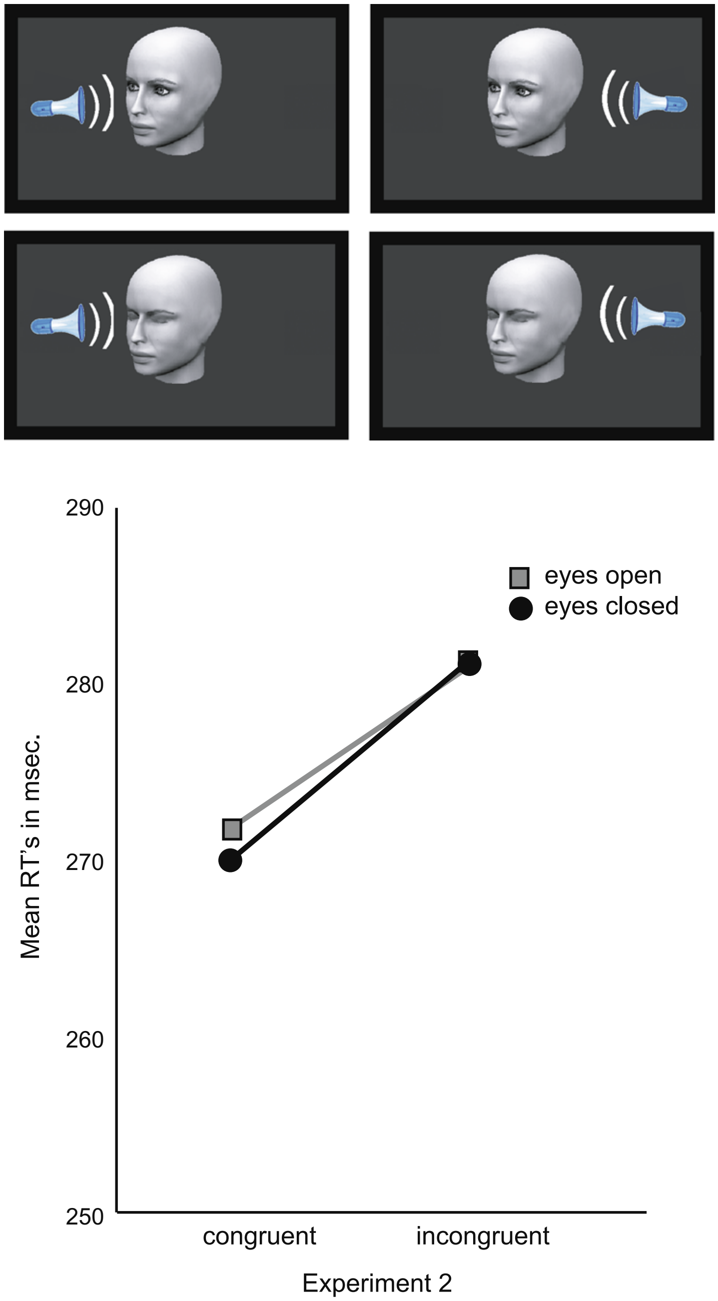


**Fig. 2.** Similarly to Experiment 1, each trial began with a fixation cross (2000-ms.) followed by the head cue (150-ms) and the auditory target randomly presented (110-ms) from either the left of the right speaker. In contrast to Experiment 1 the agent's ear here was not visible (upper panels) and the 'ear-effect' disappeared (lower panel).

*3.1.4. Discussion*

This experiment confirmed the earlier visuo-auditory effect showing that observing another's head oriented toward an auditory target location facilitates the detection of a gazed upon auditory target. However, it also showed that when the ear was removed, the facilitation occurred irrespectively of whether the agent's eyes were open or closed. In other words, by removing the ear we did remove the ear-effect seen in Experiment 1.

## 4. Experiment 3

From the second experiment one might conclude that the presence of the ear could have been responsible for the cueing pattern. One might argue that the difference between the first experiment, where the targets propagating from the ear side being facilitated, and the second experiment, where targets propagating form the missing ear side were not facilitated, might rely on the spatial ear/sound location contiguity. Such a processing argues that when the agent's eyes were open the eyes would be the main feature to trigger the observer's attention, whereas when they were closed the ear would have been the most prominent head feature. However, considering that human are able to infer other's intention, one could alternatively argue that the absence of the ear-effect in the second experiment might occur as an inferring process considering an agent without an ear as a non-hearing agent.

In order to disentangle between these two accounts we ran a third experiment where the ear would be absent, but in half of the trials is could inferred as present and functional and in the other half as present but not functional. More specifically, in one condition the agent's ear was concealed with a hat (through which one can still attend external sounds) while in another condition it was concealed with a helmet that normally is used to prevent one from attending to external sounds. We hypothesized that if seeing the agent's ear is crucial for shifting the observer's attention, neither the hat nor the helmet condition would affect the observer's alignment of attention. However, if inferring the agent's ability to auditory attends is all what observers pay attention at, only the agent with the hat would facilitate the observers' attention. Confirming this prediction would support the inferential account.

## 4.1. Methods

### 4.1.1. Participants

Sixteen participants (4 males; mean age 24 years), all students from the University of Nijmegen, received either course credits or five Euro for their participation at the experiment. All had normal (or corrected-to-normal) vision and hearing, were right handed, and naive as to the purpose of the experiment.

### 4.1.2. Stimuli, design and procedure

The apparatuses, the stimuli and the experimental design were identical to the previous experiments with exception to two changes. The visual primes were modified and depicted in half of the trials the agent's ear occluded by a hat and in the other half the agent's ear occluded by a helmet (see upper panel in Fig. 3). This change is reflected in the experimental design were the Head Cover factor (hat, helmet) was added to the existing Head Orientation (head oriented toward the speaker, head oriented away from the speaker) and Gaze Type (eyes open, eyes closed). The experiment was longer than the earlier two, consisting on 300 testing trials and 45 control (catch) trials, presented in a random order, but the rest of it, including the instruction to ignore the agent's head features and orientation and report the sound independently from its propagating location, were identical.

### 4.1.3. Results

Erroneous responses, accounting for less than the 2.0% of all trials, where excluded from the analysis. The remaining data points were submitted to repeated measures ANOVA, with Gaze Type Head Orientation, and Head Cover as within-subject factors. Although not significant, the analysis revealed a trend toward significance for both Gaze Type $F(1, 15) = 3.8$, $p > .05$ and Head Direction $F(1, 15) = 3.4$, $p > .05$, while the Head Cover effect $F(1, 15) = 1.3$, $p > .05$ was not significant. Importantly, Head Direction and Head Cover did interacted with each other, $F(1, 15) = 7.3$, $p < .05$, such that responses to head congruent targets were faster when the agent was wearing a hat (424 ms) than when wearing a helmet (430 ms).

In order to have a better insight into the data, we divided the three-way Head Orientation × Gaze Type × Head Cover interaction $F(1, 15) = 6.7$, $p < .05$ into Hat and Helmet data points. In the Hat condition, the ANOVA with Gaze Type and Head Orientation as within-subjects factors showed neither a Gaze Type nor a Head Orientation main effect (both $Fs < 1$). Importantly, however, the two factors interacted with each other $F(1, 15) = 14.2$, $p < .05$, indicating that when the agent's eyes were open, responses (420 ms) to head congruent targets were faster than the responses (426 ms) to head incongruent targets $t(15) = 2.8$, $p < .05$, whereas when the agent's eyes were closed, responses (432 ms) to head congruent trials were slower than responses (422 ms) to incongruent targets $t(15) = 2.8$, $p < .05$ (see lower panel in Fig. 3). In the Helmet condition, the Gaze Type and Head Orientation ANOVA showed a significant main effect for Head Orientation $F(1, 15) = 6.7$, $p < .05$, indicating that responses to head congruent targets, for eyes open (424 ms) and eyes closed (423 ms), were faster than responses to head incongruent targets, for eyes open (428 ms) and eyes closed (436 ms), respectively. The lack of a main effect for Gaze Type [$F(1, 15) = 3.1$, $p > .05$] suggested that the difference between the eyes open and eyes closed was overruled from the head direction, suggesting that subject's attention was triggered more by the head rather than the eye direction. No other factor, or combination of factors, reached the (alpha .05) significance level.

### 4.1.4. Discussion

In line with the two previous findings, the third experiment confirmed the existence of crossmodal orienting: observing a visual prime facilitates the detection of an auditory target. More specifically, showed that when the agent wearing the hat had the eyes closed the alignment of the observer's attention to the sounds appearing at the agent's ear side were faster than the alignment to sounds appearing at the eyes side. However, when the agent wore a helmet, the alignment of the observer's attention to sounds appearing at the eyes side was faster than the alignment to the sounds appearing at the ear side, independently from whether the agent's open or closed eyes. Considering that the agent's ear was in both conditions occluded, we exclude the possibility that this cross modal orienting derived from a perceptual (i.e., spatial ear-sound vicinity) pro-

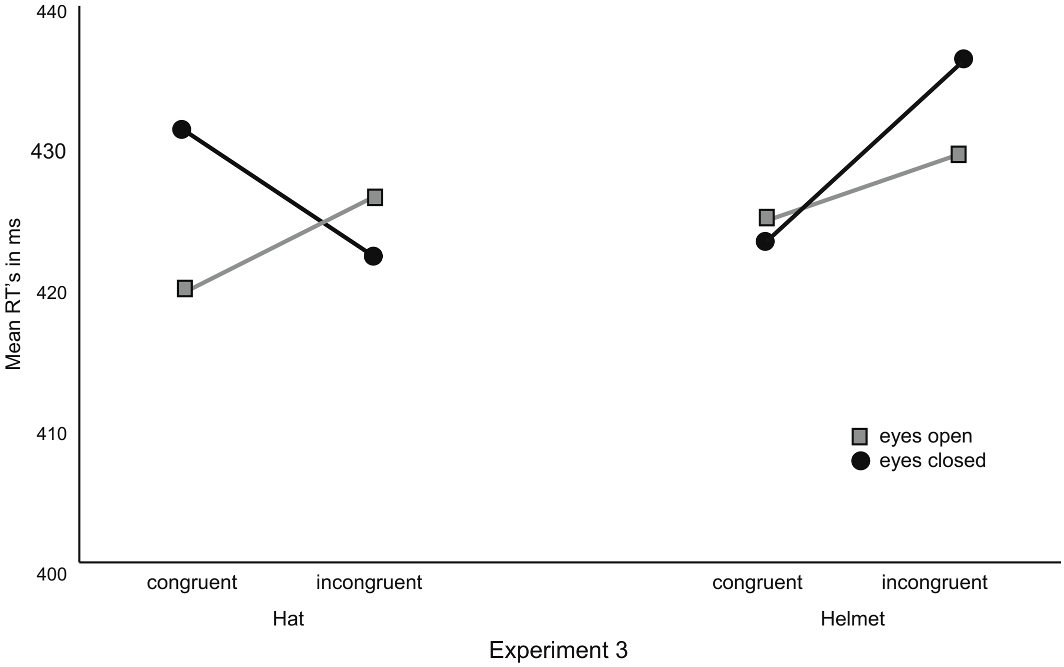


**Fig. 3.** The upper part of the figure illustrates the two types of Head Cover obstructing the agent's ear, the hat and the helmet. The lower part illustrates the findings of Experiment 3. In the hat condition, when the agent's eyes were open, the auditory targets appearing at the eye side were detected faster than the targets appearing at the ear side. When the agent's eyes were closed targets appearing at the eye side were detected slower than the targets appearing at the ear side. In contrast, in the helmet condition (right side) the targets were detected faster in the head congruent trials than in the incongruent ones.

cesses. Rather, we argue that inferring the agent as intending and being able to auditorily attend a certain sound location does facilitate the observer's attention.

## 5. General discussions

In this study, we found evidence supporting the visuo-auditory crossmodal cueing in joint attention. Observing a person gazing toward a certain location enhances the detection of an event occurring at that location even when the event is of auditory rather than visual modality. The first experiment showed that when the agent's eyes were open, the sounds propagating from the eyes side were detected faster than the sounds propagating from the ear side. In contrast, when the agent's eyes were closed, the sounds propagating from the ear side were detected faster than sounds propagating from the eyes, introducing the 'ear-effect'. In the second experiment, with the ear removed the 'ear-effect' disappeared, arguing in favor of a perceptual 'ear artifact' than 'ear-effect'. However, the last experiment showed that the ear-effect was not likely to arise from a spatial (ear-sound) processing of the scene. Here the effect was present only in the hat condition and it absent in the helmet condition, suggesting that participants selectively aligned their attention with the sound location they inferred as being attended by the agent.

The crossmodal effect in this study are in line with the crossmodal orienting literature showing that cues in one modality enhance stimuli in another modality (Bolognini, Frassinetti, Serino, & Làdavas, 2005; Mazza et al., 2006, for healthy subjects, and Lovelace, Stein, & Wallace, 2003; Pavani, Ladavas, & Driver, 2005 for neuropsychological patients). Although this study did not aim to test the time course properties of crossmodal orienting, we did comply with the rules typically associated with crossmodal orienting. According to the spatial rule (Stein & Meredith, 1993), only spatially coincident stimuli produce neuronal response enhancement in the neuron's excitatory area (Jay & Sparks, 1984; King & Hutchings, 1987). If, however, they do not overlap the visual and the auditory information won't be integrated, and the stimuli won't enhance each other. According to the temporal rule the maximal level of crossmodal enhancement is achieved when two stimuli are presented simultaneously or at very short intervals from each other (Mazza et al., 2006; McDonald et al., 2000) like in our study.

In line with our previous study (Nuku & Bekkering, 2008), we showed that observers do not strictly align their attention with another's face features (eyes or ears), but that rather they intrinsically infer whether the agent is visually or auditorily attending. The human ability to infer another's action intentions has been ascribed to the processing of a neural circuit, the mirror neuron system (Rizzolatti & Craighero, 2004), which is active both when one performs an action as well as when one observes another agent performing the same action. The circuit is activated when one plans and executes an action as well as when one observes the other agent performing it, thus in a sort of functional equivalence between observing and inferring actions (Decety & Grezes, 1999). This processing is best explicated by the simulation theory[2] (Gallese & Goldman, 1998), suggesting that humans understand other's actions by simulating the observed behavior via a direct matching process. The data of our first experiment speak in favor of the simulating account, because the observer's attention was selectively modulated when observing the attending agent with the eyes open, but not when observing the agent with the closed eyes. According to the simulation account, the observers recognized the action goals by directly mapping the observed behavior (gazing) onto a corresponding scheme in their own motor repertoire where the goal (attending) is already known.

The simulation theory relies in the assumption that the recognition of an action goal occurs by directly mapping the observed behavior onto a corresponding scheme in an existing motor repertoire. However, some events that need to be understood do not always rely on prior knowledge stored in one's motor repertoire. That is why it has been proposed that observing an event in an unusual or novel context is differently processed from observing an event in a stereotypic context. According to the mentalizing account, new events in a novel context are processed in brain areas that control the perception of social stimuli rather than in areas involved in direct matching (simulation) mechanisms (Decety & Grezes, 2006, see also Frith & Frith, 2006). Accordingly, action (intention) understanding relies on inferential processes of rationalization and mentalization (Gergely & Csibra, 2003). This alternative action understanding processes has been successfully demonstrated with infant (Gergely & Csibra, 2003). Our study is one of the first behavioral adult studies to demonstrate the same processing among adults. It showed (Experiment 2 and Experiment 3) that inferring the purpose of an unusual event in a novel context (observing an agent wearing a helmet or an agent without the ear) necessitates a great deal of active rationalization and inference to understand the agent's goal in relation to the novel context. Our finding supports an imaging study (Brass, Schmitt, Spengler, & Gergely, 2007) showing that inference-based processing of actions in novel uncommon situations does require rationalization.

In sum, our first experiment suggest that when faced with a plausible event, one's understanding about the events might occur as a consequence of simulation, or by directly mapping another's behavior onto one's own behavior repertoire. However, with the events in the last experiment being novel and uncommon, the observer's selective alignment of attention might have occurred as a consequence of rationalization rather than simulation. We believe that this conclusion needs more support before it can firmly stand alone. For example, imaging techniques showing the involvement of the mirror system (inferior frontal gyrus, inferior parietal cortex) in understanding the events occurring in a familiar context and the involvement of brain areas lacking mirror properties (superior temporal sulcus, the temporoparietal junction, the posterior cingulate cortex) in understanding novel event occurring in an uncommon context, might confirm our behavioral finding.

In conclusion this study shows that understanding a goal is just a step in the process of understanding another's the intention and that it may occur by either simulating or rationalizing (inferring) the other's intention. Furthermore we showed that visuo-auditory crossmodal cueing does occur in a joint attention context where one agent tries to understand the other's intentions. We showed that this specific crossmodal orienting does not occur as a consequence of low-level spatial processing, but that it rather relies on high-level cognitive processes, such as rationalizing and inferential mechanisms. Thus, humans simulate known event and rationalize about novel one, and these processes affect their attention.

## References

Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). Acoustical vision of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research, 160*, 273–282.

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: Inferential processes versus action simulation. *Current Biology, 17*, 2117–2121.

Carruthers, P., & Smith, P. (1996). *Theories of theories of mind*. New York: Cambridge University Press.

Decety, J., & Grezes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Science, 3*, 172–178.

Decety, J., & Grezes, J. (2006). The power of simulation: Imagining one's own and other's behavior. *Brain Research, 1079*, 4–14.

Downing, P., Dodds, C. M., & Bray, D. (2004). Why does the gaze of others direct visual attention? *Visual Cognition, 11*(1), 71–79.

Emery, N. J. (2000). The eyes have it: The neuroethnology, function and evolution of social gaze. *Neuroscience and Biobehavioural Reviews, 24*, 581–604.

[2] Although other theories, such as the 'Theory theory' (see Carruthers & Smith, 1996), the 'Interaction theory' (see Gallagher, 2001), the 'Associative learning theory' (see Schultz & Dickinson, 2000) do all regard the process of action-understanding or the understanding of other's action intentions, the simulation theory makes a clear prediction supported by neural functioning, the mirror neuron system.

Firestone, A., Turk-Browne, N. B., & Ryan, J.D. (2007). Age-related deficits in face recognition are related to underlying changes in scanning behavior. *Aging, Neuropsychology and Cognition*.
Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin and Review, 5*(3), 490–495.
Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology: Human Perception and Performance, 30*(2), 319–329.
Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*, 531–534.
Gallagher, S. (2001). The practice of mind theory: Simulation or primary interaction? *Journal of Consciousness Studies, 8*(5–7), 83–108.
Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 12*, 493–501.
Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences, 8*(9), 396–403.
Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Science, 7*, 287–292.
Gopher, D. (1973). Eye-movement patterns in selective listening tasks of focused attention. *Perception and Psychophysics, 14*, 259–264.
Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory and Cognition, 33*(1), 98–106.
Hietanen, J. K. (1999). Does your gaze direction and head orientation shift my visual attention? *NeuroReport, 10*(16), 3443–3447.
Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science, 9*, 131–134.
Jay, M. F., & Sparks, D. L. (1984). Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature, 309*, 345–347.
King, A. J., & Hutchings, M. E. (1987). Spatial response property of acoustically responsive neurons in the superior colliculus of the ferret: A map of auditory space. *Journal of Neurophysiology, 57*, 596–624.
Kobayashi, H., & Kohshima, S. (1997). Unique morphology of the human eye. *Nature, 387*, 767–768.
Langton, S. R. H., & Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition, 6*(5), 541–567.
Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences, 4*(2), 50–59.
Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: A psychophysical analysis of multisensory integration in stimulus detection. *Cognitive Brain Research, 17*, 447–453.
Mazza, V., Turatto, M., Rossi, M., & Umiltà, C. (2006). How automatic are audiovisual links in exogenous spatial attention? *Neuropsychologia, 45*(3), 514–522.
McDonald, J. J., Teder-Salerjarvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to a sound improves visual perception. *Nature, 407*, 906–908.
McDonald, J. J., & Ward, L. M. (2000). Involuntary listening aids seeing: Evidence from human electrophysiology. *Psychological Science, 11*, 167–171.
Nuku, P., & Bekkering, H. (2008). Joint attention: Inferring what others perceive and don't perceive. *Consciousness and Cognition., 17*(1), 339–349.
Pavani, F., Ladavas, E., & Driver, J. (2005). Gaze direction modulates auditory spatial deficits in stroke patients with neglect. *Cortex, 41*(2), 117–120.
Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*(A), 3–25.
Reisberg, D., Scheiber, R., & Potemken, L. (1981). Eye position and the control of auditory attention. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 318–323.
Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review Neuroscience, 27*, 169–192.
Rorden, C., & Driver, J. (1999). Does auditory attention shift in the direction of an upcoming saccade? *Neuropsychologia, 37*, 357–377.
Schmitt, M., Postma, A., & De Haan, E. (2000). Interaction between exogenous auditory and visual spatial attention. *Quarterly Journal of Experimental Psychology, 53*(A), 105–130.
Schultz, W., & Dickinson, A. (2000). Neural coding of prediction errors. *Annual Review of Neuroscience, 23*, 473–500.
Spence, C. L., & Driver, J. (1997). Audiovisual links in exogenous covert spatial attention. *Perception and Psychophysics, 59*, 1–22.
Spence, C., & Driver, J. (Eds.). (2004). *Crossmodal space and crossmodal attention*. Oxford: Oxford University Press.
Stein, B. E., & Meredith, M. A. (1993). *Merging of the senses*. Cambridge: MIT Press.
Tipples, J. (2002). Eye gaze is not unique: Automatic orienting in response to uninformative arrows. *Psychonomic Bulletin and Review, 9*, 314–318.
Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*, 675–691.
Van der Lubbe, R. H. J., & Postma, A. (2005). Interruption from irrelevant auditory and visual onsets even when attention is in a focused state. *Experimental Brain Research, 164*, 464–471.
Vecera, S. P., & Johnson, M. H. (1995). Gaze detection and the cortical processing of faces. *Visual Cognition, 2*, 59–87.
Wolters, N. C., & Schiano, D. J. (1989). On listening where we look: The fragility of a phenomenon. *Perception and Psychophysics, 45*(2), 184–186.