SPATIAL SOUND ASSISTANCE SYSTEM FOR 360 DEGREE HAZARD AWARENESS AND SAFE DRIVING

Masaki MARUYAMA
Junichi SAKAMOTO
Honda R&D Co. Ltd.
Japan

Paper Number 23-0111

ABSTRACT

In everyday driving situations, potential sources of collisions can appear from any direction around the driver. Driver assistance systems have been highly desired to assist driver's hazard awareness from all directions in order to eliminate any kinds of traffic accident fatalities. The current study addressed whether simulated spatial sounds providing directional and hazard attribute cues for potential collisions can facilitate drivers' identification of traffic hazards and reduce collision incidence in the front and rear spaces.

Forty-eight participants took part in our simulator experiment. We used a driving simulator (Honda Driving Simulator Type-DB Model S) to present them various traffic scenes with respect to the hazardous direction and recorded their driving operations. Participants' gaze directions were also recorded with an eye tracker implemented on the simulator. To provide a directional cue of hazardous traffic participants, we presented spatial sounds on the directions of hazard participants, using two speakers implemented in a driver's seat. To provide an attribute cue for hazardous objects, we classified the traffic participants into four categories (vehicles, motorcycles, bicycles, pedestrians) and presented a corresponding imitative sound for each hazard object. Presentations of monaural sound without directional cue and signal sound without attribute cue were also used as a comparison basis.

The current study observed a decrease in collision frequency and a significant reduction of onset time for pushing down the brake pedal for frontal hazard when spatialized signal sounds were presented compared with no HMI condition. A decrease in collision frequency with gazed hazards in the rear space was also observed when spatialized imitative and signal sound were presented relative to no HMI condition. The results lend to support our hypothesis that the directional cue can be effective for safer driving behaviors. On the other hand, improvements were not obtained when attribute cues were presented for both behavioral responses or the collision frequency. Significant facilitations were found in gaze responses and decelerate operations especially for rear hazards, but they did not result in a reduction of collision frequency.

Although the well-known front-rear ambiguity was confirmed in stationary sound localization, the current study observed the effectiveness of directional cue in reducing the collision frequency. It is possible that movements of spatial sound sources with hazard traffic participants could improve the resolution of front-rear sound localization. The influence of front-rear ambiguity might have also been reduced by extended spatial attention from the rear to

the front under the auditory directional cue towards the rear space. The attribute cue did not provide any effective improvements in the current study. However, we believe that in certain traffic situations where the type of hazards involved could represent more important information to the driver, the effects of attribute cue could reveal a potentially larger impact.

Our observations of the effective assistance of directional cue in spatial sound provide important references in terms of human factors for considering informative HMI that facilitates hazard awareness from all directions and help safer driving behaviors.

INTRODUCTION

In everyday driving, collision possibilities with hazardous traffic participants (e.g., pedestrians, cyclists, other vehicles) can occur from any direction around the driver. Hazardous traffic participants can thus appear outside the visual field of drivers, including behind them in the rear space. Recently, many of modern vehicles are equipped with sensor systems that possess capabilities to detect hazardous traffic participants all around the vehicle. There is also a growing interest for human machine interfaces (HMI) in driving assistant systems to provide hazard information in all direction around the drivers in order to support safer driving behaviors.

While simple alarms have often been used to inform hazard states or events, previous studies have addressed more informative cues in order to identify hazard sources. For example, a directional cue provided by a spatialized sound source location has been shown to facilitate our responses to visual targets within front space [1-4] and across front and rear spaces [5, 6]. These findings suggest that drivers can rely on the auditory spatial cue to direct their attention effectively towards the space where hazardous traffic participants appear, facilitating identifications of such potential hazards. However, such spatial auditory HMI has not been widely implemented in driver assistance systems in vehicles to date. The current study revisited the hypothesis on the effectivity of auditory directional cue, using a recent spatial sound technology which can produce sound images from any direction around drivers.

It is widely acknowledged that in everyday experiences, sound-producing events due to material interaction establish auditory informative cues for humans [7, 8]. Through common driving experiences, drivers are also able to use traffic environment sounds as auditory cues directly to distinguish between different types of potential hazards. For visual cues, selective attention based on object features such as size, color and shape have been well understood [9], however object-based auditory attention has not been well addressed, because of the lack of precise definition of object formation [10, 11] and, to our best knowledge, because of the lack of available realistic sound in experimental conditions. Using a recent interactive sound simulation technology, the current study created sounds that realistically imitated sounds emanating from actual traffic participants. We examined whether the auditory cues for selective object-based attention are effective in facilitating behavioral responses for safer driving behaviors and reducing collision possibility.

The objective of the present work is to reveal whether recent sound technologies can provide effective cues about location of potential hazard objects to drivers and enable safer driving behaviors. Considering that our findings on audiovisual attention directed in the front space cannot always be generalized to the rear space [5, 12], we evaluated the effectiveness of the cues regarding to front and rear hazards separately. Our findings provide important references in terms of human factors for considering informative HMI to assist raising awareness about potential hazards from all directions.

METHODS

Participants

Forty-eight participants (including 11 female participants) took part in this experiment. Their mean age was 53.4 years (range: 23 – 65 years). They had normal or corrected-to-normal eye sight. All participants had a valid Japanese driver's license. Self-reported frequency of driving was three days a week or less for every participant, except for two, who reported driving daily. This research received ethical approval from the Bioethics Committee of Honda R&D Co., Ltd. All participants gave their written informed consent prior to the start of the experiment. They received 14,500 JPY for taking part in this experiment, including travelling expenses.

Apparatus

An overview of the driving simulator used in this experiment (Honda Driving Simulator Type-DB Model S) is illustrated in Figure 1. The participants drove an automatic transmission vehicle using accelerator and brake pedals and a steering wheel. A gear shift and turn indicators were also available. Three flat-screen displays provided the outside view of the driver's vehicle from the viewpoint of a typical driver's position, i.e. at a 104.8 cm distance from the central display. The scope of view was 150 deg horizontal and 30 deg vertical. Images of side mirrors (50 deg horizontal visual field) and the rear view mirrors (30 deg) were shown on the displays.

To output spatialized sounds, we used a driver's seat with two speakers installed to the left and right side of the head at shoulders height location. Using interactive audio middleware software, the Vector Based Amplitude Panning sound spatialization algorithm [13] was used to control in real-time the relative level of sounds for the left and right speakers. A filter for non-individualized head related transfer function (HRTF) was also applied to implement binaural processing to generate a sense of sound source location that can distinguish the front and rear spaces, as well as distinguish the leftward and rightward directions. Monaural sounds were provided from a single speaker (JM10 pro., Conisis) placed at the center of the front panel.

To create imitation sounds, we first classified traffic participants that could cause hazards into four categories: (1) vehicles, (2) motorcycles, (3) bicycles and (4) pedestrians. Samples recorded sounds were acquired for each category. Using a layered sound approach mixed with interactive audio middleware, the samples of vehicles were simulated by combining engine sound, road noise, and wind noise. The simulated vehicle sound playback precisely reflected the effects that engine rotation speed and engine load have on the vehicle engine sounds, and the effects of the vehicle's speed on the road and wind noises, which made it possible to create imitation sounds that realistically represent the sounds of actual vehicle in the simulation. This was also the case with motorcycles. Additionally to the sound spatialization, a level attenuation effect as a function of relative distance was also applied to the simulated vehicle sounds.

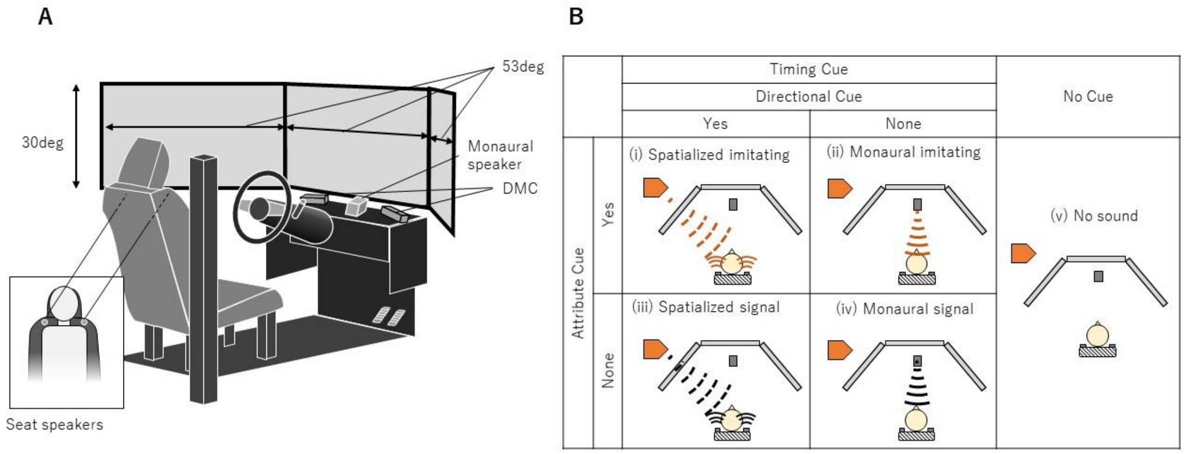

Figure. 1 A. An overview of the driving simulator used for the sound HMI experiment. The experiment participants drove an automatic transmission vehicle viewing the outside visual scenes presented on three flat screen displays. Spatial and monaural sounds were output with the seat speakers and monaural speaker respectively. Driver monitoring cameras (DMCs) were used to record participants' gaze direction. B. Summary of sound HMI conditions and cue factors considered in the present study. Presentation of spatialized sound from hazard object provided a directional cue, whereas monaural sound presented from front regardless of direction of hazard object gave no directional cue. To provide an attribute cue, traffic participants were classified into four categories (vehicles, motorcycles, bicycles and pedestrians), and an imitative sound for the hazard object category was presented. In contrast, a common signal sound was presented for all four hazard objects categories in order to evaluate sound presentation condition without any attribute cue. While the sound presentation conditions (i, ii, iii and iv) provide a timing cue of realized hazard event, the no sound HMI condition did not give any cue.

The signal sound consisted of impulsive synthetic tones of individual duration that were repeated at a rate of 0.3 seconds. The same sound was used for the different categories of traffic participants, and thus it did not provide any cues as to the attributions of hazard traffic participants. Attenuation as a function of relative distance was not applied to the signal sound.

HMI presentation

There were five variations of HMI presentation (Fig. 1B): (i) spatialized imitating sound, (ii) spatialized signal sound, (iii) monaural imitating sound, (iv) monaural signal sound and (v) no sound. A directional cue was presented in (i) and (ii) but not in (iii) and (iv). Therefore, differences in driving behaviors between (i) and (iii) and between (ii) and (iv) indicate an effect of multiple factors including directional cue. On the other hand, since an attributional cue was given in (i) and (iii) and not in (ii) and (iv), differences between (i) and (ii) and between (iii) and (iv) reveal an effect of multiple factors including the attribute cue. Whereas the conditions (i) to (iv) gave a timing cue of realized hazard to the experiment participants which was not the case in (v). Therefore, comparing the driving behaviors seen in (i) to (iv) with those in (v), we examined the effect of hazard presence cue on driving safety.

In order to assist attention allocation to hazardous traffic participants, HMI requires a distinction between hazardous and non-hazardous objects. While the development of discrimination rules in real traffic have been difficult issues, the predefined scenarios of the simulator make it possible to identify traffic participants that could become hazardous, based on the driving behavior of own vehicle. We identified the potential hazardous participants in the scenarios before conducting the experiment. To detect a transition from potential hazard to realized hazard, we used the centripetal time to collision (centripetal TTC; Fig. 2A), which extends the well-known TTC from one dimension to two dimensions. This was obtained by dividing a relative distance from the own vehicle to the object by a relative velocity component in the direction from the own vehicle to the object. To define the threshold of centripetal TTC, we conducted a preliminary test of subjective discrimination while driving the hazard prediction courses. We found that the hazard sense was usually evoked when centripetal TTC became less than 5 sec. Here, the TTC is known not to be an effective indicator when the hazard is evoked by uncertainty in behaviors of objects that are at a close distance, at a low relative velocity. It is also the case with centripetal TTC. Indeed, we found the transition often occurred when the relative distances of potential hazard objects are shorter than 5 m. Putting our preliminary findings together, we deemed hazard as realized when the centripetal TTC is shorter than 5 sec or the relative distance is shorter than 5 m. The simulator calculated the relative distance and centripetal TTC at a frequency of 100 Hz. The sound HMI was presented in the conditions (i) to (iv) when this criterion was satisfied.

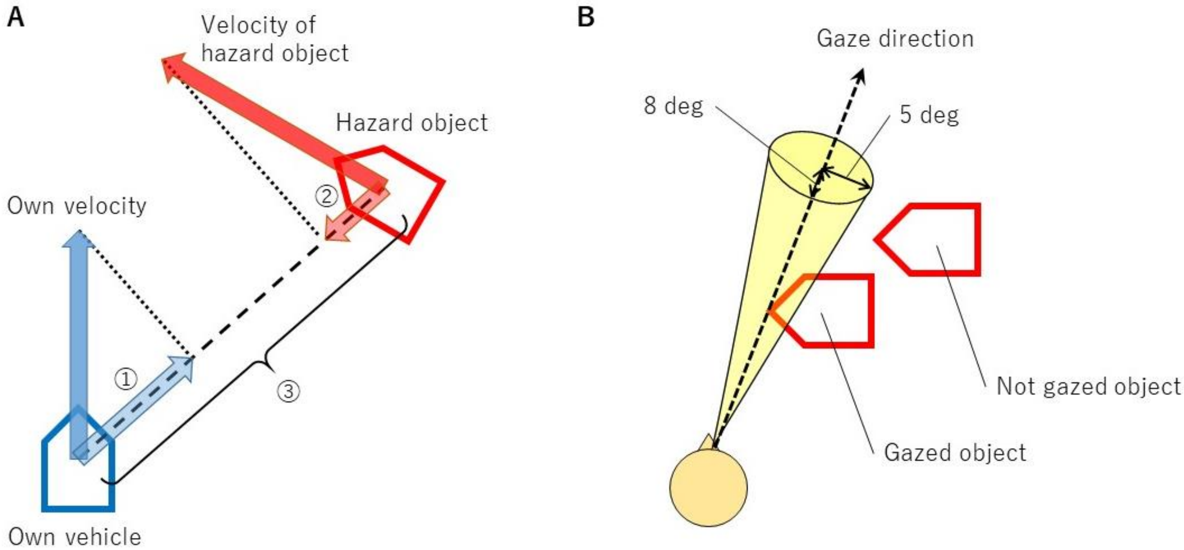

Figure. 2 A. Illustration of simulated parameters used for the estimation of time to collision (TTC) in a centripetal form, that is, relative distance to a traffic participant (①) was divided by approaching speed toward the traffic participant (②+③). B. Diagram explaining the estimation of effective visual field for spatial perception in the current study. The effective visual field is represented with a virtual cone oriented towards the gaze direction. We estimated that the experiment participants obtained spatial perception of realized hazard objects when the objects and cone overlapped for longer than 200 ms within a time window of 250 ms, which we term "gaze" in this study.

Procedure

The experiment started with a measurement of the directional accuracy of sound images. We presented the spatialized sounds of the vehicle, bicycle, pedestrian and signal from the directions of front, back, left, right and four diagonals, sequentially and in a random order. Note that the location of spatialized sound was fixed for each presentation. The time duration of each presentation was 1 sec. After every presentation, the participants indicated the direction of the sound image by marking the direction on an egocentric coordinate figure illustrated on their response sheet.

Next, the participants drove a driving course preinstalled in the simulator to familiarize themselves and verify the sensitivities of steering and pedaling and the size of their own vehicle. They also drove the 4th of six hazard prediction training courses preinstalled on the simulator, in order to familiarize themselves with inner-city driving on the simulator.

We then recorded the driving behaviors (gaze directions, accel and brake pedal operations, vehicle movements) as they drove the remaining five courses of hazard prediction training. The driving of five courses were separated by short breaks. The order of the courses (1st, 2nd, 3rd, 5th, 6th) was same for all participants, while the five Human-Machine Interface (HMI) conditions (spatialized imitating, spatialized signal, monaural imitating, monaural signal, no HMI) were given in a random order. We stopped the experiment if participants showed signs of simulator sickness

Recordings and analyses

To evaluate the effectiveness of HMI presentation on driving operations in the experiment participants, we recorded their operations of vehicle accelerator and brake pedals. We also recorded simulations of positions and velocities of their own vehicle and the twenty closest traffic participants, as well as the centripetal TTC and events of collision. The recording frequency was 100 Hz. To indicate a degree of safe driving with respect to the realized hazard traffic participants, the current study used latencies of deceleration operations from the time of hazard realization. Here, an onset of brake-pedal and offset of accel pedal after the beginning of hazard realization were respectively extracted as the latencies of deceleration operation.

The effectiveness of additional sound HMI on their attention allocation to realized hazard participants was analyzed, using the gaze direction recorded at a frequency of 20 Hz with a camera-based driver monitoring system developed by Seeingmachines company. To represent an effective visual field, we set a virtual cone towards the gaze direction (Fig. 2B). We estimated that they became aware of realized hazard objects that were overlapped with this cone. A previous study indicated that an effective visual field size of spatial perception is larger than 5 deg, when a visual target stimulus was presented for 250 ms on a dynamical background simulating a driving situation [14]. Hence, we set the horizontal radius of cone at 5 deg. Taking into account a large error of gaze recording in vertical direction, we set the vertical radius of cone at 8 deg. We deemed that they became aware of realized hazard objects when the objects and cone overlapped for longer than 200 ms, within a time window of 250 ms. In addition to the latencies of vehicle operations and gazing, the frequencies of collision were compared among the HMI conditions to show the assistance effect of HMI presentation.

In statistical analyses, we tested significance of differences in the collision frequency amongst the HMI conditions using a z-test. For the difference in the latencies of gazes and deceleration operations, we conducted a Wilcoxon signed-rank test. The current study reports uncorrected p-values in multiple comparisons.

RESULTS

Collision and hazard gaze frequencies

Collision frequency with traffic participants to the front and rear of own vehicle at the time of hazard realization have been summarized for each HMI condition in Fig. 3A and B. In the case of the front hazard, a decrease in collision frequency was observed in the conditions of spatial signal sound HMI (6.9 %) relative to the no HMI condition (11.1 %). On the other hand, in the case of the rear hazard, an increase in collision frequency was observed in the presentation of monaural signal sound (10.5 %) relative to the no HMI condition (4.9 %), whereas such increase was not found for any other conditions of sound HMI presentation (5.6 – 6.8 %).

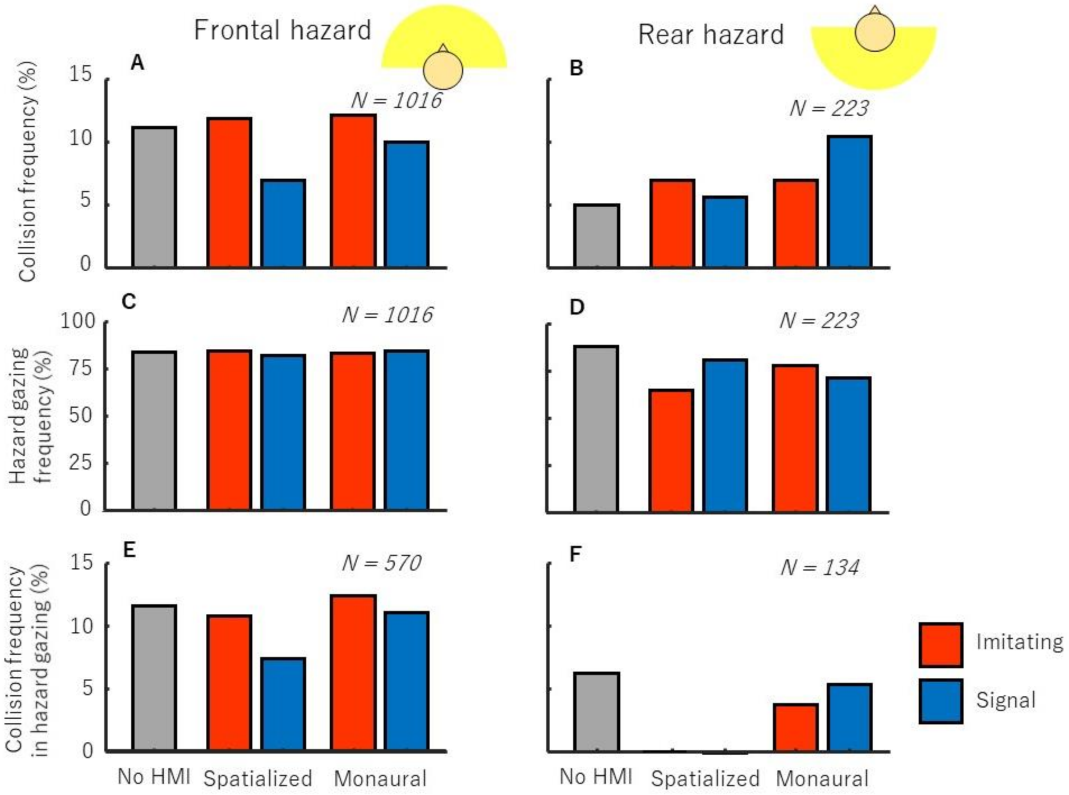

Figure. 3 A. Collision frequency with traffic participants to the front of own vehicle at the time of hazard realization. B. Collision frequency with traffic participants in the rear space. C, D. Gaze frequency of hazard traffic participants in the front and rear spaces. E, F. Collision frequency with gazed hazardous participants in the front and rear spaces, respectively. The results in each scene and participant were merged, and then the frequency was obtained in each HMI condition. The sample number is indicated on the top right in each panel. Note the differences in the frequency among the HMI conditions did not reach to a significance level in our statistical test (p > 0.1).

To find whether there was an assistance effect that helped prevent drivers from missing hazard objects, the gaze frequencies of hazard object in the front and rear spaces is illustrated in Fig. 3C and D. The frequency of front hazard gaze ranged from 82 % to 85 % in every condition of sound HMI presentation, and no significant difference was obtained relative to the no HMI condition (85 %). In the case of rear hazard, the gaze frequency in the condition of spatialized imitating sound HMI (64 %) was lower relative to the condition of no HMI (89 %) and the other conditions of sound HMI presentations (Spatialized signal: 81 %; Monaural imitating: 77 %; Monaural signal: 71 %), though the differences were not significant ($p > 0.1$).

To indicate whether the spatial perception of hazard objects with gazing was effective, the collision frequency in the condition of hazard gaze was compared under different HMI conditions. In the case of front hazard, the collision frequency obtained with spatialized signal sound (7.3%) was lower relative to the no HMI condition (11 %), whereas the other sound HMIs (11 – 12 %) did not show any significant improvement with respect to the no HMI condition (Fig. 3E). In the rear hazard scenario, no collision was observed in the spatialized sound HMIs, whereas it appeared at a frequency of 6.3 % in the condition of no HMI (Fig. 3F).

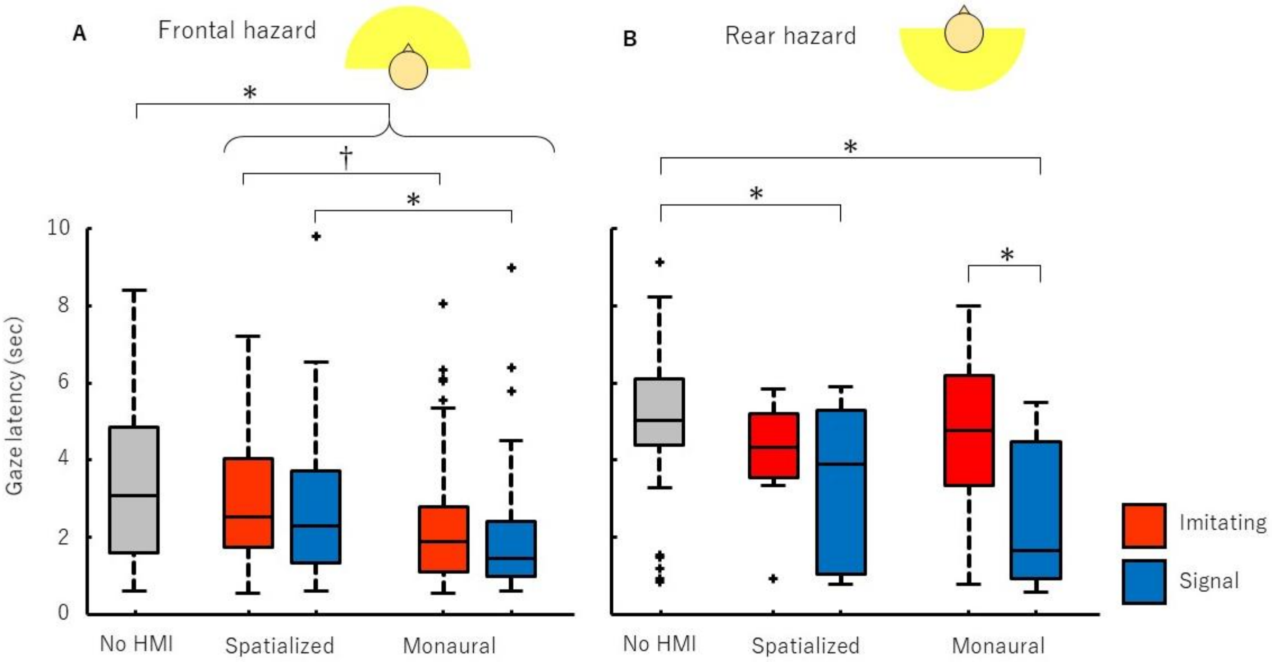

Figure. 4 A. Latencies of front hazard gaze from the hazard realization (i.e., the time when the centripetal TTC falls below 5 sec or the relative distance is shorter than 5 m.) B. Latencies of rear hazard gaze from the hazard realization. Box plots are used to indicate medians, quartiles and range of latency distributions in each HMI condition. The results in each scene and participant were merged in each distribution. Significant and marginal difference in the median between HMI conditions are marked by an asterisk (*, $p < 0.05$) and dagger (†, $p < 0.1$), respectively.

Hazard gaze latencies

Figure 4A summarizes the latencies of frontal hazard gaze in the form of boxplots. We found significantly shorter medians of latencies in every condition of sound presentation (Spatialized imitating: 2.5 s; Spatialized signal: 2.3

s; Monaural imitating: 1.9 s; Monaural signal: 1.5 s), relative to the condition of no sound presentation (3.5 s). Significantly shorter medians of latencies were also found with the signal sound presentation compared with the imitating sound presentation in both spatialized and monaural conditions. In the cases of rear hazard, the spatialized and monaural signal sounds significantly decreased the gaze latency relative to the condition of no sound presentation (Fig. 4B; Spatialized signal: 3.9 s; Monaural signal: 1.6 s; No HMI: 5.0 s; $p < 0.05$). We also found a significantly shorter median of latency with the signal sound relative to the imitating sound in the monaural condition ($p < 0.05$). No significant difference was found between the conditions of no sound presentation and 3D and monaural imitating sound presentations ($p > 0.1$).

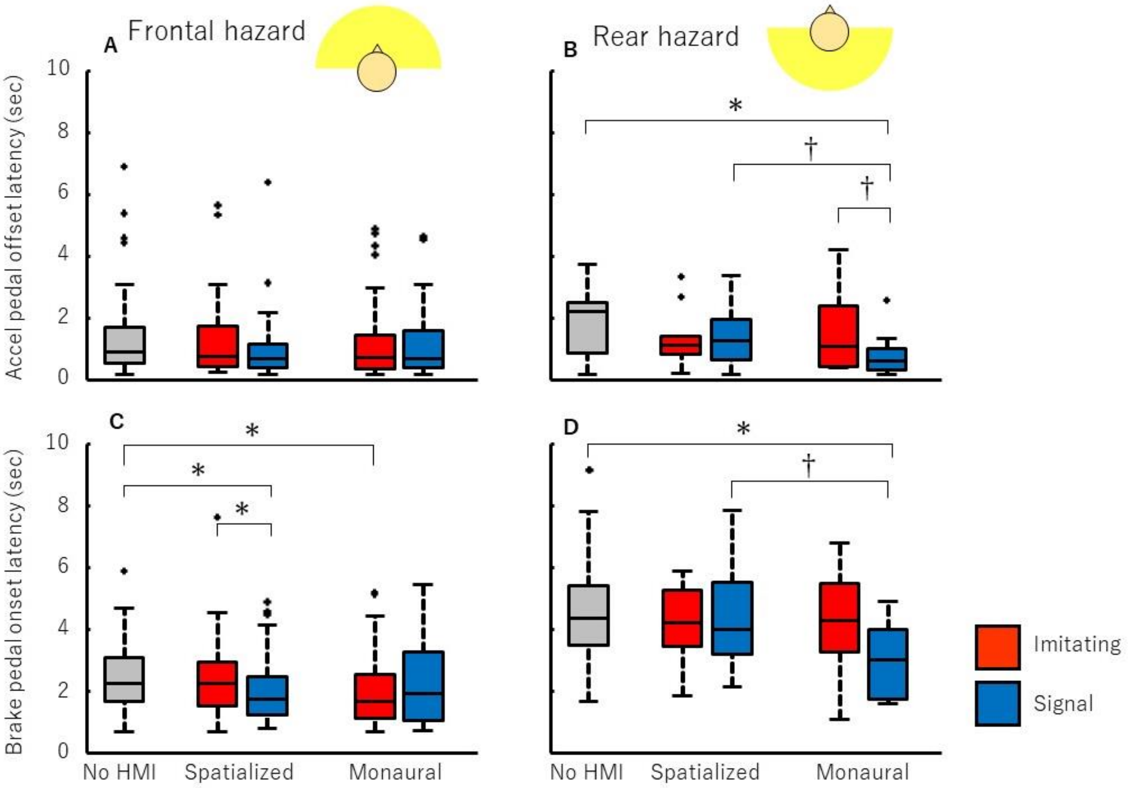

Figure. 5 A, B. Latencies of accel pedal offset from the hazard realization in the front and rear spaces, respectively. C, D. Latencies of brake pedal onset from the hazard realization in the front and rear spaces, respectively. Box plots are used to indicate medians, quartiles and range of latency distributions in each HMI condition. The results in each scene and participant were merged in each distribution. Significant and marginal difference for the median between HMI conditions are marked by an asterisk (*, $p < 0.05$) and dagger (†, $p < 0.1$), respectively.

Decelerate operation latencies

The offset latencies of acceleration pedaling relative to the front hazard realizations were summarized in Fig. 5A. No significant difference in their median was found among the HMI conditions ($p > 0.1$). When the hazard traffic participants were in the rear space (Fig. 5B), a significant decrease in the latency of acceleration pedaling was obtained with the presentation of monaural signal sound (0.6 s), relative to the no HMI condition (2.2 s; $p < 0.01$). A marginal decrease was also observed relative to the conditions of spatialized signal HMI (1.3 s; $p < 0.1$) and monaural imitating HMI (1.1 s; $p < 0.1$).

As to the onset latencies of pushing down the brake pedal, a significantly shorter median of latency was obtained in the condition of spatialized signal sound HMI (1.9 s) compared with the conditions of no HMI (2.2 s; $p < 0.01$) and spatialized imitating sound HMI (2.2 s; $p < 0.05$) in the front hazard (Fig. 5C). We also found a significantly shorter median of latency in the condition of monaural sound HMI relative (1.6 s) to the condition of no HMI ($p < 0.05$). When the hazard traffic participants were in the rear space (Fig. 5D), a decrease observed in the presentation of monaural signal sound (3.0 s) relative to the condition of no HMI was significant (4.4 s; $p < 0.05$), and that the effect relative to the condition of spatialized signal sound was marginal (4.0 s, $p < 0.1$).

Directional accuracy of 3D sound image

To estimate the sound localization accuracy of spatialized sound HMI, we used the directional responses acquired for every participant at the beginning of the experiment. We took an absolute value of angular difference between the presented spatial sound and the directional response reported by the participants. The errors of directional responses obtained from each participant were summarized for vehicle, bicycle, pedestrian, and signal sounds in Fig. 6. The median errors obtained for all participants were below the chance level of 90 deg, except for the directly in front and rear directions, where the sound image was often perceived as located in the opposite direction. The overall median across all the participants, indicated a localization accuracy of 30 deg.

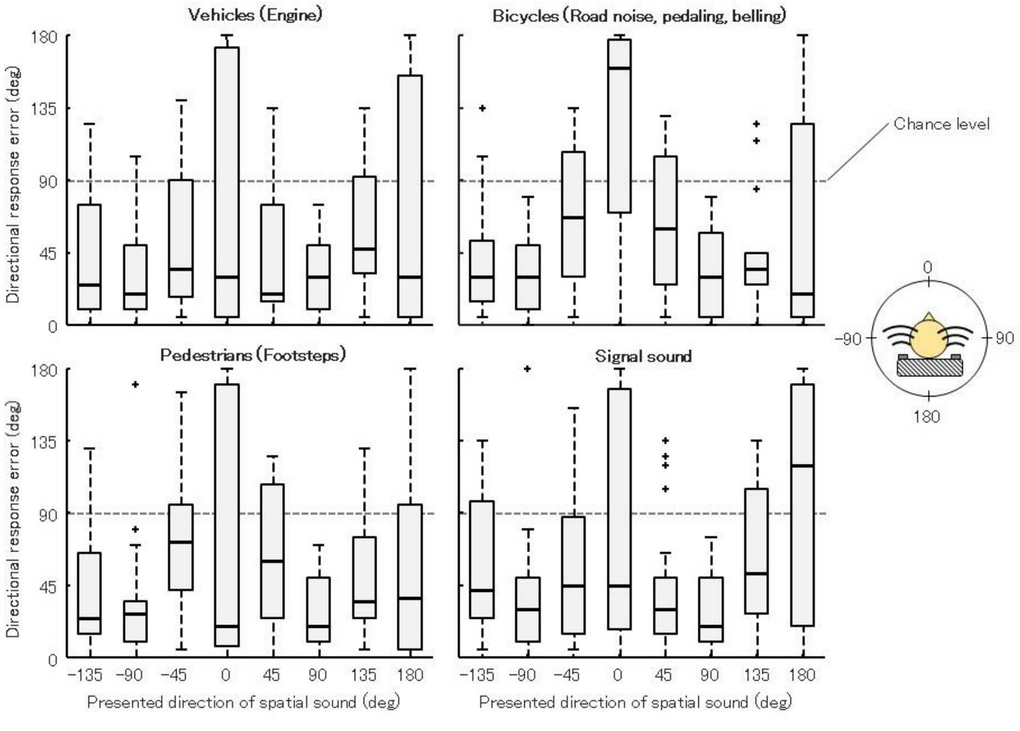

Figure. 6 Distributions of localization errors of spatial sounds presented on the directions of front, back, left, right and four diagonals. Imitative sounds of vehicle, bicycle and pedestrian and signal sound were used in the measurement of localization error. An absolute value of angular difference was taken between the presented spatial sound and the directional response reported by the participants. The chance level indicates the error size that we would obtain for random responses.

DISCUSSIONS

Timing cue assists early hazard gaze

Significantly earlier gazes of front hazard were found with every condition of sound HMI presentation compared with the no sound HMI condition (Fig. 4A). Our results confirmed that the presentation of an auditory cue on the occurrence of a visual event can facilitate earlier behavioral responses without providing any spatial or attribute cue, as indicated in the previous studies [15-17]. One might think that a cue of hazard presence, rather than a timing cue of transition event from potential hazard to realized hazard, decreased the gaze latency. If it were true, the experiment participants could have been aware of hazard presence more frequently, which could have been observed as an increase of hazard gaze frequency under the conditions of sound HMI presentation relative to the condition of no sound HMI. However, the little increase we observed (Fig. 3C) does not support this hypothesis. Our results suggest that the early cue of transition from potential hazard to realized hazard decreased the latency of gaze under the conditions of sound HMI presentations.

In the rear hazard situation, a significant decrease of gaze latency was obtained with the presentation of signal sound HMI compared with the no HMI condition (Fig, 4B), which could be related to the effect of the timing cue. On the other hand, no significant decrease of gaze latency was found with the presentation of imitation sound. In this case, because of the attenuation effect on the volume of imitation sound as a function of relative distance, an auditory awareness of the gradual onset of imitation sound might be delayed compared with the signal sound that was free from the attenuation effect. In addition, a visual awareness of the transitions from potential hazard to realized hazard in the rear space would tend to be delayed with respect to those in the front visual field while driving forward. For these reasons, no significant decrease was found in the current study with the presentation of imitation sounds for rear hazard situations.

Directional cue assists a decrease of collision frequency

A decrease of collision frequency with the frontal hazard was observed in the presentation of spatialized signal sound compared with the no HMI condition (Fig. 3A and E), which could be related to the significantly early onset of pushing down the brake pedal in the spatialized signal sound presentation relative to the no HMI condition (Fig. 5C). It was not the case where the monaural signal sound was presented. Taken together, the current study indicated that a directional cue is effective in the situation of front hazard to facilitate the earlier onset of pushing down the brake pedal, resulting in the decrease of collision frequency. The results are consistent with the previous studies in which a presence of spatial auditory cue reduced the time for visual search [2] and acceleration or braking response in driving hazard avoidance [5].

The current study also found that the latency of brake pedaling onset was significantly shorter under the condition of spatialized signal sound compared with the spatialized imitation sound. Given that the directional cue was already enough to identify a hazard in the front space, an intense sense of hazard under the signal sound would be more effective to decrease the latency of brake pedaling relative to the spatialized imitation sound attribute cue. The assumption was supported by our finding of significantly early onset of pushing down the brake pedal for the monaural imitation sound (Fig. 5C), in which the attribute cue would assist an early identification of hazard for participants.

No collisions were observed when the rear hazard was gazed for the presentations of spatialized sound HMI, whereas it was not the case with either the presentation of monaural sound HMI or no sound HMI presentation (Fig. 3F). This tendency was not observed when the collision frequency includes the cases where the realized hazard object was within the effective visual field of spatial perception [14] (Fig. 3B). In other words, neither presence perception of realized hazard object, which would be available within an effective field extending 15deg [14, 18], nor spatial perception of potential hazard object would be effective to use the directional cue. Taken together, the auditory directional cue was informative when the experiment participants obtained the visual spatial perception of hazardous object in the rear space, which was effective to reduce the collision frequency.

The previous psychophysical studies obtained a facilitation effect on earlier behavioral responses as visual targets moved towards the peripheral visual field [19] and even towards the rear space [20]. It should be also noteworthy that rapid responses were previously obtained with a presentation of close rear auditory warning signal relative to a far front auditory warning signal [21]. Taking into account that the seat speakers were close to the ears and were in the rear space, the presentation of spatial sound from the seat speakers could result in a better performance relative to the monaural sound from the front speaker. Although we did not obtain the rapid responses in both of eye movements and pedaling, a better performance was found in the form of low frequency of rear hazard collision under the presentation of spatial sound compared with the monaural sound.

Regarding the sound localization accuracy of the directional cue, front-rear ambiguity is known to occur in localization of sound source in humans. Although we can use monaural spectral cues, originating from the spectral filtering of sound by the pinnae, head, and torso in order to identify whether sound sources are in front or rear, the ambiguity remains for both actual and spatialized sound sources [22-24]. While the spectral filtering is different among individuals because of the individual anatomical variance of pinnae, head and torso, the current study used the same spectral filtering across the experiment participants (i.e., non-individualized HRTF). For these reasons, the relatively large front-rear confusion in localizing sound image occurred in our presentation of spatial sound (Fig. 6). Nevertheless, the current study observed the tendencies of effectiveness for the directional cue.

Recall that the spatial sound source from the hazardous traffic participants moved relative to the own vehicle. Considering that the resolution of front-back ambiguity in the localization of a sound source can be improved by the movement of sound source [25], the perceived directional cue in the experiment participants might have been more accurate than those we measured for the stationary directional sounds.

It should be also noteworthy that there might be neurophysiological mechanism that makes the front-rear ambiguity less serious. Receptive field (RF) of audiovisual neurons in the superior colliculus (SC), known to be involved in allocating spatial attention [26-28], can show different spatial properties in the RF extent depending on their locations. That is, the neurons in the rostral portion of the SC, responsive to stimuli presented from frontal space, have often showed visual and auditory RFs less than 10 deg and 20 deg respectively in diameter, whereas those in the caudal SC that respond to stimuli presented from the periphery have visual RFs ranging from 40 to 100 deg and auditory RFs from 60 to 135 deg in diameter [12, 19, 29]. Thus, an auditory directional cue presented in the rear-left spatial area, for instance, is possibly captured within the RF of an audiovisual neuron that has a visual RF responsive to visual stimuli in the front-left spatial area. Although the current study found the relatively large front-rear confusion in localizing sound image, the auditory directional cue towards a direction in the rear might be effective to allocate spatial attention extending from the direction to the front left, which would assist the hazard awareness not only in the lateral rear but also in the lateral front on the same side.

Early gaze and deceleration with monaural signal sound presentation for rear hazard

A significantly shorter latency of rear hazard gaze was found with the presentation of monaural signal sound compared with the no HMI condition (Fig. 4B). It was also true for the onsets to release acceleration pedal and to press brake pedal (Fig. 5B and D). While the early latencies could give a little extra time to collision, the experiment participants were required by themselves to identify the traffic participant causing the hazard status, because the monaural signal sound conveyed no spatial nor object cue. Therefore, it would be possible that the early gaze did not assist to decrease the collision frequency. The early responses of deceleration would not always help drivers from preventing the collision, since the hazard participants were in the rear space. Indeed, the current study did not observe any significant decrease of collision frequency when the rear hazard was gazed under the presentation of monaural signal sound HMI (Fig. 3F). Rather, an increase of the collision frequency was observed under the presentation of monaural signal sound compared with the other HMI conditions (Fig. 3B). Our results suggest that the signal sound could decrease the latency of gaze and deceleration, but it could not always assist collision avoidance because of the lack of directional information on hazard participant.

Early gaze of frontal hazard with monaural sound from the front

The source of monaural sound in the current study was always at the front, regardless of the directions of hazard participants. Thus, our monaural sound could not exactly provide a cue on the direction of hazard traffic participants. However, under the condition of high frequency of hazard events in the front hemifield compared with the rear hemifield, the monaural sound from the front location would be beneficial to allocate their attention on the front direction. If it was true, an effect of the valid cue should have been found as early responses, as shown in the previous studies [5, 15, 30]. Indeed, we found a significantly early gaze of hazard traffic participant under the condition of monaural sound presentation relative to the no HMI condition, when the hazard traffic participants were in the front (Fig. 4A). Overall, although the monaural sound was not informative on the hazard direction, its source location in the front space could be effective to facilitate the front hazard gaze.

CONCLUSIONS

Our observations of the effective assistance of directional cue in spatial sound provide important references in terms of human factors for considering informative HMI that facilitates hazard awareness from all directions and help safer driving behaviors.

Acknowledgements

We thank Mr. N Fujimoto for fruitful advices on the present works, Mr. K Nagura and H Ono and Managebusiness Co.,Ltd for functional extensions of driving simulator. We are grateful to Mr. T Ezaki, Y Shimura and Auto Technic Japan for their helps in acquiring data, Mr Nakayama for his helps on data processing.

References

[1] Perrott, D. R., Sadralodabai, T., Saberi, K., & Strybel, T. Z. (1991). Aurally aided visual search in the central visual field: Effects of visual load and visual enhancement of the target. Human Factors, 33, 389-400.

[2] Begault, D. (1993). Head-Up Auditory Displays for Traffic Collision Avoidance System Advisories: A Preliminary Investigation. Human factors.

[3] Corneil, B. D., van Wanrooij, M., Munoz, D. P., & van Opstal, A. J. (2002). Auditory-visual interactions subserving goal-directed saccades in a complex scene. Journal of Neurophysiology, 88, 438-454.

[4] Hancock, P. A., Mercado, J. E., Merlo, J., & Van Erp, J. B F. (2013). Improving target detection in visual search through the augmenting multi-sensory cues. Ergonomics, 56, 729-738.

[5] Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. Journal of Experimental Psychology: Applied, 11, 157-174.

[6] Lee, J., & Spence, C. (2015). Audiovisual crossmodal cuing effects in front and rear space. Frontiers in Psychology, 6:1-10.

[7] Gaver, W. W. (1993a). What in the world we hear? An ecological approach to auditory event perception. Ecological Psychology, 5, 1-29.

[8] Gaver, W. W. (1993b). How do we hear in the world? Explorations in ecological acoustics. Ecological Psychology, 5, 285-313.

[9] Desimone, R. and Duncan, J. (1995) Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. 18, 193–222.

[10] Bizley JK, Maddox RK, Lee AK (2016) Defining auditory-visual objects: behavioral tests and physiological mechanisms. Trends Neurosci 39:74–85.

[11] Shinn-Cunningham, B. G. (2008) "Object-based auditory and visual attention," Trends Cogn. Sci., 12, 182–186.

[12] Spence C., Lee J., Van der Stoep N. (2020). Responding to sounds from unseen locations: Crossmodal attentional orienting in response to sounds presented from the rear. Eur. J. Neurosci. 51:1137–1150.

[13] Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. Journal of the Audio Engineering Society. 45. 456-466.

[14] Ayama, M., Mekada, Y., Kodama, D., & Kasuga, M. (2005). Effects of visual and conversation loads on functional visual field on a dynamic background. J. Illum. Engng. Inst. Jpn. 89, 453-471.

[15] Diederich, A., & Colonius, H. (2008). Crossmodal interaction in saccadic reaction time: Separating multisensory from warning effects in the time window of integration model. Experimental Brain Research, 186, 1-22.

[16] Los, S. A., & Van der Burg, E. (2013). Sound speeds vision through preparation, not integration. Journal of Experimental Psychology: Human Perception and Performance, 39, 1612-1624.

[17] Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. Perception & Psychophysics, 59, 1-22.

[18] Kodama, D., Mekada, Y., Ayama, M., Kazuga, M. (2000). Consideration on a functional visual field in visual

and conversation tasks while driving. ITE Technical Report, 24, 7-12.

[19] Lee, J., & Spence, C. (2017). On the spatial specificity of audiovisual crossmodal exogenous cuing effects. Acta Psychologica, 177, 78-88.

[20] Perrott, D. R., Saberi, K., Brown, K., & Strybel, T. Z. (1990). Auditory psychomotor coordination and visual search performance. Perception & Psychophysics, 48, 214-226.

[21] Ho, C., & Spence, C. (2009). Using peripersonal warning signals to orient a driver's gaze. Human Factors, 51, 539-556.

[22] Asano, F., Suzuki, Y., & Sone, T. (1990). Role of spectral cues in median plane localization. Journal of the Acoustical Society of America, 88, 159-168.

[23] Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. Proceedings of the National Academy of Sciences of the USA, 111, 6104-6108.

[24] Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. Journal of the Acoustical Society of America, 94, 111-123.

[25] Wightman, F. L., & Kistler, D. J. (1999). Resolution of front–back ambiguity in spatial hearing by listener and source movement. Journal of the Acoustical Society of America, 105, 2841-2853.

[26] Ignashchenkova, A., Dicke, P. W., Haarmeier, T., & Their, P. (2004). Neuron-specific contribution of the superior colliculus to overt and covert shifts of attention. Nature Neuroscience, 7, 56-64.

[27] Kustov, A. A., & Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. Nature, 384, 74-77.

[28] Moore, T., Armstrong, K. M., & Fallah, M. (2003). Visuomotor origins of covert spatial attention. Neuron, 40, 671-683.

[29] Kadunce, D. C., Vaughan, W. J., Wallace, M. T., & Stein, B. E. (2001). The influence of visual and auditory receptive field organization on multisensory integration in the superior colliculus. Experimental Brain Research, 139, 303-310.

[30] Roberts, K. L., Summerfield, A. Q. and Hall, D. A. (2009) "Covert auditory spatial orienting: An evaluation of the spatial relevance hypothesis," J. Exp. Psychol. Hum. Percept. Perform., 35, 1178–1191.