

Optimal Time Window for the Integration of Spatial Audio-Visual Information in Virtual Environments

Jiacheng Liu* Vit Drga† Ifat Yasin‡

Dept. Computer Science, University College London, UK.

ABSTRACT

Sound duration and location may influence both auditory and visual perception with consequences for the judgement of both auditory-visual event location and integration. This study investigated audio-visual integration in a virtual environment using both short- and long-duration auditory stimuli with visual stimuli temporally offset from the start of the auditory stimulus, to investigate the effects of top-down neural effects on perception. Two tasks were used, an auditory localization task and a detection task (judgement of audio-visual synchrony). Eleven participants took part in the study using a HTC Vive Pro. The short-duration auditory stimuli (35-ms spatialized sound) and long-duration auditory stimuli (600-ms non-spatialized sound followed by 35 ms of spatialized sound) were presented at -60° , -30° , 0° , $+30^\circ$ and $+60^\circ$ degrees azimuth, with the visual stimulus presented synchronously or asynchronously with respect to the start of the auditory stimulus. Results showed that localization errors were larger for the longer-duration stimuli and judgements of audio-visual synchrony tended to be improved for stimuli presented at $\pm 30^\circ$. Top-down neural processing can affect spatial localization and audio-visual processing. Auditory localization errors and audio-visual synchrony detection may reveal the effects of underlying neural feedback mechanisms that can be harnessed to optimize audio-visual experiences in virtual environments.

Keywords: Auditory localization, Visual, Virtual, Temporal.

Index Terms: H.1.2 [Models and Principles]: User/Machine Systems—Software psychology, Human factors; H.5.1 [Information and Interfaces and presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities.

1 INTRODUCTION

Time delays (latencies) are inherently associated with virtual environments, often defined as the delay between a user's action and the updated virtual response to that action. Such system latencies can affect user presence, interaction, and ease within a virtual environment. Human perceptual processing is also known to be associated with neural latencies accompanying the processing of uni-modal or cross-modal sensory information. Understanding how perceptual processing latencies affect uni-modal or cross-modal sensory interactions allows for a) an adaptive approach to the temporal presentation of audio and visual events in virtual environments in order to optimize the perceptual experience, and b) an improved understanding of how perceptual latencies may interact with system latencies to affect presence.

Interpreting and encoding information from our immediate surroundings involves the integration of inputs from multiple senses. Such multisensory integration requires the combination of outputs from different sensory channels/pathways, in order for the human brain to produce a unitary percept and build up a coherent and accurate picture of events in the real world [6, 8]. For instance, auditory and visual information is combined to estimate object properties such as size or location [19, 3]. Some studies (using non-spatialized auditory stimuli) have shown either an enhancement or reduction in audio-visual integration, depending on the temporal relationship between auditory and visual information [1, 5]. Studies which have used spatialized auditory stimuli (using long-duration stimuli) suggest there may be different optimal times for synchronizing auditory with visual events [9, 21].

An aspect of audio-visual integration and synchrony judgement which has been overlooked is that of an internal (auditory system) response (gain) reduction which occurs with a sluggish onset delay of between 25-40 ms after the onset of a sound, and an approximate exponential time constant of 160-280 ms (for sounds above about 40 dB SPL in level and longer than about 40-ms duration) [4,25]. This sluggish auditory response reduction also appears to differentially affect the processing of sounds from different spatial locations. A particular focus of virtual reality (VR) system design is to reduce end-to-end-latency effects for optimum audio-visual effects. However, since the perceptual system itself has its own inherent auditory and visual processing latencies, then an alternative approach for VR design may be to harness the perceptual system's own internal processing delays (for different audio-visual temporal combinations and spatial locations). This may mean focusing on reducing only those system latencies which are detrimental to perception, yet retaining the latencies which could be used to enhance perception.

This study investigates the perceptual effects of triggering an internal auditory response gain reduction on both auditory sound localization and judgements of audio-visual synchrony in a virtual environment. We hypothesize that a spatialized sound (short enough in duration not to trigger the internal gain response reduction) will be easier to perceptually localize (displaying reduced localization errors) whilst a spatialized sound presented after a long-duration non-spatialized sound (long enough in duration and high enough in level to trigger the internal gain response reduction) will be more difficult to perceptually localize (displaying increased localization errors). We also hypothesize that a visual stimulus co-occurring during the spatialized auditory sound presentation will be affected by any internal response reduction. Specifically, that a visual stimulus co-occurring with the spatialized sound (in the absence of a preceding long-duration non-spatialized sound) will be judged to be more synchronous with the spatialized stimulus than the case when a preceding long-duration non-spatialized sound is presented prior to the spatialized sound.

2 METHOD

Eleven participants (7 females 4 males; age range 21-24 years) undertook the VR task. The study was approved by the UCL Ethics Committee, and participants were not paid for their participation.

The auditory broadband Gaussian noise stimulus had a total duration of 35 ms and was convolved with non-individualized HRTFs from an MIT database [11] [18], in order to present spatialized sounds to the participants from angles of $+60^\circ$, $+30^\circ$, 0° , -30° , -60° azimuth. The angles of 0° , $\pm 30^\circ$ and $\pm 60^\circ$ were chosen to span a 5-point azimuth range up to a limit of around 60° eccentricity, for studying audio-visual interactions [10]. These values were chosen so participants need deal with only a small, easily spatially discriminable set of both visual and auditory stimuli, to reduce potential effects of spatial uncertainty on judgements of temporal asynchrony. The long-duration (600-ms) non-spatialized (binaurally uncorrelated) broadband Gaussian noise stimulus was presented at 60 dB SPL. The temporal gap between the non-spatialized sound offset and spatialized sound onset was set at 0 ms. The stimuli were presented using a HTC Vive Pro.

There were two main experimental conditions. In the SHORT condition, only the short-duration spatialized sound was presented and a disc of white light (10-ms duration) was presented (i.e. visual onset) at the spatial location of the auditory stimulus, either at the onset of the sound (0 ms), 15 ms after the onset of the sound or 25 ms after the offset of the sound. The visual stimulus presentation after the offset of the short-duration spatialized sound acted as a control condition. Since the short-duration sound could not have activated any response gain reduction, the visual stimulus should not be affected by any lingering decay of the gain-reduction effect which is known to be sluggish [12]. The short-duration auditory sound should not trigger the internal response reduction so it is anticipated that the auditory localization or audio-visual synchrony judgements should not differ by a great extent between the three audio-visual conditions (0, 15, 60 ms). In the LONG condition, the short-duration spatialized sound was presented after a preceding long-duration (600 ms) non-spatialized sound. The role of the non-spatialized sound was to trigger the internal neural response reduction mechanism. In order to trigger the response the sound needs to be long enough in duration (300 ms) and high enough in level (60 dB SPL). The sound was non-spatialized in order to not confound the processing of spatial information contained within the shorter sound that followed it.

In the LONG condition, a disc of white light (10 ms duration) was presented at the onset of the non-spatialized sound (0 ms), 300 ms after onset of the non-spatialized sound, or 615 ms after the onset of the non-spatialized sound. The 615 ms lag in the LONG condition occurred 15 ms after the onset of the spatialized sound in the LONG condition, making it directly comparable to the 15 ms lag in the SHORT condition.

Since the response reduction triggered by the longer sound is sluggish in both its onset and offset it is anticipated that i) any response reduction will still have an effect on the subsequently presented short-duration spatialized sound and ii) there will be a difference between audio-visual synchrony judgements for visual stimuli presented at these three temporal positions. A schematic of the audio-visual stimuli is presented in Figure 1, with circles representing the visual stimulus.

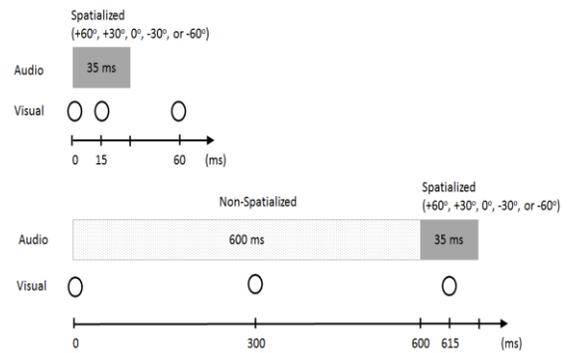


Figure 1: Schematic of the stimuli used in the task. Both auditory and visual stimuli durations and inter-stimulus temporal delays are shown. The Vive Pro visual display was run at 90Hz, with nominal lags of 0 ms, 15 ms, or 25 ms. Timing considerations due to refresh rate would not matter for the purposes of comparisons across conditions.

The current study implemented a head-turn with the nose pointing towards the sound source to measure sound localization accuracy. Many techniques have been used to indicate sound localization such as verbal indications [17]; pointing with hand/finger [22] [2] or pointing with nose [18]. In the current study, a head turn with emphasis on using the nose to point, was used as it reduces the localization errors/bias introduced by pointing with the arm/hand [16]. All participants undertook a practice session (10 mins) with the HTC Vive Pro to familiarize themselves with the VR headset and controller using a subset of the auditory stimuli (with auditory durations and angles different from those presented in the experimental task and presented without the visual stimuli). Presentation of the experimental task started with visual presentation of a cross-hair on screen to re-orient the participant to position themselves straight ahead prior to stimuli presentation. Head positions were recorded when the participant pressed the hand-held trigger button. When ready the participant would press the hand-held trigger button and progress to the next stage of the task. The initial head position (once centered) was recorded by the participant pressing the hand-held trigger button. These values were subtracted from the actual localization head position values obtained with the nose pointing towards the spatialized sound (recorded by a second hand-held trigger button press). This provided a more accurate estimate of auditory localization.

The audio or audio-visual (AV) stimulus was presented, and after a total of 3 s the participant was instructed (written instructions presented on screen) to turn their head (within 3 s) toward the location of the sound source and press their hand-held trigger button. Participants were then instructed (written instructions presented on screen) to indicate their decision (within 3 s) regarding audio-visual synchrony (pressing 1 for synchronous and 0 for not synchronous) on the hand-held controller. All trials for the conditions (SHORT or LONG) for each of the 5 angles and AV combinations were presented in pseudorandom order and each

* e-mail: jiacheng.liu.17@alumni.ucl.ac.uk

† e-mail: v.drga@ucl.ac.uk

‡ e-mail: i.yasin@ucl.ac.uk

trial repeated 4 times.

3 RESULTS

Per participant, judgements (indicated location obtained by recorded head orientation) were averaged across 4 repetitions per stimulus setting. The mean absolute value of the difference between the actual angle of presentation and the head location was calculated and is presented in Figure 2 (localization errors for the SHORT condition) and Figure 3 (localization errors for the LONG condition). Figure 2 shows that for the SHORT condition (35-ms spatialized stimulus only) localization errors differed little across the different audio-visual presentations for each angle of presentation. Per audio-visual presentation, localization errors were greater for $\pm 60^\circ$ and least for $\pm 0^\circ$ and midway for $\pm 30^\circ$. Figure 3 shows that for the LONG condition (600-ms non-spatialized stimulus followed by a 35-ms spatialized stimulus) localization errors increased the most (compared to the SHORT condition; Figure 2) for the $\pm 60^\circ$ stimuli. Synchrony judgement data is shown in Figure 4 (for the SHORT condition) and Figure 5 (for the LONG condition). Figure 4 and 5 show that AV stimuli combinations presented at $\pm 60^\circ$ were judged to be the least synchronous. In contrast, AV stimuli combinations presented in the SHORT condition (Figure 4) at $\pm 30^\circ$ and $\pm 0^\circ$ and were judged to be the most synchronous and AV stimuli combinations presented in the LONG condition (Figure 5) at $\pm 30^\circ$ and $\pm 0^\circ$ were judged to be the most synchronous when the visual stimuli was presented at the onset of the long sound, thereafter judgements of synchrony tended towards non-synchronous as the visual stimulus was presented at 300 ms and 615 ms after onset of the long-duration sound.

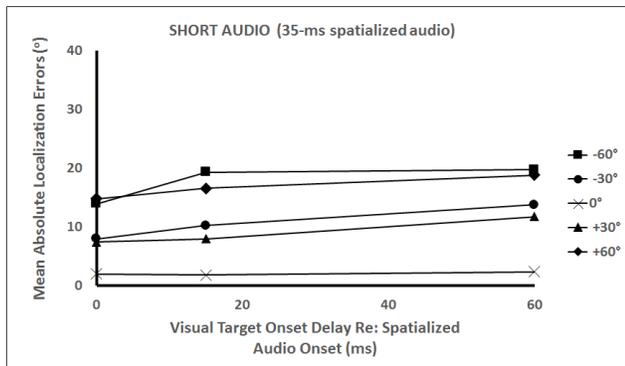


Figure 2: Mean absolute localization errors associated with visual target onset relative to the onset of the spatialized auditory stimulus in the SHORT condition.

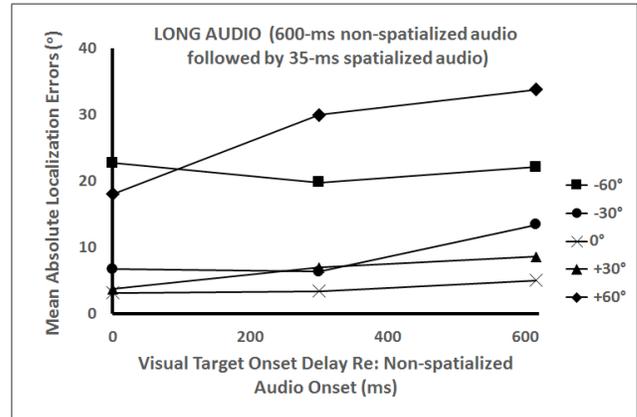


Figure 3: Mean absolute localization errors associated with visual target onset relative to the onset of the auditory non-spatialized stimulus in the LONG condition.

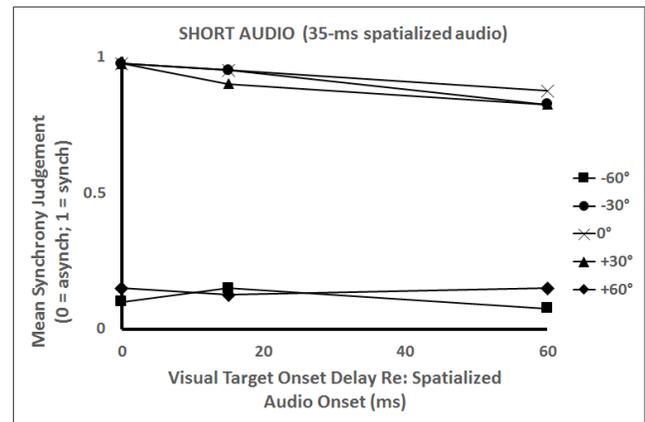


Figure 4: Mean synchrony judgement associated with visual target onset relative to the onset of the auditory spatialized stimulus in the SHORT condition.

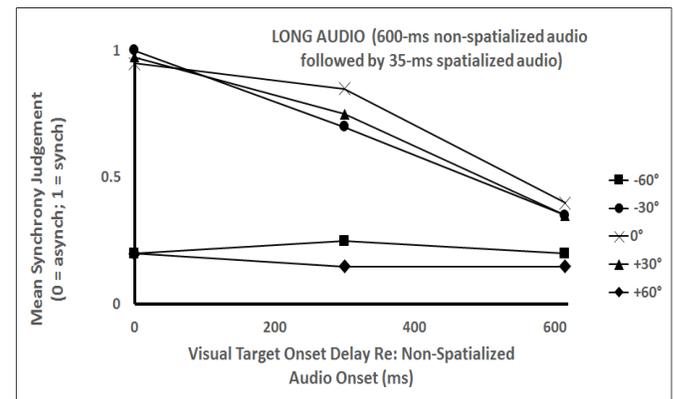


Figure 5: Mean synchrony judgement associated with visual target onset relative to the onset of the auditory non-spatialized stimulus in the LONG condition.

The results were also analyzed statistically. Firstly, to compare the localization errors in the SHORT condition (with only a 35-ms spatialized audio stimulus), a within-subject ANOVA was conducted with a factors of spatialized audio angle of presentation (five levels: +60°, +30°, 0°, -30°, -60°) and AV timing (three levels: 0 ms, 15 ms, and 60 ms). There was a significant effect of angle [$F_{(4,40)} = 5.54, p < 0.01$ with effect size, $\eta^2 = 0.36$] and AV timing [$F_{(2,20)} = 4.25, p < 0.05$ with effect size, $\eta^2 = 0.30$] but no significant interaction between angle and AV timing. *Post hoc* pairwise comparisons (Bonferroni corrected) showed that localization errors (irrespective of AV timing) for spatialized audio presented at angles of -60° and -30° was significantly greater than localization errors for spatialized audio presented at 0° [$p < 0.05$ (1-tailed)]. *Post hoc* pairwise comparisons (Bonferroni corrected) showed that localization errors (irrespective of angle of presentation) for spatialized audio presented with a visual stimulus 60 ms after onset of the audio stimulus were significantly greater than localization errors for spatialized audio when presented with a visual stimulus coincident with the onset of the audio stimulus (0 ms) [$p < 0.05$ (1-tailed)].

Similarly, to compare the localization errors in the LONG condition (presentation of 600-ms non-spatialized audio preceding a 35-ms spatialized audio stimulus), a within-subject ANOVA was conducted with factors of spatialized audio angle of presentation (five levels: +60°, +30°, 0°, -30°, -60°) and AV timing (three levels: 0 ms, 300 ms, and 615 ms). There was a significant effect of angle [$F_{(4,40)} = 13.46, p < 0.001$ with effect size, $\eta^2 = 0.57$] and AV timing [$F_{(2,20)} = 7.31, p < 0.01$ with effect size, $\eta^2 = 0.42$] and no significant interaction between angle and AV timing. *Post hoc* pairwise comparisons (Bonferroni corrected) showed that localization errors (irrespective of AV timing) for spatialized audio presented at an angle of -60° were significantly greater than localization errors for spatialized audio presented at 0° or +30° [$p < 0.05$ (1-tailed)]. Localization errors (irrespective of AV timing) for spatialized audio presented at an angle of -30° were significantly greater than localization errors for spatialized audio presented at 0° [$p < 0.05$ (1-tailed)]. Localization errors (irrespective of AV timing) for spatialized audio presented at an angle of +60° were significantly greater than localization errors for spatialized audio presented at 0°, +30° or -30° [$p < 0.05$ (1-tailed)]. *Post hoc* pairwise comparisons (Bonferroni corrected) showed that localization errors (irrespective of angle of presentation) for spatialized audio presented with a visual stimulus 615 ms after onset of the non-spatialized audio stimulus were significantly greater than localization errors for spatialized audio when presented with a visual stimulus coincident with the onset of the non-spatialized audio stimulus (0 ms) [$p < 0.01$ (1-tailed)].

To compare the localization errors for the equivalent temporal AV presentation between the SHORT (15 ms) and LONG (615 ms) conditions for all angles, an ANOVA was conducted with factors of spatialized audio angle of presentation (five levels: +60°, +30°, 0°, -30°, -60°) and overall auditory duration (two levels: SHORT, LONG). There was a significant effect of angle [$F_{(4,40)} = 12.47, p < 0.001$ with effect size, $\eta^2 = 0.56$], duration [$F_{(1,10)} = 10.61, p < 0.01$ with effect size, $\eta^2 = 0.52$] and a significant interaction between angle and duration [$F_{(4,40)} = 4.21, p < 0.05$ with effect size, $\eta^2 = 0.30$]. *Post hoc* pairwise comparisons (Bonferroni corrected) showed that localization errors were greater at -60° compared to 0° and +30° [$p < 0.05$ (1-tailed)], greater at -30° compared to 0° and +60° [$p < 0.05$ (1-tailed)] and greater at +60° compared to 0°, +30°, and -30° [$p < 0.05$ (1-tailed)]. Localization errors (irrespective of angle of presentation) were also greater for the LONG condition compared

to the SHORT condition, as hypothesized, [$p < 0.01$ (1-tailed)]. For the interaction between angle and duration, in the SHORT condition, localization errors were greater for angles of -60° and -30°, compared to 0° [$p < 0.05$ (1-tailed)]. For the LONG condition, localization errors were greater for angles of to -60° compared to 0° and +30°, and for an angle of +60° compared to -30°, 0° and +30° [$p < 0.05$ (1-tailed)].

4 DISCUSSION

Spatially congruent audio and visual stimuli were presented at differing audio-visual temporal lags in order to assess whether a preceding (non-spatialized) sound can affect judgements of both auditory localization accuracy and judgements of audio-visual synchrony. The task and effect described is distinct from the spatial ventriloquism effect which describes the bias in the localization of an auditory stimulus towards a simultaneously-presented visual stimulus that is spatially disparate from the auditory stimulus [26]. The task and effect described is also distinct from the temporal ventriloquism effect in which the ability to detect the order of two lights improves with the presentation of paired auditory stimuli [20]. Although the current study did not specifically measure response times, differences in response time with eccentricity have been noted [7]. [7] found that whereas response times to auditory stimuli presented in frontal space (central and periphery) remained similar, response times for visual targets appeared to vary across frontal space; faster response times for centrally-presented stimuli and slower response times for peripherally-presented stimuli. However, in the [7] study, a 1000-ms interval was provided to participants in which to respond. In the current study participants were provided with a much longer duration (3000 ms) in which to provide a response, which would help mitigate against any differences in response times due to eccentricity.

In the current study it was predicted that any response gain reduction triggered by the longer sound which is sluggish in both its onset and offset would increase localization errors and reduce judgements of audio-visual asynchrony. For both SHORT and LONG conditions localization errors were greater at an angle of 60° compared to 30° or 0°. Increasing localization error with a larger azimuth angle has also been shown by other studies [27]. For the SHORT condition localization errors were significantly greater (irrespective of angles of presentation) when the visual stimulus lagged the onset of the auditory stimulus by 60 ms (and 25 ms after the offset of the auditory stimulus), this didn't appear to affect judgements of synchrony, at least for audio-visual stimuli presented at angles of 0° or 30° although it is possible it could have affected judgements of synchrony for stimuli presented at 60°, if there was a perceptual trade-off between audio-visual temporal and spatial proximity effects [14]. In the LONG condition, localization errors were significantly greater (irrespective of angle of presentation) when the visual stimulus lagged the onset of the auditory (non-spatialized) stimulus by 615 ms, compared to when the visual stimulus was presented at the onset of the non-spatialized auditory sound. Audio-visual synchrony judgements were also poor at this largest audio-visual temporal gap.

If the visual stimuli presented during the longer non-spatialized auditory stimuli had acted as a spatial cue to the location (angle) of the subsequently-presented spatialized auditory stimulus, then it can be expected that the localization errors would also be reduced (despite low synchrony judgement), however, this does not appear to be the case. It is possible that an internal response gain reduction triggered by the long-duration sound

could have affected both localization accuracy and synchrony judgement.

Audio-visual stimuli presented asynchronously can be judged to be perceptually synchronous over a limited range of asynchronous presentations. However this range and synchrony judgments across modalities is affected by the task as well as stimulus parameters. Differences in stimulus presentations and parameters such as spatial congruence, respective stimulus durations, loudness/brightness, and procedural differences make it difficult to compare across studies. However, a general estimate of a just noticeable difference of around 25-50 ms has been reported [24], although this estimate is further reduced if there is a disparity in the durations of auditory and visual stimuli (as is the case for the current study) [13].

When comparing the localization errors for the case when the visual stimulus was presented 15 ms after the onset of the short spatialized audio [either without a preceding sound (SHORT condition) or with a preceding 600-ms non spatialized audio (LONG condition)], then localization errors are observed to be greater with the introduction of the preceding non-spatialized sound compared to the case when there is no preceding non-spatialized sound. As per our original hypothesis, triggering of the internal neural-mediated gain response reduction (by presentation of the longer-duration non-spatialized sound) resulted in greater localization errors in the LONG condition compared to the SHORT condition.

Future studies could investigate additional audio-visual temporal conditions for a direct statistical evaluation of the effect of a preceding auditory sound on subsequent spatialized sound, by presentation of a visual stimulus at the onset, mid-point and offset of the spatialized sound. It would also be useful to consider recording additional eye-gaze data [15] to study the effects of localization error and synchrony judgements.

An ongoing aim with the development of VR system design is to reduce end-to-end-latency effects as far as is possible in order to optimize audio-visual effects. However, human perceptual processing also displays complex neural feed-back processes associated with their own latencies, which modulate the sensory response. An alternative approach for VR design may be to capitalize on known perceptual processing latencies to enhance multisensory interactions within virtual environments, i.e., system latencies which interact positively with known perceptual latencies can be retained to further enhance the perceptual experience within a virtual environment.

5 CONCLUSION

Spatially congruent audio-visual stimuli are processed not only by bottom-up neural processes, with different inherent neural processing delays affecting perception, but as also by complex top-down neural processes which may be more sluggish and affect perception over a longer time course [23, 25]. Previous studies have shown how auditory localization errors vary with azimuth and elevation. Audio-visual processing of spatial information can be affected by top-down neural feedback affecting the spatial processing of auditory signals, which differentially affects audio-visual processing of information and synchrony judgements at some spatial locations. Inherent system latencies could either hamper multisensory perception, or in some cases be used to enhance perception for different audio-visual and spatial combinations of stimuli.

REFERENCES

- [1] W. J. Adams. The development of audio-visual integration for temporal judgements. *PLoS Computational Biology*, 12(4):e1004865, 2016.
- [2] A. Ahrens, K. D. Lund, M. Marschall, and T. Dau. Sound source localization with varying amount of visual information in virtual reality. *PloS One*, 14(3):p.e0214603, 2019.
- [3] R. Amerineni, R. S. Gupta, and L. Gupta. Multimodal Object Classification Models Inspired by Multisensory Integration in the Brain. *Brain Sciences*, 9(1):3, 2019.
- [4] B. C. Backus, and J. J. Jr. Guinan. Time-course of the human medial olivocochlear reflex. *The Journal of the Acoustical Society of America*, 119(5):2889-2904, 2006.
- [5] S. C. Boyle, S.J. Kayser, and C. Kayser. Neural correlates of multisensory reliability and perceptual weights emerge at early latencies during audio-visual integration. *European Journal of Neuroscience*, 46(10):2565-2577, 2017.
- [6] C. Chandrasekaran. Computational principles and models of multisensory integration. *Current Opinion in Neurobiology*, 43:25-34, 2017.
- [7] G. Charbonneau, M. Veronneau, C. Boudrias-Fournier, F. Lepore, and O. Collignon. The ventriloquist in periphery: impact of eccentricity-related reliability on audio-visual localization. *Journal of Vision*, 13(12):20-20, 2013.
- [8] Y. C. Chen, and C. Spence. Assessing the role of the ‘unity assumption’ on multisensory integration: A review. *Frontiers in Psychology*, 8: 445, 2017.
- [9] M. J. Crosse, G. M. Di Liberto, and E. C. Lalor. Eye can hear clearly now: inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration. *Journal of Neuroscience*, 36(38):9888-9895, 2016.
- [10] F. Frassinetti, N. Bolognini, and E. Ládavas. Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, 147(3):332-343. 2002.
- [11] W.G. Gardner, and K.D. Martin. HRTF measurements of a KEMAR Dummy-Head Microphone. MIT Media Lab Perceptual Computing Technical Report no. 280, May 1994.
- [12] A. L. James, R. J. Mount, and R. V. Harrison. Contralateral suppression of DPOAE measured in real time. *Clinical Otolaryngology and Allied Sciences*, 27(2):106-112, 2002.
- [13] I. A. Kuling, R. L. J. Van Eijk, J. F. Juola, and A. Kohlrausch. Effects of stimulus duration on audio-visual synchrony perception. *Experimental Brain Research*, 221(4):403-412. 2012.
- [14] J. Lewald, and R. Guski. Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, 16(3):468-478, 2003.
- [15] R. K. Maddox, D. A. Pospisil, G. C. Stecker, and A. K. Lee. Directing eye gaze enhances auditory spatial cue discrimination. *Current Biology*, 24(7):748-752. 2014.

- [16] P. Majdak, B. Laback, M. Goupell, and M. Mihocic. The accuracy of localizing virtual sound sources: Effects of pointing method and visual environment. *In Audio Engineering Society Convention*, 124(5). 2008
- [17] R. Mason, N. Ford, F. Rumsey, and B. De Bruyn. Verbal and nonverbal elicitation techniques in the subjective assessment of spatial sound reproduction. *Journal of the Audio Engineering Society*, 49: 366-384. 2001.
- [18] J. C. Middlebrooks, Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *Journal of the Acoustical Society of America*, 106:1493-1510. 1999.
- [19] Y. Ren, W. Yang, K. Nakahashi, S. Takahashi, and J. Wu. Audiovisual integration delayed by stimulus onset asynchrony between auditory and visual stimuli in older adults. *Perception*, 46(2):205-218, 2017.
- [20] M. D. Richards, H. C. Goltz, and A. M. Wong. Temporal ventriloquism reveals intact audiovisual temporal integration in amblyopia. *Investigative Ophthalmology and Visual Science*, 59(2):637-643, 2018.
- [21] L. Ronconi, N. A. Busch, and D. Melcher. Alpha-band sensory entrainment alters the duration of temporal windows in visual perception. *Scientific Reports*, 8(1):1-10, 2018.
- [22] B. Seeber, B. A new method for localization studies. *Acta Acustica United With Acustica*, 88:446-450. 2002.
- [23] D. Talsma, D. Senkowski, and M. G. Woldorff. Intermodal attention affects the processing of the temporal alignment of audiovisual stimuli. *Experimental Brain Research*, 198(2-3):313-328. 2009.
- [24] J. Vroomen, and M. Keetels. Perception of intersensory synchrony: a tutorial review. *Attention, Perception, and Psychophysics*, 72(4):871-884. 2010.
- [25] I. Yasin, V. Drga, and C. J. Plack. Effect of human auditory efferent feedback on cochlear gain and compression. *Journal of Neuroscience*, 34(46):15319-15326, 2014.
- [26] B. Zierul, B. Röder, C. Tempelmann, P. Bruns, and T. Noesselt. The role of auditory cortex in the spatial ventriloquism aftereffect. *Neuroimage*, 162:257-268.
- [27] M. P. Zwiers, A. J. Van Opstal, and G. D. Paige. Plasticity in human sound localization induced by compressed spatial vision. *Nature Neuroscience*, 6(2):175-181, 2017.