Does Serial Memory of Locations Benefit from Spatially Congruent Audiovisual Stimuli? Investigating the Effect of Adding Spatial Sound to Visuospatial Sequences

Benjamin Stahl
Signal Processing and Speech Communication Laboratory
Graz University of Technology
Graz, Austria
benjamin.stahl@student.tugraz.at

Georgios Marentakis
Signal Processing and Speech Communication Laboratory
Graz University of Technology
Graz, Austria
georgios.marentakis@tugraz.at

ABSTRACT

We present a study that investigates whether multimodal audiovisual presentation improves the recall of visuospatial sequences. In a modified *Corsi* block-tapping experiment, participants were asked to serially recall a spatial stimulus sequence in conditions that manipulated stimulus modality and sequence length. Adding spatial sound to the visuospatial sequences did not improve serial spatial recall performance. The results support the hypothesis that no modality-specific components in spatial working memory exist.

CCS CONCEPTS

• **Human-centered computing** → *HCI theory, concepts and models*;

KEYWORDS

Working memory, spatial audio, crossmodal space, modality-specific, supramodal

ACM Reference Format:
Benjamin Stahl and Georgios Marentakis. 2017. Does Serial Memory of Locations Benefit from Spatially Congruent Audiovisual Stimuli? Investigating the Effect of Adding Spatial Sound to Visuospatial Sequences. In *Proceedings of 19th ACM International Conference on Multimodal Interaction (ICMI'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3136755.3136773

1 INTRODUCTION

Quite often, spatial audio is used together with visual stimuli. Spatial auditory cues to visual target location improve visual search and reaction time [7, 19, 42]. In fact, congruent audiovisual (and other multimodal) stimuli are assumed to give rise to multimodal objects of perception which are processed more efficiently than unimodal stimuli by the human cognitive system [37]. Such multimodal effects have been extensively investigated within perception and attention research. In contrast, the exploration of multimodal effects within the field of working memory is a relatively young discipline, which is reviewed in Section 2.2.

ICMI'17, November 13–17, 2017, Glasgow, UK

ACM ISBN 978-1-4503-5543-8/17/11…$15.00
https://doi.org/10.1145/3136755.3136773

In this work, we investigate whether the multimodal presentation of spatial audiovisual stimulus sequences improves their memory and recall. Ways to improve serial spatial memory could find applications in communication, education and training software [15], but also in user interfaces, in which the positions of control elements need to be remembered ([34] for a review).

2 BACKGROUND

2.1 Crossmodal Perception and Attention

Crossmodal perception research has provided evidence for the interaction between different modalities [12, 37] ([38] for a review), which may be observed both in behavioral and neuroscientific (*event-related potentials, ERP*) studies. Depending on the spatial, temporal, and semantic congruency of multimodal stimuli, constructive as well as destructive interference may emerge. Improved attentional focusing and shorter reaction times to multimodal stimuli are typical behavioral results of constructive interference. These crossmodal effects are useful in supporting search, detection, and navigation tasks as well as in notifications [17, 27, 30–32].

The *spatial ventriloquism effect* describes the observation that the location of a visual stimulus biases the perceived location of a synchronous auditory stimulus [6]. This result can be considered an extension to Welch and Warren's *modality appropriateness hypothesis* [41]: The human cognitive system relies on the most suitable modality for a given task. Vision is more accurate in terms of spatial judgments and therefore dominates audition when it comes to spatial tasks. Recent research focusses on modeling perceptual judgments on multimodal stimuli by Bayesian estimates that take into account different a priori probabilities for different modalities [5, 11].

Wickens' *Multiple Resource Theory* points out that when tasks are distributed in different modalities, they occupy different resources, which can enhance dual task performance [43]. However, the spatial focus of attention is difficult to divide between modalities. When asked to filter out one of two spatially distinct spoken word streams while performing a simultaneous simulated driving task, participants performed better when the relevant auditory stream was in the direction in which visual attention was directed [39] ([13] for a review). In an *ERP* study, Eimer [14] has provided evidence suggesting the existence of a supramodal control system for spatial attention: spatial selective processing is assumed to be controlled by a single supramodal system and not by modality-specific systems.

2.2 Working Memory

2.2.1 Working Memory Components. Baddeley and Hitch [2, 3] developed the tripartite model of working memory as a precise description of the short-term store in Atkinson and Shiffrin's multi-store model (*sensory register – short-term store – long-term store*) [1]. The model differentiates between three functional entities: *central executive (CE)*, *phonological loop (PL)*, and *visuospatial sketchpad (VS)*. The *CE* coordinates the *PL* and *VS* subsystems and is responsible for focusing attention, making decisions, and switching between tasks. The *PL* is capable of storing verbal information such as spoken and written speech, lip reading, but also environmental sounds. The *VS* provides memory for visual features such as shape and color, but also for spatial features such as the location and movement of objects.

Logie [24] proposed a further partitioning of the *VS* into the *visual cache* and *inner scribe* subsystems. The visual cache provides memory for visual features such as color and shape and the inner scribe stores spatial information and movements. This theory assumes the existence of a distinct spatial component in working memory that can store spatial information from several modalities.

Lehnert and Zimmer [22] distinguish between *modality*- and *domain*-specific working memory components. *Modality* refers to the sensory input channel of mental representations, while *domain* refers to higher order information that can be extracted from the perceptual input. They identified three separate domain-specific memory systems: *verbal* (all phonological information), *object* (appearance of items), and *spatial* (location of items) domain. In contrast to the object-domain, they do not identify modality-specific components within the spatial domain. Relevant spatial information is acquired and kept in memory by shifting spatial attention, which is assumed to be supramodal [13, 14]. Thus, spatial attention is assumed to be available as a rehearsal mechanism shared by visual and auditory spatial information in working memory. In fact, Lehnert and Zimmer showed in a behavioral study that the modality in which spatial sequences were presented did not affect spatial working memory capacity [20]. Furthermore, they have found evidence for a common coding of visual and auditory spatial information in an *ERP* study [21].

2.2.2 Spatial Working Memory. Spatial working memory is often examined using the *Corsi* block-tapping task paradigm, in which subjects are asked to serially recall an observed spatial sequence of visual stimuli [9, 18, 40]. Both adaptive and fixed versions of the test have been conducted. Several computerized versions have been implemented [8]. Two measures of performance that are typically considered are 1) the *Corsi span* (maximum sequence length with at least 50% correctly recalled sequences) and 2) the *proportion of correctly recalled sequences* depending on the sequence length. Depending on the experimental setup and method, healthy subjects' average *Corsi span* varies between 5.6 and 7.2 [35].

While most experiments on spatial memory use the visual modality, research on non-visual spatial memory has also been conducted. Ruggiero and Iachini [33] performed a haptic version of the *Corsi* block-tapping test with congenitally blind, adventitiously blind, and blindfolded sighted participants. It showed that the span of recallable items (average number of recallable items: 3.7 .. 4.8, depending on participant group) was lower than the visual *Corsi span*. Adventitiously blind participants on average outperformed the other groups.

Parmentier and Jones [28] explored the spatial memory of white noise bursts from different discrete azimuth directions using fixed loudspeakers in serial and probed recall tasks. They found primacy and recency effects: serial memory of the positions of auditory stimuli is best for the first and last sequence elements. These effects typically also occur in verbal and visuospatial serial memory. Martin et al. [25] pointed out that increased spatial memory load affects also the accuracy with which sounds can be localized.

2.3 Summary, Research Question, Hypothesis

The review above suggests that visual and auditory spatial memory behave in a similar way. Multimodal combinations, however, have only been tested within object-domain memory and not within spatial memory. Delogu et al. [10] compared serial recall of verbal and non-verbal simple stimuli using auditory, visual, and semantically congruent audiovisual representations. Bimodal representations of non-verbal stimuli (environmental sounds and images) resulted in superior recall compared to unimodal ones. Similar findings in a recognition task have been reported by Heikkilä et al. [16].

Given however, that the effect of spatially congruent audiovisual stimulation on spatial working memory has not been investigated, the research question *"Does multimodal audiovisual presentation enhance the memory of visual spatial sequences?"* remains open.

Semantically congruent multimodal cues improve working memory performance due to separate but linked mental representations in the object domain. If the hypothesis concerning a supramodal spatial memory system however is correct, there are no such dual mental representations in spatial working memory: spatial memory capacity likely remains fixed irrespective of stimulus multimodality.

Does that mean that we do not expect to observe a beneficial effect of adding spatially congruent sounds to visual spatial sequences? Not necessarily. Multimodal perception research suggests that spatially congruent audiovisual stimulation leads to objects of perception that are easier to detect in comparison to unimodal ones. Despite the assumption that no dual mental representations in working memory emerge, the possibility of more efficient encoding in perception translating into a working memory benefit cannot be dismissed.

3 EXPERIMENT

To answer the research question formulated above, we designed an experiment using a modified, non-adaptive *Corsi* block-tapping task, in which participants had to serially recall spatial stimuli sequences whose length and the modality in which they were presented was varied [9, 18, 40].

There were two factors in the experiment: *stimulus modality* (visual/audiovisual) and *sequence length* (6/8). We chose sequence lengths of 6 and 8 stimuli, since we wanted to examine the modality effect at an average and a challenging sequence length. We assumed that if there was an effect of multimodal presentation, it would be more pronounced at high cognitive loads.

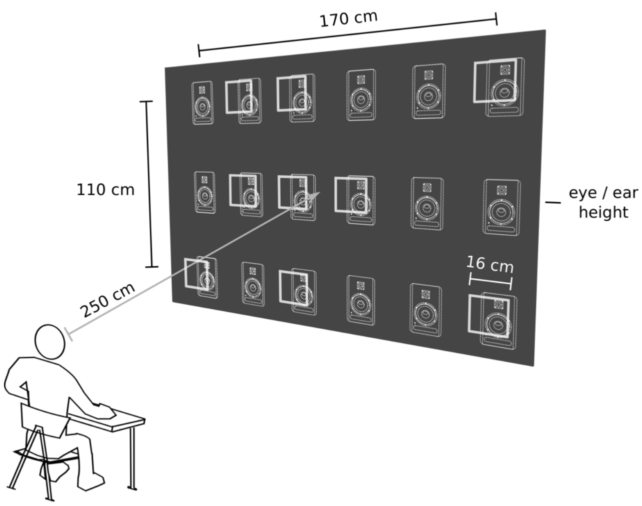


Figure 1: Experimental setup

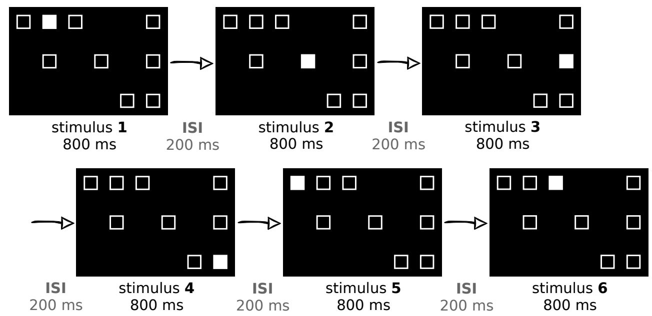


Figure 2: Sequence with 6 serial spatial stimuli

3.1 Participants

Twelve voluntary participants (aged 22-39, 3 female, 9 male) were recruited for the experiment. All were either musically trained or worked/studied in the field of audio-engineering.

3.2 Apparatus

The spatial sequence to be recalled was presented within a predefined set of 9 squares that was different in each trial. The location of each square in the set corresponded to one out of the 18 subdivisions of a 3×6 rectangular grid (see Fig. 1). We varied the set of locations out of which the spatial sequence in each trial was created in order to ensure that the task was performed using spatial memory and that no verbal coding strategies (e.g. using digits for the positions) could be used to perform the task.

During each trial, the corresponding set of squares was projected on an acoustically transparent screen in front of the participant. A loudspeaker (*Genelec 8020*) was arranged behind the center of each grid subdivision.

The maximum angular distance between the centers of the outermost squares was 37.6° horizontally and 24.8° vertically. These angles are within the minimum required binocular visual field for tasks such as driving [44]. The minimum angular distance between adjacent squares was 7.2° horizontally and 12.4° vertically, which is several times the minimum audible angles for broadband stimuli in the frontal direction [29]. Thus, positions are both auditorily and visually clearly distinguishable, which was informally tested by the authors before conducting the experiment.

The experiment took place in a dimmed and acoustically treated room. *Processing* was used for visual rendering. Sound was rendered using *Pure Data*. *OSC* was used for synchronization. Responses were given using the computer mouse.

3.3 Stimuli

Depending on the modality factor level, the spatial sequence to be recalled consisted of locations that were indicated using either unimodal visual or audiovisual stimuli. Locations were indicated visually by filling the corresponding squares with white color for 800 ms and sonically using spatially congruent white noise with sharp onsets and offsets (rectangular envelope) and a duration of 800 ms. With this choice of stimuli we aimed on introducing as little semantic content as possible. Auditory stimuli had a sound pressure level of 56 ± 0.2 dB(A) at the participant's position, measured successively for each loudspeaker with an *NTI XL2* using continuous white noise at the same level as the bursts in the experiment.

3.4 Procedure

Each participant performed two sessions, one with visual and one with audiovisual stimuli. Order (visual or audiovisual session first) was counterbalanced among participants. Between sessions, participants were given a break of at least 5 minutes.

Each session started with 4 training trials. Test trials followed. Each session was partitioned in 4 blocks with 8 repetitions per block for each sequence length. Sequence lengths were randomized within blocks. There was a 2-minute break between consecutive blocks. This resulted in a total of 2 (modality conditions) · 2 (lengths) · 4 (trial blocks) · 8 (replications) = 128 trials per participant.

In each trial, participants were presented with a spatial sequence of visual or audiovisual stimuli and were subsequently immediately asked to serially recall it. All stimulus locations were unique, i.e., appeared once in each sequence (see Fig. 2).

Participants reconstructed the sequences by left-clicking the respective (empty) squares in the recalled order. Squares lit up and in multimodal trials white-noise was played on click as in the presented sequence. A small number in the lower right corner of the clicked squares indicated its serial position in the spatial sequence as this was entered by the participant. Subjects could undo their most recent left-click by right-clicking. When done, participants clicked a button to confirm their response. Immediately after, feedback (correct/false) was given. Participants then clicked a button to continue with the next trial.

After the second session, participants were asked about the strategies they used for memorization. The whole experiment took about 70 minutes for one participant.

3.5 Results

We used the *Bernoulli* distributed dependent variable X describing the success of the event *"correctly recalled sequence"* as an indicator of serial recall performance. In the experiment, there were

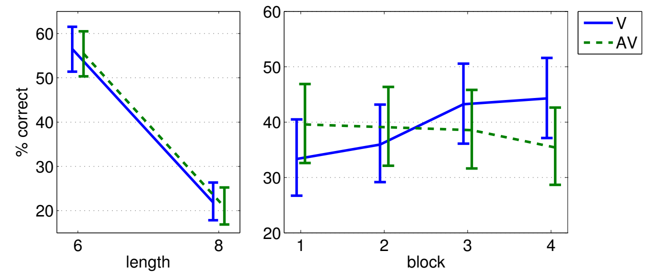
12 (participants) · 128 (trials per participant) = 1536 binary[1] observations of X. From these observations, we calculated proportions of correctly recalled sequences dependent on the factors *modality*, *sequence length* and *trial block*, as well as *Clopper-Pearson* confidence intervals for these pooled proportions (see Fig. 3).

A generalized linear mixed model [4] was fitted using *R*'s `mixed()` function from the `afex` package [36]. Effects of the *within-subjects* factors *modality*, *sequence length* and *trial block* and all of their interactions on X were estimated in a *likelihood ratio test (LRT)*.

The *LRT* showed that there is no significant main effect of the *modality* factor on X (*visual:* 39.2% *correct, audiovisual:* 38.2% *correct*) at a significance level of $\alpha = 0.05$. As expected, the *length* of the sequences clearly significantly affects the proportion of correctly recalled sequences ($df = 1$, $\chi^2 = 246.472$, $p = 10^{-16}$). Sequences with a length of 6 items were recalled correctly more than twice as likely as sequences with 8 items (*6 items:* 56.9% *correct, 8 items:* 21.4% *correct*, see Fig. 3a). There was no significant main effect of the *block* factor. Interaction effects showed to be clearly non-significant, except for the *modality* × *block* interaction, which is marginally significant ($df = 3$, $\chi^2 = 7.602$, $p = 0.055$). This can be attributed to the small improvement in performance over time (trial blocks) with unimodal stimuli vs. no improvement with audiovisual stimuli. We do not relate this marginal interaction to spatial memory and do not discuss it further.

In order to draw a stronger conclusion concerning the hypothesis that the *modality* factor has no effect, we additionally performed a *Bayes factor ANOVA* using *R*'s `anovaBF` function from the `BayesFactor` package [23, 26] on the proportions of correctly recalled sequences in all *subject-modality-length* combinations. The likelihood of a model without the *modality* factor was compared to the full *subject-modality-length* model. The *Cauchy* prior was scaled with $r = 0.5$ for the *modality* and *length* factors and with $r = 1$ for the *subject* factor (default values). Doing several simulations, we received a BF_{01} around 3.3, which we interpret as moderate evidence for the hypothesis that the *modality* factor has no effect.

Only one participant reported trying to use digits to memorize locations and did not consider this strategy helpful. This suggests a successful inhibition of verbal coding strategies and the utilization of spatial features for task performance. Interestingly, many participants reported on trying to use the *"melody"* they heard in the noise burst sequence to perform the task or verify their judgments[2].



(a) Comparison of visual (V) and audiovisual (AV) condition across length

(b) Comparison of visual (V) and audiovisual (AV) condition over time (trial blocks)

Figure 3: Graphical representation of proportions of correctly recalled sequences and 95%-confidence intervals

4 DISCUSSION AND CONCLUSIONS

We conducted an experiment in order to find out whether audiovisual (as opposed to purely visual) presentation of stimuli improves performance in a serial spatial recall task.

The comparison of the mean proportions of correctly recalled sequences in the visual and audiovisual condition (see Fig. 3a), as well as the *Bayes factor ANOVA* result, suggests that in this particular experimental setup, serial spatial recall does not benefit from presenting multimodal audiovisual stimuli. These findings could be explained by the lack of linked modality-specific components in spatial memory as suggested by Lehnert and Zimmer [22].

An alternative explanation would be that the auditory information on stimulus location was simply overwritten by the visual one. According to the *modality appropriateness hypothesis*, visual information dominates judgments of audiovisual stimuli location. Despite a benefit in detection from an additional auditory stimulus, no such benefit in the perception (and likely the memory) of the location of the audiovisual stimulus can be expected. The reported *"melody"* strategies (trying to use not the location but other features from the auditory stimuli) support this explanation. However, we cannot argue towards a specific explanation based on this experiment.

While auditory stimuli were easily *distinguishable*, they were still clearly harder/worse *localizable* than visual stimuli in this experiment. Without artificial impairment of the visual localizability, it is not possible to provide equal localizability for auditory and visual stimuli. Further research may investigate the effect of different cognitive loads for auditory and visual localization on memory.

We conclude from this study that in such clear and undisturbed scenarios, memorization of serial spatial events does not benefit from adding spatially congruent sounds to visuospatial sequences. Facilitation of detection and attention due to multimodal stimuli did not improve working memory in this particular task. However, the impact of multimodal stimulation on spatial memory needs to be examined in more detail and in scenarios which manipulate the cognitive load associated with the task using faster presentation rates, distracting events, and spatially incongruent auditory stimuli. In this connection, the effect of musical background of participants should also be investigated. Such experiments could help to illuminate the relationship between perceptive encoding and working memory performance.

ACKNOWLEDGMENTS

We thank the participants for their voluntary participation and the Institute of Electronic Music and Acoustics (IEM) Graz for providing the premises. Furthermore, we acknowledge the contribution of Marian Weger and Stefan Warum in setting up the experiment.

[1] "0" (= failure) or "1" (= success)

[2] This phenomenon can be explained by spectral coloration due to the different locations of the loudspeakers and room reflections. However, this strategy was not helpful, since timbres at different positions were very similar and did not provide any meaningful information.

REFERENCES

[1] Richard C Atkinson and Richard M Shiffrin. 1968. Human memory: A proposed system and its control processes. *Psychology of learning and motivation* 2 (1968), 89–195. https://doi.org/10.1016/S0079-7421(08)60422-3

[2] Alan D. Baddeley. 2012. Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology* 63, 1 (2012), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

[3] Alan D. Baddeley and Graham J. Hitch. 1974. Working Memory. In *The Psychology of Learning and Motivation: Advances in Research and Theory*, Gordon H. Bower (Ed.). Vol. 8. Academic Press, New York, NY, USA, 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1

[4] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, Articles* 67, 1 (2015), 1–48. https://doi.org/10.18637/jss.v067.i01

[5] Peter W. Battaglia, Robert A. Jacobs, and Richard N. Aslin. 2003. Bayesian integration of visual and auditory signals for spatial localization. *J. Opt. Soc. Am. A* 20, 7 (Jul 2003), 1391–1397. https://doi.org/10.1364/JOSAA.20.001391

[6] Paul Bertelson and Beatrice de Gelder. 2004. The psychology of multimodal perception. In *Crossmodal Space and Crossmodal Attention*, Charles Spence and Jon Driver (Eds.). OUP, Oxford, United Kingdom, Chapter 7, 141–177.

[7] Adelbert W. Bronkhorst, J. A. (Hans) Veltman, and Leo Van Breda. 1996. Application of a Three-Dimensional Auditory Display in a Flight Task. *Human Factors* 38, 1 (1996), 23–33. https://doi.org/10.1518/001872096778940859 PMID: 8682519.

[8] Riccardo Brunetti, Claudia Del Gatto, and Franco Delogu. 2014. eCorsi: implementation and testing of the Corsi block-tapping task for digital tablets. *Frontiers in psychology* 5 (2014), 939. https://doi.org/10.3389/fpsyg.2014.00939

[9] Philip Michael Corsi. 1972. *Human memory and the medial temporal region of the brain.* Ph.D. Dissertation. McGill University, Montreal, Canada. http://digitool.Library.McGill.CA:80/R/-?func=dbin-jump-full&object_id=93903&silo_library=GEN01 [Online; accessed 13-May-2017].

[10] Franco Delogu, Antonino Raffone, and Marta Olivetti Belardinelli. 2009. Semantic encoding in working memory: Is there a (multi) modality effect? *Memory* 17, 6 (2009), 655–663. https://doi.org/10.1080/09658210902998054

[11] Sophie Deneve and Alexandre Pouget. 2004. Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology-Paris* 98, 1 (2004), 249–258. https://doi.org/10.1016/j.jphysparis.2004.03.011

[12] Jon Driver and Charles Spence. 1998. Crossmodal attention. *Current opinion in neurobiology* 8, 2 (1998), 245–253. https://doi.org/10.1016/S0959-4388(98)80147-5

[13] Jon Driver and Charles Spence. 2004. Crossmodal spatial attention: evidence from human performance. In *Crossmodal Space and Crossmodal Attention*, Charles Spence and Jon Driver (Eds.). OUP, Oxford, United Kingdom, Chapter 8, 179–220.

[14] Martin Eimer. 1999. Can attention be directed to opposite locations in different modalities? An ERP study. *Clinical neurophysiology* 110, 7 (1999), 1252–1259. https://doi.org/10.1016/S1388-2457(99)00052-8

[15] Mary Hegarty and Sarah Kriz. 2008. Effects of knowledge and spatial ability on learning from animation. In *Learning with animation: Research implications for design*, Richard Lowe and Wolfgang Schnotz (Eds.). Cambridge University Press, New York, NY, USA, Chapter 1, 3–29.

[16] Jenni Heikkilä, Kimmo Alho, Heidi Hyvönen, and Kaisa Tiippana. 2015. Audiovisual Semantic Congruency During Encoding Enhances Memory Performance. *Experimental Psychology* 62, 2 (2015), 123–130. https://doi.org/10.1027/1618-3169/a000279

[17] Cristy Ho and Charles Spence. 2005. Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of experimental psychology: Applied* 11, 3 (2005), 157. https://doi.org/10.1037/1076-898X.11.3.157

[18] Roy P. C. Kessels, Martine J. E. van Zandvoort, Albert Postma, L. Jaap Kappelle, and Edward H. F. de Haan. 2000. The Corsi Block-Tapping Task: Standardization and Normative Data. *Applied Neuropsychology* 7, 4 (2000), 252–258. https://doi.org/10.1207/S15324826AN0704_8

[19] Camilla H. Larsen, David S. Lauritsen, Jacob J. Larsen, Marc Pilgaard, Jacob B. Madsen, and Rasmus Stenholt. 2014. Aurally Aided Visual Search Performance Comparing Virtual Audio Systems. In *137th International AES Convention*. Audio Engineering Society, New York, NY, USA, 1–10. http://www.aes.org/e-lib/browse.cfm?elib=17473

[20] Günther Lehnert and Hubert D Zimmer. 2006. Auditory and visual spatial working memory. *Memory & cognition* 34, 5 (2006), 1080–1090. https://doi.org/10.3758/BF03193254

[21] Günther Lehnert and Hubert D. Zimmer. 2008. Common coding of auditory and visual spatial information in working memory. *Brain Research* 1230 (2008), 158–167. https://doi.org/10.1016/j.brainres.2008.07.005

[22] Günther Lehnert and Hubert D. Zimmer. 2008. Modality and domain specific components in auditory and visual working memory tasks. *Cognitive Processing* 9, 1 (2008), 53–61. https://doi.org/10.1007/s10339-007-0187-6

[23] Jimmie Leppink, Patricia O'Sullivan, and Kal Winston. 2017. Evidence against vs. in favour of a null hypothesis. *Perspectives on Medical Education* 6, 2 (01 Apr 2017), 115–118. https://doi.org/10.1007/s40037-017-0332-6

[24] R.H. Logie. 1995. *Visuo-spatial Working Memory*. Lawrence Erlbaum, Hove, East Sussex, United Kingdom.

[25] Russell L Martin, Patrick Flanagan, Ken I McAnally, and Geoff Eberle. 2011. Memory for the locations of environmental sounds. *The Journal of the Acoustical Society of America* 129, 6 (2011), 3873–3883. https://doi.org/10.1121/1.3589254

[26] Richard D. Morey and Jeffrey N. Rouder. 2015. BayesFactor: Computation of Bayes Factors for Common Designs. (2015). https://CRAN.R-project.org/package=BayesFactor R package version 0.9.12-2.

[27] Mary Kim Ngo and Charles Spence. 2010. Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Attention, Perception, & Psychophysics* 72, 6 (2010), 1654–1665. https://doi.org/10.3758/APP.72.6.1654

[28] Fabrice BR Parmentier and Dylan M Jones. 2000. Functional characteristics of auditory temporal–spatial short-term memory: Evidence from serial order errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26, 1 (2000), 222. https://doi.org/10.1037/0278-7393.26.1.222

[29] D. Perrott and K. Saberi. 1990. Minimum audible angle thresholds for sources varying in both elevation and azimuth. *The Journal of the Acoustical Society of America* 87, 4 (1990), 1728–1731. https://doi.org/10.1121/1.399421

[30] David R Perrott, John Cisneros, Richard L McKinley, and William R D'Angelo. 1996. Aurally aided visual search under virtual and free-field listening conditions. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 38, 4 (1996), 702–715. https://doi.org/10.1518/001872096778827260

[31] David R Perrott, Kourosh Saberi, Kathleen Brown, and Thomas Z Strybel. 1990. Auditory psychomotor coordination and visual search performance. *Perception & Psychophysics* 48, 3 (1990), 214–226. https://doi.org/10.3758/BF03211521

[32] David R Perrott, Toktam Sadralodabai, Kourosh Saberi, and Thomas Z Strybel. 1991. Aurally aided visual search in the central visual field: Effects of visual load and visual enhancement of the target. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 33, 4 (1991), 389–400. https://doi.org/10.1177/001872089103300402

[33] Gennaro Ruggiero and Tina Iachini. 2010. The role of vision in the Corsi block-tapping task: evidence from blind and sighted people. *Neuropsychology* 24, 5 (2010), 674. https://doi.org/10.1037/a0019594

[34] Joey Scarr, Andy Cockburn, Carl Gutwin, et al. 2013. Supporting and exploiting spatial memory in user interfaces. *Foundations and Trends® in Human–Computer Interaction* 6, 1 (2013), 1–84. https://doi.org/10.1561/1100000046

[35] Dieter Schellig and HA Hättig. 1993. Die Bestimmung der visuellen Merkspanne mit dem Block-Board. *Zeitschrift für Neuropsychologie* 4, 2 (1993), 104–112. https://www.researchgate.net/profile/Dieter_Schellig/publication/286272982_Die_Bestimmung_der_visuellen_Merkspanne_mit_dem_Block-Board/links/569df5b008ae950bd7a78676/Die-Bestimmung-der-visuellen-Merkspanne-mit-dem-Block-Board.pdf [Online; accessed 14-May-2017].

[36] Henrik Singmann, Ben Bolker, Jake Westfall, and Frederik Aust. 2016. afex: Analysis of Factorial Experiments. (2016). https://CRAN.R-project.org/package=afex R package version 0.16-1.

[37] Charles Spence. 2007. Audiovisual multisensory integration. *Acoustical science and technology* 28, 2 (2007), 61–70. https://doi.org/10.1250/ast.28.61

[38] Charles Spence and Jon Driver. 2004. *Crossmodal Space and Crossmodal Attention*. OUP, Oxford, United Kingdom.

[39] Charles Spence and Liliana Read. 2003. Speech shadowing while driving on the difficulty of splitting attention between eye and ear. *Psychological science* 14, 3 (2003), 251–256. https://doi.org/10.1111/1467-9280.02439

[40] André Vandierendonck, Eva Kemps, Maria Chiara Fastame, and Arnaud Szmalec. 2004. Working memory components of the Corsi blocks task. *British Journal of Psychology* 95, 1 (2004), 57–79. https://doi.org/10.1348/000712604322779460

[41] Robert B Welch and David H Warren. 1980. Immediate perceptual response to intersensory discrepancy. *Psychological bulletin* 88, 3 (1980), 638. https://doi.org/10.1037/0033-2909.88.3.638

[42] Elizabeth M Wenzel, Martine Godfroy-Cooper, and Joel D Miller. 2014. Spatial Auditory Displays: Substitution and Complemenarity to Visual Displays. In *20th International Conference on Auditory Display*. Georgia Institute of Technology, Atlanta, GA, USA, 1–8. http://hdl.handle.net/1853/52061

[43] Christopher D Wickens. 1980. The structure of attentional resources. In *Attention and performance*, R.S. Nickerson (Ed.). Vol. 8. Erlbaum, Hillsdale, NJ, USA, 239–257.

[44] Joanne M Wood and Rod Troutbeck. 1992. Effect of restriction of the binocular visual field on driving performance. *Ophthalmic and Physiological Optics* 12, 3 (1992), 291–298. https://doi.org/10.1111/j.1475-1313.1992.tb00400.x