Lecture 8: Multi-Sensory Interactions

Part 03

Demonstrations and Applications: Multisensory effects

Relevance

Lederman, S.J., Klatzky, R.L., Morgan, T. and Hamilton, C., 2002, March. Integrating multimodal information about surface texture via a probe: relative contributions of haptic and touch-produced sound sources. In Proceedings 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. HAPTICS 2002 (pp. 97-104). IEEE.

Giordano, B.L., Visell, Y., Yao, H.Y., Hayward, V., Cooperstock, J.R. and McAdams, S., 2012. Identification of walked-upon materials in auditory, kinesthetic, haptic, and audio-haptic conditions. *The Journal of the Acoustical Society of America*, *131*(5), pp.4002-4012.

Spence, C. and Ho, C., 2008. Tactile and multisensory spatial warning signals for drivers. *IEEE Transactions on Haptics*, *1*(2), pp.121-129.

Ménélas, B., Picinalli, L., Katz, B.F. and Bourdot, P., 2010, March. Audio haptic feedbacks for an acquisition task in a multi-target context. In *2010 IEEE symposium on 3D user interfaces (3DUI)* (pp. 51-54). IEEE.

Sigrist, R., Rauter, G., Riener, R. and Wolf, P., 2013. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review*, *20*(1), pp.21-53.

Weisenberger, J.M. and Poling, G.L., 2004, March. Multisensory roughness perception of virtual surfaces: effects of correlated cues. In *12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2004. HAPTICS'04. Proceedings.* (pp. 161-168). IEEE.

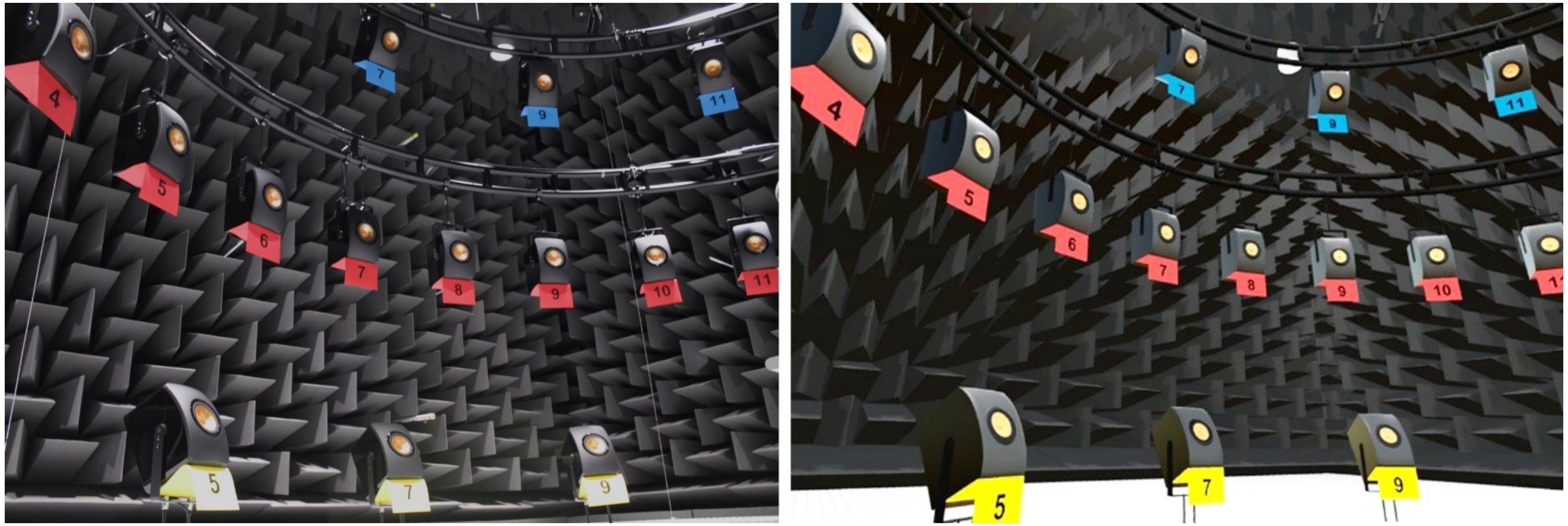
Ahrens, A., Lund, K.D., Marschall, M. and Dau, T., 2019. Sound source localization with varying amount of visual information in virtual reality. *PloS one*, *14*(3), p.e0214603.

“**Abstract:** To achieve accurate spatial auditory perception, subjects typically require personal headrelated transfer functions (**HRTFs**) and the freedom for head movements. Loudspeaker based virtual sound environments allow for realism without individualized measurements. To study audio-visual perception in realistic environments, the combination of **spatially tracked** head mounted displays (HMDs), also known as virtual reality glasses, and virtual sound environments may be valuable. However, HMDs were recently shown **to affect the subjects’ HRTFs** and thus might **influence sound localization** performance. Furthermore, due to **limitations of the reproduction of visual information on the HMD**, audio-visual perception might be influenced.

Here, a sound localization experiment was conducted both **with and without an HMD** and with a **varying amount of visual information** provided to the subjects. Furthermore, interaural time and level difference errors (ITDs and ILDs) as well as spectral perturbations induced by the HMD were analyzed and compared to the perceptual localization data.

The results showed a **reduction of the localization accuracy when the subjects were wearing an HMD and when they were blindfolded**. The HMD-induced error in azimuth localization was found to be larger in the left than in the right hemisphere. When visual information of the limited set of source locations was provided, the localization error induced by the HMD was found to be negligible. Presenting **visual information of hand-location** and room dimensions showed better sound localization performance compared to the condition with no visual information. The addition of possible source locations further improved the localization accuracy. Also adding pointing feedback in form of a virtual laser pointer improved the accuracy of elevation perception but not of azimuth perception.”

Ahrens, A., Lund, K.D., Marschall, M. and Dau, T., 2019. Sound source localization with varying amount of visual information in virtual reality. *PloS one*, *14*(3), p.e0214603.



Designed a sound localisation task with varying amounts of visual information in VR.

Photography (left) and screenshot (right) of the acoustic reproduction system in the real and in the virtual environment (RE and VE), respectively.

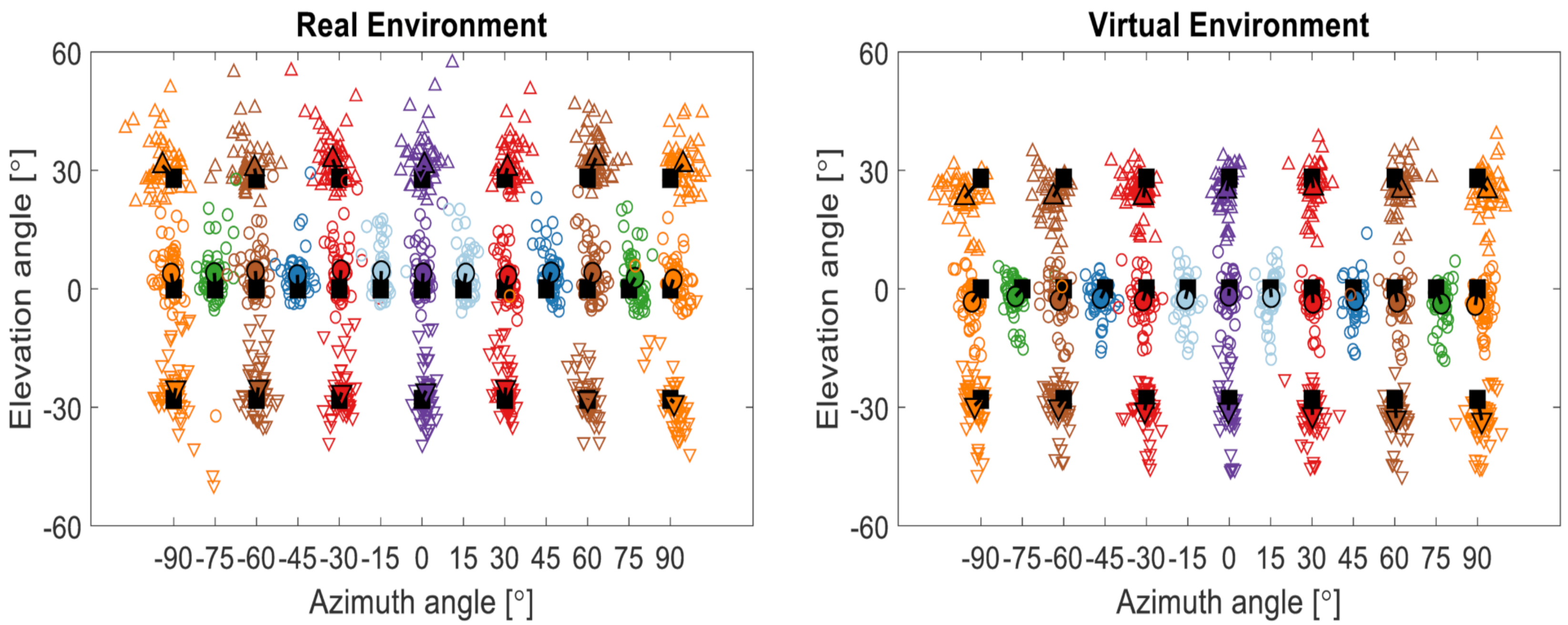
Ahrens, A., Lund, K.D., Marschall, M. and Dau, T., 2019. Sound source localization with varying amount of visual information in virtual reality. *PloS one*, *14*(3), p.e0214603.

Block	Visual information	Stimulus	HMD
I	Blind-folded	Acoustic	No
	Blind-folded	Acoustic	Yes
II	Virtual env., no loudspeaker (LS)	Acoustic	Yes
III	Virtual env., LS	Visual	Yes
	Real env.	Visual	No
	Real env.	Acoustic	No
	Virtual env., LS	Acoustic	Yes
IV	Virtual env., LS, laser pointer	Acoustic	Yes

https://doi.org/10.1371/journal.pone.0214603.t001

Table 1. Overview over the conditions considered and grouped into blocks.
Conditions were randomized within blocks.
The conditions varied in available visual information, target stimulus, and if the head-mounted display was worn or not.

Ahrens, A., Lund, K.D., Marschall, M. and Dau, T., 2019. Sound source localization with varying amount of visual information in virtual reality. *PloS one*, *14*(3), p.e0214603.



Negative angles represent:
Sources to left (azimuth angle) and sources downwards (elevation angle).
Filled black squares - 27 source locations.
Coloured symbols - Individual responses of participants.
Large open black symbols - Mean responses, averaged across participants and repetitions.

Recap:

- Responses of participants in the two visual localization experiments (RE and VE). In RE & VE participants' task was to point to centre of loudspeaker indicated on a screen.
- Results:
- Generally, participants pointed close to the correct loudspeaker, whereby the precision of the responses was generally higher for azimuth than for elevation localization.
- Found greater errors in sound localisation wearing the HMD, but errors were reduced by providing a limited set of visual information.
- Also providing hand-location and room dimensions improved localisation.

Ahrens, A., Lund, K.D., Marschall, M. and Dau, T., 2019. Sound source localization with varying amount of visual information in virtual reality. *PloS one*, *14*(3), p.e0214603.

Overall Results:

Found greater errors in sound localisation wearing the HMD, but errors were reduced by providing a limited set of visual information.

Also providing hand-location and room dimensions improved localisation.

Tidoni, E., Gergondet, P., Kheddar, A. and Aglioti, S.M., 2014. Audio-visual feedback improves the BCI performance in the navigational control of a humanoid robot. *Frontiers in neurorobotics*, *8*, p.20.

"Advancement in brain computer interfaces (BCI) technology allows people to actively interact in the world through surrogates. **Controlling real humanoid robots using BCI** as intuitively as we control our body represents a challenge for current research in robotics and neuroscience. In order to successfully interact with the environment the **brain integrates multiple sensory cues** to form a coherent representation of the world. Cognitive neuroscience studies demonstrate that multisensory integration **may imply a gain** with respect to a single modality and ultimately improve the overall sensorimotor performance. For example, reactivity to simultaneous visual and auditory stimuli may be higher than to the sum of the same stimuli delivered in isolation or in temporal sequence.
Yet, knowledge about whether **audio-visual integration may improve** the control of a surrogate is meager.
To explore this issue, we provided **human footstep sounds** as audio feedback to BCI users while **controlling a humanoid robot**. Participants were asked to steer their robot surrogate and perform a pick-and-place task through BCI-SSVEPs.
We found that **audio-visual synchrony between footsteps sound and actual humanoid's walk reduces the time required for steering** the robot. Thus, auditory feedback congruent with the humanoid actions may improve motor decisions of the BCI's user and help in the feeling of control over it. Our results shed light on the possibility to increase robot's control through the combination of multisensory feedback to a BCI user."

Tidoni, E., Gergondet, P., Kheddar, A. and Aglioti, S.M., 2014. Audio-visual feedback improves the BCI performance in the navigational control of a humanoid robot. *Frontiers in neurorobotics*, *8*, p.20.

Controlling robots using BCI

Multisensory integration of information important

Investigated the details of audio-visual integration and how it may affect control of a robot.

Tidoni, E., Gergondet, P., Kheddar, A. and Aglioti, S.M., 2014. Audio-visual feedback improves the BCI performance in the navigational control of a humanoid robot. *Frontiers in neurorobotics*, *8*, p.20.

Figure 1. Participants located in Rome (Italy) controlled an HRP-2 humanoid robot located in Tsukuba (Japan) by a SSVEPs-BCI system. The subjects guided the robot from a starting position to a first table (marked in green as "1") to grasp the bottle and then drop the bottle as close as possible to a target location marked with two concentric circles on a second table (marked in green as "2").

During the task the BCI user heard a footstep sound either synchronous or asynchronous with the real footsteps of the robot

Tidoni, E., Gergondet, P., Kheddar, A. and Aglioti, S.M., 2014. Audio-visual feedback improves the BCI performance in the navigational control of a humanoid robot. *Frontiers in neurorobotics*, *8*, p.20.

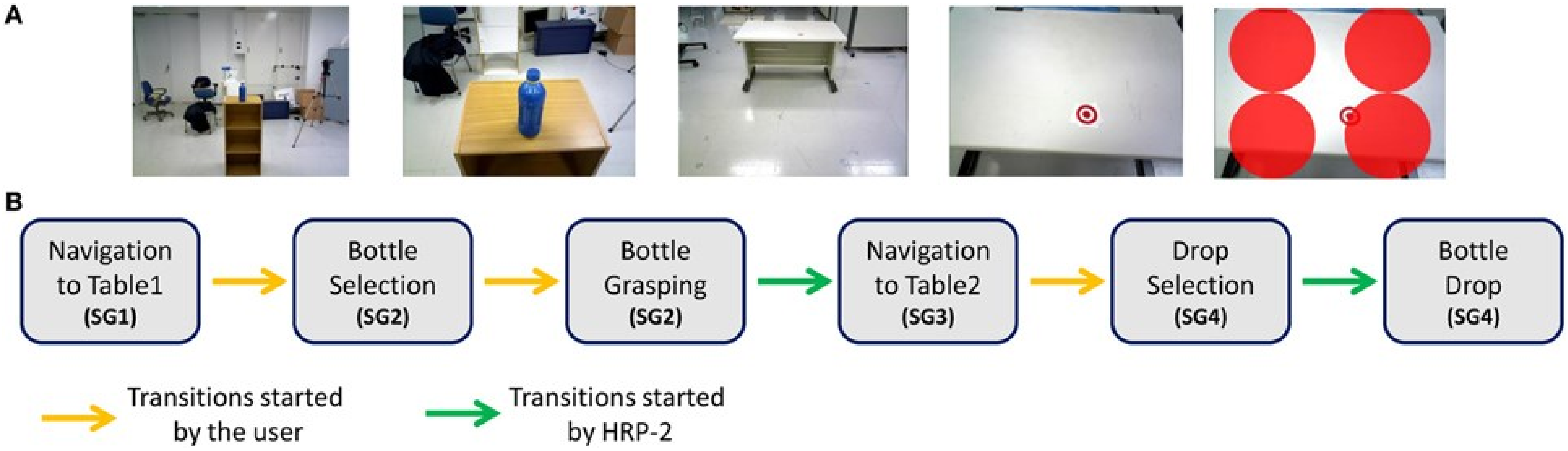
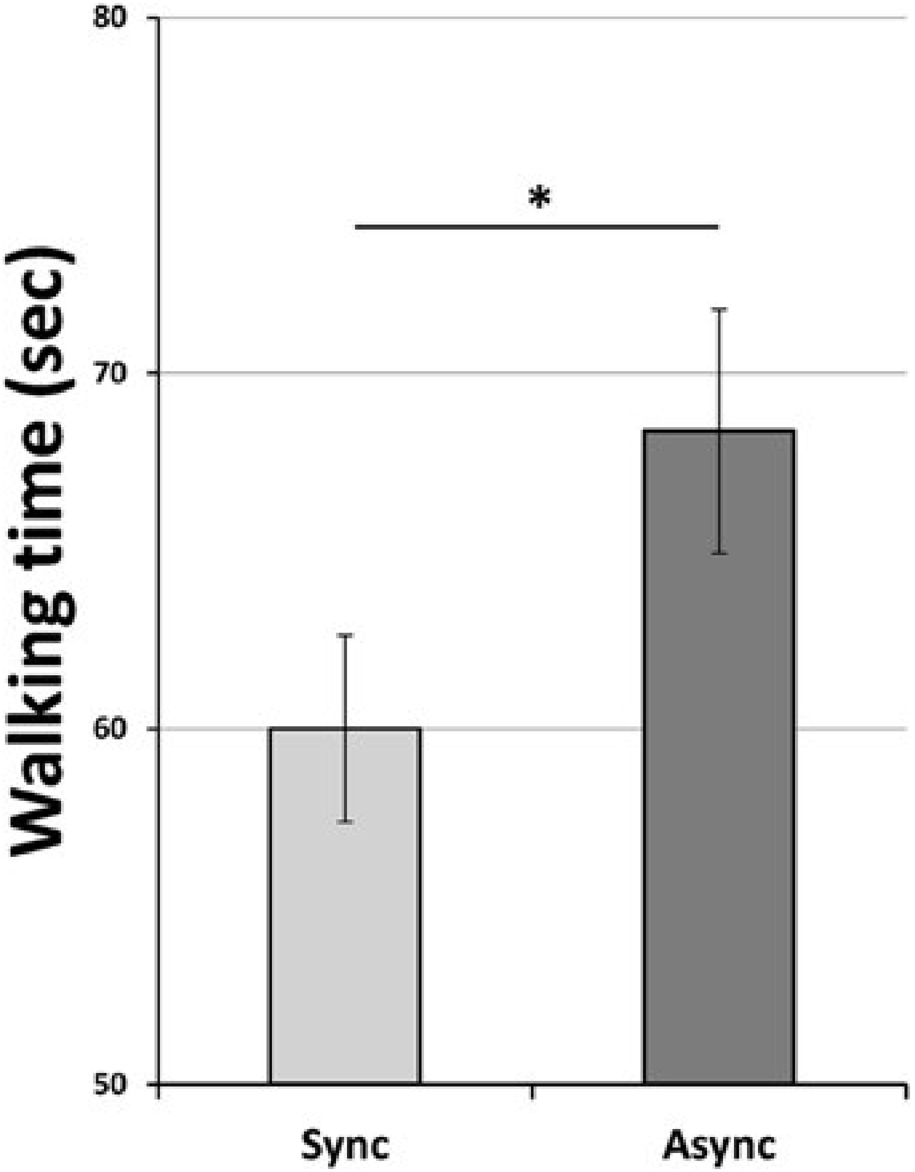


Figure 2. (A) A sequence of images depicting the different sub-goals (SGs). **(B)** A state-flow diagram showing the Finite State Machine (FSM). Yellow arrows represent transitions initiated by the user while green arrows represent transitions initiated by the robot.

Tidoni, E., Gergondet, P., Kheddar, A. and Aglioti, S.M., 2014. Audio-visual feedback improves the BCI performance in the navigational control of a humanoid robot. *Frontiers in neurorobotics*, *8*, p.20.



Results:

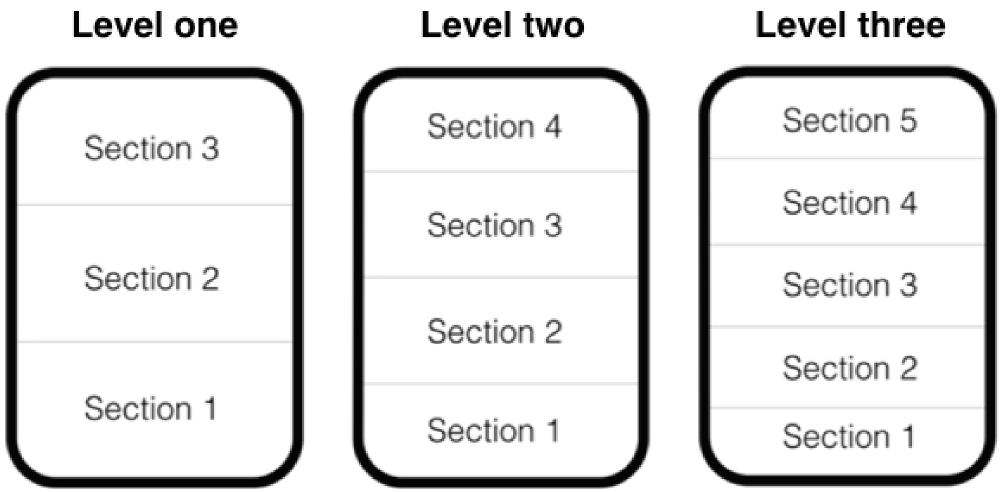
Found audio-visual synchrony information was important to the BCI user in effective robot control.

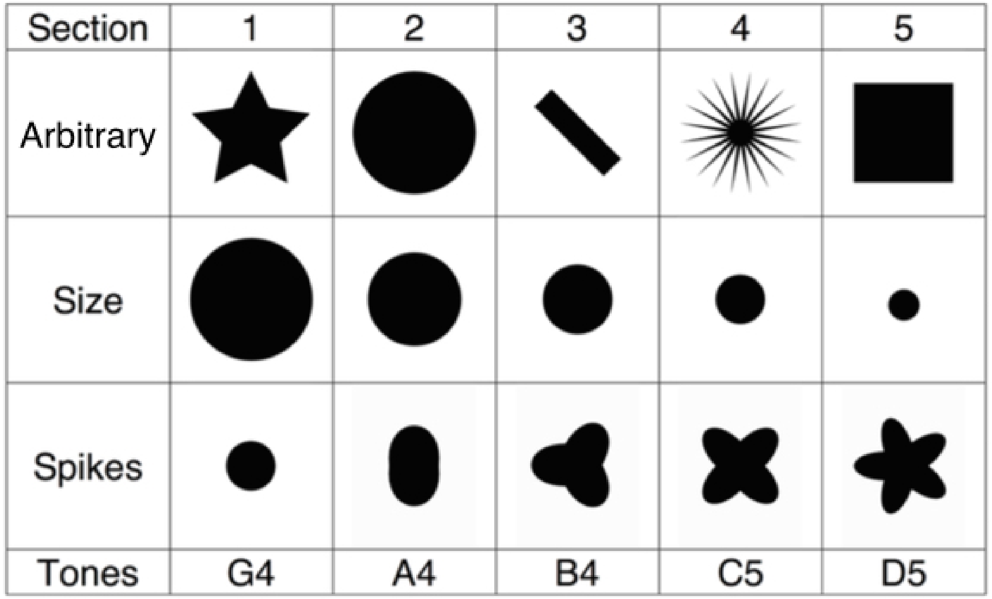
Figure 5. Mean walking time to drive the robot from the first to the second table and drop the bottle. Light-grey and dark-grey columns represent Synchronous and Asynchronous footstep sounds, respectively, heard by participants. Error bars represent s.e.m. Asterisk indicate significant comparisons ($p < 0.05$).

Metatla, O., Correia, N.N., Martin, F., Bryan-Kinns, N. and Stockman, T., 2016, May. Tap the ShapeTones: Exploring the effects of crossmodal congruence in an audio-visual interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1055-1066).

“**Abstract:** There is growing interest in the application of crossmodal perception to **interface design**. However, most research has focused on task performance measures and often ignored **user experience and engagement**. We present an examination of **crossmodal congruence** in terms of performance and engagement in the context of a **memory task** of audio, visual, and audio-visual stimuli.
Participants in a first study showed improved performance when using a visual congruent mapping that was cancelled by the addition of audio to the baseline conditions, and a subjective preference for the audio-visual stimulus that was not reflected in the objective data.
Based on these findings, we designed an **audio-visual memory game** to examine the effects of crossmodal congruence on user experience and engagement.
Results showed **higher engagement levels with congruent displays** with some reported preference for potential challenge and enjoyment that an incongruent display may support, particularly for increased task complexity.”

Metatla, O., Correia, N.N., Martin, F., Bryan-Kinns, N. and Stockman, T., 2016, May. Tap the ShapeTones: Exploring the effects of crossmodal congruence in an audio-visual interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1055-1066).





How can crossmodal perception be applied to interface design.

Created an audio-visual memory game.

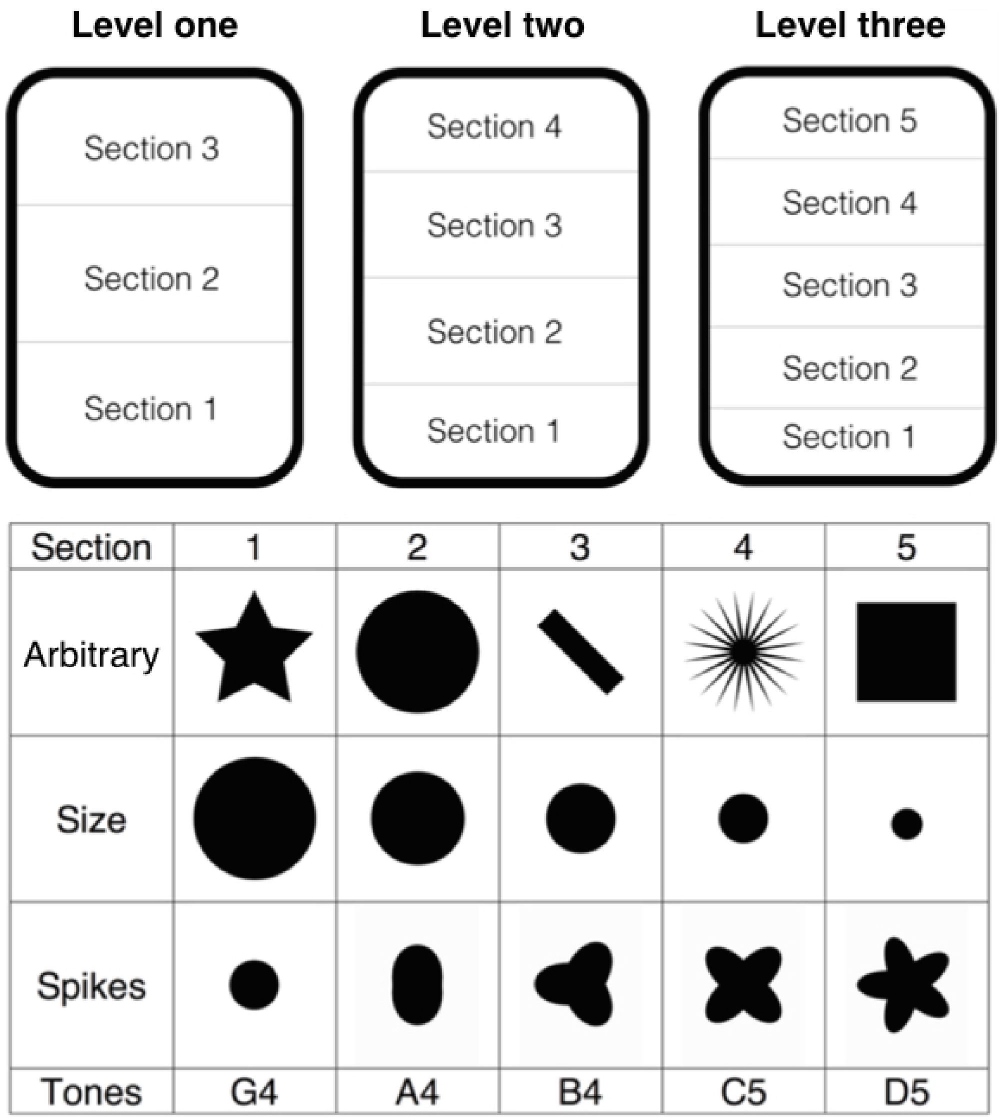
3 types of visuals “arbitrary”, “size”, “spikes”.

Spikes- basic circular shape and different number of spikes. Number of spikes corresponds to section number (e.g., spiked visual for section 3) for congruent mapping.

Size- single shape and varied its size to correspond to each screen section number (e.g., lower number sections correspond to larger objects). Since compared to spike-mapping, exact mapping from a given size to a section has to be inferred, this mapping is *semi-congruent* with the physical layout.

Arbitrary- different shapes are assigned arbitrarily to correspond to each screen section. These shapes have no obvious relationship to the section numbers (therefore incongruent with the physical layout of the screen).

Metatla, O., Correia, N.N., Martin, F., Bryan-Kinns, N. and Stockman, T., 2016, May. Tap the ShapeTones: Exploring the effects of crossmodal congruence in an audio-visual interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1055-1066).



How can crossmodal perception be applied to interface design.

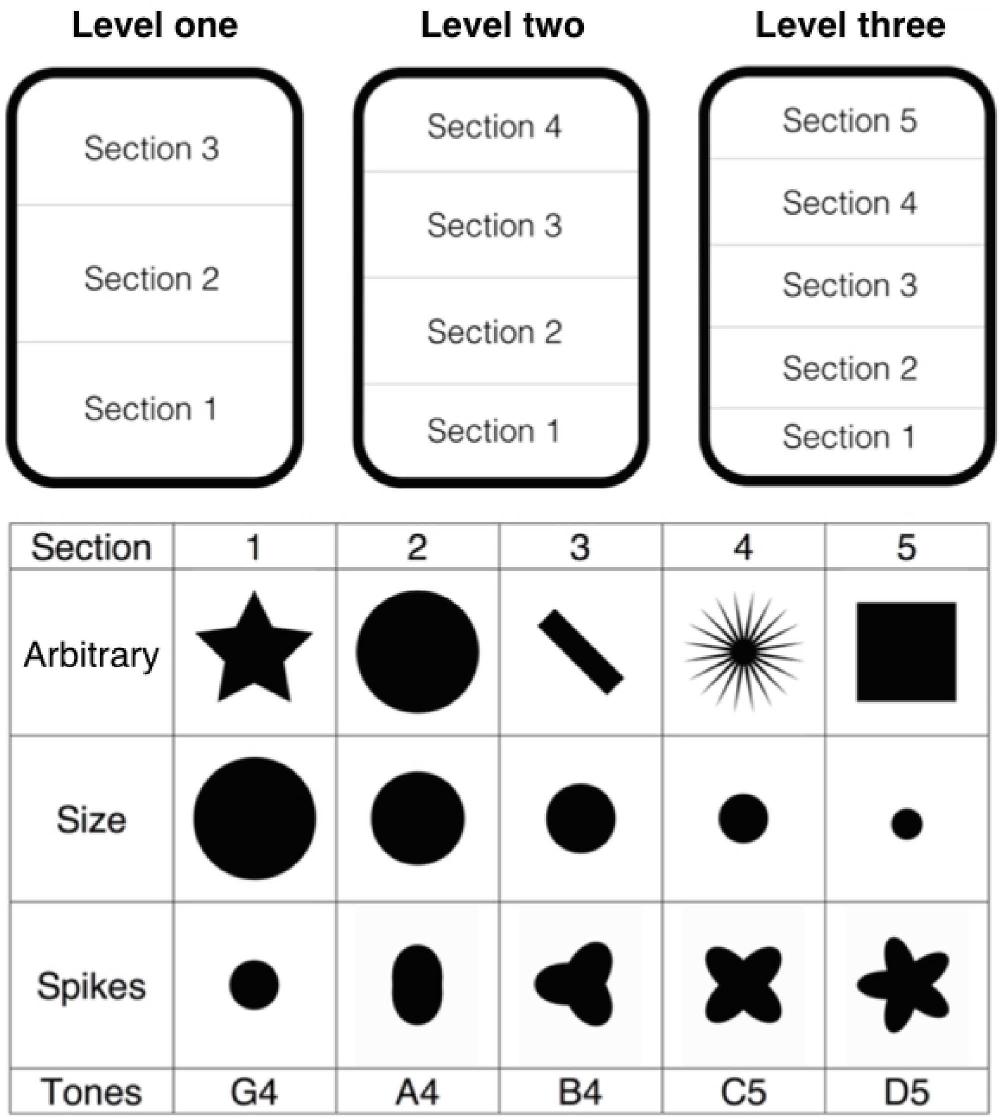
Audio

Tones were mapped vertically to screen sections:

Lower-pitched sounds corresponded to lower sections, and higher-pitched sounds corresponded to higher sections.

This mapping is based on crossmodal correspondences between vertical location and sound pitch.

Metatla, O., Correia, N.N., Martin, F., Bryan-Kinns, N. and Stockman, T., 2016, May. Tap the ShapeTones: Exploring the effects of crossmodal congruence in an audio-visual interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1055-1066).



Used touch-based input

Users interact with this application by tapping on corresponding sections on the screen to reproduce the spatial order of a sequence of items conveyed to them through the experimental programme.

Metatla, O., Correia, N.N., Martin, F., Bryan-Kinns, N. and Stockman, T., 2016, May. Tap the ShapeTones: Exploring the effects of crossmodal congruence in an audio-visual interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1055-1066).

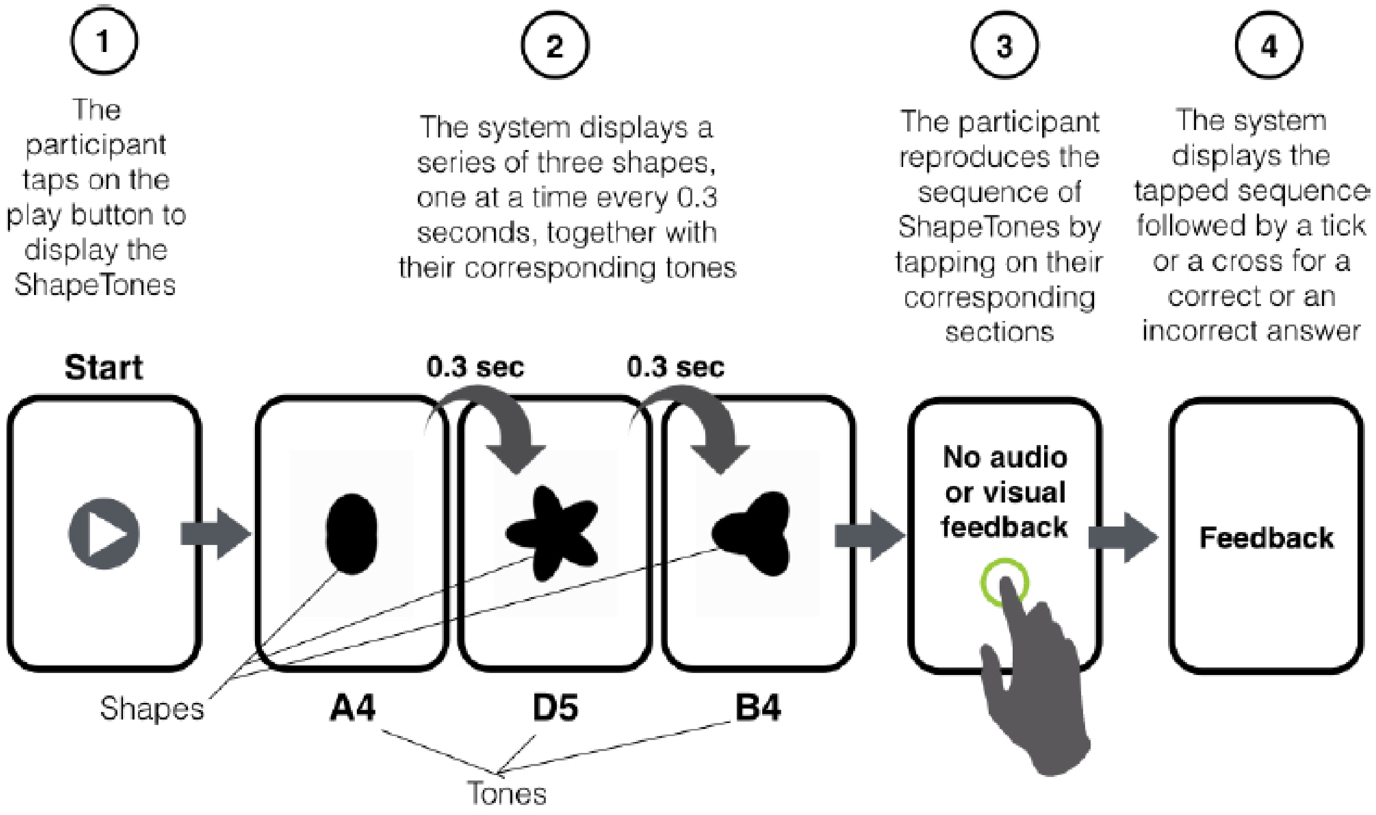


Figure 2. Experimental task

Task Design 1-4

Measured: Scores and completion times

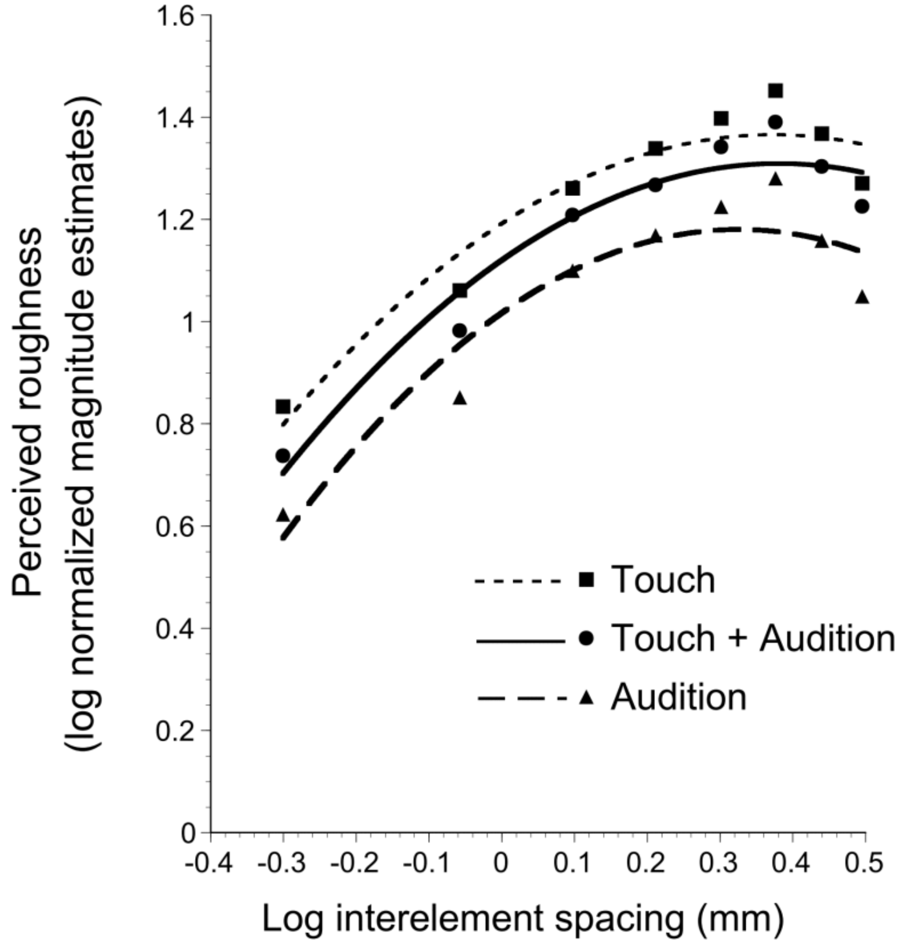
Result:

Found higher engagement with the game when there was crossmodal congruence in auditory and visual stimuli.

Lederman, S.J., Klatzky, R.L., Morgan, T. and Hamilton, C., 2002, March. Integrating multimodal information about surface texture via a probe: relative contributions of haptic and touch-produced sound sources. In *Proceedings 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. HAPTICS 2002* (pp. 97-104). IEEE.

Abstract: We experimentally assessed the relative contributions of **tactile and auditory** information to multisensory (i.e., bimodal) judgments of **surface roughness using a rigid probe**.
Participants judged the **magnitude of surface roughness** and their corresponding **confidence** in three modality conditions: touch-only, audition-only (i.e., touch-produced sounds only) and touch+audition.
The results indicated that **touch cues were weighted 62% and auditory cues 38% in the bimodal judgments**. Participants also proved to be **more confident of their judgments in the bimodal condition**. Implications for the creation of virtual roughness presented through uni- vs. multimodal interfaces is also addressed.

Lederman, S.J., Klatzky, R.L., Morgan, T. and Hamilton, C., 2002, March. Integrating multimodal information about surface texture via a probe: relative contributions of haptic and touch-produced sound sources. In *Proceedings 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. HAPTICS 2002* (pp. 97-104). IEEE.

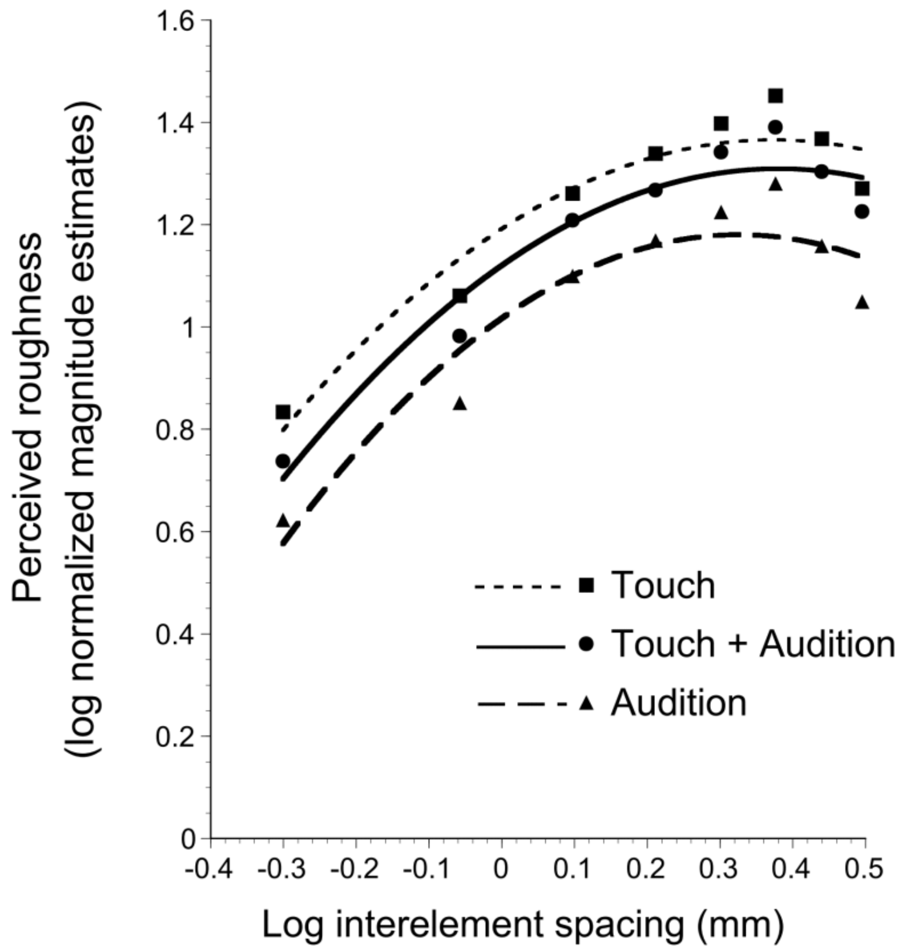


Lederman et al., (2002) Investigated relative contributions of tactile and auditory information to judgments of surface roughness using a rigid probe.

Participants judged the magnitude of surface roughness and their corresponding confidence in three modality conditions: touch-only, audition-only (i.e., touch-produced sounds only) and touch+audition.

Lederman, S.J., Klatzky, R.L., Morgan, T. and Hamilton, C., 2002, March. Integrating multimodal information about surface texture via a probe: relative contributions of haptic and touch-produced sound sources. In *Proceedings 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. HAPTICS 2002* (pp. 97-104). IEEE.

Lederman, S.J., Klatzky, R.L., Morgan, T. and Hamilton, C., 2002, March. Integrating multimodal information about surface texture via a probe: relative contributions of haptic and touch-produced sound sources. In *Proceedings 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. HAPTICS 2002* (pp. 97-104). IEEE.



Results indicated that touch cues were weighted 62% and auditory cues 38% in the bimodal judgments. Participants also proved to be more confident of their judgments in the bimodal condition.

Implications for the creation of virtual roughness presented through uni- vs. multimodal interfaces.

Lederman, S.J., Klatzky, R.L., Morgan, T. and Hamilton, C., 2002, March. Integrating multimodal information about surface texture via a probe: relative contributions of haptic and touch-produced sound sources. In *Proceedings 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. HAPTICS 2002* (pp. 97-104). IEEE.

Holloman, A.K. and Crawford, C.S., 2022. Defining scents: a systematic literature review of olfactory-based computing systems. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *18*(1), pp.1-22.

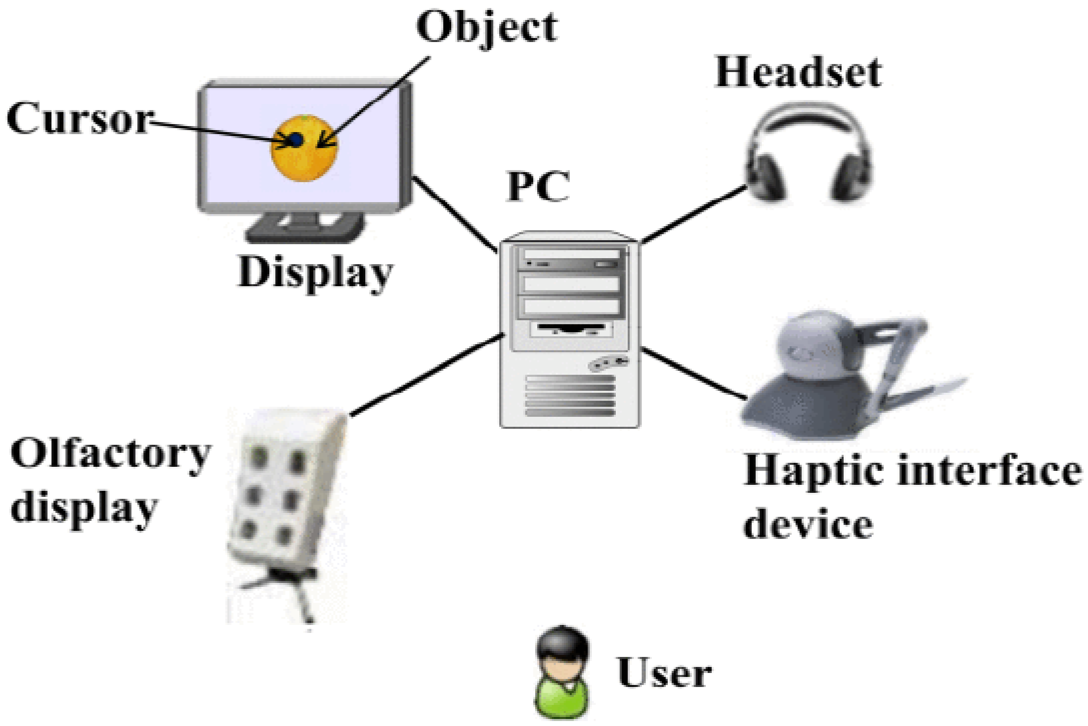
This article presents a systematic review of pertinent literature that investigates olfactory-based computing (OBC) systems in the field of Human-Computer Interaction.

Fig. 1. (A) Wearable olfactory display designed for daily activities. (B) A low-cost olfactory display designed for VR . (C) Interactive olfactory displays for art exhibit.

Sithu, M. and Ishibashi, Y., 2017, June. Identification of 3D objects with haptic, olfactory, and auditory senses in virtual environment. In *2017 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)* (pp. 235-236). IEEE.

“**Abstract:** In this paper, we objectively investigate how much correctly **3D objects can be identified** with **haptic, olfactory and auditory** senses in a **virtual environment**. We use 16 objects, which are different from each other in shape and softness. Some of the objects have sounds or smells. Assessment results show that how much largely the **identification accuracy** can be achieved with the help of haptic, olfactory, and auditory senses in the virtual environment.”

Sithu, M. and Ishibashi, Y., 2017, June. Identification of 3D objects with haptic, olfactory, and auditory senses in virtual environment. In *2017 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)* (pp. 235-236). IEEE.

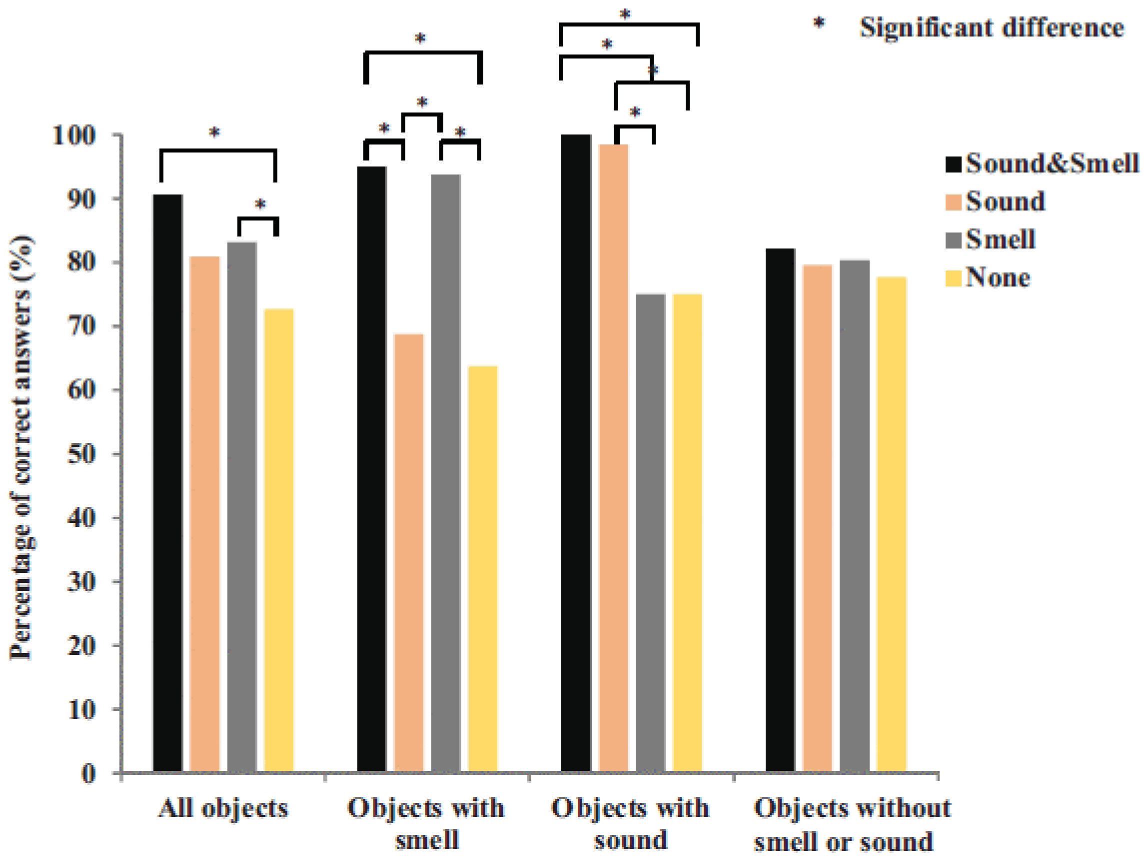


System configuration is shown in Fig. 1, where a PC has a display, a haptic interface device (Geomagic Touch), a headset, and an olfactory display (SyP@D2).

User operates haptic interface device to move a cursor (represents tip of a stylus via which the user can touch an object) in a 3D virtual space where an object is located at the centre.

16 objects. Objects differ from one another in shape and softness; some objects have sounds or smells. When user touches an object with haptic interface device, user perceives the reaction force through the device. E.g., If the object is soft user feels softness of the object when user touches it.
If an object has a sound user hears a sound via the headset when user taps object's surface.
If an object has a smell, user perceives object smell via olfactory display .
System has two modes; the *visible* (user can see each object in the virtual space) and *invisible* (user cannot see the object although it exists in the virtual space).

Sithu, M. and Ishibashi, Y., 2017, June. Identification of 3D objects with haptic, olfactory, and auditory senses in virtual environment. In *2017 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)* (pp. 235-236). IEEE.



In all the categories of the figure, we see that case Smell&Sound has the highest percentages and case None has the lowest.

Presenting combined olfactory and auditory cues in a virtual environment improves object identification accuracy.

Application to design of networked multimedia applications for collaborative work

System configuration is shown in Fig. 1, where a PC has a display, a haptic interface device (Geomagic Touch), a headset, and an olfactory display (SyP@D2).

Tan, H., Zhou, Y., Tao, Q., Rosen, J. and van Dijken, S., 2021. Bioinspired multisensory neural network with crossmodal integration and recognition. *Nature communications*, *12*(1), p.1120.

Abstract: The integration **and interaction of vision, touch, hearing, smell, and taste** in the human multisensory **neural network** facilitate high-level cognitive functionalities, such as crossmodal integration, recognition, and imagination for accurate evaluation and comprehensive understanding of the multimodal world. Here, we report a **bioinspired multisensory neural network** that integrates artificial optic, afferent, auditory, and simulated olfactory and gustatory sensory nerves. With distributed multiple sensors and biomimetic hierarchical architectures, our system can not only sense, process, and memorize multimodal information, but also **fuse multisensory data** at hardware and software level. Using **crossmodal learning**, the system is capable of **crossmodally recognizing and imagining multimodal information**, such as visualizing alphabet letters upon handwritten input, recognizing multimodal visual/smell/taste information or imagining a never-seen picture when hearing its description. Our multisensory neural network provides a promising approach towards **robotic sensing and perception.**

Tan, H., Zhou, Y., Tao, Q., Rosen, J. and van Dijken, S., 2021. Bioinspired multisensory neural network with crossmodal integration and recognition. *Nature communications*, *12*(1), p.1120.

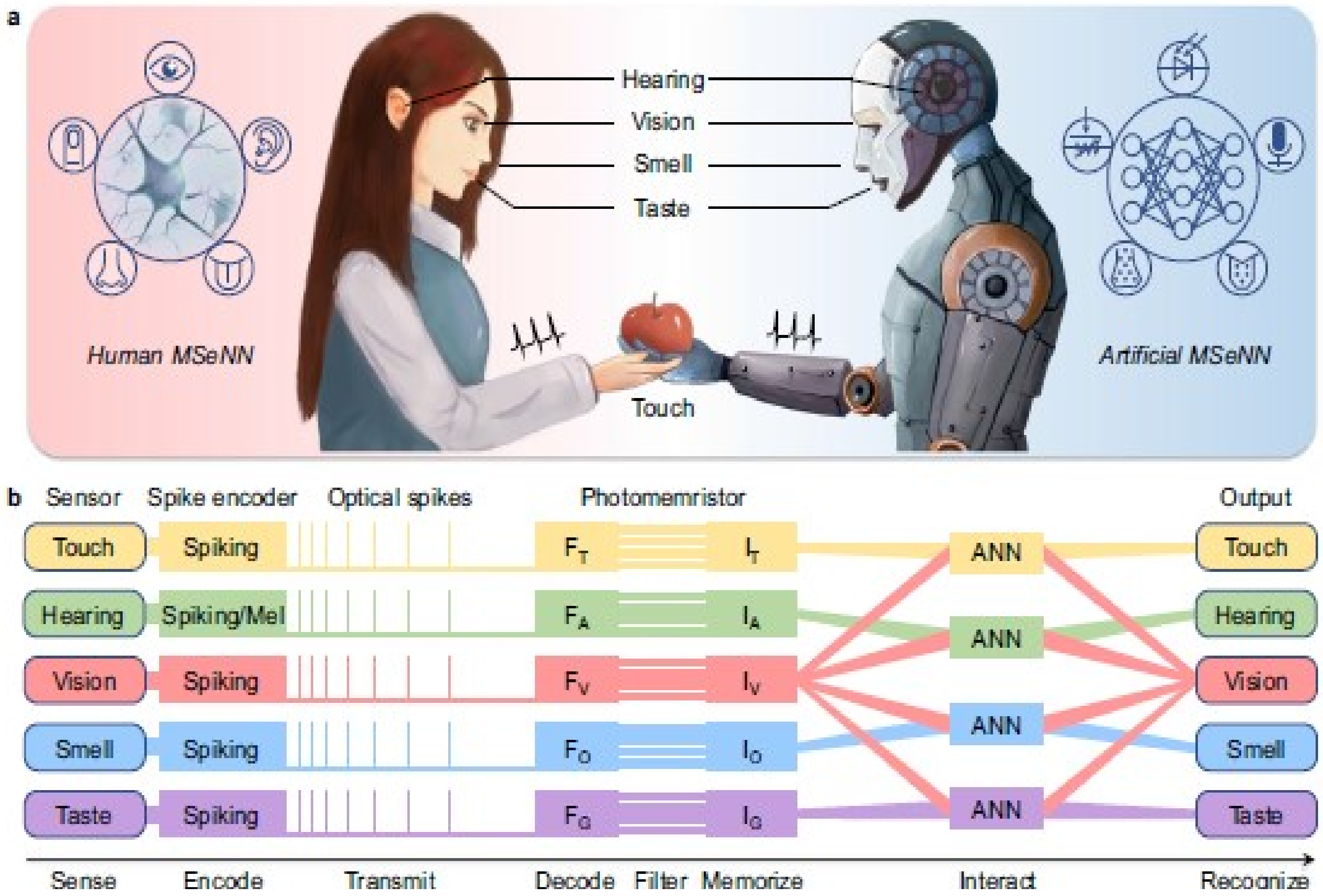


Fig. 1 Schematic of the human and artificial MSeNN. a Inspired by the five primary sensory systems (vision, touch, hearing, smell, taste) in the human MSeNN and their interaction via neural networks, the artificial MSeNN consists of five artificial sensory systems and their integration via ANNs. **b** Operational diagram of the artificial MSeNN. Sensors (photodetectors, pressure sensors, sound detectors, and simulated smell and taste receptors) convert external stimuli to potentials. Spike encoders encode potentials into optical spikes for communication. The transmitted information is decoded, filtered, and memorized by photomemristors, and the signals are crossmodally integrated and associated by ANNs for crossmodal recognition and imagination.

Present a bioinspired spiking multisensory neural network (MSeNN) that integrates artificial vision, touch, hearing, and simulated smell and taste senses with crossmodal learning via artificial neural networks (ANNs).

Resources

Essential:

Ahrens, A., Lund, K.D., Marschall, M. and Dau, T., 2019. Sound source localization with varying amount of visual information in virtual reality. *PloS one*, *14*(3), p.e0214603.

Tidoni, E., Gergondet, P., Kheddar, A. and Aglioti, S.M., 2014. Audio-visual feedback improves the BCI performance in the navigational control of a humanoid robot. *Frontiers in neurorobotics*, *8*, p.20.

Metatla, O., Correia, N.N., Martin, F., Bryan-Kinns, N. and Stockman, T., 2016, May. Tap the ShapeTones: Exploring the effects of crossmodal congruence in an audio-visual interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1055-1066).

Lederman, S.J., Klatzky, R.L., Morgan, T. and Hamilton, C., 2002, March. Integrating multimodal information about surface texture via a probe: relative contributions of haptic and touch-produced sound sources. In *Proceedings 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. HAPTICS 2002* (pp. 97-104). IEEE.

Holloman, A.K. and Crawford, C.S., 2022. Defining scents: a systematic literature review of olfactory-based computing systems. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *18*(1), pp.1-22.

Sithu, M. and Ishibashi, Y., 2017, June. Identification of 3D objects with haptic, olfactory, and auditory senses in virtual environment. In *2017 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)* (pp. 235-236). IEEE.

Tan, H., Zhou, Y., Tao, Q., Rosen, J. and van Dijken, S., 2021. Bioinspired multisensory neural network with crossmodal integration and recognition. *Nature communications*, *12*(1), p.1120.