# Auditory Display for Improving Free-hand Gesture Interaction

David Black[1,2,3], Bastian Ganze[4], Julian Hettig[4], Christian Hansen[4],

Medical Image Computing, University of Bremen, Germany[1]

Fraunhofer MEVIS, Bremen, Germany[2]

Jacobs University, Bremen, Germany[3]

Computer-Assisted Surgery, Otto-von-Guericke University Magdeburg, Germany[4]

**Abstract**

Free-hand gesture recognition technologies allow touchless interaction with a range of applications. However, touchless interaction concepts usually only provide primary, visual feedback on a screen. The lack of secondary tactile feedback, such as that of pressing a key or clicking a mouse, in interaction with free-hand gestures is one reason that such techniques have not been adopted as a standard means of input. This work explores the use of auditory display to improve free-hand gestures. Gestures using a Leap motion controller were augmented with auditory icons and continuous, model-based sonification. Three concepts were generated and evaluated using a sphere-selection task and a video frame selection task. The user experience of the participants was evaluated using NASA TLX and QUESI questionnaires. Results show that the combination of auditory and visual display outperform both purely auditory and purely visual displays in terms of subjective workload and performance measures.

## 1 Introduction

The increasing number and complexity of novel technologies demands new and innovative interfaces. Free-hand gestures allow natural and intuitive interaction that, in contrast to common devices such as mice or keyboards, permit additional degrees of freedom using the human hand, including wrist rotation and orientation and position and orientation of the fingers to each other. Free-hand gestures are by no means a new approach – however, they have only recently become practical for end users through new developments such as the Leap Motion controller. Stern et al. (2011) note that interest and research in free-hand gestures is especially prominent in the fields of medical systems and assistance, entertainment, crisis management, and human-robot interaction.

Even though free-hand gestures are promising, there are challenges facing its application as a standard tool for human-computer interaction. In addition to the fact that gesture recognition is a challenge, even with modern systems, free-hand interaction is often not preferred by users

Veröffentlicht durch die Gesellschaft für Informatik e. V. 2017 in
M. Burghardt, R. Wimmer, C. Wolff, C. Womser-Hacker (Hrsg.):
Mensch und Computer 2017 – Tagungsband, 10.–13. September 2017, Regensburg.
 https://doi.org/10.18420/muc2017-mci-0240

(Burno, 2015). The reasons are manifold: free-hand gestures are often unfamiliar; due to a lack of standards, users must learn new gestures (Stern et al., 2006). Users are frustrated by the lack of tactile feedback, which is found in almost all conventional forms of input, including mice, keyboards, joysticks, and controllers (Lee, 2015). Tactile interaction with the aforementioned devices provides instantaneous feedback during operation. In addition, most tactile feedback methods also provide audible sounds when moved or pressed. This problem has been noted in the use of touchscreens; attempts have been made to supplement such screen with tactile (Hoggan et al., 2008) and auditory (Altinsoy et al., 2009) feedback to reduce error rates and improve satisfaction.

In this work, the problem of absent secondary feedback for free-hand gestures is tackled by using auditory display. Three concepts were developed and compared: 1) auditory display with bare-bones visual feedback, 2) solely visual feedback, and 3) combined audiovisual feedback. The evaluation should help determine whether auditory feedback leads to greater speed and accuracy during free-hand gesture interaction, but also whether this leads to improved user experience and acceptance.

Previous investigations into the augmentation of free-hand gestures with feedback primarily focus on visual feedback, tactile feedback, and auditory feedback. Studies that deal with the design of free-hand gestures usually show the user the current gesture mode (Hettig et al., 2015), a 3D model (Petry and Huber, 2015) or a live video of the hand (Wachs et al., 2008) so that the user can see the hands and assess whether the gesture was correctly given or recognized by the system. Approaches that use tactile feedback often make use of gloves or similar coverings (Sharma et al. 2014), ultrasound (Carter et al., 2013), or similar technologies to aid interaction.

Lee et al. (2015) developed a radial menu to be used with a Leap Motion controller. Participants interacted with the menu to select a menu element by pointing a finger. Various auditory feedback concepts were developed and evaluated with respects to accuracy and speed of task completion. The study used visual feedback as the fundamental modality to which the various auditory feedback methods were added. One auditory feedback method triggered when the user pointed to a menu element; another triggered shortly before the element was selected. Especially interesting was a method that employed depth: the selection of a menu element was aided by a continuous tone to relay how far the finger was from the element to be selected. This concept led to higher accuracy and speed compared to other auditory feedback methods.

The majority of research into the interplay between auditory displays and user interfaces, however, focus on interaction concept without visual feedback, so-called ‘auditory user interfaces.’ The general aim of these attempts is to support users whose visual perception is already occupied (for instance, when driving a car), or whose vision is impaired. Zhao et al. (2007) present an MP3 player that can be operated solely with a touchpad and auditory cues. The auditory display gives sound feedback by playing back names of menu items when the user moves over them with a finger. Brewster et al. (2003) undertook a study to improve interaction with wearable devices by developing auditory display concepts for both 2D and 3D interaction; both were shown to positively influence user interaction. Sodnik et al. (2008) employed auditory user interfaces as a replacement for common communications devices. Performance of their interface for selecting menu items reached that of interaction using an LCD screen while significantly improving safety while operating during car driving. Kajastilan

and Lokki (2012) found that, although interaction with auditory menus was measured to be slower than with a touchscreen, it was perceived by participants to be faster.

# 2 Method

To determine whether interaction using free-hand gestures can be improved using auditory feedback, three interaction concepts were developed for evaluation using two different evaluation tasks. For gesture recognition, the Leap Motion controller was used in combination with the Unity game development platform to produce a flexible framework for gesture development. To generate real-time sound synthesis, the PureData software environment and a pair of Logitech Z120 multimedia stereo loudspeakers were employed.

## 2.1 Hand Gesture Recognition Framework

Currently, two systems dominate the market for providing relatively simple free-hand gesture recognition: the Leap Motion controller and the Microsoft Kinect. The Leap Motion controller was chosen thanks to its ability to record up to 300 images per second and up to 0.2 mm accuracy in detecting hand and fingers. The Kinect, in contrast, was found to be unsuitable due to lower resolution and accuracy (Khoshelham & Elberink, 2012). In addition to the accuracy, the Leap Motion controller offers a mature and robust high-level API for the Unity development platform.

Because the Leap Motion controller uses infrared and can only detect the hand from one side, it is prone to occlusion. In addition, the tracking area of the Leap is limited to 25 to 600 mm from the sensor. In pilot studies performed in preparation for this evaluation, a correlation between hand size and accuracy was found. The controller exhibited noticeably lower accuracy when used by participants with smaller hands. The controller is also susceptible to interference from infrared radiation, which limits its tracking quality, for example, under sunlight or halogen lamps, necessitating a testing environment free from direct sunlight or such lamps.

The ‘capsule hand’ included in the Leap Motion framework was used to represent the hand in Unity. This model consists of white cylinders and colored spheres at joints.

## 2.2 Auditory Display Framework

PureData (Puckette, 1996) is a visual programming environment that allows real-time sound to be produced. To produce auditory feedback, so-called ‘patches’ are fed with signals or data to control synthesizer elements such as oscillators, filters, or noise sources so that the changes in underlying sent data can be transformed into changes in sound and thus be heard by the user. The Open Sound Control protocol (Wright et al., 2003) was used in a client-server scenario to send data from the hand gesture recognition environment to the PureData synthesizer in real-time.

## 2.3 Interaction Concepts

To investigate the effect of auditory feedback on free-hand gesture interaction, three concepts were developed for two tasks. In Concept A, visual feedback was kept to a minimum: all gestures were supported primarily by auditory feedback. The participant sees only the position of the hand as a sphere. In Concept B, the participant receives enriched visual feedback concerning the completion of gestures and of the task. The hand is represented as a complete cylinder-and-sphere model. In Concept C, visual and auditory feedback methods are combined. The following hypotheses were derived for these concepts. Concept A (auditory feedback) should perform the weakest of the three in both subjective as well as performance measures, because no complete hand is visible during task completion. Concept B (visual feedback) should perform better than Concept A in both subjective and performance measures, because the hand is visible and errors can be quickly corrected. Concept C should outperform both other concepts, because then hand is visible, and auditory feedback compensates for the loss of tactile feedback.

## 2.4 Study Tasks

For Task 1, participants completed an experiment based on ISO 9241-420, for which 8 spheres embedded in a 3D space are oriented in a circle and the participant must select a given sphere. The sphere to be selected was colored orange, and the previously selected spheres were colored green, see Figure 1.

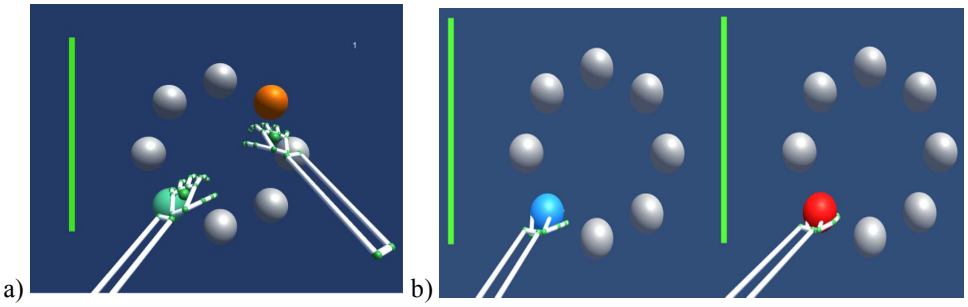


*Figure 1: Experimental task 1. A: The participant has just selected a sphere (green). The next sphere to be selected appears in orange. B: Spheres are colored blue when the hand enters, and red when incorrectly selected. The green bar represents time remaining.*

To select a sphere, participants were asked to reach into the 3D space with an outstretched hand and close their hand. If so, the sphere was colored blue. If the participant entered with the hand not outstretched, this was considered an error, in which case sphere was colored red and the participant was required to repeat the selection. Participants were asked to select as many spheres as possible with both hands within 30 seconds. The remaining amount of time was represented as a green bar on the left side of the screen.

Three feedback concepts were developed for Task 1:

**Concept A:** The participants see no representation of their hands as a 3D model, but rather as a sphere which represents their position. For audio feedback, three auditory icons were produced: a click sound when entering a sphere, an inharmonic error earcon when entering an incorrect sphere, and a harmonic earcon when selecting the correct sphere.

**Concept B:** The participants see a representation of the hands as 3D models. Spheres are colored blue when correctly selected and red when incorrectly selected.

**Concept C:** The participants receive both auditory and visual feedback from Concepts A and B, including 3D hand representation and all 3 auditory icons.

For Task 2, participants completed a more complicated assignment in which a defined image must be selected from a video. This image was shown with a red border, which should be enlarged and shifted until it fit a green border also superimposed on the image. In addition, a horizontal, transparent bar was displayed as a timeline to show current and desired image positions.

**Scroll Gesture:** To navigate through the video, the hand should be rotated 90º to the screen directly over the Leap controller. Swiping to the right advances to the next frame in the video, while swiping left advances to the previous frame. The speed of frame change depends on the angle of the hand towards the left or right.

**Shift Gesture:** To shift the image, the palm of the hand should be parallel to the monitor, help open. As soon as this gesture is recognized, the frame begins to move relative to the position of the hang.

**Zoom Gesture:** To enlarge or reduce the image, the same hand position is used as in the shift gesture, although the hand is to be in a closed fist position. Thus, the participant can ‘grab’ the image and pull out from or push into the monitor.

For each feedback concept, two test runs were completed, each with a different video. As soon as the software recognized the participants’ hands, a timer counted down from 5 to 0. After 0, the task was to be completed as quickly as possibly.

For Task 2, three feedback concepts were developed:

**Concept A:** The hand is, again, visualized as a sphere. The complexity of the task demanded a more complex auditory display. To acknowledge gesture recognition, a harmonic, 5-note ascending earcon is played back in PureData. To represent the end of a recognized gesture, the same earcon is played in descending order, from the highest to the lowest note. The frame scrolling gesture is implemented as a ‘pulse train’ for which a series of clicks are played back corresponding to each frame change. These are played back in reverse when scrolling backwards through the video. The auditory display for the shift gesture uses model-based sonification to support the participant in recognizing when images are shifted. A series of filtered noise generators gives the auditory effect of rubbing or moving a hard object over a sandy surface. The speed of gesture motion is mapped both to the variation in the speed of noise generator amplitude modulation as well as the cut-off frequency of a series of low-pass filters, so that faster gesture motions correspond to hearing faster rubbing or motion in the

model-based sonification. For the zoom gesture, model-based sonification is used again to support the participant in hearing whether the image is being enlarged or reduced. Here, an approximation of a Shepard (1064) tone is used, which gives the illusion of a never-ending increase or decrease in tone pitch. Enlargement is supported by increasing pitch, and reduction by decreasing pitch.

**Concept B:** The hand is shown as a 3D representation. The hand is colored blue for scrolling; red for shifting, and yellow for zooming (see Figure 2).

*Figure 2: Concept B (visual feedback) for Task 2. Left is the scroll gesture, middle the shift gesture, and right the zoom gesture.*

**Concept C:** In this feedback concept, the preceding two feedback concepts are combined. Thus the participant receives auditory feedback as well as a 3D model of the hands which are colored according to the current gesture.

## 2.5 Evaluation

The participants were asked whether they were colorblind or had a hearing impairment. Afterwards, their left hand was photographed for measuring its size. Participants were seated on an adjustable-height chair. For training, a test scene was produced with which participants could see their hand visualized as a 3D model. They received information concerning the operation of the Leap controller, how to avoid hand occlusion and how to stay within the tracking area of the controller. The participants were instructed using a series of slides on how to complete the two tasks and given time (ca. 10 minutes) to familiarize themselves with each feedback concept. Participants were given the opportunity to adjust volume levels to their satisfaction.

For each task, the sequence of feedback concepts A, B, and C were permutated for each participant to reduce training effects. After one feedback concept was completed, lasting 2 to

3 minutes, for each task, the participants were asked to complete both the NASA-TLX (Hart, 2006) and QUESI (Wegerich et al., 2012) questionnaires. Accuracy and performance data were gathered using the developed program. As soon as the sixth and last task was completed and the questionnaires completed, each participant was asked three questions regarding the auditory displays and acceptance.

# 3 Results

In total, 14 participants (10 M, 4 F, 20-29 yr.) completed the evaluation, none colorblind or with hearing impairment.

User experience was evaluated with two questionnaires, NASA Raw TLX for subjective workload (Hart, 2005) and QUESI for participant satisfaction (Wegerich, 2008) repeated after each of the 6 tests. For Task 1, average TLX workload scores were 34.18 for Concept A (auditory display), 35.85 for Concept B (visual display), and 28.33 for Concept C (combined feedback) on a scale of 0 to 100 where 0 is low workload and 100 is high workload. For Task 2, scores were 39.60 (A), 45.42 (B), and 39.58 (C). For the QUESI questionnaire, average Task 1 scores were 3.96 (A), 3.50 (B), and 3.75 (C), and for Task 2, 3.31 (A), 3.44 (B), and 3.75 (C), where higher scores indicate increased satisfaction on a scale of 1 to 5. Agreement to the statement, "The sound bothered me" was 1.64, on a scale of 1 to 5 where 1 indicated "completely disagree" and 5 "completely agree." For the statement "The sound helped me" agreement was 4.36.

Performance measures for Task 1 included number of spheres correctly selected, incorrectly selected spheres, and average time between selection. Of interest, the number of incorrectly selected spheres (error) averaged 1.00 (A), 1.34 (B), and 1.00 (C), and average times between selection were 1.59 s (A), 1.51 s (B), and 1.26 s (C). For Task 2, completion time was measured, resulting in average total task completion times of 48.4 s (A), 49.6 s (B), and 31.6 s (C).

Unfortunately, due to the small number of participants, typical levels of significance could not be reached, except for the comparison of Task 2 completion time of Concept C with those of Concepts A and B ($p = 0.042$) and the Task 2 error rate comparison of Concept C with Concepts A and B ($p = 0.034$).

# 4 Discussion

This work presents an investigation into the combination of auditory and visual feedback methods for two tasks: a standardized sphere-selection task and a video frame selection, enlargement, and shifting task. A user study was completed with 14 participants. Results suggest that the combination of auditory and visual feedback provide lower subjective workload, lower task completion time, and fewer errors than either solely auditory or solely visual feedback.

Free-hand gestures are not yet a standard input method for human-computer interaction. The number of degrees of freedom of the human hand and the lack of standards for gesture-based interaction methods provide substantial hindrances to its use. Designing auditory display is also not trivial; both psychoacoustic properties of human hearing as well as aesthetic concerns are burdens that prohibit a general implementation of complex auditory display as a standard means of feedback. Thanks to technological breakthroughs such as the Leap Motion controller as well as real-time sound synthesis software, the combination of both fields of research has become more approachable. However, investigations into the combination of novel gesture recognition accompanied by auditory display is sparse, but could increase in coming years through intensified focus in the field of augmented and virtual reality. Especially in medical contexts, where sterile interaction is paramount, research into gesture interaction is especially prudent.

Even though comparisons between concepts for many of the measures could not reach a typical level significance, free-hand gestures supported by auditory display appears to be a promising combination. Because of the lack of secondary tactile feedback in free-hand interaction, auditory display is essential in knowing when gestures are recognized and whether the recognized gesture is correct. Future work should especially focus on model-based auditory display in which the user receives continuous feedback for free-hand gestures. In this way, a user could 'play' the gestures similar to playing an instrument, thereby providing a finer degree of control compared to simple warning and alarm sounds that trigger when a certain action has been executed. This type of feedback was shown in Lee et al. (Lee et al., 2015) to have the highest performance, and in this work showed significant improvements in task performance time for Task 2.

# 5 Conclusion

This work describes three concepts for feedback for free-hand gesture interaction with a Leap Motion controller. Auditory, visual, and combined audiovisual feedback methods were developed to support the user in evaluating two separate screen-based tasks. Although in some cases a typical level of significance could not be reached, results show that combined audiovisual display outperforms and is preferable to purely auditory or purely visual displays. The results suggest that increased focus into the multimodal effects of auditory and visual feedback for hand gestures is warranted, especially due to the lack of secondary tactile feedback inherent in typical free-hand interaction.

# Acknowledgements

This work is partially funded by the Federal Ministry of Education and Research (BMBF) within the STIMULATE research campus (grant number 13GW0095A).

# References


Altinsoy, M. & Merchel S. (2009). Audiotactile Feedback Design for Touch Screens. In *Proceedings International Conference on Haptic and Audio Interaction Design 2009*, pp. 136-144.

Brewster, S., Lumsden, J., Bell, M., Hall, M. & Tasker, S. (2003). Multimodal 'Eyes-Free' Interaction Techniques for Wearable Devices. In *Proceedings SIGCHI Conference on Human Factors in Computing Systems 2003,* pp. 473-480.

Burno, R. (2015). *Equating User Experience and Fitts' Law in Gesture Based Input Modalities*. Tempe: Arizona State University.

Carter, T., Seah, S., Long, B., Drinkwater, B. & Subramanian, S. (2013). UltraHaptics: Multi-Point Mid-Air Haptic Feedback for Touch Surfaces. In *ACM symposium on User interface software and technology 2013*, pp. 505-514.

Hart. S. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Human Factors and Ergonomics, 50*(9), 904-908.

Hettig, J., Mewes, A,. Riabikin, O., Skalej, M., Preim, B. & Hansen, C. (2015). Exploration of 3D Medical Image Data for Interventional Radiology using Myoelectric Gesture Control. In *Visual Computing for Biology and Medicine 2015*, pp. 171-185.

Hoggan, E., Brewster, S. & Johnston, J. (2008). Investigating the Effectiveness of Tactile Feedback for Mobile Touchscreens. In *SIGCHI Conference on Human Factors in Computing Systems 2008*, pp. 1573-1582.

Kajastilan, R. & Lokki T. (2013). Eyes-free interaction with free-hand gestures and auditory menus. *Human-Computer Studies, 71*(5), 627-640.

Khoshelham, K. & Elberink. S. (2012). Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. *Sensors*, *12*(2), 1437–1454.

Lee, K., Park, Y. & Kim. J. (2015). Effects of Auditory Feedback on Menu Selection in Hand-Gesture Interfaces. *IEEE Multimedia.* PP(99), pp. 32-40.

Petry, B. & Huber, J. (2015). Towards Effective Interaction with Omni- directional Videos Using Immersive Virtual Reality Headsets. In *Augmented Human International Conference 2015*, pp. 217-218.

Puckette, M. (1996). Pure Data: another integrated computer music environment. In *Proccedings International Computer Music Conference*, pp. 37-41.

Sharma, A., Kumar, S. & Kumar, A. (2014). Haptic Feedback–A Review. In *Proceedings Recent Advances and Trends in Electrical Engineering (RATEE 2014)*, pp. 185-189.

Shepard, R. (1964). Circularity in Judgements of Relative Pitch. *Acoustical Society of America, 36*(12), 2346–2353.

Sodnik, J., Tomazic, S., Dicke, C. & Billinghurst, M. (2008). Spatial auditory interface for an embedded communication device in a car. In *Advances in Computer-Human Interaction 2008*, pp. 69-76.

Stern, H., Wachs, J. & Edan. Y. (2006). Optimal Hand Gesture Vocabulary Design Using Psycho-Physiological and Technical Factors. In *Automatic Face and Gesture Recognition, 2006*, pp. 257-262.

Wachs, J., Kölsch, M., Stern, H. & Edan, Y. (2011). Vision-Based Hand Gesture Applications: Challenges and Innovations. *Communications of the ACM, 54*(2), 60-71.

Wachs, J., Stern, H., Edan, Y., Gillam, M., Handler, J., Feied, C., & Smith. (2008). M. A Gesture-based Tool for Sterile Browsing of Radiology Images. *Am Med Inform Assoc., 15*(3), 321–323.

Wegerich, A., Löffler, D. & Maier, A. (2012) *Handbuch zur IBIS Toolbox - Evaluation Intuitiver Benutzbarkeit. Bundesministerium für Bildung und Forschung.*

Wright, M., Freed. A. & Momeni, A. OpenSound Control: state of the art 2003, In *New Interfaces for Musical Expression 2003*, pp. 153-159.

Zhao, S., Dragicevic, P., Chignel, M., Balakrishnan, R. & Baudisch, P. (2007). earPod: Eyes-free Menu Selection using Touch Input and Reactive Audio Feedback. In *SIGCHI Conference on Human Factors in Computing Systems 2007*, pp. 1395-1404.

# Authors

**Black, David**
David Black is a researcher at the University of Bremen and the Fraunhofer Institute for Medical Image Computing MEVIS in Bremen, Germany. He has a music conservatory background in classical composition and electronic music as well as expertise in usability for medical applications. He applies sound synthesis techniques to medical applications, including navigated instrument guidance, novel gesture interaction, and dataset sonification.

**Ganze, Bastian**
Bastian Ganze is a software developer from Leipzig who studied computational visualization in Magdeburg. He currently works for a startup in Düsseldorf. In his free-time he follows his passion for visualization and sound by developing computer games.

**Hettig, Julian**
Julian Hettig completed his Bachelor in Media Informatics at the RheinMain University of Applied Science in Wiesbaden. Thereafter, he completed a Master in Computer Visualization at the Otto-von-Guericke University in Magdeburg, where he currently works as a researcher in the Computer-Assisted Surgery Workgroup.

**Hansen, Christian**
Christian Hansen is a junior professor for Computer-Assisted Surgery at Otto-von-Guericke University in Magdeburg, Germany. He leads the research group “Therapy Planning and Navigation” at the research campus STIMULATE. His research interests include human-computer interaction, medical augmented reality, and medical visualization.