

Distance sonification in image-guided neurosurgery

Joseph Plazak^{1,2} ✉, Simon Drouin³, Louis Collins³, Marta Kersten-Oertel¹

¹Department of Computer Science and Software Engineering & PERFORM Centre, Concordia University, Montréal, Canada

²School of Music, Illinois Wesleyan University, Bloomington, IL, USA

³McConnell Brain Imaging Centre, Montreal Neuro, McGill University, Montréal, Canada

✉ E-mail: jplazak@iwu.edu

Published in Healthcare Technology Letters; Received on 31st July 2017; Accepted on 31st July 2017

Image-guided neurosurgery, or neuronavigation, has been used to visualise the location of a surgical probe by mapping the probe location to pre-operative models of a patient's anatomy. One common limitation of this approach is that it requires the surgeon to divert their attention away from the patient and towards the neuronavigation system. In order to improve this type of application, the authors designed a system that sonifies (i.e. provides audible feedback of) distance information between a surgical probe and the location of the anatomy of interest. A user study ($n=15$) was completed to determine the utility of sonified distance information within an existing neuronavigation platform (Intraoperative Brain Imaging System (IBIS) Neuronav). The authors' results were consistent with the idea that combining auditory distance cues with existing visual information from image-guided surgery systems may result in greater accuracy when locating specified points on a pre-operative scan, thereby potentially reducing the extent of the required surgical openings, as well as potentially increasing the precision of individual surgical tasks. Further, the authors' results were also consistent with the hypothesis that combining auditory and visual information reduces the perceived difficulty in locating a target location within a three-dimensional volume.

1. Introduction: Within neurosurgery, surgeons make frequent visual references to pre-operative patient images in order to aid in localisation of specific anatomy. Mentally mapping pre-operative images to the patient on the operating table, however, is not trivial and may be prone to error. Image-guided neurosurgery, or neuronavigation, has been used to visualise the location of a surgical probe by tracking the probe and mapping its location to pre-operative models of a patient's anatomy. However, one common limitation of this approach is that it requires the surgeon to divert their attention away from the patient and towards the neuronavigation system.

One potential solution to the limitations of visual information in the operating room (OR) involves transforming certain types of information to another sense modality, such as sounds or vibrations [1]. In particular, auditory display (i.e. data sonification) has been used in many OR applications and user studies (see [2] for a full review). Sonified data is unique from visualised data in that the ear can successfully parse multiple data streams at once, a phenomenon often referred to as the cocktail party effect [3]. Taking advantage of auditory data streams allows clinicians to reduce their dependency on visual information, thereby increasing available attention for other tasks [2] and increasing accuracy within three-dimensional (3D) localisation tasks [4]. One specific application for sonified data within image-guided surgery systems includes providing information pertaining to depth, a feature that is difficult to visualise on a 2D screen.

Despite vastly different ideas about how data should be sonified, OR sonification studies typically report similar advantages and disadvantages. For example, Parsehian *et al.* [5] utilised four different types of sonification that modified pitch, loudness, duration/tempo, and timbre within a hand guidance task, and reported similar results for each sonification type. Further, Hansen *et al.* [6] manipulated a combination of sound features, including onset frequency, tone length, and rising and falling pitch relays, within a resection line transfer task and found that the sonified data was useful for reducing diverted attention and improving the accuracy of the task. In general, many different types of sonification have been utilised within experimental studies (see [7] for a review), but basic questions still remain regarding which sonifications are best suited for a given task.

In order to examine the role of auditory feedback within image-guided neurosurgery systems, we designed an experiment utilising a

3D localisation task for measuring performance differences under auditory, visual, and audio-visual feedback. A priori, and consistent with existing literature, we hypothesised that audio information would improve the ability to locate points within a 3D volume when combined with visual information. Further, we hypothesised that tasks using audio-visual feedback would be more user-friendly than single modality feedback. To foreshadow our results, our experiment found evidence consistent with each hypothesis.

2. Methods: We developed an audio-augmented neuronavigation system that was comprised of the following elements: (i) IBIS neuronavigation software [8], (ii) custom IBIS audio plug-in software (iii) a Polaris Tracking System from Northern Digital Technologies, used for tracking, and (iv) a self-contained audio interface and powered speaker for presenting sonified distance data. A detailed description of the IBIS neuronavigation system, including the calibration, registration, and rendering is given in [9]. In order to interface the neuronavigation system with an external audio synthesiser, we developed an IBIS audio plugin capable of transmitting open sound control (OSC) messages from IBIS to the pure data audio programming environment [10]. This plug-in employed 'oscpack' [11] for handling OSC packet manipulation.

The audio portion of our system received OSC messages from IBIS, which provided real-time information on the location of the surgical probe, the location of a target point within a 3D volume of a computed tomography angiography (CTA) scan, and several other messages to help facilitate the control of the experiment (detailed below). Using this information, we calculated the real-time Euclidean distance (in three dimensions) between the tip of the pointer and a specified point within a given volume from within pure data. This distance data was then used to control five different sonifications, each of which are detailed below. Sounds were played from a self-powered speaker connected to the experiment computer at an identical volume level for all participants. Below, a brief explanation is provided on the five different types of sonifications used in our user study.

1. *Sine tone frequency matching (frequency):* This sonification utilised distance information to alter the absolute frequency of a sine

tone. As the distance decreased (i.e. as the pointer approached the target in 3D space), the frequency of the tone increased linearly from a lower limit of 130 Hz to an upper limit of 440 Hz. In order to provide a reference sound for participants, a second sine tone was played in parallel at the top limit frequency of 440 Hz. Participants were instructed that their goal was to match the frequency of the two tones, and that increases of frequency were an indication of moving closer to the desired target.

2. *Sine tone pitch mapping (pitch)*: This sonification was nearly identical to the sine tone frequency mapping sonification, with the exception that changes in distance resulted in discrete pitch changes consistent with a chromatic musical scale. The lower limit of the discrete pitch sonification was the musical tone C3 (263 Hz), whereas the upper limit was A4 (440 Hz). Therefore, there were 21 discrete pitch changes mapped in-between the lower and upper limit. Similar to the frequency mapping sonification, increases in pitch were associated with getting closer to the desired target.

3. *Pulsed tone sonification*: The pulsed tone sonification consisted of a short sine tone (400 Hz) that was pulsed at gradually faster rates as the target was approached, up until a point at which the pulses combined to form a continuous tone. This sonification has been utilised in a variety of other commercial applications, including parking assist systems. In this implementation, distance information controlled the rate at which the tone was pulsed, ranging from a slow pulse at ~55 beats/min until reaching a continuous tone as the desired target was reached.

4. *Signal-to-noise sonification (SNR)*: This sonification consisted of two sounds, white noise and a pure sine tone at 400 Hz, with the volume mixture of the two sounds being controlled by distance information. Distances equal to or greater than 600 mm resulted in the presented sound consisting only of white noise, whereas arriving at the target resulted in only hearing the pure tone. Distance change between these two limits was linear, so that at a distance of 300 mm, the presented sound consisted of 50% white noise and 50% sine tone.

5. *Binaural beat sonification (intonation)*: The binaural beats sonification consisted of two square wave tones, one of which served as a stable reference at 220 Hz, the other of which was slightly variable in frequency in order to allow the tone to be 'out-of-tune' with the reference, and therefore result in the phenomenon known as 'binaural beating'. This phenomenon is one way in which musicians are able to tune their instruments. The variable square wave could be altered within 5% of the target pitch. We chose a harmonically rich sound source for this sonification in order to maximise the presence of binaural beating.

3. Procedure: In a controlled experiment, 15 participants (five female, mean age = 30.6 years, SD = 6.67 years) were tasked with navigating a tracked surgical tool to randomly placed 3D virtual locations while being aided by visual feedback, aural feedback, or both visual and aural feedback; the 3D virtual locations randomly changed with every trial. Visual feedback consisted of a yellow point being placed on a vessel of a 3D CTA volume, along with a real-time mapping of the surgical pointer's tip onto the same surgical volume (a photograph of the experimental setup can be seen in Fig. 1). Audio feedback consisted of the five sonifications described in the section above. Eleven of the participants reported that they had some familiarity with image-guided neurosurgery systems, and four of the participants self-identified as either amateur or professional musicians.

Participants were allowed to either sit or stand for the experiment. Before data collection began, the experimenter explained the general purpose of the study and gave a demonstration of the audio feedback that would be presented within the experiment trials. Thereafter, each participant was given an opportunity to explore all five sonification types used in the study, and to ask

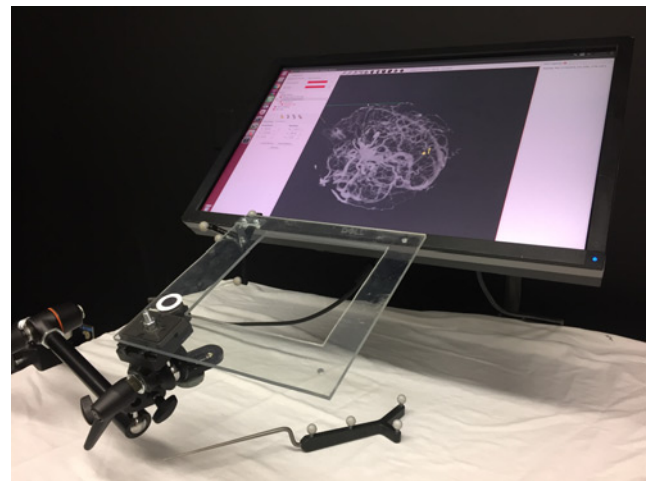


Fig. 1 Photograph of the experimental setup. For audio-visual and visual-only trials, participants navigated to a randomly placed yellow dot within a 3D volume. For audio-only conditions, the screen was blank and participants navigated to the desired point within the 3D volume via sound cues alone (sound amplifier and speaker not shown). Participants started each trial by placing the tip of the pointer within the white circle on the lower left side of the frame

any clarifying questions regarding each sound. In general, participants spent roughly 1 min exploring each of the five sonifications.

Participants were told that their goal was to navigate the surgical pointer to randomly placed target points within the CTA volume as quickly possible. Each trial lasted 15 s, and if the target was found prior to the end of the trial, the participant was asked to hold the probe steady until the trial was complete. There were a total of 11 experimental conditions: five audio-only conditions (one for each of the five sonifications), five audio-visual conditions, and one visual condition. Each participant completed two trials within each condition for a total of 22 trials per participant. The participant controlled the start of each trial by placing the tip of the pointer on small sticker on the lower left side of the frame; they were encouraged to take as much time as they needed between trials. The experiment was controlled by a small custom IBIS plug-in, which ensured that all conditions presented to the participants were randomised.

After completion of the study, the participants were asked to complete a short survey regarding their experience and the perceived utility of each sound. In total, the entire experiment lasted about 25 min, after which participants were debriefed and thanked for their participation.

4. Results: We used multiple metrics in order to explore the general utility of audio feedback within the given experimental context. Specifically, we analysed data pertaining to the average total distance between the pointer and the 3D virtual location across each 15 s trial, average end of trial distance for each trial, and also subjective ratings on the experimental conditions and the different types of sonifications. Further, in response to a feedback from participants, we also separately analysed the average z-dimension distance of pointer movement data from each trial.

Time series data was recorded during each trial. The sampling rate of this data was, on average, around 20 Hz, however, the number of data points recorded during each trial was variable. The primary cause of this variability was due to the surgical probe leaving the camera's view, which caused data recording to cease. Another source of variability was the refresh rate of IBIS, which is variable depending on workload. In order to handle this variability, each data point was time stamped, and these timestamps were used to reassemble participant's data from every trial. In some cases, data collection started before the beginning of each trial,

therefore, we trimmed each time series to include only the last 15 s of every trial.

Average distance: Our main hypothesis stated that adding audio information to visual displays would improve performance within a 3D navigation task. There are many metrics that might be used to quantify ‘improved performance’; for the present analysis, we utilised the average total distance within each trial. For each 15 s trial, we calculated the average total distance by summing the distance values between the surgical pointer and the randomly placed 3D virtual location, and then dividing by the number of samples recorded within the trial. These average trial distances were then used as data points to investigate the average trial accuracy as a function of both stimulus modality (audio, visual, and audio–visual) and sonification type.

The average distance data were utilised within a repeated measured analysis of variance (ANOVA) that investigated the effect of modality (audio, visual, and audio–visual) on accuracy. The results showed a significant effect for condition, $F(2)=35.30$, $p<0.001$; the means for each condition were: audio – 51.18 mm, visual – 40.26 mm, and audio–visual – 20.53 mm. A Tukey post-hoc test found significant differences between the audio and audio–visual condition ($p<0.001$) as well as the visual and audio–visual conditions ($p=0.006$), which is consistent with the hypothesis that audio–visual feedback resulted in greater accuracy relative to visual-only feedback. The between category differences for audio trials and visual trials failed to reach statistical significance ($p=0.20$). These results are plotted in Fig. 2. Fig. 3 presents averaged trial data for all 15 participants as a function of stimulus modality; it can be seen that audio–visual trials, on average, were more efficient than the other two modalities throughout each 15 s trial.

We also analysed the average trial distances for the different sonification types used in the study. For this analysis, we used trial data from both the audio–visual trials and the audio-only trials. Using a repeated measured ANOVA, the analysis revealed a significant effect for sonification type on accuracy, $F(4)=4.10$, $p=0.003$; the means for each condition were: binaural – 43.27 mm, pulsed tone – 38.50 mm, fixed pitch – 42.19, SNR – 20.83 mm, and continuous pitch – 34.86 mm. Post-hoc analyses revealed that the only significant differences were between the SNR sonification and the

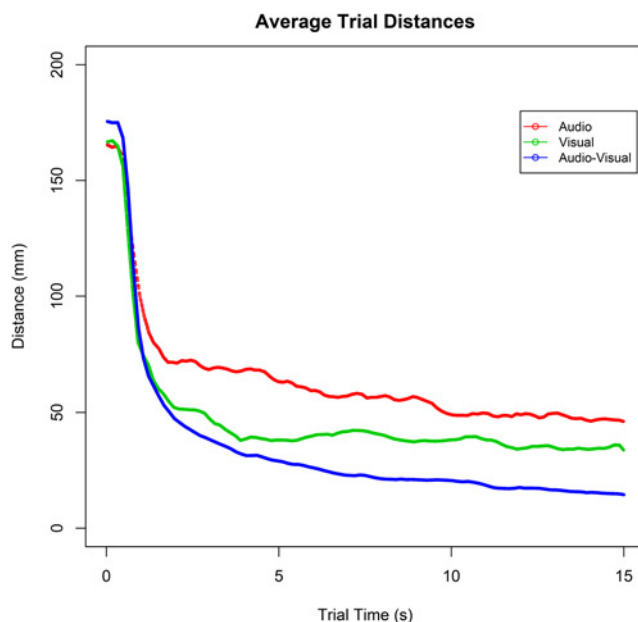


Fig. 3 Average trial distance as a function of stimulus modality and time. The red line represents the average distance (across all participants) between probe and target location for the audio-only trials. The green and blue lines represent the average distances for the visual-only and audio–visual only trials, respectively

binaural ($p=0.004$), pulsed tone ($p=0.043$) and fixed pitch ($p=0.007$) sonifications; in each case, the SNR sonification resulted in greater accuracy. These results are plotted in Fig. 4. Fig. 5 presents averaged trial data for all 15 participants as a function of sonification type; it can be seen that the SNR sonification resulted, on average, in greater accuracy for all but one of the other sonification types.

End distance to target: In addition to averaged trial data, another metric for analysing task accuracy includes the pointer’s distance to the target at the very end of the trial. While this measurement

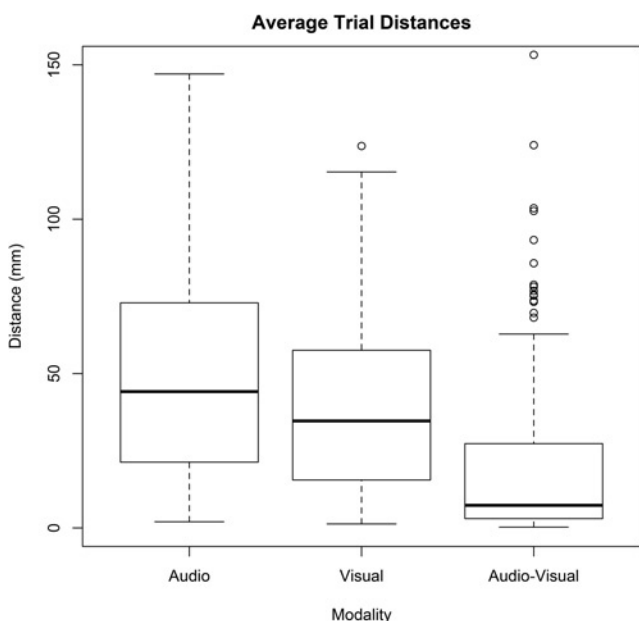


Fig. 2 Average total distance between the pointer and 3D target as a function of stimulus modality. The differences between audio–visual and audio-only conditions, as well as audio–visual and visual-only conditions, were statistically significant

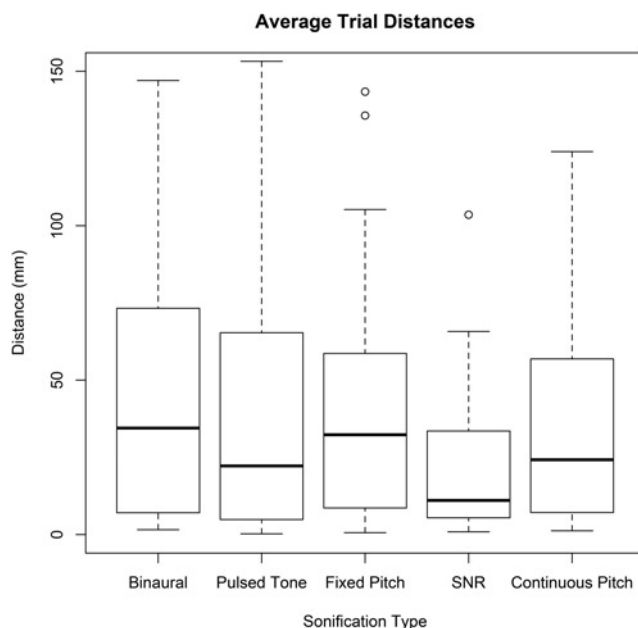


Fig. 4 Average total distance between pointer and the 3D virtual target as a function of sonification type. The SNR sonification resulted in significantly greater accuracy than the binaural, pulsed tone, and fixed pitch sonifications

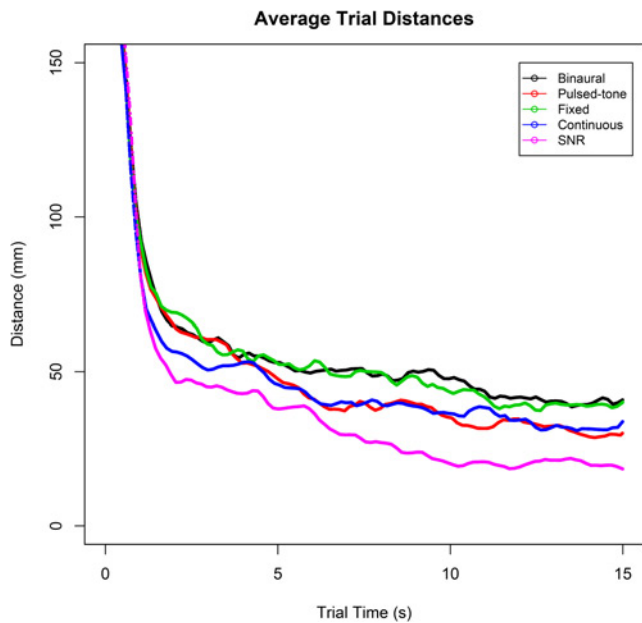


Fig. 5 Average trial distance as a function of sonification type and time. The black line represents the average distance (across all participants) between probe and target location for all binaural sonification trials (i.e. audio-only and audio-visual trials). The red, green, blue, and magenta lines represent the average distances for the pulsed-tone, fixed pitch, continuous pitch, and SNR trials, respectively

may be prone to noise, it is included here to gain further insight into the accuracy of each experimental condition. We again ran a repeated measures ANOVA and found a significant effect of modality on the average end distance to target, $F(2) = 35.79$, $p < 0.001$; the means for each condition were: audio – 51.18 mm, visual – 40.26 mm, audio-visual – 20.41 mm. A Tukey post-hoc test found significant differences between the audio and audio-visual condition ($p < 0.001$) as well as the visual and audio-visual conditions ($p = 0.005$), which is consistent with the hypothesis that audio-visual feedback resulted in greater accuracy relative to visual-only feedback. The average end distance difference between audio trials and visual trials failed to reach statistical significance ($p = 0.18$).

Using a similar approach, we also analysed the average end distances for the different sonification types used in the study. Again, using a repeated measures ANOVA, we found a significant effect of sonification type on the average end distance to the target, $F(4) = 2.735$, $p = 0.022$, with the mean distance for each type being: binaural/intonation – 42.58 mm, fixed frequency – 34.86 mm, continuous pitch – 42.2 mm, pulsed tones – 38.5 mm, and SNR – 20.83 mm. Post-hoc pairwise comparisons found only two significant differences, which included the average end distance between SNR and intonation sonifications ($p = 0.024$), as well as SNR and pitch sonifications ($p = 0.029$). In each case, the SNR sonification resulted in better accuracy.

Questionnaire responses: After the experiment, each participant completed a short questionnaire regarding the types of feedback that they found to be most useful. Participants used a 7-point Likert scale to rate the difficulty of the three conditions, and also the utility of the five different types of sonifications. A repeated measures ANOVA found that condition had a significant effect on perceived effort required for the task (i.e. ‘easiness’), $F(2) = 26.198$, $p < 0.001$. A Tukey post-hoc analysis found significant differences for the average perceived easiness score between the visual and audio-visual conditions ($p < 0.001$), as well as the audio and audio-visual conditions ($p < 0.001$), which is consistent with our second hypothesis that audio-visual feedback (relative to visual-only feedback) would result in making the task easier for participants (Fig. 6).

A separate repeated measures ANOVA found a significant effect of sonification type on perceived utility ratings, $F(4) = 3.351$, $p = 0.016$. Post-hoc pairwise comparisons did not reveal any significant differences between the five sonification types, but two comparisons were noted as being marginally significant: the pulsed sine tone sonication was rated marginally higher (i.e. more useful) than the frequency sonication ($p = 0.062$), and the SNR sonication was rated marginally higher than the frequency sonication ($p = 0.062$).

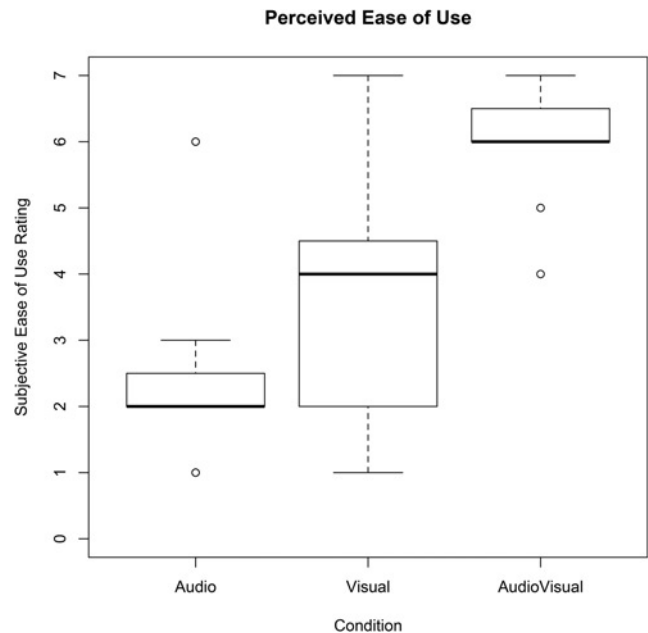


Fig. 6 Subjective ease of use/utility as a function of stimulus modality. Audiovisual feedback resulted in the task being rated as being significantly easier relative to receiving visual-only and audio-only feedback

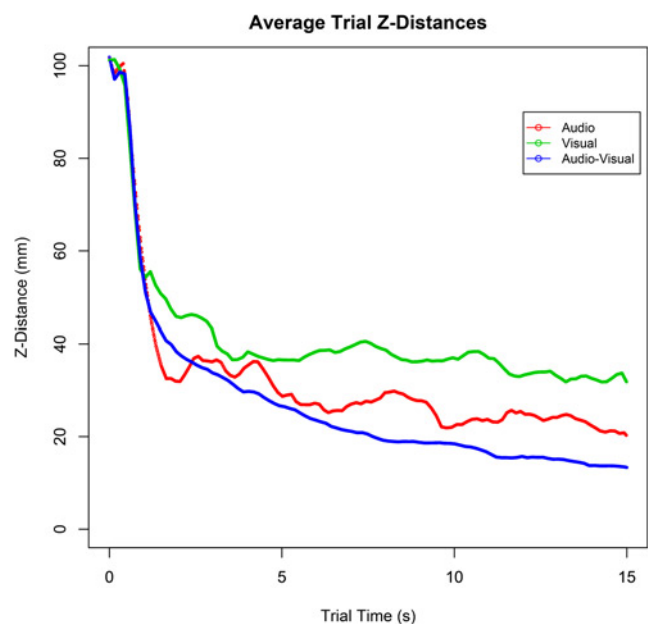


Fig. 7 Average trial z-dimension distance as a function of stimulus modality and time. The red line represents the average z-dimension distance (across all participants) between probe and target location for the audio-only trials. The green and blue lines represent the average distances for the visual-only and audio-visual only trials, respectively

Z-dimension (depth) analyses: In response to a comment made during a debriefing session, in which a participant stated that the audio feedback allowed them to ‘hear depth’, we elected to analyse the *z*-component of the average trial distance data separately. Specifically, we were interested in knowing whether audio–visual feedback resulted in greater *z*-dimension (depth) accuracy relative to visual-only and audio-only conditions. Following the same method described above for calculating average trial distance, we calculate the average *z*-dimension Euclidean distance for each trial. Using a repeated-measured ANOVA, we found a significant effect of modality on average trial *z*-distance, $F(2)=3.226$, $p=0.04$; the means for each condition were: audio – 41.49 mm, visual – 44.17 mm, and audio–visual – 33.76 mm. Despite the significant main effect, a Tukey post-hoc test found no significant differences between groups. However, the results were skewed in the predicted direction (i.e. audio–visual *z*-dimension average distances trended lower than both audio-only and visual-only averages), as can be seen in Fig. 7.

5. Conclusions: Consistent with previous work on auditory feedback within image-guided surgery, our findings present broad evidence that combining auditory distance cues with existing visual information may result in greater accuracy when locating a given target in a 3D volume. Our results are also consistent with the hypothesis that combining auditory and visual information reduces the perceived difficulty in locating a target within a 3D volume. Of the five types of sonifications employed within this study, there was evidence that the ‘signal to noise’ sonification may be better for increasing target localisation accuracy.

There were several limitations of this study that deserve special mention. In particular, because this work was an exploratory study, we made no effort to correct for multiple tests in our analyses. Next, the time series data that was collected in this study could potentially serve as a rich source of data for many questions pertaining to 3D navigation, and only a small subset of the possible analyses on these data have been presented. Future work may wish to further investigate not only the moment the pointer arrives within a target threshold, but also the different strategies that can be used to navigate within a 3D volume by using aural, visual, or multi-modal cues [5]. Finally, in seeking to understand and develop a system for expert users, our convenience sample of novice users may fail to capture some of the nuances involved in a 3D target localisation task. Feedback from one expert user (a neurovascular surgeon) familiar with our system (not reported above) suggested that one of the main uses for auditory feedback may not be to aid in target localisation, but rather as a means for remaining close to a given target after it has been found.

In general, results from this experiment are useful for comparing the utility of multimodal feedback systems against existing traditional neuronavigation systems. The results also contribute to the existent literature on sonifying OR data, by suggesting that ‘signal to noise’ sonifications may prove to be useful in future work. Finally, our experiment captured both objective and subjective measurements pertaining to sonified OR data, and found concordance between the two.

6. Funding and declaration of interests: Dr. Drouin reports grants from the Canadian Institute of Health Research, Fonds Québécois de la recherche sur la nature et les technologies and the Natural Science and Engineering Research Council of Canada during the conduct of the study.

7 References

- [1] Willems P.W.A., Noordmans H.J., van Overbeeke J.J., *ET AL.*: ‘The impact of auditory feedback on neuronavigation’, *Acta Neurochirurgica*, 2005, **147**, (2), pp. 167–173
- [2] Black D., Hansen C., Nabavi A., *ET AL.*: ‘A survey of auditory display in image-guided interventions’, *Int. J. Comput. Assist. Radiol. Surg.*, 2017, pp. 1–12
- [3] Cherry E.C.: ‘Some experiments on the recognition of speech, with one and with two ears’, *J. Acoust. Soc. Am.*, 1953, **25**, (5), pp. 975–979
- [4] Bork F., Fuers B., Schneider A.K., *ET AL.*: ‘Auditory and visio-temporal distance coding for 3-dimensional perception in medical augmented reality’. *IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR)*, 2015, 2015, pp. 7–12
- [5] Parseihian G., Gondre C., Aramaki M., *ET AL.*: ‘Comparison and evaluation of sonification strategies for guidance tasks’, *IEEE Trans. Multimed.*, 2016, **18**, (4), pp. 674–686
- [6] Hansen C., Black D., Lange C., *ET AL.*: ‘Auditory support for resection guidance in navigated liver surgery’, *Int. J. Med. Robot. Comput. Assist. Surg.*, 2013, **9**, (1), pp. 36–43
- [7] Dubus G., Bresin R.: ‘A systematic review of mapping strategies for the sonification of physical quantities’, *PLoS one*, 2013, **8**, (12), pp. e82491
- [8] Drouin S., Kochanowska A., Kersten-Oertel M., *ET AL.*: ‘IBIS: an OR ready open-source platform for image-guided neurosurgery’, *Int. J. Comput. Assist. Radiol. Surg.*, 2017, **12**, (3), pp. 363–378
- [9] Drouin S., Kersten-Oertel M., Chen S.J.S., *ET AL.*: ‘A realistic test and development environment for mixed reality in neurosurgery’. *Augmented Environments for Computer Assisted Interventions*, 2012 (LNCS, **7264**), pp. 13–23
- [10] Puckette M.: ‘Pure data: another integrated computer music environment’. *Proc. of the Second Intercollege Computer Music Concerts*, 1996, pp. 37–41
- [11] Bencina R.: ‘oscpack’, November 2006 [computer software]