Sonifying the location of an object

A comparison of three methods

Bazilinskyy, Pavlo; van Haarlem, W.; Quraishi, H.; Berssenbrugge, C.; Binda, J.; de Winter, Joost

DOI
10.1016/j.ifacol.2016.10.614

Publication date
2016

Document Version
Final published version

Published in
IFAC-PapersOnLine

Citation (APA)
Bazilinskyy, P., van Haarlem, W., Quraishi, H., Berssenbrugge, C., Binda, J., & de Winter, J. (2016). Sonifying the location of an object: A comparison of three methods. In T. Sawaragi (Ed.), *IFAC-PapersOnLine: 13th IFAC Symposium on Analysis, Design, and Evaluation of Human-Machine Systems HMS 2016* (Vol. 49 - 19, pp. 531-536) https://doi.org/10.1016/j.ifacol.2016.10.614

Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-19 (2016) 531–536

Sonifying the location of an object: A comparison of three methods

Pavlo Bazilinskyy, Wessel van Haarlem, Hashim Quraishi, Coen Berssenbrugge, Jasper Binda, Joost de Winter

Department of BioMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, the Netherlands (Tel: +31152787891; e-mail: p.bazilinskyy@tudelft.nl)

Abstract: Auditory displays are promising for informing operators about hazards or objects in the environment. However, it remains to be investigated how to map distance information to a sound dimension. In this research, three sonification approaches were tested: Beep Repetition Rate (BRR) in which beep time and inter-beep time were a linear function of distance, Sound Intensity (SI) in which the digital sound volume was a linear function of distance, and Sound Fundamental Frequency (SFF) in which the sound frequency was a linear function of distance. Participants ($N = 29$) were presented with a sound by means of headphones and subsequently clicked on the screen to estimate the distance to the object with respect to the bottom of the screen (Experiment 1), or the distance and azimuth angle to the object (Experiment 2). The azimuth angle in Experiment 2 was sonified by the volume difference between the left and right ears. In an additional Experiment 3, reaction times to directional audio-visual feedback were compared with directional visual feedback. Participants performed three sessions (BRR, SI, SFF) in Experiments 1 and 2 and two sessions (visual, audio-visual) in Experiment 3, 10 trials per session. After each trial, participants received knowledge-of-results feedback. The results showed that the three proposed methods yielded an overall similar mean absolute distance error, but in Experiment 2 the error for BRR was significantly smaller than for SI. The mean absolute distance errors were significantly greater in Experiment 2 than in Experiment 1. In Experiment 3, there was no statistically significant difference in reaction time between the visual and audio-visual conditions. The results are interpreted in light of the Weber-Fechner law, and suggest that humans have the ability to accurately interpret artificial sounds on an artificial distance scale.

Keywords: road safety, driver support, auditory display, human-machine interface, driving simulator, detecting elements

1. INTRODUCTION

Auditory displays can be of value in a broad spectrum of applications, especially in situations where visual feedback is restricted, when the visual system is overburdened, or when the message is short and calls for immediate action (Stanton and Edworthy, 1999). Adding auditory feedback to a human-machine interface may shorten visual search times and reduce the workload compared to using vision only (Perrot et al., 1990; Wickens, 1984).

Usually, auditory feedback takes the form of short warning signals (Patterson, 1982; Stanton and Edworthy, 1999). For example, auditory warnings are used in blind spot monitoring and forward collision warning systems in modern cars (Bazilinskyy et al., 2015; Jamson et al., 2008).

Auditory feedback can also be used to perceptualize objects or activity in the environment, a method which is called sonification (Hermann et al., 2011). One of the earliest known applications of sonification is an optophone. The device, used by the blind, was developed in 1913; it scans text and generates time-varying chords of tones to identify letters (Capp & Picton, 2000). One of the most successful examples of sonification is the Geiger counter, in which auditory clicks are produced to represent ionization events. The Geiger counter was developed in the early 1900s, and is still used today to measure the level of radiation in the environment (Knoll, 2010). An auditory pulse-oximeter, a device similar to the Geiger-counter, was used in hospitals in the United States in 1980's. It generated a tone that varied in pitch based on the level of oxygen in patient's blood (Kramer et al., 1999). Spain et al. (2007) investigated the implications of the use of sonified feedback during a patient monitoring task. They found that a short inter-pulse time contributes to a higher level of perceived urgency.

Sonification is also useful in the field of data analysis, in which case it is sometimes called audification or auditory graphing (Flowers, 2005). During the Voyager 2 space mission, the control encountered a problem when the spacecraft was going through the rings of Saturn. The unexpected behaviour could not be explained by means of a visual analysis of the data. When the data was played through a music synthesizer, a 'machine gun' sound was heard, leading to the conclusion that the problem was caused by collisions with electromagnetically charged micrometeoroids (Barrass and Kramer, 1999; Kramer et al., 1999).

Sensory substitution of visual information may be of value in supporting persons in locomotion tasks (e.g., Hussain et al., 2014). As early as 1936, De Florez suggested that pilots of

10.1016/j.ifacol.2016.10.614

aircrafts can benefit from the support of sonified instruments in so-called "blind flying" (De Florez, 1936). Parseihian et al. (2012) studied the mapping of the sonified distance to the actual object's location, and developed a sonified device for visually impaired persons. In the automotive industry, the parking sensor of a modern car is another example of the use of sonification, where an increasingly frequent beep is emitted to indicate that the car approaches an object. Although a parking sensor is a successful demonstration of sonification, it remains to be investigated which sonification method is the most effective for conveying information about distance or the degree of hazard.

Haas and Edworthy (1996) showed that sounds producing the highest level of perceived urgency are sounds of a high beep rate, a high intensity, and high frequency. This suggests that each of these three dimensions may be intuitive for sonification purposes. A review article of 179 publications related to sonification of physical quantities concurs that pitch (frequency), loudness (e.g., volume, intensity), and duration (e.g., beep time, inter-beep time) are the most often used auditory dimensions for sonification (Dubus and Bresin, 2013). Sanders and McCormick (1987; as cited in Stanton and Edworthy, 1999) on the other hand suggested that the auditory discrimination power of humans is rather limited, and contended that humans can identify only 2 to 3 levels of sound duration, 4 to 5 levels of sound intensity (at a given frequency), and 4 to 7 levels of sound frequency. Zahorik (2002) and Loomis et al. (1998) found that participants consistently underestimated the distance in auditory distance perception tasks. Thus, more fundamental research into the topic of mapping of given auditory cues to the distance needs to be conducted.

As mentioned above, beep time, intensity, and frequency are primary sonification dimensions. The aim of this study was to investigate which of these three sonification dimensions allows a person to most accurately indicate the location of an object. Participants completed two experiments; the first experiment involved one-dimensional distance estimation, whereas the second experiment involved the localization of an object in a two-dimensional plane. The participants were presented with sounds without visual feedback, and subsequently had to click on the screen to locate the object. In an additional Experiment 3 we sought to determine whether directional auditory feedback improves reaction times compared to visual-only feedback.

2. METHOD

Apparatus. The research was conducted using a computer program created with the Unity game engine (version 4.6.1f1). Razer Electra headphones were used.

Auditory feedback. Three types of auditory feedback were tested. The first type was Beep Repetition Rate (BRR), in which the beep time was linearly related to distance with respect to the bottom of the screen. For the closest distance (bottom of the screen), the beep time and inter-beep time were 0.05 s (i.e., 10 beeps per second). For the farthest distance (top of the screen), the beep time and inter-beep time were 0.55 s (i.e., 0.91 beeps per second). BRR resembled the feedback in a parking sensor, in that it 'beeps' faster as you are closer to an object. In the BRR condition, the sound volume was 100%, and the frequency of the beeps was 460 Hz. The volume of the laptop computer was set so that 100% sound volume generated by the software was regarded as loud but not uncomfortable.

Second, we tested Sound Intensity (SI), where the volume intensity was linearly related to the distance to the object. The volume was 0% at the top of the screen and 100% at the bottom of the screen. The frequency of the sound was 460 Hz.

Third, we tested the Sound Fundamental Frequency (SFF), where the frequency of the sound was linearly related to the distance. The frequency was 1,076 Hz at the bottom of the screen and 184 Hz at the top of the screen. The volume of the sound was 100%.

Participants. Twenty-nine persons (8 females) participated in the experiment. Most participants were students and employees of Delft University of Technology, and were on average 29.6 years old (SD = 15.7 years). None of the participants had a hearing disorder or used hearing aids.

Procedure. The participants conducted three experiments in the following order: Experiment 1: Distance estimation, Experiment 2: Distance and angle estimation, and Experiment 3: Reaction time. In Experiment 1 and Experiment 2, the participants completed three sessions, each session with a different sound condition (BRR, SI, SFF). To neutralize the effects of a learning curve, we randomized the order of the three sound conditions. We did, however, have the same order for the sessions in Experiments 1 and 2, to prevent participants from experiencing the same sound method right after each other. In Experiment 3, the participants completed two sessions: No Sound and Sound, in randomized order. Each session consisted of 10 trials. Accordingly, each of the participants completed 80 trials in total (30 in Experiment 1, 30 in Experiment 2, and 20 in Experiment 3). The three experiments are explained below.

Experiment 1: Distance estimation. In the first experiment the participant heard a sound, equally loud in both ears of the headphones. The duration of the sound was 1.0 s for SI and SFF, and 3 beeps for BRR. The participant had to locate the object as accurately as possible by clicking on the screen. Immediately afterwards, the participants were shown the chosen location (cyan square) and the actual location of the object (red square), as well as an absolute distance error score expressed as a percentage shown in the left top of the screen. The experiment was preceded by a short automated demonstration in which the participants were presented with 11 sounds from low to high intensity (0%, 10%, …, 100%), together with a corresponding red square on the screen from top to bottom.

Experiment 2: **Distance and angle estimation**. The second experiment was the same as the first, but this time the participant had to locate the object in a two-dimensional plane (Fig. 2). Not only the distance but also the azimuth angle had to be estimated. To represent the angle, we used the volume per ear linearly mapped from the azimuth angle. If the sound volume was 100% in both ears, the object was in front of the participant. If there was only sound in the right ear (right volume = 100%, left volume = 0%), the object was on the right. Sound only in the left ear (left volume = 100%, right volume = 0%) meant that the object was located on the left. Experiment 2 was also preceded by a short demonstration of different distances as in Experiment 1, followed by a presentation of 10 different angles from right to left in 20 deg increments.

Experiment 3: Reaction time. The third experiment was divided into two sessions. In one session, the participant was presented with a block on the screen (Fig. 3). It could appear on the left, on the right, or in front of the participant. The participant had to press the left, right, or up arrow key as fast as possible. During the second session the participant both heard a sound corresponding to the location of the block and was presented with a visual representation of the block. After each trial, the reaction time was shown (Fig. 3).

Self-report questionnaires. At the end of the study, participants filled out the NASA TLX for measuring workload (Hart, 2006), complemented with two extra questions "To what extent did you feel motivated while testing?" and "I experienced discomfort (eyestrain, difficulty focussing, pain in ears and/or headache)". The items consisted of 21-tick scales running from *very low* to *very high*.

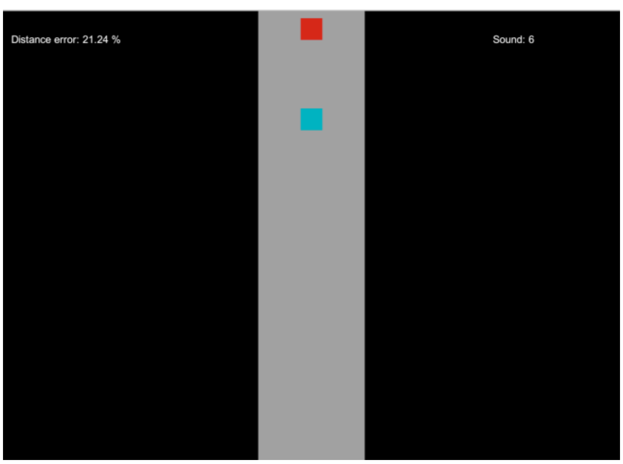

Fig. 1. Interface used in Experiment 1.

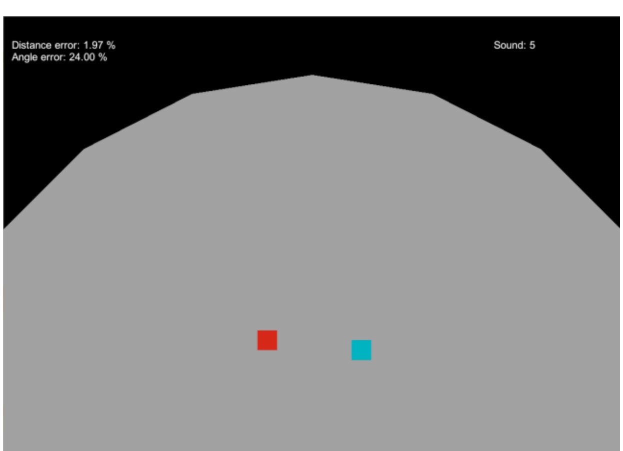

Fig. 2. Interface used in Experiment 2.

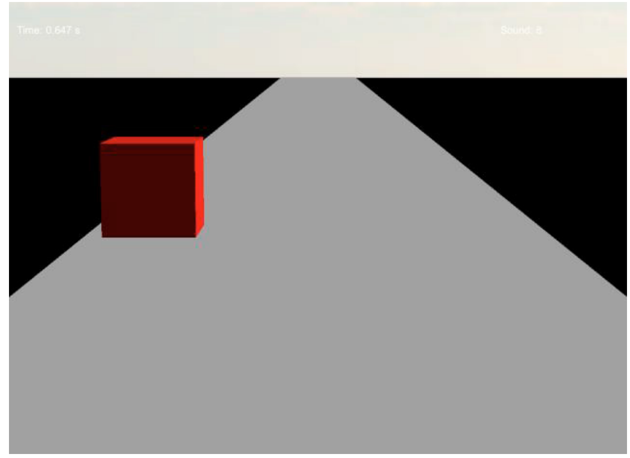

Fig. 3. Interface used in Experiment 3.

Procedure and instructions. Prior to the experiment, the participants received a leaflet explaining the three experiments. They were informed that the goal was to locate an object as accurately as possible based on the sound they heard by clicking on the screen, in a one-dimensional space (Experiment 1) or a two-dimensional space (Experiment 2). Regarding Experiment 3, the form stated: "The aim is for you to press on the left, right or up arrow as fast as possible (without making an error)."

Statistical analyses. The mean absolute distance error (Experiments 1 & 2), mean absolute angular error (Experiment 2), and mean reaction time (Experiment 3), averaged across trials of a session, were compared between sound conditions with paired t tests ($df = N - 1 = 28$).

3. RESULTS

Experiment 1: Distance estimation. The mean absolute distance error on a scale from 0 to 100 was 11.88 (SD = 5.26), 11.76 (SD = 3.87), and 12.03 (SD = 4.10), for BRR, SI, and SFF, respectively. Paired t tests revealed no significant differences between BRR and SI (p = .926), BRR and SFF (p = .906), and SI and SFF (p = .791). Figure 4 shows there were no structural under- or overestimations of the error, nor floor or ceiling effects. Figure 5 shows that BRR yielded particularly low errors when the target was near, whereas SI yielded relatively large errors in that case. Figure 6 shows the learning curves. A performance improvement was observed between trials 1–5 versus trials 6–10 (p = .001 for BRR, p = .026 for SI, p = .543 for SFF).

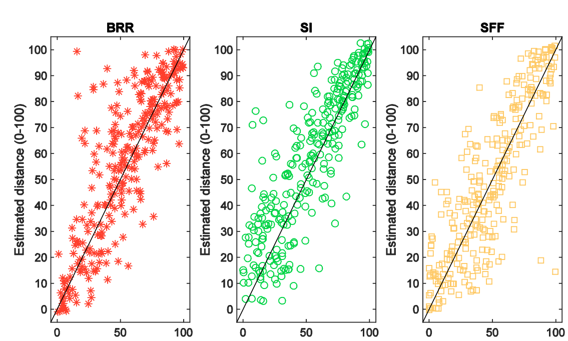

Fig. 4. Estimated distance versus actual distance for the three conditions in Experiment 1. 290 values (29 participants * 10 trials) are shown per plot.

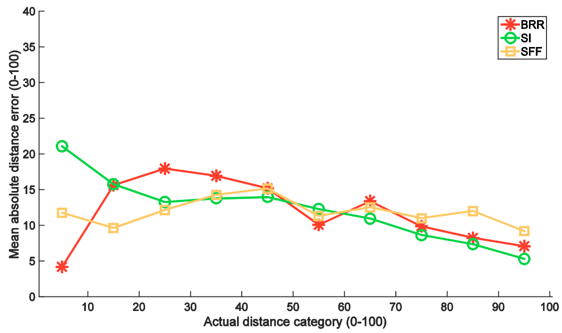

Fig. 5. Mean absolute distance error as a function of actual distance in Experiment 1. Ten categories of actual distances were created (0–10%, 10–20%, …, 90–100%).

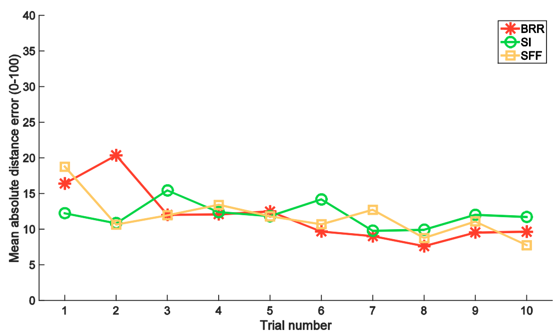

Fig. 6. Mean absolute distance error versus trial number in Experiment 1.

Experiment 2: Distance and angle estimation. The mean absolute distance error on the scale from 0 to 100 was 16.04 (*SD* = 6.35) for BRR, 20.20 (*SD* = 7.60) for SI, and 18.54 (*SD* = 7.52) for SFF. Paired *t* tests revealed a significant difference between BRR and SI (*p* = .016), and no significant difference between BRR and SFF (*p* = .174) nor between SI and SFF (*p* = .430). Figures 7 and 8 are consistent with the results of Experiment 1 (Figs. 4 and 5), in the sense that BRR performed particularly well when the distance was small, whereas SI performed relatively poor in that case. Figure 9 shows the learning curve of the distance estimation. A significant improvement was observed for SFF between trials 1–5 versus trials 6–10 (*p* = .733 for BRR, *p* = .137 for SI, *p* = .001 for SFF).

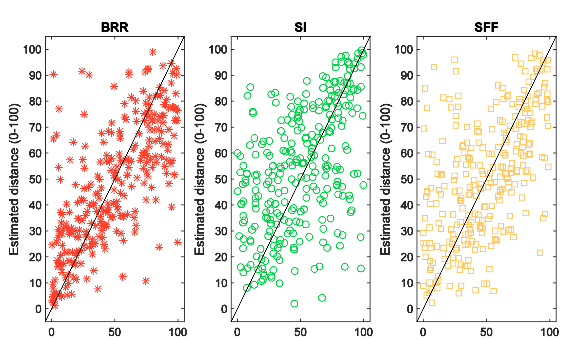

Fig. 7. Estimated distance versus actual distance for the three conditions in Experiment 2. 290 values (29 participants * 10 trials) are shown per plot.

A comparison of the mean absolute distance errors revealed significant differences between Experiments 1 and 2 (*p* < .001 for BRR, SI, & SFF), see also Figures 7 and 8 versus Figures 4 and 5.

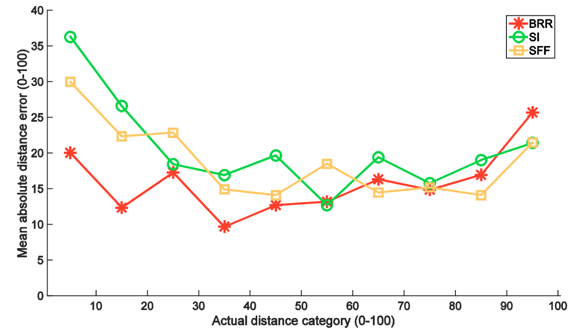

Fig. 8. Mean absolute distance error as a function of actual distance in Experiment 2. Ten categories of actual distances were created (0–10%, 10–20%, …, 90–100%).

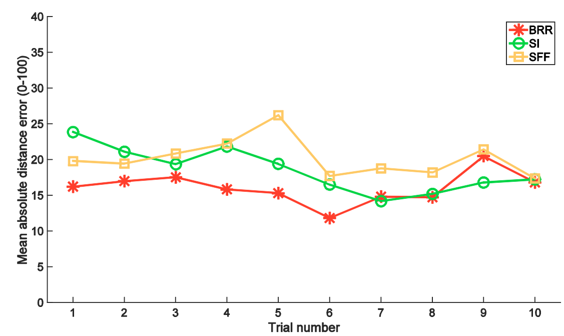

Fig. 9. Mean absolute distance error versus trial number in Experiment 2.

The mean absolute angle error was 22.74 deg (*SD* = 7.47) for BRR, 24.53 deg (*SD* = 8.90) for SI, and 24.26 deg (*SD* = 6.08) for SFF. There were no significant differences between BRR and SI (*p* = .431), between BRR and SFF (*p* = .325), and between SI and SFF (*p* = .899). The angular errors are illustrated in Figure 10, and the experience effects are illustrated in Figure 11. There was no significant performance improvement between the first five trials and the second five trials (p = .454, .893, & .564 for BRR, SI, & SFF, respectively).

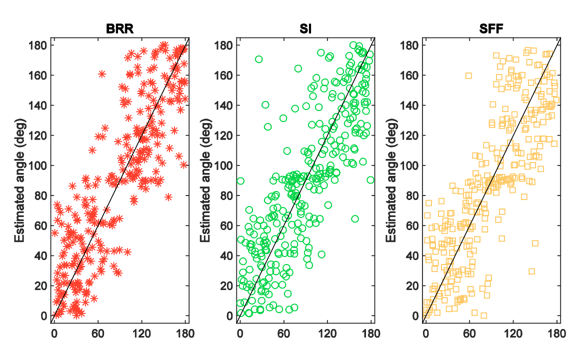

Fig. 10. Estimated angle versus actual angle for the three conditions in Experiment 2. 290 values (29 participants * 10 trials) are shown per plot. (0 deg = far right; 180 deg = far left).

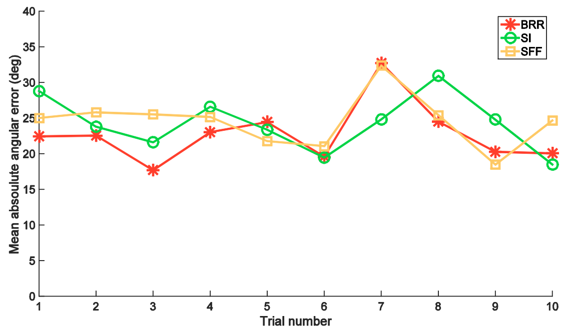

Fig. 11. Mean absolute angular error versus trial number in Experiment 2.

Experiment 3: Reaction time. The mean reaction times were 0.619 s with sound and 0.666 s without sound. The respective standard deviations among the 29 participants were 0.233 s and 0.267 s. A paired *t* test showed no significant difference in reaction time between the tests with and without sound (p = .366). The error rates were 3.8% (*SD* = 6.2%) with sound and 4.1% (*SD* = 6.3) without sound. Figure 12 illustrates the experience effect. A substantial performance improvement can be observed (a comparison of trials 1–5 with trials 6–10 yielded p = .011 for No Sound, and p = .004 for Sound).

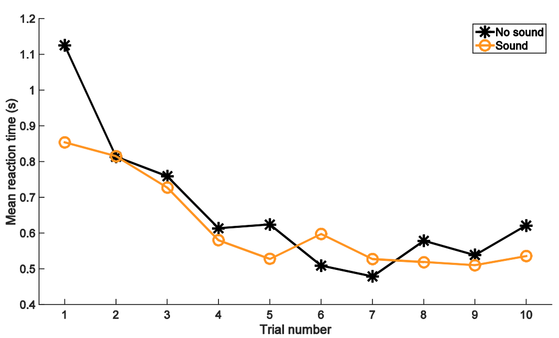

Fig. 12. Mean reaction time versus trial number in Experiment 3.

Self-report questionnaires. Figure 13 provides a boxplot of the eight questionnaire items. It can be seen that the task was regarded as somewhat mentally demanding, and that people were overall motivated.

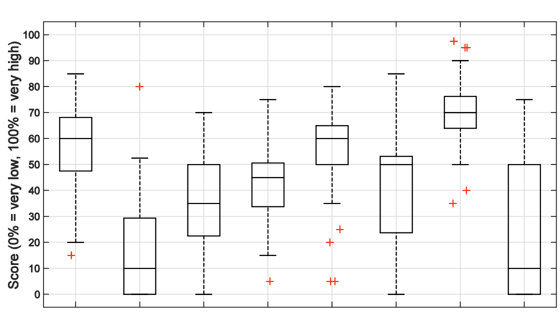

Fig. 13. Boxplots for the NASA-TLX and two additional questions. MD = Mental demand, PD = Physical demand, TD = Temporal demand, PERF = Performance, EF = Effort, FR = Frustration, MOT = Motivation, DISC = Discomfort.

4. DISCUSSION

The aim of this research was to determine which auditory method yields the smallest error between the object's actual and estimated location. For this purpose, in Experiments 1 and 2, we sonified the distance to the object along three primary dimensions: Beep Repetition Rate (BRR), Sound Intensity (SI), and Sound Fundamental Frequency (SFF). Additionally, in Experiment 2 the azimuth angle was sonified to the volume difference between the two ears.

The results revealed no clear-cut differences between the three sound conditions. The three proposed methods (BRR, SI, and SFF) yielded close to equal performance in the distance estimation task in the first experiment. However, in Experiment 2 (the distance and angle estimation task), BRR resulted in a significantly smaller percentage error than SI. BRR performed particularly well compared to SI when the actual distance was small (Figs. 5 and 8). This finding can be explained by the Weber-Fechner law, which states that the just noticeable difference between two stimuli increases linearly with stimulus intensity (e.g., Dehaene, 2003). For example, a difference between 10% and 20% volume is easier to distinguish than a difference between 90% and 100% volume. It is also possible that the sound level was saturated, and that ceiling effects may be the cause of the relatively poor performance in the SI condition when the distance was small. Moreover, according to the sone scale of loudness, how loud a sound is subjectively perceived is nonlinearly related to the physical sound intensity as well as sound frequency (Stevens, 1936). A related practical issue is that absolute sound level is difficult to control and reproduce on different desktop computers, which each have their idiosyncratic hardware and software configurations. For BRR it was beep time (rather than its reciprocal beep rate) that was linearly related to distance. At small distances, a small increase in beep time represents a large increase in beep rate (e.g., at 0% distance the beep rate was 10 Hz, and at 10% distance, the beep rate was 5 Hz). Thus the sonification of distance to beep rate may allow for sensitive discrimination at small distances. Moreover, BRR is an easy to reproduce and standardize means of sonification.

The participants became better at the tasks with increasing trial number. Experiments 1 and 2 used the same distance sonification. If the learning effect were the only factor affecting the error, one would expect the mean percentage error in the second experiment to be lower than in the first experiment. The results, however, showed the contrary. The mean distance error was statistically significantly higher in the second experiment than in the first experiment. This can be explained by the requirement to multitask in the second experiment: the participants had to divide their attention to determine both distance and angle. The difference in sound intensity in the ears could make it harder for the participants to estimate the distance of the object.

In Experiment 3, we observed no statistically significant difference in the reaction times between visual directional feedback and audio-visual feedback. This may be due to a lack of statistical power, or due to the visual dominance in these types of tasks (Posner et al., 1976).

In conclusion, with appropriate instructions and knowledge-of-results feedback, humans have a discriminating power of beep rate, sound volume, and sound frequency that allows them to map these sound dimensions to a virtual distance.

5. ACKNOWLEDGEMENT

Pavlo Bazilinskyy and Joost de Winter are involved in the Marie Curie ITN: HFAuto (PITN-GA-2013-605817).

6. REFERENCES

Barrass, S. and Kramer, G. (1999). Using sonification. Multimedia Systems, 7(1), 23–31.

Bazilinskyy, P., Petermeijer, S.M., Petrovych, V., Dodou, D., and De Winter, J. C. F. (2015). Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays. Unpublished.

Capp, M. and Picton, P. (2000). The optophone: an electronic blind aid. Engineering Science & Education Journal, 9(3), 137–143.

De Florez, L. (1936). True blind flight. Journal of Aeronautical Sciences, 3(5), 168–170.

Dehaene, S. (2003). The neural basis of the Weber–Fechner law: a logarithmic mental number line. Trends in Cognitive Sciences, 7(4), 145–147.

Dubus, G. and Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. PLOS ONE, 8(12), e82491.

Flowers, J.H. (2005). Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions. Proceedings of the International Conference on Auditory Display (ICAD2005), Limerick, Ireland, 6–9 July, 406–409.

Haas, E.C. and Edworthy, J. (1996). Measuring perceived urgency to create safe auditory warnings. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 40(16), 845–849.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(9), 904–908.

Hellier, E. and Edworthy, J. (1999). On using psychophysical techniques to achieve urgency mapping in auditory warnings. Applied Ergonomics, 30(2), 167–171.

Hermann, T., Hunt, A., and Neuhoff, J. (2011). The sonification handbook. Logos Publishing House, Berlin.

Hussain, I., Chen, L., Mirza, H.T., Xing, K., and Chen, G. (2014). A comparative study of sonification methods to represent distance and forward-direction in pedestrian navigation. International Journal of Human-Computer Interaction, 30(9), 740–751.

Jamson, A.H., Lai, F.C.H., and Carsten, O.M.J. (2008). Potential benefits of an adaptive forward collision warning system. Transportation Research Part C: Emerging Technologies, 16(4), 471–484.

Knoll, G.F. (2010). Radiation detection and measurement. John Wiley & Sons, Hoboken, NJ.

Kramer, G., Walker, B., Bonebright, T., et al. (1999). The sonification report: Status of the field and research agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display. ICAD, Santa Fe, NM.

Loomis, J.M., Klatzky, R.L., Philbeck, J.W., and Golledge, R.G. (1998). Assessing auditory distance perception using perceptually directed action. Perception & Psychophysics, 60(6), 966–80.

Parseihian, G., Katz, B.F.G., and Conan, S. (2012). Sound effect metaphors for near field distance sonification. Proceedings of International Conference on Auditory Display (ICAD) ’12, Atlanta, GA, 18–21 July, 6–13.

Patterson, R.D. (1982). Guidelines for auditory warning systems on civil aircraft (CAA paper 82017). Civil Aviation Authority, London.

Perrott, D.R., Saberi, K., Brown, K., and Strybel, T.Z. (1990). Auditory psychomotor coordination and visual search performance. Perception & Psychophysics, 48(3), 214–226.

Posner, M.I., Nissen, M.J., and Klein, R.M. (1976). Visual dominance: An information-processing account of its origins and significance. Psychological Review, 83(2), 157–171.

Sanders, M.S. and McCormick, E.J. (1987). Human Factors in Engineering and Design. McGraw-Hill, New York, NY.

Spain, R.D., Bliss, J.P., and Newlin, E.T. (2007). The effect of sonification pulse rate on perceived urgency and response behaviors. Proceedings of the Human Factors and Ergonomics Society, 61(19), 1345–1348.

Stanton, N.A. and Edworthy, J. (1999). Auditory warnings and displays: An overview. In N. A. Stanton & J. Edworthy (Eds.), Human factors in auditory warnings (pp. 3–30). Ashgate Publishing Limited, Aldershot.

Stevens, S.S. (1936). A scale for the measurement of a psychological magnitude: Loudness. Psychological Review, 43(5), 405–416.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman and D.R. Davies (Eds.), Varieties of attention (pp. 63–102). Academic Press, New York.

Zahorik, P. (2002). Assessing auditory distance perception using virtual acoustics. The Journal of the Acoustical Society of America, 111(4), 1832–1846.