EVALUATION OF REACTION TIMES TO SOUND STIMULI ON MOBILE DEVICES

György Wersényi, Hunor Nagy

Széchenyi István University
Department of Telecommunications
Egyetem tér 1, Győr, Hungary, H-9026
wersenyi@sze.hu, nhunorz@gmail.com

Ádám Csapó

Research Inst. for Computer Science and Control
3D Internet based Comm. and Control Lab.
Kende u. 13-17, Budapest, Hungary, H-1111
csapo.adam@gmail.com

ABSTRACT

When developing new devices for assistive technology, it is important to consider auditory response times to different kinds of warning and navigational sounds. Perception, processing and action following the presentation of auditory stimuli depends on several parameters, the most important being the stimuli themselves and the method used for providing feedback. With the growing market penetration of mobile devices (smartphones, tablets etc.) and increasing popularity of crowdsourced solutions, we have chosen to develop a mobile application for the measurement of reaction times with respect to a wide range of stimuli, including sine tones, speech and various kinds of clicks and noises. During tests, participants are asked to indicate the direction of sound samples by pressing the appropriate button on the touch screen. Stereo panning can be used up to five directions. In this paper, our goal is to demonstrate the viability of this approach through a set of basic (at this time, not yet crowdsourced) tests performed using the application. A rudimentary statistical evaluation of measured response times and success rates was performed. Results were compared to an earlier study using similar categories of stimuli. As in that study, some relative differences between the stimuli types were found, i.e. the 1 kHz panned sine and pink noise categories were shown to be somewhat more favorable than speech and click-trains. Future enhancements to the application will include tilt-based input control – allowing for the participation of visually impaired test subjects who cannot see the response buttons – as well as extensions allowing for the logging and analysis of large-scale crowdsourced test results.

1. INTRODUCTION

Reaction time is an important parameter in a large variety of use-case scenarios where quick response to safety issues is essential.

Test installations usually incorporate auditory and/or visual excitations and investigate subject reaction times in terms of time taken to press an appropriate button or to take an appropriate steering action. Common areas of interest include driving assistance scenarios, aviation tasks and combat simulator operation [1-5]. Assistive technologies helping impaired people in safe navigation – especially in the case of multiple sound sources – also benefit from well selected auditory stimuli in their auditory interfaces [6-13]. This application area carries with itself additional challenges inasmuch as feedback is not only required to be fast and accurate, but it is also required to take into consideration a potentially broader set of user interactions than the mere pressing of a button. For example, directional instructions can be responded to by a broad range of limb and/or full body movements. Thus, the test environments developed for the evaluation of assistive technologies can be expected to involve an equally broad range of interaction modes, including the tilting, turning, shaking or translational displacement of the mobile device.

Several authors have in the past investigated the relative advantages of various kinds of auditory stimuli in different contexts (among many others, examples can be found in [13-15]). In this paper, we present an application we have developed for such purposes. As a first step towards its evaluation, we compared the results obtained in a set of offline (i.e. not yet crowdsourced) experiments to a very recent report by Nagel *et al*. on the measurement of response times in navigational tasks [15]. We use the same categories of stimuli, including panning, noises, female TTS (text-to-speech) and sine stimuli as summarized in Table 1 [15]. The experimental setup in Nagel *et al*.'s paper was developed using MATLAB and dedicated hardware equipment for low latency measurement. The input device included three buttons corresponding to three directions. Eight listeners were tested in a 50-round random order test run after a suitable training period. Both response times and success rate were measured.

Summarized results showed overall response times of 450.8ms with a standard deviation of 180.4ms. Responses faster than 100 ms were rare, while responses slower than 1 second were frequent. Panned sine tones were confirmed as the best (M = 338 ms, SD = 118 ms, success rate = 95%) while long chords resulted in the slowest reaction times and much lower success rates (M=533 ms, SD = 198 ms, success rate = 79%). After three runs, mean response times were about 6% better than in the first run [15].

In this paper, we considered these results as a baseline for comparison. Our investigations also focus on similar excitation signals, although the hardware/software architecture used for the presentation of stimuli and for the analysis of response times was somewhat different. For this reason, we first discuss the measurement setup and procedure, along with possible advantages and disadvantages. This is followed by a presentation and comparative discussion of results. A final section on future work highlights possible enhancements.

Type	Left	Straight	Right
Chords	Major (f B=300Hz)	Sine tone (f=300Hz)	Minor (f B=300Hz)
Distinction	Click train	Sine tone (f=300Hz)	White noise
Panning	-90 deg	0 deg	+90 deg
Pitch	Low pitch sine, 80 Hz	Mid pitch sine, 200Hz	High pitch sine, 4000Hz
Speech	“Left” sine, 800Hz	“Straight” sine, 800Hz	“Right” sine, 800Hz

Table 1. Stimuli in Nagel’s experiment [15].

2. MEASUREMENT SETUP

2.1. The CrowdAudio app

The application was developed by the authors for the Android platform with the long-term goal of conducting not only laboratory but crowdsourced auditory psychophysical experiments. Through the use of the application, an important research question will be whether any kind of statistical smoothing performed on vast amounts of crowdsourced data can compensate for variabilities in hardware setup and the lack of standardized experimental protocols. In its current version, the app supports the flexible creation of tests in which the direction of wave samples linked to 2 or 3 directions can be queried from users (Figure 2). Reaction times and measures of correctness are recorded by the app into a locally stored log file (Figures 3 and 6). The parameters associated with each test include:

- *Number of directions* (2 or 3)
- *Number of test questions* (1-100)
- *Minimum delay* between test questions (in milliseconds)
- *Maximum delay* between test questions (in milliseconds)
- *Maximum response wait time* (in milliseconds).

The difference between *maximum* and *minimum delay* between test questions serves to create uncertainty as to when exactly the next wave sample will be played. Following the minimum time delay, the GUI is activated and lights up, however, the app continues to wait for a random amount of time (maximized at the difference between the *maximum* and *minimum delay*) before actually playing the next test question. The *maximum response wait time* is specified so that the test can be aborted whenever the test subject is interrupted or is ostensibly not paying attention (Figure 4).
When a test is complete, a summary of results are provided to users. An element of gamification (i.e., the use of game-like features to increase users’ level of interest and willingness to contribute), which involves displaying the top 5 scores of all time, as well as the current user’s relative performance, was introduced to further encourage correct answers at low response times (Figure 5). This game score is not used for the purposes of scientific evaluation and only serves to encourage users to participate actively. Upon the selection of appropriate settings, results are saved to a text file in .csv format (Figure 6).

2.2 Test protocol used

The application was installed on two smartphones and one tablet from the same vendor. In the case of smartphones, headphones were attached, while in case of the tablet, a set of built-in loudspeakers (on two sides of the device) were used. No special selection criteria were applied for the headphones, users were allowed to use either their own equipment or the one supplied by the administrator of the tests.
To keep the number of variables investigated to a minimum, the number of directions was set to three (left, right, front), the number of questions in each run to 15, and all maximum delay times to 2 seconds. Sound was played back from the left and right speaker in the case of *left* and *right* directions respectively. In the case of the *front* direction, both speakers radiated sound (50%-50%) in order to create a frontal sound source image. If the test is aborted by the user or automatically due to the user exceeding the maximum response wait time (Fig.4.), tests were repeated and logged data were deleted during the evaluation process. The app logged the user’s unique identifier (name), gender, age, date and time of the test. This allowed for the same users to repeat their tests any number of times, while allowing for the integrated evaluation of results. Thus, although the main goal was to test reaction times for different categories of stimuli, it would have also been possible to measure correlations across both individual users and groups of users.
In a way similar to the tests reported by *Nagel et al*, the following samples were used: 1 kHz sine, pink noise, click-train burst and female TTS speech. The current test included the Hungarian version of TTS samples “*left*”, “*right*” and “*front*” as this was the mother tongue of most users in our preliminary experiments. In further use, English samples will be provided. Two different methods were used for the generation of samples. In the first method, test samples and real-time filtering was done using SuperCollider-Android (a port of SuperCollider for Android); while in the second method, simple playback was provided of pre-filtered wave files. In the latter case, sound samples were created in 44.1 kHz / 16 bit resolution and stored as stereo wave files (e.g. in case of “left” only the left channel was active). As long as only short wave files were used, the latter approach was preferred and used in this test.
In every run, 15 random questions were presented to the users, and even though this is a multiple of 3, it was not guaranteed that all 3 directions would be presented exactly five times. Results were stored in a log file, which was later imported into Excel for further evaluation. Figure 6 shows the logged data.

Experiments were conducted using the application in normal listening rooms with staff guidance. Altogether, there were 15 participants in the experiments (9 male, 6 female, between 23 and 47 years of age, with a mean age of 34 years). All subjects were sighted persons.

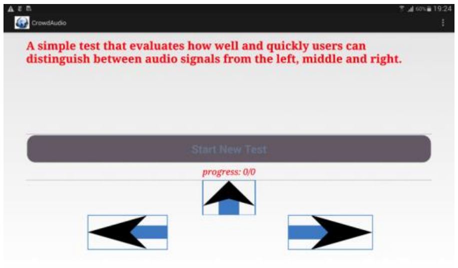


Figure 2. Basic experimental setup.

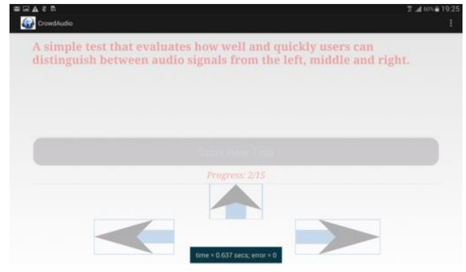


Figure 3. Feedback on correctness and measured time delay following each test question.

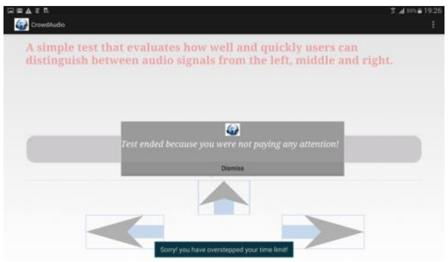


Figure 4. Tests are aborted when user is apparently not paying attention.

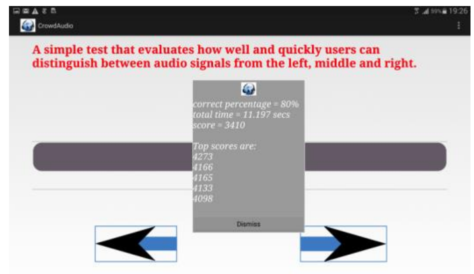


Figure 5. Results are summarized following the test.

2.3 Advantages and disadvantages of the approach

Before presenting results, we briefly discuss the advantages and disadvantages of the approach taken.

It is a well-known fact that Android devices as well as other popular mobile platforms run pre-emptive operating systems, meaning that all measurements will inherently involve some random latency that depends on a variety of factors, including number of background processes, memory availability and others. Compared with previous investigations, in which dedicated low-latency hardware was used, together with e.g. audio recording technologies to detect the times at which the stimuli became audible and the (physical) response buttons were pressed, the approach reported here involves more uncertainty.

On the other hand, the use of a relatively limited set of devices together with large numbers of (crowdsourced) data means that hopefully relative differences remain comparable. For example, if response accuracy or delay is found to be more favorable for one type of stimuli compared to another, the use of the same pair of devices, along with a sufficient number of measurements can mean that relative differences can be generalized.

An important question in the long run will be whether enough data can be collected in crowdsourced settings such that differences in devices and testing protocols can be compensated for in a statistical sense. If the answer is in the affirmative, and statistically rigorous approaches are used, then the disadvantages associated with the lack of fixed latency can perhaps be outweighed by the advantages of the availability of large amounts of data.

3. RESULTS

Altogether, participants delivered 1990 responses for evaluation. 1925 were correct (96.73%) and 65 were (3.27%) incorrect. From the correct answers 126 (6.545%) were labeled as outliers. Outliers were defined per person if measured reaction time value was at least 50% higher than of the mean reaction time of the same person. Such outliers were considered to be due to incorrect use of the touch screen, tapping outside the borders of buttons, or releasing the buttons too fast. The data was tested for normal distribution using a linear regression model.

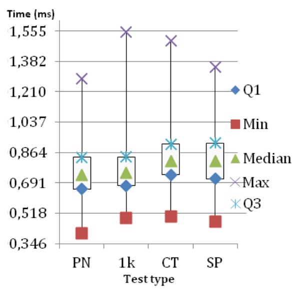
```
name,Szandi,age,24,gender,2;
correctAnswer,-1,error,0,time,1118.0;
correctAnswer,-1,error,0,time,800.0;
correctAnswer,0,error,0,time,935.0;
correctAnswer,-1,error,0,time,913.0;
correctAnswer,1,error,0,time,861.0;
correctAnswer,-1,error,0,time,932.0;
correctAnswer,0,error,0,time,970.0;
correctAnswer,0,error,0,time,918.0;
correctAnswer,0,error,0,time,912.0;
correctAnswer,1,error,0,time,879.0;
correctAnswer,0,error,0,time,842.0;
correctAnswer,1,error,0,time,854.0;
correctAnswer,0,error,0,time,815.0;
correctAnswer,-1,error,0,time,956.0;
correctAnswer,1,error,0,time,824.0;
===== year,day,hour:minute:second =
2015,41,12:35:54 ==== testType 8;
```

Figure 6. Example showing the format of logged results.

Results for the test signals (PN: pink noise; 1k: 1 kHz sine; CT: click-train; SP: female TTS speech) are shown in Table 2. Results based on gender are shown in Table 3. Five of the subjects performed each test several times. Figure 8 shows average response times for the first five runs. After the third run, users' response times became relatively stable.

	MIN	MAX	MEAN	MED	SD	ERR
PN	0,406	1,285	0,753	0,736	0,131	2,85%
1k	0,490	1,551	0,769	0,748	0,138	2,48%
CT	0,565	1,473	0,843	0,815	0,149	4,81%
SP	0,470	1,350	0,828	0,815	0,158	3,33 %

Table 2. Minimum, maximum, mean, median, standard deviation and error rates for the test signals (excluding outliers and incorrect answers).



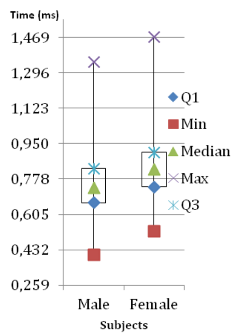


Figure 7. Box and Whisker plot for all test signals (left) and for male and female results (right).

	MIN	MAX	MEAN	MED	SD	ERR
male	0,406	1,350	0,760	0,733	0,146	3,49%
female	0,523	1,473	0,838	0,823	0,140	1,96%

Table 3. Minimum, maximum, mean, median, standard deviation values and error rate for male and female subjects (excluding outliers and incorrect answers).

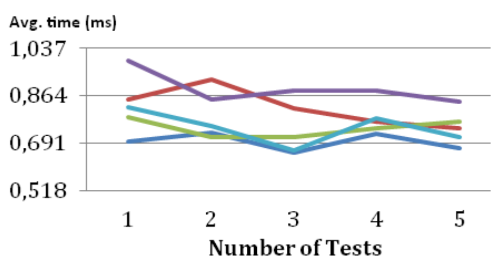


Figure 8. Average response time for five subjects in the same test after repeated trials in seconds.

4. DISCUSSION

The main purpose of the investigation was to see whether the use of the CrowdAudio application yielded realistic results when used to measure reaction times to various auditory excitation signals through touch screen interaction on mobile devices. According to Table 1 and Figure 7, signals were detected with an accuracy of 97.15 - 95.19%. Pink noise and the 1 kHz sine were detected most accurately, as in the experiments reported by Nagel *et al*. Errors using the click-train signal and speech were more frequent. The same applies to mean and median values, and SDs were also larger.

The overall mean response time for correct answers was 798ms, much larger than Nagel's 450.8ms. This is probably due to differences in equipment, that is, the difference between the use of dedicated hardware with hardware buttons in contrast to the touch screen of a mobile device. Although in general applications (e.g. for blind people) we do not expect such differences to play a significant role, they do highlight the fact that response times highly depend on the input media and the feedback action required from the user. The same observation can be made with respect to minimum and maximum response time rates in this comparison.

In order to find out whether differences in performance among stimulus categories were significant, a one-way ANOVA test was performed. Prior to that procedure, the O'Brien test was used to determine whether the data sets were sufficiently homogeneous. As the result was in the negative, a logarithmic transformation was applied as a pre-processing step, following which the criterion towards homogeneity was satisfied. The subsequent ANOVA test confirmed that differences between test types were significant ($p < 0.001$).

Left and right directions were easily identified by the users upon first use, but inexperienced subjects (that is, most subjects) asked about the meaning of the "front" direction and how they should perceive sound sources in the frontal direction. Subjects usually did not localize this direction at all (and this was not the purpose of our tests either), but they adopted and learned the link between the sound coming from both speakers and the button with the up arrow. Future work will include 5 directions instead of 3, with 45-degree directions emulated using stereo panning. It is assumed that introducing these directions between front and side will increase both error rates and reaction times. Preliminary results of test runs using 5 directions support this assumption.

In general, speech was not favorable to user performance. It was difficult for users to abstract away from the semantic meaning of spoken words, and to interpret the signals merely as directional cues. Hence, an interpretation of speech required more time. Furthermore, all other signals were active at full loudness throughout the entire presentation time, while speech samples are different for each direction (especially in Hungarian) and users needed some time to recognize even the first vocals.

Based on the results of the participants who performed the same test many times, reaction times were not significantly altered after three runs. It is suggested that about three full runs are needed for naïve participants to grow accustomed to the test, i.e. to learn the signals, the presentation method, and to optimize their strategy for obtaining the best results. This also corresponds with Nagel's experiment, who measured response times that were 6% faster after three runs. While this cannot be seen as a "real training and learning effect" in a classical sense, results nevertheless indicate a short accommodation time and a low number of trial runs to achieve maximal efficiency of inexperienced users in such tests.

It was observed that users tended to use this small application as an “unofficial” competitive game against their own as well as others’ results. After some test runs users developed their own method to maximize their chances for acting fast. This was usually done by placing the device on a table, holding the index finger as close to the screen as possible, and a bit below the “front” arrow. Although it was not determined how participants should use the device, none of the subjects were holding it in the hand(s) and/or were trying to use more than one finger. In turn, this often led to users to react in haste, and in some cases the speed of their movements was such that their pressing motions were not registered by the touch screen. Quite often this led to outliers in the data set.

5. FUTURE WORK

Plans for future work include:

- Extension of possible directions to 5 or 7 using pre-filtered stereo panning of sources in wave file format in the horizontal plane (no HRTF filtering and no vertical simulation is planned),
- Measurement of response times for feedback given by tilting the mobile device instead of pressing buttons on the screen,
- Inclusion of blind participants in the test, especially for testing alternative GUI and feedback options,
- Testing possibilities for including directions outside the frontal hemisphere (e.g. from the back) with additional filtering such as using pink noise from behind and white noise from the front,
- Testing the effect of vibration using the built-in vibration motors of devices (as an extension to the stimuli),
- Testing the effects of training, i.e. how response times and accuracy evolve after several test runs,
- Offering the application for free use and testing using crowdsourcing in order to compare lab data and data gathered from the Internet [16, 17].

6. SUMMARY

A mobile application for the Android platform was presented along with a measurement of reaction times of 15 untrained subjects using four different acoustic stimuli. The 1 kHz panned sine and pink noise produced the best results, while click-train and the female TTS speech were less favorable to performance. These results were shown to be significant. Mean response times of about 800 ms were detected, which is much slower than could be achieved using dedicated hardware feedback devices. Although there was no special protocol designed for the training of individuals, after three runs naïve users were able to use the system at their full potential.

7. ACKNOWLEDGEMENT

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 643636 “Sound of Vision”.

8. REFERENCES

[1] R. Hyman, ”Stimulus information as a determinant of reaction time,” *Journal of Experimental Psychology*, vol. 45, no. 3, pp. 188-196, March 1953.

[2] R. J. Kosinski, ”A Literature Review on Reaction Time,” http://biae.clemson.edu/bpc/bp/lab/110/reaction.htm, accessed 2015.02.02.

[3] L. H. Nikam, and J. V. Gadkari, ”Effect of age, gender and body mass index on visual and auditory reaction times in Indian population,” *Indian Journal of Physiology and Pharmacology*, vol. 56, no. 1, pp. 94-99, Jan-Mar 2012.

[4] J. Solanki, N. Joshi, C. Shah, H. B. Mehta, and P. A. Gokhle, ”A study of correlation between auditory and visual reaction time in healthy adults,” *International Journal of Medicine and Public Health*, vol. 2, no. 2, pp. 36-38, 2012.

[5] M. T. G. Pain, and A. Hibbs, ”Sprint starts and the minimum auditory reaction time,” *Journal of Sports Sciences*, vol. 25, no. 1, pp. 79-86, 2007.

[6] A. D. N. Edwards, “Auditory display in assistive technology,” in The Sonification Handbook, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany: Logos Publishing House, 2011, ch. 17, pp. 431–453. [Online]. Available: http://sonification.de/handbook/chapters/chapter17/

[7] B. Simpson, D. Brungart, R. Gilkey, and R. McKinley, “Spatial audio displays for improving safety and enhancing situation awareness in general aviation environments,” *in Proc. New Directions for Improving Audio Effectiveness Meeting RTO-MP-HFM-123*, Neuilly-sur-Seine, France, pp. 16-26, 2005.

[8] Á. Csapó, Gy. Wersényi, H. Nagy, and T. Stockman, “Survey of assistive technologies and applications for blind users on mobile platforms - a review and foundation for research,” *Journal of Multimodal User Interfaces*, vol. 9., no.3, 11 pages, 2015.

[9] J. Gillard, and M. Schutz, “Improving the efficacy of auditory alarms in medical devices by exploring the effect of amplitude envelope on learning and retention,” *in Proc. of ICAD12*, Atlanta, USA, pp. 240-241, 2012.

[10] R. Ratcliff, ”Methods for dealing with reaction time outliers,” *Psychological Bulletin*, vol. 114., no. 3, pp. 510-532, Nov 1993.

[11] K. Suzuki, H. Jansson, ”An analysis of driver’s steering behaviour during auditory or haptic warnings for the designing of lane departure warning system,” *JSAE Review*, vol. 24, no.1, pp. 65-70, January 2003.

[12] S. A. Panëels, D. Varenne, J. R. Blum, and J. R. Cooperstock, “The Walking Straight Mobile Application: Helping the Visually Impaired avoid Veering,” *In Proc. of ICAD13*, Lódz, Poland, pp. 25-32, 2013.

[13] M. Nees, and B. Walker, “Auditory Displays for In-Vehicle Technologies,” *Human Factors and Ergonomics*, vol. 7, no. 1, pp. 58-99, 2011.

[14] C. Ho, and C. Spence, “Assessing the Effectiveness of Various Auditory Cues in Capturing a Driver's Visual Attention,” *Journal of Applied Experimental Psychology*, vol. 11, no. 3, pp. 157-174, 2005.

[15] F. Nagel, F.-R. Stöter, N. Degara, S. Balke, and D. Worrall, “Fast and accurate Guidance – Response Times to Navigational Sounds,” *in Proc. of ICAD14*, New York, USA, 5 pages, 2014.

[16] M. Schoeffler, F.-R. Stöter, H. Bayerlein, B. Edler, J. Herre, “An Experiment About Estimating the Number of Instruments in polyphonic Music: A Comparison between Internet and Laboratory Results,” *in Proc. of the 14th International Society for Music Information Retrieval Conference*, Curitiba, Brazil, 6 pages, 2013.

[17] U.-D. Reips, ”Standards for Internet-based experimenting,” *Experimental Psychology*, vol. 49, no. 4, pp. 243-256, 2012.