FAST AND ACCURATE GUIDANCE - RESPONSE TIMES TO NAVIGATIONAL SOUNDS

Frederik Nagel[1]*, Fabian-Robert Stöter*[2]*, Norberto Degara*[1]*, Stefan Balke*[2]*, David Worrall*[1]

International Audio Laboratories Erlangen
[1]Fraunhofer IIS & [2]FAU Erlangen-Nürnberg
Am Wolfsmantel 33, 91058 Erlangen, Germany
Frederik.Nagel@iis.fraunhofer.de

ABSTRACT

Route guidance systems are used every day by both, sighted and visually impaired people. Systems, such as those built into cars and smart phones, usually using speech to direct the user towards their desired location. Sounds other than functional and speech sounds can, however, be used for directing people in distinct directions. The present paper compares response times with different stimuli and error rates in the detection. Functional sounds are chosen with and without intrinsic meanings, musical connotations, and stereo locations. Panned sine tones are identified as the fastest and most correctly identified stimuli in the test while speech is not identified faster than arbitrary sounds that have no particular meaning.

1. INTRODUCTION

Route guidance systems are used every day by both, sighted and visually impaired people. Those systems are built into cars and smart phones, for example, usually using speech to direct the user towards their desired location. Sounds other than functional and speech sounds can, however, be used for directing people in distinct directions and that is where Auditory Display research comes into play.

Auditory Displays are systems that transform data into sound and present this information using an interface to allow the user to interact with the sound synthesis process. This transformation of data into sound is called sonification and it can be defined as the systematic data-dependent generation of sound in a way that reflects objective properties of the input data [1].

In the context of navigation, visual substitution, and obstacle avoidance applications, sonification technology can be used to deliver location-based information to support eyes-free navigation through sound. This is a very challenging task as described in [2]. The challenge is to design a meaningful auditory display that is able to communicate relevant aspects of complex visual scenes, where aesthetics is a very important factor due to the frequent use of the display. The resulting sound must be accurate in terms of the location-based information communicated but it has to be also attractive to the user. While in everyday car driving, the state-of-the-art display is probably sufficient for most people, there are situations such as rally driving or navigation without sight where existing systems are not sufficient.

As one can expect, multiple sonification methods for visual substitution, navigation and obstacle avoidance can be found in the literature [2]. In general, these methods scan the space to look for potential obstacles and synthesize the position or other properties of the scene using different sound rendering modes. These modes include depth scanning [3], radar, and shockwave modes [4]. There are also approaches where a non-blind external operator that analyses the received image and traces the direction to be followed [5]. The sonification algorithms used to synthesize the sound are based on Parameter-Mapping [6] and Model-based sonification techniques [7].

Complete navigation systems which do not require visual or tactile interaction are comprehensively discussed in [8]. That paper particularly discusses several design choices such as timbre, silence vs. sound for conveying information, and the use of different information-sound mappings. With informal tests, the authors identify spatial information as particularly useful to guide users towards destinations. The use of spatial information in aircraft flying was further investigated in [9, 10], finding that spatial cues are easy to interpret; the benefit depended, however, of the willingness of the pilots to use it.

Despite all this work, a formal evaluation of both the accuracy and response time of the most common audio stimuli used in navigation systems has not been carried out. Accuracy refers to the precision when choosing the right direction in a navigation task and response time refers to how fast the user responds to the audio stimuli. The present paper focuses on both, the correctness in identifying different simple stimuli used for navigation and the observed response times. It reports our investigation of whether speech, panning, musical information, or [high discrimination between sounds, highly discriminated sounds] lead to lower error rates and shorter detection times. This work has been developed within the context of SONEX, a benchmark platform used to compare the efficacy of different sonification methods [11], and its application to blind navigation as proposed in [12].

Section 2 describes the experimental method used in our experiment. Section 3 presents the results which are then discussed in Section 4. Finally, conclusions and future work are discussed in Section 5.

2. METHOD AND MATERIAL

2.1. Stimuli and Stimuli Presentation

Five categories of stimuli were created exhibiting different characteristics as listed in Table 1. Major and minor chords were employed to test reaction to musical meanings. The chords consisted of three sine-tones, whereas the pitch of the bass-note is given as

f_B. No intrinsic relation of positions was expected here; similarly for a very distinctive click train and white noise. Pitch in contrast has, at least for musicians, an intrinsic meaning, considering for instance a piano where low pitches are located to the left, mid-range pitches in the middle and high pitches on the right hand side of the player. Speech and panning have a self-explanatory meaning.

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

Table 1: Stimuli used in the experiment

Except for the speech samples which were obtained from the Mac OS X text to speech function with VICKY as the speaker, all stimuli were created using MATLAB.

The samples were presented using a self-created MATLAB software with PLAYREC (`www.playrec.co.uk`) for real-time audio output. Measuring the response times is sensitive to the overall input-output latency of the measurement setup. This includes the introduced delay of the operating system and MATLAB when measuring keyboard responses of which we cannot make precise statements. Therefore we designed a simple circuit which generated short audio impulses when a button was pressed. We used a RME FIREFACE UC audio interface for the experiment with five input channels, three for each button and additionally two for the loopback channels were used. A participant had to indicate the recognized direction by pressing one of the buttons on the circuit. The resulting signal onset time was then compared to the onset time of the stimulus which was recorded on two further channels. The audio signal was hence recorded by the audio interface along with the participants' actions on the buttons. Therefore, the delay between the response signal and the stimulus do not have to be compensated, allowing for an exact measurement of the participants' reaction times towards the onsets of the presented stimuli. All but the speech samples were replayed continuously until a participant's reaction; the speech samples ended after the words were played. The input device is shown in Figure 1.

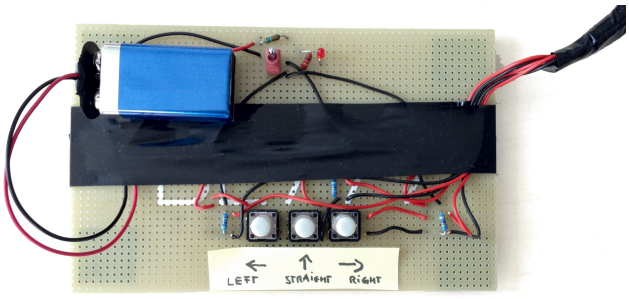

Figure 1: Circuit for participants' feedback generating analog impulses

2.2. Test Design and Procedures

The three stimuli *straight*, *left*, and *right* for each of the $k = 5$ categories were presented in random order in blocks of $l = 50$ to the participant. Additionally, an experiment was divided in two parts. In the first stage, each of the participants received three rounds of training of three possible stimuli. The training stage was supported on-screen by showing the corresponding label of the stimuli. This way the participants could match for instance the *left* direction with the *low pitched tone*.

In the second stage, the playback of the stimuli started automatically and was interrupted by the participants response. After a pause of $1s$ the next stimuli was played back. Each of the categories were presented three times by using a random permutation. In total each participant responded to $5 \cdot 50 \cdot 3 = 750$ stimuli. To complete the experiment most participants needed about 30 minutes.

2.3. Participants

The set of test subjects consisted of eight listeners with normal hearing at the age of 25 to 35 years. The median age was 29. Prior to the test, the subjects were asked to rate their musical knowledge. The employed scale ranged from zero to ten; zero meaning no musical knowledge and ten being at the level of a professional musician ($M = 6$, $SD = 1.9$).

3. RESULTS

The results of the experiment were mainly based on the response times n_Δ which were measured in ms. Additionally, the responses were categorized as *correct* or *incorrect* resulting in a binary dependent variable $success \in \{\text{true}, \text{false}\}$. The results for the dependent variable n_Δ were filtered by showing only valid and correct responses whereas *success* is based on the complete result data. From the total number of $n = 6000$ observations we removed 171 which are outside three times inter-quartile-range of the response times ($n_\Delta > 1214$ ms). From the remaining 5829 observations 5340 were considered as correct and valid.

3.1. Response Times and Success Rates

The overall response times n_Δ ($M = 450.8$ ms, $SD = 180.4$ ms) did not appear to be normal distributed with skewness of 1.2 and kurtosis of 4.6. There were only very few responses faster than 100 ms but many slower than 1 s. The main results grouped by the category of stimuli are shown in Figure 3. Panned tones resulted in the shortest overall mean response time ($M = 338$ ms, $SD = 118$ ms). The chord stimuli resulted in the longest observed time spans ($M = 533$ ms, $SD = 198$ ms).

Looking at the success rate gives the same result of panned tones having the highest success rate of 95% compared to the *Chords* stimuli with 79%.

Comparing the average response time over all presented stimuli, it turned out that the participants learned to react faster to the presented stimuli. Yielding an average response time of 466 ms during the first run, the second run already resulted in 448 ms and the third repetition ended up with 440 ms. This describes a noticeable decrease in reaction time of about 6% in the course of three repetitions. This effect is even more prominent in the complete data set (including the outliers) which includes response times of several seconds. After three runs the mean response time went

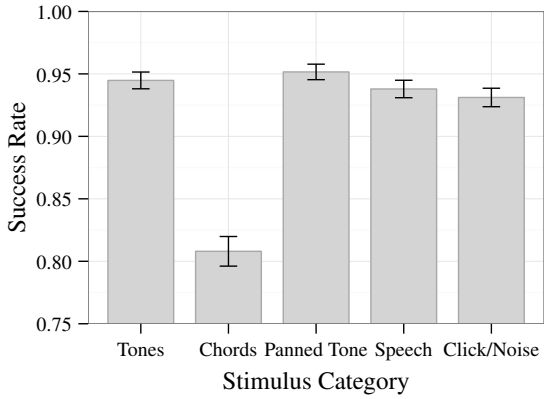

Figure 2: Success rates of the observations grouped by the stimulus category including standard errors.

down by 12%. A general comment by the subjects was that even after training due to performing the test, distinguishing minor from major chords was still a challenging task.

The level of statistical significance was evaluated by fitting a linear regression model (LM) to the observed overall response times n_Δ as follows:

$$n_\Delta \sim \text{category} \cdot \text{stimulus} + \text{musicality} + \text{run}. \quad (1)$$

We considered interaction effects of the category and stimulus and we wanted to evaluate the effect of the musical skill (musicality) as well as the learning effect represented by the independent variable run. The results of the LM fit is presented in Table 3. The fit shows that the *Panned Tone* stimuli explains a large part of the variance of the category level and differs significantly ($p < .001$) from the other categories. Furthermore, we found significant effects of the factor *musicality* and the number of *runs* the participants did. We cannot make more detailed statements as the participants were not selected by their level of musicality. Table 2 shows the effect sizes Partial η^2 of the factors.

	Partial η^2
category	0.15
stimulus	0.00
musicality	0.07
run	0.01
category:stimulus	0.03

Table 2: Effect sizes of the linear regression model based on equation1

A deeper look into the interaction effects of category and stimulus reveal that for most of the categories there is a significant difference between the stimulus representing the *left* and the one representing the *right* direction compared to *straight* which is included in the intercept of the model. As expected, the interaction of the *stimulus* and *panned tones* category explains only a small part of the Estimate in the model as shown in Table 2. On the other hand the *left* and *right* stimulus of the *Chord* category has a significant influence: the estimates of 122 ms and 168 ms indicate that the response time increases significantly when using the minor and major chords compared to the tone representing *straight*.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	638	11	55.57	**<.001**
catChord	4	12	0.32	.75
catPan	-94	12	-7.65	**<0.001**
catSpeech	-3	12	-0.22	.82
catClicks	-7	12	-0.58	.56
stimL	-40	12	-3.33	**<.001**
stimR	-47	12	-4.03	**<.001**
musicality	-24	1	-19.97	**<.001**
run2	-21	5	-3.97	**<.001**
run3	-33	5	-6.07	**<.001**
catChord:stimL	122	18	6.88	**<.001**
catPan:stimL	-36	17	-2.10	**.03**
catSpeech:stimL	36	17	2.16	**.03**
catClicks:stimL	49	17	2.88	**.004**
catChord:stimR	168	17	9.64	**<.001**
catPan:stimR	-3	17	-0.20	.84
catSpeech:stimR	60	17	3.63	**<.001**
catClicks:stimR	69	17	4.14	**<.001**

Table 3: Results of a linear regression model fit based on Equation 1. Significant p values according to the $p = .05$ level are marked in bold font

To statistically analyse the success rate as independent variable a reduced generalized linear model (GLM) has been chosen based on:

$$\text{Success} \sim \text{category} \cdot \text{stimulus}. \quad (2)$$

The fit was calculated by adding a binary logit link to the model. The results of the GLM fit can be seen in Table 5. We can see that the success rate is explained mainly by the high error rate of the presented *Chord* stimuli. The interaction of the category *Panned Tones* and the *left* and *right* stimulus show significant effects. This indicates a similar effect as seen in the results of the response times: the participants had more problems detecting the *straight* stimuli than the hard panned directions. Table 4 shows the effect sizes Partial η^2 of the factors.

	Partial η^2
category	0.59
stimulus	0.12

Table 4: Effect sizes of the binary logit regression model based on equation2

4. DISCUSSION

The reported measurements illustrate the diversity in reaction times and accuracy in the identification of sounds in auditory displays. The selection of sounds for functional purposes is thus an important criterion and not only a matter of taste. Whilst saliency and efficacy of sounds is considered in several studies [13, 14] and also cognitive load of auditory signals is investigated, the authors are not aware of any research measuring response times towards sounds in navigation or similar tasks. The simple experimental design was able to show the differences in performance after a short learning period. Learning, however, continued after the actual learning phase as the decrease in reaction time of about 6%

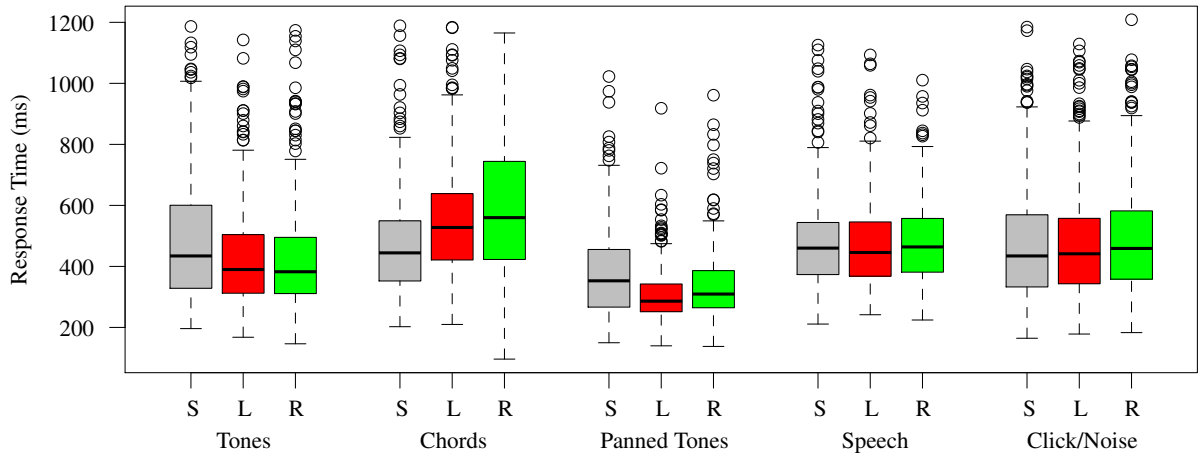

Figure 3: Response times of all correct observations grouped by the stimulus category and the stimulus direction **S**traight, **L**eft and **R**ight

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.90	0.23	12.63	**<.001**
catChord	0.01	0.33	0.03	.98
catPan	-0.85	0.28	-3.02	**.003**
catSpeech	0.10	0.33	0.32	.75
catClicks	-0.18	0.31	-0.58	.56
stimL	0.08	0.34	0.24	.81
stimR	-0.22	0.30	-0.72	.47
catChord:stimL	-2.02	0.43	-4.71	**<.001**
catPan:stimL	2.53	0.63	4.03	**<.001**
catSpeech:stimL	-0.39	0.46	-0.86	.39
catClicks:stimL	-0.10	0.45	-0.23	.82
catChord:stimR	-1.59	0.40	-3.95	**<.001**
catPan:stimR	1.65	0.45	3.64	**<.001**
catSpeech:stimR	-0.30	0.43	-0.70	.49
catClicks:stimR	-0.09	0.41	-0.22	.82

Table 5: Results of a GLM model fit based on Equation 2. Significant p values according to the $p = .05$ level are marked in bold font.

over the three repetitions showed. Future tests might therefore start with an extended learning phase. The effect of musicality could not be analyzed in detail here because the subjects reported very similar musicalities.

5. CONCLUSION AND FUTURE WORK

Our results can help to further improve both route guidance systems and navigation systems for blind people by exploiting spatial cues in directing them. As panned tones were detected more accurately and faster than speech, spoken commands may be replaced by other sounds such as sine tone. The experimental design was abstracted from a real situation. A more realistic situation might be tested applying the SONEX framework [11] in the walking game [11] in order to validate the ranking of stimuli as identified in this study,

In most navigation scenarios, speed, however, does not play a crucial role. Though, there are some cases in which speed is important. Two examples are given here: Rally driving and collision avoidance. In rallies, the co-driver usually gives commands to the driver such as "left", "right", "half-left". This could be replaced by a device which allows the co-driver to present panned sounds to headphones of the driver, of course allowing for more directions than only "left", "straight", and "right". Further experiments are needed to prove the benefit of those sounds over the speech of the co-driver and limitations in the resolution of angles in the sound panning. Visually impaired people could be quickly warned about fast approaching objects or obstacles in the way.

The current experiment measured response times towards presented stimuli. The experimental design did not allow for discrimination between recognition and processing of the perceived sounds which lead to the button presses. Following experiments will be designed for this discrimination and also investigate the mental load involved in detection and processing of the sounds. Furthermore, future experiments could reveal if adding spatial cues generally will improve the response times for all categories of stimuli. This first experiment used an arbitrary selection of a variety of stimuli; some of those stimuli were hard to distinguish, particularly the chords. It is planned for future studies to use other sounds than presented here, including noise bursts, panned speech, and more complex sounds which can convey additional information.

6. REFERENCES

[1] T. Hermann, "Taxonomy and definitions for sonification and auditory display," in *Proceedings of the 14th International Conference on Auditory Display (ICAD 2008)*, P. Susini and O. Warusfel, Eds. IRCAM, 2008.

[2] 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/

[3] M. Bujacz, P. Skulimowski, and P. Strumillo, "Naviton—a prototype mobility aid for auditory presentation of three-dimensional scenes to the visually impaired," *J. Audio Eng. Soc*, vol. 60, no. 9, pp. 696–708, 2012.

[4] D. El-Shimy, F. Grond, A. Olmos, and J. Cooperstock,

"Eyes-free environmental awareness for navigation," *Journal on Multimodal User Interfaces*, vol. 5, pp. 131–141, 2012.

[5] D. Storek, F. Rund, T. Barath, and S. Vitek, "Virtual auditory space for visually impaired - methods for testing virtual sound source localization," in *Proceedings of the International Conference on Auditory Display (ICAD)*, 2013, pp. 33–36.

[6] F. Grond and J. Berger, "Parameter mapping sonification," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany: Logos Publishing House, 2011, ch. 15, pp. 363–397. [Online]. Available: http://sonification.de/handbook/chapters/chapter15/

[7] T. Hermann, "Model-based sonification," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany: Logos Publishing House, 2011, ch. 16, pp. 399–427. [Online]. Available: http://sonification.de/handbook/chapters/chapter16/

[8] S. Holland, D. R. Morse, and H. Gedenryd, "Audiogps: Spatial audio navigation with a minimal attention interface," *Personal and Ubiquitous Computing*, vol. 6, no. 4, pp. 253–259, 2002, personal Ub Comp 1617-4909.

[9] S. Hourlier, J. Meehan, A. Léger, and C. Roumes, "Relative effectiveness of audio tools for fighter pilots in simulated operational flights," in *A Human Factors Approach. In New Directions for Improving Audio Effectiveness. Meeting Proceedings RTO-MP-HFM-123*, Neuilly-sur-Seine, France, 2005, pp. 23–1 – 23–8.

[10] B. Simpson, D. Brungart, R. Gilkey, and R. McKinley, "Spatial audio displays for improving safety and enhancing situation awareness in general aviation environments." in *New Directions for Improving Audio Effectiveness. Meeting Proceedings RTO-MP-HFM-123*, Neuilly-sur-Seine, France, 2005, pp. 26–1 – 26–16.

[11] N. Degara, F. Nagel, and T. Hermann, "Sonex: An evaluation exchange framework for reproducible sonification," in *The 19th International Conference on Auditory Display (ICAD-2013)*, Lodz, 2013, pp. 167–174.

[12] N. Degara, F. Nagel, and T. Kuppanda, "A framework for evaluating sonification methods in blind navigation," in *4th Interactive Sonification Workshop (ISon 2013)*, 2013.

[13] F. Grond, O. Kramer, and T. Hermann, "Balancing salience and unobtrusiveness in auditory monitoring of evolutionary optimization," *Journal of the Audio Engineering Society*, vol. 60, no. 7/8, pp. 531–539, 2012.

[14] J. Gillard and M. Schutz, "Impoving the efficacy of auditory alarms in medical devices by exploring the effect of amplitude envelope on learning and retention," in *Proceedings of the International Conference on Auditory Display (ICAD)*, 2012.