Using Fitts' Law to Compare Sonification Guidance Methods for Target Reaching Without Vision

Coline Fons[1,2], Sylvain Huet[2], Denis Pellerin[2] and Christian Graff[1]
[1]*Univ. Grenoble Alpes, Univ. Savoie Mont-Blanc, CNRS, LPNC, 38000 Grenoble, France*
[2]*Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France*

Keywords: Spatial Interaction, Fitts' Law, Sound Guidance, Sonification, Visual Impairment, User Performance Evaluation.

Abstract: Visually impaired people often face challenges in spatial interaction tasks. Sensory Substitution Devices assist them in reaching targets by conveying spatial deviations to the target through sound. Typically, sound guidance systems are evaluated by target reaching times. However, reaching times are influenced by target size and user-target distance, which varies across studies. We propose to explore the potential of Fitts' law for evaluating such systems. In a preliminary experiment, visually impaired and sighted participants used non-spatialized sonification to reach 3D virtual targets. Movement times were correlated with the Index of Difficulty, confirming that Fitts' law is a valuable model to evaluate target-reaching in 3D non-visual interfaces, even with non-spatialized sonification as feedback. In a second experiment, we compared two non-spatialized sonifications, one dissociating the height and azimuthal direction of the target, and the other combining them into a single 3D angle. Fitts law did allow the comparison of performance in favor of the first sonification. The potential of using Fitts' law to compare performances across studies using different experimental settings deserves exploration in future research. We encourage researchers to provide the full linear regression equations obtained when using Fitts' law, to facilitate standardized performance comparisons across studies.

1 INTRODUCTION

Visually Impaired People (VIP) may experience difficulties when they engage in spatial interaction tasks. Some assistive technologies, like Sensory Substitution Devices (SSDs) offer a solution for assisting VIP in various tasks. Guidance SSDs provide information captured by an artificial sensor through a functional sensory modality (e.g. audition or touch). In the specific case of target-reaching tasks, spatial information about the deviation between the user's position and the target is usually converted into sound, to guide the hand towards the target.

These sound guidance systems are usually evaluated by measuring performance by target-reaching time and success rate. However, experimental parameters such as target size and distance from the target can have an impact on target-reaching time, independently of the sound guidance used. We need a standardized measure, independent of experimental parameters, to evaluate and compare sound guidance systems. Fitts' law is a widely used human prediction model in HCI (Human-Computer Interaction), which predicts the time taken to point at a target depending on its size and distance to the user. In this paper, we explore the potential of Fitts' law to evaluate different sound guidance systems in an identical experimental design.

In the remainder of this section, we will introduce SSDs that use sound to guide individuals towards targets (section 1.1). We will then explain Fitts' law and how it could be used to evaluate pointing movements performance in target-reaching tasks with auditory feedback (section 1.2). Next, we will present our preliminary experiment that aims to establish whether Fitts' law applies to pointing movements without visual feedback when non-spatialized sonification is employed to guide users to a 3D target (section 2). Then, in a comparative experiment (section 3), we will demonstrate the application of Fitts' law as a standardized metric for evaluating sound guidance systems by comparing the performances of two types of sonification in a 3D virtual target reaching task. Finally, we will summarize the findings from the two experiments (section 4).

Fons, C., Huet, S., Pellerin, D. and Graff, C.
Using Fitts' Law to Compare Sonification Guidance Methods for Target Reaching Without Vision.
DOI: 10.5220/0012346400003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 1: GRAPP, HUCAPP and IVAPP*, pages 444-454
ISBN: 978-989-758-679-8; ISSN: 2184-4321

1.1 Sensory Substitution Devices for Guidance to a Target

In an Sensory Substitution Device (SSD) designed to guide a blind person to a target object through sound, the sensor collecting spatial data is usually a camera, and is also the pointer (the reference frame in which the target is positioned). The camera may be positioned on the participant's head (Katz et al., 2012; Liu et al., 2018; Thakoor et al., 2014; May et al., 2019; Zientara et al., 2017; Guezou-Philippe et al., 2018), or on their hand or forearm (Hild and Cheng, 2014; Manduchi and Coughlan, 2014; Shih et al., 2018; May et al., 2019; Zientara et al., 2017).

Once the spatial data collected, it may be encoded into sound. The sounds used can be either verbal instructions (e.g. ”turn” right", ”go forward", etc.) (Hild and Cheng, 2014; Katz et al., 2012; Manduchi and Coughlan, 2014; Shih et al., 2018; Thakoor et al., 2014; Troncoso Aldas et al., 2020; Zientara et al., 2017) or sonification. Sonification corresponds to the use of non-verbal sounds to convey perceptual information or data (Parseihian et al., 2016). There are two types of sonification (Parseihian et al., 2016):

- Spatialized sonification consists in using the natural capacities of the auditory system to locate the position of a target by virtually restituting its position (e.g. by stereo or head-related transfer function (HRTF))(Katz et al., 2012; Liu et al., 2018; Lokki and Grohn, 2005; May et al., 2019). In other words, the participant is given the impression that the sound they hear is emitted from the target's position in space, for example by adjusting the intensity of the sound in the right and left ears.
- Non-spatialized sonification uses physical characteristics of sounds, such as pitch, intensity, tempo, etc., to convert spatial information (Katz et al., 2012; Liu et al., 2018; Lokki and Grohn, 2005; Manduchi and Coughlan, 2014; May et al., 2019; Troncoso Aldas et al., 2020). For example, one can use the tempo to indicate the distance between the participant and the target, with a slow tempo when the participant is far from the target, increasing as they get closer to the target.

The auditory system's ability to localize sound sources being limited (Middlebrooks, 2015), we chose to use non-spatialized sonification instead, considering the auditory system's high performance in perceiving certain sound parameters, such as pitch, intensity, etc. (Ziemer and Schultheis (2019a) for review).

1.2 Evaluation of Sound Guidance Systems

Sound guidance systems' performances can be evaluated by real (Thakoor et al., 2014; Hild and Cheng, 2014; Manduchi and Coughlan, 2014; Shih et al., 2018; Zientara et al., 2017; Troncoso Aldas et al., 2020) or virtual (Liu et al., 2018; May et al., 2019; Lokki and Grohn, 2005) target-reaching tasks. To evaluate performance quantitatively, it is common practice to measure target-reaching times on many successive trials (Liu et al., 2018; Thakoor et al., 2014; Manduchi and Coughlan, 2014; Shih et al., 2018; May et al., 2019; Zientara et al., 2017; Troncoso Aldas et al., 2020; Lokki and Grohn, 2005), success rates (Hild and Cheng, 2014; Shih et al., 2018; May et al., 2019; Troncoso Aldas et al., 2020; Lokki and Grohn, 2005) or hand trajectories (Liu et al., 2018; May et al., 2019; Lokki and Grohn, 2005).

We focus here on another evaluation method, Fitts' law, which is an empirical model of the trade-off between speed and accuracy in target-reaching tasks. It predicts the time it takes a person to point at a target depending on the target's size and user-target distance. Fitts quantified the difficulty of movement required to reach a target and thus created an Index of Difficulty (ID):

$$ID = log_2(\frac{2A}{W}) \tag{1}$$

The unit of measurement of ID is the bit, which is the amount of binary information required to represent the difficulty of the reaching task. The letter A stands for amplitude and refers to the distance between the initial position of the pointer and the center of the target. The letter W represents the width of the target. The results of 1D target-pointing tests show a strong correlation between ID and target-reaching time. Pointing movements take longer when the target is small and remote. The complete formulation of Fitts' law is as follows:

$$MT = a + b * log_2(\frac{2A}{W}) \tag{2}$$

with MT the movement time (the time it takes to reach the target), and a- and b- constants derived empirically by linear regression. The law is associated with a performance measure known as Index of Performance (IP) or throughput, which provides information on the number of bits transferred in one second:

$$IP = \frac{ID}{MT} \tag{3}$$

Although Fitts' law was originally developed for translational movements on one dimension, it works well for 2D movements and is commonly used in the performance evaluation of pointing devices (e.g. like a mouse on a screen). Several studies also apply Fitts' law to 3D pointing (Cha and Myung, 2013; Clark et al., 2020; Barrera Machuca and Stuerzlinger, 2019; Murata and Iwase, 2001; Teather and Stuerzlinger, 2013), and pointing tasks without visual feedback (with haptic or sound feedback) (Charoenchaimonkon et al., 2010; Lock et al., 2020; Marentakis and Brewster, 2006; Hu et al., 2022). Fitts' law can therefore be used to describe the performance of a sound guidance system towards 3D targets, as it provides a good explanation of the time taken to reach the target in these experiments.

Fitts' law could therefore be an appropriate tool to evaluate and compare our sound guidance system. In a first *preliminary experiment*, we tested the applicability of Fitts' law to the characterization of 3D virtual target reach in the absence of visual feedback, using non-spatialized sonification. Participants had to reach successive virtual targets of varying size and user-target distance with their index finger, guided by sound. As this preliminary experiment showed successful for an application of Fitts' law to evaluate performance, we engaged in a *comparative experiment* to demonstrate the application of Fitts' law for evaluating sound guidance systems by comparing the performances of two types of non-spatialized sonification in an identical target-reaching task.

2 PRELIMINARY EXPERIMENT

2.1 Introduction

This first experiment aimed to test if Fitts' law could apply to 3D target reach guided by our sound guidance system. We saw that most of SSDs use a camera as both the sensor for collecting spatial data, and as a frame of reference for pointing. Here we separate the pointer from the sensor, using a motion capture system. We propose a frame of reference built around the hand, as it is central in a target-reaching task. We chose use non-spatialized sonification as a feedback, which provides more precise guidance than spatialized sonification.

To evaluate this sound guidance system, we want to use a standardized metric, independent of the target's size of the user-target distance. We saw that Fitts' law has been used to describe the performance of sound guidance system towards 3D targets (see section 1.2). However, it is uncertain whether Fitts' law applies to reaching a 3D target using only non-spatialized sonification to convey the deviations between the user and the target's position. Indeed, in (Charoenchaimonkon et al., 2010), the auditory feedback was activated only when the pointer entered or left the target. Lock et al. (2020) and Marentakis and Brewster (2006) used only spatialized sonification. Hu et al. (2022) tested the non-spatialized sonification algorithm "vOICe": video streams were sounded from left to right at a rate of one image snapshot per second. Hearing sound from the left or the right ear meant a visual object was on your left or your right side, while the pitch and the loudness was mapped to the vertical position and the brightness respectively. However, the experiment did not show that Fitts' law can be applied to this sonification, as the participants generally failed to find the targets. It is still unclear whether Fitts' law applies to reaching a 3D target both in the absence of visual feedback and using non-spatialized auditory cues to convert deviation to the target.

Taking into account not only the advice in the literature but also the cognitive load imposed on the user, we created a non-spatialized sonification in which: **(1)** The horizontal position is conveyed by the angle deviation between the direction pointed by the hand and the direction of the target. This angle is transcoded into the pitch of the sound on a continuous scale (the narrower the angle, the higher the pitch, and vice versa); **(2)** The vertical position is conveyed by the distance deviation between the height of the index finger and the height of the target. A mild binary noise is superimposed on the main auditory stream (within the same channel) when the finger is at the same height as the target; **(3)** The depth is conveyed by the distance deviation between the finger and the target. This distance is translated into sound intensity on a continuous scale (the shorter the distance, the louder the sound);

As recommended by the literature (Charoenchaimonkon et al., 2010; Lock et al., 2020; Marentakis and Brewster, 2006; Hu et al., 2022) the vertical and horizontal dimensions are orthogonal, so the participant can interpret them separately, and we used continuous scales to ensure accurate guidance. The use of binary information to encode height should reduce the amount of information to be integrated simultaneously and decrease cognitive load (Gao et al., 2022). This sonification should provide sufficiently precise guidance for the task, without imposing an excessive cognitive load on the user.

This preliminary experiment aims to test if Fitts' law applies to 3D target-reaching guided by non-spatialized sonification. To do that, we tested the sonification described above in a 3D target-reaching task in a virtual environment.

2.2 Materials and Method

Participants. Six sighted people (22.0 ± 3.6 years old) took part in the experiment. Two of them were psychology students and received a one-point bonus on a course of their choice in exchange for their participation. The other four participants were students from other disciplines recruited by the laboratory on a voluntary basis. We also recruited two visually impaired participants (57.0 ± 9.9 years old)from outside the laboratory. All participants gave their informed consent before taking part in the study.

Engineering. Tests took the form of a virtual game in which participants had to reach spheres in a 3-D space with their index finger. A Vicon optical motion capture system located reflectors that were fixed on the participant's finger and torso. Coordinates of these anatomical points were transmitted from the acquisition computer running Vicon Tracker to the pilot computer through the Virtual Reality Peripheral Network (VRPN) protocol. In the pilot computer, our C++ control software: 1) immersed the participant into the virtual environment with the target, using the OpenScene Graph (OSG) 3-D toolkit; 2) computed spatial metrics used in sound conversions; 3) transmitted them to the PureData sound system, running on the same computer, which synthesized the sounds accordingly and sent them to the participant; 4) drove the experimental protocol.

Each target was a sphere of 5.33, 8 or 12 cm diameter, depending on the condition. Targets could be placed 12 cm, 18 cm or 27 cm from the start trial button (see Figure 1, giving a total of 9 combinations of distance and target size. There were 32 different positions for each distance. Each participant performed 3 different test blocks for each target size, for a total of 9 test blocks. A test block consisted of a series of 16 consecutive targets to be reached, selected semi-randomly so that the participant reached half of all possible targets, with an equal number of targets at each distance, without repeating the same target twice for the same target size. The sequence ensured that the distance to the target varied between each trial.

Participants' Equipment. During the experiment, the participant was equipped with: **(1)** Two real objects that were fixed on their body that were located by the Vicon system using reflectors fixed on them. They defined the two points of the frame of reference: O in the middle of the torso and P at the tip of the index finger. **(2)** A bone conduction headset (Aftershokz Sportz3) connected to a receiver box (Mipro MI909R) to receive the sound feedback.

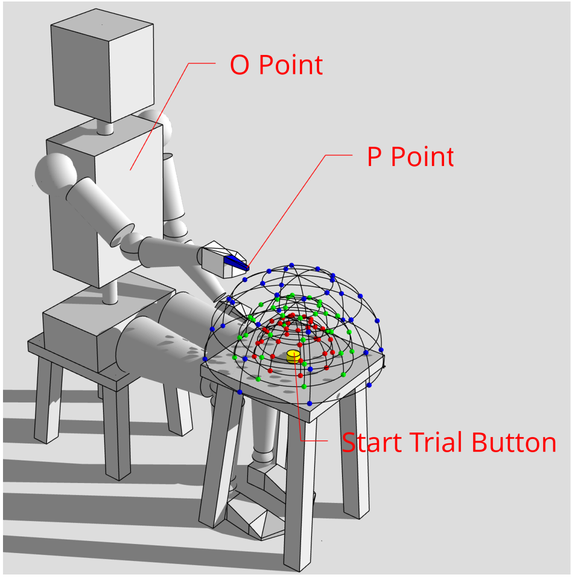


Figure 1: Illustration of the experimental setup. The $\overrightarrow{OP}$ vector define the participant's pointing direction. The distance to the target is calculated from P to the center of the target T. Possible positions of the target are represented by circles. Targets can be positioned at 12 cm (red circles), 18 cm (green circles) or 27 cm (blue circles) from the starting point (the start trial button).

Spatial Metrics. Several spatial features were extracted from the scene to be transcoded into sounds. The $\overrightarrow{OP}$ vector defined the direction of the participant's pointing and distance from target T. The spatial metrics used were:

- The angle $\theta_h = \widehat{POT_h}$, which corresponds to the projection of the angle formed by lines OP and OT on the horizontal plane parallel to the ground (angular deviation).
- The height difference $\Delta Z = |z_P - z_T|$, which corresponds to the projection of the distance [PT] on the vertical axis Z (distance deviation).
- The distance d = [PT] projected onto the horizontal plane parallel to the ground (distance deviation).

Sonification. We transcoded angle and distance deviations from the target into sound parameters. The extracted spatial information is encoded by generating a sine wave which pitch varies as a function of θ_h and intensity as a function of d. The height is encoded separately as a superimposed white noise.

The pitch f of the sine wave varies on a continuum from $fmin = 110$ Hz for $\theta_h max = 45°$ to $fmax = 440$ Hz for $\theta_h min = 2°$, so that f is within the audible spectrum while avoiding frequencies that are unpleasant with a constant stimulus.

The conversion of angle θ_h to pitch follows Stevens' (Stevens, 1957) psychophysical law, a power-law function:

$$log(f) = Af * log(\theta_h) + Bf \tag{4}$$

$$Af = \frac{log(fmax) - log(fmin)}{log(\theta_h min) - log(\theta_h max)} = 0.44 \tag{5}$$

$$Bf = log(fmin) - Af * log(\theta_h max) = 1.30 \tag{6}$$

The vertical dimension is coded by activating a mild white noise when the participant points at the same height as the target position, i.e. when $\Delta Z < r$ (i.e. the radius of the sphere).

Finally, distance from the target is encoded by sound intensity. The intensity of the sine wave varies on a continuum with a 15 dB difference between *Imin* (with *dmax* = 0.5 m) and *Imax* (with *dmin* = 0.01 m). The conversion of distance into intensity follows the following conversion function:

$$I = Ai * log(d) + Bi \tag{7}$$

$$Ai = \frac{Imin - Imax}{log(\frac{dmax}{dmin})} = -8.83 \tag{8}$$

$$Bi = Imin - Ai * log(dmax) = 82.34 \tag{9}$$

Our sonification thus dissociates the horizontal and the vertical: the horizontal dimension is coded by the pitch of the sound, while the vertical dimension is coded by the activation of an additional mild white noise when the participant's finger is at the correct height.

Additional sound cues to specify the finger's trajectory during the trials are given below.

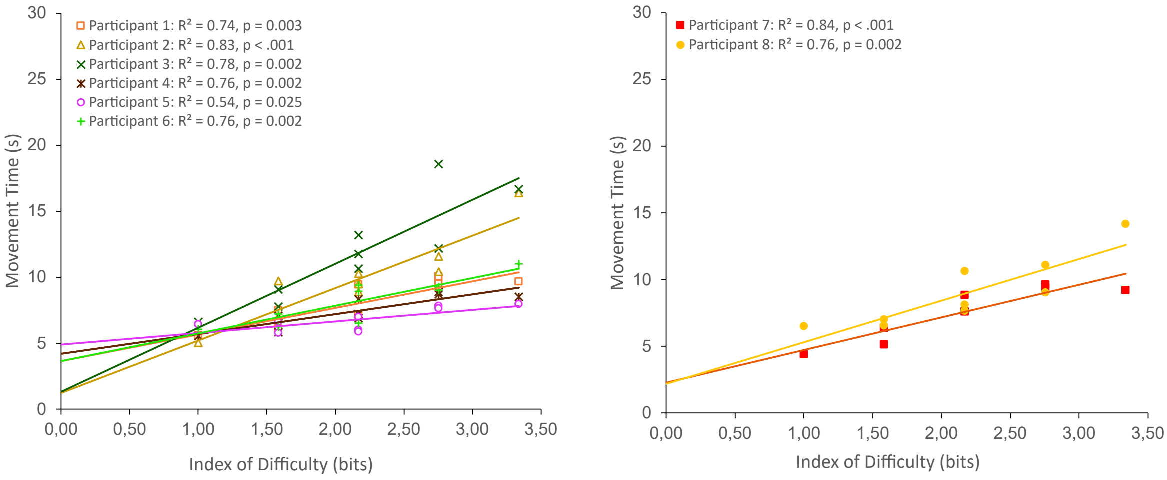
Protocol. The experiment took place in the motion capture space described above. Participants with their eyes closed had to find several targets presented one after the other. The experiment began with a training phase, consisting of a block of 4 trials per target size, starting with the largest targets and ending with the smallest. The participant could then do as many blocks of trials as they felt necessary to become comfortable with the sonification and develop a search strategy that suited them. During this phase, the participant received explanations from the experimenter on the difficulties that may arise and how to overcome them. The session continued with the completion of nine experimental blocks alternating target sizes, in a counterbalanced order between participants. Only the experimental blocks were used for statistics. From the participant's arrival to their departure, the session lasted one hour, including all the steps mentioned above.

Each block proceeded as follows: the participant sat on a chair, with a stool placed in front of them. They closed their eyes, and initiated the first trial by pressing a start command button accessible through touch, placed on the stool. A bell sound confirmed the beginning of the trial. As described in detail in section 2.2, from there, the participant heard the sound feedback that changed according to their movements. When the participant's finger passed too far over the target ([OP] > [OT]), a perfect fifth (in a musical sense) was superimposed on the sine wave. The participant therefore knew that they had passed the target, but kept access to the pitch variations and therefore to the directional information. When P (finger tip) entered the target, the sine wave was interrupted and a strong white noise was triggered. This strong white noise was clearly distinct to that indicating the correct height. When the participant's finger P remained inside the target for 100 ms (ensuring that the target was not reached by chance) a tingling sound indicated victory; the sound feedback was deactivated and the trial ended. The trial was also interrupted if the participant failed to reach the target within 60 seconds, and a different bell sound indicated defeat. After each trial, the participant started the next one by pressing the start button with their index finger P.

2.3 Results of the Preliminary Experiment

There are several ways of analyzing the results of a Fitts' task. The most common way of assessing the fit of the model to the data seems to be a linear regression between Movement Time (MT) and Index of Difficulty (ID) (see equation 2, section 1.2) by averaging MT by ID, and then interpreting the R or R^2 obtained. A high value of R or R^2 is interpreted as a good fit of the model to the data. This approach is criticized: doing this is assuming the validity of Fitts' law before testing it, as if Fitts' law is valid, different conditions of target sizes and distance to the target, but the same ID, should have the same MT (Drewes, 2010; Triantafyllidis and Li, 2021). Furthermore, the value of R or R² in itself says nothing about significance, so we used an additional F-test to verify that the R coefficient is statistically different from zero, and thus check the fit of the model. We also looked up the a- and b- constants (see equation 2). Constant a- is the reaction time, and has seconds for units. Constant b- has for units seconds per bit. It represents the time needed to process one bit of information.

Here, five ID were calculated according to the original formulation of Fitts' law (equation 1). To determine whether a Fitts-type response is occurring,



(a) Sighted participants (eyes closed)

(b) Visually impaired participants

Figure 2: Linear regression lines for mean Movement Times (MT) as a function of Index of Difficulty (ID) for each participant.

we run a linear regression between MT and ID. Reach times are averaged for each combination of size and distance (9 combinations) as Fitts did in his study (Fitts, 1954). The fit of the model to our data is confirmed, with a R^2 of 0.93 for sighted participants ($F(1,7) = 96.07$, $p < .001$), and 0.93 for visually impaired participants (($F(1,7) = 89.74$, $p < .001$). Individual regression lines of participants are represented in Fig. 2. The complete equation obtained was, for sighted participants:

$$MT = 3.17(\pm 0.59) + 2.55(\pm 0.26)ID \tag{10}$$

And for visually impaired participants:

$$MT = 2.29(\pm 0.66) + 2.75(\pm 0.29)ID \tag{11}$$

A post hoc power analysis was conducted using G*Power version 3.1.9.7 (Faul et al., 2007) for power estimation based on our results for the sighted participants. A large effect size (13.29) was calculated from the R² (.93) of the regression analysis. With a significance criterion of $\alpha = .05$, the statistical power obtained is .99. The sample size used is thus adequate to test the study hypothesis.

Finally, we computed the average IP with equation 3 for sighted and visually impaired participants. We found an average IP of 0.36 bits/s for sighted participants and 0.35 bits/s for visually impaired participants.

2.4 Discussion of the Preliminary Experiment

Our study aimed to test the applicability of Fitts’ law to characterize target-reaching in a 3D virtual environment with an auditory-feedback using non-spatialized sonification. We performed linear regressions between mean Movement Times (MT) and Index of Difficulty (ID). The results on VIP were similar, assessing the usability of the device for them. As expected, the results show that changes in ID values have a consistent effect on target-reaching time. As a result, Fitts’ law can be used to evaluate this type of sound guidance system. We could therefore compare the performance obtained in our study to other studies in the literature that used Fitts’ law. The most common way of evaluating performance using Fitts’ law is to use the Index of Performance (IP) (equation 3).

However, the use of IP to compare the performances in two different target-reaching tasks should be considered with reservation. Although the index of performance was first defined by Fitts (1954) as the average ratio of ID and MT (see equation 3), Zhai (2004) shows that its value depends not only on both a- and b- constant obtained through regression, but also on the set of ID values used. Therefore, it cannot be generalized beyond specific experimental target sizes and user-target distances. It is recommended to report the full equation of the regression analysis, with a- and b- constants, instead of using IP as the sole performance characteristic of the device (Zhai, 2004; Triantafyllidis and Li, 2021; Drewes, 2010). Fitts and Radford (1966) actually defined in a later work IP as:

$$IP = \frac{1}{b} \tag{12}$$

The b-constant corresponds to the slope of the linear relationship between ID and IP, and has the unit seconds per bit. Looking at our equations 10 and 11 for sighted and visually impaired participants results we can say that sighted participants took 2.55 seconds per bit to process, while visually impaired participants took 2.75 seconds per bit to process, or, using equation 12, that sighted participants processed 0.39 bits per second, and visually impaired participants 0.36 bits per second. These IP differ from those calculated with equation 3 and reported at the end of section 2.3.

The a- constant includes various noises in the regression, such as reaction time and motor activation time. The value of the a- constant changes according to the formula used to compute the ID. A lot of paper in HCI nowadays prefer the formulation of MacKenzie (1989):

$$ID = log_2(\frac{A}{W} + 1) \tag{13}$$

However, this formulation is questioned by several authors (Drewes, 2023; Hoffmann, 2013). Furthermore, according to Drewes (2010, 2023), the a-constant does not have the meaning of reaction time when using Mackenzie's formula. However, the interpretation of the b- constant is the same. The b-constant could therefore be a standardized metric to evaluate and compare devices' results across experimental studies, whether these studies used Fitts' formulation or Mackenzie's. Unfortunately, a lot of papers on Fitts' law do not report the full equation of the regression analysis; instead, they report only R, R² and/or IP calculated with equation 3. This is the case for Marentakis and Brewster (2006); Lock et al. (2020); Hu et al. (2022); Meijer (1992), with which a comparison would have been informative considering their studies on target-reaching guided by sound evaluated by Fitts' law.

Although using Fitts' law to compare results across studies remains a challenge, we see that the b-constant of the linear regression is a metric which is independent of experimental setting, and could therefore be used for comparison of performance between studies. Hence, we encourage authors to report the full equation of the regression analysis in future studies on Fitts' law. In order to demonstrate the use of Fitts' law as a measure of performance comparison between guidance systems, we conducted a second experiment with sighted participants. The second experiment is similar to the first, with the exception that each participant completed the task twice, once with the same sonification as in the Preliminary Experiment, and once with a new sonification, which we will describe shortly.

3 COMPARATIVE EXPERIMENT

3.1 Introduction

The aim of this second experiment was to use Fitts' law as a standardised metric to compare two sound guidance settings. We created two non-spatialized sonification and wanted to determine which one is more effective in guiding a person deprived of sight towards a virtual target in 3D space. The first sonification used was the same as in the Preliminary Experiment, which we will call Dissociated Vertical-Horizontal (DVH) as it dissociates information from the three spatial dimensions into two sound streams (within the same channel). The second sonification, Unified Vertical-Horizontal (UVH), on the contrary, encodes spatial deviations to the target on the three dimensions into a single sound stream, as recommended in the literature (Ziemer et al., 2019; Ziemer and Schultheis, 2019a,b). With UVH:

1. The horizontal and the vertical position are conveyed by the angle deviation between the direction pointed by the hand and the direction of the target. This angle is transcoded into the pitch of the sound on a continuous scale (the narrower the angle, the higher the pitch, and vice versa);
2. The depth is conveyed by the distance deviation between the finger and the target. This distance is translated into sound intensity on a continuous scale (the shorter the distance, the louder the sound);
3. An additional mild white noise is activated when the direction pointed by hand intersects the target.

Both sonifications encode the same number of spatial dimensions and contain an equivalent amount of sound information, including a sine wave and a white noise. The unification of the vertical and horizontal dimensions into a single angle is therefore the only difference between UVH and DVH.

We hypothesized that performances would be better with DVH than with UVH, as DVH allow participants to interpret the vertical and horizontal separately, and the use of binary information should reduce the amount of information to be integrated and decrease cognitive load (Gao et al., 2022).

The task was identical to that of the Preliminary Experiment, except that participants completed the experiment in two sessions, one in which the sound feedback was the UVH sonification, and one in which the sound feedback was the DVH sonification. To compare performances between the two sonifications, we used both classical reaching-times comparison, and regression analysis between MT and ID.

3.2 Materials and Method

Participants. Five sighted people took part in the experiment (19.2 ± 0.9 years old). Four of them were psychology students who received a bonus point on a course of their choice in exchange for their participation. The last one was recruited from outside the laboratory. All gave their informed consent before taking part in the study.

An a priori analysis was conducted using G*Power version 3.1.9.7 for sample size estimation, based on data from the preliminary experiment. With a significance criterion of $\alpha = .05$ and power = .95, the minimum sample size needed with the effect size obtained in the preliminary experiment (see section 2.3) is N = 5. Thus, our sample size is adequate.

Materials. For information on this section, see section 2.2 and 2.2. The only difference was that participants performed the 9 test blocks twice, once for UVH and one for DVH. The order in which sonifications were presented was counterbalanced between participants.

Spatial Metrics Spatial metrics used for DVH were identical as those described in section 2.2.

Spatial metrics used for UVH were distance d described in section 2.2, and the angle $\theta = \widehat{POT}$, which corresponds to the angle formed by lines OP and OT. Therefore, horizontal and vertical deviations were encoded into a common metric using a single auditory stream.

Sonification. DVH was identical as in the Preliminary Experiment (see section 2.2).

For UVH sonification, the spatial information extracted was encoded by generating a sine wave, just as for DVH (see section 2.2 and equations 4, 5 and 6), except that the pitch varied as a function of θ. Distance from the target was encoded by sound intensity, in the same way as for DVH (see section 2.2 and equations 7, 8 and 9). A mild white noise was superimposed on the sine wave when the OP line intersected the target.

Protocol. The same protocol as for the Preliminary Experiment (section 2.2) was repeated twice, one for the UVH sonification and one for the DVH sonification. Two participants did UVH first and two others did DVH first. The two sessions were spaced at least 24 hours apart.

3.3 Results of the Comparative Experiment

Fitts' Law. We performed a linear regression on MT as a function of ID for each sonification condition, in the same way as in the Preliminary Experiment. The plotted individual regression lines for the participants in each condition of sonification are represented in figure 3. The results show that in both conditions, MT increased linearly as a function of ID. The regression analysis of mean MT yielded significant R² values of .83 for UVH (F(1, 7) = 35.30, p < .001), and .91 for DVH (F(1, 7) = 73.27, p < .001). The complete equation obtained was, for DVH:

$$MT = 2.24(\pm 0.72) + 2.70(\pm 0.32)ID \tag{14}$$

And for UVH:

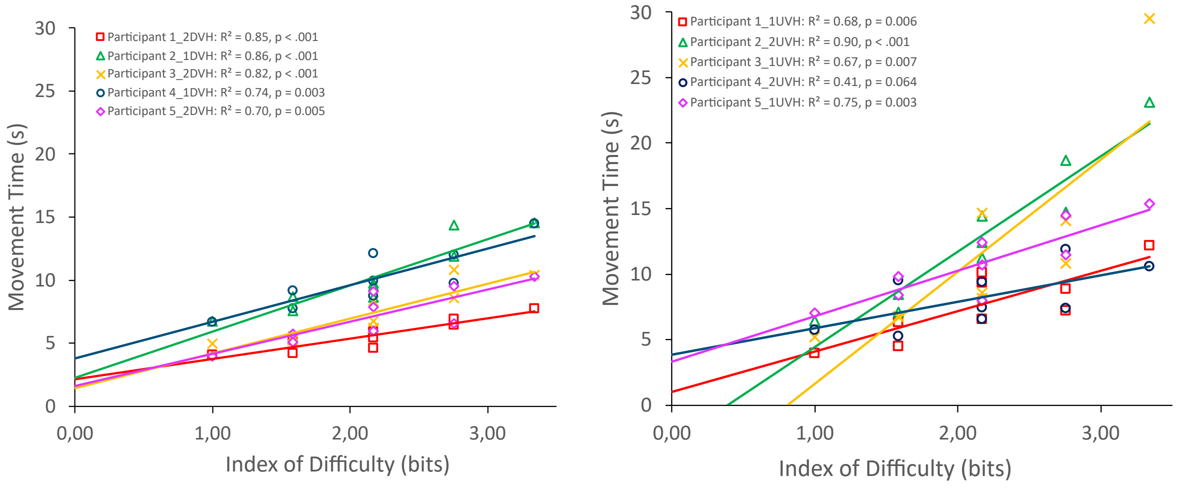
$$MT = -0.33(\pm 1.87) + 4.88(\pm 0.82)ID \tag{15}$$

Movement Times. Participants took on average 10.21 seconds (SD = 8.26 s) to reach target with UVH sonification and on average 8.09 seconds (SD = 5.09 s) with DVH sonification. To compare MT for the two types of sonification, we used a Cox model, an appropriate analysis for duration variable with skewed distribution as we have it here (Letué et al., 2018). It allows us to analyze repeated measures without averaging the data for each participant, so it accounts for intra- and inter-participant variability. To do this we used the *coxph* function in the *Survival* package of R software. Participants were significantly faster to reach targets with DVH (z = 6.14, p < .001) than with UVH.

3.4 Discussion of the Comparative Experiment

This second experiment compared performance on a 3D target reaching task with two different types of non-spatialized sonification. The following interesting points emerge from these results:

- As in our preliminary experiment, results showed that Fitts' law is a valuable model to evaluate target reaching, even with non-spatialized sonification (both DVH and UVH) as a feedback;
- Using Fitts' law, we showed that performances with DVH in the comparative experiment are similar to performances with DVH in the preliminary experiment.



(a) DVH sonification (b) UVH sonification

Figure 3: Linear regression lines for mean Movement Times (MT) as a function of Index of Difficulty (ID) for each participant.

- Using Fitts' law, we compared two different sonification conditions; in the comparative experiment, performances with DVH sonification are higher than performances with UVH sonification.

By performing a linear regression analysis between MT and ID, we were able to show Fitts' law accurately models 3D non-visual target reaching with non-spatialized sonification feedback. Indeed, in both conditions of sonification, movement times significantly increased with task difficulty (ID), which replicates the findings of the Preliminary Experiment. We find similar results for DVH sonification on this experiment as we did for sighted participants in the Preliminary Experiment, with a R^2 of 0.91 in the Comparative Experiment and 0.93 in the Preliminary Experiment, and a slope of 2.70 s/bit in the Comparative Experiment and 2.55 s/bit in the Preliminary Experiment.

We discussed in section 2.4 that the b- constant of the equation could be a good standardized metric to compare sound guidance systems. Its reciprocal (equation 12) is another version of the index of performance (IP), and is the number of bits of information processed in one second. We observe that the IP for the DVH sonification (0.37 bits/s) is almost twofold higher than the IP for the UVH sonification (0.20 bits/s). It means that participants took much more time to reach targets with UVH than with DVH as task difficulty increased. DVH sonification appears to be more efficient than UVH as a feedback, as it enables faster movement times across increasing IDs. The superiority of DVH over UVH is confirmed by comparing the MT with a Cox model, which shows significantly faster movements for DVH than for UVH.

The difference in performance between UVH and DVH replicates previous studies (Fons et al., 2023). UVH integrates spatial dimensions into a single auditory stream using continuous scales. This provides precise and fast sound guidance, but hearing all the information at the same time can create cognitive overload, amplified by the fact that the vertical and the horizontal dimensions are not orthogonal. On the other hand, DVH uses two distinct sound streams (within the same channel) to encode the vertical and the horizontal dimensions. The dimensions are therefore orthogonal, allowing the participant to interpret them separately. Despite the use of two sound streams, the fact that one of them is binary prevents participants from having to switch their attention from one to the other, decreasing the cognitive load (Gao et al., 2022). As a result, performances in the target-reaching task are better with DVH than with UVH.

4 CONCLUSION

In sound guidance systems, spatial information about the deviation between the user's position and the target is converted into sound to guide the pointing movement towards the target. These systems are usually evaluated by measuring the time taken to reach targets in a real or virtual target-reaching task in 3D space. Here, we proposed to evaluate such sound guidance systems by using Fitts' law as a standardized metric.

In a first preliminary experiment, sighted and visually impaired participants had to reach 3D virtual targets with their index finger guided by sounds. Results showed that Fitts' law is a valuable model to evaluate target-reaching, even in 3D non-visual interfaces with non-spatialized sonification as a feedback. In a second comparative experiment, we used Fitts' law as a standardized metric to compare performances of two sound guidance systems, using the slope of the linear regression between Movement Time (MT) and Index of Difficulty (ID). Results showed the advantage of using a non-spatialized sonification that dissociates the vertical and horizontal information on the position of the target into two sound streams (within the same channel) over a non-spatialized sonification that uses a single metric and sound stream for both dimensions.

Here, we have demonstrated the utility of Fitts' law in comparing the performance of different sound guidance systems within the same experimental conditions. While target-reach time, a commonly used metric for comparing guidance systems, is influenced by target size and user-target distance, these experimental parameters vary widely across studies. Therefore, the potential of utilizing Fitts' law to compare performance across studies using different experimental settings deserves exploration in future research. As a general practice, we encourage authors to provide complete regression equations when employing Fitts' law.

ACKNOWLEDGEMENTS

This work was supported by the Agence Nationale de la Recherche (ANR-21CE33-0011-01). It was authorized by the ethical committee CER Grenoble Alpes (Avis-2018-06-19-1).

This work has been partially supported by ROBOTEX 2.0 (Grants ROBOTEX ANR-10-EQPX-44-01 and TIRREX ANR-21-ESRE-0015) funded by the French program Investissements d'avenir.

REFERENCES

Barrera Machuca, M. D. and Stuerzlinger, W. (2019). The effect of stereo display deficiencies on virtual hand pointing. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–14.

Cha, Y. and Myung, R. (2013). Extended fitts' law for 3d pointing tasks using 3d target arrangements. *International Journal of Industrial Ergonomics*, 43(4):350–355.

Charoenchaimonkon, E., Janecek, P., Dailey, M. N., and Suchato, A. (2010). A comparison of audio and tactile displays for non-visual target selection tasks. In *2010 International Conference on User Science and Engineering (i-USEr)*, pages 238–243. IEEE.

Clark, L. D., Bhagat, A. B., and Riggs, S. L. (2020). Extending fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology. *International journal of human-computer studies*, 139:102413.

Drewes, H. (2010). Only one fitts' law formula please! In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, page 2813–2822, New York, NY, USA. Association for Computing Machinery.

Drewes, H. (2023). The fitts' law filter bubble. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23, New York, NY, USA. Association for Computing Machinery.

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2):175–191.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381.

Fitts, P. M. and Radford, B. K. (1966). Information capacity of discrete motor responses under different cognitive sets. *Journal of Experimental psychology*, 71(4):475.

Fons, C., Huet, S., Pellerin, D., Gerber, S., and Graff, C. (2023). Moving towards and reaching a 3-d target by embodied guidance: Parsimonious vs explicit sound metaphors. In *International Conference on Human-Computer Interaction*, pages 229–243. Springer.

Gao, Z., Wang, H., Feng, G., and Lv, H. (2022). Exploring sonification mapping strategies for spatial auditory guidance in immersive virtual environments. *ACM Transactions on Applied Perceptions (TAP)*.

Guezou-Philippe, A., Huet, S., Pellerin, D., and Graff, C. (2018). Prototyping and evaluating sensory substitution devices by spatial immersion in virtual environments. In *VISIGRAPP 2018-13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS-Science and Technology Publications.

Hild, M. and Cheng, F. (2014). Grasping guidance for visually impaired persons based on computed visual-auditory feedback. In *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, volume 3, pages 75–82. IEEE.

Hoffmann, E. R. (2013). Which version/variation of fitts' law? a critique of information-theory models. *Journal of motor behavior*, 45(3):205–215.

Hu, X., Song, A., Wei, Z., and Zeng, H. (2022). Stereopilot: A wearable target location system for blind and visually impaired using spatial audio rendering. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:1621–1630.

Katz, B. F., Kammoun, S., Parseihian, G., Gutierrez, O., Brilhault, A., Auvray, M., Truillet, P., Denis, M., Thorpe, S., and Jouffrais, C. (2012). Navig: augmented reality guidance system for the visually impaired. *Virtual Reality*, 16(4):253–269.

Letué, F., Martinez, M.-J., Samson, A., Vilain, A., and Vilain, C. (2018). Statistical methodology for the analysis of repeated duration data in behavioral studies. *Journal of Speech, Language, and Hearing Research*, 61(3):561–582.

Liu, Y., Stiles, N. R., and Meister, M. (2018). Augmented reality powers a cognitive assistant for the blind. *ELife*, 7:e37841.

Lock, J. C., Gilchrist, I. D., Gilchrist, I. D., Cielniak, G., and Bellotto, N. (2020). Experimental analysis of a spatialised audio interface for people with visual impairments. *ACM Transactions on Accessible Computing (TACCESS)*, 13(4):1–21.

Lokki, T. and Grohn, M. (2005). Navigation with auditory cues in a virtual environment. *IEEE MultiMedia*, 12(2):80–86.

MacKenzie, I. S. (1989). A note on the information-theoretic basis for fitts' law. *Journal of motor behavior*, 21(3):323–330.

Manduchi, R. and Coughlan, J. M. (2014). The last meter: blind visual guidance to a target. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3113–3122.

Marentakis, G. N. and Brewster, S. A. (2006). Effects of feedback, mobility and index of difficulty on deictic spatial audio target acquisition in the horizontal plane. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 359–368.

May, K. R., Sobel, B., Wilson, J., and Walker, B. N. (2019). Auditory displays to facilitate object targeting in 3d space. In *The 25th International Conference on Auditory Display (ICAD 2019)*. Georgia Institute of Technology.

Meijer, P. B. (1992). An experimental system for auditory image representations. *IEEE transactions on biomedical engineering*, 39(2):112–121.

Middlebrooks, J. C. (2015). Sound localization. *Handbook of clinical neurology*, 129:99–116.

Murata, A. and Iwase, H. (2001). Extending fitts' law to a three-dimensional pointing task. *Human movement science*, 20(6):791–805.

Parseihian, G., Gondre, C., Aramaki, M., Ystad, S., and Kronland-Martinet, R. (2016). Comparison and evaluation of sonification strategies for guidance tasks. *IEEE Transactions on Multimedia*, 18(4):674–686.

Shih, M.-L., Chen, Y.-C., Tung, C.-Y., Sun, C., Cheng, C.-J., Chan, L., Varadarajan, S., and Sun, M. (2018). Dlwv2: A deep learning-based wearable vision-system with vibrotactile-feedback for visually impaired people to reach objects. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1–9. IEEE.

Stevens, S. (1957). On the physiological law. *Psychol. Rev*, 64:153–181.

Teather, R. J. and Stuerzlinger, W. (2013). Pointing at 3d target projections with one-eyed and stereo cursors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 159–168.

Thakoor, K., Mante, N., Zhang, C., Siagian, C., Weiland, J., Itti, L., and Medioni, G. (2014). A system for assisting the visually impaired in localization and grasp of desired objects. In *European Conference on Computer Vision*, pages 643–657. Springer.

Triantafyllidis, E. and Li, Z. (2021). The challenges in modeling human performance in 3d space with fitts' law. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–9.

Troncoso Aldas, N. D., Lee, S., Lee, C., Rosson, M. B., Carroll, J. M., and Narayanan, V. (2020). Aiguide: An augmented reality hand guidance application for people with visual impairments. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–13.

Zhai, S. (2004). Characterizing computer input with fitts' law parameters—the information and non-information aspects of pointing. *International Journal of Human-Computer Studies*, 61(6):791–809.

Ziemer, T., Nuchprayoon, N., and Schultheis, H. (2019). Psychoacoustic sonification as user interface for human-machine interaction. *arXiv preprint arXiv:1912.08609*.

Ziemer, T. and Schultheis, H. (2019a). Psychoacoustical signal processing for three-dimensional sonification. In *The 25th International Conference on Auditory Display (ICAD 2019)*. Georgia Institute of Technology.

Ziemer, T. and Schultheis, H. (2019b). Three orthogonal dimensions for psychoacoustic sonification. *arXiv preprint arXiv:1912.00766*.

Zientara, P. A., Lee, S., Smith, G. H., Brenner, R., Itti, L., Rosson, M. B., Carroll, J. M., Irick, K. M., and Narayanan, V. (2017). Third eye: A shopping assistant for the visually impaired. *Computer*, 50(2):16–24.