Experimental Evaluation of Three Interaction Channels for Accessible Digital Musical Instruments

Nicola Davanzo(✉) and Federico Avanzini

Laboratory of Music Informatics, Department of Computer Science,
University of Milan, Via Celoria 18, 20133 Milan, Italy
`{nicola.davanzo,federico.avanzini}@unimi.it`

Abstract. Accessible Digital Musical instruments (ADMIs) dedicated to people with motor disabilities represent a relevant niche in accessibility research. The designer is often required to exploit unconventional physical interaction channels, different from hands and fingers. Although comprehensive evaluation methods for Digital Musical Instruments in general are found in literature, little has been done both in ADMIs evaluation and the analysis of suitable interaction channels from a Human-Computer Interaction perspective. In this work the performance of breath, gaze pointing and head movements is analyzed, in terms of movement speed and stability, through a simple experiment. These interaction channels could be exploited in the design of ADMIs dedicated to quadriplegic musicians. The proposed experiment has similarities with past Fitts Law evaluation tests. Results are discussed proposing possible mappings between channels and musical performance parameters. These results could also be useful to inform the design of different interface types.

Keywords: Accessible interfaces · Accessible Digital Musical Instruments · Performance metrics · Fitts' Law

1 Introduction

Among the technologies dedicated to people with physical, cognitive, and sensory impairments, ADMIs (Accessible Digital Musical Instruments) are gradually gaining attention, while the research on this topic has expanded considerably in recent years [4]. Musical activities provide benefits in terms of health, learning and concentration skills, while being an important cultural and social inclusion factor [2,11,21]. Frid's work [4] shows that a considerable percentage of existing ADMIs are dedicated to users with physical impairments. Such instruments, particularly those dedicated to quadriplegic performers, require to exploit interaction channels other than fingers, which are common to many traditional acoustic musical instruments: gaze pointing, head and face movement, breath, tongue, EEG, etc. Within the literature of Digital Musical Instruments (DMIs) in general, one challenge is the definition of proper evaluation metrics [14,17,20].

K. Miesenberger et al. (Eds.): ICCHP 2020, LNCS 12377, pp. 437–445, 2020.
https://doi.org/10.1007/978-3-030-58805-2_52

The problem is even more complex for ADMIs. In particular, little work was done to evaluate interaction channels suitable for quadriplegic users.

A widely used model in HCI research is Fitts Law, which predicts human performance in target acquisition tasks, in terms of *throughput*. Its most recent formulation [12] is as follows:

$$T = a + b \log_2 \left(\frac{D}{W} + 1 \right) \tag{1}$$

where T is the movement execution time, D and W are the target distance and width, respectively, while a and b are empirical constants. The logarithmic term is usually referred to as ID (Index of Difficulty). The ISO 9421-9 experimental procedure [12,20], often used to verify Fitts' Law, prescribes an experimental methodology employing 1D and 2D target selection tasks. In light of our specific objectives, the experiment reported next is not fully compliant with the standard. Specifically: (a) some of the analyzed channels offer only one degree of freedom; (b) an additional one would be necessary to make a selection (e.g. mouse click), which is not possible with the investigated channels; (c) we are also interested in movement *stability*; (d) the target width has been kept fixed to standardize stability measurements. The introduction of stability measurement denotes an important difference with Fitts' Law related experiments: an high stability, which is not obvious given the nature of the channels and sensors, is potentially very useful in musical interfaces interaction, e.g. for keeping a stable pitch or volume selection.

Gaze, breath, and head interactions have been analyzed in previous works. Gaze pointing is a consolidated interaction channel [18], and several gaze based ADMIs exist [3,19]. Hornof [5] and Baath [1] evaluate gaze in rhythmic tasks. Breath is widely used in traditional instruments, as well as in accessible interfaces [9,13]. Head movement is also employed for instance in wheelchair control [13] and is used in ADMIs control [6,8]. Its kinematics has been widely studied [16].

The main contribution for this work is the evaluation of three alternative channels (gaze pointing, breath, head movements), in terms of two characteristics useful for musical interaction: movement *speed* and *stability*. The evaluation occurred through a simple experiment, described in Sect. 2. Results are exposed in Sect. 3. In light of these, possible usages of the aforementioned in musical interfaces and ADMIs are discussed in Sect. 4. It should be however highlighted that these results could be useful also to inform the design of different types of interfaces other than musical ones.

2 Experiment

2.1 Procedure

The experimental GUI is shown in Fig. 1. Interaction takes place on the *interaction bar* (dark gray). A marker (in red) indicates the center of the *target zone*

Fig. 1. A screenshot of the interface used to run the experiment. (Color figure online)

(light gray). The *cursor* (in white) can only be moved horizontally within the limits of this bar, using one interaction condition.

Seven conditions were considered (see Sect. 2.2 for details): hand (considered solely as a benchmark for the remaining ones), gaze (raw or smoothed), breath, head (yaw, pitch, roll). For each condition, five target distances were tested: *D200, D300, D500, D700* and *D900* (numbers indicate distance, as a proportion of D1000). Subjects were required to (1) move the cursor to the target as quickly as possible; (2) once reached, keep it in the center of the target as stably as possible. A within-subject procedure was used. Each subject performed 15 trials for each condition. The first 5 trials were used as training and discarded in subsequent analysis, while the remaining 10 presented the 5 distances twice. Conditions and distances were randomized across subjects.

Each trial was initiated by the subject through a key press, after which a target appeared on the far left. The subject had to position the cursor inside the target for two seconds. The test then entered two main phases. **Selection phase:** an acoustic warning signalled the appearance of a new target to the right at a given distance, and the subject had 5 seconds to make a target selection (after this time, the trial was declared as failed); a selection was valid only if the cursor remained within the target for at least 1 second (not counted in the reported selection time). **Stability phase:** after selection, the subject had to keep the cursor as centered and stable as possible on the target, for 2 s, after which a new acoustic signal notified the end of the trial.

A one minute break was provided between conditions, to switch the setup. The total session time was around 30 min. At the end of the session, a questionnaire was presented with 6 questions, with answers provided on a 7-value Likert scale (higher is better), to investigate personal perception of testers on each interaction channel. Questions were as follows:

(a) **Fatigue.** Did you feel fatigue, tiredness, or pain during the execution? (High = not fatiguing)
(b) **Usability.** Did you find the interaction easy and comfortable or frustrating? (High = easy)
(c) **Precision.** Did you find the interaction accurate and precise? (High = very precise)
(d) **Involuntary movements**. During the test, did you perceive involuntary movements which affected the position of the cursor? (High = no)
(e) **Speed.** Did you find the selection method quick and fast? (High = very fast)

(f) **General opinion.** Give a general rating of these channels as interaction methods. (High = excellent)

2.2 HW/SW Setup

The experiment was run on an Apple MacBook Pro (2017) with Windows 10 (started via Bootcamp) and a dual-core Intel Core i5 CPU at 2.3 GHz, 8 GB RAM LPDDR3 at 2133 MHz, Intel Iris Graphics 640 GPU with 1536 MB VRAM. Sensors and screen were connected via a Thunderbolt/USB+VGA port adapter. A 21-inch VGA monitor at 1920 × 1080 px resolution was used. The physical length of the interaction bar on the screen was 24.7 cm (out of a 47.7 cm screen width). The software (test automation, data recording, GUI) was developed in C#.[1]

Interaction occurring through a physical channel is inextricably linked to the type, setup and quality of the used sensor. Thus, the experiment actually evaluates channel-sensor pairs.

Hands (Mouse). We used a high-end gaming mouse, namely a *Corsair M65 Pro* equipped with a 12,000 DPI resolution optical sensor. Sensitivity was set to 1500 DPI.

Gaze point (Eye Tracker). Only horizontal movements were evaluated. We used a *Eye Tribe* device, with a sampling rate of 60 Hz and an accuracy of 0.5-1.0° on the visual field [15]. Two setups were used, raw and smooth. In the latter, a natively available smoothing filter was activated.

Breath (Breath sensor). An *ad-hoc* sensor was built using a *NXP MPX5010DP* low-pressure sensor with a range of $0-10$ KPa, a sensitivity of 1 mV/mm and a response time of 1 ms. It was interfaced to the computer through an *Arduino Uno* microcontroller. A rubber tube with an interchangeable mouthpiece was connected to the sensor inlet. At zero pressure, the cursor is positioned to the left end of the interaction bar, while the right end is reached with a pressure of 5 kPa, which is a comfortable value for all subjects [10]. The sampling rate was $\sim$200 Hz.

Head (Head Tracker). An *ad-hoc* head tracker was built using the *MPU-6050 (GY-521)* 6DoF accelerometer and gyroscope integrated sensor, interfaced to the computer through an *Arduino Nano* microcontroller. For each head rotation axis, the range required to move the cursor from the left to the right ends was 40° ($[-20°, +20°]$), and the natural rest position corresponded to the cursor placed at the center. The sampling rate was $\sim$100 Hz.

3 Results

A sample of 16 subjects, aged between 22 and 47, participated in the experiment. None had previous experiences with the investigated interaction channels except

---

[1] Source code available under a GNU-GPLv3 license at https://github.com/Neeqstock/HanDMIs-TestSuite.

for the mouse, and in one case the breath (some experience in saxophone playing). The initially planned number of subjects (25–30), could not be attained due to the onset of the COVID-19 epidemic in Italy (Feb. 2020).

All subjects were able-bodied: although this is a limitation, this is a preliminary test providing a benchmark for subsequent experimentation. Moreover, it must be noted that a quadriplegic user may have the same level of control as an able-bodied user for the analyzed interaction channels.

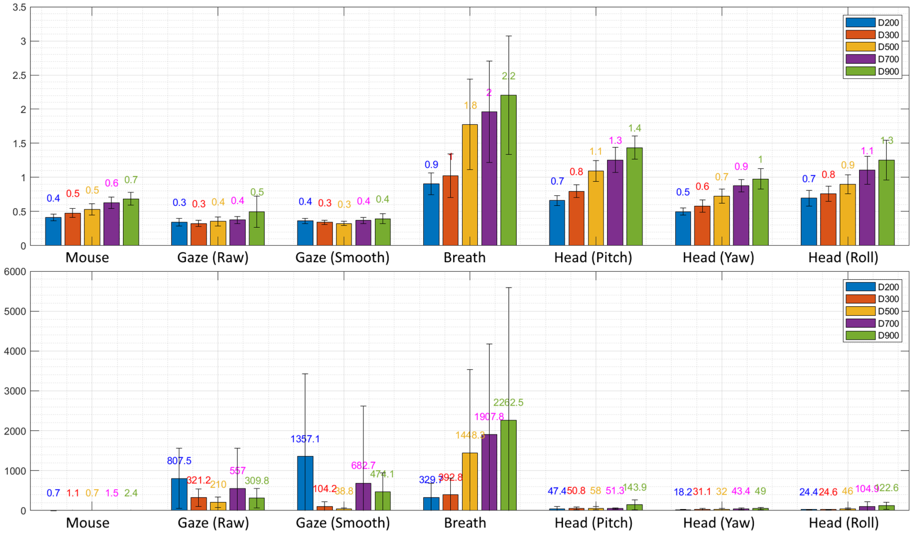

Fig. 2. Results for *selection time* (upper plot) and *selection stability* (lower plot). Error bars represent the standard deviation. Bar colors refer to the various distances (see legend). (Color figure online)

Figure 2 shows the results, where the following trials were removed as outliers: (1) trials failed due to expired time; (2) trials with wrong movements due to a reported misunderstanding of the task; (3) trials ranking outside the 10th and 90th percentile interval. Failure for expired time occurred only for Gaze (Raw) (1 for D900) and Breath (1 for D300, 2 for D500, 9 for D700, 10 for D900).

Selection times (upper plot in Fig. 2) were averaged across trials for distances and conditions. Two terms influencing these values can be noticed: a reaction time (formalized by the a parameter in the Fitts Law model) and an additive term related to the distance traveled (whose slope is defined by the b parameter). A least square linear regression was run to estimate a and b, and a subsequent ANOVA revealed an non-significant difference ($p = 0.487$) between the a parameters, suggesting comparable reaction times for all channels. In order to obtain a more accurate estimate for the b parameter, the model was then adjusted by imposing the same value for a in all channels, and linear regression on the adjusted model provided the results reported in Table 1. A subsequent ANOVA resulted in statistically significant difference for the b values ($p \ll 0.001$). Post-hoc pairwise tests with Bonferroni-Holm correction revealed non-significant differences only in the following pairs: [Mouse - Gaze (Raw)] ($p = 0.23$), [Gaze (Raw) - Gaze (Smooth)] ($p = 0.21$), [Head (Roll) - Head (Pitch)] ($p = 0.21$). All remaining pairs showed significant differences with $p < 0.05$.

Table 1. Estimates of b parameters for each interaction channel.

	Mouse	Gaze (raw)	Gaze (Smooth)	Breath	Head (Pitch)	Head (Yaw)	Head (Roll)
b	0.2304	0.1896	0.1273	0.8637	0.5315	0.3537	0.4690

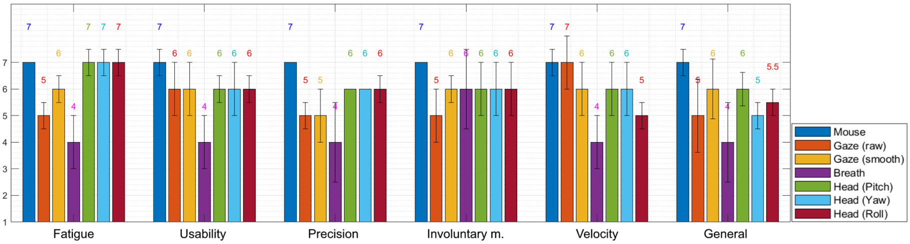

Fig. 3. Results for the questionnaire divided by question and channel-sensor pairs (higher is better). Median has been used as a measure of central tendency, while error bars are defined as interquartile range.

The stability value for each trial was computed as the variance of the cursor position within the stability phase. After outliers removal (same as above), the variances were averaged within single distances and conditions (lower plot in Fig. 2. All head related channels exhibit high stability, comparable to the mouse. The gaze point stability is negatively affected by both the sensor output and the natural instability of fixations [18]. Interestingly, the Gaze (Smooth) condition does not improve compared to Gaze (Raw). Breath is visibly more unstable especially for large distances, with large deviations. Subjects reportedly found it difficult to keep a stable pressure in the case of large distances, given that the required pressure was greater.

The questionnaires provided additional indications: results are provided in Fig. 3. With regard to *Fatigue*, Head was found not to be comparable to Mouse, while some fatigue for Gaze was indicated (in particular in the raw condition) and Breath. *Usability* of the various channels was lower than Mouse but generally comparable, except for Breath, which scored very low values. Mouse *Precision* was found to be superior, followed by Head, Gaze and Breath, consistently with results on stability. Mouse interaction was perceived free of *Involuntary movements*, while some were noted for all the other channels, notably Gaze (raw): this can be explained by the erratic movements of the cursor. *Velocity* of Gaze was perceived as comparable to Mouse, consistently with the test results, while the worst judged were Head (roll) and Breath. Excluding Mouse, *General* ratings showed a preference for Gaze (smooth) and Head (Yaw), while Breath ranked lowest.

4 Discussion and Conclusions

The channels considered in this work are independent from each other, being based on different physical degrees of freedom, and may therefore be used together for musical interactions.

Gaze interactions resulted to be very fast, and only weakly dependent on the target distance (especially in the Smooth formulation). This supports the choice of this channel for note selections (e.g., EyeHarp [19], or Netytar [3]). Although, given the nature of gaze and related sensors, stability is not particularly high, several solutions can be devised to deal with this issue [5].

On the other hand, the relatively low upper limit for the number of saccades per second [5] may lead to a reconsideration of head related channels for note selections. The three head channels have comparable performance. Yaw performs better, possibly due to (1) familiarity with this movement; (2) larger rotational range; (3) congruence between rotation direction and cursor movement, resulting in a more natural mapping. Little exploration has been carried out in this regard, even if some DMIs exploit Pitch and Yaw (e.g., Magic Flute [6] and Jamboxx [8]). Roll may instead be mapped to pitch bend, vibrato or other musical parameters.

Breath had comparatively lower scores in tests and questionnaires. The variability also appears to be very high, especially for large distances. Despite its widespread application in assistive technologies, it performed poorly in precise target selection tasks. However it can reconsidered for musical interaction given the following observations. First, breath is naturally linked to the regulation of sound intensity (e.g. while singing or playing aerophones). Furthermore, the emission of breath is associated to the sensation of energy being injected by the performer into the instrument. This doesn't happen for the other analyzed channels, as the cursor remains still in any position in absence of movement. Hunt [7] highlights how this sensation can result in a natural mapping with sound intensity. Other configurations should be explored even for the channel-sensor pairs analyzed in this work: for example, the position of the cursor could be mapped to the derivative of the head movement speed. Finally, it should be important to perform experiments with quadriplegic users. This could be achieved by proposing the same experimental setting or, given the heterogeneity of this last category, through proper case studies on differentiated users.

Acknowledgements. We would like to offer our special thanks to Daniele Zago (from University of Padova, Italy) for advising us on statistical analysis, and to Mattia Galante (from University of Milan, Italy) for the great help given by assisting the testers.

References

1. Bååth, R., Strandberg, T., Balkenius, C.: Eye tapping: how to beat out an accurate rhythm using eye movements. In: Proceedings of 11th International Conference on New Interfaces for Musical Expression (NIME 2011), NIME 2011, Oslo (May 2011)
2. Cross, I.: The nature of music and its evolution. In: Hallam, S., Cross, I., Taut, M. (eds.) The Oxford Handbook of Music Psychology, pp. 3–13. Oxford University Press, Oxford (2009)
3. Davanzo, N., Dondi, P., Mosconi, M., Porta, M.: Playing music with the eyes through an isomorphic interface. In: Proceedings of the Workshop on Communication by Gaze Interaction - COGAIN 2018, pp. 1–5. ACM Press, Warsaw (2018)
4. Frid, E.: Accessible digital musical instruments—a review of musical interfaces in inclusive music practice. Multimodal Technol. Interact. **3**(3), 57 (2019)
5. Hornof, A.J.: The prospects for eye-controlled musical performance. In: Proceedings 14th International Conference on New Interfaces for Musical Expression (NIME 2014), NIME 2014, Goldsmiths, University of London (July 2014)
6. Housemate: Magic Flute (nd). http://housemate.ie/magic-flute/. Accessed 8 June 2019
7. Hunt, A., Wanderley, M.M., Paradis, M.: The importance of parameter mapping in electronic instrument design. J. New Music Res. **32**(4), 429–440 (2003)
8. Jamboxx: Jamboxx (nd). https://www.jamboxx.com/. Accessed 8 June 2019
9. Jones, M., Grogg, K., Anschutz, J., Fierman, R.: A sip-and-puff wireless remote control for the Apple iPod. Assist. Technol. **20**(2), 107–110 (2008)
10. Lausted, C.G., Johnson, A.T., Scott, W.H., Johnson, M.M., Coyne, K.M., Coursey, D.C.: Maximum static inspiratory and expiratory pressures with different lung volumes. BioMed. Eng. OnLine **5**, 29 (2006)
11. Lubet, A.: Music, Disability, and Society. Temple University Press, Philadelphia (2011)
12. McKenzie, S.: Fitts' law. In: Norman, K.L. Kirakowski, J. (eds.) Handbook of Human-Computer Interaction, vol. 1, pp. 349–370. Hoboken (2018)
13. Mougharbel, I., El-Hajj, R., Ghamlouch, H., Monacelli, E.: Comparative study on different adaptation approaches concerning a sip and puff controller for a powered wheelchair. In: Proceedings of the 2013 Science and Information Conference, pp. 597–603. London (October 2013)
14. O'Modhrain, S.: A framework for the evaluation of digital musical instruments. Comput. Music J. **35**(1), 28–42 (2011)
15. Ooms, K., Dupont, L., Lapon, L., Popelka, S.: Accuracy and precision of fixation locations recorded with the low-cost eye tribe tracker in different experimental setups. J. Eye Mov. Res. **8**(1) (2015)
16. Sarig-Bahat, H.: Evidence for exercise therapy in mechanical neck disorders. Man. Ther. **8**(1), 10–20 (2003)
17. Shima, K., Tsuji, T., Kandori, A., Yokoe, M., Sakoda, S.: Measurement and evaluation of finger tapping movements using log-linearized Gaussian mixture networks. Sensors **9**(3), 2187–2201 (2009)
18. Sibert, L.E., Jacob, R.J.K.: Evaluation of eye gaze interaction. In: Proceedings of SIGCHI Conference on Human Factors in Computing Systems, CHI 2000, pp. 281–288. Association for Computing Machinery, The Hague (April 2000)
19. Vamvakousis, Z., Ramirez, R.: The EyeHarp: a gaze-controlled digital musical instrument. Front. Psychol. **7**, article 906 (2016)

20. Wang, F., Ren, X.: Empirical evaluation for finger input properties in multi-touch interaction. In: Proceedings of 27th International Conference on Human Factors in Computing Systems (CHI 2009), pp. 1063–1072. ACM Press, Boston (2009)
21. Williams, K.E., Berthelsen, D., Nicholson, J.M., Walker, S., Abad, V.: The effectiveness of a short-term group music therapy intervention for parents who have a child with a disability. J. Music Ther. **49**(1), 23–44 (2012)