An Evaluation of Multidimensional Controllers for Sound Design Tasks

Robert Tubb and Simon Dixon
Centre for Digital Music
Queen Mary University of London
r.h.tubb@qmul.ac.uk

ABSTRACT

This paper presents an investigation into musicians' ability to control sound synthesiser parameters using various interfaces. The principal aim was to compare separate, 1D parameter controls (touchscreen sliders) to multidimensional controllers (an XY touchpad for 2D, the Leap Motion for 3D). Subjects had to match a target sound as quickly and accurately as possible. Results show that after about two hours of practice, the XY pad is 9% faster than two sliders for no accuracy loss, and the Leap is 17% faster than 3 sliders with 9% accuracy loss. The multidimensional controllers improved most with practice. A new perspective on Fitts' index of difficulty is presented: "Index of Search Space Reduction" (ISSR). ISSR and retrospective accuracy thresholds on the search trajectory are used to obtain straight line plots and throughput values. These plots reveal that the Leap's speed improvement was mainly due to reaction time, but the XY pad traversed the space faster.

Author Keywords

Audio; Parameter Exploration; Synthesiser; Fitts' law; 3D interfaces; Leap Motion

ACM Classification Keywords

H.5.5. Sound and Music Computing: Miscellaneous
; H.5.2. User Interfaces: Evaluation/methodology

INTRODUCTION

Human-computer interaction is generally carried out in a serial fashion. Predominant interaction models tend to assume a single action at a time. However, some domains require more parallelism in the communication channel between the human and machine. One such field is music. To watch the performance of a concert pianist is to witness a virtuoso display of "space-multiplexed" [8] user input. The throughput of this interaction has been estimated at 300 bit/s, contrasting with about 50 bit/s for a good typist, and 5 bit/s for mouse pointing [4]. The increase in speed that comes with virtuosity is obvious, but usually comes at a cost in training time: some tens of thousands of hours in the pianist's case [6]. So a key question is how much practice is required to reach a throughput greater than that of a standard serial interface? This experiment investigates this issue by looking at interaction with a synthesiser, based on a simplified version of a sound design task. Controllers with 1, 2 and 3 DOF (degrees of freedom) are compared for speed, accuracy and throughput.

CHI 2015, April 18–23, 2015, Seoul, Republic of Korea.

http://dx.doi.org/10.1145/2702123.2702499

Sound Design and the Digital Musical Instrument (DMI)

The task of sound design in music, film, or computer games is a challenging one. Synthesisers and effects processors often have tens or hundreds of parameters, leading to a huge combinatorial search space. Artists search these large spaces to hand-craft instruments with complex timbres and textures. Often, only a small subset of the parameter space produces pleasing output. Parameters may interact in non-linear and unpredictable ways, and the perceived value of certain sound will change with context. This poses a challenge to interface designers: how to make the search as productive as possible?

The majority of computer music production is carried out using a Digital Audio Workstation (DAW). Despite the increasing sophistication of the sound generators, most DAW interfaces are built around metaphors that hark back to mid-20th century studio technology, namely potentiometer knobs, faders and switches. When operated with a mouse, these GUI items necessitate one at a time adjustments. DAWs are extremely powerful and flexible, but acknowledged to be less musically "expressive" than traditional instruments [17], hence the interest in using high DOF [26] controllers for Digital Musical Instruments (DMIs). This is one focus of the "New Instruments for Musical Expression" research field [20, 2]. Expressiveness is a difficult term to define, but generally implies real-time control of the dynamics and timbre of notes. The more complexity and nuance that can be imparted to a musical event, the greater the expressive range. This implies a link between expressiveness and information transfer rate. Information "throughput" has been discussed in relation to synthesiser interfaces [19], but has not yet been measured experimentally.

Synthesis parameters often correspond to, or at least affect, "perceptual dimensions". Perceptual dimensions have been shown to lie along a continuum between "integral" and "separable" [9]. Integral dimensions tend to be perceived and processed holistically [14] and not analysed in isolation, for in-

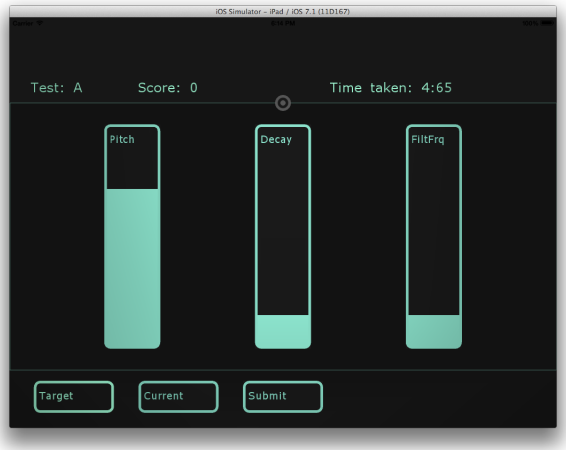

Figure 1. Screen shot of 3 slider interface during the search task. The "Target" button plays the target sound, the "Current" button plays the sound that is being adjusted using the sliders. When the user has matched the two sounds, "Submit" is pressed.

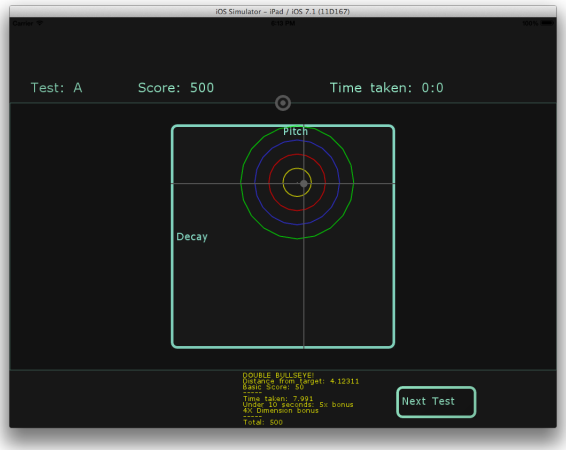

Figure 2. XY pad trial. A successfully located sound has been submitted.

stance hue and brightness. Separable dimensions are those dimensions that are perceived and most easily manipulated separately, for instance size and colour. Timbre dimensions are highly integral. This structure of perceptual space has been shown to be important for HCI by Jacob et al. [13]. This experiment revealed that the integral dimensions were best controlled by multidimensional controllers, and separate one-dimensional controllers suited separable dimensions. They also proposed a general principle that the structure of the interface must match the perceptual structure of the task domain. This has ramifications for timbre navigation and DMI design, as it implies that multidimensional controllers will be more suitable. This was tested in [22], but inaccuracies of early 3-D controllers made results inconclusive. Another influential result is [12], showing that *complex* many-to-many mappings produced better results in a sound matching task than one-to-one mappings. These complex mappings also showed more improvement with practice. A good qualitative analysis of a 4D bimanual timbre controller is found in [1].

In [23] recommendations are made for improving DMI research by borrowing tools from HCI. Fitts' law is mentioned as having potential, but has not yet been fully investigated, perhaps due to lack of a easily applicable methodology. Whilst there are many analogies between visual target pointing and sound target matching tasks, there are a number of differences and extra challenges with an auditory search. These differences must be considered when finding an analogue of Fitts' law for sound target acquisition:

1. *Delayed Assessment.* Differences in position can be assessed virtually instantaneously. Sounds, however, take time to listen to. Some control adjustments may have delayed effects, particularly time envelope controls. Differences between two sounds cannot be easily assessed with them playing simultaneously (with the exception of pitch: identifying intervals is a core part of musical training).

2. *Anisotropy.* Timbre space does not look the same in all directions. Pitch, timbre and temporal features are all very different perceptual qualities. In contrast, 3D space can be considered isotropic, although evidence for some anisotropy in pointing tasks has been found [18].

3. *Low sightedness*: Differences in sounds are harder to judge, and take far more effort to process that differences in position. Parameter spaces could be said to vary between being "sighted", where the distance and direction to the target is predictable, and "blind" where it is impossible to know which direction to move in, or how far away one is from the target. This may depend on the user's expertise.

4. Timbre space possesses a dimensionality far higher than that of ordinary space.

Research into this problem may also be applicable to many other areas where large numbers of unpredictable parameters need to be adjusted to obtain creatively satisfying results: graphic design, animation, architecture and so on.

EXPERIMENTAL METHOD

The study was a within-subjects repeated measure design. 8 subjects carried out 8 blocks of 94 sound matches. Whilst it is generally better to use more subjects for less trials, a pilot study revealed that performance was still improving after numerous runs, so a more longitudinal study was required. "Expert" participants were selected, with at least 5 years experience of music, sound synthesis or working with audio. They were paid 30 GBP for participating. To avoid fatigue, participants completed four blocks on one day, and four the following day. Table 1 shows the sequence of trials for a single block. All users conducted the trials in this order, which was designed to ramp up in difficulty, whilst balancing the multidimensional and separate slider conditions. The sequence could have been better balanced or randomised, but at the expense of a coherently gamified user experience. It was assumed that after 8 blocks the order effects would have balanced out, but this could be a limitation of the study.

REP	DIM	UI	PRC	VIS	MEM	PIT	DEC	FLT
1	1	Slidr	Y	-	-	1	-	-
1	2	XY	Y	-	Y	1	2	-
1	3	Leap	Y	Y	-	1	2	3
2	1	Slidr	-	-	-	1	-	-
2	1	Slidr	-	-	-	-	-	1
2	1	Slidr	-	-	-	-	1	-
4	2	Slidr	-	-	-	1	2	-
4	2	XY	-	-	-	1	2	-
4	2	XY	-	-	-	-	1	2
4	2	Slidr	-	-	-	-	1	2
8	3	Leap	-	-	-	1	2	3
8	3	Slidr	-	-	-	1	2	3
1	1	Slidr	-	Y	-	1	-	-
1	1	Slidr	-	Y	-	-	-	1
1	1	Slidr	-	Y	-	-	1	-
2	2	XY	-	Y	-	1	2	-
2	2	Slidr	-	Y	-	1	2	-
4	3	Slidr	-	Y	-	1	2	3
4	3	Leap	-	Y	-	1	2	3
2	1	Slidr	-	-	Y	1	-	-
2	1	Slidr	-	-	Y	-	-	1
2	1	Slidr	-	-	Y	-	1	-
4	2	XY	-	-	Y	1	2	-
4	2	Slidr	-	-	Y	1	2	-
4	2	Slidr	-	-	Y	-	1	2
4	2	XY	-	-	Y	-	1	2
8	3	Slidr	-	-	Y	1	2	3
8	3	Leap	-	-	Y	1	2	3

Table 1. The trial sequence for one block. All blocks for all users ran in this order. REP column gives the number of repetitions of this trial. PRC indicates a practice run, not scored and not included in results. Controlled conditions were: DIM: number of parameters, UI: interface type, VIS: Visible target, MEM: only one listen to target sound. PIT: indicates which control (if any) operated pitch, DEC: decay time, FLT: filter cut-off. For Multi-D controls 1, 2 and 3 correspond to X,Y and Z dimensions respectively.

The sound generator was a basic digital subtractive synthesizer, constructed in Pure-Data [21]. The sound could be described as a short "pluck" with varying pitch, duration and brightness, as often heard from classic synths such as the Minimoog. The application ran on a multi-touch tablet, the hand's coordinates being sent from the Leap via a MIDI connection. The following parameters were sent to the synth as 7-bit MIDI CC[1] values:

1. Pitch: a one octave range, midi note 40 (E3) to 52 (E4).
2. Decay time: this affected both the decay of the amplitude, and also the rate of decay of high frequencies. Maximum note length was 500ms.
3. Filter cut-off: the cut-off frequency for the resonant low-pass filter.

The Leap Motion is a device that can track the position and shape of hands and fingers. It appears to be the most spatially and temporally accurate consumer device for this purpose [24]. Skeletal hand tracking can generate at least 20 DOF, however the number of parameters was limited to 3: the XYZ position of the hand. More parameters would likely have increased the difficulty of the search beyond many participant's capabilities.

[1]Musical Instrument Digital Interface, Continuous Control

For each trial, after an initial 3 second countdown the user was presented with two sounds: the "target" and the "adjustable" sound. Both sounds parameters were randomised, but with a minimum Euclidean distance between the two[2]. The task was to alter their adjustable sound so that it matched the target sound. For example, the simplest trial featured a single slider controlling pitch: the user had to move this slider up or down until the pitches matched, and then press the submit button. Participants were told that speed and accuracy were equally important, and this was reflected in the scoring system. Controls were adjusted with the right hand, and the results heard by retriggering the sounds with the left[3]. In the standard test, either sound could be triggered whenever the user wished. In the target sound memorisation test (Table 1, MEM condition) the target button would disappear after a single listen. The intention behind this test was to more closely approximate a realistic sound design task, where the user may have a sound "in their head" that they wish to create, but was assumed to be a more difficult condition due to memory fade.

When the user was happy that their settings matched the target, they would press the "submit" button (centre bottom Fig. 1) and were given a score and a visual indication of where the target really was (Fig. 2). A small prize was offered for the best score for one block. Participants stated that "gamification" of the task increased their motivation and engagement.

A number of tests were control tests with a visual target (Table 1, VIS condition). The user simply had to line the controls up with this visual indicator, the sound being irrelevant. This was to test for interface effectiveness independent of the more complex perceptual aspects of sound matching. In the Leap motion case a 3D scene was displayed on the touch-screen, with moveable "jack" crosshairs to be aligned.

For the 3D trials, parameters 1-3 were always assigned to the x (left/right), y (forward/backward) and z (up/down) axes respectively. The 2D tests alternated between pairs of parameters 1 & 2 and 2 & 3. The 1D tests alternated between all 3 parameters. There were an equal number of trials for 2D vs. 3D tasks, sliders vs. multidimensional control types, and normal vs. target memorisation conditions.

The 1D controls were 10cm vertical sliders on the tablet screen. The 2D XY pad's height and width was also 10cm. Users did not have to pick up the position indicator from its current position before moving it. Unfortunately this meant losing data in the VIS scenario, as users could just tap the target and hardly any of the trajectory would be recorded. The iPad was directly in front of the user, and the Leap was positioned 20cm to the right of the top right corner of the iPad. The size of the Leap's active volume was 30cm cubed, 15cm above the device/table. All interaction movements and events were logged at a sample rate of 50Hz.

[2]For the Leap, the initial settings would correspond to wherever the users hand was when the test started. This start position was taken into account when calculating $ISSR$ from distance ratios.

[3]In the pilot test the sounds played automatically in alternation (reducing variability in this part of the task), but people found it too hard to determine which sound was which.

INTERPRETING THE DATA: FITTS' LAW

Fitts' law applies to rapid aimed movements in a single dimension towards a visible target. This law has been extended for more than one dimension [16, 18, 10], but it has not been investigated in non-visuospatial parameter spaces. It is a linear relation between movement time, MT, and an "index of difficulty", ID:

$$MT = a + b \times ID. \quad (1)$$

The ID is a measure of task difficulty in bits. It is calculated from the target width W, and the distance moved to reach the target D. Fitt's original formula for ID [7] can be derived (or at least motivated) by considering the movement as a series of smaller movements with iterative corrections [3, p. 53]. However, there are alternative formulae, and even power laws fit the data well in many cases. The debate continues [5, 25], but the currently accepted standard is derived by considering the nervous system as a noisy communication channel [15]:

$$ID = \log_2\left(\frac{D}{W} + 1\right) \quad (2)$$

Fitts' law can be used to predict the time taken for various common interaction tasks, such as moving a cursor to a GUI button. It can also be used to compare the effectiveness of input devices, via the "throughput": the rate at which a user can input information to the system, $TP = ID/MT$.

Throughput seems like it should be a useful measure of progress in this target acquisition task. The question is if the prerequisites for Fitts law apply for this experiment. The search is certainly not "rapid", and may not be "aimed", due to low sightedness. The size of a sonic target is impossible to specify to the user, therefore they cannot implement different accuracy levels to provide a range of values for a regression line. One can calculate W from the standard deviation of the results to obtain the "effective width" $W_e = 4.133\sigma$. However, the high variance in accuracy generates extremely low ID values (for the 3D search in this study $\sigma \approx 10, D/W_e \approx 64/40, ID \approx 2bits$), and this single error distribution would not provide a range of difficulties. On the other hand, we carried out a large number of trials, and have a record of all the search paths, many resulting in high accuracies. Therefore, it would be helpful to have a method of extracting useful information from these accurate trials without artificially selecting unrepresentative data.

Rather than taking the index of difficulty as a independent variable set by the experimenter, we need an equivalent quantity that can quantify, *post-hoc*, the amount of success achieved at a certain task in a certain time. How can we measure the information input without recourse to an end point noise distribution, and how can we extend it to an n-dimensional control space? Can we meaningfully compare information input across different control dimensionalities? We approach these problems by deriving an information measure from "search volume reduction", presented in the next section.

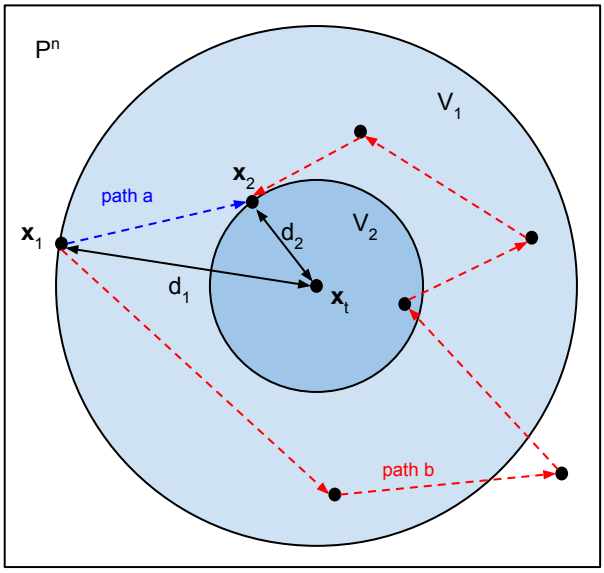

Figure 3. Search space reduction. The target is x_t, x_1 and x_2 are the start and end points of the movement along path a. V_1 and V_2 are the volumes associated with distances to target d_1 and d_2. The logarithm of the ratio between these two volumes gives a measure of information gain. Summing over all the steps of b gives the same amount of information gain as a (Eq. 5).

Index of Search Space Reduction (ISSR)

Both Fitts' and Hick's law [11] are motivated by an assumption of constant information flow through the nervous system. For a human interacting with a computer, this information then flows into the interface, and ultimately to the data artifact that is to be manipulated. The main motivation for developing ISSR is that, for content creation tasks, the ideal point at which to measure information flow is "where the rubber hits the road": precisely how it alters the data toward some desired state. Rather than looking at the capacity of the motor channel, we look at the reduction in the entropy of the point(s) in parameter space.

For an n-dimensional parameter space P^n, a start point $\mathbf{x}_1$, an end point $\mathbf{x}_2$ and a target point $\mathbf{x}_t$ we first calculate the distances to the target before and after a movement,

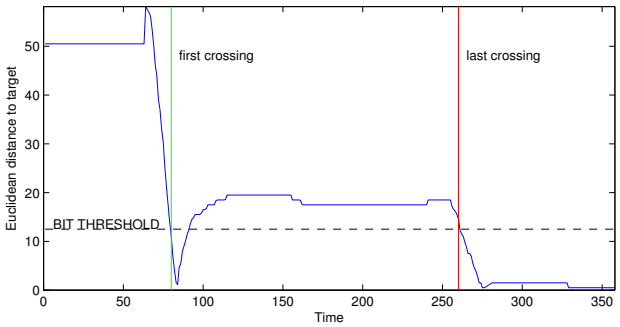

Figure 4. Time-to-threshold is calculated from the last crossing of a distance threshold. In this 1D case search space is effectively quartered, ISSR = 2 bits, MT = 260 samples (5.2 seconds).

$$d_1 = \|\mathbf{x}_t - \mathbf{x}_1\|, d_2 = \|\mathbf{x}_t - \mathbf{x}_2\|.$$

We define the search space reduction factor R as the ratio of the n-volumes corresponding to radii of the distances (see Fig. 3). Volume is calculated as $V = Cd^n$. The exact shape and size of the volume does not in fact matter, as the multiplier C cancels:

$$R = \frac{V_1}{V_2} = \frac{{d_1}^n}{{d_2}^n}.$$

For a path through a space towards a target, the remaining search space will be whatever volume the remaining search path is restricted to.

In general, any search task can be said to be a reduction of a set of possibilities. For an task involving a fixed number of options, the entropy reduction (or information conveyed) by a choosing of a subset of these possibilities will be the logarithm of the number of possible states before the choice was made divided by the number of possible states afterwards. If the remaining search volume is reduced by a factor of two, then we have successfully completed one bit of the search. This gives the "index of search space reduction" (renamed from index of difficulty to avoid confusion),

$$ISSR = log_2\left(R\right) = \log_2\left(\frac{{d_1}^n}{{d_2}^n}\right) = n\log_2\left(\frac{d_1}{d_2}\right). \quad (3)$$

Whilst a negative "difficulty" seems meaningless, it seems reasonable to say that moving away from the target results in lost information, and $ISSR < 0$ when $d_1 < d_2$. If no progress is made and $d_1 = d_2$, then $ISSR = 0$. For MT, dependence on dimensionality is simple: a constant multiplier of the gradient b,

$$MT = a + bn\log_2\left(\frac{d_1}{d_2}\right). \quad (4)$$

In fact this alternative derivation, for the one dimensional case, gives us Fitts' original equation [7]: substituting $n = 1$, $D = d_1$ and taking the target width as twice the final distance to the target centre, $W = 2d_2$ we get

$$MT = a + b\log_2\left(\frac{2D}{W}\right).$$

Interestingly, in 3D it is also yields a very similar equation to the change in entropy of a isothermally compressed gas, $\Delta S = -Nk\log\left(V_2/V_1\right)$, N and k being constants. So we might consider our interaction data set, a coalescing cloud of search paths, as analogous to the paths of a gas particles in a shrinking box.

A further reassuring property of the ISSR is that it conserves information, i.e. the total information gain of a search path

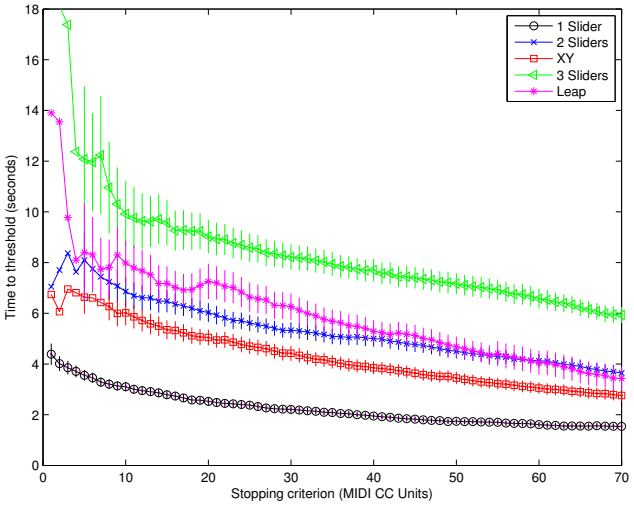

Figure 5. Average time taken to reach a given Euclidean distance threshold for each interface condition, day 2. Whilst different dimensionalities may not be directly comparable here, they are shown on the same plot for brevity. Whiskers display 95% confidence ratios at points where difference between interfaces is significant.

can be considered as the sum of the information of all its subpaths, irrespective of how it is divided. The sum of information for M steps is

$$\begin{aligned} ISSR &= \sum_{m=1}^{M-1} n\log_2\left(\frac{d_m}{d_{m+1}}\right) \\ &= n\sum_{m=1}^{M-1}\Big(\log_2\left(d_m\right) - \log_2\left(d_{m+1}\right)\Big). \end{aligned}$$

All the terms cancel except the first d_m and the last d_{m+1} term, giving

$$ISSR = n\log_2\left(\frac{d_1}{d_M}\right). \quad (5)$$

This is identical to Eq. 3 for the start and end points of the whole path. It is difficult to see how the ID in Eq. 2 can conserve information in this way.

What difference will this formula give when navigating the search space using one parameter at a time rather than with a multidimensional controller? In the separate case, the total navigation time will be just the sum of navigation times for each 1D control, from Eq. 4:

$$MT_{tot} = na_{sep} + nb_{sep}\log_2\left(\frac{d_1}{d_2}\right). \quad (6)$$

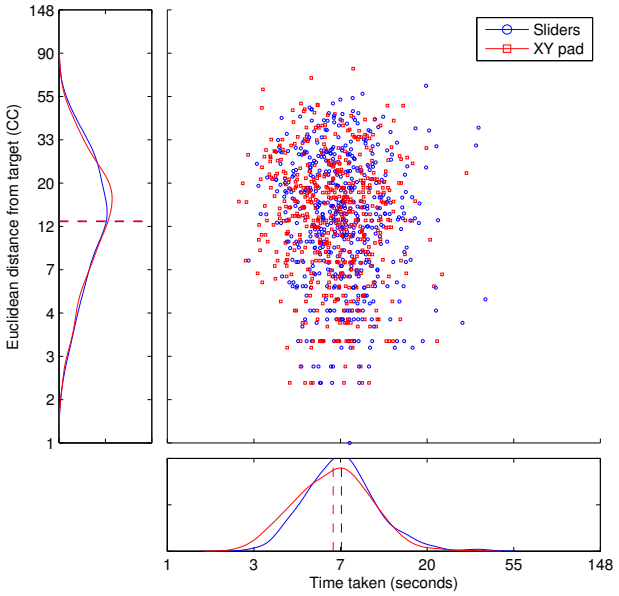

Figure 6. Distributions of log speed and accuracy results for the two parameter case, on the second day of the test. The XY-pad is as accurate as the sliders, but has more fast results under 5 seconds, resulting in a small but significant difference.

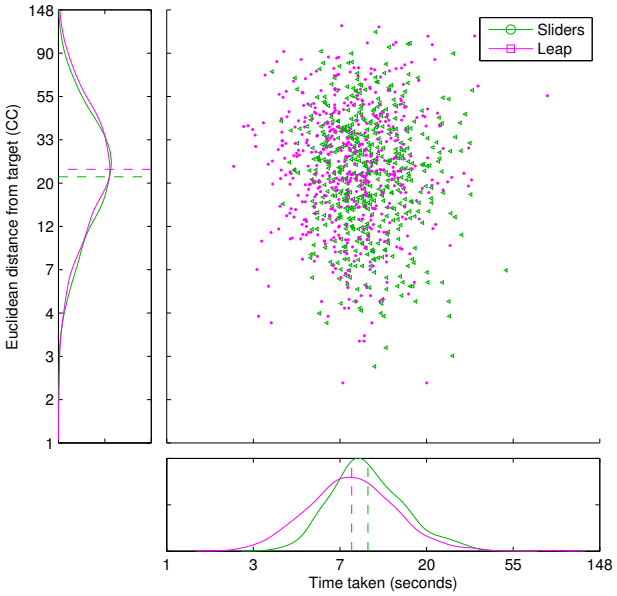

Figure 7. Speed and accuracy results for the three parameter case (day 2). Accuracy is slightly less with the leap but it yields more results under 7 seconds.

In the multidimensional case, however,

$$MT_{tot} = a_{int} + nb_{int} \log_2 \left(\frac{d_1}{d_2} \right). \quad (7)$$

If we assume the difficulty of progress through the space is the same $b_{sep} = b_{int}$, then the only difference between the two formulae will be an offset due to acquisition time. If $a_{int} < na_{sep}$, then integral controllers will be faster. In other words, any slow down seen in the separate case should be entirely explained by a constant offset time, for instance, how long it takes people to swap between sliders.

Alternatively, in light of the result in [13], it seems possible that diagonal movements require quite different cognitive processes, and $b_{int} \neq b_{sep}$.

Time-To-Threshold Plots

We would like a throughput measure that also makes use of all the search path trajectory data, rather than just its end point. As an example of this approach, Jacob et. al. [13] performed a retroactive analysis of the search trajectory that measured the time taken to reach various accuracy thresholds (or stopping criteria). This produces a series of simulated experiments with different target sizes. Figure 4 shows a plot of Euclidean distance to target for an example trial, and shows the last crossing of an accuracy threshold. We can set as many of these levels as we wish, and average many trials to get a mean time-to-threshold. One can then produce plots of time against accuracy (e.g. Fig. 5). For our purposes, these plots have a number of issues:

1. The lines often curve up steeply at smaller thresholds. Straight lines would be preferable, in order to obtain Fitts' law constants a and b.
2. The starting point is not taken into account: if the user starts close to the target, then achieving an absolute distance threshold will be easier.
3. In more dimensions the search space is larger, therefore achieving a given threshold will be harder.

We can avoid these issues by expressing the stopping criterion in terms of the $ISSR$. If sub-sections of the search path also obey Eq. 4 then plotting average MT against ISSR should give straight lines, and reflect the relative difficulties in different dimensionalities.

There is a statistical dilemma with this multiple threshold technique, however. One can include all the trials, but poor performances never reach high bit levels, and will not be represented towards the right of the plot. This will tend to make the lines curve downwards, and become unreliable at high $ISSR$. On the other hand, if the tests where the threshold was never reached are omitted entirely, the good performances are over-represented and significance decreases due to the smaller sample size. The policy here is to use the best half of all the trials for a given condition, i.e. set a threshold at the median $ISSR$ achieved. Any trial that did not reach the median no. of bits are discarded. Whilst this means the final TP values may underestimate the task difficulty as a whole, they should at least provide a *relative* comparison between experimental conditions. Higher $ISSR$s for the successful tests are not featured on the plot, therefore sample size is the same for every point along the line. This should not unfairly favour any particular control device, though it will favour the results from users more comfortable with the task. If the $ISSR$

	Pitch	Decay	Cut-off
1S	5.28	9.21	9.75
2S	4.49	13.94	12.99
XY	6.13	13.76	13.34
3S	6.67	14.79	14.24
LM	7.93	16.07	15.21

Table 2. Inaccuracies (standard deviation from the target in CC units) of individual parameters for all trials. Pitch is always most accurate.

version of Fitts' law holds, then this technique should give straight lines across a range of bit values.

RESULTS

Speed and Accuracy

Scatter plots of speed (time to submit) and accuracy (Euclidean distance to target at submission) for all 2D and 3D trials are shown in figures 6 and 7. Both axes display approximately log-normal distributions. No correlation between speed and accuracy is seen.

Overall, the decrease in completion time compared to equivaluent numbers of sliders is around 8 percent for the XY and 13 percent for the leap. However, people significantly improved across the two days (see later). If we look at results for the last 4 blocks (day 2), post practice the XY was 9% faster (paired T-test between interfaces $t(527) = 5.22, p < 0.01$). The leap was 17% faster than 3 sliders ($t(527) = 9.61, p < 0.01$), for an accuracy reduction of 9% ($t(527) = -2.36, p < 0.05$). Individual analyses for each user reveal similar patterns. Here we assume that different dimensionalities are not comparable, but if 2-way ANOVA is run for both dimensionality and interface type, speed-up is still significant ($F(1,1) = 192.4, p < .01$) and there is a significant interaction ($F(1) = 5.58, p < 0.05$).

Accuracy errors for all trials are given in Table 2, in the form of the standard deviation of the difference between the target value and the value of the parameter at submission. Not surprisingly, the accuracy for each parameter decreases the more sliders need to be set (the one exception being the good result for pitch in the 2D case). Timbre errors were around twice the size of pitch, despite a pitch range of only 1 octave, illustrating the "anisotropy" mentioned earlier.

We may already conclude that the higher DOF controllers are marginally more effective, but it would be preferable to have a single measure of throughput, and trajectory progress plots giving more insight into the cause of the differences.

Throughput

Figure 5 shows the average time taken to reach a given Euclidean distance threshold for all 2 and 3 dimensional trials. The Leap and XY pad are faster than the corresponding number of sliders for thresholds $> 5CC$. Figure 8 shows $ISSR$ plots for day 1 and day 2. Most lines now appear straighter, supporting the idea that a Fitts style law applies. Table 3 shows that if a regression is fit to the raw data, the wide distributions generate low R^2 values, but confidence bounds for the slope and intercept are reasonable.

On day 1 the leap was faster up to 3bits, but the gradient b_{LM} is obviously steeper than for 3 sliders. Day two, the gradients

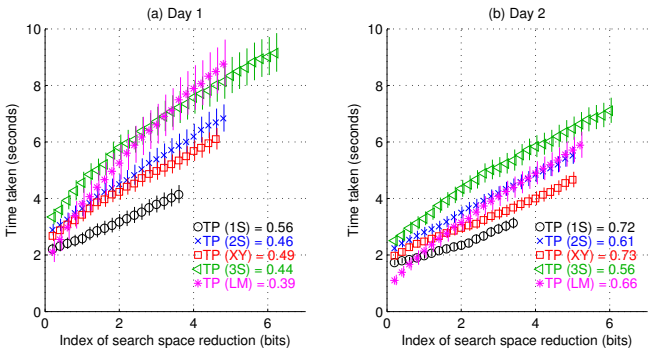

Figure 8. ISSR vs. MT, day 1 and day 2. Colors and markers consistent with Fig. 5. The gradient for the Leap improves with practice to match the sliders, but is about 1 second faster at all bit levels. To obtain cleaner plots, only trials that scored above the median $ISSR$ for that condition are included. TP values are calculated from the average of $MT/ISSR$ for every point along the line.

	1S	2S	XY	3S	LM
Intercept	1.6±0.2	1.8±0.1	2.0 ± 0.1	2.2±0.1	1.2±0.15
Slope (b)	.51±.04	.79±.03	.70±.02	.85±.02	.87±.02
TP (1/b)	1.96	1.26	1.42	1.17	1.14
R^2**(all)**	0.123	0.204	0.156	0.175	0.156
R^2**(mean)**	0.984	0.999	0.997	0.990	0.994

Table 3. Results of regression line fitting for each interface on day 2. Throughput here is taken as the reciprocal of the slope.

b_{LM} and b_{3S} appear the same, but the intercept a is lower for the Leap. This pattern is not seen in the 2D case, here a_{XY} and a_{2S} appear equal but b_{XY} is shallower than b_{2S}. The XY pad is faster even on day 1. Throughput values on the plots are calculated by averaging $ISSR/MT$ for all data points.

The intercepts can be largely explained by calculating reaction times, these are shown in Table 5. Firstly, RT is the average time from the presentation of the test until the sound is triggered. Second, listening time, LT, is taken as the time taken from the first sound trigger until the first control adjustment. RTs are the same for all interfaces (around 1s). LT is more variable. With the Leap, people start moving within 0.25s, even before they have time to listen to the sound they are adjusting. This could be just random hand waver triggering the movement threshold (set at 10CC/s), but the advantage carries through to higher accuracies, so it would appear to be real progress. The quick start also seems to explain the lower intercept on the Leap's plots. The question then becomes: what was it about the Leap that enabled people to start making progress sooner? One hypothesis is that people can *categorise* a sound quickly, and associate it with an approximate region in 3D space. On hearing the target, they can move in roughly the right direction without even listening to their current position or considering individual parameters. This

	1S	2S	XY	3S	LM
Time	-22*	-21†	-28†	-21†	-37†
Median Acc.	-6	4	9*	-3	8
Throughput	27	32*	48†	28*	70†

Table 4. Percentage difference for time taken, accuracy, and throughput between day 1 and day 2 (See Fig. 8). Two sample t-test, $\star P < 0.05$, $\dagger P < 0.01$.

	1S	2S	XY	3S	LM
RT	0.99	0.99	0.98	0.96	1.03
LT	0.85	1.26	1.05	1.39	0.24

Table 5. Reaction times (RT) and initial listening times (LT) for different interfaces. People seem to start moving much faster with the leap, explaining the lower intercepts.

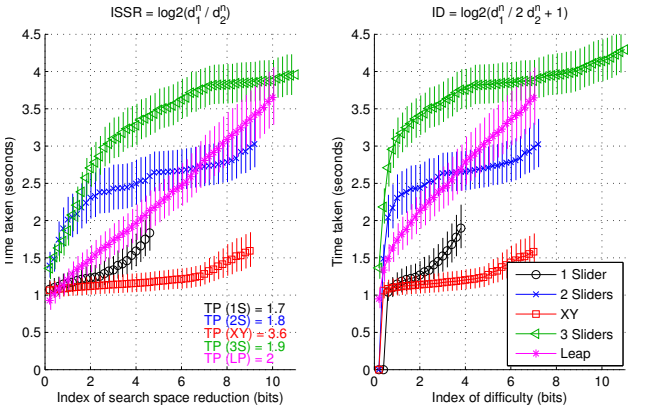

Figure 9. ISSR plot for the visible target condition (left) the kinks in the plots for the sliders are caused by having to swap controls. If we imagine the curves extrapolated onwards to higher accuracies, it seems that the 3 sliders will overtake the Leap. The right hand plot uses MacKenzie's ID, introducing sharp drops at low IDs.

would indicate a completely distinct learning process from that occurring with separate controls, certainly worth investigating further. Alternatively, one could argue that differences in reaction times reveal a flawed methodology, in which case some way of eliminating this effect should be found.

Table 4 summarises the effects of practice. The sliders show around a 21% speed improvement from day 1 to day 2, the XY improves by 28%, the Leap improves 37%. Participants keep their accuracy threshold relatively steady.

Comparisons with Visual Target Acquisition

Figure 9 shows the results for acquisition of the visual targets. Around twice the speed and twice the bit accuracy was achieved compared to the sound task. The only interface that gives a straight line is the Leap. The mostly flat lines for the 1 slider and XY plots are because users could simply tap the target, so the movement data was not recorded until their finger was on the screen for the final adjustments. The kinks in the 2 and 3 sliders' plots are probably caused by swapping time (as predicted in 6). Initial reaction times are similar to the sound task. The second plot shows the lines when a $+1$ is incorporated in the ID calculation (i.e. Eq. 2). This results in a sharp curve when $ID < 1$bit. If a regression line is fitted to the data, this reduces R^2 from 0.32 to 0.24 in the Leap's case. So the $ISSR$ formula does seem more appropriate for handling this time-to-threshold data.

People quite often needed to revisit a slider once the others were closer to the correct values. In theory, setting 3 parameters necessitates 2 slider swaps, in fact the mean number of swaps was 3.3, indicating that adjustments became more difficult if the other parameters were not set. The mean time for a swap was 0.9s. In the 2D case, number of swaps $= 1.7$ and

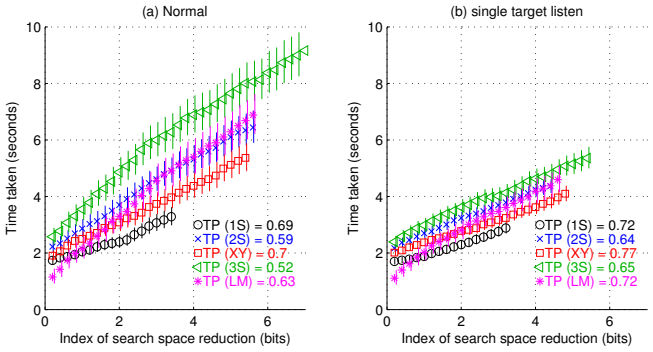

Figure 10. ISSR vs. MT, for MEM and nonMEM case, day 2 only. Not surprisingly, final accuracy has decreased, but the speed up for a given accuracy is quite surprising.

	1S	2S	XY	3S	LM
Mean Accuracy	-6*	-15†	-10†	-21†	-18†
Time to mean acc.	$-14^\star$	$-26^\dagger$	$-17^\dagger$	$-33^\dagger$	$-29^\dagger$
Throughput	8	17	8	26*	22

Table 6. Percentage change from normal to MEM condition (See Figure 10). $\star P < 0.05$, $\dagger P < 0.01$

swap time $= 0.86$s. When a visible target was present the swap times were faster: 0.65s. So an extra 0.2s was required to re-orient to another perceptual dimension in the sound task, probably to re-compare the sounds. This complicates the theory behind Eq. 6 somewhat.

Integration: Diagonal Movement

Another quantity of interest is whether people really did operate more than one dimension at a time, i.e. move diagonally in the integral controller case. Diagonal travel is also referred to as "coordination" [26] and "controller integration" [22]. The former is calculated from the correlation of different dimensions, but here, as in [13], integration was calculated from as being the ratio between the amount of time that more than one dimension was moving to the time only one dimension was moving. The speed threshold distinguishing a moving/stationary dimension was set at 10CC/s. Integration values were heavily dependent on the threshold value, but results comparing experimental conditions were not. A scatter plot of diagonality vs. completion speed (Fig. 11) shows that the amount of diagonal travel did slightly correlate with speed, however most navigation was being carried out in a city-block fashion, with integration ratios < 1.

Target Memorisation Test

Fig. 10 shows the differences in the MEM case. Accuracy worsens; participants said that auditioning the sound they were controlling degraded the memory of the target. However it is interesting that the actual time to a given bit threshold is much faster. Table 6 shows percentage differences. So for rough matches it is much faster to not keep re-listening to the target. Nevertheless, participants failed to implement this strategy when they were given the choice, indicating that they underrate their own ability to either memorise a target or predict the effect of parameter adjustments.

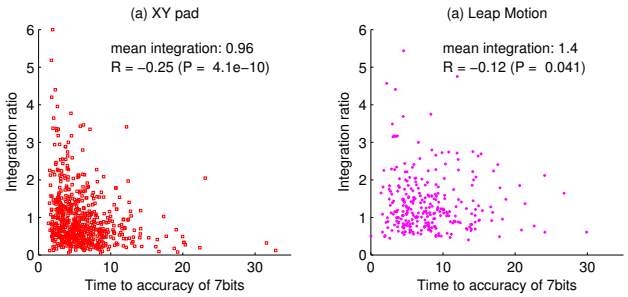

Figure 11. A small but significant correlation between amount of diagonal movement and speed.

Another interesting result was that search trajectories were more diagonal in the single target listen case. For the XY pad, the integration ratio was 1.2 (MEM), vs. 0.8 (non-MEM) $t(1022) = 6.95, p < 0.01$. For the Leap it was 2.2 (MEM) vs. 1.3 (non-MEM) $t(1022) = 7.4, p < 0.01$. It seems that if people are forced not to repeatedly compare the two sounds, they treat the dimensions in a more integral fashion. Could this be because a back and forth comparison encourages a slower, analytical mode of thinking, whereas a sound stored in a short term auditory buffer is treated in a more holistic fashion?

DISCUSSION

It seems the multidimensional controllers are more effective, though not by a huge margin. However, they were showing greater improvements with practice, so may be expected to become faster still. The reasons for the speed improvement appeared different for the different devices, however. The XY pad showed a greater throughput due to a shallower gradient: it was faster traversing the space. The speed gains with the Leap, on the other hand, seemed to be a result of faster reaction time: for some reason people felt they could start the search quicker, without waiting to compare the sounds first. We speculate that this is the result of associative learning of regions of the space. For achieving high accuracies, the sliders were still preferable to the Leap, which was 9% less accurate. Therefore, in terms of sound production work-flow, high DOF controllers would be better for early stage exploratory creativity and live performance, but individual controls better for late stage creativity and fine tuning.

There is a small correlation between diagonal movement and speed, but not yet enough to be the cause of significant speed up for multidimensional control. Far more practice seems to be needed to be able to be completely comfortable taking the shortest path through the parameter space. One user guessed that about 100 hours would be required before they had learned the perceptual space well enough to move directly to the target in 3D. Indeed, one of the most striking findings is how hard the perceptual component of this task is. Even with three simple audio parameters, experienced users, and elimination of the worst half of the results, throughput is only around 0.5bit/s, around a quarter of that for the pointing tasks.

The proposed ISSR characterisation of Fitts' law proved useful for the following reasons:

1. It provided a theoretical baseline of how difficulty should scale with dimensionality, for both integral and separable cases.
2. It measures information throughput at the point of interest: the effectiveness of the search.
3. Where varying accuracy levels cannot be specified in advance, it enables us to extract a range of difficulty values from the trajectory data.
4. For the multidimensional controllers, it generated straight lines near the intercept of movement time plots, and these intercepts agreed well with reaction time measurements.
5. It has a simple and generalisable definition, and could be easily applied to a wide variety of search task situations.
6. Information is always conserved, no matter how convoluted the search path.

What ISSR is not intended to address is accurate *prediction* of movement difficulty. Fitts' law can be used in a predictive sense, in which case subtleties of the human motor system are important, but can also be used in a comparitive/evaluative sense, where we wish to test alternative interfaces for their effectiveness. It is the second scenario that ISSR is deemed appropriate for. Further work is needed to ascertain exactly why the noisy channel approach gives a different formula.

This experiment was probably not precise enough to expose subtle cognitive effects such as integrality or separability of timbre parameters. The bulk of the disparity between interfaces seemed to be attributable to basic manipulation issues, i.e. those revealed in the visual target task. Ideally, individual experiments would be carried out to investigate each of the aspects of this experiment in isolation. More should be done to reduce the variability in participants' performance, perhaps by teaching a consistent technique. An upcoming study will attempt to treat time as the independent variable: users will match a repeating sequence of targets (this time with 6 DOF) along to metronome clicks of varying speeds. Musicians are good at predicting when regular beats will occur, so hopefully reaction times can be eliminated. Also eliminated will be the need for the user to trigger sounds themselves. This task should also be a more appropriate model of live performance. It would also be useful to compare hardware faders and knobs: these are generally preferred by musicians to touch screen controls, and they may exhibit faster acquisition times. However there is a practical difficulty specifying the target in the visual target case, and giving progress feedback to the user. Touch screens are more flexible for such graphical feedback.

There is still a long way to go to get ecologically valid results for a digital musical instrument capable of musically varied sounds. Four hours of interaction time is a fair duration for most interface evaluations, however it is tiny compared to the amount of time serious musicians spend practising. If a practise programme could be designed for a higher dimensional synthesiser interface, analysis based on Fitts' law could provide valuable insights for the instrument designer, the musician, and the HCI community at large.

REFERENCES

1. Adams, A. T., Gonzalez, B., and Latulipe, C. SonicExplorer: Fluid exploration of audio parameters. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, ACM (2014), 237–246.
2. Caramiaux, B., Tahiroğlu, K., Fiebrink, R., and Tanaka, A., Eds. *Proceedings of the International Conference on New Interfaces for Musical Expression*, Goldsmiths, University of London, UK (2014).
3. Card, S. K., Newell, A., and Moran, T. P. *The psychology of human-computer interaction*. L. Erlbaum Associates Inc., 1983.
4. Despain, A., and Westervelt, R. High performance human-computer interfaces. Tech. Rep. JSR-96-130, Advanced Research Projects Agency (ARPA), September 1997.
5. Drewes, H. Only one Fitts' law formula please! In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, ACM (2010), 2813–2822.
6. Ericsson, K. A. The influence of experience and deliberate practice on the development of superior expert performance. *The Cambridge handbook of expertise and expert performance* (2006), 683–703.
7. Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology 47*, 6 (1954), 381.
8. Fitzmaurice, G. W., and Buxton, W. An empirical evaluation of graspable user interfaces: towards specialized, space-multiplexed input. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, ACM (1997), 43–50.
9. Garner, W. R. The stimulus in information processing. *American Psychologist 25*, 4 (1970), 350.
10. Grossman, T., and Balakrishnan, R. Pointing at trivariate targets in 3d environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2004), 447–454.
11. Hick, W. E. On the rate of gain of information. *Quarterly Journal of Experimental Psychology 4*, 1 (1952), 11–26.
12. Hunt, A., and Wanderley, M. M. Mapping performer parameters to synthesis engines. *Org. Sound 7* (August 2002), 97–108.
13. Jacob, R. J., Sibert, L. E., McFarlane, D. C., and Mullen Jr, M. P. Integrality and separability of input devices. *ACM Transactions on Computer-Human Interaction (TOCHI) 1*, 1 (1994), 3–26.
14. Kemler Nelson, D. G. Processing integral dimensions: The whole view. *Journal of Experimental Psychology: Human Perception and Performance 19* (October 1993), 1105–1113.
15. MacKenzie, I. S. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction 7*, 1 (1992), 91–139.
16. MacKenzie, I. S., and Buxton, W. Extending Fitts' law to two-dimensional tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (1992), 219–226.
17. Magnusson, T., and Mendieta, E. H. The acoustic, the digital and the body: A survey on musical instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (2007), 94–99.
18. Murata, A., and Iwase, H. Extending Fitts' law to a three-dimensional pointing task. *Human Movement Science 20*, 6 (2001), 791–805.
19. Pennycook, B. W. Computer-music interfaces: a survey. *ACM Computing Surveys (CSUR) 17*, 2 (1985), 267–289.
20. Poupyrev, I., Lyons, M. J., and Fels, S. New interfaces for musical expression. In *CHI'01 Extended Abstracts on Human Factors in Computing Systems*, ACM (2001), 491–492.
21. Puckette, M. Pure Data: another integrated computer music environment. *Proceedings of the Second Intercollege Computer Music Concerts* (1996), 37–41.
22. Vertegaal, R., and Eaglestone, B. Comparison of input devices in an ISEE direct timbre manipulation task. *Interacting with Computers 8*, 1 (1996), 13–30.
23. Wanderley, M. M., and Orio, N. Evaluation of input devices for musical expression: Borrowing tools from HCI. *Computer Music Journal 26*, 3 (2002), 62–76.
24. Weichert, F., Bachmann, D., Rudak, B., and Fisseler, D. Analysis of the accuracy and robustness of the leap motion controller. *Sensors 13*, 5 (2013), 6380–6393.
25. Zhai, S. Characterizing computer input with fitts law parameters, the information and non-information aspects of pointing. *International Journal of Human-Computer Studies 61*, 6 (2004), 791–809.
26. Zhai, S., and Milgram, P. Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (1998), 320–327.