Effect of latency on performer interaction and subjective quality assessment of a digital musical instrument

Robert H. Jack
Centre for Digital Music
Queen Mary University of
London
London, UK
r.h.jack@qmul.ac.uk

Tony Stockman
Centre for Digital Music
Queen Mary University of
London
London, UK
t.stockman@qmul.ac.uk

Andrew McPherson
Centre for Digital Music
Queen Mary University of
London
London, UK
a.mcpherson@qmul.ac.uk

ABSTRACT

When designing digital musical instruments the importance of low and consistent action-to-sound latency is widely accepted. This paper investigates the effects of latency (0-20ms) on instrument quality evaluation and performer interaction. We present findings from an experiment conducted with musicians who performed on an percussive digital musical instrument with variable amounts of latency. Three latency conditions were tested against a zero latency condition, 10ms, 20ms and 10ms ± 3ms jitter. The zero latency condition was significantly rated more positively than the 10ms with jitter and 20ms latency conditions in six quality measures, emphasising the importance of not only low, but stable latency in digital musical instruments. There was no significant difference in rating between the zero latency condition and 10ms condition. A quantitative analysis of timing accuracy in a metronome task under latency conditions showed no significant difference in mean synchronisation error. This suggests that the 20ms and 10ms with jitter latency conditions degrade subjective impressions of an instrument, but without significantly affecting the timing performance of our participants. These findings are discussed in terms of *control intimacy* and instrument *transparency*.

CCS Concepts

•**Human-centered computing** → **Interaction design theory, concepts and paradigms;** *Sound-based input / output;* •**Applied computing** → **Sound and music computing;**

Keywords

Latency; digital musical instruments; perceived quality; multisensory feedback; effort; control intimacy; interaction

1. INTRODUCTION

Latency is a fundamental issue affecting digital systems and is of particular relevance to digital musical instrument design, where the fluent translation of performer action to audible output is an essential characteristic of the device. The asynchrony between a control gesture and a system's corresponding response (be it auditory, visual or tactile) can impact on user experience in ways that are both obvious and subtle. The importance of low and consistent action-to-sound latency is widely accepted when designing digital musical instruments. Wessel and Wright's 2002 recommendation that digital musical instruments should aim for a latency of less than 10ms and with a jitter of less than 1ms is a common point of reference in the community [20].

It has often been recommended that digital musical instrument designers look to acoustic instruments for examples of tools that foster a relationship between gesture and sound that is both intuitive yet complex [16]. Generally acoustic instruments produce sound in reaction to action instantaneously, as the sound producing mechanism and control interface are one and the same. There are however some exceptions when latency is built into the mechanism of an instrument – in the case of a piano, the delay between a key reaching the key bottom and the hammer striking the string can be about 35ms for *pp* notes and -5ms for *ff* notes [1]. These figures do not include the key travel time (the time elapsed between initial touch and the key reaching the key bottom) which for pressed touch can be greater than 100ms for *pp* notes and 25ms for *ff* notes [1].

With digital musical instruments latency and jitter have been identified as barriers to virtuosic engagement, obstructing a fluent interaction with the instrument [12, 16, 20]. These factors impede what Wessel and Wright describe as the development of *control intimacy* between performer and instrument [20]. Fels describes control intimacy as the perceived match of the behaviour of the instrument and the performer's control of that instrument [4], a concept deeply connected with the notion of tool *transparency* from embodied music cognition [11]: the maintenance of an ecologically valid causal link (see also [3]) between action and sound to foster embodied engagement with an instrument.

1.1 Latency in a musical context

Questions regarding the perceptual thresholds of latency between action and sound have been explored in recent research on human sensorimotor synchronisation and shall be discussed below (see Repp and Su [17] for a comprehensive

review). Thresholds of latency perception only go so far in describing the complexity of latency in a musical context. Lago and Kon [10] point out the variability of latency thresholds in a musical context and their dependence on instrument, style of music and spatial positioning: in ensemble playing latencies ranging from 10ms to 40ms are often present due to the limiting speed of sound and the distances involved between performers. This paper is concerned not with latency between players, but the latency from the player's actions to the sound of their own instrument.

Musical instruments constrain the set of possible control gestures that a performer uses, and latency effects these control strategies in different ways. Instruments with continuous gestural control, for example, have been shown to be less sensitive to latency: for a theremin, where no physical contact is made with the instrument, the threshold is 20-30ms [13]. Percussive instruments, on the other hand, are likely to be the most sensitive: studies suggest that subjects can tap to a steady beat with as low as 4ms variation [18], while listeners are able to detect timing variations of around 6ms in isochronous sequences [6]. Many previous studies point to highly trained musicians being particularly sensitive to timing differences. The standard deviation of synchronisation error is lower for highly trained musicians than for non-musicians [17]. Fujii et al. [7] find that highly trained drummers can achieve a mean synchronisation error of 2ms for a metronome at 1000ms and 500ms, and 1ms for a metronome at 300ms, with standard deviations of 10-16ms.

Latency, at heart, is a multisensory issue. It brings to the fore the mechanism of sensory integration: how impressions of simultaneity are maintained and how they can be pushed and pulled depending on the relative time of arrival of stimuli to each sensory modality. Kaaresoja et al. [8] examine the effect of multi-modal feedback on the perceived quality of touchscreen virtual buttons, finding that tactile feedback should have the lowest latency (5-50ms) followed by audio (20-70ms) and finally visuals (30-85ms). In a previous study they noted that delayed tactile stimuli can create the impression of a heavier button, requiring participants to exert more force [9]. Another relevant study was conducted by Dahl and Bresin [2] while investigating audio-tactile latency in a musical context. They found that when latency is progressively introduced musicians shift their gestures and strike ahead of the beat to align the delayed sound with a metronome, known as *anticipation* [17]; synchronisation can be maintained in this way up to around 55ms latency.

1.2 Latency and instrument quality

Though the perceptual and synchronisation effects of latency have been studied, there has been little formal investigation into the effect of latency on perceived instrument quality. Wessel and Wright's 10ms threshold appears to derive from longtime practical experience rather than controlled experiments. In this paper we present a study that examines these effects, approaching the notion of control intimacy [4] through an investigation of perceived instrument quality. Our aim is to test, in a musical context, the impact of latency on performers' subjective quality judgements of a digital musical instrument even before they become consciously aware of a delay, and in turn to see how this is reflected in the way they interact with the instrument.

To test this we designed an experiment based around a novel percussive instrument where we could have submil-

lisecond control of the amount of latency. As a methodological move we decided to deliberately mask the fact that latency was the subject of the study – participants were told that they were evaluating the quality of different settings on a novel musical instrument and latency was not mentioned. The first part of the experiment involved subjective reports of instrument quality in a free improvisation task, where a latency condition was compared against a zero latency condition according to six quality measures. The second part measured temporal and dynamic performance during a series of rhythmic tasks, again with variable amounts of latency, followed by a structured interview.

Section 2 of this paper introduces the instrument used in this study, while Section 3 explains the design of the experiment. This is followed by results from the experiment in Section 4 and finally by a discussion of the results in the context of digital musical instrument design.

2. THE INSTRUMENT

2.1 Mechanical setup

We built a self-contained percussive digital musical instrument from eight ceramic tiles of varying sizes. Ceramic tiles were chosen as the main control interface of the instrument for ecological reasons: a ceramic tile is an object which holds associations of a immediate and sharp sonic response due to the properties of the material. Ease of playability was another consideration: the affordances of the instrument encourage finger percussion and tapping, a type of instrumental control that musicians could easily achieve without a steep learning curve. The novelty of the instrument was also an important factor: our participants were to feel like they were evaluating something new, bringing fewer prior conceptions to the instrument, than they would to drum pads or other commercial interfaces.

2.2 Electronic hardware

The instrument was created using the Bela platform[1] [15], a hard real-time sensor and audio processing platform built on the BeagleBone Black which is capable of submillisecond action-to-sound latency [14]. Each of the tiles has a piezo disk mounted to the back which is fed into the analog inputs of Bela. Striking the tiles triggers samples of Gamelan percussion instruments with only one dimension of control, amplitude. All sensor and audio processing was done on Bela, with an additional computer used only to switch between settings during the study.

The piezo disks were attached to the back of the tiles using scotch pliable mounting tape. Each of the tiles was held in position on a foam and plywood mount. The natural resonance of the tiles were dampened by gluing a layer of 3 millimetre rubber foam to the back of each of them. This also helped to condition the signal we received from the tile while attenuating the acoustic sound of the impact.

2.3 Peak detection and filter group delay

Bela was running with a sample rate of 44.1KHz and an audio buffer size of 16 samples. The peak detection routine includes a DC offset filter, full-wave rectification and a moving average filter. The peak detection algorithm detects strikes on each of the tiles by looking for a downward trend in

---

[1]http://bela.io

Figure 1: Image of the instrument as it was set up for the study.

the data when the current value is above a minimum threshold. Once a peak is detected the amplitude of the strike is measured and then assigned to the sample appropriate to the tile. Our synthesis engine was capable of 40 different voices with an oldest-out voice stealing algorithm if all voices became allocated, to allow for fast repeated strikes. The peak detection and triggering routine remained constant throughout the experiment while the latency condition and sample set changed.

Group delay from the peak detection together with the audio buffering delay create a base latency of 0.8ms. We call this the "zero latency" condition as the distance between the tiles and the ears would normally contribute around 2ms of acoustic latency, and the sound of the instrument was monitored directly through headphones.

2.4 Sample sets

Four sample sets were used in the experiment. Each sample set consists of eight audio samples. The sample sets were further grouped in two treatments characterised by perceptual acoustic features of the attack transients. All sounds across both treatments were of equal duration, sounds within each treatment have equal variance of pitch and equal attack time. Treatments differed in spectral centroid during the initial strike. The two differing groups can be broadly classified as brilliant and dull (striking a metallic bar with a metal beater, striking a metallic bar with a padded beater). Samples were always arranged on the instrument according to pitch height with the lowest notes mapped to the largest tile on the left hand side, the highest note to the smallest tile of the right hand side.

2.5 Sensor logging

Throughout the experiment sensor and audio data was recorded from the instrument onto an SD card by Bela for later analysis. This included the raw signal from each of the eight piezo disks attached to the tiles, the audio output and the audio input that they were playing along with.

3. EXPERIMENTS

3.1 Experiment design

3.1.1 Participants

Eleven participants took part in the study, three female and eight male, whose age was between 26 and 35 years. Eight of the eleven participants classified themselves as instrumentalists and the other three as electronic musicians. The instrumentalists had an average of 14 years playing experience on their first instrument, the electronic musicians an average of 10 years making electronic music. All but two of the participants had used a computer to make music, with six of the participants regularly using the combination of a hardware controller and software instrument to compose and perform music. The study lasted for around one hour and consisted of two parts. Participants were video and audio recorded throughout the experiment.

3.1.2 Latency conditions

Three latency conditions were tested, always relative to the zero latency condition, *Condition A*:

- *Condition B*: 10ms latency

- *Condition C*: 20ms latency

- *Condition D*: 10ms latency $\pm$ 3ms latency (simulated jitter). Each strike was assigned a random latency between 7ms and 13ms.

These three specific latency conditions were chosen based on a recent series of measurements conducted by McPherson et al [14]. It shows that 10ms is roughly the best achievable target with microcontrollers attached to computer music software by a serial link (a common setup for the creation of digital musical instruments) and that such programs exhibit jitter at intervals of the audio block size, often several milliseconds in either direction. On Bela, which powers our instrument, inherent jitter is no more than a single sampling period (0.03ms) [15].

3.1.3 Setup

For the experiment the instrument was mounted on a stand in a sound-isolated studio (see Figure 1). Participants monitored the instrument directly, plugging noise-cancelling headphones into the back of the instrument. Throughout the study white noise was played through monitors in the room at a level where all acoustic sound from the instrument was inaudible when the participant was performing. This was to avoid participants hearing any excess sound coming through air conduction from their contact with the instrument, focusing their attention on the sound that the instrument produced and the feel of the strike.

3.2 Part 1: Quality assessment

This part of the experiment was inspired by Fontana et al.'s study on the subjective evaluation of vibrotactile cues on a keyboard [5]. In this study they assess the impact of different vibrotactile feedback routines on the perceived quality of a digital piano. Our methodology and analysis in Part 1 takes a similar route. In this first section participants were able to switch between two settings, α and β. They were asked to freely improvise until they were able to

comparatively rate the two settings according to six quality metrics. They then moved onto the next pair.

3.2.1 Stimuli and conditions

To mask the changing latency conditions the sample set was also changed between α and β. We deliberately wanted to mask the changing latency conditions to answer the following questions: firstly, are the latency conditions perceivable by the participants; secondly, what impact do the latency conditions have on quality assessments of an instrument. Participants were simply instructed that they would be comparing different settings on the instrument and that they were to compare the performance of the instrument under each one and to try and not base their ratings on sample set alone.

3.2.2 Design and procedure

Switching between α and β was controlled via a laptop which hosts a graphical user interface built in PureData[2] which communicated with the Bela board via UDP, allowing participants to switch at will. The zero latency condition (A) was randomly assigned to either α or β while the other setting in the pair would always contain a latency condition (B, C or D). Two of the four sample sets were also selected at random for α and β. There were twelve such pairs for each participant, ensuring that each sample set was in the zero latency position 3 times per participant. The order of presentation also being randomised. Participants were not informed of what was changing between α and β. Participants were advised to take around 35 minutes to complete the evaluation of the 12 pairs.

The participants were informed that they were to comparatively evaluate the two conditions according to six attributes: Responsiveness, Temporal Control, Dynamic Control, Naturalness, Engagement, General Preference. While making these choices participants were to improvise freely with no restrictions to their chosen style.

Ratings were input via slider input using a Continuous Category Rating scale (CCR), a rating widely used in subjective quality assessments of interfaces (recommendation ITU-T P.800). Participants moved the slider on the continuous scale to rate the relative merits of the two settings (see Figure 2). The scale had the following titles along its scale:

- α is much better than β
- Both α and β are equal
- β is much better than α

The six attributes (Responsiveness, Engagement, Naturalness, Dynamic Control, Temporal Control, General Preference) were selected based on recent research into instrument quality evaluations [19] and based on the qualities we hypothesised would be most relevant to the changing latency conditions.

3.3 Part 2: Timing assessment

3.3.1 Stimuli and conditions

In this part of the study the experimental setup was the same as Part 1 but in each of the following two tasks the sample set was not changed during tasks. The four latency

conditions (A, B, C and D) were presented to the participant in a random order as they completed the following tasks.

3.3.2 Task 1 - metronome

A metronome at 120 bpm was played through the headphones. The participant was instructed to tap along with the beat using one tile only in four ways: every crotchet (quarter note) which is equivalent to the 120 bpm of the metronome, then every quaver (eighth note), every semiquaver (sixteenth note), and finally only the quavers between the metronome beat. They performed each of these tapping exercises for at least four bars, paused and then moved onto the next. Participants then repeated this sequence for each of the latency conditions. Once completing the task participants rated the four conditions in terms of relative difficulty, moving four sliders that ranged from easy to hard, one to represent each condition. Our methodology in this part of the study was derived from Fujii et al.'s study on synchronisation of drum kit playing [7].

3.3.3 Task 2 - rhythmic improvisation

A backing track of conga drums at 128 bpm was played through the headphones. The participant was instructed to develop a rhythmic improvisation of approximately 30 seconds that was in time with the backing track using all the tiles. They were given around 3 minutes to do this. They then performed a version of this improvisation under each of the four latency conditions. Again, upon completing this task participants rated the four conditions in terms of relative difficulty.

3.4 Structured interview

At the end of the experiment a structured interview lasting around 6 minutes was conducted where the following themes were discussed:

1. General impression of the instrument

2. Techniques used to distinguish between α and β in Part 1, the free improvisation

3. Whether they noticed what was changing between setting, besides sample set

4. RESULTS

In this section we combine qualitative and quantitative data collected throughout the study to explore the effects of latency on instrumental interaction.

4.1 Quality assessments

The difference in subjective judgements of instrument quality was evaluated by looking at the quality ratings from Part 1 of the study. The mean quality ratings for all participants are presented in Figure 3. -100 on the scale corresponds to 'α is much better than β ' option, 100 with the 'β is much better than α' option. In Figure 3 the zero latency condition A is always α for legibility, although in the study it was randomly assigned to either α or β.

To assess agreement between participants the Lin concordance correlation was calculated for each quality and pair of participants. The average ρ_c was as follows: Responsiveness -0.05, Engagement 0.014, Naturalness -0.023, Dynamic Control 0.012, Temporal Control 0.04, General Preference

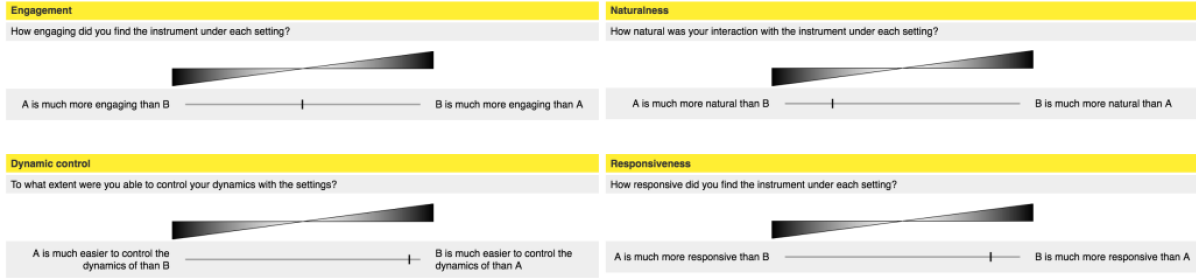Figure 2: Example of the slider input for the Continuous Category Ratings of instrument quality

Latency	Eng.	Resp.	Nat.	Dyn.	Temp.	Gen.
B	0.4	-2.5	**-20.9**	-4.4	0.9	-1.4
C	-14.8	-19.9	-15.6	-12.6	-26.3	-22.9
D	-14.4	-19.0	-18.9	-21.2	-23.0	-28.1

Table 1: Mean ratings over all participants for each quality and latency condition. Significant differences in bold ($p < 0.05$). 100 represents β is much better than α, 0 represents α is the same as β, -100 represents α is much better than β.
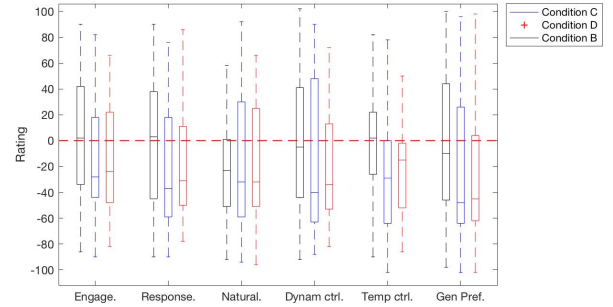

Figure 3: Quality ratings across all participants. Boxplot presenting median and quartiles for each quality and latency condition. On the y axis 100 represents β is much better than α, 0 represents α is the same as β, -100 represents α is much better than β.

-0.04. This highlights a high degree of variability in opinions between participants: for all of the quality measures there were at least two participants who disagreed in General Preference almost completely.

Participant responses were positively correlated between all quality measures. The highest correlation was observed between General Preference and Engagement (Spearman correlation $\rho_s = 0.85$), and the lowest between Engagement and Dynamic Control ($\rho_s = 0.23$). A partial correlation was observed between General Preference and the other quality measures and were as follows: $\rho_s = 0.59$ for Responsiveness, $\rho_s = 0.56$ for Naturalness, $\rho_s = 0.48$ for Dynamic Control, and $\rho_s = 0.51$ for Temporal Control.

Results are plotted in Figure 3, and the mean ratings for each quality scale and latency condition are given in Table 1. On average condition A, the zero latency condition, was rated more positively for all qualities than condition C and D, the 20ms and 10ms with jitter conditions. There is no significant difference in the quality ratings between condition A and condition B, the 10ms latency condition, aside from for Naturalness. For conditions C and D Temporal Control shows the strongest preference. This is followed by Dynamic Control, Naturalness and Responsiveness which all have very similar mean ratings for C and D. Naturalness is the only condition where B is rated significantly better than A; the others show no significant difference.

As the normality rule for Analysis of Variance was violated, a non-parametric Friedman test of differences among repeated measures was conducted. It gave a Chi-square value of 23.17 which was significant ($p < 0.01$), indicating that latency condition significantly affects quality judgements. A paired sample t-test was conducted to compare the quality ratings for each condition compared against condition A. There was a significant difference between the rating of B and the rating of C, $t(526) = 3.308$, $p = 0.001$. There was also a significant difference between the rating of B and the rating of D, $t(526) = 4.1520$, $p = 0.001$. There was how-

ever no significant difference between the rating of C and the rating of D, $t(526) = 0.5254$, $p = 0.599$.

4.2 Timing evaluation

4.2.1 Mean synchronisation error

In this paper our analysis of timing performance focuses only on task 1, playing with a metronome. For this analysis we compared the onset of the strike against the onset of the metronome tone, looking for the difference between the timing of the strike on the tile and the metronome tone rather than the audio output of the instrument, which might have added latency. The onset of each strike relative to that of the metronome was defined as the synchronization error (SE). The value was negative when the onset of the strike preceded that of the metronome and positive when the strike onset lagged behind the metronome.

Figure 4 presents the mean and standard deviation (SD) synchronisation error for the first rhythmic task for all latency conditions for all participants, as presented in Fujii et al's study of synchronisation in drum kit playing [7]. Mean synchronisation error can be seen to increase as the division of the metronome beat increases. Interestingly there is no constant anticipation displayed in relation to the latency conditions as we hypothesised. One-way ANOVAs for each latency condition showed that there were significant differences among the metronome conditions for all of the latency conditions in the crotchet, quaver, and semiquaver

metronome conditions, $F_{(2, 48)} = 14.92$, $p < 0.001$, $F_{(2, 48)} = 17.14$, $p < 0.001$, and $F_{(2, 48)} = 28.07$, $p < 0.001$, respectively. However a repeated measures ANOVA was insignificant for the interaction between latency condition and synchronisation error within each metronome condition ($F_{(2, 764)} = 0.83$, $p > 0.05$) suggesting that the difference in timing error was not significantly impacted by latency condition in the case of our participants.

Standard deviation of mean synchronisation error also shows an increase in deviation as the divisions of the metronome increase in speed. For both the crotchet and quaver metronome condition the standard deviation of mean synchronisation error for 10ms with jitter latency condition was significantly larger than that of zero latency condition ($p < 0.05$) and 10ms ($p < 0.05$) suggesting that the jitter condition led to more variation in timing than the zero and stable 10ms condition.

None of the eleven individual participants performed the rhythmic tasks with an accuracy better than the drift in latency condition D, ± 3ms, i.e. all participants had a variation in mean absolute synchronisation error of > 3 ms for all latency and metronome conditions including the zero latency condition.

4.2.2 Perceived difficulty

The 20ms latency condition C was on average rated more difficult than the other three, however the difference was not significant (see Figure 5). This was when presented with the four latency conditions one after the other in a randomised order.

4.3 Key themes from structured interview

Structured interviews were coded, here we present the major themes. This amount of latency is very subtle: only three out of the eleven participants stated that it was latency or delay that was the changing factor between settings in Part 1, even when they had the same sample set four times with four different latency conditions. What was reported across many of the participants was a *changing responsiveness* and level of *dynamic control* of the instrument, that they imagined the triggering thresholds had been changed and that the instrument was *catching less of their strikes* and that the range of dynamic control had been changed. This led them to put *more effort* into playing each individual strike, i.e. hitting harder. Four of the eleven participants reported having to play with *more weight* under certain conditions. Some participants also acknowledged that under certain conditions they were struggling to maintain timing although not able to specifically identify that a delay or latency was the cause of this perceived lack of ability to maintain timing when asked:

> ...one was very difficult to keep some sort of stable timing on, while the other one just clicked for some reason and made a lot more sense. - Participant 4
>
> On the second one (condition A) I didn't have to put much thought into it or didn't have to tap myself in or anything. It was just there under my finger tips. - Participant 10

4.3.1 Velocity of strike

A variation in striking velocity was noted for the four participants who mentioned that certain settings led them to

Latency condition:	A.	B.	C.	D
Participant 6 (dB)	-10.37	-9.42	-8.87	-9.54
Participant 10 (dB)	-9.31	-8.95	-7.84	-6.25
Participant 4 (dB)	-9.66	-7.75	-9.14	-10.41
Participant 11 (dB)	-12.15	-10.78	-9.84	-10.99

Table 2: Mean strike velocity for Part 2, task 2: rhythmic improvisations. Values are in dB. Four participants.

put more effort into playing the instrument. Upon analysis of rhythmic task two, the rhythmic improvisation, we noted that the mean striking velocity was harder for latency conditions C and D in comparison with conditions A and B. A repeated measures ANOVA was significant for the mean velocity value of both latency conditions C and D in comparison with condition A: ($F_{(1, 58)} = 3.58$, $p < 0.05$) and ($F_{(1, 58)} = 4.50$, $p < 0.05$) respectively. This suggests that for these four participants latency condition impacted upon the mean velocity they were using to strike the instrument.

This was only significant for this subset of the sample and did not hold for the sample at large: across all eleven participants there was no significant difference in striking weight across latency condition ($p > 0.05$). Mean strike velocity is presented in Table 2.

5. DISCUSSION

5.1 Latency and quality assessments

The results from Part 1 suggest that latency of 20ms and 10ms ± 3ms can degrade the perceived quality of an instrument, even when the amount of latency is too small to be perceived as a delay by the performer. The fact that latency condition D, the jitter condition, was rated in a similarly negative manner as condition C, 20ms latency, but that condition B, 10ms latency, did not receive such negative ratings, highlights the importance of stability as well as low latency, which points to an agreement with Wessel's recommendations [20]. None of the eleven participants performed with a degree of accuracy in Part 2 that was better than the jitter amount (10ms ± 3ms), yet this condition was still rated negatively. This suggests that subtle variation in the stability of the temporal response of an instrument can be detected by performers even if they cannot perform with a degree of accuracy that is less than the jitter amount. The impact of the latency conditions was identified by participants as a changing dynamic response substantially more often than as a temporal factor in the structured interviews.

Dynamic control was rated negatively for the latency conditions latency of 20ms and 10ms ± 3ms even though there was no difference in the triggering routine. This was also highlighted in participant reports of increased force needed to trigger notes.

From the structured interview we can break down responsiveness into two related areas. Firstly, the perceived effort that it takes to produce a note – participants reported having to push harder to produce the desired note under the 20ms latency condition and jitter condition. Secondly, the perceived immediacy of control of dynamic and temporal variation – whereas with the no latency condition some participants reported a feeling of ease of control, that the notes were just there under their finger tips, with the latency con-
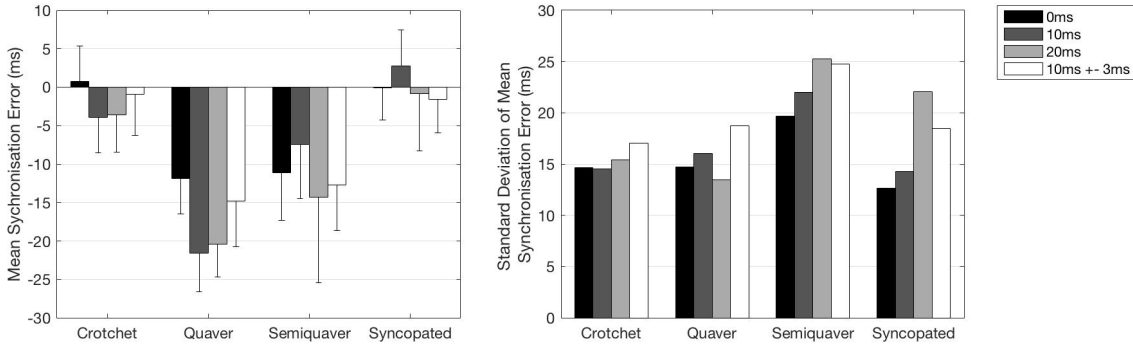
Figure 4: **Mean synchronisation error for all participants (MSE). Error bars indicate between-participants standard error (n = 11). Standard deviation (SD) of synchronization error (n=11).**

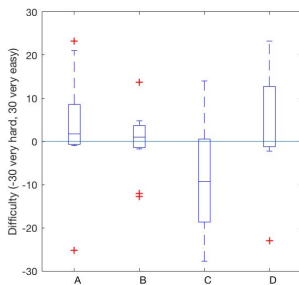

Figure 5: **Difficulty ratings from Part 2. Boxplot presenting median and quartiles for each quality and latency condition.**

ditions some reported that with fast passages the instrument would not track their gestures properly and they would trip up on notes they had already played. This, alongside participant reports of the sound being "under the fingertips" and "just there" with the zero latency condition, can help illuminate the particular form that control intimacy takes in this study, as well as the reasons why it is of central importance to digital musical instrument design at large.

5.2 Latency and timing performance

With each of the levels of latency we introduced participants continued to sync with tactile feedback rather than anticipating their strike to sync with audio as observed by Dahl [2] when evaluating a larger range of latency (0-110ms) which was progressively increased. This might suggest that progression of presentation is a factor in the anticipation of gesture to delayed sound. In the case of our study the participant seems to be able to deal with a given latency condition quickly: with relatively low levels of latency the delay seems to be taken as part of the instrument's behaviour rather than identified as a noticeable delay.

The group means of the MSE ranged from -21 to 5 ms for all metronome and latency conditions. The mean standard deviation ranged from 15 to 25 ms. Both were larger than that found by Fujii et al in their study with highly trained percussionists [7] where a mean synchronisation error of -13 to 10ms was achieved for a metronome with standard devia-

tions of 10 to 16ms. In our study our participants, although all with a high degree of musical experience, did not have a between-participant consistency of degree of specialised training as in Fujii et al's study. Future analysis will aim to look in more detail at the difference in performance between individual participants and the influence of musical training.

6. CONCLUSION

We find that 20ms latency showed significantly lower ratings of quality and a higher level of difficulty compared to the zero or 10ms latency conditions. These results lend support to Wessel's guideline [20] that digital musical instruments should aim for a latency of 10ms or less with 1ms jitter. The fact that the 10ms latency condition with ± 3ms jitter was rated similarly to the 20ms condition suggests that a measurement of mean latency is not sufficient to tell whether a digital musical instrument is sufficiently responsive: stability seems to be a crucial factor.

Consistent 10ms latency shows no significant difference in most quality ratings compared to a zero latency condition, with the sole exception of the rating Naturalness, whose difference could be statistical noise, or may be a subtle difference between conditions that would be amplified with more participants, trained professionals, or more time spent with the instrument. We also find no difference in performance accuracy under this condition compared with zero latency. Our evaluation is limited to our sample size and a further study with a larger population of musicians would allow us to draw further conclusion about what effects on interaction the difference between submillisecond and 10ms of latency could possibly have. Future analysis will include an evaluation of the difference in synchronisation error between Part 1 and Part 2 of the rhythmic task to see if latency has a more pronounced impact on performance when the participants are partaking in a more musical task. The difference in performance between participants will also be explored.

6.1 Latency and instrumental interaction

Our results from this study show that even if the level of latency is below the degree of accuracy that can be achieved by the performer on an instrument it can still impact on how the quality of that instrument is judged. In this study none of the participants were able to perform with a degree of accuracy that was better than the jitter condition (± 3ms)

yet this condition was rated negatively in comparison to the zero latency condition. This again highlights the importance of stability of latency.

6.2 Latency as a tool in instrument design

Subtle changes in the level of latency, so long as stable, also seem to demonstrate an impact on the *feel* of the instrument to the performer: it can create differences in how the performer judges the effort and perceived weight of strike that is needed to trigger a note. A parallel can perhaps be drawn to mechanical latency in acoustic instruments: a well-regulated piano will have predictable action but softer key presses exhibit quite high latency in comparison to harder strikes [1]. Once you have complete control of the amount and stability of latency in a digital musical instrument it may be possible to stop considering it as an obstacle to intimate control and to see it rather as a tool that can be deployed in the design process to create similar multisensory effects.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] A. Askenfelt and E. V. Jansson. From touch to string vibrations - the initial course of the piano tone. *Dept. for Speech Music and Hearing, Quarterly Progress and Status Report*, 29(1):31–109, 1988.

[2] S. Dahl and R. Bresin. Is the player more influenced by the auditory than the tactile feedback from the instrument? In *Proc. of the COST-G6 Workshop on Digital Audio Effects (DAFx-01), Limerick*, pages 194–197. Citeseer, 2001.

[3] G. Essl and S. O'modhrain. An enactive approach to the design of new tangible musical instruments. *Organised Sound*, 11(03):285–296, 2006.

[4] S. Fels. Designing for intimacy: Creating new interfaces for musical expression. *In Proc. of the IEEE*, 92(4):672–685, 2004.

[5] F. Fontana, H. Järveläinen, S. Papetti, F. Avanzini, G. Klauer, L. Malavolta, C. di Musica, and C. Pollini. Rendering and subjective evaluation of real vs. synthetic vibrotactile cues on a digital piano keyboard. In *Proc. of the Sound and Music Computing Conference 2015, Maynooth, Ireland*, 2015.

[6] A. Friberg and J. Sundberg. Time discrimination in a monotonic, isochronous sequence. *The Journal of the Acoustical Society of America*, 98(5):2524–2531, 1995.

[7] S. Fujii, M. Hirashima, K. Kudo, T. Ohtsuki, Y. Nakamura, and S. Oda. Synchronization Error of Drum Kit Playing with a Metronome at Different Tempi by Professional Drummers. *Music Perception: An Interdisciplinary Journal*, 28(5):491–503, 2011.

[8] T. Kaaresoja, S. Brewster, and V. Lantz. Towards the Temporally Perfect Virtual Button : Touch-Feedback Simultaneity and Perceived Quality in Mobile Touchscreen Press Interactions. *ACM Transactions on Applied Perception*, 11(2), 2014.

[9] T. Kaaresoja, E. Hoggan, and E. Anttila. Playing with tactile feedback latency in touchscreen interaction: two approaches. In *Proc. of the IFIP Conference on Human-Computer Interaction*, pages 554–571. Springer, 2011.

[10] N. Lago and F. Kon. The quest for low latency. In *Proc. of the International Computer Music Conference*, pages 33–36, 2004.

[11] M. Leman. *Embodied music cognition and mediation technology*. MIT Press, 2008.

[12] T. Magnusson and E. H. Mendieta. The acoustic, the digital and the body: A survey on musical instruments. In *Proc. of the 7th International Conference on New Interfaces for Musical Expression*, pages 94–99. ACM, 2007.

[13] T. Mäki-Patola and P. Hämäläinen. Latency tolerance for gesture controlled continuous sound instrument without tactile feedback. In *Proc. International Computer Music Conference (ICMC)*, pages 1–5, 2004.

[14] A. McPherson, R. H. Jack, and G. Moro. Action-sound latency: Are our tools fast enough? In *Proc. of the International Conference on New Interfaces for Musical Expression*, 2016.

[15] A. McPherson and V. Zappi. An environment for submillisecond-latency audio and sensor processing on beaglebone black. In *Audio Engineering Society Convention 138*. Audio Engineering Society, 2015.

[16] S. O'modhrain. A framework for the evaluation of digital musical instruments. *Computer Music Journal*, 35(1):28–42, 2011.

[17] B. H. Repp and Y. H. Su. Sensorimotor synchronization: a review of recent research (2006–2012). *Psychonomic bulletin & review*, 20(3):403–452, 2013.

[18] D. Rubine and P. McAvinney. Programmable finger-tracking instrument controllers. *Computer Music Journal*, 14(1):26–41, 1990.

[19] C. Saitis, B. L. Giordano, C. Fritz, and G. P. Scavone. Perceptual evaluation of violins: A quantitative analysis of preference judgments by experienced players. *The Journal of the Acoustical Society of America*, 132(6):4002–4012, 2012.

[20] D. Wessel and M. Wright. Problems and Prospects for Intimate Musical Control of Computers. *Computer Music Journal*, 26(3):11–14, 2002.