Action-sound Latency and the Perceived Quality of Digital Musical Instruments: Comparing Professional Percussionists and Amateur Musicians

Robert H. Jack, Adib Mehrabi, Tony Stockman, & Andrew McPherson
Queen Mary University of London, London, United Kingdom

Asynchrony between tactile and auditory feedback (action-sound latency) when playing a musical instrument is widely recognized as disruptive to musical performance. In this paper we present a study that assesses the effects of delayed auditory feedback on the timing accuracy and judgments of instrument quality for two groups of participants: professional percussionists and non-percussionist amateur musicians. The amounts of delay tested in this study are relatively small in comparison to similar studies of auditory delays in a musical context (0 ms, 10 ms, 10 ms ± 3 ms, 20 ms). We found that both groups rated the zero latency condition as higher quality for a series of quality measures in comparison to 10 ms ± 3 ms and 20 ms latency, but did not show a significant difference in rating between 10 ms latency and zero latency. Professional percussionists were more aware of the latency conditions and showed less variation of timing under the latency conditions, although this ability decreased as the temporal demands of the task increased. We compare our findings from each group and discuss them in relation to latency in interactive digital systems more generally and experimentally similar work on sensorimotor control and rhythmic performance.

Playing a musical instrument represents a highly developed sensorimotor skill, where years of training and theoretical knowledge are brought together into the nuanced and expressive control required for musical performance. Delayed feedback (be it auditory, visual or tactile) can cause disruption to this sensorimotor control. Previous studies have shown that delayed auditory feedback (DAF) during music performance can disrupt musical production, primarily by increasing the variability of timing (Pfordresher, 2006; Pfordresher & Palmer, 2002; Yates, 1963). This disruption varies with delay length and similar effects have been shown for DAF in speech (Howell, 2001).

In the field of human computer interaction, delayed feedback has mostly been studied as system latency (the asynchrony between a control gesture and a system's corresponding response) and jitter (the variability of this asynchrony). Latency is a fundamental issue affecting interactive digital systems and has long been recognized as potentially harmful to a user's experience of control (MacKenzie & Ware, 1993; Meehan, Razzaque, Whitton, & Brooks, 2003): even if accuracy of temporal performance is not impacted, the qualitative experience of the user may be negatively impacted (Kaaresoja, Anttila, & Hoggan, 2011). The way latency and jitter affect a user has been shown to vary greatly depending on the specific demands of the task and situation (Annett, 2014); for example, direct or indirect touch, tapping or swiping on a touchscreen (Jota, Ng, Dietz, & Wigdor, 2013; Ng, Lepinski, Wigdor, Sanders, & Dietz, 2012).

In this paper we present a study that investigates the effects of small amounts of action-sound latency and jitter (10 ms, 10 ± 3 ms, 20 ms) on the interaction of musicians with a digital percussion instrument. We assess both the musicians' judgments of instrument quality under different latency conditions and their timing accuracy. Two groups of participants took part in this study: non-percussionist amateur musicians and professional percussionists. Our aim with this research was to assess the impact of relatively small amounts of delay on the fluency and quality of the interaction, even when the auditory feedback is not perceived as detached from the action that produced it (the commonly accepted threshold for perceived audiotactile simultaneity can vary between 20 ms and 70 ms (Occelli, Spence, & Zampini, 2011). We also examine whether extensive rhythmical training and the demands of the musical task affect the influence of DAF on performance.

Context of Research

ALTERED AUDITORY FEEDBACK AND MUSICAL PERFORMANCE

The altered auditory feedback (AAF) paradigm has been used extensively in music psychology to investigate the importance of auditory information for the execution of control sequences on musical instruments. Delayed auditory feedback (DAF) is a common form of AAF where the onset of auditory feedback is delayed by a fixed amount in relation to the action that produced it (Black, 1951). While the contents of anticipated feedback events are usually maintained in DAF, with only the synchrony of perception and action being affected, there are other types of AAF that alter the contents of auditory feedback while maintaining synchrony. For example, experiments have been conducted on digital keyboards where the AAF consists of shifting pitches to disrupt expectations of pitch arrangements on the keyboard (Pfordresher, 2003, 2008; Pfordresher & Palmer, 2006) or randomizing pitch (Finney, 1997; Pfordresher, 2005).

Each kind of alteration to auditory feedback disrupts performance in different ways and to different extents. Recent research on delayed feedback suggests that asynchronies between action and feedback primarily disrupt the timing of actions, not their sequencing (the production of melodic sequences) (Pfordresher, 2003). Pitch alterations, on the other hand, disrupt the accuracy of production and not timing (Pfordresher, 2003). The point of maximal disruption caused by asynchronies (the amount of delay, above which no significant increase in disruption is seen) has been the focus of much research. Generally, disruption increases as the delay increases up to a certain point, and then reaches asymptote (Gates, Bradshaw, & Nettleton, 1974, found an asymptote around 270 ms in music performance). However, rather than an absolute time discrepancy, the degree of disruption caused by asynchronies depends on when it occurs in the interonset interval (IOI) in rhythmic performance, and reflects the phase relationships between onsets of auditory feedback relative to the IOI between actions (key presses for example) (Pfordresher, 2006).

THE SENSORIMOTOR CONFLICT HYPOTHESIS

A common interpretation of disruption from delayed feedback is the sensorimotor conflict hypothesis. The proposal is that delayed feedback interferes with the planned timing of actions (Pfordresher, 2006) or their execution (Howell, 2001) due to shared representations for perception and action (MacKay, 1987). Delayed feedback causes disruption by conflicting temporally with the expected timing of a planned movement (Pfordresher & Dalla Bella, 2011), in this case the expected sound that should result from an action. The magnitude of disruption most likely depends on the perceptual salience of the delayed feedback (Stenneken, Prinz, Cole, Paillard, & Aschersleben, 2006).

Our concern is with audiotactile interactions: the tight coupling between auditory and tactile feedback systems has been recognized (Occelli et al., 2011), as has its increased temporal resolution of synchrony perception in comparison to audio-visual and tactile-visual (Fujisaki & Nishida, 2009). Whereas the importance of auditory feedback for musical performance is evident, given the primary aural focus of music as a cultural practice, tactile feedback has been shown to play an important role in the control of timing during music performance (Goebl & Palmer, 2008) and expert performers have been shown to depend less on auditory feedback and more on tactile feedback than non-expert performers during the performance of sequential movements (van der Steen, Molendijk, Altenmüller, & Furuya, 2014). High temporal acuity is shared by both hearing and touch. In terms of temporal precision hearing is the most accurate of our senses: two stimuli of equal subjective intensity were perceived as being temporally discrete if separated by ca. 2 ms for monaural and binaural stimulation (Levitin et al., 2000), touch being less accurate (ca. 10–12 ms; Gescheider, 1966) but still better than sight (ca. 25 ms; Kietzman & Sutton, 1968).

MULTISENSORY INTEGRATION AND SIMULTANEITY PERCEPTION

Multisensory integration is the process by which the human nervous system merges available sensory information into unique perceptual events (Calvert, Spence, & Stein, 2004). Joining stimuli received through separate sensory channels can take place between stimuli that are temporally asynchronous, but which fall within the "temporal window" of integration (Meredith, Nemitz, & Stein, 1987). For audiotactile stimuli this can vary from tens to hundreds of milliseconds wide depending on various factors to do with the location, magnitude and content of the stimuli (Occelli et al., 2011). An important measure is the point of subjective simultaneity: the "amount of time by which one stimulus has to precede or follow the other in order for the two stimuli to be perceived as simultaneous" (Spence & Parise, 2010, p. 365).

While many studies of DAF deal with amounts of delay of 50 ms and above, research on sensorimotor synchronization (typically in the guise of non-musical finger tapping studies) has yielded sophisticated models of sensorimotor timing with auditory feedback on the

level of tens of milliseconds (see Repp & Su, 2013, for a review). Tapping studies are useful for examining how movements are synchronized with an auditory stimulus and help enrich our understanding of sensory pathways and the weighting of signals in cross-modal perception. Levitin et al. (2000) and Adelstein, Begault, Anderson, Wenzel, and Field (2003) investigated the perceptual asynchrony threshold values for an active audiotactile interaction situation (playing a drum). The threshold value of 42 ms was reported by Levitin whereas for Adelstein et al. thresholds varying between 18 ms to 31 ms depending on the stimulus duration were reported. These studies and others report that some participants had very low threshold values (ca. 10 ms), particularly musicians (Adelstein et al., 2003; Levitin et al., 2000; Repp & Doggett, 2007).

NEGATIVE MEAN ASYNCHRONY AND ADAPTATION

When tapping along to a metronome beat, participants commonly exhibit a negative mean asynchrony (NMA): they anticipate the beat and strike early by between 30 and 10 ms (see Aschersleben, 2002; Repp, 2000; Repp & Doggett, 2007). A measure of NMA and the variability of this asynchrony are a common way of assessing the temporal accuracy of a performer (Repp & Su, 2013). The bidirectional influence of auditory and movement information is evident in many simple tapping studies where auditory information guides motor timing: audiotactile stimuli with delays to auditory feedback cause anticipations to increase with the amount of delay (Aschersleben & Prinz, 1997) when delay is gradually introduced up to 70 ms in a tapping test, whereas NMA reduces with deafferented participants with only auditory and visual feedback (Stenneken et al., 2006).

This measure can be significantly affected by adaptation to asynchrony (see Vroomen & Keetels, 2010, for a review). The adaptation process is typically evaluated by measuring participants' perceptions of crossmodal simultaneity both before and after an exposure period, where there is a constant feedback delay between the stimuli presented in the two modalities. Vroomen and Keetels (2010) describe this as a widening of the temporal window for multisensory integration. The temporal window for audiotactile integration has been shown to widen in response to a relatively short exposure to asynchronously presented tactile and auditory stimuli in the case of passive tactile perception (Navarra, Soto-Faraco, & Spence, 2007).

MUSICIANS' TIMING ABILITY

Musicians have been recognized as better than nonmusicians across a range of timing dependent tasks. In duration-based tasks where the duration of two intervals are compared musicians outperform nonmusicians (Rammsayer & Altenmüller, 2006). Musicians also show a superior ability to distinguish timing changes within isochronous sequences (Lim, Bradshaw, Nicholls, & Altenmüller, 2003), which is particularly true of percussionists who demonstrate the highest accuracy of all musician groups (Ehrlé & Samson, 2005). The NMA when tapping to an isochronous sequence is also notably smaller for amateur musicians in comparison to nonmusicians (10–30 ms vs. 20–80 ms) (Aschersleben, 2002; Repp & Doggett, 2007). Further differences between instrument speciality have been demonstrated: participants with high levels of rhythm-based musical expertise (in particular, percussionists) display superior synchronization abilities (smaller NMAs with less variability in tapping tasks) when compared to other musicians and nonmusicians (Cameron & Grahn, 2014; Krause, Pollok, & Schnitzler, 2010; Manning & Schutz, 2016). Dahl (2000) reported that professional percussionists demonstrated a variation of mean synchronization error of between 10–40 ms, which equated to 2–8% of the associated tempo. Even lower synchronization error in professional drummers has been reported by Kilchenmann and Senn (2011; between 3 ms and 35 ms depending on motor effector, part of the drum kit and rhythmic "feel") and Hellmer and Madison (2015; below 5 ms).

FEEDBACK DELAYS IN HUMAN-COMPUTER INTERACTION

The effects of asynchronous multisensory feedback have also been extensively studied in the field of human-computer interaction, where latency and jitter are unavoidable side effects of digital systems and their linkages between virtual and physical worlds. Due to the current proliferation of touchscreen technologies, much recent research has focused on acceptable levels of latency in such devices: Ng et al. (2012) have shown that visual-tactile latency well under 10 ms can affect user preference, even when no delay is perceived. When examining multisensory latency in touchscreen buttons Kaaresoja, Brewster, and Lantz (2014) suggest that latency should be lowest for the tactile channel (5–50 ms), followed by audio (20–70 ms) and finally the visual (30–85 ms). However, this is task dependent: for direct touch systems the threshold of noticeable visual-tactile latency has been shown to be as low as 69 ms for tapping and 11 ms for dragging (Deber, Jota, Forlines, & Wigdor, 2015).

DELAYED AUDITORY FEEDBACK IN ACOUSTIC MUSICAL INSTRUMENTS

Auditory delays within an acoustic musical context are multifaceted and commonplace. Lago and Kon (2004)

point out the variability of the effects of such delays and their dependence on instrument, style of music, and spatial positioning: in ensemble playing, for example, delays ranging from 10 ms to 40 ms can often be present due to the distance between players and the speed of sound (Chafe & Gurevich, 2004). Lester and Boley (2007) provide a comprehensive overview of the effects of latency during a live sound monitoring situation in a recording studio and found that sensitivity to latency was highly dependent on instrument type and monitoring style (in-ear versus wedge monitors). As latency increased it became less of a spectrum affecting issue and more of a temporal perception issue (above 6.5 ms caused temporal smearing with certain instruments for in-ear monitoring). The low thresholds found in their paper are in part due to the specifics of the live monitoring situation where acoustic and delayed sound are combined.

DELAYED AUDITORY FEEDBACK IN DIGITAL MUSICAL INSTRUMENTS

The close coupling of action to sound via the virtual mechanism of the computer is of prime importance to building compelling digital musical instruments (DMIs). Wright, Cassidy, and Zbyszynski (2004) stated that a few milliseconds of latency and jitter can make the difference between a responsive, expressive, satisfying real-time computer music instrument and a rhythm-impeding frustration. Due to the complexity of the sensorimotor control that a musician has over an instrument and the high demands of musical performance, we believe that DMI design is a good testing ground for understanding the effect of latency and jitter in human computer interaction more broadly, complementing research done in relation to musical disruption caused by DAF.

Latency has been identified as a barrier to virtuosic engagement by obstructing a fluent interaction with the instrument (Magnusson & Mendieta, 2007; O'Modhrain, 2011; Wright et al., 2004). DMI designers have often been recommended to aim to create instruments that support the kind of interaction possible with acoustic instruments; tools that foster a relationship between gesture and sound that is both intuitive yet complex (O'Modhrain, 2011). Latency does not generally affect acoustic instruments that produce sound in reaction to action instantaneously as the sound producing mechanism and control interface are one and the same.[1]

---

[1] 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 around 35 ms for pp notes and -5 ms for ff notes. 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 100 ms for pp notes and 25 ms for ff notes (Askenfelt & Jansson, 1988).

Wessel and Wright (2002) suggest that DMIs should aim for a latency of less than 10 ms with less than 1 ms jitter. McPherson, Jack, and Moro (2016) demonstrated that Wessel and Wright's guideline is still not met by many toolkits commonly used to create DMIs.

With digital musical instruments there are many parts of the system that introduce latency and jitter between action and sound: buffering in hardware and software, latency in the audio code itself (from frequency domain processing for example), transmission delay between sensors and audio engine due to USB connection, latency induced by smoothing or signal conditioning of the sensor input (McPherson et al., 2016; Wright et al., 2004). These factors can combine to create a significant delay between performer action and resultant sound and impede what Wessel and Wright (2002) describe as the development of *control intimacy* between performer and instrument. Control intimacy has been described by Fels (2004) as the perceived match of the behavior of an instrument and a performer's control of that instrument, a concept that is connected to the notion of *ergoticity* (Luciani, Florens, Couroussé, & Castet, 2009): the preservation of energy through both digital and physical components of a system, maintaining an ecologically valid causal link between action and sound to foster an embodied engagement with the instrument (Essl & O'Modhrain, 2006).

During performance with digital musical instruments, latency has been shown to affect accuracy of control and to be identified 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 just noticeable difference of latency was shown to be around 20–30 ms, with latencies as high as 100 ms going undetected during the performance of slow passages (Mäki-Patola & Hämäläinen, 2004). Dahl and Bresin (2001), in a study with "in-air" percussive digital musical instruments without tactile feedback, found that a latency of 40 ms negatively impacted timing accuracy but that up to around 55 ms performers were able to compensate for the latency by increasing their anticipation (moving their strike earlier) when latency was gradually introduced.

THE PRESENT STUDY

In the present study, a group of highly trained professional percussionists and a group of non-percussionists (with varying levels of musical experience) evaluated the effects of variable levels of delayed auditory feedback on a novel digital percussion instrument. Our aim is to investigate the differences in the effects of latency and

jitter on timing accuracy and perceived instrument quality between the two groups, and to understand the influence, if any, of specialized training in rhythm based musical practices on these measures.

Method

APPARATUS

The challenge of measuring multimodal delays in interactive systems has been explored by many, mostly oriented towards touchscreen interactions (Kaaresoja & Brewster, 2010; Schultz & van Vugt, 2016; Xia et al., 2014). In the present study, in order to counter the common problems of latency in a DMI and to have sufficient control over the exact amount of latency and jitter in the system we have used the Bela platform (bela.io) as the basis of our instrument, an open source platform designed for real time, ultra-low latency, audio and sensor processing. Bela provides sub-millisecond latency and *near* jitter-free synchronization (within 25 µs) of audio and sensor data (McPherson & Zappi, 2015), making it a suitable platform for controlling the exact amount of latency present in a DMI. Measurements of its performance can be found as part of the latency tests of common platforms used to create DMIs conducted by McPherson et al. (2016).

The instrument. We built a self-contained percussive digital musical instrument, the playing surface of which consists of eight ceramic tiles of varying sizes (see Figure 1). The instrument represents a "simplest case" digital percussion instrument with one dimension of control: discrete velocity triggering of samples. Each of the ceramic tiles has a piezo disc vibration sensor attached to the back with pliable scotch mounting tape. Mounts for the tiles were created from laser cut plywood and each tile was suspended by their antinodes, allowing them to vibrate freely when struck, ensuring a strong signal to the vibration sensor (Sheffield & Gurevich, 2015). A layer of 3 mm rubber foam was glued to the back of each tile to further condition the signal while attenuating the acoustic resonance of the tile.

The piezo sensors are connected via a voltage biasing circuit to the analog inputs of the Bela board. A striking action on the tiles induces vibration in the tile, which is passed through signal conditioning routines and a peak detection algorithm (detailed in the next section) before being used to trigger samples of Gamelan percussion instruments. The intensity of the strike is measured when a peak is detected and mapped to the amplitude of the sample playback.

Filter group delay and peak detection. The peak detection routine includes a DC blocking filter, full-wave
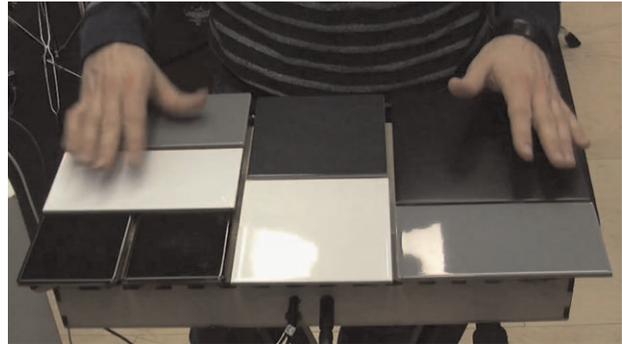

FIGURE 1. The instrument built from eight ceramic tiles with piezo discs attached to them to sense vibrations.

rectification and a moving average filter. Strikes were detected by looking for peaks in the sensor readings using an algorithm that looks for an increase in the reading followed by a downward trend once the reading is above a minimum threshold. When a peak is detected, the amplitude of the strike is measured and then assigned to the sample appropriate to the tile. Our synthesis engine had enough computation power to play 40 simultaneous samples and we used an oldest-out voice stealing algorithm if all voices became allocated, to allow for fast repeated strikes.

The audio output on Bela used a sample rate of 44.1 kHz and a buffer size of eight samples. The analog inputs used for the piezo discs were sampled at 22.05 kHz, synchronously with the audio clock. The total action-sound latency consists of the duration of the two buffers (360 µs) plus the conversion latency of the sigma-delta audio codec (430 µs). In addition, there is the group filter delay of the FIR filter (moving average) that was used to smooth the piezo signal over 20 samples before the peak detection of 1 sample, resulting in 250 µs delay. As the analog inputs and audio outputs are synchronized on an individual sample level, jitter between them is no more than 25 µs (McPherson et al., 2016). In the present study the sound of the instrument was monitored directly through noise-cancelling headphones. We conducted a test of the headphones to ensure that the noise-cancelling function was not introducing additional latency and found that when the noise cancelling was turned on there was an additional 100 µs latency in comparison to the analog signal path. We take this total of 1.2 ms latency as our Condition A, which we call the "zero latency" condition, as the distance between the tiles and the ears would normally contribute around 2 ms of acoustic latency due to the speed of sound.

Sound mapping. The use of a piezo vibration sensor naturally gives us an ergotic link (Luciani et al., 2009) between the force of a strike and the amplitude of the sound output. The response curve of this sensor is linear, unlike other commonly used sensors in electronic percussion instrument like force sensitive resistors (FSRs). For a full review of sensors commonly used in percussion instruments, see Medeiros and Wanderley (2014) and Tindale, Kapur, Tzanetakis, Driessen, and Schloss (2005). By using this sensor we were able to naturally preserve the relationship between physical energy at the input and perceived physical energy at the output by producing a linear software relationship between input level and output level.

In the present study four sample sets were used, each consisting of eight individual samples that were assigned to each of the eight ceramic tiles. All samples had equal duration, pitch variation, and perceptual attack time. The four sample sets were further divided into two groups characterized by the perceptual acoustic features of their attack transients. The difference between these two groups was in the spectral centroid during the initial strike: they could be broadly described as "bright" or "dull" sounding (striking a metallic bar with a hard beater versus striking a metallic bar with a padded beater). Pitch height was preserved on the instrument for each sample set moving with the lowest pitch notes mapped to the largest tile on the left hand side and the highest pitch note to the smallest tile of the right hand side, increasing vertically for each third of the instrument (see Figure 1).

The peak detection and triggering routine remained constant throughout the experiment while the latency condition and sample set changed. Throughout the experiment the raw signals from the instrument were recorded onto an SD card by Bela for later analysis. This included the signal from each of the eight piezo disks attached to the tiles, the audio output, and the metronome or backing track that the participants were monitoring through headphones in the second and third tasks.

Presentation. This study was conducted in a sound-isolated studio. The instrument was mounted on a keyboard stand whose height the participants could adjust for comfort. On a podium next to the instrument there was a laptop where the participants input their responses and changed the settings of the instrument. Participants monitored the instrument directly through noise cancelling headphones (Bose QC-25). White noise was played in the room through a PA system at a level (50 dB) 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 was presented through the headphones and their haptic experience of the strike.

PARTICIPANTS

Two groups of participants took part in the study we present. The first group (referred to as "non-percussionists" or "NP" from now on) consisted of 11 participants (3 female) whose age was between 26 and 45 years and who were recruited from our university department. All members of this group had musical experience but none were professional. Eight of the 11 participants classified themselves as instrumentalists and the other three as electronic musicians. None of this group had received training in percussion. These participants had varying degrees of music training (0–15 years; $M = 9.2$, $SD = 4.5$). 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/or perform music.

The second group (referred to as "professional percussionists" or "PP" from now on) consisted of 10 participants (1 female) whose age was between 26 and 35 years. They had completed at least a bachelors degree in performance specializing in percussion and were working professionally, either as a performer in an orchestra, as a session musician, or in education. This group had between 10 and 20 years of formal percussion training ($M = 13.8$, $SD = 2.5$). All participants in this group had received training on a second instrument (2–15 years; $M = 11.0$, $SD = 3.5$), most commonly piano in the case of six participants. Both groups reported normal hearing and normal or corrected-to-normal vision. This experiment met ethics standards according to our university's ethics board.

STIMULI

Four variable latency and jitter conditions were tested: 1) Condition A: 'zero' latency, 2) Condition B: 10 ms latency, 3) Condition C: 20 ms latency, and 4) Condition D: 10 ms latency $\pm$ 3 ms latency (simulated jitter).

These conditions were created by delaying the sound triggered by a detected strike by a set number of samples and were verified on an oscilloscope. In the jitter condition each strike was assigned a random latency between 7 ms and 13 ms. We chose these three specific latency conditions based on a recent series of measurements conducted by McPherson et al. (2016) of common techniques used to create digital musical

instruments. We also deliberately chose our maximum latency condition (20 ms) to be within the thresholds of simultaneity perception for audio-haptic stimuli as found by Adelstein et al. (2003) of around 24 ms. This was to focus our findings on the effects of latency and jitter when a delay is not necessarily perceived between action and sound.

DESIGN AND PROCEDURE

The study lasted for approximately one hr and 15 min and consisted of two sections followed by a structured interview. Participants were video and audio recorded throughout the experiment.

Part 1: Quality assessment. In order to evaluate the participants' subjective impression of quality of the different latency conditions on the instrument, we decided to use a method that involved participants rating the conditions in comparison to one another for a series of quality attributes. In this part of the study, latency conditions B, C, and D were always compared to condition A. This part of the experiment was inspired by Fontana et al.'s (2015) study on the subjective evaluation of vibrotactile cues on a keyboard. In their study the impact of different vibrotactile feedback routines on the perceived quality of a digital piano is assessed. Our methodology and analysis in Part 1 took a similar route. In this first section participants were presented with the instrument and advised to freely improvise while switching between two settings: α and β. Their task, for each pair of α and β, was to comparatively rate the two settings according to four quality metrics (Responsiveness, Naturalness, Temporal Control, General Preference) drawn from studies on subjective quality assessments of acoustic instruments (Saitis, Giordano, Fritz, & Scavone, 2012) and based on the qualities we hypothesized would be most relevant to the changing latency conditions. Once they had decided on the comparative ratings of the two settings they then moved onto the next pair.

Stimuli and conditions. Between α and β we changed both the latency condition and sample set. We deliberately wanted to mask the changing latency conditions to evaluate whether the latency conditions were perceivable by the participants when they were not instructed to focus on the amount of latency present. When starting the study participants were instructed simply to compare the different settings on the instrument according to the attributes and to try and not base their ratings on preference for a sample set alone. The fact that latency would be changing was not mentioned.

Experiment procedure. The instrument was self-contained, dealing with all the sensor and audio

FIGURE 2. Experimental set-up with instrument and accompanying laptop for changing settings.

processing via Bela, allowing participants to monitor the instrument directly via noise cancelling headphones. To switch between α and β, we used a separate laptop that hosted a graphical user interface built in Pure Data (https://puredata.info/), which communicates with the Bela board via UDP, allowing participants to switch between settings at will (see Figure 2). For each pair, the zero latency condition (A) was assigned to either α or β in a weighted random order while the other setting in the pair would always contain a latency condition (B, C, or D). Two different sample sets were also selected in a weighted random order for α and β. There were 12 such pairs, again presented in a weighted random order for each participant, ensuring that each sample set was in the zero latency position three times per participant. This meant that each participant rated each pair of latency conditions four times, each time with a different sample set assigned to each of the conditions. Participants were advised to take around 35 min to complete the evaluation of the 12 pairs.

Participants also input their ratings on the accompanying laptop via a graphical user interface that consisted of slider inputs for each attribute using a Continuous Category Rating scale (CCR), a rating widely used in subjective quality assessments of interfaces (recommendation ITU-T P.800). While rating the settings, participants were instructed to improvise freely with no restrictions on their chosen style. Participants moved the slider on the continuous scale to rate the relative merits of the two settings (see Figure 3). The scale had the following titles along its scale: α is much better than β, both α and β are equal, and β is much better than α.
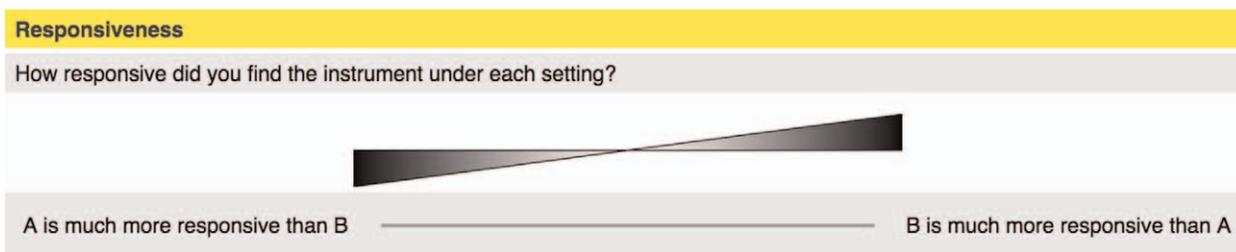
FIGURE 3. Continuous input slider for rating the settings in comparison to one another.

Part 2: Timing accuracy. In order to evaluate the impact of the latency conditions on the temporal performance of the participants, we used a synchronization task were they were instructed to play along with a metronome under each latency condition. A metronome at 120 bpm was played through the headphones. The participant was then instructed to tap along with the beat using a single tile, dividing the metronome beat into progressively smaller chunks: every crotchet (quarter note)—which is equivalent to the 120 bpm of the metronome—then every quaver (eighth note), then every semiquaver (sixteenth note). They performed each of these tapping exercises for at least four bars, paused, and then moved onto the next. They repeated the whole task three times for each latency condition and then moved onto the next condition. Each of the four latency conditions were presented in a weighted random order and the sample set remained the same across participants. Our methodology in this part of the study was derived from Fujii et al.'s study (2011) on synchronization of drum kit playing.

Part 3: Structured interview. To conclude the experiment, a structured interview lasting between 10 and 20 min was conducted. The interview was conducted in front of the instrument and demonstrations were encouraged from the participants. The following themes were discussed in each case: 1) general impression of the instrument, including the styles of playing that worked well and not; 2) techniques used to distinguish between α and β in Part 1, the free improvisation; 3) whether they noticed what was changing between settings, besides sample set; and 4) their experience of latency as an issue in musical performance.

Results

QUALITY JUDGMENTS

The differences in subjective judgments of instrument quality were evaluated by looking at the quality ratings from Part 1 of the study.

Statistics. In our analysis and in Figure 4 condition A (zero latency) is always α (zero on the *y*-axis) for legibility, although in the study it was randomly assigned to either α or β. For each *group* (professional percussionist, non-percussionist) we fitted separate linear mixed effect regression (LMER) models with fixed effects of *quality* (responsiveness, naturalness, temporal control, general preference) and *condition* (10 ms, 20 ms, 10 ms ± 3 ms), and random intercepts for each *participant*. The models were fitted using the lme4 (Bates, Mächler, Bolker, & Walker, 2015) package for R (R Core Team, 2017). We conducted a full factorial Type III ANOVA on each LMER model, with Satterthwaites's degrees of freedom approximation from the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017).

Group 1: Non-percussionists.. Figure 4a shows the median and IQR for all participants in this group (Figure 4c shows the mean and standard error). On average condition A (the zero latency condition) was rated more positively for all qualities than condition C and D, the 20 ms and 10 ms ± 3 ms conditions, respectively. We found a significant effect of condition, $F(2, 517) = 7.37$, $p < .001$, but no effect of quality. A Tukey post hoc analysis on each factor shows that the effect of condition is driven by a significant difference between 10 ms ± 3 ms and 10 ms, $Z = 3.42$, $p_{adj} < .01$, and 20 ms and 10 ms, $Z = 3.21$, $p_{adj} < .01$ (all *p* values were adjusted using the Benjamini and Hochberg false discovery rate correction; FDR = 5%).

Group 2: Professional percussionists.. Figure 4b shows the median and IQR for all participants in this group (Figure 4d shows the mean and standard error). For the professionals, we found significant effects of condition, $F(2, 470) = 4.82$, $p < .01$, and quality, $F(3, 470) = 4.98$, $p < .01$. A Tukey post hoc analysis on each factor shows that the effect of condition was driven by a significant difference between 10 ms ± 3 ms and 10 ms, $Z = 3.12$, $p_{adj} < .01$, and a significant difference between 20 ms and 10 ms, $Z = 2.54$, $p_{adj} < .05$ (all *p* values were
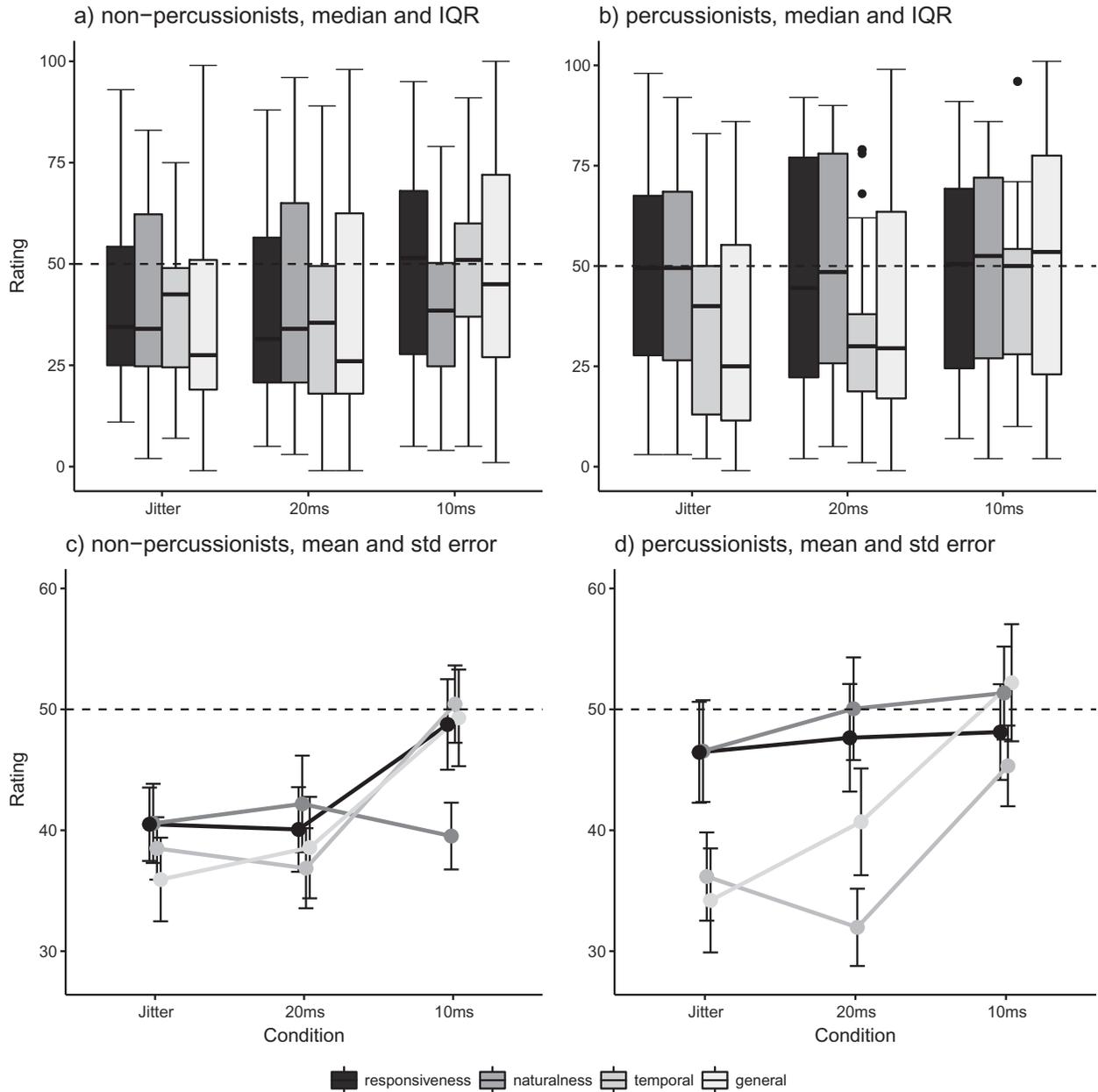
FIGURE 4. (a) and (b) show the median and IQR of all responses from both participant groups. (c) and (d) show the mean and standard error of all responses from both participant groups. 0 on the y axis corresponds to 'α is much better than β', 100 with 'β is much better than α', and 50 with 'both α and β are equal'. Note that in this representation 0 always means that condition A (zero latency) is preferred to the other latency condition it is being compared to.

adjusted using the Benjamini and Hochberg false discovery rate correction; FDR = 5%).

Influence of sample set. For both groups we tested to ensure that sample was not having an overriding effect on quality ratings (i.e., that participants were basing their ratings on sample set alone). When fitting the LMER models, we also included sample set as a fixed

effect and found no significant effect, so we were able to discount this as a factor.

TEMPORAL PERFORMANCE

In this paper our analysis of timing performance focuses only on task 2, playing with a metronome. For this analysis we compared the onset of the strike against the
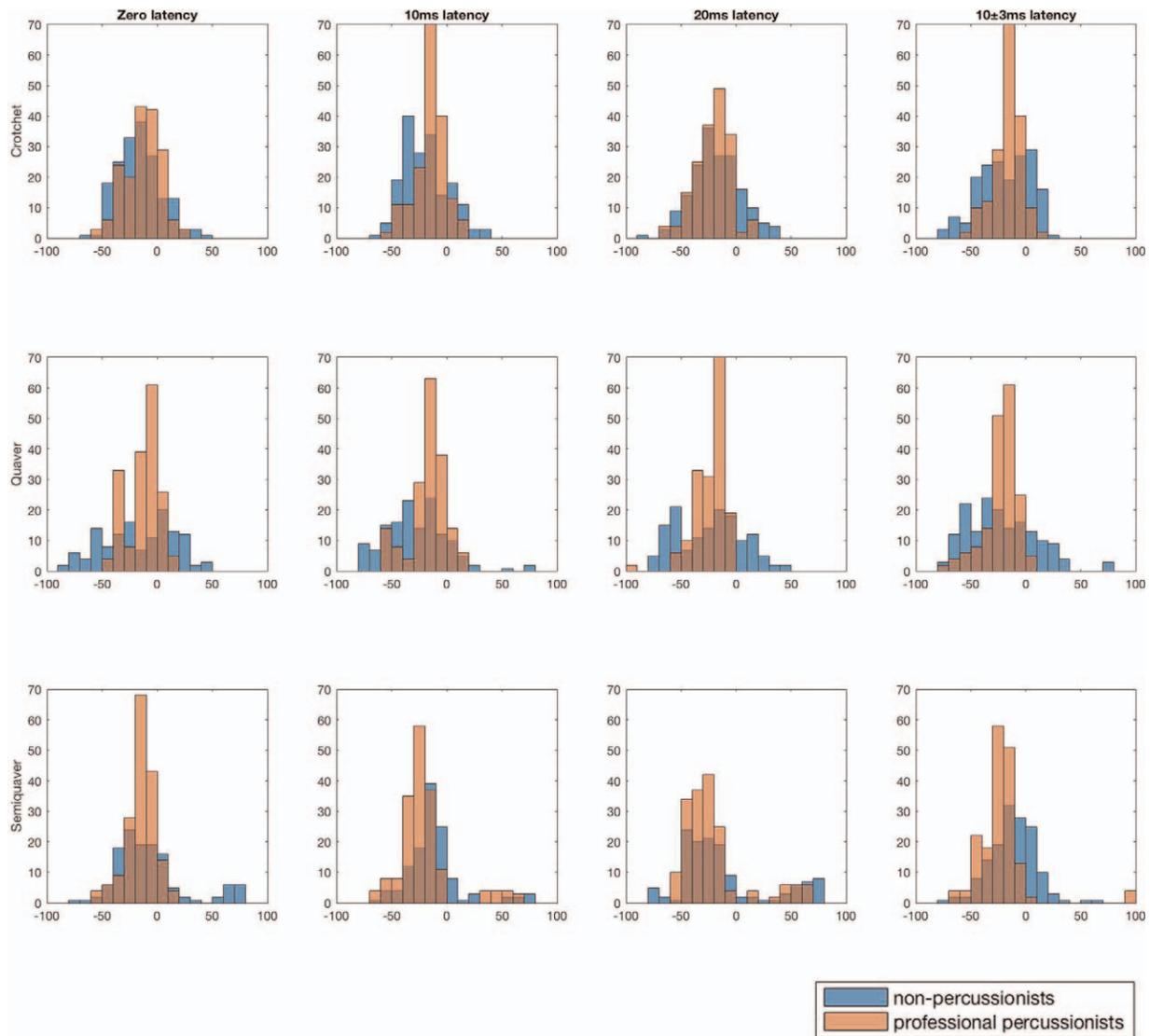
FIGURE 5. Distribution of strikes for both groups showing the spread of the timing of their strikes during the synchronization task. The medium gray bar components reflect overlap between non-percussionists and professional percussionists.

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 had added latency under certain conditions. 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.

Statistics. For the modeling we fitted an LMER model with fixed effects of *group* (non-percussionists, professional percussionists), *temporal division* (crotchet, quaver, semiquaver) and *condition* (10 ms,

20 ms, 10 ms $\pm$ 3 ms), and random intercepts for each *participant*. As with the quality judgment analysis, the significance of each fixed effect was tested using a full factorial Type III ANOVA on the LMER model, with Satterthwaites's degrees of freedom approximation.

Typical distribution. Figure 5 shows the typical distribution of strikes of both groups for each tempo measure and each latency condition. Figure 6 presents the median and interquartile range (IQR) for all latency conditions for both groups. For the NP group we excluded one participant from our analysis due to them having a mean synchronization error (MSE) of 30% greater than the group MSE giving 10 participants in
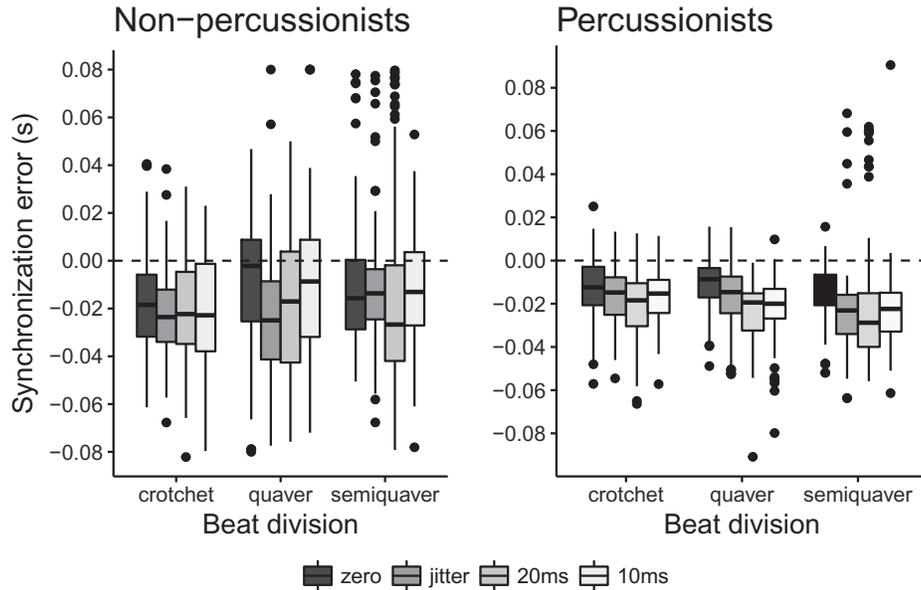
FIGURE 6. Median and IQRs of synchronization error for the first rhythmic task for all latency conditions for both groups.

this group. All 10 participants in the PP group had a MSE within this threshold.

Synchronization error. Figure 6 shows the median and IQR for all participants under each *latency condition* and *division*. We found a significant effect of *condition*, $F(3, 2289) = 5.88$, $p < .001$, and *division*, $F(2, 2289) = 3.73$, $p < .05$. A Tukey post hoc analysis showed that the effect of *condition* is driven by a significant difference between 10 ms and zero latency, $Z = 3.56$, $p_{adj} < .001$, and between 20 ms and zero latency, $Z = 3.73$, $p_{adj} < .001$. A smaller and marginally significant difference was seen between 10 ms $\pm$ 3 ms and zero latency ($p_{adj} = .08$). We also tested for interactions between each fixed effect (by fitting new models with interaction terms), and found a significant interaction between *group* and *division*, $F(2, 2288) = 12.52$, $p < .001$. This effect is shown in Figure 7, where synchronization error is negatively correlated with IOI for the NP group, and the opposite effect is observed for PP. A post hoc analysis of the interaction contrasts between all factors of group and condition was conducted using the phia package (De Rosario-Martinez, 2015) for R. This showed a significant interaction between *group* and all three division factors: *crotchet-quaver*, $\chi^2(1) = 4.38$, $p < .05$, *crotchet-semiquaver*, $\chi^2(1) = 24.98$, $p < .001$, and *quaver-semiquaver*, $\chi^2(1) = 7.82$, $p < .01$.

To assess the effect of *division* for each group, we refitted separate models for each group, with fixed effects of
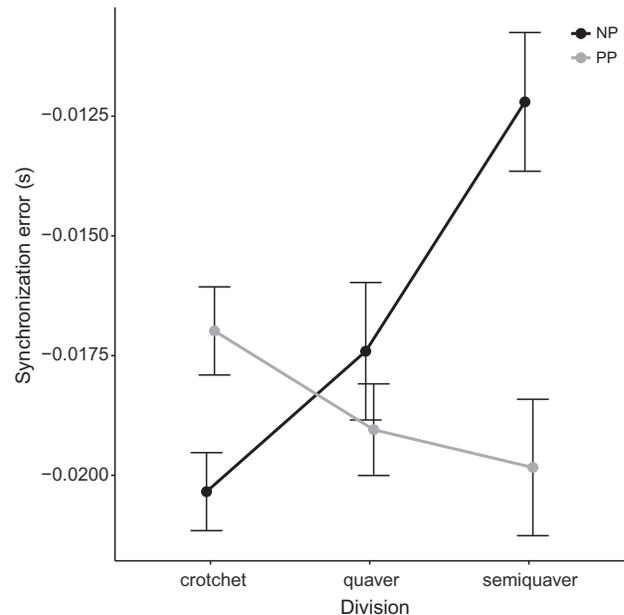
FIGURE 7. Interaction contrasts between division and group for synchronization error.

condition and *division*, and random intercepts for each *participant*. In the case of the non-percussionists this showed a significant effect of *division*, $F(2, 1483) = 17.31$, $p < .001$. A Tukey post hoc analysis showed that the effect of *division* is driven by a significant difference between *crotchet-semiquaver*, $Z = 5.94$, $p_{adj} < .001$, and

between *quaver-semiquaver,* $Z = 3.70$, $p_{adj} < .001$. In the case of the professional percussionists this showed a significant effect of *condition,* $F(3, 806) = 8.82$, $p < .001$. A Tukey post hoc analysis showed that the effect of *division* is, as expected from the interactions above, driven by a significant difference between 10 ms and zero latency, $Z = 3.04$, $p_{adj} < .01$, 20 ms and zero latency, $Z = 4.56$, $p_{adj} < .001$, and 10 ms $\pm$ 3 ms and zero latency, $Z = 4.30$, $p_{adj} < .001$.

Variability of synchronization error. To evaluate the variability of timing accuracy we refitted the above mentioned model but with standard deviation of the synchronization error as the dependent variable. We observed heteroskedasticity in the residuals of the fitted model, which we rectified using a log transform of the dependent variable (sderror). We found a significant effect of *group,* $F(1, 20) = 16.57$, $p < .001$, *division,* $F(2, 220) = 5.46$, $p < .01$, *condition,* $F(3, 220) = 4.24$, $p < .01$, and an interaction between *condition* and *division,* $F(6, 220) = 3.50$, $p < .01$. The mean standard deviation between *groups* (across all *conditions* and *divisions*) is 0.02 for non-percussionists and 0.01 for percussionists: this is a difference of over almost 50%.

Upon testing for interaction contrasts between *condition* and *division* we found the significant interactions are between 20 ms and each of the crotchet-semiquaver, $\chi^2(1) = 13.16$, $p < .01$, and *quaver-semiquaver,* $\chi^2(1) = 8.44$, $p < .05$, conditions. This can be seen in Figure 8. We noted a medium but nonsignificant positive correlation between error and standard deviation of error (i.e., as error decreases, so does its variation).

Interviews

The structured interviews conducted at the end of the study were annotated and then coded using a thematic analysis framework (Braun & Clarke, 2006). Our coding strategy aimed to identify the major themes that related to latency perception and judgments of instrument quality. Other themes that came from these interviews related to style, the constraints of the instrument and the evolution of gesture over the duration of the study have been presented elsewhere (Jack, Stockman, & McPherson, 2017).

NON-PERCUSSIONISTS

Awareness of latency. Latency perception was the first theme we investigated; whether the settings with latency were perceived as having a delay or not. Only 3 out of the 11 participants stated that there was latency or delay changing between the settings. This suggests that either the amounts of latency were small enough to not be
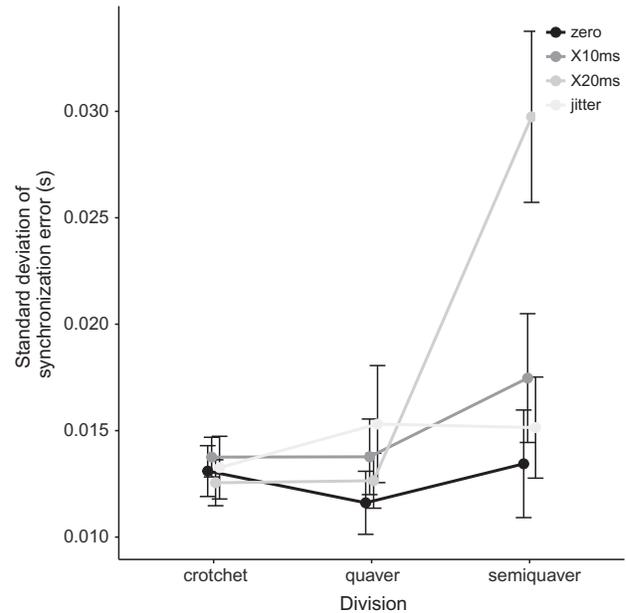

FIGURE 8. Interaction contrasts between condition and division for the standard deviation of synchronization error.

perceived as a delay for 8 of the 11 participants or that the changing sample sets masked the changing latency conditions. When asked what was changing between settings aside from the sample set participants generally reported a changing responsiveness and level of dynamic control: they described shifting triggering thresholds at times, that the instrument was catching less of their strikes under certain settings, or that the dynamic range of instrument was expanding and contracting, factors that were not in fact changing. In the quality ratings from Part 1 of the study we have seen the zero latency condition receiving more positive ratings than the 10 ms $\pm$ 3 ms and 20 ms latency condition for the attribute Temporal Control. This suggests that a disruption to the temporal behavior of the instrument was identified even if its cause was not established as delayed auditory feedback. Some participants also acknowledged that under certain conditions they were struggling to maintain timing although, again, they did not specifically identify that a delay or latency was the cause:

"...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)

". . . I was playing very fast passages and seeing if it captures all the notes. In some of the settings it wasn't tracking well but in others it was." (Participant 8)

These quotes point towards the complexity of the "response" of the instrument: this term does not seem to have been reduced to how fast the instrument responded, but rather is about how much of the participant's playing was translated into sound by the instrument: judgments seem to be based on how participants felt the instrument was reflecting the energy they put in. In addition to the above reports, there were also additional multimodal effects of the latency conditions reported, where the perceived effort required to play a note increased with latency.

Reported effects of latency during the study. Four of the 11 participants reported that under certain latency conditions they felt they needed to strike the instrument with more force to get the instrument to respond in the way they wanted.

"I also noticed that I had to put more energy into one or other of the pairs to get a sound from the instrument." (Participant 9)

For these four participants we analyzed the variation in striking velocity across latency conditions to test if their reports of increased force of strike was influenced by latency condition. We found that for these participants there was indeed an increased mean velocity of strike for 20 ms and 10 ± 3 ms latency in comparison to the zero latency condition (Jack, Stockman, & McPherson, 2016).

PROFESSIONAL PERCUSSIONISTS

Awareness of latency. In general the professional percussionists were more aware of the latency conditions than the non-percussionists, with 9 out of 10 mentioning it as the changing factor between settings.

"I felt like some of them were a bit more 'on top' [. . .]sometimes you felt like it wasn't instantaneous and you're not connected to it." (Participant 4)

"The latency also changed as well and they [the sample set] weren't necessarily related." (Participant 1)

". . . Sometimes there was a bit of a delay, sometimes the note was behind the strike." (Participant 3)

They were also more conscious of latency as an issue that faces digital musical instruments. In some cases this came from their experience of using digital samplers in live performance or from experiences of home recording with a backing track.

"I have a set of TD Roland drums, and they have latency, it's very slight but I definitely notice it, it's more than it would be from an acoustic kit definitely." (Participant 5)

Many of the participants also explicitly mentioned latency as a negative factor in an instrument's design that impedes their performance.

". . . With percussionists, we're so used to, you hit it and, bang, it's there. So any kind of delay is a bit disconcerting." (Participant 7)

"When it sits on top it's a lot more enjoyable to play. I know when that happens you tend to forget that you're playing something, and you tend to explore, you make music then, rather than trying to work out the instrument." (Participant 4)

Ability to adjust to action-sound latency. Participants also spoke of their ability to adjust to the changing latency conditions naturally and without too much active thought when they were freely improvising without a metronome.

"Because I've got experience of adjusting I was able to adjust to what I was hearing. I do that naturally. When playing acoustic instrument you listen to what's coming out and you adjust to it. It's always a tiny little difference, you can adjust naturally. You deal with it. You can get by." (Participant 3)

"We do have experience with working with delay and trying to think about that. You need to compensate so that you don't sound late. You don't really think about it much normally, it's too much if you think about it, has to be by feel." (Participant 7)

"I was adjusting very quickly. If I was doing them all with a click track I think the response I gave would be different as I would actually feel myself trying to adjust to where the beat was when there was latency, like this I just did it without thinking." (Participant 5)

Experience with acoustic instruments. One of the main differences between the two groups was their awareness and experience with latency. From the PP group there were many comparisons made between latency in a digital instrument and the timing adjustments that orchestral percussionists have to do as they

switch their position in the orchestra or switch the acoustic instrument they're playing. Talk of "sitting behind the beat," "sitting in front of the beat," and "sitting right on the beat" described how the percussionists conceptualize the microadjustments they make to their timing in order to ensure that the conductor (and audience) hears them as in time with the rest of the ensemble. When asked how they manage to adjust their playing like this, most stated that they had no idea how they actually did it, it was something that they had learned from being told by a conductor or other performers that they were coming in early or late and at this point in their careers it was just a necessary part of their role that they were able to do without thinking. They mentioned that the dress rehearsal before a concert was the most important in terms of making this adjustment as their timing needs to be tuned to their position in the ensemble and the acoustics of the room. Latencies within the range of 10 ms to 40 ms are common in ensemble playing due to the distance between players (Chafe, Cáceres, & Gurevich, 2010). The importance of playing in time is heightened by the rhythmical importance of the percussion section and the impulsive style of playing.

> "If you're sitting at the back of the orchestra the physical sound getting to the front takes longer as you're so far back. And for certain instruments this can take even longer. A lot of the time you have to play a little bit ahead or behind the beat to make sure it fits with everything else." (Participant 3)

Percussionists must also adjust to the mechanical action of the instrument they are playing. Professional percussionists are multi-instrumentalists that are expected to master and be able to switch between many different instruments in a matter of seconds. This brings with it the ability to switch playing techniques quickly and to adjust playing style to the specific action of an instrument—what the percussionists referred to as an instrument speaking *early* or *late*. Tamborine was given as an example of an instrument that sounds late, as was tubular bells and timpani. Examples of instruments that speak early included triangle and other metallic instruments played with hard beaters. The notion of how an instrument speaks seems to be related to frequency range of the instrument but also to surface hardness, the action of the instrument (triangle versus church bell for example), and striking type (hard versus soft beaters, played with the hands or not), although conflicting examples were given by different percussionists.

> "In the case of this instrument [the instrument used in this study] it's to do with the fact that it's hard.

I know that hard surfaces sound immediately, were as floppy surfaces sound later, like timpani. I guess that's just sort of Pavlovian – it's hard, it's going to sound quickly." (Participant 7)

Discussion

QUALITY JUDGMENTS

In terms of quality judgments, both groups seem to be generally in agreement. The results from Part 1 suggest that latency of 20 ms and 10 ms $\pm$ 3 ms can degrade the perceived quality of an instrument in terms of temporal control and general preference, even when the amount of latency is too small to be perceived as a delay by the performer. This is in agreement with findings from Kaaresoja et al. (2011) when evaluating the impact of audiotactile latency on user interaction with touchscreens. The fact that condition D (10 ms $\pm$ 3 ms latency) was rated in a similarly negative manner as condition C (20 ms latency) in relation to the zero latency condition, but that condition B (10 ms latency) did not receive similar negative ratings, highlights the importance of stable as well as low latency. This points to an agreement with Wessel and Wright's (2002) recommendations of 10 ms latency with 1 ms of jitter as a goal for digital musical instruments.

None of the participants in this experiment performed with a mean degree of accuracy in Part 2 that was better than the jitter amount ($\pm$ 3 ms), 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. These findings, alongside previous work (Repp & Su, 2013), suggest that the amount of acceptable latency and jitter does not correspond directly to the limits of sensorimotor accuracy possible by the player.

LATENCY PERCEPTION

In the non-percussionist group, 3 of the 11 participants identified latency, or delay, as the changing factor between settings. For the other 8 participants the difference between settings was reported as a changing triggering threshold or dynamic range, both of which remained identical throughout the study.

In the professional percussionist group, 9 of the 10 participants reported latency or delay as the changing factor between settings. It seems that this group was much more aware of latency and its causes from their experience as orchestral players, and were generally

better at discussing it, as can be seen in the examples presented from the structured interview. Making micro-adjustments to the timing of their performance in relation to an ensemble or to their instrument is a common part of professional percussionist's role as a musician. This may explain the difference between the groups, alongside the superior synchronization ability (Cameron & Grahn, 2014) and timing acuity (Ehrlé & Samson, 2005) of percussionists, whether from their extensive training on a rhythm-based instrument or natural ability.

TIMING ACCURACY

Effect of latency and beat division on mean synchronization error. We found significant differences between the effects of latency and beat division on the timing ability of both groups in this study. There was a significant effect of condition for both groups and significant differences (i.e., interactions) between group and beat division. For the NP group we found that increasing division of the beat was affecting the accuracy of their playing, whereas latency condition showed no significant effect: we observed the MSE and variation of MSE increasing as the beat division increased. This suggests that the error in their temporal performance increased as they were required to strike faster. We found no significant effect of latency condition on timing accuracy for this group.

The opposite seems to be the case for the PP group: timing accuracy was significantly affected by latency condition and not by beat division in most cases. For this group the zero latency condition had a significantly lower MSE in comparison to the other three latency conditions across beat divisions. The standard deviation of MSE under each latency condition did not differ significantly for each beat division aside from for 20 ms latency at the smallest beat division (semiquaver), as can be seen in Figure 8. This suggests that for the larger divisions, this group did not find latency disruptive to timing accuracy: when they were required to play at a speed above a certain threshold (IOI of 125 ms) the latency became detrimental to their performance. From our findings this was only the case with the 20 ms latency condition, which would equate to 16% of the IOI of 125 ms when playing semiquavers at 120 bpm, well above the variation in timing accuracy from professional percussionists that has been previously reported (Dahl & Bresin, 2001).

Mean synchronization error. Generally, we observed a higher degree of variation in the MSE of the NP group in comparison to the PP group, as can be seen in Figure 6. This agrees with the findings of Manning and Schutz (2016) that participants with high levels of rhythm-based training (particularly percussionists) show superior timing abilities (MSE and variability of MSE) and temporal acuity in comparison to other musicians and nonmusicians.

The group means of the MSE for the zero latency condition and all metronome ranged from -17 to -7 ms for NP and from -15 to -12 ms for PP. The mean standard deviation ranged from 20 to 33 ms for NP and 8 to 12 ms for PP. The MSE and SD for the NP group were larger than that found by Fujii et al. (2011) in their study with highly trained percussionists where a mean synchronization error of -13 to 10 ms was achieved for a metronome with standard deviations of 10 to 16 ms, whereas the MSE and SD of MSE for the PP are roughly aligned with these findings. The MSEs of both groups were smaller than those reported in previous tapping studies with nonmusicians in which MSE was usually around -20 to -80 ms, while for the NP group they were roughly equivalent to the performance of amateur musicians: 10 to 30 ms (Aschersleben, 2002; Repp & Doggett, 2007). The values of MSE for the PP group in this study were smaller when compared with the finger tapping study of Gerard and Rosenfeld (1995), who found an MSE of -25 ms in professional percussionists.

A further analysis step that falls beyond the scope of this paper would be to investigate systematic synchronization errors in the performances of each of our groups. In this respect part of the synchronization error that we observed could be attributed to systematic and reoccurring time deviances (Hellmer & Madison, 2015).

Adaptation and negative mean asynchrony. In the PP group we also saw an increase in negative mean asynchrony during the crotchet and quaver beat divisions that partially reflected the amount of latency being added to the instrument. We observed an increase in MSE of approximately 5 ms and 10 ms for the 10 ms and 20 ms latency condition, respectively. This can be seen quite clearly in Figure 6. It seems that there was a degree of compensation in relation to the latency condition but it was not an anticipation of the full latency amount (i.e., moving a strike 20 ms earlier when 20 ms of latency was present to bring the auditory feedback in line with the metronome). Anticipation of strike to match sound has been observed by others when introducing larger amounts of delay to auditory feedback (Aschersleben & Prinz, 1997; Dahl & Bresin, 2001; Stenneken et al., 2006). These anticipation effects were not observed with the NP group for any latency condition. We could also hypothesize that as a result of our experimental method, where the amount of latency was changed regularly between conditions, the adaptation as

reported in other studies did not have enough time to fully occur (Vroomen & Keetels, 2010).

RHYTHMIC TRAINING AND LATENCY PERCEPTION

Regardless of training, participants generally agreed on judgments of perceived instrument quality. In the case of the non-percussionists, even if the latency was not perceived as a delay, its effect on the fluency of interaction with the instrument was recognized by the participants. Timing accuracy in the non-percussionist group was not significantly affected by the latency condition, yet this group rated 20 ms and 10 ms $\pm$ 3 ms latency negatively in comparison to the zero latency condition. From the structured interviews they were reports that certain conditions felt "under the fingers," whereas with others the connection between action and sound was not as clear. This highlights the subtlety of the effects of latency and the specific demands of percussion instruments where sound is a result of direct unmediated interaction.

In general the PP group was much more aware of latency and able to identify it as the changing factor between settings, and talk explicitly about adjusting for latency. This is perhaps due to their extensive rhythmical training and expertise in switching between instruments with different actions. Some of the percussionists spoke about the changing latency conditions as the changing action of the instrument: whether the instrument would sound "late" or their playing would be "right on top" of the beat, allowing them to forget the instrument and concentrate on making music. This connects with ideas of instrument *transparency*: Nijs, Lesaffre, and Leman (2009) propose musical instruments as mediators between gesture and sound output. Transparency in this mediation is the point where the performer doesn't need to focus attention on the individual operations of manipulating an instrument, instead focusing on higher-level musical intentions. Latency in this interaction might be understood as a barrier to transparency.

Latency perception and the effects of latency vary widely dependent on the nature of the musical task, style of playing, instrument, and individual experience of the performer. From our study we cannot determine the acceptable amount of latency that digital musical instruments should aim for in general, and also, as our sample size is relatively small there needs to be a degree of caution in interpreting our results, as statistical power is necessarily limited by the amount of participants in each group. Our aim with this study, rather, is to highlight the effects of small amounts of latency on the perceived quality of an instrument, an effect that

we propose as similar to the degradation of feelings of presence in VR situations: latency as "a cause for reduction of suspension of disbelief" (Allison, Harris, Jenkin, Jasiobedzka, & Zacher, 2001). In the case of digital musical instruments the notion "presence" is perhaps best equated to the erogtic aspects of an instrument: how energy is maintained in the digital system in the translation of action to sound (Luciani et al., 2009). Latency stands a barrier to the fluency of interaction that digital musical instruments should aim to foster.

Concluding Remarks

We have presented a study that investigated the impact of latency and jitter on the temporal accuracy of performance and judgments of instrument quality for two groups of participants; professional percussionists and non-percussionists (with varying amounts of musical experience). The studies involved quality assessments of a novel percussive instrument with variable latency (zero, 10 ms, 10 ms $\pm$ 3 ms, 20 ms), temporal accuracy tests and structured interviews.

In terms of judgments of instrument quality, we found that both groups showed a preference for 0 ms in comparison to 10 ms $\pm$ 3 ms and 20 ms latency. Importantly, the 0 ms and 10 ms latency conditions show no significant difference in rating for either group. This suggests that a stable latency of 10 ms is acceptable to performers of a DMI where 20 ms is not. The 10 ms $\pm$ 3 ms latency condition was rated in a similarly negative manner to 20 ms when compared to the zero latency condition, suggesting that the addition of a random jitter of $\pm$3 ms is enough to negatively effect the perceived quality of an instrument. Our results support the recommendation put forward by Wessel and Wright (2002) that DMI designers should aim for a latency of 10 ms or below with a jitter of 1 ms or less. However, our findings cannot tell us exactly what the minimum threshold of acceptable latency is, except that it must be somewhere between 10 ms and 20 ms.

Ability to perceive latency varied between groups, as did the impact on temporal performance. Generally professional percussionists were more aware of the latency conditions and better able to adjust for them in their playing, although this ability decreased as the temporal demands of the task increased. We have seen that latency negatively affects judgments of instrument quality even when the amount of latency is not detectable as a delay and has no impact on timing performance. Latency can degrade the illusion of action translating to sound, a factor that is central to expressive and skilled

control of digital musical instruments. In this study we have demonstrated the effects of latency on two different groups of musicians and found marked differences between each group in terms of disruption to timing accuracy, and the ability to identify latency. Both groups were in agreement as to the impact of latency on the quality of the instrument in question. This suggests that the influence of latency on the perceived quality of a digital system does not hinge on the temporal acuity of the user, rather, it is something that can degrade the fluency of the interaction regardless of level of skill.

Author Note

Correspondence concerning this article should be addressed to Robert Jack, 110C Deptford High Street, London, United Kingdom, SE8 4NS. E-mail: r.h.jack@qmul.ac.uk

References

ADELSTEIN, B. D., BEGAULT, D. R., ANDERSON, M. R., WENZEL, E. M., & FIELD, M. (2003). Sensitivity to haptic-audio asynchrony. In S. Oviatt (Ed.), *Proceedings of the 5th International Conference on Multimodal Interfaces* (pp. 73-76). Vancouver, Canada: ACM.

ALLISON, R. S., HARRIS, L. R., JENKIN, M., JASIOBEDZKA, U., & ZACHER, J. E. (2001). Tolerance of temporal delay in virtual environments. In H. Takemura & K. Kiyokawa (Eds.), *Proceedings of Institute of Electrical and Electronics Engineers (IEEE) Virtual reality* (pp. 247–254). Yokohama, Japan: IEEE.

ANNETT, M., NG, A., DIETZ, P., BISCHOF, W. F., & GUPTA, A. (2014). How low should we go? Understanding the perception of latency while inking. In P. Kry & A. Bunt (Eds.), *Proceedings of Graphics Interface* (pp. 167–174). Montreal, Canada: ACM.

ASCHERSLEBEN, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition*, *48*, 66–79.

ASCHERSLEBEN, G., & PRINZ, W. (1997). Delayed auditory feedback in synchronization. *Journal of Motor Behavior*, *29*, 35–46.

ASKENFELT, A., & JANSSON, E. V. (1988). From touch to string vibrations - The initial course of the piano tone. *Department for Speech Music and Hearing, Quarterly Progress and Status Report*, *29*, 31–109.

BATES, D., MÄCHLER, M., BOLKER, B., & WALKER, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

BLACK, J. W. (1951). The effect of delayed side-tone upon vocal rate and intensity. *Journal of Speech and Hearing Disorders*, *16*, 56–60.

BRAUN, V., & CLARKE, V. (2006). Using thematic analysis in psychology. *Qualitative research in Psychology*, *3*, 77–101.

CALVERT, G., SPENCE, C., & STEIN, B. E. (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.

CAMERON, D. J., & GRAHN, J. A. (2014). Enhanced timing abilities in percussionists generalize to rhythms without a musical beat. *Frontiers in Human Neuroscience*, *8*, 1003.

CHAFE, C., CÁCERES, J.-P., & GUREVICH, M. (2010). Effect of temporal separation on synchronization in rhythmic performance. *Perception*, *39*, 982–992.

CHAFE, C., & GUREVICH, M. (2004). Network time delay and ensemble accuracy: Effects of latency, asymmetry. In B. McQuaide (Ed.), *Proceedings Audio Engineering Society (AES) Convention 117*. San Francisco, USA: AES.

DAHL, S. (2000). The playing of an accent–preliminary observations from temporal and kinematic analysis of percussionists. *Journal of New Music Research*, *29*(3), 225–233.

DAHL, S., & BRESIN, R. (2001). Is the player more influenced by the auditory than the tactile feedback from the instrument? In M. Fernstr (Ed.), *Proceedings of Digital Audio Effects (DAFX-01)* (pp. 194–197). Limerick, Ireland: DAFX.

DE ROSARIO-MARTINEZ, H. (2015). phia: Post-hoc interaction analysis [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=phia (R package version 0.2-1)

DEBER, J., JOTA, R., FORLINES, C., & WIGDOR, D. (2015). How much faster is fast enough? In B. Begole & J. Kim (Eds.), *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1827–1836). Seoul, Republic of Korea: ACM

EHRLÉ, N., & SAMSON, S. (2005). Auditory discrimination of anisochrony: Influence of the tempo and musical backgrounds of listeners. *Brain and Cognition*, *58*, 133–147.

ESSL, G., & O'MODHRAIN, S. (2006). An enactive approach to the design of new tangible musical instruments. *Organised Sound*, *11*, 285–296.

FELS, S. (2004). Designing for intimacy: Creating new interfaces for musical expression. *Proceedings of the Institute of Electrical and Electronics Engineers (IEEE)*, *92*, 672–685.

FINNEY, S. A. (1997). Auditory feedback and musical keyboard performance. *Music Perception*, *15*, 153–174.

FONTANA, F., JÄRVELÄINEN, H., PAPETTI, S., AVANZINI, F., KLAUER, G., MALAVOLTA, L., ET AL. (2015). Rendering and subjective evaluation of real vs. synthetic vibrotactile cues on a digital piano keyboard. In J. Timoney (Ed.), *Proceedings of the Sound and Music Computing Conference*. Maynooth, Ireland: SMC.

Fujii, S., Hirashima, M., Kudo, K., Ohtsuki, T., Nakamura, Y., & Oda, S. (2011). Synchronization error of drum kit playing with a metronome at different tempi by professional drummers. *Music Perception*, *28*, 491–503.

Fujisaki, W., & Nishida, S. (2009). Audio-tactile superiority over visuo-tactile and audio-visual combinations in the temporal resolution of synchrony perception. *Experimental Brain Research*, *198*(2-3), 245–259.

Gates, A., Bradshaw, J. L., & Nettleton, N. C. (1974). Effect of different delayed auditory feedback intervals on a music performance task. *Perception and Psychophysics*, *15*, 21–25.

Gerard, C., & Rosenfeld, M. (1995). Musical expertise and temporal regulation. *Annee Psychologique*, *95*, 571–591.

Gescheider, G. A. (1966). Resolving of successive clicks by the ears and skin. *Journal of Experimental Psychology, 71*, 378–381.

Goebl, W., & Palmer, C. (2008). Tactile feedback and timing accuracy in piano performance. *Experimental Brain Research*, *186*, 471–479.

Hellmer, K., & Madison, G. (2015). Quantifying microtiming patterning and variability in drum kit recordings. *Music Perception*, *33*, 147–162.

Howell, P. (2001). A model of timing interference to speech control in normal and altered listening conditions applied to the treatment of stuttering. In B. Maaesesen, W. Hulstijn, R. Kent, H. Peters, & van Lieshout, P. H. (Eds.), *Speech motor control in normal and disordered speech* (pp. 291–294). Uttgeverij: Nijmegen.

Jack, R. H., Stockman, T., & McPherson, A. (2016). Effect of latency on performer interaction and subjective quality assessment of a digital musical instrument. In J. Fagerlönn (Ed.), *Proceedings of the Audio Mostly* (pp. 116–123). Norrköping, Sweden: ACM.

Jack, R. H., Stockman, T., & McPherson, A. (2017). Rich gesture, reduced control: The influence of constrained mappings on performance technique. In M. Gillies (Ed.), *Proceedings of 4th International Conference on Movement Computing*. London, United Kingdom: ACM.

Jota, R., Ng, A., Dietz, P., & Wigdor, D. (2013). How fast is fast enough? A study of the effects of latency in direct-touch pointing tasks. In S. Brewster & S. Bødker (Eds.), *Proceedings of the Special Interest Group on Computer–Human Interaction (SIGCHI) Conference on Human Factors in Computing Systems* (pp. 2291–2300). Paris, France: ACM.

Kaaresoja, T., Anttila, E., & Hoggan, E. (2011). The effect of tactile feedback latency in touchscreen interaction. In C. Basdogan (Ed.), *IEEE World Haptics Conference* (pp. 65–70). Istanbul, Turkey: IEEE.

Kaaresoja, T., & Brewster, S. (2010). Feedback is . . . late: Measuring multimodal delays in mobile device touchscreen interaction. In W. Gao, C. Lee, & J. Yang (Eds.), *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*. Beijing, China: ACM.

Kaaresoja, T., Brewster, S., & Lantz, V. (2014). Towards the temporally perfect virtual button: Touch-feedback simultaneity and perceived quality in mobile touchscreen press interactions. *ACM Transactions on Applied Perception*, *11*(2), Article 9.

Kietzman, M. L., & Sutton, S. (1968). The interpretation of two-pulse measures of temporal resolution in vision. *Vision Research*, *8*, 287–302.

Kilchenmann, L., & Senn, O. (2011). "Play in time, but don't play time": Analyzing timing profiles in drum performances. In A. Williamon, D. Edwards, & L. Bartel (Eds.), *Proceedings of the International Symposium on Performance Science* (pp. 593–598). Toronto, Canada: ISPS.

Krause, V., Pollok, B., & Schnitzler, A. (2010). Perception in action: The impact of sensory information on sensorimotor synchronization in musicians and non-musicians. *Acta Psychologica*, *133*, 28–37.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Lago, N., & Kon, F. (2004). The quest for low latency. In M. Gurevich (Ed.) *Proceedings of the International Computer Music Conference* (pp. 33–36). Miami, USA: ICMC

Lester, M., & Boley, J. (2007). The effects of latency on live sound monitoring. *Audio Engineering Society Convention, 123*, 1–20.

Levitin, D. J., MacLean, K., Mathews, M., Chu, L., Jensen, E., & Dubois, D. M. (2000). The perception of cross-modal simultaneity (or "the Greenwich observatory problem" revisited). In D. Dubois (Ed.), *AIP Conference Proceedings* (pp. 323–329). Liege, Belgium: AIP.

Lim, V. K., Bradshaw, J. L., Nicholls, M. E., & Altenmüller, E. (2003). Perceptual differences in sequential stimuli across patients with musician's and writer's cramp. *Movement Disorders*, 18, 1286–1293.

Luciani, A., Florens, J.-L., Couroussé, D., & Castet, J. (2009). Ergotic sounds: A new way to improve playability, believability and presence of virtual musical instruments. *Journal of New Music Research*, *38*, 309–323.

MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. Berlin, Germany: Springer-Verlag.

MacKenzie, I. S., & Ware, C. (1993). Lag as a determinant of human performance in interactive systems. In B. Arnold, G. van der Veer, & T. White (Eds.), *Proceedings of the Interact'93 and Chi'93 Conference on Human Factors in Computing Systems* (pp. 488–493). Amsterdam, Netherlands: ACM

Magnusson, T., & Mendieta, E. H. (2007). The acoustic, the digital and the body: A survey on musical instruments. In C. Parkinson & E. Singer (Eds.), *Proceedings of the 7th International Conference on New Interfaces for Musical Expression* (pp. 94–99). New York, USA: NIME.

Mäki-Patola, T., & Hämäläinen, P. (2004). Latency tolerance for gesture controlled continuous sound instrument without tactile feedback. In M. Gurevich (Ed.) *Proceedings of the International Computer Music Conference* (pp. 1-5). Miami, USA: ICMC

Manning, F. C., & Schutz, M. (2016). Trained to keep a beat: Movement-related enhancements to timing perception in percussionists and non-percussionists. *Psychological Research*, 80, 532–542.

McPherson, A., Jack, R. H., & Moro, G. (2016). Action-sound latency: Are our tools fast enough? In S. Wilkie & E. Benetos (Eds.), *Proceedings of the International Conference on New Interfaces for Musical Expression*. Brisbane, Australia: NIME.

McPherson, A., & Zappi, V. (2015). An environment for submillisecond-latency audio and sensor processing on beaglebone black. In B. Kostek & U. Zanghieri (Eds.), *Audio Engineering Society Convention 138*. Warsaw, Poland: AES.

Medeiros, C. B., & Wanderley, M. M. (2014). A comprehensive review of sensors and instrumentation methods in devices for musical expression. *Sensors*, 14, 13556–13591.

Meehan, M., Razzaque, S., Whitton, M. C., & Brooks, F. P. (2003). Effect of latency on presence in stressful virtual environments. In J. Chen, B. Froehlich, B. Loftin, U. Neumann, & H. Takemura (Eds.), *Proceedings Virtual Reality* (pp. 141–148). Los Angeles, USA: IEEE.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. i. temporal factors. *Journal of Neuroscience*, 7, 3215–3229.

Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Adaptation to audiotactile asynchrony. *Neuroscience Letters*, 413, 72–76.

Ng, A., Lepinski, J., Wigdor, D., Sanders, S., & Dietz, P. (2012). Designing for low-latency direct-touch input. In H. Benko & C. Latulipe (Eds.), *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology* (pp. 453–464). Cambridge MA, USA: ACM.

Nijs, L., Lesaffre, M., & Leman, M. (2009). The musical instrument as a natural extension of the musician. In M. Castellango & H. Genevois (Eds.), *Proceedings of the 5th Conference of Interdisciplinary Musicology* (pp. 132–133). Paris, France: LAM-Institut jean Le Rond d'Alembert.

Occelli, V., Spence, C., & Zampini, M. (2011). Audiotactile interactions in temporal perception. *Psychonomic Bulletin and Review*, 18, 429–454.

O'Modhrain, S. (2011). A framework for the evaluation of digital musical instruments. *Computer Music Journal*, 35, 28–42.

Pfordresher, P., & Palmer, C. (2002). Effects of delayed auditory feedback on timing of music performance. *Psychological Research*, 66, 71–79.

Pfordresher, P. Q. (2003). Auditory feedback in music performance: Evidence for a dissociation of sequencing and timing. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 949–964.

Pfordresher, P. Q. (2005). Auditory feedback in music performance: The role of melodic structure and musical skill. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 1331–1345.

Pfordresher, P. Q. (2006). Coordination of perception and action in music performance. *Advances in Cognitive Psychology*, 2, 183–198.

Pfordresher, P. Q. (2008). Auditory feedback in music performance: The role of transition-based similarity. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 708–725.

Pfordresher, P. Q., & Dalla Bella, S. (2011). Delayed auditory feedback and movement. *Journal of Experimental Psychology: Human Perception and Performance*, 37(2), 566–579.

Pfordresher, P. Q., & Palmer, C. (2006). Effects of hearing the past, present, or future during music performance. *Attention, Perception, and Psychophysics*, 68, 362–376.

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception*, 24, 37–48.

Repp, B. H. (2000). Compensation for subliminal timing perturbations in perceptual-motor synchronization. *Psychological Research*, 63, 106–128.

Repp, B. H., & Doggett, R. (2007). *Tapping to a very slow beat: A comparison of musicians and nonmusicians, Music Perception*, 24, 367–376.

Repp, B. H., & Su, Y. H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin and Review*, 20, 403–452.

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

Schultz, B. G., & van Vugt, F. T. (2016). Tap arduino: An arduino microcontroller for low-latency auditory feedback in sensorimotor synchronization experiments. *Behavior Research Methods*, 48, 1591–1607.

SHEFFIELD, E., & GUREVICH, M. (2015). Distributed mechanical actuation of percussion instruments. In E. Berdahl (Ed.), *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 11–15). Louisiana, USA: NIME.

SPENCE, C., & PARISE, C. (2010). Prior-entry: A review. *Consciousness and Cognition, 19*, 364–379.

STENNEKEN, P., PRINZ, W., COLE, J., PAILLARD, J., & ASCHERSLEBEN, G. (2006). The effect of sensory feedback on the timing of movements: evidence from deafferented patients. *Brain Research, 1084*, 123–131.

TINDALE, A. R., KAPUR, A., TZANETAKIS, G., DRIESSEN, P., & SCHLOSS, A. (2005). A comparison of sensor strategies for capturing percussive gestures. In S. Fels (Ed.), *Proceedings of the 2005 Conference on New Interfaces for Musical Expression* (pp. 200–203). Vancouver, Canada: NIME.

VAN DER STEEN, M. M., MOLENDIJK, E., ALTENMÜLLER, E., & FURUYA, S. (2014). Expert pianists do not listen: the expertise-dependent influence of temporal perturbation on the production of sequential movements. *Neuroscience, 269*, 290–298.

VROOMEN, J., & KEETELS, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, and Psychophysics, 72*, 871–884.

WESSEL, D., & WRIGHT, M. (2002). Problems and prospects for intimate musical control of computers. *Computer Music Journal, 26*(3), 11–14.

WRIGHT, M., CASSIDY, R. J., & ZBYSZYNSKI, M. (2004). Audio and gesture latency measurements on linux and osx. In M. Gurevich (Ed.), *International Computer Music Conference* (pp. 423–429), Miami, USA: ICMC.

XIA, H., JOTA, R., MCCANNY, B., YU, Z., FORLINES, C., SINGH, K., & WIGDOR, D. (2014). Zero-latency tapping: Using hover information to predict touch locations and eliminate touch-down latency. In M. Dontcheva & D. Wigdor (Eds.), *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (pp. 205–214). Honolulu, USA: ACM.

YATES, A. J. (1963). Recent empirical and theoretical approaches to the experimental manipulation of speech in normal subjects and in stammerers. *Behaviour Research and Therapy, 1*, 95–119.