Comparing the Timing of Movement Events for Air-Drumming Gestures

Luke Dahl

Composition and Computer Technologies, Department of Music, University of Virginia lukedahl@virginia.edu

Abstract. New air-instruments allow us to control sound by moving our bodies in space without manipulating a physical object. However when we want to trigger a discrete sound at a precise time, for example by making a drumming gesture, the timing feels wrong. This work aims to understand what aspects of a performer's movement correspond to their subjective sense of when the sound should occur. A study of air-drumming gestures was conducted, and the timing of eight movement events based on movements of the hand, wrist, elbow joint, and wrist joint are examined.

Keywords: Musical gesture, air-instruments, air-drumming, new interfaces for musical expression

1 Introduction

Traditional musical instruments are tools we manipulate with our bodies and hands in order to transform bodily energy into delightful sounds: we strike a drum, pluck a string, draw a bow, or blow into a flute. Even today, most of our new instruments require the musician's touch: we press a key, push a slider, turn a knob. However our physical connection with our instruments is becoming less tactile. When we drag a mouse or swipe on a touchscreen we touch a physical object, but we don't feel the moment our pointer contacts a virtual button or plucks a virtual string. Some new instruments, beginning with the Theremin, require no contact with the instrument itself and control sound by sensing our gestures "in the air." The recent affordability of motion sensing technologies has led to a proliferation of experimental "air-instruments" which enable a performer to control sound by moving their body in free space. We still move our bodies, yet touch is missing.

1.1 The Challenges of Designing Air-Instruments

I define *instrumental air-gestures* as purposeful movements of a performer's body in free space used to control a responsive sound-generating instrument. These can be divided into *continuous air gestures*, in which some aspect of movement (e.g. the height of the hand) is continuously mapped to some sonic variable (e.g.

a filter cutoff frequency), and *discrete air-gestures* which are meant to trigger a sonic event (e.g. a note onset) at a precise time.

In my experience [3] it is difficult to design an instrument that senses discrete air-gestures in a way that allows for complex rhythmic performance. To the performer the timing often feels wrong. To design such instruments we must create an algorithm that uses movement data to decide when to trigger a sonic event. I suspect most of us do this based on some heuristic, and it seems that we are choosing the wrong heuristic.

A Subjective Movement Event I suggest that when a performer makes a discrete air-gesture, such as striking a non-existent drum, they do *something* with their body to create an internal sensation of a moment in time, and they intend the resulting sound to occur at or near the moment of this event. I call this embodied sensation the *subjective movement event*. If instrument designers understood which aspects of the performer's movement correspond to the subjective movement event we would design our instruments to detect it.

1.2 Studying Discrete Air-Gestures

I conducted a study in which ten musicians were recorded performing air-drumming gestures in time to the sound of a simple rhythm. Their movements were analyzed to detect a number of *movement events*, which are hypothetical candidates for the subjective movement event. By looking at the timing of these movement events with respect to the sonic events to which they are intended to correspond, we can begin to understand the enacted and embodied sense of discrete moments in gestures. In [2] I presented the details of this study and an analysis of movement events based on tracking the position of the hand and wrist. In this paper I summarize that work and extend it to include an analysis of movement events based on the changing angles of the elbow and wrist joints.

1.3 Related Work

Previous related work includes performance systems which enable discrete air-gestures [9][5][3][6], studies of discrete musical air-gestures [8][1], a study of snare drumming movements [4], and studies of conducting gestures [7][11]. These employ a variety of techniques for detecting the end of a striking gesture or for triggering sounds from the movement of the performer's hand or of the tip of a drumstick or baton. These include:

- Detecting when the position passes through some spatial threshold [8][9].
- Detecting extrema in position, such as vertical minima [1][11].
- Detecting when the velocity surpasses some threshold [5][3].
- Detecting positive extrema (i.e. peaks) in the magnitude acceleration [7][11]
- Detecting when the acceleration surpasses some threshold [6].
- Detecting when the third derivative of position surpasses some threshold [4].

Interestingly all of these use spatial variables or their derivatives and none employ tracking the movement of the performer's joint angles.

Sensorimotor Synchronization Gesturing in time to music is form of *sensorimotor synchronization*, which has been the topic of research for decades [10]. One of the primary findings is that when tapping in time to an audible beat (usually a metronome click), most people tap before the beat by up to 100 msec. This shows that when a performer intends a movement event to correspond to a sonic event, they may not occur at the same time, *even though to the performer they seem to.* This suggests that the subjective movement event (the event we hope to detect) may not coincide with, and is likely to precede, the sonic event it is meant to trigger.

2 A Study of Air-Drumming Gestures

The goal of this research is to understand discrete air-gestures and their timing with respect to the intended sound. However it is difficult to study this situation directly. If we record a musician performing on an air-instrument, the resulting sound may not occur at the time they intended. I address this chicken-and-egg problem by substituting a proxy activity – i.e. performing drumming gestures in time to a pre-recorded sound – for the activity we wish to understand – i.e. performing drumming gestures in order to trigger sounds.

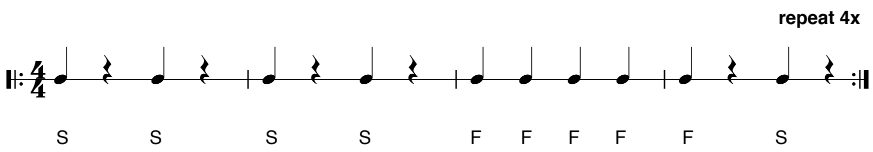

Fig. 1. The stimulus rhythm, played at 100 BPM. Slow notes are labeled ‘S’ and fast notes ‘F’

The study participants hear a recording of a drum playing the rhythm shown in figure 1, which has an equal number of “slow” and “fast” notes. Participants were instructed to gesture *as if they were performing* these sounds by striking an imaginary drum somewhere in front of them. The movements of the participant’s right hand, wrist, elbow, and shoulder were recorded with a motion capture system The drum rhythms were also recorded into the motion capture system to facilitate precise comparison of the timing of the movement and the sound.

2.1 Detecting Movement Events

In this paper I examine eight externally-detectable *movement events* which are hypothetical candidates for the subjective movement event. These are based on the movement of the performer’s hand and wrist in space, and on the changing angles of the performer's elbow and wrist joints. These movement events can be divided into *change of direction events* – moments when the body part or joint angle suddenly change direction – and *acceleration peak events*, which are moments of sudden deceleration of the body part or joint angle. The movement events and techniques for detecting them will now be described in detail.

Hand and Wrist Hits The *hand hit* is the moment at the end of the striking gesture when the performer's hand suddenly changes direction. If they were striking a physical drum, the hit would correspond to the moment when the hand (or drumstick) contacts the head of the drum, imparts energy to the instrument thus initiating the sound, and rebounds in the opposite direction.

For a gesture in free space where no physical contact occurs, how do we detect the end of the strike? If the movement were in one dimension (e.g. up and down) we could simply detect extrema (i.e. the moment of minimum height). Since the virtual drum's location and orientation were not precisely specified, participants' movements were not restricted to any particular plane. So I define the hand hit as the moment at the end of a striking gesture where the hand suddenly changes direction.

An algorithm for detecting sudden changes of direction was presented in [2]. In brief, the velocity vector of the hand is passed through two low-pass filters with different cutoff frequencies to create slow and fast estimates of the hand's velocity direction. When the hand changes direction the angle between these vectors increases. Hits are detected as peaks in the rate of change of this angle (see figure 2).

If the performer's wrist joint is not held still, the location of the hand and wrist may change direction at different times. I apply the same hit detection algorithm to the movement of the wrist (defined as the point half-way between the condyles at the distal ends of the radius and ulna) to detect the *wrist hit* events.

Acceleration Peaks of the Hand and Wrist Large peaks in the magnitude acceleration of the hand often occur close to the onset of the associated audio event. (In fact these peaks are decelerations as the participant sharply halts the movement of the hand.) For an unimpeded movement, an acceleration of is the result of a muscular force, and so an acceleration peak is a good candidate for the subjective movement event. An algorithm for detecting these *hand acceleration peaks* is described in [2], and a result is shown in figure 3. The same algorithm can be applied to the location of the wrist to detect *wrist acceleration peaks.*

Calculating Joint Angles The remaining four movement events are based on the angle of the performer's elbow and wrist joints. To calculate the elbow angle a line segment representing the upper arm is defined as passing from the top of the right shoulder, to a point half-way between the lateral and medial condyles of the elbow joint. A line segment representing the lower arm (or forearm) is

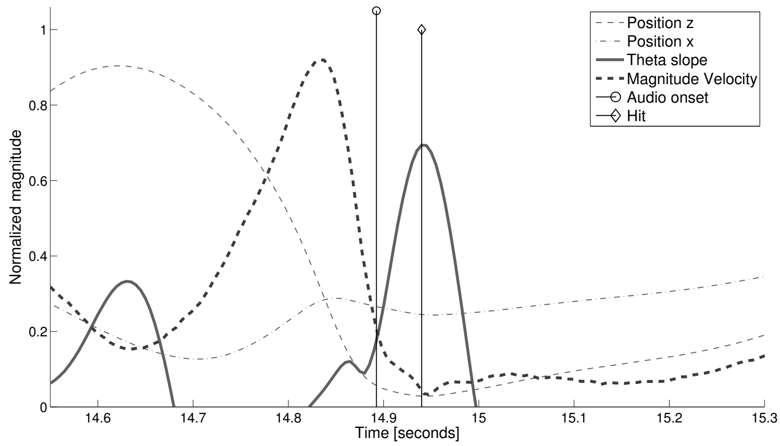

Fig. 2. Detecting the hand hit for an air-drumming gesture. Hits are detected as peaks in the rate of change (theta slope) of the angle between two estimates of the hand's velocity. In this case the hit corresponds to a minima in magnitude velocity and occurs after the onset of its associated audio event.

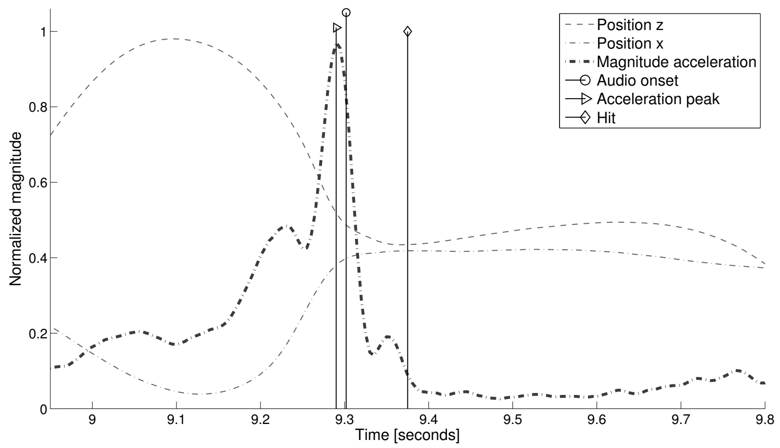

Fig. 3. Detecting the hand acceleration peak for a strike gesture.

defined as passing from the elbow to the wrist. The elbow angle is defined as the angle between the upper and lower arm segments. An angle of π radians represents a fully extended elbow joint.

Unlike the elbow, the wrist is a complex joint. It's movements can be defined in two dimensions. In flexion/extension the hand moves in the direction of the palm (flexion) or back of the hand (extension). In ulnar or radial deviation the hand moves in the direction of the little finger or thumb. I observed that for air-drumming gestures the ulnar/radial deviation has a small range, its measurements are noisy, and it tends to correlate with flexion/extension. Thus I used only the flexion/extension angle of the wrist. Two line segments are used to define the wrist angle: the lower arm segment described above, and a segment representing the hand which passes from the wrist to a the back of the hand at the base of the third finger. I define a plane whose normal passes from the ulnar side of the wrist to the radial side of the wrist. The lower arm and hand segments are projected onto this plane, and the wrist angle is taken as the angle between these two projections.

Peaks of the Elbow and Wrist Joint Angles In an air-drumming gesture, as the hand falls the elbow is extending. I define the moment near the end of the strike when the elbow changes direction from extending to flexing as the *elbow angle peak*. Similarly for most participants the wrist joint angle also changes direction near the end of the strike and this is the *wrist angle peak*.

Since the changing position of the hand and wrist take place in three-dimensional space, detecting changes-of-direction was not simple. However the angular movements of the elbow and wrist joints are, as I've defined them, one-dimensional. Therefore detecting the moment the joint changes direction is straight-forward, and is equivalent to finding peaks in the joint angle. To do so, I find all negatively-sloped zero-crossings in the angular velocity (which correspond to positive peaks in angle), and then discard those for which the joint angle is below a threshold. This ensures that only peaks near the end of the strike are detected.

Acceleration Peaks of the Elbow and Wrist Joint Angles The sudden changes in the angular velocity of the elbow and wrist joints at the end of the striking gesture are the result of torques applied to the joints. These torques are either the result of muscular effort, or they are due to the mechanical constraints of the joint itself. Figure 4 shows that both the wrist and elbow joints experience sharp negative peaks in angular acceleration before the corresponding joint angle peak. These are detected in a manner similar to that used for detecting angular peaks. First I locate positively-sloped zero-crossings in the angular jerk (the third derivative of joint angle), and then remove those for which the acceleration does not surpass a threshold. These events are then labelled as *elbow angle acceleration peaks*, and *wrist angle acceleration peaks*.

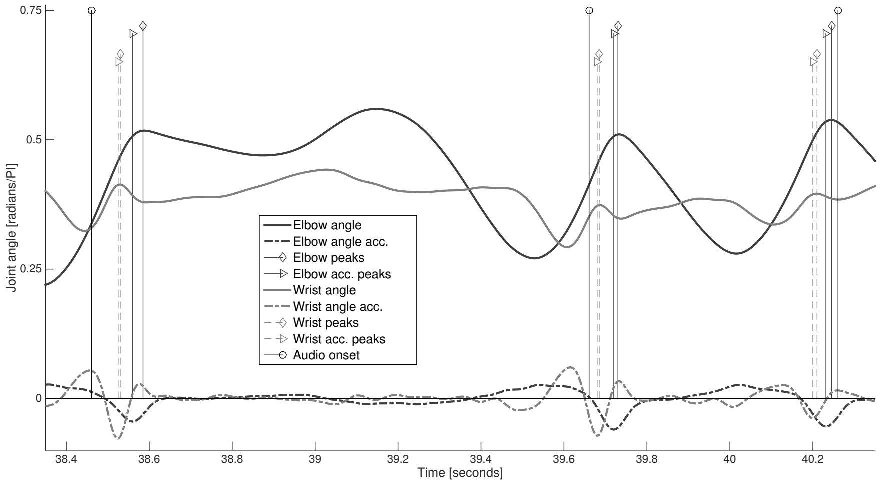

Fig. 4. Detecting joint angle peaks and angular acceleration peaks of the elbow and wrist joints for a single participant. One slow note and two fast notes are shown.

2.2 Summary of Movement Events

We now have eight movement events, summarized in table 1, which can be categorized with respect to three dimensions:

1. *Position vs. joint angle*: Four events are based on tracking the position of parts of the body, and four are based on the angles of joints.
2. *Body part*: Two events are based on the hand, two on the wrist, two on the elbow joint, and two on the wrist joint.
3. *Direction change vs. acceleration peak*: Four events detect changes of direction (in either position or joint angle), and four detect peaks in positional or angular acceleration.

Table 1. The movement events analyzed in this paper.

	Change of direction	*Acceleration peak*
Position	Hand hits	Hand acceleration peaks
	Wrist hits	Wrist acceleration peaks
Joint angle	Elbow angle peaks	Elbow angle acceleration peaks
	Wrist angle peaks	Wrist angle acceleration peaks

3 Analysis and Results

Each movement event functions as a hypothetical candidate for the subjective movement event (which is the internal sensation that the air-drummer intends to correspond to the sound). In order to evaluate a movement event I analyze its timing with respect to the sonic event to which it is meant to correspond. For each trial the audio recording is analyzed to detect the onset time of each note. This functions as the “ground truth” of when the performer intended the sound to occur. Each audio onset is paired with the movement event which occurred closest to it, for each type of movement event.

From the movement event time the associated audio onset time is subtracted to get the time offset (or asynchrony) between the audio event and the movement event. A negative offset means the movement event preceded the audio event, and a positive offset means it came after. All subsequent analysis is performed on these offset times.

3.1 Calculating Timing Statistics

For each participant the data from each trial were aggregated and then split into the slow note and fast note conditions. In order to reject bad data due to detector errors or participant mistakes, events whose offset is greater than half the time between notes were removed. Events which lay more than two standard deviations from the mean for each condition and participant were rejected as outliers.

We want to know whether the mean or standard deviation offset times differ between various conditions. For a given condition and movement event, the mean tells us by how much the event tends to precede or lag behind the audio onset, and the standard deviation can be considered a measure of the “noise” in the event detection or in the movement itself. To infer whether two conditions differ in the population, we compute the mean (or standard deviation) of each participant’s offset times for the conditions we wish to compare. Then a two-sided paired-sample t-test of the ten participants’ means (or standard deviations) for the two conditions is conducted.

3.2 Analysis of mean offsets

Figure 5 shows box plots for the eight movement events for both slow and fast notes.

Comparing change of direction events and acceleration peak events Examination of figure 5 shows that for each pair of movement events the acceleration peak event comes before the change of direction event. This appears to be true for both slow notes and fast notes, but the difference seems smaller for fast notes.

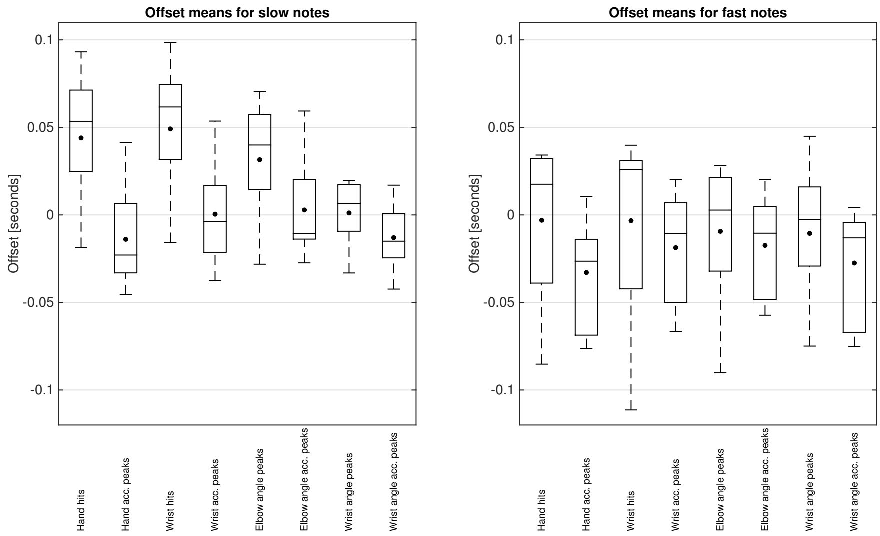

Fig. 5. Box plots of the mean offset times for all movement events. For each event, the line marks the median, the dot marks the mean, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data. Negative times indicate the movement event came before the audio onset.

A repeated measures two-way analysis of variance (ANOVA) was conducted on all offset means. The two within-subjects factors were movement event, with 7 degrees of freedom, and note speed with one degree of freedom. There were significant effects of both movement event ($F = 14.599, p = 3.9e - 11$) and note speed ($F = 8.426, p = 0.0175$). There was also a significant interaction between movement event and note speed ($F = 13.188, p = 2.5e - 10$). Post hoc comparisons were conducted between the four change of direction events and their associated acceleration peak events for both slow and fast notes using paired samples T-tests. Results are shown in table 2.

Table 2. The results of T-tests comparing the change-of-direction event with the acceleration peak event for each body location or joint angle. The 95% confidence interval is in milliseconds.

Feature	*Result*	*T(9)*	*P*	*95% CI*
	For slow notes			
Hand position	Acceleration peaks precede hits	4.884	0.0009	31 to 85
Wrist position	Acceleration peaks precede hits	4.1889	0.0023	22.4 to 75
Elbow joint angle	Acceleration peaks precede peaks	3.3777	0.0082	9.5 to 47.9
Wrist joint angle	Acceleration peaks precede peaks	3.1416	0.0119	4 to 24.3
	For fast notes			
Hand position.	Acceleration peaks precede hits	4.5294	0.0014	15 to 44.8
Wrist position	No difference found		n.s.	
Elbow joint angle	No difference found		n.s.	
Wrist joint angle	Acceleration peaks precede peaks	2.5824	0.0119	2.1 to 31.8

If we apply the conservative Bonferroni correction to these eight results we find that the joint angle results do not pass significance. Nevertheless, the trend which we observed is, for the most part, true: acceleration peak events precede their associated change of direction event, and the difference is less pronounced for fast notes and for joint angle events.

Comparing fast and slow notes Fast notes appear to occur earlier than slow notes for all movement events. This difference appears to be smaller for events based on acceleration peaks than it is for the change of direction events. To test these observations the mean offsets of slow and fast notes for all eight events are compared and the results shown in table 3

In general we find that the timing of change of direction events changes significantly between slow and fast notes, whereas acceleration peak events do not differ with note speed. The exception to this rule is the wrist angle. Like the other acceleration peak features, wrist angle acceleration peaks are not found to differ with note speed. Unlike the other change-of-direction events, wrist angle peaks are not found to be different between fast and slow notes. These results are visualized in figure 6.

Table 3. The results of T-tests comparing the offset time of fast notes with that of slow notes for each movement event. The 95% confidence interval is in milliseconds.

Movement event	*Result*	*T(9)*	*P*	*95% CI*
Hand hits	Fast notes are earlier than slow	4.1889	0.0023	22.4 to 75
Hand acceleration peaks	No difference found		n.s.	
Wrist hits	Fast notes are earlier than slow	5.2030	$5.6e-04$	9.5 to 47.9
Wrist acceleration peaks	No difference found		n.s.	
Elbow angle peaks	Fast notes are earlier than slow	4.1615	0.0024	18.7 to 63.1
Elbow angle acceleration peaks	No difference found		n.s.	
Wrist angle peaks	No difference found		n.s.	
Wrist angle acceleration peaks	No difference found		n.s.	

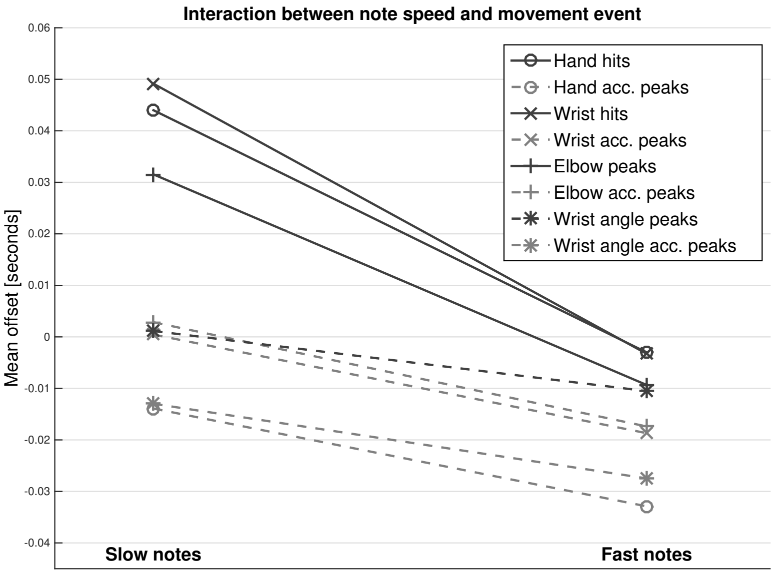

Fig. 6. Comparing fast and slow notes for all movement events. Solid lines are significant differences, and dashed are not significant. Dark lines are change of direction events, and light lines are acceleration peak events.

Linked movement events Table 4 contains the the means across the population of both the mean offsets and the offset standard deviations, and figure 6 visualizes the same data for mean offsets. These yield the following observations.

The acceleration peaks of the hand, and the angular acceleration peaks of the wrist joint occur earliest, and appear to be nearly identical for both slow notes and fast notes. T-tests comparing these events find no significant difference for either note speed. The movement of the hand is a result rotations in the shoulder, elbow, and wrist joints, and this result suggests that the last joint in the chain contributes most to the acceleration of the hand.

The movement of the wrist location is primarily a result of the elbow angle, and to a lesser degree the shoulder and trunk movements. Since it is earlier in the kinematic chain, the wrist location is not affected by movements in the wrist joint. Thus it is not surprising to see that for acceleration peaks, the wrist location and elbow joint angle timing appear to be nearly identical for both slow and fast notes. T-tests find no significant differences between these features for either note speed.

Table 4. The means of all offset means and of all offset standard deviations for each movement event. All times are in milliseconds.

	Offset mean		*Offset standard deviation*	
Movement event	Slow notes	Fast notes	Slow notes	Fast notes
Hand hits	43.0	-3.03	35.94	33.97
Hand acceleration peaks	-13.92	-32.9	31.34	25.69
Wrist hits	49.14	-3.29	40.86	32.33
Wrist acceleration peaks	0.45	-18.7	32.26	28.06
Elbow angle peaks	31.54	-9.35	36.95	29.05
Elbow angle acceleration peaks	2.84	-17.4	29.68	24.82
Wrist angle peaks	1.14	-10.53	38.86	34.28
Wrist angle acceleration peaks	-12.99	-27.48	35.46	32.44

3.3 Analysis of Noise

There are two possible sources of variability in the timing of movement events. One is due to people's inability to execute their movements with precise timing. The other is the noise or inaccuracy in the algorithms for detecting the movement events. Events which can be performed and detected more reliably will be more successful at generating the sound at the time intended by the performer.

Figure 7 shows box plots of the standard deviations for each event for both slow and fast notes. We notice that for each pair of movement event, the mean standard deviation for the acceleration peak event appears to be lower than that for the associated change of direction event. A repeated measures two-way ANOVA on the measured standard deviations finds a significant effect for movement event ($F = 3.541, p = 0.0029$), but none for note speed. Table 5

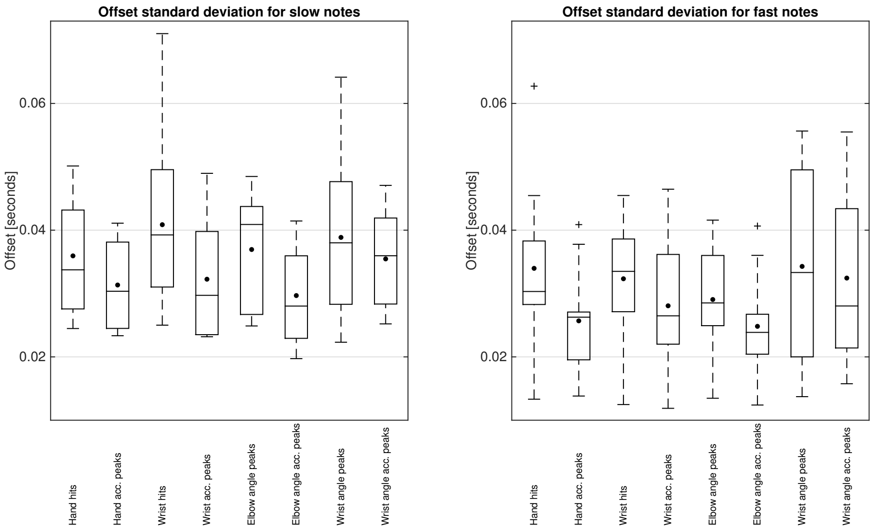

Fig. 7. Box plots of standard deviations for all movement events.

Table 5. The results of T-tests comparing the standard deviations of the change-of-direction (c.o.d.) events with the those for the acceleration peak events.

Movement event	*Result*	*T(9)*	*P*	*95% CI*
	For slow notes			
Hand position	Acc. peaks have lower noise than c.o.d.	4.5366	0.0014	1.4 to 7.8
Wrist position	Acc. peaks have lower noise than c.o.d.	3.8552	0.0039	3.6 to 13.7
Elbow joint angle	Acc. peaks have lower noise than c.o.d.	2.8488	0.0191	1.5 to 13
Wrist joint angle	No difference found		n.s.	
	For fast notes			
Hand position	Acc. peaks have lower noise than c.o.d.	2.4022	0.0398	0.5 to 16.1
Wrist position	Acc. peaks have lower noise than c.o.d.	2.5487	0.0313	0.5 to 8.1
Elbow joint angle	Acc. peaks have lower noise than c.o.d.	2.7508	0.0224	0.8 to 7.7
Wrist joint angle	No difference found		n.s.	

shows the results of post hoc T-tests comparing the change of direction and acceleration peak events.

If we apply Bonferroni correction, only the first two results would be significant. While it is not a strong result, it does seem that acceleration peaks events have less noise than change of direction events. As was found to be the case with offset means, the wrist angle events are different: the standard deviations of wrist angle peaks and wrist angle acceleration peaks were not found to differ for either note speed. For both slow and fast notes, the two events with the lowest mean standard deviations are hand acceleration peaks and elbow angle acceleration peaks.

4 Discussion

4.1 Acceleration peaks vs. change of direction

If a digital instrument builder wants to design a system to trigger sounds with air-drumming gestures which movement event should they use? There are a number of reasons why events based on acceleration peak are better.

Acceleration peaks occur earlier A real-time air-instrument needs to predict when an audio event should be triggered before it occurs, ideally early enough to account for any latencies in the system. The acceleration peaks for hand location, wrist location, and elbow angle occur before the change-of-direction event for the same body part. On average the acceleration peaks of the hand location occur earlier than any other movement event, with wrist angle acceleration peaks only slightly later.

Acceleration peaks have less noise For hand location, wrist location, and elbow angle, the acceleration peak events had less timing variability than their associated change-of-direction events. Hand acceleration peaks and elbow angle acceleration peaks had the lowest noise.

At first this seems surprising. The time at which the body part changes direction is a result of the accelerations applied to it. Shouldn't they have the same noise characteristics? The difference may simply be a result of the physics: acceleration is integrated twice to give position, so small differences in acceleration result in larger differences in position. Furthermore, I detect only the time of the peak acceleration and do not measure its shape. Two movements with identical peak acceleration time could have different acceleration curves, leading to different change of direction times. Regardless of the reason, the fact that acceleration peaks have less noise makes them better suited for triggering sounds from discrete air-gestures.

Acceleration peaks change less with note speed An ideal movement event for triggering sounds would not change timing when the speed of repetition changes. Using a movement feature which does vary with speed would require the system to infer before a gesture occurs what metric level the note is intended to operate on.

We saw that for hand location, wrist location, and elbow angle, the timing of the change-of-direction event changed significantly between fast and slow notes, with slow notes occurring much later. However for acceleration peak events any differences between fast and slow notes was not significant.

4.2 How the wrist joint is different

The wrist joint seems to behave differently than the elbow joint. For the elbow joint, the timing of peaks changes with note repetition speed but the timing of acceleration peaks do not. For the wrist joint, neither the peaks nor the acceleration peaks change with note speed. This may indicate that when we air-drum we treat the elbow and wrist joints differently. One possible difference is that the movement in the elbow joint is primarily intentionally directed, whereas the movement in the wrist is more passive. That is, we extend our elbow joint by activating the triceps, and then we decelerate that extension through a sudden deliberate activation of the biceps. However, for the wrist joint, the muscles of the forearm may be used to maintain a constant force on the wrist joint. When the elbow angle is decelerated the wrist is pulled back, but the hand is carried forward by its inertia, causing the wrist angle to change. This movement in the wrist is eventually impeded, either by the muscular tension on the joint or by the joint reaching the limits of it's range of motion.

I might call this the “flinging hypothesis,” because the hand is passively flung by the active elbow joint. That the elbow angle acceleration peaks have the least noise somewhat supports this hypothesis. If the timing of sudden decelerations of the elbow joint is what people are controlling for, it seems plausible that these would have the least variability.

Other findings from this study do not support the flinging hypothesis. In figure 4 we see that for this participant, the wrist angle peaks before the elbow angle does, and that acceleration peaks in the wrist angle also occur before acceleration peaks in the elbow angle. Table 4 confirms that this is also true on average across the population. It is interesting to note that the accelerations of the wrist joint seem to be sharper than those of the elbow (again see figure 4).

4.3 Conclusions

This work examined the timing of eight movement events for triggering sounds from air-drumming gestures, and shows that detecting sharp accelerations in the movement perform better than detecting changes of direction at the end of a gesture. This suggests that the subjective movement event – the thing we do with our bodies to create a sense of a discrete moment, and to which we intend the sound to correspond – is likely to be a sudden muscular effort.

Instead of relying on intuitively designed detection algorithms to decide when to trigger a sound from a discrete air-gesture, we can use these results to inform our design, which will hopefully result in more responsive and rhythmically accurate air-instruments. Further work implementing and evaluating real-time air-instruments will help to verify and refine these findings.

These results are useful not only for non-tactile air-instruments but for any movement-controlled musical instrument, such as traditional instruments augmented with inertial sensors. They may also have application improving the timing and user experience of any gesture controlled systems, such as those used in video games or home entertainment systems.

Acknowledgments. This research was performed as part of my PhD thesis at CCRMA, Stanford University

References

1. M. Collicutt, C. Casciato, and M. M. Wanderley. From real to virtual: A comparison of input devices for percussion tasks. In *Proceedings of NIME*, pages 4–6, 2009.
2. L. Dahl. Studying the timing of discrete musical air gestures. *Computer Music Journal*, 39(2):47–66, 2015.
3. L. Dahl and G. Wang. SoundBounce: Physical Metaphors in Designing Mobile Music Performance. In *Proceedings of the 2010 conference on New Interfaces for Musical Expression*, pages 178–181, Sydney, Australia, 2010.
4. S. Dahl. Playing the accent-comparing striking velocity and timing in an ostinato rhythm performed by four drummers. *Acta Acustica united with Acustica*, 90(4):762–776, 2004.
5. C. Havel and M. Desainte-Catherine. Modeling an air percussion for composition and performance. In *Proceedings of the 2004 conference on New interfaces for musical expression*, pages 31–34. National University of Singapore, 2004.
6. H. Kanke, Y. Takegawa, T. Terada, and M. Tsukamoto. Airstic drum: a drumstick for integration of real and virtual drums. In *Advances in Computer Entertainment*, pages 57–69. Springer, 2012.
7. G. Luck and P. Toiviainen. Ensemble musicians' synchronization with conductors' gestures: An automated feature-extraction analysis. *Music Perception*, 24(2):189–200, 2006.
8. T. Mäki-Patola. User interface comparison for virtual drums. In *Proceedings of the 2005 conference on New interfaces for musical expression*, pages 144–147. National University of Singapore, 2005.
9. M. V. Mathews. Three dimensional baton and gesture sensor, Dec. 25 1990. US Patent 4,980,519.
10. B. H. Repp. Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, 12(6):969–992, 2005.
11. Á. Sarasúa and E. Guaus. Beat tracking from conducting gestural data: a multi-subject study. In *Proceedings of the 2014 International Workshop on Movement and Computing*, page 118. ACM, 2014.